



**HAL**  
open science

# Low dimensional modeling of Double T- junctions in water distribution networks using Kriging interpolation and Delaunay triangulation

Denis Gilbert, Iraj Mortazavi, Olivier Piller, Hervé Ung

► **To cite this version:**

Denis Gilbert, Iraj Mortazavi, Olivier Piller, Hervé Ung. Low dimensional modeling of Double T- junctions in water distribution networks using Kriging interpolation and Delaunay triangulation. Pacific Journal of Mathematics for Industry, 2017, 9 (2), 19 p. 10.1186/s40736-016-0026-8 . hal-01542293

**HAL Id: hal-01542293**

**<https://hal.science/hal-01542293v1>**

Submitted on 19 Jun 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ORIGINAL ARTICLE

Open Access



# Low dimensional modeling of Double T-junctions in water distribution networks using Kriging interpolation and Delaunay triangulation

Denis Gilbert<sup>1</sup>, Iraj Mortazavi<sup>2\*</sup>, Olivier Piller<sup>1</sup> and Hervé Ung<sup>1</sup>

## Abstract

Water distribution networks are subject to potential intentional contaminations to cause harm to the consumer. Reliable transport models are needed to detect, trace and follow any contaminant transported inside the network. For now, the transport of contaminants in pipes has been mostly modeled assuming perfect mixing conditions at T-junction. However, some studies have shown that it is not always the case when crosses or double T-junctions are involved. In this paper an imperfect mixing at Double T-junction model is developed considering 3-D mixing behavior. A reduced model is then constructed in the form of a 1-D law to apply it to current 1-D transport models for water distribution networks. The methodology to create such law is detailed and can be applied to any reduced model problem including multi parameters and time consuming simulations. The procedure is composed of three steps: first calibrate the Kriging interpolation method parameters; then couple it with the Delaunay triangulation method to select simulation points with maximum gain; and finally implement a 1-D law based on the simulation results and the interpolation.

**Keywords:** Water distribution network, Imperfect mixing, Transport modeling, CFD, Laminar, Turbulent, Design of experiment, 1-D reduced model, Kriging, Delaunay triangulation

## 1 Introduction

Security is an important topic for Water Distribution Networks (WDN), as water is an essential component of society. Real-time sensors are now installed in water network to monitor and better understand the behavior of water quality through the network. Improvements in technology require improvements in the precision of water quality modeling. This work has the dual objectives to develop and enhance methods to detect any contamination in the network and to get reliable real-time transport models for source contamination identification and proper countermeasures.

Water distribution systems are modeled by a graph composed of links transporting water for several pipes in series, pumps and valves, and nodes that can represent either water sources, consumers or junctions. A

hydraulic model is used to compute the water velocities in links and the head at nodes. Then a transport model is applied to simulate the quality state of the water inside the network, such as the water chlorination concentration. The transport model equations that are usually used in water distribution hydraulic models are the 1-D advection-reaction equation and a perfect mixing law. Different resolution approaches exist. Epanet software [17] uses the Lagrangian transport model with an event-driven model [3], whereas Porteau software [16] resolves equations with an Eulerian approach and a time-driven model [9, 12]. Both are coupling a 1-D transport model for the water quality inside the links and a 0-D model for mixing at nodes.

Two imperfect mixing models exist in the literature. Ho and Khalsa [13] developed the Epanet BAM model for imperfect mixing at cross junction. They used both experiments and simulations to deduce an imperfect mixing law in between no mixing and imperfect mixing results, with a

\* Correspondence: iraj.mortazavi@cnam.fr

<sup>2</sup>M2N, IMATH, CNAM, 292 rue Saint-Martin, F-75141 Paris Cedex 03, France  
Full list of author information is available at the end of the article

scale parameter that need to be calibrated depending on each real cross. Choi et al. [7] developed the AZRED model based on an experimental data filled lookup table with interpolation or extrapolation for inputs not in the table to model both imperfect mixing at crosses and Double T-junctions. It was validated on experimental values with a large range of flows. One limitation is the need for extrapolation for Reynolds number approaching zero or infinity. Also, a 2-D resolution has been proposed in [19] to better assess the mixing at the junction by applying adjoint method and resolving Stokes equation in laminar flow (Reynolds number ( $Re$ ) < 2000).

