

**RENAN GON FERREIRA**

**MODELAGEM GEOESTATÍSTICA E APRENDIZADO DE MÁQUINA COMO  
MÉTODOS ALTERNATIVOS ÀS METODOLOGIAS TRADICIONAIS DE  
REGIONALIZAÇÃO DE VAZÕES EM BACIAS HIDROGRÁFICAS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Magister Scientiae*.

Orientador: Demetrius David da Silva

Coorientadores: Abrahão A. Alden Elesbon  
Gérson Rodrigues dos Santos

**VIÇOSA – MINAS GERAIS**

**2020**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

F383m  
2020  
Ferreira, Renan Gon, 1992-  
Modelagem geoestatística e aprendizado de máquina como  
métodos alternativos às metodologias tradicionais de  
regionalização de vazões em bacias hidrográfica / Renan Gon  
Ferreira. – Viçosa, MG, 2020.  
98 f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Demetrius David da Silva.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Desenvolvimento de recursos hídricos. 2. Hidrologia -  
Modelos estatísticos. 3. Fluxo de água. 4. Bacias hidrográficas.  
I. Universidade Federal de Viçosa. Departamento de Engenharia  
Agrícola. Programa de Pós-Graduação em Engenharia Agrícola.  
II. Título.

CDD 22 ed. 627


RENAN GON FERREIRA

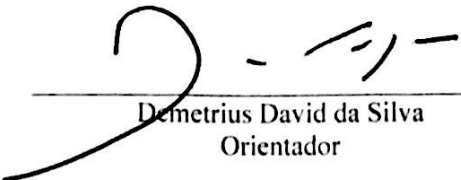
**MODELAGEM GEOESTATÍSTICA E APRENDIZADO DE MÁQUINA COMO  
MÉTODOS ALTERNATIVOS ÀS METODOLOGIAS TRADICIONAIS DE  
REGIONALIZAÇÃO DE VAZÕES EM BACIAS HIDROGRÁFICAS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Magister Scientiae*.

APROVADA: 21 de julho de 2020.

Assentimento:

  
\_\_\_\_\_  
Renan Gon Ferreira  
Autor

  
\_\_\_\_\_  
Demetrius David da Silva  
Orientador

## DEDICATÓRIA

Aos meus pais, Olália e Darci,  
pelo apoio de sempre e por cada gota de suor e de lágrima que derramaram por meus estudos.

Aos meus irmãos, Rômulo e Rafael,  
que estão sempre ao meu lado e que me incentivaram e apoiaram durante todo o mestrado.

Dedico.

## AGRADECIMENTOS

A Deus, pelo dom da Vida e por toda a saúde, força, perseverança e habilidade que me concedeu para a realização do mestrado e desta dissertação. Em todas as dificuldades, tristezas, ansiedades e frustrações, o Senhor me sustentou e me ajudou a vencer, sem me deixar desistir de qualquer batalha. Agradeço à Maria, Nossa Senhora, que esteve junto a Cristo, seu filho, Nosso Senhor e Salvador, intercedendo por mim, pela minha vida e saúde e de todos aqueles que eu amo e que estiveram longe durante o mestrado. Agradeço, Pai e Mãe.

Aos meus pais e irmãos, que estiveram presentes em todos os momentos, apesar da distância. Obrigado aos meus pais, Olália Maria Gon Ferreira e Darci Gomes Ferreira, por toda a dedicação à nossa família. O empenho de vocês foi essencial para o meu mestrado e para esta dissertação. Sou eternamente grato. Obrigado aos meus irmãos, Rômulo e Rafael Gon Ferreira, por estarem comigo em todos os momentos que precisei, aliviando alguns momentos de dificuldade e preocupação, com boas conversas e boas risadas. Agradeço por todo o apoio e incentivo que recebi, desde quando resolvi trilhar este caminho. Eu amo muito vocês, obrigado por existirem, por fazerem parte da minha vida e estarem sempre presentes.

À Tia Rita e tio Herculano, muito obrigado a vocês que estiveram presentes neste tempo, pelos bons momentos compartilhados e pelos incentivos ainda antes de eu iniciar o mestrado. Também agradeço à Sayo, Ru e Sophia pelas boas conversas e alegrias durante este período. Todos vocês foram muito importantes para a realização deste trabalho.

Ao prof. Demetrius, por toda a sua dedicação e precisão na orientação. Agradeço ao amigo e prof. Abraão, pela ótima coorientação, com enorme prontidão. Obrigado ao prof. Géron pela coorientação, e ao prof. Elpídio, que abriu as portas do LabGEO/UFV para parte da pesquisa e pelos valiosas contribuições ao trabalho. Aos amigos Micael, Carlota e Lucas, e todos aqueles do CRRH/UFV, pelas boas conversas, risadas e troca de conhecimento. A todos que direta ou indiretamente contribuíram para a realização desta dissertação, meu muito obrigado. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Agradeço, ainda, em homenagem póstuma, ao Ser Humano, poliglota, exímio pianista, vocalista, músico, compositor e maestro, Andre Matos, que fez parte de todo o período desta dissertação. Todos nós perdemos com a falta do Andre: a música, a arte e as pessoas. Certamente, suas obras e seu grande legado serão passados às próximas gerações e ficarão gravados eternamente em nossos corações. Assim, deixo a epígrafe nas mãos do maestro.

## EPÍGRAFE

“Cold is the wind and thunder struck  
On a stormy night  
But can't you see I'm by your side?  
We are marching on!...  
So, carry on,  
There's a meaning to life  
Which someday we may find...”

Música: Carry on  
Composição: Andre Matos (*in memoriam*)

## RESUMO GERAL

FERREIRA, Renan Gon, M.Sc., Universidade Federal de Viçosa, julho de 2020. **Modelagem geoestatística e aprendizado de máquina como métodos alternativos às metodologias tradicionais de regionalização de vazões em bacias hidrográfica.** Orientador: Demetrius David da Silva. Coorientadores: Abrahão A. Alden Elesbon e Gérson Rodrigues dos Santos.

A gestão e o planejamento integrado dos recursos hídricos em bacias hidrográficas de países emergentes enfrentam grandes desafios com a falta de informações hidrológicas e monitoramento deficiente. Dentre as suas ferramentas de apoio está a regionalização de vazões: um conjunto de procedimentos que exploram as informações de locais monitorados para estimar a variável de interesse em locais sem monitoramento. Ainda não existe um método universal de regionalização de vazões. Os esforços científicos sobre a adequabilidade de diferentes metodologias para contextos específicos de bacias hidrográficas visam melhorar ou possibilitar a gestão dos recursos hídricos em bacias com pouca ou nenhuma informação hidrológica, favorecendo a gestão dos recursos hídricos com alocação mais eficiente dos recursos humanos e financeiros. Assim, o presente trabalho foi dedicado à avaliação de diferentes métodos e abordagens de regionalização de vazões na bacia hidrográfica do rio Doce para fins de estimativa das vazões mínimas de referência ( $Q_{7,10}$ ,  $Q_{90}$  e  $Q_{95}$ ) e média de longa duração ( $Q_{mld}$ ). No primeiro artigo foram avaliados métodos de regionalização de vazões baseados em interpolação linear, regressões regionais e geoestatística. Objetivou-se comparar o desempenho e a viabilidade de implementação da geoestatística como alternativa às demais metodologias de regionalização de vazões. A geoestatística apresentou desempenho similar às outras abordagens, revelando-se uma alternativa promissora em situações que inviabilizam o emprego das outras metodologias, como em bacias menores e de monitoramento hidrológico precário ou inexistente. O segundo artigo explorou modelos de aprendizado de máquina para a regionalização de vazões. O objetivo foi avaliar o desempenho preditivo de um modelo linear e dos modelos não lineares Random Forest e Splines de regressões adaptativas multivariadas (Earth). O desempenho do modelo linear foi insatisfatório. A covariável mais importante aos modelos não lineares foi a vazão equivalente à precipitação com abstração de 750 mm ( $Peq750$ ). Ambos os modelos não lineares revelaram grande habilidade preditiva, constituindo alternativas poderosas e promissoras à regionalização de vazões em suporte à gestão dos recursos hídricos na bacia do rio Doce.

Palavras-chave: Hidrologia. Regionalização de vazões. Gestão de recursos hídricos.

## ABSTRACT

FERREIRA, Renan Gon, M.Sc., Universidade Federal de Viçosa, July, 2020. **Geostatistical modeling and machine learning as alternative methods to the traditional methodologies of streamflow regionalization in watersheds.** Advisor: Demetrius David da Silva. Co-advisors: Abrahão A. Alden Elesbon and Gérson Rodrigues dos Santos.

The management and integrated planning of water resources in watersheds in developing countries face great challenges due to the lack of hydrological information and deficient monitoring. Among their support tools is the streamflow regionalization: a set of procedures that explore the information of monitored sites to estimate the interest variable in locations without monitoring. There is still no universal method of streamflow regionalization. Scientific efforts about the suitability of different methodologies for specific contexts of watersheds aim to improve or enable the water resources management in basins with little or no hydrological information, favoring the water resources management with more efficient allocation of human and financial resources. Thus, the present work was dedicated to the evaluation of different methods and approaches of streamflow regionalization in the Doce river basin for the purpose of estimating the minimum reference streamflows ( $Q_{7,10}$ ,  $Q_{90}$  and  $Q_{95}$ ) and long term streamflow ( $Q_{mld}$ ). In the first article, methods of streamflow regionalization based on linear interpolation, regional regressions and geostatistics were evaluated. The objective was to compare the performance and the feasibility of implementing geostatistics as an alternative to other streamflow regionalization methodologies. Geostatistics has performed similarly to other approaches, proving to be a promising alternative in situations that make the use of other methodologies unfeasible, such as in smaller basins with precarious or nonexistent hydrological monitoring. The second article explored machine learning models for streamflow regionalization. The objective was to evaluate the predictive performance of a linear model and the non-linear models Random Forest and Splines of multivariate adaptive regressions (Earth). The linear model performance it was unsatisfactory. The most important covariate for the non-linear models was the streamflow equivalent to the precipitated volume, considering the subtraction of the precipitation abstraction factor of 750 mm ( $P_{eq750}$ ). Both non-linear models Random Forest and Earth showed great predictive ability, constituting powerful and promising alternatives to the streamflow regionalization in support of the water resources management in the Doce River basin.

Keywords: Hydrology. Streamflow regionalization. Water resources management.

## SUMÁRIO

<b>INTRODUÇÃO GERAL .....</b>	<b>10</b>
<b>REFERÊNCIAS .....</b>	<b>12</b>
<b>CAPÍTULO 1:.....</b>	<b>15</b>
<b>Regionalização de vazões utilizando metodologias tradicionais e geostatística para a bacia do rio Doce, Brasil .....</b>	<b>15</b>
1.1 INTRODUÇÃO .....	16
1.2 MATERIAL E MÉTODOS .....	17
1.2.1 Caracterização da área de estudo.....	18
1.2.2 Preparação da base de dados hidrometeorológica.....	19
1.2.3 Vazões de referência para regionalização .....	21
1.2.4 Processamento das variáveis independentes .....	21
1.2.5 Metodologias de regionalização de vazão.....	22
1.2.5.1 Método de Regressões Regionais .....	22
1.2.5.2 Interpolação Linear (ELETROBRÁS, 1985).....	24
1.2.5.3 Interpolação Linear modificada (NOVAES et al., 2007) .....	28
1.2.5.4 Chaves (CHAVES et al., 2002) .....	28
1.2.5.5 Metodologia de Chaves modificada (NOVAES et al., 2007).....	29
1.2.6 Métodos de abordagem geostatística.....	31
1.2.6.1 Análise exploratória dos dados .....	31
1.2.6.2 Modelagem variográfica .....	31
1.2.6.3 Krigagem Simples.....	32
1.2.6.4 Krigagem Ordinária .....	32
1.2.7 Análise comparativa de desempenho das metodologias de regionalização .....	33
1.3 RESULTADOS E DISCUSSÃO .....	34
1.3.1 Método de Regressões Regionais.....	34
1.3.1.1 Identificação de Regiões Hidrologicamente Homogêneas (RHH) .....	34
1.3.1.2 Obtenção das equações regionais de regressão.....	35
1.3.2 Metodologias baseadas na abordagem de interpolação linear.....	38
1.3.3 Modelagem geostatística .....	40
1.3.4 Análise comparativa de desempenho das diferentes metodologias.....	42
1.3.4.1 Vazões mínimas de referência ( $Q_{7,10}$ , $Q_{95}$ e $Q_{90}$ ).....	42

1.3.4.2	Vazão média de referência ( $Q_{mld}$ ) .....	47
1.4	CONCLUSÕES .....	51
1.5	REFERÊNCIAS .....	52
<b>CAPÍTULO 2:</b>	.....	<b>58</b>
<b>Regionalização de vazões utilizando modelos de aprendizado de máquina para a bacia do rio Doce, Brasil</b>	.....	<b>58</b>
2.1	INTRODUÇÃO .....	59
2.2	MATERIAL E MÉTODOS .....	60
2.2.1	Área de estudo .....	60
2.2.1.1	Caracterização da bacia do rio Doce.....	61
2.2.2	Dados hidrometeorológicos.....	62
2.2.3	Variáveis resposta.....	63
2.2.4	Covariáveis ambientais.....	64
2.2.5	Aplicação dos modelos de aprendizado de máquina.....	66
2.2.5.1	Seleção das covariáveis ambientais .....	66
2.2.5.2	Treinamento, seleção e validação dos modelos .....	67
2.2.6	Análise comparativa do desempenho.....	70
2.3	RESULTADOS E DISCUSSÃO .....	71
2.3.1	Eliminação de covariáveis ambientais .....	71
2.3.2	Seleção e importância das covariáveis no ajuste dos modelos preditivos.....	71
2.3.2.1	Modelos não lineares .....	71
2.3.2.2	Modelo linear .....	74
2.3.3	Análise e comparação de desempenho .....	75
2.4	CONCLUSÕES .....	83
2.5	REFERÊNCIAS .....	84
<b>CONCLUSÕES GERAIS</b>	.....	<b>91</b>
<b>APÊNDICE A</b>	.....	<b>92</b>
<b>APÊNDICE B</b>	.....	<b>93</b>
<b>APÊNDICE C</b>	.....	<b>95</b>
<b>APÊNDICE D</b>	.....	<b>96</b>

## INTRODUÇÃO GERAL

A disponibilidade de dados hidrológicos é crucial para a gestão efetiva dos recursos hídricos (ANKIT et al., 2018; NRUTHYA; SRINIVAS, 2015). Em geral, o monitoramento hidrométrico nas bacias hidrográficas de países emergentes é inexistente ou precário, em termos de qualidade e quantidade dos dados medidos (ARAUJO et al., 2018; DANANDEH MEHR et al., 2015; OCIO; BESKEEN; SMART, 2019; SEIBERT et al., 2019; ZAMOUM; SOUAG-GAMANE, 2019). Dadas as dificuldades de gestão advindas dessa realidade, a estimativa ou predição acurada dos dados hidrológicos torna-se um dos procedimentos mais importantes ao planejamento integrado e gestão adequada dos recursos hídricos (BABAEI; MOEINI; EHSANZADEH, 2019; MEHDIZADEH et al., 2019).

Dentre as inúmeras potencialidades de se ter a disponibilidade de dados hidrológicos obtidos em redes hidrométricas projetadas adequadamente, vale destacar: análise de tendências hidroclimáticas; redução de incertezas estatísticas na previsão de eventos extremos (RUHI; MESSENGER; OLDEN, 2018); aprimoramento e adequação do valor de outorga e cobrança pelo uso da água, bem como de políticas relativas à gestão dos recursos hídricos; e a execução de projetos de infraestrutura (ANKIT et al., 2018; DESPAX et al., 2016).

No Brasil, as limitações das redes de monitoramento fluviométrico são notórias, principalmente em bacias menores (BESKOW et al., 2016; COMINI et al., 2020). Enquanto diversas regiões são desprovidas de monitoramento, devido à baixa densidade de estações nas bacias hidrográficas, outras regiões apresentam séries históricas insuficientes para orientar ações dos gestores de recursos hídricos (MELATI; MARCUZZO, 2016; PIOL et al., 2019), inviabilizando o atendimento a uma ampla variedade de propósitos práticos e científicos.

Por isso, a regionalização de vazões constitui uma das alternativas contra os obstáculos oriundos de monitoramentos deficitários (SWAIN; PATRA, 2017a; TEGEGNE; KIM, 2018). Essa técnica permite explorar os dados observados numa região, visando prever ou estimar a vazão dos cursos de água em regiões sem monitoramento (JAVEED; APOORVA, 2015; LI et al., 2018; ROJAS-SERNA et al., 2016; SILVA; BLANCO; PESSOA, 2019).

Neste sentido, e devido à importância da predição de vazões em bacias não monitoradas, a Associação Internacional de Ciências Hidrológicas (IAHS) lançou, em 2003, uma iniciativa contra os problemas da falta de monitoramento, intitulada de “Década de previsões em bacias não monitoradas (PUB): 2003–2012” (KUZMIN et al., 2019; ZHANG et

al., 2020), resultando no desenvolvimento e análise de diversas abordagens de regionalização de vazões em variados contextos, regiões e condições climáticas (ARSENAULT et al., 2019).

A regionalização de vazões representa uma ferramenta tão poderosa à melhoria da gestão dos recursos hídricos quanto mais acurada e confiável for a sua predição ou estimativa. E apesar dos esforços científicos, ainda não existe uma metodologia universal de regionalização de vazões (PRUSKI et al., 2015). Logo, metodologias de diferentes abordagens foram propostas ao longo do tempo, deixando evidente que algumas são mais adequadas ou confiáveis para cada região ou bacia hidrográfica do que outras (ARAUJO et al., 2018; CECÍLIO et al., 2018; COMINI et al., 2020; SWAIN; PATRA, 2017b).

Diante do exposto, o objetivo geral deste trabalho foi avaliar metodologias de regionalização de vazões de diferentes abordagens e identificar as mais adequadas à bacia hidrográfica do rio Doce, visando possibilitar que as tomadas de decisão sejam mais seguras, rápidas e precisas, e que a gestão e o planejamento dos recursos hídricos sejam mais efetivos.

Portanto, o primeiro capítulo (Artigo 1) do trabalho, intitulado “Regionalização de vazões utilizando metodologias tradicionais e geoestatística para a bacia do rio Doce, Brasil”, apresenta uma comparação entre metodologias tradicionais, usualmente adotadas no país, com métodos baseados na abordagem geoestatística, possibilitando apontar a mais adequada para a estimativa ou predição da vazão na bacia de estudo.

O segundo capítulo (Artigo 2), intitulado “Regionalização de vazões utilizando modelos de aprendizado de máquina para a bacia do rio Doce, Brasil”, visa avaliar a acurácia de diferentes modelos de aprendizado de máquina na predição da vazão ao longo da bacia de estudo. Com isso, analisaram-se potencialidades, limitações e o ganho geral do uso de modelos da abordagem de inteligência artificial para a regionalização de vazões em apoio à gestão e planejamento integrado dos recursos hídricos na bacia hidrográfica do rio Doce.

## REFERÊNCIAS

- ANKIT, A. et al. Optimal design of hydrometric station networks based on complex network analysis. **Hydrology and Earth System Sciences**, p. 1–21, 2018.
- ARAUJO, F. C. et al. Streamflow regionalization in Piquiri river basin. **Engenharia Agrícola**, v. 4430, p. 22–31, 2018.
- ARSENAULT, R. et al. Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. **Hydrological Sciences Journal**, v. 64, n. 11, p. 1297–1311, 2019.
- BABAEI, M.; MOEINI, R.; EHSANZADEH, E. Artificial Neural Network and Support Vector Machine Models for inflow prediction of dam reservoir (Case Study: Zayandehroud Dam Reservoir). **Water Resources Management**, v. 33, n. 6, p. 2203–2218, 2019.
- BESKOW, S. et al. Potential of the LASH model for water resources management in data-scarce basins: a case study of the Fragata River basin, southern Brazil. **Hydrological Sciences Journal**, v. 61, n. 14, p. 2567–2578, 2016.
- CECÍLIO, R. A. et al. Evaluation of low-flow and long term average flow regionalization methods on Itapemirim river watershed. **Scientia Agraria**, v. 19, n. 2, p. 122–132, 2018.
- DANANDEH MEHR, A. et al. Successive-station monthly streamflow prediction using different artificial neural network algorithms. **International Journal of Environmental Science and Technology**, v. 12, n. 7, p. 2191–2200, 2015.
- DESPAX, A. et al. Considering sampling strategy and cross-section complexity for estimating the uncertainty of discharge measurements using the velocity-area method. **Journal of Hydrology**, v. 533, p. 128–140, 2016.
- JAVEED, Y.; APOORVA, K. V. Flow regionalization under limited data availability – application of IHACRES in the Western Ghats. **Aquatic Procedia**, v. 4, p. 933–941, 2015.
- KUZMIN, V. et al. Method of prediction of the stream flows in poorly gauged and ungauged basins. **Journal of Ecological Engineering**, v. 20, n. 1, p. 180–187, 2019.
- LI, Q. et al. Hydrological regionalisation based on available hydrological information for runoff prediction at catchment scale. **Proceedings of the International Association of Hydrological Sciences**, v. 379, n. 1, p. 13–19, 2018.

- MEHDIZADEH, S. et al. Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: A local and external data analysis approach. **Journal of Hydrology**, v. 579, p. 124225, 2019.
- MELATI, M. D.; MARCUZZO, F. F. N. Regressões simples e robusta na regionalização da vazão Q95 na Bacia Hidrográfica do Taquari-Antas Simple and robust regressions in flows regionalization of Q95 in Taquari-Antas river basin. **Ciencia e Natura**, v. 38, p. 722–739, 2016.
- NRUTHYA, K.; SRINIVAS, V. V. Evaluating methods to predict streamflow at ungauged sites using regional Flow Duration Curves: A Case Study. **Aquatic Procedia**, v. 4, p. 641–648, 2015.
- OCIO, D.; BESKEEN, T.; SMART, K. Fully distributed hydrological modelling for catchmentwide hydrological data verification. **Hydrology Research**, v. 50, n. 6, p. 1520–1534, 2019.
- PIOL, M. V. A. et al. Performance evaluation of flow duration curves regionalization methods. **Revista Brasileira de Recursos Hidricos**, v. 24, 2019.
- PRUSKI, F. F. et al. Low-flow estimates in regions of extrapolation of the regionalization equations: A new concept. **Engenharia Agricola**, v. 35, n. 5, p. 808–816, 2015.
- ROJAS-SERNA, C. et al. How should a rainfall-runoff model be parameterized in an almost ungauged catchment? A methodology tested on 609 catchments. **Water Resources Research**, p. 4765–4784, 2016.
- RUHI, A.; MESSENGER, M. L.; OLDEN, J. D. Tracking the pulse of the Earth's fresh waters. **Nature Sustainability**, v. 1, n. 4, p. 198–203, 2018.
- SEIBERT, J. et al. Virtual staff gauges for crowd-based stream level observations. **Frontiers in Earth Science**, v. 7, p. 1–10, 2019.
- SILVA, R. E.; BLANCO, C. J. C.; PESSOA, F. C. L. Alternative for the regionalization of flow duration curves. **Journal of Applied Water Engineering and Research**, v. 9676, 2019.
- SWAIN, J. B.; PATRA, K. C. Streamflow estimation in ungauged catchments using regionalization techniques. **Journal of Hydrology**, v. 554, p. 420–433, 2017a.
- SWAIN, J. B.; PATRA, K. C. Streamflow estimation in ungauged catchments using regional flow duration curve: Comparative study. **Journal of Hydrologic Engineering**, v. 22, n. 7, p. 1–12, 2017b.
- TEGEGNE, G.; KIM, Y. O. Modelling ungauged catchments using the catchment runoff

response similarity. **Journal of Hydrology**, v. 564, p. 452–466, 2018.

ZAMOUM, S.; SOUAG-GAMANE, D. Monthly streamflow estimation in ungauged catchments of northern Algeria using regionalization of conceptual model parameters. **Arabian Journal of Geosciences**, v. 12, n. 11, 2019.

ZHANG, Y. et al. Can remotely sensed actual evapotranspiration facilitate hydrological prediction in ungauged regions without runoff calibration? **Water Resources Research**, v. 56, n. 1, p. 1–15, 2020.

## CAPÍTULO 1:

### **Regionalização de vazões utilizando metodologias tradicionais e geoestatística para a bacia do rio Doce, Brasil**

**RESUMO:** Com foco no apoio ao planejamento e gestão dos recursos hídricos na bacia hidrográfica do rio Doce, este estudo visou comparar o desempenho de diferentes abordagens de regionalização de vazões. Os métodos adotados foram baseados nas seguintes abordagens: interpolação linear (Interpolação Linear - IL; Interpolação Linear modificado – ILmod; Chaves et al. (2002) original – C; e Chaves et al. (2002) modificado – Cmod), regressões regionais (RR) e geoestatística (krigagens ordinária e simples). As variáveis dependentes utilizadas foram as vazões mínimas de referência ( $Q_{7,10}$ ,  $Q_{95}$  e  $Q_{90}$ ) e a vazão média de longa duração ( $Q_{mld}$ ). As variáveis independentes testadas pelo método RR foram: área de drenagem (A), precipitação média anual (Pa), precipitação média do semestre seco (Pss) e chuvoso (Psc) e a vazão equivalente ao volume precipitado menos a abstração de 750 mm (Peq750). Para aplicação dos métodos de interpolação linear foram definidos cinco casos possíveis de posição relativa da seção de interesse em relação às estações fluviométricas existentes na bacia. O melhor ajuste das equações do RR ocorreu com o modelo potencial, tendo sido identificadas três regiões hidrologicamente homogêneas (RHH). Na análise geoestatística, o modelo exponencial proporcionou o melhor ajuste aos semivariogramas experimentais. Para avaliação do desempenho de cada método foram adotados os seguintes indicadores estatísticos de qualidade: erro relativo médio (ERM), viés percentual (PBIAS), raiz do erro quadrático médio (RMSE), coeficiente de Nash-Sutcliffe (NSE) e o coeficiente de determinação ( $R^2$ ). Os resultados apontaram que os métodos tiveram desempenho similar para as vazões mínimas de referência e média de longa duração, com exceção na estimativa da  $Q_{7,10}$  na RHH III, em que se observou uma acurácia de predição bastante inferior das krigagens. Em geral, a modelagem geoestatística comprovou ser uma alternativa viável e promissora aos outros métodos de regionalização de vazões. Todas as abordagens tiveram melhor desempenho para a  $Q_{mld}$ , fato esperado em função de não se tratar de evento extremo. A adoção dos métodos geoestatísticos e daqueles baseados em interpolação linear deve estar sempre associada a uma análise bastante criteriosa de fatores como a densidade e a distribuição espacial das estações de monitoramento fluviométrico na bacia hidrográfica.

**Palavras-chave:** Hidrologia; Escassez hídrica; Planejamento e gestão dos recursos hídricos.

## 1.1 INTRODUÇÃO

Nos últimos anos, os eventos de escassez hídrica tornaram-se mais frequentes e intensos em diversas regiões do mundo. Esse fato decorre da redução dos índices pluviométricos e/ou está associado às mudanças no uso e cobertura da terra, comprometendo o atendimento aos usos múltiplos da água. Com as tendências de agravamento deste cenário, indicando secas extremas nos períodos de estiagem, inúmeros estudos têm sido desenvolvidos em favor da segurança hídrica no futuro (SILVA; PEREIRA; VIEIRA, 2020; TIEZZI et al., 2019).

Um dos maiores desafios desses estudos, no âmbito da Hidrologia, é subsidiar a gestão e o planejamento integrado dos recursos hídricos em nível de pequenas bacias hidrográficas. Dado que o monitoramento hidrológico no Brasil foi concebido priorizando a geração de energia elétrica, a existência de dados fluviométricos ficou restrita aos rios principais de uma região ou aos seus tributários com maior potencial hidráulico para fins de operação de usinas hidrelétricas (BESKOW et al., 2016). Em virtude disso, e atrelado à escassez de informações hidrológicas na maior parte das bacias hidrográficas do país, o monitoramento em bacias com pequenas áreas de drenagem é inexistente ou bastante deficitário (CECÍLIO et al., 2018; COMINI et al., 2020; MELATI; MARCUZZO, 2016; PIOL et al., 2019).

A disponibilidade de dados fluviométricos é crucial para que a gestão e o planejamento dos recursos hídricos sejam eficientes (ANKIT et al., 2018; NRUTHYA; SRINIVAS, 2015; WASEEM et al., 2016; ZAMOUM; SOUAG-GAMANE, 2019). A partir deles, algumas variáveis são obtidas para evitar os conflitos pelo uso da água e garantir a segurança hídrica em períodos de seca, como as vazões mínimas de referência. No Brasil, as principais vazões mínimas de referência adotadas para fins de outorga de uso da água são: a vazão mínima média de duração de 7 dias consecutivos e período de retorno de 10 anos ( $Q_{7,10}$ ) e as vazões mínimas associadas às permanências de 95% ( $Q_{95}$ ) e 90% do tempo ( $Q_{90}$ ).

Em decorrência da escassez de dados hidrológicos, a regionalização de vazões se torna uma ferramenta útil e muito poderosa para fins de geração de informações hidrológicas em locais onde os dados observados são escassos ou inexistentes. Essa técnica permite explorar os dados existentes numa região, visando predizer ou estimar a vazão fluvial em regiões sem monitoramento, com séries históricas muito curtas, defasadas ou repletas de falhas (LI et al., 2018; ROJAS-SERNA et al., 2016; SILVA; BLANCO; PESSOA, 2019).

Os primeiros métodos de regionalização de vazões e mais tradicionais foram aqueles baseados na regressão (ARSENAULT et al., 2019) e na proporcionalidade entre vazão e área de drenagem (YILMAZ; ONOZ, 2020).

Os métodos baseados na abordagem de regressões regionais estão entre os métodos de regionalização de vazões mais utilizados no Brasil (ARAUJO et al., 2018; CECÍLIO et al., 2018; ELESBON et al., 2015; SERRANO et al., 2020). Uma das principais limitações desta abordagem é decorrente da necessidade de dados fluviométricos em quantidade mínima na área geográfica de aplicação para a realização das regressões, o que motivou a difusão de uma segunda abordagem no país, representada pelo método da Interpolação Linear (ELETROBRÁS, 1985) e suas modificações (CHAVES et al., 2002; NOVAES et al., 2007).

Estudos recentes têm sugerido um ganho de acurácia com a Geoestatística em relação a essas abordagens (FARMER, 2016; YANG et al., 2018), além de suas vantagens intrínsecas quanto à facilidade de implementação computacional e espacialização. Porém, vale ressaltar que o desempenho de cada abordagem varia em função de diversas características regionais, tais como: clima, tamanho da área de drenagem e grau de intervenções humanas (SWAIN; PATRA, 2017), além da disponibilidade de dados fluviométricos na região de estudo.

O presente trabalho teve como objetivo comparar diferentes métodos e abordagens de regionalização de vazões, analisando as suas vantagens e limitações, mediante a estimativa ou predição das vazões mínimas de referência adotadas no Brasil para fins de outorga e da vazão média de longa duração, de modo a subsidiar a gestão e o planejamento integrado dos recursos hídricos na bacia hidrográfica do rio Doce.

## **1.2 MATERIAL E MÉTODOS**

Para aplicação das metodologias de regionalização de vazões foi necessária uma série de procedimentos relativos à preparação da base de dados. Dentre as etapas metodológicas iniciais deste trabalho estão: aquisição dos dados hidrometeorológicos; definição do período base; verificação de falhas em séries históricas para seleção das estações viáveis ao desenvolvimento do estudo e análise de tendência; cálculo das vazões mínimas de referência e média de longa duração; e obtenção das variáveis explicativas.

As metodologias foram aplicadas conforme as etapas específicas de cada abordagem, como: identificação de regiões hidrologicamente homogêneas e ajuste das equações de regressão regional (Regressões Regionais – RR); enquadramento das estações fluviométricas

em casos de posição relativa em relação às estações fluviométricas existentes na bacia e aplicação das equações apropriadas a cada caso/estação na abordagem de proporcionalidade (Interpolação Linear - IL); e ajuste de semivariogramas e predição com validação cruzada para o método geoestatístico.

### **1.2.1 Caracterização da área de estudo**

O presente trabalho foi realizado na bacia hidrográfica do rio Doce, que possui área de aproximadamente 82.427 km<sup>2</sup>, dos quais 86% pertencem ao estado de Minas Gerais e o restante ao estado do Espírito Santo. A bacia está situada na região hidrográfica do Atlântico Sudeste, Brasil, e localiza-se entre 17°30'00" e 21°30'00" de latitude sul e 39°30'00" e 44°00'00" de longitude oeste. Sua área total abrange 200 municípios mineiros e 28 municípios capixabas (CBH-DOCE, 2016a) e a população residente na bacia é de, aproximadamente, 3,6 milhões de habitantes (IBGE, 2010).

