DERIVING A MATHEMATICAL MODEL OF A PAINT SHOP FROM DATA ANALYSIS

Thomas Husslein, Christian Danner, Markus Seidl, Jörg Breidbach
Department Algorithm Design and Optimization
OptWare GmbH
Prüfeninger Str. 20, D-93049 Regensburg, Germany
E-Mail: thomas.husslein@optware.de

Wolfgang Lauf
University of Applied Sciences OTH Regensburg
Universitätsstr. 31, D-93053 Regensburg, Germany
E-Mail: wolfgang.lauf@oth-regensburg.de

KEYWORDS

Paint shop, stochastic modelling, data analysis

ABSTRACT

The authors deal with the optimization of production planning in the mixed-model assembly production. Modern paint shops are highly complex facilities with a multitude of interdependent process steps. In order to describe the occurring throughput and processing times, the behavior of these times is of great interest for production planning, a modeling of the paint shop is necessary. The disruptions appearing here, which are a major source for delays and reorganizations within the car plant, are due to the complex structure and various rework. In this work, an alternate formulation of a long term reliable model is discussed: The description of the paint shop based on a stochastic model which is directly derived from data analysis of the production data.

COMPLEX STRUCTURE OF A PAINT SHOP

Due to the many successive steps the production process of painting a body can be viewed as a multi stepped flow production (Spieckermann, 2002): Figure 1 gives a schematic overview over the major technological steps of a paint shop. In general this is reduced to the following steps:

Pre-treatment: Arriving from the body shop, the body gets degreased and all metal fragments that are remnants of the production process in the body shop get cleaned away. Also some corrosion protective substances are applied.

Base coat: At the base coating, the body is immersed in a bath of electrostatic particles that coat the body due to electrostatic forces. This is predominantly for protection from corrosion and for optimizing the color application. At the end of this step the body is heated in an oven to permanently fix the coating particles on the body.

Underbody seam sealing: In this process step overlapping of the metal sheeting and gaps are sealed to prevent water intrusion. Additionally the underbody gets an extra coating against stone impact.

Filler coating: After the surface of the body has been cleaned of dust again, the filler coating is applied. The brightness of the filler coating is adjusted concerning the brightness of the final color. This is an additional protection against stone chip.

Top coating: The top coat is the layer of color that is determining the final color of the product.

Clear finish: After the top-coating the whole body is painted with a clear paint. Clear finish is used in order to protect the color against scratches and other environmental influences.

Final inspection and rework: In these steps the color is investigated for quality problems. In case of quality failures the painted body is transferred to the reworking, otherwise the painted body is clear and subsequently transferred to a storage unit. The reworking unit consists of several steps. In the inspection the severity of the problem and the specialized workstation for the repair is assessed. There are several work stations depending on the severity of the problem. The three major stations are paint removal, spot repair and extended completion. After a re-assessment of the quality of the repair work the bodies are re-introduces into the line at the appropriate position.

The detailed description of each of these production steps is very difficult. Parallel process steps and the building of uniquely-colored batches have to be described. Furthermore, the paint shop of a car manufacturer is a place of continuing change. So the

Proceedings 28th European Conference on Modelling and Simulation ©ECMS Flaminio Squazzoni, Fabio Baronio, Claudia Archetti, Marco Castellani (Editors)
ISBN: 978-0-9564944-8-1 / ISBN: 978-0-9564944-9-8 (CD)

