

ANÁLISE DE PROBLEMAS DE PARTIÇÃO DE INSTALAÇÕES EM SISTEMAS *JOB-SHOPS* POR MEIO DE MODELOS DE REDES DE FILAS

Claudio Rogerio Negri da Silva

Reinaldo Morabito*

Departamento de Engenharia de Produção
Universidade Federal de São Carlos (UFSCar)
São Carlos – SP
morabito@power.ufscar.br

* *Corresponding author* / autor para quem as correspondências devem ser encaminhadas

Recebido em 09/2006; aceito em 03/2007

Received September 2006; accepted March 2007

Resumo

Este artigo estuda o problema de projeto de fábrica focalizada envolvendo a partição da instalação (planta) em subplantas e a alocação de capacidade em cada estação de trabalho das subplantas. O sistema de manufatura *job-shop* é representado por meio de uma rede de filas aberta genérica, e aproximações baseadas em métodos de decomposição são utilizadas para avaliar e otimizar o desempenho do sistema. O objetivo é reduzir a complexidade do sistema do ponto de vista da gestão do produto ou da gestão da estação, por exemplo, limitando-se a variância dos *leadtimes* dos produtos na rede. Apresenta-se um modelo de programação não-linear inteira para o problema e um algoritmo heurístico para resolvê-lo. Aplicando-se o algoritmo em alguns problemas testes, mostra-se que a partição da instalação em subplantas pode reduzir a variância dos *leadtimes* dos produtos na rede, sem necessidade de investimentos adicionais em capacidade. Além disso, algumas vezes é possível manter (ou até melhorar) o desempenho da rede, particionando-a em subplantas que necessitam de menos capacidade do que a configuração original da rede como uma planta única.

Palavras-chave: partição da instalação; fábrica focalizada; projeto de *job-shops*; redes de filas abertas; alocação de capacidade; métodos de avaliação de desempenho.

Abstract

This paper studies the focused factory design involving the partition of the facility (plant) into sub-plants and the allocation of capacity in each workstation of the sub-plants. The job-shop system is represented by a generic open queuing network and approximations based on decomposition methods are used to evaluate and optimize the performance of the system. The aim is to reduce the system complexity from the point of view of the product or workstation management, for example, limiting the variance of the product leadtimes in the network. An integer non-linear programming model and a heuristic algorithm are presented. Applying the algorithm to some testing problems, it is shown that the partition of the facility into sub-plants can reduce the variance of the product leadtimes without requiring additional investments of capacity. Moreover, sometimes it is possible to maintain (or even improve) the network performance by partitioning the plant into sub-plants which need less capacity than the original plant.

Keywords: facility partition, focused factory; job-shop design; open queuing networks; capacity allocation; performance evaluation methods.

1. Introdução

Problemas de projeto e planejamento de sistemas discretos de manufatura podem ser classificados como (e.g. Bitran & Dasu, 1992; Govil & Fu, 1999): (i) problemas de desempenho desejado do sistema (*SP1, Strategic Problem 1*), (ii) problemas de desempenho ótimo do sistema (*SP2, Strategic Problem 2*), (iii) problemas de partição da instalação (*SP3, Strategic Problem 3*). Na classe *SP1*, o objetivo é minimizar o investimento no sistema sujeito às restrições de desempenho desejado para o sistema. Medidas de desempenho típicas são: inventário em processo (*work-in-process – WIP*), *leadtime* médio dos produtos (tempo total que um produto leva para ser produzido, ou seja, a soma dos tempos de espera em fila mais os tempos de processamento em cada estação) e utilização média das estações de trabalho. Na classe *SP2*, deseja-se otimizar o desempenho do sistema sujeito a limitações orçamentárias. Os problemas em *SP1* e *SP2* envolvem um *trade-off* (conflito) entre o capital de investimento e o capital de trabalho (ou o nível de serviço), se o *WIP* (ou o *leadtime*) for a medida de desempenho adotada. Estes problemas já foram objeto de diversos estudos anteriores em sistemas *job-shops*, por exemplo, em Bitran & Tirupati (1989), Boxma *et al.* (1990), Van Vliet & Rinnooy Kan (1991), Bitran & Sarkar (1994a), Bitran & Morabito (1995, 1996, 1999) e Silva & Morabito (2007). Outros estudos relacionados aparecem em Calabrese (1992), Frenk *et al.* (1994), Sundarraj *et al.* (1994), Brettbauer (1996), Seshadri & Pinedo (1999), Shanthikumar & Xu (2000), Souza & Ketzenberg (2002) e Souza *et al.* (2002).

A gestão de um sistema de manufatura em geral torna-se mais complexa à medida que o número de produtos produzidos na instalação aumenta. Portanto, pode ser vantajoso particionar uma instalação em subunidades menores e mais focalizadas. Nos problemas da classe *SP3* busca-se particionar a planta em subplantas (plantas dentro da planta) para reduzir a complexidade do sistema e melhorar seu desempenho global. Isto pode requerer duplicação de equipamentos e há, portanto, um *trade-off* entre o custo de adição de capacidade e a redução de complexidade gerencial do sistema. No presente artigo objetiva-se estudar modelos e métodos de solução para problemas da classe *SP3*, que envolvem o projeto de fábrica focalizada. Poucos trabalhos trataram a aplicação de modelos de redes de filas para problemas de partição de instalações em sistemas de manufatura. Estes objetivos vêm ao encontro das perspectivas para pesquisa futura apontadas em Bitran & Morabito (1996, 1999); veja também Kouvelis *et al.* (2005).

O projeto de fábrica focalizada é aqui abordado do ponto de vista de partição de uma planta *job-shop* em subplantas *job-shops* menores, e alocação de capacidade (máquinas) a cada subplanta. Para desenvolver modelos de otimização que auxiliem gerentes a particionar a instalação, Bitran & Sarkar (1994b) sugerem utilizar medidas quantitativas de complexidade do ponto de vista da gestão do produto e da gestão da estação. Além disso, sugerem analisar sob quais condições é vantajoso particionar uma instalação em subplantas focalizadas, sem necessariamente envolver aumento de capacidade na instalação, podendo, em certos casos, até reduzi-la. Os modelos de otimização apresentados baseiam-se na representação do sistema *job-shop* por uma rede de filas. Modelos de redes de filas podem ser utilizados para auxiliar no processo de tomada de decisões de projeto (ou reprojeção) e planejamento destes sistemas. As redes *job-shops* são formadas de vários sistemas de filas conectados entre si, com usuários deslocando-se entre eles para receber serviço. Os nós da rede representam as estações de trabalho (*shops*); os produtos (*jobs*) correspondem aos usuários requerendo serviço das estações e; os arcos conectando os nós, às rotas dos produtos.

Redes de filas são em geral classificadas como redes de Jackson, aquelas cujos processos de chegadas externas de produtos são Poisson e os tempos de serviço em cada estação têm distribuição exponencial, ou redes genéricas, aquelas cujos processos de chegadas externas não necessariamente são Poisson e/ou os processos de serviço não necessariamente são exponenciais, ou seja, são descritos por distribuições genéricas. As redes podem ser abertas, em que o número de *jobs* circulando na rede pode variar a cada instante, ou fechadas, em que o número de *jobs* na rede é constante, sendo as abertas o foco deste estudo. Métodos exatos de avaliação de desempenho existem somente para redes de Jackson, mas a representação de sistemas de manufatura como redes de Jackson, em geral, superestima a variabilidade do processo de serviço (e.g., Bitran & Tirupati, 1988). Portanto, métodos aproximados são comumente utilizados para estimar o desempenho de redes mais genéricas, particularmente os sistemas *job-shops*.

No presente artigo utilizam-se um modelo de otimização e um método de solução propostos em Bitran & Sarkar (1994b), combinado com um algoritmo de Boxma *et al.* (1990), para mostrar que a partição de uma rede *job-shop* em subredes *job-shops* pode reduzir a variância dos *leadtimes* dos produtos, sem necessidade de investimentos adicionais em capacidade. Em alguns casos também é possível manter (ou até melhorar) o desempenho da rede, particionando-a em subredes que necessitam de menos capacidade do que a configuração original. Este artigo está organizado da seguinte maneira: a seção 2 discute brevemente sistemas *job-shops*, o conceito de fábrica focalizada e representações de sistemas *job-shops* por meio de redes de filas abertas. A seção 3 discute medidas de complexidade do ponto de vista da gestão do produto, em seguida apresenta um modelo de programação não-linear inteira para o projeto de fábrica focalizada e um algoritmo baseado em uma heurística de economias para resolver o modelo. A seção 4 analisa os resultados computacionais obtidos ao aplicar-se o algoritmo em duas redes hipotéticas e uma rede realista conhecida da literatura. Por fim, a seção 5 discute as conclusões e perspectivas de pesquisas futuras.

2. Sistemas *job-shops*, fábrica focalizada e redes de filas

Sistemas de manufatura consistem basicamente de máquinas e estações de trabalho, onde operações são realizadas sobre partes, itens, submontagens e montagens, para produzir produtos que serão distribuídos para clientes. Os componentes de movimentação de materiais e de estocagem permitem que os itens se movam de estação para estação, que partes apropriadas estejam disponíveis para montagem e que o trabalho seja mantido até poder entrar nas estações para processamento. Os sistemas discretos de manufatura são aqueles em que cada item processado é distinto. Por simplicidade, um item, parte, submontagem ou montagem processada por uma máquina ou estação é aqui chamado simplesmente de produto. Sistemas discretos podem ser classificados em função do volume (escala) e variedade (escopo) dos produtos. Tanto a variedade quanto o volume interferem diretamente no projeto e operação do sistema.

