Interactive Visualization for Diagnosis of Industrial Model Predictive Controllers with Steady-State Optimizers

Shams Elnawawi

Department of Chemical and Biological Engineering, University of British Columbia, BC, Canada

Lim C. Siang

Department of Process Control Engineering, Burnaby Refinery, BC, Canada

Daniel L. O'Connor

Control Consulting Inc., Great Falls, MT, United States

R. Bhushan Gopaluni

Department of Chemical and Biological Engineering, University of British Columbia, BC, Canada

Abstract

Model Predictive Controllers (MPCs) are widely used in the process industries and are typically implemented with an integrated Linear Program (LP) optimizer in the form of two-stage LP-MPC systems. Despite significant control-theoretic advances in MPC design and performance evaluation in academia, there is still a gap in addressing operational issues in real-world MPC controllers. In particular, engineers and operators responsible for sustaining MPCs often need to interpret the LP solution to understand the controller's actions. Without easy interpretability, it is difficult to troubleshoot MPCs especially for large-dimensional controllers. To alleviate this difficulty, a systematic approach that facilitates LP solution diagnostics using tools from data visualization and process control is developed. The 'partial pivoting' operation - an industrial practice that has seen limited exposure in academic literature - is discussed in detail with regards to its role in LP solution diagnosis. Typical workflows for diagnosing problematic controllers are used in conjunction with data visualization principles to guide the design of new tools focused on visualizing variable constraint data that facilitate the diagnosis process. These proposed tools are designed using Munzner's "Nested Model" as a guiding framework for visualization design and evaluation. The use of these tools is demonstrated in multiple industrial examples, with comparison to current industrial methodologies.

Keywords: Data visualization, Model predictive control, Performance monitoring, Linear Programming

1. Introduction

Model Predictive Control (MPC) is an advanced process control strategy that is widely used throughout the refining and petrochemical industries [1]. Commercial MPC packages that are typically implemented in the refining and petrochemicals industries separate the MPC algorithm into two stages, comprising of a (1) steady-state target optimization stage, which uses a linear program (LP) to calculate setpoint targets, and a (2) dynamic optimization stage that receives those targets and computes control actions [2, 3]. It is important to note that the structure of this industrial two-stage MPC controller differs significantly from the typical theoretic MPC formulation found in academic literature [4, 5]. This two-stage structure has been referred to by various names, including *double-layer MPC (DLMPC)* [5], *DMC-type* or *LP-DMC controllers* [4], originating from the Dynamic Matrix Control formulation devised in the early 1980s [6, 7], *DMC-LP controllers* [8], *two-stage MPC systems* [9] and *LP-MPC cascade control systems* [10].

This two-stage MPC structure is illustrated in Figure 1, where constraints and economics are used as inputs for the steady-state optimization layer, and process models along with tuning parameters are used in the MPC layer below. For the purposes of clarity and consistency in this paper, we will refer to these two-stage LP-MPC systems as simply MPC, and explicitly refer to the steady-state optimization layer or dynamic optimization layer when additional context is required. An excellent, broad review of industrial MPC packages can be found elsewhere in the literature, covering the history of industrial MPC, implementation details and a comparison of different technologies [1, 11], as well as a comprehensive overview of current practices, challenges and opportunities in the industrial MPC landscape [12, 13].

Despite significant control-theoretic and algorithmic developments in MPC, there has been a lack of development in the tools used to visualize MPC in action [13]. Popular commercial MPC packages such as AspenTech's DMCPlus or Honey-

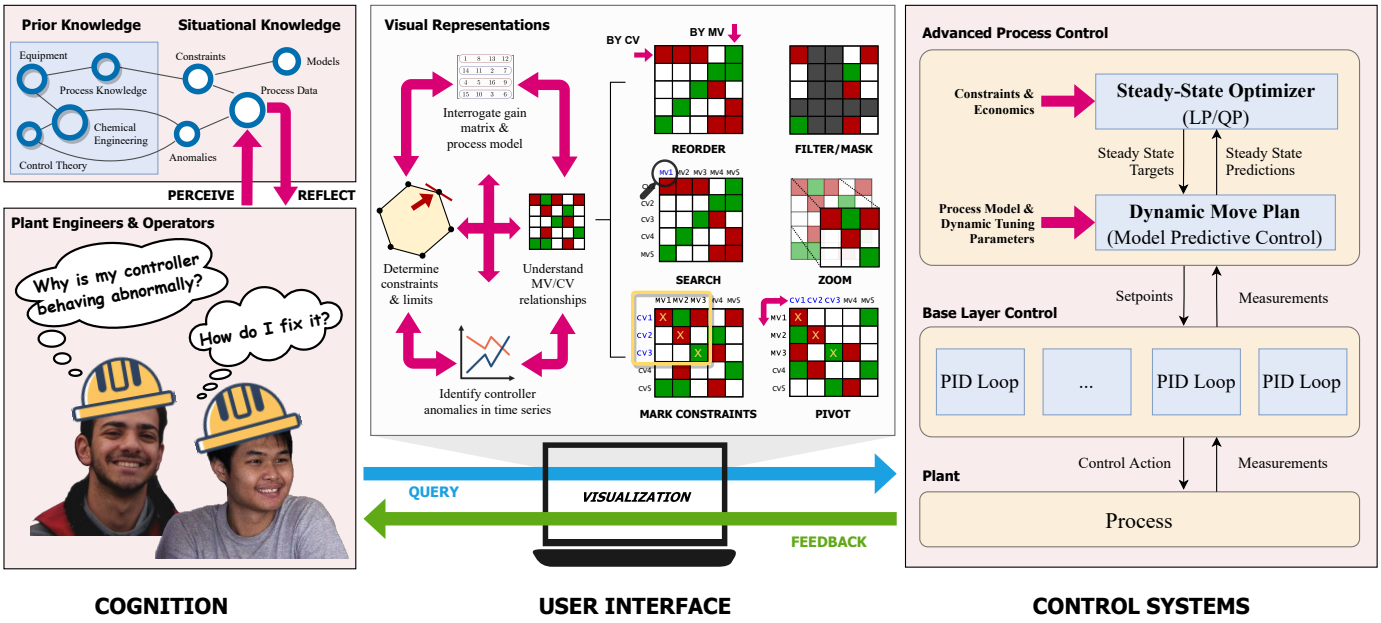

Figure 1: Essential components in the design of visualization tools for MPC systems. Engineers and operators often need to interact with the control system through a user interface to diagnose faults and abnormal controller behaviour. MPC systems store and generate large amounts of data in various forms, including process models and gain matrices, time series process data, and optimizer solutions. MPC users must successfully navigate through these complex datasets to obtain situational knowledge of the controller. To effectively determine the root causes of any undesirable controller behaviour, users must contextualize and augment situational knowledge about the controller's state with prior engineering knowledge and process understanding. The ability to rationalize complex controller actions is important to determine the solution required to restore the controller to its normal behaviour. Visual representations of the control system can facilitate this process if they are designed in a way to optimize the display of controller information.

well's RMPCT [14], often use static, tabular views and message logs to convey controller information to engineers and operators, which are often counter-intuitive and cumbersome to navigate, even for experienced users.

Industrial MPC applications often involve more than 50 process variables [15]; for instance, the MPC used to control the Fluid Catalytic Cracking (FCC) unit at the Burnaby Refinery in British Columbia, Canada has 44 manipulated and feedforward variables (MVs and FFs) and 64 controlled variables (CVs) configured in its model. For an FCC process and other complex industrial chemical processes, the counter-intuitive phenomena of a large number of CVs required for control relative to available MVs handles is well-documented in the literature [16, 17, 18, 19]. Interested readers are invited to consult a series of excellent papers by Arbel and colleagues on FCC design and control for further details [20, 21].

These large MPC datasets make navigation through tabular views extremely tedious for plant engineers and operators. Traditional MPC visualization techniques like static tables and univariate time series trends are ineffective for visualizing large MPC systems, as the large number of variables involved can be overwhelming for users, and the plots can be extremely tedious to create and use.

In LP-MPC systems, the steady-state LP optimizer is responsible for calculating a set of feasible targets using the steady-state gain matrix, while respecting constraints imposed on MVs and CVs. The LP also determines the degrees of freedom that are available for control, based on the number of unconstrained MVs, and it uses these available MVs to push CVs to their economically optimum limits. Thus, a strong understanding of the underlying steady-state gain matrix and resulting LP solution is essential to diagnose MPC issues effectively.

Despite the importance of the LP optimizer and gain matrix, effective visualization tools for the steady-state LP optimizer in the context of process control have not received much academic attention. A literature search for the term "*model predictive control*" revealed over 20,000 journal articles published over the past four decades. The majority of these articles describe successful applications of MPC on various processes or discuss control-theoretic notions such as controllability, stability, robustness or performance [4]. However, MPC papers specifically targeting industry practitioners and discussing day-to-day MPC operational issues faced by plant engineers and operators are rare.

Troubleshooting an MPC system — also known as 'controller diagnosis/diagnostics' and 'fault detection and diagnosis', refers to the analysis of process or equipment faults that can cause further issues to product quality or throughput, and the monitoring of process variables for the identification of their causes [22]. Establishing an effective mental model of the MPC troubleshooting process is a non-trivial task, even for experienced engineers and operators. Successful controller troubleshooting efforts involves a series of complex, iterative tasks, including observing undesirable controller behaviour in time series data, understanding the state of the controller, constraints and process conditions, and identifying clamped or problematic variables using a combination of controller data, process models and process knowledge. Exploratory data analysis (EDA)

and data visualization is often the first step in data analytics, but can be challenging for practitioners to set up due to the aforementioned complexities of the underlying MPC data. The complexity of the LP-MPC system necessitates the development novel visualization tools, as these large-dimensional LP solutions that cannot be effectively visualized using traditional tabular methods.

The challenges outlined by Guerlain and colleagues [15] on human factors are still prevalent in many industrial MPC systems, even though two decades have passed since its publication. An anonymous comment left on an Emerson blog article [23] provides a succinct description of real-world MPC issues:

"MPC math is simple and elegant; MPC engineering is not."

The goal of this paper is therefore to explore common human factors issues with industrial MPCs and to use them to guide the design of new visualization tools, with emphasis on interactivity and usability in the process of controller diagnosis.

This paper is structured as follows: Section 2 provides an overview of MPCs with integrated LP optimization, discussing the structure and mechanics of the LP-MPC system as well as MPC conditioning. Section 3 provides a review of previous work in this area, outlining the major existing implementations of MPC visualization tools as well as the relevant MPC analysis tools. The analysis techniques covered in this section are essential to operations presented later in this paper, especially the walkthrough of the 'partial pivoting' operation and its connection to the LP solution process. Section 4 provides the fundamentals of data visualization in how it relates to visual perception, so as to provide guidelines by which we justify our design of visual tools. Section 5 outlines each of the visualization plots used for visualizing MPC and their LP optimizers, with discussions of how these plots are used in practice. Visual tools that have seen limited implementation in industry are introduced and their associated workflows are discussed with regards to how they facilitate controller diagnosis. Section 6 provides case studies used to test the concepts designed in section 5 with real industrial examples, showing the typical workflow of engineers in the field. Finally, Section 7 discusses the limitations associated with this research from both conceptual and practical standpoints, and suggests future research directions.

2. Overview of MPCs with integrated LP optimizers

At the base level of industrial control systems, regulatory PID loops maintain setpoint tracking of flow rates, temperatures, pressures, and other variables. Advanced Process Control (APC) systems operate above the regulatory control layer in a multivariable manner, acting as a "supervisory" layer over the base control system. The APC layer can either provide setpoints for PID control loops, or it can directly manipulate actuator action [24].

