

SEAF – UM PROTÓTIPO DE UM SISTEMA ESPECIALISTA PARA ANÁLISE DE FREQUÊNCIA LOCAL DE EVENTOS HIDROLÓGICOS MÁXIMOS ANUAIS

Márcio de Oliveira Cândido¹ & Mauro Naghettini²

RESUMO --- As amostras usualmente curtas de observações hidrológicas e as grandes incertezas envolvidas na estimação de parâmetros e quantis tornam subjetiva a tarefa de selecionar uma função de distribuição de probabilidades para a análise de frequência local de máximos anuais. Os testes de aderência convencionais não são suficientemente potentes e discriminatórios para prover a sustentação objetiva ao processo de tomada de decisão, podendo levar um hidrólogo inexperiente a escolhas inapropriadas. Este artigo descreve a experiência dos autores no emprego da tecnologia de inteligência artificial e elementos de lógica difusa na construção de um sistema especialista computacional que emula os princípios de raciocínio de um especialista humano ao selecionar uma distribuição de probabilidades para proceder à análise de frequência de variáveis hidrológicas. O sistema foi aplicado a 20 amostras, relativamente longas, de alturas diárias de chuva e vazões diárias máximas anuais, observadas em estações localizadas na região sudeste do Brasil. Com o objetivo de verificar o desempenho do sistema, as mesmas amostras foram submetidas a um painel de especialistas em análise de frequência. Os resultados mostram que o sistema desempenha como um especialista e pode ser utilizado como auxílio a uma pessoa não-especialista na escolha de uma distribuição de probabilidades apropriada.

ABSTRACT --- The usually short samples of hydrological observations recorded at a site and the large uncertainties involved in parameter and quantile estimation make the task of selecting a probability distribution function for frequency analysis of annual maxima a subjective matter. Conventional goodness-of-fit tests are not powerful enough to provide the necessary objective backing to such a decision-making process and may lead a novice hydrologist to improper choices. In this paper, we describe the authors' experience in employing the technology of artificial intelligence and fuzzy-logic theory to build a computer expert system that emulates the reasoning principles of a human expert in selecting a probability distribution for hydrologic frequency analysis. Such a system has been applied to 20 relatively large samples of daily rainfall and streamflow annual maxima, recorded at gauging stations located in the Brazilian southeast. In order to check the system performance, the same samples have been submitted to a panel of experts in frequency analysis. The comparison of the results provides evidence that the computer system performs at an expert level and may be utilized to help an inexperienced to select an appropriate distribution for at-site hydrologic frequency.

Palavras-chave: análises de frequência, eventos extremos, sistema especialista.

¹ Engenheiro Hidrólogo do Serviço Geológico do Brasil - CPRM, Avenida Brasil, 1731, Funcionários, 30140-002, Belo Horizonte, MG, Brasil.
Email: marcio@bh.cprm.gov.br

² Professor da Universidade Federal de Minas Gerais, Avenida do Contorno 842, 8 andar, Centro 30110-060, Belo Horizonte, MG, Brasil.
Email: naghet@netuno.lcc.ufmg.br

1 – INTRODUÇÃO

A análise e a determinação das vazões de enchentes representam problemas correntes da Engenharia de Recursos Hídricos. Ao longo da história, não é difícil constatar a atração natural que as planícies e os vales fluviais exercem sobre as civilizações, uma vez que as áreas ribeirinhas a elas proporcionam condições favoráveis para a agricultura e vias de transporte, além de fácil acesso aos recursos hídricos, elementos indispensáveis ao desenvolvimento agrícola, industrial e urbano. Entretanto, os benefícios econômicos e sociais advindos da ocupação e uso das planícies marginais aos cursos d'água podem, com certa frequência, ser ofuscados pelos efeitos negativos dos desastres provocados por enchentes, como perda de vidas e prejuízos econômicos às propriedades ribeirinhas. Do ponto de vista geomorfológico, não há surpresa alguma no fato dos rios ocasionalmente se reapropriarem de suas próprias construções permanentes e dinâmicas, que são seus respectivos vales e planícies. Entretanto, é surpreendente constatar que as comunidades humanas por vezes ignoram o fato que habitar ou usar as planícies de inundação significa co-habitar com o risco de cheias.

A redução do risco de cheias e a mitigação de seus efeitos danosos podem ser proporcionadas por intervenções diversas no sistema fluvial natural, entre as quais cita-se a construção de reservatórios, o erguimento de diques de proteção, as estratégias de planejamento mais racional da ocupação das planícies de inundação, além de medidas de proteção das habitações e benfeitorias ribeirinhas. Da mesma forma, a análise e determinação de vazões de enchentes é parcela crucial do projeto e operação de estruturas hidráulicas destinadas ao aproveitamento de recursos hídricos, uma vez que a segurança destas estruturas depende fundamentalmente dessa informação. Em qualquer dessas situações, compete aos engenheiros e hidrólogos estimar características relevantes das enchentes como as precipitações a elas associadas, as vazões de pico, o volume e a duração dos hidrogramas, as áreas inundáveis, bem como seus valores críticos ou de projeto e/ou de operação. Existem metodologias diversas para a análise e determinação destas características das enchentes, algumas das quais são construídas em bases puramente determinísticas, enquanto outras procuram associar a magnitude das variáveis à frequência de sua superação. A preferência por uma entre estas metodologias depende de diversos fatores, tais como o objetivo do estudo, o porte das estruturas, a existência, a abundância e a qualidade dos registros plúvio-fluviométricos, entre outros.

Em diversas das circunstâncias mencionadas, é prática corrente da Engenharia de Recursos Hídricos que as variáveis hidrológicas sejam vistas como variáveis aleatórias, o que lhes confere suscetibilidade de serem analisadas sob o ponto de vista da teoria de probabilidades e da estatística matemática. Nesse sentido, a análise de frequência de variáveis hidrológicas, aqui brevemente referida como a quantificação do número esperado de ocorrências de um evento de certa magnitude, representa uma das principais aplicações da teoria de probabilidades e da estatística matemática no

campo da Engenharia de Recursos Hídricos. Os métodos de análise de frequência buscam extrair inferências quanto à probabilidade com que uma variável irá igualar ou superar um certo valor ou quantil, a partir de um conjunto amostral de ocorrências daquela variável. Se as ocorrências referem-se a observações tomadas unicamente em um ponto específico do espaço geográfico (e.g. : uma estação fluviométrica, em uma dada bacia hidrográfica), a análise de frequência é dita local. Contrariamente, se um número maior de observações da variável em questão, tomadas em diferentes pontos de uma certa região, forem empregadas conjuntamente para a inferência estatística, a análise de frequência é dita regional.

