

Energia Natural Afluente por subsistema do Brasil: uma análise de representatividade da Média de Longo Termo (MLT)

André Provenzano Naveiro

Departamento de Engenharia Industrial, PUC-Rio
Rua Marquês de São Vicente, 225 - Gávea, Rio de Janeiro - RJ
andrenaveiro@hotmail.com

Paula Medina Maçaira

Departamento de Engenharia Industrial, PUC-Rio
Rua Marquês de São Vicente, 225 - Gávea, Rio de Janeiro - RJ
paulamacaira@gmail.com

Fernando Luiz Cyrino Oliveira

Departamento de Engenharia Industrial, PUC-Rio
Rua Marquês de São Vicente, 225 - Gávea, Rio de Janeiro - RJ
cyrino@puc-rio.br

RESUMO

A Média de Longo Termo é a média das vazões naturais médias, correspondentes a um mesmo período, verificadas durante a série histórica de observações. Essa variável é de suma importância para o sistema energético Brasileiro, pois é utilizada nos modelos de otimização do despacho hidrotérmico. Este trabalho tem como objetivo a verificação da aderência dos valores mensais da Média de Longo Termo vigente com os valores observados em 2015 da série de Energia Natural Afluente. Para cumprir esse objetivo foram utilizados gráficos de séries temporais, histogramas e boxplots para verificação visual do comportamento da série além de inferência estatística com testes de hipótese. Foi possível então concluir que para todos os meses dos subsistemas Sul e Nordeste e alguns meses do subsistema Norte os valores considerados de MLT atualmente são satisfatórios, enquanto que para o subsistema Sudeste/Centro-Oeste nenhum mês apresentou aderência ao histórico de dados.

PALAVRAS CHAVE: Média de Longo Termo, Sistema Energético Brasileiro, Aderência.

Tópicos: EST, EN.

ABSTRACT

The Long Term Mean is the average of the natural inflow rate corresponding to the same period, recorded over the time series of observations. This variable is an important paramount to the Brazilian energy system, since it is used in the hydrothermal dispatch optimization models. This study aims to verify the compliance of the monthly values of the Long Term Mean currently used with the values observed in 2015 for the Natural Inflow Energy. To reach this goal it was used time series graphs, histograms and box plots to visually check the behavior of the series and statistical inference hypothesis test. It was concluded that for every month of the South and Northeast subsystems and some months of North subsystem the LTM values considered nowadays are satisfactory, while for the Southeast/Miwest subsystem no month showed adherence to the historical data.

KEYWORDS: Long Term Mean, Brazilian Energy System, Adherence.

Paper Topics: EST, EN.

1 Introdução

O Sistema Interligado Nacional (SIN), operado pelo Operador Nacional do Sistema (ONS), é o sistema de produção e transmissão de energia elétrica do Brasil, considerado um sistema hidrotérmico de grande porte, com forte predominância de usinas hidrelétricas (72,6% da geração total de energia em 2014) [ONS 2014]. Considerado único em âmbito global por conta de seu tamanho, o SIN possui múltiplos proprietários formados pelas empresas das regiões Sul, Sudeste, Centro-Oeste, Nordeste e parte da região Norte. Apenas 1,7% da energia produzida no país encontra-se fora do SIN, em pequenos sistemas isolado localizados principalmente na região amazônica [ONS 2014].

Planejar o setor elétrico brasileiro significa, basicamente, tomar decisões a cerca do despacho das hidroelétricas e termoeletricas, com o risco de perdas financeiras ou racionamento no caso de decisões equivocadas, como aconteceu em 2001, quando três dos quatros subsistemas foram afetados. Assim, o planejamento da operação energética envolve determinar metas para a geração das usinas hidrelétricas e termelétricas para cada etapa ao longo do horizonte de estudo, dada a demanda de energia, as restrições de operação das plantas e as restrições do sistema elétrico [Pereira 1989]. Como pode ser visto é bastante claro que a operação de um sistema elétrico como o brasileiro é uma tarefa bastante complexa dada a sua extensão e singularidade no mundo.

Uma das características mais importantes de um sistema de geração com predominância hidroelétrica é a forte dependência dos regimes de afluência. Como tal, a incerteza associada ao planejamento energético requer a modelagem estocástica das séries hidrológicas de forma adequada e coerente. Para o Brasil, os modelos que geram cenários hídricos, a fim de otimizar o desempenho da operação do sistema tem como variável estocástica a Energia Natural Afluente ou ENA [Oliveira, Souza e Marcato 2015]. A ENA consiste no potencial de energia que pode ser produzida a partir das vazões naturais afluentes aos reservatórios, seus valores são expressos em MW médios ou em percentual da média histórica de longo termo, MLT. Por sua vez, a partir do histórico de vazões, o ONS gera uma média de ENA para cada mês, sendo esse valor o representativo da média de longo prazo para o ano vigente.