The aim of this paper is to complement the previous models generating a generalized one-dimensional model. The approach uses a Computational Fluid Dynamics (CFD) based method [4, 18]. This study enhances the transport model with an accurate mixing model of contaminants at Double T-junctions. This is done by creating a lookup table with results from 3-D CFD simulations. These simulations are time costly, therefore an interpolating method is used to extract a global imperfect mixing behavior from this table. Two methods are needed for its creation and utilization, first a high order interpolation Kriging, and second a design algorithm based on the Delaunay triangulation. That table is finally used in a 1-D law that can be coupled to the existing transport models in real time computation.

It should be mentioned that here, the 3-D simulations are used to estimate both laminar and turbulent dynamics at Double T-junctions. They are performed solving Navier–Stokes equations with either the advection equation or mixture model. Each simulation represents the behavior of the mixing at double T-junctions at some design points depending on the 4 Double T-junction parameters: the length of the interpipe (the distance between two pipe junctions), the average interpipe Reynolds number and the input and output Reynolds number ratios (see Fig. 2). To get the solution for any other combination of the parameters' value, a Kriging interpolation is performed.

This interpolation method is used because it is suitable for cloud point distribution [20], which conforms to limited number of CFD simulations. Firstly, the Kriging interpolation needs a calibration of its parameters and is solved with a maximum likelihood function. Then the solution of the interpolation is defined as a weighted sum of the CFD simulation results with these weights found through the resolution of a linear system. An example is then given to explain its application. That interpolation is also used to define the error of interpolation that is used in the Delaunay triangulation method. This algorithm allows us to know which simulation points to do next as to minimize the interpolation error. This maximizes the gain of each new CFD

simulation, which is important as they are time consuming. The results of the CFD simulations are placed in a lookup table which is coupled to the Kriging interpolation to create a reduced model under the form of a 1-D law that can enhance current transport model reliability.

Such methodology can be used for any other reduced model problem that include multiple parameters and time consuming simulations. The two important steps in the procedure are to calibrate the Kriging interpolation parameters and to select the simulation points with the Delaunay triangulation design method.

The paper is organized as follows: we will describe first the problem equations as well as the CFD simulations procedure. Then the Kriging theory and calibration method is explained and the Delaunay triangulation design method is given. Afterwards, the generalization to the reduced 1-D model is discussed. The second part of this article is dedicated to the study of the results: first is given the Kriging calibration results on the mixing for the Double T-junction, then is explained how the Delaunay triangulation helps the design point choice and finally a discussion on CFD simulation results is given to explain the imperfect mixing behavior.

## 2 A theoretical model for mixing in WDN

In this part, first the imperfect mixing problem and equations as well as the CFD simulation parameters and procedures, are described. Then, a discussion on the design of experiment is written which explains the need for both an interpolation method and a design approach. Thereafter the Kriging interpolation is given in two parts: resolution and calibration connected to the Delaunay triangulation algorithm which is necessary to identify optimal simulation points. Finally, CFD simulation results are used to build a 1-D law that can be used in any case of flow rate and concentration for the mixing at Double T-junctions.

### 2.1 Transport model and CFD simulations

A 1-D advection-reaction equation is usually used together with the perfect mixing at junctions to simulate the propagation of a physicochemical agents inside the graph links and nodes of a water distribution network. The 1-D advection-reaction reads:

$$\begin{aligned} \frac{\partial C_i(t, x)}{\partial t} + u_i \frac{\partial C_i(t, x)}{\partial x} + K_i C_i(t, x)^{\alpha_i} \\ = 0; C_i(t, 0) = C_j \end{aligned} \quad (1)$$

With  $C_i$  is the mass concentration of the agent in link  $i$ ,  $t$  is the time,  $x$  the space,  $u_i$  the average velocity in the link  $i$ ; and  $K_i$  and  $\alpha_i$  are the kinetic constant and the order of the reaction respectively.

