

ESTIMAÇÃO DA VAZÃO DO RIO CORUMBATAÍ NA CIDADE DE PIRACICABA.

DANIEL CHRISTOFOLETTI

FERNANDO FREIRE VASCONCELOS

Resumo

Conhecer a vazão do curso d'água de onde se faz a captação para o abastecimento de uma cidade é de suma importância pois fornece dados para o estabelecimento de políticas públicas e antecipação de medidas quando necessário. No caso da cidade de Piracicaba – SP, 90 % da captação de água para o consumo na cidade é feita no Rio Corumbataí, que apresenta uma expressiva queda na vazão durante os meses de junho a outubro. Assim, a motivação para este trabalho foi a criação de ferramentas que possibilitem a previsibilidade da vazão do curso d'água nos meses críticos do ano e a probabilidade de ocorrência de seca ou não. A base de dados utilizada foi de out.1972 a abr.2021. O modelo para estimação de vazão foi elaborado para cada um dos meses através da técnica de regressão múltipla e o modelo para previsão de seca foi elaborado através da técnica conhecida como regressão logística. Os resultados obtidos através de modelos multinível neste trabalho mostraram que as estimações de vazão mensal para os meses em estudo foram adequadas e boas, se aproximando de outros modelos, como redes neurais, garantindo a viabilidade da utilização de métodos quantitativos. Para os modelos de regressão logística binária, os resultados também foram satisfatórios, obtendo previsões com boa acurácia que podem ser utilizados no planejamento governamental do município.

Palavras Chave

Modelo multinível, Regressão , Seca

ESTIMAÇÃO DA VAZÃO DO RIO CORUMBATAÍ NA CIDADE DE PIRACICABA

Introdução

A disponibilidade de recursos hídricos é uma questão de importância global e seu conhecimento é fundamental para o desenvolvimento da sociedade (Uliana et.al, 2015). Por esse motivo, a sua falta ou escassez pode ser um grande empecilho para o desenvolvimento de uma região, fazendo com que haja a necessidade do seu monitoramento (Fraga et.al, 2020).

O rio Corumbataí assume grande importância regional pelo seu potencial hídrico, capaz de abastecer municípios que estão, em sua totalidade, dentro da bacia. Apresenta ainda condições de abastecer municípios que possuem problemas quanto à qualidade de água para captação, como, por exemplo, o município de Piracicaba, que capta cerca de 90% da água para abastecimento de sua área urbana do rio Corumbataí (Moei et.al, 2016).

Contudo, vários fatores afetam a disponibilidade hídrica do país, como a significativa variação climática anual e sazonal do território brasileiro, sendo essa considerada uma das maiores condicionantes de disponibilidade hídrica (Tucci, 2001), estando esses dados ainda mais sensíveis devido as mudanças climáticas, como aconteceu na crise hídrica de 2014 e 2015 que atingiu o Estado de São Paulo. Considerada a maior seca ocorrida na região Sudeste dos últimos 70 anos, esse período de estiagem prolongada gerou impactos consideráveis como o transtorno no abastecimento público, falhas no fornecimento de energia elétrica e rebaixamento do nível d'água nos rios e aquíferos (Carvalho, 2019).

Dentre as variáveis utilizadas para monitorar a disponibilidade hídrica de uma determinada bacia hidrográfica ou região, destacam-se a precipitação pluviométrica e a vazão dos rios. A avaliação de ambas as variáveis em conjunto é de extrema importância, uma vez que mudanças bruscas nos regimes de precipitação implicam diretamente em mudanças nas vazões dos rios. Ferreira (2007), afirma que as vazões respondem razoavelmente à variabilidade temporal e espacial da precipitação, ou seja, grandes anomalias de precipitação quase sempre resultam em anomalias de descarga de rios.

Diante disso, o problema de pesquisa que motiva a elaboração desse projeto de TCC foi elaborar um modelo para estimar a vazão em m³/s do Rio Corumbataí em diferentes períodos do ano, com ênfase nos meses de maior estiagem e menor vazão do rio (junho a outubro), em função da precipitação pluviométrica nos meses antecedentes e no mês corrente, bem como a probabilidade de enfrentar períodos de seca, e assim construir uma ferramenta para fornecer informações e auxiliar na tomada de decisões para gestores públicos, empresários e produtores agrícolas que fazem uso da água do manancial.

Por tratar de um trabalho aplicado, optou-se por passar diretamente para descrição da metodologia da pesquisa.

Metodologia

Para elaboração do presente estudo, foram utilizadas series históricas de vazão e precipitação correspondentes ao período de out.1972 a abr.2021. Para vazão média mensal do rio Corumbataí, recorreu-se aos dados captados no posto de aferição de Piracicaba – SP (Prefixo 4D-021, Latitude 22°34'53" S e Longitude "47°41'01" W). Em relação aos dados de pluviosidade, valeu-se dos dados obtidos nas estações meteorológicas, todos pertencentes à rede Hidrológica Básica do Estado de São Paulo (Tabela 1).

