European Journal of Operational Research 224 (2013) 240–260

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Invited Review

Planning of capacities and orders in build-to-order automobile production: A review

Thomas Volling *, Andreas Matzke, Martin Grunewald, Thomas S. Spengler

Technische Universität Braunschweig, Institute of Automotive Management and Industrial Production, Katharinenstraße 3, 38106 Braunschweig, Germany

ARTICLE INFO

Article history:
Received 30 June 2011
Accepted 26 July 2012
Available online 8 August 2012

Keywords:
Production
Automotive industry
Planning framework
Capacity planning
Order-driven planning
Operations research

ABSTRACT

In an attempt to achieve a competitive edge, automotive companies operate global production networks to offer an ever increasing product variety, shorter and reliable lead times as well as competitively priced products. Cars are no longer exclusively produced based on standardized product configurations and stable sales plans but are increasingly build-to-order to match the needs of individual customers. Operations Research (OR) may contribute towards successful build-to-order operations. This is likewise reflected by the appreciable number of published papers on industry specific OR applications. To provide readers with an overview about these OR models and applications we identify current and future research issues based on the review of 49 works. We focus on two important planning objects which have not been considered in prior reviews: the planning of capacities and orders. To bridge the gap between conceptual works on the one hand and quantitative contributions on the other, we provide a framework for the structuring of planning tasks. Existing models are classified according to this framework and open issues that should be addressed in OR are discussed.

1. Introduction

Hardly any industry reaches up to the economic relevance of the automotive industry. 10.2% of total employment in the manufacturing sector of the EU is directly linked to automobile production. Across the EU the automotive industry operates 183 vehicle assembly and engine production plants and contributes with approximately 30 billion Euros per year to exports (ACEA, 2010).

Driven by the idea to offer an ever increasing product variety, shorter and reliable lead times as well as competitively priced products, automotive companies have been promoting customization strategies (Holweg and Pil, 2001). Only recently, Audi AG's Head of Sales compared the configuration of cars with the styling of a living room (Brückner, 2010). The main building block of these strategies is the postponement of the point of product differentiation after customer orders are received (Pil and Holweg, 2004). Postponement allows for the efficient production of low variety components which can be integrated flexibly into products that serve individual customers' needs. The corresponding strategy is referred to as build-to-order (BTO) production (MacCarthy et al., 2003).

While being attractive as a competitive strategy, BTO production fundamentally changes the mode of operations. This change is mainly related to the increased exposure to demand variability as a result of variations in the timing and specifications of customer requests, the resulting model-mix, and the total demand per period (Holweg et al., 2005; Waller, 2004).

As a consequence, automotive original equipment manufacturers (OEMs) struggle with two important challenges. Considering strategic and tactical decisions for mid- to long-term planning horizons, a first challenge is related to the identification of the right level of flexibility. To mitigate the increased exposure to demand variability it becomes attractive for OEMs to establish flexible production systems (Reichhart and Holweg, 2007). Flexibility allows for adjusting capacities in correspondence with market demand. However, flexible production systems require additional investments and result in increased production costs related to efficiency losses (Slack, 1993; **?**). To solve this trade-off, OEMs have to take decisions on investments in capacities and the deployment of capacity adjustment options. A second challenge regards short-term operational planning. Driven by the variability of demand it becomes more difficult to compile production plans which allow for an efficient mode of production. In BTO automotive manufacturing this comes down to decisions on the production of customer orders. The disposability of flexible resources supports short-term planning by offering additional degrees of freedom. However, it becomes increasingly complex to assess all possible courses of action in identifying and implementing the optimal choice.

The planning of capacities and orders can effectively be supported by operations research (OR). Throughout the last decades, tremendous advances in terms of modeling, solution approaches and computing capabilities have been reported (Bixby, 2002). As a consequence, there is a growing interest in applying OR

---

* Corresponding author. Tel.: +49 (0) 531 391 2202; fax: +49 (0) 531 391 2203.
E-mail addresses: t.volling@tu-braunschweig.de (T. Volling), andreas.matzke@tu-braunschweig.de (A. Matzke), m.grunewald@tu-braunschweig.de (M. Grunewald), t.spengler@tu-braunschweig.de (T.S. Spengler).

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2012.07.034

techniques to industrial planning problems. OR helps solving problems by structuring decision situations, mastering complexity and providing decision support.

In this paper we provide a comprehensive review of OR models and applications in the field of BTO automobile production. We focus on those planning tasks which are underrepresented yet: the planning of capacities and orders. The review thereby complements prior reviews on selected planning problems and the activity structures/planning systems used in the automotive industry such as Becker and Scholl (2006) (assembly line balancing), Boysen et al. (2009a) (mixed-model sequencing), Meyr (2004) (German OEM) and Fredriksson and Gadde (2005) (Volvo Cars). In contrast to general works on BTO production and operations planning (MacCarthy et al., 2003; Gunasekaran and Ngai, 2005, 2009; Mansouri et al., 2012), special emphasis is given to the specific requirements of the automotive industry. The main contributions of this paper can be summarized as follows. (1) Our work has identified the existing body of literature on the planning of capacities and orders in the automotive industry. (2) We provide a framework for the structuring of planning tasks and identify gaps in the research. (3) Approaches for production and operations planning in the automotive industry are particularly well developed in Germany. The review includes works which have only been reported in German and, in doing so, makes the general ideas of this research assessable to the international community. The paper is structured as follows: The planning framework is introduced in Section 2. Following some notes on the review methodology in Section 3, we use the framework to distinguish two lines of research. These are the planning of capacities (Section 4) and the planning of orders (Section 5). The paper concludes with a discussion of potentials for future work in Section 6.

2. Planning framework and scope of the review

2.1. Spanning the framework

Production and operations planning in the automotive industry covers a wide range of heterogeneous planning tasks. These can be classified according to two dimensions: the planning level and the planning object. In the following a framework will be developed based on these two dimensions. The resulting planning tasks will be discussed briefly to lay the basis for the following review of literature.

We distinguish two planning levels with respect to their scope. (1) *Production networks* result from the horizontal (in-line with the direction of the material flow) and vertical (parallel to the material flow) linking of production sites. Production sites comprise those for the production of cars (vehicle assembly sites) and components (e.g., engine sites). The resulting production networks constitute the internal supply chains of OEMs (2) A second planning level regards *production sites*. Structurally three departments (shops) can be distinguished within vehicle assembly sites: body shop, paint shop and final assembly (shop). Component sites are less generically structured, typically comprising part manufacture, surface machining and (pre-) assembly.

On each level, three planning objects can be distinguished. The object with the most long-term impact regards the planning of *structures*. The focus is on determining the spatial organization of production systems. Given a certain structure, a second planning object regards the production systems' configuration. As a result, *capacities* are set which determine the production capabilities in terms of volume and model-mix. Often the adjustment of these capacities over time is considered as well. Decisions regard the technological infrastructure as well as the workforce and the shift model. The third planning object regards operational decisions on the production program, i.e. the usage of the capacities. For the case of BTO production this relates to the question of how to fulfill individually specified customer *orders*.

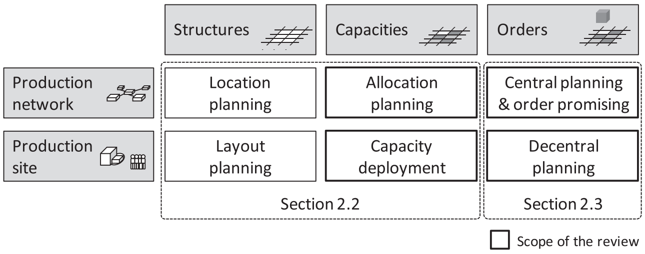


Fig. 1. Planning framework.

The framework which results from these dimensions can be used to identify typical planning tasks in BTO automobile production (Fig. 1). These tasks can be organized hierarchically and become more specific, the shorter the planning horizon. While the planning of structures is part of strategic planning, the planning of capacities is considered tactical and that of orders operational. In the following we will briefly review planning tasks and challenges based on the framework. It should be noted that the framework is focused on an important section of the automotive planning system. It can be easily extended with respect to both dimensions. For instance, production sites can be further differentiated into production lines, with the planning tasks process planning and line balancing (structures), resource planning (capacities) and sequencing (orders). As further objects the planning of budget, material flow and part requirements may be considered at each planning level. In order to avoid redundancy to prior reviews and to keep the analysis tractable this review focuses on the planning of capacities and orders. However, since decisions on structures and capacities are closely interlinked, we will briefly introduce them in an integrative manner (Section 2.2). A separate discussion on the planning of orders is following (Section 2.3).

2.2. Planning of capacities

The planning of capacities is concerned with the determination of production capabilities in terms of volume and model-mix. Given that there are close linkages between the planning of structures and capacities, we will include a brief description of problems related to the planning of structures in the following. This allows for indicating interdependencies and contributes towards an improved classification of literature in the remainder of the review. Against this background, the following characteristic planning problems can be identified:

How should production networks be set up, i.e., where should production sites be established and how should they be logistically connected? The definition of the network structure is subject of *location planning*. When assessing alternative locations, sales opportunities, differences in costs, existing production sites and resulting transport operations need to be considered as well as several international factors such as tariffs, duties, exchange rates and local content restrictions (Goetschalckx et al., 2002). For companies using BTO production, another important factor is the distance between production sites and markets. Long distances result in extended delivery times. As a consequence there is a decreasing share of customers which accepts BTO production. Planning horizons of at least one product life cycle are considered (6–12 years).

Which models should be produced at which location in which volume? Based on the network structure, *allocation planning* is concerned with defining the network configuration. This entails decisions on the assignment of products to production sites and the corresponding installation of capacities (Fleischmann et al., 2006) as well as the definition of supply and delivery structures (Escudero et al., 1999). Planning decisions influence both, mid-term investments as well as operational flexibility. Typically three network configurations are differentiated (Jordan and Graves, 1995): the fully flexible assignment, where each product can be produced at each site, the exclusive assignment, where each product is dedicated to exactly one site, and partially flexible strategies which comprise any combination of the first. Amongst the last class, chaining strategies are of particular interest. When *chaining* is applied, each product is assigned to two sites and each site to two products. As a result, an interlinked assignment structure (a spanning tree in graph theory) results which resembles a chain. The planning horizon is similar to location planning and covers at least one product life cycle.

How should production sites be physically and organizationally structured? The problem of *layout planning* regards the definition of each production site's infrastructure. For vehicle assembly sites this in particular entails the setup of body shop, paint shop and final assembly lines. Decisions regard the design and configuration of buildings and transportation infrastructure (Bundschuh, 2008). Starting point are the results of allocation planning, i.e., information on the product portfolio to be produced, and the existing infrastructure. Typical planning horizons cover one product life cycle.

How should each production site's capacity be deployed? Given a certain product allocation and maximum capacities, *capacity deployment* regards decisions on the factual capacity which is to be supplied at each production site (Sillekens et al., 2011). Flexibility instruments to adjust capacities focus on changes to the production rate/takt time and the shift model. Both instruments directly relate to workforce requirements. Often working hour accounts are employed to further increase labor flexibility. Since body and paint shops are highly automated, the focus here is on technological flexibility instruments (Askar et al., 2007). These include the activation and deactivation of parallel production stations or lines and the use of empty work piece carriers, which allow for adjusting the production rate without any consequences for line balancing. Typical planning horizons are motivated by agreements with labor unions (e.g., 1–2 years).

The planning of structures and capacities in BTO automobile production is characterized by three main challenges. These can be summarized as follows:

- *Large networks*: Almost all automotive OEMs operate globally dispersed production networks to serve global demand. Designing and operating these networks requires incorporating an immense data volume and managing complex logistical interdependencies (Fleischmann et al., 2006). In the automotive industry, special attention is given to local content requirements and the specific characteristics of production processes (e.g., decoupling points, scale effects).
- *Uncertainty*: Due to the variety of the product offering and pronounced market dynamics, the automotive industry in general and BTO automobile production in particular, are characterized by a high level of demand uncertainty (Reichhart and Holweg, 2007; Fredriksson and Gadde, 2005). This uncertainty imposes complex challenges on the evaluation and selection of mid and long-term plans.
- *Modeling of capacity*: The production output of automotive OEMs is subject to a huge number of heterogeneous constraints (Sillekens et al., 2011; Boysen et al., 2009b). These result from operational characteristics of the final assembly stage as well as up- and downstream activities. Due to flexibility most constraints are not strictly binding. Any violation will however result in financial implications, which have to be considered when planning.

2.3. Planning of orders

The planning of orders in BTO automotive production is concerned with managing the OEMs interface with its customers while ensuring efficient production in the short-term. This entails decisions on how to match customer demand for specifically configured products with the supply chain's capacities on the network and site level.

The discussion of tasks related to the planning of orders is challenged by the fact that merely all automotive OEMs which rely on BTO strategies make use of hybrid order fulfillment systems. The baseline strategy of these systems is the postponement of production activities until customers have placed their orders. As a consequence, demand uncertainty is resolved before production activities take place. Production can thus be best directed towards market needs. Yet, BTO production requires customers to accept delivery times which result from production lead times. These typically range between a few weeks and several months in the automotive industry (Rahn, 2011). In order to serve those customers that are not willing to wait, virtually all OEMs employ a second order fulfillment strategy in parallel to BTO (Brabazon and MacCarthy, 2006). In this case, production activities are started before customers place their orders following the principle of build-to-stock (BTS) production. In Europe, where most OEMs pursue BTO strategies, roughly 50% of all new cars are sold from stock. In Germany the number is smaller but still exceeds 30% (Williams, 2000).

The co-existence of both order fulfillment strategies is reflected by the planning systems of automotive OEMs. It is therefore not possible to completely disregard BTS order fulfillment when discussing the planning of orders. Against this background, we will thus include those BTS related planning tasks into the analysis which interface with BTO order fulfillment. Overall we distinguish six planning tasks (Fig. 2). These result from the two dimensions production progress (time) and informational basis (order fill rate).

The *production progress* of the customer orders produced within a single planning period (e.g., 1 day) is described by the milestones "order freeze", "end of production" and "delivery". All planning decisions prior to "order freeze" are of preliminary nature. With "order freeze" production orders are released to the production sites for all customer orders which fall into the considered planning period. This typically involves fixing the product configuration. The production sites are in the following responsible for fulfilling the production orders within the defined planning period. With the milestone "end of production" the production orders are completed and the finished cars are made available for distribution. The milestone "delivery" marks the point in time, at which the last car of the planning period is handed on to the customer.

To model the OEMs *informational basis* with respect to the considered planning period, we use the concept of the order fill rate (Meyr, 2004). The order fill rate quantifies the share of capacity which is filled by customer orders as a function of time. The fill rate increases, as more and more customers place their orders. *BTO* production is limited to the share of customer orders which is placed before "order freeze". Any other cars which are intended for production in the considered planning period are produced *BTS*. For these cars the product configuration has to be decided on in anticipation of future customer demand.

