XX SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS
ÁGUA • DESENVOLVIMENTO ECONÔMICO E SOCIOAMBIENTAL
17 – 22 DE NOVEMBRO DE 2013 – BENTO GONÇALVES / RS
ABRH
ASSOCIAÇÃO BRASILEIRA DE RECURSOS HÍDRICOS

MODELOS DE MARKOV OCULTO NA GERAÇÃO DE SÉRIES TEMPORAIS DE VAZÕES

Luis Carlos Hernández Hernández [1]* *& Dirceu Silveira Reis Junior* [2]

Resumo – Os modelos de Markov oculto (Hidden Markov Models-HMM) apresentam varias características interessantes para simular a persistência observada em series de vazões. Além disso, é possível incluir nesses modelos variáveis climáticas que influenciam a persistência, como índices climáticos. Neste trabalho foram ajustados vários HMM, para gerar series de vazões anuais, utilizando a série de observações das vazões afluentes no reservatório Orós, no estado do Ceará no nordeste do Brasil, para o período de 1911-2000. Primeiro foram ajustados modelos HMM homogêneos e depois foram ajustados HMM não homogêneos utilizando os índices climáticos DIPOLO do Atlântico e NINO3. Os resultados mostraram que HMM conseguem representar a persistência da série observada, mas apresentaram algumas deficiências ao ser comparados com os modelos tradicionais ARMA.

Palavras-Chave – Modelos de Markov Oculto, persistência, índices climáticos, vazões anuais.

HIDDEN MARKOV MODELS IN STREAMFLOWS TIME SERIES GENERATION

Abstract – Hidden Markov models (HMM) have several interesting features to simulate the observed persistence in streamflow series. In addition, you can include these models climatic variables that influence the persistence, as the climatic indices. In this work were adjusted several HMM to generate series of annual flows, using observations of the inflow streamflows of Orós reservoir in the state of Ceará in northeastern Brazil, for the period 1911-2000. First homogeneous HMM were fitted and then were fitted non-homogeneous HMM using climate indices DIPOLE Atlantic and NINO3. The results showed that the HMM can represent the persistence of the observed series, but had some shortcomings when compared with the traditional ARMA models.

Keywords – Hidden Markov Models, persistence, climate indices, annual streamflows.

1. INTRODUÇÃO

Em muitas ocasiões as séries temporais de observações hidrológicas apresentam variações entre períodos ou estados climáticos. Variações em alguns casos ao redor de uma estatística (média, mediana, etc.). Assim, dado um limiar (*threshold*) específico, é possível determinar em uma série histórica períodos ou estados climáticos, por exemplo, tomando como limiar à média se podem definir como períodos secos à sequência consecutiva de observações abaixo da média. Os períodos úmidos serão então definidos como os períodos onde as observações se encontrem acima da média (Sveinsson *et al.*, 2003).

Contabilizar o tamanho dos comprimentos desses períodos é de interesse em varias áreas. Na hidrologia, por exemplo, é importante analisar essa persistência, a qual se denominada: persistência hidrológica. Sendo útil a sua analise para ajustar modelos estocásticos e gerar dados que apresentem a variação observada entre os estados climáticos (períodos secos e úmidos) ou a persistência desses estados. Segundo Matalas, (1977) a persistência hidrológica em séries de

[1] Mestrando do Programa de Pós-Graduação em Tecnologia Ambiental e Recursos Hídricos, Universidade de Brasília-UnB. Bolsista CNPQ, Campus Universitário Darcy Ribeiro, Brasília/DF, E-mail: flecks85@gmail.com.
[2] Departamento de engenharia civil e ambiental, Universidade de Brasília-UnB. Campus Universitário Darcy Ribeiro, Brasília/DF, E-mail: dirceu.reis@gmail.com.

precipitação e vazões é o resultado da influencia de variáveis atmosféricas ou armazenamento sub-superficial na bacia. Persistência que repercute principalmente no desenho e gerenciamento de infraestrutura de abastecimento de agua. Nesse aspecto Douglas et al., (2002) sinalam que uma longa vida útil de um sistema de abastecimento de agua é possível de lograr quando a persistência das vazões é tomada em conta na etapa de desenho. Com tudo isso, a análise da persistência hidrológica ajuda a desenvolver modelos estatísticos que representem as observações históricas para se aproximar melhor à realidade, e sirvam de suporte à gestão dos recursos hídricos (Whiting, 2006).

