RBRH – Revista Brasileira de Recursos Hídricos *Volume 9 n.3 Jul/Set 2004, 69-79*

Regionalização de Parâmetros Hidrológicos Utilizando Análise Difusa

Rosana de Fátima Colaço Gibertoni, Eloy Kaviski, Miriam Rita Moro Mine
Instituto de Tecnologia para o Desenvolvimento (Lactec) - Curitiba, PR
rosana@lactec.org.br, ekaviski@lactec.org.br, mrmine.dhs@ufpr.br

Recebido: 09/05/03 revisado: 23/01/04 aceito: 09/09/04

RESUMO

O objetivo deste trabalho consiste em analisar a aplicação da análise difusa na regionalização de parâmetros hidrológicos em bacias hidrográficas de pequeno e médio porte. A análise difusa é um método de multicritérios para impactos sob incerteza, sendo adequado à modelagem e ao gerenciamento de recursos hídricos. Os estudos realizados em hidrologia deparam-se com muitas incertezas, ligadas principalmente às características das bacias hidrográficas e à imprecisão e escassez de dados hidrológicos.
As técnicas de lógica difusa utilizadas foram a análise de agrupamento de dados pelo método "c-means" e a regressão por metas difusas. Os parâmetros hidrológicos regionalizados são as séries de vazões médias diárias e as vazões máximas anuais.
As conclusões do trabalho são baseadas na aplicação de índices de desempenho sobre os resultados e na comparação destes com os resultados obtidos por outros métodos de regionalização. Através do estudo de caso foi possível verificar que a aplicação de ferramentas da análise difusa como técnica de modelagem de recursos hídricos, como a regionalização de parâmetros hidrológicos, fornece resultados satisfatórios e de alguma forma melhores que os obtidos com a utilização de métodos estatísticos padrões.

Palavras-chave: lógica difusa, regionalização, parâmetros hidrológicos.

INTRODUÇÃO

A regionalização hidrológica é um tema bastante difundido, porém, dada a sua importância no contexto atual de recursos hídricos, muitos hidrólogos têm realizado estudos que visam melhorar seus resultados. Este trabalho propõe-se a analisar a aplicação da análise difusa na regionalização de parâmetros hidrológicos em bacias hidrográficas de pequeno a médio porte (50 a 5000 km²), através da classificação difusa e regressão difusa. A área de recursos hídricos é bastante adequada para implementação da lógica difusa, dada a incerteza nos processos envolvidos, escassez de dados e existência de alternativas conflitantes na tomada de decisões no gerenciamento dos sistemas de recursos hídricos.

Os parâmetros hidrológicos focados neste artigo referem-se às séries de vazões médias diárias (média, coeficiente de variação e momentos ponderados por probabilidades), e de vazões máximas anuais (média e coeficiente de variação).

As conclusões do trabalho são baseadas na comparação dos índices de desempenho dos resultados deste com os índices de desempenho correspondentes obtidos pelo método de regionalização proposto por Kaviski (1992).

ANÁLISE DIFUSA: CONCEITOS GERAIS

A análise difusa é classificada como uma das técnicas dos Sistemas Inteligentes, os quais são inspirados no comportamento humano ou da natureza. A lógica difusa (nebulosa ou "*fuzzy*") é baseada na linguagem natural da comunicação humana, onde os conceitos não são precisos, mas aproximados. O conceito de conjunto difuso foi introduzido em 1965, por Lotfi A. Zadeh, da Universidade da Califórnia em Berkeley (Zadeh, 1965; Azevedo et al., 2000).

Comumente, a lógica difusa pode utilizar variáveis lingüísticas, as quais admitem como valores termos primários, como "muito pequeno", "altamente viável do ponto de vista econômico", "pouco frio", etc. Entretanto, a lógica difusa também permite trabalhar com variáveis numéricas, sob a forma dos chamados números difusos.

São apresentados, de forma resumida, alguns conceitos importantes da análise difusa utilizados neste trabalho. O leitor encontra em Gibertoni (2002) o assunto de forma mais aprofundada.

Os dois principais conceitos associados à lógica difusa são o conjunto difuso e a função de pertinência. A idéia do conjunto difuso pode ser tomada como uma generalização do conceito do conjunto clássico, no qual uma função qualquer pode assumir apenas os valores 1 ou 0 para cada indivíduo, caracterizando-o como membro ou não membro do conjunto que está sendo analisado. No caso dos conjuntos difusos a função analisada, denominada função de pertinência (μ), foge da solução binária, podendo assumir valores dentro de uma determinada faixa.

Um conjunto difuso, contido no conjunto dos Números Reais, que possui as propriedades de normalidade e convexidade é denominado número difuso. Os números difusos são amplamente utilizados nas aplicações práticas, constituindo o tipo mais básico de conjunto difuso, pois permitem a utilização de valores numéricos sob forma aproximada, além de serem de fácil compreensão.

O conjunto difuso convexo tem nos dois lados da função em torno do seu ponto de pertinência máxima, valores sempre crescentes ou sempre decrescentes, o mesmo não ocorrendo para o conjunto difuso não-convexo. Um conjunto difuso é dito normal se há pelo menos um ponto $x \in X$ com pertinência máxima igual a 1.

O nível de credibilidade h é um valor pré-especificado correspondente ao menor grau de uma função de pertinência a ser aceito pelo especialista. Um conceito ligado ao nível de credibilidade h é o "conjunto ao nível h", que corresponde ao intervalo de valores de x em X, tais que $\mu(x) > h$.

Muitas aplicações interessantes em recursos hídricos e ambientais podem ser encontradas em publicações técnico-científicas. Yu et al. (2000) apresenta um modelo de previsão de chuva, cujos parâmetros foram estimados através da regressão por metas difusas. Em Campana (1995) é apresentada a obtenção de um hidrograma unitário através de técnicas de regionalização baseadas na regressão linear difusa. Galvão (1995) apresenta o desenvolvimento de um método de previsão climática no nordeste para uso em operação de reservatórios, onde foram considerados conceitos representados por variáveis difusas. O trabalho desenvolvido em Gates et al. (1991) apresenta uma metodologia para tomada de decisões considerando um sistema de irrigação implantado em um contexto que possui diversos critérios de planejamento. Em Bogardi et al. (1983) foi desenvolvida uma metodologia de apoio ao gerenciamento de um aqüífero, onde objetivos de caráter ambiental são implementados com conjuntos difusos.

