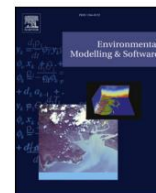


Listas de conteúdo disponíveis em [ScienceDirect](https://www.sciencedirect.com)

Modelagem Ambiental e Software

página inicial da revista: www.elsevier.com/locate/envsoft

Análise de séries temporais com variáveis explicativas: uma revisão sistemática da literatura



Paula Medina Maçaira, Antônio Marcio Tavares Thomé^y, Fernando Luiz Cyrino Oliveira, Ana Luiza Carvalho Ferrer

Pontifícia Universidade Católica do Rio de Janeiro, Industrial Engineering Department, Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, 22451-900, Brazil

INFORMAÇÕES DO ARTIGO

Palavras-chave:

Análise de regressão
Inteligência artificial
Variáveis exógenas
Cenários de previsão

RESUMO

A análise de séries temporais com variáveis explicativas engloba métodos para modelar e prever dados correlacionados levando em consideração informações adicionais, conhecidas como variáveis exógenas. Uma busca minuciosa na literatura retornou a escassez de revisões sistemáticas da literatura (SLR) sobre modelos de séries temporais com variáveis explicativas. O principal objetivo é preencher esta lacuna aplicando uma SLR rigorosa e reprodutível e uma análise bibliométrica para estudar a evolução desta área ao longo do tempo. O estudo resultou na identificação dos principais métodos de séries temporais que incorporam variáveis de entrada por área de conhecimento e metodologia. O maior número de artigos pertence às ciências ambientais, seguidas de economia e saúde. O modelo de regressão é o método com maior número de aplicações, seguido por Redes Neurais Artificiais e Máquinas de Vetor de Suporte, que tiveram um crescimento rápido e recente. Uma agenda de pesquisa em análise de séries temporais com variáveis exógenas encerra o artigo.

1. Introdução

A modelagem de séries temporais envolve a análise de um sistema dinâmico caracterizado por séries de entradas e saídas, que se relacionam com uma função. Independentemente de sua finalidade final, as diversas técnicas neste campo têm o objetivo mútuo de reproduzir a série de saída com confiabilidade e precisão a partir da estimação da função e da série de entrada. As técnicas de séries temporais podem ser essencialmente divididas em dois conjuntos de métodos: univariado e multivariado. No caso de abordagens univariadas, a série de saída é explicada por uma parcela constante e/ou tendência, sazonalidade e, em muitos casos, pela série defasada no tempo. Os métodos multivariados, por outro lado, utilizam a influência de outras variáveis no comportamento da série de saída para obter melhores resultados na representação da função de transferência.

Uma das principais áreas de aplicação de tais métodos é na ciência ambiental. [Espey et al. \(1997\)](#) estimaram um modelo econométrico para avaliar a elasticidade-preço da demanda residencial de água utilizando estrutura tarifária, localização, estação do ano e outras variáveis exploratórias. [Nunnari et al. \(2004\)](#) compararam várias técnicas estatísticas para modelar a concentração de SO₂ utilizando as informações de direção do vento, velocidade do vento, radiação solar, temperatura e umidade relativa. [Andriyas e McKee \(2013\)](#) usaram condições biofísicas nos campos dos agricultores e o sistema de irrigação durante a estação de crescimento para

antecipar pedidos de água para irrigação. [Lima et al. \(2014\)](#) desenvolveram um modelo de previsão para a entrada de água incorporando o efeito de variáveis climáticas como precipitação e El Niño.

Existem estudos que mostram as vantagens e desvantagens do uso de abordagens univariadas e multivariadas. O trabalho de [Athanasopoulos et al. \(2011\)](#) realizou uma competição entre métodos univariados e multivariados para prever a demanda internacional de turistas. [Sfetsos e Coonick \(2000\)](#) compararam as abordagens na previsão da radiação solar, e [Porporato e Ridolfi \(2001\)](#) previram os fluxos dos rios. Existem igualmente revisões de métodos e técnicas de séries temporais univariadas e multivariadas aplicadas a áreas específicas.

[Milonis e Davies \(1994\)](#), por exemplo, revisaram métodos de regressão e modelos estocásticos aplicados à análise da poluição do ar. [Ljung \(1999\)](#) descreve a teoria, metodologia, prática dos modelos ARMAX e modelos não lineares de caixa preta, entre outros. [Durbin e Koopman \(2012\)](#) publicaram uma introdução clara e abrangente à abordagem do espaço de estados para análise de séries temporais, juntamente com um pano de fundo histórico, e [Haykin \(1999\)](#) apresenta uma extensa revisão do estado da arte das redes neurais. [Young \(2011\)](#) oferece uma introdução à estimação recursiva e demonstra suas várias formas e seu uso como auxílio na

a modelagem de sistemas estocásticos e dinâmicos em séries temporais. No entanto, há uma escassez de revisões sistemáticas da literatura de métodos de análise de séries temporais usando variáveis exógenas na estrutura de modelagem,

^y Autor correspondente.

E-mail addresses: paulamaçaira@aluno.puc-rio.br (P.M. Maçaira), mt@puc-rio.br (A.M. Tavares Thomé), cyrino@puc-rio.br (F.L. Cyrino Oliveira), analferrer@gmail.com (A.L. Carvalho Ferrer).

<https://doi.org/10.1016/j.envsoft.2018.06.004> Recebido

em 16 de maio de 2017; Recebido em formulário revisado em 30 de abril de 2018; Aceito em 1 de junho de 2018 Disponível online em 22 de junho de 2018

independentemente da área de aplicação ou metodologia.

Dada a relevância da análise de séries temporais com variáveis exógenas, o principal objetivo é preencher a lacuna identificada na literatura sobre modelos que utilizam variáveis explicativas, fornecendo um “mapa da ciência” por meio de revisão sistemática da literatura (SLR) e análise bibliométrica. Segundo Small (1999, p. 799) “um mapa da ciência é uma representação espacial de como disciplinas, campos, especialidades e documentos ou autores individuais se relacionam entre si” e “podem fornecer uma visão sobre um estado contemporâneo de conhecimento”. Assim, este estudo visa orientar pesquisadores e profissionais de vários áreas de conhecimento para identificar quais são as principais áreas de aplicação do análise de séries temporais com variáveis exógenas, que são as mais prolíficas autores por área de conhecimento, quais são os trabalhos mais influentes e quais são os principais métodos utilizados.

Este artigo compreende esta introdução, seguida, na seção 2, por uma descrição dos métodos aplicados à revisão de literatura e análise bibliométrica. A Seção 3 apresenta os resultados da citação, cocitação, e análises de co-palavras; e a seção 4 engloba discussões e conclusões com implicações para a prática e pesquisas futuras.

2. Metodologia

Esta seção apresenta os métodos e estatísticas básicas extraídas de o SLR, bem como os métodos bibliométricos aplicados ao análise de áreas temáticas.

2.1. Revisão sistemática e estatísticas básicas

Thomé et al. (2016a) desenvolveu uma abordagem passo a passo para conduzir uma SLR na gestão de operações que consiste em oito etapas: (i) planejamento e formulando o problema, (ii) pesquisando a literatura, (iii) dados coleta, (iv) avaliação da qualidade, (v) análise e síntese de dados, (vi) interpretação, (vii) apresentação dos resultados e (viii) atualização da revisão.

Na Etapa 1, planejamento e formulação do problema, a equipe de pesquisa composta pelos co-autores deste artigo reuniram e definiram o escopo da revisão, a conceituação do principal tópico de pesquisa e as seguintes questões de pesquisa (QR):

- RQ 1. Quem são os autores mais prolíficos na análise de séries temporais com Variáveis explicativas?
- RQ 2. Quais são os trabalhos mais influentes na análise de séries temporais com variáveis explicativas?
- RQ 3. Quais são os principais temas na análise de séries temporais com ex variáveis planatórias e como elas evoluíram?
- RQ 4. Quais são os principais métodos aplicados na análise de séries temporais com variáveis explicativas e como elas evoluíram?

Após a segunda etapa de SLR, uma pesquisa bibliográfica foi realizada em sete etapas compreendendo a seleção das bases de dados, definição das palavras-chave, revisão dos resumos, critérios para inclusão/exclusão de artigos, revisão de texto completo e busca para trás e para frente em artigos selecionados/ referências. A base de dados Scopus foi escolhida por ser uma das maiores bancos de dados de resumos e citações de literatura de pesquisa contendo mais de 53 milhões de registros e quase 22.000 títulos de 5.000 editoras (HLWIKI, 2017). A escolha da Scopus justifica-se em razão da grande cobertura de domínios de pesquisa pretendidos pela presente SLR (ver para exemplo Mongeon e Paul-Hus, 2016). Não houve restrição de tempo para a pesquisa. A Tabela 1 mostra o número de artigos incluídos em cada fase da pesquisa de palavras-chave.

As palavras-chave devem ser amplas para não restringir artificialmente o número de estudos e específico o suficiente para trazer apenas os estudos relacionados tópico (Cooper, 2010). As primeiras palavras-chave foram “time series” e “exogenous* variable”, gerando 244 artigos. A busca foi posteriormente ampliada com sinônimos de “exógeno” na análise de séries temporais, levando para documentos de 2020. Isso foi feito para atender aos dois critérios de pesquisar estudos em SLR sugeridos por Petticrew e Roberts (2006, p. 81), sensibilidade para recuperar tudo o que é relevante e especificidade para

tabela 1

Número de artigos por busca por palavra-chave.

Pesquisa por palavra-chave	Nº de papéis
"série temporal" AND "variável exógena" "série temporal" E ("variável exógena" OR "variável explicativa") ("variável exógena" OR "variável explicativa" OU "variáveis de entrada")	244
"série temporal" AND ("variável exógena" OR "variável explicativa" OU "variáveis de entrada" OU "variável de previsão")	830
"série temporal" AND ("variável exógena" OR "variável explicativa" OU "variáveis de entrada" OU "variável de previsão" OU "explicativo variável")	1323
"série temporal" AND ("variável exógena" OR "variável explicativa" OU "variáveis de entrada" OU "variável de previsão" OU "explicativo variável" OU "variável dependente")	1575
	1579
	2020

deixar para trás o irrelevante. Os critérios de exclusão dos artigos são o idioma do artigo e o tipo de documento, resultando em 1930 artigos escritos Em inglês. A limitação do tipo de documento para artigo, revisão e artigos no prelo reduziu a seleção para um conjunto final de 1.547 documentos incluídos na análise bibliométrica. A referência bibliográfica completa é disponível mediante solicitação ao autor principal.