Assim, vários modelos estocásticos têm sido propostos na literatura para simular a persistência hidrológica, entre eles estão os modelos tradicionais *Autoregressive* (AR), *Moving Average* (MA), *Autoregressive Moving Average* (ARMA), *Fractional* ARMA (FARMA), *Fractional Gaussian Noise*, e *Broken Line*. Que resultam ser úteis para representar variabilidade de longo prazo e podem produzir mudanças aparentes de estados em séries climáticas e hidro-climáticas (Salas, 1993). Também em adição a esses modelos estão os modelos de Média Mudável (*Shifting Mean Models*), propostos por Boes e Salas (1978); e Salas e Boes, (1980); os Modelos de Markov Oculto (*Hidden Markov Models*-HMM) utilizados por Zuchinni e Guttorp,(1991); Thyer and Kuczera, (2000); Robertson et al., (2004); Lima, (2010), que também são capazes de gerar padrões de variabilidade, (Sveinsson *et al.*, 2003).

Neste trabalho foi analisado o uso dos modelos HMM na geração de séries de vazões anuais. Ajustando vários modelos para comparar o desempenho frente aos modelos tradicionalmente utilizados ARMA. Esse analise foi realizado da representação desses modelos da persistência na série observada. Persistência aqui é definida como a análise dos comprimentos de períodos úmidos e secos. Definidos por os percentis 33% e 66% utilizados no trabalho de Lima, (2010); e a mediana, utilizada por, Prairie et al., (2008) e Whiting, (2006); esse último autor menciona que a escolha da mediana reduz a influencia de dados assimétricos sobre o comprimento e magnitude dos períodos hidrológicos. Além disso, analisa-se também a influencia de fatores climáticos nas vazões analisadas.

2. MODELOS DE MARKOV OCULTO

Os Modelos de Markov Oculto (Hidden Markov Models - HMM) são modelos estatísticos nos quais a distribuição de probabilidade que gera uma observação, depende de um estado pertencente a um processo de Markov não observado ou oculto. Os HMM apresentam vários atrativos como o fácil tratamento matemático, especialmente a computação na hora de calcular os parâmetros do modelo. O que produz uma ampla flexibilidade desses modelos para múltiplas aplicações em series temporais univariadas e multivariadas, (Zucchini & MacDonald, 2009).

O interesse especial que tem surgido na hidrologia por esses modelos se deve a sua capacidade de representar a persistência de regimes climáticos ou persistência hidrológica, possibilitando, representar a persistência de períodos secos ou úmidos. Tendo aplicação na simulação e previsão da precipitação (Thyer e Kuczera, 2000); (Mallya, Tripathi, & Govindaraju, 2011); (Whiting, 2006); (Lambert, Whiting, Metcalfe, Whiting, & Metcalfe, 2003). Na geração de series vazões (Thyer e Kuczera, 2000); (Akintug e Rasmussen,

2005);(Whiting, 2006); e também na previsão de vazões (Bracken, 2011);(Fortin et al., 2004). Esses trabalhos têm utilizado series de vazões observadas anuais, e sendo utilizados inclusive em casos onde as variáveis modeladas têm influencia de fatores climáticos de longa escala e forte autocorrelação de ano para ano (Thyer e Kuczera, 2000; Whiting, 2006).