Against this background, the following planning problems result. The processing of customer orders and their transformation into production orders which are transferred to the production sites is concerned by order promising and central planning. If BTS production is used in parallel to BTO, order promising is characterized by an extended functionality. This functionality includes (a) the generation of production orders for those cars for which no customer order is known at "order freeze" and (b) the search for

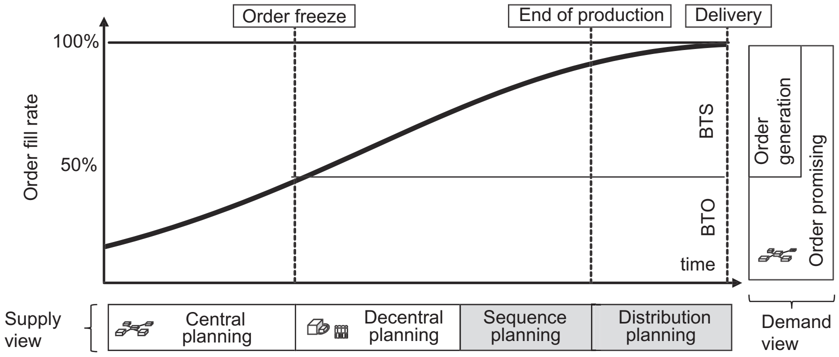


Fig. 2. Planning tasks: the planning of orders.

BTS cars which are suitable to serve incoming customer requests. Following "order freeze" decentral planning and sequence planning are executed before production. The planning of the subsequent transportation operations is subject to distribution planning. Given these planning tasks, a match of supply and demand can be obtained from integrating order promising (demand view) with the other planning tasks (supply view).

With respect to the framework developed in Section 2.1 we distinguish two planning levels. The planning tasks order promising (including order generation) and central planning regard the alignment of demand and supply for final products. In order to coordinate different sales regions and production sites, planning is done on the network level. In contrast, decentral planning generates production plans for each production site.

A brief description of the problems related to the planning of orders is given in the following. Sequence and distribution planning are not included into this review. Please refer to Boysen et al. (2009a) and Holweg and Miemczyk (2002) for details.

How shall customer requests be fulfilled? Incoming customer requests are processed via *order promising*. The objective is to determine delivery dates for individually specified customer requests. Order promising comprises two sequential tasks. In a first step, the set of fully specified BTS cars is searched for a suitable product to match the customer request. This planning task is referred to as available-to-promise check (ATP) or locating. If no suitable match is found, a second task, the capable-to-promise (CTP) check, is triggered. A new production order is generated for the customer request and inserted into the central production plan to the projected due date. This requires knowledge on capacities and therefore linkage with central planning. As a result of order promising, customers are returned a promised delivery date for their particular request (Volling and Spengler, 2011).

Which orders are to be produced in anticipation of customers requiring short lead times? To include those cars into the production plan, which are BTS, it is necessary to anticipate future demand for specified cars and to generate production orders for these cars before order freeze. The specification of these BTS production orders is the subject of *order generation*. Depending on the business model, order generation may be in the responsibility of the OEM's central sales organization or in that of intermediaries within the distribution chain (e.g., regional sales offices, local dealers). Order generation is characterized by the conflicting priorities of diversified regional market needs and the capabilities of the production system. This requires close coupling with order promising. Often both tasks are treated in an integrative manner. If the anticipation does not meet the expectation of the market, cars have to be held on inventory until a suitable customer has been identified. As a consequence inventory costs result and discounts may have to be warranted in order to compensate for product specifications which deviate from customers' expectations (Holweg and Pil, 2001). Both, inventory costs and discounts make up very significant cost in the sale of BTS cars (Waller, 2002).

Which orders are to be produced, when, at which production site? The problem of *central production planning* regards the alignment of capacities and production orders on the network level. This includes the determination of production sites and periods (e.g. weeks) for each order (Meyr, 2004). As a result, production plans are set for each production site. In addition to production, procurement and logistics costs, central planning has to incorporate service criteria. These become relevant, if projected delivery dates deviate from the initial promise (Volling and Spengler, 2011). The feasibility of production plans is subject to capacities, which result from capacity deployment. Updated information on available capacities is supplied to order promising while the resulting production plans are transferred to decentral planning. The planning horizon is determined by order lead times and covers 3–12 months.

Which lines should production orders be assigned to? The results of central planning are further detailed by *decentral planning*. For each production site the objective is to determine production plans for each shop/production line. The results are subsets of production orders which are to be produced in a certain period (e.g., shift) in a shop/on a line (Boysen et al., 2009b). This requires the synchronization of production orders for multiple production stages/shops. The objective of decentral production planning is to minimize production, procurement and due date related costs. The results are submitted to sequence planning. Planning horizons are related to procurement lead times and may cover 4 weeks.

In addition to the challenges introduced above, the most important challenges for the planning of orders in BTO automobile production are:

- *Informational dynamics*: Customer orders approach OEMs in a stochastic process (Meyr, 2004; Brabazon et al., 2010). On the network level, neither the total number, nor the specification, nor timing and destination of future orders are perfectly known, when planning is executed. This results in an open decision field and the requirement to anticipate future demand.
- *Complexity*: Production planning for individually specified orders requires solving large assignment problems for production networks and lines, typically involving several thousand planning objects (Bolat, 2003). This calls for efficient solution approaches.
- *Multi-criteria decision situation*: For short-term planning decisions, the monetary assessment of all relevant aspects is often not possible. Accordingly, planning has to incorporate non-monetary indicators (Volling and Spengler, 2011). This gives rise to the question of how to integrate potentially conflicting objectives.

- *Variations in the model-mix*: Variations in the model-mix impose challenges on the supply chain and the production system. Automotive OEMs agree on procurement quotas with their suppliers. In order to account for fluctuations in the demand for parts, some flexibility is associated with these quotas. However, this flexibility is contractually bounded. Violations of these contractual agreements may not be feasible at all or result in penalties. Similarly, production systems in the automotive industry are designed to manufacture a certain mix of cars (Boysen et al., 2009a). This mix is constraint by the takt time, i.e. by the time between the completion of two cars. Even though operating times differ in accordance with the order specification, producing at a constant takt time is possible, if operating times level out. Planning therefore has to make sure that cumulated operating times associated with a certain mix of orders do not exceed the available capacity. But even if production plans are feasible, planning should avoid inefficiencies which result from under- or over-loading capacity.

3. Review methodology

The review is based on the iterative searching procedure illustrated in Fig. 3. We first browsed the bibliographic database Scopus (http://www.scopus.com). Scopus covers the relevant journals in the fields of management science, production and operations management, operations research and supply chain management as well as related research areas. Starting with a keyword search, the results were manually filtered based on title and abstract review. We focused on quantitative approaches for the planning problems discussed in Section 2. In addition to journal articles we considered selected publications in conference proceedings. The results were complemented by German dissertations, which are well known in the German automotive research community but have not been published internationally, yet. In order to make the principle ideas behind these works assessable to the international community, we included them into the review. Additional publications were identified iteratively based on the analysis of the full texts and the reference lists of the identified works.

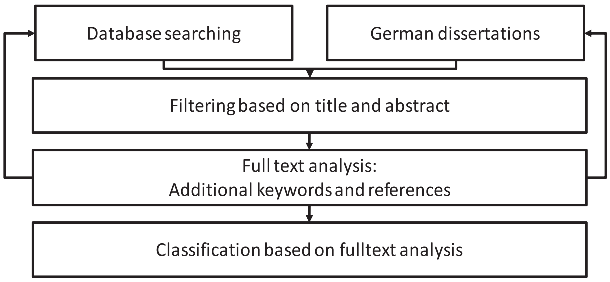


Fig. 3. Review methodology.

The literature search yielded 49 references, covering a time frame of 23 years. We distinguished two groups of works for which separate classification schemes were developed. In total 31 publications were found on capacity planning (group 1) and 17 publications on the planning of orders (group 2).

Fig. 4 reveals the number of publications per year. Most publications were released during the last decade (82%). 67% of the works have been published in peer-reviewed journals. The 32 journal articles were further analyzed with regard to their journal as depicted in Fig. 5.

4. Review of approaches for the planning of capacities

4.1. Classification scheme

In the following a classification scheme is developed to show similarities and differences in the existing approaches for the planning of capacities (Table 1). The attributes used for the classification are grouped into seven categories. An approach is described by the selection of values for each attribute.

According to the scope of the review, we distinguish the *planning tasks* allocation planning and capacity deployment. Some works explicitly consider aspects of location and layout planning. We added attributes to indicate these interfaces. *Objective functions* differ with respect to two characteristics. With respect to the objective, we differentiate approaches based on profits (p) or costs (c) vs. cash flows. When cash flows are modeled, we further distinguish approaches which seek to maximize the net present value (NPV) and those which minimize discounted cash outflows (dcf). We further indicate whether multiple criteria (mc) or other operational measures (o) are considered. Detailed information is given in the presentation of each work. Some works complement the analysis of the (expected) value of the objective function by assessing its frequency distribution depending on some stochastic parameters. To indicate this, we incorporate the attribute risk. For the differentiated assessment of the *model type*, we include seven attributes. The level of planning is indicated by the first attribute. We distinguish three levels: strategic (s), tactical (t) and operational (o) planning. In terms of the time scale we distinguish static (s) and dynamic (d) approaches. These are based on different programming techniques, namely deterministic (d) and stochastic (s) programming. In terms of the structure of the production system we differentiate single lines (sl) and multiple parallel lines (pl) as well as multi-level (ml) and single-level (sl) production processes. In order to contrast approaches in terms of their scope, we added the planning horizon (years (y), quarters (q), months (m), weeks (w)) and the granularity of time (years (y), quarters (q), months (m), weeks (w), days (d)). With respect to the *modeling of demand*, we identified four attributes. Approaches differ in terms of the information quality used for the model evaluation (deterministic (d), stochastic (s), selected scenarios (ss) or fuzzy (f)), the number of product variants (single product (sp), multiple products (mp)) and fulfillment constraints. The most common fulfillment

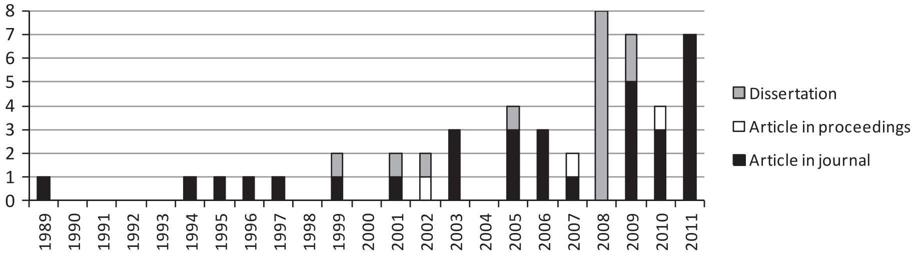


Fig. 4. Number of publications by year and type.

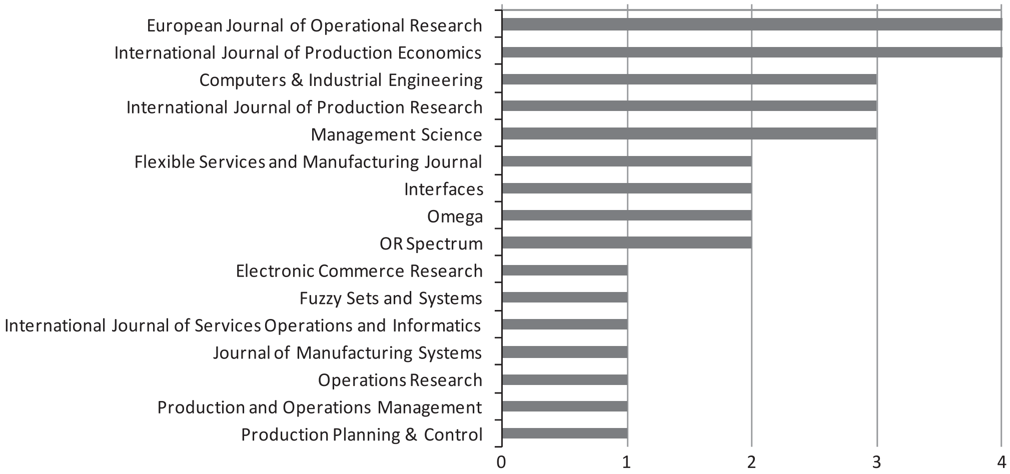


Fig. 5. Number of publications by journal.

constraint is that demand has to be fulfilled completely (f). Other approaches introduce additional degrees of freedom. Customer requests are either allowed to be denied, which leads to a selection decision (s), or lead time flexibility may be used to delay production (d).

A key feature of the reviewed works regards the modeling of *technological capacities*. These determine the maximum production output. In terms of the level of aggregation, approaches can be distinguished, which quantify capacity in units (u) and more universal ones which refer to product specific capacity requirements (cr). The approaches differ with respect to the shops considered in the models. In accordance with the general structure of automotive plants, three shops are of relevance: final assembly, paint shop and body shop. If no differentiation is being made with respect to the shops, we assume that final assembly capacity is regarded. Further, the capacity of the supply chain and the pre-assembly of modules or components may be integrated. Since technological capacity adjustments are typically followed by a ramp-up phase, learning curves are considered in some approaches. In addition to that, productivity losses may be relevant, if more than one model is produced on the same production line. Mid-term flexibility measures are available to further adjust the technological capacity in accordance with demand. These measures determine the *organizational capacity*, i.e. the factual capacity which is available for production. For the most part, these instruments are related to labor. Decisions include the number of workers, the shift model and the balance of working time accounts. Further measures are aggregated into the category "else". Examples include the deactivation of stations or lines and the usage of empty work piece carriers (Section 2.2). Lastly, decisions on capacities are subject to a set of *international factors*. We differentiate exchange rates, taxes, tariffs and local content requirements. In addition to that, there may be regional differences in terms of productivity.

4.2. Problem classification

Table 2 gives a comprehensive literature survey of approaches to support the planning of capacities. If a unique classification of a paper is not possible according to the given information, the best fit of classifying characteristics is taken. The classification exclusively covers quantitative research dealing with allocation planning and capacity deployment. Not covered are purely qualitative studies such as Alden et al. (2002) or Breitman and Lucas (1987) and studies which regard the automotive industry or the defined planning tasks peripheral (Kohler, 2009; Koether, 1986).

In total 31 works are considered, 22 of which are refereed journal articles. Five works have only been published in German dissertations. Proof of the topic's recency is the fact that more than half of the works have been published within the last four years. 24 of the 31 studies regard industrial cases or are at least motivated by a specific case. Amongst the automotive manufacturers are Daimler (12 studies), GM (5), BMW (5), Ford (1) and FAW-Volkswagen (1). Different modeling techniques are used, including deterministic, stochastic and fuzzy programming. Risk is considered explicitly in five works. All studies regard multiple products and aggregate capacity in some way.