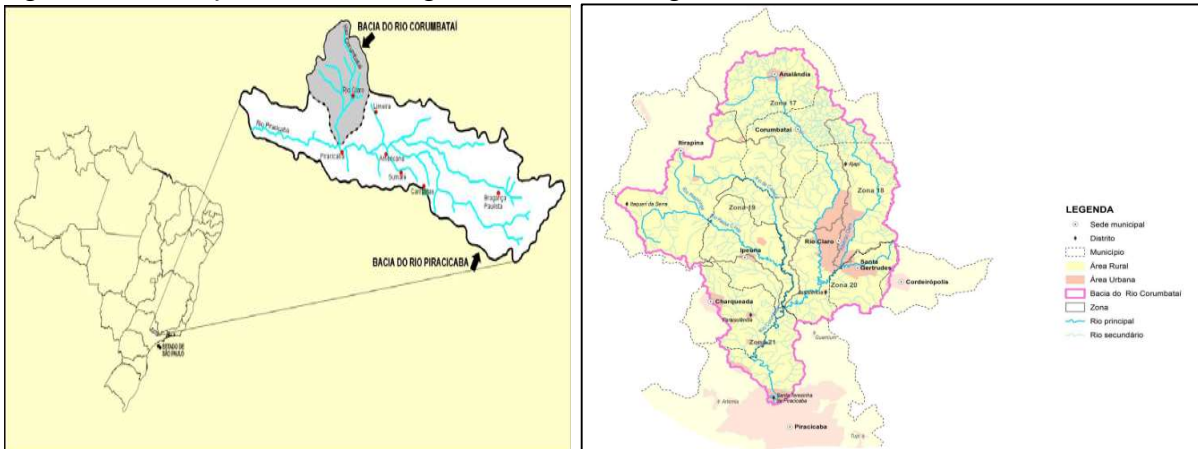
Tabela 1 - Prefixo, latitude e longitude das estações meteorológicas

Cidade	Prefixo	Latitude	Longitude
Analândia	D4-035	22°07'47" S	47°40'03" W)
Charqueada	D4-109	22°35'01" S	47°41'47" W
Corumbataí	D4-043	22°12'58" S	47°37'12" W
Itirapina	D4-036	22°18'01" S	47°44'38" W
Ipeuna	D4-074	22°26'14" S	47°43'23" W
Piracicaba	D4-104	22°43'03" S	22°43'03" S
Rio Claro	D4-012	22°24'20" S	47°33'21" W
Santa Gertrudes	D4-059	22°28'42" S	47°30'51" W

Fonte: Daee, 2022

A Bacia do Rio Corumbataí, uma das cinco sub-bacias integrantes das Bacias Hidrográficas PCJ, situa-se na porção centro-leste do Estado de São Paulo, localiza-se entre os paralelos 22° 04' 46" e 22° 41' 28" S e entre os meridianos 47° 26' 23" e 47° 56' 15" N. De acordo com Koffler (1993), sua altitude varia de 470m, na foz do Rio Corumbataí, a 1058m, na Serra do Cuscuzeiro, em Analândia (SP) e possui uma área territorial de 1.719,46 km², aproximadamente 171 mil hectares onde se encontram os municípios de Analândia, Charqueada, Corumbataí, Ipeúna, Itirapina, Rio Claro, Santa Gertrudes e Piracicaba (Santos e Garcia, 2013), como observado na figura 1.

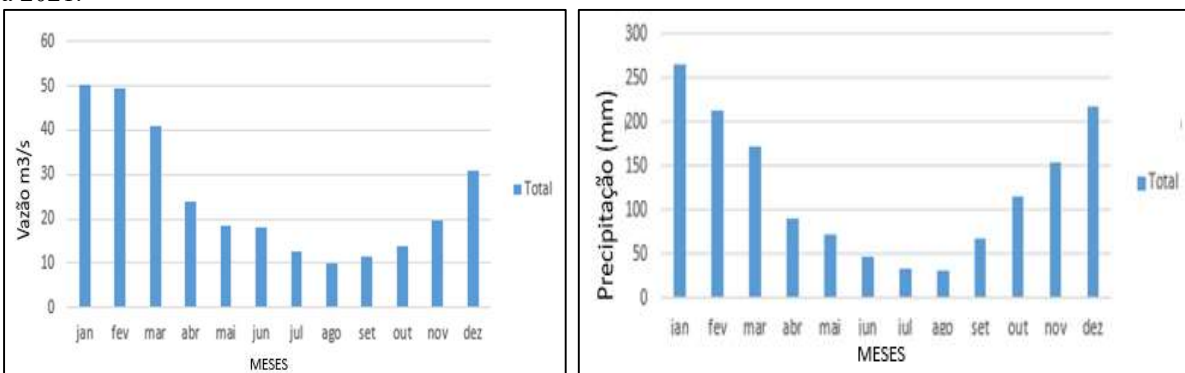
Figura 1 - Localização e área de abrangência da Bacia Hidrográfica do Rio Corumbataí/SP.



Fonte: Santos e Garcia (2013)

Os dados foram obtidos por meio de consulta do Banco de Dados Hidrológicos através do acesso ao portal do DAEE(departamento de água e energia elétrica). As figuras 2 representam a vazão média do curso d'água e a precipitação média das cidades que fazem parte da bacia.

Figura 2 - Vazão média (m³/s) e prescripção do rio Corumbataí no posto de medição de Piracicaba, de 1972 a 2021.



Fonte: Dae, 2021.

Da posse dos dados, estes foram compilados em uma tabela para serem analisados através do software R 4.1.2. Para predição dos valores, foi utilizado o modelo supervisionados de análise de dados, de regressão linear múltipla ao nível de 5% de significancia.

O termo 'regressão' foi proposto pela primeira vez por Sir Francis Galton em 1885 Francis Galton, em 1885 onde demonstrou que a altura dos filhos não tende a reflectir a altura dos pais, mas tende sim a regredir para a média da população. Atualmente, o termo "Análise de Regressão" define um conjunto vasto de técnicas estatísticas usadas para modelar relações entre variáveis e predizer o valor de uma ou mais variáveis

dependentes (ou de resposta) a partir de um conjunto de variáveis independentes (ou preditoras) (Maroco, 2003).

Temos uma regressão linear múltipla quando admitimos que o valor da variável dependente (Y) é função linear de duas ou mais variáveis explanatórias. O modelo estatístico de uma regressão linear múltipla com k variáveis explanatórias é:

$$Y_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_k X_{kj} + e_j \dots = 1, \dots, n$$

ou

$$Y_j = \alpha + \sum_{i=1}^k \beta_i X_{ij} + u_j$$

Onde: α : é o valor esperado de Y quando todas as variáveis independentes forem nulas; β_1 : é a variação esperada em Y dado um incremento unitário em X1, mantendo-se constantes todas as demais variáveis independentes; β_k : é a variação esperada em Y dado um incremento unitário em Xk, mantendo-se constantes todas as demais variáveis independentes; e_j : é o erro não explicado pelo modelo. (Maia, 2017).