When the concentration is calculated in incident links, the following 0-D perfect mixing equation is used to update the concentration at the downstream node:

$$C_j = \frac{\sum_{i \in In_j} Q_i C_i}{\sum_{i \in In_j} Q_i} \quad (2)$$

Where  $In_j$  is the set of links with water entering node  $j$ ;  $Q_i$  is the flowrate of link  $i$ ; and  $C_j$  is the resulting concentration at node  $j$ .  $C_j$  will be used as boundary condition for propagating the agent in the next adjacent links see Eq. (1).

In recent studies, Ho and Khalsa [13] and Choi et al. [7], showed that mixing at T-junctions is important with regards to security. It is therefore essential to add it when calculating the hydraulic quality state of the network. Ho and Khalsa [13] developed, and adjusted with experiments, the Bulk-Advective Mixing model BAM for the imperfect mixing at cross-junction that depends on the Reynolds number at the inlets and outlets.

In the cross-junction configuration of Fig. 1, there is a principal flow 1, where the inlet goes into the two outlet directions 3 and 4 and the secondary flow where the inlet 2 is only going to the nearest outlet 3. For this configuration and within steady state condition, the pure BAM model for cross junction element implies that:

$$\begin{cases} C_4^{bulk} = C_1 \\ C_3^{bulk} = \frac{Q_2 C_2 + (Q_1 - Q_4) C_1}{Q_3} \end{cases} \quad (3)$$

with  $Q_i$  and  $C_i$  respectively the flow rate and the concentration at inlet  $i$  of Fig. 1. The concentration at outlet 4,  $C_4^{bulk}$ , is the same as inlet 1 and the concentration at outlet 3,  $C_3^{bulk}$ , is calculated from the mass balance (from

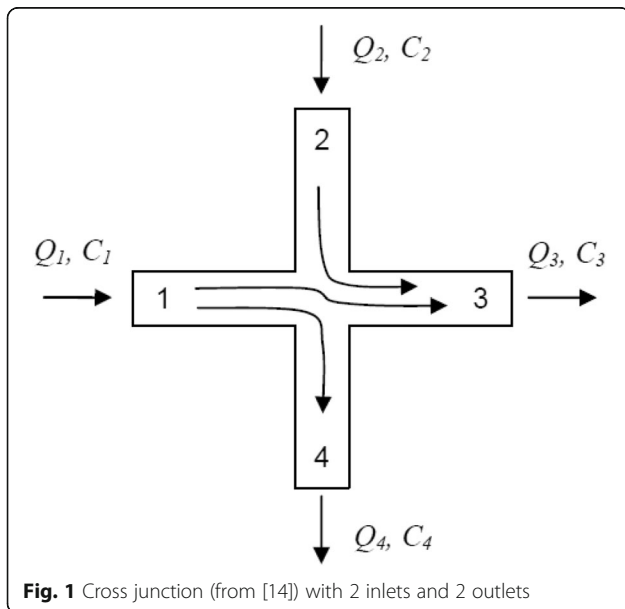


Fig. 1 Cross junction (from [14]) with 2 inlets and 2 outlets

Fig. 1, it can be seen that  $Q_4 < Q_1$ , therefore  $Q_1 - Q_4$  is positive). However, this configuration does not represent correctly the reality. The complete BAM model [13] consists of combining the pure BAM model and the perfect mixing with the following equation:

$$\begin{cases} C_j = C_j^{bulk} + s(C_j^{perfect} - C_j^{bulk}) \\ C_j^{perfect} = \frac{Q_1 C_1 + Q_2 C_2}{Q_1 + Q_2} \end{cases} \quad (4)$$

Where  $s$  is a coefficient to calibrate, which depends on the real state of the cross-junction;  $C_j$  the imperfect mixing condition concentration at either outlet  $j = 3$  or  $j = 4$  in Fig. 1;  $C_j^{perfect}$  is the concentration computed with perfect mixing hypothesis at outlets 3 and 4 using Eq. (2); and  $C_j^{bulk}$  is calculated with Eq. (3). Ho and Khalsa [13] proposed to calibrate the  $s$  coefficient from experiments. Their results for X and N-junctions (or double T-junctions) is implemented in the Epanet-BAM module [13].

