Antoine Perraudat, Stéphane Dauzère-Pérès & Scott Jennings Mason (2022) Stochastic programming approaches for an energy-aware lot-sizing and sequencing problem with incentive, International Journal of Production Research, 60:19, 5746-5768, DOI: 10.1080/00207543.2021.1969048

ARTICLE

Stochastic programming approaches for an energy-aware lot-sizing and sequencing problem with incentive

Antoine Perraudat[1,*] Stéphane Dauzère-Pérès[1,2] Scott Jennings Mason[3]

[1]Department of Manufacturing Sciences and Logistics, CMP,
Ecole des Mines de Saint-Etienne, CNRS UMR 6158 LIMOS,
F-13541 Gardanne, France
E-mail: antoine.perraudat@emse.fr, dauzere-peres@emse.fr

[2]Department of Accounting and Operations Management,
BI Norwegian Business School,
0484 Oslo, Norway

[3]College of Engineering, Computing and Applied Sciences,
Clemson University,
Clemson, SC 29634-0901, USA
E-mail: mason@clemson.edu

ARTICLE HISTORY
Compiled October 6, 2021

Abstract

Motivated by real challenges on energy management faced by industrial firms, we propose a novel way to reduce production costs by including the pricing of electricity in a multi-product lot-sizing problem. In incentive-based programs, when electric utilities face power consumption peaks, they request electricity-consuming firms to curtail their electric load, rewarding the industrial firms with incentives if they comply with the curtailment requests. Otherwise, industrial firms must pay financial penalties for an excessive electricity consumption. A two-stage stochastic formulation is presented to cover the case where a manufacturer wants to satisfy any curtailment request. A chance-constrained formulation is also proposed, and its relevance in practice is discussed. Finally, computational studies are conducted to compare mathematical models and highlight critical parameters and show potential savings when subscribing incentive-based programs. We show that the setup cost ratio, the capacity utilization rate, the number of products and the timing of curtailment requests are critical parameters for manufacturers.

KEYWORDS
Lot Sizing; Optimization; Stochastic programming; Incentive-based programs; Energy management

1. Introduction

1.1. *General introduction*

Production operations can be divided into three levels: Strategic, tactical and operational. At the strategic level, long-term supply chain design decisions (often for

* Corresponding author.

several years) are made by creating or removing facilities, selecting suppliers, and defining the targeted client segments. At the tactical level, decision makers determine production plans over a finite time horizon to fulfill customer demands while respecting objectives and guidelines imposed by the strategic level. Finally, the operational level corresponds to short-term decisions, such as job/lot scheduling. At the tactical and operational levels, energy is one of the critical raw materials necessary for manufacturing finite products, as it is used for supplying power to machines in production facilities. According to the U.S. Energy Information Administration (U.S. Energy Information Administration 2017), the industrial sector is the largest consumer of electricity in the U.S., consuming 30% of the total available electricity.

Up until a few years ago, electricity management was hardly considered as a challenge to address (Biel and Glock 2016). Today, notably due to the 50% increase in electricity prices over the last 10 years, considering electricity in production management has become a primary concern for manufacturers. Consequently, the industrial sector must manage its electricity consumption to remain competitive, in particular process industries that are among the largest electricity consumers. There exist several ways for manufacturing companies to manage consumption: 1) Lower production rates; 2) Turn off machines instead of letting them idle; and/or 3) Buy newer, more efficient machines. Unfortunately, these decisions are often risky since they can lead to a reduced quality of service and may damage machines and/or decrease their lifetime (Dai et al. 2013; Lu et al. 2017). Furthermore, while buying state-of-the-art machines may be a viable solution, the required return on investment is often significant and is not guaranteed to give desirable results if the price of electricity keeps increasing.

Another innovative way to reduce electricity costs is to study the needs of electric utilities and control or reduce electricity consumption during critical hours of the day, known as on-peak power consumption periods. During these on-peak periods, electric utilities are under pressure to provide electricity to a large number of customers while simultaneously preserving their grid integrity and limiting greenhouse gas emissions. To encourage the industrial sector to control or reduce their load during on-peak periods, electric utilities have been developing *demand response* energy programs. Among these demand response programs, two types of energy plans can be distinguished: *incentive-based programs* and *price-based programs*.

In price-based programs, the price of electricity can vary up to each hour of the day. During nights and weekends, which are the off-peak periods, electricity is cheap and affordable; in the middle of a day, an on-peak period, electricity becomes very expensive. Furthermore, the price of electricity is seasonal (i.e., it is more expensive during the summer months than during the winter months). In incentive-based programs, electric utilities send curtailment requests to manufacturing companies asking them to reduce their electric load by a short period of a few hours during the time frame when the utility expects to face an upcoming consumption peak. If manufacturing companies comply with such a request, they receive monetary rewards. As manufacturing companies receive curtailment requests during on-peak periods, compliance forces production to be rescheduled to off-peak periods (i.e., non-critical periods for electric utilities). With such energy plans, the actors within the energy supply chain collaborate to reduce costs and the impact of the change on production operations, whether at the utility level with the production of energy or at the industrial level manufacturing operations. Depending on their locations, manufacturing companies can choose among different demand response programs proposed by electric utilities. In the U.S., the Office of Energy Efficiency and Renewable Energy references demand response programs by state (Department of Energy 2018). For instance, a Californian manufacturing company that decides to opt for demand response programs proposed by Pacific Gas and Electric Company (2018) has the option between four different tariffs.

1.2. *Literature review*

In the literature, most researchers have studied energy implications at the operational level through production scheduling problems (Biel and Glock 2016; Gahm et al. 2016; Giret, Trentesaux, and Prabhu 2015; Merkert et al. 2015; Stahl, Taisch, and Kiritsis 2016) while decisions at strategic and tactical levels, which are also crucial in production management, have been rarely addressed.

Most publications, which include demand response programs in production planning (tactical level) problems, consider price-based programs. Denton et al. (1987) investigate a production planning problem where industrial firms face decisions at multiple levels. Each day, the manufacturer must define how many shifts must occur. On a weekly basis, it must determine whether it must start activities during weekends. On an annual basis, it must find the best periods (months) to manufacture products. Denton et al. (1987) consider a Time-Of-Use tariff (TOU), in which price of electricity is much more expensive during on-peak periods, as a demand response program with other energy sources. They show that the structure of the TOU tariff has a great influence on total costs when an industrial firm has the possibility to be energy-efficient by shifting production to off-peak periods. They also underline the trade-off between energy and labor costs, as labor cost is much more expensive during off-peak periods. This is a result shared by Fethke and Tishler (1990). They conclude that industrial companies whose energy costs are low and labor costs high have little interest to consider a price-based program such as TOU. Salahi and Jafari (2016) analyze a production planning problem with a one-day planning horizon where demands are random. They study energy performances with two demand response programs, a TOU tariff and a Real-Time Program (RTP) where the price of electricity varies at each hour of the day. They are notably interested in knowing how the features of a machine, such as its energy consumption and processing speed, can impact the production plan with such tariffs. Masmoudi et al. (2017) propose several energy-aware heuristics for a lot-sizing problem where the production system is organized as a flow line with N machines and M buffers. Numerous contributions that consider demand response programs in production planning study air separation plants that are energy-intensive. An air separation plant is a production facility that takes valuable natural gas from air by a separation mechanism and redistributes the gas to its customers. Karwan and Keblis (2007), Castro, Harjunkoski, and Grossmann (2011), and Mitra et al. (2012) analyze the behavior of air separation plants under price-based programs using features proper to these industries, notably the fact that switching between operating nodes causes a modification of electricity consumption and production rates between each product. Doing so, manufacturers can reduce the energy impact of production operations and costs but must pay ramp-up losses each time a new operating node is selected. Wichmann, Johannes, and Spengler (2019) model a multi-item lot-sizing and scheduling problem where the energy price is time-dependent. In their problem, energy in batteries can be saved, e.g. when energy costs are small, for later use to reduce production costs when energy costs are larger. They show that production costs for manufacturing industries can be significantly reduced when energy prices are highly volatile, machine utilization rates high and products heteregenous. Tang, Che, and Liu (2012) study a production planning problem in a steel-rolling mill production facility. The objective consists in minimizing production costs and energy consumption. The demand is assumed to be stochastic and energy consumption is assumed to be nonlinear to better model the real energy consumption of heat furnaces. A tailor-made heuristic is shown to be promising to solve the studied problem. Golpîra, Khan, and Zhang (2018) study an integrated multi-item multi-machine lot-sizing and jop-shop scheduling problem. Energy generation and distribution is included in the problem. More precisely, the production facility generate electricity, and it must be decided to either store it, use it for production or sell it to the main grid. They show that including energy generation and distribution in the problem lead to promising savings.

There are other contributions that consider fixed cost for energy with constraints on the total energy consumption by period. Giglio, Paolucci, and Roshani (2017) propose a relax-and-fix heuristic to solve an integrated lot-sizing and job-shop scheduling problem with re-manufacturing when processing times of machines are controllable. Energy consumption is considered with a fixed cost. Processing times can be compressed to save energy costs, and thus to reduce the overall production costs. Hajej and Rezg (2019) consider a single-item lot-sizing problem on multiple machines where energy costs are fixed and the demand is stochastic. In their problem, one must decide when to produce and when to perform maintenance operations to meet the demand. However, financial penalties must be paid if too much energy is consumed. Rapine et al. 2018 and Rapine, Goisque, and Akbalik (2018) consider a single-item lot-sizing problem on parallel machines where one must decide when to produce to meet time-varying demands. New machines can be started up if the production capacity is insufficient to meet the demand. However, producing and starting-up machines consume energy and a global budget on energy consumption must be respected. The global budget on energy consumption is a hard constraint in their paper. The complexity of the problem is studied and exact algorithms are also proposed to solve the problem.

To our knowledge, Latifoğlu, Belotti, and Snyder (2013) and Li and Hong (2017) are the only authors who analyze production planning problems with incentive-based programs. Latifoğlu, Belotti, and Snyder (2013) also study an air separation plant providing gas to hospitals. As hospitals have a critical need of gas, Latifoğlu, Belotti, and Snyder (2013) propose a robust framework for being able to have enough gas in inventory at any time to satisfy demands when a curtailment request occurs. In their problem, if the manufacturing company can comply with curtailment requests, it takes advantage of a discount fixed-rate energy plan. Li and Hong (2017) include an incentive-based program in a lithium-ion battery manufacturing system. To comply with curtailment requests, they opt for a reactive approach by turning machines off as soon as a notice is received. If they manage to comply, they receive a monetary reward for their reduced electricity consumption, calculated using historical electric load data. If they do not comply, there is no financial penalty.

1.3. *Contributions and organization of the paper*

The number of publications that address energy-aware industrial production planning problems considering incentive-based programs and/or lot-sizing is very limited. Moreover, the literature on stochastic lot-sizing problems is limited compared to the literature on deterministic lot-sizing problems. Also, most stochastic lot-sizing papers consider that only the demand is stochastic (Brahimi et al. 2017). However, on a short-term horizon, considering uncertain demand is not very realistic since the demand forecast is usually reliable, in particular in process industries where the demand in the next few days/weeks is often well known. There is still a lot to do when the demand is deterministic but other parameters are stochastic. Such problems are interesting to study for short-term production planning, where other critical parameters than the demand might be stochastic such as the timing of the curtailment requests. In addition, contrary to existing works on lot-sizing problems with energy constraints, we will consider incentive-based programs, which are not studied in the lot-sizing literature and little studied overall.

In this paper, we propose to pursue the work of Latifoğlu, Belotti, and Snyder (2013) and Li and Hong (2017) on incentive-based programs. We focus on an integrated stochastic lot-sizing problem with real constraints imposed by incentive-based programs. More precisely, we study how a manufacturing company can take advantage of incentive-based programs to minimize costs. Our work differs from Li and Hong (2017) because we propose a predictive approach to answer curtailment requests, where not answering a request is penalized. Because the demand is deterministic, the stochastic dimension of the problem comes from the incentive-based program. Contrary to Latifoğlu, Belotti, and Snyder (2013), monetary incentives are distributed only if a curtailment request is received, and the production facility can keep producing under a curtailment request (see Section 2.1). We also propose a chance-constrained formulation where monetary penalties are incurred when a curtailment request is not respected.

This paper is organized as follows. In Section 2, the problem is introduced and formalized. In Section 3, the stochastic modeling of the problem is introduced: (1) The modeling of scenarios for the stochastic models is presented, (2) A two-stage formulation to determine a production plan that covers all possible curtailment requests, and (3) A chance-constrained formulation where not answering some curtailment requests is allowed. In Section 4, a computational study on the two-stage formulation, inspired by the glass industry is conducted, where critical parameters and potential savings considering an incentive-based program are analyzed. In Section 5, a computational study is proposed to compare the two-stage and chance-constrained formulations. In Section 6, managerial insights are provided. Finally, in Section 7, we conclude and propose perspectives.

2. Problem modeling

2.1. *Problem description and notations*

Lot-sizing considerations. We consider a manufacturing system producing P products on a single critical machine or production line. We consider a short-term planning horizon (at most a week) with a *deterministic* demand. The planning horizon is divided into a finite number of time periods, T, and back orders are allowed. Inventory holding costs must be paid if products are in stock at the end of a time period. The setup state for a product of the machine or the production line between two consecutive time periods is kept; this feature is also known as setup carryover. If production for another product must be started, then a setup is performed.