A estimação dos parâmetros das equações de regressão linear múltipla foi baseada no método dos mínimos quadrados ordinários, que consiste em adotar os valores de α e β de modo que a soma dos erros seja igual a zero e a soma dos quadrados dos erros seja a mínima possível (Fávero e Belfiore, 2017).

A fim de melhorar a qualidade do modelo, cada mês do ano foi segmentado em 3 períodos, na seguinte arquitetura. O período um (P1) corresponde do 1º ao 10º dia do mês, o período 2 (P2) do 11º ao 20º dia e o período três (P3) do 21º ao último dia do mês. Em cada um destes períodos, calculou-se a média de chuva para cada uma das cidades e a vazão do rio no ponto de aferição do modelo. A segmentação se deu para evitar influências de eventos extremos no decorrer do mês que afetassem o modelo. Além disso, os valores foram segmentados em quartis, e aqueles com valores superiores ao 3º quartil ou inferior ao 1º quartil foram retirados da amostra para realização dos cálculos.

A partir disso, construiu-se um modelo para estimar a vazão em cada um dos meses, conforme demonstrado na tabela 2. Enquanto isso, na tabela 3 descreveu-se cada uma das variáveis empregadas no modelo para estimação da vazão.

Tabela 2. Variáveis dependentes e preditora para estimação da vazão

Modelo	Variáveis dependentes	Variáveis Predictoras
Modelo Vazao	Vazao	pt + v3ant + p3ant

Fonte: Próprio autor

Onde:

Tabela 3. Descrição das variáveis utilizadas no modelo.

Variável	Descrição
Vazao	Média vazão mensal
v3ant	Média vazão do 21º ao último dia do mês anterior
p3ant	Volume total de chuva das 8 cidades do 21º ao último dia do mês
Pt	Volume total de chuva das 8 cidades durante o mês

Fonte: Próprio autor

Em todos os modelos, empregou-se a técnica conhecida como stepwise e o teste de shapiro-francia nos resíduos a 5% de significância. Quando o teste de shapiro-francia não se mostrou significativo ($p\text{-value} < 0,05$), a variável foi submetida ao processo BOX-COX, em que a mesma foi transformada de acordo com a equação a seguir (Dal'col Lúcio et al., 2011).

$$Y_i^{(\lambda)} = \frac{(y^\lambda - 1)}{\lambda}$$

Em que: $Y_i = \acute{e}$ o valor transformado, $y =$ o valor a ser transformado e $\lambda =$ valor de lambda ideal para os dados, ou seja, aquele que maximiza a aderência dos valores transformados a normalidade.

Após ser transformada, o novo modelo foi construído e todas as variáveis preditoras submetidos ao procedimento stepwise e ao teste de shapiro franca a 5% de significancia. Isto feito, elaborou-se um grafico comparando os valores preditos com os valores encontrados pelo modelo.

Para estimar a probabilidade de ocorrência de seca, ou seja, quando a vazão for inferior a 10 m³/s, foi utilizada a técnica de regressão logística, que é uma técnica estatística utilizada para descrever o comportamento entre uma variável dependente binária e variáveis independentes métricas ou não métricas. Considerou-se o valor de 10 m³/s pois, abaixo deste valor, os especialistas indicam valor abaixo da média para o manancial (G1, 2022). Diferentemente da regressão múltipla, a regressão logística não pressupõe a existência de homogeneidade de variância e normalidade dos resíduos. Nesse tipo de regressão há o interesse em avaliar a probabilidade p de ocorrência de um determinado evento com base no comportamento de variáveis explicativas. Sendo assim, sabendo-se que a chance de ocorrência de um evento é dada por chance = (p/1-p) o modelo de regressão logística pode ser definido de acordo com o apresentado a seguir (Fávero e Belfiore, 2017).

$$\ln(chance) = Z = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj}$$

Que, ao desenvolver-se, chega-se a:

$$p = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-(\alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj})}}$$

Em que: Z: conhecido por logito; p: probabilidade estimada de ocorrência do evento de interesse; xi: são as variáveis explicativas, com i=1,2, ..., k; e α e β_i : são os parâmetros do modelo.

Enquanto na regressão múltipla baseia-se no método dos mínimos quadrados, a regressão logística consiste no método de máxima verossimilhança, ou seja, consiste em estimar os parâmetros de um modelo utilizando as estimativas que tornam máxima a probabilidade (verossimilhança) de que um evento ocorra (Batista, 2009).

Em nosso estudo, para elaboração do modelo, períodos em que a vazão média foi inferior a 10m³/s receberam a categoria “sim” para seca, ao passo que meses em que a vazão média foi superior a 10 m³/s reberam a categoria “não” para seca. A partir disso, construiu-se um modelo para estimar a probabilidade de seca no decorrer do mês, de acordo com os dados da tabela 4. A tabela 5 descreve cada uma das variáveis empregadas no modelo para estimação da vazão.

