Recebido em: 11/09/11
Aprovado em: 11/09/11

Uma proposta para a detecção do adequado período de aquecimento em modelos de simulação

Hélida Gomes Norato (UFMG – MG/Brasil) - helidanorato@gmail.com
• Av. Antônio Carlos, 6627, Pampulha, 31270-901, Belo Horizonte-MG, fone (55) 31-3409-3487
Anderson Ribeiro Duarte (UFOP – MG/Brasil) - anderson@iceb.ufop.br

Resumo
A seleção de um período de aquecimento adequado para uma simulação de eventos discretos ainda é uma escolha de difícil determinação. Alguns métodos já foram desenvolvidos e serão brevemente revistos, aqui. Ainda não existem critérios definitivos para a escolha de um método específico dentre os existentes. Buscamos discutir alguns dos métodos existentes e apresentar um método que se adapte bem à estrutura do modelo em estudo. Analisaremos o desempenho dos métodos expostos, através de um procedimento avaliativo para comparação. Os métodos serão aplicados para as saídas geradas por um modelo simples de trabalho com diferentes níveis de utilização. Algumas métricas serão utilizadas para averiguação da eficiência de cada um dos métodos, em comparação com a nova proposta implementada.
Palavras-chave: Simulação; estado transiente; estado permanente; aquecimento.

Abstract
The selection of a suitable warm-up period for a discrete-event simulation is still a difficult choice to determine. Some methods have been developed and will be briefly reviewed here. There are still no definitive criteria for choosing a particular method among the existing ones. We discuss and show some existing methods and present a method that suits the structure of the model studied. We will analyze the performance of the exposed methods by an assessment procedure for comparison. The methods are applied to the output generated by a simple work model with different levels of utilization. Some metrics will be used to investigate the efficiency of each method compared to the new proposal implemented.
Keywords: Simulation; transient state; steady state; warm-up.

1. INTRODUÇÃO

Em estudos que utilizam técnicas de simulação, exige-se um grande esforço de tempo computacional e elevados esforços no desenvolvimento e programação dos modelos. Por outro lado, em geral, é dedicado apenas um pequeno esforço para a avaliação de dados de saída de simulação, de forma a promover uma análise adequada dos resultados. Uma rodada de simulação de sistemas não terminais (aqueles que funcionam intermitentemente) é composta por um tempo de aquecimento necessário à estabilização do sistema, seguida de diversas replicações, cada uma delas fornecendo dados para a realização de estatísticas referentes ao comportamento do modelo simulado. Em procedimentos mais rústicos, é comum se deparar com uma simulação simplista, em que o tempo de aquecimento e o número de replicações são escolhidos de forma arbitrária e, em seguida, as estimativas resultantes do procedimento de simulação são observadas como as características do modelo em estudo. Considerando que os modelos de simulação a eventos discretos são conduzidos ao longo do tempo e através de amostras aleatórias de distribuições de probabilidade, podemos observar que as estimativas são apenas realizações particulares de variáveis aleatórias, que podem ter muita variabilidade.

Para duas replicações distintas, em uma simulação, é possível observar diferentes características entre elas; portanto, ao menos uma delas não deve estar representando a realidade, a qual o modelo tenta reproduzir. Assim, poderia haver uma probabilidade significativa de fazer inferências errôneas sobre o sistema em estudo. Um impedimento para a obtenção de estimativas precisas dos parâmetros ou características de um sistema modelado através de simulação, é o tempo computacional necessário para que o sistema simulado se estabilize. Isto é necessário para tornar possível coletar a quantidade necessária de dados da simulação para estudos inferenciais. Esta dificuldade ocorre muitas vezes, na simulação, em problemas de médio e grande porte. Para promover melhorias nos procedimentos de coleta de dados de simulações, visando estudos inferenciais, é de extrema importância a adoção de métodos para seleção do período de aquecimento (*warm-up*). Por período de aquecimento denominamos a detecção do período transiente inicial, em que os dados da simulação apresentam elevada variabilidade quando, comparados aos dados do período permanente.

Denominamos esse período permanente por estado estacionário. Neste período, os dados são capazes de produzir estimativas mais estáveis para os parâmetros de interesse. Dado um conjunto de realizações para um parâmetro de interesse $Y_1, Y_2,\ldots$. Seja $F_i(y \mid I_n) = P(Y_i \leq y \mid I_n)$ para i=1, 2,..., em que y é um número real e I_n representa as condições iniciais usuais para iniciar a simulação no tempo 0 (A probabilidade condicional $P(Y_i \leq y \mid I_n)$ é a probabilidade de ocorrer o evento $[Y_i \leq y]$, dadas as condições iniciais I_n). Para um sistema em estudo, I_n deve especificar o número de componentes presentes e se o servidor está ocupado ou não no tempo 0. Denotamos $F_i(y \mid I_n)$ por distribuição transiente do processo no tempo discreto i para condições iniciais I_n. Note que $F_i(y \mid I_n)$ deverá ser diferente para cada valor i e para cada conjunto de condições iniciais I_n. Para y e I_n fixos, as probabilidades $F_1(y \mid I_n)$, $F_2(y \mid I_n)$,... formam uma sequência numérica. Se $F_i(y \mid I_n) \rightarrow F(y)$ quando $i \rightarrow \infty$ para todo y e para alguma condição inicial I_n, então F(y) é chamada de distribuição estacionária para o processo estocástico $Y_1, Y_2,\ldots$. Estritamente, a distribuição estacionária F(y) é somente obtida no limite quando $i \rightarrow \infty$. Na prática, porém, muitas vezes existe um índice de tempo finito L tal que as distribuições a partir deste ponto sejam aproximadamente iguais, ou seja, "estado estacionário". Note que o estado estacionário não significa que as variáveis aleatórias $Y_{L+1}, Y_{L+2},\ldots$ tendem a assumir sempre os mesmos valores em uma simulação particular, mas sim, significa que os valores têm aproximadamente a mesma distribuição.