MPCs in the refining and petrochemical industries are typically implemented as a two-stage system with an integrated LP optimizer [4, 9, 10]. In the first stage, the LP optimizer uses economic LP costs and variable operating limits to generate a set of economically optimal targets for the MVs, while respecting constraints imposed on each variable. In the second stage, the dynamic optimizer takes the LP targets and computes a move plan for all manipulated variables. The move plan is a balancing act between minimizing control error and avoiding aggressive MV moves, and implementing the change dictated by the LP, in order to maintain process stability while reaching the economic optimum. The dynamic optimizer comprises the actual MPC algorithm, where the optimizer uses the process model and current state to predict dynamics over the prediction horizon and implement its calculated move plan.

To appreciate the complexity of the LP solution space for large-dimensional systems, we provide a high-level overview of a general LP-MPC system. Given a set of MVs and CVs, at each controller execution cycle, the LP optimizer finds a set of MV moves that minimizes costs using its objective function, subject to limits imposed on each variable. LP costs are assigned to MVs based on the economics for each control handle. CV changes are calculated as linear combinations of MV changes using its steady state gains, where $k_{i,j}$ refers to the gain for the i-th MV and j-th CV as described in Equation 1, which has a simplified notation that is commonly adopted by other industrial practitioners [4, 7], instead of typical conventions in academic literature. Detailed mathematical formulations of these LP-MPC systems can be found elsewhere in the literature [9, 10].

$$\begin{aligned} \min \quad & f(\Delta\mathbf{MV}) = \sum_i \mathrm{LPCost}_i \cdot \Delta\mathbf{MV}_i \\ \text{s.t.} \quad & \Delta\mathbf{CV}_j = \sum_i k_{i,j} \cdot \Delta\mathbf{MV}_i, \\ & \Delta\mathbf{CV}_j^{\min} \le \Delta\mathbf{CV}_j \le \Delta\mathbf{CV}_j^{\max}, \\ & \Delta\mathbf{MV}_i^{\min} \le \Delta\mathbf{MV}_i \le \Delta\mathbf{MV}_i^{\max} \end{aligned} \tag{1}$$

The LP solution space can be represented using a 2D plot for a system with 2 MVs. A sample representation of a 2×2 MIMO system is illustrated in Figure 2. Isoprofit lines show the objective function value across MV changes, and the vertex of the feasible region with the largest profit value is chosen as the solution. Note that this is the case when the objective function calculates profit; if the objective function calculates cost, the LP solution will have the lowest objective function value in the feasible region. Often the LP optimizer will attempt to maximize variables with a negative LP cost whenever possible, and those with a positive cost tend to be minimized. However, each variable's direction is mathematically determined by the gradient of the objective function hyperplane and the constraints imposed by the operating limits. As MV LP costs change, the gradient of the isoprofit line changes accordingly, so variables with negative LP costs may end up being minimized depending on how those constraints fit into the larger LP solution.

Moving operator limits may change the degrees of freedom available to the controller, which will affect the solution chosen by the LP. Changing limits in this way corresponds to moving the bounds of the feasible region in Figure 2. Additionally, LP costs need to be set carefully to ensure that variables will be

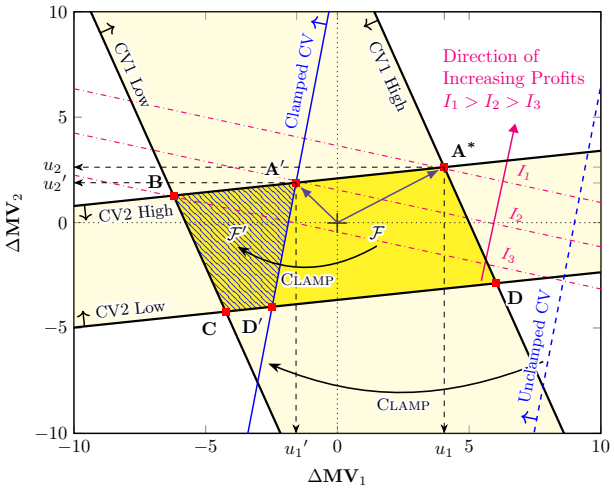

Figure 2: The feasible region, $\mathcal{F}$, of this toy LP consisting of 2 MVs and 2 CVs is highlighted in dark yellow. The red dots ($\mathbf{A}^*$, **B**, **C** and **D**) at the vertices of $\mathcal{F}$ represent feasible solutions. The lighter yellow regions indicate operating areas within the corresponding CV limits. The pink 'isoprofit' dash-dotted lines represent a constant value of the objective function for profit maximization, where $I_1 > I_2 > I_3$. The LP solution at any given time will be at the vertex that yields the largest isoprofit value. For $\mathcal{F}$, this would be $\mathbf{A}^*$ at I_1. Hypothetically, we can introduce a third CV constraint represented by the blue dashed line. Clamping the new CV limit changes the feasible region from $\mathcal{F}$ to the shaded $\mathcal{F}'$ with a new set of feasible solutions at $\mathbf{A}'$ and $\mathbf{D}'$. The new optimal solution is now at $\mathbf{A}'$ with an isoprofit line at I_2. Note that in $\mathcal{F}'$, the optimum move for $\Delta\mathbf{MV}_1$ has flipped signs and is moving in the other direction as indicated by the dark blue arrows extending from the origin. This new move is denoted as u_1', which may be physically undesirable, i.e. feed rate reductions or higher steam usage.

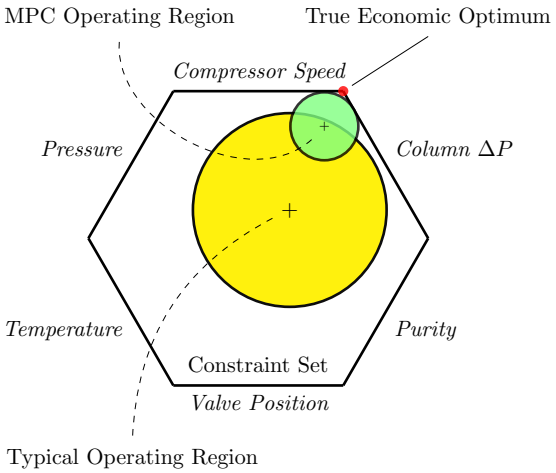

Figure 3: Stable operating region for a plant subjected to constraints, represented by a 2-dimensional polygon for illustrative purposes. The size of each operating region for human operators (yellow) and MPC systems (green) is indicative of the process variability during operation. The vertices represent possible feasible solutions in the LP formulation. The ideal operating conditioning would be persistent operation at the true economic optimum point (red) with minimum variability. Figure adapted from Brooks (2017) [26].

driven in the right direction. Sorensen et al. (1998) outline that the LP cost is calculated as a partial derivative of plant profit to a unit MV move, and that LP costs need to be up-to-date with any material balance changes in order to reflect accurate process economics [7]. As described in the figure, 'clamping' operating limits incorrectly may result in undesirable LP solution behaviour.

Generally speaking, a control system may contain any number of MVs and CVs. However, the LP formulates a solution for an equal number of MVs and CVs, as it is solving a system of simultaneous linear equations. In other words, the LP solves a 'square' problem with an equal number of MVs and CVs. Consider the example system in Figure 2: there are two MVs represented by the x- and y-axis, and two CVs represented by the two sets of constraint boundaries. Adding one more CV will add another set of constraint boundaries; if these constraint lines are within the current bounds of $\mathcal{F}$, there will be a new feasible region bounded by all three CVs, and the new operating point will be at the intersection of two CV constraint lines. This change of operating point is illustrated by the solid blue line in figure 2, which changes the feasible region to $\mathcal{F}'$. Each of the new solution points will leave one out of the three CVs unconstrained; $\mathbf{A}'$ will leave **CV1** unconstrained, $\mathbf{B}'$ and **C** leave **CV3** unconstrained, and $\mathbf{D}'$ leaves **CV1** unconstrained. Conversely, if the new CV's constraints are outside the bounds of $\mathcal{F}$ as in the dashed blue line in Figure 2, then the LP solution will remain at one of the current vertices of $\mathcal{F}$. The set of variables participating in the LP solution is denoted as the 'active constraint set' and will affect other variable relationships - this concept is discussed later in this section.

As discussed in Section 1, MPC models in the refining industry typically have more CVs than MVs, which frees the LP to move between various operating points, subject to physical limitations of the plant. MV limits are "hard constraints" that will not be violated by the LP, while CV limits are "soft constraints", because the CV setpoints cannot be set directly, and they are controlled as a result of the MVs moving. This form of control, where CVs are maintained between minimum and maximum desired values, is sometimes referred to as 'zone control' [5, 25]. If the MV movements are not be able to keep the CVs within limits, the LP solution becomes infeasible. In industrial MPC packages, CV limits can typically be violated in the event of an infeasible solution, and the specific CVs that the LP 'gives up' are determined by each CV's importance ranking [11]. In the event of an infeasibility, the LP gives up control of lower-ranked variables in favour of controlling more important ones by iteratively removing them from the solution and re-solving until the solution becomes feasible. Figure 3 illustrates the mobility of the LP solution within its feasible region using a more complicated system than that of Figure 2.

Figure 3 shows a single solution that the LP is trying to reach. Generally these constraints are at physical operating limits, and it is very difficult for a human operator to run the plant simultaneously at 5 or 10 or more limits. The LP solution always lies at variable constraints, since the hyperplane defining its objective function is linear [7]. Because of process condition changes and disturbances (feed composition changes, ambient temperature changes, rate changes etc.), the economic optimum may change rapidly, and a well-designed MPC controller can use many MVs to continuously drive the plant against the most

profitable constraints mathematically and keeping it running stably at the limits. Human operators, with the help of a good, well-tuned regulatory control layer of PID loops, will typically operate the plant conservatively with a bigger safety margin to maintain stable operations, but may have challenges running at all economic limits simultaneously, as illustrated by the circles in Figure 3. APC systems can help reduce this variability and simultaneously push the system towards the true economic optimum. As described by Maciejowski (2002), the culture of control practice in the process industries is one where human operators are entrusted with significant autonomy over the operations and control of the plant [2]. In consideration of this culture, industrial MPCs are typically implemented in a supervisory capacity over a PID control layer, giving plant operators the freedom to turn the APC layer off if it were to function erratically to maintain safe operations [2]. Therefore, a culture of trust between operators and automation systems is crucial. Tools and systems must demonstrate reliability improvements over manual operations, or risk operator distrust of automation systems [27], compromising the operational integrity of the plant.

During controller commissioning, engineers use prior process knowledge to define how each critical CV is controlled, and this is accomplished by pairing critical CVs to the primary MVs that control them in a one-to-one fashion [28]. The resulting set of MV-CV pairings is known as a set of 'pivot points' or as the 'active constraint set', and can be used to assist the diagnosis of problematic controllers, as it provides a benchmark that defines the 'normal' operation of a plant. For example, steam flow rate in the hot side of a heat exchanger (MV) would be used to control the temperature of the cold process fluid (CV). Since the LP optimizer is constraining the MPC, the active constraint set at any given time is indicative of how the LP is prioritizing variable limits, which provides important context to those troubleshooting problematic controllers.

Understanding the interplay between all of these factors is crucial to diagnosing a problematic controller. While some issues can be resolved by simply trending the necessary variables, some more complex issues may require examining the problematic variables in the larger context of LP optimization. The necessary functionality to solve such issues may be available in commercial MPC tools, but without an effective interface, users may still experience difficulty navigating through different screens to find the desired solution.

3. Literature review

This section outlines previous implementations of LP optimizer visualization in the context of APC, in order to help create guidelines that can be applied to creating new visual tools. Aside from the availability of built-in tools found in commercial MPC packages, there have been a lack of academic tools and research focusing specifically on the LP optimizer solution.