The analysis reveals three streams of literature (Table 3). *Integrative approaches* seek to provide comprehensive planning models for the long-term planning of automotive supply chains. The idea is to develop deterministic mixed integer linear programming (MILP) formulations which integrate aspects of several planning tasks. Approaches with the focus on *flexibility planning* regard allocation decisions over planning horizons spanning several years. Often aspects of capacity planning are considered simultaneously. The idea is to identify model assignments, which prove robust with respect to stochastic influence. As such, all studies incorporate stochastic information in some way. Works on *detailed capacity planning* seek to provide deterministic models and solution approaches, which very specifically reflect the planning situation of single production sites. Specific requirements comprise working time accounts and multi-level capacities. In the following, we briefly review the relevant papers according to these groups.

4.3. Integrative approaches

Integrative planning approaches consider mid- to long-term planning horizons (6–12 years) on a high level of aggregation (typically yearly planning periods). The planning problem is of global scale. The focus is on providing comprehensive model formulations, considering multiple levels of the production process and parallel lines. To keep the analysis tractable, the incorporation of stochastic influences is restricted to scenario analysis. I.e., deterministic optimization models are developed which are solved for multiple scenarios. All works make use of (mixed integer) linear programming (MILP/LP) solvers. The only problem specific solution approach is reported in Kauder (2008).

An early work on the integrative planning of capacities is Inman and Gonsalvez (2001). In the paper the problem of allocating car models to production sites and lines is considered. The idea is to differentiate two sub-models. In a first step a MILP model is

Table 1
Planning of capacities: classification scheme.

Attribute	Description
1. Planning task	
Location	Modeling of location planning decisions
Allocation	Modeling of allocation planning decisions
Layout	Modeling of layout planning decisions
Capacity	Modeling of capacity depolyment decisions
2. Objective function	
Objective	Planning objective: profits (p), costs (c), NPV (npv), discounted cash outflows (dcf), multi-criteria (mc), other (o)
Risk	Incorporation of risk
3. Model type	
Planning level	Strategic (s), tactical (t), operational (o)
Time	Time scale: static (s), dynamic (d)
Programming technique	Deterministic (d), stochastic (s), fuzzy (f)
Vertical interdependencies	Planning scope: single line (sl), parallel lines (pl)
Horizontal interdependencies	Levels incorporated into planning: single-level (sl), multi-level (ml)
Horizon	Planning horizon: years (y), quarters (q), months (m), weeks (w)
Granularity	Granularity of planning decisions: years (y), quarters (q), months (m), weeks (w), days (d)
4. Modeling of demand	
Uncertainty	Information quality: deterministic (d), stochastic (s), selected scenarios (ss), fuzzy (f)
Product	Number of products: single product (sp), multiple products (mp)
Fulfillment	Fulfillment constraints: fulfillment (f), delay (d), selection (s)
5. Capacity (technological)	
Aggregation	Level of aggregation: units of final products (u), product dependent capacity requirements (cr)
Assembly	Incorporation of assembly capacity; aggregate car production capacity if departments are not distinguished
Paint shop	Incorporation of paint shop capacity
Body shop	Incorporation of body shop capacity
Component	Incorporation of component production capacity
Supply chain	Incorporation of supply chain capacity
Learning curves	Incorporation of learning curves
Productivity	Modeling of productivity losses due to mixed-model production
6. Capacity (organizational)	
Workers	Decisions on the number of workers
Shift	Decisions on the shift model
Accounts	Modeling of working time accounts
Else	Modeling of further instruments to adjust capacity
7. International factors	
Exchange rates	Modeling of exchange rates
Taxes	Modeling of taxes
Tariffs	Modeling of tariffs
Local content	Modeling of local content restrictions
Productivity	Modeling of regional differences in productivity

proposed to determine product allocation decisions which are optimal with respect to lost sales. Given these results, in a second step a capacitated network flow formulation is used to distribute production volume to production sites such that the capacity utilization is balanced. Model features include product introductions and phase-outs as well as tooling limitations and changeovers. Capacities are considered with respect to plants, multi-model body shops and dedicated final assembly lines. Since monetary information is not incorporated, lost sales are valuated equally for all models. Investments necessary to facilitate product allocations are not considered; neither are international aspects. The work is based on the case of GM.

Motivated by the case of BMW, the global allocation of car models to production sites is considered in a series of works. The series goes back to the work of Henrich (2002). A high level MILP is developed based on the idea of a capacitated warehouse location problem. Dependent on whether capacities are scarce or not, the objective is to either maximize profits by selecting the most profitable demand or to minimize costs fulfilling all demand. Differentiated cost rates are incorporated for procurement, production and distribution. The capacities of final assembly, paint shop, and body shop are defined by the maximum output (number of cars) per year. Exchange rates, tariffs and local content restrictions are included into the model. The model supports location planning, if new potential locations are added to the set of sites.

An extension of the work of Henrich (2002) is presented in Fleischmann et al. (2006). In order to account for inter-temporal effects, discounted cash-flows are regarded in the objective function. The objective is to minimize the discounted sum of structural and product-specific investments as well as operating costs, given a demand forecast which has to be met. Technological capacities for each shop are chosen from a discrete set of pre-defined levels. The model includes aggregate capacity deployment decisions (overtime). Flexibility is implicitly modeled by means of safety factors, which are used to reserve a certain share of the nominal capacity to react to swings in demand. Time profiles for structural and product-specific investments are included into the model. International factors are considered in terms of exchange rates, tariffs, a simplified tax model and local content restrictions.

Kauder and Meyr (2009) further extend the series of works by introducing additional constraints which assure that only solutions are considered which follow the principle of chaining. Based on different assumptions regarding demand and exchange rates, seven scenarios are considered in a numerical case study. The authors conclude that while more expensive in deterministic settings, chaining effectively contributes towards improved performance when facing uncertain demand.

A problem specific heuristic for the model proposed in Kauder and Meyr (2009) is developed in Kauder (2008). This heuristic is based on problem decomposition. An integer programming (IP)

Table 2
Planning of capacities: classification.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
	Inman and Gonsalvez (2001)	Henrich (2002)	Fleischmann et al. (2006)	Kauder and Meyr (2009), Kauder (2008)	Bundschuh (2008)	Gneiting (2009)	Eppen et al. (1989)	Chandra et al. (2005)	Friese (2008)	Bihlmaier et al. (2009)	Stephan et al. (2010)	Francas et al. (2011)	Escudero et al. (1999)	Peidro et al. (2009), Peidro et al. (2010)	Zhang et al. (2011)	Kabak and Ülengin (2011)	Jordan and Graves (1995)	Boyer and Leong (1996)	Graves and Tomlin (2003)	Francas et al. (2009)	Inman and Jordan (1997)	Benyoucef et al. (2007)	Askar et al. (2007), Askar (2008)	Sillekens et al. (2011), Sillekens (2008)	Walter et al. (2011)	Roscher (2008)	Garcia-Sabater et al. (2011)
1. Planning task																											
Location	☐	☑	☑	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Allocation	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☐	☐	☐	☐	☐	☐	☐
Layout	☐	☐	☐	☐	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Capacity	☑	☐	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☐	☐	☐	☐	☐	☐	☑	☑	☑	☑	☑	☑	☑
2. Objective function																											
Objective	mc	c/p	dcf	dcf	dcf	npv	npv	npv	npv	dcf	npv	p	c/mc	c	p	p	o	o	o	o	o	n/a	dcf	c	dcf	dcf	mc
Risk	☐	☐	☐	☐	☐	☐	☑	☐	☑	☐	☐	☐	☐	☐	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐
3. Model type																											
Planning level	s/t	s	s	s	s	s	s/t	s/t	s/t	s/t	s/t	s/t	t/o	t/o	t/o	s/t	s	s	s	s	t/o	t	t	t	t	s/t	t/o
Time	d	d	d	d	d	d	d	s	d	d	d	s	d	d	d	s	s	s	s	d	s	d	d	d	d	d	d
Programming technique	d	d	d	d	d	d	s	s	s	s	s	s	s	f	s/f	f	d	d	s	s	d	d	d	d	d	d	d
Vertical interdependencies	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	pl	sl	pl	ml
Horizontal interdependencies	ml	ml	ml	ml	ml	ml	sl	ml	ml	ml	sl	sl	ml	ml	ml	ml	sl	sl	ml	sl	sl	sl	ml	ml	sl	sl	pl
Horizon	6y	12y	12y	6y	10y	6y	5y	5y	10y	7½y	15y	n/a	m/y	1-2y	n/a	10y	n/a	n/a	n/a	6y	n/a	2y	1-7y	½-2y	2y	10y	6m
Granularity	q	y	y	y	y	y	y	y	y	½ y	y	n/a	w/q	w/m	n/a	n/a	n/a	n/a	n/a	y	d	m	w/m	w	w	m	w
4. Modeling of Demand																											
Uncertainty	d	ss	ss	ss	ss	ss	s	s	s	s	s	s	s	f	s/f	f	s	s	s	s	d	d	d	d	d	s	d
Product	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	mp	sp	mp	mp
Fulfillment	s	f/s	f	f/s	f	s	s	f	s	f/s	s	s	s	d	s	s	s	s	s	s	f	d	d	f/d	d	d	d
5. Capacity (technological)																											
Aggregation	u	u	u	u	cr	cr	u/cr	u	cr	cr	cr	cr	cr	cr	cr	cr	u	u	u	cr	cr	cr	u	u	cr	cr	u
Assembly	☑	☑	☑	☑	☐	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☐	☑	☑	☑	☑	☑
Paint shop	☐	☑	☑	☑	☐	☐	☐	☑	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑	☐	☐	☐
Body shop	☑	☑	☑	☑	☐	☐	☐	☑	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑	☐	☐	☐
Component	☐	☐	☐	☐	☑	☑	☐	☑	☐	☑	☐	☐	☑	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑
Supply chain	☐	☐	☑	☑	☑	☑	☐	☑	☑	☐	☐	☐	☑	☑	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑
Learning curves	☑	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☑	☑	☐
Productivity	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☑	☑	☐
6. Capacity (organizational)																											
Workers	☐	☐	☐	☐	☑	☑	☐	☐	☐	☑	☐	☑	☐	☐	☐	☑	☐	☐	☐	☐	☑	☑	☑	☑	☑	☑	☐
Shift	☐	☐	☑	☑	☑	☑	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑	☑	☑	☑	☑
Accounts	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑	☑	☑	☐
Else	☑	☐	☐	☐	☑	☑	☐	☑	☑	☐	☐	☑	☑	☑	☐	☐	☐	☐	☐	☐	☐	☑	☑	☑	☑	☑	☐
7. International factors																											
Exchange rates	☐	☑	☑	☑	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Taxes	☐	☐	☑	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Tariffs	☐	☑	☑	☑	☑	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Local content	☐	☑	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Productivity	☐	☐	☐	☐	☑	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Table 3
Streams of literature.

Stream of literature	Number of sites	Primary planning task(s)	Modeling of demand	Works
Integrative approaches	Multiple	Allocation, capacity	Deterministic	1–6
Flexibility planning	Multiple	Allocation	Stochastic	7–20
Detailed capacity planning	Single	Capacity	Deterministic, stochastic	21–27

model is developed to determine a chaining structure for each planning period. The objective is to minimize the sum of fix product-specific investments and anticipated structural investments. Given this solution, a separate MILP is used to compute the optimal allocation of production volume to sites and the associated capacity adjustments. The objective is to minimize the sum of variable product specific investments, structural investments and operating costs. To avoid infeasible solutions, the capacity constraints of the second model are relaxed. Penalty costs are encountered, if capacity is exceeded. The heuristic relies on two steps. In a first step, the IP is solved sequentially for every period of the planning horizon. This is achieved based on the combination of a regret heuristic, which assigns car models to production sites, and a chaining heuristic, which constructs chaining structures (minimal spanning trees) from these assignments. For the latter, the approach of Dijkstra (1959) is modified such that it preferably chooses multi-site assignments for high volume models and avoids investments in body shop capacity. This is achieved by interpreting assignment costs as the sum of fixed product specific investments normalized with respect to product demand and penalty costs in case of additional body shop capacity requirements. The MILP is solved using a standard MILP solver. In a second step, a randomized Neighborhood Search with Threshold Accepting is used to improve the results of the first stage. For each candidate neighbor solution of the IP, the MILP is evaluated. To improve performance, the MILP is only solved, if the solution of the relaxed MILP meets the acceptance criteria of Threshold Accepting. While smaller problems are better solved using a MILP solver, problems of practical relevance can only be solved using the suggested heuristic.

The long-term planning of capacities for the production of engines, chassis and power trains in global production networks is regarded in Bundschuh (2008). The idea is to provide a comprehensive framework for the modeling of planning tasks on different scale (network, site) and scope (location, allocation, layout, capacity deployment). The objective is to minimize cash outflows associated with investments and operations. In this, investments are adjusted by a residual value at the end of the planning horizon. Demand for multiple levels of the bill-of-material (BOM) is considered. The most distinct features of the approach are the detailed modeling of capacity, labor requirements and floor space constraints. Towards this end, a hierarchical production model is developed. Production sites are structured in production systems, which may for instance be organized as cells or lines. Each of these systems is defined by a set of machines, the number of which may be adjusted over time. The gross capacity of a system with respect to a production task results from the working time of the chosen shift model multiplied by the number of parallel machines required for that tasks. Linearization techniques are used to improve the computational tractability of the resulting term. Based on the production model, the approach simultaneously assesses decisions on the de-/installation of production systems, technology choice (selection of production systems and machines), technological capacity (number of parallel production systems and machines) and organizational capacity (shift model). The number of direct workers is determined for every production system based on the shift model, the production volume and various (proportional) influence factors including individual working times, productivity, breaks, absenteeism and allowance times. The number may not fall below a lower boundary which is fixed for every shift model. Three kinds of indirect labor requirements are distinguished. These are derived proportionally to one of the factors production volume, number of operated machines and production area. Floor space requirements are derived from the number of machines per production system and their specific floor requirements. To assure consistency with layout planning, space floor is reserved during installation according to the maximum number of parallel machines which is to be added in later planning periods. Depending on the planning situation, three levels of aggregation are suggested. In addition to the detailed production model described above, a semi-aggregate model is proposed for which machines are grouped to pre-defined production systems. Capacity may be adjusted by choosing the kind and number of production systems and the shift model. For the aggregate model only the kind of production system and shift model remain as decision variables. The latter results in a model formulation that has similarities to the works discussed above. Case studies from BMW are presented to illustrate the application of the approach. A first study is on global location and allocation planning, two further are on capacity deployment for one production site. It is shown that the model provides similar solutions compared to those from conventional manual planning techniques used in practice.