TÉCNICAS UTILIZADAS

Classificação difusa

A classificação (análise de agrupamentos de dados) pode ser caracterizada como uma das técnicas de análise estatística multivariada, através da qual se efetua a divisão de um grupo em sub-grupos que apresentem similaridade nos fenômenos considerados. Os sub-grupos determinados são chamados de regiões homogêneas, para as quais são definidas relações únicas que expliquem os parâmetros a serem regionalizados em função das características descritivas do local (Kaviski, 1992).

O método proposto para classificação difusa, "*fuzzy c-means*" (Hall & Minns, 1999), permite trabalhar com variáveis ordinárias ou difusas. No caso de considerar que as divisões dos dados são abruptas (determinísticas), cada ponto é associado com somente uma classe. No caso da classificação difusa, cada ponto pode possuir um valor de pertinência a mais de uma classe. Na classificação difusa é definida uma família de K conjuntos difusos A_k, sendo k = 1, 2, ..., K, no universo de discurso de X, onde cada conjunto representa uma classe k. Os valores de aderência podem ser representados na forma de uma matriz, denominada matriz de partição difusa U, onde o número de linhas é igual a K e o número de colunas é igual a N. Em cada linha da matriz U, representa-se os valores das pertinências para cada um dos N pontos pertencentes à classe k. Dado que o número de valores de pertinência que cada ponto pode assumir em determinada classe é infinito, é necessário definir um critério de classificação ou função objetivo para determinar os elementos da matriz U. Na classificação difusa, a função objetivo é baseada na distância Euclidiana d_{nk} entre cada dado pontual x_n e o centro de classe ν_k:

$$d_{nk} = d(\underline{x}_n - \underline{\upsilon}_k) = \sqrt{\sum_{c=1}^{C} (x_{nc} - \upsilon_{kc})^2} \tag{1}$$

Na expressão (1), C representa o número total de características descritivas do local. A função objetivo é dada por:

$$F_{obj} = \sum_{n=1}^{N} \sum_{k=1}^{K} (\mu_{nk})^r (d_{nk})^2 \tag{2}$$

Na equação (2), r é o parâmetro que controla o nível de incerteza ("*fuzziness*") no processo de classificação; e μ_{nk} representa os valores de pertinência (elementos da matriz U). Para r = 1, o agrupamento passa a ter aderências abruptas às classes; com valores de r aumentando, a classificação torna-se cada vez mais difusa. Conforme Ross, citado por Hall & Minns (1999), um intervalo razoável para r é $1{,}25 \leq r \leq 2$.

As coordenadas do centro da classe para determinada característica c descritiva do local na classe k podem ser dadas por:

$$\upsilon_{kc} = \frac{\sum_{n=1}^{N} (\mu_{nk})^r x_{nc}}{\sum_{n=1}^{N} (\mu_{nk})^r} \tag{3}$$

Para se obter o valor difuso ótimo do número de classes K, minimiza-se a função objetivo F_{obj} expressa em (2). O nível de precisão r é pré-estabelecido, para o qual pode-se chegar à melhor solução iterativamente.

Regressão por metas difusas

A análise de regressão difusa descreve a variável dependente y na forma de um número difuso. Em todos os casos de regressão difusa, os seus coeficientes são considerados como números difusos. A variável independente x pode ou não ser tomada como um número difuso. O caminho pelo qual a regressão difusa é conduzida consiste em determinar os coeficientes da regressão que respeitem a minimização da imprecisão de y (Bardossy et al., 1990; Tanaka et al., 1982). O modelo de regressão difusa pode ser genericamente representado por:

$$y^*(t) = \underline{x}(t)^T \underline{a}^* + \varepsilon \qquad (4)$$

Na expressão (4), $y^*(t)$ representa a variável dependente, sendo que t=1...T e T é o número total de pontos amostrais; $\underline{x}(t)$ é o vetor de variáveis independentes, $\underline{x}(t)=[x_1(t),...,x_{k-1}(t)]$; $\underline{a}^*$ é o vetor de coeficientes de regressão a serem estimados; e ε denota o erro. Na regressão difusa, considera-se que o vetor $\underline{a}^*$ é composto por números difusos, sendo representado pela equação (5), onde P é o número total de coeficientes a serem estimados:

$$\underline{a}^* = [a^*(1), \ldots, a^*(P)] \qquad (5)$$

A regressão por metas difusas visa maximizar o nível de credibilidade, através da pré-definição da largura difusa (intervalo da variável prevista, definido pelo nível de credibilidade h). O problema recai na programação objetiva com múltiplas metas difusas e pode ser representado por funções de pertinência lineares dos tipos não-crescente e não-decrescente, cuja forma de obtenção é apresentada no Apêndice.

Para o conjunto de n observações (x e y) existem n metas difusas a serem analisadas no modelo de programação difusa. O problema é concebido como a maximização dos mínimos níveis de credibilidade das metas sujeitas às restrições. Cada restrição é representada por uma função de pertinência triangular, que é na verdade a interseção das funções de pertinência não-crescente e não-decrescente. Assim, cada meta restrita deve ser representada por duas inequações de restrições. O maior grau de ajuste ocorre quando o valor observado da variável dependente y é igual ao seu valor previsto.

Tomando como representação do modelo de regressão a ser resolvido a forma $y_i = \sum_{i=1}^{n} a_i x_i + e_i$, o modelo de programação de metas difusas resulta:

maximizar: $$\sum (h_{i1} + h_{i2}), \qquad (6)$$

restrito a:

$$\sum_{j=1}^{n} a_{ij} x_{ij} + h_{i1} e_i \leq y_i + e_i \qquad \forall i = 1, \ldots, T, \qquad (7)$$

$$\sum_{j=1}^{n} a_{ij} x_{ij} - h_{i2} e_i \geq y_i - e_i \qquad \forall i = 1, \ldots, T, \qquad (8)$$

$$0 \leq h_{i1}, h_{i2} \leq 1 \qquad (9)$$

Nas expressões anteriores, y_i representa a variável dependente; x_{ij} representam as variáveis independentes, onde j=1...n e n representa o número total de variáveis independentes; a_{ij} são coeficientes a serem estimados; T corresponde ao número total de pontos amostrais; h_{i1} e h_{i2} são, respectivamente, os níveis de credibilidade da função de pertinência não-crescente e não-decrescente a serem determinados na otimização; e e_i é o termo de incerteza ou intervalo de tolerância, para o qual a função de pertinência assume valores intermediários entre 0 e 1.

