Hybrid principal component analysis technique to machine-part grouping problem in cellular manufacturing system

Tamal Ghosh*

Department of Industrial Engineering,
West Bengal University of Technology,
BF 142, Sector-1, Salt Lake City, Kolkata 700064, India
E-mail: tamal.31@gmail.com
*Corresponding author

Manojit Chattopadhyay

Department of Computer Applications,
Pailan College of Management and Technology,
Sector-1, Bengal Pailan Park, Joka, Kolkata 700104, India
Fax: +91-33-2334-1030
E-mail: chattomanojit@yahoo.com

Pranab K. Dan

Department of Industrial Engineering,
West Bengal University of Technology,
BF 142, Sector-1, Salt Lake City, Kolkata 700064, India
E-mail: danpk.wbut@gmail.com

Abstract: This article portrays a hybrid principal component analysis (PCA)-based technique to construct production cells in cellular manufacturing system (CMS). The key problem in CMS is to recognise the machine cells and corresponding part families and subsequently the formation of production cells. A novel approach is considered in this study to systematise a hybrid multivariate clustering technique based on covariance analysis to form the machine cells in CMS. The intended technique is demonstrated in three segments. Firstly, a similarity matrix is developed by exploiting the covariance analysis procedure. In the second stage, the PCA is utilised to identify the potential clusters in CMS with the assistance of eigenvalue and eigenvector computation. In the last stage, an adjustment heuristic is adopted to improve the solution quality and consequently the clustering efficiency. This article states that, the addition of the adjustment heuristic approach into a traditional multivariate PCA-based clustering technique not only enhances the solution quality significantly, but also downgrades the inconsistency of the solutions achieved. The hybrid technique is tested on 24 test datasets available in published articles and it is shown to outperform other published methodologies by enhancing the solution quality on the test problems.

Keywords: cell formation; cellular manufacturing; covariance analysis; principal component analysis; PCA; adjustment heuristic.

Reference to this paper should be made as follows: Ghosh, T., Chattopadhyay, M. and Dan, P.K. (xxxx) 'Hybrid principal component analysis technique to machine-part grouping problem in cellular manufacturing system', *Int. J. Advanced Operations Management*, Vol. X, No. Y, pp.000–000.

Biographical notes: Tamal Ghosh is a PhD Research Scholar in the Department of Production Engineering, Jadavpur University, India. He has also obtained his Master of Technology (MTech) in Industrial Engineering from West Bengal University of Technology and Bachelor of Technology (BTech) in Computer Science and Engineering from National Institute of Technology Calicut, India. His current research area is cellular manufacturing system design.

Manojit Chattopadhyay is an Assistant Professor in the Department of Computer Application, Pailan College of Management and Technology India. He is a PhD scholar of Calcutta University. He has obtained his MPhil in Operations, Technology and Management from West Bengal University of Technology. He has also obtained an MBA from Visva Bharati University in Systems Management and MSc from Darjeeling Government College, India. His research and professional experience is in the area of information technology application in cellular manufacturing.

Pranab K. Dan is a Reader and the Head in the Department of Industrial Engineering and Management, West Bengal University of Technology, India. He has obtained his PhD in Production Engineering in 1996 from Jadavpur University, India. He has obtained his Bachelor of Engineering and Master of Engineering in Mechanical Engineering from Bengal Engineering College, Shibpore affiliated to Calcutta University, India in the year 1980 and 1982, respectively and His research and professional experience is in the area of industrial engineering.

1 Introduction

Conventionally in cellular manufacturing systems (CMS), group technology (GT) could be stated as a manufacturing metaphysics which recognises similar parts, therefore grouping them into part families depending on its manufacturing designs, features and geometric shapes which was first introduced by Burbidge (1963). GT is applied in CMS to develop an alternative of traditional manufacturing system. Designing manufacturing cell has been named cell formation problem (CF/CFP), consists of the following procedures: usually similar parts are grouped into part families following their processing requirements, and heterogeneous machines are grouped into manufacturing cells and subsequently part families are designated to cells. The problem encountered in CMS is construction of such cells irrespective of its type (Selim et al., 1998). Not essentially the aforementioned steps are carried out in the above order or even sequentially. Depending upon the procedures involved in CFP three methods of achieving solutions are proposed (Papaioannou and Wilson, 2010):

1 recognising part families first and consequently machines are clustered into cells depending on the processing requirement of part families

2 recognising manufacturing cells by grouping heterogeneous machines and then the part families are allocated to cells

3 part families and machine cells are developed concurrently.

Due to the NP-complete nature of the problem (Unler and Gungor, 2009), many computational techniques are heavily practised for improved solution to the CFP, a thorough discussion can be found in literature (Ghosh et al., 2010).

In present research, a new approach has been developed by exploiting principal component analysis (PCA) hybridised with an adjustment heuristic algorithm as clustering technique, which is further discussed in Section 3. The grouping efficacy measure is utilised as the performance evaluation benchmark which is elaborated in Section 4, and finally to verify and establish the effectiveness of the proposed method computational results and discussion are presented in Section 5.

2 Literature survey

Various techniques are developed to solve manufacturing cell formation problems since the last 40 years, these include similarity coefficient methods, clustering analysis, array-based techniques, graph partitioning methods, etc. The similarity coefficient approach was first suggested by McAuley (1972). The basis of similarity coefficient methods is to calculate the similarity between each pair of machines and then to group the machines into cells based on their similarity measurements. Few studies have proposed to measure dissimilarity coefficients instead of similarity coefficient for machine-part grouping problems. Dissimilarity coefficients were used by Prabhakaran et al. (2002) for generalised cell formation problems by considering the operation sequences and production volumes of parts. Most similarity coefficient methods utilised machine-part mapping chart. Few of them are single linkage clustering algorithm (McAuley, 1972), average linkage clustering algorithm (Seifoddini and Wolfe, 1986).

Clustering methods are categorised as hierarchical and non-hierarchical methods. Standard or typically designed clustering techniques could be utilised to build clusters of either components or machines. Among these, Carrie (1973), Chan and Milner (1982), Chandrasekharan and Rajagopalan (1986a, 1986b), Chu and Tsai (1990), King (1980), King and Nakornchai (1982), Kusiak (1985, 1987), McAuley (1972), McCornick et al. (1972), Mosier and Taube (1985), Seifoddini (1989), Seifoddini and Wolfe (1986), Shafer and Rogers (1993), Srinivasan and Narendran (1991), Stanfel (1985), and Waghodekar and Sahu (1984) are recognised in the literature as most prevalent approaches. Machine-part grouping problem is based on production flow analysis, in which the machine-part production cells are formed by permuting rows and columns of the machine-part mapping chart in the form of a {0-1} incidence matrix. Some of the methods are rank order clustering by King (1980), bond energy algorithm by McCornick et al. (1972), etc. Dimopoulos and Mort (2001) have proposed a hierarchical algorithm combined with genetic programming for cell formation problem.

