



Campus São Mateus
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO



A PRODUÇÃO DO CONHECIMENTO CIENTÍFICO SOBRE HIERARCHICAL TIME SERIES FORECASTING: UMA ABORDAGEM BIBLIOMÉTRICA

THE PRODUCTION OF SCIENTIFIC KNOWLEDGE ABOUT HIERARCHICAL TIME SERIES FORECASTING: A BIBLIOMETRIC APPROACH

LA PRODUCCIÓN DE CONOCIMIENTO CIENTÍFICO SOBRE LA PRONÓSTICA DE SERIES TIEMPO JERARQUICAS: UN ENFOQUE BIBLIOMÉTRICO

Tiago Silveira Gontijo¹, Ana Luiza Jentsch Nepomuceno², Bruna Peralva Lima Paiva³, Rodrigo Barbosa de Santis^{4*}, Gustavo de Souza Groppo⁵, & Marcelo Azevedo Costa⁶

¹ Universidade Federal de São João del-Rei, Campus Centro Oeste (CCO)

^{2 3 4 6} Universidade Federal de Minas Gerais, Departamento de Engenharia de Produção

⁵ Universidade Federal de Minas Gerais, Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos

⁶ Universidade Federal de Minas Gerais, Programa de Pós-Graduação em Engenharia de Produção

¹ tsg@ufsj.edu.br ² anajnepo@gmail.com ³ brunaplpaiva@gmail.com ^{4*} rodrigo.santis@live.com ⁵ macosta@ufmg.br

⁶ macosta@ufmg.br

ARTIGO INFO.

Recebido: 11.12.2023

Aprovado: 29.12.2023

Disponibilizado: 09.02.2024

PALAVRAS-CHAVE: Séries Temporais Hierárquicas; Modelos Preditivos; Bibliometria.

KEYWORDS: Hierarchical Time Series; Predictive Models; Bibliometry.

PALABRAS CLAVE: APA; Artículos científicos; BJPE.

*Autor Correspondente: Santis, R. B., de.

RESUMO

Os estudos sobre as séries temporais hierárquicas (*hierarchical time series forecasting*) têm despertado a atenção da literatura. Em geral, uma série temporal hierárquica consiste em um conjunto de informações coletadas no tempo, as quais estão organizadas através de agrupamentos como a localização geográfica, o tipo de produto e outros atributos, por exemplo. O presente artigo realizou de maneira pioneira uma pesquisa bibliométrica acerca de todas as publicações indexadas junto à base da Web of Science, sobre as séries temporais hierárquicas, durante o período de 1996 a 2020. Destaca-se que o crescimento anual das publicações sobre esse tema é igual a 13.45% e que os Estados Unidos da América se destacam como o maior polo produtor de conhecimentos sobre as séries hierárquicas, concentrando aproximadamente 30% das pesquisas. Por fim, verificou-se que métodos híbridos para a reconciliação ótima de previsões, baseadas em algoritmos de *machine learning* tem sido recorrentes em pesquisas atuais.

ABSTRACT

Studies on hierarchical time series have attracted the attention of the literature. In general, a hierarchical time series consists of the set of information collected during a certain period, which is organized through groups such as geographic location, type of product and other attributes, for example. Due to the relevance of this theme, this article pioneered bibliometric research on all publications indexed by the Web of Science, on hierarchical time series, from 1996 to 2020. It is noteworthy that the annual growth of publications on this topic is equal to 13.45% and that the United States of America stands out as the largest producer of knowledge on hierarchical series, concentrating approximately 30% of research. Finally, it was found that hybrid methods for optimal reconciliation of predictions, based on machine learning algorithms have been recurrent in current research.

RESUMEN

Los estudios sobre series de tiempo jerárquicas han llamado la atención en la literatura. En general, una serie temporal jerárquica consiste en un conjunto de información recopilada a lo largo del tiempo, que se organiza mediante agrupaciones como ubicación geográfica, tipo de producto y otros atributos, por ejemplo. Este artículo realizó, de manera pionera, una investigación bibliométrica sobre todas las publicaciones indexadas en la base de datos Web of Science, en series temporales jerárquicas, durante el período de 1996 a 2020. Es de destacar que el crecimiento anual de las publicaciones sobre este tema es equivalente al 13,45% y que Estados Unidos de América se destaca como el mayor polo de producción de conocimiento sobre series jerárquicas, concentrando aproximadamente el 30% de la investigación. Finalmente, se encontró que los métodos híbridos para la conciliación óptima de pronósticos, basados en algoritmos de aprendizaje automático, han sido recurrentes en las investigaciones actuales.

INTRODUÇÃO

Apesar de recentes, os estudos sobre as séries temporais hierárquicas (*hierarchical time series forecasting*) têm despertado o interesse da literatura nacional (Lauretto et al., 2008; Lauretto et al., 2009; Gontijo & Costa, 2020) e internacional (Fliedner, 2001; Hyndman et al., 2011; Athanasopoulos et al., 2020). Esse interesse é justificado por razões distintas: (i) trata-se de uma técnica inovadora, que se popularizou após o ano de 2010; (ii) esse método tem sido apontado como promissor e tem sido utilizado de modo incipiente nas principais competições de séries temporais do mundo, como é o caso da *M5 Competition* (Bojer et al., 2020).

Existe uma extensa literatura a respeito dos modelos preditivos utilizando as séries temporais clássicas, entretanto, para alguns tipos específicos de aplicações, os modelos tradicionais podem não ser tão eficientes para realizar previsões de modo acurado (Nielsen & Madsen, 2001; Faloutsos et al., 2019). Nesse sentido, para sistemas -em que as séries temporais estão relacionadas entre si, há relatos na literatura de que os modelos hierárquicos são promissores (Sun et al., 2018). Em geral, as séries temporais hierárquicas são aquelas que podem ser agregadas em diferentes níveis, com base em alguma característica em comum (Gontijo & Costa, 2020).

