Deterministic and Stochastic Principles to Convert Discrete Water Quality Data into Continuous Time Series

Danieli Mara Ferreira (✉ danielimaraferreira@gmail.com)

Federal University of Parana: Universidade Federal do Parana <https://orcid.org/0000-0002-6208-7178>

Marcelo Coelho

Federal University of Parana: Universidade Federal do Parana

Cristovão Vicente Scapulatempo Fernandes

Federal University of Parana: Universidade Federal do Parana

Eloy Kaviski

Federal University of Parana: Universidade Federal do Parana

Daniel Henrique Marco Detzel

Federal University of Parana: Universidade Federal do Parana

Research Article

Keywords: Water quality time series , Stochastic modelling , Deterministic modelling , SihQual model , 23
Water resources planning and management

Posted Date: April 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-359684/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Water Resources Management on July 20th, 2021. See the published version at <https://doi.org/10.1007/s11269-021-02908-1>.

Deterministic and stochastic principles to convert discrete water quality data into continuous time series

Danieli Mara Ferreira · Marcelo Coelho · Cristovão

Vicente Scapulatempo Fernandes · Eloy Kaviski · Daniel

Henrique Marco Detzel

the date of receipt and acceptance should be inserted later

Abstract Limited water quality data is often responsible for incorrect model description and misleading interpretation in water resources planning and management scenarios. This study compares two hybrid strategies to convert discrete concentration data into continuous daily values for one year in different river sections. Model A is based on an autoregressive process, accounting for serial correlation, water quality historical characteristics (mean and standard deviation) and random variability; the second approach (model B) is a regression model, based on the relationship between monitoring flow and concentrations, plus an error term. The generated series (here referred to as synthetic series) are propagated in time and space by a full deterministic model (SihQual), that solves the Saint-Venant and advection-dispersion-reaction equations. Results reveal that both approaches are appropriate to reproduce the variability of biochemical oxygen demand and organic nitrogen concentrations, leading to the conclusion that the combination of deterministic/empirical and stochastic components are compatible. A second outcome arises from the comparison of results in different time scales, supporting the need for further assessment of statistical characteristics of water quality data - which relies on monitoring plans. Nonetheless, the proposed methods are suitable to estimate multiple scenarios of interest in water resources planning and management.

Keywords Water quality time series · Stochastic modelling · Deterministic modelling · SihQual model · Water resources planning and management

D. M. Ferreira
Graduate Program on Water Resources and Environmental Engineering, Federal University of Paraná, Curitiba, Brazil
E-mail: danielimaraferreira@gmail.com

1 Introduction

The use of models for water resources management and planning purposes has been extensively reported (e. g. Yao et al., 2019; Yaseen et al., 2019). Although its importance is recognized, multiple sources of uncertainty in modelling are acknowledged, such as process description and numerical approximations (Guzman et al., 2015). Among these, input data limitations if often indicated as one of the main challenges (Strokal et al., 2019).

Although high frequency data has become more available with increasing technology (e. g. Bowes et al., 2016; Miller et al., 2016) monitoring plans remain limited by several conditions, such as those related to sample collection, integrity and laboratory analysis. Furthermore, data acquisition through sensors and remote sensing techniques is often limited due to calibration and validation issues, presence of interference and spectral resolution (Gholizadeh et al., 2016). Since concentrations of water quality parameters are affected by several external conditions (meteorological, hydrological and geomorphological characteristics, sources of pollution, land use etc), this challenge will remain as long as new compounds are produced and discharged at the environment, highlighting the importance of studies related to data requirements for modelling, management and research purposes.

In this sense, the use of statistical analysis becomes useful to understand and describe a phenomenon using past observations. In the water quality problem, usual strategies include interpolation functions (Gnauck, 2004), autoregressive models (Babu and Sure, 2016) and regression techniques (Huang et al., 2017). More recently, Tiyyasha et al. (2020) presented a thorough review of methods and examples based on artificial intelligence principles for river water quality modelling, including black-box and white-box models. While reviewing several techniques, the authors call attention to the common issue of missing data and data unavailability, and recommend integrating features of physiochemical aspects with numerical approaches to improve results.

