

CONTROLE DE PROCESSOS EM MINERAÇÃO BASEADO EM ANÁLISE POR COMPONENTES PRINCIPAIS: UM ESTUDO DE CASO REAL

Pedro Henrique Trindade Dias Cabral

Depto. de Estatística, Instituto de Ciências Exatas – ICEX, Universidade Federal de Minas Gerais
Av. Pres. Antônio Carlos, 6627 – Pampulha, 31.270-901, Belo Horizonte, MG, Brasil
pedro_eng.quimica@hotmail.com

Prof. Gustavo Matheus de Almeida

Depto. de Engenharia Química, Escola de Engenharia, Universidade Federal de Minas Gerais
Av. Pres. Antônio Carlos, 6627 – Pampulha, 31.270-901, Belo Horizonte, MG, Brasil
galmeida@deq.ufmg.br

RESUMO

Tecnologias de automação, controle e comunicação de dados, e conceitos de sistemas ciber-físicos, Internet das Coisas e computação em nuvem, têm sido integrados dentro do conceito de Indústria 4.0, cujo objetivo é tornar as organizações mais eficientes. No Brasil, várias empresas estão se adequando à Indústria 4.0. Nesse contexto, este trabalho investigou a utilização da técnica estatística multivariada, Análise por Componentes Principais (PCA), como auxílio ao controle de um sistema de moagem de uma mineradora em Minas Gerais. Foi possível observar as suas capacidades de redução do número de variáveis de processo e de determinação de regiões ótimas de operação, informação usada na definição do *set point* do sistema de controle. Usando três componentes principais como variáveis de leitura, que explicam 78% da variação total das variáveis, obteve-se um aumento médio de 1,23 ton/h na alimentação do moinho; uma redução média de 1,35% na média do teor de sólidos da polpa e de 16% em sua variabilidade; e uma redução média de 23% na variação do nível da caixa de saída do moinho.

PALAVRAS CHAVES: Análise por Componentes Principais, Processo de moagem, Banco de dados industrial, Controle de processos, Avaliação de desempenho, Indústria 4.0

1. Introdução

A qualidade dos processos industriais é encarada hoje como fator importante para as organizações que querem sobreviver em um cenário de competição agressiva, mercados globalizados, pressões social e governamental para processos produtivos mais seguros e econômicos, e evolução tecnológica contínua. Várias são as definições propostas para o termo *qualidade*, onde uma definição moderna pode ser dada como sendo uma ou mais características desejadas que um processo, produto ou serviço deva ter, sendo inversamente proporcional a variabilidade dessas características. Note que essa definição implica que se a variabilidade nas características importantes decresce, a qualidade do processo, produto ou serviço aumenta. Desta forma, a melhoria da qualidade em processos industriais é alcançada e, conseqüentemente, nos produtos gerados por esses processos, ao se minimizar a sua variabilidade (MONTGOMERY, 2016). Nesta direção, tem-se, atualmente, o termo Indústria 4.0, que engloba algumas tecnologias para a automação e compartilhamento e mineração de dados, com o intuito de deixar as unidades industriais mais eficientes e descentralizadas, utilizando conceitos de Controle Avançado de Processos, Inteligência Artificial, Sistemas Ciber-físicos, Internet das Coisas e Computação em Nuvem (AMARAL, 2016; JASPERNEITE, 2012). Tecnologias a nível Indústria 4.0 permite uma melhor visão e execução de "Fábricas *Smarts*" com os seus *layouts* modulares, a partir de sistemas ciber-físicos que monitoram os processos, gerando uma cópia virtual do mundo físico e tomando decisões antecipatórias e descentralizadas (PORTER & HEPPELMANN, 2015; SCHWAB, 2016). Esse termo foi usado pela primeira vez na Hannover Messe (<http://www.hannovermesse.de/>). O grupo de Trabalho em Indústria 4.0, presidido por Siegfried Dais (Robert Bosch GmbH) e Henning Kagermann (*German Academy of Science and Engineering*) apresentou em Outubro de 2012 um conjunto de recomendações para a implementação da Indústria 4.0 ao Governo Federal Alemão. Em abril de 2013, o relatório final desse grupo de trabalho foi apresentado novamente na Feira de Hannover (SCHWAB, 2016). Desde então, tecnologias à nível Indústria 4.0 tem sido debatidas, desenvolvidas e implementadas pelos maiores grupos econômicos do mundo.

Em 2007, um grupo de três alunos de pós-graduação da Universidade Estadual de Campinas (UNICAMP), com o intuito de comercializar tecnologia, fundou a I.Systems (<http://www.i.systems.com.br/>), uma empresa inovadora que, através de tecnologias

exclusivas de controle e automação, busca otimizar processos industriais, atuando na vanguarda da Indústria 4.0 no Brasil. Foram quatro anos de desenvolvimento até o lançamento de seu produto inicial, o *Leaf* (<http://www.i.systems.com.br/produtos/>), um *software* de controle avançado baseado em técnicas de inteligência artificial com o intuito de diminuir a variabilidade e otimizar o *set-point* de processos e equipamentos (I.SYSTEMS, 2016). Para a sua aplicação em processos industriais, o *software* deve ser configurado de acordo com cada tipo de processo e equipamento. Várias tecnologias são embarcadas no *Leaf*, onde uma programação robusta é requerida para cada aplicação. Várias técnicas estatísticas são programadas e embarcadas no *Leaf* em paralelo a estas tecnologias. Regressão linear e não linear, regressão multivariada, clusterização, e controle estatístico de processos (CEP), são algumas das técnicas estatísticas utilizadas para a configuração do *software* antes de sua atuação no controle avançado de um determinado processo. Um ponto a ser levado em consideração na aplicação do *software* é que há um limite físico de leitura e escrita em relação à quantidade de variáveis. Nesse contexto, o presente trabalho buscou ilustrar o potencial da técnica estatística multivariada, denominada Análise por Componentes Principais (PCA, *Principal Component Analysis*), como uma solução alternativa para a redução do número de variáveis de processo. Como exemplo de aplicação, tem-se a determinação de regiões ótimas de operação. O estudo de caso refere-se a um banco de dados históricos de um processo real de moagem de uma planta de beneficiamento de minério em Minas Gerais.

