An improved sampling strategy based on trajectory design for application of the Morris method to systems with many input factors

M.V. Ruano[a,*], J. Ribes[a], A. Seco[a] and J. Ferrer[b]
[a]Dep. Enginyeria Química, Universitat de València. Dr. Moliner, 50. 46100 – Burjassot. València. Spain. (E-mail: *m.victoria.ruano@uv.es; josep.ribes@uv.es; aurora.seco@uv.es*)
[b]Instituto de Ingeniería del Agua y Medio Ambiente. Universidad Polotécnica de Valencia. Camino de Vera s/n. 46022. Valencia. Spain. (E-mail: *jferrer@hma.upv.es*)

Abstract
In this paper, a revised version of the Morris approach, which includes an improved sampling strategy based on trajectory design, has been adapted to the screening of the most influential parameters of a fuzzy controller applied to WWTPs. Due to the high number of parameters, a systematic approach has been proposed to apply this improved sampling strategy with low computational demand. In order to find out the proper repetition number of elementary effects of each input factor on model output (EE_i) calculations, an iterative and automatic procedure has been applied. The results show that the sampling strategy has a significant effect on the parameter significance ranking and that random sampling could lead to a non-proper coverage of the parameter space.

Keywords
Fuzzy controllers; Morris screening; Sampling strategy; Sensitivity analysis

1. Introduction
WWTP models are used for many applications/purposes including plant design, optimisation and control. It is generally accepted that the modeling and simulation of WWTPs represents a powerful tool for control system design and tuning. However, model predictions are not free from uncertainty as these models are an approximation of reality (abstraction), and are typically built on a considerable number of assumptions. In this regard, sensitivity analysis provides useful information for the modellers as this technique attempts to quantify how a change in the model input parameters affects the model outputs. Different strategies have been applied in the literature (see for instance, Saltelli et al., 2000, Shahsavani and Grimvall, 2011, Nossent et al., 2011), which are typically classified into two main categories: global sensitivity analysis, where a sampling method is taken and the uncertainty range given in the input reflects the uncertainty in the output variables (Monte Carlo analysis; Fourier Amplitude Sensitivity Test (FAST), variance-based sensitivity analysis, Morris Screening (1991)); and local sensitivity analysis, which is based on the local effect of the parameters on the output variables (Weijers and Vanrolleghem, 1997; Brun et al., 2002).

The Morris method is a one-factor-at-a-time (OAT) method of sensitivity analysis, which calculates the so-called elementary effects, EE_i, of each input factor on model outputs. While the EE_i is in itself a local measure of sensitivity, this drawback is overcome by repeating EE_i calculations in the input space domain using Morris' efficient random sampling strategy, which is obtained via a trajectory based design (see for instance, Saltelli and Annoni, 2010). The analysis of the distribution of elementary effects, F_i, of each input factor will assess the relative importance of the input factors, which approximates well to a global sensitivity measure. One key issue of this approach is that the sampling matrix is randomly generated. This random sampling strategy can be characterised by a poor representation of the sampling space, which can lead to a non-proper screening of the non-influential parameters. For this reason, Campolongo *et al*. (2007) suggested a revised version of the elementary effects method, where an improved sampling strategy is defined by maximising the distances between the final trajectories (*r*) selected. However, this improved sampling strategy was found to be unfeasible for large models due to the high computational demand required to solve the resulting combinatorial optimisation problem (Campolongo *et al*., 2007). Apart from trajectory based designs, other sampling strategies have recently been assessed for screening purposes, such as the radial based design (Saltelli et al., 2010; Campolongo et al., 2011). With this approach, the EE of each parameter is evaluated at the same initial point in the parameter space, but with a different step size. This design differs from trajectory based designs, where the EE of each parameter is evaluated with the same step size but at different initial points in the parameter space.

Fuzzy logic based controllers have been successfully applied on wastewater treatment processes (see e.g. Ferrer et al., 1998; Serralta et al., 2002), since fuzzy sets theory offers an effective tool for the development of intelligent control systems (Zhu et al., 2009). Fuzzy control algorithms can be used to create transparent controllers that are easy to modify and extend because the fuzzy rules are written in the language of process experts and operators (Yong et al., 2006). Although these control systems have been shown to be more robust than classical controllers (Manesis et al., 1998; Traoré et al., 2005), they usually contain quite a number of parameters, which complicates their calibration. So far, these control systems have been tuned by trial and error methods, based on technical knowledge of process and controller performance (Chanona et al., 2006). Whatever optimisation method is applied, the fine-tuning of these controllers requires a previous selection of the most important parameters to be adjusted in each particular application. A systematic approach for the fine-tuning of fuzzy controllers based on model simulations was proposed by Ruano et al. (2010a) and it employs three statistical methods: (i) Monte-Carlo procedure: to find proper initial conditions, (ii) identifiability analysis: to find an identifiable parameter subset of the fuzzy controller based on local sensitivity analysis and (iii) minimisation algorithm. However, this methodology is based on local sensitivity analysis, and then requires an iterative procedure to confirm that the identifiable parameter subset does not depend on the local point in the parameter space where the identifiability study has been carried out. A global sensitivity analysis based on the Morris approach was proposed to overcome the problem of selecting the proper initial point in the parameter space (Ruano *et al*., 2010b). However, the random sampling strategy of this approach could lead to a non-proper screening of the non-influential parameters (Campolongo *et al*., 2007). In this study, the revised version of the Morris approach proposed by Campolongo *et al*. (2007) has been applied to screen out the most influential parameters of a fuzzy logic based aeration control system for WWTPs. Due to the high number of parameters, a systematic procedure has been proposed to overcome the high computational demand of this approach. Hence, an improved sampling strategy based on trajectory design is proposed for the application of the Morris method to systems with many input factors. Although this procedure does not guarantee that the final trajectories (*r*) selected present the global maximum distances between them, these distances are at least locally maximised. Finally, the results obtained with the application of the improved sampling strategy are compared with the ones obtained with a random sampling strategy.

2. Materials and Methods

2.1. Model description

The Morris method was applied to assess the sensitivity of the parameters of a fuzzy logic based control system for controlling the aeration in a nutrient removing WWTP (see Figure 1). The fuzzy controller and the WWTP model were implemented and simulated using the simulation software DESASS (Ferrer *et al*., 2008). This software includes the plant-wide model Biological Nutrient Removal Model nº 1 (BNRM1, Seco et al., 2004). The control system was previously developed by the research group and it has been applied in several full scale WWTPs (Ribes et al., 2007). The main objective of this control system is to control the oxygen in the plant by using two types of controllers: (i) dissolved oxygen and (ii) air pressure. The control system modifies the valve opening according to DO concentration and the rotational speed of the blower according to discharge pressure. As each control valve is governed by an independent DO controller, the air pressure controller is implemented in order to enhance the control system when there is more than one air valve in the same air pipeline. This controller aims to ensure that the one valve opening governed by its DO controller does not affect the air flow rate through the other valves in the same air pipeline. However, in this case study there is only one DO controller in order to simplify the aeration control system.

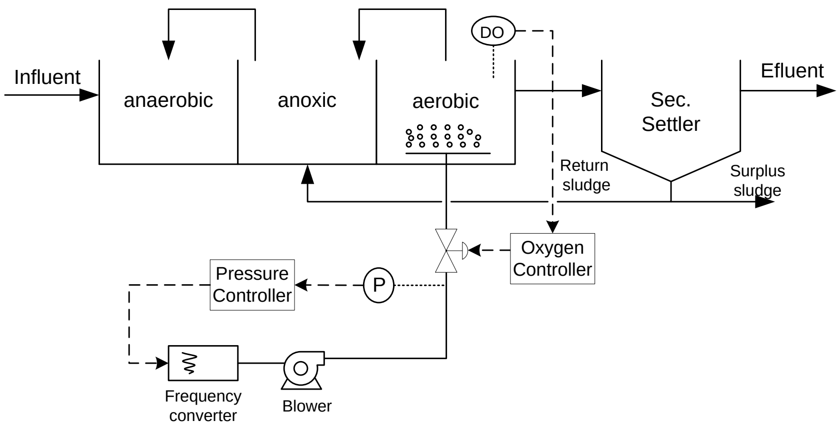
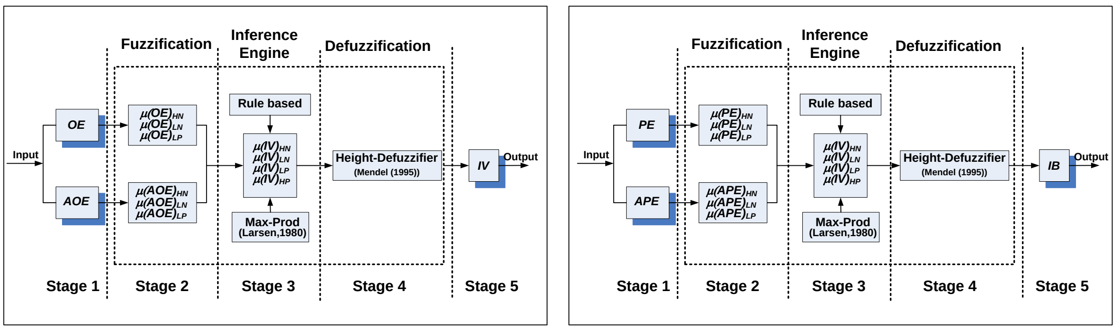


Figure 1. Flow diagram of the control system applied to a modified UCT process.

For the DO controller the input variables are the oxygen error (*OE*) and the accumulated error (*AOE*), and the output variable is the increment/decrement of the air valve opening (*IV*). For the air pressure controller the input variables are the pressure error (*PE*) and the accumulated error (*APE*), and the output variable is the increment/decrement of the rotational speed of the blower (*IB*), which is governed by a frequency converter. Both controllers are fuzzy logic based controllers, which consist of five stages. Figure 2 (a) and Figure 2 (b) show these five stages for the dissolved oxygen controller and the air pressure controller, respectively.



a) DO controller

b) Air pressure controller

Figure 2. Fuzzy control stages for the two controllers: (a) dissolved oxygen controller; and (b) air pressure controller.

The total number of parameters of both controllers (17) comes from these different stages into which fuzzy logic based controllers are divided, mainly derived from the *defuzzification* and *fuzzification* steps (Ruano *et al*., 2010a). In order to identify the parameters of this control system, acronyms for each parameter have been used. These acronyms are constructed as follows: "abbreviation of input variable"+"c/a"+"fuzzification/defuzzification membership function abbreviation". For instance, the acronym *OEaHN* means the amplitude of the High Negative membership function for the input variable Oxygen Error; and the acronym *IVcLN* means the centre of the Low Negative membership function for the output variable Increment air Valve opening. The simulation strategy consisted of a steady-state simulation to obtain proper initial conditions followed by 28 days of dynamic simulations. The last 14 days were considered for the evaluation of the control system performance. The standardised influent file for dry weather proposed by Copp (2002) was used in this study. The Integral Absolute Error (IAE, integral of the absolute value of the time dependent error function) obtained over the last 14 days for each controller was selected as the output measure (IAEO for Oxygen controller and IAEP for Pressure controller). So in this study the weighted contribution of the elementary effects obtained from both output variables was used.

