XII SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE
4 a 7 / novembro / 2014 ★ Natal ★ RN

MODELAGEM HIDROLÓGICA UTILIZANDO REGRESSÃO: CONCEITUAÇÃO MATEMÁTICA, LOGARITMIZAÇÃO E AVALIAÇÃO

Francisco F. N. Marcuzzo[1*]

Resumo – O conhecimento de regressões linearizadas e não linearizadas, visando a modelagem matemática de vazões de permanência de cursos d'água em bacias hidrográficas é de relevante importância para a gestão e o planejamento dos recursos hídricos a curto, médio e longo prazo. O objetivo deste trabalho foi estudar e discutir, para a hidrologia, a conceituação de regressão matemática, logaritmização (linearização) de modelos de regressão com uma ou duas variáveis explicativas, as possíveis diferenças entre modelos logaritmizados e "naturais", a aplicação e avaliação de modelos. Em hidrologia, os modelos matemáticos que utilizam a regressão relacionam o desempenho de uma variável dependente (como por exemplo, a vazão) com outra variável, que deve ser independente (como por exemplo, a área de drenagem). Os modelos de regressão para hidrologia sejam simples ou multivariados, simulam relacionamentos entre as variáveis explicadas (vazão) e explicativas (área de drenagem e precipitação). Essa interação deverá ser linear (equação da reta) ou não linear (equação potencial, polinomial, exponencial, etc.). Existem critérios qualitativos que podem recomendar o quanto útil ou aproximado dos dados reais coletados em campo é a curva produzida pela regressão executada. Destacam-se os seguintes critérios: Desvio (erro) dos valores estimados (*Se*); Coeficiente de determinação (r^2); Coeficiente de correlação (r).

Palavras-Chave – Logaritmo neperiano, regressão de potência, coeficiente de determinação.

Abstract – Knowledge of linearized and non-linearized regressions, aimed at mathematical modeling of flow permanence of streams in watersheds is of great importance for the management and planning of water resources in the short, medium and long term. The aim of this work was to study and discuss for hydrology, math conceptualization regression linearization of regression models with one or two explanatory variables, the possible differences between linearization and "natural" models, implementation and evaluation of models. In hydrology, mathematical models using the regression relate the performance of a dependent variable (e.g., flow) with other variable, which must be independent (for example, the drainage area). Regression models for hydrology are single or multivariate; simulate relationships between the explained variables (flow) and explanatory (drainage area and rainfall). This interaction should be linear (straight-line equation) or nonlinear (potential equation, polynomial, exponential, etc.). There are qualitative criteria that can recommend useful or approximate as the actual data collected in the field is the curve produced by the regression run. They include the following criteria: deviation (error) of the estimated values (*Se*); Coefficient of determination (*r2*); Correlation coefficient (*r*).

Keywords – Neperian logarithm, power regression, coefficient of determination.

[1*] CPRM/SGB (Companhia de Pesquisa de Recursos Minerais / Serviço Geológico do Brasil) – Rua Banco da Província, nº 105 - Santa Tereza - CEP 90840-030, Porto Alegre/RS. Tel.: (51) 3406-7324. francisco.marcuzzo@cprm.gov.br.

1. INTRODUÇÃO

São vários os recursos matemáticos disponíveis para o hidrologista no seu trabalho diário e no desenvolvimento da pesquisa hidrológica, contudo, a utilização de gráficos de dispersão de dados hidrológicos ocupa uma posição de importância dentre as inúmeras ferramentas disponíveis. Equacionar esta relação cartesiana entre dados hidrológicos e/ou dados físicos e/ou geomorfológicos, que os gráficos de dispersão destas variáveis fornecem é de suma importância para o hidrologista compreender o fenômeno presente e, se possível, estimar o mesmo fenômeno para áreas sem dados coletados em campo ou, até mesmo, prever fenômenos futuros.