2. Análise por Componentes Principais (PCA)

A técnica estatística, Análise por Componentes Principais, conhecida popularmente por PCA (sigla referente a abreviação de seu nome em inglês: *Principal Component Analysis*), é um dos métodos multivariados mais simples. Segundo Casarin *et al.* (2012), um dos seus principais usos é quando diversas variáveis ou características de um processo devem ser observadas ao mesmo tempo, como é o caso dos processos de beneficiamento mineral. O seu princípio é a explicação da estrutura da matriz de variâncias e covariâncias (isto é, da correlação espacial) de p -variáveis aleatórias, por meio da construção de combinações lineares dessas p -variáveis. Essas novas variáveis, dadas pelas combinações lineares das variáveis originais, são denominadas *componentes principais (CP's)*, sendo não correlacionadas e independentes, o que favorece a análise, principalmente quando vários fatores estão envolvidos (CASARIN, 2012; YIN *et al.*, 2014).

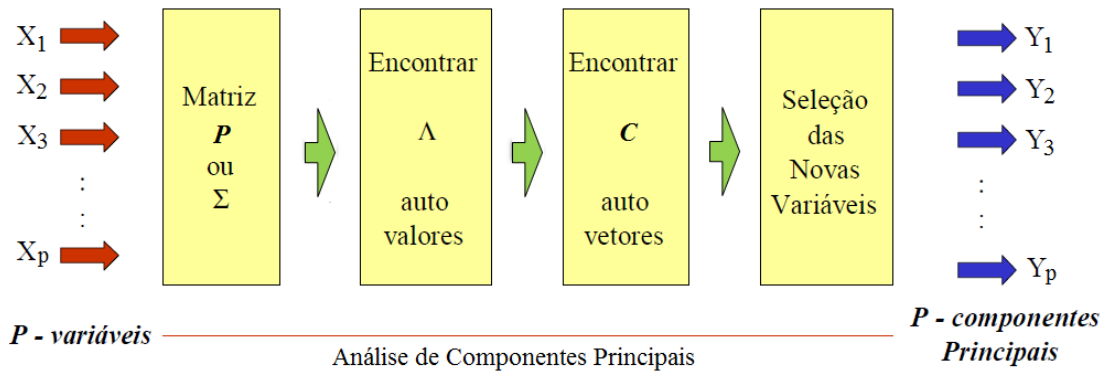
Se temos p -variáveis originais do processo, é possível obter p componentes principais. Autores como Morrison (1976), Reinsel (1993), Johnson e Wichern (1998), Souza *et al.* (2011), Vines (2000), e Jackson (2003), estudaram a aplicação de PCA com o propósito de redução e interpretação de p -variáveis originais a partir apenas de k -combinações lineares, onde essas combinações explicam a maior parte da variância total dos dados originais; $k \ll p$. Portanto, o objetivo é analisar a informação (total) contida em p -variáveis originais a partir de k -variáveis latentes (as componentes principais), principalmente quando as variáveis originais são correlacionadas. A qualidade dessa aproximação depende do número de componentes (k) que serão mantidas no sistema, o que pode ser medido através da avaliação da proporção da variância total explicada por k -componentes (MINGOTI, 2005). De acordo com Khattree e Naik (2000), a técnica proporciona, pelo estudo de correlação entre as componentes principais e as variáveis originais, o conhecimento de quais são as variáveis originais de maior relevância no sistema.

Ao utilizar a técnica PCA, relações inicialmente desconsideradas em análises univariadas podem ser reveladas (JOHNSON & WICHERN, 1998). Essa análise multivariada é realizada a partir dos escores, que são os valores assumidos pelas componentes principais. É comum utilizar os escores das componentes principais para realizar a simples ordenação (ranking) dos elementos amostrais observados, com o intuito de identificar aqueles que apresentam maiores, ou menores, valores globais das componentes (MINGOTI, 2005).

Para a determinação das componentes principais, é necessário calcular a matriz de variâncias e covariâncias (Σ) ou matriz de correlações (P), dado o conjunto de variáveis originais (X); encontrar os seus autovalores (Λ) e os autovetores (C); e então escrever as combinações lineares, que serão as novas variáveis (Y), denominadas de CP's (MARCHESAN *et al.*, 2011). A Figura 1 apresenta uma sequência para a obtenção das componentes principais.

Figura 1: Esquema da obtenção das componentes principais.

Fonte: Adaptado de Marchesan et al., 2011.



A seguir, apresentam-se os fundamentos da técnica PCA, tendo-se como referência as publicações de Mingoti (2005) e Montgomery (2016). Na prática, a matriz $\Sigma_{p \times p}$ é desconhecida, sendo estimada ($S_{p \times p}$) através dos dados amostrais coletados. Seja o vetor de variáveis originais, $X = [X_1 \ X_2 \ \dots \ X_p]'$; o vetor de médias amostrais, $\bar{X} = [\bar{X}_1 \ \bar{X}_2 \ \dots \ \bar{X}_p]$; a matriz de variâncias e covariâncias amostral, $S_{p \times p}$; e o vetor de autovalores, $[\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p]$ e a respectiva matriz de autovetores normalizados, $[\hat{c}_1, \hat{c}_2, \dots, \hat{c}_p]$. Então, a j -ésima componente principal (CP) estimada é definida conforme a Equação 1.

$$\hat{Y}_j = \hat{c}_j' \mathbf{X} = \hat{c}_{j1}X_1 + \hat{c}_{j2}X_2 + \dots + \hat{c}_{jp}X_p, \quad j = 1, 2, \dots, p \quad (1)$$

A esperança e variância da componente \hat{Y}_j são, respetivamente, conforme as Equações 2 e 3.

$$E[\hat{Y}_j] = \hat{c}_j' \bar{\mathbf{X}} = \hat{c}_{j1}\bar{X}_1 + \hat{c}_{j2}\bar{X}_2 + \dots + \hat{c}_{jp}\bar{X}_p \quad (2)$$

$$Var[\hat{Y}_j] = \hat{c}_j' S_{p \times p} \hat{c}_j = \hat{\lambda}_j \quad (3)$$

Algumas propriedades das CP's amostrais são apresentadas a seguir.

Propriedade 1: A variância estimada de \hat{Y}_j é igual a $\hat{\lambda}_j$, $j=1,2, \dots, p$.

