

Universidade Federal de Juiz de Fora  
Instituto de Ciências Exatas  
Curso de Estatística

**Isabela Pagani Heringer de Miranda**

**Comparação de diferentes Métodos de Previsão em Séries Temporais com  
valores discrepantes**

Juiz de Fora  
2014

Isabela Pagani Heringer de Miranda

**Comparação de diferentes Métodos de Previsão em Séries Temporais com valores discrepantes**

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do título de Bacharel em Estatística .

Orientador: Henrique Steinherz Hippert

Juiz de Fora

2014

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF  
com os dados fornecidos pelo(a) autor(a)

Pagani Heringer de Miranda, Isabela.

Comparação de diferentes Métodos de Previsão em Séries Temporais  
com valores discrepantes / Isabela Pagani Heringer de Miranda. – 2014.  
29 f. : il.

Orientador: Henrique Steinherz Hippert

Trabalho de Conclusão de Curso – Universidade Federal de Juiz de Fora,  
Instituto de Ciências Exatas. Curso de Estatística, 2014.

1. Métodos de Previsão. 2. Séries Temporais. 3. Combinações de  
Previsões. 4. Medianas Móveis. 5. Medidas de Erro de Previsão. I.  
Steinherz Hippert, Henrique, orient. II. Título.

Isabela Pagani Heringer de Miranda

**Comparação de diferentes Métodos de Previsão em Séries Temporais com valores discrepantes**

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do título de Bacharel em Estatística .

Aprovada em: 12/12/2014

BANCA EXAMINADORA

---

Prof. Dr. Henrique Steinherz Hippert - Orientador  
Universidade Federal de Juiz de Fora

---

Professor Dr. Augusto Carvalho Souza  
Universidade Federal de Juiz de Fora

---

Professor Dr. Ronaldo Rocha Bastos  
Universidade Federal de Juiz de Fora

*Dedico este trabalho aos alunos do curso de Estatística da Universidade Federal de Juiz de Fora, por serem dedicados e unidos, sempre ajudando uns aos outros. Com o meu trabalho, espero ajudá-los de alguma forma.*

## AGRADECIMENTOS

Agradeço aos meus pais, João Paulino e Valéria, por acreditarem mais em mim do que eu mesma. Mesmo sem saberem o que um estatístico faz, me apoiaram nessa escolha e em todas as outras que fiz e faço na vida. A eles, meu muito obrigada por isso e por tudo que representam para mim. Agradeço também as minhas irmãs, Marcela e Gabriela, e a minha sobrinha Catarina, por serem as minhas melhores amigas e sempre me alegrarem.

Ao meu avô, minhas avós, tios e primos, eu agradeço todo o carinho. Ao Talles, agradeço pela compreensão, muitas vezes o estresse com a faculdade acabava sendo descontado nele; também o agradeço por todo amor dedicado a mim.

Aos professores do departamento de Estatística da UFJF, por transmitirem, com o maior prazer, tanto conhecimento e estarem disponíveis para ajudar todos os alunos. Em especial, agradeço ao professor Henrique Hippert.

Agradeço aos amigos que fiz nesses oito anos vivendo em Juiz de Fora. Agradeço aos amigos que fiz no ICE, mesmo os que são de outros cursos como Flavinha, Isaura, Paty, Thaeny e Mariliane. Do curso de Estatística, agradeço, em especial, à Bárbara, Bethânia, Douglas e Gabriely, eles foram e são amigos dentro e fora da faculdade. Jamais esquecerei nossos estudos em véspera de natal e ano novo, como também não me esquecerei das listas de exercícios e trabalhos compartilhados, das ligações durante as madrugadas para tirarmos inúmeras dúvidas. Com eles eu sorri, chorei e estudei muito! Não tem jeito, para sempre seremos um time!

Agradeço a todas as pessoas que fizeram e fazem parte da minha vida.

E por falar em vida, meu muitíssimo obrigada a Deus! Acima de tudo, sem Ele, não iria à lugar algum.

*“Tenho somente uma luz pela qual meus pés são guiados – a luz da experiência. Não conheço nenhum modo de julgar o futuro que não seja através do passado”.*

Patrick Henry

*“Acho que é muito difícil lidar com os fatos sem nos perdermos atrás de teorias e fantasias”.*

**Inspetor Lestrade para Sherlock Holmes**

O Mistério do Vale Boscombe.

## RESUMO

As previsões são de grande importância quando nos deparamos com fenômenos nos quais a incerteza é um fator presente. Existem diversos métodos de previsão e cada um tem sua aplicabilidade, dependendo do que se está interessado estudar. Em séries temporais a amostra é um conjunto de observações ordenadas no tempo e a ordem em que são feitas as medições não pode ser ignorada. O foco desse trabalho é estudar diferentes métodos de previsão em séries temporais em que existam valores discrepantes. Dentro deste contexto, propomos um método de previsão baseado em Medianas Móveis, para séries de nível constante, baseado no fato da mediana não ser influenciada por valores discrepantes. O desempenho deste método foi comparado ao dos métodos de Amortecimento Exponencial Simples, Média Móvel Simples e a combinação linear entre Média Móvel e Mediana Móvel Simples, por meio de simulações sobre séries de dados simulados e de dados reais. As estimativas dos parâmetros foram feitas usando a função *optimize()* e implementando a função *find\_n* no *software* R. Os resíduos foram analisados e comparados, buscando o método que minimiza esses resíduos, ou seja, o método mais adequado para as séries estudadas.

Palavras-chave: Métodos de Previsão. Séries Temporais. Combinações de Previsões. Medianas Móveis. Medidas de Erro de Previsão.

## ABSTRACT

Forecasts are important when one is faced with phenomena whereupon uncertainty is present. There are various methods of forecasting and each one has its applicability, depending on the kind of series that one wants to study. In time series the sample is a set of observations ordered in time, and the order wherein the measurements are made cannot be ignored. The focus of this work is to study different methods of forecasting in time series where there are outliers. In this context, we propose a method of forecasting based on Moving Median, for constant level series, based on the fact that the median is not influenced by outliers. The performance of this method was compared to that of the Simple Exponential Smoothing methods, Simple Moving Average and a linear combination of Moving Average and Moving Median, through simulations on simulated data sets and real data. The parameter estimation was made using the function *optimize()* and implementing the function *find\_n* in the *software* R. The residues were analyzed and compared, in order to seek the method that minimizes the error, in other words, the most appropriate method for the studied series.

