

## Previsões estatísticas com base em séries temporais da cultura da laranja para o Brasil

### Statistical forecasts based on orange crop time series for Brazil

DOI:10.34117/bjdv7n4-670

Recebimento dos originais: 04/02/2021

Aceitação para publicação: 01/03/2021

**Camila Cielo**

E-mail: camila\_cielo@outlook.com

**Carla Adriana Pizarro Schmidt**

E-mail: carlaschmidt@utfpr.edu.br

**Flavio Trojan**

E-mail: trojan@utfpr.edu.br

**Caroline Cielo**

E-mail: caroline.cielo@hotmail.com

#### RESUMO

O Brasil é um dos maiores produtores de laranja do mundo, seu cultivo tem importância econômica no mercado de importação e exportação uma vez que necessita de um clima específico para produção em abundância. Nesse contexto, a pesquisa teve como objetivo realizar previsões do cenário agrícola da laranja no o Brasil, utilizando métodos estatísticos. Inicialmente, elaborou-se um panorama da citricultura no Brasil, com base na análise de componentes principais e agrupamento, realizado com o *software* Tanagra, com isso, observou-se que o estado de São Paulo tem a maior produção. A primeira componente foi capaz de explicar 70,02 % da variabilidade dos, a segunda componente explicou 20,43 % da variabilidade dos dados. Para as previsões, utilizou-se com *software* NNQ-estatística, em que realizou-se uma análise descritiva dos dados e por fim a previsão da área plantada, área colhida, da produção e do rendimento da laranja para o ano de 2019 possibilitando confrontar o resultado. Os modelos o método escolhido pelo software para área plantada e para o rendimento foi o MNM. Para área colhida e produção o método escolhido foi o MNA. As previsões encontradas para todas as variáveis foram satisfatórias, pois estavam dentro dos limites superiores e inferiores da previsão.

**Palavras-chave:** Séries Temporais. Modelos de Previsão. Laranja.

#### ABSTRACT

Brazil is one of the largest orange producers in the world, its cultivation has economic importance in the import and export market since it needs a specific climate for abundant production. In this context, the research aimed to make forecasts of the agricultural scenario of oranges in Brazil, using statistical methods. Initially, a panorama of citriculture in Brazil was elaborated, based on the analysis of main components and grouping, carried out with the Tanagra software, with that, it was observed that the state of São Paulo has the highest production. The first component was able to explain 70.02%

of the variability of, the second component explained 20.43% of the variability of the data. For the forecasts, it was used with NNQ-statistical software, in which a descriptive analysis of the data was carried out and, finally, the forecast of the planted area, harvested area, orange production and yield for the year 2019, making it possible to confront the result. The models the method chosen by the software for planted area and yield was MNM. For harvested area and production the method chosen was MNA. The predictions found for all variables were satisfactory, as they were within the upper and lower limits of the forecast.

**Key words:** Time Series. Forecasting Models. Orange.

## 1 INTRODUÇÃO

O Brasil encontra-se entre os maiores produtores de citros, com destaque para a produção de laranja. Um elevado percentual da fruta produzida no país é destinado para a indústria, que faz com que o país seja o maior fornecedor de suco de laranja do mundo, com 79% do mercado mundial e com mais de 90% da produção brasileira exportada (VITAL, 2018).

A citricultura é um dos mais tradicionais setores do agronegócio brasileiro, que passa por importantes transformações estruturais. Desde o início dos anos 2000, a cadeia da laranja tem sido desafiada por uma série de grandes mudanças que acontecem dentro e fora de seus limites, que afetam diretamente a vida de seus integrantes (NEVES, TROMBIN, 2017).

O Brasil está com uma projeção de aumento, de acordo com a USDA (2019) de 26% para a safra de 2018/2019, possibilitado pelo clima favorável que resultou em um excelente florescimento e frutificação.

Diante da importância que a produção de laranja tem para o Brasil e para o mercado mundial, na comercialização da laranja *in natura* bem como para seus processados, torna-se imprescindível conhecer e analisar seu cenário futuro, sendo que para isso pode-se adotar ferramentas que auxiliem essa análise e para ampliar o conhecimento sobre a situação atual e futura dessa cultura.