O presente estudo tem como foco o *job-shop*, que é uma forma organizacional do sistema de manufatura com capacidade de escopo (ou seja, capacidade de processamento de várias famílias de produtos), mas com limitação em capacidade de escala, devido aos problemas associados ao controle e movimentação dos produtos e ao tempo de preparação (*setup time*) das máquinas. Este sistema também é referido como *layout* por processo (ou *layout* funcional), dado que máquinas desempenhando funções semelhantes são agrupadas em um mesmo departamento ou estação. As máquinas podem ser intercambiáveis, ou seja, uma

mesma máquina pode realizar operações em produtos pertencentes a famílias (classes) diferentes, o que exige um certo tempo de preparação das máquinas. Daí, o sistema ser caracterizado em geral por longos *leadtimes* e altos níveis de *WIP*. Vantagens e desvantagens do emprego deste tipo de sistema de manufatura estão discutidas, por exemplo, em Askin & Standridge (1993), Buzacott & Shanthikumar (1993) e Krajewski & Ritzman (1996).

Segundo Skinner (1974), a abordagem mais prática do conceito de fábrica focalizada é o de divisão, organizacional e física, da instalação existente em plantas dentro de uma planta (*PWP*, do inglês, *plant within a plant*). Cada *PWP* tem suas próprias instalações podendo se concentrar em suas tarefas de produção particulares, usando abordagens de gestão de recursos humanos, controle de produção e estrutura organizacional próprias. Equipamentos e manutenção de materiais são especializados quando necessários. As vantagens são que cada *PWP* ganha experiência facilmente ao focalizar e concentrar todo elemento de seu trabalho naqueles objetivos essenciais e limitados que constituem sua tarefa de manufatura. A abordagem de *PWP* torna mais fácil o realinhamento, com o decorrer do tempo, de operações essenciais e de elementos do sistema com mudanças propostas pela estratégia corporativa e de mercado. A abordagem de fábrica focalizada pode ser aplicada em sistemas *job-shops*.

Conforme Bitran & Dasu (1992), para particionar uma planta em várias fábricas focalizadas, deve-se determinar os produtos que serão produzidos em cada planta e alocar apropriadamente capacidade a cada uma delas. Idealmente, cada planta não deveria partilhar equipamentos e cada produto deveria ser atribuído (produzido) a somente uma planta. Entretanto, se as demandas não são suficientemente altas ou se o custo dos equipamentos é muito elevado, pode ser mais vantajoso ter vários produtos partilhando o mesmo equipamento, tal como ocorre em sistemas *job-shops* de muitas fábricas, por exemplo, fábricas metal-mecânicas e de linha branca (Silva & Morabito, 2007), fábricas de semicondutores nos processos de metalização e fotolitografia (Bitran & Tirupati, 1989), etc. Mesmo assim, há várias vantagens em particionar plantas grandes em subunidades menores e mais focalizadas. Dentre as quais, podem-se citar fluxos de produtos simplificados, aperfeiçoamentos do controle, respostas mais rápidas à demanda e maior previsibilidade dos *leadtimes* de produção.

Souza *et al.* (2001) definem fábrica focalizada como uma maneira de lidar com complexidade em uma operação de manufatura. Processos ou produtos partilhando características similares são agrupados e produzidos em plantas isoladas ou em plantas dentro de uma planta. O agrupamento pode ocorrer de acordo com tecnologia de processo, demanda de mercado, volume de produto, nível de qualidade e equipamento de manufatura. Um exemplo de um problema da classe *SP3* é (Bitran & Dasu, 1992; Morabito, 1998):

(*SP3.1*) Número de produtos e *WIP* desejados em cada unidade de produção:

- *Objetivo*: minimizar o custo de aquisição de equipamentos.
- *Variáveis de decisão*: número de unidades de produção, *mix* de produtos em cada unidade, e capacidade de cada estação.
- *Restrições*: limitantes superiores para o número de produtos e o *WIP* em cada unidade.

Conforme mencionado, sistemas *job-shop* podem ser representados por redes de filas caracterizadas pelo número de estações (nós), roteiro dos produtos (seqüência de operações) e pelo tipo de rede (aberta ou fechada). O roteiro, determinístico ou probabilístico, pode ser seqüencial, seqüencial com realimentação, de montagem, cíclico e acíclico. Roteiros probabilísticos podem ser empregados para representar, por exemplo, a probabilidade de quebra de máquinas ou a necessidade de retrabalho em produtos. Uma estação pode ser

visitada mais de uma vez por um mesmo produto para realizar operações diferentes. Para especificar um nó na rede, devem ser identificados os processos que descrevem a chegada e o serviço de produtos, assim como a estrutura e disciplina de serviço dos servidores.

O processo de chegada é descrito pelos intervalos de tempo entre chegadas de produtos no sistema, e pode ser determinístico ou probabilístico. O processo de serviço é descrito pelas quantidades de tempo de processamento demandadas pelas chegadas nas máquinas (tempos de serviço), e pode ser determinístico ou probabilístico. Com relação à estrutura e disciplina de serviço, é preciso especificar: (i) capacidade da fila; (ii) número de servidores (máquinas); (iii) número de classes de produtos; (iv) disciplina de atendimento, por exemplo, *FCFS* (*first-come-first-serve*). Para mais detalhes sobre a representação e análise de sistemas discretos de manufatura por meio de modelos de redes de filas, o leitor pode consultar, por exemplo, Whitt (1983a, 1983b), Kouvelis & Tirupati (1991), Buzacott & Shanthikumar (1993), Suri *et al.* (1993), Gershwin (1994), Morabito (1998), Govil & Fu (1999), Kouvelis *et al.* (2005) e as referências neles contidas. Para trabalhos modelando sistemas logísticos de suprimento, distribuição e movimentação interna por meio de redes de filas, o leitor pode consultar, por exemplo, Kerbache & Smith (2004), Nakano & Ohno (1999) e Warsing *et al.* (2001).

Neste artigo o sistema *job-shop* é modelado por meio de uma rede de filas aberta com n nós, cada nó $j = 1, 2, \dots, n$ corresponde a um sistema de fila $G/G/m_j$, em que G significa que as distribuições dos processos de chegada e serviço dos produtos na estação podem ser genéricas e m_j é o número de máquinas (idênticas) na estação. Considera-se que haja r famílias ou classes de produtos na rede, com cada classe $k = 1, 2, \dots, r$ percorrendo um roteiro determinístico diferente que pode visitar mais de uma vez a mesma estação. Por simplicidade, no texto que segue a classe de produto k é referida simplesmente como produto k . Sejam $E(s_j)$ e cs_j a média e o coeficiente quadrático de variação dos tempos de serviço dos produtos na estação j (o coeficiente quadrático de variação de uma variável aleatória x é definido pela razão entre a variância de x e o quadrado do valor esperado de x , $cx = V(x)/E^2(x)$). Note que a média $E(s_j)$ e a variabilidade cs_j dependem de todos os produtos passando pela estação j .

Definem-se $E(a_j)$ e ca_j como a média e o coeficiente quadrático de variação dos intervalos de tempo entre chegadas de produtos na estação j . Definem-se ainda a taxa de chegadas de produtos $\mathbf{l}_j = 1/E(a_j)$ e a taxa de serviço de produtos $\mathbf{m}_j = 1/E(s_j)$ na estação j . Todos estes valores podem ser estimados por meio do método de decomposição paramétrico (e.g., Shanthikumar & Buzacott, 1981; Whitt, 1983a, 1983b, 1994, 1995; Segal & Whitt, 1989; Kouvelis & Tirupati, 1991; Suri *et al.*, 1993; Bitran & Morabito, 1996). As variabilidades ca_j e cs_j podem ser controladas por uma agregação apropriada dos produtos em grupos, e por uma atribuição de capacidade a cada estação dentro de um grupo. Cada grupo de produtos assim determinado representa uma linha de produção ou fábrica focalizada. Conforme Bitran & Sarkar (1994b), os produtos são agregados com vistas a minimizar a complexidade global do sistema.

3. Medidas de complexidade, modelo de partição da instalação e algoritmo de solução

3.1 Medidas de complexidade

Sistemas de manufatura têm se tornado mais complexos à medida que uma maior variedade de produtos compete pelos mesmos recursos, as demandas são incertas e os ciclos de vida dos produtos ficam mais curtos. Alternativas para reduzir essa complexidade têm sido

sugeridas: duplicação de recursos, reprojeto de produtos e processos de manufatura, e partição de plantas de manufatura em subplantas. A complexidade de um *job-shop* pode ser medida de várias maneiras, tais como similaridade entre roteiros de fabricação de produtos, tempos de processamento em estações, etc. Apesar de úteis, estas medidas em geral não são suficientes para fornecer uma compreensão adequada na gestão ou projeto de *job-shops*. Bitran & Sarkar (1994b) consideraram dois tipos de medidas de complexidade: uma do ponto de vista de gestão do produto e outra do ponto de vista de gestão da estação. Do ponto de vista de gestão do produto, um gerente deve ser capaz de prever o *leadtime* de produção de um produto o mais precisamente possível. A previsibilidade foi definida como a variância do tempo total gasto por um produto no sistema. A variância pode ser reduzida por meio de adição de máquinas na estação (e aí há o *trade-off* entre previsibilidade e investimento requerido), ou por meio de partição apropriada dos produtos.

Do ponto de vista de gestão da estação, quanto maior a diversidade de produtos ou incertezas na chegada em uma estação, maior deve ser a flexibilidade para operá-la, em termos de manutenção e programação. Maior flexibilidade pode ser alcançada adquirindo mais máquinas. Bitran & Sarkar (1994b) mediram a diversidade de produtos pela variância dos tempos de serviço dos produtos, ou pelo coeficiente quadrático de variação dos tempos de serviço na estação (cs_j). A incerteza da chegada de produtos é medida pelo coeficiente quadrático dos intervalos de tempo entre chegadas na estação (ca_j). A seguir são discutidas medidas de complexidade que podem compor o conjunto de restrições de problemas de otimização relacionados à partição da instalação. No presente artigo o enfoque é apenas na complexidade do ponto de vista da gestão do produto (previsibilidade). Estudos com outras medidas de complexidade do ponto de vista da gestão da estação podem ser encontrados em Bitran & Sarkar (1994b) e Silva (2005).

