A method and model for short-term shop floor performance prediction, evaluation and diagnosis

A Method and Model for Short-Term Shop Floor Performance Prediction, Evaluation and Diagnosis

Paul P.M. Stoop
J. Will M. Bertrand
Corné W.G.M. Dirne

Research Report TUE/BDK/LBS/94-08
October 1994

Graduate School of Industrial Engineering and Management Science
Eindhoven University of Technology
P.O.Box 513, Paviljoen F15
NL-5600 MB Eindhoven
The Netherlands
Phone: +31.40.474443

This research is sponsored by the Economics Research Foundation, which is part of the Netherlands Organization for Scientific Research (NWO).

A Method and Model for Short-Term Shop Floor Performance Prediction, Evaluation and Diagnosis

Paul P.M. Stoop, J. Will M. Bertrand, Corné W.G.M. Dirne
Eindhoven University of Technology
Graduate School of Industrial Engineering and Management Science
Department of Operations Planning and Control
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
Research Report TUE/BDK/LBS/94-08

Abstract

In this paper a method is described for evaluating and diagnosing the performance of shop floors with respect to production control aspects such as output and delivery reliability. The focus is directed at the short term performance. On the short term, the state of the production system to a large extent determines the possible performance of the system. For the performance evaluation, short term predictions of the performance are made that are being used as short term performance norms. The prediction model deviates from existing prediction models in that it takes into account actual available information instead of average values; the prediction rules are state-dependent. Implementation of the model in a turnery shop resulted in low prediction errors as well as low standard deviations of the prediction errors as compared to the use of either the predictions of a long term model (steady-state) or the use of external output requirements.

1. Introduction

In many companies performance measurement has become a standard procedure. At many organizational levels performance indicators are defined and being used to monitor the performance. In former days, performance measurement was restricted to financial measures only (Kaplan [1983], Andersson *et al.* [1989]). Nowadays, new non-financial performance indicators are added to the existing measures to evaluate other performance aspects such as quality and timeliness. At the shop floor level non-financial indicators predominate over financial ones (Gelders [1990], NEVEM [1989], Weston and Brothers [1984]). Examples of relevant performance indicators at the shop floor level are: mean order flow time, order delivery reliability and capacity utilization.

In general, we can observe that the actual performance in organizations on most performance indicators fluctuates over the course of time. Periodically, e.g. at the beginning of each week, the performance over the last period is evaluated and discussed. This performance evaluation is done by comparing the latest reported performance (or the trend in the performances) on each performance indicator with a performance norm. Usually, this performance norm is set as a time-independent norm for a longer period of time and is based on either the past average performance or a stated managerial goal. Deviations between the real performance and the long term performance norm may lead to actions for improvement. In practice, operations managers often state that performance deviations from the performance norm can easily be explained from recent events in the production system. Examples of these kinds of explanations are: the occurrence of unexpected operator absenteeism and unexpected machine breakdowns, and complex orders that disturb the progress of other orders. In many cases these explanations sound reasonable. However, no methods are available where these explanations are tested explicitly on their veracity. The finding of the real causes of performance deviations is the process of diagnosis. These causes can then be eliminated in order to improve the performance, or -if cause elimination is impossible- they should be taking into account in either the performance norm setting or the performance explanation.

In this paper we develop a model that can be used for predicting the short term performance of the shop floor as well as for explaining the performance that is achieved. Thereby, we only consider factors that can be quantified. The explanations for performance deviations can be used either to start up actions for performance improvement, or to make more reliable performance predictions. In section 2 a review of the literature about performance evaluation and diagnosis is given. In section 3 we discuss our method for evaluating and explaining the performance. The prediction model that is used within the method will be described in section 4. In section 5 empirical results are presented. Finally, in section 6 the main conclusions are summarized and directions for further research are suggested.

