European Journal of Operational Research 126 (2000) 422–435

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/dsw

Theory and Methodology

Optimal planning in large multi-site production networks

Christian H. Timpe, Josef Kallrath *

BASF-AG, ZOI/ZC, C13, D-67056 Ludwigshafen, Germany

Received 14 April 1998; accepted 4 May 1999

Abstract

This contribution describes a general mixed-integer linear programming model based on a time-indexed formulation covering the relevant features required for the complete supply chain management of a multi-site production network. While the actual application is taken from the chemical industry, the model provides a starting point for many applications in the chemical process industry, food or consumer goods industry. In many real world problems certain features just need to be eliminated from this general model in order to describe a current situation. The model combines aspects related to production, distribution and marketing and involves production sites (plants) and sales points. Besides standard features of lot sizing problems (raw materials, production, inventories, demands) further aspects, e.g., different time scales attached to production and distribution, the use of periods with different lengths, the modeling of batch and campaign production need to be considered. There are also new conceptual aspects in this paper, e.g., how to define the capacity of a multi-site, multi-product production network, or how to approach complex planning problems. We give a complete description of features ready for implementation, and the experience we have with the current implementation in our company. A long-term implication of this contribution might be that it will initiate further research efforts aiming to derive special cuts improving the formulation. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Decision support systems; Modeling; Production; Inventory; Transportation; Distribution; Integer programming; Branch and bound

1. Problem definition

1.1. Introduction

A problem frequently occurring in food or chemical process industry is the modeling of so-called *multi-purpose plants*, i.e. plants with machines or reactors that can be operated in different modes. In each mode it is possible to produce several products according to free or fixed recipes (joint production) leading to a general mode–product relation: in a certain mode several products are produced (with different maximal daily production capacity rates), and vice-versa, a product can be produced in different (but not all) modes. To switch between modes, a *changeover* is necessary, which results in a considerable loss of

* Corresponding author. Tel.: +49-621-60-78297; fax: +49-621-60-49463.

E-mail addresses: christian.timpe@basf-ag.de (C.H. Timpe), josef.kallrath@t-online.de (J. Kallrath).

PII: S0377-2217(99)00301-X

production time. Planning problems of this type, where the production of an item implies some discrete event, are called *lot sizing problems*. For a survey on lot sizing problems see Kuik et al. (1991).

Because of their high complexity, these problems are often seen as academic and not suitable for practical optimization. Instead, either simplifications (like linear relaxations) or simulation approaches are used, especially if there are several plants, tank inventories, and the production problem is just a part of the bigger problem of optimizing the whole supply chain with regard to a detailed description of the commercial environment, such as demands with different prices for different customers, availability of raw material or quality commitments.

We developed a rather general approach, a linear mixed integer model, that can handle all of the features mentioned and was successfully implemented into a production planning system at BASF. The system is able to produce near-optimal production plans in reasonable time on a regular basis. Examples are presented in Section 3.4.

1.2. Problem description

The problem of finding strong MIP formulations for modeling of changeovers has attracted several researchers in the past years, see for example Constantino (1996) or Wolsey (1997). For describing the general mode-product relation, we follow the formulation by Kallrath and Wilson (1997), which is shortly presented in Appendix A. We extend and generalize their formulation in order to model detailed description of commercial aspects: We now distinguish between production and marketing. The first is taking place in *production sites* or *plants,* the second in *sales points*. Because of the different accuracy of the available data, we use different time scales for plants and sales points, leading to an appropriately exact description of each real world aspect (see Fig. 1).

Plants and sales points and sales points among each other are connected by transport, so we can include the case that production sites and distribution depots are at different places, even in different countries. All types of transport are modeled by semi-continuous variables.

Production mode changes may take place at any time in a production period, but at most once. Following the classification in Belvaux and Wolsey (1999), lot sizing problems of this type are called *small bucket problems with two setups* (one before, one after mode change) *per period* (SB-2P). Our model is capable of mapping special features such as batch or campaign production. Batch production operates in integer multiples of batches where a batch is the smallest unit to be produced, e.g., 200 tons. Several batches following each other immediately establish a campaign. Some typical batch restrictions group batches into campaigns, or consider that only campaigns of a minimal size can be produced. Multi-level production or raw-material consumption can be modeled as well.

The inventory aspects are described in great detail considering fixed tanks, product-variable tanks and containers. For all products, the model keeps track of their origin to satisfy customers who have preferences for a specific production site. This is needed in the area of high performance chemicals, but also, for example, in food industry, where customers have preferences for products that come from certain areas or even preferences for certain types of packaging. Demands of different customers at the same sales points can be aggregated or described individually (e.g., with different prices) for some or all customers.

Finally, the model and its implementation supports several objective functions, the most important of which are: maximizing total sales, minimizing costs, and maximizing contribution margin. Conceptually, the maximizing sales scenario allows definition of the capacity of a multi-site, multi-product production network.

2. A mixed-integer linear programming model

We consider *production sites* (plants) located in different countries. All of them can manufacture the most important basic *products*, some can do additional rarely wanted products. To each plant or reactor belongs a set of *modes* (= status of plant), in which it can be operated. In each mode

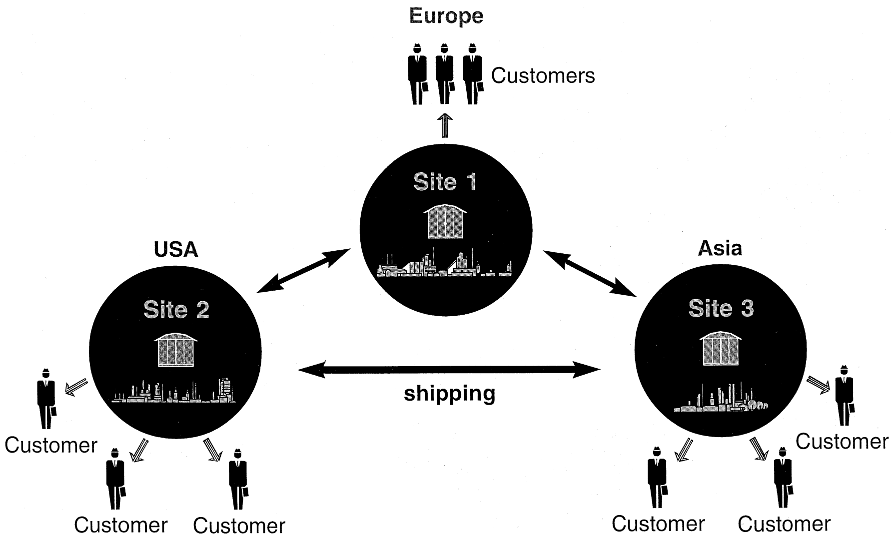

Fig. 1.

several products can be produced. A product may be produced in several modes. A plant either is in exactly one mode, or it is doing a changeover from one mode to another. Although the sites, sometimes referred to as *origins*, are similar with respect to the basic products, we cannot treat them as being quite identical. Some customers need to know exactly where their products come from, or they demand the product from a special site. To model this, we group our customers with their demands into *sales categories*. For each category, sales prices and allowed origins can be defined individually.

Because of the need for flexible and fast delivery, production and distribution networks will often have more *sales points* than sites. Each of the sales points is connected to one site by *transport* facilities, different sales points may be connected to the same site. The sales points among each other are also connected, and all of them have some (fixed or product-variable) *storage tanks* and *container* space for different products available.

For marketing aspects, we use a *commercial time scale*, whilst for production a *production time scale*. Usually marketing is less exact in forecasting than production, so we can treat each aspect with the appropriate precision and time resolution.

2.1. General framework

Throughout this model description the following set of indices will be used. Because of the two different resolutions of the time scale, we have two indices, k and t, for time, the index r is used in Section 2.2 to connect these two (Fig. 2):

$i \in \mathscr{I}$		production sites plants origins
$k \in \mathscr{K}_i$	$:= \{1,\ldots,K_i\}$	production periods
$m \in \mathscr{M}_i$		modes (site-dependent)
$p \in \mathscr{P}$		products
$r \in \mathscr{R}_{it}$	$:= \{1,\ldots,R_{it}\}$	production slices in commercial slice
$q \in \mathscr{Q}$		raw materials
$s \in \mathscr{S}$		sales points
$t \in \mathscr{T}$	$:= \{1,\ldots,T\}$	commercial periods
$c \in \mathscr{C}$		sales categories (customers)

2.2. Time discretization

The goal of the production planning system is to compute an optimal production plan (or one with guaranteed bounds) for a certain *planning horizon*. The starting point and length of the planning hori-

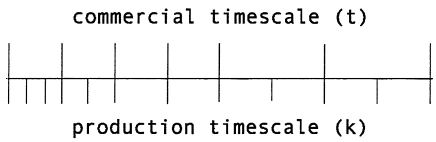

Fig. 2.

zon can be chosen by the user. Regarding delivery or sale, usually a commercial time scale of 12 periods (months) is chosen. Another possible scenario would be to cover a production plan with 16 periods: the first 12 with a length of about 30 days, and 4 additional ones with a length of 120 days. So the production plan would cover a total time of 2 years.

We divide the entire planning horizon into K_i production slices of size D_t^P/R_{it} days, where D_t^P is the length of the tth commercial period in days and R_{it} is the (integral) number of production slices embedded in that commercial period. A different number of slices can be embedded for each plant i.

In most cases, especially at the beginning of the planning horizon, the production schedule has a finer resolution than the commercial plans for sales and shipping, see the following example. This has to be done, because we can have no more than one mode change per production period. Using two time scales, the resolution is chosen adequately for the purpose of both production planners and marketing people. The function

$$k_s(i,t) := \begin{cases} 0 & \text{if } t = 1, \\ k_s(i,t-1) + R_{it-1} & \text{if } t > 1, \end{cases} \tag{2.1}$$

gives the number (minus one) of the production slice starting at the beginning of the commercial period t at site i, and, with r referring to a production slice embedded in the commercial time interval t, $k(i,t,r) := k_s(i,t) + r$ gives the absolute number $k(i,t,r)$ of that production slice within the production time scale referenced by t and r at plant i, and connects both time scales. If the production time scale and the commercial time scale are identical, we have $R_{it} = 1$, and $k := k(i,t,r) = t$.

2.3. Modes, mode-changes and production

The current model uses exactly the concepts of modes, mode-changing and coupling to production described in detail by Kallrath and Wilson (1997). For convenience of the reader, this approach based on state variables is briefly summarized in Appendix B.

Using this approach we end up with variables m_{imk}^D telling us the number of days (fractional days are allowed) in which the plant is in mode m during production period k:

$$m_{imk}^D \geqslant 0 \quad \text{number of days plant } i \text{ in period } k \text{ is in mode } m, \tag{2.2}$$

and the production variables

$$p_{ipk}^T \geqslant 0 \text{ tons of } p \text{ produced at plant } i \text{ in period } k. \tag{2.3}$$

These are connected with the help of the production rates data R_{imp}^P

$$p_{ipk}^T \leqslant \sum_{m \in \mathscr{M}_i | \exists R_{imp}^P} R_{imp}^P m_{imk}^D \quad \forall \{ipk\}. \tag{2.4}$$

The production rates R_{imp}^P are also used to indicate whether product p can be produced in mode m at site i at all. If $R_{imp}^P = 0$ this is not possible. Note that (2.4) is an inequality. Even if the plant is in a mode in which certain products could be produced there is no need to produce at full capacity.

The inclusion of campaigns of minimal size, discrete batches or the requirement that the plant has to stay in a certain mode for at least, say 5 days, over several production slices is possible. The basic idea is numbering campaigns consecutively and to give the production variables p_{ipk}^T a further index n to account for the campaign the product is produced in. The details are described in Kallrath (1999).

2.4. Coupling production to raw material flow

At site i the production of one ton of product p requires R_{iqp}^U tons of raw materials q. Note that a product may require several raw materials and that some raw materials are used by several products. Therefore the following equations are set up to couple the production products variables, p^T_{ipk}, to the variables u_{iqk} describing the usage of raw materials:

$$u_{iqk} = \sum_{p \in \mathscr{P} | R^U_{iqp} > 0} R^U_{iqp} p^T_{ipk} \quad \forall \{iqk\}. \tag{2.5}$$

The factors, R^U_{iqp}, may be interpreted as utilization factors. $1 \leqslant R^U_{iqp} \leqslant 1.1$ is a typical range. If $R^U_{iqp} = 1$, it means that the amount of raw material is completely converted into a product and no loss occurs. The coupling between the usage of raw material and the availability is established by the stock balance equations (2.26) and (2.27).

2.5. Modeling transport

Because products are manufactured in plants and sold in sales points, we have to use transport to distribute the products. Logically, this type of transport also connects the different time scales via the mass balance equations (2.22)–(2.24). Another type of transport connects the different sales points to each other. Transport is described by transport times T^I_{is} and $T^S_{s_s s}$ and further by certain minimal quantities which have to be observed. The model considers *transport variables*

$$\begin{array}{ll} \sigma_{ispk} & \text{tons/prod. period of } p \text{ shipped} \\ & \text{from site } i \text{ to sales pt. } s, \\ \rho_{ss_dpit} & \text{tons/comm. period of } p \\ & \text{shipped between sales pts. } (s \to s_d). \end{array} \tag{2.6}$$

These are semi-continuous, because transport does not make sense below some minimum amount T^{IM}_{is}. Semi-continuous variables, σ_{ispk} for instance, obey the definition

$$\sigma_{ispk} = 0 \vee T^{IM}_{is} \leqslant \sigma_{ispk} \leqslant M \quad \forall \{ispk\}. \tag{2.7}$$

It is not advantageous to define semi-continuous variables with the help of binary variables using the definition above, because the solver will usually *not* drive the binary variable to unity, if the corresponding continuous variable is above the lower limit. Instead, it will branch on the binary variable unnecessarily in the branch-and-bound process. State of the art MIP solvers give the user the opportunity to define semi-continuous variables directly.

As said, the first type of transport is used to distribute the products from a plant to the different sales points it is corresponding to. This is done by a distribution equation

$$p^T_{ipk} = \sum_{s \in \mathscr{S}} \sigma_{ispk} \quad \forall \{ipk\}. \tag{2.8}$$

The second type of transport is defined in the material balance equations (2.22)–(2.24).

2.6. Including demands

We now come to the commercial aspects of our model. Most important data are the demand data D_{spct}, telling the demand of product p at sales point s in commercial period t. To increase flexibility, we use the index c, to class articles into different *sales categories*. This allows us to define different prices for different customers for the same product, or to exclude certain origins of the product. Remember that certain customers impose severe restrictions on the properties of the product and therefore can be delivered only from a subset of the plants. If the sparse table D^{A1}_{spci} has no entry for a combination of sales point s, product p and sales category c, every origin (plant) i is possible. Otherwise only the given origins are allowed. Using this information, we define the *sales variables*

$$s^I_{spict} \mid \exists D_{spct} \wedge (\forall j \nexists D^{A1}_{spcj} \vee \exists D^{A1}_{spci}) \tag{2.9}$$

as the sales (in tons) of product p in category c from sales point s produced at i in period t. Sales variables are created only for allowed origins, and only if a demand exists.

Certainly no more must be sold than is required by the demand, i.e.,

$$\sum_{i \in \mathscr{I}} s^I_{spict} \leqslant D_{spct} \quad \forall \{spct\} \tag{2.10}$$

is a simple bound restricting the sales variables from above. A lower bound, forcing a minimum demand to be satisfied, can easily be constructed the analogous way.

Some customers can adapt to the specific quality of a product, but they want to do that only once. In these cases the solver is free to choose one of the allowed origins for the customer, but it has to be the same for the whole planning horizon. If a sales category c is specified in the sparse table D^{A2}_{spc}, shipment of p must always come from the same site. To model this last feature the sales variables s^l_{spict} are coupled with the binary variables

$$\delta^T_{spic} := \begin{cases} 1 & \text{if sale may occur} \\ 0 & \text{otherwise} \end{cases} \quad \forall\{spic\} \mid \exists D^{A2}_{spc}, \tag{2.11}$$

only generated if D^{A2}_{spc} exists. Of course there must be at least one possible origin for the demand, defined in table D^{A1}_{spci}. The binary variables are connected by the convexity constraint

$$\sum_{i\in\mathscr{I}} \delta^T_{spic} = 1 \quad \forall\{spc\}. \tag{2.12}$$

Let us now couple δ^T_{spic} and s^l_{spict}. The inequalities

$$s^l_{spict} \leqslant M\delta^T_{spic} \quad \forall\{ispct\} \tag{2.13}$$

ensure that no sale of product p in category c from origin i at sales point s is possible if $\delta^T_{spic} = 0$. If $\delta^T_{isp} = 1$ then the inequalities (2.13) produce the redundant bounds $s^l_{spict} \leqslant M$. M is a sufficiently large upper bound, the best choice is $M = D_{spct}$.

2.7. External purchase

Due to lack of capacity it may happen that not all of the demand can be covered by production. Therefore, it is very important to provide the option to consider external purchase of products. External purchase is characterized by the amount P^E_{spt} which can be purchased at most and the cost/ton C^E_{spt} for external purchase. It is very important that these costs reflect the real business process and not only are some artificial penalty costs. The external purchase variables

$$p^E_{spit} \tag{2.14}$$

describe external purchase (in tons) of product p in period t. We define an additional plant and let all products bought externally have this plant as their origin, reflecting the impossibility to sell product from the "wrong" plant to certain customers. This additional plant may also be used as origin for stock inventories whose origin has been forgotten or that are impure. If the problem is infeasible, we can allow products from all plants to be bought externally. These variables hint for the bottle neck in the current situation (not all demand can be satisfied) and may provide a platform for discussion about what to do with the current demands. Maybe a customer will accept delivery in a later period. The variables p^E_{spit} are bounded according to

$$p^E_{spit} \leqslant P^E_{spt} \quad \forall\{spit\}. \tag{2.15}$$

2.8. Product inventories

The model contains stocking of products in tanks and containers. They are associated to sales points and contain a specific product from a specific plant. At each site there is a set of $\mathscr{N}^T_s$ tanks, which can store any product. Concerning containers we only look at the sum of inventoried product from the same origin. Tanks have upper capacity limits S^T_{spn}, whilst the use of containers is only constrained by the physical space S^C_s available to keep them. Note that the tank capacity limit depends on the product: Some tanks may not be able to store certain products, and, because often products are computed as being pure although their concentration is only, say, 40%, tank capacities may be different for different products. We define the non-negative continuous *stock variables*

$$i^P_{spitn} \geqslant 0 \quad \forall\{spit\} \quad \forall n \in \mathscr{N}^T_s, \tag{2.16}$$

total stock (tons) in tank n at s of product p produced at i at the end of period t, and

$$i^C_{spit} \geqslant 0 \quad \forall\{spit\} \tag{2.17}$$

total stock (tons) in containers of p produced at i at s at the end of period t. To keep track of which product is in a tank, we define some binary variables

$$\lambda_{spitn} := \begin{cases} 1 & \text{if tank } n \text{ at } s \text{ is filled with } p \quad \forall\{spit\}, \\ 0 & \text{otherwise} \quad \forall n \in \mathscr{N}_s^T. \end{cases} \tag{2.18}$$

We ensure that the tanks contain only one product at a time by coupling the binary variables λ_{spitn} to the inventory variables i^P_{spitn} and adding the convexity constraint (2.20):

$$i^P_{spitn} \leqslant S^C_{spn} \lambda_{spitn} \quad \forall\{spit\} \quad \forall n \in \mathscr{N}_s^{Ti}. \tag{2.19}$$

This equation also serves as the capacity constraints for the tanks. Similar to (2.12) we need a constraint

$$\sum_{p\in\mathscr{P}} \sum_{i\in\mathscr{I}} \lambda_{spitn} \leqslant 1 \quad \forall\{st\} \quad \forall n \in \mathscr{N}_s^T, \tag{2.20}$$

guaranteeing that at most one of the binary variables is unity. If a tank is empty at the end of a period, the sum on the left-hand side of (2.20) is zero and the inequality is inactive.

The sum over all products stored in containers at a given site is constrained by

$$\sum_{p\in\mathscr{P}} \sum_{i\in\mathscr{I}} i^C_{spit} \leqslant S^C_s \quad \forall\{st\}. \tag{2.21}$$

2.9. Balance equations

The different parts of the model described so far are connected by material balance equations. They also connect the commercial time periods. Precisely, the balance equations connect the ends of the periods by looking at the sum of inflow and outflow. This can be interpreted in a way, that the model assumes incoming and outgoing flows in a given period to take place in such an order, that the tanks (and containers) neither overflow nor run below zero. The inflow f^I_{spit} consists in transport from plants and other sales points and external purchase:

$$\begin{aligned} f^I_{spit} = &\sum_{s_s\in\mathscr{S}|s_s\neq s} T^{SS}_{s_s spit} + T^{IS}_{ispt} + \sum_{r=1|t>T^I_{is}}^{R(i,t)} \sigma_{ispk(i,t-T^I_{is},)} \\ &+ \sum_{s_s\in\mathscr{S}|t>T^S_{s_s s}} \rho_{s_s spit-T^S_{s_s s}} + p^E_{spit} \quad \forall\{spit\}, \end{aligned} \tag{2.22}$$

where T^I_{is} the time needed to ship products from site i to sales point s and $T^S_{s_s s}$ the time for shipping from sales point s_s to sales point s. Note that because T^I_{is} and $T^S_{s_s s}$ are used in the index it is required that they are integral. If the commercial time scale is shorter than the time needed for transport, we may have transports on their ways at the time the model is run, these have to be input as T^{IS}_{ispt} and $T^{SS}_{s_s spit}$. Due to the non-zero transport times it may happen that certain containers shipped to a destination will not arrive within the planning horizon. Currently this problem is handled by the following rule: no transport is allowed if the product would not arrive at the destination within the planning horizon.

The outflow f^O_{spit} consists in transport to other sales points and sales to customers:

$$\begin{aligned} f^O_{spit} = &\sum_{s_d\in\mathscr{S}|s_d\neq s\wedge t+T^S_{ss_d}\leqslant T} \rho_{ss_d pit} \\ &+ \sum_{c\in\mathscr{C}} s^l_{spict} \quad \forall\{spit\}. \end{aligned} \tag{2.23}$$

The balance equation reads

$$\begin{aligned} &\sum_{n\in\mathscr{N}_s^T} i^P_{spitn} + i^C_{spit} \\ &\quad = \sum_{n\in\mathscr{N}_s^T} i^P_{spit-1n} + i^C_{spit-1} + f^I_{spit} - f^O_{spit} \quad \forall\{spit\}. \end{aligned} \tag{2.24}$$

The sum on the left-hand side of the equation is the total storage of product p from plant i in any available tank. The sum over all tanks on both sides of the equation implies that stock can be exchanged between compatible tank at zero cost. (For $t = 1$ we replace the first term one the right-hand side $\sum_{n\in\mathscr{N}_s^T} i^P_{spit-1n}$ by $\sum_{n\in\mathscr{N}_s^T} S^{PS}_{spin}$, the sum of the initial product stocks.)

It is possible to define a strategic minimal stock amount S^M_{spi} for products stored at any sales point:

$$\sum_{n \in \mathcal{N}_s^T} i^P_{spitn} + i^C_{spit} \geqslant S^M_{spi} \quad \forall \{spit\} \mid \exists S^M_{spi}. \tag{2.25}$$

2.10. Raw material inventories and balance equations

Raw material inventory levels s^R_{iqk} should always be above the safety stock and should never exceed the capacity limits. Raw material inventories are supplied by incoming transport at certain periods described by the table R^S_{iqk}. The raw material usage is governed by the equations

$$s^R_{iq1} = R^S_{iq1} - u^R_{iq1} \quad \forall \{iq\} \tag{2.26}$$

and

$$s^R_{iqk} = s^R_{iqk-1} + R^S_{iqk} - u^R_{iqk} \quad \forall \{iq\}, \quad k > 1. \tag{2.27}$$

Note that R^S_{iq1} represents the sum of initial raw material stock and what is supplied in the first period.

The raw material inventory levels are enforced by the bounds

$$S^{RM}_{iqk} \leqslant s^R_{iqk} \leqslant S^{RC}_{iqk} \quad \forall \{iqk\}. \tag{2.28}$$

2.11. Objective functions

The model can be used to answer a variety of questions and therefore covers several objective functions. The most important from the economical point of view maximizes the contribution margin. It includes the yield computed on the bases of production and the associated sales prices and the cost. Another one neglects the yields and just asks for the costs to satisfy all demands. If this is not possible, i.e., capacity is not sufficient, we may want to maximize total production without regard to costs or yield. Other possible scenarios would be to minimize changeovers, minimize tank stock at the end of the year, minimize transport, etc.

The yield is computed from the specific sales prices S^P_{spct}:

$$y = \sum_{s \in \mathcal{S}} \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{I}} S^P_{spct} s^l_{spict}. \tag{2.29}$$

There are a whole lot of costs that can be incorporated: variable production costs

$$c^V := \sum_{i \in \mathcal{I}} \sum_{p \in \mathcal{P}} \sum_{k=1}^{K_i} C^V_{ip} p^T_{ipk}, \tag{2.30}$$

mode changing costs

$$c^M := \sum_{i \in \mathcal{I}} \sum_{k=1}^{K_i} \sum_{m_1 \in \mathcal{M}_i} \sum_{\substack{m_2 \in \mathcal{M}_i \\ m_2 \neq m_1}} C^C_{im_1m_2} \xi_{ikm_1m_2}, \tag{2.31}$$

costs for transport between sales points and sites

$$c^{TI} := \sum_{i \in \mathcal{I}} \sum_{s \in \mathcal{S}} \sum_{p \in \mathcal{P}} \sum_{k=1}^{K_i} C^{TI}_{is} T^{IM}_{is} \sigma_{ispk}, \tag{2.32}$$

costs for transport between sales points and other sales points

$$c^{TS} := \sum_{s \in \mathcal{S}} \sum_{s_d \in \mathcal{S}} \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} C^{TS}_{ss_d} T^{SM}_{ss_d} \rho_{ss_dpit}, \tag{2.33}$$

and finally, inventory costs

$$c^{SP} := \sum_{s \in \mathcal{S}} \sum_{p \in \mathcal{P}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} C^{SP}_{sp} \left(i^C_{spit} + \sum_{n \in \mathcal{N}_s^T} i^P_{spitn} \right). \tag{2.34}$$

The objective function for the "maximize contribution scenario"

$$\max \ Z_1, \ Z_1 = y - \left(c^V + c^M + c^{TI} + c^{TS} + c^{SP}\right). \tag{2.35}$$

As is discussed in Kallrath and Wilson (1997) it is advantageous to formulate such block objective functions by replacing the block terms by their explicit equivalents.

In the case of the "minimize cost scenario" the objective function is simply

$$\max \ Z_2, \ Z_2 = -\left(c^V + c^M + c^{TI} + c^{TS} + c^{SP}\right). \tag{2.36}$$

To avoid infeasibilities due to lack of production capacity sufficient external purchase of products should be made available.

The motivation for "maximize total sales scenario" is to handle cases in which the capacity is not sufficient to satisfy all demands but to produce as much as possible with the available production capacities. If we would neglect the demands completely we would get a trivial solution: the plant systems produces the products according to highest production capacities and avoids mode changes completely. That is not what we want. Including the yield and cost structure into the problem would in fact lead to the "maximize contribution scenario" and is not compatible with what we associate with "maximum total sales" because some production is neglected because it is not attractive from the economic point of view. A consistent approach is to neglect costs (nearly) completely, and to set yields for all demands to an equal value, say, 1$, or any realistic average yield. So we maximize total sales by trying to satisfy as much of the demand as possible. Thus the objective function is simply

$$\max \sum_{s\in\mathscr{S}} \sum_{p\in\mathscr{P}} \sum_{t\in\mathscr{T}} \sum_{i\in\mathscr{I}} \left(\sum_{c\in\mathscr{C}} s^{l}_{spict} - 0.9 p^{E}_{spit} \right). \tag{2.37}$$

The only cost that has to be considered is costs for external purchase, otherwise the model will just satisfy demands by externally purchased products it gets for free. The consequence of the simplicity of the objective function is that the problem is highly dually degenerate. Therefore it produces solutions which appear as "strange", with lots of transport, for example. In case the initial and final stocks are the same maximizing sales is equivalent to maximizing production. So, this scenario can tell us the total production capacity of the whole production network with respect to a given demand pattern.

3. Implementation, optimization techniques, results

3.1. Real world aspects

The following numbers should give an impression of the size of the problem: The production network consists of four plants (sites) in three different regions with their own sales point and one additional tank farm. Each site can produce products in five different modes. Only one raw material is considered. The sales points have a fixed tank for every product and between one and six variable tanks plus container storage space. Over two hundred demands (combination of product, origin, sales category and time period) have to be fulfilled. A typical plan consists out of 12 commercial time periods, which are divided into 12 to 36 production periods, depending on the site.

3.2. The database and user-interface

The aim of our modeling is to make regular production planning possible on a PC. In order for the system to be accepted, easiness of use is its most important feature. We choose an `MS-ACCESS` database to store all the necessary data, and a friendly user-interface was also programmed in `MS-ACCESS`. The interface invokes the model generator and solver, and interprets the output after optimization. This approach allows the developer to distinguish completely between the data and the structure of the model. Communication between user-interface and solver is exclusively by `ASCII` files. This proved to be superior in speed when compared with the `ODBC` interface in `MS-ACCESS`. The interface includes a number of switches to allow using the model for different purposes. For instance, availability of raw material is usually not considered in the first production planning round, in order to determine the actual demand of raw material. When orders have been affirmed (or rejected), an optimal plan is computed with the available amounts. Or, initially, one probably would switch off the batch and campaign constraints across periods. However, in a fine tuning phase one might add this feature.

3.3. Mathematical aspects

The model lead to a mixed-integer linear programming problem (MILP) with the following problem statistics:

```
 7643      rows
15137      structural columns
49613      matrix elements
406        rhs elements
7869       bound elements
0          general integer variables
720        binary variables
7036       semi-continuous variables
14288      directives
density is.042884 percent
```

A primal simplex algorithm is used to compute the LP-relaxation; the solution is stored keeping the basis of the scenario. For the "maximize contribution" or "minimize cost" scenarios we use branch and bound with a dual simplex algorithm while the "maximize sales" scenario works much better with the primal algorithm because of its dual degeneracy.

Because of the large number of binary and semi-continuous variables in the model, we have to support B&B with directives defined for discrete variables. In the "maximize contribution margin" and "minimize sales" scenario the variables δ^T_{spic}, which assure that certain customers always get the product from the same site, should be given high priority. Next important are the plant state variables δ_{imk} described in (A.1). The solver should always try to bring them down first. Later come the binary tank variables λ_{spitn}. Also, the "down" branch should be evaluated first. If not production, but storage capacity is the main bottleneck of the system, the plant status and the tank variables may be exchanged in importance. Last come the semi-continuous transport variables. In their case the "up" branch should be the first to be explored.

3.4. Implementation and results

For a typical reference scenario (S_1) covering 12–36 production time periods we have derived production plans maximizing total sales. Using Dash's MILP-solver XPRESS-MP 10.05 (Ashford and Daniel, 1987, 1991), we got the results given in Table 1 (including the number of continuous n_c, binary b and semi-continuous variables sc, number of constraints c, integer solution number IP, number of nodes n_n, running time τ on a 166 MHz Pentium Laptop, best upper bound z^U, best lower bound z^L and integrality gap $\Delta := 100((z^U - z^L)/z^L)$).

The first feasible integer solution is usually found within 10 minutes using XPRESS-MP on a Pentium, after exploring about 500 nodes in the B&B tree. Usually, it is accepted as the solution. This heuristic is justified because our trials indicate that its associated contribution margin deviates by only a few percent from that of the continuous problem (well within the error associated with the input data) and because it eliminates the need for the time consuming complete search for the absolute optimal solution via the B&B algorithm.

The solution derived in the "maximize sales" scenario is used to support the solution process in the "maximize contribution margin" or "minimize cost" scenario. This approach is supported by two additional output data produced in the "maximize sales" scenario: the total production

$$Q_1 := \sum_{i\in\mathscr{I}} \sum_{p\in\mathscr{P}} \sum_{k=1}^{K_i} p^T_{ipk} \tag{3.38}$$

and the total unmet demands

$$Q_2 := D_{spct} - \sum_{s\in\mathscr{S}} \sum_{p\in\mathscr{P}} \sum_{i\in\mathscr{I}} \sum_{c\in\mathscr{C}} \sum_{t\in\mathscr{T}} s^l_{spict}. \tag{3.39}$$

These numbers can be used as additional constraints in the other scenarios.

In addition, the LP-relaxation of the "maximize sales" scenario is used as the initial basis to start the Simplex algorithm in the other scenarios.

Table 1

	n_c	b	sc	c	IP	n_n	τ	z^U	z^L	Δ
S_1	12397	2973	1608	8441	1	440	8^m	54479	53444	1.9
S_1					2	960	$+6^m$	54479	53724	1.4
S_1					3	1721	$+8^m$	54479	53927	1.0

3.5. Commercial impact and results

The optimal or near-optimal production plans satisfy the demand up to about one percent. Short-term wishes for extra deliveries can now be fast and reliably affirmed or rejected. Money could be saved by reducing tank storage; on the other hand, the need for certain tank stock levels can now be proven. The system also hints for bottlenecks in the plant and delivery system, long-term consequences may be the extension of the so-identified problem areas.

The model reflects the business process to a sufficient degree of reality. It turns out to be a useful tool in complex production network. The production plans are plausible and not counter-intuitive.

4. Conclusions and further research

Despite the success let us be aware that the model is less than the reality and certainly has some limitations and assumptions it is based on:

1. Only one mode change is possible per production period. This seems not to be a very serious restriction since the length of a production period can be chosen. But, since the length of the production period must be larger then the time needed to perform a mode change in that period, a problem can arise if changeover times on a plant vary strongly.
2. Production rates do not depend on time. This can obviously be extended if necessary.
3. Stock balance equations only hold for the end of each commercial period. For a more detailed accounting of product stock, tanks could be described as corresponding to production sites instead of to sales points. In that case it would also make sense to describe transport from sites to sales points in units of the production period. But if storage capacity is a very severe restriction in reality, this might still be not accurate enough. In either case transport time can only take integral values.

Apart from these assumptions, the model provides a reasonable starting point for many applications in the chemical process industry and it is open to further generalizations. At present, in some cases it is possible to prove optimality, but in most cases it provides safe bounds. The approach described in this paper will hopefully initiate further research efforts aiming to derive special cuts improving the formulation.

Appendix A. The basic states variables and uniqueness of plant status

The worldwide production network and its current state is characterized by the *state variables* $\delta_{imk} \in \{0, 1\}$,

$$\delta_{imk} := \begin{cases} 1 & \text{if plant } i \text{ is in mode} \quad \forall i \in \mathscr{I},\ \forall m \in \mathscr{M}_i, \\ & m \text{ at the end of period } k \\ 0 & \text{otherwise} \quad \forall k \in \mathscr{K}. \end{cases} \tag{A.1}$$

These variables can be used to guarantee that at the end of time interval k the plant at site i is in a unique mode, e.g., by equations of the form

$$\sum_{m \in \mathscr{M}_i} \delta_{imk} = 1 \quad \forall i \in \mathscr{I} \quad \forall k \in \mathscr{K} \tag{A.2}$$

However, as will become obvious below it is not necessary to add the equations explicitly to the systems of constraints.

Note that some initial data Δ_{im0} have to be provided to define the known status of plant i before we start planning. Of course, these initial data must satisfy the condition

$$\sum_{m \in \mathscr{M}_i} \Delta_{im} = 1 \quad \forall i. \tag{A.3}$$

Sometimes the state of all plants may be given in advance, and one may want to fix the states of all plants to the states known from another optimization run. Therefore, the model provides the option to use the states of all plants according to the bounds $\delta_{imk} = \Delta^F_{imk}$, where Δ^F_{imk} give the state of all plants during the whole planning horizon. All other variables can then be optimized with respect to the fixed modes.

If our states variables take the values $\delta_{im_1k-1} = \delta_{im_2k} = 1$ we have a mode change from mode m_1 to m_2 in time interval k.

The continuity of modes or mode changes is tracked by the binary variables

$$\xi_{ikm_1m_2} = \begin{cases} 1 & \text{if } \delta_{im_1k-1} = \delta_{im_2k} = 1 \quad \forall i \in \mathscr{I}, \\ & \qquad\qquad\qquad\qquad m \in \mathscr{M}_i, \\ 0 & \text{otherwise} \qquad\qquad\quad\; \forall k \in \mathscr{K}. \end{cases} \tag{A.4}$$

This variable is unity if at the end of period $k-1$ the plant is in mode m_1 and at the end of period k it is in mode m_2. $\xi_{ikm_1m_2}$ is a variable not only describing whether a changeover occurs or not but it tells us whether production continues. If the plant is in mode m both at the end of period $k-1$ and k then we have $\xi_{ikmm} = 1$.

The state variables and the mode changes variables will now be coupled by some additional binary variables:

$$\alpha_{imk} := \begin{cases} 1 & \text{if plant } i \text{ is in mode } m \text{ for some time in period } k, \\ 0 & \text{otherwise,} \end{cases} \tag{A.5}$$

$$\beta_{imk} := \begin{cases} 1 & \text{if mode } m \text{ is started at site } i \text{ in period } k, \\ 0 & \text{otherwise,} \end{cases} \tag{A.6}$$

and finally

$$\gamma_{imk} := \begin{cases} 1 & \text{if mode } m \text{ is terminated at site } i \text{ in period } k, \\ 0 & \text{otherwise.} \end{cases} \tag{A.7}$$

These binary variables are related to others by the constraints

$$\beta_{imk} = \sum_{m_1 \neq m} \xi_{ikm_1m} \quad \forall i, \; m \in \mathscr{M}_i \; \forall k, \tag{A.8}$$

and

$$\gamma_{imk} = \sum_{m_1 \neq m} \xi_{ikmm_1} \quad \forall i, \; m \in \mathscr{M}_i \; \forall k. \tag{A.9}$$

To express whether mode m is used at all in period k we have

$$\alpha_{imk} = \delta_{imk-1} + \delta_{imk} - \xi_{ikmm} \quad \forall i, \; m \in \mathscr{M}_i, \quad k = 2, \ldots, T, \tag{A.10}$$

and

$$\alpha_{im1} = \Delta_{im0} + \delta_{im1} - \xi_{i1mm} \quad \forall i, \; m \in \mathscr{M}_i. \tag{A.11}$$

Note that is not necessary to declare ξ, β and γ as binary variables if we have declared δ and α as binary variables.

Finally we have the following block of constraints:

$$\gamma_{imk} = \delta_{imk-1} - \xi_{ikmm} \quad \forall i, \; m \in \mathscr{M}_i, \quad k = 2, \ldots, T, \tag{A.12}$$

or

$$\gamma_{im1} = \Delta_{im} - \xi_{i1mm} \quad \forall i, \; \forall m \in \mathscr{M}_i, \tag{A.13}$$

and

$$\beta_{imk} = \delta_{imk} - \xi_{ikmm} \quad \forall i, \; \forall m \in \mathscr{M}_i, \quad k = 2, \ldots, T. \tag{A.14}$$

The constraints above complete our description of mode changes.

Now we can also see why (A.2) is not required any longer. This follows from (A.3) and inspection of (A.11)

$$\sum_{m \in \mathscr{M}_i} \alpha_{im1} = 1 + \sum_{m \in \mathscr{M}_i} \delta_{im1} - \sum_{m \in \mathscr{M}_i} \xi_{i1mm} \quad \forall i, \; \forall m \in \mathscr{M}_i. \tag{A.15}$$

If in the first period a mode change takes place then $\sum_{m \in \mathscr{M}_i} \xi_{i1mm} = 0$ but $\sum_{m \in \mathscr{M}_i} \alpha_{im1} = 2$ since production is possible in exactly two modes. Thus we have $\sum_{m \in \mathscr{M}_i} \delta_{im1} = 1$. If no mode change occurred then production was possible in only one mode which gives $\sum_{m \in \mathscr{M}_i} \alpha_{im1} = \sum_{m \in \mathscr{M}_i} \xi_{i1mm} = 1$ which again leads to $\sum_{m \in \mathscr{M}_i} \delta_{im1} = 1$.