A atividade econômica da bacia tem a participação de diversos setores, destacando-se: a indústria (siderurgia, celulose e laticínios), a agropecuária e a agroindústria, a mineração (ferro, ouro, bauxita e pedras preciosas), o comércio, os serviços de apoio dos complexos industriais e a geração de energia elétrica (ECOPLAN-LUME, 2010).

A bacia é composta por seis Unidades de Gestão de Recursos Hídricos (UGRH's) em Minas Gerais, a saber: UGRH1 Piranga; UGRH2 Piracicaba; UGRH3 Santo Antônio; UGRH4 Suaçuí; UGRH5 Caratinga; UGRH6 Manhuaçu; e cinco no Espírito Santo, que são: UGRH7 Guandu; UGRH8 Santa Maria do Doce (CBH-DOCE, 2016b), e outras três unidades de planejamento e gestão que foram criadas recentemente, como Santa Joana, Barra Seca e Foz do rio Doce (BSFRD), e Pontões e Lagoas do rio Doce (AGERH, 2020).

De acordo com o Plano Integrado de Recursos Hídricos da bacia hidrográfica do rio Doce, os principais afluentes do rio Doce pela margem esquerda são os rios do Carmo, Piracicaba, Santo Antônio, Corrente Grande, Suaçuí Grande, São José e Pancas, e pela margem direita os rios Casca, Matipó, Caratinga/Cuieté, Manhuaçu, Guandu, Santa Joana e Santa Maria do Rio Doce (ECOPLAN-LUME, 2010). Na Figura 1.1 é apresentada a localização da bacia do rio Doce e os principais afluentes do seu rio principal.

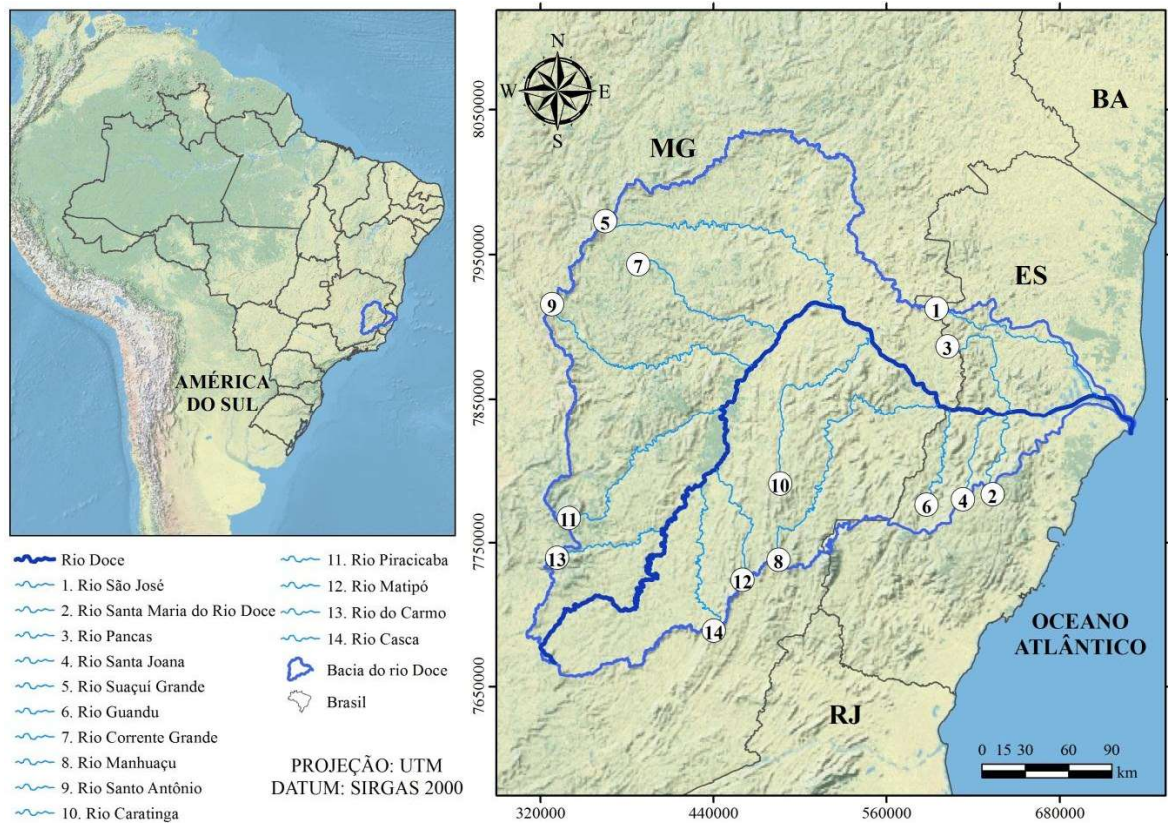


Figura 1.1. Localização da bacia do rio Doce e principais afluentes do rio principal.

### 1.2.2 Preparação da base de dados hidrometeorológica

O trabalho foi desenvolvido utilizando-se os dados de 52 estações fluviométricas e 128 estações pluviométricas pertencentes à rede hidrometeorológica da Agência Nacional de Águas (ANA). A distribuição espacial dessas estações ao longo e no entorno da área de estudo está apresentada na Figura 1.2.

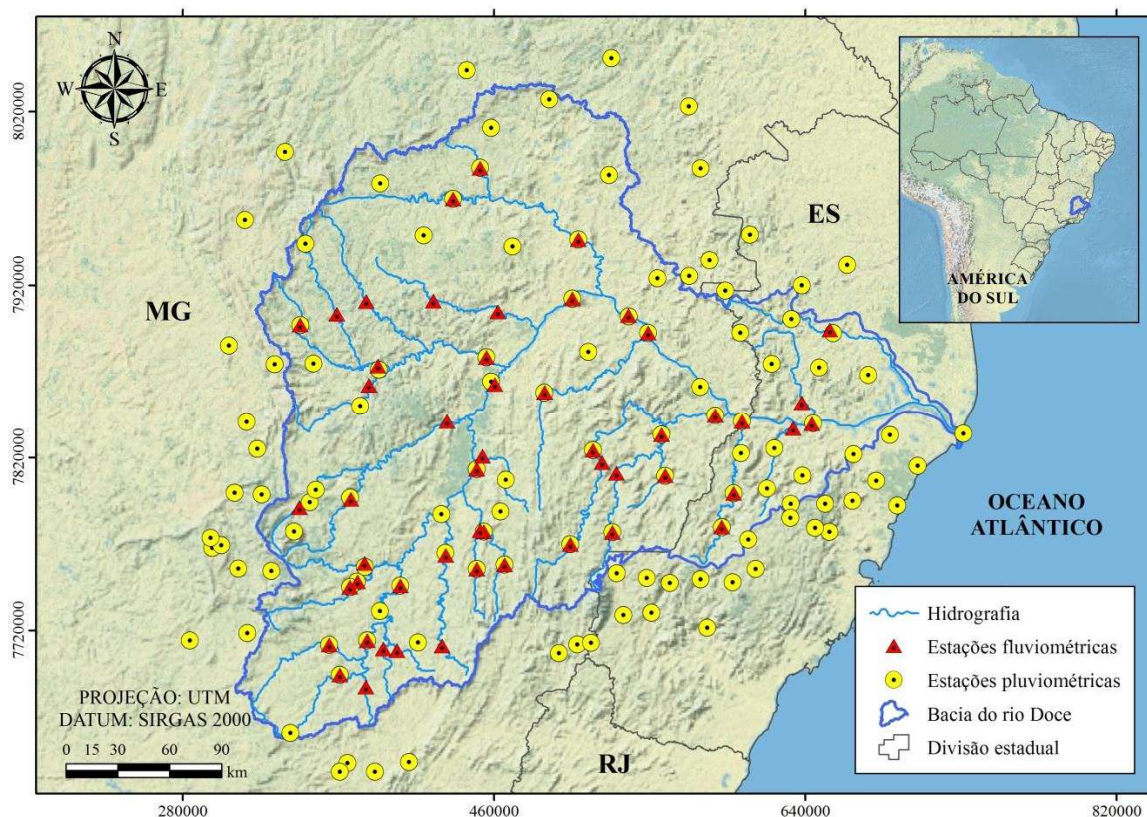


Figura 1.2. Distribuição espacial das estações utilizadas ao longo e no entorno da bacia do rio Doce.

As séries históricas utilizadas foram constituídas por dados diários de vazão e precipitação, correspondentes ao período base de outubro de 1988 a setembro de 2018, uma vez que o ano hidrológico da bacia de estudo se inicia em outubro e termina em setembro (CPRM, 2014; IGAM, 2015; MARQUES, 2010). Este período base foi adotado com o intuito de contemplar a realidade mais atual da bacia, em termos de regime das vazões fluviais.

Na etapa de seleção das estações, a série histórica de cada estação foi submetida à análise de tendência e detecção do provável ano de mudança. Foi aplicado o teste de Pettitt (PETTITT, 1979), um dos mais utilizados e com desempenho satisfatório para as variáveis de vazão e precipitação (ULIANA et al., 2015). As análises de tendência foram realizadas por meio do aplicativo XLSTAT, assumindo um nível de significância de 5%.

Adotou-se como critério a utilização de séries estacionárias com mais de 20 anos de dados e com menos de 5% de falhas em seus registros, conforme indicação de IGAM (2012). Ressalta-se que os dados brutos também foram considerados neste trabalho, após realização de análise para fins de detecção de erros e/ou inconsistências, pois o uso apenas de dados consistidos implicaria numa defasagem temporal expressiva, visto que na maior parte das

estações só se tinha dados de vazão e precipitação consistidos até o ano de 2014 e 2005, respectivamente. Informações detalhadas das estações fluviométricas e pluviométricas selecionadas estão contidas nos Apêndices A e B, respectivamente.

### 1.2.3 Vazões de referência para regionalização

As variáveis regionalizadas consistiram nas vazões mínimas de referência  $Q_{7,10}$ ,  $Q_{95}$  e  $Q_{90}$  e na vazão média de longa duração ( $Q_{mld}$ ). De posse das séries históricas de cada estação fluviométrica selecionada, procedeu-se ao cálculo dessas variáveis por meio do SisCAH 1.0 – Sistema Computacional para Análises Hidrológicas (SOUSA et al., 2009).

Para o cálculo da  $Q_{7,10}$ , que está associada a uma probabilidade de ocorrência, foi necessário verificar a distribuição de probabilidade com melhor ajuste aos dados de  $Q_7$  (vazão mínima com 7 dias de duração). Para tanto, foram considerados os seguintes modelos de distribuição de probabilidade, recorrentes em estudos de vazões mínimas: Log-normal a II e III parâmetros; Pearson tipo III, Log-Pearson tipo III e Weibull (BHATTI; KROLL; VOGEL, 2019; KESHTKAR, 2015; LANGAT; KUMAR; KOECH, 2019; NOSRATI et al., 2015).

A seleção da distribuição de densidade de probabilidade com melhor ajuste aos dados de  $Q_7$  se deu com a estimativa dos valores de vazão que apresentou a menor amplitude do intervalo de confiança dentre as estimativas obtidas em cada distribuição. Esta análise dos modelos de distribuição probabilística foi efetuada por meio do SisCAH 1.0, admitindo-se um intervalo de confiança a nível de 95%.

### 1.2.4 Processamento das variáveis independentes

A escolha das variáveis independentes ou explicativas baseou-se na literatura, priorizando a seleção daquelas mais amplamente representadas nos estudos de regionalização de vazões, como: área de drenagem (A), precipitação média do total anual (Pa), precipitação média do semestre seco (Pss) e precipitação média do semestre chuvoso (Psc) (ARAUJO et al., 2018; ELESBON et al., 2015; LOPES et al., 2016; PIOL et al., 2019).

Outra variável independente utilizada foi a vazão equivalente ao volume precipitado, subtraindo-se um fator de abstração de 750 mm ( $Peq750$ ). Essa variável demonstrou ser bem representativa da vazão em estudos de regionalização de vazões para diferentes bacias (PRUSKI et al., 2013, 2015).

A  $Peq750$  é considerada uma variável bidimensional, visto que engloba duas características da bacia (área de drenagem e precipitação). Assim, o uso da  $Peq750$  acrescenta uma unidade, em termos de graus de liberdade, à equação de regionalização, reduzindo a variância na estimativa da equação em relação às equações com múltiplas variáveis independentes (SERRANO et al., 2020). A  $Peq750$  foi calculada conforme a Equação 1.1.

$$Peq750 = \frac{(P - 750) A}{31536} \quad (1.1)$$

em que  $P$  é a precipitação média anual da área de drenagem (mm),  $A$  é a área de drenagem ( $km^2$ ) e 31536 é um fator de conversão de unidades para expressar a  $Peq750$  em  $m^3 s^{-1}$ .

A variável independente referente à área de drenagem foi calculada por intermédio do software ArcGIS 10.2 da ESRI, empregado na geração do Modelo Digital de Elevação Hidrograficamente Condicionado (MDEHC) correspondente à área de estudo. O MDEHC, com resolução de 30 metros, foi gerado a partir dos dados de radar SRTM (<http://earthexplorer.usgs.gov/>), com imposição da hidrografia produzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

Para obtenção das variáveis independentes de caráter climático os dados de precipitação média anual foram interpolados por krigagem simples, conforme sugerido por Santos (2010), empregando-se o software R (R DEVELOPMENT CORE TEAM, 2010) – pacote geoR (RIBEIRO JR; DIGGLE, 2001). Em seguida, essas variáveis foram estimadas para a área de drenagem de cada estação fluviométrica. No Apêndice C estão relacionados os valores das variáveis independentes correspondentes a cada área de drenagem considerada no estudo.

## **1.2.5 Metodologias de regionalização de vazão**

### **1.2.5.1 Método de Regressões Regionais**

O método de regressões regionais (RR) foi executado, basicamente, a partir de dois procedimentos metodológicos: i. identificação de regiões hidrologicamente homogêneas (RHH), e; ii. cálculo das equações de regressão que relacionam a variável dependente (vazão)

às variáveis independentes (características físicas e climáticas da bacia ou região hidrográfica) (ELETROBRÁS, 1985).

Com base nos possíveis modelos de regressão a serem testados, um número mínimo admissível de estações fluviométricas por RHH precisou ser estabelecido, visando assegurar que não fosse impossibilitada a aplicação das equações regionais utilizando-se, eventualmente, um número maior de variáveis independentes. O cálculo do número mínimo de estações fluviométricas por RHH ( $n$ ) baseou-se no número de graus de liberdade da regressão ( $p$ ), admitindo-se um mínimo de dois graus de liberdade para o resíduo ( $n-p-1$ ). Assim, considerando o uso de até quatro variáveis independentes nas equações regionais ( $p=4$ ), um resíduo com dois graus de liberdade ( $n-p-1=2$ ) e sabendo-se que o total da regressão é igual a  $n-1$ , foi estabelecido o número mínimo de sete estações por RHH ( $n=7$ ).

Para identificação das RHH adotou-se a análise de *cluster* (agrupamento), utilizando a metodologia k-médias. A análise de agrupamento é o método mais amplamente utilizado em estudos hidrológicos e ambientais (GOMES; BLANCO; PESSOA, 2018) e, associada à metodologia k-médias, revelou-se consistente para a área de estudo deste trabalho (CALEGARIO, 2014). Adotou-se, ainda, como critério para identificação das RHH, que as áreas de drenagem contidas em cada região homogênea fossem contíguas.

Foram testados diferentes agrupamentos, cada um a partir de uma combinação própria entre os seguintes fatores: distribuição geográfica das estações fluviométricas, valores de vazão específica e de precipitação média anual. Em seguida, empregou-se o método da conveniência geográfica como suporte à análise de agrupamento. Para garantir a consistência dos agrupamentos, verificou-se a qualidade do ajuste das equações de regressão obtidas para cada RHH com base no coeficiente de determinação ajustado ( $R^2a$ ), no erro padrão da estimativa (EPE) e na amplitude dos erros percentuais ( $\Delta ER$ ).

As equações de regionalização foram estimadas por meio do SisCoRV – Sistema Computacional para Regionalização de Vazões (SOUSA, 2009). Para ajuste das equações de regionalização foram considerados os modelos de regressão linear, potencial, exponencial, logarítmico e recíproco.

As equações de regionalização selecionadas foram aquelas que melhor reuniram os seguintes critérios de qualidade do ajuste: simplicidade da equação; maior valor de  $R^2a$ , desde que acima de 0,7 (ARAUJO et al., 2018; ELESBON et al., 2015); e menor resíduo (valores mais baixos de EPE e  $\Delta ER$ ). Estes critérios de análise, por conseguinte, serviram como métricas de aferição da consistência estatística das equações geradas para cada RHH.

### 1.2.5.2 Interpolação Linear (ELETROBRÁS, 1985)

Esta metodologia, descrita em Eletrobrás (1985b), está fundamentada na relação de proporcionalidade entre o tamanho da área da bacia e a magnitude das vazões. Pressupõe que, considerando isoladamente a influência da área de uma bacia hidrográfica, a vazão do rio principal será tanto mais elevada quanto maior for a sua área de drenagem.

O conjunto de procedimentos para realização da Interpolação Linear (IL) envolveu três etapas básicas, que foram: verificação da existência de estações de monitoramento fluviométrico (estações de apoio) nas proximidades da seção de interesse; cálculo da área de drenagem das estações de apoio e da seção de interesse, e; cálculo da vazão relacionada à área da seção de interesse com base nas vazões registradas para as áreas das estações de apoio. Porém, o cálculo da vazão na seção de interesse depende da posição relativa entre essa seção e as estações de apoio. Deste modo, o enquadramento da seção de interesse em um dos casos possíveis de posição relativa em relação às estações de apoio é uma etapa imprescindível desta abordagem de regionalização, e precede o cálculo da vazão na seção de interesse.

Neste trabalho, foi possível definir cinco casos de enquadramento. Os casos 1 e 2 (Figuras 1.3 e 1.4) referem-se às situações em que a seção de interesse situa-se numa posição a montante (Equação 1.2) ou a jusante (Equação 1.3) da estação de apoio, respectivamente.

$$Q_{si} = \left( \frac{Q_j}{A_j} \right) \times A_{si} \quad (1.2)$$

$$Q_{si} = \left( \frac{Q_m}{A_m} \right) \times A_{si} \quad (1.3)$$

em que  $Q_{si}$  é a vazão a ser calculada na seção de interesse;  $Q_j$  e  $Q_m$  são as vazões das estações de jusante e montante, respectivamente;  $A_j$  e  $A_m$  são as áreas de drenagem das estações de jusante e montante, respectivamente; e  $A_{si}$  é a área de drenagem da seção de interesse.



Figura 1.3. Caso 1 de posição relativa da seção de interesse em relação à estação de apoio.



Figura 1.4. Caso 2 de posição relativa da seção de interesse em relação à estação de apoio.

O caso 3 corresponde às situações em que a seção de interesse localiza-se entre duas estações de apoio (Equação 1.4), conforme ilustrado na Figura 1.5.

$$Q_{si} = Q_m + \left[ \left( \frac{Q_j - Q_m}{A_j - A_m} \right) \times (A_{si} - A_m) \right] \quad (1.4)$$

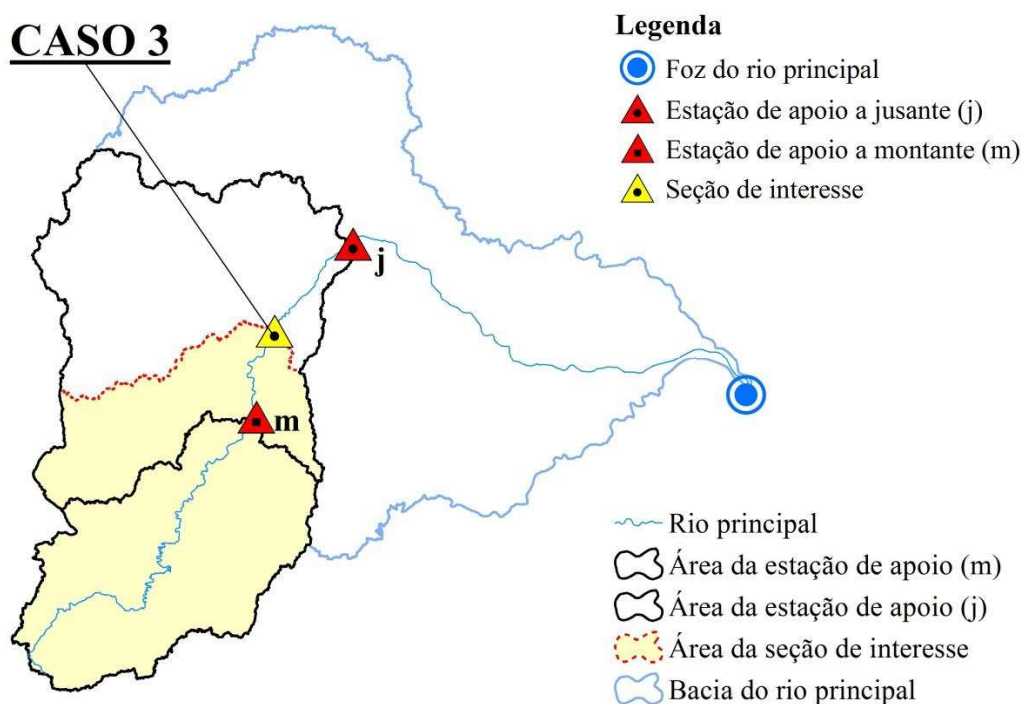


Figura 1.5. Caso 3 de posição relativa da seção de interesse em relação às estações de apoio.

No caso 4 (Figura 1.6), a seção de interesse pertence a um trecho fluvial de ordem  $i$  cuja foz situa-se entre duas estações de apoio alocadas num curso de ordem superior ( $i+1$ ). Para o cálculo da vazão na seção de interesse foi considerada uma combinação dos casos 3 e 1. Primeiro, calculou-se a vazão na confluência do trecho de ordem  $i$  com curso de ordem  $i+1$  (Equação 1.4). Em seguida, com base nesta vazão, calculou-se a vazão da seção de interesse (Equação 1.2).

Assim como no caso 4, o caso 5 (Figura 1.7) também constituiu uma combinação de duas situações (casos 1 e 2). Porém, o caso 5 baseia-se na utilização de apenas uma estação de apoio. Neste caso, a seção de interesse está situada num trecho fluvial (ordem  $i$ ) cuja confluência com um curso de ordem superior ( $i+1$ ) localiza-se a jusante da estação de apoio que, por sua vez, pertence ao curso de ordem  $i+1$ .



Figura 1.6. Caso 4 de posição relativa entre a seção de interesse e suas estações de apoio.



Figura 1.7. Caso 5 de posição relativa entre a seção de interesse e suas estações de apoio.

Por se tratar de uma metodologia interpolativa, a IL dispensa a identificação de RHH e, por isso, esta metodologia é isenta da exigência estatística relacionada ao número mínimo de graus de liberdade para o estabelecimento das equações de regressão, a qual restringe a aplicação do método de regressões regionais. Logo, por superar esta limitação do RR, a abordagem de IL é comumente utilizada para bacias com poucas estações de monitoramento.

### 1.2.5.3 Interpolação Linear modificada (NOVAES et al., 2007)

Esta metodologia modifica a IL por considerar o papel da precipitação no processo físico de formação das vazões em uma bacia. Assim, a Interpolação Linear modificada (ILmod) acrescenta a precipitação média, enquanto variável climática, nas equações propostas pela IL (NOVAES et al., 2007).

Na aplicação da ILmod, as equações utilizadas por IL e que permitem o cálculo da vazão na seção de interesse para todos os casos de posição relativa (Equações 1.2, 1.3 e 1.4), foram substituídas pelas Equações 1.5, 1.6 e 1.7, respectivamente.

$$Q_{si} = \left( \frac{Q_j}{A_j \times P_j} \right) \times (A_{si} \times P_{si}) \quad (1.5)$$

$$Q_{si} = \left( \frac{Q_m}{A_m \times P_m} \right) \times (A_{si} \times P_{si}) \quad (1.6)$$

$$Q_{si} = Q_m + \left[ \left( \frac{Q_j - Q_m}{A_j \times P_j - A_m \times P_m} \right) \times (A_{si} \times P_{si} - A_m \times P_m) \right] \quad (1.7)$$

em que  $P_{si}$ ,  $P_m$  e  $P_j$  são as precipitações médias anuais das áreas de drenagem referentes à seção de interesse e às estações de apoio de montante e jusante, respectivamente.

### 1.2.5.4 Chaves (CHAVES et al., 2002)

Chaves et al. (2002) propuseram que a metodologia IL considerasse a distância entre as estações de apoio e a seção de interesse (caso 3) e entre as estações de apoio e a confluência

(caso 4). Desta forma, a metodologia atribui um peso a cada estação de apoio, garantindo que a mais próxima da seção de interesse tenha uma influência maior na estimativa da vazão.

Na Equação 1.8 é demonstrado o cálculo, para o caso 3, da vazão na seção de interesse por meio desta metodologia, enquanto as Equações 1.9 e 1.10 correspondem à obtenção dos pesos relativos à distância entre as estações de apoio e a seção de interesse. As referidas equações também são requeridas pelo caso 4, na estimativa da vazão na foz do curso de ordem  $i$ , antes de ser aplicada a equação referente à estimativa na seção de interesse (caso 1).

$$Q_{si} = A_{si} \times \left[ \left( p_m \frac{Q_m}{A_m} \right) \times \left( p_j \frac{Q_j}{A_j} \right) \right] \quad (1.8)$$

$$p_m = \left( \frac{d_j}{d_m + d_j} \right) \quad (1.9)$$

$$p_j = \left( \frac{d_m}{d_j + d_m} \right) \quad (1.10)$$

em que  $d_m$  é a distância entre a estação de apoio a montante e a seção de interesse,  $d_j$  é a distância entre a estação de apoio a jusante e a seção de interesse, enquanto  $p_m$  e  $p_j$  são os pesos relativos à estação de apoio a montante a estação de apoio a jusante, respectivamente.

Vale ressaltar que para os demais casos de enquadramento (casos 1, 2 e 5), a aplicação desta metodologia é idêntica àquela adotada na IL, pois não pressupõem que o local de interesse da estimativa esteja situado entre duas estações fluviométricas de apoio.

#### **1.2.5.5 Metodologia de Chaves modificada (NOVAES et al., 2007)**

A metodologia de Chaves modificada (Cmod), proposta por Novaes et al. (2007), insere a precipitação média nos cálculos das metodologias IL e C. Assim, as Equações 1.5 e 1.6 foram aplicadas para os casos 1, 2 e 5. Para os casos 3 e 4, a Equação 1.8 de C foi modificada com o acréscimo da variável correspondente à precipitação média anual, sendo expressa pela Equação 1.11.

$$Q_{si} = (A_{si} \times P_{si}) \times \left[ \left( \frac{p_m \times Q_m}{A_m \times P_m} \right) \times \left( \frac{p_j \times Q_j}{A_j \times P_j} \right) \right] \quad (1.11)$$

Portanto, as metodologias baseadas na interpolação linear (IL, ILmod, C e Cmod) foram aplicadas após o devido enquadramento das estações fluviométricas quanto às suas posições relativas à seção de interesse. Foram identificadas 12 estações no caso 1, 10 no caso 2, 12 no caso 3, 13 no caso 4 e 5 no caso 5. Na Figura 1.6 pode-se observar como ficou definido este enquadramento, considerando o posicionamento de cada estação fluviométrica na bacia.

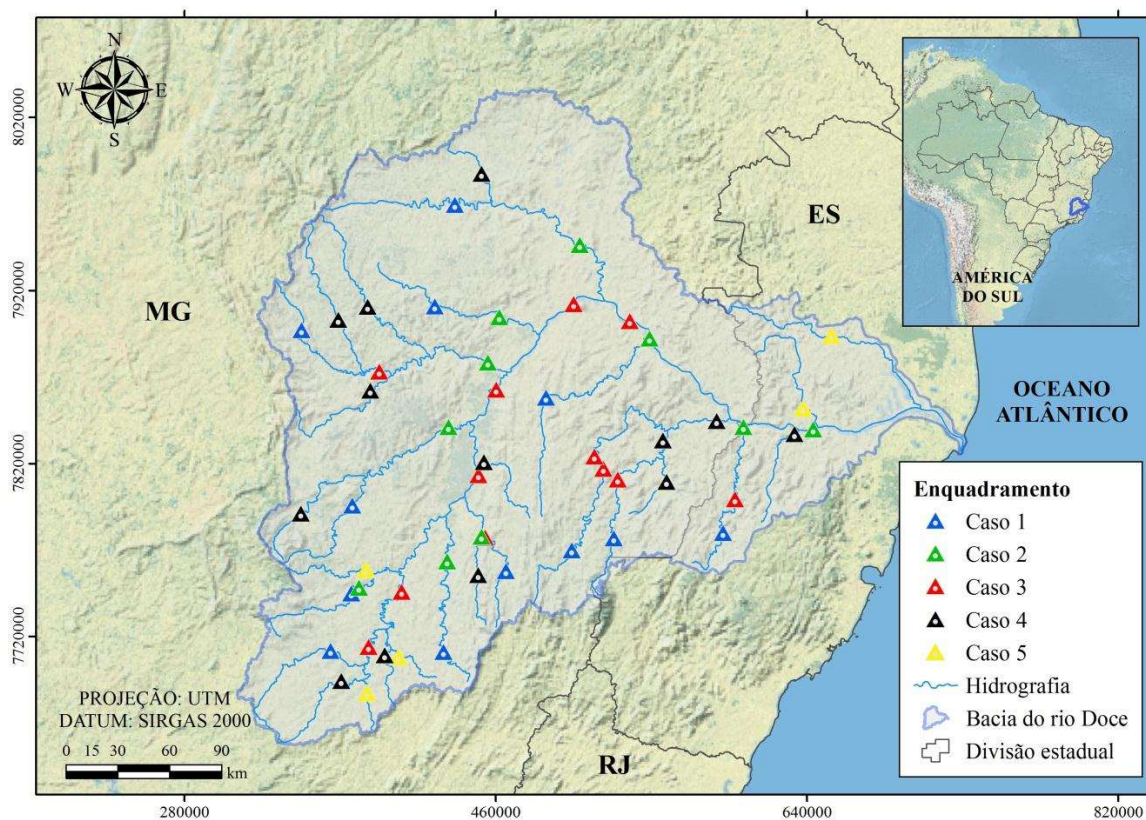


Figura 1.6. Enquadramento das estações fluviométricas conforme posição relativa à seção de interesse.

Para fins de validação das metodologias, a estimativa das variáveis dependentes se deu para cada seção de monitoramento, sendo omitida a presença da sua respectiva estação fluviométrica. Assim, para cada seção de monitoramento foram obtidos os pares de valores simulados e observados, que permitiram a avaliação do desempenho das metodologias.

## 1.2.6 Métodos de abordagem geoestatística

Dois métodos de krigagem foram adotados neste trabalho: krigagem linear simples (KS) e krigagem linear ordinária (KO). O primeiro método foi adotado por ser clássico e teoricamente mais acurado do que seus pares (krigagem ordinária e universal), conforme Santos (2010). Entretanto, a KO ainda é o preditor geoestatístico mais comumente utilizado em estudos de regionalização de vazões e, por isso, optou-se pela sua avaliação neste trabalho.

Assim como a abordagem de interpolação linear, os métodos geoestatísticos dispensam a necessidade de identificação de regiões homogêneas, procedimento que é inerente ao método RR. Neste sentido, vale destacar que as modelagens geoestatísticas foram processadas utilizando todas as 52 estações fluviométricas selecionadas para este estudo.

### 1.2.6.1 Análise exploratória dos dados

A análise estatística descritiva foi a primeira etapa que precedeu o processamento dos preditores geoestatísticos. De posse dos valores de vazão mínima e média específica ( $L s^{-1} Km^{-2}$ ), referentes a cada variável de interesse, e das coordenadas geográficas de suas respectivas seções de monitoramento, calculou-se a distância máxima entre todas as estações fluviométricas. Assim, foi possível obter a provável melhor distância para geração do semivariograma experimental.

Em seguida, estimou-se a variância dos dados de vazão específica. O valor estimado foi empregado na modelagem do semivariograma, constituindo um indicador do valor teórico do patamar. Ainda, foram analisados os seguintes descritores estatísticos dos dados: média e mediana (medidas de tendência central); dispersão e detecção de *outliers* (*boxplot*); normalidade (teste de Shapiro-Wilk, a um nível de significância de 5%); e tendência em função das abcissas, das ordenadas, distribuição em quartis e seu comportamento gráfico.

