ESTIMATING CAPACITY LOADINGS FROM WORK CENTER ARRIVAL AND DEPARTURE EVENTS[1]

Manuel D. Rossetti[†] **and Gordon M. Clark**[‡]

[†] Department of Systems Engineering
University of Virginia
Thornton Hall
Charlottesville, VA 22903

[‡] Department of Industrial, Welding, and Systems Engineering
The Ohio State University
210 Baker Systems
1971 Neil Avenue
Columbus, OH 43210

Abstract

We consider the use of bar code scanning timing information from an automatic data collection system to estimate capacity loadings within a computer integrated manufacturing environment. We evaluate the accuracy and precision of the estimators via discrete event simulation under various experimental conditions. Key factors which effect the performance of the estimators are identified. We found that the percentage of the products in the production mix can have a significant effect on estimator performance. Also, the estimator's performance can degrade significantly if the observed production mix is significantly different from the mix in the production schedule for which capacity loadings are desired. The importance of the research is the value added benefits to be gain in utilizing bar code timing information for developing estimates of capacity loadings.

[1] Working Paper SYS-MDR-5-1994, Revised June 20, 1995

1 Introduction

One of the important benefits of Computer Integrated Manufacturing (CIM) is improved control of the manufacturing system. Improved control over the manufacturing system is facilitated through the use of CIM system data. The use of CIM system data allows the functional areas of the manufacturing firm to be integrated. For the purposes of our research, CIM system data are any data available from or about the CIM system. For example, CIM system data includes part geometry data, bills of material, inventory status data, routing data, tooling data, and production activity control data. From the CIM data, information is extracted to aid in the decision making necessary to control the functions of the firm.

Two fundamental questions must be addressed in order to be able to improve the operations of a manufacturing system using available CIM system data. First, how can the data be made usable? Second, how should the data and results be used to make decisions? The complexity of many CIM systems makes addressing the above questions difficult.

Our research represents an example of how CIM system data in the form of bar code scanning data within a computerized production activity control system can be utilized as inputs to perform Capacity Requirements Planning (CRP). Computerized production activity control systems permit the collection of large amounts of event oriented data. Smith[15] presents typical examples of event oriented data, such as: material movement started, material movement completed, setup started, setup ended, processing started, processing ended, operator arrives, operator departs. Large amounts of automatically collected data offer a tremendous opportunity for the improvement of decision making within the manufacturing system. Although computers allow large amounts of data to be collected, the data may not be complete or in a directly usable form.

A common decision in managing a manufacturing system is the scheduling of production over some time horizon. An important piece of information for any scheduling method is the available capacity of the system as a whole and of the individual sub-systems. A standard method for determining the capacity of a manufacturing system is Capacity Requirements Planning (CRP) which determines whether the Material Requirements Plan (MRP) for a master production schedule is achievable and specifies the resources required to perform the MRP. CRP requires the time to perform operations by product type for each potential bottleneck facility or work center.

In order to properly estimate the capacity, one needs to know the work content of the products being produced. The work content is a direct function of how long the product spends at each of the production operations. Engineering estimates of processing time can be used to represent the work content of products; however, actual production operations experience disruptions such as facility down times for repair and additional time for rework. Operation times which refer to the total time to perform the production operation on the product more realistically represent the work content of the product since they not only include the processing time but may also include allowances for such time elements as setup, facility disruptions, and rework. Operation times can thus serve as the fundamental element in determining the capacity of manufacturing systems. Normally, a manufacturing concern would require scarce engineering resources to establish standard times for key operations which could then be used as inputs to CRP. Manual work measurement is a costly approach to solving the data collection problem. The potential exists to utilize already collected data from computerized production activity control systems to estimate inputs for CRP.

The overall goal of this research is to demonstrate the practical value of utilizing available CIM system data for estimating system characteristics, e.g. operation times, from event oriented data and then using the estimates for important decision making functions such as capacity planning and scheduling. We feel that research involving the use of available CIM system data will become increasingly important as manufacturing firms move towards automated systems with automated data collection.

To achieve our overall research goal, we develop a methodology which uses stochastic models and statistics in order to estimate mean operation times; see Rossetti and Clark[12] for more details. The operation times can then be utilized to estimate capacity loadings for the work centers of the manufacturing system. We utilize CIM system data in the form of bar code scanning records of products arriving to and departing from a work center. We would expect that:

- The accuracy of the estimates should meet the requirements for CRP.
- Observation periods as long as six months are available on shop floor operations.
- The effort required by shop floor personnel to collect data should be minimized. Ideally, the effort to collect data by shop floor personnel should be no more than they are currently doing to provide inputs for existing shop floor control systems.

In order to evaluate the estimators produced by our methodology, we evaluate the estimator's statistical properties via simulation. The objective of the simulation experiments presented in this paper is to assess the accuracy and precision of predicted capacity loadings based upon estimated mean operation times. The simulation experiments examine the effect that different system characteristics and sampling methods have upon the performance of the estimators.

The rest of this paper is organized as follows. In Section 2, we briefly describe the manufacturing system. In Section 3, we present the mathematical model and its assumptions. In Section 4, we present the simulation evaluation of the estimated capacity loadings. Finally, we summarize the research and major results.

2 System Description

The research sponsors selected circuit pack production as a system for focusing our efforts to develop and test estimation procedures which utilize CIM system data. A brief description of the product flow through work centers helps in developing a context for the problem description. Figure 1 depicts this flow where a box in the figure such as the one labeled MACHINE INSERTION indicates a work center.

