

Universidade Federal de Minas Gerais
Instituto de Geociências
Programa de Pós-graduação
Doutorado em Geografia

**Análise da Relação entre Uso do Solo e Liberação dos
Nutrientes Fósforo e Nitrogênio nas Águas Superficiais do
Rio das Velhas Utilizando Árvores de Decisão e Regressão
Múltipla**

Lília Maria de Oliveira

Belo Horizonte
2016

Lília Maria de Oliveira

**Análise da Relação entre Uso do Solo e Liberação dos
Nutrientes Fósforo e Nitrogênio nas Águas Superficiais do
Rio das Velhas Utilizando Árvores de Decisão e Regressão
Múltipla**

Tese apresentada ao Programa de Pós-
graduação do Departamento de Geografia da
Universidade Federal de Minas Gerais, como
requisito parcial à obtenção do título de Doutor
em Geografia

Área de Concentração: Análise Ambiental

Orientador: Dr. Philippe Maillard

Coorientador: Dr. Eber José de Andrade Pinto

Belo Horizonte

2016

Defesa, em 28 de Junho de 2016, pela Banca Examinadora constituída pelos doutores e professores:

Prof. Dr. Adolf Heinrich Horn

Prof. Dr. Carlos Wagner Gonçalves Andrade Coelho

Prof. Dr. Éber José de Andrade Pinto

Prof. Dr. Philippe Maillard

Prof^a Dra. Sílvia Maria Alves Corrêa Oliveira

Prof^a. Dra. Wanderlene Ferreira Nacif

Sumário

RESUMO	i
ABSTRACT	iii
RÉSUMÉ	v
SIGLAS E SÍMBOLOS	ix
1 PREFÁCIO	1
2 INTRODUÇÃO	2
2.1 Aspectos Gerais	2
2.2 Hipóteses e Objetivos	4
2.3 Organização dos artigos	6
2.4 Padronização das nomeclaturas das estações de qualidade e vazão para os Capítulos 5, 6 e 7	11
3 ÁREA DE ESTUDO	13
3.1 Aspectos Físicos	13
3.2 Aspectos Socioeconômicos	15
3.3 Qualidade das águas da bacia do Rio das Velhas	17
4 REVISÃO DA LITERATURA	21
4.1 Poluição Difusa	21
4.2 Curva de Permanência	25
4.2.1 Curva de Permanência e Qualidade das Águas	26
4.3 Árvores de Decisão	29
4.4 Árvores de Decisão em Estudos Ambientais	34
5 ARTIGO 1 : Aplicação de Árvores de Decisão na Modelagem das Concentrações de Nitrato e Fósforo Total: Estudo de Caso no Rio das Velhas	37
5.1 Resumo	37
5.2 Palavras-Chave	38
5.3 Abstract	38
5.4 Key-words:	38
5.5 Introdução	38

5.6	Princípios de árvore de decisão	40
5.7	O algoritmo C4.5 (J48)	43
5.8	Materiais e Métodos	44
5.8.1	Área de estudo	44
5.8.2	Dados de qualidade das águas e vazão da Bacia do Rio das Velhas	46
5.8.3	Período de análise e preparação da base de dados	49
5.8.4	Árvore de decisão	52
5.8.5	Validação cruzada	53
5.9	Resultados	53
5.9.1	Análise das classificações	53
5.9.2	Avaliação das regras das árvores de decisão	57
5.10	Discussão	59
5.11	Agradecimentos	60
5.12	Referências Bibliográficas	60
6	ARTIGO 2 : Modelling the effect of land use/land cover on nitrogen, phosphorous and dissolved oxygen loads in the Velhas River using the concept of exclusive contribution area	61
6.1	Abstract	61
6.2	keywords	62
6.3	Introduction	62
6.4	Methods	66
6.4.1	Study Area	66
6.4.2	Water quality and water discharge data	68
6.4.3	Statistical Analysis	71
6.5	Results	73
6.5.1	Land use / land cover per exclusive contribution areas and riparian zones	73
6.5.2	Flow stream analysis of the distribution of the WQ parameters	73
6.5.3	Correlation between LULC and WQ parameters	78
6.5.4	Multiple Regressions Models	80
6.6	Conclusions	83
6.7	acknowledgements	84
6.8	Bibliographic References	85
7	ARTIGO 3 : Application of a land cover pollution index to model non-point pollution sources in a Brazilian watershed	86
7.1	author	86
7.2	Abstract	86
7.3	KeyWords	87

7.4	Introduction	87
7.5	Material and Methods	89
7.5.1	Study Area	89
7.5.2	Water quality and water discharge data	91
7.5.3	Land cover pollution index	95
7.5.4	Statistical Analysis	95
7.6	Results and Discussion	96
7.7	Conclusion	101
7.8	Acknowledgements	103
7.9	Bibliographic References	103
8	ARTIGO 4 : Estimativa das cargas de poluição difusa com a separação das fontes, pontual e difusa, baseando-se no regime fluvial.	104
8.1	Autores	104
8.2	Resumo	104
8.3	Palavras-Chave	105
8.4	Introdução	105
8.5	Materiais e Métodos	110
8.5.1	Área de Estudo	110
8.5.2	Metodologia de Monitoramento	112
8.5.3	Análise Estatística	116
8.6	Resultados	117
8.6.1	Nitrato	118
8.6.2	Fósforo	119
8.7	Conclusão	121
8.8	Bibliographic References	123
9	CONSIDERAÇÕES FINAIS	124
9.1	Sugestões de Trabalhos Futuros	133
	REFERÊNCIAS BIBLIOGRÁFICAS	134
	Apêndice A UTEs Bacia Rio das Velhas	145
	Apêndice B Exemplo de Árvore de Decisão	146
	Apêndice C Base de Dados dos Artigos e Estatística Decritiva	152

Lista de Figuras

2.1	Montagem e consistência da base de dados: uso do solo por área de estudo (WS, ECA e B), qualidade e vazão.	7
2.2	Esquema das análises realizadas nos 4 artigos	7
2.3	Fluxograma das etapas desenvolvidas na obtenção dos dados do Artigo 1	8
2.4	Fluxograma das etapas desenvolvidas na obtenção dos dados dos Artigos 2 e 3	9
2.5	Fluxograma das etapas desenvolvidas na obtenção dos dados do Artigo 4	10
3.1	Mapa de localização da área de estudo	14
3.2	Percentual de vazões médias retiradas diariamente por tipos de usos da Bacia Rio das Velhas CBH (2015)	16
3.3	Percentual de vazões médias retiradas diariamente por tipos de usos nos trechos Alto, Médio Alto, Médio baixo e Baixo do Rio das Velhas. Adaptado de CBH (2015)	17
4.1	Exemplo de Curva de permanência (Rio Paraopeba em Ponte Nova do Paraopeba). Fonte:Naghattini e Pinto (2007)	25
4.2	Curva de permanência e cálculo da DBO média. Fonte:Adaptado de Brites (2010)	27
4.3	Concentração média de DBO para os intervalos da curva de permanência. Fonte: Adaptado de Brites (2010)	28
4.4	Curva de permanência observada e compatível com o enquadramento do Rio Paraíba do Sul. Fonte:Cunha <i>et al.</i> (2012)	28
4.5	Esquema Árvore de Decisão. Fonte:Solomatine e Dulal (2003)	30
4.6	Processo de divisão da Árvore de Decisão. Fonte:Solomatine e Dulal (2003)	31
5.1	Esquema Árvore de Decisão. Fonte:Solomatine e Dulal (2003)	40
5.2	Processo de Divisão de Árvore de Decisão. Fonte:(Solomatine e Dulal, 2003)	42
5.3	Mapa de localização da área de estudo	45
5.4	Bacia do Rio das Velhas com estações fluviométricas e de qualidade de águas	47
5.5	Tipos de Uso do Solo da Bacia do Rio das Velhas	48

5.6	Normais Climatológicas 1961-1990 para Belo Horizonte e Pirapora. Fonte: INMET (2014).	50
5.7	(a) Permanência de vazão das estações 41199998,41260000 e 41410000, (b) Permanência Nitrato - BV013 e (c) Permanência Fósforo Total - BV013	51
5.8	Etapas de geração das bases de qualidade, vazão e uso do solo	52
5.9	Percentual de instâncias classificada correta e incorretamente para Nitrato (nov-Jan/2009-2011)	55
5.10	Percentual de instâncias classificada correta e incorretamente para Fósforo Total (nov-Jan/2009-2011)	55
5.11	(a) Árvore de decisão para o nitrato Nov-Jan/2009-2011 e (b) para o fósforo total Nov-Jan/2009-2011. Legenda: I_{Q-N} - Índice que relaciona a permanência da vazão (PermQ) a permanência do nitrato, AP - Percentual de agropastoril, AU Percentual de área urbana, CA - Percentual de Campo, MI - Percentual de Mineral e I_{Q-P} - Índice que relaciona a permanência da vazão (PermQ) a permanência do fósforo total.	56
5.12	(a) Agrupamento classes do nitrato e (b) Agrupamento classes do fósforo Total	58
6.1	Map of the Velhas River Watershed showing the LULC classes, the exclusive contribution areas (ECA) and the monitoring stations (discharge and quality).	67
6.2	Illustration of how the starting date of the rainy season was selected from the daily hydrogram of each station.	68
6.3	Illustration of how the exclusive contribution areas (ECA) and the river edge strips were created from the digital surface model: (a) watershed with hydrography and quality gauging stations, (b) determination of the ECA, and (c) strips of varying width (buffers) on either side of the stream.	70
6.4	Flow diagram of the methodological steps.	72
6.5	Land use / land cover (%) per exclusive contribution and riparian zones: (a) 150 meters, (b) 300 meters, (c) 600 meters, (d) 900 meters and (e) ECA . .	74
6.6	Incremental loads for the four WQ parameters (kg / day km ²)	75
6.7	Incremental loads per unit area of (a) TAN, (b) nitrate, (c) dissolved oxygen and (d) total phosphorous in kg / km ² day	77
6.8	Plots of the two best regression models for the exclusive contribution areas (ECAs)	81
6.9	Coefficient of Determination	83
7.1	Map of the Velhas River Watershed showing the LULC categories, the exclusive contribution areas (ECA) and the monitoring stations (discharge and quality).	90

7.2	Discharge (Q100%, Q95%, Q50% and Q5% m^3/s) and drainage area of the Velhas River Watershed (km^2).	91
7.3	Illustration of how the starting date of the rainy season was selected from the daily hydrogram of each station.	92
7.4	Flowchart of the data processing steps. The left hand side shows how the water quality (WQ) and water discharge (WD) data are processed to be transformed in unit incremental loads (UIL). Because the discharge stations do not correspond to the WQ stations a regionalization of the WD was performed using the stations coordinates (x,y). The right hand side shows the processing of the spatial components (LULC, DSM, hydrological network) to generate the four types of analyses areas (three buffers and one exclusive contribution area).	94
7.5	Illustration of how the exclusive contribution areas (ECA) and the river edge strips (buffers) were created from the digital surface model: (a) watershed with hydrography and WQ gauging stations, (b) determination of the ECA, and (c) strips of varying width (buffers) on either side of the stream.	94
7.6	Unit incremental load for the five water quality parameters ($kg / day km^2$) and the 17 WQ stations: (a) - Total Ammonia Nitrogen, (b) Dissolved Oxygen, (c) Nitrate, (d) Nitrite and (e) Phosphorus.	98
7.7	Coefficient of Determination for the LCPI e LULC models: (a) TAN, (b) OD and (c) NO_2^-	101
8.1	Mapa da Bacia do Rio das Velhas contendo: classes de LULC avaliadas, área de contribuição exclusiva (ECA) e estações de monitoramento de qualidade.	111
8.2	Áreas de contribuição exclusive (ECA) e as faixas foram criadas a partir do modelo digital do terreno: (a) bacia com hidrografia e estações de qualidade de água, (b) determinação da ECA, e (c) faixas de largura variável em ambos os lados do rio.	114
8.3	Períodos de seca e chuva avaliados entre 2008 e 2013 ($S_{WQ} = 14$).	114
8.4	Detalhamento do período de 2008 a 2009 para seleção dos dados de seca e chuva ($S_{WQ} = 14$).	115
8.5	Uso e Ocupação do solo (%) por área de contribuição exclusiva e faixas: (a) 150 metros, (b) 300 metros, (c) 600 metros, (d) 900 metros and (e) ECA	120
9.1	Hipóteses, Resultados, Dificuldades e Possibilidades obtidas com o Capítulo 4.	125
9.2	Hipóteses, Resultados, Dificuldades e Possibilidades obtidas com os Capítulos 5 e 6.	129

9.3	Hipóteses, Resultados, Dificuldades e Possibilidades obtidas com o Capítulo 7	132
A.1	UTES e Regiões da bacia do Rio das Velhas. Adaptado de CBH (2015) . .	145
B.1	Ganho de Informação e Razão de Ganho para os 8 atributos de LULC avaliados	151
B.2	Árvore de Decisão (c4.5) para Razão de Ganho Maior que 0.20	151

Lista de Tabelas

2.1	Padronização dos códigos das estações de Qualidade (QA) para os Capítulos 5, 6 e 7	12
2.2	Padronização dos códigos das estações de vazão para os Capítulos 5, 6 e 7	12
3.1	Características gerais da Bacia do Rio das Velhas.	13
4.1	Síntese de estudos sobre a relação LULC x NPPS	24
5.1	Parâmetros do algoritmo J48. Adaptado de (Banon, 2013)	44
5.2	Estações de monitoramento de qualidade das águas e de vazão utilizadas	46
5.3	Uso do solo da bacia do Rio das Velhas em 2010 (as linhas em negrito mostram as categorias agrupadas ou não consideradas	49
5.4	Instâncias Classificadas corretamente e incorretamente em função da discretização utilizada nos dados de qualidade. Nov-Jan 2009-2011.	54
5.5	Instâncias Classificadas corretamente e incorretamente em função da discretização utilizada nos dados de qualidade. Dez-Fev 2009-2011.	54
5.6	Média e desvio padrão das categorias de uso do solo Agropastoril e Urbano para as faixas de I_{Q-N}	56
5.7	Intervalos de discretização das variáveis dependentes permanência do nitrato e do fósforo total (Nov-Jan).	57
6.1	Synopsis of studies that have analyzed the parameters ammonia (NH_4^+), nitrate (NO_3^-), nitrite (NO_2^-), total phosphorous (TP), total nitrogen (TN), nitrogen (N), and dissolved oxygen (DO) in relation to land use and land cover.	64
6.2	Rainy season starting dates for the 2001–2013 period for the nine daily water discharge monitoring stations	68
6.3	Watersherd contribution areas for the S_{WQ} and S_Q stations and the corresponding areas compensation factors ($CA_{WQ/Q}$) applied to the Q stations to correspond to the WQ stations.	70
6.4	Area and unit incremental loads (median) for all exclusive contribution areas (ECA) $kg/km^2 \cdot day$	76
6.5	Pearson’s correlation between the WQ parameters and the LULC classes	79

6.6	Multiple regression models	80
7.1	Area compensation factor $C'_{f_{WQ/Q}}$ and unit incremental loads (median, kg/km ² . day) for all exclusive contribution areas (ECA).	93
7.2	Results of multiple regression between Land Use and Land Cover (LULC) and Land Cover Pollution Index (LCPI) for the different analysis areas (ECAs and four B). Columns G1 through Ft and LCPI show the sign and strength of Pearson's correlation whereas the two Model columns explicitly show the regression formulas.	99
7.3	Land cover pollution index (LCPI) for the 17 stations and the five areas of analysis.	100
8.1	Vazões características das estações fluviométricas utilizadas - SQ.	112
8.2	Coeficientes de correlação (Kendall, Spearman e Pearson) entre UIDL de Nitrato e LULC (ECAs e faixas - B).	117
8.3	Carga Difusa Incremental Unitária de Nitrato	118
8.4	Resultados da regressão múltipla entre LULC para diferentes áreas de análise (ECAs e faixas - B) para o nitrato.	119
8.5	Carga Difusa de Fósforo Total (DLTP) para as estações de qualidade nos 5 períodos avaliados	121
B.1	Carga Incremental (UIL) de Fósforo total (TP) entre 2001-2013 e % das categorias de LULC em cada ECA para a faixa de 100 metros	147
B.2	Discretização da Carga Mediana Incremental (UIL) de Fósforo total (TP) e LULC faixa de 100 metros	148
B.3	Divisão de TP em função FL150	149
B.4	Entropia, Entropia Condicional, Ganho, Razão de Ganho para buffer 150 metros	150

RESUMO

A presença dos nutrientes fósforo e nitrogênio nas águas superficiais se relaciona com o uso e ocupação do solo (LULC) da bacia hidrográfica, devido à geração e deposição de substâncias no solo, que podem atingir o curso de água pela ação da chuva e do escoamento direto. Fontes poluidoras dispersas na área da bacia podem alterar a qualidade das águas superficiais e são chamadas de fontes difusas ou não-pontuais. A identificação e quantificação da carga poluidora de fontes difusas é de difícil obtenção principalmente quando ocorrem em conjunto com fontes pontuais, como o lançamento de efluentes domésticos ou industriais. A Bacia do Rio das Velhas é um exemplo desse tipo de situação, devido à presença de fontes difusas e pontuais, o que torna complexa a separação das cargas poluidoras dos dois tipos de fontes e, para fontes difusas, a determinação da sua relação com LULC. A tese tem por objetivo primário a análise da relação entre LULC e a liberação dos nutrientes fósforo e nitrogênio para as águas superficiais do Rio das Velhas, considerando potenciais e limitações das análises. Os métodos avaliados utilizam principalmente regressão múltipla e árvore de decisão para criar modelos. Para o desenvolvimento do trabalho foi proposto um índice que relaciona categorias de LULC, como fonte ou filtro de poluição para as águas superficiais, e um método para estimativa das cargas difusas e pontuais com base no fluviograma de vazões. As categorias de LULC foram obtidas considerando três diferentes áreas de análise: bacia hidrográfica, área de contribuição exclusiva e faixas nas margens do curso de água. A seleção dos dados de qualidade para períodos de chuvas e seca também utilizou diferentes abordagens: trimestre chuvoso, período chuvoso e ascensão do hidrograma. O uso de árvore de decisão se mostrou promissor no estabelecimento da relação concentração de nutrientes e LULC, atingindo eficiência de classificação superior a 80% para as duas variáveis avaliadas, nitrato e fósforo total. O uso de regressão múltipla apresentou bons resultados na relação da amônia, nitrito e oxigênio dissolvido com LULC, com coeficiente de determinação de aproximadamente 80%. Os parâmetros fósforo total e nitrato não obtiveram bons resultados na regressão. Por meio da análise do coeficiente de Pearson foi possível identificar a presença de categoria de LULC tidas como fontes de poluição (atividades agropastoris e urbanas) e como filtros de poluição (vegetação). Foi então proposto um índice correlacionando fonte e filtro de poluição, que substituiu o uso de categorias individuais de LULC. O uso do índice (LCPI) promoveu uma melhoria no coeficiente de determinação de 12 dos 15 modelos

válidos. Por fim, foi proposto um método para estimar as cargas difusas e pontuais no Rio das Velhas, baseado na análise do fluviograma de cheia. Os resultados iniciais indicaram a influência do regime de chuvas no acúmulo de fósforo e nitrato na bacia, em anos com menor precipitação/vazão, sendo possível obter modelo para o nitrato, mas impossibilitou a obtenção para o fósforo. Os resultados apontaram a necessidade do monitoramento conjunto da vazão e qualidade, com detalhamento dos primeiros hidrogramas de cheia para separação de fontes pontuais de difusas.

Palavras-Chaves: Fósforo, Nitrogênio, Poluição Difusa, Uso e Ocupação do solo, Regressão linear, Árvore de decisão.

ABSTRACT

Phosphorus and nitrogen nutrients in surface water are related to the watershed land use and land cover (LULC) due to substances generation and deposition in the soil that may reach the river by the action of rain or direct runoff. Pollution sources scattered in the basin may modify the surface water quality and are referred to as diffuse or non-point sources. Identification and quantification of pollution load from diffuse sources are hard to obtain especially when they occur jointly with point sources, as industry effluents and domestic sewage. The Velhas river watershed is an example of such a situation, due to the presence of diffuse and point sources, complicating the separation of pollutant loads from the two types of sources and, related to diffuse sources, the determination of its relationship with LULC. This thesis has as primary goal to analyze the relation between LULC and the spreading of phosphorus and nitrogen nutrients into Velhas river's surface waters, considering potential and limitations of the analysis. The evaluated methods use mainly multiple regression and decision tree to create models. We proposed an index that relates LULC categories, such as pollution source or filter for surface water, and a hydrograph-based method for estimating diffuse and point loads. We obtained the LULC categories considering three different areas of analysis: watershed, exclusive contribution area and buffers. The quality data selection for rainy and dry periods also use different approaches: rainy quarter, rainy period and hydrograph ascension. For establishing the relation between nutrient concentration and LULC, the decision tree results are promising, reaching classification efficiency greater than 80% on nitrate and phosphorous classes. Multiple regression showed good results in relate ammonia, nitrite and dissolved oxygen with LULC, whose determination coefficient is about 80%. The nitrate and total phosphorous parameters did not produce valid models. Analyzing Pearson coefficient, we identified the presence of LULC category regarded as pollution sources (agro-pastoral and urban) and as pollution filters (vegetation). Thus, we propose an index (LCPI) that relates these sources and filters substituting the individual LULC categories. Results indicate that LCPI is better than LULC to explain the relationship. In 12 out of 15 models, the coefficient of determination (R^2) increased by 11 to 155% using LCPI instead of LULC. Finally, we proposed a method for estimating the diffuse and point loads in Velhas River, based on the analysis of hydrograph in storms events. Initial results showed the influence of rainfall on phosphorus and nitrate accumulation in the basin, in years with lower precipita-

tion/flow. Thus, it was possible to obtain model for nitrate and impossible for phosphorus. The results indicated the need of jointly monitoring quality and flow, with detailing of the first hydrographs about storms events, aiming to split diffuse and point sources.

Key-words: Phosphorus, nitrogen, Diffuse pollution, Land use and land cover, Multiple regression, decision tree.

RÉSUMÉ

La présence de nutriments de phosphore et d'azote dans l'eau de surface est liée à l'utilisation et à l'occupation du sol (LULC) des bassins versants en raison de la production et du dépôt de substances dans le sol, qui peut atteindre le cours d'eau par l'action de la pluie et du ruissellement. Les sources de polluantes dispersées dans le bassin peuvent altérer la qualité de l'eau de surface et sont appelées sources diffuses ou non-ponctuelles. L'identification et la quantification de la charge de polluantes provenant de sources diffuses est difficile à obtenir en particulier quand ils proviennent de sources mixtes non-ponctuelles et ponctuelles, telles que le lancement des effluents domestiques et industriels. Le bassin de la rivière Rio das Velhas est un exemple de ce type de situation en raison de la présence de sources diffuses et ponctuelles, ce qui complique la séparation des charges en polluantes des deux types de sources, et pour les sources diffuses, la détermination de leur relation avec LULC. La thèse a pour objectif principal l'analyse de la relation entre LULC et la libération de phosphore et d'azote vers les eaux de surface du Rio das Velhas, en tenant compte des avantages et des limites de ce type d'analyse. Les méthodes évaluées utilisent principalement la régression multiple et les arbres de décision pour créer des modèles. Pour le développement du travail il est proposé un indice qui divise les catégories de LULC qui servent de source avec celle qui filtrent la pollution des eaux de surface. De plus, une méthode originale est décrite pour l'estimation des charges diffuses et ponctuelles basée sur le fluviogramme. Les catégories de LULC sont obtenues en considérant trois domaines spatiaux d'analyse: des bassins versants, la zone de contribution exclusive et des bandes de largeur différentes le long des marges du cours d'eau. La sélection des données de qualité de l'eau pour les périodes de pluie et de sécheresse utilise trois approches différentes: trimestre pluvieux, saison des pluies et la montée de l'hydrogramme. L'utilisation d'arbre de décision s'est montré prometteur pour expliquer la relation concentration de nutriments et LULC atteignant un rendement supérieur à 80% pour les deux variables évaluées: le nitrate et le phosphore total. L'utilisation de la régression multiple a également montré de bons résultats en ce qui concerne l'ammoniac, les nitrites et l'oxygène dissous avec un coefficient de détermination d'environ 80%. Les paramètres phosphore total et nitrates n'ont pas permis de bons résultats avec la régression. Grâce à l'analyse du coefficient de Pearson, il a été possible d'identifier les catégories de LULC considérées comme sources de pollution (activités agro-pastorales et urbaines) ainsi

que celles qui servent comme filtres de polluants (végétation). L'utilisation d'un indice sources/filtres (LCPI) pour remplacer l'utilisation des différentes catégories de LULC a favorisé une amélioration du coefficient de détermination de 12 des 15 modèles valides. Enfin, nous avons proposé une méthode pour estimer les charges ponctuelles et diffuses dans la rivière Velhas, basé sur l'analyse du fluviogramme. Les résultats ont montré l'influence des précipitations sur l'accumulation du phosphore et du nitrate dans le bassin, au cours des années de basses précipitations / débits, peuvent être obtenus modèle pour le nitrate, mais pas pour le phosphore. Les résultats indiquent la nécessité d'harmoniser (aux mêmes points et en même temps) les relevés de débits et de la qualité de l'eau, avec des hydrogrammes de crue précis pour la séparation des sources diffuses des sources ponctuelles.

Mots-clés: Phosphore, Azote, Pollution diffuse, Occupation du sol, Régression linéaire, Arbre de décision.

Dedicatória

Dedico este trabalho aos meus pais Geny e José Valter e a meu marido Anderson e filhos Lucas e André.

Agradecimentos

“ O rio atinge seus objetivos porque aprendeu a contornar os obstáculos. Lao-Tsé ”

À Deus, pai, que me ilumina e me dá forças para o cumprimento de minha missão.

À todos que de forma direta ou indireta me auxiliaram na conclusão desta etapa de minha vida.

Aos meus filhos que com carinho, compreensão e muito amor me incentivaram a concluir este projeto.

Ao Anderson, meu companheiro de vida, pelo amor, compreensão e paciência durante todo este percurso.

Ao Anderson, meu colega de trabalho, meu muito obrigado pelo auxílio com as ferramentas e conceitos na área de Ciência da Computação, sem a sua ajuda tudo seria muito mais difícil.

À minha mãe e irmãs: Geny, Lucimar e Luziane por todo o amor, carinho e apoio sempre.

Ao meu amado pai, que mesmo tendo cumprido sua jornada, sempre se faz presente na minha vida por meio de seus ensinamentos.

Aos meus segundos pais Geralda e Guimarães, que me deram o suporte necessário com meus filhos para que eu pudesse realizar o doutorado.

Aos meus orientadores Philippe e Eber, meu muito obrigado pelos ensinamentos e amizade em todos os momentos.

À todos os amigos e colegas do IGC, em especial: Marília, Eric, Luiza, Paulo e Evelyn.

Ao CEFET-MG pela minha liberação para realizar o doutorado.

Ao Instituto Mineiro de Gestão das Águas (IGAM) pela concessão de dados do monitoramento de qualidade de águas da Bacia do Rio das Velhas.

À Agência Nacional de Águas (ANA) e à Companhia Energética de Minas Gerais (CEMIG), pela concessão de dados de vazão.

Ao Instituto Estadual de Florestas (IEF), pela concessão das imagens do sistema RapidEye.

Ao INMET (Instituto Nacional de Meteorologia), pelo fornecimento das normais climatológicas.

SIGLAS E SÍMBOLOS

- AD: Árvores de Decisão
- AI: Inteligência Artificial
- AAM: Algoritmos de Aprendizagem de Máquina
- a_i : No pai
- A_Q : Watershed contribution area for S_Q
- A_{WQ} : Watershed contribution area for S_{WQ}
- ANA: Agência Nacional de Águas
- AP: Categoria de LULC Agropastoril
- ArcView: Software de GIS
- $AU = U_b$: Categoria de LULC Área Urbana
- AV: Estação de Monitoramento de qualidade de águas da Parte Alta da Bacia do Rio das Velhas pertencente ao IGAM
- BMP: Best Management Practices
- B: Faixa localizada nas margens do curso de água
- BV: Estação de Monitoramento de qualidade de águas da Bacia do Rio das Velhas pertencente ao IGAM
- C: Concentração dos parâmetros de qualidade
- $CA = Gl$: Categoria de LULC Campo
- CART: Algoritmo de árvore de decisão
- $CA_{WQ/Q}$: Area Compensation Factor (dimensionless)
- $Cf_{WQ/Q}$: Area Compensation Factor (dimensionless)

- CE = Sa : Categoria de LULC Cerrado
- C4.5: Algoritmo de árvore de decisão
- DBO: Demanda Bioquímica de Oxigênio
- DO: dissolved oxygen
- DQO: Demanda Química de Oxigênio
- DL: Carga de Poluição Difusa
- ECA: Área de Contribuição Exclusiva (Português) ou Exclusive Contribution Area
- ID3: Algoritmo de árvore de decisão
- ID3: Inductive Decision Tree
- IDL: Carga Difusa Incremental
- IEF: Instituto Estadual de Florestas
- ILx: incremental loads do parâmetro X
- IGAM: Instituto Mineiro de Gestão das Águas
- IQA: Índice de Qualidade das Águas
- I_{Q-QA} : Índice que relaciona a PermQ/PermQA
- I_{Q-P} : Índice que relaciona a PermQ/PermQA para o fósforo
- I_{Q-N} : Índice que relaciona a PermQ/PermQA para o nitrato
- J48: Nomeclatura atribuída ao C4.5 no WEKA
- KDD: Knowledge Discovery in Database
- Lx:Load
- C: Concentration Parameters
- LCPI: Land Cover Pollution Index
- LN_i : Carga de Nitrato (chuva (i=1) e Seca (i=2))
- LTP_i : Carga de Fósforo (chuva (i=1) e Seca (i=2))
- LULC: Uso e Ocupação do Solo (Português)ou Land Use Land Cover (Inglês)

- LU: Land Use
- L_x : Carga incremental do parâmetro x
- $M = Ft$: Categoria de LULC Floresta
- $MI = RM$: Categoria de LULC Mineração e Afloramento
- MD: Mineração de dados
- MDE: Modelo Digital de Elevação
- M_j : *No* filho
- MMA - Ministério do Meio Ambiente
- N: Nitrato
- NO_3^- : Nitrate
- NO_2^- : Nitrite
- PermQ: Permanência da Vazão
- PermQA: Permanência dos parâmetros de qualidade
- Q : Vazão
- QA : Parâmetro de Qualidade de Água
- Q_{WQ} : daily discharge
- RE: Categoria de LULC Reflorestamento
- RMSE: Erro médio quadrático
- R_f : Categoria de LULC Mata Ciliar
- R^2 : Coeficiente de Determinação
- RNA: Redes Neurais
- RMBH: Região Metropolitana de Belo Horizonte
- SRTM: Shuttle Radar Topography Mission
- SWAT: Model Soil and Water Assessment Tool
- S_{WQ} : Water quality station area

- S_Q : discharge station area
- TDIDT: Top-down induction of decision tree
- UFMG: Universidade Federal de Minas Gerais
- UIDL: Carga difusa incremental unitária
- UIDLTP: Carga difusa incremental unitária de fósforo total
- UIDLN: Carga difusa incremental unitária de nitrato
- UILx: Unit Incremental Loads of x parameters
- TAN: Total ammonia nitrogen
- TP: Fósforo Total
- UTM: Universal Transverse Mercator
- WEKA: Software Waikato Environment for Knowledge Analysis
- WQ: Water Quality
- WQD: Water Quality Data
- WDD: Water Discharge Data
- WS: Bacia Hidrográfica

Capítulo 1

PREFÁCIO

A TESE foi organizada de forma a apresentar os artigos científicos produzidos durante o período de trabalho. O texto foi estruturado com 8 capítulos. O primeiro, Introdução, fornece uma visão geral da pesquisa, a problematização realizada a partir da revisão bibliográfica, as hipóteses formuladas, os objetivos, a justificativa da escolha da área de estudo, a estrutura e encadeamento dos artigos para atender as hipóteses e objetivos previstos na tese.

O Capítulo 2 apresenta a área de estudo, considerando aspectos físicos e biológicos: clima, geologia, relevo e recursos hídricos (quantidade e qualidade) e sócio-econômicos: população e uso do solo. No Capítulo 3, é realizada a revisão da literatura sobre alguns conceitos norteadores da pesquisa relacionados à poluição difusa, curva de permanência e árvore de decisão.

A metodologia e os resultados obtidos em atendimento às hipóteses formuladas são apresentadas nos Capítulos 4, 5, 6 e 7, que trazem os artigos desenvolvidos.

O 1^o e o 2^o artigo já se encontram aprovados e publicados nas revistas RBRH - Revista Brasileira de Recursos Hídricos e Environmental Monitoring and Assessment, respectivamente. O 3^o artigo foi submetido a revista CATENA. Somente o 4^o artigo ainda não foi submetido para publicação.

O Capítulo 8 apresenta as considerações finais sobre o atendimento às hipóteses/objetivos do trabalho e sugestão de trabalhos futuros. As referências bibliográficas utilizadas nos Capítulos 1, 2, 3 e 8, estão listadas ao final da tese, uma vez que as referências dos artigos foram inseridas nos capítulos de apresentação dos mesmos. O apêndice, apresentado após as referências bibliográficas, traz um exemplo de aplicação de árvore de decisão utilizando o algoritmo C4.5 e a estatística descritiva dos dados utilizados nos artigos.

Capítulo 2

INTRODUÇÃO

2.1 Aspectos Gerais

A bacia hidrográfica pode ser entendida como a área natural de captação da precipitação responsável pela sua transformação em escoamento. O volume escoado por meio de uma dada seção exutória está relacionado com características físicas (tipo de solo, cobertura do solo e declive, entre outros) e climáticas (quantidade e intensidade da chuva, evapotranspiração, etc) da bacia e influencia a qualidade das águas superficiais e subterrâneas.

Alterações na qualidade das águas superficiais são denominadas poluição e podem ter como origem fontes pontuais e/ou difusas. As fontes pontuais são caracterizadas por despejos que apresentam um único ponto de lançamento no curso de água, como exemplo, efluentes domésticos e industriais. As fontes poluidoras difusas são aquelas distribuídas sobre a superfície da bacia, podendo atingir o curso de água em pontos diversos, por exemplo, agricultura e pecuária, e tem no escoamento superficial seu principal veículo de transporte.

No curso de água, as fontes poluidoras pontuais e difusas aparecem conjuntamente, sendo difícil sua separação em função de processos de diluição, dispersão, fixação aos sedimentos e assimilação por microorganismos e vegetais.

Alguns estudos sobre poluição difusa selecionam diferentes variáveis de qualidade de água e avaliam a variação no comportamento da concentração (ou da carga) em diferentes regimes hídricos, normalmente separando períodos de seca e chuva (Sliva e Williams, 2001; Maillard e Santos, 2008; Li *et al.*, 2015; Outram *et al.*, 2016). A variação no comportamento das variáveis de qualidade de água é então associada a características físicas da bacia (uso e ocupação do solo, declividade, entre outras) para explicar a prevaência de fontes difusas (Maillard e Santos, 2008; Gorsevski *et al.*, 2008; Davis *et al.*, 2014; Li *et al.*, 2015; Outram *et al.*, 2014).

A associação entre as variáveis de qualidade e as características físicas é realizada, em muitos estudos, por meio do uso de modelos de regressão múltipla para obter a relação entre atributos ambientais (variáveis dependentes e independentes) e desenvolver modelos prediti-

vos, (Sliva e Williams, 2001; Grunwald e Qi, 2006; Maillard e Santos, 2008). Normalmente as características físicas da bacia de contribuição são representadas por: uso e tipo de solos, declividade, distância das fontes poluidoras, cobertura vegetal; e correlacionadas com as propriedades físico-químicas da água definidas através de seus parâmetros, (Mattikalli e Richards, 1996; Basnyat *et al.*, 1999; Fisher *et al.*, 2000; Wang, 2001; Maillard e Santos, 2008).

Em estudo desenvolvido para a bacia do Rio das Velhas, Maillard e Santos (2008) avaliaram a relação entre o uso do solo e a qualidade das águas superficiais de 15 estações de monitoramento. O uso do solo foi subdividido em 6 classes (floresta ribeirinha; solo exposto; cerrado; floresta; agropastoril; floresta plantada e urbano) obtidas para 5 diferentes faixas e para a área de contribuição exclusiva (ECA) entre dois pontos de monitoramento de qualidade. A qualidade da água, tomada pela diferença na concentração entre dois pontos de monitoramento, foi avaliada para cinco parâmetros (turbidez, nitrato, nitrito, fósforo e coliformes termotolerantes) e um índice de qualidade (IQA) durante os meses de janeiro (chuva) e julho (seca).

Para o período de chuva, as classes agropastoril e urbano contribuíram para o aumento na concentração de nitrato, enquanto que floresta contribuiu para a sua redução. Os modelos obtidos para o nitrito e para o fósforo, considerando os dados do período chuvoso, produziram melhor explicação da relação qualidade de água e uso do solo, comparados aos modelos gerados com dados do período seco.

O aumento da quantidade de matéria orgânica no curso de água, no período de chuvas, foi atribuído, pelos autores, ao maior volume escoado superficialmente pelas áreas de análise (ECA e Faixas).

Em estudos recentes, buscou-se estabelecer a relação entre a variação da vazão e da qualidade da água para separar fontes pontuais de difusas, e, assim, definir métodos que possam auxiliar a separação dessas fontes (Han *et al.*, 2010; Lang *et al.*, 2013; Bowes *et al.*, 2008; 2014). Segundo Bowes *et al.* (2008), o lançamento de efluentes pontuais no curso de água produz concentrações menores para as maiores vazões em função da diluição promovida. Já a poluição difusa tende a aumentar com o aumento da vazão. O aumento da carga difusa associado ao aumento da vazão foi estudado por Lang *et al.* (2013). Os autores verificaram que precipitações de maior intensidade produziram picos de carga de nitrogênio e fósforo na água, diferente do observado para chuvas de menor intensidade, onde a infiltração promoveu o retardo e a redução destes picos. Nas chuvas mais intensas, o volume infiltrado foi menor e gerou maior volume de escoamento superficial promovendo a limpeza da superfície de bacia.

Considerando a variação na carga de poluentes em função da vazão, Bowes *et al.* (2008; 2014) desenvolveram um modelo para separar a carga de fósforo presente na água em difusa e pontual, tendo como variável independente a vazão escoada. Por meio da análise do modelo proposto, os autores identificaram, para cada estação estudada, a vazão onde a entrada de

fósforo de maneira pontual é igual à entrada de fósforo de maneira difusa e também quais as bacias como maior geração de carga difusa de fósforo.

A partir dos estudos apresentados, é possível verificar que o estabelecimento da relação uso do solo, vazão e qualidade das águas superficiais envolve várias áreas de conhecimento e trabalha com grande volume de informação, necessitando de ferramentas com capacidade de análise de volumes cada vez mais expressivo de dados.

Considerando os estudos aqui apresentados, verificamos que a busca por informações em banco de dados com grande volume de informações pode ser auxiliada por ferramentas de aprendizagem de máquina, que apresentam elevado potencial para uso neste processo, uma vez que envolvem técnicas computacionais capazes de predizer um ou vários modelos a partir de um conjunto de dados. Aplicações dessas ferramentas na área de meio ambiente vem apresentando excelentes resultados. Entretanto seu uso ainda é bastante reduzido na avaliação da relação entre fontes poluidoras e qualidade das águas (Chau, 2006).

Dentre as técnicas de aprendizagem de máquina há a árvore de decisão, que produz modelos que seguem uma estrutura hierárquica e são aplicados em várias áreas ambientais, para descobrir estruturas ocultas em conjuntos de dados complexos e prever as características de uma classe escolhida (variável dependente) a partir de um conjunto de variáveis preditoras significativas (Breiman *et al.*, 1984).

Árvores de decisão foram utilizadas em previsões e análises ecológicas para quantificar as relações entre: espécies e propriedades da paisagem ambiental (Moisen e Frescino, 2002; De'Ath, 2002); relação chuva-vazão (Solomatine e Dulal, 2003), nível de água-vazão (Bhattacharya e Solomatine, 2005) e uso do solo, precipitação, vazão com o teor de fósforo presente na água (Schärer *et al.*, 2006). Assim, considerando a discussão apresentada, foram definidas as hipóteses e objetivos apresentados na sequência.

2.2 Hipóteses e Objetivos

De modo geral, os trabalhos citados na seção anterior têm como objeto de estudo a relação entre uso do solo, como fonte de poluição difusa, e as alterações promovidas na água, apresentando os seguintes pontos de interesse comuns:

1. Conhecer a área de análise adequada para relacionar qualidade das águas e uso do solo;
2. Definir quais categorias de uso do solo funcionam como fonte ou filtro para poluição difusa;
3. Otimizar o uso das informações existentes de qualidade de água e vazão;
4. Testar diferentes abordagens para separar fontes poluidoras pontuais de difusas;

5. Avaliar como o uso do solo tomado a diferentes distâncias do curso de água pode influenciar a qualidade das águas.

Neste, contexto, se insere a presente tese que propõe um conjunto de abordagens para caracterizar a relação entre o uso do solo e as mudanças na qualidade da água de superfície na Bacia do Rio das Velhas, para os nutrientes fósforo, nitrato, nitrito e nitrogênio amoniacal. Assim, buscou-se avaliar as possibilidades e dificuldades que existem para o estabelecimento desta relação.

A Bacia do Rio das Velhas foi escolhida como área de estudo pelo fato de possuir uma série longa de informações hidrológicas e de qualidade das águas superficiais, bem como existência de fontes poluidoras (pontuais e difusas) e pela sua importância histórica, econômica e estratégica para a gestão de recursos hídricos. Para tanto foram definidas quatro hipóteses de análise:

1. É possível utilizar árvores de decisão no estudo da relação dos nutrientes nitrato e fósforo com as categorias de uso e ocupação do solo (LULC);
2. Modelos de regressão múltipla são adequados para relacionar LULC às alterações na carga de nitrogênio amoniacal, nitrito, nitrato e fósforo;
3. A adoção de diferentes áreas de análise (bacia - WS, área de contribuição exclusiva - ECA ou Faixa - B) altera os resultados obtidos para a relação LULC e nutrientes avaliados;
4. A separação de cargas difusas das pontuais pode ser efetuada baseando-se no regime fluvial;

O objetivo geral da tese é apresentar as possibilidades e dificuldades existentes na definição da relação entre as alterações na concentração e carga de fósforo, nitrogênio amoniacal, nitrito e nitrato presentes nas águas superficiais e o uso do solo da bacia. Para tanto, foram definidos os seguintes objetivos específicos:

1. Avaliar a aplicação de árvores de decisão no estudo da relação dos nutrientes nitrato e fósforo com as categorias de uso e ocupação do solo (LULC);
2. Avaliar como a definição da área de estudo (bacia - WS, área de contribuição exclusiva - ECA ou Faixa - B) altera os resultados obtidos para a relação LULC e nutrientes;
3. Propor um método de separação de cargas difusas de pontuais utilizando o hidrograma de uma seção do curso d'água;

Como a abordagem é, por natureza, distribuída, foram definidas três tipos de zonas de influência e referidas pelas siglas WS, ECA e B:

1. Bacia hidrográfica (WS) compreende toda a área da bacia de drenagem no ponto de monitoramento da qualidade avaliado;
2. Área de contribuição exclusiva (ECA) considera somente a área de bacia hidrográfica existente entre duas estações de monitoramento de qualidade de água;
3. A faixa (B) é um recorte com largura fixa de WS ou da ECA, que tem como centro o eixo principal do rio. Ou seja, uma faixa de 100 metros de largura possui 100 metros em cada margem do rio.

2.3 Organização dos artigos

Para avaliar as hipóteses definidas foram realizadas quatro abordagens, descritas nos quatro artigos apresentados nos capítulos de 4 a 7 e esquematizadas nas Figuras 2.2 a 2.5.

Inicialmente as bases de dados, pertencentes a ANA, ANEEL e CEMIG, passaram por um processo de consistência, onde verificou-se que os dados de vazão e as curvas-chave das estações fluviométricas da ANEEL foram geradas até 12/2006. Foram então definidas as curvas-chave para o período de 01/2007 a 04/2013. Na sequência, foi realizada uma análise da continuidade dos fluviogramas para verificar possíveis erros nas cotas e realizar as correções necessárias. Os procedimentos utilizados para avaliação dos dados são descritos por Pinto (2016).

Os dados de qualidade também foram analisados quanto à sua consistência, para verificar à presença de *outliers*, seguindo metodologia proposta por Naghettini e Pinto (2007). A sequência de análises é apresentada na Figura 2.1.

A Figura 2.1 também apresenta, de maneira sucinta, a geração das base de dados de relevo, drenagem, localização das estações de qualidade e vazão e uso e ocupação do solo. O relevo, a localização das estações e a hidrografia foram utilizados para obtenção das áreas de estudos (WS, ECA, B). O uso e ocupação do solo foi recortado pelas áreas de bacia hidrográfica (WS), área de contribuição específica (ECA) e faixas (B).

Os procedimentos gerais adotados para avaliação das quatro hipóteses propostas na tese são descritos na Figura 2.2. As primeira e terceira hipótese foram avaliadas com dados de concentração de nitrato (NO_3^-) e fósforo total (TP), considerando a bacia hidrográfica de 17 pontos de monitoramento, em dois trimestres chuvosos de 2009 a 2011. A ferramenta utilizada na análise foi o algoritmo C4.5 que produziu árvores de decisão para os dois parâmetros estudados.

As hipóteses 2, 3 e 4 foram avaliadas com dados de carga para os parâmetros nitrito (NO_2^-), nitrato (NO_3^-), nitrogênio Amoniacal (TAN), fósforo total (TP) e oxigênio dissolvido (OD). Efetuou-se a verificação das hipótese 2 e 3 considerando o período de estudo para análise, definido por diferentes abordagens a partir do fluviograma. No caso da hipótese

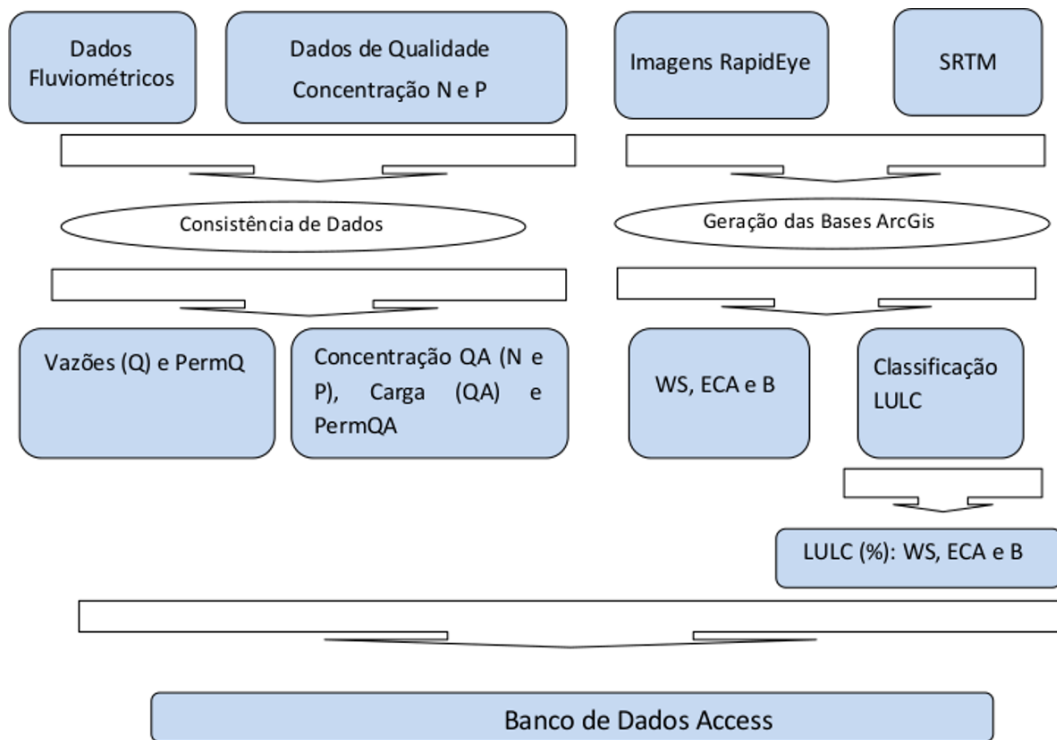


Figura 2.1: Montagem e consistência da base de dados: uso do solo por área de estudo (WS, ECA e B), qualidade e vazão.

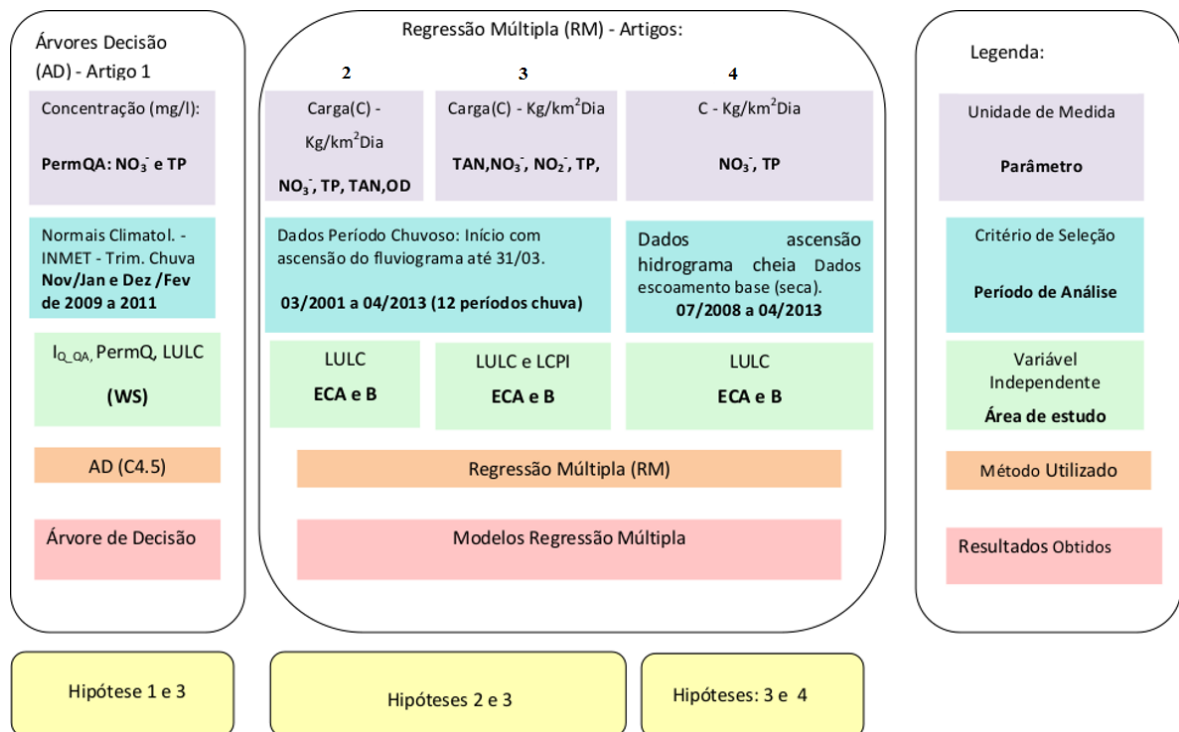


Figura 2.2: Esquema das análises realizadas nos 4 artigos

4 considerou-se todo o ano hidrológico. As áreas de estudo foram as ECAs e faixas de 17 estações de monitoramento de qualidade de água. A ferramenta utilizada foi regressão múltipla realizada no SPSS¹, com resultados apresentados por meio dos modelos obtidos.

A elaboração do primeiro artigo (Capítulo 4) seguiu os passos apresentados na Figura 2.3. O algoritmo de árvore de decisão (C4.5) (Quinlan, 1993) mostrou a relação da concentração de nitrato e fósforo total com a vazão e as categorias de uso e ocupação do solo (LULC).

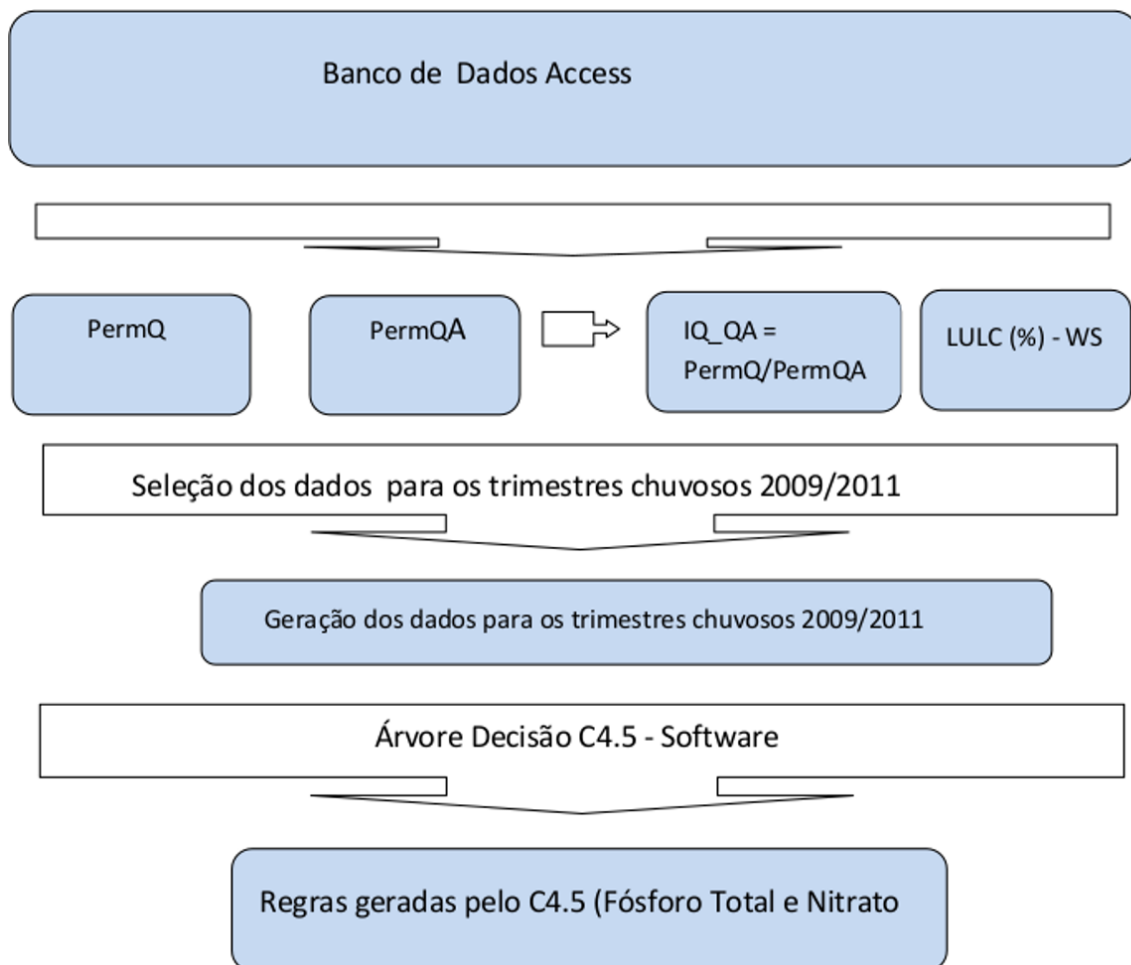


Figura 2.3: Fluxograma das etapas desenvolvidas na obtenção dos dados do Artigo 1

O segundo artigo foi elaborado utilizando a carga incremental entre dois pontos de monitoramento, para os parâmetros nitrogênio amoniacal, nitrato, fósforo total e oxigênio dissolvido. O uso e ocupação do solo foi obtido para cada área de contribuição exclusiva (ECA) e para 4 faixas com diferentes larguras (150, 300, 600 e 900 metros), recortadas em cada ECA (Figuras 2.4). Com isso, investigou-se como o uso e ocupação do solo, tomado para diferentes distâncias do curso de água, influenciou a qualidade da água. Os resultados são apresentadas no Artigo 2 (Capítulo 5).

¹<http://www-01.ibm.com/software/br/analytics/spss/>

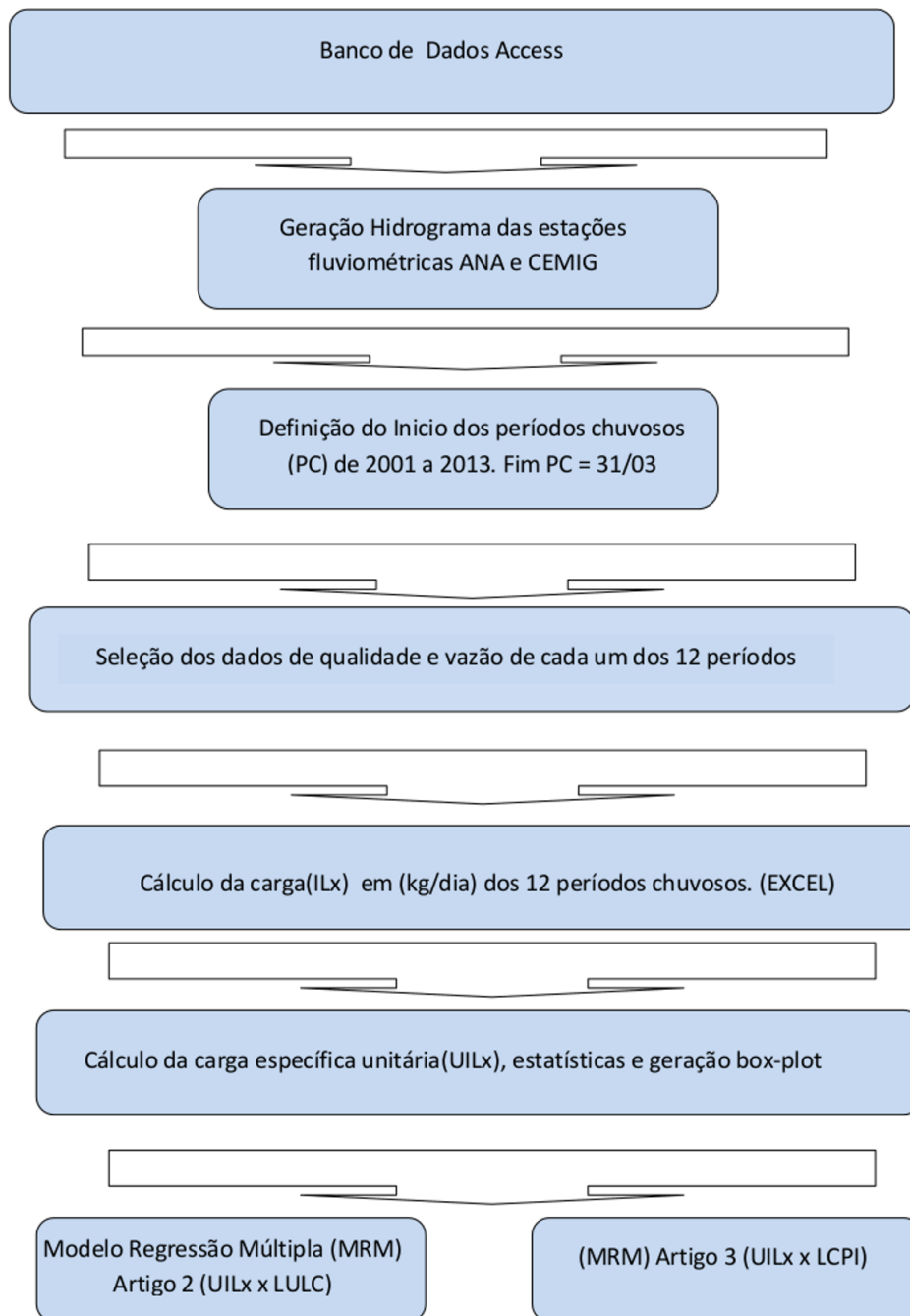


Figura 2.4: Fluxograma das etapas desenvolvidas na obtenção dos dados dos Artigos 2 e 3

O terceiro artigo (Capítulo 6) foi proposto como uma abordagem que agrupa categorias ditas fonte ou filtro de poluição difusa por meio do uso de um índice (LCPI). A classificação das categorias de uso do solo, em fonte ou filtro de poluição, foi realizada com base na análise da correlação entre o uso do solo e as variáveis de qualidade de água, para tanto foram avaliadas três medidas de correlação: os coeficientes de Kendall, Spearman e Pearson.

