Version of Record: https://www.sciencedirect.com/science/article/pii/S0009250922000525
Manuscript_6bb808e76a8c427de76a660f1468339b

A data-driven linear formulation of the optimal demand response scheduling problem for an industrial air separation unit

Morgan T. Kelley[1], Calvin Tsay[1,*], Yanan Cao[2], Yajun Wang[2], Jesus Flores-Cerrillo[2], Michael Baldea[1†]

[1] *McKetta Department of Chemical Engineering*
The University of Texas at Austin, Austin, Texas 78712, United States
[2] *Smart Operations, Center of Excellence (COE), Linde Plc.*
Tonawanda, NY 14150 USA

January 17, 2022

Abstract

Demand response (DR) has become a key element in balancing the power grid as the contribution of time-varying renewable power generation increases. Chemical plants are appealing candidates for DR programs as they offer large, concentrated and flexible loads. DR participation calls for frequent production rate changes over time scales that overlap with the dominant dynamics of the plant. Production scheduling should therefore consider the process dynamics explicitly. We present a data-driven approach for modelling the scheduling-relevant dynamics based on historical closed-loop operating data using autoregressive with extra inputs (ARX) models. We introduce a new, linear scheduling problem formulation based on the ARX representation, and demonstrate its implementation on an industrial air separation unit.

keywords: Autoregressive with extra inputs (ARX) models, air separation units (ASU), demand response, scale bridging models, production scheduling

[*] Current affiliation: Department of Computing, Imperial College London, 180 Queen's Gate, London SW7 2AZ, United Kingdom

[†] Corresponding Author: mbaldea@che.utexas.edu

1 Introduction

A sustainable future calls for a significant increase in the contribution of renewable sources (wind, solar photovoltaics) to the power generation portfolio. Renewable power generation challenges the operation of the power grid since generation rates fluctuate daily and seasonally, and are not synchronized with power demand, which is itself changing over similar time scales. Under these circumstances, managing electricity consumption in addition to generation becomes a key ingredient towards balancing supply and demand on the power grid.

In principle, any type of electricity user (residential, commercial and industrial) can engage in demand-side management/demand response (DSM/DR) programs. However, industrial users (which accounted for 32% of annual electricity consumption in the US in 2018 [1]) are particularly attractive for demand response (DR) participation for several reasons. First, they present large, localized loads that can provide significant relief on grid demand when their electricity use is reduced or shifted to off-peak times. Second, unlike commercial and residential users, they are minimally dependent on human needs and behaviors, which tend to vary considerably throughout the day. Last but not least, industrial users can be "flexible" in their power demand.

The participation of an industrial entity in a DR program involves two complementary actions occurring on a daily cycle. First, there is an increase in production rate and electricity demand during off-peak hours, when the grid load is low. During this time, product(s) are made in excess and placed in a storage facility. Then, a second action involves decreasing the production rate (and associated electricity demand) during peak power demand hours. During this time, stored product is used to make up the difference between a lower production rate and the product demand. Establishing the timing and duration of these actions is the prerogative of production scheduling calculations. From an economic perspective, and assuming that the price of the product is constant in time, the industrial entity engages in a type of arbitrage, by purchasing more electricity at low prices during off peak times, and storing it in the form of a (cheaper) product, which can be sold at the "regular" price during peak times. Additional savings are derived from reducing production when electricity prices are high during peak hours. Coordinating these decisions is the subject of short-term production scheduling.

Evidently, the aforementioned actions are conditioned by several factors: first, excess production capacity must be available. Second, the product should be easily and safely stored, should not degrade significantly while in storage, and should be easily retrievable from storage. Third, the system should be agile enough to change its production rate and thus grid load in a useful amount of time, without jeopardizing product quality or operational safety. In practice, this means that systems with time constants in the orders of minutes to hours are preferred. Nonetheless, due to the evident overlap between the dynamics of the system, and the time scales over which grid demand and renewable power generation fluctuate, the dynamics of the process must be accounted for explicitly in production scheduling calculations.

A central difficulty is related to modeling the dynamics of the process and relevant control systems; this requires balancing model accuracy, complexity and dimensionality, such that, i) the model can be built and updated easily and ii) the scheduling calculation including the dynamic model can be solved in a practical amount of time. Data-driven modeling (i.e., constructing surrogate, low-order models of the scheduling-relevant process dynamics from data collected from plant operations or from dynamic simulations) has emerged as the approach of choice to address this challenge [2], particularly in circumstances when the process is itself large-scale and complex, and the use of established model forms such as Hammerstein-Wiener structures, has been successfully demonstrated [3, 4]. Nevertheless, existing modeling approaches result in scheduling problems that are either in a nonlinear program or mixed-integer linear program form, both with their inherent solution difficulties.

Motivated by the above, the purpose of this paper is twofold: i) introduce a framework for modeling the process dynamics and for optimal demand-response scheduling, that yields a *linear* optimization problem and ii) demonstrate the implementation of these concepts with an industrial case study. Specifically, the novel contributions are:

- using autoregressive with extra inputs (ARX) models to represent the closed-loop process dynamics
- defining a representation of time that affords incorporating multiple such ARX models of arbitrary order in production scheduling models
- an industrial application considering a three-product air separation unit (ASU), where a large-scale set of operating data is used to identify the relevant models, and the solution of the DR scheduling problem results in significant operating cost savings.

While here we propose formulating the demand response problem as a linear program (LP), the ARX models introduced in this paper can be incorporated in other scheduling problem structures (including those with discrete decisions), and have the benefit of being represented by a set of linear constraints.

The paper is organized as follows: we begin with a process description of the industrial ASU, followed by a review of the relevant literature for DR scheduling. We then present the LP formulation of the scheduling problem, supported by the ARX representation of the process dynamics. Subsequently, we discuss the application to the industrial ASU, complete with optimization results and a discussion of the economic benefits of DR engagement.

2 Background and Process Description

Air separation units (ASU) are a key element in the supply chain of many industries, as suppliers of nitrogen, oxygen, argon and other rare gases. Cryogenic ASUs make up the bulk of air separation technologies due to their ability to output a high volume of high purity product. ASUs are electricity intensive, with the high power consumption due to the use of electric compressors. Cryogenic technology requires that the feed air stream be compressed to enable liquefaction; depending on the application, compression may be needed for the gaseous products as well.

The industrial process considered here is shown in Figure 1, and is designed to generate five products: gas nitrogen (GN2), liquid nitrogen (LN2), gas oxygen (GO2), liquid oxygen (LO2), and Argon (Ar). Gas and liquid forms of oxygen and nitrogen are considered as separate products since, i) they have vastly different production costs (producing a cryogenic liquid carries a significant energy penalty) and ii) they originate in different parts of the plant, as discussed below.

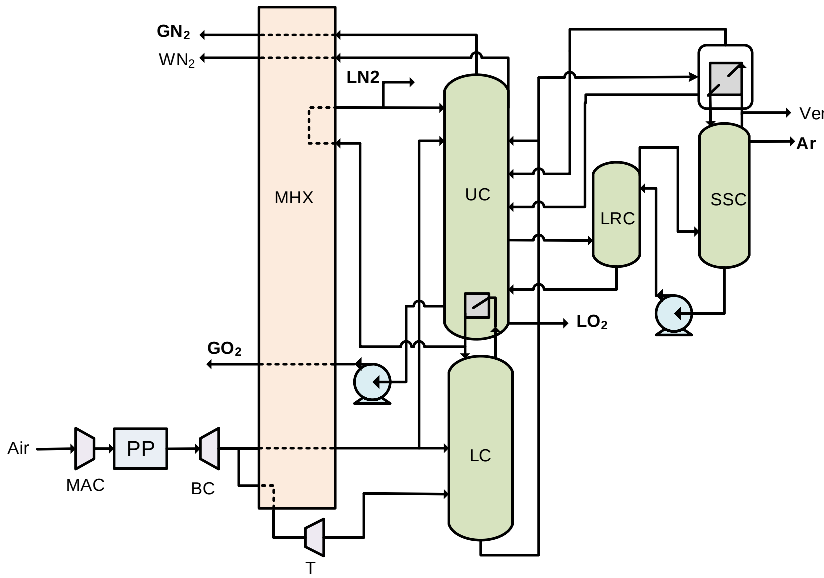

Figure 1: Overview of the industrial ASU studied in this paper, based on Cao et al. [5].