O problema de detecção do período de aquecimento adequado tem sido discutido na literatura de simulação há muito tempo. Uma escolha adequada é importante para evitar a presença de vício nas estimativas de uma simulação, quando em seu estado estacionário. Grande parte das abordagens sobre este assunto sugere a execução dos experimentos de simulação para um dado período de aquecimento e a exclusão destes dados iniciais. Entretanto, uma questão óbvia surge: "Como identificar o período adequado de aquecimento?"

Este trabalho tem por objetivos discutir algumas técnicas já existentes, para a detecção de períodos de aquecimento; estabelecer métricas de comparação entre técnicas; construir um modelo experimental de simulação para comparar técnicas e, por fim, propor uma nova metodologia para detecção de períodos de aquecimento em modelos de simulação. A seção 2 apresenta uma revisão bibliográfica acerca do assunto em estudo; a seção 3 apresenta o modelo experimental que será utilizado e descreve os métodos para detecção de períodos de aquecimento, que serão utilizados neste estudo; a seção 4 introduz uma nova técnica para detecção do período de aquecimento; avaliações numéricas comparativas entre os métodos são apresentadas na seção 5. Conclusões e propostas futuras são apresentadas na seção 6.

2. REVISÃO BIBLIOGRÁFICA

Pode-se encontrar na literatura associada à área de simulação, uma variedade de métodos para a detecção do período de aquecimento. Este fato reforça a importância do tratamento adequado do período transiente, para que este período não influencie sobremaneira as estimativas dos parâmetros de interesse no procedimento de simulação, quando este atingir o seu estado estacionário. Estes métodos possuem características que podem ser adequadas a diversos tipos de modelos de simulação, considerando as especificidades associadas a cada modelo. Em geral, duas abordagens são utilizadas: a exclusão de dados e a inicialização inteligente (NELSON, 2001). Na primeira abordagem, o modelo é executado por um período de aquecimento até atingir um estado estacionário e, antes do qual, os dados são eliminados. A maior dificuldade presente na utilização desta abordagem está em escolher a duração do período de aquecimento. Se a escolha for subestimada, haverá algum vício na estimação de variáveis de interesse no procedimento de simulação. Por outro lado, se a escolha for superestimada, alguns dados relevantes para o procedimento de estimação serão excluídos, fazendo com que os tempos de simulação se tornem elevados desnecessariamente. Na segunda abordagem, o modelo é colocado em uma condição "realista" no início da execução. O desafio é determinar o que constitui um estado realista para se iniciar a simulação. Em princípio, essas abordagens levam ao mesmo objetivo: colocar o modelo em uma condição realista, a partir do momento em que os dados de saída serão coletados. Este trabalho se concentra na primeira abordagem, ou seja, com o objetivo de determinar um período adequado de aquecimento para os modelos de simulação. Os métodos para detecção do período de aquecimento podem ser classificados em cinco grupos: Gráficos, Heurísticos, Estatísticos, Testes de vício de inicialização e Híbridos. Uma descrição de alguns destes métodos, conforme sua classificação, é um instrumento importante para a seleção adequada da metodologia, para detecção do período de aquecimento em modelos de simulação.

Os métodos Gráficos dependem de uma análise visual de séries temporais, obtidas das saídas de simulação. Em geral, se baseiam na divisão de dados em grupos, para cálculo de médias móveis. Dentre estes métodos, podemos destacar as propostas apresentadas em Robinson (2002) e Welch (1983), que se baseiam em séries temporais simples. E ainda métodos baseados em gráficos de médias cumulativas, vistos em Banks *et al.* (2001), Gordon (2001) e em Nelson (2001). São métodos que têm por vantagem a sua simplicidade e a "quase" ausência de fortes suposições acerca dos dados de saída da simulação. Entretanto, existe uma forte dependência da avaliação do analista (decisão humana), que pode levar a diferentes decisões, de acordo com a experiência do analista.

As técnicas heurísticas buscam estabelecer um conjunto de regras para a determinação da duração do período de aquecimento. Entre os métodos heurísticos, podemos citar a Regra de Conway (CONWAY, 1963), a Regra de Conway modificada (GAFARIAN *et al.*, 1978), a Regra do estimador de auto-correlação (FISHMAN, 1971), a Regra do cruzamento de médias (FISHMAN, 1973), as Regras de erros marginais MSER e MSER-n (WHITE, 1997; SPRATT, 1998). O uso de regras específicas tem por vantagem o fato de remover a subjetividade dos métodos gráficos. Por outro lado, isto pode significar que alguns padrões importantes podem não ser identificados, especialmente se uma representação gráfica dos dados não for utilizada em associação à regra em uso.

Os métodos estatísticos são técnicas que se baseiam em princípios estatísticos para o procedimento de estimação do período de aquecimento. Podemos citar o método proposto em Kelton e Law (1983), que executa um procedimento de regressão dos dados avaliados na sequência inversa de tempo, buscando até que ponto a inclinação estimada pelo procedimento de regressão seja nula. O teste de Aleatorização, proposto em Yücesan (1993), avalia a diferença entre duas médias para os dados, divididos em dois subgrupos, ao longo do tempo de execução da simulação. Trata-se de métodos mais complexos que as abordagens anteriores, dada a exigência de cálculos extensivos, bem como a validação de algumas suposições iniciais.

Os testes de vício de inicialização não são testes para determinação do período adequado de aquecimento, mas sim, testes para verificação da presença ou não de vício que seja associado à inicialização na série de dados em estudo. Como estratégia para a detecção do período adequado de aquecimento, estes testes devem ser utilizados em conjunto com algum dos métodos descritos anteriormente. Dentre possíveis formatos, podemos mencionar o Teste do Máximo (SCHRUBEN, 1982), o Teste de Schruben modificado (SCHRUBEN, 1982), o Teste Ótimo (SCHRUBEN *et al.*, 1983), o Teste do Posto (VASSILACOPOULOS, 1989). Uma família de testes desta natureza, considerando médias por subgrupos, é apresentada em Goldsman *et al.* (1994). Uma vantagem nestes métodos é a possibilidade do analista avaliar se o vício de inicialização foi completamente removido. Nestes métodos, a complexidade dos cálculos envolvidos é baixa. Por outro lado, são métodos computacionalmente instáveis.