A partir dos valores estimados de MLT são realizados os planejamentos de operação e geração de energia. Dada essa importância, o objetivo deste trabalho é o de verificar, a partir de testes estatísticos e visualização gráfica, se a MLT considerada está sendo representativa para cada subsistema ao comparar as vazões observadas em 2015 com os valores de MLT vigentes.

2 Material e Métodos

2.1 Análise Descritiva das Séries

O sistema energético brasileiro é dividido em quatro grandes subsistemas: Sudeste/Centro-Oeste (SE/CO), Sul (S), Nordeste (NE) e Norte (N). O subsistema SE/CO abrange as regiões Sudeste e Centro-Oeste do país, além dos estados de Rondônia e Acre. Já o subsistema Sul abrange a região Sul do país, enquanto que o subsistema Norte é responsável pela região Norte com exceção do estado do Maranhão. Por fim, o subsistema Norte cobre parte dos estados do Amapá, Pará, Tocantins, Maranhão e Amazonas. As séries de Energia

Natural Afluente serão tratadas de forma a representar os grandes subsistemas, sendo assim, as análises serão feitas em relação a quatro diferentes séries temporais.

Tanto as séries de Energia Natural Afluente quanto os valores da Média de Longo Termo atualmente utilizadas podem ser obtidas através do site do Operador Nacional do Sistema Elétrico [ONS 2014]. Todas as análises aqui realizadas foram feitas a partir do software R [R Core Team 2015] e dos pacotes knitr [Xie 2016] [Xie 2015] [Xie 2014], kfigr [Koochafkan 2015], TSA [Chan & Ripley 2012] e tseries [Trapletti & Hornik 2015].

No presente capítulo será realizada a análise comportamental das quatro séries de ENA onde serão apresentados gráficos para facilitar a visualização do leitor, medidas estatísticas e também testes de hipóteses. Na Figura 1 é possível observar a evolução temporal da ENA de cada subsistema no período de janeiro de 1931 a dezembro de 2015. Identifica-se primeiramente um padrão semelhante entre os subsistemas SE/CO, NE e N, em que se evidencia o caráter sazonal dessas séries. Já no subsistema Sul, esse padrão sazonal não se mostra de forma clara, onde observa-se valores extremos (outliers) em diversos períodos. Apesar das diferenças, uma constante em todos os subsistemas é o caráter não tendencioso dessas séries, de modo que no longo-prazo, não existe uma expectativa de aumento, nem de redução do nível da série.

Histórico Energia Natural Afluente (MW médio)

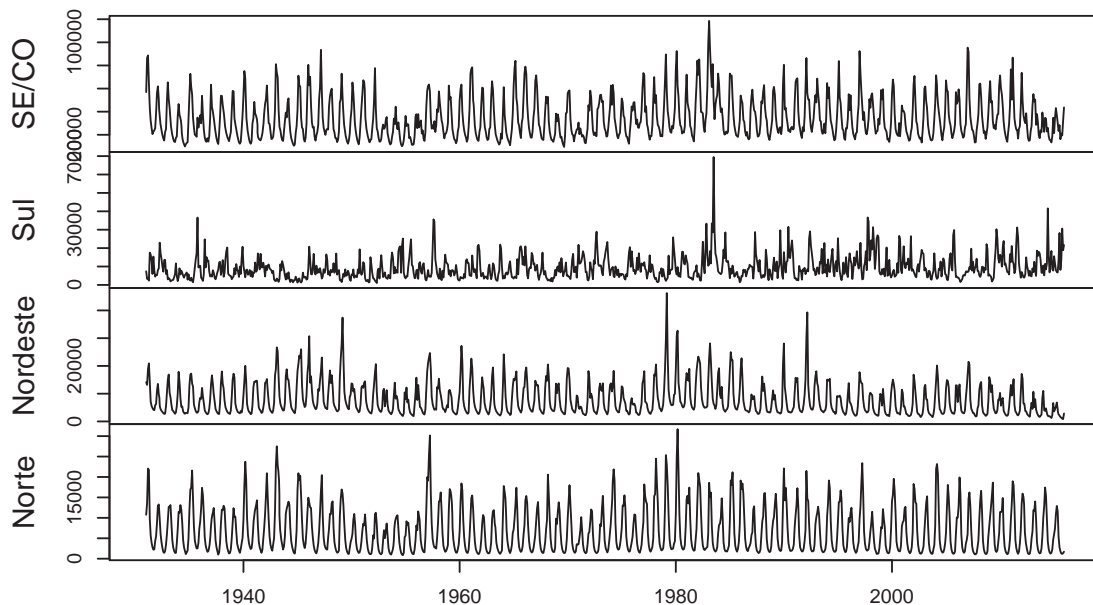


Figura 1: Gráficos das séries de Energia Natural Afluente

Na Tabela 1 são apresentadas algumas das mais importantes medidas estatísticas para cada subsistema como média, mediana, desvio padrão, coeficiente de variação, assimetria,