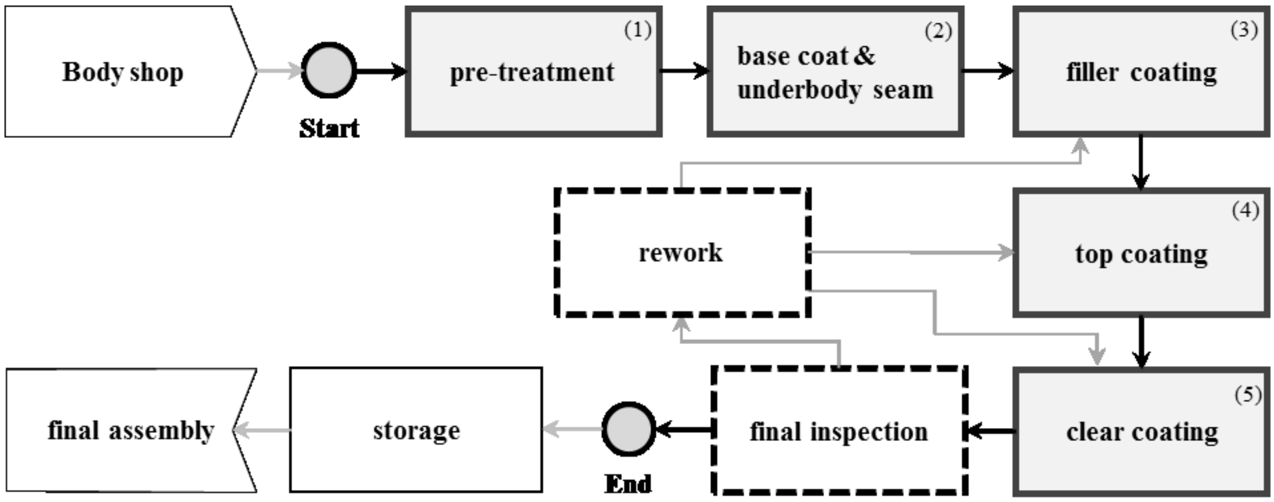

Figure 1: Basic Layout of a paint shop

formulation of long-term reliable model is nearly an impossible task.

PROBABILISTIC NATURE OF PAINT SHOP PROCESSES

Modern car production is organized by the pearl chain principle (Meyr, 2004; Weyer, 2002): The individual orders are lined up like pearls on a chain fixed in their sequence position relative to each other.

Complex processes in the final assembly are based on the production order planned. For the delivery processes just-in-time (JIT) and just-in-sequence (JIS) the stability of the production order is of central importance. As a consequence the description of the behavior of a paint shop is essential for optimal production planning and for using production control systems.

Parallelization of lines, processing of bodies in batches and rework due to system failures or violation of quality parameters cause turbulences of the pre-planned order. Especially, painted bodies which have to be reworked are separated from the assembly line and are reworked on specialized work stations.

A simulation model describing this process has to take into account this stochastic throughput.

From an abstract point of view, the paint shop can be regarded as a so called “black-box”. Bodies in white are entering the paint shop and colored bodies are leaving it. Each body has an individual throughput time depending on its path through the paint shop. The individual throughput times of the bodies can be easily obtained from data analysis.

DISTRIBUTION OF TROUGHPUT TIMES

The throughput time of a body *i* is defined as the time it takes to successfully pass through a workstation or a group of workstations (Arnold and Furmans, 2009). In this case we are interested in the throughput time through the paint shop which is defined as the timestamp a body leaves the paint shop minus the timestamp the body has entered the paint shop. The manufacturing information system of a car manufacturer stores for each produced car many production timestamps. By using the timestamps from the input and the output of the paint shop a top-down analysis is applied.

Through extensive data analysis, working time models and production interruptions have to be removed from the data set. By counting the bodies with defined throughput times random variables and corresponding distributions can be used to describe the throughput time. But, what is the correct distribution function?

To answer this question several distribution functions where fitted to the distributions from data analysis. To obtain the parameters of the distributions, on the one hand maximum likelihood estimators and on the other hand moments estimators have been used.

In general, empirical distributions of processing times are often skewed to the right, which is a result of the already mentioned production delays. So we have to concentrate on nonsymmetrical density functions.

A commonly used distribution is the exponential distribution.

$$F(x) = \begin{cases} \int_0^x 1 - e^{-\lambda y} dy & , x > 0 \\ 0 & , x \leq 0 \end{cases} \tag{1}$$

with a parameter $\lambda \in \mathbb{R}_{>0}$. The exponential distribution is practically only applied when standard deviation and mean are of roughly the same size. In order to compensate for this the distributions are generalized introducing new parameters. This leads to families of new distributions for example gamma distributions (Curry and Feldman, 2011. Manitz, 2005). The probability density function of the gamma distribution $GAM(\alpha, \lambda)$ is:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & , x > 0 \\ 0 & , x \leq 0 \end{cases} \tag{2}$$

where Γ is the gamma function and the parameters α, λ are positive real numbers.

To study the effects of turbulence on the quality of the sequence and to take into account in the model later, it is necessary to calculate the differences from the target sequence. In order to measure the sequence variations between two stations, the relative order of deviation is calculated. This is the position number at the subsequent station minus the position at the previous station.

A typical distribution for the description of sequence variations is the lognormal distribution (Meißner, 2009). The lognormal distribution is defined by

$$F(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \frac{1}{y} e^{-\frac{1}{2}\left(\frac{\ln(y)-\mu}{\sigma}\right)^2} dy & , x \geq 0 \\ 0 & , x \leq 0 \end{cases} \tag{3}$$

with parameters $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$.