Destaca-se que o rol de aplicações das previsões hierárquicas é variado. Assim, tem-se que é comum fazer previsões em diferentes níveis de agregação, por exemplo: uma empresa de varejo pode realizar as previsões de venda de seus produtos a nível nacional, estadual e em nível de localidades (Ma, Fidelis & Huang, 2016). Ainda de acordo com o referido exemplo, a empresa também pode estar interessada em prever o volume de vendas para todos os produtos comercializados, por um grupo específico de produtos ou para um produto individualmente (Hyndman & Athanasopoulos, 2018).

Sun et al. (2018) apontam que a literatura já existente sobre o assunto é restrita e utiliza, normalmente, poucos métodos hierárquicos, a saber: *Top Down* ou *Bottom Up*, partindo de séries agregadas e as desagregando ou vice-versa. Segundo Karmy & Maldonado (2019), essas abordagens são particularmente úteis na previsão de demanda de múltiplas categorias quando o nível hierárquico inferior tem dados escassos ou incompletos, o que poderia gerar previsões com ruídos.

Considerando a demanda crescente pela realização de previsões em diversos âmbitos da sociedade e a necessidade de que essas previsões sejam cada vez mais precisas, os estudos das séries temporais e da sua aplicação em modelos hierárquicos se tornam ainda mais relevantes. Ademais, tendo em vista a complexidade de sistemas reais e a dificuldade de modelar esses sistemas através de alguns modelos tradicionais, bem como a escassez de estudos direcionados para o tema, o entendimento da aplicação de séries temporais hierárquicas faz-se necessário.

Desta forma, visando um aprofundamento ainda maior no tema, objetiva-se realizar uma pesquisa bibliométrica sobre o tema, considerando publicações de artigos, congressos e livros a respeito das séries temporais hierárquicas e suas aplicações em modelos preditivos. A partir do levantamento das pesquisas, materiais publicados e da natureza deles, é possível examinar

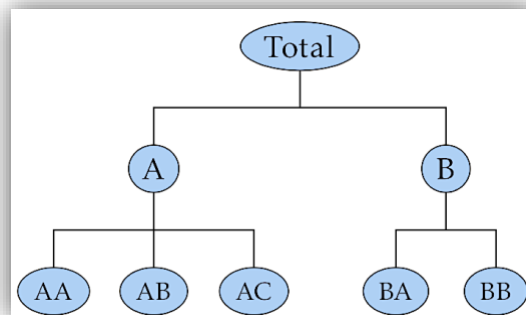
a relevância do tema, a frequência de estudo de determinadas técnicas, as áreas mais exploradas e àquelas áreas a serem examinadas em pesquisas futuras.

Esse artigo está dividido em cinco seções, iniciando por esta introdução. Na segunda seção será apresentado um referencial teórico sobre a previsão hierárquica. Na terceira seção do trabalho será definida a metodologia da pesquisa bibliométrica. Em seguida, na seção quatro serão expostos os resultados encontrados da pesquisa. Por fim, na quinta e última seção serão apresentadas as conclusões, discussões e sugestões de pesquisas futuras.

PREVISÃO HIERÁRQUICA

As previsões hierárquicas se baseiam em uma estratégia de agregar itens em famílias e partir desse sistema, é possível fornecer previsões tanto para os itens separadamente, quanto para as famílias geradas a partir desta agregação (Flidner, 2011). A Figura 1 apresenta uma estrutura hierárquica hipotética, contendo um valor total dividido em categorias principais (A e B) e em subcategorias (A, B e C).

Figura 1. Exemplo de série hierárquica.



Fonte: Hyndman & Athanasopoulos (2018).

Na literatura disponível sobre o tema, duas abordagens são mais comuns. Na primeira delas, parte-se da previsão da série agregada para, em seguida, desagregar as previsões, de forma que a previsão agregada seja igual à soma das previsões desagregadas. Essa abordagem é chamada de *Top-Down*, em que são definidas proporções que indicam como as previsões do nível mais alto devem ser desagregadas para se obter as previsões dos níveis inferiores. Existem três métodos para obter as proporções do *Top-down* pelo R package *hts* (Hyndman, Athanasopoulos & Shang, 2014):

- i. **Average historical proportions:** para cada observação, existe uma proporção, que é a relação daquela série com a série agregada, então é feito o cálculo das médias que os valores das proporções de cada série desagregada assumem com relação ao nível superior de agregação (Pinheiro, 2015);
- ii. **Proportions of historical averages:** proporção é calculada pela relação da soma dos valores assumidos pela série inferior com relação à soma de valores assumidos pela série agregada (Pinheiro, 2015);
- iii. **Forecasted proportions:** consiste na desagregação dos níveis inferiores da hierarquia com base na proporção das previsões do nível imediatamente inferior (Pinheiro, 2015).

A vantagem do método *Top-Down* é a precisão das previsões no nível hierárquico mais alto e a desvantagem é a geração de previsões viesadas, mesmo quando a previsão base é não viesada (Hyndman et al., 2007).

Já na segunda abordagem, chamada de *Bottom-Up*, é realizada a previsão em cada uma das séries desagregadas no nível hierárquico mais baixo e, posteriormente, todas essas previsões são agregadas gerando as previsões dos níveis superiores. De acordo com Veríssimo (2016), esta abordagem utiliza uma regressão linear para prever cada série do nível mais baixo da hierarquia de forma independente. Para obter as previsões dos níveis superiores, os dados são somados de acordo com a hierarquia para obter as previsões dos níveis superiores.

Ainda de acordo com a abordagem *Bottom-Up*, é possível gerar boas previsões individuais nos níveis hierárquicos mais baixos, os resultados gerados não são viesados, existe pouca perda de informação quando os dados são trabalhados em níveis desagregados, além de ser um método de fácil aplicação. Entretanto, no caso de um nível hierárquico inferior com muitas séries, pode ser que os ruídos sejam somados na agregação, prejudicando a previsão em níveis superiores (Hyndman et al., 2007).

Por serem as duas abordagens mais utilizadas ao realizar este tipo de previsão, vários trabalhos sobre a comparação das abordagens *Top-Down* e *Bottom-Up* foram realizados. Segundo Almeida et al. (2016), algumas pesquisas apontam que a abordagem *Top-Down* é mais precisa, enquanto outras mostram que o método *Bottom-Up* seria mais acurado, com base nas vantagens e desvantagens de cada método, apresentadas anteriormente.

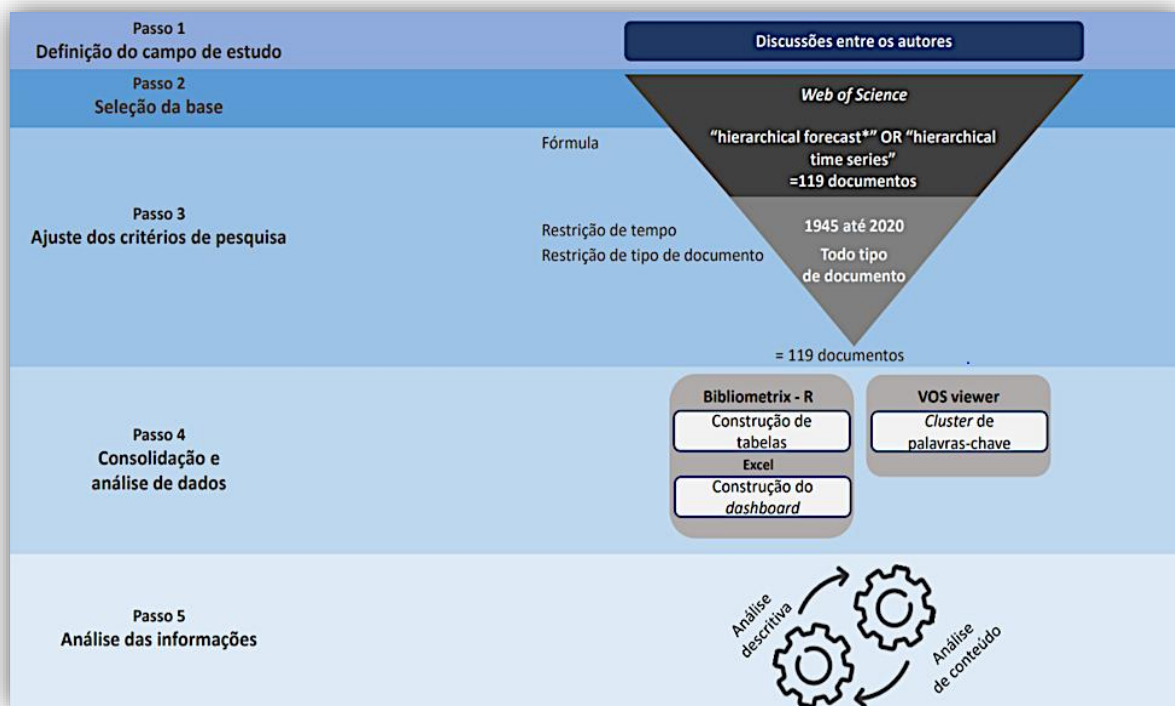
Como uma combinação dos modelos *Top-Down* e *Bottom-Up*, existe também a abordagem *Middle-Out* em que um nível intermediário é a base para serem feitas as previsões. Para o cálculo das previsões dos níveis inferiores, a desagregação é feita pela abordagem *Top-Down* e para a agregação para os níveis superiores, utiliza-se o método *Bottom-Up* (Hyndman et al., 2007).

Há, ainda, a abordagem da reconciliação ótima, em que a previsão revisada em cada nível será uma média ponderada das previsões feitas para todos os níveis da previsão base (De Freitas, 2016). O modelo introduzido por Hyndman et al. assume que a distribuição dos erros de previsão é a mesma que a estrutura hierárquica agregada. Para este método, são geradas previsões independentes para todas as séries em todos os níveis hierárquicos e, em seguida, é aplicado um modelo de regressão linear para otimizar a combinação das previsões (Wickramasuriya, Athanasopoulos & Hyndman, 2015).

METODOLOGIA

Este artigo fez uso do repositório de pesquisa *Web of Science* (WoS). Esta plataforma, originou-se na década de 1960, e trata-se na atualidade do mais antigo banco de publicações científicas do mundo. A WoS possui em seu catálogo registros superiores a 20 mil *journals* e aglutina bancos de dados, que incluem artigos, documentos de congressos, livros e documentos diversos. A escolha da WoS se deu pelo fato de que a temática das séries temporais hierárquicas é recente (Hyndman et al., 2007), logo, há um sombreamento de metadados em diferentes bases sobre o tema. Outro fator que justifica a escolha é o fato da WoS é o fato de que suas publicações são selecionadas mediante rigorosos critérios científicos (<https://clarivate.com/products/web-of-science/web-science-form/web-science-core-collection/>) (Figura 2).

Figura 2. Passo a passo do processo de busca



Fonte: Autores.