Naghettini e Pinto (2007) citam que muitas vezes, a simples visualização do diagrama de dispersão sugere a existência de uma relação funcional entre as duas variáveis. A análise de regressão é uma técnica estatística cujo escopo é investigar e modelar a relação entre variáveis. Tucci (2002) cita que inicialmente não são conhecidas as variáveis independentes que melhor explicam a variável dependente. Portanto, é necessário procurar a melhor combinação das variáveis independentes que represente a distribuição dos valores da variável dependente. Marcuzzo *et al.* (2011) demonstra que testes com diferentes hipóteses matemáticas são necessários para se encontrar o melhor modelo para determinados tipos de estudos hidrológicos. Barbosa *et al.* (2005), em seu estudo de regionalização e vazão, concluíram que os melhores modelos ajustados foram do tipo linear e potencial. Considerado o pequeno número de variáveis dependentes (somente quatro estações fluviométricas), para se evitar a superparametrização, apenas uma variável independente foi considerada em cada ajuste. Foram avaliados os modelos de máximas, médias e mínimas de várias recorrências em termos das características físicas definidas pela área de drenagem, comprimento do talvegue, declividade do curso d'água e densidade de drenagem, e das características climáticas representadas pelas chuvas máxima anual, total do semestre mais chuvoso e total anual. Para converter as chuvas no ponto em chuvas espaciais, empregou-se o método de Thiessen. Mamun *et al.* (2010) agruparam curvas semelhantes de baixa frequência de vazões na região da Malásia Peninsular e obtiveram mapas e equações regionais por meio da técnica de regressão multivariada, em função da área de drenagem, precipitação média anual e evaporação média anual. Li *et al.* (2010) propuseram um novo método de regionalização, chamado modelo de índice. Os modelos especificam relações entre as variáveis hidrológicas utilizadas na previsão de vazões, melhorando a própria previsão das vazões. O modelo foi aplicado para simular curvas de permanência de vazões em 227 microbacias no sudeste da Austrália. Costa *et al.* (2012), concluíram que os modelos de regionalização desenvolvidos em seu trabalho são ferramentas promissoras para auxiliar na solução da escassez de dados de vazão na região amazônica, podendo dar suporte ao planejamento e à gestão dos recursos hídricos da região, bem como, de outras regiões que sofram com a falta de monitoramento hidrológico. Collischonn e Dornelles (2013) observam que a regionalização hidrológica também é um procedimento relativamente complexo, quando realizado de forma cuidadosa e detalhada. Algumas formas simples de regionalização, porém, podem ser aplicadas de forma mais imediata. A forma mais simples de regionalização hidrológica é, possivelmente, o estabelecimento de uma relação linear entre vazão e área de drenagem da bacia. Virães (2013), em um estudo da vazão de 95% de permanência da sub-bacia 50, utilizou técnicas de otimização matemática para minimizar os desvios das curvas de regressão determinadas, obtendo excelentes resultados de somatório de resíduos (erros) entre os valores observados e estimados, além de elevados valores de coeficiente de determinação. Pickbrenner e Marcuzzo (2014), em um estudo da vazão de 80, 85, 90 e 95% de permanência da sub-bacia 87, observaram inúmeros problemas para se agrupar as regiões homogêneas e se ajustar as curvas obtidas frente aos valores qualitativos de coeficiente de determinação, somatório de desvio, maior desvio observado e intervalo de confiança estatístico para a linha de regressão e a para os valores previstos com as equações de regressão.

O presente trabalho objetivou estudar e discutir, para a hidrologia, a conceituação de regressão matemática, logaritmização (linearização) de modelos de regressão com uma ou duas variáveis explicativas, as possíveis diferenças entre modelos logaritmizados e “naturais”, a aplicação e avaliação de modelos.

2. CONCEITUAÇÕES, EQUAÇÕES, LINEARIZAÇÃO, APLICAÇÃO E AVALIAÇÃO

2.1 Conceituações matemática básica de regressão para hidrologia

2.1.1 Gráfico de dispersão entre os valores explicados e os explicativos

O processo básico consiste, primeiramente, na construção de um gráfico de dispersão entre fenômenos geomorfológicos e/ou hidrológicos explicativos (área de drenagem e precipitação, por exemplo) com outros fenômenos hidrológicos que se deseja explicar (vazão, por exemplo). No eixo X (das abcissas, eixo horizontal) deve ser plotada a variável independente hidrológica (área de drenagem, por exemplo) associada à variável hidrológica que, ao modificar, assume valores que não dependem dos valores da outra grandeza hidrológica (vazão, por exemplo). No eixo Y (das ordenadas, eixo vertical) deve ser registrada a variável dependente (vazão, por exemplo) associada à grandeza hidrológica que, para variar, depende de como varia a outra grandeza hidrológica (área de drenagem, por exemplo).

2.1.2 Regressão matemática simples e múltipla

Em hidrologia, os modelos matemáticos que utilizam a regressão relacionam o desempenho de uma variável dependente (como por exemplo, a vazão) com outra variável, que deve ser independente (como por exemplo, a área de drenagem). Exemplificando matematicamente, quando a função *f* que relaciona duas variáveis é do tipo:

$$f\ (Xi) = Y = \ a + b\ Xi + \varepsilon i \tag{1}$$

$$f\ (Xi) = Y = \ a + b1\ Xi + b2\ X2i + \cdots + bN\ XNi + \varepsilon i \tag{2}$$

$$a = \ \bar{Y} - b\bar{X} \tag{3}$$

$$b = \ \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2} \tag{4}$$

em que, $f\ (Xi) = Y$ é o valor estimado para um específico valor Xi; b é a inclinação da reta (coeficiente angular), ou seja, o acréscimo ou decréscimo do valor de Yi em relação à Xi; a localiza o ponto de interseção da reta em relação ao sistema de coordenada retangulares, ou seja, é ponto no qual uma linha irá interceptar o eixo Y usando valores de X e Y existentes na regressão; εi representa uma perturbação aleatória na função, ou seja, nem todos os pontos tocam a reta, e essa diferença é o erro de aproximação; e N é o tamanho da amostra; $\bar{Y}$ é a média dos valores explicados; $\bar{X}$ é a média dos valores explicativos.