A análise de frequência local de variáveis hidrológicas dispõe de um conjunto de técnicas de inferência estatística e de modelos probabilísticos, os quais têm sido objetos freqüentes de investigação, visando principalmente à obtenção de estimativas cada vez mais confiáveis. Entretanto, a inexistência de amostras suficientemente longas impõe um limite superior ao grau de sofisticação estatística a ser empregado na análise de frequência local. Nesse sentido, a análise regional de frequência representa uma alternativa que procura compensar a insuficiente caracterização temporal do comportamento de eventos extremos por uma coerente caracterização espacial da variável hidrológica em questão. Embora seja fácil constatar a preferência pelos fortes argumentos apontados pelos métodos da análise regional [e.g.: Hosking e Wallis (1997), Potter (1987) e Bobée e Rasmussen (1995)], é um fato que inúmeras decisões envolvendo eventos hidrológicos extremos sejam tomadas com base somente na análise de frequência local dos registros das variáveis associadas. Esse fato certamente está vinculado à inexistência ou à pouca profusão de registros de variáveis hidrológicas em uma dada região geográfica, ou à falta de conhecimento dos métodos de análise regional, ou mesmo à forma expedita com que se formulam soluções de engenharia para os problemas de mitigação e controle de eventos hidrológicos extremos. Não obstante tal fato e por ser intuitivamente perceptível que a análise de frequência local de variáveis hidrológicas permanecerá ainda por muito tempo como um método de uso corrente da Engenharia, focaliza-se no presente artigo as técnicas e dificuldades inerentes a esse tipo de análise e, em particular, a seleção da função de distribuição de probabilidades a ser empregada.

A análise de frequência local de variáveis hidrológicas consiste no ajuste de uma função de distribuição de probabilidades a uma amostra de observações máximas anuais da variável em questão, as quais devem ser consideradas, por princípio, como aleatórias, independentes e homogêneas; as chamadas séries de duração parcial também podem ser usadas, embora sua utilização seja muito menos freqüente e não sejam aqui consideradas. Uma vez escolhida a função de distribuição de probabilidades dentre os diversos modelos probabilísticos disponíveis, os seus parâmetros e quantis de interesse para a variável em estudo podem ser estimados. No contexto da análise de frequência local de eventos máximos anuais, as funções de distribuição de probabilidades

de uso mais comum em hidrologia podem ser agrupadas em (a) modelos de 2 parâmetros, entre os quais podem ser citados os de Gumbel, Log-Normal 2P, Gama e Exponencial 2P e (b) modelos de mais de 2 parâmetros, como os prescritos pelas distribuições Generalizada de Valores Extremos (GEV), Pearson III, Log-Pearson III e Wakeby, entre outras. Apesar da disponibilidade de um amplo conjunto de modelos probabilísticos, não há, entre os hidrólogos e especialistas da área, qualquer consenso quanto à prescrição de uma única função de distribuição de probabilidades que seja considerada adequada à análise de frequência de variáveis hidrológicas. De fato, a inexistência de leis dedutivas de seleção de uma distribuição de probabilidades para a análise de frequência de variáveis hidrológicas faz com que esse seja um procedimento *ad hoc*, baseando-se principalmente na aderência do modelo prescrito a uma amostra de observações da variável em questão.

Na prática, as amostras, geralmente curtas, de observações máximas anuais das variáveis hidrológicas tornam difícil e subjetiva a escolha de uma distribuição de probabilidades com base, principalmente, em sua aderência a um conjunto de dados. Com amostras de tamanhos típicos entre 20 e 60, é impossível afirmar categoricamente que uma certa distribuição de probabilidades, considerada aderente aos dados, irá representar o verdadeiro comportamento populacional. Os métodos convencionais de inferência estatística produzem estimativas pouco confiáveis de parâmetros e quantis, devido principalmente à grande incerteza imposta pelas amostras de pequeno tamanho. Por sua vez, os testes estatísticos de aderência, como o do Qui-Quadrado e o de Kolmogorov-Smirnov, são pouco potentes e incapazes de discriminar entre modelos probabilísticos, resultando que mais de uma distribuição pode ser considerada aderente a uma dada amostra.

Estas dificuldades fazem com que a seleção judiciosa de uma certa função de distribuição de probabilidades seja uma tarefa de especialistas, que geralmente a desempenham à luz de um conjunto de regras heurísticas formuladas de acordo com o conhecimento acumulado ao longo de anos de experiência e estudo. A abordagem heurística limita os caminhos a seguir, selecionando aqueles considerados melhores e reduzindo uma tarefa complexa a um conjunto de operações de julgamento. Por exemplo, um determinado especialista pode sugerir algumas distribuições candidatas simplesmente com base no exame do coeficiente de assimetria amostral e, na sequência, sugerir a adoção daquela que produzir a melhor aderência em papel de probabilidades. Em geral, essas regras heurísticas facilitam o trabalho do especialista, porém, se usadas indiscriminadamente, podem conduzir a resultados tendenciosos. Como exemplo disso, suponha que um evento, sabidamente muito raro em uma dada amostra, produza um coeficiente de assimetria extremamente alto. Nesse caso, se a seleção das distribuições candidatas se der com base somente no coeficiente de assimetria amostral, correr-se-ia o risco de se prescrever modelos ‘adequadamente’ assimétricos, porém incorretos. Esse exemplo vem mostrar que, longe de ser uma tarefa balizada por um conjunto sistemático de regras, a seleção judiciosa de uma distribuição de probabilidades é um procedimento

multi-critério, de caráter heurístico e inexato, o qual é passível de ser realizado por analistas especializados, com experiência e conhecimentos específicos.

Apesar de numerosos programas computacionais terem sido desenvolvidos como ferramentas auxiliares para a análise de frequência local de eventos máximos anuais, eles não são capazes de orientar o usuário sobre a escolha da função de distribuição de probabilidades ou do conjunto de distribuições mais adequadas para a amostra em questão. Os resultados apresentados são dados numéricos, os quais podem levar um hidrólogo inexperiente a encontrar dificuldades em decidir qual distribuição será adotada em sua análise, haja vista que duas ou mais distribuições podem ser aceitas pelos testes de hipóteses adotados.

Dependendo da distribuição populacional escolhida, o quantil estimado para a variável característica de projeto ou de decisão pode inviabilizar economicamente o empreendimento ou sujeitá-lo a um maior risco de falha. Dentro deste cenário, a escolha de uma distribuição de probabilidade para um conjunto de dados é importante e não pode ser realizada por um simples e único algoritmo. Ela requer a agregação de análises objetivas e subjetivas, as quais podem conduzir a resultados diferentes dependendo do padrão de raciocínio adotado por cada especialista.

De um modo geral, sempre que a solução de um problema envolve a combinação de critérios subjetivos, a utilização dos chamados Sistemas Especialistas pode ser uma alternativa interessante para a padronização e informatização do processo de análise e tomada de decisão. Um sistema especialista é um programa de computador projetado e desenvolvido para atender a uma aplicação determinada do conhecimento humano. Ele é capaz de auxiliar a tomada de uma decisão, apoiada em conhecimento justificado a partir de uma base de informações, tal qual um especialista de uma área específica do conhecimento. No caso presente, essa base de informações pode ser construída a partir da reunião de conhecimentos de profissionais qualificados e experientes em análise de frequência local de variáveis hidrológicas e compor uma lógica para a tomada de decisão sob condições subjetivas. O presente artigo constitui um relato da experiência de se conceber, implementar e testar um protótipo de um sistema especialista de auxílio à tomada de decisão, tendo como base um conjunto definido de regras heurísticas que reflitam as tendências e os resultados da pesquisa recente em análise de frequência local de variáveis hidrológicas.

