

Análise bibliométrica das cartas de controle multivariadas não paramétricas e dos métodos KNN, Kernel e Cópulas**Bibliometric analysis of nonparametric multivariate control charts, KNN, Kernel and Copulas**

DOI:10.34117/bjdv6n11-449

Recebimento dos originais: 03/10/2020

Aceitação para publicação: 20/11/2020

Tamires Fernanda Barbosa Nunes

Bacharel em Engenharia de Produção pela Universidade Federal de Pelotas
Instituição: Universidade Federal de Pelotas, Centro de Engenharias
Endereço: Rua Benjamin Constant, 989 – Porto, Pelotas - RS, Brasil - 96010-020
e-mail: tamiresfbnunes@gmail.com

Ariane Ferreira Porto Rosa

Doutora em *Automatique et Informatique Appliquées (Spécialité Génie Industriel Qualité)*, pela *Université de Nantes*, França.

Instituição: Universidade Federal de Pelotas, Centro de Engenharias
Endereço: Rua Benjamin Constant, 989 – Porto, Pelotas - RS, Brasil - 96010-020
e-mail: afprosa61@gmail.com

Rogério Royer

Doutor em Administração pela Universidade Federal do Rio Grande do Sul
Instituição: Universidade Federal de Pelotas, Centro de Engenharias
Endereço: Rua Benjamin Constant, 989 – Porto, Pelotas - RS, Brasil - 96010-020
e-mail: rogyroyer@ufrgs.br

RESUMO

As Cartas de Controle Multivariadas Paramétricas são utilizadas em processos produtivos que contenham um grande número de variáveis e correlação entre elas. Entretanto, para serem implementadas necessitam do conhecimento prévio sobre a distribuição das variáveis em estudo, suposição que muitas vezes não é satisfeita dependendo da complexidade do processo. As Cartas de Controle Multivariadas Não Paramétricas (CCMNP) se tornam uma alternativa para processos produtivos atuais, que apresentam alto grau de complexidade devido à evolução tecnológica, desenvolvimento industrial, automatização dos processos e características dos dados presentes nos processos. O presente estudo visa realizar uma revisão literária sobre a disseminação das CCMNP, dos métodos K-vizinhos mais próximos (KNN), Kernel e Cópulas, suas aplicações e adaptações nos diversos segmentos industriais. Além disso, é apresentado um estudo e análise de relevância do tema proposto por meio da realização de uma análise bibliométrica, utilizando o método de coocorrência de palavras para avaliar a expansão dos termos no período de 2006 a 2016. Por meio da análise bibliométrica pode-se observar que os métodos não paramétricos para construção de cartas de controle multivariadas são recentes, sendo sua expansão, desenvolvimento de novas aplicações e adaptações de extrema relevância para acompanhar as demandas industriais atuais.

Palavras-chave: Cartas de Controle Multivariadas, Não Paramétricas, Análise Bibliométrica.

ABSTRACT

Parametric Multivariate Control Charts are used in production processes that contain a large number of variables and correlation between them. However, to be implemented, they need prior knowledge about the distribution of the variables under study, an assumption that is often not satisfied depending on the complexity of the process. Non-parametric Multivariate Control Charts (CCMNP) become an alternative for production processes current, which present a high degree of complexity due to technological evolution, industrial development, process automation and data characteristics present in the processes. The present study aims to carry out a literary review on the dissemination of CCMNP, of the nearest K-neighbor methods (KNN), Kernel and Copulas, their applications and adaptations in the various industrial segments. In addition, a study and analysis of the relevance of the proposed theme is presented through a bibliometric analysis, using the word co-occurrence method to assess the expansion of terms in the period from 2006 to 2016. Through bibliometric analysis it can be seen that non-parametric methods for building multivariate control charts are recent, being its expansion, development of new applications and adaptations of extreme relevance to keep up with current industrial demands.

Keywords: Multivariate Control Charts, Nonparametric, Bibliometric Analysis.

1 INTRODUÇÃO

A qualidade sempre foi um fator determinante para o sucesso de produtos e serviços. Desta forma a busca pela melhoria da qualidade apresenta desenvolvimento evolutivo contínuo, para que assim consiga atender as demandas oriundas das necessidades dos clientes, do desenvolvimento tecnológico e industrial e da crescente competitividade enfrentada pelas empresas. A competitividade no meio industrial possui crescimento constante, seja pelo alto grau de desenvolvimento tecnológico atual ou pela globalização, o que faz com que as empresas se empenhem na busca para excelência de todos seus processos (SILVEIRA, WERNER, 2011).

Silva et al. (2007) considera que à medida que a performance de máquinas e equipamentos aumenta, aumenta também os requisitos em projetos de produtos, exigindo cada vez mais exatidão dos projetos, em sua execução e maior controle durante a linha de produção. Com isso, ferramentas de Controle Estatístico do Processo (CEP) são fundamentais para avaliar o desempenho dos processos e manter o nível de qualidade exigido para que o produto seja entregue ao cliente com excelência.