O objetivo das restrições de previsibilidade é reduzir a variância do tempo que cada produto permanece no sistema. Se $V(T_k)$ é a variância do *leadtime* de produção do produto k na rede, T_k , então se requer que:

$$V(T) = \sum_{k=1}^r w_k V(T_k) \leq T^* \quad (1)$$

onde w_k é um peso (adimensional) associado ao produto k (este peso pode ser arbitrado pelo gerente dependendo do valor ou importância relativa do produto k), e T^* é um limitante superior para a variância ponderada do *leadtime* de produção da rede, $V(T)$. Note que se $w_k = 1$ para todo k , então $V(T)$ é simplesmente a variância do *leadtime* de produção da rede. A metodologia descrita nas próximas seções é diretamente aplicável em situações com múltiplas restrições de complexidade como (1), ou com limitantes individuais para cada produto k . Observe que quanto maior o valor de $V(T)$ (lado esquerdo de (1)), menor o potencial de previsibilidade do sistema.

No método de decomposição paramétrica, cada estação da rede é tratada como se fosse estocasticamente independente das demais. Sejam n_k o número de operações no roteiro do produto k e n_{kl} a estação visitada para operação l do roteiro do produto k . Por simplicidade, a seguir considera-se que os tempos de serviço s_{kl} da operação l do roteiro do produto k (i.e., na estação n_{kl}) sejam determinísticos, portanto, a variância destes tempos é nula (i.e., $V(s_{kl}) = cs_{kl} = 0$). Como resultado, a variância $V(T_k)$ pode ser aproximada simplesmente como a soma das variâncias de tempos de espera nas filas do roteiro do produto k :

$$V(T_k) \approx \sum_{l=1}^{n_k} \sum_{j=1}^n V(Wq_{n_{kl}}) 1_{\{n_{kl}=j\}} \quad (2)$$

em que $1_{\{\cdot\}}$ é uma função indicadora igual a 1 se $\{\cdot\}$ é verdadeiro e 0 caso contrário; e $V(Wq_{n_{kl}})$ é a variância do tempo de espera na estação n_{kl} . Esta variância pode ser avaliada utilizando-se aproximações de tráfego pesado, em que as utilizações médias de capacidade $\mathbf{r}_j = \mathbf{I}_j / m_j \mathbf{m}_j$ são altas, isto é, próximas de 1 (Kleinrock, 1975; Tijms, 1986):

$$V(Wq_j) \approx \frac{(ca_j + cs_j)}{4} E^2(Wq_j)_{M/M/m_j} \quad (3)$$

onde $E^2(Wq_j)_{M/M/m_j}$ representa o quadrado do tempo médio de espera para um sistema de fila $M/M/m_j$ em equilíbrio. Um modelo de projeto de fábrica focalizada utilizando a restrição de complexidade (1) é apresentado em seguida.

3.2 Definição do problema de redução de complexidade (PRC)

Bitran & Sarkar (1994b) propuseram um problema para o projeto de fábrica focalizada, com o objetivo de encontrar a partição da instalação que reduz o custo total de máquinas em todas as subplantas (linhas) e atende uma determinada medida de complexidade. Sejam:

$$x_{kp} = \begin{cases} 1, & \text{se o produto } k \text{ é atribuído à linha } p \\ 0, & \text{caso contrário} \end{cases}$$

m_{jp} : número de máquinas na estação j da linha p .

O modelo PRC a seguir envolve decidir a atribuição de cada produto k a uma linha p , e quantificar o número de máquinas m_{jp} em cada estação j da linha p :

$$\min \sum_{j=1}^n \sum_{p=1}^r c_{jp} m_{jp} \quad (4a)$$

$$sa : \text{restrição de complexidade} \quad (4b)$$

$$\sum_{p=1}^r x_{kp} = 1, \quad k = 1, 2, \dots, r \quad (4c)$$

$$\mathbf{I}_{jp} / \mathbf{m}_{jp} = \sum_{k=1}^r \sum_{l=1}^{n_k} \mathbf{I}'_k E(s_{kl}) 1_{\{n_{kl}=j\}} x_{kp} < m_{jp}, \quad j = 1, \dots, n; \quad p = 1, \dots, r \quad (4d)$$

$$x_{kp} \in \{0, 1\}; \quad m_{jp} \text{ inteiro}, \quad k, p = 1, \dots, r; \quad j = 1, \dots, n \quad (4e)$$

em que $E(s_{kl})$ é o tempo médio de serviço do produto k na l -ésima operação do seu roteiro. Note que \mathbf{I}_{jp} é a taxa média de chegada de produtos na estação j da linha p , definida por:

$\mathbf{I}_{jp} = \sum_{k=1}^r \sum_{l=1}^{n_k} \mathbf{I}'_k 1_{\{n_{kl}=j\}} x_{kp}$, onde \mathbf{I}'_k é a taxa externa de chegada de produtos k na rede. O parâmetro c_{jp} é o custo unitário de cada máquina na estação j da linha p . Por simplicidade, no presente trabalho considera-se $c_{jp}=1$ ($j=1, \dots, n; p=1, \dots, r$), ou seja, o objetivo (4a) do modelo passa a ser simplesmente minimizar o número total de máquinas em todas as estações e linhas. As restrições (4c) e (4e) garantem que todo produto deve ser atribuído a somente uma linha de produção. A restrição (4d) considera que o sistema esteja em equilíbrio, ou seja, o nível médio de utilização na estação j da linha p satisfaz: $\mathbf{r}_{jp} = \mathbf{I}_{jp} / m_{jp} \mathbf{m}_{jp} < 1$ para toda estação j e linha p . Em implementações computacionais, é desejável que esta restrição satisfaça uma certa tolerância: $\mathbf{r}_{jp} \leq 1 - \epsilon$, com $\epsilon > 0$. Observe que o modelo admite que

podem existir tantas linhas p quantos forem os produtos k , uma vez que p varia de 1 a r . A depender da medida de complexidade que se pretende utilizar, o PRC tem uma formulação diferente. Denominamos problema de redução de complexidade PRC-1 quando a restrição de complexidade (4b) é dada por (1) em que:

$$V(T_k) \approx \sum_{l=1}^{n_k} \sum_{j=1}^n \sum_{p=1}^r V(Wq_{jp}) 1_{\{n_{kl}=j\}} x_{kp} \quad (5)$$

$$V(Wq_{jp}) \approx \frac{(ca_{jp} + cs_{jp})^2}{4} E^2(Wq_{jp})_{M/M/m_j} \quad (6)$$

onde, similarmente a (3), $V(Wq_{jp})$, ca_{jp} e cs_{jp} são a variância do tempo de espera em fila, a variabilidade do processo de chegada e a variabilidade do processo de serviço na estação j da linha p , respectivamente, e (Kleinrock, 1975; Tijms, 1986):

$$E(Wq_{jp})_{M/M/m_j} = \frac{(I_{jp}/m_{jp})^{m_{jp}} p_{jp}(0)}{m_{jp} m_{jp} (1 - r_{jp})^2 m_{jp}!} \quad (7)$$

Outros PRC's envolvendo outras medidas de complexidade diferentes de (1) foram estudados em Bitran & Sarkar (1994b) e Silva (2005).

3.3 Solução do problema PRC-1

O PRC-1 (4a)-(4e) com as restrições (1), (5), (6) e (7) é um problema de programação não-convexo inteiro e difícil para resolver otimamente. Bitran & Sarkar (1994b) propuseram um método aproximado de solução baseado em um algoritmo para alocar produtos às linhas de produção, requerendo o menor número de máquinas (admitindo $c_{jp}=1$ em (4a)). *Grosso modo*, este algoritmo utiliza idéias similares às do algoritmo de economias de Clarke-Wright para problemas de roteirização de veículos (vide Larson & Odoni, 1981). Basicamente, o procedimento do tipo guloso é iniciado atribuindo uma linha para cada produto, em que se calcula o número mínimo de máquinas necessárias para atender às restrições. Em seguida, agrega-se o par de linhas (e, conseqüentemente, os produtos) que proporciona maior redução (economia) no número de máquinas. A partição da instalação é atualizada e, novamente, agrega-se o par de linhas que proporciona maior redução no número de máquinas. Este procedimento continua enquanto houver economia no número de máquinas com a agregação de um par de linhas qualquer, conforme descrição a seguir:

Algoritmo:

Passo 0: Faça $iter=0$. Aloque cada produto a uma linha separada. Aloque máquinas às estações de modo a satisfazer as restrições com o menor número de máquinas possível (o método utilizado para alocar máquinas é definido adiante). Calcule o número de máquinas requeridas.

Passo 1: Faça $iter=iter+1$. Escolha duas linhas quaisquer e agregue-as. Aloque máquinas às estações de modo a satisfazer as restrições com o menor número de máquinas possível (o método utilizado para alocar máquinas é definido adiante). Seja $\Delta m(p_1, p_2)$ a mudança no número de máquinas quando as linhas p_1 e p_2 são agregadas e, conseqüentemente, os respectivos produtos pertencentes às linhas. Calcule $\Delta m(p_1, p_2)$ para cada par de linhas da rede.

Passo 2: Se $\Delta m(p_1, p_2) < 0$, então faça $(p_1^*, p_2^*) = \arg \min_{p_1, p_2} \Delta m(p_1, p_2)$ (i.e., (p_1^*, p_2^*) é o par de linhas que minimiza $\Delta m(p_1, p_2)$), agregue as linhas p_1^* e p_2^* , atualize a lista de linhas e volte para o passo 1. Senão, pare. A agregação dos produtos nas linhas da iteração anterior (partição da instalação) é a que resulta no menor número de máquinas.

