Flex Serv Manuf J (2017) 29:184–202
DOI 10.1007/s10696-016-9244-8

Solving integrated production and condition-based maintenance planning problems by MIP modeling

Fahimeh Shamsaei[1] · **Mathieu Van Vyve**[2]

Published online: 3 May 2016

Abstract It has been demonstrated that integrating maintenance scheduling and production planning can lead to substantial savings. However in these works the associated optimization problems are solved by enumerating all (exponentially many) maintenance schedules. This has led these researchers to either solve instances of very small sizes, or consider cyclic maintenance schedules only to limit the possibilities to a small (polynomial) number. We show here how to formulate these problems as (strong) mixed-integer linear programs, and then solve them using off-the-shelve MIP solvers. We demonstrate the efficiency of the proposed approach to solve problems of up to 10 products and 24 time periods, sizes that were simply unreachable before. We also illustrate the value of using non-cyclic maintenance schedules when the demand varies over time, with savings of up to 41 % compared to cyclic schedules.

Keywords Maintenance scheduling · Integrated models · Lot sizing · Non-cyclic maintenance · Production planning

1 Introduction

Performing maintenance of productive equipment is a standard and natural practice in almost all industries. Without maintenance equipment will usually gradually consume more raw material and energy, operate at lower speed, lead to more defect products or to more down-time. The increase of consumption in raw material and

✉ Fahimeh Shamsaei
fahimeh.shamsaei@gmail.com

1 Louvain School of Management, Université catholique de Louvain, Louvain-La-Neuve, Belgium

2 Center for Operations Research and Econometrics, Université catholique de Louvain, Louvain-La-Neuve, Belgium

energy, the increase of defect products and the corrective maintenance after failure will all increase the production cost per unit, while the lower speed of operation, the increase of defect products and the increase of down-time after failures will all decrease the nominal production capacity of the production system. Therefore, performing regular preventive maintenance to restore the equipment to an 'as-good-as-new' state is usual practice in most production systems. However, performing a maintenance is itself costly because it might necessitate the usage of specialized equipment or human skills and will also usually lead to a duration of reduced or zero capacity. Therefore plant managers do not want to perform maintenance too often either.

Lot sizing is the optimal timing and sizing of production lots in the face of dynamic, time-varying demand. The most important feature of lot sizing models are economies of scale related to production. More precisely, larger batches lead to lower production cost because of the presence of fixed costs, or to higher production capacity through the reduction of setup times. So the key trade-off in lot sizing models is between just-in-time, low-inventory but high-cost production plans, and production plans in large batches leading to low production cost and high capacity utilization but high inventory. Production planning generalizes lot sizing by also considering multiple products and multiple machines. An excellent reference book on this topic is Pochet and Wolsey (2006).

Maintenance scheduling and production planning are interdependent because the former essentially determines the production capacity that is used as input data of the latter. However, performing these two functions sequentially remains the usual approach in the industry, probably for simplicity and the belief that the inefficiency is small among other reasons. Nonetheless, it has been known since at least the work of Iravani and Duenyas (2002) that making maintenance and production decisions separately can be rather costly and that there are significant benefits for making these decisions in an integrated fashion. The reason is intuitive. If the maintenance schedule is determined without taking into account the full knowledge available about time-varying level of customer demand, it could be that a machine is in maintenance status during a period where full capacity is most needed.

The present paper makes the following contributions. We give the first mixed integer linear programming (MILP) formulation for non-cyclic maintenance and the effect of production interruptions as a result of maintenance on the lot sizing problem. We also provide several alternative capacity models (i.e. different ways in which production decisions influence the production capacity) that substantially broaden the range of applicability of the proposed models. Next we show that these MILP formulations enable us to solve much larger problem instances (24 periods and 10 products) than could previously be solved (8 time periods and 2 products). Finally, we demonstrate the benefits of allowing non-cyclic maintenance schedules over cyclic ones.

The structure of this paper is as follows. In the next section we discuss the different types of maintenance and give a brief literature review. Section 3 is the key section and presents the integrated maintenance and production planning problems and several variants and extensions, each time with the associated MILP formulation. An illustrative example is presented in Sect. 4. We report on extensive computational experiments in Sect. 5, where we also show the benefits of non-cyclic over cyclic maintenance schedules. Finally, concluding remarks are given in Sect. 6.

2 Types of maintenance and literature review

During last decade and even today, reactive maintenance (fixing or replacing equipment after it fails) or proactive maintenance (assuming a certain level of performance degradation, with no input from the machinery itself, and servicing equipment on a routine schedule whether service is actually needed or not) have been existing widely in mechanical industry.

Proactive maintenance is done without being caused by some form of failure. Proactive maintenance can take one of two forms, condition-based maintenance (CBM) or preventive maintenance (PM) while corrective maintenance (CM) is a form of reactive maintenance. PM is a set of activities that are regularly performed on plant equipment, machinery, and systems before the occurrence of a failure in order to protect them and to prevent or eliminate any degradation in their operating conditions. In contrast, CBM typically involves active monitoring of the state of the equipment. When the condition is deemed too low, a maintenance is performed and the machine is restored to its nominal state (Garg and Deshmukh 2006).

The objective of any form of maintenance is to buy the reliable utilization of a machine for some period of time at a low price. In practice, the cost of reactive maintenance is much greater than a form of proactive maintenance that will provide an equivalent period of machine utilization (Hu et al. 2012). Therefore, considering proactive maintenance in decision making in maintenance scheduling usually provides more benefits than reactive maintenance (although the latter cannot be completely eliminated).

CBM involves some form of activity aimed at identifying the presence of deterioration or defining the extent of deterioration that already exists. At the end of this task, more is known about the machine, but the condition of the machine has not been changed. Actual maintenance is then performed if needed. Since CBM has been implemented to continuously deliver data related to the machines status and performance, the decisions making in proactive maintenance focus on CBM which promises to reduce downtime, spares inventory, maintenance costs, hazards and accidents.

In the model we consider here, the deterioration of the equipment is modeled by a gradual reduction of the capacity. This is used to take various effects into account like increased defect products, decrease of production rate or increased needs of corrective maintenance after failure. Since we assume in the model that we know the capacity of the equipment at all times, the model we consider is closer to the concept of CBM. However, CM is also taken into account (although in expectation only) by the gradual reduction of capacity.

The impact of machine breakdown and CM on the economic lot sizing decisions was first considered by Groenevelt et al. (1992); however, less attention has been paid to it during the last two decades (Chakraborty et al. 2008). Plenty of works have been devoted to integrated production planning and maintenance scheduling where failure rate is deterministic or stochastic. The constant failure rate was considered by Groenevelt et al. (1992) where it is assumed that a certain fraction of the items produced will be diverted to the safety stock in order to satisfy the demand during repair time of the machine. Since then several researchers Kim et al. (1997) have extended the above seminal work to fit into various realistic situations. Various probability functions have been considered to reflect stochasticity of failure rates among which general distribution (Ben-Daya 2002; Wang and Meng 2009; Chakraborty et al. 2008), exponential distribution (Lee and Rosenblatt 1989; Groenevelt et al. 1992; Berg et al. 1994; Lin and Gong 2006; Zhang et al. 2014), Weibull distribution (Fituhi and Nourelfath 2012; Lin et al. 2011), Gamma distribution (Aghezzaf et al. 2007) are the most frequent ones. Stochasticity of integrated production planning and maintenance scheduling, on the other hand, can be attributed to the required repair time to bring a machine back into an in-control state where distribution function of the repair time is exponential (Berg et al. 1994), or uniform (Chakraborty et al. 2008; Kim et al. 1997).

Many of the researchers have considered cyclic (or periodic) maintenance in their PM models (Aghezzaf et al. 2007). We will not propose an extensive literature review of this topic here, but the interested reader can find a recent one in Grigorieva et al. (2006). In periodic PM we deal with optimizing how often each task in a set of predefined tasks should be executed. Moreover, there is a relatively limited literature on models presenting a general (not necessarily periodic) preventive maintenance policy (Fituhi and Nourelfath 2012; Aghezzaf and Najid 2008). The objective of these models is to determine either the best time for doing preventive replacements by new items, i.e., perfect PM (Yao et al. 2004), or the optimal sequence for imperfect maintenance actions (Levitin and Lisnianski 2000).

Since the integrated maintenance scheduling and production planning falls in the realm of combinatorial optimization, developing efficient solution methods is as important as developing efficient models. Hence, beside mathematical programming-based approaches (Ferik 2008; Chakraborty et al. 2009; Ashayeri et al. 1996; Lee and Chen 2002; Aramon Bajestani et al. 2014), Lagrangian relaxation and decomposition technique (Aghezzaf and Najid 2008; Lu et al. 2013) are used to find bounds for these problems. Moreover, meta-heuristics such as genetic algorithm (Sortrakul et al. 2005; Levitin and Lisnianski 2000), simulated annealing (Fituhi and Nourelfath 2014) have been extensively used to find good quality solutions.

In Ashayeri et al. (1996) a mixed integer linear programming model is developed to simultaneously plan preventive maintenance and production on multiple production lines, where each line has one bottleneck machine. The proposed model only indicates whether or not to produce a certain product in a certain period on a certain production line. The model schedules production jobs and preventive maintenance jobs, while minimizing costs associated with production, back orders, corrective maintenance and preventive maintenance. A branching solution procedure is suggested and its performance is discussed.

Aramon Bajestani et al. (2014) have formulated an integer programming model for the problem of integrated production planning and maintenance scheduling in a multi-machine production system in which a Markov decision process model is used to determine the maintenance schedule. Xiang et al. (2014) have used the same process in Aramon Bajestani et al. (2014) for the stochastic demand.

The model presented in Lee and Chen (2002) deals with the scheduling of production and preventive maintenance on a set of parallel machines where each machine must be maintained once during the planning horizon. The objective is to schedule jobs and maintenance activities so that the total weighted completion time is minimized. They studied two cases; in the first, the machines can be maintained simultaneously if necessary, and in the second only one machine can be maintained at any given time. They showed that, even when all jobs have the same weight, both cases of the problem are NP-hard. They also proposed a branch-and-bound algorithm based on column generation for solving both cases of the problem.

A three-stage heuristic including a Lagrangian based heuristic is proposed by Lu et al. (2013) to solve integrating CLSP and run-based preventive maintenance scheduling problem which is restricted by some reliability constraints.

Sortrakul et al. (2005) have used genetic algorithm (GA) in an integrated single-machine preventive maintenance scheduling and production planning. The goal is to find the optimal PM actions and job sequence minimizing the total weighted expected completion time.

Fituhi and Nourelfath (2012) have considered an integrated non-cyclic preventive maintenance and tactical production model for a single machine which determines simultaneously the optimal production plan and the instants of preventive maintenance actions. The problem is solved by enumerating all possible maintenance schedules in which the number of possibilities increases exponentially. They show substantial benefits for allowing non-cyclic maintenance schedules compared to cyclic ones.

Our work differs from these previous results as follows:

- The problem we consider is the first to integrate the following elements: lot-size decision in each period, economies of scale in production, time-varying demand and scheduling of maintenance. In particular, most previous works only consider time-varying demand but only by assuming that demand is a random variable drawn from a given (time-independent) distribution. However, in industrial practice, knowledge of future demand is usually much more precise than that. Firm orders or pre-orders usually constitute indeed a large fraction of total demand in the short-term. This knowledge has to be incorporated in the model to be able to place maintenance when capacity is least needed. The demand has to be truly time-varying. Also, fixed cost and economies of scale are ubiquitous in productions systems. Therefore, our model is more usable in practice.
- Our model enables us for the first time to compute optimal solutions for problems of more practical size (5 products, 10 periods) than before. Previous works that argued integrating maintenance scheduling decisions with lot-size decisions is important, did so by restricting to either cyclic maintenance schedules or problems of very small sizes. Therefore we broaden significantly the support for integrating these two decisions by demonstrating that this conclusion generalizes to a much larger class of problems.

- The solution approach we propose here (MIP modeling) is much simpler yet more effective than previous works. Developing decomposition heuristics or meta-heuristics for a similar problem requires a substantial amount of programming. The model we describe is just twenty lines of code in a high-level modeling language. Also, any improvement in the MILP solver will automatically translate in improved computing times or solution quality. These features makes it an attractive solution approach in industrial applications.

3 Problem statement

3.1 Motivation and description

The proposed model in this paper has been inspired from a real problem in Lime Industry where each kiln[1] is utilized to produce different products. The kiln should be gradually heated to reach a predefined temperature required to feed the kiln by raw materials. This high temperature gradually weakens the kiln's wall which is made of refractory bricks. Moreover, the kiln's capacity decreases progressively because of the deposit in its bottom. So, it becomes necessary to perform a global maintenance of the kiln either because the productivity is too low, or because of the rift in the walls.

A production cycle begins with a new system (an in-control state) in which the kiln can use its full capacity to produce items of acceptable quantity and quality. As time goes by, the capacity of the kiln decreases. However, after a period of time in production, the process may shift to an out-of-control state, in which the production cycle must end. The production cycle can also be terminated earlier, to restore the capacity to its nominal value by a CBM.

The capacity decrease of the kiln is in first approximation proportional to the incremental deposit at the bottom of the kiln, which is proportional to the production amount in the last period. Assuming production is at full residual capacity, this implies an exponential decrease of the capacity. This is the first and most stylized capacity model that we will consider in Sect. 3.2.1. Several generalizations will be considered next.

The exponentially decreasing capacity model that we consider is very similar to the one discussed in Aghezzaf et al. (2007), where they present a natural model of production with preventive maintenance. In the latter work, the probability of machine failure and loss of capacity increases over time, unless a preventive maintenance is performed. Since everything is dealt with in expectation, they eventually obtain a model with capacity decreasing exponentially over time, until maintenance is performed. Because they only consider cyclic maintenance schedules, there are only as many different maintenance schedules as time periods. An efficient solution procedure is then to enumerate all possible maintenance schedules. The main scientific contribution of this paper is the improved

[1] An industrial oven

formulation for integrated production and maintenance planning problem when the capacity decreases over time. Moreover, as we consider general (non-cyclic) maintenance, there are an exponential number of possible schedules. Therefore enumerating all maintenance schedules will not be an efficient approach in our case.

3.2 Mathematical models

We consider a single machine on which a set of products P must be produced during a given planning horizon T consisting of N fixed periods with length τ, i.e., $T = N\tau$. For each product $i \in P$, a demand d_{it} must be satisfied at the end of period $t \in T$. The machine has a nominal capacity c_{max} which decreases exponentially with $0 \leq \alpha \leq 1$ in a production cycle. It is also assumed that if maintenance is performed in period $t \in T$ it starts from the very beginning of period t. Our integrated maintenance and production problem consists of a multi-product single machine capacitated lot sizing problem and a maintenance problem. The decisions involve determination of lot sizes in each period and the periods in which maintenance is performed. Our objective is minimizing the total production, inventory, set up, and maintenance cost, while satisfying the demands for all products over the planning horizon. The required sets, parameters, and decision variables used in our model are the following.

Sets and parameters

T: Set of periods in the planning horizon
P: Set of products
α: Capacity reduction coefficient
τ: Length of each period
f_{it}: Fixed production cost of product i in period t
p_{it}: Variable production cost per unit of product i in period t
h_{it}: Inventory holding cost per unit of product i by the end of period t
d_{it}: Demand for product i in period t
m_t: Maintenance cost in period t
c_{max}: Maximum (nominal) capacity of the machine

Decision variables

x_{it}: Lot size of product i in period t
I_{it}: Inventory level of product i at the end of period t
c_t: Available capacity in period t

$$y_{it} : \begin{cases} 1 & \text{if product i is produced in period t} \\ 0 & \text{otherwise} \end{cases}$$

$$z_t : \begin{cases} 1 & \text{if maintenance is performed in period t} \\ 0 & \text{otherwise} \end{cases}$$

3.2.1 The integrated non-cyclic maintenance and production model (NCMP)

$$\min \sum_{t \in T} \left(\sum_{i \in P} (f_{it} y_{it} + p_{it} x_{it} + h_{it} I_{it} \right) + m_t z_t) \quad (1)$$

Subject to:

$$x_{it} + I_{i,t-1} = d_{it} + I_{it}; \quad \forall t \in T \ , \ i \in P \quad (2)$$

$$x_{it} \leq \left(\sum_{s \in T, s \geq t} d_{is} \right) y_{it}; \quad \forall t \in T \ , \ i \in P \quad (3)$$

$$\sum_{i \in P} x_{it} \leq c_t; \quad \forall t \in T \quad (4)$$

$$c_t = max(\alpha c_{t-1}, c_{max} z_t); \quad \forall t \geq 2 \quad (5)$$

$$x_{it} \geq 0 \ , \ I_{it} \geq 0 \ , \ c_t \geq 0 \ , \ y_{it} \in \{0,1\} \ , \ z_t \in \{0,1\}; \quad \forall t \in T, i \in P \quad (6)$$

The objective function (1) consists of the total production, inventory, set up, and maintenance costs. The first set of constraints (2) is the flow conservation constraints defined for all products $i \in P$ and all periods $t \in T$. They guarantee that summation of the available inventory of a product and its produced quantity in each period is sufficient to satisfy demand of the pertinent period; the remainder inventory is obviously stocked for the subsequent periods. The second set of constraints (3) enforces $x_{it} = 0$ if $y_{it} = 0$ and does not impose a constraint on x_{it} if $y_{it} = 1$. In these constraints the quantity $\sum_{s \in T, s \geq t} d_{is}$ is an upper bound for x_{it}. Constraints (4) are the capacity restrictions defined for each period. They guarantee that the produced quantity in each period does not exceed the available capacity in the related period. Constraints (5) describe the evaluation of the capacity. If no maintenance is performed in period t ($z_t = 0$), c_t decreases exponentially at a rate of α where $0 \leq \alpha \leq 1$. If maintenance is performed in period t ($z_t = 1$), c_t is restored to the nominal capacity c_{max}.

Note that the same problem without maintenance and given fixed capacities c_t is the well-known multi-item big-bucket lot sizing problem. It is strongly NP-Hard, and can be very difficult in practice (Miller et al. 2003).

The non-linear capacity function(5) can be linearized as follows.

$$c_t \leq c_{max}; \quad \forall t \in T \quad (7)$$

$$c_t \leq \alpha c_{t-1} + c_{max} z_t; \quad \forall t \geq 2 \quad (8)$$

3.2.2 More complex capacity models

We describe here more complicated capacity models, and show that the proposed approach can accommodate many different situations. In particular, we consider situations where the capacity decrease depends on the previous production level, on the presence of a set up, or also follows an arbitrary given law between maintenances (i.e. other than exponential).

In some situations when a single machine is used to produce different products, the capacity of each period depends not only on the capacity of the previous period but also on the type of products and their lot sizes. For example, a kiln in lime industry can be used to produce various types of lime. Each product has its own recipe[2], and has a specific affect on capacity reduction. It can be mathematically modeled as:

$$c_t = max\left(\alpha c_{t-1} - \sum_{i \in P} \beta_i x_{i,t-1}, c_{max} z_t\right); \quad \forall t \geq 2 \tag{9}$$

where capacity is additionally reduced by a term proportional to the production level $x_{i,t-1}$ and a given parameter $0 \leq \beta_i \leq 1$. This constraint is linearized by following the same technique we used for constraint (5):

$$c_t \leq c_{max}; \quad \forall t \in T \tag{10}$$

$$c_t \leq \alpha c_{t-1} - \sum_{i \in P} \beta_i x_{i,t-1} + c_{max} z_t; \quad \forall t \geq 2 \tag{11}$$

In the previous models, it was shown how capacity and production quantity of one period affect the capacity of the following periods. It is necessary to accurately model capacity utilization to obtain feasible production plans. This often requires one to model the capacity consumed when a machine starts a production batch, or when a machine switches from one product to another. In these cases, we obtain so-called set-up or start-up time models, changeover time models, or models with sequencing restrictions. In some applications, there is wear-and-tear especially when the equipment is started up. This is typical in the case of engines. To incorporate this into our model we introduce a new binary variable w_t equal to 1 if the equipment is started up in period t. Following Pochet and Wolsey (2006, p 323), the start-up can be modeled as:

$$w_t \geq y_{it} - y_{i,t-1}; \quad \forall i, t \tag{12}$$

Indeed, a start-up has to be incurred if production of some item i occurs in period t, but not in period t-1. Then our capacity function can be altered as follows:

$$c_t = max(\alpha c_{t-1} - \gamma w_{t-1} c_{t-1}, c_{max} z_t); \quad \forall t \geq 2 \tag{13}$$

where $0 \leq \gamma \leq 1$ is a non-negative coefficient represents loss of capacity. This can be similarly linearized as

$$c_t \leq \alpha c_{t-1} - \gamma c'_{t-1} + c_{max} z_t; \quad \forall t \geq 2 \tag{14}$$

where

[2] A recipe specifies the percentage of each ingredient of lime

$$c_t' = c_t w_t; \qquad \forall t \tag{15}$$

Constraint (12) forces $w_t = 1$ if $y_{it} = 1$ and does not enforce any value on w_t if $y_{it} = 0$. The capacity in period t (c_t) is the maximum value of c_{max} and $(\alpha - \gamma)c_{t-1}$ if $w_t = 1$ and the maximum value of c_{max} and αc_{t-1} if $w_t = 0$.

$$c_t' \le c_t; \qquad \forall t \tag{16}$$

$$c_t' \le c_{max} w_t; \qquad \forall t \tag{17}$$

$$c_t' \ge 0; \qquad \forall t \tag{18}$$

$$c_t' \ge c_t - c_{max}(1 - w_t); \qquad \forall t \tag{19}$$

The set of constraints (16)–(19) linearizes constraint (15). They enforce $c_t' = c_t$ if $w_t = 1$ and $c_t' = 0$ if $w_t = 0$.

Constraints (11) and (14) can of course be combined to obtain the more general form:

$$c_t = max(\alpha c_{t-1} - \sum_{i \in P} \beta_i x_{i,t-1} - \gamma c_t', c_{max} z_t); \qquad \forall t \ge 2 \tag{20}$$

In some applications, the required time to do maintenance is more than one period, i.e., for a couple of periods during maintenance there is no production. Since the required time for maintenance might occupy a few full periods plus a fraction of one period, the capacity of the first possible period after maintenance might be a fraction of the nominal capacity. For instance, in lime industry the maintenance actions take more than 2 weeks. Moreover, after the maintenance the kiln cannot produce with its full capacity.

We introduce the following new parameters to model these features:

p_k: the percentage of the nominal capacity available after k periods after the start of the last maintenance

Let us define the following new variable:

$$w_{tk} : \begin{cases} 1 & \text{if in period t, the last maintenance is performed in period k } (k \le t) \\ 0 & \text{otherwise} \end{cases}$$

and the associated constraint

$$w_{tk} = w_{t-1,k}(1 - z_t); \qquad \forall t > k \tag{21}$$

Constraint (21) forces $w_{tk} = 0$ if $z_t = 1$, which means that the last maintenance in period t has been performed in the same period. The linear constraints (22)–(25) are equivalent to the set of the non-linear constraint (21) show this point better.

$$w_{tk} \le w_{t-1,k}; \qquad \forall t > k \tag{22}$$

$$w_{tt} = z_t; \qquad \forall t > k \tag{23}$$

$$w_{tk} \leq 1 - z_t; \qquad \forall t > k \tag{24}$$

$$w_{tk} \geq w_{t-1,k} - z_t; \qquad \forall t > k \tag{25}$$

The general capacity function can then be written as follows:

$$c_t = (p_1 w_{tt} + p_2 w_{t,t-1} + p_3 w_{t,t-2} + \cdots) c_{max} \tag{26}$$

It implies that if we are in period t and maintenance is performed in this period then the capacity in period t is $p_1 c_{max}$, i.e., the machine has restored $p_1\%$ of its full capacity.

3.2.3 Cyclic maintenance schedules (CMP)

Several authors have described similar problems but with the additional restriction that the maintenance schedule must be cyclic. For all the capacity models presented, we show how to model this situation by adding the following constraints:

$$z_t + z_{t+k} \leq z_{t+2k} + 1; \qquad \forall t \in T, \;\; t + 2k \leq N\tau \tag{27}$$

$$z_t + z_{t+k} \leq z_{t-k} + 1; \qquad \forall t \in T, \;\; t - k \geq 1, \;\; t + k \leq N\tau \tag{28}$$

The set of constraints (27), (28) forces maintenance in period $t + 2k$ and $t - k$ when maintenance occurs in period t and $t + k$. Therefore, only maintenance schedules of the form $\tau, \tau + \delta, \tau + 2\delta, \tau + 3\delta, \ldots,$ where δ is the periodicity and $\tau \leq \delta$ is the period of the first maintenance, are feasible in this model.

Clearly, a cyclic schedule is somewhat contradictory to the idea of a CBM, where the decision to perform a maintenance depends on the actual state of the machine. However, it is interesting to see that modeling cyclic schedules is rather simple, and we also use it in the next sections to compare our models and results with earlier works.

4 Numerical example

We illustrate our modeling and algorithmic approach on a numerical example taken from Aghezzaf et al. (2007). Let us consider a production process with the planning horizon of 10 production periods, each with an available maximal capacity of $c_{max} = 50$. The machine has to produce two kinds of products in lots so that the demands are satisfied. For each product, the periodic demands, set up, production, holding and maintenance costs are presented in Table 1. As Table 1 shows, the periodic demands have been generated randomly in a specific range.

The value of using non-cyclic maintenance action when the demand varies from one period to another is illustrated. To get the optimal solution, this problem is

Table 1 Problem's data

Product	d_{it}	f_{it}	p_{it}	h_{it}	m_t
1	(0,30)	40	10	2	50
2	(0,30)	40	10	2	50

Table 2 Optimal production and maintenance planning for CMP and NCMP

Period	CMP					NCMP				
	x_{1t}	x_{2t}	y_{1t}	y_{2t}	z_t	x_{1t}	x_{2t}	y_{1t}	y_{2t}	z_t
1	16	19	1	1	1	16	19	1	1	1
2	27	12.4	1	1	0	22	9	1	1	0
3	32	0	1	0	0	27	0	1	0	0
4	0	25.6	0	1	0	21	29	1	1	1
5	11	34	1	1	1	0	22	0	1	0
6	38	0	1	0	0	26.4	0	1	0	0
7	0	32	0	1	0	11.6	14	1	1	0
8	20	0	1	0	0	20	30	1	1	1
9	25	23	1	1	1	16	23	1	1	0
10	0	0	0	0	0	9	0	1	0	0
Total cost			4041.6					3990.8		

solved with a linear mixed integer solver (Xpress). The following are the summary results for CMP and NCMP (Table 2).

For 10 periods, we have only 10 possible cyclic maintenance schedules knowing that a maintenance is operated at the beginning of the planning horizon $(z_1 = 1)$. The total cost of an optimal integrated production and cyclic maintenance plan is 4041.6 and is obtained for $Z = (1, 0, 0, 0, 1, 0, 0, 0, 1, 0)$. In other words, the optimal integrated production and cyclic maintenance plan suggests that a maintenance is performed for the machine every 4 periods. On the other hand, if we remove the periodicity constraint in CMP, the total cost of an optimal integrated production and non-cyclic maintenance planning is 3990.8, and it is obtained for the combination of $Z = (1, 0, 0, 1, 0, 0, 0, 1, 0, 0)$. The optimal total cost obtained for NCMP is then lower than that for CMP. In fact, the removal of the maintenance periodicity constraint reduces the total cost by 1.26 % in this case.

In order to validate the results of our modeling approach, we use the solution algorithm in Aghezzaf et al. (2007) where the integrated maintenance and production problem is solved by enumerating all possible maintenance schedules. Table 3 shows the results of implementing the latter algorithm for our example.