Key-words: Forecasting Methods. Time Series. Forecast Combining. Moving Median. Forecasting Error Measures.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Quatro componentes de uma série . . . . .	12
Figura 2 – Cálculo das médias móveis simples . . . . .	17
Figura 3 – Gráfico de uma das séries simuladas . . . . .	21
Figura 4 – Gráficos dos módulos das séries simuladas com os valores discrepantes.	23
Figura 5 – Gráficos das Séries Reais . . . . .	25

## SUMÁRIO

1	INTRODUÇÃO . . . . .	10
2	SÉRIES TEMPORAIS . . . . .	12
3	MATERIAIS E MÉTODOS . . . . .	14
3.1	SÉRIES . . . . .	14
3.2	MÉTODOS DE OTIMIZAÇÃO . . . . .	15
3.2.1	Função <i>Optimize()</i> . . . . .	15
3.2.2	Função <i>Find_n</i> . . . . .	15
3.3	MÉTODOS DE PREVISÃO . . . . .	15
3.3.1	Previsão por Média Móvel Simples . . . . .	15
3.3.2	Previsão por Mediana Móvel Simples . . . . .	17
3.3.3	Previsão pela combinação linear de Média Móvel e Mediana Móvel . . . . .	18
3.3.4	Previsão por Amortecimento Exponencial Simples . . . . .	18
3.4	MEDIDAS DE ERRO DE PREVISÃO . . . . .	19
3.4.1	Erro MAPE . . . . .	20
3.4.2	Erro MSE . . . . .	20
4	RESULTADOS . . . . .	21
5	DISCUSSÃO . . . . .	27
6	CONCLUSÕES . . . . .	28
	REFERÊNCIAS . . . . .	29

## 1 INTRODUÇÃO

Para testar uma teoria, um pesquisador faz previsões sobre os valores futuros de uma variável de interesse, com base no modelo que desenvolveu; se estas previsões se confirmam, dizemos que a teoria recebeu corroboração empírica.

O avanço da Estatística e da computação tem permitido o surgimento de novos métodos de previsão com uso em diversas áreas, bem como a combinação de algum deles, sempre buscando a melhor estimativa, minimizando os erros. Esse avanço se deve principalmente ao número cada vez maior de informações disponíveis. Porém, nem sempre as previsões são fáceis de obter, e é necessário fazer uso de softwares apropriados.

Previsões são de grande importância nas empresas já que podem melhorar a eficiência dos projetos, antecipar problemas e desenvolver planos para resolver estes problemas. É essencial para o planejamento – uma organização não pode planejar se não consegue prever com razoável grau de precisão que demanda haverá para o produto ou serviço que oferece, e que recursos estarão disponíveis para a produção.

Os planejamentos de curto, médio e principalmente longo prazo são de difícil elaboração, pois sempre nos deparamos com fenômenos onde existe incerteza, impedindo o conhecimento exato do comportamento futuro. Nesses casos, pode-se perceber com clareza a necessidade de produzir previsões. Com base no passado da série, podemos fazer previsões por meio de diversos métodos. A combinação de previsões de médio ou longo prazo com previsões de curto prazo ajuda a identificar pontos de decisão críticos onde intervenções são necessárias, pois permite identificar se o sistema está indo na direção que foi prevista [10].

Em alguns casos, como em finanças onde o ambiente é instável e competitivo, as previsões podem gerar lucros ou prejuízos. Estimativas superestimadas (previsão acima do valor real) acarretam o chamado “custo de excesso”. Já estimativas subestimadas (previsão abaixo do valor real) ocasionam o chamado “custo de falta”. Ambas podem gerar prejuízos. Por isso, o objetivo é gerar previsões que estejam o mais próximo possível da realidade, ou seja, criar modelos que possuam os menores erros de previsão possíveis.

Geralmente existe um melhor método de previsão para cada tipo de série. O objetivo deste estudo é analisar e comparar quatro tipos de métodos de previsão em séries onde valores discrepantes surgem aleatoriamente. Os métodos de previsão aqui testados foram Amortecimento Exponencial Simples, Média Móvel Simples, Mediana Móvel Simples e a Combinação Linear de Média Móvel com Mediana Móvel.

O método de previsão por Medianas Móveis não tem sido comumente usado para fazer previsões, porém sabendo que a média tem como desvantagem o fato de ser influenciada por valores discrepantes, é válido o estudo do método de previsão por Medianas

Móveis buscando corrigir possíveis desvios já que a mediana não é afetada por estes valores. Outro método válido de estudar é a combinação de previsões; este método é usado com a finalidade de melhorar a precisão das previsões.

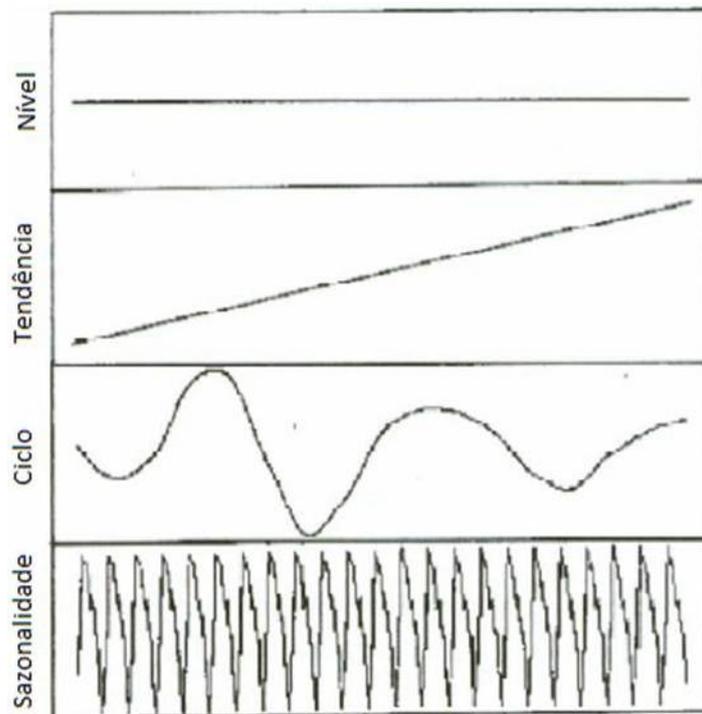
Para avaliar o desempenho dos quatro métodos mencionados, faremos comparações empíricas de seus erros de previsão, sobre séries com valores discrepantes, tanto de dados reais quanto simulados. Espera-se que o método de previsão por Mediana Móvel obtenha um desempenho superior quando comparado aos demais métodos aqui estudados.

## 2 SÉRIES TEMPORAIS

Uma série temporal é um conjunto de observações de uma determinada variável feitas em períodos sucessivos de tempo, ao longo de um determinado intervalo. Estas séries podem ser tratadas como amostras aleatórias ordenadas no tempo; a ordem em que são feitas as medições é fundamental e não pode ser esquecida. Existe diferença entre séries temporais e dados transversais (*cross-section*). Os dados provenientes de uma *cross-section* possuem apenas dependência espacial e não linear; por isso, um valor alto de um dado pode não dizer nada a respeito do valor de outro dado. A análise de séries temporais, por outro lado, mostra que existe dependência entre os dados, havendo correlação entre as observações em diferentes instantes.

