Workload control under diagnosis

Soepenberg, Gerrit Dinant

Publication date:
2010

Chapter 3.

A framework for diagnosing delivery reliability performance of make-to-order companies

3.1. Introduction

Achieving high delivery reliability is increasingly important for make-to-order (MTO) companies. Their production planning and control (PPC) decisions should contribute to achieving this aim. In order to improve PPC decisions, a diagnosis of delivery reliability performance is vital. Generally, the performance diagnosis phase can be regarded as a phase within an improvement cycle where the causes of undesired performance are determined (see Figure 3.1). This paper provides a systematic framework for supporting this phase in order to indicate those PPC decisions that have a negative influence on the delivery reliability of MTO companies. As such, it provides directions for appropriate action in the performance improvement phase.

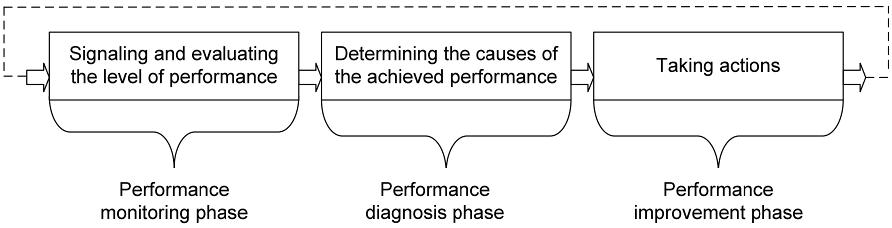

Figure 3.1. Performance monitoring, diagnosis and improvement (adapted from Stoop and Bertrand (1997).

The importance of the performance diagnosis phase in achieving improvements is recognised in the literature on operations management. For example, Neely et al. (1997) note that, in the event of unsatisfactory delivery reliability performance, a company should *"investigate reasons for late delivery, set up problem-solving teams to eliminate root causes"*. Detecting the causes of negative influences on performance in the performance diagnosis phase requires both general problem-solving knowledge as well as domain knowledge (Wagner, 1993). General problem-solving knowledge focuses on acquiring an adequate process to navigate from the achieved performance to the causes underlying that performance (see, for example, Cowan, 1986; Pokras, 1990; Wagner, 1993). Knowledge from the specific domain should be used to gain a clear focus on the diagnosis process, for which insights from the scientific literature can be useful (see, for example, Hopp et al., 2007).

To the best of our knowledge, achieving a structured diagnosis of delivery reliability, incorporating both general problem-solving knowledge and domain knowledge, has not been explicitly considered in the literature. A notable approach has been reported by the Institute of Production Systems and Logistics (IFA) in Hannover. Von Cieminski and Nyhuis (2007) discuss a diagnostic software tool that includes an adapted version of the logistical interrelationship model developed by Lödding (2005). The software user is requested to provide relevant logistical information through an interactive questionnaire. Unfortunately, the diagnosis process is only explained superficially. Nyhuis and Wiendahl (2009) present a quantitative bottleneck-oriented logistic analysis that can be carried out provided quantitative order-progress data are available. They show in detail how a diagnosis can be performed that indicates relevant improvement opportunities for reducing average throughput times (and thus average lateness) with the help of logistic operating curves and throughput diagrams. They also point out that the variance in lateness in a certain production stage could have PPC-related causes, although the steps needed to identify these specific causes are not discussed in detail.

An initial contribution, that specifically focused on diagnosing variances Soepenberg et al., (2008), clearly showed how variance could be linked to specific PPC decisions with the help of order progress diagrams. This new study goes much further and aims to build a comprehensive diagnosis framework for MTO companies, enabling them to navigate from the recorded delivery reliability to the underlying PPC decisions. The framework combines insights from general problem-solving literature with domain-specific literature. The diagnosis makes use of quantitative order-progress data, which ought to be generally available from a company's ERP system. Further, because of time-phased representation in supportive diagnosis diagrams, insights can also be gained into the dynamics involved in MTO companies.

This paper is split in three main sections. Sections 2 and 3 respectively discuss the theoretical background to the framework and the final framework. Section 4 assesses the application of the framework in three different MTO companies. Finally, Section 5 draws conclusions.

3.2. Theoretical backgrounds of framework

This section discusses the theoretical background to the framework. The final framework will be presented in the next section as a diagnosis tree, and incorporates domain-specific content enabling a structured diagnosis process. In developing the process steps in the framework, Wagner (1993) is an important reference. Wagner reviewed the literature in order to find suitable domain-independent diagnosis concepts and strategies to structure the diagnosis process. In this respect, Wagner indicates that a breadth-first strategy should be a distinctive feature of a good diagnosis, and avoids getting bogged down early in the diagnosis process. This suggestion was adopted in developing our framework. In order to provide content for the diagnosis tree, we start with Wagner's suggested upfront system description, combined with a discussion on potential causes of problems based on knowledge from domain-specific PPC literature. This simplifies an effective and complete search to arrive at the root causes within the diagnosis tree. First, we briefly discuss the unravelling of delivery performance in terms of the main indicators. Second, the relevant processes underlying this performance are distinguished, together with the expected influence of relevant PPC decisions related to these processes.

3.2.1. Indicators underlying delivery reliability performance

Undesirable delivery reliability performance can be signalled by many indicators (see, for example, NEVEM-workgroup, 1989). A key indicator often used in practice is the percentage of orders delivered late. Lateness is defined as the difference between the promised delivery time and the realised throughput time of an order. We can distinguish positive lateness (an order delivered late) and negative lateness (an order delivered early). A useful distinction related to the lateness performance indicator found in the literature is its subdivision into average lateness and variance of lateness (see, for example, Baker, 1974; Lödding, 2005). Both indicators can imply a high percentage of late order deliveries (see Figure 3.2). However, lateness distributions B and C in Figure 3.2 will require different approaches to achieve improvements.

It should be noted that the distributions plotted in Figure 3.2 only provide a stationary view of performance. In a practical situation both the average and the variance could change considerably over time, and this should be accounted for in the performance diagnosis. A time-based view on delivery reliability performance can be fostered with the help of supportive diagrams, as have been developed in the literature. In this respect, throughput diagrams (Wiendahl, 1995) and order progress diagrams (Soepenberg et al., 2008) are useful for indicating how the average lateness and variance of lateness respectively evolve over time.

Finally, one should recognise that useful conclusions can only be drawn from the lateness distribution plotted in Figure 3.2 for relatively homogeneous order populations. As such, differences between orders resulting from the market (for example, markets with more or less pressure on delivery times) or from order characteristics (for example, routings, processing times, lot sizes) should be taken into account in the diagnosis. Some of these differences may be noticeable in the lateness distribution (for example, in a distribution with two peaks). In other situations, company characteristics may trigger order subsets (such as having two types of orders that are produced in different departments).

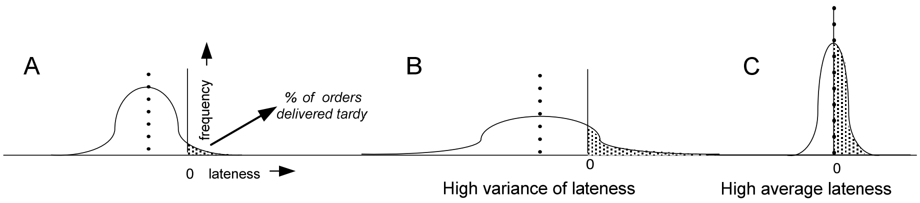

Figure 3.2. The influence of the average lateness and the variance of lateness.