Zucchini and MacDonald, (2009), descreve um HMM para o caso discreto, mencionando que similarmente pode ser formulado um HMM para o caso continuo, ele descreve que um HMM de T observações $R_{1:T} = \{R_1, \dots, R_t, \dots, R_T\}$, é uma forma particular de uma mistura dependente; com $R_t: \{t = 1, 2, \dots\}$ e $S_t: \{t = 1, 2, \dots\}$ representando os históricos (ou series), do tempo 1 ate o tempo t, assim o modelo pode ser então descrito como:

$$P_r(S_t|\boldsymbol{S}_{1:t-1}) = P_r(S_t|S_{t-1}), \qquad t = 1,2,\dots,T \qquad (1)$$

$$P_r(\boldsymbol{R}_t|\boldsymbol{R}_{1:t-1}, \boldsymbol{S}_{1:t}) = P_r(R_t|S_t), \qquad t \in \mathbb{N} \qquad (2)$$

E, portanto o modelo consiste em dois processos. (1) Um processo (parameter process) não observado ou oculto (Hidden) $S_t: \{t = 1, 2, \dots\}$ (equação 1). Processo que consiste em uma Cadeia de Markov, satisfazendo a propriedade de Markov (equação 2). (2)Um processo estocástico dependente dos estados St, (state-dependent process) $R_t: \{t = 1, 2, \dots\}$, tal que quando o estado St é conhecido, a distribuição de Rt depende só do estado atual St e não a estados prévios (equação 1), a Figura 1 representa graficamente um HMM.

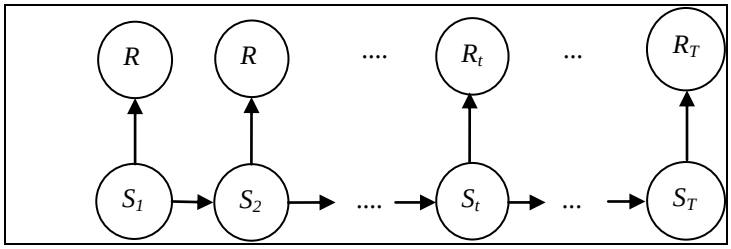


Figura 1 - Representação gráfica de um Modelo de Markov Oculto, HMM, (Zucchini e MacDonald, 2009)

Se a Cadeia de Markov $\{St\}$ tem m estados. Então $\{Rt\}$ será um HMM de m estados ocultos. Tendo uma distribuição estacionaria $\boldsymbol{\pi}$ com elementos $\pi_i = P_r(S_1 = i)$ e matriz de probabilidades de transição $\boldsymbol{\Gamma}$, de elementos $\gamma_{ij} = P_r(S_t = i|S_{t-1} = j)$ regulando a transição entre os estados $\{St\}$ da Cadeia de Markov.

Define-se também para o conjunto de observações discretas, $i = 1, 2, \dots, m$:

$$p_i(r) = P_r(R_t = r|S_t = i) \qquad (3)$$

Onde $p_i(r)$ é a função massa de probabilidade de R_t se a cadeia de Markov está no estado i no tempo t. Analogamente se define para o caso continuo $p_i(r)$ como a função de densidade de probabilidade (PDF) de R_t, se a cadeia de Markov está no estado i no tempo t. Assim as m distribuições $p_i(r)$ serão então as distribuições dependentes dos estados (state-dependent distributions).

No caso em que algumas variáveis (X) influenciam as observações R_t. Por exemplo, a chuva pode ser influenciada por fatores atmosféricos como a temperatura, velocidade do vento; entre outras variáveis. Ou influenciada por processos climáticos de larga escala resultantes de anomalias nas temperaturas no oceano. Dessa forma, essas X variáveis podem ser entradas (inputs) num HMM homogêneo e fazer que esse processo não seja mais homogêneo, influenciando as probabilidades de transição de estados, e por tanto influenciando as distribuições de probabilidade que geram as observações, (Kirshner, 2005). Os modelos que tomam por conta o anterior são chamados Modelos de Markov Oculto Não Homogêneos (Non-homogeneous Hidden Markov Models - NHMM). Inicialmente esses modelos foram descritos por Hughes e Guttorp, (1994) e Hughes et al., (1999). São descritos por Kirshner, (2005), como segue:

Dado um conjunto D-dimensional de variáveis de entrada para o modelo $X_{1:T} = \{X_1, \ldots, X_t, \ldots, X_T\}$, que convertem o HMM em não homogêneo ao fazer que a probabilidade do estado oculto S_t dependa do estado S_{t-1} e também dependa do valor de X_t, redefine-se a equação 1 e, portanto, se tem que:

$$P(S_t|\boldsymbol{S}_{1:t-1}, \boldsymbol{X}_{1:T}) = P(S_t|\boldsymbol{S}_{1:t-1}, \boldsymbol{X}_{1:t}), \quad t = 1,2,\ldots,T \qquad (4)$$

A Figura 2., representa graficamente a estrutura desse modelo.

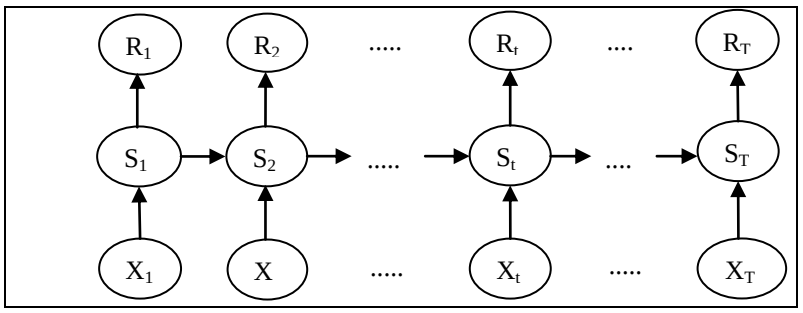


Figura 2. - Representação gráfica de um Modelo de Markov Oculto Não Homogêneo, NHMM (Kirshner, 2005) .

Além disso, as probabilidades de transicao $\gamma_{ij}(r)$ e as distribuicoes iniciais $\pi_i(r)$ dependem da variavel X_t, assim:

$$\pi_i(r) = P(S_1 = i|\boldsymbol{X}_1 = \boldsymbol{x}) \qquad (5)$$
$$\gamma_{ij}(r) = P(S_t = i|S_{t-1} = j, \boldsymbol{X}_t = \boldsymbol{x}) \qquad (6)$$

E para modelar a transição de estados se utiliza uma regressão multinomial logística, assim:

$$P(S_t = i|S_{t-1} = j, \boldsymbol{X}_t = \boldsymbol{x}) = \frac{\exp(\sigma_{ji} + \rho_i x^t)}{\sum_{i=1}^{m} \exp(\sigma_{ji} + \rho_i x^t)} \quad para\ t = 2, \ldots, T \tag{7}$$

$$P(S_1 = i|\boldsymbol{X}_1 = \boldsymbol{x}) = \frac{\exp(\lambda_i + \rho_i x^t)}{\sum_{i=1}^{m} \exp(\lambda_i + \rho_i x^t)} \quad para\ t = 1 \tag{8}$$

Onde m representa o numero de total de estados e $\lambda_i, \sigma_{ij} \in \mathbb{R}\ \rho_i \in \mathbb{R}^D$, logo, se define a $\Theta = (\Omega, \Gamma)$ como o conjunto total de parâmetros do modelo NHMM, onde Ω, representa a matriz dos parâmetros λ_i, σ_{ij} e ρ_i de tamanho $1xm$ e Γ a matriz de probabilidades de transição.

3. DADOS

As vazões médias anuais utilizadas neste trabalho para a geração de séries sintéticas de vazões anuais. Foram obtidas das vazões médias mensais afluentes no reservatório Orós para o período 1911-2000. Reservatório que toma as aguas do rio Jaguaribe no estado do Ceará (CE), Nordeste do Brasil (NEB).Os dados dos valores mensais das séries dos índices climáticos NINO3 e DIPOLO do Atlântico para o período 1910-2000. Foram tomados do site do *International Research Institute for Climate and Society* (IRI). Dados que correspondem a uma versão estendida dos trabalhos de Reynolds and Smith, (1994) e Kaplan *et al.*, (1998). O site do IRI entrega valores de anomalias da temperatura da superfície do mar (*Surface Sea Temperature*-SST).