Observa-se na primeira equação acima (1) um modelo de regressão linear simples. Sabe-se que a variável X (vazão) é a variável independente da equação, enquanto Y (área de drenagem) é a variável dependente (ou seja, $Y = f\ (X)$), das variações que ocorrem em X. Observa-se na segunda equação acima (2) um modelo de regressão linear múltipla.

O modelo de regressão matemática é chamado de simples (equação 1), quando envolve uma relação "aleatória" entre duas variáveis, uma explicada (vazão) e outra explicativa (precipitação). O modelo de regressão é múltiplo (equação 2), ou multivariado, quando envolve uma relação aleatória dependente com mais de duas variáveis, como por exemplo, a vazão (dependente) com a área de drenagem e a precipitação (independente), ou seja, quando o desempenho de uma variável dependente é explicado por mais de uma variável independe.

Os modelos de regressão para hidrologia sejam simples ou multivariados, simulam relacionamentos entre as variáveis explicadas (vazão) e explicativas (área de drenagem e precipitação). Essa interação deverá ser linear (equação da reta) ou não linear (equação potencial, polinomial, exponencial, etc.).

Em estudos hidrológicos, que utilizam a modelagem matemática por meio de regressão, os modelos mais utilizados podem ser classificados simplificadamente em modelos: linear simples, não linear simples, linear multivariado e não linear multivariado.

Os modelos de regressão em hidrologia são normalmente utilizados em duas situações.

A primeira, quando hidrólogos interessados em simular as implicações sobre uma variável dependente, como vazão, em consequência de alterações engranzadas nos valores de uma variável independente, como comprimento do talvegue do rio. Exemplificando: como a vazão (*Y*) de um rio é alterada quando se aumenta ou diminui a precipitação (*X*) na área de drenagem a montante.

A segunda, quando hidrólogos executam previsões sobre o comportamento futuro de algum fenômeno hidrológico frente a um fenômeno físico e/ou também hidrológico, como a vazão frente à intensidade de precipitação. Para esta situação, extrapolam-se para o futuro as relações entre a causa (intensidade de precipitação) e o efeito (aumento da vazão), já analisadas no passado por meio de uma série histórica, entre as variáveis explicada (vazão) e explicativa (intensidade de precipitação).

Em quase sua totalidade, as relações entre duas grandezas hidrológicas ou uma grandeza geomorfológica e hidrológica não é linear. Com isso é necessário descobrir qual o tipo de equação melhor caracteriza a relação entre as variáveis explicadas e explicativas (MARCUZZO e MELATI, 2014). Para isso pode-se lançar mão da linearização por meio de logaritmos.

2.1.3 Construção da equação de regressão pelo método dos mínimos quadrados

O método dos mínimos quadrados é o mais utilizado para a construção de equações (*retas*, *curvas*) de regressão. Este método definirá uma reta que minimizará a soma das distâncias ao quadrado entre os pontos plotados explicativos (*X*), como por exemplo, área de drenagem, e explicados (*Y*), como por exemplo, vazão, e a reta, ou seja, os valores obtidos pela equação de regressão (*X’*,*Y’*). Pelo método dos mínimos quadrados calculam-se os parâmetros “*a*“ e “*b*” da reta que minimiza estas distâncias ou as diferenças (denominado também de desvios ou erros) entre *Y* e *Y’*. O método consiste nos seguintes passos:

Calcular a diferença entre o valor observado em campo (explicados, *Y*), como por exemplo, vazão, e o mesmo valor estimado *Y’* pela equação de regressão:

$$Erro = E = (Y - Y') \quad (5)$$

O objetivo do modelo de regressão é fornecer uma reta (curva) com r o somatório dos erros menor possível:

$$E1^2 + E2^2 + E3^2 + \ldots + En^2 = Mínimo \quad (6)$$

A hipótese de trabalho é a seguinte:

$$Erro\ Total = \boldsymbol{\Sigma}(Y - Y')^2 \quad (7)$$

Equação de regressão (da reta) que minimiza o erro (ou os desvios):

$$Y' = a + b\,X \quad (8)$$

Em seguida substitui a equação de regressão (8) na equação de hipótese de trabalho (7):

$$\Sigma(Y - a - b\,X)^2 \quad (9)$$

Objetivando-se que a soma dos quadrados dos desvios (erros dos valores explicados coletados em campo e os estimados pela equação de regressão) tenha um valor mínimo possível, devem-se aplicar os conceitos de cálculo diferencial com derivadas parciais. Como as incógnitas do problema são os coeficientes “*a*“ e “*b*” estrutura-se um sistema de duas equações. Portanto, aplicando os conceitos descritos acima, arma-se o sistema de equações normais que consentirá determinar os valores dos coeficientes “*a*“ e “*b*”, sendo *N* o tamanho da amostra:

$$\frac{\partial e}{\partial a} = -2\,\Sigma\,(Y - a - bX) \quad (10)$$

$$-2\,\Sigma\,Y + 2\,\Sigma\,a + 2\,\Sigma\,bX \quad (11)$$

$$\Sigma\,Y = \Sigma\,a + \Sigma\,bX \quad (12)$$

$$\Sigma Y = Na + b\,\Sigma\,X \quad (13)$$

$$a = \frac{\Sigma\,Y + b\,\Sigma\,X}{N} \quad (14)$$

$$\Sigma\,XY = a\Sigma\,X + b\Sigma X^2 \quad (15)$$

$$b = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2} \quad (16)$$

Os valores a e b acima correspondem aos parâmetros da equação de regressão que minimiza as diferenças entre os valores de *Y* (levantados) e os de *Y’* (estimados pela regressão). Portanto, o problema de “*fitting*” (ajustar) uma reta que melhor se adeque à nuvem de dados se reduz em calcular os parâmetros a e b da equação de regressão. Um detalhamento exemplificado pode ser obtido em Naghettini e Pinto (2007).

2.2 Transformações de equações naturais em equações linearizadas

Por heurística entende-se que a relação entre duas grandezas hidrológicas quase sempre não é linear, e é essencial descobrir de que tipo é e quais são os parâmetros que a caracterizam. Numa relação linear é muito simples o processo de se definir os parâmetros envolvidos da equação, logo, quando se observa pelo gráfico de dispersão que os valores envolvidos não formam uma reta, pode-se linearizá-lo utilizando uma mudança de variáveis, transformando em equações de retas mesmo regressões que formam curvas visivelmente complexas. Este método, de transformar uma equação de regressão que gera uma curva em uma equação que gera uma reta, denomina-se linearização ou logaritimização.

Sugere-se um conhecimento mediano com as representações gráficas das principais funções matemáticas utilizadas em regressão, já que se deve ter uma base teórica sobre que tipo de função matemática de regressão poderia gerar uma curva parecida a indicada pela sequência de pontos formados por variáveis explicativas e explicadas no gráfico de dispersão.

Observa-se que, para hidrologia em geral e, em específico a regionalização de vazões, a escala logarítmica é muito útil quando estamos tratando com funções do tipo exponencial (17) e, principalmente, potência (18), que é o tipo de equação mais comum em regressões para se estimar vazão em rios. Estas funções sempre podem ser linearizadas com o uso de escalas logarítmicas.

$$Y' = a\ e^{b\ X} \tag{17}$$

$$Y' = a\ X^{b} \tag{18}$$

em que, Y' é o valor explicado obtido pela equação (como por exemplo, vazão); a é i coeficiente determinado pela equação da linha de tendência, sendo uma constante, que representa o declive (coeficiente angular de inclinação) da reta; X é o valor da variável explicativa (como por exemplo, área de drenagem); b é o coeficiente exponencial do valor de interseção da equação no gráfico produzido entre a variável explicativa e a variável explicada.

2.2.1 Logaritmização da função exponencial

Em hidrologia, um tipo de relação entre duas grandezas física muito comum e bem simples é a exponencial (equação 17).

A equação exponencial pode ser linearizada através de uma mudança de variáveis ou então plotando um gráfico em papel com eixos milimétricos, colocando os valores explicados Y' no eixo das ordenadas e os valores $e^{b\ X}$ no eixo da abscissa e não as medidas explicativas X. Outra possibilidade é utilizar um papel onde um dos eixos tem escala logarítmica e o outro linear.

Os passos da transformação da função exponencial (equação 17) na função linear (segunda equação do Quadro 1) logaritmizada são:

1º. Transformar os valores Y' em ln Y'.
2º. Calcular os coeficientes da reta de regressão que são o coeficiente angular de inclinação da reta a e o coeficiente de interseção da equação no gráfico b, além do coeficiente de determinação r^2.
3º. Em seguida determinam-se os coeficientes a e b da seguinte maneira:
  I. A interseção da reta da transformação exponencial é o logaritmo neperiano de a, portanto para calcular o coeficiente a deve-se fazer $a = e^b$ $(ln\ a = b)$.
  II. b = valor de interseção da equação no gráfico.

2.2.2 Logaritmização da função logarítmica

Os passos da transformação da função logarítmica (terceira equação do Quadro 1) na função linear são:

1º. Transformar os valores observados X em ln X.
2º. Calcular os coeficientes da reta de regressão que são o coeficiente angular de inclinação da reta a e o coeficiente de interseção da equação no gráfico n, além do coeficiente de determinação r^2.
3º. Em seguida determinam-se os coeficientes a e b da seguinte maneira:
  I. a = coeficiente angular de inclinação da curva no gráfico.
  II. b = valor de interseção da equação no gráfico.