A consulta ao Web of Science (WoS) foi realizada no dia 8 de outubro de 2020 para identificar as publicações relacionadas ao tema: "Séries Temporais Hierárquicas e Previsão Hierárquica". O critério de busca foi a expressão em inglês "hierarchical forecast*" OR "hierarchical time series", em que o caracter * permite a identificação de todas as variações do termo *forecast*. Durante a pesquisa, não houve restrição de idioma ou de tipo de documentos. Além disso, o intervalo de tempo estipulado foi de 1945 até 2020.

A partir do repositório Web of Science extraiu-se uma base de dados do tipo "txt.", contendo todos os trabalhos já publicados sobre o tema, para a janela temporal analisada. O arquivo armazenou algumas informações daquelas publicações selecionadas, a saber, o nome dos autores, o idioma de publicação, o título do trabalho de pesquisa o ano de sua publicação, o resumo e as referências bibliográficas. No tratamento destes dados, utilizou-se o *software* R® para realizar uma análise bibliométrica. Especificamente, este artigo utilizou o pacote Bibliometrix do referido *software*.

Segundo Aria & Cuccurullo (2017) o pacote Bibliometrix é flexível, pois trata-se de uma interface interativa que pode ser integrada com outros pacotes estatísticos. Essa interface do Biblioshiny foi carregada utilizando o comando "biblioshiny", dentro do ambiente R, habilitando assim, uma página *web* para a utilização dos principais recursos do pacote Bibliometrix. Através da interface é possível realizar a importação, coleta e a filtragem de dados, bem como realizar Analytics e Plots abordando as fontes, autores e documentos mais importantes. Com a interface carregada, a base extraída do Web of Science foi importada a fim de realizar as devidas análises com o Biblioshiny.

Além do R, também foi utilizado o *software* VOSviewer (www.vosviewer.com) para a elaboração de um mapa de palavras. Os mapas de palavras-chave fornecem visões gerais para identificar a estrutura de um tópico (Van Eck & Waltman, 2010). Dessa forma, é possível visualizar a relação entre termos amplamente citados nas publicações e sua relevância. O tamanho de cada palavra é proporcional à quantidade de vezes que foi citado e a localização das palavras se relaciona sua a similaridade com as demais, ou seja, palavras com alta similaridade são mais próximas do que palavras com baixa ou nenhuma similaridade, mostrando uma tendência de palavras que são citadas nas mesmas publicações. Além disso, o mapa também permite visualizar como tem sido o uso de cada termos ao longo do tempo, de acordo com a data que as publicações foram feitas.

Diante do exposto, foi possível aplicar um arcabouço de métodos matemáticos e estatísticos às publicações acadêmicas (Diodato; Gellatly, 2013; Thompson; Walker, 2015), avaliado assim, diversos aspectos quantitativos da base de dados selecionada.