As anomalias da SST para construir a série mensal do DIPOLO do Atlântico para o período de 1910-2000 foram tomadas do site do IRI disponíveis no seguinte link http://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED/.v2/. Para as coordenadas 5° N - 20° N, 60° W - 30° W (Souza Filho e Lall, 2003). Para a série de anomalias da SST do Oceano Atlântico Sul Tropical (TAS) as coordenadas 0° - 20° S, 30° W - 10°. Logo, a série do DIPOLO do Atlântico foi obtida como a diferença aritmética entre a série das anomalias TAN e a série das anomalias TAS, (Souza Filho e Lall, 2003; Brabo Alves et al., 2009). Os valores médios mensais do indicador climático NINO3 para o período 1910-2000. Foram obtidas do site do IRI disponíveis do seguinte link http://iridl.ldeo.columbia.edu/SOURCES/.Indices/.nino/.EXTENDED/.NINO3/. Esses valores correspondem a anomalias da SST da região com coordenadas geográficas 5° S – 5° N, 150° W – 90° W, (Souza Filho e Lall, 2003).

4. RESULTADOS

Foram analisadas realizadas 1000 simulações dos HMM de séries de 180 anos tomando os últimos 90 anos de cada series. As métricas analisadas foram os comprimentos definidos na figura 3 e a estatística T definida por Lima, (2010) como:

$$T = \sum_{l=2}^{6} n_c(l) \tag{9}$$

Onde $\mathrm{n_c}(\mathrm{l})$ indica o número total de clusters com l anos consecutivos no estado seco. Por exemplo, se existe uma série com quatro anos consecutivos de eventos seco, o valor de T será:

$$T = n_c(2) + n_c(3) + n_c(4) + n_c(5) + n_c(6) = 3 + 2 + 1 + 0 + 0 = 6$$

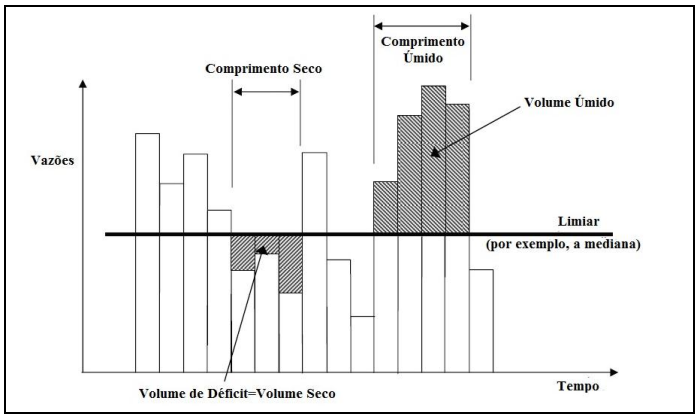
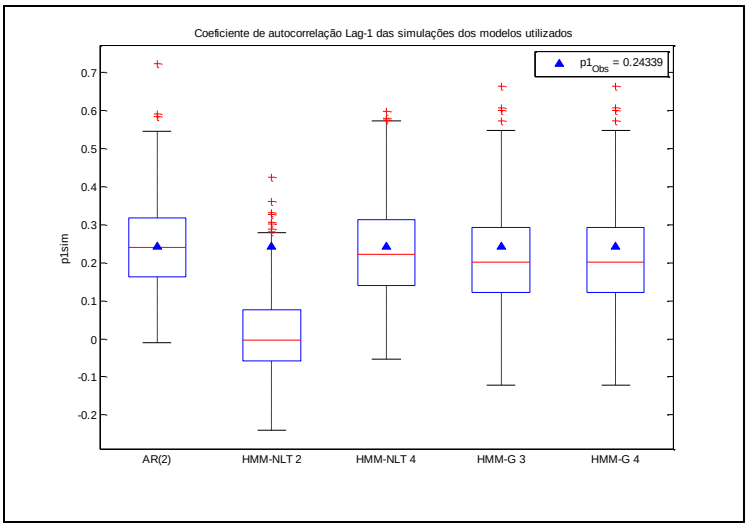
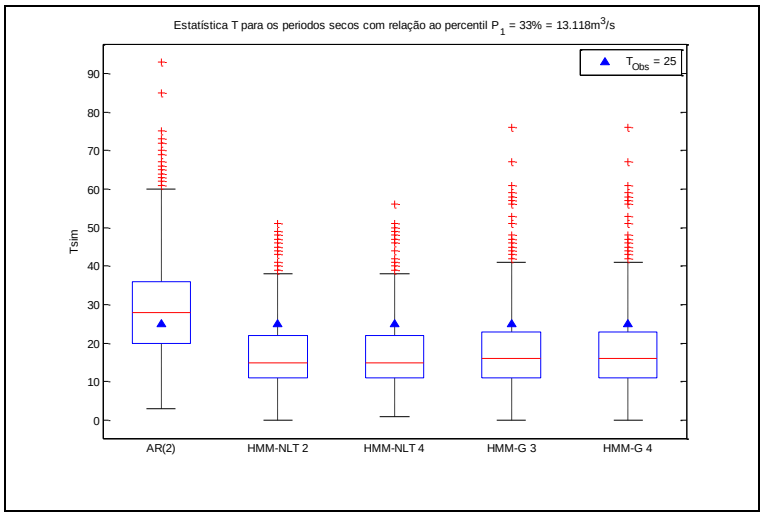
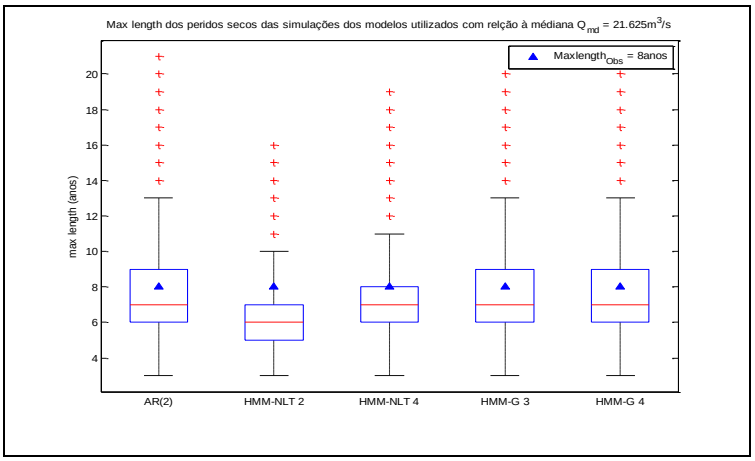


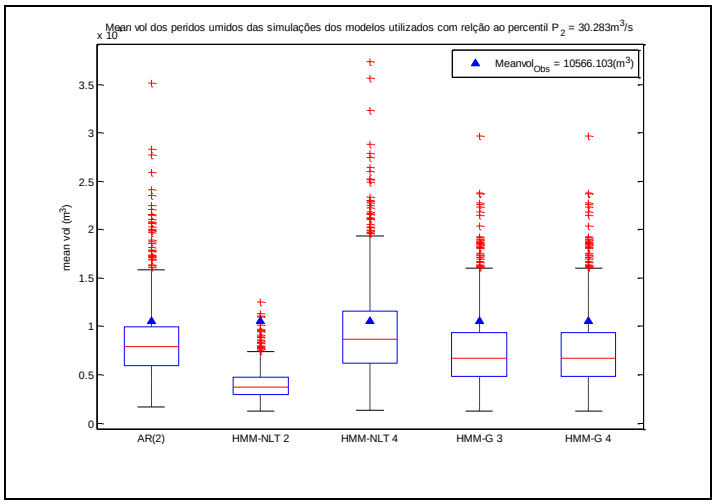
Figura 3. - Definição de Estatísticas de períodos secos e úmidos. (Prairie *et al.*, 2008,traduzido)

Além disso, foi comparada a autocorrelação e as estatísticas como média, desvio padrão etc. Assim, a seguir apresentam-se os resultados dessas simulações em forma de boxplots, onde os triângulos azuis representam os valores observados.