Propriedade 2: As componentes principais são não correlacionadas, ou seja, a covariância entre as componentes \hat{Y}_j e \hat{Y}_k é igual a zero, para todo $k \neq j$.

Propriedade 3: A proporção da variância total explicada pela j -ésima CP é o j -ésimo autovalor $\hat{\lambda}_j$. Portanto, a proporção de variabilidade nos dados originais explicada pela j -ésima CP é dada conforme a Equação 4.

$$\hat{\lambda}_j / (\hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_p) \quad (4)$$

Assim, pode-se obter o quanto da variabilidade total é explicada pela retenção de apenas k das p -CP's, calculando-se a soma dos autovalores para essas k componentes e comparando-a com a soma de todos os p autovalores. Na prática, em uma análise de dados, selecionam-se as k componentes principais mais relevantes (com maiores variâncias), e calculam-se os escores dessas componentes para cada elemento amostral.

Propriedade 4: A correlação estimada entre a j -ésima CP e a variável aleatória X_i , $i=1,2, \dots, p$ é dada pela Equação 5, onde s_{ii} é a variância amostral da variável aleatória X_i .

$$r_{\hat{Y}_j, X_i} = \left(\hat{c}_{ji} \sqrt{\hat{\lambda}_j} \right) / \sqrt{s_{ii}} \quad (5)$$

Propriedade 5: Pelo teorema da decomposição espectral, a matriz de variâncias e covariâncias amostral ($S_{p \times p}$) pode ser expressa conforme a Equação 6.

$$S_{p \times p} = C \Lambda C' = \sum_{j=1}^p \hat{\lambda}_j \hat{c}_j \hat{c}_j' \quad (6)$$

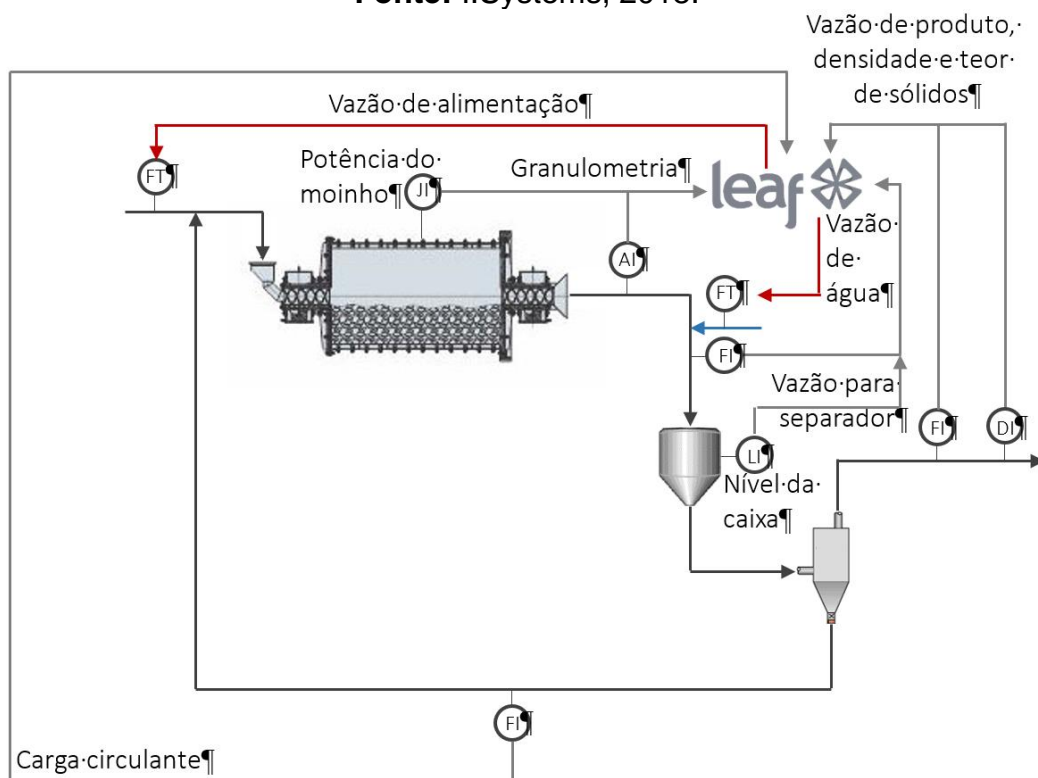
É usual calcular as CP's usando variáveis padronizadas, isto é, com média zero e variância 1. Desse modo, as componentes principais são obtidas a partir da matriz de variâncias e covariâncias das variáveis originais padronizadas, o que é equivalente a extrair-se as CP's a partir da matriz de correlações das variáveis originais. A razão para isso é que as variáveis originais de processo são, na maioria das vezes, expressas em escalas de grandezas diferentes e, conseqüentemente, podem ter magnitudes bem diferentes. Assim, uma variável pode parecer contribuir significativamente para a variabilidade total do sistema apenas porque sua unidade de medida é de uma grandeza consideravelmente maior do que das demais variáveis (MONTGOMERY, 2016).

3. Estudo de Caso

Neste trabalho, o banco de dados do estudo de caso diz respeito às operações de um moinho de beneficiamento de minério de ferro de uma mineradora no Brasil. O Tratamento ou Beneficiamento de minérios consiste de operações que buscam modificar a granulometria, a concentração relativa e/ou a forma dos minerais, sem modificação de suas identidades química e física. Desta maneira, visa-se, através de operações de concentração, remover os minerais de ganga (conjunto de minerais não aproveitáveis de um minério) dos minerais de minério (minerais valiosos, passíveis de serem explorados economicamente). A Figura 2 é um esquema do moinho do estudo de caso. O sistema de moagem é composto pelo moinho de bolas (Potência: 2486 cv; Dimensões: 6110 mm x 4400 mm, e Rotação: 15,18 rpm) e por uma caixa, onde a saída de material do moinho é depositada. Em seguida, este material é direcionado para um sistema de classificação composto por peneiras de alta frequência (Dimensões: 1,22 m x 3,04 m; Abertura da tela: 0,150 mm – poliuretano; Potência Instalada: 2,5 cv por conjunto de 5 pineiras). A granulometria e o teor de sólidos são os dois parâmetros de qualidade desse processo, funcionando como restrição à taxa de alimentação.

Figura 2: Fluxograma do processo de beneficiamento mineral.

