

# Modelagem da propagação da variância em processos multivariados correlacionados em estágios

Josiane Da S Jesus, Paulo H C Maranhão\*, D.Sc.

Departamento de Ciência e Tecnologia, Instituto Militar de Engenharia, Rio de Janeiro, Brasil.

\*maranhao@ime.eb.br

**RESUMO:** A maioria dos processos produtivos consiste de um número significativo de estágios. Assim, modelar a variação das características de qualidade de produtos é uma estratégia fundamental para reduzir variação na produção e, por conseguinte, reduzir os custos das empresas responsáveis pelo processo. A ideia do estudo da transmissão da variação em processos com múltiplos estágios é identificar as fontes de variação em cada estágio no intuito de agir preventivamente em casos de erros no processo, ou seja, dado que o processo apresentou problemas, busca-se identificar exatamente em qual estágio ocorreu. Nesse sentido, este trabalho propõe um modelo para avaliar a propagação da variância em processos multivariados correlacionados em estágios. O estudo foi realizado por meio de simulações.

**PALAVRAS-CHAVE:** Processos multivariados correlacionados, propagação da variância, estágios.

**ABSTRACT:** Most industrial processes consist of a significant number of stages. Thus, modeling and monitoring the variation of the quality characteristics of products is a key to reducing variation in production and therefore reduce the costs of company responsible for the strategy process. The idea of studying the propagation of variation in multi-stage process is to identify the sources of variation in each stage in order to take preventive action in cases of errors in the process, e.g., since the case presented problems, how to identify exactly which stage occurred. In this sense, this work proposes a model to evaluate the propagation of variance in multivariate correlated processes in stages. The study was carried out through simulations.

**KEYWORDS:** Correlated multivariate processes, propagation of variance, stages.

## 1. INTRODUÇÃO

O mundo atual, competitivo e globalizado, provoca uma busca constante por qualidade, eficiência, diminuição de custos e maior retorno financeiro em toda e qualquer organização. Esse conjunto de objetivos se transformou em necessidade básica para a sobrevivência de empresas e indústrias.

Um produto ou serviço pode ser afetado negativamente pela variabilidade durante o seu processo de fabricação ou execução. Qualquer diferença entre o produto obtido e o produto desejado é denominada de desvio e uma medida bastante utilizada para mensurar essa diferença é a variância. Além disso, grande parte dos processos produtivos é composta por dois ou mais estágios. Em geral, observa-se que a variabilidade de um processo em uma etapa posterior ao primeiro estágio divide-se em duas partes: a primeira, adicionada pela própria etapa e, a segunda, transmitida pela etapa anterior. Sendo assim, o estudo da contribuição de cada etapa para a variância final é de suma importância para a melhoria do processo. Assim, o desenvolvimento de estudos sobre a propagação de erros surge como uma ferramenta imprescindível para qualquer organização, já que uma vantagem imediata do monitoramento da variabilidade é a redução do custo.

Geralmente, a avaliação de um processo leva em consideração mais de uma variável de interesse. Porém, não é raro ocorrer casos em que as variáveis de interesse mensuradas em cada estágio também tenham um grau elevado de correlação e nesse caso, técnicas multivariadas são necessárias para o desenvolvimento de modelos estatísticos que consigam avaliar a propagação da variabilidade.

Dessa forma, o principal objetivo deste trabalho é propor um modelo que consiga avaliar e mensurar a propagação de erros em processos multivariados em estágios, cujas características de qualidade de interesse estejam correlacionadas. Assim, o artigo proposto encontra-se organizado em 8 seções. Na seção 2, são tratados os principais trabalhos relacio-

nados. Nas seções 3, 4 e 5, são abordados o modelo autoregressivo de propagação, o método de análise de componentes principais e o modelo autoregressivo nas componentes principais, respectivamente. A seção 6 descreve a metodologia aplicada nas simulações. A sétima seção apresenta a análise de resultados. Por fim, a oitava seção contém a conclusão do trabalho.