MÉTODO PROPOSTO

Concepção geral

O objetivo da análise foi verificar o comportamento de algumas técnicas de lógica difusa na regionalização. O método desenvolvido tem como premissa regionalizar parâmetros de séries de vazões médias diárias e de vazões máximas anuais de maneira objetiva e que faça uso de dados básicos de fácil obtenção. A unidade hidrológica focada foi a bacia hidrográfica com área de drenagem variando de 50 a 5000 km². Os parâmetros de séries de vazões médias diárias regionalizados foram: média das vazões médias diárias $\overline{Q}$, coeficiente de variação das vazões médias diárias CV[Q], momentos ponderados por probabilidades de ordem 1 a 3 das vazões médias diárias, $MP_1[Q]$, $MP_2[Q]$ e $MP_3[Q]$. Estes parâmetros serão, neste trabalho, referidos como "Grupo Med". Os parâmetros referentes às séries de vazões máximas anuais foram: média das vazões máximas anuais $\overline{Q}_{max}$ e coeficiente de variação das vazões máximas anuais $CV[Q_{max}]$. Estes parâmetros serão, a partir daqui, concebidos como "Grupo Max".

A fim de possibilitar a avaliação desta aplicação em termos de melhoria nos resultados, selecionou-se para a análise de comparação um método de regionalização já proposto. O método escolhido para realizar a análise de comparação dos resultados foi implementado em Kaviski (1992). Foram utilizados os mesmos dados de entrada, as mesmas técnicas não difusas complementares, os mesmos estudos de caso e as mesmas aplicações para análise dos resultados do método proposto em Kaviski (1992). Este procedimento visou assegurar que as diferenças nos resultados se justificassem exclusivamente pelas técnicas de lógica difusa implementadas.

Foi utilizada a análise de agrupamentos difusos para definir K regiões homogêneas em termos das variáveis regionalizadas. Através da análise de agrupamentos foi possível determinar as variáveis classificadoras, ou seja, aquelas que foram utilizadas em uma aplicação de caso na etapa denominada de análise discriminante. Na regionalização dos parâmetros de séries de vazões médias diárias foram utilizadas as seguintes variáveis: latitude x_F, longitude y_F, altitude z_F, área de drenagem A, precipitação total média anual de longo período $\overline{P}$ e coeficiente de variação das precipitações totais anuais CV[P]. Estas variáveis também foram utilizadas na regionalização de parâmetros de séries de vazões máximas anuais, além da declividade média do talvegue principal s, $\overline{Q}$, CV[Q], a média $\overline{P}_{max1}$ e o coeficiente de variação CV[P_{max1}] dos totais precipitados máximos anuais de um dia de duração. Como $\overline{Q}$ e CV[Q] são dados regionalizados, fez-se necessário implementar primeiramente a regionalização de parâmetros de séries de vazões médias diárias.

Para cada uma das regiões homogêneas foram definidas, através da regressão difusa, as equações que determinam os parâmetros regionalizados em função das características da bacia. Os dados utilizados na obtenção das equações de regressão foram os dados observados ou regionalizados das estações do agrupamento de cada região homogênea.

Após o desenvolvimento do método de regionalização, foram implementadas as técnicas utilizadas para a aplicação do método em uma bacia hidrográfica sem informações hidrológicas situada na região do estudo. Os dados coletados referentes aos locais sem informações hidrológicas são: latitude, longitude, altitude e área de drenagem. Em função da localização e altitude desta bacia, foram determinados por regionalização (interpolação), a precipitação total média anual de longo período e o coeficiente de variação das precipitações totais anuais.

Por meio das variáveis classificadoras determinadas na análise de agrupamentos, foi realizada a análise discriminante para definir as possibilidades de que o local em análise pertença a cada uma das K regiões homogêneas. Enfim, os parâmetros regionalizados foram determinados pela equação de regressão definida para a região homogênea em que se enquadrou o local estudado.

Dados utilizados

Foram coletados dados climatológicos, fisiográficos e fluviais, de locais situados na região de estudo, no caso, o Estado do Paraná. Os dados são referentes ao ponto de instalação das estações fluviométricas selecionadas ou às bacias hidrográficas representadas por estas estações.

Foram utilizadas 55 estações fluviométricas localizadas dentro da área de interesse. Destas estações, interessaram os dados de vazões médias diárias e de vazões máximas anuais, ou seja, os parâmetros do "Grupo Med" e do "Grupo Max"; dados relacionados ao ponto de instalação da estação: x_F (graus decimais), y_F (graus decimais) e z_F (m); e dados relacionados às bacias hidrográficas drenadas pelas estações: A (km²) e s (m/km). O último ano de dados utilizado destas estações é 1982. Cabe ressaltar que não foi efetuada uma atualização dos dados a fim de assegurar que as diferenças nos resultados entre os métodos comparados pudessem ser avaliadas exclusivamente pela implementação de técnicas de análise difusa.

Foram também utilizadas diversas estações pluviométricas localizadas nos limites e no entorno da área de interesse. Das informações pluviométricas, foi extraída a $\overline{P}$ (mm), CV[P], $\overline{P}_{max1}$ (mm), CV[P_{max1}], além da latitude x_P (graus decimais), longitude y_P (graus decimais) e altitude z_P (m) da estação. Para o estudo de parâmetros de séries de vazões médias diárias e de séries de vazões máximas anuais foram utilizadas, respectivamente, 94 e 69 estações pluviométricas. Os valores de $\overline{P}$ e CV[P], de cada estação pluviométrica, foram obtidos dos dados observados entre os anos de 1954 e 1964.

Foram considerados os valores de $\overline{P}$ e CV[P] regionalizados para as estações fluviométricas obtidos diretamente de Kaviski (1992). Os valores de $\overline{P}$ e CV[P], disponíveis para as estações pluviométricas, foram transferidos para as estações fluviométricas através do método de regionalização implementado em Kaviski (1992). Neste trabalho, a regionalização de $\overline{P}$ e CV[P] foi efetuada de acordo com os seguintes critérios: para estações fluviométricas localizadas em regiões de interpolação, os valores foram obtidos por superfície "*spline*"; e para estações fluviométricas localizadas em regiões de extrapolação foi utilizada a interpolação ponderada. Como CV[P] não se encontrava disponível para todas as estações pluviométricas, este parâmetro foi regionalizado para as estações fluviométricas em função do número reduzido de dados amostrais.

Definição das regiões homogêneas

A análise de agrupamentos foi realizada usando-se a classificação difusa "*c-means*" (Hall & Minns, 1999). A técnica foi utilizada com o auxílio de ferramenta específica, implementada no software MATLAB através do "*Fuzzy Logic Toolbox*" (MathWorks, 1998).

