Improving scheduling accuracy by reducing data inconsistencies in production control

G. Schuh, T. Potente, C. Thomas, F. Brambring (f.brambring@wzl.rwth-aachen.de)
Laboratory for Machine Tools and Production Engineering (WZL)
RWTH Aachen University, Steinbachstr. 19, 52074 Aachen, Germany

Abstract
The quality of production feedback data gathered on the shop-floor is regularly reduced by various data inconsistencies and errors which impair detailed scheduling accuracy. In this paper, an approach for reducing these inconsistencies in production feedback data by deriving error-specific integrity rules is presented and validated through a simulation study.

Keywords: production planning and control, data inconsistencies

Introduction
Companies of the manufacturing industry in high-wage countries, especially machinery and plant engineering companies, are facing a multitude of interacting challenges: driven by a rising competition from emerging countries and by increasing customers’ demands concerning individualized product features, manufacturing companies strive for providing highly customized products at short delivery times (Brecher 2012).

A study conducted by the Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University in 2013 shows, that manufacturing companies claiming to pursue a high adherence to promised delivery dates as their main logistic target, can consistently outperform their competitors’ with respect to sales and profit (Schuh and Stich 2013). In the context of the described accelerating market dynamic, pursuing a high adherence to promised delivery dates has proved to be a recipe for high customer satisfaction while simultaneously realizing lower costs due to lower stocks and saved capacities (Maravelias and Sung 2009).

In order to achieve a high adherence to promised delivery dates, excellent Production Planning and Control (PPC) processes are required. However, PPC has become an evermore complex task in recent years: the customers’ risen demands for individualized products lead to a high product variety associated with increasingly complex process chains and material flows organized in a job-shop environment (Stevenson et al. 2005). Another complexity driver for PPC processes are the constantly changing circumstances in a manufacturing environment such as unplanned machine downtime, employees calling in sick or last minute wishes by the customers’ concerning product features and delivery dates (Stevenson et al. 2005). Hence, manufacturing companies deploy a variety of specialized IT-systems like planning modules of ERP-systems and Advanced Planning and Scheduling (APS) systems (Schuh et al. 2007). In order to provide essential information for updating the near- and middle-term scheduling, a high number of high-
resolution data concerning manufacturing and assembly processes is gathered on the shop-floor by so called Production Data Acquisition (PDA) systems (Perzyk et al. 2008). The gathered production feedback data are comprised of information concerning each process step of all active production orders such as the start and end dates, utilized resources with needed set up and processing durations and the produced quantities (Buzacott et al. 2013).

The quality of these production feedback data is typically impaired by a wide range of different data inconsistencies and errors (Kwak and Kim 2012, Schuh et al. 2013). Utilizing these compromised data for updating near and middle term scheduling of production orders has a negative impact on the results of planning processes and ultimately leads to a low adherence to promised delivery dates due to longer lead times. Consequently, production feedback data with high integrity serve as an important prerequisite for high quality scheduling results.

In this paper, a methodology for estimating the negative impact which data inconsistencies have on scheduling results is presented. For this end, integrity rules for the most common data inconsistencies found in production feedback data are formulated and validated through a simulation study.

The rest of the paper is organized as follows. First, an overview of the state of the art concerning processes and IT-systems as well as data inconsistencies in production control and other disciplines is given in section 2. Section 3 describes the process of generating and utilizing production feedback data in detail and presents typical data inconsistencies and errors. Integrity Conditions dealing with the identified inconsistencies and errors are presented in Section 4 and validated through a simulation study. The paper concludes with an outlook on future research.

State of the art

The processes and IT-systems which companies use in order to gather information on the shop floor and employ these data have been described and discussed extensively by researchers. Especially scheduling methodologies and algorithms have been of high interest, ranging from single and multi criteria scheduling simulations to artificial intelligence (AI) approaches (Chan and Chan 2004, Höppe et al. 2013).

However, approaches for dealing with data inconsistencies which are specifically designed for production feedback data are scarce in literature. In their broad literature review concerning the architecture of manufacturing scheduling systems, Framinan and Ruiz (2010) name only 1 of 62 considered studies which focuses on checking input data while most studies are dedicated to scheduling algorithms. McKay and Wiers (2013) propose that data consistency can be improved by a decision support system which empowers one integrated planner in being responsible for planning, scheduling and dispatching the production orders. Mayer and Poege (2013) acknowledge the fact that production feedback data is stored inhomogeneous and incomplete in databases which requires a methodology for checking the plausibility of these data but do not propose an own approach in dealing with this challenge.

Although, problems caused by data inconsistencies have not been extensively discussed in regard to manufacturing scheduling, a broad variety of literature concerning data consistency can be found in other fields. Especially with the growing importance of the world wide web in the 1990s, computer scientists have discussed ways for ensuring data consistency in shared data bases. Sheth and Rusinkiewicz (1990) define consistency requirements for interdependent data in multiple data repositories. Pitoura and Bhargava (1995) introduce a two-level consistency model for dealing with inconsistencies caused by disconnections in mobile distributed environments.
Data inconsistencies in emerging data grids are addressed by Domenici et al. (2004) and different consistency models are presented.

