2. A FUZZY IMPLEMENTATION

The collection of rules extracted from the previous flow diagrams constitutes the knowledge base of the risk model. The uncertainty associated with the description of heuristic knowledge is tackled by using fuzzy logic in the development of the heuristic IF-THEN rules. Membership functions were developed for each variable considered in the flow diagrams while a fuzzy rule base or decision matrix must be defined for each operational problem.

Figure 5 illustrates how this new knowledge-based model works to be able to obtain the risk of microbiologically related settling problems. Some data from the simulation output is fuzzified, then the fuzzy rule-based system launches those rules whose antecedents (IF part) are satisfied and finally the defuzzifcation takes place to provide the user with a numerical value (between 0 and 1, or 0 and 100%) for the risk of each one of the operational problems considered.

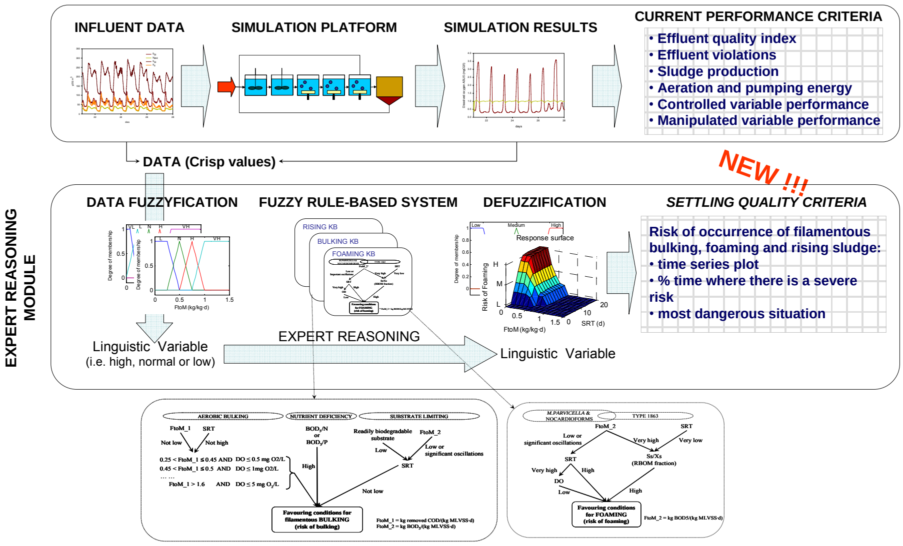

Figure 5. Scheme of the operations performed by the risk model.

Next the fuzzification, rule base definition and defuzzification steps are detailed using standard modelling descriptions and equations.

2.1 Fuzzification

For each input and output variable selected, we define two or more membership functions (MF), normally three but can be more. We have to define a qualitative category for each one of them, for example: low, normal or high. The shape of these functions can be diverse but we will usually work with triangles and trapezoids (actually usually pseudo-trapezoids) (see Figure 6). For this reason we need at least three (for triangles) or four (for trapezoids) points to define one MF of one variable.

Example 1: If we take **x** like a variable and low, normal and high as trapeizodal, triangle and trapezoidal MFs, respectively (Figure 6),

- the MF **low** will be defined by three points: (x_1, x_2, x_3). However, in order to have a real trapezoid, we need a four point at the left of x_1 (**any** negative one, e.g x_0)

- following the same reasoning, the MF **high** will be defined by four points: (x_3, x_4, x_5, x_6) (x_6 **any** positive > x_5, being x_5 the higher possible value for x)

- finally, the MF **normal** (like any other triangular MF) will be defined by three points: (x_2, x_3, x_4).

