Influencing factors of job waiting time variance on a single machine

Xueping Li*

Department of Industrial and Information Engineering,
University of Tennessee,
408 East Stadium Hall,
Knoxville, TN 37996-0700, USA
Fax:+1-865-974-0588
E-mail: Xueping.Li@utk.edu
*Corresponding author

Nong Ye and Xiaoyun Xu

Department of Industrial Engineering,
Arizona State University,
Tempe, AZ 85287-5906, USA
Email: nongye@asu.edu Email: xiaoyun.xu@asu.edu

Rapinder Sawhey

Department of Industrial and Information Engineering,
University of Tennessee,
Knoxville, TN 37996-0700, USA
Email: sawhney@utk.edu

Abstract: When a batch of jobs are waiting for services from a machine or resource, sometimes it is desirable to minimise the variance of job waiting times Waiting Time Variance (WTV) for service stability to all the jobs in the batch so that the jobs have about the same waiting times. Many factors, including the sum of the jobs' processing times, the probability distribution of job processing times and the scheduling method may influence the variance of job waiting times. In this paper, we use multivariate exploratory techniques such as Principal Components Analysis (PCA) and Correspondence Analysis (CA) along with other statistical analysis techniques to investigate these factors. We prove that the expected WTV can be predicted given characteristics of the jobs. These findings can be applied to achieve a desire level of WTV for service stability.

Keywords: job scheduling; waiting time variance; WTV; quality of service; QoS; statistical analysis; principal components analysis; PCA; correspondence analysis; CA.

Reference to this paper should be made as follows: Li, X., Ye, N., Xu, X. and Sawhey, R. (2007) 'Influencing factors of job waiting time variance on a single machine', *European J. Industrial Engineering*, Vol. 1, No. 1, pp.56–73.

Biographical notes: Xueping Li is the director of the Intelligent Information Engineering and Systems Laboratory (IIESL) and an Assistant Professor in the Department of Industrial and Information Engineering at the University of

Received 3 July 2006, Accepted 21 September 2006

Tennessee-Knoxville. He received an MS in Computer Science and a BS in Automatic Control from Nankai University, China. His research interests include information assurance, web Mining, sensor networks, scheduling and operations research. He is a member of IEEE and INFORMS.

Nong Ye is a Professor of Industrial Engineering at Arizona State University (ASU). She received a PhD in Industrial Engineering from Purdue University, West Lafayette, Indiana and an MS and a BS in Computer Science from the Chinese Academy of Sciences and Peking University, China, respectively. She is the Director of the Information and Systems Assurance Laboratory at ASU. Her research work focuses on security, dependability and Quality of Service (QoS) of computer and network systems. She is an Editor for IEEE Transactions on Systems, Man and Cybernetics and an Associate Editor for IEEE Transactions on Reliability and Information, Knowledge, Systems Management.

Xiaoyun Xu is a PhD student of Industrial Engineering at Arizona State University. He received a BS and an MS in Industrial Engineering. He is now working at Information and Systems Assurance Laboratory at ASU and his research focuses on assuring QoS in Information Systems and scheduling of both discrete and dynamic processes.

Rapinder Sawhey is an Associate Professor in the Department of Industrial and Information Engineering at the University of Tennessee-Knoxville. His research interests include manufacturing planning and control systems, lean production systems and simulation modelling.

1 Introduction

The single machine Waiting Time Variance (WTV) minimisation problem, denoting by $1\|\text{WTV}$, is to minimise the WTV of jobs in a batch on a single machine as follows:

$$\min_{\lambda\in\Pi} \text{WTV} = \frac{1}{n-1}\sum_{i=1}^{n}[W_i(\lambda) - \overline{W}(\lambda)]^2 \tag{1}$$

where Π is the set of all the sequences or permutations of n jobs, $W_i(\lambda)$ is the waiting time of the job on position i of job sequence λ and $\overline{W}(\lambda)$ is the mean waiting time of jobs in sequence λ. The minimisation of WTV is critical for the computer and network resources to provide stable Quality of Service (QoS), particularly stable timeliness, because the minimisation of the second moment of the waiting time leads to the service predictability and dependability (Ye, 2002; Ye et al., 2005b). For example, it is desirable to minimise jitter which is closely related to WTV in streaming real-time applications (Mansour and Patt-Shamir, 1989). The authors in Merten and Muller (1972) argue that it is important to minimise the variance of file retrieval times in computer file organisation systems and prove the antithetic relationship between WTV and Completion Time Variance (CTV). The WTV problem also has applications in many other fields. Close relationships between WTV or CTV minimisation and Mean Squared Deviation (MSD) problems and Just-in-Time concept in production planning are discussed by Panwalker et al. (1982), Bagchi et al. (1987), Cheng and Gupta (1987), Cheng and Kovalyov (1996) and Cheng et al. (2002) that CTV is equivalent to the unconstrained version of the MSD minimisation problem.

WTV minimisation has received considerable attention since it was initiated in Merten and Muller (1972) and there are extensive studies in this area. The V-Shape property of the optimal sequence for WTV is proven in Eilson and Chowdhury (1977), Mittenthal et al. (1995) and Cai (1996). The optimal sequences for WTV problems with up to five jobs are given by Schrage (1975). The conjecture about the positions of the three largest jobs in Schrage (1975) is proven by Hall and Kubiak (1991). Some heuristic methods are developed for CTV problems in Eillon and Chowdhury (1977), Kanet (1981) and Vani and Raghavachari (1987). Most of the above studies mainly focus on developing job scheduling methods for WTV or CTV. In previous study by Ye et al. (2005b), we noticed the characteristics of the jobs, particularly the probability distribution of job processing times and Sum of the Processing Times (SOPT) of the jobs, had impacts on WTV along with the job scheduling methods. It is of interest to figure out these influencing factors and how they impact the WTV. For example, different applications on the internet have different QoS requirements (Chen et al., 2003). In order to reach the level of service stability for certain applications, the administrators of a network domain need to adjust the values of the factors influencing the level of WTV. With the knowledge of the relationship between SOPT of a batch of jobs (data packets) and WTV, the network administrator of a router can set the proper length of the buffer for the incoming data packets (Belenki, 2002; Bianchi et al., 2003; Cheng et al., 1999; Grossglauser and Tse, 1999, 2003; Mao and Habibi, 2002; Peha and Sutivong, 2001; Ye et al., 2005, a,b). In other application fields where service stability and WTV are concerned, the relationship between the influencing factors and WTV can be helpful to design the application system, to predict WTV and to realise a desired level of WTV.