Shop floor control systems exist which can track the flow of work through the work centers using bar code labels on each item of work. Figure 2 illustrates the work center structure to be examined in this paper. This parallel facility structure closely approximates the structure of the functional test work center in our circuit pack production system. The work center consists of a number of parallel identical facilities. The facilities are parallel in that any one of them can perform the necessary work and identical in the sense that the mean operation times are the same for each facility. That is, the mean operation times only depend on the product code. A single queue feeds these machines and the order of items of work in the queue can not be accurately predicted from the work center arrival times. In Rossetti and Clark[12], we utilize the bar code scanning data to estimate the operation times for this system configuration. The following section indicates how the estimated operation times can be used to estimated capacity loadings for the work center.

3 Mathematical Model

In this section, we present a least squares model which can be used to estimate capacity loadings given estimated operation times. We refer the reader to Rossetti and Clark[12] for more details of the model and for the evaluation of the estimated operation times.

We assume that the shop floor control system and other data sources can provide the busy time for a work center and the production counts by product type over a data collection

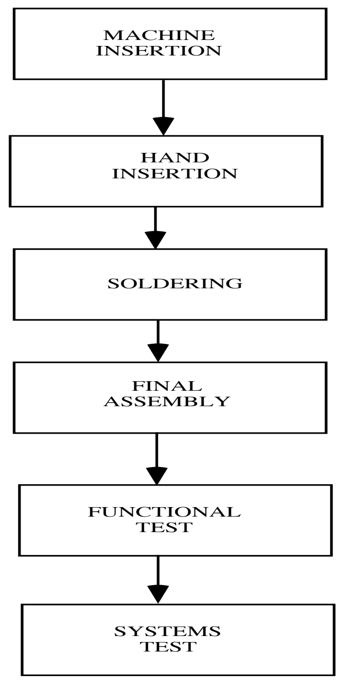

Figure 1: Circuit Pack Work Flow

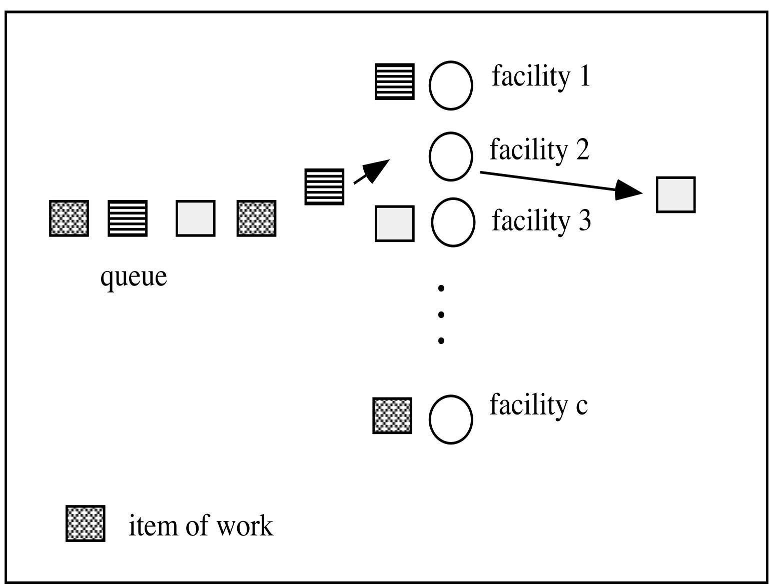

Figure 2: Work Center with Parallel Facilities

period which has been divided into fixed length time buckets. We define θ_i to be the mean operation time for product i. For a work center, the available data are:

Let Y_j = the total facility busy time for time bucket j

Let X_{ij} = the amount of product i produced during time bucket j

Let K = the number of product types observed during the data collection period

Let N = the number of time buckets observed during the data collection period

The value of Y_j represents the total amount of busy time for a work center. Because θ_i is the mean time to produce one unit of product i, the multiplication, $X_{ij} \times \theta_i$, represents the mean time to produce X_{ij} products of type i during time bucket j. If we sum the quantities $X_{ij} \times \theta_i$, over all the products, we obtain the total amount of expected busy time for time bucket j. So, we propose the following linear model:

$$Y_j = \sum_{i=1}^{K} X_{ij}\theta_i + \epsilon_j \qquad j = 1, \ldots, N \tag{1}$$

The term, ϵ_j, represents the error in using $\sum_{i=1}^{K} X_{ij}\theta_i$ to estimate Y_j. Estimates for the parameters, θ_i, can be obtained by minimizing the error in a least squares sense. The least squares problem can be represented as finding those values of θ_i which satisfy the following:

$$\min \text{SS} = \sum_{j=1}^{N} \left(Y_j - \sum_{i=1}^{K} X_{ij}\theta_i \right)^2 \tag{2}$$

$$\text{subject to:} \quad \theta_i > 0 \qquad i = 1, \ldots, K \tag{3}$$

The number of parameters for this model is K, so that $N \geq K$ is required for proper estimation.

Throughout the following sections we will denote an estimator with the symbol, " ˆ ", for example, an estimator of θ_i is denoted by $\hat{\theta}_i$. We note that the capacity work load requirements for a given amount of product to produce is just the sum over all products of the amount of product to produce times the mean operation times. For example, suppose we have two products with mean operation times $\theta_1 = 20$ and $\theta_2 = 10$ minutes per product. Furthermore, suppose that the production schedule calls for 100 type 1 products and 200 type 2 products to be produced for a week, then the capacity work load requirement for that week would be $100 \times 20 + 200 \times 10 = 4000$ minutes. If the estimated values of θ_1 and θ_2 are $\hat{\theta}_1 = 19$ and $\hat{\theta}_2 = 11$, then the predicted capacity work load requirement would be $100 \times 19 + 200 \times 11 = 4100$ minutes. We note that the predicted capacity work load based on estimated mean operation times is just an application of the linear model given in Equation 1 to predict Y given values for the X's.

