



Produto & Produção, vol. 17 n. 4, p. 80-93, dez. 2016

RECEBIDO EM 15/07/2016. ACEITO EM 11/10/2016.

Murilo Gattás Arthus

Universidade Estadual de Campinas – Unicamp
mugarthus@gmail.com

Juliane Andressa Camatti

Universidade Estadual de Campinas – Unicamp
juli_camatti@hotmail.com

Wagner Bortoletto

Universidade Estadual de Campinas – Unicamp
wagner.bortoletto@gmail.com

Orestes Ferrari Junior

Universidade Estadual de Campinas – Unicamp
orestesfj@gmail.com

Paulo Sérgio de Arruda Ignácio

Universidade Estadual de Campinas – Unicamp
paulo.ignacio@fca.unicamp.br

Planejamento da safra de soja no Oeste do Paraná

RESUMO

A soja é o grão mais produzido no Brasil e tem dado ao país o título de maior exportador desta commodity. A pouca quantidade de produto disponível internamente e a alta do dólar tem elevado o seu preço a valores recorde, com desdobramentos na área de cultivo, que vem aumentando com o passar dos anos. Este artigo tem o objetivo de aplicar modelos de previsão de demanda para o cenário de rendimento de soja no Oeste do Paraná e discutir os seus resultados quanto aos erros em relação aos valores reais. Dados referentes ao rendimento do grão por hectare no Oeste do Paraná, entre os anos de 1980 e 2014, foram coletados e submetidos a quatro testes de hipóteses para determinar os modelos de previsão mais adequados para comparação. Os resultados obtidos validaram a modelagem, que pode servir de base para o planejamento dos produtores rurais em relação às estimativas de rendimento de soja no curto prazo.

Palavras- chave: Soja, previsão de demanda, séries temporais.

ABSTRACT

Soybean is the most produced grain in Brazil and has given the country the title of largest exporter of this commodity. The small amount of internally available product and the Dollar high exchange rates have raised its the price to record levels, with extensions in the harvesting area, which has been increasing over the years. This article aims to apply forecasting methods for soybean yield scenario in western Parana and to discuss the results of the errors against the original values. Data on the grain hectare yield in the west of Parana were collected, between the years of 1980 and 2014, and four tests of hypotheses were applied to point out the most appropriate forecasting methods for comparison. The results validated the model, which may become the basis for the farmers' crop planning regarding soybean yield in the short-term period.

Keywords: Soybean, forecasting methods, time series.

1. Introdução

De acordo com o relatório “Indicadores IBGE” divulgado em Março/2016, o PIB brasileiro em 2015 foi 3,8% menor que o ano anterior. Na contramão deste movimento, a agropecuária cresceu 1,8% no mesmo período, principalmente por conta do desempenho da agricultura, com destaques para o crescimento das produções de soja (11,9%) e milho (7,3%). Atualmente, o PIB agropecuário corresponde a aproximadamente 21,5% do valor adicionado ao PIB total no ano. Dentro da agropecuária, a agricultura representa 68% do valor do PIB do agronegócio enquanto a pecuária fica com os outros 32% (CEPEA/USP, 2016). Assim, a cadeia de suprimentos voltada para a agricultura foi responsável diretamente por 15% do PIB nacional em 2015.

Segundo o CONAB, 56% da produção de soja de 2015 foram destinadas ao mercado externo. Considerando a demanda mundial estimada em 300 milhões de toneladas/ano para 2015, as produções brasileiras de soja representaram quase um terço do que foi consumido no mundo, sendo que 75% da soja exportada pelo Brasil foram absorvidas pelo mercado chinês (CONAB, 2016).

Em termos de área, o Brasil apresenta cerca de 40% de seu território disponível para a agropecuária (IBGE, 2016 [2]). Considerando apenas o território utilizado por lavouras, o plantio de soja representa 41% de toda a área utilizada na agricultura, sendo que os principais estados produtores de soja atualmente são Mato Grosso, Paraná e Rio Grande do Sul (IBGE, 2016 [3]).

Nos últimos dez anos, o rendimento médio brasileiro das plantações de soja, em toneladas por hectare, aumentou 27% (IBGE, 2016 [3]). Comparando-se os rendimentos das unidades federativas brasileiras, quem se destaca é o Paraná, cujo rendimento médio é o maior dentre os estados produtores. Em 2015, o estado produziu aproximadamente 35 milhões de toneladas de soja, ficando atrás apenas do Mato Grosso (IBGE, 2016 [1]).

A importância deste grão não está somente relacionada à fabricação de farelos para animais e óleo: atualmente, seu uso como biocombustível e sua fonte proteica para a dieta humana é notável (MASUDA, GOLDSMITH, 2009).

As previsões são muito relevantes e utilizadas na área econômica e financeira. Quanto melhor a previsão, melhores são as decisões tomadas com base nesses números (DIEBOLD, 1996). Com o aumento da população mundial, a demanda por alimentos segue a mesma tendência. Prever o que será produzido nos próximos anos, o rendimento das plantações, as mudanças climáticas que ocorrem e como elas impactam na agricultura, é necessário para o cumprimento da demanda no futuro (TILMAN *et al.*, 2011).

Estudos relativos à previsão de safras já foram feitos: tecnologias de sensoriamento remoto no Brasil (GUSSO, 2012) e Estados Unidos (JOHNSON, 2013) e aplicação dos modelos de previsão de Winters (PATINO *et al.*, 2013) e Box-Jenkins (MASUDA; GOLDSMITH, 2009).

A partir das conjunturas apresentadas, o objetivo deste artigo é aplicar um modelo de previsão para rendimento de soja (toneladas por hectare), a partir de dados de produção no Oeste do estado brasileiro do Paraná, e analisar os resultados.

