# Project-based Production Planning for Modular Products in Configure-to-order Environments

Description of the planning problem

---

## 1 Disclaimer

Lieber Florian,
das hier beschriebene Planungsproblem ist als multi-mode RCMPSP formuliert und entspricht somit nicht exact dem RCMPSP, welches wir mit Dir anschauen möchten. Allerdings kannst Du diese Modellbeschreibung nutzen, um ein grundsätzliches Verständnis des Problems zu erlangen und um Dir die Notation in der Beispiel-Instanz klar zu machen.
Viele Grüße, Franziska

## 2 Definition of key words

| | |
|---|---|
| Product | Each customer order consists of one product that is derived from the customer requirements. The product is specified in a configure-to-order (CTO) process and contains one or multiple sub-products. |
| Sub-product | A sub-product is part of a product that provides a certain function to the product. |
| Module | A module is a part of the sub-products with a predefined function, interfaces, installation space, bill of material, and routing. |
| Module kit | A module kit contains all existing modules. Standardized interfaces and installation space allow modules to be combined flexibly. |
| Product structure | The product structure is a tree representation of the product portfolio. It defines how modules from the module kit can be combined to achieve a product. |
| Project | The production process of a customer order is depicted as a single project in which one product is considered. Since multiple customer orders have to be processed at a certain time, a multi-project environment needs to be considered. |
| Activity (Master) | An activity depicts a certain work content that needs to be processed without preemption. Activities have to be scheduled, leading to defined activity start and finish times. For each activity, the duration and the resource requirements are given. |
| Production area | A production area is one part of the production with a dedicated set of workstations and workers who process the activities. The organization of the production areas can follow different structures, such as workshop assemblies or mixed-model assembly lines, which are considered in this work. |
| Synchronization | The synchronization is the activity in the production process before the delivery to the customer. Here, products of one solution are integrated and tested. |
| Resource | The resources are renewable and are assigned to one or more production areas. Each resource is able to perform a set of activities. |
| Activity (Sub) | In the granular sub-problems, the activities already considered in the master problem are subdivided into activities providing more detailed information on the required equipment, necessary skills of the workers, and the used non-renewable resources. Furthermore, the duration of the activities is deterministic and given. |

## 3 Planning problem for CTO products

The considered production planning problem is a two-level problem with one master problem that schedules activities in a multi-project environment with constrained resources and multiple sub-problems. The master problem and the sub-problems are hierarchically connected through the activity schedule. This hierarchy also depicts two levels of information granularity. Early in the planning process, only limited information on the activities is given; thus, the master problem has to work with aggregate information. Later, the details were clarified, and more specific information on the activities can be used for operational planning in the sub-problems. The master problem determines the start and finish times of activities that need to be respected in the sub-problems. Two main types of sub-problems that can occur multiple times in the production processes are the scheduling of activities in a workshop assembly and the sequencing of sub-products in a mixed-model assembly line. The master problem is located at a tactical planning level with a time horizon of several months and uses aggregated information. The sub-problems are operational planning problems with a shorter planning horizon, more granular planning periods, and more detailed information.

A project is used to plan the production process for a customer order. Since multiple customer orders have to be processed at a certain time, a multi-project environment needs to be considered. These projects are depicted as activity-on-nodes (AON) networks. The release and due dates for the projects are given and result from the order arrival and customer requirements. For each activity, the duration and the resource requirements are given. The duration is deterministic but may depend on the production area in which the activity is processed. Each customer order is defined as one product. The product can contain multiple sub-products that are specified through the configure-to-order (CTO) process, leading to multiple possible product configurations. After the customer order is confirmed, order-specific engineering work is needed to specify the technical details. Only thereafter, the Bill-of-Materials (BOM) and the work plan can be determined.

Each customer order is organized as described in the product structure, which represents the entire product portfolio. It defines how modules from the module kit are combined to achieve the sub-products within a product. The considered product structure can be depicted as a tree structure resulting in multi-stage BOMs that follow the manufacturing structure. Consequently, the product structure can be flipped and used as a base for the project network, resulting in outtree-intree structures (see Brucker (2007)). Additional activities, such as project management and the project start and finish nodes, have to be added to the project network.

