

8. Referências Bibliográficas

AKAIKE, H., 1974. A New Look at the Statistical Model Identification. *I.E.E.E. Transactions on Automatic Control*, pp. 716-723.

AKAIKE, H., 1976. Canonical Correlation Analysis of Time Series and the Use of an Information Criterion. In: R. K. M. a. D. G. Lainotis, ed. *System Identification: Advances and Case Studies*. New York: Academic Press, pp. 52-107.

ALMEIDA, A., ELLIAN, S. & NOBRE, J., 2008. Modificações e Alternativas aos Testes de Levene e de Brown e Forsythe para Igualdade de Variâncias e Médias. *Revista Colombiana Estadística*, 31(2), pp. 241-260.

ANDERSON, P. L. & VECCHIA, A., 1993. Asymptotic results for periodic autoregressive moving-average process. *Journal of Time Series Analysis*, 14(1), pp. 1-18.

ANDERSON, T. W. & DARLING, D. A., 1952. Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann. Math. Stat.*, Volume 23, pp. 193-212.

BARTLETT, M. S., 1946. On the theoretical specification and sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society*, Volume 27, pp. 27-41.

BARTOLINI, P., SALAS, J. & OBEYSEKERA, J. T. B., 1988. Multivariate periodic ARMA(1,1) process. *Water Resources Research*, 25(8), pp. 1237-1246.

BASAWA, I. V. & LUND, R., 2001. Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis*, 22(6), pp. 651-663.

BELLMAN, R. E., 1957. *Dynamic Programming*. s.l.:Princeton University Press.

BENDERS, J. F., 1962. Partitioning Procedures for Solving Mixed Variables Programming Problems. *Numerische Mathematik*, Volume 4, pp. 238-252.

BOSHNAKOV, G. N., 1996. Recursive Computation of the parameters of periodic autoregressive moving-average process. *Journal of Time Series Analysis*, 17(4), pp. 333-249.

BOX, G. E. P., JENKINS, G. M. & REINSEL, G. C., 2008. *Time Series analysis: forecasting and control*. s.l.:Wiley.

BRANDI, R. B. S., 2011. *Processo Iterativo de Construção da Função de Custo Futuro na Metodologia PDE-ConvexHull*. Juiz de Fora: Dissertação de Mestrado, Departamento de Engenharia Elétrica, Universidade Federal de Juiz de Fora.

BURNHAM, K. P. & ANDERSON, D. R., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. ed. s.l.:Springer-Verlag.

CARPENTER, J. & BITHELL, J., 2000. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, Volume 19, pp. 1141-1164.

CASELA, G. & BERGER, R. L., 2001. *Statistical Inference*. 2nd ed. s.l.:Duxbury Press.

CASTRO, C. M. B., 2012. *Planejamento Energético da Operação de Médio Prazo Conjugando as Técnicas de PDDE, PAR(p) e Bootstrap*. Juiz de Fora: Tese de Doutorado em Engenharia Elétrica, Universidade Federal de Juiz de Fora.

CEPEL, 2001. *Manual de referência do Modelo NEWAVE*, Rio de Janeiro: s.n.

CEPEL, 2012. *Eletrobras Cepel*. [Online] Available at: <http://www.cepel.br/servicos/descprog.shtm> [Acesso em 01 Novembro 2012].

CHARBENEAU, R. J., 1978. Comparison of the Two and Three-Parameter Log Normal Distributions Used in Streamflow Synthesis. *Water Resources Research*, 14(1), pp. 149-150.

CHIRALAKSANAKUL, A., 2003. *Monte Carlo Methods for Multi-stage Stochastic Programs*. Austin(Texas): PhD Dissertation, The University of Texas at Austin.

CHVÁTAL, V., 1983. *Linear Programming*. New York: W. H. Freeman.

CONOVER, W. U., 1971. *Practical Nonparametric Statistics*. New York: John Wiley & Sons.

CONSIGLI, G. & DEMPSTER, M. A. H., 1998. Dynamic stochastic programming for asset-liability management. *Annals of Operations Research*, Volume 81, pp. 131-161.

da COSTA, J., de OLIVEIRA, G. & LEGEY, L., 2006. *Reduced Scenario Tree Generation for Mid-term Hydrothermal Operation Planning*. Stockholm, Sweden, 9th International Conference on Probabilistic Methods Applied to Power Systems, pp. 1-7.

da COSTA, J. P., 2007. *Árvore de Cenários para o Planejamento da Operação de Médio Prazo*. Rio de Janeiro: Tese de Doutorado, COPPE/PPE.

da SILVA, E. L. & FINARDI, E. C., 2001. Planning of Hydrothermal Systems Using a Power Plant Individualistic Representation. *IEEE Porto Power Tech Conference*.

DIAS, B. H. et al., 2010. Stochastic Dynamic Programming Applied to Hydrothermal Power Systems Operation Planning Based on the Convex Hull Algorithm. *Mathematical Problems in Engineering*, Volume 2010, pp. 1-21.

EFRON, B. & TIBSHIRANI, R. J., 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall.

ENGLE, R. F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), pp. 987-1007.

FERREIRA, P. G. C., 2013. *A estocasticidade associada ao Setor Elétrico Brasileiro e uma nova abordagem para a geração de afluências via Modelos Periódicos Gama*. Rio de Janeiro: Tese de Doutorado em Engenharia Elétrica, PUC-Rio.

GAMERMAN, D., 1997. *Markov Chain Monte Carlo – Stochastic Simulation for Bayesian Inference*. s.l.:Chapman & Hall.

GELLENS, D., 2002. Combining regional approach and data extension procedure for assessing GEV distribution of extreme precipitation in Belgium. *Journal of Hydrology*, Volume 268, pp. 113-126.

GIBBONS, J. D. & CHAKRABORTI, S., 2003. *Non Parametric Statistical Inference*. Fourth Edition ed. New York: Taylor & Francis.

GIBBS, W. R., RICHARDSON, S. & SPIEGELHALTER, D., 1998. *Markov Chain Monte Carlo in Practice*. s.l.:Chapman & Hall.

GLADYSEV, E. G., 1961. Periodically correlated random sequences. *Sov. Math. Dokl.*, Volume 2, pp. 385-388.

HANNAN, E. J. & QUINN, B. G., 1979. The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society B*, Issue 41, pp. 190-195.

HIPEL, K. W. & McLEOD, A. I., 1994. *Time series modelling of water resources and environmental systems*. Amsterdan: Elsevier.

HURVICH, C. M. & TSAI, C. L., 1989. Regression and time series model selection in small samples. *Biometrika*, Issue 76, pp. 297-307.

HURVICH, C. M. & TSAI, C. L., 1991. Bias of the corrected AIC criterion for underfitted regression. *Biometrika*, 78(3), pp. 499-509.

HUYNH, T. H., LAI, S. V. & SOUMARÉ, I., 2008. *Stochastic Simulation and Applications in Finance with MATLAB programs*. Chichester: John Willey & Sons Ltd.