2.2. Morris screening with the improved sampling strategy

The Morris method (1991) evaluates the so-called distribution of Elementary Effects (EE) of each input factor on model outputs, from which basic statistics are computed to derive sensitivity information. In this case study the scaled elementary effects SEE_i proposed by Sin and Gernaey (2009) were applied. The finite distribution of elementary effects associated with each input factor denoted as F_i is obtained by randomly sampling different X from the parameter space. Nevertheless, this random sampling from X can imply a limited coverage of the space. Therefore, we applied the improved sampling strategy proposed in Campolongo *et al*. (2007). This idea consists in selecting the r trajectories in such a way as to maximise their dispersion in the input space. At first, a high number of random Morris trajectories M are generated and then the highest spread r trajectories are chosen. This spread is defined following the definition of distance between a couple of trajectories m and l defined by the following equation:

$$d_{ml} = \begin{cases} \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \sqrt{\sum_{z=1}^{k} \left[X_i^m(z) - X_j^l(z) \right]^2} \end{cases} \text{ for } m \neq l \text{ otherwise } d_{ml} = 0 \quad (1)$$

where $X_i^m(z)$ indicates the zth coordinate of the ith point of the mth Morris trajectory. Consequently, the best r trajectories out of M are selected by maximising the distance d_{ml} among them, and thus the quantity D, which is the sum of all the distances d_{ml} between the couple of trajectories belonging to the combination. This D quantity must be calculated for each possible combination of r trajectories. Consequently, the evaluation of all the possible combinations results in a high computational demand. To solve this combinatorial optimisation problem we developed an alternative methodology which does not take into account all the possible combinations, but it gets a combination of r trajectories out of M that are really close to the highest spread ones and with low computational demand. This approach, which has been programmed in Matlab, consists of different iterative steps that are shown in Figure 3, where the r trajectories are selected from a group of all the possible combinations. As Figure 3 shows, firstly the distance matrix, DM, between the initial M trajectories (see Eq.2) is calculated.

$$DM = \begin{bmatrix} 0 & d_{12} & ..... & d_{1M} \\ d_{21} & 0 & ..... & d_{2M} \\ ..... & ..... & ..... & ..... \\ d_{M1} & d_{M2} & .... & 0 \end{bmatrix} \quad (2)$$

Each row of this matrix represents all the geometric distances, d_{ml}, between the trajectory corresponding to the number of the row, m, and the number of the column, l. Then, iteratively from i = 1 to i = r-1 the following procedure is carried out:

1. From each row of the DM matrix, m, the i columns whose d_{ml} are the highest ones are selected, $[n_1, n_2 .... n_i]$. Then, the quantity $D_{n_1,n_2 \ldots n_i,m}$ is calculated for each row of the DM matrix, considering the i trajectories selected in each row. Thus, M values of D are obtained, generating the matrix D_{i+1} defined in the following expression:

$$D_{i+1} = \begin{bmatrix} D_{n_1,n_2 \ldots n_i,1} \\ D_{n_1,n_2 \ldots n_i,2} \\ \cdots\cdots\cdots \\ D_{n_1,n_2 \ldots n_i,M} \end{bmatrix} \quad (3)$$

where the sub index i+1 corresponds to the total number of trajectories considered. Then, the maximum value of D_{i+1} is selected, which represents the highest spread i+1 trajectories of these matrix: $n_{H_1}, n_{H_2} \ldots n_{H_{i+1}}$

2. The next step is the selection of the r-(i+1) trajectories. Subsequently, iteratively, for k = 1 to k = r-i, the matrix D_{i+k+1} defined in the following expression is calculated:

$$D_{i+k+1} = \begin{bmatrix} D_{n_{H_1},n_{H_2} \ldots n_{H_{i+k}},1} \\ D_{n_{H_1},n_{H_2} \ldots n_{H_{i+k}},2} \\ \cdots\cdots\cdots \\ D_{n_{H_1},n_{H_2} \ldots n_{H_{i+k}},M} \end{bmatrix} \quad (4)$$

where each row represents the quantity D for the trajectories: $n_{H_1}, n_{H_2} \ldots n_{H_{i+k}}, m$. Then, the maximum value of D_{i+k+1} is obtained, and the corresponding i+k+1 trajectory is selected.

3. At last, for the specific i considered, a combination of r trajectories out of M is obtained: $n_{H_1}, n_{H_2} . n_{H_{r_i}}$.

Once the r-1 iterations are executed, r-1 combinations of r trajectories will be generated, which are represented in the matrix D_r defined by the following expression:

$$D_r = \begin{bmatrix} D_{n_{H_1},n_{H_2} \ldots n_{r_1}} \\ D_{n_{H_1},n_{H_2} \ldots n_{r_2}} \\ \cdots\cdots\cdots \\ D_{n_{H_1},n_{H_2} \ldots n_{r_{r-1}}} \end{bmatrix} \quad (5)$$

Finally, the maximum value of D_r will define the final r trajectories selected, $D_{n_{H_1},n_{H_2} \ldots n_{r_i}}$.

As an example, Figure 4 compares the empirical distributions (sampled values) for k = 4 parameters that are respectively obtained via the random sampling strategy proposed by Morris (1991) (Figure 4a), via the revised sampling strategy proposed by Campolongo et al. (2007) (Figure 4b), and via the sampling strategy proposed in this paper (Figure 4c). The theoretical distributions in this example are the same as those proposed in Campolongo et al. (2007), i.e. discrete uniform with p = 4 levels and for r = 20 trajectories. The empirical distribution illustrated in Figure 4c is not as close to the theoretical shape of the discrete uniform distributions as the one obtained in Figure 4b, but it significantly improves the results in comparison with the ones obtained with the random sampling strategy (Figure 4a). The sampling strategy proposed in this paper considerably reduces the computational demand required to select the optimal r trajectories out of M. For instance, for M = 100 initial trajectories and r = 5, the computational cost of the rigorous method was 8 minutes whilst the method proposed in this paper gave the results in less than 2 seconds (obtained with Matlab® using a PC with a 2.53 GHz Intel® Core™ i5 processor). This computational demand depends mainly on the number of trajectories to be selected (r) from the initial ones (M). For instance, for selecting r = 20 trajectories out of M = 1000, the computational cost of the proposed strategy is about 5 minutes, whilst applying the rigorous approach it is unfeasible due to the high number of combinations to be evaluated (almost $3.4 \cdot 10^{41}$). Although the proposed sampling strategy does not guarantee that the final trajectories (r) selected present the global maximum distances between them, these distances are at least locally maximised. For instance, for the simple analytical example mentioned above (r = 5 trajectories out of M = 100 with k = 4 parameters and a grid of p = 4), the distance between the final r trajectories obtained with the methodology proposed in this paper was D = 95.6 and with the rigorous approach was D = 97.6 (the global maximum distance). This difference in terms of distance can be acceptable enough, provided that when the random sampling strategy was repeated 10 times the highest distance obtained was D = 70.4.

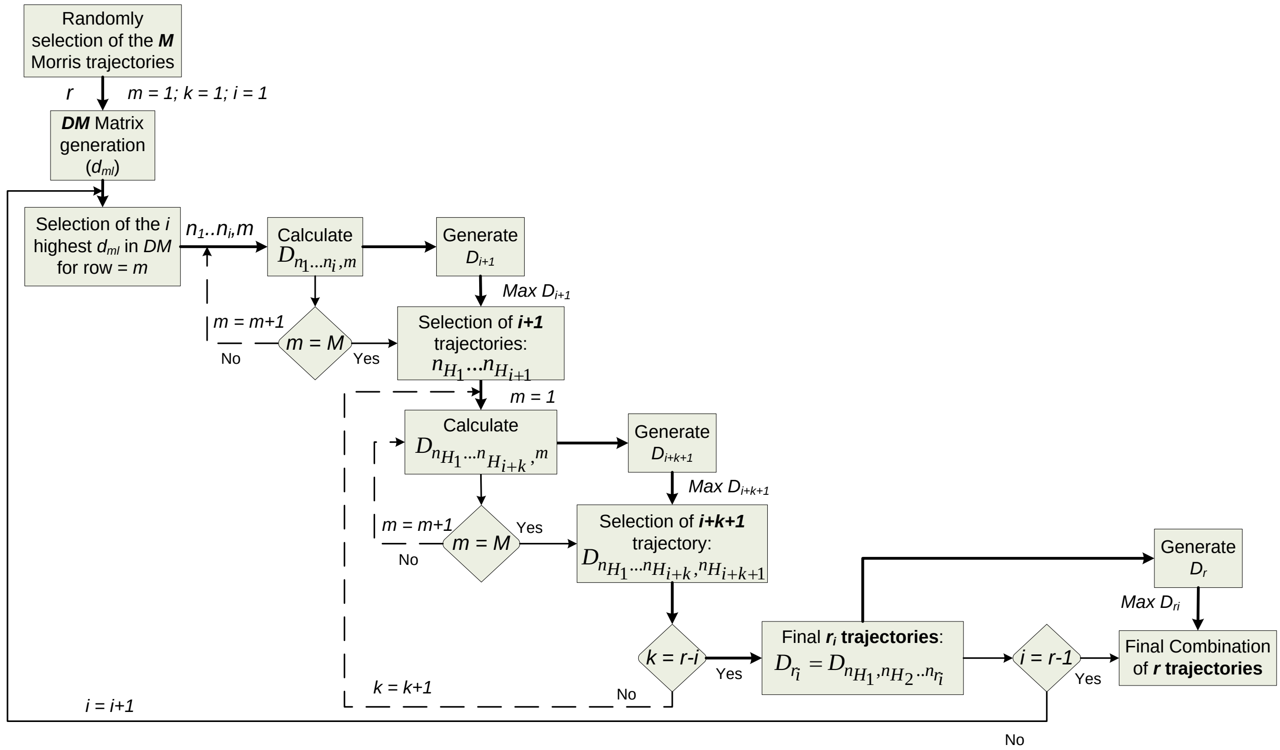


Figure 3. Flow diagram of the proposed methodology to find r high spread trajectories out of M.

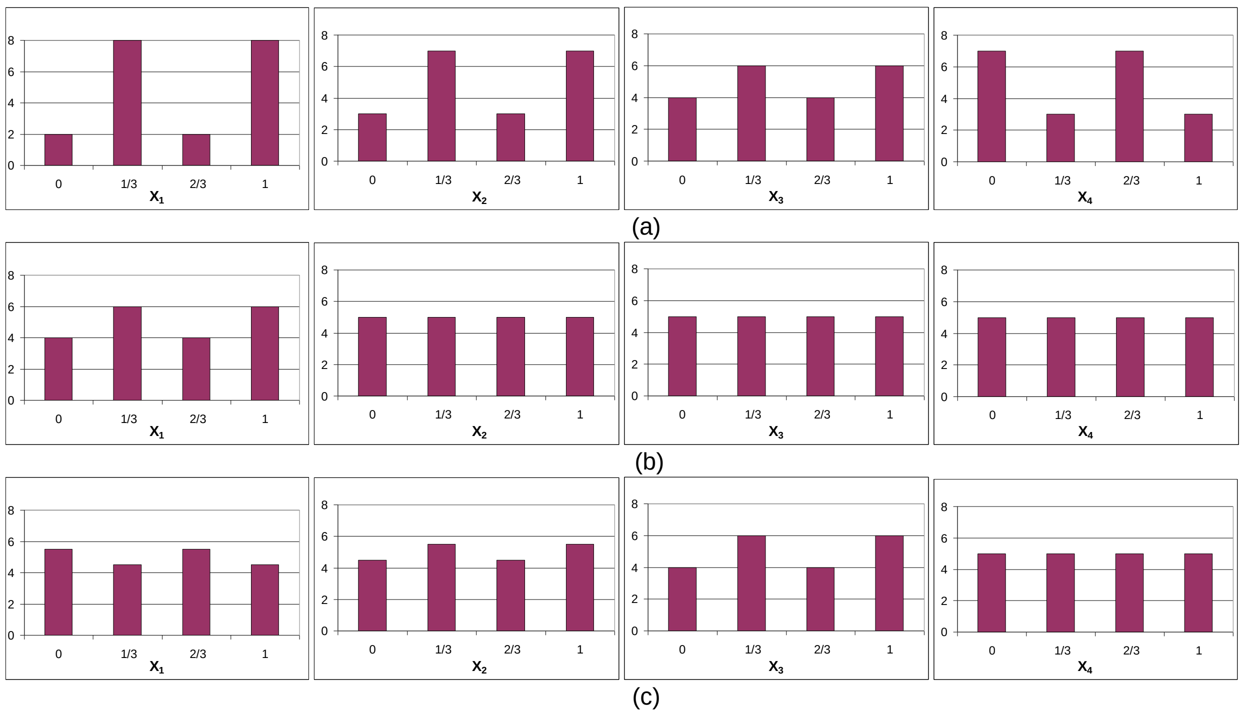


Figure 4. Empirical distributions for k = 4 parameters, X_1, X_2, X_3, X_4 whose theoretical distributions are uniform discrete with 4 levels, sample size r = 20. The samples are obtained using the random sampling strategy (a); the revised sampling strategy proposed by Campolongo et al. (2007) (b); and the sampling strategy proposed in this paper (c).

With regard to the sensitivity measures, the mean (μ), the standard deviation (σ) and the absolute mean (μ^*) of the SEE_i values of each F_i are considered (Saltelli *et al*., 2004). The measure μ^* has been used to rank the parameters in order to systematically identify the non-influential parameters (low μ^*) from the influential ones (high μ^*). An optimal setting of r (r_{opt}) has been searched for with a constant resolution of p = 8. To this end, the repetition number of elementary effect calculations (r) of each distribution F_i was increased until the ranking of parameters (based on μ^*) remained more or less stable, i.e. the type II error is minimised (type II error: indentifying an important factor as insignificant). This stability has been numerically evaluated with a numerical index proposed in this paper (the position factor, $PF_{ri\rightarrow rj}$). For given rankings obtained by r_i and r_j, we define the index $PF_{ri\rightarrow rj}$ by the following expression:

$$PF_{r_i \rightarrow r_j} = \sum_{k=1}^{k} \frac{P_{k,i} - P_{k,j}}{\mu_{P_{k,i},P_{k,j}}} \qquad (6)$$

where $P_{k,i}$ is the position of the kth parameter in the ranking obtained by r_i, and $\mu_{P_{k,i},P_{k,j}}$ is the average of the kth parameter positions in the ranking obtained by r_i and r_j. The index $PF_{ri\rightarrow rj}$ indicates how different the rankings calculated by sampling sizes r_i and r_j are, i.e. a low value means that most of the parameters remain in the same or nearly the same position in the ranking (If $PF_{ri\rightarrow rj}$ = 0 then Ranking r_i = Ranking r_j). Moreover, this index decreases the importance of a change in the position of parameters located at the bottom of the ranking. This criterion allows an optimal value for r (r_{opt}) to be found. Once the r_{opt} was found, the graphical Morris approach was used to find the significant parameters. In order to evaluate the effect of this improved sampling strategy, the Morris approach based on a random sampling strategy has also been applied and the results obtained with both sampling strategies have been compared.