Energy considerations. Energy considerations are explored by considering that the machine has two states (i.e., idle and busy) that consume energy, but in different quantities, and by considering an incentive-based program, which is subscribed annually or for multiple years (strategic decision).

Firstly, an electricity-consuming manufacturer ("the manufacturer") pays a fixed rate per kW-hr based on electricity usage (kWh) measured by the electric utility at regular time intervals (i.e., 15-minute or 30-minute time intervals). Secondly, the manufacturer can receive curtailment requests, which occur in a very precise time interval known as the performance period. The uncertainty in the problem comes from the fact that, as it is the case in practice, the occurrence of a curtailment request is known very late (i.e., between 15 minutes and an hour before), and a curtailment request can arrive at *any time* period during the performance period. However, the duration of the performance period is bounded (e.g., between 2:00 PM and 6:00 PM), the number of requests is limited to at most one per day and the duration of a curtailment request is a predetermined constant (e.g., four hours). The two last assumptions are consistent with most incentive-based programs, which are designed so that the manufacturer does not have to endure more than one curtailment request per day (the total number of curtailment requests by year is also limited to a fixed number) that must be answered and whose duration is fixed.

When the manufacturer receives a curtailment request, it must reduce and maintain its electric load below a predetermined level, the *Firm Service Level* (FSL). Production can then continue if the electric load is below the firm service level. The firm service level is defined as the difference between the historical electric load level and the *Load Reduction Threshold* (THR). The historical electric load level is the electric load consumed by the manufacturer and measured by the electric utility before subscribing to an incentive-based program. The manufacturer has the possibility to select the most appropriate load reduction threshold for its production facility, but it must respect the minimum load reduction threshold imposed by the demand response program.

For each time interval under a curtailment request, if the electric load exceeds the specified firm service level, the manufacturer must pay financial penalties. The typical length of a time interval in the performance period is 15 or 30 minutes. Otherwise, the manufacturer gets monetary rewards, which are proportional to the electric load reduction (\$/kW). The electric load reduction is equal to the difference between the historical electric load level and the current electric load. The rewarded electric load reduction is an average over the curtailment request. The rewards are also limited to a Demand Saving Goal, denoted by DSG, by curtailment request. DSG is always greater or equal to THR. If the manufacturer receives γ curtailment requests in the horizon, it is rewarded γ times, and rewards are therefore limited to $\gamma \times DSG$. In practice, electric utilities may not always reward load reduction beyond THR (Pacific Gas and Electric Company 2015; Southern California Edison 2014).

Finally, we assume that there is no energy storage.

Figure 1 illustrates the energy considerations for a demand saving goal of 25 kW, a load reduction threshold of 10 kW, and a historical electric load of 30 kW. The performance period starts 12:00 PM and ends at 2:30 PM. A curtailment request is received at 12:30 PM and lasts to 2:00 PM. The manufacturer must therefore reduce its electric load so that it does not exceed 30 - 10 = 20 kW.

Objective. The objective is to determine a feasible production plan that respects the capacity constraints, considers the curtailment requests, and minimizes the total cost minus the monetary rewards. The three following assumptions are considered.

- Curtailment requests occur at the beginning of a time period during the performance period.
- Curtailment requests last an integer number of time periods.

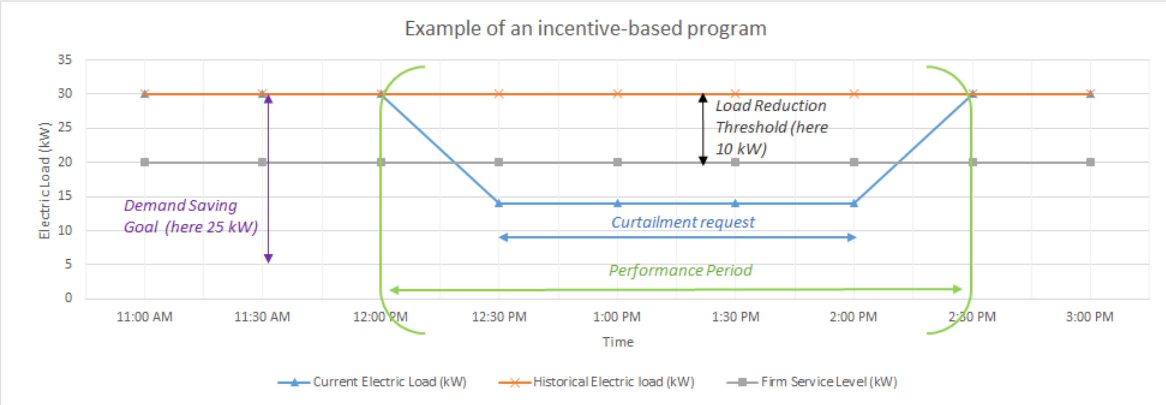

Figure 1. Example of the studied incentive-based program.

- Idle times do not break setup states.

Indices and sets:

i: Index for products, $i \in \{1, ..., P\}$,
t: Index for periods, $t \in \{1, ..., T\}$.

Parameters:

d_{it}: Demand for product i at period t,
C_t: Capacity available (in time units) at period t,
a_i: Production time (in time units) per unit of product i,
b_i: Setup time (in time units) of product i,
Ep_i: Energy consumed (kWh) per unit of product i,
Es_i: Energy consumed (kWh) for setting up product i,
Cp_{it}: Production cost (\$/unit) per unit of product i at period t,
Cs_{it}: Setup cost (\$) for product i at period t,
h^+_{it}: Holding cost (\$/unit) per unit of product i at the end of period t,
h^-_{it}: Backlogging cost (\$/unit) per unit of product i at the end of period t,
P^{idl}: Electric load (kW) when the machine is idle,
f: Fixed rate (\$/(kWh)) to pay for consuming electricity,
f^{rwd}: Rewards (\$/kW) received for reducing electric load during curtailment requests,
THR: Load reduction threshold (kW),
S_t: Is equal to 1 if a curtailment requests starts at period t, and 0 otherwise,
DSG: Demand saving goal (kW) by curtailment request,
H_t: Historical electric load (kW),
I^{end}: Maximum units of demand that can be backlogged at the end of the planning horizon,
$I^+_{i,0}$: Initial inventory of for product i at the beginning of the planning horizon,
$I^-_{i,0}$: Initial backlog of product i at the beginning of the planning horizon.

Decision variables:

X_{it}: Quantity of product i (lot size) to be manufactured at period t,
$Y_{it} \in \{1, ..., P\}$: Is equal to 1 if there is a setup for product i at period t, and 0 otherwise,
I^+_{it}: Inventory level of product i at the end of period t,
I^-_{it}: Backlog of product i at the end of period t,
IDL_t: Time duration (in time units) the machine must be idle in period t,

$W_{it} \in \{1, ..., P\}$: Is equal to 1 if the setup state of product i is carried over from period $t-1$ to period t, and 0 otherwise,
E_t: Energy consumption (kWh) in period t,
RLR: Rewarded Load Reduction (kW).

Additionally, the set $\Omega_{t'}$ is introduced, which describes the set of consecutive periods where a curtailment request is active after starting in period t', i.e. $\Omega_{t'} = \{t', t'+1, ..., t'+N_{t'}-1\}$.

2.2. *Mathematical formulation*

Minimize

$$\sum_{i,t}(Cp_{it}X_{it}+Cs_{it}Y_{it})+\sum_{i,t}(h^{+}_{it}I^{+}_{it}+h^{-}_{it}I^{-}_{it})$$
$$+f\sum_{t}E_t-f^{rwd}RLR \tag{1}$$

Subject to

$$I^{+}_{i,t-1}-I^{-}_{i,t-1}+X_{it}=d_{it}+I^{+}_{it}-I^{-}_{it} \qquad \forall i,\forall t>2 \tag{2}$$
$$I^{+}_{i,0}-I^{-}_{i,0}+X_{i,1}=d_{i,1}+I^{+}_{i,1}-I^{-}_{i,1} \qquad \forall i \tag{3}$$
$$I^{-}_{iT}\le I^{end} \qquad \forall i \tag{4}$$
$$X_{it}\le\sum_{l=1}^{T}d_{il}(W_{it}+Y_{it}) \qquad \forall i,\forall t \tag{5}$$
$$\sum_{i}W_{it}\le 1 \qquad \forall t\ge 2 \tag{6}$$
$$W_{it}\le Y_{i,t-1}+W_{i,t-1} \qquad \forall i,\forall t\ge 2 \tag{7}$$
$$W_{it}+W_{i,t-1}+Y_{j,t-1}\le 2+Y_{i,t-1} \qquad \forall i,\forall j\ne i,\forall t\ge 2 \tag{8}$$
$$W_{i1}=0 \qquad \forall i \tag{9}$$
$$C_t=\sum_{i}(a_iX_{it}+b_iY_{it})+IDL_t \qquad \forall t \tag{10}$$
$$E_t=\sum_{i}(Ep_iX_{it}+Es_iY_{it})+P^{idl}IDL_t \qquad \forall t \tag{11}$$
$$THR\le H_{t'}-\frac{E_{t'}}{C_{t'}} \qquad \forall t \text{ s.t. } S_t=1,\forall t'\in\Omega_t \tag{12}$$
$$RLR\le DSG\sum_{t}S_t \tag{13}$$
$$RLR\le\sum_{t,N_t\ne 0}\frac{1}{N_t}\sum_{t'=t}^{t+N_t-1}(H_{t'}-\frac{E_{t'}}{C_{t'}}) \tag{14}$$
$$RLR\ge 0 \tag{15}$$
$$X_{it}\ge 0 \qquad \forall i,\forall t \tag{16}$$
$$I^{+}_{it}\ge 0, I^{-}_{it}\ge 0 \qquad \forall i,\forall t \tag{17}$$
$$IDL_t\ge 0 \qquad \forall t \tag{18}$$
$$Y_{it},W_{it}\in\{0,1\} \qquad \forall i,\forall t \tag{19}$$

The objective function (1) includes manufacturing, setup, holding, backlogging and energy costs, and monetary rewards. Constraints (2)-(10) are the traditional lot-sizing constraints, while Constraints (11)-(12) are constraints related to the incentive-based program. Constraint (2) and (3) ensure the balance flow. Constraint (4) defines a limit on the backlogged demands at the end of the planning horizon. Constraint (5) ensures that a product i can only be manufactured in period t if a setup operation is conducted for i at period t, or if the setup state for i is carried over from the preceding period.

Constraint (6) limits to one the number of setup states that can be carried over from period $t-1$ to period t. Constraints (7)-(8) model the multi-period setup carryovers. In particular, Constraint (8) ensures that, if a setup state is carried over from period $t-2$ to period t, and if there is a setup operation for another product at period $t-1$, then the machine needs to be setup again for the first product at period $t-1$. Constraint (9) corresponds to the initial conditions for setup states. Constraint (10) corresponds to the production capacity constraints. It is also used to compute how long the machine is idle in a period. Constraint (11) calculates the electricity consumption for each time period. Constraint (12) is the load reduction threshold constraint that forces the electric load in period t, $\frac{E_t}{C_t}$, to be reduced by a given threshold THR for each t under curtailment request. Constraints (13) and (14) calculate the rewarded load reduction. Constraint (13) limits the rewarded load reduction to the demand saving goal DSG multiplied by the number of curtailment requests received in the planning horizon as $S_t = 1$ if a curtailment request is received at period t. Constraint (14) compute the load reduction as an average of the curtailment amount during the periods where a curtailment request is active. It also considers the number of curtailment requests received in the horizon. Constraint (15) ensures that the model is bounded. Finally, Constraints (16)-(19) are the non-negativity and integrality constraints.

Let us discuss below some important characteristics of our problem:

- Production costs consists of labor costs and raw material purchasing costs.
- If idling time do not consume electricity, it can be simpler to combine production costs with electricity consumption costs.
- We assume that the processing of two products require different electricity consumption on the same machine (or production line). This is possible, for instance, when products can require different ramp-up and ramp-down temperatures and processing times.

2.3. *Illustrative example*

Let us provide an example to illustrate the consequences on production plans of an incentive-based program, especially how curtailment requests are handled. In this example, production is planned for 24 hours for three products. During a curtailment request, the FSL is set up to 50% of the historical electric load and the average capacity utilization rate to 75%.

There are two options for respecting the load reduction threshold and answering curtailment requests:

(1) A front-loaded production plan where lots of each product are stored before the beginning of the performance period, thus incurring holding costs,
(2) A back-loaded production plan where some of the demands are backlogged during the performance period, thus incurring backlogging costs.

