Integrated production and distribution planning for perishable food products

Poorya Farahani · M. Grunow · H.-O. Günther

Abstract Certain types of food, such as catering foods, decay very rapidly. This paper investigates how the quality of such foods can be improved by shortening the time interval between production and delivery. To this end, we develop an approach that integrates short-term production and distribution planning in an iterative scheme. Further, an aggregation scheme is developed as the interface between the production scheduling and distribution problem. The production scheduling problem is solved through an MILP modeling approach which is based on a block planning formulation. Our implementation shows promising results, elaborated in a numerical investigation, which recollects the real settings of a catering company located in Denmark.

Keywords Production scheduling · Block planning · Vehicle routing · Catering industry · Food quality

1 Introduction

A commodity is called perishable if its quality or quantity is subject to deterioration. A group of perishable products comprises goods in which product

P. Farahani (✉) · M. Grunow
Department of Business Administration-Production and Supply Chain Management, Technical University of Munich, Munich, Germany
e-mail: poorya.farahani@tum.de

M. Grunow
e-mail: martin.grunow@tum.de

H.-O. Günther
Department of Production Management, Technical University of Berlin, Berlin, Germany
e-mail: hans-otto.guenther@tu-berlin.de

quality is subject to a continuous change after the production stage. Fresh foods, bakery products and blood products are some well-known examples of this category. Fresh foods create unique supply chain challenges. Foods such as fruits, vegetables, and meat need to be nurtured to ensure that they remain in good quality. This includes monitoring the time that this delicate cargo spends in stock and during transport in the supply chain—the most critical time for all fresh foods. It is essential to deliver such products right after production to limit quality degradation. Therefore, a tight coordination between the production and distribution systems is necessary which can be realized by integrating the planning of these two stages of the supply chain.

Motivated by a case study from a food caterer in Copenhagen, we develop mathematical models and a heuristic solution method to investigate the effect of integrating production scheduling and distribution decisions on the total costs and the quality of the delivered products. The distribution problem involves designing vehicle routes and schedules for picking up prepared foods and delivering them to customers. On the production site, food production orders must be scheduled on available production lines. Moreover, the setup cost for producing each food type strongly depends on the product type produced in the previous production run, i.e. the setup costs are sequence-dependent.

In such a supply chain, product quality is determined not only by the production processes, but also through the coordination of the production and distribution decisions. To ease this coordination, food preparation sites are usually directly connected to customers by a fleet of vehicles operated by the caterer. However, the analysis of practices at the Copenhagen-based caterer shows that currently the production and distribution stages are handled independently, causing large operational and quality problems. In this paper, we therefore investigate how an integrated planning approach can be used to harmonize production and distribution decisions aiming at a trade-off between the quality of delivered products and the total costs.

The contribution of our study is in integrating fresh food production and distribution planning. An iterative solution framework is developed to investigate the effect of such integration on total costs as well as on the quality of delivered products. We reduce the problem complexity through developing a hierarchical modeling approach, which splits the overall planning problem into a production and a distribution sub-problem. In addition, an aggregation procedure is applied to group customer orders into production batches. A practical approach called block planning is employed for developing the mixed-integer linear programming (MILP) formulation of the production scheduling problem. The distribution problem is solved through a large neighbourhood search algorithm (LNS). Experimental scenarios, designed for critical aspects of the planning problem, reveal the impact of the integrative approach on the overall solution quality.

The remainder of this paper is organized as follows. In the next section, a brief review of the relevant literature on production scheduling and distribution planning in fresh food industries is presented. In Sect. 3, the practical planning problem is presented in detail and the motivating catering case study is explained. In Sect. 4, our hierarchical modeling approach consisting of a heuristic order batching

procedure, the production scheduling model which is based on the block planning principle, and the distribution planning heuristic is introduced. In addition, the iterative solution procedure is presented. The results of our numerical studies are presented in Sect. 5. Finally, conclusions are drawn and further research directions are discussed in the last section.

2 Literature review

The production scheduling problem in fresh food industries, as briefly discussed in Sect. 1, represents a famous class of problems referred to as scheduling with sequence-dependent setups that are well known to be NP-hard (Sun et al. 1999). A complete survey of production scheduling problems with setup consideration is presented in Zhu and Wilhelm (2006) and in Allahverdi et al. (2008). A common difficulty in these problems is the complexity of the resulting models. To overcome such difficulty, Günther and Neuhaus (2004) developed an approach, named block planning. Applications of the block planning concept have not been widely investigated in literature, although this policy is easy to implement and provides decision makers with managerial insights. Günther and Neuhaus (2004) introduced the simultaneous lot sizing and scheduling approach in a so-called make-and-pack production environment. They investigated fast moving consumer goods industries including foods and beverages and realized that a set of technical considerations are most often observed in these businesses. By incorporating such considerations, they could significantly reduce the complexity of production scheduling in these industries and at the same time better reflect real world conditions. Later, Lütke Entrup et al. (2005) developed MILP models based on the block planning concept for production of yogurt that investigate the effect of integrating shelf-life into production planning and scheduling. Günther et al. (2006) presented two different implementations of the block planning concept by considering a continuous representation of time. One utilizes the Production Planning/Detailed Scheduling module of the SAP APO software and the other approach is based on a mixed-integer linear programming formulation. They concluded that the suggested approach is computationally very efficient and provides the flexibility to model a variety of application specific features.

Several aspects of food distribution are investigated in literature as different variants of the vehicle routing problem with or without time windows (VRP/VRPTW). Transportation planning of dairy products is investigated based on a hierarchical approach by Adenso-Diaz et al. (1998). Considering an integrated distribution network, they make decisions on demand assignment, frequency of visiting each customer per week and vehicle routes. They considered a variant of VRPTW with the objective of minimizing the total distribution costs while clients are allocated among vendors in a fair way. Kiranoudis and Tarantilis (2001, 2002) developed a threshold-accepting based algorithm for a heterogeneous fixed-fleet VRP applied to a fresh milk and a fresh meat distribution environment with the goal of minimizing the total transportation time. Belenguer et al. (2005) considered the distribution operations of meat and developed a variant of VRPTW with the goal of

minimizing the lateness in servicing customers and minimizing the total distance traveled (multi-objective). They proposed a 2-phase constructive algorithm and a tabu search based improvement algorithm to solve the resulting model. Ambrosino and Sciomachen (2007) developed an asymmetric capacitated VRP model for delivery of a combination of fresh and frozen foods through vehicles with two compartments. Hsu et al. (2007) studied a VRPTW problem for perishable food delivery with focus on customer response and energy consumption cost as a sustainability measure. Zheng et al. (2007) considered a special transportation infrastructure in Beijing city and developed a VRPTW model for fresh food delivery in cities with circular transportation networks. Osvald and Stirn (2008) investigated distribution operations of fresh vegetables and similar perishable foods. They developed a solution algorithm for the corresponding VRPTW paying attention to the linear quality degradation of products. Akkerman et al. (2010) performed a thorough survey of food distribution studies in which a separate section is devoted to transportation planning.