Array-based methods consider the rows and columns of the machine-part incidence matrix as binary patterns and reconfigure them to obtain a block diagonal cluster formation. The rank order clustering algorithm is the most familiar array-based technique for cell formation (King, 1980). Substantial alterations and enhancements over rank order clustering algorithm have been described by King and Nakornchai (1982), and

Chandrasekharan and Rajagopalan (1986a). The direct clustering analysis (DCA) has been stated by Chan and Milner (1982), and bond energy analysis is performed by McCornick et al. (1972).

Graph theoretic approach depicts the machines as vertices and the similarity between machines as the weights on the arcs. Rajagopalan and Batra (1975) proposed the use of graph theory to form machine cells. Chandrasekharan and Rajagopalan (1986a) proposed an ideal seed non-hierarchical clustering algorithm for cellular manufacturing. Srinivasan (1994) implemented a method using minimum spanning tree (MST) for the machine-part cell formation problem. A polynomial-time algorithm based on a graph theoretic approach was developed by Veeramani and Mani (1996), named as vertex-tree graphic matrices.

The PCA is one of the first-born techniques in multivariate statistical analysis (Preisendorfer, 1988). Pearson (1901) first familiarised it in the perspective to reorganise linear regression analysis into a new dimension. Subsequently, it was developed by Hotelling (1933) in the psychometry domain and it was named as 'Hotelling transform'. In the numerous real world problems, the PCA is repeatedly practiced to analyse the data with some innate complications (Tuncer et al., 2008; Horenko et al., 2006; Rothenberger et al., 2003). Application of PCA has been utilised to signify the data using trivial count of variables (Wall et al., 2003).

PCA has been adopted in cell formation recently (Hachicha et al., 2006; Chattopadhyay et al., 2010) and demonstrated its ability to form efficient manufacturing cells. Although PCA itself is principally a statistical analysis method which is not essentially originated to form machine-part cells. Therefore, the novelty of this research lies in inclusion of an adjustment heuristic technique which is primarily inspired from Zolfaghari and Liang (2003) and modified considerably. The competence of this hybrid method is significant in terms of solution quality and computational complexity when exploited as a cell formation technique.

3 Solution approach

In this study, covariance analysis have been utilised as a similarity coefficient method to analyse the similarity and dissimilarity between parts and machines and deployed to generate the similarity matrices. Further, PCA is used as to group the machines into machine cells and parts into part families. In the last stage, the adjustment heuristic is stated as a method to find the optimal cluster and improved assignment of part families to the machine cells in order to improve the grouping efficacy measure.

3.1 Standardisation of data and generating covariance matrix

The first step consists of the development of similarity matrix based on covariance method. The initial machine-part incidence matrix $A = [a_{ij}]_{m\times p}$ is shown in Figure 1 which is a {0-1} incidence matrix where rows represent machines and columns represent parts. In the incidence matrix $a_{ij} = 1$ if machine i processes part j and $a_{ij} = 0$ otherwise. The example of Figure 1 is obtained from Waghodekar and Sahu (1984). Cell formation problem could be reflected as a dimension reduction problem. A substantial number of correlated machines are grouped into subsets of production cells and a substantial number of correlated parts are clustered into families. To make the initial matrix (A) adequately

expressive and important, its standardisation is needed to minimise the mean square errors of approximating the data (Miranda et al., 2008). The rows standardised matrix $B = [b_{ij}]_{m\times p}$ is computed using:

$$X_i^s = \frac{X_i - \bar{X}}{\sigma_i} \tag{1}$$

where X_i is the original row vector obtained from incidence matrix $A = [a_{ij}]_{m\times p}$ $X_i = [a_{1i}, a_{2i}, \ldots, a_{pi}]$.

$\bar{X}$ can be calculated from the mean of X_i:

$$\bar{X} = \frac{\sum_{k=1}^{p} a_{ik}}{p} \tag{2}$$

σ_i is the standard deviation:

$$\sqrt{\frac{1}{p}\sum_{k=1}^{p}\left(a_{ik} - \bar{X}\right)^2} \tag{3}$$

The covariance matrix C could be obtained using:

$$C = \frac{1}{p}\sum B.B^T = \frac{1}{p}\sum_{k=1}^{p} b_{ik}b_{jk} \tag{4}$$

where $c_{ii} = 1$.

Figure 1 The input matrix of 5 × 7 size

	p1	*p2*	*p3*	*p4*	*p5*	*p6*	*p7*
m1	1				1	1	1
m2		1	1	1	1		
m3			1	1	1	1	
m4	1	1	1	1			
m5		1		1	1	1	

Figure 2 Similarity matrix of parts obtained using covariance analysis

	p1	*p2*	*p3*	*p4*	*p5*	*p6*	*p7*
p1	1						
p2	–0.167	1					
p3	–0.167	0.167	1				
p4	–0.612	0.612	0.612	1			
p5	–0.612	–0.408	–0.408	–0.25	1		
p6	–0.167	–0.667	–0.667	–0.408	0.612	1	
p7	–0.612	–0.612	–0.612	–1	0.25	0.408	1

Figure 2 and Figure 3 demonstrate the similarity matrices obtained for parts and machines using the covariance method stated above. These are the triangular matrices and utilised to accomplish the PCA, which is elaborated in next subsection.

Figure 3 Similarity matrix of parts obtained using covariance analysis

	m1	*m2*	*m3*	*m4*	*m5*
m1	1				
m2	–0.75	1			
m3	–0.167	0.417	1		
m4	–0.75	0.417	–0.167	1	
m5	–0.167	0.417	0.417	–0.167	1

3.2 Principal component analysis

Principal component method pursues to maximise the sum of squared loadings of each factor obtained. The principal component factor can describe more variance than any other loadings obtained from any other techniques of factorising. The obtained PCs satisfy two clauses:

PC1 components are orthogonal

PC2 the first component has the maximum variance, the second component has the second maximum variance and so on.

The PCA from the standpoint of statistical pattern identification has the rational importance as it is a competent method for dimension reduction (Haykin, 2008). The other benefit of practicing PCA is that, as soon as these patterns in the data are obtained, the data can be compacted by elimination of the number of dimensions without significant damage of information. The principal components analysis, first obtains the set of orthogonal eigenvectors of the covariance matrix of the variables. The matrix of principal components is the product of the eigenvector matrix with the matrix of independent variables. The first principal component describes the major fraction of the total data variance. The second principal component describes the next major fraction of the total data variance, and so on. The aim is to elucidate the largest volume of variance with the least number of components.

