Journal of Environmental Management 86 (2008) 158–170

Journal of Environmental Management

www.elsevier.com/locate/jenvman

A spatial-statistical approach for modeling the effect of non-point source pollution on different water quality parameters in the Velhas river watershed – Brazil

Philippe Maillard[a,*], Nádia Antônia Pinheiro Santos[b]

[a]Departamento de Cartografia, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Belo Horizonte, Minas Gerais, 31270-901, Brazil
[b]Instituto Mineiro de Gestão das Águas (IGAM), Rua Santa Catarina, 1354, Belo Horizonte, Minas Gerais, 30170-081, Brazil

Received 27 September 2005; received in revised form 27 November 2006; accepted 1 December 2006
Available online 20 February 2007

Abstract

In this article, a methodology for evaluating the effect of land use/land cover on the quality of nearby stream water in a semiarid environment is described and tested on a large watershed in Southeastern Brazil. The approach aims at identifying the width of the riparian area having the strongest effect on different water quality parameters. The land use/land cover data were generated from remotely sensed data while water quality point data were supplied by a government agency. Testing was conducted for both the rainy and dry seasons in an effort to understand the direct effect of surface runoff. The approach combines cartographic modelling using a geographical information system (GIS) and statistics to establish the strength of the relationship between water quality, land use and the distance from the stream. Results suggest a strong relationship between land use/land cover and turbidity, nitrogen and fecal coliforms. They also suggest that each of these parameters has a unique behavior when distance from the stream is considered. Finally, although it was expected that the models would apply better during the wet season, some parameters had the opposite behavior and displayed a better fit during the dry season.

Keywords: Water quality; Non-point pollution; Statistical modelling; Remote sensing

1. Introduction

A watershed is a natural spatial entity for land use planning and environmental management. Given a long time scale (geological), almost everything within the watershed will be deposited in the stream that drains it. Even on a scale of a few hours (a precipitation event), surface runoff can wash unconsolidated sediments downstream and, with it, nutrients and chemicals that will directly affect the quality of the receiving waters. As such, surface runoff is a major source of non-point pollution and is primarily responsible for the relationship between land use/land cover (LULC) and water quality (WQ).

In a natural densely vegetated watershed, most rainfall is absorbed by the land surface and vegetation and released over a long period of time and there is relatively little direct surface runoff with few nutrients being "lost" to the stream (Karr and Schlosser, 1978; McCulloch and Robinson, 1993). The removal of the natural vegetation and its replacement by agricultural or pastoral land promotes surface runoff and sediment transport (Bruijnzeel, 1990; McCulloch and Robinson, 1993). The conversion of natural vegetation to agriculture and pasture is intense in Brazil. This is especially true for the *cerrado* (wooded savana), the second largest biome in Brazil having a rate of "conversion" exceeding that of the Amazon forest (Ratter et al., 1997).

The impact of this conversion to agriculture, pasture and other uses such as eucalyptus plantations, open mining and urbanization can be considered severe in the Velhas river watershed (Euclydes and Ferreira, 2002). It had drastic

*Corresponding author. Tel.: +55 31 3499 5427; fax: +55 31 3499 5410.
E-mail addresses: philippe@cart.igc.ufmg.br (P. Maillard), psnadia@yahoo.com.br (N.A. Pinheiro Santos).

0301-4797/$ - see front matter
doi:10.1016/j.jenvman.2006.12.009

consequences on the quality of water and has caused the disappearance of most fish species downstream of Belo Horizonte (the capital of the state of Minas Gerais) (Polignano et al., 2001). Many point pollution sources are the cause of the problem and local authorities are taking measures such as the implementation of sewage water treatment plants and stricter control of polluting industries but comparatively little is done with regard to non-point source pollution. One initiative has taken the form of a pilot project for the restoration of riparian vegetation of tributaries of the São Francisco river by EMATER (2005), and the Velhas river is one such tributary. The exact effectiveness of riparian vegetation to reduce diffuse pollution and sediment transport still needs to be further studied. The combination of the level of degradation of the Velhas river and the initiative to restore riparian vegetation has triggered a lot of interest and motivated the present study.

2. Research objective

The objective of this study is to develop and test a simplified methodology to determine the effect of LULC on WQ on a large scale watershed in a semiarid environment. The work is based on the analysis of the Velhas river watershed in Minas Gerais—Brazil. Since it is known that the effect of LULC is greater closer to the stream in the so-called riparian zone, the approach incorporates this spatial relation through analytical cartography (in a GIS environment) by creating buffer zones corresponding to riparian zones (RZ) for each sampling point of WQ. It is hoped that this approach will help determine the effect of distance on different WQ parameters. The methodology is also bi-seasonal: wet and dry. The dry season data are used as a means of control since it is expected that the relation between LULC and WQ should decrease when no precipitation has been recorded for some time and there is no surface runoff.

The approach adopted is a direct consequence of the difficulty and cost of acquiring data over such an extensive area. As in most of Brazil, topographical maps of the Velhas river watershed are only available on a scale of 1:100 000 and have not been updated for over twenty-five years. On such a coarse scale, slope analysis would be strongly biased, especially if only the riparian zones are considered. It is hoped that developing a simpler approach that still provides reliable results would promote more studies on the effect of non-point pollution sources on the WQ of Brazilian streams. The approach is also based on the assumption that each WQ sampling point can be treated as independent from the others when processed appropriately. This is also an effort to disregard point (e.g. sewers, industries) and linear (e.g. roads) pollution sources based on the fact that rural areas are much more extensive than urban ones. Therefore, only a few water sampling points are directly affected by point pollution sources and do not strongly bias the majority of the sampling points located in rural areas.

3. Background

3.1. Water quality and land use

Land use plays a complex multi-faceted role in the hydrological cycle (Karr and Schlosser, 1978; Olsson and Pilesjo, 2002). The vegetation intercepts, and re-evaporates (evapotranspiration) part of the precipitation and has an effect on other hydrological parameters such as percolation and surface runoff (Tong and Chen, 2002). Cleared land and some agricultural practices can promote overland flow and erosion (Nisbet, 2001). The construction of impervious surfaces also causes erosion and prevents replenishing of the water table (Cornish, 2001). All of these effects eventually influence the water quality of the nearby streams and rivers. It has been well reported that the use of agricultural fertilizers can result in an increase of levels of nitrate and phosphorus (Karr and Schlosser, 1978; Foster et al., 1989; Meybeck et al., 1989; Mattikalli and Richards, 1996; Basnyat et al., 1999). Similarly, stock grazing and dairies may increase the presence of fecal bacteria in the water (Moore et al., 1989), provoke erosion problems and increase the turbidity of stream waters (Stout et al., 2000; Evans, 2005).

3.2. GIS and remote sensing in water quality studies

The use of GIS technology for integrating LULC data, WQ data and even hydrological models is increasingly popular in the scientific community and numerous articles on WQ issues can be quoted to that effect. In many of these projects, remote sensing is used for mapping and monitoring LULC. Mattikalli and Richards (1996) combined land use data from maps and remotely sensed images with an "export coefficient model" and a GIS to evaluate nitrogen loss and compared them with measured values in the River Glen watershed (UK). Although their estimates generally agree with real measured values, significant differences were reported. Fisher et al. (2000) used a GIS as a means of spatially grouping WQ stations and correlated their results with agricultural activities (poultry, dairy and beef production) and conservation practices. In their discussion they showed that levels of nitrogen, phosphorus, turbidity and fecal coliform bacteria could be directly associated with the localization of certain agricultural activities.