Also the normal distribution has to be mentioned. Although it is unsuitable for production processes that involve quality control and rework because of the symmetry, it is nevertheless an important distribution for production processes (Bayer et al., 2003). The normal distribution is defined by

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy \tag{4}$$

with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}, \sigma > 0$.

Another important issue is the mapping between sequence variation and throughput time. Using cycle time *ZZ*, average stock WIP and sequence variation RFA, it is possible to convert these values into each other (Meißner, 2009). For the throughput time M, one obtains:

$$M = (RFA + WIP)\, ZZ \tag{5}$$

In Figure 2, the densities of the measured and calculated cycle time are shown; they are almost identical (Danner, 2013).

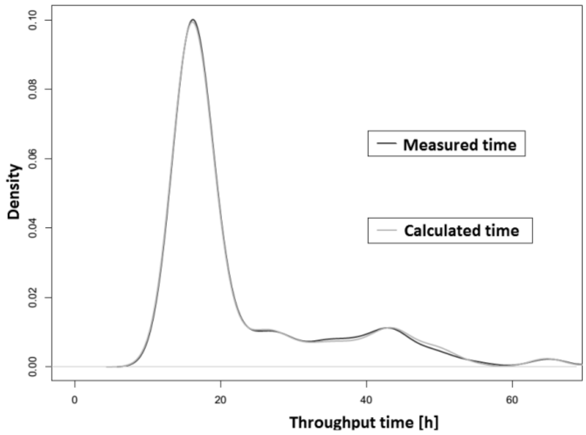

Figure 2: Calculated and measured throughput time

MODELLING THROUGHPUT TIME

In addition to understand and also to refine the distributions derived in the top-down analysis, we did a bottom-up analysis of the paint shop, based on detailed invariant points within the plant and the associated data.

By using the timestamps recorded by the manufacturing information system within the paint shop at each workstation, the paths of individual bodies through the paint shop are evaluated.

Using directed graphs an "empirical topological structure" of the paint shop can be visualized. An example is shown in Figure 3.

The analysis shows that the individual paths through the paint shop are branched and crossing each other in a complex fashion. Yet there are several nodes that all vehicles are passing through. The graph of the paint shop is divided by these common nodes into different sectors. The throughput time distribution is then examined in more detail for these sectors individually.

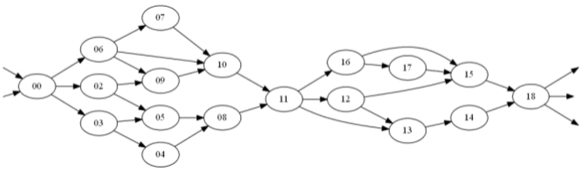

Figure 3: A part of paint shop structure

Data analysis of the throughput time distribution between the nodes suggests a possible separation of the paint process into five major sub-processes. These can be mapped to the big process steps initially described: Pre-treatment, Base coat, Filler coating, Top coating and clear finish, Final inspection and rework.

Throughput time has been analyzed and statistically clustered concerning the color of the top coating (Danner, 2013). For the throughput time we use $M(c)$ depending on the color c of the top coat. Typically, the number of colors is 20.

The throughput time for the paint shop is then leading to a refined model

$$M(c) = m(c) + \sum_{j=1}^{5} \mathrm{X}_j(c) \tag{6}$$

where $m(c)$ is the fixed time for the overall transport on conveyors in the paint shop and $\mathrm{X}_j(c)$ is the random variable for the duration of production step j of color c (Danner, 2013). In order to incorporate the significant influence of rework on the distribution of the throughput time we split the process time $\mathrm{X}_j(c)$ into a process time $\mathrm{X}_{j1}(c)$ for bodies with no re-working performed and a process time $\mathrm{X}_{j2}(c)$ with re-working. This leads to:

$$\mathrm{X}_j(c) = p_j(c)\,\mathrm{X}_{j1}(c) + \left(1 - p_j(c)\right)\mathrm{X}_{j2}(c) \tag{7}$$

where $p_j(c)$ is the probability for an individual body of color c to pass production step j. We assume the random variables to be independent. This is valid only because we neglect effects due to systematic errors like machine misconfigurations or color batches out of specification.From the layout of the paint shop and the distribution analysis of the processes we derive, that the processes X_1 and X_2 have no rework thus we set $p_1 = p_2 = 1$. Additionally we combine the two processes to one random variable Y_1. In a second simplification we combine the three processes with rework $\mathrm{X}_3, \mathrm{X}_4, \mathrm{X}_5$ to one random variable Y_2. Thus we get:

$$M(c) = m(c) + \sum_{k=1}^{2} Y_k(c) \tag{8}$$

Obviously, the simplification reduces the number of unknown parameters of the color depending distributions significantly.

NUMERICAL RESULTS

In order to validate the various models, the real throughput times are compared with the throughput times of the models for every color c by calculating error sums.

To obtain the deviations the time is discretized and then the quadratic differences between the density of the data and the considered distribution M(c) are summed. The most import error sums of the top-down analysis are shown in Table 1 (Danner, 2013).

	$M(c) \sim N(\mu,\sigma)$	$M(c) \sim GAM(\alpha,\lambda)$	$M(c) \sim LN(\mu,\sigma)$
mean	1,15	0,41	0,29
standard deviation	0,19	0,13	0,10
minimum	0,87	0,23	0,15
maximum	1,61	0,65	0,45

Table 1: Error sums of top-down analysis

As expected from the skewedness of the observed distributions, the normal distribution is not suitable for fitting throughput times of a paint shop. The gamma and the lognormal distribution on the other hand show promising results.

Figure 4 shows a typical histogram of the throughput time for one selected color. Also the fit with a lognormal distribution (Danner, 2013) is shown in this graph. The differences between the histogram and the lognormal distribution are very small.

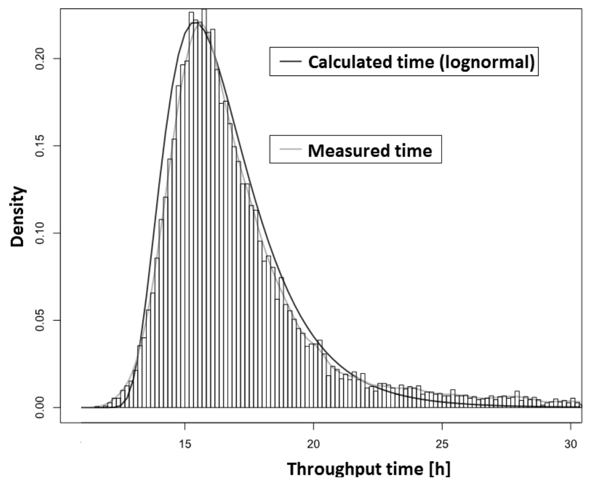

Figure 4: Throughput time of a paint shop fitted with a lognormal distribution

Next the procedure and the results of the bottom-up method are verified and validated. We assume the processes 1 and 2 to follow a normal or gamma distribution. We use the folding invariance of these distributions (Hübner, 2009) to yield a gamma function again. Note that we set the parameter λ to be equal for both gamma functions which is a prerequisite for this relation.

In Figure 5, the distribution of the throughput time for the process Y_1 is compared to the measurement. From this we conclude that there is no rework in processes 1 and 2. Thus the combination of the two processes is a valid simplification of the model in combination with the assumption of a normal or gamma distribution.

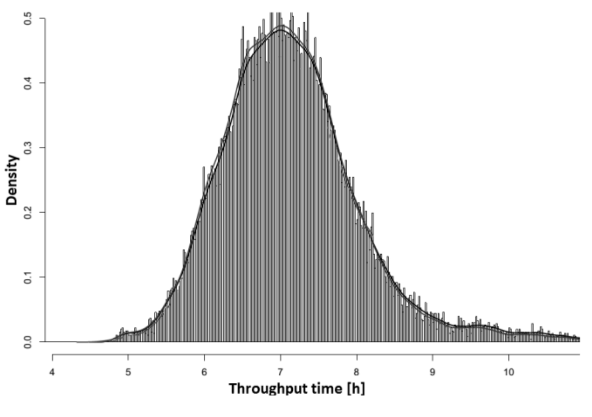

Figure 5: Histogram and density of throughput time $\mathbf{Y_1}$

For the steps 3 to 5 (Y_2) the concept of folding of integrals to unify the three distributions into one cannot be applied (Danner, 2013). For logarithmic normal distributions the folding can result in other functions than the logarithmic normal distribution. For the Gamma function we cannot assume the parameter λ to be equal for all three production steps due to the very different nature of the reworks. Nevertheless, we obtained good results by simplifying the last three processes to Y_2 and fitting with suitable distributions (Danner, 2013).

In Table 2, typical results or the error sums using the bottom-up method are shown (Danner, 2013).

	$Y_1 \sim N(\mu_1, \sigma_1)$, $Y_2 \sim LN(\mu_2, \sigma_2)$	$Y_1 \sim N(\mu, \sigma)$, $Y_2 \sim GAM(\alpha, \lambda)$	$Y_1 \sim GAM(\alpha, \lambda)$, $Y_2 \sim LN(\mu, \sigma)$
mean	0,12	0,23	0,17
standard deviation	0,08	0,09	0,07
minimum	0,07	0,14	0,09
maximum	0,36	0,50	0,36

Table 2: Error sums of bottom-up analysis

From studying various simulation models that have been created with the bottom-up analysis, we found that the most appropriate combination of distributions for the two sub processes Y_1, Y_2 is: $Y_1 \sim N(\mu_1, \sigma_1)$ and $Y_2 \sim LN(\mu_2, \sigma_2)$.

CONCLUSION

The correct description of the behavior of a paint shop forms the basis for an optimal simulation of modern automotive production. Precise reproduction of the processes is not recommended due to the complex structure and the volatility of the processes. Therefore, a methodology was developed to „measure" the behavior of the paint shop using production data. In a top-down analysis of the throughput time a distribution was calculated based on modelling of these datasets using distribution functions. Here the log-normal distribution gives a good description. In a bottom-up analysis, the description was further refined. As a result from our data analysis we showed that a convolution of a normal and a log-normal distribution provides a good stochastic description for a paint shop.

REFERENCES

Arnold, D., Furmans, K., 2009. "Materialfluss in Logistiksystemen". 6., expanded edition, Springer, Berlin, Heidelberg

Bayer, J., Collisi, T., Wenzel, S., 2003. “Simulation in der Automobilproduktion”. Springer, Berlin, Heidelberg.

Curry, G. L., Feldman, R. M., 2011. “Manufacturing Systems Modeling and Analysis”. Second Edition, Springer, Berlin, Heidelberg.

Danner C., 2013. “Optimales Scheduling unter Berücksichtigung stochastischer Wiederbeschaffungszeiten”. master thesis at the University of Applied Sciences OTH Regensburg, Professorship of Informatics/Mathematics.

Hübner, G., 2009. “Stochastik”. 5., improved edition, Springer, Berlin, Heidelberg.

Manitz, M., 2005. “Leistungsanalyse von Montagesystemen mit stochastischen Bearbeitungszeiten“. Kölner Wissenschaftsverlag, Köln.

Meißner, S., 2009. „Logistische Stabilität in der automobilen Variantenfließfertigung“. Thesis at the TU München, Professorship of Materials Handling, Material Flow and Logistics.

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

Spieckermann, S., 2002. “Neue Lösungsansätze für ausgewählte Planungsprobleme in Automobilrohbau und –lackiererei“. Shaker, Aachen.

Weyer, M., 2002. “Das Produktionssteuerungskonzept Perlenkette und dessen Kennzahlensystem”. Helmesverlag, Karlsruhe.

AUTHOR BIOGRAPHIES

THOMAS HUSSLEIN was born in Munich, Germany and went to University of Regensburg, where he studied computational physics and obtained his degree in 1993. He obtained his PhD in physics from the University of Regensburg in 1996. Afterwards he did his Postdoc at the IBM Research Center Yorktown Heights and the University of Pennsylvania. In 1999 the founded OptWare in order to transfer mathematical methods into application which he is leading ever since. Since 2012 he has a teaching assignment for Operations Research at the University of Applied Sciences OTH Regensburg.

JÖRG BREIDBACH was born in Bendorf/Rhein, Germany and went to the University of Kaiserslautern where he obtained his degree in physics in 1998. He got his PhD in physical chemistry in the field of numerical simulation of charge migration in molecules from the University of Heidelberg in 2006. In 2004 he joined OptWare where he is now heading the department for algorithmic development and optimization. Since 2009 he has a teaching assignment at the University of Applied Sciences OTH Regensburg.

MARKUS SEIDL was born in Regensburg, Germany. He studied Mathematics at the University of Applied Sciences OTH Regensburg and obtained his Diploma degree in 2006. Also in 2006 he joined OptWare where he is currently leading the group for product development.

CHRISTIAN DANNER was born in Regensburg, Germany. He studied Mathematics at the University of Applied Sciences OTH Regensburg and obtained his Master degree in 2013. In 2013 he joined OptWare as a member of the group for product development.

WOLFGANG LAUF was born in Kiel, Germany. He studied Mathematics at the Universities of Tübingen and Würzburg, where he obtained his PhD in mathematics in 1994. From 1998 to 2002 he gained a lot of experience in air traffic related optimization and simulation problems at Lufthansa. In 2002 he got a position as a professor for mathematics at the University of Applied Sciences OTH Regensburg. His main research interests are Operations Research, Function Theory and Financial Mathematics.