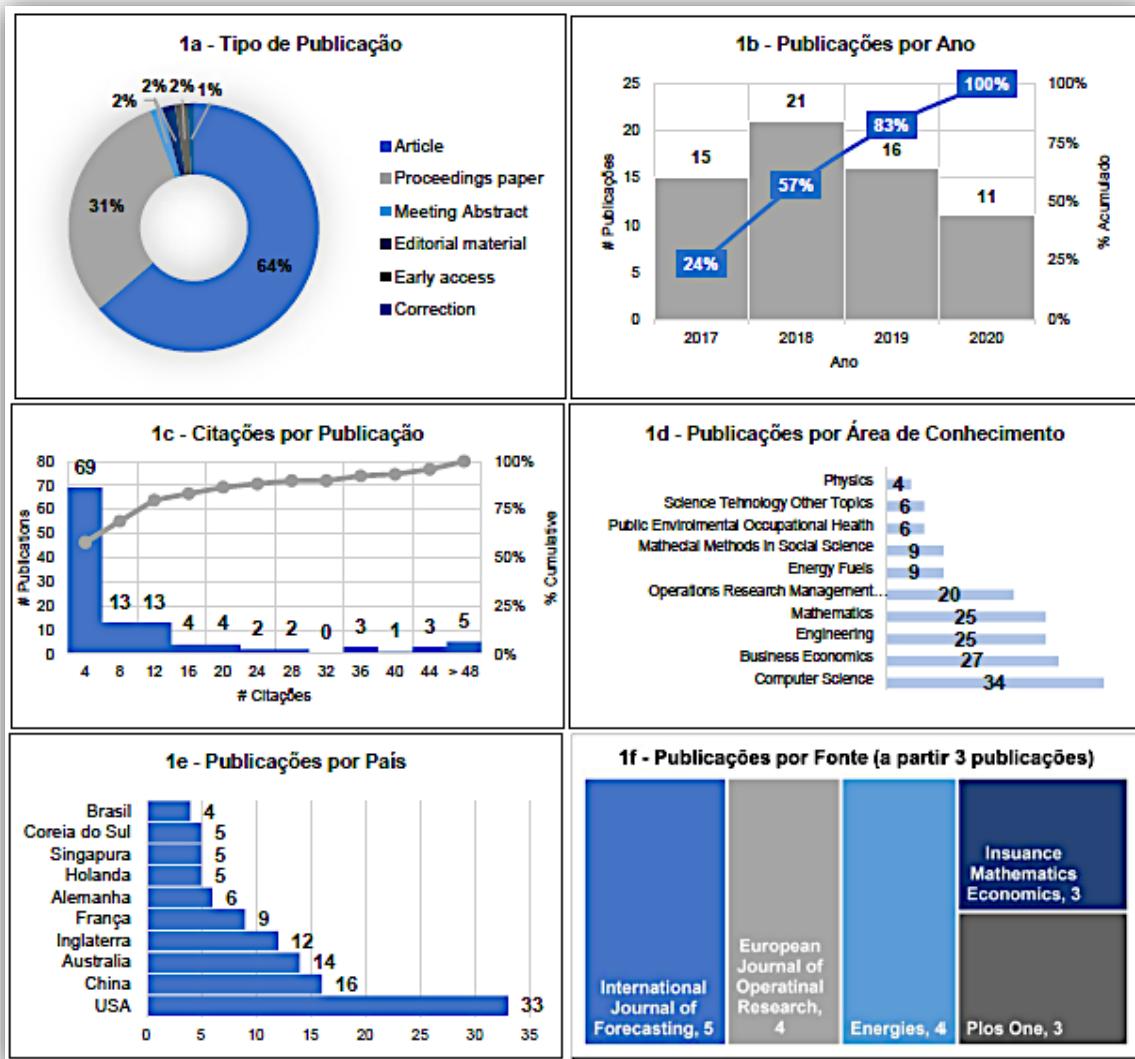
RESULTADOS E DISCUSSÃO

Foram identificadas 119 publicações na pesquisa sobre Séries Temporais Hierárquicas e Previsões Hierárquicas utilizando os parâmetros de pesquisa descritos na seção anterior. O material de pesquisa pioneiro foi publicado em 1996 o que mostra o quanto a área de pesquisa estudada no presente artigo é recente (Berliner, 1996). Essa pesquisa pioneira foi publicada por Mark Berliner na forma de um *proceeding paper* de título “*Hierarchical Bayesian Time Series Model*”.

Em relação ao tipo de publicação, os artigos são o mais comum, representando 64% do total de publicações, seguido dos *proceeding papers*, com 31% (Figura 1a), representando juntos expressivos 95% de todo material publicado. Na Figura 1 pode-se verificar que 2018 foi o ano com mais publicações nos últimos quatro anos, todavia vale ressaltar que o ano de 2020 ainda está em andamento e que é possível que sejam publicados novos registros sobre o tema. Na Figura 1c está representado o número de citações por publicação, em que é possível ver que a maioria das publicações (58%) foram citadas entre 0 a 4 vezes. Por outro lado, cinco artigos concentram o maior número de citações, indicando que esses podem ser considerados referências no tema.

Com 34 publicações, Ciência da Computação é a área do conhecimento predominante nas publicações sobre Séries Hierárquicas (Figura 1d). Além disso, o país com mais publicações sobre o tema é Estado Unidos, totalizando 33 registros (Figura 1e) e o International Journal of Forecasting, com cinco publicações (Figura 1f), é a fonte mais expressiva em número de publicações, seguido do European Journal of Operational Research (Figura 3).

Figura 3. Destaque sobre a pesquisa sobre as publicações já feitas sobre Séries Temporais Hierárquicas e Previsões Hierárquicas



Fonte: Autores.

Das 119 publicações encontradas, foi feito um levantamento visando encontrar pesquisas chave no âmbito das Séries Hierárquicas, que possivelmente apresentariam maior contribuição para a construção do trabalho sobre o tema. Para isso, foram usados os seguintes critérios:

- Publicação pioneira;
- Publicação mais citada;
- Publicação mais recente.

Quadro 1, Artigo pioneiro, artigo mais citado e artigo mais recente.

	Referência	Citações	Título	Contribuição
Pioneiro	Berliner (1996)	147	Hierarchical bayesian time series models	Modelo hierárquico geral para análise de séries temporais, discutindo formulações de tempo discreto e contínuo
Mais citado	Sedgh et al. (2016)	233	Abortion incidence between 1990 and 2014: global, regional and subregional levels and trends	Estimativa da incidência de aborto entre 1990 e 2014 em níveis sub-regionais, regionais e globais por meio de um modelo de séries temporais hierárquicas
Mais recente	Smyl (2020)	-	A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting	Formulação de um método de previsão híbrido e hierárquico (apresentação vencedora da M4 Forecasting Competition)

Fonte: Autores.

O artigo pioneiro (Berliner, 1996) reflete que a área de pesquisa estudada no presente artigo é recente, conforme exposto. O artigo *Hierarchical Bayesian Time Series Models* traz o tema pela primeira vez, abordando de forma geral, focando nos aspectos matemáticos e teóricos dos modelos e não tanto na aplicação em algum case específico. Esta pesquisa obteve 147 citações entre sua publicação e a data de consulta.

Já a publicação mais citada (Sedgh et. al., 2016), data de elaboração do presente artigo, possui 233 citações, um número elevado principalmente considerando que é uma pesquisa recente. Intitulada de *“Abortion incidence between 1990 and 2014: global, regional, and subregional levels and trends”*, a publicação não se trata especificamente do assunto das séries temporais hierárquicas, mas sim da aplicação deste método no estudo de um tema específico, que é a incidência de abortos induzidos no período compreendido entre 1990 e 2014. Dessa forma, para realizar as análises sobre o tema proposto, um modelo de série temporal hierárquica bayesiana foi utilizado para estimar a incidência.

Por fim, a pesquisa mais recente (Smyl, 2020) foi publicada no mesmo ano de elaboração do presente artigo (2020) e ainda não possui citações. *“A hybrid method of exponential smoothing and recurrent neural networks for Time Series Forecasting”* se trata da apresentação do trabalho vencedor da competição M4. A Competição segue as três competições M anteriores, cujo objetivo é aprender com a evidência empírica como melhorar a precisão das previsões e como tal aprendizado poderia ser usado para fazer avançar a teoria e a prática da previsão (Makridakis, Spiliotis & Assimakopoulos, 2020). Na apresentação vencedora retratada no artigo é utilizado um sistema que une um modelo de suavização exponencial padrão com redes de memória de longo prazo em uma só estrutura. O resultado é um método de previsão híbrido e hierárquico (Smyl, 2020).