In the PCA approach, the very next step computes the eigenvalues from C and diagonal matrix D is obtained using:

$$V^{-1}CV = D \tag{5}$$

V is a matrix formed of eigenvectors which utilises the diagonal form of the covariance matrix C. From the abovementioned rules, the number of principal components can be determined. In the eigen analysis of the covariance matrix, the eigenvalues are equal to the variances of the principal components. The principal components with higher eigenvalues will be traced in the analysis. The principal components are obtained using the aforementioned technique. The components grounded on the amount of justified variance would be the decisive factor of the number of principal components.

Figure 4 Eigen analysis of covariance matrix of parts

	Component 1	*Component 2*	*Component 3*	*Component 4*
Eigenvalues	3.686	2.022	0.833	0.458
Percentage	52.663	28.889	11.905	6.543
Cum. percentage	52.663	81.552	93.457	100
	Component 5	*Component 6*		*Component 7*
Eigenvalues	0	0		0
Percentage	0	0		0
Cum. percentage	100	100		100

Figure 5 Eigen analysis of covariance matrix of machines

	Component 1	*Component 2*	*Component 3*	*Component 4*	*Component 5*
Eigenvalues	2.431	1.588	0.583	0.271	0.126
Percentage	48.618	31.762	11.677	5.428	2.526
Cumulative percentage	48.618	80.38	92.046	97.474	100

Figure 6 PCA component loading chart obtained for parts

	Component 1	*Component 2*	*Component 3*	*Component 4*	*Component 5*	*Component 6*	*Component 7*
p1	–0.184	–0.652	0	0.18	0.35	–0.129	0.608
p2	0.387	–0.085	–0.707	–0.194	0.507	0.122	–0.182
p3	0.387	–0.085	0.707	–0.196	0.507	0.122	–0.182
p4	0.478	0.238	0	0.311	–0.155	0.571	0.518
p5	–0.252	0.561	–0.001	–0.532	0.286	–0.106	0.496
p6	–0.383	0.364	0.001	0.644	0.507	0.122	–0.182
p7	–0.478	–0.238	0	–0.311	–0.027	0.776	–0.127

Figure 7 PCA component loading chart obtained for machines

	Component 1	*Component 2*	*Component 3*	*Component 4*	*Component 5*
m1	–0.586	0.238	0	–0.008	0.775
m2	0.584	0.109	–0.001	–0.697	0.401
m3	0.268	0.563	–0.707	0.332	0.033
m4	0.415	–0.545	0	0.542	0.486
m5	0.268	0.563	0.707	0.331	0.033

Once the PCA method is adopted for the machine-part clustering problem, it could be explained visually with the help of Figure 4 to Figure 9. The variance in each variable is determined in terms of eigenvalue. For the problem matrix under consideration, the eigenvalues are presented in Figure 4 and Figure 5 for the parts and machines respectively. Two principal components are adequate to analyse correlation between machines and parts. There should be high association (positive loading) among machines/parts strongly related with the same cell, and low association (negative loading)

among machines/parts that are related with different cells. Therefore, to illustrate the above issues component 1 and component 2 are opted to analyse the grouping of machine-part. The first principal component's scores are computed from the original data using the loading values from Figure 6 and Figure 7.

Figure 8 Scattered plot obtained for parts (see online version for colours)

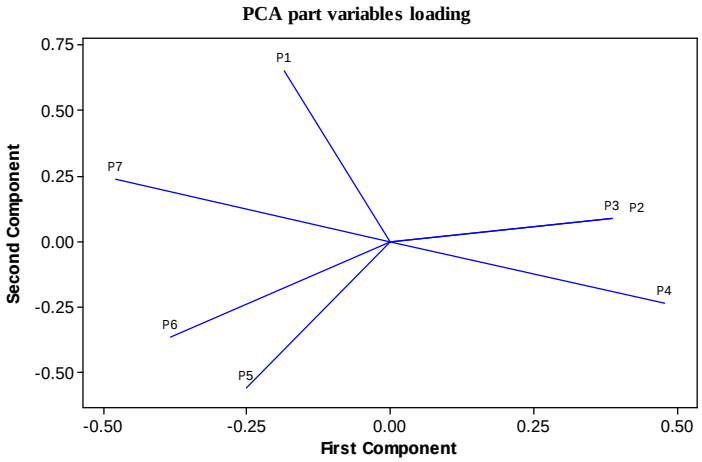
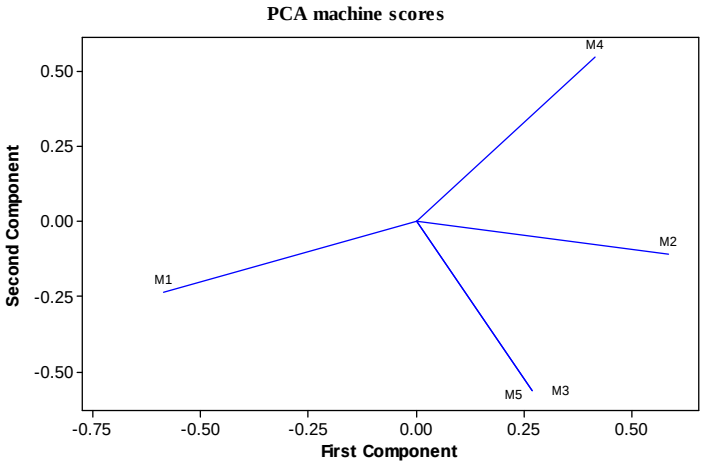


Figure 9 Scattered plot obtained for machines (see online version for colours)



From the parts chart of Figure 6 observation can be made that p1, p5, p6, p7 are having large negative loading on component 1 and therefore clustered in one group and p2, p3, p4 are having large positive loading on component 1 and therefore those are clustered in group 2. In the similar manner from the machines chart of Figure 7, m1 is clustered in group 1 and m2, m3, m4 and m5 are clustered in group 2. In Figure 8 and Figure 9, the obtained scattered plots clearly show the visible groups/clusters/cells in each occasion.

Therefore grouping the rows (machines) and columns (parts) of the original data matrix, two visible groups or cells are obtained block diagonally as shown in Figure 10.