curtose, mínimo, máximo e quartis. Para maiores detalhes ver [Morettin e Bussab 2004]. Constata-se que o subsistema SE/CO possui as maiores medidas centrais (média e mediana), o que é consistente dado que é o subsistema que apresenta os maiores valores de ENA para todos os períodos. Nota-se também que apesar do subsistema SE/CO ter o maior valor de desvio padrão, o subsistema Norte é o que apresenta a maior variância relativa, com um coeficiente de variação de 0.82, superando os demais subsistemas.

Tabela 1: Análise descritiva das séries de Energia Natural Afluente

	SE/CO	S	NE	N
Média	34398.92	9371.5	8032.95	6988.22
Mediana	28911.00	7658.0	5704.00	4827.00
Desv. Pad.	17936.25	6550.1	5965.31	5697.59
Coef. Var.	0.52	0.7	0.74	0.82
Assimetria	2.10	2.1	2.10	2.10
Curtose	4.99	5.0	4.99	4.99
Min.	9501.00	1043.0	846.00	868.00
Max.	118545.00	69521.0	46244.00	31706.00
Q1	20755.00	4735.5	3622.50	2276.00
Q3	45493.75	11825.0	11390.75	10831.50

Os valores de assimetria quantificam quão simétrica é a distribuição dos dados, enquanto a medida de curtose quantifica se o formato da distribuição dos dados coincide com a distribuição Gaussiana. Se a assimetria for igual a 0, então os dados são perfeitamente simétricos, porém é aceitável que valores entre -1 e 1 também indiquem simetria dos dados. É esperado que uma distribuição Gaussiana possua valor de curtose igual a 3 (mesocúrtica), sendo assim, valores abaixo de 3 indicam uma distribuição mais “achatada” (platicúrtica) enquanto curtose acima de 3 representa distribuições mais “altas” (afuniladas) e são chamadas de leptocúrtica. Sendo assim, como verifica-se que todas as séries possuem valores semelhantes de assimetria (2.1) e curtose (4.99) espera-se que a distribuição dos dados não seja simétrica e possua um formato mais “afunilado”.