In this study, we investigate the WTV influencing factors, including scheduling methods, the probability distribution of the processing times of the jobs and SOPT and how they influence WTV through statistical analysis techniques such as general regression models, Principal Components Analysis (PCA) and Correspondence Analysis (CA). The results of the analysis will provide a guideline to researchers and practitioners. The rest of this paper is organised as follows. We define the concepts in this study in Section 2. The methodology used to discover the relationships between the factors and WTV is discussed in Section 3. Results and remarks are presented in Section 4. Section 5 concludes this study.

2 Problem definition

Given a batch of jobs to be processed on a single machine, we assume that all the jobs are ready to be released at time zero and the processing times of the jobs are given. The machine can process one job at one time. No preemption is allowed, which means a job cannot be interrupted after it starts its service on the machine. We adopt the same assumptions as (Eilon and Chowdhury, 1977; Merten and Muller, 1972; Schrage, 1975). The notations of the influencing factors of WTV are defined as follows:

n: The number of jobs to be processed on a single machine or resource.

p_i: The processing time of job i, $i = 1, 2, \ldots, n$, which is the size of the job divided by the service rate of the machine.

Φ: SOPT of the jobs in the batch, $\Phi = \sum_{i=1}^{n} p_i$.

S: Scheduling rules or methods, for example, First-In-First-Out (FIFO) or First-Come-First- Serve (FCFS), Shortest Processing Time first (SPT), Verified Spiral (VS) Balanced Spiral (BS) (Ye et al., 2005b), etc.

ψ_s: $\{q_1^{(s)}, q_2^{(s)}, \ldots, q_n^{(s)}\}$, the sequence of the jobs in the batch produced by scheduling method s, where $q_i^{(s)}$ is the processing time of the job at position i, $i = 1, 2, \ldots, n$.

$w_i^{(s)}$: Waiting time of the job at position i of sequence ψ_s under scheduling method S, $w_1^{(s)} = 0$, $w_i^{(s)} = \sum_{j=1}^{i-1} q_j^{(s)}$, $i = 2, 3, \ldots, n$.

$\overline{w^{(s)}}$: Mean waiting time of the jobs in sequence $\psi_{(s)}$ under scheduling method S, $\overline{w^{(s)}} = (\sum_{i=1}^{n} w_i^{(s)})/n$.

D: The probability distribution of job processing times such as the normal distribution, exponential distribution and so on. Furthermore, we denote $Para\_M$ and $Para\_std$ for the mean and standard deviation parameters of the probability distribution D.

$\mathrm{WTV}^{(s)}$: WTV of the jobs if scheduling method S is used.

$$\mathrm{WTV}^{(s)} = \frac{\sum_{i=1}^{n} (w_i^{(s)} - \overline{w^{(s)}})^2}{n-1} \tag{2}$$

Our objective is to determine the relationship between WTV and factors SOPT, S and D:

$$\mathrm{WTV}^{(s)} \propto f(\Phi, S, D) + \epsilon \tag{3}$$

where f is the unknown relationship between WTV and Φ, S and D and ϵ is a term that represents other sources of variability not accounted for in f.

In addition to the above factors, the service rate of the machine has an impact on WTV for a given batch of jobs. The relationship of service rate and WTV is $\mathrm{WTV}_{K_1}/\mathrm{WTV}_{K_2} = K_2^2/K_1^2$ for a given job sequence, where K_1 and K_2 are the service rates and WTV_{K_1} and WTV_{K_2} are the WTV of the jobs under two different service rates of the machine, K_1 and K_2, respectively. The relationship can be shown as follows. We introduce the following notations:

P_{K_1}: The processing time of job i if the service rate of the resource is K_1

P_{K_2}: The processing time of job i if the service rate of the resource is K_2

W_{K_1}: The waiting time of job i if the service rate of the resource is K_1

W_{K_2}: The waiting time of job i if the service rate of the resource is K_2.

It can be easily obtained that P_{K_1}=$P_{K_2} \times K_2/K_1$ and W_{K_1}=$W_{K_2} \times K_2/K_1$. Using (2), we have $\mathrm{WTV}_{K_1} = \mathrm{WTV}_{K_2} \times K_2^2/K_1^2$ for a certain sequence. Thus, increasing the service rate of a resource yields smaller WTV in a non-linear fashion. For instance, doubling the service rate of a machine would make WTV one fourth of the original for the same job sequence. When investigating the factors of SOPT, S and D, we fix the service rate of the machine at the level of one unit.

3 Methodology

3.1 Experimental design settings

It is proven that WTV minimisation problem is NP-hard (Kubiak, 1993). Many researchers have developed scheduling methods to tackle this problem. Six scheduling methods are investigated in this study: FIFO, SPT, VS, BS, EC11 and EC12. EC11 and EC12 refer to the scheduling methods 1.1 and 1.2 at (Eilon and Chowdhury, 1977), respectively. FIFO is commonly used scheduling method on computer and network systems and many other fields. Since this study focuses on a deterministic scheduling problem, we assume the order that the jobs are generated as the order the jobs come into the system such that we can apply FIFO to them. SPT gives the optimal mean waiting time for a batch of jobs. VS, BS, EC11 and EC12 have been shown to reduce WTV effectively (Eilon and Chowdhury, 1977; Ye et al., 2005b).

Four probability distributions are used to generate the processing times of the jobs: the normal distribution, exponential distribution, uniform distribution and Pareto distribution (von Seggern, 1993). The normal distribution is commonly used in many application fields to present the job size, as is the exponential distribution. The uniform distribution lets the sizes of the jobs fall in a range of a lower bound and an upper bound. Recently, Pareto distribution becomes increasingly popular in the internet systems (Arlitt and Williamson, 1996).

To evaluate the possible impacts from each of the above probability distributions, we need to further specify its parameters. Since there are infinite choices for the value of the parameters which we cannot study entirely, we choose four ordinal levels for our study. Specifically, for the normal distribution we select four levels of its mean parameter μ and under each level of μ we select four levels of its standard deviation parameter σ. Similarly, we vary the mean parameter for the exponential distribution, lower bound and upper bound parameters for the uniform distribution and the shape and scale parameters for the Pareto distribution. Note that we deliberately choose the parameters for the four distributions for later comparison study such that they have the same levels of mean parameters. Thus, we have four levels of mean parameters for Para_M and four levels of standard deviation parameters for Para_std in each distribution.