3.2.2. Processes underlying delivery reliability performance

The system under study consists of the process between a delivery date enquiry by a customer and the actual delivery of that order to the customer. Within this process, activities and decisions are carried out that influence the progress of each order. Since these activities and decisions will be case-specific, a disaggregation of the overall process into sub-processes should be structured around generic characteristics that apply to MTO companies.

The disaggregation into sub-processes is based on domain knowledge, i.e., insights from scientific literature on PPC that is related to MTO companies. As a guide, we make use of the structure provided by the workload control (WLC) concept, a concept specifically designed to meet the needs of MTO companies (Stevenson et al., 2005). It distinguishes PPC decisions at three main points in the order flow, of which the order acceptance and the release decisions are the most relevant in our disaggregation.

The remainder of this section focuses on the generic processes that can be distinguished in MTO companies, together with a concise discussion on the PPC decisions involved in these processes. A first level of disaggregation can be found at the point where orders enter the company. This coincides with the order acceptance decision, the first decision distinguished in the WLC concept (Kingsman and Hendry, 2002). Two main sub-processes can be distinguished at this point: the process of promising delivery times and the realisation process (see Figure 3.3)

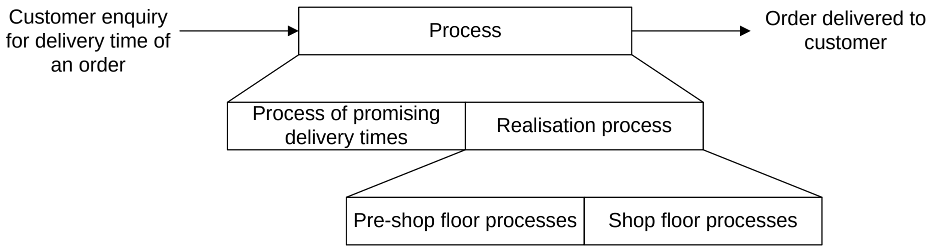

Figure 3.3. System description on various aggregation levels.

3.2.2.1. Process of Delivery time Promising (DP)

In the process of delivery time promising (DP), a customer's enquiry for a delivery time is translated into a promised delivery time. Although this process could be regarded as amounting to the realised throughput time, its main role regarding delivery reliability is setting the promised delivery time. We are interested whether the decisions that are made in this process result in realistic promised delivery times. Land and Gaalman (2009) show that poor delivery performance can be the result of not using relevant information in determining promised delivery times in many MTO companies. The review by Cheng (1989) distinguishes several types of information that can be used in establishing promised delivery times that can be categorised into shop-floor-status-related information and order-related information. The average lateness is typically influenced by the failure to properly use shop-floor status information, for example, current work-in-process (WIP). The variance in lateness is more specifically influenced by failing to adequately consider order-related information, such as the number of operations (NOP) required.

3.2.2.2. Realisation process (RP)

After the promised delivery time has been specified, the realisation process (RP) starts. Achieving the on-time delivery of orders requires the control of input and output moments of orders in different stages of the RP. Generally, this realisation process can further be subdivided into pre-shop-floor processes and shop-floor processes (see Figure 3.3).

Input to the pre-shop-floor stage is triggered by the order acceptance decision. The influence of this decision is highly interrelated with the DP. For example, the potentially negative influence on delivery reliability of accepting further orders beyond what can reasonably be completed (assuming constant capacity) can be remedied by promising longer delivery times for these extra orders.

The pre-shop-floor and shop-floor stages are separated by the release decision, which regulates when orders are allowed to be processed on the shop floor. One of the main objectives of the release decision is to balance loads across machine centres (capacity groups) and over time (Bergamaschi et al., 1997). It enables controlled progress of orders on the shop floor. Further, the relative urgencies of orders should be taken into account, to avoid the need to expedite orders on the shop floor.

Pre-shop-floor processes

Whereas the sub-processes distinguished above are applicable to most MTO companies, it is difficult to distinguish generic processes within both the pre-shop-floor and the shop-floor processes. The pre-shop-floor process in some companies can be seen as no more than the queue of orders that are waiting to be processed on the shop floor. Most MTO companies have to perform preliminary activities before operations on the shop floor can be commenced. The number of activities performed after receiving an order is often related to the level of customisation a company offers (see, for example, Amaro et al., 1999). For example, when customisation is high, it may be necessary to purchase customer-specific materials and carry out extensive process planning. These kinds of decisions are largely ignored in the PPC concepts (Land and Gaalman, 2009) where the main focus is on shop-floor operations.

Shop-floor processes

The shop-floor process normally consists of performing several operations. At this level, distinguishing relevant sub-processes will be company-specific. Often, certain operations can be grouped to define production units (Bertrand et al., 1990) that are responsible for certain sub-processes. After release to the shop floor, the pattern of an order's progress through its operations can be influenced by detailed scheduling or dispatching rules. Some rules focus on shorter average throughput times (for example, shortest-processing time-first: SPT), others in reducing the variance of lateness (for example, earliest-due-date-first: EDD).

Besides controlling the input moment of orders to the shop floor, output can also be controlled. Output control regulates the capacity which is made available so that orders can proceed to the next stage and finally be completed. To reduce the average lateness, additional capacity can be allocated to resources where congestion occurs. Further, peak requirements for specific orders can also be managed by adjusting capacity. Capacity changes, resulting from output control, can be usefully considered at the same time as making important input control decisions (Land and Gaalman, 1996). Table 3.1 summarises the influences of all the input and output control decisions on both the average lateness and the variance of lateness. Throughput diagrams and order progress diagrams can help to identify the influence of individual PPC decisions on delivery reliability performance, a topic extensively discussed in Soepenberg et al. (2008).

Table 3.1. How PPC decisions enable control of the average lateness and the variance of lateness.

PPC decisions	Controlling average lateness	Controlling variance of lateness
Delivery time promising	Promised delivery times based on shop-floor status (e.g., current WIP) of the company	Taking order-related information (e.g., NOP) into account when promising delivery times
Order acceptance	Controlling the number of orders accepted within a certain period	Controlling the number of orders with specific characteristics accepted
Release	Considering those orders that provide capacity groups with good load balance over time	Considering the relative urgency of orders at release
Priority dispatching	Using priority rules that focus on accelerating throughput (e.g., SPT/WINQ)	Using priority rules that focus on reducing variance of lateness (e.g., EDD/ODD)
Output control	Dedicating capacity to those resources where orders are congested	Dedicating capacity to meet the peak requirements of specific orders

3.3. Framework for diagnosing delivery reliability

This section discusses how we incorporated the theoretical building blocks discussed in the previous section in a supportive diagnosis framework that structures the diagnosis process using a diagnosis tree. We explicitly distinguish two layers in this diagnosis tree.

In the branches comprising the upper layer of the tree, diagnosis steps are taken sequentially in order to determine relevant problem areas that underlie delivery reliability performance. Four main problem areas are distinguished, resulting from the four possible combinations of the average or variance of lateness on the one hand and the control of the DP or the RP on the other. The breadth-first search strategy among the main problem areas enables one to gain a comprehensive overview of delivery reliability performance and avoids getting too involved in the detail in an early phase of the diagnosis process (Wagner, 1993).

In the lower layer of the diagnosis tree, more-detailed diagnosis steps are carried out in those problem areas that are identified as having a negative influence on delivery reliability. The diagnosis process ends when causes have been identified that are related to PPC decisions within each problem area. The remainder of this section elaborates on how the full diagnosis process can be carried out.