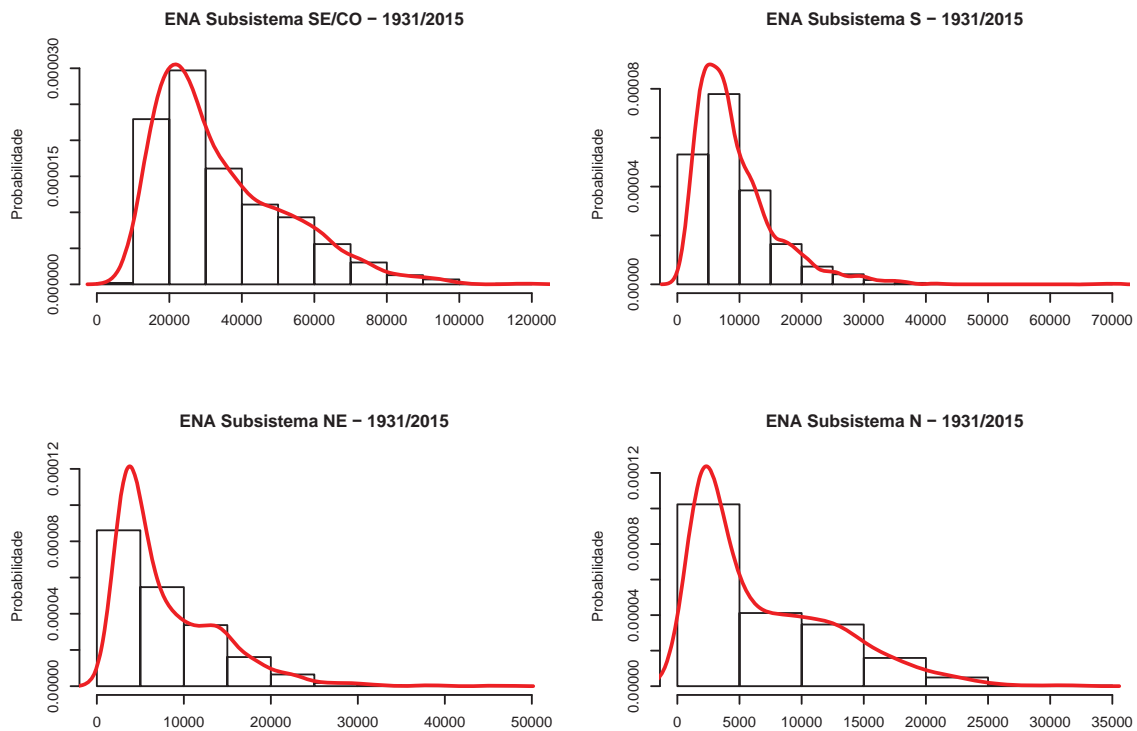


Figura 2: Histograma das séries de Energia Natural Afluente

O comportamento descrito pode ser confirmado pela Figura 2 que apresenta o histograma das ENA's conjuntamente com as funções de probabilidade de densidade adequadas. Concluindo assim que as séries dos subsistemas são assimétricas positivas e leptocúrticas, caracterizadas por distribuições de caudas pesadas, que apresentam uma distribuição afunilada e mais concentrada que a distribuição Normal (Gaussiana).

Para se analisar a distribuição das ENA observadas, foi realizado o teste de Wilcoxon-Mann-Whitney [Corder e Foreman 2014], cujo propósito consiste em determinar se duas distribuições de probabilidade subjacentes diferem uma da outra ou se assemelham. Isto é, trata-se do seguinte teste de hipótese: H_0 série de dados segue distribuição X versus H_1 série de dados não segue distribuição X .

Sendo assim, adotando nível de significância de 5%, se o pvalor encontrado for menor que 0.05 rejeita-se a hipótese dos dados possuírem distribuição X , enquanto que pvalor maior que 5% não rejeita-se a hipótese nula dos dados seguirem a distribuição X .

Tabela 2: Teste de aderência Wilcoxon-Mann-Whitney

	SE/CO	Sul	Nordeste	Norte
Normal	0.075	0.00	0.045	0.36
Qui-Quadrado	0.000	0.00	0.000	0.00
Gama	0.968	0.52	0.560	0.72

	SE/CO	Sul	Nordeste	Norte
Weibull	0.000	0.00	0.000	0.00
LogNormal	0.000	0.00	0.000	0.00

De acordo com a Tabela 2 e o teste de hipóteses descrito acima, ao nível de 5% de significância todas as séries de ENA obtiveram valores significantes para aderência a distribuição Gama, isto é, não rejeita-se que os dados em estudo possuam uma distribuição Gama.

Outra informação que pode ser extraída da Tabela 1 é referente ao comportamento dos quartis das séries. Um quartil é qualquer um dos três valores que divide o conjunto ordenado de dados em quatro partes iguais, e assim cada parte representa 1/4 da amostra ou população, assim quartil 25% = Q1, quartil 50% = Q2 = Mediana e quartil 75% = Q3. Por exemplo, para o caso do subsistema SE/CO, 25% dos dados são menores que 20.755 MW médio. Esse tipo de estatística auxilia na análise de assimetria dos dados e pode ser representada visualmente pelo gráfico de Boxplot, como na Figura 3. Note que para todos os subsistemas a média está mais próxima do quartil 25% do que do quartil 75%, identificando uma assimetria já confirmada. Destaca-se também a dissimilaridade em valores absolutos do subsistemas Sudeste/Centro-Oeste para os demais.