Hence, we have four categorical factors and one continuous factor for investigation. The four categorical factors are the scheduling rule, distribution, Para_M and Para_std. The SOPT Φ is a continuous factor which can be estimated from distribution, Para_M and Para_std factors if the number of jobs in a batch is given. From the design of experiment viewpoint, it is a full factorial design in which scheduling rule factor has six levels and distribution, Para_M and Para_std factors have four levels, respectively. To practically investigate these factors, we generate 384 cases according to the above proposed settings. We implement the six scheduling methods and apply each of them to the jobs to calculate WTV.

3.2 Principal components analysis

As we can see that the influencing factors include categorical and continuous variables and some of them may correlate to and interact with each other. This makes the interpretation of the data and the detection of its structure difficult. PCA transforms original variables into a number of orthogonal variables which could be helpful to reveal the relationship among the factors.

PCA seeks to map a p-dimensional input space $\Re^p$ to some q-dimensional output space $\Re^q$, where $q \ll p$,

$$G(x) : \Re^p \to \Re^q \tag{4}$$

producing a low-dimensional encoding $z = G(x)$ for every input vector x (Cherkassky and Mulier, 1998; Hastie et al., 2001; Jambu, 1991).

An ideal mapping G acts as an encoder of the original distribution. There exists another inverse mapping

$$F(z) : \Re^q \rightarrow \Re^p \tag{5}$$

producing the decoding $x' = F(z)$ of the original input x. Thus an encode-decode mapping process is

$$x' = F(G(x)) \tag{6}$$

Here we use a linear transformation on the input data; that is, $z = G(x)$ in Equation (4) is a linear approximation to the data. Denote the observations by $x_1, x_2, \ldots, x_N$ and consider a linear model to represent them as

$$G(\lambda) = \mu + V_q \lambda \tag{7}$$

where μ is a location vector in $\Re^p$, V_q is a $p \times q$ matrix with q orthogonal unit vectors as columns and λ is a vector of q parameters. Equation (7) is the parametric representation of an affine hyperplane of rank q. We define the loss function to minimise the squared error distortion or reconstruction error as

$$\min_{\lambda_i, \mu, V_q} \sum_{i=1}^{N} \| x_i - \mu - V_q \lambda_i \|^2 \tag{8}$$

where $\|$ denotes the usual L_2 norm. We can partially optimise to obtain

$$\hat{\mu} = \overline{x}$$
$$\hat{\lambda_i} = V_q^T (x_i - \overline{x})$$

Thus we only need to find the orthogonal matrix V_q

$$\min_{V_q} \sum_{i=1}^{N} \| (x_i - \overline{x}) - V_q V_q^T (x_i - \overline{x}) \|^2 \tag{9}$$

The $p \times p$ matrix $H_q = V_q V_q^T$ is a projection matrix which maps a data point x_i onto its rank-q reconstruction $H_q x_i$. The solution can be expressed as construction of a Singular Value Decomposition (SVD) of an $N \times p$ matrix X:

$$X = UDV^T \tag{10}$$

This is a standard decomposition in numerical analysis. Here U is an $N \times p$ orthogonal matrix such that $U^T U = I_p$. Columns u_j are called the left singular vectors. V is a $p \times p$ orthogonal matrix ($V^T V = I_p$) with columns v_j called the right singular vectors. D is a $p \times p$ diagonal matrix with diagonal elements $d_1 \geq d_2 \ldots \geq d_p \geq 0$ which are known as the singular values. The columns of UD are the principal components of X.

Denote the set of q positive eigenvalues of the matrix $X'X$ as $\{\lambda_1, \lambda_2, \ldots, \lambda_q\}$ and corresponding set of q eigenvectors as $\{\mu_1, \mu_2, \ldots, \mu_q\}$. Through locating the coordinates of the variables on the principal components space, the relationship among the variables will be revealed. The coordinate of the jth variable on the αth factor can be computed as:

$$\text{Factor coordinate}_{\alpha j} = (\lambda_\alpha)^{1/2} \mu_{\alpha j}$$

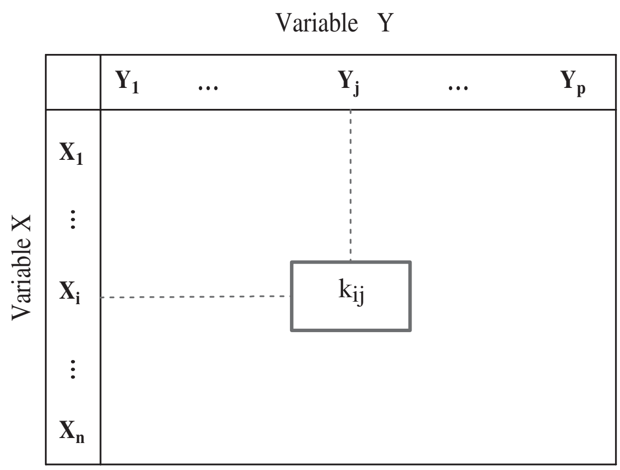
Note that dimensionality reduction is not our goal of PCA. Here we are rather interested in detecting the relationship of the WTV influencing factors. Specifically, we use the influencing factors as the so-called active variables to compute the principal components which are orthogonal to each other. Then, we figure out the factor coordinates of the influencing factors and project WTV, the supplement variable, to the principal components space. Thus, both WTV influencing factors and corresponding WTV are mapped to principal components space with the principal components acting as a 'bridge'. Hence, we are able to find out their relationship by identifying how strongly they associate with the principal components.

3.3 Correspondence analysis

CA is an exploratory technique designed to analyse two-way and multiway crosstabulation tables containing some measure of correspondence between the rows and columns. PCA is a generalisation of Pearson's geometry which is ordinary Euclidean (Pearson, 1901). CA differs from PCA in that it is based on multidimensional weighted Euclidean distance to identify the subspaces of lower dimensionality that best contain the set of points. CA was originally developed by Jean-Paul Benzécri in the early 1960s and 1970s (Benzécri, 1973), but it only gains increasing popularity lately (Carrol et al., 1986; Hoffman and Franke, 1986). A comprehensive description of this method can be found at the classic text by Greenacre (1984). Note that similar techniques were independently developed known as optimal scaling, reciprocal averaging, optimal scoring or homogeneity analysis.

CA is powerful to explore the structure of categorical variables in a table. The most common kind of table of this type is the two-way frequency crosstabulation table. Figure 1 illustrates a model of CA for our study to investigate two categorical factors.