3. Results and Discussion

The Morris method with the proposed improved sampling strategy was applied to a different number of trajectories (r), chosen from M = 1000 initial Morris trajectories, until the parameter significant ranking remained more or less stable, quantitatively measured by the index $PF_{ri\rightarrow rj}$. Similarly, the Morris approach based on random sampling was also applied. Table 1 shows the resulting sensitivity measures (μ^* and σ) for the different number of elementary effects calculated with the improved sampling strategy. Table 2 shows the resulting $PF_{ri\rightarrow rj}$ values for each pair of compared rankings: (a) the improved sampling strategy, and (b) the random sampling strategy. As can be seen in this table, at a low number of trajectories for the improved sampling strategy (from r = 5 to r = 40) the tendency of the $PF_{ri\rightarrow rj}$ value is not monotonic. This behaviour reveals that, for this case study, values of r below 40 do not provide a suitable estimation of the sensitivity measures, which can be obtained when either a highly nonlinear model is used or a large input uncertainty is defined. These results are in contrast to previous applications of the Morris method since most of those studies used a low repetition number, e.g. r = (10~20) (Campolongo *et al*, 2007, Ruano *et al*., 2010b). Then, for higher values of r, the index manifests a downward trend as the number of trajectories is increased, which demonstrates a closer similarity between the positions of the parameters in the compared rankings. In contrast, the $PF_{ri\rightarrow rj}$ values for the random sampling strategy decreases as the number of trajectories is increased, except for r = 40→50 and for r = 60→70. Surprisingly, working with considerably high values of r, an increase in the number of trajectories from r = 60 to r = 70 does not imply an improvement in the $PF_{ri\rightarrow rj}$ index. Moreover, the increase in this index is mainly due to a position change of the parameters located at the top of the ranking (data not shown) such as: *PEcHN* (*High Negative membership function centre of the Pressure Error*). These results can be interpreted as a result of a non-optimal coverage of the sampling space obtained by a random strategy, which could lead to a Type II error, i.e. failing in the identification of a parameter of considerable influence in the model; and a Type I error, as well, i.e.

considering a factor as significant when it is not. Thus, a suitable scan of the input space, such as the improved sampling strategy applied in this study, can lead to more realistic results.

Table 1. Sensitivity measures of the control parameters at the different *r* evaluated with the improved sampling strategy.

***r* = 5**			***r* = 10**			***r* = 15**			***r* = 30**		
Parameter	**μ^***	**σ**	**Parameter**	**μ^***	**σ**	**Parameter**	**μ^***	**σ**	**Parameter**	**μ^***	**σ**
PEaHN	2.401	3.356	*PEcHN*	3.101	3.895	*APEaHN*	2.451	3.557	*APEaHN*	2.141	3.246
APEcHN	1.720	2.522	*IVcHN*	1.505	2.030	*IBcHN*	1.682	2.889	*IBcHN*	1.675	2.449
IVcHN	1.668	2.643	*PEaHN*	1.434	2.538	*PEcHN*	1.586	3.049	*PEcHN*	1.579	2.835
IBcHN	1.436	2.364	*APEcHN*	1.264	1.813	*IVcHN*	1.497	2.318	*PEaHN*	1.162	2.101
PEcHN	1.144	2.248	*IBcHN*	1.102	2.215	*IBcLN*	1.356	2.035	*PEcLN*	0.860	1.639
PEcLN	0.681	1.123	*APEaHN*	1.041	1.430	*PEaHN*	1.043	1.700	*IBcLN*	0.809	1.697
OEcHN	0.601	0.779	*RT*	0.948	1.220	*RT*	1.022	1.652	*AOEaHN*	0.774	1.927
AOEaHN	0.493	1.015	*OEaHN*	0.901	1.591	*OEcHN*	0.952	1.793	*APEcHN*	0.736	1.164
APEaHN	0.424	0.487	*IBcLN*	0.872	1.648	*PEcLN*	0.942	1.964	*IVcHN*	0.706	1.352
OEaHN	0.408	0.651	*APEcLN*	0.770	1.397	*APEcHN*	0.641	1.417	*RT*	0.684	1.504
RT	0.316	0.405	*PEcLN*	0.416	0.901	*OEaHN*	0.632	1.298	*OEcHN*	0.603	1.264
APEcLN	0.289	0.506	*OEcHN*	0.385	0.860	*IVcLN*	0.487	1.074	*OEaHN*	0.586	1.318
OEcLN	0.213	0.309	*IVcLN*	0.332	0.651	*APEcLN*	0.288	0.712	*APEcLN*	0.433	1.213
IBcLN	0.182	0.263	*AOEaHN*	0.305	0.591	*AOEcHN*	0.180	0.287	*AOEcLN*	0.274	0.905
AOEcHN	0.108	0.193	*OEcLN*	0.245	0.453	*AOEcLN*	0.156	0.370	*IVcLN*	0.261	0.953
AOEcLN	0.074	0.137	*AOEcLN*	0.222	0.540	*OEcLN*	0.151	0.275	*OEcLN*	0.237	0.621
IVcLN	0.019	0.028	*AOEcHN*	0.146	0.309	*AOEaHN*	0.149	0.293	*AOEcHN*	0.174	0.326

***r* = 40**			***r* = 50**			***r* = 60**			***r* = 70**		
Parameter	**μ^***	**σ**	**Parameter**	**μ^***	**σ**	**Parameter**	**μ^***	**σ**	**Parameter**	**μ^***	**σ**
IBcHN	1.848	2.584	*APEaHN*	2.010	3.263	*APEaHN*	2.272	3.498	*APEaHN*	2.222	3.208
APEcHN	1.681	2.571	*IBcHN*	1.981	3.030	*PEcHN*	1.973	3.282	*PEcHN*	1.909	3.249
APEaHN	1.412	2.067	*PEcHN*	1.720	2.466	*IBcHN*	1.697	2.477	*IBcHN*	1.680	2.511
PEcHN	1.394	2.180	*IVcHN*	1.204	2.121	*PEaHN*	1.397	2.299	*PEaHN*	1.315	2.257
IVcHN	1.319	2.424	*PEaHN*	1.042	1.775	*IVcHN*	1.108	2.100	*APEcHN*	1.085	2.138
OEaHN	1.285	2.816	*APEcHN*	0.994	2.068	*APEcHN*	1.007	1.745	*IVcHN*	0.941	1.669
PEaHN	1.023	1.805	*PEcLN*	0.912	1.698	*OEaHN*	0.900	1.744	*PEcLN*	0.891	1.682
OEcHN	0.957	1.694	*OEcHN*	0.771	1.483	*RT*	0.818	1.699	*RT*	0.861	1.604
PEcLN	0.798	1.628	*RT*	0.763	1.445	*IBcLN*	0.816	1.703	*IBcLN*	0.761	1.486
IBcLN	0.752	1.649	*OEaHN*	0.744	1.317	*OEcHN*	0.779	1.453	*OEcHN*	0.721	1.425
AOEaHN	0.493	1.210	*IBcLN*	0.712	1.350	*PEcLN*	0.741	1.403	*OEaHN*	0.688	1.356
RT	0.431	0.713	*AOEaHN*	0.349	0.859	*AOEcHN*	0.432	1.300	*APEcLN*	0.301	0.880
OEcLN	0.341	0.887	*IVcLN*	0.308	1.452	*AOEaHN*	0.393	1.165	*AOEaHN*	0.282	0.741
AOEcHN	0.335	1.204	*APEcLN*	0.272	0.723	*AOEcLN*	0.312	1.273	*AOEcHN*	0.270	0.792
APEcLN	0.218	0.726	*OEcLN*	0.128	0.299	*APEcLN*	0.295	1.164	*IVcLN*	0.170	0.724
AOEcLN	0.161	0.541	*AOEcHN*	0.122	0.367	*IVcLN*	0.132	0.517	*OEcLN*	0.164	0.561
IVcLN	0.083	0.222	*AOEcLN*	0.039	0.090	*OEcLN*	0.121	0.294	*AOEcLN*	0.123	0.393

Table 2. Position factors, $PF_{ri \rightarrow rj}$, for the *r* calculated: (a) improved sampling strategy; (b) random sampling strategy.

$r_i \rightarrow r_j$	5→10	10→15	15→30	30→40	40→50	50→60	60→70
(a) $PF_{ri \rightarrow rj}$	7.5	7.8	4.2	7.8	5.4	3.5	1.9
(b) $PF_{ri \rightarrow rj}$	10.1	7.6	5.2	2.1	2.7	2.3	3.7

As a result, *r* = 70 was selected as the optimal setting for this case study. The overall model evaluation cost was, therefore, 1190 simulations. We considered *r* = 70 as the optimal one, not only due to the low $PF_{60 \rightarrow 70}$ value but also due to the significant stability of the parameters located at the top of the rankings (see Table 1). Figure 5a shows the graphical Morris approach for the optimal number of repetitions obtained for the improved sampling strategy. In addition, Figure 5b shows the same graph for r = 70 obtained with the random sampling strategy. This figure was used to screen out the non-influential parameters of the control system (i.e., the six parameters that are not labelled in Figure 1a). From the eleven influential parameters, *RT* (Response Time), *OEcHN* (*High Negative centre of the Oxygen Error*), *APEaHN* (*High Negative amplitude of the Accumulated Pressure Error*), *IVcHN* (*High Negative centre of the Increment of the air Valve opening*) and *IBcLN* (*Low Negative centre of the Increment of the rotational speed of the Blower*) presented a high mean and a low standard deviation, lying outside the wedge formed by two lines corresponding to $\mu_i = \pm 2SEM_i$. Thus, the effect of these parameters on the output variables are expected to be linear and additive, which is desirable for parameter estimation based on optimisation algorithms. In comparison with the results obtained for the random sampling strategy, the following could be said: (i) the resulting non-influential parameters agreed with the improved sampling strategy; (ii) the sensitivity measures of the eleven influential parameters are different from the ones obtained with the improved sampling strategy, which reflects that the sampling strategy has a significant effect on the parameter significance ranking. In addition, the necessity of finding out the optimal repetition number for SEE_i calculations (r) has been underlined. Thus, a non-optimal selection of r would lead to Type I or Type II errors, as well.

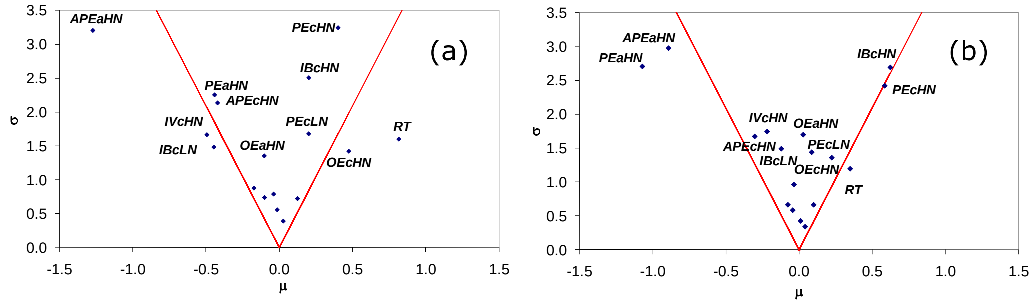


Figure 5. (a) μ versus σ, for r_{opt} = 70, for the improved sampling; (b) μ versus σ, for r = 70, for the random sampling. Lines correspond to $\mu_i = \pm 2SEM_i$;

4. Conclusions

The Morris method with the improved sampling strategy proposed by Campolongo *et al*. (2007) has been applied to a fuzzy logic based control system of a WWTP. A systematic approach has been proposed in order to apply this improved sampling strategy based on trajectory design to large models with low computational demand. In order to find out the proper repetition number of SEE_i calculations (r_{opt}), an iterative and automatic procedure has been applied. The optimal repetition number found in this study is in direct contrast with previous applications of the Morris method, which usually uses a low number of repetitions, e.g. r = (10~20). This high r value can be explained by either a highly nonlinear behaviour of the system, or the definition of a large input uncertainty. The results show that the sampling strategy has a significant effect on the parameter significance ranking and that the random sampling strategy could lead to a non-proper coverage of the sample space. Working with a non-proper sampling matrix and a non-proper sample size (r) could lead to either Type I or Type II errors. Overall, the improved sampling strategy proposed and the iterative and automatic procedure to find out the proper repetition number of SEE_i calculations to apply the Morris approach, provides a good approximation of a global sensitivity measure, helping engineers to calibrate large models with many input factors such as the fuzzy control system used in this study.