The total number of mode changes during the entire planning horizon at site i can be forced upper-bounded to M_i^C by

$$\sum_{k=1}^{K_i} \sum_{m_1 \in \mathcal{M}_i} \sum_{\substack{m_2 \in \mathcal{M}_i \\ m_2 \neq m_1}} \xi ikm_1 m_2 \leqslant M_i^C \quad \forall i. \tag{A.16}$$

Appendix B. Modeling the production requirements

The most important data for modeling production are H_{ik}, the number of days available for production and mode-change in production period k, and R_{imp}^P, the production rate (and mode–product relation) at plant i. Notice that H_{ik} depends on k which gives us the opportunity to model temporary shutdowns (maintenance, test runs, etc.) The production rates R_{imp}^P will also be used to indicate whether product p can be produced in mode m at site i at all. If $R_{imp}^P = 0$ this is not possible. It is possible that a certain product can be produced in several modes.

From H_{ik} we compute $H_{im_1m_2k}^R$, the number of days left for production if a mode change from m_1 to m_2 occurs, by $H_{im_1m_2k}^R = \max\{0, H_{ik} - M_{im_1m_2}^T\}$, with the mode-change duration $M_{im_1m_2}^T$. Note that we do not require H_{ik} or $H_{im_1m_2k}^R$ to be an integer quantity.

With this data we connect the state variables to the production variables:

$$p_{ipk}^T \geqslant 0 \text{ tons of } p \text{ produced at plant } i \text{ in period } k. \tag{B.1}$$

For that we introduce a variable telling us the number of days (fractional days are allowed) in which the plant is in mode m during production period k:

$$m_{imk}^D \geqslant 0 \text{ number of days plant } i \text{ in period } k \text{ is in mode } m. \tag{B.2}$$

We first note that

$$m_{imk} < H_{immk}^R \alpha_{imk} \quad \forall \{imk\} \tag{B.3}$$

is a valid upper bound. Remember, that α_{imk} carries the information whether mode m is used in period k or not. If the plant is never in mode m during period k then $\alpha_{imk} = 0$ and the plant indeed spends zero days in mode m. Otherwise, the inequality reduces to $m_{imk} < H_{ik}$. Fractional values of α_{imk} during the LP relaxation or within the tree reduce the time the plant can be in mode m leading to smaller amounts of the products being produced in this mode. This is one aspect of tightening the relaxation.

Next, we want to compute the available capacities subject to mode-changes. These constraints read

$$m_{imk} \leqslant \sum_{m \in \mathcal{M}_i} H_{im_2mk}^R \xi_{ikm_2m} + \sum_{\substack{m_2 \in \mathcal{M}_i \\ m_2 \neq m}} H_{imm_2k}^R \ \xi_{ikmm_2} \quad \forall i,\ m \in \mathcal{M}_i,\ \forall k, \tag{B.4}$$

and

$$\sum_{m \in \mathcal{M}_i | H_{ik} \neq 0} m_{imk} \leqslant H_{ik} - \sum_{m_1 \in \mathcal{M}_i} \sum_{\substack{m_2 \in \mathcal{M}_i \\ m_2 \neq m}} \Delta_{im_1m_2} \xi_{ikm_1m_2} \quad \forall i,\ \forall k. \tag{B.5}$$

Constraint (B.4) defines the upper limit on the number of days of production of product p at site i in period k as a function of the mode-change variables. Days for mode m become available in period k only if either the site status switches *to* mode m in period k (first term in the right-hand side of (B.4)) or the site status switches *from* mode m in period k (i.e., has status p at the end of period $k-1$). When the mode of the site is m at the end of period $k-1$ and k (i.e., $\xi_{ikmm} = 1$ and no change-over occurs in period k), the value of the available capacity H_{ikmm}^R is counted once in the first terms. In any integer solution at most two modes have a positive upper bound on available days in each period.

Constraint (B.5) defines the global available capacity in period k to be equal to the number of days available minus the number of days used for mode-change. The second term can only be applied if $\Delta_{im_1m_2} \leqslant H_{ik}$. Otherwise, we should fix $\xi_{ikm_1m_2} = 0$ because there is not enough capacity in period k to perform the mode-change from mode m_1 to mode m_2.

References

Ashford, R.W., Daniel, R.C., 1987. LP-MODEL XPRESS-LP's model builder. Institute of Mathematics and its Application Journal of Mathematics in Management 1, 163–176.

Ashford, R.W., Daniel, R.C., 1991. Practical aspects of mathematical programming. In: Munford, A.G., Bailey, T.C. (Eds.), Operational Research Tutorial Papers. Operational Research Society, Birmingham, pp. 105–122.

Belvaux, G., Wolsey, L.A., 1999. Lot-sizing problems: Modelling issues and a specialised branch-and-cut system BC-PROD. CORE Discussion Paper 9849 09/1998, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Constantino, M., 1996. A cutting plane approach to capacitated lot-sizing with start-up costs. Mathematical Programming 75, 353–376.

Kallrath, J., 1999. Nonlinear Mixed Integer Applications. In: OR in Industry. Macmillan, Stockton, pp. 60–97.

Kallrath, J., Wilson, J.M., 1997. Business Optimisation Using Mathematical Programming. Macmillan, UK.

Kuik, R., Solomon, M., van Wassenhove, L.N., 1991. Batching decisions: structure and models. European Journal of Operational Research 75, 243–263.

Wolsey, L., 1997. MIP modelling of changeovers in production planning and scheduling problems. European Journal of Operational Research 99, 154–165.