JACKMAN, S., 2009. *Bayesian Analysis for the Social Sciences*. s.l.:Wiley.

KAPLAN, D., 2000. *Structural equation modeling: Foundation and extensions*. Thousand Oaks, CA: Sage Publications.

KATZ, R. W., 1999. Extreme value theory for precipitation. *Advances in Water Resources*, Volume 23, pp. 133-139.

KATZ, R. W., PARLANGE, M. B. & NAVEAU, P., 2002. Statistics of extremes in hydrology. *Advances in Water Resources*, Volume 25, pp. 1287-1304.

KONISHI, S. & KITAGAWA, G., 2008. *Information criteria and statistical modeling*. New York: Springer.

KUHA, J., 2004. AIC and BIC : Comparisons of Assumptions and Performance. *Sociological Methods & Research*, 33(2), pp. 188-229.

LI, W. K. & HUI, Y. V., 1988. An algorithm for the exact likelihood of periodic autoregressive moving-average models. *Commun. Statist. Simula*, 17(4), pp. 1483-1484.

LOPES, H. C. V., 2012. *Simulação Estocástica: notas de aula*. [Online] Available at: <http://www-di.inf.puc-rio.br/~lopes/courses/inf2035---simulacao-estocas.html>

[Acesso em 29 11 2012].

LUND, R. B. I. V., 2000. Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis*, 21(1), pp. 75-93.

LUND, R., SHAO, Q. & BASAWA, I., 2006. Parsimonious Periodic Time Series Modeling. *Australian & New Zealand Journal of Statistics*, Volume 48, pp. 33-47.

MACEIRA, M. E. P., 1989. *Operação Ótima de Reservatórios com Previsão de Afluências*. Rio de Janeiro: Dissertação de Mestrado, COPPE/UFRJ.

MACEIRA, M. E. P. et al., 2002. Chain of Optimization models for Setting the Energy Dispatch and Spot price in the Brazilian System. *14th PSCC*.

MARCATO, A. L. M., 2002. *Representação híbrida de sistemas equivalentes e individualizados para o planejamento da operação a médio prazo de sistemas de potência de grande porte*. Rio de Janeiro: Tese de Doutorado, DEE, PUC-Rio.

MARSAGLIA, G., 1995. *DIEHARD: a battery of tests of randomness*. [Online] Available at: <http://www.stat.fsu.edu/pub/diehard> [Acesso em 29 11 2012].

MARTINS, E. S. & STEDINGER, J. R., 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3), pp. 737-744.

MATOS, V. L., 2008. *Análise comparativa entre as modelagens de reservatório equivalente de energia agregado por Subsistema e por cascata no problema do Planejamento anual da operação energética*. s.l.:Dissertação de Mestrado, Departamento de Engenharia Elétrica, UFSC.

MATSUMOTO, N. & NISHIMURA, T., 1998. Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1), pp. 3-30.

MAYDEU-OLIVARES, A. & GARCIA-FORERO, C., 2010. Goodness-of-Fit Testing. *International Encyclopedia of Education*, Volume 7, pp. 190-196.

MAZZOTTI, B. F., 2010. *Co-projeto de hardware/software do filtro de partículas para localização em tempo real de robôs móveis*. São Carlos: Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. Dissertação de Mestrado em Ciências de Computação e Matemática Computacional.

McLEOD, A. I., 1994. Diagnostic checking of periodic autoregression models with application. *Journal of Time Series Analysis*, 15(2), pp. 221-233.

METROPOLIS, N. U. S., 1949. The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247), pp. 335-341.

MORRISON, J. & SMITH, J. A., 2002. Stochastic modeling of flood peaks using the generalized extreme value distribution. *Water Resources Research*, 38(12), pp. 1-12.

MULAIK, S. A. et al., 1989. Evaluation of Goodness-of-Fit Indices for Structural Equation Models. *Psychological Bulletin*, 105(3), pp. 430-445.

NI, E., GUAN, X. & LI, R., 1999. Scheduling Hydrothermal Power Systems with Cascaded and Head-Dependent Reservoirs. *IEEE Trans. Power Systems*, 14(3), pp. 1127-1132.

OBEYSEKERA, J. T. B. & SALAS, J. D., 1982. ARMA model identification of hydrologic time series. *Water Resources Research*, 18(4), pp. 1001-1021.

OLIVEIRA, F. L. C., 2010. *Nova abordagem para geração de cenários de afluências no planejamento da operação energética de médio prazo*. Rio de Janeiro: Dissertação de Mestrado, DEE, PUC-Rio.

OLIVEIRA, F. L. C. & SOUZA, R. C., 2011. A new approach to identify the structural order of par (p) models. *Pesquisa Operacional*, 31(3), pp. 487-498.

OLIVEIRA, W. L., 2007. *Redução ótima de cenários em programação estocástica. Aplicação às vazões afluentes aos aproveitamentos hidroelétricos*. s.l.:Dissertação de Mestrado, IMPA.

ONS, 2012. www.ons.org.br. [Online] [Acesso em Agosto 2012].

PAGANO, M., 1978. On periodic and multiple autorregressions. *The Annals of Statistics*, 6(6), pp. 1310-1317.

PARZEN, E. & PAGANO, M., 1979. A approach to modeling seasonally stationary time series. *Journal of Econometrics*, Volume 9, pp. 137-153.

PENNA, D. D. J., 2009. *Definição da árvore de cenários de afluências para o planejamento da operação energética de médio prazo*. Rio de Janeiro: Tese de Doutorado, DEE, PUC-Rio.

PEREIRA, M., CAMPODÓNICO, N. & KELMAN, R., 1998. Long-term Hydro Scheduling based on Stochastic Models. *Proceedings of EPSOM Conference*.

PEREIRA, M. V. F., 1989. Optimal stochastic operations scheduling of large hydroelectric systems. *International Journal of Electric Power and Energy Systems*, 11(3), pp. 161-169.

PEREIRA, M. V. F. & PINTO, L. M. V. G., 1985. Stochastic Optimization of a Multireservoir Hydroelectric System: a Decomposition Approach. *Water Resources Research*, June, 21(6), pp. 779-792.

PEREIRA, M. V. F. & PINTO, L. M. V. G., 1991. Multi-Stage Stochastic Optimization Applied to Energy Planning. *Mathematical Programming*, 52(3), pp. 359-375.

PEREIRA, M. V. P., 1985. Optimization Scheduling of Hydrothermal Systems – an Overview. *IFAC Symposium on Planning and Operation of Electric Energy Systems*.

PINTO, L. M. V. G. & PEREIRA, M. V. F., 1983. Application of Decomposition Techniques to the Mid- and Short Term Scheduling of Hydrosystems. *IEEE Trans. Power Appar. Syst.*, 102(11), pp. 3611-3618.

PLACKETT, R. L., 1983. Karl Pearson and the Chi-Squared. 51(1), pp. 59-72.