Como exposto, o artigo que ocupa o primeiro lugar no Quadro 2 obtém um total de 233 citações e o décimo possui 34 citações. Percebe-se que o rol de aplicações sobre *Hierarchical Time Series Forecasting* é amplo, ou seja, há publicações sobre o tema nas áreas da Administração, Economia, Engenharias, Estatística, Medicina e outras. Ainda, as contribuições podem ser divididas entre artigos teóricos, com contribuições no método e nos algoritmos de cálculo, bem como há um significativo número de publicações práticas que buscam a solução de problemas reais.

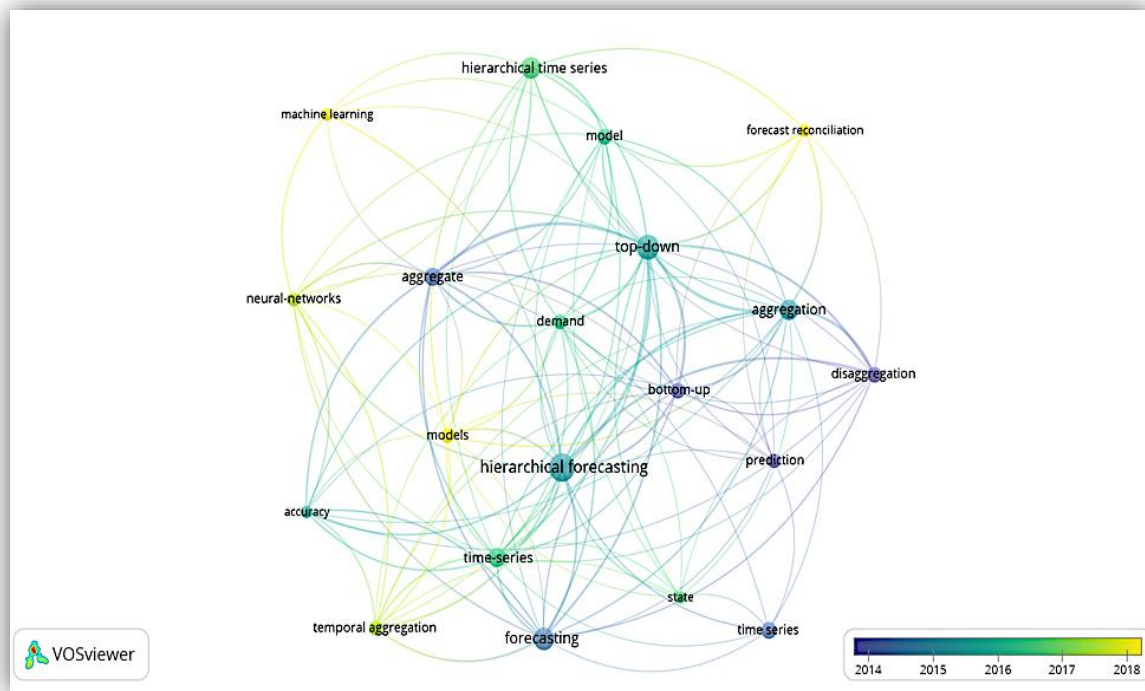
Quadro 2. Os dez artigos sobre *Hierarchical Time Forecasting* que possuem mais citações.

Título da publicação	Autores	Citações	Objeto de investigação
Abortion incidence between 1990 and 2014: global, regional, and subregional levels and trends	Sedgh et al. (2016)	233	Estimativa da incidência do número de abortos entre 1990 e 2014 em níveis sub-regionais, regionais e globais por meio de um modelo de séries temporais hierárquicas.
Hierarchical bayesian time series models	Berliner (1996)	147	Apresenta um modelo hierárquico geral para análise de séries temporais, discutindo formulações de tempo discreto e contínuo.
Global, regional, and subregional trends in unintended pregnancy and its outcomes from 1990 to 2014: estimates from a bayesian hierarchical model	Bearak et al. (2014)	131	Estimativa de gravidez entre 1990 e 2014 em nível global, regional e subregional utilizando o desenvolvimento de um modelo de séries temporais hierárquicas.
Optimal combination forecasts for hierarchical time series	Hyndman et al. (2011)	103	É proposto uma nova abordagem para previsões hierárquicas baseado na previsão independente de todas as séries em todos os níveis da hierarquia, seguido do uso de um modelo de regressão para combinar e reconciliar as previsões de forma otimizada.
Hierarchical forecasts for australian domestic tourism	Athanasopoulos, Ahmed & Hyndman (2009)	67	Previsão de curto prazo para a demanda turística da Austrália, organizando os dados hierarquicamente baseado na desagregação de acordo com as regiões geográficas e objetivos de viagem.
An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation	Fliedner (1999)	43	Análise de estratégias e diretrizes para especificar famílias agregadas em sistemas de previsões hierárquicas.
Forecasting with temporal hierarchies	Athanasopoulos et al. (2017)	42	São discutidas implicações organizacionais das previsões hierárquicas temporais usando um estudo de caso do departamento de acidente e emergência.
National, regional, and global sex ratios of infant, child, and under-5 mortality and identification of countries with outlying ratios: a systematic assessment	Alkema (2014)	41	Estimativa das taxas de mortalidade por sexo para bebês e crianças até 5 anos para todos os países entre os anos de 1990 e 2012, utilizando modelo de séries temporais hierárquicas.
Hierarchical forecasting: issues and use guidelines	Fliedner (2001)	40	Resume as diretrizes práticas para a estruturação de sistemas de previsão hierárquica dos estudos já publicados.
Seasonal patterns in time series of pertussis	De Greeff (2009)	34	Uso de modelos de séries temporais hierárquicas para analisar a incidência da doença Pertussis de Janeiro/1996 até Junho/2006 por faixa etária.