Os melhores agrupamentos dos parâmetros hidrológicos do "Grupo Med" foram determinados em função de x_F, y_F, z_F, A, $\overline{P}$ e CV[P]. A análise de agrupamentos dos parâmetros hidrológicos do "Grupo Max" foi realizada em função de x_F, y_F, z_F, A, s, $\overline{P}$, CV[P], $\overline{P}_{max1}$, CV[P_{max1}], $\overline{Q}$ e CV[Q]. Foram realizadas todas as combinações possíveis, envolvendo as variáveis classificadoras a fim de determinar os melhores agrupamentos.

A classificação implementada pelo algoritmo "*fuzzy c-means*" consiste em determinar os elementos da matriz de partição difusa U, através da minimização da função objetivo dada por (2). Os elementos da matriz são os valores de pertinência que cada local pode assumir em cada grupo. O número de linhas da matriz U corresponde ao número de classes e o número de colunas corresponde ao número total de estações fluviométricas.

O método de "desfuzificação", processo pelo qual o valor da saída difusa é convertido em um único valor numérico, implementado no MATLAB, busca agrupar cada estação na classe em que possui a maior aderência.

O método de classificação "*fuzzy c-means*" não permite a definição a priori do número ideal de classes. Logo, este valor foi determinado iterativamente, analisando-se a vantagem de se passar de K para K+1 grupos através da verificação da conseqüência hidrológica. Nesta técnica, avalia-se a razão entre a variação entre agrupamentos V_B e a variação dentro dos agrupamentos V_W, dos parâmetros hidrológicos a serem regionalizados. Os valores de V_B e V_W são obtidos por:

$$V_B = \frac{1}{K-1}\sum_{k=1}^{K}(\overline{\theta}_{k\circ} - \overline{\theta}_{\circ\circ})^2 \tag{10}$$

$$V_W = \frac{1}{N-K}\sum_{k=1}^{K}\sum_{l=1}^{L_k}(\theta_{kl} - \overline{\theta}_{k\circ})^2 \tag{11}$$

sendo: $$\overline{\theta}_{k\circ} = \sum_{l=1}^{L_k}\theta_{kl} / L_k \tag{12}$$

$$\overline{\theta}_{\circ\circ} = \sum_{k=1}^{K}\sum_{l=1}^{L_k}\theta_{kl} / N \tag{13}$$

Nestas expressões N é o número de elementos a serem agrupados; K é o número de grupos; L_k é o número de elementos reunidos no grupo k; e θ é o parâmetro que está sendo avaliado. A passagem de K para K + 1 agrupamentos se justifica, caso a relação V_B/V_W seja crescente.

Para evitar a formação de agrupamentos com números de estações muito discrepantes entre si, o valor ideal de K foi determinado também conforme (Kaviski, 1992):

$$L_k \geq n + 1, \tag{14}$$

Para a regionalização dos parâmetros de séries de vazões médias diárias, o agrupamento adotado foi o gerado para 7 grupos em função das variáveis classificadoras x_F, y_F, $\overline{P}$ e CV[P]. Este agrupamento obteve o maior valor da média de V_B/V_W entre os agrupamentos que verificam a equação expressa em (14). Os valores da relação V_B/V_W, obtidos considerando o agrupamento adotado, para os parâmetros hidrológicos $\overline{Q}$, CV[Q], MP_1[Q], MP_2[Q] e MP_3[Q] são, respectivamente, iguais a: 1,555, 0,454, 0,801, 0,935 e 1,025. Estes resultados são melhores se comparados aos valores obtidos em Kaviski (1992), onde o agrupamento adotado considerou x_F e y_F como variáveis classificadoras, o valor de K resultou igual a 5 e os valores de V_B/V_W determinados para os parâmetros $\overline{Q}$, CV[Q], MP_1[Q], MP_2[Q] e MP_3[Q] são respectivamente iguais a: 1,260, 0,076, 0,152, 0,188 e 0,215. A Tabela 1 apresenta os agrupamentos obtidos para os métodos propostos, através de seus valores centrais.

Tabela 1 - Parâmetros para utilização de técnicas de análise discriminante – Vazões médias diárias

Grupos	Método proposto (análise difusa)			
	x_F (°)	y_F (°)	$\overline{P}$ (mm)	CV[P]
Grupo 1	23,40	50,67	1253,14	0,1783
Grupo 2	24,91	50,03	1527,86	0,1992
Grupo 3	24,53	53,03	1613,57	0,1948
Grupo 4	25,89	49,39	1448,08	0,1927
Grupo 5	26,13	51,02	1591,57	0,1689
Grupo 6	25,73	51,99	1981,40	0,1757
Grupo 7	24,94	49,25	1187,80	0,1992

Grupos	Método proposto em Kaviski (1992)	
	x_F (°)	y_F (°)
Grupo 1	23,40	50,73
Grupo 2	24,92	49,70
Grupo 3	24,53	53,03
Grupo 4	25,89	51,95
Grupo 5	25,89	49,50

Na regionalização dos parâmetros de séries de vazões máximas anuais, obteve-se o melhor agrupamento considerando-se 3 grupos em função dos parâmetros classificadores s, $\overline{P}$ e $\overline{P}_{max1}$. Os valores da relação V_B/V_W assumidos para os parâmetros hidrológicos $\overline{Q}_{max}$ e CV[Q_{max}] são, respectivamente, iguais a: 2,231 e 0,692. Em

Kaviski (1992), as variáveis classificadoras são A, $\overline{Q}$ e CV[Q], o valor de K resultou igual a 4 e os valores de V_B/V_W determinados para os parâmetros $\overline{Q}_{max}$ e $CV[Q_{max}]$ são respectivamente iguais a: 1,710 e 0,754. A Tabela 2 apresenta os agrupamentos obtidos para os métodos propostos, através de seus valores centrais.

Conforme indicação de Ross, citado por Hall & Minns (1999), os valores de r, parâmetro que especifica o nível de incerteza e é utilizado na função expressa em (2), são recomendados para variar de 1,25 a 2,0. O valor de r foi adotado igual ao máximo indicado (r=2,0), retratando assim um alto nível de incerteza a respeito da ambigüidade dos dados.