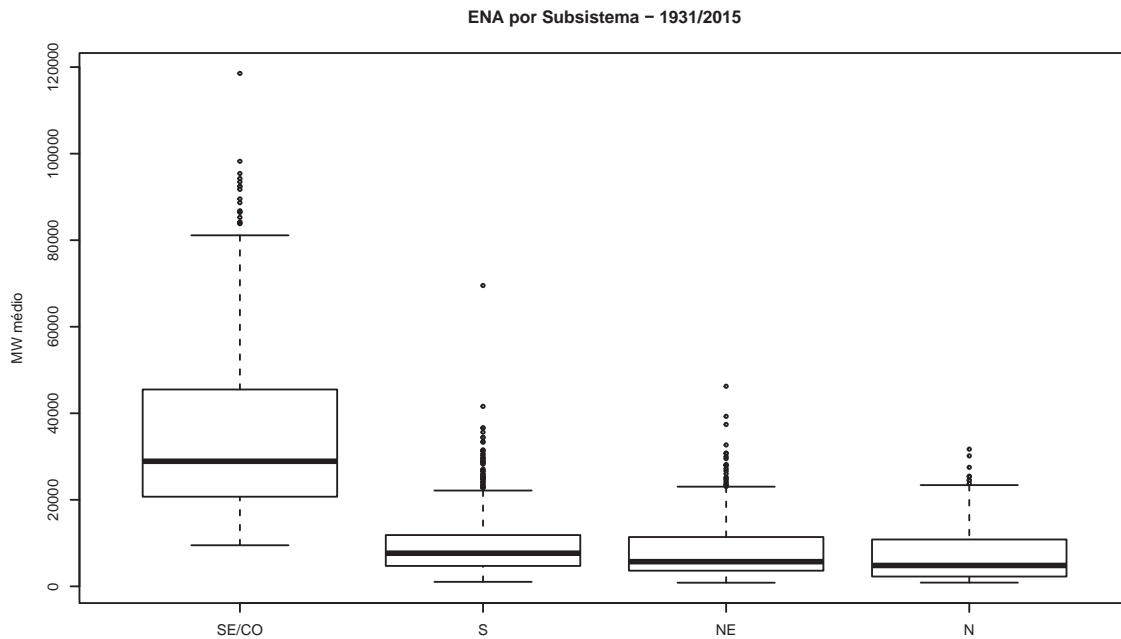


Figura 3: Boxplot das séries de Energia Natural Afluente

A série de Energia Natural Afluente, conforme visto anteriormente, é uma variável hidrológica que sofre total influência dos regimes de chuva, com isso é possível observar períodos ditos secos e úmidos. Sendo assim, é interessante analisar os valores históricos para cada mês.

Na Figura 4 estão representadas mensalmente os valores de ENA para cada um dos subsistemas em formato de boxplot. Pode-se observar que em três subsistemas, SE/CO, Nordeste e Norte, o caráter sazonal da ENA se faz presente onde nos meses entre Maio e Novembro há um período de seca e de Dezembro a Abril há um período mais úmido.

Uma grande singularidade no Brasil é a grande diferença dos regimes de chuva entre os subsistemas. Note que o subsistema Sul se difere amplamente dos demais, não contendo nenhum mês tipicamente seco ou úmido, fato evidenciado na grande semelhança dos boxplot de todos os meses. Esse comportamento, é um indicativo da falta de sazonalidade desse subsistema, se contrapondo ao restante do país.