Another research study, The AZRED model [7] proposes the use of a lookup table with interpolation or extrapolation for inputs not in the table to model both imperfect mixing at crosses and Double T-junctions. It was validated on experimental values with a large range of flows. One limitation is the need for extrapolation for Reynolds number approaching zero or infinity. What is proposed is to complete the two previous models into another one-dimensional model. The approach is to use a Computational Fluid Dynamics (CFD) based method with some results published in [4, 18].

A double T-junction (Fig. 2) is a special kind of junction composed of an interpipe connecting two T-junctions. In case of two inflows on one side and two outflows on the other side, the mixing may not be perfect depending on four parameters: The first one is the length of the interpipe, noted  $L$ , the longer it is the more likely the mixing will be perfect as the two inlet streams will have time to mix; the second is  $Re_{average}$ , the average Reynolds number in the interpipe, which also corresponds through the flow rate balance to the sum of both Reynolds numbers at the inlets and also at the outlets; the third one corresponds to how much flow comes from both inlets. And because the averaged Reynolds numbers can vary, it was decided to use the ratio:

$$R_1^{in} = \frac{Re_{In1}}{Re_{average}} \times 100 = \frac{Re_{In1}}{Re_{In1} + Re_{In2}} \times 100 = 100 - R_2^{in}$$

Therefore, the third parameter chosen is  $R_1^{in}$ , which represents how much flow goes in inlet 1 compared

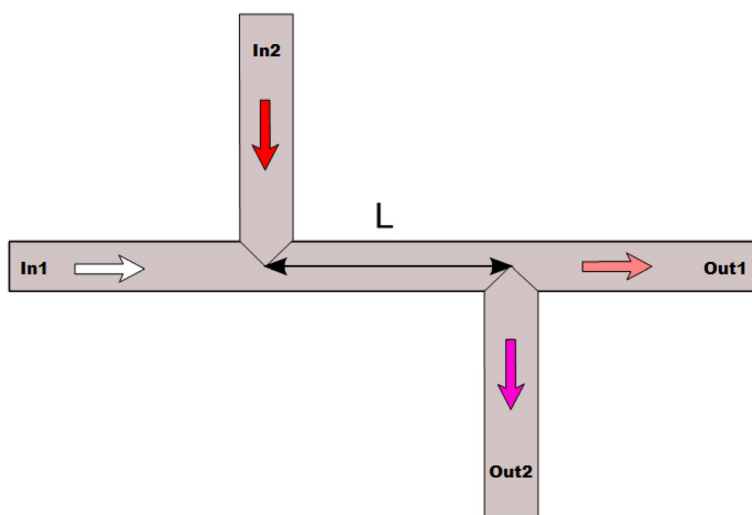


Fig. 2 Double T-junction configuration with 2 inlets and 2 outlets

to the average Reynolds number. By fixing this parameter, it also determines  $R_2^{in}$ . Finally, the fourth parameter is like the third one and corresponds to the outlets, it is noted  $R_1^{out}$ . To summarize, the four parameters which are expected to modify the behavior of the mixing in the simulations, are:  $L$ ,  $Re_{average}$ ,  $R_1^{in}$  and  $R_1^{out}$ .

It is proposed to use 3-D CFD simulations to create a 1-D law based on a lookup table and a 4-D Kriging interpolation method. Code Saturne [2] has been used to simulate laminar flows and ANSYS Fluent for turbulent flows. Some common simulations have been completed on both software for turbulent cases and have given similar results, which justifies that they can both be used to model the phenomenon. They have been computed on the Mésocentre de Calcul Intensif Aquitain (MCIA) and Irstea computation grid respectively. The Jade supercomputer of Centre Informatique National de l'Enseignement Supérieur (CINES) has been used to assure the validity of the grid convergence.

The CFD simulations that have been carried out are composed of 2 types:

- The first one is composed of the laminar flow ( $Re \leq 2000$ ) and transitional flow ( $2000 < Re < 4000$ ) cases. Direct Numerical Simulations (DNS) have been used and the transport model is the resolution of the advection equation for a concentration. In Fig. 2, the boundary conditions are 0 mg/L for the straight inlet (In1) and 1 mg/L for the other one (In2). The velocities at 3 boundary surfaces are calculated from the laminar velocity formula:

$$v(r) = 2 * vm * \left(1 - \frac{r^2}{R^2}\right)$$

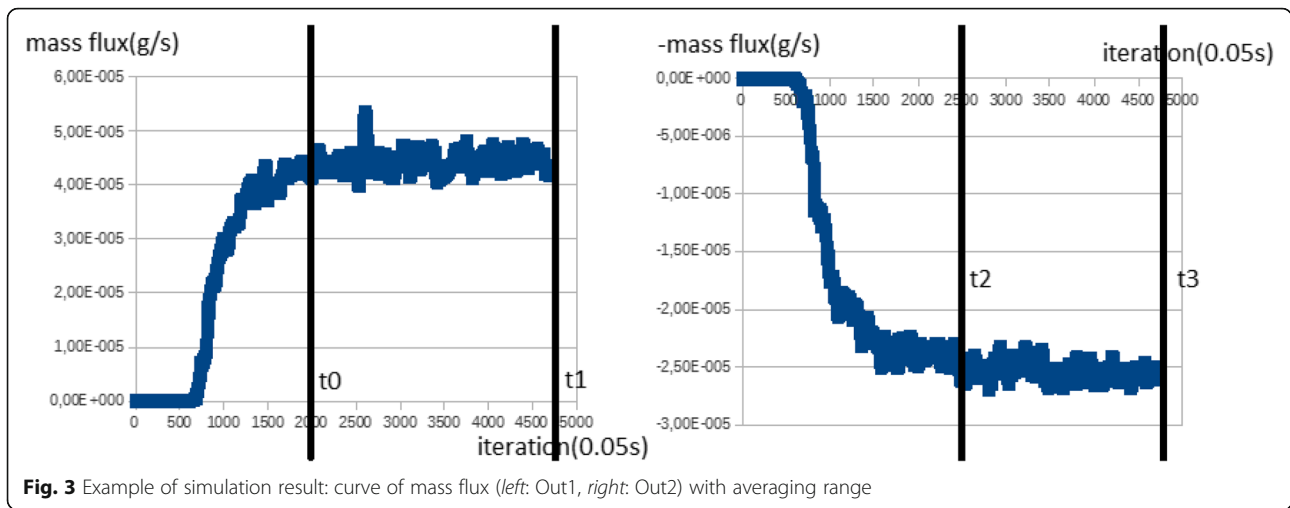
With  $v$  the velocity that is calculated,  $vm$  the average velocity,  $r$  the radius from the center of the pipe and  $R$  the internal radius of the pipe. The last boundary surface is set as pressure outflow.

- The second one is composed of the turbulent cases ( $Re \geq 4000$ ). Large Eddy Simulation (LES) has been applied with Smagorinsky model. The Mixture model of ANSYS Fluent has been used. Two fluids are defined to be water with similar properties but different limit conditions, the volume fraction of each fluid is set to 0 and 1 respectively at each input. Turbulent profiles have been developed and injected at the inlet and flow percentages are defined for each output.

Additionally, LES and DNS simulations are transient and therefore results need to be averaged for a sufficient period (e.g., several dozen minutes) when converged (see Fig. 3).

It was determined in a previous study [13] that for  $Re > 10,000$ , no further change in behavior is observed. Additionally, it has been shown by [13] that after 20 diameters for the length between 2 T-junctions, perfect mixing occurs (Eq. 2). A discussion with operating partners has permitted fixing the low boundary for distance between T-junctions. Indeed, crosses and double T-junctions cannot be compared easily, because for double T-junctions some space is needed to put a valve at the interpipe. Therefore, we have chosen the initial domain for parameters as in Table 1.