2. Performance evaluation and diagnosis

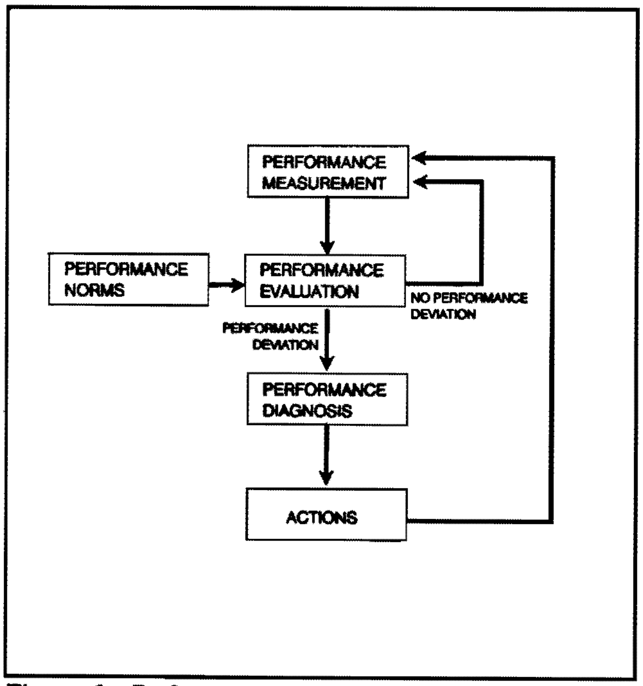

Figure 1. Performance measurement, evaluation and diagnosis.

An important reason to measure and monitor the performance is to improve the performance; performance measurement is directed at continuous improvement. Another reason for performance measurement is to evaluate how well people control the tasks they are responsible for. In this paper we assume that the performance is periodically reported and evaluated. The measured performance is compared with performance norms (the performance evaluation) and in case a deviation is found a diagnosis is started to find the causes for the deviations. The results of the diagnosis are then being used to start up actions for performance improvement. This continueing process of measurement, evaluation and diagnosis is depicted in Figure 1.

In the Operations Management literature we can observe that the models mainly focus on performance predicting. For example, Bertrand *et al.* [1990] state that performance models are "models used to estimate the performance of a controlled system". Another characteristic of most OM models is that they are directed at estimating the steady state performance of a system. The short term performance however, varies over time due to the dynamic nature of all processes that take place at the shop floor. Further, these steady state models do not take into account the actual state of the system as well as the expected input for the short term, which may have a considerable influence on the short term performance. This, for example, is experienced in the field of scheduling, where these dynamic processes in the system may lead to poor predictions (e.g. Gary *et al.* [1994], Kempf *et al.* [1993], McKay *et al.* [1989]). The underlying reasons for focusing our attention on the short term performance are twofold. First, a short term performance prediction (that can serve as a short term performance norm) helps operations managers (and their executives) to see what performance is likely to be achieved in the short term, and to what extent this predicted performance will differ from the performance norm that is set for a fixed long term period. This will avoid lengthy discussions about performance deviations which could not have been avoided anyway. The second reason is that short term (or state and input dependent) predictions of the performance enable operations managers to evaluate their planned decisions for the next measurement period. This enables them to control some of their activities in advance.
With short term performance measurement and evaluation operations management also automatically have a quick feedback of their decisions of the foregoing measurement period. Short feedback loops make the search for causes of performance deviations easier, because over a short time period possible causes for deviations can be obtained more easily (Conellan [1978]).

The literature about performance diagnosis is very scarce. A short overview of the literature about diagnosis is given by Wagner [1993]. He makes a distinction between "causal diagnosis" (cf. Smith [1989]) and "situation understanding" (cf. Bouwman [1983]). Causal diagnosis denotes the task of determining a problem's cause, while situation understanding interprets diagnosis as the identification of the state of the underlying system on the basis of a set of observable symptoms. Causal diagnosis is triggered by an undesired situation, whereas situation understanding only results in an interpretation of the current state of the system without knowledge about its desired state. In the remainder of this paper we use the definition of causal diagnosis, thereby speaking of only diagnosis.

One of the few available models that has been developed for diagnosis purposes in production shops is the model developed by Wiendahl and Ludwig [1991]. Their model is based on expert learning and therefore called a knowledge based model. Based on the observed performance on different measures, their model gives a list of possible causes for performance deviations. In contrast to other performance models, their model is primarly developed for monitoring functions instead to predicting functions. Because the list of possible causes is always limited, and because the real knowledge about a shop's performance comes from the people working in the shop and not from a (knowledge based) model, we think this model can never be complete.