O Controle Estatístico do Processo, introduzido na década de 20, realizado por meio de gráficos de controle, é uma das principais ferramentas de monitoramento e controle que buscam a redução da variabilidade dos processos. Inicialmente as abordagens de Cartas de Controle Paramétricas, tanto univariadas quanto multivariadas, na qual a distribuição da variável sob análise é conhecida, atendiam as características e necessidades dos processos. À medida em que novas tecnologias são incorporadas nos processos, a quantidade de características a serem monitoradas aumenta, assim como a complexidade de execução dos processos. Neste contexto, as Cartas de Controle Não Paramétricas,

nas quais os dados de aprendizagem são considerados sem nenhuma suposição de distribuição das variáveis, tornaram-se uma opção mais robusta para o CEP.

O presente estudo é impulsionado pela busca, análise e compreensão de ferramentas estatísticas utilizadas no controle estatístico de processos multivariados, com particular interesse no estudo das Cartas de Controle Multivariadas Não Paramétricas e nos métodos não paramétricos K-vizinhos mais próximos, Kernel e Cópulas. O estudo foi desenvolvido por meio de uma pesquisa bibliográfica e análise bibliométrica contendo o estudo e análise das publicações recentes do tema proposto, relativo aos últimos 10 anos (2006-2016), para melhor entender a relevância, importância e sua influência na busca da melhoria contínua dos processos.

2 REVISÃO DA TEORIA

A revisão teórica apresentada visa fazer uma breve explanação sobre a fundamentação teórica acerca das Cartas de Controle Multivariadas Não Paramétricas e dos métodos K-vizinhos mais próximos, Kernel e Cópulas.

2.1 CARTAS DE CONTROLE MULTIVARIADAS NÃO PARAMÉTRICAS

As Cartas de Controle Multivariadas Não Paramétricas (CCMNP) abrangem técnicas que não pressupõem que determinado conjunto de dados siga alguma distribuição particular. Segundo He e Wang (2007) as CCMNP são construídas a partir dos dados de aprendizagem do processo, sem nenhuma suposição sobre a distribuição das variáveis.

Para Verdier e Rosa (2011) a implementação de cartas de controle multivariadas paramétricas deixa implícito que determinado conjunto de observações segue uma distribuição de probabilidade normal multidimensional. Porém em processos industriais a hipótese de normalidade geralmente pode ser de difícil verificação, tornando as CCMNP uma alternativa para o controle e monitoramento do desempenho destes processos. A aplicação de procedimentos não paramétricos ou livres de distribuição possui a vantagem de não requerer nenhuma hipótese sobre a distribuição das variáveis em estudo.

As CCMNP são úteis quando não queremos que a função de probabilidade seja modelada parametricamente, seja devido à falta de conhecimento de sua verdadeira distribuição de probabilidade ou para que seja verificado o modelo paramétrico adotado.

2.2 K – VIZINHOS MAIS PRÓXIMOS

A regra k vizinhos mais próximos (KNN, do inglês, K Nearest Neighbors rule) é um método de classificação supervisionada, no qual um objeto novo é classificado por meio da análise das distâncias de amostras de treinamento de vizinhos mais próximos num determinado espaço de características.

Assim amostras sem rótulos são classificadas com base em suas semelhanças com as amostras de um conjunto de treinamento. A regra de classificação supervisionada KNN é um método de classificação baseado na similaridade, no qual a observação tende a pertencer a uma mesma classe (VERDIER, ROSA, 2011).

O método de classificação de amostras KNN têm sido amplamente utilizado em aplicações na área de mineração de dados, reconhecimento de padrões, no processamento de imagens, entre outros. No âmbito industrial a regra KNN tem sido aplicada na detecção de falha nos processos. Entretanto, a regra KNN apresenta algumas limitações quando aplicada no controle e detecção de falhas, pois no método de classificação as amostras falhas são definidas com antecedência para que novas amostras sejam rotuladas a partir delas, o que não ocorre no caso de detecção de falhas, por não se ter um conhecimento prévio das falhas a serem detectadas. Por isso a regra KNN não pode ser aplicada de forma direta nesses casos.

A distância Euclidiana é a métrica mais usual para aplicação da regra KNN. Porém, outras escolhas são possíveis, visando uma representação apropriada dos dados. No contexto de classificação, a escolha da distância para os k vizinhos mais próximos é a questão mais importante. A escolha correta da métrica de distância a ser utilizada pode melhorar significativamente o desempenho do mesmo (VERDIER, ROSA, 2011). Diversas aplicações propõem adaptações da métrica utilizada pela regra KNN para detecção de falhas a fim de melhorar seu desempenho de previsão das amostras como a de Biau et al. (2012), Zhou e Chen (2006) e Wang et al. (2007).

A literatura apresenta diversas aplicações do método como as publicações de Magnussen et al. (2010) que propõem um método de calibração para aumentar a acurácia das previsões realizadas pela técnica do k - vizinho mais próximo, Gharun et al. (2015) que testam métodos para previsão do rendimento de água após um incêndio na área florestal, Towler et al. (2008) que aplicam a técnica k -vizinho mais próximo para contornar os problemas causados pela variabilidade significativa na qualidade da água superficial, Holsbach et al. (2014) que utilizam a regra KNN e a ferramenta Análise discriminante para categorizar observações em um estudo de seleção de variáveis para o diagnóstico precoce de câncer de mama, He e Wang (2007) e Verdier e Rosa (2011) que utilizam o método KNN em processos de fabricação de semicondutores.

2.3 KERNEL

A estimativa de densidades Kernel é um método para estimação da função densidade de probabilidades de uma variável aleatória, no qual as densidades são estimadas através de um núcleo, denominado Kernel. O método de estimação Kernel é uma técnica não paramétrica capaz de estimar

curvas de densidades, onde as distâncias em relação ao valor central (núcleo) realizam a ponderação para cada observação.

