

Eixo Temático: Estratégia e Internacionalização de Empresas

PREVISÃO DA ÁREA PLANTADA DE ARROZ NA REGIÃO SUL DO BRASIL

FORECAST AREA PLANTED RICE IN THE REGION SOUTH OF BRAZIL

Deise Scheffer e Adriano Mendonça Souza

RESUMO

A cultura do arroz é responsável por grande parte da economia do Rio Grande do Sul, considerado pela Companhia Nacional de Abastecimento (Conab) como um dos principais produtos da cesta básica. Entre as culturas produzidas no estado o arroz ocupa o terceiro lugar em relação à área plantada, sua importância pode relacionar-se à diversidade da sua utilização. Este trabalho tem como objetivo conhecer o processo gerador da série temporal da área plantada, por meio da análise descritiva dos dados, gráficos e tabelas. Verificar a presença de autocorrelação e aplicar a metodologia *Box-Jenkins* através do MAPE o melhor modelo para a série em estudo, para a realização das previsões a horizontes futuros, através dos dados históricos. O modelo ARIMA(2,1,0) apresentou-se mais adequado entre os modelos que foram encontrados para gerar previsões. O *software* utilizado nas análises estatísticas foi o *EViews 8*.

Palavras-chave: Séries Temporais, Metodologia Box-Jenkins, Arroz.

ABSTRACT

The rice crop accounts for much of the Rio Grande do Sul economy, considered by the National Supply Company (Conab) as one of the main products of the basic food basket. Among the crops grown in the state rice ranks third in relation to planted area, its importance can relate to diversity will use. This work aims to know the generating process of the time series of the planted area, using descriptive data analysis, graphs and tables. Check for autocorrelation and apply the Box-Jenkins methodology through MAPE the best model for the series under study, to carry out forecasts of future horizons, through historical data. The ARIMA model (2,1,0) appeared more suitable between models that have been found to generate forecasts. The software used in the statistical analysis was the *EViews 8*.

Keywords: Time Series, Box-Jenkins methodology, Rice.

1 INTRODUÇÃO

O arroz é considerado o produto de maior importância econômica em muitos países em desenvolvimento, constituindo-se alimento básico para cerca de 2,4 bilhões de pessoas. É uma cultura que apresenta grande capacidade de adaptação a diferentes condições de solo e clima. Cultivado e consumido em todos os continentes, o arroz se destaca pela produção e área de cultivo, desempenhando papel estratégico tanto em nível econômico quanto social para os povos das nações mais populosas da Ásia, África e América Latina (EMBRAPA, 2010).

Em 2014 o total colhido de arroz chegou a 12,2 milhões de toneladas, superando em 3,3% do que foi colhido em 2013. Para o IBGE, a safra do arroz em 2014 foi favorecida devido as condições climáticas, com elevados níveis de mananciais de irrigação na Região Sul. O Rio Grande do Sul é o maior produtor dessa cultura, com 67,8% da produção nacional (IBGE, 2014).

Segundo a EMBRAPA, o arroz é cultivado e consumido em todos os continentes, o arroz destaca-se pela produção e área de cultivo, desempenhando papel estratégico tanto no aspecto econômico quanto social.

Uma das principais fontes alimentícias em diversos países, o arroz, junto com o trigo e o milho é um dos principais cereais cultivados no mundo. Com a maior parcela do consumo nos países asiáticos, o arroz se tornou um alimento básico para grande parte da população mundial. Segundo dados da FAO (2004, apud FERREIRA et al., 2005), no asiático grande contingente da população obtém de 60% a 70% do consumo de energia do arroz e de seus derivados. Ademais, é o alimento que apresenta as maiores taxas de crescimento de consumo na África nas últimas décadas, além da importância para a segurança alimentar em países em desenvolvimento.

Considerado uma cultura pioneira no Brasil, o arroz avançou em todo o processo de ocupação das fronteiras agrícolas no país, antecedendo outras culturas agrícolas e a pecuária bovina. Esta prática é utilizada ainda hoje, nos processos de ocupação dos cerrados e das franjas da Amazônia. Porém é uma produção marginalizada, utilizada em maior parte para auto-abastecimento local, sem muita importância no sistema agroindustrial como um todo (GIORDANO; SPERS, 1998).