Os métodos híbridos empregam os testes para vício de inicialização, associados aos testes gráficos ou heurísticos. Dois dos principais métodos são: o método sequencial, visto em Pawlikowski (1990) e o método da escala invariante, proposto por Jackway e DeSilva (1992). São testes mais complexos que exigem a utilização de uma análise espectral de variância. O método proposto em Jackway e DeSilva (1992) exige uma estimativa da variância e pode precisar de uma grande quantidade de dados, para um bom desempenho, visto que é baseado em teoria assintótica.

3. MATERIAIS E MÉTODOS

Diversos testes já foram realizados sobre os métodos anteriormente descritos, obtendo-se estimativas para o período de aquecimento, com diferentes graus de sucesso. Em muitos casos, isto se deve a uma dependência da natureza dos dados analisados. Não se pode afirmar sobre a existência de um método, que possa ser recomendado acima de quaisquer outros. De fato, a literatura sugere a não existência de métodos de uso comum. Apesar da grande variedade de abordagens que têm sido propostas ao longo dos últimos anos, existe uma necessidade tanto de testar estas abordagens mais profundamente como de elaborar novos métodos, que poderiam ser adotados amplamente por usuários da simulação.

Considerando alguns métodos de grande utilização, o trabalho em Mahajan e Ingalls (2004) apresenta uma comparação entre métodos para a detecção do período de aquecimento. Este trabalho utiliza um modelo experimental e algumas medidas de desempenho, para avaliação da qualidade de alguns métodos de vasta utilização para detecção de período de aquecimento. Estas análises não são completamente conclusivas; algumas deficiências entre os métodos são verificadas; entretanto, não é apresentada uma nova proposta mais adequada para a detecção do período de aquecimento. Utilizamos este modelo experimental como referência para a construção de um novo modelo, para avaliar os métodos comparados e também, o novo método que será proposto neste trabalho.

3.1. Modelo Experimental

No modelo experimental, os métodos foram testados em modelos simples de trabalho. São cinco procedimentos em série; cada procedimento tem diferentes números de máquinas (recursos) e existem três tipos distintos de peças a serem produzidas. As chegadas são distribuídas conforme Poisson (λ) e os tempos de serviço para os procedimentos são exponenciais com médias $\mu i,j$, em que i representa a peça e j, o procedimento. O modelo foi construído e executado através do *software* Arena. As peças que chegam são divididas nos três tipos, com probabilidades: 0,5; 0,3 e 0,2, respectivamente. As peças são processadas em tempos de serviço diferentes. Após o tratamento, as peças saem do sistema. Os métodos foram aplicados ao modelo com diferentes níveis de utilização, como segue:

- Tipo I: Modelo tem um alto nível de utilização. A utilização média de todos os recursos, esta próxima de 90%, com utilizações individuais entre 80% e 95%;
- Tipo II: Modelo tem um nível moderado de utilização. A utilização média de recursos, esta próxima de 70%, com utilizações individuais entre 65% e 80%;
- Tipo III: Modelo tem um baixo nível de utilização. A utilização média de recursos, esta próxima de 50%, com utilizações individuais entre 45% e 65%;

Quanto às condições de execução da simulação, o sistema foi iniciado vazio e ocioso. Foi considerado um tamanho inicial para cada simulação de 5000 horas. O parâmetro de interesse θ avaliado foi o tempo médio de permanência das entidades no sistema. Trata-se de um parâmetro de desempenho escolhido de forma arbitrária, com intuito apenas de testar os métodos. Para outros parâmetros, a análise pode revelar resultados diferentes. Com o interesse de estabelecer uma estratégia comparativa e expor uma nova técnica, apresentaremos, de forma suscinta cinco métodos que serão comparados à nova proposta.

3.2. Método de Welch

Trata-se de uma técnica bastante simples para a determinação do adequado período de aquecimento. Esta técnica se inclui no grupo das técnicas gráficas. Podemos descrever o Método de Welch (WELCH, 1983), através dos seguintes passos:

1. Execute n replicações do procedimento de simulação, cada uma destas replicações terá tamanho m. Seja Y_{ji} a i-ésima observação da variável de interesse na j-ésima replicação, neste caso i=1,2,...,m e j = 1, 2,..., n;

2. Calcule as médias de cada uma das observações entre as replicações; estas médias serão denotadas por $\overline{Y}_i$, em que: $\overline{Y}_i = \sum_{j=1}^{n} \frac{Y_{ji}}{n}$ para i=1, 2,..., m;

3. Defina a média móvel $\overline{Y}_i(w)$ para suavizar as oscilações de alta frequência em $\overline{Y}_1$, $\overline{Y}_2$,..., $\overline{Y}_m$, neste caso, w será a medida de janela que é um inteiro positivo, a proposta utiliza $w \leq \frac{m}{4}$. A média móvel $\overline{Y}_i(w)$ será definida por:

$$\overline{Y}_i = \frac{\sum_{s}^{w} = -w\overline{Y}_{i+S}}{2w+1} \quad \text{para i} = w+1, w+2,..., m-w; \qquad (1)$$

4. Plote os valores $\overline{Y}_i(w)$ para $= w+1, w+2,..., m-w$ e escolha L como o valor i, no qual $\overline{Y}_i(w)$ parece convergir. Detalhes mais profundos sobre a convergência podem ser obtidos em Welch (1983).