Acknowledgments

This research study has been supported by the Spanish Research Foundation (CICYT Project reference CTM2005-06919-C03-01/TECNO), which is gratefully acknowledged.

References

Brun, R., Kühni, M., Siegrist, H., Gujer, W., Reichert, P., 2002. Practical identifiability of ASM2d parameters-systematic selection and tuning of parameter subsets. Water Res. 36(16), 4113-4127.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environ. Mod. Soft. 22 (10), 1509-1518.

Campolongo, F., Saltelli, A., Cariboni, J, 2011. From screening to quantitative sensitivity analysis. A unified approach. Comput. Phys. Commun. 182, 978-988.

Chanona, J., Pastor, L., Borras, L., Seco, A., 2006. Fuzzy algorithm for pH control in a struvite crystallization reactor. Wat. Sci. Tech. 53(12), 161-168.

Copp J., 2002. The COST Simulation Benchmark – Description and Simulator Manual. COST (European Cooperation in the field of Scientific and Technical Research), Brussels, Belgium.

Ferrer, J., Rodrigo, M.A., Seco, A., Penya-Roja, J.M., 1998. Energy saving in the aeration process by fuzzy logic control. Wat. Sci. Tech. 38(3), 209-217.

Ferrer, J., Seco, A., Serralta, J., Ribes, J., Manga, J., Asensi, E., Morenilla, J.J., Llavador F., 2008. DESASS: a software tool for designing, simulating and optimising WWTPs. Environ. Mod. Soft. 23, 19-26.

Manesis, S.A., Sapidis, D.J., King R.E., 1998. Intelligent control of wastewater treatment plants. Artif. Intell. Eng. 12 (3), 275-281.

Morris, M. (1991). Factorial sampling plans for preliminary computational experiments. Technometrics. 33(2), 239-245.

Nossent J., Elsen P., Bauwens W., 2011. Sobol' sensitivity analysis of a complex environmental model. Environ. Modell. Softw. 26(12), 1515-1525.

Ribes, J., Serralta, J., Seco, A., Ferrer, J., Lloret, R., Bernácer, I., Morenilla J.J., 2007. Implementation of a supervisory fuzzy logic based control system in the Denia WWTP. In: 10th IWA Conference on Design, Operation and Economics of Large Wastewater Treatment Plants. 77-80. Vienna.

Ruano, M.V., Ribes, J., Sin, G., Seco, A., Ferrer J., 2010a. A systematic approach for fine-tuning of fuzzy controllers applied to WWTPs. Environ. Modell. Softw. 25(5), 670-676.

Ruano, M.V., Ribes, J., Ferrer, J., Sin G., 2010b. Application of the Morris method for screening the influential parameters of fuzzy controllers applied to WWTPs. Wat. Scien. Tech. (In press).

Saltelli, A, Scott, EM, Chen, K, editors, 2000. Sensitivity analysis. Chichester: Wiley.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity analysis in practice: a guide to assessing scientific models. Chicester: Wiley.

Saltelli A., Annoni P., 2010. How to avoid a perfunctory sensitivity analysis. Environ. Modell. Softw. 25(12), 1508-1517.

Saltelli, A., Annoni, P., Azzini. I., Campolongo, F., Ratto M., Tarantola, S, 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181, 259-270.

Seco, A., Ribes, J., Serralta, J., Ferrer J., 2004. Biological Nutrient Removal Model No.1. Water Sci. Technol. 50(6), 69-78.

Serralta, J., Ribes, J., Seco, A., Ferrer, J., 2002. A supervisory control system for optimising nitrogen removal and aeration energy consumption in wastewater treatment plants. Wat. Sci. Tech. 45(4/5), 309-316.

Shahsavani D., Grimvall A., 2011. Variance-based sensitivity analysis of model outputs using surrogate models. Environ. Modell. Softw. 26(6), 723-730.

Sin, G., Gernaey, KV., 2009. Improving the Morris method for sensitivity analysis by scaling the elementary effects. In proceedings of the 19th European Symposium on Computer Aided Process Engineering (ESCAPE 19), eds. J. Jezowski and J. Thullie. Elsevier, pp 925-930.

Traoré, A., Grieu, S., Corominas, L., Thiery, F., Polit, M., Colprim J., 2005. Fuzzy control of dissolved oxygen in a sequencing batch reactor pilot plant. Chem. Eng. Jour. 111(1), 13–19.

Yong, M., Yong-zhen, P., Xiao-lian, W., Shu-ying W., 2006. Intelligent control aeration and external carbon addition for improving nitrogen removal. Environ. Modell. Softw. 21, 821-828.

Weijers, S.R., Vanrolleghem, P.A., 1997. A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data. Wat. Sci. Tech. 36(5), 69–79.

Zhu G., Peng Y., Ma B., Wang Y., Yin C., 2009. Optimization of anoxic/oxic step feeding activated sludge process with fuzzy control model for improving nitrogen removal. Chem. Eng. J. 151(1-3), 195-201.

Improved sampling strategy based on trajectory design for the application of Morris method to systems with many input factors

M.V. Ruano[a,*], J. Ribes[a], A. Seco[a] and J. Ferrer[b]
[a]Dep. Enginyeria Química, Universitat de València. Dr. Moliner, 50. 46100 – Burjassot. València. Spain. (E-mail: *m.victoria.ruano@uv.es; josep.ribes@uv.es; aurora.seco@uv.es*)
[b]Instituto de Ingeniería del Agua y Medio Ambiente. Universidad Polotécnica de Valencia. Camino de Vera s/n. 46022. Valencia. Spain. (E-mail: *jferrer@hma.upv.es*)

Abstract
In this work, a revised version of Morris approach, which includes an improved sampling strategy based on trajectory design, has been adapted to the screening of the most influential parameters of a fuzzy controller applied to WWTPs. Due to the high number of parameters, a systematic approach to apply this improved sampling strategy with low computational demand has been proposed. In order to find out the proper repetition number of elementary effects of input factor to model outputs (EE_i) calculations, an iterative and automatic procedure has been applied. The results show that the sampling strategy has a significant effect on the parameter significance ranking and that random sampling could lead to a non-proper coverage of the parameter space.

Keywords
Fuzzy controllers; Morris screening; Sampling Strategy; Sensitivity analysis

1. Introduction
WWTP models are used for many applications/purposes including plant design, optimisation and control. It is generally accepted that the modeling and simulation of WWTPs represents a powerful tool for control systems design and tuning. However, the model predictions are not free from uncertainty as these models are an approximation of reality (abstraction), and are typically built on a considerable number of assumptions. In this regard, sensitivity analysis provides useful information for the modellers as this technique attempts to quantify how a change in the input model parameters affects the model outputs. Different strategies have been applied in the literature (see for instance, Saltelli et al., 2000), which are typically classified in two main categories: global sensitivity analysis, where a sampling method is taken and the uncertainty range given in the input reflects the uncertainty in the output variables (Monte Carlo analysis; Fourier Amplitude Sensitivity Test (FAST), Morris Screening (1991)); and local sensitivity analysis, which is based on the local effect of the parameters in the output variables (Weijers and Vanrolleghem, 1997; Brun et al., 2002).

Morris method is a one-factor-at-a-time (OAT) method of sensitivity analysis, which calculates the so-called elementary effects, EE_i, of input factor to model outputs. While the EE_i is in itself a local measure of sensitivity, this drawback is overcome by repeating EE_i calculations in the input space domain using Morris' efficient random sampling strategy which is obtained via a trajectory based design. The analysis of the distribution of elementary effects, F_i, of each input factor will assess the relative importance of the input factors, which approximates well a global sensitivity measure. One key issue of this approach is that the sampling matrix is randomly generated. This random sampling strategy can be characterised by a poor representation of the sampling space which can lead to a non-proper screening of the non-influential parameters. For this reason, Campolongo *et al*. (2007) suggested a revised version of the elementary effects method, where an improved sampling strategy is defined by maximizing the distances between the final trajectories (r) selected. However, this improved sampling strategy was found to be unfeasible for large models due to the high computational demand required to solve the resulting combinatorial optimisation problem (Campolongo *et al*., 2007). Besides the trajectory based design other sampling strategies have been recently assessed for screening purposes such as the radial based design (Saltelli et al., 2010; Campolongo et al., 2011). With this approach, the EE of each parameter is evaluated at the same initial point in the parameter space but with different step size. This design differs from the trajectories based design, where the EE of each parameter is evaluated with the same step size but at different initial points in the parameter space.

Fuzzy logic based controllers have been successfully applied on wastewater treatment processes (see e.g. Ferrer et al., 1998; Serralta et al., 2002), since fuzzy sets theory offers an effective tool for the development of intelligent control systems (Zhu et al., 2009). Fuzzy control algorithms can be used to create transparent controllers that are easy to modify and extend because the fuzzy-rules are written in the language of process experts and operators (Yong et al., 2006). Although these control systems have shown to be more robust than classical controllers (Manesis et al., 1998; Traoré et al., 2005), they usually contain quite a number of parameters which complicates their calibration. So far, these control systems have been tuned by trial and error methods, based on technical knowledge on the process and controller performance (Chanona et al., 2006). Whatever the optimisation method is applied, the fine-tuning of these controllers requires a previous selection of the most important parameters to be adjusted in each particular application. A systematic approach for fine tuning of fuzzy controllers based on model simulations was proposed by Ruano et al. (2010a) that employs three statistical methods: (i) Monte-Carlo procedure: to find proper initial conditions, (ii) Identifiability analysis: to find an identifiable parameter subset of the fuzzy controller based on local sensitivity analysis and (iii) minimization algorithm. However, this methodology is based on local sensitivity analysis, and then requires an iterative procedure to confirm the identifiable parameter subset does not depend on the local point in the parameter space where the identifiability study has been carried out. A global sensitivity analysis based on the Morris approach was proposed to overcome the problem of selecting the proper initial point in parameter space (Ruano *et al*., 2010b). However, the random sampling strategy of this approach could lead to a non-proper screening of the non-influential parameters (Campolongo *et al*., 2007). In this work, the revised version of Morris approach proposed by Campolongo *et al*. (2007) has been applied to screen out the most influential parameters of a fuzzy logic based aeration control system for WWTPs. Due to the high number of parameters, a systematic procedure has been proposed to overcome the high computational demand of this approach. Hence, an improved sampling strategy based on trajectory design is proposed for the application of Morris method to systems with many input factors. Although, this procedure does not guaranty that the final trajectories (*r*) selected present the global maximum distances between them, these distances are at least locally maximized. Finally, the results obtained with the application of the improved sampling strategy are compared with the ones obtained with a random sampling strategy.

2. Materials and Methods

2.1. Model description

Morris method was applied to assess the sensitivity of the parameters of a fuzzy logic based control system for controlling the aeration in a nutrient removing WWTP (see Figure 1). The fuzzy controller and the WWTP model were implemented and simulated using the simulation software DESASS (Ferrer *et al*., 2008). This software includes the plant-wide model Biological Nutrient Removal Model nº 1 (BNRM1, Seco et al., 2004). The control system was previously developed by the research group and it has been applied in several full scale WWTPs (Ribes et al., 2007). The main objective of this control system is to control the oxygen in the plant by using two types of controllers: (i) dissolved oxygen and (ii) air pressure. The control system modifies the valve opening according to DO concentration and the rotational speed of the blower according to discharge pressure. As each control valve is governed by an independent DO controller, the air pressure controller is implemented in order to enhance the control system when there is more than one air valve in the same air pipeline. This controller aims that the one valve opening governed by its DO controller does not affect the air flow rate through the other valves in the same air pipeline. However, in this case study there is only one DO controller in order to simplify the aeration control system.