The production system consists of several production areas. A production area is one part of the production with a dedicated set of workstations and workers who process the activities. The organization of the production areas can follow different structures, such as workshop assemblies or mixed-model assembly lines, which are considered in this work. While a tactical perspective enables coordination between the production areas, the operational planning is done for each production area individually. Since the production areas may be located on different sides, a transfer time as a placeholder for transportation effort between the production areas needs to be respected. Figure 1 shows an example of a production system with multiple production areas. In the example, the assembly of each product starts in workshop assembly 1 before being passed to one of the following production areas. After the product has passed either one of the mixed-model assembly lines or workshop assembly 2, it is always finished in workshop assembly 3.

The resources in the master problem are renewable and are assigned to one or more production areas. Each resource is able to perform a set of activities. For each production system, the capacity needs to be considered on an aggregated level. Uncertainties in the considered planning problems arise from the stochastic arrival of orders and the actual delivery dates of the supplier parts. The objective of the master problem is to schedule the activities of the projects in a way that reduces the weighted lateness of the projects.

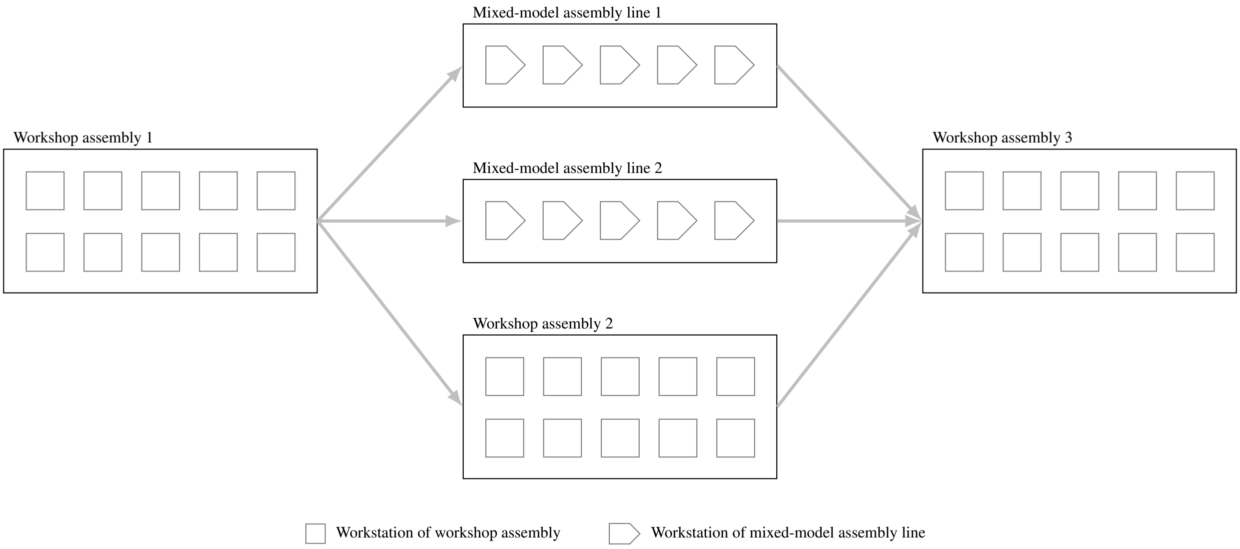


**Fig. 1:** Production system with multiple production areas

# 4 Model

## *4.1 Tactical problem (Master problem)*

Let $\mathscr{P}$ be a set of projects with a set of activities $\mathscr{V}_p$, which need to be scheduled over the horizon $T$ of several months. The planning horizon is divided into individual periods $t$ representing weeks. Set $\mathscr{V}$ contains the activities and is the unification set of the activities of project $p$ that are given in the sets $\mathscr{V}_p$ as $\mathscr{V} = \bigcup_{p \in \mathscr{P}} \mathscr{V}_p$. The finish activities, which are the activities without successors in the sets $\mathscr{V}_p$, are grouped in set $\mathscr{V}^F$. Each activity can be processed in multiple modes, which depict the processing in different production areas. For each activity $j$, the set $\mathscr{M}_j$ contains all modes $m$ in which the activity might be executed. The set $\mathscr{M}_a^A$ contains the modes referring to production area $a \in \mathscr{A}$. The finish-start precedence constraints that connect the activities $i$ and $j$ from the set $\mathscr{V}$ are summarized in the set $\mathscr{E}$. The set $\mathscr{W}$ consists of pools of workers with different skills. Each worker is part of only one pool. The workers of one pool are qualified to work in specific production areas. The capacity of these pools $w$ has to be assigned to the