2 – O SISTEMA SEAF

O Sistema Especialista de Análise de Frequência (SEAF) tem como objetivo implementar alguns critérios de seleção da distribuição de probabilidades, utilizando estatísticas amostrais baseadas nos momentos-L, a fim de padronizar o processo de escolha de uma distribuição de probabilidades na análise de frequência local de eventos hidrológicos máximos anuais. Vale ressaltar que não é objetivo do sistema identificar a distribuição populacional dos dados, mas

selecionar, entre um grupo de distribuições candidatas, aquela (ou aquelas) distribuição (distribuições) de probabilidade mais apropriada (s) para modelar os dados amostrais sob análise.

O conjunto de distribuições candidatas, do qual se extraem as possíveis escolhas do sistema, é formado pelas distribuições Normal, Log-Normal de 2 parâmetros ou Log-Normal 2p, Gumbel, Generalizada de Valores Extremos (GEV), Exponencial, Generalizada de Pareto (GPA), Pearson III e Log-Pearson III. Com exceção da distribuição Normal, que aqui é empregada como paradigma para algumas decisões, o conjunto de distribuições candidatas agrupa aquelas mais utilizadas em análise de frequência local de eventos hidrológicos máximos anuais.

O programa foi desenvolvido para trabalhar em ambiente Windows. A arquitetura do sistema é dividida em duas partes: a primeira refere-se à interface gráfica com o usuário e aos cálculos matemáticos, tendo sido desenvolvida no aplicativo Delphi 4.0 Professional, da Borland. A segunda refere-se ao armazenamento da chamada 'base de conhecimento do sistema', bem como às interpretações e análise, tendo sido desenvolvida em linguagem FuzzyCLIPS, a qual é mantida e distribuída pelo "Integrated Reasoning Group do Institute for Information Technology do National Research Council" do Canadá. Os dois módulos constituintes do sistema são interligados entre si e foram projetados para serem acionados automaticamente pelo programa sem que seja necessária uma intervenção do usuário.

De modo resumido, o sistema primeiramente extrai a informação numérica dos dados amostrais, analisando-a, na seqüência, à luz do conjunto interno de regras heurísticas fundamentadas em conhecimento, transformando-a, finalmente, em declarações lingüísticas de decisão. Nesse sentido, o programa SEAF, em interface com a *shell* FuzzyCLIPS, emprega a tecnologia de inteligência artificial e elementos de lógica difusa na emulação dos princípios de raciocínio de um especialista humano, ao selecionar uma distribuição de probabilidades para proceder à análise de frequência local de variáveis hidrológicas. No julgamento da plausibilidade de uma dada distribuição, o sistema baseia-se em um conjunto de regras fundamentadas em conhecimento contemporâneo que constituem diretrizes razoavelmente similares àquelas empregadas por um especialista humano ao selecionar uma distribuição de probabilidades para uso na análise de frequência hidrológica.

Para descrever a variabilidade presente nos dados amostrais, bem como para inferir quanto à estimação de parâmetros e às características distributivas de forma, foram empregados o método dos momentos-L e os quocientes de momentos-L. O emprego dessas estatísticas encontrou justificativa em argumentos sobre sua relativa superioridade, argumentos estes que têm sido enfocados na literatura especializada [e.g.: Hosking e Wallis (1997), Muhara (2001), Davis e Naghettini (2001), Vogel e Fennessey (1993), Zvi e Azmon (1997)].

As regras heurísticas que formam a base de conhecimento do sistema SEAF podem ser resumidas nas seguintes etapas:

1. Verificação da condição de aleatoriedade, homogeneidade, independência serial e a presença de *outliers* nas series de dados anuais respectivamente por meio dos testes de Mann-Whitney, Wilcoxon, ρ de Spearman e Grubbs-Beck. Os testes não paramétricos executados pelo sistema são apenas de caráter informativo e não afetam o processo de escolha e classificação das distribuições. O sistema apenas informa ao usuário quanto à possível presença de correlação serial, heterogeneidade e *outliers*. A partir daí, cabe ao usuário decidir sobre a continuidade ou descontinuidade da análise em curso.
2. Cálculo dos momentos-L e/ou quocientes de momentos-L amostrais: l_1 , l_2 , t_2 (L-CV), t_3 (L-Assimetria) e t_4 (L-Curtose), de acordo com a formulação descrita por Hosking e Wallis (1997).
3. Cálculo dos parâmetros das seguintes distribuições de 2 parâmetros: Normal, Log-Normal, Exponencial e Gumbel, bem como as respectivas variâncias amostrais assintóticas $\text{Var}^{\text{Dist}}(\tau_3)$, calculadas por simulação de Monte Carlo.
4. Cálculo dos parâmetros das seguintes distribuições de 3 parâmetros: Pearson III, Log-Pearson III, GEV e Generalizada de Pareto, bem como as respectivas variâncias amostrais assintóticas $\text{Var}^{\text{Dist}}(\tau_4)$, calculadas por simulação de Monte Carlo.
5. Assumindo que τ_3 é $N[\tau_3, \text{Var}^{\text{Dist}}(\tau_3)]$, o sistema define o intervalo de confiança $[\tau_3^{0,025}; \tau_3^{0,975}]$ para cada uma das distribuições de 2 parâmetros.
6. Com base no intervalo de confiança $[\tau_3^{0,025}; \tau_3^{0,975}]$ e nas assimetrias-L amostrais t_3 , o sistema seleciona todas as distribuições de 2 parâmetros plausíveis e associa a cada uma delas um nível de confiança preliminar calculado por uma função difusa do tipo sino, conforme descrito por Candido (2003).
7. Assumindo que τ_4 é $N[\tau_4, \text{Var}^{\text{Dist}}(\tau_4)]$, o sistema define o intervalo de confiança $[\tau_4^{0,025}; \tau_4^{0,975}]$ para cada uma das distribuições de 2 parâmetros.
8. Com base no intervalo de confiança $[\tau_4^{0,025}; \tau_4^{0,975}]$ e nas curtoses-L amostrais t_4 , o sistema seleciona todas as distribuições de 3p plausíveis e associa a cada uma delas um nível de confiança preliminar calculado por uma função difusa do tipo sino, conforme descrito por Candido (2003).
9. Cálculo da estatística amostral r de Filliben e do intervalo de confiança $[\rho_{0,005}, 1]$, o qual é calculado por simulação de Monte Carlo, para cada uma das distribuições analisadas.