Guerlain et al. (2002) [15] proposed the MPC Elucidator: a MPC visualization tool that aimed to address issues with MPC interfaces through the use of 'Representation Aiding Strategies', a set of well-designed visual properties meant to help reduce the cognitive effort required by engineers and operators for situational awareness in the context of an APC system. Their design is guided by results of cognitive task and work analyses. However, discussions with industry partners and contacts revealed that the product was not implemented as widely as expected. Many issues raised by the Elucidator paper are still present today in many commercial packages, and there has been a lack of attention on the human factors aspect of MPC interfaces in the last two decades.

Notably, Guerlain and colleagues highlighted that some MPC variables may have their limits tightened (or 'clamped') due to instrumentation issues or local operating conditions. For large-scale MPC systems, operating conditions may take days or weeks to return to normal, in which time engineers or operators may inadvertently fail to unclamp those variables. If the typical limits are not reinstated when operating conditions return to normal, the clamped variables can limit the ability of the controller to control the process. Each incorrectly clamped variable can take up one degree of freedom from the controller, forcing it to drive the process towards the desired operating point using potentially undesirable moves - e.g., by changing feed rate or other economic handles in the process. It is often difficult to understand such undesirable MV moves as a consequence of the LP solution, without detailed understanding of the process and an investigation of process data, especially for large-dimensional controllers. This clamping effect is illustrated in Figure 2.

Peterson et al. (2011) [29] concur in their MPC solution analysis process patent that there is a demand for flexible, real-time tools that can analyze LP solutions to provide operators meaningful instructions on unclamping variables, which they've referred to as 'constraint set relief'. They developed a technique to address variable clamping, providing a quantitative method to inform operators on the variables and limits to adjust to shift the LP solution to the correct targets.

In their patent, they also highlight the importance of the 'partial pivoting' operation: a common industrial practice used to combine real-time constraint states with steady-state gain data to reveal how the LP will behave under constrained conditions. The key insight is that the size of the active constraint set is determined by the number of unconstrained MVs. The active constraint set is comprised of the unconstrained MVs (MV_u) that are available for control, and the constrained CVs (CV_c) that are being actively driven to a setpoint. Some other authors refer to these constrained variables as *'active'* or *'binding'*, and the unconstrained variables as *'inactive'*, *'non-binding'* [4, 30]. The constrained MVs that are at their limits are not available for control, while the unconstrained CVs are just floating between their limits due to MV movements, and the LP optimizer is not actively driving them to any setpoint. This partial pivoting operation is accomplished by rearranging the gain matrix to form a square submatrix of constrained MVs and CVs in the upper-left corner, then inverting only the square submatrix. The raw gain matrix describes the relationship between independents and dependents; the pivoted gain matrix describes that between *constrained* and *unconstrained* variables. Steady-state gains are calculated in open-loop, meaning that they do not show any in-

formation related to active constraints; the gain matrix therefore needs to be combined with constraint data in order to provide any explanation for LP behaviour.

The LP is highly dependent on the process gain matrix, since it is only concerned with steady-state variable changes. As described in Hoffman et al. (2010) [28], "the number of manipulated variables floating between their limits will be exactly equal to the number of controlled variables at their limits", since the LP optimizer will be limited by the number of available control handles in the system. This set of constrained CVs and unconstrained MVs is the basis with which we perform the partial pivoting operation, by moving this set to create a square submatrix at the top-left. This operation creates four submatrices, split by variables' constraint statuses. Each submatrix is then transformed to create its closed-loop counterpart; a summary of the operation is shown in Figure 4.

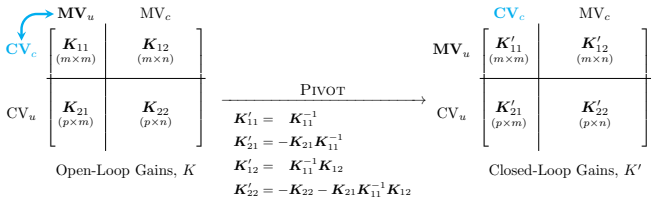

Figure 4: Pivoting operation based on a constraint set of (MV_u, CV_c). The open-loop gain matrix, K can be split into 4 submatrices $\boldsymbol{K}_{11}$, $\boldsymbol{K}_{12}$, $\boldsymbol{K}_{21}$ and $\boldsymbol{K}_{22}$. The constraint set in $\boldsymbol{K}_{11}$ will always form a square matrix of dimensions $m \times m$. Note that $\boldsymbol{K}_{12}, \boldsymbol{K}_{21}$, and $\boldsymbol{K}_{22}$ can be empty matrices with dimensions $p = 0$ or $n = 0$ if all MVs or all CVs are in the typical constraint set. Pivoting swaps the MVs and CVs to generate a closed-loop matrix, K' where the constrained CVs are the independent variables and the constrained MVs are the dependent variables.

By pivoting the open-loop gain matrix using the active constraint set, we obtain an 'effective' gain matrix that accounts for both steady-state changes and active constraints. Essentially, the closed-loop matrix shows the relationships between MVs and CVs in the inverse, swapping the dependents and independents, and describes the LP optimizer actions. Users can therefore employ the closed-loop gain matrix in combination with variable trends to get a holistic understanding of LP behaviour.

The open-loop matrix describes the relationship between the MV-CV pairings, but caution must be taken when using it for troubleshooting the LP solution. Depending on the gain values, active constraints and pivot points, the pivoted gains in the closed-loop matrix may experience a sign flip, indicating the MV would now move opposite of the expected direction for that particular CV or have their pivoted gains transformed to zero, indicating that the MV will no longer be used as a handle for that particular CV. APC practitioners typically spend significant time during the controller design and commissioning phase to review and adjust gains in the closed-loop matrix to enforce 'structure' in the controller (based on material balance and other physical constraints) and avoid undesirable and erratic MV movements [28].

Performance assessment and monitoring of two-stage LP-MPC systems have been considered by several authors. These insights may aid in designing visualization strategies. Kozub (2002) discusses several methods of monitoring MPCs, specifically with regards to their performance and connecting the steady-state and dynamic optimization components. Relevant issues that we can be wary of include the high volume and complexity of MPC data, and overlooking the LP aspect of the solution [31]. Novel Key Performance Indicators (KPIs) have been proposed by Sun et al. (2011) [30] by considering CV priorities. A more recent paper by Godoy et al. (2017) present an excellent overview and critique of typical KPIs used in industry, and suggest utilizing a novel economic performance indicator [4].

Contrary to LP-MPC system visualization, there have been more significant academic advances in implementing user-focused visualization tools in the field of alarm management and analytics. Hu et al. (2018) provide a taxonomy of existing alarm visualizations and propose new innovative plots based on design requirements for visual techniques, which include bubble charts, treemaps, ranking charts, and spiral graphs [32]. Further, the concise structure and layout shown by Hu et al. (2018) served as a strong basis adapted in this paper, due to the similarity of presented ideas.

Applying lessons learned from novel visualization work such as that by Hu et al. (2018) requires that data visualization fundamentals be understood clearly, such that any created tools can be optimized for operator workflow. Our literature review revealed a noticeable gap of academic work addressing the usability and operational issues of industrial MPC systems faced by practitioners. In the following sections, we discuss the challenges around LP interpretability and MPC diagnostics and propose novel solutions using data visualization principles.

4. Data and Information Visualization

The purpose of data visualization is to communicate information through visual media, allowing users accessible insight into their data. The properties of a given dataset generally dictate which kinds of visual encodings are more effective, and Perin et al. (2019) describe three types of data: quantitative, ordinal, and nominal. Quantitative data have exact numbers without an inherent order (e.g. mass, distance); ordinal data can be numerical or categorical but have an associated order (e.g. low-medium-high classifications); nominal data comprise everything else (e.g. categorical, non-numerical data). In visualization literature are also several 'retinal' variables: methods of visually encoding data that utilize different visual channels in order to convey information [33]. Mackinlay (1986) provides a summary of different visual encodings and their associated effectiveness, as shown in Figure 5.

Using the information in Figure 5 requires a systematic method that involves consistent iteration and evaluation against the visualization targets. We apply Munzner's 'Nested Model' in our work, which breaks the problem of visualization development into four layers: domain characterization, data and task abstraction, encoding and interaction design, and algorithm design. Progressing through the layers takes us to less abstract levels of visualization design [35].

The design process begins with domain characterization, where we identify target users and their needs. Potential errors

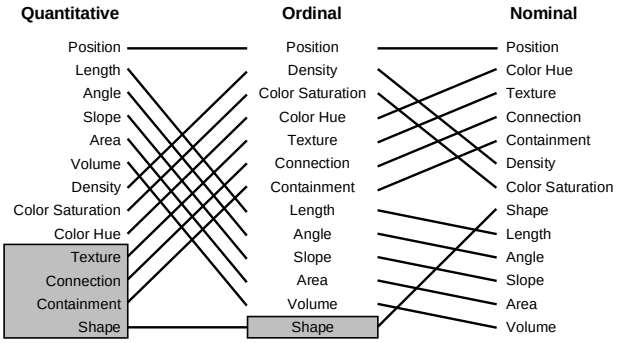

Figure 5: Mackinlay's ranking of retinal variables. The grey boxes indicate encodings that do not apply to the given data type. Figure adapted from Mackinlay (1986) [34].

here are primarily due to mischaracterizing the problem, so we must ensure that the problems discussed are actually faced by the target users - possibly through user studies and discussions. Next is data and task abstraction, where we identify the relevant data and tasks that are to be improved using visualization. We also employ a 'multi-level typology' described by Brehmer et al. (2013) in order to abstract the tasks a user typically goes through into a more generic format that is used for visualization design [36]. Care must be taken to ensure that we are involving the correct data and that we are accurately portraying user tasks in their abstraction. Then, encoding and interaction design are the steps concerned with the actual visualization design, incorporating retinal variables for encoding data as well as interactive parameters to assist the users. This layer requires making suitable choices for visual encodings, as choosing less fitting visual methods will create more work for the users. Finally, algorithm design involved the creation of suitable algorithms to run the visualizations. This layer is also concerned with the choice of software used to create the visualizations, as well as performance parameters such as speed and memory capacity.

Essential to this process is the validation of design decisions as they are made. As this is an iterative process, there is no single validation step taken after design milestones; rather, validation is woven into each step to create a more continual process of improvement. Each layer has associated threats to its validity and we must take these into account throughout the design.

5. Design of visualization plots

In this section, we discuss the typical visualization plots used for MPC controllers and LP optimizers in action. We review traditional methods that are common in industry and identify their limitations. We then introduce novel visualization techniques that address these limitations.

5.1. Time series

Time series data, in the context of MPC controller diagnostics, represents key LP optimizer parameters monitored at every controller execution cycle in the time domain. Each MV and CV configured in the controller will have four key parameters that are representative of the LP optimizer solution: process value (PV, or measurement), steady-state target (SS), upper limit (UL), and lower limit (LL). These 4 variables are important because LP must identify the system's current condition through its process value to output a set of SS targets that respects the limits of each variable.

The 4 parameters for each controller variable are essentially a series of snapshots of the LP optimizer solution over time. A typical setup of time series plots used in practice is shown in Figure 6, along with the colour scheme used in this paper.

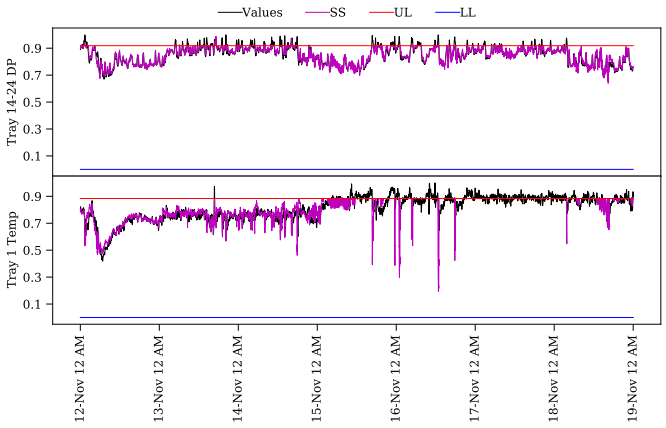

Figure 6: Sample time series plot showing two FCC variables: Tray 14-24 DP and Tray 1 Temperature in the fractionation column. The process values or measurements are plotted in black. The steady-state targets calculated by the LP optimizer are shown in magenta. The red and blue lines denote the upper and lower operating limits, respectively. All plot values are normalized to lie between 0 and 1.