3.3. Método baseado no Controle estatístico de processo

O método *Statistical Process Control* – SPC, proposto em Robinson (2002), é outra técnica gráfica, que necessita de múltiplas replicações, podendo ser descrito em quatro etapas:

1. Defina Y_{ji} e as médias $\overline{Y}_i$ para i = 1,2,...,m, como no método anterior;

2. Agora, as médias serão agrupadas em b grupos de tamanho k e representaremos as médias por grupo como $\overline{\overline{Y}}_x$ para $i = 1,2,...,b$. O tamanho destes grupos é selecionado de forma a garantir que as médias por grupo sejam aceitas no teste de aderência para normalidade de *Anderson-Darling* e no teste de correlação de *Von Neuman*. Recomenda-se, inclusive, um mínimo de 20 grupos. Após a divisão em grupos, denotamos a série temporal resultante por:

$$Y_{(k)} = \{\overline{\overline{Y}}_1(k), \overline{\overline{Y}}_1(k), \ldots, \overline{\overline{Y}}_b(k)\} \quad (2)$$

3. Será gerada então, uma carta de controle para estas séries temporais. Consideram-se as estimativas de média populacional μ e desvio padrão σ, obtidas da última metade da série $Y_{(k)}$. De posse destas estimativas, os seguintes limites de controle LC são obtidos: LC= $\hat{\mu}\hat{\sigma} \pm z\frac{\hat{\sigma}}{\sqrt{\frac{b}{2}}}$ para z = 1,2,3;

4. A carta de controle é construída com os três limites de controle. São seguidas as seguintes regras para a verificação de saída de controle:

a. Existência de pontos fora do limite de controle de 3σ;
b. Dois pontos fora do limite de controle de 2σ de três consecutivos;
c. Quatro pontos fora do limite de controle de 1σ de cinco consecutivos;
d. Existência de oito pontos consecutivos acima da média ou então, oito pontos consecutivos abaixo da média.

Entende-se que o processo atingiu o equilíbrio, quando está em controle.

3.4. Teste de Aleatorização

Um método estatístico baseado em testes de Aleatorização é apresentado em Yücesan (1993). O problema foi formulado como um conjunto de testes de hipóteses acerca da média do processo. São executados testes de Aleatorização, visando testar a validade da hipótese nula de que a média se mantém sem alteração durante toda a simulação. Uma grande vantagem sobre este método seria a possibilidade de relaxar a hipótese de normalidade. O método pode ser descrito, através dos seguintes passos:

1. Execute a simulação, considerando um tamanho pré-estabelecido de m horas. Defina Y_i como a i-ésima observação da variável de interesse na simulação;

2. Divida a série $Y_i, \ldots, Y_m$ em b grupos de tamanho k e obtenha médias $\overline{Y}_1, \ldots, \overline{Y}_b$;

3. Particione o conjunto de médias em dois conjuntos C_1 e C_2. Para a primeira iteração, o primeiro conjunto conterá apenas $\overline{Y}_1$ e o segundo conjunto, as outras $b - 1$ médias. Obtenha as médias $\overline{Y}_{C_1}$ e $\overline{Y}_{C_2}$ dentro dos dois conjuntos;

4. Para cada iteração, as médias $\overline{Y}_{C_1}$ e $\overline{Y}_{C_2}$ são comparadas. Se existir diferença estatisticamente significativa entre elas, então a hipótese nula é rejeitada;

5. Se a hipótese nula for rejeitada, os dois conjuntos C_1 e C_2 são modificados; é acrescentada a média $\overline{Y}_2$ ao conjunto C_1 enquanto o conjunto C_2 passa a ser composto por $b - 2$ elementos. O teste é executado novamente, agora para os novos conjuntos C_1 e C_2;

6. Enquanto a hipótese nula estiver sendo rejeitada, o conjunto C_1 é aumentado como no passo anterior. Quando a hipótese nula for aceita pela primeira vez, entende-se que o grupo C_2 é composto pelas observações do estado estacionário.

3.5. Regra do cruzamento de médias

Esta regra é um procedimento heurístico, apresentado em Fishman (1973), para a determinação do período de aquecimento adequado. O objetivo é obter a quantidade de vezes em que o valor para a variável de interesse saiu de um valor menor que a média e foi para um valor maior que a média, em apenas um passo e, também, quando a variável de interesse saiu de um valor maior que a média e foi para um valor menor que a média, em apenas um passo. Se esta contagem atinge um valor pré-estabelecido l, então se assume que o processo se tornou estacionário. Entretanto, este valor pré-estabelecido é uma escolha do analista. De uma forma mais prática, podemos descrever o procedimento, através dos seguintes passos:

1. Execute a simulação, considerando um tamanho pré-estabelecido de m horas. Defina Y_i como a i-ésima observação da variável de interesse na simulação;

2. Obtenha a série temporal $Y_1, Y_2, \ldots, Y_m$ e a média $\overline{Y}_m = \dfrac{\sum_{i=1}^{m} Y_i}{m}$;

3. Para $i = 1, 2, \ldots, m - 1$ defina w_i tal que $w_i = \begin{cases} 1 & se \quad Y_{i-1} < \overline{Y}_m < Y_i \\ 1 & se \quad Y_i < \overline{Y}_m < Y_{i-1} \\ 0 & caso\ contrário \end{cases}$;

4. O número de vezes que a série cruza a média, será dado por $\Omega_m = \sum_{i=1}^{m-1} w_i$;

5. Calcule $\Omega_1, \Omega_2, \ldots, \Omega_k$ até algum k, tal que Ω_k atinja o valor pré-estabelecido l;

3.6. Regra de erros marginais (MSER)

Trata-se de um procedimento heurístico, apresentado por White (1997), denominado (MSER) e uma versão modificada (MSER-n) apresentada em Spratt (1998). O método MSER baseia-se em obter o período de aquecimento adequado d^* que melhor equilibra o balanço entre a ausência de vício (alta precisão) e redução do tamanho da amostra (baixa precisão), considerando a série $Y_{d^*+1}, Y_{d^*+1}, \cdots, Y_m$ obtida de uma execução de tamanho m com as observações $Y_1, Y_2, \cdots, Y_m$. Os dois casos são métodos que escolhem um ponto de truncamento d^*, que minimiza a largura do intervalo de confiança marginal sobre a média da amostra truncada (note que dada a série truncada de observações, existe uma correlação sequencial, portanto o intervalo de confiança marginal não é uma estimativa adequada para a média truncada. Este intervalo está sendo usado, aqui, como uma medida de homogeneidade para a série truncada que será analisada). Dada uma série de saída finita $Y_1, Y_2, \cdots, Y_m$, o ponto de truncamento ideal para a sequência é dado por:

$$d^* = argmin \left[\frac{1}{(m-d)^2} \sum_{i=d+1}^{m} (Y_i - \overline{Y}_{m,d})^2 \right], \text{ em que } \overline{Y}_{m,d} = \frac{\sum_{i=d+1}^{m} Y_i}{(m-d)} \quad (3)$$