2.2.3 Logaritmização da função potencial (normalmente a mais utilizada em regionalização de vazões mínimas)

Os passos da transformação da função potência (quarta equação do Quadro 1) na função linear logaritmizada são:

1º. Transformar os valores Y' em ln Y' e X em ln X.
2º. Calcular os coeficientes da reta de regressão que são o coeficiente angular de inclinação da reta a e o coeficiente de interseção da equação no gráfico n, além do coeficiente de determinação r^2.
3º. Em seguida obtêm-se os coeficientes a e b da seguinte maneira:

I. A interseção da reta da transformação potencial é o logaritmo neperiano de a, portanto para calcular o coeficiente a deve-se fazer $a = e^b$ $(ln\ a = b)$.
II. b = valor de interseção da equação no gráfico.

2.2.4 Quadro resumo de linearização de equações

O Quadro 1 mostra um resumo com os diferentes tipos de equações mais utilizadas em estudos hidrológicos e suas respectivas transformações lineares por logaritmos.

Quadro 1. Resumo dos tipos de equações e suas respectivas transformações lineares por logaritmos.

Tipo de Equação		Equação	Linearização (logaritmização)	Variáveis Observadas	
				X	Y
1	Linear	$Y' = a + bX$	$Y' = a + bX$	X	Y
2	Exponencial	$Y' = a\ e^{bX}$	$ln\ Y' = ln\ a + bX$	X	$ln\ Y$
3	Logarítmica	$Y' = a + b\ ln\ X$	$Y' = a + b\ ln\ X$	$ln\ X$	Y
4	Potencial	$Y' = a\ X^b$	$ln\ Y' = ln\ a + b\ ln\ X$	$ln\ X$	$ln\ Y$

2.3 Utilização de programas de planilha de cálculo nas transformações de equações naturais em equações linearizadas

Na utilização de programas com planilhas de cálculo (como por exemplo, o CALC® do Apache Open Office® ou o EXCEL® da Microsoft®) deve-se estabelecer o gráfico de dispersão dos valores observados explicativos (X) e explicados (Y). Para mudar a escala de cada eixo clique com o botão direito do cursor sobre o eixo X, por exemplo, e depois em "formatar eixo". Nas opções que aparecem, basta selecionar o quadro "escala logarítmica" e definir a base desejada (a mais convencional para logaritmos é a base 10, para o caso de uma equação exponencial (como a segunda do Quadro 1), emprega-se a base 2,718.

Os programas com planilhas de cálculo dispõem do comando linha de tendência dentro do ambiente de gráficos. A linha das equações de regressão (tendência) pode ser ajustada de forma automática nos gráficos de barras horizontais, colunas, de linhas ou de dispersão entre as variáveis explicativas e explicadas.

Passos para a construção das curvas das equações de regressão (tendência):

1º. Construir um gráfico de dispersão da série de dados X (explicativos) e Y (explicados). Não devem ser incluídos título nem legenda, ajustando o gráfico para ocupar todo o espaço disponível e facilitar a construção da linha de regressão (tendência) desejada.
2º. Clica-se duas vezes dentro do gráfico obtido para se ativar os comandos do ambiente gráfico.
3º. Em seguida seleciona-se a trajetória dos pontos do gráfico, clicando uma vez num dos pontos do gráfico.
4º. Ir para o menu Inserir e escolhe-se Linha de Tendência recebendo a caixa de diálogo.
   a. A caixa de diálogo contém duas partes: Tipo e Opções.
      i. Em Tipo: seleciona-se primeiro o tipo de curva desejado
      ii. Em Opções: escolhe-se dentre as seguintes alternativas:
         1. Linha de Tendência: escolhe-se um nome para ser incluído no gráfico.
            a. Escolhe-se Automática, o programa dá um nome à curva de ajuste baseado no tipo de curva selecionada e a série associada com ela.
            b. Pode-se personalizar colocando um nome qualquer com até 256 caracteres.

2. Opção de Previsão. Depois de construir a linha de regressão (tendência) no intervalo dos pontos desenhados no gráfico, o comando Previsão admite construir mais pontos antes e depois desse intervalo de dados disponibilizados, podendo ser aplicada apenas paras as curvas de ajustes de regressão.
   a. Opção Prospectiva. Pode-se informar o número de períodos, ou unidade para o gráfico de dispersão dos dados X (explicativos) e Y (explicados) que o comando linha de tendência projetará para o futuro.
   b. Opção Retrospectiva: Pode-se informar o número de períodos, ou unidade para o gráfico de dispersão dos dados X (explicativos) e Y (explicados) que o comando linha de tendência projetará para o passado.
      2. Definição do ponto de Interseção (coeficiente b das equações demonstradas anteriormente). Caso seja de interesse para o estudo a ser executado, visando o ajuste refinado da curva, pode-se definir o ponto onde a curva ajustada deve cortar o eixo das ordenadas Y (explicados).
         3. Marque Exibir a equação de regressão no gráfico de dispersão com a curva.
         4. Marque Exibir valor do coeficiente de determinação (r^2).