Monitoring a small number of controller variables, for instance, two variables, as illustrated in Figure 6 is relatively simple. There is little difficulty for the user in keeping track of what the LP optimizer is doing at any given time. Each plot has enough space to clearly show the time series trends for both variables. In fact, for a system with just 2 MVs, the entire LP solution space can be visualized and interrogated on paper, as discussed earlier in Figure 2.

In practice, these time series plots are typically monitored regularly by APC engineers using data historians and constantly by plant operators using the DCS. Certain anomalies are relatively easy to spot visually, such as SS targets bouncing between limits, which is indicative of design issues such as plant-model mismatches, poor numerical conditioning in the gain matrix, as well as operational issues, such as clamped variables or instrumentation malfunctions.

However, many controller anomalies are not visually obvious in the absence of broader contextual information and understanding of the APC design intent and controller objectives. For example, which controlled variables should typically be pushed to their limits? Which manipulated variables are typically used to move those controlled variables? The column tray DP in Figure 6 appears to be both constrained and unconstrained in the plot, but are both conditions normal behaviour for the controller? If the aforementioned CVs are not at the correct con-

straint status, how do we determine which other MV or CV to unclamp to shift the LP solution and provide constraint set relief?

Evidently, to understand the behavior of the MPC controller and LP optimizer, relying solely on the time series trends visually is insufficient. Users must also possess a strong understanding of the MPC design intent, controller objectives and gain matrix.

Remark 1. Real world industrial MPC systems are much larger, which exacerbates the problem of information overload. For example, the FCC controller at the Burnaby refinery has a total of 104 variables, meaning that there are over 400 individual trends of interest that will potentially require investigation during abnormal situations. As more variables are trended, it becomes more difficult to understand how the trends are changing, and how variables are related to each other. The user's screen size also imposes a physical limitation on the number of trends that can be included. Consequently, users may need to spread trends over multiple pages, which can make the investigative work more tedious and time-consuming to navigate, and users quickly lose the ability to systematically compare variables that are now spread across different pages. How do we decide which variables to plot on the same page? Furthermore, as shown in Figure 6, the spread of PV and SS for Tray 14-24 DP is low compared to the full variable span. Variables that stay relatively stable may be harder to explore, since the LL-UL range will 'compress' the PV and SS trends, requiring the LL or UL lines to be turned off to improve the visualization plot. Techniques that can help systematically cut down the number of time series trends to investigate will be of immense value to plant engineers and operators.

5.2. Steady-state gain matrix in tabular form

Relationships between MVs and CVs are described by the steady-state gain matrix. The steady-state gain matrix encodes the open loop behavior of CVs in the system, in response to an increase in the MVs by one engineering unit. These values can be represented in a table, as shown in Figure 7.

Visualizing the gain matrix in this way is noteworthy because in practice, APC engineers typically use traditional spreadsheet software like Excel as the default, in-house, general-purpose data analytics tool [37]. Based on discussions with industry partners and contacts, APC engineers in industry have developed many sophisticated, custom-built Excel spreadsheets for various process control tasks, such as gain matrix visualization, LP simulations, matrix pivoting and APC controller monitoring. These spreadsheets are typically proprietary tools that are owned by their respective companies, and not available in the public domain or open sourced.

Consider a hypothetical problem where a user is monitoring the time series trends and noticed that **CV7** breaches its upper limit. Using the open-loop gain matrix, we see that **MV1** and **MV2** are the only variables with nonzero gain, making them the only directly related variables. Looking at those relationships, there are four unique CVs with nonzero gains that could potentially be causing the issue. This process can be iterated until a solution is found or all variables are checked. Iteratively checking variables in this way can be thought of as a 'naive' search process, as we are not incorporating process knowledge at all. In reality, the variables most commonly affecting **CV7** from a practical standpoint will be checked first. Furthermore, not all variables will participate in the LP optimizer solution, and the MV and CV constraints should be used to eliminate irrelevant variables to reduce the search space.

	MV1	**MV2**	**MV3**	**MV4**	**MV5**	**MV6**
CV1	0	0	0	0	0	2.68
CV2	3.47	-4.17	0	-0.91	-1.61	0
CV3	0	0	3.16	4.24	3.8	0
CV4	4.66	0	0	0	4.16	0
CV5	4.44	0	-4.0	0	0	0
CV6	0	0	3.72	0	0	4.65
CV7	-0.96	0.81	0	0	0	0
CV8	-0.67	0.71	0.9	0	0	0

Figure 7: Sample 8×6 gain matrix with CVs as rows and MVs as columns, displayed as a numerical table. For each MV-CV pair, the sign of the gain indicates the directionality, and the gain magnitude determines the strength. The matrix is sparse, and a gain of zero indicates that there is no direct relationship between an MV-CV pair in the process model.

In the context of the LP optimizer solution, the magnitude of each gain is relatively less important - more important is the sign of the gain value, which determines its optimization and control direction. Systematically filtering or reordering of the rows and columns can help users easily identify problematic variables with a smaller margin of error. By sectioning off those variables that do not need further investigation, users can spend less time sifting through rows/columns and rapidly obtain the information that they need.

Remark 2. The prominence of Excel and other spreadsheet tools in the engineering profession makes it a natural baseline for evaluating visualization techniques in the process industry. For alternative techniques and tools to be successful, they must also provide the user-friendliness, widespread availability and ease of access that Excel offers. However, Excel is a poor tool for visualization of the gain matrix. The numerical values for the gains can be hard to read and imposes a lot of cognitive load on the user to process this information. Furthermore, the static, tabular display in Excel makes it difficult to identify the variables of interest. For a large, sparse gain matrix, there may be a lot of zeros that the user has to ignore and sift through to obtain the information that they need. These issues regarding the gain matrix table in Figure 7 become aggressively more prevalent as the system grows, making it very difficult to trace through sparse rows/columns to find related variables. Moreover, it may be difficult for users to keep track of the variables and gain values they are dealing with at any given point, when they have an

overwhelming numbers of variables to keep in mind.

Notably, since the gain matrix only shows steady-state gains, process dynamics are not available in this visualization. However, this is not a concern in the context of visualizing the LP optimizer solution, since the LP only uses the steady-state gains. More accurately, there is only a subset of variables that participate in the LP solution. Thus, not every element in the open-loop matrix is useful in understanding the LP optimizer's actions. By identifying the constraint set consisting of the unconstrained MVs and the constrained CVs, the open-loop matrix can be pivoted to the closed-loop gain matrix, which provides more relevant information in the context of LP solution diagnosis, since it considers the CVs that are actively driven to a setpoint under the influence of the unconstrained CVs. Partially pivoting the gain matrix requires isolating the pivot points: a set of pairings between unconstrained MVs and constrained CVs, where each unconstrained MV is expected to control its paired CV under typical operating conditions. The ability to reorder, filter and sort the gain matrix to eliminate irrelevant variables will be very useful for engineers or operators during a troubleshooting exercise.

5.3. Interactive gain matrix heat map

A heatmap is a data visualization technique that encodes the values in a 2D matrix, typically as colors. The rows and columns in a heatmap represents the dimension of the data. In the context of a gain matrix heatmap, the rows would be the CVs and the columns the MVs, and the color encodes the gain value. However, data pre-processing is an important step in the construction of the heatmap and its resulting utility and practicality.

In a broader data visualization context, there are many options for how we can encode data, and we are not limited to colors. The effectiveness of each method of encoding depends on the data being encoded. Since gain values are continuous and numeric (i.e. gain is not a categorical or ordered form of data), visual encodings that are inherently quantitative like object length, size, or position are better suited to encode them [34]. For example, we can add an opaque bar to each cell and use each bar's length to represent the gain value. This way, users can visually compare different gains more easily than if they had to retain their exact numerical values.

The gain matrix itself can be sparse due to many MV-CV pairs without direct relationships, and have a high variance in values due to the usage of different engineering units. For instance, the full range of Burnaby's raw FCC gain matrix is over 1,000, while the majority of gain values are relatively close to zero. Naively encoding these raw gain values as colors makes the vast majority of the gains indistinguishable from each other. For this reason, we need to be able to transform the data to 'squeeze' the distribution to a smaller range, in order for the encoded colours to be distinct.

In the context of understanding the LP optimizer solution, we are primarily concerned with the gain directions and the exact magnitudes are relatively less important. Therefore, it is natural to apply a sign function to the raw gain matrix, to preserve just the gain directions, and visualize the transformed gain matrix as a heatmap. This transformation addresses the issue of indistinguishable colors, and we can encode positive values in green, and negative gains in red, and white for zero gains, as illustrated in Figure 8.

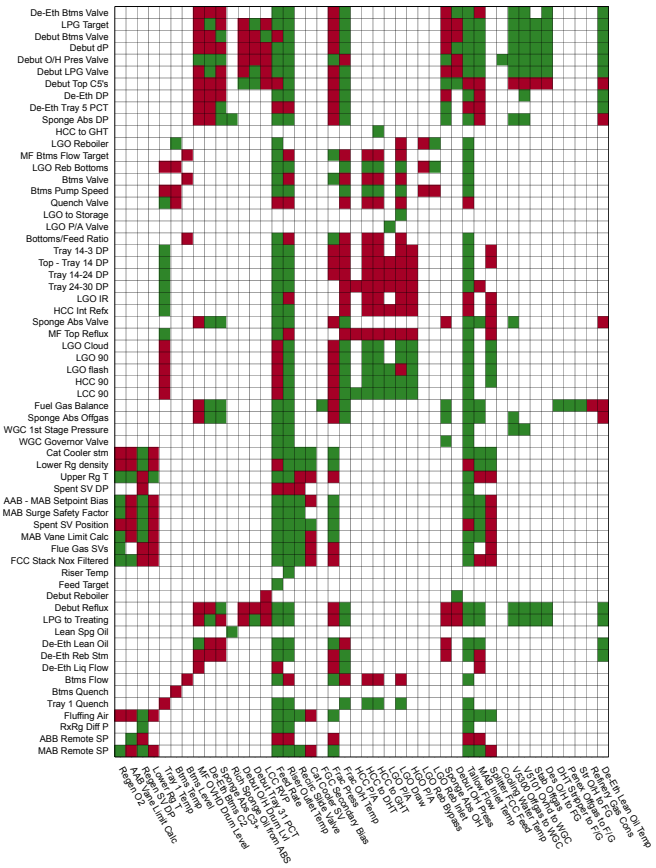

Figure 8: The FCC gain matrix at the Burnaby refinery visualized as a heatmap with CVs as rows and MVs as columns. Negative gains are shown in red, zero gains in white/transparent, and positive gains in green. The ability to visually sort, reorder and filter variables interactively in the matrix will provide a usability improvement over static, numerical tables. For example, by sorting an MV column, users can quickly determine a list of CVs the MV handle is controlling, and divide the CVs into 2 separate groups based on the gain directions.

To gain further insights on the variable relationships and address the restrictive nature of static, tabular displays, we can construct an interactive and reorderable heatmap. Perin et al. (2019) discuss the general benefits of reorderable matrices using Jacques Bertin's physical reorderable heat map [38], revealing underlying structure in the matrix data that are otherwise hidden [33]. By sorting the gain matrix using the gains of a single row/column, one can visually filter variables that are worth investigating in the event of a process upset.

Visualizing the gain matrix as an interactive heat map presents numerous benefits over the raw values. A numerical display of data requires a higher cognitive load to process compared to colors in a heatmap. Comparing two numbers in a table requires a small amount of cognitive work, which is trivial for small datasets but quickly grows as the gain matrix gets larger.