Tendo uma série temporal, a primeira coisa a se fazer é a representação gráfica da série com o objetivo de analisar e identificar as características relevantes para o estudo da série. Uma série temporal pode ser analisada em quatro componentes: nível (a média em torno da qual os valores da série flutuam), tendência (quando o nível apresenta comportamento ascendente ou descendente por um longo período de tempo), ciclo (quando o nível possui variações ascendentes ou descendentes em intervalos não regulares de tempo) e sazonalidade (quando padrões cíclicos de variação se repetem em intervalos relativamente constantes de tempo). A figura 1 mostra os quatro componentes de uma série.

Figura 1 – Quatro componentes de uma série



Fonte: Desenvolvido pela autora

Neste estudo vamos adotar as seguintes notações:

- Valor da variável  $Z$  no instante  $t$ :  $Z_t$
- Previsão feita no instante  $t$ , do valor de  $Z$  no instante  $t+k$ :  $\hat{Z}_{t+k}$
- Previsão um passo a frente:  $\hat{Z}_{t+1}$

### 3 MATERIAIS E MÉTODOS

#### 3.1 SÉRIES

Para testar os métodos de previsão de Amortecimento Exponencial Simples, Médias Móveis Simples, Medianas Móveis Simples e Combinação Linear entre os métodos de Médias e Medianas, primeiro simulamos cinco séries de tamanho  $n = 200$  cada. Essas séries foram simuladas com média constante e com o erro assumindo uma distribuição normal padrão, seguindo o modelo  $Z_t = a + e_t$ , onde  $a = 5$  e  $e_t \sim N(0, 1)$

Fizemos essas séries receberem dez valores discrepantes em diferentes posições selecionadas aleatoriamente. Para as três primeiras fizemos o desvio padrão dos discrepantes serem iguais a dez vezes o desvio padrão do resto da série; para as outras duas séries, fizemos o desvio padrão dos discrepantes receber um valor vinte vezes superior ao desvio padrão do restante da série.

Simulamos mais cinco séries também de tamanho  $n = 200$  e com média constante igual a zero, e calculamos o módulo dos valores observados para que pudéssemos usar apenas os valores positivos de cada série. Na primeira série criamos dez valores discrepantes em diferentes posições selecionadas aleatoriamente e fizemos o desvio padrão desses discrepantes receberem um valor dez vezes o desvio padrão do restante da série. Na segunda série, aumentamos o número dos valores discrepantes para trinta e fizemos o desvio padrão destes discrepantes receberem um valor cinco vezes o desvio padrão da série. Na terceira, quarta e quinta série, fizemos o mesmo que na segunda, porém o desvio padrão dos discrepantes recebeu um valor dez, quinze e vinte vezes o desvio padrão do restante da série, respectivamente.

Depois de feitas as previsões e calculados os erros para as séries simuladas, fizemos o mesmo para as séries reais *flow*, *google*, *days*, *larain* e *robot* do pacote TSA do *software* R e a série *precip BA*. Essas séries foram escolhidas pelo fato da presença de valores discrepantes ao longo delas e também pela média da maioria delas ser praticamente constante ao longo do tempo.

Descrição das séries:

**Flow:** Fluxo mensal do rio Iowa.

**Google:** Retornos diários das ações do Google.

**Days:** Números de dias entre o pagamento a Winegard Corp.

**Larain:** Precipitação anual em Los Angeles.

**Robot:** A distancia de um robô a partir de uma posição definida.

**Precip BA:** Precipitação média mensal em várias cidades da Bahia.

## 3.2 MÉTODOS DE OTIMIZAÇÃO

### 3.2.1 Função *Optimize()*

Para fazer a busca do valor ótimo da constante do método de amortecimento exponencial simples, e dos pesos usados na combinação de métodos, usamos a função *optimize()* do R. Esta função faz a minimização de uma função de custo escolhida (usamos o erro médio quadrático), usando como default o método de Nelder & Mead [12]. A função permite também o uso de outros métodos de minimização, como os quasi-Newton ou os de gradiente conjugado, mas estes não foram empregados neste trabalho. Com o argumentos de entrada desta função, é necessário fornecer valores iniciais para os parâmetros, o intervalo de busca, a função a minimizar e a função de custo.

### 3.2.2 Função *Find\_n*

Para definir o tamanho da média móvel e da mediana móvel, implementamos no *software* R a função *find\_n*. Essa função faz uma busca em grade [11] em um vetor com os possíveis valores do tamanho da média móvel e da mediana móvel, simulando as previsões com cada valor e comparando os erros; o tamanho escolhido é aquele que resulta no menor erro de previsão.

## 3.3 MÉTODOS DE PREVISÃO

### 3.3.1 Previsão por Média Móvel Simples

O método de médias móveis simples utiliza como previsão a média das observações antigas. Este método é diferente do método de previsão por amortecimento exponencial simples, pois aqui as observações usadas para encontrar a previsão do valor futuro recebem o mesmo peso cada uma para o cálculo da previsão.

O método supõe que a série segue um modelo constante, dado por:

$$Z_t = \mu_t + \varepsilon_t$$

Onde:  $\varepsilon_t$  são variáveis independentes e identicamente distribuídas.

$$E(\varepsilon_t) = 0 \text{ e } Var(\varepsilon_t) = \sigma^2, \forall t.$$

Se o nível é constante, temos:

$\mu_t$ : nível médio no instante t.  $\mu_t = a$

$$E(Z_t) = a \text{ e } Var(Z_t) = \sigma^2$$

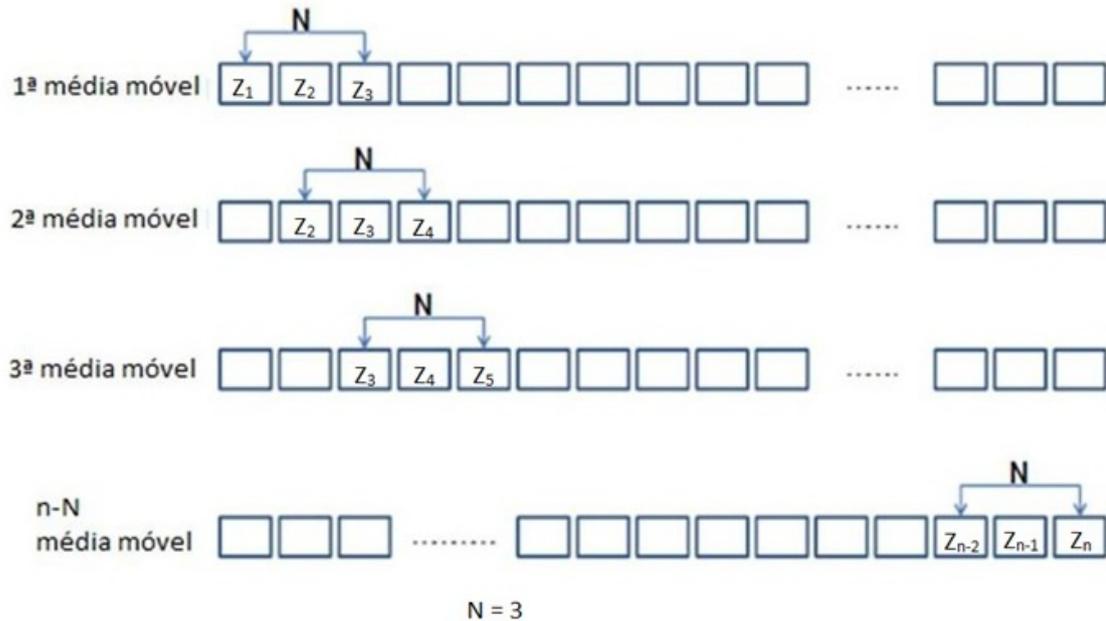
A equação de previsão do método de médias móveis simples é dada por:

$$\hat{Z}_{T+k/T} = \frac{Z_T + Z_{T-1} + \dots + Z_{T-N+1}}{N} \quad (3.1)$$

Onde:  $\hat{Z}_{T+k/T}$  é a previsão do valor no instante  $T+k$ , feita no instante  $T$ .  $N$  é o número de observações a ser utilizado para o cálculo da média móvel simples.  $Z_T$  é a observação mais recente,  $Z_{T-1}$  é a penúltima observação mais recente e assim por diante até chegar na observação  $Z_{T-N+1}$ .

A Figura 2 mostra como é feito o cálculo dessa média móvel, supondo uma série com  $n$  observações:

Figura 2 – Cálculo das médias móveis simples



Fonte: Desenvolvido pela autora

Essa figura é um exemplo para o cálculo de uma média móvel de tamanho três. Este tamanho será definido utilizando a função *find\_n* no *software* R versão 3.1.0. Percebe-se no exemplo que para cada nova janela de três dias a ser formada, a observação mais antiga é desprezada, e uma nova observação é inserida no conjunto N para o próximo cálculo.

### 3.3.2 Previsão por Mediana Móvel Simples

Não existe na literatura registro do uso do modelo de medianas móveis para fazer previsões. O que se pode encontrar é a aplicação desse método no amortecimento de uma série, com a finalidade de extrair sua tendência [13]. Apesar de ser um método novo, acreditamos que pelo fato da mediana não ser influenciada por valores discrepantes, este método possa ser útil para fazer previsões em séries com valores discrepantes.

A equação do método de previsão por mediana móvel simples é dada por:

$$\hat{Z}_{T+k/T} = \text{mediana}(Z_T, Z_{T-1}, \dots, Z_{T-N+1}) \quad (3.2)$$

Se N for ímpar, a mediana será o valor central do vetor com as observações ordenadas. E se N for par, não existe apenas um valor central e sim dois valores de centro. Neste caso a mediana será a média desses dois valores do centro.

Da mesma maneira que é feito no método da média móvel, o tamanho de N é definido usando a função *find\_n* implementada no *software* R versão 3.1.0.

### 3.3.3 Previsão pela combinação linear de Média Móvel e Mediana Móvel

O primeiro artigo sobre combinação de previsões foi publicado pelos pesquisadores Bates e Granger em 1969 [2]. Neste artigo, os autores apresentaram métodos de combinação linear de dois previsores, considerando as hipóteses de que um previsor talvez seja baseado em várias observações ou informações que o outro previsor não considerou, ou de que um previsor faça suposições diferentes sobre a relação entre as variáveis.

No presente estudo, usamos o método *outperformance* de combinação, baseado em dois artigos publicados na década de 1970 desenvolvido por Bunn [3] [4]. Esse método propõe que dadas duas previsões  $p_1$  e  $p_2$ , a previsão final é:

$$P = \alpha_1 p_1 + \alpha_2 p_2 \quad (3.3)$$

Nas simulações que fizemos, usamos a notação:

$\alpha_1$  é o peso atribuído ao método de previsão por médias móveis simples;

$\alpha_2$  é o peso atribuído ao método de previsão por medianas móveis simples;

$p_1$  e  $p_2$  são as previsões pelos métodos de média e mediana móvel respectivamente;

$P$  é a nova previsão pela combinação desses modelos.

Os pesos obedecem às restrições:

$$\alpha_1 + \alpha_2 = 1$$

$$0 \leq \alpha_1 \leq 1$$

O valor ótimo de  $\alpha_1$  e conseqüentemente de  $\alpha_2$  ( $\alpha_2 = 1 - \alpha_1$ ), foi definido usando a função *optimize()* do *software* R, versão 3.1.0.

No modelo de previsão por combinação, a probabilidade de um dos modelos é interpretada como uma avaliação sobre a eficiência desse modelo em relação ao outro utilizado na combinação.

### 3.3.4 Previsão por Amortecimento Exponencial Simples

A idéia principal do método de amortecimento exponencial simples é “ponderar” a média, ou seja, atribuir um peso diferente a cada observação da série. Quanto mais velha a observação, menor o peso que ela recebe.

O método também se baseia no modelo constante, definido na seção 3.2.1.

A previsão feita pelo método de amortecimento exponencial simples é dada por:

$$\hat{Z}_{T+1} = \alpha Z_T + (1 - \alpha)\hat{Z}_T \quad (3.4)$$

Onde:  $\hat{Z}_{T+1}$  representa a previsão do valor  $Z$  no instante  $T+1$ .

Para inicializar o método, precisamos escolher um valor inicial para  $\hat{Z}_0$ ; utilizaremos a média das primeiras observações.

O peso utilizado na ponderação é  $\alpha$ , ( $0 < \alpha < 1$ ); valores pequenos de  $\alpha$  produzem previsões que dependam mais das observações mais antigas, pois assim, será dado um peso maior a estas observações e altos valores de  $\alpha$  (próximo de 1) produzem previsões que dependam mais das observações mais recentes, dando a estas observações um peso maior. Se o valor de  $\alpha$  for igual a 1, temos o método de previsão naive, onde a previsão é dada apenas pela última observação.

$Z_T$  é a observação mais recente, ou seja, é o último valor observado e  $\hat{Z}_T$  é o valor da previsão anterior.

Da mesma maneira que foi feito no modelo de previsão pela combinação de média móvel e mediana móvel, o  $\alpha$  foi definido usando a função *optimize()* do *software* R versão 3.1.0.

### 3.4 MEDIDAS DE ERRO DE PREVISÃO

As medidas de erro de previsão são usadas para avaliar o desempenho de um método de previsão. A equação do erro de previsão  $k$  passos à frente, feita no instante  $t$ , é dada por:

$$e_{t+k/t} = Z_{t+k/t} - \hat{Z}_{t+k/t} \quad (3.5)$$

Onde:  $Z_{t+k/t}$  é o valor observado da série e  $\hat{Z}_{t+k/t}$  é o valor previsto da série.