Figure 1 A CA model with contingency data set. For example, X_i can be the scheduling rule factor such as FIFO, SPT, BS, etc., Y_j can be the probability distribution factor such as the normal distribution, exponential distribution and so on. k_{ij} is the corresponding WTV value



Suppose we have a matrix **N** with non-negative numbers. The correspondence matrix **P** is defined as the matrix of elements of **N** divided by the grand total of **N**. Denote **r** and **c** the vectors of row and column sums of **P**, respectively. Denote the diagonal matrices of these sums by $\mathbf{D}_r$ and $\mathbf{D}_c$, respectively. The notations are defined as follows.

Data matrix: $\mathbf{N}(I \times J) = [n_{ij}], n_{ij} \geq 0; i = 1 \ldots I$ and $j = 1 \ldots J$

Correspondence matrix: $\mathbf{P} = (1/n_{..})\mathbf{N}$, where $n_{..} = \mathbf{1^T N 1}$

Row and column sums: $\mathbf{r} = \mathbf{P1}$ and $\mathbf{c} = \mathbf{P}^T 1$, where

Diagonal matrics of the sums: $\mathbf{D}_r = \text{diag}(r)$ and $\mathbf{D}_c = \text{diag}(c)$

Matrices of row and column profiles:

$$\mathbf{R} = \mathbf{D}_r^{-1}\mathbf{P} = \begin{bmatrix} \tilde{r}_1^T \\ \vdots \\ \tilde{r}_I^T \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \mathbf{D}_c^{-1}\mathbf{P}^T = \begin{bmatrix} \tilde{c}_1^T \\ \vdots \\ \tilde{c}_J^T \end{bmatrix}$$

Note that $\mathbf{1} = [1 \ldots 1]^T$ denotes an I-vector of I or J ones. $\mathbf{P}$ can be considered as a probability density on the cells of the $I \times J$ matrix and $\mathbf{r}$ and $\mathbf{c}$ the marginal densities. The row and column profiles $\mathbf{R}$ and $\mathbf{C}$ define two clouds of points in respective J- and I-dimensional weighted Euclidean spaces. The centroids of the row and column clouds are

Row centroid: $\mathbf{c} = \mathbf{R^{T\,r}}$ and

Column centroid: $\mathbf{r} = \mathbf{C^{T\,c}}$.

In the terminology of CA, the row and column sums of the matrix of relative frequencies are called the *row mass* and *column mass*, which are $\mathbf{D}_r$ and $\mathbf{D}_c$, respectively. So, $\mathbf{P}$ shows how one unit of mass is distributed across the cells.

The overall spatial variation of the cloud points is quantified by their total inertia which is a term in CA that is used by analogy with the definition in applied mathematics of 'moment of inertia'. It stands for the integral of mass times the squared distance to the centroid. The total inertia of row points and column points are the weighted sum of squared distances from the points to their respective centroids as follows:

$$\text{Total intertia of row points: in}(I) = \sum_i \left(\mathbf{r_i \tilde{r}_i - 1c}\right)^{\mathbf{T}} \mathbf{D_c^{-1}} \left(\mathbf{\tilde{r}_i - c}\right) \tag{11}$$

$$\text{Total intertia of column points: in}(J) = \sum_{\mathbf{j}} \left(\mathbf{c_j \tilde{c}_j - 1r}\right)^{\mathbf{T}} \mathbf{D_r^{-1}} \left(\mathbf{\tilde{c}_j - r}\right) \tag{12}$$

or, they can be expressed as:

$$\text{in}(I) = \text{trace}\left[\mathbf{D_r}\left(\mathbf{R - 1c^T}\right)\mathbf{D_c^{-1}}\left(\mathbf{R - 1c^T}\right)^{\mathbf{T}}\right]$$

$$\text{in}(J) = \text{trace}\left[\mathbf{D_c}\left(\mathbf{C - 1r^T}\right)\mathbf{D_r^{-1}}\left(\mathbf{C - 1r^T}\right)^{\mathbf{T}}\right]$$

The $\text{in}(I)$ and $\text{in}(J)$ denote the inertia of row and column clouds of points, respectively.

For the two-way table, inertia is defined as the total Pearson Chi-square (χ^2). According to the well-known formula for computing the Chi-square statistic for two-way tables, the expected frequencies or measures in a table, where the columns and rows are independent of each other, are equal to the respective column total times the row total, divided by the grand total. Hence, CA can also be considered as a method to decompose the overall Chi-square statistic in which inertia is defined as $\chi^2/n_{..}$ by identifying a small number of dimensions to represent the deviations from the expected values. If we denote $e_{ij} = n_{i.}n_{.j}/n_{..}$, based on Equations (11) and (12), we have

$$\chi^2 = \sum_i \sum_j \frac{\left(n_{ij} - e_{ij}\right)^2}{e_{ij}}$$

and

$$\text{in}(I) = \text{in}(J) = \sum_i \sum_j \frac{\left(p_{ij} - r_i c_j\right)^2}{r_i c_j} = \chi^2/n \tag{13}$$

that is,

$$= \text{trace}\left[\mathbf{D_r^{-1}}\left(\mathbf{P} - \mathbf{rc^T}\right)\mathrm{D}_\mathbf{c}^{-1}\left(\mathbf{P} - \mathbf{rc^T}\right)^\mathbf{T}\right]$$

Our objective is to identify the K^*-dimensional optimal subspace which minimises the weighted sum of squared distances to the points. There are infinite bases for the optimal subspace. However, we do not have to resort to the use of optimisation techniques to solve the problem since our particular choice of fit in terms of weighted squared distances leads to algebra simplification which is embodied in the concepts of SVD.

Let the generalised SVD of $\mathbf{P} - \mathbf{rc^T}$ be:

$$\mathbf{P} - \mathbf{rc^T} = \mathbf{AD}_\mu\mathbf{B^T} \quad \text{where} \quad \mathbf{A^T D_r^{-1} A} = \mathbf{B^T D_c{-}1B} = \mathbf{I} \tag{14}$$

The singular values are $\mu_1 \geq \ldots \geq \mu_k > 0$. The columns of **A** and **B** will define the principal axes of the column and row clouds, respectively (Greenacre, 1984). More computational details about SVD can be found at (Marshall and Olkin, 1979).

4 Results and discussions

4.1 WTV influencing factor analysis through PCA

We have six factors to conduct PCA: S, D, Para_M, Para_std, Φ and WTV. S, D, Para_M and Para_std are treated as active variables which involve the calculation of the principal components. Φ and WTV are supplement variables which are projected onto the factor space computed from the active variables. Based on the settings in Section 3, we have a 384×6 matrix X for PCA.