Integration of production scheduling and distribution planning has received a lot of attention among researchers particularly in the recent years. Chen (2010) classified and reviewed a rather considerable number of articles on general integrated production and distribution scheduling. However, there are very few studies specifically developed for the food industry. Arbib et al. (1999) developed a three-stage matching model to investigate a scheduling problem with the goal of reducing the time gap between distribution and the best-before date. Higgins et al. (2006) considered the integration of production and distribution scheduling for a sugar supply chain. Bilgen and Günther (2010) presented an integrated production scheduling and truck routing model for a fruit juice supply chain. They considered different transportation modes and described the production system based on the block planning approach which establishes cyclical production patterns with regard to the definition of setup families.

To the best of the authors' knowledge, Chen et al. (2009) is the only study that analyses the integration of prepared food production scheduling with short-term distribution planning, including the solution of the associated vehicle routing problem. However, they used a special approach and a set of fundamental assumptions, which are not applicable to our case. While we are considering a more complex production environment with multiple production lines and sequence-dependent setup times and costs, Chen et al. (2009) assumed a production system with a single production line and no setup time and costs. Further, by assuming stochastic demand following a given distribution function, they focused on inventory and shortage costs in their model, while we face a deterministic environment in which customer orders and, as a result, production quantities are given.

3 Problem description

The catering industry is the main food sector dealing with production and distribution of fresh foods. Caterers are concerned with the provision of prepared

food and drinks ready for consumption away from home. The main customers for the catering industry are restaurants, cafes, take-a-ways, hospitals, schools, prisons, residential homes, hotels and other premises, which order food for immediate consumption (Taylor 2008). Catering is by far more complex than other sectors of food industry in terms of the network structure and the underlying constraints. Due to the major concern on delivering foods freshly (and most often hot), direct links between food preparation and distribution stages are prevalent, usually without any intermediate stage.

Fresh food supply chains often consist of three stages (see Fig. 1). The first stage usually comprises batch processing of raw materials into food products, which are packaged in the second and immediately distributed in the third stage. Merging the packaging and the production stages is also frequently observed in fresh food supply chains. Production and distribution planning of fresh food, such as in the catering industry, is certainly a challenging task. The planner has, for example, to cope with a high number of food variants and sequence-dependent setup costs and times of processing equipment as well as fast quality degradation of goods.

In the problem at hand, the caterer operates a production center to produce a multitude of variants of food products ordered from different service organizations. To produce orders on time and in an economical way, a batch production system is adopted to combine orders into production batches. In general, a set of orders can be combined into a batch if the respective food menus require the same temperature level and the same processing time at the ovens.

The setup cost and time in the catering company result from heating up and cooling down of ovens. Since orders need to be processed at different temperatures, the setup cost and time depend on the sequence of producing orders on ovens. Following, a list of other specifications of this production environment is presented.

- Products are produced at a single production site that operates several identical ovens. These ovens form the bottleneck of the production system.
- At the beginning of a production day, all ovens start from the turned-off state.
- A set of customer orders is known a day prior to their realization and has to be processed and delivered within certain time windows.
- Each customer order includes one type of food menu. (A food menu is defined as a unique combination of temperature requirement and processing time.)
- Idle times can be scheduled on ovens, and it is assumed that temperature of ovens does not change during their idle times.
- Food quality decay starts right after the completion of production operations and continues linearly until the delivery time.
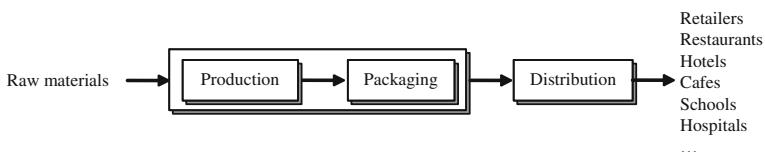

Fig. 1 A sample fresh food supply chain

- The time difference between the completion of production operations of each food and its delivery time represents its quality decay.

The above characteristics of the production environment result in a complex production scheduling problem with sequence dependency in setup costs and times. Moreover, prepared foods are distributed to customers through a fleet of vehicles, which is run by the catering company. Therefore, the caterer is responsible for both production scheduling and distribution decisions. The distribution operations involve transportation of highly perishable products within tight time windows at customer locations. The principal attributes of distribution planning in this company can be summarized as follows:
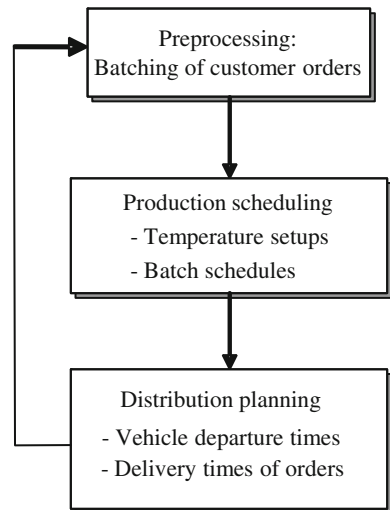
- The distribution problem is defined in the form of planning direct deliveries from the production site to customers (cf. Figure 1) without any intermediate storage.
- Vehicles are homogeneous, and they all start from and return to the central production site known as depot in the vehicle routing literature.
- Transportation costs are given based on the travel distances for delivering the orders.
- The earliest departure time of a vehicle at the production site is determined by the latest finishing times of batches to be delivered by the vehicle.
- For the delivery of orders at customer locations time windows (defined as earliest and latest visiting times) have to be considered.

The entire planning problem comprises decisions on the scheduling of production orders on the various ovens and the distribution of customer orders by use of the company's fleet of vehicles. In particular, this production and distribution planning problem involves the following decisions:

- grouping of customer orders with similar temperature requirements and processing times and with compatible delivery and vehicle departure times into batches for joint processing at the same oven,
- assignment of batches to ovens and sequencing and time phasing of individual batches at each oven,
- determination of temperature levels which have to be realized on each oven according to the assigned batches and their specific temperature requirements,
- assignment of customer orders to vehicles and sequencing and time phasing of individual deliveries to customer locations.

The caterer aims at developing a production scheduling and distribution plan that allows to fulfill customer orders within the given delivery time windows. Major constraints arise from the available production capacities of the ovens and the distribution capacities determined by the fleet of vehicles. All production operations of batches must be completed such that the delivery due dates of the respective customer orders can be met. As for the objective of coordinating daily production and distribution activities, the caterer seeks solutions that ensure an appropriate trade-off between the monetary objective of minimizing total production setup and transportation costs and the non-monetary objective of minimizing the total quality decay of products.

Fig. 2 Hierarchical modelling
framework


```
┌─────────────────────────┐
│      Preprocessing:      │
│ Batching of customer orders │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│   Production scheduling   │
│   - Temperature setups    │
│   - Batch schedules       │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│   Distribution planning   │
│ - Vehicle departure times │
│ - Delivery times of orders │
└─────────────────────────┘
```

4 Hierarchical modeling approach

Due to its complexity, the integrated production scheduling and distribution planning problem presented in the previous section is computationally intractable. Therefore, we propose a hierarchical modeling approach (see Fig. 2), which subdivides the entire planning problem into sub-problems of considerably reduced complexity. In a preprocessing stage, an aggregation procedure is applied, which creates batches of customer orders with similar temperature and processing requirements and compatible delivery and vehicle departure times (Sect. 4.1). This way, the number of entities in the production scheduling stage is considerably reduced. For processing the various batches on the ovens, a so-called block planning approach is proposed which schedules the batches according to a pre-defined sequence of temperature levels (Sect. 4.2). Given the finishing times of the batches, a heuristic solution algorithm is applied to solve the distribution planning problem (Sect. 4.3). Finally, an iterative framework is introduced to coordinate the production and distribution schedules (Sect. 4.4).