QUADROS, L. E., QUEIROZ, M. M. F. & VILAS BOAS, M. A., 2011. Distribuição de frequência e temporal de chuvas intensas. *Acta Sci., Agron. (Online)*, 33(3), pp. 401-410.

RAMESH, N. I. & DAVISON, A. C., 2002. Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology*, Volume 256, pp. 106-119.

RAMOS, T. P., 2011. *Modelo Individualizado de Usinas Hidrelétricas Baseado em Técnicas de Programação Não Linear com o Modelo de Decisão Estratégica*. Juiz de Fora: Dissertação de Mestrado, Departamento de Engenharia Elétrica, Universidade Federal de Juiz de Fora.

RASMUSSEN, R. F. et al., 1996. Estimation and validation of contemporaneous PARMA models for streamflow simulation. *Water Resources Research*, 32(10), pp. 3151-3160.

REIS, R. L., 2013. *Modelos autorregressivos periódicos para previsão e geração de séries de vazões médias mensais*. São Paulo: Tese de Doutorado, Universidade de São Paulo.

ROBERT, C. P. & CASELLA, G., 2004. *Monte Carlo Statistical Methods*. New York: Springer.

RUKHIN, A. et al., 2010. *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. s.l.:National Institute of Standards and Technology.

SALAS, J. D., BOES, D. C. & SMITH, R. A., 1982. Estimation of ARMA models with seasonal parameters. *Water Resources Research*, 18(4), pp. 1006-1010.

SCHWARZ, G., 1978. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), pp. 461-464.

SHAHJAJAN MONDAL, M. W. S. A., 2006. Generating and forecasting monthly flows of the Ganges river with PAR model. *Journal of Hydrology*, Volume 323, pp. 41-46.

SILVA, D. N., 1995. *O método Bootstrap e aplicações à Regressão Múltipla*. Campinas: Dissertação de Mestrado, IMEEC, UNICAMP.

SILVA, E. L., 2001. *Formação de preços em mercados de Energia Elétrica*. s.l.:Sagra Luzzato.

SOARES, M. P., 2006. *Otimização multicritério da operação de sistemas hidrotérmicos utilizando algoritmos genéticos*. Rio de Janeiro: Dissertação de Mestrado, DEE, PUC-Rio.

SOUZA, R. C. & CAMARGO, M. E., 2004. *Análise e Previsão de Séries Temporais: os modelos ARIMA*. Ijuí: SEDIGRAF.

SOUZA, R. C., MARCATO, A. L. M., DIAS, B. H. & OLIVEIRA, F. L. C., 2012. Optimal Operation of Hydrothermal Systems with Hydrological Scenario Generation through Bootstrap and Periodic Autorregressive Models. *European Journal of Operational Research*, 222(3), pp. 606-615.

STEPHENS, M. A., 1977. Goodness of fit for the extreme value distribution. *Biometrika*, 64(3), pp. 583-599.

SUGIURA, N., 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods*, 7(1), pp. 13-26.

TERRY, L. A. et al., 1986. Coordinating the Energy Generation of the Brazilian National Hydrothermal Electrical Generating System. *Interfaces*, Volume 16, pp. 16-38.

THOMAS, H. A. & FIERING, M. B., 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. *Design of Water Resource Systems*, pp. 459-463.

TROUTMAN, B. M., 1979. Some results in periodic autorregression. *Biometrika*, Volume 66, pp. 219-228.

ULA, T. A., 1993. Forecasting of multivariate periodic autoregressive moving average process. *Journal of Time Series*, 14(6), pp. 645-657.

UNNY, T. E. & COVER, K., 1985. Application of Computer Intensive Statistics to parameters uncertainty in streamflow synthesis. *Symposia on Statistics in Honours of Professor V. W. Josni's 70th Birthdat*.

VAN SLYKE, R. & WETS, R. J. B., 1969. L-shaped linear programs with applications to optimal control and stochastic linear programs. *Journal of Applied Mathematics*, Volume 17, pp. 638-663.

VECCHIA, A. V., 1985a. Maximum likelihood estimation for periodic autorregressive moving average models. *Technometrics*, Volume 27, pp. 375-384.

VECHIA, A. V., 1985b. Periodic autoregressive moving average (PARMA) modeling with application to water resources. *Water Resources Bulletin*, Volume 21, pp. 721-730.

VOGEL, R. M. & WILSON, I., 1996. Probability Distribution of Annual Maximum, Mean, and Minimum Streamflows in the United States. *Journal of Hydrologic Engineering*, 1(2), pp. 69-76.

VOSE, D., 2008. *Risk Analysis: a quantitative guide*. Third Edition ed. Chichester, England: John Wiley & Sons.

YAKOWITZ, S., 1982. Dynamic programming applications in water resources. *Water Resources Research*, 18(4), pp. 673-986.

9. Apêndices

9.1 Apêndice A – Estrutura de Configurações do PMO

Neste Apêndice é apresentada a estrutura de formação das configurações do PMO, conforme discutido no Capítulo 5. Maiores detalhes acerca do assunto estão disponíveis em (FERREIRA, 2013).

A estrutura do histórico de ENA utilizada no modelo NEWAVE tem características bastante peculiares. Embora esteja disponível apenas uma série histórica de dados, no planejamento da operação de médio prazo são consideradas “configurações” desta série histórica, que variam a cada PMO em função da inclusão/exclusão e/ou ampliação/redução de usinas que compõem os subsistemas equivalentes. Desta forma, uma “configuração” é uma série histórica correspondente a um dado período de tempo, à qual são acrescentados ou decrementados valores da variável estocástica em questão. Todos os meses, em cada PMO, uma nova série de configurações é disponibilizada para os períodos de planejamento. A partir destas séries serão simuladas as árvores de cenários.

Para, mensalmente, compor as configurações de ENA, é necessário (i) transformar as séries de vazões naturais em ENA – conforme descrito na equação (3.12) –, agregá-las por subsistema e compor as séries históricas correspondentes às configurações do parque gerador.

A “montagem” das configurações, embora não envolva conceitos da modelagem estocástica, é importante para o entendimento da abordagem dada pelo SEB à utilização do modelo $PAR(p)$ na geração dos cenários da árvore, conforme apresentado a seguir.

A Câmara de Comercialização de Energia Elétrica disponibiliza mensalmente o PMO, elaborado pelo ONS. No mesmo é possível identificar quando usinas entrarão (ou sairão) em operação nos próximos estágios do planejamento, bem como as novas capacidades de geração e produtividade. A cada mês em que novas unidades geradoras hidrelétricas se tornam operacionais,

tem-se uma nova configuração hidrelétrica do sistema, isto é, a série histórica de ENA para o período de janeiro de 1931 (início das séries de ENA) a dezembro de dois anos antes do ano corrente, o que ocorre em função da consolidação dos dados das séries de afluições. É importante ressaltar que o impacto dessa defasagem diz respeito ao conjunto de dados que será utilizado para ajuste dos parâmetros do modelo $PAR(p)$ e não para definir as condições correntes da hidrologia necessárias para gerar os cenários. Para tal, é informado ao modelo um conjunto de dados até o período imediatamente anterior ao estudo que está sendo feito. Por exemplo, para o PMO de junho, estão disponíveis os dados históricos de afluição (ou de ENA) até maio.

Desta forma, o termo “configuração” é uma série histórica de ENA energia correspondente a um dado período de tempo fixo. É válido ressaltar que nem todos os estágios de planejamento apresentam variações das configurações. Muitas vezes, quando não há incremento/decremento nos valores das séries, as mesmas se mantêm constantes até que haja uma alteração.

Imaginando que a cada estágio será disponibilizada uma nova série e o horizonte de planejamento seja de cinco anos (60 meses), o conjunto dessas formará o PMO, cujo número de configurações poderá variar de entre 60 (no caso de janeiro) a 49 (no caso de dezembro). Desta forma, o número de configurações de um dado PMO é dado por (em que $NPER$ representa o número de períodos considerados no horizonte de planejamento):

$$NCONF = NPER - mes_PMO + 1 \quad (9.1)$$

Ainda, neste, existem dois termos definidos pelo SEB que são o *pré-estudo* e o *pós-estudo*. O primeiro corresponde, basicamente, a primeira configuração e serve como ponto inicial para a estimação dos modelos, já o pós-estudo é a última configuração do sistema e os modelos definidos nessa configuração serão replicados, caso deseje-se criar cenários com horizonte maior que 60 estágios..

No contexto do módulo NEWAVE, os períodos de pré-estudo e pós-estudo são períodos nos quais as configurações do parque gerador e os dados de demanda são considerados constantes. Estes períodos servem para eliminar os efeitos do estado inicial (pré-estudo) e para obter uma informação acerca da FCF a partir do período de interesse (FCF diferente de zero). Estes períodos são utilizados tanto

no planejamento da expansão quanto no planejamento da operação. A diferença é que para o planejamento da expansão, o que se quer é definir as configurações do parque gerador para todo o período de planejamento, que compreende 10 anos ou mais, e no planejamento da operação a configuração é supostamente conhecida.

De forma ilustrativa, considere um PMO de janeiro/2011, 60 estágios de simulação e que em todos estes há alterações nas séries temporais de ENA (ou mesmo que não haja, uma dada configuração pode ser replicada até que sejam observadas atualizações nos dados). A Figura 9.1, adaptada de (FERREIRA, 2013), apresenta o esquema de configurações deste PMO.



Figura 9.1 – Pré-estudo, Pós-estudo e as diferentes configurações

Costuma-se referenciar as configurações com o nome do mês e ano em que a mesma corresponde, isto é, caso esteja-se trabalhando com o PMO de janeiro de 2011, por exemplo, referencia-se a configuração 1 como configuração de janeiro de 2011, a configuração 2 como configuração de fevereiro de 2011 e assim sucessivamente até a última que, neste caso, seria a configuração de dezembro de 2016 (configuração 60).

É válido ressaltar que cada configuração fornecerá um modelo estocástico, com ordem, parâmetros e resíduos, que será empregado na simulação das realizações das variáveis estocásticas de ENA que compõem a árvore de cenários.

De acordo com (FERREIRA, 2013), para estudos de planejamento da operação, o período de interesse são os cinco primeiros anos. No contexto do NEWAVE, adicionalmente, em geral consideram-se cinco anos adicionais, perfazendo um total de 10 anos de cenários gerados, para levar em conta os efeitos de final de horizonte, isto é, o sistema deve continuar operando após os cinco anos iniciais. Para o planejamento da expansão, em geral o horizonte é maior, 10 anos de interesse, podendo chegar a 30. De qualquer forma, em ambos os estudos,

operação e planejamento, sempre se consideram cinco anos adicionais, destinados a construir uma função de custo futuro para a fronteira do horizonte de interesse.

Nessa situação, em que se deseja gerar séries sintéticas de ENAs para períodos superiores ao número de configurações, mantêm-se a configuração denominada pós-estudo para todo o horizonte de análise, em outras palavras, o SEB assume não ter conhecimento da dinâmica do parque hídrico após o período de 5 anos e o admite constante.

9.2

Apêndice B – Caracterização das Distribuições de Probabilidade

Neste apêndice serão apresentadas as principais distribuições contínuas passíveis de serem ajustadas aos resíduos do modelo $PAR(p)$ em função do suporte que comportam. Para cada uma das distribuições será realizada uma descrição teórica e logo em seguida serão apresentadas as características estatísticas, como momentos, parâmetros, função densidade de probabilidade e outros, quando cabível.

9.2.1

Distribuição Normal

A distribuição de Gauss ou Normal é uma distribuição de probabilidade para variáveis aleatórias contínuas, caracterizada pela média e desvio padrão. Os valores de uma série que segue a distribuição Normal se distribuem simetricamente em relação à média e, portanto, apresentam o coeficiente de assimetria igual a zero.

A caracterização matemática da distribuição Normal está resumida na Tabela 9.1.

Tabela 9.1 - Características da Distribuição Normal

Suporte	$-\infty \leq x \leq \infty$
Parâmetro de localização	μ
Parâmetro de escala	σ
Função Densidade de Probabilidade	$f_X(x) = \frac{1}{\sigma(2\sigma)^{1/2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$
Coefficiente de Assimetria	0
Coefficiente de Curtose	3

A seguir, na Figura 9.2, exemplos de distribuições Normais com diferentes parâmetros de localização e escala.

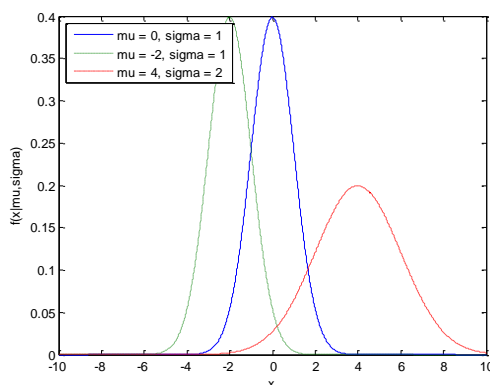


Figura 9.2 – Exemplos da Distribuição Normal

9.2.2 Distribuição Logística

A distribuição Logística se assemelha à Normal, porém suas caudas são mais densas. A literatura mostra exemplos de aplicação em hidrologia para a modelagem de dados de precipitação de chuvas.

Tabela 9.2 - Características da Distribuição Logística

Suporte	$-\infty \leq x \leq \infty$
Parâmetro de localização	a
Parâmetro de escala	$b^2 = k^2 \sigma^2 / 3$
Parâmetro de escala alternativo	$k = 3^{1/2} b / \pi$
Função Densidade de Probabilidade	$f_x(x) = \frac{\exp[(x-a)/k]}{k\{1 + \exp[(x-a)/k]\}^2}$
Coefficiente de Assimetria	0
Coefficiente de Curtose	4.2

A seguir, na Figura 9.3, um exemplo da função de distribuição de dados simulados de uma distribuição Logística.

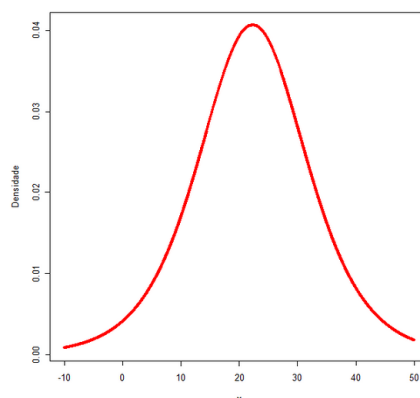


Figura 9.3 – Dados Simulados de uma Distribuição Logística

9.2.3 Distribuição de Gumbel

Também conhecida como Distribuição de Eventos Extremos, a distribuição de Gumbel surge quando se toma o logaritmo de uma variável com a distribuição Weibull. Em hidrologia, a distribuição de Gumbel é usada para analisar variáveis como valores máximos mensais e anuais de precipitação e volumes de descarga de rios, além de descrever secas. Em (STEPHENS, 1977), o autor apresenta um tratamento específico para o ajuste da Distribuição de Eventos Extremos.

Tabela 9.3 - Características da Distribuição de Gumbel

Suporte	$-\infty \leq x \leq \infty$
Parâmetro de localização	a
Parâmetro de escala	b
Média	$a + b\Gamma'(1) = a - 0.57721$
Variância	$b^2 \pi^2 / 6$
Função Densidade de Probabilidade	$f_x(x) = \frac{1}{b} \exp \left\{ \left[\frac{x-a}{b} \right] - \exp \left[\frac{x-a}{b} \right] \right\}$
Coefficiente de Assimetria	1.14
Coefficiente de Curtose	2.4

A seguir, na Figura 9.4, exemplo da função de distribuição de dados simulados de uma distribuição de Gumbel.

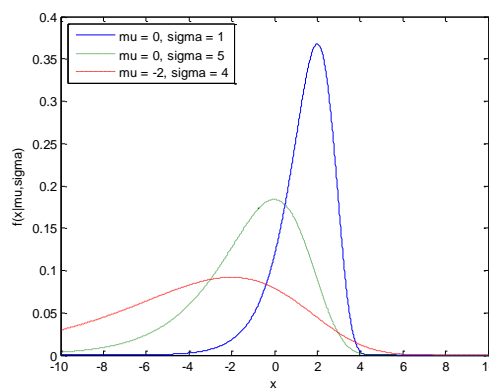


Figura 9.4 – Distribuição de Gumbel

9.2.4 Distribuição Generalizada de Valores Extremos

Na teoria da Probabilidade e Estatística, a distribuição Generalizada de Valores Extremos (GEV) é uma família de distribuições de probabilidade contínuas desenvolvidas dentro da Teoria do Valor Extremo para combinar as distribuições de Gumbel, Fréchet e Weibull, também conhecidas como tipos I, II e III de distribuições de valores extremos.

A GEV tem facilitado muitas aplicações e tem crescente aceitação para descrição dos eventos máximos naturais, principalmente dados de cheias máximas anuais. Esta distribuição é utilizada em muitas áreas, sendo uma das principais, a Engenharia Estrutural, em que o objetivo principal é projetar estruturas que resistam aos níveis mais extremos de certos processos ambientais. Na prática, a distribuição GEV é usada para modelar uma extensa variedade de extremos naturais, como cheias, chuvas, velocidade do vento, temperaturas e outros extremos, conforme mostram (VOGEL & WILSON, 1996). Pode também ser aplicada na modelagem de ruídos.

Tabela 9.4 - Características da Distribuição GEV

Suporte	$\begin{cases} x \in [\mu - \sigma / \xi, +\infty) & \text{se } \xi > 0 \\ x \in (-\infty, +\infty) & \text{se } \xi = 0 \\ x \in (-\infty, \mu - \sigma / \xi] & \text{se } \xi < 0 \end{cases}$
Parâmetro de localização	μ
Parâmetro de escala	σ
Parâmetro de forma	ξ
Média	$\begin{cases} \mu + \sigma \frac{\Gamma(1-\xi)}{\xi} & \text{se } \xi \neq 0, \xi < 0 \\ \mu + \sigma & \text{se } \xi = 0 \\ \infty & \text{se } \xi \geq 1 \end{cases}$
Variância	$\begin{cases} \sigma^2 \frac{(g_2 - g_1^2)}{\xi^2} & \text{se } \xi \neq 0, \xi < \frac{1}{2} \\ \sigma^2 \frac{\pi^2}{6} & \text{se } \xi = 0 \\ \infty & \text{se } \xi \geq \frac{1}{2} \end{cases}$ $g_k = \Gamma(1 - k\xi)$
Função Densidade de Probabilidade	$\frac{1}{\sigma} t(x)^{\xi+1} \exp[-t(x)]$ $t(x) = \begin{cases} \left[1 + \left(\frac{x - \mu}{\sigma} \right) \xi \right]^{-1/\xi} \\ \exp[-(x - \mu / \sigma)] \end{cases}$
Coeficiente de Assimetria	$\begin{cases} \frac{g_3 - 3g_1g_2 + 2g_1^3}{(g_2 - g_1^2)^{3/2}} & \text{se } \xi \neq 0 \\ \frac{12\sqrt{6}\zeta(3)}{\pi^3} & \text{se } \xi = 0 \end{cases}$ <p>Onde $\zeta(x)$ é a <i>Função Zeta de Riemann</i></p>
Coeficiente de Curtose	$\begin{cases} \frac{g_4 - 4g_1g_3 + 6g_2g_1^2 - 3g_1^4}{(g_2 - g_1^2)^2} - 3 & \text{se } \xi \neq 0 \\ \frac{12}{5} & \text{se } \xi = 0 \end{cases}$

A seguir, na Figura 9.5, exemplo da função de distribuição de dados simulados de uma distribuição de GEV.

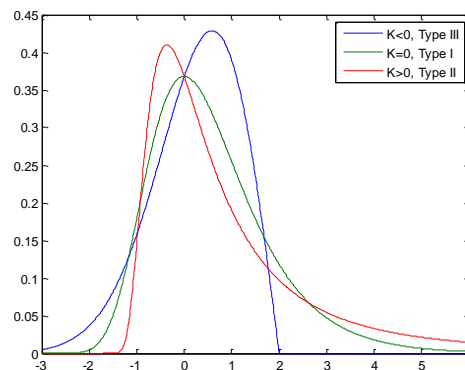


Figura 9.5 – Distribuição GEV

9.2.5 Distribuição do Tipo *Location-scale*

A Distribuição *t* de Student é bastante semelhante à Normal, porém com caudas mais largas, ou seja, uma simulação desta distribuição pode gerar valores mais extremos que uma simulação da Normal. O único parâmetro que a define e caracteriza a sua forma é o número de graus de liberdade (GL, representando pela letra grega ν). Quanto maior for esse parâmetro, mais próxima da Normal ela será.

