Large-scale analytical water quality model coupled with GIS for simulation of point sourced pollutant discharges

Fernando Mainardi Fan a,*, Ayan Santos Fleischmann a, Walter Collischonn a, Daniel P. Ames b, Daniel Rigo c

a Instituto de Pesquisas Hidráulicas (IPH), Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves, 9500, CEP 91501-970, 15029, Porto Alegre, RS, Brazil
b Brigham Young University, Department of Civil and Environmental Engineering, 368 Clyde Building, Provo, UT, 84602, USA
c Universidade Federal do Espírito Santo, Centro Tecnológico, Departamento Engenharia Ambiental, Av. Fernando Ferrari 5/N, Bairro Goiabeiras, CEP 29060-970, 01901, Vitoria, ES, Brazil

Article info
Article history:
Received 19 March 2014
Received in revised form
11 November 2014
Accepted 14 November 2014
Available online

Keywords:
Water quality
Modeling
SIAQUA-IPH

A B S T R A C T

Mathematical modeling is an important tool for water quality studies, and the integration of water quality models with geographic information systems (GIS) is very useful for information extraction and for results interpretation. In this context, this work presents the development of a water quality model coupled with GIS (MapWindow GIS) for representing impacts of point-sourced pollutants released with distinct durations under different flow scenarios, allowing a systemic view of the entire basin, and capable of being used with low data availability. The model is called SIAQUA-IPH and uses a pollutograph convolution scheme to represent multiple discharges and confluences in the basin, based on analytical solutions of the longitudinal advection-dispersion equation. Operational tests presented a full operational performance from all technical solutions adopted, and a representation of plumes considered satisfactory in comparison to observations. Additionally, a simple sensitivity analysis is presented, that gives useful insights about the model application.

© 2014 Elsevier Ltd. All rights reserved.

Software availability

Name of software: SIAQUA-IPH
First available year: 2013
Software requirements: Windows, MapWindow GIS
Programming language: VB.Net
Program availability and cost: The software application SIAQUA-IPH is a free plugin of MapWindow GIS. Current version is fully in Portuguese. An English version may be issued on demand.
Contact persons: Fernando Mainardi Fan, Ayan Santos Fleischmann, Walter Collischonn
E-mails: fernando.fan@ufrgs.br, ayan.fleischmann@gmail.com, collischonn@iph.ufrgs.br

1. Introduction

Water quality modeling consists in the use of mathematical equations to simulate the fate and transport processes of pollutants spilled into water bodies. Currently, there are several models developed for water quality simulation, which are generally used to assess impacts of pollutant releases on relatively small river reaches. They are usually applied for water resources planning, industrial and urban wastewater control, and evaluation of ecological issues. Typical models include QUAL2kw, WASP7, MIKE-11, ISIS and RWQM1 (Riecken, 1995; Rauch et al., 1998; Cox, 2003; Pelletier et al., 2006; Benedini, 2011; Cardona et al., 2011; Sharma and Kansal, 2013; Black et al., 2014).
These models have been successfully used for representing relatively small regions of water bodies, and can be referred to as reduced spatial scale models. On the other hand, there are water quality problems that must be analyzed in larger spatial scales, covering basins of several tens of km², or hydrographic networks with rivers reaches from hundreds to thousands of km. Examples of such problems are the accidental releases of hazardous substances, such as the disruption of mining tailings dams or accidents involving trucks carrying hazardous materials.

In Brazil, where the model presented in this work was developed and is intended to be applied, these kinds of problems are not uncommon: in March 2003, for example, after the failure of a tailing dam in the city of Cataguases, state of Minas Gerais, the pollution reached the small Cagados River and then the Pomba River, which in turn contaminated the Paraíba do Sul River in the State of Rio de Janeiro, leaving people without water supply for days at different points in the basin. Environmental authorities had to deal with impacts on water quality at great distances from the pollutant source (Gonçalves et al., 2007). Furthermore, the growing expansion of Brazilian road network has created a demand for the development of dangerous material transport management plans, with the identification of critical points, such as river crossing points, and the analysis of accidents scenarios and environmental risks (IPR/DNIT, 2005), for which water quality modeling is a demanded tool.

Scale-related problems also occur in water resources planning, where models representing large basins are usually necessary. In this case, there are several points of pollutant releases and results are needed for different development scenarios at many points of interest.

Furthermore, the release points of pollutants can be located on small tributaries of large rivers, and for this reason it is not enough to simulate only the main rivers in mathematical models, but it is also necessary to represent, although in a simplified way, the entire river network of the basin.

The coupling of water quality models with geographic information systems (GIS) can be especially interesting in these large-scale problems, because it allows an easier assessment of the full possible impacted region with a relatively low cost and an integrated view of the region in terms of physical and occupation aspects. For instance, when the location of a pollutant release in the basin is not known a priori, as in the case of accidental contaminant releases, integration of models and GIS is especially useful. Likewise, in most cases it is not initially clear where results are desired, or if water quality information (e.g. monitoring locations) will be necessary at different places of the basin.