### 1.2.6.2 Modelagem variográfica

A partir de múltiplas subdivisões da distância máxima foram gerados e analisados diversos semivariogramas experimentais. Foi selecionado o semivariograma experimental que melhor expressou as seguintes características gráficas e numéricas: alinhamento dos pontos, menor efeito pepita e estabilização na variância dos dados (VIEIRA, 2000). Sobre a estrutura

de dependência espacial das variáveis dependentes, considerou-se que o comportamento dos dados de vazão segue um padrão isotrópico, conforme assumido por Elesbon et al. (2015).

Posteriormente, obteve-se o modelo teórico de semivariograma de melhor ajuste. Os modelos teóricos de semivariograma foram ajustados pelos métodos dos mínimos quadrados ordinários (OLS), mínimos quadrados ponderados (WLS), máxima verossimilhança (ML) e máxima verossimilhança restrita (REML). Para aferição da qualidade do ajuste foi efetuada a autovalidação “*Leave-One-Out Cross Validation (LOOCV)*”, sendo o melhor ajuste aquele com média nula e variância unitária em relação aos erros padronizados (VIEIRA, 2000).

Outra informação importante para análise do ajuste foi a normalidade dos resíduos, verificada por meio do teste de Shapiro-Wilk, a um nível de significância de 5%. Ainda, visando avaliar o desempenho da modelagem, foi estabelecida uma regressão linear simples entre valores preditos e observados. Admitiu-se que o modelo era tão confiável quanto mais próximo o intercepto de zero ( $\beta_0=0$ ) e o coeficiente angular da unidade ( $\beta_1=1$ ).

### 1.2.6.3 Krigagem Simples

A principal característica da krigagem simples (KS) é que a média da variável deve, necessariamente, ser conhecida. Assim, o preditor empregado no processamento da KS pode ser descrito pela Equação 1.12 (OLEA, 1999).

$$Z_{KS}^*(x_0) = m + \sum_{i=1}^{n(u)} \lambda_i [Z(x_i) - m] \quad (1.12)$$

em que  $Z_{KS}^*(x_0)$  é o valor predito na seção de interesse  $x_0$ ;  $Z(x_i)$  é o valor observado da variável num dado local  $x_i$ ;  $\lambda_i$  é o peso atribuído aos valores próximos da seção de interesse;  $m$  é a média constante em toda a área, e;  $n(u)$  é o número de observações, ou estações de apoio. Tal predição é não enviesada, pois  $E [Z(x_i) - m] = 0$  e  $E [Z^*(x_0)] = m = E [Z(x_0)]$ .

### 1.2.6.4 Krigagem Ordinária

A krigagem ordinária (KO) também assume que a média da variável é constante em toda a área. Porém, no caso da KO, considera-se que essa média é desconhecida. Logo, o

processamento da KO utilizou o preditor da Equação 1.13 (YAMAMOTO; LANDIM, 2013). Vale ressaltar que a KO não tem viés em virtude da restrição imposta pela Equação 1.14 sobre a Equação 1.13 (JOURNEL; HUIJBREGTS, 1978).

$$Z_{KO}^*(x_0) = \sum_{i=1}^{n(u)} \lambda_i [Z(x_i)] \quad (1.13)$$

$$\sum_{i=1}^{n(u)} \lambda_i = 1 \quad (1.14)$$

A análise de semivariograma e processamento das predições por KS e KO foram realizados por meio do software R (R DEVELOPMENT CORE TEAM, 2010) – pacote geoR (RIBEIRO JR; DIGGLE, 2001).

### 1.2.7 Análise comparativa de desempenho das metodologias de regionalização

Os indicadores estatísticos de erro utilizados para análise e comparação da qualidade e acurácia dos métodos na regionalização das vazões foram os seguintes: a) erro relativo médio (ERM) entre os valores observados da série histórica e os regionalizados (Equação 1.15); b) percentual de tendência ou viés (PBIAS), conforme Equação 1.16; c) raiz do erro quadrático médio (RMSE) pela Equação 1.17; d) coeficiente de determinação ( $R^2$ ); e) o coeficiente de eficiência de Nash-Sutcliffe (NSE) pela Equação 1.18 (NASH; SUTCLIFFE, 1970).

$$ERM = 100 \times \left| \frac{\sum(Qobs - Qest)}{\sum(Qobs)} \right| \quad (1.15)$$

$$PBIAS = 100 \times \left[ \frac{\sum(Qest - Qobs)}{\sum(Qobs)} \right] \quad (1.16)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \times \sum (Q_{\text{obs}} - Q_{\text{est}})^2} \quad (1.17)$$

$$\text{NSE} = 100 \times \left[ 1 - \frac{\sum (Q_{\text{obs}} - Q_{\text{est}})^2}{\sum (Q_{\text{obs}} - \overline{Q_{\text{obs}}})^2} \right] \quad (1.18)$$

em que  $n$  é o número de estações fluviométricas;  $Q_{\text{obs}}$  é a vazão observada ( $\text{m}^3 \text{s}^{-1}$ );  $Q_{\text{est}}$  é a vazão regionalizada ( $\text{m}^3 \text{s}^{-1}$ ); e  $\overline{Q_{\text{obs}}}$  é a média das vazões observadas ( $\text{m}^3 \text{s}^{-1}$ ).

O método mais acurado e, portanto, com melhor desempenho, foi aquele que melhor atendeu, em conjunto, aos critérios de classificação dos indicadores estatísticos utilizados, a saber: valores mais baixos de ERM e RMSE, valor de PBIAS mais próximo de zero, valor mais próximo de 100 para NSE e  $R^2$  mais próximo de 1.

As análises comparativas foram realizadas, separadamente, no domínio de cada região homogênea identificada com o método RR. Assim, tais regiões padronizaram os domínios geográficos para avaliação do desempenho dos métodos de regionalização de vazões.

## 1.3 RESULTADOS E DISCUSSÃO

### 1.3.1 Método de Regressões Regionais

#### 1.3.1.1 Identificação de Regiões Hidrológicamente Homogêneas (RHH)

Após as diferentes tentativas de agrupamento, verificou-se que a distribuição geográfica das estações fluviométricas em conjunto com seus respectivos valores de vazão específica foi a melhor combinação testada, a partir da qual a bacia de estudo foi individualizada em três regiões homogêneas. Para todas as vazões mínimas de referência e média de longa duração, a análise de agrupamento identificou as mesmas regiões homogêneas.

Após a delimitação das regiões com a análise de agrupamento, complementada pela aplicação do método de conveniência geográfica, foram contabilizadas 22 estações fluviométricas pertencentes à primeira RHH (RHH I), enquanto a segunda (RHH II) e a

terceira região (RHH III) abrangeram 13 e 17 estações, respectivamente. Na Figura 1.8 pode ser observada a divisão da bacia de estudo nas referidas regiões homogêneas.

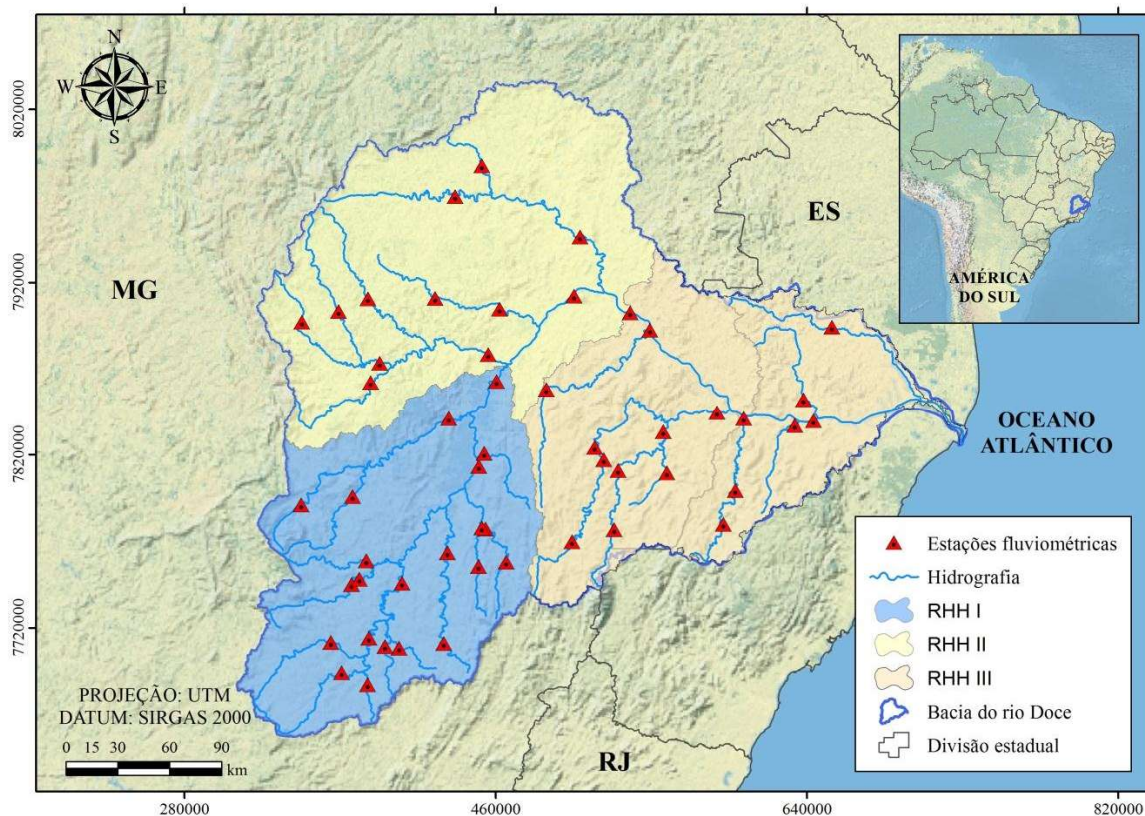


Figura 1.8. Delimitação das regiões hidrologicamente homogêneas (RHH I, II e III)

Os índices estatísticos para avaliação da confiabilidade do processo de identificação das RHH serão relacionados adiante, junto com a análise das equações regionais de regressão obtidas pelo método RR, uma vez que a qualidade do ajuste estatístico dessas equações está diretamente atrelada à consistência da delimitação das respectivas regiões homogêneas.

### 1.3.1.2 Obtenção das equações regionais de regressão

A variável independente mais representativa das vazões de referência foi a área de drenagem (A). Este resultado é similar ao encontrado por outros autores em estudos de regionalização de vazões (ARAÚJO et al., 2018; ELESBON et al., 2015; LOPES et al., 2016). Ainda, em conjunto com a área de drenagem, outra variável explicativa que contribuiu com a melhoria da qualidade das equações foi a precipitação total anual média ( $P_a$ ), tal como observado por Lopes et al. (2017) na regionalização de vazões mínimas.

Na Tabela 1.1 são apresentadas as melhores equações regionais obtidas nas três RHH para cada vazão considerada no estudo, bem como seus indicadores de ajuste estatístico e o intervalo de validade de área de drenagem para aplicação na bacia do rio Doce.

Tabela 1.1 – Equações de regionalização para a estimativa das vazões de referência, em  $\text{m}^3 \text{s}^{-1}$ , nas três regiões hidrologicamente homogêneas da bacia do rio Doce

RHH	Equação de regionalização*	R <sup>2</sup> a	EPE	$\Delta$ ER (%)	Intervalo de validade (Km <sup>2</sup> )
I	$Q_{7,10} = (3,3851 \times 10^{-13}) A^{0,9635} Pa^{3,2327}$	0,95	0,29	94,17	272,85 a 24209,52
	$Q_{95} = (3,4627 \times 10^{-13}) A^{0,9632} Pa^{3,2656}$	0,96	0,27	96,41	
	$Q_{90} = 0,3111 Peq750^{0,9741}$	0,96	0,24	99,44	
	$Q_{mld} = 0,8980 Peq750^{0,9620}$	0,99	0,12	51,07	
II	$Q_{7,10} = (6,0533 \times 10^{-21}) A^{1,0948} Psc^{5,6514}$	0,96	0,37	124,56	302,76 a 55042,15
	$Q_{95} = (8,3425 \times 10^{-20}) A^{1,0463} Psc^{5,3849}$	0,96	0,32	111,80	
	$Q_{90} = (7,0846 \times 10^{-17}) A^{1,0463} Psc^{4,4566}$	0,98	0,25	85,90	
	$Q_{mld} = (2,6771 \times 10^{-15}) A^{0,9701} Psc^{4,2001}$	0,99	0,14	43,99	
III	$Q_{7,10} = (3,7406 \times 10^{-27}) A^{1,1244} Psc^{7,7305}$	0,89	0,58	303,79	386,04 a 76382,10
	$Q_{95} = (3,8256 \times 10^{-20}) A^{1,0895} Pa^{5,3697}$	0,96	0,28	117,53	
	$Q_{90} = (6,5175 \times 10^{-17}) A^{1,0563} Pa^{4,3972}$	0,96	0,25	84,58	
	$Q_{mld} = (3,5028 \times 10^{-9}) A^{0,9480} Pa^{2,1851}$	0,99	0,14	50,04	

\*Variáveis dependentes em  $\text{m}^3 \text{s}^{-1}$ ; A em  $\text{km}^2$ ; Peq750 em  $\text{m}^3 \text{s}^{-1}$ ; Pa e Psc em mm.

O modelo de regressão potencial foi aquele que se ajustou melhor em todas as regiões homogêneas e para todas as vazões de referência, tal como observado em trabalhos similares (BAZZO et al., 2017; IGAM, 2012).

Com base nas melhores equações obtidas, pode-se constatar que, como esperado, o método RR foi mais bem sucedido no ajuste das equações que se referem à estimativa da  $Q_{mld}$ , independentemente da região homogênea. Além do R<sup>2</sup>a mais elevado para esta vazão, as equações para a  $Q_{mld}$  apresentaram os menores valores de EPE e  $\Delta$ ER em todas as RHH. Resultados similares foram obtidos por Barbosa et al. (2005). Ainda, vale destacar a ocorrência do bom ajuste estatístico de todas as equações regionais para a estimativa das vazões mínimas, reforçando sua aplicabilidade para as regiões homogêneas identificadas.

Dentre os possíveis fatores que explicam o melhor ajuste das equações para as regiões homogêneas I e II está a maior heterogeneidade climática ao longo da área da RHH III,

implicando em elevados valores de erro e a menor aderência no ajuste das equações, apontada pelos valores mais baixos de  $R^2$ .

As equações regionais confirmam a relação praticamente linear entre a vazão e a área de drenagem, em virtude dos expoentes da área serem ao redor da unidade, e a baixa amplitude de variação relativa entre estas variáveis. Esta amplitude de variação indica a capacidade de variação da vazão com o aumento da área. Quanto mais próxima da unidade, menor é esta amplitude de variação. Logo, com base nas equações obtidas, a baixa amplitude de variação é apontada pelo valor do expoente da área de drenagem, assim como observado por outros autores (ARAUJO et al., 2018; CECÍLIO et al., 2018; SILVA et al., 2017).

Analisando-se apenas as equações regionais para  $Q_{7,10}$ , o expoente das variáveis explicativas aponta que o excedente hídrico é, em média, maior na RHH I e menor nas RHH II e III. As condições climáticas da RHH I são bastante discrepantes em relação às RHH II e III, que apresentam baixos índices pluviométricos e deficiência hídrica mais pronunciada (LOUZADA; XAVIER; PEZZOPANE, 2018). Com isso, o ajuste das equações nas RHH II e III requer que o expoente relativo à área de drenagem e à precipitação tenham maiores magnitudes, com vistas a compensar a dificuldade nestas regiões de se sustentar o deflúvio superficial ao longo do ano, como discutido por Silva et al. (2017). Neste sentido, enquanto a equação da  $Q_{7,10}$  na RHH I foi ajustada utilizando a precipitação média anual ( $P_a$ ), o melhor ajuste desta equação nas RHH II e III foi obtido com a utilização da precipitação média do período chuvoso ( $P_{sc}$ ). Esta distinção de variáveis climáticas para o melhor ajuste das equações de  $Q_{7,10}$  aponta a desigualdade de regime pluviométrico entre as regiões homogêneas e respalda a compensação estatística às equações das RHH II e III com expoentes mais altos para a área de drenagem.

Ainda, outro forte indicativo das melhores condições gerais na RHH I para a produção de deflúvio superficial foi o valor do coeficiente linear ( $a$ ) nas equações de regressão, considerando o modelo potencial  $Q = a X^b$ . Quanto maior o valor deste coeficiente, maior será a capacidade de produção de vazão na região. Assim, analisando-se este coeficiente para a  $Q_{7,10}$ , pode-se verificar a maior capacidade de produção de vazão da RHH I em comparação às RHH II e III (SILVA et al., 2017).

### 1.3.2 Metodologias baseadas na abordagem de interpolação linear

Todas as metodologias desta abordagem (IL, ILmod, C e Cmod) obtiveram estimativas mais acuradas para as seções alocadas entre duas estações de apoio (caso 3) e, em seguida, para aquelas referentes à posição relativa do caso 1. Resultados similares foram obtidos por Cecílio et al. (2018), que destacaram o fato do caso 1 ter apresentado melhores estimativas do que o caso 2, apesar de serem calculados de forma similar, com a diferença que o primeiro se baseia na vazão de jusante. Na Tabela 1.2 podem-se verificar estes resultados, a variação dos valores de ERM entre os cinco casos considerados no presente trabalho e o desempenho de cada metodologia na estimativa das vazões mínimas de referência e média de longa duração.

Tabela 1.2. Erro relativo médio (%) na estimativa das vazões mínimas de referência e média de longa duração referente aos casos de posição relativa, segundo o enquadramento requerido pelas metodologias baseadas na interpolação linear

Vazões de referência	Caso	ERM (%) por metodologia baseada na interpolação linear			
		IL	ILmod	C	Cmod
Q <sub>7,10</sub>	1	38,76	39,43	38,76	39,43
	2	43,03	38,90	43,03	38,90
	3	12,12	11,37	9,40	9,72
	4	108,45	93,43	115,66	97,65
	5	491,54	429,58	491,54	429,58
Q <sub>95</sub>	1	27,67	26,27	27,67	26,27
	2	38,64	34,61	38,64	34,61
	3	8,45	8,02	5,82	6,01
	4	50,15	41,51	58,21	45,94
	5	75,64	60,12	75,64	60,12
Q <sub>90</sub>	1	21,91	20,44	21,91	20,44
	2	30,31	26,74	30,31	26,74
	3	8,79	8,39	5,33	5,49
	4	39,43	31,80	46,30	35,32
	5	48,74	37,11	48,74	37,11
Q <sub>mld</sub>	1	14,10	11,53	14,10	11,53
	2	22,48	17,23	22,48	17,23
	3	6,52	5,21	4,29	3,65
	4	19,52	15,14	22,47	15,48
	5	11,89	14,33	11,89	14,33

Os maiores valores de ERM correspondentes ao caso 5 são devidos, em grande parte, aos elevados erros relativos verificados nas seções fluviais das estações 56995500 e 56998400 situadas na RHH III. Constatou-se que os erros nestas seções estão associados à grande discrepância de dimensão das duas áreas de drenagem envolvidas no cálculo das vazões

(ALMEIDA, 2010; LI et al., 2019; SILVA; MARQUES; LEMOS, 2009), gerando superestimativas de 767,40 a 1577,30%.

O caso 3 considera a utilização de maior número de estações de apoio e, por isso, apresentou melhores resultados em comparação aos demais. Os casos 1 e 2 são aplicados de forma semelhante e se distinguiram em acurácia, principalmente, devido às diferenças de área de drenagem entre as seções de interesse e as respectivas seções de apoio. Observou-se que, em média, essas diferenças de área, para o caso 2, foram equivalentes ao dobro daquela correspondente ao caso 1, aproximadamente, o que explica o desempenho inferior do caso 2.

Os casos 4 e 5 apresentaram os maiores valores de ERM na estimativa das vazões mínimas. Como estes casos se baseiam numa combinação de outros casos, os seus desempenhos estão sujeitos a um acúmulo de erros, acarretando a perda de acurácia. O caso 5 foi bastante problemático na estimativa da  $Q_{7,10}$  em razão da fragilidade desta abordagem quando a seção de interesse e sua seção de apoio possuem tamanho de área e condições hidrometeorológicas muito distintas. Com isso, a pior estimativa do caso 5 ocorreu na seção da estação 56998400, cuja área é cerca de 74 vezes menor do que a área da estação de apoio.

Em geral, como retratado pela Tabela 1.2, os erros da estimativa pelas metodologias utilizadas tendem a ser menores à medida que se eleva a magnitude dos valores da vazão de interesse. Neste sentido, tal como obtido e discutido por Novaes et al. (2007) e Cecílio et al. (2018), o ERM diminuiu à medida que a estimativa se referia às vazões menos restritivas.

O desempenho da ILmod foi superior à IL em praticamente todas as situações, sendo inferior apenas à estimativa da  $Q_{7,10}$  (caso 1) e  $Q_{mld}$  (caso 5). Para o caso 3, a metodologia C obteve a maior acurácia na estimativa de todas as vazões mínimas de referência, mais foi menos acurada que Cmod para a  $Q_{mld}$ . A acurácia superior das metodologias de C e Cmod no caso 3 é decorrente da ponderação em relação às distâncias que elas consideram na aplicação.

Na Figura 1.9 é apresentado o enquadramento dos trechos fluviais pertencentes à rede de drenagem monitorada do presente estudo em seus devidos casos de posição relativa, considerando a distribuição espacial e o respectivo posicionamento das estações fluviométricas selecionadas para este trabalho ao longo da bacia do rio Doce.

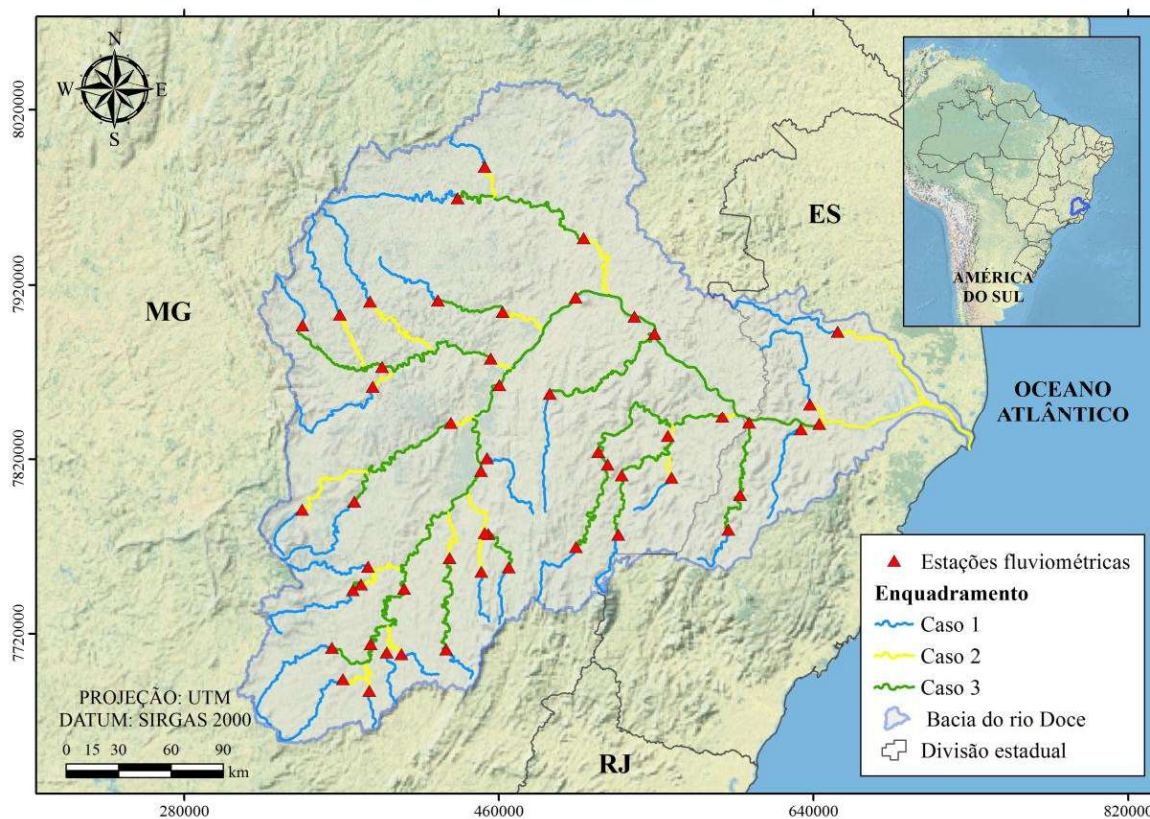


Figura 1.9. Enquadramento dos trechos fluviais conforme o posicionamento das estações fluviométricas ao longo da bacia do rio Doce.

Para fins de proposição em suporte à gestão e planejamento integrado dos recursos hídricos na bacia do rio Doce, e seguindo os devidos procedimentos metodológicos, pode ser observado na Figura 1.9 que o enquadramento dos trechos fluviais dotados de monitoramento de vazões compreendeu os casos 1, 2 e 3, considerando o posicionamento relativo entre os trechos de interesse à estimativa e as seções com estações de monitoramento fluviométrico.

### 1.3.3 Modelagem geoestatística

O melhor ajuste dos semivariogramas experimentais ocorreu como modelo exponencial. Outros autores, em estudos de predição de vazão, também selecionaram este modelo (ELESBON et al., 2014; LACHANCE-CLOUTIER; TURCOTTE; CYR, 2017). O método que contribuiu para o melhor ajuste foi a Máxima Verossimilhança (ML). Com o emprego da autovalidação LOOCV detectou-se que os semivariogramas ajustados foram superiores, em termos de qualidade do ajuste, visto que obtiveram os valores mais próximos de 0 e 1 para a

média (EM) e desvio padrão dos erros padronizados (SD), respectivamente, conforme recomendado por Vieira (2000).

Na Figura 1.10 são apresentados os semivariogramas ajustados que refletiram os melhores indicadores de ajuste e seus respectivos semivariogramas experimentais.

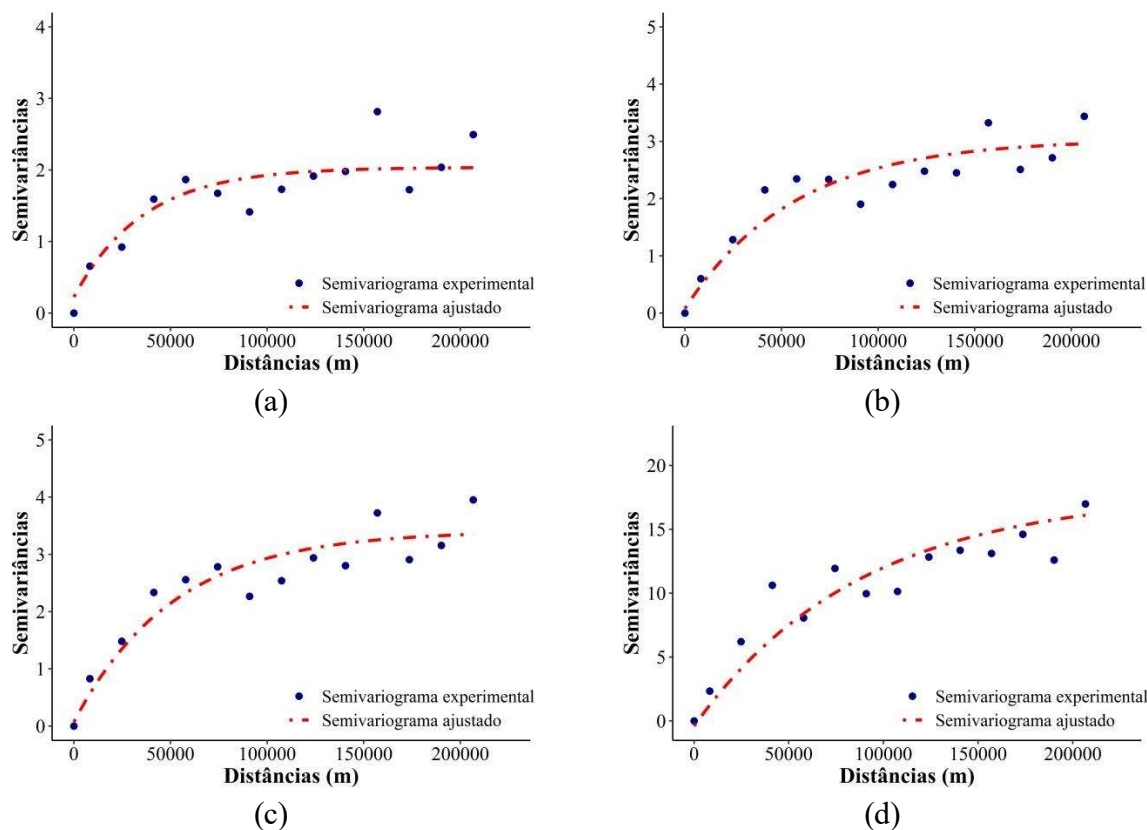


Figura 1.10. Semivariogramas gerados para aplicação das técnicas de krigagem às vazões específicas mínimas  $q_{7,10}$  (a),  $q_{95}$  (b), e  $q_{90}$  (c) e média de longa duração  $q_{mld}$  (d).

Comprovou-se que o pressuposto geoestatístico relacionado à normalidade dos resíduos também foi atendido, segundo o teste de Shapiro-Wilk ao nível de significância de 5%. Da mesma forma, a regressão linear estabelecida entre os valores preditos e observados reforçou a confiabilidade dos modelos ao retornar valores próximos de 0 e 1 para o intercepto ( $\beta_0$ ) e coeficiente angular ( $\beta_1$ ), respectivamente.

Os valores estimados dos parâmetros do modelo exponencial e dos referidos erros oriundos da autovalidação LOOCV no ajuste dos semivariogramas podem ser observados na Tabela 1.3.

Tabela 1.3. Estimativa dos parâmetros e erros na geração dos semivariogramas ajustados

Variável	Modelo	Patamar	Alcance (m)	Efeito pepita	EM	SD	$\beta_0$	$\beta_1$
<b>q7,10</b>	Exponencial	1,92	35928,39	0,06	-0,03	1,00	-0,03	0,99
<b>q95</b>	Exponencial	2,63	36952,93	0,00	-0,02	1,03	0,09	0,96
<b>q90</b>	Exponencial	2,93	33989,44	0,00	-0,01	1,03	0,15	0,96
<b>qmld</b>	Exponencial	12,86	42259,98	0,00	0,01	0,97	-0,41	1,03

Outros autores obtiveram resultados similares, com o modelo exponencial sendo aquele associado aos melhores parâmetros de ajuste do semivariograma, tal como erro médio com valor próximo de zero e desvio padrão dos erros padronizados com valor próximo da unidade (SWAIN; PATRA, 2019; Xu et al., 2018).

### 1.3.4 Análise comparativa de desempenho das diferentes metodologias

#### 1.3.4.1 Vazões mínimas de referência ( $Q_{7,10}$ , $Q_{95}$ e $Q_{90}$ )

Os indicadores estatísticos apontaram que os métodos, em geral, forneceram predições mais acuradas para as vazões mínimas menos restritivas  $Q_{95}$  e  $Q_{90}$  em comparação à  $Q_{7,10}$ , conforme também obtido por Cecílio et al. (2018). Por outro lado, todas as vazões mínimas estiveram associadas a maiores erros de predição do que a  $Q_{mld}$ , independentemente do método de regionalização e da região homogênea. Além disso, percebeu-se um padrão de variação na acurácia dos métodos em função da região homogênea, ocorrendo maior acurácia na RHH I e erros maiores de predição na RHH III, como observado para o RR em relação ao ajuste das equações regionais.

Na Tabela 1.4 são apresentados os valores de PBIAS, calculados com base nas estimativas das vazões mínimas de referência ( $Q_{7,10}$ ,  $Q_{95}$  e  $Q_{90}$ ) por cada método, nas diferentes RHH.

Tabela 1.4. PBIAS nas estimativas das vazões mínimas pelos diferentes métodos em cada Região Hidrológicamente Homogênea

Métodos de regionalização	PBIAS (%)								
	Q <sub>7,10</sub>			Q <sub>95</sub>			Q <sub>90</sub>		
	Regiões Hidrológicamente Homogêneas (RHH)								
	I	II	III	I	II	III	I	II	III
<b>IL</b>	-3,41	0,23	20,84	-1,15	0,15	11,17	-1,12	-0,98	14,18
<b>ILmod</b>	-3,57	0,60	16,63	-1,30	0,34	7,42	-1,28	-0,66	10,31
<b>C</b>	-2,27	2,89	22,96	0,09	4,19	13,06	-0,17	2,60	16,07
<b>Cmod</b>	-2,69	1,68	17,66	-0,42	2,73	8,20	-0,68	1,27	11,10
<b>RR</b>	-7,03	-7,90	11,39	-5,92	-7,82	11,83	-3,87	-7,44	7,310
<b>KO</b>	-16,16	-26,90	-55,46	-9,28	-20,37	-42,61	-11,25	-19,69	-37,57
<b>KS</b>	-15,62	-25,13	-53,36	-10,40	-19,00	-41,00	-10,79	-18,65	-36,12

As metodologias baseadas na interpolação linear revelaram, de modo geral, um viés de estimativa mais sutil para as vazões mínimas nas RHH I e II. Na RHH III, porém, o RR foi o método que se apresentou menos enviesado para a maioria das vazões, sendo superado apenas quanto à estimativa da Q<sub>95</sub> pelas metodologias de ILmod e Cmod. Verificou-se que o viés do RR na RHH III decorreu da superestimativa na seção da estação 56994500.