4 Simulation Investigation of Estimator Performance

To assess the statistical properties of the estimates for capacity loadings, we simulated the parallel work center structure under a variety of experimental conditions. The simulation model of the parallel work center generates the busy time and product count data for the least squares model. The operation times are then estimated from the simulated data. Representative production schedules are generated. Then, using the estimated operation times and the production schedules, estimates of the work center capacity loadings are obtained. The estimates of capacity loadings are then evaluated against the actual capacity loadings based on knowledge of the true value of the operation times. The overall modeling assumptions and parameter settings for the simulated work center are given in Rossetti and Clark[12]. The rest of this section is structured as follows. First, we give the experimental inputs and notation, and then we present a description of a typical experiment. The experimental outputs and notations are then presented along with a discussion of the major results.

4.1 Experimental Inputs and Notation

The major factors examined include the number of product types, the number of time buckets, the arrival process parameters, and the service process parameters. For a complete discussion of the factors and the experimental parameter settings, see Rossetti and Clark[12]. Throughout the following discussion, we will use the following notation to represent the various factors and parameter settings used within the experiments.

Notation for Experimental Inputs

R total number of replications

K number of products

N number of time buckets

π_i probability that an arriving product is of type i

Let $M_j = [\pi_1, \pi_2, \ldots, \pi_K]$ be the simulated product mix number j

d Euclidian distance of randomly generated product mix from simulated mix.

θ_i mean of the operation times for product type i

δ_i^2 variance of the operation times for product type i

CV_i, coefficient of variation of the operation times for product type i, where $\text{CV}_i = \delta_i/\theta_i$

We refer to the collection of products which can arrive as the product mix. The probability that an arrival is of type i is the proportion of the total number of products which arrive that are of type i. More specifically, we will refer to the product mix as the set of probabilities, $\pi_i, i = 1, \ldots, K$. An important problem in estimating the capacity loadings for a production schedule is the fact that the product mix observed during the estimation of the mean operation times may not necessarily be the same as the product mix in the production schedule.

4.2 Description of a Typical Experiment

An experiment consists of the simulation of the parallel work center at the specified factor levels, the estimation of the product operation times based on the data obtained from the simulation, and the estimation of work center capacity requirement loadings for 23 randomly generated production schedules based on 23 randomly generated product mixes. The size of each time bucket was fixed at a value of 480 minutes. Each experiment was simulated for a total of $(N + 20) \times 480$ minutes, where the first 20 time buckets (9600 minutes) were discarded as a warm up period. Each experiment was replicated $R = 50$ times yielding 50 estimates of the mean operation time for each product. Twenty three production schedules were randomly generated with different proportions of products in their product mixes as compared to the product mix used to simulate the work center. The linear model was evaluated at the 50 estimates of the mean operation times to yield 50 estimates of work center capacity requirement loadings for each of the 23 randomly generated production schedules. A production schedule refers to an amount of each product to be produced over a fixed time horizon. The major characteristic of a production schedule which can effect the prediction of capacity work load requirements is the proportion of products of each type to be produced, i.e. the product mix.

4.2.1 Production Schedule Generation from Product Mixes

As noted in Section 3, the predicted capacity work load requirement based on estimated mean operation times is just an application of the linear model given in Equation 1 to predict Y

given values for the X's. A production schedule is a specification of the X's for a given product mix. The purpose of generating production schedules based on different product mixes is to evaluate the performance of the estimation process when the product mix in the production schedule is different than the product mix used to develop estimates of the mean operation times.

The production schedules were generated according to the algorithm given in Section 7.1 of the Appendix and described as follows. The aggregate arrival rate for the simulated mix is scaled to a weekly basis. Each individual arrival probability in the simulated mix is multiplied by the weekly aggregate arrival rate to yield a weekly arrival rate for each product. The individual weekly arrival rates for each product are then multiplied by a scale factor to yield an upper limit and divided by the same scale factor to yield a lower limit to be used as parameters of a uniform distribution for that product. Each uniform distribution is sampled from to yield a uniformly distributed arrival rate for the corresponding product. The arrival rates are summed to yield a total rate. The individual rates are then divided by the total rate to yield the proportion of the total which can be attributed to each product. The above process yields a product mix for a production schedule. The product mix generation process was performed for the following scale factors (1.5,2,3,4,5,6,7,8,9,10) yielding 10 randomly generated product mixes. In order to generate the production schedules for each mix, the aggregate arrival rate was multiplied by the individual arrival probabilities for each of the generated mixes yielding an arrival rate for each product. The arrival rates were then used as the parameters for Poisson distributions which generated the amount of product to arrive for that week.

Ten additional product mixes were generated by inverting the previous ten mixes such that if product i had the highest proportion and product j had the lowest proportion in the generated mix then the new generated mix would have product i and j switched (the second highest was switched with the second lowest proportion, third highest with the third lowest, etc.). The actual simulated product mix and its inverse were also used along with a product mix with all products having equal proportions. Thus, a total of 23 product mixes were generated for each experimental setting. A production schedule was generated for each generated product mix and used to evaluate the linear model's ability to predict the work centers loading.

The method used to generate the product mixes allows product mixes to be generated which are "close" or "far" from the simulated mix as controlled by the scale factor. We measure distance using Euclidean point to point distance,

$$d = \left(\sum_{i=1}^{K} |\pi_i - \hat{\pi}_i|^2 \right)^{\frac{1}{2}} \tag{4}$$

where π_i is the product mix probability for product type i in the simulated mix and $\hat{\pi}_i$ is a generated product mix probability for product type i. Because π_i and $\hat{\pi}_i$ are probabilities the range of d is $0 \leq d \leq \sqrt{2}$. In later experiments, we will consider those mixes which have $d \leq (0.25)\sqrt{2} \approx 0.35$ as representative of those product mixes which may occur in a slowly changing product mix environment. In the next section, we present the notation used for the experimental outputs.