Figure 10 Initial block diagonal cell formation of 5 × 7 matrix

	p1	*p5*	*p6*	*p7*	*p2*	*p3*	*p4*
m1	1	1	1	1	0	0	0
m2	0	1	1	0	0	1	1
m3	0	1	0	0	1	1	1
m4	1	0	0	0	1	1	1
m5	0	1	1	0	1	0	1

3.3 Adjustment heuristic method

The PCA technique adopted in this study is principally a statistical analysis method which is not fundamentally designed to form machine-part cells. Therefore, this technique is unacquainted of the processing requirements of parts through the machines. In order to incorporate this phenomena, an adjustment heuristic technique is proposed in this study which is primarily inspired from Zolfaghari and Liang (2003) and modified substantially. This method is helpful to evaluate the cell formation obtained via PCA. This phenomenon is based on identifying a machine cell which processes the part for a maximum number of operations than any other machine cell and assigning the corresponding part into that cell. Therefore, parts are assigned to the cells which further form tangible part families using membership index given as,

$$D_{cj} = \frac{m_{cj}}{k_c} \times \frac{m_{cj}}{n_j} \times \frac{1}{v} \tag{6}$$

D_{cj} membership index of part j to cell c

m_{cj} number of machines in cell c which process part j

k_c total number of machines in cell c

n_j total number of machines required by part j

v total number of zeros (voids) in the cells in obtained block diagonal matrix.

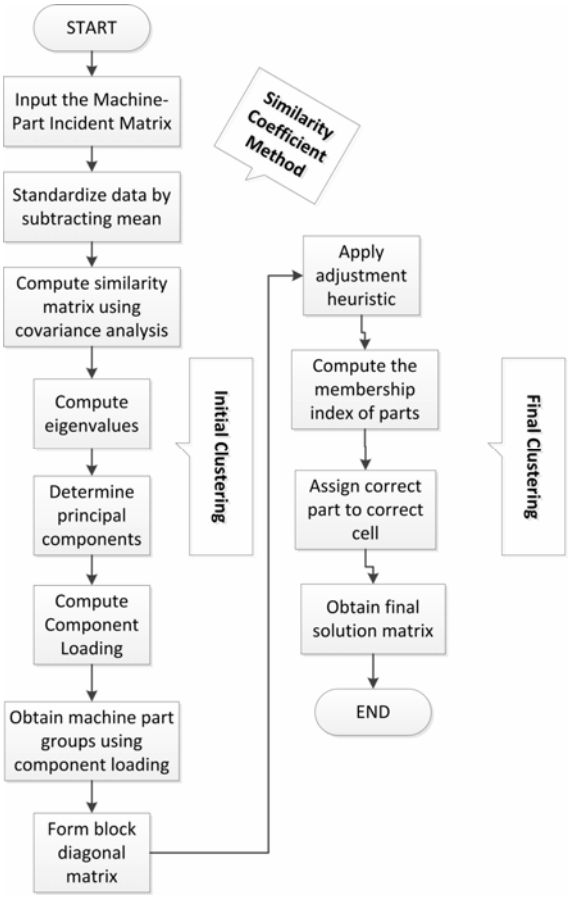
Table 1 Computing the membership index values for parts

Parts	*Membership index values*	
	Cell 1	*Cell 2*
P1	*0.167*	0.021
P2	0.167	*0.375*
P3	0	*0.375*
P4	0	*0.5*
P5	0.125	*0.187*
P6	*0.11*	0.066
P7	*0.5*	0

Figure 11 Final block diagonal cell formation of 5×7 matrix

	p1	*p6*	*p7*	*p2*	*p3*	*p4*	*p5*
m1	1	1	1	0	0	0	1
m2	0	0	0	1	1	1	1
m3	0	1	0	0	1	1	1
m4	0	1	0	1	0	1	1
m5	1	0	0	1	1	1	0

Figure 12 Flowchart of the hybrid PCA technique



In the above mathematical formula, the count of voids has been introduced, which implies the number of zeros in the cells. Using (6) the membership index value of each part can be computed. Larger the membership index value of a part for a particular cell, will subsequently assign it to that cell to obtain the final part families. The computed membership index values for Figure 10 are depicted in Table 1.

The above analysis illustrates the final cell formation, p1, p6, p7 are grouped into part family 1 and p2, p3, p4, p5 are grouped into part family 2 and they are assigned to cell 1 and cell 2, respectively. Therefore, the final block diagonal structure is obtained and presented in Figure 11. The flow chart of the proposed technique is presented in Figure 12.

4 Performance measure

To measure the goodness of solutions, different performance measures have been proposed by researchers since past few decades. Various measures can be obtained from the critical survey of performance measures (Sarker and Mondal, 1999). In this study, grouping efficacy (Kumar and Chandrasekharan, 1990) has been considered which is heavily adopted by researchers to measure the efficiency of their solutions and it is specified as:

$$\tau = \frac{E - E_e}{E + E_v} \tag{7}$$

where

E total number of 1s in matrix A

E_e total number of exceptional elements (EE) (1s outside the cluster block)

E_v total number of voids (0s inside the cluster block).

5 Computational results

The hybrid PCA model is tested with a set of 24 problems that have been published in the past literature. All the datasets were transcribed from the original articles to avoid the inconsistency in data. Out of 24 test datasets two groups are formed. First, 16 test datasets (Table 2), referred as datasets 1–16, are experimented using hybrid PCA method and the obtained results are compared with the published results to demonstrate the effectiveness of the proposed technique. Published results are obtained from Unler and Gungor (2009), and James et al. (2007). The comparison is exhibited in Table 3. The size of first 16 datasets ranges from 5 × 7 to 30 × 50. For the remaining eight problems (referred as datasets 17–24), since there is no comparison exists in literature using grouping efficacy measure, these computed results of experiments are presented in Table 4 and Table 5 in the Appendix A. Proposed hybrid PCA method is simulated with multivariate statistical analysis toolbox and MATLAB 7.0 and tested on a Intel®Core™2Duo 2.1 GHz processor and 2 GB of RAM. For the problems solved with hybrid PCA technique to obtain optimal solution, the grouping efficacy value is better or equal in all instances. All the solutions are obtained with negligible computational time (< 10 sec.). This observation designates that this hybrid technique is substantially competent and less complex because of its minimalism in simulation. The hybrid PCA method is shown to outperform the standard techniques in seven instances, and equal in nine instances, which further illustrates 43.75% enhanced result than published result which is noteworthy in

terms of solution quality and time and space complexities [all 16 solution matrices are presented in Tables 6(a) to 6(d) in Appendix B].

Table 2 Experimental datasets for the result comparison

#	*Dataset*	*Size*
1	Waghodekar and Sahu (1984)	5 × 7
2	King and Nakornchai (1982)	5 × 7
3	Seifoddini (1989)	5 × 18
4	Kusiak and Cho (1992)	6 × 8
5	Boctor (1991)	7 × 11
6	Kusiak and Chow (1987)	7 × 11
7	Car and Mikac (2006)	8 × 10
8	Seifoddini and Wolfe (1986)	8 × 12
9	Chandrasekharan and Rajagopalan (1986a)	8 × 20
10	Chu and Hayya (1991)	9 × 9
11	Mosier and Taube (1985)	10 × 10
12	Chan and Milner (1982)	10 × 15
13	Goncalves and Resende (2004)	15 × 12
14	Jayakrishnan and Narendran (1998)	20 × 8
15	Masnata and Settineri (1997)	25 × 10
16	Stanfel (1985)	30 × 50

Table 3 Comparison of the proposed technique with published results

#	*Grouping efficacy value*		*Cell*	*EE*	*Void*	*Improvement from previous result*
	Best result found in literature	*Proposed hybrid PCA*				
1	62.5	*69.56*	2	6	2	*12.10%*
2	73.68	73.68	2	2	3	0.00%
3	79.59	79.59	2	7	3	0.00%
4	76.92	76.92	2	2	4	0.00%
5	70.37	70.37	3	2	6	0.00%
6	53.13	*59.26*	4	7	4	*11.54%*
7	66.67	*68.96*	4	7	2	*3.43%*
8	68.3	68.3	3	7	6	*0.00%*
9	85.24	*86.67*	3	8	0	*1.68%*
10	73.53	*74.28*	3	6	3	*1.02%*
11	76.47	76.47	3	0	8	0.00%
12	92	92	3	0	3	0.00%
13	86.67	86.67	4	0	6	0.00%
14	83.87*	82.25	3	9	1	0.00%
15	63.93	*70.27*	4	9	13	*9.91%*
16	59.43	*60.12*	11	25	42	*1.16%*

Note: *Inconsistent result shown in Unler and Gungor (2009), the actual computed value is 82.25.

6 Conclusions

This empirical article proposes a hybrid PCA-based clustering technique that combines an adjustment heuristic approach to cell formation problem in cellular manufacturing.

Experimental results presented in Section 5 demonstrate that the proposed hybrid technique outperforms the other techniques, and delivers improved results in comparison with the published results. This article states that, the inclusion of the adjustment heuristic approach into a traditional multivariate PCA-based clustering technique can improve the solution quality substantially, but it also reduces the variability of the solutions obtained. The proposed method obtains better quality solutions by consuming lesser computational time and resources than that of the traditional complex methodologies. It is also shown that the proposed technique performs at least as well as, and often better than the available algorithms for the cell formation on all problems tested. Therefore, it is verified as a promising method in aforestated area. Further work can be done by utilising this technique in more large-scale and realistic and complex cell formation problem which deals with ratio data of production volume, operational time, worker assignment by considering multi-objective factors. Future work can also be done by combining soft computing techniques in the proposed model, such as fuzzy mathematics, particle swarm optimisation or artificial bees colony techniques, which would be the extension of the present research.

References

Boctor, F.F. (1991) 'A linear formulation of the machine-part cell formation problem', *International Journal of Production Research*, Vol. 29, No. 2, pp.343–356.

Burbidge, J.L. (1963) 'Production flow analysis', *Production Engineer*, Vol. 42, No. 12, pp.742–752.

Car, Z. and Mikac, T. (2006) 'Evolutionary approach for cell formation problem in cell manufacturing', *Advanced Engineering Informatics*, Vol. 20, No. 3, pp.227–232.

Carrie, A.S. (1973) 'Numerical taxonomy applied to group technology and plant layout', *International Journal of Production Research*, Vol. 11, No. 4, pp.399–416.

Chan, H.M. and Milner, D.A. (1982) 'Direct clustering algorithm for group formation in cellular manufacturing', *Journal of Manufacturing Systems*, Vol. 1, No. 1, pp.65–75.

Chandrasekharan, M.P. and Rajagopalan, R. (1986a) 'An ideal seed non-hierarchical clustering algorithm for cellular manufacturing', *International Journal of Production Research*, Vol. 24, No. 2, pp.451–464.

Chandrasekharan, M.P. and Rajagopalan, R. (1986b) 'MODROC: an extension of rank order clustering for group technology', *International Journal of Production Research*, Vol. 24, No. 5, pp.1221–1233.

Chattopadhyay, M., Chattopadhyay, S. and Dan, P.K. (2012) 'Application of principal component analysis in machine-part cell formation', *Management Science Letters*, Vol. 2, No. 4, pp.1175–1188.

Chu, A.H. and Hayya, J.C. (1991) 'A fuzzy clustering approach to manufacturing cell formation', *International Journal of Production Research*, Vol. 29, No. 7, pp.1475–1487.

Chu, C.H. and Tsai, M. (1990) 'A comparison of three array-based clustering techniques for manufacturing cell formation', *International Journal of Production Research*, Vol. 28, No. 8, pp.1417–1433.

Dimopoulos, C. and Mort, N. (2001) 'A hierarchical clustering methodology based on genetic programming for the solution of simple cell-formation problems', *International Journal of Production Research*, Vol. 39, No. 1, pp.1–19.

Ghosh, T., Sengupta, S., Chattopadhyay, M. and Dan, P.K. (2010) 'Meta-heuristics in cellular manufacturing: a state-of-the-art review', *International Journal of Industrial Engineering Computations*, Vol. 2, No. 1, pp.123–140, DOI: 10.5267/j.ijiec.2010.03.005, (in press).

Goncalves, J. and Resende, M. (2004) 'An evolutionary algorithm for manufacturing cell formation', *Computers & Industrial Engineering*, Vol. 47, Nos. 2–3, pp.247–273.

Hachicha, W., Masmoudi, F. and Haddar, M. (2006) 'Principal component analysis model for machine-part cell formation problem in group technology', Paper presented at *The International Conference on Advances in Mechanical Engineering and Mechanics, ICAMEM 2006*, Tunisia.

Haykin, S. (2008) *Neural Networks a Comprehensive Foundation*, 2nd ed., Pearson Education, India, Delhi.

Horenko, I., Dittmer, E. and Schütte, C. (2006) 'Reduced stochastic models for complex molecular systems', *Computing and Visualization in Science*, Vol. 9, No. 2, pp.89–102.

Hotelling, H. (1933) 'Analysis of a complex of statistical variables into principal components', *The Journal of Educational Psychology*, Vol. 24, No. 7, pp.498–520.