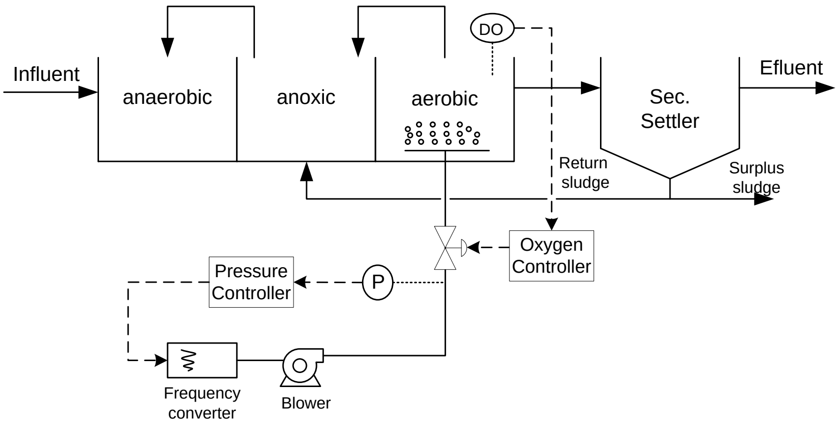
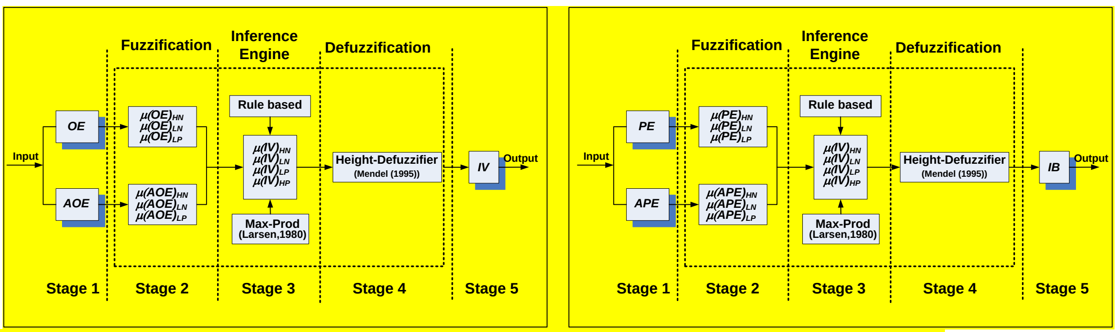


Figure 1. Flow diagram of the control system applied to a modified UCT process.

For the DO controller the input variables are the oxygen error (*OE*) and the accumulated error (*AOE*) and the output variable is the increment/decrement of the air valve opening (*IV*). For the air pressure controller the input variables are the pressure error (*PE*) and the accumulated error (*APE*) and the output variable is the increment/decrement of the rotational speed of the blower (*IB*), which is governed by a frequency converter. Both controllers are fuzzy logic based controllers, which consist of five stages. Figure 2 (a) and Figure 2 (b) show these five stages for the dissolved oxygen controller and the air pressure controller, respectively.



a) DO controller b) Air pressure controller

Figure 2. Fuzzy control stages for the two controllers: (a) dissolved oxygen controller; and (b) air pressure controller.

The total number of parameters of both controllers (17) comes from these different stages in which are divided fuzzy logic based controllers, mainly derived from the *defuzzification* and *fuzzification* steps (Ruano *et al.*, 2010a). In order to identify the parameters of this control system, acronyms for each parameter have been used. These acronyms are constructed as follows: “abbreviation of input variable”+ “c/a”+“fuzzification/defuzzification membership function abbreviation”. For instance, the acronym *OEaHN* means the amplitude of the High Negative membership function for the input variable Oxygen Error; and the acronym *IVcLN* means the centre of the Low Negative membership function for the output variable Increment air Valve opening. The simulation strategy consisted of a steady-state simulation to obtain proper initial conditions followed by 28 days dynamic simulations. The last 14 days were considered to evaluate the performance of the control system. The standardised influent file for dry weather proposed by Copp (2002) was used in this study. The Integral Absolute Error (IAE, integral of the absolute value of the time dependent error function) obtained along the last 14 days for each controller was selected as output measure (IAEO for Oxygen controller and IAEP for Pressure controller). So, in this study the weighted contribution of the elementary effects obtained from both output variables was used.

2.2. Morris screening with the improved sampling strategy

The method of Morris (1991) evaluates the so called distribution of Elementary Effects (EE) of each input factor to model outputs, from which basic statistics are computed to derive sensitivity information. In this case study the scaled elementary effects SEE_i proposed by Sin and Gernaey (2009) was applied. The finite distribution of elementary effects associated with each input factor denoted as F_i is obtained by randomly sampling different X from the parameter space. Nevertheless, this random sampling from X can imply a short coverage of the space. Therefore, we applied the improved sampling strategy proposed in Campolongo *et al*. (2007). This idea consists of selecting the r trajectories in such a way as to maximise their dispersion in the input space. At first, a high number of random Morris trajectories M are generated and then the highest spread r trajectories are chosen. This spread is defined following the definition of distance between a couple of trajectories m and l defined by the following equation:

$$d_{ml} = \begin{cases} \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \sqrt{\sum_{z=1}^{k} \left[X_i^m(z) - X_j^l(z)\right]^2} & \text{for } m \neq l \text{ otherwise } d_{ml} = 0 \end{cases} \quad (1)$$

where $X_i^m(z)$ indicates the zth coordinate of the ith point of the mth Morris trajectory. Consequently, the best r trajectories out of M are selected by maximising the distance d_{ml} among them, and thus, the quantity D, which is the sum of all the distances d_{ml} between couple of trajectories belonging to the combination. This D quantity must be calculated for each possible combination of r trajectories. Consequently, the evaluation of all the possible combinations results in a high computational demand. To solve this combinatorial optimisation problem we developed an alternative methodology which does not take into account all the possible combinations, but it gets a combination of r trajectories out of M that are really close to the highest spread ones and with low computational demand. This approach, which has been programmed in Matlab, consists in different iterative steps that are shown in Figure 3, where the r trajectories are selected from a group of all the possible combinations. As Figure 3 shows, firstly the distance matrix, DM, between the initial M trajectories (see Eq.2) is calculated.

$$DM = \begin{bmatrix} 0 & d_{12} & ..... & d_{1M} \\ d_{21} & 0 & ..... & d_{2M} \\ ..... & ..... & ..... & ..... \\ d_{M1} & d_{M2} & .... & 0 \end{bmatrix} \quad (2)$$

Each row of this matrix represents all the geometric distances, d_{ml}, between the trajectory corresponding to the number of the row, m, and the number of the column, l. Then, iteratively from i = 1 to i = r-1 the following procedure is carried out:

1. From each row of the DM matrix, m, the i columns whose d_{ml} are the highest ones are selected, [n_1,n_2....n_i]. Then, the quantity $D_{n_1,n_2...n_i,m}$ is calculated for each row of the DM matrix, considering the i trajectories selected in each row. Thus, M values of D are obtained, generating the matrix D_{i+1} defined in the following expression:

$$D_{i+1} = \begin{bmatrix} D_{n_1,n_2....n_i,1} \\ D_{n_1,n_2....n_i,2} \\ .... \\ D_{n_1,n_2....n_i,M} \end{bmatrix} \quad (3)$$

where the sub index i+1 corresponds to the total number of trajectories considered. Then, the maximum value of D_{i+1} is selected, which represents the highest spread i+1 trajectories of these matrix: $n_{H_1}, n_{H_2}...n_{H_{i+1}}$

2. The next step is the selection of the r-(i+1) trajectories. Subsequently, iteratively, for k=1 to k= r-i, the matrix D_{i+k+1} defined in the following expression is calculated:

$$D_{i+k+1} = \begin{bmatrix} D_{n_{H_1},n_{H_2}....n_{H_{i+k}},1} \\ D_{n_{H_1},n_{H_2}....n_{H_{i+k}},2} \\ .... \\ D_{n_{H_1},n_{H_2}....n_{H_{i+k}},M} \end{bmatrix} \quad (4)$$

where each row represents the quantity D for the trajectories: $n_{H_1}, n_{H_2}...n_{H_{i+k}}, m$. Then, the maximum value of D_{i+k+1} is obtained, and the corresponding i+k+1 trajectory is selected.

3. At last, for the specific i considered, a combination of r trajectories out of M is obtained: $n_{H_1}, n_{H_2}.n_{H_{r_i}}$.

Once the r-1 iterations are executed, r-1 combinations of r trajectories will be generated, which are represented in the matrix D_r defined by the following expression:

$$D_r = \begin{bmatrix} D_{n_{H_1},n_{H_2}....n_{r_1}} \\ D_{n_{H_1},n_{H_2}....n_{r_2}} \\ .... \\ D_{n_{H_1},n_{H_2}....n_{r_{r-1}}} \end{bmatrix} \quad (5)$$

Finally, the maximum value of D_r will define the final r trajectories selected, $D_{n_{H_1},n_{H_2}....n_{r_i}}$.

As an example, Figure 4 compares the empirical distributions (sampled values) for k=4 parameters that are obtained respectively via the random sampling strategy proposed by Morris (1991) (Figure 4a), via the revised sampling strategy proposed by Campolongo et al. (2007) (Figure 4b), and via the sampling strategy proposed in this work (Figure 4c). The theoretical distributions in this example are the same as proposed in Campolongo et al. (2007), i.e. discrete uniform with p=4 levels and for r = 20 trajectories. The empirical distribution illustrated in Figure 4c is not as closer to the theoretical shape of the discrete uniform distributions as the one obtained in Figure 4b, but it improves significantly the results compared to the ones obtained with the random sampling strategy (Figure 4a). The sampling strategy proposed in this work reduces considerably the computational demand required to select the optimal r trajectories out of M. For instance, for M = 100 initial trajectories and r = 5, the computational cost of the rigorous method was 8 minutes whilst the method proposed in this work gave the results in less than 2 seconds (obtained with Matlab® using a PC with a 2.53 GHz Intel® Core™ i5 processor). This computational demand depends mainly on the number of trajectories to be selected (r) from the initial ones (M). For instance, for selecting r = 20 trajectories out of M= 1000, the computational cost of the proposed strategy is about 5 minutes whilst applying the rigorous approach is unfeasible due to the high number of combinations to be evaluated (almost $3.4 \cdot 10^{41}$). Although the proposed sampling strategy does not guaranty that the final trajectories (r) selected present the global maximum distances between them, these distances are at least locally maximized. For instance, for the simple analytical example abovementioned (r = 5 trajectories out of M = 100 with k= 4 parameters and a grid of p = 4) the distance between the final r trajectories obtained with the methodology proposed in this work was D = 95.6 and with the rigorous approach was D = 97.6 (the global maximum distance). This difference in terms of distance can be acceptable enough, provided that when the random sampling strategy was repeated 10 times the highest distance obtained was D = 70.4.

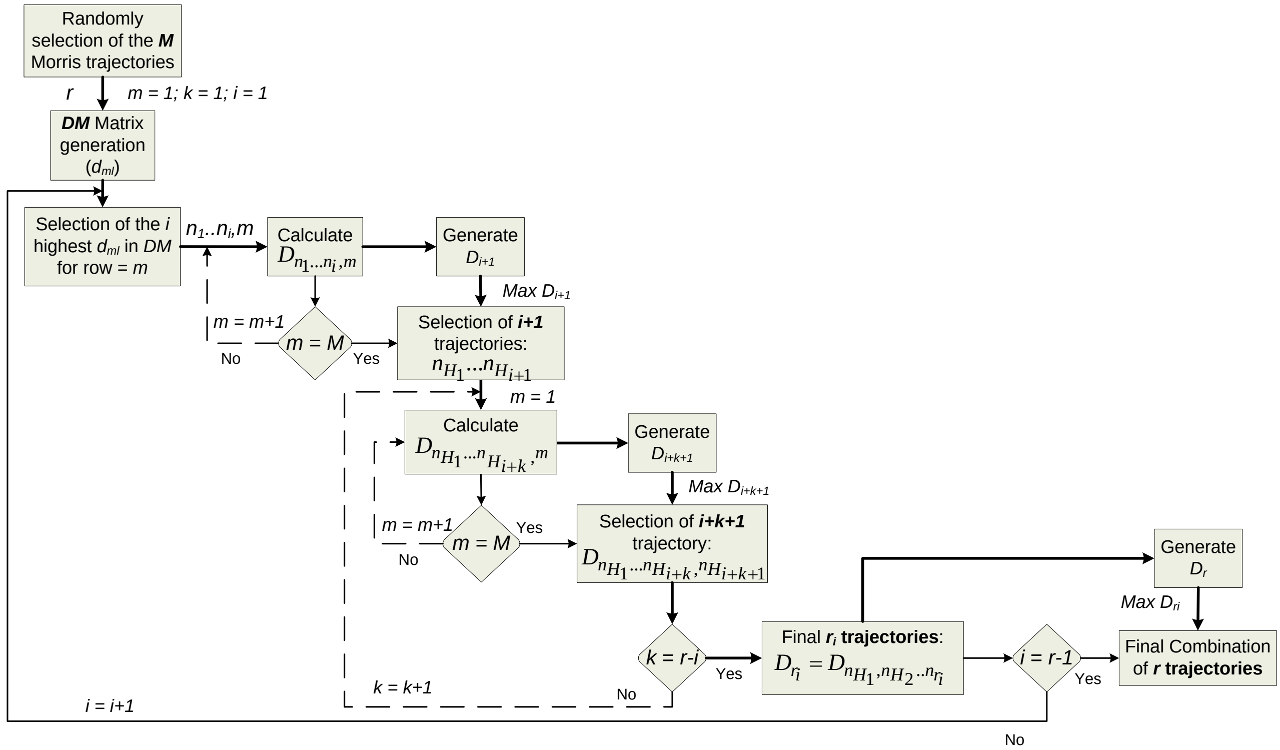


Figure 3. Flow scheme of the proposed methodology to find a high spread r trajectories out of M.