10. Com base no intervalo de confiança $[\rho_{0,005}, 1]$ e na estatística r , o sistema seleciona todas as distribuições plausíveis e associa a cada uma delas um segundo nível de confiança calculado por uma função difusa do tipo S, conforme descrito por Candido (2003).
11. O sistema inicia o processo de conversão das informações numéricas em declarações linguísticas, as quais são agrupadas de acordo com as distribuições de probabilidades correspondentes. Com base nestas declarações, o sistema atribui um nível de confiança a cada distribuição analisada. O nível de confiança é calculado como a média aritmética dos níveis de confiança de suas declarações correspondentes. O sistema pode apresentar até duas declarações linguísticas para cada distribuição analisada. No caso de existir apenas uma declaração linguística para uma determinada distribuição, o sistema atribui a ela um nível de confiança igual à média entre o valor do nível de confiança de sua declaração e o valor 0,5. Isto permite a penalização daquela distribuição específica por não ter sido bem sucedida em um dos testes de aderência. Caso não haja nenhuma declaração linguística para uma certa distribuição, ela é descartada da análise.
12. O próximo passo da análise é a verificação se há alguma razão para se rejeitar uma dada distribuição de probabilidades anteriormente classificada. Para isto, o sistema SEAF compara os limites de aplicação de cada distribuição ajustada com os limites mínimo e máximo amostrados e descarta a sua utilização nas seguintes situações: (i) quando a distribuição for limitada na direção do máximo e (ii) quando o mínimo amostral for menor que 90% do valor do limite mínimo da distribuição ajustada.
13. Antes de apresentar as distribuições selecionadas, o sistema verifica a parcimônia estatística entre as distribuições de mesma família, através da comparação de seus respectivos níveis de confiança. Caso uma distribuição de 2 parâmetros tenha um nível de confiança maior do que o de uma distribuição de 3 parâmetros da mesma família, o sistema descarta a distribuição de 3 parâmetros, uma vez que a incerteza na estimação de mais um parâmetro não contribuiu para melhorar o nível de confiança desta distribuição em relação à sua parente de 2 parâmetros. As distribuições foram agrupadas em quatro famílias, a saber: Família 1, composta pela Normal e Pearson III, Família 2, composta pela Log-Normal 2p e Log-Pearson III, Família 3, composta pela Gumbel e GEV, e Família 4, composta pela Exponencial e GPA.
14. Finalmente, o sistema apresenta as distribuições selecionadas com os seus respectivos níveis de confiança. O sistema SEAF recomenda o emprego da distribuição que apresentar o maior nível de confiança.

O arquivo executável do sistema SEAF e um breve manual de instruções para seu uso encontram-se disponíveis para *download* a partir da URL <http://www.ehr.ufmg.br>.

3 - ANÁLISE DE DESEMPENHO DO SISTEMA ESPECIALISTA

Objetivando avaliar o desempenho do sistema SEAF foram selecionadas 20 amostras de máximos anuais de alturas diárias de precipitação e vazões médias diárias para serem submetidas à análise pelo sistema. Estas amostras foram obtidas a partir dos registros diários observados em estações hidrológicas, pertencentes à Agência Nacional de Águas e localizadas na região sudeste do Brasil.

O critério utilizado para a seleção das estações foi o de ter em mãos amostras com mais de 35 anos de registros contínuos de observações diárias, cujos dados já estivessem consistidos e classificados como de boa qualidade. A justificativa para tal critério é a constatação de que amostras de tamanhos muito pequenos ou compostas por dados de má qualidade poderiam conduzir a maiores erros na estimação dos parâmetros amostrais e, conseqüentemente, induzir o sistema à seleção de uma distribuição de probabilidades imprópria. No caso das estações pluviométricas, observou-se também o critério de vazões naturais, ou seja, a inexistência de regularização significativa a montante. As Tabelas 1 e 2 apresentam as listas das estações pluviométricas e fluviométricas utilizadas na verificação do sistema, com suas respectivas coordenadas geográficas.

Tabela 1 – Estações pluviométricas utilizadas na verificação do sistema

Código	Nome	Coordenadas		Altitude M
		Latitude	Longitude	
01544012	São Francisco	15° 56' 58"S	44° 52' 05"W	448
01645000	São Romão	16° 22' 18"S	45° 04' 58"W	472
01943000	Mineração Moro Velho	19° 45' 58"S	43° 51' 00"W	770
01944004	Ponte Nova do Paraopeba	19° 57' 20"S	44° 18' 24"W	721
01944007	Fazenda Escola Forestal	19° 52' 47"S	44° 25' 18"W	745
02044012	Ibirité	20° 02' 34"S	44° 02' 36"W	1073
02045005	Lamounier	20° 28' 20"S	45° 02' 10"W	738
02244038	Ponte do Souza	22° 16' 15"S	44° 23' 26"W	918
01943009	Vespasiano	19° 41' 14"S	43° 55' 15"W	676
02243004	Conservatória	22° 17' 16"S	43° 55' 44"W	530

Tabela 2 – Estações fluviométricas utilizadas na verificação do sistema

Código	Nome	Rio	Coordenadas		Área m ²
			Latitude	Longitude	
40025000	Vargem Bonita	São Francisco	20° 19' 43"S	46° 21' 58"W	299
40050000	Iguatama	São Francisco	20° 10' 12"S	45° 42' 57"W	4846
40100000	Porto das Andorinhas	São Francisco	19° 16' 50"S	45° 16' 06"W	13087
40680000	Entre Rio de Minas	Brumado	20° 39' 37"S	44° 04' 19"W	469
41250000	Vespasiano	Da Mata	19° 41' 14"S	43° 55' 14"W	676
40800001	Ponte Nova do Paraopeba	Paraopeba	19° 56' 57"S	44° 18' 19"W	5663
56028000	Piranga	Piranga	20° 41' 17"S	43° 18' 02"W	1395
56075000	Porto Firme	Piranga	20° 40' 13"S	43° 05' 30"W	4251
56415000	Rio Casca	Casca	20° 13' 34"S	42° 39' 00"W	2036
56500000	Abre Campo	Santana	20° 17' 56"S	42° 28' 41"W	272

O sistema SEAF identificou a presença de *outliers* em 11 das 20 amostras analisadas, conforme a Tabela 3. Verifica-se que os testes utilizados pelo sistema para a identificação da presença de *outliers* mostraram-se eficientes, mas conservadores quando avaliados isoladamente.

Contudo, considerando-se os casos de resultados consensuais, em ambos os testes, o número de acertos foi maior em relação ao número de amostras identificadas com suspeitas de *outliers*, quando comparados com o exame visual.

Tabela 3 – Resultados dos testes para verificação da presença de *outliers*

Amostras	Avaliação da presença de <i>outliers</i>		
	Grubbs e Beck	Assimetria	Visual
01544012	*		
01944004		*	
01944007	*		
02244038	*	*	
01943000	*	*	*
40050000		*	
40100000	*	*	*
40680000	*	*	*
41250000		*	
56075000	*	*	*
56500000	*	*	*

As distribuições aprovadas nos testes estatísticos, utilizados pelo sistema SEAF, estão listadas na Tabela 4.

Tabela 4 – Resultados dos testes estatísticos para as amostras selecionadas

Amostra	Distribuições aprovadas							
	NOR	LNR	GUM	EXP	PE3	LP3	GEV	GPA
01544012		*	*	*	*	*	*	
01645000		*	*	*	*	*	*	*
01943000	*	*	*		*	*	*	*
01944004		*	*		*	*	*	
01944007				*	*	*	*	*
02044012		*	*	*	*	*	*	*
02045005		*	*	*	*	*	*	*
02244038			*	*	*	*	*	*
01943009		*		*	*	*	*	
02243004		*	*	*	*	*	*	*
40025000				*	*	*	*	*
40050000		*	*	*	*	*	*	
40100000		*		*			*	
40680000						*	*	*
41250000		*	*	*	*	*	*	*
40800001		*	*	*	*	*	*	*
56028000		*		*	*	*	*	*
56075000				*	*	*	*	*
56415000		*	*	*	*	*	*	*
56500000		*		*			*	
TOTAL	1	15	12	17	17	18	20	14

Nota: NOR – Distribuição Normal, LNR – Distribuição Lognormal, GUM – Distribuição de Gumbel, EXP – Distribuição Exponencial, PE3 – Distribuição Pearson III, LP3 – Distribuição Log-Pearson III, GEV – Distribuição Generalizada de Valores Extremos, GPA – Distribuição Generalizada de Pareto.

Observa-se que, em média, 5 das 8 distribuições analisadas podem representar os dados amostrais, com base apenas nestes testes. Além disto, com exceção da distribuição Normal, todas as distribuições analisadas foram aprovadas no mínimo em 12 das 20 amostras selecionadas, sendo que a distribuição Generalizada de Valores Extremos ajustou-se em todas elas. As distribuições de 3 parâmetros foram aprovadas em um número maior de amostras do que as distribuições de dois

parâmetros, uma vez que elas possuem um grau de liberdade a mais, o que permite que elas se ajustem melhor aos dados amostrais devido à adição do parâmetro de forma em sua modelagem. Por outro lado, a adição de mais um parâmetro à modelagem pode provocar um acréscimo nas incertezas dos parâmetros populacionais estimados. Além disso, erros de amostragem podem induzir erroneamente a seleção de modelos com maior número de parâmetros. Com isto, vê-se a importância da adoção de outros critérios, além dos testes de ajuste tradicionais, para a seleção e classificação das distribuições de probabilidades analisadas, uma vez que tais testes não são capazes de fazer qualquer distinção entre as distribuições aprovadas. Um exemplo de tal fato encontra-se ilustrado na Tabela 4, na qual foram aprovadas mais de uma distribuição de probabilidades para modelagem dos dados de cada amostra.

Em conformidade com a Tabela 5, observa-se que, em média, 2 distribuições satisfizeram os critérios adotados pelo SEAF para sua exclusão da lista de distribuições classificadas, sendo o critério de verificação dos limites amostrais responsável por 84,4% dos casos de rejeição. Vale ressaltar que as distribuições Exponencial e GPA representam, juntas, 57,8% de todos os casos de rejeição.

Tabela 5 - Distribuições rejeitadas pelo sistema

Amostra	CRITÉRIO			
	Limites superior e/ou inferior			Parcimônia
01544012	EXP			
01645000	EXP	GPA	GEV	LP3
01943000	GPA	GEV		LP3
01944004	GEV	LP3		
01944007				GPA
02044012	GPA	PE3	EXP	
02045005	GPA	EXP		GEV
02244038				GPA
01943009	PE3	EXP		
02243004	GPA	EXP		
40025000	GPA	EXP	PE3	
40050000	EXP	LP3	PE3	GEV
40100000	EXP			
40680000	GPA			
41250000	GPA	LP3	PE3	EXP
40800001	EXP	GPA	GEV	LP3
56028000	PE3			
56075000	GPA			
56415000	GPA	EXP		LP3
56500000	EXP			

Nota: NOR – Distribuição Normal, LNR – Distribuição Lognormal, GUM – Distribuição de Gumbel, EXP – Distribuição Exponencial, PE3 – Distribuição Pearson III, LP3 – Distribuição Log-Pearson III, GEV – Distribuição Generalizada de Valores Extremos, GPA – Distribuição Generalizada de Pareto.

Os resultados finais das análises realizadas pelo sistema SEAF estão apresentados na Tabela 6, em ordem decrescente de níveis médios de confiança atribuídos pelo SEAF a cada distribuição não rejeitada anteriormente. Na Tabela 6, a primeira coluna após a de identificação da amostra, lista

as distribuições com maiores níveis de confiança associados durante as análises realizadas pelo sistema, ou sejam, as escolhas primeiras do sistema.

Tabela 6 – Distribuições classificadas pelo sistema para as amostras selecionadas

Amostra	Distribuições classificadas				
	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a
01544012	GEV	LP3	GUM	PE3	LNR
01645000	GUM	PE3	LNR		
01943000	LNR	GUM	PE3	NOR	
01944004	GUM	LNR	PE3		
01944007	GEV	EXP	LP3	PE3	
02044012	GEV	LP3	GUM	LNR	
02045005	PE3	GUM	LP3	LNR	
02244038	EXP	PE3	GEV	LP3	GUM
01943009	GEV	LP3	LNR		
02243004	GEV	LP3	PE3	GUM	LNR
40025000	GEV	LP3			
40050000	GUM	LNR			
40100000	GEV	LNR			
40680000	GEV	LP3			
41250000	GEV	LNR	GUM		
40800001	GUM	LNR	PE3		
56028000	GPA	EXP	GEV	LP3	LNR
56075000	GEV	LP3	EXP	PE3	
56415000	GEV	LNR	GUM	PE3	
56500000	LNR	GEV			

Nota: NOR – Distribuição Normal, LNR – Distribuição Lognormal, GUM – Distribuição de Gumbel, EXP – Distribuição Exponencial, PE3 – Distribuição Pearson III, LP3 – Distribuição Log-Pearson III, GEV – Distribuição Generalizada de Valores Extremos, GPA – Distribuição Generalizada de Pareto.

Conforme pode ser visto na Tabela 4, de modo geral, para 5 das 8 distribuições analisadas, não foram obtidas evidências estatísticas que as excluíssem da análise. Para dar continuidade ao processo de classificação, as distribuições aprovadas nos testes estatísticos foram submetidas a uma análise subjetiva, a qual envolve um conjunto de critérios pessoais, os quais podem conduzir a diferentes escolhas, quando comparadas àquelas obtidas por outros especialistas.