A Tabela 2 apresenta as 10 principais áreas com maior concentração de artigos de 28 áreas de conhecimento identificadas. Como esperado, há uma grande variedade de áreas temáticas, que vão da ciência da computação à medicina. Outro resultado consistente é a concentração de 20% dos trabalhos em área de meio ambiente (Ciência Ambiental & Terra e Ciências) dada a complexidade das séries dos fenômenos naturais e a necessidade de usar modelos que incorporem informações externas.

A Ciência Ambiental, a Matemática e a Medicina são responsáveis por aproximadamente 31% do total de trabalhos. Junto com Ciências Sociais e Economia, Econometria e Finanças, os cinco primeiros áreas correspondem a aproximadamente 50% do total.

A Figura 1 ilustra o número de artigos publicados por ano. O primeiro artigo apareceu em 1967 (Scott Jr. e Heady, 1967), sobre a demanda para novos investimentos em edifícios agrícolas. Nos próximos três anos, houve não havia publicações. Vale a pena notar que até 1995 o total número de artigos publicados por ano foi consistentemente inferior a 20. As publicações chegaram a 27 em 1996. Cinco anos depois, em 2001, o número de publicações atingiu a marca de mais de 40 artigos publicados por ano. Desde então, tem havido uma tendência de rápido crescimento, atingindo um recorde de 150 publicações em 2015 e 124 publicações em 2016.

Ao analisar a distribuição dos artigos por periódicos, uma grande surge uma variedade de 151 fontes diferentes, como esperado em um campo tão multidisciplinar. Os 20 principais periódicos em número de citações estão em Tabela 3. Juntos, eles respondem por 73% do total de citações. O periódico mais bem classificado em número de citações refere-se a a economia, onde os modelos de séries temporais são amplamente utilizados. o Journal of Econometrics é também o periódico com o maior número de

mesa 2

Dez principais áreas temáticas.

Categorias do Scopus	Nº de papéis	Porcentagem de contribuição
Ciência ambiental	314	12%
Matemática	266	10%
Remédio	250	9%
Ciências Sociais	247	9%
Economia, Econometria e Finança	242	9%
Ciências da Terra e Planetárias	203	8%
Ciências Agrárias e Biológicas	202	8%
Engenharia	196	7%
Ciência da Computação	190	7%
Negócios, Gestão e Contabilidade	127	5%

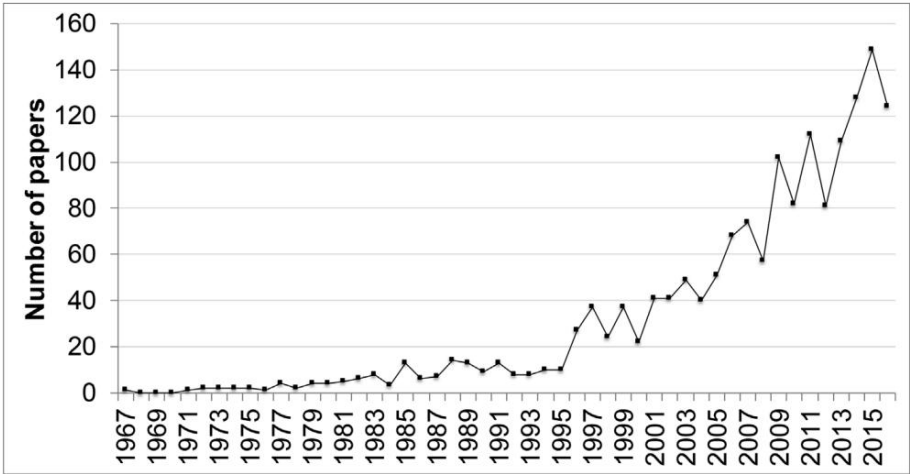


Fig. 1. Número de artigos publicados por ano.

Tabela 3
Números de artigos e citações por fonte.

Fonte	Nº de papéis	Nº de citações
Revista de Econometria	24	1923
Revista Americana de Ciência Política	4	1186
Jornal da Associação Americana de Estatística	23	1066
Pesquisa de Recursos Hídricos	20	812
Transações IEEE em Sistemas de Energia	9	701
Jornal do Clima	8	625
Neurocomputação	16	548
Jornal Internacional de Previsão	19	521
Sensoriamento Remoto do Ambiente	11	489
Estatística em Medicina	6	455
Revisão Anual da Ciência Política	2	424
Revista de Hidrologia	22	376
Revista Internacional de Climatologia	12	374
Modelagem Ecológica	6	351
Sistemas especialistas com aplicativos	11	342
Revisão de Ciência Política Americana	2	309
Revista Europeia de Pesquisa Política	5	282
Organização Internacional	3	247
Ambiente Atmosférico	11	231
Pesquisa de Pesca	11	217
Pesquisa de Água	3	195
Ciências da Informação	3	189
Revista CIEM de Ciências Marinhas	7	187
Medicina Acadêmica de Emergência	3	186
Economia Ecológica	2	184
Análise e Prevenção de Acidentes	9	156
Gestão de Recursos Hídricos	6	155
Gerência de Turismo	5	153
Modelagem Ambiental e Software	8	152
Computação Neural e Aplicações	6	152

publicações no tema. Também vale a pena mencionar a grande presença de periódicos relacionados ao meio ambiente (Environmental Mod elling and Software, Water Resources Research, Journal of Climate, Sensoriamento Remoto do Ambiente, Journal of Hydrology, International Jornal de Climatologia, Modelagem Ecológica, Ambiente Atmosférico, Pesquisa da Água, Economia Ecológica e Recursos Hídricos Gerenciamento). Juntos, eles representam cerca de 30% do total número de citações e 39% de artigos, respectivamente. No entanto, a correlação de posto de Spear man entre o número de publicações e a número total de citações é igual a 0,57, um valor médio mostrando que alguns dos periódicos mais prolíficos não são necessariamente os mais influentes.

A terceira e quarta etapa do SLR é a coleta de dados e qualidade avaliação. Um modelo de computador foi usado para a coleta de dados e a cálculo de frequências e médias simples. A restrição de seleção

artigos para revistas revisadas por pares forneceram uma medida inicial da qualidade dos artigos recuperados para análise. A quinta etapa do SLR são os dados análise e síntese. Uma abordagem indutiva para o conteúdo qualitativo foi adotada a análise (Seuring e Gold, 2012; Thomé et al., 2016a) e combinado com análise bibliométrica quantitativa de cocitações e coocorrência de palavras-chave, descritas na subseção 2.2. Posteriormente, a sexta etapa, interpretação, refere-se a uma pesquisa qualitativa síntese aliada a indicadores bibliométricos da cocitação e análises de co-palavras. Este estudo consiste na sétima etapa do SLR abordagem, apresentação de resultados. A etapa oito, atualização da revisão, está além do escopo deste estudo.

2.2. Análise bibliométrica

Medidas e indicadores bibliométricos podem ser empregados para análises de desempenho e gerar mapas científicos. Esses procedimentos quantificar e medir o desempenho e o impacto da pesquisa científica (Cobo et al., 2011). São realizadas três análises específicas: análise de citação, cocitação e análise de co-palavras.

A análise de citações é a primeira e mais simples análise realizada no base de dados. Esta análise envolve o estudo dos autores mais influentes, medindo o número de artigos e citações por conhecimento áreas, instituição e país. A co-ocorrência é o estudo da aparecimento de pares de unidades em vários registros bibliográficos. Existem diferentes tipos de análise bibliométrica de coocorrência: as análises de cocitação, coautoria e co-palavra são as mais comuns (Persson et al., 2009). Por exemplo, se o documento B e C citam o documento A, ambos aparecem como co-citando A. Este método pode ser usado para descrever a espinha dorsal de uma área de pesquisa, mapeando a rede de cocitação entre os autores mais influentes. Permite a identificação do principais correntes de pesquisa.

2.2.1. Disponibilidade de software e dados

Existe um grande número de softwares disponíveis para análise bibliométrica em revisões de literatura, como evidenciado por Cobo et al. (2012). Para o análise de citação, cocitação e co-palavra, software BibExcel (Persson et al., 2009) (<http://homepage.univie.ac.at/juan.gorraiz/bibexcel/>), Pajek (De Nooy et al., 2005) (<http://mrvar.fdv.uni-lj.si/pajek/>), e SciMAT (Cobo et al., 2012) (<http://sci2s.ugr.es/scimat/>) foram usados. Tudo software está disponível como freeware para uso acadêmico sem fins lucrativos. Bib Excel foi usado para preparar a matriz de cocitação e gerar os clusters, Pajek para preparar a análise de rede e SciMAT para analisar dinâmica mapas de análise longitudinal co-palavra de temas usando a temática de Callon gráficos bipartidos ou diagramas estratégicos (Cobo et al., 2012) , densidade e centralidade das coocorrências das palavras-chave do autor. O cheio conjunto de dados contendo 1547 documentos incluídos na bibliometria

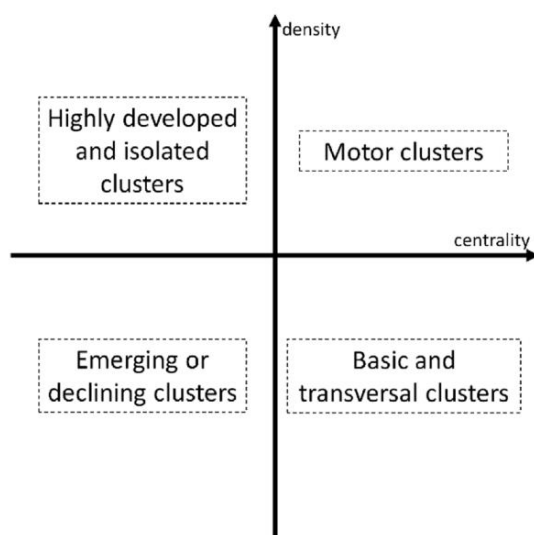


Fig. 2. Diagrama estratégico de Callon.
Fonte: Adaptado de Cobo et al. (2012).

a análise é disponibilizada no repositório Mendeley (Thomé et al., 2017).

2.2.2. Diagrama estratégico de Callon: índices bibliométricos

A Fig. 2 mostra o diagrama estratégico de Callon (Callon et al., 1991).

Segundo o diagrama estratégico de Callon, na Fig. 2, os grupos de motores (superior direito) representam os temas centrais da área de pesquisa, com alta centralidade e alta densidade. Os clusters básicos e transversais (parte inferior direita) são temas centrais que também são fundamentais para o campo de pesquisa, mas não são bem desenvolvidos, pois combinam alta centralidade com baixa densidade. Temas emergentes ou em declínio (canto inferior esquerdo) apresentam baixa centralidade e densidade baixa; ou seja, eles não se relacionam bem com outros temas e não estão bem representados na área de pesquisa. Clusters altamente desenvolvidos e isolados (canto superior esquerdo) geralmente são áreas bem pesquisadas com alta densidade, comumente denotando temas clássicos na área de pesquisa.

O índice de similaridade mede a co-ocorrência. Ele considera o número de documentos em que ocorrem duas palavras-chave e o número de documentos em que cada um ocorre. Os clusters são formados no SciMAT com a aplicação do “algoritmo de centro simples” (Cobo et al., 2011; Thomé et al., 2016b). No “algoritmo de centro simples”, a centralidade e densidade são calculados como uma função linear do índice de similaridade. O índice de similaridade é calculado como $e_{ij} = C_{ij} / \sqrt{C_i \cdot C_j}$, sendo o número de documentos em que duas palavras-chave i e j coocorrem, C_i e C_j sendo o número de documentos em que cada um ocorre. Por um lado, a centralidade é uma medida da interação entre redes de palavras calculadas como $C = 10 \cdot K/h$, onde k é uma palavra-chave pertencente ao tema e h é uma palavra-chave pertencente ao outro tema. No outro lado, a densidade mede a força interna da rede ou tema calculado como $d = 100 \cdot \frac{\sum_{i,j} e_{ij}}{w}$, onde i e j são palavras-chave da rede e w é o número total de palavras-chave no tema.

3. Resultados das análises de citação, cocitação e co-palavra

3.1. Análise de citações

Na tentativa de responder à primeira questão de pesquisa, “Quem são os autores mais prolíficos na análise de séries temporais com variável explicativa?”, a análise de citações escrutina o número de citações e número de artigos por autor. A Tabela 4 mostra a classificação dos autores por número de citações, número de artigos, instituição, país e assunto área. Na Tabela 4, o número de artigos por autor e o número de citações por autor é o número de artigos e a soma de todas as citações

de um determinado autor que aparece no banco de dados, respectivamente.

O autor mais prolífico (Beck, N.) maior número de citações; no entanto, o número de documentos e número de citações pode não se correlacionar. De fato, a relação de classificação de Spearman entre o número de artigos e citações não é significativa ($p = 0,13$). Uma outra olhada na Tabela 4 revela que as principais áreas para esses autores são Ciências Sociais e Economia e estão localizados principalmente nos EUA e no Reino Unido.

O trabalho mais citado de Beck encontrado no banco de dados (Beck et al., 1998) é co-autoria com Katz e Tucker. Apresenta uma solução para analisar dados de corte transversal de séries temporais usando análises logit e aplica a metodologia proposta para estimar a relação entre independência econômica, democracia e paz. Vale a pena mencionar que Beck, Katz e Tucker são coautores frequentes em estudos com o tempo dados de corte transversal em série. Meese e Rogoff (1983) têm o segundo artigo mais citado na base de dados; seu artigo compara a precisão da previsão fora da amostra de vários modelos de taxa de câmbio estruturais e de séries temporais. A obra de Anderson e Hsiao (1982) vem próximo em número de citações; apresenta uma análise estatística do tempo modelos de regressão em série para dados longitudinais com e sem atraso variáveis dependentes. Wilby, Wigley e Conway foram coautores do trabalho que tem mais citações na base de dados para cada autor (Wilby et al., 1998); ele calibra uma variedade de diferentes modelos estatísticos de downscaling usando modelos de circulação observada e geral para gerar e comparar séries temporais de precipitação diária. Como esperado, mais frequentemente autores citados publicados anteriormente, uma vez que esses autores tiveram mais tempo para obter citações do que os autores mais recentes.

3.2. Análise de cocitação

A análise de cocitação aborda a segunda questão de pesquisa: “Quais são os trabalhos mais influentes na análise de séries temporais com variáveis explicativas?” A co-citação apresenta os mais influentes trabalhos em análise de séries temporais com variáveis exógenas, ou seja, o espinha dorsal da área estudada.

A rede na Fig. 3 ilustra as relações de cocitação entre os 30 principais documentos com maior número de citações, ou seja, os trabalhos mais influentes na análise de séries temporais com variáveis explicativas. Os documentos aparecem na Fig. 3 cronologicamente de cima para baixo. Os tamanhos da circunferência dos círculos são proporcionais ao número de citações e a largura das linhas que conectam a rede é proporcional ao número de co-ocorrências.

A seguir descreve a maior contribuição de cada nó do árvore de cocitação. Granger (1969) definiu a diferença entre causalidade e feedback usando o exemplo simples de modelos bivariados. Ele introduziu o uso de defasagem causal e força causal para analisar a direção da causalidade entre duas variáveis relacionadas. Nash e Sutcliffe (1970) discutiu os fluxos de rios de chuva, evaporação e outros fatores observando o coeficiente R^2 de uma regressão linear.

O trabalho popular de Box e Jenkins (1970) introduziu os modelos de séries temporais de média móvel integrada auto-regressiva (ARIMA) para caracterizar e prever observações de séries temporais em intervalos igualmente espaçados períodos. Esse modelo clássico foi revisado pelos mesmos autores em 1976. Existem três métodos diferentes derivados do modelo ARIMA:

autoregressivo (AR), média móvel (MA) e média móvel AR modelos (ARMA). Os modelos AR são aplicados a uma série temporal que pode ser representado por seus próprios valores passados, enquanto os modelos MA são usados para representar uma série temporal onde os erros passados (perturbações) determinam sua valores futuros. Os modelos ARMA são apropriados quando uma série é uma função de erros não observados, bem como seu próprio comportamento passado. A diferença entre ARIMA e ARMA é que o primeiro é aplicado a séries que não são estacionárias ao longo do tempo, ou seja, com alguma tendência e, portanto, precisam ser diferenciado (o termo “I”) para se tornar estacionário.

O critério de informação de Akaike (AIC) foi originalmente desenvolvido sob o nome de “critério de informação” por Akaike (1973) e recebeu sua denominação plena de AIC em 1974, juntamente com um

Tabela 4
Número de citações e artigos publicados por autores, instituição, país e área temática.

Autores	Nº de citações	Nº de papéis	Instituição	País	Área de estudo
Beck N.	1668	5	Universidade de Nova York	cervo	Ciências Sociais
Katz J.	1300	3	Instituto de Tecnologia da Califórnia	cervo	Ciências Sociais
Meese R.	1267	1	Escola de Negócios da UC Berkeley Haas	cervo	Economia
Rogoff K.	1267	1	Universidade de Harvard	cervo	Economia
Tucker R.	1064	1	Universidade Vanderbilt	cervo	Ciências Sociais
Hsiao C.	669	2	Universidade do Sul da Califórnia	cervo	Economia
Anderson T.	656	1	Universidade de Stanford	cervo	Matemática
Wigley T.	583	3	Centro Nacional de Pesquisa Atmosférica	cervo	Ciências da Terra e Planetárias
Wilby R.	566	2	Universidade de Loughborough	Reino Unido	Ciência ambiental
Conway D.	428	2	London School of Economics and Political Science	Reino Unido	Ciência ambiental

definição de conceitos. O AIC auxilia na seleção do modelo para estimar a qualidade relativa de um modelo baseado em perda de informação e parcimônia.

Schwarz (1978) tentou resolver o problema de selecionar um de um número de modelos de diferentes dimensões, e Ljung e Box (1978) apresentou uma modificação no teste de falta de ajuste proposto por Box e Pierce (1970).

O teste de Dickey-Fuller deriva representações para as distribuições limitantes do estimador do coeficiente AR e testa a hipótese nula se uma raiz unitária está presente (Dickey e Fuller, 1979). Mais tarde, Philips e Perron (1988) desenvolveram um teste de raiz unitária denominado teste de Phillip s-Perron que faz uma correção não paramétrica da estatística do teste t presente no teste de Dickey-Fuller.

Heterocedasticidade nas perturbações de uma linearidade devidamente especificada modelo leva a estimativas de parâmetros ineficientes, o que resulta em falhas inferências ao testar hipóteses estatísticas. White (1980) desenvolveu o chamado teste de White para verificar se há heterocedasticidade na perturbações de um modelo linear.

Ainda tratando da heterocedasticidade, Engle (1982) introduziu uma nova classe de processo estocástico chamado heterocedasticidade condicional AR (ARCH) que não assume variância constante para o termo de erro. Esse tipo de processo assume que a variância do erro segue um modelo AR. Dentro 1986, Bollerslev (1986) estendeu a ideia de Engle (1982) ao introduzir uma classe mais geral de processo chamada heterocedasticidade condicional AR generalizada (GARCH), que permite uma estrutura de defasagem muito mais flexível e admite um modelo ARMA para o erro estrutura.

Entrando na área da inteligência artificial (IA), a primeira referência principal surge em 1985 com o estudo de Takagi e Sugeno (1985), que sugeriu uma ferramenta matemática para descrever um sistema que pode

representam relações altamente não lineares de forma difusa. A principal contribuição de o estudo trata da representação do sistema fuzzy como um sistema linear.

Igualmente, na área de IA, Rumelhart et al. (1986) apresentaram um procedimento chamado propagação de erro, pelo qual o gradiente pode ser determinado por unidades individuais da rede com base apenas na formação disponível localmente.

Trabalhando com a definição de co-integração, Engle e Granger (1987) sugeriu um teste para estimar as relações de co-integração usando regressão e combinando os problemas de testes de raiz unitária e teste com parâmetros não identificados sob o valor nulo. No estudo de Johansen (1988), em vez de trabalhar com estimativas de regressão para testar a cointegração, o autor derivou um estimador de máxima verossimilhança que leva em consideração a estrutura de erro do processo subjacente, o que não é possível com as estimativas de regressão. O método proposto anteriormente foi modificado por adicionando restrições lineares nos vetores de cointegração e pesos AR no método de Johansen e Juselius (1990). Johansen (1991) apresentou estimadores de máxima verossimilhança e testes de razão de verossimilhança para cointegrações em modelos AR vetoriais gaussianos, que permitem a introdução de dummies de termo constante e sazonal no modelo.