Fonte: I.Systems, 2018.



3.1 Banco de Dados

A Tabela 1 contém o conjunto de oito variáveis operacionais do estudo de caso. A variável Y1-MV e Y2-MV são as variáveis controladas pelo *software*, ISTO É, as variáveis de escrita. As demais são variáveis de processo e de distúrbio, ou seja, variáveis de leitura, sendo as variáveis X1-PV e X2-PV aquelas de maior interesse por serem as indicadores de qualidade do processo de beneficiamento de minério. A base dessa seleção é a literatura, o conhecimento prático de operários e de engenheiros da unidade de mineração, e a equipe da I.Systems. O banco de dados é referente a 19 dias de operação contínua, com amostragem de 5 em 5 segundos, com um total de 259.189 registros por variável.

Tabela 1: Variáveis operacionais da do sistema de moagem do estudo de caso.

Descrição	Código	Unidades
Vazão de Alimentação do moinho	Y1-MV	Ton/h
Vazão de água da caixa de saída do moinho	Y2-MV	m ³ /h
Teor de sólidos da polpa	X1-PV	%
Granulometria 100#	X2-PV	#
Nível da caixa de saída do moinho	X3-DV	%
Densidade da polpa	X4-DV	g/cm ³
Carga Circulante	X5-DV	%
Eficiência do moinho	X7-DV	%

3.2 Definição do Problema

O sistema de controle avançado (*Leaf*), integrado neste processo, observa a granulometria do material (X2-PV) para corrigir a vazão de alimentação do moinho (Y1-MV). Quando a granulometria está baixa, ele aumenta a taxa de alimentação. O *software* também considera a faixa de operação normal do teor de sólidos e da densidade para ajustar a taxa de alimentação. Ele verifica se é possível aumentar ou não a taxa e monitora a consequente variação nesses parâmetros. Estas variáveis estão relacionadas com o controle de vazão de água para a caixa. O nível da caixa também é monitorado para evitar que o material transborde na área. O nível está ligado ao controle da vazão de polpa que segue para as peneiras. O *Leaf* controla a vazão de polpa de forma a manter o nível em uma faixa segura de operação. Esta é a variável mais importante para o controle da bomba. A quantidade de polpa a ser enviada para as peneiras também é monitorada, para não exceder o limite de operação do equipamento. Todas as variáveis deste estudo de caso são possíveis de serem monitoradas, e como o *Leaf* é um *software* que trabalha com análise multivariada

para prever os distúrbios, quanto maior o número de variáveis, melhor a estratégia de controle para a redução de variabilidade. Devido a limitação do número de variáveis que podem ser lidas pelo *software*, espera-se que a técnica PCA consiga reduzir o número de variáveis sem perda significativa da informação disponível sobre o processo. Além disso, espera-se que a técnica consiga indicar para o *software* regiões onde as variáveis operam com menor variabilidade. Sem o controle e a otimização do processo pelo *Leaf*, a taxa de alimentação apresenta uma média de 134,05 ton/h e um desvio de 5,36 ton/h (Figura 3), e o teor de sólidos, uma média de 51,91% e um desvio padrão de 2,03% (Figura 4).

Toda a análise estatística e simulações neste trabalho foram conduzidos no *software* R (<http://www.r-project.org/>) e com o *dashboard* do *Leaf*, que é uma ferramenta gráfica e de simulação usado pelo grupo de consultoria de controle de processos da I.Systems para estimar o benefício de aplicação do *software* nos processos industriais. Como o *Leaf* não apresenta a técnica de análise de componentes principais embarcado no seu *dashboard*, todo o estudo da técnica foi realizado no R. Todo o script (com os pacotes utilizados) em R, desenvolvido neste trabalho, encontra-se em Anexo.

Figura 3: Histograma da Taxa de Alimentação (Y1-MV).

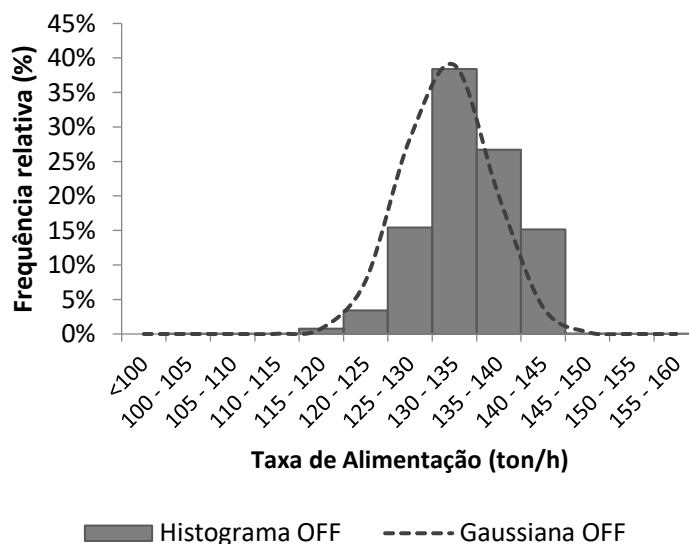
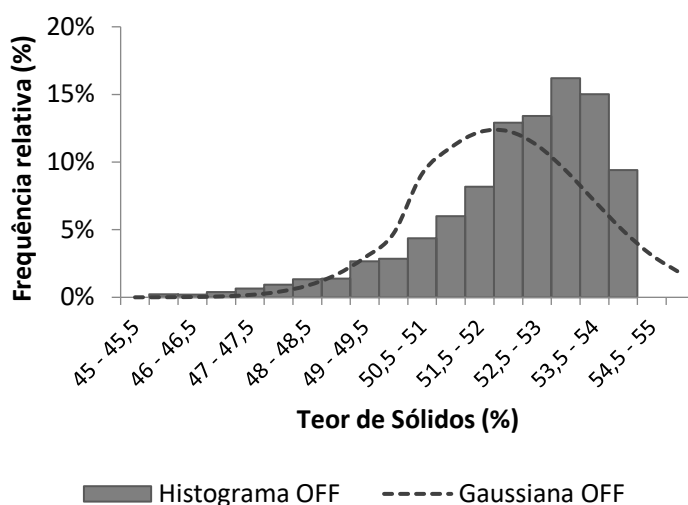


Figura 4: Histograma do Teor de Sólidos (X1-PV).