In case the MF are trapezoids (or pseudo-trapezoids) (in this case ‘low’ and ‘high’), the MF can be defined as:

$$y^{low}{}_{(trap)}(x; x_0, x_1, x_2, x_3) = \max\left(\min\left(\frac{x - x_0}{x_1 - x_0}, 1, \frac{x_3 - x}{x_3 - x_2}\right), 0\right)$$

$$y^{high}{}_{(trap)}(x; x_3, x_4, x_5, x_6) = \max\left(\min\left(\frac{x - x_3}{x_4 - x_3}, 1, \frac{x_6 - x}{x_6 - x_5}\right), 0\right)$$

In case the MF are triangles (in this case 'normal'), the MF can be defined as:

$$y^{normal}{}_{(tri)}(x; x_2, x_3, x_4) = \max\left(\min\left(\frac{x - x_2}{x_3 - x_2}, \frac{x_4 - x}{x_4 - x_3}\right), 0\right)$$

It is important to emphasize that the computation of all the functions/equations for all the MFs of all variables has to be done every time the shape and interval of the MFs are changed (on the contrary, once computed for the first time, these computations do not have to be done if MFs are not changed).

How the fuzzification step works

Next question to be solved is how to fuzzificate all the real values of the variable **x**. First, for a given value of **x,** for example $\mathbf{x_n}$ which can belong to one or more MF we calculate the **y** value for each of the MF/s which $\mathbf{x_n}$ belong**.** This **y** value has to be between **0** and **1**. For example: Consider three MF: low, normal and high and a given value of $\mathbf{x_n}$, then the degrees of membership to each MF (y values) for $\mathbf{x_n}$ can be, for example: 0.6 for the MF low and 0.4 for the MF normal (see Figure 6). Likewise, we can fuzzificate all the values of any variable. Any of the values will belong to at least one MF with a certain degree of membership.

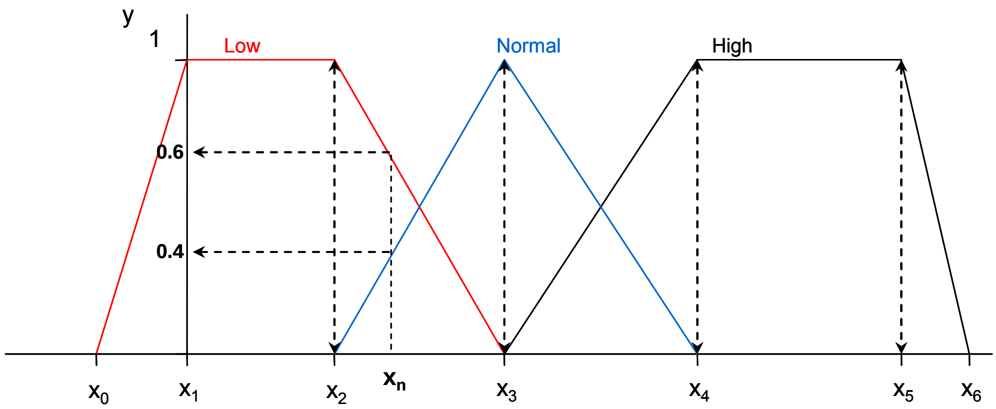

Figure 6 Example of the three MF for a given input.

2.2 Rule base (decision matrix) definition

Once the input and output variables and MF are defined, we have to design the rule-base (or decision matrix of the fuzzy knowledge-base) composed of *expert* IF <antecedents> THEN <conclusions> rules. These rules transform the input variables to an output that will tell us the risk of operational problems (this output variable, risk of a problem, also have to be defined with MF, usually low, normal and high risk). Depending on the number of MF for the input and output variables, we will be able to define more or less potential rules. The easier case is a rule-base concerning only one input and one output variable.

Example 2: For a given variable **x** involved in the development of a problem, we could have this kind of "theoretical" rule:

IF **x** is **normal** THEN risk of problem is **normal.**

The more variables we have, the more rules we have to define in order to make the inference reliable.

Example 3: Let's suppose that we have these variables, **x** and **y,** both having three MF, low, normal and high. Each variable can belong to a different MF. Depending on the expert knowledge, we can have several inputs and several outputs.