Em 1989, Harvey (1989) publicou um livro que e teoria abrangente de modelos estruturais de séries temporais e incluiu um tratamento detalhado do filtro de Kalman. Várias aplicações ilustraram as propriedades dos modelos e técnicas metodológicas.

Voltando às referências que tratam de métodos de IA, Hornik et al. (1989) deu o primeiro passo em uma rigorosa investigação geral sobre a capacidades e propriedades de redes feedforward multicamadas, estabelecendo que esse tipo específico é uma classe de aproximação universal.

O livro escrito por McCullagh e Nelder (1989) forneceu uma

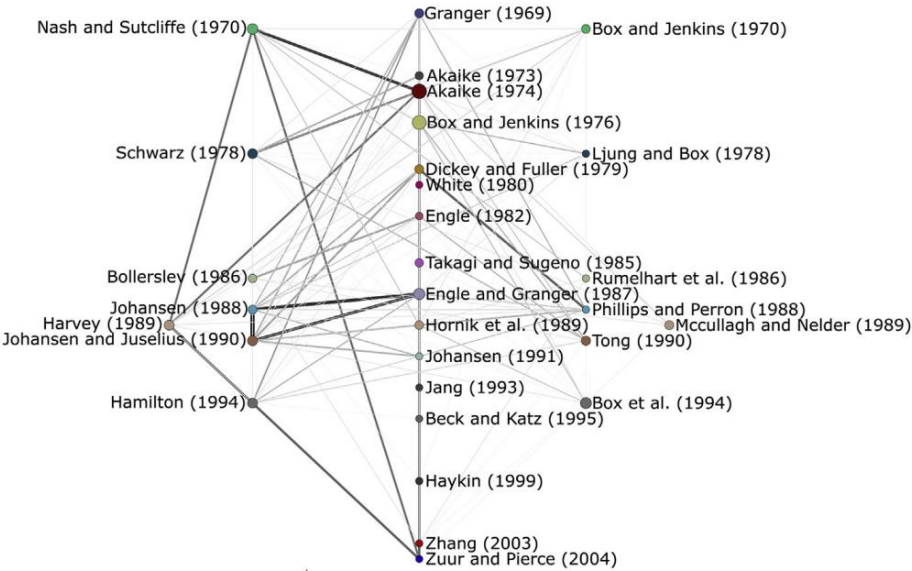


Fig. 3. Rede de cocitação de análise de séries temporais com variáveis exógenas.

tratamento unificado de métodos para a análise de diversos tipos de dados. O trabalho se concentrou em examinar a maneira como uma variável de resposta depende de uma combinação de variáveis explicativas.

Em seu livro, *Non-Linear Time Series: A Dynamical System Approach*, Tong (1990) introduziu a teoria de séries temporais não lineares e forneceu pesquisa de ponta na época. Seu livro preencheu a lacuna entre análise de séries temporais lineares e caóticas, tornando-se uma das principais referências em séries temporais não lineares.

A modelagem e identificação difusa, explorada pela primeira vez por Takagi e Sugeno (1985), é a principal inspiração para a arquitetura denominada sistema de inferência fuzzy baseado em rede adaptável (ANFIS) desenvolvido por Jang (1993). Usando um procedimento de aprendizagem híbrido, o ANFIS proposto pode construir um mapeamento de entrada-saída com base em estudos de saúde conhecimento e pares de dados de entrada-saída estipulados.

Em 1994, dois livros que cobriam as ferramentas para modelagem e análise de séries temporais e apresentavam os últimos desenvolvimentos que ocorreram no campo na última década foram publicados. Caixa et al. (1994) publicou uma nova edição do clássico *Time Series Analysis: Previsão e controle* de Box e Jenkins (1970), enquanto Hamilton (1994) explorou a primeira edição de um livro que sintetiza os avanços em séries temporais econômico-financeiras.

A obra de Beck e Katz (1995) é a primeira referência que aparece na rede de cocitação que trata do estudo transversal, uma técnica amplamente utilizada em pesquisas médicas e ciências sociais. Seus trabalho examinou algumas questões na estimativa de séries temporais cross-sec modelos tradicionais e propôs um novo método.

Um livro apresentando uma análise detalhada de redes neurais artificiais (RNAs) não apareceram até 1999, quando Haykin (1999) apresenta uma extensos métodos e antecedentes de RNA de última geração.

Com o crescimento do poder computacional de processamento e consequentemente o uso de métodos de IA em séries temporais, modelos híbridos, como o de Zhang (2003), que combinam os modelos clássicos de séries temporais, ARIMA, AI e ANNs para fornecer melhores medidas de erro em séries temporais previsão apareceu. Essa metodologia, em particular, aproveita da força única dos modelos ARIMA e ANN em modelagem linear e não linear.

A referência mais recente na rede de cocitação é de Zuur e Pierce (2004), que utilizou uma técnica denominada análise fatorial dinâmica, apresentada em Zuur et al. (2003), para estimar tendências comuns no tempo série de captura de lulas por unidade de esforço em vez de aplicar ARIMA modelos. O método também permite determinar a relação entre variáveis independentes e exploratórias, como a superfície do mar temperatura e a influência do índice de Oscilação do Atlântico Norte na série temporal de captura de lulas.

Com a análise de cocitação, é possível concluir que o trabalhos que influenciam principalmente as pesquisas na área estudada são também as referências mais importantes em métodos de séries temporais e fornecidas os principais testes estatísticos. A árvore de cocitação da Fig. 3 oferece uma guia cronológico para o estudo de métodos de séries temporais com variáveis.

3.3. Análise de co-palavras

A análise de co-palavras pretende responder às duas últimas pesquisas perguntas: “Quais são os principais temas na análise de séries temporais com variáveis explicativas e como elas evoluíram?” e “Quais são os principais métodos aplicados na análise de séries temporais com variáveis explicativas e como eles evoluíram?” Para permitir uma melhor compreensão a evolução de temas e métodos e retratar pesquisas que mais recentes na área, os documentos foram subdivididos em quatro períodos sucessivos: 1967–1998, 1999–2007, 2008–2012 e 2013–2016, com 246, 423, 434 e 444 documentos, respectivamente. A escolha do duração dos períodos foi feita a partir da equalização dos números de artigos do mais recente ao mais antigo, fornecendo os três últimos períodos com uma média de cerca de 400 artigos e o último com cerca de 250. Os principais temas na área de pesquisa por período estão na Tabela 5,

Tabela 5

Número de artigos e citações por período, agrupamento temático e tipo de documento mento.

Período	Cluster temático	Número de documentos			
		Núcleo	Secundário	Total	Core com até 80% citações
Previsão de 1967–1998	Estudos de saúde	16	33	49	6
	Desenvolvido países	10	17		3
	Filtro Kalman	27	25	7	1
		2	5	7	1
	Subtotal	30	60	90	11
1999–2007	Estudos de saúde	57	61	118	27
	Matemático modelos	47	83	130	18
	Poluente do ar	15	55	70	8
	Subtotal	119	199	318	53
Estudos de saúde de 2008–2012	Previsão	42	54	96	18
	Algoritmos	63	64	127	24
	Modelos estatísticos	7	32	39	4
		6	38	44	3
	Subtotal	118	188	306	49
Estudos de saúde de 2013–2016	Previsão	32	46	78	15
	Árvores de decisão	55	87	142	17
	Custos	4	6	10	2
	Algoritmos	5	32	37	1
		4	36	40	2
	Subtotal	100	207	307	37
Total		367	654	1021	150

com o número de documentos principais e secundários recuperados para análise.

Os documentos principais apresentam pelo menos duas co-ocorrências de palavras-chave, enquanto os documentos secundários apresentam uma única coocorrência de palavras-chave no cluster temático. A leitura do texto completo dos documentos principais correspondendo a até 80% do total de citações dentro do cluster, desde que a base para a análise de cada cluster temático. Por exemplo, no período 1967-1998, os seis dos 16 documentos principais reagrupando até 80% do total de citações para o cluster de previsão (primeira linha da Tabela 5) foram analisados de conteúdo para a identificação de temas e métodos para o cluster de previsão neste período. A Fig. 4 mostra o número de palavras-chave por período e apresenta sua evolução mostrando o número de palavras-chave de saída e entrada e o número e porcentagem de palavras-chave que permanecem de um período para o outro.

Como esperado, o número de palavras-chave cresce ao longo dos períodos, paralelamente ao aumento do número de documentos ao longo dos anos. o número de palavras-chave aumenta de 130 no primeiro período (1967-1998) para 377 no quarto período (2013–2016), um crescimento de 290%. Dos 130 palavras que aparecem no primeiro período, 92 (71%) permanecem para o segundo período, e 194 são adicionados, totalizando 286 palavras. Para o terceiro período, 218 (76%) permanecem e 171 novas palavras são incluídas, totalizando 389 no total. Para o quarto período, 241 (62%) palavras-chave do terceiro

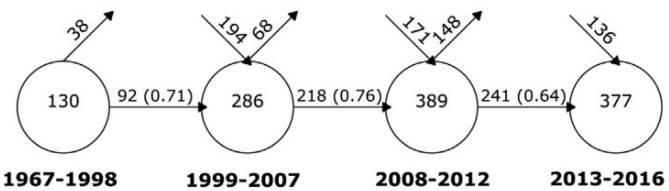


Fig. 4. Evolução das palavras-chave por períodos.

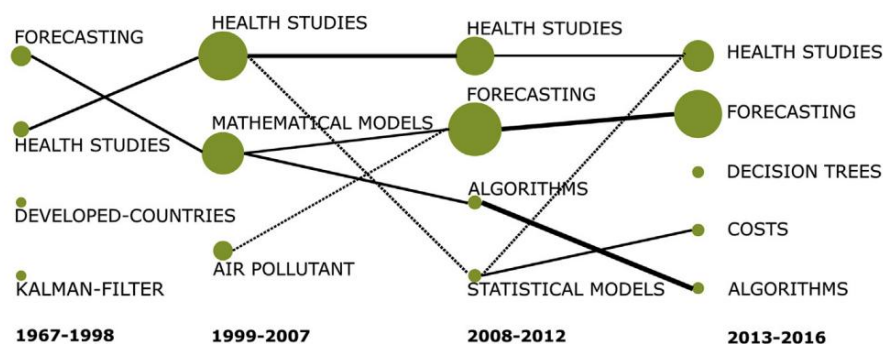


Fig. 5. Evolução dos clusters temáticos.