Remark 3. The most useful functionality provided to modify the gain matrix heatmap is filtering, reordering, and transforming the gain data. Filtering the gain matrix is necessary to reduce the number of variables the user needs to manage at any given time. Taking the Burnaby refinery FCC as an example, this gain matrix has 44 columns and 64 rows. Despite the matrix being relatively sparse, having to trace through this many variables can be tedious and can potentially give room for more human error. By selecting the necessary variables, such as those specifically causing problems in the controller, helps users greatly reduce the data they must trace through at any given point. Visualizing the gain matrix as an interactive heatmap also addresses the limitations of static, tabular views inherent in spreadsheets, as discussed in the previous subsection. Clustergrammer [39] is an example of an open-source interactive heat map implementation developed to support bioinformatics research, and includes the ability to cluster and sort the displayed matrix by rows and columns. The code is open-sourced and can be adapted for process control visualization tasks. Aside from filtering and reordering, transforming the gain matrix is important as it allows the user to focus the visualization on a specific feature of the data; an example of this is how we took the sign function in order to show only the gain directions. If, in further iterations, it is necessary to see the gain direction and magnitude, there are numerous transformations available that can accentuate various data features while still keeping the colors distinguishable, such as hyperbolic tangents and log transformations. Selecting or creating a suitable transformation is a separate design decision that requires investigation to understand what information the user should try to emphasize and try to hide [40].

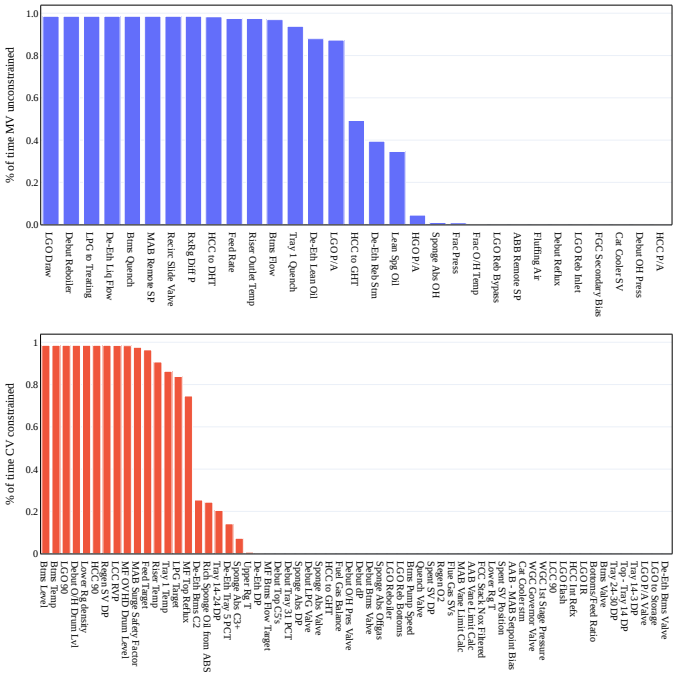

Figure 9: Percent Constraint Activity Plot (PCAP) developed in Kozub (2002) plotted using sample data from the Burnaby refinery FCC controller. For a particular time window, the top subplot is a bar chart sorted in descending order by the % of time an MV is *unconstrained*, and the bottom subplot shows the % of time a CV is *constrained*. An MV that is frequently constrained and unavailable for control (MV_c) is not particularly useful in a process control context. These MVs do not directly participate in the LP optimizer solution since they offer no degrees of freedom to control the system. Similarly, a CV that is typically unconstrained (CV_u) is not being actively driven to any setpoints by the LP directly, but they are moving as a consequence of other MV movements and their relationship in the gain matrix.

5.4. Controller performance monitoring plots

The active constraint set is the set of variables (MVs and CVs) that are at their LP limits during a particular controller execution cycle. A typical task in MPC monitoring and diagnostics is to identify anomalies in the constraint set, which could serve as indicators for clamped variables or process upsets, and help isolate the problematic variables for further action. Kozub (2002) developed a set of visualization tools that can help with controller diagnostics.

The first plot is the Percent Constraint Activity Plot (PCAP), a bar chart that visualizes the fraction of time that a CV is constrained and the fraction of time that an MV is freely available to move. As explained by Kozub (2002), this plot can help identify CVs that are often floating between LP limits, which essentially leaves them in an open-loop state, compared to constrained CVs, which indicates that the feedback controller is actively driving them to a setpoint. This plot can also help identify MVs that are often unconstrained, and can be utilized for feedback control.

Godoy et al. (2017) identify typical KPIs used for monitoring MPC performance that are similar to the PCAP: the *pCVac* and *pMVac*, which define the percentage of CV active constraints to total MVs, and the percentage of MV active constraints to the total MVs, respectively. Contrary to Kozub, Godoy et al. (2017) argue that these KPIs are too simplistic to capture important details for LP operation, such as proximity of inactive limits [4].

The second plot from Kozub (2002) is a bar chart called a Dynamic Constraint Activity Trend (DCAT), which plots the number of CVs that are at their LP limit over time, and the number of MVs that are free for control (not at their LP limits) over time. Chattering in these plots indicate possible LP stability issues.

Remark 4. The constrained CVs and unconstrained MVs form the 'pivot points'. These are MV-CV pairings, where the CVs are actively driven to their limits by the LP solution using the unconstrained MVs for control. This set of pivot points is also sometimes referred to as the 'state' of the multivariable controller in industry (which is different from the mathematical definition of a 'state' in control theory), and these are the variables of interest from a monitoring perspective. As discussed earlier, pivoting can help derive further insights on controller directionality and isolate the variables that are actively participating in the LP solution. A combination of information from the first and second plot can lead to insights on problematic variables that requires further investigation. Variables in the PCAP with percentages not at 0 or 100% during the investiga-

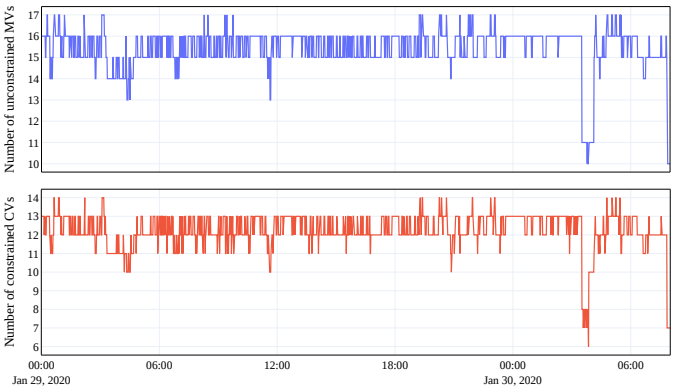

Figure 10: Dynamic Constraint Activity Trends (DCAT) developed in Kozub (2002) plotted using sample data from the Burnaby refinery FCC controller. The 2 subplots provides an overview of variable constraint statuses in the time domain. At any particular time index, the subplot shows the number of MVs that are *unconstrained* and available for control, and the bottom subplot shows the number of CVs that are *constrained* and actively being driven to a setpoint. A sudden drop in the number of MVs that are unconstrained, and consequently, a reduction in the number of CVs being controlled, is indicative of potential controller faults and anomalies that may require further investigation.

tion window are likely to be the variables of concern. For instance, if a given CV is at its limit 80% of the time, that means it dropped from its constraint and therefore can be a problematic variable. Time series trends of the variables and their associated limits and steady state targets can be used to obtain univariate performance indicators for that particular variable, such as their setpoint tracking ability and settling time.

For a large-dimensional controller, the analysis and interpretation of these plots is a nontrivial task due to the volume of time series data and number of variables involved. The two plots presented by Kozub provides an excellent overview of APC performance for monitoring purposes, but is limited by the lack of a direct, visual connection between the PCAP plot and DCAT plot that will facilitate controller troubleshooting and help us drill down to the variables of interest.

5.5. Dynamic Constraint Map (DCM)

To address the limitations discussed in the previous subsections, we now present a novel technique that facilitates LP optimizer solution diagnosis. We propose a new visualization tool called the Dynamic Constraint Map (DCM) that provides an overview of the LP solution by tracking the changes in constraint sets in the time domain, combining information in the PCAP and DCAT plots, but providing more clarity on the individual variables while taking up less screen space compared to the full time series trends.

This dynamic heatmap visualization plot consists of these features:

- Variables are plotted as rows and the timestamp as columns. For clarity, the rows can be sorted with the CVs at the top, followed by the MVs, or can be reordered to group variables with similar patterns together.
- The color in the heatmap encodes the constraint status of that variable in a particular point in time, with time moving in the left to right direction.
- One or more time series trends with the same investigation window can be added to the top of the heatmap to provide additional details for selected variables, and facilitate visual tasks to correlate time series trends with variable constraint statuses.

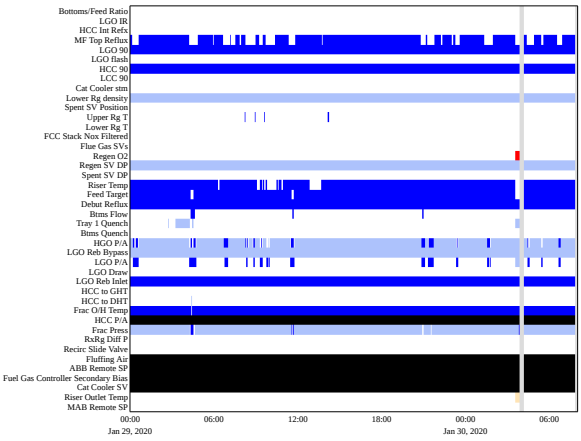

Figure 11: Dynamic Constraint Map (DCM) plotted using sample data from the Burnaby refinery FCC controller. The color encoded for each constraint status is provided in Table 1. Abrupt changes in colors for a variable or a set of variables could be indicative of a potential controller fault.

Remark 5. The DCM plot is designed to condense large volumes of information into a single plot, providing practitioners with a dynamic overview of active constraint sets over time. The plot can be interpreted in multiple ways. By studying a particular variable across time, univariate chattering activity and LP instability issues can be identified. Similarly, by studying the heatmap at a particular time point, the active constraint set and associated MVs and CVs can be identified. Users need not sift through numerous individual trends for a specific variable - the DCM provides data relevant to the LP solution without cluttering the display with numerous trends and intersecting lines. Additionally, for important process variables like FCC feed rate, the DCM clearly shows when they may exhibit abnormal behaviour. For example, FCC feed rate is typically driven towards its upper limit - going through its row in the DCM, if the user encounters cells with different colours, they quickly find where the problem begins. They can then isolate a vertical window and look for changes in other related cells to quickly and easily dig through the related data.

Comparing the heatmap across time will help the user identify anomalies and undesirable changes in the active constraint set. For example, for a typical distillation column, it is typically desirable to maximize throughput and operate at maximum reflux and minimum column pressure to increase profits and reduce energy usage. The user can identify regions in time where this ideal active constraint set were not met, and utilize time series trends for each variable to further investigate the root causes.

Furthermore, the concepts of heatmap sorting, filtering and re-ordering introduced in the previous section can be applied here to derive further insights on controller performance. From a univariate analysis perspective, abrupt changes in the heatmap color for a variable, typically held constant at a certain constraint, could indicate a possible controller fault or abnormal situation that may need to be addressed. The root causes and variables that contribute to this fault can be determined through a multivariate analysis perspective. In the context of controller diagnostics, we can reasonably assume that if the plant is running at steady-state without any operational upsets and posture changes, there will be a particular active constraint set that is most common and active for the majority of the time. Variables deviating from this common constraint set are candidates for further investigation.

In the next section, we explain a general troubleshooting framework using the Dynamic Constraint Map as a visualization tool for controller diagnostics to eliminate variables that are most likely to be irrelevant, and rapidly drill down to variables that matter.