IF **x** is **normal** *and* **y** is **very high** THEN risk of problems is **high.**

Once we have defined the realistic rules according to the expert knowledge, these rules will become the knowledge base of each of the problems considered in the risk model. It is necessary to say that the whole knowledge does not necessarily have to be translated in rules, sometimes some of the rules can be redundant. Let's suppose the following *invented* decision matrix that contains the expert knowledge to detect the risk of a problem from the inputs X and Y:

Table 1. Representation of an imaginary example of a decision matrix.

		Input Y			
		LOW	NORMAL	HIGH	VERY HIGH
Input X	LOW	Low	High	High	High
	NORMAL	Low	Low	Medium	Medium
	HIGH	Low	Low	Low	Low
	VERY HIGH	Low	Low	Low	Low

How the rule base works

Next raising question is to compute the degree of membership to the MF (low, normal or high) of the output (the risk of the problem). As explained in the fuzzification part, once a variable is fuzzified it takes a value between 0 and 1 indicating degree of membership to a given MF of that specific variable. The degrees of membership of the input variables have to be combined to get the degree of membership of the output variable.

Example 4: For a given variable **x** involved in the development of a problem (the risk-output have its own MF, low, normal and high risk), we can for example have a rule-base "saying" that:

IF **x** is **low** THEN risk of problem is **low.**

IF **x** is **normal** THEN risk of problem is **normal.**

IF **x** is **high** THEN risk of problem is **high.**

According to these rules, if we suppose that the degree of membership for **x** is 0.6 to the MF low, then the risk of the problem will be 0.6 low, too.

In case we have more than one input variable (in fact, the normal case), the degree of membership for the output value will be the **minimum value of the degree of membership for the different inputs**.

Example 5: Looking at the Figure 7, for a set of 9 rules resulting from the decision matrix (see Table 1, above), let's say that **Input X** = 0.55 has a membership degree of 0.8 to the MF 'normal' (rules 4, 6 and 7), and a membership degree of 0.2 to the MF 'high' (rule 8). On the other hand, **Input Y** = 6.5 has a membership degree of 0.2 to the MF 'high' (rules 1 and 7) and a membership degree of 0.9 the MF 'normal' (rules 3 and 4). When a rule is totally satisfied (the antecedent is satisfied, those with (1) in Figure 7, rules: 4, 7 and 8), it will have an output with a certain membership degree to an output MF. These are the rules satisfied in this example:

IF **Input X** is **normal** (degree of 0.8) *and* **Input Y** is **normal** (degree of 0.9) THEN Risk of problem is **low** (degree of 0.8) **(Rule 4)**

IF **Input X** is **normal** (degree of 0.8) *and* **Input Y** is **high** (degree of 0.2) THEN Risk of problem are **medium** (degree of 0.2) **(Rule 7)**

IF **Input X** is **high** (degree of 0.2) THEN Risk of problem is **low** (degree of 0.2) **(Rule 8)**

The MF of the output will have a degree of membership equal to the lower among the inputs.