Another relevant aspect in this context is proper communication, one of the challenges for model and theories acceptance in real-life applications. Taherdoost (2018) stated that numerous studies do not provide clear guidelines for operational aspects. Moreover, mechanisms to identify data anomalies and correlations may be useful to distinguish and communicate simulations conclusions, such as information on system accuracy, comparative assessment, visual data enhancement, hierarchically and sequentially arrangement of spatial-temporal patterns (Marrin, 2017; Tiyyasha et al., 2020). This is especially relevant

to assist acceptability by stakeholders and managers, since the temporal analysis is not usually accounted for water resources legislation in most countries, including Brazil. Although the management approach is often interested in long-term averages, one of the benefits of describing a system on fine temporal scales is an estimate of possible variance and uncertainty (Uusitalo et al., 2015).

In this context, the present study advances in the subject of conversion of discrete data into continuous information, comparing two simplified hybrid methods. Both are based on an integration of deterministic and stochastic principles, using a historical monitoring dataset of flows, biochemical oxygen demand and organic nitrogen concentrations gathered since the 1980s in five sections along 85 km of the Iguaçu river, located at Paraná state, Brazil. The generated series of each model, here refereed as synthetic series, are input for the deterministic module in the SihQual model (Hydrodynamic and Water Quality Simulation), to test their physical meaning – this integration of stochastic and deterministic approaches have been consolidated in a previous study (Ferreira et al., 2019). Comparisons are made in terms of overall data distribution and transgression of legal limits against observed data; the goal of grouping is to identify general patterns and representation gaps.

While it is often recognized in the literature that data representation is a key aspect in modelling studies, the contribution brought by this study is the exploration of different approaches to convert temporal scales of water quality information, and to show how this process should be improved; although the methods are simple, synthetic series generation must be carefully assessed. This study also presents recommendations towards the application of these strategies to assess water quality over time and space, aiming for more reliable analysis for water resources planning and management purposes. Furthermore, the analysis points to monitoring needs and research efforts.

2 Modelling strategy

The experiments are based on the integration of five simulation modules, as presented in figure 1: (i) deterministic hydrodynamic model, (ii) deterministic water quality model under unsteady state, (iii) kinetics rates model, (iv) stochastic model A and (v) stochastic model B. The integrated analysis seeks to convert historical water quantity and quality data – originated from discrete sampling – to continuous information. This highlights the use, differences and role of the stochastic models A and B.

1 The first module solves the Saint-Venant equations, providing the advection field and cross sections
2 areas where the pollutants are diluted and transported. This information is applied to solve the one-
3 dimensional water quality equation, which is a mass balance considering the processes of advection, dis-
4 persion and reaction. Other required inputs in this module are upstream and lateral boundary conditions,
5 dispersion and reaction coefficients. The stochastic model A assists the determination of transformation
6 rates, besides providing a time series as input for the deterministic water quality module. Model B is also
7 tested as an auxiliary to generate the upstream boundary condition.

8 Method A is based on an autoregressive process, that accounts for series correlation, statistical traits
9 (mean and standard deviation) and a random noise – which deals with the inherent uncertainty related to
10 data representativeness and external factors, leading to multiple possible combinations. This latter aspect
11 is solved with an algorithm to select one series most similar to observed data. The second approach is
12 a regression model, in which concentrations are generated from observed flows. It is also a stochastic
13 approach, since a lognormal probability density function is fitted to the error of the regression model;
14 hence, an infinite number of synthetic water quality time series can be generated from a unique streamflow
15 time series.

16 Both methods are parsimonious, aiming at simplified estimates for changes (specifically concentration
17 statistical metrics and flow conditions) in overall water quality variability in river control sections.