Unsurprisingly, the CMP by our modeling and Aghezzaf et al. (2007) both yield the same results indicated in Table 3. However, the optimal solution in Aghezzaf et al. (2007) is determined by varying the length of maintenance (k) as a parameter and solving 10 optimization problems whereas the cycle length is implicitly

Table 3 Optimal production and maintenance planning for each maintenance senario

Period	Available capacities									
	$k = 1$	2	3	4	5	6	7	8	9	10
1	50	50	50	50	50	50	50	50	50	50
2	50	40	40	40	40	40	40	40	40	40
3	50	50	32	32	32	32	32	32	32	32
4	50	40	50	25.6	25.6	25.6	25.6	25.6	25.6	25.6
5	50	50	40	50	20.5	20.5	20.5	20.5	20.5	20.5
6	50	40	32	40	50	16.4	16.4	16.4	16.4	16.4
7	50	50	50	32	40	50	13.1	13.1	13.1	13.1
8	50	40	40	25.6	32	40	50	10.5	10.5	10.5
9	50	50	32	50	25.6	32	40	50	8.4	8.4
10	50	40	50	40	20.5	25.6	32	40	50	6.7
Maintenance cost	500	250	200	150	100	100	100	100	100	50
Production cost	3800	3808	3846	3891.6	–	–	–	–	–	–
Total cost	4300	4058	4046	4041.6	–	–	–	–	–	–

– infeasible solution

incorporated into our proposed model as a decision variable. This implies that our optimization problem is required to be solved only once.

5 Computational experiments

In the previous sections, we have proposed novel MIP models for integrated lot sizing and maintenance scheduling. In this section, our goal is to evaluate the performance of the proposed algorithmic approach (use a MIP solver to solve the proposed models). The computational experiments are divided into two parts. In the first part, sensitivity analysis is performed for NCMP in order to study the effect of changes of different parameters on the quality of the solution found (measured by the duality gap obtained at the branch-and-cut procedure) and the number of nodes. Moreover, the impact of adding Wagner-Within (WW) constraint for different sets of parameters is considered. In the second part, we compare the impact of using WW constraint for the other CM and NCM problems.

5.1 NCMP

In order to validate these conclusions, we have randomly generated many instances by varying some important parameters of the problem. Six parameters are considered for the analysis: number of products P, number of periods in the planning horizon T, capacity tightness w, the exponential decrease rate of capacity α, the set up cost f, and the maintenance cost m. The number of product P is selected

Table 4 Computational results of the medium size data set for NCMP

Case	Problem						Node	Time (s)	Gap (%)
	T	P	α	w	f_{it}	m_t			
1	12	5	0.8	1.6	[300,500]	[300,500]	10641	23.78	<0.01
2	12	5	0.7	1.6	[300,500]	[300,500]	23414	99.52	<0.01
3	12	5	0.8	1.2	[300,500]	[300,500]	50897	327.98	<0.01
4	12	5	0.7	1.2	[300,500]	[300,500]	70796	557.95	<0.01
5	12	5	0.8	1.6	[5000,8000]	[300,500]	1991	3.61	<0.01
6	12	5	0.7	1.6	[5000,8000]	[300,500]	784	1.44	<0.01
7	12	5	0.8	1.2	[5000,8000]	[300,500]	16419	59.16	<0.01
8	12	5	0.7	1.2	[5000,8000]	[300,500]	13552	52.15	<0.01
9	12	5	0.8	1.6	[300,500]	[5000,8000]	18244	51.88	<0.01
10	12	5	0.7	1.6	[300,500]	[5000,8000]	8187	25.57	<0.01
11	12	5	0.8	1.2	[300,500]	[5000,8000]	78880	620.86	<0.01
12	12	5	0.7	1.2	[300,500]	[5000,8000]	34363	172.9	<0.01

from {5,10}, the number of periods T is selected from {12,24}. The demand for each product is assumed to be a random number in [0,40].

The factor w is introduced to control the capacity tightness. The nominal capacity is calculated according to $\sum_{i \in P, t \in T} d_{i,t}/T * w$, so that values of w slightly above 1 imply a tight production system. The cost structure is assumed to have the following properties: the set up cost for each product f_{it} is selected randomly from two different intervals [300,500] for low set up cost problem and [5000,8000] for high set up cost problems. The maintenance cost is selected randomly from [300,500] for low maintenance cost problems and [5000,8000] for high maintenance cost problems. The production cost is zero and holding costs are uniformly equal to 1.

For each combination of set up cost, maintenance cost and capacity tightness 5 test problems are generated. Each test problem is solved within a time limitation of 600 seconds to obtain the best feasible solution (Z_{BS}) and a lower bound (Z_{LB}) on the optimal of the problem. The duality gap is computed as follows:

$$Duality\ Gap = \frac{Z_{BS} - Z_{LB}}{Z_{LB}} * 100\,\% \qquad (29)$$

Xpress-MP 7.1 using default settings is used to solve the problems on a single processor with 4Gb of RAM and clocked at 2.5GHz. Tables 4 and 5 show the summary of the results obtained for model NCMP for different values of α and w.

The main message to be taken from Table 4 is that these medium-sized problems can all be solved to optimality in a few seconds to minutes. Moreover very high quality solutions are always available after a few seconds.

Other interesting aspects to notice are the following:

- when the ratio of set up cost to maintenance cost is very high the problem is easier to solve (e.g. case 1 vs. case 5 & 9),

Table 5 Computational results of the large size data set for NCMP

Case	Problem						Basic model		+WW ineq.	
	T	P	α	w	f_{it}	m_t	Node	Gap (%)	Node	Gap (%)
1	24	10	0.8	1.6	[300,500]	[300,500]	25274	14.71	25694	11.92
2	24	10	0.7	1.6	[300,500]	[300,500]	25694	14.47	31568	10.88
3	24	10	0.8	1.2	[300,500]	[300,500]	40513	14.26	43152	12.56
4	24	10	0.7	1.2	[300,500]	[300,500]	40572	14.53	39401	12.55
5	24	10	0.8	1.6	[5000,8000]	[300,500]	28198	9.99	29192	8.86
6	24	10	0.7	1.6	[5000,8000]	[300,500]	43012	10.89	47087	8.98
7	24	10	0.8	1.2	[5000,8000]	[300,500]	22425	16.89	20890	15.07
8	24	10	0.7	1.2	[5000,8000]	[300,500]	29783	17.08	28767	15.89
9	24	10	0.8	1.6	[300,500]	[5000,8000]	23478	9.73	24727	8.89
10	24	10	0.7	1.6	[300,500]	[5000,8000]	20307	8.13	17351	7.2
11	24	10	0.8	1.2	[300,500]	[5000,8000]	21598	4.94	20323	4.64
12	24	10	0.7	1.2	[300,500]	[5000,8000]	30675	4.97	28072	4.33

- for different setup and maintenance costs, a faster rate of decrease of capacity (smaller α) is accompanied with a reduction in running time(e.g. case 9 vs. case 10),
- when setup cost is as large as maintenance cost, a faster rate of capacity reduction makes the problem more difficult to solve(e.g. case 1 vs. case 2),
- tightly capacitated systems are much more difficult to solve.