Note no algoritmo que ao agregar um par de linhas qualquer, é necessário calcular o número mínimo de máquinas requeridas para calcular a mudança no número de máquinas com a agregação. Para uma dada partição (x_{kp} conhecidos para $k=1, \dots, r$, $p=1, \dots, r$), ca_{jp} e cs_{jp} podem ser determinados e a variância em (6) e (7) é convexa em m_{jp} . Entretanto, o PRC-1 permanece um modelo não-linear e inteiro em m_{jp} , o que inviabiliza sua resolução direta por meio de um *solver* de programação linear inteira para determinar o número mínimo de máquinas requerido para a partição dada. Por isso, no presente trabalho propõe-se utilizar o algoritmo de Boxma *et al.* (1990), com uma pequena variação no seu índice de prioridade (veja adiante). Em outras palavras, conhecidos x_{kp} , o algoritmo de Boxma *et al.* (1990) é aplicado para determinar o número mínimo de máquinas nas estações das linhas. O algoritmo começa com o menor número de máquinas possível para cada estação, satisfazendo (4d). Em cada iteração, adiciona-se uma máquina na estação com o menor índice de prioridade, ou seja, aquela estação que conduz à máxima redução na medida de variância ponderada. O algoritmo pára tão logo a restrição (4b) (i.e., a expressão (1)) seja satisfeita. O índice de prioridade do algoritmo de Boxma *et al.* (1990), que determina a estação j de uma linha p na qual uma máquina será adicionada, é definido por:

$$PI_{jp}(m_j) = \frac{\Delta m_{jp}}{-\Delta \sum_{k=1}^r w_k V(T_k)} \quad j=1, \dots, n \quad (8)$$

em que $\Delta m_{jp} = (m_{jp} + 1) - (m_{jp}) = 1$ representa a adição de uma máquina na estação j da linha p (note que no caso de c_{jp} ser diferente de 1 em (4a), este termo seria $c_{jp}(m_{jp} + 1) - c_{jp}(m_{jp}) = c_{jp}$) e $\Delta \sum_{k=1}^r w_k V(T_k) = \sum_{k=1}^r w_k V(T_k)(m_{jp} + 1) - \sum_{k=1}^r w_k V(T_k)(m_{jp})$ representa a mudança na variância ponderada do *leadtime* de produção da rede $V(T)$ com a adição da máquina.

O método aproximado de decomposição paramétrico foi empregado para avaliar o desempenho da rede em cada configuração. A menos que seja feita alguma menção em contrário, foi utilizado o sistema linear nas variáveis ca_{jp} composto pelas seguintes aproximações:

$$ca_{jp} \approx \sum_{i=0}^n \frac{I_{ijp}}{I_{jip}} ca_{ijp} \quad j=1, \dots, n; \quad p=1, \dots, r \quad (9)$$

$$cd_{jp} \approx \frac{r_{jp}^2}{\sqrt{m_{jp}}} cs_{jp} + [1 - r_{jp}^2] ca_{jp} + r_{jp}^2 \left[1 - \frac{1}{\sqrt{m_{jp}}} \right] \quad j=1, \dots, n; \quad p=1, \dots, r \quad (10)$$

$$cd_{jip} \approx q_{jip} cd_{jp} + (1 - q_{jip}) q_{jip} ca_{jp} + (1 - q_{jip})^2 ce_{jp} \quad i, j=1, \dots, n; \quad p=1, \dots, r \quad (11)$$

em que I_{ijp} , ca_{ijp} , cd_{jp} , q_{jip} são, respectivamente, a taxa de chegadas da estação i para a estação j , a variabilidade das chegadas da estação i para a estação j , a variabilidade das

partidas da estação j , a proporção de produtos que vão da estação i para a estação j , todos na linha p . O termo ce_{jp} em (11) é a média das variabilidades das chegadas externas ponderada pelas taxas de chegadas externas na estação j da linha p . Para mais detalhes deste sistema, o leitor pode consultar e.g. Segal & Whitt (1989), Morabito (1998) e Silva (2005).

4. Resultados computacionais

Nesta seção analisa-se a adequação e aplicabilidade da abordagem da seção 3 para o problema de partição da instalação. Inicialmente o modelo PRC-1 foi aplicado a duas redes ilustrativas, uma composta de uma única estação $G/G/m_j$ e dois produtos e a outra de duas estações $G/G/m_j$ e dois produtos. Em seguida, aplicou-se o modelo a uma rede mais realista com 13 estações $G/G/m_j$ e 10 produtos, utilizada em Bitran & Sarkar (1994b). As redes ilustrativas são exemplos simples e foram aqui criadas para ilustrar sob quais condições é mais vantajoso particionar ou não particionar a instalação. As Figuras 1 e 2 apresentam as redes ilustrativas, denominadas doravante por Rede 1 e Rede 2, respectivamente. Os números nos arcos indicam os produtos 1 e 2. As Tabelas 1 e 2 mostram os dados referentes às Redes 1 e 2, respectivamente. Nas Tabelas, I'_k e ca'_k correspondem a taxa (em produtos por hora) e a variabilidade (adimensional) das chegadas externas do produto k . Conforme mencionado na seção 3.1, por simplicidade, os tempos de serviço s_{kl} do produto k na estação n_{kl} foram considerados determinísticos (i.e., $cs_{kl}=0$).

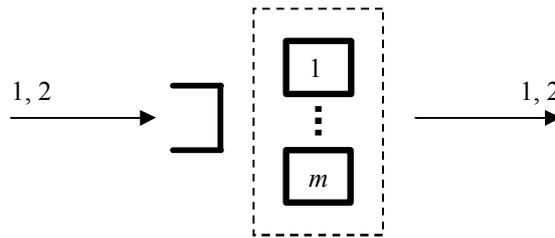


Figura 1 – Rede formada por um estágio único $G/G/m_j$, processando dois produtos (Rede 1).

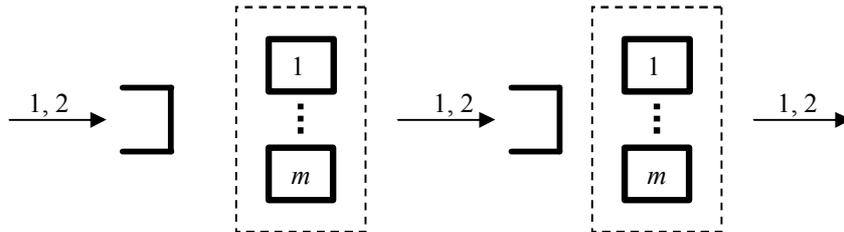


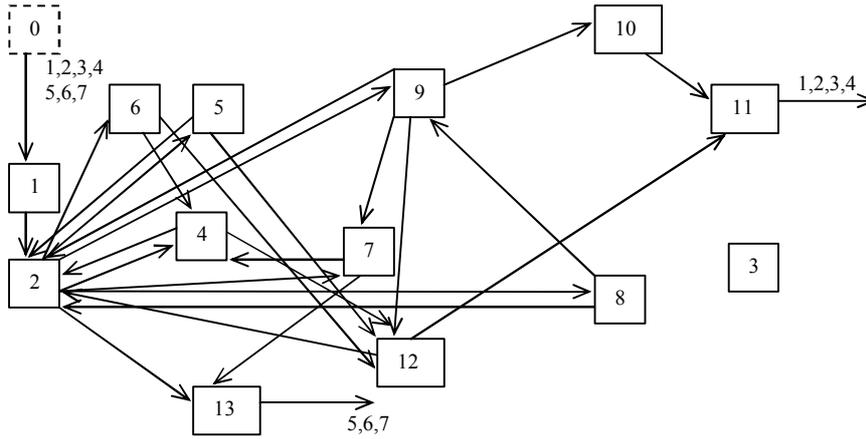
Figura 2 – Rede formada por dois estágios $G/G/m_j$, processando dois produtos (Rede 2).

Tabela 1 – Dados para os dois produtos na Rede 1.

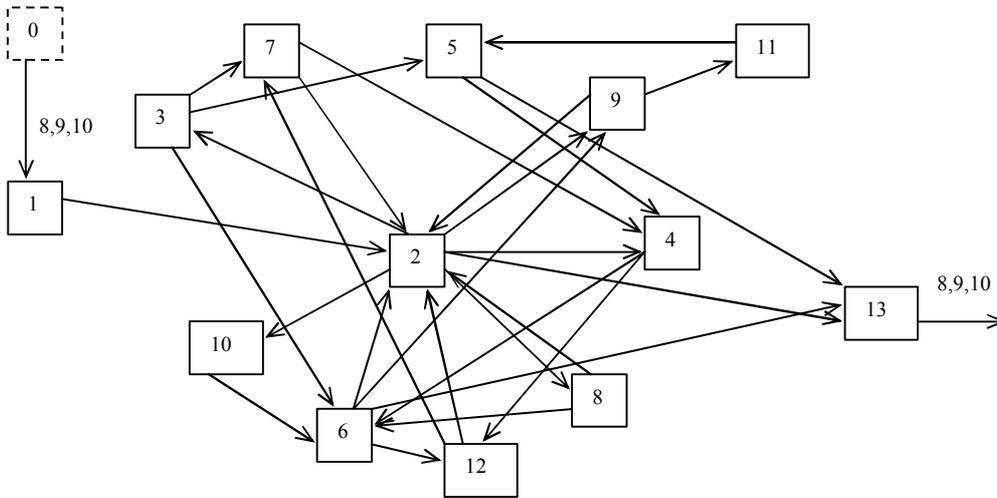
Produto	I'_k (prod/h)	$E(s_{kl})$ (h/prod)	ca'_k	cs_{kl}	w_k
1	4	0,2	2,0	0	1
2	1	0,9	0,1	0	1

Tabela 2 – Dados para os dois produtos na Rede 2.

Produto	I'_k (prod/h)	$E(s_{k1})$ (h/prod)	$E(s_{k2})$ (h/prod)	ca'_k	cs_{k1}	cs_{k2}	w_k
1	4	0,15	0,15	0,5	0	0	1
2	1	0,85	0,85	0,1	0	0	1



(a)



(b)

Figura 3 – Rede formada por vários estágios $G/G/m_j$, processando os produtos 1,2,...,7 (Figura 3a) e 8,9,10 (Figura 3b) de produtos na Rede 3.

A Figura 3 ilustra a rede composta por um conjunto de estações $G/G/m_j$ que doravante é chamada de Rede 3. Os números nos nós indicam as estações e os números nos arcos indicam os produtos. As características desta rede referentes aos processos de chegadas e roteiros dos

produtos são dadas na Tabela 3. A Tabela 4 apresenta os tempos de serviço dos produtos nas estações da rede. Da mesma forma que nas Redes 1 e 2, considerou-se na Rede 3 s_{kl} determinístico, ou seja, $cs_{kl}=0, k=1,\dots,10; l=1,\dots,n_k$, conforme em Bitran & Sarkar (1994b).

