Solving River Pollution Problems by Means of Fuzzy Fault Tree Analysis

FRAGISKOS BATZIAS, ATHANASIA BOUNTRI, CHRISTINA SIONTOROU
Department of Industrial Management & Technology
University of Piraeus
Karaoli & Dimitriou 80, 18534 Piraeus
GREECE
fbatzi@unipi.gr

Abstract: In this work, an algorithmic procedure, based on Fault Tree Analysis (FTA) in its fuzzy version (to count for uncertainty), has been developed for solving river pollution problems. The main steps followed are: (a) determination of metrological requirements, necessary to maintain the present ecological status of the river, (b) combinatorial relaxation to obtain acceptable tolerance intervals for each ecological sub-system, (c) sensitivity analysis to quantify the impact of the parameter(s) values deviation from the normally expected values, (d) synthesis of the corresponding fault tree and assignment of weights, (e) FTA by using fuzzy input, (f) formation of representative alternatives by combining the most influential final events acting as ultimate causes that contribute to the occurrence of the ‘top event’, and fuzzy multicriteria ranking of them, and (g) sensitivity/robustness analysis, of the ranked first alternative and suggestions for corrective action. A case example is presented, where the top event is “high BOD at site G”, downstream of site F where wastewater is discharged into the river. The input independent/explanatory variables are given as fuzzy trapezoid numbers and the results (obtained as crisp numbers, after defuzzification) are discussed.

Key-Words: River Pollution, Fuzzy Fault Tree Analysis, Multicriteria Ranking, Environmental Modeling, Polluter Pays Principle, Biochemical Oxygen Demand

1 Introductory Analysis

Polluted rivers may cause several problems, especially to a downstream (in relation to the point/region of waste discharge) user of water. We can solve such problems by determining/identifying the ultimate cause and either eliminating it or diminishing its impact. To facilitate this determination, we may adopt a model of river behavior and try to identify the parameter mostly responsible for the problem. Subsequently, it is easy to suggest remedial/preventive action and test the effectiveness [1-3]. Nevertheless, these models may not be adequately effective, since they do not take into consideration the interaction of variables/parameters, especially in the case of river ecosystems that appear to be rather sensitive to such interactions. To analyze this situation, we may refer to the most usual river model [4], according to which the following one-dimensional expression is representative of material transport through a cross-sectional area perpendicular to the direction of stream flow:

$$\frac{\partial c}{\partial t} = \frac{1}{A(x,t)} \frac{\partial}{\partial x} \left[ E(x,t) A(x,t) \frac{\partial c}{\partial x} \right] - \frac{1}{A(x,t)} \frac{\partial}{\partial x} \left[ Q(x,t) c \right] - S(c,x,t)$$

For constant $A$ and volumetric flow rate $Q$, the transport of dissolved oxygen concentration $C$ is given by:

$$\frac{\partial C}{\partial t} = - \frac{Q}{A} \frac{\partial C}{\partial x} + k_2 (C^* - C) - k_1 L(x,t) - S_R(x,t)$$

where $c$ is the concentration of transported material, $A$ is the cross-sectional area, $E$ is the dispersion coefficient, $Q$ is the rate of volume transport of fluid, $t$ is time, $x$ is the distance, and $S$ represents the sources and sinks of transported material.

Neglecting longitudinal dispersion and assuming velocity as the only significant component of such flux, we obtain:

$$\frac{\partial c}{\partial t} = - \frac{1}{A(x,t)} \frac{\partial}{\partial x} \left[ Q(x,t) c \right] - S(c,x,t)$$

where $c$ is the concentration of transported material, $A$ is the cross-sectional area, $E$ is the dispersion coefficient, $Q$ is the rate of volume transport of fluid, $t$ is time, $x$ is the distance, and $S$ represents the sources and sinks of transported material.
\[
\frac{\partial L}{\partial t} = -\frac{Q}{A} \frac{\partial L}{\partial x} - k_1 L(x, t) - k_3 L(x, t) + L_a(x, t) \quad (4)
\]

where \( k_3 \) is the rate constant for BOD removal through sedimentation and/or adsorption, and \( L_a \) is the rate of addition of BOD by local runoff or by resuspension of organics from bottom deposits.