Table 1 gives the principal components coordinates of the WTV factors. In the terminology of the factor analysis, the principal components coordinates are also referred to as ‘factor loadings’ which imply the correlations between the variable and the principal component axes. Since the number of principal components is four which is the same as the number of WTV factors, the vector subspace has the same dimensionality as the original vector space. This means we capture the variance in its entirety. It can be seen that S and D are correlated with principal components 1 and 4, Para_M with principal component 2 and Para_std with principal component 3. The supplement variables Φ and WTV are associated with principal components 2 and 3.

Figure 2 shows the plot of factor coordinates of variables which makes the interpretation of the WTV factors easier. The graph shows a unit circle because the analysis is based on correlations. The largest WTV factor versus principal components coordinate is equal to 1.0. Also, the sum of squared factor coordinates for a WTV factor cannot exceed

1.0. Hence, all WTV factor coordinates will fall with the unit circle in Figure 2 which gives a visual indication of how well a variable is represented by the current set of the principal components. Figure 2 indicates that WTV factors Para_M and Para_std are closely associated with principal components 2 and 3. Since WTV is also closely correlated with principal components 2 and 3, Para_M and Para_std correlate with WTV. This leads us to derive Theorem 4.1 in Section 4.3.

Table 1 Factor coordinates of the variables based on correlations. The active variables are scheduling rule, distribution, Para_M and Para_std. Φ and WTV are supplement variables which are marked with a'*'

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*
Scheduling rule	−0.702	0.018	−0.042	0.711
Distribution	0.641	0.157	0.377	0.650
Para_M	−0.238	0.827	0.455	−0.229
Para_Std	−0.201	−0.539	0.806	−0.138
Φ*	−0.251	0.767	0.463	−0.240
WTV*	0.058	0.624	0.403	−0.297

Figure 2 Projection of the variables on the factor-plane. It shows that WTV influencing factors Para_M and Para_std are well represented by principal components 2 and 3, so are the supplement variables WTV and Φ. Supplementary variables are marked with a '*' which are not involved in the derivation of the principal components

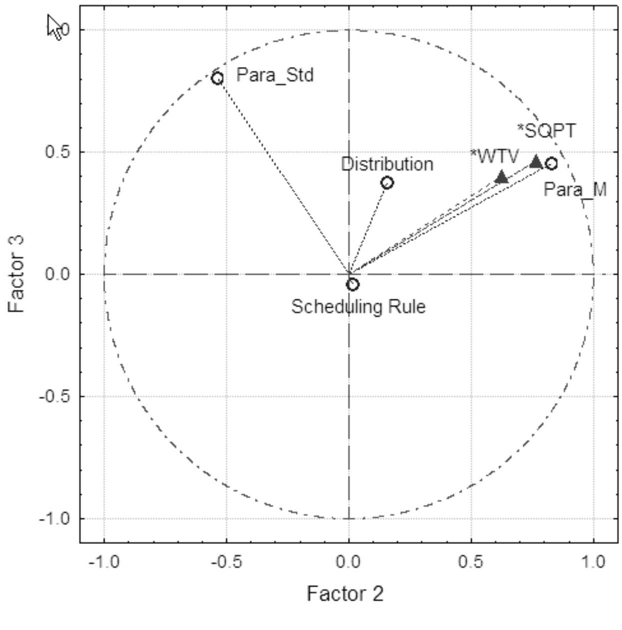
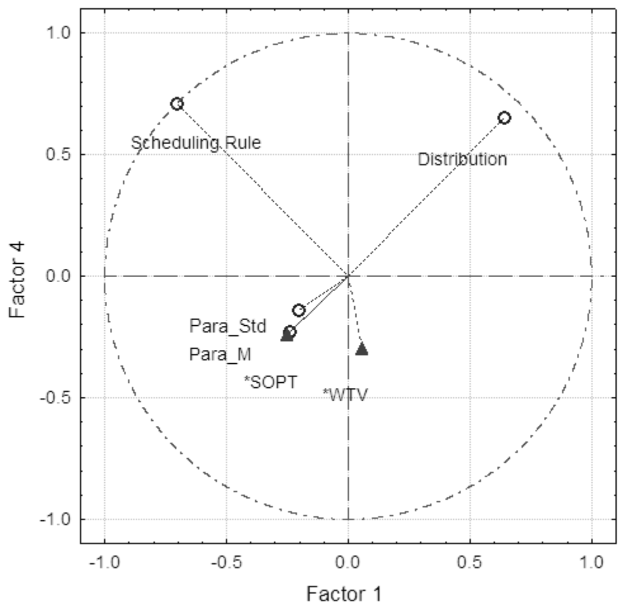


Figure 3 depicts the projection of the WTV influencing factors on the principal components plane. It can be easily observed that *S* and *D* are closely correlated with principal components 1 and 4. The influences of *S* and *D* on WTV will be discussed later using CA.

Figure 3 Projection of the variables on the factor-plane. It shows that WTV influencing factors scheduling rule **S** and distribution **D** are well represented by principal components 1 and 4



4.2 WTV influencing factor analysis through CA

The previous section shows that WTV closely correlates with influencing factors Para_M and Para_std which are well represented by principal components 2 and 3. Influencing factors *S* and *D* are closely associated with principal components 1 and 4. Here we further reveal how *S* and *D* influence with WTV and whether they interact with each other through CA using the method proposed in Section 3.3.

We put the scheduling rule *S* as the row variable the distribution *D* as the column variable and the WTV values under each level of *S* and *D* as the measure of the performance. Thus, we have a 2-dimensional table as the CA data set.

Table 2 gives the row coordinates and contributions to inertia for the row variables. The coordinates of the *S* factor are given on a 2-dimensional space. It can be easily seen that EC11, EC12, BS and VS schduling rules have the similar coordinates which indicates that these scheduling rules have the similar performance with regard to WTV reduction. This confirms the study at Ye et al. (2005b). The mass of each variable is shown in Table 2 as well which is used to compute the matrix of the conditional probabilities. The quality column contains the information concerning how well the quality of the coordinate system in the new vector space represent the respective row points. The quality of each row variable is 1.0 which indicates the new space contains all the information as the original space while conducting the CA.

The relative inertia is also presented in Table 2. It represents the proportion of the contribution of that point to the overall inertia. As we can see our solution not only represents the data points well (high quality) but also offers much contribution to the overall inertia. The relative inertia for each dimension is given in Table 2 indicating a higher proportion of inertia at the first dimension.