Os dois primeiros coeficientes (Kendall e Spearman) são baseados no ordenamento dos

dados e avaliam as relações monotônicas entre as variáveis e, também são resistentes a presença de "outliers". O coeficiente de Pearson mede a correlação linear entre as variáveis, que é um tipo específico de relação monotônica (Helsel e Hirsch, 1992). Os métodos adotados e os resultados obtidos são apresentados nas Figuras 2.2 e 2.4.

O quarto artigo apresenta um método para estimar a carga difusa gerada em uma dada bacia (Figura 2.5), considerando períodos de seca e cheia na produção de poluentes difusos e pontuais. Os resultados são apresentados no Capítulo 7.

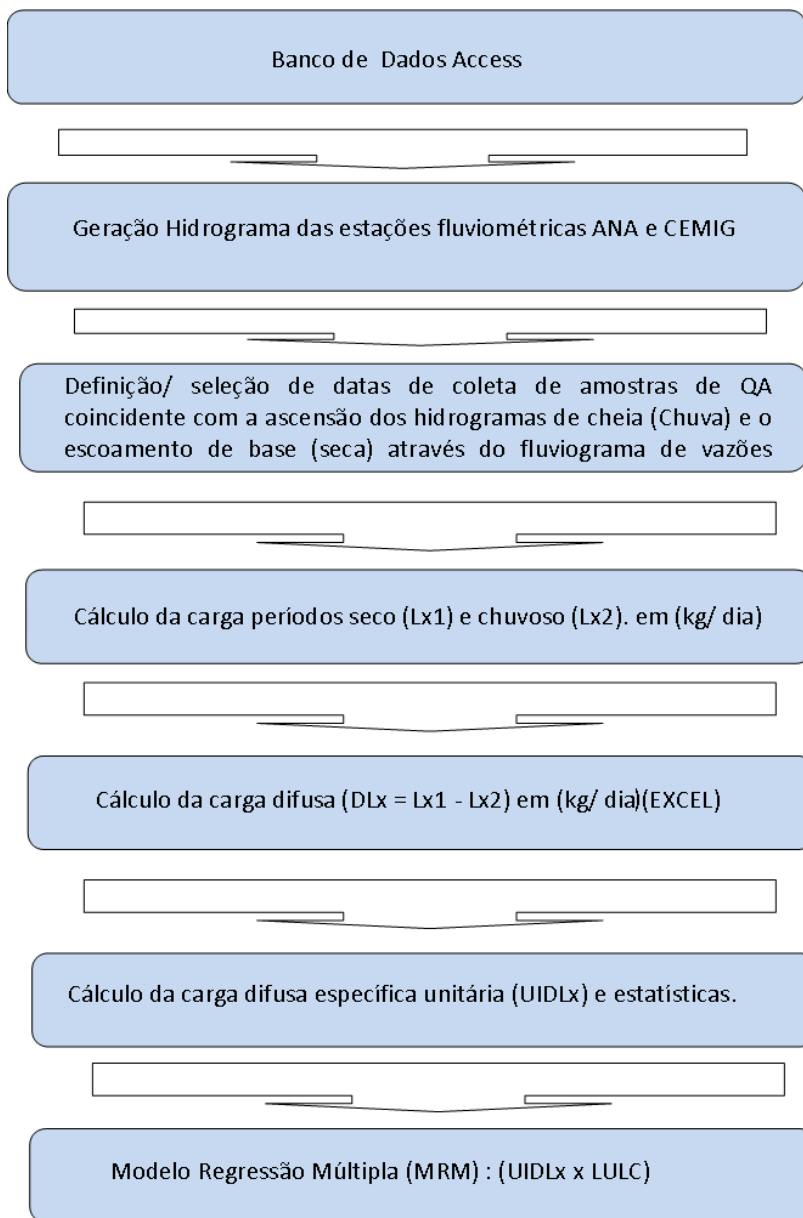


Figura 2.5: Fluxograma das etapas desenvolvidas na obtenção dos dados do Artigo 4

2.4 Padronização das nomenclaturas das estações de qualidade e vazão para os Capítulos 5, 6 e 7

O primeiro artigo (Capítulo 4) apresenta os códigos das estações de qualidade e vazão, conforme definidos pelo IGAM e ANA, respectivamente. Os três últimos artigos foram submetidos a revistas internacionais, assim, optou-se por renomear as estações de qualidade de água e vazão para facilitar o entendimento. Foram atribuídos novos códigos, em ordem sequencial crescente, para as estações de qualidade e vazão. A Tabela 2.1 e 2.2 apresenta as estações de qualidade e de vazão utilizadas no estudo com seus códigos no IGAM e ANA e os novos códigos atribuídos.

Considerando as 17 estações apresentadas na Tabela 2.1, temos que 16 localizam-se na calha principal do Rio das Velhas, somente a segunda estação (BV035) localizasse em um afluente. A estação BV035 foi utilizada para um melhor detalhamento da área de cabeceira da Bacia do Velhas.

Conforme pode-se observar na Figura 2.2, o período de análise dos dados de monitoramento, obtidos pelo IGAM, foi diferente em função da abordagem realizada em cada artigo. Algumas estações não aparecem em todos os artigos em função de não possuírem dados ou destes não apresentarem a discretização necessária no período de análise adotado. Duas das estações utilizadas no Artigo 1 (BV150 e BV151) foram retiradas dos artigos 2 (Capítulo 5) e 3 (Capítulo 6), que utilizaram dados a partir de 2001, visto que essas estações só entraram em operação em 2008. Por sua vez, foram inseridas duas outras estações, BV035 (Rio Itabirito) e a estação BV139 (Rio das Velhas) mantendo assim um total de 17 estações no 2^o e 3^o artigos.

O método de trabalho utilizado no Artigo 4 (Capítulo 7) não possibilitou o uso das estações BV013 e BV035, uma vez que não foi possível definir hidrogramas para as estações de vazão 41152000 e 41180500, com os dados diários de vazão média. Para a estação BV037, a data de coleta dos dados de qualidade não coincidiu com a ascensão e pico de hidrograma, desta forma iniciou-se a análise a partir da estação BV139.

Tabela 2.1: Padronização dos códigos das estações de Qualidade (QA) para os Capítulos 5, 6 e 7
Códigos Estações de QA

IGAM	Descrição	Latitude	Longitude	Capítulos 5 e 6 Artigos 2 e 3	Capítulo 7 Artigo 4
BV013	Rio das Velhas a montante da foz do Rio Itabirito	-20.21	-43.74	1	
BV035	Rio Itabirito a jusante da cidade de Itabirito	-20.23	-43.80	2	
BV037	Rio das Velhas a jusante da foz do Rio Itabirito	-20.13	-43.80	3	
BV139	Rio das Velhas a montante da ETA/COPASA, em Bela Fama	-20.06	-43.82	4	1
BV063	Rio das Velhas a jusante do Ribeirão Água Suja	-19.98	-43.81	5	2
BV067	Rio das Velhas a montante do Ribeirão Sabará	-19.94	-43.83	6	3
BV083	Rio das Velhas logo a jusante do Ribeirão Arrudas	-19.85	-43.87	7	
BV105	Rio das Velhas logo a jusante do Ribeirão do Onça	-19.80	-43.88	8	5
BV153	Rio das Velhas a jusante do Ribeirão da Mata	-19.71	-43.84	9	6
BV137	Rio das Velhas na Ponte Raul Soares, em Lagoa Santa	-19.54	-43.9	10	7
BV156	Rio das Velhas a jusante do Rio Jabuticatubas	-19.28	-44.01	11	8
BV141	Rio das Velhas na cidade de Santana do Pirapama	-19.02	-44.04	12	9
BV142	Rio das Velhas a jusante do Ribeirão Santo Antônio	-18.54	-44.18	13	10
BV152	Rio das Velhas entre os Rios Paraúna e Pardo Grande	-18.31	-44.24	14	11
BV146	Rio das Velhas a jusante do Rio Pardo Grande	-18.21	-44.35	15	12
BV148	Rio das Velhas na cidade de Várzea da Palma	-17.61	-44.70	16	13
BV149	Rio das Velhas a montante da sua foz no Rio São Francisco em Guaicuí	-17.2	-44.82	17	14

Tabela 2.2: Padronização dos códigos das estações de vazão para os Capítulos 5, 6 e 7

ANA	Descrição	Códigos Estações Fluviométricas		
		Latitude	Longitude	Capítulo 7 Artigo 4
41152000	Fazenda Vertentes	-20.23	-43.74	1
41180500	AVH-040	-20.22	-43.80	2
41199998	Honório Bicalho - Montante	-20.02	-43.82	3
41260000	Pinhões	-19.71	-43.81	4
41340000	Ponte Raul Soares	-19.56	-43.91	5
41410000	Jequitibá	-19.22	-44.02	6
41600000	Pirapama	-19.01	-44.04	7
41818000	Santo Hipólito	-18.31	-44.23	8
41990000	Várzea da Palma	-17.59	-44.71	9

Capítulo 3

ÁREA DE ESTUDO

A BACIA do Rio das Velhas com área de 28.887 km^2 ¹, localiza-se na região central do Estado de Minas Gerais (Figura 3.1), onde estão inseridos 51 municípios que abrigam uma população de aproximadamente 4.4 milhões de habitantes, sendo que 44 tem sede na bacia e 7 possuem a sede municipal fora da bacia.

O curso principal do Rio das Velhas, percorre uma extensão de 806.84 km, com nascente no município de Ouro Preto, a uma altitude de aproximadamente 1.500 m e foz na Barra do Guaicuí, município de Várzea da Palma, a uma altitude de 478 m (Tabela 3.1).

Tabela 3.1: Características gerais da Bacia do Rio das Velhas.

Extensão do curso principal	806.84 km
Localização	Região Central de Minas Gerais
Ocupação	51 municípios
População (IBGE, 2010)	4.409.896 habitantes no interior da bacia

Fonte: (CBH, 2015)

3.1 Aspectos Físicos

Os principais tributários do Rio das Velhas são os rios: Paraúna, Itabirito, Taquaraçu, Bicudo e Ribeirão da Mata. A drenagem da bacia é mais densa entre os afluentes da margem direita, fato associado às características topográficas e geológicas da região. Na margem esquerda, predominam as formações calcário-ardosianas que favorecem a circulação subterrânea (Santos, 2005). O padrão da rede de drenagem da maioria dos cursos d'água da bacia é do tipo dendrítico, comum a regiões de rochas cristalinas ou rochas do embasamento (IGAM, 2005).

A precipitação diminui de montante para jusante, com uma média anual de 1700 - 1600

¹Foi verificada uma diferença de 1% entre a área obtida na tese e a apresentada em CBH (2015) de 27.850 km^2 , valor considerado insignificante, sendo a diferença justificada em função da base de dados utilizada em cada estudo.

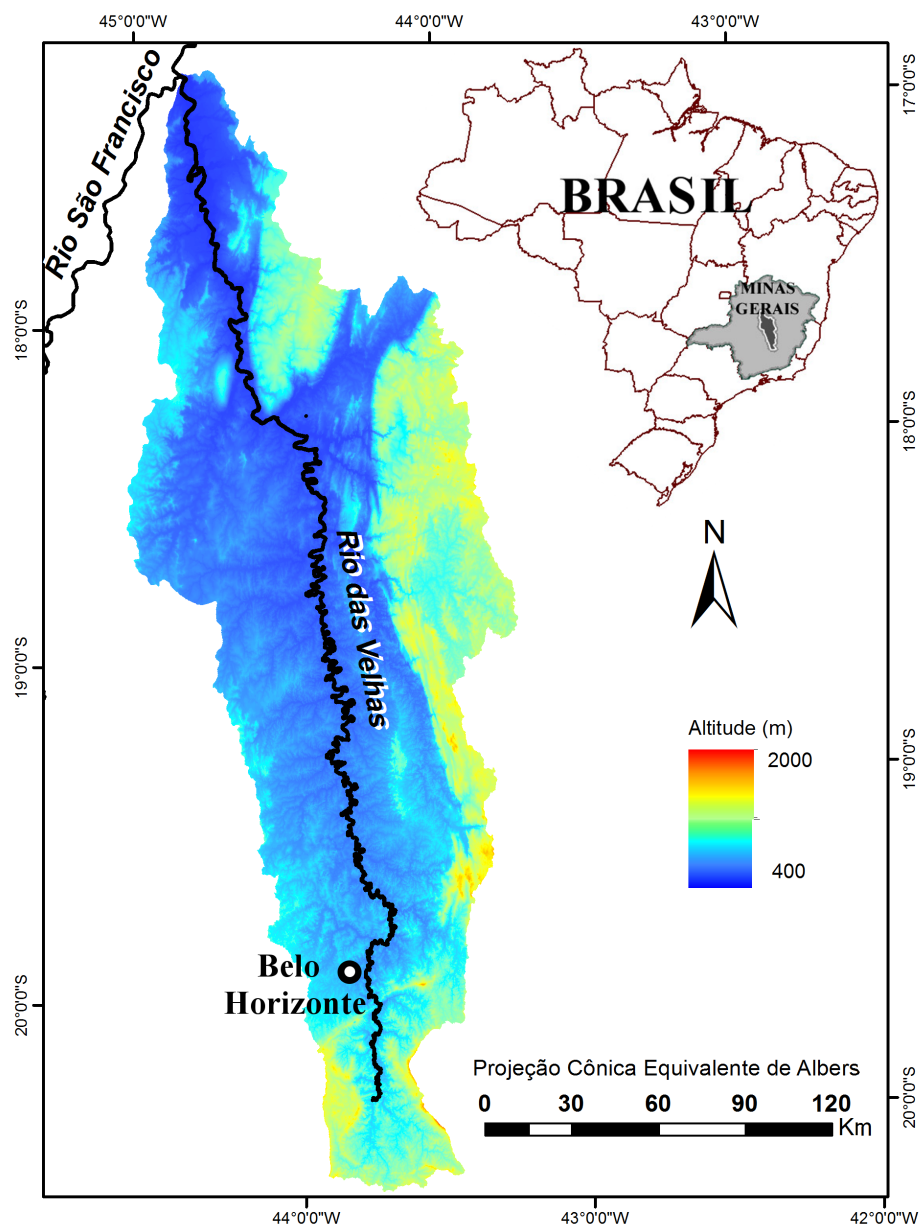


Figura 3.1: Mapa de localização da área de estudo

mm próximo a nascente do rio para cerca de 1100 mm na sua foz (Pinto *et al.*, 2011). O período chuvoso normalmente tem início em outubro e se estende até março.

A temperatura média anual varia entre 18^oC, na região de cabeceiras, e 23^o C, no extremo norte, junto à foz no Rio São Francisco. Entretanto, há alguns núcleos com temperaturas médias entre 19^o e 20^oC, ao redor de áreas elevadas da Serra do Espinhaço (IGAM, 2005).

A umidade relativa do ar na bacia varia entre 70 e 80%, com menores valores na baixa bacia e os maiores nos limites ocidentais da média bacia. Nas demais áreas, predominam os valores médios da ordem de 75%. Sazonalmente, as menores médias ocorrem no trimestre de julho a setembro, enquanto as maiores ocorrem entre os meses de novembro e abril, com

maiores incidências nos meses de dezembro e janeiro (IGAM, 2015).

As formas de relevo da bacia podem ser agrupadas em três categorias:

1. áreas ou formas aplainadas;
2. áreas dissecadas;
3. formas cársticas.

As áreas aplainadas, quando sobre planaltos, constituem superfícies tabulares ou superfícies onduladas. Tratam-se de antigas superfícies de aplainamento sobrelevadas e delimitadas por escarpamentos. As áreas aplainadas ocorrem também no interior de amplas depressões, caracterizando-se por vales denominados superfícies aplainadas ou onduladas. As formas dissecadas são encontradas nos compartimentos intermediários dos planaltos, no interior das depressões e nos maciços antigos. A caracterização das formas atuais está calcada nas mudanças climáticas do Período Quaternário. As formas cársticas ocorrem em rochas calcárias, apresentando uma morfologia peculiar, resultante de processos especiais de evolução e baseadas principalmente na dissolução e corrosão. Condicionam uma evolução de relevo com predominância de processos superficiais e subterrâneos. A geomorfologia cárstica e a hidrologia cárstica introduzem processos de corrosão como parâmetros de gênese e evolução, e condicionam as formas cársticas tais como cavernas, maciços residuais, dolinas, uvalas etc.

As unidades geomorfológicas da bacia hidrográfica do Rio das Velhas são : Planaltos do São Francisco, Depressão do São Francisco, Serra do Espinhaço e Quadrilátero Ferrífero (IGAM, 2005).

A bacia do Rio das Velhas oferece grande variedade de bens minerais entre metálicos e não-metálicos, com importantes reservas minerais de ferro, manganês, cobre, antimônio, arsênio, ouro, alumínio e urânio, mais concentradas na parte alta da bacia. No médio curso do Rio das Velhas, tem-se a exploração de calcário como matéria prima para a indústria de cimento, enquanto a extração de areia ocorre em toda bacia.

3.2 Aspectos Socioeconômicos

O Plano de Diretor da Bacia do Rio das Velhas apresentado em 2015, propõem uma nova divisão da bacia, sub-dividindo a mesmas nos trechos alto, médio alto, médio baixo e baixo de forma a tornar mais eficiente a sua caracterização, diagnóstico e prognóstico da qualidade de suas águas, assim como a efetivação dos planos propostos (Apêndice A).

De acordo com CBH (2015), a atividade agropecuária é uma importante fonte de contaminação difusa, geração de processos erosivos e retirada de cobertura vegetal natural. O manejo inadequado do solo gerou áreas degradadas, especialmente de pastagens. A lavoura

temporária, a lavoura permanente e a pecuária são as atividades inseridas nesta categoria. O primeiro caso diz respeito às culturas de curta e média duração, que após a colheita deixam o terreno disponível para novo cultivo ou para o aproveitamento da pecuária. Destacam-se as culturas de milho, feijão e cana-de-açúcar. A lavoura permanente agrupa todas as culturas de ciclo longo que permitem colheitas sucessivas, sem necessidade de novo plantio. Destacam-se as produções de banana, café e laranja. A pecuária corresponde às áreas abertas visando a produção de gado (de corte e leiteiro), pastoreio de equinos e outros animais em menor quantidade na bacia. Nestas áreas predominam vegetação campestre natural ou pastagem plantadas para o alimento dos animais.

A extensa mancha urbana representada pela RMBH, associada a concentração de atividade industrial, representa atualmente o principal fator de pressão sobre a qualidade dos recursos hídricos na bacia. A atividade industrial e de mineração são fonte de geração de carga poluidora. As lavras a céu aberto são fontes de erosão e as barragens não contam com procedimentos de segurança apropriados (CBH, 2015).

Diariamente são retirados do Rio das Velhas (calha principal e afluentes) uma vazão de $24.2 \text{ m}^3/\text{s}$. Destes 47.2, 26.6, 15.7 e 10.5 %, considerando os trechos alto, médio alto, médio baixo e baixo do Rio das Velhas.

O uso dos recursos hídricos na Bacia Hidrográfica do Rio das Velhas é variável conforme observa-se na Figura 3.2.

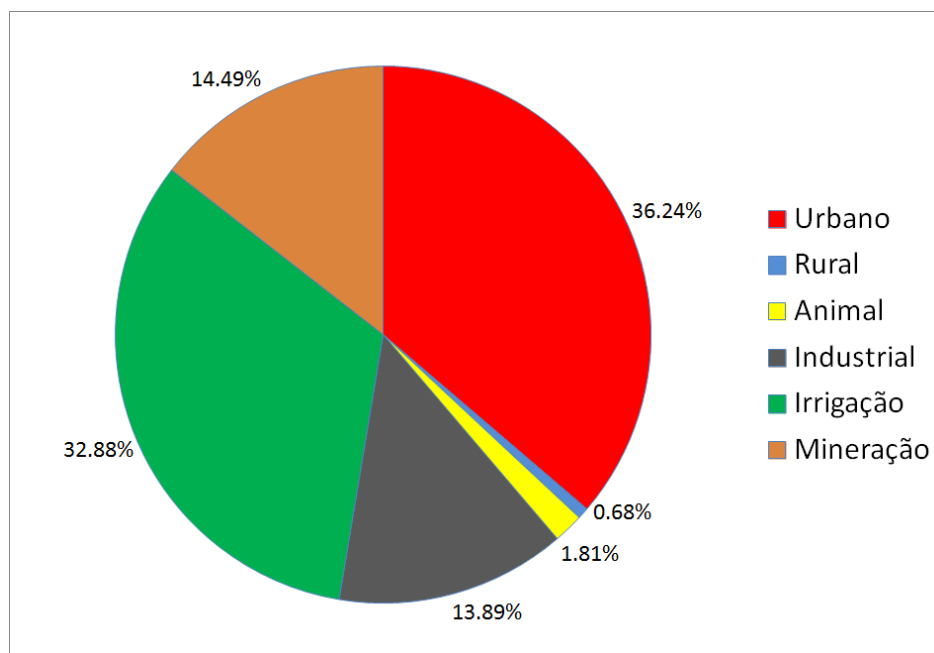


Figura 3.2: Percentual de vazões médias retiradas diariamente por tipos de usos da Bacia Rio das Velhas CBH (2015)

No Alto Rio das Velhas, esta vazão está direcionada ao atendimento de demanda com

abastecimento urbano, industrial e mineração. No trecho médio alto, os usos se concentram em rural, industrial e irrigação. No trecho médio baixo, os usos se concentram em rural, animal e irrigação. Já no trecho baixo, os maiores usuários são a irrigação e a dessedentação animal (Figura 3.3).

Na Figura 3.3 observa-se que no trecho médio alto, o uso do solo industrial possui uma importante participação em função da sua proximidade com a RMBH.

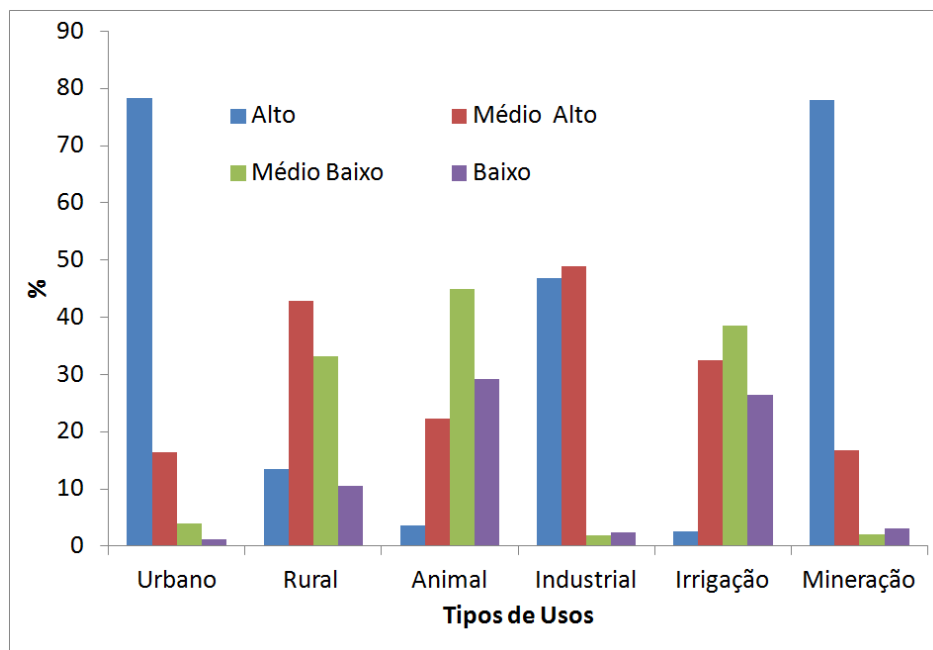


Figura 3.3: Percentual de vazões médias retiradas diariamente por tipos de usos nos trechos Alto, Médio Alto, Médio baixo e Baixo do Rio das Velhas. Adaptado de CBH (2015)

3.3 Qualidade das águas da bacia do Rio das Velhas

A qualidade das águas do Rio das Velhas é acompanhada em vários projetos em andamento na bacia entre os quais destacam-se os Projetos *Manuelzão* e *Águas de Minas*. O projeto Manuelzão, com cunho mais social, utiliza o peixe como bioindicador. Já o projeto Águas de Minas, com cunho mais científico, avalia a qualidade do rio por meio de análise dos parâmetros de qualidade.

O *Projeto Revitalização da Bacia do Rio das Velhas - Meta 2010* foi proposto pelo Projeto Manuelzão em parceria com a Secretaria de Meio Ambiente de Minas Gerais - SEMAD e buscou agrupar interesses do Governo do Estado, das Prefeituras dos municípios localizados na bacia, da sociedade civil organizada e da população em geral. Teve início em 2003, com a expedição “Manuelzão desce o Rio das Velhas” que mobilizou diversos moradores ao longo de todo o seu curso. A partir das observações obtidas com a expedição, foi possível

identificar os principais focos de degradação e as ações necessárias a sua reversão, elaborando assim a Meta 2010. O lançamento de esgoto sanitário doméstico sem tratamento foi identificado como a principal causa da degradação, sendo os Ribeirões Arrudas e Onça foram apontados como os epicentros da degradação, seguidos de perto pelo Ribeirão da Mata.

Em 2007, a Meta 2010 passou a ser um dos Projetos Estruturadores do Governo de Minas Gerais e assim, com a adesão do Governo do Estado, passou a ter uma dimensão mais abrangente unindo esforços e recursos públicos e privados para comprovar a todos os agentes envolvidos sua viabilidade técnica, relevância social e racionalidade estratégica, além de convocar a sociedade para um objetivo com prazos e metas definidos.

Do ponto de vista do Estado, após a consolidação do processo de monitoramento da qualidade das águas em elaboração pelo Igam por meio do *Projeto Águas de Minas*, ficou claro que não bastava verificar a qualidade das águas, mas seriam também necessárias ações visando sua melhoria.

De acordo com Polignano *et al.* (2014), vários foram os resultados positivos obtidos pela Meta 2010, sendo que o maior, mais visível e simbólico foi à volta dos peixes. Algumas espécies já podem ser capturadas na região próxima a Lagoa Santa.

Em 2009 foi realizada uma grande expedição pelo Rio das Velhas, saindo da Cachoeira das Andorinhas - Ouro Preto até a foz em Barra do Guaicuí-Várzea da Palma, oportunidade na qual foi possível comprovar a volta dos peixes na região do médio Velhas (Polignano *et al.*, 2014).

Segundo Polignano *et al.* (2014) a Meta 2010 atingiu 60% do esperado, pois o rio não só deixou de piorar, da forma vegetativa esperada, mas apresentou uma melhoria na sua qualidade, considerando a volta do peixe ao rio. Segundo a ONG, isto demonstrou na prática que a sociedade pode reverter o processo de degradação, desde que estabeleça esse objetivo como uma meta política pactuada entre sociedade e Estado.

O biomonitoramento realizado nos últimos dez anos pelo Projeto Manuelzão, pela Universidade Federal de Minas Gerais - UFMG constatou que os peixes que subiam cerca de 200 km na bacia em 2000, já foram identificados ao longo de 580 km do rio em 2010, chegando bem próximo às áreas consideradas mais degradadas na RMBH. O principal fator responsável pelos resultados obtidos pela Meta 2010 foi o volume de esgoto tratado pela Companhia de Saneamento de Minas Gerais (COPASA) na bacia do Rio das Velhas, que passou de 41 milhões de m^3 em 2003 para 85 milhões de m^3 em 2008, atingindo a meta de 127 milhões de m^3 de esgoto tratado em 2010.

Assim em 14 de agosto de 2010, foi assinado novo Termo de Compromisso para dar continuidade ao processo de revitalização com criação da Meta 2014 que tem como objetivo principal assegurar a volta dos peixes e a possibilidade de nadar na região metropolitana do Rio das Velhas. Para tanto, a Meta 2014, propõe três focos de atuação: recuperação da região mais degradada da calha do Velhas, que é a RMBH, conservação da bacia do Cipo-

Paraúna (reserva biológica natural da bacia do Velhas) e desenvolver ações de preservação e recuperação das dezenas de sub-bacias do Velhas, envolvendo todas as prefeituras e empresas das respectivas áreas hidrográficas e o Comitê da Bacia Hidrográfica do Rio das Velhas (CBH Velhas).

O monitoramento das águas na bacia do Rio das Velhas, instrumento importante para a Meta 2014, é realizado pelo *Projeto Águas de Minas* (IGAM) desde 1997. Atualmente, existem 82 estações de qualidade das águas na bacia do Rio das Velhas, distribuídas em diferentes projetos de Avaliação da Qualidade das Águas Superficiais (Projeto Pampulha, Projeto Alto Velhas, Meta 2014, entre outros) (IGAM, 2015).

A contaminação por tóxicos (CT) avaliada no Projeto classificou 22 estações com alta contaminação, durante as campanhas de medição de qualidade da água realizadas em 2015, ou seja, estas estações representam as piores condições de CT no estado de Minas Gerais. Vale destacar que das 22 estações listadas 15 encontram-se na bacia do Rio das Velhas IGAM (2015). O índice de qualidade de água (IQA) variou na bacia do Rio das Velhas de Muito Ruim (4.6%), Ruim (35%), Médio (40%), Bom (20%) Excelente (0.6%) . As principais fontes poluidoras indicadas como responsáveis pela degradação da qualidade das águas são: falta de tratamento de efluentes domésticos e industriais, pecuária e mineração.

Os dados produzidos pelo *Projeto Águas de Minas* são utilizados em diversos estudos dentre os quais tem-se: o tratamento estatístico das variáveis de monitoramento (Nonato *et al.*, 2007), a avaliação das alterações na qualidade das águas em função do uso do solo da bacia (Raposo *et al.*, 2011), e avaliação da presença de Cianobactérias, decorrentes dos índices de qualidade de águas medidos (Jardim, 2011), entre outros.

A identificação de áreas produtoras de sólidos em suspensão foi realizada por Raposo *et al.* (2011) para a bacia do Rio das Velhas. Foram mapeadas fontes pontuais e difusas de poluição em seis sub-bacias do alto Rio das Velhas. A turbidez no curso de água principal foi avaliada para as seis sub-bacias estudadas. Os resultados indicaram como fontes poluidoras: voçorocamentos, mineração, usos agropastoris e o lançamento de efluentes, os quais foram relatados como principais provedores de sedimentos aos rios. Nas bacias dos ribeirões da Prata e Sabará, a preservação de áreas verdes se refletiu numa baixa turbidez. Os autores destacam a necessidade do uso criterioso do parâmetro turbidez, pois a dinâmica sedimentar de algumas bacias é muito influenciada pela mineração, sendo caracterizada pela abundância de carga de leito, a qual não é passível de ser constatada pelo monitoramento da turbidez.

O diagnóstico realizado por CBH (2015) avaliou que o índice de qualidade (IQA) apontou para uma discreta melhoria na qualidade das águas do Rio das Velhas do período 1997/2003 para 2004/2012, sobressaindo a redução no percentual de nível Muito Ruim de 10% para 3.9% e o aumento do nível Médio de 31.2% para 38.2%, respectivamente. O diagnóstico ainda verificou que houve uma influência sazonal acentuada, aspecto também observado no estudo realizado por Calazans (2015). De acordo com CBH (2015) o IQA apresenta pior

condição na época de chuvas, sendo que cerca de 50% dos valores de IQA indicaram categoria Ruim nos dois períodos de tempo avaliados (1997/2003 e 2004/2012). Na época de estiagem, predominaram os níveis Médio e Bom.

Visando otimizar a rede de monitoramento, o estudo conduzido por Nonato *et al.* (2007) utilizou técnicas de análise estatística multivariada e análise exploratória, especificamente, análise de agrupamentos e análise de componentes principais, para 39 parâmetros físico-químicos monitorados no período de 2003-2004 no alto curso do Rio das Velhas.

Os resultados obtidos indicaram que apenas 22 parâmetros de qualidade descrevem as 4 componentes principais mais importantes, explicando 80% da variabilidade dos dados. Os autores concluíram que o lançamento de esgoto é a principal fonte poluidora do alto Rio das Velhas, haja visto que os principais parâmetros selecionados por componentes principais foram: nitrogênio amoniacal, nitrato, nitrito, fósforo, estreptococos, coliformes fecais, DBO e DQO. A presença de atividade mineradora na área foi caracterizada pelos parâmetros: sólidos em suspensão, turbidez e Ferro total.

Almeida (2013) avaliou a distribuição espacial da rede de monitoramento de qualidade das águas superficiais do Rio das Velhas (MG), utilizando o método da entropia, para estabelecer níveis de prioridade de manutenção das estações. Foram selecionados cinco parâmetros para estudo, Coliformes Termotolerantes, Arsênio Total, Fósforo Total, Turbidez e Oxigênio Dissolvido, considerando os maiores percentuais de violação à DN COPAM/CERH N°01/08 e 19 estações de monitoramento localizadas na calha do Rio das Velhas. Os resultados apontaram que as estações mais importantes na manutenção da rede foram a BV067, BV141, BV142 e BV149 e as que obtiveram a pior classificação e que, teoricamente, poderiam ser desativadas foram as estações BV013, BV139, BV150 e BV148.

Capítulo 4

REVISÃO DA LITERATURA

4.1 Poluição Difusa

O controle das fontes pontuais de poluição, por meio do tratamento de efluentes, possibilita verificar a existência de fontes poluidoras de origem difusa em uma bacia. Em decorrência disto, os Estados Unidos, em 1979, alteraram a seção 208 da *Federal Water Control Pollution Act* introduzindo medidas de controle da poluição difusa. A alteração da norma decorreu do reconhecimento de que esforços no controle da poluição pontual não ocasionaram a melhoria da qualidade das águas superficiais em função da existência de fontes difusas de poluição na bacia (Wicker, 1979; Wozniak, 1980; Wiens, 1980).

Segundo Wicker (1979) 92% dos sólidos em suspensão; 79% de nitrogênio, 53% de fósforo e 98% de coliformes termotolerantes eram decorrentes de fontes difusas de poluição, sendo a agricultura, pecuária, remoção da cobertura vegetal, construções, mineração, atividades reconhecidas com geradoras de grande volume de poluição para os rios nos Estados Unidos.

Assim mudanças nas práticas de uso e ocupação do solo como por exemplo: plantio em curvas de nível, para redução do escoamento e potencial de erosão do solo e manutenção da vegetação nas margens dos rios foram identificadas como melhores práticas de manejo (Best Management Practices -BMP's) e passaram a fazer parte de programas de controle da poluição difusa (Wozniak, 1980).

A partir deste momento diversos foram os estudos conduzidos para avaliar os tipos e a localização das fontes poluidoras difusas, as variáveis envolvidas no transporte de poluentes, entre outros aspectos. Modelos matemáticos e ferramentas computacionais passaram a auxiliar pesquisas nesta área (Phillips, 1989; Arnold *et al.*, 1998; Huang e Hong, 2010; Zhenyao *et al.*, 2012). Entretanto, ainda hoje, várias questões persistem e são objeto de investigação, tais como: qual a influência do uso e ocupação do solo na liberação de nutrientes para os cursos de água, considerando sua distribuição no espaço e no tempo, como a chuva e a vazão modificam a concentração e a carga dos nutrientes na água (Sliva e Williams, 2001; Maillard

e Santos, 2008; Bowes *et al.*, 2014; Li *et al.*, 2015).

Diferentemente de fontes poluidoras pontuais, a NPPS podem atingir grandes áreas em função do efeito de transporte pela água de chuva, irrigação ou neve levando poluentes para rios, lagos, reservatórios, oceanos. Dependendo das características do poluente, este pode ser levado por longa distância e ter efeitos drásticos na qualidade das águas (Maillard, 2010).

Isto torna o estudo da poluição difusa complexo, pois depende do uso do solo da bacia de contribuição, sendo também influenciado pelo escoamento superficial e sub-superficial, ou seja, o caráter sazonal e a capacidade de atingir extensas áreas são fatores que dificultam o seu controle.

Vários estudos buscam estabelecer a relação entre LULC e NPPS, considerando as diferenças sazonais de umidade, e nestes identificam-se algumas questionamentos comuns nas pesquisas (Arheimer e Liden, 2000; Sliva e Williams, 2001; Maillard e Santos, 2008; Chang, 2008; Ouyang *et al.*, 2010; Huang e Hong, 2010; Bowes *et al.*, 2014; Li *et al.*, 2015):

1. Quais os parâmetros de qualidade adequados à definição da relação LULC e NPPS;
2. Qual a área adequada para análise: bacia, área de contribuição exclusiva entre pontos de monitoramento - ECA ou faixas - localizadas nas margens dos rios;
3. Como as condições de umidade (seca e/ou chuva) influencia a relação LULC/NPPS ;
4. Qual a unidade de medida a ser considerada para os parâmetros de qualidade (concentração ou carga).

Sliva e Williams (2001) avaliaram a relação entre uso do solo, declividade e pedologia e a qualidade das águas em 3 bacias afluentes ao lago de Ontario - Canadá. Os autores utilizaram dados de 12 estações de qualidade de águas e 3 estações de vazão, monitoradas pelo Ministério do Meio Ambiente no período de 1990-1993, na análise de 10 parâmetros de qualidade. A seleção do período de análise levou em consideração a variação das vazões no curso d'água: primavera (janeiro-abril), verão (maio-agosto) e outono (setembro-dezembro), com vazões maiores na primavera e menores no verão. O uso do solo foi dividido em quatro categorias: campo, floresta, agricultura e urbano. A pedologia foi caracterizada por quatro tipos de depósitos: areia-cascalho-silte, areia, argila-areia e silte-areia. O parâmetro declividade foi avaliado pelo seu desvio padrão. A área de análise compreendeu a bacia de drenagem definida nos 12 pontos de monitoramento e a existência de uma faixa de 100 metros tomada a partir do curso de água.

Na primavera os melhores resultados, avaliados por meio do uso do coeficiente de determinação R^2 , foram obtidos para bacia hidrográfica quando comparada ao uso da faixa de 100 metros. Durante o verão os resultados obtidos para 5 dos 10 parâmetros avaliados, para ambas as áreas estudadas (faixa e a bacia), foram similares. Para coliformes termotolerantes

os resultados foram melhores quando se utilizou a faixa. As variáveis oxigênio dissolvido e temperatura não puderam ser explicadas por nenhum dos modelos utilizados (bacia e faixa).

O processo de análise utilizou inicialmente 9 variáveis explicativas, entretanto os modelos utilizaram somente 6 destas variáveis identificadas como relacionadas aos parâmetros de qualidade de água avaliados. Considerando cada parâmetro de qualidade de maneira isolada no máximo duas variáveis foram utilizadas. A variável explicativa mais importante foi uso do solo urbano, responsável pela produção de fontes não pontuais. Floresta e campo, foram relacionados como filtro de poluição. A presença de fontes pontuais de poluição dificultou o estabelecimento de um modelo adequado para poluição difusa, devido a dificuldade de separação das duas fontes.

A Tabela 4.1 apresenta alguns estudos realizados nos últimos 15 anos com o objetivo de avaliar a relação entre LULC e NPPS, e responder aos questionamentos apresentados anteriormente relativos as diferenças obtidas com relação à sazonalidade climática, às áreas adequadas à análise (bacia, ECA e faixas) e tipos de poluentes pesquisados na água.

Em síntese, verifica-se nos estudos apresentados na Tabela 4.1 que, para estabelecer a relação uso do solo/qualidade das águas, são utilizados vários parâmetros de qualidade (físicos, químicos e biológicos), entretanto o nitrogênio e fósforo são os que mais aparecem nos estudos, haja vista que estes estão ligados a várias fontes difusas de poluição.

Mais da metade dos estudos avaliados buscaram incorporar o efeito da sazonalidade, separando os dados para períodos de seca e chuva. A bacia hidrográfica foi a área mais estudada, seguida das faixas.

Tabela 4.1: Síntese de estudos sobre a relação LULC x NPPS

Autores	País	N ¹	P ²	Outros Parâmetros	N e P	Unidade	Área de Análise ³	LULC ⁴	Outras Variáveis ⁴	Período
Sliva e Williams (2001)	Canada	X	X	Al,Cl,Cu,OD,S ⁵ ,CTe ⁶ ,T	NH_4^+ , NO_3^- e PO_4	Kg/ha mês	WW e B	Ft,Ub,AP	Ge e s	Seca e Chuva
Gardi (2001)	Italy	X			NO_3^-	mg/l	WS	Ft,Ub,BS,AP	Q, So, s	chuva
Quinn e Stroud (2002)	New Zealand	X	X	Al,S,T,pH,COD,Tu	NH_4^+ , NO_3^- , TN, PRD e TP	mg / m ³	WS	Ft,Pa		Annual
Yang <i>et al.</i> (2007)	China	X			NH_4^+ , NO_3^- e TN	Kg/ha ano	WS	Ft, Ag, Ub		Annual
Poor e McDonnell (2007)	USA	X			NO_3^-	Kg /ha evento	WS	Ub, Ag e Ft	Q	eventos chuvosos
Maillard e Santos (2008)	Brazil	X	X	CTe, Tu	NO_3^- , NO_2^- e TP	mg/l	ECA e B	Ub,AP,BS,Ft,RF,Sa		Seca e Chuva
Diebel <i>et al.</i> (2009)	USA	X	X		TN e TP	Kg / Km ² ano	WS e B	Ub, Ag, AP, Ft, RF	s, Dd	Annual
Ouyang <i>et al.</i> (2010)	China	X	X		Po, Po(sed), NO e NO_3^-	Kg/ha ano	WS	FL, Ag e Pa	MP	Annual
Zucco <i>et al.</i> (2012)	Brazil	X	X		TN, NH_4^+ , NO_3^- , TP, PO_4 e NO_2^-	Kg/ha ano	ECA	Ub, Ag, Ft, Pa		Annual
Pinheiro <i>et al.</i> (2014)	Brazil	X	X	Al,OD,CTe	NH_4^+ , NO_3^- , NO_2^- e TP	mg/l	WS	Ft, Ag, Ub		Annual
Bu <i>et al.</i> (2014)	China	X	X	pH,OD,CE,S,Cl, SO_4 , DBO_5 , COD, Mn, e SiO_4	NH_4^+ , NO_3^- , NO_2^- , TN e TP	mg/l	WS	Ub, Ag, AP, Ft, RF, BS		Seca e Chuva
Li <i>et al.</i> (2015)	China	X	X		NH_4^+ , NO_3^- e TN	mg/l	WS e B	Pa, Ag, Ft, Ub, W	MP	Seca e Chuva
Outram <i>et al.</i> (2016)	United Kingdom	X	X		NH_4^+ , NO_3^- e TP	Kg/ha ano	WS	Ag, Ub, Pa, Ft, W	Seca e chuva	Chuva

N - Nitrogênio, P - Fósforo, Al - Alcalinidade, Cl-Cloreto, Cu - Cobre, OD - Oxigênio Dissolvido, S- Sólidos, T- Temperatura, pH - Potencial Hidrogênico.

CE- Condutividade Elétrica, COD carbono Orgânico Dissolvido, Col - Coliformes, Tu - Turbidez, SiO_4 - Silica.

¹TN- Nitrogênio Total, NH_4^+ - Amônia, NO_3^- - Nitrato, NO_2^- - Nitrito, NO - Nitrogênio Orgânico ²TP - Fósforo Total, PO_4 - Fosfato, PRD - Fósforo Reativo Dissolvido, Po - Fósforo Orgânico, Po(Sed)- Fósforo Orgânico incorporado sedimento, ³ Bacia(WS), Área de Contribuição Exclusiva (ECA), Faixa(B) ⁴ Sa-Cerrado; Ub-Urbano; Ag-Agricultura; Ft-Floresta; AP-Agricultura e Pastagem; Pa-Pastagem, RF - Floresta ripária, W-água, BS-Solo exposto, Geologia (Ge), Declividade (s), Métricas da Paisagem (MP), Solo - So, Vazão- Q, HR - Hierarquia Fluvial ⁵ST - Sólidos Totais, SS - Sólidos Suspensão, SSV - Sólidos Suspensão Voláteis, DBO - Demanda Bioquímica de Oxigênio ⁶ CTe - Coliformes Termotolerantes

4.2 Curva de Permanência

Tucci (1993) define curva de permanência com uma função hidrológica que relaciona a vazão ou nível de água de um rio à probabilidade de ocorrerem valores (vazão ou nível) maiores ou iguais ao valor da ordenada.

De acordo com Naghettini e Pinto (2007) a curva de permanência é uma variação do diagrama de frequências acumulada, onde a frequência de não superação é substituída pela porcentagem de um intervalo de tempo específico onde o valor da variável, apresentada na abscissa, foi igualado ou superado (Figura 4.1).

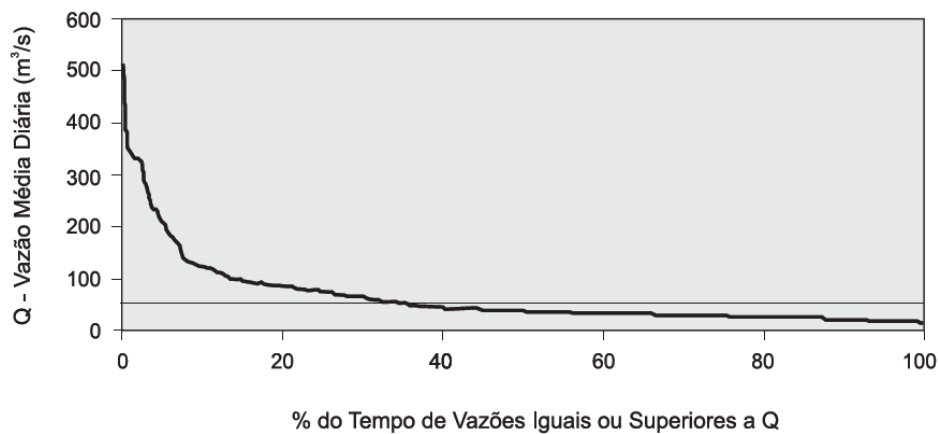


Figura 4.1: Exemplo de Curva de permanência (Rio Paraopeba em Ponte Nova do Paraopeba). Fonte: Naghettini e Pinto (2007)

Existem dois procedimentos normalmente utilizados para obtenção da curva de permanência: empírico e ajuste de função matemática (Tucci, 1993). O procedimento empírico consiste na separação das vazões em intervalos de classe (entre 20 a 30) bem distribuídos, e definição da curva de permanência como um histograma cumulativo de vazões, com base nos intervalos de classe.

Entretanto, com o avanço da tecnologia pode-se utilizar toda a série de dados sem a necessidade de separação por intervalos. A construção da curva de permanência empírica segue quatro passos principais:

- as vazões observadas Q_i são ordenadas em ordem decrescente; para produzir um conjunto de dados indexados por $i = 1, 2, 3, \dots, N$, em que N é o comprimento amostral, e Q_1 e Q_N são, respectivamente, o maior e o menor valores observados;
- a cada valor ordenado Q_i deve ser atribuído sua respectiva ordem de classificação;
- associar a cada vazão ordenada Q_i a sua respectiva frequência de ser igualada ou superada $P(Q \geq Q_i)$ estimada pela razão i/N ;

- traçar um gráfico das vazões ordenadas e suas respectivas porcentagens $100i/N$ de serem igualadas ou superadas, (Naghetini e Pinto, 2007).

A Figura 4.1 apresenta a Curva de Permanência das Vazões do Rio Paraopeba em Ponte Nova do Paraopeba, por meio da qual pode-se observar que uma vazão de $50 \text{ m}^3/\text{s}$ possui uma probabilidade de ser igualada ou superada em 35% do tempo, ou seja, possui uma permanência de 35%.

Segundo Tucci (1993), o ajuste de uma função matemática a uma curva de permanência empírica permite uma sintetização maior de informações, entretanto nem sempre a função matemática retrata a forma da distribuição empírica, principalmente nos seus extremos, e cria uma tendenciosidade nesta região.

4.2.1 Curva de Permanência e Qualidade das Águas

Diversos estudos apresentados na literatura relativos à poluição das águas superficiais comprovam que a vazão está diretamente associada à concentração de poluentes na água (Chesterikoff *et al.*, 1992; Poor e McDonnell, 2007; Roque *et al.*, 2008; Davis *et al.*, 2014; Outram *et al.*, 2014; 2016).

Poor e McDonnell (2007) estudaram a variação da concentração de nitrato em três bacias nos Estados Unidos, com distintas classes de uso do solo (agropastoril, floresta e urbano). Foram selecionados três eventos chuvosos e as maiores concentrações de nitrato foram observadas para maiores vazões. As maiores taxas de exportação de nitrato ocorreram para as bacias com uso do solo urbano (0.131 kg/ha), agrícola (0.121 kg/ha) e floresta (0.012 kg/ha). O uso do solo teve papel preponderante na avaliação das taxas de nitrato, se comparado com os aspectos morfológicos das bacias estudadas.

A bacia com uso por floresta mostrou exportação mínima nos três eventos monitorados e a bacia com uso urbano apresentou alta taxa de exportação para os três eventos. Já a bacia com uso agrícola exibiu exportação elevada no outono, similar à bacia hidrográfica urbana. Entretanto, durante o restante do período chuvoso (inverno e primavera) a exportação de nitrato diminuiu progressivamente na bacia com uso agrícola.

Sliva e Williams (2001) verificaram que em bacias com maior ocorrência de fontes pontuais de poluição, a concentração de poluentes é maior durante o período de menores vazões (seca). A exportação de nutrientes como nitrato e fósforo foi mais acentuada no período chuvoso e associada ao uso do solo por agricultura.

A geração de nutrientes, em diferentes condições de umidade, foi também abordada nos estudos de Outram *et al.* (2014; 2016), por meio da combinação do monitoramento detalhado de nitrato, fósforo total (TP), dados hidrológicos e informações sobre a aplicação de fertilizantes, produzindo uma nova compreensão dos fatores que controlam o transporte de nitrato (N) e fósforo total (TP) para a foz do rio com uso do solo predominante por agricultura. Os

autores identificaram que a precipitação total anual e a infiltração foram importantes para o transporte do nitrato e de fósforo total, sendo os fluxos escoados de N e TP, em um ano chuvoso, o dobro dos registrados em um ano seco.

A medida que se tenta associar a vazão a qualidade das águas, conforme realizado nos estudos de Poor e McDonnell (2007), Sliva e Williams (2001), Outram *et al.* (2014; 2016) é natural que se busque alternativas que associem as duas variáveis, assim o uso de curvas de permanência com esta finalidade passa a ser um alternativa para este processo.

Diferentes aplicações da curva de permanência associando vazão a qualidade de água podem ser encontradas na literatura (Mendiondo, 2008; Henrique, 2009; Brites, 2010; Cunha *et al.*, 2012).

Brites (2010) utilizou curvas de permanências para associar a Demanda Bioquímica de Oxigênio (DBO) a dados de vazão, medidos simultaneamente. Assim para cada intervalo da curva de permanência de vazões foi obtido o valor médio da DBO (Figura 4.2).

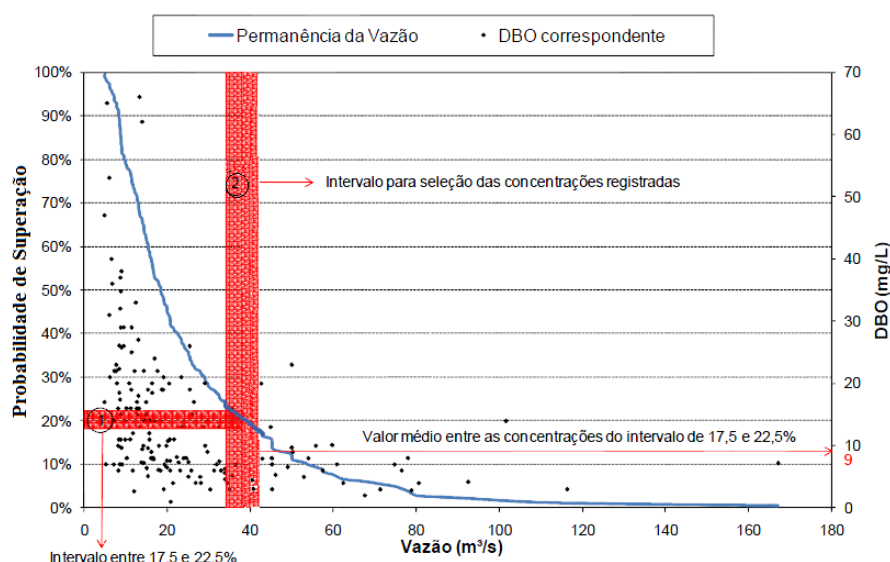


Figura 4.2: Curva de permanência e cálculo da DBO média. Fonte: Adaptado de Brites (2010)

Na Figura 4.2 observa-se que para o intervalo de 17.5 a 22.5 % da curva de permanência o valor médio da DBO é 9 mg/l.

Após obtenção dos valores médios de DBO para cada intervalo da curva de permanência foi possível o ajuste de uma função estatística (Figura 4.3), associando o valor da DBO a permanência.

Cunha *et al.* (2012) integrou dados de vazão a dados de qualidade das águas para o Rio Paraíba do Sul, considerando a carga poluidora monitorada e os valores de referência da Resolução CONAMA nº 357/2005 (Figura 4.4).

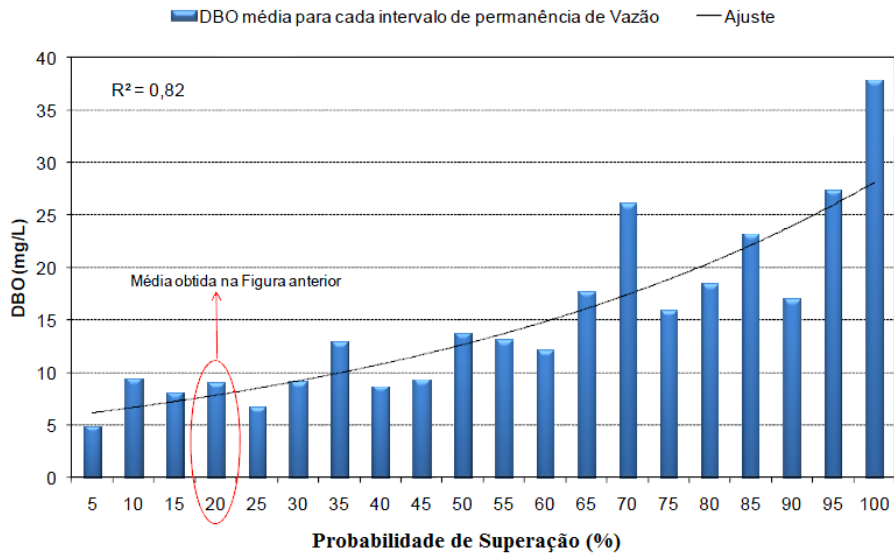


Figura 4.3: Concentração média de DBO para os intervalos da curva de permanência. Fonte: Adaptado de Brites (2010)

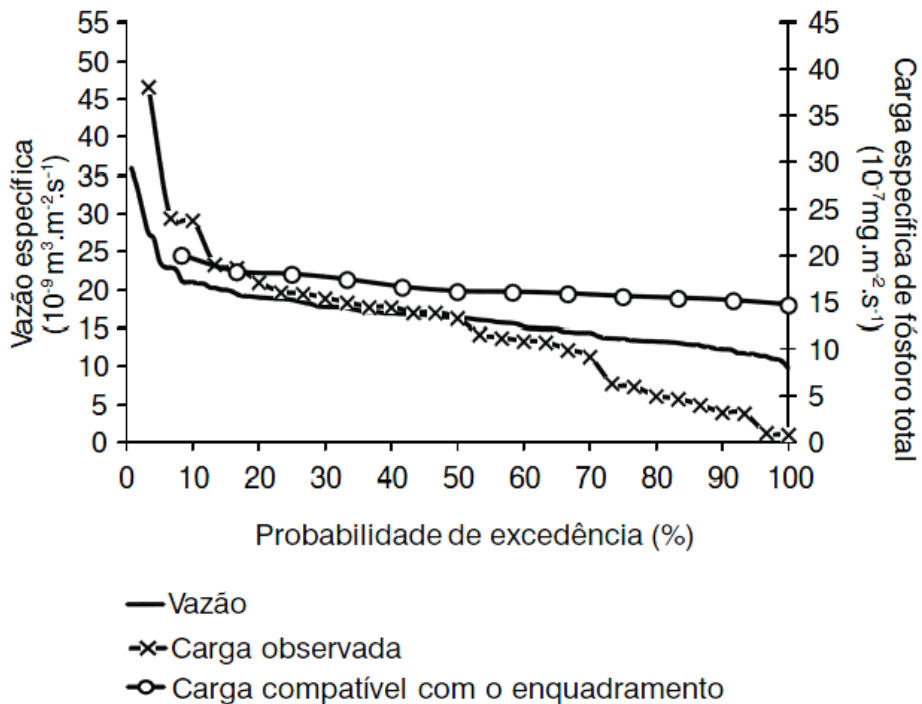


Figura 4.4: Curva de permanência observada e compatível com o enquadramento do Rio Paraíba do Sul. Fonte: Cunha *et al.* (2012)

As cargas específicas observadas foram obtidas pela multiplicação das vazões específicas do sistema aquático pelas respectivas concentrações observadas de Fósforo total na água.

Para obtenção das cargas compatíveis com o enquadramento utilizou-se as vazões características das diferentes estações do ano associadas a uma concentração única, estabelecida para a variável conforme Resolução do CONAMA 357/2005. A Figura 4.4 indica que a carga de fósforo observada excede a carga compatível com o enquadramento que em 15% do tempo.

De acordo com Cleland (2003) a curva de permanência pode ser utilizada para estimar as cargas existentes ou permitidas e para o desenvolvimento de estudos sobre a Carga Total Máxima Diária (CTMD) de parâmetros de qualidade de água. A CTMD especifica a quantidade máxima de poluente que o corpo de água pode receber e também define seu padrão de qualidade da água.

4.3 Árvores de Decisão

O monitoramento da qualidade das águas da bacia do Rio das Velhas, realizado desde de 1997 pelo IGAM, gera um volume considerável de dados, e torna possível definir técnicas que permitam identificar informações relevantes, entender melhor o significado destes dados e estipular procedimentos de tomada de decisão.

O uso da Descoberta do Conhecimento em Banco de Dados, termo derivado do inglês *Knowledge Discovery in Databases* (KDD), é bastante adequado para esta aplicação (Banon, 2013). O KDD pode ser definido como um processo, não trivial, de identificar em dados, padrões que sejam válidos, novos (agreguem conhecimento), potencialmente úteis e compreensíveis. A Mineração de Dados (MD) é um passo fundamental dentro deste processo maior, conhecido como KDD, e compreende a seleção e a aplicação de métodos para identificar regras e padrões de interesse nos dados em análise, sempre buscando o melhor ajuste dos parâmetros no algoritmo escolhido.

Algoritmos de aprendizagem de máquina (AM) são utilizados para resolver problemas em diversas áreas de conhecimento (Schärer *et al.*, 2006; Atkins *et al.*, 2007; Liao e Sun, 2010). Normalmente, aplicados para: (1) mineração de dados, quando se tem um grande volume de dados para os quais se deseja obter relações implícitas que podem ser descobertas automaticamente; (2) em áreas ainda pouco compreendidas pelo ser humano, onde relações podem ser obtidas a partir do estabelecimento de algoritmos (ex. reconhecimento facial por imagens) e (3) domínios onde os programas devem se adaptar a mudanças dinâmicas (ex. qualidade das águas e uso do solo) (Basgalupp, 2010).

Vários algoritmos de aprendizagem de máquina utilizam conceitos de inteligência artificial (AI) e estatística para aprenderem a partir de um conjunto de exemplos de dados. Cada exemplo desse conjunto de dados é constituído por atributos (ou características). Os atributos são as variáveis observáveis e independentes, com domínio contínuo (numérico) ou discreto (categórico). A variável dependente é denominada classe e tem seu valor determinado a

partir dos atributos (Basgalupp, 2010). Os exemplos contidos neste conjunto de dados são denominados dados de treinamento, a partir dos quais pode-se realizar inferências indutivas, que conduzam a hipóteses verdadeiras ou falsas.

O aprendizado indutivo pode ser dividido em supervisionado e não supervisionado. Quando se conhece o valor da classe dos exemplos do conjunto de dados de treinamento, tem-se um processo indutivo supervisionado. Se o domínio do atributo classe é discreto tem um problema de classificação, se é numérico o problema passa a ter solução por regressão ou aproximação de funções (Basgalupp, 2010). O resultado, ou conhecimento adquirido, pode ser apresentado de várias formas: árvores de decisão (AD), redes neurais (RNA), redes bayesianas, entre outros, (Liao *et al.*, 2012).

A obtenção de árvores de decisão para classificação ou regressão, se baseia em particionamento recursivo normalmente binário, uma vez que, o nó pai (a_i) nesta árvore, possui dois nós filhos (M_j) (Figura 4.5) e recursivo por que o processo pode ser repetido por tratamento de cada nó filho como um nó pai. Neste processo, podem ser utilizados dados categóricos ou contínuos, ou uma combinação de ambos, para prever a classe do dado ou o seu valor absoluto. Assim, as regras de divisão são identificadas e otimizadas em cada nível da árvore (Grunwald *et al.*, 2009).

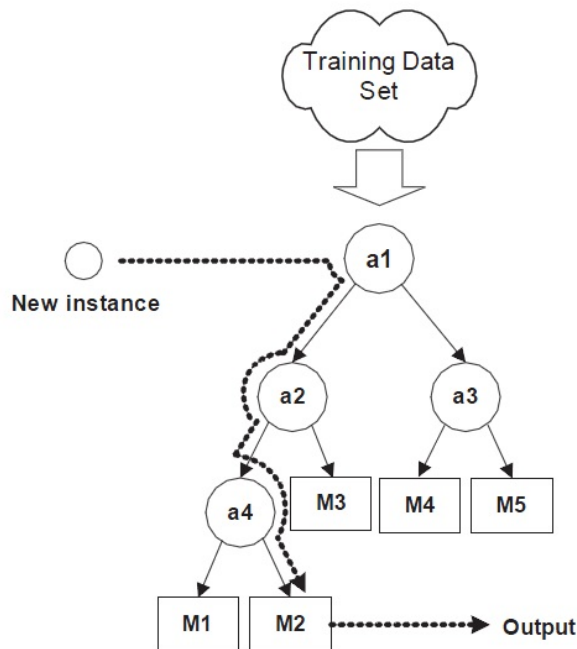


Figura 4.5: Esquema Árvore de Decisão. Fonte:Solomatine e Dulal (2003)

As árvores de decisão constituem uma ferramenta poderosa e amplamente utilizada em problemas de classificação devido, principalmente, ao fato que o conhecimento adquirido é representado por meio de regras, o que facilita o seu entendimento e aplicação por vários

usuários (acadêmicos, privados e público) (Basgalupp, 2010; Liao e Sun, 2010).

Diversas são as maneiras de estruturar uma árvore de decisão a partir de um conjunto de atributos. O número de possíveis árvores de decisão cresce fatorialmente à medida que se aumenta o número de atributos, assim, normalmente, é impraticável definir uma estrutura ótima para um determinado problema, devido ao elevado custo computacional empregado na sua busca.

Neste sentido, algoritmos baseados em heurísticas³ têm sido desenvolvidos para indução de árvores de decisão e vêm apresentando resultados satisfatórios com tempo e recursos computacionais factíveis.

De acordo com Basgalupp (2010), o algoritmo *Top-down Induction of Decision Tree (TDIDT)* é utilizado como base para muitos algoritmos de indução de árvores de decisão: ID3 (Quinlan, 1986), CART (Breiman *et al.*, 1984) e C4.5 (Quinlan, 1993).

De maneira geral, pode-se dizer que os algoritmos do tipo TDIDT são algoritmos recursivos, ou seja, produzem regras de decisão de forma implícita numa árvore de decisão, a qual é construída por sucessivas divisões dos exemplos, de acordo com os valores de seus atributos (variável explicativa), buscando dentro do conjunto de atributos, aquele que melhor divide o conjunto de exemplos em sub-conjuntos (Figura 4.6).

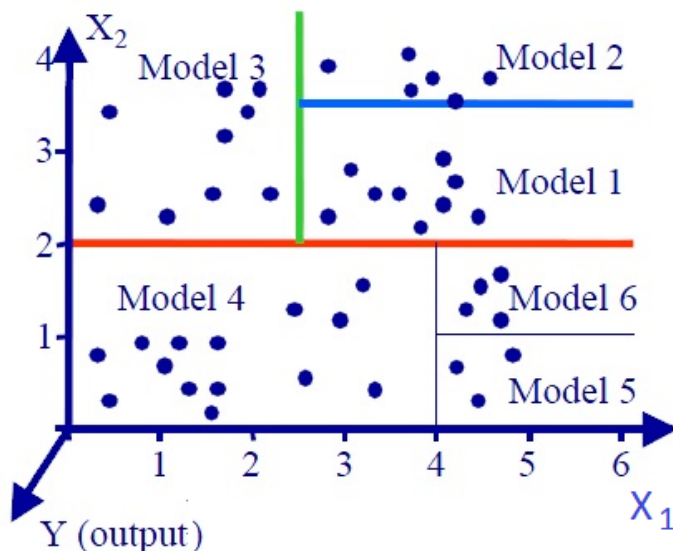


Figura 4.6: Processo de divisão da Árvore de Decisão. Fonte:Solomatine e Dulal (2003)

Este processo se repete até que todos os exemplos estejam classificados, ou até que todos os atributos preditivos tenham sido utilizados.

Os métodos para selecionar o melhor atributo preditivo, que possa realizar a melhor

³Técnica utilizada para resolver problemas de elevado nível de complexidade. Não há garantia de que a solução obtida seja a ótima, Basgalupp (2010).

divisão da árvore de decisão se baseiam em diferentes medidas: impureza, distância e independência. De maneira geral, a maior parte dos algoritmos de indução busca dividir os dados de um nó-pai de forma a minimizar o grau de impureza dos nós-filhos, pois quanto menor o grau de impureza, mais desbalanceada é a distribuição de classes.

Em um determinado nó, se a impureza é nula, todos os exemplos contidos nele irão pertencer a uma mesma classe. Por outro lado, se o grau de impureza é máximo no nó, cada classe terá o mesmo número de exemplos. As medidas mais comumente utilizadas na divisão das árvores de decisão são: ganho de informação, razão de ganho e Gini.

O ganho de informação utiliza a entropia (H) (Equação 4.1) como medida de impureza, comparando o grau de entropia do nó-pai ($H(D)$ - antes da divisão) com a entropia dos nós-filhos ($H(D_X)$ e $H(D_Y)$ - após a divisão). O atributo que gera maior ganho de informação é escolhido como condição de teste. O algoritmo ID3 (Quinlan, 1986) é um dos pioneiros no uso desta medida de divisão.

A entropia mede a quantidade de desordem ou incerteza num sistema. Na classificação se uma partição tem baixa entropia (ou baixa desordem) ela é relativamente pura, isto é, a maioria dos exemplos pertencem a uma mesma classe. Por outro lado, uma partição com maior entropia (ou mais desordem), os exemplos estarão distribuídos em várias classes. A entropia de um conjunto de exemplos D é definida por Zaki e Meira Jr (2014) na Equação 4.1. Se todos os exemplos pertencerem a uma mesma classe, a entropia será zero. Por outro lado, se os exemplos estão distribuídos com igual probabilidade em cada classe, $P(C_i | D) = \frac{1}{k}$, então a entropia terá seu maior valor, $H(D) = \log_2 k$.

$$\text{Entropia : } H(D) = - \sum_{i=1}^k P(C_i | D) \log_2 P(C_i | D) \quad (4.1)$$

Onde $P(C_i | D)$ é a probabilidade da classe C_i em D ($i=1, \dots, k$) e k é o número de classes.

Suponha que o conjunto de exemplos de D seja dividido em dois subconjuntos D_X e D_Y . A entropia desta divisão será obtida da entropia ponderada de cada uma das partições resultantes, dada por $H(D_X, D_Y)$ equação 4.2.

$$\text{Entropia Divisão: } H(D_X, D_Y) = \frac{n_X}{n} H(D_X) + \frac{n_Y}{n} H(D_Y) \quad (4.2)$$

Onde n é o número de exemplos de D , n_X e n_Y são o número de exemplos em D_X e D_Y , respectivamente.

O ganho de informação obtido pela partição de D em D_X e D_Y será obtido pela equação 4.3. Quanto maior o ganho de informação, maior será a redução na entropia, sendo possível considerar a divisão dos exemplos adequada. Assim, dados exemplos para divisão de D e suas partições correspondentes, pode-se escolher o ponto adequado a divisão de forma a maximizar o ganho de informação.

$$\text{Ganho de Informação} : (D, D_X, D_Y) = H(D) - H(D_X, D_Y) \quad (4.3)$$

O máximo ganho de informação é interessante para particionar os exemplos gerando bons resultados. Entretanto, ele tem uma tendência (bias) em favor de testes com muitos valores. Para solucionar esta situação pode-se avaliar a razão de ganho de informação gerada pela partição dos exemplos de D em subconjuntos. A Razão de Ganho (Equação 4.4) nada mais é que o ganho de informação relativo (ponderado).

$$\text{Razão de Ganho} := \frac{(D, D_X, D_Y)}{H(D_Y, D_X)} \quad (4.4)$$

Outro aspecto a ser considerado é a representação do nó, uma vez que influencia o desempenho das árvores de decisão. A representação do nó para o particionamento dos dados pode ser avaliada em função do tipo de atributo categórico (ordinal ou não-ordinal) ou contínuo. A partição mais comum é aquela onde é criada uma aresta para cada valor do atributo usado na condição de teste. O algoritmo C4.5 (Quinlan, 1993) utiliza este tipo de partição para atributos categóricos não-ordinais. Como principal desvantagem deste tipo de partição, tem-se que produz árvores de decisão mais complexas.

Uma partição binária também pode ser utilizada, por exemplo a Solução de Hunt, onde um dos valores é atribuído a uma das arestas e todos os outros valores à outra aresta. Como desvantagem deste tipo de partição tem-se o não aproveitamento de todo o poder de discriminação do atributo em questão. O algoritmo ID3, utiliza este tipo de partição (Quinlan, 1986).

Para atributos categóricos ordinais, é possível realizar uma partição binária, com o valor de um atributo em uma aresta e os demais em outra aresta, como exemplo de uso tem-se o algoritmo CART (Breiman *et al.*, 1984).

Os atributos contínuos permitem maior variedade de testes, implicando em maior complexidade de cálculo. De acordo com Basgalupp (2010), pode-se citar: testes simples ou pesquisa exaustiva, testes múltiplos (segmentação global e segmentação ao nível de nó) e combinação linear de características.

O algoritmo C4.5 (Quinlan, 1993) utiliza o teste simples, sendo a divisão sempre binária, conforme apresentada no exemplo do Apêndice B.

A árvore de decisão consiste em uma seqüência de perguntas que divide o espaço de parâmetros de acordo com a variável resposta. Isso resulta em um conjunto de ramos que subdividem o espaço de parâmetros em hiperespaços disjuntos (folhas) da variável de previsão. Em uma árvore de decisão, os nós se subdividem até que um critério de terminação, ou poda, seja atingido (Pappenberger *et al.*, 2006).

Com relação aos métodos de poda, estes permitem excluir arestas ou sub-árvores que representam um aprendizado muito específico, que não deve ser generalizado. A poda pode

ser classificada em pré-poda ou pós-poda. A pré-poda é realizada durante o processo de construção da árvore, onde o processo pode ser interrompido, parando de dividir o conjunto de elementos, transformando, assim, o nó corrente em um nó folha da árvore.

O ganho de informação pode ser utilizado como critério de poda. Caso todas as divisões possíveis utilizando um atributo “A” produzam ganhos menores que um valor pré-definido, o nó virá folha. Entretanto é difícil estabelecer o valor limite para a poda, pois um valor muito alto gera árvores muito simplificadas e um valor muito baixo pode simplificar pouco uma determinada árvore (Basgalupp, 2010).

O pós-poda é realizado após a construção da árvore de decisão. Para cada nó interno, o algoritmo calcula a taxa de erro caso a árvore seja podada e caso não seja podada. Se a diferença entre os dois valores for menor que um valor pré-definido, a árvore é podada. Este processo se repete progressivamente, gerando um conjunto de árvores podadas. A acurácia do conjunto de dados independentes de cada árvore é calculada e seleciona-se a árvore com maior acurácia.

O tratamento de valores desconhecido na construção de árvores de decisão pode ser realizado de diversas maneiras: substituição por valor (média ou moda); tratamento por exclusão, que não utiliza a informação da classe com valor desconhecido; indução para valor desconhecido, todos os valores desconhecidos são direcionados para uma dada aresta; entre outras.

Um exemplo com a aplicação destes conceitos em dados da tese é apresentado no Apêndice B.

4.4 Árvores de Decisão em Estudos Ambientais

A modelagem baseada em árvore de decisão é aplicada em várias áreas ambientais para descobrir estruturas ocultas em conjuntos de dados complexos e prever as características de uma classe escolhida (variável dependente), a partir de um conjunto de variáveis de predição significativas, (Breiman *et al.*, 1984).

De’Ath (2002) utilizou árvores de decisão para previsões e análises ecológicas e indicou como fator positivo a possibilidade do uso de um número ilimitado de variáveis explicativas numéricas e categóricas, o que também foi obtido na quantificação das relações entre as espécies e propriedades da paisagem ambiental por Moisen e Frescino (2002).

A complexidade da simulação numérica de fluxo de um dado curso de água e seus parâmetros de qualidade gera um aumento na demanda pelo uso de inteligência artificial (AI) na solução desses problemas, uma vez que a técnica pode reduzir o número de variáveis explicativas necessárias (Chau, 2006).

A modelagem da qualidade da água, por meio de modelos hidrodinâmicos, foi realizada por Baird e Whitelaw (1992) e verificou que o comportamento das algas tinha relação com a taxa de respiração das algas e com a temperatura da água. Para simplificar o modelo a

simulação da eutrofização foi então realizada considerando somente a intensidade da luz solar recebida pelo curso de água, haja visto que esta variável influencia a taxa de respiração das algas e a temperatura da água.

A situação apresentada por Baird e Whitelaw (1992) reflete como a inteligência humana reduz o número de parâmetros a ser utilizado na modelagem e como esta manipulação pode aumentar a eficiência do modelo. Desta forma, fixam-se vários parâmetros e varia um ou dois parâmetros. As técnicas de AI são capazes de imitar este comportamento humano e complementar suas deficiências. Segundo Chau (2006), a integração da tecnologia AI e modelos hidrodinâmicos para simulação da qualidade de água possibilita um melhor aproveitamento dos modelos e melhora sua interface com os usuários.

Liao e Sun (2010) avaliaram a qualidade da água do lago Chao na China, integrando redes neurais e árvores de decisão para prever a qualidade de água em um tempo (T+1), considerando a qualidade da água obtida no tempo (T). Foi então utilizado um algoritmo de árvore de decisão melhorado com redes neurais e os resultados obtidos foram considerados satisfatórios com uma acurácia de classificação de 78.75 %, que comparado aos resultados do algoritmo C4.5 atingiu uma perda de 3 % na acurácia, mas reduziu a quantidade de “nós” da árvore de decisão em 50 %.

Solomatine e Dulal (2003) utilizaram árvores de decisão em estudos com modelos chuva-vazão, os resultados obtidos com o uso de árvores de decisão permitiram a construção de vários modelos de baixa complexidade e elevada precisão (99%) para vazão estimada para 1 hora após o início da chuva.

Para modelar a relação entre nível e vazão de um curso de água Bhattacharya e Solomatine (2005), utilizaram um modelo obtido por árvore de decisão. Os resultados indicaram um erro médio quadrado (RMSE) inferior ao obtido por meio do uso de metodologias convencionais (ex. curva-chave).

A relação entre 14 variáveis ambientais (precipitação, área irrigada, tipo de solo, tipos de cultura, entre outros) e a carga de fósforo foi objeto de análise por Grunwald *et al.* (2009) em 4 bacias no sul da Florida. Os resultados obtidos indicaram que o uso de árvore de decisão foi bem sucedido na identificação da relação entre a carga poluidora P e as variáveis de previsão ambiental, obtendo coeficiente de determinação (R^2) entre 0.69 e 0.99. As variáveis ambientais de maior importância foram : demanda para irrigação, percentual de área ocupado por cana de açúcar, localização da bacia, quantidade de chuva, razão chuva/ escoamento superficial, nível de água no canal, percentual de área de inundação, considerando as 4 bacias estudadas.

Árvores de decisão foram também utilizadas para prever a quantidade de fósforo filtrado $P < 0.45\mu\text{m}$ e particulado exportado $P > 0.45\mu\text{m}$ em 11 bacias localizadas na Suíça, com uso do solo predominantemente agrícola. As variáveis explicativas utilizadas foram: índice de fluxo de base (relação entre os fluxos de base e total), chuva anual, área ocupada por agri-

cultura e declividade. O método foi capaz de identificar as bacias hidrográficas com elevada exportação fósforo filtrado e reproduziu o padrão de exportação para os períodos avaliados (seca e chuva). No período de chuva houve maior exportação de fósforo particulado (Schärer *et al.*, 2006).

Capítulo 5

ARTIGO 1 : Aplicação de Árvores de Decisão na Modelagem das Concentrações de Nitrato e Fósforo Total: Estudo de Caso no Rio das Velhas

Lília Maria de Oliveira¹ and Philippe Maillard²

5.1 Resumo

O uso e ocupação do solo da bacia hidrográfica são responsáveis pela perda ou manutenção da qualidade de suas águas superficiais, especialmente quando tratamos de fontes de poluição difusa. A árvore de decisão é uma ferramenta de aprendizagem de máquina que tem se mostrado promissora para este tipo de estudo. O modelo proposto pela técnica segue uma estrutura hierárquica na busca de descobrir estruturas ocultas em conjuntos de dados complexos e prever o comportamento de uma classe (variável dependente) a partir do conjunto de variáveis independentes. A metodologia aplicada neste trabalho avalia o uso desta técnica no estudo da relação vazão - uso do solo - poluição difusa. A Bacia do Rio das Velhas e os parâmetros de qualidade nitrato e fósforo total foram escolhidos como objeto de investigação. Foram utilizados dados de qualidade de 17 estações de monitoramento e de 8 estações fluviométricas de vazão implantadas na calha principal do Rio das Velhas. O uso do solo, sub-dividido em 8 categorias, foi obtido a partir de um mosaico de 72 imagens RapidEye. As árvores de decisão foram geradas utilizando o algoritmo C4.5 de Quinlan que produziu regras para associar a vazão e as categorias de uso do solo aos parâmetros de qualidade, buscando identificar as categorias de uso do solo responsáveis pela alteração da qualidade

¹CEFET-MG, Av. Amazonas, 5354, Belo Horizonte, Brasil Tel.: +55-31-33770361 email: lilia@deii.cefetmg.br

²Universidade Federal de Minas Gerais, Instituto de Geociências, Av. Antônio Carlos, 6627, Belo Horizonte, Brasil; Tel.: +55-31-33770361; email: philippermaillard@yahoo.com.br

das águas superficiais. As árvores de decisão propostas atingiram eficiência de classificação acima de 80% para as duas variáveis avaliadas nitrato e fósforo total. Entretanto, os estudos também apontaram fragilidades na metodologia associadas ao baixo volume de informações existentes para a Bacia do Rio das Velhas.

5.2 Palavras-Chave

Aprendizagem de máquina, algoritmo C4.5, nitrato, fósforo e poluição das águas.

5.3 Abstract

Land use in a watershed can be responsible for the loss or preservation of the quality of its superficial waters. This especially true for non-point water pollution. Decision trees are machine learning techniques with promising applications in this type of research. A decision tree model follows a hierarchical structure that aims at finding hidden patterns within complex databases and predict the behavior of a class (explained variable) from the independent variables. The methodology used in this article evaluates the applicability of decision trees to study the relationship water discharge and land use and nonpoint pollution. The Velhas River watershed and the water quality parameters nitrate and phosphorous were chosen as the objects of investigation. Water quality data from 17 stations and discharge data from 8 fluviometric stations, all within the main river channel of the Velhas River were used to carry out our study. Eight categories of land use data were extracted from mosaic of 72 RapidEye images. The decision trees were generated using Quinlan's C4.5 algorithm which created sets of rules associating land use categories to variations of nitrate and phosphorous therefore identifying which ones are the main responsible for loss of water quality. The resulting decision trees were able to correctly classify over 80% of the instances of discrete nitrate and phosphorous classes. However, some shortcomings were also identified and were mainly attributed to the relatively low volume of water quality data.

5.4 Key-words:

Machine learning, C4.5 algorithm, nitrate, phosphorous and water pollution

5.5 Introdução

A ALTERAÇÃO na qualidade das águas superficiais é denominada poluição e tem como origem fontes poluidoras pontuais e/ou difusas. As fontes pontuais são caracterizadas por despejos que apresentam um único ponto de lançamento no curso de água, por exemplo

efluentes domésticos e industriais. As fontes poluidoras difusas são aquelas distribuídas sobre a superfície da bacia, podendo atingir o curso de água em qualquer ponto tendo o escoamento superficial como veículo de transporte de poluentes. A agricultura e a pecuária são atividades que contribuem com a produção de poluição difusa para os cursos de água.

Com o objetivo de avaliar a relação entre o uso do solo e a qualidade das águas superficiais, muitos estudos já foram realizados (Cleland, 2003; Maillard e Santos, 2008; Gorsevski *et al.*, 2008; Sharifi e Hosseini, 2011; Pinheiro *et al.*, 2014). Nesses, o uso do solo é normalmente apresentado como o percentual de área ocupada por cada categoria de uso e é considerado fonte difusa ou uma forma de controlar a poluição difusa para o curso de água.

A seleção da área de influência do uso do solo, normalmente segue duas linhas distintas:

1. a bacia de contribuição
2. faixas de largura variada ao longo do curso de água

A qualidade das águas por sua vez é avaliada pelos parâmetros físicos, químicos e biológicos obtidos por meio do monitoramento sistemático (Maillard e Santos, 2008; Gorsevski *et al.*, 2008; Sharifi e Hosseini, 2011; IGAM, 2013b).

A avaliação da poluição difusa a partir de modelos matemáticos pode ser realizada por meio do uso de diversos softwares disponíveis e aplicações como demonstrado em Huang e Hong (2010), Sharifi e Hosseini (2011) e Zhenyao *et al.* (2012). A maior dificuldade na utilização de modelos matemáticos se apresenta em função da baixa disponibilidade de dados (Sharifi e Hosseini, 2011).

Uma outra abordagem bastante comum é o uso de regressão múltipla entre qualidade de água, representada pelos parâmetros de monitoramento (ex. turbidez, nitrato, fósforo, DBO, OD, entre outros), e as categorias de uso do solo (Maillard e Santos, 2008; Ouyang *et al.*, 2010; Huang e Hong, 2010; Sharifi e Hosseini, 2011).

Recentemente, tem-se utilizado árvores de decisão para obter regras que associem variáveis independentes (ex. uso do solo) a dependentes (ex. parâmetro de qualidade de água). Apesar do seu uso recente em Recursos Hídricos, a mesma vem apresentando resultados promissores na obtenção de regras associando chuva a vazão, nível de água a vazão e uso do solo a qualidade das águas (Solomatine e Dulal, 2003; Bhattacharya e Solomatine, 2005; Grunwald *et al.*, 2009; Liao e Sun, 2010).

Árvore de decisão é uma técnica de aprendizado de máquina que, a partir de um conjunto de instâncias (dados de treinamento), compostas por dois ou mais atributos (variáveis independentes e dependente), extrai regras sobre um subconjunto das variáveis independentes, usando, geralmente, uma medida de ganho de informação (Chau, 2006; Basgalupp, 2010), para inferir a variável dependente.

A árvore de decisão é construída de forma hierárquica (Figura 5.1), onde seleciona-se, a partir de um conjunto de instâncias, o atributo (a, b, c) com o maior ganho de informação

para dividir esse conjunto de instâncias. A cada divisão seleciona-se novamente um novo atributo para dividi-la até que um critério de poda seja atingido. Ao final deste processo obtém-se regras que associam as variáveis independentes (a, b, c) a variável dependente (M).

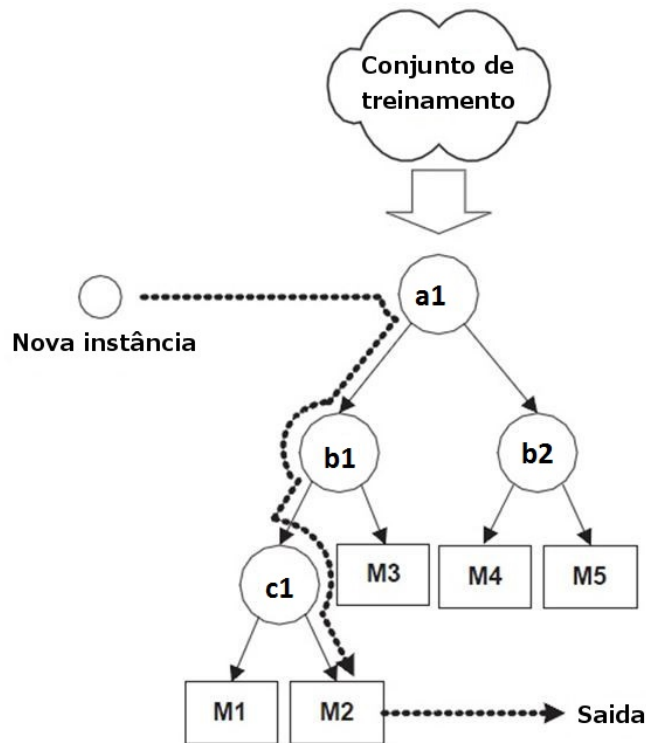


Figura 5.1: Esquema Árvore de Decisão. Fonte:Solomatine e Dulal (2003)

Considerando a importância de entender melhor a relação entre o uso do solo e a qualidade das águas, para fontes difusas, este trabalho tem como objetivo avaliar o potencial de árvores de decisão, por meio do uso do algoritmo C4.5 (Quinlan, 1986), para modelar a relação entre as concentrações de nitrato e fósforo total, vazão e o uso do solo na bacia do Rio das Velhas.

Desta forma, avaliamos através de regras, produzidas por árvores de decisão, a associação entre os parâmetros nitrato e fósforo (variáveis dependentes) e uso do solo e vazão (variáveis independentes). As regras geradas dividiram as instâncias, de forma a relacionar uso do solo e vazão a concentração de nitrato e fósforo. A eficiência de classificação, medida pelo número de instâncias classificadas corretamente, foi superior 80 %.

5.6 Princípios de árvore de decisão

Árvores de decisão utilizam conceitos de inteligência artificial (IA) e de estatística para aprenderem a partir de um conjunto de dados. O conjunto dos dados é constituído de uma

série de subconjuntos, definidos por atributos ou características. Os atributos são as variáveis dependentes e independentes, com domínio contínuo, discreto ou categórico. Os atributos são definidos pelos parâmetros de qualidade de água (dependente), de vazão e de uso do solo (ambos independentes). A variável dependente é geralmente denominada de “classe”, e tem seu valor determinado a partir das variáveis independentes (Basgalupp, 2010).

As instâncias contidas nesse conjunto de dados servem de exemplos de treinamento. A partir destes dados realizam-se inferências indutivas que conduzam a condições verdadeiras ou falsas na forma: se (condição) então (ação).

O aprendizado indutivo pode ser dividido em supervisionado e não supervisionado. Quando se conhece o valor da classe, pertencente ao conjunto de exemplos de treinamento, o processo é chamado de indutivo supervisionado. Se o domínio da classe é discreto o problema é de classificação e se é numérico o problema passa a ter solução por regressão ou aproximação de funções.

A obtenção de árvores de decisão por classificação ou regressão, se baseia em particionamento recursivo binário, uma vez que o “nó” pai (a_i) é dividido em dois “nós” filhos (b_i) (Figura 5.1) recursivamente, por que o processo pode ser repetido por tratamento de cada “nó” filho como um “nó” pai, até que se atinjam as folhas, que indicam as classes (M_j) (Grunwald *et al.*, 2009).

Um mesmo conjunto de atributos pode gerar diversas estruturas de árvore de decisão. O número de árvores de decisão possíveis cresce fatorialmente à medida que se aumenta o número de atributos. Assim normalmente é impraticável definir uma estrutura ótima para um determinado problema, devido ao elevado custo computacional empregado na sua busca.

Nesse sentido, algoritmos baseados em heurística, que é uma técnica utilizada para resolver problemas de elevado nível de complexidade, sem que haja garantia de que a solução obtida seja a ótima, têm sido desenvolvidos para indução de árvores de decisão, apresentando resultados promissores em relação ao tempo e aos recursos computacionais necessários.

De acordo com Basgalupp (2010), o algoritmo Top-down induction of decision tree (TDIDT) é utilizado como base para muitos algoritmos de indução de árvores de decisão, por exemplo, ID3 (Quinlan, 1986), CART (Breiman *et al.*, 1984) e C4.5 (Quinlan, 1993). De maneira geral, os algoritmos do tipo TDIDT são algoritmos recursivos, ou seja, produzem regras de decisão de forma implícita, em uma árvore de decisão, construída por sucessivas divisões das instâncias de treinamento de acordo com os valores de suas variáveis independentes (X_1 e X_2 na Figura 5.2). Assim, busca-se, dentro do conjunto de atributos aquele que melhor divide as instâncias de treinamento em subconjuntos.

Este processo se repete até que todos os exemplos de treinamento da classe (variável dependente) estejam classificados, ou até que todos os atributos preditivos tenham sido utilizados.

Os métodos usados para selecionar o melhor atributo preditivo para divisão da árvore

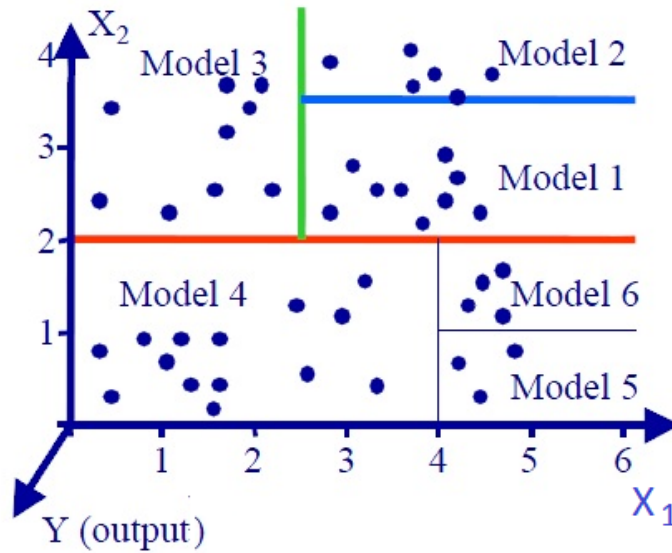


Figura 5.2: Processo de Divisão de Árvore de Decisão. Fonte:(Solomatine e Dulal, 2003)

de decisão se baseiam em medidas de impureza, de distância e de independência. Assim, a maior parte dos algoritmos de indução busca dividir os exemplos de um “nó-pai”, de forma a minimizar o grau de impureza dos “nós-filhos”, pois, quanto menor o grau de impureza, mais desbalanceada é a distribuição de classes. Em um determinado “nó”, se a impureza é nula, todas as instâncias de treinamento contidas nele irão pertencer a uma mesma classe. Por outro lado, se o grau de impureza é máximo no “nó”, cada classe terá o mesmo número de instâncias de treinamento.

Entre as medidas mais comumente utilizadas na divisão das árvores de decisão estão o ganho de informação (Equação 5.1) e a razão de ganho que é obtida pela divisão do ganho pela entropia (Equação 5.2 e 5.3).

$$\text{Ganho de Informação : } \textit{ganho} = \textit{entropia}(\textit{pai}) - \sum_{j=1}^n \frac{N(v_j)}{N} * \textit{entropia}(v_j) \quad (5.1)$$

Onde

$$\text{Entropia : } \textit{entropia}(\textit{nó}) = - \sum_{i=1}^c p(i/\textit{nó}) * \log_2[p(i/\textit{nó})] \quad (5.2)$$

$$\text{Razão de Ganho : } \textit{razão de ganho} = \frac{\textit{ganho}}{\textit{entropia}(\textit{nó})} \quad (5.3)$$

onde: n é o número de valores do atributo, ou seja, número de nós-filhos, N é o número total de objetos do nó-pai, N(vj) é o número total de exemplos associados ao nó-filho(vj),

$p(i/nó)$ é a fração de instâncias de treinamento pertencentes à classe i no nó e c é o número de classes.

O ganho de informação utiliza a entropia como medida de impureza, comparando o seu valor para o “nó-pai” (antes da divisão) com a entropia dos “nós-filhos” (após a divisão). O atributo que gerar o maior ganho é escolhido como condição de teste.

A entropia está relacionada com a distribuição dos valores de uma variável, ou seja, um valor elevado quer dizer que há uma distribuição mais uniforme, ao contrário, uma entropia pequena quer dizer que os valores estão concentrados em um ou vários conjuntos (“clusters”).

Assim, pode-se dizer que a árvore de decisão consiste em uma sequência de condições que divide o espaço dos dados de acordo com a variável dependente. Isso resulta em um conjunto de ramos que subdividem o espaço dos dados em hiperespaços disjuntos (folhas) da variável a ser prevista. Em uma árvore de decisão, os “nós” se subdividem até que um critério de terminação, ou poda, seja atingido (Pappenberger *et al.*, 2006).

A poda permite excluir arestas ou sub árvores que representam um aprendizado muito específico e que não deveria ser aplicado em todas as situações. A poda pode ser classificada em pré-poda ou pós-poda.

A pré-poda é realizada durante o processo de construção da árvore, quando o processo pode ser interrompido, parando de dividir o conjunto de instâncias, transformando, assim, o “nó” corrente em uma folha da árvore. O ganho de informação pode ser utilizado como critério de poda. Caso todas as divisões possíveis utilizando-se um atributo “A” produzam ganhos menores que um limiar predefinido, o “nó” vira folha (Basgalupp, 2010).

A pós-poda é realizada após a construção da árvore de decisão. Para cada “nó” interno, o algoritmo calcula a taxa de erro caso a árvore seja podada ou não. Se a diferença entre os dois valores for menor que o limiar, a árvore é podada. Este processo se repete progressivamente, gerando uma árvore podada.

5.7 O algoritmo C4.5 (J48)

Uma implementação em Java deste algoritmo pode ser encontrada na ferramenta de domínio público WEKA - Software Waikato Environment for Knowledge Analysis (Witten e Frank, 2005), com o nome de J48.

A Tabela 5.1 lista os parâmetros utilizados para a construção da árvore de decisão do algoritmo J48 (WEKA).

O parâmetro `binarySplits (B)` serve para determinar se as classes são nominais (verdadeiro) ou numéricas, onde utiliza um operador de comparação do tipo “<” ou “>” (falso).

`ConfidenceFactor (C)` é o parâmetro de fator de confiança que tem o objetivo de analisar a precisão das regras geradas, pois uma redução no valor padrão (0.25) implica um aumento

Tabela 5.1: Parâmetros do algoritmo J48. Adaptado de (Banon, 2013)

J48			
Parâmetro	ID	Significado	Padrão
binarySplits	B	Recurso para utilizar a árvore binária	Falso
confidenceFactor	C	Define o fator que será usado para a poda	0.25
minNumObj	M	Define o número mínimo de amostras por folha	2
numFolds	N	Define o número de partições usadas para poda com redução de erros. Apenas uma partição é usada para poda	3
reducedErrorPruning	R	Recurso para realizar a poda com redução de erros	Falso
subtreeRaising	S	Recurso para utilizar a subárvore durante a poda	Verdadeiro
unpruned	U	Recurso para desabilitar o recurso de poda	Falso
useLaplace	L	Recurso para suavização do número de folhas baseando-se em Laplace	Falso

da poda da árvore.

Os parâmetros `minNumObj` e `numFolds` estão associados a recursos de poda da árvore de decisão. Cabe destacar que o `numFolds` com valor padrão (3) implica a separação de uma parte dos dados para poda e duas partes para construção da árvore. Aplicações deste algoritmo envolvendo recursos hídricos podem ser encontradas em Solomatine e Dulal (2003) Bhattacharya e Solomatine (2005), Schärer *et al.* (2006), Grunwald *et al.* (2009), Liao e Sun (2010), Cheng *et al.* (2012) e Ali *et al.* (2013).

5.8 Materiais e Métodos

5.8.1 Área de estudo

A Bacia do Rio das Velhas possui uma área de 27.887 km^2 e está localizada na região central do estado de Minas Gerais (Figura 5.3), onde estão inseridos 51 municípios, que abrigam uma população de, aproximadamente, 4,8 milhões de habitantes (destes, aproximadamente 89 % residem em distritos e municípios integralmente inseridos na bacia), sendo que 44 destes estão totalmente inseridos na bacia (CBH, 2015).

O curso principal do Rio das Velhas percorre uma extensão de 802 km, com nascente no município de Ouro Preto, a uma altitude de, aproximadamente, 1.500 m, e foz na Barra do Guaicuí, município de Várzea da Palma, a uma altitude de 478 m, onde é afluente do Rio São Francisco.

Entre os projetos existentes na Bacia do Rio das Velhas é importante mencionar os projetos Manuelzão e Águas de Minas (Instituto Mineiro de Gestão das Águas - IGAM) ambos iniciados em 1997 (Manuelzão, 2014). O biomonitoramento realizado pelo Projeto Manuelzão e pela Universidade Federal de Minas Gerais (UFMG) constatou que os peixes subiam apenas 200 km, a partir da foz do Rio das Velhas em 2000. Em 2010, foi verificado que os peixes já avançavam 580 km no rio, chegando bem próximo às áreas consideradas mais degradadas na Região Metropolitana de Belo Horizonte (RMBH). O principal fator responsável pelos resultados obtidos foi o volume de esgoto tratado pela Companhia de Saneamento de Minas Gerais (COPASA) na Bacia do Rio das Velhas, que passou de 41 milhões de m^3 em 2003 para 85 milhões de m^3 em 2008, atingindo 127 milhões de m^3 de esgoto tratado em

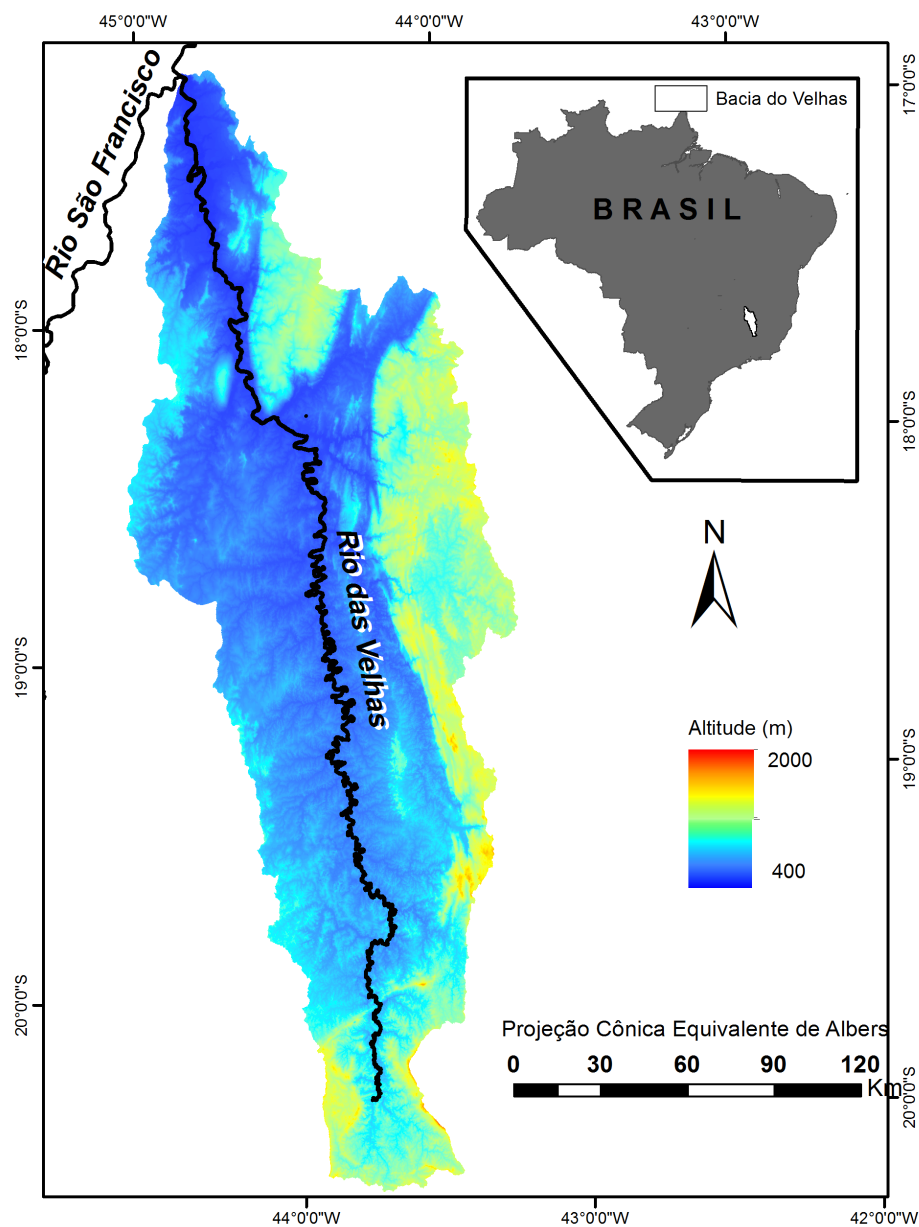


Figura 5.3: Mapa de localização da área de estudo

2010 (Manuelzão, 2014).

O monitoramento da qualidade das águas, por meio da obtenção dos parâmetros físicos, químicos e biológicos na Bacia do Rio das Velhas, é conduzido pelo Projeto Águas de Minas - IGAM desde 1997. Atualmente, existem 82 estações de qualidade das águas na Bacia do Rio das Velhas, distribuídas em diferentes projetos de Avaliação da Qualidade das águas Superficiais: Projeto Pampulha, Projeto Alto Velhas, Meta 2014 (IGAM, 2013b). Em 2013, as principais fontes poluidoras, indicadas pelo IGAM como responsáveis pela degradação da qualidade das águas, foram: efluentes sanitários e industriais, pecuária e mineração; esses dois últimos associados à poluição difusa.

5.8.2 Dados de qualidade das águas e vazão da Bacia do Rio das Velhas

Os parâmetros de qualidade Nitrato e Fósforo total foram obtidos mensalmente para o período de 2009 a 2011 em 17 estações monitoradas (Tabela 5.2) pelo IGAM e localizadas na calha principal do Rio das Velhas.

As informações relacionadas às vazões monitoradas nas datas de medição de qualidade foram obtidas para 8 estações, sendo 7 pertencentes à Agência Nacional de Águas (ANA) e 1 à Companhia Energética de Minas Gerais (CEMIG), também localizadas na calha principal do Rio das Velhas.

A Tabela 5.2 apresenta para cada estação de qualidade a correspondente estação de vazão utilizada no estudo.

A espacialização das estações, ao longo da calha do Rio das Velhas, é apresentada na Figura 5.4.

Tabela 5.2: Estações de monitoramento de qualidade das águas e de vazão utilizadas

Estações de Qualidade	Estações de Vazão	
BV013	41152000	¹
BV037	41152000	
BV063	41199998	
BV067	41199998	
BV083	41199998	
BV105	41260000	
BV153	41260000	
BV137	41340000	
BV141	4160000	¹ Estação CEMIG
BV142	41818000	
BV146	41818000	
BV150	41818000	
BV152	41818000	
BV148	41990000	
BV149	41990000	
BV151	41990000	
BV156	41410000	

Uso do solo da bacia do Rio das Velhas

O mapeamento do uso do solo foi realizado pelo laboratório de Sensoriamento Remoto do Instituto de Geociências da UFMG em parceria com o IGAM. As imagens RapidEye do ano 2010 foram obtidas junto ao Instituto Estadual de Florestas (IEF).

Este tipo de imagem é adquirida de uma constelação de 5 satélites em 5 bandas espectrais, do azul ao infravermelho próximo, e resolução de 5 metros. A partir de 2011 estas imagens passaram a ser compradas continuamente para todo o território nacional pelo Ministério

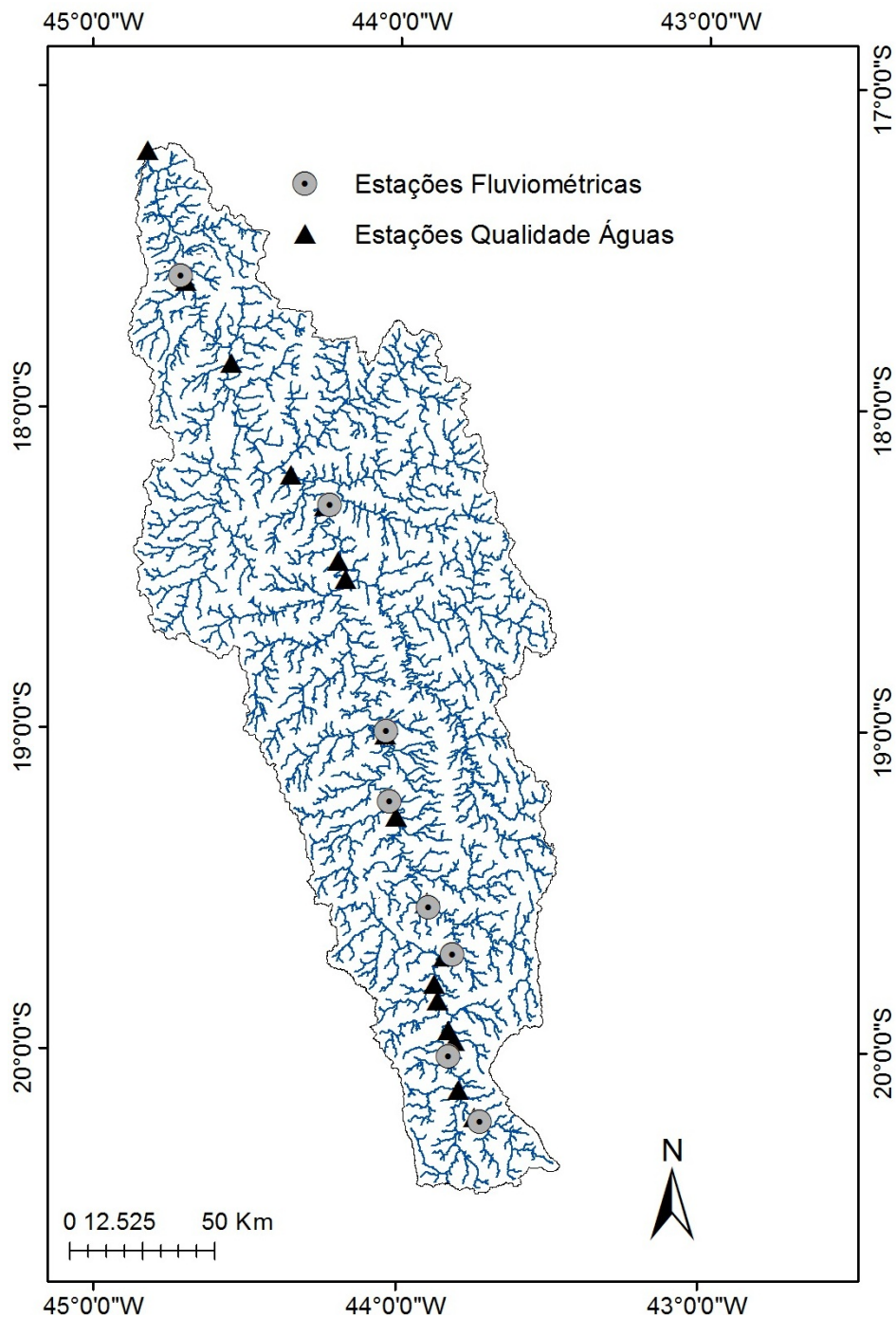


Figura 5.4: Bacia do Rio das Velhas com estações fluviométricas e de qualidade de águas

do Meio Ambiente (MMA) e disponibilizadas para instituições públicas, universidades e institutos tecnológicos (Ministério do Meio Ambiente, 2008).

O processo de classificação de imagens foi realizado com o software Trimble \mathcal{R} eCognition Developer 8.4 com utilização de regras baseadas, principalmente, na resposta espectral de cada alvo. As categorias mapeadas são apresentadas na Tabela 5.3 e Figura 5.5.

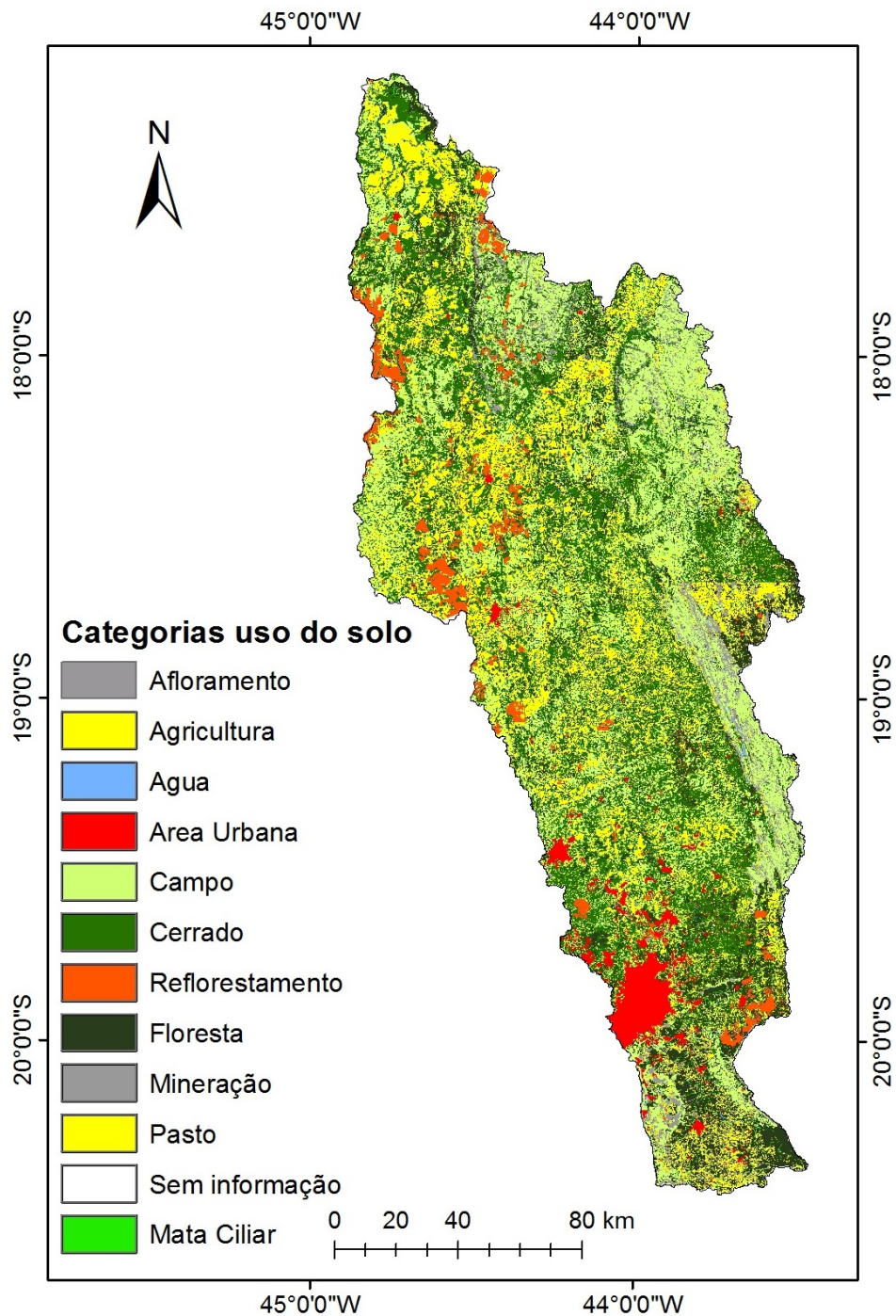


Figura 5.5: Tipos de Uso do Solo da Bacia do Rio das Velhas

As categorias *pasto e agricultura* foram agrupadas gerando a categoria *Agropastoril*. O agrupamento foi realizado devido a dificuldade de diferenciar as duas categorias numa área tão grande como a da Bacia do Rio das Velhas, não sendo possível diferenciar a grande variedade de tipos de culturas, de forma a separa-la da categoria *pasto*, dificuldade acentuada considerando que as culturas se encontram em várias fases de crescimento e quase sem

Tabela 5.3: Uso do solo da bacia do Rio das Velhas em 2010 (as linhas em negrito mostram as categorias agrupadas ou não consideradas)

Categorias	(%)	Categorias Finais	(%)
Cerrado (CE)	26.4	Cerrado (CE)	26.4
Campo (CA)	21.5	Campo (CA)	21.5
Pasto (P)	25	Agropastoril (AP)	25.2
Agricultura (A)	0.2		
Mata (M)	16.1	Mata (M)	16.1
Urbano (AU)	2.9	Urbano (AU)	2.9
Mineração (MI)	0.2	Mineral (MI)	4.1
Afloramento (AF)	3.9		
Reflorestamento (RE)	3.3	Reflorestamento (RE)	3.3
Água AG)	0.3	Não consideradas	0.5
Sem Informação	0.2		

biomassa verde logo após a safra. Finalmente muitas plantações não são espectralmente diferenciáveis com apenas cinco bandas espectrais dos dados RapidEye.

Adicionalmente, como a categoria *mineração* ocupava apenas 0.2 % da área, ela foi mesclada com *Afloramento* dando origem à categoria Mineral. As categorias *Água e Sem informação* não foram utilizadas nas análises por não terem contribuição previsível como fonte difusa.

5.8.3 Período de análise e preparação da base de dados

As fontes de poluição difusa dependem da existência do escoamento superficial para atingir o curso de água, conseqüentemente a relação “uso do solo - poluição difusa - vazão” é mais acentuada nos meses chuvosos.

Realizamos as análises considerando um período hidrológico com vazões maiores que a Q50%. Dessa forma, os dois trimestres mais chuvosos foram selecionados para análise: dezembro-fevereiro e novembro-janeiro dos anos hidrológicos 2009-2010 e 2010-2011. As normais climatológicas de Belo Horizonte (curso médio) e de Pirapora (fz) serviram de base na escolha desses trimestres e foram obtidas do Instituto Nacional de Meteorologia Instituto Nacional de Meteorologia (2004) e são apresentados na (Figura 5.6). Os anos foram selecionados considerando-se a representatividade do uso e da ocupação do solo obtido para o ano de 2010.

As 17 estações de qualidade selecionadas possuem área de drenagem variando de 549.9 km^2 (BV013) a 27853 km^2 (BV149), próximo à nascente e à foz respectivamente. Essas estações apresentam valores de Nitrato e Fósforo Total bastante variáveis em função das fontes poluidoras e da dimensão da sua área de contribuição. O mesmo ocorre com a vazão, onde as estações de monitoramento fluviométrico possuem área de drenagem variando de

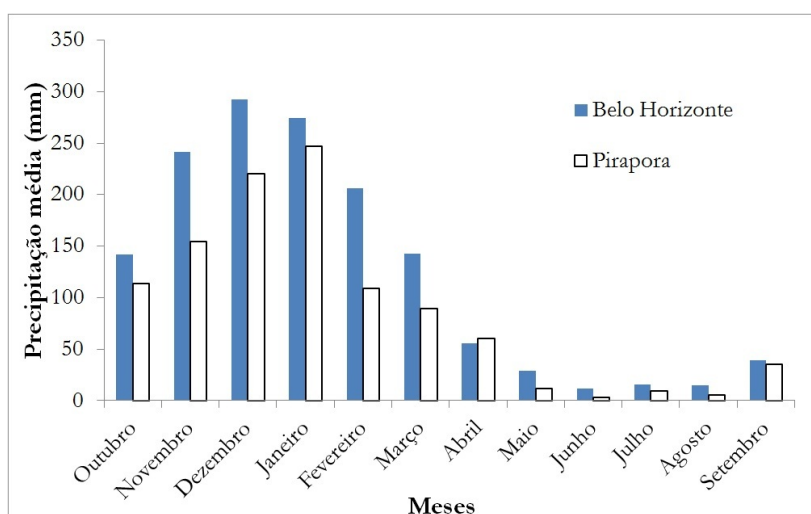


Figura 5.6: Normais Climatológicas 1961-1990 para Belo Horizonte e Pirapora. Fonte: INMET (2014).

490.3 km^2 (41152000) a 26019.4 km^2 (41990000).

A variação das áreas de drenagem e da proporção de usos em cada área impossibilita o adequado relacionamento das variáveis vazão, nitrato e fósforo total se forem utilizados seus valores absolutos. Por este motivo, realizamos a normalização destas informações, substituindo as mesmas pela sua permanência no tempo. As curvas de permanência de vazão, nitrato e fósforo total foram traçadas para cada uma das estações avaliadas.

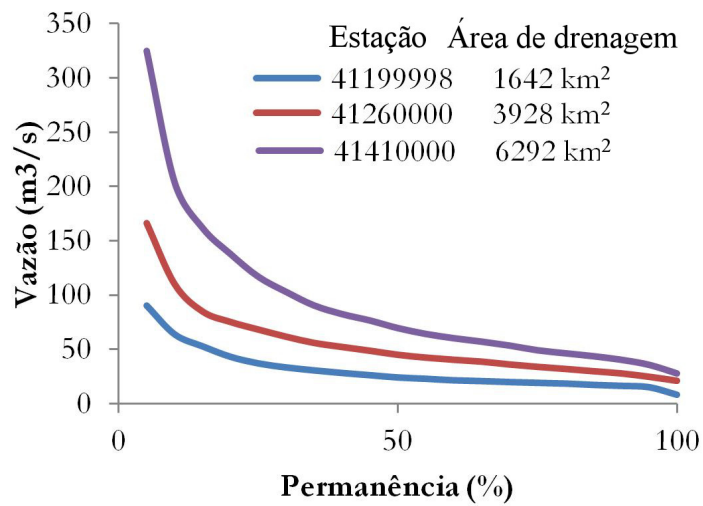
A análise da permanência da vazão no tempo é uma abordagem bastante consolidada em recursos hídricos. Por outro lado, o uso desta mesma abordagem para parâmetros de qualidade é mais recente. A aplicação da permanência de parâmetros de qualidade de águas superficiais podem ser encontradas em Cleland (2003), Brites (2010), Formigoni *et al.* (2011) e Cunha *et al.* (2012).

Como o uso do solo e a permanência da vazão foram utilizados como variáveis independentes e a permanência do nitrato e do fósforo total como variáveis dependentes, selecionamos um período comum para análise destas variáveis. O uso do solo do ano de 2010 foi tido como o centro deste período de análise, com dois anos a mais e a menos, como representativos deste uso do solo.

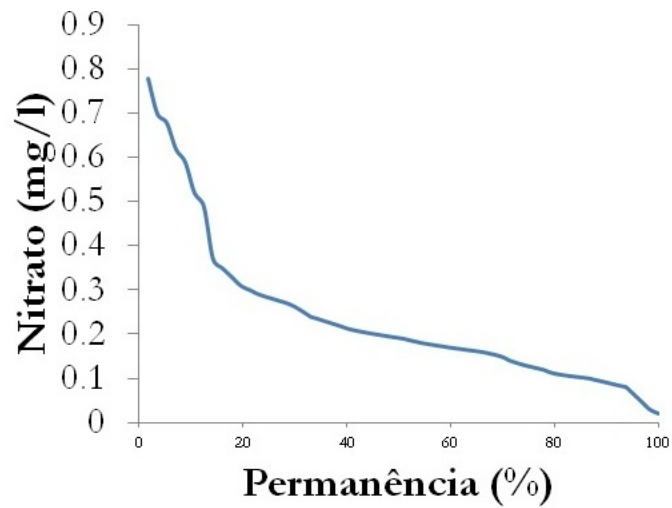
Os dados de qualidade e vazão de 2008 a 2012 foram selecionados para gerar as curvas de permanência (Figura 5.7).

Considerando que fontes poluidoras difusas tendem a ter o seu comportamento atrelado ao escoamento superficial, ou seja, em bacias com um número elevado de fontes difusas presentes, um aumento no volume escoado (vazão) é acompanhado da perda de qualidade das águas se a poluição difusa for significativa.

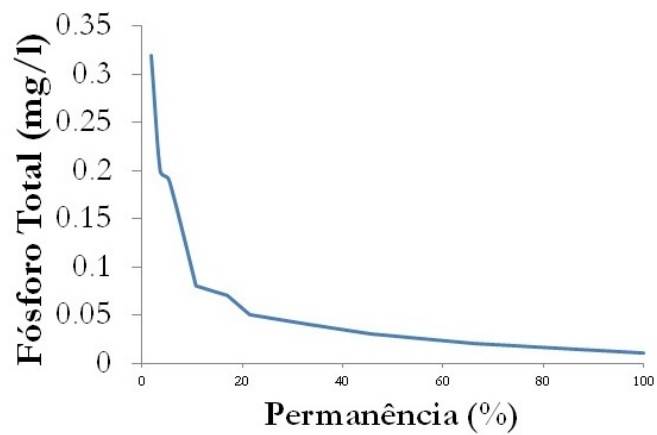
Para avaliar este efeito, neste trabalho, é proposto um índice (I_{Q-QA}) que relaciona a



(a)



(b)



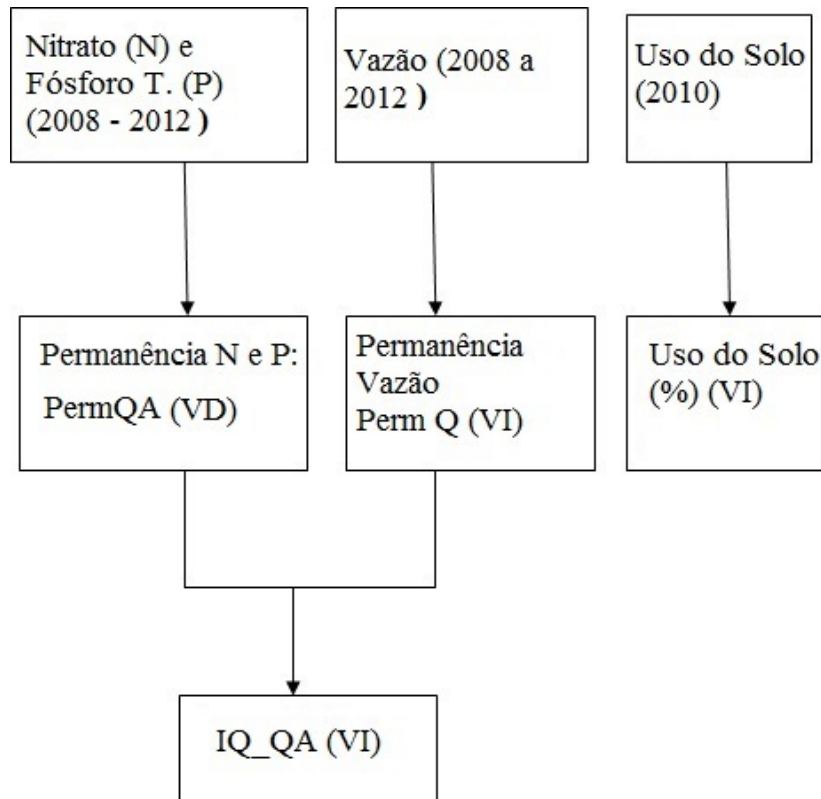
(c)

Figura 5.7: (a) Permanência de vazão das estações 41199998, 41260000 e 41410000, (b) Permanência Nitrito - BV013 e (c) Permanência Fósforo Total - BV013

permanência da vazão à permanência da qualidade (Equação 5.4).

$$I_{Q-QA} = \frac{PermQ}{PermQA} \quad (5.4)$$

onde: PermQ é a permanência da vazão (%) e PermQA é a permanência do parâmetro de qualidade (%) e I_{Q-QA} é a relação obtida pela Equação 5.4. As etapas de transformação da base de dados são apresentadas na Figura 5.8.



VD- Variável Dependente VI- Variável Independentes

Figura 5.8: Etapas de geração das bases de qualidade, vazão e uso do solo

O uso e a ocupação do solo foram avaliados em termos percentuais para as bacias de drenagem dos 17 pontos de qualidade estudados (Figura 5.8).

5.8.4 Árvore de decisão

As duas variáveis dependentes avaliadas foram permanência do nitrato e do fósforo total, enquanto as variáveis independentes foram: permanência da vazão (PermQ), I_{Q-QA} (Nitrato: I_{Q-N} e Fósforo Total: I_{Q-P}) e 8 categorias de uso do solo (Tabela 5.3). O algoritmo C4.5 (J48) trabalha com a variável dependente (classe) discretizada, sendo o número

de classes sugerido no WEKA igual a 10, com intervalos iguais. Como a entropia pode variar em função da discretização escolhida adotamos 6 discretizações dos dados em 2, 3, 4, 5, 7 e 10 intervalos, buscando avaliar as diferenças encontradas pelo algoritmo com cada uma delas. As variáveis independentes não necessitam ser discretizadas, sendo que o uso solo foi transformado em valores percentuais para as bacias dos pontos de monitoramento de qualidade das águas. A vazão foi avaliada pela sua permanência no tempo e o I_{Q-QA} , conforme definido na Equação 5.4. Os parâmetros do algoritmo C4.5 (J48), mostrados na Tabela 5.1, foram utilizados com seus valores padrões.

5.8.5 Validação cruzada

Os algoritmos de indução de Árvores de Decisão constroem geralmente estruturas com mais ramificações que o necessário. O processo de “poda” da árvore busca reduzir os “nós” que trazem pouco ganho de informação.

No C4.5 a poda é realizada após a construção da árvore e determinação da taxa de erro da classificação produzida pela árvore. Se houver dados suficiente, uma parte é utilizada para construção da árvore e outra é reservada para testar e calcular a taxa de erro após a construção da árvore.

Caso os dados sejam escassos, pode-se utilizar um esquema de Validação Cruzada. Neste caso, os dados são divididos em n blocos de dimensão semelhante. A aprendizagem faz-se com recurso a ni iterações, em que a cada iteração são utilizados $n-1$ blocos para aprendizagem (treinamento) e o bloco que sobrou para teste, sendo este diferente a cada iteração. A técnica de validação cruzada foi adotada em função do número reduzido de instâncias.

Neste trabalho, foram utilizados 95 instâncias para o trimestre Dez-Fev e 104 instâncias para Nov-Jan. A diferença no número de instâncias se deu em função da inexistência de dados em algumas estações em determinadas datas.

5.9 Resultados

5.9.1 Análise das classificações

As Tabelas 5.4 e 5.5 apresentam os resultados das seis discretizações testadas através do percentual de instâncias classificadas corretamente e incorretamente para os trimestres Nov-Jan de 2009 a 2011 (Tabela 5.4) e Dez-Fev dos mesmos três anos (Tabela 5.5).

Os melhores resultados da classificação do trimestre Nov-Jan foram obtidos para a discretização do nitrato e do fósforo total em 3 classes, obtendo-se 83% e 82% das amostras classificadas corretamente, respectivamente, para o período de Nov-Jan 2009-2011.

Os resultados do trimestre Dez-Fev de 2009-2011 (Tabela 5.5) apresentaram um número inferior de amostras classificadas corretamente em relação a Nov-Jan (Tabela 5.4). A melhor

Tabela 5.4: Instâncias Classificadas corretamente e incorretamente em função da discretização utilizada nos dados de qualidade. Nov-Jan 2009-2011.

Discretização	Nitrato		Fósforo Total	
	Inst. Cor.(%)	Inst. Inc.(%)	Inst. Cor.(%)	Inst. Inc.(%)
2	67	33	69	31
3	83	17	82	18
4	49	51	72	28
5	53	47	67	33
7	38	62	51	49
10	24	76	43	57

Tabela 5.5: Instâncias Classificadas corretamente e incorretamente em função da discretização utilizada nos dados de qualidade. Dez-Fev 2009-2011.

Discretização	Nitrato		Fósforo Total	
	Inst. Cor.(%)	Inst. Inc.(%)	Inst. Cor.(%)	Inst. Inc.(%)
2	48	52	65	35
3	59	41	80	20
4	62	38	66	34
5	57	43	57	43
7	47	53	46	54
10	27	73	42	58

árvore de decisão para o nitrato foi obtida com 4 discretizações (Tabela 5.5), classificando corretamente 62% das instâncias. Para o fósforo total o intervalo de 3 discretizações também obteve melhor resultado com 80% das instâncias classificadas corretamente (Tabela 5.5).

A variação no número de instâncias classificadas corretamente e incorretamente para as 6 discretizações da variável dependente (permanência de nitrato e fósforo total) é menor para o parâmetro fósforo que para o nitrato (Figuras 5.9 e 5.10), considerando os dois trimestres avaliados. As Figuras 5.9 e 5.10 apresentam os resultados para o período de Nov-Jan.

Uma vez que as árvores geradas (Nitrato e Fósforo total) no período de Nov-Jan de 2009-2011 (Figura 5.11 a e b) tiveram melhor resultado de classificação as demais análises apresentadas se referem a este período.

Considerando a Figura 5.11, verifica-se que a separação dos dados ocorreu inicialmente através do uso do índice I_{Q-QA} (equação 4) representado pelos símbolos I_{Q-N} (nitrato) e I_{Q-P} (fósforo total).

O segundo parâmetro utilizado foi o uso do solo, representado pelas categorias: AP, AU, CA e MI para o nitrato e pelas categorias: AU e AP para o fósforo total, Figuras 5.11a e b, respectivamente.

No modelo do nitrato (Figura 5.11a) para $I_{Q-N} > 0.72$, o algoritmo não conseguiu encontrar categorias de uso do solo e de permanência da vazão (PermQ), que pudessem explicar a permanência do nitrato no intervalo -INF a 40.48%. Os dados desse intervalo

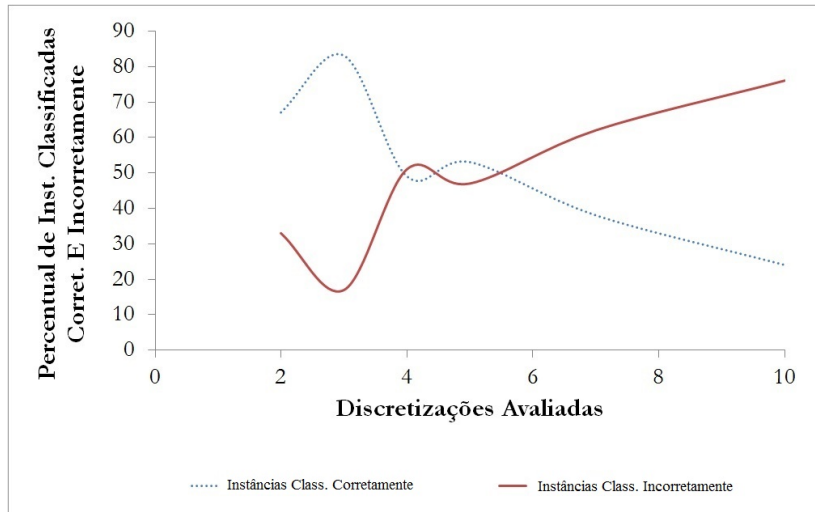


Figura 5.9: Percentual de instâncias classificada correta e incorretamente para Nitrato (nov-Jan/2009-2011)

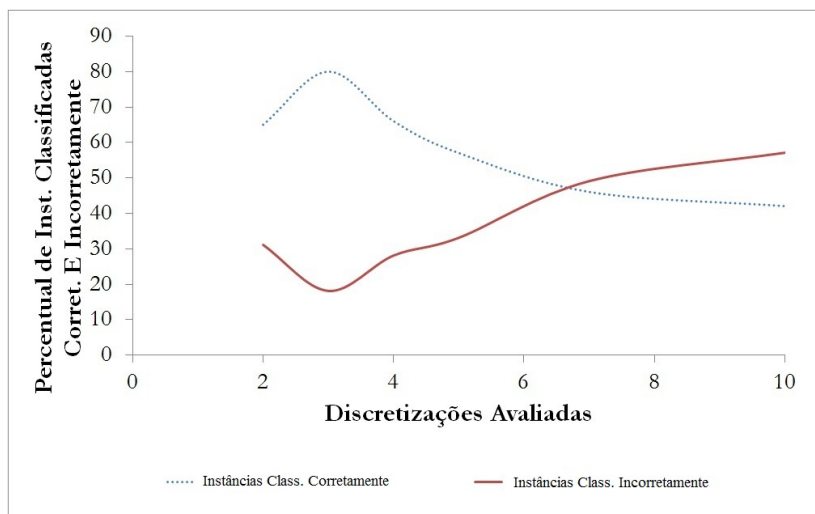


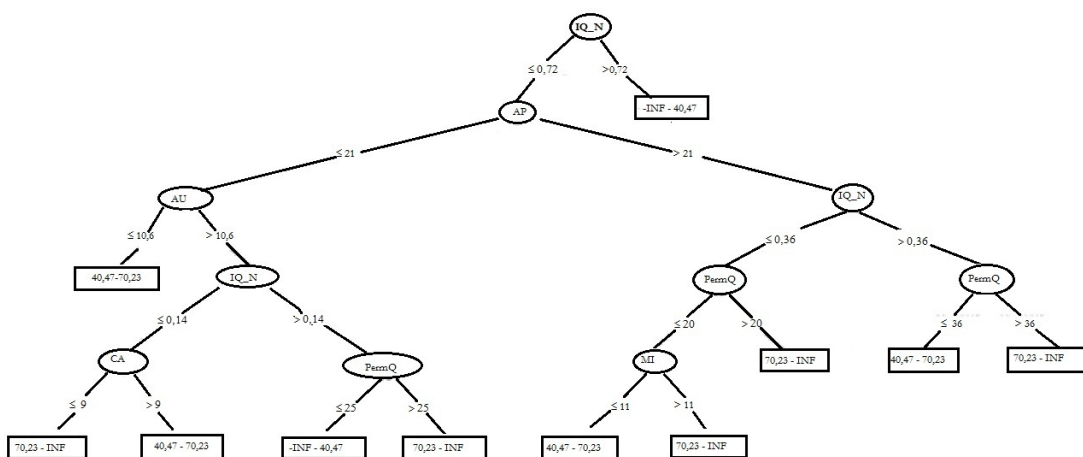
Figura 5.10: Percentual de instâncias classificada correta e incorretamente para Fósforo Total (nov-Jan/2009-2011)

apresentam uma média e um desvio padrão do I_{Q-N} de 0.99 e 0.32, respectivamente. Já para $I_{Q-N} \leq 0.72$ a média e o desvio padrão do I_{Q-N} foram de 0.45 e 0.15, respectivamente.

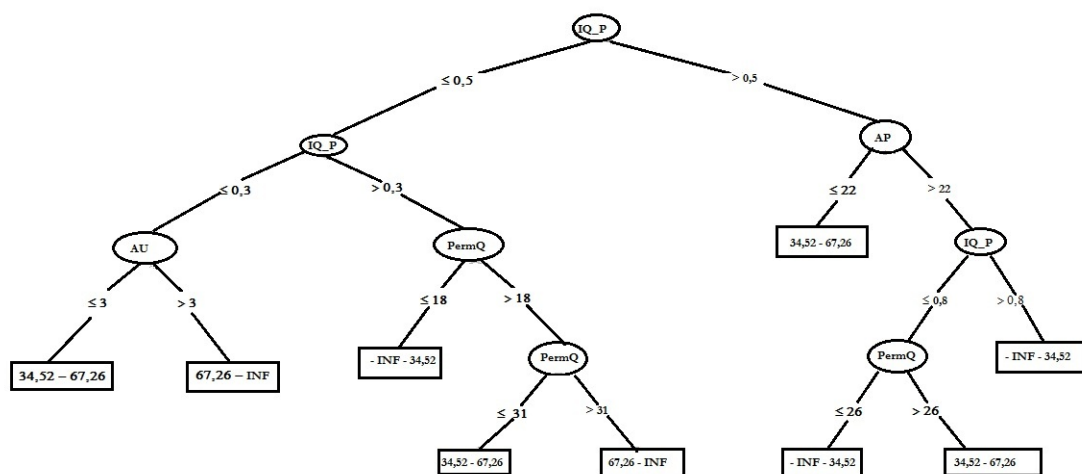
Tem-se ainda que os valores de $I_{Q-N} > 0.72$ ocorreram para uma permanência da vazão, em média, de 15%, contra 30% para $I_{Q-N} \leq 0.72$. Isso significa que o pressuposto de que a poluição difusa é principalmente observável com vazões elevadas é válido.

Com relação ao uso do solo (urbano e agropastoril), uma elevada dispersão nos dados para $I_{Q-N} > 0.72$ é observada (Tabela 5.6), o que faz com que a relação entre estas categorias de uso do solo e permanência do nitrato não seja estabelecida.

O uso do solo agropastoril (AP), urbano (AU) e o aumento da vazão foram responsáveis



(a)



(b)

Figura 5.11: (a) Árvore de decisão para o nitrato Nov-Jan/2009-2011 e (b) para o fósforo total Nov-Jan/2009-2011. Legenda: I_{Q-N} - Índice que relaciona a permanência da vazão (PermQ) a permanência do nitrato, AP - Percentual de agropastoril, AU Percentual de área urbana, CA - Percentual de Campo, MI - Percentual de Mineral e I_{Q-P} - Índice que relaciona a permanência da vazão (PermQ) a permanência do fósforo total.

Tabela 5.6: Média e desvio padrão das categorias de uso do solo Agropastoril e Urbano para as faixas de I_{Q-N}

I_{Q-N}	Agropastoril		Urbano	
	Média	Desvio	Média	Desvio
$\leq 0,72$	18.9	1.7	15.7	1.2
$> 0,72$	23.2	2.6	7.2	5.3

pelo aumento da concentração de nitrato representada pelo intervalo -INF a 40.48%.

As bacias de contribuição, das estações de qualidade avaliadas, com menor proporção de área urbana, foram responsáveis pela melhor qualidade da água em relação ao nitrato (40.48

a 70.23%). Tal efeito também foi verificado para a presença de uso do solo das categoriais campo (CA) e mineral (MI), ficando a permanência do nitrato entre 40.48 a INF.

Para o fósforo (Figura 5.11b), o uso do solo agropastoril (AP) foi associado à maior concentração de fósforo total, -INF a 34.52% para $I_{Q-P} > 0.5$.

As estações de qualidade com bacias de contribuição com menor percentual de área urbana foram responsáveis por menor concentração de fósforo total, resultando em sua maior permanência no tempo (34.52- 67.26%).

Para I_{Q-P} entre 0.3 e 0.5 não foi possível associar usos do solo à permanência do fósforo total nos três intervalos avaliados e o algoritmo trabalhou somente com as variações da vazão (PermQ), permitindo observar, neste intervalo que o aumento da vazão foi associado ao aumento da concentração de fósforo (Figura 5.11 b).

5.9.2 Avaliação das regras das árvores de decisão

Ao todo foram avaliadas 104 instâncias (Figura 5.12) relacionando a permanência de vazão à permanência do nitrato (entre 10.71% e 100%) e do fósforo total (entre 1.78 e 100%), respectivamente.

A permanência do nitrato e do fósforo total foi discretizada em 3 classes (melhor resultado obtido para classificação) C1, C2 e C3, com diferente número de instâncias (TA) enquadradas em cada classe (Tabela 5.7).

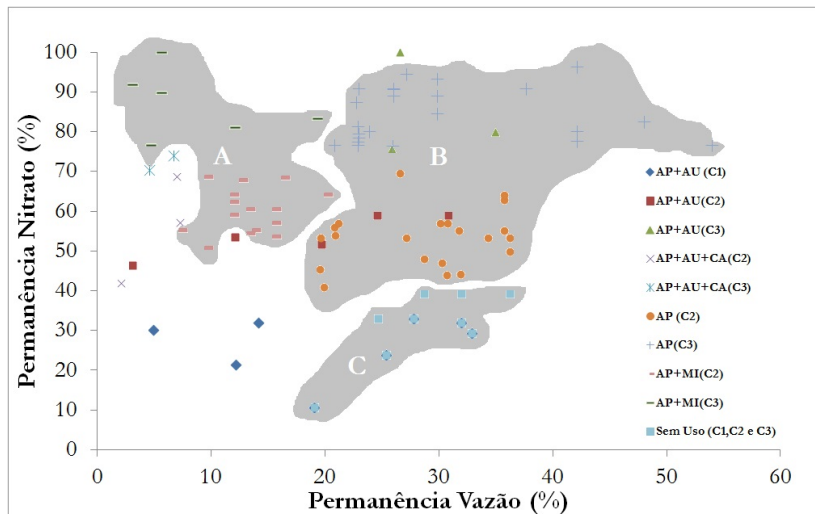
Tabela 5.7: Intervalos de discretização das variáveis dependentes permanência do nitrato e do fósforo total (Nov-Jan).

Parâmetro	C1/(TA)	C2/(TA)	C3/(TA)
Nitrato	-INF - 40.5 (18)	40.5 - 70.2 (46)	70.2 - INF (40)
Fósforo Total	-INF - 34.7 (36)	34.5 - 67.3 (22)	67.3 - INF(46)

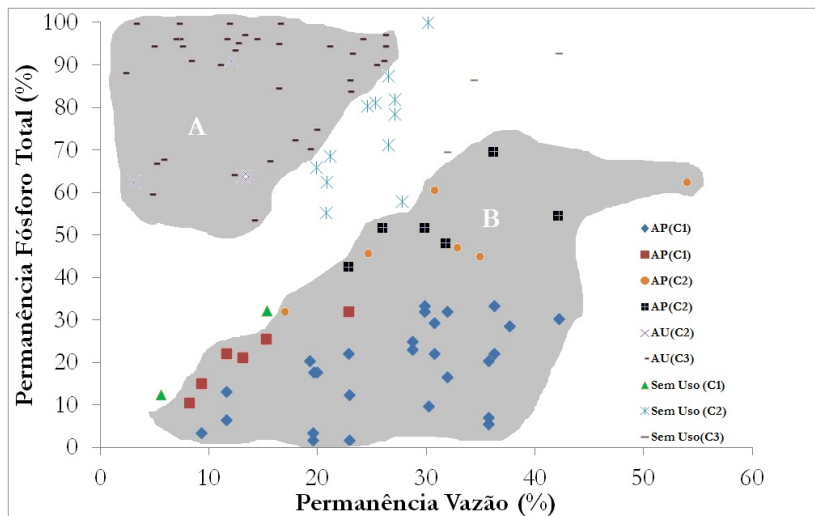
Legenda: C_i - Intervalo da classe TA - Número de Instâncias

As classes C1 e C3 representam alta e baixa concentração de nitrato e fósforo total avaliados na estações pesquisadas para o período de Nov-Dez de 2009-2011, respectivamente. A classe C2 possui valores intermediários (médios).

Os agrupamentos das classes para a permanência do nitrato e do fósforo total são apresentados na Figura 5.12. O uso do solo é representado pela cor dos símbolos pontuais. O uso do solo agropastoril (AP) associado ao mineral (MI) foi relacionado à média e baixa concentração de nitrato, agrupamento “A” da Figura 5.12a. A associação do uso urbano (AU) com agropastoril (AP) e campo (CA) apresentou confusão com os demais usos em todas as classes avaliadas (C1, C2 e C3). O agrupamento “B” apresentou os dados classificados com uso do solo agropastoril, também vinculados às classes C2 e C3 do nitrato. A classe C1 do nitrato teve baixa representatividade, em função do número reduzido de amostras nesta classe (TA=18) (Tabela 5.7).



(a)



(b)

Figura 5.12: (a) Agrupamento classes do nitrato e (b) Agrupamento classes do fósforo Total

Para o parâmetro nitrato, o maior problema esteve associado ao desbalanceamento entre as classes da Tabela 5.7. Ou seja, as classes C2 e C3 contêm a maior parte das instâncias, restando somente 18 instâncias para C1, o que se refletiu na dificuldade de associação da variável uso do solo para esta classe.

Para o fósforo total, o agrupamento “A” da Figura 5.12 b, contém o uso do solo urbano (AU) vinculado às classes de permanência C2 e C3, ou seja, com concentração variando de média a baixa de fósforo total. O uso agropastoril, agrupamento “B”, foi relacionado às classes C1 e C2 com concentração variando de média a alta.

5.10 Discussão

A aplicação do algoritmo C4.5 para o estabelecimento de regras de comportamento entre a permanência da qualidade dos parâmetros nitrato e fósforo total, a permanência da vazão e o uso do solo permitiu obter sucessos de classificação de 83% e 82%, respectivamente.

A eficiência da classificação foi maior para o trimestre chuvoso de Nov-Jan 2009-2011, com vazões com permanência entre 2 e 52%. Para o trimestre Dez-Jan 2009-2011 as vazões tiveram permanência variando entre 3 e 68%. Assim, verificou-se que os meses com maiores vazões obtiveram melhor resultado na classificação, colaborando com a associação da poluição difusa ao aumento da vazão.

Outro aspecto que reforça esta relação vazão-uso do solo-poluição difusa foi a seleção do índice I_{Q-QA} , como variável inicial para a divisão dos dados, haja visto que no período chuvoso, o I_{Q-QA} deve ser maior para fontes de poluição difusa e menor para fontes pontuais. Em se tratando de fontes pontuais a permanência da vazão diminui e a permanência do parâmetro de qualidade aumenta, indicando uma melhor qualidade das águas, aspecto representado pela permanência do fósforo no cluster A da Figura 5.12b.

Entretanto, foi possível observar que, a discretização utilizada para as variáveis dependentes (permanência de nitrato e fósforo total) influencia a classificação. O aumento no número de intervalos utilizados na discretização promoveu uma queda na eficiência da classificação. Tal fato ocorre, basicamente, em função do reduzido número de instâncias (dados disponíveis).

O monitoramento mensal das variáveis de qualidade não permite a construção de sequências temporais consistentes, para que as variações possam ser acompanhadas. Isto promove um desbalanceamento das classes dificultando a classificação. No caso do nitrato a distribuição de frequência das classes foi desbalanceada, e o algoritmo teve maior dificuldade de prever a classe mais rara (C1 - Tabela 5.7), confirmando o exposto por Breiman *et al.* (1984) e Witten e Frank (2005).

Apesar desta dificuldade as árvores de decisão confirmaram que os usos urbano e agropastoril estão associados à alterações de fósforo total e nitrato, verificadas na Bacia do Rio das Velhas, o que corrobora com as fontes poluidoras indicadas por IGAM (2013). Para o nitrato os clusters A e B (Figura 12a) estão associados ao uso do solo agropastoril, com a permanência do nitrato variando de 40.47 a INF. O modelo definido pela árvore de decisão não conseguiu prever o uso do solo associado a permanência do nitrato de -INF a 40.47 devido a número reduzido de instâncias nesta faixa de permanência (Figura 5.12a).

A permanência do fósforo total entre 34.5-INF ocorreu para vazões com permanência entre 10 e 25% (cluster A - Figura 5.12b). O uso do solo urbano foi associado a esta situação, onde o aumento da vazão esta vinculado a uma redução na concentração do fósforo total. O uso do solo agropastoril foi associado a menor permanência do fósforo total (-INF a

34.5) e permanência de vazões entre 8 e 53% (cluster B - Figura 5.12b). Ou seja, o aumento do escoamento superficial esteve associado ao aumento do fósforo total.

Alguns dados de fósforo total não puderam ser vinculados a nenhuma categoria de uso do solo e são visualizados na Figura 5.12b, através dos pontos não agrupados em clusters.

O uso de árvores de decisão se mostrou viável no processo de identificação das fontes poluidoras na Bacia do Rio das Velhas. Entretanto é necessário o aumento do volume de informações disponíveis para aprimorar a eficiência da ferramenta.

A obtenção de um maior número de observações, através da intensificação das campanhas de monitoramento, permitiria uma melhor exploração das possibilidades do uso de árvores de decisão.

5.11 Agradecimentos

Os autores agradecem ao Instituto Mineiro de Gestão das Águas (IGAM) pela concessão de dados do monitoramento de qualidade de águas da Bacia do Rio das Velhas; à Agência Nacional de Águas (ANA) e à Companhia Energética de Minas Gerais (CEMIG), pela concessão de dados do monitoramento de vazão; ao Instituto Estadual de Florestas (IEF), pela concessão das imagens do sistema RapidEye, ao INMET (Instituto Nacional de Meteorologia), pelo fornecimento das normais climatológicas e ao Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) pela liberação do professor para capacitação.

5.12 Referências Bibliográficas

Ver Referências Bibliográficas na página 134.

Capítulo 6

ARTIGO 2 : Modelling the effect of land use/land cover on nitrogen, phosphorous and dissolved oxygen loads in the Velhas River using the concept of exclusive contribution area

Lília Maria de Oliveira¹ and Philippe Maillard² and Éber José de Andrade Pinto³

6.1 Abstract

Non-point source water pollution is a major problem in most parts of the World, but is also very difficult to quantify and control since it is not easily separated from point sources and can theoretically originate from the whole watershed. In this article we evaluate the relationship between land use and land cover and four water pollution parameters in a watershed in Southeast Brazil. The four parameters are nitrate, total ammonia nitrogen, total phosphorous, and dissolved oxygen. To help concentrate on non-point source pollution, only data from the wet seasons of the time period (2001–2013) were analysed, based on the fact that precipitation causes runoff which is the main cause of diffuse pollution. The parameters measured were transformed into loads, which were in turn associated with an exclusive contribution area, so that every measuring station could be considered independent. Analyses were also performed on riparian zones of different widths to verify if the effect of the land cover on the water quality of the stream decreases with the increased distance.

¹CEFET-MG, Av. Amazonas, 5354, Belo Horizonte, Brasil Tel.: +55-31-33770361 email: lilia@deii.cefetmg.br

²Universidade Federal de Minas Gerais, Instituto de Geociencias, Av. Antônio Carlos, 6627, Belo Horizonte, Brasil; Tel.: +55-31-33770361; email: philippermaillard@yahoo.com.br

³Universidade Federal de Minas Gerais, Escola de Engenharia, Av. Antônio Carlos, 6627, Belo Horizonte, Brasil; Tel.: +55-31-33770361; email: eber.andrade@cprm.gov.br

Pearson correlation coefficients indicate that urban areas and agriculture/pasture tend to worsen water quality (source). Conversely, forest and riparian areas have a reducing effect on pollution (sink). The best results were obtained for total ammonia nitrogen and dissolved oxygen using the whole exclusive contribution areas with determination coefficients better than $R^2 \approx 0.8$. Nitrate and total phosphorous did not produce valid models. We suspect that the transformation delay from total ammonia nitrogen to nitrate might be an important factor for the poor result for this parameter. For phosphorous, we think that the phosphorous sink in the bottom sediment might be the most limiting factor explaining the failure of our models.

6.2 keywords

Water quality, Non point pollution, Regression models, Land use land cover, Nitrogen, Dissolved oxygen, Phosphorous.

6.3 Introduction

The monitoring of water quality (WQ) in Brazil is performed with various objectives, one of which is the determination of the quality of a stream for the classification of its possible use. But it is also the understanding of the spatial and temporal evolution of pollutant sources that is essential for Brazilian water agencies (IGAM, 2013b). Regardless of the specific objectives, monitoring is performed by measuring the concentrations of three sets of parameters: 1) physical, 2) chemical, and 3) microbiological. These concentrations are associated with pollutants from both point and non-point sources that can reach the stream directly through sewers or through natural processes such as leaching in the soil or runoff (McDowell *et al.*, 2001; Quinn e Stroud, 2002). Consequently, land use and land cover (LULC) can have, directly or indirectly, a strong effect on the WQ. As such, the watershed and its subdivisions are the most effective unit areas to model the effect of LULC on WQ parameters.

The concentrations of these parameters are highly dependent on the environmental conditions of the contribution area within the watershed (Pinheiro e Deschamps, 2008). These data are useful for comparing the quality of a stream with pre-established standards for the type of use to which it is destined. However, since concentrations are dependent on discharge it is highly variable and does not enable a fair comparison between different areas. Conversely, pollutant loads are obtained from the product of the concentrations by the discharge, and are usually measured in weight over time (Tucci *et al.*, 1998; Dingman, 2004, p. 544 and 65 respectively). Loads are considered better for comparing watersheds, and for establishing the relationship between pollutants and other physical factors such as land use

(Quilbé *et al.*, 2006). Many authors have also preferred loads over concentrations to convey their research (Yang *et al.*, 2007; Diebel *et al.*, 2009; Ouyang *et al.*, 2010; Zucco *et al.*, 2012; Outram *et al.*, 2016).

Any distributed study of WQ of a watershed should consider the complex transformation processes (physical, chemical and microbiological) that can take place through space and time. For instance, when sewers are the main source of pollution, the nitrogen cycle involves a series of transformations: from a given form of nitrogen (organic, ammonium, nitrite, or nitrate), one can deduce whether the pollution stage is remote (nitrate) or recent (organic nitrogen or ammonium) (Even *et al.*, 2007).

Land use can contribute in polluting a stream as a non-point source but it can also act as a filter and keep pollutants from entering the stream, by leaching or by biological fixation. Vegetation, especially in the riparian zone, can play an important role as a filter (Sliva e Williams, 2001; Maillard e Santos, 2008). WQ parameter loads can be related to LULC by trying to identify which categories act to reduce pollution and which increase it. Examples of this type of approach can be found throughout the specialized literature. Buck *et al.* (2004) found that the amount of pasture in two watershed in New Zealand are well correlated with concentrations of total nitrogen and nitrate in one watershed and phosphorous and turbidity in the other. Sliva e Williams (2001) found that urban land use was positively correlated with various water quality parameters in three watersheds in Ontario while forest appeared as a mitigating factor. Salvia-Castellví *et al.* (2005) calculated that nitrogen loads were seven to eight times higher in agriculture watershed than in forested ones and mixed watersheds produced an intermediate value. Poor e McDonnell (2007) reached similar findings but also included residential areas which produced even more pollutants than the agriculture areas. Ouyang *et al.* (2010) in the upper Yellow River watershed found that agriculture contributed more to increasing organic nutrients loads than grassland and forest. Li *et al.* (2015) studied the JinJing watershed in China and characterized agricultural and residential land use classes as pollutant “sources” while forested areas acted as “sink”. We also opted to use these same two terms to characterize our LULC classes.

Another aspect generally taken into consideration is the determination of the area of influence of the LULC. Studies have used the whole watershed, the exclusive contribution area (ECA), or bands of varying width on either side of the stream—generally called buffers (B)—or even combinations of the three. We considered in our experimental design that the strength of the relationship between the LULC and the WQ parameters may change as the distance to the stream increases. Furthermore the way in which this relation changes may be different for each parameter. Some authors have also reached this conclusion (Johnson *et al.*, 1997; Buck *et al.*, 2004; Chang, 2008; Li *et al.*, 2015).

The river discharge (Q) should also be considered because it has a strong influence on WQ parameters, not only because of the increased dilution of a greater discharge but also

because of the better autodepuration capacity it provides (Roque *et al.*, 2008) (Chesterikoff *et al.*, 1992). In regions with a well defined rainy season, like many parts of Brazil, runoff can be null for many consecutive months during which non-point pollution sources usually have a reduced effect on the stream water quality (Stutter *et al.*, 2008; Maillard e Santos, 2008; Outram *et al.*, 2014; 2016). As such, changes in the hydrological regime should be considered when trying to determine the effect on non-point pollution within a watershed, especially when making predictions or explanatory models. The study of the hydrogram is usually the best way to determine the start of the rainy season.

As a means of summarizing recent studies relating WQ and land cover, Table 6.1 shows a selection of articles published on this theme. These articles all have at least one WQ parameter in common with the ones we have studied in our research. For each study, the table presents the time frame used and whether the authors used the whole watershed (WS), an exclusive contribution area (ECA), or a buffer area on both sides of the stream (B). The LULC classes used in each study are also shown.

Tabela 6.1: Synopsis of studies that have analyzed the parameters ammonia (NH_4^+), nitrate (NO_3^-), nitrite (NO_2^-), total phosphorous (TP), total nitrogen (TN), nitrogen (N), and dissolved oxygen (DO) in relation to land use and land cover.

Authors	Country	WQ parameter	Study area ¹	LULC classes ²
Sliva e Williams (2001)	Canada	NH_4^+, NO_3^-	$WS_1 = 10.16$ ha; $WS_2 = 33.28$ ha, $WS_3 = 28.65$ ha & B=100 m	Ft,Ub,AP
Gardi (2001)	Italy	NO_3^-	WS=273 ha	Ft,Ub,BS,AP
Quinn e Stroud (2002)	New Zeland	NH_4^+, NO_3^- and TN	WS_i (i= 1 to 8) from 1.1 to 30 ha	Ft,Pa
Yang <i>et al.</i> (2007)	China	NH_4^+, NO_3^- and TN	WS= 4.07 ha	Ft, Ag, Ub
Poor e McDonnell (2007)	USA	NO_3^-	$WS_1, WS_2, WS_3 \approx 0.50$ ha	$WS_1=Ub$ $WS_2=Ag$ $WS_3=Ft$
Maillard e Santos (2008)	Brazil	NO_3^-, NO_2^-, TP	16 ECA & B= 90, 150, 210, 300, 510 m	Ub,Ag,BS,Ft,RF
Diebel <i>et al.</i> (2009)	USA	TN, TP	WS = 616 ± 784 (km ²) and B= 30 and 90 m	Ub, Ag, AP, Ft, RF
Ouyang <i>et al.</i> (2010)	China	N, TP and NO_3^-	WS = 34 ha	FL, Ag and Pa
Jarvie <i>et al.</i> (2010)	United Kingdom	TP, NO_3^-	$WS_1=25$ ha; $WS_2=3.14$ ha; $WS_3=6.3$ ha	Ag,Pa
Zucco <i>et al.</i> (2012)	Brazil	TN, TP	$ECA_1 = 72.9$ km ² ; $ECA_2 = 119.2$ km ² , $ECA_3 = 140.4$ km ² ; $ECA_4 = 158$ km ²	Ub,Ag,Ft,Pa
Sun <i>et al.</i> (2013)	China	TN	WS=318.000km ²	Ub,Ag,Ft,Pa,W
Evans <i>et al.</i> (2014)	USA	TN, TP, NO_3^-, NH_4^+	WS=966km ²	Ag,Ft,BS
Pinheiro <i>et al.</i> (2014)	Brazil	$NH_4^+, NO_3^-, NO_2^-, OD$ and TP	$WS_1 = 21.67$ km ² ; $ECA_2 = 9.79$ km ² ; $ECA_3 = 14.99$ km ²	$ECA_1=Ft,Ag;$ $ECA_2 =Ft,Ag;$ $ECA_3 =$ Ft,Ag,Ub
Li <i>et al.</i> (2015)	China	NH_4^+, NO_3^- , TN and TP	18 ECA and B= 250, 500 and 750 m	Pa,Ag,Ft,Ub,W
Bowes <i>et al.</i> (2015)	United Kingdom	NO_3^-	WS=148 km ²	Ft,Ag,Ub
Outram <i>et al.</i> (2016)	United Kingdom	NH_4^+, NO_3^- , TP	WS=570km ²	Ag,Pa,Ft

¹ Study Area: Watershed (WS), Exclusive contribution area (ECA), Riparian buffer (B) ²Land use/land cover: Ub – Urban; Ag – Agriculture; Ft – Forest; AP – Agriculture and pasture; Pa – Pasture, RF – Riparian forest, W – Water and BS – Bare soil

N and P are normally encountered in unaltered stream water in various forms: organic

or inorganic, soluble or in particles, available or not for phytoplankton (Golterman, 2007). Sewers are normally the main source of N in organic form or as ammonium. Organic N can undergo sedimentation and be converted into ammonium in the sediments or in the water itself. In time, the ammonia will transform into nitrite and then nitrate. Although this can occur before the sewer reaches the stream (during transport or in the water treatment plant), it can also happen in the stream, thus causing the latter forms (nitrite and nitrate) to be found downstream from the early forms (Chesterikoff *et al.*, 1992). This will depend on a number of factors, including the velocity of the current and the discharge (Haggard *et al.*, 2001). The conversion of ammonium into nitrate consumes large quantities of oxygen: 4.3 mg/L of O₂ are needed to convert 1 mg/L of ammonium; a process that can take up to many kilometres depending on the temperature, pH, DO, and turbulence (Von Sperling, 2007).

Domestic sewers are important sources of inorganic P in the form of orthophosphate and polyphosphate. While the former is readily available for biological organisms, the latter are complex molecules that slowly transform into orthophosphate (Von Sperling, 2007). The P found in streams can be either in organic or inorganic form, in the water or in the sediments. Sedimentation of organic P causes it to be removed from the water as a measurable parameter and be fixed in a particle form (Golterman, 2007). Conversely, organic P in sediment can migrate to dissolved inorganic P causing an increase in concentrations in the water (Gibson e Meyer, 2007).

Agricultural activities are important sources of P. Even though supplies of P from urban runoff are twice as much per unit area than from agriculture (Von Sperling, 2007), the much higher proportion of area dedicated to agriculture generally makes it the prime source of P (Grizzetti *et al.*, 2008). However, the migration of P from sediments to water and vice-versa can make this relation difficult to verify.

The understanding of the transformation between the different forms of these pollutants (*e.g.* from ammonia to nitrate) and the relation between some of the parameters themselves (*e.g.* the absorption of DO by the transformation of N) is essential for establishing a causal relation between LULC and WQ, especially when considering the separation between point- and non point-pollution sources (Grizzetti *et al.*, 2008).

In this article, we evaluate how the land use and land cover (LULC) affect the changes in loads of total ammonia nitrogen (TAN), nitrate (NO_3^-), total phosphorous (TP), and dissolved oxygen (DO) in the Velhas River in Minas Gerais, Southeast Brazil. To do so we have considered exclusive contribution areas (ECAs) for each monitoring station. Concentrations were transformed in “incremental loads” to avoid a cumulative effect from stations situated upstream from each other. We also considered strips of varying width (buffer areas) on either side of the stream to verify how the hydrological distance¹ affects the strength of the rela-

¹“Hydrological distance” considers both horizontal and vertical distances between dry land and the stream (Nardi *et al.*, 2008; Van Nieuwenhuysen *et al.*, 2011)

tion. WQ data from 17 stations situated along 800 km of the Velhas River were used in 12 successive rainy seasons between 2001 and 2013.

6.4 Methods

6.4.1 Study Area

The Velhas River watershed covers an area of 27,887 km² in the central part of the State of Minas Gerais (Figure 6.1) and is home to 4.4 million people, most of whom live in the capital city of Belo Horizonte and its metropolitan area. The Velhas River Watershed is responsible for 74% of the water supply of the city of Belo Horizonte (CBH, 2015).

The climate is semi-humid tropical of altitude, with a well defined wet (summer) and dry (winter) season. Precipitation decreases as one moves downstream: from an annual average of 1700–1600 mm near the source of the river, to about 1100 mm at the mouth (Pinto *et al.*, 2011). The rainy season normally starts in October and extends until March. Average temperatures range from 18⁰C in the south end and reach 23⁰C in the north end (IGAM, 2005).

Near the source of the Velhas River, the terrain is hilly to mountainous with narrow valleys and gorges and streams of low sinuosity and high structural control. Downstream, towards the mid- and low-watershed, the river starts meandering and the terrain becomes smoother with rolling hills and a wide and flat river valley with terraces. The river source is situated at an altitude of 1500 m whereas its mouth is at 478 m (IGAM, 2013a).

The watershed LULC was classified into the nine following categories with their percent areas: savanna (Sa, 26.4%), grassland (Gl, 21.3%), agriculture and pasture (AP, 25.2%), forest (Ft, 15.3%), riparian forest (Rf, 0.8%), rock outcrop and open mines (RM, 4.2%), tree plantations (TP, 3.3%), urban (Ub, 2.9%), and water, clouds and other (WC, 0.6%) (Figure 6.1). The LULC mapping was done using image segmentation and classification of a mosaic of 72 RapidEye images from 2010. RapidEye is a constellation of five Earth observing optical satellites with a ground resolution of five meters and five spectral bands ranging from the blue to near-infrared wavelengths (BlackBridge,<http://blackbridge.com/rapideye>). The classification was performed using Trimble eCognition Developer 8.4 software.

It would have been ideal to use consecutive version of the LULC for the period of analysis (2001-2013). Although we did find older versions of the LULC map — one from 2002 (Maillard e Santos, 2008) and one from 2004 (IGAM, 2005) — the differences in scale and LULC classes made them unpractical to use. We then opted to use a single date 2010. Although suboptimal, we argue that this shortcoming (*i.e.* the changes in LULC that might have occurred between 2001 and 2010 and between 2001 and 2013) will fall into the unex-

plained variance of our models.

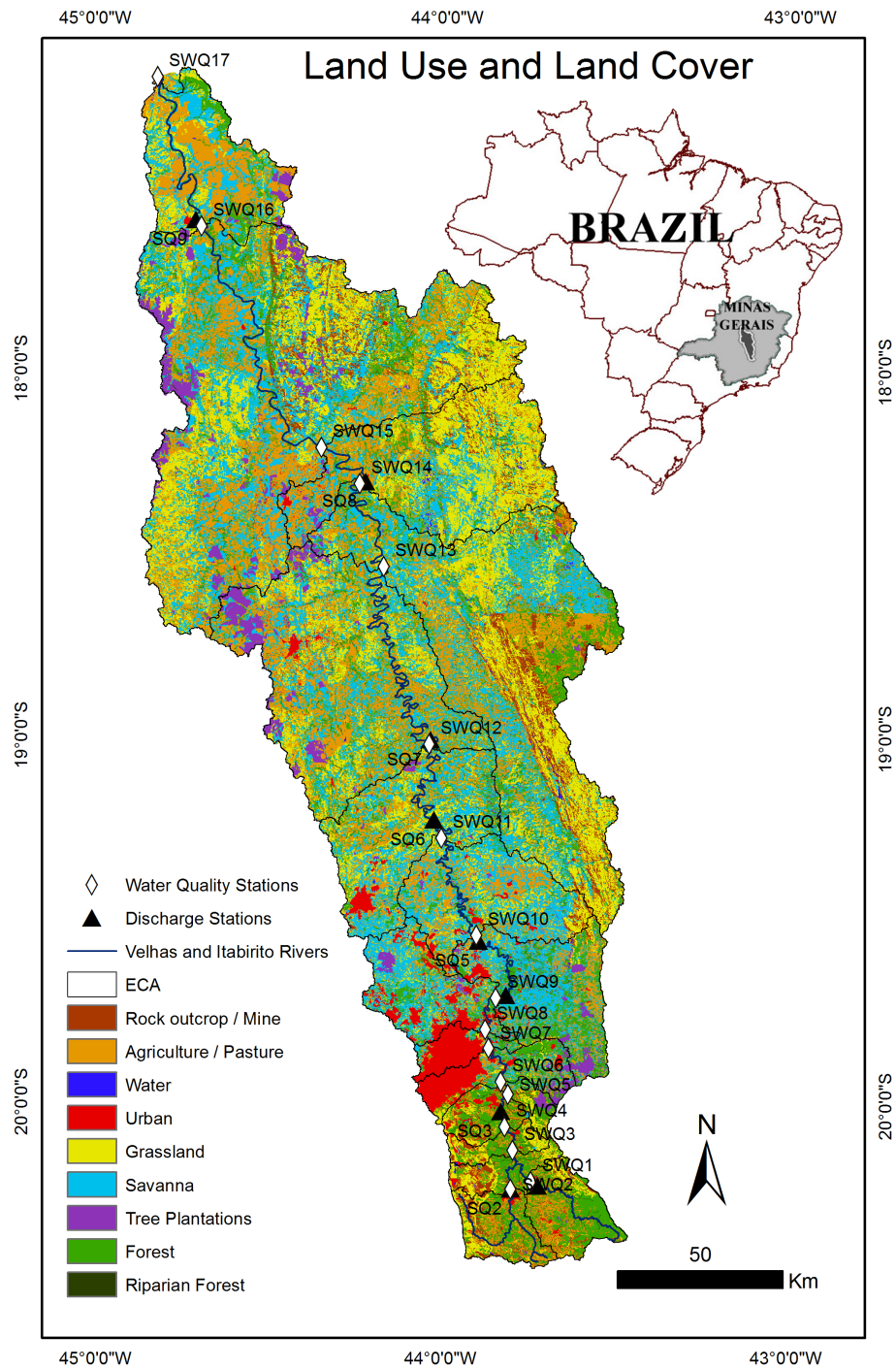


Figura 6.1: Map of the Velhas River Watershed showing the LULC classes, the exclusive contribution areas (ECA) and the monitoring stations (discharge and quality).

6.4.2 Water quality and water discharge data

The beginning of the rainy season was defined by finding a significant break upward in the daily hydrogram (Figure 6.2). Table 6.2 shows these starting dates for each station and each year. The end was set at 31 March regardless of the year. Unlike the beginning of the rainy season, the end is progressive and not easily pinpointed so that an arbitrarily selected date does not have any strong effect on the data. Like other authors, (Stutter *et al.*, 2008; Outram *et al.*, 2014; 2016) we opted for using only the rainy season data because this is when runoff occurs and create the strongest effect of non-point pollution.

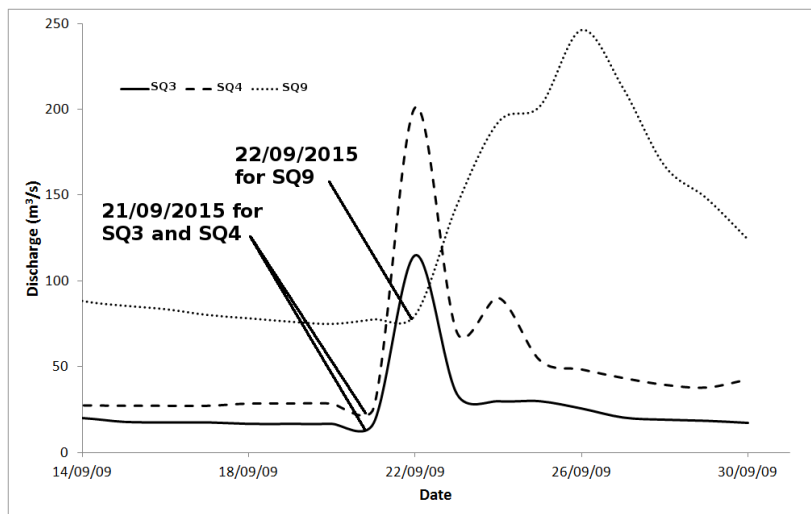


Figura 6.2: Illustration of how the starting date of the rainy season was selected from the daily hydrogram of each station.

Tabela 6.2: Rainy season starting dates for the 2001–2013 period for the nine daily water discharge monitoring stations

Period	Water Discharge Monitoring Stations		
	SQ_1-SQ_3	SQ_4-SQ_6	SQ_7-SQ_9
2001–02	15/9/01	15/9/01	18/9/01
2002–03	7/11/02	23/10/02	2/11/02
2003–04	4/11/03	2/11/03	26/10/03
2004–05	4/10/04	4/10/04	8/10/04
2005–06	24/9/05	8/9/05	21/9/05
2006–07	21/9/06	20/9/06	25/9/06
2007–08	18/10/07	19/10/07	24/10/07
2008–09	15/9/08	16/9/08	19/9/08
2009–10	21/9/09	21/9/09	22/9/09
2010–11	29/9/10	30/9/10	2/10/10
2011–12	14/10/11	5/10/11	8/10/11
2012–13	5/11/12	26/09/12	3/11/12

Our approach required that the data from each WQ monitoring station be analyzed independently so that the load from one station does not directly affect the data in the downstream stations. To do so, we created for each station an exclusive contribution area (ECA) based on the topography in the watershed. Basically, an ECA was created for each monitoring station that excluded the watershed from the previous station upstream. Then, successive strips of varying width were defined on either side of the stream in an effort to determine the strength of the relationship between the LULC and WQ with the increase in hydrological distance. Figure 6.3 illustrates this process. A digital surface model (DSM) with a ground cell resolution of 30 m from the *shuttle radar topographic mission* (SRTM) was used to generate these areas. Unfortunately, the water discharge stations did not coincide with the WQ stations, so we calculated a compensation factor ($CA_{WQ/Q}$) to regionalize the daily mean discharge data to correspond to the coordinates of the WQ stations (Eq. 6.1-6.2). Table 6.3 shows these compensation factors for all WQ stations (S_{WQ}).

$$CA_{WQ/Q} = \frac{A_{WQ}}{A_Q} \quad (6.1)$$

$$Q_{WQ} = Q \times CA_{WQ/Q} \quad (6.2)$$

Where:

- $CA_{WQ/Q}$ is the area compensation factor (dimensionless);
- Q_{WQ} is the daily discharge at the S_{WQ} station (m^3/s);
- Q is the daily discharge at the S_Q station (m^3/s);
- A_{WQ} is watershed contribution area for the S_{WQ} station (km^2);
- A_Q is watershed contribution area for the S_Q station (km^2).

The daily mean discharge was obtained by averaging the daily monitoring carried out at 7 am and 5 pm. Data for the four parameters considered in our study (total ammonia nitrogen, nitrate, total phosphorous, and dissolved oxygen) were downloaded from the Minas Gerais water agency (*Instituto Mineiro de Gestão de Águas - IGAM*). The data was collected every three months between 2001 and 2007 then monthly from 2008 to 2013. The concentrations of the four parameters were then converted into loads for the 17 monitoring stations (Figure 6.1) using the following relation (Eq. 6.3):

$$L = K * C * Q_{WQ} \quad (6.3)$$

Where:

- L is the load in kg/day

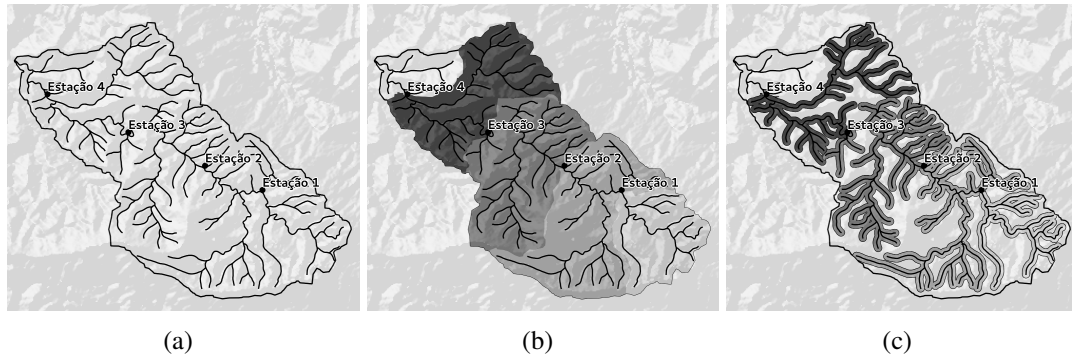


Figura 6.3: Illustration of how the exclusive contribution areas (ECA) and the river edge strips were created from the digital surface model: (a) watershed with hydrography and quality gauging stations, (b) determination of the ECA, and (c) strips of varying width (buffers) on either side of the stream.

Tabela 6.3: Watersherd contribution areas for the S_{WQ} and S_Q stations and the corresponding areas compensation factors ($CA_{WQ/Q}$) applied to the Q stations to correspond to the WQ stations.

S_{WQ}	A_{WQ}	S_Q	A_Q	$CA_{WQ/Q}$
1	549.9	1	490.3	1.12
2	467.0	2	467.0	1
3	1201.9	3	1678.5	0.72
4	1513	3	1678.5	0.90
5	1829.8	3	1678.5	1.10
6	1964.1	3	1678.5	1.17
7	2497.1	3	1678.5	1.48
8	2754.8	4	3735.1	0.74
9	3668.2	4	3735.1	0.98
10	4967.5	5	4852.5	1.02
11	6244	6	6244	1
12	8030	7	8030.8	1
13	11659.9	8	16638	0.7
14	16638.2	8	16638	1.0
15	19064.0	8	16638	1.15
16	26391.6	9	26019.4	1.01
17	27853	9	26019.4	1.07

- K is a unit conversion constant (86.4)
- C is the concentration in mg/l
- Q_{WQ} is the discharge in m^3/s

Since we only had LULC data for one year and we considered it constant for the whole period, the median loads of the four WQ parameters were calculated for each hydrological year (j) and for each of the 17 stations (i). This generated 17 median values (one for each

station) for each parameter. The median was selected as the measure of central tendency because it reduces the weight given to extreme values (Miller, 1993).

The loads then became $Lx_{i,j}$, where x is one of the four parameters. These loads were then transformed into incremental loads (ILx), meaning that for each station, the load from the previous station (upstream) were subtracted (Eq. 6.4) except for the first two, which do not have any upstream station (Eq. 6.5). Station 3 was a particular case since it had two upstream stations from two different rivers (Eq. 6.6)

$$ILx_{i,j} = Lx_{i,j} - Lx_{i-1,j} \quad \text{for } i=4,\dots,17 \quad j=1,\dots,12 \quad (6.4)$$

$$ILx_{i,j} = Lx_{i,j} \quad \text{for } i=1,2 \quad j=1,\dots,12 \quad (6.5)$$

$$ILx_{3,j} = Lx_{3,j} - (Lx_{1,j} + Lx_{2,j}) \quad \text{for } i=3 \quad j=1,\dots,12 \quad (6.6)$$

These loads can vary widely since each station is influenced by a different contributing area. These areas are shown in Table 6.4 in the Results section. To normalise these incremental loads, we divided each one by the ECA to produce unit incremental loads (UIL) in $kg \cdot km^{-2} day^{-1}$ (Eq. 6.7).

$$UILx_{i,j} = ILx_{i,j} / ECA_i \quad \text{for } i=1,\dots,17 \quad j=1,\dots,12 \quad (6.7)$$

where ECA_i is the exclusive contribution area of station i .

With this procedure, illustrates by Figures 6.4, we aimed at verifying whether the LULC between two monitoring stations had an increasing or decreasing effect on the WQ parameters, without scale problems.

6.4.3 Statistical Analysis

Before doing any test or model, the data was analysed using bar graphs, box plots and maps. These tools were meant to add spatial context to the analysis and offer a means for a qualitative as well as a quantitative evaluation of the data in order to help interpret the results in their context.

Bar graphs were used to analyse the spatial distribution of the LULC data within the ECA and within the riparian strips of different width. This is likely to help understand why certain regression models have retained or excluded some LULC class. Box plots were produced to understand the dynamics of the WQ data along the stream. Box plots are useful for easily evaluating a series of distribution parameters, including central tendency, dispersion, asymmetry, outliers, and the general spatial behavior of the WQ parameter, by plotting the stations in downstream order (Naghetini e Pinto, 2007).

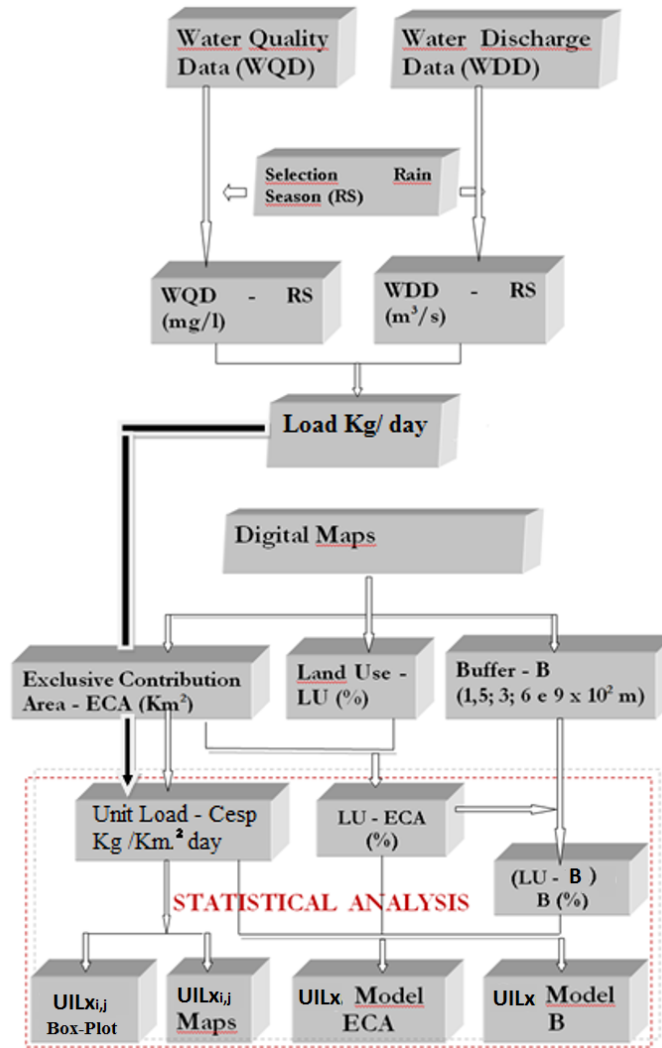


Figura 6.4: Flow diagram of the methodological steps.

The Pearson correlation coefficient was used for two distinct reasons. First to gain knowledge on what to expect as to which LULC variables are more likely to be correlated with the WQ parameters and to verify the sign (\pm) of the relation. Secondly to verify the degree of linear correlation between the independent variables that may compromise the multiple regression models.

The relation between LULC and WQ was defined for the four parameters using multiple regression models. The models take the form $Y=b_0 + b_1X_1 + \dots + b_nX_n$ where the X represent the independent (explaining) variables, Y is the dependent (explained) variable; the WQ parameters here and the $b_{1,2,\dots,n}$ are the beta weights for each LULC variable used in the model. These models were built for the whole of the ECAs and the different riparian strips as well. For each regression model, only the LULC variable with a significant correlation with the WQ parameter were included.

Lastly, an analysis of variance was performed to test the null hypothesis that the beta weights are null ($H_0 : b_{1,2,\dots,n} = 0$; $H_1 : b_{1,2,\dots,n} \neq 0$). A significance level of 5% was adopted to reject the null hypothesis. The residuals were also evaluated to check their independence and normal distribution.

6.5 Results

6.5.1 Land use / land cover per exclusive contribution areas and riparian zones

To represent the nine land cover classes presented in Figure 6.1, we overlaid the 17 exclusive contribution zones (ECAs) corresponding to the 17 stations and created a cross classification map so that we could define the proportion of each class within each zone (Figure 6.5e). Then the same thing was done for each of the four riparian zones (150, 300, 600 and 900 metres) and shown in Figure 6.5a to 6.5d. To facilitate the interpretation, we used the same color coding as in the land cover map shown in Figure 6.1.

The analysis of Figure 6.5 for the five zone types revealed a number of observations that are critical for understanding the results presented in the following sections.

- Agriculture and pasture (AP) are evenly represented all five zones;
- Urban (Ub) occupies a larger proportion in the three riparian zones for stations 4 through 9;
- Grassland (Gl) is well represented in the ECAs but not much in the riparian zones;
- Savanna (Sa) is mostly represented from station 6 downstream;
- Tree plantations represent a very small proportion of the land in all riparian zones and in the whole ECAs;
- Forests are mainly concentrated in the upper watershed between ECAs of stations 1 through 6 while ECAs 7 to 9 have very little forest mostly because this is where Belo Horizonte is located;
- With only 0.8% in the whole watershed, riparian forest is sparsely represented but has a significant importance for being the last barrier between the dry land and the stream.

6.5.2 Flow stream analysis of the distribution of the WQ parameters

The incremental loads for the four WQ parameters were analyzed sequentially in the upstream–downstream order and are presented as box plots in Figure 6.6 and Table 6.4. Our

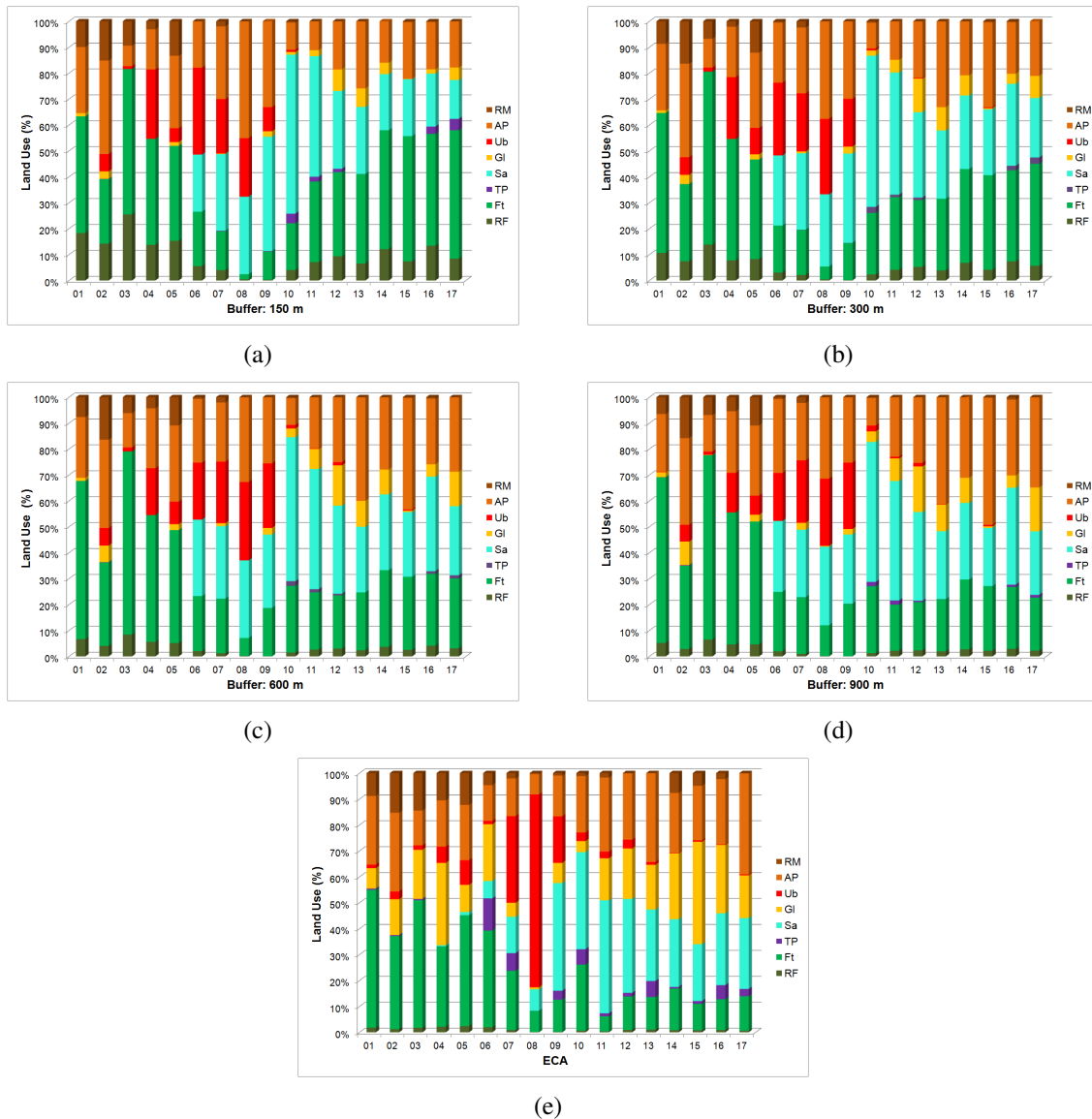


Figure 6.5: Land use / land cover (%) per exclusive contribution and riparian zones: (a) 150 meters, (b) 300 meters, (c) 600 meters, (d) 900 meters and (e) ECA

analysis revealed a number of facts that are summarized below. It should be noted that these figures represent the incremental loads, meaning that the loads per unit area of the previous station upstream were subtracted from each station (except the first one which does not have a previous station). This explains why some stations registered negative values.

- Incremental loads of total ammonia nitrogen (TAN) vary widely between stations 6 and 12 and are very high for stations 7 and 8 because of the situation of these two stations near Belo Horizonte, which contributes the most to point pollution through sewers.
- Incremental loads of nitrate (NO_3^-) are more stable compared with ammonia. Station 8 shows a strong reduction compared with station 7, explained by the absence of the

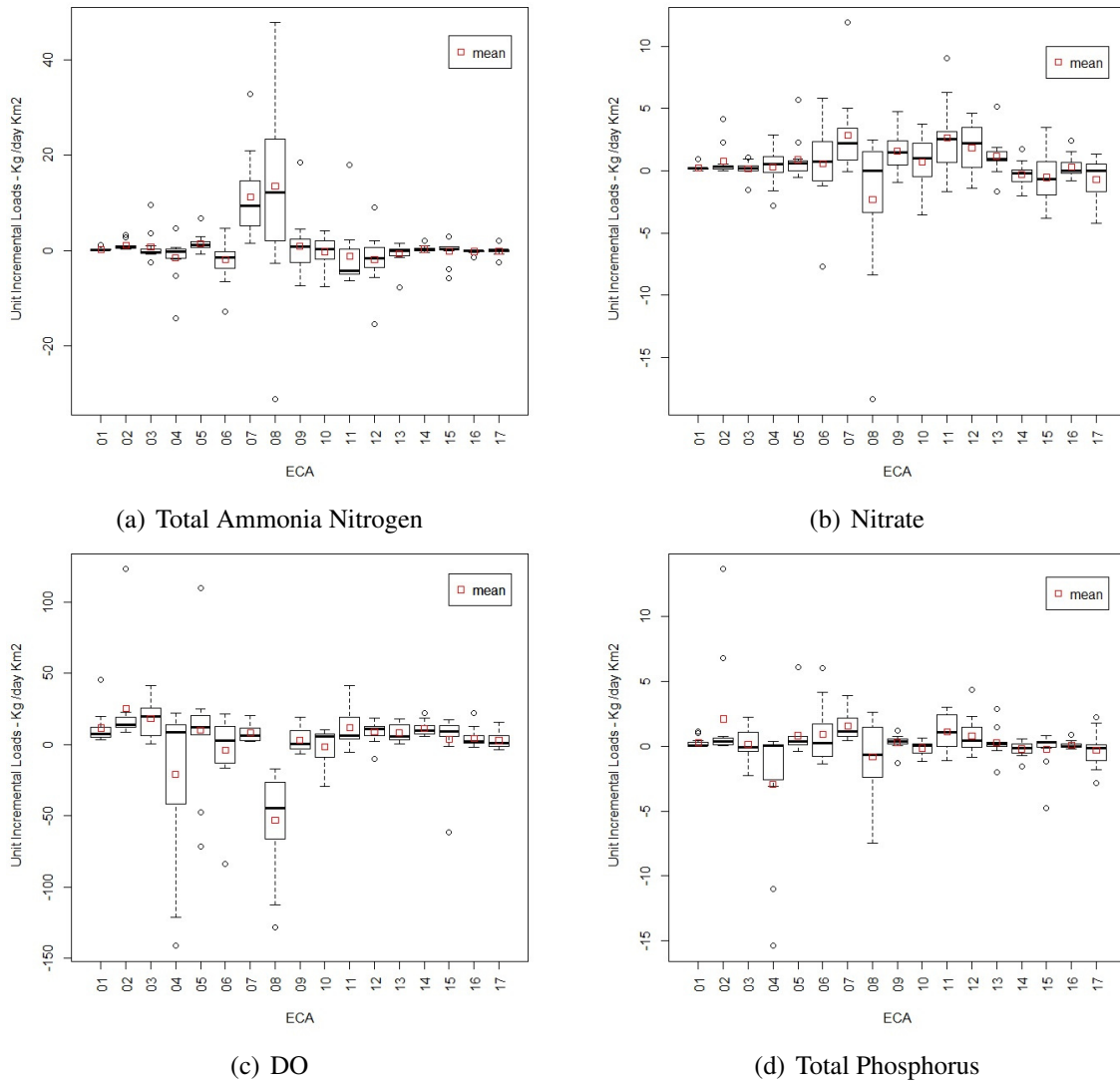


Figure 6.6: Incremental loads for the four WQ parameters (kg / day km²)

OD necessary to convert ammonia into nitrate. The Incremental loads of nitrate are almost always positive for the first 13 stations, showing a gradual increase of nitrate downstream up to station 13. We attribute this to the time delay needed for the conversion of ammonia into nitrate. Since the successive stations are downstream from one another, this conversion time translates to distance. After station 13, the Paraúna River with an average discharge of $72.4m^3s^{-1}$ merges with the Velhas River improving significantly its quality.

- For dissolved oxygen, station 8 stands out as being the only negative value, explained by the presence of the metropolitan area of Belo Horizonte with sewers as the main source of point pollution, causing a high consumption of oxygen through a series of chemical processes, one of which is the conversion of ammonia into nitrate.
- The total phosphorous is generally less variable and no consistent pattern is observable.

As explained above, part of the difficulty of analysis comes from the fact that only the water is analysed and not the sediments, in which some P can be fixed and excluded from the analysis.

It is difficult to make a thorough analysis of the distribution of these parameters because of the number of factors that can interfere in the process. Still we feel the need to outline the strong effect that the urban area of Belo Horizonte has for the four parameters. This is mainly caused by point pollution sources. Conversely we can infer that the agriculture and pasture (as well as some areas of grassland and savanna that can be used as pasture) contribute to non point-pollution, but since these areas are much more spread out than urban areas, their direct effect is harder to identify.

Tabela 6.4: Area and unit incremental loads (median) for all exclusive contribution areas (ECA) kg/km²·day

ECA	Area (km ²)	UIL			
		TAN	NO ₃ ⁻	TP	OD
1	549.9	0.20	0.21	0.07	7.56
2	467.1	0.62	0.31	0.35	14.00
3	288.6	-0.29	0.20	-0.05	20.08
4	207.4	-0.25	0.52	0.03	8.47
5	307.8	1.11	0.60	0.39	12.44
6	143.3	-1.36	0.76	0.27	2.89
7	533.0	9.43	2.20	1.17	6.39
8	257.7	12.11	-0.01	-0.68	-44.25
9	913.5	0.78	1.50	0.34	0.34
10	1299.3	0.24	1.02	0.03	5.72
11	1376.6	-4.16	2.53	1.06	6.12
12	1686.2	-1.69	2.20	0.46	11.22
13	3629.6	0.05	0.97	0.21	6.09
14	4978.4	0.12	-0.23	-0.12	10.19
15	2425.8	0.24	-0.69	0.28	9.20
16	7327.6	-0.06	-0.002	0.01	2.38
17	1461.3	0.10	0.03	-0.14	0.91

Positive incremental loads in **bold**

In order to better visualize the spatial distribution of the parameters and the trend of the incremental loads over the whole watershed, four thematic maps (one for each parameter) were compiled using the median values of the ECAs and are presented in Figure 6.7. It can be observed that the WQ generally tends to improve towards the river's mouth (North) and that the region around stations 5 through 7 has the worst WQ. The total ammonia nitrogen and nitrate tend to improve downstream from the Belo Horizonte region, but it also appears to cause the consumption of oxygen. It is worth mentioning that the lower part of the watershed is also a region with more intense agriculture and that the presence of pollutants can be more

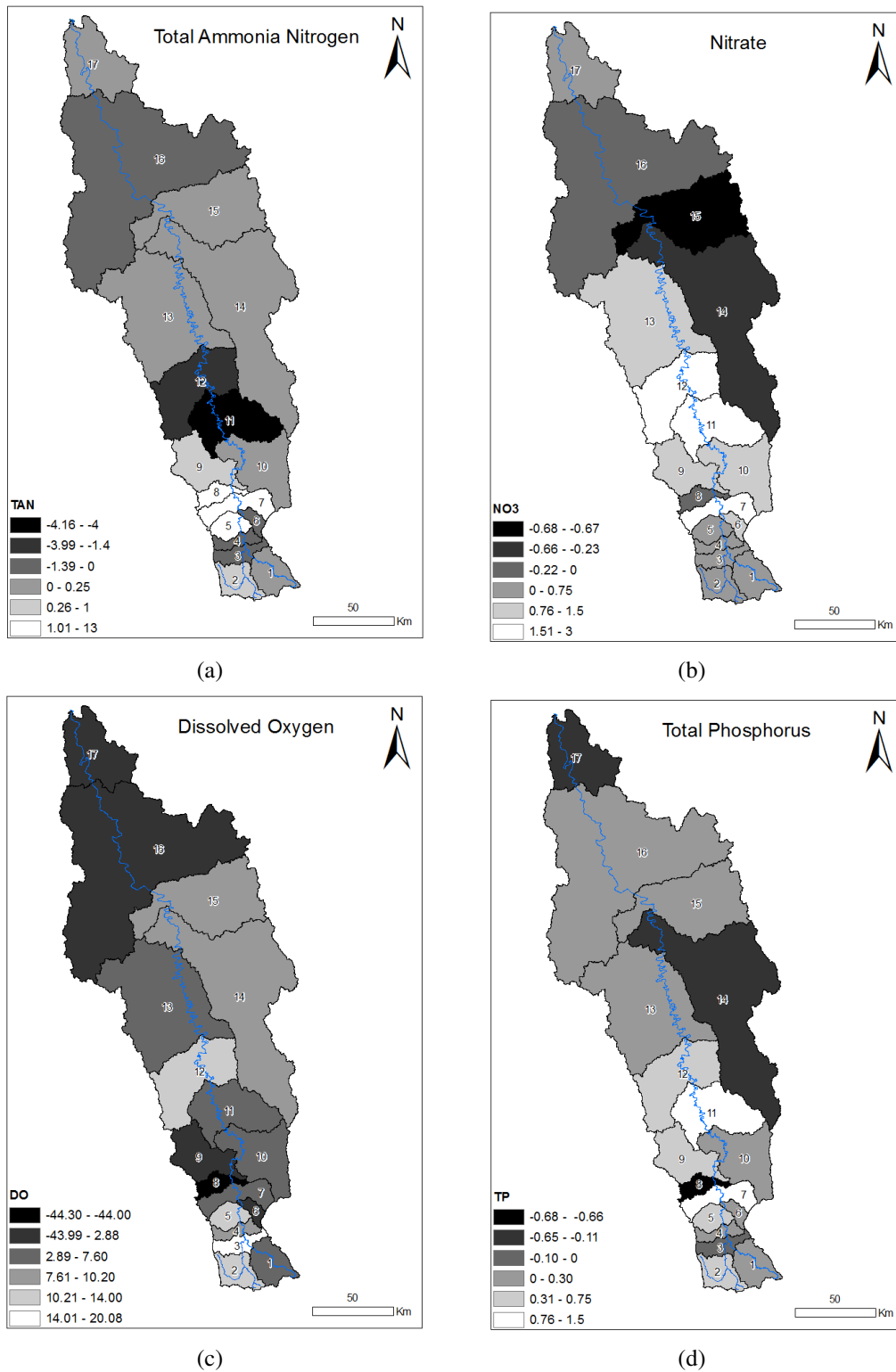


Figure 6.7: Incremental loads per unit area of (a) TAN, (b) nitrate, (c) dissolved oxygen and (d) total phosphorous in kg / km² day

attributed to agricultural practices and non point processes and not sewers (although those are still present).

6.5.3 Correlation between LULC and WQ parameters

Pearson's correlation was employed to evaluate the linear relation between each of the LULC classes with each WQ parameter and considering all area types (whole ECAs and riparian buffers). The analysis of Table 6.5 revealed a number of observations described below. These are organized by parameter.

Total ammonia nitrogen:

- Positively correlated with the proportion of agriculture and pasture (AP); significant for the ECAs and the 150 m and 300 m riparian buffers.
- Negatively correlated with Forest (F_t) but only significant for 150 m.
- Strongly correlated positively with the urban class (U_b) for three buffers (except 150 m) and for the ECAs. The relationship becomes stronger with the distance to the stream (the effect of the capital city is possibly responsible for this since larger areas incorporate more urban areas).
- Although the correlations with riparian forest are all negative, none are significant.
- No significant correlation ($p \leq 0.05$) is observed with the remaining classes.

Dissolved oxygen:

- DO is negatively correlated with (AP) for the 150 m buffer (all other areas were not significant).
- Riparian forest (R_f) is positively and significantly correlated with all riparian buffers and with the ECAs, showing its role as a sink and contributes to water quality.
- Forest (F_t) plays the same role as R_f but the correlations are only significant for the 150 m, 300 m and 600 m buffers.
- The Urban class (U_b) is a consumer of oxygen, being negatively and significantly correlated with three buffer areas (except 150 m) and for the ECAs.
- No significant correlations were observed with the other classes.

Tabela 6.5: Pearson's correlation between the WQ parameters and the LULC classes

Zone ¹	Parameter ²	Pearson's correlation ³							
		G_t	S_a	AP	U_b	TP	RM	R_f	F_t
150 m	TAN	-0.331	+0.100	+0.650**	+0.433	+0.242	-0.044	-0.318	-0.531*
	OD	+0.205	-0.331	-0.521*	-0.378	-0.018	+0.382	+0.563*	+0.651**
	NO_3^-	+0.286	+0.562**	-0.041	+0.088	+0.121	-0.205	-0.282	-0.342
	TP	+0.094	+0.239	-0.102	-0.006	-0.051	+0.039	-0.097	-0.091
300 m	TAN	-0.36	+0.014	+0.489*	+0.577*	-0.248	-0.021	-0.358	-0.423
	OD	+0.190	-0.289	-0.397	-0.512*	-0.006	+0.356	+0.570*	+0.612**
	NO_3^-	+0.310	+0.475	-0.168	+0.113	+0.128	-0.180	-0.320	-0.327
	TP	+0.111	+0.211	-0.072	-0.038	-0.061	+0.085	-0.126	-0.115
600 m	TAN	-0.370	+0.002	+0.203	+0.656**	-0.307	-0.035	-0.379	-0.304
	OD	+0.189	-0.314	-0.220	-0.571*	+0.018	+0.358	+0.563*	+0.559*
	NO_3^-	+0.271	+0.398	-0.355	+0.174	+0.254	-0.174	-0.316	-0.285
	TP	+0.107	+0.149	-0.095	-0.016	+0.065	+0.072	-0.13	-0.102
900 m	TAN	-0.321	-0.006	+0.057	+0.656**	-0.359	-0.037	-0.390	-0.206
	OD	+0.182	-0.319	-0.164	-0.536*	+0.032	+0.373	+0.572*	+0.477
	NO_3^-	+0.264	+0.391	-0.418	+0.231	+0.343	-0.189	-0.300	-0.281
	TP	+0.124	+0.128	-0.105	+0.049	+0.181	+0.067	-0.101	-0.133
ECA	TAN	+0.522*	-0.290	+0.527*	+0.908**	-0.017	-0.171	-0.263	-0.114
	OD	+0.434	-0.033	+0.383	-0.854**	-0.005	+0.492*	+0.513*	+0.407
	NO_3^-	-0.386	+0.451	-0.005	+0.053	+0.242	-0.375	-0.195	-0.188
	TP	-0.026	+0.246	+0.121	-0.231	+0.218	-0.072	+0.010	-0.038

1-Zone: ECA = Exclusive Contribution Area; 150 m–900 m = riparian buffers

2- Nitrate (NO_3^-), Total ammonia nitrogen (TAN), Dissolved Oxygen (DO), Total Phosphorous(TP)

3- **0.01 significance; *0.05 significance

Nitrate and total phosphorous: The only significant correlation observed for nitrate was savanna (S_a) at the 150 m riparian buffer and we could not offer a valid explanation for this relation other than the possible use of savanna as pasture. No significant correlation were observed for total phosphorous.

An increase of total ammonia loads and decrease of dissolved oxygen are normally considered pollution factors and are associated with the presence of either urban areas, agriculture, or pasture. The urban areas appear to have an effect even at larger distances. Conversely, the correlation between pollution (more ammonia, less oxygen) and the AP class weakens with distance.

Natural forested vegetation (R_f and F_t) plays a definite role in improving water quality, acting as sink for the ammonia and as a source for dissolved oxygen. The filtering effect for the total ammonia nitrogen lessens rapidly with distance (at for the F_t class; see Table 6.5).

6.5.4 Multiple Regressions Models

In defining the regression models, the dependent (explained) variables chosen were the unit incremental loads of each ECA for the four WQ parameters, while the independent (explaining) ones were the LULC proportions over the four riparian buffers and the ECAs. For the latter, only the classes with a significant correlation ($p \leq 0.05$) and a correct sign (making sure that the direction of the relation is as would be expected) were included. Furthermore, when a strong correlation was observed between two LULC classes (collinearity), only the one with the stronger correlation was included in the model. Although stepwise regression could have solved colinearity problems we did not adopt this approach for two reasons: 1) our tests with stepwise regression tended to produce a very simple model with but one variable; 2) we wanted the models to pick at least one LULC class as pollution source and one as a pollution sink.

Table 6.6 shows the multiple regression equations, the coefficients of determination, significance, and F test (ANOVA) results for the 11 valid (significant) models. The models are shown for total ammonia nitrogen (TAN), dissolved oxygen (DO) and nitrate (NO_3^-), and for the four riparian buffers and the ECAs. For nitrate, only the riparian buffer of 150 m was retained, the remaining being not significant. Figure 6.8 shows the plots for the two best regression models, both referring to the ECAs.

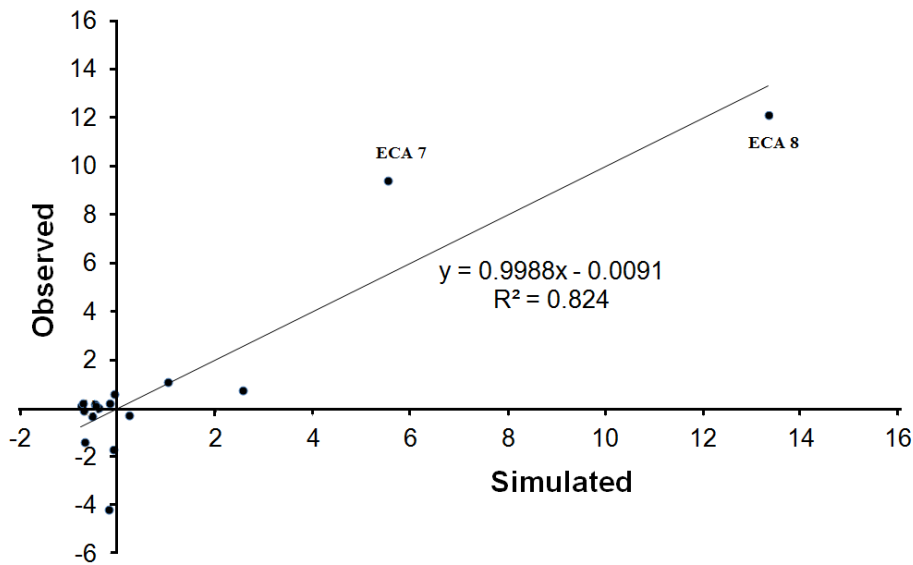
Tabela 6.6: Multiple regression models

Zone	Y	Model	R^2	F	Sig.
150 m buffer	TAN	$-0.638 + 0.21 AP - 0.085 F_t$	0.48	6.48	0.010
	OD	$-3.014 - 0.376 AP + 0.555 F_t$	0.48	6.352	0.011
	NO_3^-	$0.120 + 0.035 S_a$	0.31	6.914	0.019
300 m buffer	TAN	$-3.631 + 0.152 AP + 0.168 U_b$	0.41	4.9	0.024
	OD	$-5.091 - 0.389 U_b + 0.447 F_t$	0.45	5.792	0.015
600 m buffer	TAN	$-0.917 + 0.244 U_b$	0.43	11.325	0.004
	OD	$-1.191 - 0.564 U_b + 0.35 R_f$	0.48	6.539	0.010
900 m buffer	TAN	$-0.939 + 0.262 U_b$	0.43	11.335	0.004
	OD	$+0.025 - 0.481 U_b + 3.247 R_f$	0.42	5.086	0.022
ECA	TAN	$-1.15 + 0.193 U_b + 0.016 AP$	0.83	32.9	0.000
	OD	$5.137 - 0.557 U_b + 4.427 R_f$	0.77	23.639	0.000

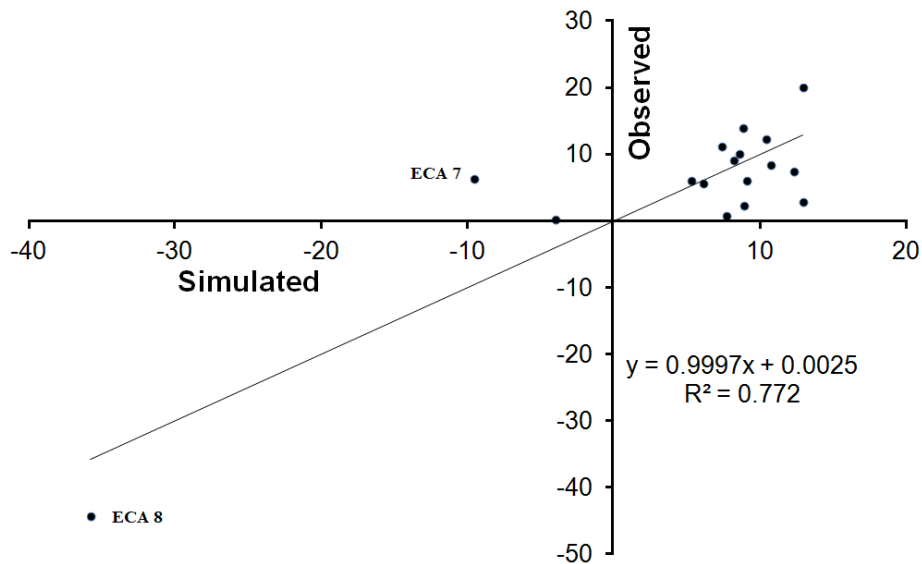
Y - Dependent Variable; X - Independent Variable: U_b - Urban, F_t - Forest, R_f - Riparian Forest, S_a - Savanna and AP - Agriculture and Pasture.

For all the models shown in Table 6.6, we were able to reject the null hypothesis (H_0) stating that the betas are equal to zero with a 0.05 significance.

For the TAN parameter, based on Pearson's correlations, we selected the three LULC classes with a significant correlation for inclusion in the model: AP and U_b as sources of pollution and F_t as sink. However, depending on the zone studied (one of the riparian buffers



(a) Total Ammonia Nitrogen



(b) DO

Figura 6.8: Plots of the two best regression models for the exclusive contribution areas (ECAs)

or the whole ECA), the relative weights of these LULC classes vary widely. The following points are tentative interpretations.

- For the 150 m riparian buffer, the agriculture and pasture class (*AP*) were the only pollution source used, with forest as a sink showing the importance of natural vegetation in acting as a diffuse pollution sink. Urban (*U_b*) was not included, simply because it was poorly represented within the 150 m buffer. An R^2 of 0.48 was significant but cannot be considered strong.

- For the 300 m buffer, only the pollution sources (AP and U_b) were selected by the model. As for the previous model, an R^2 below 0.5, even if significant, is still rather low. Because TAN is mainly associated with sewers and point pollution, this might explain why F_t did not enter into the model.
- For both the 600 m and 900 m buffers, only U_b was selected, again this might reflect the fact that TAN is mainly associated with point source pollution, for which the filtering effect of vegetation does not have a significant effect.
- The ECA brought the AP class as a pollution source. Because at this scale the amount of AP is much larger than U_b , it was included in the model. Still, its weight in the model is 12 times smaller than U_b , somewhat confirming the point pollution explanation.
- The ECA model shows a very strong coefficient of determination of 0.83, being almost twice the average value obtained with the buffer zones.

The results for OD are quite different from the TAN models. Although these models usually include the same source classes (AP and U_b), unlike TAN models, they always include a sink class (F_t or R_f), which are always positive and with almost a stronger absolute weight than AP or U_b (except for the 600 m buffer). In the extreme case this relation is about 9:1 in the case of the ECA (see Figure 6.9).

- The 150 m buffer included the AP class as a source whereas the U_b class was chosen for the other modeled zones. F_t acted as a sink.
- The 300 m buffer included the U_b class as a source. F_t and RF acted as sink (RF riparian forest being linearly correlated with F_t , it was not included in the model).
- The 600 m, 900 m and ECA zones all used the same two LULC classes: U_b and R_f but the latter two give a much stronger weight to the R_f as a sink.
- As for the TAN parameter, the coefficient of determination of 0.77 for the ECA is much stronger than that for the riparian buffer zones, which average 0.46.

The TAN models only included vegetation for the 150 m model and we argue that this is because TAN is mainly associated with point pollution sources. Nitrogen in ammonia form tends to be more associated with point pollution (sewers). Riparian vegetation might be effective as a sink but only for non-point sources.

The nitrate model was only significant for the 150 m buffer zone and only achieved $R^2 = 0.31$. It identified savanna (S_a) as a source of pollution, which is not consistent with a reasonable interpretation. For these reasons, we did not consider the nitrate model valid.

No valid model was obtained for total phosphorous and we suspect that this might be due to the fact that a significant portion is fixed in the sediments and not measured properly. This portion might be released at a later stage, making a time-based analysis difficult (Arheimer e Liden, 2000; Munson, 2015). A similarly poor result was obtained by Maillard e Santos (2008) in the same watershed.

For both the TAN and DO parameters, which produced valid models, the coefficient of determination was far superior for the whole ECA than for the riparian buffers (see Figure 6.9). And for the latter, no clear trend was observable other than that the R^2 are rather constant.

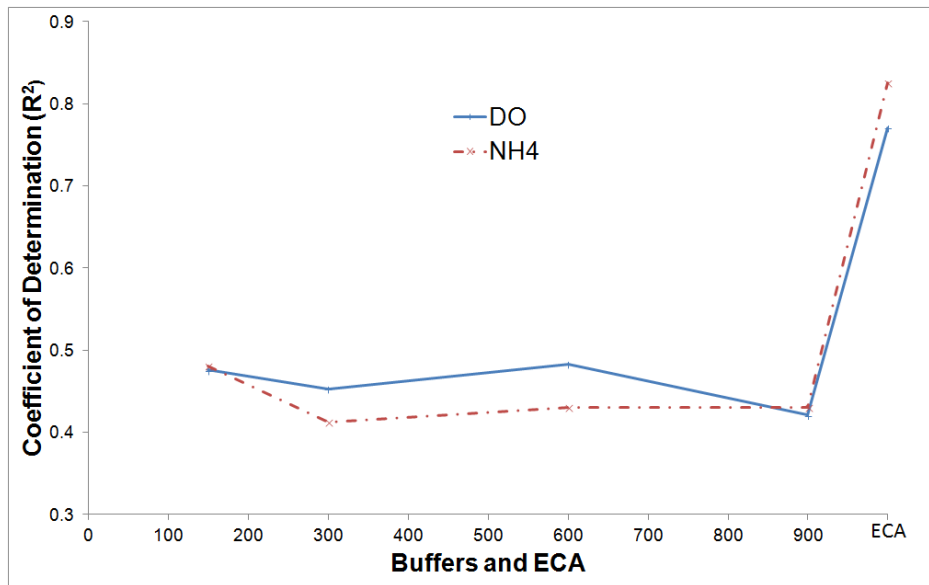


Figura 6.9: Coefficient of Determination

6.6 Conclusions

Our study demonstrated the feasibility of establishing a relationship between water quality (WQ) and land use / land cover using incremental loads per exclusive contribution areas (ECA) instead of using WQ parameter concentrations. This enabled an understanding of the spatial distribution of these parameters within the watershed by isolating what happens between two WQ stations from the events upstream. This was also consistent with findings from Ouyang *et al.* (2010).

We proposed an original approach for selecting WQ data, based on the rainy season period, applicable to places where there is a distinct wet season. This approach analyses how the discharge hydrogram progresses in order to determine a starting date for the rainy season. This is especially consistent with studies focusing on diffuse pollution.

A series of valid models were produced for the total ammonia nitrogen (TAN) and for dissolved oxygen (DO) for four riparian buffers (150 m, 300 m, 600 m, 900 m) and for the ECA. For both parameters, the ECA yielded the best results, with $R^2 = 0.83$ and $R^2 = 0.77$ for TAN and OD, respectively. This is also consistent with findings from Meynendonckx *et al.* (2006) and Li *et al.* (2015).

Out of eight LULC classes, only four entered into the different models, and never more than two for any particular model (TAN and OD). A similar reduction of independent variables was obtained by Li *et al.* (2015), who generated models with only one LULC class from the eight classes they originally considered. This approach was used mainly because of the existing collinearity between the LULC classes, which would have weakened the models. On the positive side, this made it possible to identify the main classes responsible for the variations in water quality. The agriculture and pasture class (AP) and the urban class (U_b) entered the model as increasing (positive beta) the total ammonia nitrogen and decreasing the dissolved oxygen (DO) and are being considered both as pollutant sources and as worsening the water quality. Conversely, forest (F_t) and riparian forest (R_f) had a negative impact on TAN and a positive one on OD, and acted to improve the water quality. This is consistent with previous findings by Maillard e Santos (2008), also for the Velhas River.

Our study attempted to provide physical and chemical explanations for some problems inherent in the measurement of water quality parameters, for instance, the transformations of ammonia into nitrite and nitrate, which occurs over time and might bring about a delay between the ammonia input and the measurements of its derivatives downstream (Chestrikoff *et al.*, 1992). This depends on local conditions of temperature, pH and OD, that vary with time (Von Sperling, 2005; 2007). Likewise, there are difficulties associated with the measurement of phosphorous, which gets fixed in the sediments and might be released at a later time. We think these explanations can partially explain similar problems experienced by Maillard e Santos (2008) for their rainy season models.

Our study outlined the necessity for the incorporation of other variables not linearly correlated with the simple land use land cover classes. The latter are, to a certain extent, expected to be correlated, since their sum is always equal to 100%. Ideally, it would be desirable to create a specific variable that would enable the unequivocal separation between point and non-point pollution sources. One possibility we are considering is to use the relation between the pollutant loads in the dry and rainy seasons to estimate the point pollution using the fact that non-point pollution is strongly reduced when no precipitation occurs.

6.7 acknowledgements

The authors would like to acknowledge the following institutions: the *Instituto Mineiro de Gestão das Águas* (IGAM) for the water quality data, the *Agência Nacional de Águas*

(ANA) and the *Companhia Energética de Minas Gerais* (CEMIG) for the discharge data and the *Instituto Estadual de Florestas* (IEF) for providing the RapidEye scenes from which the LULC map was based. This article is an inherent part of L.M. Oliveria's PhD thesis under the supervision of P. Maillard, with E.J. de Andrade Pinto acting as co-supervisor.

6.8 Bibliographic References

The bibliography in page 134.

Capítulo 7

ARTIGO 3 : Application of a land cover pollution index to model non-point pollution sources in a Brazilian watershed

7.1 author

Líliá Maria de Oliveira¹ and Philippe Maillard² and Éber José de Andrade Pinto³

7.2 Abstract

Non-point and point source pollution are a water quality problem in most parts of the World. Many studies have used the land use and land cover (LULC) to model non-point pollution sources. In this article we evaluate the relationship between LULC and five water quality parameters using different zones of analysis: riparian buffers (B) and exclusive contribution areas (ECA). The five parameters are nitrate, nitrite, total ammonia nitrogen, total phosphorous and dissolved oxygen. Analyses were performed on riparian zones of different widths and ECAs to verify if the effect of the land cover on the water quality of the stream decreases with the increased distance. The urban and agricultural/pasture categories of LULC were characterized as pollution sources while vegetation (Forest and Riparian Forest) as pollution sink (filter). We proposed a Land Cover Pollution index (LCPI) which is a ratio between source and sink to substitute the individual LULC categories. The source and sink categories were selected considering our knowledge of the LULC relationship and

¹CEFET-MG, Av. Amazonas, 5354, Belo Horizonte, Brasil Tel.: +55-31-33770361 email: li-lia@deii.cefetmg.br

²Universidade Federal de Minas Gerais, Instituto de Geociencias, Av. Antônio Carlos, 6627, Belo Horizonte, Brasil; Tel.: +55-31-33770361; email: philippermaillard@yahoo.com.br

³Universidade Federal de Minas Gerais, Escola de Engenharia, Av. Antônio Carlos, 6627, Belo Horizonte, Brasil; Tel.: +55-31-33770361; email: eber.andrade@cprm.gov.br

the sign of the Pearson's Correlation Coefficient. The LCPI varied between 0.11 in the 150 m buffer (higher filter effect) and a maximum of 27.99 (largest source effect). Additionally, we transformed the water quality data in incremental loads per unit area so that each station could be considered independent from the others and be compared between themselves independently of the contribution area they represent. Our method also included only data from the rain season, the beginning of which was determined using the hydrograms of discharge data. Results indicate that the index is better than the individual LULC classes to explain this relationship, especially in riparian zones. In 12 out of 15 models the coefficient of determination (R^2) increased by 11 to 155%, when we use the index instead of the LULC classes. We believe that these good results can be attributed to the LCPI but also to the special processing of the water quality data making it more sound for statistical processing.

7.3 KeyWords

Water quality, Land use/ land cover, Diffuse pollution, LCPI Index, Brazil, Regression.

7.4 Introduction

The monitoring of water quality (WQ) parameters in Brazil is performed with various objectives, one of which is the determination of the quality of a stream for the classification of its possible use. The parameter concentrations are associated with pollutants from both point and non-point sources that can reach the stream directly through sewers or through natural processes such as leaching in the soil or runoff (McDowell *et al.*, 2001; Quinn e Stroud, 2002; Łowicki, 2012). These data are useful for comparing the quality of a stream with pre-established standards for the type of use to which it is destined. However, they are insufficient for comparing watersheds in terms of pollutant loads (Quilbé *et al.*, 2006). Loads are considered better for comparing different watersheds and for establishing the relationship between pollutants and other physical factors such as land use (Ouyang *et al.*, 2010).

Sewers are normally the main source of nitrogen (N) in organic form or as ammonium. With time, the ammonia is transformed into nitrite and then nitrate. Although this can occur before the sewer reaches the stream, it can also happen in the stream, thus causing the latter forms (nitrite and nitrate) to be found downstream from the early forms (Chesterikoff *et al.*, 1992). The conversion of ammonium into nitrate consumes large quantities of oxygen (4.3 mg/L of O₂ are needed to convert 1 mg/L of ammonium) a process that can take up to many kilometers depending on the temperature, pH, dissolved oxygen (DO), and turbulence (Von Sperling, 2007).

The phosphorus (P) found in streams can be either in organic or inorganic form, in the water or in the sediments. Sedimentation of organic P causes it to be removed from the

water as a measurable parameter and be fixed in a particle form (Golterman, 2007). Conversely, organic P in sediment can migrate to dissolved inorganic P causing an increase in concentrations in the water (Gibson e Meyer, 2007).

Agricultural activities are important sources of P. Even though supplies of P from urban runoff are twice as much per unit area than from agriculture (Von Sperling, 2007), the much higher proportion of area dedicated to agriculture generally makes it the prime source of P (Grizzetti *et al.*, 2008). However, the migration of P from sediments to water and vice-versa can make this relation difficult to verify. The understanding of the transformation between the different forms of these pollutants is essential for establishing a causal relation between land use and land cover (LULC) and water quality (WQ), especially when considering the separation between point- and non point-pollution sources (Grizzetti *et al.*, 2008).

Land use can contribute in polluting a stream as a non-point source but it can also act as a filter and keep pollutants from entering the stream, by leaching or by biological fixation. Vegetation, especially in the riparian zone, can play an important role as a filter (Sliva e Williams, 2001; Poor e McDonnell, 2007). WQ parameter loads can be related to LULC by trying to identify which categories act to reduce pollution and which increase it. Examples of this type of approach can be found throughout the specialized literature. The agriculture/pasture and urban classes are usually positively correlated with various water quality parameters while forest stands as a mitigating factor (Buck *et al.*, 2004; Salvia-Castellví *et al.*, 2005; Poor e McDonnell, 2007; Li *et al.*, 2015; de Oliveira *et al.*, 2016). Using these principles, we can consider all LULC classes to be either pollution source, sink or neutral. Disregarding neutral classes (usually defined by a non-significant correlation) all other LULC categories either contribute to pollute or mitigate the effects of pollution. The inclusion of LULC classes with a non-significant correlation often results in weakening multivariate models and many authors end up excluding them from their models (Bu *et al.*, 2014; Li *et al.*, 2015).

The fact that many authors used the classification of LULC classes as either source or sink brought us to take one more step forward and take advantage of this dichotomous relationship to create an index that captures both ends of the effect of LULC on water pollution. By dividing sources by sinks we can expect to propose univariate models to try to explain the LULC/WQ relationship.

Another aspect generally taken into consideration is the determination of the area of influence of the land use and land cover (LULC). Studies have used the whole watershed, the exclusive contribution areas (ECA), or bands of varying width on either side of the stream - generally called buffers (B) - or even combinations of the three (Johnson *et al.*, 1997; Buck *et al.*, 2004; Chang, 2008; Li *et al.*, 2015).

In this article, we evaluate how the LULC affects the changes in loads of total ammonia nitrogen (TAN), nitrate (NO_3^-), nitrite (NO_2^-), total phosphorous (TP), and dissolved oxygen (DO) in the Velhas River in Minas Gerais, Southeast Brazil. To do so we have considered

exclusive contribution areas (ECAs) for each monitoring station as well as four riparian buffers of different width. Concentrations were transformed in "incremental loads" per unit area to avoid a cumulative effect from stations situated upstream from each other. Finally we also proposed and test a Land Cover and Pollution Index (LCPI) that combines pollution sources and sinks to substitute the individual LULC categories.

7.5 Material and Methods

7.5.1 Study Area

The Velhas River Watershed covers an area of 27,887 km^2 in the central part of the State of Minas Gerais (Figure 7.1) and is home to 4.4 million people, most of whom live in the capital city of Belo Horizonte and its metropolitan area (CBH, 2015).

Using satellite imagery, the watershed LULC was classified into the nine following categories with their respective percent areas: woody savanna (Sa, 26.4 %), grassland (Gl, 21.3 %), agriculture and pasture (AP, 25.2 %), forest (Ft, 15.3 %), riparian forest (Rf, 0.8 %), rock outcrop and mines (RM, 4.2 %), tree plantations (TP, 3.3 %), urban (Ub, 2.9 %), and water, clouds and other (WC, 0.6 %) (Figure 7.1).

The land use and land cover (LULC) mapping was done using image segmentation and classification of a mosaic of 72 RapidEye image tiles from 2010. RapidEye is a constellation of five Earth observing optical satellites with a ground resolution of five meters and five spectral bands ranging from the blue to near-infrared wavelengths (BlackBridge, <http://blackbridge.com/rapideye>). The classification was performed using Trimble eCognition Developer 8.4 package.

Although woody savanna, grassland and pasture were labeled separately we are aware that confusion can remain between these classes as pasture can grow shrub given its appearance of woody savanna while grassland can be used for stock grazing. This resulted in creating unnatural straight line limits in the classified mosaic due to different months of image acquisition. Other consequences of these possible confusion will be discussed in the results section.

Near the source of the Velhas River, the terrain is hilly to mountainous with narrow valleys, gorges and streams of low sinuosity and high structural control. Downstream, towards the mid- and low-watershed, the river meanders more and the terrain becomes smoother with rolling hills and a wide and flat river valley with terraces. The river source is situated at an altitude of 1500 m whereas its mouth is at 478 m (IGAM, 2015).

The climate is semi-humid tropical of altitude, with a well defined wet (summer) and dry (winter) season. Yearly precipitation decreases from the source downstream, from an average of 1600-1700 mm, to about 1100 mm near the mouth of the river (Pinto et al., 2011).

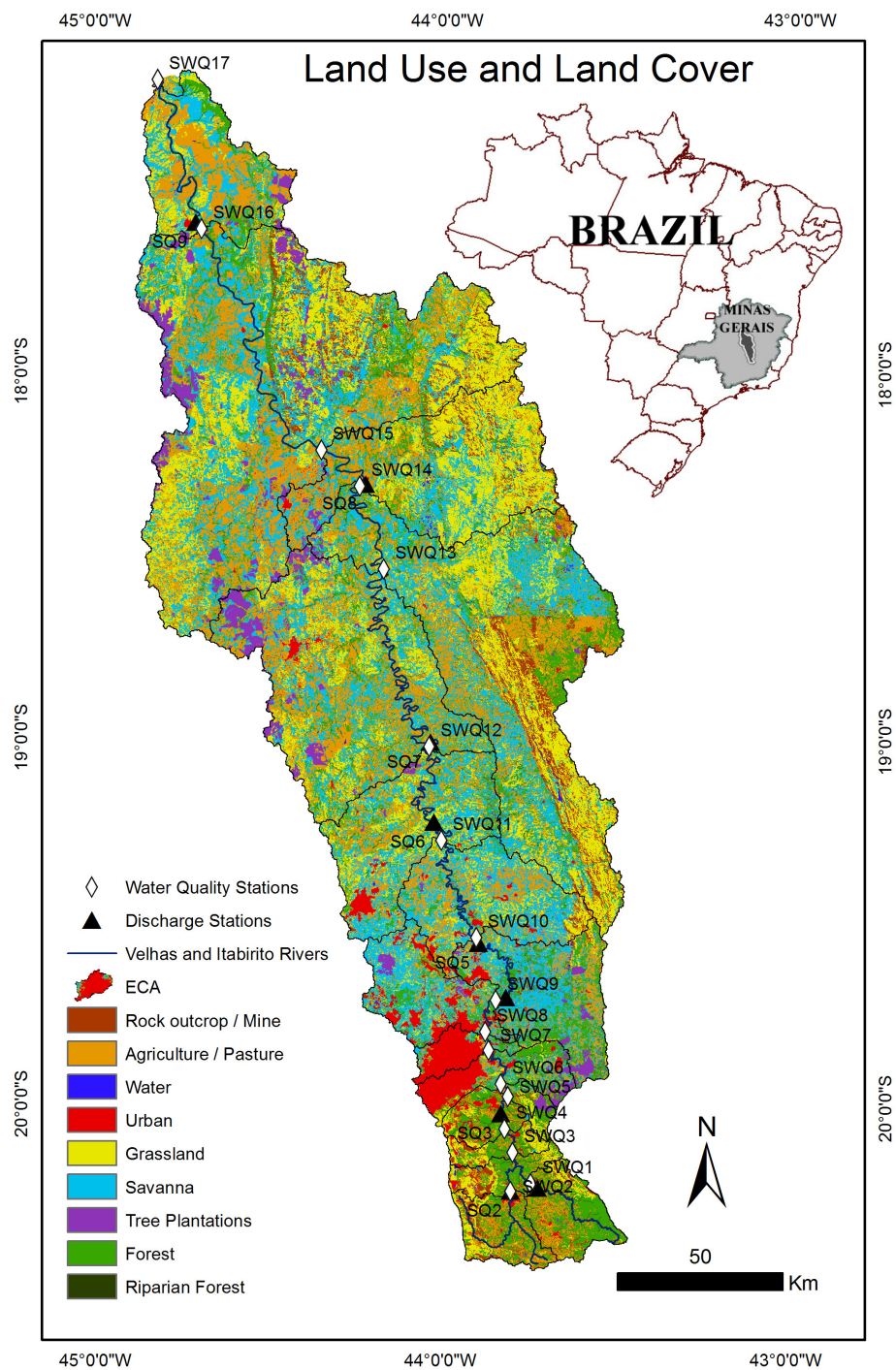


Figura 7.1: Map of the Velhas River Watershed showing the LULC categories, the exclusive contribution areas (ECA) and the monitoring stations (discharge and quality).

The rainy season normally starts in October and extends until March.

The minimum discharge (Q100%) of the Velhas River varies between 1.39 m^3/s and 48.8 m^3/s from the source (closest station) to its mouth (Figure 7.2).

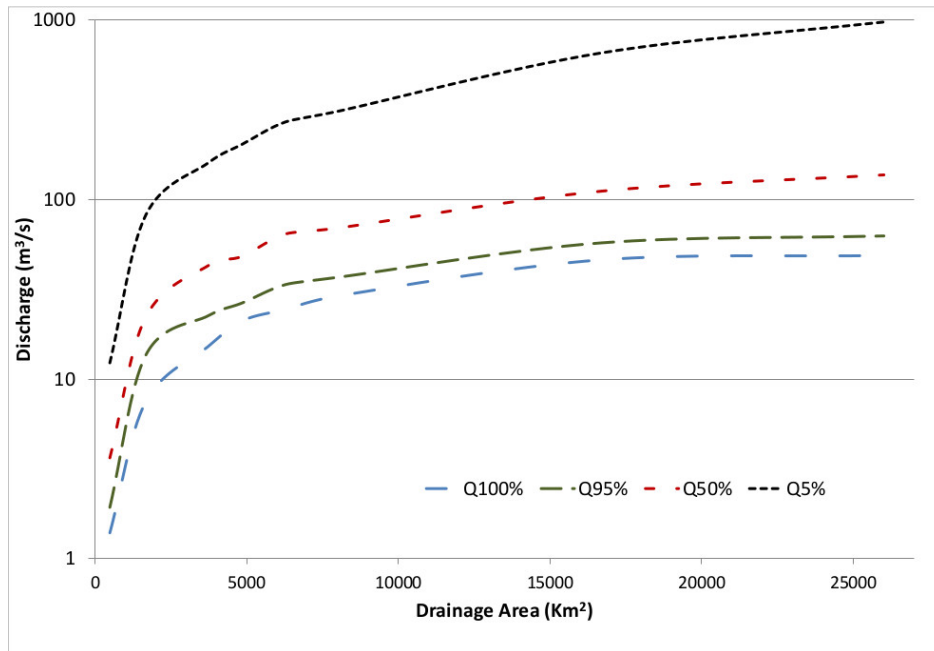


Figura 7.2: Discharge (Q100%, Q95%, Q50% and Q5% m^3/s) and drainage area of the Velhas River Watershed (km^2).

7.5.2 Water quality and water discharge data

Like other studies have demonstrated, (Stutter *et al.*, 2008; Outram *et al.*, 2014; 2016), we opted for using only data from the rainy season when runoff occurs and create the strongest effect of non-point pollution. The beginning of the rainy season was defined by finding a significant break upward in the daily hydrogram (Figure 7.3). While other authors use a fixed period defining the rain season, we found that this approach is more realistic. The end was set at 31 March regardless of the year. Unlike the beginning of the rainy season, the end is progressive and not easily pinpointed so that an arbitrarily selected date does not have any strong effect on the data.

The daily mean discharge was obtained by averaging the daily monitoring carried out at 7 am and 5 pm. Data from the five parameters considered in our study (total ammonia nitrogen, nitrite, nitrate, total phosphorous and dissolved oxygen) were downloaded from the Minas Gerais water agency (Instituto Mineiro de Gestão de Águas - IGAM). The data was collected every three months between 2001 and 2007 then monthly from 2008 to 2013. The concentrations of the five parameters were then converted into loads for the 17 water quality monitoring stations (Figure 7.1) using the following relation (Eq. 7.1).

$$L = K * C * Q_{WQ} \quad (7.1)$$

Where L is the load (kg/day), K is a unit conversion constant (86.4), C is the concentration (mg/l) and Q_{WQ} is the discharge (m^3/s).

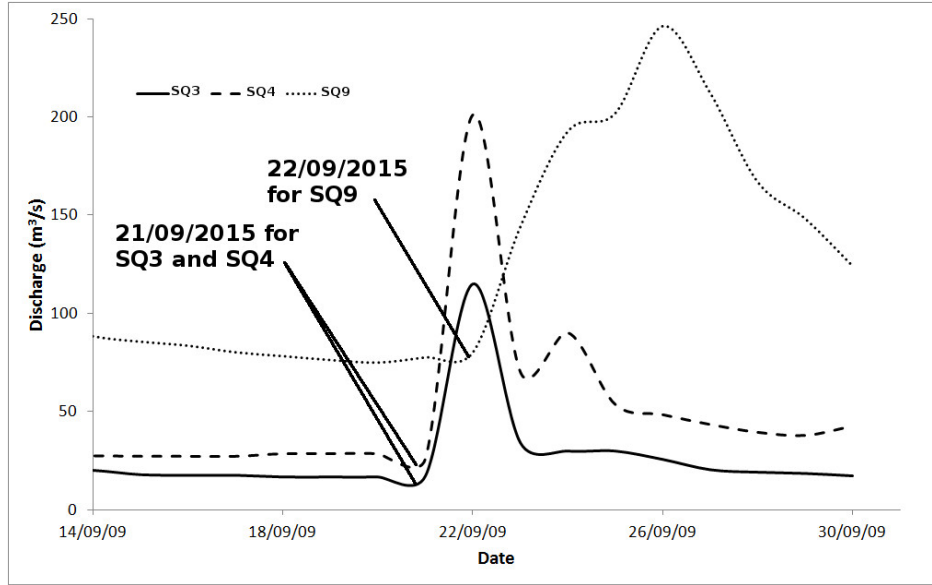


Figure 7.3: Illustration of how the starting date of the rainy season was selected from the daily hydrogram of each station.

Unfortunately, the locations of the water discharge stations (S_Q) and the WQ stations (S_{WQ}) did not coincide, so we calculated a compensation factor ($Cf_{WQ/Q}$) to regionalize the daily mean discharge data to correspond to the coordinates of the WQ stations (Equation 7.2, (Harvey et. al. 2012)). Table 7.1 shows these compensation factors for all water quality stations (S_{WQ}). Then the discharge data was multiplied by its respective compensation factor (Equation 7.3).

$$Cf_{WQ/Q} = \frac{A_{WQ}}{A_Q} \quad (7.2)$$

$$Q_{WQ} = Q \times Cf_{WQ/Q} \quad (7.3)$$

Where $Cf_{WQ/Q}$ is the area compensation factor (dimensionless), Q_{WQ} is the daily discharge at the S_{WQ} station (m^3/s), Q is the daily discharge at the S_Q station (m^3/s), A_{WQ} and A_Q are the sub-watershed areas for the S_{WQ} station (km^2) and the S_Q station (km^2) respectively. The left side of the flowchart in Figure 7.4 illustrates these successive steps.

Our approach required that the data from each water quality monitoring station be analyzed independently so that the loads from each station are not directly affected from the water quality (WQ) of stations located upstream. This was performed in three separate steps. First

Tabela 7.1: Area compensation factor $Cf_{WQ/Q}$ and unit incremental loads (median, kg/km²·day) for all exclusive contribution areas (ECA).

ECA	Area (km ²)	$Cf_{WQ/Q}$	U nit Incremental Loads (UIL)				
			TAN	NO_3^-	NO_2	TP	OD
1	549.9	1.12	0.20	0.21	0.006	0.07	7.56
2	467.1	1	0.62	0.31	0.028	0.35	14.00
3	288.6	0.72	-0.29	0.20	0.010	-0.05	20.08
4	207.4	0.90	-0.25	0.52	0.024	0.03	8.47
5	307.8	1.10	1.11	0.60	0.099	0.39	12.44
6	143.3	1.17	-1.36	0.76	-0.049	0.27	2.89
7	533.0	1.48	9.43	2.20	0.314	1.17	6.39
8	257.7	0.74	12.11	-0.01	1.101	-0.68	-44.25
9	913.5	0.98	0.78	1.50	0.343	0.34	0.34
10	1299.3	1.02	0.24	1.02	-0.146	0.03	5.72
11	1376.6	1	-4.16	2.53	0.340	1.06	6.12
12	1686.2	1	-1.69	2.20	-0.269	0.46	11.22
13	3629.6	0.70	0.05	0.97	0.000	0.21	6.09
14	4978.4	1.0	0.12	-0.23	-0.002	-0.12	10.19
15	2425.8	1.20	0.24	-0.69	-0.004	0.28	9.20
16	7327.6	1.01	-0.06	-0.002	0.001	0.01	2.38
17	1461.3	1.07	0.10	0.03	0.024	-0.14	0.91

Positive incremental loads in **bold**

an exclusive contribution area (ECA) was created for each station that excludes all areas of the other stations upstream (Figure 7.4 and 7.5). Then, for each station (except for the two first stations, located at the source) the loads from the station upstream were subtracted so that each station represents only its ECA. These new loads (IL_x) are called "incremental" since they represent the difference from the previous station upstream (Equation 7.4).

$$ILx_{i,j} = Lx_{i,j} - Lx_{i-1,j} \quad \text{for } i=1,\dots,17 \quad j=1,\dots,12 \quad (7.4)$$

Where x is one of the five parameters, i represent the WQ stations and j the hydrological year. Finally, to standardize these incremental loads, we divided each one by the ECA to produce a "unit incremental loads"(UIL) in kg km⁻² day⁻¹ (Eq. 7.5).

$$UILx_{i,j} = ILx_{i,j}/ECA_i \quad \text{for } i=1,\dots,17 \quad j=1,\dots,12 \quad (7.5)$$

where ECA_i is the exclusive contribution area of station i .

With this procedure, we aimed at verifying whether the land use and land cover (LULC) between two monitoring stations had an improving or deteriorating effect on the WQ parameters, independently of scale (Table 7.1). The procedure resulted in creating 17 independent stations of water quality (WQ) data of "unit incremental loads". This process can generate negative load values that can be interpreted as a WQ improvement.

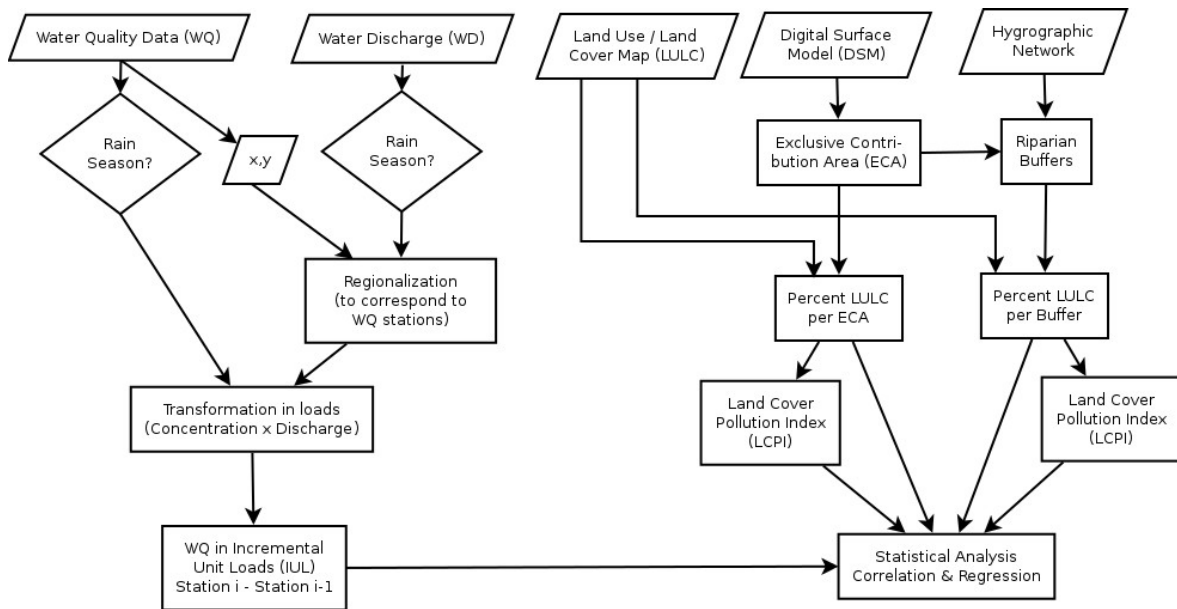


Figure 7.4: Flowchart of the data processing steps. The left hand side shows how the water quality (WQ) and water discharge (WD) data are processed to be transformed in unit incremental loads (UIL). Because the discharge stations do not correspond to the WQ stations a regionalization of the WD was performed using the stations coordinates (x,y). The right hand side shows the processing of the spatial components (LULC, DSM, hydrological network) to generate the four types of analyses areas (three buffers and one exclusive contribution area).

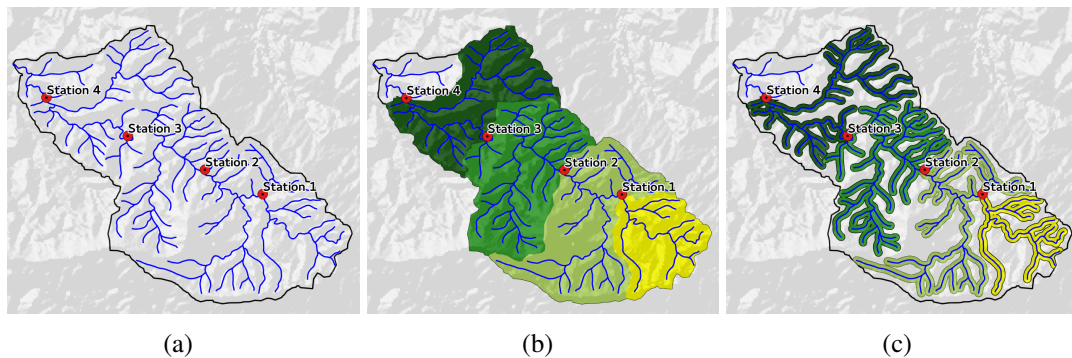


Figure 7.5: Illustration of how the exclusive contribution areas (ECA) and the river edge strips (buffers) were created from the digital surface model: (a) watershed with hydrography and WQ gauging stations, (b) determination of the ECA, and (c) strips of varying width (buffers) on either side of the stream.

Then, successive strips of varying width (buffers) were defined on either side of the stream in an effort to determine the strength of the relationship between the land use and land cover (LULC) and water quality (WQ) with the increase in hydrological distance ¹. Figure 7.4 and 7.5 illustrate this process. A digital surface model (DSM) with a ground cell

¹Hydrological distance is intended to be the length of the path that a raindrop would follow on the surface of the terrain to reach the outlet of the basin Nardi *et al.* (2008); Van Nieuwenhuysen *et al.* (2011)

resolution of 30 m from the shuttle radar topographic mission (SRTM) was used to generate these areas.

Since we only had LULC data for one year and we considered it constant for the whole period, the median loads of the five WQ parameters were calculated for each hydrological year and for each station. This generated 17 median values (one for each station) for each parameter. The median was selected as the measure of central tendency because it reduces the weight given to extreme values (Miller, 1993).

7.5.3 Land cover pollution index

The watershed LULC was classified into nine categories that can be classified as either contributing to improve (Sink) or deteriorate (Source) water quality (WQ). Based on this, we constructed a "Land Cover Pollution Index"(LCPI) of the ratio between Source and Sink (Equation 7.6).

$$LCPI = Source/Filter \quad (7.6)$$

7.5.4 Statistical Analysis

The Pearson correlation coefficient was used for two distinct reasons. First to gain knowledge on what to expect as to which LULC variables are more likely to be correlated with the WQ parameters and to verify the sign (\pm) of the relation which indicates which variables (categories) contributes as a Source or a Sink of pollution. Secondly to verify the degree of linear correlation between the independent variables that may compromise the multiple regression models. A statistical significance of $p < 0.01$ and $p < 0.05$ levels (2-tailed), was used to keep or reject the individual LULC classes.

Furthermore, the relation between land use and land cover (LULC) and water quality (WQ) was analyzed for the five parameters using multiple regression models. The models take the form $Y=b_0 + b_1 * X_1 + \dots + b_n * X_n$ where X represent the independent (explaining) variables and Y is the dependent (explained) variable. The $b_{1;2;\dots;n}$ are the beta weights for each LULC variable used in the model. These models were built considering the ECAs and the different riparian buffers as well (B150, B300, B600 and B900). Regression models were calculated using both the individual LULC categories and the Land Cover Pollution Index (LCPI). Lastly, an analysis of variance was performed to test the null hypothesis that the beta weights are null ($H_0 : b_{1;2;\dots;n} = 0$; $H_1: b_{1;2;\dots;n} \neq 0$). A significance level of 5% was adopted to reject the null hypothesis. The residuals were also evaluated to check their independence and normal distribution. These statistical analysis also gave us a means to evaluate the performance of the index we created.

7.6 Results and Discussion

The unit incremental loads (UIL) for the five water quality (WQ) parameters were calculated between 2001 and 2013, for the 17 water quality monitoring stations, according to Eq. 7.1 and were analyzed sequentially in the upstream - downstream order as presented in the box plots of Fig. 7.6.

UILs of total ammonia nitrogen (TAN) vary widely between stations 5 and 12 and are very high for stations 7 and 8 because of their location near Belo Horizonte, which contributes the most to point pollution sources through sewers. Likewise, the dissolved oxygen in station 8 stands out as being the only negative value also explainable by the presence of sewers, causing a high consumption of oxygen through a series of chemical processes, one of which is the conversion of ammonia (TAN) into nitrite (NO_2^-) and then nitrate (NO_3^-). The reduction of the Dissolved Oxygen (DO) in station 8 also explains the reduction of nitrate.

Unit incremental loads (UIL) of nitrate are almost always positive for the first 13 stations (except station 8 with a small negative value), showing a gradual increase of nitrate downstream up to station 9. We attribute this to the time delay needed for the conversion of ammonia into nitrate. Since the successive stations are downstream from one another, this conversion time translates into distance. After station 13, the Paraúna River with an average discharge of $72.4 m^3/s$ merges with the Velhas River improving significantly its quality. The UIL of nitrite and ammonia have similar behavior up until station 10. After this, nitrite becomes more variable with no consistent pattern observable. Downstream from station 10 the release of ammonia, originating from the sewage, is strongly reduced (very few urban areas).

The total phosphorous (TP) is generally less variable with no observable pattern. As explained above, part of the difficulty of analysis comes from the fact that only the water is analyzed and not the sediments, in which some P can be fixed and excluded from the analysis.

The regression results for the five water quality (WQ) parameters are presented in Table 7.2. Pearson's correlation was employed to evaluate the linear relation between each of the LULC categories and each WQ parameter considering all analysis area types (Exclusive Contribution Area - ECAs and Buffer-B). We included in the models only the LULC categories with a significant correlation and a correct sign (although no inversion of sign was observed in the significant correlations).

Some LULC classes were very poorly represented in terms of area (like grassland) and were either merged with other classes with similar characteristics or excluded from the models. Furthermore, when a strong correlation was observed between two LULC categories (collinearity), only the one with the strongest correlation was considered.

Analysis of Table 7.2 reveals that total ammonia nitrogen (TAN) is positively correlated with the proportion of "Agriculture and Pasture"(AP) and "Urban"(Ub) and negatively cor-

related with natural forested vegetation (F_t and R_F). A similar effect was (mostly) observed for dissolved oxygen (DO), but with inverted signs since DO is considered a desirable quality. Nitrite (NO^{2-}) was positively correlated with the proportion of AP but only for B150. For the remaining buffers (B300, B600 and B900) and the ECAs, nitrite was positively correlated only with the Urban category (U_b).

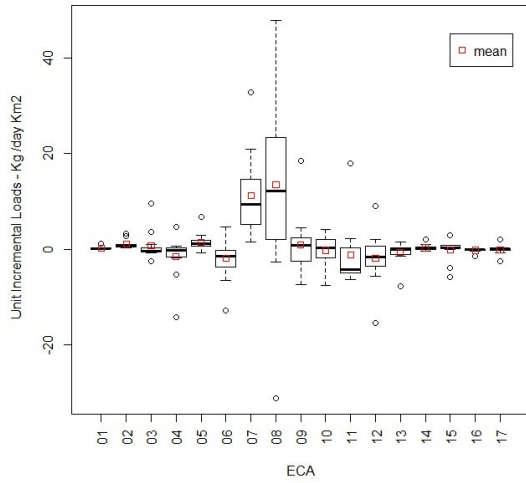
Table 7.2 shows the multiple regression equations, (with the coefficients of determination and the sign of Pearson's correlations) for the 15 significant models out of the potential 25 (5 parameters and 5 area types). For all the models shown in Table 7.2, we were able to reject the null hypothesis (H_0) stating that the betas are equal to zero with a $P \leq 0.05$ significance.

The total phosphorous (TP) is generally less variable and no consistent pattern is observable. Part of the difficulty of analyzing total phosphorous TP comes from the fact that only the water is analyzed and not the sediments, in which some total phosphorus (TP) can be fixed and excluded from the analysis.

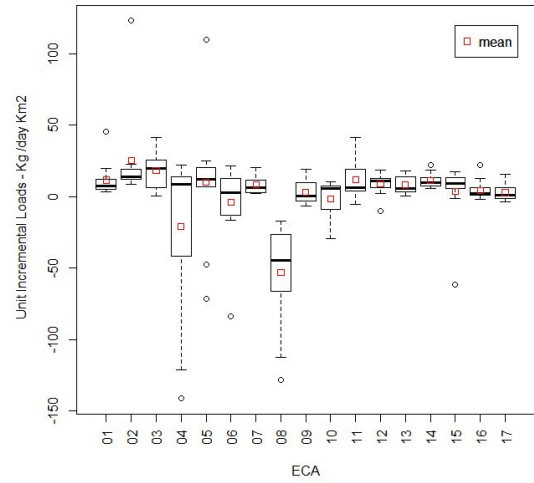
Incremental loads of nitrate (NO^{3-}) are more stable compared with ammonia nitrogen. Station 8 shows a strong reduction compared with station 7, explained by the absence of the dissolved oxygen (DO) necessary to convert ammonia into nitrate (Table 7.1). The Incremental loads of nitrate are almost always positive for the first 13 stations, showing a gradual increase of nitrate downstream up to station 13. We attribute this to the time delay needed for the conversion of ammonia into nitrate. Since the successive stations are downstream from one another, this conversion time translates to distance. After station 13, the Paraúna River with an average discharge of $72,4 \text{ m}^3\text{s}^{-1}$ merges with the Velhas River provoking significant improvement in its quality.

Pearson's correlation coefficients indicated that U_b and AP were the main LULC categories responsible for an increase in pollution (source) for three parameters: total ammonia nitrogen (TAN), nitrite (NO^{2-}), and dissolved oxygen (DO), whereas vegetation (F_t and R_F) functioned as a pollution sink, in both Exclusive Contribution Area (ECA) and Buffer (B). We then tested our proposed index as an independent variable to uniquely substitute the four LULC individual categories (agriculture and pasture - AP, forest - F_t , riparian forest - R_F and Urban - U_b) in the following relation (Eq. 7.7).

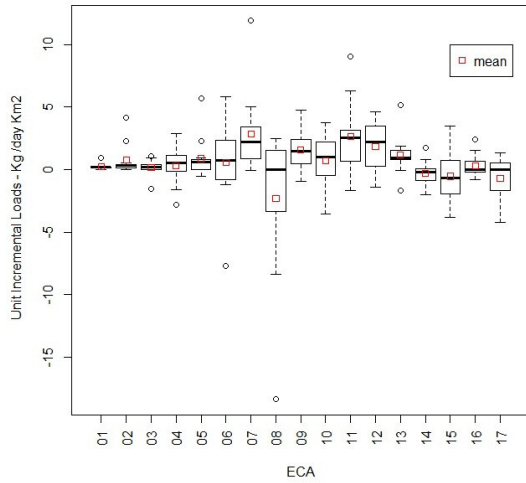
$$LCPI = \frac{U_b + AP}{R_F + F_t} \quad (7.7)$$



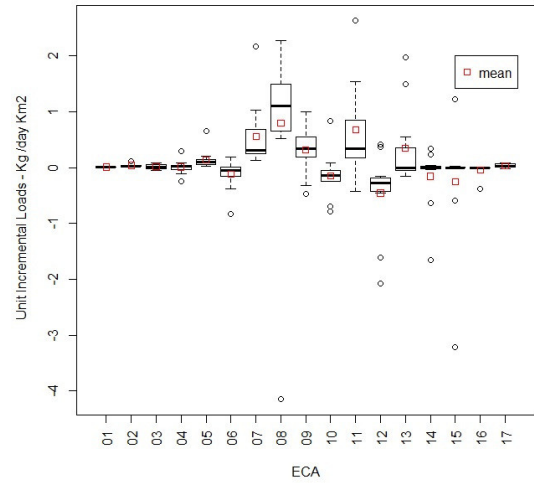
(a)



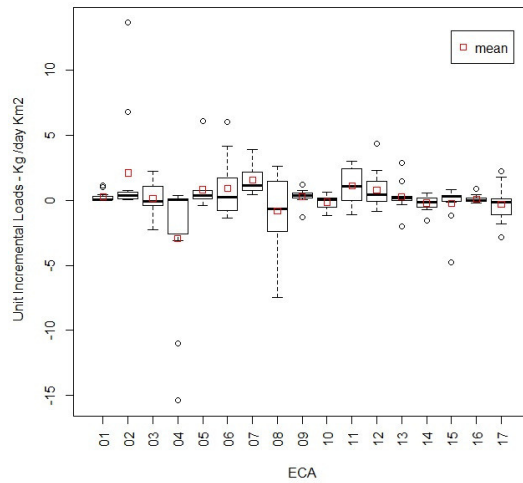
(b)



(c)



(d)



(e)

Figura 7.6: Unit incremental load for the five water quality parameters (kg / day km²) and the 17 WQ stations: (a) - Total Ammonia Nitrogen, (b) Dissolved Oxygen, (c) Nitrate, (d) Nitrite and (e) Phosphorus.

Tabela 7.2: Results of multiple regression between Land Use and Land Cover (LULC) and Land Cover Pollution Index (LCPI) for the different analysis areas (ECAs and four B). Columns GI through Ft and LCPI show the sign and strength of Pearson's correlation whereas the two Model columns explicitly show the regression formulas.

Zone	Y	AP	U_b	R_f	F_t	Model(LULC)	R^2	LCPI	Modelo (LCPI)	R^2
150 m	TAN	+	+	-	-	$-0.638 + 0.21 AP - 0.085 F_t$	0.48	+	$-0.126 + 0.538 LCPI$	0.60
	OD	-	+	+	+	$-3.014 - 0.376 AP + 0.555 F_t$	0.48	-	$9.51 - 1.932 LCPI$	0.88
	NO_2^-	+	+	-	-	$0.049 + 0.014A - 0.008F_t$	0.46	+	$0.007 + 0.040 LCPI$	0.78
300 m	TAN	+	+	+	+	$-3.631 + 0.152 AP + 0.168 U_b$	0.41	+	$-0.786 + 1.058 LCPI$	0.62
	OD	-	+	+	+	$-5.091 - 0.389 U_b + 0.447 F_t$	0.45	-	$12.108 - 4.67 LCPI$	0.87
	NO_2^-	+	+	-	-	$-0.018 + 0.016 U_b$	0.31	+	$-0.050 + 0.092 LCPI$	0.79
600 m	TAN	+	+	+	+	$-0.917 + 0.244 U_b$	0.43	+	$1.633 + 0.367 LCPI$	0.60
	OD	-	+	+	+	$-1.191 - 0.564 U_b + 0.35 R_f$	0.48	-	$1.517 - 6.396 LCPI$	0.88
	NO_2^-	+	+	-	-	$-0.041 + 0.019 U_b$	0.43	+	$-0.103 + 0.133 LCPI$	0.77
900 m	TAN	+	+	+	+	$-0.939 + 0.262 U_b$	0.43	+	$-2.627 + 2.52 LCPI$	0.48
	OD	-	+	+	+	$+0.025 - 0.481 U_b + 3.247 R_f$	0.42	-	$20.567 - 10.996 LCPI$	0.76
	NO_2^-	+	+	-	-	$-0.040 + 0.02 U_b$	0.41	+	$-0.218 + 0.226 LCPI$	0.65
ECA	TAN	+	+	+	+	$-1.15 + 0.193 U_b + 0.016 AP$	0.83	+	$-1.018 + 0.934 LCPI$	0.31
	OD	-	+	+	+	$5.137 - 0.557 U_b + 4.427 R_f$	0.77	-	$15.77 - 5.099 LCPI$	0.77
	NO_2^-	+	+	-	-	$-0.030 + 0.015 U_b$	0.80	+	$-0.135 + 0.111 LCPI$	0.74

Y - Dependent Variable: Total Ammonia Nitrogen (TAN), Dissolved Oxygen (DO) e Nitrite (NO_2^-)

Independent Variable: U_b (Urban), F_t (Forest), R_f (Riparian Forest), AP (Agriculture and Pasture) and LCPI.

Analysis Areas: Exclusive Contribution Area (ECA), Buffer (B).

***0.01 level (correlation is significant; two tailed) and **0.05 level (correlation is significant; two tailed).

+ for positive correlation, - for negative correlation, and no symbol represents no linear correlation.

By using the LCPI index (Table 7.3), we were able to reduce the four LULC categories to a single variable while eliminating problems of collinearity.

Of the 15 models shown in Table 7.2, 12 have had their coefficients of determination increased significantly by using the index instead of the individual LULC categories and one remained equal. This was especially noticeable for all the riparian buffers. Conversely, none of the three exclusive contribution area (ECA) models (Total Ammonia Nitrogen- TAN, Dissolved Oxygen - DO and Nitrite - NO_2^-) were improved. This however makes sense since it is expected that as one increases the area of analysis, the number of factors affecting the relation also increases therefore reducing the explaining strength of any single variable. The relative effect of the index on the coefficients of determination can be observed graphically in Fig. 7.7 for three water quality parameters (Total Ammonia Nitrogen, Dissolved Oxygen and Nitrite). Models for phosphorous and nitrate were not represented because they were not significant or had a low R^2 (<0.4).

Tabela 7.3: Land cover pollution index (LCPI) for the 17 stations and the five areas of analysis.

WQ Station	Land cover pollution index (LCPI)				
	150	300	600	900	ECA
1	0.40	0.40	0.35	0.32	0.51
2	1.09	1.15	1.13	1.14	0.90
3	0.11	0.16	0.19	0.20	0.41
4	0.77	0.79	0.76	0.70	0.58
5	0.64	0.84	0.78	0.66	0.68
6	1.94	2.42	2.00	1.89	0.38
7	2.58	2.45	2.09	2.03	2.02
8	27.88	12.43	8.86	4.78	9.96
9	3.79	3.36	2.71	2.51	2.68
10	0.51	0.41	0.43	0.48	0.96
11	0.29	0.46	0.81	1.18	5.05
12	0.44	0.71	1.12	1.28	2.09
13	0.63	1.05	1.62	1.88	2.60
14	0.28	0.48	0.84	1.05	1.39
15	0.39	0.82	1.42	1.83	1.96
16	0.32	0.47	0.79	1.09	1.94
17	0.31	0.47	0.96	1.53	2.83

The least gain made by using the LCPI index for the Total Ammonia Nitrogen, Dissolved Oxygen and nitrite models was 11% (0.43 to 0.48) for Total Ammonia Nitrogen at B900 while the largest was 155% (0.31 to 0.79) for nitrite at B300. On average, including all good models ($R^2 > 0.4$) the gain produced by using the index was of 51%.

We outline the fact that the contribution of the natural vegetation is especially strong as one gets nearer the stream and diminishes rapidly with the increasing distance. This explains

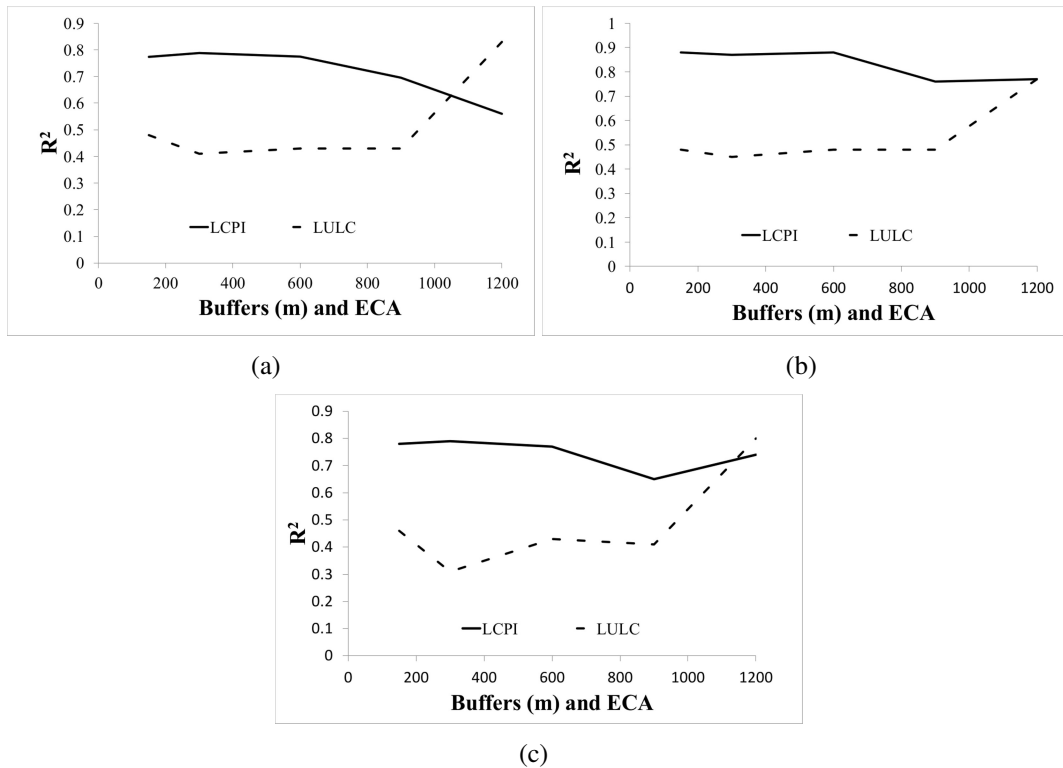


Figure 7.7: Coefficient of Determination for the LCPI e LULC models: (a) TAN, (b) OD and (c) NO_2^- .

why the index (having natural forested vegetation as its denominator) did not perform well while using the whole Exclusive Contribution Areas (ECAs). This also tends to agree with many studies stating the importance of the riparian zone for maintaining water quality (Lorwance *et al.*, 1997; Sliva e Williams, 2001; Poor e McDonnell, 2007; Tu, 2011; Bu *et al.*, 2014; Li *et al.*, 2015; de Oliveira *et al.*, 2016).

7.7 Conclusion

Non point or diffuse pollution sources are difficult to control and to model because it involves complex relations that occur within the whole watershed. It is also not easily separated from point pollution and many pollutants can originate from either sources. Because non point pollution requires transport through runoff, it is naturally associated with precipitation so, as a first measure to concentrate on non point pollution we restricted our analysis on periods with significant precipitation and defined the start of the rain season using the hydrograms of discharge stations. We also transformed the pollutants concentrations in unit incremental loads (UIL) where the loads of the station upstream are subtracted from each station and then divided by the contribution area making the data more independent and comparable. Non point pollution is essentially associated with land use and land cover (LULC)

which can represent difficult to use in statistical inference because of frequent collinearity between the LULC categories used as independent variables. LULC categories are usually used as proportions of land coverage so that they are not strictly independent from each other. This is easily observable using stepwise regression that usually rejects most LULC variable to concentrate on one or two categories (Sliva e Williams, 2001; Buck *et al.*, 2004; Pratt e Chang, 2012; Bu *et al.*, 2014; Shen *et al.*, 2014; Li *et al.*, 2015). This often makes any regression model non replicable in any other context.

In this article we introduce a Land Cover Pollution Index (LCPI) that combines different LULC categories (four in our case) by inferring on their known ability to serve either as a pollution "source" or "sink". The index is generic in nature and can be represented by the simple relation: "source / sink". In our case the index was created using the following relation (Eq. 7.8).

$$LCPI = \frac{(Urban + Agriculture + Pasture)}{(RiparianForest + NativeForest)} \quad (7.8)$$

We used both the individual LULC categories and the index to compare the results for five water quality parameters: total ammonia nitrogen (TAN), dissolved oxygen (DO), total phosphorous (TP), nitrate (NO_3^-) and nitrite (NO_2^-). Five analysis areas were used, four riparian buffers (150, 300, 600 and 900 m) and the whole exclusive contribution area (ECA) for 17 stations of the *Velhas* River watershed in South-East Brazil.

Good significant multiple regression models were obtained for total ammonia nitrogen, dissolved oxygen and nitrite using the individual LULC categories with coefficients of determination between 0.41 and 0.80 and for all five areas of analysis. Phosphorus and nitrate did not yield valid models and we tried to explain this by the specific characteristics of these parameters. We then substituted the individual LULC categories by the Land Cover Pollution Index (LCPI) index in simple regression models to assess its potential. In 12 out of 15 valid models the coefficient of determination (R^2) increased by 11 to 155%. Only two models had inferior R^2 and one remained equal. Interestingly all 12 models improved with the LCPI were created using the four riparian buffers and the remaining three models are all from the exclusive contribution areas. This can be understood by the increased complexity of analyzing the Exclusive Contribution Area (ECA) instead of a restricted riparian buffer. This also strengthens the accepted fact that the riparian zone should be considered the most important filtering area for pollutants from non point sources.

Because every region is inherently unique, any effort of comparison is always difficult if not impossible. Land use and land cover categories differ widely in different parts of the world and so are their effect on non point pollution. The advantage of the Land Cover Pollution Index resides in its generic quality by simplifying the relationship between land use and land cover (LULC) categories and water quality.

7.8 Acknowledgements

The authors would like to acknowledge the following institutions: the Instituto Mineiro de Gestão das Águas (IGAM) for the water quality data, the Agência Nacional de Águas (ANA) and the Companhia Energética de Minas Gerais (CEMIG) for the discharge data and the Instituto Estadual de Florestas (IEF) for providing the RapidEye scenes from which the LULC map was based. This article is an inherent part of L.M. Oliveria's PhD thesis under the supervision of P. Maillard with E.J. de Andrade Pinto acting as co-supervisor.

7.9 Bibliographic References

The bibliography in page 134.

Capítulo 8

ARTIGO 4 : Estimativa das cargas de poluição difusa com a separação das fontes, pontual e difusa, baseando-se no regime fluvial.

8.1 Autores

Líliã Maria de Oliveira¹ and Éber José de Andrade Pinto² and Philippe Maillard³

8.2 Resumo

Neste trabalho propomos um método de separação das fontes de poluição difusa e pontual baseando-se no fluviograma e, também, a estimativa da carga difusa em 14 estações de qualidade de água localizadas na bacia do Rio das velhas. O escoamento superficial é responsável pela “limpeza” inicial do solo da bacia, e traz consigo grande quantidade de poluentes difusos presentes no solo, sendo esta quantidade função do uso e ocupação do solo da bacia. Por outro lado em período sem ou com reduzida ocorrência de precipitação, não há geração de escoamento superficial, e os poluentes transportados são de origem pontual, advindos do lançamento de efluentes no curso de água. Neste trabalho aplicamos o seguinte conceito: durante a ascensão do hidrograma de cheia são transportados poluentes de origem pontual e difusa e as menores vazões do período de estiagem, que caracterizam o escoamento de base, são responsáveis pelo transporte de poluição pontual. Para os fluviogramas de 14 estações de monitoramento de qualidade, gerados para o período de abril de 2008 a

¹CEFET-MG, Av. Amazonas, 5354, Belo Horizonte, Brasil Tel.: +55-31-33770361 email: lilia@deii.cefetmg.br

²Universidade Federal de Minas Gerais, Escola de Engenharia, Av. Antônio Carlos, 6627, Belo Horizonte, Brasil; Tel.: +55-31-33770361; email: eber.andrade@cprm.gov.br

³Universidade Federal de Minas Gerais, Instituto de Geociências, Av. Antônio Carlos, 6627, Belo Horizonte, Brasil; Tel.: +55-31-33770361; email: philippermaillard@yahoo.com.br

março de 2013, identificamos as cargas de nitrato e fósforo produzidas durante a ascensão (cheia) e as cargas geradas durante o transporte de poluentes pelo escoamento de base (seca). A diferença entre as cargas de cheia e seca, produziram uma estimativa da poluição difusa em cada estação, para os 5 anos avaliados. Posteriormente calculamos a carga de poluição difusa incremental unitária entre duas estações e a correlacionamos com o uso do solo da área de contribuição específica (ECA), tomado para diferentes distâncias do curso de água (ECA e faixas). Apesar de ser uma aproximação inicial, os resultados obtidos foram bastante promissores e melhoram o desempenho do modelo de nitrato em até 51% para o coeficiente determinação em uma faixa 150 metros, quando comparado ao modelo desenvolvido com dados de todo período chuvoso. Para o fósforo total, não foi possível gerar modelo, em função da sua retenção no solo durante anos mais secos e posterior liberação em anos chuvosos.

8.3 Palavras-Chave

Qualidade das águas, Uso e ocupação do solo, Poluição difusa, Fluviograma, Escoamento superficial.

8.4 Introdução

A qualidade das águas de um rio, medida por meio de seus parâmetros de qualidade, é influenciada pelo seu regime de escoamento, que pode promover o aumento ou a redução da concentração de nutrientes na água (Gardi, 2001; Poor e McDonnell, 2007).

Kronvang e Bruhn (1996) avaliaram a concentração de fósforo (P) e nitrogênio (N) em dois rios da Dinamarca durante dois anos (1987-1989). A quantidade de fósforo total (TP) transportado em condições de seca e chuva apresentou valores diferentes. Durante o primeiro ano de estudo, 67% e 43% de TP foi transportado durante as chuvas nas duas bacias avaliadas, respectivamente. No segundo ano de estudo este valor foi para 71 e 37%. Já o nitrogênio total (TN) teve sua concentração aumentada sempre que havia aumento na vazão.

A concentração das frações de fósforo reagiu de forma diferente durante os períodos de seca e de chuva em ambos os rios. A relação vazão-concentração para Fósforo inorgânico dissolvido (DIP) foi razoavelmente bem definida durante condições de seca (menor vazão) em um dos rios e em condições de cheia a concentração de DIP não esteve relacionada com a vazão. O comportamento da vazão e da concentração para fósforo inorgânico particulado (PIP) foram semelhantes nos dois rios, mas não foi possível estabelecer uma relação entre as duas variáveis na seca ou na chuva.

De acordo com Davis *et al.* (2014) em anos mais secos, ou seja, com menor volume de escoamento superficial, ocorre menor lixiviação do solo promovendo o acúmulo de nitrogênio (N) e fósforo (P) nos solos e sedimentos. Assim a ocorrência de eventos de precipitação,

posteriores a estes anos secos, faz com que o transporte destes nutrientes resultem em aumento na concentração de N e P nas águas receptoras. Por outro lado, anos mais úmidos, com melhor distribuição das chuvas, resultam em menor concentração de N e P no curso de água devido à contínua lavagem dos nutrientes pelo escoamento superficial.

Segundo Donner e Scavia (2007) parte da bacia do rio Mississippi foi responsável pelo fluxo de nitrato para o Golfo do México, entre os anos de 1980 e 2002. A área avaliada de 3.2 milhões de Km^2 , possuía 60 % da ocupação dada pelo cultivo de soja e milho.

O estudo revelou que a precipitação ocorrida nesta área foi responsável pela maior parte do nitrato disponibilizado para o Golfo do México. A precipitação média anual sobre a área de estudo foi de 961 mm, 19% maior que média anual da chuva quando toda a área da bacia do rio Mississippi foi considerada.

A análise de regressão estabeleceu uma relação linear entre o fluxo de nitrato e a precipitação durante os meses de maio a junho ($R^2 = 0.57$). Considerando o aumento na precipitação, nos meses de novembro a maio, obteve-se uma relação mais forte ($R^2 = 0.64$).

Stutter *et al.* (2008) avaliaram a variação na concentração de frações de fósforo (fosfato, fósforo orgânico dissolvido - DOP e fósforo particulado - PP), Sedimento - SS e nitrato em 5 eventos chuvosos que ocorreram em duas pequenas bacias (4 e 51 km^2) na Escócia, e foram precedidos por diferentes condições de umidade antecedente.

O primeiro evento de chuva ocorreu em agosto após um período mais longo de estiagem e foi acompanhado do aumento na concentração de fosfato, DOP e PP durante a ascensão e pico do hidrograma, durante a recessão ocorreu a redução da concentração do fósforo. O comportamento da concentração do nitrato foi semelhante, entretanto com menor variação na concentração durante ascensão, pico e recessão do hidrograma.

Os eventos 2 e 3, ocorreram no mês de outubro, após as chuvas de agosto. A concentração de nitrato foi maior que o valor obtido para o mês de agosto, considerando ascensão, pico e recessão do hidrograma. Já a concentração de fosfato, DOP e PP apresentou menor variação ao longo do hidrograma. O quarto e quinto eventos foram monitorados em março, mês com menor precipitação, sendo que o período que antecedeu o quarto evento foi mais chuvoso quando comparado ao período antecedente ao quinto evento. As maiores concentrações de fosfato e DOP ocorreram durante a ascensão e pico do hidrograma do 5º evento estudado. Já PP apresentou maior concentração no início da ascensão do hidrograma do 4º evento. A concentração de nitrato apresentou uma queda do 4º para o 5º evento.

Considerando os 5 eventos, a concentração de fosfato e DOP apresentaram maiores valores para o 1º evento (agosto) e para o nitrato os maiores valores ocorreram no 2º e 3º eventos. Em todos os eventos avaliados a concentração de PP acompanhou a variação da concentração de sedimento - SS, que por sua vez reproduziu o hidrograma aumentando na ascensão, com maior concentração no pico e redução durante a recessão.

Pesquisas de manejo de bacias nos Estados Unidos e no Báltico sugerem que o acúmulo

de nutrientes no solo, decorrente de lançamentos realizados ao longo do anos, resulta em uma estacionariedade biogeoquímicas destes nutrientes na bacia. A existência destas áreas com armazenamento de nutrientes amortecer quaisquer variações biogeoquímica que ocorram na bacia, tais como, redução na entradas de fertilizantes (Green *et al.*, 2007; Basu *et al.*, 2010).

Poor e McDonnell (2007) utilizaram o monitoramento de eventos de chuva, vazão e concentração de nutrientes, para determinar se o aumento da vazão produzia a “diluição” de um nutriente o que pode indicar a dominância de uma fonte pontual, ou o aumento da vazão implicava em aumento na “concentração” do nutriente na água o que pode indicar a descarga de uma fonte difusa.

A pesquisa de Poor e McDonnell (2007) foi conduzida em 3 bacias com características similares (área, geologia, clima, etc) mas diferentes usos do solo (floresta, agricultura e urbano), os resultados mostraram diferenças marcantes nas taxas de exportação de nitrato, em períodos de maior precipitação (outono, inverno e primavera) nas três bacias hidrográficas monitoradas. A bacia com uso por floresta mostrou exportação mínima nos três eventos monitorados e a bacia com uso urbano apresentou alta taxa de exportação para os três eventos. Já a bacia com uso agrícola exibiu exportação elevada no outono, similar à bacia hidrográfica urbana. Entretanto, durante o restante do período chuvoso (inverno e primavera) a exportação de nitrato diminuiu progressivamente na bacia com uso agrícola, os autores atribuíram este resultado devido ao fato do nitrato ter sido lançado na seca e as primeiras chuvas do outono levaram maior quantidade de nitrato para o curso de água.

Outram *et al.* (2014; 2016) avaliaram o Rio Wensum que possui 75 km de comprimento, no leste da Inglaterra. A bacia Wensum (570 km²) é dividida em 20 sub-bacias hidrográficas, uma das quais, com 19.7 km² é a sub-bacia Blackwater, localizada na cabeceira da bacia de Wensum sendo intensamente monitorada como parte do projeto Wensum DTC. O uso do solo na sub-bacia é predominantemente agricultura (74%), seguido de pastagem (12%) e florestas (11%).

O monitoramento foi realizado na foz da sub-bacia de Blackwater com registro dos dados de: vazão, chuva, águas subterrâneas, concentração de amônia, nitrato, fósforo total (TP) e fósforo reativo total (TRP), realizado a cada 30 minutos.

A combinação do monitoramento detalhado de nitrato, TP e dados hidrológicos com dados de aplicação de fertilizantes produziram uma nova compreensão dos fatores que controlam o transporte de nitrato e fósforo total (TP) para a foz do rio. A precipitação total anual e a infiltração foram importantes para o transporte do nitrato e de TP, sendo os fluxos escoados de nitrato e fósforo total em um ano chuvoso o dobro do registrado em um ano seco.

Os autores avaliaram que eventos chuvosos subsequente a períodos secos levaram a uma elevação temporária de nitrato e particularmente das cargas de TP, em função da interação do escoamento superficial sobre reservas antigas de nutrientes. O calendário de aplicação

de fertilizantes teve pouco impacto sobre o transporte de nutrientes e reduções realizadas na entrada anual de fertilizante em função da rotação de culturas, não resultaram em semelhante redução na disponibilização de nitrato ou TP.

Segundo Bowes *et al.* (2008; 2014) em rios onde as fontes pontuais são a forma preponderante do poluição, com a quantidade de fósforo lançada suposta constante no tempo, tem-se que a concentração de fósforo diminui com o aumento da vazão. De maneira inversa, em rios onde o fósforo provém principalmente de fontes difusas o aumento da vazão é acompanhado de um aumento na concentração de fósforo na água. Estas suposições também foram realizadas nos estudos de Jarvie *et al.* (2006); Jordan *et al.* (2007); Wood *et al.* (2005).

Considerando estas suposições Bowes *et al.* (2008) propôs um modelo, baseado em uma função de potência, para estimar a quantidade de fósforo derivado de fontes difusas e pontuais a partir da vazão escoada pelo curso de água. O modelo foi calibrado e validado utilizando dados semanais de monitoramento de fósforo e vazão durante um período de 2 anos em 3 bacias do Reino Unido.

Os resultados obtidos na estimativa da carga difusa e pontual de fósforo foram considerados adequados para os pontos estudados. Entretanto, um detalhamento maior na coleta de dados para pontos com preponderante poluição por fontes pontuais é necessário, haja visto ter ocorrido nestes locais a maior dificuldade em estimar as cargas difusas. O monitoramento de eventos específicos de precipitações intensas, assim como a inclusão de dados de chuva no modelo foram apontados com possíveis soluções para este problema.

Por outro lado os autores atentam para o problema de que os dados produzidos pela Agência Ambiental do Reino Unido e outras agências governamentais terem frequência mensal o que não é adequado ao estabelecimento da relação vazão-carga de fósforo, pois não são capazes de retratar adequadamente a variação da vazão-fósforo no tempo.

O modelo desenvolvido assume que o fósforo total (TP) é relativamente constante. Entretanto a presença de TP na água está sujeita a processos que podem alterar a sua concentração, tais como interações com sedimentos e biota, e a suspensão/re-suspensão de fósforo particulado (Bowes *et al.*, 2008).

O modelo pareceu ser capaz de lidar com estas perturbações por meio do ajuste dos seus coeficientes, mas pode não obter bons resultados para rios com altas taxas de processamento do fósforo, ou seja, sítios onde a retenção ou libertação TP dentro do canal respondem pela maior proporção da carga total TP.

Os vários estudos aqui apresentados avaliaram a concentração de nutrientes transportados no curso de água em diferentes condições de umidade antecedente e indicaram que o comportamento do nitrato e fósforo foi dependente da condição de umidade.

Estudos como os desenvolvidos por Poor e McDonnell (2007); Stutter *et al.* (2008); Oustram *et al.* (2014; 2016) ainda são de difícil realização no Brasil, uma vez que o monitoramento detalhado de nutrientes (ex. nitrato e fósforo) e da vazão na discretização temporal

adequada para eventos de chuva são de elevado custo financeiros e desta forma são realizados em poucas bacias. Os dados de qualidade da água, obtidos pelo estado de Minas Gerais, passaram a ter maior frequência de monitoramento a partir de 2008 para as estações localizadas na calha do Rio das Velhas (IGAM, 2015).

Considerando a geração de nutrientes em diferentes condições de umidade, ou seja, para diferentes volumes escoados no curso de água e também para diferentes condições de saturação do solo, alguns estudos tentam estabelecer a relação do fósforo na água e o uso e ocupação do solo da bacia (Sliva e Williams, 2001; Howarth *et al.*, 2002; Buck *et al.*, 2004; Salvia-Castellví *et al.*, 2005; Maillard e Santos, 2008; Ouyang *et al.*, 2010; Li *et al.*, 2015; de Oliveira *et al.*, 2016).

de Oliveira *et al.* (2016) avaliou a relação entre a carga unitária incremental ($\text{kg/ km}^2 \text{ dia}$) de nitrato, amônia, oxigênio dissolvido (OD) e fósforo total (TP) monitorada em 12 períodos chuvosos ocorridos entre 2001 e 2013 em 17 estações localizadas na calha principal do Rio das Velhas. Foram delimitadas áreas de contribuição exclusiva entre duas estações (ECAS) e obtidas faixas de diferentes larguras (150, 300, 600 e 900 metros) para as ECAs. O uso do solo era bastante diversificado entre as categorias: mineração, agricultura, pecuária, urbano e floresta, sendo obtido o percentual de cada categoria para as faixas e ECAs.

Os autores conseguiram relacionar as categorias de LULC às alterações na carga incremental unitária de amônia e OD, sendo os melhores resultados obtidos para a ECA, com coeficiente de determinação de 0.83 e 0.77, respectivamente. Entretanto, para o nitrato somente foi possível estabelecer um modelo para a faixa 150 m ($R^2= 0.31$) e para o fósforo total, nenhum modelo foi estabelecido. A dificuldade em estabelecer modelos para o nitrato e para o fósforo foi atribuída à presença de fontes pontuais (esgoto) e a retenção e liberação de fósforo pelo canal, respectivamente.

A relação entre categorias de LULC e aspectos da paisagem foram avaliadas por Li *et al.* (2015) para a amônia, nitrato e nitrogênio total. Foram utilizados 18 pontos de monitoramento e 8 categorias LULC. Os modelos obtidos retiveram somente a categoria urbano como variável explicativa, as demais categorias de LULC não foram utilizadas. O autor justificou a dificuldade na definição dos modelos à presença de fontes pontuais na bacia, reforçando assim a necessidade de estimar-se separadamente a carga proveniente de fontes pontuais e difusas.

O presente estudo apresenta um método para estimativa das cargas difusas e pontuais em 14 pontos de monitoramento da bacia do Rio das Velhas considerando os parâmetros nitrato e fósforo total. O método se baseia nos seguintes apontamentos realizados pelos estudos anteriormente descritos:

- As maiores cargas de poluição difusa ocorrem durante a ascensão dos primeiros hidrogramas após um período de estiagem, devido a lixiviação inicial de nutrientes depositados no solo da bacia;

- É possível estimar a carga difusa de nitrato e fósforo total de uma bacia, em um dado período de chuvas, considerando a subtração da “carga do período chuvoso”, a qual representa a soma das contribuições difusa e pontual, da “carga do período seco anterior”, que representa preponderantemente a carga pontual;
- O uso do solo das áreas de contribuição exclusiva (ECA) e de faixas de diferentes larguras, estabelecidas nas margens dos rios, influencia de maneira diferente a carga incremental unitária da ECA.

Assim, o método de separação das cargas difusas e pontuais se apoia nas seguintes afirmações:

- Durante o período de chuva as cargas de nitrato e fósforo total são decorrentes da entrada de fontes pontuais e difusas;
- A carga do período seco, anterior às chuvas, avaliadas nas estações de monitoramento do Rio das Velhas, é decorrente das fontes pontuais existentes na bacia.

8.5 Materiais e Métodos

8.5.1 Área de Estudo

A Bacia do Rio das Velhas com área de 27,887 km^2 está localizada na região central do estado de Minas Gerais - Brasil (Figura 8.1) abrangendo uma população de aproximadamente 4.4 milhões de habitantes (CBH, 2015), sendo que a maior parte desta população vive na capital do estado e entorno, definida como Região Metropolitana de Belo Horizonte (RMBH).

O clima predominante na bacia é o tropical semi-úmido de altitude, com estações de seca e chuva bem definidas. A precipitação diminui de montante para jusante, com média anual de 1600-1700 mm próximo as nascentes, e cerca de 1100 mm na foz da bacia (Pinto *et al.*, 2011). O semestre mais chuvoso na bacia compreende os meses de outubro a março, período no qual se verifica a elevação do nível das águas dos cursos de água (IGAM, 2005).

O relevo apresenta vales estreitos e fortemente encaixados na parte alta da bacia e grande quantidade de meandros no trecho médio e baixo do rio que percorre uma extensão de 802 km, nascendo a uma altitude de aproximadamente, 1.500 m e com foz no Rio São Francisco a uma altitude de 478 m (IGAM, 2013a).

O uso do solo da bacia foi classificado em 8 classes principais: 26.4 % Cerrado (Sa), 21.3 % Campo (Gl), 25.2 % Agropastoril (AP), 15.3 % Floresta (Ft), 0.8% Mata Ciliar (Rf), 4.2% Afloramento/Mineração (RM), 3.3 % Reflorestamento (TP), 2.9 % Urbano (U_b) e 0.6% Água, Nuvens e Sem informação (WC), (Figura 8.1).

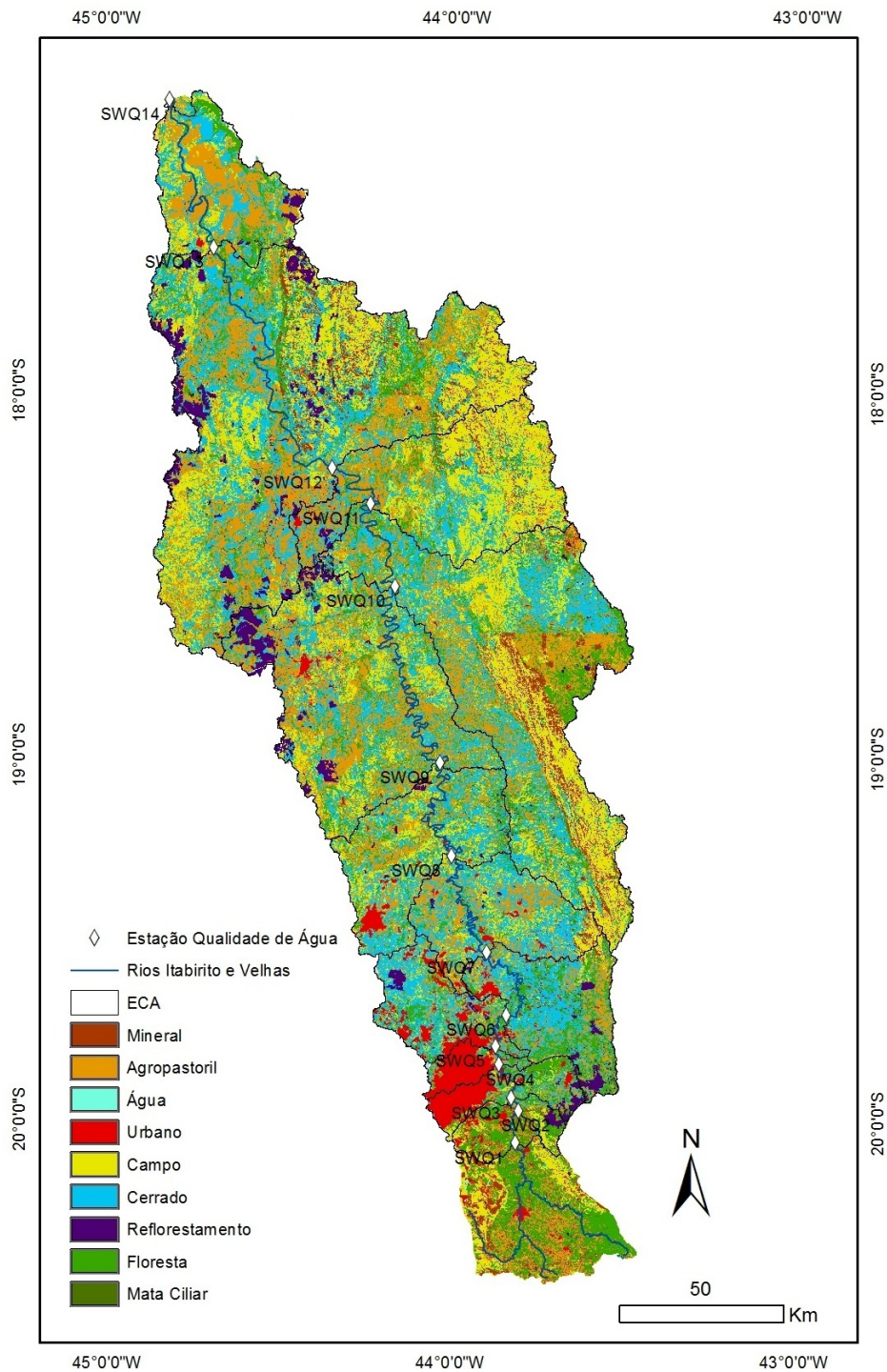


Figura 8.1: Mapa da Bacia do Rio das Velhas contendo: classes de LULC avaliadas, área de contribuição exclusiva (ECA) e estações de monitoramento de qualidade.

O mapeamento do uso do solo foi realizado a partir do uso de imagens RapidEye do ano 2010 com resolução de 5 metros. A classificação das imagens com utilização de regras baseadas principalmente na resposta espectral de cada alvo foi realizada com o software Trimble eCognition Developer 8.4.

Foram utilizadas 14 estações de monitoramento de qualidade das águas e 7 estações de monitoramento de vazão, inseridas na calha principal do Rio das Velhas (Figura 8.1). Os dados de qualidade das águas possuem frequência de amostragem mensal e os de vazões frequência diária. O período de análise dos dados foi de 2008 a 2013.

As estações de vazão consideradas no cálculo da carga de nitrato e fósforo total apresentam as vazões médias diárias características do período de 04/2008 a 04/2013 conforme Tabela 8.1. A coleta de dados foi realizada diariamente (7:00 e 17:00 hs) e os dados disponibilizados pela Agência Nacional de Águas (ANA).

Tabela 8.1: Vazões características das estações fluviométricas utilizadas - SQ.

Estação (SQ)	Área Drenagem (km^2)	Q100%	Q95%	Q50%	Q5%
1(41199998)	1678.5	8.26	15.2	24.3	90
2(41260000)	3735.1	21.1	24.8	45.2	162
3(41340000)	4852.5	22.7	30.6	53.4	207
4(41410000)	6296	27.6	32.8	61.4	271
5(41600000)	8046.8	31.3	41.2	72.8	338
6(41818000)	16658	47	56.2	111	608
7(41990000)	26019.4	54.4	63.8	157	942

8.5.2 Metodologia de Monitoramento

A seleção das 14 estações de qualidade (SWQ) e do período de análise (2008 a 2013) foi realizada em função da necessidade de possuímos dados simultâneos de nitrato e fósforo total em períodos coincidentes com a ascensão dos hidrogramas. Como o monitoramento dos níveis dos rios para estimativa das vazões é realizado duas vezes ao dia (07 e 17 horas), não foi possível trabalhar com as estações fluviométricas com áreas de drenagem menor que $1500 km^2$ (que possui menor tempo de concentração), uma vez que esta frequência de monitoramento não possibilita o traçado do hidrograma de cheia. Os dados de qualidade da água anteriores a 2008 não foram utilizados, pois a frequência de amostragem era trimestral (4 dados/ano).

Os dados de nitrato e fósforo total foram obtidos junto ao IGAM, na forma de concentração (mg/l) e a vazão (m^3/s) junto a ANA. Como algumas estações de vazão não coincidiam com as estações de monitoramento de qualidade foi realizada a transferência dos dados para os pontos de monitoramento de qualidade de água por correlação entre áreas de drenagem, conforme equações 8.1 e 8.2 e exemplo apresentado por Diebel *et al.* (2009).

$$Cf_{wQ/Q} = \frac{A_{wQ}}{A_Q} \quad (8.1)$$

$$Q_{wQ} = Q \times Cf_{wQ/Q} \quad (8.2)$$

onde:

- $C_{fwQ/Q}$ é o fator compensatório de área (adimensional);
- A_{WQ} é a área de contribuição na estação de qualidade da água (km^2);
- A_Q é a área de contribuição na estação de monitoramento das vazões (km^2);
- Q_{WQ} é a vazão na estação de qualidade da água (m^3/s);
- Q é a vazão na estação fluviométrica (m^3/s).

Os dados de nitrato e fósforo total foram coletados mensalmente de junho de 2008 a junho de 2013. A concentração foi então convertida em carga para as 14 estações de monitoramento usando a equação 8.3:

$$L = K * C * Q_{WQ} \quad (8.3)$$

onde:

- L é a carga em kg/dia;
- K é uma constante de conversão de unidades (86.4);
- C é a concentração em mg/l;
- Q_{WQ} é a vazão em m^3/s .

Para que cada estação de monitoramento de qualidade pudesse ser analisada de maneira independente, a sua carga não deveria ser afetada pela estação de montante, para tanto foram utilizados dois procedimentos. Primeiro criamos para cada estação uma área de contribuição exclusiva (ECA), através da subtração da área de drenagem da estação de jusante da estação de montante, considerando as 14 estações de qualidade, Figura 8.2. Segundo realizamos a subtração da carga difusa da estação de jusante da carga difusa da estação de montante, gerando a carga difusa incremental (IDL) e dividimos este valor pela área da respectiva ECA, obtendo a carga difusa incremental unitária (UIDL).

As datas com medidas de concentrações e das vazões para cálculo das cargas nas épocas úmidas e de estiagem, relativa aos 5 períodos avaliados, foram definidas por meio da análise visual do fluviograma nos 14 pontos de monitoramento de qualidade de águas, conforme exemplo da Figura 8.3.

Nos períodos de seca selecionamos as cargas dos parâmetros considerando as datas de coleta com os menores valores de vazão, supondo a ocorrência somente do escoamento de base neste período de monitoramento. A Figura 8.4 exemplifica a situação de seca na Estação

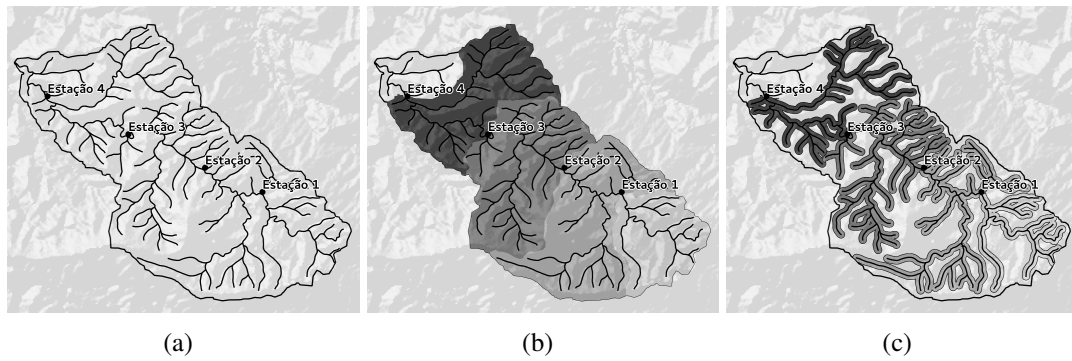


Figura 8.2: Áreas de contribuição exclusiva (ECA) e as faixas foram criadas a partir do modelo digital do terreno: (a) bacia com hidrografia e estações de qualidade de água, (b) determinação da ECA, e (c) faixas de largura variável em ambos os lados do rio.

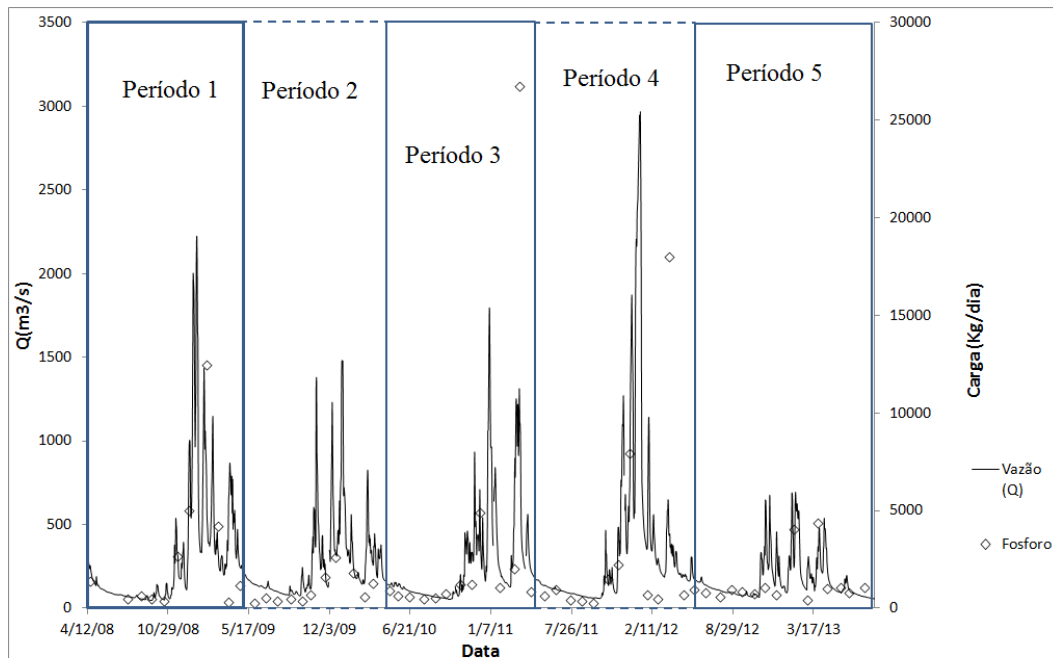


Figura 8.3: Períodos de seca e chuva avaliados entre 2008 e 2013 ($S_{WQ} = 14$).

14, durante o período 2008 a 2009, onde foram selecionados três valores de carga de fósforo na seca.

Para o período chuvoso de 2008 a 2009 foram selecionadas três valores de carga medidas durante os períodos de ascensão dos hidrogramas nas datas indicadas pelas setas na Figura 8.4.

O mapeamento do uso do solo realizado com as informações de 2010 foi considerado como representativo de todo o período de análise, ou seja, não ocorreram alterações significativas no uso e ocupação do solo. Assim, pode se adotar um valor representativo da carga difusa incremental unitária (*UIDL*) de cada ECA para o período do estudo, de 2008 a 2013.

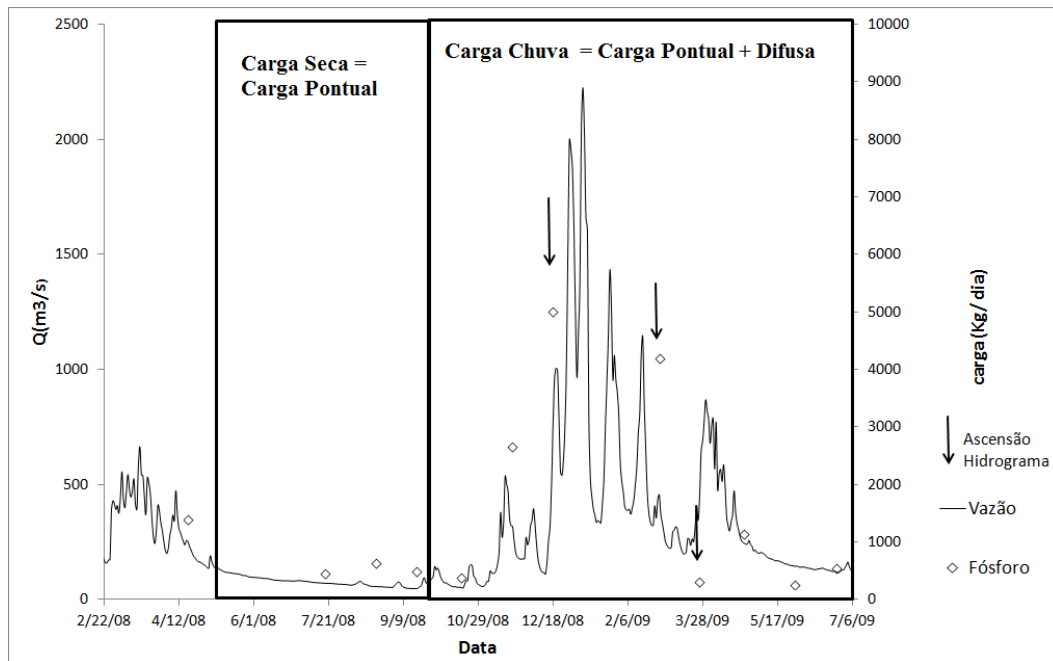


Figura 8.4: Detalhamento do período de 2008 a 2009 para seleção dos dados de seca e chuva ($S_{WQ} = 14$).

A estimativa da carga difusa incremental unitária ($UIDL$) em, $kg\ km^{-2}\ dia^{-1}$, nas 14 ECA's foi realizada da seguinte forma:

- Inicialmente, em cada estação de QA , selecionamos as datas com medidas de concentrações e de vazões para cálculo das cargas nas épocas úmidas e de estiagem, relativa aos 5 períodos avaliados como descrito nos parágrafos anteriores;
- Com os dados obtidos no item anterior, calculamos as cargas (L) de cada parâmetro empregando a equação 8.3 e, se necessário, as equações 8.2 e 8.1;
- Em cada um dos 5 períodos (j), como ilustrado na figura 8.3, para as épocas úmidas e de estiagem calculamos a carga mediana de nitrato e fósforo total em cada estação (i). A mediana foi selecionada por ser uma medida de tendência central que reduz o peso dado aos valores extremos (Miller, 1993). O valor mediano da carga de cada um dos 5 períodos considerados é representado por $(Lxy_{i,j})$, onde x representa o parâmetro avaliado (nitrato e fósforo total), y representa seca (1) ou chuva (2), conforme exemplo da Figura 8.4;
- Em cada estação, de posse das cargas medianas das épocas seca e úmida, calculamos a carga de poluição difusa ($DL_{xi,j}$) de cada período aplicando a equação 8.4;
- A carga difusa foi então transformada em carga incremental ($IDLx$), pela subtração da carga da estação de jusante pela carga de montante (Eq. 8.5) exceto para a primeira

estação, que não possuía estação de montante (Eq.8.6).

- Como a carga difusa pode variar amplamente em cada estação em função das diferentes áreas de contribuição, normalizamos a carga incremental ($IDLx$) dividindo seu valor pela área da respectiva ECA e produzimos a carga difusa incremental unitária ($UIDL$) em $kg\ km^{-2}\ dia^{-1}$ (Eq. 8.7).
- Finalmente, para cada uma das 14 estações, o valor representativo da carga difusa incremental unitária (UIDL) para o período de 2008 a 2013, foi obtido adotando o valor mediano dos 5 períodos avaliados. Com este procedimento nós pudemos verificar se o LULC existente para a ECA, definida entre duas estações de monitoramento, produzia um aumento ou uma redução na carga difusa incremental, e evitamos a ocorrência de problemas relativo a escala das áreas incrementais.

$$DLx_{i,j} = Lx_{2,i,j} - Lx_{1,i,j} \quad \text{para } i=1,\dots,14 \quad j=1,\dots,5 \quad (8.4)$$

$$IDLx_{i,j} = DLx_{i,j} - DLx_{i-1,j} \quad \text{para } i=2,\dots,14 \quad j=1,\dots,5 \quad (8.5)$$

$$IDLx_{i,j} = DLx_{i,j} \quad \text{para } i=1 \quad j=1,\dots,5 \quad (8.6)$$

$$UIDLx_{i,j} = IDLx_{i,j}/ECA_i \quad \text{para } i=1 \dots 14 \quad j=1,\dots,5 \quad (8.7)$$

8.5.3 Análise Estatística

A análise da correlação entre as 8 categorias de uso e cobertura do solo (LULC) e a carga difusa incremental unitária (UIDL) de nitrato e fósforo total foi realizada com base em três medidas de correlação: os coeficientes de Kendall, Spearman e Pearson. Os dois primeiros coeficientes (Kendall e Spearman) são baseados no ordenamento dos dados e avaliam as relações monotônicas entre as variáveis e, também são resistentes a presença de "outliers". O coeficiente de Pearson mede a correlação linear entre as variáveis, que é um tipo específico de relação monotônica (Helsel e Hirsch, 1992).

A correlação entre UIDL e LULC foi avaliada considerando duas razões distintas. Primeiro avaliar a relação de cada categoria de LULC com os parâmetros de qualidade (nitrato e fósforo) e o sinal desta relação. Segundo para verificar o grau da correlação linear entre as variáveis independentes selecionadas para o modelo de regressão. O software SPSS foi utilizado para determinar a correlação com significância estatística de $p < 0.01$ e $p < 0.05$. A relação entre LULC e os dois parâmetros de qualidade foi definida utilizando regressão múltipla. Os modelos na forma de $Y=b_0 + b_1 * X_1 + \dots + b_n * X_n$ onde Y é a variável

dependente, X é a variável independente e $b_{1;2;...;n}$ são os coeficientes de cada categoria de LULC utilizada no modelo. Os modelos foram construídos para a ECA e para diferentes faixas ripárias. Para cada modelo, somente a variável LULC com correlação significativa com o parâmetro de qualidade foi incluída. Finalmente a análise de variância foi realizada para testar a hipótese nula ($H_0 : b_{1;2;...;n} = 0$; $H_1: b_{1;2;...;n} \neq 0$). O nível de significância adotado foi de 5% para rejeitar a hipótese nula (0.05). Os resíduos foram avaliados para checar a sua independência e a normalidade.

8.6 Resultados

A correlação entre Carga Difusa Incremental Unitária de Nitrato (UIDLs) e LULC foi avaliada utilizando os coeficientes de correlação de Kendall, de Spearman e de Pearson Tabela 8.2. O coeficiente de Kendall apresentou valores menores, quando comparado aos obtidos os coeficientes de Spearman e Pearson. Entretanto, considerando o exposto por Helsel e Hirsch (1992), menores valores de correlação obtidos para Kendall e Spearman não significam que estes sejam menos sensíveis que o coeficiente de Pearson, mas simplesmente que existe uma diferença na escala de medida da mesma correlação.

Tabela 8.2: Coeficientes de correlação (Kendall, Spearman e Pearson) entre UIDL de Nitrato e LULC (ECAs e faixas - B).

Zona	Coeficiente	AP	U_b	F_t
150 m	Pearson		0.659*	-0.529*
	Kendall		0.384	-0.407*
	Spearman		0.528	-0.547*
300 m	Pearson		0.612*	-0.532*
	Kendall		0.398	-0.429*
	Spearman		0.502	-0.596*
600 m	Pearson		0.544*	-0.363
	Kendall		0.358	-0.407*
	Spearman		0.506	-0.516
ECA	Pearson	-0.586*		
	Kendall	-0.341		
	Spearman	-0.451		

Y - Variável Dependente: Carga Difusa Incremental Unitária de Nitrato (UIDL)

X - Variável Independente: Sa(Cerrado), U_b (Urbano), F_t (Floresta), AP (Agropastoril)

“*” para $P < 0.05$. “+” para correlação positiva, “-” para correlação negativa.

a (Oliveira et. al. (2016))

Os coeficientes de Kendall e Spearman avaliam a correlação através de ordenamento dos dados, de forma que cada dado seja ranqueado. Embora o coeficiente de Kendall seja inferior ao de Spearman em magnitude, os seus p-valores para significância foram bastante semelhantes. A correlação com significância estatística de $p < 0.05$ foi obtida entre as categorias de uso do solo e a UIDL do nitrato, sendo os resultados semelhantes para os três coeficientes de correlação avaliados Tabela 8.2. As diferenças de magnitude que ocorreram (valor do coeficiente) foram função da escala do coeficiente utilizado, não tendo havido mudança de sinal.

Para o Fósforo Total os três métodos não indicaram correlação significativa entre (UIDLs) e LULC.

8.6.1 Nitrato

A carga difusa incremental unitária ($UIDLx$) para o nitrato é apresentada na Tabela 8.3. O coeficiente de correlação de Pearson foi utilizado para avaliar a correlação linear entre cada categoria de LULC e o nitrato, considerando as cinco áreas de estudo (ECA e 4 faixas). Nós incluímos nos modelos somente as categorias de LULC com correlação significativa ($p < 0.05$).

Tabela 8.3: Carga Difusa Incremental Unitária de Nitrato

Estações (x)	UIDLx ($kg\ km^{-2}\ dia^{-1}$)
1	0.40
2	0.93
3	4.39
4	6.30
5	-0.88
6	-0.56
7	0.68
8	1.14
9	0.55
10	1.20
11	-1.31
12	-0.14
13	0.53
14	-0.05

A Tabela 8.4 apresenta os modelos obtidos para o nitrato no presente estudo e em estudo anterior desenvolvido por de Oliveira *et al.* (2016).

No estudo desenvolvido por de Oliveira *et al.* (2016) verifica-se que o nitrato apresentou correlação positiva com a categoria cerrado, na faixa de 150 metros, neste estudo foram utilizados dados referentes aos períodos chuvosos de 2001 a 2013.

A estimativa da carga difusa, aplicando a metodologia descrita anteriormente, para os anos de 2008 a 2013, permitiu a obtenção de modelos para a ECA e 3 faixas (150, 300 e 600 metros). Sendo que, para as 3 faixas o uso do solo “Urbano” apresentou correlação positiva com o nitrato. A categoria “Floresta” foi negativamente correlacionada com nitrato para as faixas de 150 e 300 metros. Quando se analisou a faixa de 900 metros nenhum modelo foi obtido. Na análise da ECA a categoria “Agricultura e Pastagem” teve correlação negativa com o nitrato.

A Tabela 8.4 mostra as equações de regressão múltipla, com coeficientes de determinação e o sinal da correlação de Pearson, para os 4 modelos propostos e o resultado obtido por

de Oliveira *et al.* (2016) para o nitrato (NO_3^-). Para os modelos apresentados foi possível rejeitar a hipótese nula (H_0) para betas iguais a zero com $P < 0.05$ de significância.

Tabela 8.4: Resultados da regressão múltipla entre LULC para diferentes áreas de análise (ECAs e faixas - B) para o nitrato.

Zona	Y	Sa	AP	U_b	F_t	Modelo(LULC)	R^2
150 m	NO_3^-	***				$0.12 + 0.035 Sa$	0.31 ^a
	NO_3^-			+	-	$1.31 + 1.21U_b - 0.046 F_t$	0.47
300 m	NO_3^-			+	-	$1.47 + 0.096U_b - 0.05 F_t$	0.40
600 m	NO_3^-			+		$-0.097 + 0.109U_b$	0.30
ECA	NO_3^-		-			$4.532 - 0.158AP$	0.34

Y - Variável Dependente; X - Variável Independente: Sa (Cerrado), U_b (Urbano), F_t (Floresta), AP (Agropastoril)

*** para $R < 0.01$ and ** para $P < 0.05$. "+" para correlação positiva, "-" para correlação negativa.

a (Oliveira et. al. (2016))

de Oliveira *et al.* (2016), para a mesma área de análise, obtiveram somente um modelo para o nitrato na faixa de 150 metros (Tabela 8.4). No presente estudo conseguimos melhorar este resultado com a entrada das classes U_b e F_t , em 51% (o coeficiente de determinação foi de 0.31 para 0.47). Também foi possível obter modelos para a ECA e as faixas de 300 e 600 metros. Para a faixa de 900 metros não foi possível obter modelo.

As estações 3 e 4 foram responsáveis pelas maiores cargas positivas difusas incrementais de nitrato (Tabela 8.3), e possuem grande percentual de LULC Urbano (U_b) nas 3 primeiras faixas como ilustra a Figura 8.5. Na faixa de 900 metros o uso urbano se concentra nas estações de 3 a 6, entretanto as duas primeiras apresentam elevada carga difusa positiva e as duas últimas elevada carga difusa negativa, dificultando o estabelecimento da relação com o urbano.

Apesar de ocorrer pequena variação no uso do solo da faixa de 600 m para a faixa 900 m, a mesma não possibilitou a definição de modelo para o nitrato na faixa de 900 metros.

Para a ECA o uso urbano se concentra nas estações de 4 a 6, dificultando também a relação com esta categoria. No caso do modelo de área de contribuição específica (ECA) a correlação entre o nitrato e a categoria de LULC agricultura e pastagem é negativa, o que pode parecer estranho uma vez que a agricultura é fonte de nitrato. Entretanto a categoria de LULC agricultura e pastagem (AP, na ECA, ocorre a uma maior distância do curso de água. O aumento nesta distância faz com que haja um filtro maior para a poluição entre o curso de água e as plantações propriamente ditas. Por outro lado a pastagem afastada do curso de água, é relacionada ao efeito de filtro de poluição, em função da retenção de nutrientes escoados com o escoamento superficial, efeito este também verificado por Ouyang *et al.* (2010).

8.6.2 Fósforo

Para o fósforo não foi possível definir modelo que representasse a carga difusa (DL) para nenhuma das 5 áreas avaliadas (faixas e ECA). Isto ocorreu em função da existência de

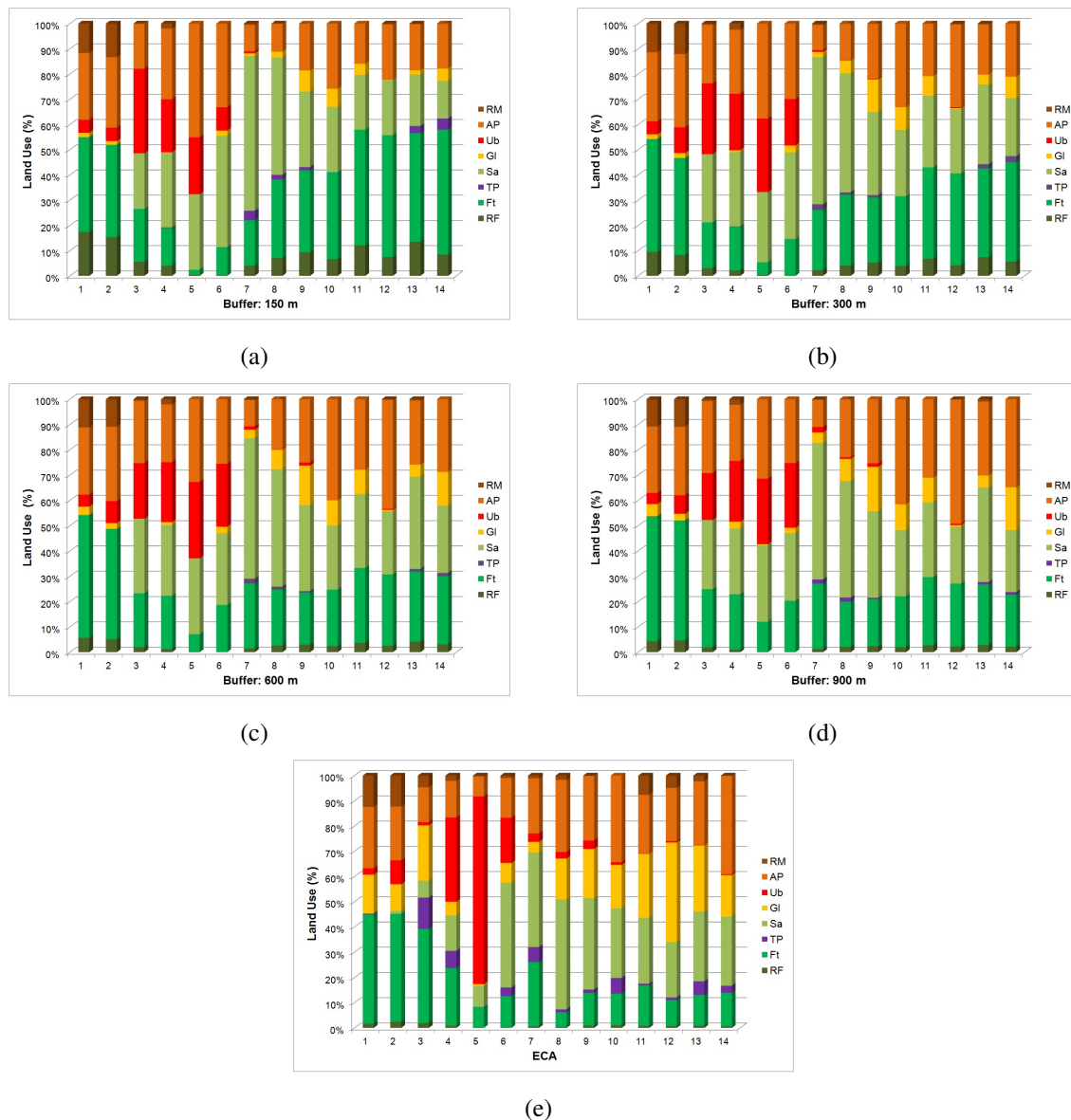


Figura 8.5: Uso e Ocupação do solo (%) por área de contribuição exclusiva e faixas: (a) 150 metros, (b) 300 metros, (c) 600 metros, (d) 900 metros and (e) ECA

cargas difusas negativas quando aplicamos a equação 8.4, ou seja, a carga pontual (seca) foi superior a carga difusa + pontual (chuva).

A Tabela 8.5 mostra que as cargas difusas “negativas” ocorreram nos períodos 2, 3 e 5 da Figura 8.3, que foram períodos de menor vazão. Este resultado vem de encontro aos resultados obtidos por Outram *et al.* (2016) que indicaram a queda acentuada no transporte de fósforo (P) durante anos mais secos.

Segundo Davis *et al.* (2014) e Outram *et al.* (2016), em anos mais secos ocorre o acúmulo de nutrientes no solo e esses são posteriormente liberados em anos mais chuvosos. Na Tabela 8.5 é possível observar que o (4º) quarto ano (chuvoso), ocorreu após dois anos secos (2º e 3º anos), e resultou nas maiores cargas obtidas nos 5 períodos, considerando as estações de

Tabela 8.5: Carga Difusa de Fósforo Total (DLTP) para as estações de qualidade nos 5 períodos avaliados

Período Avaliado (2008/2013)	Estações de Qualidade (WQ)						
	1	2	3	4	5	6	7
1	972.7	2003.9	1391.2	402.4	2163.0	310.8	302.7
2	192.8	294.6	132.4	-718.9	-379.1	-2355.0	-831.2
3	99.7	109.5	355.8	-674.9	-643.8	495.8	-374.3
4	511.0	530.0	227.3	391.2	413.9	4133.8	868.8
5	1363.8	1339.8	883.5	603.1	-176.6	-843.3	-1113.8
Continuação	8	9	10	11	12	13	14
1	901.0	3282.7	511.1	1164.8	9705.1	5352.6	4117.7
2	583.4	149.9	781.3	-347.8	915.3	201.1	-45.9
3	-646.3	8561.6	-1263.2	4532.3	3614.6	4252.9	2967.4
4	4668.1	8501.7	1309.1	7287.2	6354.2	6490.1	7602.0
5	177.6	278.8	85.7	1473.3	819.6	630.2	-48.2

6 a 14, exceto para a estação 12.

Analisando a Tabela 8.5, verifica-se que dos 15 valores negativos obtidos para a carga difusa de fósforo nos períodos de 2, 3 e 5, sete (7) valores estão contidos nas estações de 4 a 6 com maior proporção de uso do solo urbano U_b e os outros oito (8) valores ocorrem nas estações: 7, 8, 10, 11 e 14, com uso do solo na ECA preponderante por AP, S_a e G_t .

8.7 Conclusão

O estudo indica que a estimativa da carga difusa de nitrato, por meio da subtração das cargas de nutrientes do período chuvoso (ascensão hidrograma) das cargas do período seco pode ser uma boa aproximação para a carga difusa transportada em uma dada área de contribuição exclusiva (ECA) com diferentes categorias de uso e cobertura vegetal do solo (LULC).

Considerando as categorias de LULC, tomadas a diferentes distâncias do curso de água (ECA e faixas) e a carga difusa incremental unitária de nitrato ($UIDL_N$) geradas por estas áreas foram definidos modelos para o nitrato que apresentaram uma melhora substancial quando comparados ao trabalho de Oliveira *et al.* (2016) que utilizou dados de todo o período chuvoso. Para três das quatro faixas avaliadas (150, 300 e 600 metros) foi possível obter modelos, rejeitando a hipótese nula (H_0) para betas iguais a zero com $P < 0.05$ de significância. Nas 3 faixas o uso do solo “Urbano” apresentou correlação positiva com o nitrato. A categoria “Floresta” foi negativamente correlacionada com nitrato para as faixas de 150 e 300 metros.

Os modelos definidos para as faixas mais próximas ao curso de água (150 e 300 m) apresentaram floresta como categoria correlacionada negativamente com o nitrato, isto ressalta a importância da recomposição da mata ciliar nas margens dos cursos de água, para que a mesma funcione como um filtro de poluentes.

Para a faixa de 900 metros e a ECA a maior proporção uso urbano ocorreu nas estações 4 e 6. Por sua vez, as maiores cargas difusa incremental unitárias ($UIDL_N$) de nitrato positivas ocorreram para as estações 3 e 4 (4.39 e 6.30 kg/ km² dia) e negativas para as estações 5 e 6 (-

0.88 e -0.56 kg/ km² dia), respectivamente, o que contribuiu para que a variável U_b perdesse força no modelo.

Para a ECA o uso por agricultura e pastagem (AP), localizado a uma maior distância do curso de água ampliou a área de filtro entre curso de água e fonte poluidora estabelecendo uma relação negativa com as cargas unitárias de nitrato.

Quando avaliamos a UIDLN (Kg/ km² dia) produzida em determinada ECA, a existência de atividades de pastagem afastadas do curso de água podem auxiliar na retenção de nutrientes, o que também pode ser responsável pela relação negativa com as alterações do nitrato, resultado semelhante foi obtido por Ouyang *et al.* (2010).

Outro aspecto importante com relação à abordagem da poluição difusa pela diferença entre as carga de nitrato na chuva (LN_1) e na seca (LN_2) é que obtivemos 4 possíveis modelos, em contraposição a 1 modelo obtido por de Oliveira *et al.* (2016) que utilizou os dados de todo o período chuvoso. Os modelos obtidos selecionaram categorias de uso do solo relacionadas como fontes para poluição em diversos estudos presentes na literatura (Maillard e Santos, 2008; Ouyang *et al.*, 2010; Li *et al.*, 2015).

Para a faixa de 150 metros o modelo obtido com as categorias Ub e Ft, melhora em 51% o coeficiente de determinação obtido por de Oliveira *et al.* (2016) (de 0.31 para 0.47). Também foi possível obter modelos para as faixas de 300 e 600 metros e para a ECA, diferente de de Oliveira *et al.* (2016), que não encontrou modelos para estas áreas.

Para o nitrato o melhor resultado obtido foi para a faixa de 150 metros, o que indica que quando tratamos da UIDLN, ou seja, a carga produzida na área entre dois pontos de monitoramento, o uso mais próximo ao curso de água apresenta um melhor potencial de explicação para o nitrato, resultado semelhante aos obtidos por Maillard e Santos (2008); Li *et al.* (2015).

Para o fósforo não foi possível obtermos modelos que relacionassem as categorias de LULC com a carga difusa incremental unitária do fósforo ($UIDL_{TP}$), em função da existência de cargas difusas negativas em algumas estações quando aplicamos a equação 4, ou seja, a carga pontual (seca - L_{TP2}) foi superior a carga difusa + pontual (chuva - L_{TP1}). Entretanto julgamos os resultados importantes pois condizem com os resultados e afirmações apresentadas por Davis *et al.* (2014) e Outram *et al.* (2016).

As cargas difusas “negativas” ocorreram nos períodos 2, 3 e 5 da Figura 8.3, que foram períodos de menor vazão, períodos mais secos. O que condiz com os resultados obtidos por Outram *et al.* (2016) que indicaram uma queda acentuada no transporte de P durante o período de chuva em anos mais secos.

Segundo Davis *et al.* (2014) e Outram *et al.* (2016), em períodos mais secos ocorre o acúmulo de nutrientes no solo e esses são posteriormente liberados em anos mais chuvosos, o que também pode ser verificado pelos nossos dados.

A Tabela 8.5 mostra que o (4^o) quarto ano (chuvoso), ocorreu após dois anos secos (2^o)

e 3^o anos), e resultou nas maiores cargas obtidas nos 5 períodos, considerando as estações de 6 a 14, exceto para a estação 12.

Uma possibilidade para melhorar os resultados do fósforo total seria ampliar o período de coleta de dados e separar os anos secos dos anos chuvosos, entretanto, como o monitoramento mensal dos dados teve início somente em 2008 e os anos de 2013/2014 e 2014/2015 também foram anos secos na Bacia do Rio das Velhas é necessário que se aguardem anos mais chuvosos para avaliarmos a relação entre a carga incremental unitária de TP e LULC.

O modelo desenvolvido assumiu que o fósforo total (TP) é relativamente constante no curso de água, entretanto, conforme exposto por Bowes *et al.* (2008; 2014) a presença de TP, na água, está sujeita a processos que podem alterar a sua concentração, como interações com sedimentos e biota, e a suspensão/ressuspensão de fósforo particulado, efeito que também pode ter dificultado a obtenção de bons resultados para o Fósforo gerado de maneira difusa.

Outra abordagem possível é tentar avaliar a distribuição das chuvas nas diversas ECAs e tentar correlacionar a chuva a $UIDL_{TP}$, ao invés da vazão, conforme abordagem de Donner e Scavia (2007).

Sugerimos ainda que sejam realizadas coleta de amostras simultâneas de qualidade da água e vazão, quando da ocorrência de chuvas que geram o escoamento superficial, ou seja, provocam a ascensão dos níveis dos rios. Este monitoramento permitirá um maior conhecimento da contribuição da carga difusa.

Consideramos adequada a proposta de estimativa das cargas realizada neste trabalho, subtraindo a carga obtida para os períodos de ascensão dos hidrogramas na época chuvosa da carga obtida no período seco, uma vez que ela consegue se aproximar de resultados obtidos em trabalhos com maior nível de detalhamento dos dados, ou seja, com frequência de coleta superior a mensal (Davis *et al.*, 2014; Outram *et al.*, 2016). Desta forma pode ser um importante método para aplicação em bacias com escassez de dados de monitoramento.

8.8 Bibliographic References

The bibliography in page 134.

Capítulo 9

CONSIDERAÇÕES FINAIS

O objetivo principal desta tese foi analisar a relação entre o uso do solo e a liberação dos nutrientes fósforo e nitrogênio para as águas superficiais do Rio das Velhas, investigando as possibilidades e dificuldades dos métodos empregados. A tese teve três contribuições consideradas inovadoras: a aplicação de árvores de decisão, o índice LCPI e a proposta de um método para estimar a carga difusa e pontual, por meio da análise do fluviograma.

Para avaliar o atendimento às hipóteses formuladas no trabalho elaboramos os fluxogramas das Figuras 9.1, 9.2 e 9.3, que ilustram como cada artigo contribuiu para a análise e resultados obtidos. Assim, para lembrar o exposto na introdução, a tese se propôs a investigar as seguintes hipóteses:

1. É possível utilizar árvores de decisão no estudo da relação dos nutrientes nitrogênio e fósforo com as categorias de uso e ocupação do solo (LULC);
2. Modelos de regressão múltipla são adequados para relacionar LULC às alterações na carga de nitrogênio e fósforo;
3. A adoção de diferentes áreas de análise (bacia - WS, área de contribuição exclusiva - ECA ou Faixa - B) altera os resultados obtidos para a relação LULC e nutrientes avaliados;
4. A separação de cargas difusas das pontuais pode ser estimada a partir dos hidrogramas de cheia.

O uso de árvores de decisão (AD) mostrou-se promissor no estabelecimento das relações entre: o nitrato ($NO-3$) e o fósforo total (TP) com a vazão e as categorias de uso e ocupação do solo (LULC) Figura 9.1.

O atributo vazão foi analisado por meio de sua permanência no tempo ($PermQ$) e de um índice I_{Q-QA} , que relaciona a permanência da vazão com a permanência da concentração do parâmetro de qualidade avaliado (fósforo total ou nitrato). Definimos o índice I_{Q-QA} com base nas abordagens realizadas por Poor e McDonnell (2007); Bowes *et al.* (2008; 2014).

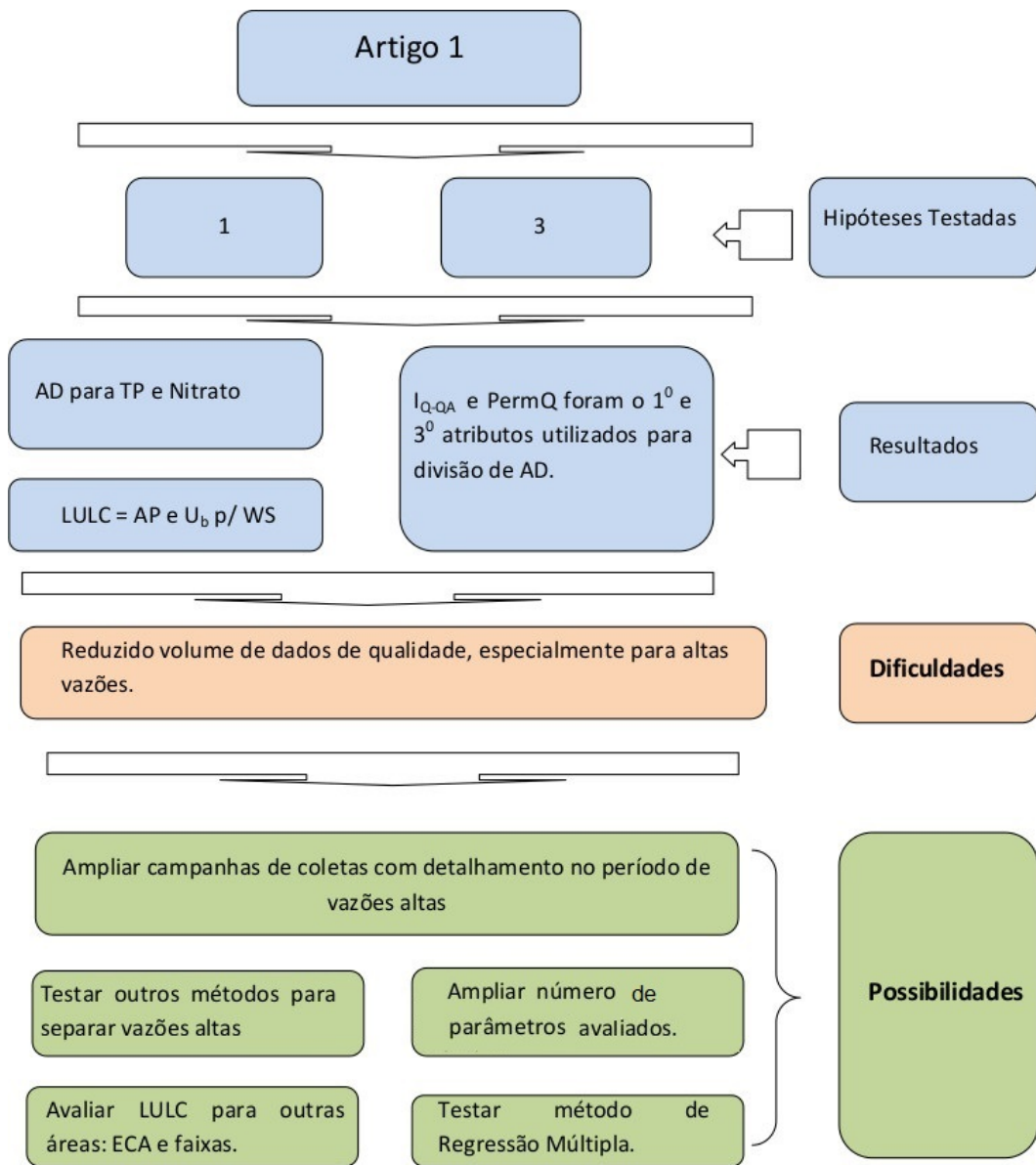


Figura 9.1: Hipóteses, Resultados, Dificuldades e Possibilidades obtidas com o Capítulo 4.

Os autores utilizaram dados de monitoramento de vazão e concentração de nutrientes durante eventos de chuva, para avaliar se o aumento da vazão produzia a “diluição” de um nutriente, o que pode indicar a dominância de uma fonte pontual, ou se, o aumento da vazão implicava em “concentração” da poluição, o que pode indicar a prevaência de fontes difusas.

O índice I_{Q-QA} e a permanência da vazão ($PermQ$) foram selecionados como primeiro e terceiro atributos para divisão dos dados e obtenção das árvores de decisão para o fósforo total (TP) e nitrato, respectivamente. O índice I_{Q-QA} foi o atributo que gerou o maior ganho de informação, seguido das categorias de uso do solo (segundo atributo)(Figura 9.1).

Em condição de cheia, as vazões possuem permanência baixa e os parâmetros de qualidade de água (nitrato e fósforo) vão apresentar permanência alta, quando sofrem diluição,

e permanência baixa, quando ocorre o aumento da concentração. Se considerarmos que as proporções entre a permanência da vazão e permanência da qualidade da água são semelhantes, teríamos $I_{Q-QA} \approx 1$ para fontes difusas e $I_{Q-QA} \ll 1$ para fontes pontuais.

O uso e ocupação do solo (LULC) por bacia hidrográfica possibilitou a escolha dos atributos Agropastoril (AP) e Urbano (U_b) como fontes poluidoras. Analisando as árvores geradas e os atributos selecionados para sua divisão (I_{Q-QA} , U_b , AP e $PermQ$) verificou-se que algumas folhas caracterizam o aspecto pontual da fonte poluidora (diluição - baixa $PermQA$) ou difuso (aumento concentração - alta $PermQA$).

A divisão da classe $PermQA$ utilizando o atributo I_{Q-QA} , com valor menor que 1, define a situação representativa de fonte pontual na chuva, onde o aumento do escoamento implica em redução na concentração do nutriente. Esta situação foi melhor definida para o fósforo total (TP), uma vez que valores menores ou iguais a 0.3 indicaram a classe urbano (U_b) como fonte poluidora. Um valor de $I_{Q-QA} \approx 1$ ocorreu para o fósforo total ($I_{Q-P} > 0.8$), associando fontes difusas à classe Agropastoril (AP). Para $0.3 < I_{Q-P} < 0.5$ não houve seleção de categoria de LULC, entretanto, o aumento da vazão ($PermQ \leq 18$) implicou em aumento na concentração de fósforo ($PermQA \leq 34.52$) característica de fonte difusa.

O reduzido número de dados de qualidade de água (QA), principalmente para as maiores vazões, dificultou a obtenção da relação entre as variáveis analisadas, principalmente para o nitrato. Assim, realizar campanhas com monitoramento simultâneo de vazão e QA, como o detalhamento da subida do fluviograma, é desejável para melhoria da classificação.

No primeiro artigo foram avaliados os dois trimestres mais chuvoso na Bacia do Rio das Velhas (Dez-Fev e Nov-Jan). Para esses dois trimestres foram selecionados os dados das variáveis de qualidade de água (fósforo total e nitrato) utilizados na elaboração das árvores de decisão. Em trabalhos futuros podemos alterar o modo de selecionar os dados, utilizando fluviogramas para identificar a ascensão da vazão de cheia, selecionando os dados de (QA) obtidos neste período de ascensão. Tal abordagem foi testada nos três últimos artigos e indica um caminho promissor.

A seleção de outras variáveis de qualidade é interessante de ser avaliada em novos testes com árvore de decisão (AD). Nos Artigos 2 e 3, inserimos novas variáveis na análise por regressão múltipla, o que nos auxiliou na interpretação dos resultados de fósforo total (TP) e nitrato.

No Artigo 1 o detalhamento do uso do solo com sua obtenção para outras datas não se mostrou necessário, haja visto termos utilizados somente 3 anos, centrados nos dados obtidos para LULC em 2010. Entretanto, em análises futuras, realizadas para períodos mais extensos, é desejável realizar a classificação de LULC para outras datas.

O recorte das categorias de LULC foi realizado tendo como base as bacias hidrográficas de cada ponto de monitoramento. Em trabalhos futuros a concentração poderia ser substituída por carga.

Para contrapor os resultados obtidos com a classificação realizada por árvore de decisão (AD), sugere-se o uso de outros métodos como regressão múltipla, análise de componentes principais e análise de cluster.

O uso da árvore de decisão (AD) demonstrou seu grande potencial para aplicação nesse tipo de análise. Entretanto, necessita de um volume maior de dados de monitoramento (vazão, qualidade). As sugestões de novos trabalhos utilizando AD são apresentadas como possibilidades na Figura 9.1 e itemizadas no final desta seção.

De posse dos resultados iniciais alterou-se a abordagem utilizada na seleção dos dados do período de chuvas. Assim, ao invés de trabalhar com trimestres chuvosos, optou-se pela separação dos dados de chuva por meio do uso da ascensão dos primeiros fluviogramas, estabelecendo esta data como o início do período chuvoso. Tal abordagem se mostrou mais adequada uma vez que o início do período de chuvas pode ser antecipado ou retardado em cada ano, assim a análise dos fluviogramas pode resultar em maior precisão na seleção dos dados de qualidade durante o período de maiores vazões.

No Artigo 2 foram adicionados dois novos parâmetros (Amônia - TAN e oxigênio dissolvido - OD), o uso do solo foi reamostrado por área de contribuição específica e faixas, e substituímos a concentração de cada parâmetro pela sua carga.

Para obter a carga, em cada ponto de monitoramento de qualidade, nos deparamos com a falta de dados de vazão em algumas estações. Para contornar este problema realizamos a transferência de vazões dos postos fluviométricos para os pontos de monitoramento da qualidade de água (QA), por meio de correlação entre áreas de drenagem. Na seleção da estação fluviométrica a ser utilizada para cada estação de QA, estabelecemos que a relação entre suas respectivas áreas de drenagem ficasse entre 0.7 e 1.3, de acordo com o sugerido por Euclides (2005). Esta abordagem foi selecionada por ser mais conservadora que a proposta por Brasília (2000), que prevê o uso de estações com relação entre áreas de drenagem de 0.25 a 4 e também que a proposta por Diebel *et al.* (2009), que prevê a seleção de estações com relação entre áreas de drenagem de 0.5 a 1.5.

A Bacia do Rio das Velhas apresenta duas regiões hidrologicamente heterogêneas e um volume expressivo de captações ao longo da calha principal do rio e em seus afluentes (Euclides, 2005). Assim para realizar a transferência dos dados de vazão entre dois pontos e garantir que não ocorram valores super ou sub-estimados de vazão e também da carga obtidas para os parâmetros avaliados, utilizamos a carga por área de contribuição exclusiva (ECA), denominada de carga incremental unitária ($UILx$). A ($UILx$) foi obtida da subtração da carga de jusante da carga de montante e posterior divisão pela área de cada ECA, onde x representa os parâmetros de qualidade de água avaliados nos três últimos artigos. Esta abordagem permitiu compreender a evolução das cargas ao longo do Rio das Velhas em cada ECA, ou seja, se a área existente entre dois pontos de monitoramento de qualidade de água promove o aumento ou a redução na carga poluidora, eliminando assim possíveis erros de

vazão.

Outra modificação realizada foi a substituição do uso de árvore de decisão por regressão múltipla, utilizando o valor mediano de cada parâmetro de qualidade durante o período avaliado (2001-2013) como variável dependente, ao invés da permanência da concentração. Apesar das alterações realizadas os resultados obtidos confirmam e ampliam as análises do primeiro artigo.

Os testes de *Kolmogorov - Smirnov* e *Shapiro-Wilk* foram utilizados para analisar se as variáveis dependentes apresentavam distribuição normal, tendo sido atestada a distribuição normal das cargas incrementais unitárias de: amônia, OD e nitrato.

A regressão múltipla entre a carga incremental unitária e LULC gerou 11 modelos lineares para: amônia, OD e nitrato. Os modelos foram gerados considerando os percentuais de LULC para a ECA e para as faixas, sendo possível verificar as categorias de LULC que funcionavam como filtro ou fontes de poluição (Figura 9.2).

Uma dificuldade encontrada no segundo artigo diz respeito ao efeito de colinearidade entre as variáveis independentes definidas pelas categorias de LULC, este problema foi maior para definir os modelos das faixas. Nas faixas as proporções entre as categorias de LULC, para cada estação, foram bastante semelhantes. Desta forma, os modelos obtidos para OD, nitrato e amônia nas faixas apresentaram baixo R^2 .

Para o nitrato foi possível obter somente um modelo, para a faixa de 150 metros e para fósforo total (TP) nenhum modelo foi obtido. As considerações para estes baixos resultados serão realizadas após os comentários do terceiro artigo.

Para melhorar os resultados obtidos listamos como possibilidades (Figura 9.2) a realização de campanhas simultâneas com coletas dos parâmetros de qualidade de água e vazão, com o detalhamento das vazões de cheia o que possibilitará a melhor definição da concentração para as maiores vazões; obtenção do uso do solo para outras datas e avaliação do uso de um índice relacionando categorias de LULC como fonte e filtro de poluição, o que minimiza o efeito de colinearidade.

Em termos de ganho obtido do Artigo 1 para o Artigo 2, temos que ao se trabalhar com a carga, em contraposição a permanência da concentração de cada parâmetro de qualidade de água, continuamos com a indicação de AP e U_b como fontes de poluição. Ou seja, a obtenção dos dados para o trimestre mais chuvoso e para o período de chuvas, com início determinado pelo fluviograma, apontou as mesmas categorias de uso do solo como fontes de poluição. Entretanto, ao considerarmos o período chuvoso, delimitado com o auxílio do fluviograma, foi possível identificar a vegetação (F_t e R_F) como filtro de poluição para TAN, NO_3^- , TP e OD, considerando o coeficiente de Pearson, embora o coeficiente tenha sido significativo somente para TAN e OD. Os coeficientes Tau Kendall e Rho de Spearman também foram calculados e indicaram sinal semelhantes a Pearson, entretanto, a magnitude foi menor para ambos.

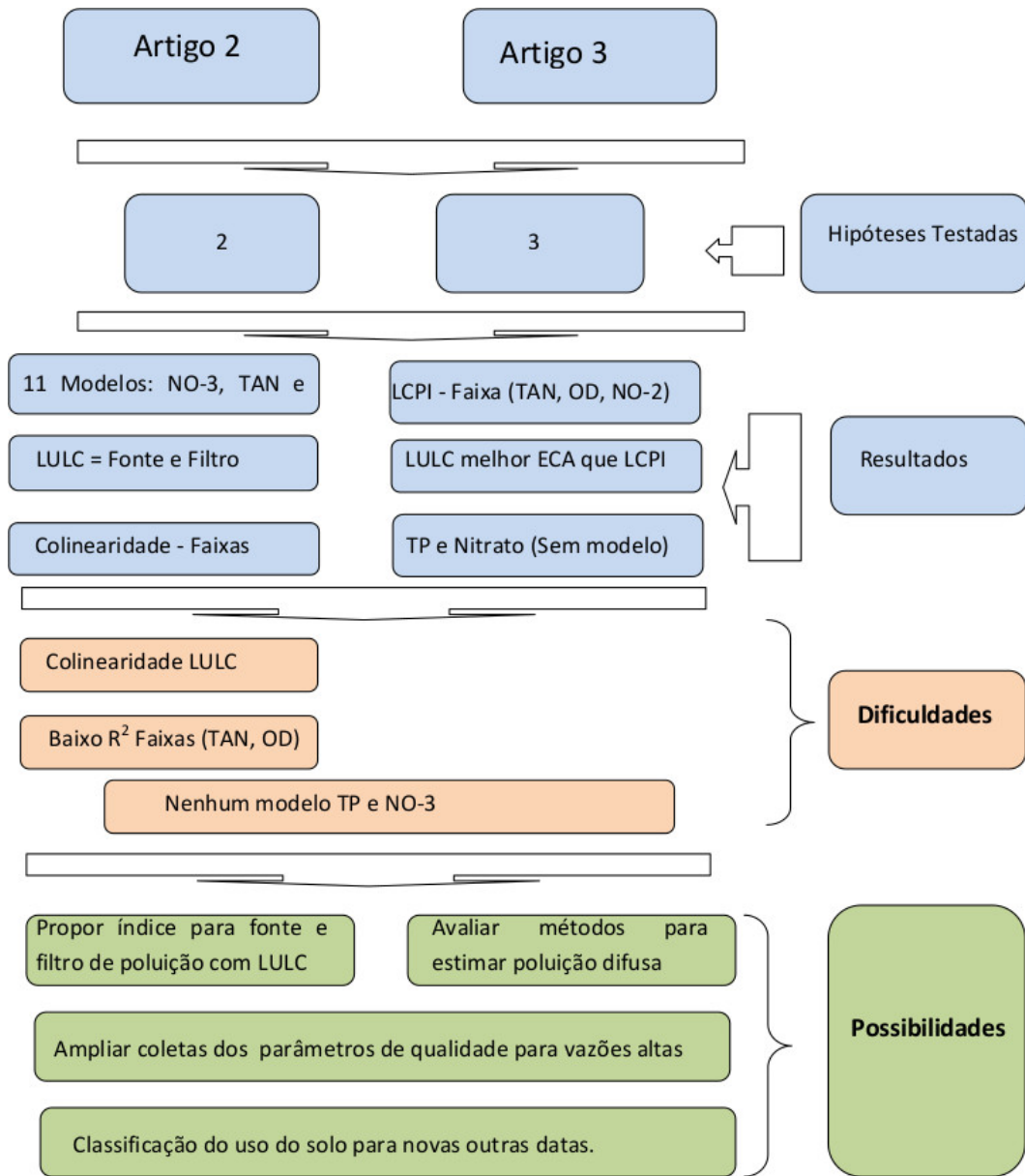


Figura 9.2: Hipóteses, Resultados, Dificuldades e Possibilidades obtidas com os Capítulos 5 e 6.

Conforme exposto por Helsel e Hirsch (1992), menores valores de correlação obtidos para Tau Kendall e Rho de Spearman não significam que estes sejam menos sensíveis que R de Pearson, mas simplesmente que existe uma diferença na escala de correlação utilizada nos três métodos.

A abordagem por área de contribuição exclusiva (ECA) e faixa (B), ao invés de bacia (WS), manteve o indicativo das fontes de poluição (AP e U_b) e acrescentou os filtros (F_t e R_F). Entretanto, a análise do coeficiente de Pearson indicou a existência de relação entre algumas variáveis independentes associadas as categorias de uso do solo avaliadas para as

faixas, o que gerou colineariedade entre as variáveis. Para resolver este problema selecionamos somente uma das categorias tidas como filtro de poluição (F_t e R_F) para elaborar os modelos. Esta solução nos garantiu termos fontes e filtros de poluição para análise.

O terceiro artigo resultou da tentativa de melhorar os resultados obtidos com o segundo artigo, levando em conta que as análises conduzidas indicaram (Artigo 2) que as categorias AP e U_b foram selecionadas como fontes de poluição e F_t e R_F como filtros e também no fato de ser necessário a redução do efeito de colineariedade entre as categorias de uso do solo. Assim optou-se por propor um índice que relacionasse fonte e filtro de poluição.

Outro aspecto ampliado com o terceiro artigo foi a abordagem de outra forma do nitrogênio, o nitrito ($NO - 2$), com a intenção de verificar se este parâmetro reproduziria o comportamento do nitrato ou da amônia. Para amônia (TAN) obtivemos bons resultados com a abordagem para o período de chuvas, mas não conseguimos os mesmos resultados para o nitrato. Como o nitrito representa uma fase intermediária da transformação da amônia em nitrato, optou-se por sua inserção na análise.