Tabela 3 – Parâmetros referentes ao processo de chegada (I'_k, ca'_k), ao peso da variância do *leadtime* (w_k) e ao roteiro (n_k, n_{kl}) dos 10 produtos da Rede 3.

Produto	I'_k (prod/h)	ca'_k	w_k	n_k	n_{kl}
1	0,10	0,333	0,5	7	1, 2, 4, 2, 9, 10, 11
2	0,10	0,500	0,5	8	1, 2, 5, 2, 8, 9, 10, 11
3	0,10	0,333	0,5	8	1, 2, 6, 4, 2, 9, 12, 11
4	0,10	0,333	0,5	8	1, 2, 7, 4, 2, 9, 10, 11
5	0,10	0,250	0,5	8	1, 2, 4, 12, 2, 9, 2, 13
6	0,10	0,500	0,5	8	1, 2, 5, 12, 2, 9, 7, 13
7	0,10	0,250	0,5	8	1, 2, 6, 12, 2, 8, 2, 13
8	0,10	0,333	0,5	12	1, 2, 3, 7, 4, 12, 2, 8, 6, 9, 2, 13
9	0,10	0,050	0,1	13	1, 2, 3, 5, 4, 6, 12, 2, 8, 2, 10, 6, 13
10	0,10	0,100	0,1	13	1, 2, 3, 6, 2, 4, 12, 7, 2, 9, 11, 5, 13

Tabela 4 – Parâmetros referentes ao tempo médio de serviço $E(s_{kl})$ do produto k e operação l , para os 10 produtos da Rede 3.

Produto	Operação (h/prod)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0,2	0,1	0,2	0,2	0,9	3,4	3,8						
2	0,6	0,2	3,0	0,6	1,8	0,9	0,8	2,4					
3	0,8	0,4	1,8	3,5	0,6	0,4	1,0	0,6					
4	1,8	0,4	1,8	0,2	0,2	0,4	0,8	0,4					
5	0,5	0,3	1,5	0,6	0,2	0,8	0,5	0,6					
6	0,6	0,7	0,7	2,6	0,7	0,8	0,4	3,0					
7	0,7	0,2	1,0	1,2	0,3	1,8	0,3	0,7					
8	0,3	0,2	0,6	1,8	0,2	0,7	0,7	0,6	2,2	0,4	0,8	1,3	
9	1,0	0,8	8,0	15,0	1,2	2,8	2,8	1,2	14,0	1,6	9,6	2,8	3,6
10	1,3	0,8	21,6	8,8	1,2	12,0	4,0	14,0	0,8	9,2	1,6	15,0	6,8

4.1 Experimentos com a Rede 1

Nesta seção são apresentados os resultados dos experimentos com a Rede 1, formada por um único estágio $G/G/m_j$. O modelo (4a)-(4e) com a restrição (1) foi solucionado empregando o algoritmo da seção 3, juntamente com o algoritmo de Boxma *et al.* (1990) para alocar capacidade às linhas de produção. Inicialmente resolveu-se a Rede 1 com limitante superior para a variância do *leadtime* $T^*=1,0 h^2$. Note na Tabela 1 que o produto 2 é bem comportado, ou seja, a variabilidade do processo de chegadas é pequena ($ca'_2=0,1$), diferente do produto 1 ($ca'_1=2,0$).

A Tabela 5 mostra o número mínimo de máquinas requeridas ($M = \sum_{j=1}^n \sum_{p=1}^r m_{jp}$) e a variância do *leadtime* da rede $V(T)$ ao aplicar o algoritmo na Rede 1. O algoritmo pára na iteração 0 com a partição [1],[2], com 2 máquinas. A primeira coluna da tabela indica a iteração do algoritmo, as demais colunas p_1 e p_2 indicam os produtos contidos em cada uma das linhas [1],[2] e o número de máquinas (entre parênteses). Nota-se que são necessárias no mínimo 2 máquinas (uma em cada linha) para que a variância ponderada do *leadtime* da rede não seja superior a 1,0 ($V(T)=0,804$). Se os dois produtos forem agregados em uma linha única [1,2], esta “não partição” requer, no mínimo, 3 máquinas para que a variância do *leadtime* seja inferior a 1,0. Se o número de máquinas na linha única for igual a 2, obtém-se $V(T)=2,070$, requerendo, portanto, uma terceira máquina para que $V(T) \leq T^* = 1,0$. Logo, este exemplo ilustra uma situação em que particionar a instalação é mais vantajoso do que não particioná-la.

A Tabela 6 mostra algumas medidas obtidas para as configurações [1],[2] e [1,2] da Rede 1 quando o número de máquinas na rede é $M = 2$ máquinas. Nota-se na tabela que, para a mesma capacidade ($M = 2$), particionar a rede é melhor do que não particioná-la do ponto de vista da variância do *leadtime* $V(T)$ (i.e., 0,804 ao invés de 2,070 h²), do ponto de vista da média do *leadtime* $E(T_1)$, $E(T_2)$ e $E(T) = (I_1' E(T_1) + I_2' E(T_2)) / (I_1' + I_2')$ (i.e., 1,000 e 1,305 h, ao invés de 1,217 e 1,917 h, respectivamente, para os produtos 1 e 2, e 1,061 e 1,357 h para a rede considerando as configurações [1],[2] e [1,2], respectivamente), e do ponto de vista do número médio de produtos na rede $E(L) = E(L_1) + E(L_2)$ (i.e., 5,305 produtos, ao invés de 6,786 produtos, respectivamente).

Tabela 5 – Número de máquinas (M) e variância da rede $V(T)$ obtidos para as diferentes partições em cada iteração (*iter*) para $T^* = 1,0$ na Rede 1.

<i>Iter</i>	Produtos alocados às linhas de produção (p_1, p_2) em cada partição e número de máquinas em cada linha (entre parênteses)		M	$V(T)$
	p_1	p_2		
0	1 (1)	2 (1)	2	0,80 4

Tabela 6 – Medidas obtidas para as configurações [1],[2] (partição) e [1,2] (linha única) na rede com $M = 2$.

Medidas	Partição [1],[2]	Linha única [1,2]
$V(T)$ (h ²)	0,804	2,070
$E(L_1)$ (prod)	4,000	4,869
$E(L_2)$ (prod)	1,305	1,917
$E(T_1)$ (h)	1,000	1,217
$E(T_2)$ (h)	1,305	1,917
$E(T)$ (h)	1,061	1,357

Realizou-se outros experimentos com a Rede 1, com o objetivo de avaliar o efeito de ca'_k em $V(T)$ e, conseqüentemente, na configuração da rede. Experimentos com as configurações [1],[2] e [1,2] foram conduzidos mantendo $M = 2$, conforme mostra a Tabela 7. Observa-se na Tabela que, quando a variabilidade do processo de chegadas externas para os dois

produtos ca'_1 e ca'_2 é pequena (tendendo a um processo de chegadas determinístico), convém particionar a instalação (caso (a)) em termos de $V(T)$. Por outro lado, quando ca'_1 e ca'_2 assumem valores maiores (casos (b) e (c)), a partição da instalação é bem menos conveniente que a agregação das linhas. O caso (d) mostra que o fato dos coeficientes serem bastante distintos, isoladamente, não implica que a partição [1],[2] seja mais conveniente do que a linha única [1,2] (compare com o caso base da Tabela). Por isso, outros experimentos com a Rede 1 foram conduzidos de modo a avaliar a influência do nível de utilização do produto k , $r_k = I'_k E(s_{k1})$ e do nível de utilização da estação r ao particionar ou não a instalação. Note que para todos os casos (a)-(d) da Tabela 7 tem-se: $r_1=0,8$, $r_2=0,9$ e $r=0,85$.

Tabela 7 – Variância do *leadtime*, $V(T)$, ao mudar a variabilidade do processo de chegadas ca'_k da Rede 1, mantendo $M = 2$ máquinas.

Variabilidade do processo de chegadas externas	$V(T)$	
	Partição [1],[2]	Linha única [1,2]
Base: $ca'_1=2,0$ e $ca'_2=0,1$	0,804	2,070
a) $ca'_1=0,1$ e $ca'_2=0,1$	0,166	0,240
b) $ca'_1=0,5$ e $ca'_2=0,5$	4,141	0,544
c) $ca'_1=1,0$ e $ca'_2=1,0$	16,563	1,103
d) $ca'_1=0,1$ e $ca'_2=2,0$	65,612	0,526

Os experimentos ilustrados na Tabela 8 foram realizados também se mantendo $M = 2$ máquinas. Observando os casos (e) e (f), pode-se notar que há uma pequena vantagem em agregar os produtos em termos de $V(T)$. Entretanto, quando a utilização atinge níveis mais altos, por exemplo, $r = 0,8$ (caso (g)), particionar a instalação torna-se mais vantajoso do que não particioná-la. O caso (h) ilustra que, quando as estações da partição [1],[2] estão desbalanceadas (i.e., níveis de utilização r_k bem diferentes) e a estação com o produto menos comportado (aquela com maior ca'_k , vide Tabela 1) possui nível de utilização mais alto, é conveniente não particionar a instalação. Por outro lado, se as estações da partição [1],[2] estão desbalanceadas, mas o nível de ocupação da estação menos comportada é mais baixo (caso (i)), então convém particionar.

Tabela 8 – Variância do *leadtime*, $V(T)$, ao mudar o nível de utilização das estações da Rede 1, mantendo $M = 2$ máquinas.

Partição [1],[2]		Linha única [1,2]	
Utilização das estações	$V(T)$	Utilização das estações	$V(T)$
Base: $r_1=0,8$ e $r_2=0,9$	0,804	Base: $r=0,85$	2,070
e) $r_1=0,4$ e $r_2=0,4$	0,005	a) $r=0,40$	0,002
f) $r_1=0,6$ e $r_2=0,6$	0,053	b) $r=0,60$	0,043
g) $r_1=0,8$ e $r_2=0,8$	0,666	d) $r=0,80$	0,771
h) $r_1=0,8$ e $r_2=0,4$	0,640	f) $r=0,60$	0,027
i) $r_1=0,4$ e $r_2=0,8$	0,030	g) $r=0,60$	0,081

De uma forma geral, observa-se na Rede 1 que é conveniente particionar os produtos [1],[2] quando pelo menos um dos produtos tem baixo ca'_k (isto é, ele é bem comportado em termos de variabilidade), e tem alto nível de utilização (veja casos base, (a), (g) e (i) nas Tabelas 7 e 8). Outros experimentos foram realizados com redes formadas por um estágio único $G/G/m_j$ e com múltiplos produtos, e os resultados obtidos foram similares aos anteriores. Por motivo de economia de espaço, eles não são aqui apresentados; o leitor interessado pode consultar Silva (2005) para mais detalhes.