Os valores mais extremos de PBIAS ocorreram na estimativa da Q<sub>7,10</sub> com a aplicação das krigagens e C. As krigagens apresentaram as maiores tendências de subestimar as vazões mínimas, enquanto o viés de superestimativa destas vazões foi maior com a aplicação de C na RHH III. Para a gestão de recursos hídricos, a tendência de superestimativa é um dos maiores problemas relacionados aos métodos de regionalização, pois pode acarretar em atendimento a demandas hídricas que extrapolam a disponibilidade dos mananciais, gerando conflitos de uso (MOREIRA; DAVID, 2014; SOUZA et al., 2012). Por outro lado, também é preciso considerar os problemas relacionados às subestimativas, especialmente quanto à concessão de outorgas, que pode ocasionar a subutilização dos recursos hídricos, originando impactos socioeconômicos e conflitos entre usuários e órgão gestor (BORK, 2018; OLIVEIRA, 2008).

Apesar dos resultados de PBIAS apontarem para o melhor desempenho das metodologias de interpolação linear e RR, vale destacar que este índice estatístico pode ocasionar enganos na avaliação dos métodos de regionalização de vazões se avaliado isoladamente, tendo em vista que tendências opostas se anulam, gerando valores de PBIAS próximos de zero ainda que as estimativas sejam ruins (MORIASI et al., 2015). Logo, seus valores serão analisados em conjunto dos demais indicadores para uma avaliação de desempenho mais confiável.

Os valores de ERM (Tabela 1.5), por sua vez, indicaram que os métodos mais acurados para as vazões mínimas foram ILmod e RR, sendo o primeiro mais adequado à RHH I e o segundo às RHH II e III. As predições geoestatísticas obtiveram, em geral, valores de ERM bastante similares às estimativas das metodologias de interpolação linear e/ou RR. Além disso, em certas regiões homogêneas, as krigagens foram superiores a outras metodologias.

Tabela 1.5. ERM nas estimativas das vazões mínimas pelos diferentes métodos em cada Região Hidrologicamente Homogênea

Métodos de regionalização	ERM (%)								
	Q <sub>7,10</sub>			Q <sub>95</sub>			Q <sub>90</sub>		
	Regiões Hidrologicamente Homogêneas (RHH)								
	I	II	III	I	II	III	I	II	III
<b>IL</b>	14,91	48,47	232,37	15,59	42,06	56,49	13,15	31,87	42,61
<b>ILmod</b>	14,06	45,26	203,74	12,26	40,24	47,35	10,29	30,60	34,60
<b>C</b>	14,63	51,84	233,74	15,59	44,13	59,20	13,39	33,30	44,01
<b>Cmod</b>	14,06	48,43	203,37	12,55	41,22	48,19	10,84	30,86	34,33
<b>RR</b>	21,84	26,02	45,17	18,76	24,70	20,96	16,80	18,94	19,76
<b>KO</b>	20,42	63,46	188,10	15,48	56,35	53,43	15,06	42,87	41,38
<b>KS</b>	19,84	65,72	195,74	17,30	57,32	54,42	14,75	42,53	42,45

Analisando-se os valores de RMSE (Tabela 1.6), verifica-se que os métodos de maior acurácia foram, em geral, C, Cmod e ILmod. No caso da Q<sub>7,10</sub> e Q<sub>90</sub>, estes métodos apresentaram os menores valores de RMSE para RHH I, II e III, respectivamente. Para a Q<sub>95</sub>, enquanto C apresentou-se mais acurado na RHH I, ILmod foi o método mais adequado às demais regiões. Por este indicador, o RR superou os métodos IL (Q<sub>7,10</sub> e Q<sub>90</sub> na RHH III) e C (Q<sub>7,10</sub> e Q<sub>90</sub> na RHH III e Q<sub>95</sub> na RHH II), apresentando pior desempenho geral na estimativa do que os métodos baseados na interpolação linear. Os valores de RMSE das krigagens apontam para baixo desempenho das predições geoestatísticas, apesar de superarem o RR na estimativa da Q<sub>7,10</sub> na RHH I. Estes resultados decorrem da sensibilidade que o RMSE possui em relação a erros de maiores magnitudes (*outliers*).

Tabela 1.6. Valores de RMSE nas estimativas das vazões mínimas pelos diferentes métodos em cada Região Hidrologicamente Homogênea

Métodos de regionalização	RMSE ( $\text{m}^3 \text{s}^{-1}$ )								
	Q <sub>7,10</sub>			Q <sub>95</sub>			Q <sub>90</sub>		
	Regiões Hidrologicamente Homogêneas (RHH)								
	I	II	III	I	II	III	I	II	III
<b>IL</b>	1,55	4,95	8,25	1,16	6,01	5,35	1,48	5,91	10,22
<b>ILmod</b>	1,55	5,22	6,36	1,15	5,40	3,18	1,47	5,41	7,26
<b>C</b>	1,31	3,88	8,29	0,89	8,05	5,46	1,20	6,38	10,30
<b>Cmod</b>	1,41	4,77	6,38	1,00	6,38	3,20	1,36	5,12	7,28
<b>RR</b>	3,14	9,00	7,34	3,13	7,80	9,92	2,57	10,33	8,14
<b>KO</b>	7,27	22,79	32,98	5,07	22,00	32,91	7,35	27,29	37,24
<b>KS</b>	7,18	22,04	32,08	6,01	21,39	32,01	7,22	26,53	36,24

A magnitude dos desvios absolutos com o emprego do RR e das krigagens foi, em geral, substancialmente maior em comparação àqueles correspondentes às metodologias de interpolação linear, gerando os valores de RMSE mais altos para RR, KO e KS. Constatou-se que os desvios mais acentuados com o emprego do RR e das krigagens ocorreram nas seções das estações fluviométricas 56539000 (RHH I), 56920000 (RHH II) e 56994500 (RHH III).

Os valores elevados de RMSE do RR decorreram, principalmente, dos desvios nas RHH II e III. A maioria das estações da RHH II está situada em locais chuvosos, enquanto a seção de interesse 56920000 se localiza em local seco. Assim, como a equação foi ajustada em função de muitos locais chuvosos, os valores estimados em locais secos foram menores que os observados. Esta subestimativa é tão majorada quanto maior a diferença entre a pluviosidade capturada no ajuste da regressão e a correspondente à área da seção de interesse.

As krigagens apresentaram valores mais acentuados de RMSE nas RHH II e III. Como o desempenho dos preditores geoestatísticos depende diretamente da densidade da rede de estações fluviométricas, a qualidade da predição nestas regiões foi mais afetada. Vale salientar, ainda, que na RHH III os desvios foram mais acentuados do que na RHH II e isto foi devido à má distribuição espacial das estações, fator que induziu as krigagens aos desvios nas predições, conforme a comprovação no estudo de Elesbon et al. (2014).

Os indicadores de qualidade NSE e  $R^2$  (Tabela 1.7) apontam que os métodos apresentaram boas estimativas, não havendo diferenças expressivas entre seus desempenhos, salvo algumas exceções para as krigagens. Ressalta-se que o uso de  $R^2$  para se avaliar a acurácia de métodos de modelagem somente é admitido se os dados estiverem condicionados à linha de regressão ajustada em 1:1. Neste estudo, como esta condição não existe, o papel do  $R^2$  foi capturar a eficiência dos métodos em representar a tendência de comportamento e a

proporção de variação dos valores de vazão observada entre as seções de monitoramento fluviométrico.

Tabela 1.7. Valores de NSE e R<sup>2</sup> nas estimativas das vazões mínimas pelos diferentes métodos em cada Região Hidrologicamente Homogênea

Métodos de regionalização	NSE (%)								
	Q <sub>7,10</sub>			Q <sub>95</sub>			Q <sub>90</sub>		
	Regiões Hidrologicamente Homogêneas (RHH)								
	I	II	III	I	II	III	I	II	III
<b>IL</b>	99,37	98,89	95,86	99,78	98,94	98,91	99,76	99,39	97,51
<b>ILmod</b>	99,37	98,77	97,54	99,78	99,15	99,61	99,76	99,48	98,74
<b>C</b>	99,55	99,32	95,82	99,87	98,10	98,86	99,84	99,28	97,47
<b>Cmod</b>	99,48	98,97	97,52	99,83	98,81	99,61	99,80	99,54	98,74
<b>RR</b>	97,41	96,33	96,72	98,38	98,21	96,24	99,28	98,12	98,42
<b>KO</b>	86,19	76,47	33,79	95,73	85,80	58,62	94,11	86,91	66,95
<b>KS</b>	86,53	78,00	37,37	94,00	86,57	60,85	94,32	87,63	68,71
Métodos de regionalização	R <sup>2</sup>								
	Q <sub>7,10</sub>			Q <sub>95</sub>			Q <sub>90</sub>		
	Regiões Hidrologicamente Homogêneas (RHH)								
	I	II	III	I	II	III	I	II	III
<b>IL</b>	0,9973	0,9898	0,9988	0,9981	0,9892	0,9995	0,9978	0,9950	0,9996
<b>ILmod</b>	0,9971	0,9878	0,9989	0,9981	0,9912	0,9996	0,9978	0,9954	0,9997
<b>C</b>	0,9972	0,9932	0,9989	0,9986	0,9833	0,9993	0,9983	0,9931	0,9996
<b>Cmod</b>	0,9964	0,9893	0,9990	0,9983	0,9888	0,9996	0,9979	0,9953	0,9998
<b>RR</b>	0,9897	0,9771	0,9993	0,9933	0,9951	0,9993	0,9955	0,9941	0,9995
<b>KO</b>	0,9309	0,9354	0,9258	0,9766	0,9642	0,9837	0,9655	0,9729	0,9889
<b>KS</b>	0,9321	0,9371	0,9364	0,9617	0,9635	0,9853	0,9661	0,9726	0,9898

O R<sup>2</sup> não constituiu um bom indicador de qualidade, em termos de diferenciação dos métodos quanto ao desempenho. Para todos os métodos, independentemente da vazão de referência e da região homogênea, as estimativas atingiram valores de R<sup>2</sup> muito próximos a 1, todos acima de 0,90. Já o NSE demonstrou-se mais útil à distinção dos métodos quanto à acurácia nas estimativas, apesar de também indicar a boa qualidade de todos os métodos para a regionalização das vazões mínimas de referência. De acordo com Moriasi et al. (2007), os métodos apresentaram “ótimo” desempenho (NSE>75%), com exceção das krigagens na RHH III, cujas predições tiveram “bom” desempenho para Q<sub>90</sub> (65%<NSE≤75%), “satisfatório” para Q<sub>95</sub> (50%<NSE≤65%) e “insatisfatório” para Q<sub>7,10</sub> (NSE≤50%).

Silva et al. (2017) alertam que os gestores devem dedicar especial atenção às regiões que apresentam mananciais com menor capacidade de regularização natural. Neste sentido, vale salientar a necessidade de atenção no gerenciamento dos recursos hídricos situados na

RHH III, cujo menor desempenho dos métodos na estimativa das vazões mínimas é inerente às suas características climáticas, refletidas em menor pluviosidade média e excedente hídrico. Logo, destaca-se que o método das regressões regionais é o mais recomendado para as vazões mínimas de referência na RHH III, respeitada sua aplicação às áreas de validade.

De maneira geral, os resultados apontaram que os métodos mais acurados na estimativa das vazões mínimas de referência foram aqueles baseados na interpolação linear, seguidos pelo RR e, por último, pelos preditores geoestatísticos (KO e KS). Vale ressaltar que, no entanto, como discutido por outros autores (CECÍLIO et al., 2018; NOVAES et al., 2007), a simplicidade do processo de estimativa e a viabilidade de implementação não devem ser ignorados na comparação entre os métodos de regionalização de vazões.

Apesar do comparativo de desempenho, ressalta-se que o emprego dos métodos de interpolação linear ao invés do RR deve ser avaliado com cuidado, pois o número de dados de entrada e a complexidade do processo de estimativa podem não justificar sua adoção (CECÍLIO et al., 2018). Os métodos de interpolação linear não utilizam regiões homogêneas e, assim como o RR, ainda estão altamente sujeitos a erros em decorrência das captações superficiais ao longo da hidrografia. Portanto, não se recomenda a sua utilização sem haver previamente uma análise criteriosa dos fatores que afetam seu desempenho, principalmente a quantidade e distribuição espacial das estações fluviométricas ao longo da bacia hidrográfica.

Do mesmo modo, o uso da geoestatística deve ser avaliado com cautela para fins de gerenciamento dos recursos hídricos, pois seu desempenho varia com a distribuição espacial e densidade das estações fluviométricas. Porém, com exceção da RHH III, esta abordagem apresentou bons indicadores de qualidade, apesar da baixa densidade de estações fluviométricas na bacia. Logo, a geoestatística pode constituir uma ferramenta importante à gestão dos recursos hídricos na bacia do rio Doce, especialmente pela sua facilidade de espacialização das vazões para as bacias menores e mais desprovidas de dados fluviométricos.

#### **1.3.4.2 Vazão média de referência ( $Q_{mld}$ )**

Na Tabela 1.8 são apresentados os valores de PBIAS para as estimativas da  $Q_{mld}$  pelos diferentes métodos em cada RHH. Pode-se notar que o RR foi superior aos outros métodos em relação ao PBIAS nas RHH I e III. Os preditores geoestatísticos, por sua vez, apesar de demonstrarem desempenhos inferiores aos demais métodos nas RHH I e II, geraram

estimativas na RHH III com viés percentual menor do que aquelas provenientes dos métodos baseados na interpolação linear.

Tabela 1.8. PBIAS nas estimativas da vazão média de longa duração ( $Q_{mld}$ ) pelos diferentes métodos em cada Região Hidrologicamente Homogênea

Métodos de regionalização	PBIAS (%)		
	Regiões Hidrologicamente Homogêneas (RHH)		
	I	II	III
<b>IL</b>	0,46	2,25	6,79
<b>ILmod</b>	0,21	2,32	3,27
<b>C</b>	1,27	6,18	8,66
<b>Cmod</b>	0,62	4,68	3,99
<b>RR</b>	0,28	-4,44	-0,08
<b>KO</b>	-2,29	-10,92	0,30
<b>KS</b>	-2,32	-11,07	0,89

Em relação ao ERM (Tabela 1.9), o método RR apresentou os menores valores na estimativa da  $Q_{mld}$ , perdendo somente para a ILmod na RHH I. As krigagens obtiveram valores de ERM próximos dos demais métodos, superando IL e C nas RHH I e III.

Tabela 1.9. ERM nas estimativas da vazão média de longa duração ( $Q_{mld}$ ) pelos diferentes métodos em cada Região Hidrologicamente Homogênea

Métodos de regionalização	ERM (%)		
	Regiões Hidrologicamente Homogêneas (RHH)		
	I	II	III
<b>IL</b>	12,17	21,97	13,66
<b>ILmod</b>	8,66	19,65	11,50
<b>C</b>	12,10	21,30	14,94
<b>Cmod</b>	8,77	18,51	11,39
<b>RR</b>	8,76	10,30	9,27
<b>KO</b>	10,51	27,24	11,61
<b>KS</b>	10,26	24,89	12,52

Como discutido por outros autores, os erros relativos na estimativa da  $Q_{mld}$  foram inferiores àqueles observados na estimativa das vazões mínimas em decorrência, principalmente, do achatamento dos valores extremos quando da obtenção dos valores médios, exibindo magnitudes de variação menores do que as vazões mínimas (ARAÚJO et al., 2018; MOREIRA; DAVID, 2014; NOVAES et al., 2007).

Os valores de RMSE (Tabela 1.10) foram mais elevados nas estimativas realizadas na RHH II, indicando desempenho inferior dos métodos nesta região, tal como apontando pelos

demais indicadores de qualidade. Em média, o RMSE na RHH II foi cerca de seis vezes maior do que aquele associado às estimativas na RHH I e quase três vezes superior àquele referente às estimativas na RHH III. Na estimativa da  $Q_{mld}$ , a maior acurácia dos métodos também ocorreu na RHH I, assim como se constatou para as vazões mínimas de referência.

Tabela 1.10. Valores de RMSE nas estimativas da vazão média de longa duração ( $Q_{mld}$ ) pelos diferentes métodos em cada Região Hidrologicamente Homogênea

Métodos de regionalização	RMSE ( $m^3 s^{-1}$ )		
	Regiões Hidrologicamente Homogêneas (RHH)		
	I	II	III
<b>IL</b>	4,51	19,42	18,96
<b>ILmod</b>	2,58	15,82	11,22
<b>C</b>	4,80	20,44	19,23
<b>Cmod</b>	2,92	15,34	11,22
<b>RR</b>	2,05	21,35	2,98
<b>KO</b>	6,92	45,97	3,75
<b>KS</b>	6,80	45,55	3,92

Quanto aos métodos de interpolação linear, os valores elevados de RMSE na RHH II se devem em grande parte à menor acurácia da estimativa nas seções fluviais referentes ao caso 2 de posição relativa. Os erros mais acentuados na estimativa da  $Q_{mld}$  e, conseqüentemente, com maior contribuição para a elevação do RMSE, ocorreram nas seções das estações 56920000 (RHH II) e 56978000 (RHH III), cujas características pluviométricas e a densidade e distribuição espacial das estações exerceram grande influência na magnitude dos desvios, tal com observado para as vazões mínimas.

Embora tenha sido notada uma diferença numérica entre os valores de  $R^2$  e NSE entre as regiões homogêneas (Tabela 1.11), ambos os indicadores retrataram um ótimo desempenho de todos os métodos aplicados. O indicador  $R^2$  apresentou valores muito próximos de 1 e o valor mais baixo de NSE foi cerca de 94,66%. Portanto, além de seguirem uma tendência de variação em proporções similares, os valores estimados de  $Q_{mld}$  pelos métodos aplicados ainda refletiram uma notável proximidade numérica em relação aos valores observados.

Tabela 1.11. Valores de NSE e R<sup>2</sup> nas estimativas da vazão média de longa duração (Q<sub>mld</sub>) pelos diferentes métodos em cada Região Hidrologicamente Homogênea

Métodos de regionalização	NSE (%)		
	Regiões Hidrologicamente Homogêneas		
	I	II	III
<b>IL</b>	99,66	99,05	98,89
<b>ILmod</b>	99,89	99,37	99,61
<b>C</b>	99,62	98,94	98,85
<b>Cmod</b>	99,86	99,41	99,61
<b>RR</b>	99,93	98,85	99,97
<b>KO</b>	99,21	94,66	99,96
<b>KS</b>	99,23	94,75	99,95

Métodos de regionalização	R <sup>2</sup>		
	Regiões Hidrologicamente Homogêneas		
	I	II	III
<b>IL</b>	0,9963	0,9903	0,9996
<b>ILmod</b>	0,9988	0,9936	0,9997
<b>C</b>	0,9959	0,9909	0,9995
<b>Cmod</b>	0,9985	0,9946	0,9997
<b>RR</b>	0,9993	0,9934	0,9997
<b>KO</b>	0,9914	0,9836	0,9996
<b>KS</b>	0,9918	0,9838	0,9996

Na estimativa da Q<sub>mld</sub> não houve qualquer desempenho considerado insatisfatório, segundo os valores de NSE propostos por Moriasi et al. (2007). De modo geral, apesar dos métodos ILmod, Cmod e RR terem se destacado entre os demais no que tange à comparação numérica dos indicadores de qualidade, todos os métodos estudados exibiram resultados promissores para a regionalização da Q<sub>mld</sub> na bacia do rio Doce. Portanto, vale destacar que assim como para as vazões mínimas de referência, a adoção dos métodos de interpolação linear ao invés do RR também deve ser analisada com muito critério para estimativa da Q<sub>mld</sub>.

Destaca-se, ainda, que considerando os dados deste estudo, as krigagens demonstraram expressiva capacidade de prever com bom desempenho as vazões mínimas de referência e média de longa duração, não apresentando bom desempenho apenas para a Q<sub>7,10</sub> na RHH III. Outros autores obtiveram resultados similares, apontando que as krigagens apresentaram desempenho tão bom ou superior aos demais métodos de regionalização de vazões (MERZ; BLÖSCHL, 2004; PARAJKA; MERZ; BLÖSCHL, 2005). Os resultados também revelaram que as krigagens podem representar uma ferramenta promissora à gestão dos recursos hídricos em áreas de drenagem menores, minimizando a inviabilidade das demais abordagens de serem aplicadas em bacias hidrográficas com monitoramento hidrológico precário ou inexistente.

## 1.4 CONCLUSÕES

A partir dos resultados obtidos pode-se concluir que:

- Os métodos avaliados para fins de regionalização de vazões (Interpolação Linear, Interpolação Linear modificado, Chaves et al. (2002) original, Chaves et al. (2002) modificado, regressões regionais, krigagens ordinária e simples) apresentaram desempenho similar na estimativa das vazões mínimas de referência e média de longa duração na bacia do rio Doce, sendo a única diferença expressiva a menor acurácia dos métodos geoestatísticos na predição da  $Q_{7,10}$  em uma das três regiões hidrologicamente homogêneas da bacia do rio Doce, caracterizada por distribuição espacial desuniforme das estações fluviométricas.
- Todas as abordagens de regionalização de vazões apresentaram desempenho bastante similar e superior para a vazão média de longa duração em relação às vazões mínimas de referência.
- As abordagens com foco em interpolação linear obtiveram estimativas mais acuradas para as seções alocadas entre duas estações (caso 3) e piores estimativas nas situações em que a seção de interesse se situa a jusante da estação de apoio (caso 2) ou em trecho afluente do rio monitorado com confluência a jusante da estação de apoio (caso 5), devido às grandes diferenças de área de drenagem entre as seções de interesse e as respectivas seções de apoio.
- A modelagem geoestatística demonstrou bom desempenho preditivo, constituindo uma alternativa promissora aos outros métodos de regionalização de vazões em bacias menores e com precariedade de dados fluviométricos.
- A adoção dos métodos geoestatísticos e baseados em interpolação linear deve estar sempre associada a uma análise bastante criteriosa de fatores como a densidade e a distribuição espacial das estações de monitoramento fluviométrico na bacia hidrográfica.

## 1.5 REFERÊNCIAS

- ALMEIDA, M. M. D. E. **Regionalização de vazões médias de longo período e de volumes de reservatórios de regularização**. 2010. 209 f. Dissertação (Mestrado em Engenharia Ambiental) – Universidade Federal do Espírito Santo, Vitória, 2010.
- AGERH. **Comitês de bacias hidrográficas**. Disponível em:  
<<https://agerh.es.gov.br/documentos-dos-comites>>. Acesso em: 09 jul. 2020.
- ANA. **Base hidrográfica ottocodificada da bacia do rio Doce 1:50.000/1.100.000**. Disponível em: <<http://metadados.ana.gov.br/geonetwork/srv/pt/main.home>>. Acesso em: 31 mai. 2019.
- ANKIT, A. et al. Optimal Design of Hydrometric Station Networks Based on Complex Network Analysis. **Hydrology and Earth System Sciences**, n. March, p. 1–21, 2018.
- ARAÚJO, F. C. et al. Streamflow regionalization in Piquiri river basin. **Engenharia Agrícola**, v. 4430, p. 22–31, 2018.
- ARSENAULT, R. et al. Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. **Hydrological Sciences Journal**, v. 64, n. 11, p. 1297–1311, 2019.
- BARBOSA, S. E. et al. 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**, v. 10, n. 1, p. 64–71, 2005.
- BAZZO, K. R. et al. Regionalização da vazão Q95: comparação de métodos para a bacia hidrográfica do Rio Taquari-Antas, RS. **Revista Ambiente & Água**, v. 12, n. 5, p. 855–870, 2017.
- BESKOW, S. et al. Potential of the LASH model for water resources management in data-scarce basins: a case study of the Fragata River basin, southern Brazil. **Hydrological Sciences Journal**, v. 61, n. 14, p. 2567–2578, 2016.
- BHATTI, S. J.; KROLL, C. N.; VOGEL, R. M. Revisiting the Probability Distribution of Low Streamflow Series in the United States. **Journal of Hydrologic Engineering**, v. 24, n. 10, p. 1–11, 2019.
- BORK, C. K. **Regionalização de vazões mínimas para o estado do Rio Grande do Sul**. 2018. 156 f. Dissertação (Mestrado em Recursos Hídricos) – Universidade Federal de Pelotas, Londrina, 2018.
- CALEGARIO, A. T. **Identificação de regiões hidrologicamente homogêneas e**

- comportamento das vazões mínima e média regionalizadas**. 2014. 80 f. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Federal de Viçosa, Viçosa, 2014.
- CBH-DOCE. **A bacia do rio Doce**. Disponível em:  
<<http://www.cbhdoce.org.br/institucional/a-bacia>>. Acesso em: 28 fev. 2019a.
- CBH-DOCE. **Deliberação Normativa CBH-Doce nº 51/2016**. 2016. Disponível em:  
<<http://www.cbhdoce.org.br/wp-content/uploads/2016/12/Delibera%C3%A7%C3%A3o-051-Ad-Referendum-Aprova-Realoca%C3%A7%C3%A3o-do-PAP.pdf>>. Acesso em: 28 fev. 2020b.
- CECÍLIO, R. A. et al. Evaluation of low-flow and long term average flow regionalization methods on Itapemirim river watershed. **Scientia Agraria**, v. 19, n. 2, p. 122–132, 2018.
- CHAVES, H. M. L. et al. Regionalização de Vazões Mínimas em Bacias Através de Interpolação em Sistemas de Informação Geográfica. **Revista Brasileira de Recursos Hídricos**, v. 7, n. 3, p. 43–51, 2002.
- COMINI, U. B. et al. Hydrological modelling in small ungauged catchments. **Anais da Academia Brasileira de Ciências**, v. 92, n. 2, p. 1–16, 2020.
- CPRM. **Acompanhamento da estiagem na região Sudeste do Brasil – Relatório 6: Área de atuação da Superintendência Regional da CPRM de Belo Horizonte**. 2014.
- ECOPLAN-LUME. **Plano Integrado de Recursos Hídricos da Bacia Hidrográfica do Rio Doce - Volume I**. Consórcio ECOPLAN-LUME, 2010. 478 p.
- ELESBON, A. A. A. et al. Methodological proposal for hydrometric network design: Part II - Exclusion, rearrangement and inclusion of monitoring stations. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 18, n. 10, p. 1023–1030, 2014.
- ELESBON, A. A. A. et al. Multivariate statistical analysis to support the minimum streamflow regionalization. **Engenharia Agrícola**, v. 35, n. 5, p. 838–851, 2015.
- ELETROBRÁS. **Manual de Minicentraís Hidrelétricas**. Rio de Janeiro: Eletrobrás, 1985.
- FARMER, W. H. Ordinary kriging as a tool to estimate historical daily streamflow records. **Hydrology and Earth System Sciences**, v. 20, p. 2721–2735, 2016.
- GOMES, E. P.; BLANCO, C. J. C.; PESSOA, F. C. L. Regionalization of precipitation with determination of homogeneous regions via fuzzy c-means. **Revista Brasileira de Recursos Hídricos**, v. 23, n. 51, p. 1–19, 2018.
- IBGE. **Censo Demográfico 2010**. Disponível em:  
<<https://censo2010.ibge.gov.br/resultados.html>>. 2010. Acesso em: 28 fev. 2020.

- IGAM. **Estudo de regionalização de vazão para o aprimoramento do processo de outorga no Estado de Minas Gerais**. Belo Horizonte: IGAM, 2012. 415 p.
- IGAM. **Relatório hidrometeorológico: período chuvoso 2014/2015**. Belo Horizonte: IGAM. 2015. 30 p.
- JOURNAL, A. G.; HUIJBREGTS, C. J. **Mining Geostatistics**. New York: Academic Press, 1978.
- KESHTKAR, A. R. Low flow frequency analysis by L-moments method (Case study: Iranian Central Plateau River Basin). **Desert**, v. 20, n. 2, p. 167–175, 2015.
- LACHANCE-CLOUTIER, S.; TURCOTTE, R.; CYR, J. Combining streamflow observations and hydrologic simulations for the retrospective estimation of daily streamflow for ungauged rivers in southern Quebec (Canada). **Journal of Hydrology**, v. 550, p. 294–306, 2017.
- LANGAT, P. K.; KUMAR, L.; KOECH, R. Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow. **Water (Switzerland)**, v. 11, p. 1–24, 2019.
- LI, Q. et al. Hydrological regionalisation based on available hydrological information for runoff prediction at catchment scale. **Proceedings of the International Association of Hydrological Sciences**, v. 379, n. 1, p. 13–19, 2018.
- LI, Q. et al. A combined method for estimating continuous runoff by parameter transfer and drainage area ratio method in ungauged catchments. **Water**, v. 11, n. 5, p. 1–18, 2019.
- LOPES, T. R. et al. Regionalização de vazões máximas e mínimas para a bacia do rio Ivaí - PR. **Irriga**, v. 21, n. 1, p. 188–201, 2016.
- LOPES, T. R. et al. Regionalization of maximum and minimum flow in the Teles Pires basin, Brazil. **Journal of the Brazilian Association of Agricultural Engineering**, v. 37, n. 1, p. 54–63, 2017.
- LOUZADA, F. L. R. D. O.; XAVIER, A. C.; PEZZOPANE, J. E. M. Climatological water balance with data estimated by tropical rainfall measuring mission for the Doce river basin. **Engenharia Agrícola**, v. 38, n. 3, p. 376–386, 2018.
- MARQUES, F. A. **Sistema de controle dinâmico para a gestão dos usos múltiplos da água**. 2010. 234 f. Tese (Doutorado em Engenharia Agrícola) – Universidade Federal de Viçosa, Viçosa. 2010.
- MELATI, M. D.; MARCUZZO, F. F. N. Regressões simples e robusta na regionalização da vazão Q95 na Bacia Hidrográfica do Taquari-Antas Simple and robust regressions in

- flows regionalization of Q95 in Taquari-Antas river basin. **Ciencia e Natura**, v. 38, p. 722–739, 2016.
- MERZ, R.; BLÖSCHL, G. Regionalisation of catchment model parameters. **Journal of Hydrology**, v. 287, n. 1–4, p. 95–123, 2004.
- MOREIRA, M. C.; DAVID, D. Análise de métodos para estimativa das vazões da bacia do rio Paraopeba. **Revista Brasileira de Recursos Hídricos**, v. 19, n. 2, p. 313–324, 2014.
- MORIASI, D. N. et al. Model evaluation guidelines for systematic quantification of accuracy in watershed simulation. **Transactions of the ASABE**, v. 50, n. 3, p. 885–900, 2007.
- MORIASI, D. N. et al. Hydrologic and water quality models: Performance measures and evaluation criteria. **Transactions of the ASABE**, v. 58, n. 6, p. 1763–1785, 2015.
- NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models part I — A discussion of principles. **Journal of Hydrology**, v. 10, n. 3, p. 282–290, 1970.
- NOSRATI, K. et al. Regional low flow analysis in Sefidrood Drainage Basin, Iran using principal component regression. **Hydrology Research**, v. 46, n. 1, p. 121–135, 2015.
- NOVAES, L. F. DE et al. Avaliação do desempenho de cinco metodologias de regionalização de vazões. **Revista Brasileira de Recursos Hídricos**, v. 12, n. 2, p. 51–61, 2007.
- NRUTHYA, K.; SRINIVAS, V. V. Evaluating methods to predict streamflow at ungauged sites using regional flow duration curves: a case study. **Aquatic Procedia**, v. 4, p. 641–648, 2015.
- OLEA, R. A. **Geostatistics for engineers and earth scientists**. Boston: Kluwer Academic Publishers, 1999.
- OLIVEIRA, F. A. **Procedimentos para aprimorar a regionalização de vazões: estudo de caso da bacia do rio Grande**. 2008. 173 f. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Federal de Viçosa, Viçosa, 2008.
- PARAJKA, J.; MERZ, R.; BLÖSCHL, G. A comparison of regionalisation methods for catchment model parameters. **Hydrology and Earth System Sciences**, v. 9, p. 157–171, 2005.
- PETTITT, A. N. A Non-Parametric approach to the Change-Point problem. **Applied Statistics**, v. 28, n. 2, p. 126–135, 1979.
- PIOL, M. V. A. et al. Performance evaluation of flow duration curves regionalization methods. **Revista Brasileira de Recursos Hidricos**, v. 24, 2019.
- PRUSKI, F. F. et al. Improved regionalization of streamflow by use of the streamflow equivalent of precipitation as an explanatory variable. **Journal of Hydrology**, v. 476, p.