Tabela 4. Variáveis dependentes e preditora para estimação da seca

Modelo	Variáveis dependentes	Váriaveis Predictoras
Modelo seca	seca	pt + v3ant + p3ant

Fonte: Próprio autor

Onde:

Tabela 5. Descrição das variáveis utilizadas no modelo

Variável	Descrição
seca	Ocorrência de valor inferior a 10m ³ /s na média de vazão mensal.
v3ant	Média vazão do 21º ao último dia do mês anterior
p3ant	Volume total de chuva das 8 cidades do 21º ao último dia do mês
Pt	Volume total de chuva das 8 cidades durante o mês

Fonte: Próprio autor

Resultados e discussão

Primeiramente serão apresentados os resultados referentes aos modelos elaborados com objetivo de estimar a vazão do rio corumbatai em cada dos meses estudados. Na tabela 6 são apresentados os resultados finais dos modelos de regressão multipla com os coeficientes para cada uma das variáveis predictoras, para cada um dos meses do estudo

Tabela 6. Valores dos coeficientes para cada um dos meses em estudo

mês	intercepto	pt	v3ant	p3ant	lamba	R ²	sf
jun	3,682209	0,006364 *	0,493743*			0,8142	0,4312*
jul	0,3369426	0,006209 *	0,765851*			0,8377	0,4334*
ago	2,5084619	0,0036974 *	0,6037286*	-0,00521*		0,6624	0,3529*
set	1,24	-0,00025 *	-0,02475*		-0,32079	0,7354	0,7935*
out	0,4771098	0,008128 *	0,408428*			0,7823	0,4348*

Fonte: Resultado de pesquisa *P valor < 0,05

R² = coeficiente de determinação para cada um dos modelos.

Sf = Valor obtido através do teste de Shapiro-francia.

Lambda = Valor de lambda encontrado na transformação box-cox.

Pela análise da tabela é possível inferir que, para o rio corumbatai, nos meses de junho a outubro, as variáveis chuva do mês e a vazão de um período anterior são significantes em relação a vazão do mês, ou seja, o mês anterior ao mês corrente exerce influência na vazão.

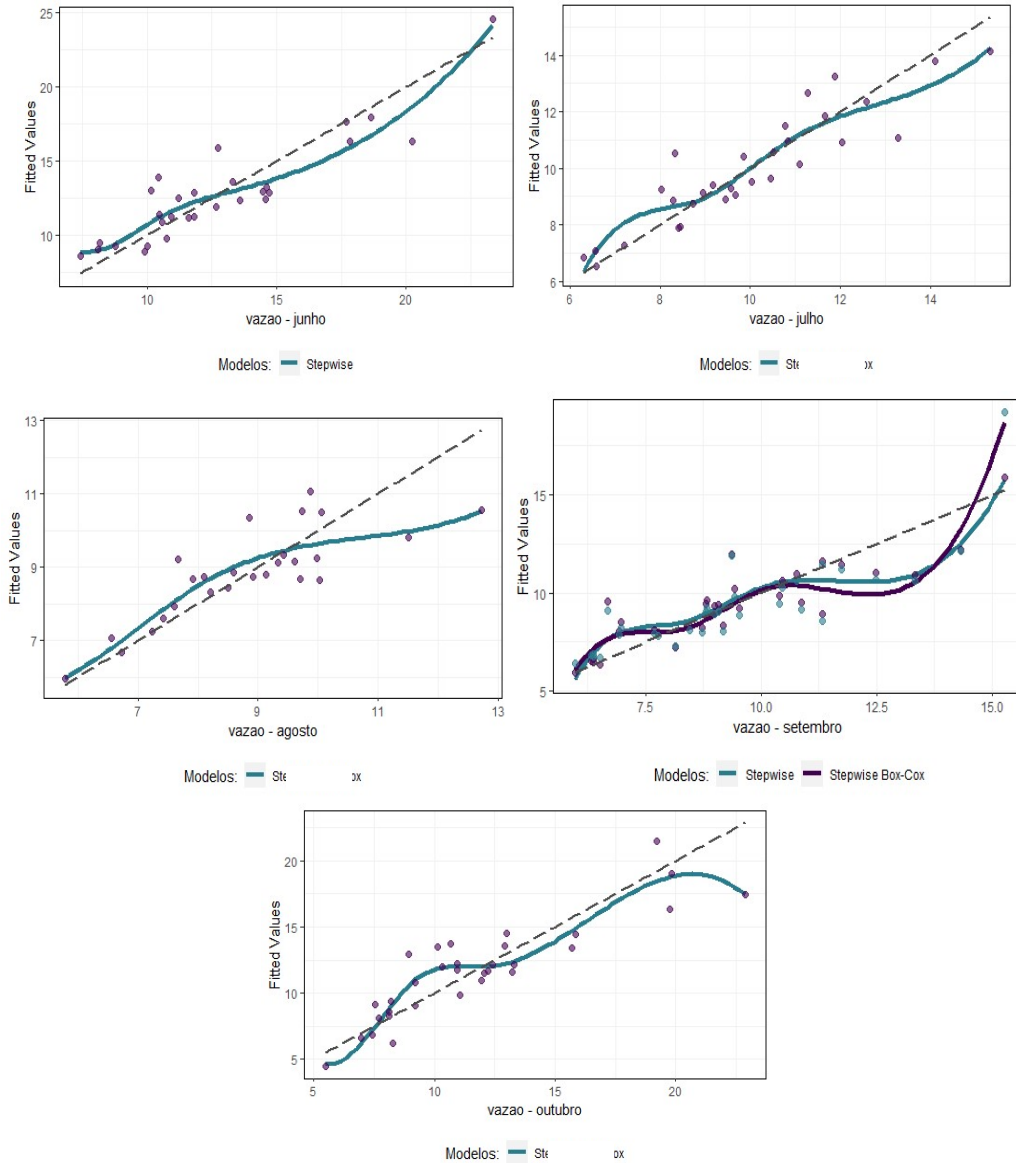
Malfatti et.al (2018), na construção do Modelo Empírico Linear para Previsão de Vazão de Rios na Usina Hidrelétrica de Itaipu – Bacia do Rio Paraná, concluíram que modelos mensais apresentaram melhores desempenhos na previsão de vazão e precipitação quando comparados a series mais longas. Além disso, o estudo mostrou que a vazão da usina itaipu é indicada como preditora com maior frequência ao longo dos meses, principalmente nas defasagens mais curtas.