A previsão de safras e o rendimento das plantações são importantes por vários motivos, que incluem aspectos tanto macroeconômicos, referentes ao planejamento de utilização de recursos, quanto informativos, referentes à expectativa de produção para o próprio agricultor (JOHNSON, 2013). É importante que o produtor tenha conhecimento de como a previsão de demanda é calculada, pois ele faz parte da cadeia de suprimentos e precisa ter confiança nos números que são divulgados, mesmo que venham de órgãos governamentais. Além disso, a previsão de safra é importante em virtude das variações entre os rendimentos de uma safra e outra. Essas variações podem resultar de diferenças nas áreas de produção (GOOD, IRWIN, 2006), estratégias de plantio, irrigação, fertilização e utilização de fitossanitários (JOHNSON, 2013).

2. Revisão da literatura

2.1 Previsão de demanda

A previsão de demanda é uma técnica coerente de investigação de informações referentes ao número de vendas futuras de um produto. Na medida do possível, fornece conhecimentos sobre qualidade e local onde serão essenciais os itens no futuro (MOREIRA, 2014). A Figura 1 apresenta a classificação mais comum em relação as abordagens dos métodos de previsão.

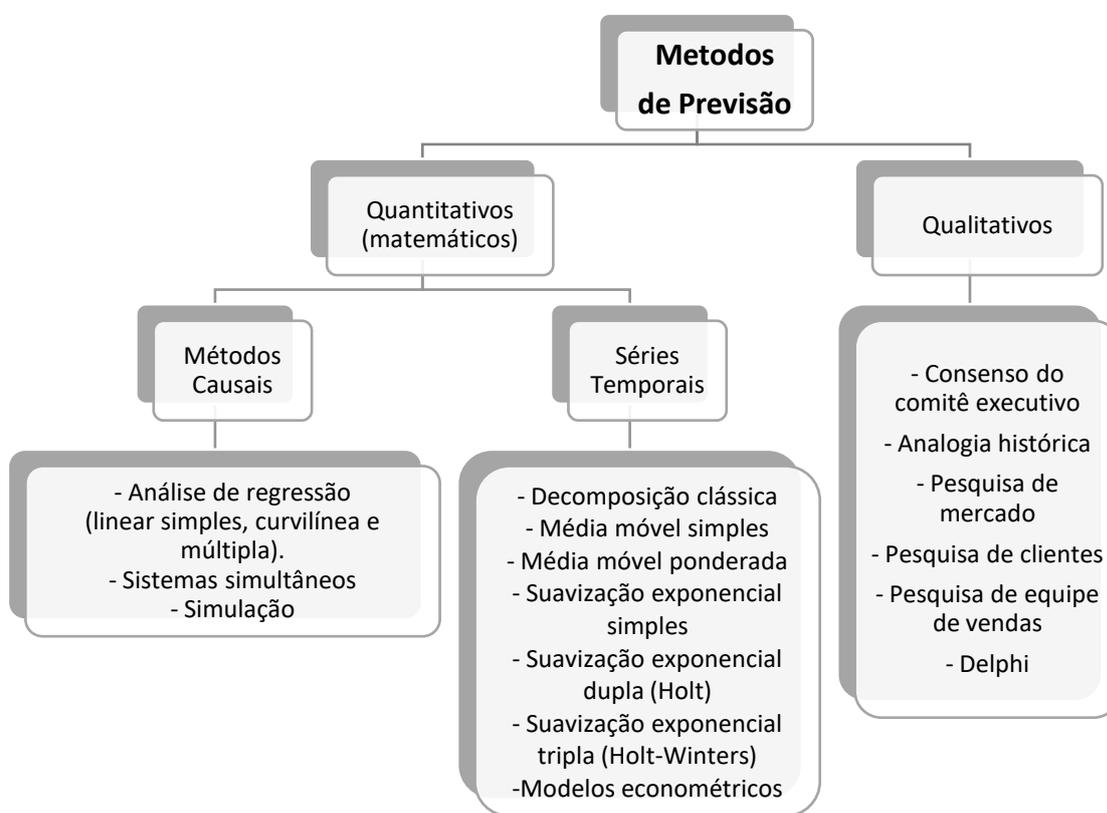


Figura 1- Classificação dos métodos de previsão.

Fonte: Adaptado de Moreira (2014), Fernandes e Filho (2010).

Os modelos de abordagem qualitativos têm natureza subjetiva sustentada principalmente no veredito do tomador de decisões para execução da previsão. É utilizada especialmente quando não há dados históricos acessíveis ou quando são de caráter qualitativo (FERNANDES; FILHO, 2010).

Segundo Moreira (2014), modelos quantitativos, também conhecidos por matemáticos, são aqueles que empregam modelagem matemática para se obter os valores futuros. Os métodos matemáticos se dividem em:

- ✓ Modelos causais: a busca de um elemento ou um conjunto de elementos é relativa a uma ou mais variáveis internas ou externas a organização, chamadas de variáveis causais. Se houver uma ótima estimativa de valor, é factível a obtenção da projeção almejada.

- ✓ Series temporais: a observação de séries temporais não requer nada além do entendimento dos valores precedentes da demanda. O estado de série temporal comunica unicamente um agrupamento de valores da demanda tomados em relação ao tempo. A probabilidade esperada é que o padrão visto nos valores passados propicie dados satisfatórios para a previsão futura dos valores.

2.2 Métodos quantitativos - séries temporais

A suposição inicial para previsão de demanda é de que a demanda futura pode ser presumida em relação aos valores passados. Assim a demanda futura decorre de uma análise dos dados passados no decorrer do tempo (MOREIRA, 2014). O método de série temporal utiliza dados históricos em relação a variável dependente. Por isso, este método é fundamentado na teoria de que o padrão antecedente da desta variável persistirá no futuro (KRAJEWSKI *et al.*, 2009).

De acordo com Moreira (2001), em razão a um período de tempo consideravelmente longo, o padrão da demanda pode ser classificado de acordo com quatro comportamentos ou efeitos relacionados à série temporal, disponíveis na Tabela 1.

Tabela 1 - Características de uma série temporal.