2.4 Avaliação da utilidade das equações de regressão

A equação de regressão que se obtém, seja ela linearizada ou não linearizada, utilizando o método dos mínimos quadrados é apenas uma aproximação dos valores reais. A reta produzida é um modo útil para indicar a tendência dos dados explicados por meio dos dados explicativos. Contudo deve-se avaliar a equação produzida de modo a se ter um ou mais parâmetros qualitativos da curva de regressão para se aprovar ou não, ou, até mesmo, escolher a que representa melhor os dados explicados por meio de comparação com outros tipos de regressão, conjuntos de dados, etc.

Existem critérios qualitativos que podem recomendar o quanto útil ou aproximado dos dados reais coletados em campo é a curva produzida pela regressão executada. Destacam-se os seguintes critérios:

a. Desvio (erro) dos valores estimados (*Se*);
b. Coeficiente de determinação (r^2).
c. Coeficiente de correlação (r).

2.4.1 Desvio (erro) dos valores estimados (*Se*)

O desvio padrão (*Se*) dos valores estimados, ou erro padrão, mede o desvio médio entre os valores reais de Y e os valores estimados Y'. Este dado qualitativo informa de modo aproximado a extensão do desvio (erro) entre os valores obtidos das estimativas explicadas e os valores de Y captados em campo. O desvio (erro) é medido na unidade da ordenada Y. O objetivo é perseguir o menor valor possível de S_e.

Também se pode interpretar o S_e como um desvio padrão dos resíduos, pois assumindo que estes resíduos (erros) são normalmente distribuídos, pode-se dizer então que uma percentagem (68%**)** dos pontos (plotados) encontra-se dentro dos “um” de desvio possível (-1 ≤ S_e ≥ 1); e que outra percentagem (95%) dos pontos encontram-se dentro de (dois) desvios padrão (-2 ≤ S_e ≥ 2).

Sendo os desvios normalmente distribuídos a fórmula de S_e é obtida da definição da variância da amostra S_e^2, com *n-2* graus de liberdade:

$$Var(e_i) = \sigma_e^2 = E(e_i^2) - E^2(e_i) = E(e_i^2) \qquad (19)$$

$$\sigma\hat{}_e^2 = s_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} \qquad (20)$$

$$\sigma\hat{}_e = s_e = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} \qquad (21)$$

No ajuste da curva produzida pela equação de regressão, se obtém uma melhor performance da explicação do conjunto de dados recolhidos. Se os dados estivem todos contidos numa reta tem-se uma reta de regressão coincidente com os dados coletados. Nesse caso a somatória dos desvios ao quadrado será zero e, o ajuste da reta será exato. Com isso a equação da curva de regressão explica perfeitamente a relação entre os dados *X* (explicativos) e *Y* (explicados).

O desvio padrão (erro padrão) existirá sempre que a capacidade de explicação da curva produzida pela regressão não for exata. O valor do erro constitui então que a existência de outros fatores que intervêm no comportamento dos dados explicados e dos valores explicativos.

2.4.2 Coeficiente de determinação (r^2).

Naghettini e Pinto (2007) citam que após a estimativa dos coeficientes da reta de regressão, é necessário verificar se os dados amostrais são descritos pelo modelo das equações, além disso, determinar a parcela da variabilidade amostral que foi, de fato, explicada pela reta de regressão. O coeficiente de determinação é calculado da seguinte maneira:

$$y_i = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}_i) + \bar{y} \qquad (22)$$

A partir da equação (22), é possível demonstrar que:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \qquad (23)$$

O primeiro membro da equação (23) pode ser interpretado como proporcional à variância total de *Y*, enquanto o segundo membro reflete a soma de termos proporcionais às suas variâncias residuais e explicada pelo modelo de regressão. Em outros termos:

$$SQT = SQ\ Res + SQ\ Reg \qquad (24)$$

em que, *SQT* é a soma quadrática total; *SQ Res* é a soma dos quadrados dos resíduos e *SQ Reg* é a soma dos quadrados devidos à regressão.

O coeficiente de determinação é dado pela relação entre a soma dos quadrados devidos a regressão (*SQ Reg*) e a soma total dos quadrados (*SQT*), ou seja:

$$r^2 = \frac{Variância\ Explicada}{Variância\ Total} = \frac{SQ\ Re\ g}{SQT} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (25)$$

em que, r^2 é o coeficiente de determinação ($0 \leq r^2 \leq 1$), y_i é o valor observado da variável dependente, $\hat{y}_i$ é o valor estimado da variável dependente e $\bar{y}$ é a média da variável dependente.

Sabe-se que o coeficiente de determinação é sempre positivo e deve ser interpretado como a proporção da variância total da variável dependente Y que é explicada pelo modelo de regressão e que também pode ser estimado por:

$$r^2 = b^2 \frac{s_X^2}{s_y^2} \qquad (26)$$

em que, s_X^2 é a variância amostral de *X*; s_y^2 é a variância amostral de *Y* e b é o coeficiente angular da reta de regressão calculado pela equação 4.