Um das ferramentas para aumentar a assertividade das decisões conseguindo prever certos acontecimentos, tendências e sazonalidades é a previsão estatística. De acordo com o Tubino (2009) as previsões possuem uma função muito importante nos processos de planejamento dos sistemas de produção, permitindo que seus gestores antevêm o futuro e planejem adequadamente suas ações, para isso é preciso adotar ferramentas que auxiliam essa análise e ampliam o conhecimento.

Krajewski, Ritzman e Malhotra (2009) explicam que as ferramentas utilizadas se baseiam em modelos matemáticos de previsão e são deveras importantes, pois têm papel fundamental como guias para o planejamento estratégico da produção, finanças e vendas de um produto.

Após mais de 50 anos de uso generalizado, a suavização exponencial ainda é um dos métodos estatísticos de previsão mais relevantes dentre os atualmente disponíveis (GOODWIN, 2010). Assim, os modelos de suavização exponencial são utilizáveis e altamente relevantes na prática, possuem uma base teórica sólida, o que torna qualquer tentativa de melhorar a precisão das previsões realizadas por esse método bastante trabalhosa, exigindo um esforço bem considerável (BERGMEIR, HYNDMAN, BENÍTEZ, 2016).

Diante do exposto, esse trabalho teve como finalidade realizar previsões estatísticas de algumas das variáveis referentes ao cultivo da laranja no Brasil, utilizando modelos que podem ser aplicáveis às séries temporais.

## 2 METODOLOGIA

Esta pesquisa pode ser classificada como básica, pois de acordo com a definição de Gil (2010), reúne estudos com o propósito de preencher uma lacuna de conhecimentos, buscando resolver problemas globais.

Quanto à abordagem, caracteriza-se como pesquisa quantitativa, pois se refere àquilo que pode ser quantificável, ou seja, as opiniões e informações sobre determinado evento ou fenômeno são traduzidas em números, para que assim possam ser classificadas e analisadas. Para tanto, faz-se necessário a utilização de recursos e de técnicas estatísticas, tais como: porcentagem, média, moda, mediana, desvio-padrão, coeficiente de correlação (KAUARK, MANHÃES, MEDEIROS, 2010).

O método de pesquisa utilizado pode ser classificado como documental, uma vez que se utilizou de uma coleta de dados restrita a documentos, escritos ou não, a partir de fontes primárias de informação, como é o caso das fontes estatísticas de informação governamentais utilizadas nesse estudo (MARCONI, LAKATOS 2008).

A princípio foram coletados os dados históricos da área plantada, área colhida, da produção e do rendimento médio do cultivo da laranja entre os anos de 2014 e 2018, diretamente da base de dados do Instituto Brasileiro de Geografia e Estatística (IBGE, 2019) para todo o Brasil. Primeiramente elaborou-se um panorama da citricultura no Brasil, para tanto utilizou-se de análise de componentes principais e agrupamento,

realizado com apoio do *software* Tanagra, utilizado como ferramenta para a mineração de dados.

Após a coleta desses dados, eles foram agrupados por meio do Software Microsoft® Excel. Os índices sazonais foram obtidos por decomposição dos dados e os diversos modelos de suavização exponencial possíveis foram gerados e testados por meio do software NNQ, um add in do Microsoft® Excel.

O NNQ-Estatística na qual as siglas representam Núcleo de Normalização e Qualimetria tem como objetivo desenvolver aplicações abordando controle estatístico de qualidade, estatística industrial e de previsão. O Software NNQ é vinculado e localizado no Departamento de Engenharia de Produção e Sistemas da Universidade Federal de Santa Catarina-UFSC, (NNQ, 2019).

Na sequência realizou-se uma análise descritiva dos dados e os resultados estatísticos de previsão foram apresentados, juntamente com uma análise das séries temporais, realizada por meio das técnicas de modelagem estatística de suavização exponencial. Devido à sua simplicidade e transparência, bem como a sua capacidade de se adaptar a muitas situações diferentes os modelos de suavização exponencial são muito utilizados (BERGMEIR, HYNDMAN, BENÍTEZ, 2016).

