

# COMPARATIVO DE ALGUNS MODELOS DE PREVISÃO PARA O TRANSPORTE DE GRÃOS DA REGIÃO OESTE DO PA- RANÁ ATRAVÉS DO MODAL FERROVI- ÁRIO PARA OS ANOS DE 1997 A 2001

**Cicero Aparecido Bezerra\***

**Juvenile Wolf\*\***

**Robert Wayne Samohyl\*\*\***

RESUMO: O Brasil possui posição de destaque mundial na produção de grãos. Em relação aos principais grãos como soja, milho e trigo, o país ocupa segundo e terceiro lugares respectivamente. Neste contexto, o Estado do Paraná é responsável por 22% do total da produção nacional, sendo que 35% vem da região oeste. Para o escoamento desta produção até o porto de Paranaguá, aproximadamente 70% do transporte é realizado através do modal rodoviário e o restante, ferroviário. Apesar das vantagens deste último modal, existem várias dificuldades enfrentadas pelos produtores que utilizam este meio, dentre elas, a falta de vagões no pico da safra. Este problema pode estar relacionado com inadequadas estimativas na demanda futura por este serviço. Desta forma, o presente estudo comparou o desempenho de três modelos de previsão (Holt-Winters, regressão múltipla e ARIMA) aplicados à problemática apresentada.

---

\* Graduado em Informática pela Universidade do Vale do Rio dos Sinos - UNISINOS; Mestre em Engenharia de Produção pela Universidade Federal de Santa Catarina - UFSC; Doutor em Engenharia de Produção pela Universidade Federal de Santa Catarina - UFSC; Docente Adjunto nível I da Universidade Federal do Paraná - UFPR. E-mail: cicero.bezerra@ufpr.br

\*\* Mestrando em Engenharia de Produção pela Universidade Federal de Santa Catarina - UFSC. E-mail: jwolf@hotmail.com

\*\*\* Graduado em Economia pela University of Houston System; Mestre em Economia pela Rice University; Doutor em Economia pela Rice University; Pós-doutor pela Virginia Tech; Pós-doutor pela University Of Illinois At Urbana-Champaign; Pós-doutor pela University Of Texas At Austin; Docente Titular da Universidade Federal de Santa Catarina - UFSC. E-mail: samohyl@deps.ufsc.br

Os resultados obtidos mostram que, na medida em que os dados referentes ao transporte de grãos através do modal ferroviário se estabilizam, os modelos apresentam uma capacidade preditiva mais acurada, em especial o modelo Holt-Winters.

**PALAVRAS-CHAVE:** Transporte de Grãos; Modal Ferroviário; Modelos de Previsão de Demanda.

## **COMPARATIVE PREVIEW OF SOME MODELS FOR THE TRANSPORTATION OF GRAIN FROM WEST OF PARANA THROUGH THE RAIL MODAL BE- TWEEN THE YEARS 1997 TO 2001**

**ABSTRACT:** Brazil has a prominent position in world grain production. With respect to the main grain like soybean, corn and wheat, the country occupies the second and third places respectively. In this context, the state of Paraná is responsible for 22% of the total of the national production, being that, 35% comes from the west region. For disposal of this production in the Paranaguá Harbour, approximately 70% of the transportation is achieved through the road modal and the remainder by railroad. Despite the advantages of the latter modal, there are many problems faced by producers who use it, among them the lack of wagons at the peak of the season. This problem may be related to the inadequate estimates on future demand for this service. Thus, this study compared the performance of three forecasting models (Holt-Winters, multiple regression and ARIMA) applied to the problem presented. The results show that, in those that the data for the grain transportation by rail modal stabilized it, the models provide a more accurate predictive capability, especially the Holt-Winters model.

**KEYWORDS:** Grain Transportation; Railway Modal; Forecasting Models.

## INTRODUÇÃO

O Brasil possui posição de destaque na produção mundial de grãos. No caso de soja e milho, chega a ocupar segundos e terceiro lugares, respectivamente, segundo o Atlas Econômico do Rio Grande do Sul (RIO GRANDE DO SUL, 2010). Do total nacional produzido, o Estado do Paraná é responsável por 21,78% (em média), sendo que a região oeste participa com 35% do total, de acordo com Andretta (2001). O escoamento desta produção em direção ao porto de Paranaguá se dá através de duas modalidades de transporte: rodoviário (responsável por aproximadamente 70%) e ferroviário.

Apesar do modal ferroviário apresentar algumas vantagens com relação ao rodoviário, entre os vários problemas que dificultam o transporte nesta modalidade, pode-se identificar a falta de vagões no pico da safra. Este problema pode estar relacionado diretamente às adequadas estimativas da demanda do transporte com suficiente antecipação para a tomada de decisão de forma mais segura. Neste sentido, estudos envolvendo modelos de previsão, aplicados à problemática descrita, podem auxiliar no dimensionamento adequado do número de vagões.

Este, portanto, é o objetivo do presente trabalho: comparar os resultados obtidos para a previsão da demanda de transporte de grãos da região oeste do Paraná, através do modal ferroviário entre os anos de 1997 e 2001, fornecidos pelos modelos Holt-Winters, regressão múltipla e ARIMA. Para atingir o objetivo proposto, o artigo estará estruturado em seções que tratam, através de um referencial teórico, da produção e transporte de grãos da região oeste do Paraná, bem como dos modelos de previsão abordados e suas medidas de precisão. Em seguida serão descritos os procedimentos metodológicos utilizados, a análise e discussão dos resultados obtidos com a aplica-

## 2 REFERENCIAL TEÓRICO

Este tópico apresenta a pesquisa bibliográfica no que se refere aos temas abordados no presente artigo: dados a respeito do transporte de grãos do oeste do Paraná, através do modal ferroviário, entre os anos de 1997 e 2001 e, alguns modelos de previsão de demanda. É importante salientar que não se pretende esgotar o assunto, mas sim identificar os principais pontos para a contextualização dos temas.