2.4.3 Coeficiente de correlação (*r*).

O coeficiente de correlação é igual à raiz quadrada do coeficiente de determinação (equação 27), que por sua vez é igual ao quadrado do coeficiente de correlação. Assim a partir do valor do coeficiente de determinação pode-se obter o valor do coeficiente de correlação.

O coeficiente de determinação (equação 25) é sempre positivo, variando de 0 a 1, enquanto que o coeficiente de correlação (*r*) pode admitir valores negativos e positivos, variando de –1 a 1, ou seja, $-1 \geq r \leq +1$. Valores de r igual ou próximos de 1 ou –1 indica que determina uma intensa relação entre as variáveis: no primeiro caso a relação é direta, enquanto que no segundo a relação é inversa. Valores iguais ou próximos de zero, significa que existe nenhuma ou pouca relação entre as variáveis explicativas e explicadas.

Sabe-se que o coeficiente de determinação indica o quanto a curva de regressão explica o ajuste da curva, enquanto que o coeficiente de correlação deve ser usado como uma medida de força da relação entre as variáveis. Entende-se que o coeficiente de correlação tem um valor único para o conjunto de dados testados e que padroniza dentro dos seus limites as variações da covariância.

O coeficiente de correlação amostral (*r*) está relacionado ao coeficiente de determinação, r^2, através da seguinte equação:

$$r = \pm\sqrt{r^2} \qquad (27)$$

em que, o sinal do r é o mesmo do sinal encontrado no coeficiente b.

Pode-se também determinar o coeficiente de correlação pela seguinte relação estatística:

$$r = \frac{Covariancia\ (X,Y)}{\sigma X\ \sigma Y} \tag{28}$$

em que, σ é o desvio padrão da variável estudada.

3. CONCLUSÃO

No ajuste do modelo da equação de regressão, se obtém uma melhor performance da explicação do conjunto de dados hidrológicos recolhidos. A equação de regressão que se obtém, seja ela linearizada ou não linearizada, utilizando o método dos mínimos quadrados é apenas uma aproximação dos valores reais. Deve-se avaliar a equação produzida com um ou mais parâmetros qualitativos da curva de regressão para se aprovar ou não, ou, escolher a que representa melhor os dados explicados por meio de comparação com outros tipos de regressão, conjuntos de dados, etc.

AGRADECIMENTO

O autor agradece a CPRM/SGB (Companhia de Pesquisa Recursos Minerais / Serviço Geológico do Brasil - Empresa Pública do Ministério de Minas e Energia) pelo fomento que possibilitou o desenvolvimento deste estudo.

REFERÊNCIAS

BARBOSA, E. da S.; BARBOSA JUNIOR, A. R.; DA SILVA, G. Q.; CAMPOS, E. N. B.; RODRIGUES, B. de CARVALHO. Geração de modelos de regionalização de vazões máximas, médias de longo período e mínimas de sete dias para a bacia do rio do Carmo, Minas Gerais. **Engenharia Sanitária e Ambiental.** Rio de Janeiro, v. 10, n. 1, Mar. 2005. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-41522005000100008&lng=en&nrm=iso>. Acesso em: 11 Jun. 2014.

BRASIL. Agência Nacional de Águas. Hidroweb: Sistema de informações hidrológicas. Disponível em: <http://www.hidroweb.ana.gov.br>. Acesso em: 29 mai. 2013.

COLLISCHONN, W.; DORNELLES, F. **Hidrologia para engenharia e ciências ambientais**. Porto Alegre: Associação Brasileira de Recursos Hídricos (ABRH), 2013. 336 p.

COSTA, A. S. CARIELLO, B. L.; BLANCO, C. J. C.; PESSOA, F. C. L. Regionalização de curvas de permanência de vazão de regiões hidrográficas do estado do Pará. **Revista Brasileira de Meteorologia**, São Paul, v. 27, n. 4, dez. 2012. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-77862012000400005&lng=pt&nrm=iso>. Acesso em 11 jun. 2014.

KÖPPEN, W. Das geographisca System der Klimate. In: KÖPPEN, W.; GEIGER, G. Handbuch der Klimatologie. **C. Gebr, Borntraeger**, Berlin, 1936. p. 1–44. Disponível em: <https://www.climond.org/Public/Data/Publications/Koeppen_1936_GeogSysKlim.pdf>.

LI, M.; SHAO, Q.; ZHANG, L.; CHIEW, F. H. S. A new regionalization approach and its application to predict flow duration curve in ungauged basins. **Journal of Hydrology**, v. 389, n. 1-2, p. 137-145, 2010. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0022169410003240>. Acesso em 11 jun. 2014.

MAMUN, A.A.; HASHIM, A.; DAOUD, J. I. Regionalization of low flow frequency curves for the Peninsular Malaysia. **Journal of Hydrology**, v. 381, n.1-2, p. 174-180, 2010. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0022169409007598>. Acesso em 11 jun. 2014.