Os modelos de suavização exponencial se caracterizam por decompor uma série temporal em componentes, suavizar seus valores passados e depois recompor as componentes para fazer as previsões. Os métodos tratados pelo software são: ANA, MNA, AAA, MAA, AAdA, MAdA, MNM, MAM, MAdM, MMM, MMdM, nos quais a primeira letra de cada sigla significa o tipo de correção de erros, podendo ser aditiva ou multiplicativa, a segunda letra corresponde a presença de tendência (aditiva, aditiva amortecida, multiplicativa e multiplicativa amortecida) e a última letra condiz com o tipo de sazonalidade, que pode ser aditiva ou multiplicativa.

Com a utilização do NNQ-Estatística foi possível propor um método de suavização exponencial que mais se aproximou do ideal para realizar previsões futura para os conjuntos de dados, sendo que neste caso, utilizou-se o valor de Akaike (AIC) para minimizar o erro amostral, conforme indicado pelo próprio *software*.

Com o intuito de validar o modelo escolhido ao conjunto de dados, comparou-se as previsões realizadas para o ano de 2019 com os dados reais já disponíveis para este ano.

### 3 RESULTADOS E DISCUSSÃO

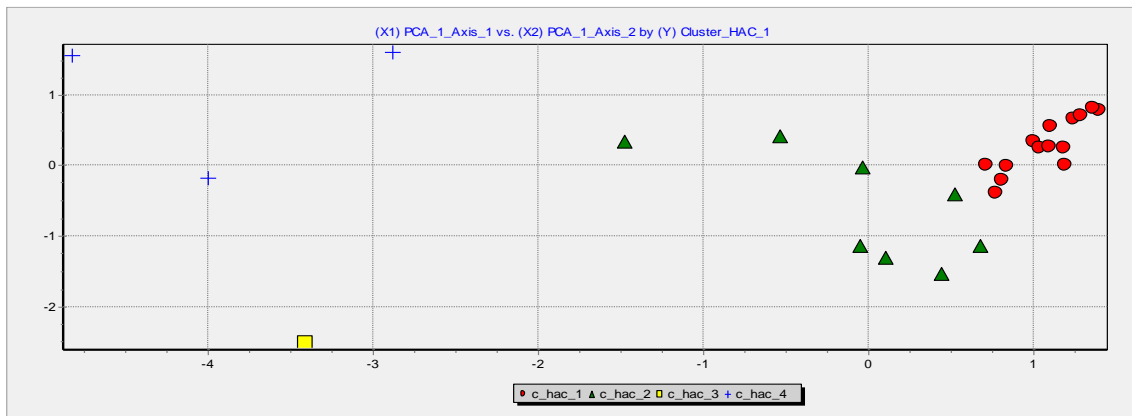
Inicialmente notou-se que a laranja é produzida em todos os estados do Brasil, independente do clima, existe área plantada com laranjas ao longo de todo o território nacional. O estado de São Paulo se destacou dos demais e por conta dos valores das variáveis encontradas para o estado de São Paulo terem sido *Outliers* devido a sua área destinada ao cultivo e produção de laranjas ser muito superior aos demais estados, este foi mantido fora do agrupamento. Notou-se que quatro grupos puderam ser identificados sobre um gráfico de componentes principais (Figuras 1A e 1B).

A primeira componente foi capaz de explicar 70,02 % da variabilidade dos dados e se correlacionou mais com as variáveis produções: (-0,97), área colhida (-0,96) e plantada (0,95), enquanto que a segunda componente explicou 20,43 % da variabilidade dos dados e se correlacionou mais apenas com a produtividade das plantas (-0,82), todos os valores de correlação calculados foram negativos.