De acordo com (JACKMAN, 2009), a Distribuição *t* de Student pode ser generalizada para uma distribuição de três parâmetros da família de distribuições conhecida como *Location-scale*. Para tal, basta introduzir um parâmetro de localização μ e um parâmetro de escala σ através da relação $X = \mu + \sigma T$. Desta forma pode-se definir a Distribuição do tipo *Location-scale* ou Distribuição *t* de Student não padronizada.

Se a variável *X* pertence a uma Distribuição do tipo *Location-scale* com parâmetros μ , σ e ν , então a relação $\frac{x - \mu}{\sigma}$ segue uma Distribuição *t* de Student com ν graus de liberdade.

A Distribuição do tipo *Location-scale* é útil para a modelagem de dados com caudas mais pesadas (mais propensas à presença de *outliers*) do que a distribuição Normal.

Tabela 9.5 - Características da Distribuição do Tipo *Location-scale*

Suporte	$-\infty \leq x \leq \infty$
Parâmetro de localização	μ
Parâmetro de escala	σ
Parâmetro de forma (GL)	ν
Variância	$\sigma^2 \frac{\nu}{\nu-1} \quad \nu > 2$
Função Densidade de Probabilidade	$f_x(x) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)(\pi\nu)^{1/2}\sigma} \left(1 + \frac{1}{\nu} \left(\frac{x-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2}$
Coeficiente de Assimetria	0
Coeficiente de Curtose	0

É importante notar que o parâmetro de escala σ não corresponde a um "desvio padrão", o qual, na realidade, pode nem existir, e nem é o desvio padrão da distribuição Normal subjacente, que é desconhecido. O referido parâmetro simplesmente ajusta a escala global da distribuição.

Na Figura 9.6, exemplo da função de distribuição de dados simulados de uma distribuição *t* de Student com 2 GL e uma Normal Padrão, para fins de comparação.

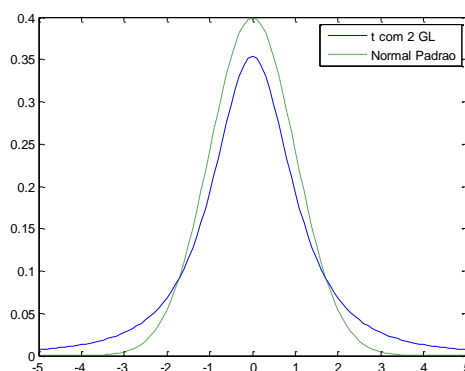


Figura 9.6 – Distribuição *t* de Student vs Normal

9.3

Apêndice C – Testes e Critérios de Ajuste, Estimação Paramétrica e Seleção de Distribuições

Haja vista que as principais distribuições de probabilidade contínuas aplicadas aos resíduos do modelo estocástico periódico, com suporte definido no conjunto dos números reais, foram apresentadas, faz-se necessário realizar testes e/ou critérios estatísticos que sejam capazes de comparar e/ou testar a hipótese de aderência dos dados às distribuições apresentadas. A seguir serão apresentados alguns testes utilizados no ajuste e estimação paramétrica de distribuições de probabilidade e os Critérios de Informação, que, conforme apresentado na seção 5.2, serão aplicados contexto deste trabalho.

9.3.1

Testes de Inferência Estatística Não Paramétrica

Os testes estatísticos clássicos de aderência de distribuição são não-paramétricos e correspondem a uma classe dos testes de hipóteses que têm como função verificar a forma de uma distribuição de probabilidade, isto é, verificar se os dados em questão se adaptam à curva de um modelo distributivo hipotético. Os testes clássicos, conhecidos como "*Tests of Goodness of Fit*" são Chi-Quadrado, e Kolmogorov-Smirnov e Anderson-Darling. Os dois primeiros são descritos a seguir e detalhes do último podem ser encontrados em (ANDERSON & DARLING, 1952).

9.3.1.1

Teste Chi-Quadrado

Considere uma amostra de N observações independentes de uma variável aleatória X . Essas N observações são agrupadas em c classes e o número de elementos em cada classe é representado na forma de uma Tabela de Contingência, (CONOVER, 1971), da seguinte forma, em que O_j denota o número de observações na classe $j = 1, 2, \dots, c$.

Tabela 9.6 - Tabela de Contingência

	Classes				Total
	1	2	...	c	
Frequências observadas	O_1	O_2	...	O_c	N

Considere agora $F(x)$ como função de distribuição desconhecida da variável aleatória X e $F^*(x)$ como alguma distribuição completamente especificada (incluindo a estimação de seus respectivos parâmetros). O teste de hipótese a ser realizado é:

$$\begin{aligned} H_0 : F(x) &= F^*(x) \\ H_1 : F(x) &\neq F^*(x) \end{aligned} \tag{9.2}$$

Em outras palavras:

$$\begin{aligned} H_0: & \text{a função de distribuição da variável aleatória observada é } F^*(x) \\ H_1: & \text{a função de distribuição da variável aleatória observada é diferente de } F^*(x) \end{aligned} \tag{9.3}$$

A estatística de teste empregada considera a probabilidade de uma dada observação pertencer a uma classe específica (p_j^*), sob a hipótese de que $F^*(x)$ é a função de distribuição da variável aleatória X , (PLACKETT, 1983). Logo, o número esperado de observações na classe j quando H_0 é verdadeira é dado por:

$$E_j = p_j^* N \quad j = 1, 2, \dots, c \tag{9.4}$$

A estatística de teste, T , é dada por:

$$T = \sum_{j=1}^c \frac{(O_j - E_j)^2}{E_j} = \sum_{j=1}^c \frac{(O_j)^2}{E_j} - N \tag{9.5}$$

A distribuição aproximada de T , válida para grandes amostras, é uma Chi-Quadrado com $(c-1)$ graus de liberdade. Assim, a região crítica de tamanho α corresponde aos valores de T maiores que $x_{1-\alpha}$. O quantil $(1-\alpha)$ de uma distribuição Chi-Quadrado com $(c-1)$ graus de liberdade é obtido a partir de tabelas padronizadas. A regra de decisão é: rejeitar H_0 se T excede $x_{1-\alpha}$. Em caso contrário, H_0 não é rejeitada.

9.3.1.2 Teste Kolmogorov-Smirnov

Este teste observa a máxima diferença absoluta entre a função de distribuição acumulada assumida para os dados e a função de distribuição empírica dos mesmos. Como critério, comparamos esta diferença com um valor crítico, para um dado nível de significância.

Considere uma amostra aleatória simples X_1, X_2, \dots, X_n , de uma população com função de distribuição acumulada contínua F_X desconhecida. A estatística utilizada para o teste é:

$$D_n = \sup_x |F(x) - F_n(x)| \quad (9.6)$$

Esta função corresponde a distância máxima vertical entre os gráficos de F_X e $F_n(x)$ sobre a amplitude dos possíveis valores de x . Em D_n temos que:

- F_X representa a função de distribuição acumulada assumida para os dados.
- $F_n(x)$ representa a função de distribuição acumulada empírica dos dados.

Neste caso, queremos testar a hipótese $H_0 : F_n(x) = F(x)$ contra a hipótese alternativa $H_1 : F_n(x) \neq F(x)$. Para isto, tomamos $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ as observações aleatórias ordenadas de forma crescente da população com função de distribuição contínua $F_n(x)$.

A função de distribuição acumulada assumida para os dados é definida por $F_n(x(i)) = P(X \leq x(i))$ e a função de distribuição acumulada empírica é definida por uma função escada, dada pela função indicadora:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{(-\infty, x]\}}(x(i)) \quad (9.7)$$

Observe que a função da distribuição empírica $F_n(x)$ corresponde à proporção de valores menores ou iguais a x .

A distribuição assintótica da estatística de teste K-S é válida quando temos conhecimento completo sobre a distribuição de H_0 . Entretanto, na prática, H_0 especifica uma família de distribuições de probabilidade. Neste caso, a distribuição assintótica da estatística de Kolmogorov-Smirnov não é conhecida e foi determinada via simulação.

Como a função de distribuição empírica $F_n(x)$ é descontínua e a função de distribuição hipotética é contínua, consideram-se duas outras estatísticas para o cálculo da estatística de teste:

$$D^+ = \sup_{x(i)} |F(x(i)) - F_n(x(i))| \quad (9.8)$$

$$D^- = \sup_{x(i)} |F(x(i)) - F_n(x_{(i-1)})| \quad (9.9)$$

Essas estatísticas medem as distâncias (vertical) entre os gráficos das duas funções, teórica e empírica, nos pontos $(x_{(i-1)})$ e $(x_{(i)})$. Com isso, pode-se utilizar como estatística de teste:

$$D_n = \max(D^+, D^-) \quad (9.10)$$

Se D_n é maior que o valor crítico, rejeita-se a hipótese nula com $(1 - \alpha) * 100\%$ de confiança. Caso contrário, a hipótese nula é aceita.

Embora bastante consolidados, os testes não paramétricos de *Goodness of Fit* podem apresentar alguns inconvenientes em determinadas aplicações e devem ser utilizados com cautela.

Na próxima seção são apresentados os Critérios de Informação empregados na seleção de distribuições.

9.3.2 Critérios de Informação

O uso de Critérios de Informação com vistas à comparação de modelos é bastante difundido na literatura. Diversos estão disponíveis e os mais comuns são: BIC (*Bayesian Information Criterion*), AIC (*Akaike Information Criterion*), AICc (*Akaike Information Criterion Corrected*), HNIC (*Hanna-Quinn Information Criterion*), dentre outros. As principais referências primárias são: (AKAIKE, 1974), (AKAIKE, 1976), (HANNAN & QUINN, 1979) e (SCHWARZ, 1978).

Basicamente, se uma boa estimativa para a log verossimilhança esperada dos modelos em questão puder ser obtida através dos dados observados, esta estimativa poderá ser utilizada como um critério para comparar modelos. Assim um modo de comparar n modelos, $g_1(x; \theta_1), g_2(x; \theta_2), \dots, g_n(x; \theta_n)$, é simplesmente comparar as magnitudes da função suporte maximizada $L(\hat{\theta}_i)$. Contudo, tal método não fornece uma verdadeira comparação, haja vista que, em não conhecendo o verdadeiro modelo $g(x)$, primeiramente o método da máxima verossimilhança estima os parâmetros θ_i de cada modelo $g_i(x)$ e posteriormente são utilizados os mesmos dados para estimar $E_G[\ln f(x; \hat{\theta})]$, isto introduz um viés em $L(\hat{\theta}_i)$, sendo que a magnitude deste viés varia de acordo com a dimensão do vetor de parâmetros.

Deste modo, os Critérios de Informação são construídos para avaliar e corrigir o viés ($b(G)$) da função suporte. Segundo (KONISHI & KITAGAWA, 2008), um critério de informação tem a forma que se segue:

$$IC(X_n; \hat{G}) = -2 \sum_{i=1}^n \ln f(X_i; \hat{\theta}(X_n)) + 2(b(G)) \quad (9.11)$$

Em (MULAIK, et al., 1989), os autores apresentam formas alternativas em substituição aos testes não paramétricos clássicos de testes de *goodness of fit*, haja vista que esses não conseguem avaliar a parcimônia dos modelos. No citado, o AIC é tratado com um índice de *lack-of-fit* ao invés de *goodness of fit*, uma vez que a regra de decisão da estatística seria “quanto menor, melhor”.

O mesmo ocorre em outros trabalhos recentes, como (VOSE, 2008) e (MAYDEU-OLIVARES & GARCIA-FORERO, 2010), que apresentam os Critérios de Informação como soluções eficientes em substituição às técnicas não paramétricas tradicionais para verificar a aderência de distribuições.

Dessarte, serão empregados os Critérios de Informação para seleção das distribuições de probabilidades mais aderentes aos resíduos do modelo PAR empregado no ajuste dos dados de ENA. Importante ressaltar que o ajuste, *rankeamento* e estimação paramétrica das distribuições de probabilidade dos resíduos do modelo estocástico são fundamentais para a simulação dos resíduos empregados no passo *Backward*, conforme sequência definida na seção 4.1.

Neste trabalho serão considerados para *rankear* as distribuições ajustadas os Critérios de Informação BIC, AIC e AICc. O último, AIC “corrigido”, revisto por (HURVICH & TSAI, 1989) e recomendado por (BURNHAM & ANDERSON, 2002), foi proposto por (SUGIURA, 1978), faz uma correção no AIC clássico em função de se adequar melhor a pequenas amostras (é mister lembrar que a amostra de resíduos do modelo PAR tem tamanho finito). Em (KUHA, 2004), o autor argumenta que a situação ideal para comparação dos modelos ou distribuições mais aderentes é de que todos os Critérios de Informação apontem para a mesma decisão.

Assim, tendo em vista o exposto, a seguir serão apresentados os Critérios AIC, BIC e AICc.

9.3.2.1 AIC

A literatura apresenta diversas versões do AIC, embora este critério tenha sido proposto originalmente por (AKAIKE, 1974) com a seguinte estrutura:

$$AIC = -2\ln(L) + 2k \quad (9.12)$$

Em que L é o valor obtido para a Máxima Verossimilhança do modelo e k é o número de parâmetros do mesmo.

O modelo ou distribuição mais aderente aos dados terá o menor valor de AIC. Em (KAPLAN, 2000), o AIC é classificado como um índice de *badness of fit*, de forma análoga ao apresentado em (MULAİK, et al., 1989).