### 2.1 PRODUÇÃO E TRANSPORTE DE GRÃOS NO OESTE DO PARANÁ

De acordo com a Secretaria da Agricultura e do Abastecimento do Paraná (PARANÁ, 2002), a produção dos principais grãos na safra 2001, no oeste do Paraná foi assim dividida, em relação ao Brasil: milho (30,25%), soja (23,05%) e trigo (51,05%). De toda a produção de grãos do Paraná, a região oeste contribui com 35% deste total, segundo a Estrada de Ferro Oeste Paraná S.A. (2002), onde aproximadamente 71,69% do total produzido pela região em 2001 foram enviadas ao porto de Paranaguá.

Apesar de ser responsável por 67% do escoamento dos grãos até o porto de Paranaguá, o transporte rodoviário vem apresentando alguns problemas que influenciam diretamente na competitividade dos produtos. O uso de caminhões pesados levanta questões ambientais, riscos de acidentes, além da perda de grãos e o custo dos pedágios, chegando a 47 milhões de reais (equivalente a 65,6 mil hectares de soja), conforme Bezerra (2000). Já o modal ferroviário apresenta problemas menores. O custo por tonelada transportada por 1.000 quilômetros é de US\$ 20,00 (contra US\$ 50,00 no rodoviário). Outros fatores favorecem o modal ferroviário: menor tráfego, mais toneladas transportadas em uma viagem, menor perda de grãos, menor risco de acidentes. Apesar das vantagens citadas, constata-se a

precariedade da infra-estrutura dos serviços e, principalmente, falta de vagões para atender à demanda no pico da safra, segundo Bezerra (2000).

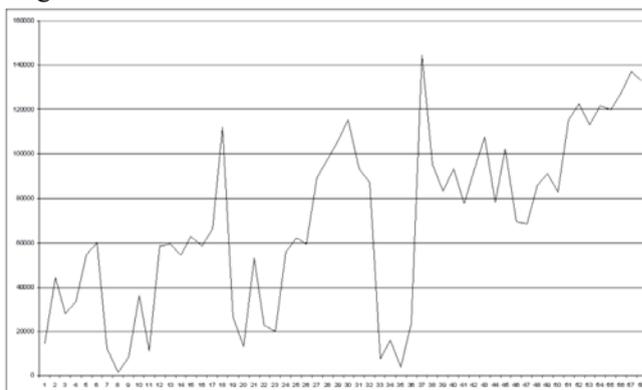
Entre os períodos compreendidos de março de 1997 a dezembro de 2001, o modal ferroviário transportou a seguinte quantidade de grãos (valores fornecidos em toneladas, já transformados para meses de 30 dias), conforme apresentado através da Tabela 1:

**Tabela 1** Escoamento de grãos – modal ferroviário

	1997	1998	1999	2000	2001
<b>JAN</b>	-	11 216	20 012	3 972	68 295
<b>FEV</b>	-	58 597	55 606	24 011	85 925
<b>MAR</b>	14 513	59 397	62 360	144 568	90 961
<b>ABR</b>	44 220	54 181	59 469	95 250	82 737
<b>MAI</b>	27 895	62 892	89 579	83 275	115 481
<b>JUN</b>	33 247	58 486	97 354	93 193	122 811
<b>JUL</b>	54 498	66 064	105 841	77 616	113 066
<b>AGO</b>	59 981	112 018	115 442	92 627	121 901
<b>SET</b>	12 312	26 426	93 300	107 467	119 820
<b>OUT</b>	1 247	13 233	87 170	78 278	127 866
<b>NOV</b>	8 420	53 033	7 590	102 100	137 361
<b>DEZ</b>	36 223	22 751	15 959	69 611	132 936

Fonte: adaptado de Estrada de Ferro Oeste Paraná S.A. (2002)

O comportamento dos dados ao longo dos anos pode ser observado na Figura 1:



**Figura 1** Escoamento de grãos – modal ferroviário

Fonte: adaptado de Estrada de Ferro Oeste Paraná S.A. (2002)

É possível verificar que houve um aumento médio no transporte de grãos, no período compreendido entre 1997 e 2001, da ordem de 275,75%, sendo que o maior aumento ocorreu no biênio 2000/2001, com um percentual de 35,72%. A média mensal é de aproximadamente 68.821 toneladas, com desvio padrão de 5.209 toneladas, sendo que os anos de 1999 e 2000 apresentaram a maior variação mensal (desvio padrão de 10.636 e 10.662, respectivamente). Já os anos de 1997 e 2001 apresentaram a menor variação (6.319 e 6.431 toneladas, respectivamente).

### 2.3 PREVISÃO DE DEMANDA A PARTIR DA ANÁLISE DE SÉRIES TEMPORAIS

Com a dinâmica de mercado cada vez mais acirrada, a utilização de modelos de previsão nas mais diversas áreas tem conquistado mais espaço na tomada de decisões. A tendência que, no passado, era basicamente intuitiva, é o aumento da utilização de procedimentos sistematizados de previsão, entre eles, aqueles que envolvem a análise de séries temporais (JAIN, 2003).