Segundo Chapelle et al. (2006) os métodos não paramétricos Kernel estão sendo cada vez mais utilizados, devido a simplicidade de seu conceito, suas propriedades teóricas e o bom desempenho de suas aplicações, se mostrando atraente quando direcionada especificamente à aprendizagem semi-supervisionada.

A funcionalidade de sua aplicação é demonstrada em trabalhos como de He et al. (2016) que apresentam uma abordagem de controle de processo multivariável, na qual o monitoramento do processo é aplicado à problemas de classificação não linear, Wang e Yao (2015) que utilizam o método Kernel para resolver os problemas de correlações não lineares em dados de processos em lotes normais e defeituosos, Khediri et al. (2011) que utilizam a técnica para o controle on-line de processos multivariáveis não lineares não estacionários, o modelo é aplicado a uma simulação de dois processos de benchmark.

Outras publicações com aplicações do método de estimação Kernel são apresentadas por Theljani et al. (2015) que aplica o algoritmo Kernel para o monitoramento de processos químicos, Marcondes Filho et al. (2011) que utilizam o método Kernel para o monitoramento de processos industriais em batelada, Hoffmann (2006) que utiliza a abordagem Kernel ACP (Análise de Componentes Principais) em dados bidimensionais e em dados reais de manuscritos e mamacitologia do câncer, Choi et al. (2008) que utiliza o monitoramento multi escala baseado em Kernel para detecção de falhas e em Fan et al. (2014) que utilizam a técnica Kernel no monitoramento dos processos com qualificação de filtração.

2.4 CÓPULAS

Segundo Moretin (2011, v.2, p.329) “a teoria de cópulas teve um desenvolvimento explosivo nos últimos anos, notadamente nas aplicações em estatística, finanças, gestão de riscos e modelagem de dependência em carteiras de seguro”. Métodos baseados em Cópulas permitem a modelagem da dependência dos dados separadamente, em densidades marginais e em dependência conjunta. Como a distribuição das variáveis nem sempre é conhecida devido à dimensão dos dados e complexidade das variáveis envolvidas nos processos industriais, muitas vezes se torna necessário modelar a dependência das variáveis para monitorar o desempenho dos processos. A modelagem de cópulas é uma ferramenta estatística de análise multivariável muito eficiente, com diversas aplicações em finanças, ciências atuariais, entre outras (VERDIER, 2013).

Aplicações da modelagem de cópulas pode ser observadas nos trabalhos de Verdier (2013) que utiliza cópulas aplicada a casos multivariáveis não normais, Chorós et al. (2010) que apresentam uma

pesquisa expondo os procedimentos de estimação de modelos cópulas, Danaher e Smith (2011) que utilizam os modelos de cópulas de probabilidade multivariada aplicados à literatura de marketing, Chen et al. (2017) que aplicam a abordagem cópulas para estimar a confiabilidade do tempo de viagem em artérias urbanas, Grimaldi et al. (2016) que realizam um estudo não paramétrico através de cópulas sobre estruturas de dependência das respostas hidrológicas e em Durante e Okhrin (2015) que utilizam modelos baseados em cópulas intercambiáveis aplicados à hidrologia.

3 PROCEDIMENTOS METODOLÓGICOS

O presente estudo consiste em uma pesquisa exploratória, fundamentada por uma revisão bibliográfica teórica na qual foram levantadas publicações recentes relacionadas as Cartas de Controle Multivariadas Não Paramétricas e dos métodos K-vizinhos mais próximos, Kernel e Cópulas. Além da revisão da literatura uma análise bibliométrica foi desenvolvida para compreender a relevância, importância e sua influência na busca da melhoria contínua dos processos.

A bibliometria consiste na aplicação de técnicas estatísticas à produção bibliográfica, caracterizando-se como produto de interdisciplinaridade que descreve o pensamento científico e o pensamento humanístico atualmente (FONSECA, 1986). A bibliometria permite a visualização da situação atual da produção científica, identificando a existência de vácuos que possam vir a ser preenchidos com a criação de novas pesquisas científicas ou a análise bibliométrica pode servir de ponto de partida para novos estudos (RODRIGUES et al., 2016).

Segundo Bellis (2009) a avaliação quantitativa de materiais de cunho bibliográfico, remete a ideia de tangibilidade do progresso científico, assim a análise estatística das publicações é capaz de dimensionar a ciência e seu progresso (BELLIS, 2009). Por meio da análise bibliométrica podemos acompanhar o desenvolvimento das contribuições científicas publicadas ao longo do tempo considerando-as como indicador de apoio a tomada de decisão para prospecções científicas e tecnológicas.

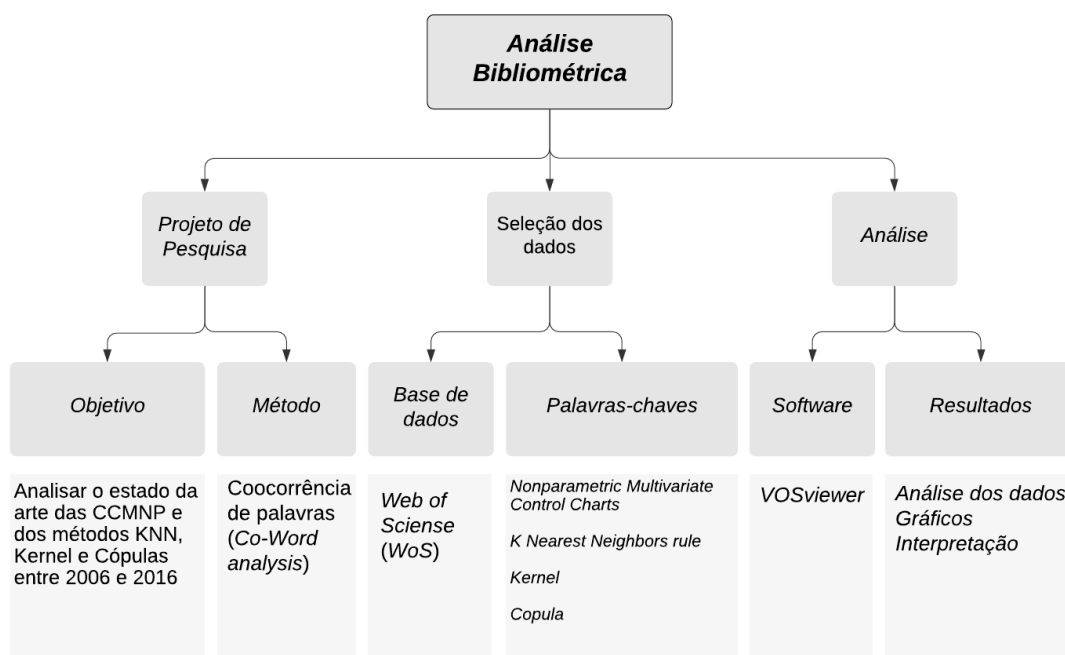
Para realização de uma análise bibliométrica primeiramente é necessário definir o objetivo da pesquisa e o método apropriado para que estes sejam alcançados. Para seleção dos dados é preciso definir a base de dados a ser utilizada na busca e coleta de artigos. Após estratégias de busca para coleta das publicações científicas devem ser definidas. Para realização da análise bibliométrica é preciso definir um *software* a ser utilizada para análise dos dados, para então realizar a análise bibliométrica propriamente dita. A metodologia proposta para desenvolvimento da análise bibliométrica é apresentada na Figura 1.

Tendo em vista as demandas atuais dos processos industriais que necessitam de técnicas estatísticas de controle de qualidade robustas para acompanhar sua complexidade, a presente análise

bibliométrica visa analisar o estado da arte das Cartas de Controle Multivariadas Não Paramétricas e dos métodos KNN, Kernel e Cópulas por meio da análise de publicações recentes, considerando o período de 2006 a 2016.

A análise das publicações científicas por meio da análise bibliométrica foi realizada a partir da análise de coocorrência de palavras, de forma a quantificar a frequência das palavras Cartas de Controle Multivariadas Não Paramétricas, k vizinhos mais próximos, Kernel e Cópulas nos artigos científicos coletados. A análise de coocorrência pressupõe que as palavras que se repetem em diversos documentos possuem conceitos relacionados.

Figura 1. Metodologia utilizada para desenvolvimento da análise bibliométrica



Fonte: Elaborado pelos autores

4 RESULTADOS

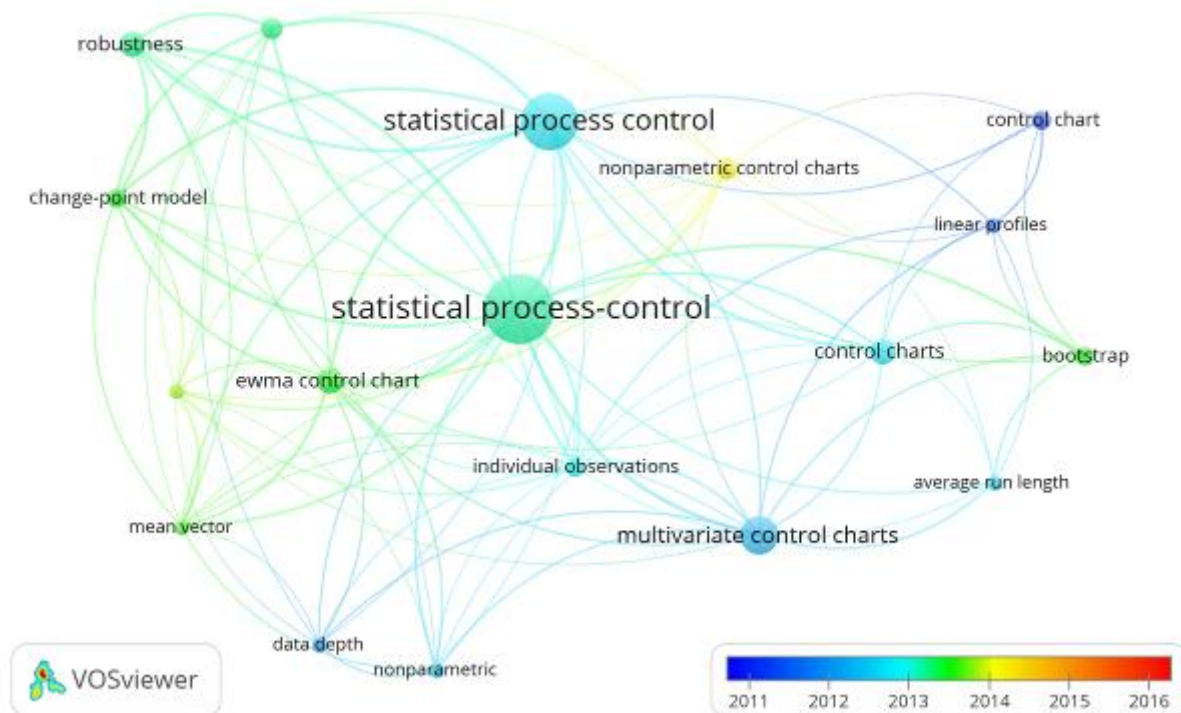
A análise bibliométricas por meio da técnica de coocorrência de palavras foram realizadas de forma individual para Cartas de Controle Multivariadas Não Paramétricas e dos métodos KNN, Kernel e Cópulas. Os resultados das análises realizadas no *software VOSviewer* são apresentadas a seguir.

A análise gráfica dada pelo *software VOSviewer* permite visualizar o número de palavras-chave definidos pelo *software* por meio de clusters (grupos de afinidades), onde cada esfera representa uma palavra-chave e cada *cluster* é representado por uma cor.

O número total de palavras-chave contabilidade para as Cartas de Controle Multivariadas Não Paramétricas foi de 302, definindo-se um mínimo de ocorrência de 5 vezes obteve-se como resultado

que 18 termos/palavras se repetiram no mínimo 5 vezes em cada publicação. Por meio da análise gráfica é possível verificar o ano médio de publicação de cada palavra-chave. A escala representa a incidência das publicações de 2011 a 2016, pois os artigos analisados não apresentaram incidência nos anos anteriores. A Figura 2 nos permite observar que as publicações tiveram maior incidência entre 2012 e 2013.

Figura 2. Análise bibliométrica para o termo Cartas de Controle Multivariadas Não Paramétricas

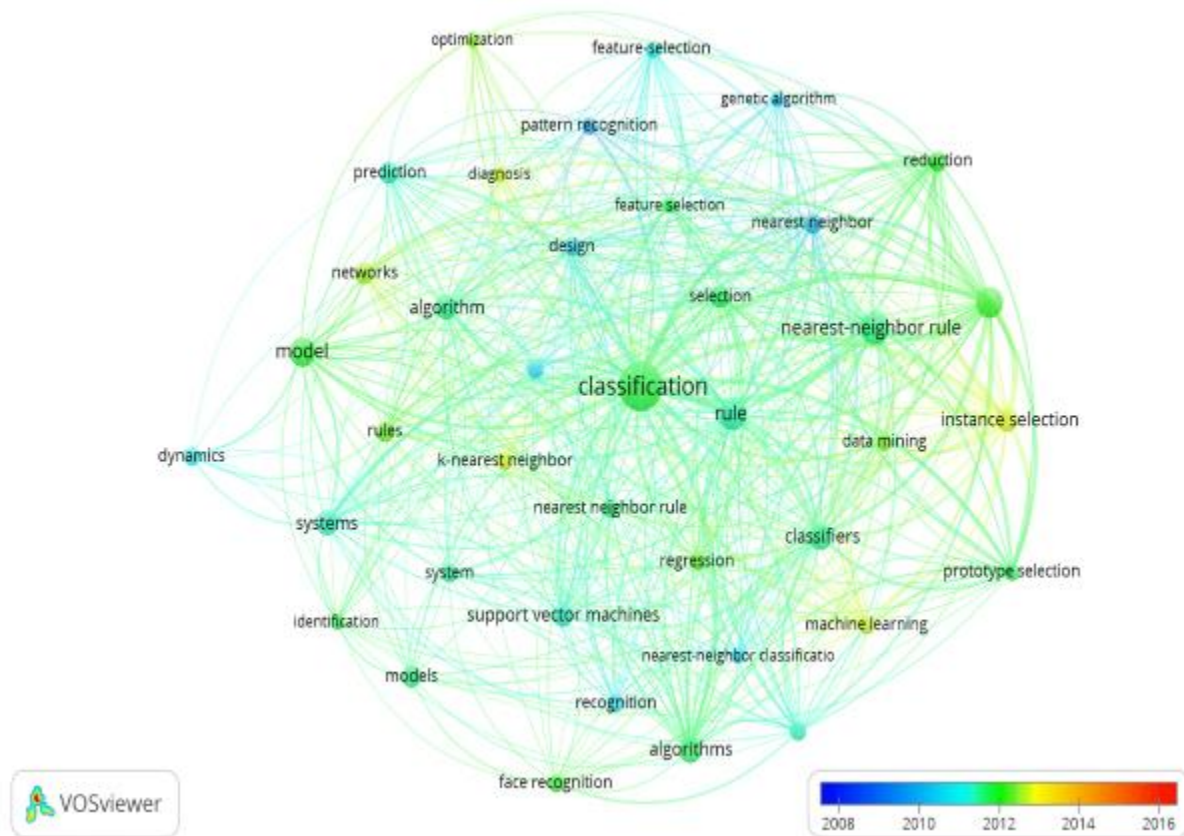


Fonte: Elaborado pelos autores

Os termos que tiveram maior destaque neste período e que apresentam forte relação com a área na qual a pesquisa tem foco foram: as duas representações de controle estatístico do processo (em inglês, *statistical process control*) e cartas de controle multivariadas (em inglês, *multivariate control charts*).

A análise bibliométrica para a regra K-vizinhos mais próximos contabilizou 6.028 palavras-chave. Definindo-se um mínimo de ocorrência de 25 vezes obteve-se como resultado que 39 termos/palavras se repetiram no mínimo 25 vezes. Estes 39 termos foram divididos em quatro clusters, conforme afinidades entre as palavras. A Figura 3 apresenta a análise gráfica para os termos KNN, observa-se pela escala temporal que os termos apresentam maior incidência nas publicações entre 2011 e 2012, e o termo regra K- vizinho mais próximo (em inglês, *nearest neighbor rule*) passa a ter incidência nas publicações a partir de 2012.

Figura 3. Análise bibliométrica para o termo KNN



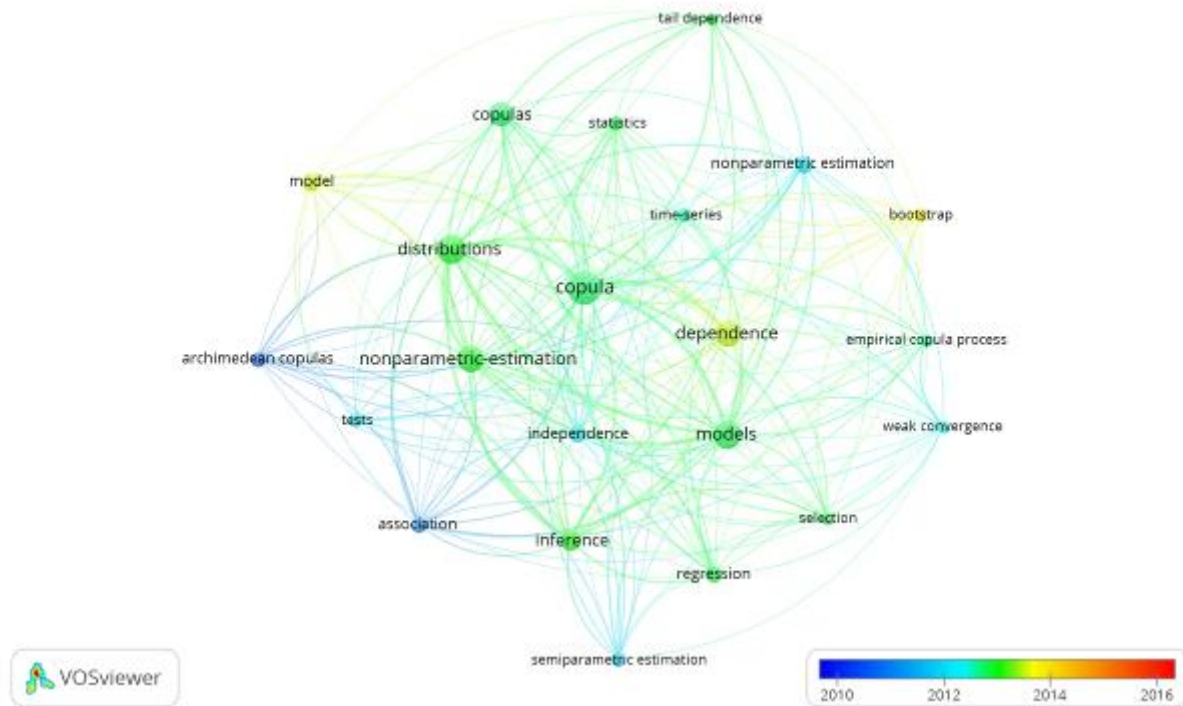
Fonte: Elaborado pelos autores

Considerando o período de 2006 a 2016 observa-se que os termos que tiveram maior destaque neste período e que apresentam forte relação com a área na qual a pesquisa tem foco foram: classificação (em inglês, *classification*) e regra K- vizinho mais próximo (em inglês, *nearest-neighbor rule*).

Durante a análise dos artigos referentes ao termo Kernel o *software* contabilizou 4.752 palavras-chave, definindo um mínimo de ocorrência de 20 vezes teve-se como resultado que 49 termos/palavras se repetiram no mínimo 20 vezes. Estes 49 termos foram divididos em quatro clusters, conforme afinidades entre as palavras. É possível observar que a incidência dos termos relacionados a palavras Kernel ocorre com maior frequência entre 2009 e 2013, Figura 4.

Os termos que tiveram maior destaque neste período e que apresentam forte relação com a área na qual a pesquisa tem foco foram: estimação de densidade (em inglês, *density-estimation*), estimação de densidade não paramétrica (em inglês, *nonparametric density estimation*) e regressão não paramétrica (em inglês, *nonparametric regression*).

Figura 5. Análise bibliométrica para o termo Cópulas



Fonte: Elaborado pelos autores

A Tabela 1 apresenta em síntese alguns dados importantes considerados durante a análise, como o número de artigos encontrados para cada um dos termos pesquisados, o número total de palavras-chave contidos nos artigos de cada termo, a definição de mínimo de ocorrências para cada termo e o número dos termos que se repetiram conforme o mínimo estabelecido.

Tabela 1. Síntese dos resultados das buscas e análise bibliométrica

Termo pesquisado	Nº de artigos	Nº de palavras-chave	Nº mínimo de ocorrência das palavras-chave	Nº de termos repetidos
CCMNP	47	302	5	18
KNN	1.224	6.028	25	39
Kernel	1.093	4.552	20	49
Cópula	357	1.852	15	22

Fonte: Elaborado pelos autores

Pode-se observar que a pesquisa na base de dados *WOS* para os termos CCMNP resultou em 47 artigos, os 47 artigos totalizaram 302 palavras-chaves e definindo um mínimo de ocorrência de 5 vezes obteve-se 18 termos/palavras se repetiram no mínimo 5 vezes. A escolha do mínimo de ocorrência levou em consideração a frequência de termos repetidos, visando que estes fossem minimizados para

que a análise gráfica fosse composta por termos que possuíssem maior afinidade e não limitada a termos repetitivos.

A partir da Tabela 1, verifica-se que o termo com maior incidência de publicações no período de 2006 a 2016 foi o termo KNN e Kernel. Enquanto as publicações com o termo CCMNP e Cópula tiveram menor incidência.

Os resultados da análise bibliométrica demonstraram os termos que tiveram maior destaque entre 2006 a 2016, e sua evolução ao longo deste período. Como visto os termos pesquisados se mostram recentes, as publicações do termo CCMNP datam o período de 2012 a 2014, a regra KNN 2010 e 2013, o método Kernel 2009 e 2014 e o método cópula 2010 e 2014.

5 CONCLUSÕES

O presente estudo propôs a realização de uma análise bibliométrica sobre o estado da arte das Cartas de Controle Multivariadas Não Paramétricas e os métodos K- vizinhos mais próximos, Kernel e Cópula. A análise teve como objetivo pesquisar publicações recentes, dos últimos 10 anos (2006-2016) referente aos termos, onde se buscou compreender sua disseminação ao longo dos anos e sua relevância como ferramentas estatísticas atuais com potencial para aplicações em diversas áreas principalmente como ferramentas estatísticas direcionadas a melhoria e otimização de processos.

A análise de coocorrência de palavras foi utilizada para analisar a expansão dos termos Cartas Multivariadas Não Paramétricas, K vizinhos mais próximos, Kernel e Cópulas. A base de dados *Web of Science* foi utilizada para busca das publicações. E a análise bibliométrica foi realizada por meio *software* de licença livre *VOSviewer*.

Além de tornar possível a visualização da expansão de um dado tema, a análise bibliométrica nos permite delinear tendências de mercado, identificar lacunas na evolução da pesquisa científica, além de proporcionar a identificação de nichos de mercado ou pesquisa promissores. A partir da análise pode-se visualizar a incidência de publicações de cada termo nos últimos dez anos. Assim como o ano de publicação média para cada termo. Pode-se observar que os termos objeto de análise são recentes, com incidência nas publicações científicas entre 2008 e 2015, ressaltando a atualidade das técnicas estatísticas.

Os métodos para implementação destas técnicas estão definidos na literatura, porém para que haja continuidade no desenvolvimento de estudos científicos sobre estas técnicas é necessário implementá-las e adaptá-las conforme as necessidades dos processos industriais, para que assim seus conceitos sejam expandidos a mais processos e a outros contextos. Fatores que podem justificar a limitação destas implementações são os recursos necessários para que estes métodos estatísticos sejam utilizados como uma base de dados atualizada sobre as variáveis do processo, investimento em

tecnologia para que se possa armazenar os dados e gerar o monitoramento e controle do desempenho dos processos de forma efetiva, além da qualificação dos funcionários para utilizar estas técnicas estatísticas robustas que são relativamente complexas. Entretanto, os recursos tecnológicos como *big data*, armazenamento em nuvem, internet das coisas vêm impulsionando o uso de técnicas estatísticas aplicadas a um grande número de dados.

As técnicas estão se expandindo, porém lentamente se comparadas com a evolução dos processos que aumentam sua complexidade rapidamente. A partir da análise bibliométrica o mapeamento da expansão dos temas foi realizado, demonstrando que são campos de pesquisa recentes. Que possuem lacunas em seu desenvolvimento, pois o pico de pesquisas científicas sobre os termos ocorreu entre os anos de 2008 e 2015.

A evolução de ferramentas estatísticas que melhor se adaptem as características dos processos atuais são essenciais para o desenvolvimento da qualidade, confiabilidade, otimização, aumento da eficiência e eficácia dos processos. Para trabalhos futuros há a possibilidade de ampliação dos artigos publicados, sem limitação de período, para abranger um maior número de registros para posterior análise bibliométrica, assim como pesquisa em outras bases de dados para comparar a frequência das publicações.

REFERÊNCIAS

- BELLIS, N. **Bibliometrics and citation analysis**. EUA: Scarecrow Press, 2009.
- BIAU, G. *et al.* **An affine invariant k -nearest neighbor regression estimate**. Journal of Multivariate Analysis, v.112, p. 24-34, 2012.
- CHAPELLE, O. *et al.* **Semi-supervised learning**. [S.I.]: MIT press, v.2. Cambridge, 2006.
- CHEN, M. *et al.* **A copula-based approach for estimating the travel time reliability of urban arterial**. Transportation Research Part C, v.28, p. 1-23, 2017.
- CHOI, S. W.; MORRIS, J.; LEE, I.B. **Nonlinear multiscale modelling for fault detection and identification**. Chemical Engineering Science, v. 63, p. 2252-2266, 2008.
- CHORÓS, B.; IBRAGIMOV, R.; PERMIAKOVA, E. **Copula Estimation**. Copula Theory and its Applications, p. 77-91, 2010.
- DANAHER, P. J.; SMITH, M. S. **Modeling multivariate distributions using copulas: Applications in marketing**. Marketing Science Celebrating 30 Years, v. 30, n. 1, p. 4-21, 2011.