5.5.1. General Troubleshooting Steps Using DCM

- Using time series data, determine the investigation time window and identify the timestamp when the controller was misbehaving and the problematic variable that requires further investigation (i.e. feed rate dropping)
- Plot the DCM for all variables vertically below the time series. Use the time window determined in the previous step. Note that the start and end time should encompass the period where the controller was misbehaving, as well as periods where the controller was functioning normally.
- Eliminate variables from the DCM that do not have constraint changes in the time window. These variables most likely did not contribute directly to the fault, but they can be checked later to identify the root cause
- Apply a hierarchical clustering algorithm to the truncated DCM to reorder the variables. Select an appropriate clustering algorithm and distance metric that will group similar variables together and provide further insights on the constraint trends.
- Observe the patterns in the DCM and identify variables with constraints that correlate with constraints on the problematic variable
- Filter the open loop gain matrix by retaining only variables present in the truncated DCM. Observe the relationships between the variables, and identify variables that have a direct relationship with the faulty variable. Plot these variables as a time series
- Use a combination of the time series plots and process knowledge as a heuristic to eliminate the selected variables and identify the root cause and understand the controller actions. Verify the controller actions using the closed-loop gain matrix if necessary

6. Industrial Case Studies

To illustrate the effectiveness of techniques developed in this work, we present two industrial case studies to describe the procedures and mechanics of these techniques and highlight improvements over conventional troubleshooting methods. Both case studies are applications on a real APC system at the Burnaby refinery. The colors used in the DCM plots are provided in Table 1 below.

Table 1: Color legend for DCP plots in all industrial case studies presented here.

Color	Constraint Type
White	Unconstrained
Dark Blue	Constrained - Upper Limit
Light Blue	Constrained - Lower Limit
Yellow	Constrained - Ramp or others
Red	Gave Up (Infeasibility)
Black	Out of Service (Unused)

6.1. Case Study 1: FCC Feed Rate

Consider a fluidized catalytic cracking (FCC) unit at the Burnaby Refinery in British Columbia. The FCC APC controller has been designed with 44 independent variables (MVs and FFs) and 64 dependent variables. The data set includes about 32 hours of APC data with a sampling rate of 1 minute. The data set contains an anomaly - a drop in FCC feed rate target (CV), as shown in the combined time series and constraint heatmap in Figure 12. We can observe that the feed rate target switched from an upper limit constraint to unconstrained in the region highlighted in red, accompanied by changes in the constraint status of other variables as indicated in the heatmap.

The goal of this case study is two-fold. First, we want to rapidly identify a smaller, more manageable subset of the 108 variables that are the most likely contributors to this anomaly. Second, through this smaller subset, we want to use a combination of the gain matrix and process knowledge to identify the most likely root cause.

We apply data pre-processing techniques to improve the appearance and utility of this constraint heatmap. First, we identify and remove variables that had no changes in its constraint in the time window of interest. If those variables did not undergo any constraint changes during the period of anomaly, we can reason that they most likely did not directly contribute to the drop in feed rate target. The remaining variables are those that experienced constraint changes in the dataset. Second, we apply a clustering algorithm to group similar rows together. The clustering algorithm reorders the APC variables, such that similar variables are closer to each other, based on a pre-defined distance metric. Reordering the variables will help us identify patterns visually. The exact distance metric (e.g. Euclidean, Hamming or Jaccard) and clustering algorithm (e.g. single-linkage, complete-linkage or Ward's) can be a tuning handle and selected based on the nature of the APC dataset under investigation. By using an interactive zoom slider, we can rapidly

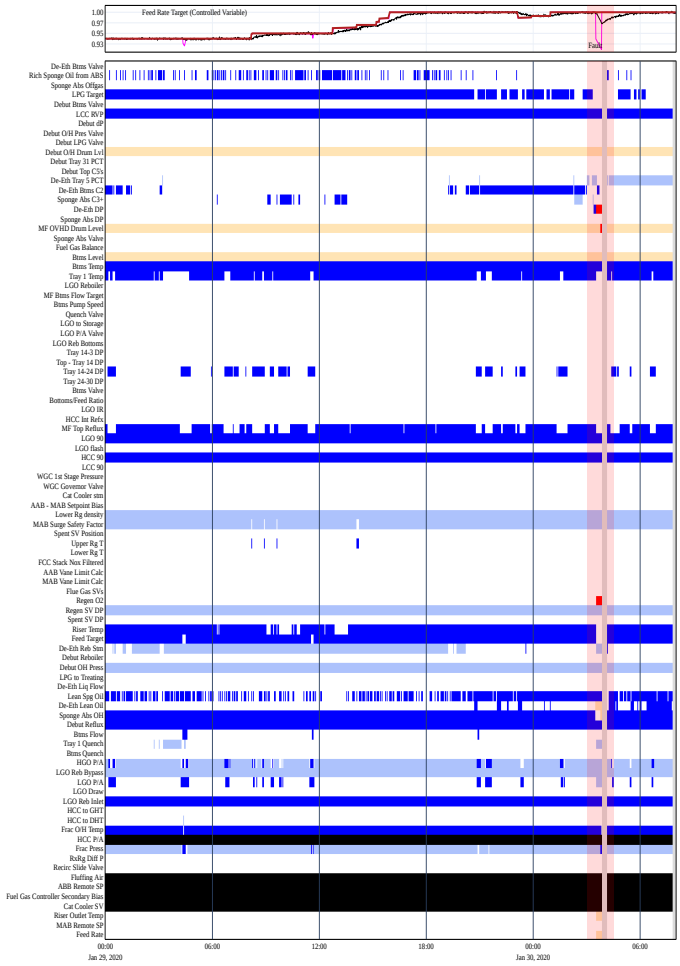

Figure 12: Combined DCM and time series trends for Case Study 1, investigating a process upset in the FCC feed rate. The time window under investigation is highlighted in light red. We can observe that the majority of variables did not experience any constraint changes. These variables are unlikely to be direct contributors to the feed rate drop, and can be removed to narrow down the list of variables for further investigation.

drill down to the sections of interest, and select only the relevant variables within the highlighted period with constraint changes for further investigation.

This technique allows us to rapidly hide all the irrelevant variables without constraint change that had no contributions to the feed rate drop. We have reduced the original 108 variables by 3 fold to a subset of about 30 suspicious MVs and CVs that require further investigation. Once these suspicious variables have been identified, we construct a *gain submatrix* of the original open-loop gain matrix by retaining only the suspicious variables to interrogate their relationships, and eliminating all other rows and columns as shown in Figure 13.

We then plot the time series for those suspicious variables and perform a visual inspection. The time series trends reveals abnormal situations for the De-Eth DP, De-Eth C2 Bottoms and De-Eth Reb Stm variables immediately before the feed rate drop, as shown in Figure 14. The most salient observation in the time series is that the feed rate drop correlates with a sharp increase in De-Eth DP. To verify a relationship between the Feed

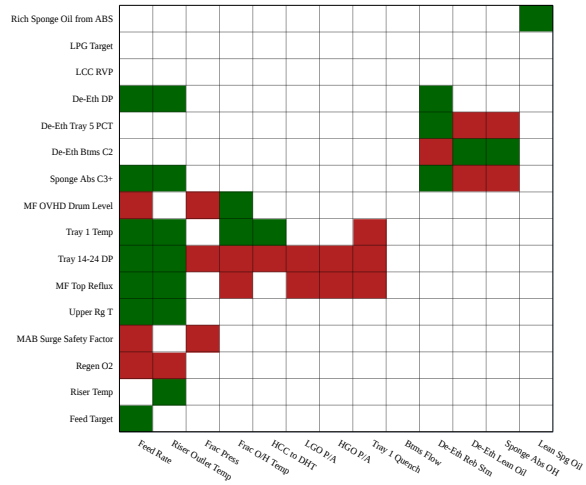

Figure 13: Submatrix from the FCC gain matrix showing the MVs and CVs that need investigation as outlined in figure 12. Negative gains are shown in red, zero gains in white and positives in green. Since feed rate is the issue, we can find that it is related to 11 CVs by looking at the leftmost column. The next step would be to examine the behaviour of these related variables.

Rate and the De-Eth DP, we use the gain submatrix and observe that the De-Eth DP is controlled by the Feed Rate, De-Eth Reb Stm and the Riser Outlet Temp.

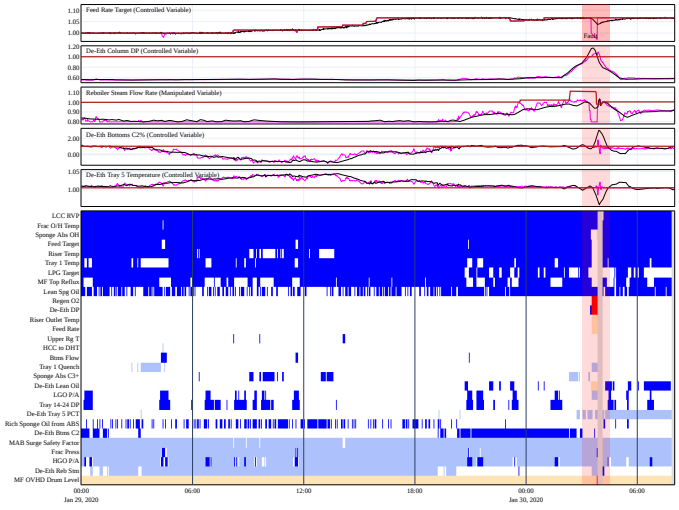

Figure 14: Case study 1 combined DCM with relevant time series plots after clustering based on variable constraint changes before and during the investigative window. Note the abnormal behaviour in each of the time series above the DCM. For novice users, it may be difficult to distinguish effects of the upset from their causes, which is where expert interpretation comes in. Variables that are known by SMEs to cause feed rate issues will be the first ones they examine, while novice users may consider all options likely.

To probe for the root cause, we then ask, what could have caused the sharp increase in the De-Eth DP? Based on the gain submatrix and process knowledge, the pressure in the de-ethanizer column would be directly correlated with the De-Eth Reb Stm. Using the time series, we observe that the De-Eth Reb Stm displays abnormal behaviour, increasing for several hours prior to the feed rate drop. Based on the gain submatrix and process knowledge, by firing the reboiler steam harder, the C2 Bottoms

will decrease, and the Tray 5 temperature will increase. They are inversely correlated because a higher Tray 5 temperature near the bottom of the de-eth column will boil off the light ends, and decrease the C2 composition of the bottom stream.

Evidently, the controller was trying to control the C2 Bottoms and Tray 5 temperature within limits by increasing De-Eth Reb Stm. The drastic increase in steam rate over several hours resulted in a surge in column dP, which then forced the APC controller to cut feed. These variable relationships can be confirmed by using closed-loop gain matrix using the pivot points defined by the active constraint set when the feed rate started dropping. Based on discussions with plant engineers, a lighter crude slate and an accumulation of excess C2s in the column was the most likely physical cause of this incident.

In this case study, we have demonstrated how to apply the techniques described in this paper to rapidly isolate the key variables out of 108 variables in a few minutes based on its constraint changes. The conventional APC troubleshooting methodology requires strong process knowledge and domain expertise on how the specific APC controller was designed. Without expert knowledge, a novice APC engineer will have to perform a tedious naive search to investigate almost all 108 variables and understand their intricate relationships through the open loop gain matrix. Supposed that the investigation was performed using conventional tools like Excel and simple time series trends spread across multiple pages, a novice user may not realize that a limitation in the de-ethanizer plant in the downstream section of the FCC is directly contributing to a feed rate cut at the front end, and may inadvertently fail to notice the De-Eth DP anomaly while screening through over 100 variables.

By using a combination of well-designed visualization tools and systematic techniques such as clustering, filtering and variable elimination, the APC engineer can focus on just the key variables for further investigation and rapidly arrive at the solution compared to conventional methods.