4.1 Batching of customer orders

The main idea behind the heuristic batching procedure is to reduce the size of the production scheduling model by grouping consistent customer orders into a set of production batches. As a result, the original problem is reduced to scheduling a considerably smaller number of entities (batches vs. orders). Customer orders can be grouped into a batch if they have the same temperature and processing time requirements and satisfy the following consistency conditions:

Their delivery times and their corresponding vehicle departure times do not differ by more than a pre-defined threshold value.

HEURISTIC BATCHING ALGORITHM
1   *function* Batching ($i \in$ {orders})
2   *sort* all orders in list **L** as follows:
3   Based on temperature in ascending order
4   Based on vehicle departure times in ascending orders for orders with the same temperature
5   *while* | **L**|>0 *do*
6   *create* a new **batch** and *set* **batch**.Load=0
7   **i** = first order in **L**
8   *assign* **i** to the created **batch** and *add* **i**.Demand to current **batch**.Load
9   *remove* **i** from **L**
10   **j** = first order in **L** and *set* **A** = | **L**|
11   *while* (**A** > 0) *do*
12   *if* ((**j**.Temperature == **i**.Temperature) &
(**j**.ProcessingTime == **i**.ProcessingTime) &
(**j**.DeliveryTime < **i**.DeliveryTime + **threshold**) &
(ABS(**j**.DepartureTime - **i**.DepartureTime) < **threshold**) &
(batch.Load+**j**.Demand <=   Capacity))
13   *assign* **j** to the current **batch** and *add* **j**.Demand to current **batch**.Load
14   *remove* **j** from **L**
15   **j**= next order in **L** and *set* **A** = **A-1**
16   *else*
17   **j** = next order in **L** and *set* **A** = **A-1**
18   *endwhile*
19   *endwhile*
Note: ABS refers to absolute value

Fig. 3 Heuristic batching algorithm

The respective food products can be processed on an oven within the given oven capacity.

Details of the heuristic batching algorithm are given in Fig. 3. Depending on the considered case, different sorting criteria, e.g. temperature level, processing time, vehicle departure time, delivery time, and demand volume could be chosen for sorting the orders in list *L*. In our case, we first rank orders based on their temperature levels in ascending order. As a secondary ranking criterion, the earliest vehicle departure time of orders is chosen. An important parameter in the algorithm is the acceptable time gap between the vehicle departure and delivery times of orders that are grouped into one batch. Later in the numerical analysis (see Sect. 5), we investigate the influence of this threshold value on the final solution as well as on the algorithm performance.

4.2 Production scheduling

Principally, lot sizing and scheduling models can be formulated using a discrete or a continuous representation of time. In a discrete time-scale model, which is still predominant in the academic literature, the start and ending of production and setup activities are restricted by the period boundaries. Obviously, a dense time-grid is

needed to adequately reflect the timing of production activities. In a continuous time based model, however, the succession and detailed timing of production activities can be reflected more realistically.

In the practical application at hand, an issue which further increases the complexity is the sequence dependency of setups. A practical approach which exploits the advantages of both the continuous time representation and the use of predefined setup sequences is the so-called block planning concept (cf. Günther et al. 2006, and Bilgen and Günther 2010). This concept is widely adopted in production environments with a single bottleneck stage, such as in large parts of the consumer goods industry. For the problem at hand, this method not only reflects the practical settings but also significantly reduces the complexity of the problem. In this approach, an important aspect is to define a block as a pre-set sequence of setups based on human expertise and technological requirements. Afterwards, production activities are scheduled based on the pre-defined setup sequence. For instance, in food processing industries, production managers often sequence production runs from the less intensive taste of a food product to the stronger or from the brighter color of a product to the darker.

In catering environments, a setup operation refers to setting up a certain temperature on an oven. As a result, a sequence of temperature increases and/or decreases (a block) has to be defined on each oven. However, since heating up an oven is much quicker than cooling it down, orders are scheduled on ovens based on their temperature requirements in an ascending order. Therefore, in our catering problem, we define a block as a sequence of all required temperature levels in ascending order.

Setup costs and times for heating up ovens are assumed to be additive: Heating up an oven from temperature T_1 to T_3 is the same as heating it up from T_1 to T_2 and then from T_2 to T_3 (with $T_1 < T_2 < T_3$). Therefore, after defining blocks (temperature patterns), the total setup costs of an oven (production line) can be defined as the cost of heating up the line to its highest realized temperature. Accordingly, an adequate setup strategy tends to assign orders with lower temperature requirements to a line, which produces other orders with higher temperatures, thereby yielding the most economical schedules. Based on this definition, no additional setup cost is incurred if a new order can be inserted into an existing schedule without increasing the maximum temperature requirement of the schedule. However, it is essential that all orders are processed in ascending order of their temperature requirements. This constraint might enlarge the time gap between the production and delivery time of certain orders and impair the total quality measure. Therefore, an appropriate trade-off has to be sought between setup costs and the quality decay objective.

By means of the block planning concept we are able to capture sequence dependencies in setup costs and times and simplify the scheduling problem. Next, an MILP model based on the block planning concept is formulated, assuming that initial distribution decisions are given and the heuristic batching algorithm has already been applied to group orders into a set of batches. We further assume that batches are numbered such that the required temperature levels are non-decreasing. The notation is introduced in Tables 1, 2 and 3.

Table 1 Indices and sets

$i, i' \in S \cup \{O, \bar{O}\}$	Customer orders (S), production site ($\{O\}$:starting depot), production site ($\{\bar{O}\}$: ending depot)
$k \in J$	Batches
$i \in J'_k$	Customer orders processed in batch k
$k \in J_h$	Batches requiring temperature level h
$v \in V$	Vehicles
$l \in L$	Production lines (ovens)

Table 2 Parameters

W_1, W_2, W_3	Weights for cost terms used in the objective function
$c_{i',i}$	Transportation cost for delivering order i immediately after i'
$c'_{h,l}$	Setup cost for realizing temperature h on line l
B_i	Delivery date for customer order i (given from the distribution model)
$X_{i',i,v} = 1$	Binary parameter (given from the distribution model) indicating if order i is delivered immediately after i' by vehicle v (0, otherwise)
θ_i	Quality decay rate of customer order i per unit of time
t_k	Processing time of batch k
D_k	Due time for producing batch k
M	Sufficiently large number
g_k	Setup time for producing batch k after batch $k-1$

Table 3 Decision variables

$Y_{h,l} = 1$	If temperature h is realized on line l (0, otherwise)
$Z_{k,l} = 1$	If batch k is realized on line l (0, otherwise)
$F_{k,l} \geq 0$	Finishing time of producing batch k on line l (with $F_{0,l}$ given)
$F'_k \geq 0$	Finishing time of producing batch k
$SC_l \geq 0$	Setup cost on line l

Model formulation:

$$Min\ W_1 \cdot \sum_{l \in L} SC_l + W_2 \cdot \sum_{k \in J} \sum_{i \in J'_k} (B_i - F'_k) \cdot \theta_i + W_3 \cdot \sum_{v \in V} \sum_{i \in S \cup \{\bar{O}\}} \sum_{i' \in S \cup \{O\}} c_{i',i} \cdot X_{i',i,v}$$