The most recent integrative model has been proposed by Gneiting (2009). A MILP for the long-term planning of capacities is developed. The objective is to maximize the NPV of cash in- and outflows. Unfulfilled demand results in lost revenues and additional penalty costs for losses in goodwill. The baseline model is comparable to that of Bundschuh (2008), yet on a higher level of aggregation. The most special features are two additional models for logistics and pre-assembly planning. Starting point of the logistics model are inter- and intra-plant transportation demands derived from the baseline model. The model distinguishes direct transport and indirect transports which are decoupled by a warehouse. A direct transport is possible, if the supplying production process is capable of providing products in the required sequence. Otherwise, a warehouse with sequencing capability is required. For each transportation demand the model determines the mode of transportation (direct/indirect) and the carrier (e.g., railway, truck, forklift truck). The model also determines the configuration of the logistics network, i.e., the capacity of the carriers, the installation of warehouses and whether sequencing is done at the warehouse. Aggregate floor space constraints and lead times are considered. The model may either be integrated into the master model or solved sequentially. The pre-assembly model is based on the idea of line balancing. A high level of detail is chosen. This includes direct and indirect labor requirements, technology choice, floor space constraints and operations which span multiple stations. Two model variants are suggested, depending on whether the pre-assembly is coupled to the main assembly line or not. For the case that the lines are coupled, it is assumed that the production sequence is determined by the main assembly line. To assure that any production sequence is feasible, the most complex product must be producible within the station boundaries. If the line is decoupled, station boundaries may overlap. The pre-assembly model is solved successively or iteratively with the master model. A case study from a German OEM (Daimler) is presented to validate the application of the approach. The analyzed case is rather small, a benchmark is missing. Solution times are not reported. A sensitivity analysis shows that variations in demand, penalty costs, interest rate and wages influence the results considerably.

4.4. Flexibility planning

Flexibility planning has received considerable attention in the automotive industry. The planning horizon is similar to that of integrative approaches. Information on uncertainty is incorporated based on distribution functions and fuzzy numbers. For the non-fuzzy approaches, sampling is used to reduce complexity. The central questions related to flexibility planning are the determination of split-models and multi-model production lines. The former relates to the definition of models which are assigned to more than one production site. While this redundancy is related to higher investment and operating costs, it allows for production volume to be shifted and thus increases (mix and volume) flexibility. Similarly, multi-model lines require higher investments and are characterized by a productivity which is lower than that of dedicated lines. However, multi-model lines allow for adjusting the model-mix in accordance with demand (mix flexibility) (Sethi and Sethi, 1990).

Three streams of literature can be distinguished. A first may be understood as stochastic version of the integrative planning models presented above. The objective is to evaluate decisions on capacities in the context of uncertain influences. A more short-term perspective is regarded in papers of a second stream. The objective is to determine supply and delivery structures in the supply chain. The network configuration is assumed to be fixed, such that investment decisions can be disregarded. In a third stream, high level models are used. The objective is to evaluate the value of flexibility from a strategic point of view. The costs of achieving this flexibility are not considered.

Six works may be distinguished in the *first stream* of literature. These are presented in the following. A two-stage stochastic MILP with recourse is developed by Eppen et al. (1989). Products are allocated to production sites. Price and demand are regarded as uncertain. On the first stage, tactical investment decisions on capacities are made for each model and production site. Similar to the integrative models discussed above, capacity is chosen from a discrete set of pre-defined levels. On the second stage, operational decisions on production volume are made. To this end, a capacitated network flow problem is solved, which incorporates information on the particular demand realization. Investment decisions are feasible in any period but only once per production site. Demand may be transferred to similar products or is lost, if capacities do not allow for this transfer. The most special feature of the approach is the incorporation of risk. A downside risk measure is introduced to constrain the expected deviations to a profit target. The model is based on a high level capacity representation and therefore does not capture a series of relevant characteristics of the automotive industry. It is, however, the first work to emphasize on the importance of incorporating uncertainty into the capacity planning of automotive companies. A MILP solver is used. The case of GM is considered.

Chandra et al. (2005) regard the investment in flexibility enablers like supplier base, capacity adjustments and common parts capacity. The model is formulated as a non-linear stochastic program and is solved using simulation based optimization. The objective is to maximize the expected NPV. The approach is unique in terms of the incorporation of marketing cost. These are given as a non-linear function of actual demand and forecasted demand. The authors develop a differentiated capacity model, which distinguishes three kinds of capacity. The maximum number of cars per production site is further detailed by the maximum number of models and the capacity of the supply base, which may be in-house or external. In addition to that, in-house parts supply is separated in unique parts, which require dedicated resources, and common parts, which can be produced using the same resources. As such, pre-assembly, body and paint shop may be modeled. Several details such as overtime, maintenance, tooling and the compliance with fuel efficiency standards are considered. Numerical examples from Ford are given to illustrate the applicability of the approach.

In Friese (2008) a two-stage stochastic MILP with recourse is developed for the global planning of capacities. Uncertainty is represented by demand scenarios. On the first stage, products are allocated to production sites. This allocation is static for the planning horizon and independent of the demand realization. For each period of the planning horizon, the second stage considers decisions on the actual capacity level as well as production and transportation volumes dependent on the demand scenario. Capacity is chosen from a discrete set of pre-defined levels. An additional summand is incorporated to capture the capacity flexibility of transferring workers between production lines. The objective is to maximize the expected NPV. A case study for engine production is presented to validate the applicability of the approach. A commercial MILP solver is applied. The work is synchronized with a German OEM (Daimler).

A fourth model within the stream is developed in Bihlmaier et al. (2009). The authors propose a two-stage stochastic MILP to support product allocation planning and capacity deployment under uncertain demand. The model extends the work of Santoso et al. (2005) with respect to the requirements of the automotive industry. Some typical features of integrative models are included. These comprise multiple production stages, learning curves, transportation operations, and productivity losses due to multi-model production. The most special feature is the linear approximation of workforce and shift planning when evaluating alternative capacity plans. The objective is to minimize the expected sum of cash outflows. To solve the model, an accelerated Benders decomposition approach is presented. The number of split-models and the level of process flexibility result from the model solution. The approach is applied to an academic example, which is based on Jordan and Graves (1995), as well as a more realistic example based on the case of a German OEM (Daimler).

A multi-stage stochastic dynamic programming (DP) approach to support flexibility planning in automobile production networks is proposed by Stephan et al. (2010). The work may be understood as an attempt to account for the limited capability of two-stage approaches to model the full scope of future capacity adjustment options. To model the evolution of demand a Markov demand model is developed. A comparably long planning horizon is considered (15 years). The network configuration is assumed to be exogenously given. Decisions regard capacity adjustments and production volume for each model to be produced at a certain production site for a particular market. The objective is to maximize the NPV of capacity dependent, switching and operating costs as well as revenues. Variable costs as well as revenues are assumed to be constant. Results of a case study as well as a simplified numerical example are presented. These are based on the setting of a German OEM (Daimler). As compared to two-stage approaches, the multi-stage approach provides the largest benefit for production networks with limited flexibility such as the exclusive assignment described in Section 2.2.

Special attention to the complementary effects of split-models and labor flexibility is given in Francas et al. (2011). The authors develop a single-period two-stage stochastic LP with recourse to optimize profits. For a given network configuration, first stage decisions include investments in permanent workforce and capacities. At the second stage the allocation of production volume to production sites and the deployment of labor flexibility are regarded. Two labor flexibility instruments are analyzed being the employment of temporary workers and worker transfer between production sites. Based on numerical studies and the analytical analysis of an idealized model the authors conclude that the number of permanent workers is positively influenced by personnel transfers and negatively by temporary employment. Personnel transfers are most effective, if the number of production sites is high and allocation flexibility is low. If chaining is applied, the potential of labor flexibility diminishes.

Papers within a *second stream* investigate the problem of identifying supply and delivery structures for the OEM and its suppliers and distributors subject to capacity constraints. This requires allocating demand quantities from different markets to production sites over the planning horizon and from this to derive network flows and stock levels with respect to procurement, manufacturing, assembly and distribution activities. While intended for mid- to long-term planning, neither investment nor capacity deployment decisions are modeled explicitly. Likewise discounting effects are disregarded.

To identify supply and delivery structures, Escudero et al. (1999) present a stochastic capacitated network flow formulation. The authors develop a two-stage stochastic program with recourse where only the network flows of the planning periods within the first stage are to be implemented. Demand, production and procurement costs are modeled as stochastic input parameters. A cost based and a service oriented formulation are proposed for the objective function. Various details are incorporated. These include alternative supply/production modes, intervals in which components may be assembled and lead times. Lost sales are computed as percentage of backlog and demand. The alternative supply/production modes may be interpreted as linear approximation of convex production cost functions and therefore reflect the principle idea of organizational capacities. An implementation oriented variable reduction scheme is introduced, however, without giving details on the model solution. Results from the application of the approach are missing.

In a series of two papers Peidro et al. (2009) and Peidro et al. (2010) develop a fuzzy mixed-integer program (FMIP) for supply chain planning. Decisions on over- and under-time and therefore capacity deployment are included into the model. The objective is to minimize the costs for regular and overtime production, inventory holding, procurement, transport and demand backlog. Fuzzy coefficients are used to model the uncertainty inherent to demand, lead times, capacities and costs. Based on the assumption of trapezoidal fuzzy numbers the model is converted into an equivalent MILP and embedded into an interactive solution approach. An application to the supply chain of an automotive supplier for seats is reported. The supply chain includes 47 companies.

Another work within in the stream is presented in Zhang et al. (2011). The objective of the proposed stochastic FMIP is to maximize the overall expected supply chain profit. The most special features are the incorporation of service level constraints and the hybrid modeling of uncertainty. Uncertainty is incorporated in terms of (fuzzy) procurement prices and (stochastic) demand. To solve the model, the authors make use of simulation optimization (scatter search), fuzzy programming and chance-constrained programming. As opposed to other works of the stream, the model does not incorporate decisions on capacity deployment. Illustrative results are reported for both, independent and correlated demand. The work is motivated by the case of FAW Volkswagen.

A slightly different perspective is taken in the paper of Kabak and Ülengin (2011). Again a supply chain planning problem is regarded. In contrast to other works of the stream, a static probabilistic network flow model is developed. The focus is on decision hierarchies and the comprehensive modeling of uncertainty. Fuzzy coefficients are used to incorporate uncertainty with respect to demand and production yield rates. In addition to that, the authors introduce fuzzy decision variables to reflect hierarchical decision making, i.e. long-term decisions are intended to be adjusted (defuzzified) with respect to more reliable information in the medium- to short-term. A profit maximizing objective is complemented by a second objective to control the fuzziness of the results. Both objectives are combined using a simple weighted sum approach. To solve the model, the probabilistic formulation is converted into an LP. In this, the authors assume triangular fuzzy numbers and use normalization to compute the value of the objective function. Results from an application at Mercedes–Benz Türk (Daimler) are reported.

The underlying assumption of papers within the *third stream* on flexibility planning is that most practical settings do not allow for the determination of valid investments and operating costs within mathematical programs. The idea is to support practitioners with information on the value of alternative supply chain configurations in the light of uncertainty. To reduce complexity, a high level of aggregation is chosen. All works focus on non-financial criteria.

The stream goes back to the seminal work of Jordan and Graves (1995). In this work the authors investigate a single-level production system with multiple products and production sites in a static environment. The network configuration, i.e., the assignment of products to sites is modeled using a bi-partite graph. A capacitated network flow formulation is introduced to compute the optimal share of demand to be produced at each site. The objective is to maximize the number of produced cars. Three principles are postulated to heuristically add product-production-site-links. Based on the results of a simulation and an analytical analysis the authors conclude that chaining structures are characterized by almost the same level of flexibility as compared to fully flexible strategies, although only employing a fraction of links between products and sites. The case of GM is considered.

Boyer and Leong (1996) develop a static MILP formulation to determine the optimal allocation of products to sites. The model extends the work of Jordan and Graves (1995) in that it includes decisions on activating additional product-site links and productivity losses which result from multi-model production. As in Jordan and Graves (1995) the objective is to maximize the number of produced cars. Demand scenarios are simulated to evaluate the effect of adding optional product – production site links for different assumptions on capacity losses that go along with multi-model production. The settings considered in the simulation are taken from Jordan and Graves (1995). According to the analysis the benefits of split models are valid, even if productivity losses are considered. The most important link is the one closing the chaining structure.

In the same line of research Graves and Tomlin (2003) evaluate the flexibility of supply chains. Based on the LP formulation of Jordan and Graves (1995) the authors analytically derive a flexibility measure which indicates whether a multi-stage supply chain is prone to inefficiencies. Configuration guidelines are developed thereupon. The authors distinguish two sources of inefficiencies in supply chains: floating and stage-spanning bottlenecks. Floating bottlenecks are a typical problem of BTO production, since capacity requirements depend on the dynamically changing portfolio of accepted orders and their configuration. Bottlenecks might therefore occur anywhere in the supply chain and change over time. Stage-spanning bottlenecks occur, if bottlenecks propagate horizontally (up- or downstream) along the supply chain. The authors conclude that the advantages of chaining remain valid in multi-stage supply chain environments, if chaining is applied to every single stage.

A dynamic generalization of the work of Jordan and Graves (1995) is presented in Francas et al. (2009). The principle idea is to analyze network configurations in the context of life-cycle demand. The authors present a two-stage stochastic model with recourse. The objective is to minimize lost sales. On the first stage, the assignment of products to production sites is determined; on the second volume decisions are made. Capacities are regarded fixed and neither production lines and logistics nor international factors are considered. The number of split-models and multi-model lines result from the model solution. Results of two numerical case studies are presented. These are based on the setting of a German OEM (Daimler). Following the results, the benefits of flexible network configurations might be substantially misjudged, if product life-cycles are not considered. Even if life-cycles are considered, the potential of chaining strategies remains robust.

4.5. Detailed capacity planning approaches

In works on detailed capacity planning the focus is put on single production sites. Of particular importance is the deployment of flexibility instruments to account for swings in demand. Flexibility instruments affect technological capacities (takt time) and organizational capacities (workforce, shift model). The proposed models are characterized by a very high level of detail. For the sake of computational tractability, the works focus on deterministic planning models. Solution approaches are developed in most works. The typical planning horizon covers some years.

Inman and Jordan (1997) present a static non-linear deterministic mixed-integer program (MIP) to minimize the number of workers for a given production program. Decisions regard the mid-term assignment of models and workers to production lines as well as the takt time and therefore both, organizational and technological capacities. Change-over times are considered. Shifts and working time accounts are not included. To solve the model, a constructional heuristic is developed, which uses branch & bound techniques for improvement. The heuristic performs well for small instances. The performance deteriorates approximately linearly with the size of the problems. The case of GM is considered.

Benyoucef et al. (2007) develop a model to support detailed capacity planning based on a sales forecast, a given takt time, the current inventory level and workforce constraints. The latter comprise the limited use of non-working days, overtimes and third shifts. As a result, shift models and production quantities are derived for each assembly line. Mixed-model production is not considered. To determine capacity adjustments, seven rules are formulated in an order of decreasing usage priority. A dynamic programming approach is developed, which uses the rules to construct a feasible solution. To illustrate its application, the example of an automotive OEM is used.