Pela tabela, podemos observar valores de R² que variam de 0,6624 a 0,8377. O coeficiente R², ou coeficiente de determinação, é uma medida de qualidade do ajustamento da equação aos dados, podendo ser interpretado como a variabilidade do Yj que é estatisticamente explicada por meio das variáveis predictoras, ou seja, com as variáveis utilizados, explica-se a proporção da variável dependente (Hoffman, 2016). No entanto, valores de R² elevado não significam que um modelo é melhor do que outro modelo que possui R² com valores menores. Um R² baixo significar a não lineariedade do modelo ou que existem outros fatores (inclusive a aleatoriedade) que afetam a variabilidade da variável explicada, não que as estimações estejam incorretas. Quinino et al. (1991), apresentaram uma tabela com valores mínimos de R² para considerar um modelo útil a 5% de significancia. Considerando 3 variáveis predictoras, para dados com tamanho amostral no valor de 15, R² superior a 0,49 indicam que o modelo é util, ou seja, os modelos apresentados neste estudo atingem os valores mínimos.

Dos Santos et al. (2013), utilizando o modelo SWAT para simulação chuva-vazão na bacia do rio tapacura, estado de pernanbuco, obteve R² nos valores de 0,73 e 0,9, julgando assim que o modelo apresentou desempenho satisfatório. Em outro estudo, através da metodologia tank model, para modelagem chuva-vazão na bacia hidrográfica do ribeirão santa bárbara em goiás, alcançaram coeficiente de determinação R² igual a 0,8 (Costa e Honório, 2020).

Na figura 3 são apresentados os graficos relacionando os valores reais de vazão e os valores preditos pelo modelo, bem como valores numéricos reais de vazão e os preditos pelo modelo. A figura 6 mostra um comparativo dos valores reais (vazão) e valores previstos (yhat_step_modelo).

Figura 3 - Valores previstos (fitted values) x valores reais para cada mês



Fonte: Resultados da pesquisa

Comparando os dados, é possível verificar, em muitos momentos, a grande proximidade dos valores reais de vazão aos valores preditos pelo modelo, bem como uma boa captura do modelo quando observamos os gráficos. No entanto, notamos que alguns valores se distanciam da curva representada pelo modelo, sendo estes valores, na maioria das vezes, aqueles que apresentam poucas repetições durante as séries históricas, como exemplo, a ocorrência de eventos extremos.

Gorodetskaya et al. (2018), utilizaram redes neurais para previsão de vazão em função de observações de precipitação e vazão antecedentes, obtendo resultados satisfatórios para previsões de mais curto prazo. Entretanto, observou-se que os picos de vazão do rio não foram previstos com desejável precisão.

Welerson (2019), em estudo realizado para realizar previsões sazonais para o rio São Francisco, através do modelo MGB – IPH, obteve resultados semelhantes de R^2 para modelos que dados de meses antecedentes.

Para 1 mês de antecedência, os valores de R^2 foram de em torno de 0,7, coincidindo com o modelo aqui apresentado.

Em outro estudo, Uliana et al. (2020), apresentou um modelo hidrológico híbrido, utilizando Redes Neurais Artificiais (RNAs) combinadas aos modelos hidrológicos conceituais IPH II e SAC-SMA, de forma a obter um modelo híbrido para estimativa de vazões dos cursos de água da bacia do rio Piracicaba-MG. Nos resultados apresentados, o modelo proposto apresentou bom desempenho. Porém, quando observamos os gráficos, vazões observadas com valores extremos não conseguiram ser previstas pelo modelo, da mesma forma que o modelo de regressão múltipla. Na tabela 7 são apresentados os coeficientes de cada uma das variáveis do modelo de regressão logística binária

Tabela 7. Valores dos coeficientes para cada um dos meses em estudo modelo seca.

mês	intercepto	pt	v3ant	p3ant
jun	8,1917	-0,007804*	-0,652351*	
jul	23,117097	-0,016673*	-1,803986*	
ago	7,69495	-0,005102*	-0,516526*	0,004928*
set	26,226131	-0,01396*	-2,194432*	
out	11,203489	-0,015698*		

* Teste t (p-valor < 0,05)

Fonte: Resultado de pesquisa

Como pode ser observado, a vazão do curso da água no período antecedente ao mês e o volume das chuvas no mês em estudo assumem grande influência para a ocorrência ou não de secas, ao passo que a chuva do período anterior foi significativa em apenas um dos meses.

Após a elaboração dos modelos, destes foram avaliados a sensibilidade, especificidade e acurácia. Segundo Favero e Belfiore (2017), entende-se como sensibilidade o percentual de acerto dos casos de ocorrência do evento de interesse (verdadeiro positivo), especificidade como percentual de acerto dos casos de não ocorrência do evento (verdadeiro negativo) e acurácia, designada como a taxa de acerto do modelo, ou seja, é a probabilidade do modelo em fornecer resultados corretos, ou seja, ser positivo para seca e negativo para não seca. A tabela 8 apresenta os dados de sensibilidade, especificidade e acurácia do modelo.

Tabela 8. Sensibilidade, especificidade e acurácia (cutoff – 0,5)