The integration of water quality models with GIS is a field that has developed in recent decades, with works such as Hession and Shanholz (1988), Olivieri et al. (1991), Joao and Walsh (1992) and Tim et al. (1992), which used GIS platforms for generating model input data. Subsequently, there were studies using GIS platforms as model interfaces, such as Srinivason and Arnold (1994), USEPA, 2003, USACE (2003), Jayakrishnan et al. (2005), Ruelland et al. (2007) and Lin et al. (2011). Finally, the recent trends also include the development of fully integrated applications with GIS, as shown in the works of Foster and Mcondon (2000), Di Luzio et al. (2004), Strager et al. (2010), Zhang et al. (2011), Ferrer et al. (2012), Samuels et al. (2014), Formetta et al. (2014), and Bhatt et al., (2014), who developed models and methods coupled into GIS platforms structure (or in the opposite way, with GIS platforms coupled within models structure). Some of these couplings have been developed for non-point source pollution analysis, focusing mainly on agricultural issues (Srivastava et al., 2001; Usery et al., 2004; Borahm and Bera, 2003).

The authors Liu et al. (2013) suggest that these methods of coupling GIS and models can be classified in three categories: (i) loose coupling, when the model works unrelated to GIS, which is used only for pre-, post-processing and visualization of spatial features and/or results; (ii) tight coupling, when the model works within the GIS, with all functions of data generation, processing and visualization running through the GIS. However, the modeling process works in parallel, and the GIS is necessary to start it (the model interface is GIS platform); (iii) fully integrated coupling, when the model equations are programmed within the GIS own code, or when a GIS structure is programmed within a model itself. This last form of coupling, therefore, has two subdivisions, where Sui and Maggio (1999) actually present both as separated classes. It is important to mention that Liu et al. (2013) highlights that it is difficult to conclude that one type of integration is better than the others. And, therefore, the benefit of one or another can be dependent to the application.

This paper presents a water quality model fully coupled to GIS developed to simulate the impacts of point-source pollutants discharges in large basins in situations of relative lack of data. It is a typical situation in Brazilian water courses, where flow-rate in each river section is rarely known, and general cross-sections characteristics that are necessary to calculate water velocity are usually unavailable (part of this problem comes from the continental size of the country). Similarly, when accidental spilling occurs, there is no detailed information on the amount of material released.

This model simulates the propagation of point-source pollution releases using a technique based on analytical solutions of the dispersion equation applied to each reach of a vector drainage network, considering different flow scenarios (low, medium and high flow rates). It is applicable to large basins where the effluent discharge points and the points of interests are not known a priori, and operates completely integrated with a GIS. The GIS used is the open source MapWindow GIS, that allows the development of tools as plugins and has the advantage of being free, little resource-intensive, and have customizability options.

The model was named SIAQUA-IPH, which is an acronym for Analytical Simulator of Water Quality (in Portuguese), followed by the abbreviation of Hydraulic Research Institute (IPH, also in Portuguese), where the model was developed. The components and structure of SIAQUA-IPH are presented together with the results of a performance analysis of the model using data from a tracer test carried out by Rigo (1992) in the Paráiba do Sul River.

2. Methods

SIAQUA-IPH model application involves five steps: (i) pre-processing of geospatial data, (ii) definition of hydraulic attributes, (iii) inclusion of pollutant releases, (iv) pollutant propagation, (v) parameters calibration (if necessary) and (vi) post-processing for results displaying. These steps are described in detail in the following sections: “Geospatial Data Pre-Processing”; “Hydraulic Attributes Setting”; “Inclusion of Pollutant Releases”; “Pollutant Fate and Transport”; “Model Coefficients Estimation and Calibration”; and “Computational implementation Integrated with GIS”, which addresses the post-processing routines and the available tools in the model interface.

2.1. Geospatial Data Pre-Processing

SIAQUA-IPH model represents the watershed river network with connected river reaches (Fig. 1). If the vector drainage network is not available, it can be obtained in a prior step, using GIS tools applied to the analysis of a digital elevation model (DEM) data, such as the Arc Hydro toolset (Maidment, 2002) and the TauDEM toolset (Tarboton, 2002).

The drainage network is composed of two kinds of river reaches: incremental reaches and headwater ones. One incremental reach is defined as the river length between two confluentes. A headwater reach is defined as the extension of a watercourse from the springhead to the first junction (see Fig. 1 for an example of a drainage network). The springhead positions are defined by minimum drainage area criteria, from which concentrated flow occurs (Tarboton, 2002). Each river reach has an identification code and seven other characteristics or attributes required for pollutants spread calculation: length, slope, width, drainage area, streamflow, average water velocity, and the identifier code of the river stretch located right downstream.
The river reach identification code is an arbitrary number associated with the stretch during preprocessing. An important attribute is the downstream reach code, which defines the connection between the component parts of the drainage network. This attribute is automatically set in the preprocessing step using tools like Arc Hydro (Maidment, 2002). Fig. 1 presents a simplified example of a drainage network with river reaches encoded. In Fig. 1 the river reach 94 is located upstream from reach 119, so the water and pollutants spilled on 94 go straight to 119.