Table 2 Row coordinates and contributions to inertia. It shows scheduling rule FIFO differs from SPT and others. SPT slightly differs from EC11, EC12, BS and VS

	Coordin. Dim. 1	*Coordin. Dim. 2*	*Mass*	*Quality*	*Relative inertia*	*Inertia Dim. 1*	*Cosine*^{2} *Dim. 1*	*Inertia Dim. 2*	*Cosine*^{2} *Dim. 2*
FIFO	−0.220	0.012	0.203	1.000	0.516	0.517	0.997	0.280	0.003
SPT	−0.109	−0.019	0.188	1.000	0.120	0.117	0.971	0.695	0.029
EC11	0.107	0.002	0.152	1.000	0.091	0.092	1.000	0.006	0.000
EC12	0.107	0.002	0.152	1.000	0.091	0.092	1.000	0.006	0.000
BS	0.107	0.002	0.152	1.000	0.091	0.092	1.000	0.006	0.000
VS	0.107	0.002	0.152	1.000	0.091	0.092	1.000	0.006	0.000

Table 2 gives the quality or squared correlations with each dimension. It is denoted by Cosine^2, a term that refers to the fact that it is the squared cosine value of the angle the point makes with the respective dimension. The Cosine^2 shows that the row variables are well represented by the first dimension.

Table 3 gives the column coordinates and contributions to inertia of the column variables. We observe a negative coordinate on dimension 1 for the exponential distribution while other distributions have positive coordinates. This indicates that the exponential distribution differs from others concerning the WTV values. We find that the normal and uniform distributions have similar coordinates. The Pareto distribution is slightly away from the normal and uniform distributions on dimension 1.

Table 3 Column coordinates and contributions to inertia. It indicates the differences between the exponential distribution and other probability distributions. The normal and uniform distributions are close to each other

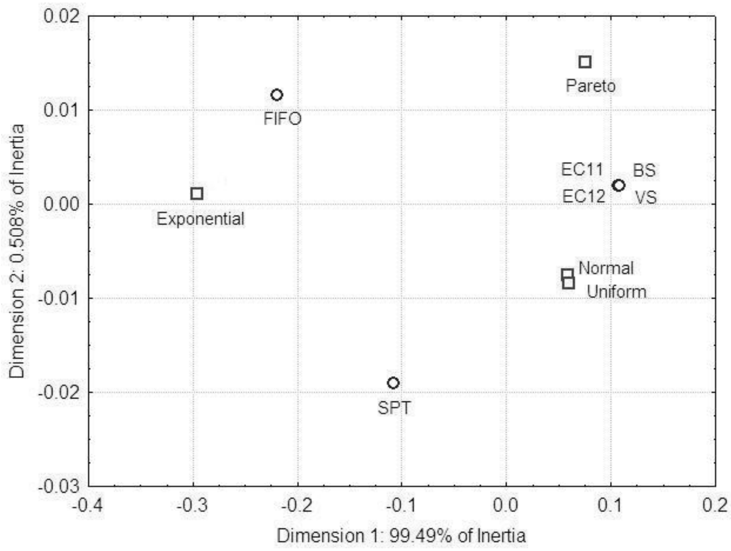
	Coordin. Dim. 1	*Coordin. Dim. 2*	*Mass*	*Quality*	*Relative Inertia*	*Inertia Dim. 1*	*Cosine*^{2} *Dim. 1*	*Inertia Dim. 2*	*Cosine*^{2} *Dim. 2*
Normal	0.058	−0.007	0.276	1.000	0.049	0.049	0.984	0.158	0.016
Exponential	−0.297	0.001	0.178	1.000	0.816	0.820	1.000	0.002	0.000
Uniform	0.059	−0.008	0.272	1.000	0.051	0.050	0.980	0.196	0.020
Pareto	0.075	0.015	0.274	1.000	0.084	0.081	0.961	0.644	0.039

Figure 4 depicts a visual presentation of the row and column coordinates on a 2-dimension plane. The first dimension mainly distinguishes the WTV influencing factors since it accounts about 99.49% of the inertia. It can be easily seen that the scheduling rules EC11, EC12, BS and VS consist of a cluster which indicates their similar impacts on WTV. FIFO and SPT form a solo cluster separately. The normal and uniform distributions make up another cluster indicating that they influence WTV similarly. The Pareto distribution is slightly different from the normal and uniform cluster on dimension 1 while the exponential distribution is far from it.

4.3 Statistical analysis and discussions

Sections 4.1 and 4.2 give the relationship between WTV and its influencing factors through multivariate exploratory techniques. Here we further reveals the relationship to give a closed-form solution through statistical analysis.

Figure 4 Two-dimension plot of row and column coordinates. It classifies the scheduling rules and the distributions



We prove the following theorem that the expected WTV of a batch of jobs is determined by the mean and standard deviation parameter of the distribution of the job processing times given certain conditions.

Theorem 4.1: *Given n jobs whose processing times follow a normal distribution with mean* μ *and variance* σ^2 *, the expected WTV of the jobs is* $n(n+1)/12\mu^2 + (n+1)/6\sigma^2$ *if scheduling rule FIFO is applied.*

Proof: Given a batch of jobs whose processing times follow a normal distribution with mean μ and variance σ^2, assume that the jobs are independent of each other. For a sequence of n jobs, the waiting time of the job at position i is expressed as $W_i = \sum_{j=1}^{i-1} p_j$, $i = 1$ to n and let $p_0 = 0$ and $W_1 = 0$. The expected waiting time, W_i, can be shown as:

$$E(W_i) = (i-1)\mu \geq 0 \qquad i = 1, 2, \ldots, n \tag{15}$$

The expected variance of the waiting time W_i is $\text{Var}(W_i) = (i-1)\sigma^2$, $i = 1, 2, \ldots, n$.

The mean waiting times of the jobs is

$$\overline{W} = \frac{\sum_{i=1}^{n} W_i}{n} = \frac{\sum_{i=1}^{n} (n-i)p_i}{n}$$