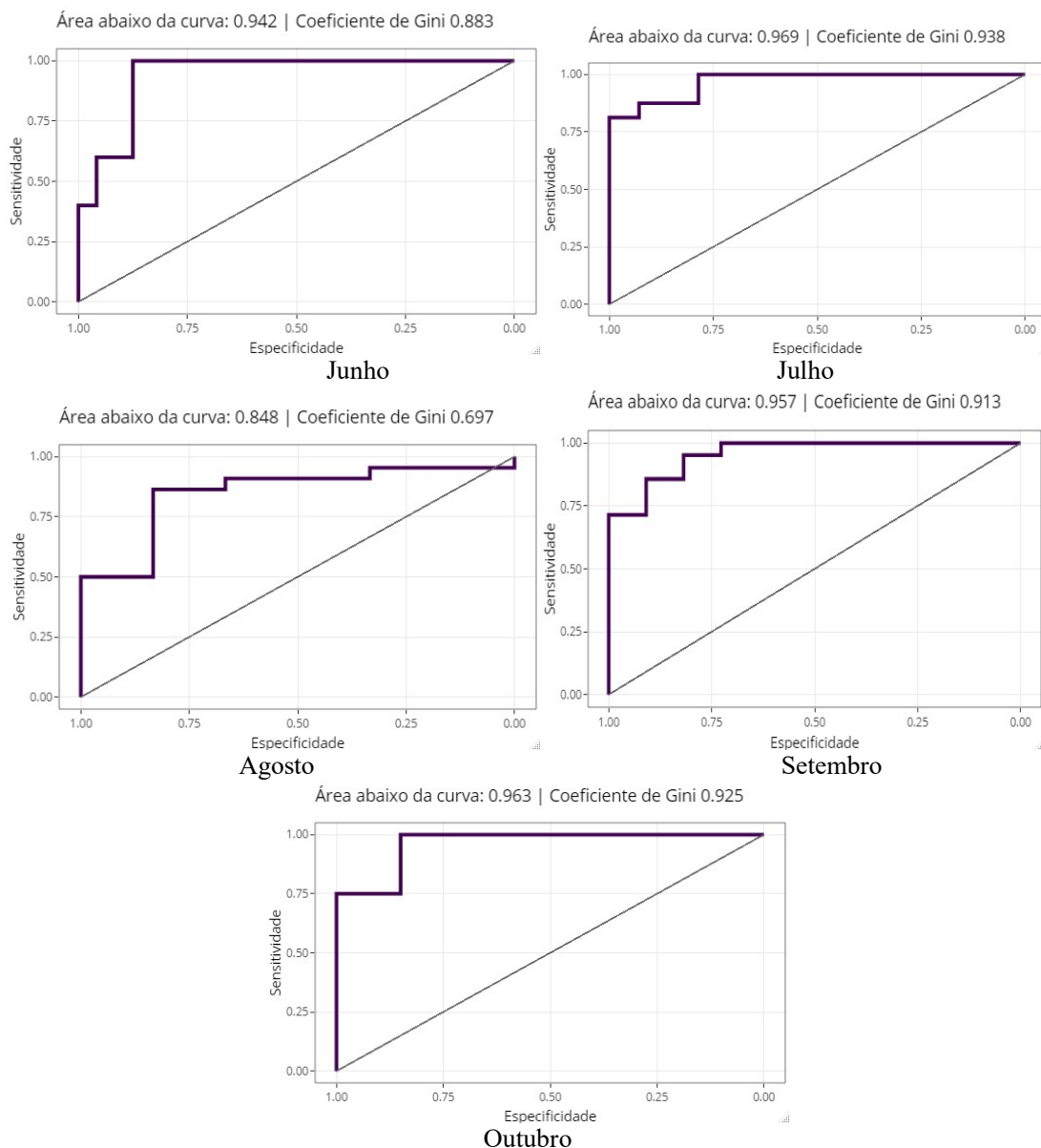
mês	sensibilidade	especificidade	acurácia
jun	0,6	0,95	0,89
jul	0,875	0,8571	0,866
ago	0,9545	0,5	0,8571
set	0,9523	0,81	0,90625
out	0,83	0,9	0,875

Fonte: Resultados da pesquisa

Como pode ser observado na tabela 8, para todos os meses a acurácia do modelo foi superior a 85%. Segundo Resende e Duarte (2007), valores de acurácia de 85% representa que o modelo possui uma classe de precisão alta, ao passo que, valores superiores a 90% indicam uma classe de previsão muito alta, ou seja, os modelos elaborados contam com uma alta precisão. Apesar da alta acurácia, em alguns meses, como junho e agosto, a sensibilidade e a especificidade não foram elevadas, respectivamente.

Além dos parâmetros de sensibilidade, especificidade e acurácia, elaborou-se também uma curva ROC (*Receiver Operating Characteristic*) para cada um dos meses, que é construída variando os pontos de corte (cutoff) ao longo da amplitude dos escores fornecidos pelos modelos. Desse modo, a curva ROC é construída tendo no seu eixo horizontal os valores de (1 – especificidade), ou seja, a proporção de não eventos classificados como eventos e no eixo vertical a sensibilidade, que é a proporção de eventos classificados como eventos (Silva, 2008). A figura 4 demonstra os resultados da curva ROC para cada um dos meses.

Figura 4 - Curva ROC para cada mês



Fonte: Resultados da pesquisa

Para interpretação das curvas ROC, temos que avaliar a área abaixo da curva ou a distância da diagonal. Quanto maior a área abaixo da curva (mais próximo de 1) e, conseqüentemente maior a distância da diagonal, melhor o modelo. A curva ROC mostra a eficiência global do modelo independente do cutoff (Speckhann, 2018). A curva ROC possibilita acompanhar a variação da sensibilidade e da especificidade. Quanto mais a curva ROC se aproxima do canto superior esquerdo, melhor é a qualidade do teste quanto à capacidade para discriminar os grupos. (Polo e Miot, 2020).

Em nosso estudo, todos os modelos apresentaram curvas ROC com área abaixo da curva superior a 0,84 e com grande proximidade do canto superior esquerdo. Fávero e Belfiore (2017), apresentaram um quadro informando que curvas ROC com área maior que 0,8 abaixo da curva garantem ao modelo uma discriminação

excelente dos grupos. Nesse caso, os modelos apresentados contam com uma discriminação excelente, uma vez que todos possuem área sobre a curva superior a 0,8, alguns alcançando valores de 0,969.