Fig. 1 Schematic of modelling integration; the dashed line highlights the main features of interest in this study

2.1 Stochastic model A

The first order autoregressive process, described by Loucks and Beek (2017), is the basis of model A to convert discrete information into a continuous dataset:

$$C_{j+1} = \mu + \rho (C_j - \mu) + z_j \sigma \sqrt{1 - \rho^2} \quad (1)$$

where C_j represents concentrations at the interval j , μ is the mean and σ defines the standard deviation of C ; ρ is the autocorrelation parameter (deterministic component), and z_j represents a gaussian random noise (stochastic component).

The stochastic component is a set of a thousand series with normal random distribution, zero mean and unit variance. Because the historical dataset is not continuous, the coefficient ρ is assumed – it represents the sample correlation parameter, which indicates the dependency between lagged concentrations.

The model results were limited to a calculated limit for extremely high values, which is the third quartile (75th percentile) plus 3 times the interquartile range:

$$lim = Q1 + 3(Q3 - Q1) \quad (2)$$

In equation 2, *lim* is the limit defined, while *Q1* and *Q3* represent first and third quartiles, respectively. This is one of the standard criteria to define outliers, although other methods could be used (Helsel et al., 2019).

An upper limit is needed to control the generation of values without physical meaning. Also, it guarantees that generated data is not too far from observations, while critical events of maximum concentrations are still accounted for.

Because multiple daily time series are generated, a series similar to the data measured in the simulated period is then selected considering the minimum difference between observed and fitted series quantiles: quartiles 1, 2 and 3 and concentration of 10 and 90% of occurrence.

Even though this method does not represent variability patterns of trend and daily cycles, for example – although test TD accounts for seasonality –, the model has been successfully applied to reproduce diverse study cases, including daily data (Ferreira et al., 2019).

2.2 Stochastic model B

A linear regression model is established between flows and concentrations, generating concentrations from observed flows. It is also a stochastic approach, since a lognormal probability density function (pdf) is fitted to the error of the regression model; hence, an infinite number of synthetic water quality time series can be generated from a unique streamflow time series – a number of 1000 is set in this study. Since the model is applied on a real daily flow time series, if variability patterns (trends, cycles and/or shifts) are present, they will be partially represented in the synthetic concentrations.

The generation of synthetic daily concentrations from daily flows was performed in three main steps: i) filling gaps of missing data in the daily flow time series; ii) building a regression model for flows and concentrations and; iii) generation of synthetic daily concentrations.

The first step is based on a relationship among different stations, which is detailed in the supplementary material. In the following phase, a quadratic polynomial regression model is fitted to flows and concentrations of each station:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + e \quad (3)$$

A 3-parameter lognormal pdf (LN3) was fitted to the errors of the model (e) – this was the best fitted condition, as discussed in Coelho (2019).

In the last step, time series of synthetic daily concentrations were generated by applying the regression model on the daily flow time series of each station. Negative values were avoided by performing a new draw when the model resulted negative. The model also limits the generated values according to equation 2.

2.3 SihQual – Hydrodynamic and Water Quality Simulation

The SIHQAL model solves the hydrodynamic problem coupled with water quality. It is a tool to study the fate and transport of pollutants in rivers under steady and unsteady state, implementing deterministic and statistical strategies (Ferreira et al., 2016, 2019).

The deterministic module refers to the Saint-Venant equations and the one-dimensional advection-dispersion-reaction, while the second approach explores simplified methods of a synthetic series generation – models A and B in this study. This latter module was first developed to meet the temporal scale required for numerical solution, since water quality data is often unavailable as high frequency samples. Any intermediary scale compatibility for the numerical solution is achieved through piecewise cubic interpolation.

The numerical methods applied to solve the partial differential equations are explicit finite differences, which is an advantage to implement different modelling configurations – as those introduced in this study.

In the water quality component, a constant dispersion coefficient is assumed, while point and diffuse contributions from the basin are calculated using population data and export coefficients – details are in Ferreira et al. (2016). The calibration techniques, named TRATS (Transformation Rates Time Series) generates daily decay rates as a result of: (i) attributes of the river station, (ii) reference values from the literature, and (iii) random variability. Details are presented in Ferreira et al. (2020).

The water quality module is tested simulating organic matter and nutrient concentrations, given by the parameter biochemical oxygen demand (BOD), and organic nitrogen (N-org), respectively. The

following transformation processes are represented in the mass balance: deoxygenation, BOD removal through sedimentation, organic nitrogen sedimentation, and conversion of organic nitrogen to ammonia.

3 Study case and dataset

The study is conducted for 83.5 km of the Iguaçu river, located at Paraná state, south of Brazil. Flow and water quality data are from two databases in five control sections, IG2 to IG6 - stations 65009000, 65017006, 65019980, 65025000 and 65028000, respectively (IAT, 2019; Fernandes, 2019). A summary of the dataset is presented in the supplementary material, while the study area is shown in figure 2.

While deterministic simulations are performed for the year 2010, an important assumption is that the historical observed dataset represents the expected range of concentrations in this period of interest. Furthermore, one of the main premises in model A is that such information is representative of statistical characteristics (mean and standard deviation); model B assumes that the relationship concentration-flow is well established by the observations in each control river section and that no significant trends, cycles and/or shifts are present.

Other data required for deterministic simulations include cross section geometry, rating curves and flow as daily records for the period of interest, also available at IAT (2019).

Fig. 2 Upper Iguaçu watershed and monitoring points

4 Results and discussion

Two cases were selected for comparison in this study: (i) test TD (Daily Test): daily data generated with $\rho = 0.5$ and seasonal μ and σ of the historical dataset; (ii) test TH (Hourly Test): hourly data with the same μ and σ of the full monitoring dataset and $\rho = 0.9$. As hourly data have higher persistence than daily values, ρ for TH is higher than TD. The specified parameters were chosen after extensive testing, which was discussed in Ferreira et al. (2019). The parameters β_0 , β_1 and β_2 of the model B were estimated by the Ordinary Least Squares method, as presented in the supplementary material.

The two stochastic models were applied to the first river section (IG2), and then propagated by the SihQual model downstream of the Iguau river. For comparison, models A and B were also applied for the other control points (IG3 to IG6). Therefore, nine datasets were compared in the river sections IG3 to IG6: monitoring data, results from the deterministic model propagating four stochastic series – two of model A and two of model B (differently than model A, in which one series among the multiple choices is selected, two random solutions generated by model B were chosen) –, and two synthetic series of each model A and B. This setting is summarized in table 1.

Hydrodynamic simulations and other intermediary results and not focus on this study – details can be found in (Ferreira et al., 2016, 2019). Additionally, the criterion to limit the generation of extreme values by models A and B did not alter the time series, since these rejected data represent 0 to 10% of total values in each synthetic times series.

Table 1 Summary of compared datasets

Series number	Dataset
1	Monitored data
2	First set of random daily series from model B
3	Deterministic simulation result with series 2 as UBC ⁽¹⁾
4	Second set of random daily series from model B
5	Deterministic simulation result with series 4 as UBC
6	Daily series from model A (series TD)
7	Deterministic simulation result with series 6 as UBC
8	Daily series from model A (series TH)
9	Deterministic simulation result with series 8 as UBC

⁽¹⁾ Upstream boundary condition

4.1 Deterministic water quality modelling integrated to stochastic models A and B

2

3 Results are compared as boxplots and duration curves in figures 3 and 4. The series of models A
4 and B are daily values, while results from the SihQual model are every 50 s (which is the time step for
5 numerical solution). The goal of this comparison is to verify the concentration range represented by each
6 approach for the simulated year.

7 Overall, models A and B showed to be adequate to represent the upstream boundary condition
8 for the deterministic model. The first one generated similar variability (series 6 and 8) for biochemical
9 oxygen demand; for organic nitrogen, option TH (hourly data generation) produced boxplots with higher
10 interquartile than expected, although still close to data.

11 Deterministic solutions are characterized by a high persistence in the series, which explains the
12 proximity of outliers and duration curves (series 3, 5, 7 and 9), when compared to the series from the
13 stochastic model (series 2, 4, 6 and 8). The high persistence in deterministic data is a consequence of the
14 explicit numerical solution, in which a value is calculated with previous time steps.