Segundo Silva (2004), o arroz é uma das mais importantes culturas anuais produzidas no Brasil, sendo cultivado em praticamente todos os Estados e consumido por todas as classes sociais do país, ocupando posição de destaque social e econômico, pelo fato de gerar renda e suprir a população brasileira com um significativo conteúdo de nutrientes e calorias na dieta básica da população, principalmente nos estratos de menor renda.

Dentre os maiores produtores mundiais de arroz, o Brasil se coloca na nona posição desta commodity, com uma produção de 12,1 milhões de toneladas de arroz em casca na safra 2007/08, chegando a alcançar cifra de 13,3 milhões de toneladas no ano de 2004/05 CONAB (2008, apud SANTOS et al., 2009), recorde histórico de produção no país. Além disso, apresentou um consumo per capita médio de aproximadamente 55 kg por habitante entre 1994 e 2003 FAO (2008, apud SANTOS et al., 2009).

No que diz respeito às safras de arroz no Brasil, Adami (2005) retrata que a colheita mensal do arroz está distribuída de forma que os períodos de maior volume se concentram entre os meses de março e abril, ocorrendo o período de entressafra, ou seja, quando não há colheita de arroz, entre os meses de agosto a dezembro.

Devido à importância da plantação de arroz na economia do Rio Grande do Sul. Assim como para o país, este estudo irá desenvolver um modelo para realizar previsões da área plantada deste produto nos próximos anos, através de modelos econométricos.

2 METODOLOGIA

Os dados empregados para realização deste trabalho foram adquiridos na Companhia Nacional de Abastecimento (Conab) e, corresponde ao período de 1976 a 2014, da área plantada de arroz na Região Sul do Brasil.

Primeiramente foi feita uma análise descritiva da série. Logo, foi verificada a presença de autocorrelação. Na etapa inicial é preciso verificar sua estacionariedade, onde é necessário a média, variância e a autocovariância serem constantes, caso contrário, aplica-se as transformações para estacioná-la.

Constatada a estacionariedade da série, a próxima fase será a identificação da estrutura do modelo, pelas funções: autocorrelação (ACF) e autocorrelação parcial (PACF).

Na próxima etapa, estimação, onde a notação utilizada para designar o modelo ARIMA(p,d,q), onde AR(p) corresponde a parte autoregressiva de ordem p é explicada por $y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varepsilon_t$, φ indica os parâmetros reais e ε_t ruído branco. O processo de médias móveis MA(q) tem como equação $y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$. Onde, θ indica constantes reais e ε_t o ruído branco. Conhecido como um conjunto de variáveis aleatórias com distribuição e variância constantes e média igual a zero. A parte I(d) do modelo ARIMA é referente a diferenciação, que por sua vez é a subtração da segunda observação pela primeira demonstrado na fórmula $\Delta^d y = y_t - y_{t-1}$, o valor diferenciado é o valor de y no período t-1, representado por d (MORETTIN, 2008).

Depois de encontrado o modelo faz-se a validação, através da análise dos resíduos para verificar se os mesmos são ruído branco, isto é, se possuem média zero, variância constante e não são autocorrelacionados, além de se utilizar as estatísticas de ajustes AIC e BIC, para determinar o melhor modelo.

$$AIC = \ln \hat{\sigma}_{\square}^2 + \frac{2(p+q)}{n}$$

$$BIC = \ln \hat{\sigma}_{\square}^2 + \frac{2(p+q)\ln(n)}{n}$$

Em seguida, modelos Box-Jenkins determinará, através da MAPE o melhor modelo para a série em estudo, com o objetivo de realizar uma previsão para a área plantada de arroz na região Sul do país nos anos seguintes. O software utilizado para análise das medidas estatísticas foi o *EViews 8*.

Para orientar os estudos e as análises das séries temporais George Box e Gwilyn Jenkins em 1976 desenvolveram um procedimento de análise de séries temporais que se tornou mundialmente conhecido como metodologia de Box & Jenkins, sendo que a previsão é um dos principais fatores da popularidade deste modelo. (Souza et al., 2011).