4. Resultados e Discussão

4.1 Etapa 1 – Identificação de um Modelo PCA

O objetivo é construir um modelo aplicando a análise de componentes principais (PCA) sobre o banco de dados de trabalho. *A priori*, foi avaliado a matriz de correlações das variáveis (Figura 5).

Figura 5: Matriz de correlações das variáveis.

	X3-DV	X2-PV	X4-DV	X1-PV	X7-DV	X5-DV
X3-DV	1,0000					
X2-PV	-0,1038	1,0000				
X4-DV	0,4134	-0,1781	1,0000			
X1-PV	0,4164	-0,1235	0,9821	1,0000		
X7-DV	-0,0989	0,1330	0,0026	0,0666	1,0000	
X5-DV	0,3131	-0,1865	0,5238	0,4684	-0,1036	1,0000

Pode-se observar a forte correlação positiva entre a variável X4-DV e X1-PV (maior que 98%). Isso se deve ao fato que quanto maior a densidade da polpa, maior é a quantidade do teor de sólidos da polpa. Outra correlação significativa é entre X3-DV com X4-DV, X1-PV e X5-DV, significando que há uma correlação positiva entre o nível da caixa de saída

do moinho (X3-DV) com a densidade da polpa (X4-DV), o teor de sólidos da polpa (X1-PV) e a carga circulante do sistema (X5-DV). Também há uma correlação moderada entre a densidade da polpa (X4-DV) com a carga circulante do sistema (X5-DV), e entre o teor de sólidos da polpa (X1-PV) e a carga circulante do moinho (X5-DV). A granulometria 100# (X2-PV) apresentou uma pequena correlação negativa com a carga circulante (X5-DV), a densidade da polpa (X4-DV), e com o teor de sólidos da polpa (X1-PV). A variável eficiência do moinho (X7-DV) não apresentou correlação significativa com nenhuma outra variável de distúrbio ou processo.

Na sequência, aplicou-se a análise por componentes principais a partir da matriz de correlações (Figura 5). Nesta etapa, o uso da técnica de análise de componentes principais, sem a remoção da variável eficiência do moinho (X7-DV), foi realizada com o intuito de investigar sua importância nas componentes principais (CPs). A Tabela 6 apresenta os resultados.

Figura 6: Resultado numérico da análise por componentes principais (CP).

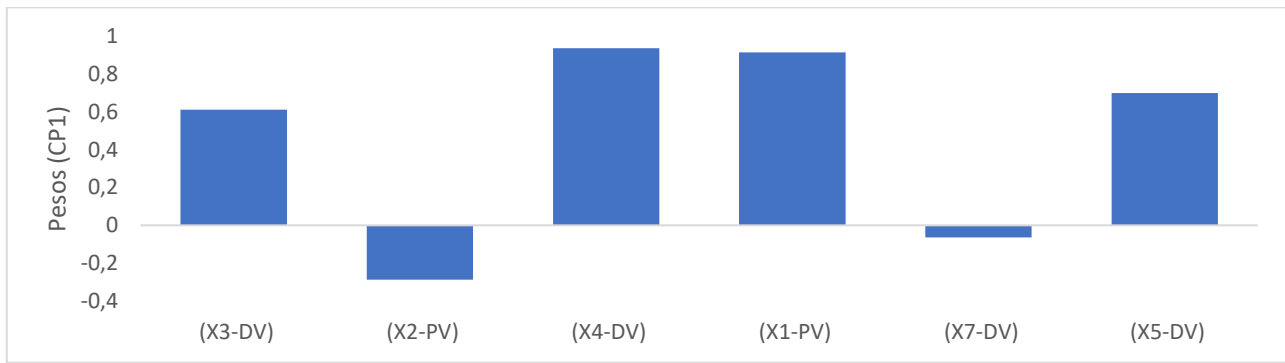
Variável	CP1	CP2	CP3	CP4	CP5	CP6
Nível Cx do Moinho (X3-DV)	0,6123	0,1006	0,3503	0,6931	0,1090	0,0017
Granulometria # 100 (X2-PV)	-0,2885	-0,5893	0,7288	-0,1870	0,0578	0,0035
Densidade da Polpa (X4-DV)	0,9385	-0,1680	-0,0297	-0,1552	-0,2433	0,0823
Teor de Sólidos (X1-PV)	0,9171	-0,2502	-0,0175	-0,1209	-0,2738	-0,0801
Eficiência do Moinho (X7-DV)	-0,0644	-0,8228	-0,4653	0,2347	0,2174	0,0044
Carga Circulante (X5-DV)	0,7006	0,1463	0,0138	-0,2949	0,6330	-0,0051
Autovalor	2,675	1,147	0,872	0,696	0,597	0,013
Variância Explicada (%)	44,583	19,109	14,531	11,601	9,956	0,221
Variância Explicada Acumulada (%)	44,583	63,692	78,223	89,824	99,779	100,000

Decidiu-se por empregar três componentes principais (CP), ou seja, CP1, CP2 e CP3, cuja capacidade conjunta de explicação da variância total dos dados originais é de aproximadamente 78%, um valor satisfatório em engenharia industrial. Portanto, pôde-se representar o conjunto de seis variáveis originais de leitura (Tabela 1), a partir de apenas três variáveis latentes (as componentes principais), sem perda significativa da informação. Essa possibilidade é útil por facilitar o manuseio e a interpretação dos dados, ou seja, de interpretação do sistema sob análise. Na sequência, com o objetivo de validação do modelo PCA, ou seja, de verificação de sua coerência com o sistema físico, realizou-se a sua interpretação.