15 Most of the differences among duration curves are for higher concentrations, with less than 20%
16 occurrence; the other discrepancies tend to be small when compared to typical uncertainties associated
17 with BOD and organic nitrogen data (Coelho, 2019). The differences in the higher percentiles probably
18 happen because it is the region of extreme events. As hydrological models have difficulties in representing
19 flow peaks (Onyutha, 2019), water quality concentrations that happen at small frequencies have larger
20 variances – therefore, greater uncertainty.

Fig. 3 Boxplots and duration curves of compared BOD series

Fig. 4 Boxplots and duration curves of compared N-org series

4.2 Significance in water resources management

Although yearly concentration variability was well reproduced by all strategies, an alternative analysis showed that data frequency is a key aspect when dealing with tools to convert discrete into continuous information.

Figure 5 compares the number of days in which BOD limit concentrations were exceeded during the simulated year. Since results from the deterministic model are available every fifty seconds (i.e., 1728 values per day), three approaches were considered to allow daily comparisons: first, second and third quartiles of each day in the simulated year.

The limits for comparison are based on the legal regulation CONAMA nº 357/2005 in Brazil (CONAMA, 2005), which establishes water quality classes 1, 2, 3 and 4 for BOD concentration with a maximum threshold of 3, 5, 10 and larger than 10 mg-O₂/L, respectively. For organic nitrogen, the limits were defined as 3.7 mg-N/L for Classes 1 and 2, and 13.3 mg-N/L for classes 3 and 4.

BOD

Fig. 5 Number of days in the simulated year – N – for which the river stations are classified as Class 1, 2, 3 or 4 according to BOD and N-org limits; rows differ by the measure to take daily values from deterministic simulations: quartiles 1, 2 and 3, respectively

Results reveal that the approaches differ in the indication of water quality impairment during the year. For most of the comparisons, deterministic simulations tend to indicate the need for more restrictive actions in the watershed than stochastic approaches, since their results show more days in which the river is classified as class 3 or 4.

This behavior may be explained with the assistance of violin plots, that were build to illustrate the distribution of each complete series (figure 6). This type of visualization is similar to boxplots, but it combines the box shape with a density trace (kernel density plot, understood as a smoothed histogram), revealing structure found within the data (Hintze and Nelson, 1998).

Although the central tendency of both datasets is similar (vertical displacement is due to calibration aspects), differences in data distribution are expected, since daily time series (stochastic models) are being compared to sub-daily values (50s from the deterministic model). Nonetheless, the comparison of violin shapes shows that data distribution is compatible with the patterns of exceeding daily concentration limits observed in figures 5.

Most differences between stochastic and fully deterministic solutions for BOD are observed for section IG3 and IG4 (figures 3, 4 and 5), that receive higher pollution loads than IG5 and IG6. This observation suggests how relevant is a future representation of sub-daily effects in sections where the temporal dynamic is relevant.

Fig. 6 Violin plots for BOD series (left column) and N-org series (right column); continuous line indicates legal concentration limits

5 Conclusions

This study aims for more reliable representations of pollutants transport over time in rivers. Because field samplings usually do not match the required model input, two approaches to generate continuous data from discrete sampling were compared – models A and B. The verification of generated data was achieved through integration with a physical-based model and analysis of their meaning in real-world applications.

Conversion of temporal scales of water quality data is a challenging research topic, since it requires taking into account several processes of low occurrence frequency – which are usually missed by discrete field sampling, such as daily and weekly patterns, or rain-driven events. This study presented a comparison of two models to convert sparse historical monitoring data into daily time series for one year, combining the observed data information with deterministic and stochastic representations.

Results indicate that there is no significant difference in generating time series from a concentration or a concentration-flow dataset. Therefore, the combination of deterministic and stochastic parts of models A and B are equivalents and suitable to reproduce the overall variability of historical monitoring datasets in different river sections. Synthetic series showed a consistent behavior in reproducing mass transport, since, as input in the deterministic model, they produced reasonable simulations.