Segundo Morettin e Tolo (2006), a estratégia para a construção do modelo será baseada em um ciclo iterativo, onde a escolha da estrutura do modelo é baseada nos próprios dados. Os estágios do ciclo iterativo são:

- (a) uma classe geral de modelos é considerada para análise (especificação);
- (b) há identificação de um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- (c) estimação, na qual os parâmetros do modelo identificado são estimados;
- (d) e finalmente, há a verificação ou diagnóstico do modelo ajustado, por meio de uma análise de resíduos, para se saber se este é adequado ou não.

Segundo SOUZA (1996), a metodologia de Box-Jenkins foi baseada em um ciclo iterativo, conforme na Figura 1, na qual a estrutura do modelo é formada pelos próprios elementos da série:

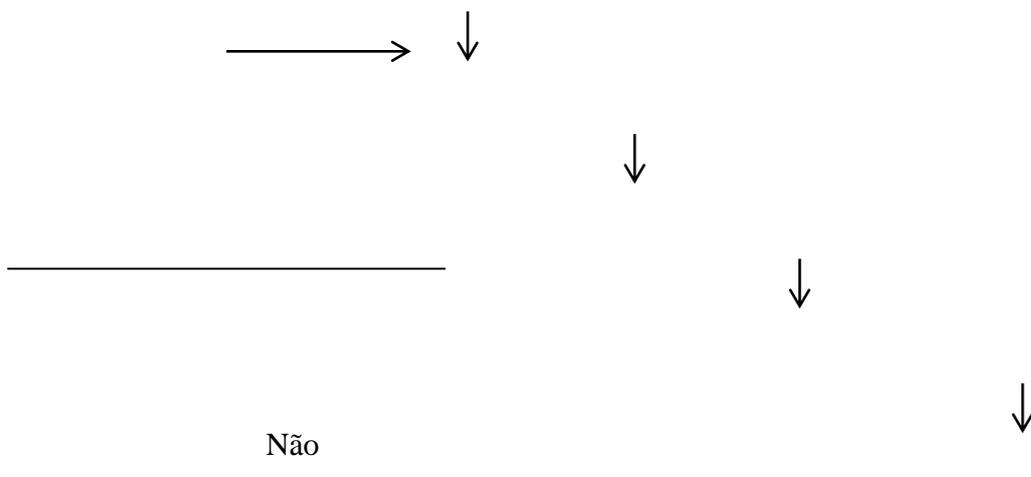


Figura 1: Fluxograma do Ciclo iterativo de Box-Jenkins

Caso o modelo não seja adequado, o ciclo é repetido, voltando-se a fase de identificação. Um procedimento que muitas vezes é utilizado é identificar não só um único modelo, mas alguns modelos concorrentes que serão então estimados e verificados. Logo, quando se obtém um modelo satisfatório, passa-se para a última etapa da metodologia de Box-Jenkins, ou seja, a realização de previsões.

As métricas de comparação usadas correspondem ao erro percentual médio absoluto (MAPE), que penaliza mais os erros maiores; e o coeficiente U de Theil, que avalia o desempenho da previsão em relação à previsão ingênua ou trivial, ou seja, previsão ingênua ou trivial significa que a estimativa do valor futuro é igual ao valor atual.

A métrica MAPE, *Mean Absolute Percentage Error*, indica o valor médio do erro percentual das previsões e é definido pela seguinte fórmula:

$$MAPE = \frac{100}{N} \times \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right|$$

Onde, N é o número de previsões realizadas e x_i e \hat{x}_i os valores observado e previsto para o índice i , respectivamente.

O coeficiente U de Theil é definido como:

$$U = \frac{\sqrt{\sum_{i=1}^N (a_k - y_k)^2}}{\sqrt{\sum_{i=1}^N (a_k - a_{k-1})^2}}$$

O coeficiente acima analisa a qualidade de uma previsão através dos seguintes valores:

$U > 1$, significa que o erro do modelo é maior do que da previsão ingênua;

$U < 1$, significa que o erro do modelo é menor que da previsão ingênua.