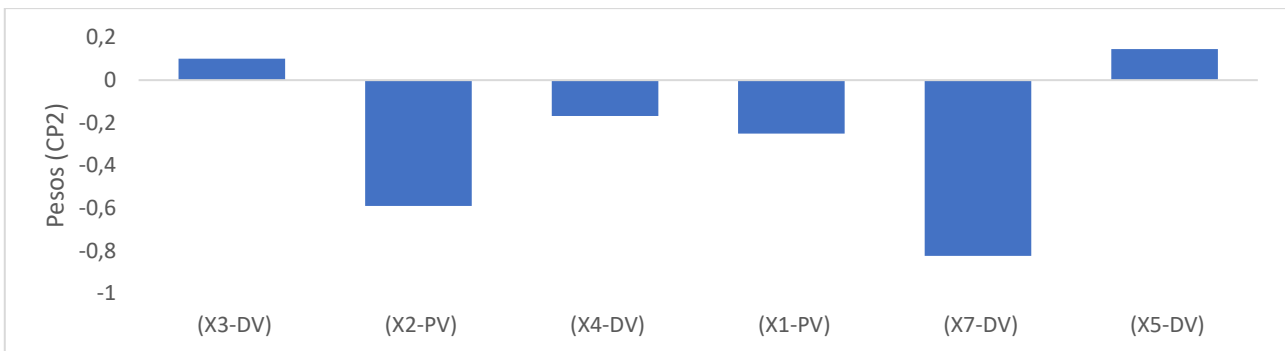
4.2 Etapa 2 – Interpretação e Validação do Modelo PCA

A Figura 7 mostra gráficos em barras para os pesos das componentes principais: CP1, CP2 e CP3. Inicialmente, pode-se observar que a CP1 é um indicador da quantidade de carga de material no sistema, devido aos pesos relativamente altos para as variáveis: nível da caixa do moinho (X3-DV), densidade de polpa (X4-DV), teor de sólidos (X1-PV) e carga circulante (X5-DV) (Figura 7a). Na estratégia de controle, esta componente, seria aquela observada para o controle da alimentação de matéria-prima; altos valores (os escores) de CP1 significam que o sistema está demandando baixos valores de vazão de alimentação (Y1-MV), e vice-versa, para baixos valores de CP1, um alto valor de vazão de alimentação será demandado pelo sistema. A componente CP2 é um indicador da eficiência do moinho (X7-DV) e da porcentagem de material retido na peneira de 100# (X2-PV) (Figura 2b). Para altos valores de eficiência do moinho e da porcentagem de material retido na granulometria de 100#, um baixo valor de CP2 será observado. Na estratégia de controle, um baixo valor de CP2 indicará um aumento de produtividade, levando o sistema de controle a aumentar a alimentação do moinho (Y1-MV) e a vazão de água da caixa de saída do moinho (Y2-MV). Como a qualidade de produção consiste em obter um material de granulometria 100%, ter uma maior assertividade dos indicadores dessa variável é fundamental. Como a seleção de apenas duas componentes principais captura apenas aproximadamente 64% da variância total do conjunto de dados, CP3 foi selecionada para aumentar a porcentagem de explicação do modelo. Além disso, a terceira componente principal auxilia a componente CP2 a explicar a variável granulometria 100# (X2-PV); para altos valores de porcentagem de material retido na peneira de granulometria 100#, um CP3 alto é estimado devido ao alto peso (positivo) dado a variável granulometria 100#. Esta coerência entre o modelo e o sistema é importante para a sua validação prática.

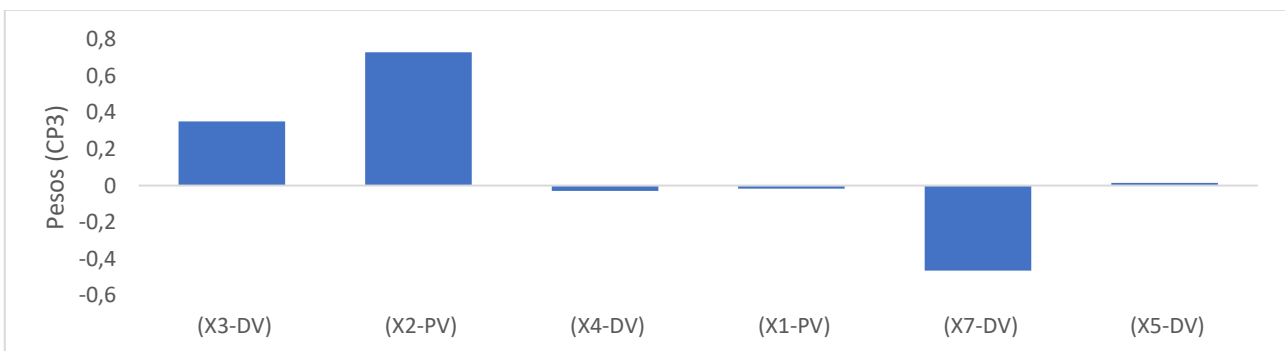
Figura 7: Pesos das componentes principais: **(a)** CP1, **(b)** CP2 e **(c)** CP3.



(a)



(b)