Let vectors

$$X' = [n-1, n-2, \ldots, 1]$$

and

$$Y' = \left[p_1, p_2, \ldots, p_{n-1}\right]$$

such that

$$\overline{W} = \frac{X'Y}{n}$$

and

$$\overline{W}^2 = \overline{W}\,\overline{W}' = \frac{1}{n^2}(X'Y)(X'Y)' = \frac{1}{n^2}X'YY'X$$

we obtain:

$$\begin{aligned}
E\left(\overline{W}^2\right) &= \frac{1}{n^2}X'E(YY')X \\
&= \frac{1}{n^2}X'E\begin{bmatrix} p_1^2 & p_1p_2 & \dots & p_1p_{n-1} \\ p_2p_1 & p_2^2 & \dots & p_2p_{n-1} \\ \dots & \dots & \dots & \dots \\ p_{n-1}p_1 & p_{n-1}p_2 & \dots & p_{n-1}^2 \end{bmatrix} X \\
&= \frac{1}{n^2}X'\begin{bmatrix} \mu^2+\sigma^2 & \mu^2 & \dots & \mu^2 \\ \mu^2 & \mu^2+\sigma^2 & \dots & \mu^2 \\ \dots & \dots & \dots & \dots \\ \mu^2 & \mu^2 & \dots & \mu^2+\sigma^2 \end{bmatrix} X \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n-1}\left(i^2\sigma^2\right)\right) + \mu^2 X'\begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \end{bmatrix} X \\
&= \frac{(n-1)^2}{4}\mu^2 + \frac{(n-1)(2n-1)}{6n}\sigma^2
\end{aligned}$$

Since $\mathrm{Var}(W_i) = E(W_i^2) - E^2(W_i)$, we have $E(W_i^2) = (i-1)^2\mu^2 + (i-1)\sigma^2$, $i = 1, 2, \dots, n$.

Hence, $\mathrm{WTV} = 1/(n-1)\left(\sum_{i=1}^{n} W_i^2 - n\overline{W}^2\right)$

$$\begin{aligned}
E(\mathrm{WTV}) &= \frac{1}{n-1}\left[\sum_{i=1}^{n} E\left(W_i^2\right) - E\left(n\left(\overline{W}\right)^2\right)\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}\left[(i-1)^2\mu^2 + (i-1)\sigma^2\right] - nE\left(\overline{W}^2\right)\right]
\end{aligned}$$

Finally, it concludes the proof as

$$E(\mathrm{WTV}) = \frac{n(n+1)}{12}\mu^2 + \frac{n+1}{6}\sigma^2 \tag{16}$$

The relationship between the expected WTV and Φ for a given batch of jobs can be derived as follows.

Corollary 4.1: *Given n jobs whose processing times follow a normal distribution with mean μ and variance σ^2, the expectated WTV the jobs is $(n+1)/12n[E(\Phi)]^2 + (n+1)/6nVar(\Phi)$, where $E(\Phi)$ and $Var(\Phi)$ are the expected mean and variance of Φ, respectively and FIFO is the scheduling rule.*

Proof: The SOPT of the jobs is $\Phi = \sum_{i=1}^{n} p_i$ with $E(\Phi) = n\mu$ and $\text{Var}(\Phi) = n\sigma^2$. Substitute μ with $E(\Phi)/n$ and σ^2 with $\text{Var}(\Phi)/n$ in Equation (16), we get

$$E(\text{WTV}) = \frac{n+1}{12n}[E(\Phi)]^2 + \frac{n+1}{6n}\text{Var}(\Phi)s \tag{17}$$

This concludes the proof.

Theorem 4.1 concludes that the expected WTV of a batch of job can be derived from Para_M and Para_std given that S be FIFO and D be the normal distribution. We can see that larger mean and variance of job processing times yields a larger expected WTV. This theorem is very useful to estimate the WTV value in practice. Furthermore, the equation $\overline{W} \pm 3(\text{WTV})^{1/2}$ can be used to predict the waiting time of a jobs.

Corollary 4.1 predicts the WTV of a batch of jobs if Φ is known. In reality, Φ can be treated as the workload of a batch of jobs. For example, for a batch of parts to be processed on a machine in product system, Φ will be the total processing times of the parts.

Several observations can be made based on Theorem 4.1 and Corollary 4.1.

Firstly, both mean and variance parameters have influence on the expected WTV. The coefficient of the squared mean parameter is $n/2$ times of that of the variance parameter. The mean parameter dominates Equation (16) with the increasing number of jobs. Also, the number of jobs, thus Φ, has impacts on WTV as well. Corollary 4.1 is very helpful for the system administrator to achieve a desired level of WTV. For instance, one could use admission control to allow a certain amount of jobs to get into the system and thus a certain level of WTV can be obtained. Since WTV related to service stability and user satisfaction, this is of great significance in practice such as for production planning and web services.

Secondly, Equation (16) provides an upper bound of expected WTV if any one of EC11, EC12, BS or VS scheduling rules is applied. This is because EC11, EC12, BS or VS produces V-shaped job sequences which reduce WTV than FIFO and SPT scheduling rules.

Thirdly, Equation (16) confirms the findings at the PCA that Para_M and Para_std are closely correlated with WTV.

Finally, Theorem 4.1 and Corollary 4.1 can be extended from the normal distribution to the uniform distribution. This is based on the CA that both distributions belong to one cluster. This is verified through extensive testings.

5 Conclusion

WTV minimisation is closely related to service stability and of great significance in many areas. In this study, we investigate the WTV influencing factors including scheduling rules, the distribution parameters of the job processing times and the SOPT. We use multivariate exploratory techniques including PCA and CA to find out the relationship among the factors and how they impact the WTV performance. We prove a theorem that the expected WTV of a batch of jobs can be predicted by the mean and standard deviation parameters of the probability distribution of the processing times of the jobs for given scheduling rules. A corollary is proven to predict the expected WTV given the SOPT of the jobs when FIFO is applied and the job processing times follow the normal or uniform distribution. These findings provide a guideline to researchers and practitioners to achieve a desired level of WTV through manipulating these factors.

The WTV influencing factors may correlate and interact with each other which makes it difficult to find how they influence WTV. Using PCA and CA, we transform these factors to some orthogonal variables and project them onto the coordinates of the new space.

It shows that the mean and standard deviation parameters of the probability distribution of the jobs strongly contribute to the WTV. This leads us to give a closed-form solution to predict the expected WTV. The CA classifies the influencing factors into clusters. It reveals that the uniform distribution has similar influence on WTV as the normal distribution which extends the application of the theorem. It also confirms that scheduling rules like EC11, EC12, BS and VS are able to reduce WTV more effectively than FIFO and SPT.

In this study we assume the single-machine environment for WTV problems. Future studies can be carried out to investigate the factors influencing WTV in parallel-machine context. We address four probability distributions in this study. There are also jobs whose processing times follow different probability distributions in some applications of WTV problems. It is of interest to identify the distributions and to investigate their influences on WTV. We will address these issues in our future studies.

References

Arlitt, M.F. and Williamson, C.L. (1996) 'Web server workload characterization: the search for invariants', *Measurement and Modeling of Computer Systems. ACM SIGMETRICS'96*, pp.126–137.

Bagchi, U., Sullivan, R.S. and Chang, Y.L. (1987) 'Minimizing mean squared deviation of completion times about a common due date', *Management Science*, Vol. 33, No. 7, pp.894–906.