The inlet air enters the process through the main air compressor (MAC), driven by an electric motor. The prepurifier (PP) consists of a pressure swing adsorption (PSA) system that removes impurities (e.g. water, hydrocarbons, CO_2) before the air is compressed again in the booster compressor (BC). From here, the air stream is separated into two streams entering the the multistream heat exchanger (MHX), both of which are cooled alongside warming cryogenic products. After cooling, one air stream is split, with one portion entering the upper column (UC) and the other entering the high-pressure lower column (LC). The other portion of the feed stream is expanded in a turbine (T) to generate additional refrigeration before entering the LC. In the LC, the air is separated into an oxygen-rich bottoms stream and a nearly pure nitrogen stream at the top (shelf gas). The shelf gas is used to provide reboiler duty for the upper column (UC) before it is split into a return stream (for reflux into the LC) and a stream entering the MHX. From here, some of the nitrogen is removed as product (LN2) and the rest serves as the reflux stream in the UC. The bottoms product from the LC is used as a condensing stream for the argon super staged column (SSC) prior to being passed onto the UC. A "waste" stream of nitrogen (WN2) is removed near the top of the UC and utilized as an additional cooling stream for the MHX before being vented to the atmosphere. Purified oxygen product is removed from the bottom of the UC, with a portion of it remaining liquid product (LO2) and the rest vaporized in the MHX. A pump is needed to pressurize the liquid in order to obtain a high pressure GO2 product (rather than relying on an O2 compressor). Argon accumulates towards the bottom of the UC and is removed and sent to the low ratio column (LRC) and then the SSC. A high-purity argon stream is obtained at the top of the SSC.

In the context of an ASU, DR amounts to utilizing chemical products (oxygen, nitrogen and argon (to a lesser extent)) as a medium for storing electricity, such that production rate (and thereby electricity usage) is increased during off-peak times and extra product is sent to storage. ASUs generally meet the requirements for DR participation laid out above in terms of dynamic agility, production rate flexibility and product storage capacity.

As is the case with many chemical plants, the dynamics of ASUs are typically slow, with dominant time constants in the order of several hours. As a consequence, DR production scheduling models must explicitly account for the dynamics of the process [6, 7, 8].

An accurate first-principles representation of an ASU's dynamics requires nonlinear models which, in many practical situations, are large-scale (with thousands of variables and equations) and poorly conditioned. Their integration in optimal scheduling models adds considerable computational complexity and makes the corresponding optimization problems difficult to solve in a useful amount of time.

The development of computationally-tractable dynamic models has received extensive attention in the context of ASUs involving nitrogen as a product [9, 5, 10, 3, 11]. For multiproduct ASUs (O_2, N_2 and Ar), limited work is available; this is in no small measure due to the complexity of the process itself, and in the difficulty to capture its dynamics accurately. Cao et al. [5] utilized reduced-order dynamical modeling, and Misra et al. [12] and Zhang et al. [13] used operating points which were quasi-stationary, effectively reducing model complexity. Mitra et al. [14] utilized a state task network to model constraints on production, which was later extended by Zhang et al. [15]. Such studies have enabled application of similar modeling paradigms to clusters of ASUs operating together in service of a single site [16, 17]. Zhou et al. [17] used convex hulls to represent a series of operating modes observed in historical data from which an optimal scheduling problem could pick. Zhao et al. [16] simultaneously scheduled two ASUs using a state transition network model. Basán et al. [18] also utilized state transitions, enhancing existing work by having multiple transition states to better capture transition times and employing such representations on an ASU scheduling problem with a moving horizon. Kender et al. [19] demonstrated the applicability of boundary scenarios for determining if shutdown and restart are beneficial in terms of flexible operation. We note that these works typically do not represent the dynamics of the process explicitly. Rather, approximations such as rate of change constraints are employed.

To our knowledge, previous literature references that consider explicitly the dynamics of a multi-product ASU in the DR scheduling calculation is very limited: Caspari et al. [20] applied economic MPC (eNMPC) to an ASU, Dias et al. [10] proposed a simulation-based optimization construct that integrated MPC in production scheduling, and Tsay et al. [21] utilized historized operational data from an industrial ASU to identify Hammerstein-Wiener dynamic models that were then embedded in a DR scheduling calculation Kelley et al. [4].

The participation of ASUs in DR programs has strong economic incentives; for example, Tsay et al. [21] demonstrated an operating cost savings potential of 13% in real-time markets and 3% in day-ahead markets for a multiproduct ASU, relative to steady-state operation. Thus, ensuring that optimal DR production scheduling calculations are computationally tractable in a practical amount of time is key to broadening the participation industrial entities in such programs; this requirement is typically met by developing techniques for reducing model size and complexity. Below, we describe the rationale for doing so, and review existing results both in the general case and as they pertain to ASUs.

Computationally tractable DR scheduling formulations generally rely on low-order representations of the scheduling-relevant closed-loop dynamics of the process. Pattison et al. [3] found that a small subset of the process variables can be used to represent the dynamics in the scheduling problem. These variables are *scheduling-relevant*, and they are (i) at or near their constraints during normal operation, and/or (ii) are in the scheduling objective function (e.g., they affect the operating cost of the process). The models used to represent these scheduling-relevant variables bridge the time scales between the (long-term) scheduling calculation and (relatively shorter term) process dynamics/process control, and will be referred to as "scale-bridging models" (SBMs). Focusing particularly on DR scheduling of an industrial air separation unit (ASU), Pattison et al. [3] demonstrated that using the data-driven SBMs brings significant computational benefits. An extensive case study showed reductions in computation time of two orders of magnitude, compared to using a first-principles model. Further reductions in computation time for DR scheduling of an ASU were achieved by Kelley et al. [22] via an exact linearization of a class of SBMs, which rendered the DR scheduling problem a mixed-integer linear program (MILP) that could be solved in a matter of minutes.

3 Production Scheduling for Demand Response

Production scheduling for DR operation seeks to exploit time-sensitive electricity prices (an example is shown in Figure 7), aiming to minimize operating cost while, (i) meeting product demand and, (ii) abiding by all process constraints related to, e.g., product quality and safety. In the context of chemical process operations, DR scheduling is part of a broader decision-making structure illustrated in Figure 2.

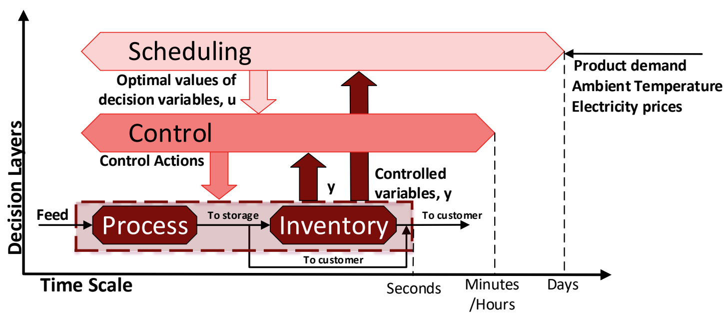

Figure 2: Hierarchy of decisions for chemical process operation. u are the production setpoints/targets and y are the values of the controlled variables.

The role of the (DR) scheduling layer is to provide a time sequence of production targets (e.g. a production rate, defined in terms of setpoints $\bar{u}$ for the control layer) in response to fluctuations in product demand and electricity prices. The problem can be described as:

$$\begin{aligned} \min_{\bar{u}} J = \quad & \text{Operating Cost} \\ \text{s.t.} \quad & \text{Process dynamic model} \\ & \text{Storage system model} \\ & \text{Power consumption model} \\ & \text{Process constraints} \\ & \text{Demand satisfaction constraints} \end{aligned} \tag{1}$$

The setpoint sequence produced by the scheduling layer must be *dynamically feasible*, in the sense that it can be implemented/tracked by the control system without violating any of the process constraints over the time horizon considered. As a consequence, the process dynamic model in (1) must include the dynamics of the process itself, the storage system, *and* the dynamics of the controller, i.e., provide a representation of the *closed-loop* dynamics of the process.

4 A Linear Programming Formulation of the DR Scheduling Problem

Du et al. [23] introduced the notion of Scale-Bridging Model (SBM) as a (low-order) representation of the *closed-loop* dynamics of the process that are relevant to scheduling calculations. Scheduling-relevant variables can be defined as variables included in key process, quality, and demand constraints as well as variables affecting the cost of operation. In the original developments described by Du et al. [23], an explicit expression of the closed loop dynamics was available owing to the use of a nonlinear input-output linearizing controller. On the other hand, Pattison et al. [3] advocated the use of data-driven SBMs, obtained from closed-loop operating data via system identification techniques. In previous works, we have demonstrated the applicability of (nonlinear) Hammerstein Wiener (HW) models for slower dynamics, and (some) linear finite step response (FSR) models for the faster dynamics [22, 24, 25].

While beneficial from a solution time point of view, these approaches resulted in optimization problems that were either in MIDO (mixed-integer dynamic optimization) or MILP formats. Motivated by the above, a key contribution of the present work is to introduce the use of Auto Regressive with eXtra inputs (ARX) model as SBMs for DR scheduling calculations. The key advantage of ARX model is their linearity, which, as we will demonstrate below, results in a *linear* formulation of the DR scheduling problem.

4.1 Model Format

It is typical for SBMs to take a multiple input, single output (MISO) form, where the evolution of a single scheduling-relevant variable is predicted based on multiple variables (which could be inputs that can be deliberately defined by operators, such as production setpoints/targets, or disturbances such as ambient temperature). Thus, we consider ARX models of the form:

$$A(q)y(t) = \sum_{s=1}^{n_u} B_s(q)u_s(t - n_{k_s}) + e(t) \tag{2}$$

where $A(q)$ is a polynomial of order n_a corresponding to the number of poles in the model, expressed in time shift operator form, q^{-1}. $B(q)$ is a polynomial of order n_b corresponding to the number of zeros, $y(t)$ is the model output, $u(t)$ is the model input, n_u is the number of inputs to the model, n_k is the delay, and e(t) is a term assumed to be white noise. Graphically, an ARX model with four inputs is represented in Figure 3.

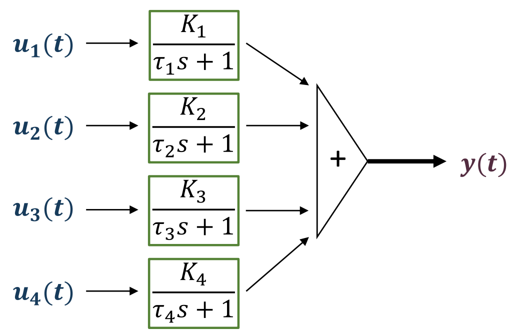

Figure 3: Representation of a MISO ARX model in (2), where u are the production set-points/targets, y are the values of the controlled variables, and K and τ are model parameters. Here, first order transfer functions are used for illustration purposes.

4.2 Representation of time

We define the DR scheduling problem based on a discrete representation of time. The decision making hierarchy of DR scheduling (Figure 2) evinces three distinct time horizons, which we account for in our definition of the time grid. First, we consider a scheduling horizon of N_D days, represented by index d. Second, within each day there are N_I scheduling time slots of length T_I, represented by index i. Third, within each scheduling time slot, there are N_J time points of index j, with length T_J, which consist the sample times (which are T_J apart) for the ARX models used to represent the process dynamics. A schematic of this time representation is shown in Figure 4.

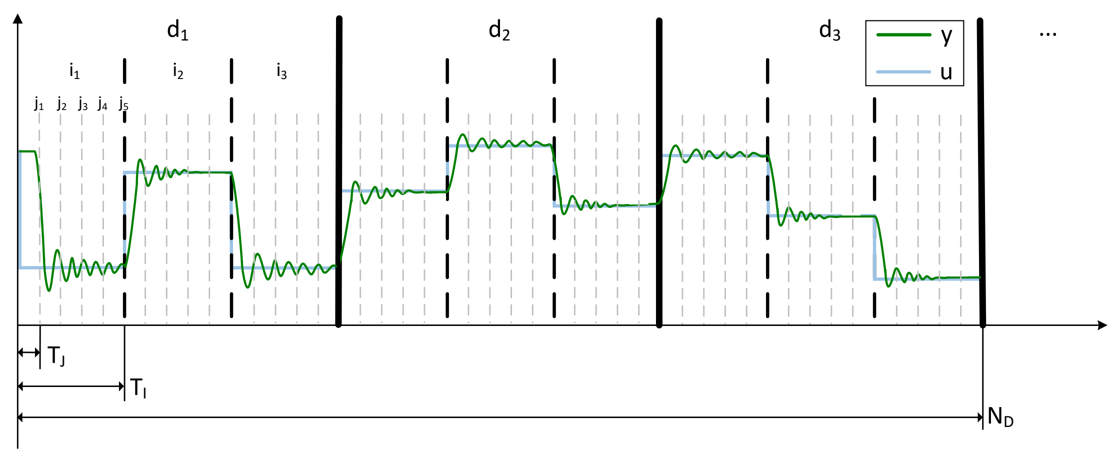

Figure 4: Depiction of the timescales, showing days, d, in the time horizon, scheduling slots, i, and sample times, j, of the dynamic model, along with the trajectory of a scheduling-relevant variable, y in response to changes in an input u. Here, u is assumed to be the setpoint of a controller whose controlled variable is y, and hence y tracks the values of u.

4.2.1 State Continuity Constraints for ARX SBMs

Figure 4 illustrates an important feature of DR scheduling, and a consequent key requirement for SBM representations of any kind. Specifically, the outputs y are physical variables that reflect the state of a continuous system; as such, they are continuous in time. This continuity must be preserved by their models, even when the inputs u are *discontinous*. DR scheduling formulations are often predicated on assuming that the inputs u are piecewise constant signals, whose values change in a stepwise fashion in every scheduling time slot.

In view of these observations, the use of ARX structures as SBMs requires several adjustments to the scheduling problem. These are explained by considering a scalar ARX model with input u_i and output y_i in time slot i, having na poles and nb zeros.

$$A_1 y_{i,j} + A_2 y_{i,j-1} + ... + A_{na+1} y_{i,j-na} = B_1 u_{i,j} + B_2 u_{i,j-1} + ... + B_{nb+1} y_{i,j-nb} + e_{i,j} \tag{3}$$

Imposing state continuity when using this ARX model in the DR scheduling problem calls for a particular set of continuity conditions that link the value of the state across scheduling slots i. To this end, we introduce a new parameter, N_F, as is the number of *functional* timepoints within each scheduling slot, i.e. time points which are not used specifically for the purpose of imposing continuity conditions for each SBM. As an example, if the length of each scheduling slot, i, were one hour, and the sample time T_J were one minute, the number of functional time points N_F within the scheduling slot would be 60.

The continuity constraints are related to the structure (poles and zeros) of the model (3), such that the last time points (j) of scheduling timeslot $i-1$ overlap with the first time points of scheduling timeslot i. In the problem formulation, some extra time points, j, referred to as the *continuity tail*, are added to each time slot, in excess of N_F. Thus, the total number of time points, N_J per scheduling slot, i, can be determined by the following relation:

$$N_J = N_F + \max(n_a, n_b) \tag{4}$$

Because the length of the continuity tail depends on the model order, and a scheduling problem is likely to feature multiple ARX SBMs of different orders, it is possible that each SBM (for each scheduling-relevant variable) will have a continuity tail of different length. Uniformity in the number of time points throughout the scheduling problem is imperative to having a cohesive and generic formulation. Therefore, the number of timepoints per time slot, N_J, for the *overall* scheduling problem is set to the highest N_J (calculated by (4)) for all the ARX models. Any timepoints in excess of what is needed for each ARX SBM are left unused; as this problem is implemented in modeling software, the corresponding variables and equations will be eliminated at the preprocessing step. Figure 5 illustrates the concepts of functional time points (4) and continuity tail for a case where to ARX models, of second and third order, respectively, are used.

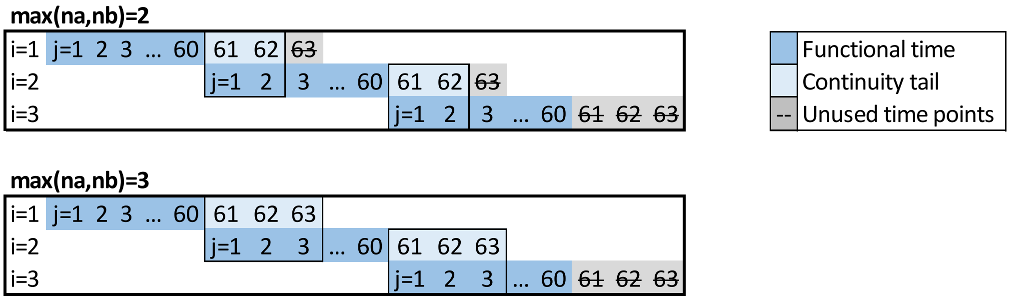

Figure 5: Illustration of the continuity tail, functional time, and unused time for an optimization problem including a second and third order ARX model where $N_I = 3$, $N_J = 63$, $N_F = 60$. Each model has the same number of functional time points.

Based on the continuity tail structure introduced above, we can define the continuity equations between scheduling time slots:

$$y_{i-1,N_J-n_a+j} = y_{i,j} \qquad \forall i > 1 \; and \; j = \{1, ..., n_a\} \tag{5}$$

$$u_{i-1,N_J-n_a+j} = u_{i,j} \qquad \forall i > 1 \; and \; j = \{1, ..., n_a\} \tag{6}$$

To stipulate that inputs may only change once per scheduling timeslot, i, we impose an additional constraint for the inputs (*within* each time slot), u, not already included in the input continuity constraints:

$$u_{i,j} = u_{i,j-1} \; \forall i > na \tag{7}$$

Relationships (5)-(6) and (7) are used in place of merely indexing the input u over i (rather than over both i and j) for the case where one input u is used for multiple SBMs of different orders and therefore data for *all* time points (continuity tail and functional) must be included.

Note that this framework easily accommodates upper (U^U, Y^U) and lower (U^L, Y^L) bounds on process inputs, u, and scheduling-relevant variables, y, which may be required to capture constraints related to process safety, equipment capacity and product quality.

$$Y^L \leq y_{i,j} \leq Y^U \tag{8a}$$
$$U^L \leq u_{i,j} \leq U^U \tag{8b}$$

One of the challenges that arises when using linear models and linear objective functions in optimal control (and scheduling) problems is that they result in “bang-bang” solutions, or the abrupt switching between the maximum and minimum [26] which can be difficult for the control-layer MPC to track. This tendency can be mitigated by imposing rate of change constraints on inputs of selected ARX models:

$$|u_{i,j} - u_{i-1,j}| \leq U^{\Delta} \tag{9}$$

5 Application to Industrial ASU

5.1 Data set and processing

The data set available for this effort consisted of recordings of all process variables stored in the process historian. The sample frequency was one minute and the data set spanned one year of operation, which included several plant startup and shut down events. At all relevant time points, the process was operating under model predictive control (MPC). No deliberate excitation was implemented other than changes in the setpoints of the MPC system imposed by the operators as part of normal operation.

Given the size of the data set and the measurement frequency, some preprocessing was required. First, the time frames covering start up and shut down events were removed based on process knowledge. Following this, a first-order Savitzky-Golay filter [27] with a frame length of 15 samples was applied, with the aim of smoothing the noisy data.

Subsequently, outliers were removed from the filtered data. Outliers were defined as any points whose value was more than 3 scaled median absolute deviations (MAD*) away from the median value (10). The MAD is scaled using factor c, which is calculated using the inverse complementary error function (erfc^{-1}).

$$\mathrm{MAD} = median(|y_i - median(y)|)\ \forall i = 1, 2, ..., N \tag{10a}$$

$$\mathrm{MAD}^* = c \times \mathrm{MAD} \tag{10b}$$

$$c = \frac{\text{-1}}{\sqrt{2}\mathrm{erfc}^{-1}(\frac{3}{2})} \tag{10c}$$

The amount of outlier data points removed for each model is given in Table 1. Note that each percent value provided is relative to the number of points in the “full” filtered data set, i.e., the process for fitting the models began with the filtered data set, the relevant input and output data series were selected, and the aforementioned outlier removal procedure was applied to these selected series. The process was repeated for each model.

For select variables, additional filtering was needed due to periodic (and relatively high frequency) fluctuations caused by the switching events of the pressure-swing adsorption (PSA) cycles. For these variables, a simple moving average filter was applied, with the span equal to the period of fluctuation for the PSA.

5.2 Identification of ARX scale-bridging models

Ten scheduling-relevant variables were selected using the criteria outlined by Pattison et al. [3] combined with domain knowledge. These are the product flowrates (F_{GN2}, F_{LN2}, F_{GO2}, F_{LO2}, and F_{Ar}), feed air flowrate (F_{Air}), Power consumption ($P_i, i \in \{1, 2, 3\}$) for the two compressors and an internal motor, and an internal flow rate (C_F) essential to process and quality safety. We defined one measured disturbance, ambient temperature, T. The ARX model parameters were estimated in MATLAB [28] using the arx function, with the data set being divided equally into training and test data sets. The quality of the models is characterized using the Normalized Mean-Square Error (NMSE) (11), where y are the observed data, and $\hat{y}$ are the model predictions. Table 1 gives the model orders and the resulting (unscaled) NMSE values.

$$\mathrm{NMSE} = \frac{||y - \hat{y}||_2^2}{||y||_2^2} \tag{11}$$

Table 1: ARX model details for the scheduling-relevant variables, where the overbar denotes an MPC setpoint and $\phi(\cdot)$ denotes a function of its inputs/arguments; the forms of these functions are not presented to protect confidential industrial information. The percentage of outliers removed from the dataset for each model is also given.

Variable	Input	Outliers (%)	na	nb	NMSE (training)	NMSE (test)
P_1	PC_1, PC_2, T	3.9	2	2	7.8561e-09	8.2419e-09
P_2	PC_1, PC_2, T	4.0	2	2	1.8987e-09	1.6896e-09
P_3	PC_1, PC_2, T	3.9	2	2	2.2687e-08	2.4331e-08
F_{GN2}	$\phi(\bar{F}_{Air})$	4.1	3	3	7.8075e-08	5.9252e-08
F_{LN2}	$\bar{F}_{LN2}$	1.9	3	3	4.2225e-10	3.5797e-10
F_{GO2}	$\bar{F}_{GO2}$	2.4	3	3	4.988e-10	4.551e-10
F_{LO2}	$\phi(\bar{F}_{Air}, \bar{F}_{GO2})$	18	1	1	1.1259e-07	1.7783e-07
F_{Ar}	$\bar{F}_{Ar}$	1.4	3	3	4.0058e-09	1.8106e-09
F_{Air}	$\bar{F}_{Air}$	4.3	3	1	1.1312e-10	6.6157e-11
C_F	$\phi(\bar{F}_{Air})$	12	3	3	1.0325e-09	9.2544e-10

A number of the scheduling-relevant variables were controlled by the process MPC (F_{GN2}, F_{Air}, F_{GO2}, F_{Ar}) and their respective setpoints ($\bar{F}_{GN2}$, $\bar{F}_{Air}$, $\bar{F}_{GO2}$, $\bar{F}_{Ar}$)were used as inputs to the models. For the power consumption models, principal component analysis (PCA) was used on the plant flow rates to reduce the number of inputs to the power models, which also incorporated measured disturbance, T. The first two principal components, PC_1 and PC_2 were used as inputs as they captured 94.9% and 3.48% of the variance, respectively. For variables without direct setpoints, such as F_{LO2}, F_{LN2}, and C_F, linear functions based on other setpoints were used as model inputs.

Plots of the responses of the ARX models are shown in Figure 6 to demonstrate visually the model fit to the training and test data as well as the training/test split for each variable. The data have been scaled using the corresponding upper and lower bounds from the optimization problem (discussed later) to preserve confidentiality.

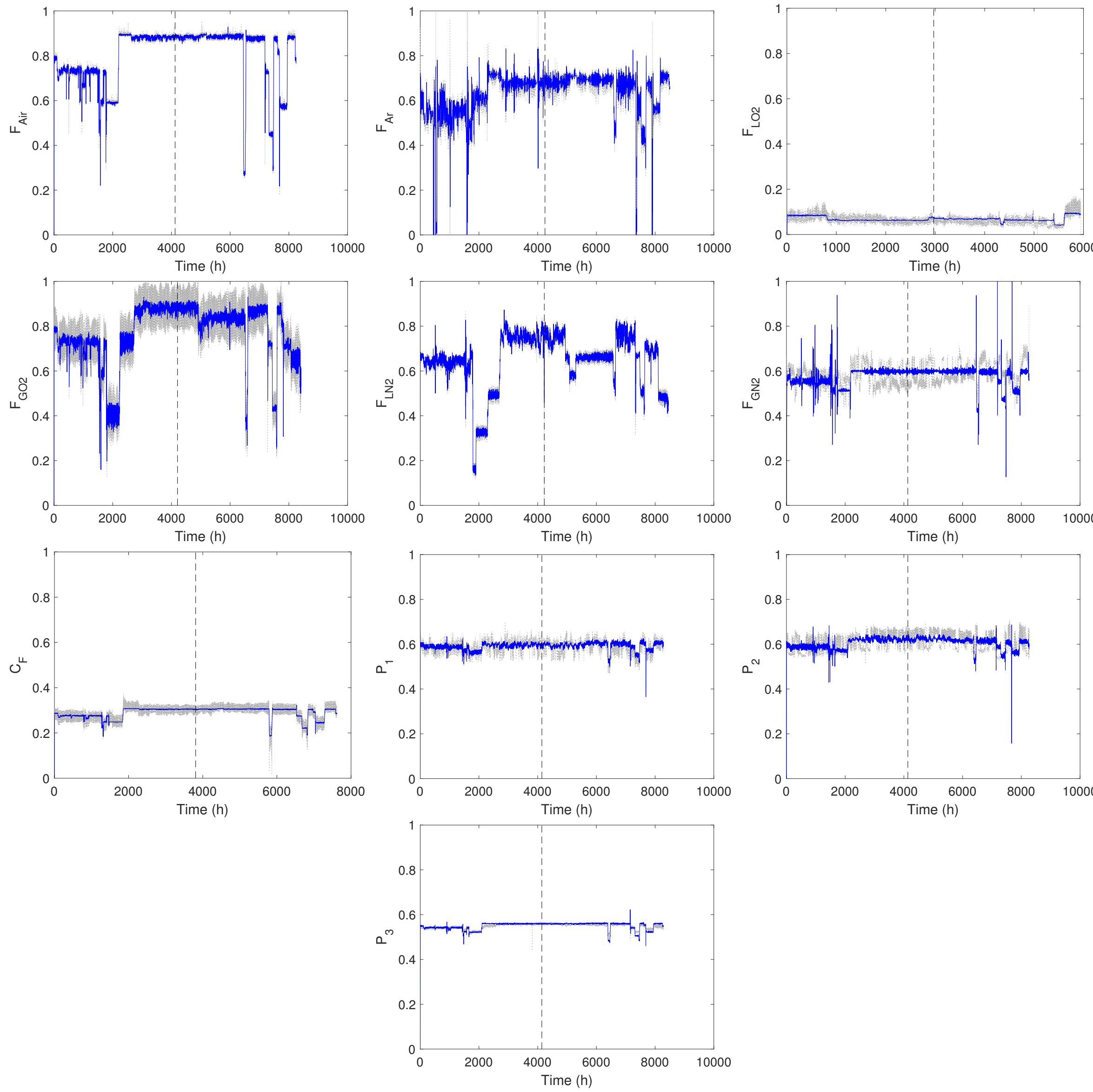

Figure 6: Plots of the data and ARX scale bridging model predictions, where the grey dashed lines correspond to the data, the blue dashed lines correspond to the predictions, and the vertical black dashed line serves as a delimiter between the testing and the training data. Data have been scaled to preserve confidentiality.

A key challenge related to using data-driven models in an industrial context is that they will inherently include measurement errors or biases from plant sensors. These biases will then be reflected in solutions of the DR scheduling problem (which will be described in detail later). The immediate consequence is that these solutions may not be physically meaningful, in the sense that the material balance may not close. To prevent this, we developed the following heuristic approach. First, we verified the overall and component balances based on the available data, and computed the relevant errors. Subsequently, we developed additional ARX models that predicted the material balance error as a function of the relevant scheduling inputs. Subsequently, these models were included in the DR scheduling problems as additional constraints, to enforce closure of the material balance. Due to confidentiality reasons, we cannot present the explicit forms of these models, but we flag the use of this approach in order to provide a complete account of using data-driven SBMs to represent industrial entities for optimal (DR) scheduling problems, and in the interest of supporting further implementations.

5.3 Product storage and demand modeling

The plant includes storage tanks for LN2 and LO2, from which product can be removed to meet demand during peak electricity pricing hours or where excess product can be diverted when production rates are increased during off-peak hours. The amount of gas product, $p = \{GO2, GN2\}$, removed from storage ($S^{out}_{p,i,j}$) is governed by:

$$S^{out}_{p,i,j} + F_{p,i,j} \geq D_{p,i} \tag{12}$$

where $F_{p,i,j}$ is the flow of product p from the process at time points (i, j) and $D_{p,i}$ is the demand of product p during scheduling slot i. To represent inventory holdup, we utilize a simple integrator:

$$S_{p,i,j+1} = (S^{in}_{p,i,j+1} - S^{out}_{p,i,j+1}) + S_{p,i,j} \tag{13a}$$

$$S^{in}_{p,i,j} = F_{p,i,j} \tag{13b}$$

$$S_{p,1,1} = S^{o}_{p} \tag{13c}$$

where the flow into storage is equivalent to the liquid product from the process (LN2, LO2). Due to the discretized nature of (13a), we impose a continuity constraint:

$$S_{p,i,N_J} = S_{p,i+1,j=1} \tag{14}$$

such that the inventory holdup at the end of timeslot i is equal to the state at the beginning of timeslot $i + 1$. We also impose an *endpoint* constraint on the inventory holdup such that the inventory at the end of the time horizon is greater than or equal to the starting inventory. This constraint is to ensure there is no system depletion.

$$S_{p,i,N_J} \geq S^{o}_{p} \tag{15}$$

We do not consider Ar storage (at least as it pertains to DR operation) in our scheduling model. While Ar is an energy intensive product, the time constant of the process dynamics involving modulating Ar production is far too long for such modulation to be practical for demand response in the day-ahead electricity market. Nevertheless, in order to account for the fact that the Ar product flow rate does fluctuate as other production changes are implemented in the plant, we require that the *average* Ar production be equal to Ar demand. This allows for small fluctuations in F_{Ar} to occur in order to ensure that the mass balance does close while the other streams fluctuate (particularly F_{Air}) to reduce electricity cost. The corresponding constraint is formulated as:

$$\frac{\sum_{i,j} F^{Ar}_{i,j}}{N_I N_J} \geq D_{Ar} \tag{16}$$

5.4 Demand response scheduling problem

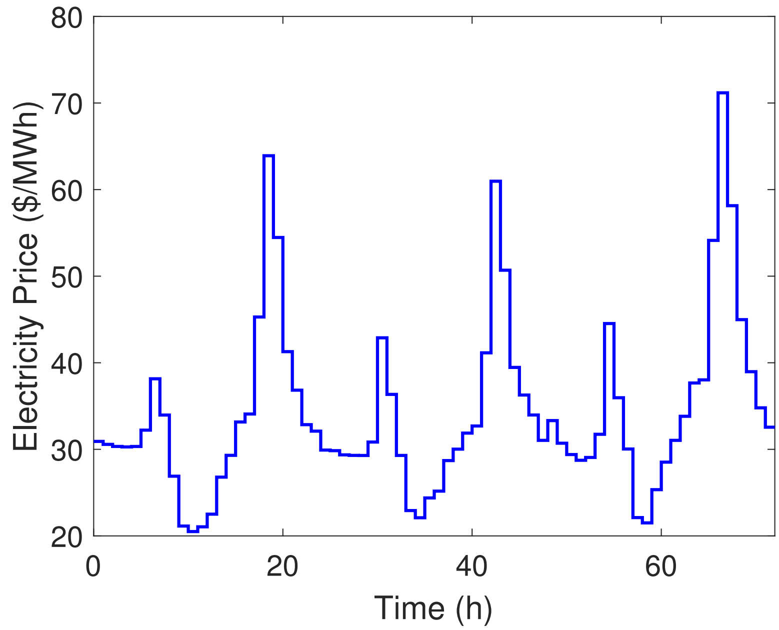

Figure 7: Hourly electricity prices for an aggregate node in Fresno, California during a 72-hour period in September 2020, provided by the California Independent System Operator (CAISO) [29].

The DR scheduling problem follows the format of (1):

$$\begin{aligned} \min_{u_{i,j}} \quad & \sum_{i=1}^{N_I} \sum_{j=1}^{N_J} \sum_{n=1}^{3} P_{n,i,j} C_i \\ \text{s.t.} \quad & \text{Process dynamic models (3)-(7)} \\ & \text{Storage system model (12)-(15)} \\ & \text{Process constraints (8), (9)} \\ & \text{Demand satisfaction constraints (12),(16)} \end{aligned} \tag{17}$$

where C_i is the hourly electricity price shown in Figure 7 and $P_{n,i,j}$ is the power consumption of the three compressors (P_1, P_2, and P_3) as predicted by the ARX SBMs described earlier. We have imposed an ROC constraint (based on (9)) on the Argon flowrate (F_{Ar}) to ensure the MPC layer is able to track it.

$$|\bar{F}_{Ar_{i,j}} - \bar{F}_{Ar_{i-1,j}}| \leq \bar{F}_{Ar}^{\Delta} \tag{18}$$

The scheduling-relevant variables were constrained to lie between the upper/lower bounds (8) of the industrial data used to fit the models in Table 1, to ensure that the resulting DR schedule does not exceed the domain of validity of the scale-bridging models.

6 Results

The DR scheduling problem (17) was implemented in GAMS 33.2.0 [30] and solved using CPLEX 12.10.0 [31] on a 64-bit Windows system with Intel Core-7-2600 CPU at 3.40 GHz and 16 Gb RAM.

The resulting problem size was 224,652 equations and 184,568 variables. Ambient temperature T (which is a model input as a measured disturbance), was assumed to be constant. The case of steady-state operation of the plant serves as the baseline for estimating the cost savings derived from DR operation (i.e., from solving problem (17)). The solution indicates that for the price profile and three day time considered, DR operation yields a 8.91% reduction in operating cost, with a solution time of 1.98 min. Figure 8 shows the optimal (scaled the same way as the original data in Figure 6) trajectories for the scheduling-relevant variables (as well as the relevant setpoints) listed in Table 1. We note here that while the case study is centered on an “offline” solution of the DR problem, it is likely that this problem would in practice be solved online in a moving horizon fashion, as price information and process state data are updated [24]. Nevertheless, given that the optimization problem is an LP, we expect that the solution times would be similar to those presented here for other sets of price and process data values. We also note that these solution times compare favorably with those reported in our previous works [22, 24, 25], which relied on mixed-integer models of ASUs that were less complex (and had lower-dimensional models) than the plant considered in this work.

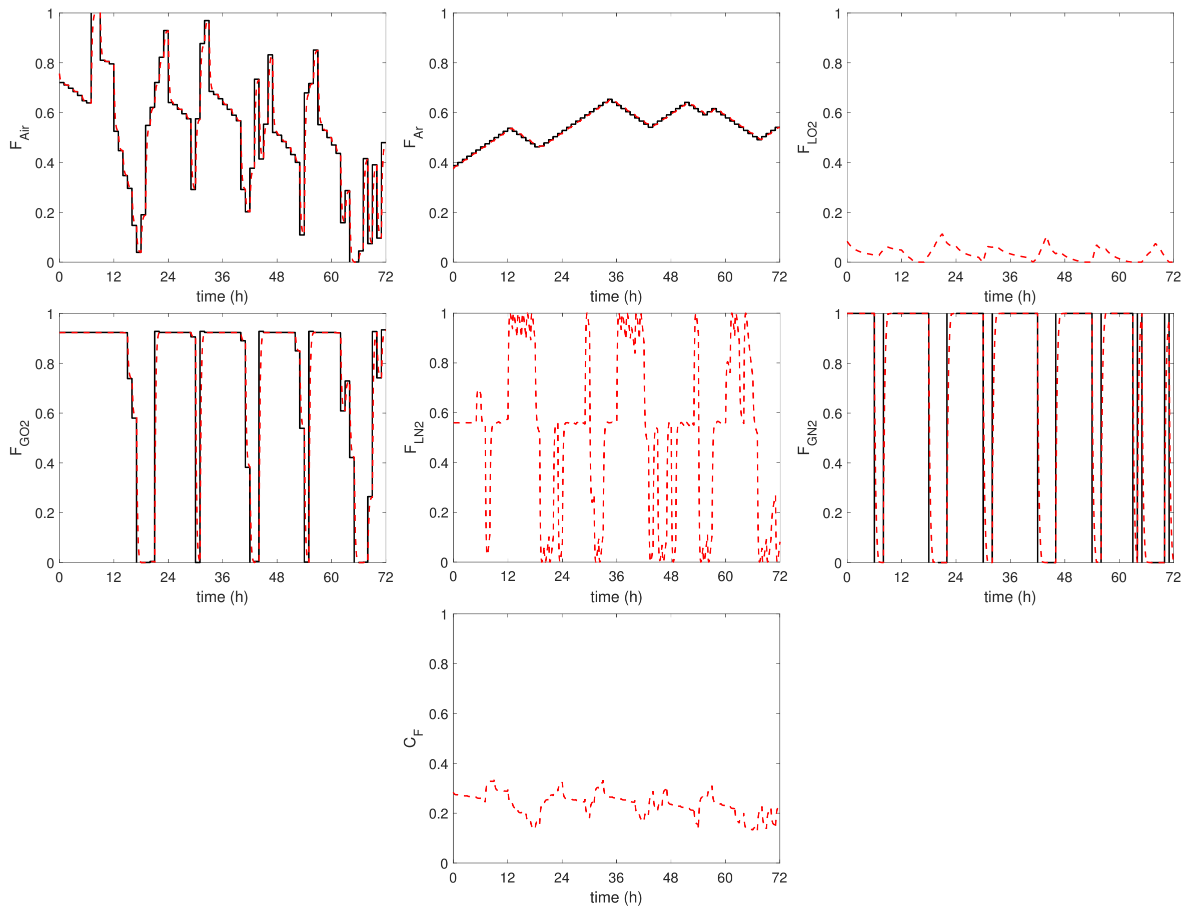

Figure 8: Plots of the optimal trajectories of scheduling-relevant variables from the solution of (17), where the black solid line represents the optimal setpoint (MPC input) trajectory, and the dashed red line represents the variable response calculated by the ARX model. Data have been scaled to preserve confidentiality.

The flowrates in Figure 8 fluctuate in time, where the lower flow rate values align in time with the peak pricing hours shown in Figure 7. The Argon flowrate (F_{Ar}) is limited by the rate-of-change constraint imposed (18), as is evinced by the "stair step" pattern.

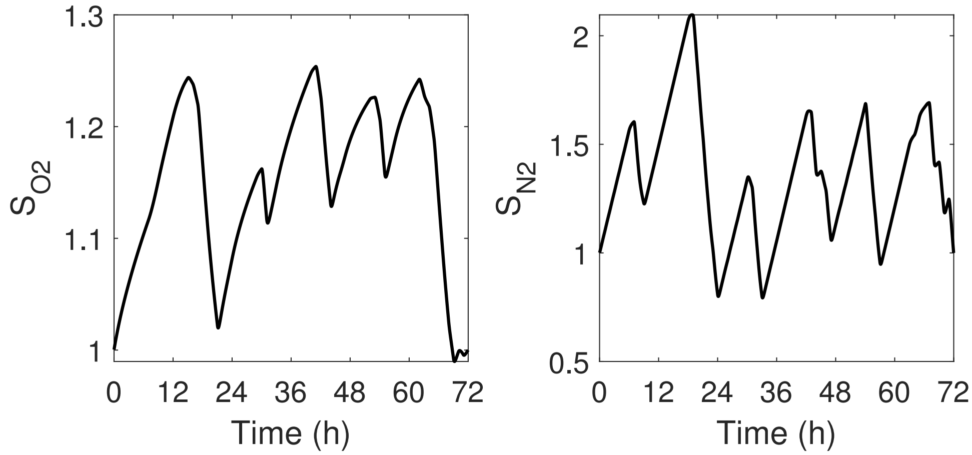

Figure 9: Inventory levels corresponding to the optimal schedule.

Figure 9 shows the dynamics of the inventory in the storage tanks, scaled such that the initial value is set at 1 and the y-axis shows a fractional increase based on the initial value. As shown in the figure, the endpoint constraints (15) are satisfied. The nature of the plant operations is such that there is a non-zero flow of LO2 and LN2 into the storage system at almost all times, the storage holdup rarely drops below its initial inventory level. Akin to the trends noticed in the flow rates of Figure 8, the storage holdup fluctuates inversely with the electricity prices in Figure 7.

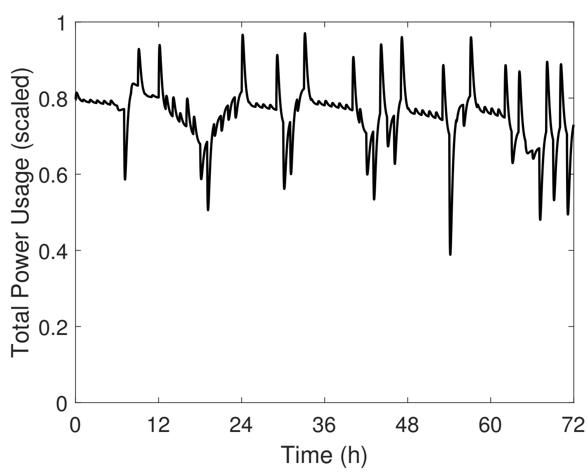

Figure 10: Total power consumption for the DR scheduling problem (17), scaled using the combined upper/lower bounds from P_1, P_2, and P_3 in the data.

Lastly, we show the total power consumption in Figure 10, which peaks during off-peak pricing hours and lowers during peak hours. The total power shown in the plot is a sum of the three power outputs in Table 1.

7 Conclusions

In this work, we solved a demand response (DR) optimal scheduling problem under time-sensitive electricity price for an industrial air separation unit (ASU). The DR scheduling scheme reported here has potential to increase grid reliability by responding to fluctuations in electricity prices with counteractive electricity consumption rates. Specifically, the production rate of the ASU is raised above demand during off-peak pricing hours and extra product (nitrogen) is sent to storage for later use during peak pricing hours when production rate is lowered below demand and product from storage is sent to customers. To include dynamic considerations in the optimal scheduling problem, scale-bridging models (SBMs) were used, which have significant computational benefits and thus open the door for consideration of complex industrial processes here.

The central contribution of this work is the utilization of data-driven autoregressive with extra inputs (ARX) SBMs identified from historical operating data for representation of a true industrial process. These models were incorporated into a DR scheduling problem. While the LP formulation of the scheduling problem using ARX models is complex in the way of continuity constraints and timescales, the computational time advantages of LPs far outweigh the additional complexity introduced by ARX continuity constraints. Namely, the DR scheduling problem solved in 1.98 minutes for a 3-day time horizon. Furthermore, this study demonstrated a significant cost-reduction potential for DR operation for ASUs; for the scenario considered, a cost reduction of 8.91% was possible. While this value may vary from day to day as electricity prices fluctuate, it provides a strong argument in favor of industrial DR participation.

8 Acknowledgments

This material is based upon work supported by the U.S. Department of Energy under Award DE-EE0007613. M.T.K. was supported by a Department of Energy Computational Science Graduate Fellowship (Award DE-FG02-97ER25308).

9 Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

[1] US EIA. Monthly Energy Review March 2019. Technical report, Washington, D.C., 2019. URL https://www.eia.gov/totalenergy/data/monthly/pdf/mer.pdf.

[2] C. Tsay and M. Baldea. 110th Anniversary: Using Data to Bridge the Time and Length Scales of Process Systems. *Industrial & Engineering Chemistry Research*, 58(36):16696–16708, sep 2019. ISSN 0888-5885. doi: 10.1021/acs.iecr.9b02282. URL https://pubs.acs.org/doi/10.1021/acs.iecr.9b02282.

[3] R. C. Pattison, C. R. Touretzky, T. Johansson, I. Harjunkoski, and M. Baldea. Optimal Process Operations in Fast-Changing Electricity Markets: Framework for Scheduling with Low-Order Dynamic Models and an Air Separation Application. *Industrial & Engineering Chemistry Research*, 55(16):4562–4584, apr 2016. ISSN 0888-5885. doi: 10.1021/acs.iecr.5b03499.

[4] M. T. Kelley, R. C. Pattison, R. Baldick, and M. Baldea. An efficient MILP framework for integrating nonlinear process dynamics and control in optimal production scheduling calculations. *Computers & Chemical Engineering*, 110:35–52, feb 2018. ISSN 00981354. doi: 10.1016/j.compchemeng.2017.11.021. URL http://linkinghub.elsevier.com/retrieve/pii/S0098135417304222.

[5] Y. Cao, C. L. E. Swartz, J. Flores-Cerrillo, and J. Ma. Dynamic modeling and collocation-based model reduction of cryogenic air separation units. *AIChE Journal*, 62(5):1602–1615, may 2016.

[6] C. T. Maravelias and C. Sung. Integration of production planning and scheduling: Overview, challenges and opportunities. *Computers and Chemical Engineering*, 33(12):1919–1930, 2009. ISSN 00981354. doi: 10.1016/j.compchemeng.2009.06.007.

[7] I. Grossmann. Enterprise-wide optimization: A new frontier in process systems engineering. *AIChE Journal*, 51(7):1846–1857, jul 2005. ISSN 0001-1541. doi: 10.1002/aic.10617.

[8] M. Baldea and I. Harjunkoski. Integrated production scheduling and process control: A systematic review. *Computers & Chemical Engineering*, 71:377–390, dec 2014. ISSN 00981354. doi: 10.1016/j.compchemeng.2014.09.002.

[9] Y. Cao, L. E. Swartz, Christoper, and M. Baldea. Design for dynamic performance: Application to an air separation unit. In *American Control Conference*, pages 2683–2688. McMaster University, 2011.

[10] L. S. Dias, R. C. Pattison, C. Tsay, M. Baldea, and M. G. Ierapetritou. A simulation-based optimization framework for integrating scheduling and model predictive control, and its application to air separation units. *Computers and Chemical Engineering*, 113:139–151, 2018. ISSN 00981354. doi: 10.1016/j.compchemeng.2018.03.009. URL https://doi.org/10.1016/j.compchemeng.2018.03.009.

[11] P. Schäfer, A. Caspari, K. Kleinhans, A. Mhamdi, and A. Mitsos. Reduced dynamic modeling approach for rectification columns based on compartmentalization and artificial neural networks. *AIChE Journal*, 65(5):e16568, may 2019. ISSN 0001-1541. doi: 10.1002/aic.16568. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/aic.16568.

[12] S. Misra, M. Kapadi, R. D. Gudi, and R. Srihari. Energy-Efficient Production Scheduling of a Cryogenic Air Separation Plant. *Industrial & Engineering Chemistry Research*, 56(15): 4399–4414, apr 2017. ISSN 0888-5885. doi: 10.1021/acs.iecr.6b04585. URL https://pubs.acs.org/doi/10.1021/acs.iecr.6b04585.

[13] Q. Zhang, I. E. Grossmann, C. F. Heuberger, A. Sundaramoorthy, and J. M. Pinto. Air separation with cryogenic energy storage: Optimal scheduling considering electric energy and reserve markets. *AIChE Journal*, 61(5):1547–1558, may 2015. ISSN 00011541. doi: 10.1002/aic.14730. URL http://doi.wiley.com/10.1002/aic.14730.

[14] S. Mitra, J. M. Pinto, and I. E. Grossmann. Optimal multi-scale capacity planning for power-intensive continuous processes under time-sensitive electricity prices and demand uncertainty. Part I: Modeling. *Computers & Chemical Engineering*, 65:89–101, jun 2014. ISSN 00981354. doi: 10.1016/j.compchemeng.2014.01.016. URL http://dx.doi.org/10.1016/j.compchemeng.2014.01.016https://linkinghub.elsevier.com/retrieve/pii/S0098135414000283.

[15] Q. Zhang, A. Sundaramoorthy, I. E. Grossmann, and J. M. Pinto. A discrete-time scheduling model for continuous power-intensive process networks with various power contracts. *Computers & Chemical Engineering*, 84:382–393, jan 2016. ISSN 00981354. doi: 10.1016/j.compchemeng.2015.09.019. URL https://linkinghub.elsevier.com/retrieve/pii/S0098135415003142.

[16] S. Zhao, M. P. Ochoa, L. Tang, I. Lotero, A. Gopalakrishnan, and I. E. Grossmann. Novel Formulation for Optimal Schedule with Demand Side Management in Multiproduct Air Separation Processes. *Industrial & Engineering Chemistry Research*, 58(8):3104–3117, feb 2019. ISSN 0888-5885. doi: 10.1021/acs.iecr.8b04964. URL https://pubs.acs.org/doi/10.1021/acs.iecr.8b04964.

[17] D. Zhou, K. Zhou, L. Zhu, J. Zhao, Z. Xu, Z. Shao, and X. Chen. Optimal scheduling of multiple sets of air separation units with frequent load-change operation. *Separation and Purification Technology*, 172:178–191, 2017. ISSN 18733794. doi: 10.1016/j.seppur.2016.08.009. URL http://dx.doi.org/10.1016/j.seppur.2016.08.009.

[18] N. P. Basán, M. E. Cóccola, R. G. Dondo, A. Guarnaschelli, G. A. Schweickardt, and C. A. Méndez. A reactive-iterative optimization algorithm for scheduling of air separation units under uncertainty in electricity prices. *Computers and Chemical Engineering*, 142, 2020. ISSN 00981354. doi: 10.1016/j.compchemeng.2020.107050.

[19] R. Kender, F. Kaufmann, F. Rößler, B. Wunderlich, D. Golubev, I. Thomas, A.-M. Ecker, S. Rehfeldt, and H. Klein. Development of a digital twin for a flexible air separation unit using a pressure-driven simulation approach. *Computers & Chemical Engineering*, 151:107349, 2021. ISSN 00981354. doi: 10.1016/j.compchemeng.2021.107349. URL https://doi.org/10.1016/j.compchemeng.2021.107349.

[20] A. Caspari, C. Offermanns, P. Schäfer, A. Mhamdi, and A. Mitsos. A flexible air separation process: 2. Optimal operation using economic model predictive control. *AIChE Journal*, 65 (11), nov 2019. ISSN 0001-1541. doi: 10.1002/aic.16721. URL https://onlinelibrary.wiley.com/doi/10.1002/aic.16721.

[21] C. Tsay, A. Kumar, J. Flores-Cerrillo, and M. Baldea. Optimal demand response scheduling of an industrial air separation unit using data-driven dynamic models. *Computers and Chemical Engineering*, 126:22–34, 2019. ISSN 00981354. doi: 10.1016/j.compchemeng.2019.03.022. URL `https://doi.org/10.1016/j.compchemeng.2019.03.022`.

[22] M. T. Kelley, R. C. Pattison, R. Baldick, and M. Baldea. An MILP framework for optimizing demand response operation of air separation units. *Applied Energy*, 222:951–966, jul 2018. ISSN 03062619. doi: 10.1016/j.apenergy.2017.12.127. URL `http://linkinghub.elsevier.com/retrieve/pii/S0306261917318524`.

[23] J. Du, J. Park, I. Harjunkoski, and M. Baldea. A time scale-bridging approach for integrating production scheduling and process control. *Computers & Chemical Engineering*, 79:59–69, aug 2015. ISSN 00981354. doi: 10.1016/j.compchemeng.2015.04.026.

[24] M. T. Kelley, R. Baldick, and M. Baldea. An empirical study of moving horizon closed-loop demand response scheduling. *Journal of Process Control*, 92:137–148, aug 2020. ISSN 09591524. doi: 10.1016/j.jprocont.2020.05.011. URL `https://linkinghub.elsevier.com/retrieve/pii/S0959152420302183`.

[25] M. T. Kelley, R. Baldick, and M. Baldea. Demand response scheduling under uncertainty: Chance-constrained framework and application to an air separation unit. *AIChE Journal*, 66(9), sep 2020. ISSN 0001-1541. doi: 10.1002/aic.16273. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/aic.16273`.

[26] L. Sonneborn and F. Van Vleck. The Bang-Bang Principle for Linear Control Systems. *Society for Industrial and Applied Mathematics Series A Control*, 2:151–159, 1965.

[27] S. J. Orfanidis. *Intorduction to Signal Processing*. 2010. ISBN 9780080511986.

[28] MATLAB. MATLAB 2018a, 2018.

[29] CAISO. California Independent System Operator, 2020. URL `http://www.caiso.com/Pages/default.aspx`.

[30] GAMS. General Algebraic Modeling System (GAMS) Release 33.2.0, 2020.

[31] IBM. CPLEX 12.10.0, 2019.