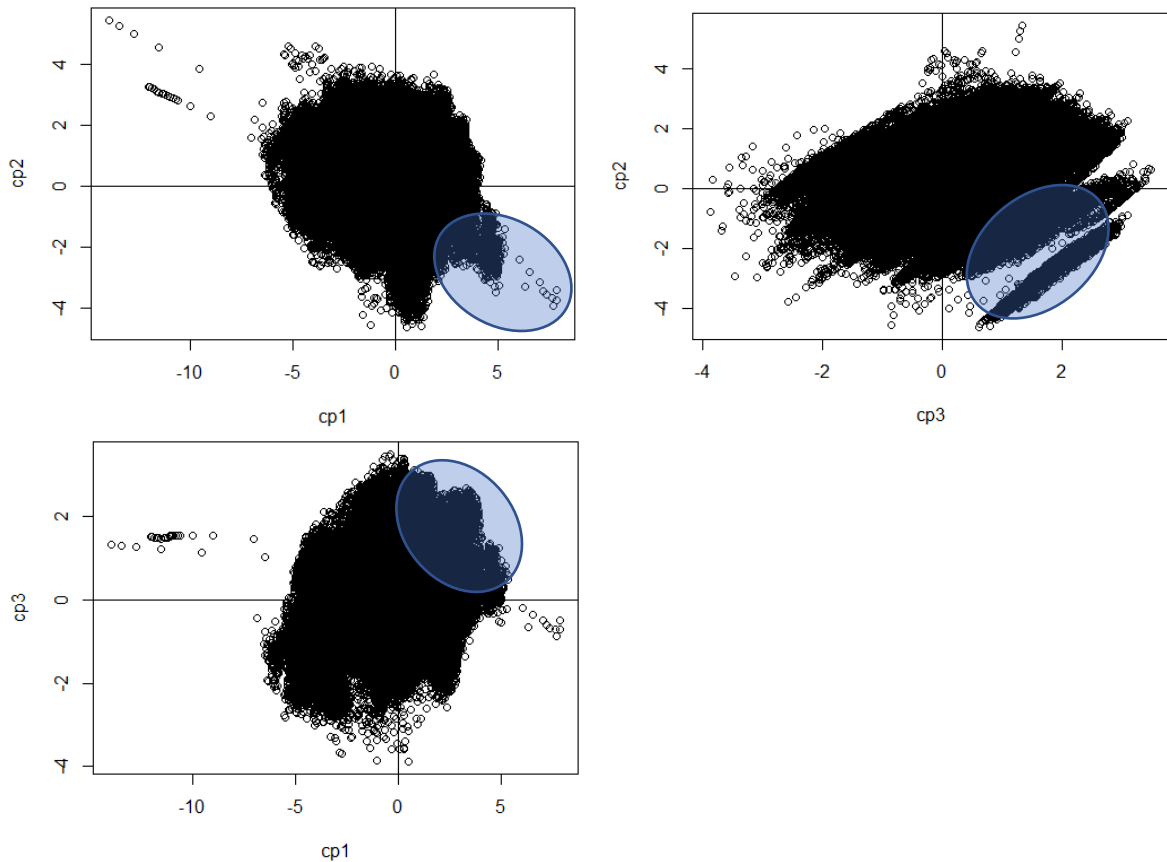


(c)

4.3 Etapa 3 – Definição de Regiões Operacionais Ótimas

Como já previamente discutido na Etapa 4.2, devemos operar este sistema com valores altos de CP1, baixos de CP2 e altos de CP3. Portanto, as regiões ótimas de operação são o quarto quadrante do gráfico CP1 vs CP2, o primeiro quadrante do gráfico CP1 vs CP3, e o quarto quadrante do gráfico CP2 vs CP3.

Figura 8: Gráfico de escores para as combinações entre as componentes principais: CP1 vs CP2, CP1 vs CP3 e CP2 vs CP3.



4.4 Etapa 4 – Simulação de Melhoria da Implementação da Técnica no Software Leaf

Utilizando a técnica de PCA no banco de dados deste trabalho, com os objetivos de minimizar o número de variáveis de leitura e de estimar regiões ótimas de operação do processo, foi realizada uma simulação de controle do processo pelo *dashboard* do *Leaf*. A comparação das variáveis de qualidade do processo antes e depois da atuação do *software*, ao se empregar as componentes principais (Figura 6) ao invés das variáveis originais (Tabela 1), é apresentado, de modo visual, na Figura 9, enquanto que a Tabela 2 apresenta o resultado de modo quantitativo.

Figura 9: Histograma da Taxa de Alimentação (Y1-MV) com o software *Leaf On e Off*.

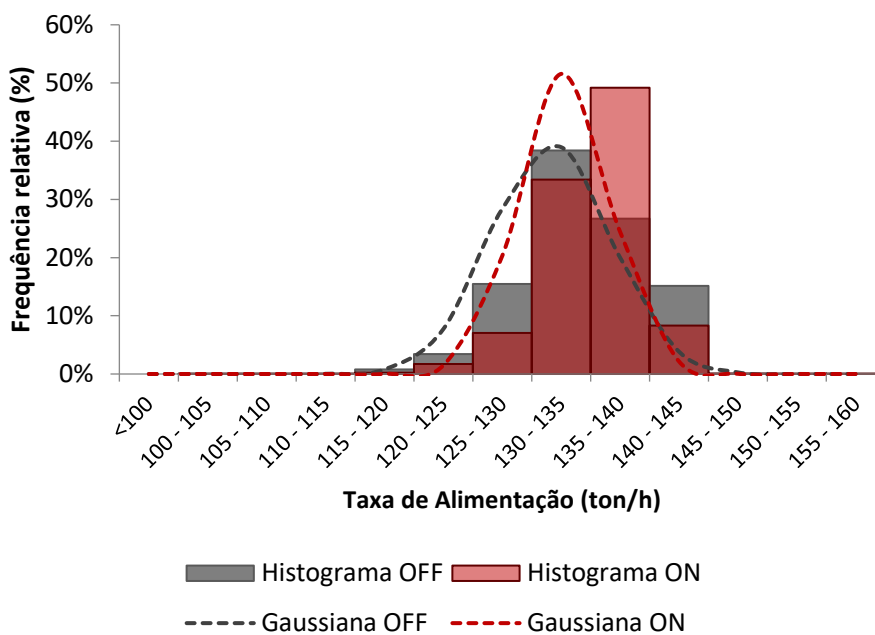


Figura 10: Histograma da porcentagem de Teor de Sólidos (X1-PV) com o software *Leaf On e Off*.

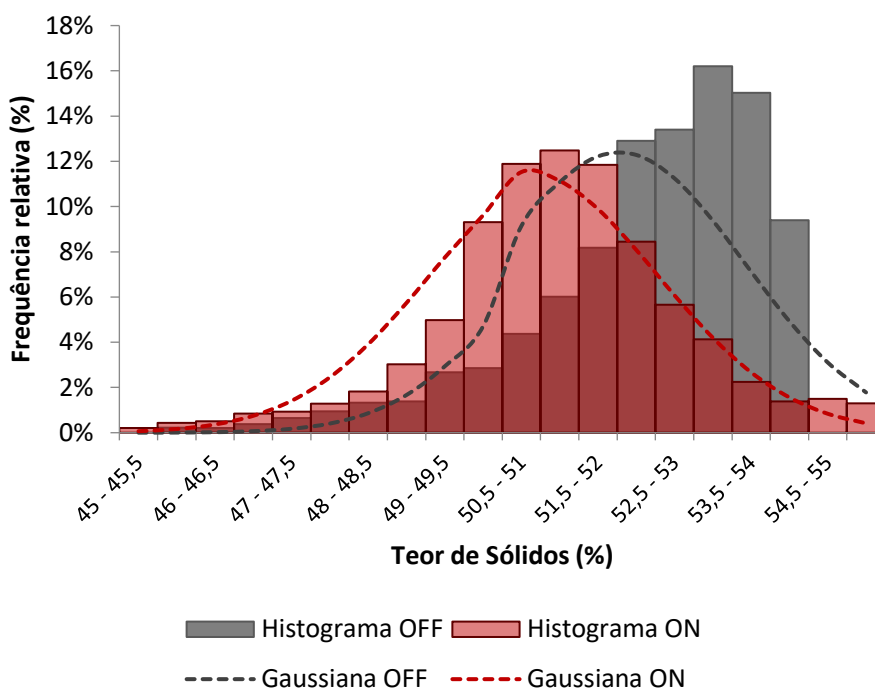


Tabela 2: Estatística descritiva das variáveis de processo: Alimentação do moinho (Y1-MV), Nível da Caixa de Saída do Moinho (X3-DV) e Teor de Sólidos da Polpa (X1-PV), antes e depois da atuação do *software Leaf*, com o emprego das componentes principais (On) e das variáveis originais (Off).