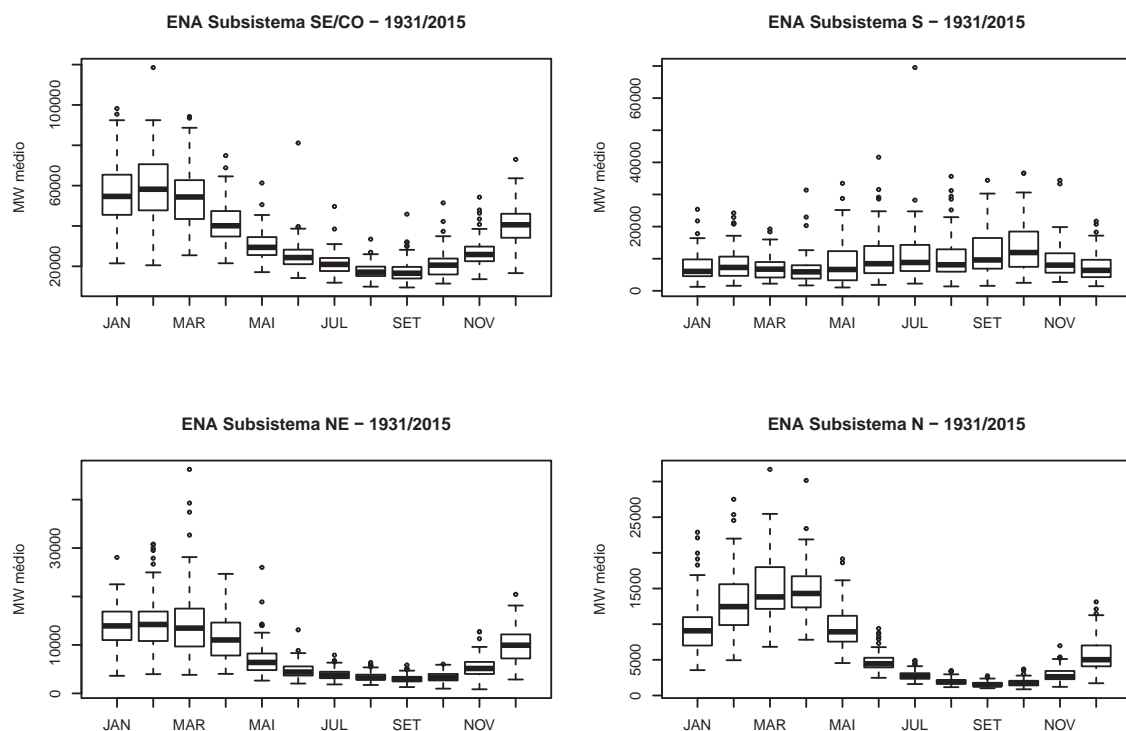


Figura 4: Boxplot das séries de Energia Natural Afluente

2.2 Aplicação e Resultados

Como dito na Introdução, o principal objetivo deste trabalho é a verificação da aderência da Média de Longo Termo vigente com os dados observados de Energia Natural Afluente para o ano de 2015. Sendo assim o primeiro passo será a comparação entre a ENA observada em 2015 e a MLT vigente utilizando o gráfico de barras conforme pode ser visto na Figura 5. Note que, com exceção do subsistema Sul, a MLT superestima todos os meses quando comparada com a ENA observada, chegando a uma diferença de ser realizada somente 15% da Energia Natural Afluente esperada pela MLT do mês de novembro no subsistema Nordeste como pode ser visto na Tabela 3.

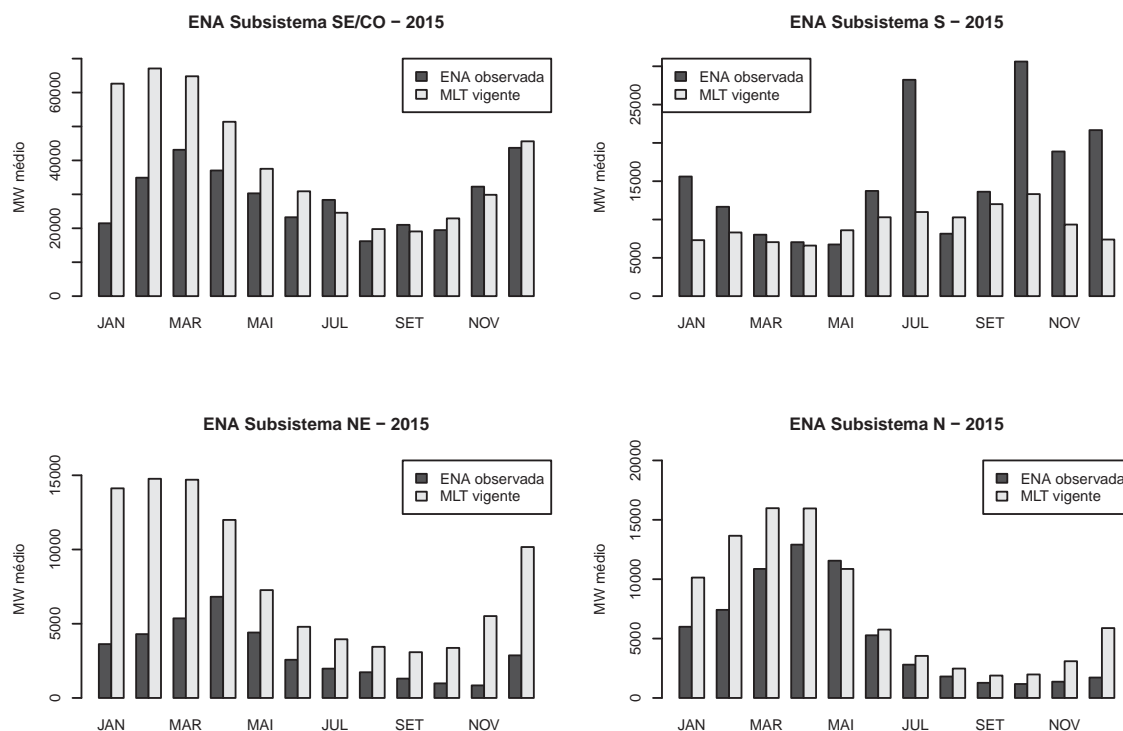


Figura 5: Comparação entre: ENA observada e MLT vigente

Tabela 3: % da ENA observada em relação a MLT vigente

	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
SE/CO	34	52	67	72	81	75	115	82	110	85	108	96
Sul	214	140	114	107	78	133	257	79	113	230	202	293
Nordeste	26	29	36	57	61	54	50	50	42	29	15	28
Norte	59	54	68	81	106	92	79	73	67	59	44	29

Visualmente concluiu-se que a média de longo termo não é representativa da Energia Natural Afuente, pois superestima a ENA esperada para três dos quatro subsistemas. Para uma primeira confirmação da hipótese de sobrestimação será construído um intervalo interquartilico (entre os quartis 1 e 3) para cada mês de cada subsistema utilizando o histórico da série de 1931 até 2014. Caso o valor da MLT vigente daquele mês se encontre dentro do intervalo interquartilico então pode-se dizer que os valores de MLT são válidos, caso contrário, os valores de MLT não são adequados.