Tabela 2 - Parâmetros para utilização de técnicas de análise discriminante – Vazões máximas anuais

Grupos	Método proposto (análise difusa)		
	s (m/km)	$\overline{P}$ (mm)	$\overline{P}_{max1}$ (mm)
Grupo 1	3,8877	1338,72	85,92
Grupo 2	6,3657	1697,77	100,32
Grupo 3	12,7135	2214,75	145,83

Grupos	Método proposto em Kaviski (1992)		
	A (km²)	$\overline{Q}$ (l/s/km²)	CV[Q]
Grupo 1	1678,88	26,64	1,5948
Grupo 2	3571,00	19,58	0,8649
Grupo 3	1153,46	13,78	1,1255
Grupo 4	831,17	22,09	0,9256

Determinação das equações de regressão

O modelo de regressão adotado neste trabalho foi o por metas difusas (Chang et al., 1996). Os parâmetros hidrológicos regionalizados foram determinados por expressões que relacionam estes parâmetros com as características fisiográficas e climatológicas da bacia estudada. Estas expressões foram determinadas uma para cada grupo homogêneo definido na análise de agrupamentos. Todos os parâmetros considerados na regressão foram lognormalizados. O modelo de regressão difusa pode ser generalizado por:

$$\ln(y_i) = a_0^* + \sum_{j=1}^{p} a_j^* \ln(x_{i,j}) + \varepsilon, \qquad i=1,...,L_k. \tag{15}$$

Na expressão (15) y é a variável regionalizada descrita na forma difusa, representando os parâmetros do "Grupo Med" ou "Grupo Max"; x_j, $j = 1,\ldots,p$, representam as características fisiográficas e climatológicas da bacia e podem ser: x_F, y_F, z_F, A, $\overline{P}$ e CV[P] na regionalização de parâmetros de séries de vazões médias diárias, e x_F, y_F, z_F, A, s, $\overline{P}$, CV[P], $\overline{P}_{max1}$, $CV[P_{max1}]$, $\overline{Q}$ e CV[Q] na regionalização de parâmetros de séries de vazões máximas anuais; a_j^*, $j = 0,\ldots,p$, são os coeficientes difusos a serem determinados na regressão difusa; ε é o erro; e L_k é o número de estações do agrupamento considerado. O modelo de programação difusa é representado pelas expressões (6) a (9).

Os modelos de regressão ajustados foram avaliados por fim, pelo grau de ajuste médio $\overline{h}$ obtido pela seguinte expressão:

$$\overline{h} = \frac{\sum_{i=1}^{L_k} \min[h_{i1}, h_{i2}]}{L_k} \tag{16}$$

Os valores de h_{i1} e h_{i2} da expressão (16) correspondem aos níveis de credibilidade otimizados na estimativa dos coeficientes da regressão. Foram analisados todos os possíveis casos de combinação entre parâmetros utilizados como variáveis independentes, sendo adotado o caso que apresentou o maior grau de ajuste médio $\overline{h}$.

Através de testes de sensibilidade foi definido o termo de incerteza e, expresso nas equações (7) e (8). Para a obtenção das equações de regressão dos parâmetros o valor de e considerado é igual a 60% de y.

Não foi possível determinar as equações de regressão por metas difusas para ln(CV[Q]), pois os resultados apresentaram-se inconsistentes (valores negativos ou demasiadamente superestimados).

As Tabelas 3 a 8 apresentam os modelos de regressão obtidos através do método proposto para $\ln(\overline{Q})$, $\ln(MP_1[Q])$, $\ln(MP_2[Q])$, $\ln(MP_3[Q])$, $\ln(\overline{Q}_{max})$ e $\ln(CV[Q_{max}])$, respectivamente. Estas equações apresentaram o maior grau de ajuste médio $\overline{h}$ expresso na equação (16), cujo valor é constante destas tabelas. Os melhores ajustes foram determinados para equações com termo independente, cujo grau de ajuste médio $\overline{h}$ resultou igual a 1 na maioria dos ajustes. Porém estas equações não possuem bom comportamento na aplicação em locais que não foram utilizados na construção do modelo, estimando valores muito diferentes dos observados. Uma possível razão para isto, é que estas equações não possuindo nenhum grau de liberdade, e adaptando-se perfeitamente aos dados utilizados na sua construção, podem estimar valores degenerados.

Em Kaviski (1992), os parâmetros do "Grupo Med" foram determinados por interpolação ponderada, e os parâmetros do "Grupo Max" foram determinados por meio de equações de regressão lineares determinadas por mínimos quadrados ponderado, onde foi suposto que $\overline{Q}_{max}$ e $CV[Q_{max}]$ seguem o modelo log-normal com 2 parâmetros.

Tabela 3 - Equações de regressão para ln($\overline{Q}$).

Grupos	Parâmetros x_F	CV[P]	$\overline{h}$
Grupo 1	1,382737	4,352804	0,8975
Grupo 2	0,415806	0,166438	0,9191
Grupo 3	0,410614	-0,034672	0,9362
Grupo 4	0,560754	0,641649	0,9495
Grupo 5	0,450606	0,099260	0,9515
Grupo 6	0,418548	-0,076868	0,9483
Grupo 7	0,291288	-0,457575	0,8154

Tabela 4 - Equações de regressão para ln(MP_1[Q]).

Grupos	Parâmetros A	y_F	$\overline{P}$	$\overline{h}$
Grupo 1	0,059531	-2,120181	0,945005	0,8918
Grupo 2	0,032944	-0,839680	0,237092	0,8697
Grupo 3	-0,190526	1,436730	-0,776316	0,7751
Grupo 4	0,000094	-0,193749	-0,057171	0,9349
Grupo 5	-0,085446	0,691150	-0,473780	0,9534
Grupo 6	0,026604	-0,642959	0,137976	0,9591
Grupo 7	-0,046872	0,670079	-0,473142	0,9776

Tabela 5 - Equações de regressão para ln(MP_2[Q]).

Grupos	Parâmetros A	y_F	$\overline{P}$	$\overline{h}$
Grupo 1	0,072272	-2,747671	1,194922	0,9023
Grupo 2	-0,079109	-1,191433	0,438386	0,8775
Grupo 3	-0,248446	2,065428	-1,151989	0,8007
Grupo 4	-0,003238	-0,414945	-0,018313	0,9339
Grupo 5	-0,127156	0,615400	-0,491602	0,9574
Grupo 6	0,026985	-0,912008	0,188603	0,9608
Grupo 7	-0,052838	0,686764	-0,553602	0,9799

Tabela 6 - Equações de regressão para ln(MP_3[Q]).