NÍVEL - O nível da demanda representa a quantidade de vendas da série temporal das demandas anteriores, não levando em conta a variação sazonal e aleatória. Se apresenta tanto no padrão estacionário como não estacionário no decorrer da interpretação da série temporal.
TENDÊNCIA – É a taxa de crescimento ou decrescimento entre os dados seguidos de uma série histórica.
SAZONALIDADE – É o parâmetro de variação que se repete ao longo do tempo, que pode ser previsto e interpretado. Se caracteriza por ser de padrão repetitivo e não aleatório. Os produtos que apresentam mais vendas em determinada época do ano como brinquedo, agasalho e sorvete, tendem a ser sazonais. Já os produtos de higiene, o arroz e carne são de consumo frequente e não tendem a sazonalidade.
ALEATORIEDADE – Correspondente ao número alto de fatores, a sua demanda denota elementos aleatórios, que por isso não consegue ser prevista por um modelo de previsão, mas permite realizar comparação de erro entre a previsão e a demanda passada.

Fonte: Adaptado de Peinado e Graeml (2007) e Chopra e Meindl (2011).

Corroborando a ideia de Moreira (2001), Enders (2008) afirma que uma série temporal pode ser decomposta nos componentes de tendência, sazonalidade, ciclicidade e componente irregular.

$$Z_t = T_t + S_t + I_t \text{ (Método Aditivo)}$$

$$Z_t = T_t * S_t * I_t \text{ (Método Multiplicativo)}$$

Onde:

T_t é a componente de tendência no período t ;

S_t é a componente sazonal no período t ;

I_t é a componente irregular no período t ;

Z_t é a previsão para o período t .

Sendo assim, para se realizar previsões com maior acuracidade, é necessário conhecer se a série temporal em estudo possui tais características.

Um aspecto importante em relação às séries temporais está relacionado com sua estacionariedade. Para ser estacionária, os dados devem crescer de forma contingente, em volta de média e variância constantes. Porém, a maior parte das séries é não estacionária, ou seja, tem variância não estável. Neste caso, os dados precisam ser trabalhados por diferenciações, até que a série passe a

ser estacionária (MORETTIN; TOLOI, 2006). A Tabela 2 traz um resumo sobre os métodos de previsão de demanda para séries temporais:

Tabela 2 - Alguns modelos de previsão de séries temporais.

Média Móvel – Aplicabilidade se dá em demanda sem tendência e sem sazonalidade. É usada para previsão a partir da média aritmética de n números seguidos de demanda real, anteriores à previsão.

Alisamento Exponencial Simples – Aplicabilidade em demanda sem tendência e sem sazonalidade. Esse método assume a previsão do período anterior e a complementa com um ajuste para ter a previsão do período seguinte.

Modelo de Holt – Aplicabilidade em demanda com tendência e sem sazonalidade. É pertinente sua utilização quando a demanda tem nível e tendência no componente sistemático. A demanda e o tempo tem relação linear.

Modelo de Winter – Aplicabilidade em demanda com tendência e sazonalidade. O modelo de Winter é utilizado em organizações cuja sua demanda apresenta variação no seu aspecto de nível, tendência e sazonalidade.

Modelo ARIMA – A forma mais geral do modelo não considera a sazonalidade. O modelo ARIMA ajusta os valores observados para que a diferença entre esses valores e os valores produzidos no modelo seja próximo de zero. Modelo avalia a autocorreção e a autocorrelação parcial entre os dados dentro dos valores críticos e distingue os padrões aleatórios (ruído branco).

Fonte: Adaptado de Moreira (2001); Pindyck e Rubinfeld (2004); Peinado e Graeml (2007); Sato (2013).

Os modelos citados possuem suas equações características e de acordo com Makridakis *et al.* (1998) e Montgomery *et al.* (2008) construiu-se o Quadro 1 a seguir.

Quadro 1- Equações dos modelos de previsão de demanda

Método de Previsão	Equações
Média Móvel Simples (MMS)	$MMS = Z_t = \frac{(Y_t + Y_{t-1} + \dots + Y_{t-n+1})}{n}$
Suavização Exponencial Métodos de Holt-Winters (HW)	$L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1})$ $b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$ $S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s}$ $HW = Z_{t+m} = L_t + b_t m + S_{t-s+m}$
Autoregressivos Integrados de Médias Móveis (ARIMA)	$AR = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p}$ $MA = c + \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ $I_1 = Y'_t = Y_t - Y_{t-1}$ $I_2 = Y''_t = Y'_t - Y'_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$ $ARIMA = Z_t = AR + I + MA$

Fonte: Adaptado de Makridakis *et al.* (1998); Montgomery *et al.* (2008).

Onde:

Y_t é o valor observado no período t ;

Z_t são os valores previstos para o período t ;

α, β e γ são os pesos e/ou coeficientes das expressões;

L_t é o componente do cálculo relativo ao nível da série;

b_t é o componente do cálculo relativo à tendência da série;

S_t é o componente do cálculo relativo à sazonalidade da série;

s refere-se ao tamanho do ciclo sazonal;

m é o período a frente a ser previsto;
 c é a constante ou ponto de intercepto;
 ϕ_p é o coeficiente do fator autoregressivo;
 θ_q é o coeficiente do fator de média-móvel;
 Y'_t são os componentes das diferenças tomadas da série;
 ε_t são os erros aleatórios;
 n é o número de observações.

Nos casos onde não se aplica a tendência e sazonalidade, ou seja, uma suavização simples, basta ignorar os componentes b_t e S_t do modelo de Winters. No caso do modelo de Holt, ignora-se apenas a componente b_t .

2.3 Erros de previsão

Um método de previsão, para ser considerado bom, precisa apresentar um erro estatístico semelhante à característica da imprevisibilidade da demanda (PEINADO; GRAEML, 2007). A análise do erro é usada para indicar o método de previsão e se este pode prever com precisão o elemento sistemático da demanda. O plano de contingência necessita declarar o erro de previsão. Estes erros precisam estar dentro das deduções históricas, para que a utilização do modelo prossiga (CHOPRA; MEINDL, 2011). No Quadro 2, há um sumário das medidas de erros de previsão, com suas equações correspondentes.

Quadro 2 – Equações dos erros de previsão.