O coeficiente U de Theil menor do que 1 já indica uma previsão melhor que a previsão ingênua; quanto mais próximo o mesmo for de zero, melhor será o resultado da previsão (BARROSO, 2010)

3 RESULTADOS E DISCUSSÃO

Com o intuito de gerar um modelo de previsões para a série estudada, primeiramente reuniu-se os dados históricos do ano/safra de 1976/1977 a 2014/2015 da área plantada no Rio Grande do Sul.

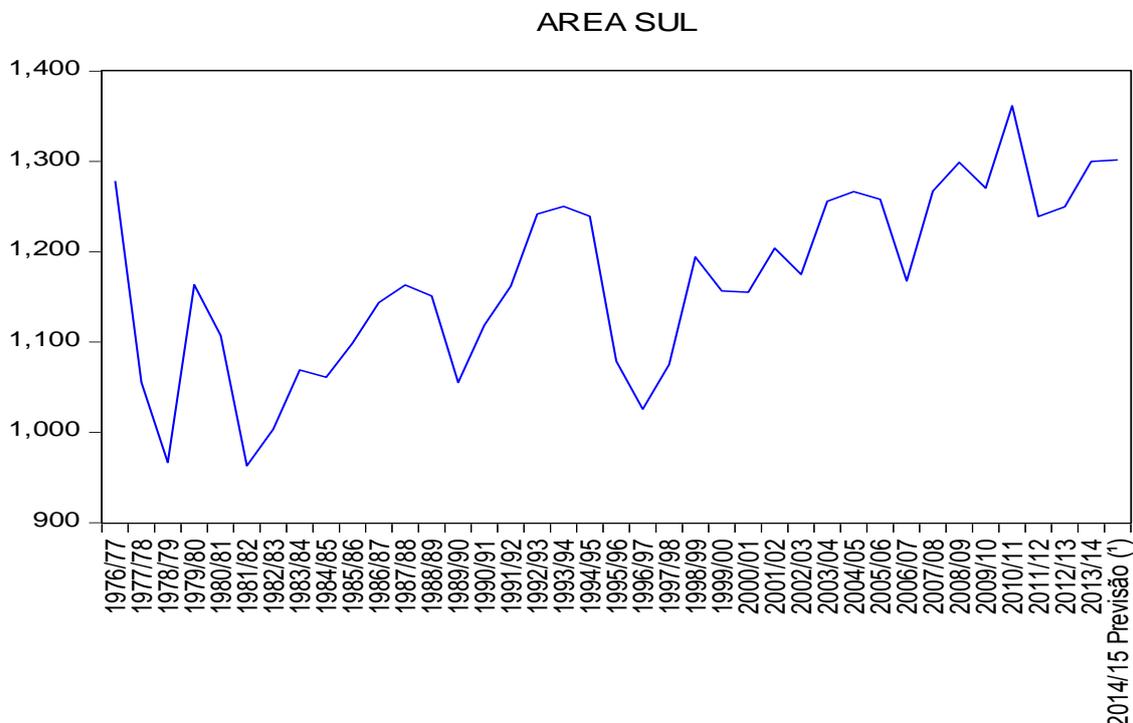


Figura 1: Gráfico da Série de área plantada de arroz do estado do Rio Grande do Sul (1976/77-2014/15)

Na análise visual do gráfico da série de área plantada de arroz do Rio Grande do Sul (Figura 1), foi possível observar que a Série não assume um comportamento estacionário por apresentar uma tendência crescente e picos. Logo, sabe-se que a análise visual não é suficiente e recorreu-se aos testes Dick-Fuller (ADF) e KPSS, através da análise das Funções de Autocorrelação da Série Original, e aos testes de significância, para se concluir se a Série apresenta tendência.

Pode-se observar visualmente que a série não é estacionária, neste caso devemos aplicar uma diferença, mas para confirmarmos esta não estacionariedade fizemos o teste de estacionariedade Kwiatkowski, Phillips, Schmidt and Shin (KPSS), e o Teste da raiz unitária Augmented Dickey-Fuller (ADF).

O próximo passo é analisar as funções de autocorrelações (FAC) e de autocorrelações parciais (FACP) da série da área plantada de arroz, sendo que o comportamento dessas funções auxilia na verificação da estacionariedade e na proposição do modelo.