3. Method for evaluation and diagnosis

The method we propose for performance evaluation and diagnosis is depicted in Figure 2 and will be explained in the remainder of this section. To evaluate the reported performance of the shop (and implicitly the quality of the decisions that were made) one needs performance norms. A good performance norm should be achievable. A performance norm that is set for a longer period of time will generally not always be achievable each measurement period; due to the fluctuations in the actual state and input of the system, there will be periods in which the actual performance is far better than the long term norm, and there will be periods that the real performance is worse. For these reasons, a performance norm is needed for the short term, or more specific a performance norm for the next measurement period. This performance norm will be calculated in a prediction model. This prediction is among others based on the actual state of the shop in terms of the characteristics of the current work in progress, the expected available capacity, and the expected work supply. A more detailed description of the prediction model is given in the next section. Besides the predicting function of the prediction model, this prediction model is also used for evaluation as well as for explaining the short term performance, as we will see in the remainder of this section.

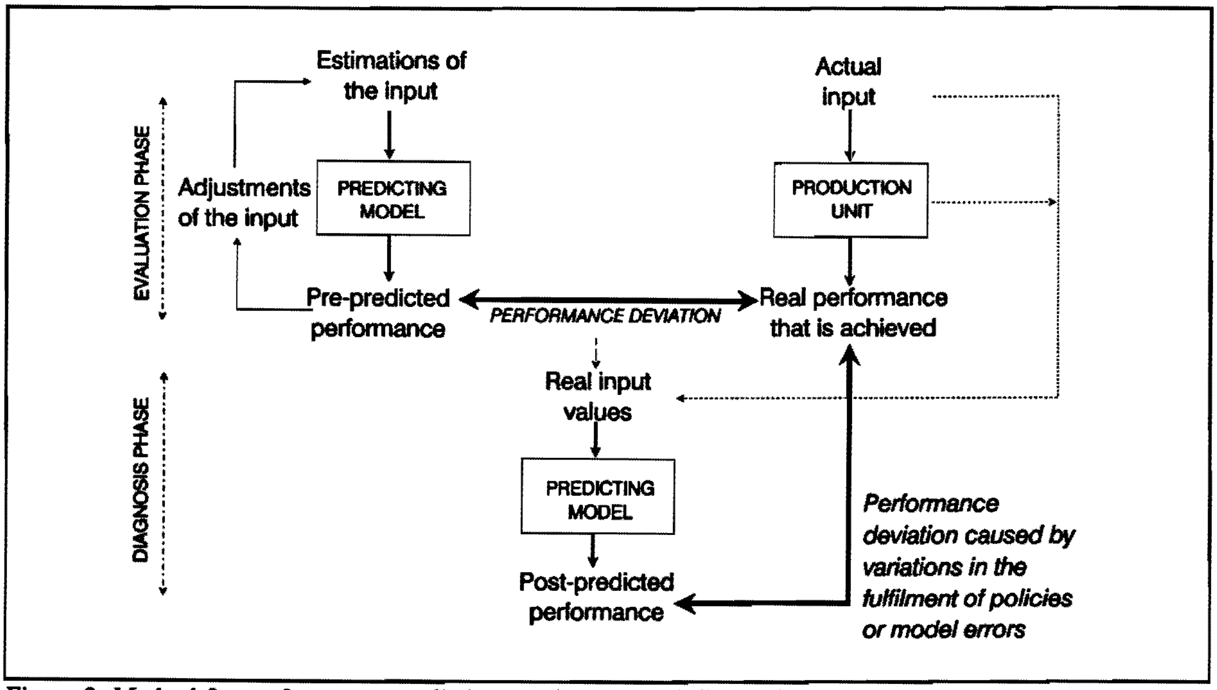

Figure 2. Method for performance prediction, evaluation and diagnosis.

Another property of the prediction model is its function for evaluating the influence of planned decisions about capacity allocation and order release. The evaluation of these kinds of decisions is the first part of the evaluation and diagnosis method. Based on the actual state and expectations or plans about capacity allocation and worksupply, one is able to foresee problems in the short term, such as full capacity utilization that will result in long waiting times and long order flow times. The impact of a planned capacity allocation and order release can be evaluated and changed till an acceptable performance results. This predicted performance will be called the pre-prediction. The pre-predicted performance serves as the short term performance norm; given the actual state of the system and the expectations about the input, this performance can be achieved.

At the end of each measurement period the shop's performance of the foregoing period is reported and fed back to all whom it may concern. For example, operators may get information about their efficiency, shop floor managers about last period's delivery reliability, and financial managers about machine utilization levels. The actual performance will now be compared with the predicted one, that served as the performance norm for the evaluation. In many cases a performance gap can be observed which has to be explained in the diagnosis phase.