An important issue is the choice of boundary conditions (cases when either  $R_1^{in}$  or  $R_1^{out}$  equals 0 or 100):



**Fig. 3** Example of simulation result: curve of mass flux (left: Out1, right: Out2) with averaging range

contrary to previous studies, the boundary conditions are fixed to be the perfect mixing case (Eq. 2).

Initially 180 simulations were carried out but some grid refinement was necessary for better capturing the variations of the concentration at outlets. In this paper, we have placed iteratively new simulation points to minimize the error of interpolation by the Delaunay method. The interpolation is first described and kriging method is used.

### 2.2 Kriging interpolation

The Kriging method is a well-known method [5, 11, 14] invented in geological sciences by Matheron and Krige. It gives an effective way of interpolating non-structured points outside the initial Cartesian grid.

Spatial inference of the mass flux  $Z$ , at an unobserved location  $s_0$  in the space of parameters defined in Table 1, is calculated from a linear combination of the observed values  $z_i = Z(s_i)$  and weights  $\lambda_i$  as follows:

$$\hat{Z}(s_0) = \sum_i \lambda_i(s_0, s_i) Z(s_i)$$

There are two objectives for estimating the weights: to obtain minimal variance of estimation and to constrain the minimal variance objective.

Kriging is a minimum variance method based on the definition of a probability distribution at each of the  $n$  sample points  $x_i$ . It is based on the theory of regionalized variables

**Table 1** Parameters' range

Parameters	Values
Reynolds average: Re	1000; 2000; 5000; 10,000; 20,000;
Interpipe length: L (Diameters)	5; 8; 10; 20;
Inlet Reynolds ratio: $R_1^{in}$	0; 30; 50; 70; 100
Outlet Reynolds ratio: $R_1^{out}$	0; 30; 50; 70; 100

and provides best results when data points are not evenly spread. Compared to radial-based interpolation, it gives more predominance to closer nodes [20]. It also has the advantage to be a generalization of radial-based function [10]. A comparison of 12 different types of interpolations is accomplished in [6] and Kriging interpolation gives a very good approximation but is quite slow. However, the time of interpolation can be neglected in this study compared to the CFD simulation time, therefore Kriging interpolation is a good choice. The method chosen is the universal Kriging that we detail below.

#### 2.2.1 Kriging interpolation solution

$Z$  is taken as the sum of a polynomial function  $F$  and a stochastic scalar function  $Y$ . The first part explains the mean behavior of the function we interpolate and the second part is a stochastic function with known mean and variance. Considering  $n$  points of observation  $s_i$  and the interpolation point  $s$  the following stochastic model reads:

$$Z(s) = \sum_{l=1}^p f_l(s)\beta_l + Y(s) \tag{5}$$

With  $f_l$  the polynomial component functions,  $\beta_l$  the unknown coefficients of the polynomial and  $p$  the polynomial order.

We set

$$\mathbf{F}(s) = (f_1(s), \dots, f_p(s)) \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

Therefore (5) can be written,

$$Z(s) = Y(s) + \mathbf{F}(s)\boldsymbol{\beta}$$

Remark: if the row-vector  $\mathbf{F}(s)$  is only composed of one element equal to 1, the method is called ordinary Kriging, otherwise it is universal Kriging.

In the theory, it is also supposed that the stochastic part is of expectation zero.

$$E(\mathbf{Y}(s)) = 0$$

It implies:

$$\begin{aligned} Cov(\mathbf{Z}) &= Cov(\mathbf{Y}) = E[\mathbf{Y}\mathbf{Y}^T] \\ E[Z(s)] &= \mathbf{F}(s)\boldsymbol{\beta} \end{aligned}$$

Where  $\mathbf{Z}$  and  $\mathbf{Y}$  are the column n-vectors with components  $(Z(s_i))$  and  $(Y(s_i))$

Finally, the covariance of the noise  $\mathbf{Y}$  is set as:

$$\begin{aligned} \mathbf{K} &= Cov(\mathbf{Y}) = \sigma^2\mathbf{R}(\mathbf{Y}) \\ R(Y(s_i), Y(s_j)) &= \prod_{k=1}^m \exp(-\theta_k |s_{ik} - s_{jk}|^{\gamma_k}) \end{aligned} \tag{6}$$

Where  $m$  is the dimension of the problem (here 4) and  $\theta_k$  and  $\gamma_k$  are to be estimated. The same equation is used when calculating  $R(Y(s_0), Y(s_j))$  with the coordinates of the estimation point  $s_0$  instead of any observation point  $s_i$ .