Length, slope and drainage area of each reach are calculated by an algorithm similar to that presented by Paz et al. (2008). In this algorithm the length is derived by analyzing a flow direction raster file and accumulating distances between cells, which depend on the size of the cells and on the travel direction between them. Slope is computed by dividing the difference in altitude values between the most upstream and downstream cells by the reach length. Lastly, drainage area is obtained by summing areas of all upstream cells for the reach, which are defined with the flow direction grid. All these routines are performed automatically during preprocessing.

2.2. Hydraulic attribute settings

A steady one-dimensional streamflow was adopted in SIAQUA-IPH model, i.e., the flow rates in each river reach do not change over time. Thus, the model computes reference flow values ($Q_{90}$, $Q_{70}$, $Q_{50}$, $Q_{30}$ and $Q_{10}$), where $Q_9$ is a flow with a probability of exceedance of $x \%$) for each river reach, and the user can choose between low ($Q_{90}$ and $Q_{70}$), medium ($Q_{50}$) and high flow scenarios ($Q_{30}$ and $Q_{10}$) to be simulated. The chosen scenario is adopted for the entire basin, i.e., if one selects the setting of $Q_{90}$ then it is considered that each river reach in the basin has a flow equal to its respective $Q_{90}$. Pollutant propagation is calculated for a complete mixing scenario, considering the length of the non-complete mixing zone calculated with equations proposed by Taylor (1954) and Fischer (1975), and that pollutant propagation results can only be seen downstream of this length.

Reference streamflow values can be obtained by two ways: streamflow regionalization or data reading from an input file. In the case of streamflow regionalization, the reference flows of each river reach are computed using its drainage area, through a previously developed equation that relates the drainage area of a basin to its reference flow. Such equations are available for many regions of Brazil. Streamflow values can also be obtained by reading a data file. At this point, SIAQUA–IPH model uses the same type of discretization in small catchments and river reaches adopted in the hydrological rainfall-runoff model MGB-IPH (Collischonn et al., 2007; Paiva et al., 2013). The MGB-IPH model is a large-scale hydrological model commonly used in Brazil that calculates the river streamflows into a basin from precipitation and other climate variables data (Collischonn et al., 2007). Recent developments of the model (Paiva et al., 2013) also includes the consideration of hydrodynamic flow routing. In MGB-IPH model, routines are included for calculation of reference flows in each river reach, which generate a file with $Q_{90}$, $Q_{90}$, $Q_{90}$, $Q_{90}$ and $Q_{90}$ values that can be read by SIAQUA-IPH. The average velocity of the water is automatically calculated assuming the Manning’s equation, based on the flow rate defined for each river reach and considering that the flow is steady and uniform. Manning’s $n$ coefficient can be altered while using MGB-IPH, and by default it is usually adopted a value of 0.035. The cross section can be chosen between rectangular and trapezoidal shapes, which are commonly used ones.

The riverbed width of each river reach is necessary for velocity calculation and is obtained from a relationship between the river width and the drainage area of the basin at the end of the considered reach. These relationships can be drawn from cross section information from river flow gauging stations, being this method generally valid for watersheds with homogeneous physical characteristics. Two examples where such geomorphological relations were applied satisfactorily for hydrological studies can be found in Collischonn et al. (2007) and in Paiva et al. (2013). The model interface allows the user to edit manually the regression parameters; otherwise, predetermined values obtained are used for width estimation.

2.3. Inclusion of pollutant releases

According to pollutant release duration, three types of entries can be simulated within the model: a) continuous releases; b) instantaneous spills; c) limited duration releases. Fig. 2 shows these three types of releases. Continuous type of pollutant release is the one that occurs continuously without changes in quantity and concentration over time. Instantaneous release is the one which occurs when a pollutant load is abruptly discharged over a very short period of time. The transient limited duration release is the one that begins in a moment of time and has a finite duration over which the amount and concentration of the pollutant is constant, and ends in a later instant of time (t2).

SIAQUA-IPH was also developed to simulate sewer effluent discharges in the river reaches using Streeter-Phelps model (Streeter and Phelps, 1925) if one would like to consider it as a continuous release, but permanent in time considering that the discharge does not have an initial time specified, and longitudinal dispersion effects are not considered.

2.4. Pollutant fate and transport

The calculations of pollutant fate and transport along the drainage network are performed for each river reach using analytical solutions of the transport equation. The equations implemented in the model consider processes of advection, dispersion and decay along a reach, and dilution effects due to incremental flows and confluences. A convolution scheme to represent multiple discharges and confluences in the basin was adopted, and is the main principle of the model.

Given an instantaneous release of a pollutant with mass M at a point x = 0 on a river, the concentration at any downstream point located at x, where x is the distance in meters from the discharge point, and at a moment t, can be calculated by Equation 1 (Chapra, 1997; Chanson, 2004):

$$c(x, t) = \frac{M}{A \sqrt{4 \pi \cdot E_L \cdot t}} \exp \left( - \frac{(x - u - \frac{t}{E_L} u)^2}{4 \cdot E_L \cdot t} - k \cdot t \right)$$

(1)

Where $c(x, t)$ is the concentration at location x at time t (mg L$^{-1}$); $E_L$ is the longitudinal dispersion coefficient (m$^2$ s$^{-1}$); u is the mean velocity of water in the longitudinal direction (m s$^{-1}$); M is the mass injected into the system (g); A is the cross-sectional area where the longitudinal dispersion occurs (m$^2$); and $k$ is the coefficient of first order decay ($s^{-1}$). This equation simulates the processes of advection, dispersion and first-order decay of solutes. It is a one-dimension formulation, and thus is only applicable where complete lateral mixing is attained. Some rivers can have significant lateral fluxes, what is not here considered. Another physical process that is neglected is that of transient storage, where solutes are temporarily stored in dead zones and hyporheic zone, leading to pollutants with long tails and smaller peak concentrations. Lastly, Equation (1) assumes that dispersion coefficient and mean velocity are constant in space and time (Runkel and Bengala, 1995; Bengala and Walters, 1983).

As an example for SIAQUA-IPH modeling scheme, consider the river basin presented in Fig. 1 and a pollutant spill of mass $M_0$ (g) in the river reach number 86. The pollutant dispersion on this reach is performed directly through the analytical solution of the longitudinal advection–dispersion equation. The result of this process is the transformation of a single mass spill that occurs in only one time step in a plume propagated by a part of the river reach (transformation of the pollutograph in Fig. 3b to the pollutograph in Fig. 3c).
To determine concentrations at the start of reach number 94, immediately downstream to the reach of discharge, two calculations are made: (i) decay and dispersion by travel time until the end of the of the reach 86; and (ii) dilution calculated at the end of section 86 due to incremental flows and confluences. Fig. 3 presents this process.

The calculation to be made for obtaining the pollutographs on the others river reaches, for example in the reach 119, cannot be done in the same manner shown for the two upstream parts (86 and 94), since the input information in the advection-dispersion equation analytical solution is the amount of mass discharged into the reach, and the information available, arising upstream, is a concentration profile scattered in time, in units of concentration.

To enable this dispersion calculation, it was adopted in SIAQUA-IPH a strategy where the model divides the upstream pollutograph (at the start of the reach number 94) in a finite number of parts, each part corresponding to one minute time step, as presented in Fig. 4b.

These small parts, which constitute the discretized pollutograph in the river reach 94, are multiplied by the flow rate, transforming them on time-distributed mass units. These units are then considered as instantaneous individual releases in the stream, and therefore are used as discharges occurring temporally distributed in this reach of the river, enabling the application of the analytical solution of the longitudinal advection-dispersion equation. Next, the resulting set of individual pollutographs can be summed to obtain the resultant pollutograph at the end of the reach number 94 (4c).

To calculate the concentration of pollutants in the river reach number 119, downstream of reach 94, after the sum of pollutographs, the model also makes the dilution of concentrations calculated with the flow of incremental river reach basin 94 and the flow from the reach number 112. Fig. 4 presents a simplified calculation of concentrations in the reach 119.

The same method is applied successively to other downstream river reaches until the end of the basin is reached. The result is a pollutograph calculated for each reach, viewed by an observer standing at a point downstream of the release. This process of discretization and propagation of individual releases is the main principle adopted in SIAQUA-IPH to calculate the pollutants fate and transport and can be called superposition or convolution of pollutographs, as commented by some authors (Jobson, 1997; Chanson, 2004; Ribeiro et al., 2011). According to them, the advection-dispersion differential equation is linear, so it is possible to apply a convolution scheme, using this principle of superposition to estimate the combined effect of a number of different releases. However, the sum of pollutographs is generally only used for representation of a series of spills, and not for the representation of river confluences, as assumed in SIAQUA-IPH. Convolution schemes are normally used in hydrological modeling for unit hydrograph calculations.

For releases with other durations than instantaneous (continuous and intermittent releases) the model uses the same calculation method through the convolution of pollutographs. However, it uses the analytical solutions for intermittent durations (O’Loughlin and Bowmer, 1975; Runkel, 1996; Chapra, 1997). Equation (2) is used only for continuous releases, while Equations (2) and (3) are used for intermittent releases.

\[
c(x, t) = \frac{C_0}{2} \left[ e^{-\frac{x^2}{4\sigma^2}} \text{erfc} \left( \frac{x - u \cdot t - \Gamma \cdot \sqrt{t}}{2\sqrt{\sigma^2 \cdot t}} \right) + e^{-\frac{x^2}{4\sigma^2}} \text{erfc} \left( \frac{x + u \cdot t + \Gamma \cdot \sqrt{t}}{2\sqrt{\sigma^2 \cdot t}} \right) \right]
\]

\[
c(x, t) = \frac{C_0}{2} \left[ e^{-\frac{x^2}{4\sigma^2}} \text{erfc} \left( \frac{x - u \cdot t - \Gamma \cdot \sqrt{t}}{2\sqrt{\sigma^2 \cdot t}} \right) - \text{erfc} \left( \frac{x - u \cdot (t - \tau)}{2\sqrt{\sigma^2 \cdot (t - \tau)}} \right) + e^{-\frac{x^2}{4\sigma^2}} \text{erfc} \left( \frac{x + u \cdot (t - \tau)}{2\sqrt{\sigma^2 \cdot (t - \tau)}} \right) - \text{erf} \left( \frac{x + u \cdot (t - \tau)}{2\sqrt{\sigma^2 \cdot (t - \tau)}} \right) \right]
\]