available production areas $\mathscr{A}$. The assigned capacity $q_{a,w,t}$ of pool of workers $w$ to production area $a$ in period $t$ are defined as continuous variables. The start of the activities and their assignment to a specific production area is represented by the step variable $x_{j,m,t}$ being 1 if activity $j$ is processed in mode $m$ and started at the beginning of period $t$ or before. The definition of this step variable and step variables in the following problem formulation in this work follow Artigues, Koné, Lopez, and Mongeau (2015). The objective $z_{Master}$ (1) minimizes the weighted lateness of the projects, which is measured as earliness $e_j$ and tardiness $t_j$ with weight $\alpha_j$ for each activity in the project $p \in \mathscr{P}$. The weight $\alpha_j$ depicts the different importance of the activities in which strategic aspects, such as the strengthening of specific customer relations, can be considered. $\beta^e$ and $\beta^t$ reflect the penalties for earliness and tardiness of the projects.

| | |
|---|---|
| **Indices** | |
| $p$ | Project |
| $j$ | Activity |
| $m$ | Mode |
| $a$ | Production area |
| $w$ | Pool of workers |
| $t$ | Planning period |
| **Sets** | |
| $\mathscr{V}$ | Set of activities |
| $\mathscr{V}^F$ | Set of finish activities of the projects |
| $\mathscr{V}_m$ | Set of activities executable in mode $m$ |
| $\mathscr{M}_j$ | Set of modes of activity $j$ |
| $\mathscr{A}$ | Set of production areas |
| $\mathscr{E}$ | Set of precedence relations between activities |
| $\mathscr{W}$ | Set of pools of workers |
| $T$ | Set of planning periods |
| **Parameters** | |
| $p_{j,m}$ | Processing time of activity $j$ |
| $\alpha_j$ | Weight for activities $j \in \mathscr{V}^F$ |
| $\beta^e, \beta^t$ | Penalty for earliness and tardiness |
| $d_j$ | Due date of activity $j \in \mathscr{V}^F$ |
| $r_{j,w,m}$ | Resource requirement of activity $j$ for pool of workers $w$ in mode $m$ |
| $\overline{q}_{w,a}$, $\underline{q}_{w,a}$ | Upper bound and lower bound for allocation of the pool of workers $w$ in production area $a$ |
| $b_{w,t}$ | Capacity of worker pool $w$ at time $t$ (measured in hours) |
| $ES_j, LF_j$ | Earliest start and latest finish time of activity $j$ |
| $\delta_{m,n}$ | Transfer time if activity is processed in mode $n$ while predecessor was processed in mode $m$ |
| $T_{max}$ | Planning horizon of several months |
| **Continuous variables** | |
| $e_j, t_j$ | Earliness and tardiness of activity $j$ |
| $q_{a,w,t}$ | Capacity of worker pool $w$ allocated to production area $a$ in period $t$ (measured in hours) |
| **Binary variables** | |
| $x_{j,m,t}$ | $\begin{cases} 1, & \text{if activity } j \text{ starts in mode } m \text{ at the beginning of period } t \text{ or before (step variable)} \\ 0, & \text{otherwise} \end{cases}$ |

**Table 1:** Summary of notation — Master problem

$$z_T = \min \quad \sum_{j \in \mathscr{V}^F} \alpha_j \cdot (\beta^e \cdot e_j + \beta^t \cdot t_j) \tag{1}$$