(1)

subject to

$$SC_l \geq c'_{h,l} \cdot Y_{h,l} \quad h \in H, l \in L \tag{2}$$

$$\sum_{k \in J_h} Z_{k,l} \leq M \cdot Y_{h,l} \quad h \in H, l \in L \tag{3}$$

$$\sum_{l \in L} Z_{k,l} = 1 \quad k \in J \tag{4}$$

$$F_{k,l} \geq F_{k-1,l} + g_k + t_k \cdot Z_{k,l} \quad k \in J, \, l \in L \tag{5}$$

$$F_{k,l} \leq D_k + M \cdot (1 - Z_{k,l}) \quad k \in J, \, l \in L \tag{6}$$

$$F'_k \leq F_{k,l} + M \cdot (1 - Z_{k,l}) \quad k \in J, \, l \in L \tag{7}$$

The objective function (1) minimizes the weighted sum of total setup costs, total food decay and transportation costs. According to Eq. 2, setup costs depend on the highest temperature level realized on each line. Food decay is measured by the time difference between the delivery time of an order and its production finishing time. Since each order is assigned to a batch, the finishing time of an order is the same as the finishing time of its corresponding batch. Transportation costs are calculated based on the traveled distance of vehicles for visiting all customers. In this model, the transportation part of the objective function is a constant given as the solution to the distribution problem. However, for the sake of uniformity, we keep this term in the objective function. The objective function consists of monetary as well as of non-monetary terms. Thus, an appropriate trade-off is pursued between cost components and the quality measure in the numerical tests. Constraints (3) prevent realization of a batch on a line unless its corresponding temperature level is used on that line. Constraints (4) guarantee that each batch will be scheduled on one line. Constraints (5) indicate the relationship between production finishing times of two successive batches. In these constraints, the processing time of a batch t_k, depends on the processing time of orders assigned to it by the heuristic batching algorithm. Constraints (6) ensure that when a batch is produced on a line, its production finishing time on that line is smaller than its due time. It should be noted that due to the negative sign of the F-variables in the objective function orders are scheduled as late as possible. Since a preliminary distribution plan with the departure time of vehicles is assumed to be given, constraints (6) impose a proper upper bound on the production finishing time of each batch on each line. Constraints (6) cause infeasibility when the available production capacity is not sufficient to produce orders before their due times (D_k). Constraints (7) determine the production finishing time of each batch depending on its assigned production line. Variable domains are already given in Table 3 with the definition of the decision variables.

The application of the block planning concept and the heuristic batching algorithm significantly simplifies the production scheduling problem. As a result, the developed model is easy to solve using standard optimization software.

4.3 Distribution planning

Given the production scheduling decisions, the distribution problem determines the vehicles' tours and schedules. It seeks solutions that minimize a trade-off between transportation cost and the quality decay component of the objective function. Further, it ensures that the developed production schedules are observed in determining the vehicle tours. The distribution problem in our catering case can be formulated as a variant of the vehicle routing problem with time windows (VRPTW). Since it is a classical model (with a different objective function) and we

aim at solving this problem using a heuristic approach, we do not present the mathematical model for the distribution problem.

Several heuristic algorithms were developed to deal with the VRPTW, which mainly work on the exchange of two or a few nodes between routes. Such node exchange algorithms can be considered as small neighborhood search methods. However, it is believed that such small neighborhood explorations, even when combined with meta-heuristic methods, are not effective in avoiding local optimality thereby making it difficult to jump from one promising area of the solution space to another, when a tightly constrained problem is solved (Ropke and Pisinger 2006). In the following, we develop a large neighborhood search (LNS) algorithm that obtains good quality solutions for this problem.

The LNS algorithm was first applied to the VRPTW by Shaw (1997) showing its superiority over other heuristic methods. This algorithm iteratively relaxes and re-optimizes the current solution. In VRPTW applications, the relaxation phase refers to removing a certain number of orders from the available routes and the re-optimization stage inserts them back in the feasible positions based on some optimization criteria. Therefore, in designing an LNS algorithm, the number of orders to remove in each iteration (denoted by q) and the removal and insertion operators must be determined. In this paper, we adopt two operators for removal and perform the insertion based on the greedy insertion method. The first removal operator is taken from Shaw (1997) and described in Fig. 4.

The relatedness measure $R_{i',i}$ between orders i' and i is defined in Eq. 8. This equation determines the relatedness of every pair of orders based on four criteria, namely travelling cost $c_{i',i}$, delivery time difference $|B_{i'} - B_i|$, being served by the same vehicle $\left(T'_{i',i} = 1\right)$ or not $\left(T'_{i',i} = 0\right)$, and load difference $|d_{i'} - d_i|$. The corresponding weights for these criteria are set to $\alpha = 0.75$, $\beta = \delta = 0.1$ and $\gamma = 1$ based on the results of preliminary test runs.

$$\frac{1}{R_{i',i}} = \alpha c_{i',i} + \beta |B_{i'} - B_i| + \gamma T'_{i',i} + \delta |d_{i'} - d_i| \tag{8}$$

Further, we consider a simple worst-case removal operator as a second removal operator to diversify the search scope. This operator removes orders based on their influence on the objective function. First, a random number is generated and is

Shaw's removal procedure
1.   *randomly* choose one of the orders from the routes to be removed
2.   *while* not enough orders are removed *do*
3. i = an order chosen randomly from the set of removed orders
4.   *Rank* all unremoved orders based on their relatedness to order i
5. r = a random number in [0,1)
6.   *Remove* the order which is r^i of the way through the rank
7.   *endwhile*

Fig. 4 Shaw's removal procedure

Large Neighborhood Search algorithm
1 *function* LNS ($s \in$ {solutions} , q)
2   *while* stopping criterion is not met *do*
3         $s' = s$
4   r = a random number in [0,1]
5   *while* q orders have not been removed *do*
6   if ($r < p$)
7   *remove* one order from s' based on Shaw operator
8   *else*
9   *remove* one order from s' based on worst case operator
10   *endwhile*
11   *reinsert* removed orders into s' based on greedy insertion operator
12   if ($f(s') < f(s)$)
13               $s = s'$
14   *endwhile*
15   *return* s

Fig. 5 Large neighborhood search algorithm

compared to a parameter p. If the generated number is less than p, Shaw's operator is applied. Otherwise, the worst case operator is chosen. Considering two different operators incorporates a random component into the algorithm while orders are still removed systematically. The appropriate value of p depends on characteristics of each case study. In the case at hand, we set p to 0.75 as it results the best solution in our numerical implementations. In each operation of the LNS algorithm, the greedy insertion operator ranks the removed orders based on their influence on the objective function and selects the one which increases the total costs the least. In order to increase the feasible solution space for the algorithm, time-windows at customer locations are considered as soft constraints that can be violated subject to high penalty costs. The developed LNS algorithm stops after a certain number of iterations and returns the best incumbent solution as described in Fig. 5.