4.2 Experimentos com a Rede 2

O algoritmo da seção 3 foi aplicado à Rede 2 com limitante superior para a variância do *leadtime* $T^* = 0,2 h^2$. Conforme a Tabela 2, o produto 2 é mais comportado ($ca'_2=0,1$) que o produto 1 ($ca'_1=0,5$), e também possui taxa de chegada I'_2 menor e tempos médios de serviço $E(s_{2l})$ maiores que o produto 1. A Tabela 9 ilustra as configurações obtidas ao aplicar o algoritmo na Rede 2. O algoritmo pára na iteração 0 com a partição [1],[2], com $V(T) = 0,067$ (menor que $T^*=0,2$) e $M = 4$ máquinas. A Tabela 10 mostra as variabilidades dos processos de chegadas ca_{jp} e os níveis de utilização r_{jp} das estações das duas linhas, obtidos com a aproximação (9)-(11) do método de decomposição paramétrico (seção 3). Note na tabela que a partição [2] apresenta estações com ca_{jp} baixos e r_{jp} altos. Espera-se, portanto, que o processo de agregação das linhas [1] e [2] em uma linha única [1,2] seja desvantajoso, conforme foi observado com a Rede 1.

Tabela 9 – Número total de máquinas (M) e variância do *leadtime* da rede $V(T)$ obtidos para as diferentes partições em cada iteração (*iter*) do algoritmo para $T^*=0,2$ na Rede 2.

Iter	Produtos alocados às linhas de produção (p_1, p_2) em cada partição e número de máquinas em cada linha – entre parênteses		M	V(T)
	p_1	p_2		
0	1 (2)	2 (2)	4	0,067

Tabela 10 – Coeficiente de variação do processo de chegadas e nível de utilização das estações das linhas [1] e [2] na Rede 2.

Estação j	Partição [1]		Partição [2]	
	ca_{j1}	r_{j1}	ca_{j2}	r_{j2}
1	0,500	0,60	0,100	0,85
2	0,320	0,60	0,028	0,85

De fato, ao agregar as duas linhas em uma única, o número de máquinas necessárias para que $V(T) \leq 0,2$ aumenta de 4 para 5. Ora, conforme observado para a Rede 1, a partição é vantajosa quando pelo menos uma das linhas possui baixa variabilidade no processo de chegadas (tendendo a um processo de chegadas determinístico) e suas estações operam com

alto nível de utilização. Ou seja, a partição da linha [1,2] em duas linhas [1],[2] na Rede 2 reduz o número de máquinas (de 5 para 4) necessárias para manter a variância $V(T)$ abaixo de $T^* = 0,2$. Além disso, particionar a rede reduz $V(T)$ de 0,097 para 0,067 h^2 , além de reduzir o *leadtime* médio da rede $E(T)$ de 0,831 para 0,715 h, e o número médio de produtos na rede $E(L)$ de 4,154 para 3,577 produtos (Tabela 11).

Tabela 11 – *Leadtime*, $E(T_k)$, e número médio de produtos, $E(L_k)$, obtidos para os produtos e para a rede ($E(T)$ e $E(L)$) nas partições [1],[2] e [1,2], com $T^*=0,2$ na Rede 2.

Produto	Medidas de Desempenho para as Partições			
	$E(T_k)$ (h)		$E(L_k)$ (prod)	
	[1],[2]	[1,2]	[1],[2]	[1,2]
1	0,392	0,551	1,569	2,203
2	2,008	1,951	2,008	1,951
Rede	0,715	0,831	3,577	4,154

Aumentando-se o coeficiente de variação do processo de chegadas externas, por exemplo, de $ca_2=0,1$ para $ca_2=0,25$ na Rede 2, e aplicando-se o algoritmo, a partição da instalação torna-se menos vantajosa, pois também necessita de 5 máquinas (com $V(T) = 0,033$), assim como a opção de não particionar (com $V(T) = 0,101$). Conforme mencionado, para estimar ca_{jp} foi empregado o sistema de aproximações (9)-(11). Outros experimentos foram conduzidos na Rede 2 empregando outros sistemas de aproximações para o método de decomposição, com o objetivo de avaliar a sensibilidade dos resultados alcançados com o sistema utilizado. Por exemplo, a aproximação que descreve o processo de desagregação de partidas (11) foi substituída pela aproximação alternativa descrita em Bitran & Tirupati (1988), que explora efeitos das interferências entre produtos nas estações. Os resultados para a Rede 2 e para a Rede 2 modificada (i.e., com $ca_2=0,25$ ao invés de $ca_2=0,1$) utilizando as duas aproximações foram idênticos.

4.3 Experimentos com a Rede 3

O algoritmo da seção 3 foi aplicado à Rede 3 com $T^* = 8,0 h^2$, conforme em Bitran & Sarkar (1994b). As partições com menor número de máquinas M em cada iteração do algoritmo, com as respectivas variâncias ponderadas do *leadtime* de produção $V(T)$, são mostradas na Tabela 12. Nota-se na tabela que o algoritmo inicia ($iter=0$) com um produto em cada linha, requerendo $M = 86$ máquinas para que $V(T)=7,698 \leq 8,0$, e termina ($iter=8$) com a partição [1,2,3,4,5,6,7,8],[9,10], requerendo $M = 41$ máquinas para que $V(T)=7,908 \leq 8,0$. A Tabela 13 mostra o *leadtime* médio $E(T_k)$ e o número médio de produtos $E(L_k)$ de cada produto k na rede para a partição [1,2,3,4,5,6,7,8],[9,10], e o *leadtime* médio e o número médio de produtos na rede, $E(T) = \sum_k I_k E(T_k) / \sum_k I_k$ e $E(L) = \sum_k E(L_k)$ (note que ambos são iguais dado que $I_k = 0,1$ para todo $k = 1, \dots, 10$).

Tabela 12 – Número total de máquinas (M) e variância da rede $V(T)$ obtidos para as diferentes partições em cada iteração (*iter*) do algoritmo para $T^*=8,0$ na Rede 3.

Iter	Produtos alocados às linhas de produção (p_1, \dots, p_{10}) em cada iteração e número de máquinas em cada linha – entre parênteses										M	$V(T)$
	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}		
0	1 (6)	2 (7)	3 (7)	4 (7)	5 (6)	6 (7)	7 (6)	8 (10)	9 (13)	10 (17)	86	7,698
1	1 (6)	2 (7)	3 (7)	4 (7)	5 (10)	6 (7)	7-8 (6)		9 (13)	10 (17)	80	4,740
2	1 (6)	2 (7)	3 (7)	4 (7)	5-7-8 (10)	6 (7)			9 (13)	10 (17)	74	4,752
3	1-4 (7)	2 (7)	3 (7)		5-7-8 (10)	6 (7)			9 (13)	10 (18)	69	4,798
4	1-4 (7)	2 (7)	3-5 7-8 (11)			6 (7)			9 (13)	10 (18)	63	5,582
5	1-3 4-5 7-8 (12)	2 (7)				6 (7)			9 (13)	10 (18)	57	6,759
6	1-2 3-4 5-7 8 (14)					6 (7)			9 (13)	10 (18)	52	6,917
7	1-2 3-4 5-6 7-8 (15)								9 (13)	10 (18)	46	6,516
8	1-2 3-4 5-6 7-8 (15)								9-10 (26)		41	7,908

Tabela 13 – *Leadtime* médio, $E(T_k)$, número médio de produtos k , $E(L_k)$, e total na Rede 3, $E(T)$ e $E(L)$, na partição [1,2,3,4,5,6,7,8],[9,10].

Produto	$E(T_k)$ (h)	$E(L_k)$ (prod)
1	8,959	0,896
2	11,662	1,166
3	10,484	1,048
4	10,555	1,056
5	9,474	0,947
6	11,300	1,130
7	11,172	1,117
8	15,358	1,536
9	92,161	9,216
10	87,491	8,749
Rede	26,862	26,862

Comparando a partição referente à iteração 8 ($iter=8$) da Tabela 12 com a opção de não particionar (linha única, Tabela 14), verifica-se que esta opção, além de requerer as mesmas 41 máquinas, apresenta redução da variância da rede $V(T)$ (de 7,908 para 6,489 h^2), assim como redução do *leadtime* médio da rede $E(T)$ (de 26,862 para 25,886 h, vide Tabela 14) e do número médio de produtos $E(L)$ (de 26,862 para 25,886, vide Tabela 14). A Tabela 14 também mostra o *leadtime* médio $E(T_k)$ e o número médio de produtos $E(L_k)$ de cada produto k para a Rede 3 com a linha única. Portanto, para a Rede 3 a partição da instalação não é vantajosa. Convém salientar que o algoritmo parou na iteração 8 porque não encontrou nenhuma outra solução com número de máquinas M menor do que 41, conforme o critério de parada do passo 3 do algoritmo. Uma simples modificação neste critério permitiria que o algoritmo continuasse procurando por soluções alternativas do ponto de vista do número de máquinas M , porém melhores do ponto de vista da variância $V(T)$. No entanto, esta modificação aumentaria o número de iterações e o tempo computacional de execução até o término do algoritmo.

Tabela 14 – *Leadtime* médio, $E(T_k)$, número médio de produtos k , $E(L_k)$, e para a Rede 3, $E(L)$ e $E(T)$, na linha única [1-10].