The quantity of the renewable resources $q_{a,w,t}$ being assigned to the production area $a$ has to cover the demand of all activities that are processed in period $t$ as defined in Constraint (2). Constraints (3) and (4) ensure that the resource quantity $q_{a,w,t}$ assigned to $a$ holds the upper and lower bounds $\overline{q}_{w,a}, \underline{q}_{w,a}$ for resource $r$ and production area $a$ in all time periods $t$. The lower bound reflects the assumption that a production area requires a specific minimum capacity to function correctly. In contrast, the upper bound represents the maximum capacity of the production area, which may be limited for space reasons, for example. Additionally, Constraint (5) ensures that the total quantity of the resources $q_{a,w,t}$ assigned to the production areas $a \in \mathscr{A}$ does not exceed the total available amount $b_{w,t}$ for all pools of workers $w \in \mathscr{W}$ and in each period.

$$\sum_{j \in \mathscr{V}_m} r_{j,w,m} \cdot (x_{j,m,t} - x_{j,m,t-p_{j,m}}) \leq q_{a,w,t} \qquad \forall a \in \mathscr{A}, m = a, w \in \mathscr{W}, t \in T \tag{2}$$

$$q_{a,w,t} \leq \overline{q}_{w,a} \qquad \forall a \in \mathscr{A}, w \in \mathscr{W}, t \in T \tag{3}$$

$$q_{a,w,t} \geq \underline{q}_{w,a} \qquad \forall a \in \mathscr{A}, w \in \mathscr{W}, t \in T \tag{4}$$

$$\sum_{a \in \mathscr{A}} q_{a,w,t} \leq b_{w,t} \qquad \forall w \in \mathscr{W}, t \in T \tag{5}$$

Further, Constraint (6) depicts the minimal time lag between the start of the activities as disaggregated precedence constraint (see Artigues et al. (2015)). We use the disaggregated formulation of this constraint as it leads to tighter bounds of the LP-relaxation (see ....), which is helpful when solving the MIP. Constraint (7) defines that each activity $j \in \mathscr{V}$ must not be started before the earliest start $ES_j$ while it must have been started in the respective latest start period $LS_j$ in one of the available modes $m \in \mathscr{M}_j$ as described in Constraint (8). Further, Constraint (9) ensures that the variables $x_{j,m,t}$ only can change from 0 to 1, which describes the characteristics of a step variable.

$$\sum_{m \in \mathscr{M}_i} x_{i,m,t-p_{i,m}-\delta_{m,n}} - x_{j,n,t} \geq 0 \qquad \forall (i,j) \in \mathscr{E}, n \in \mathscr{M}_j, t \in T \tag{6}$$

$$x_{j,m,ES_j-1} = 0 \qquad \forall j \in \mathscr{V}, m \in \mathscr{M}_j \tag{7}$$

$$\sum_{m \in \mathscr{M}_j} x_{j,m,t} = 1 \qquad \forall j \in \mathscr{V}, t \geq LS_j \tag{8}$$

$$x_{j,m,t} - x_{j,m,t-1} \geq 0 \qquad \forall j \in \mathscr{V}, m \in \mathscr{M}_j, t \in T \tag{9}$$

Constraint (10) calculates the earliness $e_j$ or tardiness $t_j$ of each activity $j$. Constraints (11) - (13) ensure the required value range of the variables.

$$\sum_{m \in \mathscr{M}_j} \sum_{t \in T} (t + p_{j,m}) \cdot (x_{j,m,t} - x_{j,m,t-1}) + e_j - t_j = d_j \qquad \forall j \in \mathscr{V}^F \tag{10}$$

$$e_j, t_j \geq 0 \qquad \forall j \in \mathscr{V}^F \tag{11}$$

$$q_{a,w,t} \geq 0 \qquad \forall a \in \mathscr{A}, w \in \mathscr{W}, t \in T \tag{12}$$

$$x_{j,m,t} \in \{0,1\} \qquad \forall j \in \mathscr{V}, m \in \mathscr{M}_j, t \in T \tag{13}$$

## 5 Instances

To investigate the interaction between tactical and operational scheduling, a new instance set is developed based on the largest available real-world reference instance from the dataset used by Brachmann and Kolisch (2021). The primary goal is to preserve the structural integrity of the original network while introducing additional information for the operational level and controlled variation to key planning parameters. Transformation details can be found in section **??**. The following paragraphs describe the resulting instance structure and parameterization in detail.