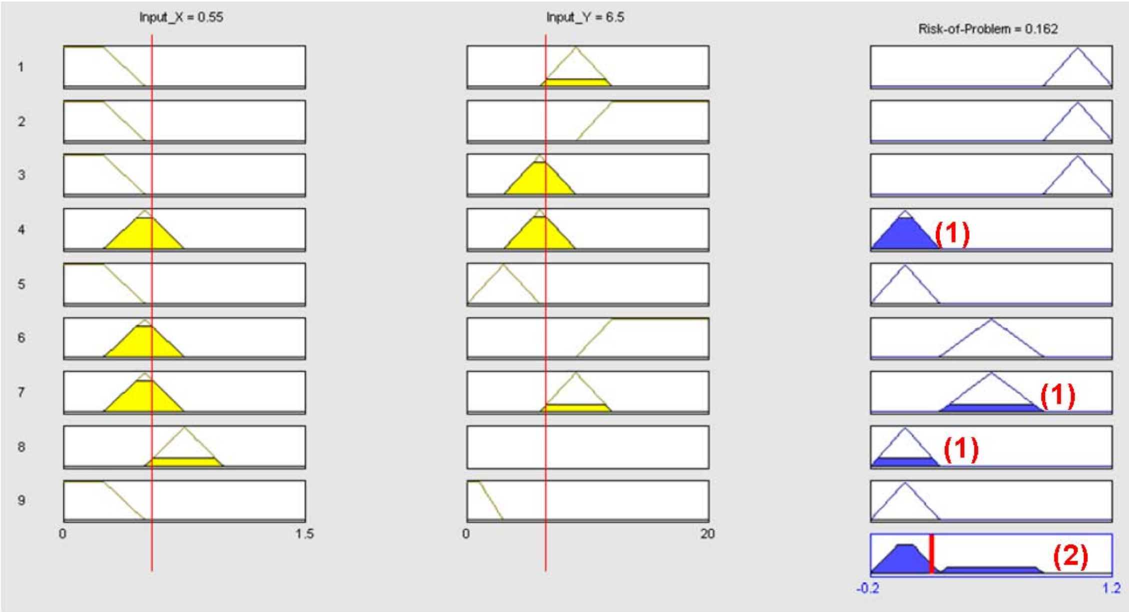

Figure 7 Example of the rules for the determination of a hypothetic risk of problem (detected by the rules of table 1).

From here, we will only look at those satisfied rules (4, 7 and 8). The resulting figure for the output (2) has the MF 'low' due to rules 4 and 8. Moreover, it has the MF 'normal' due to rule 7. To sum up, the final output figure (2) is the integration (sum) of the MF from the satisfied rules (1). Among the satisfied rules, the membership degree of each output MF will be the higher among the rules that have as a result that MF. It means that the degree of membership of the MF 'normal' (0.2) (in (2)) is due to rule 7 and that the degree of membership of the MF 'low' (0.8) (in (2)) is due to the higher among rules 4 and 8 (those that have as a conclusion the MF 'low').

2.3 Defuzzyfication

The MFs of the output have always the same shape and configuration in our risk model: the risk of any problem has the same ranks for the MFs of the output: low, normal and high, and always without overlapping. Figure 8 shows the shape of each MF of the output variable (risk on any problem considered in the risk model).

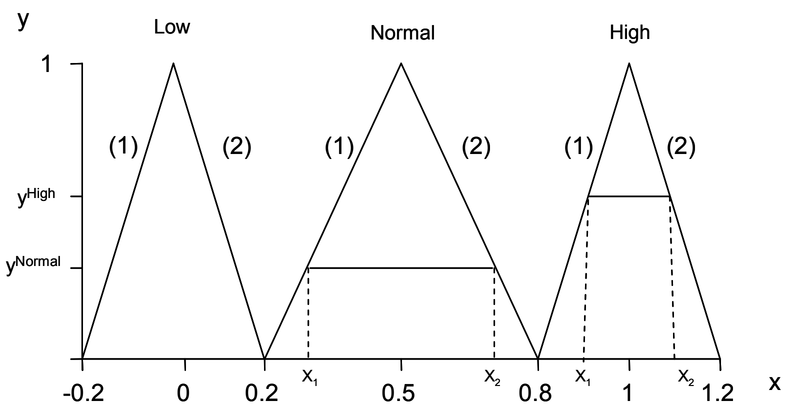

Figure 8. MFs of the risk of any problem within the risk model.

In order to obtain a percentage of risk of a problem, the output must be defuzzyfied. The equations of the straight lines of each MF of the output have to be calculated. The calculations for each of the MFs are presented next:

For the MF *'Low'*:

$$y^{low} = m_1 \cdot x_1 + n_1$$
$$y^{low} = m_2 \cdot x_2 + n_2$$
$$m_1 = (1-0)/(0.2-0) = 5$$
$$m_2 = (1-0)/(0-0.2) = -5$$

Now, to find the n, the point (0, 1) is substituted on both straight lines to obtain the two equations:

$$y^{Low} = 5 \cdot x_1 + 1$$
$$y^{low} = -5 \cdot x_2 + 1$$

A similar calculation is done for the 'Normal' and 'High' MFs, obtaining finally:

$$y^{Normak} = \frac{1}{0.3} \cdot x_1 - 0.665 \qquad y^{High} = 5 \cdot x_1 - 4$$
$$y^{Normal} = -\frac{1}{0.3} \cdot x_2 + 2.665 \qquad y^{High} = -5 \cdot x_2 + 6$$

As in the fuzzification step, it is important to emphasize that the computation of these equations for the MFs of the output has to be done every time the shape and interval are changed (on the contrary, once computed for the first time, these computations do not have to be done if MFs are not changed).

How the defuzzyfication step works

The next step involves calculating the area of the resulting figures for each MF, taking into account that these areas will not be triangle in all cases (most of the times they are triangles and trapezoids). If the degree of membership is not equal to one, the area will have a trapezoidal shape instead of a triangle.

For the 'low' and 'high' MF:

$$A_{Low/High} = \frac{\left[0.4 + (x_2 - x_1)\right] \cdot h_{Low/high}}{2}$$

(0.4 is the length of the big base of the trapezoid: (0.2 – (-0.2)) for the 'low' MF or (1.2 – 0.8) for the 'high' MF).

And for the 'normal' MF:

$$A_{Normal} = \frac{\left[0.6 + (x_2 - x_1)\right] \cdot h_{Normal}}{2}$$

(0.6 is the length of the big base of the trapezoid: (0.8– 0.2)).

In the equations above, **h**, which is the height of the trapezoid or the triangle of the MF of the output, is equal to **y** and so it will be equal to the degrees of membership of the MFs of the output.

The last step to defuzzyfy an output is to calculate the x of the centroid (x_c) of the integrated figure (integrating the areas calculated for each one of the MFs). This step will be done by the following equation:

$$x_c = \frac{A_{low} \cdot 0 + A_{normal} \cdot 0.5 + A_{high} \cdot 1}{A_{low} + A_{normal} + A_{high}}$$

Where represent the output of the variable, once defuzzyfied. Concerning the Risk Model, after multiplying by 100, the result will indicate the percentage of risk to experience a problem related to activated sludge suspended solids separation. In all cases, the MF or MFs of the output, as well as the degree of membership for each MF, will always be known.

Example 7: The MFs of a given output variable are *normal* and *high* with a degree of membership of 0.4 and 0.6, respectively. For each one we have to calculate x_1 and x_2, which are the limits of the small base of the trapezoid (see Figure 8).

'Normal':

$$\begin{cases} y_1^{Normal} = \frac{1}{0.3} \cdot x_1 - 0.665 \\ y_2^{Normal} = -\frac{1}{0.3} \cdot x_2 + 2.665 \end{cases}$$

$$\begin{cases} x_1^{Normal} = 0.38 \\ x_2^{Normal} = 0.62 \end{cases}$$

'High':

$$\begin{cases} y^{High} = 5 \cdot x_1 - 4 \\ y^{High} = -5 \cdot x_2 + 6 \end{cases}$$

$$\begin{cases} x_1^{High} = 0.92 \\ x_2^{High} = 1.08 \end{cases}$$

With this pair of values it is possible to obtain the area of each MF by means of the following formula, where the *h* is equal to the respective degrees of membership:

$$A_{Normal} = \frac{[0.6 + (x_2 - x_1)] \cdot h_{Normal}}{2} \qquad A_{High} = \frac{[0.4 + (x_2 - x_1)] \cdot h_{High}}{2}$$

$$A_{Normal} = \frac{[0.6 + (0.62 - 0.38)] \cdot 0.4}{2} \qquad A_{High} = \frac{[0.4 + (1.08 - 0.92)] \cdot 0.6}{2}$$

$$A_{Normal} = 0.168 \qquad A_{High} = 0.168$$

and finally,

$$x_c = \frac{A_{low} \cdot 0 + A_{normal} \cdot 0.5 + A_{high} \cdot 1}{A_{low} + A_{normal} + A_{high}}$$

$$x_c = \frac{0 \cdot 0 + 0.168 \cdot 0.5 + 0.168 \cdot 1}{0 + 0.168 + 0.168} = 0.75$$

In the Risk Model, this $x_c = 0.75$ represents a certainty of risk of an specific settling problem with a percentage of 75 %.