3.3.1. Diagnosis process: Part 1 - determining the relevant problem areas

Step 1: Analyse the distribution of lateness

The diagnosis starts by determining the percentage of orders delivered tardy in a certain period (see Figure 3.4). For this purpose, quantitative data on promised delivery times and realised throughput times have to be gathered for the set of orders delivered in that period. Based on the lateness values of all orders, a distribution of lateness can also be constructed. Studying this distribution of lateness, as shown in Figure 3.2, can help in deciding whether to focus on the average lateness branch and/or the variance of lateness branch shown in Figure 3.4. It should be noted that Figure 3.2 shows rather clear situations whereas, in most practical situations, neither the average lateness nor the variance of lateness can be ignored in the subsequent diagnosis.

Step 2: Analyse differences among order subsets

When analysing the average lateness, one should check to see if a limited number of order subsets are responsible for the high average. This may allow one to narrow down the later analysis to these subsets. In terms of the variance of lateness, two situations may emerge: (i) some order subsets have a high internal variance of lateness which would lead to continuing the analysis of variance within these subsets; or (ii) the internal variance within subsets is limited and the variance of lateness is mainly the result of variance between subsets. In the latter case, the analysis should proceed by focussing on those subsets where the average lateness is high, and therefore continue in the average-oriented part of Figure 3.4.

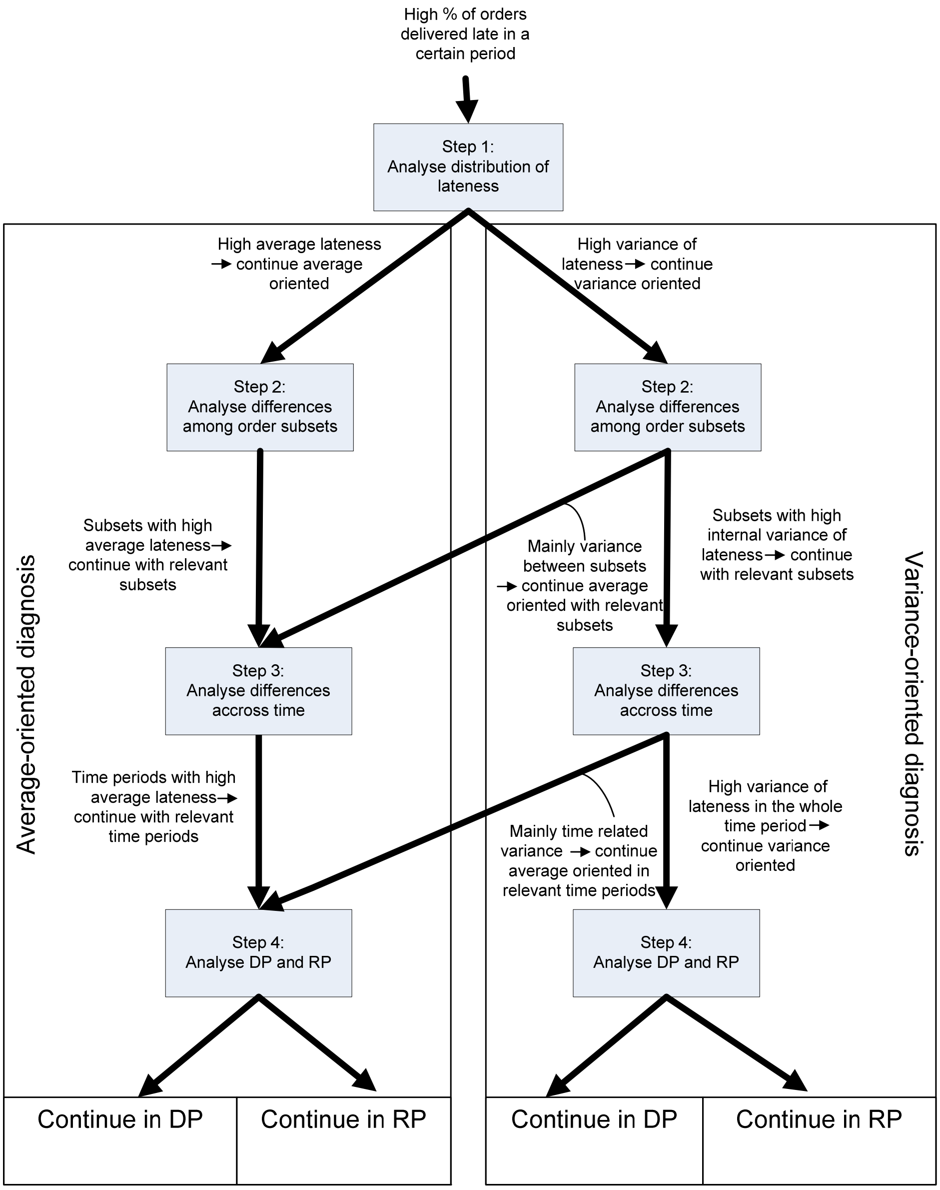

Figure 3.4. Upper part of diagnosis tree.

Step 3: Analyse differences over time

To gain insights into time dependency, the diagnosis continues by analysing how lateness changes over time. The throughput diagram and order progress diagram discussed in detail below, in the fourth diagnosis step, can be used in support. Analysing the average lateness could, for instance, reveal that specific time periods show relatively high lateness values compared to others, suggesting specific focus on these periods. In terms of variance, two situations may become apparent: (i) if variance of lateness is mainly time-dependent, the diagnosis process may switch from analysing the variance of lateness to the average lateness. Here, the high average lateness in certain time periods becomes the focus of the analysis in the later steps. Alternatively, (ii) if the variance of lateness is independent of time, the diagnosis continues along the branches related to the variance of lateness.

Step 4: Analyse DP and RP

Depending on the decisions made in the previous steps, the diagnosis will be either average-oriented or variance-oriented (see Figure 3.4). The fourth step is to determine, whichever orientation has been identified, whether the diagnosis should continue in the process of delivery time promising (DP) and/or in the realisation process (RP). This decision can be supported with throughput diagrams (Wiendahl, 1995) and order progress diagrams (Soepenberg et al., 2008). Figure 3.5 illustrates how these diagrams can be used to decide whether both or only one (and if so which) process warrants further investigation. Figure 3.5a and Figure 3.5b are illustrative of throughput diagrams that might lead one to favour the RP or the DP respectively in an average-oriented diagnosis. Figure 3.5c and Figure 3.5d are order progress diagrams that could lead one to opt for the RP and the DP respectively in a variance-oriented diagnosis. This is explained in more detail below.

Average-oriented diagnosis: DP or RP?

A throughput diagram is particularly useful for gaining insights, making the decision whether to focus on the DP and/or the RP regarding the average lateness. The horizontal axis in a throughput diagram represents time. The vertical axis shows, for each curve in the throughput diagram, the cumulative input and output of work regarding a process in the order flow. Work can be measured in units of orders or processing time units, with the latter measurement being the most commonly used. Curves can be drawn for each input/output moment in the flow of an order. At this stage of the diagnosis, we are taking an aggregated overview (see Figure 3.5a and Figure 3.5b) with just two or three curves.