The Kriging function is solved by setting the estimation of  $Z$  at the point  $s_0$  by taken it as a linear combination of the value  $Z$  takes on the observation points.

$$\hat{Z}(s_0) = \sum_i \lambda_i(s_0, s_i)Z(s_i) \tag{7}$$

Where the  $\lambda_i$  are the unknowns of the problem. Moreover, we consider an unbiased predictor:

$$E[\hat{Z}(s_0)] = E[Z(s_0)]$$

This leads to:

$$\forall s_0, \sum_i \lambda_i \mathbf{F}(s_i)\boldsymbol{\beta} = \mathbf{F}(s_0)\boldsymbol{\beta}$$

Which is satisfied if:

$$\forall l \in [1, p], \sum_i \lambda_i f_l(s_i) = f_l(s_0)$$

We seek  $\lambda_i$  that minimizes the variance of the error term:

$$\text{mine}(\boldsymbol{\lambda}) = Var[\hat{Z}(s_0) - Z(s_0)]$$

By expanding the error term:

$$\begin{aligned} \hat{Z}(s_0) - Z(s_0) &= \sum_i \lambda_i (Y(s_i) + \mathbf{F}(s_i)\boldsymbol{\beta}) - Y(s_0) - \mathbf{F}(s_0)\boldsymbol{\beta} \\ &= \sum_i \lambda_i Y(s_i) - Y(s_0) \end{aligned}$$

It follows:

$$\begin{aligned} Var[\hat{Z}(s_0) - Z(s_0)] &= Var\left[\sum_i \lambda_i Y(s_i) - Y(s_0)\right] \\ &= E\left[\left(\sum_i \lambda_i Y(s_i) - Y(s_0)\right)^2\right] - \left(E\left[\sum_i \lambda_i Y(s_i) - Y(s_0)\right]\right)^2 \\ &= E\left[\left(\sum_i \lambda_i Y(s_i) - Y(s_0)\right)^2\right] \\ &= E\left[\left(\sum_i \lambda_i Y(s_i)\right)^2\right] - 2\sum_i \lambda_i E[Y(s_i)Y(s_0)] + E[(Y(s_0))^2] \\ &= \sum_i \sum_j \lambda_i \lambda_j Cov(Y(s_i), Y(s_j)) - 2\sum_i \lambda_i Cov(Y(s_i), Y(s_0)) \\ &\quad + Var(Y(s_0), Y(s_0)) \end{aligned}$$

We recall that:

$$K_{ij} = Cov(Y(s_i), Y(s_j))$$

We have found that the function to minimize is quadratic in  $\lambda$  with Hessian  $\mathbf{K}$  that is symmetrical definite positive (under mild conditions of  $\boldsymbol{\theta}$ ). It is therefore strongly convex.

The convex minimization problem states:

$$\left\{ \begin{aligned} \min Var[\hat{Z}(s_0) - Z(s_0)] \\ \forall l \in [1, p], \sum_i \lambda_i f_l(s_i) = f_l(s_0) \end{aligned} \right. \tag{8}$$

It is possible to define a problem without constraint by defining the Lagrangian function and the Lagrange multipliers  $\mu_1$  such that:

$$\begin{aligned} L(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \sum_i \sum_j \lambda_i \lambda_j K_{ij} - 2\sum_i \lambda_i k_i(s_0) + k_{00} \\ &\quad + 2\sum_l \mu_l \left(\sum_i \lambda_i f_l(s_i) - f_l(s_0)\right) \end{aligned} \tag{9}$$