Assuming that \( Q/A = U \), \( k_1 \), \( k_3 \), and \( L_a \) remain unchanged along the river segments under examination, we obtain the following expression for steady state (i.e., \( \frac{\partial L}{\partial t} = 0 \)):

\[
L(x) = L_o F_1 + \frac{L_a}{k_1 + k_3} (1 - F_1) \quad (5)
\]

\( F_1 = \exp \left[ -\left( k_1 + k_3 \right) \frac{x}{U} \right] \) and \( L_o \) is BOD at \( x = 0 \).

Eq. (5) signifies a BOD profile along the river, which is independent of oxygen deficit \( D = (C^* - C) \), since the rate constant \( k_2 \) is not taken into account. This is not in accordance with what is observed/measured in real situations, since activated biomass (which is the biological agent for the BOD removal) concentration (i.e., the population of microorganisms per unit volume) depends on oxygen concentration, especially when approaching deficit points/regions or oxygen sinks. To cope with this problem, we adopted a methodological framework based on Fault Tree Analysis (FTA) in its fuzzy version to count for uncertainty. The main characteristic of the pollution problem is set as ‘top event’ and the remedial proposals are the results of combining FTA with multicriteria ranking of alternative solutions.

2 Methodology

The methodological framework, designed/developed under the form of an algorithmic procedure for solving river pollution problems by means of Fuzzy Fault Tree Analysis (FFTA), includes 22 activity stages and 6 decision nodes, as described below (for their interconnection, see Fig. 1).

1. Selection of data referring to the river usage and its ecosystems.
2. Determination of metrological requirements, necessary to maintain the present ecological status of the river.
3. Experimental design to obtain additional data at the required information granularity level at minimum cost.
4. Data/information acquisition by performing the designed measurements.
5. Combinatorial relaxation to obtain acceptable tolerance intervals for each ecological subsystem of the river.
6. Application of these models to data and estimation of the corresponding statistical parameters.
7. Selection of the best model, according to goodness of fitting criteria.
8. Determination/identification of the parameter(s)

![Figure 1. The methodological framework designed/developed by the authors for solving river pollution problems by means of Fuzzy Fault Tree Analysis.](image-url)
mainly responsible for the pollution problem.

9. Sensitivity analysis to quantify the impact of the parameter(s) values deviation from the normally expected value.

10. Synthesis of remedial proposal(s) accompanied by suggestions for successful implementations.

11. Selection of variables/parameters and their relations for ontological mapping of the river ecological sub-systems.

12. Discrimination of sub-structures performing taxonomy/partonomy operations and thinning within the obtained ontology.

13. Set up of the causal chains, putting emphasis upon their inter-relation to form the respective specific network related to the pollution problem under investigation.

14. Synthesis of the corresponding fault tree by following both procedures, top-down by deduction and bottom-up by induction.

15. Assignment of weights or indices to indicate the strength of causal relation between successive events connected through the inclusive OR gate.

16. FTA by using fuzzy input (after partitioning variables/parameters by means of linguistic terms) to count for uncertainty.

17. Formation of representative alternatives, made by combining the most influential final events acting as ultimate causes that contribute (through the chains developed/identified at stage 13) to the occurrence of the 'top event'.

18. Selection of criteria.

19. Fuzzy multicriteria ranking of the alternatives formed at stage 17.

20. Testing of the alternative ranked (among the alternatives not examined so far) first to identify its influence on the 'top event'.


22. Searching in external KBs for data mining by means of an Intelligent Agent, according to [5].

A. Do the existed and acquired data satisfy these requirements?

B. Are there technological equipment, skilled personnel, and economic resources available to carry out the required additional measurements?

C. Is further relaxation of the metrological requirements (i.e., width increase of the corresponding tolerance intervals) possible/ permitted?

D. Are there mechanismic or semi-empirical (derived from relevant theory or dimensional analysis, where the resulting dimensionless groups have a kind of physical meaning) models simulating the river’s behaviour?

E. Is it a successful confirmation test?

F. Is there at least one alternative left unexamined?

Sensitivity and robustness analysis are used for the testing described in stage 20 (as regards the impact of both, criteria weights and preference matrix fuzzy elements).

3 Implementation

The case of measuring high BOD (in relation to the corresponding expected value) at stream site G, when a user is discharging at the upstream site F, has been analyzed. The intermediate/final events of the corresponding fault tree, where the top event (with code number 1) is the mentioned BOD deviation, are described subsequently while their interconnection is depicted in Fig. 2.