\[
\Gamma = \sqrt{1 + 4\eta}
\]

\[
\eta = \frac{k \cdot E_t}{u^2}
\]
Where \(c(x, t)\) is the concentration at the site \(x\) at time \(t\) (mg L\(^{-1}\)); \(c_0\) is the initial concentration of the discharge in the watercourse (mg L\(^{-1}\)); \(E_L\) is the longitudinal dispersion coefficient (m\(^2\) s\(^{-1}\)); \(t\) is the time instant at which one wish to know the concentration (s); \(t_2\) is the time instant at which stops the transient release (s); \(x\) is the point where one want to know the concentration (m); \(u\) is the mean flow velocity in the longitudinal direction (m s\(^{-1}\)); \(k\) is the environmental decay coefficient of the constituent (s\(^{-1}\)); and \(erfc(\cdot)\) is the complementary error function. Equation (2) is used when \(t\) is less than or equal to \(t_2\), and Equation (3) is used when \(t\) is greater than \(t_2\). These equations have the same applicability as Equation (1).

### 2.5. Model coefficients estimation and calibration

An important parameter for pollutant dispersion simulation is the longitudinal dispersion coefficient \((E_L)\), presented in Equations (1)–(3). In SIAQUA-IPH each river reach has its own longitudinal dispersion coefficient, what allows the representation of the river network variety of physical characteristics.

By default, \(E_L\) value of each river reach is initially estimated using an automated decision process based on physical (width and slope) and hydraulic (velocity) characteristics of the reach sections and on the range of applicability of 18 empirical equations inserted into the model (Fan, 2013). These equations were obtained from the works of Bowie et al. (1985), Roussis and Rodriguez-Mirasol (1998), and Kashifpour and Falconer (2002), which present comprehensive reviews of various existing empirical equations for determining \(E_L\). The standardized method inserted in SIAQUA-IPH selects the empirical equation that best suits the characteristics of the reach, aiming to reduce model complexity related to the choice of the dispersion coefficient value by users. However, these equations can also be manually selected in the model, or the user can manually insert the value of the coefficient.

Besides \(E_L\) coefficient, all other model parameters can be changed. These parameters are kinetic coefficients of the releases and physical information of river reaches, such as water velocity and river channel shape.

### 2.6. Computational implementation integrated with GIS

The SIAQUA-IPH model was implemented by programming a supplementary code ("plugin") which is added to the GIS software MapWindow GIS® (Ames et al., 2008; Aburizaiza and Ames, 2009). The MapWindow GIS is a free and open source program, has a set of basic tools of GIS, and allows the development of additional tools in the VB.Net or C# language. Fig. 5 presents one of the main windows of data entry in SIAQUA-IPH, where the main GIS menu can also be seen, through which it is possible to access the water quality model tools developed (highlighted).

After providing geospatial and hydraulic data for the model, the user inserts the pollutants in the main window of the GIS interface by clicking on the river reaches in the map and informing the release type and other characteristics, such as concentration, temporal extent and total load. Then, the model is run, calculating the pollutants fate and transport. Finally, results can be presented in the form of colored maps (Fig. 6), pollutographs (Fig. 7) or longitudinal profiles. Tables with the results can be exported.

### 3. Model application results

#### 3.1. Study case

In order to test SIAQUA-IPH, a comparison was performed between the model results and the concentrations obtained from a tracer test carried out by Rigo (1992). Quantification of the errors due to the model approximations was analyzed. The aim of the test was to examine SIAQUA-IPH performance and the improvement obtained by model calibration, considering that it is run with just few detailed data.

The tracer test was carried out in the Paraíba do Sul river basin, which is located in southeastern Brazil, comprising the territories of three Brazilian states. It is marked by a situation of diversified water use, including public water supply, power generation, disposal of sewage, irrigation and industries. It is a basin of great economic importance to Brazil, and a potential application field for SIAQUA-IPH as a system for water quality control, due to the presence of both industries and water intakes and a history of accidental spills.

The tracer test consisted in the release of a known quantity of Uranine at a point called S0 and the monitoring of concentrations curves passing by other points (S1 to S6), showed in Fig. 8. However results were only properly assessed at points S2, S3, S4 and S6, because at S1 the complete mixing of the tracer was not identified.
in the river, and point S5 is a replica of point S4 for quality control purposes.

Regarding the cross sections of the monitoring points they have a general U-shaped format with river widths around 120 m and 130 m, and depths around 3 m—5 m in the range of the tested flows. These sections were represented in the model using the regionalization approach described earlier in this text.

The monitoring points are presented in Fig. 9 and the test data are shown on Table 1. Further detailed description of the field data and tests can be found on Rigo (1992).

Fig. 5. One of the main windows of SIAQUA-IPH data entry and the MapWindow GIS® main menu, from where tools are accessed for water quality modeling (highlighted).

Fig. 6. Hypothetical example of a spill simulated by SIAQUA-IPH. River reach colors indicate the degree of impact. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. 7. Pollutograph resulting from a hypothetical benzene spill simulation by SIAQUA-IPH.