Grupos	Parâmetros A	Y_F	$\overline{P}$	$\overline{h}$
Grupo 1	0,079954	-3,233353	1,397587	0,9067
Grupo 2	-0,156076	-1,467178	0,597113	0,8837
Grupo 3	-0,275471	2,166568	-1,241952	0,8157
Grupo 4	-0,007982	-0,617256	0,039698	0,9333
Grupo 5	-0,153062	0,372047	-0,403196	0,9585
Grupo 6	0,019100	-0,995517	0,178125	0,9623
Grupo 7	-0,063332	0,726463	-0,616237	0,9769

Tabela 7 - Equações de regressão para ln($\overline{Q}_{max}$).

Parâmetros	Grupos Grupo 1	Grupo 2	Grupo3
x_F	0,658119	...	2,348600
s	...	0,173615	...
$\overline{Q}$	...	1,551947	...
CV[P]	...	-1,389160	...
$\overline{P}_{max1}$	0,548671	-0,487210	...
CV[P_{max1}]	...	0,158054	1,449501
$\overline{h}$	0,8906	0,9252	0,8012

... Parâmetro não utilizado na equação de regressão

Tabela 8 - Equações de regressão para ln(CV[Q_{MAX}]).

Parâmetros	Grupos Grupo 1	Grupo 2	Grupo3
x_F	-0,562072	-0,2010125	...
y_F	...	3,1806229	...
z_F	...	...	0,540900
A	...	-0,0954916	...
S	...	-0,2081925	...
$\overline{Q}$	...	-0,313342	...
CV[Q]	...	0,7759554	...
$\overline{P}$	...	-0,3685427	...
CV[P]	...	0,8308353	...
$\overline{P}_{max1}$	0,239364	-1,3728395	-0,762994
CV[P_{max1}]	...	0,1961126	-0,003393
$\overline{h}$	0,6942	0,7963	0,8623

... Parâmetro não utilizado na equação de regressão

ANÁLISE DOS RESULTADOS

O método de regionalização desenvolvido neste trabalho foi aplicado em alguns locais, a fim de efetuar a análise dos resultados dos estudos de caso. Desta maneira, foi possível verificar a aplicabilidade do método através da comparação com os dados observados e verificar se o método proposto fornece melhorias nos resultados através de comparações com o método apresentado em Kaviski (1992).

Para possibilitar a análise dos resultados, os locais estudados são estações fluviométricas que possuem dados observados, porém não foram utilizadas no desenvolvimento dos estudos de casos.

Os índices estatísticos utilizados para avaliar os métodos que foram comparados são o desvio médio quadrático $\overline{d}_q$, e absoluto $\overline{d}_a$, e são aplicados para cada parâmetro regionalizado p, quais sejam:

Tabela 9 - Estações fluviométricas analisadas na aplicação do estudo de caso.

Código	Nome	Período Início	Fim	Latitude	Longitude	Altitude (m)	Área (km²)
64659000	Barbosa Ferraz	1975	1984	24º01'	51º57'	320	3294
64673000	Quinta do Sol	1975	1984	23º49'	52º11'	315	1534
64680000	Jussara	1977	1984	23º37'	52º28'	275	727
64682000	Japurá	1977	1984	23º26'	52º35'	285	807
64771500	Porto Guarani	1976	1990	24º51'	52º45'	360	4223
64815000	Fazenda Uberaba	1979	1984	24º07'	53º19'	310	2941
65190000	Balsa do Potinga	1942	1953	25º52'	50º49'	760	222
65295000	Sta. Cruz do Timbó	1974	1975	26º23'	50º52'	750	2614
65690000	Leonópolis	1958	1967	25º41'	51º12'	960	358
81100000	Passo do Assungui	1936	1945	25º01'	49º29'	390	1573
82111000	Mergulhão	1945	1966	25º19'	48º43'	9	344
82121000	Limoeiro	1930	1952	25º15'	48º46'	8	350
82170000	Morretes	1938	1982	25º29'	48º50'	3	208

Tabela 10 - Parâmetros estimados.

Código	Dados amostrais $\overline{Q}$ (l/s/km²)	$\overline{Q}_{max}$ (l/s/km²)	$CV[Q_{max}]$	Método proposto $\overline{Q}$ (l/s/km²)	$\overline{Q}_{max}$ (l/s/km²)	$CV[Q_{max}]$	Kaviski $\overline{Q}$ (l/s/km²)	$\overline{Q}_{max}$ (l/s/km²)	$CV[Q_{max}]$
64659000	24,27	422,83	0,4184	21,25	101,61	0,5109	21,29	310,16	0,4969
64673000	24,95	159,06	0,4837	21,19	102,12	0,5096	19,38	177,61	0,2904
64680000	19,81	89,68	0,2734	20,76	97,94	0,5040	18,19	133,97	0,6452
64682000	22,45	78,66	0,5824	20,55	92,69	0,4953	17,32	108,07	0,6815
64771500	27,24	353,31	0,3580	22,82	299,93	0,5288	22,06	197,60	0,3823
64815000	22,45	87,33	0,3436	21,62	102,83	0,4492	19,92	128,92	0,3861
65190000	26,80	214,98	0,3295	22,15	193,19	0,3630	21,34	139,26	0,4981
65295000	27,07	169,47	...	22,80	102,84	0,4687	23,93	310,16	0,4969
65690000	27,32	322,31	0,3271	24,00	252,20	0,3573	22,84	170,73	0,3217
81100000	...	75,29	0,5360	20,79	96,41	0,4767	13,79	153,23	0,4883
82111000	...	262,66	0,0880	25,65	265,76	0,0817	24,11	71,20	0,4883
82121000	...	229,93	0,0640	25,61	240,65	0,0766	20,96	123,62	0,2885
82170000	...	605,42	0,3430	17,34	96,93	0,4704	32,93	90,45	0,2848

$$\overline{d}_q(p) = \{\sum_{i=1}^{n}[(y_i - \hat{y}_i)/y_i]^2/n\}^{1/2} \quad (17)$$

$$\overline{d}_a(p) = [\sum_{i=1}^{n}(|y_i - \hat{y}_i|/y_i)]/n \quad (18)$$

Nas expressões (17) e (18), y_i corresponde ao valor do parâmetro p na estação i obtido através dos dados históricos observados em i; $\hat{y}_i$ corresponde ao valor do parâmetro p estimado pelo método de regionalização proposto neste trabalho ou em Kaviski (1992); e n é o número de locais analisados.

A aplicação do método de regionalização foi efetuada nos locais das estações fluviométricas apresentadas na Tabela 9. Todas estas estações drenam áreas menores que 5000 km². O conjunto de estações selecionadas apresenta distintos tamanhos de séries e qualidade.