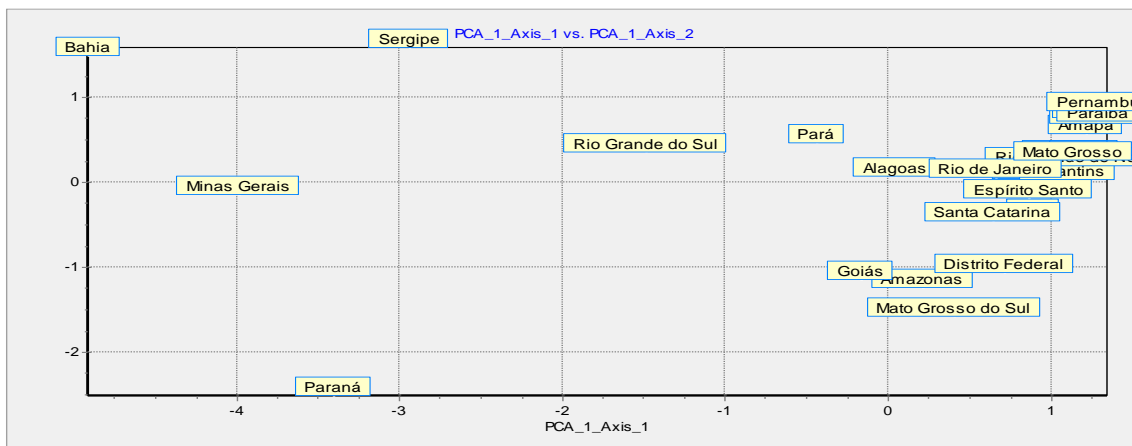
Por mais que o Paraná tenha área destinada ao cultivo próxima a dos outros estados, sua produtividade foi mais elevada, o que pode ter sido o fator determinante para que este não fosse agrupado a nenhum outro ficando isolado em um grupo, denominado grupo 1 (Figura 1a).

Nos outros grupos ficaram agrupados respectivamente três, oito e quatorze estados. A área de cultivo foi determinante no agrupamento dos três estados (Grupo 2 – Bahia, Sergipe e Minas Gerais). A produtividade foi semelhante e principal responsável pelo agrupamento dos oito estados agrupado no grupo 3 (Amazonas, Pará, Alagoas, Santa Catarina, Rio Grande do Sul, Mato Grosso do Sul, Goiás e Distrito Federal) enquanto que a produção dos demais estados foi a principal responsável pela formação do grupo 4.

Figura 1 – Gráficos ilustrativos da dispersão dos estados obtidos por meio de análise de Componentes Principais. (A) Gráfico de agrupamento dos estados. (B) Identificação dos estados no agrupamento.



(A)



(B)

Com os dados históricos coletados realizou-se uma análise da estatística descritiva para a área plantada e colhida, para a produção e para o rendimento de produção de laranja no Brasil (Tabela 1).

Tabela 1 – Estatística descritiva da série anual de dados de 2014 até 2018 da área plantada e colhida, da produção e rendimento da laranja.

	Área Plantada	Área Colhida	Produção	Rendimento
Mínimo	666173	605752	13742120	22048
Média	724238	661955	16106134	24395
Mediana	722338	666524	16257008	23227
Máximo	791140	721180	18666928	29922
Desvio Padrão da Média	4039	4437	152855	292
Desvio Padrão	31290	34365	1184010	2264
Coeficiente de Variação	4,3%	5,2%	7,4%	9,3%

Fonte: Autoria própria (2019).

As medidas de tendência central, média e mediana conseguem resumir em um único valor o que ocorre tipicamente com a série de dados temporais. Quando essas medidas apresentaram diferenças, mostra-se que a série de dados contém valores que fogem da tendência central, ou seja, que não representam os dados como um todo. Neste caso, pode-se dizer que as medidas são próximas.

O desvio padrão é uma medida de dispersão que indica a variabilidade que os dados apresentam entre si. Se todos os dados são iguais, não há dispersão. Para valores próximos uns dos outros, temos uma pequena dispersão. E se os dados são muito diferentes entre si, a dispersão é grande. Neste caso, temos um desvio padrão que pode ser considerado alto, isso ocorre por se tratarem de valores muito grandes, entretanto o coeficiente de variação foi baixo e nesse caso percebe-se que não se nota grande dispersão entre os dados mensais coletados para os cinco anos estudados.