James, T.L., Brown, E.C. and Keeling, K.B. (2007) 'A hybrid grouping genetic algorithm for the cell formation problem', *Computers & Operations Research*, Vol. 34, No. 7, pp.2059–2079.

Jayakrishnan, G.N. and Narendran, T.T. (1998) 'CASE: a clustering algorithm for cell formation with sequence data', *International Journal of Production Research*, Vol. 36, No. 1, pp.157–179.

King, J.R. (1980) 'Machine-component grouping in production flow analysis: an approach using a rank order-clustering algorithm', *International Journal of Production Research*, Vol. 18, No. 2, pp.213–232.

King, J.R. and Nakornchai, V. (1982) 'Machine-component group formation in group technology: review and extension', *International Journal of Production Research*, Vol. 20, No. 2, pp.117–133.

Kumar, S.C. and Chandrasekharan, M.P. (1990) 'Grouping efficacy: a quantitative criterion for goodness of block diagonal forms of binary matrices in group technology', *International Journal of Production Research*, Vol. 28, No. 2, pp.233–243.

Kusiak, A. (1985) 'The part families problem in flexible manufacturing systems', *Annals of Operations Research*, Vol. 3, No. 6, pp.279–300.

Kusiak, A. (1987) 'The generalized group technology concept', *International Journal of Production Research*, Vol. 25, No. 4, pp.561–569.

Kusiak, A. and Chow, W.S. (1987) 'Efficient solving of the group technology problem', *Journal of Manufacturing Systems*, Vol. 6, No. 2, pp.117–124.

Kusiak. A. and Cho, M. (1992) 'Similarity coefficient algorithm for solving the group technology problem', *International Journal of Production Research*, Vol. 30, No. 11, pp.2633–2646.

Masnata, A. and Settineri, L. (1997) 'An application of fuzzy clustering to cellular manufacturing', *International Journal of Production Research*, Vol. 35, No. 4, pp.1077–1094.

McAuley, J. (1972) 'Machine grouping for efficient production', *Production Engineer*, Vol. 51, No. 2, pp.53–57.

McCornick, W.T., Schweitzer, J.P.J. and White, T.W. (1972) 'Problem decomposition and data reorganization by a clustering technique', *Operations Research*, Vol. 20, No. 5, pp.993–1009.

Miranda, A.A., Borgne, Y.A.L. and Bontempi, G. (2008) 'New routes from minimal approximation error to principal components', *Neural Processing Letters*, Vol. 27, No. 3, pp.197–207.

Mosier, C.T. and Taube, L. (1985) 'Weighted similarity measure heuristics for the group technology machine clustering problem', *OMEGA*, Vol. 13, No. 6, pp.577–583.

Nair, G.J. and Narendran, T.T. (1998) 'CASE: a clustering algorithm for cell formation with sequence data', *International Journal of Production Research*, Vol. 36, No. 1, pp.157–179.

Papaioannou, G. and Wilson, J. M. (2010) 'The evolution of cell formation problem methodologies based on recent studies (1997–2008): review and directions for future research', *European Journal of Operational Research*, Vol. 206, No. 3, pp.509–521.

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

Prabhakaran, G., Janakiraman, T.N. and Sachithanandam, M. (2002) 'Manufacturing data based combined dissimilarity coefficient for machine cell formation', *International Journal of Advanced Manufacturing Technology*, Vol. 19, No. 12, pp.889–897.

Preisendorfer, R.W. (1988) *Principal Component Analysis in Meteorology and Oceanography*, Elsevier Science, New York, USA.

Rajagopalan, R. and Batra, J.L. (1975) 'Design of cellular production system – a graph theoretic approach', *International Journal of Production Research*, Vol. 13, No. 6, pp.567–579.

Rothenberger, M.A., Dooley, K.J., Kulkarni, U.R. and Nada, N. (2003) 'Strategies for software reuse: a principal component analysis of reuse practices', *IEEE Transactions on Software Engineering*, Vol. 29, No. 9, pp.825–837.

Sarker, B.R. and Mondal, S. (1999) 'Grouping efficiency measures in cellular manufacturing: a survey and critical review', *International Journal of Production Research*, Vol. 37, No. 2, pp.285–314.

Seifoddini, H. (1989) 'Duplication process in machine cells formation in group technology', *IIE Transactions*, Vol. 21, No. 9, pp.382–388.

Seifoddini, H. and Wolfe, P.M. (1986) 'Application of the similarity coefficient method in group technology', *IIE Transactions*, Vol. 18, No. 3, pp.271–277.

Selim, M.S., Askin, R.G. and Vakharia, A.J. (1998) 'Cell formation in group technology: review evaluation and directions for future research', *Computers & Industrial Engineering*, Vol. 34, No. 1, pp.3–20.

Shafer, S.M. and Rogers, D.F. (1993) 'Similarity and distance measures for cellular manufacturing', *International Journal of Production Research*, Vol. 31, No. 5, pp.1131–1142.

Sofianopoulou, S. (1999) 'Manufacturing cells design with alternative process plans and/or replicate machines', *International Journal of Production Research*, Vol. 37, No. 3, pp.707–720.

Srinivasan G. (1994) 'A clustering algorithm for machine cell formation in group technology using minimum spanning trees', *International Journal of Production Research*, Vol. 32, No. 4, pp.2149–58.

Srinivasan, G. and Narendran, T.T. (1991) 'GRAFICS: a non-hierarchical clustering algorithm for group technology', *International Journal of Production Research*, Vol. 29, No. 3, pp.463–478.

Stanfel, L.E. (1985) 'Machine clustering for economic production', *Engineering Costs and Production Economics*, Vol. 9, No. 2, Part 1, pp.73–81.

Sudhakarapandian, R. (2007) 'Application of soft computing techniques for cell formation considering operational time and sequence', PhD thesis, NIT Rourkela, India.

Tuncer, Y., Tanik, M.M. and Alison, D.B. (2008) 'An overview of statistical decomposition techniques applied to complex systems', *Computational Statistics & Data Analysis*, Vol. 52, No. 5, pp.2292–2310.

Unler, A. and Gungor, Z. (2009) 'Applying K-harmonic means clustering to the part-machine classification problem', *Expert Systems with Applications*, Vol. 36, Nos. 1–3, pp.1179–1194.

Veeramani, D. and Mani, K. (1996) 'A polynomial-time algorithm for optimal clustering in a special class of {0, 1}-matrices', *International Journal of Production Research*, Vol. 34, No. 9, pp.2587–2611.

Venugopal, V. and Narendran, T.T. (1992) 'Cell formation in manufacturing systems through simulated annealing', *European Journal of Operations Research*, Vol. 63, No. 3, pp.409–422.