Basnyat et al. (1999) went further and proposed the integration of GIS, ecological modelling and remote sensing as an effective decision support tool for land management. They computed two multiple regression models based on the proportion of different LULC classes to explain nitrate (NO_3) levels in a medium-size watershed (138 km^2) which was divided into eight different basins where WQ data were collected over a two-year period. While the frst model had a relatively low determination coefficient (r^2) of 0.186 because it integrated whole basins, the second model only considered a model-based riparian zone and showed a strong r^2 of 0.959. Many aspects of the current study were inspired by Basnyat et al. and, for consistency reasons, some terms (e.g. LULC, contribution zone) were also borrowed from their study.

In another study (Wang, 2001), twenty-two river catchments near headwaters were used as independent processes to which were associated the proportions of land use and their relative position to water treatment plants. He used Pearson's correlation and multiple regression to reveal the relationship between biological indicators (fish and macro-invertebrates) and land use. Multiple regression results were conclusive for the fish bio-indicator with urban and wooded areas as strong predictors. Gardi (2001) used a GIS and a crop simulation model (CropSyst) to characterize the relationship between cropping practices, land use, soils and nitrate loss as well as concentrations of herbicides in the water. Although he was not able to establish the relationship between nitrate loss and cropping practices, he demonstrated that the practice of row crops had a negative effect on concentrations of herbicides and pesticides in the Centonara watershed (Italy). In an integrated study Tong and Chen (2002) found a significant correlation between land use and a series of WQ variables on a regional scale (the State of Ohio). They then used these correlations to calibrate a hydrological model (BASINS) which was used in combination with a GIS to simulate runoff and levels of various WQ parameters on a local watershed scale.

3.3. Riparian vegetation

All the studies quoted above (except the study by Basnyat et al. (1999)) used the whole watershed as an input. Although using the whole of the watershed processes to explain the hydrological and ecological condition of the stream is valid, the riparian zone has a disproportionate influence (Petersen, 1992; Schuft et al., 1999; Paula Lima (de) and Brito Zakia, 2001). In particular, riparian forests have the following effect on streams, they:

- stabilize river banks through their intertwining root system,
- serve as a buffer zone and filter between the higher watershed and the aquatic ecosystem, absorbing nutrients and other chemicals from surface runoff,
- decrease and filter the sediment load of surface runoff,
- help maintain the thermal stability of the water by intercepting part of the solar radiation and,
- supply organic matter for aquatic organisms.

As an example of the importance of riparian forests, Sparovek et al. (2002) showed that from an economic point of view, to maintain soil loss at a "sustainable" level (blocking 80% of the current sediment yield value), in a Southeastern Brazilian watershed, a minimum width of 52 meters of forest should be maintained on each side of the river (current legislation requires only 30 m). In an attempt to take the importance of riparian zones into account, Schuft et al. (1999) have proposed a methodology for extracting a series of landscape metrics to characterize riparian stream networks to accommodate the dichotomy of WQ point data and spatial LULC information. Their approach is based on defining sampling areas around each observation site. First, concentric circles of equal area (1 km^2) are drawn, then linear buffer zones are defined on each side of the stream section within the rings. The sampling areas are defined by the intersection of the buffer zones and the concentric circles. A similar, yet different sampling approach has been adopted in the present article and is described in Section 4.3. The idea of isolating the land use in riparian zones is an effort to help determine the width of riparian forest that should be preserved (in a legislative sense) to reduce non point pollution sources in agricultural areas.

4. Material and method

4.1. The Velhas river watershed

With 27,867 square kilometers, the Velhas river watershed is one of the most important in the state of Minas Gerais (Fig. 1). Situated in the central part of the state it also hosts the capital Belo Horizonte and the greater metropolitan area with over three million people. Having suffered repeated degradation for the past 400 years, it is the subject of various restoration projects. Point and non-point pollution sources are unevenly distributed in the watershed. The upper watershed concentrates very important urban areas and equally important mining operations, but agriculture and grazing are also found. The land use in the middle and lower Velhas river is mainly dedicated to agriculture and grazing. Although eucalyptus plantations are found everywhere, the majority of the activities are concentrated in the middle section. While the headwater region is characterized by a rugged topography (average altitude: 1500 m) and receives about 2000 mm of rainfall each year, the river mouth is mostly flat (at an altitude of 500 m) and receives only about 1100 mm of rain.

4.2. Data and data generation

In this study, data were obtained from four different sources: (1) Landsat images, (2) digital topographic maps, (3) in situ field surveys and (4) water quality data. The first three sources were used to produce LULC digital maps. The LULC map was generated from a mosaic of five Landsat 7 ETM+ images. Digital maps and in situ field data were used both as ground truth and to complement the classified mosaic with elements that could not be extracted from the images. The five Landsat images were geometrically rectified and registered to a UTM cartographic projection. Table 1 shows the orbit/scene reference of each image along with its date of acquisition and root