A ideia é que o modelo escolhido minimize uma medida baseada numa função deste erro. Na literatura podem-se encontrar diversas destas medidas. A mais simples é erro médio (ME) é uma média de todos os valores  $e_t$ , cuja equação é dada por:

$$ME = \frac{1}{N} \sum_{t=1}^N e_t \quad (3.6)$$

No presente estudo as medidas de erro MAPE e MSE foram calculadas e através delas comparamos os modelos de previsão aqui testados.

### 3.4.1 Erro MAPE

O *erro absoluto percentual médio* (MAPE) é um dos mais usados na Estatística por ser de fácil compreensão. Esta medida expressa a porcentagem média dos erros (em valor absoluto) cometidos na previsão da série temporal; quanto menor esta medida, melhor será a previsão.

O MAPE possui uma desvantagem quando, no estudo, têm-se séries que passam pelo valor zero; se algum dos valores observados for igual à zero, o MAPE tenderá ao infinito.

A equação do MAPE é dada por:

$$MAPE = \frac{100}{N} \sum_{t=1}^N \frac{|e_t|}{|Z_t|} \quad (3.7)$$

Onde:  $|e_t|$  é o módulo do t-ésimo erro e  $|Z_t|$  é o módulo do valor da variável Z no instante t.

### 3.4.2 Erro MSE

O *erro quadrático médio* (MSE) é a média dos quadrados dos erros; quanto maiores forem os erros, maior será o MSE, e se os erros forem pequenos, ele também será pequeno. Assim como o MAPE, o MSE também é muito usado na Estatística, mas sua interpretação é mais complicada do que a interpretação do MAPE, pois a unidade de medida do MSE é a unidade original de Z, elevada ao quadrado.

Sua expressão é dada por:

$$MSE = \frac{1}{N} \sum_{t=1}^N e_t^2 \quad (3.8)$$

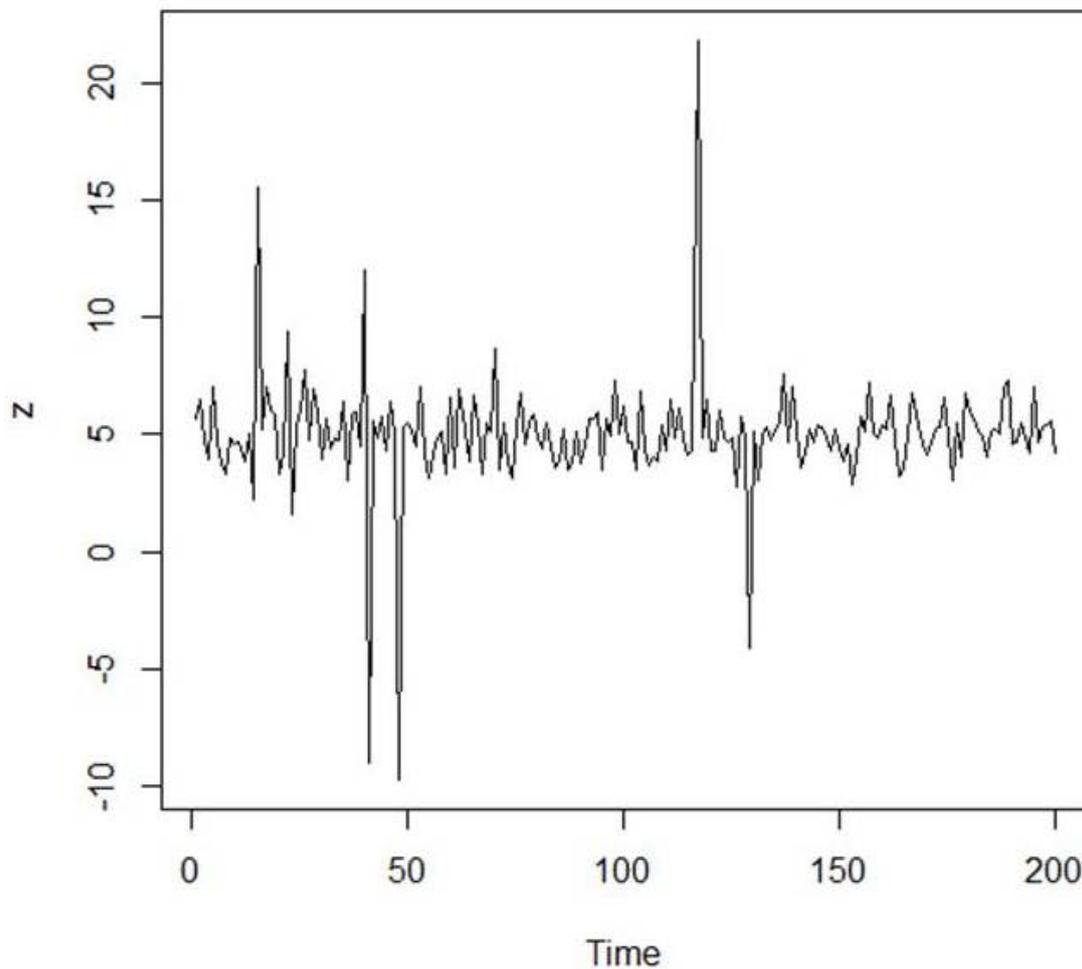
## 4 RESULTADOS

### Séries Simuladas I

Os resultados foram obtidos testando quatro modelos de previsão em quinze séries. Primeiro simulamos cinco séries como foi definido na seção 3.1.

A Figura 3 mostra o gráfico de uma dessas séries.

Figura 3 – Gráfico de uma das séries simuladas



Fonte: Desenvolvido pela autora

Os quatro métodos de previsão (Amortecimento Exponencial Simples, Média Móvel Simples, Mediana Móvel Simples e Combinação linear da Média e Mediana) foram aplicados nessas séries e as medidas de erro MSE e MAPE foram calculados. Na Tabela 1 estão os valores desses erros.

Tabela 1 – Erros MSE e MAPE das previsões das séries simuladas.