MARCUZZO, F. F. N.; ANDRADE, L. R.; MELO, D. C. R. (2011). Métodos de Interpolação Matemática no Mapeamento de Chuvas do Estado do Mato Grosso. **Revista Brasileira de Geografia Física**, v. 4, n. 4, p. 793-804. Disponível em: http://www.ufpe.br/rbgfe/index.php/revista/article/view/197/204. Acesso em 20 de agosto de 2012.

MARCUZZO, F. F. N.; MELATI, M. D. Nível de significância no intervalo de confiança de linha de regressão e valor previsto em regressão da Q50 com duas variáveis explicativas. In: SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE, 12. 2014 , Natal. **Anais**... Natal: ABRH, 2014. p. 1-10. CD-ROM.

NAGHETTINI, M.; PINTO, É. J. de A. Correlação e regressão. In: NAGHETTINI, M.; PINTO, É. J. de A. **Hidrologia estatística.** Belo Horizonte: CPRM, 2007b. cap. 9, p. 355-400. Disponível em: <http://www.cprm.gov.br/publique/media/livro_hidro_estatistica.zip>. Acesso em: 3 jan. 2014.

PICKBRENNER, K.; MARCUZZO, F. F. N. **Regionalização de Vazões nas Bacias Hidrográficas Brasileiras:** estudo da vazão de 80, 85, 90 e 95% de permanência da sub-bacia 87 – Bacias da Lagoa dos Patos, do Lago Guaíba, do Litoral Médio e dos rios Camaquã, Caí, Gravataí, Jacuí, Sinos e Tramandaí. Porto Alegre: CPRM, 2014. 1 DVD. Projeto Disponibilidade Hídrica do Brasil - Estudos de Regionalização nas Bacias Hidrográficas Brasileiras. Levan. da Geodiversidade.

PINTO, E. J. de A.; AZAMBUJA, A. M. S. de; FARIAS, J. A. M.; SALGUEIRO, J. P.de B.; PICKBRENNER, K. (Coords.). (2011). **Atlas pluviométrico do Brasil: isoietas mensais, isoietas trimestrais, isoietas anuais, meses mais secos, meses mais chuvosos, trimestres mais secos, trimestres mais chuvosos**. Brasília: CPRM, 2011. 1 DVD. Escala 1.5:000.000. Equipe Executora: Andressa M. S. de Azambuja; Margarida R. da Costa; Carlos Eduardo de O. Dantas; José Alexandre M. Farias; Érica C. Machado; Francisco Fernando Noronha Marcuzzo; Vanesca S. Medeiros; Denise C. de Rezende Melo; Jean R. da S. do Nascimento; Paulo de Tarso R. Rodrigues; André Luis M. R. dos Santos; Adriana B. Weschenfelder; Sistema de Inf. Geográfica-SIG - versão 2.0 - atualizada em novembro/2011; Prog. Geol. do Brasil; Lev. da Geodiversidade. Disponível em: <http://www.cprm.gov.br/publique/media/Isoietas_Totais_Anuais_1977_2006.pdf>. Acesso em 11 jun. 2014.

SIMON, F. W.; PICKBRENNER, K.; MARCUZZO, F. F. N. Estudo do regime pluvial e fluvial em bacia hidrográfica com precipitação homogênea. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 20. 2013, Bento Gonçalves. **Anais...** Bento Gonçalves: ABRH, 2013. p. 1-8. CD-ROM. Disponível em: <http://www.abrh.org.br/SGCv3/UserFiles/Sumarios/22de4a642c2c18259e4809409096e0ff_6f2356d4ea7d3fcaba0d55bad04ebea4.pdf>. Acesso em: 30 dez. 2013.

TSCHIEDEL, A. da F.; PICKBRENNER, K.; MARCUZZO, F. F. N. Analise hidromorfológica da sub-bacia 87. In: SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE, 11. 2012 , João Pessoa. **Anais...** João Pessoa: ABRH, 2012. p. 1-20. CD-ROM. Disponível em: <http://www.cprm.gov.br/publique/media/Evento_Analise_Marcuzzo.pdf>. Acesso em: 3 abr 2013.

TUCCI, C. E. M. **Regionalização de vazões.** Porto Alegre: UFRGS, 2002. 256 p.

VIRÃES, M. V. **Regionalização de Vazões nas Bacias Hidrográficas Brasileiras:** estudo da vazão de 95% de permanência da sub-bacia 50 – Bacias dos rios Itapicuru, Vaza Barris, Real, Inhambupe, Pojuca, Sergipe, Japaratuba, Subaúma e Jacuípe. Recife: CPRM, 2013. 1 DVD. Projeto Disponibilidade Hídrica do Brasil - Estudos de Regionalização nas Bacias Hidrográficas Brasileiras. Levantamento da Geodiversidade. Disponível em: <http://www.cprm.gov.br/rehi/regionalizacao/sub_bacia_50/relatorio_sub_bacia50.pdf>. Acesso em 2 fev. 2014.