A derivação matemática do AIC unifica os conceitos teóricos de Entropia e do Método de Máxima Verossimilhança, tornando o AIC bastante robusto. Por outro lado, o cálculo do índice é bastante simples, uma vez conhecido o valor da log verossimilhança dos modelos utilizados para comparação. Evidentemente, assim como no caso do BIC, dado que o cálculo não preconiza normalizações, não é possível interpretar um valor isolado do AIC, devendo ser utilizado como critério de desempate no *ranking* de distribuições ajustadas aos dados. Importante lembrar que os Critérios de Informação são medidas descritivas e não testes de hipóteses do ponto de vista do arcabouço da Inferência Estatística.

9.3.2.2 BIC

Proposto por (SCHWARZ, 1978), o BIC é muito similar ao seu antecessor, o AIC. O autor propõe um tratamento Bayesiano para o índice, uma vez que num modelo de dimensão conhecida, os estimadores de Máxima Verossimilhança podem ser obtidos como os limites dos estimadores de Bayes de grandes amostras para as distribuições *a priori*. Assim, o autor sugere uma modificação no cálculo do Critério de Informação, propondo o seguinte índice:

$$BIC = -2\ln(L) + k \ln(n) \quad (9.13)$$

Nota-se que o BIC que o segundo termo da equação (9.13), conhecido na literatura como “termo penalizador” é maior no caso do BIC do que do AIC, tornando o primeiro mais conservador.

Em seu trabalho, (KUHA, 2004) realiza um extenso estudo comparativo entre o desempenho dos índices AIC e BIC em diferentes simulações e conclui que ambos fornecem resultados satisfatórios. A sugestão do autor é que dois ou

mais Critérios sejam utilizados conjuntamente, dando robustez à escolha do modelo probabilístico mais aderente.

Ao longo dos anos, diversos outros Critérios de Informação surgiram e adaptações nos já existentes também. Neste contexto, a seguir será apresentado o AIC_C , uma evolução do AIC, que pode ser mais adequado em determinadas situações e vem sendo amplamente utilizado na literatura e nos softwares comerciais como critério de seleção de modelos.

9.3.2.3 **AICc**

A correção de viés no AIC foi proposta por (SUGIURA, 1978) e revisitada posteriormente nos trabalhos de (HURVICH & TSAI, 1989) e (HURVICH & TSAI, 1991), chamando o índice de AIC corrigido ou AICc.

Ambos os critérios podem ser vistos como estimadores da informação esperada de Kullback-Leibler. A literatura mostra que o AICc reduz drasticamente o viés do índice original, sugerindo que aquele fornece modelos substancialmente melhores quando comparado com a seleção via AIC.

Segundo (HURVICH & TSAI, 1989), o AICc, nos casos de modelos de Regressão Linear e Séries Temporais, é assintoticamente eficiente. No primeiro modelo, o índice é também não viesado, enquanto que em modelos de Regressão Não Linear e Séries Temporais é apenas aproximadamente não viesado. Em todos os casos, a redução do viés é alcançada sem incremento da variância, uma vez que o AICc é dado pela soma do AIC e um termo não estocástico.

Finalmente, de acordo com (HURVICH & TSAI, 1991), o AICc é recomendado em amostras reduzidas e se mostrou eficiente e com melhor performance em comparação ao AIC.

Isto posto, o AICc é dado por:

$$AICc = AIC + \frac{2(k+1)(k+2)}{N-k-2} \quad (9.14)$$

9.4 Apêndice D – Análise de Sequências

No intuito de apurar e complementar a avaliação de desempenho dos modelos de geração propostos, neste item são introduzidas novas variáveis aleatórias escolhidas a fim de aferir a capacidade dos modelos em reproduzir as correspondentes distribuições de frequências obtidas do histórico. Nesta seção, as variáveis selecionadas estão relacionadas com a representação de períodos críticos, como as secas registradas ao longo dos anos. Neste contexto é empregado o conceito de sequência negativa: uma sequência negativa é o período de tempo em que as vazões afluentes estão continuamente abaixo de valores predeterminados, por exemplo, as médias mensais (utilizadas neste trabalho), precedidos e sucedidos por valores acima destes limites. Vale lembrar que podem ser determinadas também sequências positivas, cujo conceito é o oposto do anterior. De forma ilustrativa, o exemplo a seguir considera apenas as sequências negativas.

A Figura 9.7 ilustra este conceito. A linha contínua representa afluências e a linha pontilhada um limite prestabelecido – neste trabalho utilizou-se a média histórica. Os intervalos $(t_1 - t_2)$ e $(t_3 - t_4)$ correspondem às sequências negativas, isto é, as vazões estão abaixo dos limites.

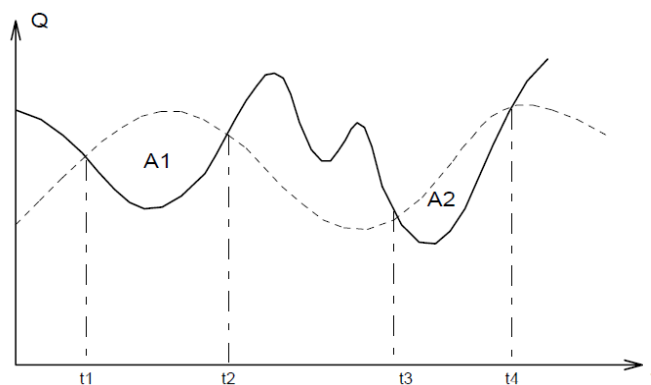


Figura 9.7 - Esquema de uma sequência negativa

Fonte: adaptado de (CEPEL, 2001)

Para cada sequência negativa estão associadas três variáveis de interesse: comprimento, soma e intensidade de sequência, da seguinte forma:

Tabela 9.7 – Modelagem das sequências negativas

Variável Aleatória	Descrição	Cálculo
Comprimento	Corresponde ao comprimento dos intervalos $(t_1 - t_2)$, $(t_3 - t_4)$	$C = (t_2 - t_1)$
Soma	Corresponde à área abaixo do limite durante a sequência. Na figura anterior, equivale às áreas A1 e A2.	$S = \sum_{i=t_1}^{t_2} (Z_i - \mu_i)$
Intensidade	Corresponde ao valor médio abaixo do limite, isto é, a soma de sequência dividida pelo respectivo comprimento de sequência.	$I = \frac{S}{C} = \frac{\sum_{i=t_1}^{t_2} (Z_i - \mu_i)}{(t_2 - t_1)}$

Fonte: adaptado de (CEPEL, 2001)

Repetindo estes cálculos para cada sequência negativa encontrada, é possível obter, para a série histórica, amostras das três variáveis aleatórias mencionadas. De maneira análoga, o algoritmo é realizado para a série sintética (utilizando o mesmo limite predeterminado), obtendo-se outras três amostras das mesmas variáveis.

Visto que amostras de mesmas variáveis foram obtidas, para o histórico e para os cenários, é possível, portanto, testar a hipótese de as amostras serem provenientes de mesmas distribuições por meio dos testes estatísticos de aderência descritos Apêndice C.