The input curve (1) in both figures refers to order acceptances. The cumulative output curve (2) reflects order completions. These cumulative curves increase respectively in value by the capacity requirements of the order at the time of its acceptance or its completion. In addition, Figure 3.5b shows a promised output curve (3), which increases in value by the capacity requirements of the order each time an order is promised to be delivered to a customer. The vertical distance between the order acceptance curve and the order completion curve (A) depicts the work-in-process (WIP) in the company and the horizontal distance (B) relates to average throughput times of the RP. The horizontal distance between the order acceptance curve and the promised output curve (C) relates to average promised delivery times. The horizontal distance between the promised output curve and the realisation curve (D) reflects the average lateness.

An average lateness that would justify further investigation would be indicated in the throughput diagram by an order completion curve that lies close to, or even below, the promised output curve over a considerable period of time, as shown in Figure 3.5b. It is often hard to fully exclude either the RP or the DP, which respectively influence the order completion curve and the promised output curve, from further diagnosis. Nevertheless some guidelines can be used to decide which of these processes to focus on after constructing a throughput diagram.

An uncontrolled DP is often indicated by irregularities over time in the promised output curve. An example of this is shown in Figure 3.5b where there is a sudden shift at a certain time, resulting in a high (positive) average lateness from that time on. An uncontrolled RP can often be indicated by irregularities in the order completion curve (see, for example, Figure 3.5a), unless these fluctuations follow the promised output curve. The latter situation indicates that the RP is trying to satisfy the irregular demand rather than being uncontrolled. For an extensive discussion on these patterns, we refer to Lödding (2005).

It should be noted that although a stable level of order completions may suggest adequate control of the RP, average throughput times could still be perceived as too long. For the pre-shop-floor stage this could be the result of relatively long average material delivery times, whereas on the shop floor it could result from unnecessary waiting times caused by excessive buffering. For the latter we refer the reader to Nyhuis and Wiendahl (2009), who extensively discuss a diagnostic approach that can lead to reduced average throughput times.

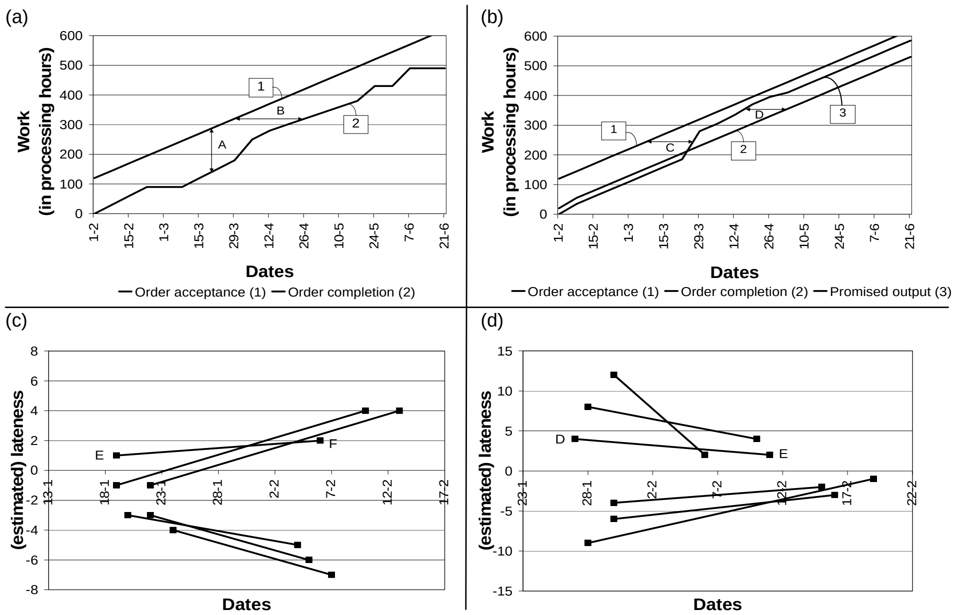

Figure 3.5. Key diagrams for diagnosis process.

Variance-oriented diagnosis: DP or RP?

Order progress diagrams are helpful in gaining a clear understanding of whether to focus on the DP and/or the RP regarding the variance of lateness. The horizontal axis in the order progress diagram measures time, and the vertical axis indicates (estimated) lateness. Each order is plotted as an individual curve in this diagram, which shows its progress from the start to the end of a process. Dots on the order curve indicate the start or completion of a process, and can be included for each of the underlying processes in the RP. Again we present an aggregated overview here.

Figure 3.5c and Figure 3.5d both show the progress of six orders from their acceptance through to their completion. The horizontal positions of the left and right dots in each curve represent the times of acceptance and completion of the order respectively. The vertical position of the right dot for each curve represents the lateness at order completion. The vertical position of the left dot for each curve represents the estimated lateness at order acceptance. An estimated positive or negative lateness at acceptance indicates that an order would be delivered late or early respectively, if throughput times in the realisation process were equal to average throughput times for this process. Orders with a positive and negative estimated lateness at acceptance (or at intermediate stages) would commonly be referred to as urgent and non-urgent orders respectively. The vertical spreads of the left and right dots respectively indicate the variance of estimated lateness at acceptance and actual lateness at order completion. Curves with an upward trend in the realisation process are orders that are delayed in that process, whereas curves with a downward trend indicate orders that are accelerated.

To decide whether to focus on the DP or the RP, a comparison of the variance of estimated lateness at order acceptance and the variance of lateness at order completion should be made. This is illustrated using Figure 3.5c and Figure 3.5d. Figure 3.5d shows a situation in which the variance of estimated lateness is high at the time of acceptance. In this situation the RP, although accelerating the most urgent orders, cannot compensate for the variability in promised delivery times. This conclusion would direct further investigation towards the DP. Figure 3.5c, on the other hand, shows a situation in which the variance of estimated lateness at the time of acceptance is relatively low. However, urgent orders are slowed down in the RP, while less urgent orders are accelerated. As a result, the variance in (estimated) lateness increases during the RP, resulting in the late delivery of some orders. This suggests focussing further investigation on the RP.

The simple examples in this subsection show how both throughput diagrams and order progress diagrams can be helpful in deciding which problem area to investigate further.

3.3.2. Diagnosis process: Part 2 - determining PPC causes within problem areas.

The diagnosis continues within the identified relevant problem areas. Here, this subsection is confined to a concise discussion on the diagnosis process within the two problem areas related to the RP since the diagnosis process in the DP problem areas is rather straightforward. That is, for the average-oriented DP diagnosis, the average promised delivery time is assumed to be too short in certain periods. Therefore, one should determine whether shop-floor status is taken into account when promising delivery times (see also delivery time promising in Table 3.1). With a variance-oriented diagnosis, one should investigate whether relevant order characteristics are considered when promising delivery times. Neglecting characteristics that lead to longer or shorter delivery times (such as longer or shorter routings) will typically increase lateness variability.

Figure 3.6 provides an overview of the diagnosis process for the RP regarding both average lateness and variance of lateness issues.