A fim de avaliar a performance do sistema SEAF foram realizados dois experimentos. No primeiro, as mesmas 20 amostras de alturas diárias de chuva e vazões médias diárias máximas anuais observadas em estações pluviométricas e fluviométricas localizadas na região sudeste do Brasil, que haviam sido objeto de análise pelo sistema SEAF, foram também submetidas à análise de frequência por parte de um painel de peritos. Em um segundo experimento, aplicou-se o sistema SEAF a amostras de tamanho típico retiradas de populações sintéticas correspondentes às distribuições candidatas.

O primeiro experimento para avaliar a performance do sistema SEAF consistiu na formação inicial de um painel composto por 4 profissionais especialistas em análise de frequência. A cada membro do painel foram enviados, por correio eletrônico, as 20 amostras selecionadas, os respectivos gráficos de quantis classificados versus probabilidades empíricas, as principais estatísticas amostrais calculadas por momentos convencionais e momentos-L, bem como uma carta

contendo as orientações necessárias à condução do experimento. Além disto, foi enviada uma descrição da versão preliminar das regras heurísticas do sistema; a esse respeito, salientou-se na carta que caberia aos peritos decidir sobre a leitura e a emissão de eventuais comentários ou avaliações sobre as regras heurísticas.

Em sua formação final, o painel de especialistas ficou composto por dois estatísticos, um engenheiro-hidrólogo e um engenheiro professor universitário, todos com conhecimento específico e experiência na área de análise de frequência de variáveis hidrológicas e/ou geofísicas.

Cada membro do painel analisou o conjunto de amostras e selecionou, para cada amostra, uma distribuição de probabilidades com base em suas regras heurísticas pessoais; 3 dos 4 peritos escreveram breves justificativas a respeito de suas escolhas. Os resultados enviados por cada especialista estão apresentados na Tabela 7. Conforme norma de conduta pré-estabelecida, os membros do painel de especialistas não foram nominalmente identificados.

Tabela 7 – Resultados enviados pelos membros do painel de especialistas

Amostra	Especialista				n ^o de Coincidências	
	1 ^o	2 ^o	3 ^o	4 ^o	≥ 2	= 3
01544012	LNR	GEV	GEV	EXP	GEV	
01645000	LNR	GEV	GUM	GUM	GUM	
01943000	LNR	GEV	GUM	LNR	LNR	
01944004	LNR	GEV	GUM	EXP		
01944007	LNR	GEV	GEV	EXP	GEV	
02044012	LP3	GEV	GEV	EXP	GEV	
02045005	LP3	GEV	GUM	LP3	LP3	
02244038	LP3	GEV	GEV	LP3		
01943009	LNR	GEV	GEV	GEV	GEV	GEV
02243004	LNR	GEV	GEV	EXP	GEV	
40025000	LP3	GEV	GPA	GEV	GEV	
40050000	LNR	GEV	GEV	EXP	GEV	
40100000	LNR	GEV	GEV	LP3	GEV	
40680000	LP3	GEV	GEV	LP3		
41250000	LP3	GEV	GEV	GEV	GEV	GEV
40800001	LNR	GEV	GUM	GUM	GUM	
56028000	LP3	GEV	GPA	LP3	LP3	
56075000	LNR	GEV	GEV	GEV	GEV	GEV
56415000	LNR	GEV	GEV	LP3	GEV	
56500000	LNR	GEV	GEV	GEV	GEV	GEV

Nota: NOR – Distribuição Normal, LNR – Distribuição Lognormal, GUM – Distribuição de Gumbel, EXP – Distribuição Exponencial, PE3 – Distribuição Pearson III, LP3 – Distribuição Log-Pearson III, GEV – Distribuição Generalizada de Valores Extremos, GPA – Distribuição Generalizada de Pareto.

De acordo com os resultados, para nenhuma amostra ocorreu uma escolha consensual entre os todos especialistas. Trata-se de um fato já esperado, uma vez que a formação do painel é composta por especialistas com diferentes experiências em análise de frequência, bem como com diferentes expectativas e visões em relação aos produtos da análise em si, tendo cada um conduzido a escolha do modelo apropriado de forma diferenciada.

O especialista-1 fez sua análise com o auxílio do aplicativo “GenStat”, no qual a medida da qualidade de ajuste é expressa em termos de uma ‘estatística de desvio’ ou “*deviance statistic*”.

Quanto menor o valor desta estatística, tanto melhor é o seu ajuste. O critério utilizado pelo especialista-1 foi objetivamente escolher a distribuição que tivesse a menor “*deviance statistic*”. O universo de distribuições utilizadas na análise feita pelo aplicativo “GenStat” não foi o mesmo daquele considerado pelo SEAF. Para efeito da presente avaliação, somente foram consideradas as distribuições analisadas pelo “GenStat” que também fazem parte do conjunto de distribuições analisadas pelo SEAF.

O especialista-2 não acredita que seja possível identificar um único modelo como a distribuição populacional de uma série, cujo tamanho amostral seja da ordem de 50 anos de dados. Por outro lado, ele preconiza o uso de informações regionais para a identificação da distribuição apropriada à série em questão. Em sua análise, ele presumiu não haver diferenças climáticas ou geomorfológicas entre as estações que pudessem proporcionar significativa heterogeneidade regional do ponto de vista da frequência de ocorrência das variáveis em estudo. Em consequência, os seus resultados foram obtidos através de análise de frequência regional, considerando separadamente os dados de precipitação e vazão, ambos sob a premissa de uma única região homogênea. A metodologia utilizada pelo especialista-2 é aquela apresentada em Hosking e Wallis (1997).

O especialista-3 utilizou critérios subjetivos de análise similares aos implementadas pelo sistema SEAF. Seus resultados estão fundamentados basicamente na análise visual das distribuições ajustadas em papel de probabilidades e das razões-L amostrais grafadas no diagrama de momentos-L. Além disto, ele utiliza os testes de Filliben e Kolmogorov-Smirnov para verificar a aderência da distribuição ajustada à empírica.

O especialista-4 conduziu o processo de análise de modo semelhante ao empregado pelo especialista-3, contudo dando maior ênfase ao ajuste dos dados à cauda superior das distribuições.

Os resultados obtidos foram agrupados em dois grupos para avaliação da performance do sistema. O primeiro grupo, onde estão reunidas as amostras com duas ou mais soluções coincidentes entre os especialistas, está apresentado na coluna de título “ ≥ 2 ” da Tabela 7. Comparando os resultados deste grupo a seus correspondentes obtidos pelo sistema SEAF, verifica-se que estes valores são em 76,5% dos casos coincidentes com a primeira escolha do sistema. Considerando as duas primeiras escolhas do SEAF este valor aumenta para 82,4% chegando a 94,1%, quando todas as distribuições classificadas estiverem sendo consideradas. Já o segundo grupo, formado pelas amostras com três soluções coincidentes entre os especialistas, está apresentado na coluna de título “ $= 3$ ” da Tabela 7. Comparando os resultados deste grupo aos obtidos pelo SEAF, os valores, acima apresentados para o primeiro grupo, passam para 75%, 100% e 100%, respectivamente. Um resumo desta mesma avaliação, com os resultados individuais dos especialistas, está apresentado na Tabela 8.