It can be concluded, that processes and IT-systems related to PPC tasks and more specifically with manufacturing scheduling, have been and still are an extensively discussed topic within the production management research community. Even though, the impediment which inconsistent production feedback data poses to realistic scheduling results has been identified, there has been little research towards this topic.

This paper presents integrity rules for major data inconsistencies found in production feedback data. These integrity rules are used to impute values in inconsistent data sets. The effect this approach has on scheduling results is estimated through a simulation study.

**Data inconsistencies in production control**

*Generating production feedback data*

The basic process of generating and utilizing data is depicted in Figure 1. ERP-systems usually include modules for planning production orders under capacity constraints. These rough planning results are the basis for the detailed scheduling done by the APS-systems which is updated every night with the newly gathered information on the shop floor concerning the progress of the currently active production orders (Höppe et al. 2013).

![Figure 1 – Process of generating and utilizing production feedback data](image)

Modern CNC work stations often possess an interface for directly uploading the data to the companies’ databases via Ethernet. However, the before mentioned study carried out by the WZL of RWTH Aachen University showed that only 41 % of the participating large companies and only 23 % of all participating small and mid sized companies already employ automatic production data acquisition (Schuh and Stich 2013). With 67 % the majority of all participating large companies mentioned manual production data acquisition via specially designed terminals as their leading system for acquiring production feedback data while for small and mid sized companies the most utilized method of production data acquisition is still hand written
documentation with 57 % (Schuh and Stich 2013). While these diverse technologies can lead to media breaks and are a major reason for data inconsistencies in production feedback data, it cannot be assumed that companies will manage a complete switch to more sophisticated systems in the near future. Through the introduction of new sensors and technologies the inhomogeneity within the production feedback datasets might even become larger and introduce new types of inconsistencies into the data. Where data inconsistencies and errors cannot be ruled out, a functionality for checking and assuring data consistency is of paramount importance given the high impact poor production feedback data quality has on scheduling results. An overview of today’s most frequent inconsistencies and errors found in production feedback data is given in the following sub section.

Data inconsistencies and errors in production feedback data

Regarding faulty information gathered on the shop floor, we distinguish between inconsistencies and errors. The German Institute of Standardization (DIN) defines errors as the “non-fulfillment of a requirement” (DIN 2005). In distinction to errors, inconsistencies occur whenever two bits of information which are believed to be correct, contradict each other and therefore cannot be true at the same time (Nguyen 2008).

Table 1 – Data inconsistencies and errors in production feedback data

<table>
<thead>
<tr>
<th>Inconsistencies</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1 work machines</strong></td>
<td></td>
</tr>
<tr>
<td>• Mismatch of planned work stations</td>
<td>• No feedback concerning employed work</td>
</tr>
<tr>
<td>• Process steps reported repeatedly</td>
<td>stations</td>
</tr>
<tr>
<td>a) for identical work station</td>
<td></td>
</tr>
<tr>
<td>b) for different work stations</td>
<td></td>
</tr>
<tr>
<td>• Process steps reported repeatedly for identical work station with inconsistent</td>
<td></td>
</tr>
<tr>
<td>time stamps</td>
<td></td>
</tr>
<tr>
<td><strong>Group 2 time stamps</strong></td>
<td></td>
</tr>
<tr>
<td>• 5 cases of inconsistent time stamps</td>
<td>• Missing start time</td>
</tr>
<tr>
<td></td>
<td>• Missing end time</td>
</tr>
<tr>
<td></td>
<td>• Missing start and end time</td>
</tr>
<tr>
<td><strong>Group 3 master data</strong></td>
<td></td>
</tr>
<tr>
<td>• Increasing production order quantity</td>
<td>• Incorrect set up time</td>
</tr>
<tr>
<td></td>
<td>• Incorrect processing time</td>
</tr>
</tbody>
</table>

In production feedback data, errors most commonly occur when information is missing completely while inconsistencies are usually, but not exclusively, caused by incorrect time stamps which would indicate overlapping process steps for example. In this paper, we concentrate on 16 data inconsistencies and errors which typically occur in production feedback data gathered on the shop floor. These 16 data inconsistencies and errors are clustered in three groups according to their common subject (e.g. work station, time stamps and master data) and similarities of the respective integrity conditions. Table 1 gives an overview of these inconsistencies and errors with their classification.

Group 1 is comprised of inconsistencies and errors concerning the employed work stations. Group 2 contains all theoretically possible cases of inconsistent time stamps and missing information regarding start and end times. Finally, Group 3 covers inconsistencies and errors observed in production feedback data related to information retrieved from the respective master data.
Improving scheduling accuracy by reducing data inconsistencies

Deriving integrity conditions for production feedback data
Integrity conditions have been derived for all data inconsistencies and errors presented in Table 1. They reach from heuristical approaches such as cold deck imputation to rather elaborated algorithms like an adaptation of association rule induction for production control (Schuh et. al. 2013).