Nonetheless, data representativeness is a key aspect to guarantee reliable results, since all the statistical information should be from a representative period of the complete time series. In model A, preservation of statistical metrics and a consistent description of natural persistence in the required time scale are essential – however, it is difficult to determine this latter characteristic due to irregular and low frequency data; model B relies mainly on a well established relationship between discharges and mass distributions – which can be highly variable and challenging because of a scarcity of concentration measurements during high flow events.

Both hybrid approaches are suitable to reproduce historical water quality variability, which is a valuable analysis in water resources planning and managing. In addition, these methods yield quick multiple scenarios estimation: changes in metrics that characterize the system, in case A - average, standard deviation, quartiles and concentrations of 10 and 90% of frequency - or in flow conditions, in model B. In this context, the full deterministic model is suggested for integrated analysis in the watershed and cause-effect studies.

The analysis of overall patterns indicated that, to properly convert information into different time scales and support decisions in water resources management, data distribution should be observed. It is necessary to understand and represent cyclical effects (such as daily, weekly and seasonal patterns), and statistical characteristics of water quality series, such as trends – this is imperative to assess validity periods of water quality synthetic series, for example.

In this context, all approaches here presented have the potential to represent these aspects in future efforts. Furthermore, statistical validations and uncertainty analysis are also recommended to increase model reliance.

Declarations

Ethical Approval: Not applicable.

Consent to Participate: Not applicable.

Consent to Publish: Not applicable.

Authors contributions: D.M.F, M.C., C.V.S.F., E.K. and D.H.M.D. contributed to the study conception and design. Material preparation, data collection and analysis were performed by D.M.F and M.C. The first draft of the manuscript was written by D.M.F; M.C., C.V.S.F., E.K. and D.H.M.D. commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding: No funding was received to assist with the preparation of this manuscript.

Conflicts of interest/Competing Interests: The authors have no conflicts of interest to declare.

Availability of data and material: All data and material used in this study are available on request and at <http://www.iat.pr.gov.br/Pagina/Sistema-de-Informacoes-Hidrologicas>.

References

- Babu CN, Sure P (2016) Partitioning and interpolation based hybrid arima-ann model for time series forecasting. *Sādhanā* 41(7):695–706
- Bowes M, Loewenthal M, Read D, Hutchins M, Prudhomme C, Armstrong L, Harman S, Wickham H, Gozzard E, Carvalho L (2016) Identifying multiple stressor controls on phytoplankton dynamics in the river thames (uk) using high-frequency water quality data. *Science of the Total Environment* 569:1489–1499

- Coelho M (2019) Uncertainty analysis in the statistical and stochastic context of water quality time series. PhD thesis, Federal University of Parana
- CONAMA (2005) (Conselho Nacional do Meio Ambiente). Resolução n 357/05. URL <http://www2.mma.gov.br/port/conama/legiabre.cfm?codlegi=459>
- Fernandes CVS (2019) INTEGRA 2: Bases Técnicas para a Integração de Instrumentos de Gestão de Recursos Hídricos - Estudo de Caso da Bacia do Alto Iguaçu e Bacia do Alto Tietê (INTEGRA 2: Technical Bases for the Integration of Water Resources Management System Instruments - Case Study of the Upper Iguaçu Basin and Alto Tietê.). Tech. rep., University of Zurich, Department of Informatics
- Ferreira DM, Fernandes CVS, Kaviski E (2016) Curvas de permanência de qualidade da água como subsídio para o enquadramento de corpos d'água a partir de modelagem matemática em regime não permanente. RBRH 21(3):479–492
- Ferreira DM, Fernandes CVS, Kaviski E, Fontane D (2019) Water quality modelling under unsteady state analysis: Strategies for planning and management. *Journal of Environmental Management* 239:150–158, DOI 10.1016/j.jenvman.2019.03.047
- Ferreira DM, Fernandes CVS, Kaviski E, Fontane D (2020) Transformation rates of pollutants in rivers for water quality modelling under unsteady state: A calibration method. *Journal of Hydrology* p 124769
- Gholizadeh MH, Melesse AM, Reddi L (2016) A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors* 16(8):1298
- Gnauck A (2004) Interpolation and approximation of water quality time series and process identification. *Analytical and bioanalytical chemistry* 380(3):484–492
- Guzman JA, Shirmohammadi A, Sadeghi AM, Wang X, Chu ML, Jha MK, Parajuli PB, Harmel RD, Khare YP, Hernandez JE (2015) Uncertainty considerations in calibration and validation of hydrologic and water quality models. *Transactions of the ASABE* 58(6):1745–1762
- Helsel DR, Hirsch RM, Ryberg KR, Archfield SA, Gilroy EJ (2019) Statistical methods in water resources. Tech. rep., US Geological Survey
- Hintze JL, Nelson RD (1998) Violin plots: a box plot-density trace synergism. *The American Statistician* 52(2):181–184
- Huang H, Wang Z, Xia F, Shang X, Liu Y, Zhang M, Dahlgren RA, Mei K (2017) Water quality trend and change-point analyses using integration of locally weighted polynomial regression and segmented regression. *Environmental Science and Pollution Research* 24(18):15827–15837