Fonte: Autores.

A relação de coocorrência entre duas palavras-chave é determinada pelo número de artigos em uma base de documentos em que ambas ocorrem conjuntamente, seja no título, no resumo ou na lista de palavras-chave (Van Eck & Waltman, 2014). As redes formadas pela relação entre essas palavras permite mapear possíveis temas de pesquisas sobre Previsões Hierárquicas. O tamanho de cada nó é proporcional à sua incidência nos artigos publicados e a localização das palavras se relaciona sua a relação com as demais. Através da escala de cores, também é possível visualizar como tem sido a evolução do tema com o passar dos anos (Figura 4).

Figura 4. Rede de co-ocorrência das palavras-chave em publicações sobre Previsões Hierárquicas.



Fonte: Autores.

Analisando a rede de coocorrência das palavras-chave nas 119 publicações levantadas, é possível perceber com o *cluster* roxo no canto inferior direito que as publicações mais antigas retratadas na rede utilizavam palavras-chave relacionadas aos conceitos fundamentais das séries temporais como “*bottom-up*” e “*disaggregation*” e, relacionado a isto está também a palavra “*prediction*”. Com isso, é possível identificar que os artigos mais antigos nesta escala se tratavam da apresentação inicial do tema e da sua relevância para a realização de previsões acuradas.

O mesmo comportamento pode ser identificado nos primeiros anos subsequentes conforme podemos ver no *cluster* azul e nos em tons de verde nos quais também são recorrentes palavras referentes aos conceitos fundamentais como “*top-down*”, “*aggregation*” e “*forecasting*”. Já nos nós verdes e, principalmente no *cluster* amarelo, ou seja, nas publicações mais recentes, são encontradas palavras variadas que sugerem a aplicação do tema na construção de modelos para diferentes áreas do conhecimento.

CONSIDERAÇÕES FINAIS

O interesse da literatura internacional pelo estudo de Séries Temporais Hierárquicas, apesar de recente, é crescente. Esta observação ancora-se no fato de que existem várias aplicações nas quais as informações podem ser organizadas em uma estrutura hierárquica. Dessa forma, é possível agregar observações através de distintos critérios, como por exemplo, a localização geográfica, os diferentes tipos de fontes de energia, a tipologia de produtos e outras características específicas.

Apesar da evidente relevância teórica, destaca-se que o debate na literatura nacional sobre esse tema ainda é incipiente. Diante do exposto, este artigo realizou de modo pioneiro uma pesquisa bibliométrica voltada para todas as publicações sobre séries temporais hierárquicas,

durante o período de 1996 a 2020. Assim sendo, selecionou-se aquelas publicações indexadas junto à base da Web of Science, incluindo livros, artigos de periódico e artigos de conferência. A análise bibliométrica realizada foi de destacada importância no aprofundamento dos autores com relação à recorrência e relevância das publicações sobre séries temporais hierárquicas. Dessa forma, através do uso de ferramentas de *big data*, como o *text mining*, foi possível tornar o processo mais ágil, o que possibilitou um amplo levantamento de registros bibliográficos.

Com isso, os resultados obtidos indicaram que o tema das séries hierárquicas é recente, tendo sua primeira pesquisa sido publicada apenas no ano de 1996. Apesar desse fato, percebe-se que grande parte das pesquisas foi publicada entre os anos de 2017 a 2020, o que denota a relevância atual do assunto. Paralelamente, aproximadamente 64% das publicações sobre HTS foram direcionadas a artigos de revistas. Nas publicações encontradas, inicialmente eram tratados assuntos de natureza teórica sobre os principais conceitos e abordagens de séries temporais hierárquicas e, à medida que o tema avançou no tempo, as pesquisas se voltaram para a aplicação em diferentes cenários, visto que a necessidade de realizar previsões é cada vez mais recorrente.

De maneira complementar, este artigo verificou que o debate sobre o assunto das séries temporais hierárquicas é basicamente centralizado em poucas revistas, dentre elas, *International Journal of Forecasting* (n = 5), *European Journal of Operational Research* (n = 4) e *Energies* (n = 4). Especificamente, no tocante aos países líderes em publicações sobre o assunto, situam-se nas três primeiras colocações: Estados Unidos da América (n = 33); China (n = 16); e Austrália (n = 14), respectivamente. Destaca-se ainda que as pesquisas sobre HTS centraram-se em aplicações voltadas aos campos da Computação (n = 34), Economia (n = 27) e Engenharia (n = 25).

Esta pesquisa possui limitações. A primeira delas consiste no fato da busca de registros bibliográficos ter sido feita única e exclusivamente na base de dados da Web of Science. Portanto, a título de trabalhos futuros, recomenda-se a investigação de publicações em paralelo a outras bases, como a Scopus e IEEE Xplore. Por fim, mediante a construção de uma base unificada, com a devida exclusão de duplicatas, recomenda-se o desenvolvimento de uma revisão sistemática de literatura sobre esse assunto, explorando as relações de parceria entre os autores, as instituições de pesquisa, bem como na confecção de redes de citação bibliográfica. Isso poderá contribuir para a identificação de *gaps* de pesquisa nessa área.