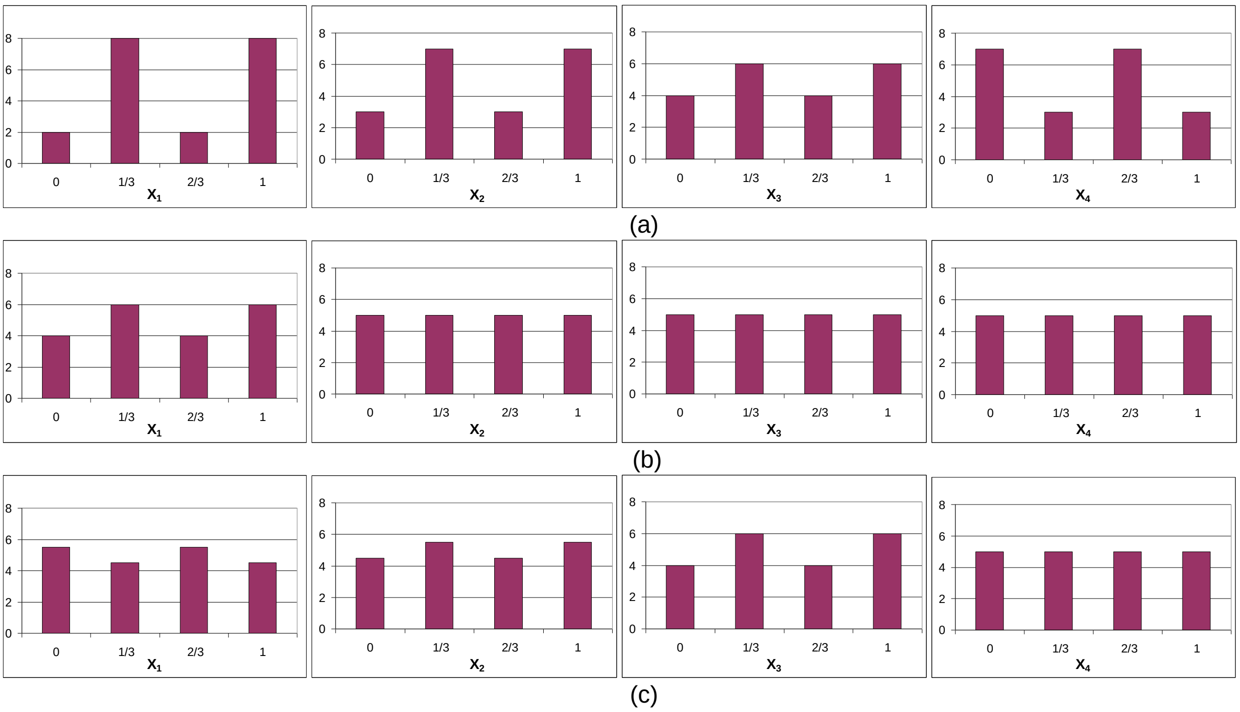


Figure 4. Empirical distributions for k=4 parameters, X_1, X_2, X_3, X_4 whose theoretical distributions are uniform discrete with 4 levels, sample size r=20. The samples are obtained using the random sampling strategy (a); the revised sampling strategy proposed by Campolongo et al. (2007) (b); and the sampling strategy proposed in this work (c).

With regard to the sensitivity measures, the mean (μ), the standard deviation (σ) and the absolute mean (μ^*) of the SEE_i values of each F_i are considered (Saltelli *et al*., 2004). The measure μ^* has been used to rank the parameters, in order to identify systematically the non-influential parameters (low μ^*) from the influential ones (high μ^*). An optimal setting of r (r_{opt}) has been searched with a constant resolution of p=8. To this end, the number of repetitions of elementary effects calculations (r) of each distribution F_i was increased until the ranking of parameters (based on μ^*) remains more or less stable, i.e. the type II error is minimised (type II error: indentifying an important factor as insignificant). This stability has been numerically evaluated with a numerical index proposed in this work (the position factor, $PF_{ri \rightarrow rj}$). For given rankings obtained by r_i and r_j, we define the index $PF_{ri \rightarrow rj}$ by the following expression:

$$PF_{r_i \rightarrow r_j} = \sum_{k=1}^{k} \frac{P_{k,i} - P_{k,j}}{\mu_{P_{k,i},P_{k,j}}} \quad (6)$$

where $P_{k,i}$ is the position of the kth parameter in the ranking obtained by r_i, and $\mu_{P_{k,i},P_{k,j}}$ is the average of the positions of the kth parameter in the ranking obtained by r_i and r_j The index $PF_{ri \rightarrow rj}$ indicates how different are the rankings calculated by sampling sizes r_i and r_j, i.e. a low value means that most of the parameters remains in the same or near position in the ranking (If $PF_{ri \rightarrow rj} = 0$ then Ranking r_i = Ranking r_j). Moreover, this index decreases the importance of a change in the position of parameters located at the bottom of the ranking. This criterion allows an optimal value for r (r_{opt}) to be found. Once the r_{opt} was found, the graphical Morris approach was used to find the significant parameters. In order to evaluate the effect of this improved sampling strategy, the Morris approach based on a random sampling strategy has also been applied and the results obtained with both sampling strategies have been compared.

3. Results and Discussion

The Morris method with the proposed improved sampling strategy was applied to different number of trajectories (r), chosen from M=1000 initial Morris trajectories, until the parameter significant ranking remained more or less stable, quantitatively measured by the index $PF_{ri \rightarrow rj}$. Similarly, the Morris approach based on the random sampling was also applied. Table 1 shows the resulting sensitivity measures (μ^* and σ) for the different number of elementary effects calculated with the improved sampling strategy. Table 2 shows the resulting $PF_{ri \rightarrow rj}$ values for each pair of compared rankings: (a) the improved sampling strategy, and (b) the random sampling strategy. As can be seen on this table, at low number of trajectories for the improved sampling strategy (from r= 5 to r =40) the tendency of the $PF_{ri \rightarrow rj}$ value is not monotonic. This behaviour reveals that, for this case study, values of r below 40 do not provide a suitable estimation of the sensitivity measures, which can be obtained when either a highly nonlinear model is used or a large input uncertainty is defined. These results are in contrast to previous applications of the Morris method since most of these studies used a low repetition number, e.g. r=(10~20) (Campolongo *et al*, 2007, Ruano *et al*., 2010b). Then, for higher values of r, the index manifests a downward trend as the number of trajectories is increased, which demonstrates a closer similarity between the positions of the parameters in the compared rankings. In contrast, the $PF_{ri \rightarrow rj}$ values for the random sampling strategy decreases as the number of trajectories is increased, except for r = 40→50 and for r = 60→70. Surprisingly, working with considerable high values of r, an increase of the number of trajectories from r=60 to r=70 does not imply an improvement in the $PF_{ri \rightarrow rj}$ index. Moreover, the increase of this index is mainly due to a position change of the parameters located at the top of the ranking (data not shown) such as: *PEcHN* (*High Negative membership function centre of the Pressure Error*). These results can be interpreted as a result of a non-optimal coverage of the sampling space obtained by a random strategy, which could lead to Type II error, i.e. failing in the identification of a parameter of considerable influence in the model; and Type I Error, as well, i.e. considering a factor as significant when it is not. Thus, a suitable scan of the input space, such as the improved sampling strategy applied in this work, can lead to more realistic results.

Table 1. Sensitivity measures of the control parameters at the different *r* evaluated with the improved sampling strategy.

***r* = 5**			***r* = 10**			***r* = 15**			***r* = 30**		
Parameter	μ^*	σ	**Parameter**	μ^*	σ	**Parameter**	μ^*	σ	**Parameter**	μ^*	σ
PEaHN	2.401	3.356	*PEcHN*	3.101	3.895	*APEaHN*	2.451	3.557	*APEaHN*	2.141	3.246
APEcHN	1.720	2.522	*IVcHN*	1.505	2.030	*IBcHN*	1.682	2.889	*IBcHN*	1.675	2.449
IVcHN	1.668	2.643	*PEaHN*	1.434	2.538	*PEcHN*	1.586	3.049	*PEcHN*	1.579	2.835
IBcHN	1.436	2.364	*APEcHN*	1.264	1.813	*IVcHN*	1.497	2.318	*PEaHN*	1.162	2.101
PEcHN	1.144	2.248	*IBcHN*	1.102	2.215	*IBcLN*	1.356	2.035	*PEcLN*	0.860	1.639
PEcLN	0.681	1.123	*APEaHN*	1.041	1.430	*PEaHN*	1.043	1.700	*IBcLN*	0.809	1.697
OEcHN	0.601	0.779	*RT*	0.948	1.220	*RT*	1.022	1.652	*AOEaHN*	0.774	1.927
AOEaHN	0.493	1.015	*OEaHN*	0.901	1.591	*OEcHN*	0.952	1.793	*APEcHN*	0.736	1.164
APEaHN	0.424	0.487	*IBcLN*	0.872	1.648	*PEcLN*	0.942	1.964	*IVcHN*	0.706	1.352
OEaHN	0.408	0.651	*APEcLN*	0.770	1.397	*APEcHN*	0.641	1.417	*RT*	0.684	1.504
RT	0.316	0.405	*PEcLN*	0.416	0.901	*OEaHN*	0.632	1.298	*OEcHN*	0.603	1.264
APEcLN	0.289	0.506	*OEcHN*	0.385	0.860	*IVcLN*	0.487	1.074	*OEaHN*	0.586	1.318
OEcLN	0.213	0.309	*IVcLN*	0.332	0.651	*APEcLN*	0.288	0.712	*APEcLN*	0.433	1.213
IBcLN	0.182	0.263	*AOEaHN*	0.305	0.591	*AOEcHN*	0.180	0.287	*AOEcLN*	0.274	0.905
AOEcHN	0.108	0.193	*OEcLN*	0.245	0.453	*AOEcLN*	0.156	0.370	*IVcLN*	0.261	0.953
AOEcLN	0.074	0.137	*AOEcLN*	0.222	0.540	*OEcLN*	0.151	0.275	*OEcLN*	0.237	0.621
IVcLN	0.019	0.028	*AOEcHN*	0.146	0.309	*AOEaHN*	0.149	0.293	*AOEcHN*	0.174	0.326

***r* = 40**			***r* = 50**			***r* = 60**			***r* = 70**		
Parameter	μ^*	σ	**Parameter**	μ^*	σ	**Parameter**	μ^*	σ	**Parameter**	μ^*	σ
IBcHN	1.848	2.584	*APEaHN*	2.010	3.263	*APEaHN*	2.272	3.498	*APEaHN*	2.222	3.208
APEcHN	1.681	2.571	*IBcHN*	1.981	3.030	*PEcHN*	1.973	3.282	*PEcHN*	1.909	3.249
APEaHN	1.412	2.067	*PEcHN*	1.720	2.466	*IBcHN*	1.697	2.477	*IBcHN*	1.680	2.511
PEcHN	1.394	2.180	*IVcHN*	1.204	2.121	*PEaHN*	1.397	2.299	*PEaHN*	1.315	2.257
IVcHN	1.319	2.424	*PEaHN*	1.042	1.775	*IVcHN*	1.108	2.100	*APEcHN*	1.085	2.138
OEaHN	1.285	2.816	*APEcHN*	0.994	2.068	*APEcHN*	1.007	1.745	*IVcHN*	0.941	1.669
PEaHN	1.023	1.805	*PEcLN*	0.912	1.698	*OEaHN*	0.900	1.744	*PEcLN*	0.891	1.682
OEcHN	0.957	1.694	*OEcHN*	0.771	1.483	*RT*	0.818	1.699	*RT*	0.861	1.604
PEcLN	0.798	1.628	*RT*	0.763	1.445	*IBcLN*	0.816	1.703	*IBcLN*	0.761	1.486
IBcLN	0.752	1.649	*OEaHN*	0.744	1.317	*OEcHN*	0.779	1.453	*OEcHN*	0.721	1.425
AOEaHN	0.493	1.210	*IBcLN*	0.712	1.350	*PEcLN*	0.741	1.403	*OEaHN*	0.688	1.356
RT	0.431	0.713	*AOEaHN*	0.349	0.859	*AOEcHN*	0.432	1.300	*APEcLN*	0.301	0.880
OEcLN	0.341	0.887	*IVcLN*	0.308	1.452	*AOEaHN*	0.393	1.165	*AOEaHN*	0.282	0.741
AOEcHN	0.335	1.204	*APEcLN*	0.272	0.723	*AOEcLN*	0.312	1.273	*AOEcHN*	0.270	0.792
APEcLN	0.218	0.726	*OEcLN*	0.128	0.299	*APEcLN*	0.295	1.164	*IVcLN*	0.170	0.724
AOEcLN	0.161	0.541	*AOEcHN*	0.122	0.367	*IVcLN*	0.132	0.517	*OEcLN*	0.164	0.561
IVcLN	0.083	0.222	*AOEcLN*	0.039	0.090	*OEcLN*	0.121	0.294	*AOEcLN*	0.123	0.393

Table 2. Position factors, $PF_{ri \rightarrow rj}$, for the *r* calculated: (a) improved sampling strategy; (b) random sampling strategy.