Diversos estudos envolvendo hidrologia fizeram utilização da curva ROC para avaliação dos modelos. Melo (2021), em estudo sobre aplicação de informações espaciais de alta resolução para estudo de ocorrências de inundações urbanas no estado de Pernambuco, fez uso da curva ROC para avaliar o modelo HAND quanto a assertividade de inundações, alcançando valores de curva ROC de 0,99. Outro estudo em que se utilizou da curva ROC para avaliação do modelo foi o realizado por Speckhann (2018), também para avaliação de inundação, obtendo valores semelhantes aos apresentados neste estudo.

Considerações Finais

Com base nos resultados obtidos pode-se concluir que, para o rio Corumbataí, nos meses de maio a outubro, as vazões do mês anterior bem como a chuva do mês presente interferem na vazão média e conseqüentemente na possibilidade de ocorrência de seca. Além disso, os modelos aqui apresentados, tanto para previsão de vazão mensal quanto para probabilidade de seca, quando comparados a outros modelos presentes na literatura apresentaram resultados satisfatórios.

No entanto, através de novos estudos, este modelo pode ser aperfeiçoado através do emprego de diferentes variáveis preditoras, como uso do solo e ocupação urbana. Assim, este estudo pode ser utilizado como um início para o aprofundamento de pesquisas, elevando assim a assertividade dos modelos.

Referências

Agência das bacias PCJ (2020). Estudo do Uso dos Recursos Hídricos na Bacia Hidrográfica do Rio Corumbataí, Relatório Final Volume I – Tomo I.

Almeida, H.M, de. (2014). Análise de regressão linear múltipla com estudo relacionado a horas de máquinas paradas na linha de produção de uma indústria de calçados. 31f. Trabalho de Conclusão de Curso (Graduação em Estatística)- Universidade Estadual da Paraíba, Campina Grande.

Batista, J.L.F. (2009). Verossimelhança e Máxima verossimelhança.

Carvalho, G.B. (2019). Diagnóstico do uso de águas subterrâneas na bacia do rio Corumbataí: Uma contribuição para a gestão sustentável. Rio Claro.

Chierice Júnior, N. (2013). Análise da chuva e vazão da bacia hidrográfica do Rio Pardo.

Costa, I. A.; da Silva Honório, M. (2020). Modelagem do processo chuva-vazão na bacia hidrográfica do ribeirão santa barbara em Goiás. *Tecnia*, 5(1) :102-118.

DAEE – Departamento de Águas e Energia Elétrica. (2022). Disponível em: < > Acesso em 03 mar. 2022

Dal'col Lucio , A. et al. (2011). Transformação box-cox em experimentos com pimentão em ambiente protegido. *Horticultura Brasileira*, v. 29, n. 1: 38-42.

Dos Santos, J. Y. G.; Carvalho Neto, J. G.; da Silva, R. M.; Montenegro, S. M. G. L. & Santos, C. A. G. (2013). Aplicação do modelo SWAT para a simulação chuva-vazão na Bacia do Rio Tapacurá, Estado de Pernambuco.

Fávero, L.P. & Belfiore. (2017). P. Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata. Rio de Janeiro.

Fraga, M. S.; Abreu, M. C.; Reis, G. B.; Tozi, T.; Pinheiro, S. A. R. (2020). Análise de tendência em séries históricas de precipitação e vazão na UGRH2 Piracicaba, Minas Gerais. *Revista Ibero Americana de Ciências Ambientais*, v.11, n.2:136-144.

G1. Piracicaba e Região. Efeito da estiagem, Rio Corumbataí em Piracicaba registra vazão 41% abaixo da média para 2022. (2022) Disponível em < <https://g1.globo.com/sp/piracicaba-regiao/noticia/2022/07/14/efeito-da-estiagem-rio-corumbatai-em-piracicaba-registra-vazao-41percent-abaixo-da-media-para-2022.ghtml> > Acesso em: 01 ago. 2022

Gorodetskaya, Y.; da Fonseca, L. G.; de Melo Ribeiro, C. B. (2018). Previsão de vazão de curto prazo utilizando redes neurais artificiais. estudo de caso: Rio paraíba do sul.

Hoffmann, R. (2016) Análise estatística de relações lineares e não lineares.

Maia, A.G. (2017). *Econometria: conceitos e aplicações*. Cap. 6.

Maroco, J.; (2003). *Análise Estatística – Com utilização do SPSS*, 2ª edição; Edições Sílabo.

Malfatti, M. G. L.; Cardoso, A. D. O.; Hamburger, D. S. (2018). Modelo empírico linear para previsão de vazão de rios na usina hidrelétrica de Itaipu–Bacia do Rio Paraná. *Revista Brasileira de Meteorologia*, 33: 257-268.

Melo, V.B. (2021). *Aplicação de informações espaciais de alta resolução para estudo de ocorrências de inundações urbanas*. Dissertação de Mestrado. Universidade Federal de Pernambuco.

Mori, G.B.; Cassiano, C. C.; Ferraz, S.; Camargo, A. (2016). Bacia do rio Corumbataí: uso do solo e características limnológicas.

Polo, T. C. F.; Miot, H. A. (2020). Aplicações da curva ROC em estudos clínicos e experimentais. *Jornal Vascular Brasileiro*, 19.

Quinino, R. C.; Reis, E. A. & Bessegato, L. F. (1991). O coeficiente de determinação R² como instrumento didático para avaliar a utilidade de um modelo de regressão linear múltipla. Belo Horizonte: UFMG.

Resende, M. D. V. de. & Duarte, J. B. (2007). Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Embrapa Florestas-Artigo em periódico indexado*.

Santos, S. C. dos, & Garcia, G. J. (2013). Gestão De Bacias Hidrográficas: Desenvolvimento Integrado E Políticas Públicas Para A Bacia Hidrográfica Do Rio Corumbataí. *Geoambiente On-Line*, (3): 01-21 .

Silva, P. H. F. (2008). *Medidas do Valor Preditivo de Modelos de Classificação Aplicados a Dados de Crédito*.