REFERÊNCIAS

- Alkema, L., Chao, F., You, D., Pedersen, J., & Sawyer, C. C. (2014). National, regional, and global sex ratios of infant, child, and under-5 mortality and identification of countries with outlying ratios: a systematic assessment. *The Lancet Global Health*, 2(9), e521-e530. [https://doi.org/10.1016/S2214-109X\(14\)70280-3](https://doi.org/10.1016/S2214-109X(14)70280-3)
- Aria, M. & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of informetrics*, 11(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1), 60-74. <https://doi.org/10.1016/j.ejor.2017.02.046>
- Athanasopoulos, G., Gamakumara, P., Panagiotelis, A., Hyndman, R. J., & Affan, M. (2020). *Hierarchical forecasting*. In *Macroeconomic Forecasting in the Era of Big Data* (pp. 689-719). Springer, Cham. https://doi.org/10.1007/978-3-030-31150-6_21
- Bearak, J., Popinchalk, A., Alkema, L., & Sedgh, G. (2018). Global, regional, and subregional trends in unintended pregnancy and its outcomes from 1990 to 2014: estimates from a Bayesian hierarchical model. *The Lancet Global Health*, 6(4), e380-e389. [https://doi.org/10.1016/S2214-109X\(18\)30029-9](https://doi.org/10.1016/S2214-109X(18)30029-9)
- Berliner, L. M. (1996). *Hierarchical Bayesian time series models*. In *Maximum entropy and Bayesian*

- methods* (pp. 15-22). Springer, Dordrecht. https://doi.org/10.1007/978-94-011-5430-7_3
- Bojer, C. & Meldgaard, J. P. (2020). The M5: A Preview from Prior Competitions. *Foresight: The International Journal of Applied Forecasting*, (58), 17-23.
- Freitas, T., de. (2016). *Modelos e aplicações para séries temporais hierárquicas: abordagens de reconciliação ótima e proporções de previsão*. 91 p. Dissertação (Mestrado em Engenharia de Produção) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- Greeff, S. C., de, Dekkers, A. L., Teunis, P., Rahamat-Langendoen, J. C., Mooi, F. R., & Melker, H. E., de. (2009). Seasonal patterns in time series of pertussis. *Epidemiology & Infection*, 137(10), 1388-1395. <https://doi.org/10.1017/S0950268809002489>
- Faloutsos, C., Gasthaus, J., Januschowski, T., & Wang, Y. (2019, June). Classical and contemporary approaches to big time series forecasting. In *Proceedings of the 2019 International Conference on Management of Data* (pp. 2042-2047). <https://doi.org/10.1145/3299869.3314033>
- Fliedner, G. (1999). An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers & operations research*, 26(10-11), 1133-1149. [https://doi.org/10.1016/S0305-0548\(99\)00017-9](https://doi.org/10.1016/S0305-0548(99)00017-9)
- Fliedner, G. (2001). Hierarchical forecasting: issues and use guidelines. *Industrial Management & Data Systems*. <https://doi.org/10.1108/02635570110365952>
- Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational statistics & data analysis*, 55(9), 2579-2589. <https://doi.org/10.1016/j.csda.2011.03.006>
- Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Karmy, J. P. & Maldonado, S. (2019). Hierarchical time series forecasting via support vector regression in the European travel retail industry. *Expert Systems with Applications*, 137, 59-73. <https://doi.org/10.1016/j.eswa.2019.06.060>
- Lauretto, M., Nakano, F., Pereira, C. A. B., & Stern, J. M. (2008, November). Hierarchical Forecasting with Functional Trees. In *AIP Conference Proceedings* (Vol. 1073, No. 1, pp. 317-324). *American Institute of Physics*. <https://doi.org/10.1063/1.3039015>
- Lauretto, M. S., Nakano, F., Pereira, C. A. B., & Stern, J. M. (2009). Hierarchical forecasting with polynomial nets. In *New Advances in Intelligent Decision Technologies* (pp. 305-315). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-00909-9_30
- Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra-and inter-category promotional information. *European Journal of Operational Research*, 249(1), 245-257. <https://doi.org/10.1016/j.ejor.2015.08.029>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54-74. <https://doi.org/10.1016/j.ijforecast.2019.04.014>
- Nielsen, H. A. & Madsen, H. (2001). A generalization of some classical time series tools. *Computational Statistics & Data Analysis*, 37(1), 13-31. [https://doi.org/10.1016/S0167-9473\(00\)00061-X](https://doi.org/10.1016/S0167-9473(00)00061-X)
- Pinheiro, S. M. (2015). *Previsão Hierárquica Aplicada às Políticas Públicas de Transporte Rodoviário*. Monografia. 61 p. Trabalho de Conclusão de Curso (Especialização em Estatística) - Universidade Federal de Minas Gerais, Brasil.
- Sedgh, G., Bearak, J., Singh, S., Bankole, A., Popinchalk, A., Ganatra, B., ... & Alkema, L. (2016). Abortion incidence between 1990 and 2014: global, regional, and subregional levels and trends. *The Lancet*, 388(10041), 258-267. [https://doi.org/10.1016/S0140-6736\(16\)30380-4](https://doi.org/10.1016/S0140-6736(16)30380-4)
- Silveira G., T. & Azevedo C., M. (2020). Forecasting hierarchical time series in power generation. *Energies*, 13(14), 3722. <https://doi.org/10.3390/en13143722>
- Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, 36(1), 75-85. <https://doi.org/10.1016/j.ijforecast.2019.03.017>
- Sun, Y., Zhang, X., & Wang, S. (2020). A hierarchical forecasting model for China's foreign trade. *Journal of Systems Science and Complexity*, 33(3), 743-759. <https://doi.org/10.1007/s11424-020-8070-y>
- Van Eck, N. J. & Waltman, L. (2014). Visualizing bibliometric networks. In *Measuring scholarly impact* (pp. 285-320). Springer, Cham. https://doi.org/10.1007/978-3-319-10377-8_13
- Veríssimo, F. V. (2016). *Modelo de Séries Temporais Hierárquicas de Previsão de Vendas Aplicado à Indústria do Calçado*. 79 p. Dissertação (Mestrado em Modelação, Análise de Dados e Sistemas de Apoio à Decisão) - Faculdade de Economia do Porto, Portugal.