52–71, 2013.

PRUSKI, F. F. et al. Low-flow estimates in regions of extrapolation of the regionalization equations: a new concept. **Engenharia Agrícola**, v. 35, n. 5, p. 808–816, 2015.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. 2010. Vienna: R Foundation for Statistical Computing. Disponível em: <<https://www.rproject.org/>>. Acesso em: 05 nov. 2019.

RIBEIRO JR, P. J.; DIGGLE, P. J. geoR: A package for geostatistical data analysis using the R software. **R-News**, v. 1, n. 2, p. 15–18, 2001.

ROJAS-SERNA, C. et al. How should a rainfall-runoff model be parameterized in an almost ungauged catchment? A methodology tested on 609 catchments. **Water Resources Research**, n. June, p. 4765–4784, 2016.

SANTOS, G. R. **Hierarquização geométrica dos preditores geoestatísticos**. 2010. 200 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2010.

SERRANO, L. D. O. et al. Low-flow seasonality and effects on water availability throughout the river network. **Water Resources Management**, p. 1–16, 2020.

SILVA, D. D.; MARQUES, F. A.; LEMOS, A. F. Avaliação de metodologias de regionalização de vazões mínimas de referência para bacia do rio São Francisco. **Revista Engenharia na Agricultura - REVENG**, v. 17, n. 5, p. 392–403, 2009.

SILVA, D. D.; PEREIRA, S. B.; VIEIRA, E. O. Integrated water resources management in Brazil. In: VIEIRA, E. DE O. et al. (Eds.). **Integrated water resource management**. 1. ed. Davis: Springer, Cham, 2020. 151 p.

SILVA, R. E.; BLANCO, C. J. C.; PESSOA, F. C. L. Alternative for the regionalization of flow duration curves. **Journal of Applied Water Engineering and Research**, v. 9676, n. May, 2019.

SILVA, L. A. et al. Minimum and reference discharges and specific yield for the state of Minas Gerais, Brazil. **Revista Brasileira de Ciências Agrárias**, v. 12, n. 4, p. 543–549, 2017.

SOUSA, H. T. DE et al. **SisCAH 1.0 – Sistema Computacional para Análises Hidrológicas**. Brasília, DF: Agência Nacional das Águas, 2009.

SOUSA, H. T. **Sistema computacional para regionalização de vazões**. 2009. 86 f. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Federal de Viçosa, Viçosa, 2009.

- SOUZA, F. A. O. et al. Caracterização das vazões em uma pequena bacia hidrográfica do Distrito Federal , Brasil Discharge characterization in a watershed of Distrito Federal , Brazil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 16, n. 1, p. 10–17, 2012.
- SWAIN, J. B.; PATRA, K. C. Streamflow estimation in ungauged catchments using regionalization techniques. **Journal of Hydrology**, v. 554, p. 420–433, 2017.
- SWAIN, J. B.; PATRA, K. C. Impact assessment of land use/land cover and climate change on streamflow regionalization in an ungauged catchment. **Water & Climate Change**, v. 10, n. 3, p. 554–568, 2019.
- TIEZZI, R. et al. Trends of streamflow under climate change for 26 Brazilian basins. **Water Policy**, v. 21, n. 6, p. 206–220, 2019.
- ULIANA, E. M. et al. Análise de tendência em séries históricas de vazão e precipitação: uso de teste estatístico não paramétrico. **Revista Ambiente e Agua**, v. 10, n. 1, p. 82–88, 2015.
- VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: **Tópicos em ciência do solo**. Sociedade Brasileira de Ciência do Solo, 2000. p. 1–54.
- WASEEM, M. et al. Development and evaluation of an extended inverse distance weighting method for streamflow estimation at an ungauged site. **Hydrology Research**, v. 47, n. 2, p. 333–343, 2016.
- XU, P. et al. A kriging and entropy-based approach to raingauge network design. **Environmental Research**, v. 161, p. 61–75, 2018.
- YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística: conceitos e aplicações**. São Paulo: Oficina de Textos, 2013. 215 p.
- YANG, X. et al. Runoff prediction in ungauged catchments in Norway: comparison of regionalization approaches. **Hydrology Research**, v. 49, n. 2, p. 487–505, 2018.
- YILMAZ, M. U.; ONOZ, B. A Comparative Study of Statistical Methods for Daily Streamflow Estimation at Ungauged Basins in Turkey. **Water**, v. 12, n. 2, p. 1–23, 2020.
- ZAMOUM, S.; SOUAG-GAMANE, D. Monthly streamflow estimation in ungauged catchments of northern Algeria using regionalization of conceptual model parameters. **Arabian Journal of Geosciences**, v. 12, n. 11, 2019.

## CAPÍTULO 2:

### **Regionalização de vazões utilizando modelos de aprendizado de máquina para a bacia do rio Doce, Brasil**

**RESUMO:** Nos últimos anos, os hidrólogos têm explorado cada vez mais o aprendizado de máquina na descrição e previsão dos regimes de vazões em bacias hidrográficas, dada a infinidade de variáveis ambientais que interagem entre si, afetando os diversos componentes do ciclo hidrológico. O presente trabalho visou avaliar diferentes modelos de aprendizado de máquina na regionalização de vazões para a bacia do rio Doce, analisando suas vantagens e limitações, com vistas a verificar sua viabilidade de implementação e o ganho geral para a gestão dos recursos hídricos. Os algoritmos aplicados foram o Random Forest, o Earth e um modelo linear. As variáveis resposta foram as vazões mínimas de referência ( $Q_{7,10}$ ,  $Q_{95}$  e  $Q_{90}$ ) e a vazão média de longa duração ( $Q_{mld}$ ). O conjunto de covariáveis ambientais foi constituído por um grupo de 76 covariáveis preditoras, classificadas nas seguintes categorias: morfometria, topografia, clima, uso e cobertura da terra e condições de superfície. Cada covariável ambiental se referiu à área a montante de cada estação fluviométrica localizada na bacia e utilizada no estudo. De posse das covariáveis ambientais foram efetuados dois processos de eliminação de covariáveis: eliminação por correlação e eliminação por importância. Os modelos foram treinados após cada processamento do RFE (Recursive Feature Elimination). A fase de validação se deu com a aplicação do modelo melhor ajustado pelo treinamento, processado apenas com a amostra de validação. Os indicadores estatísticos da validação foram o coeficiente de Nash-Sutcliffe (NSE), viés percentual (PBIAS), índice de concordância de Willmott (d), coeficiente de determinação ( $R^2$ ), raiz do erro quadrático médio (RMSE), erro médio absoluto (MAE) e erro relativo (ER). O modelo linear foi insatisfatório para a predição de todas as vazões de referência. Os modelos não lineares apresentaram bom desempenho, tendo como covariável preditiva de maior importância a vazão equivalente ao volume precipitado com abstração de 750 mm ( $Peq_{750}$ ). Em geral, os modelos Random Forest e Earth apresentaram desempenhos similares e grande habilidade em prever vazões mínimas de referência e média de longa duração na bacia do rio Doce, constituindo alternativas poderosas e promissoras à regionalização de vazões em suporte à gestão e ao planejamento integrado dos recursos hídricos em nível de bacia hidrográfica.

**Palavras-chave:** Ciclo hidrológico; Inteligência artificial; Vazões de referência

## 2.1 INTRODUÇÃO

Na última década, o aumento da capacidade computacional contribuiu para o estudo e desenvolvimento de modelos de aprendizado de máquina cada vez mais robustos e confiáveis. Isto propiciou a criação de um campo de pesquisa focado no desenvolvimento e aprimoramento de métodos avançados de predição da vazão fluvial (KISI et al., 2019; SHORTRIDGE; GUIKEMA; ZAITCHIK, 2016).

Para estudos hidrológicos, ainda não há um modelo de aprendizado de máquina mais adequado a todas as situações (AL-SUDANI; SALIH; MUNDHER, 2019). Além disso, apesar dos recentes avanços metodológicos referentes à regionalização de vazões e da crescente pesquisa sobre inteligência artificial, tais modelos ainda são pouco explorados na predição de vazões mínimas em bacias sem monitoramento (WORLAND; FARMER; KIANG, 2018). Apesar desta realidade, nos últimos anos alguns modelos desta abordagem têm recebido cada vez mais atenção no campo científico da Hidrologia, dentre eles o *Random Forest* (RF) e o *Support Vector Machine* (SVM) (GHORBANI et al., 2016; PAPACHARALAMPOUS; TYRALIS; KOUTSOYIANNIS, 2019; PATEL; RAMACHANDRAN, 2015; PRIETO et al., 2019; TONGAL; BOOIJ, 2018). A popularidade desses modelos se deve ao fato de possuírem ampla aplicabilidade, grande capacidade em relacionar múltiplas variáveis e, ainda, apresentarem, de forma geral, elevada acurácia nas predições (JURKOVIC et al., 2018; WUEST et al., 2016; ZHANG; GOH, 2016).

Outra abordagem de notório potencial e de uso crescente na predição e previsão de vazões é o *Multivariate Adaptive Regression Splines* (MARS) (ADNAN et al., 2019; YIN et al., 2018). Suas principais vantagens competitivas são a eficiência e a robustez. Além de processar inúmeras relações não lineares com rapidez, esta abordagem tem grande capacidade em capturar a relação intrínseca entre os dados e, apesar da complexidade envolvida, fornecer modelos mais simples e de fácil interpretação (ADNAN et al., 2019; ZHANG; GOH, 2016).

A simplicidade dos modelos é uma das características mais importantes para os gestores dos recursos hídricos. Em um dos estudos mais completos a comparar métodos de predição de vazões, os modelos lineares superaram seus concorrentes de maior complexidade em determinadas condições, como aquelas em que há poucos dados disponíveis para o estudo (ELSHORBAGY et al., 2010a, 2010b). Assim, fica evidente a relevância de serem elencados modelos com níveis distintos de complexidade em estudos comparativos desta natureza.

Este estudo visa avaliar diferentes modelos de aprendizado de máquina na regionalização de vazões para a bacia do rio Doce, analisando as vantagens e limitações dos modelos aplicados e, baseado na acurácia de predição de cada um, apontar a sua viabilidade de implementação e o ganho geral para a gestão dos recursos hídricos.

## 2.2 MATERIAL E MÉTODOS

### 2.2.1 Área de estudo

A bacia hidrográfica do rio Doce delimita uma área aproximada de 82.427 km<sup>2</sup>, sendo 86% pertencente ao estado de Minas Gerais e o restante ao estado do Espírito Santo. Inserida na região hidrográfica brasileira do Atlântico Sudeste, a bacia envolve 200 municípios mineiros e 28 municípios capixabas (CBH-DOCE, 2016a). Na Figura 2.1 é apresentada a localização da bacia e seus principais cursos de água, conforme o Plano Integrado de Recursos Hídricos da bacia do rio Doce (ECOPLAN-LUME, 2010).

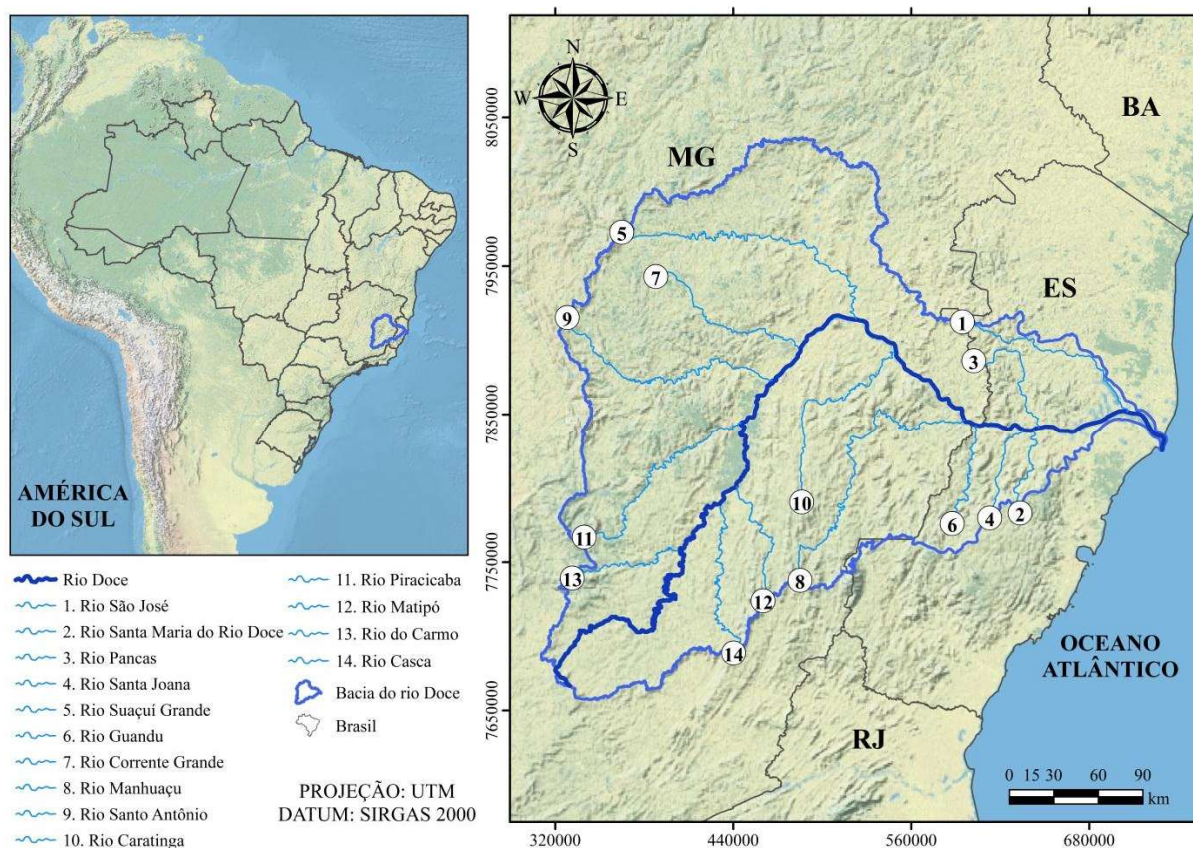


Figura 2.1. Localização da bacia do rio Doce e principais afluentes do rio principal.

A população residente da bacia é de aproximadamente 3,6 milhões de habitantes (IBGE, 2010). Sua atividade econômica é bastante diversificada e desempenha um importante papel para a economia do país, destacando-se os seguintes setores: indústria (siderurgia, celulose e laticínios), agropecuária/agroindústria, mineração (ferro, ouro, bauxita e pedras preciosas), comércio, serviços de apoio dos complexos industriais, e a geração de energia hidrelétrica (ECOPLAN-LUME, 2010).

Para fins de planejamento e gestão dos recursos hídricos, a bacia do rio Doce foi dividida em: Alto, Médio e Baixo Doce. O Alto Doce se refere à área a montante da confluência do rio Doce com o rio Piracicaba. O Médio Doce delimita-se a partir dessa confluência até a divisa entre Minas Gerais e Espírito Santo. Já o Baixo Doce corresponde à parcela capixaba da bacia do Rio Doce (ECOPLAN-LUME, 2010).

A bacia é composta por seis Unidades de Gestão de Recursos Hídricos (UGRH's) em Minas Gerais, a saber: UGRH1 Piranga; UGRH2 Piracicaba; UGRH3 Santo Antônio; UGRH4 Suaçuí; UGRH5 Caratinga; UGRH6 Manhuaçu; e cinco no Espírito Santo, que são: UGRH7 Guandu; UGRH8 Santa Maria do Doce (CBH-DOCE, 2016b), e outras três unidades de planejamento e gestão que foram criadas recentemente, como Santa Joana, Barra Seca e Foz do rio Doce (BSFRD), e Pontões e Lagoas do rio Doce (AGERH, 2020).

### **2.2.1.1 Caracterização da bacia do rio Doce**

O clima típico da bacia é o tropical úmido e está submetido a forte influência orográfica e de continentalidade (CUPOLILLO; ABREU; VIANELLO, 2008). A bacia está inserida numa região de alta variabilidade climática, decorrente de fenômenos de grande escala, como a interação entre a Alta Subtropical do Atlântico Sul (ASAS), as Zonas de Convergência do Atlântico Sul (ZCAS) e os sistemas frontais (CAVALCANTI et al., 2009).

Na maior parte da bacia, a média anual da temperatura do ar está compreendida entre 20-26°C. As temperaturas mais altas são registradas entre os meses de dezembro e março (LOUZADA; XAVIER; PEZZOPANE, 2018). A precipitação média na bacia é mal distribuída e seu regime pluviométrico apresenta um ano hidrológico com duas estações bem definidas: seca (150-250mm) e chuvosa (800-1300mm). Tais estações constituem os períodos de abril a setembro e outubro a março, respectivamente (ECOPLAN-LUME, 2010).

Os solos predominantes na bacia do rio Doce pertencem às classes dos Latossolos e Argissolos. Os Latossolos Vermelho-Amarelos estão presentes em cerca de 66% da bacia e os Argissolos Vermelhos correspondem a 23% da sua área (SOUZA et al., 2018b). Outras classes de solos ocorrem em menores proporções ao longo da bacia, como os Latossolos húmicos, os solos litólicos e os Cambissolos (OLIVEIRA; QUARESMA, 2017).

O relevo da bacia é conhecido como “mar de morros”, impondo restrições às atividades econômicas e ocupação urbana. Os locais para instalação dos setores produtivos e de moradias foram condicionados pelo relevo, o que favoreceu a ocupação próxima às margens fluviais, altamente suscetíveis aos eventos de inundações (ANA, 2016).

### **2.2.2 Dados hidrometeorológicos**

Os dados mais básicos à modelagem hidrológica, considerando o processo físico de formação das vazões fluviais, e mais requeridos pelos diversos métodos de regionalização de vazões são os dados de chuva e da própria vazão. Esses dados foram adquiridos junto à rede hidrometeorológica da Agência Nacional de Águas (ANA), coletados através do Portal HydroWeb/Sistema Nacional de Informações sobre os Recursos Hídricos (SNIRH). As séries históricas utilizadas foram constituídas por dados diários de vazão e precipitação, correspondentes ao período base de outubro de 1988 a setembro de 2018, uma vez que o ano hidrológico da bacia de estudo se inicia em outubro e termina em setembro (CPRM, 2014; IGAM, 2015; MARQUES, 2010). Este período base foi adotado com o intuito de contemplar a realidade mais atual da bacia, em termos de regime das vazões fluviais.

Na etapa de seleção das estações, a série histórica de cada estação foi submetida à análise de tendência e detecção do provável ano de mudança. Foi aplicado o teste de Pettitt (PETTITT, 1979), um dos mais utilizados e com desempenho satisfatório para as variáveis de vazão e precipitação (ULIANA et al., 2015). As análises de tendência foram realizadas por meio do aplicativo XLSTAT, assumindo um nível de significância de 5%.

Adotou-se como critério a utilização de séries estacionárias com mais de 20 anos de dados e com menos de 5% de falhas, conforme indicação de IGAM (2012). Ressalta-se que os dados brutos também foram considerados neste trabalho, após realização de análise para fins de detecção de erros e/ou inconsistências, pois o uso apenas de dados consistidos implicaria numa defasagem temporal expressiva, visto que na maior parte das estações só se tinha dados de vazão e precipitação consistidos até o ano de 2014 e 2005, respectivamente.

O trabalho foi desenvolvido utilizando-se os dados de 52 estações fluviométricas e 128 estações pluviométricas, cuja distribuição espacial ao longo e no entorno da área de estudo está apresentada na Figura 2.2. Informações detalhadas das estações fluviométricas e pluviométricas selecionadas estão contidas nos Apêndices A e B, respectivamente.

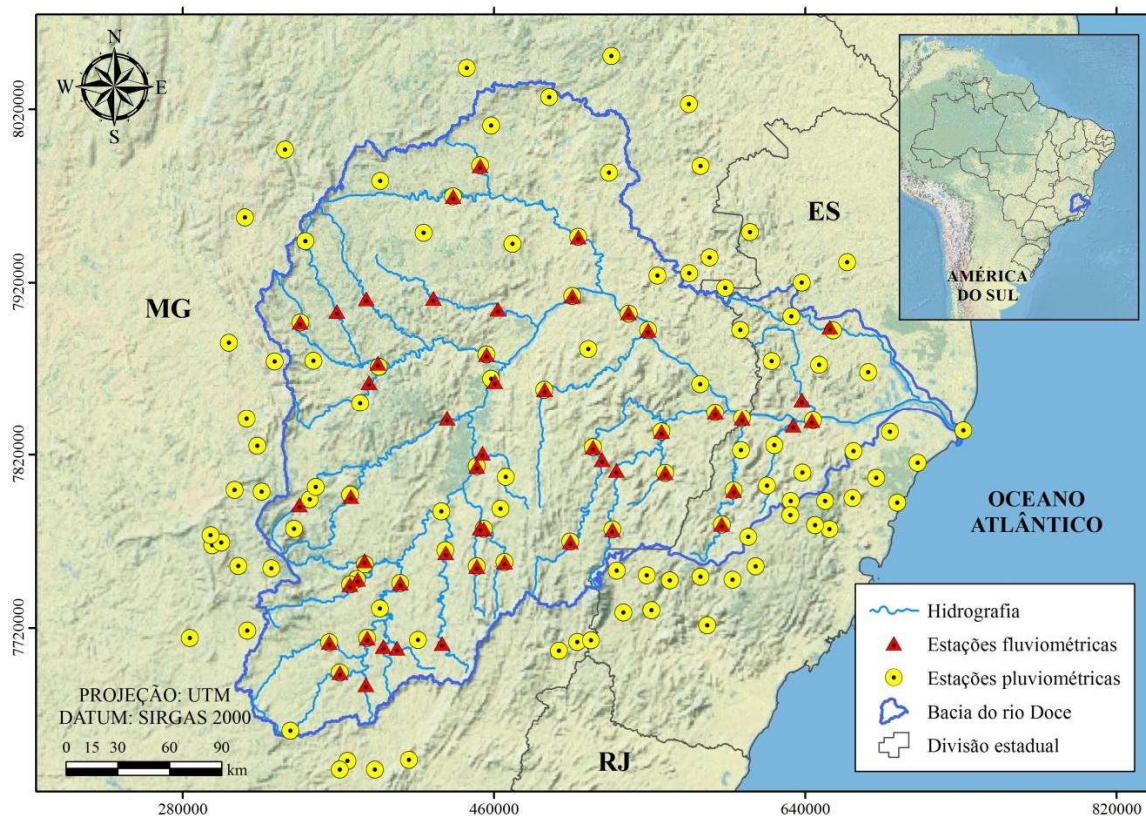


Figura 2.2. Distribuição espacial das estações utilizadas ao longo e no entorno da bacia do rio Doce.

### 2.2.3 Variáveis resposta

As variáveis de interesse do estudo foram as vazões de referência, amplamente adotadas no âmbito da gestão de recursos hídricos no Brasil: vazão mínima de 7 dias consecutivos e período de retorno de 10 anos ( $Q_{7,10}$ ), vazões mínimas associadas às permanências de 95 e 90% do tempo ( $Q_{95}$  e  $Q_{90}$ , respectivamente) e vazão média de longa duração ( $Q_{mld}$ ). Tais variáveis foram calculadas utilizando o SisCAH 1.0 – Sistema Computacional para Análises Hidrológicas (SOUSA et al., 2009), o mesmo software empregado na análise do percentual de falhas anuais das séries históricas, considerando o ano hidrológico da bacia.

### 2.2.4 Covariáveis ambientais

Com base em estudos de modelagem hidrológica em bacias hidrográficas (ARAUJO et al., 2018; ELESBON et al., 2015; PEÑAS et al., 2014; PEÑAS; BARQUÍN; ÁLVAREZ, 2018; PRIETO et al., 2019; SHORTRIDGE; GUIKEMA; ZAITCHIK, 2016; WORLAND; FARMER; KIANG, 2018), considerou-se um conjunto de 76 variáveis ambientais com potencial preditivo para a bacia de estudo (Tabela 2.1), cujas estatísticas descritivas estão no Apêndice D. As covariáveis visaram representar as características da bacia quanto à morfometria, clima, topografia, condições de superfície e uso e cobertura da terra.

Tabela 2.1. Grupos de covariáveis utilizadas como potenciais preditores das variáveis de interesse (MF: morfometria; TP: topografia; CL: clima; UCT: uso/cobertura da terra; S: solo)

MF	TP	CL	MF/CL	UCT*	S
Área	Hmédia	DHmédia	Peq – IDW	Floresta natural (FN)	NC – seco
Perímetro	Hmax	DHmax	Peq – KS	Floresta plantada (FP)	NC – médio
	Hmin	DHmin	Peq750 – IDW	Formação savânica (FS)	NC – úmido
	$\Delta H$	ETmédia	Peq750 – KS	Formação campestre (FC)	USmédia
	Dmédia	ETmin		Pastagens (P)	USmax
	Dmax	ETmax		Agropecuária (AP)	USmin
		ETPmédia		Culturas agrícolas (CAgr)	Latossolo
		ETPmax		Mineração (Min)	Argissolo
		ETPmin		Urbano (Urb)	Cambissolo
		Tmax		Afloramento rochoso (R)	Neossolo
		Tmin		Rios e Lagos (RL)	
		Pa – IDW			
		Pss – IDW			
		Psc – IDW			
		Pa – KS			
		Pss – KS			
		Psc – KS			
		Pmax – IDW			
		Pmax – KS			
		Pmin – IDW			
		Pmin – KS			

\*Cada classe foi representada individualmente pelos anos de 1992, 2002 e 2012.

Hmédia, Hmax e Hmin: altitude média, máxima e mínima.  $\Delta H$ : amplitude altimétrica; Dmédia e Dmax: declividade média e máxima. DHmédia, DHmax e DHmin: deficiência hídrica média, máxima e mínima. ETmédia, ETmax e ETmin: evapotranspiração real média, máxima e mínima. ETPmédia, ETPmax e ETPmin: evapotranspiração potencial média, máxima e mínima. Tmax e Tmin: temperatura do ar máxima e mínima. Pa – IDW, Pss – IDW e Psc – IDW: precipitação média anual (Pa), média do semestre seco (Pss) e média do semestre chuvoso (Psc) interpoladas pelo inverso da distância ao quadrado (IDW). Pa – KS, Pss – KS e Psc – KS: precipitação média anual (Pa), média do semestre seco (Pss) e média do semestre chuvoso (Psc) interpoladas por krigagem simples (KS). Peq – IDW e Peq750 – IDW: Peq e Peq750 com Pa interpolada por IDW; Peq – KS e Peq750 – KS: Peq e Peq750 com Pa interpolada por KS; NC – seco, NC – médio e NC – úmido: número da curva em condição de solo seco, mediana e úmido. USmédia, USmax e USmin: umidade do solo média, máxima e mínima.

Todos os valores das covariáveis foram obtidos para a área de drenagem pertinente a cada uma das 52 estações fluviométricas, requerendo previamente suas respectivas delimitações. Para isso, obteve-se o Modelo Digital de Elevação Hidrograficamente Condicionado (MDEHC) do HydroSHEDS, na resolução de 90 metros (LEHNER; VERDIN; JARVIS, 2006). Após a delimitação das bacias calcularam-se dois atributos físicos: área de drenagem e perímetro. Os dados de elevação do HydroSHEDS também forneceram os dados topográficos das áreas de drenagem, como a variação altimétrica e as declividades do relevo.

A aquisição dos dados pluviométricos se deu através da plataforma HidroWeb/ANA, e outros dados relativos ao clima e ao solo, como temperatura do ar, deficiência hídrica, evapotranspiração real e potencial e umidade do solo foram adquiridos junto ao TerraClimate (ABATZOGLOU et al., 2018). Os dados pluviométricos cedidos pela ANA foram interpolados com base nos métodos do inverso do quadrado da distância (IDW) e por krigagem simples, gerando dados matriciais com resolução de 0,9 km. A resolução espacial dos dados de elevação e precipitação foi adotada com fins de economia de tempo e esforço computacional, tendo em vista a viabilidade de eventual espacialização futura das predições.

Para a extração dos dados de uso e cobertura da terra foram utilizados os mapas da “Coleção 4” do projeto MapBiomas (MAPBIOMAS, 2019). Como o presente estudo tem um período base de 30 anos (outubro de 1988 a setembro de 2018), foram considerados os mapas de uso e cobertura da terra que correspondem aos anos de 1992, 2002 e 2012.

Os percentuais de cada classe de solo presente na bacia foram extraídos do mapeamento realizado para todo o território brasileiro, na escala de 1:5.000.000 (IBGE; EMBRAPA, 2001). Outra característica relativa à superfície da bacia e fornecida aos modelos foi o número da curva, após a aquisição dos dados do Global Curve Number, disponibilizados em resolução de 250 metros (GCN250). Os dados matriciais do GCN250 estão disponíveis apenas nesta resolução e foram produzidos com base em dados de superfície e de propriedades do solo, como textura e profundidade, constituindo uma importante base de informações para a modelagem de vazões (JAAFAR; AHMAD; EL BEYROUTHY, 2019).

Outras duas covariáveis adotadas foram a vazão equivalente ao volume precipitado (Peq) e a vazão equivalente ao volume precipitado menos um fator de abstração correspondente a 750 mm (Peq750) (PRUSKI et al., 2013, 2015). Essas variáveis englobaram duas características da bacia (área de drenagem e precipitação média anual) e têm apresentado bons resultados em estudos recentes de regionalização de vazões (SERRANO et al., 2020).

## 2.2.5 Aplicação dos modelos de aprendizado de máquina

O delineamento das etapas metodológicas foi concebido com a integração de diferentes metodologias, tendo em vista a otimização do processo de predição pelos modelos de aprendizado de máquina. De modo geral, a estrutura metodológica foi constituída pelos seguintes processos interligados: seleção de covariáveis, separação das amostras de validação, ajuste e seleção dos modelos, predição, cálculo dos indicadores estatísticos e classificação qualitativa dos modelos com base nos critérios de Moriasi et al. (2015) (Figura 2.3).

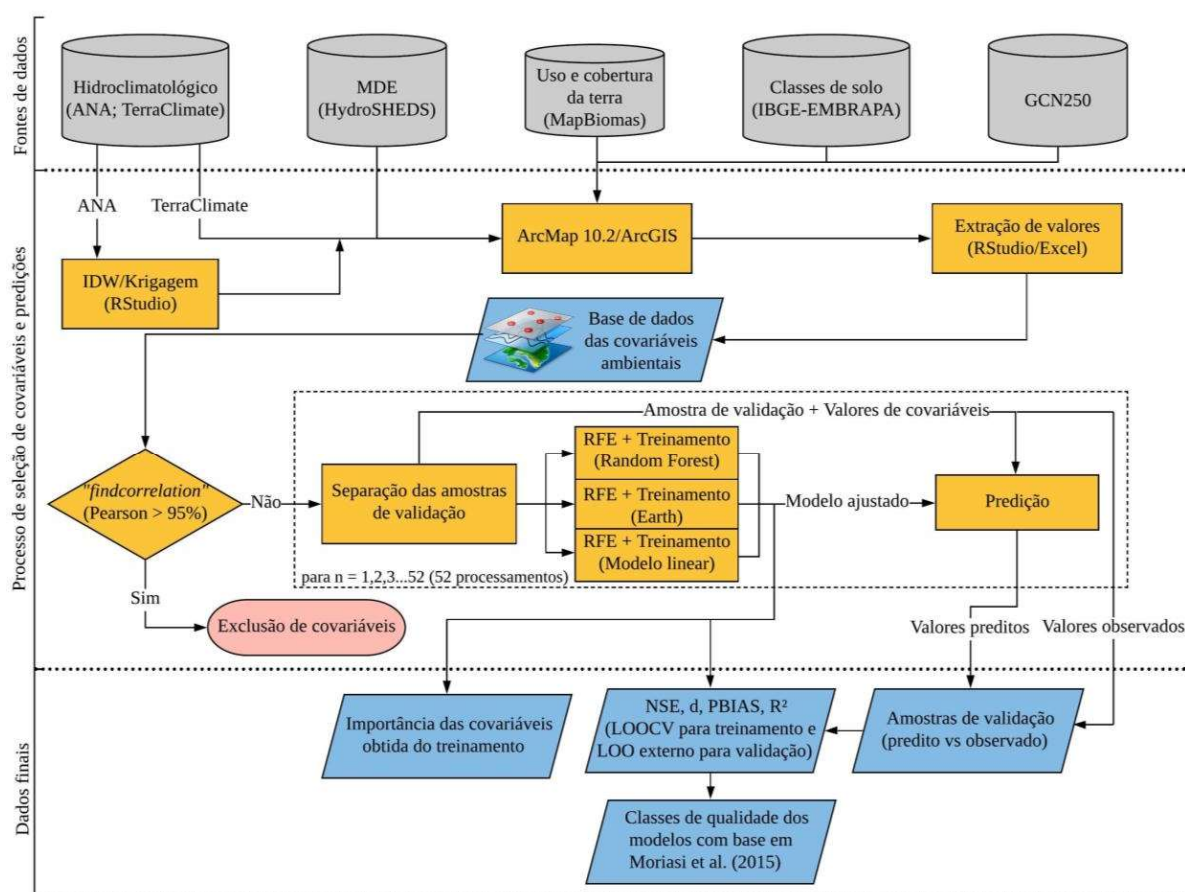


Figura 2.3. Fluxograma metodológico com os processos adotados na avaliação dos modelos

### 2.2.5.1 Seleção das covariáveis ambientais

De posse das covariáveis ambientais e após a extração dos seus valores, a seleção de covariáveis foi realizada por meio de duas etapas subseqüentes: a exclusão de amostras altamente correlacionadas e a eliminação por importância. Estes procedimentos foram

adotados com vistas à maior simplicidade dos modelos de predição de vazões e à economia de tempo e esforço computacional requerido nas etapas posteriores, como o treinamento e a validação dos algoritmos (FORKUOR et al., 2017; KUHN; JOHNSON, 2013).