Medidas de erro da previsão
<p>Erro quadrático médio (EQM) – está conectado com a variância do erro de previsão.</p> $EQM_n = \frac{1}{n} \sum_{t=1}^n E_t^2$
<p>Desvio absoluto – é o número absoluto do erro no tempo t.</p> $A_t = E_t $
<p>Desvio médio absoluto (DMA) – é média do desvio absoluto por todo o ciclo.</p> $DMA_n = \frac{1}{n} \sum_{n=t}^n A_t$
<p>Erro percentual absoluto médio (EPAM) – é o erro absoluto médio com uma porcentagem de demanda.</p> $EPAM_n = \frac{1}{n} \sum_{t=1}^n \left \frac{E_t}{Y_t} \right * 100$

Fonte: Adaptado de Chopra e Meindl (2011).

Onde:

$E_t = Y_t - Z_t$ – Erro de previsão para o período t ;

Z_t é a previsão para o período t ;

Y_t é a demanda real observada no período t ;

n é o número de observações.

3. Método

3.1. Classificação da Pesquisa

O estudo apresenta natureza aplicada, cujo propósito é promover conhecimento para aplicação prática, conduzindo a resolução de situações específicas. A abordagem é quantitativa, pois os resultados serão deduzidos em números, opiniões e informações (KAUARK *et al.* 2010). O objetivo da pesquisa se classifica como exploratório, pela familiaridade com o problema e pelo aperfeiçoamento de ideias em prol da tentativa de redigir uma hipótese (GIL, 2002).

Os procedimentos técnicos envolvem: pesquisa bibliográfica, cujos materiais gráficos e informatizados buscam esclarecer uma objeção ou contrair conhecimento sobre um assunto em específico; e levantamento de coleta de dados atrelado à análise do objeto de estudo em seu espaço (LEÃO, 2016).

3.2 Descrição do modelo

A coleta de dados deste artigo envolveu o levantamento de informações referentes à produção de soja por unidade de área na região Oeste do Paraná. Para tanto, os dados foram obtidos no banco de dados do IPARDES (Instituto Paranaense de Desenvolvimento Econômico e Social) e foram considerados os valores de rendimento médio, em tonelada/hectare, da produção anual de soja entre os anos de 1980 e 2014, configurando 35 pares de dados (tempo e rendimento).

Para escolher os métodos de previsão a serem comparados, foram feitos quatro testes de hipóteses na série de dados para determinar suas características, de acordo com a Figura 2. Os testes foram realizados no software STATA 14.

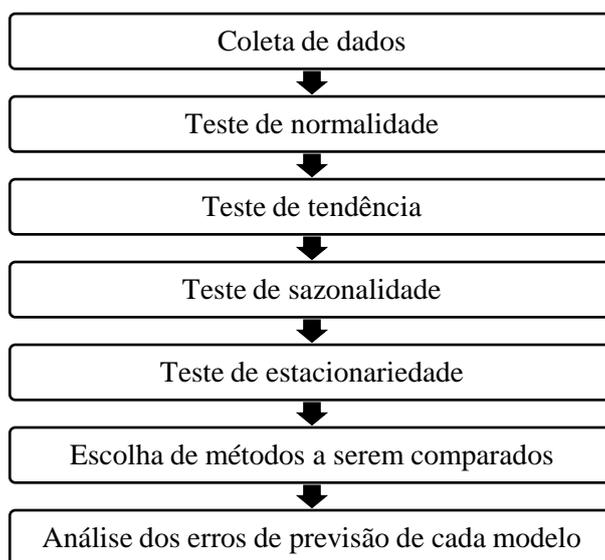


Figura 2 - Sequência de desenvolvimento para análise

O teste de Shapiro-Wilk averigua os desvios acerca de uma população de dados e o resultado do teste aponta se os dados apresentam ou não distribuição normal (FARREL; STEWART, 2006). Se a hipótese nula de que os dados provêm de uma população normal for rejeitada, tem-se que os dados futuros desta série estarão fora dos limites já preestabelecidos na série. Se a hipótese nula do teste não for rejeitada, tem-se que os dados futuros da série estão nos limites do intervalo de confiança adotado para essa população e tem o mesmo comportamento dos dados anteriores. (ROYSTON, 1995).

Inicialmente, o teste de Shapiro-Wilk foi desenvolvido para amostras de até 50 pares de dados (SHAPIRO; WILK, 1965). Entretanto, Royston (1995), aperfeiçoou o teste para amostras no intervalo entre $3 < n < 5000$. Comparando com os testes de Kolmogorov-Smirnov, Lilliefors e Anderson-Darling com o de Shapiro-Wilk, este último é considerado o mais robusto para avaliação de normalidade (RAZALI, 2011).

Um teste para a verificação da presença de tendência em uma série temporal foi proposto por Cuzick (1985). Este é uma extensão do teste não-paramétrico de soma de postos de Wilcoxon, no qual se verifica a presença de movimento de tendência nos pares de dados obtidos. O teste é considerado

robusto dada a suposição de normalidade dos dados mesmo para pequenos tamanhos de amostras (CUZICK, 1985).

Outro teste importante para a suposição de tendência é o teste baseado no coeficiente de correlação de *Spearman*. Este teste mede a intensidade da relação entre variáveis ordenadas. Usa, em vez do valor observado, apenas a ordem das observações. (CONOVER, 1999).

Outro componente importante de uma série temporal é a presença de sazonalidade. Neste cenário, a média dos dados dentro de um ciclo sazonal é diferente, e isso acontece com repetições regulares. Caso contrário, infere-se que todos os períodos possuem a mesma média e que a previsão de demanda independe dos períodos anteriores (BARKSKY; MIRON, 1988). A sazonalidade pode ser determinística, quando os padrões sazonais são regulares ou estáveis no tempo; ou estocástica, quando a componente sazonal da série varia com o tempo (MORETTIN; TOLOI, 2006).