To solve larger instances, it is necessary to improve the formulation of the problem. A well-known way to tighten the formulation of lot sizing problems is to add the so-called Wagner-Within (WW) inequalities to the original formulation (Pochet and Wolsey 2006):

$$I_{i,t-1} \geq \sum_{u=t}^{l} \left(d_{i,u} \left(1 - \sum_{v=t}^{u} (y_{i,v}) \right) \right); \qquad \forall t \in T \ , \ t \leq l \leq t + M \tag{30}$$

We now experiment with instances with 10 items and a planning horizon of 24 periods. The other parameters are similar as before. For each combination, 5 test problems are solved by setting 600 seconds as the maximum running time.

The results in Table 5 demonstrate that increasing number of periods and products have significant effects on the duality gap. According to Table 5, tighter capacity ($w = 1.2$ compared to $w = 1.6$) yields larger average gaps when set up cost is at least as large as maintenance cost. On the other hand, tighter capacity yields smaller average gaps when set up cost is less than maintenance cost. Moreover, decreasing the value of α seems not to have much impact on the average gap. The results indicate that using WW constraint as an additional constraint in all mathematical models yields a better value for the objective function and narrower duality gap.

Table 6 Computational results for comparing CMPs with NCMPs

Problem							CMP			CMP_WW (600 s)		
α	$\bar{\beta_i}$	γ	$\bar{p_k}$	w	f_{it}	m_t	Node	Obj. fun.	Gap (%)	Node	Obj. fun.	Gap (%)
0.8	0.31	–	–	1.2	[300,500]	[300,500]	17624	25629.2	11.81	23559	24594.5	7.75
0.8	–	0.35	–	1.2	[300,500]	[300,500]	55617	23945.4	11.03	59990	23397.3	8.84
0.9	–	–	0.62	1.6	[300,500]	[300,500]	50568	23886.1	12.02	54702	22683.1	6.55

Problem							NCMP			NCMP_WW (600 s)		
α	$\bar{\beta_i}$	γ	$\bar{p_k}$	w	f_{it}	m_t	Node	Obj. fun.	Gap (%)	Node	Obj. fun.	Gap (%)
0.8	0.31	–	–	1.2	[300,500]	[300,500]	55128	24857.4	8.88	65286	21337.6	5.41
0.8	–	0.35	–	1.2	[300,500]	[300,500]	52444	22817	16.18	56506	22072.9	12.55
0.9	–	–	0.62	1.6	[300,500]	[300,500]	55601	16958.8	14.61	55474	16525.6	12.19

5.2 Computational results for other CMPs with NCMPs

In this section, we experiment with the more complex capacity models of Sect. 3.2.2. Our goal is to see whether the same performance can be expected as for the more basic capacity models. Again, we attempt to solve large instances with 24 time periods and 10 different products with the time limitation of 600 s. The results can be found in Table 6. $\bar{\beta_i}$ corresponds to the rate of capacity decrease per unit of production, γ is the relative decrease of capacity per set up, and $\bar{p_k}$ shows the average of the percentage of the capacity available after k periods after the last maintenance. Each line of the table corresponds to the average of 5 randomly generated instances. We also run each instance twice: once where general non-cyclic maintenance is allowed, and a second time where cyclic maintenance is imposed. Also, we experiment the impact of adding WW inequalities to improve the formulation.

The results indicate that

- More complicated capacity models do not make the problem more difficult (the gaps are similar),
- Wagner-Whitin inequalities substantially improve the results, but even with the inequalities added, solving instances of that size to optimality is challenging,
- based on our experience with instances of smaller size, it is likely that the solutions obtained are close to optimum, even if the resulting gaps remain substantial (i.e. for small size instances, the optimum solution is usually found quickly, and the bulk of the computational time is used to improve the lower bound and prove optimality)
- when demand and costs are dynamic, using cyclic maintenance schedule can result in a total cost increase of up to 41 % compared to non-cyclic maintenance schedules.

6 Conclusion

Several models for the multi-product, multi-period joint production planning and (condition-based) maintenance problems that has applications in many different areas have been proposed in this paper. The motivation comes from a production system in lime industry, but the model can be used to represent many other aspects like the increase of defect products and the increase of downtime from corrective maintenance. We have described several models, mostly differing by the way capacity is impacted by production decisions. In the most general model, it is assumed that the capacity of the machine in each period is a function of the capacity of the previous period, lot size of the previous period and the presence of a setup in the previous period. A series of test problems were generated and solved with an off-the-shelf MIP solver. The computational results have demonstrated that the algorithms are capable of generating high-quality solutions for problems of size up to 24 time periods and 10 products, and solving to optimality instances with 12 time periods and 5 products. Moreover, the higher ratio of set up cost to maintenance cost and the smaller rate of decrease of capacity make the problem easier to solve. Since the maintenance actions are very costly in lime industry, the advantage of using non-cyclic maintenance schedules instead of cyclic ones is egregious. Specially, non-cyclic maintenance is more advantageous than cyclic maintenance when demand fluctuations are more serious. Our improvement of the formulation results in an optimal solution for the small size instances in a reasonable time. For large size instances, however, the duality gap is positive and could further decrease by resorting to other solution methods. An extension of the proposed model to a production system consisting of multiple machines can be investigated. Moreover, the model could be extended to include transportation planning.