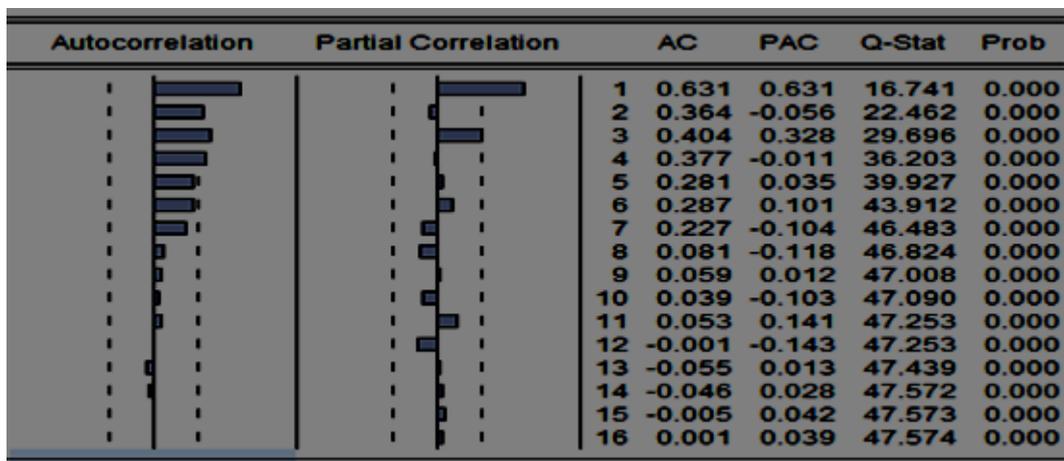


Figura 2: Correlogramas das FAC e FACP da série de área plantada de arroz no Sul

A figura 2 apresenta a FAC e a FACP da série de área de arroz plantada na região Sul do Brasil, desta forma, obtém-se que as autocorrelações da FAC apresentam um leve decaimento exponencial, e o correlograma da FACP apresenta as duas primeiras defasagens (lags) diferentes de zero. Portanto, há uma indicação de que a ordem do modelo auto-regressivo é $k=2$, sendo um modelo ARIMA(2,1,0).

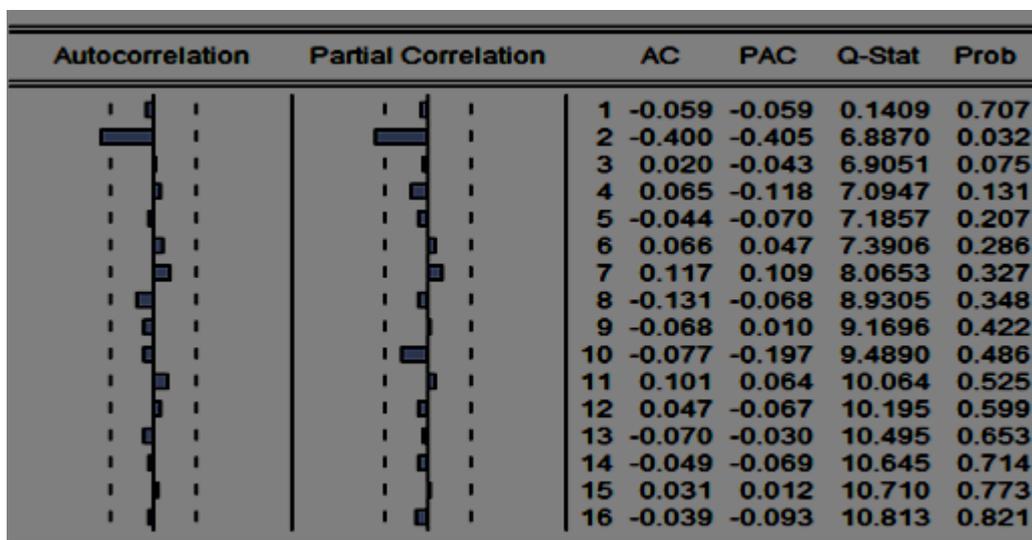


Figura 3: FAC e a FACP da série de arrecadação do ICMS no estado do Rio Grande do Sul, em primeiras diferenças.

Foram analisados vários modelos concorrentes, com e sem sazonalidade, porém foram escolhidos os mais representativos, cujos principais resultados podem ser observados na Tabela 1. Os demais modelos não relacionados juntamente com os abaixo foram excluídos por apresentar coeficientes não significativos ou porque tinham variância residual maior que os relacionados.

Tabela 1 – Modelos concorrentes para a área plantada de arroz no sul do país

ÁREA	Coef	Std. Erro	t-std	p-valor	AIC	BIC	RB
AR(1)	0,9981	0,011	86,86	0,00	11,69	11,74	
AR(1)	0,9435	0,151	6,24	0,00	11,55	11,64	sim
AR(2)	0,0603	0,151	0,39	0,00			
MA(1)	0,9601	0,021	45,44	0,00	15,69	15,73	
MA(1)	1,3825	0,043	32,12	0,00	14,71	14,79	
MA(2)	0,9212	0,032	28,47	0,00			
AR(1)	1,003	0,0123	80,94	0,00	11,45	11,54	
SAR(12)	0,06	0,1591	0,37	0,70			
MA(1)	0,979	0,008	121,74	0	14,58	14,67	
SMA(12)	0,822	0,044	18,47	0			

Fonte:

Elaborado pelos autores

Através da estimação de vários modelos, observou-se os modelos significativos encontrados bem como seus parâmetros. Como regra de decisão para optar pelo melhor modelo, usou os Critérios o AIC e o BIC, que definem os modelos parcimoniosos, por meio dos menores valores encontrados. Desta forma, o modelo mais significativo é o ARIMA(2,1,0).

O modelo ARIMA(2,1,0) foi o que mais se ajustou a série de área plantada no Sul do Brasil, pois analisando os critérios AIC (Akaike info criterion) e BIC (Schwartz Bayesian Criteria) nos modelos concorrentes, o modelo ARIMA(2,1,0) apresentou os menores valores para esses critérios, 11,55 e 11,64, respectivamente.

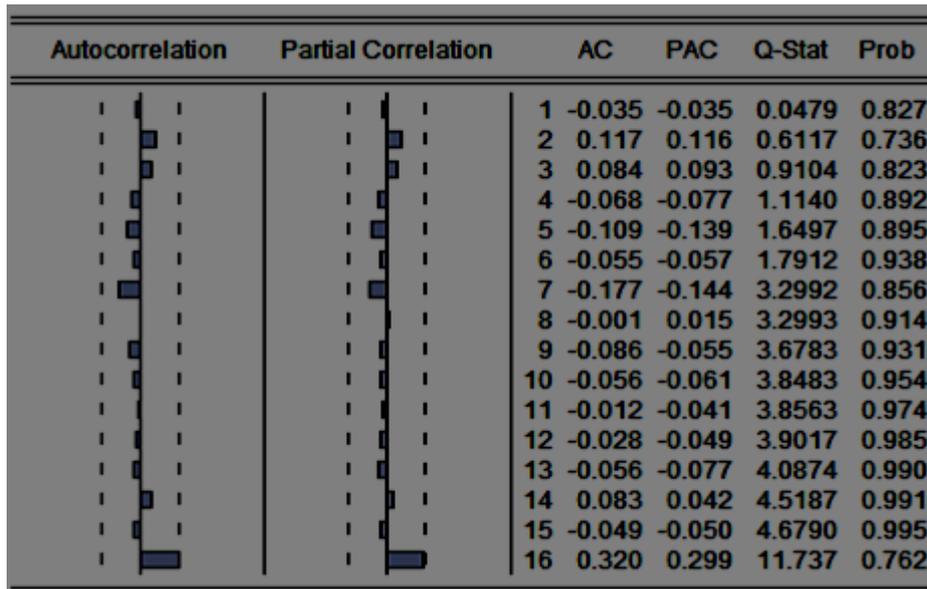


Figura 5 - FAC e FACP da série de resíduos do modelo ARIMA(2,1,0).

A FAC residual do modelo mostra que a autocorrelação dos resíduos estão dentro do limite de confiabilidade de 95%, assim como a FACP, indicando que estes resíduos não são correlacionados.

Como forma de observar o comportamento da série ajustada e dos resíduos, observa-se, na Figura 6, que a série ajustada converge para a série real, o que gera resíduos próximos de zero, indicando que o modelo ajustado é um bom representante da área plantada de arroz no Sul do Brasil e validando o mesmo para que se façam previsões futuras.

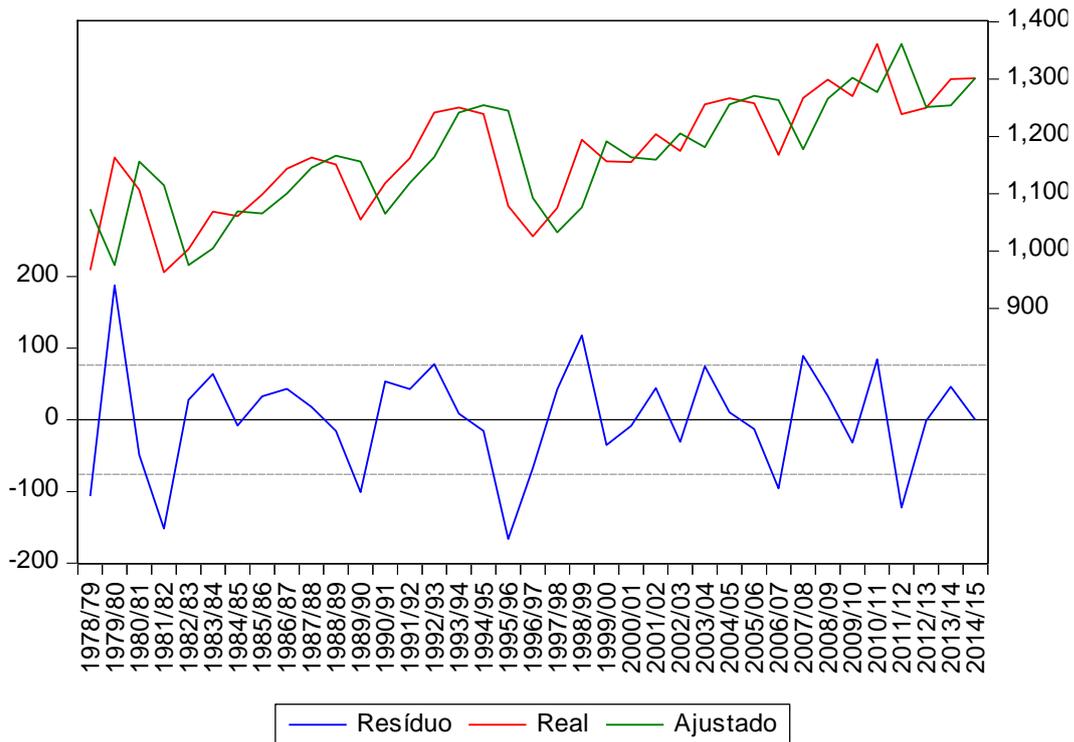


Figura 6 – Gráfico representativo da série real e da série ajustada

Utilizando o modelo selecionado, realizou-se a previsão dentro da amostra para o período de análise, conforme Figura 7.