6.2. *Case Study 2: FCC Riser Outlet Temperature*

In this section, we apply the same techniques to a second case study. We investigate the reasons for a sudden drop in the riser outlet temperature (ROT) of a FCC unit. In the FCC system, the 'riser' is a reaction chamber in which fluidized catalyst and raw crude oil are mixed and 'cracked' to form shorter chain hydrocarbon products. The severity of cracking is determined by the riser temperature, which can be adjusted to shift the FCC products and yields.

The data set includes 24 hours of APC data, also with a sampling rate of 1 minute. The data set contains an anomaly - repeated drops in the riser outlet temperature, as shown in the combined time series and constraint heatmap in Figure 15. We can observe that the ROT target erratically switched from an upper limit constraint to both unconstrained and lower limit constraint in the region highlighted in red, accompanied by changes in the constraint status of other variables as indicated in the heatmap.

We apply the same data pre-processing techniques as described in Case Study 1, by removing the variables without constraint changes to identify only the relevant variables, and apply

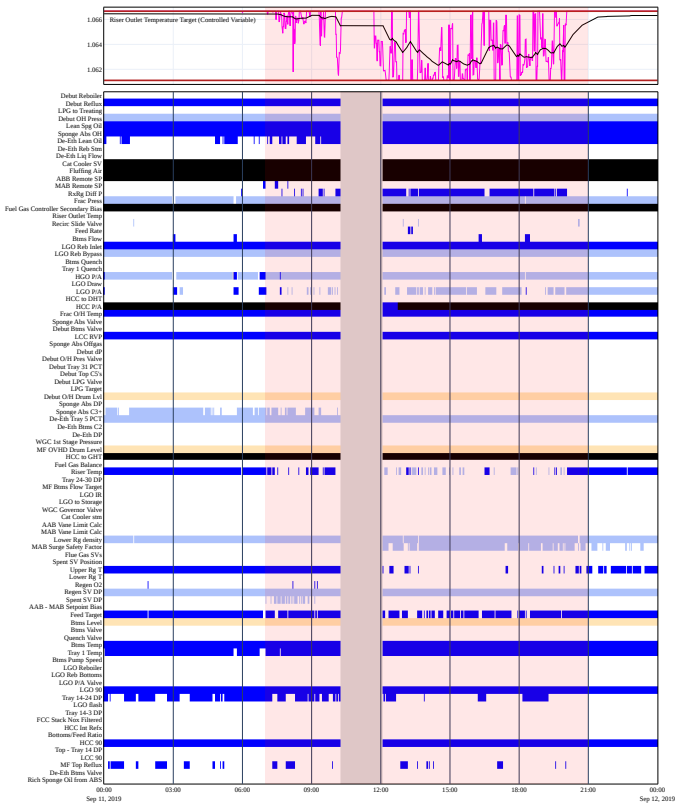

Figure 15: Combined DCM and time series trends for case study 2, investigating a process upset in FCC ROT. The investigative time window is highlighted in light red. Compared to case study 1, there is a much wider time window to investigate, corresponding to about 14 hours. In this window, there are many variables whose constraint status stays the same, meaning that they can be eliminated from our analysis.

hierarchical clustering to reorder the rows such that variables with similar constraint behavior are grouped together. The results of clustering are shown in Figure 17 and a gain submatrix of the remaining variables plotted in Figure 16.

Based on visual inspection of the constraint set heatmap, it appears that the ROT variable constraints are correlated with the *RXRG dP* variable. The clustered heatmap displays the ROT and RXRG dP in adjacent rows, and shows that the riser temp fails to reach its typical upper limit constraint whenever the RXRG dP is constrained, as highlighted in Figure 17.

An FCC unit is a delicate pressure balance operation. Differential pressure between different vessels drives the fluidized catalysts in one direction. The RXRG dP variable is the differential pressure between the reactor section and the upper regenerator section. Physically, the RXRG dP is controlled by a slide valve that regulates the flue gas existing the regenerator. Since the reactor pressure is set by the main fractionator column top pressure, a higher dP would indicate a higher regenerator pressure [41].

Discussions with plant engineers revealed that in that time frame, the FCC plant experienced an upset condition, requiring higher regenerated catalyst slide valve (RCSV) dP, the differential pressure measured across the upper regenerator and riser, to maintain catalyst circulation. From a process safety

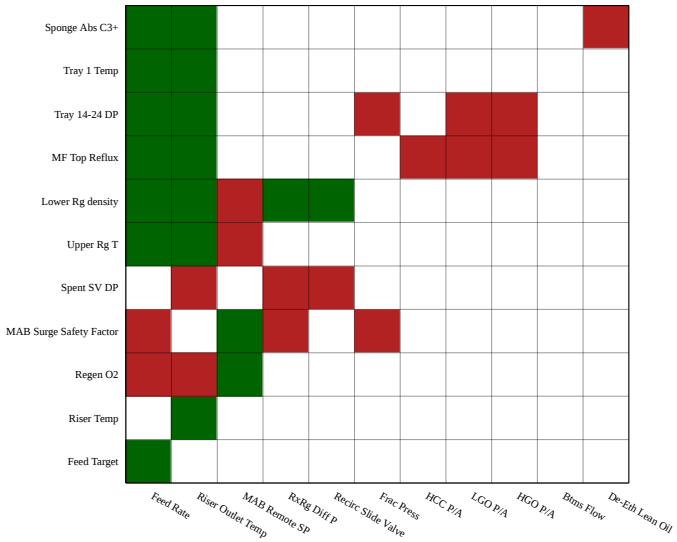

Figure 16: Submatrix from the FCC gain matrix showing the relevant variables requiring investigation based on figure 15. Since ROT is exhibiting the upset, we can see that there are 9 variables directly related to ROT that may need investigation.

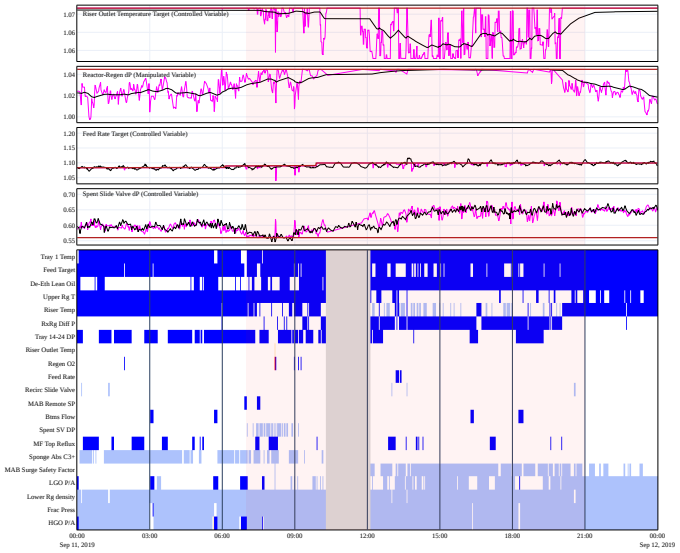

Figure 17: Case study 2 combined DCM with relevant time series plots after clustering based on variable constraint changes before and during the investigative window. Note that the time series other than ROT here exhibit more subtle constraint changes. The Spent Slide valve DP for example, corresponding to the time series at the bottom, only touches its LL for a brief period and then floats between its limits. In contrast, RXRG DP (the second time series) clearly ramps up and stabilizes at its UL.

perspective, the RCSV dP must always be kept above a safe limit to prevent flow reversal, in which the hydrocarbons will flow backwards from the riser into the regenerator, leading to possible explosions or other safety hazards [41].

By using the gain submatrix, we can rationalize using process knowledge that since the reactor-regen DP handle was constrained, and because the RCSV dP was at the lower limit, the LP had to pick the next best available handle to attempt to keep the RCSV dP above the safety limit. There are only 3 handles, the RXRG dP, the ROT or the feed rate. Based on the LP costs set during the design of the APC controller, the next available handle was the ROT. This reasoning can be confirmed by using closed-loop gain matrix using the pivot points defined by the active constraint set when the ROT started dropping.

The feed rate case study involves a single sharp drop in feed rate, whereas the ROT case study exhibits multiple gradual rises and falls. Consequently, there is a wider time span within which other variables must be investigated. We have shown that the techniques used for the first case study can also be successfully adapted for this slightly more complex situation.

A novice user troubleshooting this issue using Excel and other conventional methods may struggle with the cognitive load of performing a naive search of over 100 variables. Furthermore, the larger time window of oscillating targets, compared to the first case study, which was a single, sharp drop, makes the troubleshooting process more difficult. By combining effective visualization tools with process knowledge, we demonstrate how the techniques described in this paper can help plant engineers efficiently troubleshoot real world APC issues.

7. Limitations and future work

The tools presented here are meant to highlight typical operational issues faced by engineers and operators in the process industry and suggest visualizations for solving these problems. We also introduce the idea of human-centered design to the field of APC, to encourage the development of MPC visualization tools with improved usability and utility.

One major limitation is that the presented tools only guide the process of controller diagnosis, aiming to point the users in the right direction to investigate further. With these tools there are still numerous manual steps that the user needs to go through, which have not yet been automated. Further work on the computational aspects of the analyses presented here can significantly reduce the amount of work required by users. One option is to develop analytical methods or machine learning tools to analyze constraint data, where these tools would provide suggestions for variables to unclamp with associated probabilities that each suggestion is 'right'. It is important to note that this tool is not a replacement for a full LP simulation typically used by practitioners during the controller design and commissioning process, but a visualization tool using existing, historical data that augments the user's troubleshooting capabilities.

Another limitation is that the case studies provided in this paper are all faults or disturbances occurring within a single MPC controller. Typically there can be numerous MPC controllers for different process sections that have interconnected feed streams, meaning that faults from upstream controllers could propagate to downstream controllers as unmeasured disturbances or feedforward variables. Process upsets occurring in the FCC, for example, may cause subsequent upsets in downstream units such as the Gasoline Hydrotreater or Alkylation plants, and a troubleshooting exercise conducted from the Alkylation controller's perspective may not yield a root cause. Engineers may

need to communicate with the wider plant personnel in order to have the full picture, meaning that diagnosing this kind of upset would require a larger scope of process knowledge. A higher supervisory control layer, such as an RTO (Real-Time Optimization) system, that describes the interaction between various plants and MPC systems, can be useful in augmenting the required process knowledge.

Finally, the tools presented here lack usage of the full, dynamic process models, and only the steady-state gains. Consequently, these tools do not take into account the dynamics of the process. The gain matrix provides steady-state data, and so the tool cannot provide information regarding how quickly or aggressively one MV would affect a particular CV. Incorporating process dynamics may provide some additional benefits in understanding moves made by the dynamic optimizer, as opposed to just the LP solution emphasized by this paper. However, if required, engineers can conduct dynamic simulations that are typically included in commercial MPC packages to diagnose these issues.

In terms of future work, there are three major areas that we wish to emphasize in order to effectively improve this research. First, users can greatly benefit through the automation of the controller diagnosis process. The idea here is for a tool that can read the controller state and all of its changes during a provided time window - the tool would analyze constraint set changes and process trends to provide suggestions as to which variables may be causing the upset, with each suggestion having an associated probability of being the right one. Such a tool could use a database of case studies as in this research to build a machine learning algorithm, where the algorithm learns what data to look at and what patterns to look for.

Secondly, since the tools presented here aim to mitigate the human factors issues with common commercial MPC packages, a formal usability study can help to give more structured and detailed pinch points that we can focus on to improve. We've solicited informal user feedback from plant engineers in the design and evaluation of visualization tools presented in this paper, but not a formal usability study. Conducting such a study would involve what Munzner (2009) refers to as a *formative evaluation*, which identifies areas where the visualization can be improved to achieve its intended purpose [35]. This process requires regular consultation with target users and re-evaluation of the design with respect to the original objectives. Additionally, testing the designed visual tools using a wider variety of datasets can help to ensure that the tools are flexible enough to be used in different contexts.