Primeiramente, foi efetuada a análise discriminante sobre as estações da Tabela 9. O objetivo da análise discriminante é alocar uma dada bacia hidrográfica em um dos grupos definidos na análise de agrupamentos. A alocação é realizada com base nas características fisiográficas e/ou climatológicas. Na análise de agrupamentos são definidas as n características fisiográficas e/ou climatológicas utilizadas como variáveis classificadoras na análise discriminante. Existem diversas regras discriminantes que podem ser utilizadas. Neste trabalho foi utilizada a mesma regra discriminante utilizada em Kaviski (1992), ou seja, a regra discriminante de máxima verossimilhança para $\underline{x}$.

Para todas as estações da Tabela 9, foram estimados, pelo método proposto os parâmetros do "Grupo Max". Os parâmetros do "Grupo Med", exceto CV[Q], foram estimados para as 9 primeiras estações desta tabela. Como já explanado, não foi possível determinar as equações de regressão por metas difusas para CV[Q]. Todas as tentativas de estimar este parâmetro geraram resultados inconsistentes, ou seja, valores muito distorcidos em relação aos observados. A determinação deste parâmetro foi efetuada pelo método utilizado em Kaviski (1992), ou seja, por interpolação ponderada. Na Tabela 10 são apresentados os parâmetros estimados pelos dados amostrais, pelo método de regionalização proposto nesta pesquisa e pelo método apresentado em Kaviski (1992).

Para a maioria das estações não foi possível efetuar as comparações com dados amostrais em termos dos parâmetros CV[Q], MP_1[Q], MP_2[Q] e MP_3[Q], pois estes dados não estão disponibilizados. Desta maneira, as estimativas efetuadas para estes parâmetros não são aqui apresentadas. Apenas uma única estação possui as estimativas amostrais destes parâmetros.

Para verificar se haveria, efetivamente, alguma melhoria na aplicação do estudo de caso apenas pela utilização da classificação difusa, foi efetuada uma análise complementar para a estimativa de $\overline{Q}$, onde não foi considerada a regressão difusa e, sim, a interpolação ponderada. Esta última técnica foi utilizada em Kaviski (1992) e considera a probabilidade de que o local sem dados pertença a um determinado grupo. A Tabela 11 apresenta os valores de $\overline{d}_q$ e $\overline{d}_a$ para o métodos apresentados neste trabalho e para o método apresentado em Kaviski (1992). De maneira geral, considerando estes dois índices, o método proposto neste trabalho apresenta um grau de melhoria na regionalização dos parâmetros, até mesmo ao considerar-se a utilização da lógica difusa somente na classificação.

Tabela 11 - Valores comparativos de $\overline{d}_q$ e $\overline{d}_a$.

Métodos		$\overline{Q}$	$\overline{Q}_{max}$	CV[Q_{max}
Classificação difusa + regressão difusa	$\overline{d}_q$	0,127	0,372	0,334
	$\overline{d}_a$	0,118	0,278	0,252
Classificação difusa	$\overline{d}_q$	0,154	...	...
	$\overline{d}_a$	0,140	...	...
Kaviski	$\overline{d}_q$	0,168	0,400	1,290
	$\overline{d}_a$	0,160	0,355	0,794

CONCLUSÕES E RECOMENDAÇÕES

O objetivo geral deste trabalho consiste em analisar a aplicação da análise difusa em uma modelagem em recursos hídricos. Através do estudo de caso foi possível atingir este objetivo, ou seja, verificar que a aplicação de ferramentas da lógica difusa em uma técnica de modelagem de recursos hídricos, no caso, regionalização de parâmetros hidrológicos, pode fornecer resultados melhores que os obtidos com a utilização de métodos tradicionais.

O método proposto apresentou bastante simplicidade na sua conceituação matemática e desenvolvimento, confirmando uma das principais vantagens citadas da análise difusa. Esta característica foi verificada em ambas as técnicas de lógica difusa utilizadas nesta pesquisa. Outra vantagem comum aos sistemas difusos é a utilização da experiência do especialista no desenvolvimento do método. Esta característica foi também verificada em ambas as técnicas de lógica difusa implementadas.

Na classificação difusa o especialista pode intervir no parâmetro que controla o nível de incerteza r ("*fuzziness*") no processo de particionamento, tomando como base o intervalo de variações nos parâmetros a serem classificados. Pode-se atribuir um valor menor para r, caso ocorram na região estudada locais com características fisiográficas, meteorológicas e fluviais bastante diferenciadas entre si e/ou os parâmetros fluviais possuam forte dependência dos parâmetros fisiográficos e meteorológicos. Desta maneira, os agrupamentos formados possuem um maior grau de certeza, fato este já conhecido previamente.

No caso da regressão difusa, os intervalos de tolerância são, em geral, entre os parâmetros afetados pela inferência do especialista, os de maior sensibilidade. Porém, verificou-se no estudo de caso, que, para os parâmetros hidrológicos relacionados às séries de vazões médias diárias, as variações nos intervalos de tolerância não interferem na definição das equações.

A classificação difusa apresentou melhores resultados que a análise de agrupamentos pelo método convencional conhecido como "*K-means*", tanto no desenvolvimento do método proposto como na aplicação do estudo de caso. Pela comparação dos valores da relação de variâncias V_B/V_W, pode-se observar que há um aumento da relação de variâncias para os parâmetros classificados pelo método "*fuzzy c-means*", exceto para CV[Q_{max}]. Pelos valores de $\overline{d}_q$ e $\overline{d}_a$ apresentados na Tabela 11, pode-se também constatar que somente a utilização da classificação difusa apresenta uma melhora na regionalização dos parâmetros.

A regressão difusa forneceu ajustes perfeitos (grau de ajuste médio $\overline{h}$ = 1) para a maioria dos parâmetros hidrológicos analisados, quando utilizado o termo independente. Porém, estas equações não possuem bom comportamento na aplicação em locais que não foram utilizados na construção do modelo.

Com base nos resultados apresentados, pode-se concluir que a análise difusa pode ser utilizada na estimativa de parâmetros hidrológicos em regionalização. As principais vantagens são maior simplicidade, rapidez e bons resultados. A classificação difusa apresentou-se como um método bastante consistente e poderia substituir outras técnicas de análise de agrupamentos.

Como no caso do modelo tradicional de regressão por mínimos quadrados, a regressão difusa considera que os erros são independentes e possuem variâncias iguais. A regressão difusa carece de maiores investigações e pode ser utilizada na complementação de outros métodos. Sempre que possível deve-se tomar o cuidado de verificar as estimativas para locais com dados e que não foram utilizados na construção do modelo.