O método MSER-n considera novamente a equação anterior, entretanto com as observações divididas em b grupos de tamanho k, são consideradas as b médias dos grupos.

3.7. Uma avaliação inicial sobre os métodos

Apesar dos métodos buscarem um objetivo comum, as abordagens podem apresentar grande diferença de formulação. Existem técnicas mais complexas e algumas mais simplistas. Como já citado anteriormente, é bastante comum se deparar com analistas que optam por escolhas de período de aquecimento de forma completamente arbitrária. Acreditamos que este fato se deve a dois motivos principais: em primeiro lugar, algumas regras são bastante complexas, podendo, inclusive, requerer um tempo computacional significativo para sua utilização; em segundo lugar, a grande maioria destes procedimentos exige do analista a escolha de alguns parâmetros para seu funcionamento e estas escolhas levam novamente a alguma arbitrariedade indesejada. Baseado nestas considerações, será feita uma análise preliminar sobre as técnicas até aqui descritas.

O método de Welch (WELCH, 1983) necessita da definição do parâmetro w, associado ao comprimento da janela; espera-se que este parâmetro não influencie os resultados de forma significativa. Por outro lado, definir de forma específica um critério de convergência para a série das médias, é um problema pontual da utilização deste método. Modificações na escolha destes critérios afetam, de forma significativa, a solução proposta através desta técnica.

O método baseado no Controle estatístico de processo (ROBINSON, 2002) parece bastante justificável, entretanto pode se tornar difícil definir a quantidade de grupos que garantiria sucesso em testes de normalidade e correlação. Em particular, pode ser necessária a execução de um tempo maior de simulação até que se consiga adequar a amostra simulada e a divisão em grupos para os referidos testes.

O Teste de Aleatorização (YÜCESAN, 1993) necessita da escolha da quantidade *b* de grupos, na qual a amostra será dividida. Em particular, para este método, uma variação na quantidade de grupos pode alterar de forma significativa a conclusão do teste. Vale ressaltar que, se o comprimento proposto inicialmente para cada grupo for superior ao adequado período de aquecimento para os dados, automaticamente a solução apresentada será superestimada, em relação a este adequado período de aquecimento.

A Regra do cruzamento das médias (FISHMAN, 1973) é bastante dependente do parâmetro *l* a ser escolhido pelo analista. O valor *l* determina quantos cruzamentos da média, em apenas um passo, se deve observar para que se considere ter atingido a estacionariedade.

A Regra dos erros marginais (WHITE, 1997) é um procedimento que na sua forma original se torna livre da escolha de parâmetros pelo analista. Entretanto, tende a subestimar o adequado período de aquecimento. Este fato é uma das principais razões para a postulação da técnica MSER-n (SPRATT, 1998), dividindo em *n* grupos. Entretanto, voltamos ao problema da subjetividade da escolha do valor *n*. Novamente, se trata de um parâmetro que pode afetar sobremaneira as conclusões do método.

Desta análise preliminar, concluímos pela necessidade de um método simples quanto a sua utilização, robusto quanto a sua eficiência ao estimar o comprimento adequado para o período de aquecimento e livre da escolha de parâmetros pelo analista, ou pelo menos, muito pouco dependente de parâmetros que devam ser escolhidos previamente.

4. UMA NOVA METODOLOGIA PARA DETECÇÃO DO COMPRIMENTO DE PERÍODOS DE AQUECIMENTO

Dentre os métodos aqui analisados, e mesmo considerando outros métodos para o mesmo fim, nota-se que, em geral, os métodos se baseiam em observar o comportamento dos dados, em relação à média. De fato, ao considerarmos um conjunto de replicações e a série temporal das médias entre replicações, seria bastante razoável admitir que ao atingirmos uma "quase" estacionariedade, o lastro de dependência entre os termos da série temporal das médias seria pequeno o suficiente para fazer valer o Teorema Central do Limite. Não buscamos contradizer este caminho que já foi seguido pela maioria dos métodos. Também seguiremos este raciocínio, entretanto, através de uma formulação simples e evitando a dependência da escolha de parâmetros pelo analista.

4.1. Método do cruzamento das médias iterado

O método que será proposto avalia os cruzamentos ocorridos em um passo, em relação à média, mas não da forma original proposta no método do cruzamento de médias. Este procedimento será feito, agora, de uma forma iterada. Para tanto, considere *n* replicações do procedimento de simulação, cada uma delas de tamanho *m*. Defina as médias das observações entre replicações por $\overline{Y}_i$ com

$i = 1, 2, \ldots, m$. Denote por $\overline{\overline{Y}}_j$ a média das últimas $m - (j - 1)$ médias entre replicações, neste caso $\overline{\overline{Y}}_j = \frac{\sum_{i=j}^{m} \overline{Y}_i}{m - (j-1)}$. É fácil ver que para os valores $\overline{Y}_i$ com $i = 1, 2, \ldots, m$, algumas das observações iniciais são pertencentes ao período transiente. Em geral, todas essas observações iniciais são menores ou então, são todas maiores que $\overline{\overline{Y}}_1$. O objetivo é descartar estas observações de uma forma iterada, como veremos a seguir, através dos seguintes passos:

1. Execute n replicações da simulação, considerando um tamanho pré-estabelecido m para as replicações e obtenha os valores $\overline{Y}_i$ para $i = 1, 2, \ldots, m$ e o valor $\overline{\overline{Y}}_1$;

2. Verifique se $\overline{Y}_1$ é inferior ou superior à $\overline{\overline{Y}}_1$. Se inferior, determine o menor k, tal que $\overline{Y}_{k+1}$ seja superior a $\overline{\overline{Y}}_1$ e descarte as k primeiras observações, o procedimento é análogo para $\overline{Y}_1$ superior a $\overline{\overline{Y}}_1$;

3. Considere agora a média das médias $\overline{\overline{Y}}_k$ e a nova série $\overline{Y}_i$ para os valores $i = k + 1, k + 2, \ldots, m$, repita o procedimento até que o número de valores em sequência, que sejam inferiores ou então superiores à média das médias, seja limitado por 1, ou seja, considerando a média das médias dos últimos valores, é possível cruzar a média em um passo. Neste momento, o método atingiu sua estimativa para o comprimento do período de aquecimento.

O método aqui exposto, tem uma formulação bastante simples, sem a dependência de escolha de parâmetros pelo analista. A eficiência de seus resultados poderá ser verificada através dos resultados apresentados na seção de Avaliações numéricas.

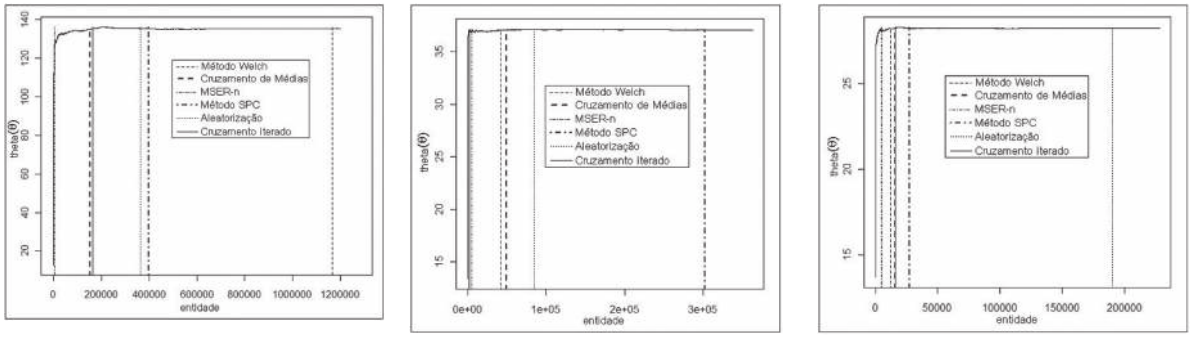
5. ANÁLISE DE DADOS E RESULTADOS

Os cinco métodos descritos e o novo método proposto foram avaliados através de execuções do modelo experimental descrito anteriormente. Em particular, a métrica de comparação utilizada foi o erro quadrático médio para o estimador do parâmetro de interesse θ, dado por $EQM(\hat{\theta}) = [\mathbb{E}(\theta - \hat{\theta})]^2 + Var(\hat{\theta})$. Obviamente, o valor θ é desconhecido; para tanto, utilizamos uma estimativa deste parâmetro, considerando um período de aquecimento que descarta a metade inicial da série, ou seja, uma escolha completamente conservadora, para que a análise comparativa fosse justa e adequada. Já os valores $\hat{\theta}$ são obtidos considerando a média das realizações, descartando o período de aquecimento sugerido por cada um dos métodos. Vale ressaltar que em situações nas quais o erro quadrático médio apresenta valores muito próximos de zero, a comparação através desta métrica poderia não ser suficientemente conclusiva; para tanto, utilizamos uma segunda métrica, a variação relativa do erro quadrático médio da estimativa $\hat{\theta}$, fornecida por cada método, em relação ao erro quadrático médio, considerando a estimativa $\hat{\theta}_{ini}$, que leva em conta todas as observações da simulação. Definiremos esta métrica por:

$$V_r(EQM(\hat{\theta})) = \frac{EQM(\hat{\theta}) - EQM(\hat{\theta}_{ini})}{EQM(\hat{\theta}_{ini})}.$$

A simulação foi executada considerando que o sistema foi iniciado vazio e ocioso. Foram executadas 30 replicações, cada uma com 5000 horas de simulação para cada nível de utilização do modelo experimental. O parâmetro θ avaliado foi o tempo médio de permanência das entidades no sistema. A avaliação foi conduzida considerando os diferentes tipos de utilização para o modelo experimental e também, as diferentes técnicas para detecção do comprimento do período de aquecimento. Denotaremos por L, o comprimento detectado por cada uma das técnicas em estudo.

Figura 1 – Apresentação dos limites obtidos por cada um dos métodos com a utilização do modelo experimental com alto, moderado e baixo nível de utilização, respectivamente.



Fonte: Resultados da pesquisa.

5.1. Modelo experimental Tipo I, com alto nível de utilização

Considerando o modelo experimental Tipo I, com alto nível de utilização do sistema a ser simulado e considerando todas as observações da simulação, obtivemos uma estimativa inicial para o parâmetro em avaliação dada por $\hat{\theta}_{ini} = 134{,}7813$. Quando se considerou o descarte da primeira metade da série, obtivemos $\theta = 135{,}0477$. Nestas condições o erro quadrático médio para $\hat{\theta}_{ini}$ foi $EQM(\hat{\theta}_{ini})$= 4,155896. A tabela 1 apresenta os resultados destas medidas e, a redução percentual do erro quadrático médio em relação ao $EQM(\hat{\theta}_{ini})$.

Tabela 1 – Resultados obtidos para o modelo experimental com alto nível de utilização.