Belenki, S. (2002) 'An enforced inter-admission delay performance-driven connection admission control algorithm', *ACM SIGCOMM Computer Communication Review*, Vol. 32, No. 2, pp.31–41.

Benzécri, J-P. (1973) *L'Analyse des Données*, Dunod, Paris.

Bianchi, G., Blefari-Melazzi, N., Chan, P.M.L., Holzbock, M., Hu, Y.F., Jahn, A. and Sheriff, R.E. (2003) 'Design and validation of QoS aware mobile internet access procedures for heterogeneous networks', *Mobile Networks and Applications*, Vol. 8, No. 1, pp.11–25.

Cai, X. (1996) 'V-shape property for job sequences that minimize the expected completion time variance', *European Journal of Operational Research*, Vol. 91, No. 1, pp.118–123.

Carrol, J., Green, P. and Schaffer, C. (1986) 'Interpoint distance comparisons in correspondence analysis', *Journal of Marketing Research*, Vol. 23, pp.271–280.

Chen, Y., Farley, T. and Ye, N. (2003) 'QoS requirements of network applications on the internet', *Information, Knowledge, Systems Management*, Vol. 4, No. 1, pp.55–76.

Cheng, R-G., Chang, C-J. and Lin, L-F. (1999) 'A QoS-provisioning neural fuzzy connection admission controller for multimedia high-speed networks', *IEEE/ACM Transactions on Networking*, Vol. 7, No. 1, pp.111–121.

Cheng, T., Chen, Z.L. and Shakhlevich, N.V. (2002) 'Common due date assignment and scheduling with ready times', *Computers and Operations Research*, Vol. 29, No. 14, pp.1957–1967.

Cheng, T. and Gupta, M. (1989) 'Survey of scheduling research involving due-date determination decisions', *European Journal of Operational Research*, Vol. 38, pp.156–166.

Cheng, T. and Kovalyov, M.Y. (1996) 'Batch scheduling and common due-date assignment on a single machine', *Discrete Applied Mathematics*, Vol. 70, pp.231–245.

Cherkassky, V. and Mulier, F. (1998) *Learning from Data: Concepts, Theory, and Methods. Adaptive and Leaning Systems for Signal Processing, Communications, and Control*, John Wiley & Sons.

Eilon, S. and Chowdhury, I.G. (1977) 'Minimizing waiting time variance in the single machine problem', *Management Science*, Vol. 23, No. 6, pp.567–575.

Greenacre, M. (1984) *Theory and Applications of Correspondence Analysis*, Academic Press.

Grossglauser, M. and Tse, D.N.C. (1999) 'A framework for robust measurement-based admission control', *IEEE/ACM Transactions on Networking*, Vol. 7, No. 3, pp.293–309.

Grossglauser, M. and Tse, D.N.C. (2003) 'A time-scale decomposition approach to measurement-based admission control', *IEEE/ACM Transactions on Networking*, Vol. 11, No. 4, pp.550–563.

Hall, N.G. and Kubiak, W. (1991) 'Proof of a conjecture of Schrage about the completion time variance problem', *Operations Research Letters*, Vol. 10, pp.467–472.

Hastie, T., Tibshirani, R. and Friedman, J. (2001) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction Statistics*, Springer.

Hoffman, D. and Franke, G. (1986) 'Correspondence analysis: graphical representation of categorical data in marketing research', *Journal of Marketing Research*, Vol. 13, pp.213–227.

Jambu, M. (1991) *Exploratory and Multivariate Data Analysis*, Academic Press.

Kanet, J.J. (1981) 'Minimizing variation of flow time in single machine systems', *Management Science*, Vol. 27, pp.1453–1459.

Kubiak, W. (1993) 'Completion time variance on a single machine is difficult', *Operations Research Letter*, Vol. 12, pp.49–59.

Mansour, Y. and Patt-Shamir, B. (1998) 'Jitter control in QoS networks', *Thirty Ninth Annual IEEE Symposium on Foundations of Computer Science*, pp.50–59.

Mao, G. and Habibi, D. (2002) 'Loss performance analysis for heterogeneous on-off sources with application to connection admission control', *IEEE/ACM Transactions on Networking*, Vol. 10, No. 1, pp.125–138.

Marshall, A. and Olkin, I. (1979) *Inequalities: Theory of Majorization and its Applications*, New York: Academic Press.

Merten, A. and Muller, M. (1972) 'Variance minimization in single machine sequencing problems', *Management Science*, Vol. 18, No. 9, pp.518–528.

Mittenthal, J., Raghavachari, M. and Rana, A. (1995) 'V- and a-shaped properties for optimal single machine schedules for a class of non-separable penalty functions', *European Journal of Operational Research*, Vol. 86, No. 2, pp.262–269.

Panwalker, S.S., Smith, M.L. and Seidmann, A. (1982) 'Common due date assignment to minimize total penalty for the one machine scheduling problem', *Operations Research*, Vol. 30, No. 2, pp.391–399.

Pearson, K. (1901) 'One lines and planes of closest fit to a system of points in space', *Philosophical Magazine and Journal of Science*, Vol. 6, No. 2, pp.559–572.

Peha, J.M. and Sutivong, A. (2001) 'Admission control algorithms for cellular systems', *Wireless Networks*, Vol. 7, No. 2, pp.117–125.

Schrage, L. (1975) 'Minimizing the time-in-system variance for a finite jobset', *Management Science*, Vol. 21, No. 5, pp.540–543.

Vani, V. and Raghavachari, M. (1987) 'Deterministic and random single machine sequencing with variance minimization', *Operations Reserach*, Vol. 35, No. 1, pp.111–120.

von Seggern, D. (1993) *CRC Standard Curves and Surfaces*, CRC Press, Inc., pp.252–253.

Ye, N. (2002) 'QoS-centric stateful resource management in information systems', *Information Systems Frontiers*, Vol. 4, No. 2, pp.149–160.

Ye, N., Bashettihalli, H., Li, X. and Farley, T. (in press) 'Batch scheduled admission control for service dependability of computer and network resources', *Information, Knowledge, Systems Management*.

Ye, N., Gel, E., Li, X., Farley, T. and Lai, Y. (2005a) 'Web QoS models: applying scheduling rules from production planning', *Computers and Operations Research*, Vol. 32, pp.1147–1164.

Ye, N., Li, X., Farley, T. and Xu, X. (2005b) 'Job scheduling methods for reducing waiting time variance', *Computers and Operations Research*.