O primeiro passo para este procedimento, conforme proposto por Libonati e colaboradores (2004), é a análise gráfica, visto sua capacidade em determinar padrões presentes na série. Neste sentido, Samohyl e Rocha (2001) indicam a decomposição clássica (em suas formas multiplicativa ou aditiva), na qual um determinado conjunto de dados é formado pelo produto (ou soma) de padrões referentes à sazonalidade ( $S_t$ ), tendência ( $T_t$ ) e resíduos ( $E_t$ ), conforme equações a seguir:

$$Y_t = S_t \times T_t \times E_t \quad (1)$$

$$Y_t = S_t + T_t + E_t \quad (2)$$

A decomposição multiplicativa é utilizada naquele conjunto de dados que apresenta comportamento exponencial e a aditiva, linear. Segundo Makridakis, Wheelwright e Hyndman (1998), ambas as

equações partem da obtenção da tendência através da média móvel centrada e, a partir daí, removendo-se estes valores dos dados observados, estima-se o componente sazonal através da média de cada observação mensal, assumindo que sejam constantes de ano para ano. Finalmente, subtraindo-se tanto a tendência, quanto a sazonalidade, da série temporal, obtêm-se os resíduos. Desta forma, em função dos padrões encontrados, é possível escolher os modelos de previsão de demanda mais adequados aos dados fornecidos.

Ainda que uma enorme gama de métodos de previsão possa ser empregada (entre elas, as previsões qualitativas), este estudo optou por verificar a aderência dos modelos Holt-Winters, ARIMA e regressão múltipla – uma vez que, preliminarmente, parecem serem mais adequados ao comportamento da série temporal apresentada.

### 2.1.1 Modelo Holt Winters

Em 1960, Peter R. Winters estendeu o modelo linear de Charles C. Holt para capturar a sazonalidade, baseado em 3 equações (multiplicativa ou aditiva) – uma para o nível, uma para a tendência e outra para a sazonalidade – aplicadas de acordo com o comportamento da série. O modelo multiplicativo é composto das seguintes equações (CARVALHO; LOIOLA; COELHO, 2001):

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (3)$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \quad (4)$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s} \quad (5)$$

$$F_{t+m} = (L_t + b_t m)S_{t-s+m} \quad (6)$$

Já o modelo aditivo é expresso a partir das seguintes equações (YAFEE; MCGEE, 2000):

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (7)$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \quad (8)$$

$$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s} \quad (9)$$

$$F_{t+m} = L_t + b_{t-m} + S_{t-s+m} \quad (10)$$

Onde, para ambos os modelos,  $s$  representa o período sazonal,  $L_t$  representa o nível da série,  $b_t$  denota a tendência,  $S_t$  é o componente sazonal e,  $F_{t+m}$  corresponde à previsão para  $m$  períodos adiante (CARVALHO; LOIOLA; COELHO, 2001).

Estes modelos utilizam as constantes de amortecimento  $a$ ,  $b$  e  $g$ . Segundo Makridakis, Wheelwright e Hyndman (1998), estes valores variam entre 0 e 1 e são fixados pelo analista. Para o  $a$ , quanto maior seu valor, mais rápida será a reação do modelo a uma variação real dos dados observados. O mesmo se aplica ao  $b$ , porém relacionado à tendência da série e ao  $g$  (relacionado à flutuação sazonal). Os modelos Holt-Winters indicam que determinado conjunto de dados pode ser extrapolado  $m$  períodos adiante a partir do momento em que se obtêm a tendência, a sazonalidade e o nível presente no conjunto de dados.

### 2.1.2 Modelo ARIMA

O modelo baseado na metodologia desenvolvida por George E. P. Box e Gwilyn M. Jenkins, apresentada em 1970, consiste na explicação de uma variável através de seus valores passados e de seus termos de erros estocásticos (SARTORIS, 2003; GUJARATI, 2006). Este modelo pode resultar da combinação de três componentes: auto-regressivo (AR), integrado (I) e de médias móveis (MA). O modelo autoregressivo (AR) contempla  $p$  edições de correlação entre as observações sucessivas da série (SILVA; SAMOHYL;

COSTA, 2002), sendo expresso por:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t \quad (11)$$

Indicando que, a partir de um dado observado ( $Y_t$ ) é possível estimar os parâmetros do modelo estimado ( $\varphi_p$ ), avaliando-se o comportamento dos resíduos ( $e_t$ ).

Já o modelo de médias móveis (MA) pode ser expresso pela seguinte equação:

$$Y_t = \mu - \varphi_1 e_{t-1} - \varphi_2 e_{t-2} - \dots - \varphi_q e_{t-q} + e_t \quad (12)$$

Onde, de acordo com Silva, Samohyl e Costa (2002),  $\mu$  indica o nível do processo,  $\varphi_{t-i}$  são os termos de erros correlacionados e,  $e_t$  é o termo de erro aleatório.

Por sua vez, tomando-se as  $d$ -ésimas diferenças entre os valores observados ( $\Delta^d Y_t$ ), obtém-se um modelo integrado (I) de ordem  $d$ , através de um processo autoregressivo (AR) de ordem  $p$  e de médias móveis (MA) de ordem  $q$ , expresso pela seguinte equação (SARTORIS, 2003):

$$\Delta^d Y_t = \varphi_1 \Delta^d Y_{t-1} + \varphi_2 \Delta^d Y_{t-2} + \dots + \varphi_p \Delta^d Y_{t-p} + e_t - \varphi_1 e_{t-1} - \varphi_2 e_{t-2} - \dots - \varphi_q e_{t-q} \quad (13)$$

Como o modelo pode apresentar muitas combinações de equações válidas para o problema, um critério que seleciona as mais adequadas, de maneira satisfatória (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998), é o *Akaike's Information Criterion* (AIC), calculado pela equação a seguir:

$$AIC \approx n(1 + \log(2\pi)) + n \log s^2 + 2m \quad (14)$$

Onde  $s^2$  é a variância dos resíduos,  $n$  é o número de observações da série e  $m$  é o número de termos estimados no modelo ( $p+q+P+Q$ ).

Segundo Matos (2000), o foco principal do modelo ARIMA é a previsão de curso prazo.