4.4 Iterative solution procedure

To solve the integrated production and distribution problem, an iterative solution framework is developed as illustrated in Fig. 6. We initialize the solution algorithm by assuming a set of very early production finishing times as the starting point. Afterwards, an iterative procedure is employed starting with the distribution model and followed by the production scheduling model. The main reason for this initialization is that every set of early production finishing times allows obtaining a feasible schedule for the distribution problem. Afterwards, the production schedule is determined by solving the production scheduling model using the departure times of vehicles as inputs. Due to the time windows around lunch time, the departure times are such that a feasible solution for the production scheduling problem can always be obtained. On the other hand, the distribution problem is facing tight time windows at customers, which limits the possibility of accepting any arbitrary
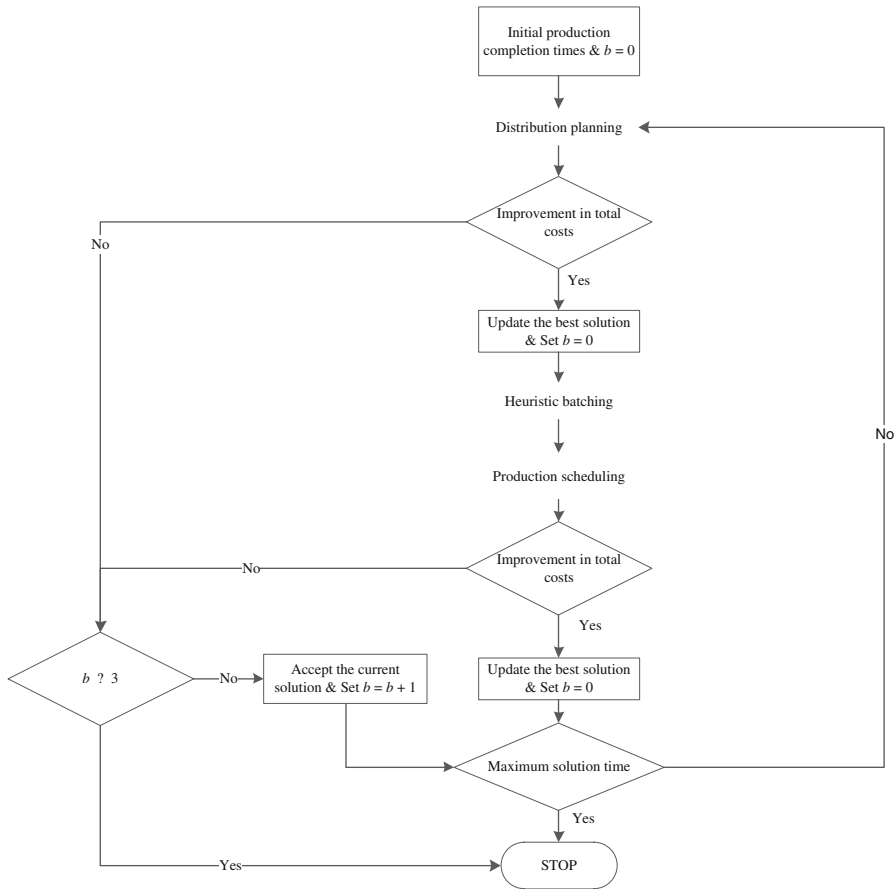
Fig. 6 Iterative solution method

distribution pattern as the starting point. Therefore, different starting points are utilized to initialize the algorithm by different early production finishing times.

To stop the algorithm, we used two termination criteria: (1) maximum solution time, as a common stopping criterion in designing iterative algorithms, and (2) maximum number of iterations in which the model cannot achieve any improvement in the objective function. The reason for the second criterion is that either a new solution of the heuristic batching algorithm or a degenerate solution of the production scheduling problem is obtained. At each iteration, the heuristic batching algorithm may find a new set of batches resulting in a different solution space of the production scheduling problem. Therefore, it re-initializes the algorithm with a new initial solution, which may lead to further improvements in total costs in subsequent iterations. Also, in the case of problem degeneracy, the production scheduling problem may find solutions with the same objective function. Therefore, we let the algorithm continue unless the same objective value is obtained b times in a row (with b being the maximum number of acceptable iterations without any

improvement in the objective function). In the numerical investigations, we noticed that setting b to be 3 is sufficient for our case study.

Figure 7 illustrates one iteration of the solution algorithm showing how the production scheduling problem is simplified by using the heuristic batching algorithm and the decomposition into the production and distribution problem. This figure considers a simple instance with eleven orders, three temperature levels and three ovens. For simplicity, all order processing times are assumed to be the same and a single threshold value of five time units is used for both delivery and departure time-differences. Further, we assume that all order sizes are 3 units and the capacity of the ovens is 9 units.

Given a set of production schedules, the LNS algorithm is called to solve the distribution problem generating vehicle routes and departure times as depicted in the first box of Fig. 7. As a result, vehicle departure times (Sv) and customer delivery times are determined. These variables serve as input in running the heuristic batching algorithm, which mainly comprises two steps shown in the second and third box. In the second box, orders are sorted in a list (L) in ascending order based on their temperature requirements and their due times (their corresponding vehicle departure times). The list of sorted orders, L, represents the priority of orders for being grouped into batches. The third box shows the grouping of orders into batches based on the consistency conditions in the following way. The first order in L, order 2, is selected and removed from the list to form batch 1. Then, considering the two combination conditions of the heuristic batching algorithm (see Sect. 4.1), the batching procedure combines consistent orders from the top of the list into batch 1. Next, it generates a new batch (batch 2) which includes the first order that cannot be combined with batch 1, and combines the remaining consistent orders with this order with regard to the two combination conditions. In the illustrated example, first the consistency of order
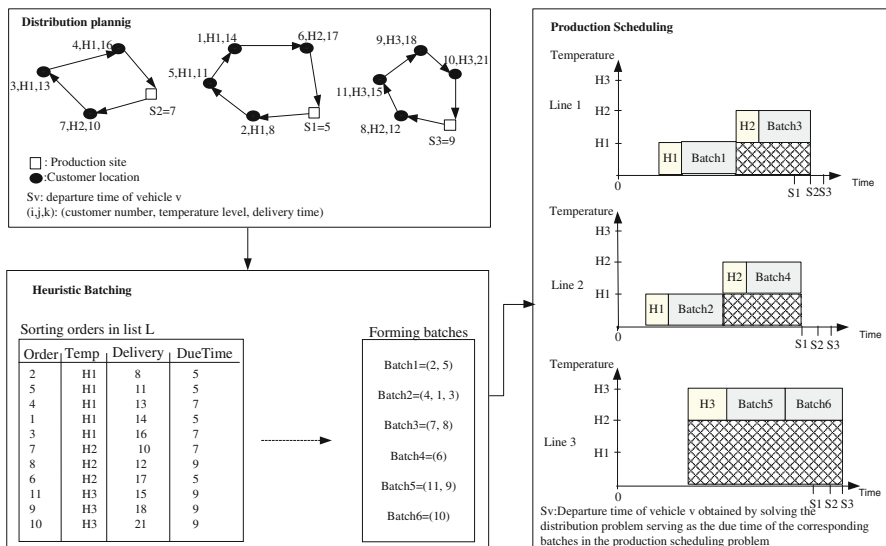

Fig. 7 An illustrative example of the solution method

5 with order 2 is checked. Since its processing time and temperature requirement are the same as of order 2, it is consistent with order 2. Besides, its delivery and departure times are closer than 5 time units to those of order 2 (condition 1) and the aggregate load of the batch does not violate the oven capacity (condition 2). Therefore, order 5 is removed from the list and added to batch 1. However, the remaining orders cannot be combined with order 2. For example, order 4 cannot be added to batch 1 because its delivery time is not closer than 5 time units to that of order 2 (condition 1 is violated). Therefore, batch 2 is created for the first remaining order of the list, order 4. Later, order 7 cannot be added to batch 2 because its temperature requirement is not the same as that of order 4. The algorithm continues until all orders are grouped into a set of batches. In this small instance, the algorithm reduces the number of items to be scheduled from 11 to 6.