Considerando as faixas, o índice LCPI (índice de poluição por cobertura vegetal) melhorou os resultados obtidos para TAN, OD e $NO - 2$, quando comparado ao uso das categorias de uso do solo isoladamente. Os ganhos variaram de 11 a 155%, considerando os valores obtidos para R^2 . Entretanto, não obtivemos os mesmos resultados para a abordagem por ECAs, principalmente para a amônia. Este resultado já era esperado um vez que o coeficiente de Pearson para a ECA indica que as fontes poluidoras são AP e U_b . Entretanto, U_b apresenta uma forte correlação com a amônia, indicando ser a fonte pontual importante para as alterações na amônia na ECA e assim nenhum filtro foi selecionado, haja visto que para fontes pontuais o efeito da vegetação é insignificante.

O Artigo 3 apresentou uma solução adequada para as faixas, pois nestas áreas normalmente as categorias Floresta e Mata Ciliar vão apresentar comportamento semelhante ao longo do rio, o que produz o efeito de colineariedade, assim o agrupamento das categorias na forma de um índice melhorou os resultados. Em se tratando da ECA, os efeitos de colineariedade foram menores e o índice não mostrou melhoria nos resultados.

Os modelos obtidos para o nitrato no segundo e terceiro artigos, apresentaram baixo valor de R^2 o que podem estar associado a conversão da amônia em nitrato ao longo do rio, o que faz com que a carga incremental unitária de nitrato, por ECA, não tenha correspondência com as categorias de LULC da ECA. O fato de trabalhar com a carga incremental unitária (UIL) para cada ECA, busca minimizar este efeito promovido principalmente pelas fontes pontuais. Entretanto esta abordagem não foi suficiente para minimizar esse problema. Para o fósforo a sua adsorção ao sedimento e posterior liberação, pode também ser a causa da dificuldade de estabelecer sua relação com LULC, o que condiz com os estudos de Golterman (2007); Gibson e Meyer (2007); Bowes *et al.* (2008; 2014).

Para minimizar ou eliminar as dificuldades encontradas, avaliamos como possibilidade,

de melhoria dos resultados, a realização de campanhas de coletas dos parâmetros de qualidade de água, mais detalhadas para vazões máximas, assim como obter o uso do solo para outras datas. A avaliação de novos métodos para a obtenção da carga difusa de nitrato e TP, também pode ser uma alternativa.

Desta forma no 4^o artigo (Fig. 9.3) buscou-se uma nova abordagem para os parâmetros nitrato e fósforo baseada nos estudos de Donner e Scavia (2007); Poor e McDonnell (2007); Stutter *et al.* (2008); Outram *et al.* (2014); Bowes *et al.* (2008; 2014) que avaliaram o transporte de nitrato e fósforo durante eventos de chuva e encontraram que o transporte da maior quantidade desses nutrientes ocorre durante a ascensão do hidrograma. Já Bowes *et al.* (2008; 2014) propuseram um modelo para estimar as cargas difusa e pontual de nitrato e fósforo utilizando dados de concentração e vazão. Assim, optou-se por utilizar o hidrograma para obter as datas de monitoramento da qualidade (QA) coincidentes com a ascensão do hidrograma (chuva) e com o rio em escoamento de base (seca).

Supomos que durante a ascensão do hidrograma tínhamos poluentes de origem pontual e difusa e na situação de escoamento de base somente poluentes pontuais. Assumindo que os lançamentos pontuais são constantes (Bowes *et al.*, 2008; 2014), propusemos uma abordagem em que a carga de nitrato e fósforo obtida para o período de chuva pudesse ser subtraída da carga obtida para a seca, de forma a obtermos somente a carga difusa de um dado parâmetro para cada ponto avaliado.

A estimativa da carga difusa de nitrato, por meio da subtração das cargas do período chuvoso (ascensão hidrograma) das cargas do período seco (menor vazão na seca) foi uma boa aproximação. A Carga Difusa Incremental Unitária (UIDL_x) manteve o sinal da Carga Incremental Unitária (UIL) obtida nos artigos 2 e 3, para 12 das 14 estações avaliadas. Somente duas estações que apresentavam ECAs com carga positiva de nitrato no Artigo 2 passaram a ter carga negativa com a nova abordagem. E uma passou de negativa a positiva.

Para o fósforo não foi possível obter modelos que relacionassem as categorias de LULC com a carga difusa incremental unitária do fósforo ($UIDL_{TP}$), uma vez que obtivemos cargas difusas “negativas” para TP. As cargas difusas “negativas” ocorreram nos períodos chuvosos com menor vazão, o que condiz com os resultados obtidos por Davis *et al.* (2014) e Outram *et al.* (2016) que indicaram uma queda acentuada no transporte de fósforo total (TP) durante o período de chuva, em anos mais secos. Os autores afirmaram que em anos com menor quantidade de precipitação ocorre o acúmulo de nutrientes no solo e esses são posteriormente liberados em anos mais chuvosos, o que também pode ser verificado pelos nossos dados com o aumento da carga de TP difusa no 5^o ano.

De maneira geral, o Artigo 4 indica que, para a continuidade dos estudos de estimativa das cargas difusas de nitrato e fósforo total, deve-se avaliar inicialmente a dinâmica de acúmulo de nutrientes no solo em anos de menor precipitação, principalmente, com relação ao fósforo total (TP). Uma solução é a realização de um monitoramento de dados dinâmico,

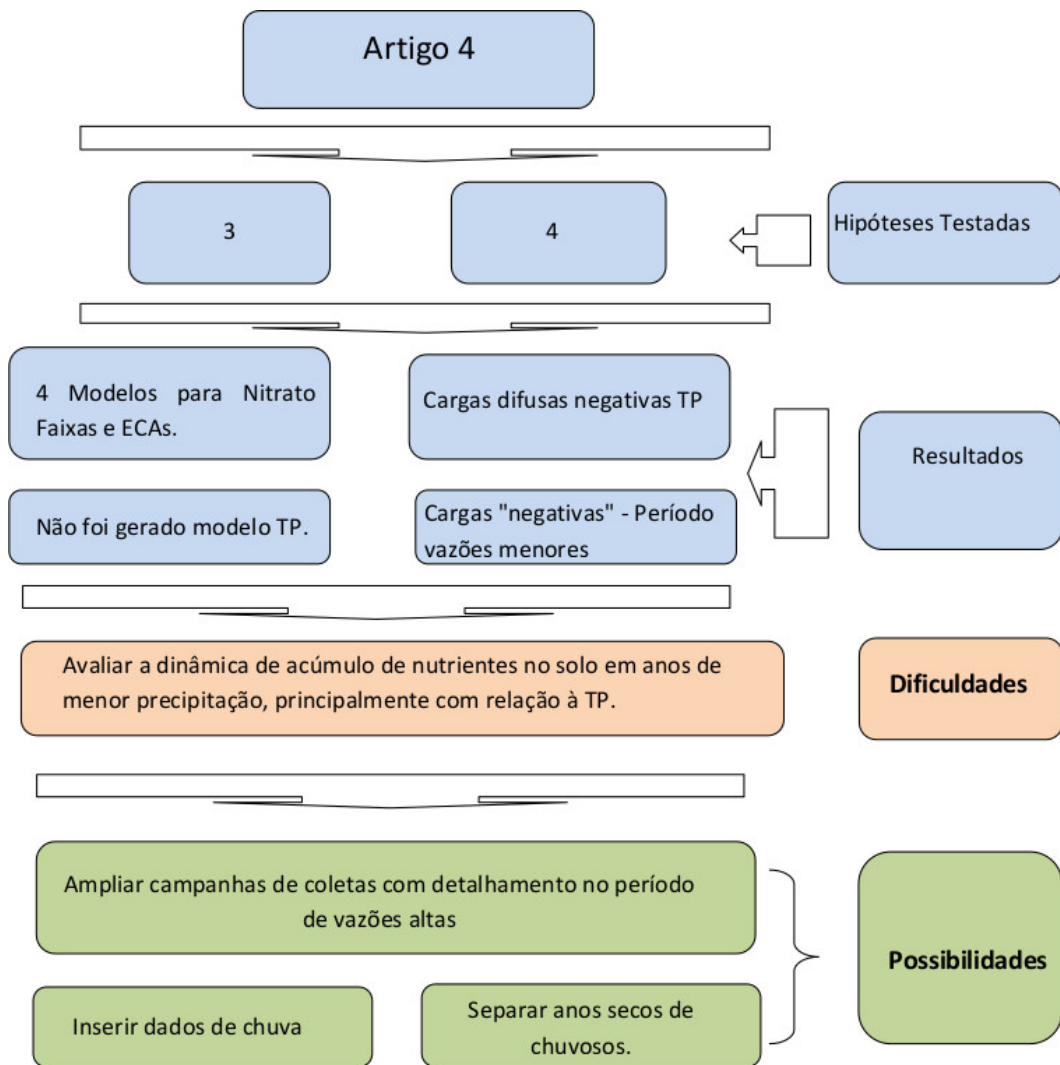


Figura 9.3: Hipóteses, Resultados, Dificuldades e Possibilidades obtidas com o Capítulo 7

por meio do uso de sistemas automatizado para coleta de informações de vazão (Q) e qualidade de água (QA). Conhecer a vazão em tempo real e realizar campanhas extras de monitoramento de qualidade de água, possibilitará avaliarmos o comportamento das variáveis em condições extremas (seca e chuva). Inserir dados de climatologia, como a chuva efetiva e prevista, pode auxiliar na avaliação de quais anos foram mais ou menos chuvosos, possibilitando inclusive a separação destes períodos e também pode ser utilizada para definir equipe de campo para a intensificação da coleta de dados, respectivamente.

Com base nas considerações apresentadas nesta seção, elaboramos a seção seguinte sintetizando todas as nossas sugestões para trabalhos futuros.

9.1 Sugestões de Trabalhos Futuros

Conforme indicado nas Figuras 9.1, 9.2 e 9.3, elaboradas com base nos 4 artigos desenvolvidos, identificamos como dificuldade a baixa quantidade de dados de qualidade de água (QA), mas especificamente quando da ocorrência de altas vazões; a necessidade de monitoramento conjunto de vazão-qualidade e a obtenção do uso do solo para outras datas.

Em 2008 o IGAM ampliou a frequência de monitoramento na bacia para mensal e está incorporando ao monitoramento da qualidade de água à medição de vazão. As imagens RapidEye, desde 2011, são compradas continuamente para todo o território nacional pelo Ministério do Meio Ambiente (MMA) e disponibilizadas para instituições públicas, universidades e institutos tecnológicos (MMA, 2014).

Assim, sugerimos como possibilidades de continuidade dos estudos as seguintes propostas:

1) - Para o primeiro artigo pode-se testar:

- Substituir o critério de trimestre chuvoso pela separação dos dados de cheia por meio da análise dos fluviogramas, identificando as datas de monitoramento de QA que coincidam com a ascensão do hidrograma;
- Testar outros parâmetros de qualidade para obtenção das árvores de decisão;
- Gerar árvores de decisão substituindo a concentração pela carga de cada parâmetro;
- Incorporar dados de chuva nas análises.

2)- Para o segundo e terceiro artigos sugere-se:

- Incorporar outros parâmetros de qualidade na análise de regressão múltipla;
- Adicionar dados de chuva às análises;
- Utilizar análise de componentes principais na seleção das categorias de uso e ocupação do solo, em substituição ao coeficiente de Pearson.

3)- As análises realizadas no quarto artigo podem ser ampliadas com:

- Inserção de dados de chuva;
- Aplicação da metodologia de Bowes *et al.* (2008) para obter as cargas difusa e pontual entre 2008 e 2013 e comparar os resultados com os obtidos neste trabalho;
- Incorporar outros parâmetros de qualidade na análise de regressão múltipla;
- Utilizar análise de cluster para tentar agrupar as estações em função das cargas difusas e pontuais geradas em cada ponto de monitoramento.

Referências Bibliográficas

- Ali, M., A. M. Qamar, e B. Ali (2013). Data analysis, discharge classifications, and predictions of hydrological parameters for the management of rawal dam in pakistan. In *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, Volume 1, pp. 382–385. IEEE.
- Almeida, K. C. d. B. (2013). *Avaliação da rede de monitoramento de qualidade das águas superficiais da Bacia do Rio das Velhas utilizando o método da entropia*. Belo Horizonte: Universidade Federal de Minas Gerais - UFMG.
- Arheimer, B. e R. Liden (2000). Nitrogen and phosphorus concentrations from agricultural catchments influence of spatial and temporal variables. *Journal of Hydrology* 227(1), 140–159.
- Arnold, J. G., R. Srinivasan, R. S. Muttiah, e J. R. Williams (1998). Large area hydrologic modeling and assessment part I: Model development. *JAWRA Journal of the American Water Resources Association* 34(1), 73–89.
- Atkins, J. P., D. Burdon, e J. H. Allen (2007). An application of contingent valuation and decision tree analysis to water quality improvements. *Marine Pollution Bulletin* (55), 592–602.
- Baird, J. e K. Whitelaw (1992). Water quality aspects of estuary modeling. *Water quality modeling*, 119–126.
- Banon, L. C. (2013). *Árvores de decisão aplicadas à extração automática de redes de drenagem*. Ph. D. thesis, Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos - Brasil.
- Basgalupp, M. P. (2010). *LEGAL-Tree: Um algoritmo genético multi-objetivo lexicográfico para indução de árvores de decisão*. Ph. D. thesis, ICMC-USP. (In Portuguese).
- Basnyat, P., L. D. Teeter, K. M. Flynn, e B. G. Lockaby (1999). Relationships between landscape characteristics and nonpoint source pollution inputs to coastal estuaries. *Environmental Management* 23(4), 539–549.

- Basu, N. B., G. Destouni, J. W. Jawitz, S. E. Thompson, N. V. Loukinova, A. Darracq, S. Zarnardo, M. Yaeger, M. Sivapalan, A. Rinaldo, *et al.* (2010). Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity. *Geophysical Research Letters* 37(23).
- Bhattacharya, B. e D. P. Solomatine (2005). Neural networks and m5 model trees in modeling water level–discharge relationship. *Neurocomputing* 63, 381–396.
- Bowes, M., H. Jarvie, S. Halliday, R. Skeffington, A. Wade, M. Loewenthal, E. Gozzard, J. Newman, e E. Palmer-Felgate (2015). Characterising phosphorus and nitrate inputs to a rural river using high-frequency concentration–flow relationships. *Science of the Total Environment* 511, 608–620.
- Bowes, M. J., H. P. Jarvie, P. S. Naden, G. H. Old, P. M. Scarlett, C. Roberts, L. K. Armstrong, S. A. Harman, H. D. Wickham, e A. L. Collins (2014). Identifying priorities for nutrient mitigation using river concentration–flow relationships: The thames basin, uk. *Journal of Hydrology* 517, 1–12.
- Bowes, M. J., J. T. Smith, H. P. Jarvie, e C. Neal (2008). Modelling of phosphorus inputs to rivers from diffuse and point sources. *Science of the total Environment* 395(2), 125–138.
- Brasília, E. C. E. B. S. (2000). *Diretrizes para Estudos e Projetos de Pequenas Centrais Hidrelétricas. Capítulo 6 - Estudos Básicos*. belo Horizonte, Brasil: Eletrobrás.
- Breiman, L., J. Friedman, C. J. Stone, e R. A. Olshen (1984). *Classification and regression trees*. CRC press.
- Brites, A. P. Z. (2010). *Enquadramento dos corpos de água através de metas progressivas: probabilidade de ocorrência e custos de despoluição hídrica*. Ph. D. thesis, Universidade de São Paulo.
- Bu, H., W. Meng, Y. Zhang, e J. Wan (2014). Relationships between land use patterns and water quality in the taizi river basin, china. *Ecological Indicators* 41, 187–197.
- Buck, O., D. K. Niyogi, e C. R. Townsend (2004). Scale-dependence of land use effects on water quality of streams in agricultural catchments. *Environmental Pollution* 130(2), 287–299.
- Calazans, G. M. (2015). *Avaliação e proposta de adequação da rede de monitoramento da qualidade das águas superficiais das sub-bacias do Rio das Velhas e do Rio Paraopeba utilizando técnicas estatísticas multivariadas*. Belo Horizonte: Universidade Federal de Minas Gerais - UFMG.

- CBH (2015). *Plano diretor de recursos hídricos da Bacia Hidrográfica do Rio das Velhas: resumo executivo dezembro 2015*. Belo Horizonte: Comitê da Bacia Hidrográfica do Rio das Velhas.
- Chang, H. (2008). Spatial analysis of water quality trends in the han river basin, south korea. *Water Research* 42(13), 3285–3304.
- Chau, K.-w. (2006). A review on integration of artificial intelligence into water quality modelling. *Marine Pollution Bulletin* 52(1), 726–733.
- Cheng, K.-F., L. Cheng, e Y.-D. Huang (2012). Water quality evaluation method based on j48 decision tree algorithm [j]. *Computer Engineering* 11, 080.
- Chesterikoff, A., B. Garban, G. Billen, e M. Poulin (1992). Inorganic nitrogen dynamics in the river seine downstream from paris (france). *Biogeochemistry* 17(3), 147–164.
- Cleland, B. R. (2003). Tmdl development from the “bottom up”–part iii: Duration curves and wet-weather assessments. *Proceedings of the Water Environment Federation* 2003(4), 1740–1766.
- Cunha, D. G. F., M. do Carmo Calijuri, e E. M. Menciondo (2012). Integração entre curvas de permanência de quantidade e qualidade da água como uma ferramenta para a gestão eficiente dos recursos hídricos. *Engenharia Sanitaria e Ambiental* 17(4), 369–376.
- Davis, C. A., A. S. Ward, A. J. Burgin, T. D. Loecke, D. A. Riveros-Iregui, D. J. Schnoebelen, C. L. Just, S. A. Thomas, L. J. Weber, e M. A. St Clair (2014). Antecedent moisture controls on stream nitrate flux in an agricultural watershed. *Journal of environmental quality* 43(4), 1494–1503.
- de Oliveira, L. M., P. Maillard, e É. J. de Andrade Pinto (2016). Modeling the effect of land use/land cover on nitrogen, phosphorous and dissolved oxygen loads in the velhas river using the concept of exclusive contribution area. *Environmental Monitoring and Assessment* 188(6), 1–19.
- De’Ath, G. (2002). Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology* 83(4), 1105–1117.
- Diebel, M. W., J. T. Maxted, D. M. Robertson, S. Han, e M. J. Vander Zanden (2009). Landscape planning for agricultural nonpoint source pollution reduction iii: assessing phosphorus and sediment reduction potential. *Environmental Management* 43(1), 69–83.
- Dingman, S. L. (2004). *Physical hydrology*. Waveland press.

- Donner, S. D. e D. Scavia (2007). How climate controls the flux of nitrogen by the mississippi river and the development of hypoxia in the gulf of mexico. *Limnology and Oceanography* 52(2), 856–861.
- Euclides, H. P. O. (2005). Atlas digital das águas de minas; uma ferramenta para o planejamento e gestão dos recursos hídricos. *RURALMINAS ISBN85-7601-082-8*, 78.
- Evans, D. M., S. H. Schoenholtz, P. J. Wigington Jr, S. M. Griffith, e W. C. Floyd (2014). Spatial and temporal patterns of dissolved nitrogen and phosphorus in surface waters of a multi-land use basin. *Environmental monitoring and assessment* 186(2), 873–887.
- Even, S., G. Billen, N. Bacq, S. Théry, D. Ruelland, J. Garnier, P. Cugier, M. Poulin, S. Blanc, F. Lamy, *et al.* (2007). New tools for modelling water quality of hydrosystems: An application in the seine river basin in the frame of the water framework directive. *Science of the Total Environment* 375(1), 274–291.
- Fisher, D., J. Steiner, D. Endale, J. Stuedemann, H. Schomberg, A. Franzluebbbers, e S. Wilkinson (2000). The relationship of land use practices to surface water quality in the Upper Oconee Watershed of Georgia. *Forest Ecology and Management* 128, 39–48.
- Formigoni, Y., A. P. Brites, C. Fernandes, e M. Porto (2011). Análise crítica da curva de permanência de qualidade de água com base em dados históricos. In *XIX Simpósio Brasileiro de Recursos Hídricos - Maceió-Brasil*.
- Gardi, C. (2001). Land use, agronomic management and water quality in a small northern italian watershed. *Agriculture, ecosystems & environment* 87(1), 1–12.
- Gibson, C. A. e J. L. Meyer (2007). Nutrient uptake in a large urban river. *Journal of the American Water Resources Association (JAWRA)* 43(3), 576–587.
- Golterman, H. L. (2007). *The chemistry of phosphate and nitrogen compounds in sediments*. Springer Science & Business Media.
- Gorsevski, P., E. G. J. Boll, e E. S. Brooks (2008). Dynamic riparian buffer widths from potential non-point source pollution areas in forested watersheds. *Forest Ecology and Management* 256, 664–673.
- Green, M. B., J. L. Nieber, G. Johnson, J. Magner, e B. Schaefer (2007). Flow path influence on an n: P ratio in two headwater streams: A paired watershed study. *Journal of Geophysical Research: Biogeosciences* 112(G3).
- Grizzetti, B., F. Bouraoui, e G. De Marsily (2008). Assessing nitrogen pressures on european surface water. *Global Biogeochemical Cycles* 22(4), 1–14.

- Grunwald, S., S. Daroub, T. Lang, e O. Diaz (2009). Tree-based modeling of complex interactions of phosphorus loadings and environmental factors. *Science of the total environment* 407(12), 3772–3783.
- Grunwald, S. e C. Qi (2006). Gis-based water quality modeling in the sandusky watershed, ohio, usa1. *JAWRA Journal of the American Water Resources Association* 42(4), 957–973.
- Haggard, B. E., D. E. Storm, e E. H. Stanley (2001). Effect of a point source input on stream nutrient retention. *Journal of the American Water Resources Association* 37(5), 1291–1299.
- Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, e I. H. Witten (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter* 11(1), 10–18.
- Han, J., Z. Li, P. Li, e J. Tian (2010). Nitrogen and phosphorous concentrations in runoff from a purple soil in an agricultural watershed. *Agricultural Water Management* 97(5), 757–762.
- Helsel, D. R. e R. M. Hirsch (1992). *Statistical methods in water resources*, Volume 49. Elsevier.
- Henrique, C. S. (2009). *Regionalização hidrológica de vazões e de cargas difusas de sólidos suspensos totais para o programa ProdutorES de Água eo pagamento de serviços ambientais. Estudo de caso: bacia hidrográfica do rio Benevente, ES*. Ph. D. thesis, Universidade de São Paulo.
- Howarth, R. W., E. W. Boyer, W. J. Pabich, e J. N. Galloway (2002). Nitrogen use in the united states from 1961-2000 and potential future trends. *AMBIO: A Journal of the Human Environment* 31(2), 88–96.
- Huang, J. L. e H. H. Hong (2010). Comparative study of two models to simulate difusse nitrogen and phosphorus pollution in a medium-sized watershed, southeast china. *Estuarine, Coastal and Shelf Science* 1(86), 387–394.
- IGAM (2005). *Plano diretor de recursos hídricos da bacia hidrográfica do Rio das Velhas: resumo executivo dezembro 2004/ Luíza de Marillac Moreira Camargos (coord.)*. Belo Horizonte: Instituto Mineiro de Gestão das Águas, Comitê da Bacia Hidrográfica do Rio das Velhas.
- IGAM (2013a). *Identificação de municípios com condição crítica para a qualidade de água na Bacia do Rio das Velhas*. IGAM - Instituto Mineiro de Gestão das Águas. Estado de Minas Gerais - Brasil.

- IGAM (2013b). *Monitoramento da Qualidade das Águas Superficiais no estado de Minas Gerais - 1^o Trimestre de 2013*. Belo Horizonte: Instituto Mineiro de Gestão das Águas -IGAM.
- IGAM, I. M. d. G. a. d. A. (2015). Relatório de monitoramento das Águas superficiais 2015 - resumo executivo 2015. belo horizonte.
- Instituto Nacional de Meteorologia (2004). <http://www.inmet.gov.br>. Página da internet visitada em 28 de outubro.
- Jardim, B. F. M. (2011). *Variação dos parâmetros físicos e químicos das águas superficiais da Bacia do Rio das Velhas-MG e sua associação com as florações de cianobactérias*. Ph. D. thesis, Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos - UFMG, Belo Horizonte - Brasil.
- Jarvie, H., P. Withers, M. Bowes, E. Palmer-Felgate, D. Harper, K. Wasiak, P. Wasiak, R. Hodgkinson, A. Bates, C. Stoate, *et al.* (2010). Streamwater phosphorus and nitrogen across a gradient in rural–agricultural land use intensity. *Agriculture, ecosystems & environment* 135(4), 238–252.
- Jarvie, H. P., C. Neal, e P. J. Withers (2006). Sewage-effluent phosphorus: a greater risk to river eutrophication than agricultural phosphorus? *Science of the Total Environment* 360(1), 246–253.
- Johnson, L., C. Richards, G. Host, e J. Arthur (1997). Landscape influences on water chemistry in midwestern stream ecosystems. *Freshwater Biology* 37(1), 193–208.
- Jordan, P., A. Arnscheidt, H. McGrogan, e S. McCormick (2007). Characterising phosphorus transfers in rural catchments using a continuous bank-side analyser. *Hydrology and Earth System Sciences Discussions* 11(1), 372–381.
- Kronvang, B. e A. Bruhn (1996). Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. *Hydrological processes* 10(11), 1483–1501.
- Lang, M., P. Li, e X. Yan (2013). Runoff concentration and load of nitrogen and phosphorus from a residential area in an intensive agricultural watershed. *Science of the Total Environment* 458, 238–245.
- Li, H., L. Liu, e X. Ji (2015). Modeling the relationship between landscape characteristics and water quality in a typical highly intensive agricultural small watershed, dongting lake basin, south central china. *Environmental monitoring and assessment* 187(3), 1–12.

- Liao, H. e W. Sun (2010). Foresting and evaluating water quality of chao lake based on an improved decision tree method. *Procedia Environmental Sciences* (2), 970–979.
- Liao, S.-H., P.-H. Chu, e P.-Y. Hsiao (2012). Data mining techniques and applications—a decade review from 2000 to 2011. *Expert Systems with Applications* 39(12), 11303–11311.
- Łowicki, D. (2012). Prediction of flowing water pollution on the basis of landscape metrics as a tool supporting delimitation of nitrate vulnerable zones. *Ecological Indicators* 23, 27–33.
- Lowrance, R., L. S. Altier, J. D. Newbold, R. R. Schnabel, P. M. Groffman, J. M. Denver, D. L. Correll, J. W. Gilliam, J. L. Robinson, R. B. Brinsfield, *et al.* (1997). Water quality functions of riparian forest buffers in chesapeake bay watersheds. *Environmental Management* 21(5), 687–712.
- Maillard, P. (2010). *Non-point pollution sources*. Encyclopedia.
- Maillard, P. e N. A. P. Santos (2008). A spatial-statistical approach for modeling the effect of non-point source pollution on different water quality parameters in the velhas river watershed–brazil. *Journal of Environmental Management* 86(1), 158–170.
- Manuelzão, P. (2014). *Sem ações sustentáveis os rios morrem. Edição especial*. Projeto Manuelzão.
- Mattikalli, N. M. e K. S. Richards (1996). Estimation of surface water quality changes in response to land use change: Application of the export coefficient model using remote sensing and geographical information system. *Journal of Environmental Management* 48, 263–282.
- McDowell, R., A. Sharpley, L. Condon, P. Haygarth, e P. Brookes (2001). Processes controlling soil phosphorus release to runoff and implications for agricultural management. *Nutrient Cycling in Agroecosystems* 59(3), 269–284.
- Mendonça, E. (2008). Challenging issues of urban biodiversity related to ecohydrology. *Brazilian Journal of Biology* 68(4), 983–1002.
- Meynendonckx, J., G. Heuvelmans, B. Muys, e J. Feyen (2006). Effects of watershed and riparian zone characteristics on nutrient concentrations in the river scheldt basin. *Hydrology and Earth System Sciences Discussions Discussions* 3(3), 653–679.
- Miller, J. N. (1993). Tutorial review-outliers in experimental data and their treatment. *Analyst* 118(5), 455–461.

- Ministério do Meio Ambiente (2008). <http://www.mma.gov.br>. Ministério do Meio Ambiente, Brasil. Último acesso: 09/01/2008.
- Moisen, G. G. e T. S. Frescino (2002). Comparing five modelling techniques for predicting forest characteristics. *Ecological modelling* 157(2), 209–225.
- Munson, K. M. (2015). *Modeling the Sensitivity of Total Phosphorus in an Urban Stream to Climate Variability*. Ph. D. thesis, TUFTS UNIVERSITY.
- Naghetini, M. e E. J. A. Pinto (2007). *Hidrologia estatística*. CPRM:Belo Horizonte.
- Nardi, F., S. Grimaldi, M. Santini, A. Petroselli, e L. Ubertini (2008). Hydrogeomorphic properties of simulated drainage patterns using digital elevation models: the flat area issue/propriétés hydro-géomorphologiques de réseaux de drainage simulés à partir de modèles numériques de terrain: la question des zones planes. *Hydrological Sciences Journal* 53(6), 1176–1193.
- Nonato, E. A., Z. G. Viola, K. C. B. Almeida, e H. H. R. Schor (2007). Tratamento estatístico dos parâmetros da qualidade das águas da bacia do alto curso do rio das velhas. *Química Nova* 30(4), 797.
- Outram, F., C. Lloyd, J. Jonczyk, C. M. Benskin, F. Grant, M. Perks, C. Deasy, S. Burke, A. Collins, J. Freer, *et al.* (2014). High-frequency monitoring of nitrogen and phosphorus response in three rural catchments to the end of the 2011–2012 drought in england. *Hydrology and Earth System Sciences* 18(9), 3429–3448.
- Outram, F. N., R. J. Cooper, G. Sünnerberg, K. M. Hiscock, e A. A. Lovett (2016). Antecedent conditions, hydrological connectivity and anthropogenic inputs: Factors affecting nitrate and phosphorus transfers to agricultural headwater streams. *Science of The Total Environment* 545, 184–199.
- Ouyang, W., A. Skidmore, A. Toxopeus, e f. Hao (2010). Long-term vegetation landscape pattern with non-point source nutrient pollution in upper stream of yellow river basin. *Journal of Hydrology* 389(1), 373–380.
- Pappenberger, F., I. Iorgulescu, e K. J. Beven (2006). Sensitivity analysis based on regional splits and regression trees (sars-rt). *Environmental Modelling and Software* 21(7), 976–990.
- Phillips, J. D. (1989). Nonpoint source pollution control effectiveness of riparian forests along a coastal plain river. *Journal of Hydrology* 110(3), 221–237.

- Pinheiro, A. e F. C. Deschamps (2008). Transporte de ortofosfato e de nitrato na microbacia do ribeirão fortuna, sc. *Revista Brasileira de Engenharia Agrícola e Ambiental* 12(3), 318–325.
- Pinheiro, A., C. Shoen, J. Schultz, K. G. H. Heinz, I. G. Pinheiro, e F. C. Deschamps (2014). Relação entre o uso do solo e a qualidade da água na bacia hidrográfica rural no bioma mata atlântica. *Revista Brasileira de Recursos Hídricos* 19(3), 127–139.
- Pinto, C. C. (2016). *Importância da Avaliação de Dados de Concentração e de Carga de Poluentes em Bacias Hidrográficas: Exemplo da Bacia do Rio Verde*. Belo Horizonte: Universidade Federal de Minas Gerais - UFMG.
- Pinto, E. d. A., A. d. Azambuja, J. Farias, J. d. B. Salgueiro, e K. Pickbrenner (2011). *Atlas pluviométrico do Brasil: isoietas mensais, isoietas trimestrais, isoietas anuais, meses mais secos, meses mais chuvosos, trimestres mais secos, trimestres mais chuvosos*. Brasília: CPRM.
- Polignano, M. V., E. Ibrahim, G. Alves, e N. Vieira (2014). *Meta 2014- Revitalização da Bacia do Rio das Velhas*. Projeto Manuelzão.
- Poor, C. J. e J. J. McDonnell (2007). The effects of land use on stream nitrate dynamics. *Journal of Hydrology* 332(1), 54–68.
- Pratt, B. e H. Chang (2012). Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales. *Journal of hazardous materials* 209, 48–58.
- Quilbé, R., A. N. Rousseau, M. Duchemin, A. Poulin, G. Gangbazo, e J.-P. Villeneuve (2006). Selecting a calculation method to estimate sediment and nutrient loads in streams: application to the beaurivage river (québec, canada). *Journal of Hydrology* 326(1), 295–310.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning* 1(1), 81–106.
- Quinlan, J. R. (1993). *C4. 5: programs for machine learning*, Volume 1. Morgan kaufmann.
- Quinn, J. M. e M. J. Stroud (2002). Water quality and sediment and nutrient export from new zealand hill-land catchments of contrasting land use. *New Zealand Journal of Marine and Freshwater Research* 36(2), 409–429.
- Raposo, A. A., L. F. de Paula Barros, e A. P. M. Júnior (2011). O uso de taxas de turbidez da bacia do alto rio das velhas–quadrilátero ferrífero/mg–como indicador de pressões humanas e erosão acelerada. *Revista de Geografia (Recife)* 27(3. Esp), 34–50.

- Roque, T. V. P., L. N. d. Andrade, R. Zorzal, e A. S. F. Mendonça (2008). *Metodologia para apoio à análise de solicitação de outorga de diluição de efluentes - análise e aplicação*. II Simpósio Brasileiro de Recursos Hídricos do Sul-Sudeste - Rio de Janeiro - Brasil.
- Salvia-Castellví, M., J. F. Iffly, P. Vander Borgh, e L. Hoffmann (2005). Dissolved and particulate nutrient export from rural catchments: a case study from luxembourg. *Science of the Total Environment* 344(1), 51–65.
- Santos, N. A. P. (2005). *Uma abordagem metodológica para determinar a influência do uso e da cobertura do solo como fonte de poluição difusa na alteração da qualidade da água na Bacia do Rio das Velhas*. Ph. D. thesis, Departamento de Geociências - UFMG, Belo Horizonte - Brasil.
- Schärer, M., T. Page, e K. Beven (2006). A fuzzy decision tree to predict phosphorus export at the catchment scale. *Journal of hydrology* 331(3), 484–494.
- Sharifi, S. e S. M. Hosseini (2011). Methodology for identifying the best equations for estimating the time of concentration of watersheds in a particular region. *Journal of irrigation and drainage engineering* 137, 712–719.
- Shen, Z., X. Hou, W. Li, e G. Aini (2014). Relating landscape characteristics to non-point source pollution in a typical urbanized watershed in the municipality of beijing. *Landscape and Urban Planning* 123, 96–107.
- Sliva, L. e D. D. Williams (2001). Buffer zone versus whole catchment approaches to studying land use impact on river water quality. *Water Resource* 35, 3462–3472.
- Solomatine, D. P. e K. N. Dulal (2003). Model trees as an alternative to neural networks in rainfall runoff modelling. *Hydrological Sciences Journal* 48(3), 399–411.
- Stutter, M., S. Langan, e R. Cooper (2008). Spatial contributions of diffuse inputs and within-channel processes to the form of stream water phosphorus over storm events. *Journal of Hydrology* 350(3), 203–214.
- Sun, R., L. Chen, W. Chen, e Y. Ji (2013). Effect of land-use patterns on total nitrogen concentration in the upstream regions of the haihe river basin, china. *Environmental management* 51(1), 45–58.
- Tu, J. (2011). Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. *Applied Geography* 31(1), 376–392.
- Tucci, C. (1993). *Hidrologia: Ciência e Aplicação*. Porto Alegre, RS, BR: 1-654, p.

- Tucci, C. E. *et al.* (1998). *Modelos hidrológicos*. Porto Alegre, RS, BR: UFRGS.
- Van Nieuwenhuysse, B. H., M. Antoine, G. Wyseure, e G. Govers (2011). Pattern-process relationships in surface hydrology: hydrological connectivity expressed in landscape metrics. *Hydrological Processes* 25(24), 3760–3773.
- Von Sperling, M. (2005). *Introdução à qualidade das águas e ao tratamento de esgotos*. Belo Horizonte, MG, Brasil: 1-452, p.
- Von Sperling, M. (2007). *Estudos e modelagem da qualidade da água de rios*. Belo Horizonte, MG, Brasil: 1-588, p.
- Wang, X. (2001). Integrating water-quality management and land-use planning in a watershed context. *Journal of Environmental Management* 61, 25–36.
- Wicker, W. C. (1979). Enforcement of section 208 of the federal water pollution control act amendments of 1972 to control nonpoint source pollution. *Land & Water L. Rev.* 14, 419.
- Wiens, J. (1980). Agricultural runoff and water pollution. *Canadian Water Resources Journal* 5(3), 78–89.
- Witten, I. H. e E. Frank (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Wood, F., A. L. Heathwaite, e P. Haygarth (2005). Evaluating diffuse and point phosphorus contributions to river transfers at different scales in the taw catchment, devon, uk. *Journal of Hydrology* 304(1), 118–138.
- Wozniak, C. C. (1980). Nonpoint source water pollution and the imposition of effluent limitations on point sources. *Rutgers Computer & Tech. LJ* 8, 117.
- Yang, J.-L., G.-L. Zhang, e Y.-G. Zhao (2007). Land use impact on nitrogen discharge by stream: a case study in subtropical hilly region of china. *Nutrient Cycling in Agroecosystems* 77(1), 29–38.
- Zaki, M. J. e W. Meira Jr (2014). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.
- Zhenyao, S., C. Lei, L. Qian, L. Ruimin, e H. Qian (2012). Impact of spatial rainfall variability on hydrology and nonpoint source pollution modeling. *Journal of Hydrology* 472-473, 205–215.
- Zucco, E., A. Pinheiro, F. C. Deschamps, e P. A. Soares (2012). Metodologia para estimativa das concentrações em cursos de água para vazões de referência: uma ferramenta de suporte e apoio ao sistema de gestão de bacias hidrográficas. *REGA* 9(1), 25–37.

Apêndice A

UTES Bacia Rio das Velhas

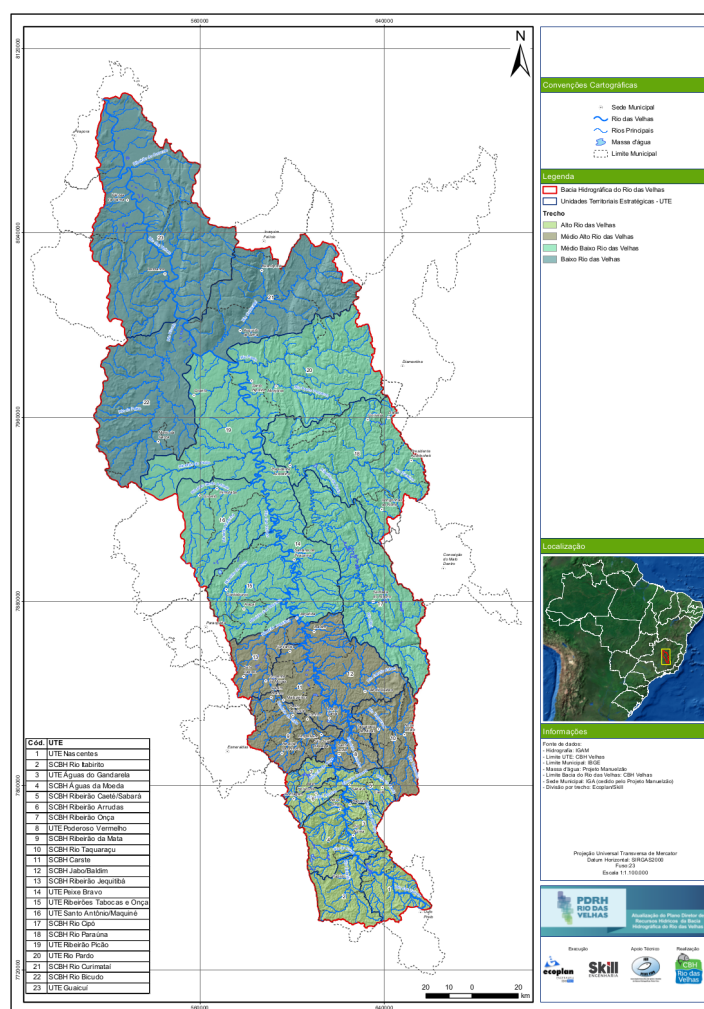


Figura A.1: UTES e Regiões da bacia do Rio das Velhas. Adaptado de CBH (2015)

Apêndice B

Exemplo de Árvore de Decisão

O exemplo desenvolvido na sequência busca apresentar o processo realizado para obtenção de árvores de decisão utilizando o algoritmo C4.5, proposto por Quinlan (1993). A base de dados de Fósforo Total (TP) e o uso do solo para a faixa de 100 metros, produzida para a elaboração do 2^o artigo da tese (tabela B.1), foi utilizada. O algoritmo C4.5 é disponibilizado no pacote de aplicativos WEKA (*The Waikato Environment for Knowledge Analysis*), em conjunto com uma série de algoritmos de aprendizagem de máquina e outras ferramentas para pré-processamento de dados (Hall *et al.*, 2009). No Weka, o algoritmo C4.5 é renomeado para J48.

As etapas de pré-processamento e geração da árvore de decisão são apresentadas na sequência:

1^a Etapa: Discretização dos dados de UIL de Fósforo Total (Variável Dependente) e cálculo da entropia do nó pai.

O algoritmo C4.5 trabalha com a variável dependente (classe) discretizada em k intervalos consecutivos, chamados *bins*. As variáveis independentes (atributos) podem ser utilizadas com ou sem discretização. Neste exemplo e no 1^o artigo, não foi realizada discretização das variáveis independentes.

De acordo com Zaki e Meira Jr (2014), a discretização converte atributos numéricos em categóricos, sendo normalmente aplicada para métodos de mineração de dados que não trabalham com dados numéricos. Também é útil para reduzir o número de valores de um atributo, especialmente ruídos nos dados numéricos, permitindo assim ignorar diferenças pequenas e irrelevantes nos valores.

Dado um atributo numérico X e uma amostra aleatória $\{x_i\}$ de tamanho n , com $i = 1, \dots, n$, retirada de X , a tarefa da discretização é dividir os valores de X em k intervalos consecutivos, encontrando $k - 1$ valores limitadores do intervalo: V_1, V_2, \dots, V_{k-1} , gerando intervalos do tipo:

$$[x_{min}, V_1]; (V_1, V_2]; \dots; (V_{k-1}, x_{max}]$$

Onde, os limites extremos de X são dados por $x_{min} = \min \{x_i\}$ e $x_{max} = \max \{x_i\}$.

Tabela B.1: Carga Incremental (UIL) de Fósforo total (TP) entre 2001-2013 e % das categorias de LULC em cada ECA para a faixa de 100 metros

ECA	UIL (Kg / km^2 dia)	% de LULC							
		TP ¹	CA ²	CE	AP	AU	MC	FL	MI
1	0.07	1.19	0.00	24.04	0.00	17.31	42.43	9.38	0.01
2	0.35	2.92	0.00	36.03	6.69	14.20	24.88	15.14	0.07
3	-0.05	0.00	0.00	7.65	1.03	24.62	54.31	9.08	0.00
4	0.03	0.00	0.00	13.81	23.81	12.23	36.38	2.74	0.00
5	0.39	1.26	0.00	24.30	4.62	13.29	31.86	11.56	0.00
6	0.27	0.00	19.70	15.99	29.99	4.96	18.76	0.00	0.00
7	1.17	0.21	26.15	24.81	18.61	3.54	13.29	1.74	0.15
8	-0.68	0.00	26.04	39.24	19.66	0.37	1.75	0.00	0.00
9	0.34	1.72	36.94	27.73	7.77	0.22	9.15	0.00	0.13
10	0.03	0.87	48.60	8.33	0.65	3.18	14.31	0.39	2.95
11	1.06	1.85	37.49	8.99	0.00	5.71	25.04	0.00	1.50
12	0.46	6.54	23.35	14.41	0.02	7.26	25.33	0.00	0.98
13	0.21	5.61	20.30	20.30	0.00	5.16	27.05	0.00	0.01
14	-0.12	2.99	14.58	10.76	0.00	8.15	30.89	0.00	0.00
15	0.28	0.12	13.67	13.54	0.06	4.61	30.04	0.19	0.00
16	0.01	1.04	12.86	11.52	0.00	8.45	27.16	0.15	1.77
17	-0.14	3.56	10.99	13.10	0.00	6.19	36.39	0.00	3.22

¹Variável Dependente: TP (UIL de TP em (Kg / km^2 dia)

²Variáveis Independentes: CA- Campo, CE - Cerrado, AP - Agropastoril, U_b - Urbano , MC - Mata Ciliar, FL - Floresta ; MI - Mineral; RE - Reflorestamento

Os intervalos resultantes, que abrangem todos os valores de X , são usualmente renomeados para valores simbólicos para o novo atributo categórico X . Podem ser criados intervalos de largura ou frequência constante. Os intervalos de largura constante (ω) resultam da divisão dos valores de X por k conforme equação B.1.

$$\omega = \frac{x_{max} - x_{min}}{k} \quad (B.1)$$

Assim, o i ésimo intervalo será dado pela equação B.2.

$$v_i = x_{min} + i\omega, \quad \text{Para } i = 1, \dots, k - 1 \quad (B.2)$$

Os intervalos de frequência constante dividem X em intervalos que contém (aproximadamente) o mesmo número de dados. Intervalos de igual frequência podem não ser possíveis devido a existência de valores repetidos. No primeiro artigo utilizamos intervalos de mesma largura em função da existência de um substancial número de valores repetidos na base de dados.

Como o número de dados destes exemplo é bastante reduzido optou-se por utilizar dois intervalos de igual largura para discretizar o fósforo total ($k = 2$). Assim para os dados Tabela B.1, obtivemos $\omega = 0.927$, que resultou nos intervalos $[-INF, 0.245]$; $(0.245, +INF]$ para

divisão dos dados (exemplos) de TP. Os dois intervalos foram então renomeados da seguinte forma $A = [-INF, 0.245]$ e $B = (0.245, +INF]$ e são apresentados na Tabela B.2.

Tabela B.2: Discretização da Carga Mediana Incremental (UIL) de Fósforo total (TP) e LULC faixa de 100 metros

Discretização	ECA	UIL (Kg / km^2 dia)	% de LULC							
			TP	CA	CE	AP	AU	MC	FL	MI
A	8	-0.68	0.00	26.04	39.24	19.66	0.37	1.75	0.00	0.00
A	17	-0.14	3.56	10.99	13.10	0.00	6.19	36.39	0.00	3.22
A	14	-0.12	2.99	14.58	10.76	0.00	8.15	30.89	0.00	0.00
A	3	-0.05	0.00	0.00	7.65	1.03	24.62	54.31	9.08	0.00
A	16	0.01	1.04	12.86	11.52	0.00	8.45	27.16	0.15	1.77
A	10	0.03	0.87	48.60	8.33	0.65	3.18	14.31	0.39	2.95
A	4	0.03	0.00	0.00	13.81	23.81	12.23	36.38	2.74	0.00
A	1	0.07	1.19	0.00	24.04	0.00	17.31	42.43	9.38	0.01
A	13	0.21	5.61	20.30	20.30	0.00	5.16	27.05	0.00	0.01
B	6	0.27	0.00	19.70	15.99	29.99	4.96	18.76	0.00	0.00
B	15	0.28	0.12	13.67	13.54	0.06	4.61	30.04	0.19	0.00
B	9	0.34	1.72	36.94	27.73	7.77	0.22	9.15	0.00	0.13
B	2	0.35	2.92	0.00	36.03	6.69	14.20	24.88	15.14	0.07
B	5	0.39	1.26	0.00	24.30	4.62	13.29	31.86	11.56	0.00
B	12	0.46	6.54	23.35	14.41	0.02	7.26	25.33	0.00	0.98
B	11	1.06	1.85	37.49	8.99	0.00	5.71	25.04	0.00	1.50
B	7	1.17	0.21	26.15	24.81	18.61	3.54	13.29	1.74	0.15

As duas variáveis categorias (A e B) foram geradas para o atributo TP. O intervalo “A” possui 9 exemplos e “B” 8 exemplos de TP. A entropia do nó pai foi então obtida para A=9 e B=8 de acordo com a equação B.4.

$$\text{Entropia do nó Pai TP: } H(TP) = \left(-\frac{9}{17} * \log_2[9/17]\right) + \left(-\frac{8}{17} * \log_2[8/17]\right) = 0.9975 \quad (\text{B.3})$$

2ª Etapa: Divisão dos atributos (Variáveis explicativas)

Se um atributo X assume valores reais (numéricos), é gerado um teste binário cujos resultados são $X \leq Z$ e $X > Z$. O limite de Z pode ser encontrado da seguinte forma no algoritmo C4.5:

- Os exemplos de X são inicialmente ordenados do menor para o maior valor, conforme discretização realizada conforme Tabela B.2.
- Há apenas um conjunto finito de valores, que podemos denotar (em ordem) por $\{v_1, v_2, \dots, v_L\}$.
- Qualquer limite caindo entre v_i e v_{i+1} tem o mesmo efeito que particionar o atributo X que encontra-se em $\{v_1, v_2, \dots, v_i\}$ e em $\{v_{i+1}, v_{i+2}, \dots, v_L\}$.

- Assim, existem apenas $L - 1$ divisões possíveis para o atributo X , cada uma devendo ser examinada.
- Isso pode ser obtido (uma vez ordenados os valores) em uma única passagem, atualizando as distribuições de classes para a esquerda e para a direita do limite Z durante o processo para todos os 8 atributos avaliados.
- Alguns indutores podem escolher o valor de limite como sendo o ponto médio de cada intervalo $Z = \frac{(v_i+v_{i+1})}{2}$
- O algoritmo C4.5 escolhe o maior valor de Z entre todo o conjunto de treinamento que não excede o ponto médio acima, assegurando que todos os valores que aparecem na árvore de fato ocorrem nos dados, conforme exemplo apresentado na Tabela B.3 onde Z foi calculado igual a 34.12 e adotado como 31.86.

A Tabela B.3 apresenta o atributo floresta (FL150) ordenado, o cálculo de Z e nova distribuição dos intervalos da classe de TP em função do atributo FL150 ordenado (Percentual de Floresta na faixa de 150 metros).

Tabela B.3: Divisão de TP em função FL150

FL150	TP	Z
1.75	A	
9.15	B	5.45
13.29	B	11.22
14.31	A	13.80
18.76	B	16.53
24.88	B	21.82
25.04	B	24.96
25.33	B	25.18
27.05	A	26.19
27.16	A	27.11
30.04	B	28.60
30.89	A	30.47
31.86	B	31.38
36.38	A	34.12
36.39	A	36.38
42.43	A	39.41
54.31	A	48.37

A partir de $Z \geq 34.12$, TP é classificado como A, e para $Z < 34.12$, TP possui as classes A e B. Entretanto, como o algoritmo C4.5 não utiliza o ponto médio do intervalo $Z=(v_i + v_{i+1})/2$, definimos Z como sendo o maior valor dos dados que não excede Z médio, ou seja, ele adota o maior valor de Z entre todo o conjunto de treinamento que não exceda o ponto médio acima, assegurando que todos os valores que aparecem na árvore de fato ocorrem nos dados. Assim, Z de FL150 será adotado igual a 31.86.

Fazendo $Z \leq 31.86 = FL_1$, temos 5 amostras classificadas como A e 7 amostras classificadas como B (5A,7B) e para $Z > 31.86 = FL_2$, 4 amostras são classificadas como A e zero (0) amostras como B (4A,0B) (Tabela B.3).

3ª Etapa: Procedimento de cálculo da Entropia da divisão dos dados (Entropia do nó filho a partir da divisão do nó pai), Ganho de Informação e Razão de Ganho para os atributos para Floresta no buffer de 150 metros.

A entropia de FL_1 e a FL_2 são obtidas pelas equações B.4 e B.5:

$$H(FL_1) = -\left(\frac{5}{13} * \log_2[5/13]\right) - \left(\frac{7}{13} * \log_2[7/13]\right) = 0.96 \quad (B.4)$$

$$H(FL_2) = -\left(\frac{4}{4} * \log_2[4/4]\right) - \left(\frac{0}{4} * \log_2[0/4]\right) = 0 \quad (B.5)$$

O cálculo da entropia da divisão (entropia dos nós-filhos) promovida por (FL150) com relação ao nó pai (TP) é obtida pela equação B.7. A Tabela B.4 apresenta os resultados obtidos para os oito atributos testados

$$H(FL_1, FL_2) = \frac{13}{17} * 0.9612 + \frac{4}{17} * 0 = 0.7350 \quad (B.6)$$

Tabela B.4: Entropia, Entropia Condicional, Ganho, Razão de Ganho para buffer 150 metros

Parâmetro	CA	CE	AP	U_b	MC	FL	AM	RE
Z	5.6	37.5	36	23.8	14.2	31.86	9.34	1.50
H(TP)	0.9975							
$H(FL_1, FL_2)$	0.9305	0.9305	0.9305	0.9305	0.8795	0.7350	0.8567	0.8113
Ganho	0.0669	0.0669	0.0669	0.0669	0.1179	0.2624	0.1407	0.1861
Razão de Ganho	0.0719	0.0719	0.0719	0.0719	0.1341	0.3570	0.1643	0.2294

O ganho de informação utiliza a entropia (H) (Equação B.7) como medida de impureza, comparando o grau de entropia do nó-pai ($H(TP)$ - antes da divisão) com a entropia dos nós-filhos ($H(FL_1, FL_2)$ - após a divisão).

$$(TP, FL_1, FL_2) = H(TP) - H(FL_1, FL_2) = 0.9975 - 0.7350 = 0.2625 \quad (B.7)$$

Os oito atributos de LULC foram testados para avaliar o que gera maior ganho de informação, sendo este escolhido como condição de teste.

Quanto maior o ganho de informação, maior será a redução na entropia, sendo possível considerar a divisão dos exemplos adequada.

A Razão de Ganho (Equação B.8) nada mais é que o ganho de informação relativo (ponderado) Figura B.1.

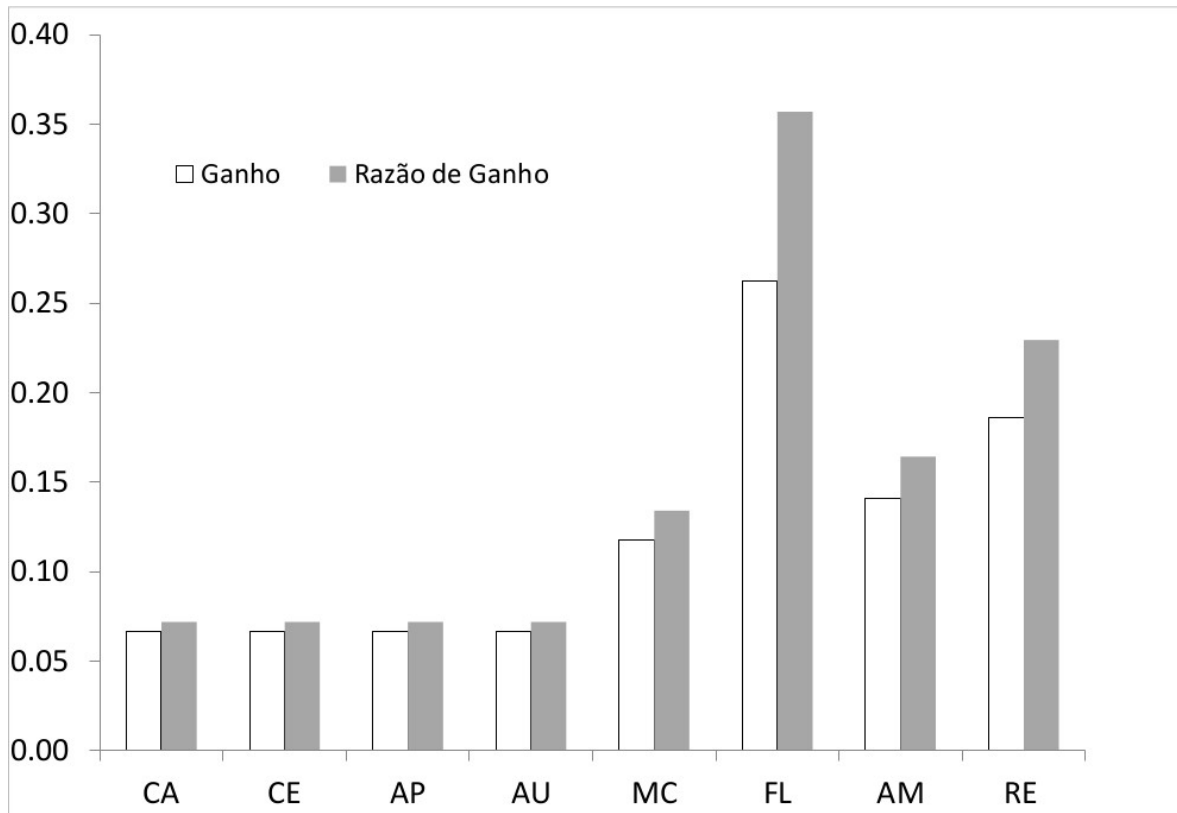


Figura B.1: Ganho de Informação e Razão de Ganho para os 8 atributos de LULC avaliados

$$\text{Razão de Ganho} : = \frac{(TP, FL_1, FL_2)}{H(FL_1, FL_2)} = 0.2625/0.7350 = 0.357 \quad (\text{B.8})$$

Na Tabela B.4 e figura B.1, podemos verificar que os atributos Floresta (FL) e Reflorestamento (RE) obtiveram os maiores ganhos de informação, 0.26 e 0.19, respectivamente.

Para o traçado da árvore de decisão, selecionamos, como critério de pré-poda, a razão de ganho superior a 0, 20 e obtivemos a árvore de decisão da Figura B.2.

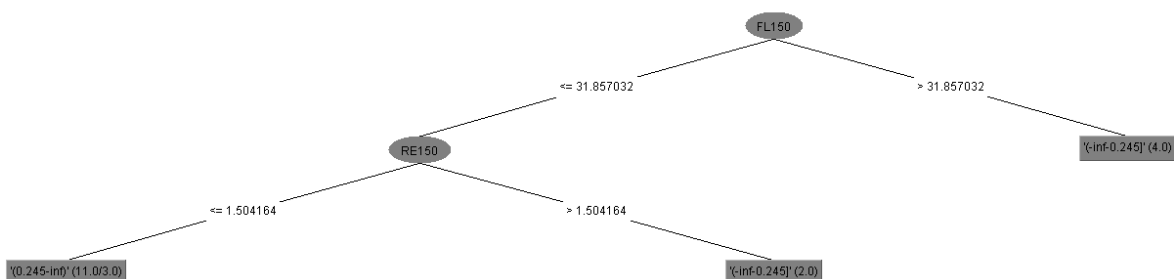


Figura B.2: Árvore de Decisão (c4.5) para Razão de Ganho Maior que 0.20

Apêndice C

Base de Dados dos Artigos e Estatística Decritiva

A base dados utilizada nos 4 artigos produzidos na tese se encontranda disponibilizada em CD anexo ao trabalho final.