### *5.1 Tactical Problem*

At the tactical level, projects with multiple activity execution modes and shared worker types need to be coordinated. Each activity $j$ belongs to a project $p$ and is scheduled in one of several available modes $m$ within a finite planning horizon $T^{max}$. Worker availability is modeled via worker types $w$, which can be flexibly allocated to production areas $a$. Available worker types are shown in Table 2. The set of activities $\mathscr{V}$ is complemented by a set of precedence relations $\mathscr{E}$, a set $\mathscr{V}^F$ indicating project finish activities, and a mapping $\mathscr{M}_j$ specifying available modes per activity. Modes are directly linked to production areas.

For each activity $j$ and selected mode $m$, a processing time $p_{j,m}$ is specified. Processing times $p_{j,m}$ are derived from expert knowledge and historical data. Pre-assembly activities have durations from the set {2,3} weeks, with each duration having the same probability. For machine assembly, durations are aligned with assembly line capacities and therefore amount to 6 weeks. The synchronization activities have durations of 6 weeks. The activity durations remain the same across different modes. Start times ($ES_j$) and finish times ($LS_j$) are computed using the precedence relations in $\mathscr{E}$. Due dates $d_j$ are defined for all finish activities, accompanied by a weight $\alpha_j$, reflecting the project's importance, and penalty coefficients $\beta^e = 1$ and $\beta^t = 2$ for earliness

| w | Worker Type |
|---|---|
| 1 | Project planning |
| 2 | Project management |
| 3 | Mechanical engineering hardware |
| 4 | Electrical engineering |
| 5 | Mechanical pre-assembly |
| 6 | Electrical pre-assembly |
| 7 | Blister machine assembly |
| 8 | Cartoner assembly |
| 9 | Electrical engineering software |
| 10 | Tools engineering |
| 11 | Tools assembly |
| 12 | Layout design |
| 13 | Synchronization |

**Table 2:** Overview of worker types

and tardiness, respectively. The resource demand $r_{j,w,m}$ denotes the requirement of worker type $w$ in mode $m$, while the time-dependent availability of worker types is given by $b_{w,t}$. Production areas impose upper and lower capacity bounds $\overline{q}_{w,a}$ and $\underline{q}_{w,a}$, which govern the allocation of worker capacity. Dependencies between the activities, worker types and production areas are depicted in Table 3. The temporal structure is further defined by transfer times $\delta_{m,n}$ to model logistical effort between execution modes. Transfer times $\delta_{m,n}$ are set to one for production areas in a remote location and zero otherwise.

| **Mode** $m$ | **Production Area** $a$ | **Activity Type** | **Required Worker Type** $w$ | **Processing Time** $p_{j,m}$ **(weeks)** |
|---|---|---|---|---|
| 1 | Assembly workshop pre-assembly (building 1) | Pre-assembly | Mechanical pre-assembly (5), Electrical pre-assembly (6) | 2 or 3 |
| 2 | Assembly workshop pre-assembly (building 2) | Pre-assembly | Mechanical pre-assembly (5), Electrical pre-assembly (6) | 2 or 3 |
| 3 | Assembly workshop pre-assembly (abroad) | Pre-assembly | Mechanical pre-assembly (5), Electrical pre-assembly (6) | 2 or 3 |
| 4 | Assembly line blister machines | Machine assembly | Blister machine assembly (7) | 6 |
| 5 | Assembly line cartoners | Machine assembly | Cartoner assembly (8) | 6 |
| 6 | Workshop special machines | Machine assembly | Blister machine assembly (7), Cartoner assembly (8) | 6 |
| 7 | Synchronization workshop (building 3) | Synchronization | Synchronization (13) | 6 |
| 8 | Synchronization workshop (building 4) | Synchronization | Synchronization (13) | 6 |

**Table 3:** Relation of modes, production areas, activity types, worker types, and processing times

The key decision variables include the step variable $x_{j,m,t}$, which indicates whether an activity starts in mode $m$ at or before period $t$, and the continuous allocation variable $q_{a,w,t}$, which assigns worker capacity to production areas. Earliness $e_j$ and tardiness $t_j$ control the projects' finish time.