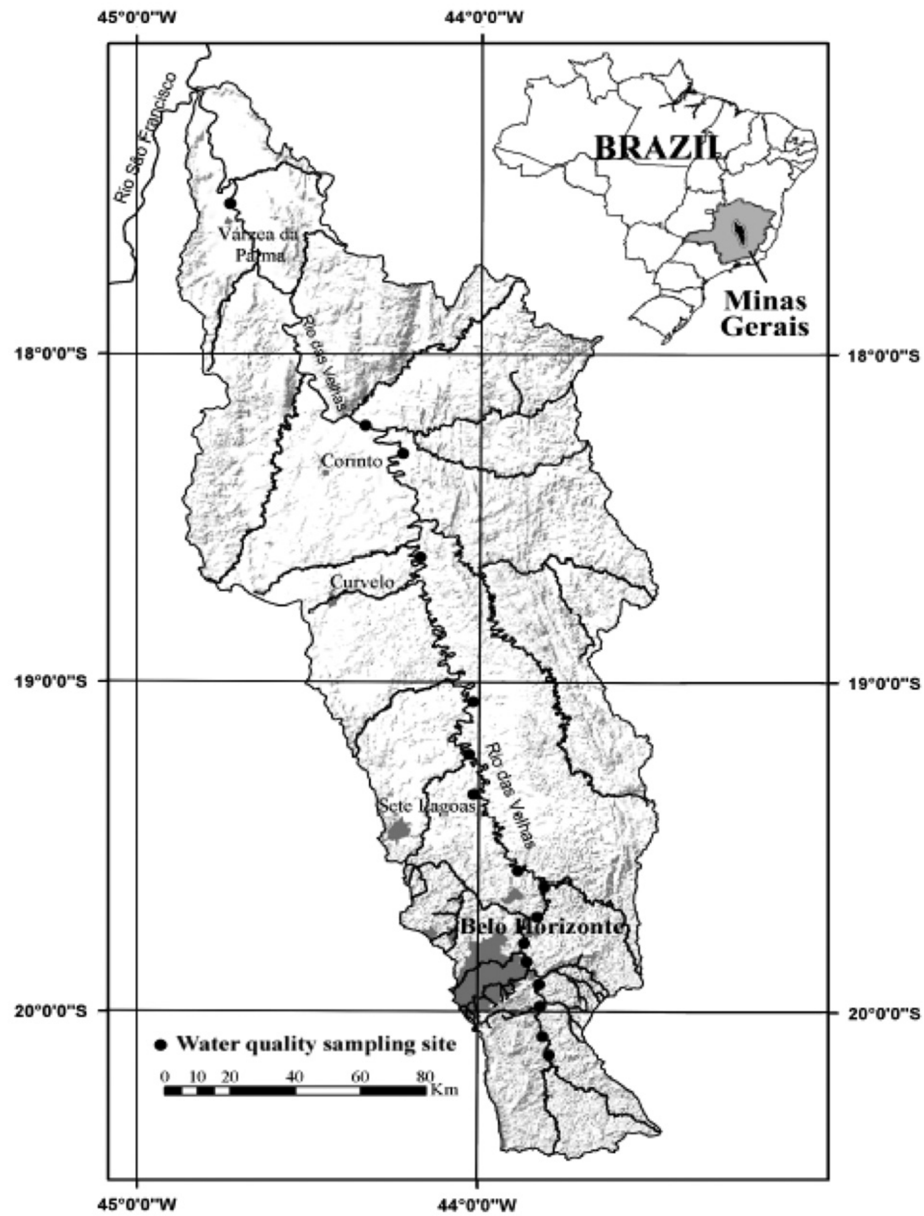

Fig. 1. The Velhas river watershed.

mean square (RMS) registration error. The scenes were first radiometrically corrected using the "improved dark-object subtraction technique"[1] to correct for atmospheric differences (Chavez Jr., 1988). The results from this correction were not completely satisfactory and further *ad hoc* radiometric adjustments were necessary to reduce spectral response differences between the images. These additional corrections were small linear histogram adjustments. The mosaic was constructed by juxtaposing the five images while always leaving the image with the least atmospheric effect on top for overlap. The resulting mosaic

Table 1
Listing of the orbit/scene reference (WRS) for the five Landsat 7 ETM + scenes with their respective date of acquisition, RMS registration error and standard deviation (σ)

WRS orbit/scene	Date of acquisition	RMS error (m)	σ (m)
218/72	22 October 2002	24.30	13.21
218/73	06 October 2002	49.59	24.96
218/74	06 October 2002	44.27	24.57
219/72	13 October 2002	66.29	26.15
219/73	13 October 2002	16.56	12.00

[1]The dark object subtraction technique consists of subtracting a constant value from all the pixels in a band of an image. The constant is determined by identifying an object that should either absorb all electromagnetic energy (*e.g.* a clear water lake) or not receive any direct light (*e.g.* a very steep slope). If such an object has a value of 20, then 20 is subtracted from all the pixels in that band. The process is repeated for all the other bands. In the improved version, the constants for all the bands are not independent but computed from a look-up table and an estimate of the quality of the atmosphere.

was sufficiently uniform to classify as a single data source using training data acquired in situ and, in some exceptional cases, from maps. The thermal and panchromatic bands were excluded from the classification process. Urban areas were first extracted from topographic maps and then visually updated using the Landsat mosaic. The mosaic was over 99% cloud free. Clouds were treated as "no data" and removed from the analysis. The watershed limits were used as a mask to clip onto the classified mosaic.

Classification was performed using maximum likelihood from the Purdue/NASA MultiSpec software package that is available at no cost from the MultiSpec website (http://dynamo.ecn.purdue.edu/biehl/MultiSpec/). The MultiSpec classification scheme makes it possible to classify any combination of the training areas, testing areas and/or the entire image and reports the results with a wide range of statistics. Biehl and Landgrebe (2002) give a complete description of the MultiSpec package. The LULC map accuracy was tested using only in situ data which were collected using a near-random sampling scheme. A total of 250 random sampling points were defined within the watershed to guide field work. Using a global positioning system (GPS) the sampling sites were then visited (by the authors) on the basis of the "nearest accessible site" within a one kilometer radius of the randomly generated coordinates. Each site generated a bundle of pixels (from 9 to 25 depending on the homogeneity of the terrain) that were used either as training (about 20%) or as test areas. Although sub-optimal, this approach had the benefit of expediency and avoiding land owners' authorization problems. The hydrographic network was extracted from the topographic maps and manually adjusted where needed (and possible). The LULC classification system used was inspired by the Anderson system (Anderson et al., 1976) and adapted for Brazil according to Sokolonski's "Sistema de Classificação de Uso Atual da Terra" (Sokolonski, 1999). The classified mosaic was merged with the adjusted hydrographic network, the updated urban areas and the "no data" zones to produce a single LULC map. Overall accuracy was of 89.8% (kappa $\kappa = 0.883$). Table 2 shows the two-level class system used in this study and gives the producer's and user's accuracies for each LULC class along with sample size. For statistical modelling purposes the classification system was simplified to the following classes: riparian forest, forest, savanna, planted forest, agro-pastoral and barren (the class *urban* was extracted from maps and/or interpreted from the imagery and was not tested for accuracy).

The WQ data were obtained from the *Projeto Águas de Minas* of the *Instituto Mineiro de Gestão das Águas* (IGAM) with 29 water collecting stations for which 13 physical, chemical and biological parameters have been measured on a quarterly basis since 1992. Based on the known relationship between the parameters and LULC, five of these were retained for the statistical analysis: turbidity, nitrate, nitrite, phosphorus, and fecal coliforms. The DELPHI water quality index (Bollmann and Marques, 2000) was also used even though it incorporates some parameters that are more related to point pollution than non-point. The index was computed by IGAM using the following formula:

$$WQI = \prod_{i=1}^{9} q_i^{w_i},$$

where q is the parameter and w is the parameter weight as defined in Table 3.

Of the 29 stations, only 16 were directly from the Velhas river (and not from a tributary). Only data from 2002 were used as this was the year corresponding to the

Table 2
The LULC classification system adapted from Anderson et al.(1976) and Sokolanski (1999) along with the sample size and producer's and user's accuracy values from the image classification

Level I	Level II	Samples	Accuracy (%)	
			Producer's	User's
1 Urban		+		
2 Agro-pastoral	21 Agriculture	*		
	22 Pastoral	382	81.7	76.5
3 Vegetation	31 Primary forest	–		
	32 Secondary forest	667	97.5	95.6
	33 Wooded savanna	269	77.0	73.9
	34 Grassland savanna	67	61.2	100
	35 Dry deciduous forest	272	74.6	90.2
	36 Riparian forest	237	76.8	71.1
Planted forest	Eucalyptus plantation	441	74.4	81.2
Barren	Exposed soil	197	95.4	82.5
	Open soil	+		
	Rock outcrop	857	99.4	97.6
Water		124	99.2	100

Note that the samples were based on bundles of 9 (3×3) to 25 (5×5) pixels
+, interpreted; *, interpreted and/or merged with pastoral; –, none observed.

LULC data. As a reference, Table 4 shows total precipitation for the month of January and July for the year 2002. It should be noted however that we have no knowledge of the exact day in the month when the water samples were collected. Table 5 shows the five WQ parameters and the WQ index for each of the 16 Velhas river stations and for both the wet and dry seasons.

4.3. Determining contributing areas

Since all the WQ data are from the same watershed and most are located on the same stream (the Velhas river), they are statistically highly correlated and, to some extent, all the upstream points contribute to the measurements of any observation point. This is statistically undesirable and would produce strongly biased results. To solve this problem, the water quality data from the sampling point upstream were subtracted from each point. For the approach to be consistent, the first sampling point near the watershed source had to be eliminated (no sampling point upstream) and only the sampling points directly on the Velhas river could be used. Of course, this resulted in producing some negative values for the water quality parameters. For each remaining point (15 in all), a series of *riparian zones* (RZ) of various widths were defined. To do so, a sub-watershed was defined that contributes exclusively to each water sampling point. Then each sub-watershed is successively overlaid on riparian buffer zones of varying width to produce the *exclusive contribution zones* (ECZs). Buffers were created for a maximum width of 510 m and all were multiples of 30 m to match the LANDSAT image resolution. This approach produced 17 buffer zones which were then reduced to five using a process of elimination. The first two buffers were eliminated (30 and 60 m) for being too close to the stream and possibly affected by positioning errors. Ten of the remaining 15 buffers were eliminated for being uninformative (linearly distributed) based on analyzing the plots of coefficients of determination (see Section 5). Fig. 2 illustrates the process of creating the RZs. Since the topography of riparian zones does not vary as much as in the whole watershed, this procedure also had the effect of reducing the impact of relief variations; an important hydrologic variable for watershed modelling that is not considered in the present study. Although incorporating the relief would probably improve the models, the only available medium scale topographic maps (1:100.000 and a few 1:50.000) could not yield enough precision within the restricted riparian zones.

4.4. Statistical modelling

With only 15 sampling points, a thorough statistical analysis is difficult. Still, considering the high cost of collecting water samples systematically and performing

Table 3
Parameters and parameter weights used in the calculation of the DELPH1 water quality index

Parameter	Parameter weight
Dissolved oxygen (%)	0.17
Fecal coliform (NMP/100 ml)	0.15
PH	0.12
Biochemical oxygen demand (mg/L)	0.10
Nitrate (mg/LNO_3)	0.10
Phosphorus (mg/LPO_4)	0.10
Temperature variation (°C)	0.10
Turbidity (UNT)	0.08
Total solid (mg/L)	0.08

Table 4
Total precipitation (mm) for the months of January and July of the year 2002 for the Velhas river watershed

Month	High	Mid	Low
January	306.8	398.4	226.9
July	8.0	6.3	0.0

The data are presented for the high-, mid- and low-watershed.

Table 5
Water quality parameters measured in the 16 Velhas river stations for the wet (top) and dry (bottom) season of 2002

WQ	Velhas river station: WET season: January 2002															
Parameter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Turbidity (NTU)	94.8	603.0	531.0	505.0	436.0	415.0	1088.0	915.0	738.0	627.0	644.0	923.0	726.0	690.0	366.0	434.0
Nitrate (mg/L NO_3)	0.09	0.04	0.04	0.22	0.24	0.25	0.28	0.46	0.66	0.34	0.57	0.67	0.43	0.52	0.48	0.38
Nitrate (mg/L NO_2)	0.005	0.004	0.005	0.011	0.007	0.010	0.105	0.121	0.000	0.051	0.014	0.000	0.008	0.007	0.005	0.006
P (mg/L PO_4)	0.02	0.22	0.01	0.01	0.03	0.14	0.05	0.04	0.31	0.01	0.05	0.03	0.08	0.05	0.09	0.21
Fecal col. (NMP/100 L)	—	—	—	50.0	50.0	90.0	160.0	160.0	50.0	7.0	30.0	17.0	2.2	2.8	0.7	1.7
WQI	77.3	57.4	61.7	41.3	41.9	38.1	30.1	28.9	30.0	38.6	37.5	39.8	46.5	44.3	49.9	44.7
DRY season: July 2002																
Turbidity (NTU)	6.5	8.17	9.29	12.2	9.83	54.9	71.0	55.3	82.4	35.3	6.88	20.7	20.5	11.6	7.72	3.83
Nitrate (mg/L NO_3)	0.16	0.24	0.29	0.38	0.41	0.27	0.02	0.02	0.03	0.18	1.76	2.90	1.83	1.82	1.87	1.02
Nitrite (mg/L NO_2)	0.003	0.022	0.023	0.048	0.046	0.027	0.006	0.006	0.005	0.022	0.223	0.245	0.015	0.019	0.032	0.008
P (mg/L PO_4)	0.02	0.04	0.05	0.08	0.06	1.50	0.95	1.05	0.96	0.34	0.19	0.14	0.12	0.11	0.04	0.04
Fecal col. (NMP/100 L)	0.00	7.0	3.0	7.0	8.0	160.0	160.0	160.0	160.0	13.0	0.13	0.001	0.05	0.03	0.01	0.02
WQI	76.1	62.6	65.2	60.3	61.3	19.6	15.7	14.6	16.3	36.2	59.7	66.0	59.5	55.9	79.5	77.4

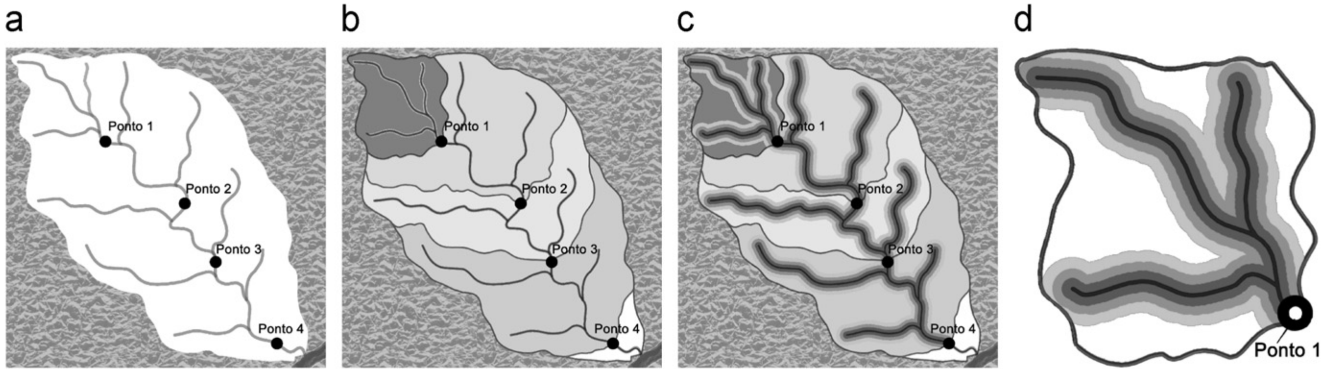

Fig. 2. Illustration of how the riparian zones are defined; (a) the water quality sampling points, (b) definition of the exclusive contribution zone for each sub-watershed, (c) overlaying the riparian buffer zones of different width, and (d) illustration of the riparian zones for point 1.

chemical analyses, the practice of statistical modelling with few samples has been considered acceptable and many other studies have adopted a similar approach. For example, Mattikalli and Richards (1996) and Fisher et al. (2000) used 17 and 10 water sampling sites respectively while Basnyat et al. (1999) and Wang (2001) used 8 and 22 small catchment watersheds respectively for their analysis. To compensate for this weakness and strengthen the results, the analysis was carried out for two different months: January (wet season) and July (dry season). The argument to support this approach was the following: during the wet season, surface runoff is more often active and LULC should have a stronger influence upon WQ than during the dry season. There are however some problems with this statement which will be considered in due time. One such problem is that farmers rarely use fertilizers and nutrients during the wet season because much of them will be washed away by the rain. Another problem is that the soil is more exposed during the dry season and therefore more susceptible to erosion.