References

Aghezzaf EH, Jamali MA, Ait-Kadi D (2007) An integrated production and preventive maintenance planning model. Eur J Oper Res 181(2):679–685

Aghezzaf EH, Najid N (2008) Integrated production planning and preventive maintenance in deteriorating production systems. Inf Sci 178(17):3382–3392

Aramon Bajestani M, Banjevic D, Beck C (2014) Integrated maintenance planning and production scheduling with Markovian deteriorating machine conditions. Int J Prod Res 52(24):7377–7400

Ashayeri J, Teelen A, Selen W (1996) A production and maintenance planning model for the process industry. Int J Prod Res 34(12):3311–3326

Ben-Daya M (2002) The economic production lot-sizing problem with imperfect production processes and imperfect maintenance. Int J Prod Econ 76(3):257–264

Berg M, Posner MJM, Zhao H (1994) Production inventory system with unreliable machines. Oper Res 42(1):111–118

Chakraborty T, Giri BC, Chaudhuri KS (2008) Production lot sizing with process deterioration and machine breakdown. Eur J Oper Res 185(2):606–618

Chakraborty T, Giri BC, Chaudhuri KS (2009) Production lot sizing with process deterioration and machine breakdown under inspection schedule. Omega 37(2):257–271

El Ferik S (2008) Economic Production lot-sizing for an unreliable machine under imperfect age-based maintenance policy. Eur J Oper Res 186(1):150–163

Fituhi MC, Nourelfath M (2012) Integrating noncyclical preventive maintenance scheduling and production planning for a single machine. Int J Prod Econ 136(2):344–351

Fituhi MC, Nourelfath M (2014) Integrating noncyclical preventive maintenance scheduling and production planning for multi-state systems. Reliab Eng Syst Saf 121:175–186

Garg A, Deshmukh SG (2006) Maintenance management: literature review and directions. J Qual Maint Eng 12(3):205–238

Grigorieva A, Van de Klundertb J, Spieksma FCR (2006) Modeling and solving the periodic maintenance problem. Eu J Oper Res 172(3):783–797

Groenevelt HA, Pintelon L, Seidmann A (1992a) Production lot sizing with machines breakdowns. Manag Sci 38(1):104–123

Groenevelt H, Pintelon L, Siedmann A (1992b) Production batching with machine breakdown and safety stocks. Oper Res 40(5):959–971

Hu J, Zhang L, Liang W (2012) Opportunistic predictive maintenance for complex multi-component systems based on DBN-HAZOP model. Process Saf Environ Prot 90(5):376–388

Iravani SMR, Duenyas I (2002) Integrated maintenance and production control of a deteriorating production system. IIE Trans 34(5):423–435

Kim CH, Hong Y, Kim SY (1997) An extended optimal lot sizing model with an unreliable machine. Prod Plan Control 8(6):577–585

Lee CY, Chen ZL (2002) Scheduling jobs and maintenance activities on parallel machines. Naval Res Logist 47(2):145–165
Lee HL, Rosenblatt MJ (1989) A production and maintenance planning model with restoration cost dependent on detection delay. IIE Trans 21(4):368–375
Levitin G, Lisnianski A (2000) Optimization of imperfect preventive maintenance for multi-state systems. Reliab Eng Saf Syst 67(2):193–203
Lin GC, Gong DC (2006) On a production-inventory system of deteriorating items subject to random machine breakdowns with a fixed repair time. Math Comput Model 43(7–8):920–932
Lin YH, Chen JM, Chen YC (2011) The impact of inspection errors, imperfect maintenance and minimal repairs on an imperfect production system. Math Comput Model 53:1680–1691
Lu Z, Zhang Y, Han X (2013) Integrating run-based preventive maintenance into the capacitated lot sizing problem with reliability constraint. Int J Prod Res 51(5):1379–1391
Miller AJ, Nemhauser GL, Savelsbergh MWP (2003) On the polyhedral structure of a multi-item production planning model with set up times. Math Program 94(2–3):375–405
Pochet Y, Wolsey L (2006) Production planning by mixed integer programming, Springer series in operations research and financial engineering, New York
Sortrakul N, Nachtmann CR, Cassady CR (2005) Genetic algorithms for integrated preventive maintenance planning and production scheduling for a single machine. Comput Ind 56(2):161–168
Wang CH, Meng FC (2009) Optimal lot size and offline inspection policy. Comput Math Appl 58(10):1921–1929
Xiang Y, Cassady CR, Jin T, Zhang CW (2014) Joint production and maintenance planning with machine deterioration and random yield. Int J Prod Res 52(6):1644–1657
Yao X, Fernandez EG, Fu MC, Marcus SI (2004) Optimal preventive maintenance scheduling in semiconductor manufacturing. IEEE Trans Semicond Manuf 17(3):345–356
Zhang Y, Lu Z, Xia T (2014) A dynamic method for the production lot sizing with machine failures. Int J Prod Res 52(8):2436–2447

Fahimeh Shamsaei has obtained her doctoral degree in Management and Economics in 2015 with the specialization in Supply Chain Management form Universite catholique de Louvain (UCL). She participated on ARC project (Management Resources in Supply Chain) at Center for Operations Research and Econometrics (CORE).

Mathieu Van Vyve is a professor in Operational Research at the Louvain School of Management. He is currently Director of the Center of Supply Chain Management (CESCM) and chairman of the department Operations and Information Systems (OIS). He is a member of the Center for Operations Research and Econometrics (CORE). His research interests are combinatorial optimization, Planning of production and transport, and the organization of the electricity markets.