The definition of a valid integrity condition for every observed data inconsistency and error is a three step approach: the starting point is always an analysis of the missing or inconsistent data and what other information is available to fill these gaps. Within the defined groups presented in Table 1, the integrity conditions can be rather similar since they address data inconsistencies and errors with a common subject. As a second step, the available data needs to be linked in order to make an estimation about the missing or inconsistent information. Afterwards this logic has to be transferred into an algorithm in order to automate the process.

Validating the approach
In order to validate the practicability of the proposed integrity conditions, a simulation study was conducted, using an original production feedback data set by a German manufacturer of clutches and brakes for special applications such as seafaring. This company is a typical representative of the German machinery and plant engineering industry with highly customized products and job shop production. The data set covered a time span of 16 months with roughly 16,000 production orders and 78,000 process steps on 165 work machines.

Table 2 – Data inconsistencies and errors in sample data set

<table>
<thead>
<tr>
<th>Data Inconsistency or Error</th>
<th>Rate of occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect processing time</td>
<td>8.8 %</td>
</tr>
<tr>
<td>Incorrect set-up time</td>
<td>5.3 %</td>
</tr>
<tr>
<td>No feedback concerning employed work stations</td>
<td>3.1 %</td>
</tr>
<tr>
<td>Inconsistent time-stamp (case #3)</td>
<td>2.3 %</td>
</tr>
<tr>
<td>Missing start and end time</td>
<td>1.3 %</td>
</tr>
<tr>
<td>Mismatch of planned work stations</td>
<td>0.5 %</td>
</tr>
<tr>
<td>Sum of the rest</td>
<td>0.4 %</td>
</tr>
</tbody>
</table>

Naturally, this data set was not flawless. Hence, data quality was impaired by various data inconsistencies and errors which are presented with their respective rate of occurrence in Table 2. For example, 8.8 % of all process steps within the original data set were registered with incorrect information concerning the processing time. In total, almost 22 % of data sets are in some way impaired by data inconsistencies and errors.

By imputing missing or inconsistent information into the data through the implementation of integrity conditions, we aim to obtain a more realistic representation of current production order statuses which ultimately lead to better results of detailed scheduling. However, since the prerequisite for this research is that no complete data set exists, the actual processes which took place are unknown. Therefore, the validity of the integrity conditions has to be proven in an indirect way.

The proposed methodology for validating the applicability of the derived integrity conditions is depicted in Figure 2. First, the raw data are analyzed for the rate of occurrence of the defined data inconsistencies and errors. The data are then cleaned simply by deleting all
faulty data sets which leads to the so called basis data. This basis data will later be interpreted as the representation of reality in the simulation study. In a third step, new raw data is created by inducing inconsistencies and errors into the basis data according to their previously analyzed rates of occurrence. This new raw data is then once again cleaned by deleting all faulty data sets to obtain comparison data set 1. On another copy of this new raw data, the integrity conditions are applied, in order to obtain the so called comparison data set 2. This leaves us with three databases with which the simulation can actually work, since faulty data sets would lead to a termination of the simulation run.

For the simulation study, a simulation model of the above mentioned company’s shop floor was built up and fed with restrictions such as shift schedules and the planned production program directly from the three compiled databases. With the premise that identical data sets will lead to identical results, the simulation results were compared in respect to five important logistic targets: capacity utilization, adherence to promised delivery dates, through put, work in progress (WIP) and lead time (Schuh et. al. 2012). The results are depicted in Figure 3.

The simulation study results for all three data sets are computed relatively to the basis data which therefore reaches 100 % in every logistic target. The comparison data set 1 in which we simply ignored all data inconsistencies and errors by merely deleting them deviates by over 20 % in three of the five logistic targets. However, the application of the integrity conditions on comparison data set 2 shows a positive effect for all targets with a maximum deviation of 12 %. These results are a clear indication towards the practicability of the derived integrity conditions.

Figure 2 – Preparation of data for simulation study
In this paper, the idea of deriving integrity conditions for typical data inconsistencies and errors in production control has been presented. The approach was validated with a simulation study based on an original production feedback data set by a German mid sized mechanical engineering company. The superior goal is to achieve a higher scheduling accuracy in production control.

Future research needs to be conducted in several directions. In order to obtain a complete picture of the most severe data inconsistencies in production control, additional production feedback data sets need to be analyzed. For each of these data inconsistencies exist a high number of theoretically valid integrity conditions, so that further research needs to be carried out in order to determine the best performing ones. Additionally, the results obtained by applying specific integrity conditions might influence the results of others which makes the interdependency of the integrity conditions another research topic for the near future.

Acknowledgments

This new approach of improving scheduling accuracy by reducing data inconsistencies in production control is being investigated by the Laboratory of Machine Tools and Production Engineering (WZL) within the publicly funded research and development project: "Cluster of Excellence - Integrative Production Technology for High Wage Countries" (German Research Foundation, DFG).

References

Chan, F., Chan, H. 2004. A comprehensive survey and future trend of simulation study on FMS scheduling. Journal
of Intelligent Manufacturing. 15:87-102.