## 2. TRABALHOS RELACIONADOS

Os primeiros estudos sobre a transmissão da variância foram realizados em meados dos anos noventa. Dessa forma, em 1998 foi realizado um estudo da variabilidade em um sistema flexível de montagem de portas de veículos através de layout fornecido pelo programa CAD. Nesse trabalho, os padrões de variação dos componentes pré-montados e a capacidade de compensação do sistema de montagem são modelados usando uma geometria de layout do localizador (dados CAD) e por meio de uma abordagem obtida da estatística multivariada, conhecida como a Análise de Componentes Principais [1].

Em 1999, foi proposto um modelo autoregressivo de ordem 1 (AR1) para estimar a propagação da variação em estágios de um processo produtivo em relação a uma única variável de interesse, identificando qual etapa produtiva contribuía significativamente para a variabilidade do processo [2]. Em 2002, foi desenvolvido um método para avaliar a propagação da variação em um processo de usinagem por meio de um modelo de estado-espço, baseado nas informações do projeto do produto e processo. Também foi utilizado o conceito virtual de operação para isolar as falhas e determinar as causas das mesmas [3].

Posteriormente, foi empregado um algoritmo de otimização que combina elementos finitos e programação não linear para determinar a melhor posição na fixação de componentes em um processo de montagem, de forma a minimizar a va-

riabilidade do mesmo. Com os elementos finitos foi possível construir a matriz de sensibilidade que estabelece relação linear entre os desvios de entrada com os de saída [4]. Em outro trabalho a metodologia Seis Sigma foi empregada em um processo produtivo de veículos, em especial, na Chaparia e Montagem da porta traseira, através da análise da folga, com o intuito de reduzir os desvios nas etapas produtivas [5].

Em 2007, foi apresentada uma metodologia do fluxo de variação baseada na descrição do projeto (CAD) e do processo. Essa metodologia utiliza a modelagem de espaço-estado linear que estabelece a relação entre os erros e suas causas, faz uma análise de sensibilidade para verificar o grau de impacto dos desvios na qualidade de cilindros automotivos e desenvolve a otimização do processo a fim de minimizá-lo [6].

Mais recentemente, alguns autores afirmaram que o controle de um processo é feito, otimizando cada fase do mesmo, ou seja, minimizando a transferência de desvios entre os diferentes estágios de fabricação. Para isso, foi necessária uma coordenação eficaz, capaz de dividir a produção em lotes com características semelhantes [7].

Ao longo dos últimos anos, percebe-se um aumento de trabalhos sobre a variabilidade em processos produtivos, já que o acúmulo de desvios ao longo dos diferentes estágios de fabricação pode ser extremamente prejudicial ao produto ou peça final. Assim, constatou-se um fato comum a todos esses trabalhos apresentados aqui: a existência de propagação da variância ao longo dos diferentes processos. Portanto, estudar o processo da transmissão da variabilidade ainda é um tema atual e relevante.

### 3. MODELO AUTOREGRESSIVO DE PROPAGAÇÃO

Em 1999, foi proposto um modelo autorregressivo (AR(1)) para estudar a transmissão da variação em processos de múltiplos estágios [2]. A escolha deste modelo é justificada visto que se desejava estudar a influência da variância entre etapas subsequentes. Dessa forma, a equação a seguir ilustra o modelo AR(1):

$$Y_i = \alpha_i + \beta_i Y_{i-1} + \varepsilon_i \quad i = 2, \dots, k \quad (1)$$

em que  $Y_i$  é a variável que representa a medida observada na etapa  $i$  e  $Y_{i-1}$  é a medida observada na etapa imediatamente anterior, onde é suposto que  $Y_i$  segue uma distribuição normal com média  $\mu_i$  e variância  $\sigma_i^2$ , os termos  $\alpha_i$  e  $\beta_i$  representam respectivamente, os coeficientes linear e angular do modelo. Os  $\varepsilon_i$ 's representam os erros aleatórios do modelo e também seguem uma distribuição normal com média zero e variância constante  $\sigma_\varepsilon^2$ . Além disso, supõe-se que os erros aleatórios e  $Y_i$  têm covariância igual à zero, ou seja, são independentes.

O principal atributo dos modelos AR (1) é que o valor atual está fortemente relacionado com o imediatamente anterior, isto é, há uma correlação na primeira defasagem entre  $Y_i$  e  $Y_{(i-1)}$ . Isso contribui para que os erros possam ser obtidos em cada defasagem considerada.