All measured variables that may have an impact on the performance and that differ from the expected values should now be considered as possible causes for the observed performance deviation. A so-called post-prediction is made by replacing expected variable values by realized variable values. The starting point for the post-prediction is the initial state of the shop (at the beginning of the foregoing measurement period). In this way the effect of the observed deviations between expected and real input on the performance can be determined. The post-predicted performance may still differ from the actual performance. This deviation will then be due to variations in the fulfilment of certain policies or unformalized decisions. Behavior deviations are difficult to quantify and thus also difficult or even impossible to model, although they may have an important influence on the performance (see for instance O'Leary-Kelly and O'Leary-Kelly [1993]). A substantial gap between the post-predicted performance and the performance that has been achieved thus indicates that the real behavior deviates from the behavior one is expected to show. However, it may also be the case that the system that has been modelled is not completely understood. this in turn will lead to model errors, which are another possible cause for the observed deviations in performance.

4. Prediction model

The short term predictions are made to get an idea of the performance that can be achieved in the short term. This predicted performance will generally differ from the long term performance norm because the predictions are based on the actual state instead of the "average state" of the shop. Further, short term expectations about, for example, work supply and capacity are more rational than the average values. In other words, it is assumed that state-dependent predictions perform better than state-independent predictions. Therefore, the short term predicted performance can very well serve as performance norm in the evaluation phase.

The state-dependent prediction rules that were developed should have to meet the following demands. The rules should be:

- easily to understand by practitioners, in other words they should be in accordance with the line of thought of practitioners;
- as simple as possible;
- robust in the sense that slightly different input values of the model will not result in extreme differences in the output of the model;
- perform better than steady state predictions or externally set performance norms (in terms of average prediction error and -more important- variance of prediction error).

It should be noted that the development of optimal prediction rules is not our main purpose; prediction rules that are "good enough" suffice.

An important characteristic of our prediction methods is that it is time-aggregated in nature, that is that it does not take into account the specific moment within a prediction period an event happens. For example, machine breakdowns (as an example of capacity availability) are represented in an aggregated way; only the total number of hours of breakdowns during the period are being used in the calculations, and not each breakdown apart. There are several reasons for this simplification. First, the prediction rules keep simple and understandable by practitioners. Second, it will be more difficult to predict when each event takes place than to predict the total impact of one type of an event; this relates to the robustness of the model. Therefor, a discrete event simulation model is not considered here. If no aggregation is used, a distribution of specific event times would be a first requirement. However, each input derived from this distribution for the next prediction period has a very little probability of occurrence, so the exact prediction of future events is impossible.

The following variables are considered to influence the short term performance:

$I_i(t)$: the actual work in process at workcenter i at the beginning of period t (hours)

$C_i(t)$: the available capacity at workcenter i during period t (hours)

$W_i(t)$: the work supply to workcenter i during period t (hours)

$S_i(t)$: the amount of scrap at workcenter i during period t (hours)

The capacity at a workcenter is a function of available operator capacity and available machine capacity. The aggregate measures for the total shop (indexed with an s) can easily be obtained by adding the values of all workcenters.

$$C_s(t) = \sum_{i=1}^{n} C_i(t) \tag{1}$$

$$W_s(t) = \sum_{i=1}^{n} W_i(t) \tag{2}$$

$$S_s(t) = \sum_{i=1}^{n} S_i(t) \tag{3}$$

where n is the total number of workcenters in the shop.

For the predictions, estimations of available capacity, work supply and scrap are being used. These will be denoted by $\hat{C}_i(t)$, $\hat{W}_i(t)$, and $\hat{S}_i(t)$ respectively. The predicted work in process is denoted as $\hat{I}_i(t)$. The work supply of a workcenter consists of new orders released for production (external work supply) and orders coming from other workcenters (internal work supply). Using a list of all work in process and a list of orders to be released in the future and their remaining routings, an estimation of the expected work supply can be made. When there is no insight in which orders are going to be released, the long term average work supply with mean order characteristics can be used as an estimator for the external work supply.