Produto	$E(T_k)$ (h)	$E(L_k)$ (prod)
1	10,405	1,041
2	12,742	1,274
3	11,481	1,148
4	8,819	0,882
5	7,050	0,705
6	13,133	1,313
7	8,555	0,855
8	14,569	1,457
9	69,451	6,945
10	102,658	10,266
Rede	25,886	25,886

Outros experimentos foram realizados com a Rede 3 fazendo-se mudanças em alguns parâmetros de entrada das Tabelas 3 e 4. A Tabela 15 reporta o número de máquinas (M) e a variância da rede ($V(T)$) para a partição gerada pelo algoritmo, e para a opção de não particionar a Rede 3 em cada uma dessas mudanças. Os produtos 9 e 10 foram escolhidos para alteração nos parâmetros de entrada porque, conforme a Tabela 12, o algoritmo pára com uma linha dedicada para os produtos 1 a 8 e outra dedicada aos produtos 9 e 10, e que ao agregar estas duas linhas o número de máquinas não é reduzido. Isso sugere que os produtos 9 e 10 possuem características em comum e que a agregação da linha a qual pertencem com a linha [1,2,3,4,5,6,7,8] é pouco vantajosa. Note que os produtos 9 e 10 possuem tempos médios de serviço $E(s_{kl})$ em geral maiores que os demais produtos, mas coeficientes de variação do processo de chegadas externas ca'_k menores (vide Tabelas 3 e 4). Por isso, com o objetivo de tornar os produtos 9 e 10 mais distintos dos demais produtos, foram consideradas duas modificações: os processos de chegadas externas dos produtos 9 e 10 foram considerados determinísticos (segunda linha da Tabela 15), e os tempos médios de serviço em cada operação foram duplicados (terceira linha da Tabela 15).

Percebe-se que ao tornar o processo de chegadas dos produtos 9 e 10 determinístico, a partição com menor capacidade ($M=40$ máquinas) é composta de duas linhas [1,2,3,4,5,6,7,8],[9,10], ao invés de uma linha única ($M=41$ máquinas), apesar desta diferença não ser muito significativa (menor que 2,5%). A Tabela 16 apresenta os coeficientes de variação do intervalo de tempo entre chegadas das estações (ca_{jp}) e os níveis de utilização das estações (r_{jp}) nas linhas [1,2,3,4,5,6,7,8] e [9,10]. Esta tabela reforça os resultados obtidos para as Redes 1 e 2: é vantajoso particionar a rede se pelo menos uma linha possui baixas variabilidades ca_{jp} nos processos de chegadas e suas estações operam com altos níveis de utilização r_{jp} . Note na Tabela 16 que a linha [9,10] possui estações com baixas variabilidades ca_{jp} e relativamente altas utilizações r_{jp} (exceção para as estações 1, 11 e 13 que possuem baixos r_{jp}).

Tabela 15 – Número de máquinas (M) e variância ($V(T)$) da Rede 3 com $T^*=8,0$ para a partição obtida pelo algoritmo e para a configuração com uma linha única.

Mudanças nos parâmetros de entrada das Tabelas 3 e 4	Partição obtida pelo algoritmo			Linha única [1-10]	
	M	$V(T)$	Partição	M	$V(T)$
Caso base	41	7,908	[1-8], [9,10]	41	6,489
$ca'_9 = 0$ e $ca'_{10} = 0$	40	7,125	[1-8] [9,10]	41	6,292
$2 * E(s_{9,i})$ e $2 * E(s_{10,i})$	62	7,050	[1-8] [9,10]	66	6,824

Tabela 16 – Coeficiente quadrático de variação (ca_{jp}) e nível de utilização (r_{jp}) das 13 estações das linhas [1,2,3,4,5,6,7,8] e [9,10] para o caso com $ca'_9 = 0$ e $ca'_{10} = 0$ na Rede 3 modificada com $T^*=8,0$.

Estação j	Partição [1,2,3,4,5,6,7,8]		Partição [9,10]	
	ca_{j1}	r_{j1}	ca_{j2}	r_{j2}
1	0,354	0,55	0,000	0,23
2	0,370	0,38	0,101	0,64
3	0,346	0,06	0,053	0,74
4	0,338	0,28	0,108	0,66
5	0,349	0,37	0,098	0,75
6	0,348	0,50	0,139	0,72
7	0,349	0,40	0,103	0,70
8	0,353	0,42	0,029	0,70
9	0,350	0,46	0,029	0,92
10	0,335	0,50	0,029	0,96
11	0,375	0,36	0,005	0,16
12	0,354	0,61	0,185	0,68
13	0,355	0,56	0,166	0,52

Ao duplicar os tempos de processamento dos produtos 9 e 10 (terceira linha da Tabela 15), a partição [1,2,3,4,5,6,7,8],[9,10] obtida pelo algoritmo necessita de 62 máquinas, enquanto a linha única [1-10] necessita de 66 máquinas. Portanto, para a Rede 3, acentuar a diferença entre os produtos 9 e 10 e os demais, duplicando os tempos médios de serviço, favoreceu mais o particionamento da instalação do que tornar o processo de chegadas dos produtos 9 e 10 determinístico (segunda linha da Tabela 15). A Tabela 17 mostra os valores de ca_{jp} e r_{jp} para as estações das linhas [1,2,3,4,5,6,7,8], [9,10] e linha única [1-10]. Note na tabela que, novamente, a linha [9,10] possui estações com relativamente baixas variabilidades ca_{jp} e relativamente altas utilizações r_{jp} (exceção para as estações 1 e 11 que possuem baixos r_{jp}).

Tabela 17 – Coeficiente quadrático de variação (ca_{jp}) e nível de utilização (r_{jp}) e nas 13 estações das linhas [1,2,3,4,5,6,7,8] e [9,10] para o caso com $2 * E(s_{9,i})$ e $2 * E(s_{10,i})$ na Rede 3 modificada com $T^*=8,0$.

Estação j	Partição [1,2,3,4,5,6,7,8]		Partição [9,10]		Linha única [1-10]	
	ca_{j1}	r_{j1}	ca_{j2}	r_{j2}	ca_{j1}	r_{j1}
1	0,354	0,55	0,075	0,46	0,298	0,51
2	0,370	0,38	0,217	0,64	0,461	0,68
3	0,346	0,06	0,172	0,74	0,349	0,66
4	0,338	0,28	0,236	0,66	0,388	0,53
5	0,349	0,37	0,256	0,86	0,349	0,71
6	0,348	0,50	0,291	0,72	0,362	0,68
7	0,349	0,40	0,243	0,70	0,349	0,53
8	0,353	0,42	0,127	0,70	0,372	0,54
9	0,350	0,46	0,127	0,61	0,397	0,46
10	0,335	0,50	0,127	0,64	0,476	0,48
11	0,375	0,36	0,238	0,32	0,511	0,52
12	0,354	0,61	0,292	0,68	0,445	0,66
13	0,355	0,56	0,301	0,69	0,355	0,53

Uma observação na Tabela 17 é que, ao agregar as partições [1,2,3,4,5,6,7,8],[9,10] em uma linha única [1-10], as variabilidades das chegadas nas estações ca_{jp} em geral aumentam, uma vez que as variações dos tempos de serviço nas estações cs_{jp} também aumentam com a agregação dos produtos 1-8, cujos tempos médios de serviço $E(s_{ki})$ são bem menores do que os dos produtos 9 e 10. Ora, o aumento dos coeficientes ca_{jp} e cs_{jp} aumenta a variância do tempo de espera em fila $V(Wq_j)$ nas estações j (expressão (6)) que, por sua vez, aumenta a variância do *leadtime* $V(T_k)$ dos produtos k (expressão (5)), implicando em um número maior de máquinas para que a variância da rede $V(T)$ seja menor que T^* . Por isso, para este exemplo, particionar a instalação em duas linhas [1,2,3,4,5,6,7,8],[9,10] requer um número menor de máquinas. Note que a agregação de produtos com tempos de serviços bem diferentes em uma mesma linha aumenta ca_{jp} e cs_{jp} nas estações que, por sua vez, aumenta a variância do *leadtime* e requer mais capacidade para mantê-la a níveis inferiores a T^* . No caso deste exemplo, a partição da instalação é vantajosa se comparada a não particioná-la.

Outros experimentos foram realizados com a Rede 3 empregando outros sistemas de aproximações para o método de decomposição além do descrito em (9)-(11), com o objetivo de avaliar a sensibilidade dos resultados alcançados com o sistema utilizado. Similarmente

aos experimentos com a Rede 2, a aproximação que descreve o processo de desagregação de partidas (11) foi substituída pela aproximação descrita em Bitran & Tirupati (1988), que explora efeitos das interferências entre produtos nas estações. O algoritmo da seção 3 com esta modificação resultou em 42 máquinas, ao invés das 41 máquinas conforme anteriormente (Tabela 12), ou seja, uma diferença menor que 2,5%, o que é aceitável para as decisões envolvidas. Experimentos adicionais ainda foram realizados com a Rede 3 utilizando o algoritmo com outras aproximações da literatura. Em alguns casos, os resultados sugerem certa sensibilidade do algoritmo em relação às aproximações utilizadas. Em parte destes casos, as vantagens de particionar a instalação desapareceu em relação à não particionar. Estes resultados estimulam pesquisa futura, além do escopo do presente trabalho, envolvendo o projeto de experimentos com diferentes configurações de rede e aproximações de decomposição para avaliar e otimizar o seu desempenho. O objetivo seria identificar melhor quais sistemas de aproximação são mais adequados para determinadas configurações de rede.

5. Conclusões

Neste artigo estudaram-se modelos e métodos de solução para problemas da classe *SP3* (partição da instalação), que envolve o projeto de fábrica focalizada. Realizou-se uma pesquisa exploratória inicial acerca das características da rede que sugerem quando é vantajoso ou não particioná-la em subredes. Ao aplicar os modelos e métodos em alguns exemplos testes, os resultados mostraram que, em certos casos, particionar a rede em subredes pode ser mais vantajoso do que não particioná-la do ponto de vista da redução da variância do *leadtime* de produção, e que a partição não necessariamente requer capacidade adicional. Ao contrário, a partição da instalação pode requerer até menos capacidade do que a opção de não particioná-la.