4.3 Experimental Outputs and Notation

In this section, we present the notation used to represent the statistical properties of predicting the capacity work load requirements for the work center based upon the estimated mean operation times and randomly generated production schedules. We assess the quality of the predicted capacity work load requirements relative to true work load requirements in terms of estimates of bias and variance. The performance of the predicted capacity requirements is also evaluated by estimating the probability that the estimator will be within $\pm\gamma \times 100\%$ of the true value. We will use the following notation and terms to describe the experimental outputs.

Notation for Experimental Outputs

$\hat{\theta}_{ijr}$ an estimate of the mean operation time for product i on replication r of experiment j

η_{jn} the true capacity work load requirement for a production schedule generated from the product mix used to simulate experiment j, where if we let Z_{in} be the planned amount of product i to be produced for production schedule n, we have

$$\eta_{jn} = \sum_{i=1}^{K} Z_{in}\theta_i$$

$\hat{\eta}_{jnr}$ an estimate of the true capacity load requirements for a randomly generated production schedule n on replication r of experiment j, where,

$$\hat{\eta}_{jnr} = \sum_{i=1}^{K} Z_{in}\hat{\theta}_{ijr}$$

$\widehat{\text{PLE}}_{jnr}$ the predicted load error (PLE), i.e. an estimate of the error in estimating the true capacity load requirement, η_{jn}, with $\hat{\eta}_{jnr}$, for production schedule n on experiment j of replication r, where

$$\widehat{\text{PLE}}_{jnr} = \eta_{jn} - \hat{\eta}_{jnr}$$

$\widehat{\text{PLRE}}_{jnr}$ the predicted load relative error (PLRE), i.e. an estimate of the error in estimating the true capacity load requirement, η_{jn}, with $\hat{\eta}_{jnr}$, for production schedule n on experiment j of replication r, where

$$\widehat{\text{PLRE}}_{jnr} = \frac{\eta_{jn} - \hat{\eta}_{jnr}}{\eta_{jn}}$$

$\widehat{\text{Bias}}$ an estimate of the bias of the estimator

$\hat{\sigma}$ an estimate of the standard deviation of the estimator

$\widehat{\text{mse}}$ an estimate of the mean squared error of the estimator

$\hat{p}(\gamma)$ an estimate for the probability that an estimator is within $\pm\gamma \times 100\%$ of the true value

The following section presents the details of the experimental results.

4.4 Summary of Experimental Results and Discussion

In this section, we present a summary of the results from the experiments which give evidence as to the overall accuracy and precision of the predicted capacity load requirement estimators. We will also discuss the trends and important factors identified in the results. For further results, we refer the reader to Rossetti[10].

4.4.1 Discussion of Predicted Capacity Requirements Estimators

For Tables 1 and 2, the sample consists of observations of the individual estimates of the error in estimating the true capacity load requirement for a given randomly generated production schedule, experiment, and replication combination. For example, in Table 2, the sample mean reported for each level of K is

$$\bar{x} = \frac{1}{23 \times 60 \times 50} \sum_{j=1}^{60} \sum_{n=1}^{23} \sum_{r=1}^{50} \widehat{\text{PLRE}}_{jnr}$$

The overall sample standard deviation of 0.0599 reported in Table 2 presents a measure of the variability of the individual observations of the predicted load relative error, PLRE_{jnr}. Tables 1 and 2 illustrate the accuracy of using the estimated mean operation times to predict the load for the randomly generated production schedules. According to Table 1, 86% of the time the predicted load was within 5% of the true load, and 94% of the time the predicted load was within 10% of the true load. The overall mean relative error was only -0.000034 indicating that in relation to the true load the error is extremely small on average. Figure 3 presents a histogram of the relative error in estimating the true capacity load requirement. A bar in the figure represents the percentage of observations of $\widehat{\text{PLRE}}_{jnr}$ which fell into the intervals identified by their midpoints on the horizontal axis. Note that the histogram has the appearance of a normal probability distribution.

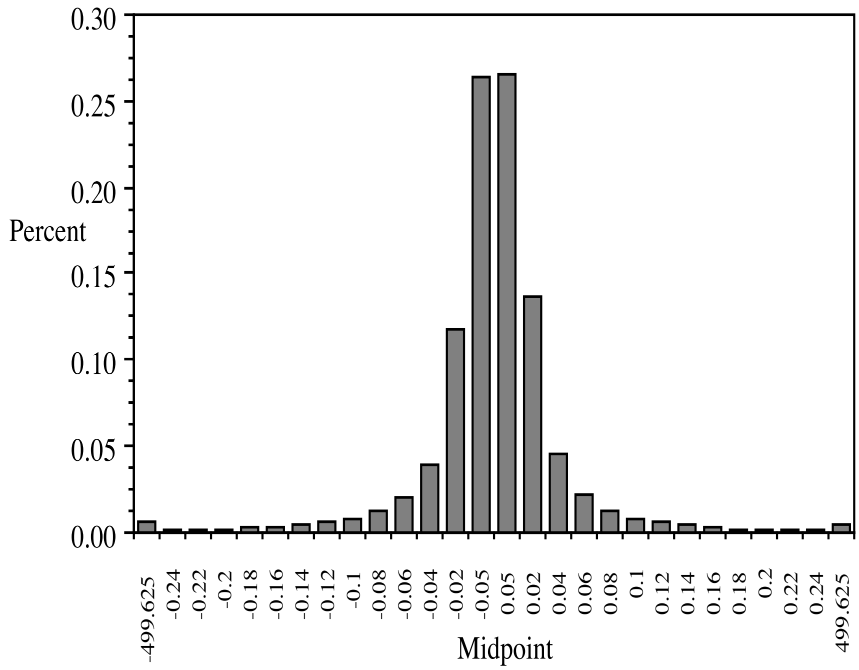

Figure 3: Histogram for Predicted Load Relative Error ($\widehat{\text{PLRE}}_{ijr}$)

Table 1: Predicted Load Error Summary Statistics

	$K = 2$	$K = 4$	$K = 8$	Overall
sample† mean	24.715	65.6606	-4.8761	28.4998
sample std. dev.	508.7245	644.2482	188.7018	487.1574
maximum	5270.12	7116.137	2831.858	7116.137
upper quartile	116.146	174.057	27.041	82.9266
median	-5.6122	15.819	0.0841	1.6563
lower quartile	-108.063	-87.2927	-29.4795	-67.6393
minimum	-7626.91	-6256.63	-2988.80	-7626.91
sample size	69000	69000	69000	207000
$\hat{p}(0.05)$	0.9278	0.8339	0.8327	0.8648
$\hat{p}(0.10)$	0.9786	0.9171	0.9271	0.9409
$\hat{p}(0.15)$	0.9917	0.9526	0.9605	0.9683

†Note: The sample consists of observations of $\widehat{\text{PLE}}_{ijr}$ for 60 experiments $\times 50$ replications $\times 23$ mixes.