Assim, o AR (1) é bastante utilizado na identificação de falhas devido a sua característica preditora, ou seja, é possível comparar os valores encontrados pelo modelo com os dados reais.

Dessa maneira, a variabilidade transmitida entre dois estágios é estabelecida pela equação a seguir:

$$\sigma_i^2 = \beta_i^2 \sigma_{i-1}^2 + \sigma_{i,A}^2 \quad (2)$$

onde  $\sigma_i^2$  é a variância total da etapa  $i$ ;  $\sigma_{i,A}^2$  é a variância adicionada na etapa atual;  $\sigma_{i-1}^2$  é a variância na etapa anterior;  $\beta_i^2$  é o coeficiente responsável pela transmissão da variabilidade entre os dois estágios.

A Eq. 2 pode ainda ser generalizada pela seguinte expressão:

$$\sigma_k^2 = \beta_k^2 \sigma_{k-1,A}^2 + \dots + \beta_k^2 \beta_{k-1}^2 \dots \beta_2^2 \sigma_1^2 + \sigma_{k,A}^2 \quad (3)$$

### 4. ANÁLISE DE COMPONENTES PRINCIPAIS

A análise de Componentes Principais (ACP) é uma técnica multivariada, em que um número de variáveis relacionadas é transformado em um conjunto menor de variáveis não correlacionadas, chamadas de Componentes Principais (CP). Cada CP é uma combinação linear das variáveis originais. Assim, nesse método são geradas tantas componentes quanto variáveis, mas a grande vantagem da técnica é que, em geral, poucas componentes explicam a maior parte da variabilidade do conjunto de variáveis originais. Dessa forma, se  $p$  componentes representam a variabilidade total de um conjunto de  $p$  variáveis originais, a maior parte desta variabilidade pode ser explicada por  $k$  componentes ( $k < p$ ). Então,  $k$  componentes podem substituir as  $p$  variáveis iniciais, reduzindo a dimensão do conjunto de dados.

Seja  $X$  um vetor aleatório com vetor de médias  $\mu$  e matriz de covariâncias ou de correlação,  $\Sigma_{p \times p}$ , com  $\lambda_i$  autovalores e  $e_i$  os autovetores (normalizados) que obedecem às condições:  $e_i' e_j = 0$  ( $i \neq j$ );  $e_i' e_i = 1$  (para todo  $i$ ) e  $\Sigma_{p \times p} e_i = \lambda_i e_i$  (para todo  $i$ ). A  $j$ -ésima componente principal é dada pela Eq. 4.

$$Y_j = e_j' X = e_{j1} X_1 + e_{j2} X_2 + \dots + e_{jp} X_p \quad (4)$$

A Fig. 1 retrata um esquema de como funciona o método.



Fig. 1 – Resumo do Método ACP. Fonte: [próprios autores].

Dessa forma, como grande parte da variabilidade das variáveis originais é explicada por poucas componentes principais, pode-se dizer que, se mais de 80% da variância total pode ser explicada por duas ou três componentes, então estas componentes podem substituir as variáveis originais sem perda de informação [8].

### 5. MODELO AUTOREGRESSIVO NAS COMPONENTES PRINCIPAIS

O modelo autorregressivo de ordem um, descrito na seção anterior, servirá de base para o modelo autorregressivo nas componentes principais. Dessa forma, a componente principal do estágio  $i$  pode ser modelada em função da componente

principal da etapa anterior, conforme pode ser visto na Eq. 5 a seguir:

$$CP_i = \alpha_i + \beta_i CP_{i-1} + \varepsilon_i \quad (5)$$

Aplicando a variância na i-ésima componente principal, tem-se:

$$\sigma_{CP_i}^2 = \beta_i^2 \sigma_{CP_{i-1}}^2 + \sigma_{\varepsilon_i}^2 \quad (6)$$

Por analogia, todas as suposições e definições aplicadas ao modelo autorregressivo dado pela Eq. 1 continuam válidas para a Eq. 5. Assim, os termos  $\alpha_i$  e  $\beta_i$  representam também os coeficientes linear e angular do modelo e  $\varepsilon_i$  representa o i-ésimo erro aleatório, também conhecido por resíduo. Considera-se que os resíduos e as CP's sejam independentes.