Nos experimentos com as Redes 1 e 2 com dois produtos e um e dois estágios $G/G/m_j$, verificou-se que a partição da rede pode ser vantajosa se pelo menos uma linha possuir baixa variabilidade do processo de chegadas (i.e., próxima de um processo determinístico) e estações com alto nível de utilização (vide Tabelas 5, 9 e 10). Isto também foi observado nos experimentos com a Rede 3 com dez produtos e treze estágios $G/G/m_j$. Além disso, também se observou que a agregação de produtos com tempos de processamento bastante diferentes em uma mesma linha aumenta os valores das variabilidades dos processos de chegada e serviço nas estações que, por sua vez, aumentam a variância do *leadtime* de produção na rede. Portanto, para estes casos, particionar a instalação pode ser mais vantajoso do que não particionar do ponto de vista da redução da variância do *leadtime*. Outros experimentos realizados com variações da Rede 3 mostraram que os resultados podem ser sensíveis ao sistema de aproximações utilizado no método de decomposição para avaliação do desempenho da rede.

Uma perspectiva interessante para pesquisa futura é estudar melhor em projeto de experimentos a relação entre as características de redes de filas abertas e a aplicabilidade dos diversos sistemas de aproximações disponíveis na literatura. O objetivo seria identificar melhor quais sistemas de aproximação são mais adequados para determinadas configurações de rede. Ainda, poder-se-ia comparar os benefícios atingidos ao aplicar o modelo de programação não-linear inteira PRC-1 com os benefícios propostos pela filosofia de produção enxuta que, de certa forma, preconiza a partição da instalação para formação de células de manufatura para minimizar a movimentação desnecessária e facilitar a gestão, reduzindo o inventário em processo, resultados esperados com a redução de complexidade.

Também seria interessante pesquisar outras formas de resolver o modelo PRC-1. Por exemplo, melhorando a heurística construtiva descrita no algoritmo da seção 3, ou utilizando outros métodos tais como heurísticas de busca local e metaheurísticas do tipo algoritmo genético, busca tabu, etc. Além disso, poder-se-ia considerar a adaptação e aplicação dos modelos e métodos de solução aqui empregados em outras áreas além de sistemas de manufatura, tais como redes de serviço (por exemplo, centrais de chamadas), redes computacionais e redes de comunicação.

Agradecimentos

Os autores agradecem aos três revisores anônimos pelos seus úteis comentários e sugestões.

Referências Bibliográficas

- (1) Askin, R. G. & Standridge, C. R. (1993). *Modeling and analysis of manufacturing systems*. John Wiley & Sons, New York.
- (2) Bitran G. R. & Dasu, S. (1992). A review of open queueing network models of manufacturing systems. *Queueing systems*, **12**, 95-134.
- (3) Bitran, G. R. & Morabito, R. (1995). Modelos de otimização de redes de filas abertas para projeto e planejamento de *job-shops*. *Pesquisa Operacional*, **15**(1), 1-22.
- (4) Bitran, G. R. & Morabito, R. (1996). Open queueing networks: Optimization and performance evaluation models for discrete manufacturing systems. *Production and Operations Management*, **5**(2), 163-193.
- (5) Bitran, G. R. & Morabito, R. (1999). An overview of trade-off curves in manufacturing systems design. *Production and Operations Management*, **8**(1), 56-75.
- (6) Bitran, G. R. & Sarkar, D. (1994a). Targeting problems in manufacturing queueing networks – An iterative scheme and convergence. *European Journal of Operational Research*, **76**, 501-510.
- (7) Bitran, G. R. & Sarkar, D. (1994b). Focused factory design: Complexity, capacity and inventory trade-offs. *Technical Memorandum*, AT&T Bell Lab., Holmdel, NJ.
- (8) Bitran, G. R. & Tirupati, D. (1988). Multiproduct queueing networks with deterministic routing: Decomposition approach and the notion of interference. *Management Science*, **34**(1), 75-100.
- (9) Bitran, G. R. & Tirupati, D. (1989). Tradeoff curves, targeting and balancing in manufacturing queueing networks. *Operations Research*, **37**(4), 547-564.
- (10) Boxma, O. J.; Rinnooy Kan, A. & Van Vliet, M. (1990). Machine allocation problems in manufacturing networks. *European Journal of Operational Research*, **45**, 47-54.
- (11) Bretthauer, K. M. (1996). Capacity Planning in manufacturing and computer networks. *European Journal of Operational Research*, **91**, 396-394.
- (12) Buzacott, J. A. & Shanthikumar, J. G. (1993). *Stochastic models of manufacturing systems*. Prentice-Hall, Englewood Cliffs, New Jersey.

- (13) Calabrese, J. M. (1992). Optimal workload allocation in open networks of multiserver queues. *Management Science*, **38**(12), 1792-1802.
- (14) Frenk, H.; Labbe, M.; Van Vliet, M. & Zhang, S. (1994). Improved algorithms for machine allocation in manufacturing systems. *Operations Research*, **42**(3), 523-530.
- (15) Gershwin, S. B. (1994). *Manufacturing systems engineering*. Prentice-Hall, Englewood Cliffs, NJ.
- (16) Govil, M. K. & Fu, M. C. (1999). Queueing theory in manufacturing: A survey. *Journal of Manufacturing Systems*, **18**(3), 214-240.
- (17) Kerbache, L. & Smith, J. M. (2004). Queueing networks and the topological design of supply chain systems. *International Journal of Production Economics*, **91**, 251-272.
- (18) Kleinrock, L. (1975). *Queueing systems, Volume 1*. John Wiley & Sons, New York. 417pg.
- (19) Krajewski, L. J. & Ritzman, L. P. (1996). *Operations management: strategy and analysis*. 4^a ed. Addison-Wesley Publishing Company, Massachusetts.
- (20) Kouvelis, P. & Tirupati, D. (1991). Approximate performance modeling and decision making for manufacturing systems: A queueing network optimization framework. *Journal of Intelligent Manufacturing*, **2**, 107-134.
- (21) Kouvelis, P.; Chambers, C. & Yu, D. Z. (2005). Manufacturing operations manuscripts published in the first 52 issues of POM: Review, trends and opportunities. *Production and Operations Management*, **14**(4), 450-467.
- (22) Larson, R.C. & Odoni, A.R. (1981). *Urban operations research*. Prentice-Hall, Englewood Cliffs, New Jersey.
- (23) Morabito, R. (1998). Análise de curvas de trade-off baseada em teoria de redes de filas para o projeto e planejamento de sistemas discretos de manufatura. São Carlos, tese de livre-docência, EESC/USP, 136pg.
- (24) Nakano, M. & Ohno, K. (1999). Decomposition algorithm for performance evaluation of AGV systems. *Production and Operations Management*, **8**(2), 193-205.
- (25) Segal, M. & Whitt, W. (1989). A queueing network analyzer for manufacturing. **In**: *Teletraffic science for new-effective system, networks and services*, ITC-12 [edited by M. Bonatti], Elsevier, North-Holland, Amsterdam, 1146-1152.
- (26) Seshadri, S. & Pinedo, M. (1999). Optimal allocation of resources in a job shop environment. *IIE Transactions*, **31**, 195-206.
- (27) Shanthikumar, J. G. & Buzacott, J. A. (1981). Open queueing network models of dynamic job shops. *International Journal of Production Research*, **19**(3), 255-266.
- (28) Shanthikumar, J. G. & Xu, S. H. (2000). Strongly asymptotically optimal design and control of production and service systems. *IIE Transactions*, **32**, 881-890.
- (29) Silva, C. R. N. (2005). Aplicação de modelos de redes de filas abertas no projeto e planejamento de sistemas discretos de manufatura. São Carlos, tese de doutorado, PPG-EP/UFSCar, 272pg.

- (30) Silva, C. R. N. & Morabito, R. (2007). Aplicação de modelos de redes de filas abertas no planejamento do sistema *job-shop* de uma planta metal-mecânica. *Gestão & Produção*, **14**(2), 393-410.
- (31) Skinner, W. (1974). The focused factory. *Harvard Business Review*, May-June, 113-121.
- (32) Souza, G. S.; Wagner, H. M. & Whybark, D. C. (2001). Evaluating focused factory benefits with queueing theory. *European Journal of Operational Research*, **128**, 597-610.
- (33) Souza, G. S.; Ketzenberg, M. E. & Guide, V. D. R. (2002). Capacitated remanufacturing with service level constraints. *Production and Operations Management*, **11**(2), 231-248.
- (34) Souza, G. S. & Ketzenberg, M. E. (2002). Two-stage make-to-order remanufacturing with service-level constraints. *International Journal of Production Research*, **40**(2), 477-493.
- (35) Sundarraj, R. P.; Sundararaghavan, P. S. & Fox, D. R. (1994). Optimal server acquisition in open queueing networks. *Journal of the Operational Research Society*, **45**(5), 549-558.
- (36) Suri, R.; Sanders, J. L. & Kamath, M. (1993). Performance evaluation of production networks. **In:** *Handbooks in OR/MS* [edited by S. C. Graves], vol.4, Elsevier, North-Holland, Amsterdam.
- (37) Tijms, H. C. (1986). *Stochastic modeling and analysis: A computational approach*. John Wiley & Sons, New York.
- (38) Van Vliet, M. & Rinnooy Kan, A. (1991). Machine Allocation Algorithms for Job-Shop Manufacturing. *Journal of Intelligent Manufacturing*, **2**, 83-94
- (39) Warsing, D. P.; Souza, G. C. & Greis, N. P. (2001). Determining the value of dedicated multimodal cargo facilities in a multi-region distribution network. *European Journal of Operational Research*, **133**, 81-93.
- (40) Whitt, W. (1983a). The queueing network analyzer. *The Bell System Technical Journal*, **62**(9), 2779-2815.
- (41) Whitt, W. (1983b). The queueing network analyzer. *The Bell System Technical Journal*, **63**(9), 1911-1979.
- (42) Whitt, W. (1994). Towards better multi-class parametric-decomposition approximations for open queueing networks. *Annals of Operations Research*, **48**, 221-248.
- (43) Whitt, W. (1995). Variability functions for parametric decomposition approximations of queueing networks. *Management Science*, **41**(10), 1704-1715.