Tabela 8 – Comparação entre os resultados do painel e do SEAF

Critério de escolha	Especialista				n ^o de Coincidências	
	1 ^o	2 ^o	3 ^o	4 ^o	≥ 2	= 3
1 ^a	10,0	55,0	70,0	35,0	76,5	75,0
1 ^a ou 2 ^a	50,0	60,0	85,0	50,0	82,4	100,0
entre as Classificadas	85,0	70,0	90,0	65,0	94,1	100,0

No segundo experimento foram geradas, pelo método de Monte Carlo, amostras sintéticas com base nos momentos-L e razões-L correspondentes às 20 amostras selecionadas. Para cada conjunto de momentos-L e razões-L amostrais, foram geradas 8 séries sintéticas correspondentes às formas paramétricas das distribuições candidatas, exceto nos casos onde a distribuição ajustada fosse limitada na cauda superior. No total, foram geradas 139 séries sintéticas com distribuições populacionais conhecidas, as quais foram submetidas ao sistema SEAF com o intuito de avaliar o número de acertos do sistema e se há qualquer tendência em suas análises. Os resultados por este experimento estão apresentados na Tabela 9.

Tabela 9 – Resultados obtidos pelo SEAF na avaliação das séries sintéticas

Amostras sintéticas	Porcentagem de acertos				
	Distribuição			Família	
	1 ^a	2 ^a	Classificadas	1 ^a	2 ^a
NOR	50,0	70,0	95,0	90,0	100,0
LNR	25,0	65,0	95,0	30,0	75,0
GUM	40,0	55,0	95,0	70,0	80,0
EXP	20,0	30,0	95,0	35,0	60,0
PE3	15,0	45,0	90,0	20,0	50,0
LP3	6,3	31,3	81,3	25,0	50,0
GEV	56,3	75,0	81,3	62,5	93,8
GPA	0,0	57,1	57,1	28,6	85,7
TOTAL	28,8	53,2	89,2	46,8	73,4

Nota: NOR – Distribuição Normal, LNR – Distribuição Lognormal, GUM – Distribuição de Gumbel, EXP – Distribuição Exponencial, PE3 – Distribuição Pearson III, LP3 – Distribuição Log-Pearson III, GEV – Distribuição Generalizada de Valores Extremos, GPA – Distribuição Generalizada de Pareto.

Comparando os resultados obtidos pelo sistema SEAF, relativos às séries sintéticas, verifica-se que em 53,2% dos casos o sistema identifica, na primeira ou na segunda escolha, a distribuição que gerou a série sintética analisada. Considerando todas as distribuições classificadas pelo sistema, este número aumenta para 89,2% dos casos. Observa-se também que, em 73,4% dos casos, a distribuição que gerou a série sintética, ou sua ‘parente’ da mesma família, estão classificadas pelo sistema como a primeira ou a segunda escolha.

Vale ressaltar que as amostras extraídas das séries sintéticas modeladas pela distribuição Generalizada de Pareto tiveram em sua maioria o parâmetro de forma positivo em seus ajustes. Tal fato é condição para que haja, neste modelo, um limite máximo à direita, o que explica a baixíssima performance do sistema ao modelo, principalmente em relação à primeira escolha, uma vez que o sistema SEAF foi construído para rejeitar distribuições limitadas na direção dos máximos.

À exceção da distribuição Generalizada de Pareto, verifica-se que, em acima de 80,0% dos casos, o modelo populacional das amostras geradas está presente na lista de distribuições classificadas pelo sistema SEAF, o que mostra a coerência dos resultados apresentados.

Avaliando a primeira escolha do sistema, verifica-se que o nível de acerto está abaixo dos 50%, à exceção das amostras sintéticas geradas pelas distribuições Normal e GEV, as quais apresentaram níveis de acerto de 50% e 56,3%, respectivamente. Conforme os resultados apresentados na Tabela 10, observa-se que não há uma tendência ou preferência por um modelo de distribuição de probabilidades específico.

Tabela 10 – Primeira escolha do SEAF para as amostras sintéticas

Amostras sintéticas	1ª escolha do SEAF							
	NOR	LNR	GUM	EXP	PE3	LP3	GEV	GPA
NOR	50,0	5,0	5,0	0,0	40,0	0,0	0,0	0,0
LNR	5,0	25,0	10,0	0,0	40,0	5,0	15,0	0,0
GUM	0,0	25,0	40,0	0,0	5,0	0,0	30,0	0,0
EXP	0,0	0,0	5,0	20,0	35,0	5,0	20,0	15,0
PE3	5,0	5,0	30,0	15,0	15,0	0,0	20,0	10,0
LP3	0,0	18,8	12,5	6,3	12,5	6,3	37,5	6,3
GEV	0,0	12,5	6,3	0,0	12,5	12,5	56,3	0,0
GPA	0,0	0,0	14,3	28,6	0,0	14,3	42,9	0,0
TOTAL	8,6	12,2	15,8	7,2	22,3	4,3	25,2	4,3

Nota: NOR – Distribuição Normal, LNR – Distribuição Lognormal, GUM – Distribuição de Gumbel, EXP – Distribuição Exponencial, PE3 – Distribuição Pearson III, LP3 – Distribuição Log-Pearson III, GEV – Distribuição Generalizada de Valores Extremos, GPA – Distribuição Generalizada de Pareto.

De acordo com os dois experimentos realizados é possível afirmar que o desempenho do sistema SEAF foi satisfatório, podendo ser comparado àquele de um especialista em análise de frequência de variáveis hidrológicas. Nesse sentido, o SEAF apresenta-se como uma ferramenta bastante útil de auxílio à escolha de uma distribuição de probabilidade entre diversos modelos candidatas. Conforme apresentado na Tabela 4, em média, 5 das 8 distribuições candidatas podem estatisticamente representar as amostras selecionadas. Entretanto, o ato de escolher uma entre delas envolve a aplicação de critérios subjetivos, os quais dependem do especialista envolvido. Isto justifica a total falta de consenso entre os membros do painel. Evidentemente, o sistema não dispensa a capacitação técnica e o discernimento próprio do ser humano no processo de seleção de uma distribuição de probabilidades. Entretanto, os resultados mostram que o protótipo SEAF demonstrou sua utilidade como ferramenta de auxílio à decisão e instrumento de valor para o melhor entendimento dos métodos de análise de frequência por parte de estudantes e hidrólogos não-especialistas.

4 – CONCLUSÕES E RECOMENDAÇÕES

Neste artigo, descreveu-se um protótipo de um sistema especialista para auxiliar na condução da análise de frequência local de eventos máximos anuais de variáveis hidrológicas e hidrometeorológicas, particularmente no que se refere à escolha do modelo paramétrico mais

apropriado. O sistema, denominado SEAF (Sistema Especialista de Análise de Frequência), (i) é instrumentado para analisar a veracidade das premissas necessárias para a análise de frequência convencional, a saber: as hipóteses de homogeneidade e independência serial, (ii) possui meios para identificar a presença de *outliers* na amostra, e (iii) tem instrumentos para extrair das informações numéricas calculadas as justificativas necessárias para selecionar um ou um pequeno número de modelos paramétricos apropriados, dentre o conjunto formado pelas seguintes distribuições de probabilidades: Normal, Log-Normal de 2 parâmetros, Gumbel, Generalizada de Valores Extremos, Exponencial, Generalizada de Pareto, Pearson III e Log-Pearson III.