Finally, there is a need to delve deeper into concepts of Human-Computer Interaction (HCI). Understanding the basis of how humans interact with computers from a psychological and logistical viewpoint may help to provide better-designed components that are optimized to meet performance requirements. For example, formulating the problem using Norman's Execution and Evaluation cycle can help to pinpoint where the users experience difficulty in navigating or understanding the system, and can help to minimize the presence of redundant affordances [42][43]. Incorporating these concepts will help improve the visualization design, and streamline the tool's workflow to eliminate potential sources of confusion or error.

8. Conclusions

This paper outlines the usability issues common to commercial MPC packages in industry, which have been overlooked in favour of developing increasingly complex and robust MPC algorithms. Based on literature and discussions with researchers and industry practitioners, the past two decades have seen limited academic research on usability and operational issues of MPC systems. MPCs generally have large amounts of data that may be relevant, so this paper summarizes the process of diagnosing problematic controllers and presents visual tools that can help users to easily identify where there may be control issues. The mechanics of controller diagnosis are discussed in detail, including essential industrial practices like the partial pivoting operation, which have been difficult to come by in academic literature. By isolating relevant data and transforming it into visual patterns the user can identify, users can obtain insights into their data more easily and perform less work to get to the same solution. This paper provides an overview of existing tools for MPC visualization and their pitfalls in order to guide the design of new visual tools, and tests these designs on two industrial case studies. Visual tools here like the DCM, interactive gain matrix heat map, and PCAP/DCAT are not especially prevalent in industrial practices, even though they summarize MPC data essential to the controller diagnosis process. The efficacy of these new tools is discussed in comparison to the current process in which engineer diagnose problematic controllers.

Our aim is to develop visualization tools for assisting control engineers with their daily MPC monitoring and troubleshooting tasks. However, there are many different facets of the LP-MPC system that we have yet to explore. These aspects include, but are not limited to: LP costs, the objective function, dynamic MPC optimization, as well as integration with real-time process historian data and evaluating other multivariate data visualization techniques such as parallel coordinate plots. The tools we present here scratch the surface of how we can improve day-to-day MPC operation using data visualization, and ultimately, the goal of this paper is to highlight this research gap and encourage further exploration of this topic by the wider controls research community.

9. Acknowledgements

The authors would like to gratefully acknowledge support from the Burnaby Refinery team and in particular, staff members in the Process Control department: Jin Li, Amy Chiu, Paul Herar, Hao Zhou, David Beaudoin, Eric Loong, Ken Leduc and Ashish Malhotra, who provided instrumental data, guidance, and examples that we used to develop our research. We would also like to thank Prof. Greg Jamieson and Prof. Stephanie Guerlain for helpful comments and insights related to the Elucidator paper and Dr. Antony Hilliard for pointing us to the intriguing comments on the Emerson blog. Bhushan Gopaluni would like to acknowledge the financial support from Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] J. H. Lee, Model predictive control: Review of the three decades of development, International Journal of Control, Automation and Systems 9 (3) (2011) 415–424.

[2] J. M. Maciejowski, Predictive control: with constraints, Pearson education, 2002.

[3] E. F. Camacho, C. Bordons, Model Predictive control, Advanced Textbooks in Control and Signal Processing, Springer London, 2007. doi:10.1007/978-0-85729-398-5.

[4] J. Godoy, A. Ferramosca, A. González, Economic performance assessment and monitoring in LP-DMC type controller applications 57 26–37. doi:10.1016/j.jprocont.2017.06.007.

[5] H. Sun, T. Zou, J. Liu, M. Wang, Double-layer model predictive control integrated with zone control, ISA transactions 114 (2021) 206–216.

[6] C. R. Cutler, B. L. Ramaker, Dynamic matrix control - A Computer control algorithm, 1979.

[7] R. C. Sorensen, C. R. Cutler, LP integrates economics into dynamic matrix control, Hydrocarbon Processing 77 (9) (1998) 57 – 66.

[8] A. Ishikawa, M. Ohshima, M. Tanigaki, A practical method of removing ill-conditioning in industrial constrained predictive control, Computers & chemical engineering 21 (1997) S1093–S1098.

[9] C.-M. Ying, B. Joseph, Performance and stability analysis of LP-MPC and QP-MPC cascade control systems, AIChE Journal 45 (7) (1999) 1521–1534, _eprint: https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.690450714. doi:https://doi.org/10.1002/aic.690450714.

[10] A. Nikandrov, C. L. Swartz, Sensitivity analysis of lp-mpc cascade control systems, Journal of Process Control 19 (1) (2009) 16–24.

[11] S. Qin, T. A. Badgwell, A survey of industrial model predictive control technology, Control Engineering Practice 11 (7) (2003) 733–764. doi:10.1016/S0967-0661(02)00186-7.

[12] M. L. Darby, M. Nikolaou, Mpc: Current practice and challenges, Control Engineering Practice 20 (4) (2012) 328–342.

[13] M. G. Forbes, R. S. Patwardhan, H. Hamadah, R. B. Gopaluni, Model predictive control in industry: Challenges and opportunities, IFAC-PapersOnLine 48 (8) (2015) 531–538. doi:10.1016/j.ifacol.2015.09.022.

[14] M. Kano, M. Ogawa, The state of the art in chemical process control in japan: Good practice and questionnaire survey, Journal of Process Control 20 (9) (2010) 969–982.

[15] S. Guerlain, G. Jamieson, P. Bullemer, R. Blair, The MPC elucidator: a case study in the design for human-automation interaction, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 32 (1) (2002) 25–40. doi:10.1109/3468.995527.

[16] L. F. Lautenschlager Moro, D. Odloak, Constrained multivariable control of fluid catalytic cracking converters, Journal of Process Control 5 (1) (1995) 29–39. doi:https://doi.org/10.1016/0959-1524(95)95943-8.

[17] I.-S. Han, J. B. Riggs, C.-B. Chung, Dynamic matrix control of a fluidized catalytic cracking process, IFAC Proceedings Volumes 34 (25) (2001) 281–286, 6th IFAC Symposium on Dynamics and Control of Process Systems 2001, Jejudo Island, Korea, 4-6 June 2001. doi:https://doi.org/10.1016/S1474-6670(17)33837-5.

[18] M. V. Kothare, R. Shinnar, I. Rinard, M. Morari, On defining the partial control problem: Concepts and examples, AIChE Journal 46 (12) (2000) 2456–2474.

[19] S. Goodhart, J. Nishizawa, K. Yano, H. Yada, Advanced control in cogeneration utility management, Computing & Control Engineering Journal 11 (6) (2000) 273–282.

[20] A. Arbel, I. H. Rinard, R. Shinnar, Dynamics and control of fluidized catalytic crackers. 3. designing the control system: choice of manipulated and measured variables for partial control, Industrial & Engineering Chemistry Research 35 (7) (1996) 2215–2233. doi:10.1021/ie9507080.

[21] A. Arbel, I. H. Rinard, R. Shinnar, Dynamics and control of fluidized catalytic crackers. 4. the impact of design on partial control, Industrial & engineering chemistry research 36 (3) (1997) 747–759.

[22] Y.-J. Park, S.-K. S. Fan, C.-Y. Hsu, A review on fault detection and process diagnostics in industrial processes, Processes 8 (9) (2020). doi:10.3390/pr8091123.

[23] J. Cahill, Perspectives on big matrix and unit-level MPC applications (2008).

[24] C. Ramos, J. S. Senent, X. Blasco, J. Sanchis, LP-DMC CONTROL OF a CHEMICAL PLANT WITH INTEGRAL BEHAVIOUR 35 (1) 423–428. doi:10.3182/20020721-6-ES-1901.00639.

[25] D. Fernandes, M. E. Haque, S. Palanki, S. G. Rios, D. Chen, Dmc controller design for an integrated allam cycle and air separation plant, Computers & Chemical Engineering 141 (2020) 107019. doi:https://doi.org/10.1016/j.compchemeng.2020.107019.

[26] K. Brooks, Linear model predictive control – really not so bad?, in: South African Council for Automation and Control, 2017.

[27] C. Lindscheid, A. Bremer, D. Haßkerl, A. Tatulea-Codrean, S. Engell, A test environment to evaluate the integration of operators in nonlinear model-predictive control of chemical processes, IFAC-PapersOnLine 49 (32) (2016) 129–134. doi:10.1016/j.ifacol.2016.12.202.

[28] D. W. Hoffman, D. L. O'Connor, K. H. Rasmussen, Durable model predictive control, in: AIChE, 2010.

[29] T. J. Peterson, A. R. Punuru, K. F. Emigholz, R. K. Wang, D. Barrett-Payton, Model predictive controller solution analysis process, library Catalog: Google Patents (2011).

[30] Z. Sun, S. J. Qin, A. Singhal, L. Megan, Control performance monitoring of LP-MPC cascade systems, in: Proceedings of the 2011 American Control Conference, IEEE, 2011, pp. 4422–4427.

[31] D. J. Kozub, Controller performance monitoring and diagnosis. industrial perspective, IFAC Proceedings Volumes 35 (1) (2002) 405–410. doi:10.3182/20020721-6-ES-1901.01621.

[32] W. Hu, A. W. Al-Dabbagh, T. Chen, S. L. Shah, Design of visualization plots of industrial alarm and event data for enhanced alarm management, Control Engineering Practice 79 (2018) 50–64. doi:https://doi.org/10.1016/j.conengprac.2018.07.005.

[33] C. Perin, J.-D. Fekete, P. Dragicevic, Jacques bertin's legacy in information visualization and the reorderable matrix, Cartography and Geographic Information Science 46 (2) (2019) 176–181. doi:10.1080/15230406.2018.1470942.

[34] J. Mackinlay, Automating the design of graphical presentations of relational information, ACM Trans. Graph. 5 (2) (1986) 110–141. doi:10.1145/22949.22950.

[35] T. Munzner, A nested model for visualization design and validation, IEEE Transactions on Visualization and Computer Graphics 15 (6) (2009) 921–928. doi:10.1109/TVCG.2009.111.

[36] M. Brehmer, T. Munzner, A multi-level typology of abstract visualization tasks, IEEE Trans. Visual. Comput. Graphics 19 (12) (2013) 2376–2385. doi:10.1109/TVCG.2013.124.

[37] L. Zhai, B. Sun, L. Guo, M. Liu, The Application of Excel Software in Chemical Thermodynamics Calculation, in: International Conference on Information and Communication Technology for Education, Toronto, Canada, 2016.

[38] J. Bertin, Semiology of Graphics: Diagrams, Networks, Maps, ESRI Press, 2011.
URL https://books.google.ca/books?id=X5caQwAACAAJ

[39] N. F. Fernandez, G. W. Gundersen, A. Rahman, M. L. Grimes, K. Rikova, P. Hornbeck, A. Ma'ayan, Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data, Scientific Data 4 (1) (2017) 1–12, number: 1 Publisher: Nature Publishing Group. doi:10.1038/sdata.2017.151.

[40] M. Eisemann, G. Albuquerque, M. Magnor, Data Driven Color Mapping, in: S. Miksch, G. Santucci (Eds.), EuroVA 2011: International Workshop on Visual Analytics, The Eurographics Association, 2011. doi:10.2312/PE/EuroVAST/EuroVA11/005-008.

[41] R. Sadeghbeigi, Fluid catalytic cracking handbook: An expert guide to the practical operation, design, and optimization of FCC units, Butterworth-Heinemann, 2020.

[42] Y. Rogers, New theoretical approaches for human-computer interaction, Annual Review of Information Science and Technology 38 (1) (2004) 87–143, _eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/aris.1440380103. doi:https://doi.org/10.1002/aris.1440380103.

[43] J. Fogarty, A. Fiannaca, L. Milne, S. Kawas, K. Munsell, CSE 440: Introduction to HCI - user interface design, prototyping and evaluation, in: University of Washington.