Table 2: Predicted Load Relative Error Summary Statistics

	$K = 2$	$K = 4$	$K = 8$	Overall
sample† mean	-0.0019	0.0015	-0.00059	-0.000034
sample std. dev.	0.0339	0.0742	0.0641	0.0599
maximum	0.5008	0.8169	0.6336	0.8169
upper quartile	0.0061	0.0133	0.0107	0.0098
median	-0.00033	0.00134	0.000033	0.00025
lower quartile	-0.0065	-0.0075	-0.0110	-0.00809
minimum	-0.9358	-1.5131	-0.898	-1.5131
sample size	69000	69000	69000	207000

†Note: The sample consists of observations of $\widehat{\text{PLRE}}_{ijr}$ for 60 experiments $\times 50$ replications $\times 23$ mixes.

For Table 3, the sample consists of observations of the statistics computed on the predicted load for the $R = 50$ replications of each experiment for each randomly generated production schedule. The replication statistics illustrate the properties of the replication sample statistics and their distributions. For example in Table 3, the sample mean reported for $\hat{\sigma}$ of 263.1533 is computed as follows. Let J be the total number of experiments, see Table 1 in Rossetti and Clark[12], $J = (\#\ \mathrm{K} \times \#\ \mathrm{CV} \times \#\ \mathrm{Mixes} \times \#\ \mathrm{N}) = (3 \times 3 \times 5 \times 4) = 180$, then the sample mean $\bar{x}$ is,

$$\bar{x} = \frac{1}{23 \times J} \sum_{j=1}^{J} \sum_{n=1}^{23} \hat{\sigma}_{jn}$$

where

$$\hat{\sigma}_{jn} = \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} (\hat{\eta}_{jnr} - \bar{\hat{\eta}}_{jn})^2}$$

where

$$\bar{\hat{\eta}}_{jn} = \frac{1}{R} \sum_{r=1}^{R} \hat{\eta}_{jnr}$$

Thus, the value of 263.1533 for the sample mean of $\hat{\sigma}$ reports the average over the experiments of the sample standard deviation of the estimator based on 50 replications. As such, it presents a measure of the average variability of the estimators over all experimental conditions. From the quartile information of $\hat{p}(0.05)$ for the predicted load estimate, we note that while on average 86% of the time the estimates were within 5% of the true mean operation time, we had 75% of the experiments with a $\hat{p}(0.05)$ greater than 0.82.

For Tables 4 and 5, the sample is the same as in Table 3 except now the statistics are grouped by the factor d which represents how close the product mix used to generate the production schedule is to the simulated mix as defined by Equation 4 of Section 4.2. For Tables 4 and 5, a value of $d \leq 0.35$ can be considered as a relatively small change in the product mix ($d \leq 0.35$ translates into the prediction mix being less than 25% of the maximum possible change in the product mix). The classification of the observations by d indicates that d has a major effect on the performance of the predicted load estimates. In Table 4, with $d \leq 0.35$, 98% of the time the predicted load was within 5% of the true load. In Table 5, with $d > 0.35$, 70% of the time the predicted load was within 5% of the true load (a drop of 28% in performance). Figure 4 presents the performance of $\hat{p}(0.05)$ versus the distance from the simulated mix. Each experiment has a corresponding simulated mix, M_j, see Table 1 and Table 2 in Rossetti and Clark[12] for the actual parameter settings. For every experiment, twenty three production schedules were randomly generated based upon a randomly generated product mix. For each of the 23 randomly generated product mixes, the value of d was computed according to Equation 4. Each point in the figure corresponds to an observation of $\hat{p}(0.05)$ based on the 180 experiments with a total of 4140 data points. Note that the performance of the estimator in terms of $\hat{p}(0.05)$ degrades sharply as d increases past 0.35. The plot was made using the "S" statistical package with the smoothed line added with the *lowess* smoothing command; see reference[1].

Table 3: Predicted Load Replication Summary Statistics

	$\widehat{\text{Bias}}$	$\widehat{\text{PLRE}}$	$\hat{\sigma}$	$\hat{p}(0.05)$	$\hat{p}(0.10)$	$\hat{p}(0.15)$
sample mean	28.4998	-0.00034	263.1533	0.8648	0.9409	0.9682
sample std. dev.	185.8876	0.0121	371.0758	0.2343	0.1522	0.1066
maximum	1305.875	0.0467	2727.595	1.0	1.0	1.0
upper quartile	32.3211	0.0038	317.6255	1.0	1.0	1.0
median	0.2076	0.000048	116.4514	1.0	1.0	1.0
lower quartile	-27.9875	-0.0032	44.2828	0.82	1.0	1.0
minimum	-770.802	-0.0946	2.4636	0.02	0.06	0.10
sample size	4140	4140	4140	4140	4140	4140