1. Deviation of oxygen consumption/transfer rate constants, resulting to high BOD of stream at G.

1.1.1 High deoxygenation rate constant, \( k_1 \).

1.1.2 Low re-aeration rate constant, \( k_2 \).

1.1.3 Low removal (through sedimentation and/or adsorption) rate constant, \( k_3 \).

1.2 Underestimation (in the system’s analysis phase) of exogeneously determined critical parameter values.

1.2.1 High local surface runoff, mainly due to agricultural activities.

1.2.2 High resuspension of organics from bottom sludge deposits.

1.2.3 High decomposition of organics from bottom sludge deposits to organic acids, subsequently released in water.

1.2.4 High eutrification in water body and on stream banks.

1.3 Underestimation (in the system’s analysis phase) of BOD at G.

1.3.1 Wrong simulation model of oxygen-sag profile function.

1.3.1.1 Small number of explanatory independent variables and parameters.

1.3.1.2 Lack of information about the role/quantification of certain parameters.

1.3.1.3 Fusion of variables/parameters.

1.3.2 Big number of explanatory independent variables and parameters for sake of completeness although some of them are of ambiguous significance and some others are supported by inadequate data/information.

1.3.3 Mostly theory-based model structure (actually a mechanismic model), implying some kind of weakness in functioning under...
real conditions, especially within multi-affected environments.

1.3.1.4 Mostly practice-based model structure (actually a ‘black box' model), implying some kind of weakness in functioning under real conditions, especially when certain parameter values of the working environment are far from the values used for validating the simulation model.

1.3.2 Problematic data set.
1.3.2.1 Inadequate data for fitting a spatial distribution along the stream within a critical time interval (cross section analysis).
1.3.2.2 Inadequate data for time series analysis.
1.3.2.3 Decreased repeatability of measurements.
1.3.3 Decreased numerical approximation in parameter values estimation.
1.3.3.1 Linearization of non-linear model for sake of simplicity in data processing.
1.3.3.2 Convergence of the error/objective function to local minimum.

1.3.3.3 Stepwise/partial instead of total/synchronous optimization (e.g., when attempting to obtain high information granularity for parameter identification purposes).

1.4 Underestimation (in the system’s analysis phase) of BOD $L_F$ at F.

1.4.1 Overestimation of the river flow rate $Q_r$ at F, before mixing.

1.4.2 Underestimation of the river BOD $L_r$ at F, before mixing.

1.4.3 Underestimation of the flow rate $Q_u$ of the upstream user, since $Q_u << Q_r$, normally.

1.4.4 Underestimation of the BOD $L_u$ of the upstream user, since $L_u Q_u + L_r Q_r = L_F (Q_u + Q_r)$.

1.4.4.1 Underestimation of the inflow wastewater BOD $L_i$ in the user’s treatment facility.

1.4.4.2 High deviation of processing parameter values during secondary treatment in comparison with the corresponding values.
1.4.4.2.1 Inhomogeneity of the wastewater suspension in the CFSTR.
1.4.4.2.1.1 Insufficient agitation.
1.4.4.2.1.2 Appearance of coagulated activated sludge in the recycle flow from the secondary thickener.
1.4.4.2.2 Overestimation of the value of the BOD removal rate constant.
1.4.4.2.3 Underestimation of the BOD asymptotic value in the kinetic equation.
1.4.4.2.4 Overestimation of the order (i.e., the exponent, even if it is not an integer) of the kinetic equation.

A sample of fuzzy inference in the case under consideration is shown in Fig. 3, where the values of input-output variables are represented by trapezoid fuzzy numbers, i.e., the most likely to occur value corresponds to a region rather than to a point (in which case the fuzzy number is represented by a triangle). The input values (rounded to the nearest integer) are (19, 26, 31, 40) for the \( k_1 \)-input, (15, 23, 29, 34) for the \( k_2 \)-input, (34, 41, 47, 52) for the \( k_3 \)-input, and (28, 39, 50, 61), with membership function 0.5499 for the \( M \) linguistic term, and 55, 67, 100 with membership function 1.0000 for the \( H \) linguistic term) for the \( L_G \)-output; the corresponding crisp values are \( k_1 = 29.10 \), \( k_2 = 25.12 \), \( k_3 = 43.42 \), \( L_G = 62.3 \), found after defuzzification according to the method of centroid, which gives the crisp number \( x^* \) (abscissa of gravity centre) as a function of the four typical points \( a, b, c, d \) of the trapezoid as follows:

\[
x^* = \frac{c^2 + d^2 + cd - a^2 - b^2 - ab}{3(c + d - a - b)}
\]

where \((b-a)\) and \((d-c)\) are the Left and Right margins of the fuzzy number (in the usual \( L, R \)-form), respectively. In terms of the code used in the fault tree, the corresponding numbers (representing final/intermediate events) are 1.1.1, 1.1.2, 1.1.3 for input and 1.1 for output.