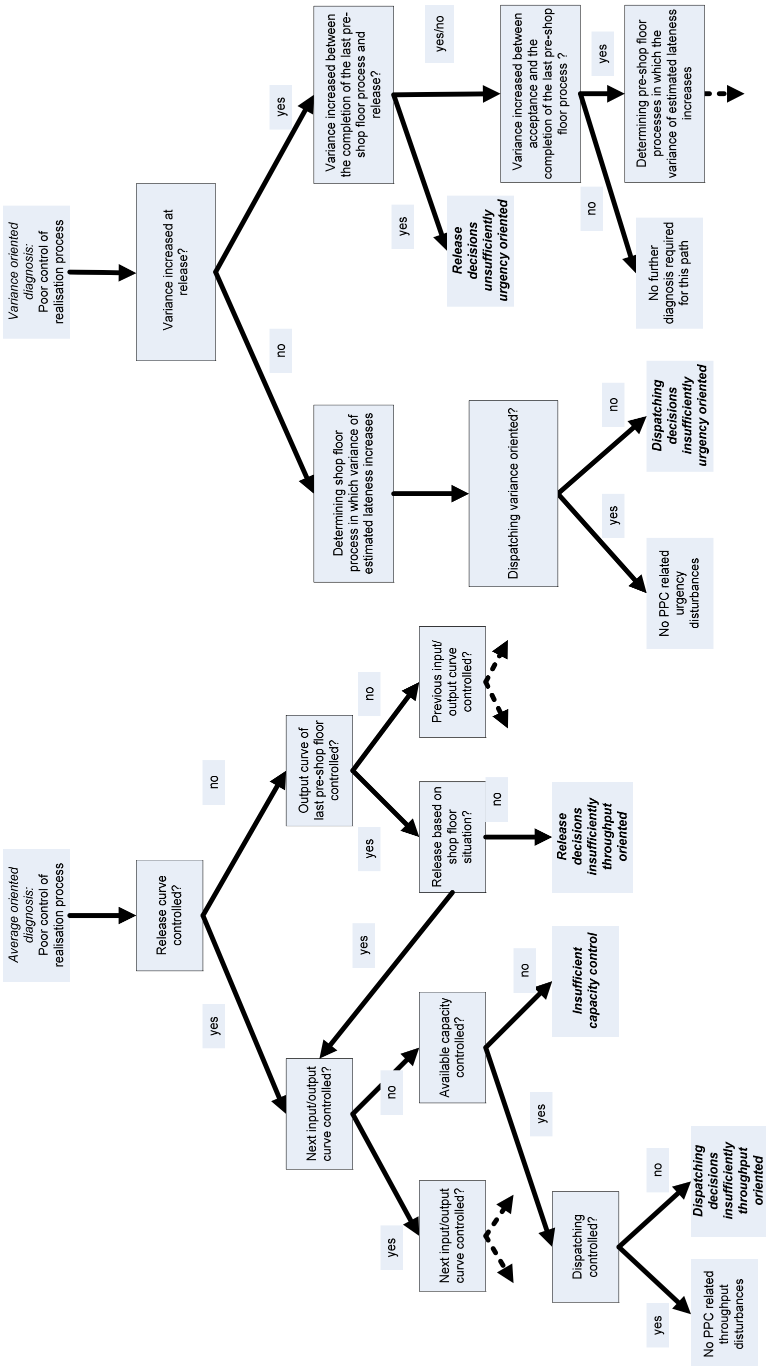

Figure 3.6. Average-oriented diagnosis (left) and variance-oriented diagnosis (right) within the RP.

Average-oriented diagnosis within the RP

Identifying this as a problem area is the result of several earlier conclusions drawn during the diagnosis process. First, the average lateness was considered excessive in the first diagnosis step, which could be linked to a specific order subset or time period as identified in steps 2 and 3. Furthermore, the fourth step would have revealed that the order completion curve in the throughput diagram had undesirable irregularities. This leads to continuing with an average-oriented diagnosis within the RP.

The subsequent diagnosis within this problem area focuses on detecting those PPC decisions that led to the undesired irregularities in the order completion curve. A first step is to diagnose whether the release curves of resources are controlled over time (see Figure 3.6). Naturally, only resources with the potential of being bottlenecks have to be considered here. This diagnosis step can be supported with throughput diagrams.

Controlled order release curve

If the order release curve is well controlled, the diagnosis continues by considering the shop-floor processes. The diagnosis steps involved try to detect the first input/output curve of a (potential bottleneck) shop-floor process that is uncontrolled after release, for example, a curve which shows irregularities over time.

Next, it should be determined whether the irregularities in the process completion curve are the result of PPC decisions regarding capacity. This could lead to the conclusion that changes in the assigned capacity of a resource explain the irregularities in completion times. As such, at this point, delivery performance could be traced back to a specific PPC decision, which is the aim of our diagnosis. Further diagnosis could focus on whether it is, for example, caused by decisions regarding total capacity over time and/or whether it is caused by differences in capacity dedications across operations over time.

If the available capacity is considered to be sufficiently controlled, an alternative step is to consider the influence of priority dispatching decisions on average lateness. It should be noted that the influence of these decisions is often limited where the released curve is controlled. Nevertheless, dispatching decisions that are not throughput-oriented, such as selecting orders with the longest processing time first, could still be identified as a cause of increasing average throughput times. If capacity and dispatching decisions cannot explain poorly controlled input/output curves, performance must be being affected by factors outside the scope of PPC. It could be the result of disturbances such as machine breakdowns or staff illness.

Uncontrolled order release curve

If an uncontrolled order release curve is detected, it could be the result of poor control of the pre-shop-floor processes or its causes could be related to the release decision itself.

First, irregularities in the order release curve could be the result of poor control of the pre-shop-floor processes, which then fail to provide an appropriate set of orders to choose from at release. If this is the case, the diagnosis should continue by analysing the input/output curves of the pre-shop-floor processes. When an input/output curve in the pre-shop-floor stage is regarded as uncontrolled, a further diagnosis can be carried out similar to that used in the situation where an uncontrolled input/output curve is detected on the shop floor.

Second, if the output of the pre-shop-floor processes is controlled, then irregularities in the order release curve should be attributed to the release decision itself. Two causes related to the main order release could underlie irregularities in the release curve: (i) the irregularities could result from release decisions that are not aimed at balancing loads; (ii) the irregularities could result from adjusting released amounts of work based on observed deviations from planned output. The latter situation suggests seeking further insights into the origin of these output disturbances on the shop floor.

Variance-oriented diagnosis within the RP

Diagnosis related to an excessive variance in lateness focuses on identifying those decisions that led to the increase in variance of (estimated) lateness between order acceptance and order completion. Again here, the diagnosis can be confined to a specific order subset or time period.

The diagnosis within this problem area focuses on analysing variance differences between acceptance and release and between release and completion. As such, the diagnosis should focus on the sub-processes in which the variance of (estimated) lateness increases significantly. This type of diagnosis can be supported by order progress diagrams.

Increased variance between order release and order completion

If the variance of (estimated) lateness increases between order release and order completion, the diagnosis should focus on the shop-floor processes. More specifically, it should be determined during which of the shop-floor sub-processes the estimated lateness increases. For each of these processes it can then be investigated whether the increased variance is the result of poor dispatching decisions or decisions beyond the scope of PPC.

Increased variance between order acceptance and order release

If an increased variance is found at order release, the diagnosis should aim to detect whether this increase can be blamed on the release decision. Such a conclusion could be drawn if the variance of estimated lateness increases between the moment that an order is available for release and the moment that it is actually released since this would suggest that release decisions do not properly consider order urgency. Further diagnosis could then focus on why non-urgent orders are accelerated at the expense of less urgent orders at the point of release: it could be the result of poorly made (and sometimes even non-existent) release decisions or of release decisions that explicitly aim to balance loads. This latter situation could indicate a situation in which the release of urgent orders is delayed to improve load balancing.