2.4 Results of the Risk Model

Results from the risk model allow the degree of truthfulness to be obtained on the risk of microbiologically-related activated sludge solids separation problems (bulking, foaming, rising sludge and/or deflocculation) during the evaluation period (7th to the 14th day in the BSM1). (The aim of these model is to quantify whether the evaluated control strategies could lead the process to the favorable conditions for them to arise or not).

The risk model results are reported and quantified in three different ways for each operational problem risk (R), $R_{Bulking}$, $R_{Foaming}$, R_{Rising} and $R_{Deflocculation}$:

i) A time series plot showing the risk during the evaluation period; Figure 9 presents this type of results obtained when applying the risk model to test and control two different control strategies;

ii) The percentage of time that the plant is experiencing severe risk (>0.8 over 1); and,

iii) The worst situation computed as the largest time interval the plant is exposed to an uninterrupted severe risk. If results from the risk model indicate that conditions for more than one cause for the same operational problem are fulfilled, the worst conditions (higher value) are selected.

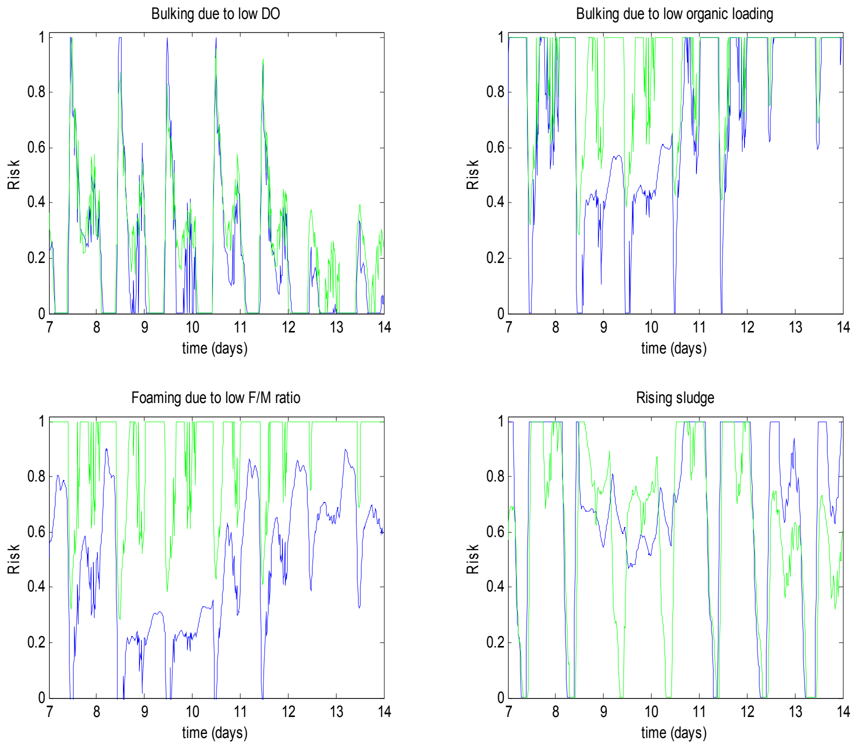

Figure 9. Risk of low DO bulking, low F/M bulking, low F/M foaming and rising under rainy influent conditions (blue: BSM1 default control strategy; green: same as the default one but with RAS flow rate = 2·Inflow rate).

An example of all the computations for the fuzzy implementation (membership functions, fuzzification, rule base, defuzzification) is presented from pages 12 to 18. For further details about the risk model, you can also refer to Comas *et al.* (2006).

Work in progress with documents available shortly:

- In order to demonstrate the correct performance of the proposed risk model, a **sensitivity analysis** is now being evaluated under different control strategies, for long term influent files, different disturbances and for different configurations.