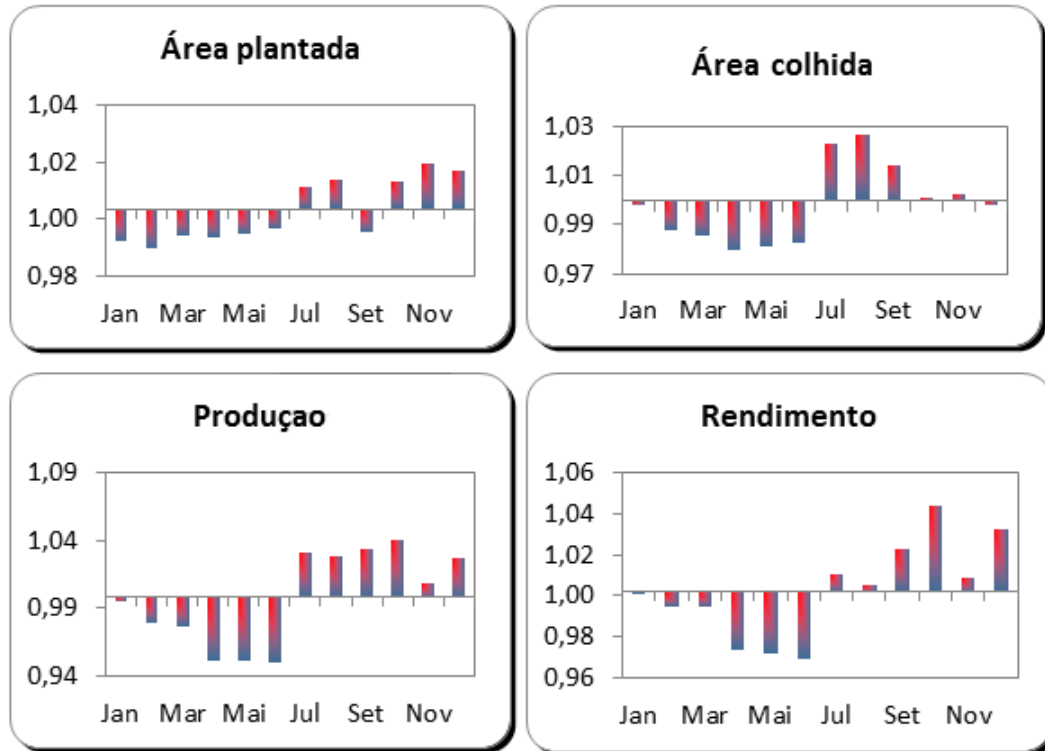
A diferença entre o menor valor da amostra (mínimo) e o maior valor da mostra (máximo) ilustra que os valores de área plantada, colhida, produção e rendimento foram variáveis ao longo dos anos acompanhados.

Na sequência realizou-se as previsões futuras para as variáveis: área plantada, área colhida, produção e rendimento da laranja no Brasil, para o ano de 2019, com base em suas séries temporais históricas.

Para escolha do melhor método de suavização exponencial foram aplicadas algumas análises de decomposição nos dados a fim de identificar as variações cíclicas que a série sofre em um período de tempo. Através da Figura 1 pode-se visualizar a decomposição dos dados em índices sazonais.

Com a decomposição dos índices sazonais é possível perceber que de janeiro a junho tanto a área plantada, a área colhida, produção e o rendimento apresentam baixas, já de julho a dezembro exibem um aumento significativo.

Figura 1 – Decomposição do conjunto de dados da laranja em índices sazonais anual.



Fonte: Autoria própria (2019).

A seguir os dados foram analisados por meio dos diversos modelos possíveis de suavização exponencial, com auxílio do software NNQ. O software avalia os modelos ajustados a partir de quatro tipos de erro, o R1 (autocorrelação), o U de Theil e o Akaike (AIC). Todos os valores de U de Theil dos modelos apresentados na Tabela 2 foram inferiores a 1,0 o que significa que todos os modelos calculados apresentam previsões melhores, que um modelo ingênuo de previsão, os valores dos quatro tipos de erros foram pequenos.