An increase in variance between order acceptance and order release could also be the result of poor priority-setting in pre-shop-floor processes. If this is thought to be the case, the diagnosis continues in a similar way to that proposed for the shop-floor processes. As such, it will be investigated whether the increased variance is caused by poor priority dispatching decisions or decisions beyond the scope of PPC.

3.4. Applying the framework in practice

The framework discussed in the previous section has been applied in three companies and found to support the diagnosis of delivery reliability performance. From the companies, quantitative data on both promised delivery times and order progress were gathered for a period of approximately three months in a standardised way. A scanning system was implemented in each company to measure order progress at each relevant step of the RP: from order acceptance to order completion. The collected data have been analysed by the researchers with the support of the proposed framework. The outcomes of these diagnoses were presented in a workshop with the managers of each company to verify the outcomes and to obtain further clarification on those factors that could not be measured quantitatively.

Each case description starts with a brief overview of the main order routing scheme in the company. Then the diagnosis process applied to each company is discussed. Given that a full discussion of each diagnosis process would be very lengthy, the discussions focus on the search process related to one major cause of delays detected at each company. The diagnosis steps leading to the identification of these causes are shown in Figure 3.7. The subsections that follow discuss the selected diagnosis steps.

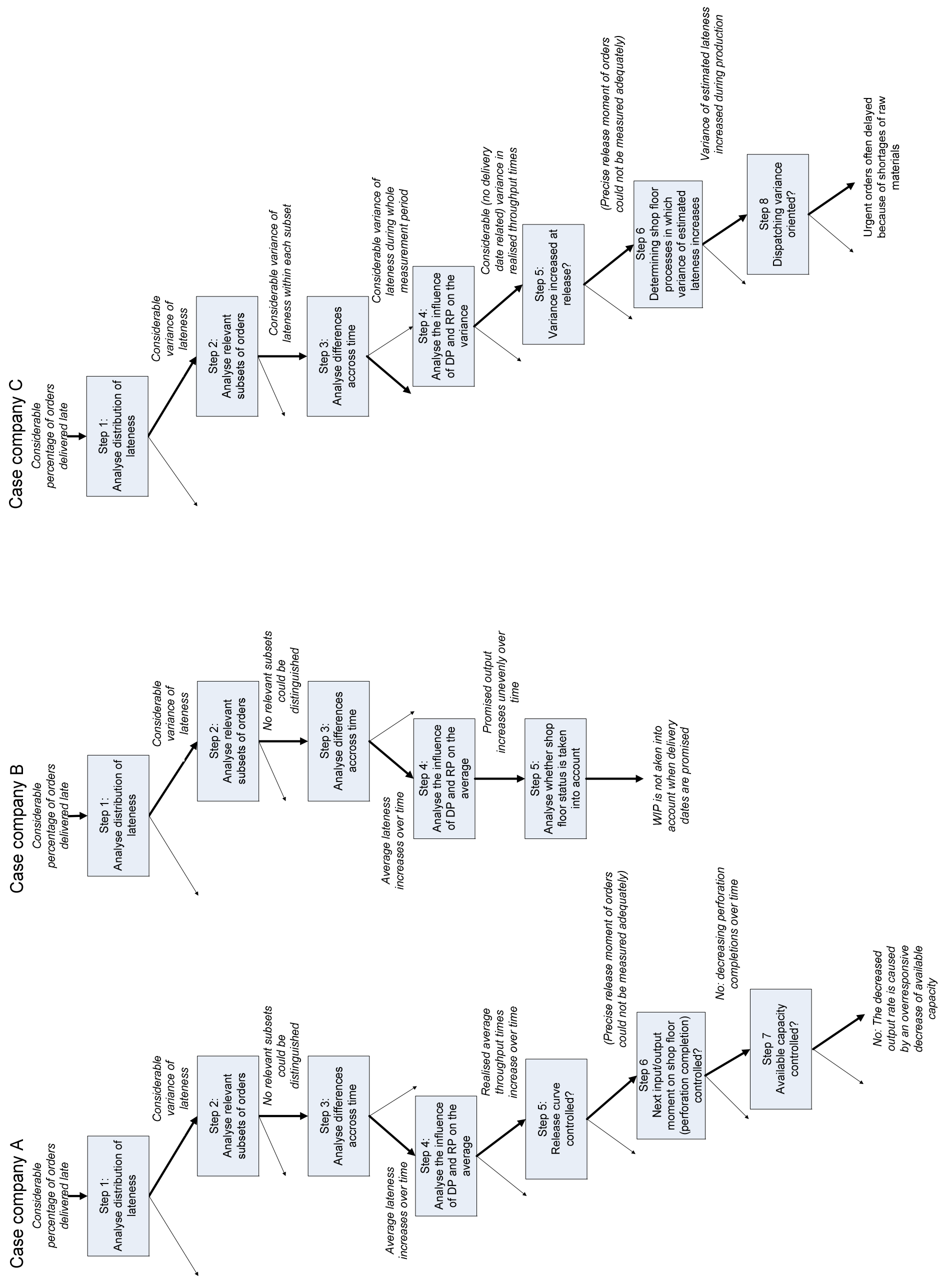

Figure 3.7. Overview of diagnosis steps and intermediate conclusions within each company.

3.4.1. Company A

Company A produces a large variety of perforated metal sheets, which are supplied as semi-finished products or end-products. Six main operations can be distinguished in the RP of an order. Figure 3.8 reflects the dominant flow. The need for the finishing operations depends on the product; it may include cutting and setting operations.

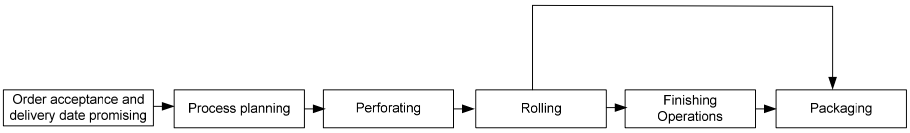

Figure 3.8. Main order routing in Company A.

The diagnosis exercise in this company revealed several causes of late delivery. An important cause was an over-responsive cut in capacity that led to an increase in average throughput times. The left diagnosis tree in Figure 3.7 shows the successive diagnosis steps undertaken and the conclusions drawn from using the framework for this company. After a brief reference to the other steps, we will focus on step 6 and provide the throughput diagrams that supported the diagnosis.

At step 1, a considerable variance in lateness is observed. The associated increase in average lateness cannot be related to a specific subset of orders (step 2). Further diagnosis of the variance of lateness reveals that this variance can mainly be explained by a change in average lateness over time (step 3). Therefore, the subsequent diagnosis focuses on the average lateness instead of the variance of lateness. In the next two diagnosis steps we attribute the increase in lateness to increasing throughput times (step 4), and more specifically to the decreased output from the perforation operation (step 6). It should be noted that the available data did not allow us to accurately determine the release moment in step 5.

Our findings in step 6 are a result of analysing the throughput diagram shown in Figure 3.9. The vertical axis shows work measured in processing time (hours) on the main machines in the perforating department. Two curves are drawn: the order acceptance curve (1) and the completion curve of the department (2). The horizontal distance between the two curves, indicated by the double-ended arrows, reflects the average throughput time that was realised in a certain period. Here, this horizontal distance is increasing over time as a result of the fact that the slopes of both the input and output curves decrease simultaneously at the beginning of April. This pattern indicates a quick response in terms of output rate to a reduced acceptance rate. This quick response meant that the WIP did not decrease. As Little (1961) proved, a steady amount of WIP combined with a decreased completion (output) rate must mean longer average throughput times.