Through this illustrative example, note that the first option is widely utilized while the second one is not exploited. The first option is favored because the capacity utilization rate is low enough to enable a large inventory for each product. Moreover, doing this is rather inexpensive since average hourly inventory costs are often negligible in industrial contexts. Figure 2 shows that the model builds a large inventory of product 3 from period 6 to period 15 in order to satisfy the demands between periods 16 and 24. This has a two-fold impact: The electric load is reduced during the curtailment

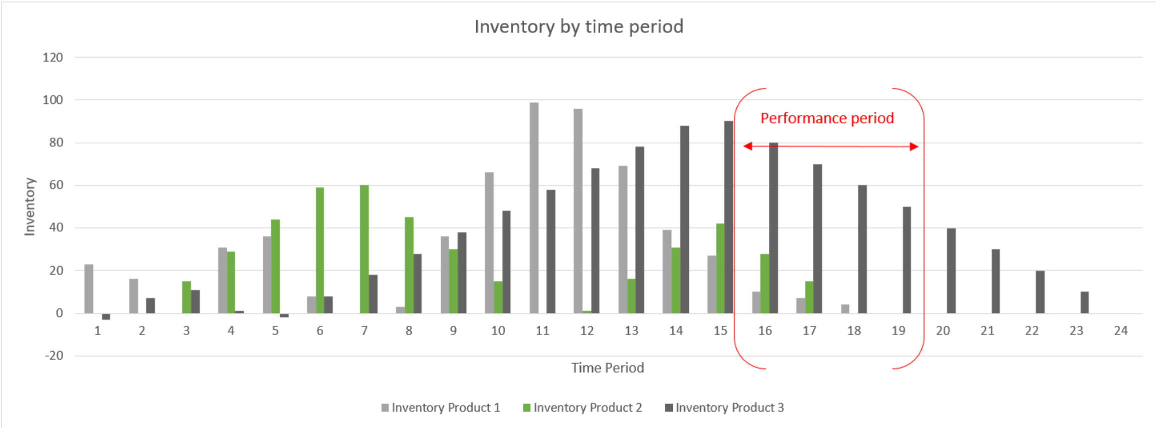

Figure 2. Illustrative example - Inventory profile, with request.

request and setup operations, which are very expensive, are saved. A large inventory of product 3 is built because this is the product for which the machine consumes the most electricity to manufacture lots. Similar to product 3, for other products, lots in inventory are used to satisfy most of the demands during the performance period and to reduce the electric load. Finally, note that the inventory profile illustrates the traditional trade-off in lot-sizing problems between setup costs, inventory costs, and manufacturing costs. The second option, which would consist in backlogging some of the demands during the performance period, has not been exploited. This is because, in contrast to average hourly holding costs, backlogging is usually very costly.

Figure 3 compares the electric load when there is a curtailment request ("Request"), and when there is no curtailment request ("No request"). With a curtailment request, the electric load over the performance period is constant, the load reduction threshold is satisfied in a very tight way, which is caused by the fact that $THR = DSG$ and there is no point in saving more than DSG kW since the electric utility will not reward additional savings. Respecting load reduction constraints tightly is risky and could lead to a tense situation in practice, since it is always possible to deviate from the initial production plan. Hence, a decision maker might want an additional margin. For example, he/she could use a larger load reduction threshold but states a lower one to the electric utility to make sure the production system will be able to answer a curtailment request.

3. Stochastic modeling

In practice, it is often difficult, if not impossible, to know exactly when curtailment requests are active. It is only known where they can occur (during the performance period), typically between 11:00 AM and 7:00 PM for week days (Pacific Gas and Electric Company 2015), weekends being request-free days. Moreover, it is known that the number of curtailment requests and their cumulative duration are limited over a year. Finally, there can be at most a single curtailment request per day. However, it is more difficult to evaluate in practice when electric utilities send curtailment requests and how long they last.

To address this, a simple way to generate scenarios, which describe where curtailment requests occur and how long they last, and compute their associated probabilities

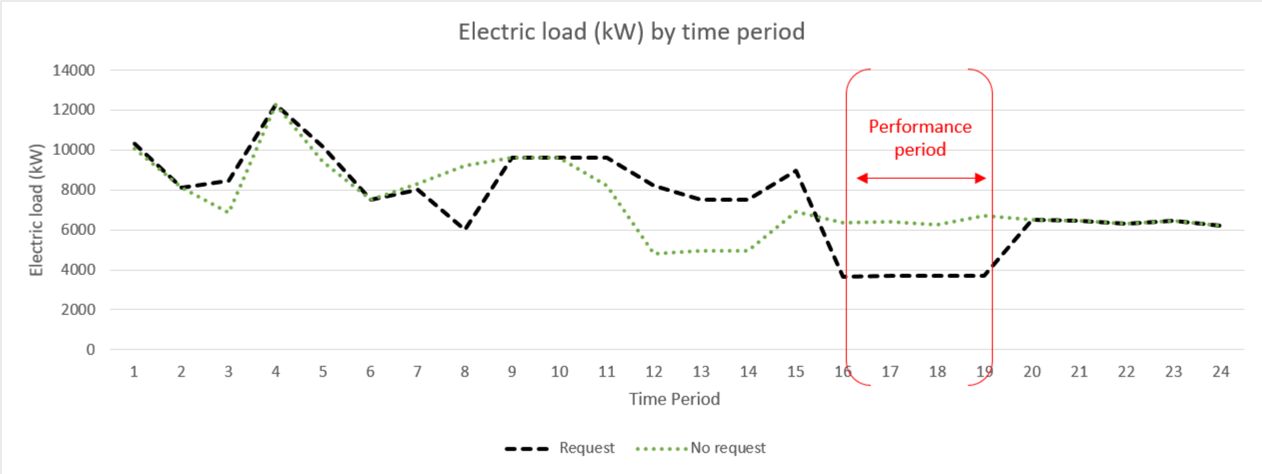

Figure 3. Illustrative example - Electric load profiles.

is proposed in Section 3.1. Two stochastic optimization models are then proposed in Section 3.2 to cover both the case where all the curtailment requests must be answered and the case where not answering some curtailment requests is tolerated.

In the remainder of the paper, let us use the subscript $s \in \{1, ..., S\}$ for scenarios with an associated probability p_s.

3.1. *Scenario management*

3.1.1. Scenario generation

Even though it is difficult to predict when curtailment requests occur, it is possible to generate scenarios with two reasonable assumptions.

The first one is on the duration of the requests. When electric utilities state in their contract that they can send curtailment request that can last up to X hours, they will send requests that last exactly X hours when they are not limited by the end of the performance period of the incentive-based program. This assumption is reasonable since electric utilities are limited by the demand response programs to a limited number of curtailment requests by year, and because electric utilities do not reward the electric consumption reduction ($/kWh), but the average electric load reduction ($/kW). For instance, consider that the performance period starts at 12:00 PM and ends at 6:00 PM, and that a curtailment request can last up to 6 hours. If the electric utility sends a curtailment request at 12:00 PM, then it is assumed the curtailment request will last exactly 6 hours. In this case, there is no reason that curtailment requests should be smaller than 6 hours, as it should help electric utilities reduce their operating costs and limit greenhouse gas emissions. Similarly, if a curtailment request is sent at 12:30 PM, the curtailment request is assumed to last 5 hours and 30 minutes. In other words, all curtailment requests that do *not* last exactly X hours or do not end at the end of the performance period are not considered.

The second assumption is on the time at which the requests occur. The electric utility measures energy usage at regular time intervals, e.g., every 15 minutes (Pacific Gas and Electric Company 2015) or 30 minutes (Southern California Edison 2014), and not on a continuous time basis. Then, it can be assumed that at the beginning of each time period during the performance period, which corresponds to the beginning of a time interval, a curtailment request starts being active. This assumption is very close to those we formulated in Section 2 for the deterministic model. In addition, this assumption implies that the curtailment requests last a multiple integer of time intervals within the performance period.

Based on these two assumptions, generating scenarios is performed as follows: At the beginning of each time interval in the performance period, a scenario is generated and valid if it either last the maximum duration of curtailment requests or ends at the very last period of the performance period. We give an example of such a scenario generation in Figure 4 for a five-hour performance period between 11:00 AM and 4:00 PM when electricity usage is measured every 30 minutes and requests last 4 hours. In total, there are 11 scenarios, 10 of them contain a curtailment request.

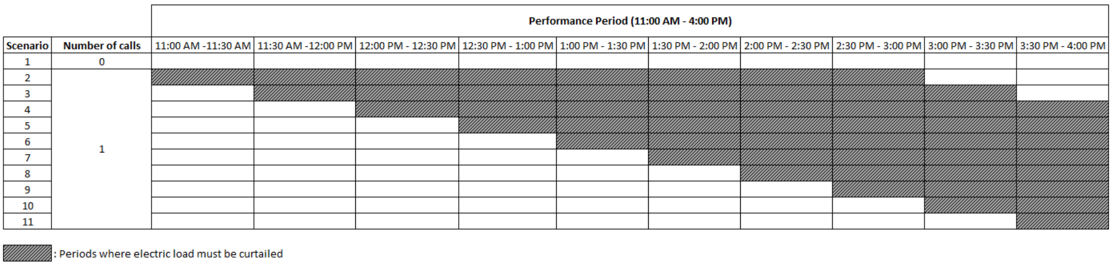

Figure 4. Scenario generation - Example.

We also show the scenario generation for the incentive-based programs that can be found in Pacific Gas and Electric Company (2015) and Southern California Edison (2014) for an interruptible day because they will be studied in the computational study. Pacific Gas and Electric Company 2015 contains 35 scenarios (see Appendix B), and Southern California Edison (2014) 15 scenarios (see Appendix A). For a planning horizon of a week with five interruptible days, Pacific Gas and Electric Company (2015) contains $35^5 = 52,521,875$ scenarios, and Southern California Edison (2014) contains $13^5 = 371,293$ scenarios.

As curtailment requests are assumed to last a multiple integer of time intervals and are assumed to start at the beginning of a time interval, curtailment requests follow a discrete distribution probability as the number of outcomes is finite. In Section 3.1.2, we elaborate on the probabilities associated to scenarios.

3.1.2. Discrete distribution probabilities

In an incentive-based program, it is often known:

(1) How many interruptible days there are in a year d_{int},
(2) How many curtailment requests an electric utility sends in a year N_{year},
(3) How many curtailment requests an electric utility can send in a day N_{day}.

When $N_{day} = 1$ (Pacific Gas and Electric Company 2015; Southern California Edison 2014), we propose to use this information to compute a simple way to estimate the probability of receiving a curtailment request in an interruptible day. Let us denote P this probability and p_1 the probability of not receiving any request in an interruptible day:

$$p_1 = \frac{d_{int} - N_{year}}{d_{int}} \tag{20}$$

$$P = \sum_{s \geq 2} p_s = 1 - p_1 \tag{21}$$

Numerical examples are provided in Table 1. Example 1 corresponds to the contracts proposed by Pacific Gas and Electric Company (2015) and Southern California Edison (2014), and examples 2, 3, and 4 are provided for comparison.

Table 1. Numerical example: Estimating probability distributions

Example	d_{int}	N_{year}	N_{day}	p_1	P
1	260	30	1	0.885	0.115
2	260	50	1	0.808	0.192
3	200	30	1	0.850	0.150
4	200	50	1	0.750	0.250

Recall that the subscript s is used for scenarios. P is the probability of receiving a curtailment request in an interruptible day, but it is not the probability of an actual curtailment request p_s for $s \geq 2$. Under the assumption that all curtailment requests are equiprobable, $p_s = \frac{P}{N}$ for $s \geq 2$, where N is the number of possible curtailment requests in the performance period.

3.2. *Stochastic optimization models*

The deterministic formulation shows how curtailment requests and load reduction constraints can be modeled. In real-life environments, this formulation is limited because it does not consider the underlying uncertainty related to curtailment requests, especially the fact that the time at which a request occurs is unknown. The occurrence of a curtailment request is known very late (i.e., between 15 minutes and an hour before). Hence, we propose two stochastic formulations in this section. The first one assumes that every load reduction threshold constraint must be respected, i.e. the manufacturer wants to answer any curtailment request. This is motivated by the fact that the electric utility might break the contract if the manufacturer does not respect the load reduction constraint even for a single time period. The second formulation assumes that penalties are acceptable for a limited number of time periods when the load reduction threshold constraints are violated. In both formulations, uncertainty is represented through scenarios with discrete distribution probabilities (see Section 3.1 to see how scenarios are managed).

3.2.1. Two-stage formulation: Penalties are forbidden

The objective of the first stochastic model is to propose a formulation that determines production plans that can tackle uncertainty on curtailment requests by answering any curtailment request. Hence, there is no risk of a possible curtailment request violation. Such an optimization problem can be modeled by using a two-stage or multi-stage modeling. In a two-stage formulation, decision variables are partitioned into two sets. The first set contains the first-stage decision variables, also known as *here-and-now* decisions, which must be taken before any realization of uncertainty for the *whole* planning horizon. The second set contains the second-stage decision variables, which are also known as the *wait-and-see* decisions, and represents an available recourse when uncertainty unfolds. In a multi-stage formulation, the uncertainty is revealed over time. Production decisions are taken at different stages based on the knowledge of uncertainty at the stage (Shapiro, Dentcheva, and Ruszczynski 2009). In our model, the first-stage decisions are the lot-sizing and sequencing decisions, and the second-stage decisions are decisions related to the incentive-based programs.

In this paper, we propose a two-stage model to evaluate incentive-based programs. A two-stage model does not fit all industries, but it is realistic for manufacturing systems with little operational flexibility such as process industries. This is why the glass industry is considered in the computational study in Section 4.