Multiple regression using the least square rule has been chosen for the present study. Multiple regression produces linear models in the form $Y = b_0 + b_1X_1 + b_2X_2 + \ldots b_nX_n$ where Y is the explained variable, $X_{1,2\ldots n}$ are the independent or predictor variables, $b_{1,2\ldots n}$ are the coefficients and n is the number of variables. It was preferred to correlation because LULC is made up of a mosaic of different classes in all the exclusive contribution zones of the WQ sampling sites. Taken one at a time, the LULC classes would not have a significant impact on the WQ. This is also consistent with the fact that together these classes make up 100% of the contribution zones (except for a few clouds classified as "no data"). All the models were calculated using proportions of each LULC and not actual areas in units of measure. Only models with coefficients of determination (r^2) superior to 0.6 were retained. To further test the validity of the models, a level of significance $p \geqslant 0.9$ has been considered. Finally, an analysis of variance was performed on the betas ($b_{1,2,\ldots,n}$) to test if they are equal to zero ($H0$: $b_{1,2\ldots,n} = 0$; $H1$: $b_{1,2,\ldots n} \neq 0$). A level of 90% (0.1) was also used for this inference test. These precautions also made sure that the results were valid despite the small sample used. Scatterplots of the residuals were also produced to make sure these were randomly distributed.

5. Results and discussion

Tables 6 and 7 show the coefficients of determination of the various models. The tables were formatted to include information on the models that were rejected on the basis of their low coefficients of determination (italic) or on not being able to reject the null hypothesis ($H0$) of the analysis of variance stating that the betas are equal to zero at 90% (values with asterisk). This was the case for the following WQ parameters: phosphorus (P) and water quality index (WQI). Valid models are in bold type. These include nitrite for the wet season and turbidity and fecal coliforms for the dry season.

Models with a plus sign (+) mean that the $H0$ of the analysis of variance could be rejected at 85%. Nitrite and nitrate during the dry season are such models and will be discussed. Although WQI has some valid models ($H0$ rejected at 85%), it was not retained for discussion, because it is an index and not a measured value, and its interpretation would, therefore, be too subjective. Fig. 3 shows a series of plots of the coefficients of determination and how they vary with the buffer width (RZ) for the WQ parameters retained for discussion. The results for these parameters are discussed separately in the following subsections.

5.1. *Turbidity*

High turbidity is normally associated with the wet season when surface runoff transports sediments from the soil and carries them to the stream. During the wet season the high waters are also much more turbulent and do not allow the sediments to settle on the river bed. Suspended sediments can therefore travel long distances and all the tributaries also contribute to the turbidity. This is also the case here and turbidity is much higher during the wet season (see Table 5).

For modelling purposes, turbidity values were log transformed to obey a more linear behavior. The January

Table 6
Coefficients of determination (r^2) of the wet season multiple regression models for the five water quality parameters and for the different riparian zones (RZ) including the whole exclusive contribution zone (ECZ, last column)

Water quality	WET season: January 2002					
Parameters	RZs					ECZ
	90 m	150 m	210 m	300 m	510 m	
Turbidity	*0.194*	*0.198*	*0.194*	*0.198*	*0.214*	*0.222*
Nitrate	*0.513*	*0.504*	*0.482*	*0.457*	*0.431*	*0.507*
Nitrite	**0.823**	**0.823**	**0.820**	**0.815**	**0.803**	**0.822**
P	**0.690 +**	**0.647***	**0.647***	*0.580*	*0.543*	*0.543*
Fecal col. (× 1000)	**0.755***	**0.735***	**0.725***	**0.709***	**0.691***	**0.789***
WQI	*0.453*	*0.461*	*0.459*	*0.464*	*0.487*	*0.502*

Bold: $r^2 \geqslant 0.6$ betas$\neq 0$ (90%); +: $r^2 \geqslant 0.6$ betas$\neq 0$ (85%)
*: $r^2 \geqslant 0.6$ betas = 0; italic: $r^2 < 0.6$.

Table 7
Coefficients of determintation (r^2) of the dry season multiple regression models for the five water quality parameters and for the different riparian zones (RZ) including the whole exclusive contribution zone (ECA, last column)

Water quality	DRY season: January 2002					
Parameters	RZs					ECZ
	90 m	150 m	210 m	300 m	510 m	
Turbidity	**0.874**	**0.906**	**0.918**	**0.925**	**0.895**	**0.831**
Nitrate	**0.726 +**	**0.752**	**0.779**	**0.809**	**0.836**	**0.708**
Nitrite	**0.713 +**	**0.729 +**	**0.744**	**0.700**	**0.772**	*0.621*
P	*0.507*	*0.515*	*0.515*	*0.493*	*0.441*	*0.343*
Focal col. (× 1000)	**0.850**	**0.849**	**0.834**	**0.801**	**0.726**	**0.621***
WQI	**0.677***	**0.673***	**0.664***	**0.649***	**0.616***	**0.474**

Bold: $r^2 \geqslant 0.6$ betas$\neq 0$ (90%); $r^2 \geqslant 0.6$ betas$\neq 0$ (85%).
*: $r^2 \geqslant 0.6$ betas = 0; italic: $r^2 < 0.6$.

(wet) model was not conclusive with r^2 values below 0.3 for the RZ buffers and below 0.4 for the ECZ. This is understandable for the wet season since suspended sediment intake before the ECZ are generally high and may come from a long way upstream. The contribution of the immediate neighborhood of the water sampling points is therefore relatively too small to affect the model. This is somewhat confirmed by the fact that the r^2 is larger for the ECZ than the RZs.

During the dry season, the regression model is quite conclusive with high coefficients of determination. The analysis of variance on the betas also made it possible to reject $H0$: betas = 0. In that period, relatively little sediment is taken in, and, when this occurs, it is not transported very far. This is also the season during which farmers irrigate their fields and when soils are most vulnerable to erosion. The RZ model shows the strongest relationship (0.925) at a buffer distance of 300 m declining to 0.895 at 510 m and 0.831 for the ECZ (Fig. 3a). The actual model (Table 8) shows that the class that contributes most to increased turbidity is barren land followed by agro-pastoral and planted forest. Here the class *forest* is abnormally counted as a contributor but with a much lower weight than barren soil. The strongest inhibitor is predictably riparian forest. Savanna is also an inhibitor but with a rather low weighting factor. Although imperfect, the model still maintains a predictable behavior for most classes (except forest) and appears to be a valid predictor Tables 9,10,11.