$r_i \rightarrow r_j$	5→10	10→15	15→30	30→40	40→50	50→60	60→70
(a) $PF_{ri \rightarrow ri}$	7.5	7.8	4.2	7.8	5.4	3.5	1.9
(b) $PF_{ri \rightarrow ri}$	10.1	7.6	5.2	2.1	2.7	2.3	3.7

As a result, *r*=70 was selected as the optimal setting for this case study. The overall model evaluation cost was, therefore, 1190 simulations. We considered *r*=70 as the optimal one, not only due to the low $PF_{60 \rightarrow 70}$ value but also due to the significant stability of the parameters located at the top of the rankings (see Table 1). Figure 5a shows the graphical Morris approach for the optimal number of repetitions obtained for the improved sampling strategy. In addition, Figure 5b shows the same graph for r=70 obtained with the random sampling strategy. This figure was used to screen out the non-influential parameters of the control system (i.e., the six parameters that are not labelled in Figure 1a). From the eleven influential parameters, *RT* (Response Time), *OEcHN* (*High Negative centre of the Oxygen Error*), *APEaHN* (*High Negative amplitude of the Accumulated Pressure Error*), *IVcHN* (*High Negative centre of the Increment of the air Valve opening*) and *IBcLN* (*Low Negative centre of the Increment of the rotational speed of the Blower*) presented high mean and low standard deviation, laying outside of the wedge formed by two lines corresponding to $\mu_i=\pm 2SEM_i$. Thus, the effect of these parameters on the output variables are expected to be linear and additive, which is desirable for parameter estimation based on optimisation algorithms. Compared to the results obtained for the random sampling strategy, the following could be said: (i) the resulting non-influential parameters agreed with the improved sampling strategy; (ii) the sensitivity measures of the eleven influential parameters are different from the ones obtained with the improved sampling strategy, which reflects that the sampling strategy has a significant effect on the parameter significance ranking. In addition, the necessity of finding out the optimal repetition number for SEE_i calculations (r) has been underlined. Thus, a non-optimal selection of r would lead to Type II error and Type I Error, as well.

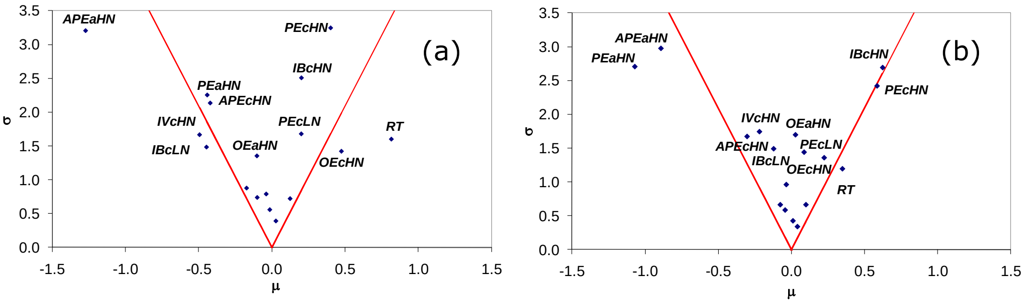


Figure 5. (a) μ versus σ, for r_{opt}= 70, for the improved sampling; (b) μ versus σ, for r=70, for the random sampling. Lines correspond to $\mu_i=\pm 2SEM_i$;

4. Conclusions

The Morris method with the improved sampling strategy proposed by Campolongo *et al*. (2007) has been applied to a fuzzy logic based control system of a WWTP. A systematic approach has been proposed to be able to apply this improved sampling strategy based on trajectory design to large models with low computational demand. In order to find out the proper repetition number of SEE_i calculations (r_{opt}), an iterative and automatic procedure has been applied. The optimal repetition number found in this study is in direct contrast with previous applications of Morris method, which usually uses low number of repetition, e.g. r=(10~20). This high r value can be explained by either a highly nonlinear behaviour of the system, or a large input uncertainty defined. The results show that the sampling strategy has a significant effect on the parameter significance ranking and that the random sampling strategy could lead to a non-proper coverage of the sample space. Working with a non-proper sampling matrix and a non-proper sample size (r) could lead to either Type I or Type II errors. Overall, the improved sampling strategy proposed and the iterative and automatic procedure to find out the proper repetition number of SEE_i calculations to apply the Morris approach, provides a good approximation of a global sensitivity measure, helping engineers to calibrate large models with many input factors such as the fuzzy control system used in this work.

Acknowledgments

This research work has been supported by the Spanish Research Foundation (CICYT Project reference CTM2005-06919-C03-01/TECNO), which is gratefully acknowledged.

References

Brun, R., Kühni, M., Siegrist, H., Gujer, W., Reichert, P., 2002. Practical identifiability of ASM2d parameters-systematic selection and tuning of parameter subsets. Water Res. 36(16), 4113-4127.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environ. Mod. Soft. 22 (10), 1509-1518.

Campolongo, F., Saltelli, A., Cariboni, J, 2011. From screening to quantitative sensitivity analysis. A unified approach. Comput. Phys. Commun. 182, 978-988.

Chanona, J., Pastor, L., Borras, L., Seco, A., 2006. Fuzzy algorithm for pH control in a struvite crystallization reactor. Wat. Sci. Tech. 53(12), 161-168.

Copp J., 2002. The COST Simulation Benchmark – Description and Simulator Manual. COST (European Cooperation in the field of Scientific and Technical Research), Brussels, Belgium.

Ferrer, J., Rodrigo, M.A., Seco, A., Penya-Roja, J.M., 1998. Energy saving in the aeration process by fuzzy logic control. Wat. Sci. Tech. 38(3), 209-217.

Ferrer, J., Seco, A., Serralta, J., Ribes, J., Manga, J., Asensi, E., Morenilla, J.J., Llavador F., 2008. DESASS: a software tool for designing, simulating and optimising WWTPs. Environ. Mod. Soft. 23, 19-26.

Manesis, S.A., Sapidis, D.J., King R.E., 1998. Intelligent control of wastewater treatment plants. Artif. Intell. Eng. 12 (3), 275-281.

Morris, M. (1991). Factorial sampling plans for preliminary computational experiments. Technometrics. 33(2), 239-245.

Ribes, J., Serralta, J., Seco, A., Ferrer, J., Lloret, R., Bernácer, I., Morenilla J.J., 2007. Implementation of a supervisory fuzzy logic based control system in the Denia WWTP. In: 10th IWA Conference on Design, Operation and Economics of Large Wastewater Treatment Plants. 77-80. Vienna.

Ruano, M.V., Ribes, J., Sin, G., Seco, A., Ferrer J., 2010a. A systematic approach for fine-tuning of fuzzy controllers applied to WWTPs. Environ. Modell. Softw. 25(5), 670-676.

Ruano, M.V., Ribes, J., Ferrer, J., Sin G., 2010b. Application of the Morris method for screening the influential parameters of fuzzy controllers applied to WWTPs. Wat. Scien. Tech. (In press).

Saltelli, A, Scott, EM, Chen, K, editors, 2000. Sensitivity analysis. Chichester: Wiley.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity analysis in practice: a guide to assessing scientific models. Chicester: Wiley.

Saltelli, A., Annoni, P., Azzini. I., Campolongo, F., Ratto M., Tarantola, S, 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181, 259-270.

Seco, A., Ribes, J., Serralta, J., Ferrer J., 2004. Biological Nutrient Removal Model No.1. Water Sci. Technol. 50(6), 69-78.

Serralta, J., Ribes, J., Seco, A., Ferrer, J., 2002. A supervisory control system for optimising nitrogen removal and aeration energy consumption in wastewater treatment plants. Wat. Sci. Tech. 45(4/5), 309-316.

Sin, G., Gernaey, KV., 2009. Improving the Morris method for sensitivity analysis by scaling the elementary effects. In proceedings of the 19th European Symposium on Computer Aided Process Engineering (ESCAPE 19), eds. J. Jezowski and J. Thullie. Elsevier, pp 925-930.

Traoré, A., Grieu, S., Corominas, L., Thiery, F., Polit, M., Colprim J., 2005. Fuzzy control of dissolved oxygen in a sequencing batch reactor pilot plant. Chem. Eng. Jour. 111(1), 13–19.

Yong, M., Yong-zhen, P., Xiao-lian, W., Shu-ying W., 2006. Intelligent control aeration and external carbon addition for improving nitrogen removal. Environ. Modell. Softw. 21, 821-828.

Weijers, S.R., Vanrolleghem, P.A., 1997. A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data. Wat. Sci. Tech. 36(5), 69–79.

Zhu G., Peng Y., Ma B., Wang Y., Yin C., 2009. Optimization of anoxic/oxic step feeding activated sludge process with fuzzy control model for improving nitrogen removal. Chem. Eng. J. 151(1-3), 195-201.

Figure(s)

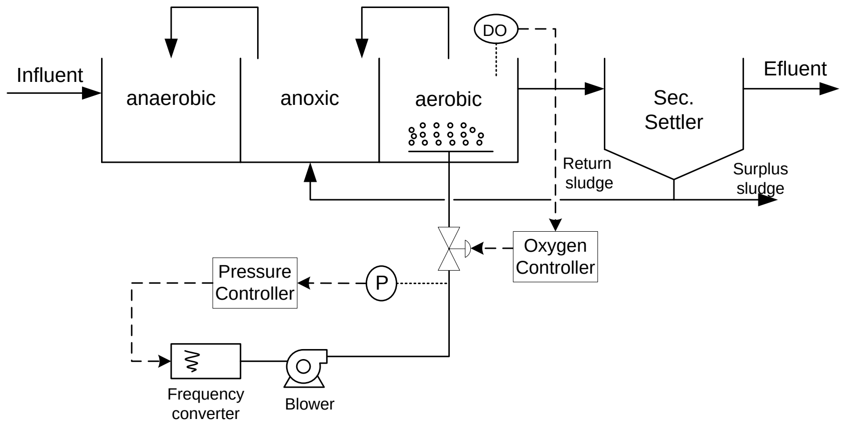
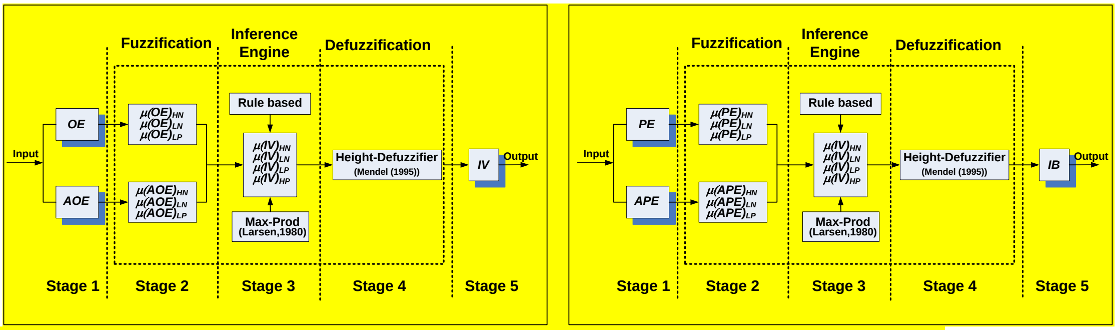


Figure 1. Flow diagram of the control system applied to a modified UCT process.



a) DO controller b) Air pressure controller

Figure 2. Fuzzy control stages for the two controllers: (a) dissolved oxygen controller; and (b) air pressure controller.