Waghodekar, P.H. and Sahu, S. (1984) 'Machine-component cell formation in group technology', *International Journal of Production Research*, Vol. 22, No. 6, pp.937–948.

Wall, M.E., Rechtsteiner, A. and Rocha, L.M. (2003) *A Practical Approach to Microarray Data Analysis*, Kluwer, Norwell, MA.

Won, Y. and Lee, K.C. (2001) 'Group technology cell formation considering operation sequence and production volumes', *International Journal of Production Research*, Vol. 39, No. 13, pp.2755–2768.

Zolfaghari, S. and Liang, M. (2003) 'A new genetic algorithm for the machine/part grouping problem involving processing times and lot sizes', *Computers and Industrial Engineering*, Vol. 45, No. 4, pp.713–731.

Appendix A

Test result of the additional eight problem datasets, ranges from 4 × 5 to 9 × 9 are presented in Table 4. Solution matrices are provided in Table 5. Some of the matrices are presented in transposed form due to space constraints.

Table 4 Test result of additional 8 problems from literature

#	*Dataset*	*Size*	*Proposed hybrid PCA*	*Cell*	*EE*	*Void*
			Grouping efficacy value			
17	Sofianopoulou (1999)	4 × 5	80	2	0	2
18	Won and Lee (2001)	5 × 5	78.57	2	1	2
19	Sudhakarapandian (2007)-1	5 × 7	69.56	2	5	2
20	Venugopal and Narendran (1992)-1	5 × 8	100	2	0	0
21	Sudhakarapandian (2007)-2	6 × 8	69.56	3	5	2
22	Nair and Narendran (1998)	7 × 7	85	3	3	0
23	Venugopal and Narendran (1992)-2	7 × 11	59.26	4	7	4
24	Venugopal and Narendran (1992)-3	9 × 9	70.59	3	7	3

Table 5 Solution matrices of 8 additional test datasets

	m1	*m3*	*m2*	*m4*
p2	1	1		
p4	1	1		
p5	1			
p1			1	
p3			1	1

Sofianopoulou (1999) (4 × 5)

	m1	*m3*	*m5*	*m2*	*m4*
p2	1	1			
p3	1	1	1		
p5	1		1	1	
p1				1	1
p4				1	1

Won and Lee (2001) (5 × 5)

	m1	*m2*	*m4*	*m3*	*m5*
p1	1	1	1		
p5	1	1	1		
p3	1		1		1
p7		1	1		1
p2		1		1	1
P4		1		1	1
p6	1			1	1

Sudhakarapandian (2007)-1 (5×7)

	m1	*m2*	*m4*	*m3*	*m5*	*m6*
p4	1		1			
p6	1	1		1		
p7		1				1
p8		1	1			
p1		1		1	1	1
p2				1	1	1
p3				1	1	1
p5				1		1

Sudhakarapandian (2007)-2 (6×8)

	m1	*m2*	*m3*	*m4*	*m5*	*m6*	*m7*
p1	1	1	1				
p2	1	1					
p3			1	1			
p4			1	1			
p5					1	1	1
p6			1		1	1	1
p7				1	1	1	1

Nair and Narendran (1998) (7×7)

	m1	*m4*	*m5*	*m2*	*m3*
p1	1	1	1		
p4	1	1	1		
p7	1	1	1		
p2				1	1
p3				1	1
p5				1	1
p6				1	1
p8				1	1

Venugopal and Narendran (1992)-1 (5×8)

	m2	*m5*	*m3*	*m6*	*m1*	*m4*	*m7*
p5	1	1					
p11	1		1				
p10			1	1			
p1	1			1		1	
p4				1			1
p8		1		1			
p9				1			1
p2					1		
p3					1	1	1
p6						1	1
p7					1		1

Venugopal and Narendran (1992)-2 (7×11)

	m1	*m5*	*m2*	*m6*	*m9*	*m3*	*m4*	*m7*	*m8*
p1	1	1	1						
p4		1					1		1
p5	1	1							1
p2	1		1	1			1		
p6			1	1	1				
p9			1	1	1				
p3						1	1	1	1
p7						1		1	1
p8		1				1	1	1	1

Venugopal and Narendran (1992)-3 (9×9)

Appendix B

Table 6(a) 16 solution matrices

	m1	*m2*	*m3*	*m5*	*m4*
p1	1				1
p7	1				
p6	1		1	1	
p4		1	1	1	1
p3		1	1		1
p5	1	1	1	1	
p2		1		1	1

Waghodekar and Sahu (1984) (5 × 7)

	m1	*m4*	*m2*	*m3*	*m5*
p1	1	1	1		
p3	1	1	1		
p6	1	1	1		
p8	1	1	1		
p11	1	1	1		
p12	1	1	1		
p13	1	1	1		
p2	1	1			
p5	1	1			
p14	1	1			
p16	1	1			
p17	1	1			
p4			1	1	1
p7			1	1	
p10			1	1	1
p15			1	1	1
p18			1	1	1
p9					1

Seifoddini (1989) (5 × 18)

	m1	*m4*	*m2*	*m3*	*m5*
p2	1	1			
p4	1	1			
p5	1				1
p6	1	1		1	
p1			1	1	1
p3			1	1	
p7				1	1

King and Nakornchai (1982) (5 × 7)

	m1	*m4*	*m6*	*m2*	*m3*	*m5*
p4	1	1	1			
p7	1	1	1	1		
p1				1		1
p2	1			1		
p3				1	1	1
p5				1		1
p6				1	1	1
p8				1	1	1

Kusiak and Cho (1992) (6 × 8)

	m3	*m4*	*m5*	*m6*	*m7*	*m1*	*m2*
p11	1		1				
p3	1	1	1				
p7	1	1					
p4			1	1			
p5				1	1		
p8					1		
p10				1	1		
p1	1					1	
P2						1	1
p6						1	1
p9							1

Boctor (1991) (7 × 11)

Table 6(b) 16 solution matrices

	m4	*m5*	*m6*	*m3*	*m7*	*m8*	*m1*	*m2*
p7	1	1	1	1				1
p8	1	1	1	1				
p9	1	1	1	1				
p10	1	1						1
p6	1			1				1
p11			1		1	1		
p12					1	1		
p1							1	1
p2							1	
p3				1			1	1
p4				1			1	1
p5				1				1

Seifoddini and Wolfe (1986) (8 × 12)

	m1	*m6*	*m7*	*m8*	*m5*	*m2*	*m3*	*m4*
p6	1	1	1	1				
p7	1	1						
p5		1	1	1	1			
p2			1	1				
p8			1	1	1			
p10		1		1	1			
p4								1
p3							1	1
p1					1	1	1	1
p9							1	1