Speckhann, G. A. (2018). Uma proposta de mapeamento de risco de inundação na bacia do rio Itajai-Açu utilizando descritores de terreno.

Tucci, C. E. M. (2001). *Gestão da água no Brasil*. Brasília: UNESCO, 156p.

Uliana, E. M.; Silva, D. D. D.; Rodrigues, B. S. & Corrêdo, L. D. P. (2015). Análise de tendência em séries históricas de vazão e precipitação: uso de teste estatístico não paramétrico. *Revista Ambiente & Água*, 10: 82-88.

Uliana, E. M.; Silva, D. D. D.; Moreira, M. C.; Pereira, D. D. R. & Almeida, F. T. D. (2020). Modelo Hidrológico Híbrido para Previsão de Vazões na Bacia do Rio Piracicaba-MG. *Revista Brasileira de Meteorologia*, 34: 471-480.

Wellerson, C. C. (2020). Previsões sazonais de vazão para o rio São Francisco com base em previsões numéricas de precipitação.

Apêndice

Código para elaboração do modelo multinível.

```
##### BASE DE DADOS #####

library(readxl)
dataset_corumbatai <- read_excel("corumbatai1.13.xlsx")
dataset_corumbatai_1 <- na.omit(dataset_corumbatai)
dataset_corumbatai_1_1 <- dataset_corumbatai_1

mes <- dataset_corumbatai_1_1 %>%
  filter(mes %in% c(' '))

##### Estimando a Regressão Múltipla #####

modelo_vazao <- lm(formula = vazao ~ pt + v3ant + p3ant,
  data = mes)

#Aplicando o procedimento Stepwise, temos o seguinte código:
step_vazao <- step(modelo_vazao, k = 3.841459)
summary(step_vazao)

TESTE DE VERIFICAÇÃO DA ADERÊNCIA DOS RESÍDUOS À NORMALIDADE #
# SHAPIRO-FRANCIA #

#Shapiro-Francia: n > 30
sf.test(step_vazao$residuals) #função sf.test do pacote nortest

##### TRANSFORMAÇÃO DE BOX-COX #####

#Para calcular o lambda de Box-Cox

lambda_BC <- powerTransform(mes$vazao) #função powerTransform do pacote car#
lambda_BC

#Inserindo o lambda de Box-Cox na base de dados para a estimação de um novo modelo
mes$bcvazao <- (((mes$vazao ^ lambda_BC$lambda) - 1) /
  lambda_BC$lambda)
#Estimando um novo modelo múltiplo com variável dependente transformada por Box-Cox

modelo_bc <- lm(formula = bcvazao ~ p1 + v3ant + p3ant,
  data = mes)
#Aplicando o procedimento Stepwise

step_modelo_bc <- step(modelo_bc, k = 3.841459)

#Verificando a normalidade dos resíduos do modelo step_modelo_bc
sf.test(step_modelo_bc$residuals) #função sf.test do pacote nortest

#Ajustes dos modelos: valores previstos (fitted values) X valores reais com bc
```

```

mes %>%
  ggplot() +
  geom_smooth(aes(x = vazao3, y = yhat_step_vazao, color = "Stepwise"),
    method = "lm", se = F, formula = y ~ splines::bs(x, df = 5), size = 1.5) +
  geom_point(aes(x = vazao3, y = yhat_step_vazao),
    color = "#440154FF", alpha = 0.6, size = 2) +
  geom_smooth(aes(x = vazao3, y = yhat_step_modelo_bc, color = "Stepwise Box-Cox"),
    method = "lm", se = F, formula = y ~ splines::bs(x, df = 5), size = 1.5) +
  geom_point(aes(x = vazao3, y = yhat_step_modelo_bc),
    color = "#287D8EFF", alpha = 0.6, size = 2) +
  geom_smooth(aes(x = vazao3, y = vazao3), method = "lm",
    color = "grey30", size = 1.05,
    linetype = "longdash") +
  scale_color_manual("Modelos:",
    values = c("#287D8EFF", "#440154FF")) +
  labs(x = "vazao - setembro", y = "Fitted Values") +
  theme(panel.background = element_rect("white"),
    panel.grid = element_line("grey95"),
    panel.border = element_rect(NA),
    legend.position = "bottom")

```

Códigos para elaboração do modelo logístico binário.

```
##### ESTIMAÇÃO DE UM MODELO LOGÍSTICO BINÁRIO #####
```

```
## Criação da variável binária
```

```
dataset_corumbatai_arrumado %>%
mutate(seca = ifelse(vazao < 10, yes = "sim",
  no = "não"),
  seca = factor(seca)) -> dataset_corumbatai_arrumado
view(dataset_corumbatai_arrumado)
```

```
#Estimando o modelo logístico binário
```

```
base <- dataset_corumbatai_arrumado %>% filter(mes %in% c('6'))
modelo_seca <- glm(formula = seca ~ pt + v3ant + p3ant,
  data = base,
  family = "binomial")
```

```
#Procedimento Stepwise
```

```
step_seca <- step(object = modelo_seca,
  k = qchisq(p = 0.05, df = 1, lower.tail = FALSE))
summary(step_seca)
```

```
#####
EXEMPLO 01 - CONSTRUÇÃO DA CURVA ROC
#####
```

```
#função roc do pacote pROC
```

```
ROC <- roc(response = base$seca,
  predictor = modelo_seca$fitted.values)
```

```
ggplotly(
  ggroc(ROC, color = "#440154FF", size = 1) +
  geom_segment(aes(x = 1, xend = 0, y = 0, yend = 1),
    color="grey40",
    size = 0.2) +
  labs(x = "Especificidade",
    y = "Sensitividade",
    title = paste("Área abaixo da curva:",
      round(ROC$auc, 3),
      "|",
      "Coeficiente de Gini",
      round((ROC$auc[1] - 0.5) / 0.5, 3))) +
  theme_bw()
```