Para a sazonalidade, utilizou-se o teste não-paramétrico de Kruskal-Wallis, que traz resultados referentes à presença de ciclos sazonais determinísticos em uma série temporal. O teste supõe que as variáveis dentro de cada amostra são independentes e que as amostras são independentes entre si. Uma alternativa para o teste de Kruskal-Wallis é o Teste de Friedman – também não-paramétrico - para amostras que relacionam períodos com anos. Neste caso, os períodos (Meses, Trimestres) são considerados “tratamentos” e os anos são considerados “blocos” (MORETTIN; TOLOI, 2006).

A identificação da série em estudo como estacionária e não-estacionária é de fundamental importância. O procedimento desenvolvido por Dickey-Fuller (1979) testa se uma variável tem raiz unitária e equivalente a um passeio aleatório. A hipótese nula é de que a variável possui raiz unitária e, portanto, é não estacionária, ou de que ela possui uma tendência estocástica.

A estacionariedade está relacionada com a média e variâncias desses dados. A avaliação do comportamento desses dois aspectos da série revela se os dados da série apresentam um comportamento aleatório ou não. Para uma série estacionária, a covariância dos dados depende da defasagem entre os períodos de tempo, enquanto que uma série não estacionária, a covariância é independente, prescrevendo um comportamento aleatório (SATO, 2013).

4. Resultados e Discussão

Os resultados para os testes de normalidade, tendência, sazonalidade e estacionariedade para a série de dados de rendimento de soja (t/ha) são apresentados na Tabela 4.

Tabela 3 - Resultados dos testes executados sobre a amostra de dados de rendimento de soja (t/ha) no período entre 1980 e 2014.

Teste	Estatística	Crítico	P-valor	Classificação
Normalidade (<i>Shapiro-Wilk</i>)	W = 0,963	$W_{0,05;35} = 0,934$	0,287	População Normal
Tendência (<i>Cuzick</i>)	Z = 4,61	$Z_{0,05;35} = 1,96$	0,000	Possui Tendência
Tendência (<i>Spearman</i>)	$\rho = 0,790$	$\rho_{0,05;35} = 0,335$	0,000	Possui Tendência
Sazonalidade (<i>Kruskall-Wallis</i>)	H = 33,98	$\chi^2_{0,05;30} = 43,77$	0,282	Não há Sazonalidade
Estacionariedade (<i>Dickey-Fuller 1</i>)	$\tau = -2,045$	$\tau_{0,05;34} = -2,975$	0,267	Não é Estacionária
Estacionariedade (<i>Dickey-Fuller 2</i>)	$\tau = -4,385$	$\tau_{0,05;34} = -3,564$	0,0023	É Estacionária

O resultado do teste de normalidade apontou que os dados de rendimento obtidos possuem um comportamento normal. O p-valor 0,287 apontado pelo teste de Shapiro-Wilk não rejeita a hipótese nula, indicando que os dados estão dentro da curva normal com 95% de confiabilidade. Essa distribuição possui assimetria de -0,083 e um coeficiente de curtose no valor de -1,073, indicando que os dados são praticamente simétricos em relação à média e possuem características de uma distribuição platicúrtica.

O teste de Cuzick mostrou que os dados de rendimento de soja apresentam tendência – p-valor 0,000 é menor que o alfa de 0,05, rejeitando a hipótese nula. Este resultado delimitou os modelos de previsão de demanda mais adequados para aqueles que consideram movimentos de tendência.

O teste de Kruskal-Wallis para sazonalidade não rejeitou a hipótese nula de amostra não sazonal, considerando que o resultado de 0,231 para p-valor é maior que o valor alfa de 0,05 adotado como desejável. Assim, foi necessário considerar um modelo de previsão de demanda que não considera sazonalidade. No estudo, não foi aplicado o teste de Friedman devido a série não possuir tratamentos.

O teste de Dickey-Fuller [1] foi realizado testando a série sem considerar o componente de tendência. O teste não rejeitou a hipótese nula de que os dados são não-estacionários em nível. O resultado de 0,267 para p-valor é maior que 0,05 assumido pelo alfa. Logo, o modelo mais adequado de previsão deve considerar que os dados são não-estacionários. Porém, em uma segunda avaliação, considerando tendência, o teste de Dickey-Fuller [2] resultou em um p-valor = 0,0023, indicando que a série é estacionária na inclinação.

De acordo com os resultados obtidos nos testes de hipóteses, foram feitas simulações com três modelos de previsão: o modelo ARIMA, o modelo de média móvel e o modelo de suavização exponencial corrigido pela tendência (Holt).

O modelo ARIMA é indicado para previsão sobre dados não estacionários. Entretanto, como a série foi considerada estacionária na inclinação, utilizamos o ARIMA sem o componente “I”, equivalente a um modelo ARMA estacionário.

O motivo da escolha é como a série possui tendência e não possui sazonalidade, os resultados do ARIMA foram comparados com o modelo de média móvel, que não contempla a tendência, e de suavização exponencial (Holt), que é adequado para modelos que possuem tendência. Entretanto, ambos os métodos tratam de dados estacionários.

Essa comparação foi escolhida pois os modelos de média móvel e Holt são de fácil modelagem, e não necessitam de software específico. Como a aplicabilidade do artigo está em situações reais de baixa complexidade (rendimento de soja por unidade de área) e a previsão de rendimento não precisa de um alto nível de precisão, os dois modelos para casos estacionários trouxeram resultados comparáveis de erro.

Os resultados referentes ao erro de previsão de cada um dos modelos de previsão avaliados estão dispostos na Tabela 5. Como se pode notar, a variação entre os resultados é de menos de um ponto percentual entre o melhor resultado (ARIMA [2,1,0], 10,31%) e o pior (Modelo de Holt, 11,19%).