Na Tabela 2 encontram-se os valores de R1 e do Akaike (AIC) para cada uma das variáveis estudadas. Utilizou-se o menor valor do critério de Akaike (AIC) para escolha e indicação do melhor modelo.

Observando-se os resultados apresentados na Tabela 2 foi possível verificar que os coeficientes de autocorrelação foram baixos e levando-se em conta o menor valor do AIC o método escolhido pelo *software* para área plantada e para o rendimento foi o MNM. Para área colhida e produção o método escolhido foi o MNA.

O modelo MNM realiza a correção de erros de forma multiplicativa, trata a sazonalidade também de forma multiplicativa e não faz nenhum tratamento para a tendência. O modelo MNA realiza a correção do erro de forma multiplicativa, trata a sazonalidade de forma aditiva e não identifica nenhuma tendência para tratar.



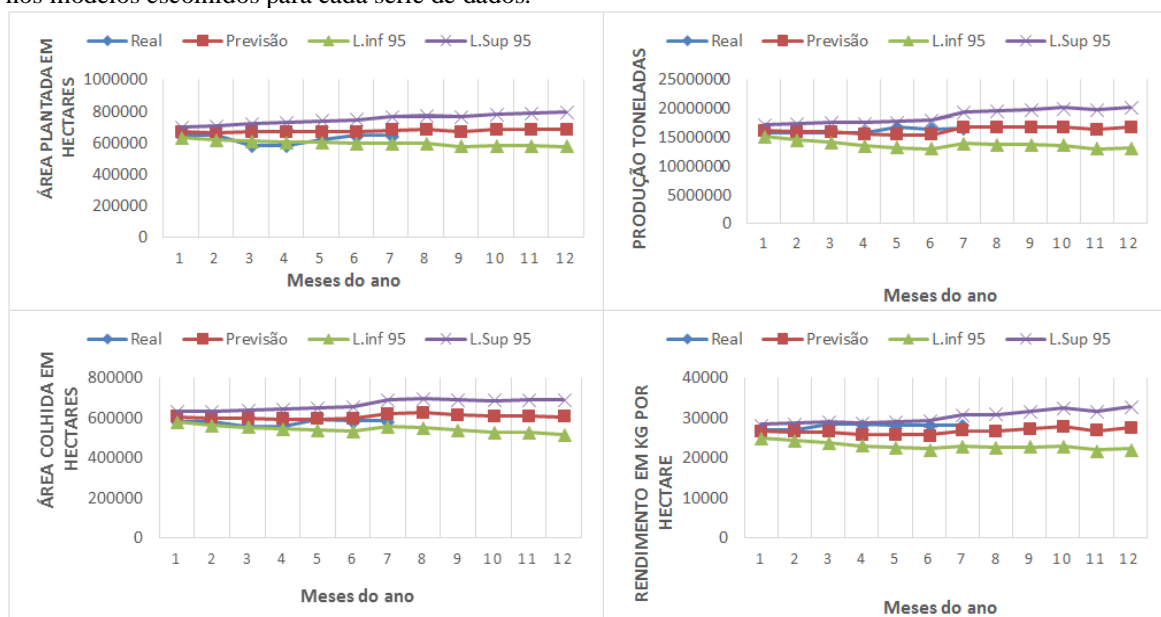
A Tabela 3 apresenta em maiores detalhes esses modelos e os valores de *alfa*, *beta*, *gama* e *fi* calculados pelo NNQ. Percebeu-se que não foi necessário o uso das constantes de tendência nem de amortecimento para nenhum dos modelos escolhidos e que os maiores valores encontrados foram para correção de nível e os menores para a sazonalidade sendo que se pode notar que o que mais influenciou na variação dos valores dessas variáveis estudadas referentes ao cultivo da laranja no Brasil foi o nível. O nível é o valor médio da observação no período estudado (valor observado retirando a sazonalidade se houver, e o erro aleatório).

Tabela 2 – Valores calculados para o coeficiente de autocorrelação e critério de Akaike.

Métodos	Área Plantada		Área Colhida		Produção		Rendimento	
	r1	AIC	r1	AIC	r1	AIC	r1	AIC
ANA	0,000	1451,76	-0,009	1426,22	0,001	1858,25	-0,027	1075,40
MNA	-0,029	1450,10	-0,014	1423,87	0,003	1855,25	-0,016	1071,82
AAA	0,005	1456,06	0,006	1430,02	0,001	1862,80	-0,012	1079,06
MAA	-0,024	1454,49	0,001	1427,89	0,004	1859,54	0,001	1074,83
AAdA	0,001	1457,85	0,004	1432,09	0,001	1864,13	-0,012	1080,57
MAdA	-0,028	1456,22	-0,011	1430,00	0,004	1861,09	0,000	1076,97
MNM	-0,031	1450,07	-0,015	1424,84	0,002	1858,31	-0,019	1070,98
MAM	-0,026	1454,45	0,000	1428,87	0,002	1862,65	-0,005	1074,10
MAdM	-0,030	1456,19	-0,012	1430,97	0,003	1864,16	0,000	1075,83
MMM	-0,024	1454,42	0,000	1428,75	0,002	1862,76	-0,005	1074,36
MMdM	-0,038	1455,93	-0,002	1431,01	0,012	1862,44	0,012	1074,43

Fonte: Autoria própria (2019).

Figura 2 – Valores previstos e limites de confiança para a previsão ao longo do ano de 2019, com base nos modelos escolhidos para cada série de dados.



Fonte: Autoria própria (2019).

Tabela 3 – Melhores modelos escolhidos para as previsões e valores calculados pelo NNQ para os coeficientes dos modelos.

Variáveis	Modelos	Coeficientes calculados para os modelos			
		Alfa Nível	Beta Tendência	Gama Sazonalidade	Fi Amortecimento
Área Plantada	MNM	0,92	0,00	0,01	0,00
Área Colhida	MNA	0,95	0,00	0,01	0,00
Produção	MNA	0,97	0,00	0,01	0,00
Rendimento	MNM	0,87	0,00	0,01	0,00

Fonte: Autoria própria (2019).

Percebeu-se através dos resultados obtidos, a proximidade dos valores previstos com os valores reais da área colhida da produção e do rendimento da laranja para o ano de 2019, nota-se também que em nenhum momento a demanda ultrapassou os limites inferiores e superiores confirmando assim a eficácia dos modelos de previsão escolhidos para prever o comportamento dos dados.

Porém notou-se um comportamento inesperado com os valores de área colhida e principalmente da área plantada real da laranja nos meses de março e abril de 2019, esse comportamento pode ser explicado através da análise do cenário, na qual por meio do IBGE (2019) foi possível observar que o estado de Minas Gerais não obteve coleta de dados nos meses de janeiro a março de 2019, voltando a coletar os dados apenas no mês de maio, este acontecimento pode ser oriundo a tragédia do rompimento da barragem que ocorreu no estado no início do ano de 2019.

Diante deste cenário, a quantidade de área plantada e também colhida sofreu uma queda inesperada, que não pode ser prevista de forma perfeita pelos dados históricos quantitativos, fazendo com que os dados reais ficassem muito próximos e até abaixo da linha de mínimo prevista pelos modelos.

#### 4 CONCLUSÃO

Estudos quantitativos são muito importantes com vista a realização de previsões futuras com base em dados passados, mas a avaliação do contexto político e econômico não pode ser ignorado nas previsões. Todos os estados brasileiros cultivam laranjas em maior ou menor quantidade, o cultivo se estende por todo o território nacional, sendo São Paulo o que apresenta maior área destinada ao cultivo e produção, enquanto o Paraná se destaca em produtividade sendo esta superior inclusive à produtividade de São Paulo.

Com os resultados obtidos foi possível estabelecer o melhor modelo de suavização exponencial para cada caso. O software NNQ mostrou-se capaz de auxiliar na realização

das previsões pelos diversos métodos, bem como na indicação dos melhores modelos com base nos valores dos erros, U de Theil e Akaike que indicou como um tratamento estatístico eficiente dos dados brutos e os modelos MNM e MNA como os melhores modelos de suavização exponencial para essa análise e previsão.

Os modelos utilizados para as previsões estatísticas mostraram-se eficazes, pois obteve-se resultados muito próximos aos limites esperados e quando desvios foram observados a justificativa foi clara pois a ausência de dados de um estado que dedica uma área significativa ao cultivo da laranja é um problema sério que pode afetar a qualidade da previsão. A escolha do melhor modelo de previsão para cada caso, foi capaz de permitir uma análise futura da área plantada, área colhida, produção e rendimento médio dessa cultura com reduzidos valores de erros de previsão que variaram entre 1,36 % até 2,25 %.

Espera-se que os resultados encontrados nas previsões possam servir de auxílio para futuras decisões e estratégias do produtor, da agroindústria, do consumidor e principalmente para o governo, uma vez que a laranja é um cultivo de impacto no mercado Brasileiro tanto nas exportações quanto nas importações.

## REFERÊNCIAS

BERGMEIR, Christoph; HYNDMAN, Rob j.; BENÍTEZ, José m.. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. **International Journal Of Forecasting**, [s.l.], v. 32, n. 2, p.303-312, abr. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.ijforecast.2015.07.002>.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010. 184 p.

GOODWIN, P. The Holt–Winters approach to exponential smoothing: 50 years old and going Strong Foresight: **The International Journal of Applied Forecasting**, 19, pp. 30-33. 2010.

IBGE. **Produção Agrícola Municipal**: Produção Agrícola Municipal. 2019. Disponível em: <<https://sidra.ibge.gov.br/tabela/1618>>. Acesso em: 15 ago. 2019

KAUARK, Fabiana da Silva; MANHÃES, Fernanda Castro; MEDEIROS, Carlos Henrique. **Metodologia da pesquisa**: Um guia prático. Itabuna / Bahia: Litterarum, 2010. 89 p. Disponível em: <[http://197.249.65.74:8080/biblioteca/bitstream/123456789/713/1/Metodologia da Pesquisa.pdf](http://197.249.65.74:8080/biblioteca/bitstream/123456789/713/1/Metodologia_da_Pesquisa.pdf)>. Acesso em: 16 maio 2018.

KRAJEWSKI, Lee J.; RITSMAN Larry P.; MALHOTRA Manoj K. **Administração de Produção e Operações**. São Paulo, 2009.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Técnicas de pesquisa**. 7. ed. São Paulo: Atlas, 2008. 277 p.

NEVES, Marcos Fava; TROMBIN, Vinicius Gustavo. **Anuário da Citricultura 2017**. 2017. Disponível em: <[http://www.citrusbr.com/download/biblioteca/CitrusBR\\_Anuario\\_2017\\_alta.pdf](http://www.citrusbr.com/download/biblioteca/CitrusBR_Anuario_2017_alta.pdf)>. Acesso em: 25 ago. 2019

NNQ- Núcleo de Normalização e Qualimetria -. **Previsão**. 2019. Disponível em: <<https://qualimetria.ufsc.br/publicacoes/software/previsao/>>. Acesso em: 22 ago. 2019.

TUBINO, Dalvio Ferrari. **Planejamento e controle da Produção**. 2. ed. São Paulo: Atlas, 2009. 190 p.

USDA – United States Department Of Agriculture. **Citrus: World Markets and Trade**. 2019. Disponível em: <<https://apps.fas.usda.gov/psdonline/circulars/citrus.pdf>>. Acesso em: 17 ago. 2019

VIDAL, Maria de Fatima. **Citricultura na Área de Atuação do BNB**. 2018. Disponível em: <[https://www.bnb.gov.br/documents/80223/4049480/41\\_Citrus\\_2018.pdf/11e22002-](https://www.bnb.gov.br/documents/80223/4049480/41_Citrus_2018.pdf/11e22002-)