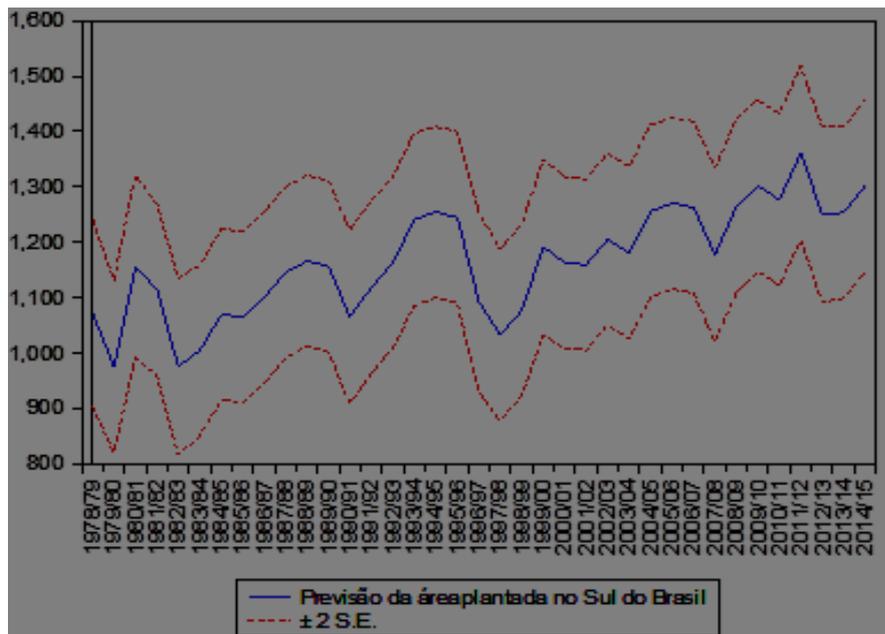


Figura 7 – Gráfico de Previsão da área plantada de arroz na Região Sul do país.

Observa-se que os valores previstos apresentam o comportamento semelhante à série original com estatísticas de ajustes de erro percentual médio absoluto de 5,036% e coeficiente U-Theil = 0,031.

4 CONCLUSÃO

O estudo realizado teve como objetivo encontrar um modelo de previsão confiável para a previsão da área plantada de arroz, na região sul do Brasil na safra 2014/2015, pois sendo considerado o terceiro principal produto na economia sul do país, destaca-se a importância não só para os produtores como também para a economia do país como um todo.

Assim através dos resultados obtidos verifica-se que a previsão de melhor qualidade ou que melhor se ajustou aos dados, segundo os critérios utilizados neste presente estudo, foi o modelo ARIMA(2,1,0).

Por meio do ajuste deste modelo foi possível verificar e prever uma tendência crescente da área plantada de arroz no sul do país, logo, é possível auxiliar nas decisões futuras, como novos investimentos e possivelmente novos projetos para empresários no ramo.

5 REFERÊNCIAS BIBLIOGRÁFICAS

BARROSO, S. N. O. **Abordagem HPA para a modelagem de dados de alta frequência: aplicação à previsão horária de carga elétrica.** 2010. 24f. Dissertação (Mestrado em Engenharia Elétrica) – PUC-Rio, Rio de Janeiro, 2010.

BAYER, F. M.; Souza, A. M. **Wavelets e modelos tradicionais de previsão: Um estudo comparativo.** Revista Brasileira de Biometria. v.28, p.40-61, 2010.

BOX, G. E. P. e JENKINS, G. M., 1970. **Time Series Analysis, Forecasting and Control.** Holden-Day

Companhia Nacional de Abastecimento (CONAB): <http://www.conab.gov.br/>

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento. Disponível em: <https://www.embrapa.br/>
Acesso em 01/06/2015.

IBGE - Instituto Brasileiro de Geografia e Estatística. Disponível em:
<http://www.ibge.gov.br/home/> Acesso em 01/06/2015.

GUJARATI, D. N. **Econometria Básica.** Rio de Janeiro: 4ª Ed. AMGH Editora Ltda, 2006.

MORETTIN, P.A . & TOLOI, C.M.C. **Análise de Séries Temporais.** São Paulo, Edgard Blücher, 2006.

PINTO, P. L. A. ; MAIA, S. F. . **Previsão do preço das commodities agrícolas brasileiras: uma abordagem por modelos univariados.** In: XLVI Congresso da Sociedade Brasileira de Economia e Sociologia Rural - SOBER Amazônia mudanças globais e agronegócio: O desenvolvimento em questão. 2008, Rio Branco - Acre. Anais do XLVI Congresso da

Sociedade Brasileira de Economia e Sociologia - SOBER Rural Amazônia mudanças globais e agronegócio: O desenvolvimento em questão., 2008.

SOUZA, r.c.; CAMARGO, M.E., **Análise de Séries Temporais: Os Modelos ARIMA, SEDIGRAF**, 1996

SOUZA, Francisca Mendonça; ALMEIDA, Silvana Gonçalves; SOUZA, Adriano Mendonça; LOPES, Luis Felipe Dias; ZANINI, Roselaine Ruviaro; **Previsão do preço da gasolina para a região sul do Brasil**. IJIE – Ibero american Journal of industrial Engineering. v.3, n.1, pg 234-248, julho, 2011.