Método	L estimado	$\hat{\theta}$	$EQM(\hat{\theta})$	$V_r(EQM(\hat{\theta}))$
Welch	1164905	135,18100	0,01844658	-0,9955613
SPC	399450	135,00940	0,01103734	-0,9973442
Aleatorização	365390	135,01630	0,01132434	-0,9972751
Cruzamento	154428	135,11140	0,06317613	-0,9847984
MSER-n	5013	134,86980	0,83913980	-0,7980845
Cruz. iterado	163101	135,11260	0,06359186	-0,9846984

Fonte: resultados da pesquisa

Podemos observar que os melhores resultados foram obtidos através do Método SPC; entretanto, este método propõe o descarte de uma quantidade de dados extremamente grande se comparada aos demais; este fato decorre da necessidade de se obter grupos que atendam aos requisitos de normalidade e ausência de correlação. No caso específico desta simulação, para que estes requisitos fossem satisfeitos, foi necessário construir grupos muito extensos. Neste caso, a exclusão de uma pequena quantidade de grupos pode levar uma exclusão de um volume muito grande de dados.

É fácil ver que, se a quantidade de dados descartada for muito grande, as estimativas produzidas tendem a ser muito próximas do valor θ, que considera apenas a segunda metade da série; por outro lado, se for possível produzir estimativas satisfatórias para θ, com um volume significativamente menor de dados descartados, o tamanho total da simulação poderia ser bastante inferior. Desta forma, não podemos considerar a solução proposta através do método SPC, como uma solução adequada.

Os métodos de Welch e de Aleatorização também apresentam resultados bons, mas sofrem de um problema semelhante ao método SPC, com um descarte excessivo de dados. O gráfico apresentado na figura 1 ilustra estes fatos.

A regra dos erros marginais MSER-n apresenta os piores resultados. A regra do cruzamento das médias e a nova metodologia apresentam resultados bastante semelhantes. Dentre as métricas utilizadas para a avaliação, seria difícil escolher uma das técnicas. Se considerarmos o fato da regra do cruzamento das médias exigir a calibração do parâmetro associado a quantos cruzamentos devem ser obtidos, até que se entenda por completo o período de aquecimento, enquanto a regra do cruzamento das médias iterado é livre da calibração deste parâmetro, isto nos leva a optar pela utilização da nova metodologia proposta.

5.2. Modelo experimental Tipo II, com moderado nível de utilização

Considerando o modelo experimental Tipo II, com moderado nível de utilização e todas as observações da simulação, obtivemos uma estimativa inicial $\hat{\theta}_{ini} = 37{,}07618$. Quando considerando o descarte da primeira metade da série, obtivemos $\theta = 37{,}07334$. O erro quadrático médio para $\hat{\theta}_{ini}$ foi $EQM(\hat{\theta}_{ini}) = 1{,}1143684 \times 10^{-2}$. A tabela 2 apresenta os resultados para estas medidas e também, a redução percentual do erro quadrático médio em relação ao $EQM(\hat{\theta}_{ini})$.

Tabela 2 – Resultados obtidos para o modelo experimental, com médio nível de utilização.

Método	L estimado	$\hat{\theta}$	$EQM(\hat{\theta})$	$V_r(EQM(\hat{\theta}))$
Welch	42506	37,08409	$3{,}667991 \times 10^{-4}$	-0,9679283
SPC	301988	37,07547	$3{,}829701 \times 10^{-5}$	-0,9966514
Aleatorização	84645	37,08309	$3{,}025090 \times 10^{-4}$	-0,9999272
Cruzamento	49273	37,08427	$3{,}685085 \times 10^{-4}$	-0,9677788
MSER-n	5000	37,07906	$9{,}238176 \times 10^{-4}$	-0,9192244
Cruz. iterado	1900	37,07860	$9{,}901725 \times 10^{-4}$	-0,9134225

Fonte: resultados da pesquisa

Novamente, o método SPC apresenta o melhor resultado obtido através de um descarte excessivo de dados; isto pode ser ilustrado através do gráfico da figura 1. Podemos observar que os melhores resultados foram obtidos através do Método de Welch, Método da Aleatorização e da Regra do cruzamento das médias; entretanto, novamente estes métodos propõem o descarte de uma quantidade de dados excessiva, se comparada aos demais. O volume de dados descartados é pelo menos 20 vezes maior que através da nova metodologia proposta. Por outro lado, mesmo com este volume bem inferior de dados descartados, a técnica do cruzamento das médias iterado fornece resultados bastante adequados considerando as métricas avaliadas.

A regra dos erros marginais MSER-n apresenta resultados similares aos da nova metodologia; entretanto, para obter estes resultados, foi necessário um descarte de 2,5 vezes mais dados que na nova metodologia proposta.

Novamente, frisamos que boas estimativas com um pequeno comprimento de período de aquecimento, levam um custo computacional menor para a execução de todo o procedimento.

5.3. Modelo experimental Tipo III, com baixo nível de utilização

Considerando o modelo experimental Tipo III, com baixo nível de utilização, obtivemos uma estimativa inicial para o parâmetro em avaliação dada por $\hat{\theta}_{ini}$ = 28,26932. Quando considerando o descarte da primeira metade da série, obtivemos θ = 28,27379. Nestas condições, o erro quadrático médio para $\hat{\theta}_{ini}$ foi $EQM(\hat{\theta}_{ini}) = 4{,}985163 \times 10^{-3}$. A tabela 3 apresenta os resultados para estas medidas e também, a redução percentual do erro quadrático médio em relação ao $EQM(\hat{\theta}_{ini})$.

Tabela 3 – Resultados obtidos para o modelo experimental com baixo nível de utilização.

Método	L estimado	$\hat{\theta}$	$EQM(\hat{\theta})$	$V_r(EQM(\hat{\theta}))$
Welch	12030	28,27354	$7{,}422078 \times 10^{-5}$	-0,9851117
SPC	27163	28,27323	$5{,}230056 \times 10^{-5}$	-0,9895088
Aleatorização	190120	28,27549	$3{,}341882 \times 10^{-5}$	-0,9932963
Cruzamento	15975	28,27370	$6{,}995527 \times 10^{-5}$	-0,9859673
MSER-n	4935	28,27256	$1{,}887736 \times 10^{-4}$	-0,9623411
Cruz. iterado	16022	28,27370	$6{,}995527 \times 10^{-5}$	-0,9859673

Fonte: resultados da pesquisa

Os melhores resultados foram obtidos pelo método de Aleatorização, mas com um volume de dados descartado muito maior que os outros métodos. A Regra dos erros marginais MSER-n apresentou resultados satisfatórios, mas ligeiramente inferiores aos demais métodos. Excluindo estes dois métodos, observamos resultados semelhantes entre os demais.

6. CONSIDERAÇÕES FINAIS

A estratégia de avaliação utilizada neste trabalho tem uma implementação bastante simples e pode ser utilizada para testar diversos métodos. As métricas aqui utilizadas, descrevem de forma bastante clara, uma classificação entre os métodos. Não buscamos concluir um método de senso comum, acima de quaisquer outros, entretanto, o método aqui proposto, mostrou-se bastante eficiente para todos os níveis de utilização do modelo experimental. São apresentadas conclusões preliminares, que buscam atestar as qualidades e deficiências de cada um dos métodos em análise. É possível notar que a influência do analista na calibração de alguns parâmetros associados ao método de detecção do período de aquecimento, tende a afetar sobremaneira a qualidade de resultados. Este é o motivo preponderante para a busca de uma técnica livre da calibração de parâmetros pelo analista.

Foi escolhido, de forma arbitrária, um parâmetro associado ao sistema em estudo, bem como um modelo experimental específico. Uma proposta de continuidade lógica é um estudo com a estimação de outros parâmetros de interesse e a postulação de modelos experimentais mais complexos. Acreditamos que exista uma ligação entre a complexidade do modelo e a adequação da técnica para detecção do período de aquecimento. Esta ligação somente pode ser avaliada, através da construção de mais modelos e da execução de mais experimentos.

A considerar o modelo experimental em uso neste trabalho, atestamos a grande capacidade de adequação da técnica proposta, em todos os níveis de utilização. Os resultados obtidos são bastante promissores e nos levam a um estudo mais profundo, acerca da nova técnica, aqui, apresentada.

7. REFERÊNCIAS BIBLIOGRÁFICAS

BANKS, J.; CARSON, J. S.; NELSON, B. L.; NICOL, D. M. **Discrete-Event System Simulation**. 3. ed. Upper Saddle River – NJ: Prentice-Hall, 2001.

CONWAY, R. W. Some tactical problems in digital simulation. **Management Science**, n.10:1, p.47–61, 1963.

FISHMAN, G. S. Estimating sample size in computing simulation experiments. **Management Science**, n.18:1, p. 21–38, 1971.

FISHMAN, G. S. **Concepts and Methods in discrete Event Digital Simulation**. New York: John Wiley and Sons, 1973.

GAFARIAN, A. V.; ANCKER, C. J.; MORISAKU, T. Evaluation of commonly used rules for detecting steady-state in computer simulation. **Naval Research Logistics Quarterly**, n.25, p.511–529, 1978.

GOLDSMAN, D.; SCHRUBEN, L. W.; SWAIN, J. J. Tests for transient means in simulated time series. **Naval Research Logistics Quarterly**, n.41, p. 171–187, 1994.

GORDON, G. **System Simulation**. Upper Saddle River – NJ: Prentice-Hall, 2001.

JACKWAY, P. T.; DESILVA, B. M. A methodology for initialization bias reduction in computer simulation output. **Asia-Pacific Journal of Operational Research**, n.9, p. 87–100, 1992.

KELTON, W. D.; LAW, A. M. A new approach for dealing with the startup problem in discrete event simulation. **Naval Research Logistics Quarterly**, n.30, p. 641–658, 1983.

MAHAJAN, P. S.; INGALLS, R. G. Evaluation of methods used to detect warmup period in steady state simulation. *In*: INGALLS, R. G.; ROSSETTI, M. D.; SMITH, J. S.; PETERS, B. A. (org.). **Proceedings of the 2004 Winter Simulation Conference**, Los Alamitos – CA: IEEE Computer Society, p. 663–671, 2004.

NELSON, B. L. Statistical analysis of simulation results. *In*: SALVENDY, G. (org.). **Handbook of Industrial Engineering**, New York: John Wiley and Sons, cap. 94, 2001.

PAWLIKOWSKY, K. A survey of problems and solutions. **Computing Surveys**, n.22:2, p.123–170, 1990.

ROBINSON, S. A statistical process control approach for estimating the warmup period. *In*: YÜCESAN, E.; CHEN, C. H.; SNOWDON, J. L.; CHARNES, J. M. (org.). **Proceedings of the 2002 Winter Simulation Conference**, San Diego - CA: Computer Society , p. 439– 446, 2002.

SPRATT, S. C. **Heuristic for startup problem**, Dissertação (Mestrado em Engenharia de Sistemas) – University of Virginia, Virginia, 1998.

SCHRUBEN, L. W. Detecting initialization bias in simulation output. **Operations Research**, n.30:3, p. 569–590, 1982.

SCHRUBEN, L. W.; SINGH, H.;TIERNEY, L. Optimal tests for initialization bias in simulation output. **Operations Research**, n.31:6, p.1167–1178, 1983.

VASSILACOPOULOS, G. Testing for initialization bias in simulation output. **Simulation**, n.52:4, p. 151–153, 1989.

WELCH, P. The statistical analysis of simulation results. *In*: LAVENBERG, S. (org.). The computer modeling handbook, New York: Academic Press, p. 268–328, 1983.

WHITE, K. P. An effective truncation heuristic for bias reduction in simulation output. **Simulation**, n.69:6, p. 323–334, 1997.

YÜCESAN, E. Randomization tests for initialization bias in simulation output. **Naval Research Logistics Quarterly**, n.40, p. 643–663, 1993.