Further diagnosis of the decrease in perforation completion rate reveals that available capacity was purposefully decreased by the production manager in response to the reduction in accepted orders (step 7). As such, the underlying cause can clearly be attributed to a PPC decision. Management did not realise that their responsive adjustment to capacity would lead to an increase in average throughput times and thus detrimentally affect delivery reliability.

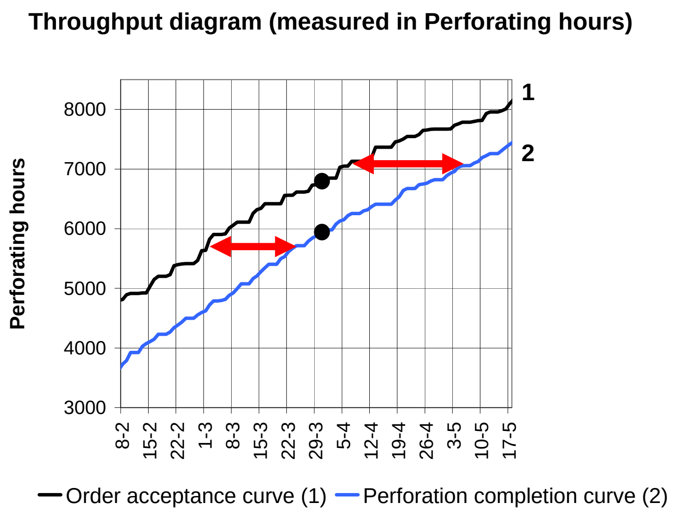

Figure 3.9. Throughput diagram of perforating department in Company A.

3.4.2. Company B

Company B produces blades and special components for gas and steam turbines (see Figure 3.10).

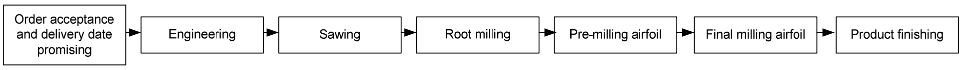

Figure 3.10. Main order routing in company B.

The diagnosis undertaken for this company revealed that one of the main causes of late deliveries was the lack of adequate support for sales; resulting in them promising unrealistic delivery times (see Figure 3.7). The starting point in this discussion is the conclusion, at step 3, that average lateness was increasing over time. Figure 3.11 shows this in a throughput diagram of root-milling hours. Three curves are drawn: the order acceptance curve (1), the order completion curve (2) and the promised output curve (3). The throughput diagram shows that the horizontal distance between the promised output curve (3) and order completion curve (2) grows, indicating an increasing average lateness.

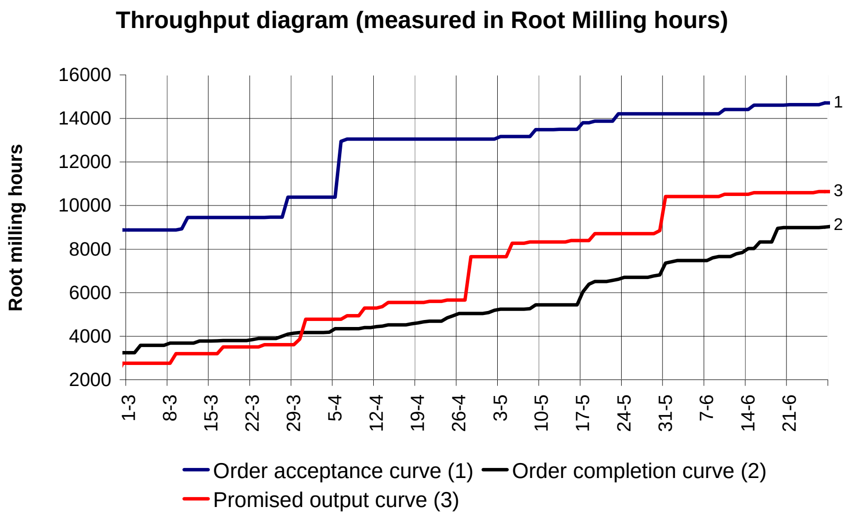

Figure 3.11. Throughput diagram of root-milling in Company B.

In step 4 it is determined whether the focus should be on the DP and/or the RP. Here, we observe a difference in the inclinations of the order completion curve (2) and the promised output curve (3). Whereas the inclination of the order completion curve (2) is fairly stable over time, the promised output curve suddenly increases at the end of March, resulting in an increase in average lateness from that time on. This indicates that more hours of work were being promised than could reasonably be achieved in this period. This makes it likely that the cause of the increased average lateness can be found in the DP. However, it should be noted that the RP should not be totally excluded from the diagnosis since the final stage of the graph shows that order completions are also increasing to some extent. This suggests that the RP may have had unexploited flexibility to react to the increase in promised output in March.

Diagnosis step 5 focuses on identifying whether the shop-floor status, in terms of WIP, is taken into account when delivery dates are promised. To check this explicitly, an additional diagram is created that visualises the relationship between WIP and promised delivery times (see Figure 3.12). WIP is defined in this figure as the root-milling load that has been accepted but not yet processed. On the same graph, promised delivery times are plotted against the order acceptance date of an order. Here, both WIP and promised delivery times are measured in working days. Figure 3.12 shows that no clear relationship exists between WIP level and promised delivery times. Between May 10 and May 24, several orders had promised delivery dates of no more than ten working days, despite the existing WIP already being equivalent to around 40 days of work. A discussion with the sales department and company management confirmed our conclusion that WIP is not adequately taken into account when promising delivery dates.

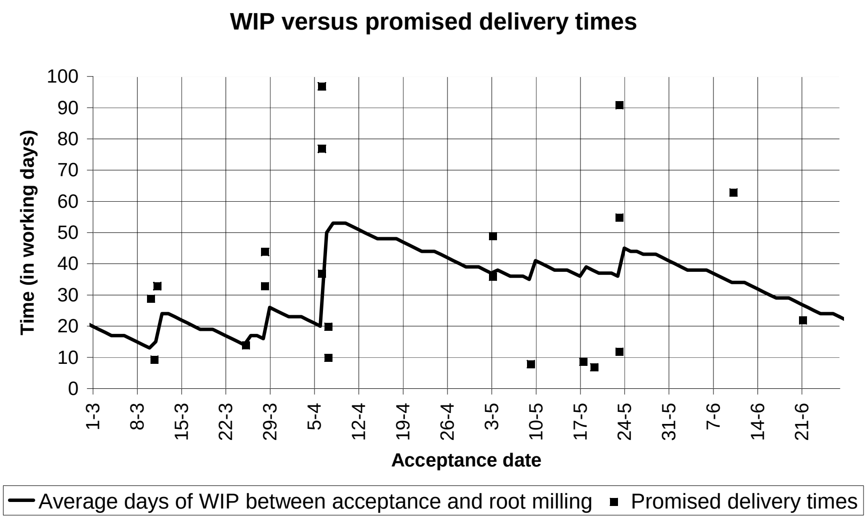

Figure 3.12. WIP vs. Promised delivery times in Company B.

3.4.3. Company C

Company C develops and produces mechanical and electronic devices for measuring temperature, air pressure and humidity. The main routings are visualised in Figure 3.13. The company's development process is mainly used to develop special enclosures.