Askar et al. (2007) regard the detailed planning of technological and organizational capacities. Based on a single-line baseline model, a comprehensive approach is developed which uses approximate DP to solve separate non-linear MIP models for final assembly, paint shop and body shop. Each shop might encounter multiple production lines. The objective is to determine cost minimal mid-term capacity plans which are operationally consistent. Two planning levels are distinguished within an integrated framework. On the top level, technological and organizational capacities are planned for macro periods (e.g., weeks). The technological capacities of each line result from decisions on takt times and the choice of a pre-defined line configuration. Organizational capacities result from decisions on shift models and the amount of regular and temporary workers as well as the usage of working time accounts. Capacity adjustment options for the paint and body shop are incorporated in terms of the activation and deactivation of parallel production lines and the use of empty work piece carriers (Section 2.2). On the lower level, the authors introduce micro periods (e.g., shifts) within each macro period. The micro periods are used to assure that capacity plans are feasible with respect to inter-shop buffer constraints. The objective is to minimize discounted costs encountered from capacity changes and operations. If multiple shops are considered, a separate DP is solved successively for each line in every shop beginning with final assembly. A rule-based approach is presented to coordinate the solutions.

In Askar (2008) the DP proposed in Askar et al. (2007) is extended with respect to (a) a more detailed capacity model, (b) an approximation scheme for the DP and (c) two further approaches to coordinate the successive planning of the shops. (a) The extended capacity model includes the effects of learning curves and non-productive idle-times. Labor productivity and line productivity are reduced following changes of the takt time. This effect goes back to learning curves which are approximated by stepwise linear functions. To reflect non-productive idle-times, the labor productivity of a given assembly line configuration is further decreased proportionally with increasing takt time. As a result of a decreasing labor productivity, the labor requirements per car increase. To solve the extended DP, (b) an approximation scheme is developed. The idea is to reduce the dimensionality of the decision and state space. A combination of discretization and problem specific expertise is used. Based on this expertise, the number of alternative shift models can be reduced. Also, decisions on the production volume can be restricted to either producing at maximum capacities of the chosen shift model (minimize costs for canceling shifts) or synchronous to demand (minimize holding costs). (c) In addition to the rule-based approach introduced in Askar et al. (2007), Askar (2008) presents two further approaches to coordinate the successive planning of the shops. (1) Given the macro period demand of the succeeding shop, a linear production planning model is solved for every state of the DP. From the solution of this model, violations to the inter-shop buffer constraints are determined. States are deleted from the DP, if the shortfall is larger than a pre-defined threshold. If no feasible state remains, the planning of the succeeding shop is iterated to adjust the macro period demand. In the case of parallel lines, rules are used to allocate buffer capacity and demand to each line a priori. (2) An approach based on Lagrange relaxation and subgradient procedure is presented to avoid the exogenous definition of rules and threshold values. The idea is to integrate the violations of buffer constraints into the single-line DPs and to iteratively solve the DPs based on that. Both approaches are significantly slower than the original rule-based approach but are less restrictive. The case study of Daimler is reported.

Sillekens et al. (2011) extend the work of Askar (2008). A MILP formulation for a single-level, single-line setting is presented. The objective is to minimize total costs. While productivity losses due to capacity changes are neglected, special attention is given to modeling labor demand and labor costs. The micro-period concept is extended to assure the operational consistency of plans with respect to working time accounts and capacity. The most special feature of the approach is the presentation of a linear approximation scheme to include constraints arising from working time accounts into mid-term capacity planning. Results of a numerical study suggest that flexibility options in terms of alternative takt times and shift models result into reduced costs while making more use of working time accounts.

A multi-level, parallel-line extension of the model is proposed in Sillekens (2008). The work includes worker transfers between lines and presents a solution approach which combines problem specific pre-solving techniques with an adopted branch & bound procedure. Pre-solving is based on the idea of pre-setting/constraining binary and integer variables. All combinations of technical and organizational capacity configurations are eliminated from the solution space that require more than the maximum number of workers or fall below demand. Primal heuristics are integrated into a branch & bound procedure to improve the computation of boundaries. In addition to standard techniques (LP-and-Fix, Relax-and-Fix), a heuristic is proposed which relies on constraining capacity adjustments in case that the demand of two consecutive periods is similar.

The approach is evaluated based on the case study of Daimler. For a three line setting, no optimal solution could be found within 6–9 hours of computation time. The duality gap ranged between 4% and 14.7%. While the relative performance of the acceleration techniques varies, they improve the performance of the solution approach. For a given budget of computing time the results of the integrated model are worse as compared to a sequential solution approach which solves single-line models starting from final assembly according to the ideas of Askar (2008). Generally, the DP approach of Askar (2008) performs superior for small instances (Sillekens et al., 2011). If multi-level production and change costs are considered, the branch & bound approach performs better.

In the same line of research Walter et al. (2011) investigate the potential of different capacity adjustment options and their interactions. Towards this end, the authors apply a fully-factorial sensitivity analysis to a capacity planning model for a single final assembly line. The model has close similarities to that developed in Sillekens et al. (2011). Two parameter levels are considered for five selected capacities adjustment options. These are the introduction of an additional takt time, the shift model, the allowance of more extra hours, less restrictive regulations of the working time accounts, and the allowance of additional temporary workers. To solve the problem, the approximate dynamic programming approach of Askar (2008) is adopted. The case of Daimler is considered. For this case, the flexibility instruments takt time, shift model and temporary workers showed promising potential. Interaction effects between the instruments could not be found.

The detailed capacity planning of multi-line final assembly shops is considered in Roscher (2008). The objective of the proposed non-linear MIP is to minimize discounted costs. Technological and organizational capacities are modeled on a level of detail which is comparable to that used in Askar (2008) and Sillekens (2008). The most special feature of the approach is the detailed modeling of learning curves. Two learning effects are distinguished. A first effect regards mid-term learning depending on the cumulative production volume of a product on an assembly line. A logarithmic-linear term is used to compute the increase in productivity. The parameters of the learning curve are adjusted, if the same or similar products are produced at the same line/site. A second effect regards reductions in the labor productivity following changes of the line configuration. In contrast to Askar (2008), a detailed model is used. Towards this end, a second logarithmic-linear term is introduced to model the labor productivity depending on the time passed after the last change of the takt time. An approximate dynamic programming formulation is suggested to solve the model. Several possibilities to reduce the dimensionality of the state and decision space are discussed. These are based on domain expertise (e.g., adding/canceling shift groups one by one, prohibiting increasing takt times during ramp-up, prioritizing assembly lines and worker groups) and heuristic ideas (discretisation, iterative solution of single-line problems). To incorporate risk, the expected performance is complemented by information on the worst case performance. The approach is illustrated based on the case study of Daimler.

Garcia-Sabater et al. (2011) develop a planning framework which supports the mid- to short-term planning of an engine assembler. The framework integrates two MILP models for 6-month capacity deployment and 4-week production planning. The most interesting features of the approach are on the one hand the integration of capacity deployment with short-term operational planning, materials requirements planning and distribution planning. On the other hand, the approach incorporates several specific industry requirements. These include the leveling of production quantities and deviations to target stock levels as well as the planning stability as compared to previous plans. A simple additive weighting approach is used to include multiple criteria into the objective function. To solve the model, both MILPs are treated sequentially using standard MILP solvers. To illustrate the applicability of the approach, the authors present some results from an industrial application.

The match of the works on the planning of capacities with the requirements provided in Section 2 is provided in the discussion (Section 6). Based on this match directions for future work will be derived.

5. Review of approaches for the planning of orders

5.1. Classification scheme

To review works on the planning of orders, a classification scheme is developed in the following (Table 4). The proposed scheme classifies publications according to four attributes. A first attribute regards the *planning tasks*. According to the framework presented in Section 2, the focus is on works which concern order promising and generation as well as central and decentral production planning. The second attribute characterizes the proposed *models*. As compared to the planning of capacities, the classification is less differentiated. The underlying reason is that all models considered in the review are deterministic. Accordingly, the attributes objective function, model type and modeling of demand are aggregated into the single attribute model. The characteristics have been described in Section 4.1. We have added a new attribute labeled planning object to indicate whether planning is based on vehicle models (m) or orders (o). With respect to the *technological capacity,* differentiated shop capacities are no longer considered as separate attributes, since none of the approaches found in the literature incorporates any of these. At the same time, a more differentiated modeling of final assembly capacity was found in the literature. The analysis revealed three approaches: the modeling of model-mix constraints (mm), station workload (swl) and aggregate cars (c). With the exception of Boysen et al. (2009b)

Table 4
Planning of orders: classification scheme.

Attribute	Description
1. Planning task	
Order promising	Modeling of order promising decisions
Order generation	Modeling of order generation decisions
Central production planning	Modeling of central production planning decisions
Decentral production planning	Modeling of decentral production planning decisions
2. Model	
Time	Time scale: static (s), dynamic (d)
Objective function	Criteria used within objective function: financial (f), multi-criteria (mc), else (e)
Planning object	Planning object: models (m), orders (o)
Fulfillment	Fulfillment constraints: fulfillment (f), delay (d), selection (s)
Vertical interdependencies	Planning scope: single line (sl), parallel lines (pl)
Horizontal interdependencies	Levels incorporated into planning: single-level (sl), multi-level (ml)
3. Capacity (technological)	
Final assembly	Incorporation of final assembly capacity: station workload (swl), cars (c), model-mix (mm), none (n/a)
Component	Incorporation of component availability
4. Evaluation	
Uncertainty	Incorporation of demand uncertainty
Informational dynamics	Incorporation of informational dynamics
Criteria	Criteria used for evaluation: financial (f), multi-criteria (mc), solution quality (sq), else (e), none (n/a)

Table 5
Planning of orders: classification.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Biller et al. (2002)	Bish et al. (2005)	Biller et al. (2005)	Biller and Swann (2006)	Gruß (2008)	Hayler (1999)	Stautner (2001)	Brabazon and MacCarthy (2006)	Brabazon et al. (2010)	Volling and Spengler (2011), Volling (2009)	Hindi and Ploszajski (1994)	Bolat (2003)	Ding and Tolani (2003)	Dörmer et al. (2010)	Boysen et al. (2009b), Boysen (2005)	Gans (2008)
1. Planning task																
Order promising	☐	☐	☐	☐	☑	☐	☐	☑	☑	☑	☐	☐	☐	☐	☐	☐
Order generation	☐	☐	☐	☐	☐	☑	☑	☑	☑	☐	☐	☐	☐	☐	☐	☐
Central production planning	☑	☑	☑	☑	☑	☐	☐	☑	☑	☑	☐	☐	☐	☐	☑	☐
Decentral production planning	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑	☑	☑	☑	☑
Sequencing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☑	☐
2. Model																
Time	d	d	d	s	d	s	s	d	d	d	s	s	d	d	d	s
Objective function	e	e	f	f	f	n/a	e	e	e	mc	e	f	e	e	f	e
Planning object	m	m	m	m	o	o	m	o	o	o	o	o	m	o	o	o
Fulfillment	s	s	d	d	s	n/a	n/a	d	d	d	f	d	f	d	f/d/s	f
Vertical interdependencies	pl	pl	sl	pl	sl	n/a	n/a	sl	pl	sl	sl	sl	sl	sl	sl	pl
Horizontal interdependencies	sl	sl	sl	ml	ml	n/a	n/a	ml	ml	sl	sl	sl	sl	ml	sl	sl
4. Capacity (technological)																
Assembly	c	c	c	c	c/swl	n/a	n/a	c	c	mm	mm	swl	n/a	swl	mm	swl
Component	☑	☑	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐
6. Evaluation																
Uncertainty	☑	☑	☑	☐	☐	☐	☐	☑	☑	☑	☐	☐	☐	☐	☐	☐
Informational dynamics	☐	☑	☐	☐	☐	☐	☐	☑	☑	☑	☐	☐	☐	☐	☐	☐
Criteria	mc	mc	f	f	f	n/a	n/a	mc	mc	mc	sq	sq	sq	sq	sq	e

organizational capacities have not been incorporated into existing planning approaches so far. We have therefore not integrated the attribute into the classification scheme. Based on the review of works, we have added a further attribute which is used to classify the approach used to evaluate the performance of the proposed models. The performance *evaluation* is challenging due to demand uncertainty, informational dynamics and complexity. In BTO automobile production customer requests/orders approach the company in a dynamic process. Neither the timing nor the specifications of future customer requests are known with certainty. Also, the informational basis is evolving dynamically (Section 2.3). This results in multiple plans being generated for the same planning period. Instead of directly interpreting the objective function values of single model runs this requires assessing aggregate performance criteria which concern multiple planning iterations. Often multiple criteria (mc) are used for the evaluation. Works which address the complexity issue follow a different evaluation approach. Typically the solution quality obtained from a particular heuristic (sq) is reported. Details on the modeling techniques and solution approaches used are given in the description of each approach.

5.2. Problem classification

Table 5 gives a comprehensive survey of approaches for the planning of orders. The analysis is focused on the planning problems introduced in Section 2.3. We do not consider qualitative or empirical works such as Childerhouse et al. (2008) or Graf (2006). In total 18 works are considered, 11 of which are refereed journal articles. 12 of the 18 studies regard industrial cases or are at least motivated by a specific case from the automotive industry. Amongst the OEMs considered are GM (4) BMW (2), Daimler (2), Ford (2) and Jaguar (1).

As compared to the planning of capacities, the field is more fragmented. Two main branches can be identified in the literature. These are approaches focusing on central planning and order promising on the one hand (Section 5.3) and on decentral planning on the other hand (Section 5.4). We classify a work central, if either the assignment of orders to production sites or the interface to order promising is modeled explicitly. Decentral approaches assume a given pool of accepted orders and focus on the compilation of production plans from these orders.

5.3. Central planning and order promising

We distinguish two classes of works on central planning. A first class of works (1–7) provides decision support for selected aspects of central production planning and order promising/order generation. The integration of the planning tasks is regarded in a second class of papers (8–10).

Within the *first class*, Biller et al. (2002) and Bish et al. (2005) investigate the impact of different allocation policies used for central planning. Biller et al. (2002) provide a three stage framework to integrate central planning with mid- to long-term aspects. The objective of central planning (base-level) is to dynamically assign observed demand for products to production sites such that unsatisfied demand is minimized and the model-mix is leveled. To this end, a quadratic lexicographic multi-objective IP is developed. To determine target production levels for each site, three allocation policies (mid-level) are developed. These policies support the mid-term allocation of production volume to production sites for different product models according to (1) allocation priorities, (2) fixed ratios or (3) allocation priorities and profits. The authors further suggest using the proposed central planning approach to evaluate alternative network configurations (top-level). The approach is evaluated based on two simulation studies. Following an illustrative setting, the setting of Jordan and Graves (1995) is adapted, taking correlated demand into account. The best performance in terms of unsatisfied demand and inventory is obtained from combing the fixed ratio policy with the quadratic IP. The work has been synchronized with GM.

A very similar, yet more stylized decision situation is treated in Bish et al. (2005). Given a set of customer requests, prioritizing policies are used to determine a production site for each request. The policies in principle follow the idea of those presented in Biller et al. (2002), however, extended by a component which takes the customer location into account. In contrast to Biller et al. (2002), allocation is done solely based on these policies. Demand is lost, if it cannot be served given the available capacity. Consequences with regard to sales, inbound and outbound logistics, as well as inventory are evaluated by means of analytical and numerical analysis. A stylized two-plant, two-product model is used. Each product is assumed to require one specific component. Backlogging is not considered and demand is considered identically and independently distributed (IID). The authors conclude that order allocations based on the proximity between customers and production sites and prioritized according to the contribution margin may result in higher variability in production and reduced profits. Amongst others, the effect is moderated by contribution margins, lead times and inventory holding costs. Again, the work has been synchronized with GM.

The integration of demand management techniques to improve central planning is analyzed in a series of three works. In the first work, Biller et al. (2005) regard a BTO environment where customers order their cars at a price set by the OEM. A simplified version of a multi-period capacity and inventory planning model is extended to include pricing as additional decision variable. To model price-sensitive demand, linear functions are used. Informational dynamics are not considered. The focus is on a single car model. However, the model can easily be adapted to a multi-model setting, if demand diversion among different models is ignored. A greedy heuristic is developed to solve the resulting non-linear IP. The model is applied to a mid-term planning horizon covering several seasons (months). Capacity is modeled in terms of cars. Results from a numerical study are presented. These suggest that the profit potential from a small number of price changes may be significant. In addition dynamic pricing may help in absorbing demand variability.

Similarly, Biller and Swann (2006) analyze the contribution of pricing towards compliance with environmental legislation. The model of Biller et al. (2005) is extended to incorporate aggregate capacity constraints for critical components (engines, transmissions) and vehicle assembly lines. The most interesting feature is the modeling of corporate average fuel efficiency (CAFE) legislation. To solve the non-linear IP a commercial solver is used. The case of GM is considered.

A planning approach which embeds a more realistic demand model is provided by Gruß (2008). The idea is to integrate a multi-nominal LOGIT-model to integrate customer-choice into a capacitated production planning framework. The basic idea is to identify a set of product configurations (the offer set) to present to a certain customer (segment) dependent on time to production and the availability of capacity. Based on the LOGIT-model, purchase probabilities are quantified in dependence of the offer set. Two capacity models are introduced: a first with a high level of aggregation and a second to incorporate multiple resources. For the first model a stochastic dynamic programming approach is suggested, for the second a deterministic approximation. For the simple setting, an illustrative example is reported. While the potential of adopting the approach to high variance BTO environments seems to be promising, the author concludes that the computational burden and the data availability strongly limit the applicability.

The problem of order generation is regarded in a series of two works. These have not been published in international research outlets yet. Hayler (1999) develops a decision support tool for generating BTS car configurations to support car dealers. Starting point is the definition of a number of orders to be generated in different price classes. The sequential order generation approach is based on association rules, historic orders and expert knowledge. Starting from the model, colors are identified, which have often been ordered with the model. The configuration is successively completed by adding typically combined interior trims and further options. All permutations of model, color, interior trim and options are considered, which fulfill minimum confidence constraints within the association analysis. Starting from the configuration with the highest deviation from the desired model-mix, orders are generated until the required number of orders is available within each price class. If the permutation procedure does not result in sufficiently many configurations for low price classes, those options are deleted from the generated configurations which have the lowest association confidence. If high value orders are missing, expert knowledge generated from a non-recurrent survey of dealer perceptions is used to suggest additional options. To improve performance, the procedure is divided into an offline procedure, which determines all configurations that fulfill minimum confidence constraints, and an online procedure, which uses these configurations to compile the required number of orders. The work is based on the case of BMW.

A very similar approach is provided by Stautner (2001). The main difference is that suggestions are generated based on fully specified orders as opposed to the synthetic construction of orders from independent associations between options proposed by Hayler (1999). The idea is to use cluster analysis to derive representative orders from recent historic orders. Partially specified orders are identified with a *k*-means cluster algorithm. Association rules are used to add further options to these orders. In this, associations between five to ten options are considered simultaneously. The approach is embedded into a three stage procedure. First, representative partially specified orders are identified. In a second step, forecasts are generated for each representative order. In a third step, BTS orders are generated manually based on the representative orders, the forecasts and the association rules. The approach is illustratively applied to the US market of BMW.

Within the *second class*, integrated models for central planning and order promising are developed. The focus is on the behavior of these models in the light of informational dynamics. Planning policies are developed, which are evaluated based on simulation. These policies are based on non-financial criteria and are derived from deterministic models. Since order promising decisions are modeled explicitly, customer requests may be delayed or rejected.

Brabazon and MacCarthy (2006) study the fundamental behavior and performance of order fulfillment using a holistic simulation model. In addition to BTO order fulfillment the authors consider BTS as a second parallel fulfillment mode. The formal assessment of such systems is given in Brabazon and MacCarthy (2010). The order fulfillment process is divided into three segments: the stock of unsold cars, the frozen production plan and the pool of customer orders waiting for being considered in the frozen production plan. Cars are produced at a fixed rate determined by capacity. It is assumed that central planning adds production orders to the frozen production plan in first-come-first-served mode. If no customer order is available, central planning generates production orders with randomly chosen configuration. To model demand the authors randomly generate customer requests for specifically configured cars. Within order promising, at first the production plan and the stock of finished cars are searched for a car which matches the requested configuration and is unsold, i.e., has not been assigned to another customer request. If no such car is available, an order is generated to fill the customer request and is added to the order pool. Accordingly, a combination of BTO and BTS order fulfillment is considered. Effects of product variety, the number of orders in the production plan and the direction of searching BTS cars (starting from the last BTS order added to the plan vs. starting from the oldest car on stock) on the delivery times and the stock level are examined. Model structure and planning processes are synchronized with the case of Ford.

Using an extended version of the model, Brabazon et al. (2010) analyze the flexibility of combined BTO/BTS order fulfillment strategies as means to realize short lead times and enable high product variety. More advanced policies are used for order generation. The idea is to align the model-mix in the production plan and on stock with the models' take rates. Those product configurations are preferred in the generation procedure for which the deviation from the anticipated take rate is largest. A critical assumption is that the take rates, which are used order generation, perfectly match the true demand – regardless of the potentially large product variety and the associated uncertainty. Multiple dealers are considered constituting vertical interdependencies. Two types of flexibility are compared. These are the reconfiguration of BTS cars in the production plan and interdealer trading. Experiments reveal that both types of flexibility can bring reductions regarding stock level and delivery time. In this, reconfiguration flexibility dominates interdealer trading. Results strongly depend on the degree of variability.

In the same class of papers, Volling and Spengler (2011) consider the integration of central planning with order promising. Two MILPs are developed. One to determine delivery dates in direct response to customer requests (online order promising) and a second to generate production plans from the set of quoted orders (offline central planning). Simple additive weighting is used to incorporate customer service and efficiency criteria into the objective functions. The most interesting feature is that both models are integrated into a hierarchical framework. Anticipation and feedback techniques are developed to improve the coordination of both models in the face of informational dynamics. To evaluate the approach, simulation is used, which is based on data from a German OEM. An ARMA demand process is implemented in order to account for the lack of internal consistency attributed to IID-figures. According to the analysis, the proposed coordination schemes may contribute to solve the trade-off between the leveling of mix-model production, responsiveness and holding.

An extension of the work is provided in Volling (2009). A simulation-based optimization approach is presented to identify configurations of the mathematical models which result into a good performance in a rolling horizon setting. A model configuration results from defining values for the weights used in the objective functions of the MILPs. A combination of ranking-and-selection (to dynamically control the number of replications used to evaluate each neighborhood) and tabu-search (to control the variations of the parameters) is used. In each step of the iterative search procedure, the application of all model configurations which define one neighborhood is simulated to derive estimates on aggregated performance criteria. Based on the results, the weights are adjusted and the next neighborhood is defined according to a tabu-search procedure. While computationally intensive, the approach supports the definition of model parameters in the face of informational dynamics.

5.4. Decentral planning

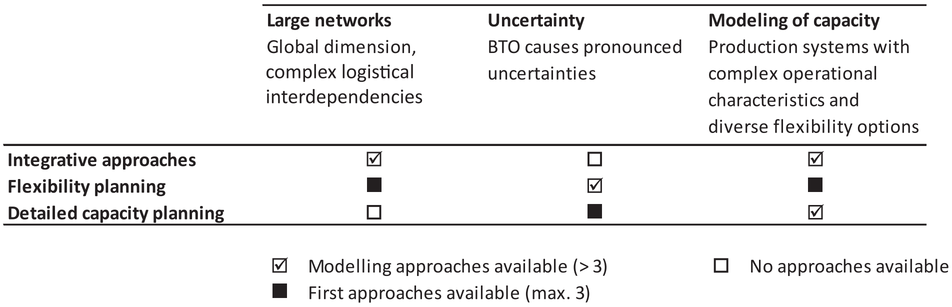
Starting point for decentral planning is a pool of accepted orders with specific configuration. For these orders the objective is to determine production lines and periods. With the exception of Bolat (2003) and Boysen et al. (2009b), the approaches use non-financial criteria and assume that all orders have to be fulfilled. Most approaches focus on single assembly lines and do not incorporate multi-level production. The focus is on the development of solution approaches. The obtained solution quality is reported as central result of the evaluation.

Hindi and Ploszajski (1994) consider the selection of a subset of cars from the order bank for production in the next planning period. Product options are used to model capacity requirements and availability. The objective of the IP is to generate a production plan which maximizes the production output, i.e., the total number of fulfilled orders. Minimum production volumes are set for each market segment. A second model is developed for sequencing. However, both models are not integrated. A solution approach is developed based on Lagrangian relaxation coupled with a subgradient procedure and simple list processing. A shortcoming of the approach is its tendency for cherry picking. Since orders are solely included into the production plan according to resource criteria, unacceptably long lead times may result. The work was synchronized with Jaguar.

Bolat (2003) extend the ideas of Hindi and Ploszajski (1994) with respect to a cost-based model formulation. The objective of the IP is to select orders for the next production period such that delivery date related costs are minimized. These comprise tardiness and earliness cost which are encountered, if an order is not scheduled to the promised date/period. To incorporate the model-mix, constraints are imposed on the maximum and minimum capacity load derived from detailed capacity planning. To solve the model, a relaxed branch & bound procedure complemented by a heuristic to construct a feasible solution is developed. The solution approach is evaluated based on a static numerical analysis. Again, the approach has a tendency for cherry picking. Since only costs for those orders are considered in the objective function, which are included into the production plan, the approach has a tendency to exclude orders, if they have high penalty costs due to delayed production.

The dynamic extension of the static problem considered so far is treated in Ding and Tolani (2003). The authors regard a situation, where models are to be assigned to days of production. Customer orders are not modeled explicitly. In order to allow for just in time production, two model-mix objectives are combined into a vector optimization approach. The first regards the sum of squared deviations to the average daily production quantity of the production system and the second the sum of squared deviations to the average daily production quantity of each model. A weighted sum approach is used to aggregate both criteria. The objective is to minimize the weighted sum. To solve the resulting quadratic MIP, local search heuristics are developed. Capacities are not considered.

A more comprehensive approach for BTO automobile production is developed by Dörmer et al. (2010). A multi-period MILP for the assignment of individual orders to production periods is developed. Capacities of a single-stage production system are considered as well as the availability of critical parts. In order to achieve a leveled model-mix, the spread between the minimal and maximal capacity utilization of assembly stations and parts is minimized. Time windows are used to control the adherence of due dates. To solve the model, a two-stage heuristic is proposed. Two methods are suggested for the determination of an initial solution, being a list-scheduling-procedure and a MILP-based-approach. To improve the solution, a local search procedure is used. Numerical results show, that the approach outperforms a commercial MILP solver in terms of speed and solution quality, if



	Large networks Global dimension, complex logistical interdependencies	**Uncertainty** BTO causes pronounced uncertainties	**Modeling of capacity** Production systems with complex operational characteristics and diverse flexibility options
Integrative approaches	☑	☐	☑
Flexibility planning	■	☑	■
Detailed capacity planning	☐	■	☑

☑ Modelling approaches available (> 3)
■ First approaches available (max. 3)
☐ No approaches available

Fig. 6. Matching planning approaches with requirements: planning of capacities.

the solver is interrupted after one hour. The work was synchronized with the case of Daimler.

The integration of decentral planning into a comprehensive framework is provided by Boysen et al. (2009b). The focus is put on the interplay of production planning, reconfiguration planning and sequencing. In addition to that, model extensions to incorporate order promising and detailed capacity planning are discussed briefly. However, there is no explicit differentiation between central and decentral planning. Since the focus is put on compiling production plans from promised orders, we consider the work decentral. The objective of the approach is to minimize the cost related to assigning a particular order to a planning period. Different adoptions of the resulting multi-resource generalized assignment problem are proposed. These regard the cost of operations and setups. Both might vary, if the line configuration is changed. Additional constraints are added to incorporate model-mix aspects. While the focus of Boysen et al. (2009b) is put on the model formulation, solution approaches are developed and evaluated in Boysen (2005). Two solution strategies are compared. A first builds on the idea to solve a LP-relaxation of the original problem and to construct a feasible integer solution from that. Two alternatives are analyzed. For the first alternative, the LP-relaxation of the assignment problem is solved. For those orders, which are not assigned to exactly one period, the original MILP is solved. For the second alternative, Lagrange relaxation is applied to the assignment constraint and combined with a subgradient procedure to compute the Lagrange parameters. A construction heuristic is used to generate a feasible solution from that. A second solution strategy builds on Threshold-Accepting. According to the analysis, the best performance in terms of solution quality and speed is obtained from the LP relaxation. Informational dynamics are not addressed.

All approaches considered so far do not incorporate parallel production lines. A first approach for the assignment of orders to assembly lines is provided in Gans (2008). The approach is embedded into a line balancing context. The basic assumption is that efficiency gains can be reaped in line balancing, if the variance of processing times can be reduced by decentral production planning. To this end, a capacitated *k*-means clustering approach is developed. For each of the parallel assembly lines a separate cluster is defined. The objective is to group orders to these clusters such that the variations of the processing times are minimized. Processing times are considered for each station of the line. Clustering is based on a rule-based multi-start clustering heuristic. The analysis gives evidence that the variance can be reduced by 60% on average. The work was synchronized with the case of Daimler.

6. Discussion and conclusions

In this survey we have provided a comprehensive review on research streams in the field of capacity planning and the planning of orders in BTO automobile production. The development of domain specific planning approaches can be traced back to eight major requirements, which have been introduced in Section 2. In the following we will match these requirements with the approaches reviewed so far. This will lay the basis for the identification of open issues. In addition to that, we will raise issues for future work, which go beyond the scope of the scheme.

6.1. Planning of capacities

The available models for the planning of capacities are specifically designed towards selected aspects of the planning tasks. The question of how to configure production networks in the face of uncertainty is regarded in approaches for flexibility planning. High level models are proposed. While some works address the coordination of multi-level material flows along the supply chain, there is significantly less attention given to the modeling of international factors as well as technological and organizational capacities. The modeling of capacity receives much attention in research on detailed capacity planning. While effectively incorporating technological and organizational capacities, existing models assume deterministic demand. The focus is on single production sites. Planning decisions derived from those models are thus prone to local optima. This might explain the growing interest in integrative models which allow finding well balanced decisions for the involved subproblems in global production networks. Integrative models would allow for synchronizing resources on all planning levels and setting the right level of flexibility. Existing approaches provide modeling techniques for integrating planning tasks from different planning levels and incorporate international factors. They, however, fall short on properly integrating flexibility and capacity aspects.

The match between requirements of BTO automobile production and the available research for the planning of capacities is summarized in Fig. 6. Based on this match the following directions can be identified for the development of more comprehensive models.

The most important extension of integrative approaches regards the improved incorporation of uncertainty. Given the dynamics of the marketplace, it is vital for automotive companies to assess the risk inherent to mid- to long-term planning decisions. Works on the stochastic version of integrative planning models (Section 4.4) may be understood as a first step in this direction. However, existing models miss numerous important aspects of global production. Adding these aspects into existing models will improve the model quality but will likewise increase solution times. Solution times surge even further, if uncertainty is incorporated. This requires for efficient solution approaches. Initial effort in this direction is made in the work of Kauder (2008).

Important directions for future work on flexibility planning are the incorporation of international factors and the improved modeling of capacities. A first attempt to model international factors is made by Friese (2008). A detailed capacity model is introduced by Bihlmaier et al. (2009). Making use of hierarchical planning techniques, the work offers a promising concept to incorporate

	Large networks	Uncertainty	Modeling of capacity	Informational dynamics	Complexity	Multi-criteria	Model-mix
Central planning and order promising	■	☐	☐	■	☐	■	■
Decentral Planning	☒	☒	☐	☒	☑	■	☑

☑ Modelling approaches available (> 3)
■ First approaches available (max. 3)
☐ No approaches available
☒ Not relevant

Fig. 7. Matching planning approaches with requirements: planning of orders.

operational characteristics into mid- and long-term planning. Integrating both aspects would be an attempt to better align decisions on network, site and production line level. Further open fields for future work are solution approaches and the incorporation of risk. Interestingly, the notion of risk is addressed in the early work of Eppen et al. (1989) but has not been adapted until very recently.

Future work on detailed capacity planning should concentrate on two aspects. Since multiple assignments of models to production sites become common practice in the automotive industry, detailed capacity planning should firstly incorporate network effects. First ideas in this direction are presented in Wittek et al. (2011). Secondly, more detailed capacity models are promising. This in particular regards the reconfiguration of production lines, which is widely used for mid-term capacity adjustment in industry (Altemeier et al., 2010; Boysen et al., 2009b). Additional improvements may be obtained from incorporating stochastic influences. If, however, capacity adjustment decisions can be changed without any significant costs, the expected utility from integrating these influences will effectively be small.

In order to provide decision support for the planning of capacities, some general opportunities for future work can be identified. These aspects have not sufficiently been addressed in prior work.

- *International factors*: The integration of international factors into planning models is far from comprehensive. In particular the impact of tax payments and reimbursements on production networks may be very significant and requires further attention. The same holds true for tariffs.
- *Implicit part requirements*: A challenge which has not been addressed in existing works for the planning of capacities regards implicit part requirements. In the automotive industry part requirements may result explicitly from single product options or implicitly from the combination of different options (Gebhardt et al., 2004). The choice of a seat heating in combination with a high-performance radio might for instance require the installation of a high-capacity battery. The number of these implicit dependencies may be as large as 100,000 for a single car model. Implicit part requirements link demand for one part or production resource to the availability of other parts or production resources on the same and different stages of the supply chain. Implicit part requirements may therefore cause an additional source of inefficiency in supply chains which deserves further attention (Section 4.4). If, for instance, the assembly of seat heatings is constrained, this will likewise limit material requirements with respect to heating elements and high-capacity batteries.
- *Portfolio planning*: The planning of capacities is tightly linked with decisions on the introduction and phase-out of products in the course of product portfolio planning. Existing models assume portfolio planning to be decoupled from capacity planning and therefore miss to identify synergies. A first work to integrate both aspects is Xu et al. (2009).
- *Data supply*: A topic which has not received much attention in existing works yet is the supply of data necessary to formulate models for the planning of capacities. Consider the following example. Simple models for planning global networks comprise index sets for production and sales regions, products and time. If only five regions are considered for production and sales, five products and five time periods are considered, 625 data points have to be generated for the outbound logistics costs. This illustrative example gives rise to the important question of how to set up efficient processes for the generation and management of data in global networks.
- *Incorporation of organizational constraints*: When optimizing globally dispersed production networks, local autonomies and expertise have to be adequately considered. This in particular becomes relevant for planning approaches involving multiple independent companies. Coordinating such networks requires incentive and controlling structures which guide the behavior of each actor towards the global optimum. Approaches of distributed decision making may prove helpful in advancing existing models. These have, however, not been discussed in existing works.

6.2. Planning of orders

As compared to the planning of capacities, the planning of orders has received considerably less attention in the literature. In addition to that, the approaches available are more fragmented in terms of scope, assumptions and modeling techniques. We have distinguished two main classes of works. Approaches for decentral planning provide decision support for the compilation of production plans from orders which have already been accepted. The main focus is on the development of solution approaches to address the complexity challenge. Approaches for central planning and order promising explicitly consider the interface between customers and the production system. These approaches provide decision support for plant assignment and order promising. First works are available for order generation and demand management. The match between the requirements and planning approaches is given in Fig. 7.

Works in the domain of central planning and order promising may be understood as important building blocks for the planning of orders in BTO automobile production. First approaches are available to incorporate informational dynamics, plant assignment decisions, multiple criteria and model-mix constraints. Problem specific solution approaches are missing by now as are models to cope with uncertain demand. Generally there is a lack of models which address the requirements in a comprehensive manner. In addition to that, works on the integration of the specific approaches are missing. For instance the question is open, to what extend order generation and demand management techniques contribute towards an improved performance of BTO order fulfillment systems. This results into several open issues for future work.

- *Large networks*: Models which capture the effects of decisions on orders on the costs of logistics activities as well as customer service (e.g., lead times) are missing. Given the geographical extension of existing BTO production networks, these aspects considerably influence operational performance. International factors may be another interesting aspect. The works of Biller et al. (2002) and Bish et al. (2005) may be understood as a first step into this direction. A more comprehensive model of inbound logistics is discussed in Florian et al. (2010).
- *Uncertainty*: Existing models for order promising are deterministic and process orders first come first served. The competition which exists between a current customer request and future requests is therefore not considered. Given that the willingness to pay differs widely across customers, there should be a potential to more effectively integrate techniques from demand management (Waller, 2004; Voigt et al., 2008). A thorough assessment of benefits and costs of such systems in different situations, i.e., lost sales for different order fulfillment strategies, is missing (Biller et al., 2005). In addition to that, existing models are targeted towards a small number of different products. Given the vast product variety of BTO automobile production, more differentiated demand management techniques are needed. Quantity based approaches such as capacity controls have proven helpful in other industries (Hintsches et al., 2010). An adaption to the automotive industry could be helpful in maximizing the value contribution of resource utilization. Uncertainty is likewise not sufficiently incorporated into works on order generation. More rigorous system analysis, model development and evaluation will be necessary to understand how order generation can be effectively supported. In order to better incorporate uncertain demand information, research on information capturing techniques as well as marketing-related approaches like customer-choice modeling including up- and downward buying behavior may be helpful.
- *Capacity*: The existing models are based on aggregate capacity models. Flexibility to adjust capacities is not incorporated. This flexibility is quite significant given the lead times of BTO automobile production which range around several weeks. We would therefore expect an additional potential to improve planning from better coordinating the planning of orders and detailed capacity planning. This in particular holds true for the compilation of production plans for each production site (central planning) as well as the anticipative specification of production orders (order generation).
- *Complexity*: In existing works for central planning and order promising simplified models are formulated. Accordingly, special solution approaches are not necessary. The integration of one or more aspects discussed above will result into a significantly increased complexity. Solving the problems will therefore require more effort. Approaches from decentral planning and detailed capacity planning may serve as a starting point for efficient solution approaches.
- *Multi-criteria*: While first approaches for the planning of orders in BTO automotive production consider multiple criteria, there is no common understanding on which criteria are the most relevant. Empirical studies might be helpful in exploring this important question.

Approaches for decentral planning focus on single sites and assume customer orders to be known with certainty. Accordingly, the list of requirements can be narrowed down to four criteria as depicted in Fig. 7. Amongst these criteria complexity and model-mix aspects have been pivotal to the development of existing approaches. Areas for future work regard the modeling of capacity and the incorporation of multiple criteria.

- *Capacity*: As compared to central planning, more detailed representations of capacity are used within decentral planning. However, only one work explicitly addresses the distribution of orders among multiple parallel lines. Vertical interdependencies among departments are not considered at all. The same holds true for flexibility to adjust capacities. Similar to central planning we expect additional potential to improve planning from better coordinating the planning of orders and detailed capacity planning.
- *Multi-criteria*: Most works for decentral planning address the trade-off between the leveling of the model-mix and the corresponding resource utilization as well as customer service criteria. However, there is no mutual consensus on how to incorporate these criteria into optimization. Similar to the central case, empirical studies might be helpful in evaluating the potential of alternative multi-criteria decision making techniques.

Further opportunities to advance the planning of orders relate to supply chain aspects and implicit part requirements. The *supply chain* for a single car model typically involves more than 150 suppliers and 2000 dealers (Fredriksson and Gadde, 2005). Since decisions on orders often affect the profits of different members in the supply chain differently, coordination mechanisms are required to motivate cooperative behavior and truthful information sharing. For the planning to be effective, in particular reliable information on part availability is required on the supply side. The question of how to encourage this information sharing is a field for future work. A first step into this direction is described by Oh et al. (2010). For a BTS setting, the study proposes a simplified version of a capacitated production planning model which integrates trust into procurement decisions. On the demand side, issues of risk sharing with dealers are of high relevance for BTO automotive production. If demand is lower than capacity, the OEM might want to stimulate additional demand. If capacity is scarce, it is in the interest of the OEM to accept only those orders that maximize profits. Currently quota systems are used in industry to coordinate the sales channel (Meyr, 2004). However, these systems are prone to inefficiencies. This in particular holds true, if the dealers have heterogeneous cost structures or demand is not perfectly correlated at dealer level. A second opportunity for future work arises from *implicit requirements* and reconfiguration flexibility. Current models for the planning of orders are either based on output oriented product representations based on options or directly consider station workload. Part requirements are not explicitly considered. The models therefore do not allow for the anticipation of bottlenecks which result from the combination of certain product options (e.g. seat heatings in combination with high power radio). In addition to that, current models assume demand to be inflexible. There may be degrees of freedom to substitute certain product options in case of bottlenecks. Integrating these degrees of freedom increases flexibility and should therefore improve performance.

7. Conclusions

In this paper we have developed a comprehensive overview of OR models and applications for the planning of capacities and orders in BTO automobile production. From the review of works we conclude that the topic constitutes a recent field of research which is of high practical relevance and still evolves dynamically. The existing body of literature comprises 49 works, the vast majority of which has been published in the last decade, more than one half within the last four years. Three quarter of the works (36 publications) regard industrial cases.

Amongst the works considered have been 26 from the German speaking community. From our point of view this provides evidence that approaches for production management in the automotive industry are particularly well developed in this community. The dissemination of the research results varies. Nine works have exclusively been published in German dissertations yet. The main characteristic of these works is a very detailed analysis of the real world planning situation. In order to make the general ideas behind these works assessable for the international research community, we have extended the review to include these dissertations.

From a thematic point of view, the review paints a mixed picture. While consistent streams of literature can be identified for selected planning tasks, the tasks itself appear to be rather isolated from each other. And even within most tasks we could not identify evidence for a common understanding of the problems. Most approaches focus on a single or a few selected requirements. In addition to that, frequently specific requirements are incorporated, often based on some empirical analysis. This in particular becomes evident with respect to integrative approaches for the planning of capacities and in works on central planning. As a consequence, short streams of literature result, in some cases singular works. Exceptions are the domains of flexibility planning and decentral planning. Due to the high degree of abstraction (flexibility) or the clearly confined problem setting (decentral planning) the works better integrate with each other. A future challenge will be to better integrate the approaches and the domains of research. This will allow for better complying with the requirements of industry.

In showing similarities and differences between existing works and planning tasks, the review aims at contributing towards a common understanding of production management in the automotive industry. We likewise hope to support readers facing a particular problem in identifying suitable modeling techniques and efficient solution approaches. Several open issues have been identified. These comprise issues for research (e.g., more comprehensive model formulations, solution approaches) as well as issues which essentially determine the applicability and benefit in industry (e.g., incorporation of organizational constraints, data management). From this we conclude that planning of capacities and orders in the automotive industry is a topic that deserves further research efforts.

Acknowledgements

The authors sincerely thank the co-ordinating editor and three anonymous referees for their extensive, very constructive and helpful comments on the earlier version of the manuscript which helped to improve the paper.

References

ACEA, 2010. The Automotive Industry Pocket Guide 2010.

Alden, J.M., Costy, T., Inman, R.R., 2002. Product-to-plant allocation assessment in the automotive industry. Journal of Manufacturing Systems 21 (1), 1–13.

Altemeier, S., Helmdach, M., Koberstein, A., Dangelmaier, W., 2010. Reconfiguration of assembly lines under the influence of high product variety in the automotive industry – a decision support system. International Journal of Production Research 48 (21), 6235–6256.

Askar, G., Sillekens, T., Suhl, L., Zimmermann, J., 2007. Flexibility planning in automotive plants. In: Günther, H.-O., Mattfeld, D.C., Suhl, L. (Eds.), Management logistischer Netzwerke. Entscheidungsunterstützung, Informationssysteme und OR-Tools. Physica-Verlag, Heidelberg, pp. 235–255.

Askar, G., 2008. Optimierte Flexibilitätsnutzung in Automobilwerken. Dissertation Technische Universität Clausthal. Shaker, Aachen.

Becker, C., Scholl, A., 2006. A survey on problems and methods in generalized assembly line balancing. European Journal of Operational Research 168 (3), 694–715.

Benyoucef, L., Jain, V., Rezg, N., Xie, X., 2007. Production planning in an assembly system with social and industrial constraints: a dynamic programming approach. International Journal of Services Operations and Informatics 2 (2), 131–151.

Bihlmaier, R., Koberstein, A., Obst, R., 2009. Modeling and optimizing of strategic and tactical production planning in the automotive industry under uncertainty. OR Spectrum 31 (2), 311–336.

Biller, S., Bish, E., Muriel, A., 2002. Impact of manufacturing flexibility on supply chain performance in the automotive industry. In: Song, J.-S., Yao, D.D. (Eds.), Supply Chain Structures. Springer, Berlin, pp. 73–118.

Biller, S., Chan, L., Simchi-Levi, D., Swann, J., 2005. Dynamic pricing and the direct-to-customer model in the automotive industry. Electronic Commerce Research 5 (2), 309–334.

Biller, S., Swann, J., 2006. Pricing for environmental compliance in the auto industry. Interfaces 36 (2), 118–125.

Bish, E.K., Muriel, A., Biller, S., 2005. Managing flexible capacity in a make-to-order environment. Management Science 51 (2), 167–180.

Bixby, R.E., 2002. Solving real-world linear programs. Operations Research 50 (1), 3–15.

Bolat, A., 2003. A mathematical model for selecting mixed models with due dates. International Journal of Production Research 41 (5), 897–918.

Boyer, K.K., Leong, G.K., 1996. Manufacturing flexibility at the plant level. Omega 24 (5), 495–510.

Boysen, N., Fliedner, M., Scholl, A., 2009a. Sequencing mixed-model assembly lines: survey, classification and model critique. European Journal of Operational Research 192 (2), 349–373.

Boysen, N., Fliedner, M., Scholl, A., 2009b. Production planning of mixed-model assembly lines: overview and extensions. Production Planning & Control 20 (5), 455–471.

Boysen, N., 2005. Produktionsprogrammplanung bei Variantenfließfertigung. Zeitschrift für Planung und Unternehmenssteuerung 16 (1), 53–72.

Brabazon, P.G., MacCarthy, B., 2006. Fundamental behaviour of virtual-build-to-order systems. International Journal of Production Economics 104 (2), 514–524.

Brabazon, P.G., MacCarthy, B.L., 2010. On Markovian approximations for virtual-build-to-order systems. Journal of the Operational Research Society 61 (10), 1471–1484.

Brabazon, P.G., MacCarthy, B., Woodcock, A., Hawkins, R.W., 2010. Mass Customization in the Automotive Industry: comparing interdealer trading and reconfiguration flexibilities in order fulfillment. Production and Operations Management 19 (5), 489–502.

Breitman, R.L., Lucas, J.M., 1987. PLANETS: a modeling system for business planning. Interfaces 17 (1), 94–106.

Brückner, F., 2010. Der Audi A1 macht uns finanziell richtig viel Spaß. <http://www.handelsblatt.com/unternehmen/industrie/der-audi-a1-macht-uns-finanziell-richtig-viel-spass/3487830.html>.

Bundschuh, M., 2008. Modellgestützte strategische Planung von Produktionssystemen in der Automobilindustrie. Dissertation Universität Augsburg. Verlag Dr. Kovac, Hamburg.

Chandra, C., Eversonb, M., Grabisa, J., 2005. Evaluation of enterprise-level benefits of manufacturing flexibility. Omega 33 (1), 17–32.

Childerhouse, P., Disney, S.M., Towill, D.R., 2008. On the impact of order volatility in the European automotive sector: special section on competitive advantage through global supply chains. International Journal of Production Economics 114 (1), 2–13.

Dijkstra, E.W., 1959. A note on two problems in connection with graphs. Numerische Mathematik 1, 269–271.

Ding, F.-Y., Tolani, R., 2003. Production planning to support mixed-model assembly. Computers & Industrial Engineering 45 (3), 375–392.

Dörmer, J., Günther, H.O., Gujjula, R., Friedrich, K., 2010. Master production scheduling for high-variant mixed-model assembly lines. In: EurOMA (Ed.), Proceedings of the 17th International Annual EurOMA Conference, Porto.

Eppen, G.D., Martin, R.K., Schrage, L., 1989. A scenario approach to capacity planning. Operations Research 37 (4), 517–527.

Escudero, L.F., Galindo, E., Garcia, G., Gomez, E., Sabau, V., 1999. Schumann, a modeling framework for supply chain management under uncertainty. European Journal of Operational Research 119 (1), 14–34.

Fleischmann, B., Ferber, S., Henrich, P., 2006. Strategic planning of BMWs global production network. Interfaces 36 (3), 194–208.

Florian, M., Kemper, J., Sihn, W., Hellingrath, B., 2010. Concept of transport-oriented scheduling for reduction of inbound logistics traffic. In: 43rd CIRP International Conference on Manufacturing Systems, Vienna.

Francas, D., Kremer, M., Minner, S., Friese, M., 2009. Strategic process flexibility under lifecycle demand. International Journal of Production Economics 121 (2), 427–440.

Francas, D., Löhndorf, N., Minner, S., 2011. Machine and labor flexibility in manufacturing networks. International Journal of Production Economics 131 (1), 165–174.

Fredriksson, P., Gadde, L.-E., 2005. Flexibility and rigidity in customization and build-to-order production. Industrial Marketing Management 34 (7), 695–705.

Friese, M., 2008. Planung von Flexibilitäts- und Kapazitätsstrategien für Produktionsnetzwerke der Automobilindustrie. Dissertation Leibniz Universität Hannover. PZH Produktionstechnisches Zentrum, Garbsen.

Gans, J.E., 2008. Neu- und Anpassungsplanung der Struktur von getakteten Fließproduktionssystemen für variantenreiche Serienprodukte in der Montage. Dissertation Universität Paderborn. Paderborn.

Garcia-Sabater, J.P., Maheut, J., Garcia-Sabater, J.J., 2011. A two-stage sequential planning scheme for integrated operations planning and scheduling system using MILP. Flexible Services and Manufacturing Journal, online first.

Gebhardt, J., Rugheimer, F., Detmer, H., Kruse, R., 2004. Adaptable Markov models in industrial planning. IEEE International Conference on Fuzzy Systems. Proceedings of IEEE Service Center, Piscataway, NJ, pp. 475–479.

Gneiting, P.R., 2009. Supply Chain Design für modulare Fahrzeugarchitekturen. Dissertation ETH Zürich. Zürich.

Goetschalckx, M., Vidal, C.J., Dogan, K., 2002. Modeling and design of global logistics systems: a review of integrated strategic and tactical models and design algorithms. European Journal of Operational Research 143 (1), 1–18.

Graf, H., 2006. Innovative logistics is a vital part of transformable factories in the automotive industry. In: Dashchenko, A.I. (Ed.), Reconfigurable Manufacturing Systems and Transformable Factories. Springer, Berlin, Heidelberg, pp. 423–457.

Graves, S.C., Tomlin, B.T., 2003. Process flexibility in supply chains. Management Science 49 (7), 907–919.

Gruß, C.M.F., 2008. Revenue-Management-Konzept in der Automobilindustrie: Vorgehenskonzept zur online Neuwagendirektdistribution und Produktionsharmonisierung. Dissertation Universität Cottbus. Gabler, Wiesbaden.

Gunasekaran, A., Ngai, E.W.T., 2005. Build-to-order supply chain management: a literature review and framework for development. Journal of Operations Management 23 (5), 423–451.

Gunasekaran, A., Ngai, E.W.T., 2009. Modeling and analysis of build-to-order supply chains. European Journal of Operational Research 195 (2), 319–334.

Hayler, C., 1999. Ein regelbasiertes System zur Generierung von Orders für Lagerfahrzeuge: Fallstudie bei einem deutschen Automobilhersteller. Dissertation Friedrich-Schiller Universität Jena. Jena.

Henrich, P., 2002. Strategische Gestaltung von Produktionssystemen in der Automobilindustrie. Shaker, Aachen.

Hindi, K.S., Ploszajski, G., 1994. Formulation and solution of a selection and sequencing problem in car manufacture. Computers & Industrial Engineering 26 (1), 203–211.

Hintsches, A., Spengler, S.T., Volling, T., Wittek, K., Priegnitz, G., 2010. Revenue management in make-to-order manufacturing: case study of capacity control at ThyssenKrupp VDM. BuR – Business Research 3 (2), 173–190.

Holweg, M., Miemczyk, J., 2002. Logistics in the three-day car age: assessing the responsiveness of vehicle distribution logistics in the UK. International Journal of Physical Distribution and Logistics Management 32 (10), 829–850.

Holweg, M., Pil, F.K., 2001. Successful build-to-order strategies start with the customer. MIT Sloan Management Review 43 (1), 74–83.

Holweg, M., Disney, S.M., Hines, P., Naim, M.M., 2005. Towards responsive vehicle supply: a simulation-based investigation into automotive scheduling systems. Journal of Operations Management 23 (5), 507–530.

Inman, R.R., Jordan, W.C., 1997. Integrated assembly line loading, design, and labor planning. Journal of Manufacturing Systems 16 (5), 315–322.

Inman, R.R., Gonsalvez, D.J.A., 2001. A mass production product-to-plant allocation problem. Computers & Industrial Engineering 39 (3-4), 255–271.

Jordan, W.C., Graves, S.C., 1995. Principles on the benefits of manufacturing process flexibility. Management Science 41 (4), 577–594.

Kabak, Ö., Ülengin, F., 2011. Possibilistic linear-programming approach for supply chain networking decisions. European Journal of Operational Research 209 (3), 253–264.

Kauder, S., Meyr, H., 2009. Strategic network planning for an international automotive manufacturer. OR Spectrum 31 (3), 507–532.

Kauder, S., 2008. Strategische Planung internationaler Produktionsnetzwerke in der Automobilindustrie. Dissertation Wirtschaftsuniversität Wien. Wien.

Koether, R., 1986. Verfahren zur Verringerung von Modell-Mix-Verlusten in Fliessmontagen. Dissertation Universität Stuttgart. Springer, Berlin.

Kohler, K., 2009. Global supply chain design. Dissertation Julius-Maximilians-Universität. Würzburg.

MacCarthy, B., Brabazon, P.G., Bramham, J., 2003. Fundamental modes of operation for mass customization. International Journal of Production Economics 85 (3), 289–304.

Mansouri, A.S., Gallear, D., Askariazad, M.H., 2012. Decision support for build-to-order supply chain management through multiobjective optimization. International Journal of Production Economics 135 (1), 24–36.

Meyr, H., 2004. Supply chain planning in the German automotive industry. OR Spectrum 26 (4), 447–470.

Oh, S., Ryu, K., Moon, I., Cho, H., Jung, M., 2010. Collaborative fractal-based supply chain management based on a trust model for the automotive industry. Flexible Services and Manufacturing Journal 22 (3-4), 183–213.

Peidro, D., Mula, J., Poler, R., Verdegay, J.-L., 2009. Fuzzy optimization for supply chain planning under supply, demand and process uncertainties. Fuzzy Sets and Systems 160 (18), 2640–2657.

Peidro, D., Mula, J., Jiménez, M., del Mar Botella, M., 2010. A fuzzy linear programming based approach for tactical supply chain planning in an uncertainty environment. European Journal of Operational Research 205 (1), 65–80.

Pil, F.K., Holweg, M., 2004. Linking product variety to order-fulfillment strategies. Interfaces 34 (5), 394–403.

Rahn, C., 2011. Audi buyers wait months as factories struggle to keep up with orders. <http://www.bloomberg.com/news/2011-02-08/audi-buyers-wait-months-as-factories-at-full-steam-struggle-to-keep-pace.html>.

Reichhart, A., Holweg, M., 2007. Creating the customer-responsive supply chain: a reconciliation of concepts. International Journal of Operations and Production Management 27 (11), 1144–1172.

Roscher, J., 2008. Bewertung von Flexibilitätsstrategien für die Endmontage in der Automobilindustrie. Dissertation Universität Stuttgart. Stuttgart.

Santoso, T., Ahmed, S., Goetschalckx, M., Shapiro, A., 2005. A stochastic programming approach for supply chain network design under uncertainty. European Journal of Operational Research 167 (1), 96–115.

Sethi, A.K., Sethi, S.P., 1990. Flexibility in manufacturing: a survey. International Journal of Flexible Manufacturing Systems 2 (4), 289–328.

Sillekens, T., 2008. Aggregierte Produktionsplanung in der Automobilindustrie unter besonderer Berücksichtigung von Personalflexibilität. Dissertation Universität Paderborn. Paderborn.

Sillekens, T., Koberstein, A., Suhl, L., 2011. Aggregate production planning in the automotive industry with special consideration of workforce flexibility. International Journal of Production Research 49 (17), 5055–5078.

Slack, N., 1993. Flexibility as a manufacturing objective. International Journal of Operations and Production Management 3 (3), 4–13.

Stautner, U., 2001. Kundenorientierte Lagerfertigung im Automobilvertrieb: ergänzende Ansätze zum Supply-chain-Management. Dissertation Universität Göttingen. Dt. Univ.-Verl., Wiesbaden.

Stephan, H.A., Gschwind, T., Minner, S., 2010. Manufacturing capacity planning and the value of multi-stage stochastic programming under Markovian demand. Flexible Services and Manufacturing Journal 22 (3-4), 143–162.

Voigt, K.I., Saatmann, M., Schorr, S., 2008. Flexibility and revenue management in the automotive industry. Journal of Enterprise Information Management 21 (4), 424–439.

Volling, T., 2009. Auftragsbezogene Planung bei variantenreicher Serienproduktion, Dissertation Technische Universität Braunschweig. Gabler, Wiesbaden.

Volling, T., Spengler, T.S., 2011. Modeling and simulation of order-driven planning policies in build-to-order automobile production. International Journal of Production Economics 131 (1), 183–193.

Waller, B., 2002. Sales channel for build to order. In: Sponsor Report The International Car Distribution Programme, Solihull.

Waller, B., 2004. Market responsive manufacturing for the automotive supply chain. Journal of Manufacturing Technology Management 15 (1), 10–19.

Walter, M., Sommer-Dittrich, T., Zimmermann, J., 2011. Evaluating volume flexibility instruments by design-of-experiments methods. International Journal of Production Research 49 (6), 1731–1752.

Williams, G., 2000. Progress towards customer pull distribution. Research paper 4/2000, The International Car Distribution Programme, Solihull.

Wittek, K., Volling, T., Spengler, T., Gundlach, F.-W., 2011. Tactical planning in flexible production networks in the automotive industry. In: Hu, B., Morasch, K., Siegle, M., Pickl, S. (Eds.), Operations Research Proceedings 2010. Springer, Berlin, pp. 429–434.

Xu, W.L., Nelson, B.L., Hopp, W.J., Owen, J.H., 2009. Assessing the benefits of integrated vehicle life cycle planning. <http://users.iems.northwestern.edu/~nelsonb/Publications/XuHoppNelsonOwenDraft3.pdf>.

Zhang, G., Shang, J., Li, W., 2011. Collaborative production planning of supply chain under price and demand uncertainty. European Journal of Operational Research 215 (3), 590–603.