DURANTE, F.; OKHRIN, O. **Estimation procedures for exchangeable Marsall copulas with hydrological application.** Stochastic Environmental Research and Risk Assessment, v. 29, p.205-226, 2015.

FAN, J.; QUIN, S. J.; WANG, Y. **Online monitoring of nonlinear multivariate industrial processes using filtering KICA-PCA.** Control Engineering Practice, v. 22, p. 205-216, 2014.

FONSECA, E N. **Bibliometria: Teoria e Prática.** São Paulo: Cultrix, 1986.

GHARUN, M.; AZMI, M.; ADAMS, M. A. **Short-term forecasting of water yield from forested catchments after bushfire: A case study from Southeast Australia.** Water, v. 7, 2015.

GRIMALDI, S. et al. **Catchment compatibility via copulas: A non-parametric study of the dependence structures of hydrological responses.** Advances in Water Resources, v. 90, p. 116-133, 2016.

HE, Q.P.; WANG, J. **Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes.** IEEE Transactions on semiconductor manufacturing, vol. 20, n. 4, 2007.

HE, S.; WEI, J.; DENG, H. **A distance-based control chart for monitoring multivariate processes using support vector machine.** Annals of Operations Research, 13 de abril de 2016. DOI: 10.1007/s10479-016-2186-4.

HOFFMANN, H. **Kernel PCA for novelty detection.** Pattern Recognition, v. 40, p. 863-874, 2007.

HOLSBACH, N.; FOGLIATTO, F. S.; ANZENELLO, M. J. **Método de mineração de dados para identificação de câncer de mama baseado na seleção de variáveis.** Ciência saúde coletiva, v. 19, n. 4, 2014.

KHEDIRI, I. B.; LIMAN, M.; WEIHS, C. **Variable window adaptative kernel principal componente analysis for nonlinear nonstationary process monitoring.** Computers & Industrial Engineering, v. 61, p. 437-446, 2011.

MARCONDES FILHO, D.; FOGLIATTO, F. S.; OLIVEIRA, L. P. L. **Gráficos de controle multivariados para monitoramento de processos não lineares em bateladas.** Produção, v. 21, n. 1, p. 132-148, 2011.

MAGNUSSEN, S.; TOMPPONEN, E.; MCROBERTS, R. E. **A model-assisted k-nearest neighbour approach to remove extrapolation bias.** Scandinavian Journal of Forest Research, v. 25, p. 174-184, 2010.

MORETIN, P. A. **Econometria financeira – Um curso em séries temporais financeiras.** 2 ed. São Paulo: Blucher, 2011.

RODRIGUES, M. C. N. **A influência da engenharia de produção nos serviços de atendimento à saúde: Estudo bibliométrico focado em técnicas operacionais.** Revista Produção Online, v. 16, n. 1, jan/mar. de 2016.

SILVA, W. V. **Garantia da qualidade do café solúvel com o uso do gráfico de controle de somas acumuladas.** Produção Online, v. 7, n. 2, Agosto 2007.

SILVEIRA, E. F.; WERNER, L. **Proposta de método de priorização de processos a serem monitorados no controle estatístico de processo: Uma aplicação em trocador de calor.** Revista Produção Online, v.11, n.1, mar. 2011.

THELJANI, F. **A new kernel-based classification algorithm for systems monitoring: Comparison with statistical process control methods.** Arabian Journal for Science and Engineering, v. 40, p. 645-658, 2015.

TOWLER, E. *et al.* **Simulating Ensembles of source water quality using a k-nearest neighbor resampling approach.** Environmental Science & Technology, v.43, p. 1407-1411, 2009.

VERDIER, G. **Application of copulas to multivariate control charts.** Journal of statistical planning and inference, v. 143, p. 2152-2159, 2013.

VERDIER, G.; ROSA, A.F.P. **Adaptive Mahalanobis Distance and k-Nearest Neighbor Rule for Fault Detection in Semiconductor Manufacturing.** IEEE Transactions on Semiconductor Manufacturing, v.24, n.1, p.59-68, 2011.

WANDERLEY, M. F. B. **Estudos de estimação de densidade por kernel: Métodos de seleção de características e estimação do parâmetro suavizador.** Tese. Programa de Pós-Graduação em Engenharia Elétrica – Universidade Federal de Minas Gerais, 2013.

WANG, H.; YAO, M. **Fault detection of batch processes based on multivariate functional kernel principal component analysis.** Chemometrics and Intelligent Laboratory Systems, v. 149, p. 78-89, 2015.

WANG, J.; NESKOVIC,P.; COOPER, L. N. **Improving nearest neighbor rule with a simple adaptive distance measure.** Pattern Recognition Letters, v. 28, p.207-213, 2007.

ZHOU, C. Y.; CHEN, Y. Q. **Improving nearest neighbor classification with cam weighted distance.** Pattern Recognition, v. 39, p. 635-645, 2006.