### 2.1.3 Regressão Múltipla

O modelo de regressão múltipla pressupõe a presença de variáveis que explicam o comportamento da série analisada e que, portanto, devem ser incluídas no modelo (HILL; GRIFFITHS; JUDGE, 2003). A regressão múltipla pode ser expressa através da equação abaixo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_k + e_i \quad (15)$$

Onde  $Y_t$  é o valor a ser previsto,  $b$  é o coeficiente angular das variáveis ( $X$ ) que explicam o valor a ser previsto e  $e$ , o termo do erro. Antes de se proceder à determinação dos coeficientes angulares é importante verificar o quanto as variáveis  $X$  explicam a variável  $Y$ . Este cálculo é obtido através do coeficiente de correlação, conforme equação a seguir:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}} \quad (16)$$

Onde  $n$  é o número de observações. Valores próximos a 1 indicam que existe uma relação linear positiva perfeita entre as variáveis. Valores próximos a -1 indicam uma relação linear negativa perfeita, ou seja, para cada unidade que é aumentada em  $X$ , uma unidade equivalente é decrescida em  $Y$ . Valores próximos a 0 indicam que não existe relação entre as variáveis (REID; SANDERS, 2005).

Segundo Makridakis, Wheelwright e Hyndman (1998), é possível aplicar duas técnicas na regressão múltipla. A primeira consiste em analisar apenas as variáveis incluídas no modelo e na segunda técnica assume-se a influência que determinado atributo possui em relação a variável a ser explicada. Por exemplo: presumindo-se que

o componente sazonal permanece inalterado a cada ano, é possível incluir no modelo uma nova variável explanatória (*dummy*), para cada mês, possuindo somente valores 0 e 1 (1 se o mês for janeiro, caso contrário 0; 1 se o mês for fevereiro, caso contrário 0 e assim sucessivamente) para testar o impacto de cada mês no valor a ser previsto.

## 2.2 Medidas de precisão

Para determinar a precisão dos modelos, autores como Reid e Sanders (2005), Bezerra (2006), Martins e Laugeni (2006) e Gonçalves (2007) indicam o uso de medidas padrões obtidas pela diferença entre os valores observados e previstos, conforme definido pela equação a seguir:

$$e_t = Y_t - F_t \quad (17)$$

Onde  $e_t$  é o erro no período  $t$ ,  $Y_t$  é o dado observado e  $F_t$  é a previsão para o período  $t$ . Se existirem observações e previsões para  $n$  períodos, é possível determinar medidas estatísticas padrão, para o conjunto de períodos:

$$ME = \frac{1}{n} \sum_{t=1}^n e_t \quad (18)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (19)$$

O valor fornecido pelo erro médio (18) tende a ser pequeno, uma vez que erros positivos e negativos encontrados durante os períodos tendem a se anularem. Seu mérito é informar se a previsão foi sistematicamente acima ou abaixo do observado. O erro médio absoluto torna os erros positivos para então calcular a média, o que parece fornecer uma informação mais precisa quanto à sua amplitude. Estas estatísticas lidam com medidas de precisão cujo tamanho depende da

escala dos dados, porém não facilitam a comparação entre diferentes séries temporais e diferentes intervalos de tempo (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998). Tal situação é contornada pelas medidas relativas de erros, a partir das seguintes equações:

$$MPE = \frac{1}{n} \sum_{t=1}^n \left( \frac{Y_t - F_t}{Y_t} \right) \quad (20)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - F_t}{Y_t} \right| \quad (21)$$

Porém, da mesma forma que o erro médio (18), o percentual de erro médio (20) tende a ser pequeno. Já o percentual de erro médio absoluto – *MAPE* (21) é significativo somente se a escala possuir uma origem significativa.

Para a comparação entre os métodos, Makridakis, Wheelwright e Hyndman (1998) propõem a estatística U de Theil, levando em consideração os valores previstos para as próximas<sup>2</sup> ocorrências, comparadas com os valores que efetivamente ocorrerem. A equação a seguir representa a estatística U de Theil:

$$U^2 = \frac{\sum_{t=1}^{n-1} \left[ \left( \frac{Y_{t+1} - Y_t}{Y_t} \right)^2 + \left( \frac{F_{t+1} - Y_t}{Y_t} \right)^2 \right]}{\sum_{t=1}^{n-1} \left( \frac{Y_{t+1} - Y_t}{Y_t} \right)^2} \quad (22)$$

Sendo  $Y_t$  o valor observado e  $F_t$  o valor previsto no momento  $t$ . O coeficiente (neste caso, decimal) gerado indica a desigualdade entre os valores observados e previstos. O valor de U será igual a 0, quando a previsão for exatamente igual à observação. Valores maiores ou iguais a 1 indicam que um método ingênuo de previsão é tão preciso quanto a técnica de previsão utilizada (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998).

### 3 PROCEDIMENTOS METODOLÓGICOS

De maneira geral, os procedimentos metodológicos empregados serão aqueles propostos por autores como Pellegrini e Fogliatto (2000) e Samohyl e Rocha (2001) para estudos desta natureza, a saber:

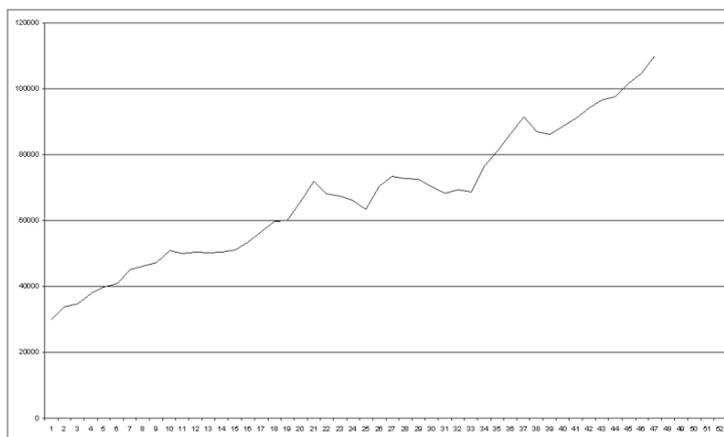
- Definição do problema: já apresentada na introdução deste trabalho. Envolve a compreensão do contexto no qual a previsão será utilizada, bem como a disponibilidade dos dados.
- Coleta de informações: apesar de existirem, pelos menos, dois tipos de informações disponíveis (qualitativas e quantitativas), em função do objetivo deste estudo optou-se por coletar somente os dados estatísticos históricos do transporte de grãos – já abordado na seção 2.1.
- Análises preliminares: a partir do comportamento gráfico do transporte de grãos será utilizada a decomposição clássica para a descoberta de padrões, com a finalidade de distinguir modelos de previsão mais compatíveis com os dados.
- Ajuste dos modelos: visto que cada modelo envolve um ou mais parâmetros que devem ser estimados, a partir dos dados históricos conhecidos será utilizado o software Microsoft Excel 2003 ® para a determinação destes elementos.
- Avaliação do modelo: tendo sido o modelo selecionado criteriosamente e os parâmetros estimados apropriadamente, o comportamento do modelo será avaliado através de seu MAPE.

Finalmente, uma vez que, segundo Makridakis, Wheelwright

e Hyndman (1998) e Pellegrini e Fogliatto (2000), um modelo só poderá ser considerado válido após os dados para o período previsto estarem disponíveis para a comparação, as últimas quatro observações serão reservadas e comparadas com as previsões formadas a partir do período 54, supondo que estes dados ainda não ocorreram. Desta forma, será possível avaliar o poder preditivo dos modelos, bem como compará-los entre si utilizando a estatística U de Theil.

#### 4 ANÁLISE E DISCUSSÃO

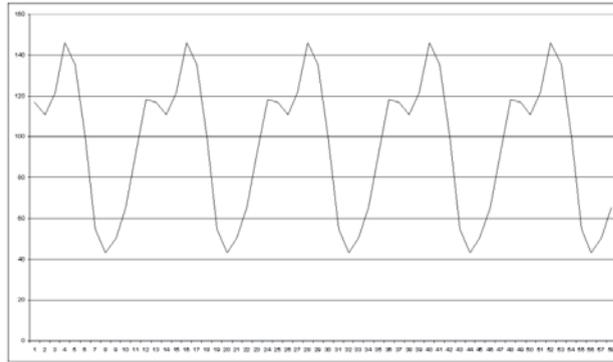
A análise inicia-se com a procura por padrões nos dados observados. Neste sentido optou-se pela decomposição clássica aditiva, visto o comportamento linear dos dados (conforme Figura 1, apresentada na seção 2.1). O primeiro passo é a determinação da tendência, através da utilização da média móvel centrada de 12 períodos, cujo resultado pode ser visualizado na Figura 2:



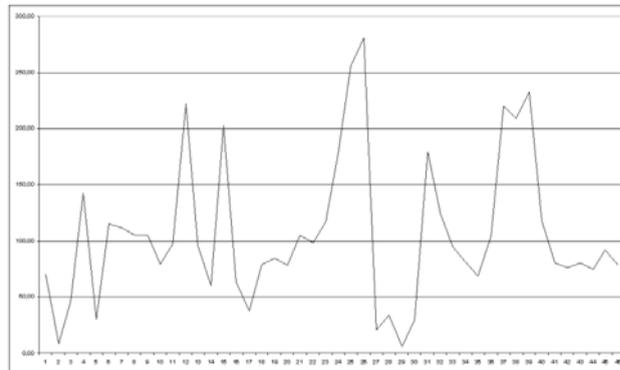
**Figura 2** Tendência

Removendo-se o componente que determina a tendência, dos dados observados, obtém-se a sazonalidade, conforme observado na

Figura 3:

**Figura 3** Sazonalidade

Finalmente, subtraindo-se tanto a tendência, quanto a sazonalidade, dos dados observados, extraem-se os resíduos, ou série irregular, conforme a Figura 4:

**Figura 4** Resíduos

Através da decomposição clássica dos dados observados, pode-se concluir que os mesmos apresentam tendência e sazonalidade bem definidos, o que determina o uso de modelos de previsão que

possam incorporar estas características.

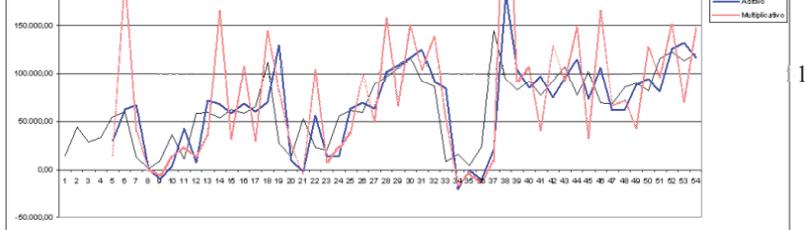
#### 4.1 HOLT-WINTERS

O modelo Holt-Winters comporta a presença de tendência e sazonalidade, como é o caso apresentado. Desta forma, para um período sazonal de 4 meses, adotando-se o critério de minimização dos erros médios percentuais absolutos (*MAPE*) e utilizando o suplemento Solver, da planilha Microsoft Excel 2003 ®, para a determinação das constantes de amortecimento, obteve-se os valores observados na Tabela 2:

**Tabela 2** Constantes de amortecimento

	$\alpha$	$\beta$	$\gamma$	MAPE
Multiplicativo	1	0,3322	1	96,51
Aditivo	0,9314	0,3676	1	83,07

Graficamente, os modelos Holt-Winters aditivo e multiplicativo se comportaram conforme visualizado na Figura 5:



**Figura 5** Holt-Winters

Em virtude de elevados *MAPE*'s, o modelo Holt-Winters não se mostrou adequado para os dados apresentados nas 54 primeiras observações, ainda que o modelo aditivo tenha apresentado um melhor desempenho sobre o multiplicativo.