### 5.2 Systematic Parameter Variation

To generate a diverse and informative set of test instances, a systematic parameter variation is applied following a ceteris paribus approach. Each instance is constructed by fixing all but one parameter and varying the remaining one across three predefined levels. This controlled variation allows for isolating the effects of individual planning parameters on the model behavior and solution quality. Six parameters are considered in the variation process: number of projects, due date tightness, worker type scarcity, mode availability, and worker type allocation flexibility. All parameters are discretized into three levels, as shown in Table 4. Each level induces specific structural adaptations during instance generation.

The number of projects determines the scale of the instance. Depending on the selected value (20, 40, or 60), between 5 and 36 activities are assigned per project, resulting in 306, 666, or 945 tactical activities, respectively. Due date tightness influences the temporal constraints on project completion. For each project $p \in \mathscr{P}$, the due date is computed by scaling the critical path duration with a factor from the set $\{1.1, 1.2, 1.3\}$ and adding it to the project release time. This results in more or less restrictive due dates, affecting the feasible scheduling windows. Worker type scarcity modifies the overall availability of worker capacity. The capacity $b_{w,t}$ for each worker type $w$ and period $t$ is scaled by a scarcity factor from $\{1.3, 1.35, 1.4\}$. Higher scarcity levels reduce capacity availability and increase the likelihood of resource conflicts. Mode availability controls the flexibility of the activity assignment to production areas. For each activity $j$, the set of feasible modes $\mathscr{M}_j$ is reduced according to the selected availability level. In the high setting, all modes remain available. The medium setting retains on average 50% of the modes per activity, while the low setting restricts each activity to exactly one mode. The filtering is performed randomly but with uniform probability. Finally, worker pool allocation flexibility affects how worker capacity is distributed across production areas. The upper bounds $\overline{q}_{w,a}$ for resource assignment are scaled by a flexibility factor from $\{1, 1.1, 1.2\}$. This adjustment reflects more or less restrictive area-level assignment constraints and has a direct impact on workforce balancing in the operational problems.

The operational activity tightness governs the level of detail between tactical and operational planning. This parameter controls the slack between the processing time of tactical activities and the aggregate duration of the corresponding operational activities. For each project $p \in \mathscr{P}$, a project-specific tightness factor $u_p$ is sampled from a truncated normal distribution $\mathscr{N}(1, 0.1)$ and restricted to lie within bounds defined by the current tightness level. The bounds are selected from one of three intervals: $[0.95, 1.05]$, $[0.9, 1.1]$, or $[0.85, 1.15]$. Each sampled value $u_p$ is applied to all tactical activities of the corresponding project and scales the resource demand during operational activity generation. This parameter determines the degree of alignment and slack between tactical and operational levels.

| Level | Number of Projects | Due Date Tightness | Worker Type Scarcity | Mode Availability | Worker Type Alloc. Flex. | Op. Activity Slack |
|---|---|---|---|---|---|---|
| 1 | 20 | 1.1 | 1.3 | low | 1.0 | [0.95, 1.05] |
| 2 | 40 | 1.2 | 1.35 | medium | 1.1 | [0.90, 1.10] |
| 3 | 60 | 1.3 | 1.4 | high | 1.2 | [0.85, 1.15] |

**Table 4:** Instance generation parameters and their variation levels

## References

Artigues, C., Koné, O., Lopez, P., & Mongeau, M. (2015). Mixed-integer linear programming formulations. In C. Schwindt & J. Zimmermann (Eds.), *Handbook on project management and scheduling vol.1* (pp. 17–41). Cham: Springer International Publishing.

Brachmann, R., & Kolisch, R. (2021). The impact of flexibility on engineer-to-order production planning. *International Journal of Production Economics*, *239*, 108183. Retrieved from `https://www.sciencedirect.com/science/article/pii/s0925527321001596`

Brucker, P. (2007). *Scheduling algorithms* (5th ed. ed.). Berlin and New York: Springer.