Veja na Figura 6 que para os subsistemas Sul e Nordeste nenhum valor da MLT violou o intervalo interquartilico, enquanto que para os subsistemas Sudeste/Centro-Oeste e Norte uma sequência de meses da MLT ultrapassou o limite superior, confirmando uma superestimação dos valores esperados de Energia Natural Afuente.

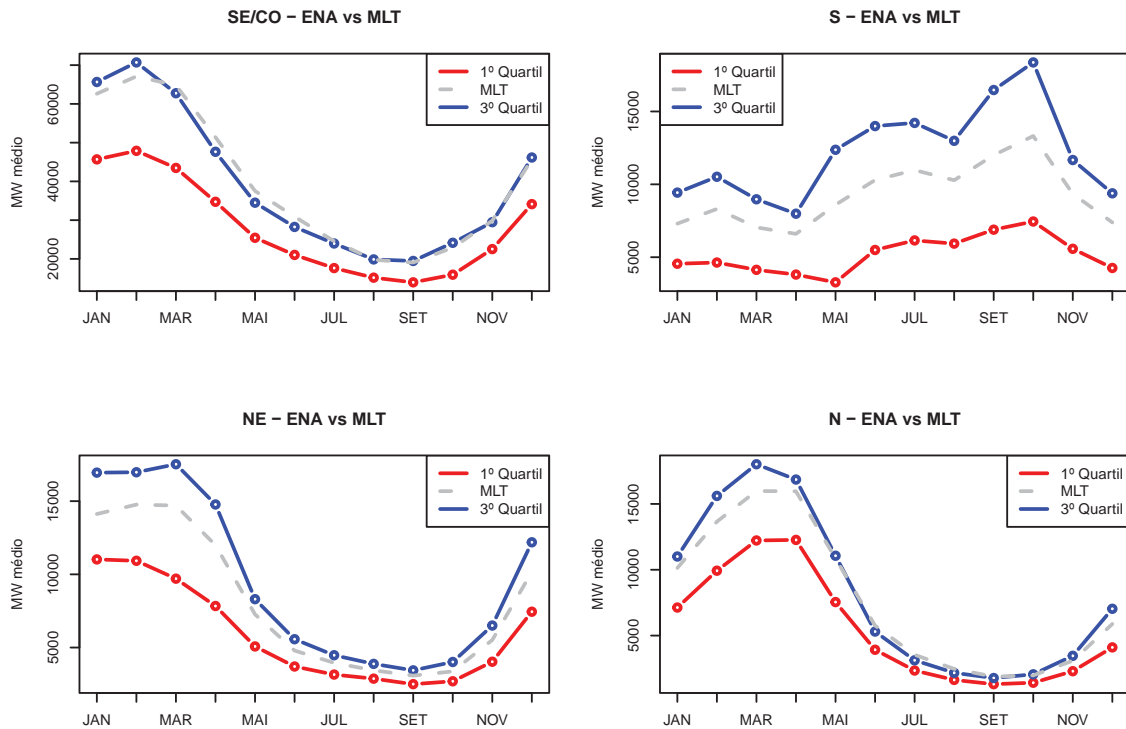


Figura 6: Intervalo interquartilício versus Média de Longo Termo

Um último teste para verificar a aderência dos valores vigentes de MLT a série histórica da ENA é o teste para a média (teste T-Studente), mais detalhes em [McDonald 2008]. As seguintes hipóteses são testadas:

$$H_0: \text{média ENA } m\hat{e}s_i = MLT_i$$

$$H_1: \text{média ENA } m\hat{e}s_i \neq MLT_i$$

Considerando, mais uma vez, um nível de significância de 5% observa-se na Tabela 4 que a Média de Longo Termo vigente para todos os meses dos subsistemas Sul e Nordeste é aderente a média da série histórica de ENA, para o subsistema Nordeste 7 de 12 meses (58.3%) da MLT não foram significativamente iguais a média histórica da ENA. Já para o subsistema Sudeste/Centro-Oeste, segundo o teste de hipóteses, nenhum mês da MLT foi representativo em relação à série histórica.