5.2. Nitrate and nitrite

In the aquatic environment, nitrogen can be found in various forms: molecular nitrogen, organic nitrogen, ammonium nitrogen, nitrite and nitrate. Domestic and industrial sewers along with animal excrement and fertilizers are the main sources of nitrogen. Nitrite is usually associated with active biological processes influenced by organic pollution (IGAM, 2002). Nitrate is more associated with the use of organic and inorganic fertilizers (Meybeck et al., 1989; Mattikalli and Richards, 1996; Basnyat et al., 1999). Although these two nitrogen sources are often treated together, since they were measured separately, two different sets of models were created. Except for four sampling sites in the middle of the watershed, nitrite and nitrate are higher in the dry season

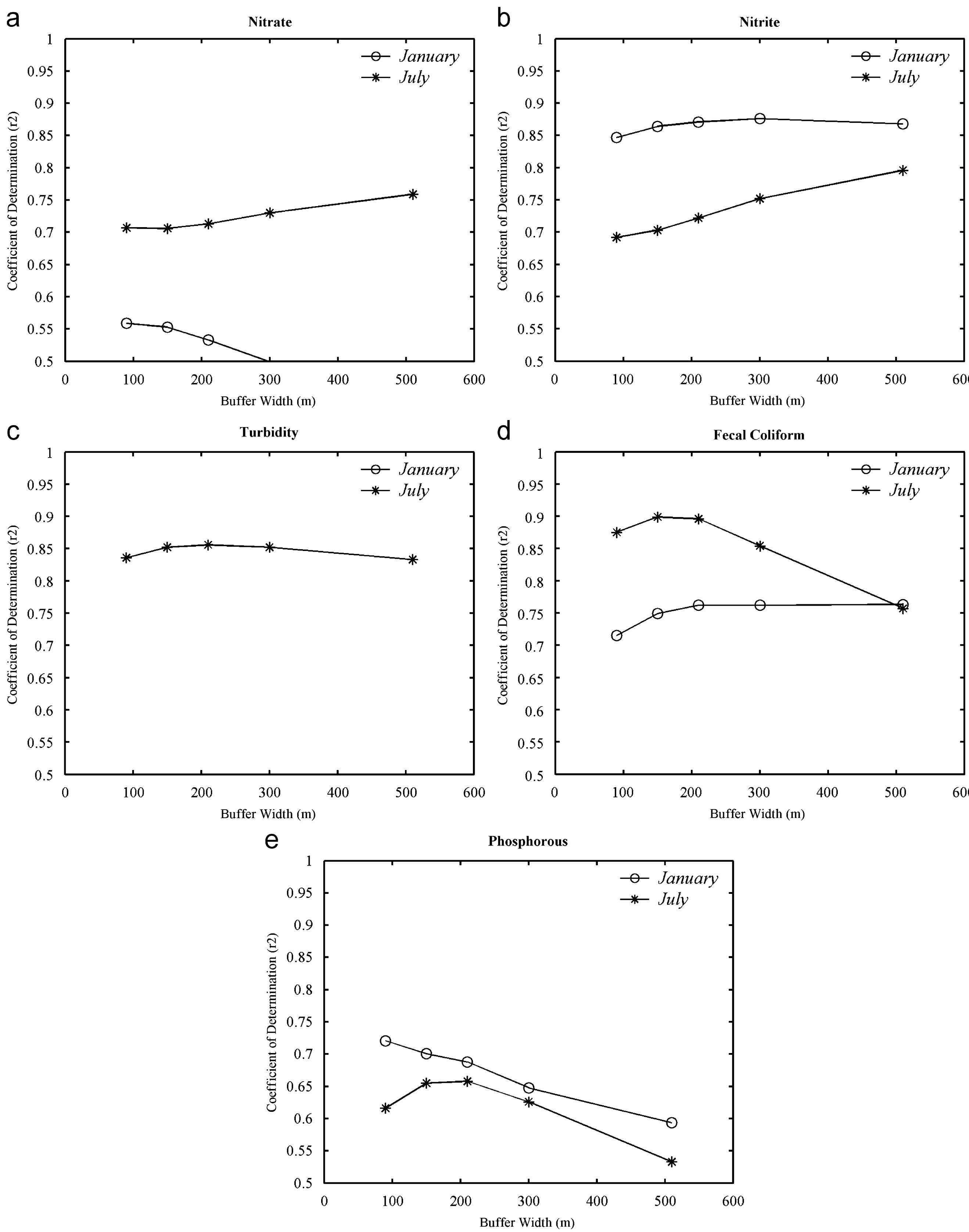

Fig. 3. Plosts of r^2 according to riparian buffer zone width.

Table 8
Multiple regression model for turbidity (dry season/300 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	−5.985	−3.98E-03	0.180	−4.75E-04	7.740E-02	8.868E-02	8.433E-02	5.728E-02
Std. err.	2.220	0.029	0.038	0.024	0.029	0.037	0.024	0.022

C = constant; RF = Riparian Forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted Forest; U = Urban.

Table 9
Multiple regression model for nitrate (dry season/510 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	−3.294	−2.53E-02	2.486E-02	1.032E-02	5.234E-02	8.074E-02	6.804E-02	2.976E-02
Std. err.	1.918	0.022	0.030	0.018	0.022	0.027	0.017	0.016

C = constant; RF = Riparian forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted forest; U = Urban.

Table 10
Multiple regression model for nitrate (wet season/90 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	1.186E-02	−1.18E-04	1.756E-04	−6.32E-04	−3.42E-04	4.045E-04	−7.82E-04	9.974E-04
Std. err.	0.090	0.001	0.001	0.001	0.001	0.001	0.001	0.001

C = constant; RF = Riparian forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted forest; U = Urban.

Table 11
Multiple regression model for nitrite (dry season/510 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	−0.313	−4.28E-04	−3.49E-03	4.368E-04	6.766E-03	8.894E-03	7.811E-03	3.259E-03
Std. err.	0.248	0.003	0.005	0.003	0.003	0.004	0.003	0.002

C = constant; RF = Riparian forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted forest; U = Urban.

(Table 5). The reason for these generally higher concentrations[2] can partly be explained by the fact that high waters of the wet season dilute these chemicals more. Another by no means negligible reason for this seasonal behavior is that farmers do not use much fertilizer during the rainy season as most of it would be washed away.

The models for nitrite show a consistent behavior for both seasons but with a stronger relationship in January (wet) that culminates at 300 m. The dry season models are weaker and less significant in terms of analysis of variance results. The curve shows a steady increase from 90 to 510 m suggesting that it might have continued to increase if a larger buffer had been considered. These results suggest that runoff during the rainy season might be responsible for bringing an increased amount of organic matter from the riparian zone into the stream.

In all three models riparian forest acts as an inhibitor confirming its essential filtering role. For the nitrate, all the other classes have a positive contribution. The fact that forest and savanna have positive weights is not an expected result and probably comes from the various sources of error and the simplified model approach. The wet season nitrite model has, in decreasing order of importance, planted forest, savanna, forest and riparian forest as inhibitors. Here the distance from the stream appears to be less important since the determination coefficients do not decrease much with the increased distance. This is perhaps why the riparian forest is the weakest inhibitor. Urban is the strongest contributor followed by agro-pastoral and barren soil. During the dry season, the model behaves in a less understandable way but the strongest contributors (agro-pastoral, planted forest and urban) are consistent with the expected behavior. Since this model has a weaker coefficient of determination (maximum of 0.772 at 510 m), it is therefore less reliable.