#### 4.2 REGRESSÃO MÚLTIPLA

Para a análise de regressão, levaram-se em consideração que as variáveis que mais impactam no transporte de grãos através do modal ferroviário são os valores do dólar (venda), soja, milho, trigo e petróleo. Desta forma, analisou-se, utilizando-se o suplemento Análise de Dados da planilha Microsoft Excel 2003®, a (1) relação de cada variável isolada em relação ao transporte, (2) todas as variáveis e (3) todas as variáveis mais o impacto de cada mês (com o uso de variáveis *dummy*). Os valores obtidos podem ser visualizados na Tabela 3:

**Tabela 3** Coeficientes de correlação

Dólar	Soja	Milho	Trigo	Petróleo	Todas	Dummy
0,6152	0,2422	0,2101	0,6870	0,4215	0,7715	0,8356

Os valores indicam que, isoladamente, cada variável possui pouco impacto no transporte de grãos (vale ressaltar, entretanto, os valores do dólar e do trigo que se destacam). Porém, juntas a relação se mostra mais consolidada. O mesmo suplemento mostra, na Tabela 4 (pág. 52), os valores de cada coeficiente angular (neste caso, optou-se por mostrar somente aquelas que obtiveram os dois maiores coeficientes de correlação).

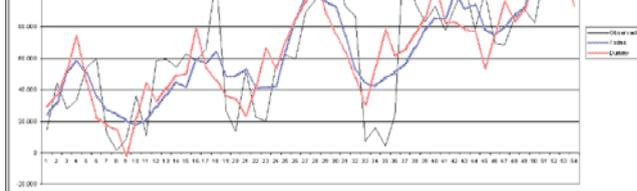
Usando o *MAPE* como critério para avaliação do melhor modelo

de regressão, obteve-se 105,007% para o modelo que utiliza os valores do dólar, soja, milho, trigo e petróleo e 107,858% para o modelo que, além das anteriores, incorpora as variáveis *dummy*.

**Tabela 4** Coeficientes angulares

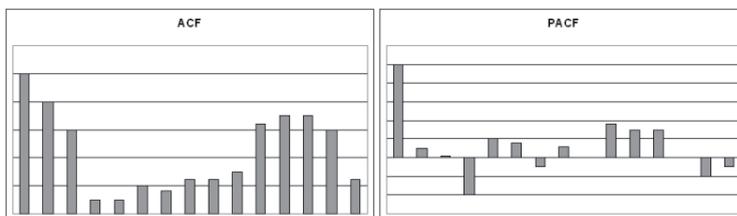
VARIÁVEIS	TODAS	DUMMY
Intercepto	4699,0632	-4842,9125
Dólar	-4106,2912	10134,769
Soja	-396903,19	-294430,06
Milho	-268388,57	-106713,88
Trigo	1116255,3	759835,43
Petróleo	171,71968	145,49051
dummy1	-	18877,671
dummy 2	-	1871,907
dummy 3	-	4085,1176
dummy 4	-	6106,0405
dummy 5	-	8222,4914
dummy 6	-	25567,354
dummy 7	-	998,33244
dummy 8	-	-10612,252
dummy 9	-	-9627,4396
dummy 10	-	-12328,096
dummy 11	-	-26139,346

Desta forma, na Figura 6, a seguir, pode-se visualizar o resultado da análise de regressão nas primeiras 54 observações:



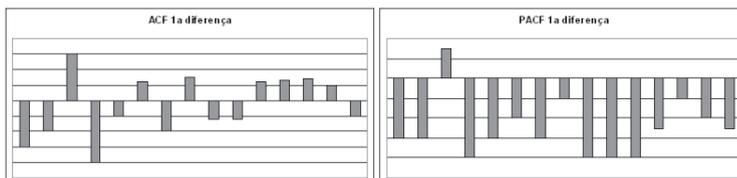
### 4.3 ARIMA

Inicialmente, em virtude da presença da sazonalidade nos dados observados, o modelo ARIMA assume a notação  $ARIMA(p, d, q)$  ( $P, D, Q$ ). O primeiro passo no processo de identificação do modelo ARIMA é a análise do gráfico das observações examinando a existência de estacionaridade na média e na variância da série, através da função de autocorrelação (ACF), que indica padrões remanescentes nos erros, e da função de parcial de autocorrelação (PACF), que determina o grau de associação entre os valores previstos quando os efeitos de outros períodos de tempo são removidos. A Figura 7 mostra estes gráficos:



**Figura 7** Estacionaridade e variância

Pelas observações, verifica-se que a média, aparentemente, não apresenta estacionaridade. Desta forma torna-se necessário aplicar a primeira diferença para observar a existência de outras possíveis correlações na série. O efeito da primeira diferença é observado na Figura 8:



**Figura 8** 1<sup>as</sup> diferenças na estacionaridade e variância

Na aplicação da primeira diferença pode se observar que o ACF e o PACF não apresentam valores significativos nas primeiras defasagens. Existe apenas um valor significativo na quarta defasagem dos dois gráficos. Isto sugere um modelo  $ARIMA(0,1,0)(P,0,Q)4$  ou uma combinação de um valor significativo na quarta defasagem e uma possível sazonalidade de 12 meses.

O gráfico do ACF dos dados observados não captura a não-estacionaridade da média, mas apresenta três valores significantes nas primeiras defasagens e um valor na primeira defasagem do PACF sugerindo modelos  $ARIMA(p,0,q)$  onde  $p$  pode ser 0 ou 1 e  $q$  0, 1, 2 ou 3. O valor na 12<sup>a</sup> defasagem do ACF pode indicar uma sazonalidade de 12 meses. Por outro lado, o PACF apresenta um valor significativo na quarta defasagem e assim é possível determinar a sazonalidade examinando os seguintes modelos:  $ARIMA(p,0,q)(P,0,Q)12$  e  $ARIMA(p,0,q)(P,0,Q)4$  onde  $p$  pode ser variado até 4 e  $q$  até 3.