Table 4: Predicted Load Replication Summary Statistics $d \leq 0.35$

	$\widehat{\text{Bias}}$	$\widehat{\text{PLRE}}$	$\hat{\sigma}$	$\hat{p}(0.05)$	$\hat{p}(0.10)$	$\hat{p}(0.15)$
sample mean	-3.6362	-0.00026	86.6021	0.9828	0.9977	0.9995
sample std. dev.	55.63227	0.0040	96.4222	0.0644	0.0204	0.0075
maximum	381.1124	0.0208	788.0306	1.0	1.0	1.0
upper quartile	10.6329	0.0014	116.7326	1.0	1.0	1.0
median	-0.1927	-0.00006	57.1291	1.0	1.0	1.0
lower quartile	-18.8588	-0.0018	26.2463	1.0	1.0	1.0
minimum	-332.719	-0.0261	2.4636	0.20	0.42	0.70
sample size	2388	2388	2388	2388	2388	2388

Table 5: Predicted Load Replication Summary Statistics $d > 0.35$

	$\widehat{\text{Bias}}$	$\widehat{\text{PLRE}}$	$\hat{\sigma}$	$\hat{p}(0.05)$	$\hat{p}(0.10)$	$\hat{p}(0.15)$
sample mean	72.3017	-0.00044	499.706	0.7039	0.8635	0.9256
sample std. dev.	272.2713	0.01808	464.4856	0.2816	0.2093	0.1538
maximum	1305.875	0.0467	2727.595	1.0	1.0	1.0
upper quartile	195.8407	0.0110	681.3475	0.98	1.0	1.0
median	4.0758	0.0014	352.7021	0.78	0.98	1.0
lower quartile	-50.7687	-0.0099	166.133	0.47	0.80	0.94
minimum	-770.802	-0.0946	6.1933	0.02	0.06	0.10
sample size	1752	1752	1752	1752	1752	1752

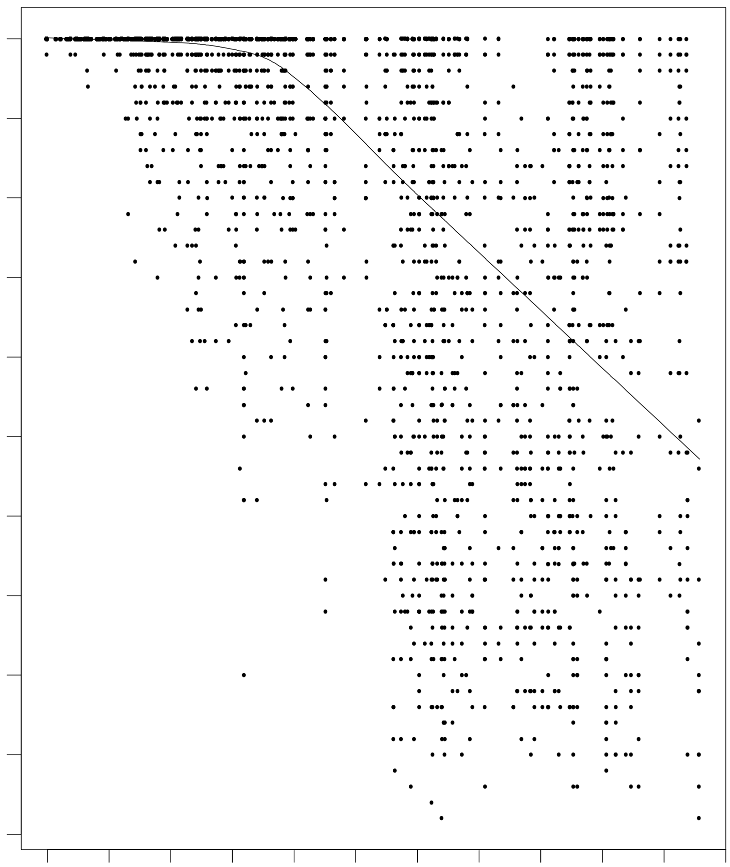

Figure 4: Predicted Load $\hat{p}(0.05)$ vs d

4.4.2 Trends Associated with Experimental Factors

Based on exploratory data plots of the performance of the capacity loading estimates versus the factors in the experiments, we identified the following trends.

1. The data plots did not indicate a problem with the bias of the predicted load estimators with respect to the individual factors K, N, CV and d.

2. Plots of the estimates for the standard deviation of the predicted load estimators, $\hat{\sigma}^{PL}$, versus various factors indicated that

   (a) $\hat{\sigma}^{PL}$ decreases as N increases

   (b) $\hat{\sigma}^{PL}$ increase as d increases

   (c) $\hat{\sigma}^{PL}$ decreases as CV decreases

Based on the exploratory plots and summary statistics, we conclude that the estimator's bias properties are within acceptable limits for use with capacity requirements planning.

5 Planning Model and Example

This section presents the results of our efforts to build a planning model for $\hat{\sigma}^{PL}$ in terms of the important factors within the experiments. The regression model can be used to plan the amount of data necessary to achieve a certain level of estimator performance under various factors.

5.1 Regression Model for $\hat{\sigma}^{PL}$

Based upon the trends discussed in Section 4.4, and based upon some theoretical considerations, we examined the following multiplicative model for $\hat{\sigma}^{PL}$,

$$\hat{\sigma}^{PL} = \omega_0 N^{\omega_1} \mathrm{CV}^{\omega_2} K^{\omega_3} (d+1)^{\omega_4} \epsilon \tag{5}$$

Taking the natural logarithm of Equation (5) transforms the model to

$$Z = \ln \hat{\sigma}^{PL} = \ln \omega_0 + \omega_1 \ln N + \omega_2 \ln \mathrm{CV} + \omega_3 \ln K + \omega_4 \ln(d+1) + \ln \epsilon \tag{6}$$

which is linear in the parameters and can be estimated via standard linear regression procedures. The value of 1.0 was added to d in order to prevent ln(0), and to ensure that $\hat{\sigma}^{PL}$

would not be zero if $d = 0$. Using regression procedures, see reference [13], the parameters were estimated as

$$\hat{\omega}_0 = 4981.57 \quad \hat{\omega}_1 = -0.5238 \quad \hat{\omega}_2 = 0.7164 \quad \hat{\omega}_3 = -0.9566 \quad \hat{\omega}_4 = 4.9383$$

Plots of the residuals did not indicate any strong departures from the standard normality assumptions used in linear regression. The adjusted R^2 value of the regression model was 0.9186 and the parameters all had p-values of 0.0001. We refer the reader to the appendix for the ANOVA results.