All input-output variables are defined in the normalized 0 – 100 domain and its partition with the \( \text{Low, Medium, High} \) (\( L, M, H \), respectively) linguistic terms, in order to apply the IF… THEN… fuzzy rules, has been performed by experts as shown in [6,7].

**Figure 3.** Partition of the domain of each input/output variable, represented as trapezoid fuzzy number, by means of the L, M, H linguistic terms; the shadowed trapezia represent the input fuzzy numbers \( k_1, k_2, k_3 \), and the output fuzzy number \( L_G \) (events 1.1.1, 1.1.2, 1.1.3, 1.1, respectively, in the fault tree), corresponding to the case example considered herein.
4 Discussion & Concluding Remarks

In case of a positive output from decision node D, as a result of avoiding approach to critical oxygen deficit points/regions or sinks, Eq. (3) can be written as follows:

\[
\frac{\partial D}{\partial t} = -\frac{Q}{A} \frac{\partial D}{\partial x} - k_2 D + k_1 L(x,t) + S_R(x,t) \tag{7}
\]

which at steady state (i.e., \(\partial D / \partial t = 0\)) becomes:

\[
U dD/dx = -k_2 D + k_1 L(x) + S_R(x) \tag{8}
\]

By combining Eq. (8) with Eq. (5), we obtain:

\[
dD = -D \frac{k_2}{U} + \left[\frac{L_0 F_1}{k_1 + k_3} (1-F_1) \frac{k_1}{U} + \frac{S_R}{U}\right] \tag{9}
\]

By further assuming that \(k_2\) and \(S_R\) remain unchanged along the river segment under examination, the last equation can be integrated and solved for the BOD \(L_u\) of the upstream user effluent to give:

\[
L_u = \left[\frac{D - G - D_0 F_2}{(k_1/(k_2 - (k_1 + k_3)))(1-F_2)} + \frac{L_u}{k_1 + k_3}\right] \tag{10}
\]

\[
\frac{Q_0 + Q_u}{Q_u} = \frac{L_u L_r}{Q_u} \tag{11}
\]

where \(G = S_R/k_2 + k_1 L_u/k_2 (k_1 + k_3)(1-F_2)\), \(F_2 = \exp(-k_3 x/U)\), and \(D_0 = D\) at \(x=0\), while \(L_r\) is the river BOD at \(F\) before mixing and \(Q_0, Q_u\) are the corresponding volumetric flow rates of the river and the user at the same point. Comparing this predicted \(L_u\)-value with the corresponding measurements obtained in ‘real time’, we can conclude about ‘underestimation’ as a fault/event possibly causing the top event (see node 1.4.4).

Other models can be similarly synthesized and the best can be chosen according to (i) goodness of fitting criteria and (ii) sensitivity analysis, also quoted in the action nodes/stages 6-10 shown in Fig. 1. Both approaches, FFTA and deterministic/mechanistic, can be stored in the KB of stage 21 to be used subsequently for Case Based Reasoning and Model Based Reasoning (CBR and MBR, respectively) to contribute in solving similar problems.

In conclusion, the methodological framework, we have designed/developed under the form of an algorithmic procedure for solving river pollution problems by means of Fuzzy Fault Tree Analysis (FFTA) is proved to be successful, at least in cases where (i) we may assume that the values of all critical variables/parameters remain unchanged or change within a very limited region, permitting (a) consideration of the validity of steady state conditions and (b) integration of the differential equations that describe the river’s behavior when moderate polluting activities take place, and (ii) in absence of analytic expressions serving as models, we have (or can obtain) adequate information to formulate empirical relations (possibly by means of dimensional analysis) connecting successive events within a causal path leading to the top event.

5 Acknowledgements

Financial support provided by the Piraeus University Research Center is kindly acknowledged.

References:


