Optimal Production and Scheduling of a Process with Decaying Catalyst

Véronique M. Bizet and Ignacio E. Grossmann
Dept. of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

Nikola M. Juhasz
Arkema Inc., King of Prussia, PA 19406

DOI 10.1002/aic.10342
Published online in Wiley InterScience (www.interscience.wiley.com).

A multiperiod mixed-integer nonlinear programming (MINLP) optimization model is presented to schedule catalyst changeovers and determine the best operating policy for a chemical process with decaying performance. Two solution strategies are proposed to reduce the effect of nonconvexities and to reduce the solution time: (1) a search strategy that relies on partitioning the time horizon and (2) a strategy that makes use of the Generalized Benders Decomposition (GBD). Examples are presented for different time horizons to illustrate the performance of the proposed methods. The same objective value is reached with both strategies, which find better solutions than the full-space MINLP. For small models, the GBD algorithm is faster than the partitioning search strategy, whereas the latter method is faster for larger models. Comparison with the operating strategy used in a real plant is also presented. © 2005 American Institute of Chemical Engineers *AIChE J,* 51: 909–921, 2005
Keywords: catalyst deactivation, production planning, mixed integer planning, optimization

Introduction

The operation of chemical processes with catalysts having decaying performance over time gives rise to a challenging modeling and optimization problem (for example, see Xiong and Jutan, 2003). As the catalyst activity decays over time, process shutdowns for catalyst changeovers must be planned to restore the process performance. It is therefore necessary to develop a kinetic model that can predict the production, taking into account the catalyst deactivation. Process operations are optimized based on the deactivation and process economics. The trade-off is between the high production rates achieved from maintaining frequently renewed, high-functioning catalyst loads, and the loss in production and maintenance costs arising from catalyst changeovers.

Correspondence concerning this article should be addressed to I. E. Grossmann at ig0c@andrew.cmu.edu.

Catalyst deactivation has received attention with kinetic studies at the reactor level (Ho, 1984; Kittrel, 1982) and at the pilot-plant level (Krishnaswamy and Kittrel, 1979). Deactivation kinetic models are determined based on data sets consisting of declining conversion at constant temperature or of increasing temperature at constant conversion. Sapre (1997) proposed an alternate technique where space velocity is adjusted to maintain constant conversion. This technique, at constant temperature, gives more reliable rate information for both the catalyst deactivation and the primary catalytic reaction. The empirical kinetic models are used to determine the temperature–time relationship required under various process constraints to maintain constant conversion.

The development of optimization models for planning and scheduling of chemical processes has received significant attention over the last 10 years (Grossmann et al., 2002). The problem of scheduling multiple feeds on parallel units with decaying performance has been addressed by Jain and Grossmann (1998). This work considered a constant demand over an

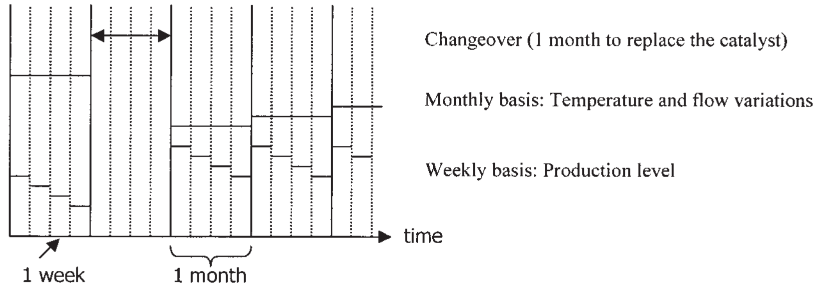

Figure 1. Discrete time representation.

infinite time horizon. Houze et al. (2003) formulated a multiperiod optimization model for the production and scheduling of catalyst changeovers in a process with decaying performance for short time horizons over which the demand varies. This model is a nonconvex mixed-integer nonlinear programming (MINLP) problem and therefore may lead to suboptimal solutions. The model proposed later makes more extensive use of disjunctive constraints and can handle much longer time horizons. In addition, two solution strategies, partitioning and Generalized Benders Decomposition, are presented to tackle the problem of nonconvexities and to reduce the solution time. The efficiency of both strategies is discussed in the section on numerical examples, where the examples are from a real-world plant.

Problem Statement

We consider a chemical process involving the use of a catalyst that has decaying performance with time. Conversion can be adjusted by manipulating the reactor temperature and the flow through the reactor. For the problem addressed in this article, we assume that we are given:

(1) a fixed time horizon (several years)

(2) the maximum number of catalyst loads that can be used over the time horizon

(3) the expected minimum and maximum life of a catalyst load

(4) the time required to shut down the process, replace the catalyst, and restart the process

(5) a kinetic empirical model that can predict the process production, taking into account the catalyst deactivation

(6) process constraints for temperature, flow, and production level

(7) inventory capacity

(8) costs for storage and changeover (variable seasonal penalty)

(9) the cost and the amount of catalyst loaded in the reactor

(10) the net profit obtained by selling the product

(11) seasonal demand figures

The optimization model must then determine:

(1) the number of catalyst loads to be used within the given time horizon

(2) the best time to schedule catalyst changeovers

(3) the best operating conditions for process parameters (flow, temperature)

(4) the level of production and inventories at each time period

In the model presented here, the time required to shut down the process, replace the catalyst load, and restart the process is estimated to be one month. The full time horizon is divided into monthly time periods. The discrete time representation considers changeovers as well as flow and temperature variations on a monthly basis but the production level is estimated on a weekly basis to take into account the catalyst deactivation (Figure 1).

In addition, we assume that the product demands are not necessarily met. The only constraint is that the sales in each time period are less than or equal to the customer demand for this period. We also assume steady-state operations for each time period.

The chemical reaction involved in the process is of the following form: $A + B \rightarrow C + D$, where C is the desired chemical product. Another compound X is fed into the reactor to minimize the catalyst deactivation. The predictive production model was determined using daily-averaged data from the plant for the production (*prod*), the temperature (T), the flow, the pressure (P), and the molecular ratios of component B to C (B/C) and X to C (X/C). The kinetic model is similar to the one presented by Houze et al. (2001). The general form of this model (KM) is the following

$$prod = K \cdot \exp\left(\frac{-Ea}{RT}\right) \cdot flow^{nf} \cdot P^{np} \cdot \left(\frac{B}{C}\right)^{nb} \cdot \left(\frac{X}{C}\right)^{nx} \cdot Act \quad \text{(KM)}$$

Here Ea is the activation energy, R is the perfect gas constant, and Act is the catalyst activity described by a first-order deactivation (DM) in the following differential equation

$$\frac{d\,Act}{d\,catage} = -kd \cdot Act \quad \text{(DM)}$$

The catalyst age (*catage*) used in this relationship is estimated by the cumulative production.

The kinetic parameters (K, Ea, nf, np, nb, nx, and kd) are determined with a nonlinear regression by the least-squares minimization. The fit-to-plant data ($R^2 = 0.847$) is illustrated in Figure 2.

The problem involves optimizing the process over all time periods and has as an objective to maximize the profit subject to constraints. Constraints include inventory balance, process requirements, kinetic equations, linking constraints, and logical relationships between decision variables. These decision vari-

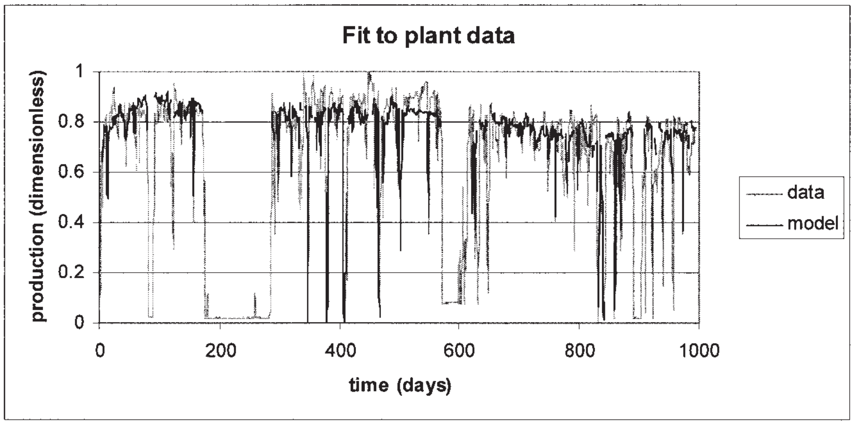

Figure 2. Fit-to-plant data.

ables are binary variables essentially used to state which catalyst load is used at a particular time period and when the changeovers occur. Constraints are both linear and nonlinear, whereas variables are discrete (binary decision variables) and continuous (production, flow, temperature, etc.).

The optimization model is first formulated as a disjunctive multiperiod model (Raman and Grossmann, 1994) before being converted into an MINLP problem (Grossmann, 2002). This model is presented in the following section.

MINLP Model

Consider optimizing the process described in the previous section over time periods (months) $i = 1 \cdots nm + 1$. A general model for the production and the scheduling optimization of such a process with decaying performance can be formulated as will be described below. We first define the sets, parameters, and variables. Then, background for the conversion of disjunctions into mixed-integer constraints is given before the optimization model itself is presented.

Definitions

Sets

- I = set of time periods (months)
- W = set of weeks
- C = set of catalyst loads

Indices

- i = time period in set I ($i = 1 \cdots nm + 1$)
- w = week in set W ($w = 1 \cdots 4$)
- c = catalyst load in set C ($c = 1 \cdots ncl$)

Parameters

- P^{lo} = weekly production lower bound
- P^{up} = weekly production upper bound
- $difP$ = weekly production increasing and decreasing bound
- T^{lo} = temperature lower bound
- T^{up} = temperature upper bound
- $increaseT$ = temperature increasing rate upper bound
- $difT$ = temperature decrease lower bound
- $flow^{lo}$ = flow lower bound
- $flow^{up}$ = flow upper bound
- Vol_c = catalyst load
- C_c = catalyst load c cost
- pen_i = changeover penalty during period i
- VM = net profit
- Cs_i = monthly storage cost
- $stock^{lo}$ = inventory level lower bound
- $stock^{up}$ = inventory level upper bound
- $stock$ = initial inventory level
- $demand_i$ = monthly demand figures
- β = penalty coefficient for unmet demand
- $minlife$ = minimum life of a catalyst load (months)
- $maxlife$ = maximum life of a catalyst load (months)
- nm = number of months in the time horizon
- $nm1$ = $nm + 1$
- ncl = number of allowed catalyst loads over the time horizon

Continuous Variables

- $Pw_{i,w,c}$ = production obtained in period i, week w with catalyst load c
- $Pm_{i,c}$ = production obtained in period i with catalyst load c
- $penalty_{i,c}$ = production loss in period i arising from catalyst load c changeover
- $CP_{i,w,c}$ = cumulative production for catalyst load c in period i and week w
- $catage_{i,w,c}$ = catalyst load c age in period i, week w
- p_i = production obtained in period i
- $stock_i$ = inventory level in period i
- $sales_i$ = sales in period i
- $unmetD_i$ = unmet demand in period i
- $T_{i,c}$ = reactor temperature in period i for catalyst load c
- $flow_i$ = flow in period i
- $Tcused_c$ = last month of use for catalyst load c

[Note that all these continuous variables are defined as nonnegative variables.]

Binary Variables

- u_c = catalyst load c changed before $i = nm$
- $y_{i,c}$ = catalyst load c changeover in period i
- $z_{i,c}$ = catalyst load c used in period i
- $a_{i,c}$ = used to define $z_{i,c}$
- $b_{i,c}$ = used to define $z_{i,c}$

Background

Binary variables appear in several disjunctions (sets of constraints of which at least one is true) involving continuous variables x (Raman and Grossmann, 1994). They are represented here with the use of logic operators OR ($\vee$) and NOT ($\neg$) as follows

$$\begin{bmatrix} z_{i,c} \\ h_{i,c}(x) \leq 0 \end{bmatrix} \vee \begin{bmatrix} \neg z_{i,c} \\ B_{i,c}x = 0 \end{bmatrix} \qquad \forall\, i, c$$

The above representation can be interpreted as follows: if catalyst load c is being used during period i ($z_{i,c} = 1$, that is, $z_{i,c}$

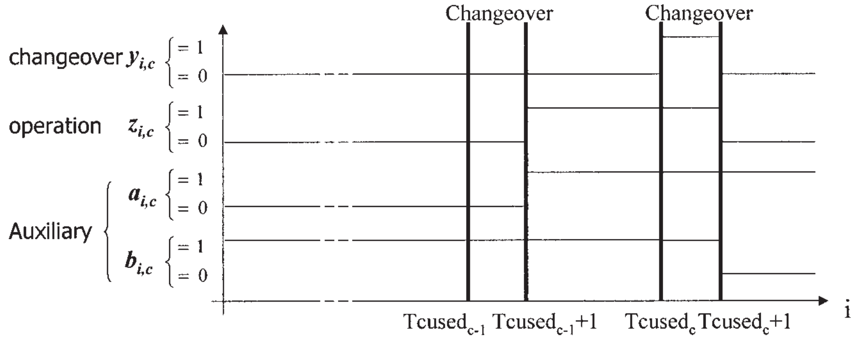

Figure 3. Binary variables definition.

is True), then enforce process constraints $h_{i,c}(x)$. If catalyst load c is not being used during period i ($z_{i,c} = 0$, that is, $z_{i,c}$ is False), a subset of continuous variables is fixed to zero through the matrix $B_{i,c}$.

These disjunctions are converted into mixed-integer constraints by using the "big-M" or the convex hull formulations. The "big-M" formulation introduces a big parameter M in the inequalities. M enforces the inequality if the corresponding binary variable equals 1 and renders it redundant otherwise. The convex hull formulation (Balas, 1985) requires disaggregating continuous variables to have a variable for each term in the disjunction. The convex hull yields a tighter formulation than the "big-M" formulation but introduces more variables into the model (Turkay and Grossmann, 1996). In the proposed model, the "big-M" formulation is used when the disjunction involves the nonlinear kinetic model ($Pweek_{i,w,c}$). The convex hull is used in the other disjunctions if an improvement is observed.

Model

We will first describe the constraints related to process decisions and the corresponding logical relationships before providing the process constraints. In the following, unless specified otherwise, all disjunctions and equations are defined for time periods, $i = 1 \cdots nm$, and for up to ncl catalyst loads.

The binary variables are defined in reference to the continuous variable $Tcused_c$. This variable is equal to the value of i during the last month of use of catalyst load c before its changeover. The dummy period $nm + 1$ is a device for the formulation. By convention, when a catalyst load c is not used during the given time horizon, the corresponding $Tcused_c$ is set to nm and the changeover occurs during the dummy period $nm + 1$.