Com base nas autocorrelações e nas autocorrelações parciais, foram definidos os componentes AR e MA do modelo ARIMA. O melhor desempenho demonstrado por este método era esperado devido à sua robustez. O modelo sem diferenciação ARMA resultou um valor EPAM de 10,83%. Porém, seus coeficientes AR não foram estatisticamente significativos. Sendo assim, optou-se em realizar o ARIMA diferenciando uma vez para estacionar os níveis, visto que ele é o método indicado para dados não-estacionários. Os resultados se mostraram melhores com simulações com um, dois e três processos auto-regressivos, embasados nas autocorrelações parciais da série. O melhor resultado foi obtido com dois processos auto-regressivos, tendo em vista que somente dois coeficientes (AR) foram estatisticamente significativos.

Os resultados obtidos pelo modelo ARIMA e a os valores reais de rendimento estão mostrados na Figura 3. É possível notar que as linhas de duas e três auto-regressões tem o mesmo comportamento, evidenciando o fato de somente ser necessário fazer duas auto-regressões para chegar ao modelo de previsão com melhor resultado.

Tabela 4 - Percentual de erro para os modelos de previsão avaliados

TIPO DE DADO	TESTE	ERRO
Dados não-estacionários	ARIMA (3,1,0)	10,35%
	ARIMA (2,1,0)	10,31%
	ARIMA (1,1,0)	10,47%
Dados estacionários	ARIMA (1,0,0)	10,83%
	Média Móvel 2 períodos	10,82%
	Média Móvel 3 períodos	10,67%
	Holt (0,5 e 0,3)	11,19%

Holt (0,6 e 0,1)	10,45%
------------------	---------------

Comparando os métodos para dados estacionários, o melhor resultado aparece para o método de suavização exponencial corrigido pela tendência (Holt) com coeficiente alfa igual a 0,6 e beta igual a 0,1, cujo resultado de MAPE é de 10,41%. Este resultado é superior ao método ARIMA com uma auto-regressão apenas.

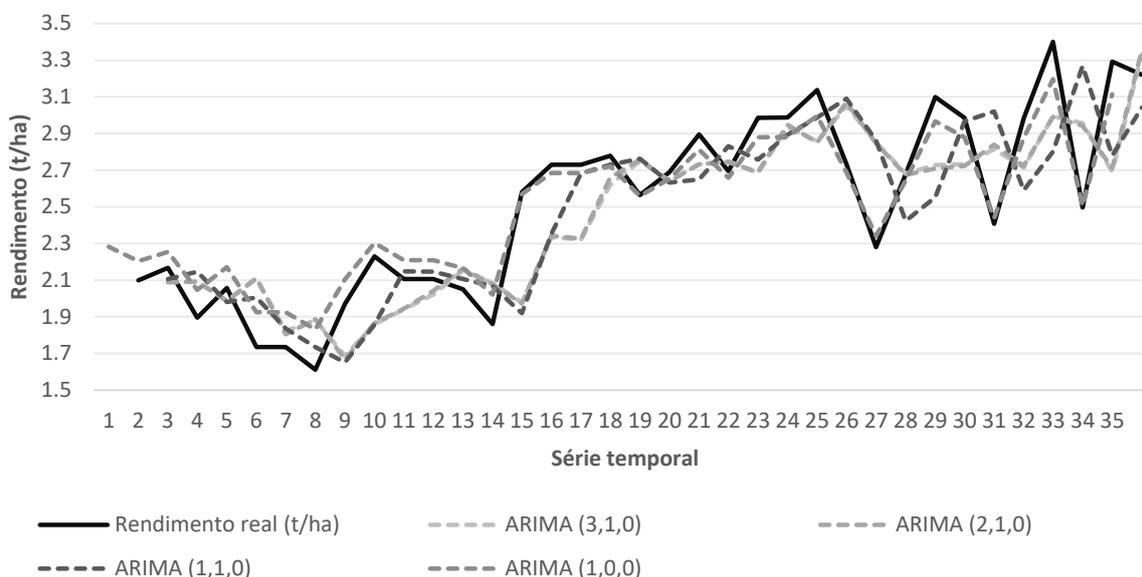


Figura 3 - Comparação entre rendimento de soja real e métodos de previsão ARIMA.

O valor de 0,6 para alfa indica que a previsão é bastante impactada pelos erros da demanda anterior ($t-1$), enquanto que o valor de 0,1 para beta indica uma baixa influência da tendência. Ou seja, o resultado da previsão está mais atrelado aos erros dos dados anteriores do que ao comportamento de tendência evidenciado na série.

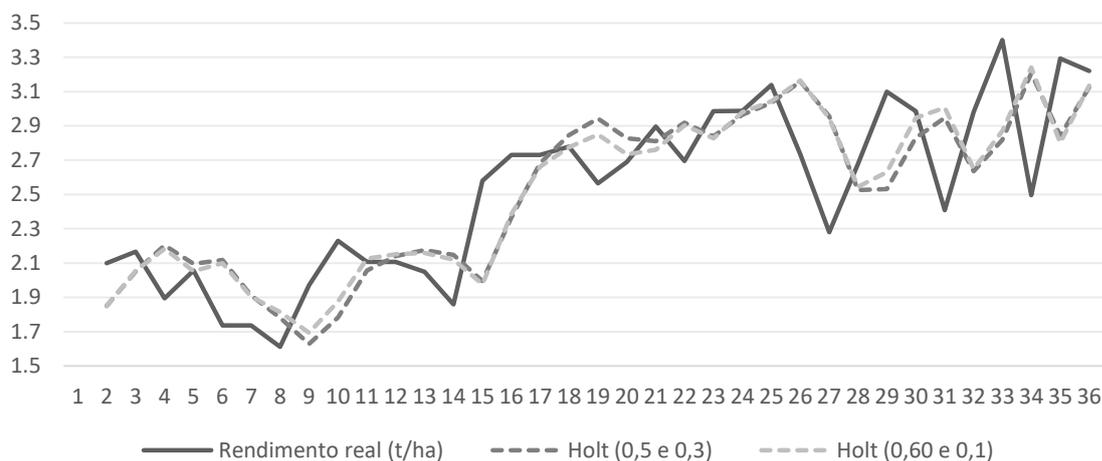


Figura 4 - Comparação entre rendimento real e métodos de suavização exponencial simples (Holt).