O primeiro procedimento foi efetuado com a função *findcorrelation* – pacote *caret* (KUHN; JOHNSON, 2013). Esta função quantifica e analisa o valor de correlação entre todos os pares possíveis de covariáveis preditoras, simultaneamente. Então, em cada par com correlação de Pearson acima de 95%, foi excluída a covariável que apresentou o maior somatório de correlações com as demais covariáveis (SOUZA et al., 2018a).

O segundo procedimento realizado foi o *Recursive Feature Elimination* (RFE) – pacote *caret* – sendo um método de seleção de covariáveis do tipo *backward*. Os modelos foram aplicados com todas as combinações possíveis entre as covariáveis preditoras e, por meio do RFE, identificou-se o poder explicativo de cada preditor sobre a variável resposta. Em seguida, o algoritmo ordenou as covariáveis por ordem decrescente de importância e eliminou, iterativamente, aquelas menos promissoras ou explicativas, segundo os critérios de importância admitidos pelos próprios modelos (KUHN; JOHNSON, 2013).

Os modelos de aprendizado de máquina contemplados neste trabalho foram baseados em diferentes abordagens, a saber: modelo baseado em árvores de decisão (Random Forest - RF), Splines de regressão adaptativa multivariada (Earth) e modelo linear (ml). Ressalta-se que cada um dos modelos pode atribuir um nível de importância diferente para um mesmo preditor. Logo, a aplicação do RFE para a seleção das covariáveis ótimas se deu individualmente para cada modelo. Assim, as covariáveis mais importantes para o modelo RF foram selecionadas utilizando RFE-RF, aquelas mais importantes para o Earth foram selecionadas com RFE-Earth e para o modelo linear com RFE-ml (GOMES et al., 2019).

### **2.2.5.2 Treinamento, seleção e validação dos modelos**

O treinamento dos modelos foi realizado após cada processamento do RFE, totalizando 52 etapas de treinamento por modelo. A operação do treinamento se deu da seguinte forma: para a predição de uma amostra  $i$  (omitida do conjunto total de 52 amostras) foi gerado um modelo  $i$ , com treinamento e otimização de hiperparâmetros realizados com base num subconjunto  $i$  de 51 amostras; para a predição de uma amostra  $i+1$  (omitida do conjunto total de dados) foi gerado outro modelo  $(i+1)$  otimizado, baseado em outro subconjunto  $(i+1)$  de 51

amostras; e assim por diante para as 52 amostras. Os processos de eliminação utilizando RFE, treinamento e seleção do modelo ajustado se deram para cada algoritmo, separadamente.

O ajuste dos modelos de aprendizado de máquina pode ser melhorado com certas alterações em seus hiperparâmetros. Os hiperparâmetros do Random Forest, como o número de árvores (ntree) e o número de covariáveis preditoras em cada nó (mtry), assim como aqueles relativos à complexidade de interações (degree) e regularização (penalty) do Earth, foram otimizados automaticamente pelo pacote caret, utilizando-se *grid-search* e *Leave-One-Out Cross-Validation* (LOOCV). O indicador adotado para otimizar os hiperparâmetros, selecionar as covariáveis e treinar os modelos foi o coeficiente de Nash-Sutcliffe (NSE). Ressalta-se que, dos modelos utilizados, apenas o RF e Earth apresentam hiperparâmetros.

Em suma, o treinamento dos modelos foi realizado com a retirada da respectiva amostra de interesse para predição, seguida pela otimização dos hiperparâmetros. Assim, a aferição do desempenho dos modelos nesta etapa se deu pela validação cruzada *Leave-One-Out* (LOOCV) e cálculo dos seguintes indicadores estatísticos: coeficiente de Nash-Sutcliffe (NSE) (NASH; SUTCLIFFE, 1970), conforme Equação 2.1; viés percentual (PBIAS) (Equação 2.2.); índice de concordância *d* proposto por Willmott (1981) (Equação 2.3); coeficiente de determinação ( $R^2$ ) (Equação 2.4); raiz do erro quadrático médio (RMSE) (Equação 2.5); e erro médio absoluto (MAE) (Equação 2.6).

$$NSE = 100 \times \left[ 1 - \frac{\sum(Qobs - Qpred)^2}{\sum(Qobs - \overline{Qobs})^2} \right] \quad (2.1)$$

$$PBIAS = 100 \times \left[ \frac{\sum(Qpred - Qobs)}{\sum(Qobs)} \right] \quad (2.2)$$

$$d = 1 - \left[ \frac{\sum(Qobs - Qpred)^2}{\sum(|Qpred - \overline{Qobs}| + |Qobs - \overline{Qobs}|)^2} \right] \quad (2.3)$$

$$R^2 = \frac{[\sum(Qpred - \overline{Qpred}) \times (Qobs - \overline{Qobs})]^2}{[\sum(Qpred - \overline{Qpred})^2] \times [\sum(Qobs - \overline{Qobs})^2]} \quad (2.4)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \times \sum (Q_{\text{obs}} - Q_{\text{pred}})^2} \quad (2.5)$$

$$\text{MAE} = \frac{1}{n} \times \sum |Q_{\text{pred}} - Q_{\text{obs}}| \quad (2.6)$$

em que  $Q_{\text{obs}}$  é a vazão observada ( $\text{m}^3 \text{s}^{-1}$ );  $Q_{\text{pred}}$  é a vazão predita ( $\text{m}^3 \text{s}^{-1}$ );  $\overline{Q_{\text{obs}}}$  é a média das vazões observadas ( $\text{m}^3 \text{s}^{-1}$ );  $\overline{Q_{\text{pred}}}$  é a média das vazões preditas ( $\text{m}^3 \text{s}^{-1}$ ).

Outro indicador estatístico utilizado para avaliar o desempenho dos modelos foi o Erro Relativo (ER), em percentual, calculado conforme a Equação 2.7. Este indicador permitiu avaliar como a qualidade dos modelos variou em função da magnitude dos valores observados de vazão, possibilitando a extração de informações relevantes, em termos de aplicabilidade.

$$\text{ER (\%)} = 100 \times \left| \frac{(Q_{\text{obs}} - Q_{\text{pred}})}{(Q_{\text{obs}})} \right| \quad (2.7)$$

A etapa de validação foi realizada para cada algoritmo, separadamente, sendo operacionalizada por um *Leave-One-Out* externo (External LOO) que partiu das predições dos modelos de melhor ajuste selecionados na fase de treinamento. O processamento da etapa de validação se deu da seguinte maneira: as covariáveis pertinentes à área de drenagem de uma estação fluviométrica  $i$  foram inseridas no modelo ajustado, gerando um valor predito  $i$ ; depois, as covariáveis pertinentes à área de drenagem de outra estação fluviométrica  $i+1$  foram inseridas no modelo ajustado, gerando outro valor predito  $i+1$ ; e assim sucessivamente para todas as seções com estação fluviométrica. Logo, foram obtidos 52 pares de valores preditos e observados, permitindo o cálculo dos indicadores de qualidade da predição e a classificação qualitativa de Moriasi et al. (2015). O intuito com este tipo de validação foi avaliar a eficiência e aplicabilidade dos modelos em situação próxima à real, considerando a carência de monitoramento fluviométrico em bacias hidrográficas. A adoção e processamento do External LOO na etapa de validação foram baseados em estudo de Souza et al. (2018a).

Resumindo, a fase de treinamento dos modelos se deu com o treinamento de 51 amostras para cada local de predição, com realização das devidas otimizações de hiperparâmetros e seleção de modelo com melhor ajuste baseada no indicador NSE. Já a fase

de validação foi realizada para cada amostra, em que seus valores de covariáveis foram inseridos no modelo ajustado, obtendo-se os pares de valores preditos e observados. Destaca-se que apesar da descrição detalhada de cada etapa ter sido feita separadamente, os processos de treinamento e validação se deram subsequentemente e de forma iterativa e simultânea.

Os indicadores NSE, PBIAS,  $d$  e  $R^2$ , propostos por Moriasi et al. (2015), bem como RMSE e MAE, foram calculados com o pacote ‘hydroGOF’ (ZAMBRANO-BIGIARINI, 2014). Os modelos de aprendizado de máquina foram processados no software R, por meio do pacote ‘randomForest’ (LIAW; WIENER, 2002), ‘earth’ (MILBORROW, 2019) e função “lm” do pacote ‘stats’ para o modelo linear (R DEVELOPMENT CORE TEAM, 2019).

### 2.2.6 Análise comparativa do desempenho

O modelo de melhor desempenho foi aquele que atendeu, em conjunto, aos critérios de classificação dos indicadores estatísticos adotados, a saber: valor de PBIAS mais próximo de zero, valor mais próximo de 100 para NSE, valor de  $R^2$  e  $d$  mais próximos de 1, e valores mais baixos de RMSE e MAE. A classificação qualitativa do desempenho de cada modelo foi auxiliada pelos limites e classes da Tabela 2.2, propostos por Moriasi et al. (2015).

Tabela 2.2 – Classificação qualitativa de índices estatísticos para regionalização de vazões

Indicador	Critério de avaliação do desempenho			
	Muito bom	Bom	Satisfatório	Insatisfatório
NSE (%)	> 80	$70 < NSE \leq 80$	$50 < NSE \leq 70$	$NSE \leq 50$
$R^2$	> 0,85	$0,75 < R^2 \leq 0,85$	$0,60 < R^2 \leq 0,75$	$R^2 \leq 60$
PBIAS (%)	< $\pm 5$	$\pm 5 < PBIAS \leq \pm 10$	$\pm 10 < PBIAS \leq \pm 15$	$PBIAS \geq \pm 15$
$d$	> 0,90	$0,85 < d \leq 0,90$	$0,75 < d < 0,85$	$d \leq 0,75$

Fonte: adaptado de Moriasi et al. (2015).

O RMSE foi incluído como métrica de erro em virtude de constituir um dos indicadores mais utilizados na avaliação da eficiência de modelos (SHORTRIDGE et al., 2016), caracterizando-se pela sensibilidade a *outliers* (MEKANIK; IMTEAZ; TALEI, 2016), além de ser apresentado na mesma unidade de medida da variável resposta. Já o MAE, apesar de também ser apresentado na mesma unidade de medida da variável resposta, não apresenta a referida sensibilidade do RMSE a *outliers*, tendo sido adotado em razão da simplicidade e facilidade de sua interpretação (SHORTRIDGE et al., 2016).

## 2.3 RESULTADOS E DISCUSSÃO

### 2.3.1 Eliminação de covariáveis ambientais

A partir dos processos de eliminação por correlação e importância (RFE), a maior parte das covariáveis ambientais foi excluída. Na Tabela 2.3 são apresentadas apenas as covariáveis que se apresentaram relativamente importantes à predição das vazões mínimas de referência e média de longa duração, utilizando os modelos de aprendizado de máquina.

Tabela 2.3. Covariáveis selecionadas com importância preditiva para as vazões de referência

Modelo	Covariáveis ambientais selecionadas*			
	TP	CL	MF/CL	UCT**
Random Forest	Hmédia	DHmax	Peq750 – KS	CArg – 92
	$\Delta H$	Pa – IDW		CArg – 12
	Dmax	Pmax – IDW	Min – 92	
			Min – 02	
			R – 12	
			RL – 92	
			RL – 02	
			RL – 12	
Earth		Peq750 – KS		
Modelo linear	$\Delta H$	ETmédia		R – 12
	Hmédia	ETmin		AP – 02
	Hmax	ETmax		AP – 12

\*TP: topografia; CL: clima; MF: morfometria; UCT: uso e cobertura da terra.

\*\*CArg – 92 e CArg – 12: cobertura da terra com culturas anuais e perenes mapeada em 1992 e 2012. Min – 92 e Min – 02: atividade de mineração referente aos anos de 1992 e 2002; R – 12: afloramento rochoso do ano de 2012; RL – 92, RL – 02 e RL – 12: espelho d'água de rios e lagos correspondente aos anos de 1992, 2002 e 2012. AP – 02 e AP – 12: cobertura da terra com atividades agropecuárias nos anos de 2002 e 2012.

### 2.3.2 Seleção e importância das covariáveis no ajuste dos modelos preditivos

#### 2.3.2.1 Modelos não lineares

As covariáveis que demonstraram maior poder preditivo para as vazões mínimas de referência ( $Q_{7,10}$ ;  $Q_{95}$  e  $Q_{90}$ ) e média de longa duração ( $Q_{mld}$ ) com o modelo Random Forest foram as seguintes: Peq750 com precipitação média anual interpolada por krigagem simples (Peq750 – KS), cobertura da terra por rios e lagos (RL) referente aos anos de 1992 (RL – 92) e 2002 (RL – 02) e cobertura da terra com culturas anuais e perenes do ano de 1992 (CArg –

92). Na Figura 2.4 são apresentadas as importâncias normalizadas correspondentes a cada covariável ambiental, considerando o emprego do modelo Random Forest.

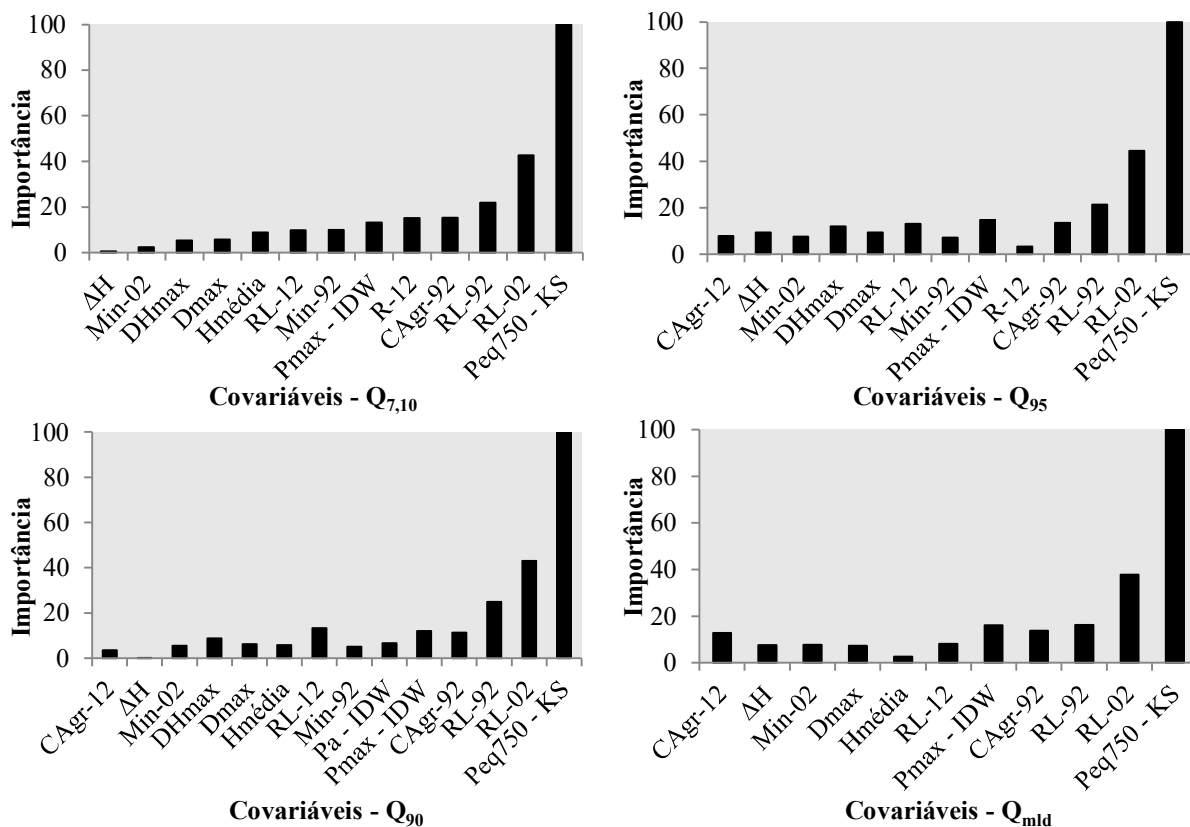


Figura 2.4. Importância normalizada das covariáveis ambientais no modelo Random Forest

Os resultados apresentados na Figura 2.4 sustentam a importância de executar os processos de seleção de covariáveis por correlação e aplicar RFE na seleção das covariáveis importantes à predição. Como a Peq750 está altamente relacionada com área e precipitação, assim como está intrinsecamente associada à Peq, a alta correlação entre tais variáveis foi identificada, e a Peq750 foi detectada como a covariável ambiental que melhor explicou as vazões de referência. Além disso, vale notar o nível de importância atribuído à Peq750, que confirmou a sua estreita relação com o processo físico de formação das vazões, destacada em outros estudos (CESCONETTO et al., 2018; PRUSKI et al., 2013; SERRANO et al., 2020).

As outras covariáveis mais importantes além da Peq750 foram representadas pelos rios e lagos presentes nas áreas das bacias, nos anos de 1992 e 2002. A explicação para estes resultados, com este tipo de covariável ambiental, se apoia na proporção de água superficial ao longo das áreas de drenagem de cada estação fluviométrica, seja pelo escoamento nas calhas dos rios ou armazenamento em reservatórios superficiais. Assim, supõe-se que a

aprendizagem do modelo com esta covariável ambiental seguiu a seguinte relação: quanto mais corpos hídricos na área, conseqüentemente maior a disponibilidade hídrica.

Conforme a Figura 2.4, vale assinalar que a  $Q_{95}$  foi a única vazão de referência que teve a altitude média descartada em termos de importância para a predição pelo modelo Random Forest. Da mesma forma, a deficiência hídrica máxima foi desconsiderada por este modelo na predição da  $Q_{mld}$ . A precipitação média anual ( $P_a$ ) foi explicitamente importante apenas para a predição da  $Q_{90}$ , enquanto sua importância para as demais vazões de referência foi expressada de forma implícita na seleção e importância atribuída à covariável ambiental  $Peq750$ .

A precipitação máxima anual revelou-se substancialmente mais importante à predição das vazões de referência do que as precipitações mínimas. Novaes et al. (2009a) explicam que as vazões em geral não dependem das precipitações do período seco, pois além de serem em menor quantidade, o solo se encontra em elevado déficit de umidade, fazendo com que a precipitação fique retida na zona não saturada do solo e, assim, não contribua com a recarga subterrânea. Assim, conforme Dakhllalla et al. (2016), foi representada a maior capacidade das precipitações máximas em sustentar os escoamentos de base e superficial ao longo do tempo.

A importância atribuída às culturas anuais e perenes, especialmente para a predição das vazões mínimas, oferece um indício da contribuição dessas culturas à permanência da água nas bacias, aumentando a infiltração da água no solo e recarregando os aquíferos. Esta suposição é sustentada pelas constatações do estudo de Schilling et al. (2008), conduzido na bacia agrícola do rio Racoon, região Centro-Oeste de Iowa. Os autores observaram que a expansão de culturas agrícolas nem sempre leva à redução da disponibilidade hídrica da região e, pelo contrário, pode favorecer o seu incremento quando as taxas de evapotranspiração são reduzidas e os índices pluviométricos se mantêm inalterados. Ressalta-se que, apesar do respaldo científico, esta explicação parte de uma análise de indícios. Portanto, estudos adicionais sobre estes processos são requeridos para melhor entendimento.

Apesar de constituir uma variável fisicamente importante no processo hidrológico de formação das vazões (ELESBON et al., 2015; LELIS et al., 2020; LOPES et al., 2017), a precipitação média anual foi considerada apenas para a predição da  $Q_{90}$  pelo Random Forest. Entende-se que a razão da sua eliminação para a maioria das vazões de referência foi a correlação com a  $Peq750$ , cuja importância superou todas as demais covariáveis ambientais.

A classe de afloramento rochoso foi considerada pelo Random Forest apenas na predição da  $Q_{7,10}$  e  $Q_{95}$ . Esta ocorrência expõe certa fragilidade do modelo quanto à capacidade de generalizar a representatividade física que se observou para a  $Peq750$ . No

entanto, apesar dos afloramentos rochosos serem mais explicativos para deflúvios superficiais de maiores magnitudes (NAGESWARA RAO, 2020), este resultado denota a captura da relação existente entre a permeabilidade do solo de uma bacia hidrográfica e seu tempo de concentração. Neste sentido, constatou-se que o modelo foi capaz de considerar a redução do tempo de concentração ao longo da bacia em decorrência de superfícies pouco permeáveis e, assim, associar a classe de afloramento rochoso aos valores observados de vazões mínimas.

Para o algoritmo Earth, verificou-se que o ajuste do modelo para todas as vazões mínimas de referência e média de longa duração se deu com apenas uma covariável, a Peq750. Este resultado reforça a representatividade física e o poder preditivo que esta covariável possui em relação às vazões, tal como constatado para o modelo Random Forest.

### 2.3.2.2 Modelo linear

As covariáveis mais importantes para o ajuste do modelo linear foram bastante diferentes daquelas observadas para os modelos não lineares (Random Forest e Earth). A Peq750 foi a única covariável importante para o ajuste do Earth e a mais importante para o ajuste do Random Forest, mas foi descartada para o ajuste do modelo linear (Figura 2.5).

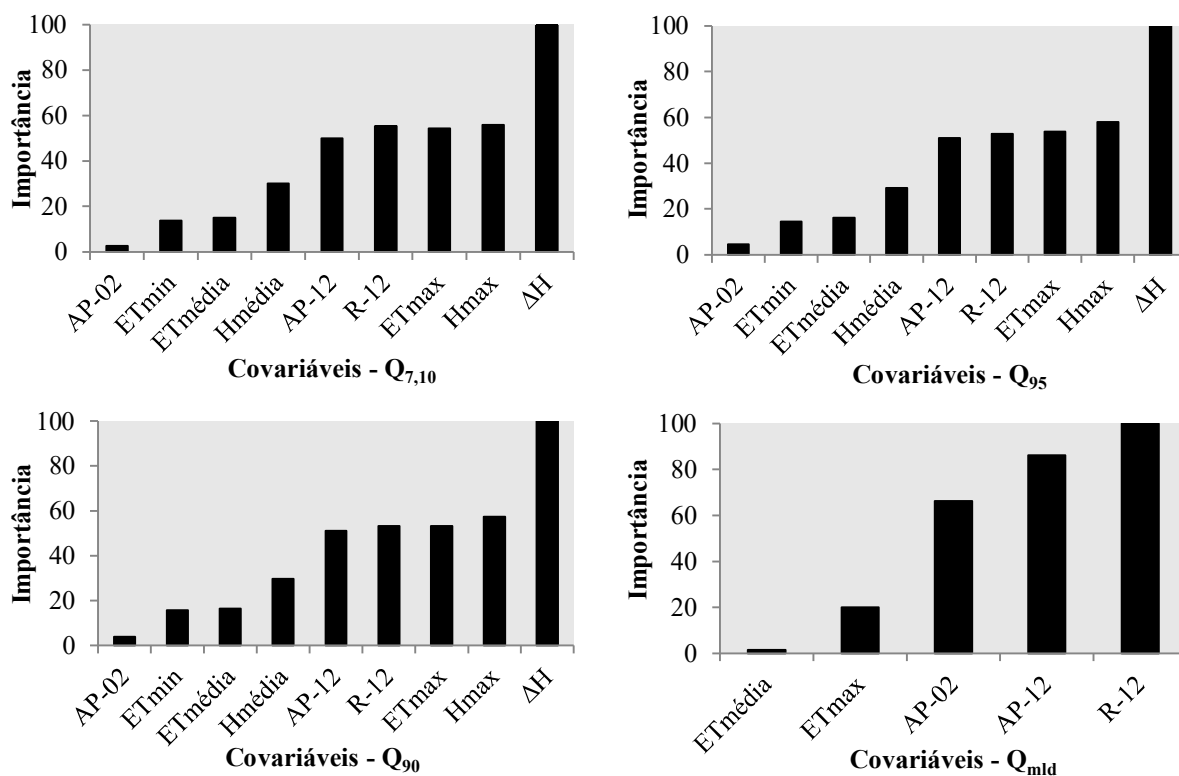


Figura 2.5. Importância normalizada das covariáveis ambientais no modelo linear

Observa-se na Figura 2.5 que o modelo linear foi ajustado com as mesmas covariáveis para todas as vazões mínimas de interesse ( $Q_{7,10}$ ;  $Q_{95}$  e  $Q_{90}$ ), atribuindo as maiores importâncias à amplitude altimétrica seguida da altitude máxima, ETmax e uso e cobertura da terra (R – 12 e AP – 12). Para a predição da  $Q_{mld}$ , porém, as covariáveis correspondentes ao uso e cobertura da terra foram os mais importantes à otimização do ajuste do modelo linear.

O modelo linear capturou relações entre as características topográficas das bacias e a permanência das vazões ao longo do tempo. Estas variáveis mantêm uma relevante interação física entre si, especialmente em razão da influência das elevações sobre os regimes pluviométricos (GONZALEZ-FERRERAS; BARQUÍN, 2017; SUN; THOMPSON; CROKE, 2011), o que revela certa habilidade do modelo para seleção das covariáveis mais importantes. No entanto, os relacionamentos não lineares que governam os processos hidrológicos não são considerados por este modelo. Isto está associado ao grau de distinção entre o conjunto de covariáveis selecionadas por um modelo de abordagem linear e pelos modelos não lineares. Pode-se observar que o modelo linear descartou diversas variáveis que são relatadas por diversos estudos de modelagem hidrológica como altamente explicativas dos regimes de vazões, como área de drenagem e precipitação (ARAUJO et al., 2018; CECÍLIO et al., 2018; ELESBON et al., 2015; IGAM, 2012; PRUSKI et al., 2015; SERRANO et al., 2020).

Nota-se que o modelo linear teve a classe de afloramento rochoso entre as mais importantes para a predição das vazões mínimas de referência, revelando que também foi capaz de compreender a relação existente entre permeabilidade superficial da bacia e o seu tempo de concentração, tal como obtido para o modelo Random Forest.

Para a  $Q_{mld}$ , o modelo linear teve seu ajuste determinado, majoritariamente, pelas covariáveis de uso e cobertura da terra. Estes tipos de covariáveis são bastante importantes à explicação das vazões médias em canais fluviais, pois possuem forte associação física com os componentes hidrológicos (GEBREMICAEL; MOHAMED; VAN DER ZAAG, 2019).

### **2.3.3 Análise e comparação de desempenho**

Para a base de dados utilizada, e conforme os critérios de classificação propostos por Moriasi et al. (2015), os modelos não lineares apresentaram desempenho significativamente superior à predição das vazões em comparação ao desempenho do modelo linear (Tabela 2.4).

Tabela 2.4. Classificação de desempenho dos modelos na predição das vazões de referência

Vazões de referência	Classificação dos modelos preditivos*		
	Random Forest	Earth	Modelo Linear
$Q_{7,10}$	B	S	Ins.
$Q_{95}$	B	MB	Ins.
$Q_{90}$	B	B	Ins.
$Q_{mld}$	B	MB	Ins.

\*MB= “Muito Bom”; B= “Bom”; S= “Satisfatório”; Ins.= “Insatisfatório”.

Os tipos de covariáveis que o modelo linear considerou mais importantes à predição da  $Q_{mld}$  possuem, de fato, uma estreita relação empírica com a física dos processos hidrológicos envolvidos na formação e variação dos regimes de vazões (REN et al., 2002; YANG et al., 2012). Neste sentido, o modelo sugere a forte influência do uso e cobertura da terra sobre a diminuição das vazões médias na bacia do rio Doce constatada por Lyra e Rigo (2019). Por outro lado, é válido ressaltar que o modelo linear teve ajuste muito ruim e aplicabilidade não confiável, sendo considerado insatisfatório, dada a baixa capacidade inerente a estes modelos em lidar com variáveis de comportamento não linear (ARISTIZÁBAL-GIRALDO; VÉLEZ-UPEGUI; MARTÍNEZ-CARVAJAL, 2016; ELLIS; POST, 2004; KIM et al., 2020).

Os valores de NSE pertinentes ao modelo linear foram negativos para a maioria das variáveis de interesse à predição, sendo positivo apenas ao predizer a  $Q_{95}$ . A ocorrência de valores negativos de NSE permite presumir que o simples cálculo da média dos valores observados proporciona predições mais confiáveis do que o modelo avaliado (FARMER, 2016; NKONYA et al., 2014). Apesar da  $Q_{95}$  ter sido predita com NSE positivo, o valor desta métrica foi de 17,14%, inferior ao limite mínimo (50%) sugerido por Moriasi et al. (2015).

Na predição da  $Q_{7,10}$ , o valor de PBIAS do Random Forest foi equivalente a cerca de 0,2%, o menor viés percentual obtido no estudo dentre os modelos avaliados (Tabela 2.5). Já os valores de PBIAS associados aos modelos Earth e linear denotam um viés de subestimativa considerável, sinalizando alguns riscos na adoção pela gestão dos recursos hídricos. Dentre os possíveis problemas de modelos com viés de subestimativa está a redução das concessões de outorgas e conseqüente subutilização da água, gerando impactos socioeconômicos e conflitos entre usuários e órgãos gestores de recursos hídricos (BORK, 2018; OLIVEIRA, 2008).

Tabela 2.5. PBIAS dos modelos de aprendizado de máquina usados nas predições das vazões mínimas de referência e média de longa duração na bacia do rio Doce

Vazões de referência	PBIAS (%)		
	Random Forest	Earth	Modelo Linear
Q <sub>7,10</sub>	0,20	-11,00	-11,40
Q <sub>95</sub>	-1,50	0,30	2,80
Q <sub>90</sub>	-0,60	3,30	-2,40
Q <sub>mld</sub>	-2,30	-3,60	11,30

Com base na Tabela 2.5 pode-se verificar que o Random Forest foi superior ao Earth na predição da Q<sub>7,10</sub> em relação ao PBIAS. Entretanto, ambos os modelos demonstraram desempenho relativamente similar, com baixo viés percentual na predição da Q<sub>95</sub>, Q<sub>90</sub> e Q<sub>mld</sub>.

Na Figura 2.6 são apresentados os valores dos indicadores estatísticos d, NSE e R<sup>2</sup> obtidos pelos modelos de aprendizado de máquina na predição das vazões mínimas de referência e média de longa duração na bacia do rio Doce.

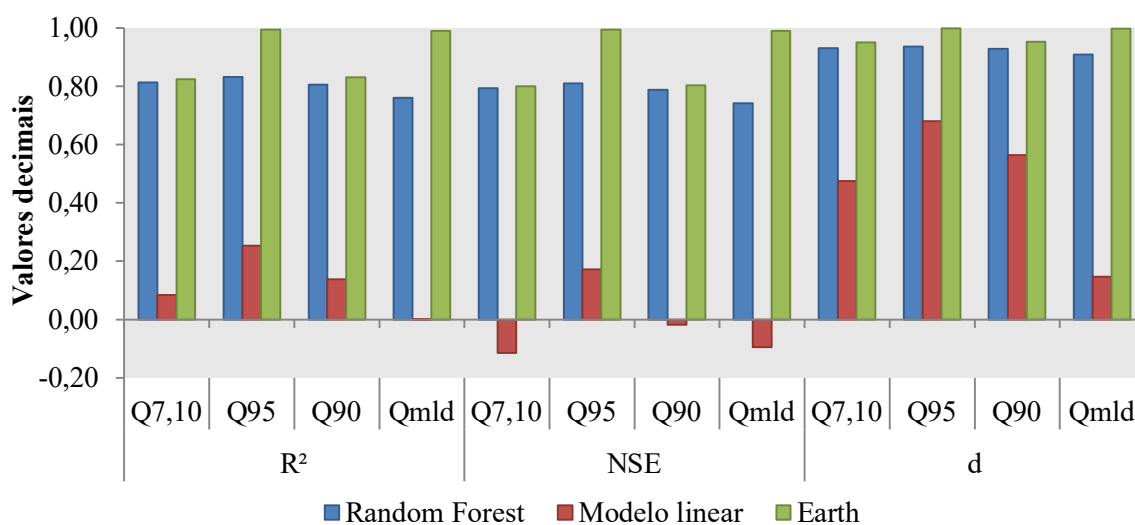


Figura 2.6. Valores dos indicadores estatísticos (d, NSE e R<sup>2</sup>), em decimais, obtidos pelos modelos de aprendizado de máquina na predição das vazões mínimas de referência e média de longa duração na bacia do rio Doce.