A two-stage approach is also relevant because violating a curtailment request may have serious consequences on the relationship with the electric utility and break the program. Therefore, a two-stage model is more reliable as it does not rely on operational response time as much as a multi-stage model. It is always possible to force the shutdown of a piece of running equipment but at the expense of its lifetime and by increasing the probability of scraps and reworks. For process industries such as the steel, glass, or semiconductor industries where the capital investment can be significant, rescheduling short-term production is not desirable and is often not acceptable. In this case, managers may be reluctant to assess potential benefits of incentive-based programs with a multi-stage model. With a two-stage model, there is still a risk on the first possible curtailment request of the performance period due to production variability. Nevertheless, managers can decide to add an additional margin, for example fifteen minutes, to answer the curtailment request. Additional margins cannot be defined in a multi-stage formulation as the uncertainty is not yet realized. For this problem, the two-stage formulation produces robust solutions because it ensures that any curtailment request is satisfied.

We present the two-stage mathematical formulation below. Constraints involving the first-stage decision variables remain unchanged compared to the deterministic formulation. To model the second-stage, we introduce new decision variables RLR_s. These decision variables calculate the Rewarded Load Reduction for each scenario s. We assume that the first scenario, i.e. when $s = 1$, is the scenario where there is no curtailment request. Similarly to the set $\Omega_{t'}$, the set $\Omega_{s,t'}$ is introduced to model the set of consecutive periods where a curtailment request is active after starting in period t' in scenario s, i.e. $\Omega_{s,t'} = \{t', t'+1, ..., t' + N_{st'} - 1\}$. Observe that $|\Omega_{s,t'}| = N_{s,t'}$.

It is interesting to note that many constraints in Constraints (12) of the deterministic model would be redundant in a two-stage formulation if there were indexed by s. This is because a period can be covered by more than one curtailment request (scenario). In other words, if the load reduction threshold is (not) respected in period t for a given scenario, it is actually (not) respected in t for all scenarios, and Constraints (12) do not need to be indexed by s in a two-stage formulation. In addition, there is no possible recourse action as production must be planned before the realization of the curtailment requests, and the electricity consumption E_t is identical to all scenarios. RLR_s is only used to compute the outcome for each scenario. The two-stage formulation can therefore be formulated as follows:

$$
\begin{aligned}
&\text{Minimize} \\
&\sum_{i,t}(Cp_{it}X_{it} + Cs_{it}Y_{it}) + \sum_{i,t}(h^{+}_{it}I^{+}_{it} + h^{-}_{it}I^{-}_{it}) \\
&+ f\sum_{t}(E_t) - f^{rwd}\sum_{s\geq 2}(p_s RLR_s) && (22)\\
&\text{Subject to} \\
&(2)-(11),(16)-(19) \\
&THR \leq H_t - \frac{E_t}{C_t} && \forall t \in \Omega' \quad (23)\\
&RLR_s \leq DSG\sum_{t} S_{st} && \forall s \quad (24)\\
&RLR_s \leq \sum_{t,N_{st}\neq 0}\frac{1}{N_{st}}\sum_{t'=t}^{t+N_{st}-1}(H_{t'} - \frac{E_{t'}}{C_{t'}}) && \forall s \quad (25)\\
&RLR_s \geq 0 && \forall s \quad (26)
\end{aligned}
$$

The objective function (22) is the expected value of total costs when all scenarios representing the uncertainty on curtailment requests are considered. Constraints (23) ensure that curtailment requests are satisfied. Similarly to the deterministic model, Constraints (24) and (25) calculate the rewarded load reduction for each scenario. Constraints (24) limit the rewarded load reduction to the demand saving goal DSG multiplied by the number of curtailment requests received in the planning horizon and scenario s. Constraints (25) compute for scenario s the load reduction as an average of the curtailment amount during the periods where a curtailment request is active for scenario. It also considers the number of curtailment requests received. Constraint (26) is the non-negativity constraint for decision variables RLR_s.

In the case where $THR = DSG$, as all curtailment requests must be respected, we necessarily have $DSG\sum_t S_{st} = RLR_s \;\; \forall s$, and Constraints (24) and (25) are no longer relevant. The optimization model can be simplified as follows:

$$
\begin{aligned}
&\text{Minimize} \\
&f_1 = \sum_{i,t}(Cp_{it}X_{it} + Cs_{it}Y_{it}) + \sum_{i,t}(h^{+}_{it}I^{+}_{it} + h^{-}_{it}I^{-}_{it}) \\
&+ f\sum_{t}(E_t) - f^{rwd}\sum_{s\geq 2,t}(p_s S_{st} DSG) && (27)\\
&\text{Subject to} \\
&(2)-(11),(16)-(19),(23)
\end{aligned}
$$

The objective function (27) is the expected value of total costs when all scenarios representing the uncertainty on curtailment requests are considered. However, $f^{rwd}\sum_{s\geq 2,t}(p_s S_{st} DSG)$ is now a constant term, and can be computed before (or after) solving the optimization problem.

3.2.2. Chance-constrained formulation: Penalties are authorized

Mathematical formulation. In the chance-constrained model, not answering curtailment requests is now tolerated with a risk level $\alpha \in [0, 1]$. In our problem, the only constraint candidate that can be probabilistic is Constraint (12). The other constraints are not subject to uncertainty or are used to calculate an outcome. Constraint (12) is modeled as a joint probabilistic constraint as follows:

$$Pr[THR \leq H_{t'} - \frac{E_{t'}}{C_{t'}} \qquad \forall s, \forall t \text{ s.t. } S_{st} = 1, \forall t' \in \Omega_{s,t}] \geq 1 - \alpha \tag{28}$$

If a curtailment request is violated, the manufacturer pays financial penalties proportional to the excessive electric usage. We propose the following mathematical formulation which models the financial penalties and the probabilistic constraint based on the fact that scenarios follow a discrete probability distribution.

Similar to the two-stage model, there is no need to introduce many additional variables to model the joint probabilistic constraint. This is because, if the load reduction threshold is not respected in a period for a given scenario, it is actually not respected for all scenarios that contains this period. Consequently, only new variables indexed by t, and not by s are necessary to model the joint probabilistic constraint.

To compute the joint probabilistic constraint and better evaluate the expected production costs, penalties and rewards, the periods where the electric load is correctly curtailed have to be distinguished from the periods where the electric load is not correctly curtailed. Binary and continuous indicator variables are introduced for this purpose. Let us introduce the following new parameters and decision variables.

New parameters:
f^{pen}: Penalty rate (\$/kWh) for excessive electric usage,
$M_t^+ = M_t^- = \frac{\sum_{t',i}(Ep_i d_{it'} + Es_i) + P^{idl} C_t}{C_t}$: Upper bounds on electric load at period t.

New decision variables:
RLR_t: Rewarded Load Reduction (kW), which is positive if the curtailment request at period t is respected, and equal to 0 otherwise,
E_t^{excess}: Excessive electric usage (kWh) at period t,
Δ_t^-: Is positive (unit in kW) if the load reduction threshold at period t is violated, and equal to 0 otherwise,
Δ_t^+: Is positive (unit in kW) if the load reduction threshold at period t is respected, and equal to 0 otherwise,
z_t: Is equal to 1 if curtailment request at period t is respected, and equal to 0 otherwise.

Minimize

$$f_2 = \sum_{i,t}(Cp_{it}X_{it} + Cs_{it}Y_{it}) + \sum_{i,t}(h^+_{it}I^+_{it} + h^-_{it}I^-_{it}) + f\sum_{t}(E_t)$$

$$- f^{rwd} \sum_{s\geq 2,t,N_{st}\neq 0} (p_s \sum_{t'=t}^{t+N_{st}-1} \frac{RLR_{t'}}{N_{st}}) + f^{pen} \sum_{s\geq 2,t} (p_s E^{excess}_t) \tag{29}$$

Subject to

$$(2) - (11), (16) - (19)$$

$$p_1 + \sum_{t,s\geq 2,N_{st}\neq 0} p_s \sum_{t'=t}^{t+N_{st}-1} (\frac{z_{t'}}{N_{st}}) \geq 1 - \alpha \tag{30}$$

$$H_t - THR - \frac{E_t}{C_t} = \Delta^+_t - \Delta^-_t \qquad \forall t \in \Omega' \tag{31}$$

$$\Delta^+_t \leq M^+_t z_t \qquad \forall t \in \Omega' \tag{32}$$

$$\Delta^-_t \leq M^-_t (1 - z_t) \qquad \forall t \in \Omega' \tag{33}$$

$$RLR_t \leq THR \qquad \forall t \in \Omega' \tag{34}$$

$$RLR_t \leq DSG z_t \qquad \forall t \in \Omega' \tag{35}$$

$$E^{excess}_t \geq C_t \Delta^-_t \qquad \forall t \in \Omega' \tag{36}$$

$$E^{excess}_t, RLR_t, \Delta^+_t, \Delta^-_t \geq 0 \qquad \forall t \in \Omega' \tag{37}$$

$$z_t \in \{0, 1\} \qquad \forall t \in \Omega' \tag{38}$$

In the objective function (29), $\sum_{i,t}(Cp_{it}X_{it} + Cs_{it}Y_{it}) + \sum_{i,t}(h^+_{it}I^+_{it} + h^-_{it}I^-_{it}) + f\sum_t(E_t)$ are the classical production costs, $f^{pen}\sum_{s\geq 2,t}(p_s E^{excess}_t)$ is the summation term that penalizes the excessive electric usage over the planning horizon, and $-f^{rwd}\sum_{s\geq 2,t,N_{st}\neq 0}(p_s \sum_{t'=t}^{t+N_{st}-1} \frac{RLR_{t'}}{N_{st}})$ is the summation term that rewards the respect of curtailment requests over the planning horizon. In this summation term, $RLR_{t'}$ is divided by N_{st}, which corresponds to the number of periods in the performance period affected by scenario s, not to overestimate the expected rewards when the electric load could only be correctly curtailed for only a smaller number of periods than N_{st}. This implies that, for scenario s, rewards are fully paid, i.e. $\sum_{t'=t}^{t+N_{st}-1} \frac{RLR_{t'}}{N_{st}} = DSG\sum_t S_{st}$, by the electric utility only if the electric load is curtailed for all the periods in Ω_{st} and therefore N_{st} periods.

Constraints (30)-(33) model the risk. Constraint (30) models the probabilistic constraint and ensures that the maximum risk level is satisfied. Note that p_1 is out of the summation term since it is the scenario that contains no request. Constraint (31) calculates the margin to the load reduction threshold, i.e. sets Δ^+_t greater than 0 if the margin is positive, and sets Δ^-_t greater than 0 if the margin is negative. Constraint (32) ensures that the binary variable z_t is equal to 1 if the load reduction threshold constraint at period t is respected. Otherwise, z_t is equal to 0 through constraint (33). If z_t is equal to one, then it increases the value of $\sum_{t,s\geq 2,N_{st}\neq 0} p_s \sum_{t'=t}^{t+N_{st}-1}(\frac{z_{t'}}{N_{st}})$, which is used to ensure that the maximum risk level is satisfied. Scenario s contributes totally, i.e. by an amount of p_s, to Constraint (30) when $z_t = 1$ for the periods associated to scenario s. Otherwise, scenario s contributes to Constraint (30) by an amount that is equal to $\frac{1}{N_{st}}$ for each period where the electric load is correctly curtailed. For these periods, the manufacturer receives rewards. For non curtailed periods, the manufacturer must pay penalties. Constraints (34)-(36) model the penalty and reward aspects. Constraint (34) calculates the saved electric load at period t. Constraint (35) limits the rewarded load reduction RLR_t to the demand saving goal and sets RLR_t to 0 when the load reduction threshold constraint is not respected for period t. Constraint (36) calculates the financial penalty associated to when the load reduction threshold is not respected at period t. Constraints (37)-(38) are the non-negativity and binary constraints.