Entre os modelos para séries temporais, o pior resultado também foi obtido na simulação com o modelo de Holt. Para valores de alfa e beta iguais a 0,5 e 0,3, respectivamente, obteve-se um EPAM de 11,19%. Os resultados para a simulação com o método de Holt estão representados na Figura 4.

Os modelos de média móvel trouxeram resultados intermediários dentre as opções de simulação para dados estacionários. Já em relação às simulações do método ARIMA, a média móvel obteve resultados inferiores.

Foram feitas simulações com dois e três períodos a partir da observação gráfica da variação do rendimento de soja ao longo do período. A Figura 5 apresenta a comparação entre o rendimento real de soja (t/ha) e as duas simulações de média móvel.

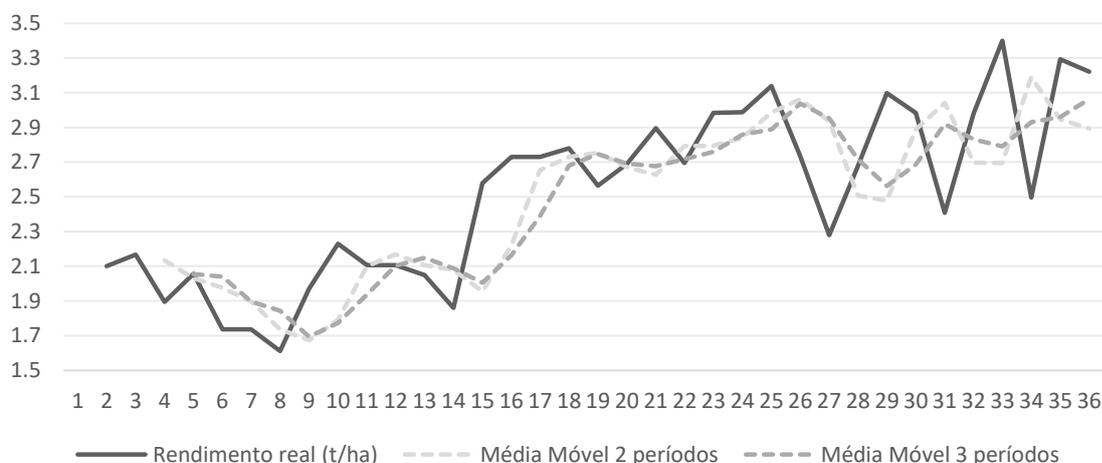


Figura 5 - Comparação entre rendimento real e método de média móvel

Considerando o EPAM individual de cada modelo, os valores de erro entre 10% e 11%, aproximadamente, são representativos para o planejamento da cadeia de suprimentos da soja. Considerando uma área de plantio de 53.000 ha, que foi a área plantada em 2014, e um rendimento de 3,2 toneladas por hectare colhido (IPARDES, 2016), o erro das modelagens representa uma variação de aproximadamente 17.000 toneladas de soja, entre os valores mínimo e máximo previstos, que impactam na logística de recebimento, armazenagem e escoamento dos grãos.

Comparando os EPAMs de cada modelo, para os produtores, notou-se que a diferença entre o valor maior e menor não é tão relevante do ponto de vista prático do problema. Para quem produz, esta variação de um ponto percentual representa 32 kg de soja em uma área de um hectare, que produziu cerca de 3.200 kg na última safra (IPARDES, 2016). Logo, se a escolha do modelo mais adequado para a previsão avalia os métodos pesquisados, o resultado final está atrelado à facilidade de obtenção e avaliação dos resultados, sob a ótica de recursos (*softwares* específicos) e conhecimento técnico. Como os EPAMs são muito parecidos, a dificuldade de operação e avaliação dos resultados é mais relevante para o produtor do que a variação do erro entre eles.

5. Considerações finais

O foco da pesquisa foi encontrar uma alternativa prática e economicamente viável para que tanto os produtores quanto as cooperativas de soja da região possam prever o rendimento das regiões produtoras de soja, melhorando o planejamento da cadeia de suprimentos em torno da *commodity*. Durante o desenvolvimento do artigo, não foi identificado nenhum estudo ou aplicação de modelos de previsão de demanda para a produção de soja em uma região pequena e específica, com características de plantio pouco heterogêneas.

Dentre os resultados de desempenho apresentados pelos modelos comparados, o modelo de Holt se apresentou como uma solução viável prática e economicamente para os produtores agrários. O erro para a simulação com alfa igual a 0,6 e beta igual a 0,1 é o terceiro melhor resultado entre todas as comparações, mas não necessita de *software* e conhecimento específico para fazer a previsão do período seguinte e avaliar seu resultado.

Porém, os erros de previsão na casa dos 10% são relevantes para o planejamento de toda a cadeia de suprimentos da soja. Para tanto, os modelos de previsão que envolvem estimativas para recebimento de grãos, por exemplo, devem ser estudadas com mais profundidade, para encontrar um modelo cujo resultado seja mais competitivo frente aos apresentados nesta pesquisa.

O estudo em questão considera uma série histórica de 35 pares de dados. Para avaliar modelos de previsão de maneira robusta, essa é a quantidade mínima necessária para validar um modelo de

previsão. Esta limitação está atrelada a disponibilidade de dados medidos referentes ao plantio de soja na região analisada. Considerando séries históricas com mais dados, os modelos de previsão podem ser melhor avaliados e comparados.

Ademais, existem variáveis exógenas que interferem no rendimento, como, por exemplo, as práticas de agricultura e as condições do solo no período de plantio. A agricultura de precisão é um forte influenciador dos dados de rendimento e este aspecto não é levado em consideração quando modelos de previsão são estudados a partir de séries históricas. Essas intervenções que interferem na qualidade do plantio não estão relacionadas com o tempo, portanto, são imprevisíveis e não apresentam um comportamento que pode ser explicado pelo histórico temporal dos dados.

Referências

BARSKY, R.B.; MIRON, T. A. *The seasonal cycle and the business cycle*. NBER working paper series. N.2688. 1988.

CEPEA/USP – Centro de Estudos Avançados em Economia Aplicada. Disponível em: <http://cepea.esalq.usp.br/pib/>. Acessado em: 13/06/2016.

CHOPRA, S.; MEINDL, P. **Gestão da Cadeia de Suprimentos: estratégia, planejamento e operações**. São Paulo: Pearson Prentice Hall, 2011.

CONAB - Companhia Nacional de Abastecimento, 2016.

CONOVER, W. J. *Practical nonparametric statistics*, 3rd ed. Wiley. New York, p. 250-257, 1999.

CUZICK, J. *A Wilcoxon-type test for trend*. **Statistics in Medicine** 4: 87–90. 1985.

DICKEY, D. A.; FULLER, W. A. *Distribution of the estimators for autoregressive time series with a unit root*. **Journal of the American Statistical Association** 74: 427–431. 1979.

DIEBOLD, F.X.; LOPEZ, J.A. *Forecast Evaluation and Combination*. *Technical Working Paper Series*. National Bureau of Economic Research. Cambridge, 1996.

ENDERS, Walter. *Applied econometric time series*. John Wiley & Sons, 2008.

FARREL, P. J. STEWART, K. R. *Comprehensive Study Of Tests For Normality And Symmetry: Extending The Spiegelhalter Test*. **Journal of Statistical Computation and Simulation**. v. 76, n. 9, pp. 803–816, 2006.

FERNANDES, F. C. F.; FILHO, M. G. **Planejamento e Controle da Produção: dos fundamentos ao essencial**. São Paulo: Atlas S.A., 2010.

GIL, A. C. **Como Elaborar Projetos de Pesquisa**. 4.ed. São Paulo: Atlas S. A., 2002.

GOOD, D.L.; IRWIN, S.H. *Understanding USDA Corn and Soybean Production Forecasts: Methods, Performance and Market Impacts over 1970 – 2005*. 2006.

GUSSO, A.; FORMAGGIO, A.R.; RIZZI, R.; ADAMI, M.; RUDORFF, B.F.T. *Soybean Crop Area Estimation by Modis/Evi Data*. **Pesquisa Agropecuária Brasileira**. Brasília. v.47, n. 3, p. 425-435, 2012.

IBGE - Instituto Brasileiro de Geografia e Estatística [1]. *Indicadores IBGE*. Disponível em: ftp://ftp.ibge.gov.br/Contas_Nacionais/Contas_Nacionais_Trimestrais/Fasciculo_Indicadores_IBGE/pib-vol-val_201504caderno.pdf. Acessado em: 13/06/2016.

- IBGE - Instituto Brasileiro de Geografia e Estatística** [2]. *Censo Agropecuário 2006*. 2016.
- IBGE - Instituto Brasileiro de Geografia e Estatística** [3]. *Levantamento Sistemático da Produção Agrícola*. Disponível em: [ftp://ftp.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_\[mensalFasciculo/lspa_201603.pdf](ftp://ftp.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_[mensalFasciculo/lspa_201603.pdf). Acessado em: 26/04/2016.
- IPARDES – Instituto Paranaense de Desenvolvimento Social e Econômico**. *Base de Dados do Estado*. Disponível em: <http://www.ipardes.gov.br/imp/index.php>. Acessado em: 26/04/2016.
- JOHNSON, D. M. *An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States*. **Remote Sensing of Environment**, 2013.
- KAUARK, F.; MANHÃES, F. C.; MEDEIRAOS, C. H. **Metodologia da Pesquisa: guia prático**. Itabuna: Via Litterarum, 2010.
- KRAJEWSKI, L.; RITZMAN, L.; MALHOTA, M. **Administração de Produção e Operações**. São Paulo: Pearson Prentice Hall, 2009.
- LEÃO, L. M. **Metodologia do Estudo e Pesquisa: facilitando a vida dos estudantes, professores e pesquisadores**. Petrópolis: Vozes, 2016.
- MAKRIDAKIS, S. G.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. *Forecasting: methods and applications*. 3. ed. New York: John Wiley & Sons, 1998.
- MASUDA, T.; GOLDSMITH, P. D. *World Soybean Production: Area Harvested, Yield, and Long-Term Projections*. **International Food and Agribusiness Management Review**. V. 12, N. 4, 2009.
- MONTGOMERY, D.C.; JENNINGS, C. L.; KULAHCI, M. *Introduction to time series analysis and forecasting*. Hoboken: John Wiley & Sons, 2008.
- MOREIRA, D. A. **Administração da Produção e Operações**. 2.ed. São Paulo: Cengage Learning, 2014.
- MOREIRA, D. A. **Administração da Produção e Operações**. São Paulo: Pioneira Thomson Learning, 2001.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2. ed. São Paulo: Blucher, 2006.
- PATINO, M.T.O.; MACHADO, M.F.; DO NASCIMENTO, G.T.; DE ALCÂNTARA, M.R. *Analysis and forecast of the storage needs of soybeans in Brazil*. **Engenharia Agrícola, Jaboticabal**. v. 33, n. 4, p. 834-843, Julho/Agosto 2013.
- PEINADO, J.; GRAEML, A. R. **Administração da Produção: operações industriais e de serviços**. Curitiba: Unicenp, 2007.
- PINDYCK, R. S.; RUBINFELD, D. L. *Econometria: modelos e previsões*. ed. 4. São Paulo: Campus – Elsevier. 2004.
- RAZALI, Nornadiah Mohd et al. *Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests*. **Journal of statistical modeling and analytics**, v. 2, n. 1, p. 21-33, 2011.
- ROYSTON, Patrick. Remark AS R94: *A remark on algorithm AS 181: The W-test for normality*. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, v. 44, n. 4, p. 547-551, 1995.

SHAPIRO, S. S.; WILK, M. B. *An analysis of variance test for normality (complete samples)*. **Biometrika**. e. 52. p. 591. 1965.

TILMAN, D.; BALZER, C.; HILL, J.; BEFORT, B.L. *Global food demand and the sustainable intensification of agriculture*. **PNAS**. v. 108. n. 50. 2011.