Quanto ao indicador R<sup>2</sup>, ressalta-se que seu emprego para avaliar a acurácia de métodos de modelagem somente é admitido se os dados estiverem condicionados à linha de regressão ajustada em 1:1. Neste estudo, como esta condição não existe, o papel do R<sup>2</sup> foi capturar a eficiência dos métodos em representar a tendência de comportamento e a proporção de

variação dos valores de vazão observada entre as seções de monitoramento. Portanto, consideraram-se os indicadores NSE e índice de concordância de Willmott (d) como apropriados ao propósito de comparar o desempenho dos modelos preditivos em relação à proximidade entre valores preditos e observados.

Com base na Figura 2.6, verifica-se que o modelo Earth superou os outros modelos quanto ao  $R^2$ , NSE e d para todas as vazões de referência. Ainda, constatou-se que, em geral, o Earth também superou os outros modelos quanto ao RMSE e MAE na predição das vazões mínimas de referência e média de longa duração (Figura 2.7). Logo, à exceção do viés de subestimativa considerável da  $Q_{7,10}$  (Tabela 2.5), o Earth foi o modelo que se demonstrou estatisticamente mais confiável à predição das vazões de interesse na bacia do rio Doce.

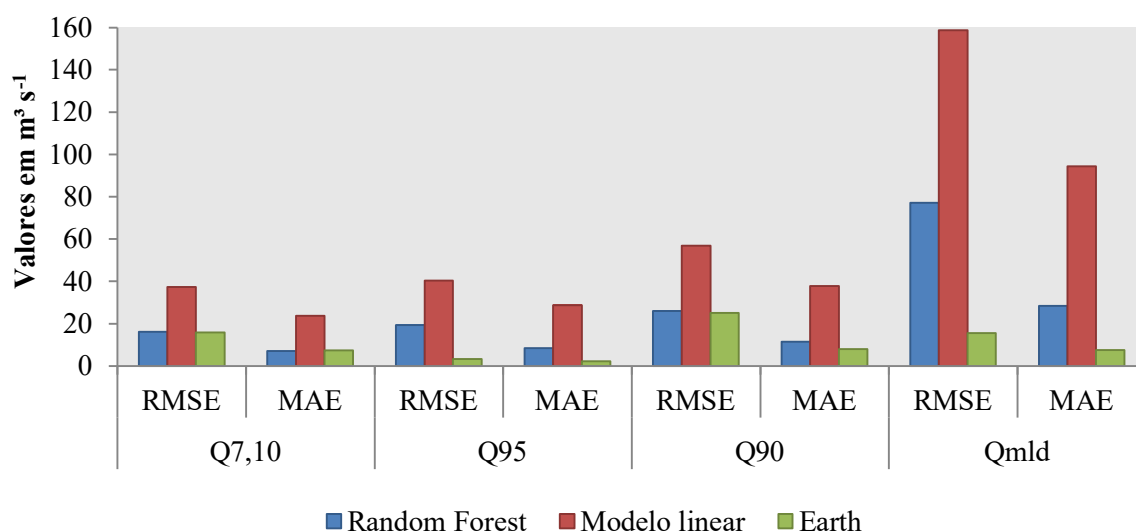


Figura 2.7. Valores dos indicadores RMSE e MAE, em  $m^3 s^{-1}$ , obtidos pelos modelos de aprendizado de máquina na predição das vazões mínimas de referência e média de longa duração na bacia do rio Doce.

Vale notar que algumas predições da  $Q_{7,10}$  e  $Q_{90}$  pelo modelo Earth resultaram em valores negativos, o que não ocorreu com as predições do modelo Random Forest (Figura 2.8). Embora seja óbvio, vale salientar a impossibilidade física de ocorrerem vazões negativas na natureza, conforme destacado por Irving et al. (2018). Neste sentido, o Earth revelou que não é recomendado para cerca de 21,15 e 7,69% das seções fluviais consideradas na predição da  $Q_{7,10}$  e  $Q_{90}$ , respectivamente, dada a ocorrência de vazões negativas nestas áreas.

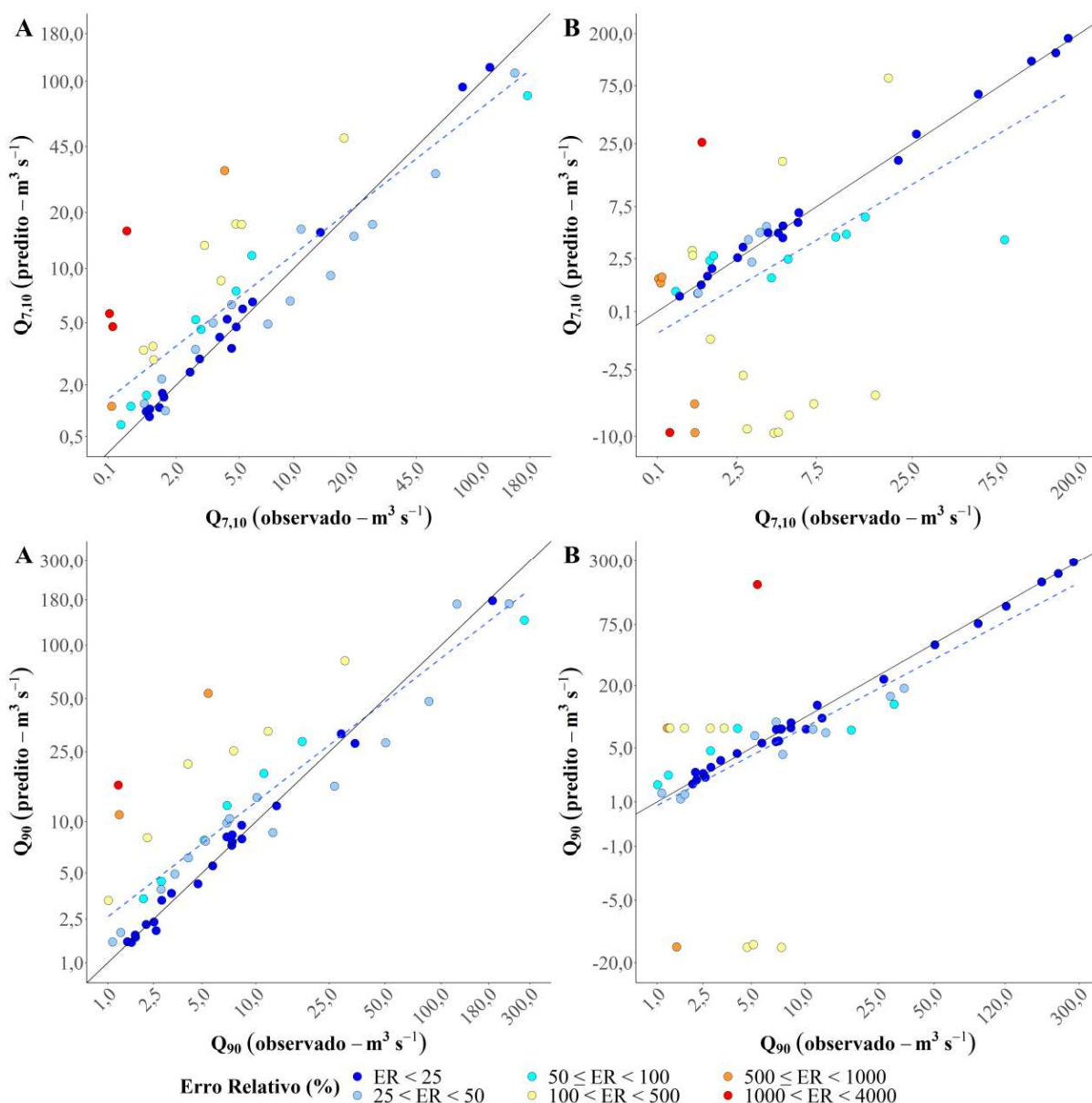


Figura 2.8. Gráficos de dispersão com valores de ER pertinentes às predições da  $Q_{7,10}$  e  $Q_{90}$  utilizando os modelos Random Forest (A) e Earth (B) (linhas sólidas – linhas de inclinação 1:1; linhas tracejadas – tendência da regressão linear).

Conforme a Figura 2.8, as predições da  $Q_{7,10}$  pelo modelo Earth forneceram, em geral, valores preditos menores do que os valores observados. A dispersão dos dados para esta vazão mínima de referência sugere que o modelo Earth apresenta um viés de subestimativa dos valores observados considerável, enquanto reflete um baixo viés de superestimativa da  $Q_{90}$ , concordando com os valores obtidos de PBIAS. Ainda pode-se verificar na Figura 2.8, tal como na Figura 2.9, a ocorrência de certa variação de desempenho dos modelos Random Forest e Earth em função das magnitudes dos valores de vazão observada.

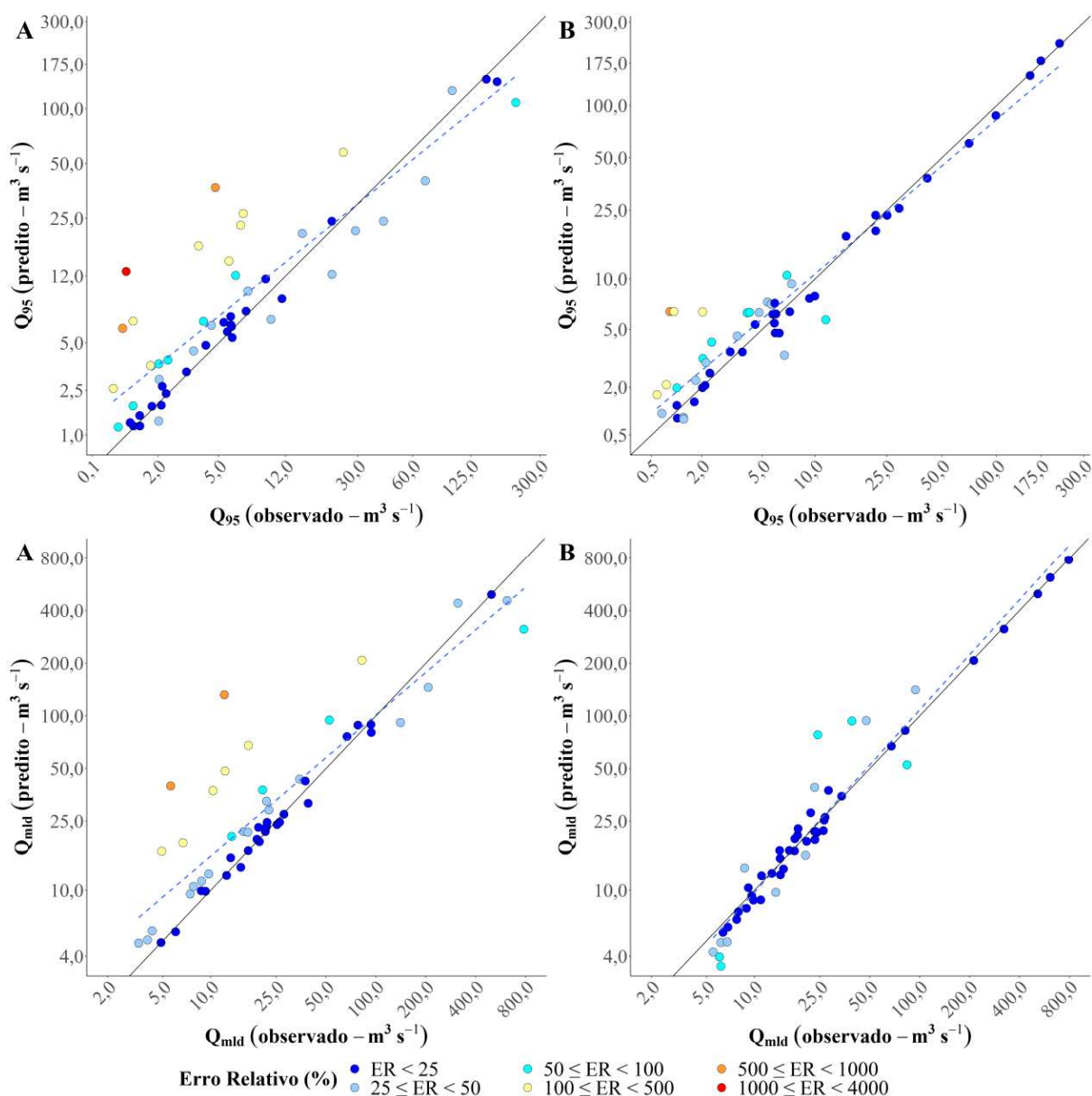


Figura 2.9. Gráficos de dispersão com valores de ER pertinentes às predições da  $Q_{95}$  e  $Q_{mld}$  utilizando os modelos Random Forest (A) e Earth (B) (linhas sólidas – linhas de inclinação 1:1; linhas tracejadas – tendência da regressão linear).

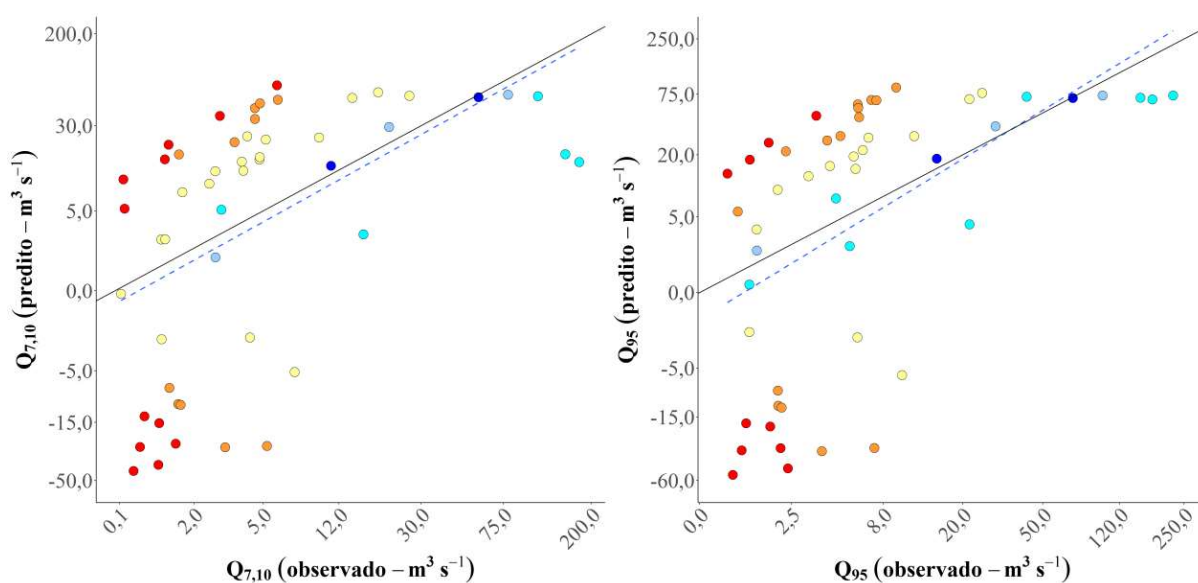
Em geral, os maiores valores de ER ocorreram em seções de canais fluviais com valores mais baixos de vazão observada, localizados em grande parte nas proximidades das regiões de cabeceira, em concordância com os resultados obtidos por outros autores (LISBOA et al., 2008; NOVAES et al., 2009b). A explicação desta ocorrência está associada ao fato de que a regularização natural das bacias tende a ser mais pronunciada em áreas de drenagem maiores, de modo que a expectativa em áreas de drenagem menores é que ocorram maiores variações

dos deflúvios superficiais (CRUZ; TUCCI, 2008; MOREIRA; SILVA; 2014). Constatou-se, ainda, por análise espacial, que estes erros estiveram fortemente relacionados a regiões com baixa densidade de estações fluviométricas, como o norte do estado de Minas Gerais e nas áreas próximas à foz da bacia, no estado do Espírito Santo. Isto faz sentido, tendo em vista que o desempenho dos modelos de aprendizado de máquina será tanto melhor quanto maior o número de amostras de treinamento (KAMPICHLER et al., 2010).

Com base nas Figuras 2.8 e 2.9, a semelhança de inclinação e proximidade entre as linhas de tendência da regressão linear em relação às linhas de inclinação 1:1 sustentam a classificação qualitativa que se obteve com base nos critérios propostos por Moriasi et al. (2015). Verificou-se que o modelo Random Forest superou o modelo Earth, em termos de desempenho, apenas na predição da  $Q_{7,10}$  e em certos locais da bacia na predição da  $Q_{90}$ , onde o modelo Earth apresentou erros de predição que resultaram em valores negativos de vazão.

Pode-se observar, ainda, nas Figuras 2.8 e 2.9, que os modelos não lineares não apresentaram enviesamento expressivo de subestimativa ou superestimativa, exceto o modelo Earth na predição da  $Q_{7,10}$ , corroborando com os resultados referentes ao PBIAS. Isto reforça a sugestão de que estes modelos revelaram-se ferramentas potencialmente adequadas ao apoio da gestão de recursos hídricos na bacia do rio Doce, tendo em vista seus bons desempenhos, mesmo para uma rede de monitoramento fluviométrico relativamente pouco densa.

As predições do modelo linear foram as mais críticas para todas as vazões mínimas de referência e média de longa duração, o que foi apontado tanto pela classificação qualitativa de desempenho do modelo, quanto pelos valores elevados de erro relativo das predições, refletidos na discrepância entre as linhas da regressão linear e as linhas 1:1 (Figura 2.10).



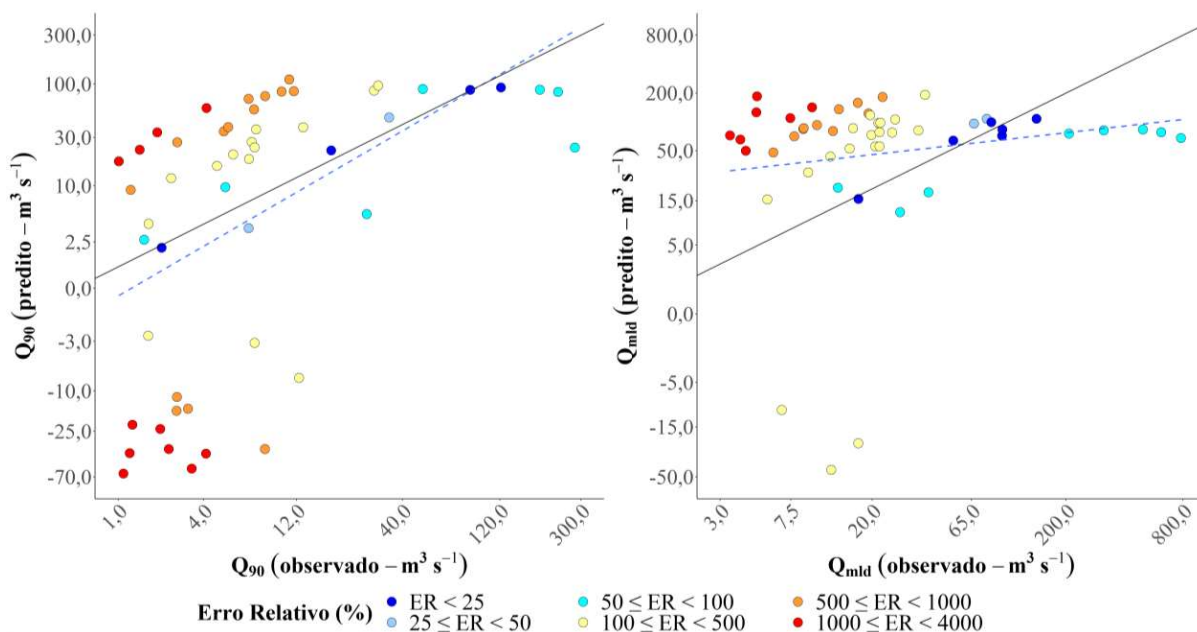


Figura 2.10. Gráficos de dispersão com valores de ER pertinentes às predições das vazões mínimas de referência e média de longa duração utilizando o modelo linear (linhas sólidas – linhas de inclinação 1:1; linhas tracejadas – tendência da regressão linear).

A Figura 2.10 permite constatar como o modelo linear foi ineficiente na predição dos valores observados de cada vazão de referência. A dispersão dos valores preditos e observados permite verificar que o ajuste do modelo linear foi muito ruim. De acordo com os indicadores estatísticos obtidos, o modelo linear apenas satisfaz o limite “satisfatório” dos critérios de Moriasi et al. (2015) quanto ao PBIAS. Cabe ressaltar que a análise isolada deste indicador não é confiável, pois pode ocasionar enganos na avaliação dos modelos, dado que seus valores se anulam em casos de tendência opostas, gerando um valor de viés percentual próximo de zero, ainda que as predições tenham sido muito ruins (MORIASI et al., 2007).

Os resultados do presente trabalho comprovam que os modelos de aprendizado de máquina Random Forest e Earth apresentaram bom desempenho. Entretanto, estudos adicionais são recomendados para uma investigação mais aprofundada a respeito dos impactos no desempenho destes modelos ao serem ajustados para bacias de áreas distintas e com diferentes densidades de estações. Além de complementar os resultados deste trabalho, tais estudos seriam de grande utilidade à gestão dos recursos hídricos, especialmente em bacias menores de países em desenvolvimento, onde o monitoramento hidrológico é precário.

## 2.4 CONCLUSÕES

Com base nos resultados obtidos conclui-se que:

- O modelo de aprendizado de máquina linear não apresentou desempenho satisfatório na predição de vazões mínimas de referência e média de longa duração, sendo contraindicada a sua utilização para fins de regionalização das vazões na bacia do rio Doce.
- A vazão equivalente ao volume precipitado com abstração de 750 mm (Peq750) foi a covariável ambiental que apresentou, para todas as vazões, a contribuição mais importante ao melhor desempenho dos modelos Random Forest e Earth.
- Os modelos Random Forest e Earth apresentaram desempenhos similares e grande habilidade em prever vazões mínimas de referência e média de longa duração na bacia do rio Doce, porém, recomenda-se a adoção do Random Forest para a predição da  $Q_{7,10}$ , devido ao seu melhor desempenho, bem como para a predição da  $Q_{90}$  em locais específicos, onde ocorreram predições de valores negativos pelo Earth, apesar deste modelo ter sido ligeiramente superior para as demais situações.
- Os modelos não lineares de aprendizado de máquina constituem alternativas poderosas e promissoras para fins de regionalização de vazões em apoio à gestão e ao planejamento integrado dos recursos hídricos em nível de bacia hidrográfica.

## 2.5 REFERÊNCIAS

- ABATZOGLOU, J. T. et al. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. **Scientific Data**, v. 5, p. 1–12, 2018.
- ADNAN, R. M. et al. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. **Journal of Hydrology**, 2019.
- AGERH. **Comitês de bacias hidrográficas**. Disponível em:  
<<https://agerh.es.gov.br/documentos-dos-comites>>. Acesso em: 09 jul. 2020.
- AL-SUDANI, Z. A.; SALIH, S. Q.; MUNDHER, Z. Development of multivariate adaptive regression spline integrated with differential evolution model for stream flow simulation. **Journal of Hydrology**, v. 573, p. 1–12, 2019.
- ANA. **Encarte especial sobre a bacia do rio Doce: rompimento da barragem em Mariana/MG Conjuntura dos Recursos Hídricos no Brasil - Informe 2015**. 2016. 49 p.
- ANA. **Base hidrográfica Ottocodificada da bacia do rio Doce 1:50.000/1.100.000**. Disponível em: <<http://metadados.ana.gov.br/geonetwork/srv/pt/main.home>>. Acesso em: 31 maio. 2019.
- ARAÚJO, F. C. et al. Streamflow regionalization in Piquiri river basin. **Engenharia Agrícola**, v. 4430, p. 22–31, 2018.
- ARISTIZÁBAL-GIRALDO, E. V.; VÉLEZ-UPEGUI, J. I.; MARTÍNEZ-CARVAJAL, H. E. A comparison of linear and nonlinear model performance of SHIA\_Landslide: A forecasting model for rainfall-induced landslides. **Revista Facultad de Ingeniería**, v. 80, p. 74–88, 2016.
- BORK, C. K. **Regionalização de vazões mínimas para o estado do Rio Grande do Sul**. 2018. 156 f. Dissertação (Mestrado em Recursos Hídricos) – Universidade Federal de Pelotas, Londrina, 2018.
- CAVALCANTI, I. F. DE A. et al. **Tempo e clima no Brasil**. 1. ed. São Paulo: Oficina de Textos, 2009. 464 p.
- CBH-DOCE. **A bacia do rio Doce**. Disponível em:  
<<http://www.cbhdoce.org.br/institucional/a-bacia>>. Acesso em: 28 fev. 2019a.
- CBH-DOCE. **Deliberação Normativa CBH-Doce nº 51/2016**. 2016. Disponível em:

- <<http://www.cbhdoce.org.br/wp-content/uploads/2016/12/Delibera%C3%A7%C3%A3o-051-Ad-Referendum-Aprova-Realoca%C3%A7%C3%A3o-do-PAP.pdf>>. Acesso em: 28 fev. 2020b.
- CECÍLIO, R. A. et al. Evaluation of low-flow and long term average flow regionalization methods on Itapemirim river watershed. **Scientia Agraria**, v. 19, n. 2, p. 122–132, 2018.
- CESCONETTO, L. B. et al. Potentiality of sugarcane expansion under irrigation conditions considering natural and potential water availability. **Agricultural Water Management**, v. 203, p. 162–171, 2018.
- CPRM. **Acompanhamento da estiagem na região Sudeste do Brasil – Relatório 6: Área de atuação da Superintendência Regional da CPRM de Belo Horizonte**. 2014.
- CRUZ, J. C.; TUCCI, C. E. M. Estimativa da disponibilidade hídrica através da curva de permanência. **Revista Brasileira de Recursos Hídricos**, Porto Alegre, v.13, n.1, p.111-124, 2008.
- CUPOLILLO, F.; ABREU, M. L. DE; VIANELLO, R. L. Climatologia da Bacia do Rio Doce e sua Relação com a Topografia Local. **Geografias**, v. 4, n. 1, p. 45–60, 2008.
- DAKHLALLA, A. O. et al. Evaluating the impacts of crop rotations on groundwater storage and recharge in an agricultural watershed. **Agricultural Water Management**, v. 163, p. 332–343, 2016.
- ECOPLAN-LUME. **Plano Integrado de Recursos Hídricos da Bacia Hidrográfica do Rio Doce - Volume I**. Consórcio ECOPLAN-LUME, 2010. 478 p.
- ELESBON, A. A. A. et al. Multivariate statistical analysis to support the minimum streamflow regionalization. **Engenharia Agrícola**, v. 35, n. 5, p. 838–851, 2015.
- ELLIS, A. M.; POST, E. Population response to climate change: linear vs. non-linear modeling approaches. **BMC Ecology**, v. 4, p. 1–9, 2004.
- ELSHORBAGY, A. et al. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology. **Hydrology and Earth System Sciences**, v. 14, p. 1931–1941, 2010a.
- ELSHORBAGY, A. et al. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2: Application. **Hydrology and Earth System Sciences**, v. 14, p. 1943–1961, 2010b.
- FARMER, W. H. Ordinary kriging as a tool to estimate historical daily streamflow records. **Hydrology and Earth System Sciences**, v. 20, p. 2721–2735, 2016.

- FORKUOR, G. et al. High resolution mapping of soil properties using Remote Sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models. **PLoS ONE**, v. 12, n. 1, p. 1–21, 2017.
- GHORBANI, M. A. et al. A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. **Environmental Earth Sciences**, v. 75, p. 1–14, 2016.
- GOMES, L. C. et al. Modelling and mapping soil organic carbon stocks in Brazil. **Geoderma**, v. 340, p. 337–350, 2019.
- GONZALEZ-FERRERAS, A. M.; BARQUÍN, J. Mapping the temporary and perennial character of whole river networks. **Water Resources Research**, v. 53, p. 6709–6724, 2017.
- IBGE. **Censo Demográfico 2010**. Disponível em:  
<<https://censo2010.ibge.gov.br/resultados.html>>. Acesso em: 28 fev. 2020.
- IBGE; EMBRAPA. **Mapa de solos do Brasil**. Rio de Janeiro: IBGE, 2001. Disponível em:  
<[http://www.dpi.inpe.br/Ambdata/mapa\\_solos.php](http://www.dpi.inpe.br/Ambdata/mapa_solos.php)> Acesso em : 18 dez. 2019.
- IGAM. **Estudo de regionalização de vazão para o aprimoramento do processo de outorga no Estado de Minas Gerais**. Belo Horizonte: IGAM, 2012. 415 p.
- IGAM. **Relatório hidrometeorológico: período chuvoso 2014/2015**. Belo Horizonte: IGAM. 2015. 30 p.
- IRVING, K. et al. Data Descriptor: a high-resolution stream flow and hydrological metrics dataset for ecological modeling using a regression model. **Scientific Data**, v. 5, p. 1–14, 2018.
- JAAFAR, H. H.; AHMAD, F. A.; EL BEYROUTHY, N. GCN250, new global gridded curve numbers for hydrologic modeling and design. **Scientific data**, v. 6, n. 145, p. 1–9, 2019.
- JURKOVIC, Z. et al. A comparison of machine learning methods for cutting parameters prediction in high speed turning process. **Journal of Intelligent Manufacturing**, v. 29, p. 1683–1693, 2018.
- KAMPICHLER, C. et al. Ecological Informatics Classification in conservation biology: a comparison of five machine-learning methods. **Ecological Informatics**, v. 5, p. 441–450, 2010.
- KIM, Y. J. et al. Comparison of linear and nonlinear statistical models for analyzing determinants of residential energy consumption. **Energy and Buildings**, v. 223, p. 1–13, 2020.

- KISI, O. et al. Incorporating synoptic-scale climate signals for streamflow modelling over the Mediterranean region using machine learning models. **Hydrological Sciences Journal**, v. 64, n. 10, p. 1240–1252, 2019.
- KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. New Yourk: Springer, 2013.
- LEHNER, B.; VERDIN, K.; JARVIS, A. **HydroSHEDS Technical Documentation** Washington, DC World Wildlife Fund (WWF). 2006. Disponível em: <<http://hydrosheds.cr.usgs.gov>> Acesso em: 18 dez. 2019.
- LELIS, L. C. et al. Assessment of hydrological regionalization methodologies for the upper Jaguari River basin. **Journal of South American Earth Sciences**, v. 97, p. 1–10, 2020.
- LIAW, A.; WIENER, M. Classification and regression by randomForest. **R News**, v. 2, p. 18–22, 2002.
- LOPES, T. R. et al. Regionalization of maximum and minimum flow in the Teles Pires basin, Brazil. **Journal of the Brazilian Association of Agricultural Engineering**, v. 37, n. 1, p. 54–63, 2017.
- LOUZADA, F. L. R. D. O.; XAVIER, A. C.; PEZZOPANE, J. E. M. Climatological water balance with data estimated by tropical rainfall measuring mission for the Doce river basin. **Engenharia Agrícola**, v. 38, n. 3, p. 376–386, 2018.
- LISBOA, L.; MOREIRA, M. C.; SILVA, D. D. da.; PRUSKI, F. F. Estimativa e regionalização das vazões mínimas e média na bacia do rio Paracatu. **Engenharia na Agricultura**, Viçosa, v.16, n.4, p.471-479, 2008.
- LYRA, B. U.; RIGO, D. Deforestation impact on discharge regime in the Doce River Basin. **Revista Ambiente & Água**, v. 14, n. 4, p. 1–11, 2014.
- MAPBIOMAS. **Coleção 4 da série anual de mapas de cobertura e uso de solo do Brasil**. 2019. Disponível em: <<https://mapbiomas.org/>>. Acesso em: 15 maio. 2019.
- MARQUES, F. A. **Sistema de controle dinâmico para a gestão dos usos múltiplos da água**. 2010. 234 f. Tese (Doutorado em Engenharia Agrícola) – Universidade Federal de Viçosa, Viçosa. 2010.
- MEKANIK, F.; IMTEAZ, M. A.; TALEI, A. Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals. **Climate Dynamics**, v. 46, p. 3097–3111, 2016.
- MILBORROW, S. **earth: Multivariate Adaptive Regression Splines R Package**, 2019. Disponível em: <<https://cran.r-project.org/web/packages/earth/earth.pdf>> Acesso em: 28 nov. 2019.