The binary variable u_c, defined by disjunction (D1), is true (=1) if the catalyst load c is used and replaced before the end of the time horizon ($Tcused_c \leq nm - 1$). Otherwise, $Tcused_c$ is set to nm when u_c is false (=0). The convex hull formulation is given by Eqs. 1 to 3

$$\begin{bmatrix} u_c \\ Tcused_c \leq nm - 1 \end{bmatrix} \vee \begin{bmatrix} \neg u_c \\ Tcused_c = nm \end{bmatrix} \qquad \text{(D1)}$$

$$Tcused_c = Tcused1_c + Tcused2_c \qquad (1)$$

$$Tcused1_c \leq (nm - 1) \cdot u_c \qquad (2)$$

$$Tcused2_c = nm \cdot (1 - u_c) \qquad (3)$$

Equation 4 states that catalyst load $c + 1$ is used after catalyst load c and Eq. 5 ensures a minimum life for a catalyst load c between two consecutive changeovers (such as several months)

$$u_c \geq u_{c+1} \qquad \forall\ c < ncl \qquad (4)$$

$$Tcused_{c+1} \geq Tcused_c + minlife \cdot u_c \qquad \forall\ c < ncl \qquad (5)$$

Figure 3 illustrates how binary variables $y_{i,c}$, $z_{i,c}$, $a_{i,c}$, and $b_{i,c}$ are defined using $Tcused_c$.

The binary variable $y_{i,c}$ is used to identify in which period a changeover will occur. $y_{i,c}$ equals 1 when catalyst load c is being replaced during period i and 0, otherwise. $y_{i,c}$ is related to $Tcused_c$ with a sum (Eq. 6)

$$Tcused_c = \sum_{i=1}^{i=nm+1} [(i - 1) \cdot y_{i,c}] \qquad (6)$$

Equation 7 ensures that there is at most one changeover per time period. Equation 8 states that there is one changeover per catalyst load (including dummy time period $i = nm + 1$), and Eq. 9 states that this changeover occurs before the end of the time horizon if the corresponding u_c equals 1

$$\sum_{c=1}^{c=ncl} y_{i,c} \leq 1 \qquad (7)$$

$$\sum_{i=1}^{i=nm+1} y_{i,c} = 1 \qquad (8)$$

$$\sum_{i=1}^{i=nm} y_{i,c} = u_c \qquad (9)$$

The binary variable $z_{i,c}$ is used to identify those periods i over which the reactor operates with catalyst load c. $z_{i,c}$ equals

1 when catalyst load c is being used or changed during period i, and equals 0, otherwise. $z_{i,c}$ can be defined directly by disjunction (D2) for the first catalyst load ($c = 1$) but for the following ones ($c > 1$), we need two auxiliary binary variables $a_{i,c}$ and $b_{i,c}$ to define $z_{i,c}$. $a_{i,c}$ equals 1 for all periods i after catalyst load c-1 changeover and 0, otherwise. $b_{i,c}$ equals 1 for all periods i before the use of catalyst load $c + 1$ and 0, otherwise. Then, $z_{i,c}$ equals 1 when both $a_{i,c}$ and $b_{i,c}$ equal 1. $a_{i,c}$ and $b_{i,c}$ are defined with disjunctions (D3) and (D4), respectively. The "big-M" formulation is used to convert disjunctions (D2), (D3), and (D4) into mixed-integer constraints (Eqs. 10–15) because no improvement is observed with the convex hull formulation. Here, the parameter M ($nm2$) is equal to $nm + 2$. This value gives the tightest possible relaxation. Equation 16 enforces the logical relationship $a_{i,c} \wedge b_{i,c} \Rightarrow z_{i,c}$

$$\begin{bmatrix} z_{i,c} \\ i \le Tcused_c + 1 \end{bmatrix} \vee \begin{bmatrix} \neg z_{i,c} \\ i \ge Tcused_c + 2 \end{bmatrix} \qquad c = 1 \tag{D2}$$

$$i \le Tcused_c + 1 + nm2 \cdot (1 - z_{i,c}) \qquad c = 1 \tag{10}$$

$$i \ge Tcused_c + 2 - nm2 \cdot z_{i,c} \qquad c = 1 \tag{11}$$

$$\begin{bmatrix} a_{i,c} \\ i \ge Tcused_{c-1} + 2 \end{bmatrix} \vee \begin{bmatrix} \neg a_{i,c} \\ i \le Tcused_{c-1} + 1 \end{bmatrix} \qquad \forall\, c > 1 \tag{D3}$$

$$i \ge Tcused_{c-1} + 2 - nm2 \cdot (1 - a_{i,c}) \qquad \forall\, c > 1 \tag{12}$$

$$i \le Tcused_{c-1} + 1 + nm2 \cdot a_{i,c} \qquad \forall\, c > 1 \tag{13}$$

$$\begin{bmatrix} b_{i,c} \\ i \le Tcused_c + 1 \end{bmatrix} \vee \begin{bmatrix} \neg b_{i,c} \\ i \ge Tcused_c + 2 \end{bmatrix} \qquad \forall\, c > 1 \tag{D4}$$

$$i \le Tcused_c + 1 + nm2 \cdot (1 - b_{i,c}) \qquad \forall\, c > 1 \tag{14}$$

$$i \ge Tcused_c + 2 - nm2 \cdot b_{i,c} \qquad \forall\, c > 1 \tag{15}$$

$$1 - a_{i,c} + 1 - b_{i,c} + z_{i,c} \ge 1 \qquad \forall\, c > 1 \tag{16}$$

Equation 17 states that only one catalyst is being used or changed during a time period. Equation 18 enforces at least one $z_{i,c}$ to be equal to 1 if the corresponding catalyst is used and changed before the end of the time horizon ($u_c = 1$). Equation 19 ensures the maximum life of a catalyst load c

$$\sum_{c=1}^{c=ncl} z_{i,c} = 1 \tag{17}$$

$$\sum_{i=1}^{i=nm+1} z_{i,c} \ge u_c \tag{18}$$

$$\sum_{i=1}^{i=nm} z_{i,c} \le maxlife + 1 \tag{19}$$

The constraints that apply on the weekly production and the temperature when a catalyst load c is used or not during a period i, are represented by disjunction (D5), which is converted into mixed-integer constraints using the "big-M" formulation (Eqs. 20–30). Here, we use the weekly production upper bound (P^{up}) and the temperature range ($T^{up} - T^{lo}$) for the parameters M. These values give the tightest possible relaxation. When the catalyst load c is being used or replaced during period i, the corresponding weekly production is estimated from the predictive production model (Eqs. 20 and 21). The weekly production and the temperature are forced to stay between their bounds (Eqs. 22 and 23 for the production, Eq. 28 for the temperature) and some constraints apply on their increasing and decreasing rates from one month to the following one (Eqs. 24–27 for the production, and Eqs. 29 and 30 for the temperature). Otherwise, when the catalyst load c is not being used or replaced during period i, the weekly production is fixed to zero (Eq. 22) and the temperature is fixed to its lower bound (Eq. 28).

$$\begin{bmatrix} z_{i,c} \\ Pw_{i,w,c} = f(T_{i,c}, flow_i, catage_{i,w,c}) \\ P^{lo} \le Pw_{i,w,c} \le P^{up} \\ -difP \le Pw_{i,w+1,c} - Pw_{i,w,c} \le difP \qquad \forall\, w < 4 \\ [y_{i,c}] \vee \begin{bmatrix} \neg y_{i,c} \\ -difP \le Pw_{i+1,w1,c} - Pw_{i,w,c} \le difP \end{bmatrix} \qquad w = 4 \qquad \forall\, i < nm \\ T^{lo} \le T_{i,c} \le T^{up} \\ T_{i+1,c} - T_{i,c} \le increaseT \qquad \forall\, i < nm \\ T_{i,c} \ge T_{j,c} - difT \qquad \forall\, j < i \end{bmatrix} \vee \begin{bmatrix} \neg z_{i,c} \\ Pw_{i,w,c} = 0 \\ T_{i,c} = T^{lo} \end{bmatrix} \tag{D5}$$

$$Pw_{i,w,c} \le f(T_{i,c}, flow_i, catage_{i,w,c}) + P^{up} \cdot (1 - z_{i,c}) \tag{20}$$

$$Pw_{i,w,c} \ge f(T_{i,c}, flow_i, catage_{i,w,c}) - P^{up} \cdot (1 - z_{i,c}) \tag{21}$$

$$Pw_{i,w,c} \le P^{up} \cdot z_{i,c} \tag{22}$$

$$Pw_{i,w,c} \ge P^{lo} - P^{up} \cdot (1 - z_{i,c}) \tag{23}$$

$$Pw_{i,w+1,c} - Pw_{i,w,c} \leq difP + P^{up} \cdot (1 - z_{i,c}) \qquad \forall\, w < 4 \tag{24}$$

$$Pw_{i,w+1,c} - Pw_{i,w,c} \geq -difP - P^{up} \cdot (1 - z_{i,c}) \qquad \forall\, w < 4 \tag{25}$$

$$Pw_{i+1',w1',c} - Pw_{i,w,c} \leq difP + P^{up} \cdot (1 - z_{i,c} + y_{i,c}) \qquad w = 4 \qquad \forall\, i < nm \tag{26}$$

$$Pw_{i+1',w1',c} - Pw_{i,w,c} \geq -difP - P^{up} \cdot (1 - z_{i,c} + y_{i,c}) \qquad w = 4 \qquad \forall\, i < nm \tag{27}$$

$$T_{i,c} \leq T^{lo} + (T^{up} - T^{lo}) \cdot z_{i,c} \tag{28}$$

$$T_{i+1,c} - T_{i,c} \leq increaseT \cdot z_{i,c} \qquad \forall\, i < nm \tag{29}$$

$$T_{i,c} \geq T_{j,c} - difT - (T^{up} - T^{lo}) \cdot (1 - z_{i,c}) \qquad \forall\, j < i \tag{30}$$

Note that the following bounds are fixed in the model independent of the value of the binary variables

$$0 \leq Pw_{i,w,c} \leq P^{up} \tag{31}$$

$$T^{lo} \leq T_{i,c} \leq T^{up} \tag{32}$$

The temperature lower bound T^{lo} allows reduction of the number of constraints needed to derive disjunctions (D5) above and (D7) in the following.

Disjunction (D6) estimates for each catalyst load the loss in production arising from a changeover when $y_{i,c}$ equals 1 and the monthly production obtained otherwise when there is no changeover ($y_{i,c} = 0$). This production is nil if $z_{i,c} = 0$ for the corresponding time period ($Pw_{i,w,c} = 0$). The derivation of the convex hull is given by Eqs. 33 to 37

$$\begin{bmatrix} y_{i,c} \\ penalty_{i,c} = \sum_{w} Pw_{i,w,c} \end{bmatrix} \vee \begin{bmatrix} \neg y_{i,c} \\ Pm_{i,c} = \sum_{w} Pw_{i,w,c} \end{bmatrix} \tag{D6}$$

$$Pmonth_{i,c} = \sum_{w} Pw_{i,w,c} = Pm1_{i,c} + Pm2_{i,c} \tag{33}$$

$$Pm1_{i,c} \leq 4 \cdot P^{up} \cdot y_{i,c} \tag{34}$$

$$Pm2_{i,c} \leq 4 \cdot P^{up} \cdot (1 - y_{i,c}) \tag{35}$$

$$penalty_{i,c} = Pm1_{i,c} \tag{36}$$

$$Pm_{i,c} = Pm2_{i,c} \tag{37}$$

Note that the weekly production upper bound P^{up} is multiplied by 4 to obtain the monthly production upper bound.

The last disjunction (D7) enforces the initial temperature just after a changeover. This disjunction is converted into mixed-integer constraints with the convex hull formulation (Eqs. 38–40)

$$\begin{bmatrix} y_{i,c} \\ T_{i+1,c+1} = T^{lo} \qquad \forall\, i < nm \qquad \forall\, c < ncl \end{bmatrix} \vee [\neg y_{i,c}] \tag{D7}$$

$$T_{i+1,c+1} = T1_{i+1,c+1} + T2_{i+1,c+1} \qquad \forall\, i < nm \qquad \forall\, c < ncl \tag{38}$$

$$T1_{i+1,c+1} \leq T^{lo} \cdot y_{i,c} \qquad \forall\, i < nm \qquad \forall\, c < ncl \tag{39}$$

$$T2_{i+1,c+1} \leq T^{up} \cdot (1 - y_{i,c}) \qquad \forall\, i < nm \qquad \forall\, c < ncl \tag{40}$$

We need additional constraints to estimate the cumulative production (Eq. 41) used to define the catalyst age (Eq. 42) appearing in the production model, the monthly production (Eq. 43) used in the inventory balance (Eqs. 44a and 44b), and the unmet demand (Eq. 45) appearing in the objective function

$$CP_{i,w,c} = \sum_{ip \leq i} Pm_{ip,c} + \sum_{wp \leq w} Pw_{i,wp,c} \tag{41}$$

$$catage_{i,w,c} = \frac{CP_{i,w,c}}{Vol_c} \tag{42}$$

$$p_i = \sum_{c} Pm_{i,c} \tag{43}$$

$$stock0 + p_i = stock_i + sales_i \qquad i = 1 \tag{44a}$$

$$stock_{i-1} + p_i = stock_i + sales_i \qquad \forall\, i > 1 \tag{44b}$$

$$unmetD_i = demand_i - sales_i \tag{45}$$

Finally, the following bounds are provided for the sales, the inventory, and the flow through the reactor

$$sales_i \leq demand_i \tag{46}$$

$$stock^{lo} \leq stock_i \leq stock^{up} \tag{47}$$

$$flow^{lo} \leq flow_i \leq flow^{up} \tag{48}$$

The inventory lower bound ($stock^{lo}$) is identical for all periods of the time horizon except the last one. Indeed, it may be worth providing a higher value at the end of the time horizon because the model does not consider the future changeovers or demand variations. The value of this inventory lower bound at the end of the time horizon can be determined by studying the inventory cyclic profile in the results of the optimization model without a specific last inventory level lower bound.

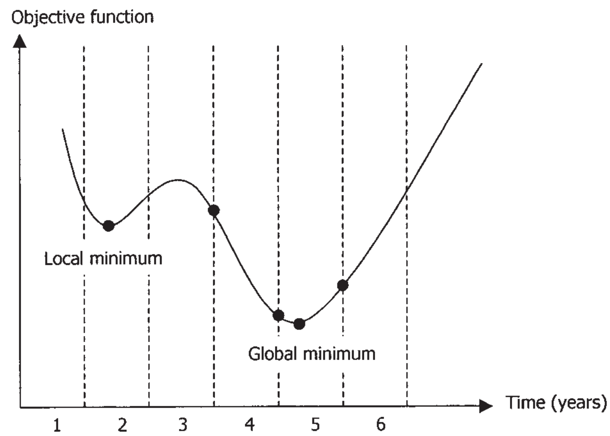

Figure 4. Partitioning search strategy.

Based on the above definitions and constraints, the objective function for minimizing the negative of the profit (that is, maximizing the profit) is given below. The profit is equal to the income from the sales minus costs for storage and changeover, and a penalty for unmet demand. The objective is subject to all the constraints previously defined

$$\min -VM\sum_{i=1}^{i=nm} sales_i + \beta \cdot VM \sum_{i=1}^{i=nm} unmetD_i + \sum_{i=1}^{i=nm}\left(pen_i \sum_{c=1}^{c=ncl} y_{i,c}\right) + \sum_{c=1}^{c=ncl}(Cc_c \cdot u_c) + \sum_{i=1}^{i=nm}(Cs_i \cdot stock_i)$$

s.t. Eqs. 1–48

Solution Strategies

In contrast to the model developed by Houze et al. (2001) for shorter time horizons, the proposed model can handle much longer time horizons because it involves fewer "big-M" constraints and incorporates the convex hull of some of the disjunctions (D1, D6, and D7). However, the resulting MINLP has two computational difficulties. The first is that the nonlinear kinetic equations introduce nonconvexities into the model, and therefore this can lead to suboptimal solutions. To overcome this problem, we propose a search strategy that partitions the time horizon of the multiperiod problem. In addition to this difficulty, the MINLP becomes very large as the number of time periods increases because the number of constraints and variables increases with each additional time period. To tackle this problem we propose to use the Generalized Benders Decomposition (Geoffrion, 1972). Both solution strategies are presented in the next two sections.

Partitioning search strategy

The optimization model is not convex; therefore, finding the global optimum cannot be guaranteed. DICOPT, the MINLP solver used in the modeling system GAMS (Brooke et al., 1998) may find suboptimal solutions depending on the initial point provided. This solver makes use of the outer-approximation (OA) algorithm (Duran and Grossmann, 1986; Viswanathan and Grossmann, 1990), which involves iterating between a nonlinear programming (NLP) subproblem where the binary variables are held fixed and a mixed-integer linear programming (MILP) master problem where the model is linearized at the NLP solution point. To avoid using deterministic global optimization methods we propose a search strategy that partitions the time horizon of the multiperiod problem. Although not rigorous, it requires much less computational effort than a rigorous global search and often yields near global optimum solutions.

The basic idea is to divide the full time horizon into one-year intervals. The strategy constitutes two steps:

Step 1. Solve MINLP subproblems with DICOPT for selected intervals. The changeovers are forced to occur within those selected intervals. All possible intervals considering the minimum and maximum life of a catalyst load are successively tested (Figure 4). A local optimum is found for each selection. For two or more changeovers, the enumeration of intervals must be performed in a tree search to account for combinations of intervals over which the changeovers may occur (Figure 5).

Step 2. The optimum solution is selected as the best one over the intervals of terminal nodes in the tree search.

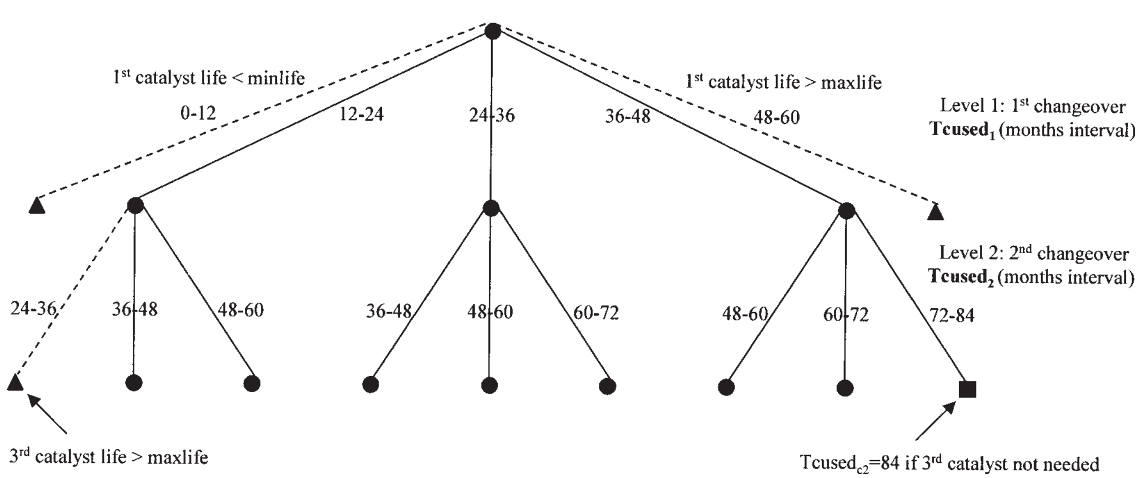

Figure 5. Tree search.

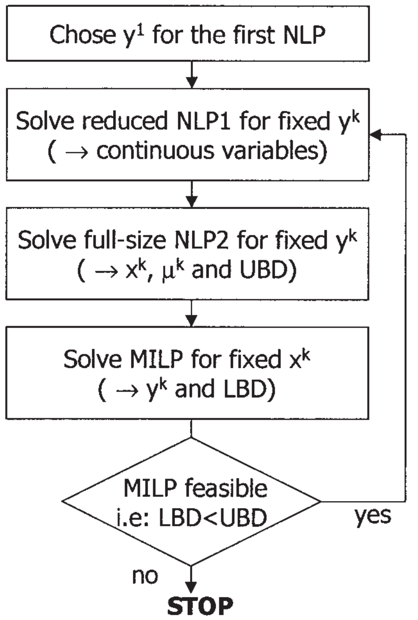

Figure 6. Steps of GBD algorithm.

Figure 4 illustrates the partitioning of the time horizon for the case of two catalyst loads and one changeover. The black dots represent the local optima found for each interval where the changeover is forced to occur. In this example, the global optimum corresponds to a changeover during the fifth year.

When we have more than two catalyst loads, and thus two or more changeovers, the combination of intervals is generated with a tree where each terminal node is enumerated and its corresponding subproblem solved. Figure 5 illustrates the tree when we consider three catalyst loads over a 7-year time horizon (84 months). The minimum life for a catalyst load is 1 year (12 months) and its maximum life is 4 years (48 months). The structure of the tree is as follows. Level 1 contains the intervals to be tested for changeover of the first catalyst load. The second level is for changeover of the second catalyst load and so on. The circle and square nodes represent all the tested intervals for $Tcused_c$, the last months of use of catalyst loads c. The corresponding changeovers occur during the following month. The triangle nodes represent examples of the intervals that are not tested because of the maximum or minimum catalyst life. This tree search does not require use of the three available catalyst loads. If the third catalyst load is not necessary, this solution will be found at the square node. In that case, the last month of use of catalyst load 2 is scheduled during the last period of the time horizon ($Tcused_2 = 84$) and the second changeover does not occur (it is scheduled during month 85). Note that in the tree search of Figure 5, a total of eight subproblems must be solved.

Generalized Benders Decomposition

The algorithm presented in Figure 6 consists in solving successively an NLP subproblem in two steps providing an upper bound (UBD) and an MILP master problem providing a lower bound (LBD). Consider that the MINLP presented earlier is written as follows

$$\min Z = c^T y + f(x)$$

$$\begin{aligned} \text{s.t.}\quad & h(x) = 0 \\ & g(x) + Ay \le 0 \\ & By \le b \\ & y = \{0, 1\}^m \\ & x \in X = \{x | Dx \le d, x^L \le x \le x^U\} \end{aligned}$$

The NLP subproblem corresponding to this MINLP with fixed binary variables is as follows

$$\min \quad Z = c^T y^k + f(x)$$

$$\begin{aligned} \text{s.t.}\quad & h(x) = 0 \\ & g(x) + Ay^k \le 0 \ \rightarrow \ \mu^k \\ & x \in X \end{aligned}$$

Note that the logical relationships ($By \le b$) between binary variables are not needed and that Lagrange multipliers μ^k are obtained to derive the cuts in the MILP master problem. When all binary variables are held fixed, the value of the continuous variable $Tcused_c$ is also fixed. Therefore, all the constraints involving only $Tcused_c$ and binary variables can be suppressed from the NLP subproblem. To save computational time, this simplified version, NLP1, is first solved to find the continuous variables' optimal values (x^k), which are provided as the initial point to solve NLP2, the complete NLP. It is necessary to solve NLP2 to obtain all the Lagrange multipliers μ^k used to define the Benders cut in the master problem. NLP2 includes Eqs. 1–3, 5, 6, 10–15, and 20–48 from the initial MINLP, whereas NLP1 includes only Eqs. 20–48. Equations 1–3, 5, 6, and 10–15 are not needed for NLP1 because they involve only $Tcused_c$ and binary variables as explained previously. In addition to the Benders cuts defined with fixed x^k and μ^k, the MILP master problem given below includes the integer constraints from the initial MINLP. To strengthen the lower bound, it also contains all linear constraints from the initial MINLP. It helps to find a feasible integer solution for the NLP subproblem

$$Z = \min \alpha$$

$$\begin{aligned} \text{s.t.}\quad & \alpha \ge c^T y + f(x^k) + (\mu^k)^T [g(x^k) + Ay] \qquad k = 1, 2 \cdots K \\ & \alpha \le UBD^{best} - \varepsilon \\ & By \le b \\ & \left.\begin{aligned} & \alpha \ge c^T y + f(x) \\ & Dx + Ey \le d \end{aligned}\right\} \text{additional constraints} \\ & y \in \{0, 1\}^n, x \in X, \alpha \in \Re^1 \end{aligned}$$

where $Dx + Ey \le d$ is the set of linear constraints from the initial MINLP.

The algorithm stops when the lower bound given by the

Table 1. Example 1: Comparison between the Empirical Strategy and Optimization Results

	Empirical Strategy	Optimization
Changeovers	2	1
Profit	127.502	137.761
Storage cost	11.842	8.193

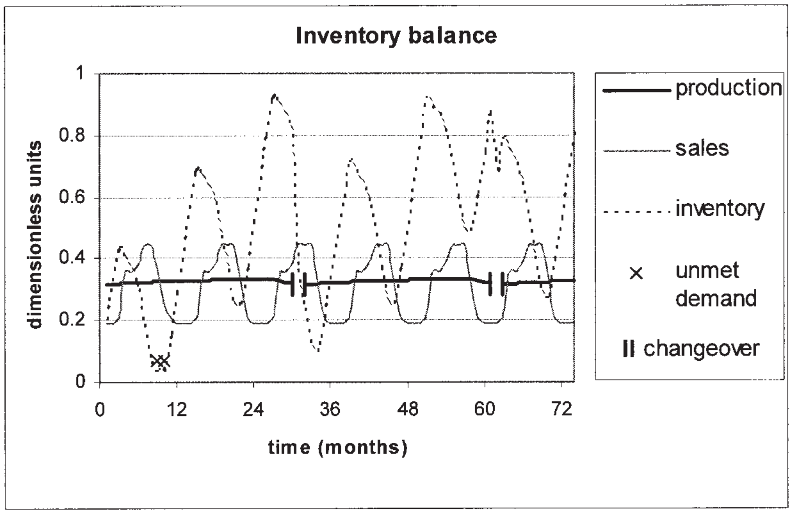

Figure 7. Empirical strategy, inventory balance.

master problem exceeds the upper bound given by the NLP subproblem.

Numerical Examples

The performance of both solution strategies is tested with two examples presented in this section. The results of the first example are compared with the empirical strategy used at the plant. Finally, the computational results are compared between the proposed methods and the simultaneous solution obtained by solving the full-space MINLP with the OA algorithm. For confidentiality reasons, the objective function value (profit) is reported in dimensionless form.

The examples were implemented and solved on GAMS 21.1 (Brooke et al., 1998) with an Intel P-4, 2.4-GHz processor with 512 MB of RAM. The code CPLEX 8.1 was used for solving MILP problems and CONOPT3 for the NLP subproblems.

Example 1: comparison with current strategy

In this section, we can compare the optimization results with the empirical strategy currently used at the plant to show the improvement that can be made by using the multiperiod MINLP model. The empirical strategy consists in maintaining a constant conversion by increasing the reactor temperature. The catalyst load is replaced when the production starts to decrease significantly.

In this first example, we consider a 74-month time horizon. Table 1 and Figures 7 to 10 illustrate the comparison between the empirical strategy and the optimization results. The empirical strategy performs two changeovers and the corresponding profit is about 127.5. By contrast, the optimization model schedules only one changeover and the profit is significantly higher at 137.8. This difference in the profit value is attributed to the number of changeovers and the storage cost.

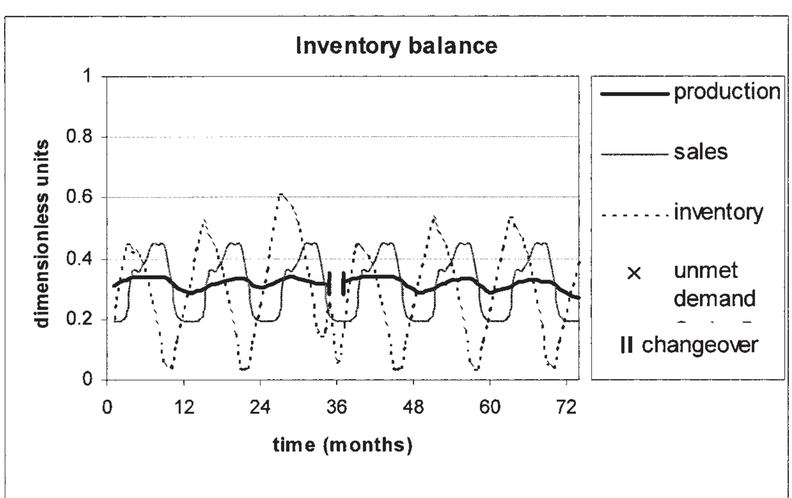

Figure 8. Optimization results (Example 1), inventory balance.

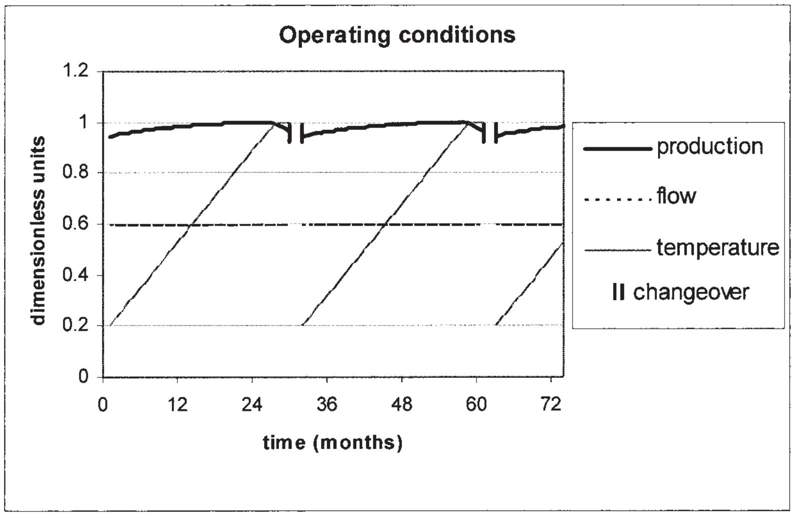

Figure 9. Empirical strategy, operating conditions.

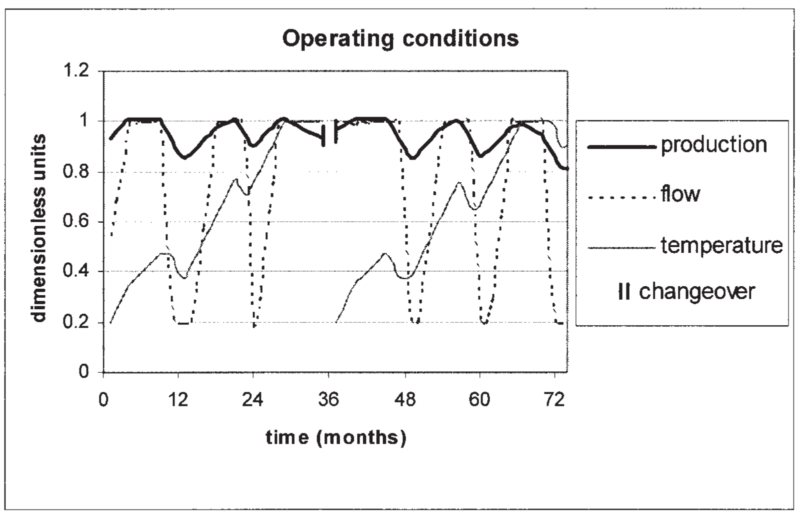

Figure 10. Optimization results (Example 1), operating conditions.

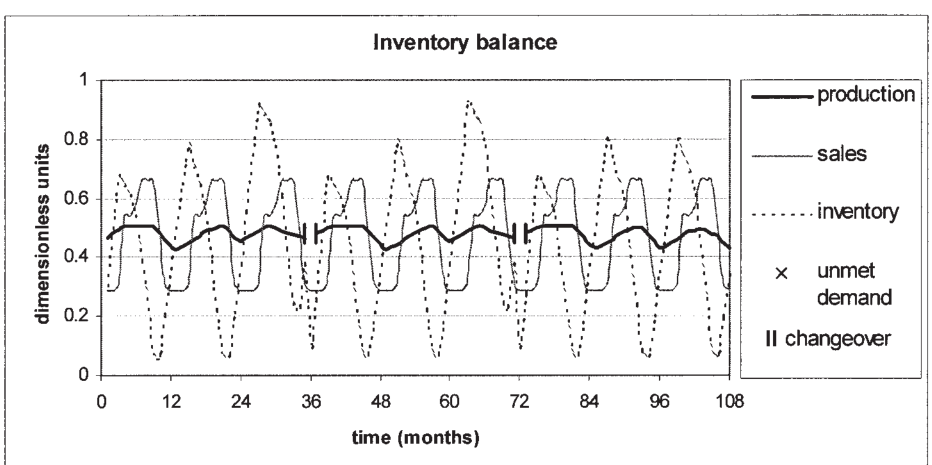

Figure 11. Inventory balance (Example 2).

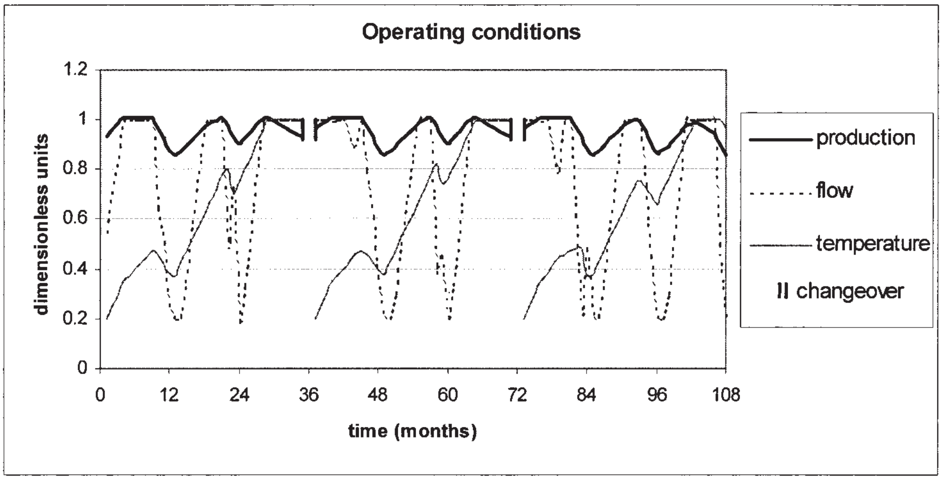

Figure 12. Operating conditions (Example 2).

Table 2. Model Sizes

Example	Number of Catalyst Loads	Time Periods	Discrete Variables	Continuous Variables	Constraints	Nonlinear Nonzeroes
1	2	74	334	3,717	12,596	3,552
2	3	108	816	8,013	33,230	7,776

The production, inventory, temperature, and flow profiles shown in Figures 7–10 for both cases provide an explanation of the results shown in Table 1. These profiles look quite different because the empirical strategy does not take into account the seasonal demand figures and the storage cost. The production level predicted by the optimization model follows the seasonal demand variations thanks to the optimal process conditions for temperature and flow (Figure 10). In particular, the production level is reduced during the low demand seasons but at a level that ensures that the sales can continue during the shutdown period. With the empirical strategy, a slowly increasing temperature maintains the production approximately constant (Figure 9). Consequently, the average inventory level (Figure 8) is higher than that obtained after optimization (Figure 7) because the production level is too high during the low demand season. In addition, the catalyst loads are used more quickly. By contrast, the optimization model manages to save production capacity during the low demand seasons and then to increase the catalyst life compared to the empirical strategy.

Example 2

Example 2 considers three catalyst loads over a 9-year time horizon. The optimization results are shown in Figures 11 and 12, which provide the corresponding production, inventory, sales, temperature, and flow profiles. There is no unmet demand in this case; the sales are equal to the demand figures for each time period. The cycles observed for the production and the inventory levels are attributed to the fact that the production level follows the seasonal demand variations (Figure 11). In addition, the inventory level increases before the two changeovers to meet the demand during the shutdown period. In that case, there is no specific lower bound for the inventory level at the end of the time horizon (Eq. 47). Then, there is a bound effect at the end of the time horizon because the model does not consider the following changeover, which we can predict. As for the first two loads, the inventory level should increase at the end of the time horizon. This problem is avoided by increasing the last inventory level lower bound. The value of this inventory lower bound at the end of the time horizon can be determined by studying the inventory cyclic profile in the results of the optimization model without this specific last inventory lower bound.

Table 3. Example 1: Two Catalyst Loads over a 74-Month Time Horizon

Solution Strategy	Changeover (period)	Profit	CPU Time (s)	Iterations/Subproblems
Simultaneous	37	137.688	93.80	4 major iterations (DICOPT)
Partition	36	137.761	132.95	4 subproblems
GBD	36	137.761	58.36	10 major iterations

The optimal operating conditions for temperature and flow (Figure 12) are quite similar but not identical for the first two cycles. They lead to the same production profile because several combinations of temperature and flow can produce exactly the same level of production. Adding some constraints on the flow or the temperature avoids these differences, but in that case the model is more restrictive. The third cycle is somewhat different because of the bound effect.

Computational results

Finally, we will compare the results obtained with both solution strategies and by solving the full-space MINLP for the two examples presented in the previous sections. Table 2 compares the size of both models. The number of constraints

Table 4. Example 2: Three Catalyst Loads over a 108-Month Time Horizon

Solution Strategy	Changeovers (period)		Profit	CPU Time (s)	Iterations/Subproblems
Simultaneous	48	—	191.771	351.35	3 major iterations (DICOPT)
Partition	36	72	200.370	1880.06	15 subproblems
GBD	36	72	200.370	3369.80	54 major iterations

Table 5. Example 1: Partitioning Search Strategy

Tcused Interval (period)	Changeover (period)	Profit	Major Iterations	CPU Time (s)
18–30	25	136.65	4	35.64
30–42	36	137.76	5	30.3
42–54	48	135.03	5	33.25
54–66	59	124.45	5	33.76

and variables (continuous and discrete) in the second example is about twice that in the first one.