- 1 IAT (2019) Instituto Água e Terra (Water and Earth Institute). URL <http://www.iat.pr.gov.br/Pagina/Sistema-de-Informacoes-Hidrologicas>
- 2 Loucks DP, Beek EV (2017) Water Resource Systems Planning and Management. Springer International Publishing, DOI 10.1007/978-3-319-44234-1
- 3 Marrin D (2017) Pattern-based approaches to evaluating water quality. In: Multidisciplinary Digital Publishing Institute Proceedings, vol 2
- 4 Miller MP, Tesoriero AJ, Capel PD, Pellerin BA, Hyer KE, Burns DA (2016) Quantifying watershed-scale groundwater loading and in-stream fate of nitrate using high-frequency water quality data. Water Resources Research 52(1):330–347
- 5 Onyutha C (2019) Hydrological model supported by a step-wise calibration against sub-flows and validation of extreme flow events. Water 11(2):244
- 6 Strokal M, Spanier JE, Kroeze C, Koelmans AA, Flörke M, Franssen W, Hofstra N, Langan S, Tang T, van Vliet MT, et al. (2019) Global multi-pollutant modelling of water quality: scientific challenges and future directions. Current opinion in environmental sustainability 36:116–125
- 7 Taherdoost H (2018) A review of technology acceptance and adoption models and theories. Procedia manufacturing 22:960–967
- 8 Tiyyasha, Tung TM, Yaseen ZM (2020) A survey on river water quality modelling using artificial intelligence models: 2000–2020. Journal of Hydrology 585:124670
- 9 Uusitalo L, Lehtikoinen A, Helle I, Myrberg K (2015) An overview of methods to evaluate uncertainty of deterministic models in decision support. Environmental Modelling & Software 63:24–31
- 10 Yao J, Wang G, Xue W, Yao Z, Xue B (2019) Assessing the adaptability of water resources system in shandong province, china, using a novel comprehensive co-evolution model. Water Resources Management 33(2):657–675
- 11 Yaseen ZM, Sulaiman SO, Deo RC, Chau KW (2019) An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. Journal of Hydrology 569:387–408

Figures

Figure 1

Schematic of modelling integration; the dashed line highlights the main features of interest in this study

Figure 2

Upper Iguazu watershed and monitoring points Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 3

Boxplots and duration curves of compared BOD series

Figure 4

Boxplots and duration curves of compared N-org series

BOD

Figure 5

Number of days in the simulated year – N – for which the river stations are classified as Class 1, 2, 3 or 4 accordingly to BOD and N-org limits; rows differ by the measure to take daily values from deterministic simulations: quartiles 1, 2 and 3, respectively

Figure 6

Violin plots for BOD series (left column) and N-org series (right column); continuous line indicates legal concentration limits

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Lettereditor.docx](#)
- [Letterreviewers.docx](#)
- [Supplementarymaterial.docx](#)
- [graphicalabstract.docx](#)