5.2 Example for Regression Model for $\hat{\sigma}^{PL}$

There are a variety of ways that the regression model for $\hat{\sigma}^{PL}$ can be utilized for planning purposes. For example, under normality assumptions, we can assert with probability $1 - \alpha$ that if

$$N \geq \left(\frac{E}{z_{\alpha/2}\, \omega_0 \mathrm{CV}^{\omega_2} K^{\omega_3} (d+1)^{\omega_4}} \right)^{1/\omega_1}$$

the error in estimating η_i with $\hat{\eta}_i$ will be less than E. For a specific example, suppose that

$$K = 2 \quad \mathrm{CV} = 0.25$$

and that we would like to be 95% certain that the error is less than 100 minutes. Using the estimated regression coefficients, $\hat{\omega}_0$, $\hat{\omega}_1$, $\hat{\omega}_2$, $\hat{\omega}_3$, $\hat{\omega}_4$, and noting that $d = 0$, we have

$$\begin{aligned} N &\geq \left(\frac{100}{(1.96)(4981.57)(0.25)^{(0.72)}(2)^{(-0.96)}} \right)^{(1/-0.52)} \\ &\geq 270.2 \approx 271 \text{ time buckets} \end{aligned}$$

The regression model can also be used for computing an approximate bound on the error in predicting the load for a different product mix. For example, suppose that we have estimated θ_i under the following conditions

$$N = 300 \quad K = 2 \quad \mathrm{CV} = 0.25 \quad \pi_1 = 0.7 \quad \pi_2 = 0.3$$

and we would like to have an approximate upper bound on the error in predicting a mix with $\pi_1 = 0.4$ and $\pi_2 = 0.6$. If the regression model is true then with probability $1 - \alpha$

$$E \leq z_{\alpha/2}\, \omega_0 N^{\omega_1} \mathrm{CV}^{\omega_2} K^{\omega_3} (d+1)^{\omega_4}$$

For the given data and estimates of $\omega_i, i = 1,2,3,4$ and $d = 0.42$, so that with probability 0.95

$$\begin{aligned} E &\leq (1.96)(4981.57)(300)^{-0.52}(0.25)^{0.72}(2)^{-0.96}(0.42+1)^{4.94} \\ &\leq 542.6 \text{ minutes} \end{aligned}$$

6 Summary and Conclusions

In this paper, we presented and evaluated a methodology which utilizes CIM system data in the form of bar code scanning data to estimate the capacity loadings for a work center with a parallel facility structure. The simulation results indicate that the method can achieve the accuracy and precision necessary for use in CRP. The sensitivity to various experimental factors was explored. We found that the percentage of the products in the production mix can have a significant effect on estimator performance. Also, the estimator's performance can degrade significantly if the observed production mix used to develop the mean operation time estimates is significantly different from the mix in the production schedule for which capacity loadings are desired. Section 5 presented a planning model which can be used to estimate the amount of data required under the various experimental conditions.

As with many information value added applications, there exists a trade off between the amount and quality of information gained and the cost and/or complexity of obtaining the information. The estimation of capacity loadings depends upon the estimation of operation times which can be estimated via automatic data collections systems based on bar code scanning technology. In Rossetti and Clark[12], we developed and evaluated estimators for operation times using both simulated and actual shop floor control data. Thus, traditional methods, such as manual time study, can be reduced. An integrated systems approach to the design of the data collection system can significantly improve planning and control. We note that the possibility exists to estimate other shop floor quantities such as queue times and lead times.

Finally, we note that the entire area of utilizing CIM system data for decision making is an exciting area of research as exemplified by the following quote. During the 1990 Philip McCord Morse Lecture at the ORSA/TIMS Joint National Meeting, Professor John Little addressed the new directions that OR/MS must take in order to face up to the challenge of continuing to contribute to industrial productivity. One of the challenges that Prof. Little identified was:

"... to find models and methods for extracting decision making information out of massive amounts of automatically collected data."[7]

We feel that our research addresses this fundamental challenge.

7 Appendix

7.1 Production Schedule Generation Algorithm

This appendix section presents the algorithm used to generate production schedules based on a product mix for an experiment as described in Section 4.2 of this document.

1. Let $M = [\pi_1, \pi_2, \ldots, \pi_K]$ be the simulated mix
2. Let α_i mean time to the next arrival given the arrival is of type i
3. Let Z_{ij} be the amount of product i in production schedule.j
4. Let $F_n, n = 1, \ldots, 10$ be the scale factor,
   where $\vec{F} = \{1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
5. Let $C = \frac{60 \text{ min}}{1 \text{ hour}} \times \frac{24 \text{ hours}}{1 \text{ day}} \times \frac{5 \text{ days}}{1 \text{ week}}$
6. Compute $\alpha = \sum_{i=1}^{K} \pi_i \alpha_i$
7. Let $\lambda = C/\alpha$
8. For $i = 1$ to K, let $\lambda_i = \pi_i \lambda$
9. For $n = 1$ to 10,
   (a) Let $f = F_n$
   (b) For $i = 1$ to K, let $\hat{\lambda}_i \sim \text{Uniform}(\lambda_i/f, \lambda_i f)$
   (c) Let $\hat{\lambda} = \sum_{i=1}^{K} \lambda_i$
   (d) For $i = 1$ to K,
      i. Let $\hat{\pi}_i = \hat{\lambda}_i / \hat{\lambda}$
      ii. Let $Z_{in} \sim \text{Poisson}(\lambda \hat{\pi}_i)$
   (e) Let $\hat{M}_n = [\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_K]$