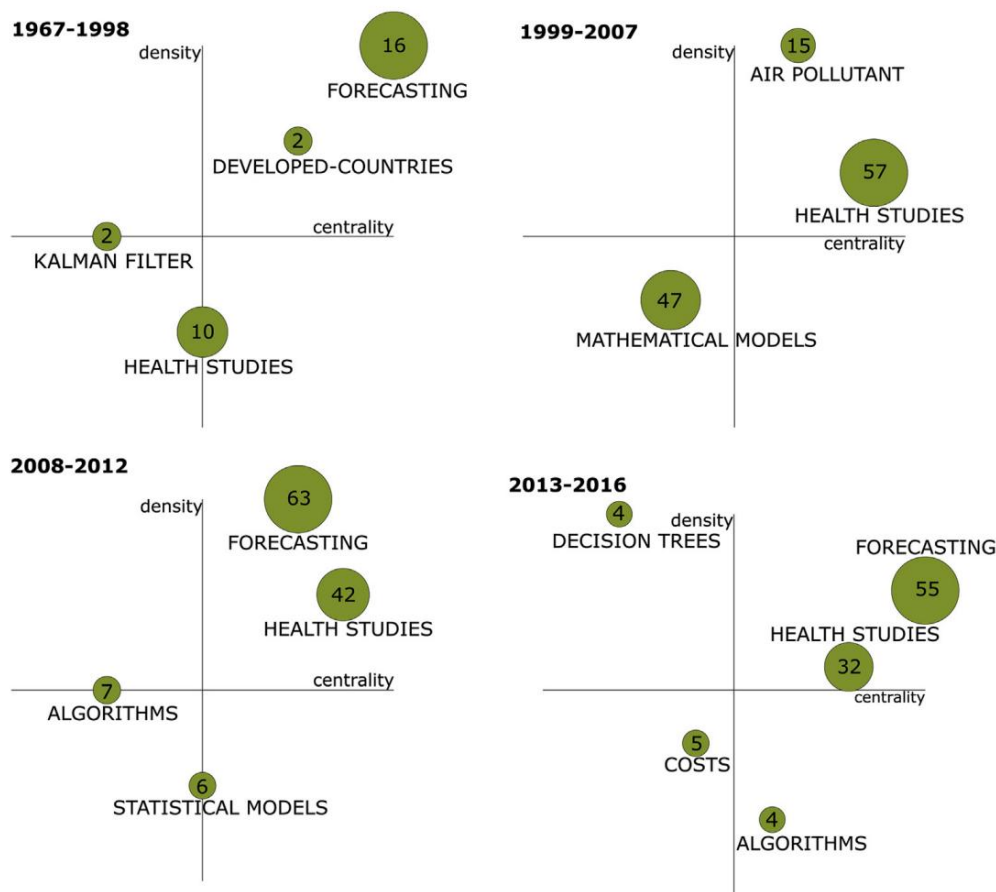


Fig. 6. Diagramas estratégicos.

período permanecem, e 136 novas palavras aparecem, resultando em 377 palavras-chave. Isso pode ser indicativo de que as áreas que utilizam variáveis explicativas em séries temporais estão se diversificando e aumentando ao longo do tempo e, portanto, isso é ainda não é um campo consolidado.

A palavra-chave mais usada em todos os períodos é humano, seguido de previsão, análise de regressão, redes neurais e algoritmos. As palavras modelos matemáticos, demografia e desenvolvimento países aparecem com alta frequência no primeiro período. Poluição do ar, ozônio, temperatura, mudanças climáticas e monitoramento ambiental são palavras mais usadas do segundo ao último período. No último período, as palavras avaliação de risco e crescimento econômico são usadas mais com destaque do que nos períodos anteriores.

A Fig. 5 mostra a evolução dos temas por período, com base na força da associação entre os temas de um período para outro. O índice de inclusão mede a força da associação, representada pela espessura das linhas que ligam cada cluster. o

índice de inclusão varia de zero a um, onde zero significa que o áreas temáticas não estão conectadas, e é medido como onde $\frac{\# UV \cap V}{\min(\#U, \#V)}$, U e V são conjuntos disjuntos (temas em períodos diferentes). O contínuo linhas representam uma associação de nomes entre clusters temáticos, ou seja, dois clusters têm o mesmo nome em períodos consecutivos ou um cluster temático contém o outro, e as linhas pontilhadas representam associações sobre outros aspectos que não o nome. Os tamanhos das esferas correspondem ao número de documentos centrais presentes no cluster.

Da Tabela 5 e Fig. 5, em 1967-1998, a previsão representa 53% do número total de documentos principais no período, enquanto em Estudos de saúde e modelos matemáticos de 1999-2007 são os principais temas responsáveis por 48% e 39% dos documentos centrais, respectivamente. Nos dois períodos seguintes, 2008–2012 e 2013–2016, os principais tema volta a ser previsão, respondendo por 53% e 55% do núcleo documentos, respectivamente. O segundo tema principal são os estudos em saúde ambos os períodos, responsáveis por 36% e 32% dos documentos essenciais,

respectivamente.

O cluster de previsão em 1967-1998 continua como modelos matemáticos em 1999-2007 e como previsão para os dois períodos restantes.

A área de estudos da saúde mantém-se ao longo de todos os períodos mas funde-se com os modelos estatísticos no terceiro período, que se desdobram em custos e estudos em saúde no período seguinte. No primeiro período, dois clusters isolados estão presentes: países desenvolvidos e filtro de Kalman. Um tema isolado, as árvores de decisão, aparece apenas no último período. O cluster de algoritmos em 2008–2012 tem origem em modelos matemáticos e permanece com o mesmo nome no próximo período.

O tema previsão reúne estudos que focam na avaliação de métodos de previsão, estudos em saúde concentram-se em trabalhos relacionados ao comportamento de saúde, países desenvolvidos reúne trabalhos aplicados a países como Suécia e Austrália e o filtro de Kalman foca trabalhos que utilizam o filtro. Artigos que aplicaram técnicas de IA formam os modelos matemáticos e clusters de algoritmos. O cluster de poluentes atmosféricos concentra-se em trabalhos que tratam da influência da qualidade do ar no desenvolvimento de doenças. Os trabalhos que aplicaram a metodologia das árvores de decisão formam o tema das árvores de decisão e cluster de custos apresentam trabalhos relacionados a preços e custos.

A Fig. 6 sintetiza a evolução dos diagramas temáticos estratégicos para os quatro períodos, onde os tamanhos das esferas são proporcionais ao número de documentos centrais em cada cluster e período.

De acordo com o posicionamento no diagrama de Callon da Fig. 6, os agrupamentos motores evoluíram de países desenvolvidos e de previsão em 1967-1998 para estudos de saúde e poluentes atmosféricos em 1999-2007, para estudos e previsões de saúde em 2008-2012 e 2013-2016. Desde 2008, os núcleos temáticos surgem com uma clara concentração na área da saúde e nos métodos e técnicas de previsão.

3.4. Principais métodos na análise de séries temporais

Para identificar os principais métodos aplicados a séries temporais com variáveis explicativas, foi realizada uma leitura minuciosa de 150 artigos correspondentes a documentos centrais que responderam a 80% do total de citações.

Foram revisados apenas os métodos que incorporam variáveis explicativas, ou seja, se um artigo utilizou suavização exponencial, regressão linear múltipla e regressão dinâmica, apenas os dois últimos métodos foram atribuídos a esse artigo.

Por meio da análise, foi possível identificar 30 tipos diferentes de métodos presentes nos documentos centrais. O método mais aplicado foram os modelos de regressão, seguidos de RNAs, Box e Jenkins' ARIMA com a incorporação de variável explicativa (ARIMAX), máquinas de vetores de suporte (SVMs) e modelos estruturais, que juntos estão presentes em 70% dos artigos. A Fig. 7 mostra a evolução temporal dos cinco principais métodos. Apesar da predominância de modelos de regressão

e RNAs, o uso de SVMs está aumentando ao longo do tempo.

Os cinco principais métodos são resumidos e discutidos a seguir.

3.4.1. Modelos de regressão

De acordo com a análise de co-palavras, no período 1967-1998, os modelos de regressão foram o segundo método mais aplicado, liderado por Goldstein et al. (1994), que propuseram um modelo que pode incorporar variáveis explicativas a dados de séries temporais, onde as medidas são feitas próximas no tempo, resultando em uma possível correlação nos resíduos.

Goldstein et al. (1994) foram classificados no cluster de estudos em saúde.

De 1999 até os dias atuais, os modelos do tipo regressão lideram o número de aplicações. Em 1999-2007, o trabalho mais citado que utiliza modelos de regressão é Marcellino et al. (2006), que compararam previsões de modelos lineares univariados e bivariados para prever séries temporais macroeconômicas mensais. Em 2008-2012, o estudo de Sistrom et al. (2009), o trabalho mais citado entre os modelos de regressão nesse período, foi classificado como estudos em saúde. Sistrom et al. (2009) usaram regressão linear por partes para determinar os efeitos da entrada computadorizada de pedidos com suporte integrado à decisão sobre o crescimento de procedimentos ambulatoriais. Em 2013-2016, o artigo de Ishak et al. (2013) liderou o trabalho em modelos de regressão e SVMs, focado no desenvolvimento de um modelo híbrido que utiliza dois modelos de regressão, RNAs e SVMs, para melhor estimar a velocidade do vento utilizando parâmetros hidrometeorológicos.

3.4.2. Redes neurais artificiais (RNA)

Em 1967-1998 as RNAs são lideradas pelo trabalho de Jorquera et al. (1998), classificado no cluster de previsão, que comparou métodos de séries temporais lineares, RNAs e modelos fuzzy para prever níveis máximos diários de ozônio. Nos períodos seguintes, González et al. (2005) propuseram um modelo de Markov oculto de entrada-saída para analisar e prever preços spot de eletricidade. Coman et al. (2008) avaliaram um modelo dinâmico versus um estático, usando o multilayer perceptron (MLP) no contexto da previsão de concentrações horárias de ozônio. Fernández et al. (2009) usaram redes neurais difusas para melhorar a previsão de vazão de águas residuais. Guresen et al. (2011) avaliaram a eficácia dos modelos de RNA nas previsões do mercado de ações. Nourani et al. (2013) aplicaram uma rede neural de feed-forward para modelar um processo de chuva-escoamento em um dia e em várias etapas à frente

escala de tempo.