5.3. Phosphorus

Consumption of phosphorus does not appear to affect human health but high phosphorus levels in the water can

[2]The concentrations in Table 5 are considered acceptable for human consumption ($< 5\,\mathrm{mg} = L$) but are significantly higher than would be found in a natural environment.

have dramatic effects on aquatic life. Concentrations of over 0.03 mg/L can provoque excessive plant growth, increase demand of oxygen and reduce fish stock (MOE, 1984). Most concentrations presented in Table 5 are over 0.03 mg/L and can be considered excessive. Phosphorus in surface water can come from a variety of sources both point and non-point. It is associated with sewers, animal excrement, fertilizers and some detergents.

Unlike nitrogen, the difference in levels between the wet and dry seasons are small. However, the model for the dry season could not be considered reliable with coefficients of determination well below the 0.6 mark and a level of significance below 90%. The wet season model on the other hand, can be considered acceptable for the first three buffer widths (90, 150 and 210 m). The model for the 90 m buffer is in Table 12 and Fig. 3e and shows that all the land uses except one (agro-pastoral) contribute positively to the level of phosphorus, the strongest being bare soil and planted forest (eucalyptus). Since the model coefficients can only be considered different from zero at 85% (betas≠0 (85%)), we decided to restrict our interpretation to the simple fact that distance from the stream seems to be a determining factor for levels of phosphorus.

5.4. Fecal coliforms

Except for three water sampling stations (and the first three sampling sites with no data in January), the fecal coliform counts are higher during the wet season. Stations 6–8 have the highest counts which are related to the urban area of Belo Horizonte and its surrounding fringe. The lower values during the dry season are probably related to the almost complete absence of rain and runoff.

The models for January are not reliable because the null hypothesis that the betas are equal to zero could not be rejected. The July models, however, were very conclusive with a peak at $r^2 = 0.9$ and rejected $H0$ at 99.5% (Fig. 3f). This peak is reached at 150 m from the stream and then the relationship decreases steadily to $r^2 = 0.76$ at 510 m. The even lower $r^2 = 0.54$ for the ECZ suggests that this trend continues with the increased distance from the stream. Stock grazing near the stream is probably the main explanation for this behavior. It has been observed for instance that in the absence of riparian forest, stock tends to enter the stream more often. This further shows the importance of maintaining or restoring riparian forests.

As can be seen in Table 13, the strongest LULC contribution to fecal coliform counts comes from barren land followed by planted forest, agro-pastoral land and urban areas. These contributors can be merged into "human impact" areas. The inhibitors are riparian forest and savanna showing again the importance of maintaining natural vegetation near streams.

5.5. Discussion

The simplified approach used here was able to show a consistent link between LULC and WQ and that the former can be used to model the latter to some extent. Isolating LULC in riparian zones of different width made it possible to identify some patterns in the spatial behavior of the various water quality parameters with respect to stream distance. It did not, however, bring a significant increase in all the coefficients of determination of the regression models.

The models have also helped identify the LULC classes that have the most impact on WQ. In that respect, riparian forest is the single land use class that shows the most consistent behavior. It is a good inhibitor for turbidity, nitrate, nitrite and fecal coliforms. This behavior is consistent with that observed in many other studies although the exact weight varies widely which can probably be explained by regional differences such as climate, topography, soils, lithology, etc. Another very consistent finding is that barren land is a systematic contributor to the five WQ parameters selected. Its relative role in turbidity

Table 12
Multiple regression model for phosphorus (wet season/90 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	−0.661	1.50E-02	1.958E-02	11.133E-03	5.976E-03	−3.57E-03	1.720E-02	3.856E-03
Std. err.	0.554	0.009	0.008	0.006	0.007	0.010	0.006	0.005

C = constant; RF = Riparian forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted forest; U = Urban.

Table 13
Multiple regression model for fecal coliform (dry season/90 m, $n = 15$)

Variable	C	RF	B	S	F	AP	PF	U
B	−179283	2386.6	7513.6	−1899.4	1563.0	2078.3	5116.4	1717.7
Std. err.	152749.7	2409.9	2232.4	1675.8	1840.6	2686.0	1772.1	15001.4

C = constant; RF = Riparian forest; B = Barren; S = Savanna; F = Forest; AP = Agro-Pastrol; PF = Planted forest; U = Urban.

exceeds all the other LULC classes. The approach taken here did not however enable us to show this relation for the wet season. During the dry season, the agro-pastoral and planted forest classes consistently contribute to worsening WQ showing that when they are near streams, land conservation practices should be carefully applied. Savanna and secondary forest have a more complex behavior and are sometimes inhibitors, sometimes contributors. It should be considered here that both these classes of mainly open vegetation are often encountered in varying degrees of degradation and we were not able to make the LULC mapping carry this information. The urban class (taken here as a non-point source since only its area is considered) is a consistent contributor but with relatively small weight in the models retained (except phosphorus). This might come as an unexpected behavior but is explained by the fact that only two sampling points are directly affected by large urban centers and by the fact that, being mainly a point pollution source, its area might be of little relevance.

The riparian zone model used in this study was inspired by the study of Basnyat et al. (1999) but was simplified and adapted through cartographic modelling inside a GIS environment. However, unlike the results from Basnyat et al., the differences between the whole ECZs and the RZs are generally small. By subtracting water quality data from the previous point upstream and defining an exclusive contribution zone, we were able to analyze each sampling point as an independent process. The RZs were defined using the main river stream and its major tributaries but other smaller tributaries were left out.

The specific spatial behavior of each parameter with relation to the distance to the stream is perhaps the most interesting finding of the current study. Although by itself, this information does not help in defining legislation for riparian zone preservation, it can be useful for agricultural and environmental agencies giving advice on ideal width depending on the main activities. For instance a wider riparian zone should be preserved where large quantities of nitrogen-based fertilizers are used. On the other hand, the present legislation (30 m on each side of small streams) might be sufficient for grazing land. During the dry season, while the relationship between LULC and turbidity reaches it peak at 300 m, the fecal coliform coefficient of determination culminates at 90 m (or less but this could not be determined) to decrease steadily thereafter. On the contrary, the strength of the relation between LULC and nitrogen (nitrate and nitrite) increases with distance and reaches its peak at the maximum RZ buffer size. This last relation might be due to the fact that the main activities responsible for this parameter are not as significant near the stream. We feel that this aspect of the research, i.e. the spatial behavior of the relation between LULC and the different WQ parameters, merits more attention and should guide specific restrictions in land use at varying distances from the stream instead of the adoption of a single criterion (e.g. preserving 30 m of riparian vegetation) for all activities.

6. Conclusions

The strength of the statistical approach presented here lies in its simplicity and in the fact that it takes advantage of existing WQ analysis and a "coarse" resolution LULC map. It relies on the assumptions that the relation between WQ and LULC can be modelled without taking point pollution sources into account and without considering important hydrological variables like slope and precipitation. Slope could not, in this case, be properly modelled due to the lack of reliable data. Precipitation could not be adequately incorporated into the approach because the water samples were collected throughout the month (January or July) and not during a single day.