The expected work in process at a workcenter for the next period, $\hat{I}_i(t+1)$, is determined by the starting work in process level, $I_i(t)$, the expected work supply and the expected capacity. In formula:

$$\hat{I}_i(t+1) = \max\,(0,\ I_i(t) + \hat{W}_i(t) - \hat{C}_i(t)) \tag{4}$$

We consider the predicted output (hours) of the shop and the predicted completion times of orders as the main performance measures. Other measures like mean order flow time, delivery reliability, and capacity utilization can be deduced easily from these two measures. The predicted output, $\hat{O}_i(t)$, for a certain period is determined by comparing the actual work in process plus the expected work supply with the available capacity. The result is adjusted for expected scrap.

$$\hat{O}_i(t) = \min(\ \hat{C}_i\ (t),\ I_i(t) + \hat{W}_i(t) - \hat{S}_i(t)\ ) \tag{6}$$

For the prediction of order completion times, the following variables are needed.

$V_{i,j}$	=	collection of orders at workcenter j with a higher priority than order i
$Z_{i,j}$	=	collection of orders with a higher priority than order i, that are expected to arrive at workcenter j within the coming measurement period

$F_{i,j}$ = expected flow time of order i at workcenter j
$MW_{i,j}$ = minimal waiting time of order i at workcenter j
$EW_{i,j}$ = expected extra waiting time of order i at workcenter j
$p_{i,j}$ = processing time of order i at workcenter j

The calculation of the prediction of the completion times is split into two parts. First, the expected flow time of the order at the current workstation is calculated. This is the summation of:

(a) the processing times of orders with a higher priority that are waiting in the same queue;
(b) the processing times of orders with a higher priority that are expected to arrive within the period of time calculated in (a);
(c) the processing time of the order at the current workcenter.

In formula:

$$F_{i,j} = MW_{i,j} + EW_{i,j} + p_{i,j} \tag{6}$$

where

$$MW_{i,j} = \sum_{x \in V_{i,j}} p_{x,j} \tag{7}$$

$$EW_{i,j} = \sum_{y \in Z_{i,j}} p_{y,j} \tag{8}$$

The estimation of $Z_{i,j}$ is done as follows. For each workcenter - except for workcenter j- an estimation is made which orders will be processed at these workcenters based on the comparison of the current queue at the workcenters and the available capacity at the workcenters. This collection of orders is reduced by taking out the orders which do not have their subsequent processing at workcenter j. This implies that only orders are considered that are one workcenter before workcenter j. Finally, it is determined which of the selected orders has got a higher priority than order i.
Having calculated the expected flow time at the first workcenter in the remaining routing, we now predict the expected flow time at the remaining workcenters in the routing of an order. The expected waiting times at the other workcenters is approximated by the amount of work in process at these workcenters at the moment this order arrives. So, the expected flow time at the second workcenter, k, in the routing of order i is determined by

$$F_{i,k} = \hat{I}_k(F_{i,j}) + p_{i,k} \tag{9}$$

The amount of work in process for a workcenter in future measurement periods is estimated by using the iteration method of de Kok [4]. Because it is very difficult to predict which specific orders will be at a workcenter at a specific moment, we only predict the amount of work at the workcenter. This amount of work divided over the total number of parallel machines in the workcenter is expected to be the waiting time of the order at the subsequent workcenter. In fact, a First Come First Serve approximation is assumed in this calculation, just like most steady state models do.

At the end of the measurement period (or the beginning of the next period) the real values of the variables are known. It is assumed that the differences between the expected values and the real ones are caused by the unexpected variations that enter the shop and cause variations in the performance.
The differences between the actual and planned or expected values of capacity and work supply can be modelled as follows:

$$\underline{\theta_i}(t) = C_i(t) - \hat{C}_i(t) \qquad (10)$$

$$\underline{\psi_i}(t) = W_i(t) - \hat{W}_i(t) \qquad (11)$$

$$\underline{v_i}(t) = S_i(t) - \hat{S}_i(t) \qquad (12)$$

where $\underline{\theta}_i$, $\underline{\psi}_i$, and $\underline{v}_i$ are distributed with known averages and standard deviations. The type of distribution can be obtained by fitting observed data in practice with known distributions.

A factor that is not mentioned before, but that may very well influence the performance of a shop is the difference between calculated processing times and real processing times. So,

$$\underline{\epsilon_{i,j}} = p_{i,j} - \hat{p}_{i,j} \qquad (13)$$

where $\underline{\epsilon}_i(t)$ is a stochastic variable that represents the difference between pre- and post-calculation at workcenter i, with known average and standard deviation, for example $\underline{\epsilon}_i(t) \sim NID(\mu_\epsilon, \sigma_\epsilon)$.

A last remark bout the model is that the predictions of the model are point estimations; only the first moment of the predictions is considered. The accuracy of the predictions can be obtained from the prediction errors of the past.

5. Empirical results

The model described in the previous section has been implemented in a production department which is part of the supply chain of a Dutch aircraft manufacturer. The material flow over the entire supply chain is controlled by an MRP system. In the production department, that is structured as a job shop, 9 functional workcenters can be distinguished. The work is done in two shifts of 8 hours with 12 operators per shift. The department produces all kinds of turnery products such as guards, screws, pins and spacers.

Upon the implementation of the model the department management stated that the shop floor's performance should be measured and evaluated on the two performance measures: output for the next week and order completion dates. The output measure could be used by the production department management to evaluate the short term capacity utilization levels and the internal capacity allocation over the workcenters. The predictions of the order completion dates could be used by MRP order planners to evaluate the progress of individual orders and to inform customers timely about deviations between planned and expected order completion dates to facilitate the work planning at the succeeding production phases..

The prediction rules that were used for the production department were developed under the following assumptions:

- The measurement period was set at one week.
- Orders were released based on input-output control. So, the amount of work to be released for the next period equals the output of the current measurement period.
- Orders were processed at the workcenters in order of operation due date.
- Initially, only capacity and work supply were regarded as the major causes for performance deviations. The other possible causes for performance deviations (scrap and differences between expected and real processing times) were not included into the model, because these factors could not yet be measured automatically. Besides, the production department management expected that work supply and capacity deviations were the main causes for performance deviations.

The quality of the performance predictions was expressed in terms of average prediction errors (planned or predicted performance minus real performance) and the standard deviation of the prediction errors. The prediction results are compared with the performance as planned by MRP, that can be regarded as a prediction based on long term data.

In Figure 3 the results for the output predictions of eight subsequent weeks are presented. The average output in a week was 290 hours. As is shown in Figure 3, we may conclude that the state-dependent prediction rules outperform the state-independent (MRP) rules in average error as well as standard deviation of the prediction error. Further, we see that the real output performance could not be explained completely by deviations in capacity and work supply; the average post-predicted prediction error is about 20 hours. On the average, an error of about 10 hours is caused by deviations between expected and real capacity and work supply. Additional research showed that the sequencing at the workcenters was not in

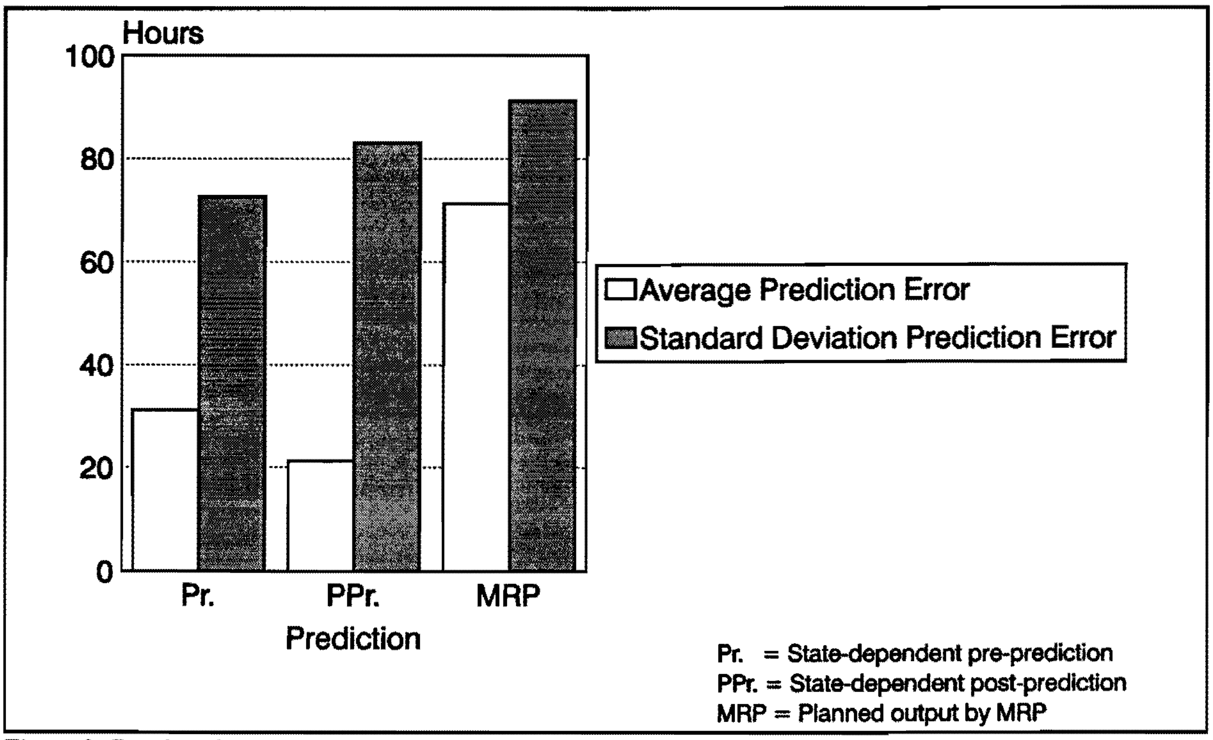

Figure 3. Results of output predictions.

order of operation due date, but more at random. Although this behavior deviation may have an impact on the performance, production management thought that other causes had a greater impact, namely rework and deviations between expected and real processing times. These possible causes are now subject for further research. The learning process has been started to find the real causes for performance deviations, which is the main purpose of the method for evaluation and diagnosis.

The quality of the predictions of the order completion dates are presented in Figure 4. The average flow time in the shop is about 2 weeks. The predictions are classified into horizons which indicate the number of weeks the predictions are made for. For example, in the class horizon 1-2 all the predictions are gathered that expect orders to be completed between 1 and 2 weeks. Again, we see that the state-dependent prediction rules result in more reliable completion dates than the MRP-planned completion dates. The high reliability makes it possible for MRP order planners to timely give reliable promises about order arrivals at their customers. Finally, it is remarkable that the post-predictions do not significantly perform better than the pre-predictions, or in other words, deviations in expected and real capacity and work supply do not have an influence on the order progress. This might be due to the random order in which the orders are processed at the workcenters. Further research is directed at the explanation of these findings.