3.4.3. ARIMAX Em

1967-1998, o método mais prevalente em número de artigos é o ARIMAX, onde o trabalho de Yang et al. (1996) é o artigo mais citado.

Em seu trabalho, classificado no cluster de previsão, é proposta uma nova abordagem de programação evolutiva para identificar o modelo ARMAX para previsões de demanda de carga horária de um dia a uma semana à frente. O estudo de Low et al. (2006), é o trabalho mais citado que utiliza ARIMAX

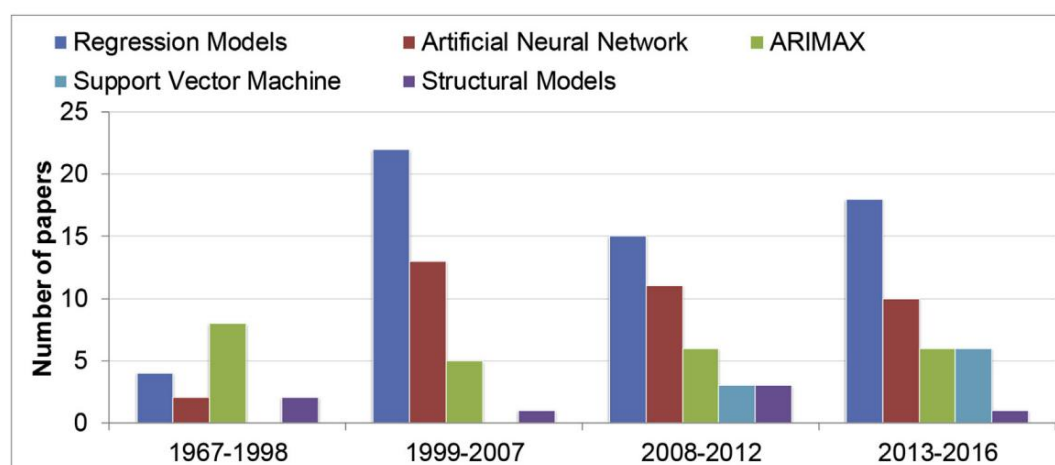


Fig. 7. Número de artigos por métodos por período.

metodologia no segundo período, utilizando as admissões por AVC como variável de resposta e dia da semana, feriados, 11 de setembro e outras contagens e níveis como variáveis explicativas. No terceiro período, [Pisoni et al. \(2009\)](#) usaram um modelo AR não linear com variáveis exógenas para prever os níveis de pico de poluição do ar. O estudo de [Marcilio et al. \(2013\)](#), é o principal trabalho em métodos ARIMAX no último período, usou modelos lineares generalizados, equações de estimação generalizadas e média móvel sazonal integrada de AR com e sem o efeito da temperatura média diária como variável preditiva para prever o serviço de emergência diário visitas.

3.4.4. Máquina de vetores de suporte (SVM)

De acordo com a análise de co-palavras, os dois trabalhos mais proeminentes que aplicam SVR são [Lu et al. \(2009\)](#), que aplicou análise de componentes independentes e Regressão Vetorial de Suporte para prever séries temporais financeiras, e [Ishak et al. \(2013\)](#) que usam SVR para estimar com precisão a velocidade do vento.

3.4.5. Modelos estruturais

Em modelos estruturais, o trabalho mais proeminente no período 1967-1998 é de [Velicer et al. \(1996\)](#), que desenvolveram um modelo teórico que tenta definir conjuntos multivariados de variáveis dependentes mais apropriados no estudo da mudança de comportamento em saúde, alocados no cluster de estudos em saúde. O segundo período é liderado por [Hernán et al. \(2002\)](#), que usaram modelos estruturais marginais para estimar o efeito da terapia com zidovudina na contagem média de CD4 entre homens infectados pelo HIV. No terceiro período, [Athanasopoulos e Hyndman \(2008\)](#) usaram uma estrutura de regressão para estimar relações econômicas importantes para a demanda de turismo doméstico na Austrália e modelos de espaço de estado de inovação para prever as mesmas. O principal trabalho que aplicou modelos estruturais no quarto período é de [Bergel-Hayat et al. \(2013\)](#), que destacou a ligação entre as condições meteorológicas e o risco de acidente rodoviário.

A [Tabela 6](#) resume os pontos fortes e fracos de cada método apresentado na [Fig. 7](#), bem como os artigos mais proeminentes para os mesmos períodos, extraídos da [Tabela 5](#) e da [Fig. 5](#).

4. Conclusão

Este estudo aplicou um protocolo SLR rigoroso e reproduzível resultando na seleção de 1547 artigos relacionados à análise de séries temporais

com variáveis explicativas. A análise de séries temporais com variáveis exógenas não é uma metodologia recente, com a primeira publicação datando de 1969. No entanto, há um crescimento exponencial no número de publicações após 1996. O crescente número de áreas temáticas de pesquisa que aplicam os modelos, variando da ciência da computação à medicina, mostra a proeminência da análise de séries temporais com variáveis explicativas em uma gama muito diversificada de disciplinas. O maior número de artigos pertence às ciências ambientais, como esperado dada a complexidade de modelagem e previsão de dados como fenômeno El Niño, mancha solar, escoamento de água, aflúências e vazões, concentração de CO2, qualidade do ar.

Quatro questões de pesquisa nortearam a análise. A análise de citação foi aplicada para responder à primeira pergunta “Quem são os autores mais prolíficos na análise de séries temporais com variáveis explicativas?” A análise de citações mostra que o autor mais prolífico na área de pesquisa é Beck, N., com base no número de artigos. É igualmente o autor com maior número de citações. No entanto, não há uma correlação direta entre o número de artigos e citações. A análise também evidenciou as pesquisas mais citadas em séries temporais com variáveis exógenas de ciências sociais e pesquisas econômicas originadas em instituições sediadas nos Estados Unidos e no Reino Unido.

A análise de cocitação respondeu à segunda pergunta “Quais são os trabalhos mais influentes na análise de séries temporais com variáveis explicativas?” Os resultados apresentam os trabalhos mais influentes na área de interesse em ordem cronológica e mostram que eles são as referências para os métodos de séries temporais mais proeminentes e os principais testes estatísticos. A análise de cocitação mostra a espinha dorsal do trabalho seminal na análise de séries temporais com variáveis explicativas e pode ser um guia cronológico para qualquer pesquisador que queira se aprofundar no assunto.

As duas últimas perguntas, “Quais são os principais temas na análise de séries temporais com variáveis explicativas e como eles evoluíram?” e “Quais são os principais métodos aplicados na análise de séries temporais com variáveis explicativas e como evoluíram?” foram respondidas por meio de análise de co-palavras e divisão do banco de dados em quatro períodos distintos. A análise de co-palavras mostra as principais correntes de pesquisa contemporâneas e sua evolução. Os temas de previsão e estudos de saúde são os temas predominantes nos quatro períodos. Quanto aos principais métodos, o ARIMA de Box e Jenkins com a incorporação de variáveis explicativas liderou no primeiro período (1967-1998), seguido por modelos de regressão. Atualmente (2013–2016), os modelos de regressão mantêm a

Tabela 6
Pontos fortes e fracos dos métodos de séries temporais.

Método	Forças	Fraquezas	Referências
Modelos de regressão	Não requer alto poder computacional. A relação entre as variáveis exógenas e de resposta é aberta e, na maioria das vezes, interpretável.	É sensível a outliers. Pressupõe que os erros do modelo são independentes e seguem uma distribuição normal com média zero e variação constante.	Goldstein et al. (1994) Marcelino et al. (2006) Sistrom et al. (2009) Ishak et al. (2013)
ANN	Alto poder de processamento de dados devido à sua estrutura massivamente distribuída e sua capacidade de aprender e, portanto, de generalizar, produzindo saídas adequadas para entradas que não estavam presentes durante o treinamento.	Requer um alto poder computacional e uma grande quantidade de dados no processo de treinamento. É considerado um modelo caixa-preta, uma vez que a relação entre variáveis exógenas e de resposta é impossível de ser interpretada.	Jorquera et al. (1998) González et al. (2005) Guresen et al. (2011) Nourani et al. (2013)
ARIMAX	Com base em uma metodologia já estabelecida na análise de séries temporais. Fácil de encontrar vários testes formais para verificar a adequação do modelo.	Assume que a relação entre as variáveis é linear. Pressupõe que os erros do modelo são independentes e seguem uma distribuição normal com média zero e variação constante.	Yang et al. (1996) Low et al. (2006) Pisoni et al. (2009) Marcilio et al. (2013)
SVM	Eficiente na resolução de problemas não lineares. Robustez para um pequeno número de pontos de dados.	Requer um alto poder computacional e uma grande quantidade de dados no processo de treinamento. Falta de transparência nos resultados.	Lu et al. (2009) Ishak et al. (2013)
Modelos estruturais	Lida com dados ausentes de maneira eficiente. Lida facilmente com quebras estruturais, mudanças e parâmetros que variam no tempo.	A determinação dos valores iniciais dos parâmetros pode ser um desafio. Assume que a relação entre as variáveis é linear. Pressupõe que os erros do modelo são independentes e seguem uma distribuição normal com média zero e variância constante.	Velicer et al. (1996) Hernán et al. (2002) Athanasopoulos e Hyndman (2008) Bergel-Hayat et al. (2013)

Tabela 7
Respostas-chave às Questões de Pesquisa (QR).

Análises RQ	Principais respostas
1	Citação O autor com maior número de artigos é Beck, N., mas aquele com maior índice h é Wigley, T., mostrando que os autores com maior número de artigos na área não são necessariamente os mais citados.
2	Cocitação Os trabalhos mais influentes são as referências para os métodos de séries temporais mais proeminentes (Box e Jenkins, 1970; Haykin, 1999; Takagi e Sugeno, 1985 etc.) e os principais testes estatísticos (Akaike, 1974; Ljung e Box, 1978; Philips e Perron, 1988 etc.).
3	Co-palavra Quatro períodos (1967–1998, 1999–2007, 2008–2012 e 2013–2016) são suficientes para analisar a evolução dos temas. Os principais temas em todos os períodos são a previsão e os estudos de saúde.
4	Co-palavra Os principais métodos, em número de aplicações, são os modelos de Regressão, Redes Neurais Artificiais, ARIMAX, Máquina de Vetores de Suporte e Modelos Estruturais. No primeiro período, o ARIMAX lidera em número de trabalhos, mas perde lugar para o modelo Regressivo. Support Vector Machine cresce após o terceiro período.

primeiro lugar como método de escolha e as RNAs ocupam o segundo lugar em número de aplicações, seguidas pelas SVMs. O crescimento dos métodos de IA pode ser atribuído ao aumento da capacidade de processamento dos computadores e suas características atrativas, sendo máquinas de aprendizado semiparamétricas, permitindo a aproximação universal de funções lineares e não lineares arbitrárias a partir de exemplos sem suposições a priori sobre a estrutura do modelo e muitas vezes superando as convencionais. métodos de abordagem estatística (Crone et al., 2006).

A Tabela 7 resume as respostas encontradas para cada uma das Pesquisas Questões (RQ) elaboradas na subseção 2.1.

A incorporação de variáveis exógenas nos modelos de previsão são mais prevalentes em pesquisas que visam melhorar a precisão da previsão, identificando as covariáveis mais importantes que afetam a resposta (por exemplo, causalidade, relações entre variáveis), verificando a elasticidade entre a variável de saída e sua entradas, e gerar cenários sintéticos para a variável dependente levando em consideração diferentes cenários para as variáveis independentes. Um esforço considerável é igualmente dedicado à escolha do próprio conjunto de variáveis exógenas, bem como aos métodos de previsão e simulação de diferentes cenários.

Esta revisão apresenta algumas limitações devido ao amplo escopo que abrange diferentes áreas de pesquisa, à escolha dos parâmetros de busca (palavras-chave e critérios de exclusão) e das bases de dados de citações (somente Scopus), oferecendo, portanto, uma amostra limitada de artigos relevantes. Essas limitações abrem caminho para novos caminhos de pesquisa. Primeiro, como a revisão não está focada em uma área de pesquisa específica, uma análise mais aprofundada e mapeamento dos métodos de séries temporais com variáveis exógenas nas áreas de pesquisa mais prevalentes de ciências ambientais, economia e saúde parecem ser oportunas. Em segundo lugar, os principais métodos descritos aqui podem ser descritos com mais detalhes com orientações sobre quando e em quais casos tais métodos devem ou não ser aplicados. Em terceiro lugar, a análise da evolução de temas e métodos de séries temporais por disciplina e períodos poderia ser prosseguida. Poderia levar a uma análise muito mais detalhada dos diferenciais na pesquisa frente a séries temporais com variáveis explicativas em diferentes disciplinas. Quarto, a combinação de outras bases de dados, além da Scopus, pode levar a uma revisão mais exaustiva, apesar da grande interseção entre documentos que figuram em diferentes bases de dados de citações. Quinto, a inclusão de material não revisado por pares, como teses e dissertações, pode enriquecer a análise e levar à inclusão de novos métodos e aplicações ainda a serem publicados em periódicos revisados por pares, contornando o risco de viés de publicação associado à revisão por pares. literatura.

Reconhecimento

Os autores agradecem o apoio das seguintes agências brasileiras: Conselho Nacional de Desenvolvimento Científico e Tecnológico, Bolsas nº 304931/2016-0, 404682/2016-2 e 443595/2014-3, e Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Bolsas nº E-26/203.252/2017 e 202.806/2015.

Apêndice A. Dados Suplementares

Dados suplementares relacionados a este artigo podem ser encontrados em <http://dx>.

doi.org/10.1016/j.envsoft.2018.06.004.

Referências

Akaike, H., 1973. A teoria da informação como extensão do princípio da máxima verossimilhança. Em: Segundo Int. Sintoma Inf. Teoria, pp. 267-281.

Akaike, H., 1974. Um novo olhar sobre a identificação do modelo estatístico. Autom. Ao controle. Trans. IEEE 19, 716-723. <https://doi.org/10.1109/TAC.1974.1100705>.

Anderson, TW, Hsiao, C., 1982. Formulação e estimação de modelos dinâmicos usando dados em painel. J. Econom. 18, 47-82. [https://doi.org/10.1016/0304-4076\(82\)90095-1](https://doi.org/10.1016/0304-4076(82)90095-1).

Andriyas, S., McKee, M., 2013. Técnicas de particionamento recursivo para modelagem do comportamento da irrigação. Ambiente. Modelo. Software 47, 207-217. <https://doi.org/10.1016/j.envsoft.05.05.2013>.

Athanasopoulos, G., Hyndman, RJ, 2008. Modelagem e previsão do turismo doméstico australiano. Gestao de Turismo. 29, 19-31.

Athanasopoulos, G., Hyndman, RJ, Song, H., Wu, DC, 2011. A competição de previsão de turismo. Int. J. Previsão. 27, 822-844. <https://doi.org/10.1016/j.ijforecast.2010.04.009>.

Beck, N., Katz, J., Tucker, R., 1998. Em: Beck, Nathaniel, Katz, Jonathan N., Tucker, Richard (Eds.), Levando o tempo a sério: corte transversal da série temporal Análise com uma variável dependente binária, vol. 42. Publicado por: Midwest Political Science Association Stable, pp. 1260–1288. <http://www.jstor.org/stable/2>.

Beck, N., Katz, JN, 1995. O que fazer (e não fazer) com dados de corte transversal de séries temporais. Sou. Policial. Sci. Associação 89, 347-634.

Bergel-Hayat, R., Debbas, M., Antoniou, C., Yannis, G., 2013. Explicando o risco de acidente rodoviário: efeitos climáticos. Ácido. Anal. Anterior 60, 456-465.

Bollerslev, T., 1986. Heteroscedasticidade condicional autoregressiva generalizada. J. Economia 31, 307-327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1).

Box, GEP, Jenkins, GM, 1970. Análise, Previsão e Controle de Séries Temporais. Holden Day, São Francisco.

Box, GEP, Jenkins, GM, Reinsel, GC, 1994. Análise de Séries Temporais: Previsão e Controle. Prentice Hall.

Box, GEP, Pierce, DA, 1970. Distribuição de Autocorrelações Residuais em Modelos de séries temporais de média móvel integradas autorregressiva, vol. 65. pp. 1509-1526.

Callon, M., Courtial, JP, Laville, F., 1991. Análise de co-palavras como ferramenta para descrever a rede de interações entre pesquisa básica e tecnológica: o caso da química de polímeros. Cienciometria 22 (1), 155–205.

Cobo, M., López-Herrera, A., Herrera-Viedma, E., Herrera, F., 2011. Uma abordagem para detectando e visualizando a evolução de um campo de pesquisa: uma aplicação prática ao campo da Teoria dos Conjuntos Fuzzy. J. Informer 5.

Cobo, MJ, López-Herrera, AG, Herrera-Viedma, E., Herrera, F., 2012. SciMAT: uma nova ferramenta de software de análise de mapeamento científico. Geléia. Soc. Inf. Sci. Tecnol. 63, 1609-1630.

Coman, A., Ionescu, A., Candau, Y., 2008. Previsão horária de ozônio para um horizonte de 24 horas usando redes neurais. Ambiente. Modelo. Software 23, 1407-1421. <https://doi.org/10.1016/j.envsoft.2008.04.004>.

Cooper, HM, 2010. Síntese de Pesquisa e Meta-análise: uma abordagem passo a passo, quarta ed. Sálvia, Mil Carvalhos. Mil Carvalhos.

Crone, SF, Guajardo, J., Weber, R., 2006. Um estudo sobre a capacidade de suporte vetorial re regressão e redes neurais para prever padrões básicos de séries temporais. IFIP Int. Fed. Inf. Processo 217, 149-158.

De Nooy, W., Mrvar, A., Bagatelj, V., 2005. Análise de Rede Exploratória com Pajek. Cambridge University Press.

Dickey, DA, Fuller, WA, 1979. Distribuição das estimativas de tempo autoregressivo série com raiz unitária. Geléia. Estado. Associação 366, 427-431.

Durbin, J., Koopman, SJ, 2012. Análise de séries temporais por métodos de espaço de estados. Oxf. Estado. Sci. 38.

Engle, RF, 1982. Heteroscedasticidade condicional autorregressiva com estimativas da variação da inflação do Reino Unido. Econometria 50, 987-1008.

Engle, RF, Granger, CWJ, 1987. Co-integração e correção de erros: representação, estimativa e teste. Econometria 2, 251-276.

Espey, M., Espey, J., Shaw, WD, 1997. Elasticidade-preço da demanda residencial por água: uma meta-análise. Recursos Hídricos. Res. 33, 1369-1374. <https://doi.org/10.1029/97wr00571>.

Fernandez, FJ, Seco, A., Ferrer, J., Rodrigo, MA, 2009. Uso de redes neurofuzzy para melhorar a previsão de vazão de águas residuais. Ambiente. Modelo. Software 24, 686-693.

Goldstein, H., Healy, MJR, Rasbash, J., 1994. Modelos de séries temporais multinível com aplicações para dados de medidas repetidas. Estado. Med. 13, 1643-1655.

González, AM, San Roque, AM, García-González, J., 2005. Modelagem e previsão de preços de eletricidade com modelos ocultos de entrada/saída de Markov. *Trans. IEEE Sistema de energia* 20, 13-24.

Granger, CWJ, 1969. Investigando relações causais por modelos econométricos e métodos espectrais. *Econometria* 37, 424-438.

Guresen, E., Kayakutlu, G., Daim, TU, 2011. Usando modelos de redes neurais artificiais na previsão do índice do mercado de ações. *Sistema especialista. Aplic.* 38, 10389-10397.

Hamilton, JD, 1994. *Análise de Séries Temporais*. Imprensa da Universidade de Princeton.

Harvey, AC, 1989. *Previsão, Modelos Estruturais de Séries Temporais e o Filtro de Kalman*. Cambridge University Press, Cambridge.

Haykin, S., 1999. *Redes Neurais: uma Fundação Abrangente*. Prentice Hall, Englewood Cliffs.

Hernán, MA, Brumback, BA, Robins, JM, 2002. Estimando o efeito causal da zido vudina na contagem de Ccd com um modelo estrutural marginal para medidas repetidas. *Estado. Com.* 21, 1689-1709.

HLWIKI, 2017. HLWIKI Canadá.

Hornik, K., Stinchcomb, M., White, H., 1989. Redes multicamadas feed-forward são aproximadores universais. *Rede neural.* 2, 359-366.

Ishak, AM, Remesan, R., Srivastava, PK, Islam, T., Han, D., 2013. Correção de erros modelagem da velocidade do vento através de parâmetros hidrometeorológicos e modelo de mesoescala: uma abordagem híbrida. *Recursos Hídricos. Gerenciar* 27, 1-23.

Jang, JSR, 1993. ANFIS: sistema de inferência fuzzy baseado em rede adaptativa. *Trans. IEEE Sistema Cara. Cybern* 3, 665-685.

Johansen, S., 1988. *Análise estatística de vetores de cointegração*. J. Eco. Dinâmica. Contr. 12, 231-254. [https://doi.org/10.1016/0165-1889\(88\)90041-3](https://doi.org/10.1016/0165-1889(88)90041-3).

Johansen, S., Juselius, K., 1990. Estimativa de máxima verossimilhança e inferência sobre integração de moedas com aplicações à demanda por moeda. *xford Bull. Eco. Estado* 52, 169-210.

Johansen, SJ, 1991. Estimação e teste de hipóteses de vetores de cointegração em modelos de autoregressão vetorial gaussiana. *Econometria* 59, 1551-1580.

Jorquera, H., Pérez, R., Cipriano, A., Espejo, A., Letelier, MV, Acuña, G., 1998. Previsão dos níveis máximos diários de ozônio em santiago, Chile. *Atmos. Ambiente.* 32, 3415-3424.

Lima, LMM, Popova, E., Damien, P., 2014. Modelagem e previsão de re fluxos de servoir através de modelos lineares dinâmicos. *Int. J. Previsão.* 30, 464-476. <https://doi.org/10.1016/j.ijforecast.2013.12.009>.

Ljung, GM, Box, GE, 1978. Em uma medida de falta de ajuste em modelos de séries temporais. *Biometria* 65. <https://doi.org/10.1093/biomet/65.2.297>.

Ljung, L., 1999. *Identificação do Sistema: Teoria para o Usuário*. Prentice Hall PTR.

Low, RB, Qureshi, AI, Dunn, V., Stuhlmiller, DFE, Dickey, DA, 2006. A relação de admissões por acidente vascular cerebral com clima recente, alérgenos transportados pelo ar, poluição do ar, estações do ano, infecções respiratórias superiores e incidência de asma, 11 de setembro , 2001 e dia da semana. *AVC* 37, 951-957.

Lu, C.-J., Lee, T.-S., Chiu, C.-C., 2009. Previsão de séries temporais financeiras usando análise de componentes independentes e regressão vetorial de suporte. *Decis. Sistema de suporte* 47, 115-125.

Marcellino, M., Stock, JH, Watson, MW, 2006. Uma comparação de métodos Ar multipasso direto e iterado para previsão de séries temporais macroeconômicas. *J. Econom.* 135, 499-526.

Marçílio, I., Hajat, S., Gouveia, N., 2013. Previsão de visitas diárias ao serviço de urgência utilizando variáveis de calendário e leituras de temperatura ambiente. *Acad. Emerg. Med.* 20, 769-777.

McCullagh, P., Nelder, JA, 1989. *Generalized Linear Models*. Chapman & Hall/CRC.

Meese, RA, Rogoff, K., 1983. Modelos empíricos de taxa de câmbio dos anos setenta. Eles se encaixam fora da amostra? *J. Int. Eco.* 14. [https://doi.org/10.1016/0022-1996\(83\)90017-X](https://doi.org/10.1016/0022-1996(83)90017-X).

Milonis, AE, Davies, TD, 1994. Regressão e modelos estocásticos para poluição do ar-I. Críticas, comentários e sugestões. *Atmos. Ambiente.* 28, 2801-2810. [https://doi.org/10.1016/1352-2310\(94\)90083-3](https://doi.org/10.1016/1352-2310(94)90083-3).

Mongeon, P., Paul-Hus, A., 2016. A cobertura da revista Web of Science e Scopus: uma análise comparativa. *Cienciometria* 106, 213-228.

Nash, JE, Sutcliffe, JV, 1970. previsão de vazão do rio através de modelos conceituais Parte Ia discussão de princípios*. *J. Hydrol* 10, 282-290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).

Nourani, V., Baghanam, AH, Adamowski, J., Gebremichael, M., 2013. Usando mapas auto-organizados e transformadas wavelet para pré-processamento espaço-tempo de precipitação de satélite e dados de escoamento em rede neural baseada em precipitação- modelagem de escoamento. *J. Hydrol* 476, 228-243.

Nunnari, G., Dorling, S., Schlink, U., Cawley, G., Foxall, R., Chatterton, T., 2004. Modelagem da concentração de SO2 em um ponto com abordagens estatísticas. *Ambiente. Modelo. Software* 19, 887-905. <https://doi.org/10.1016/j.envsoft.2003.10.003>.

Persson, O., Danell, R., Wiborg Schneider, J., Astrom, R., Danell, R., Larsen, B., Wiborg Schneider, J., 2009. Como usar o Bibexel para vários tipos de análise bibliométrica. In: *Celebrando Estudos de Comunicação Acadêmica: um Festschrift para Olle Persson em seu 60º aniversário*.

Petticrew, R., Roberts, H., 2006. *Revisões Sistemáticas em Ciências Sociais*. Um prático Guia. Blackwell, Malden, MA.

Philips, PCB, Perron, P., 1988. Teste para uma raiz unitária na regressão de séries temporais. *Biometria* 75, 335-346.

Pisoni, E., Farina, M., Carnevale, C., Piroddi, L., 2009. Modelos narrativos usando a previsão de picos de poluição do ar. *Eng. Aplic. Artif. Intel.* 22, 593-602.

Porporato, A., Ridolfi, L., 2001. Predição não linear multivariada de fluxos fluviais. *J. Hydrol* 248, 109-122. [https://doi.org/10.1016/S0022-1694\(01\)00395-X](https://doi.org/10.1016/S0022-1694(01)00395-X).

Rumelhart, DE, Hinton, GE, Williams, RJ, 1986. Aprendizagem de representações internas por propagação de erros. In: *Processamento Distribuído Paralelo: Explorações na Microestrutura da Cognição*. MIT Press, Cambridge, pp. 318-362.

Schwarz, G., 1978. Estimando a dimensão de um modelo. *Ana Estado.* 6, 461-464. <https://doi.org/10.1214/aos/1176344136>.

Scott Jr., JT, Heady, EO, 1967. Demanda regional para edifícios agrícolas nos Estados Unidos Estados. *Sou. J. Agric. Eco.* 49.

Seuring, S., Gold, S., 2012. Realização de revisões de literatura baseadas em análise de conteúdo na gestão da cadeia de suprimentos. *Gestão da Cadeia de Suprimentos*. Um Int. J. 17, 544-555.

Stetsos, a., Coonick, AH, 2000. Previsão univariada e multivariada da radiação solar horária com técnicas de inteligência artificial. *Sol. Energia* 68, 169-178. [https://doi.org/10.1016/S0038-092X\(99\)00064-X](https://doi.org/10.1016/S0038-092X(99)00064-X).

Sistrom, CL, Dang, PA, Weilburg, JB, Dreyer, KJ, Rosenthal, DI, Thrall, JH, 2009. Efeito da entrada de pedidos computadorizada com suporte integrado à decisão no crescimento dos volumes de procedimentos ambulatoriais: análise de séries temporais de sete anos. *Radiologia* 251, 147-155.

Small, H., 1999. Visualizando a ciência por mapeamento de citações. *Geléia. Soc. Inf. Sci.* 50, 799-813.

Takagi, T., Sugeno, M., 1985. Identificação difusa de sistemas e sua aplicação à modelagem e controle. *Trans. IEEE Sistema Homem, Cybern* 15, 116-132.

Thomé, AMT, Scavarda, LF, Scavarda, A., 2016a. Realização de revisão sistemática da literatura em gestão de operações. *Prod. Plano Contr.* 27, 408-420.

Thomé, A.M.T., Scavarda, L.F., Scavarda, A., Thomé, F.S.S., 2016b. Similarities and contrastes de complexidade, incerteza, riscos e resiliência em cadeias de suprimentos e projetos multiorganizacionais temporários. *Int. J. Proj. Gerenciar* 34, 1328-1346.

Thomé, A.M.T., Maçaira, P., Cyrino Oliveira, F.L., Ferrer, A.L.C., 2017. Citation Data in Time Series, Mendeley Data, V1. <http://dx.doi.org/10.17632/4svhzhzw8d.1>.

Tong, H., 1990. *Séries Temporais Não-lineares: uma Abordagem de Sistemas Dinâmicos*. Oxford University Press, Nova York.

Velicer, WF, Rossi, JS, Prochaska, JO, Diclemente, CC, 1996. Um modelo de mensuração de critérios para mudança de comportamento em saúde. *Viciado. Comportamento* 21, 555-584.

White, H., 1980. Um estimador de matriz de covariância consistente com heterocedasticidade e um teste direto para heterocedasticidade. *Econometria* 48, 817-838.

Wilby, RL, Wigley, TML, Conway, D., Jones, PD, Hewitson, BC, Main, J., Wilks, DS, 1998. Redução estatística da saída do modelo de circulação geral: uma comparação de métodos. *Recursos Hídricos. Res.* 34, 2995-3008. <https://doi.org/10.1029/98wr02577>.

Yang, H.-T., Huang, C.-M., Huang, C.-L., 1996. Identificação do modelo armax para previsão de carga de curto prazo: uma abordagem de programação evolutiva. *Trans. IEEE Sistema de energia* 11, 403-408.

Young, PC, 2011. *Estimativa Recursiva e Análise de Séries Temporais: uma Introdução para a Estudante e Praticante*. Springer-Verlag, Berlim.

Zhang, GP, 2003. Previsão de séries temporais usando um modelo híbrido ARIMA e rede neural . *Neurocomputing* 50, 159-175.

Zuur, AF, Fryer, RJ, Jolliffe, IT, Dekker, R., Beukema, JJ, 2003. Estimativa de tendências comuns em séries temporais multivariadas usando análise fatorial dinâmica. *Ambientes* 14, 665-685.

Zuur, AF, Pierce, GJ, 2004. Tendências comuns em séries temporais de lulas no nordeste do Atlântico. *J. Ver Res.* 52, 57–72.