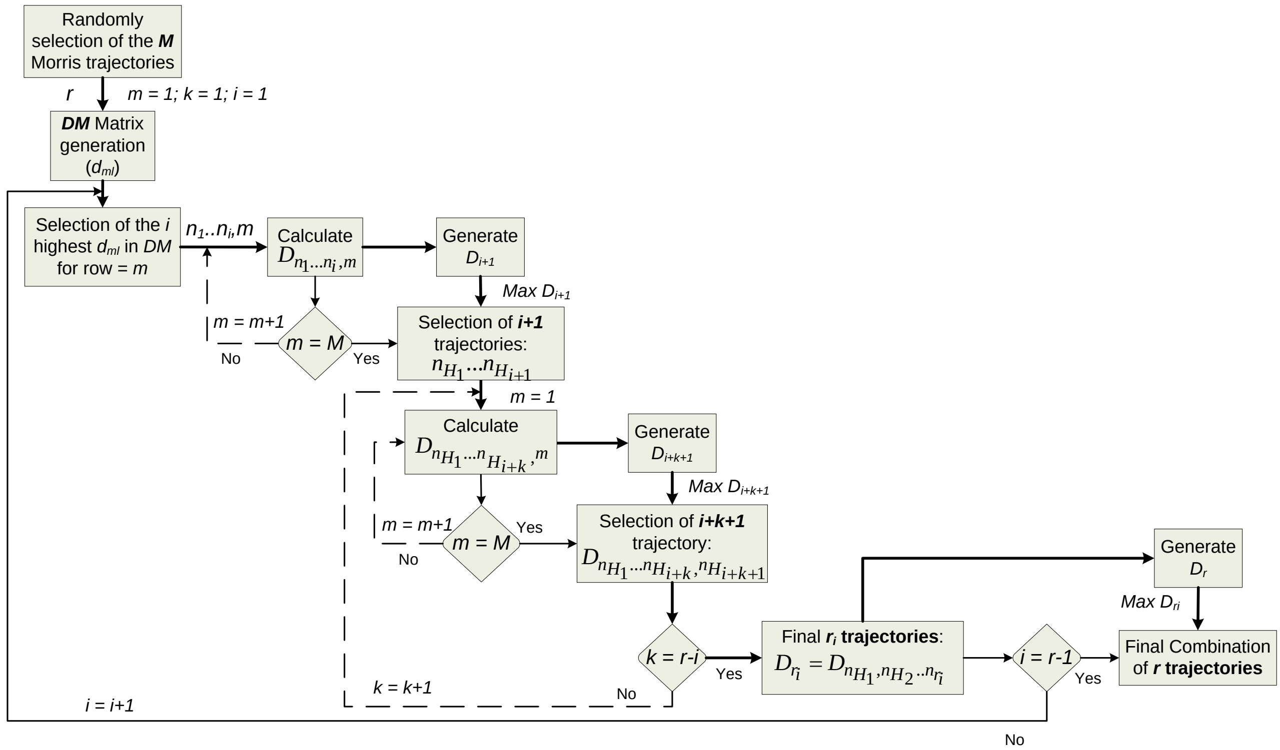


Figure 3. Flow scheme of the proposed methodology to find a high spread r trajectories out of M.

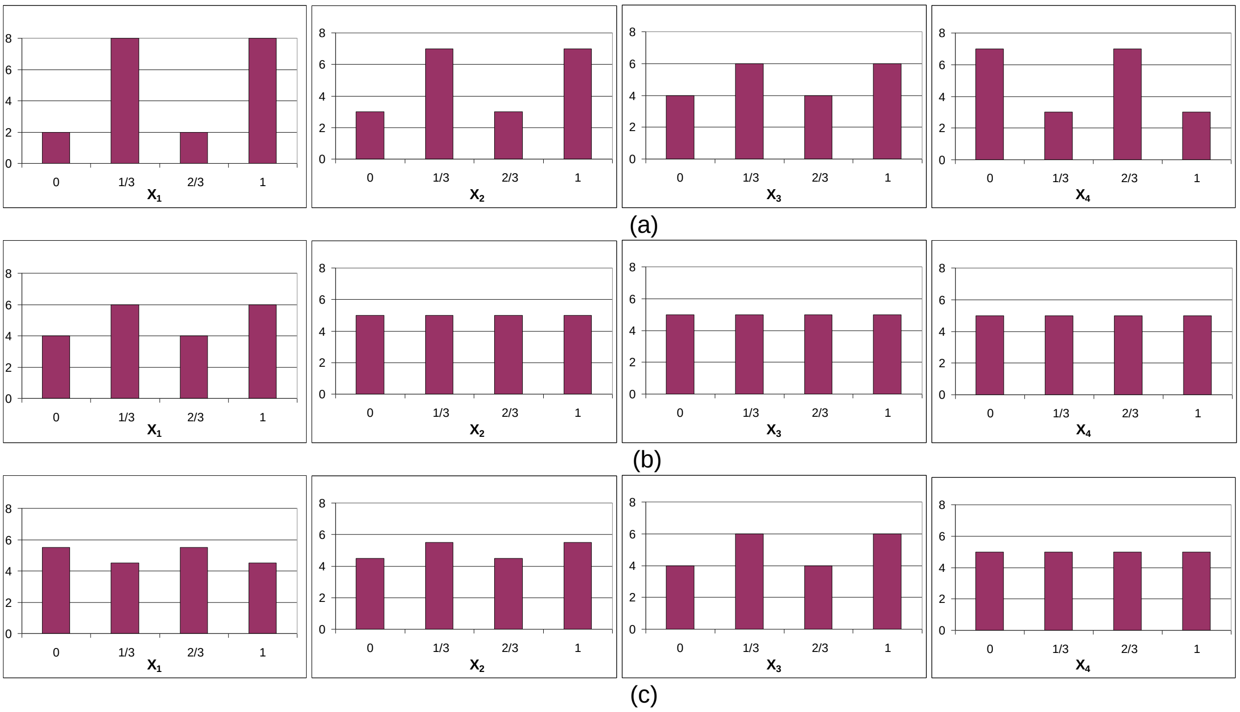


Figure 4. Empirical distributions for k=4 parameters, X_1, X_2, X_3, X_4 whose theoretical distributions are uniform discrete with 4 levels, sample size r=20. The samples are obtained using the random sampling strategy (a); the revised sampling strategy proposed by Campolongo et al. (2007) (b); and the sampling strategy proposed in this work (c).

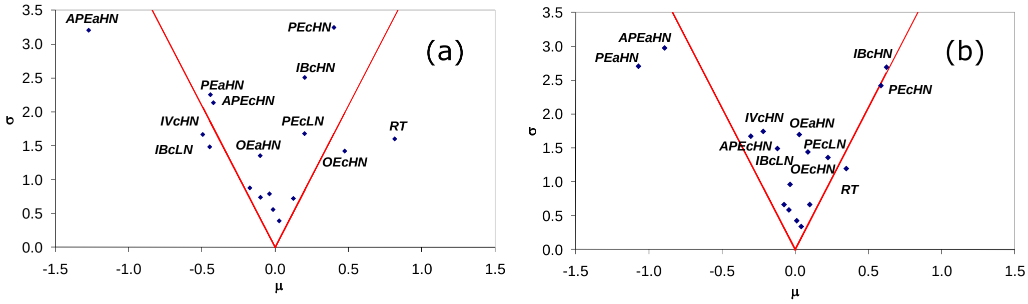


Figure 5. (a) μ versus σ, for r_{opt}= 70, for the improved sampling; (b) μ versus σ, for r=70, for the random sampling. Lines correspond to $\mu_i=\pm 2SEM_i$;

Table(s)

Table 1. Sensitivity measures of the control parameters at the different *r* evaluated with the improved sampling strategy.

***r* = 5**			***r* = 10**			***r* = 15**			***r* = 30**		
Parameter	**μ^***	**σ**	**Parameter**	**μ^***	**σ**	**Parameter**	**μ^***	**σ**	**Parameter**	**μ^***	**σ**
PEaHN	2.401	3.356	*PEcHN*	3.101	3.895	*APEaHN*	2.451	3.557	*APEaHN*	2.141	3.246
APEcHN	1.720	2.522	*IVcHN*	1.505	2.030	*IBcHN*	1.682	2.889	*IBcHN*	1.675	2.449
IVcHN	1.668	2.643	*PEaHN*	1.434	2.538	*PEcHN*	1.586	3.049	*PEcHN*	1.579	2.835
IBcHN	1.436	2.364	*APEcHN*	1.264	1.813	*IVcHN*	1.497	2.318	*PEaHN*	1.162	2.101
PEcHN	1.144	2.248	*IBcHN*	1.102	2.215	*IBcLN*	1.356	2.035	*PEcLN*	0.860	1.639
PEcLN	0.681	1.123	*APEaHN*	1.041	1.430	*PEaHN*	1.043	1.700	*IBcLN*	0.809	1.697
OEcHN	0.601	0.779	*RT*	0.948	1.220	*RT*	1.022	1.652	*AOEaHN*	0.774	1.927
AOEaHN	0.493	1.015	*OEaHN*	0.901	1.591	*OEcHN*	0.952	1.793	*APEcHN*	0.736	1.164
APEaHN	0.424	0.487	*IBcLN*	0.872	1.648	*PEcLN*	0.942	1.964	*IVcHN*	0.706	1.352
OEaHN	0.408	0.651	*APEcLN*	0.770	1.397	*APEcHN*	0.641	1.417	*RT*	0.684	1.504
RT	0.316	0.405	*PEcLN*	0.416	0.901	*OEaHN*	0.632	1.298	*OEcHN*	0.603	1.264
APEcLN	0.289	0.506	*OEcHN*	0.385	0.860	*IVcLN*	0.487	1.074	*OEaHN*	0.586	1.318
OEcLN	0.213	0.309	*IVcLN*	0.332	0.651	*APEcLN*	0.288	0.712	*APEcLN*	0.433	1.213
IBcLN	0.182	0.263	*AOEaHN*	0.305	0.591	*AOEcHN*	0.180	0.287	*AOEcLN*	0.274	0.905
AOEcHN	0.108	0.193	*OEcLN*	0.245	0.453	*AOEcLN*	0.156	0.370	*IVcLN*	0.261	0.953
AOEcLN	0.074	0.137	*AOEcLN*	0.222	0.540	*OEcLN*	0.151	0.275	*OEcLN*	0.237	0.621
IVcLN	0.019	0.028	*AOEcHN*	0.146	0.309	*AOEaHN*	0.149	0.293	*AOEcHN*	0.174	0.326

***r* = 40**			***r* = 50**			***r* = 60**			***r* = 70**		
Parameter	**μ^***	**σ**	**Parameter**	**μ^***	**σ**	**Parameter**	**μ^***	**σ**	**Parameter**	**μ^***	**σ**
IBcHN	1.848	2.584	*APEaHN*	2.010	3.263	*APEaHN*	2.272	3.498	*APEaHN*	2.222	3.208
APEcHN	1.681	2.571	*IBcHN*	1.981	3.030	*PEcHN*	1.973	3.282	*PEcHN*	1.909	3.249
APEaHN	1.412	2.067	*PEcHN*	1.720	2.466	*IBcHN*	1.697	2.477	*IBcHN*	1.680	2.511
PEcHN	1.394	2.180	*IVcHN*	1.204	2.121	*PEaHN*	1.397	2.299	*PEaHN*	1.315	2.257
IVcHN	1.319	2.424	*PEaHN*	1.042	1.775	*IVcHN*	1.108	2.100	*APEcHN*	1.085	2.138
OEaHN	1.285	2.816	*APEcHN*	0.994	2.068	*APEcHN*	1.007	1.745	*IVcHN*	0.941	1.669
PEaHN	1.023	1.805	*PEcLN*	0.912	1.698	*OEaHN*	0.900	1.744	*PEcLN*	0.891	1.682
OEcHN	0.957	1.694	*OEcHN*	0.771	1.483	*RT*	0.818	1.699	*RT*	0.861	1.604
PEcLN	0.798	1.628	*RT*	0.763	1.445	*IBcLN*	0.816	1.703	*IBcLN*	0.761	1.486
IBcLN	0.752	1.649	*OEaHN*	0.744	1.317	*OEcHN*	0.779	1.453	*OEcHN*	0.721	1.425
AOEaHN	0.493	1.210	*IBcLN*	0.712	1.350	*PEcLN*	0.741	1.403	*OEaHN*	0.688	1.356
RT	0.431	0.713	*AOEaHN*	0.349	0.859	*AOEcHN*	0.432	1.300	*APEcLN*	0.301	0.880
OEcLN	0.341	0.887	*IVcLN*	0.308	1.452	*AOEaHN*	0.393	1.165	*AOEaHN*	0.282	0.741
AOEcHN	0.335	1.204	*APEcLN*	0.272	0.723	*AOEcLN*	0.312	1.273	*AOEcHN*	0.270	0.792
APEcLN	0.218	0.726	*OEcLN*	0.128	0.299	*APEcLN*	0.295	1.164	*IVcLN*	0.170	0.724
AOEcLN	0.161	0.541	*AOEcHN*	0.122	0.367	*IVcLN*	0.132	0.517	*OEcLN*	0.164	0.561
IVcLN	0.083	0.222	*AOEcLN*	0.039	0.090	*OEcLN*	0.121	0.294	*AOEcLN*	0.123	0.393

Table 2. Position factors, $PF_{ri \rightarrow rj}$, for the r calculated: (a) improved sampling strategy; (b) random sampling strategy.

$r_i \rightarrow r_j$	5→10	10→15	15→30	30→40	40→50	50→60	60→70
(a) $PF_{ri \rightarrow rj}$	7.5	7.8	4.2	7.8	5.4	3.5	1.9
(b) $PF_{ri \rightarrow rj}$	10.1	7.6	5.2	2.1	2.7	2.3	3.7