5. CONCLUSÕES

REFERÊNCIAS

Akintug, B., & Rasmussen, P. F. (2005). A Markov switching model for annual hydrologic time series. *Water Resources Research*, *41*(9), 1–10.

Brabo Alves, J., Servain, J., & Campos, J. (2009). Relationship between ocean climatic variability and rain-fed agriculture in northeast Brazil. *Climate Research*, *38*(May), 225–236.

Bracken, C. W. (2011). *Seasonal to Inter-Annual Streamflow Simulation and Forecasting on the Upper Colorado River Basin and Implications for Water Resources Management*. University of Colorado.

Douglas, E. M., Vogel, R. M., & Kroll, C. N. (2002). Impact of Streamflow Persistence on Hydrologic Design. *JOURNAL OF HYDROLOGIC ENGINEERING*, *7*(June), 220–227.

Fortin, V., Perreault, L., & Salas, J. D. (2004). Retrospective analysis and forecasting of streamflows using a shifting level model. *Journal of Hydrology*, *296*(1-4), 135–163.

Kaplan, A., Cane, M. A., Kushni, Y., Clement, A. C., Blumenthal, M. B., Rajagopalan, B., & Bottoraley, C. (1998). Analyses of global sea surface temperature 1856-1991. *Journal of Geophysical Research*, *103*(C9), 18,567–18,589.

Kirshner, S. (2005). *Modeling of Multivariate Time Series Using Hidden Markov Models*. UNIVERSITY OF CALIFORNIA, IRVINE.

Lambert, M. F., Whiting, P., Metcalfe, V., Whiting, J. P., & Metcalfe, A. V. (2003). A non-parametric hidden Markov model for climate state identification A non-parametric hidden Markov model for climate state identification, *7*(5), 652–667.

Lima, C. H. R. (2010). ANÁLISE E MODELAGEM DA SÉRIE HISTÓRICA DE FORTALEZA POR MEIO DE UM MODELO DE MARKOV ESCONDIDO NÃO-HOMOGÊNEO. *X Simpósio de Recursos Hídricos do Nordeste*, 1–15.

Mallya, G., Tripathi, S., & Govindaraju, R. S. (2011). HIDDEN MARKOV MODEL BASED PROBABILISTIC ASSESMENT OF DROUGHTS. *World Environmental and Water Resources Congress 2011: Bearing Knowledge for Sustainability ASCE 2011*, 1282–1291.

Prairie, J., Nowak, K., Rajagopalan, B., Lall, U., & Fulp, T. (2008). A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data. *Water Resources Research*, *44*(6), 1–11.

Reynolds, R. W., & Smith, T. M. (1994). Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation. *Jornal of Climate*, *7*, 929–948.

Robertson, A. W., Kirshner, S., & Smyth, P. (2004). Downscaling of Daily Rainfall Occurrence over Northeast Brazil Using a Hidden Markov Model. *Journal Of Climate*, *VOLUME 17*, 4407–4424.

Souza Filho, F. A., & Lall, U. (2003). Seasonal to interannual ensemble streamflow forecasts for Ceara , Brazil : Applications of a multivariate , semiparametric algorithm. *Water Resources Research*, *39*(11), 1–13.

Sveinsson, O. G. B., Salas, J. D., Boes, D. C., & Pielke Sr., R. A. (2003). Modeling the Dynamics of Long-Term Variability of Hydroclimatic Processes. *JOURNAL OF HYDROMETEOROLOGY*, *4*, 489–505.

Thyer, M., & Kuczera, G. (2000). Modeling long-term persistence in hydroclimatic time series using a hidden state Markov model. *Water Resources Research*, *36*(11), 3301–3310.

Whiting, J. P. (2006). *IDENTIFICATION AND MODELLING OF HYDROLOGICAL PERSISTENCE WITH HIDDEN MARKOV MODELS*. University of Adelaide, Australia.

Zucchini, W., & MacDonald, I. L. (2009). *Hidden Markov Models for Time Series: An Introduction using R*. *South African Actuarial Journal* (Chapman &., Vol. 10, p. 265). Boca Raton, FL, USA.