To avoid unnecessary summation of many terms in the objective function and Constraint (30), which can lead to a large consumption of memory for a large number of scenarios when the model is implemented by using the libraries of standard solvers (that we experienced), both expressions can be reformulated. Reformulating both expressions is possible because: (1) if the load reduction threshold is not respected for a period in performance period, then it is not respected for all scenarios that contain this period, and (2) it is possible to count for each period in performance period the number of scenarios (and their associated probabilities) that contain the period. Mathematically, the reformulation is provided through Equations (39)-(41):

$$
\begin{aligned}
& -f^{rwd} \sum_{s \geq 2, t, N_{st} \neq 0} (p_s \sum_{t'=t}^{t+N_{st}-1} \frac{RLR_{t'}}{N_{st}}) + f^{pen} \sum_{s \geq 2, t} (p_s E_t^{excess}) = \\
& -f^{rwd} \sum_{t \in \Omega'} (RLR_t \beta_t) + f^{pen} (1 - p_1) \sum_t (E_t^{excess}) \qquad (39)
\end{aligned}
$$

$$
\sum_{t, s \geq 2, N_{st} \neq 0} p_s \sum_{t'=t}^{t+N_{st}-1} (\frac{z_{t'}}{N_{st}}) = \sum_{t \in \Omega'} (z_t \beta_t) \qquad (40)
$$

$$
\beta_t = \sum_{s \geq 2, t' \in \Omega' | t \in \Omega_{s,t'}} \frac{p_s}{|\Omega_{s,t'}| \sum_{t''} S_{st''}} \qquad \forall t \qquad (41)
$$

Discussion on practicability of the chance-constrained formulation. In real conditions, it is likely that the load reduction threshold constraints are violated for some time intervals during the performance period. A violation leads to incurring financial penalties calculated using a fixed rate (\$/kWh) multiplied by the excess electricity usage. This excessive electricity usage is not insignificant, and it may have serious consequences on the relationship between the manufacturer and the electric utility. Typically, this means that the first time the manufacturer does not comply with a curtailment request, it is asked to change its FSL and is tested for the new value of the parameter. If the test is conclusive, the manufacturer's incentive-based program stays active. If not, the manufacturer is asked to change this parameter until the test is conclusive or the electric utility decides to break the contract. This operation is exceptional and cannot be repeated more than a few times in practice because otherwise, the electric utility cannot count on the manufacturer for reducing its electric load when facing an upcoming power consumption peak (see e.g., Pacific Gas and Electric Company (2015) and Southern California Edison (2014). Thus, although financial penalties are authorized, they must not be ubiquitous. Financial penalties are also extremely expensive as f^{pen} is much larger than f, costing 6.50 \$/kWh (see e.g., Pacific Gas and Electric Company 2015), contrary to at most a few tens of cents for f. In other words, a mathematical model for production planning that allows not answering curtailment requests has little chance to be relevant as violating financial penalties will be too expensive to be practicable.

4. Computational study without penalties

The objectives of the computational study are to emphasize the potential of an incentive-based program for energy-intensive industries and to highlight critical parameters that should be considered when evaluating such a program.

We propose to carry out a computational study where production is scheduled for a week with daily demands. Backlogging and holding costs are incurred at the end of each day. The two-stage model is used to solve the instances.

Two metrics are used to evaluate incentive-based programs: The savings (%) and the Price of Uncertainty (PoU) in dollars. The savings represent what the manufacturer can save with the new production plan by considering an incentive-based program. They are evaluated on a yearly basis (incentive-based programs are subscribed annually) by assuming that the optimized week is representative of the year. The price of uncertainty is generally interpreted as the reachable gain if a company had access to better forecasting technologies (Gorissen, Yanıkoğlu, and den Hertog 2015). In our problem, this corresponds to having an early notice of the curtailment request, e.g., one week instead of 30 minutes, or, for each week knowing the days where the curtailment requests are received. Metrics are evaluated for each experimental case and instance size. Mathematically, both metrics are defined as follows:

$$\text{Savings (\%)} = 100 \times \frac{52 \times f^{NoRequest} - (52 \times f^{Request} - N_{year}RWD)}{52 \times f^{NoRequest}}$$

$$\text{Price of Uncertainty (\%)} = \frac{f^{Request}_{stoch.} - f^{Request}_{determ.}}{f^{Request}_{determ.}}$$

$f^{NoRequest}$ is the objective function when production is planned without an incentive-based program. $f^{Request}_{stoch.}$ are the production costs of the two-stage program with an incentive-based program. $f^{Request}_{determ.}$ are the production costs of the deterministic program with an incentive-based program. More precisely, production costs correspond to the costs without the monetary rewards and is equal to $\sum_{i,t}(Cp_{it}X_{it} + Cs_{it}Y_{it}) + \sum_{i,t}(h^+_{it}I^+_{it} + h^-_{it}I^-_{it}) + f\sum_t(E_t)$. Note that, $f^{Request}$ is either equal to $f^{Request}_{stoch.}$ or $f^{Request}_{determ.}$, depending on whether savings are computed for the deterministic case or the stochastic case. To better estimate RWD, which are the estimated monetary rewards in the planning horizon, instead of using probabilities, we will use the fact that the deterministic and two-stage programs cover all possible requests. Consequently, $RWD = f^{rwd}RLR^*$, where RLR^* is the value of the decision variable RLR of the deterministic program.

Similarly, as both the deterministic and two-stage programs cover all possible curtailment requests in the horizon, it is sufficient to compare production costs to compute the PoU as they have the same rewards.

The achieved savings are expected to cover the additional production costs caused by curtailment requests and to lower production costs significantly. The PoU is expected to be relatively high because in the deterministic case, it is possible to anticipate curtailment requests and adapt production plans.

Note that the Expected Value of Perfect Information (EVPI) and the Value of the Stochastic Solution (VSS, Birge 1982) could also be used to assess the performance of the stochastic approach. However, evaluating the EVPI can be quickly computationally intractable because it requires to solve S deterministic models of a NP-Hard problem. The VSS consists in comparing the objective function when optimizing a linear program for the expected values of the stochastic parameters with the objective function of the stochastic program. In our case, the VSS is difficult to measure for the first two-stage formulation for multiple reasons. First, the average scenario is not easy to define, both in terms of timing and time range. The average scenario is easier to define when the demand or the capacity are stochastic. Moreover, in our case, the first two-stage formulation answers any curtailment request. Even if an average scenario (e.g., Wednesday in the middle of the time range) could be defined for curtailment requests, finding a production plan for an average scenario does not guarantee that the manufacturer answers all curtailment requests, which is not compatible with the goal of the first two-stage model in this paper. For these reasons, we evaluate the stochastic program in terms of PoU. Although the PoU is more often used in robust optimization (Gorissen, Yanıkoğlu, and den Hertog 2015), it can be used in the computational study because the first two-stage model tends to determine robust solutions as it seeks to satisfy any curtailment request.

4.1. *Computational study settings*

Generation of parameters. The computational study is inspired from the glass industry, which is known to be an energy-intensive industry (Worrell 2008). In actual systems that manufacture container glass, modern production lines are able to produce about "200 containers" per minute (Worrell 2008). Assuming that on average a container weights 0.3 kg, one lot of 200 containers weights 60 kg. In the report, it is also indicated that producing one ton of container glass requires at least 2,000 kWh for the manufacturing process. If 2,000 kWh are required for one ton, it is assumed that 2kWh are required for one kg and consequently, 120 kWh for a lot. It is also assumed that the only source of energy that is used is electricity, which implies that energy costs represent about 30-50% of the total cost. Based on Worrell (2008), Tables 2 and 3 illustrate how parameters are randomly generated for simulating different products in the glass industry.

Table 2. Computational study - Traditional lot-sizing parameters.

Parameters	Cp_{it}	Cs_{it}	a_i (min)	b_i (min)	h^+_{it}	h^-_{it}
Value	U[15,22]	CpU[25,50]	U[1,3]	U[2,6]	$\frac{0.3Cp_{it}}{365}$	$10Cp_{it}$

Table 3. Computational study - Electric parameters.

Parameters	Es_i	Ep_i	P^{idl}	f ($)
Value	U[12,18]	U[12,18]*10	U[1,2]*25	0.08

Equation (43) is used to approximate the average capacity utilization rate CUR of the tool set for a day (setup times are not considered).

$$CUR = \frac{\text{Average Manufacturing Time Required To Satisfy Demands}}{\text{Total Capacity Available}} \quad (42)$$

$$CUR = \frac{\sum_i a_i \bar{d}_i}{1,440} \quad (43)$$

We generate the demands to meet a given average capacity utilization rate using Equations (42) and (43): Given a_i the manufacturing time of a lot of product i and P the number of products, equation (43) gives the average daily demand for product i. More precisely, d_{it} is drawn in the interval $[0.90 * \frac{1{,}440 \times CUR}{P \times a_i}, 1.10 \frac{1{,}440 \times CUR}{P \times a_i}]$, where 1,440 is the number of minutes in a day. Finally, the historical electric load H_t is assumed to be constant over the planning horizon: $H_t = \frac{\sum_{i,t'}(Ep_i d_{it'})}{\sum_{t'} C_{t'}}$.

For all experiments, we assume that all demands must be satisfied by the end of the planning horizon, i.e., $I^-_{i,end} = 0 \quad \forall i$. The initial inventory, $I^+_{i,0}$, and backlog, $I^-_{i,0}$, are both set to zero.

In the deterministic case, curtailment requests are assumed to be received on Wednesdays. They correspond to Scenario 2 in Appendix A and Appendix B.

Scenarios are generated using the method described in Section 3.1.1. Probabilities of receiving one curtailment request in a day are computed by using Table 1. We further assume that each curtailment request within the performance period is equiprobable. Relaxing this assumption is left for future research.

Experimental cases. We want to evaluate the influence of the incentive-based program for 12 different experimental cases (Table 4) and for 6, 9 and 15 products. These experimental cases show the sensitivity of the savings and the PoU when most the significant parameters in lot-sizing problems vary. The values of f^{rwd} and N_{st} are inspired from actual incentive-based programs (Pacific Gas and Electric Company 2015; Southern California Edison 2014). Experiments 1-3 analyze the influence of FSL with the program proposed by Southern California Edison (2014). Similarly, experiments 4-6 analyze the program of Pacific Gas and Electric Company (2015). Experiments 7-9 study the sensitivity of the setup cost ratio, and experiments 10-12 analyze the sensitivity of the average capacity utilization rate, on the additional charge and savings combined with the incentive-based program proposed by Southern California Edison (2014). Note that for studied contracts, $THR = DSG$.

Table 4. Computational study: Penalties are forbidden - Experimental cases.

Experiment	CUR(%)	Cs_{it} ($\times$ Cp_{it})	THR (%H_t)	f^{rwd}	Duration of curtailment requests (hours)	Electric contract
1	80	50	30	20	6	Southern California Edison (2014)
2	80	50	50	20	6	Southern California Edison (2014)
3	80	50	70	20	6	Southern California Edison (2014)
4	80	50	30	8	4	Pacific Gas and Electric Company (2015)
5	80	50	50	8	4	Pacific Gas and Electric Company (2015)
6	80	50	70	8	4	Pacific Gas and Electric Company (2015)
7	80	25	50	20	6	Southern California Edison (2014)
8	80	60	50	20	6	Southern California Edison (2014)
9	80	150	50	20	6	Southern California Edison (2014)
10	60	50	50	20	6	Southern California Edison (2014)
11	75	50	50	20	6	Southern California Edison (2014)
12	85	50	50	20	6	Southern California Edison (2014)

Sizes of instance: The number of time periods depends on the performance period and the length of time intervals at which the electric utility measures electricity usage. In the contract of Pacific Gas and Electric Company (2015), electricity usage is measured every 15 minutes and the performance period is between 11:00 AM and 7:00 PM. In the contract of Southern California Edison (2014), electricity usage is measured every 30 minutes and the performance period is between 12:00 AM and 6:00 PM. During the performance period, a time interval is equal to one period. Otherwise, hours out of the performance period are gathered into a single time period. We illustrate the division of the planning horizon in Table 5 and Figure 5 for the two-stage (and chance-constrained) formulation. In order to model daily demands, demands are equal to zero for all time periods except for the last period of a day (e.g. Figure 5).

Table 5. Number of time periods per day per contract.

Contract	Sunday and Saturday	Monday, ..., Friday
Pacific Gas and Electric Company (2015)	1	34
Southern California Edison (2014)	1	14

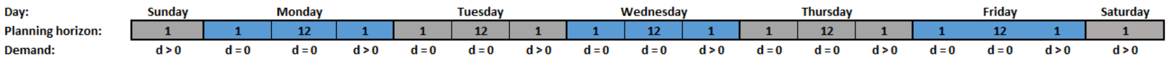

Figure 5. Division of the planning horizon for Southern California Edison (2014)

For the case where there is no curtailment request, which is used to compute potential savings, dividing the horizon into 7 periods of 24 hours is sufficient for a horizon of one week. For the deterministic case, it is assumed that a single curtailment request is received on Wednesday, which corresponds to scenario 2 in Appendix A and in Appendix B. The horizon can therefore be divided into 6 periods for Sunday, Monday, Tuesday, Thursday, Friday, and Saturday, and 14 or 18 periods depending on the incentive-based program, for Wednesday. Both cases contain much fewer periods than the stochastic formulations.

4.2. *Numerical results*

Ten data replications are generated for each instance (experimental case and size of instance). They are solved by using the Java libraries of the standard solver IBM ILOG CPLEX (version 12.9) and the strong facility location formulation, which gives tighter linear relaxations (Brahimi et al. 2017), and is presented in Appendix C. The computational time limit is set to one hour. All other settings are the default settings. An Intel-Xeon CPU E3-1240 V5 @3.5 GHz with 32GB of RAM is used. Tables 6, 7, 8 and 9 illustrate numerical results for experiments presented in Table 4.

Influence of Firm Service Level (FSL) on savings. The FSL has a strong impact on savings and the PoU.

In the deterministic case, increasing the FSL leads to more savings. Consider the contract proposed by Southern California Edison (2014) through Experiments 1, 2 and 3 (Table 6). For instance, consider the case with six products, savings are equal to 7.99% when the FSL is equal to 30% and reach 18.27% when the FSL reaches 70%. Savings are also interesting for the contract proposed by Pacific Gas and Electric Company (2015) although savings are smaller. For six products, savings reach 7.07% when the FSL reaches 70%. Savings are smaller as the number of products increases as more setups are required to satisfy the demand. This also can be explained by the fact that f^{rwd} is more than three times higher in Southern California Edison (2014) than in Pacific Gas and Electric Company (2015), and the performance period is shorter. As f^{rwd} is much higher, the additional charge incurred due to building inventories are compensated. Secondly, as the performance period is shorter, the planning horizon has fewer time periods (Table 5), which makes the second contract easier to solve with a standard solver.