Car and Mikac (2006) (8 × 10)

	m6	*m5*	*m2*	*m3*	*m1*	*m4*	*m7*
p8	1	1					
p9	1						1
p4	1						1
p1	1		1			1	
p5		1	1				
p10	1			1			
p11			1	1			
p6						1	1
p7					1		1
p2					1		
p3					1	1	1

Kusiak and Chow (1987) (7×11)

	m1	*m3*	*m7*	*m8*	*m2*	*m4*	*m5*	*m6*
p2	1	1						
p8	1	1						
p9	1	1						1
p11	1	1	1					
p13	1	1						
p14	1	1			1			
p16	1	1						
p17	1	1					1	
p19	1	1						
p20			1	1	1	1		1
p18			1	1	1	1		
p6			1	1	1	1	1	
p7			1	1	1	1		
p3	1		1	1	1	1		
p4			1	1	1	1		
p1							1	1
p5							1	1
p10						1	1	1
p12			1				1	1
p15							1	1

Chandrasekharan and Rajagopalan (1986a) (8 × 20)

	m1	*m5*	*m3*	*m7*	*m8*	*m4*	*m2*	*m6*	*m9*
p1	1	1					1		
p5	1	1			1				
p4		1			1	1			
p3			1	1	1	1			
p8		1	1	1	1	1			
p7			1	1	1				
p6							1	1	1
p2	1					1	1	1	1
p9							1	1	1

Chu and Hayya (1991) (9 × 9)

Table 6(c) 16 solution matrices

	m2	*m7*	*m9*	*m10*	*m1*	*m4*	*m5*	*m6*	*m3*	*m8*
p2			1	1						
p3	1	1	1	1						
p4	1		1	1						
p8	1	1		1						
p1					1	1		1		
p10					1	1	1	1		
p7						1	1	1		
p5									1	
p6									1	1
p9										1

Mosier and Taube (1985) (10 × 10)

	m2	*m5*	*m8*	*m1*	*m7*	*m10*	*m3*	*m4*	*m6*	*m9*
p3	1	1	1							
p5	1	1	1							
p8	1	1	1							
p13	1	1	1							
p15	1	1	1							
p10				1	1	1				
p11				1	1	1				
p12				1	1	1				
p7				1	1	1				
p2				1	1	1				
p14							1	1	1	1
p9							1	1	1	1
p1							1	1	1	
p4								1	1	1
p6							1		1	1

Chan and Milner (1982) (10 × 15)

	p1	*p4*	*p11*	*p2*	*p9*	*p3*	*p6*	*p8*	*p10*	*p5*	*p7*	*p12*
m1	1	1										
m4	1	1	1									
m6	1	1	1									
m12		1	1									
m13	1		1									
m2				1	1							
m8				1	1							
m11				1	1							
m3						1	1	1				

Table 6(c) 16 solution matrices (continued)

	p1	*p4*	*p11*	*p2*	*p9*	*p3*	*p6*	*p8*	*p10*	*p5*	*p7*	*p12*
m5						1	1	1				
m9						1	1	1				
m7						1		1				
m10									1	1		1
m14										1	1	1
m15									1	1	1	1

Goncalves and Resende (2004) (15 × 12)

	p5	*p6*	*p2*	*p4*	*p7*	*p8*	*p1*	*p3*
m1	1	1						
m5	1	1						
m10	1	1		1				
m12	1	1			1			
m15	1			1				
m3			1	1	1	1	1	
m4			1	1	1	1		
m6	1		1	1	1	1		
m7			1	1	1	1		
m18		1	1	1	1	1		
m20							1	1
m2							1	1
m8							1	1
m13							1	1
m16							1	1
m17	1						1	1
m19							1	1
m9		1					1	1
m11					1		1	1
m14			1				1	,

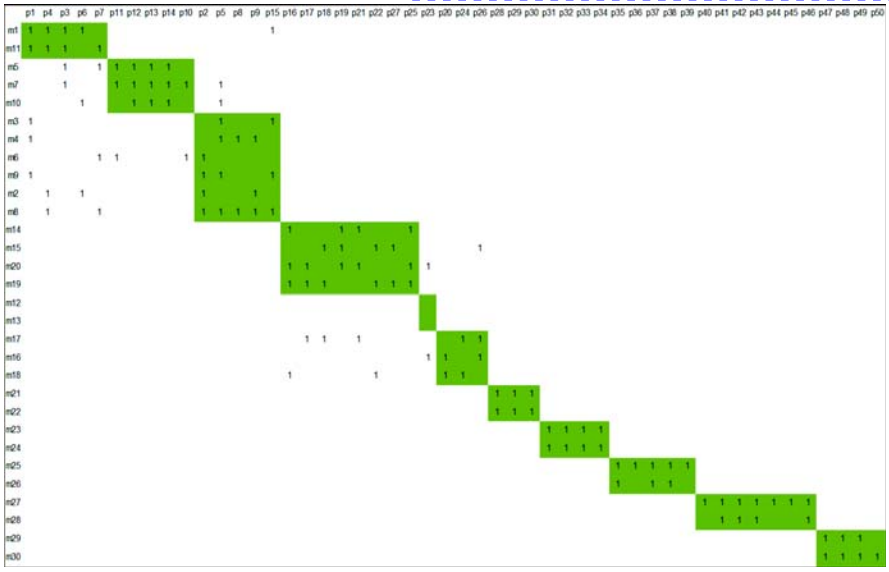
Jayakrishnan and Narendran (1998) (20 × 8)

Table 6(d) 16 solution matrices (see online version for colours)

	m1	*m2*	*m8*	*m13*	*m16*	*m22*	*m24*	*m21*	*m4*	*m6*	*m11*	*m25*	*m3*	*m14*	*m23*	*m9*	*m10*	*m12*	*m7*	*m5*	*m15*	*m17*	*m19*	*m20*	*m18*
p1	1	1	1	1	1	1		1																	
p7	1	1	1	1	1	1	1				1														
p9	1		1			1	1	1																	1
p3							1		1	1	1	1													
p5		1						1	1	1		1													
p4													1	1	1	1	1		1						
p10													1	1	1			1							
p8																1	1		1	1	1	1		1	
p6																			1	1	1	1	1		
p2																		1		1	1	1	1	1	1

Masnata and Settineri (1997) (25 × 10)

Table 6(d) 16 solution matrices (continued) (see online version for colours)



Stanfel (1985) (30 × 50)

Comment [t1]: Author: Please provide an MS Word file editable version of this table so we can format using the Inderscience guidelines.