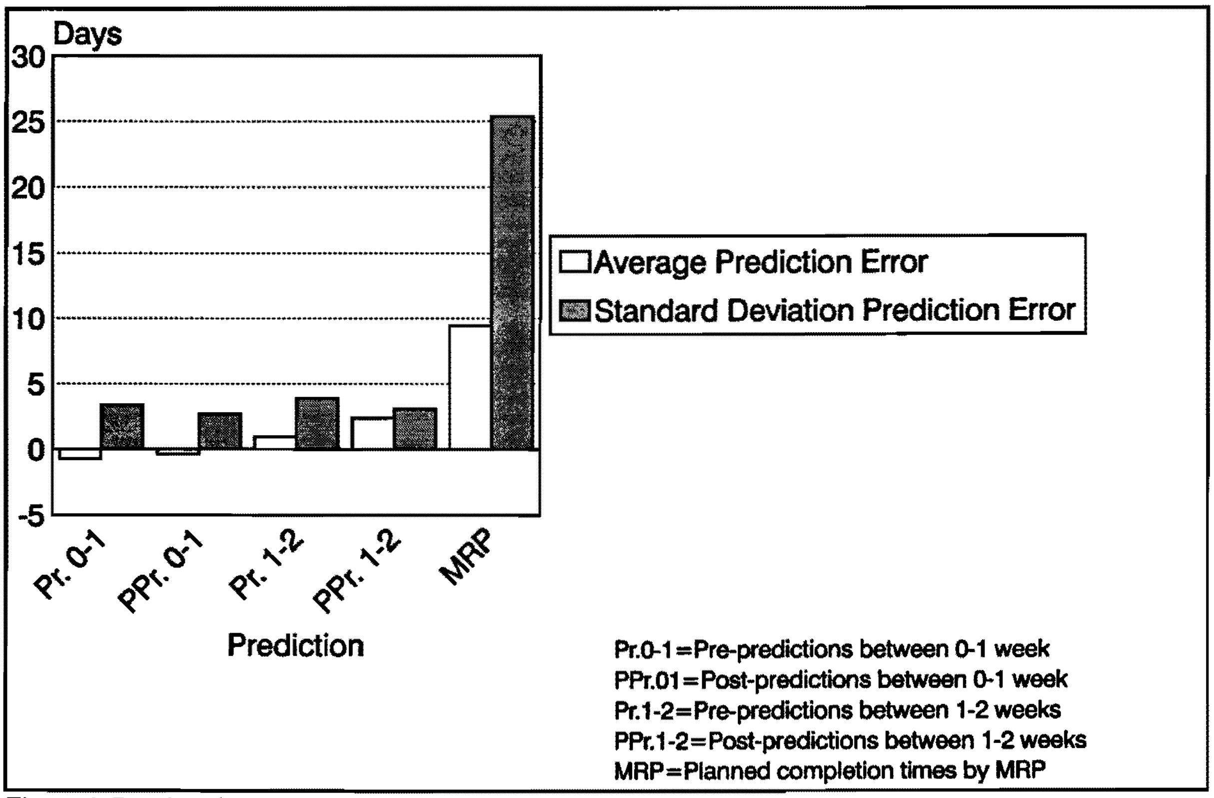

Figure 4. Results of completion time predictions.

6. Conclusions

In this paper a method is described to evaluate and diagnose the short term performance of job shop structured production departments. Within the method a prediction model is used to predict the expected performance of the shop and to predict what performance could have been achieved, knowing what actually has been happened. The model is state-dependent, which means that the actual state of the shop and the expected input for the shop are taken into account. This characteristic makes the model suitable for short term predictions, which are needed at shop floor management level to improve their performance. The prediction model was tested in a real life situation and the results were quite impressive. The state-dependent prediction rules performed far better than the state-independent or planned performances in terms of average and standard deviation of the prediction errors. However, more empirical research is needed to explain an achieved performance with the help of models.

References

Andersson P., H. Aronsson, and N.G. Storhagen, 1989. Measuring Logistics Performance. *Engineering Costs and Production Economics*, vol.17, pp.253-262.

Bertrand, J.W.M., J.C. Wortmann, and J. Wijngaard, 1990. *Production Control. A Structural and Design Oriented Approach*. Elsevier Science Publishers B.V., Amsterdam.

Bouwman, M.J., 1983. Human Diagnostic Reasoning by Computer: an Illustration from Financial Analysis. *Management Science*, vol.29, iss.6, pp.653-672.

Conellan, Th., 1978. *How to improve human performance?* Harper & Row.

Gary, K., R. Uszoy, S.P. Smith, and K. Kempf, 1994. *Measuring the Quality of Manufacturing Schedules*. To appear in: Intelligent Scheduling Systems, D.E. Brown, W.T. Schener (eds.), Kluwer Academic Publishers.

Gelders, L., 1990. Shop Floor Performance Indicators. *Logistics Information Managment*, vol.3, iss.1, pp.45-48.

Kaplan, R.S., 1983. Measuring Manufacturing Performance: A New Challenge for Managerial Accounting Research. *The Accounting Review October*, vol.58, iss.4, pp.686-705.

Kempf, K., R. Uszoy, S. Smith, and K. Gary, 1993. *Evaluation and Comparison of Production Schedules*. Research Memorandum no.93-14, Purdue University.

Kok, A.G. de, 1989. A Moment-Iteration Method for Approximating the Waiting-Time Characteristics of the GI/G/1 Queue. *Probability in the Engineering and Informational Sciences*, vol.3, pp.273-287.

McKay, K.N., J.A. Buzacott, and F.R. Safayeni, 1989. *The Scheduler's Knowledge of Uncertainty: The Missing Link*. Knowledge Based Production Management Systems, J. Browne (ed.), Elsevier Science Publishers B.V. (North-Holland), IFIP.

NEVEM-workgroup, 1989. *Performance indicators in logistics*. Springer-Verlag Berlin-Heidelberg-New York-Tokyo, IFS Publications, UK.

O'Leary-Kelly, S.W. and A.M. O'Leary-Kelly, 1993. *Organizational Behavior Contributions to the Field of Operations Management*. Proceedings of the 24th Annual Meeting of the Decision Sciences Institute, Washington D.C.

Smith, G.F., 1989. Defining Managerial Problems: a Framework for Prescriptive Theorizing. *Management Science*, vol.35, iss.8, pp.963-981.

Wagner, C., 1993. Problem Solving and Diagnosis. *OMEGA International Journal of Management Science*, vol.21, no.6, pp.645-656.

Weston, F.C., and W.S. Brothers, 1984. Productivity Management by the Numbers. *Production and Inventory Management*, vol.25, iss.3, pp.54-67.

Wiendahl, H.P. and E. Ludwig, 1991. *A Knowledge Based Diagnostic System for Production Processes - Basics and Concepts*. Advances in Production Management Systems, E. Eloranta (ed.), Elsevier Science Publishers B.V. (North-Holland), IFIP.