No SEAF, a plausibilidade das distribuições candidatas de 2 e 3 parâmetros, quantificada por meio da atribuição de um nível de confiança, é feita com base, primeiramente, nas respectivas propriedades da assimetria-L e curtose-L obtidas por simulação de Monte Carlo. O raciocínio adotado no SEAF prossegue com o teste de aderência de Filliben, o qual serve para atribuir um novo grau de confiança, independente do anterior, para a distribuição em análise; ressalte-se aqui que, durante o processo, o nível de confiança previamente atribuído a cada uma das distribuições candidatas não é alterado, o que faz com que o modo de raciocínio possa ser considerado indutivo e monotônico. Finalmente, a combinação matemática de todos os níveis de confiança já atribuídos e do número de parâmetros estimados fornece o critério de parcimônia estatística para discriminar entre distribuições da mesma família. Tal como menção anterior, o SEAF classifica as distribuições que não foram descartadas durante o decurso da análise por ordem decrescente de seus respectivos níveis de confiança médios.

O sistema teve seu desempenho analisado por meio de dois experimentos. No primeiro, o SEAF foi aplicado a 20 amostras relativamente longas de alturas diárias de chuva e vazões médias diárias máximas anuais observadas em estações pluviométricas e fluviométricas localizadas na região sudeste do Brasil, as quais foram também submetidas à análise de frequência por parte de um painel de peritos. Em um segundo experimento, aplicou-se o sistema SEAF a amostras de tamanho típico retiradas de populações sintéticas correspondentes às distribuições candidatas. Conforme demonstram os resultados em ambos os experimentos o sistema SEAF desempenhou-se satisfatoriamente. Não obstante a falta de consenso entre especialistas e as complexidades inerentes à análise de frequência local de eventos máximos anuais de variáveis hidrológicas, é uma conclusão deste artigo que o sistema SEAF, quando utilizado em amostras de tamanho não muito inferior, digamos, a 30 anos, é capaz de fornecer importantes diretrizes a hidrólogos não-especialistas, sobre a escolha de uma distribuição de probabilidades apropriada, entre um conjunto de possíveis modelos candidatos. Cabe lembrar, entretanto, que as regras heurísticas para a seleção de uma distribuição de probabilidades, implementadas no sistema SEAF, representam aproximações de um certo padrão de raciocínio que reflete as convicções e as preferências dos autores deste trabalho.

Dada a subjetividade, sempre presente em tais regras e inerente à problemática em si, outras convicções e preferências podem ser implementadas em outros sistemas semelhantes. De qualquer modo, o presente trabalho vem demonstrar, por meio do protótipo SEAF, a possibilidade de construção de um sistema de auxílio à decisão para a análise de frequência local de eventos hidrológicos máximos anuais, com benefícios prováveis no que concerne a correta prescrição de variáveis hidrológicas características para o projeto e a operação de estruturas de aproveitamento de recursos hídricos, assim como a mitigação dos danos devidos às cheias.

Os resultados representam um primeiro passo para a construção de um sistema especialista completo para análise de frequência de variáveis hidrológicas. Um futuro sistema completo poderia contemplar decisões interdependentes sobre:

1. Se a análise deve ser local ou regional, com base no cotejo entre a quantidade de informações disponíveis e o tempo de retorno até o qual os quantis devem ser estimados.
2. Se a análise deve se processar com base em séries anuais ou parciais e, nesse último caso, para qual valor limiar de referência as excedências deverão ser consideradas.
3. Quais outros modelos poderiam pertencer ao elenco de distribuições candidatas.
4. O aperfeiçoamento de metodologias para identificação de *ouliers* e mecanismos de definição de ações para sua permanência ou retirada da amostra.
5. A inclusão de informações históricas úteis para a redefinição de probabilidades empíricas com menor grau de incerteza.
6. Uma vez que há divergência entre os padrões de raciocínios adotados entre os especialistas, implementar um sistema que agrupe vários padrões de análises e que dê ao usuário a liberdade em decidir qual o padrão de raciocínio será utilizado em suas análises.
7. O desenvolvimento de critérios mais objetivos para a definição das formas das funções de pertinência usadas para atribuição de níveis de confiança.
8. A flexibilização à escolha do usuário, ou mesmo a definição, dos limites inferiores acima dos quais a pertinência é atribuída.
9. Uma melhor análise dos efeitos da adoção do critério de monotonicidade ao padrão de raciocínio do sistema.
10. O desenvolvimento de uma *interface* mais informativa sobre a extensão e conseqüências de cada uma das decisões que estão sendo tomadas pelo sistema, de modo a garantir maior envolvimento e maior elaboração da análise em curso, por parte do usuário, entre outras.

BIBLIOGRAFIA

- BOBÉE, B.; RASMUSSEN, P. (1995). “*Recent advances in flood frequency analysis*”. U.S. National Report to IUGG, 1991-1994, Rev. Geophysics, v. 33 Suppl. (<http://earth.agu.org/revgeophys/bobee01/bobee01.htm>), 1995.
- CANDIDO, M. O. (2003). “*SEAF – Um protótipo de um sistema especialista para análise de frequência local de eventos hidrológicos máximos anuais*”. EEUFMG, Belo Horizonte-MG, 2003, 174p.
- DAVIS, E. G.; NAGHETTINI, M. C. (2001). *Estudo de Chuvas Intensas: Projeto Rio de Janeiro*. CPRM, Belo Horizonte, 2001.
- HOSKING, J. R. M.; WALLIS, J. R. (1997). *Regional Frequency Analysis - An Approach Based on L-Moments*. Cambridge University Press, Cambridge, Reino Unido, 1997.
- MUHARA, G. (2001). “*Selection of flood frequency model in Tanzania using L-moments and the region of influence approach*” in Second WARFSA/WaterNet Symposium: Integrated Water Resources Management - Theory, Practice, Cases. Cape Town, África do Sul, 2001.
- POTTER, K. W. (1987). “*Research on flood frequency analysis : 1983-1986*”. Rev. Geophysics, V. 26, n. 3, p. 113-118, 1987.
- VOGEL, R. M.; FENNESSEY, N. M. (1993). “*L-moment diagrams should replace product moment diagrams*”. Water Resources Research, v. 29, n. 6, p. 1745-1752, 1993.
- ZVI, A. B.; AZMON, B. (1997). “*Joint use of L-moment diagram and goodness-of-fit test: a case study of diverse series*”. Journal of Hydrology, 198, p. 245-259, 1997.