	Vazão de alimentação do moinho (ton/h)		Nível da caixa de saída do moinho (%)		Teor de sólidos da polpa (%)	
	LEAF OFF	LEAF ON	LEAF OFF	LEAF ON	LEAF OFF	LEAF ON
Média	134,05	135,28	65,63	74,34	51,91	50,56
Desvio-padrão	5,36	3,86	16,49	12,71	2,03	1,70

Pode-se observar nas Figuras 9 e 10, e na Tabela 2, que, após a simulação da atuação do *software Leaf*, utilizando as componentes principais: CP1, CP2 e CP3, como variáveis de leitura, houve um aumento médio de 1,23 ton/h na alimentação do moinho; uma redução média de 1,35% na média do teor de sólidos da polpa e de 16% em sua variabilidade; e uma redução média de 23% na variabilidade do nível da caixa de saída do moinho. Monetizando apenas o ganho com o aumento da taxa de alimentação para este sistema de beneficiamento mineral, um aumento de 1,23 ton/h na moagem média representa um ganho de aproximadamente R\$ 500 mil/ano, com base nos valores de produção dado pela mineradora, que disponibilizou o banco de dados para este trabalho.

5. Conclusões

Investigou-se a redução do número de variáveis e a estimação de regiões ótimas de operação, em processos de beneficiamento de minério, via análise por componentes principais (PCA). De modo a alcançar percepções práticas, ilustrou-se a aplicação comum caso industrial real. Após o entendimento do processo e da estratégia de controle realizada pelo *software Leaf*, foi identificado e validado um modelo PCA, que minimizou em 50% o número de variáveis a serem processadas pelo *software*. Foi possível também, a determinação de indicadores e de seus possíveis valores ótimos de operação, com o consequente uso dessa região ótima de operação como *set-point* para a atuação do *software Leaf* na estratégia de controle avançado de processos do sistema. Portanto, pôde-se atestar a aplicabilidade da técnica estatística multivariada, denominada de Análise por Componentes Principais (PCA), para a minimização do número de variáveis de processo e

a identificação e determinação de regiões ótimas de operação em processos de beneficiamento mineral. A sua extensão para outros equipamentos e processos industriais é direta.

6. Agradecimentos

Os autores agradecem a I.Systems e a mineradora pela cessão do banco de dados sobre as operações de um de seus moinhos de beneficiamento de minério.

7. Referências

AMARAL, L. M. **O conceito de reindustrialização, indústria 4.0 e a política industrial para o século XXI - O caso português.** Política industrial para o século XXI. Confederação Empresarial de Portugal. Ciclo de debates CIP, 2016.

I.SYSTEMS. **Consultoria em Controle Avançado de Processos.** Campinas, SP. Disponível em: <http://www.i.systems.com.br/>. Acesso em: 07/01/2018.

JACKSON, J. E. **A user's guide to principal components.** New York: John Wiley & Sons, 2003.

JASPERNEITE J. **Was hinter Begriffen wie Industrie 4.0 steckt.** In **Computer & Automation.** Massachusetts, EUA. Harvard Business Review. Dezember 2012.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Belo Horizonte: Editora UFMG, 2005.

MONTGOMERY, D.C. **Introdução ao controle estatístico da qualidade.** 7 ed., Rio de Janeiro: LTC, 2016.

MORRISON, D.F. **Multivariate statistical methods.** 2 ed., New York: McGraw Hill, 1976.

PORTER, M. E.; HEPPELMANN, J. E. **How Smart, Connected Products Are Transforming Companies.** Massachusetts, EUA. Harvard Business Review, pp. 96–112, 114, October 2015.

SCHWAB, K. **A quarta revolução industrial.** World Economic Forum, São Paulo, SP. Edipro, 1a ed, 2016.

ANEXO – Script em R

```
setwd("C:/Users/pedro/Desktop/Pesquisa/Modelo PCA")
library(readxl)
dados <- read_excel("C:/Users/pedro/Desktop/Pesquisa/dados.xlsx")
#normalizar a matriz de dados
dataN <- scale(dados,center = TRUE,scale=TRUE)
#criar data.frame de dataN para aplicar a função PCA do pacote Mvar.pt
dataN <- data.frame(dataN)
#Carregar a biblioteca do pacote Mvar.pt
library(Mvar.pt)
#aplicando a função PCA
data2<-dataN[,c(3:8)]
pca <- PCA(data2, Type = 2)
MatrixMC <- pca$MatrixMC #Matriz de covariância ou de correlação|Depende do
"Type": 1=covariância, 2=correlação
MatrixCCP <- pca$MatrixCCP #Matriz da correlação das componentes principais com
as variáveis originais
MatrixCCP <-data.frame(t(MatrixCCP))
MatrixAutoVlr <-pca$MatrixAutoVlr # Matriz de autovalores (variâncias) com as
proporções e proporções acumuladas
MatrixAutoVlr <- data.frame(MatrixAutoVlr)
MatrixEsc <- pca$MatrixEsc #Matriz com os escores dos componentes principais.
MatrixEsc
#Gráficos para a determinação de regiões ótimas de peração.
cp1=MatrixEsc[,1]
cp2=MatrixEsc[,2]
cp3=MatrixEsc[,3]
#PC2vsCP1
plot(cp1,cp2)
abline(v=0)
abline(h=0)
#PC2vsCP3
plot(cp3,cp2)
abline(v=0)
abline(h=0)
#PC3vsPC1
plot(cp1,cp3)
abline(v=0)
abline(h=0)
```