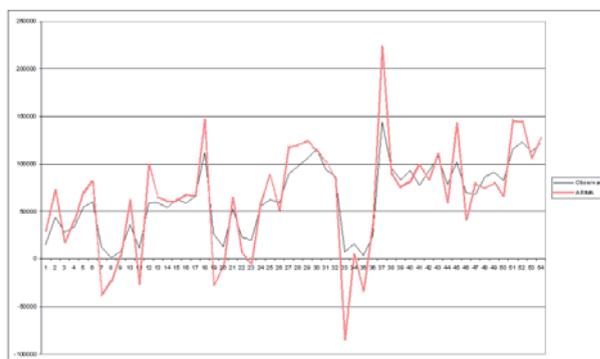
Na tentativa de encontrar um modelo apropriado pode-se testar um número limitado fazendo pequenas variações nos parâmetros  $p$ ,  $q$ ,  $P$  e  $Q$ . Desta forma, utilizando o suplemento Solver, foi possível encontrar vários modelos aparentemente viáveis. Faz-se necessário, portanto, aplicar critérios adicionais para escolher o melhor. O primeiro critério é a eliminação daqueles que contêm parâmetros não significativos. Assim sendo, sobraram apenas os seguintes para uma análise mais apurada:  $ARIMA(1,0,0)$ ,  $ARIMA(0,0,3)$ ,  $ARIMA(0,0,3)(0,0,1)12$ ,  $ARIMA(1,0,0)(1,0,0)4$ ,  $ARIMA(0,1,0)(1,0,0)4$  e  $ARIMA(4,1,3)(1,0,0)12$ . Para os modelos remanescentes foi calculado o  $AIC$  de cada um, com o intuito de estimar o mais adequado, conforme observado na Tabela 5 (pág. 55).

Como uma diferença de até 2 no valor de  $AIC$  não é considerada substancial, opta-se pelo  $ARIMA(1,0,0)(1,0,0)4$ . Assim sendo, é possível visualizar o desempenho do modelo escolhido através da

Figura 9 (pág. 55).

**Tabela 5** AIC's calculados

MODELO	m	$\Sigma$	AIC
ARIMA(1,0,0)	1	30354,84	583,18
ARIMA(0,0,3)	3	29739,72	586,22
ARIMA(0,0,3)(0,0,1)2	4	27998,25	585,39
ARIMA(1,0,0)(1,0,0)4	2	29441,38	583,74
ARIMA(0,1,0)(1,0,0)4	1	30099,58	582,78
ARIMA(4,1,3)(1,0,0)2	8	24406,00	586,95



**Figura 9** ARIMA

Embora o *AIC* tenha fornecido um valor significativo para a escolha do modelo ARIMA(1,0,0)(1,0,0)4, seu *MAPE*, para as primeiras 54 observações, foi elevado, apresentado um valor de 122,793%. Apesar disso, é possível perceber que este modelo conseguiu capturar os picos e vales do transporte de grãos, ainda que, com este comportamento, os valores obtidos fossem além dos valores observados.

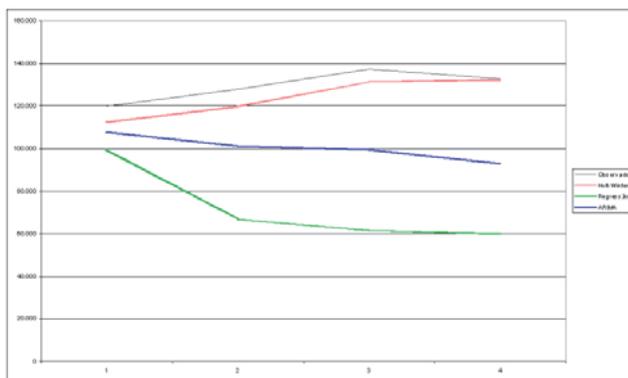
#### 4.3 COMPARAÇÃO ENTRE MODELOS

Apesar do fraco desempenho dos modelos (com *MAPE's* acima de 80%), determinaram-se as previsões, baseadas no período 54, comparando-as com os dados reservados (períodos 55 a 58) e verificando os resultados entre os modelos abordados através do *MAPE* e *U* de Theil. Desta forma, a Tabela 5 apresenta os valores previstos (arredondados), os *MAPE's* e os *U's* de Theil para os próximos 4 períodos:

**Tabela 6** Comparação entre modelos

Observado	Holt-Winters			Regressão Múltipla			ARIMA		
	Previsto	MAPE	U-Theil	Previsto	MAPE	U-Theil	Previsto	MAPE	U-Theil
119.820	112.355	6,23	0,998	99.564	16,91	7,576	107.951	9,91	3,301
127.866	119.837	6,29	0,623	66.906	47,68	7,979	101.305	20,77	3,976
137.361	131.450	4,30	0,097	61.598	55,16	16,468	99.612	27,48	9,062
132.936	132.507	0,32	-	60.063	54,82	-	92.837	30,16	-

Visualmente, os modelos se comportaram conforme mostrado na Figura 10:



**Figura 10** Comparação entre os modelos

É possível perceber que, para os dados reservados, os modelos tiveram um comportamento mais adequado em relação aos 54 primeiros valores observados. Na média, o modelo Holt-Winters apresentou maior precisão, obtendo um *MAPE* médio de 4,4836% e um

U de Theil médio de 0,5724, contra 43,6385% e 10,6745, respectivamente, para o modelo de regressão múltipla e 22,0809% e 5,4462 para o modelo ARIMA.