SÉRIE	MÉTODO DE PREVISÃO	MSE	MAPE
1	Amortecimento Exponencial Simples ( $\alpha = 0,025$ )	<b>7,95</b>	26,65
	Média Móvel Simples (n=10)	8,62	30,09
	Mediana Móvel Simples (n=10)	8,54	<b>26,41</b>
	Combinação de Média e Mediana	8,53	26,67
2	Amortecimento Exponencial Simples ( $\alpha = 0,028$ )	3,33	24,04
	Média Móvel Simples (n=6)	1,68	18,90
	Mediana Móvel Simples (n=6)	1,63	18,69
	Combinação de Média e Mediana	<b>1,62</b>	<b>18,52</b>
3	Amortecimento Exponencial Simples ( $\alpha = 0,030$ )	5,99	56,84
	Média Móvel Simples (n=5)	5,95	56,99
	Mediana Móvel Simples (n=5)	<b>5,58</b>	53,84
	Combinação de Média e Mediana	<b>5,58</b>	<b>53,56</b>
4	Amortecimento Exponencial Simples ( $\alpha = 0,029$ )	<b>11,54</b>	<b>25,78</b>
	Média Móvel Simples (n=10)	14,14	31,87
	Mediana Móvel Simples (n=10)	13,02	26,27
	Combinação de Média e Mediana	12,20	26,44
5	Amortecimento Exponencial Simples ( $\alpha = 0,031$ )	37,69	39,49
	Média Móvel Simples (n=6)	37,80	28,79
	Mediana Móvel Simples (n=9)	<b>30,11</b>	<b>23,53</b>
	Combinação de Média e Mediana	30,12	23,54

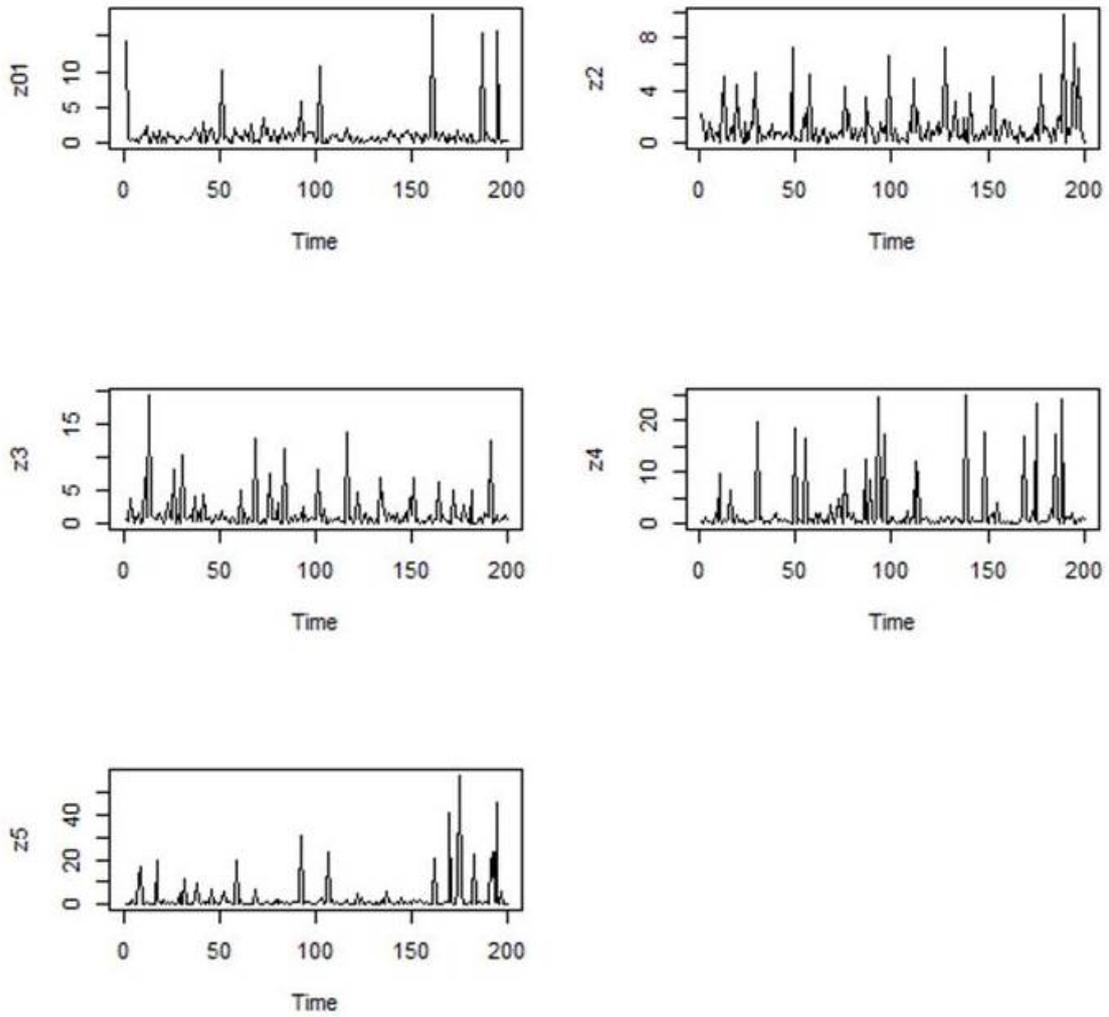
1,2 e 3 - Séries simuladas com 10 valores discrepantes selecionados aleatoriamente e com valor do desvio padrão dos discrepantes 10 vezes o desvio padrão da série.

4 e 5 - Séries simuladas com 10 valores discrepantes selecionados aleatoriamente e com valor do desvio padrão dos discrepantes 20 vezes o desvio padrão da série.

### Séries Simuladas II

Simulamos mais cinco séries também definidas na seção 3.1. Podemos ver os gráficos na Figura 4.

Figura 4 – Gráficos dos módulos das séries simuladas com os valores discrepantes.



Fonte: Desenvolvido pela autora

Novamente aplicamos os quatro modelos de previsão e calculamos os erros MSE e MAPE que estão representados na Tabela 2.

Tabela 2 – Erros MSE e MAPE das previsões do módulo das séries simuladas.

SÉRIE	MÉTODO DE PREVISÃO	MSE	MAPE
1	Amortecimento Exponencial Simples ( $\alpha = 0,035$ )	<b>3,11</b>	2816,47
	Média Móvel Simples (n=5)	3,56	2070,17
	Mediana Móvel Simples (n=5)	3,37	<b>1676,25</b>
	Combinação de Média e Mediana	3,33	1769,24
2	Amortecimento Exponencial Simples ( $\alpha = 0,032$ )	<b>5,14</b>	1024,22
	Média Móvel Simples (n=8)	8,04	807,80
	Mediana Móvel Simples (n=8)	6,06	<b>715,38</b>
	Combinação de Média e Mediana	5,48	737,89
3	Amortecimento Exponencial Simples ( $\alpha = 0,031$ )	<b>7,02</b>	440,55
	Média Móvel Simples (n=1)	8,78	476,59
	Mediana Móvel Simples (n=5)	7,60	<b>266,45</b>
	Combinação de Média e Mediana	7,59	287,06
4	Amortecimento Exponencial Simples ( $\alpha = 0,009$ )	<b>22,14</b>	513,76
	Média Móvel Simples (n=1)	47,20	649,02
	Mediana Móvel Simples (n=6)	25,44	<b>239,79</b>
	Combinação de Média e Mediana	25,34	261,32
5	Amortecimento Exponencial Simples ( $\alpha = 0,017$ )	<b>27,23</b>	426,51
	Média Móvel Simples (n=6)	32,83	545,67
	Mediana Móvel Simples (n=9)	31,05	<b>174,49</b>
	Combinação de Média e Mediana	30,28	296,94

1 - Série simulada com 10 valores discrepantes selecionados aleatoriamente e com valor do desvio padrão dos discrepantes 10 vezes o desvio padrão da série.

2 - Série simulada com 30 valores discrepantes selecionados aleatoriamente e com valor do desvio padrão dos discrepantes 5 vezes o desvio padrão da série.

3 - Série simulada com 30 valores discrepantes selecionados aleatoriamente e com valor do desvio padrão dos discrepantes 10 vezes o desvio padrão da série.

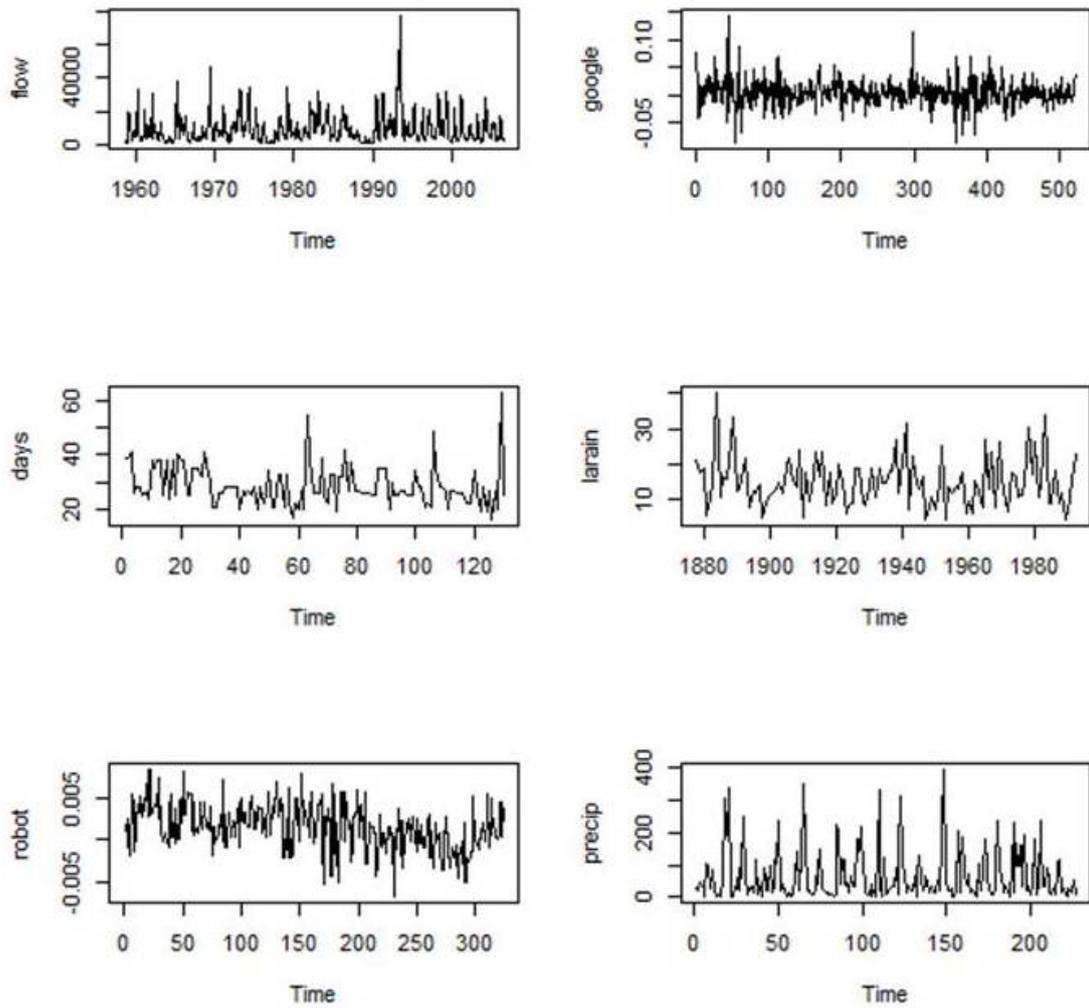
4 - Série simulada com 30 valores discrepantes selecionados aleatoriamente e com valor do desvio padrão dos discrepantes 15 vezes o desvio padrão da série.

5 - Série simulada com 30 valores discrepantes selecionados aleatoriamente e com valor do desvio padrão dos discrepantes 20 vezes o desvio padrão da série.

#### Séries de dados reais

Por último, aplicamos os métodos de previsão e calculamos as medidas de erro para as séries reais *flow*, *google*, *days*, *larain* e *robot* do pacote TSA e para a série *precip BA*. Os gráficos dessas séries estão representados na Figura 5.

Figura 5 – Gráficos das Séries Reais



Fonte: Desenvolvido pela autora

As previsões foram feitas usando cada um dos quatro métodos. Os erros MSE e MAPE foram calculados, e seus valores são apresentados na Tabela 3.

Tabela 3 – Erros MSE e MAPE das previsões das séries reais.

SÉRIE	MÉTODO DE PREVISÃO	MSE	MAPE
Flow	Amortecimento Exponencial Simples ( $\alpha = 0,13$ )	46424,0	50,23
	Média Móvel Simples (n=1)	<b>46789,0</b>	48,59
	Mediana Móvel Simples (n=1)	<b>46789,0</b>	48,58
	Combinação de Média e Mediana	<b>46789,0</b>	<b>48,57</b>
Google	Amortecimento Exponencial Simples ( $\alpha = 0,14$ )	0,00072	238,62
	Média Móvel Simples (n=6)	0,00073	227,99
	Mediana Móvel Simples (n=9)	0,00073	240,06
	Combinação de Média e Mediana	<b>0,00071</b>	<b>220,96</b>
Days	Amortecimento Exponencial Simples ( $\alpha = 0,15$ )	45,48	19,17
	Média Móvel Simples (n=9)	<b>44,10</b>	18,53
	Mediana Móvel Simples (n=9)	51,46	<b>18,45</b>
	Combinação de Média e Mediana	<b>44,10</b>	18,52
Larain	Amortecimento Exponencial Simples ( $\alpha = 0,084$ )	46,61	50,39
	Média Móvel Simples (n=5)	<b>40,63</b>	<b>46,47</b>
	Mediana Móvel Simples (n=5)	45,41	46,54
	Combinação de Média e Mediana	<b>40,63</b>	<b>46,47</b>
Robot	Amortecimento Exponencial Simples ( $\alpha = 0,11$ )	<b>6,23e-06</b>	-
	Média Móvel Simples (n=8)	6,30e-06	-
	Mediana Móvel Simples (n=7)	6,89e-06	-
	Combinação de Média e Mediana	6,30e-06	-
Precip BA	Amortecimento Exponencial Simples ( $\alpha = 0,033$ )	<b>6685,21</b>	487,10
	Média Móvel Simples (n=1)	8912,83	352,53
	Mediana Móvel Simples (n=9)	8523,72	<b>279,95</b>
	Combinação de Média e Mediana	6717,04	303,40

Para a série robot não calculamos os valores do MAPE, pois essa série passa pelo valor zero e assim quando um valor observado for igual a zero esta medida tenderá ao infinito.

Para cada método de previsão, e para cada série, foi feito o teste de Kolmogorov-Smirnov para testar normalidade dos erros, e gráficos das funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) para verificar a independência dos erros. Todos os testes rejeitaram a hipótese de normalidade e independência dos erros, o que já era esperado devido aos valores discrepantes.

## 5 DISCUSSÃO

Para as cinco primeiras séries simuladas o método de previsão por Medianas Móveis (MnN) se mostrou como melhor previsor que o método por Médias Móveis Simples (MM). Podemos perceber este fato quando analisamos as medidas de erros MSE e MAPE na Tabela 1. O método de previsão por MnM obteve em todas as séries do primeiro grupo o MSE e o MAPE menores do que essas medidas de erro para o método de previsão por MM. Este último método não se destacou como o melhor em nenhuma das cinco primeiras séries simuladas.

Nas outras cinco séries simuladas onde foram considerados apenas os valores positivos das séries, o método de previsão por MnM apresentou em todos os casos a menor medida de erro MAPE, mostrando que para esse tipo de série e quando o objetivo é minimizar o erro MAPE o método de previsão por MnM é o melhor. Para as mesmas séries o método de Amortecimento Exponencial Simples (AES) apresentou as melhores previsões quando considerado a medida de erro MSE, como pode ser observado na Tabela 2.

Entre as séries reais estudadas, os resultados foram menos claros. As MnM conseguiram os melhores MAPEs nas séries *days* e *precip BA*. A combinação entre MnM e MM obteve melhores resultados, tanto em MAPE quanto em MSE, na série *google*, e o melhor MSE na série *days*. As MM obtiveram melhores MAPEs e MSEs na série *larain*. O AES obteve os melhores MSEs nas séries *robot* e *precip BA*.

Se analisarmos estes resultados de acordo com as características de cada série, verificamos:

- para a única série não-estacionária na média (*robot*), o melhor método foi o de AES.
- para as duas séries com sazonalidade (*flow* e *precip BA*), as MM foram igualadas ao previsor naive ( $n=1$ ); as MnM tiveram bons resultados na série *precip BA*.
- para as três séries sem autocorrelação, equivalentes a ruídos brancos (*google*, *days* e *larain*), os melhores métodos foram os de MM, MnM, e a combinação destes dois. Não parece ser possível julgar, a partir das características de cada uma destas séries, qual método irá obter o melhor desempenho preditivo.

## 6 CONCLUSÕES

O foco desse estudo é comparar quatro métodos de previsão em série com valores discrepantes, principalmente o método de previsão por Medianas Móveis Simples que é um método novo, e não se tem registros na literatura do seu uso para fazer previsões.

Acreditávamos que este método poderia obter bons resultados em séries onde existem valores muito discrepantes; isto de fato ocorreu, como verificado empiricamente nas séries simuladas I e II. Nos experimentos com séries reais, porém, os resultados não foram muito claros. As séries reais usadas tinham outras características, que as séries simuladas não tinham (por exemplo, sazonalidade, flutuação do nível), que tornaram impossível prever se o método de Medianas Móveis conseguiria o desempenho esperado.

Como proposta para estudos futuros, podemos sugerir um estudo mais aprofundado das características das séries reais, visando identificar aquelas que tornam estas séries mais adequadas à previsão por Medianas Móveis do que por Médias Móveis ou por Amortecimento Exponencial. Sugerimos também, o estudo da dependência das medidas de erro (MSE e MAPE) quanto aos métodos estudados, para verificarmos o fato de que o método por Medianas Móveis se destacou quando analisamos a medida de erro MAPE.

## REFERÊNCIAS

- [1] ALMEIDA, Phillippo Morais. **Avaliação de ferramentas para análise de séries temporais e métodos de previsão**. Trabalho de Conclusão de Curso (Graduação em Engenharia de Produção) – Universidade do Estado do Rio de Janeiro, RJ, 2008.
- [2] BATES, J. M., GRANGER, C. W. J., combination of forecasts", Journal of the Operational Research Society, v. 20, n. 4, pp. 451-468, 1969.
- [3] BUNN, D. W; A Bayesian approach to the linear combination of forecasts", Operational Research Quarterly, v. 26, pp. 325-329, 1975.
- [4] BUNN, D. W;The synthesis of forecasting models in decision analysis", Birkhauser. Basel, 1978.
- [5] CARNEIRO, Ana Cláudia Mancini da Silva. **Previsão de consumo de energia elétrica a curto prazo, usando combinações de métodos univariados**. Dissertação de Mestrado em Modelagem Computacional) – Programa de Pós-Graduação em Modelagem Computacional, Universidade Federal de Juiz de Fora, MG, 2014.
- [6] **Dados de precipitação média mensal em várias cidades do estado da Bahia**. Disponível em:<<http://www.inmet.gov.br/portal/>>. Acessado em: 06/08/2014 às 14:30.
- [7] FARIA, Elisângela Lopes; ALBUQUERQUE, Marcelo Portes; ALFONSO, Jorge Luiz González; ALBUQUERQUE, Márcio Portes; CAVALCANTE José Thadeu Pinto. **Previsão de Séries Temporais utilizando Métodos Estatísticos**. Rio de Janeiro, 2008.
- [8] HIPPERT, Henrique Steinherz. **Aulas de Séries Temporais**. 15 mar. 2013, 20 jul. 2013. Notas de Aula.
- [9] LEMOS, Fernando de Oliveira. **Metodologia para seleção de métodos de previsão de demanda**. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, RS, 2006.
- [10] LINDBERG, E; ZACKRISSON, U. Deciding about the Urcentain: The Use of Forecast as an Aid to Decision-making. **Scandinavian Journal of Management**. v. 7, n. 4, p. 271-283, 1991.
- [11] MAKRIDAKIS, S., WHEELWRIGHT, S. C., HYNDMAN, R. J., Forecasting - Methods and Applications. John Wiley Sons, 3a ed., 1998.
- [12] Nelder, J. A. and Mead, R. (1965) A simplex algorithm for function minimization. Computer Journal 7, 308–313.
- [13] Tukey, John (1977). Exploratory data analysis. Addison-Wesley.