10. For each $\hat{M}_n, n = 1, \ldots, 10$, do

    (a) Find the largest and smallest $\hat{\pi}_i$

    (b) Let ℓ be the index of largest

    (c) Let s be the index of the smallest

    (d) Switch order of $\hat{\pi}_\ell$ and $\hat{\pi}_s$ in $\hat{M}_n$

    (e) Repeat (d) for next largest and next smallest until all K are examined

    (f) Let $\hat{M}_n^*$ be the reordered $\hat{M}_n$, with $\hat{M}_n^* = [\hat{\pi}_1^*, \hat{\pi}_2^*, \ldots, \hat{\pi}_K^*]$

11. For each $\hat{M}_n^*, n = 1, \ldots, 10$, do

    (a) Let $j = 11$

    (b) For $i = 1$ to K, let $Z_{ij} \sim \text{Poisson}(\lambda \hat{\pi}_i^*)$

    (c) Add 1 to j

12. Let $j = 21$

13. For $i = 1$ to K, let $Z_{ij} \sim \text{Poisson}(\lambda \pi_i)$

14. Let M^* be the reordered M, with $M^* = [\pi_1^*, \pi_2^*, \ldots, \pi_K^*]$

15. Let $j = 22$

16. For $i = 1$ to K, let $Z_{ij} \sim \text{Poisson}(\lambda \pi_i^*)$

17. For $i = 1$ to K, let $\hat{\pi}_i = 1/K$

18. Let $j = 23$

19. For $i = 1$ to K, let $Z_{ij} \sim \text{Poisson}(\lambda \hat{\pi}_i)$

7.2 ANOVA Results for Planning Models

This appendix section presents typical SAS/STAT statistical results for the regression planning model developed in Section 5.

Table 6: Regression Results for $\hat{\sigma}^{PL}$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob> F
Model	4	6924.16158	1731.04040	11678.941	0.0001
Error	4135	612.88538	0.14822		
C Total	4139	7537.04696			

Root MSE = 0.38499
R-square = 0.9187
Adj R-sq = 0.9186

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H_0 Parameter=0	Prob > $\lvert T \rvert$
$\ln \omega_0$	1	8.513460	0.04619620	184.289	0.0001
ω_1	1	-0.523797	0.00772096	-67.841	0.0001
ω_2	1	0.716405	0.00907750	78.921	0.0001
ω_3	1	-0.956629	0.01061454	-90.124	0.0001
ω_4	1	4.938272	0.02838330	173.985	0.0001

Acknowledgments

The authors would like to thank AT&T Network Systems in Columbus, Ohio for their support of this research especially, Martin Gussman, our project liaison.

8 References

[1] Becker R. A., and Chambers, J. M., 1984. *"S": An interactive environment for data analysis and graphics*, Bell Laboratories, Inc, Murray Hill, New Jersey, Wadsworth Statistics/Probability Series.

[2] Browne, Jimmie, John Harhen and James Shivnan. 1988. *Production Management Systems A CIM Perspective.* Addison-Wesley Publishing Company, Inc.

[3] Duffy E. M., 1992. "Southern Utility Company Energized By Automatic Data Collection System," *Industrial Engineering*, October, Volume 24, No. 10, pg. 29.

[4] Hill, J. M., 1987. *Automatic Identification Systems*, in the *Production Handbook*, John A. White, Editor, Fourth Edition, John Wiley & Sons, Inc.

[5] *IMSL Math/Library Fortran Subroutines for Mathematical Applications*, 2500 ParkWest Tower One, 2500 CityWest Boulevard, Houston, Texas, 77042, version 1.1, 1989.

[6] Law, A.M., and Kelton, W.D., 1991. *Simulation Modeling & Analysis*, Second Edition, McGraw Hill.

[7] Little, John D. C. 1990. "Operations Research in Industry New Opportunities in a Changing World." *The 1990 Philip McCord Morse Lecture*, ORSA/TIMS Joint National Meeting, Philadelphia, PA.

[8] Mabert, V. A., 1992. *Shop Floor Monitoring and Control Systems*, in the *Handbook of Industrial Engineering*, Gavriel Salvendy, Editor, Second Edition, John Wiley & Sons, Inc.

[9] Neter J., Wasserman W., and Kutner, M. H., 1989.*Applied Linear Regression Models*, 2nd Edition, Irwin.

[10] Rossetti, M. D., "Queueing Service Parameter Estimation From Event Oriented Production Control Data," PhD Dissertation, The Ohio State University, Columbus, Ohio, 1992.

[11] Rossetti, M. D., and Clark, G. M., 1995. "Estimating Expected Sojourn Times for an Exiting Markov Renewal Process," *Communications in Statistics - Theory and Methods*, Vol. 24, No. 2, pp. 553-579.

[12] Rossetti, M. D., and Clark, G. M., "Estimating Operation Times From Work Center Arrival and Departure Events," Working Paper SYS-MDR-4-1994, Department of Systems Engineering, University of Virginia, Thornton Hall, Charlottesville, VA 22903.

[13] SAS User's Guide: Statistics. Version 5 ed. Cary, N. C.: SAS Institute, 1985.

[14] Simmons, D. M., 1975. *Non-Linear Programming for Operations Research*, Prentice Hall Inc., Englewood Cliffs, New Jersey, 1975.

[15] Smith, Spencer B. 1989. *Computer Based Production and Inventory Control.* Englewood Cliffs, New Jersey: Prentice Hall Inc.

[16] Vollmann, T. E., Berry, W. L., and Whybark, D. C., 1988. *Manufacturing Planning and Control Systems*, Dow Jones Irwin.