## 5 CONSIDERAÇÕES FINAIS

O comportamento dos dados, ao longo dos períodos observados, indica a ocorrência de fenômenos que não foram captados pelos modelos, principalmente nos anos de 1999 e 2000. Porém observa-se que a série tende a se estabilizar a partir de dezembro de 2000, o que permitiu a obtenção de melhores resultados a partir deste período, explicando a proximidade das previsões aos valores observados, para todos os modelos, em especial o da categoria Holt-Winters.

Cada modelo utilizado apresentou características distintas. O modelo Holt-Winters, por exemplo, ignora fatores que podem influenciar a série. A regressão múltipla, apesar de ser bastante abrangente em relação à explicação da série, justamente por este motivo requer um número muito grande de dados que possam ser associados ao fenômeno estudado. Além disso, informa a correlação entre variáveis dependentes e independentes, mas não se existe esta casualidade. Finalmente, o modelo ARIMA demanda o uso de bom senso por parte do analista, visto o grau de complexidade na sua implementação e interpretação.

Finalmente, é importante notar que a precisão não é o único critério para a confiabilidade de quaisquer modelos. Segundo Makridakis, Wheelwright e Hyndman (1998), o estímulo para a ação na organização é que irá determinar o sucesso da previsão. No caso apresentado, os processos de estimativa da demanda podem contribuir com a minimização da falta de vagões para o transporte de grãos da região oeste do Paraná.

## REFERÊNCIAS

ANDRETTA, G. C. **Valor bruto da produção agropecuária**. Curitiba, PR: SEAB-DERAL-DEB, 2001.

RIO GRANDE DO SUL. **Atlas Socioeconômico do Rio Grande do Sul**. Agricultura. Porto Alegre, RS: DEPLAN, [S. d.]. Edição Eletrônica. Disponível em <<http://www.scp.rs.gov.br/atlas/atlas.asp?menu=31#>>. Acesso em: 19 maio 2010.

BEZERRA, C. A. Evaluation of Holt-Winters models in the solid residua forecasting: a case study in the city of Toledo – PR. In: INTERNATIONAL CONFERENCE ON PRODUCTION RESEARCH – AMERICAS’ REGION, 3, 2006, Curitiba. **Anais...** Porto Alegre, RS: Abepro, 2006. 1 Cd-Rom.

BEZERRA, J. A. Aperto na estrada. **Globo Rural**, ano 15, n. 176, p. 60-66. 2000.

CARVALHO, L. M. L.; LOIOLA, E. M.; COELHO, P. S. S. Métodos quantitativos de previsão aplicados ao caso de uma seguradora nacional. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 23, 2001, Viçosa. **Anais...** Rio de Janeiro, RJ: Sobrapo, 2001. 1 Cd-Rom.

ESTRADA DE FERRO OESTE PARANÁ S.A. **Área de influência**. Curitiba, PR: Ferroeste, 2002. Relatório técnico.

GONÇALVES, F. **Excel avançado 2003/2007 Forecast**: análise e previsão de demanda. Rio de Janeiro, RJ: Ciência Moderna, 2007.

GUJARATI, D. N. **Econometria básica**. Rio de Janeiro, RJ: Campus, 2006.

HILL, R. C.; GRIFFITHS, W. E.; JUDGE, G. G. **Econometria**. 2.

ed. São Paulo, SP: Saraiva, 2003.

JAIN, C. L. Business forecasting in the 21st century. **Journal of Business Forecasting Methods & Systems**, Chicago, v. 22, n. 3, p. 3-6, 2003.

LIBONATI, J. J. et al. Séries temporais. In: CORRAR, L. J.; THEÓFILO, C. R. (Coord.). **Pesquisa operacional para decisão em contabilidade e administração**. São Paulo, SP: Atlas, 2004.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. J. **Forecasting methods and applications**. 3. ed. New York: John Wiley & Sons, 1998.

MARTINS, P. G.; LAUGENI, F. P. **Administração da produção**. 2. ed. São Paulo, SP: Saraiva, 2006.

MATOS, O. C. **Econometria básica**. 3. ed. São Paulo, SP: Atlas, 2000.

PARANÁ. Secretaria da Agricultura e do Abastecimento. **Principais culturas: Comparativo de área e produção – Paraná / Brasil e ranking – safra 2001**. Disponível em: <<http://www.pr.gov.br/seab/deral/cprbr.xls>>. Acesso em: 13 jan. 2002.

PELLEGRINI, F. R.; FOGLIATTO, F. Estudo comparativo entre modelos de Winters e de Box-Jenkins para a previsão de demanda sazonal. **Revista Produto & Produção**, v. 4, p. 72-85, 2000.

REID, R. D.; SANDERS, N. R. **Gestão de operações**. Rio de Janeiro, RJ: LTC, 2005.

SAMOHYL, R. W.; ROCHA, R. Utilização do método de Holt-Winters para previsão do leite entregue às indústrias catarinenses.

In: ESCOLA DE SÉRIES TEMPORAIS E ECONOMETRIA, 9, 2001, Belo Horizonte. **Anais...** Belo Horizonte, MG: UFMG, 2001. 1 Cd-Rom.

SARTORIS, A. **Estatística e introdução à econometria**. São Paulo, SP: Saraiva, 2003.

SILVA, W. V.; SAMOHYL, R. W.; COSTA, L. S. Comparação entre os métodos de previsão univariados para o preço médio da soja no Brasil. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 12, 2002, São Paulo. **Anais...** Porto Alegre, RS: Sobrapo, 2002. 1 Cd-Rom.

YAFEE, R.; MCGEE, M. **An introduction to time series analysis and forecasting**: with applications of SAS and SPSS. Burlington: Academic Press, 2000.

*Recebido em: 23 Julho 2009*

*Aceito em: 20 Outubro 2009*