Na prática, as equações 5 e 6 podem ser estimadas por meio das equações de 7 a 11, que representam respectivamente os estimadores da covariância das CP's entre os estágios, da variância das CP's, dos coeficientes angular e linear e da variância adicionada no próprio estágio.

$$S_{i-1,i} = \frac{\sum_{j=1}^n (CP_{j,i-1} - \overline{CP}_{i-1})(CP_{j,i} - \overline{CP}_i)}{n} \quad (7)$$

$$S_{ii} = \frac{\sum_{j=1}^n (CP_{j,i} - \overline{CP}_i)^2}{n} \quad (8)$$

$$\beta_i = \frac{S_{i-1,i}}{S_{i-1,i-1}} \quad (9)$$

$$\alpha_i = \overline{CP}_i - \beta_i \overline{CP}_{i-1} \quad i=2, \dots, k \quad (10)$$

$$\sigma_{\varepsilon_i}^2 = S_{ii} - \beta_i^2 (S_{i-1,i-1}) \quad i=2, \dots, k \quad (11)$$

## 6. METODOLOGIA

Neste tópico, apresenta-se a metodologia utilizada nas simulações. Dessa forma, a ideia é simular um processo produtivo formado por dois estágios, em que cada estágio é composto por quatro variáveis conforme a Fig. 2, em que as mesmas são correlacionadas em diferentes níveis. Assim, para cada estágio foram geradas 100 amostras dessas quatro variáveis, todas seguindo uma distribuição normal com média zero e variância um.

Considerou-se ainda, três tipos de correlação entre as variáveis de cada estágio, a saber: forte, moderada e fraca, e mais três tipos de relacionamento entre os estágios, conforme pode ser visto na Fig. 3. Assim, para a primeira simulação, considera-se que as variáveis em cada estágio são fortemente correlacionadas entre si, considerando ainda três tipos de correlação entre os estágios. Para a segunda e terceira simulação foram estabelecido respectivamente relações moderadas e fracas entre as variáveis e ainda foram levadas em conta as referidas correlações entre os estágios.

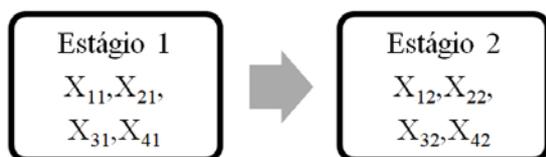


Fig. 2 – Representação de variáveis nos estágios. Fonte: [próprios autores].

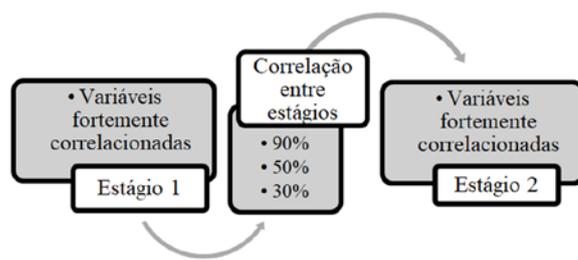


Fig. 3 – Esquema dos casos simulados considerando variáveis em cada estágio fortemente correlacionadas. Fonte: [próprios autores].

A partir de cada grupo de variáveis em cada estágio foram obtidas as componentes principais, selecionando a componente que mais explica a variabilidade das variáveis originais. Em seguida, é aplicado o modelo proposto entre as componentes selecionadas de cada estágio, isto é, a componente do segundo estágio é regredida em função da componente do primeiro estágio, conforme Fig. 4.

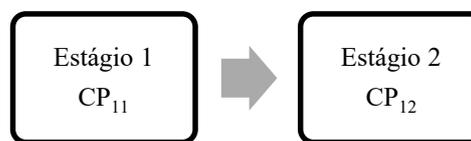


Fig. 4 – Representação de componentes principais nos estágios. Fonte: [próprios autores].

Embora sejam obtidas tantas componentes como variáveis, pela teoria da ACP, as primeiras componentes serão sempre responsáveis pela maior explicação da variabilidade das variáveis originais. Assim, o modelo proposto utilizará apenas as primeiras componentes principais de cada estágio.

## 7. RESULTADOS

Nesta seção, são apresentados os resultados referentes à metodologia descrita na seção anterior.

### 7.1 Resultados da simulação 1

Depois de geradas as variáveis fortemente correlacionadas em cada estágio de acordo com a metodologia, foram obtidas as componentes principais para o primeiro e o segundo estágios, considerando as três relações entre os estágios. Pela Tab. 1, observa-se que 95,80% da variância total, referente às quatro variáveis do estágio 1, é explicada quase totalmente por uma única componente principal, ou seja, a estrutura de variância e covariância das quatro variáveis originalmente simuladas pode ser representada por apenas uma componente principal, sem perda significativa de informação. A Tab. 1 apresenta o percentual de variância explicada por cada componente.

Tabela 1: Percentual total da variância explicada pelas CP's Simulação 1 Estágio 1.

% CP1	% CP2	% CP3	% CP4
95.8%	4.20%	0.00%	0.00%

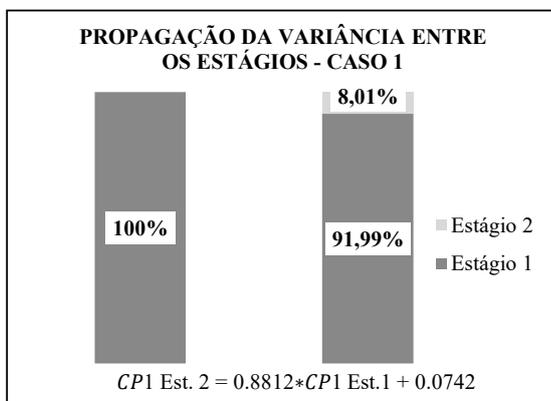
Para o estágio 2, conforme descrito na metodologia, foram considerados três casos de relacionamento entre os estágios. Assim, de acordo com a Tab. 2. Para uma correlação de 90%, a primeira componente principal representa 83,10% do

total da variância explicada. No caso 2, a CP1 responde por 46,40% das variações presentes no estágio 2. Enquanto que, no caso 3, esse valor cai para 39,10%, devido principalmente a menor correlação entre os estágios.

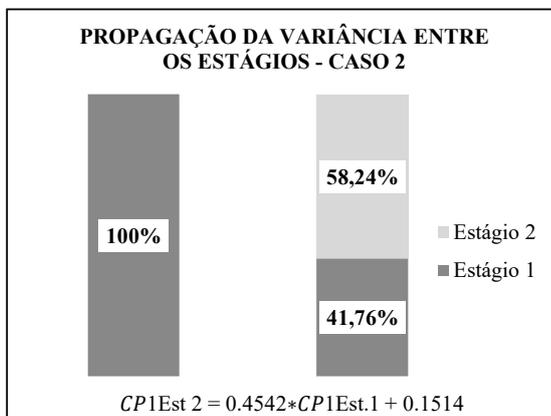
**Tabela 2: Percentual total da variância explicada pelas CP's Simulação 1 Estágio 2.**

Caso	% CP1	% CP2	% CP3	%CP4
1 (90%)	83.10%	8.70%	5.00%	3.20%
2 (50%)	46.40%	24.20%	17.90%	11.50%
3 (30%)	39.10%	26.50%	21.00%	13.40%

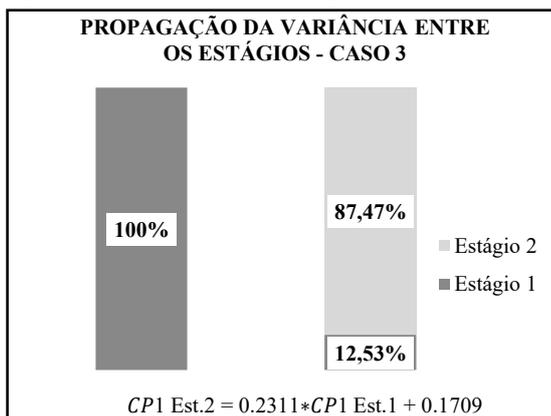
Analisando as tabelas 1 e 2, observa-se que quando as variáveis originalmente geradas são fortemente correlacionadas, as primeiras componentes principais de cada estágio se tornam responsáveis por quase a totalidade da variabilidade das variáveis originais, principalmente no estágio 1. Dessa forma, aplicou-se o modelo proposto, considerando as primeiras componentes em cada estágio. As figuras 5, 6 e 7 revelam como se propaga a variância entre estágios.



**Fig. 5 – Variabilidade entre estágios - Simulação 1 caso 1.**



**Fig. 6 – Variabilidade entre estágios - Simulação 1 caso 2.**



**Fig. 7 – Variabilidade entre estágios - Simulação 1 caso 3.**

A Fig. 5 apresenta que 91,99% da variabilidade presente no estágio 2 foi originada no estágio anterior. Isso significa que um erro no estágio anterior é quase totalmente propagado para a etapa seguinte. É possível ainda notar por meio das figuras 6 e 7 que, conforme a relação entre os estágios diminui, cai também a transmissão do erro.

## 7.2 Resultados da simulação 2

As variâncias explicadas pelas componentes principais no estágio 1 para a simulação 2 são apresentadas na Tab. 3. Já a Tab. 4 apresenta os valores das variâncias explicadas no estágio 2 de acordo com cada nível de correlação entre os estágios. Assim, analisando a Tab. 3, verifica-se que a primeira componente principal representa 78,60% da variabilidade do conjunto das quatro variáveis, o que indica que as quatro variáveis originais podem ser representadas por uma única componente principal.

**Tab 3: Percentual total da variância explicada pelas CP's Simulação 2 Estágio 1.**

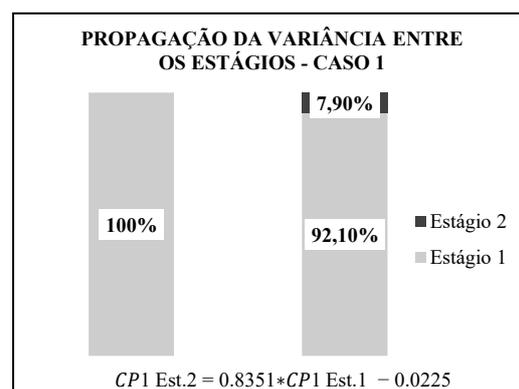
% CP1	% CP2	% CP3	%CP4
78.60%	21.40%	0.00%	0.00%

Além disso, observando a Tab. 4, é possível afirmar que no primeiro caso, a CP1 corresponde a 68,10% da variância total, enquanto que a CP1 do segundo caso, explica somente 34,60% e para o terceiro caso, tem-se que apenas 31% do sistema pode ser representados por uma única CP. No entanto, as primeiras CP's de cada caso serão adotadas para simulação do modelo proposto, visto que, no primeiro estágio, foi utilizada apenas uma componente.

**Tabela 4: Percentual total da Variância explicada pelas CP's simulação 2 Estágio 2.**

Caso	% CP1	% CP2	% CP3	%CP4
1 (90%)	68.10%	22.30%	5.40%	4.20%
2 (50%)	34.60%	28.00%	21.60%	15.80%
3 (30%)	31.00%	26.00%	24.60%	18.40%

Observe que, embora as variáveis nos estágios estejam moderadamente correlacionadas, as primeiras componentes principais ainda conseguem um alto grau de explicação no primeiro estágio, especialmente quando existe um forte relacionamento entre os estágios. As figuras 8, 9 e 10 retratam a propagação da variância entre os dois estágios em relação à primeira componente principal, visto que a mesma representa a maior parte da variabilidade das variáveis.



**Fig. 8 – Variabilidade entre estágios - Simulação 2 caso 1.**

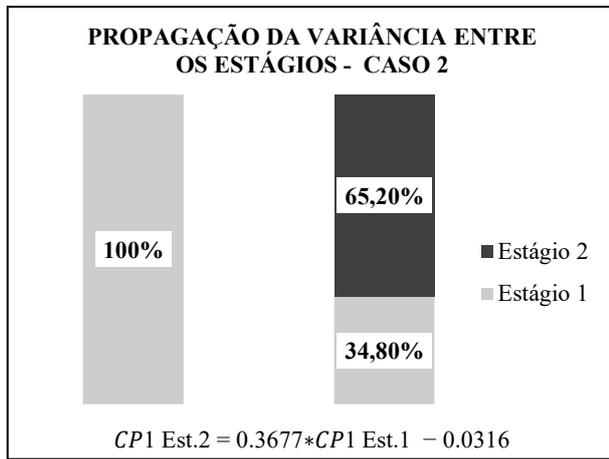


Fig. 9 – Variabilidade entre estágios - Simulação 2 caso 2.

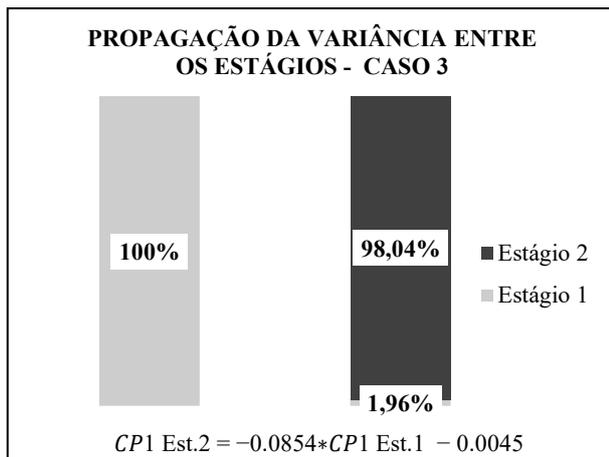


Fig. 10 – Variabilidade entre estágios - Simulação 2 caso 3.

Pela Fig. 8, observa-se que a propagação da variabilidade entre estágios foi de 92,10% para o caso 1 da simulação 2. Já pela Fig. 9, tem-se que a transmissão da variação foi de 34,80% no caso 2. Por fim, a Fig. 10 mostra que, no segundo estágio, apenas 1,96% da variância foi proveniente do estágio anterior e que 98,04% dos erros foram gerados no próprio estágio.

### 7.3 Resultados da simulação 3

Conforme a Tab. 5. Percebe-se, que na simulação 3, tal como foi observado nas simulações anteriores, também se pode utilizar uma única CP, visto que a mesma representa mais de 70% da variância total da variáveis originais no primeiro estágio.

Tab 5: Percentual total da variância explicada pelas CP's Simulação 3 Estágio 1.

% CP1	% CP2	% CP3	%CP4
75.10%	24.90%	0.00%	0.00%

A Tab. 6 apresenta os valores da variabilidade total explicada por cada componente principal no estágio 2 para cada tipo de correlação especificada. Nesse caso, ainda serão utilizadas apenas as primeiras CP's.

Tabela 6: Percentual total da variância explicada pelas CP's Simulação 3 Estágio 2.

Caso	% CP1	% CP2	% CP3	%CP4
1 (90%)	62.50%	25.60%	6.70%	5.20%
2 (50%)	31.90%	28.00%	22.40%	17.70%
3 (30%)	29.10%	26.10%	24.60%	20.20%

Analisando a Fig. 11, observa-se que a propagação de erros entre os estágios foi de 91,28%.

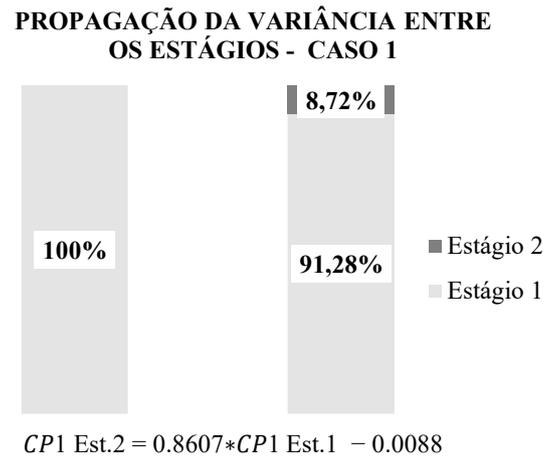


Fig. 11 – Variabilidade entre estágios - Simulação 3 caso 1.

A Fig. 12 mostra que a variância transmitida entre os estágios foi de 38,95% para a primeira componente no caso 2 da simulação 3 e, no caso 3 (Fig. 13), a propagação de desvios foi somente de 0,5%, conforme era esperado, na medida que a correlação entre as variáveis era de apenas 30% e a relação entre os estágios era fraca.

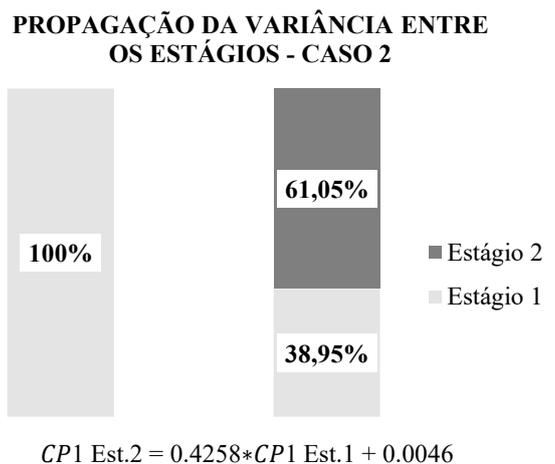


Fig. 12 – Variabilidade entre estágios - Simulação 3 caso 2.

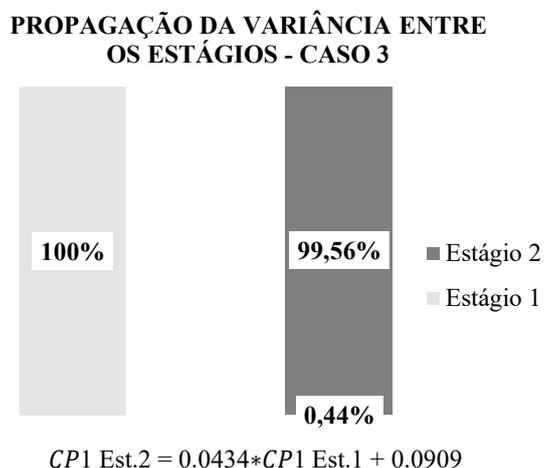


Fig. 13 – Variabilidade entre estágios - Simulação 3 caso 3.

## 8. CONCLUSÃO

Após analisar os casos simulados, conclui-se que o modelo autoregressivo nas Componentes Principais foi capaz de captar a variação vinda do estágio anterior, mostrando-se promissor para aplicação em casos reais. Além disso, observou-se que quanto maior for a correlação entre as variáveis de interesse, maior será a capacidade da técnica de Análise de Componentes Principais em reduzir o número de variáveis, devido ao elevado percentual da variância explicada pelas componentes, em especial pela primeira componente. Em consequência, maior será a captação da propagação dos desvios entre as etapas.

Em resumo, pode-se afirmar que os modelos autoregressivos nas componentes principais são válidos para avaliar o comportamento da propagação da variância em processos multivariados em estágios. Por meio deles, é possível identificar qual a etapa é responsável pela maior variabilidade do processo.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Ceglarek, D.J. Multivariate Analysis and Evaluation of Adaptive Sheet Metal Assembly Systems. The University of Michigan. 1998.
- [2] Lawless, J.F., Mackay, R.J., Robinson, J.A. Analysis of variation transmission in manufacturing processes – Part 1. *Journal of Quality Technology*, v.31, n.2, 1999.
- [3] Huang, Q., Zhou, S., Shi, J. Diagnosis of multi-operational machining processes through variation propagation analysis. *Robotics and Computer Integrated Manufacturing* 18, p.233-239, 2002.
- [4] Camelio, J. A., Hu S. J., Ceglarek, D. Impact of Fixture Design on Sheet Metal Assembly Variation. *Journal of Manufacturing Systems*, v.23, n.3, 2004.
- [5] Gaio, A. P.C., Sá, F.C. O Uso da Metodologia Seis Sigma em um Processo da Indústria Automotiva por Meio de um Estudo de Caso. Trabalho Graduação em Engenharia de Produção. UERJ. Resende. 2004.
- [6] Zhang, M., Djurdjanovic, D., Ni, J. Diagnosibility and sensitivity analysis for multi-station machining processes. *International Journal of Machine Tools & Manufacture* 47, p.646-657, 2007.
- [7] Han, K., Wang, K. Coordination and control of batch-based multistage processes. *Journal of Manufacturing Systems* 32, p.372-381, 2013.
- [8] Johnson, R.A., Wichern, D.W. *Applied Multivariate Statistical Analysis*. 6ª ed. Pearson Prentice Hall. 2007