APÊNDICE

A regressão por metas difusas visa maximizar o nível de credibilidade, através da pré-definição da largura difusa. A obtenção das equações de regressão recai na programação objetiva com múltiplas metas difusas e pode ser simplificado pelas expressões abaixo (Chang et al., 1996), onde CX representa a variável prevista:

$$CX_i \lesssim f_1 \qquad (1)$$

$$CX_j \gtrsim f_2 \qquad (2)$$

As inequações (1) e (2) representam as restrições objetivas. Pode-se representá-las por funções de pertinência lineares dos tipos não-crescente e não-decrescente, onde CX é definido no eixo das abcissas. O intervalo de valores no eixo de CX_i, para o qual os valores de pertinência μ_i são decrescentes, denominado 'intervalo de tolerância', é igual a δ_i. Analogamente, o intervalo de valores no eixo de CX_j para o qual os valores de pertinência μ_j são crescentes é igual a δ_j. Os intervalos de tolerância são, entre os valores afetados pela inferência do especialista, os de maior sensibilidade (Chang et al., 1996).

As expressões das funções de pertinência não-crescente e não-decrescente são dadas, respectivamente, por:

$$\mu_i(CX_i) = \begin{cases} 1 & \text{se } CX_i \leq f_{1i}, \quad (3) \\ 1-\dfrac{(CX_i - f_{1i})}{\delta_i} & \text{se } f_{1i} < CX_i < f_{1i}+\delta_i, \quad (4) \\ 0 & \text{se } CX_i \geq f_{1i}+\delta_i. \quad (5) \end{cases}$$

$$\mu_j(CX_j) = \begin{cases} 1 & \text{se } CX_j \geq f_{2j}, \quad (6) \\ 1-\dfrac{(f_{2j} - CX_j)}{\delta_j} & \text{se } f_{2j} < CX_j < f_{2j}+\delta_j, \quad (7) \\ 0 & \text{se } CX_j \leq f_{2j}+\delta_j. \quad (8) \end{cases}$$

REFERÊNCIAS

AZEVEDO, Fernando M. de, BRASIL, Lourdes M. e OLIVEIRA, Roberto C. L. de (2000). *Redes Neurais com aplicações em controles e em sistemas especialistas*. Florianópolis: Bookstore.

BARDOSSY, Andras, BOGARDI, Istvan & DUCKSTEIN, Lucien (1990). Fuzzy regression in hydrology. *Water Resources Research*, Washington, 7(26):1497-1508, julho.

BOGARDI, Istvan, BARDOSSY, Andras & DUCKSTEIN, Lucien (1983). Regional management of na aquifer for mining under fuzzy environmental objectives. *Water Resources Research*, Washington, 6(19):1394-1402, dezembro.

CAMPANA, Néstor A. (1995). *Regionalização de hidrograma unitário usando regressão fuzzy*. Anais do XI Simpósio Brasileiro de Recursos Hídricos e II Simpósio de Hidráulica e Recursos Hídricos dos Países de Língua Oficial Portuguesa, Volume I – Hidrologia, Recife, 1995, pp. 33-38.

CHANG, Ni-Bin, CHEN, Y. L. & YANG, H. H. (1996). A fuzzy goal regression model for the construction cost estimation of municipal waste incinerators. *International Journal of Systems Science*, 5(27):433-445.

GALVÃO, Carlos de O. (1995). *Modelagem fuzzy da previsão climática no nordeste para uso em operação de reservatórios*. Anais do XI Simpósio Brasileiro de Recursos Hídricos e II Simpósio de Hidráulica e Recursos Hídricos dos Países de Língua Oficial Portuguesa, Volume II – Água Subterrânea e Semi-Árido, Recife, 1995, pp. 57-62.

GATES, Timothy K., HEYDER, Walter E., FONTANE, Darrell G. & SALAS, José D. (1991). Multicriterion strategic planning for improved irrigation delivery. I: Approach. *Journal of Irrigation and Drainage Engineering*, Reston, 6(117):897-913, nov./dez.

GIBERTONI, Rosana C. (2002). *Regionalização de parâmetros hidrológicos utilizando análise difusa*. Dissertação de Mestrado, UFPR, Programa de Pós-Graduação em Engenharia de Recursos Hidrícos e Ambiental, Curitiba: outubro.

HALL, M. J. & MINNS, A. W. (1999). The classification of hydrologically homogeneous regions. *Hydrological Sciences Journal*, Wallingford, 44(5):693-704, outubro.

KAVISKI, Eloy (1992). *Métodos de regionalização de eventos e parâmetros hidrológicos*. Dissertação de Mestrado, UFPR, Curso de Pós-Graduação em Engenharia Hidráulica, Curitiba: novembro.

MATHWORKS (1998). *Fuzzy Logic Toolbox – User's Guide Version 2*, (Manual do Usuário do MATLAB).

TANAKA, Hideo, UEJIMA, Satoru & ASAI, Kiyoji (1982). Linear regression analysis with fuzzy model. *IEEE Transactions on Systems, Man and Cybernetics*, 6(SMC-12):903-907, nov./dec..

YU, Pao-Shan, CHEN, Chia-Jung & CHEN, Shiann-Jong (2000). Application of gray and fuzzy methods for rainfall forecasting. *Journal of Hydrologic Engineering*, Reston, 4(5):339-345, outubro.

ZADEH, Lotfi A. (1965). Fuzzy sets. *Information and Control*, 8(3):338-353.

Regionalization of hydrologic parameters with fuzzy analysis

ABSTRACT

This paper addresses the application of the fuzzy sets theory in regionalization of hydrologic parameters in small and medium basins. Fuzzy analysis is appropriate for multi criteria modeling of impacts under uncertainty for water resources management and modeling. In hydrology the uncertainties are mostly related to physiographic characteristics of the basins and also to imprecision and scarcity of hydrologic data.

The following techniques of fuzzy logic were used: analysis of clustering of numerical data by the "c-means" algorithm and fuzzy goal regression. The investigation focused on regionalization of mean daily discharge (mean, coefficient of variation and probability weighted moments); and of maximum annual flow (mean and coefficient of variation).

The conclusions are based on performance indices of the results and comparison of the results of this dissertation with other regionalization methods. Through the case study it was possible to verify that the application of fuzzy tools as a water resources modeling technique, in this case, regionalization of hydrologic parameters, is justified by better results than those obtained with standard statistical methods.

Key words: fuzzy logic, regionalization, hydrologic parameters.