Results can be quite different in the stochastic case. In the stochastic case, potential savings reach 11.07% for the contract Southern California Edison (2014) with six products, 4.76% for nine products (Table 6). The case with fifteen products is to be analyzed carefully since the gaps for the stochastic case become too large (>10%). For the contract Southern California Edison (2014), potential savings vary between -0.64% and -2.48%. Similarly, for the contract Pacific Gas and Electric Company (2015), potential savings vary between -3.85% and -30.04% (Table 7) no matter the number of products. For six and nine products, and experiments 4 and 5, as gaps are large (greater than 10%) and potential savings slightly less than 10%, there might be room for improvement by giving more time to the solver, or by developing tailored solution approaches. However, for experiment 6 for any number of products, as gaps are smaller than 8% whereas potential savings are smaller than -11%, potential savings are arguably negative. Therefore, subscribing to an incentive-based program can cost a lot of money to the manufacturer if the FSL is not well selected, in particular under the settings proposed in experiments 6, or for a large number of products, or if there is no effective method to generate a good production plan. In the deterministic case, potential savings are much higher, and actually positive, which means that being noticed of a curtailment request in advance lead to much better savings.

Although potential savings in the stochastic case can be quite significant, the PoU significantly increases when the FSL increases. For experiments 1, 2 and 3 and with six and nine products, the PoU is approximately multiplied by three when the FSL increases from 30% to 70%. For experiments 4, 5 and 6 and with six products, the PoU is also multiplied by three. With nine and fifteen products, the PoU is multiplied by almost four. This indicates that there is a large gap between potential savings in the deterministic case and the stochastic case. Additional savings could be further reached if the manufacturer knew in advance when to answer a curtailment request.

These results show that the potential savings can be significant if FSL is selected in a clever way. Note that although it is rather natural to subscribe an incentive-based program where curtailment requests are short, it can be preferable to subscribe an incentive-based program where curtailment requests are long because they are associated to a higher reward rate.

Influence of setup cost ratio on savings. As building inventories for a large number of different products requires setup operations, the higher the setup cost ratio, the lower the savings (Table 8). The larger the number of products, the more this trend is pronounced, since there are more setup operations to perform. In the stochastic case, potential savings vary between 12.95% and -8.39%.

Similarly to experiments 1 through 6, when the setup cost ratio increases, the PoU increases. However, the PoU increases to a larger extent. Between experiments 7 and 9, the PoU is approximately multiplied by five. Although the PoU becomes larger, potential savings can still be interesting either for a small setup cost ratio or a small number of products. However, experiments 8 and 9 show that negative savings can be reached if the number of products is equal to 9 or 15.

Experiments 7, 8 and 9 show that there is no point in subscribing to an incentive-based program when the setup cost ratio is large, especially for production facilities with a high number of products. Therefore, setup costs are parameters that should be carefully considered when evaluating possible gains associated to demand response programs.

Influence of the capacity utilization rate on savings. The capacity utilization rate is a critical parameter for lot-sizing problems, especially in our problem since it influences the capability of production systems to build inventories and make front-loaded production plans before possible curtailment requests. In the stochastic case, the higher the capacity utilization rate, the lower the savings (Table 9). This is due to the fact that the additional charge is increasingly expensive: For a high capacity utilization rate, large inventories have to be made and part of the demand has to be backlogged. This effect is increasingly pronounced as the number of products increases.

In the stochastic case, potential savings reach 12.03% when the capacity utilization rate is equal to 60% for six products and attain -0.24% when the capacity utilization rate is equal to 85% for nine products.The capacity utilization rate has less influence on savings in the deterministic case than in the stochastic case since curtailment requests are scheduled and it is relatively easy to adapt the production plan. Although gaps are relatively large, most often larger than 5%, in the stochastic gap, subscribing an incentive-based program can be interesting as savings can be larger than 5%.

Similarly to experiments 10, 11 and 12, when the capacity utilization rate increases, the PoU increases. For six products and nine products, the PoU is multiplied by two and three between experiments 10 and 12. For fifteen products, the savings are negative because gaps are larger than 15%. With larger computational times, it is probable that larger savings could be achieved.

Observations on optimality gaps. The deterministic case is "easily" optimized by the solver because gaps are always smaller than 1.0%, except for experiment 10 and fifteen products where the gap is equal to 1.01%. The stochastic case is much more difficult to optimize, gaps are always greater than 3%, often go beyond 10%, and the computational time limit of one hour is always reached. With a more practical meaning, it is much more difficult to plan production when short notices are sent by electric utilities. However, although stochastic solutions are suboptimal and much smaller than in the deterministic case, savings can still be interesting for the manufacturer to reduce its production costs under different parameter settings. In contrast, there may exist some settings such as in experiments 1, 2 and 3 with fifteen products, where potential savings could be positive if near optimal solutions were reached.

The difference between the deterministic and stochastic cases can be explained by several factors. In general, for large instance sizes, the studied problem is too difficult for standard solvers. This is first due to the complexity of the basic problem since the multi-item Capacited Lot-Sizing Problem (CLSP) is NP-Hard in the strong sense (Bitran and Yanasse 1982; Chen and Thizy 1990). Moreover, the feasibility problem of the CLSP when setup times are considered, as it is the case in our study, is already NP-Complete (Trigeiro, Thomas, and McClain 1989). Also, the regular time intervals used to measure electricity consumption heightens the problem complexity by adding a very large number of time periods (up to 172 periods for a week!) overcoming the benefits brought by the facility location formulation. Hence, for problems of industrial size, a tailor-made solution approach is required to determine solutions of good quality in a reasonable computational effort.

Table 6. Numerical results for experiments 1, 2 and 3.

		Case Without Request			Deterministic case			Stochastic case			Deterministic	Stochastic	
Instance ($P \times T$)	Experiment	Gap (%)	Obj.	CPU (sec.)	Gap (%)	Obj.	CPU (sec.)	Gap (%)	Obj.	CPU (sec.)	Yearly Gain (%)	Yearly Gain (%)	PoU ($)
	1	0.01	161,180	1	0.01	138,384	17	6.62	142,305	3,601	7.99	5.43	3,921
6×92	2	0.01	161,180	1	0.01	123,248	24	5.20	127,826	3,601	13.25	10.28	4,578
	3	0.01	161,180	1	0.02	108,482	375	4.14	119,521	3,602	18.27	11.07	11,039
	1	0.01	167,538	4	0.01	146,739	645	10.64	152,233	3,601	6.93	3.51	5,495
9×92	2	0.01	167,538	4	0.01	133,077	804	9.56	142,276	3,601	11.43	5.62	9,198
	3	0.01	167,538	4	0.10	119,415	714	5.30	137,443	3,602	15.94	4.76	18,028
	1	0.01	200,332	143	0.86	180,257	3,603	17.06	192,870	3,601	5.66	-0.65	12,613
15×92	2	0.01	200,332	144	0.79	167,403	3,603	15.45	185,743	3,602	9.15	-0.14	18,340
	3	0.01	200,332	145	0.53	155,016	3,602	9.27	184,423	3,602	12.42	-2.48	29,407

Table 7. Numerical results for experiments 4, 5 and 6.

		Case Without Request			Deterministic case			Stochastic case			Deterministic	Stochastic	
Instance ($P \times T$)	Experiment	Gap (%)	Obj.	CPU (sec.)	Gap (%)	Obj.	CPU (sec.)	Gap (%)	Obj.	CPU (sec.)	Yearly Gain (%)	Yearly Gain (%)	PoU ($)
	4	0.01	161,180	1	0.01	152,122	21	11.84	163,008	3,600	3.15	-3.85	10,886
6×172	5	0.01	161,180	1	0.01	146,106	15	9.90	159,776	3,601	5.25	-3.64	13,669
	6	0.01	161,180	1	0.01	140,491	98	3.25	169,074	3,602	7.07	-11.69	28,583
	4	0.01	167,538	4	0.04	159,489	1,172	17.89	178,471	3,601	2.60	-9.27	18,983
9×172	5	0.01	167,538	4	0.01	154,067	605	12.65	175,930	3,601	4.37	-9.12	21,862
	6	0.01	167,538	4	0.01	148,837	225	3.90	193,643	3,602	6.03	-21.80	44,806
	4	0.01	200,332	144	0.87	192,511	3,602	35.18	273,135	3601	2.15	-39.61	80,624
15×172	5	0.01	200,332	144	0.62	187,439	3,603	24.64	250,657	3,601	3.51	-28.27	63,219
	6	0.01	200,332	144	0.58	183,218	3,464	7.06	251,001	3,602	4.45	-30.04	67,783

Table 8. Numerical results for experiments 7, 8 and 9.

		Case Without Request			Deterministic case			Stochastic case			Deterministic	Stochastic	
Instance ($P \times T$)	Experiment	Gap (%)	Obj.	CPU (sec.)	Gap (%)	Obj.	CPU (sec.)	Gap (%)	Obj.	CPU (sec.)	Yearly Gain (%)	Yearly Gain (%)	PoU ($)
6×92	7	0.01	144,127	0	0.01	106,009	368	2.55	108,763	3,602	14.97	12.95	2,754
	8	0.01	167,919	1	0.01	130,011	26	7.48	136,564	3,601	12.70	8.62	6,553
	9	0.01	224,468	1	0.01	186,696	44	13.56	201,132	3,601	9.44	2.82	14,436
9×92	7	0.01	141,935	1	0.01	107,057	172	5.20	111,624	3,603	13.83	10.46	4,567
	8	0.01	177,625	7	0.16	143,382	1,101	10.80	153,473	3,601	10.65	4.74	10,091
	9	0.01	256,072	6	0.16	223,317	1,650	18.24	249,513	3,601	6.78	-3.71	26,197
15×92	7	0.01	158,695	43	0.50	125,349	3,379	10.33	134766	3602	11.83	5.74	9,417
	8	0.01	215,118	99	0.60	182,250	3,603	17.03	203,487	3,601	8.49	-1.66	21,237
	9	0.01	327,582	35	1.00	297,054	3,588	24.59	339,885	3,601	4.85	-8.39	42,831

Table 9. Numerical results for experiments 10, 11 and 12.

		Case Without Request			Deterministic case			Stochastic case			Deterministic	Stochastic	
Instance ($P \times T$)	Experiment	Gap (%)	Obj.	CPU (sec.)	Gap (%)	Obj.	CPU (sec.)	Gap (%)	Obj.	CPU (sec.)	Yearly Gain (%)	Yearly Gain (%)	PoU ($)
6×92	10	0.01	117,482	6	0.01	88,851	12	4.05	90,837	3,601	13.79	12.03	1,986
	11	0.01	149,098	1	0.01	113,696	398	6.36	117,666	3,601	13.31	10.53	3,970
	12	0.01	174,124	0	0.01	133,833	60	5.40	142,648	3,602	13.03	7.75	8,814
9×92	10	0.01	121,041	2	0.01	94,493	276	9.97	98,107	3,601	12.51	9.45	3,614
	11	0.01	154,255	5	0.01	121,754	422	9.93	127,737	3,601	11.81	7.75	5,982
	12	0.01	183,914	11	0.07	147,081	1393	6.33	158,879	3,603	11.18	4.37	11,798
15×92	10	0.01	141,614	29	1.01	116,260	3601	14.85	122,396	3,600	10.22	5.82	6,136
	11	0.01	182,763	308	0.99	151,855	3,603	15.04	164,195	3,601	9.42	2.53	12,340
	12	0.01	220,993	165	0.57	186,082	3,603	11.51	205,568	3,603	8.78	-0.24	19,486

5. Computational study without penalties

This computational study is designed to compare the two-stage and chance-constrained formulations, and in particular compare what would be production costs, including monetary rewards and penalties, when the occurrence of curtailment requests is unknown as it would be the case in practice. In particular, we want to study the influence of the risk level α on expected savings proposed by the chance-constrained formulation, and compare savings of the two-stage and chance-constrained formulations. Parameters are generated the same way as in Section 4.

Two metrics are used for to compare formulations: The amount of Financial Penalties (FP) in dollars that must be paid if some curtailment requests are not answered, and the Difference between the Expected Objective Function (DEOF) of both formulation. Mathematically, both metrics are defined as follows:

$$\mathrm{FP} = f^{pen} \sum_{s \geq 2, t} (p_s E_t^{excess}) = f^{pen}(1 - p_1) \sum_t (E_t^{excess})$$

$$\mathrm{DEOF}(\%) = 100 \times \frac{f_1 - f_2}{f_1}$$

FP can only be computed for the chance-constrained formulation as the two-stage formulation ensures that the manufacturer answers any curtailment request. If $DEOF$ is positive, the chance-constrained formulation leads to better production plans. If it is negative, two-stage formulation proposes better production plans.

5.1. *Experimental cases*

We want to evaluate the influence of the risk level α for 4 different experimental cases and for 6, 9 and 15 products (see Table 10).

Table 10. Computational study: Penalties are tolerated - Experimental cases.

Experiment	CUR(%)	Cs_{it} ($\times$ Cp_{it})	THR ($\%H_t$)	f^{rwd}	Duration of curtailment requests (hours)	Electric contract	f^{pen} (\$/kWh)	α
13	80	50	50	20	6	Southern California Edison (2014)	6.50	1
14	80	50	50	20	6	Southern California Edison (2014)	6.50	0.2
15	80	50	50	20	6	Southern California Edison (2014)	6.50	0.1
16	80	50	50	20	6	Southern California Edison (2014)	6.50	0.05

5.2. *Numerical results*

Similarly to the first computational study, ten data replications are generated for each instance (experimental case and size of instance). They are solved by using the Java libraries of the standard solver IBM ILOG CPLEX (version 12.9) and the strong facility location formulation, which gives tighter linear relaxations (Brahimi et al. 2017), and is presented in Appendix D. The computational time limit is set to one hour. All other settings are the default settings. An Intel-Xeon CPU E3-1240 V5 @3.5 GHz with 32GB of RAM is used.

Table 11 illustrate numerical results for experiments presented in Table 10.

It can be observed that expected production costs between production plans optimized by the chance-constrained and two-stage formulations are actually very close.

The absolute value of $DEOF$ is actually smaller than 1.05%. For six products, $DEOF$ is positive for experiments 13, 14 and 16, which means that the chance-constrained model leads to better production plans in terms of expected production costs, without paying any financial penalties ($FP = 0$), even for experiments 13 where $\alpha = 1$, i.e. when answering curtailment requests is not enforced. Financial penalties are equal to \$57 for experiments 14, which is very small. In addition, it is worth observations that $DEOF$ is actually very small, varying between 0.23% and 0.44%. For experiments 15, the two-stage formulation leads to better production plans in terms of expected production costs although production plans are not much better than the ones proposed by the chance-constrained formulation as $DEOF = -0.10\%$. Similar observations can be made when nine or fifteen products are considered. For nine products, the two-stage formulation is better for experiments 13, 14 and 15 and $DEOF$ varies between -0.33% and -0.07%. The chance-constrained formulation is slightly better for experiments 16, $DEOF = 0.29\%$ and $FP = \$6$. For fifteen products, the two-stage formulation is better and $DEOF$ varies between -1.05% and -0.36% for all experiments.

Although the chance-constrained formulation is more flexible because it can allow not answering some curtailment requests, which could theoretically lead to smaller production costs, optimized production plans rarely violate potential curtailment requests because financial penalties are too costly. Observe that financial penalties are smaller than \$200, which is arguably very small. This means that production plans optimized by both formulations are actually very close in terms of costs. Because violating curtailment requests is costly and may prevent the manufacturer from continuing to participate to incentive-based programs, we argue that using the two-stage formulation is better in practice.

Table 11. Numerical results for experiments 13, 14, 15 and 16.

		Chance-constrained model			Two-stage model				
Instance ($P \times T$)	Experiment	Gap (%)	Obj.	CPU (sec.)	Gap (%)	Obj.	CPU (sec.)	DEOF (%)	FP (\$)
6×92	13	5.80	154,262	3,602	6.41	154,550	3,601	0.23	0
	14	5.60	154,199	3,602	6.41	154,550	3,601	0.26	57
	15	6.40	154,723	3,602	6.41	154,550	3,601	-0.10	0
	16	5.29	153,936	3,602	6.41	154,550	3,601	0.44	0
9×92	13	11.16	164,354	3,601	10.64	163,814	3,601	-0.33	21
	14	10.55	164,175	3,601	10.64	163,814	3,601	-0.18	0
	15	10.34	163,953	3,601	10.64	163,814	3,601	-0.07	50
	16	10.06	163,382	3,601	10.64	163,814	3,601	0.29	6
15×92	13	17.15	206,229	3,600	17.25	204,443	3,601	-0.97	47
	14	16.83	205,285	3,601	17.25	204,443	3,601	-0.47	80
	15	17.66	206,435	3,600	17.25	204,443	3,601	-1.05	124
	16	16.98	204,965	3,601	17.46	204,461	3,601	-0.36	97

6. Managerial insights

In this section, managerial insights are provided from computational studies to guide decision-makers when assessing incentive-based programs for their manufacturing system:

- Manufacturing systems with a small setup cost ratio and a small number of products can significantly benefit from incentive-based programs,

- Manufacturing systems with low capacity utilization rates (e.g. smaller than 80%) benefit more than manufacturing systems with high capacity utilization rates (e.g. larger than 90%) because it is easier to reschedule production to satisfy curtailment requests *and* demand,
- Incentive-based programs with longer notices (a few days), or short notices but knowing in advance days when curtailment requests are received, lead to better savings than incentive-based programs with short notices (a few hours) because production is more easily planned (smaller computational complexity). If possible, negotiate with electric utilities to have large notices even if not initially proposed in the incentive-based program,
- Counter-intuitively, incentive-based programs with longer curtailment requests should be favored as they are often associated to larger monetary rewards,
- Incentive-based programs with larger time intervals in the performance period should be favored as production is more easily planned (smaller computational complexity),
- The FSL should be selected in a clever way. A large value may constrain the production system and reduce savings.
- Computational studies show that allowing the violation of curtailment requests may actually only lead to small savings on production costs while potentially preventing manufacturers from continuing to participate to incentive-based programs. It is recommended that manufacturers plan production to satisfy any curtailment request.

It is important to note that if electric utilities do not offer long notices in their incentive-based programs, it might be because predicting peaks in electric load in a few days or weeks is unreliable. However, this is arguably the most profitable aspect for manufacturing systems because they can more easily answer to curtailment requests.

7. Conclusions and industrial perspectives

In this paper, we propose new energy-aware lot-sizing models for dealing with electricity and reducing costs by collaborating with actors in the energy supply chain, especially electric utilities. Our study is focused on incentive-based demand response programs that have to be subscribed for at least a whole year. This paper is motivated by the fact that electricity has become a critical aspect of manufacturing operations and it is surprising to see that few papers have dealt with lot-sizing problems while they are as important as scheduling problems.

Firstly, we present a deterministic mathematical formulation and an example illustrating how curtailment requests can be handled. However, the practicality of the deterministic model is limited, because curtailment requests are not always scheduled well in advance. Two new stochastic models are then proposed. The first one, a two-stage model, can be used when a manufacturer needs or wants to answer any curtailment request. The second stochastic model, a chance-constrained model, can be used when not answering some curtailment requests is tolerated. A method to generate scenarios, representing possible curtailment requests, is developed by using real constraints on energy measurement in incentive-based programs after making reasonable assumptions.

Thirdly, inspired by the glass industry, two computational studies are performed. The first computational study is designed to show how incentive-based programs can reduce production costs. The study shows that high firm service levels, high setup costs and a large number of products or high capacity utilization rate make front-loaded production plans too expensive to make curtailment requests profitable. In such configurations, manufacturing systems with little operational flexibility are likely not to take advantage of incentive-based programs. Otherwise, the computational study shows that interesting savings can be achieved. The second computational is designed to compare the two-stage and chance-constrained formulations, and in particular compare what would be production costs, including monetary rewards and penalties, when the occurrence of curtailment requests is unknown as it would be the case in practice. We show that as financial penalties are very large, the optimized production plans almost never violate potential curtailment requests even tough it is allowed.

Finally, managerial insights are provided to guide decision-makers when assessing incentive-based programs for manufacturing systems.

As all curtailment requests must necessarily be answered and production cannot be rescheduled in the same day, the two-stage formulation is designed for industries with little operational flexibility. For more flexible industries where daily production rescheduling is possible, the two-stage linear program can still be used without resorting to multistage stochastic linear programming. This is possible by applying a rolling horizon approach, i.e., each time a new parameter is revealed, the two-stage linear program is solved on the remainder of the planning horizon. Another option is to develop a new two-stage stochastic programming model where production decisions until time period t are the first-stage variables and production decisions from t to T (end of the planning horizon) are the second-stage variables. We refer readers to Clark and Clark (2000), Balasubramanian and Grossmann (2004), and Curcio et al. (2018) for details on both possible approaches.

In addition, note that some industries, such as the container glass industry, are known to have sequence-dependent setups (Guimarães, Klabjan, and Almada-Lobo 2014). In this case, not including sequence-dependent setups may have an impact on production plans, thus affecting total costs and energy consumption, therefore potential savings induced by incentive-based programs. We decided to consider setup carryovers, which can also have an impact on production plans, and the inclusion of sequence-dependent setups is left for future research.

Finally, this work should be seen as a starting point for other research work on energy-aware lot-sizing problems with incentive-based programs. We believe the following perspectives are worth investigating in the future:

- Improving the modeling by, for instance, adapting the formulation when manufacturing or setup times exceed the duration of time intervals used to measure electricity consumption, and including sequence-dependent setups, which are common features for energy-intensive industries,
- Comparing the potential savings between price-based, incentive-based and both types of programs combined,
- Working on the incentive-based program by including several performance periods in a day with different reward rates, considering a time-varying price of electricity, and considering settings where $THR \neq DSG$,
- Improving the stochastic modeling by extending the two-stage formulation to a multi-stage formulation for more flexible production lines,
- Evaluating a rolling horizon approach based on the two-stage formulation without resorting to multistage stochastic linear programming to manage curtailment requests,

- Evaluating an approach where production decisions until time period t are the first-stage variables and production decisions from t to T are the second-stage variables,
- Developing new solution approaches to solve problems, for instance by proposing heuristics and meta heuristics,
- Differentiating the probability of occurrence of a curtailment request based on the hour of a day. Some hours may be more critical for electric utilities than others.

Acknowledgments

The authors would like to thank the three anonymous referees whose remarks and suggestions improved the quality of the article.

References

Balasubramanian, J, and IE Grossmann. 2004. "Approximation to multistage stochastic optimization in multiperiod batch plant scheduling under demand uncertainty." *Industrial & engineering chemistry research* 43 (14): 3695–3713.

Biel, Konstantin, and Christoph H. Glock. 2016. "Systematic literature review of decision support models for energy-efficient production planning." *Computers & Industrial Engineering* 101: 243–259.

Birge, John R. 1982. "The value of the stochastic solution in stochastic linear programs with fixed recourse." *Mathematical programming* 24 (1): 314–325.

Bitran, Gabriel R, and Horacio H Yanasse. 1982. "Computational complexity of the capacitated lot size problem." *Management Science* 28 (10): 1174–1186.

Brahimi, Nadjib, Nabil Absi, Stéphane Dauzère-Pérès, and Atle Nordli. 2017. "Single-item dynamic lot-sizing problems: An updated survey." *European Journal of Operational Research* 263 (3): 838 – 863.

Castro, Pedro M., Iiro Harjunkoski, and Ignacio E. Grossmann. 2011. "Optimal scheduling of continuous plants with energy constraints." *Computers & Chemical Engineering* 35 (2): 372–387.

Chen, W-H, and J-M Thizy. 1990. "Analysis of relaxations for the multi-item capacitated lot-sizing problem." *Annals of operations Research* 26 (1): 29–72.

Clark, Alistair R, and Simon J Clark. 2000. "Rolling-horizon lot-sizing when set-up times are sequence-dependent." *International Journal of Production Research* 38 (10): 2287–2307.

Curcio, Eduardo, Pedro Amorim, Qi Zhang, and Bernardo Almada-Lobo. 2018. "Adaptation and approximate strategies for solving the lot-sizing and scheduling problem under multistage demand uncertainty." *International Journal of Production Economics* 202: 81–96.

Dai, Min, Dunbing Tang, Adriana Giret, Miguel A Salido, and W D Li. 2013. "Robotics and computer-integrated manufacturing energy-efficient scheduling for a flexible flow shop using an improved genetic-simulated annealing algorithm." *Robotics and Computer-Integrated Manufacturing* 29: 418–429.

Denton, Frank T, Kevan L Jefferies, Dean C Mountain, A Leslie Robb, and Byron G Spencer. 1987. "The response of an industrial firm to alternative electricity an optimization model for simulation applications." *Resources and Energy* 9: 327–346.

Department of Energy. 2018. "Energy Incentive Programs." `https://www.energy.gov/eere/femp/energy-incentive-programs`.

Fethke, Gary, and Asher Tishler. 1990. "The optimal choice of inputs under time-of-use pricing and fixed-proportions technology. An application to industrial firms." *Resources and Energy* 11 (3): 241–259.

Gahm, Christian, Florian Denz, Martin Dirr, and Axel Tuma. 2016. "Energy-efficient scheduling in manufacturing companies: A review and research framework." *European Journal of Operational Research* 248 (3): 744–757.

Giglio, Davide, Massimo Paolucci, and Abdolreza Roshani. 2017. "Integrated lot sizing and energy-efficient job shop scheduling problem in manufacturing/remanufacturing systems." *Journal of cleaner production* 148: 624–641.

Giret, Adriana, Damien Trentesaux, and Vittal Prabhu. 2015. "Sustainability in manufacturing operations scheduling: A state of the art review." *Journal of Manufacturing Systems* 37: 126–140.

Golpîra, Hêriş, Syed Abdul Rehman Khan, and Yu Zhang. 2018. "Robust smart energy efficient production planning for a general job-shop manufacturing system under combined demand and supply uncertainty in the presence of grid-connected microgrid." *Journal of cleaner production* 202: 649–665.

Gorissen, Bram L, İhsan Yanıkoğlu, and Dick den Hertog. 2015. "A practical guide to robust optimization." *Omega* 53: 124–137.

Guimarães, Luis, Diego Klabjan, and Bernardo Almada-Lobo. 2014. "Modeling lotsizing and scheduling problems with sequence dependent setups." *European Journal of Operational Research* 239 (3): 644–662.

Hajej, Zied, and Nidhal Rezg. 2019. "An optimal integrated lot sizing and maintenance strategy for multi-machines system with energy consumption." *International Journal of Production Research* 1–21.

Karwan, Mark H., and Matthew F. Keblis. 2007. "Operations planning with real time pricing of a primary input." *Computers & Operations Research* 34 (3): 848–867.

Latifoğlu, Çağrı, Pietro Belotti, and Lawrence V Snyder. 2013. "Models for production planning under power interruptions." *Naval Research Logistics* 60 (5): 413–431.

Li, Yi-Chang, and Seung Ho Hong. 2017. "Real-Time Demand Bidding for Energy Management in Discrete Manufacturing Facilities." *IEEE Transactions on Industrial Electronics* 64 (1): 739–749.

Lu, Chao, Liang Gao, Xinyu Li, Quanke Pan, and Qi Wang. 2017. "Energy-efficient permutation flow shop scheduling problem using a hybrid multi-objective backtracking search algorithm." *Journal of Cleaner Production* 144: 228–238.

Masmoudi, Oussama, Alice Yalaoui, Yassine Ouazene, and Hicham Chehade. 2017. "Lot-sizing in a multi-stage flow line production system with energy consideration." *International Journal of Production Research* 55 (6): 1640–1663.

Merkert, Lennart, Iiro Harjunkoski, Alf Isaksson, Simo Säynevirta, Antti Saarela, and Guido Sand. 2015. "Scheduling and energy - Industrial challenges and opportunities." *Computers & Chemical Engineering* 72: 183–198.

Mitra, Sumit, Ignacio E. Grossmann, Jose M. Pinto, and Nikhil Arora. 2012. "Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes." *Computers & Chemical Engineering* 38: 171–184.

Pacific Gas and Electric Company. 2015. "Base interruptible program." `https://www.pge.com/tariffs/tm2/pdf/ELEC_SCHEDS_E-BIP.pdf`.

Pacific Gas and Electric Company. 2018. "Available demand response programs." `https://www.pge.com/en_US/business/save-energy-money/energy-management-programs/demand-response/demand-response.page`.

Rapine, Christophe, Guillaume Goisque, and Ayse Akbalik. 2018. "Energy-aware lot sizing problem: Complexity analysis and exact algorithms." *International Journal of Production Economics* 203: 254–263.

Rapine, Christophe, Bernard Penz, Céline Gicquel, and Ayse Akbalik. 2018. "Capacity acquisition for the single-item lot sizing problem under energy constraints." *Omega* 81: 112–122.

Salahi, Niloofar, and Mohsen A. Jafari. 2016. "Energy-Performance as a driver for optimal production planning." *Applied Energy* 174: 88–100.

Shapiro, Alexander, Darinka Dentcheva, and Andrzej Ruszczynski. 2009. *Lectures on Stochastic Programming: Modeling and Theory.* Philadelphia, PA, USA: Society for Industrial and

Applied Mathematics.

Southern California Edison. 2014. "Base interruptible program." https://www.sce.com/NR/sc3/tm2/pdf/CE195-12.pdf.

Stahl, Bojan, Marco Taisch, and Dimitris Kiritsis. 2016. "Energy management in manufacturing : From literature review to a conceptual framework." *Journal of Cleaner Production* 1–26.

Tang, Lixin, Ping Che, and Jiyin Liu. 2012. "A stochastic production planning problem with nonlinear cost." *Computers & Operations Research* 39 (9): 1977–1987.

Trigeiro, William W, L Joseph Thomas, and John O McClain. 1989. "Capacitated lot sizing with setup times." *Management science* 35 (3): 353–366.

U.S. Energy Information Administration. 2017. "Statistics and analysis of energy habits in the U.S." https://www.eia.gov/.

Wichmann, Matthias Gerhard, Christoph Johannes, and Thomas Stefan Spengler. 2019. "Energy-oriented Lot-Sizing and Scheduling considering energy storages." *International Journal of Production Economics* 216: 204–214.

Worrell, Ernst. 2008. "Energy Efficiency Improvement and Cost Saving Opportunities for the Glass Industry. An ENERGY STAR Guide for Energy and Plant Managers." *Lawrence Berkeley National Laboratory* .

Appendix A. Scenario generation for Southern California Edison (2014)

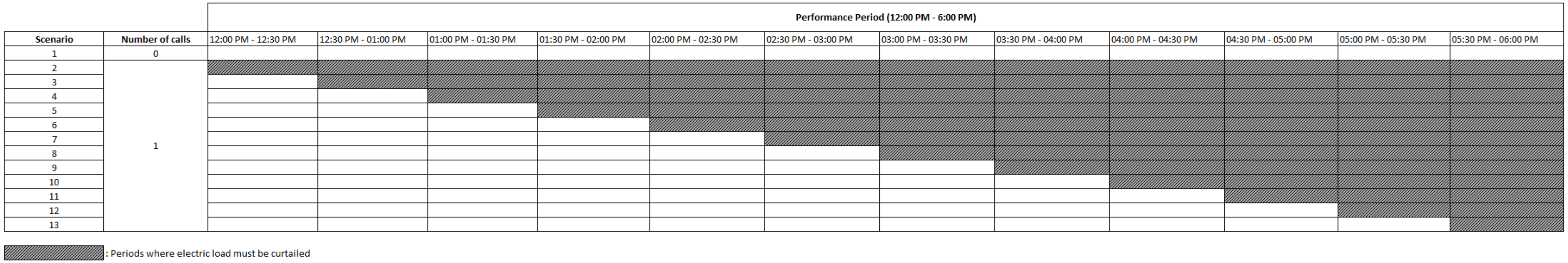

Figure A1. Scenario generation for Southern California Edison (2014)

Appendix B. Scenario generation for Pacific Gas and Electric Company (2015)

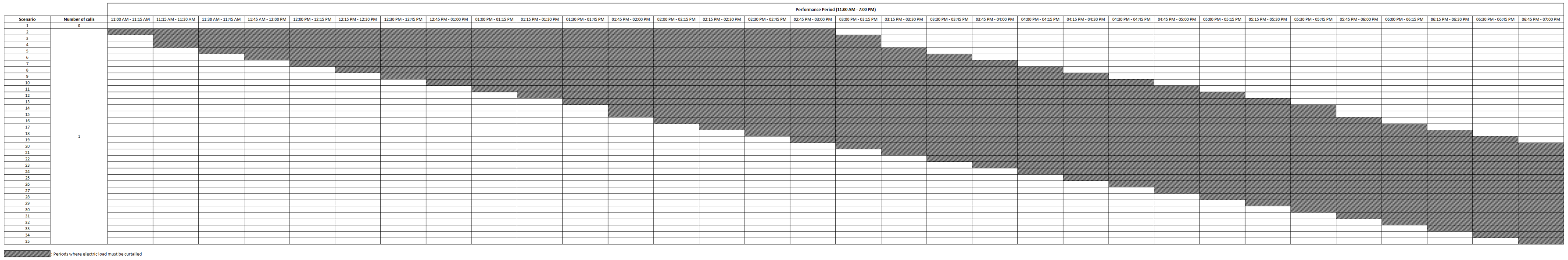

Figure B1. Scenario generation for Pacific Gas and Electric Company (2015)

Appendix C. Facility location reformulation: Penalties are forbidden

In this section, the two-stage facility location reformulation used in the computational study performed in Section 4 is presented. A new decision variable $Z_{i,t,t'}$ is introduced, which represents the quantity of product i made in period t to cover the demand at period t'. Let $hd^{+}_{i,t,t'}$ be the inventory cost for producing one unit of item i in period t and holding it until period t'. $hd^{+}_{i,t,t'} = \sum_{u=t}^{t'-1} h^{+}_{i,u}$ when $t < t'$, and is equal to 0 otherwise. Similarly, let $hd^{+}_{i,t,t'}$ be the backlogging cost for delaying the production of one unit of item i in period t for t'. $hd^{-}_{i,t,t'} = \sum_{u=t'}^{t-1} h^{-}_{i,u}$ when $t > t'$, and is equal to 0 otherwise.

$$
\begin{aligned}
&\text{Minimize} \\
&\sum_{i,t}\sum_{t'=t}^{T}(hd^{+}_{i,t,t'}Z_{i,t,t'}) + \sum_{i,t}\sum_{t'=1}^{t}(hd^{-}_{i,t,t'}Z_{i,t,t'}) \\
&+ \sum_{i,t}(Cs_{it}Y_{it} + \sum_{k=1}^{T}(Cp_{it}Z_{itk})) \\
&+ f\sum_{t}(E_t) - f^{rwd}\sum_{s\geq 2}(p_s S_{st} DSG) && && \text{(C1)} \\
&\text{Subject to} \\
&\sum_{i} W_{it} \leq 1 && \forall t && \text{(C2)} \\
&W_{it} \leq Y_{i,t-1} + W_{i,t-1} && \forall i, \forall t \geq 2 && \text{(C3)} \\
&W_{it} + W_{i,t-1} + Y_{j,t-1} \leq 2 + Y_{i,t-1} && \forall i, \forall j \neq i, \forall t \geq 2 && \text{(C4)} \\
&W_{i1} = 0 && \forall i && \text{(C5)} \\
&Z_{i,t,t'} \leq d_{i,t'}(Y_{p,t} + W_{p,t}) && \forall i, \forall t, \forall t' && \text{(C6)} \\
&\sum_{k=1}^{T} Z_{i,k,t} = d_{i,t} && \forall i, \forall t && \text{(C7)} \\
&C_t = \sum_{i}(b_i Y_{i,t} + \sum_{k=1}^{T} a_i Z_{i,t,k}) + IDL_t && \forall t && \text{(C8)} \\
&E_t = \sum_{i}(Es_i Y_{i,t} + \sum_{k=1}^{T} Ep_i Z_{i,t,k}) + P^{idle} IDL_t && \forall t && \text{(C9)} \\
&THR \leq H_t - \frac{E_t}{C_t} && \forall t \in \Omega' && \text{(C10)} \\
&Z_{i,t,t'} \geq 0 && \forall i, \forall t, \forall t' && \text{(C11)} \\
&W_{i,t} \in \{0,1\} && \forall i, \forall t && \text{(C12)} \\
&Y_{i,t} \in \{0,1\} && \forall i, \forall t && \text{(C13)}
\end{aligned}
$$

Equation (C1) includes, in order, inventory costs, backlogging costs, production costs, setup costs, energy costs, and expected financial rewards. Constraints (C2)-

(C5) are setup carryover constraints. Constraint (C6) is the new setup constraint. Constraint (C7) is the balance inventory constraint. With this constraint, all demand must be satisfied by the end of the planning horizon. Constraint (C8) ensures that capacity constraints are respected. Constraint (C9) computes the energy consumption at each period of the planning horizon. Constraint (C10) is the load reduction threshold constraint. Finally, constraints (C11)-(C13) are the binary and non-negativity constraints.

Appendix D. Facility location reformulation: Penalties are tolerated

In this section, the chance-constrained facility location reformulation used in the computational study performed in Section 5. Similarly to Appendix C, a new decision variable $Z_{i,t,t'}$ is introduced, which represents the quantity of product i made in period t to cover the demand at period t'. Let $hd^{+}_{i,t,t'}$ be the inventory cost for producing one unit of item i in period t and holding it until period t'. $hd^{+}_{i,t,t'} = \sum_{u=t}^{t'-1} h^{+}_{i,u}$ when $t < t'$, and is equal to 0 otherwise. Similarly, let $hd^{+}_{i,t,t'}$ be the backlogging cost for delaying the production of one unit of item i in period t for t'. $hd^{-}_{i,t,t'} = \sum_{u=t'}^{t-1} h^{-}_{i,u}$ when $t > t'$, and is equal to 0 otherwise.

$$
\begin{aligned}
&\text{Minimize} \\
&\sum_{i,t} \sum_{t'=t}^{T} (hd^{+}_{i,t,t'} Z_{i,t,t'}) + \sum_{i,t} \sum_{t'=1}^{t} (hd^{-}_{i,t,t'} Z_{i,t,t'}) \\
&+ \sum_{i,t} (Cs_{it} Y_{it} + \sum_{k=1}^{T} (Cp_{it} Z_{itk})) \qquad\qquad (D1) \\
&+ f \sum_{t} (E_t) - f^{rwd} \sum_{t \in \Omega'} (RLR_t \beta_t) + f^{pen} (1 - p_1) \sum_{t} (E^{excess}_t) \\
&\text{Subject to} \\
&(C2) - (C9), (C11) - (C13), (30) - (38)
\end{aligned}
$$

Equation (D1) includes, in order, inventory costs, backlogging costs, production costs, setup costs, energy costs, and expected financial rewards and penalties.