The partitioning search strategy and the Generalized Benders Decomposition lead to the same objective for both examples. The solution is slightly better than the simultaneous solution in Example 1 and significantly better in Example 2. The simultaneous solution was obtained by solving the full-space MINLP with DICOPT. In addition, the GBD algorithm was tested with all possible binary combinations to initialize the first NLP. The same solution is always obtained but the number of iterations and the solution time required to converge vary slightly. Therefore, both strategies seem to yield the global optimum, although there is no rigorous guarantee.

The efficiency or solution time of both methods depends on the model size. For small models (Example 1, Table 3), the GBD algorithm is faster. It is twice as fast as the partitioning search strategy (58.36 CPUs compared to 132.95 CPUs), and it is even faster than the simultaneous solution (93.80 CPUs), which leads to a slightly worse objective function value. Indeed, both the partitioning search and GBD solution strategies schedule the changeover during period 36 to reach a profit of 137.761, whereas the simultaneous solution schedules the changeover during period 37 to reach a profit of 137.688.

When we consider larger models (Example 2, Table 4), the partitioning search strategy is much faster than the GBD algorithm. Solution of Example 2 takes only 1880.06 CPUs compared to 3369.8 CPUs for the GBD algorithm. This is explained by the number of iterations required to make the GBD algorithm converge: 54 iterations are needed for Example 2 compared to only 10 for Example 1. Indeed, with three catalyst loads (Example 2), the number of binary combinations is quite high and the algorithm adds only one cut at each iteration.

For this second example the simultaneous solution is much faster because it requires only 351.35 CPUs, although it gives a solution that has lower profit than the other two solution strategies (191.771 vs. 200.37). The simultaneous solution schedules only one changeover during period 48 to reach a profit of 191.771, whereas the other two methods schedule two changeovers during periods 36 and 72 to reach a profit of 200.370, which is substantially higher. Therefore, although both the partitioning search and GBD strategies require more computational effort, they provide higher-quality results.

Finally, Tables 5 and 6 show the results of the partitioning search strategy for both examples. The minimum and maximum life of a catalyst load are 18 and 60 months, respectively. For each of the tested intervals, the tables provide the changeover period, the corresponding profit, the number of major iterations needed to solve the model, and the solution time. The results show how the profit decreases when the changeover periods depart from the best solution. In addition, the number of intervals to test is of course much higher when more than one changeover is considered. Indeed, we need to test only four intervals in the search strategy for Example 1, which considers only one changeover; however, 15 tests are necessary for Example 2, which considers up to two changeovers.

Conclusions

A multiperiod mixed-integer nonlinear programming model has been proposed to optimize process operations and to sched-

Table 6. Example 2: Partitioning Search Strategy

$Tcused_c$ Intervals		Changeovers (period)				
1st Catalyst	2nd Catalyst	$c1$	$c2$	Profit	Major Iterations	CPU Time (s)
18–30	36–48	24	49	190.92	4	160.31
	48–60	24	61	199.17	4	101.66
	60–72	25	61	199.25	4	92.87
	72–84	25	73	197.61	5	138.1
30–42	48–60	36	61	199.27	4	120.02
	60–72	36	72	200.37	4	139.12
	72–84	36	73	200.3	4	115.23
	84–96	37	85	197.38	4	107.28
42–54	60–72	48	72	197.77	5	144.22
	72–84	48	73	197.78	4	115.5
	84–96	48	85	197.55	4	128.08
	96–108	48	97	194.05	5	121.22
54–66	72–84	59	84	187.22	4	129.19
	84–96	59	85	187.21	5	169.61
	96–108	59	—	189.89	4	97.65

ule catalyst changeovers in a process with decaying performance. Based on seasonal demand figures, the model optimizes the reactor temperature, the flow through the reactor, and the inventory level, and schedules the shutdowns to replace the catalyst. Two solution strategies, a partitioning search strategy and Generalized Benders Decomposition, were developed to reduce the effect of nonconvexities and to reduce the solution time. Two examples from a real-world plant process have been solved using the proposed methods. Numerical results have illustrated the improvements achieved by using the optimization model compared to the empirical strategy used at the plant, and the benefits of applying the proposed solution strategies. The partitioning search strategy was faster than the Generalized Benders Decomposition for the large example, but for the smaller example this trend was reversed.

Acknowledgments

The authors gratefully acknowledge the financial support from TOTAL and the Center for Advanced Process Decision-Making at Carnegie Mellon University.

Literature Cited

Balas, E., "Disjunctive Programming and a Hierarchy at Relaxations for Discrete Optimization Problems," *SIAM J. Algebra Discr. Methods*, **6**, 466 (1985).

Brooke, A., D. Kendrick, A. Meeraus, and R. Raman, *GAMS: A User's Guide.* GAMS Development Corp., Washington, DC (1998).

Duran, M. A., and I. E. Grossmann, "An Outer Approximation Algorithm for a Class of Mixed-Integer Nonlinear Programs," *Math. Program.*, **36**, 307 (1986).

Geoffrion, A. M., "Generalized Benders Decomposition, " *J. Opt. Theory Appl.*, **10**, 237 (1972).

Grossmann, I. E., "Review of Non-Linear Mixed Integer and Disjunctive Programming Techniques for Process Systems Engineering, " *Opt. Eng.*, **3**, 227 (2002).

Grossmann, I. E., S. A. van den Heever, and I. Harjunkoski, "Discrete Optimization Methods and Their Role in the Integration of Planning and Scheduling," *AIChE Symp. Ser. 32*, **98**, 150 (2002).

Ho, T. C., "Some Aspects of the Constant Conversion Policy Dealing with Catalyst Deactivation, " *J. Catal.*, **86**, 48 (1984).

Houze, M., I. E. Grossmann, and N. Juhasz, "Optimization Model for the Production and Scheduling of the Catalyst Loads Changeouts in a Process with Decaying Performance, " In: Grossmann, I. E., and C. M. McDonald, eds. *Proc. of FOCAPO2003*, pp. 311–314 (2003).

Jain, V., and I. E. Grossmann, "Cyclic Scheduling of Continuous Parallel Process Units with Decaying Performance, " *AIChE J.*, **44**, 1623 (1998).

Kittrel, J. R., "Modeling Laboratory Deactivation Data, " *Ind. Eng. Chem. Process. Des. Dev.*, **21**, 778 (1982).

Krishnaswamy, S., and J. R. Kittrel, "Analysis of Temperature–Time Data for Deactivating Catalyst, " *Ind. Eng. Chem. Process. Des. Dev.*, **18**, 399 (1979).

Raman, R., and I. E. Grossmann, "Modeling and Computational Techniques for Logic Based Integer Programming, " *Comput. Chem. Eng.*, **18**, 563 (1994).

Sapre, A. V., "Catalyst Deactivation Kinetics from Variable Space-Velocity Experiments, " *Chem. Eng. Sci.*, **52**, 4615 (1997).

Turkay, M., and I. E. Grossmann, "Disjunctive Programming Techniques for the Optimization of Process Systems with Discontinuous Investment Costs—Multiple Size Regions," *Ind. Eng. Chem. Res.*, **35**, 2611 (1996).

Viswanathan, J., and I. E. Grossmann, "A Combined Penalty Function and Outer-Approximation Method for MINLP Optimization, " *Comput. Chem. Eng.*, **14**, 769 (1990).

Xiong, Q., and A. Jutan, "Continuous Optimization Using a Dynamic Simplex Method, " *Chem. Eng. Sci.*, **58**, 3817 (2003).

Manuscript received Mar 2, 2004, and revision received July 6, 2004.