The results indicate that the LULC classes can be used to model turbidity, fecal coliforms, nitrogen and, to a lesser extent, phosphorus. The dry season appears to be a better choice for modelling these WQ parameters but more extensive research is needed to determine the effect of precipitation on the simplified approach presented here. They also indicate that each WQ parameter can have a distinct pattern with relation to distance from the stream and that these patterns can help define guidelines for watershed management.

The simplified approach that takes advantage of existing WQ data and LULC of medium-coarse resolution represents an important tool in the Brazilian context. As a developing country, adequate data for monitoring LULC changes and its effect on water quality are often lacking. Future research efforts will focus on incorporating precipitation and slope into the models in a smaller watershed having less disparate characteristics in order to better isolate the effect of LULC.

Acknowledgements

The research presented here is part of the Master's dissertation of Nádia Antônia Pinheiro Santos under the supervision of Dr. Philippe Maillard. The authors are extremely grateful to the *Instituto Mineiro de Gestão das Águas* (IGAM) for making available their water quality data and for providing logistic support and to Luciano V. Dutra from the *Instituto Nacional de Pesquisa Espacial* (INPE) for providing the LANDSAT data. Finally we would like to thank the three anonymous referees who have helped improve the present article with their most appropriate suggestions.

References

Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E., 1976. A land use and land cover classification system for use with remote sensing data. Geological Survey Professional Paper 964, US Government Printing Office for the US Department of the Interior, Geological Survey, Washington, D.C.

Basnyat, P., Teeter, L., Lockaby, B., Flynn, K., 1999. The use of remote sensing and GIS in watershed level analyses of non-point source pollution problems. Forest Ecology and Management 128, 65–73.

Biehl, L., Landgrebe, D., 2002. Multispec—a tool for multispectral - hyperspectral image data analysis. Computers and Geosciences 28 (10), 1153–1159.

Bollmann, H.A., Marques, D.M., 2000. Bases para a estrutura»c~ao de indicadores de qualidade de águas. Revista Brasileira de Recursos Hídricos 5 (1), 37–60.

Bruijnzeel, L.A., 1990. Hydrology of Moist Tropical Forests and Effects of Conversion: A State of Knowledge Review. Free University, Amsterdam.

Chavez Jr., P.S., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispetral data. Remote Sensing of Environment 24, 459–479.

Cornish, P.M., 2001. The effect of roading, harvesting and forest regeneration on streamwater turbidity levels in a moist eucalypt forest. Forest Ecology and Management 152, 293–312.

EMATER, 2005. Programas de Desenvolvimento. Empresa de Assistência Técnica e Extensão Rural do Estado de Minas Gerais—Secretaria de Estado da Agricultura—MG, 〈http://www.emater.mg.gov.br/siteemater/ServProd/Progdesenv.asp〉 (last accessed August 29, 2005).

Euclydes, H.P., Ferreira, P.A., 2002. Recursos Hídricos e Suporte Ecológico a Projetos Hidroagrícolas: Sub-Bacia do Alto e Médio São Francisco. Universidade Federal de Vicosa (Boletim técnico, 6), Vicosa, MG, Brazil.

Evans, R., 2005. Curtailing grazing-induced erosion in a small catchment and its environs, the Peak District, Central England. Applied Geography 25, 81–95.

Fisher, D., Steiner, J., Endale, D., Stuedemann, J., Schomberg, H., Franzluebbers, A., Wilkinson, S., 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.

Foster, I.D.L., Ilbery, B.W., Hinton, M.A., 1989. Agriculture and water quality: a preliminary examination of the Jersey nitrate problem. Applied Geography 9, 95–113.

Gardi, C., 2001. Land use, agronomic management and water quality in a small Northern Italian watershed. Agriculture, Ecosystems and Environment 87, 1–12.

IGAM, 2002. Qualidade das Águas Superficiais no Estado de Minas em 2001. Instituto Mineiro de Gestão das Águas. Karr, J.R., Schlosser, J, Water resources and the land-water interface. Science 201 (4352), 229-234.

Karr, J.R., Schlosser, I.J., 1978. Water resources and the land-water interface. Science 201, 229–234.

Mattikalli, N.M., Richards, K.S., 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.

McCulloch, J.S.G., Robinson, M., 1993. History of forest hydrology. Journal of Hydrology 150, 189–216.

Meybeck, M., Chapman, D., Helmen, P., 1989. Global freshwater quality: a first assessment. Global Environment Monitoring System/UNEP/WHO, Geneva.

MOE, 1984. Water Management: Goals, Policies, Objectives and Implementation Procedures of the Ministry of the Environment. Ministry of the Environment, Government of Canada.

Moore, J.A., Smyth, J., Baker, E., Miner, J., Moffitt, D., 1989. Modeling bacteria movement in livestock manure systems. Transactions of the American Society of Agricultural Engineers 32 (3), 1049–1053.

Nisbet, T.R., 2001. The role of forest management in controlling diffuse pollution in UK forestry. Forest Ecology and Management 143, 215–226.

Olsson, L., Pilesjo, P., 2002. Environmental Modelling with GIS and Remote Sensing, 1st Edition. Approaches to spatially distributed hydrological modelling in a GIS environment, Taylor and Francis, London, GB, Ch. pp. 167–199.

Paula Lima (de), W., Brito Zakia, M.J., 2001. Mata Ciliares: Conservação e Recuperação, 2nd Edition. Universidade de São Paulo: Fapesp, São Paulo, SP, Ch. Hidrologia de matas ciliares, pp. 33–44.

Petersen, R.C., 1992. The RCE: the riparian, channel, and environmental inventory for small streams in the agricultural landscape. Freshwater Biology 27, 295–306.

Polignano, V.M., Alves, A.L., Machado, A.T.M., Lisboa, A.H., 2001. Uma viagem ao Projeto Manuelzão e à Bacia do Rio das Velhas. Universidade Federal de Minas Gerais—Projeto Manuelzão, Belo Horizonte, Brazil.

Ratter, J.A., Ribeiro, J.F., Bridgewater, S., 1997. The brazilian cerrado vegetation and threats to its biodiversity. Annals of Botany 80, 223–230.

Schuft, M.J., Moser, T.J., Wigington Jr., P.J., Stevens Jr., D.L., McAllister, L.S., Chapman, S.S., Ernst, T.L., 1999. Development of landscape metrics for characterizing riparian-stream networks. Photogrammetric Engineering and Remote Sensing 65 (10), 1157–1168.

Sokolonski, H.H., 1999. Manual Técnico de Uso da Terra. IBGE: Departamento de Recursos Naturais e Estudos Ambientais, Rio de Janeiro – RJ.

Sparovek, G., Ranieri, S.B.L., Gassner, A., De Maria, I.C., Schnug, E., Santos (dos), R.F., Joubert, A., 2002. A conceptual framework for the defnition of the optimal width of riparian forests. Agriculture, Ecosystems and Environment 90, 169–175.

Stout, W., Fales, S., Muller, L., Schnabel, R., Weaver, S., 2000. Water quality implications of nitrate leaching from intensively grazed pasture swards in the northeast US. Agriculture, Ecosystems and Environment 77, 203–210.

Tong, S.T.Y., Chen, W., 2002. Modeling the relationship between land use and surface water quality. Journal of Environmental Management 66, 377–393.

Wang, X., 2001. Integrating water-quality management and land-use planning in a watershed context. Journal of Environmental Management 61, 25–36.