Then, the production scheduling problem is solved for the generated batches such that each batch is finished before its due date. (The due date of a batch is the departure time of the vehicle that carries orders of that batch). As shown in Fig. 7, lines 1 and 3 start later than line 2. This is because the due dates of their batches allow them to postpone the production time to reduce the time gap between the production and delivery times of orders. Finally, the resulting production schedules are released into the distribution module as input for the next iteration, where the production finishing times provide bounds on the vehicle departure times.

5 Numerical results

In this section, the proposed planning approach is evaluated through a number of numerical tests. We developed a main test set consisting of 200 orders based on the real setting of the catering company at hand. Three additional test sets are derived from the main set for testing the algorithm on different problem sizes (see Sect. 5.1). To obtain problem instances, we modified Solomon's R201 test instance (Solomon 1987) regarding customer locations and delivery durations. In case 1 and 2, Solomon's R201 data are used by considering the first 50 and 100 orders. In case 3 and 4, the first 100 orders from the second test instance are copied and location data are randomly generated for the next 50 and 100 orders such that they locate in the defined area of the first 100 customers. Customer order size is randomly generated in the range of 40–60 units based on a uniform distribution function. Orders are randomly assigned to one of the available temperature levels, processing times and decay rates. Production setup costs and times are determined based on the difference in temperature levels while transportation costs depend on the distance between customer locations. Further, setup costs, transportation and quality decay costs are weighed 0.2, 0.3 and 0.5, respectively, in the objective function. We also determined the number and capacity of ovens and vehicles based on the generated customer demand. Furthermore, to better represent the company's case, we set the same time window of 11 a.m. to 1 p.m. for delivering all orders. Table 4 briefly presents the other production and transportation parameters used in the different test sets.

To implement the algorithm, we coded the iterative framework, heuristic batching and LNS algorithm using VC++ 6 and modeled the production

Table 4 Problem parameters

Test set number	1	2	3	4
# of orders	50	100	150	200
# of ovens	2	3	4	5
# of vehicles	8	15	20	25
# of temperature levels	3	4	4	5
# of processing times	3	3	3	3
Decay rates	{3,6,9}	{3,6,9}	{3,6,9}	{3,6,9}

scheduling problem with ILOG OPL 6.1.1. CPLEX 11.2 was used as the solver. All programs are run on a PC with Intel P4 3.2GH CPU and 1 GB RAM.

In designing the LNS algorithm, two parameters are to be set: The number of nodes to be removed and reinserted at each iteration (q), and the number of iterations. Selecting a large value for q will lead to a redesign of a large part of available routes at each iteration. However, since a very myopic operator is used for the insertion procedure, the chance to generate a feasible solution would be very small. Further, such setting inflates the computational time. On the other hand, selecting a very small value for q will limit the chance of exhaustively exploring the potential neighborhoods. Based on the problem settings and time windows, we found a value of 3 to be appropriate for q. The number of iterations is set to 8,000 as the algorithm rarely achieves a significant solution improvement thereafter.

Since the selection of a certain starting point might affect the way the algorithm proceeds in the next stages, we considered different sets of initial solutions. In determining the starting points for the algorithm, we took into account the requirements of the distribution problem in terms of the time-windows at customer locations. As a result, different sets of potential earliest and latest production finishing times are considered. For each set, once we ran a test instance by assigning an identical production finishing time to all orders. Further, we produced other sets of starting solutions by assigning different production finishing times to different orders as random numbers between the potential earliest and latest finishing times. For running the numerical analysis, we chose the starting solution set, which produced the best results.

The next two parameters that need to be determined are the threshold parameters in the heuristic batching algorithm. They influence the number of batches to be generated and significantly affect the performance of the algorithm. We set the threshold parameters to a value of 70 min and investigate the sensitivity of these values in Sect. 5.3. Finally, we should mention that we limited the total solution time to 40 min through which most problem instances are solvable, regardless of the chosen threshold values.

The principal questions addressed in the numerical studies are as follows:

1. What is the effect of integrating production and distribution decisions?
2. How does the perishability of products impact the overall solution quality?
3. How do different parameter settings for the heuristic batching algorithm influence the final solution?

4. How can an appropriate balance be found between the conflicting objectives of minimizing quality decay and minimizing setup and transportation costs?

5.1 Sequential versus integrated planning approach

As a benchmark, we consider a sequential approach similar to the procedure currently used by the company. This approach performs the production and distribution planning tasks separately and only based on their related cost terms, i.e. the quality decay is not optimized in this approach. Production activities start at the beginning of the shift every day (6:00 am), and we consider a realistic fixed due time (10:00 am) based on staffs' experience to finish all production activities. To determine the production schedule in the sequential approach, the concept of block planning is applied. We first run the heuristic batching by only considering combination condition 2 of the heuristic batching procedure. After running the heuristic batching, production schedules are determined imposing constraints on the distribution problem. Then, the distribution problem is solved with regard to the established production schedules.

Table 5 shows that the relative improvement in the quality decay of the integrated approach over the sequential approach increases considerably with problem size. In the problem instance with 200 customers (main problem), integration resulted in a significant decrease of 40.9% in quality decay. At the same time, the increase in transportation and production costs is about 18.7%. The magnitude of quality improvement justifies the integrated planning approach despite the increase in production and transportation costs. In problems with a small number of orders, the number of consistent orders that can be combined into a batch decreases considerably. The heuristic batching procedure hence creates a large number of small-sized production batches. Therefore, the oven's capacity is not well utilized leading to an increase in makespan. This leads to high quality decays in batches that must be produced early. However, the algorithm still works considerably better than the sequential approach for all problem instances. In Sect. 5.3, we further investigate other settings that yield a large number of small batches and show how they impair the quality of the final solution.

5.2 Integrated decision making and product characteristics

We now analyze how the degree of product perishability influences the importance of integrated decision making. In this regard, three variants of the main problem are

Table 5 Comparison of the integrated versus the non-integrated approach

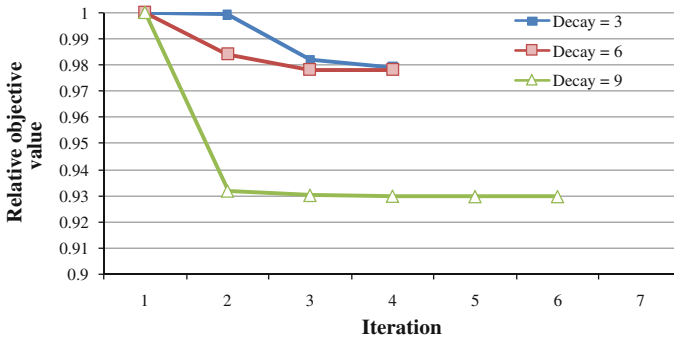| Number of customers	50	100	150	200
Relative change in the quality decay value	−18	−27.7	−36	−40.9
Relative change in transportation and production costs	+12.3	+16.5	+18.3	+18.7
Relative change in objective function	−9.5	−14.7	−22.3	−28.2

Fig. 8 Decay rate analysis

analyzed each with all products having the same decay rate of 3, 6 and 9, respectively. It can be expected that a higher reduction in the objective function value is conceivable by integrating production and distribution decisions when products are more perishable. Figure 8 shows that the relative objective value improves more when products are more perishable. However, in our problem instance, the final improvement for decay rates 3 and 6 is almost identical.