Fig. 8. Monitoring points for the Paraíba do Sul river.
3.2. Input data preparation using GIS

The spatial data used in the test were obtained through the TauDEM Tools package, available in MapWindow GIS. The total area of the Paraíba do Sul river basin is approximately 55,400 km², and by applying the toolkit it was divided on 13,878 stream reaches, with an average length of 2 km each. After the discretization, the input files for the water quality model was prepared using SIAQUA-IPH preprocessing tools.

The hydraulic input data for SIAQUA-IPH were obtained by using the MGB-IPH model. The hydrological model was calibrated for the study region using information of gauging stations. The division of the hydrographic network in the region and locations of gauging stations used are presented in Fig. 10.

The prepared input data were added to the model through the developed GIS coupled solutions. Subsequently, the tracer releases were inserted by selecting the appropriate river reach through the software interface.

3.3. Simulated events: model results, calibration and validation

The reported events were simulated using the above mentioned input data. As an example, the pollutographs obtained for the test of day 01/08/1987 are presented in Fig. 11. S1 point results are only showed as a reference point, because complete mixing (3.6 km downstream of S0) was not observed at this location.

Two simple metrics were used to evaluate model results: the differences between peak concentrations and the differences between peak times found for observed and simulated plumes. These metrics were chosen because they enable an evaluation of the two most important aspects of the model usage, namely the correct prediction of pollutants travel time and concentration values (in this case represented by peak concentration). Table 2 contains a summary of these metrics applied to model results.

In general the simulated plumes presented a delay in the peak concentration arrival, with peak time errors ranging from 2% to 17%. It was also obtained a greater dispersion than that observed in field, leading to lower peak concentrations and errors from \( \frac{1}{C_0} \times 15 \% \) to \( \frac{1}{C_0} \times 50 \% \) for the peak concentration difference. These inaccuracies are probably related to an overestimation of velocities (less expressive) and longitudinal dispersion coefficients (more expressive).

On the other hand, the model run without any detailed information about the river and its basin, and thus the results can be considered acceptable. Errors found in this comparison are useful to estimate typical mistakes that can be made when the model is applied under such conditions. Also, it is possible to improve the representation of curves through the calibration of SIAQUA-IPH parameters, obtaining a more reliable representation of plumes of pollutants.

For this study, a calibration strategy was adopted to identify general trends in the results, applying two coefficients to two model variables: one multiplier coefficient (C1) to correct water velocity \( u \) and other (C2) to correct longitudinal dispersion coefficient \( D_L \). These two parameters were chosen because travel time and dispersion were verified as important and calibrated information.

Table 1

<table>
<thead>
<tr>
<th>Release point, hour and date</th>
<th>Localization</th>
<th>Streamflow (m³/s)</th>
<th>Mass (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0; 6:35am 09/27/86</td>
<td>Paraíba do Sul River</td>
<td>272</td>
<td>10.0</td>
</tr>
<tr>
<td>S0; 6:35am 01/08/87</td>
<td>Paraíba do Sul River</td>
<td>342</td>
<td>5.0</td>
</tr>
<tr>
<td>S0; 7:00am 01/29/87</td>
<td>Paraíba do Sul River</td>
<td>639</td>
<td>5.0</td>
</tr>
<tr>
<td>S0; 7:00am 12/15/87</td>
<td>Paraíba do Sul River</td>
<td>294</td>
<td>5.0</td>
</tr>
</tbody>
</table>
independently and manually. It is important to note that the dispersion in the pollutograph is dependent on the dispersion coefficient, and also on the flow velocity, since a plume traveling for a longer time will also have more dispersion. Then, the calibration processes were conducted taking this factor into account, and aiming to decrease the two greatest biases in the results, related to delay and excess of dispersion.

Calibration was conducted for the 09/27/86 and 01/08/1987 events. The final pair of values obtained were $C_1 = 0.92$ and $C_2 = 0.2$. This means that after calibration velocities and dispersion coefficients were overall reduced by 8% and 80% respectively. This calibration agrees with the observed metrics results obtained by the raw model application, where velocities were less underestimated than peak concentrations, primarily due to excess of
dispersion. The comparison results obtained after calibration are presented on Fig. 12 for the 01/08/1987 test and on Table 3. The main benefits observed in the calibration are related to the reduction of the error metrics module. For example, considering the point S4 in the 09/27/86 experiment, peak concentration errors were reduced from −16% to 8%, and peak time errors from 10% to 3%. Moreover, considering the point S6 in the 01/08/87 release, peak concentration errors diminished from −50% to 11%, and peak time errors from 3% to −2%.

Following calibration, a validation analysis was conducted for the calibrated model using the 01/29/87 and 12/15/87 events datasets, not considered in the calibration procedure. The comparison results obtained after application of C1 and C2 coefficients are presented on Fig. 13 for the 01/29/1987 test and on Table 3.

Validation results are compatible with results obtained in the calibration, where peak concentration errors and peak time errors modules were reduced. This was expected since the kinds of errors were the same in both calibration and validation datasets (delayed plume and excess of dispersion). Yet the final module of errors was greater on the validation than in the calibration. (see Table 4).

As an example, considering the point S6 in the 01/29/87 validation release, the peak concentration errors were reduced from −41% to −35%, and peak time errors were reduced from 5% to 2%, while for the point S6 in the 01/08/87 calibration release the peak concentration errors were diminished from −50% to 9%, and peak time errors from 9% to −7%.

In general, the application of the correction coefficients decreased the model errors. All plumes were represented despite simplifications adopted, and the results can be considered promising.

These calibration/validation tests indicate that velocities and dispersion coefficients can be used to better represent releases occurred in rivers through an overall bias removing strategy. Then, following calibration a simple sensitivity analysis was conducted for both variables, intending to demonstrate and discuss how the estimative of these variables can affect the system.

### Table 2

Summary of differences between peak concentrations and peak times in results.

<table>
<thead>
<tr>
<th>Release</th>
<th>Flow (m³/s)</th>
<th>Point</th>
<th>Peak concentration difference (μg/L)</th>
<th>Peak concentration error (%)</th>
<th>Difference between peak times (h)</th>
<th>Peak time error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 kg; 09/27/86</td>
<td>272</td>
<td>S2</td>
<td>−7.46</td>
<td>−58</td>
<td>0.3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S3</td>
<td>−0.55</td>
<td>−15</td>
<td>1.1</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S4</td>
<td>−0.40</td>
<td>−16</td>
<td>1.2</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S6</td>
<td>−0.45</td>
<td>−23</td>
<td>2.9</td>
<td>17</td>
</tr>
<tr>
<td>5 kg; 01/08/87</td>
<td>342</td>
<td>S2</td>
<td>−2.02</td>
<td>−29</td>
<td>0.1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S4</td>
<td>−1.22</td>
<td>−46</td>
<td>−0.1</td>
<td>−1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S6</td>
<td>−0.95</td>
<td>−50</td>
<td>0.4</td>
<td>3</td>
</tr>
<tr>
<td>5 kg; 01/29/87</td>
<td>639</td>
<td>S2</td>
<td>−0.55</td>
<td>−11</td>
<td>0.3</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S4</td>
<td>−0.55</td>
<td>−25</td>
<td>0.1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S6</td>
<td>−0.75</td>
<td>−41</td>
<td>0.6</td>
<td>5</td>
</tr>
<tr>
<td>5 kg; 12/15/87</td>
<td>294</td>
<td>S2</td>
<td>−4.20</td>
<td>−49</td>
<td>0.1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S3</td>
<td>−1.86</td>
<td>−43</td>
<td>0.8</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S4</td>
<td>−1.99</td>
<td>−55</td>
<td>0.3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S6</td>
<td>−1.10</td>
<td>−50</td>
<td>1.2</td>
<td>9</td>
</tr>
</tbody>
</table>

Fig. 12. Simulation results of the 5 kg of Uranine release at the point S0, conducted at 01/29/1987 by Rigo (1992), for validation.
3.4. Sensitivity analysis

The sensitivity analysis consisted on varying velocities values by ±5%, ±10% and ±20% and dispersion coefficients by ±5%, ±10%, ±20%, ±30%, and ±50%. The test was applied to the 09/27/86 release, and results are presented on Fig. 13 for the S2 and S6 points.

Sensitivity analysis results provide some useful insights into the model usage. Firstly about velocity, it was observed that the velocity of the plume propagation affects directly the dispersion on a certain point, since a plume propagating for more time will also have more time to suffer dispersion. Second insight into velocities is given in the comparison between results of S2 and S6, where results of S6

![Fig. 13. Sensitivity analysis applied to the 09/27/86 release. Results for S2 and S6 locations.](image)
(more distant from the release source) are more affected by the variation in velocities than results of S2 (near to the release source). On S2 a change of ±20% on velocities led to errors of less than one hour on propagation of the peak, and on S6 the same errors corresponded to more than 2 h. In other words, a general bias in the velocities estimates tends to give greater discrepancies in more distant points, as the errors propagate for more time. The 20% alteration on velocity gave a great divergence in results for S6 points, and thus further tests were not conducted. This evidences the importance of good velocities estimates. In the applied case study this variable had a good initial guess given by field observation.

About longitudinal dispersion coefficient sensitivity, results show that $E_t$ only affects the plumes concentrations, as expected. A lower coefficient of dispersion leads to greater peaks, and a greater coefficient to lower and more spread peaks, but have no significant in the plume location (small displacements may be verified due to the time discretization for the convolution application). A variation of ±50% in the coefficient led to variations around ±29% on S2 plumes concentration peak and around ±28% on S6 ones, while a variation of ±20% issued for values around ±14% on S2 and around ±11% on S6. This information indicates that the variation of $E_t$ is not directly equal to the variation found on plume peaks. Also, it suggests that the impact depends on the distance to the release source, being more expressive near to the release (S2), probably because the plume is more concentrated at this point.

These results show that a calibration of the model is not trivial, since the variables perform differently with distance. Also, sensitivity results correspond perfectly to the performed calibration, where changes of 80% were induced in the $E_t$ to compensate errors from up to 50% in the peak concentrations. As it was shown in the previous analysis, the relationship between coefficient changes is not a forty-five degrees line.

4. Discussions and limitations

We believe that the developed software presented a useful performance for the study case related to plumes issuing, what could be also improved with a calibration strategy. The sensitivity analysis allowed a better understanding about the model two most important parameters, namely velocity and dispersion. It was demonstrated that general bias in velocities of 20% can create delays of 2 h on results, and errors in $E_t$ of around 50% can lead to errors of 28% on plumes peaks, indicating that a good estimative of both variables is an important point for reliable modeling. On the other hand, the software itself is designed as a tool to quickly evaluate approximate travel time and peak concentration of pollutant plumes in accidental spill events in Brazilian rivers, under the consideration of being applicable on data-poor situations, where in many cases the catchment hydraulic conditions will not be perfectly issued.

As highlighted by Bennet et al. (2013), in these data-poor situations it may be more interesting to carry out a qualitative model performance evaluation than a quantitative one, providing trends and system behavior instead of quantitative outputs. This means that the SIAQUA-IPH model usage can be more valuable if the possible errors are taken into account in a way similar to the here presented sensitivity analysis, performing multiple model runs considering possible inaccuracies. This is a possibility for the model application, since simulations based on the convolution of analytical solutions are fast, and variations on parameters values can be done by setting $C_1$ and $C_2$ values.

Finally, as a model developed based on the one dimensional solutions of the advection-dispersion equation, SIAQUA-IPH has some limitations that need to be considered in its application. The most important ones are that the current version of the system can only be applied to one-dimensional problems (rivers), and that features as reservoirs are not represented. Reaches with downstream effects in the streamflow (e.g. tidal rivers) and where transmission losses are significant (flows decrease as they progress downstream, for example, rivers flowing through deserts or arid zones) cannot also be modeled. Lastly, the adopted equations do not consider physical processes that may be important in the system description, such as lateral inflows and transient storage zones.

5. Conclusions

This paper presents the development of a water quality model called SIAQUA-IPH, which works coupled with the MapWindow GIS platform and needs just a few input data. The model is aimed to be used in large Brazilian river basins for pollution assessment, where data availability and model usability are important issues. Moreover, it has a versatile structure that can be used in any country and be adapted for small basins. The MapWindow GIS open source software makes the model more accessible in terms of costs, and also easily adaptable in terms of computational needs.

In the mathematical framework, it was decided that the model would use primarily analytical solutions, through a pollograph convolution scheme. The topology adopted in the model is the subdivision of the basin into a large number of small river reaches, where the results are calculated for each reach. This subdivision can be performed automatically using existing tools in GIS platforms, which is the first reason for the model integration with GIS.

As a second reason for the integration with GIS, SIAQUA-IPH was developed completely coupled with the MapWindow GIS platform, being free to use, thereby increasing the functionality and applicability of the model to facilitate the preparation of input information, spills data entry and results interpretation and visualization.

The adopted GIS coupling method can be classified as full coupling. In this case, the coupling permits the full use of GIS functionality in terms of layers and spatial variables information management, and speed up the information fluxes between the GIS and the model equations inputs/outputs.

To evaluate the operability of the model and all proposed methods, a tracer test was carried out in the Paraíba do Sul River in Brazil. The tests presented an adequate performance of the adopted methods and coupling solutions. The observed plumes were represented by the model with a small delay in the concentration peak arrival and a greater dispersion than that observed on field. After calibration and in the verification, we believe that satisfactory plumes were obtained. These results can be considered as promising in these cases, where despite the simplifications adopted, all tracer plumes were reasonably represented.

Also, the sensitivity analysis for the tested case provided some useful insights into the model usage. General bias in the velocities of flow showed to be important, and more significant as one move downstream in the analysis, for both correct timing and concentration prediction. And dispersion coefficients showed to be important for correct peak concentrations assessment and plumes duration, where the relationship between changes in values and bias issued was shown to be dependent on distance from the source not directly equal to the variation found on plume peaks.

Water quality models that are developed to represent different types of releases, including accidental ones, need to quickly provide results in order to support decisions such as the interruption of a water intake located downstream of the spill source. Although this is not the only intended use for SIAQUA-IPH, it is believed that this model can be very useful for such cases, as a decision-making tool for members of technical groups of environmental and sanitation companies in Brazil.
The software developed until now is fully operational in a technical perspective, what represents a foundation for further developments and testing. For future developments, additional features of the model can be the representation of reservoirs in the river reaches, thermal plumes simulation, adoption of a non-steady state flow regime, incorporation of diffuse pollution inputs and the creation of an uncertainty analysis system, allowing a probabilistic result of the simulations, testing of other kind of equations for plumes simulation, such as dead zones approaches, and conduct a benchmark evaluation between SIAQUIA-IPH and other available software in terms of results, time of processing, and numerical dispersion issues. Lastly, more studies to evaluate the model results with tracer tests may be conducted.

Acknowledgments

The first author and the second author of this paper thank the Brazilian CNPQ and Brazilian FAPERGS institutions for the scholarships provided during the research development. We also gratefully thank this paper editor and the two anonymous reviewers for the suggestions that helped to improve the quality of our work.

References