- MOREIRA, M. C.; DAVID, D. Análise de métodos para estimativa das vazões da bacia do rio Paraopeba. **Revista Brasileira de Recursos Hídricos**, v. 19, n. 2, p. 313–324, 2014.
- MORIASI, D. N. et al. Model evaluation guidelines for systematic quantification of accuracy in watershed simulation. **Transactions of the ASABE**, v. 50, n. 3, p. 885–900, 2007.
- MORIASI, D. N. et al. Hydrologic and water quality models: Performance measures and evaluation criteria. **Transactions of the ASABE**, v. 58, n. 6, p. 1763–1785, 2015.
- NAGESWARA RAO, K. Analysis of surface runoff potential in ungauged basin using basin parameters and SCS-CN method. **Applied Water Science**, v. 10, n. 47, p. 1–16, 2020.
- NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models part I — A discussion of principles. **Journal of Hydrology**, v. 10, n. 3, p. 282–290, 1970.
- NKONYA, E. et al. **Assessing the economic benefits of sustainable land management practices in Bhutan**. Washington, DC: IFPRI. 2014. 54 p.
- NOVAES, L. F. DE et al. Modelo para a Quantificação da Disponibilidade Hídrica: Parte 2 - Análise do Comportamento do Modelo para a Estimativa da Q7,10 na Bacia do Paracatu. **Revista Brasileira de Recursos Hídricos**, v. 14, n. 1, p. 27–39, 2009a.
- NOVAES, L. F.; PRUSKI, F. F.; PEREIRA, S. B.; QUEIROZ, D. O.; RODRIGUEZ, R. G. Gestão de recursos hídricos: uma nova metodologia para a estimativa das vazões mínimas. **Engenharia na Agricultura**, Viçosa, MG, v.17, n.1, p.62-74, 2009b.
- OLIVEIRA, F. A. **Procedimentos para aprimorar a regionalização de vazões: estudo de caso da bacia do rio Grande**. 2008. 173 f. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Federal de Viçosa, Viçosa, 2008.
- OLIVEIRA, K. S. S.; QUARESMA, V. Temporal variability in the suspended sediment load and streamflow of the Doce River. **Journal of South American Earth Sciences**, v. 78, p. 101–115, 2017.
- PAPACHARALAMPOUS, G.; TYRALIS, H.; KOUTSOYIANNIS, D. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. **Stochastic Environmental Research and Risk Assessment**, v. 33, p. 481–514, 2019.
- PATEL, S. S.; RAMACHANDRAN, P. A Comparison of Machine Learning Techniques for Modeling River Flow Time Series: The Case of Upper Cauvery River Basin. **Water Resources Management**, v. 29, p. 589–602, 2015.
- PEÑAS, F. J. et al. The influence of methodological procedures on hydrological classification performance. **Hydrology and Earth System Sciences**, v. 18, p. 3393–3409, 2014.

- PEÑAS, F. J.; BARQUÍN, J.; ÁLVAREZ, C. A comparison of modeling techniques to predict hydrological indices in ungauged rivers. **Limnetica**, v. 37, n. 1, p. 145–158, 2018.
- PETTITT, A. N. A Non-Parametric approach to the Change-Point problem. **Applied Statistics**, v. 28, n. 2, p. 126–135, 1979.
- PRIETO, C. et al. Flow Prediction in Ungauged Catchments Using Probabilistic Random Forests Regionalization and New Statistical Adequacy Tests. **Water Resources Research**, v. 55, p. 4364–4392, 2019.
- PRUSKI, F. F. et al. Improved regionalization of streamflow by use of the streamflow equivalent of precipitation as an explanatory variable. **Journal of Hydrology**, v. 476, p. 52–71, 2013.
- PRUSKI, F. F. et al. Low-flow estimates in regions of extrapolation of the regionalization equations: A new concept. **Engenharia Agrícola**, v. 35, n. 5, p. 808–816, 2015.
- R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. 2019. Vienna: R Foundation for Statistical Computing. Disponível em: <<https://www.rproject.org/>>. Acesso em: 28 nov. 2019.
- REN, L. et al. Impacts of human activity on river runoff in the northern area of China. **Journal of Hydrology**, v. 261, p. 204–217, 2002.
- SCHILLING, K. E. et al. Impact of land use and land cover change on the water balance of a large agricultural watershed: historical effects and future directions. **Water Resources Research**, v. 44, p. 1–12, 2008.
- SERRANO, L. D. O. et al. Low-flow seasonality and effects on water availability throughout the river network. **Water Resources Management**, p. 1–16, 2020.
- SHORTRIDGE, J. E.; GUIKEMA, S. D.; ZAITCHIK, B. F. Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. **Hydrology and Earth System Sciences**, v. 20, p. 2611–2628, 2016.
- SOUSA, H. T. et al. **SisCAH 1.0 - Sistema Computacional para Análises Hidrológicas**. Brasília, DF: Agência Nacional das Águas, 2009.
- SOUZA, C. M. P. et al. Multivariate analysis and Machine Learning in properties of ultisols (argissolos) of Brazilian Amazon. **Revista Brasileira de Ciência do Solo**, v. 42, p. 1–20, 2018a.
- SOUZA, E. et al. Spatial and temporal potential groundwater recharge: the case of the Doce river basin, Brazil. **Revista Brasileira de Ciência do Solo**, v. 43, p. 1–27, 2018b.

- SUN, X. Y.; THOMPSON, C. J.; CROKE, B. **Using a logistic regression model to delineate channel network in southeast Australia**. Proc. of 19th International Congress on Modelling and Simulation. **Anais...**Perth (Australia): 2011.
- TONGAL, H.; BOOIJ, M. J. Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. **Journal of Hydrology**, v. 564, p. 266–282, 2018.
- ULIANA, E. M. et al. Análise de tendência em séries históricas de vazão e precipitação: uso de teste estatístico não paramétrico. **Revista Ambiente e Agua**, v. 10, n. 1, p. 82–88, 2015.
- WILLMOTT, C. J. On the validation of models. **Physical Geography**, v. 2, n. 2, p. 184–194, 1981.
- WORLAND, S. C.; FARMER, W. H.; KIANG, J. E. Environmental Modelling & Software Improving predictions of hydrological low- flow indices in ungaged basins using machine learning. **Environmental Modelling and Software**, v. 101, p. 169–182, 2018.
- WUEST, T. et al. Machine learning in manufacturing: Advantages, challenges, and applications. **Production and Manufacturing Research**, v. 4, n. 1, p. 23–45, 2016.
- YIN, Z. et al. Design and evaluation of SVR, MARS and M5Tree models for 1, 2 and 3-day lead time forecasting of river flow data in a semiarid mountainous catchment. **Stochastic Environmental Research and Risk Assessment**, v. 32, p. 2457–2476, 2018.
- ZAMBRANO-BIGIARINI, M. **hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series, version 0.4-0R Package**, 2014. Disponível em: <<https://cran.r-project.org/package=hydroGOF>>. Acesso em 12 mai. 2020.
- ZHANG, W.; GOH, A. T. C. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. **Geoscience Frontiers**, v. 7, p. 45–52, 2016.

## CONCLUSÕES GERAIS

- As abordagens avaliadas para fins de regionalização de vazões (proporcionalidade entre vazão e área de drenagem, regressões regionais e geoestatística) revelaram desempenho similar na bacia do rio Doce, sendo a única diferença expressiva a menor acurácia dos métodos geoestatísticos na predição da  $Q_{7,10}$  em uma das três regiões hidrologicamente homogêneas, caracterizada por distribuição espacial desuniforme das estações fluviométricas.
- As abordagens tradicionais de regionalização de vazões e os métodos de abordagem geoestatística apresentaram desempenho bastante similar e superior para a vazão média de longa duração em relação às vazões mínimas de referência.
- As abordagens com foco em interpolação linear obtiveram estimativas mais acuradas para as seções alocadas entre duas estações (caso 3) e piores estimativas nas situações em que a seção de interesse se situa a jusante da estação de apoio (caso 2) ou em trecho afluente do rio monitorado com confluência a jusante da estação de apoio (caso 5), devido às grandes diferenças de área de drenagem entre as seções de interesse e as respectivas seções de apoio.
- A modelagem geoestatística demonstrou bom desempenho preditivo, constituindo uma alternativa promissora aos outros métodos de regionalização de vazões em bacias menores e com precariedade de dados fluviométricos.
- A adoção dos métodos geoestatísticos e baseados em interpolação linear deve estar sempre associada a uma análise bastante criteriosa de fatores como a densidade e a distribuição espacial das estações de monitoramento fluviométrico na bacia hidrográfica.
- O modelo linear de aprendizado de máquina não apresentou desempenho satisfatório na predição de vazões mínimas de referência e média de longa duração, sendo contraindicada a sua utilização para fins de regionalização das vazões na bacia do rio Doce.
- A vazão equivalente ao volume precipitado com abstração de 750 mm ( $Peq_{750}$ ) foi a covariável ambiental que apresentou, para todas as vazões, a contribuição mais importante ao melhor desempenho dos modelos não lineares Random Forest e Earth.
- Os modelos não lineares apresentaram desempenhos similares e bons para a bacia do rio Doce, revelando-se alternativas promissoras para fins de regionalização de vazões em apoio à gestão e planejamento dos recursos hídricos em nível de bacia hidrográfica, porém, para a bacia do rio Doce, recomenda-se a adoção do Random Forest para a predição da  $Q_{7,10}$ , devido ao seu melhor desempenho, e para a  $Q_{90}$  em locais específicos, onde ocorreram valores negativos pelo Earth, apesar deste modelo ter sido ligeiramente superior nas demais situações.

**APÊNDICE A – Informações de identificação das estações fluviométricas utilizadas**

<b>Código</b>	<b>Nome da estação</b>	<b>Lat. (°)</b>	<b>Long. (°)</b>
56028000	PIRANGA	-20,69	-43,30
56055000	BRÁZ PIRES	-20,85	-43,24
56065000	SENADOR FIRMINO	-20,91	-43,10
56075000	PORTO FIRME	-20,67	-43,09
56085000	SERIQUEITE	-20,72	-42,92
56090000	FAZENDA VARGINHA	-20,71	-43,00
56110005	PONTE NOVA JUSANTE	-20,38	-42,90
56240000	FAZENDA PARAÍSO	-20,39	-43,18
56335001	ACAIACA JUSANTE	-20,36	-43,14
56337000	FAZENDA OCIDENTE	-20,27	-43,10
56385000	SÃO MIGUEL DO ANTA	-20,70	-42,67
56415000	RIO CASCA	-20,23	-42,65
56460000	MATIPÓ	-20,28	-42,33
56484998	RAUL SOARES MONTANTE	-20,10	-42,44
56500000	ABRE CAMPO	-20,30	-42,48
56510000	INSTITUTO FLORESTAL RAUL SOARES	-20,10	-42,46
56539000	CACHOEIRA DOS ÓCULOS MONTANTE	-19,78	-42,48
56570000	PINGO D'ÁGUA	-19,71	-42,45
56610000	RIO PIRACICABA	-19,93	-43,17
56640000	CARRAPATO (BRUMAL)	-19,97	-43,46
56696000	MARIO DE CARVALHO	-19,52	-42,64
56719998	BELO ORIENTE	-19,33	-42,38
56750000	CONCEIÇÃO DO MATO DENTRO	-19,01	-43,45
56765000	DOM JOAQUIM	-18,96	-43,24
56775000	FERROS	-19,23	-43,02
56787000	FAZENDA BARRACA	-19,33	-43,07
56800000	SENHORA DO PORTO	-18,89	-43,08
56825000	NAQUE VELHO	-19,19	-42,42
56845000	FAZENDA CORRENTE	-18,89	-42,71
56846000	PORTO SANTA RITA	-18,95	-42,36
56850000	GOVERNADOR VALADARES	-18,88	-41,95
56860000	SÃO PEDRO DO SUAÇUÍ	-18,36	-42,60
56870000	SANTA MARIA DO SUAÇUÍ	-18,20	-42,45
56891900	VILA MATIAS MONTANTE	-18,57	-41,92
56920000	TUMIRITINGA	-18,97	-41,64
56935000	DOM CAVATI	-19,37	-42,10
56940002	BARRA DO CUIETÉ JUSANTE	-19,06	-41,53
56960005	FAZENDA VARGEM ALEGRE	-20,17	-41,96
56976000	FAZENDA BRAGANÇA	-19,74	-41,79
56978000	SANTO ANTÔNIO DO MANHUAÇU	-19,68	-41,84
56983000	DORES DE MANHUMIRIM	-20,11	-41,73
56988500	IPANEMA	-19,80	-41,71
56989001	MUTUM	-19,81	-41,44
56989400	ASSARAI MONTANTE	-19,59	-41,46
56990000	SÃO SEBASTIÃO DA ENCRUZILHADA	-19,49	-41,16
56990990	AFONSO CLÁUDIO MONTANTE	-20,08	-41,12
56991500	LARANJA DA TERRA	-19,90	-41,06
56992000	BAIXO GUANDU	-19,52	-41,01
56993551	JUSANTE CÔRREGO DA PIABA	-19,56	-40,73
56994500	COLATINA	-19,53	-40,63
56995500	PONTE DO PANCAS	-19,42	-40,69
56998400	BARRA DE SÃO GABRIEL	-19,04	-40,53

**APÊNDICE B – Informações de identificação das estações pluviométricas utilizadas**

<b>Código</b>	<b>Nome da estação</b>	<b>Lat. (°)</b>	<b>Long. (°)</b>
1741006	LADAINHA (EFBM)	-17,63	-41,74
1741007	PEDRO VERSIANI (EFBM)	-17,88	-41,31
1742014	CAPELINHA	-17,69	-42,53
1742017	MALACACHETA	-17,85	-42,08
1742019	ÁGUA BOA	-17,99	-42,39
1840000	ÁGUIA BRANCA	-18,99	-40,75
1840004	BARRA DE SÃO FRANCISCO	-18,75	-40,89
1840007	ÁGUA DOCE	-18,55	-40,98
1840010	CEDROLÂNDIA	-18,81	-40,69
1840019	CÓRREGO DA BOA ESPERANÇA	-18,70	-40,44
1841001	VILA MATIAS - MONTANTE	-18,57	-41,92
1841003	CAMPANÁRIO	-18,24	-41,75
1841006	VARGEM GRANDE	-18,68	-41,20
1841007	FIDELÂNDIA - MONTANTE	-18,20	-41,25
1841009	MANTENÓPOLIS	-18,84	-41,11
1841011	TUMIRITINGA	-18,98	-41,64
1841018	CENTRAL DE MINAS	-18,76	-41,31
1841019	DIVINO DAS LARANJEIRAS	-18,78	-41,48
1841020	GOVERNADOR VALADARES	-18,88	-41,95
1842004	SÃO PEDRO DO SUAÇUÍ	-18,36	-42,60
1842005	COROACI	-18,61	-42,28
1842007	GUANHÃES	-18,77	-42,93
1842008	SANTA MARIA DO SUAÇUÍ	-18,20	-42,46
1842020	SÃO JOÃO EVANGELISTA	-18,55	-42,76
1843002	GOUVEIA	-18,47	-43,74
1843003	MENDANHA - MONTANTE	-18,11	-43,52
1843011	SERRO	-18,59	-43,41
1843012	RIO VERMELHO	-18,28	-43,00
1939002	POVOAÇÃO	-19,58	-39,79
1940000	ITARANA	-19,87	-40,87
1940001	SÃO JOÃO DE PETRÓPOLIS	-19,81	-40,68
1940002	SANTA CRUZ - LITORAL	-19,96	-40,15
1940003	RIACHO	-19,75	-40,04
1940005	CAVALINHO	-19,69	-40,40
1940006	COLATINA - CORPO DE BOMBEIROS	-19,53	-40,62
1940007	FUNDÃO	-19,94	-40,40
1940009	PANCAS	-19,22	-40,85
1940010	VALSUGANA VELHA - MONTANTE	-19,95	-40,55
1940012	ITAIMBÉ	-19,66	-40,84
1940013	NOVO BRASIL	-19,24	-40,59
1940016	BARRA DE SÃO GABRIEL	-19,06	-40,52
1940020	CALDEIRÃO	-19,96	-40,74
1940021	ARACRUZ	-19,83	-40,27
1940022	JACUPEMBA	-19,59	-40,20
1940023	RIO BANANAL	-19,27	-40,32
1941000	IPANEMA	-19,80	-41,71
1941003	BAIXO GUANDU	-19,52	-41,01
1941004	RESPLENDOR - JUSANTE	-19,34	-41,25
1941005	BARRA DO CUIETÉ - JUSANTE	-19,06	-41,53
1941006	ASSARAI - MONTANTE	-19,59	-41,46
1941008	LARANJA DA TERRA	-19,90	-41,06

continua...

## APÊNDICE B – continuação

<b>Código</b>	<b>Nome da estação</b>	<b>Lat. (°)</b>	<b>Long. (°)</b>
1941009	IBITUBA	-19,69	-41,02
1941010	SÃO SEBASTIÃO DA ENCRUZILHADA	-19,49	-41,16
1941011	SANTO ANTÔNIO DO MANHUAÇU	-19,68	-41,84
1941012	ALTO RIO NOVO	-19,06	-41,03
1941018	ITANHOMI	-19,16	-41,86
1941019	MUTUM	-19,81	-41,44
1942002	BOM JESUS DO GALHO	-19,83	-42,32
1942006	VERMELHO VELHO	-20,00	-42,35
1942008	DOM CAVATI	-19,37	-42,11
1942029	MARIO DE CARVALHO	-19,52	-42,64
1942030	BELO ORIENTE	-19,32	-42,40
1942031	CACHOEIRA DOS ÓCULOS - MONTANTE	-19,78	-42,48
1942032	NAQUE VELHO	-19,19	-42,42
1943001	RIO PIRACICABA	-19,92	-43,18
1943002	CONCEIÇÃO DO MATO DENTRO	-19,02	-43,44
1943003	FERROS	-19,25	-43,01
1943004	JABOTICATUBAS	-19,52	-43,74
1943006	SABARÁ	-19,89	-43,82
1943007	SANTA BÁRBARA	-19,95	-43,40
1943008	SANTA MARIA DO ITABIRA	-19,44	-43,12
1943010	CAETÉ	-19,90	-43,67
1943023	TAQUARAÇU	-19,66	-43,69
1943025	MORRO DO PILAR	-19,22	-43,37
1943027	USINA PETI	-19,88	-43,37
1943035	VAU DA LAGOA	-19,22	-43,59
1943042	FAZENDA CARAIBAS	-19,12	-43,84
1944020	PIRAPAMA	-19,01	-44,04
2040007	SANTA MARIA DO JETIBÁ	-20,03	-40,74
2040008	GARRAFÃO (DNOS)	-20,14	-40,98
2040010	SANTA LEOPOLDINA	-20,10	-40,53
2040018	CACHOEIRA SUIÇA (DNOS)	-20,08	-40,61
2040023	SÃO RAFAEL	-20,30	-40,93
2041002	CASTELO	-20,61	-41,20
2041005	CAIANA	-20,69	-41,92
2041008	DORES DO MANHUMIRIM	-20,11	-41,73
2041011	CONCEIÇÃO DO CASTELO (DNOS)	-20,35	-41,24
2041013	IÚNA	-20,35	-41,54
2041014	DORES DO RIO PRETO	-20,69	-41,85
2041016	IBITIRAMA	-20,54	-41,67
2041017	SANTA CRUZ - CAPARAÓ	-20,32	-41,70
2041018	USINA FORTALEZA	-20,37	-41,41
2041019	ITAICI	-20,53	-41,51
2041020	ARACE (FAZENDA MODELO)	-20,37	-41,06
2041023	AFONSO CLÁUDIO - MONTANTE	-20,08	-41,12
2041048	FAZENDA VARGEM ALEGRE	-20,17	-41,96
2042008	RAUL SOARES - MONTANTE	-20,10	-42,44
2042010	ABRE CAMPO	-20,30	-42,48
2042011	RIO CASCA	-20,21	-42,65
2042014	BICUIBA	-20,77	-42,30
2042016	SÃO MIGUEL DO ANTA	-20,68	-42,81
2042017	MATIPÓ	-20,28	-42,33

continua...

**APÊNDICE B – continuação**

<b>Código</b>	<b>Nome da estação</b>	<b>Lat. (°)</b>	<b>Long. (°)</b>
2042018	PONTE NOVA JUSANTE	-20,38	-42,90
2042031	FAZENDA CACHOEIRA D'ANTAS	-20,01	-42,67
2043002	LAGOA GRANDE (MMV)	-20,18	-43,94
2043005	CONSELHEIRO LAFAIETE	-20,63	-43,75
2043009	ACAIACA - JUSANTE	-20,36	-43,14
2043010	PIRANGA	-20,69	-43,30
2043011	FAZENDA PARAÍSO	-20,39	-43,18
2043014	PORTO FIRME	-20,67	-43,09
2043025	USINA DA BRECHA	-20,52	-43,02
2043026	BRAZ PIRES	-20,85	-43,24
2043027	FAZENDA OCIDENTE	-20,29	-43,10
2043042	REPRESA DAS CODORNAS (MMV)	-20,16	-43,89
2043043	REPRESA DO MIGUELÃO (MMV)	-20,13	-43,95
2043056	FAZENDA ÁGUA LIMPA JUSANTE	-20,30	-43,62
2043059	COLÉGIO CARAÇA	-20,10	-43,49
2043060	ITABIRITO LINIGRAFO	-20,29	-43,80
2044007	ENTRE RIOS DE MINAS	-20,66	-44,07
2142000	ASTOLFO DUTRA	-21,31	-42,86
2142004	FAZENDA UMBAÚBAS	-21,05	-42,52
2142009	JUSSARA	-20,91	-42,35
2143000	USINA ITUERÊ	-21,31	-43,20
2143001	GUARANI	-21,36	-43,05
2143003	DESTERRO DO MELO	-21,15	-43,52
2143005	CAMPOLIDE	-21,28	-43,82
2143006	BARROSO	-21,19	-43,98
2143017	TABULEIRO	-21,36	-43,25

**APÊNDICE C – Valores das variáveis independentes de cada estação fluviométrica**

<b>Código</b>	<b>A (km<sup>2</sup>)</b>	<b>Pa (mm)</b>	<b>Pss (mm)</b>	<b>Psc (mm)</b>	<b>Peq750 (m<sup>3</sup> s<sup>-1</sup>)</b>
56028000	1397,06	1433,17	200,95	1237,35	30,26
56055000	1091,11	1496,19	205,48	1284,63	25,82
56065000	299,73	1323,59	196,94	1125,28	5,45
56075000	4259,82	1415,99	198,90	1216,50	89,96
56085000	341,50	1228,69	196,71	1035,06	5,18
56090000	328,61	1264,67	195,88	1074,84	5,36
56110005	6224,67	1367,56	197,79	1171,01	121,90
56240000	855,20	1371,60	199,73	1188,62	16,86
56335001	1374,10	1386,61	200,87	1202,02	27,74
56337000	529,65	1478,99	207,95	1288,24	12,24
56385000	525,92	1309,82	214,39	1097,64	9,34
56415000	2033,12	1276,26	201,75	1078,50	33,93
56460000	616,19	1275,26	197,18	1089,63	10,26
56484998	1355,50	1245,94	186,38	1068,63	21,32
56500000	271,85	1267,46	193,04	1100,07	4,46
56510000	1870,60	1249,04	186,26	1075,19	29,60
56539000	15858,15	1324,40	194,69	1133,39	288,84
56570000	818,57	1197,71	172,57	1026,84	11,62
56610000	1161,64	1537,91	204,80	1337,92	29,02
56640000	611,77	1633,50	200,41	1441,93	17,14
56696000	5295,39	1407,93	190,79	1215,23	110,48

continua...

**APÊNDICE C – continuação**

<b>Código</b>	<b>A (km<sup>2</sup>)</b>	<b>Pa (mm)</b>	<b>Pss (mm)</b>	<b>Psc (mm)</b>	<b>Peq750 (m<sup>3</sup> s<sup>-1</sup>)</b>
56719998	24209,52	1327,55	191,17	1137,92	443,37
56750000	301,76	1341,88	176,53	1170,90	5,66
56765000	974,49	1325,74	182,47	1146,43	17,79
56775000	4090,03	1373,51	182,34	1195,67	80,86
56787000	1252,17	1319,23	176,88	1147,52	22,60
56800000	1520,22	1275,48	178,49	1102,99	25,33
56825000	10187,03	1311,27	180,55	1134,28	181,31
56845000	1062,16	1212,05	172,05	1043,74	15,56
56846000	1975,82	1217,52	177,87	1044,32	29,29
56850000	40519,01	1301,45	186,28	1117,77	708,52
56860000	2584,76	1156,70	163,13	1007,42	33,33
56870000	621,18	1064,79	154,42	923,36	6,20
56891900	9748,65	1098,60	168,29	938,99	107,76
56920000	55042,15	1236,93	180,28	1060,78	849,87
56935000	775,84	1179,82	172,72	1015,18	10,57
56940002	3224,69	1087,47	162,33	933,75	34,51
56960005	1090,84	1320,32	211,20	1114,06	19,73
56976000	2279,32	1254,60	194,29	1070,22	36,47
56978000	2368,52	1251,14	193,17	1067,84	37,64
56983000	386,05	1373,36	221,17	1157,10	7,63
56988500	1407,29	1253,26	196,92	1063,94	22,46
56989001	1178,61	1142,69	188,22	961,67	14,68
56989400	3193,85	1185,96	189,05	1003,84	44,15
56990000	8711,42	1152,33	180,84	980,09	111,14
56990990	434,39	1209,40	228,16	976,99	6,33
56991500	1331,40	1149,38	205,23	943,98	16,86
56992000	2141,12	1066,74	184,72	883,02	21,50
56993551	879,48	1033,80	178,37	852,56	7,91
56994500	76382,10	1190,25	178,31	1016,90	1066,32
56995500	920,33	1079,05	215,31	860,78	9,60
56998400	1026,69	1032,25	205,07	785,50	9,19

**APÊNDICE D – Estatísticas descritivas das covariáveis ambientais utilizadas no estudo**

<b>Covariáveis ambientais</b>	<b>Média</b>	<b>Máximo</b>	<b>Mínimo</b>	<b>SD*</b>	<b>CV**</b>
Área (Km <sup>2</sup> )	5943,40	76401,69	271,54	13849,31	2,33
Perímetro (Km)	442,67	2364,66	113,22	451,06	1,02
Hmédia (m)	732,66	1135,82	300,01	143,69	0,20
Hmax (m)	1700,42	2618,00	885,00	443,87	0,26
Hmin (m)	370,62	732,00	23,00	197,96	0,53
ΔH (m)	1329,81	2595,00	272,00	526,38	0,40
Dmédia (%)	22,99	30,72	16,64	3,48	0,15
Dmax (%)	164,92	256,97	56,12	50,63	0,31
DHmédia (mm)	193,44	289,06	148,91	29,94	0,15
DHmax (mm)	364,08	665,06	266,78	104,44	0,29
DHmin (mm)	85,83	159,18	44,74	22,29	0,26
ETmédia (mm)	904,94	921,19	863,32	13,60	0,02
ETmin (mm)	788,31	837,15	626,65	45,85	0,06
ETmax (mm)	1004,85	1070,72	951,40	23,98	0,02
ETPmédia (mm)	1098,48	1175,79	1055,79	30,45	0,03

continua...

**APÊNDICE D – continuação**

<b>Covariáveis ambientais</b>	<b>Média</b>	<b>Máximo</b>	<b>Mínimo</b>	<b>SD*</b>	<b>CV**</b>
ETPmax (mm)	1203,64	1318,66	1128,76	46,88	0,04
ETPmin (mm)	85,83	159,18	44,74	22,29	0,26
Tmax (°C)	26,09	28,16	24,23	0,87	0,03
Tmin (°C)	14,15	17,24	12,26	1,04	0,07
Pa – IDW (mm)	1269,02	1601,21	1030,22	118,48	0,09
Pss – IDW (mm)	189,40	231,45	152,31	17,31	0,09
Psc – IDW (mm)	1081,56	1410,91	797,96	117,56	0,11
Pa – KS (mm)	1269,39	1632,70	1032,01	127,24	0,10
Pss – KS (mm)	190,72	228,00	154,41	15,08	0,08
Psc – KS (mm)	1083,38	1442,01	786,16	124,24	0,11
Pmax – IDW (mm)	1446,80	1871,73	1112,61	211,91	0,15
Pmax – KS (mm)	1445,61	1864,51	1112,28	213,53	0,15
Pmin – IDW (mm)	1145,35	1448,87	807,77	154,36	0,13
Pmin – KS (mm)	1141,85	1465,41	809,77	152,45	0,13
Peq – IDW (m³/s)	235,64	2886,69	10,86	538,33	2,28
Peq – KS (m³/s)	235,79	2883,32	10,91	538,58	2,28
Peq750 – IDW (m³/s)	94,29	1069,67	4,40	209,78	2,22
Peq750 – KS (m³/s)	94,44	1066,31	4,45	210,08	2,22
Floresta natural 1992 (%)	25,97	56,14	8,48	12,27	0,47
Floresta natural 2002 (%)	25,89	57,40	8,95	10,89	0,42
Floresta natural 2012 (%)	28,06	57,51	9,61	9,79	0,35
Floresta plantada 1992 (%)	2,04	15,96	0,00	3,43	1,68
Floresta plantada 2002 (%)	2,16	15,50	0,02	3,53	1,64
Floresta plantada 2012 (%)	3,25	18,31	0,34	3,69	1,14
Formação savânica 1992 (%)	0,36	3,28	0,00	0,65	1,78
Formação savânica 2002 (%)	0,33	2,16	0,00	0,52	1,58
Formação savânica 2012 (%)	0,34	2,36	0,00	0,54	1,59
Formação campestre 1992 (%)	0,30	4,77	0,00	0,85	2,83
Formação campestre 2002 (%)	0,36	5,57	0,00	1,01	2,79
Formação campestre 2012 (%)	0,32	5,17	0,00	0,93	2,87
Pastagens 1992 (%)	54,41	77,68	4,23	15,51	0,28
Pastagens 2002 (%)	55,55	78,33	4,50	15,98	0,29
Pastagens 2012 (%)	47,91	69,34	3,98	13,00	0,27
Agropecuária 1992 (%)	15,13	24,06	6,37	4,46	0,29
Agropecuária 2002 (%)	13,12	23,50	0,25	5,39	0,41
Agropecuária 2012 (%)	17,09	34,45	0,49	6,19	0,36
Culturas agrícolas 1992 (%)	0,00	0,01	0,00	0,00	2,86
Culturas agrícolas 2002 (%)	0,31	16,04	0,00	2,20	7,11
Culturas agrícolas 2012 (%)	0,58	30,08	0,00	4,13	7,07
Mineração 1992 (%)	0,00	0,08	0,00	0,01	2,94
Mineração 2002 (%)	0,01	0,12	0,00	0,02	2,41
Mineração 2012 (%)	0,02	0,14	0,00	0,03	1,89
Urbano 1992 (%)	0,23	1,54	0,00	0,28	1,20
Urbano 2002 (%)	0,33	2,54	0,00	0,41	1,25
Urbano 2012 (%)	0,44	2,99	0,04	0,51	1,15
Afloramento rochoso 1992 (%)	1,22	7,88	0,00	1,91	1,57
Afloramento rochoso 2002 (%)	1,51	9,86	0,00	2,27	1,50
Afloramento rochoso 2012 (%)	1,61	11,97	0,00	2,47	1,54
Rios e Lagos 1992 (%)	0,09	0,44	0,00	0,13	1,45
Rios e Lagos 2002 (%)	0,11	0,45	0,00	0,13	1,22
Rios e Lagos 2012 (%)	0,11	0,49	0,00	0,14	1,26

continua...

**APÊNDICE D – continuação**

<b>Covariáveis ambientais</b>	<b>Média</b>	<b>Máximo</b>	<b>Mínimo</b>	<b>SD*</b>	<b>CV**</b>
NC – seco (adm.)	65,28	67,82	61,23	1,51	0,02
NC – médio (adm.)	81,47	83,23	78,58	1,07	0,01
NC – úmido (adm.)	91,93	92,84	90,18	0,57	0,01
USmédia (mm)	1794,73	2961,81	968,05	449,23	0,25
USmax (mm)	2227,39	3859,90	1346,25	494,61	0,22
USmin (mm)	1176,23	2054,62	325,29	442,71	0,38
LX (%)	71,73	100,00	0,06	25,01	0,35
AX (%)	15,47	66,58	0,00	18,64	1,20
CX (%)	9,88	88,70	0,00	18,39	1,86
NX (%)	1,77	18,07	0,00	4,07	2,31

\*SD: desvio padrão (unidade de medida dos dados).

\*\*CV: coeficiente de variação (decimal).