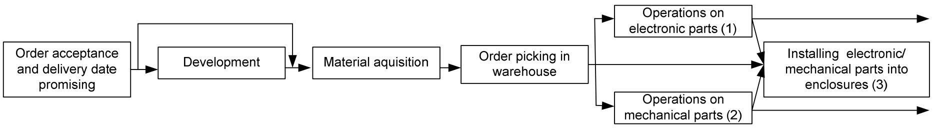

Figure 3.13. Main order routing in company C.

Several causes of late delivery were detected in the diagnosis process for this company. An important reason was that a shortage of raw materials led to delays with urgent orders in the production process (see Figure 3.7). We elaborate here on diagnosis steps 2-6, which were preceded by detecting a considerable variance in lateness in step 1. In diagnosis step 2, a subdivision of the order set was made into order types based on the required resources. Three subsets of orders could be distinguished in this company, denoted by numbers 1 to 3 in Figure 3.13. Order types 1 and 2 respectively require operations on electronic parts and mechanical parts. Orders of type 3 involve enclosures, normally completed with parts from the other resources. Figure 3.14 illustrates the lateness distributions within each order group. The figure shows that substantial variances of lateness exist within each order group, suggesting a need for further diagnosis within each order group. For reasons of brevity, we will confine the discussion to the diagnosis related to the electronic parts (1), which made up 59% of the orders in the measurement period.

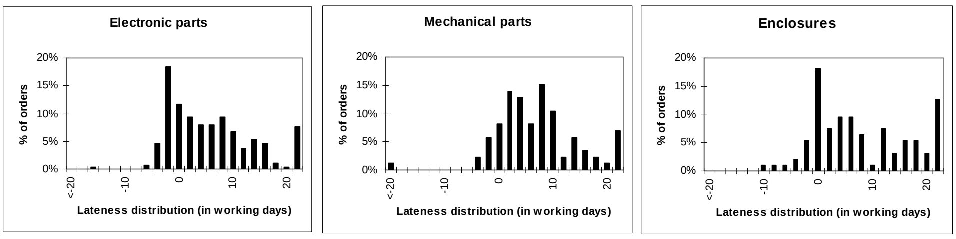

Figure 3.14. Lateness of orders for three order types in Company C.

As the lateness variance could not be explained by changes over time in the third step, diagnosis step 4 is used to gain insights into the influence of both the DP and the RP on the variance of lateness. This step can be supported by using order progress diagrams. In a simplified approach, the relationship between promised delivery times and realised throughput times is shown in Figure 3.15. The diagonal reflects the ideal situation in which realised throughput times equal promised delivery times. Orders positioned above this diagonal are delivered late. The figure shows that the high variance in lateness is the result of high variances in both promised delivery times (horizontal scatter) as well as in realised throughput times (vertical scatter). Both aspects therefore require further diagnosis. Again for brevity, we will here only discuss the diagnosis related to the search for causes related to the RP.

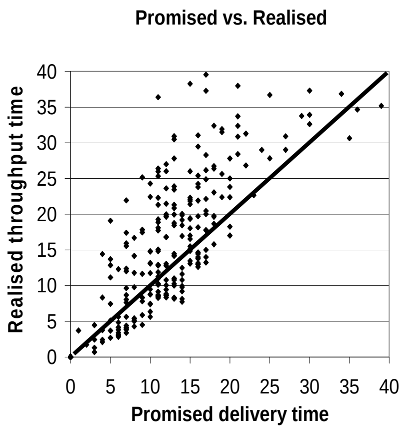

Figure 3.15. Promised delivery times vs. realised throughput times in Company C.

As with Company A, the release moment (step 5) could not be accurately recorded in this company, and so we focused on determining whether the variance of lateness increased on the shop floor (step 6). This step was executed with the aid of order progress diagrams. In Figure 3.16, the right-end dots of each curve represent the lateness at the time the production of the order was completed. The first (left-end) dot of each curve represents the estimated lateness of the order when production commenced. Curves with a positive estimated lateness are referred to as urgent production orders. Upward and downward directions of a curve respectively represent a relatively long or short waiting time in the production process.

Some refinements, as in Soepenberg et al. (2008), had to be made to the conventional calculations in the order progress diagram. To correct for the significant influence of processing times on throughput times in the production process, estimated lateness is based on constant waiting times rather than on constant throughput times. The median waiting time is taken as a reference to distinguish between delays and accelerations since the more-conventional mean would be very sensitive to the extreme outliers that existed. Figure 3.16 shows only delayed urgent orders and accelerated non-urgent orders, and it is notable that these made up half of the orders in the measurement period. A large number of delayed urgent orders combined with several non-urgent orders that were accelerated was responsible for the found increase in the variance of lateness in this process.

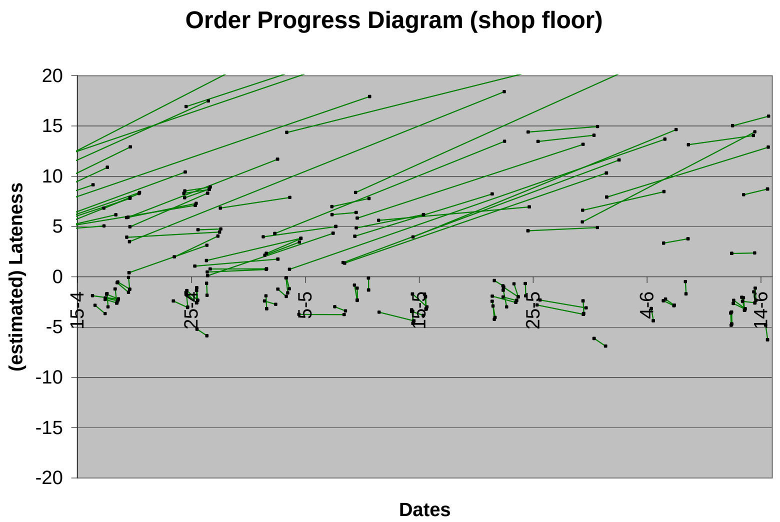

Figure 3.16. Order progress diagram of production process for electronic parts in Company C.

A further (qualitative) diagnosis revealed that an important factor in the delay of many urgent orders was material shortages, only detected after production commenced. Therefore, after identifying this process as a major issue, further diagnosis steps focused on the control of the material acquisition process.

3.5. Conclusions

Production planning and control (PPC) has a considerable influence in achieving a high delivery reliability, an order winning performance criterion for many make-to-order (MTO) companies. This research has its origins in the perceived lack of literature support for diagnosing the influence of PPC decisions on delivery performance, a process which is a prerequisite for detecting improvement opportunities. For this purpose, a framework has been developed which aims to assist in identifying those PPC decisions that have negatively influenced a company's performance.

The framework integrates insights from several literature domains. Domain-specific literature on PPC is used to provide a comprehensive diagnosis structure. The applied breadth-first search strategy, taken from general problem-solving literature, avoids getting bogged down in detail too early in the diagnosis. The final result is a diagnosis tree, which structures the diagnosis process from delivery performance to specific PPC decisions. In addition, recently developed research tools such as throughput diagrams and order progress diagrams are incorporated to support the navigation through this diagnosis tree.

Three case studies have shown that applying the framework in practice can lead to the identification of PPC issues that undermine delivery performance. Its application resulted in the detection of several PPC-related causes that reduced delivery reliability performance. In line with Hopp et al. (2007), we believe that incorporating scientific knowledge in such an intuitive and easy-to-use framework can foster the transfer of academic work in the operations management field to practitioners. Future research could focus on refining the diagnosis framework based on new scientific insights and practitioners' needs.