Tabela 4: Teste de t-Student para comparação de médias

	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
SE/CO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.03	0.00	0.00
Sul	0.98	0.98	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.98	0.98
Nordeste	1.00	1.00	0.99	0.99	0.99	0.97	0.94	0.93	0.93	0.97	0.99	1.00
Norte	0.41	0.23	0.08	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.52

3 Conclusão

Dada a importância da variável Média de Longo Termo, este trabalho teve como objetivo a aplicação de testes estatísticos para averiguar a aderência de tal variável a série histórica de Energia Natural Afluente. Para atingir tal objetivo foram empregadas técnicas estatísticas de visualização via gráficos de séries temporais, histogramas e boxplots além de testes de hipóteses.

Em um primeiro momento foi estudado o comportamento das quatro séries de ENA, uma para cada subsistema. Conclui-se que, com exceção do subsistema Sul, as séries temporais possuem comportamento sazonal, identificando os períodos ditos secos (menos chuva) e úmidos (mais chuva). Também foi constatado que a distribuição de probabilidade dos quatro subsistemas era semelhante e confirmado via teste de hipóteses que tais distribuições poderiam ser consideradas advindas da distribuição Gama.

Para início da comparação entre os valores atualmente utilizados de Média de Longo Termo e a Energia Natural Afluente observada em 2015 foi utilizado um gráfico de barras, mostrando que para três dos quatro subsistemas o valor da MLT sobrestimou o realizado. Em seguida foi construído um intervalo de confiança a partir dos intervalos interquartílicos mensais das séries históricas e comparado com o valor vigente mensal da MLT. Tal exercício mostrou mais uma vez que a variável MLT para os subsistemas Sul e Nordeste é completamente satisfatória enquanto que para os subsistemas SE/CO e Norte vários meses não foram contemplados com a MLT. A última confirmação da aderência foi feita a partir do teste de comparação de média (t-student) onde a hipótese nula é de igualdade entre a média histórica da ENA e o valor vigente da MLT daquele mês. Mais uma vez o subsistema Sul e Nordeste obteve valores significantes para igualdade entre as médias, alguns meses subsistema Norte apresentaram o mesmo resultado, porém o subsistema SE/CO rejeitou a variável MLT para todos os meses.

Conclui-se então que a partir das análises e testes empregados a variável Média de Longo Termo é aderente às séries históricas dos subsistemas Sul e Nordeste, porém tal variável deve ser revista para os demais subsistemas, dado que os testes apontaram a não aderência da MLT para esses casos.

Referências

- Trapletti, A. and Hornik, K. 2015. tseries: Time Series Analysis and Computational Finance. R package version 0.10-34.
- Corder, G. W. and Foreman, D. I. 2014. "Nonparametric Statistics: A Step-by-Step Approach". Wiley. ISBN 978-1118840313.
- Chan, K. and Ripley, B. 2012. TSA: Time Series Analysis. R package version 1.01. <https://CRAN.R-project.org/package=TSA>
- McDonald, J. H. 2008. Handbook of Biological Statistics Sparky House Publishing. Baltimore.
- Koohafkan, M. C. 2015. kfigr: Integrated Code Chunk Anchoring and Referencing for R Markdown Documents. R package version 1.2. <https://CRAN.R-project.org/package=kfigr>
- Morettin, P. A. and Bussab, W. de O. 2004. Estatística Básica. 5ª edição, Editora Saraiva.

- Oliveira, F. L. C., Souza, R. C. and Marcato, A. L. M. 2015. A Time Series Model for Building Scenarios Trees Applied to Stochastic Optimisation. *International Journal of Electrical Power & Energy Systems* 67:16-38.
- ONS. 2014. Operador Nacional Do Sistema Elétrico. www.ons.com.br.
- ONS. 2015. Avaliação de Curto Prazo da Operação, Energias Naturais Afluentes dos Subistemas. Operador Nacional do Sistema Elétrico. www.ons.org.br/operacao/enas_subistemas.aspx
- Pereira, M.V.F. 1989. Optimal Stochastic Operations Scheduling of Large Hydroelectric Systems. *International Journal of Electric Power and Energy Systems* 11: 161-69.
- R Core Team. 2015. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. www.R-project.org/.
- Xie, Y. 2016. knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.12.3.
- Xie, Y. 2015. *Dynamic Documents with R and knitr*. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963
- Xie, Y. 2014. knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC. ISBN 978-1466561595