5.3 Parameter settings in the heuristic batching algorithm

As explained in Sect. 4.1, the heuristic batching algorithm uses threshold parameters to identify consistent orders that can be grouped into batches. Generally, a threshold parameter defines a certain acceptable distance from a base value of a parameter. In this section, we analyze the effect of threshold parameters by running the main test set under different values for both the delivery and departure thresholds. We tested a range of values between the two extreme settings of 0 and 120 min for the threshold parameters. In our main problem instances, the value of 120 min for the threshold parameters is an extreme setting because the total time-window over which a customer may be served is between 11 a.m. and 1 p.m.

Setting the threshold value to zero prevents order combinations. Hence, orders have to be scheduled sequentially based on the list L (see Fig. 3). In such situation, since the number of production lines is limited, early scheduling of the first batches is unavoidable. This solution is indeed very poor in terms of quality decay. In the other extreme, setting a too large threshold value reduces the number of batches, but combines orders that better not be produced together. Thus, some orders will again be produced earlier than necessary and the quality decay will be high. As illustrated in Fig. 9, the threshold parameters mainly influence the decay cost, while other objectives are impacted much less. We found the best results using a value of 70 min for the threshold. The figure also depicts that, in general, high threshold values are more appropriate than very low values. Although a high threshold may combine more orders than necessary, a too low threshold value has a stronger impact on decay costs by generating too many batches and increasing the total
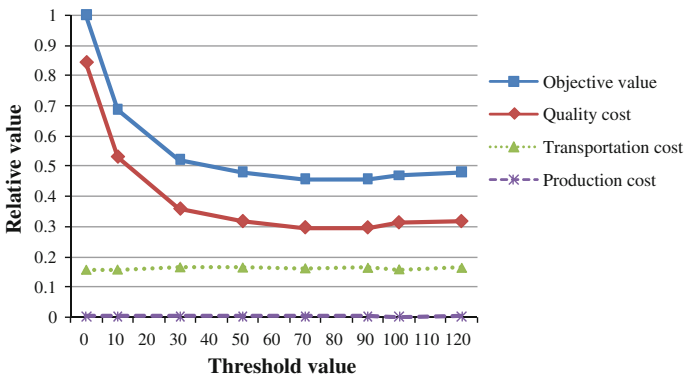
Fig. 9 The effect of threshold value

production time. Such extra production time forces the whole schedule to start earlier, thus yielding a higher quality decay cost.

The total production time can be reduced to a certain extent by setting higher threshold parameters. On the other hand, a direct relationship is inherent between the length of processing times and the total production time such that the longer the processing times are, the longer will be the total production time and vice versa. Therefore, in a problem instance where all parameters are fixed, if processing times are increased, the optimal decision is most likely to increase the threshold values to compensate for the newly added production times.

5.4 Finding an appropriate trade-off between quality decay and other cost terms

The last part of the numerical analysis addresses the question of how an appropriate trade-off can be found between quality decay and production and transportation costs. To investigate the relationship between these conflicting objectives, we solved the main test problems with five different weights for decay cost and considered equal weights for production and transportation costs. In our problem setting, setup and transportation costs are relatively smaller than decay cost. Therefore, by increasing the decay cost weight the model compromises on the setup and transportation costs and adapts distribution decisions accordingly to improve the quality term. In Fig. 10, the corresponding unweighted values of quality decay, production and transportation costs are compared. As illustrated in this figure, by increasing the decay weight from 0.1 to 0.5 a continuous decrease in the decay objective can be achieved without exorbitantly increasing total transportation and production costs. The two dashed lines represent the share of production and transportation in the total costs.

By increasing the decay weight beyond 0.5 still a major improvement in food quality is achieved. However, the achieved improvement in food quality goes along with a significant increase in the other cost terms. Thus, the cost related consequences of such quality improvement should be carefully discussed with managers of the company for setting the quality decay weight in such situations.
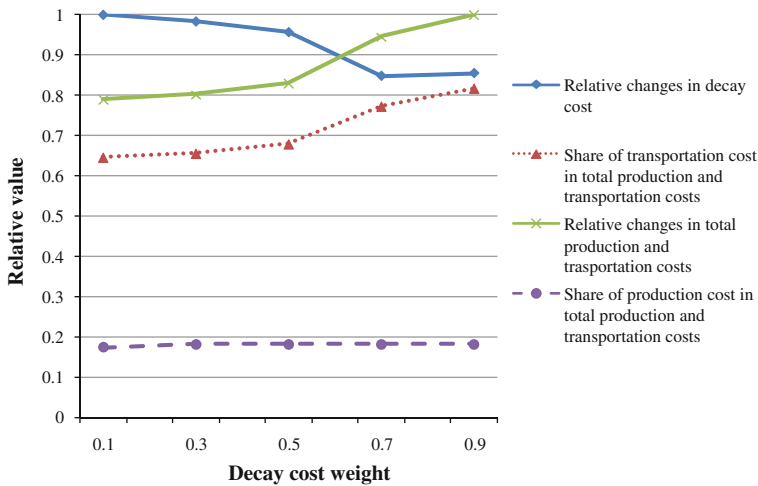
Fig. 10 Decay weight analysis

One possibility for further improving quality cost without increasing the production and transportation costs would be to change the block definition. Currently, production blocks are defined based on an incremental temperature pattern, because heating up ovens is much easier than cooling them down. However, if we could change this technical limitation, we could for example obtain block definitions that allow for cooling down ovens at the end of their production schedules. Such consideration could significantly increase the flexibility of the production system.

5.5 An alternative quality improvement approach

As an alternative approach, we first run a pre-optimization procedure, through which a problem instance is solved with quality decay being optimized by a min–max function. Contrary to the previous approach in which the difference between the highest and lowest product quality might be very large, the new method ensures that all customers enjoy a minimum level of quality. In order to implement this approach we introduce a new decision variable and a new set of constraints as follows:

$$\varphi \geq 0 \quad \text{Maximum quality decay among all delivered products}$$

$$\varphi \geq \sum_{\forall k \in J} \left(B_i - F_k''\right) \cdot \theta_i \quad \forall i \in S \tag{9}$$

Further, we modify the objective function accordingly to incorporate the newly defined variable as follows:

$$Min\, W_1 \cdot \sum_{l \in L} SC_l + W_2' \cdot \varphi + W_3 \cdot \sum_{v \in V} \sum_{i \in S \cup \bar{O}} \sum_{i' \in S \cup O} c_{i',i} \cdot X_{i',i,v} \tag{10}$$

Table 6 Comparative results between with and without the pre-optimization stage

	Impact of the pre-optimization step
Total quality decay	1.0001
Maximum perceived quality decay	0.86
Minimum perceived quality decay	0.96

Constraints (9) set φ as the maximum quality decay among all delivered products and equivalently the minimum quality level perceived by customers. Further, the quality component is defined in the objective function (10) such that it minimizes the value of φ resulting in a better lower bound on the quality level. The same analysis as in Sect. 5.4 is required to find the appropriate weight W_2'.

In a pre-optimization stage, constraints (9) are added to both the production scheduling and transportation planning problems, and they are solved with the modified objective function (10) considering φ as a decision variable. Afterwards, the modified problem is solved again, with objective function (1) considering the value of φ as a given parameter. This parameter provides an upper bound on the quality decay that a customer might observe. To investigate the influence of the pre-optimization stage, we solved the main problem instance (with 200 customers) and compared its results with those of the previously introduced method. Table 6 presents the ratio of the results achieved when the pre-optimization stage is considered versus the option without pre-optimization. The comparison reveals a 14% improvement in maximum perceived quality decay (minimum quality level) and a 4% improvement in minimum perceived quality decay for a virtually unchanged total quality decay (increase of 0.01%).

6 Summary and conclusions

Among several specific characteristics of fresh food industry, the quality concern is perhaps the most prominent one. In the fresh food industry, food quality is strongly influenced by the two interrelated stages of production and distribution. However, due to the complexity of these problems, research attention is usually limited to solving them separately thereby neglecting their strong interrelation. Being aware of such complexity, we incorporated some simplification steps to develop an integrated approach for production scheduling and distribution planning of fresh foods such that the model is still a good representation of the real world situation. To this end, we focus on decomposing the production model into two successive stages and designing a solution algorithm to appropriately integrate them with the distribution model. As a result, we were able to reduce the problem complexity considerably and to solve the integrated problem in a reasonable time.

Results of our numerical investigation indicate that the integrated approach significantly improves food quality without considerably increasing production and distribution costs. Further, we showed that an appropriate minimum quality level can be guaranteed through our alternative approach which imposes a measure of

balance among customers service levels. We also showed that the improvement in food quality is larger when foods are more perceptible to quality decay. However, focusing on food quality improvement is sound and commercially viable only as long as the total costs do not exceed certain limits. We took into account the particular characteristics of a catering problem in developing the mathematical model, designing the algorithm, and setting the parameters. The developed models reflect major issues arising in catering industry based on the explained case study. Nevertheless, their main features are also relevant for other types of fresh food industries like bakeries. Analyzing more complex cases with multiple producers is the subject of our future research.

References

Adenso-Diaz B, Gonzalez M, Garcia E (1998) A hierarchical approach to managing dairy routing. Interfaces 28(2):21–31

Akkerman R, Farahani P, Grunow M (2010) Quality, safety and sustainability in food distribution: a review of quantitative operations management approaches and challenges. OR Spectr 32:863–904

Allahverdi A, Ng CT, Cheng TCE, Kovalyov MY (2008) A survey of scheduling problems with setup times and costs. Eur J Oper Res 187:985–1032

Ambrosino D, Sciomachen A (2007) A food distribution network problem: a case study. IMA J Manag Math 18:33–53

Arbib C, Pacciarelli D, Smriglio S (1999) A three-dimensional matching model for perishable production scheduling. Disc Appl Math 92:1–15

Belenguer JM, Benavent E, Martinez MC (2005) RutaRep: a computer package to design dispatching routes in the meat industry. J Food Eng 70:435–445

Bilgen B, Günther H-O (2010) Integrated production and distribution planning in the fast moving consumer goods industry: a block planning application. OR Spectr 32:927–955

Chen ZL (2010) Integrated production and outbound distribution scheduling: review and extensions. Oper Res 58(1):130–148

Chen HK, Hsueh CF, Chang MS (2009) Production scheduling and vehicle routing with time windows for perishable food products. Comput Oper Res 36:2311–2319

Günther H-O, Neuhaus U (2004) Advanced planning and scheduling in the consumer goods industry: realizing block planning concepts for make-and-pack production using MILP modelling and SAP's APO software. Oper Manag Change Agent 2:757–766

Günther H-O, Grunow M, Neuhaus U (2006) Realizing block planning concept in make-and-pack production using MILP modelling and SAP APO. Int J Prod Res 44:3711–3726

Higgins A, Beashel G, Harrison A (2006) Scheduling of brand production and shipping within a sugar supply chain. J Oper Res Soc 57(5):490–498

Hsu CI, Hung SF, Li HC (2007) Vehicle routing problem with time-windows for perishable food delivery. J Food Eng 80:65–475

Kiranoudis CD, Tarantilis CT (2001) A meta-heuristic algorithm for the efficient distribution of perishable foods. J Food Eng 50:1–9

Kiranoudis CD, Tarantilis CT (2002) Distribution of fresh meat. J Food Eng 51:85–91

Lütke Entrup M, Günther H-O, Van Beek P, Grunow M, Seiler T (2005) Mixed-integer linear programming approaches to shelf-life-integrated planning and scheduling in yoghurt production. Int J Prod Res 43:5071–5100

Osvald A, Stirn LZ (2008) A vehicle routing algorithm for the distribution of fresh vegetables and similar perishable food. J Food Eng 85:285–295

Ropke S, Pisinger D (2006) An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. Transp Sci 40:455–472

Shaw P (1997) A new local search algorithm providing high quality solutions to vehicle routing problems. Technical report, University of Strathclyde, Scotland

Solomon MM (1987) Algorithms for the vehicle routing and scheduling problems with time windows constraints. Oper Res 35:254–265

Sun X, Noble JS, Klein CM (1999) Single-machine scheduling with sequence dependent setup to minimize total weighted squared tardiness. IIE Trans 31:113–124

Taylor E (2008) A new method of HACCP for the catering and food service industry. Food Control 19:126–134

Zheng W, Hu X, Zeng AZ (2007) A two-stage solution procedure for food delivery decisions in cities with circular transportation networks. IEEE Int Conf Serv Oper Logist Inform. doi:10.1109/SOLI.2007.4383892

Zhu X, Wilhelm WE (2006) Scheduling and lot sizing with sequence-dependent setup: a literature review. IIE Trans 38:987–1007

Author Biographies

Poorya Farahani is currently a post-doc researcher at the chair of Business Administration—Production and Supply Chain Management at Technical University of Munich. He did his PhD at the Department of Management Engineering, at Technical University of Denmark focusing on distribution management in food supply chains. During his PhD project, Mr. Farahani considers a quantitative approach towards food distribution management and investigates the principal features and challenges of using advanced planning systems in this area.

M. Grunow is the Chair of Business Administration—Production and Supply Chain Management at Technical University of Munich. Previously, he was the head of the Operations Management group at the Technical University of Denmark and leader of the theme "Catering and convenience, freshness and supply chains" in the FoodDTU Research Center. His research interests are in production and logistics management with a focus on supply chain management in the process industries, including food. He has coauthored more than 100 publications amongst others in International Journal of Production Economics, International Journal of Production Research, European Journal of Operational Research, CIRP Annals, and OR Spectrum and acts as an editor for the journals OR Spectrum and Flexible Services and Manufacturing (FSM).

H.-O. Günther is professor (Chair) of Production Management at the Technical University of Berlin (TU Berlin), Germany, since 1993. He has coauthored more than 300 scientific papers amongst them more than 40 papers in refereed international journals in the area of Production and Operations Management, Operations Research and Industrial Engineering and co-edited numerous special issues of journals and several book volumes on container terminal logistics, production and operations management and supply chain planning. Since 2009, he is editor-in-chief of the "Flexible Services and Manufacturing" journal. From 2001 to 2007 he was editor-in-chief of the journal "OR Spectrum". His current research interests include production and supply chain planning, logistics operation management, simulation and optimization of assembly in the electronics and the automotive industry.