Or

$$\begin{aligned} L(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \langle \boldsymbol{\lambda} | \mathbf{K} \boldsymbol{\lambda} \rangle - 2\langle \boldsymbol{\lambda} | \mathbf{k}(s_0) \rangle + \sigma^2 \\ &\quad + 2\langle \boldsymbol{\mu} | \sum_i \lambda_i \mathbf{F}^T(s_i) - \mathbf{F}^T(s_0) \rangle \end{aligned}$$

With  $k_i(s_0) = Cov(Y(s_0), Y(s_i))$  and  $k_{00} = Cov(Y(s_0), Y(s_0)) = \sigma^2$  are given by Eq. (6).

The sufficient and necessary optimality conditions are:

$$\left\{ \begin{aligned} \forall l \in [1, p], \sum_i \hat{\lambda}_i f_l(s_i) = f_l(s_0) \\ \forall i \in [1..n], \sum_j \hat{\lambda}_j K_{ij} - k_i(s_0) + \sum_l \hat{\mu}_l f_l(s_i) = 0 \end{aligned} \right. \tag{10}$$

These last equations can be written as the saddle point equation:

$$\begin{pmatrix} \mathbf{K} & \mathbf{F} \\ \mathbf{F}^T & 0 \end{pmatrix} \begin{pmatrix} \hat{\lambda} \\ \hat{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{K}_0 \\ \mathbf{F}(s_0)^T \end{pmatrix} \tag{11}$$

Using the Schur complement of  $\mathbf{K}$  in the full matrix, it leads to the expression of optimal  $\lambda$  and  $\mu$ :

$$\hat{\mu} = [\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F}]^{-1} (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{K}_0 - \mathbf{F}(s_0)^T) \tag{12}$$

$$\hat{\lambda} = \mathbf{K}^{-1} (\mathbf{K}_0 - \mathbf{F} \hat{\mu}) \tag{13}$$

Moreover, we also obtain an estimation of the error variance by using Eq. (10) in the Eq. (9):

$$Var[\hat{Z}(s_0) - Z(s_0)] = \hat{\sigma}^2 - \hat{\lambda} \mathbf{K}_0 - \hat{\lambda}^T \mathbf{F} \hat{\mu}$$

The parameters  $\sigma$ ,  $\beta$ ,  $\theta$  (Eqs. 5 and 6) are estimated at maximum-likelihood sense and calibrated in different process that is detailed in Appendix. The  $\gamma_k$  are fixed to 1 s; a regularization term  $\alpha$  is added to the diagonal of the correlation matrix.

The interpolation is used in the 1-D law based on the lookup-table of the results, and to determine the errors of interpolation used with a Delaunay triangulation to choose new points of simulation from an initial set.

### 2.3 Delaunay triangulation

One limitation of such a study is the limited number of simulations to be performed. Indeed, DNS and LES simulations are computationally demanding, therefore a method is needed to find the best point of simulation. One well-known method is the Latin-Square design [1] consisting in dividing the space into a fixed number of squares and then randomly picking points in each square. However, for numerical simulations, a major drawback is that it requires a lot of simulation points, at least one for each region defined. In this study we have selected the Delaunay triangulation method. It consists of the calculation of interpolation error sums (based on the Kriging as described before) in the triangulation domain to define the region the most susceptible to be poorly modelled. It was made to enrich the database of principal components analysis basis functions used in oscillating airfoils in a compressible flow context. Points of simulations could have been chosen from structured method or Latin Hypercube method [15] which is a generalization of the Latin-Square design, however more points of simulation would have been needed.

The Delaunay triangulation is in general used for space partitioning. It defines a list of triangles from a list of points where no points are strictly inside any circum-circle of any triangle. Here it is used to determine new points of simulations by a greedy algorithm:

- For each point of the design plan, we define the interpolation error as the absolute difference

between its value and the interpolated value at that point when it is not considered;

- For every Delaunay triangle the sum of the interpolation error of its vertices multiplied by its area is calculated;
- The best point candidates to include in the design plan are the centers of gravity for triangles with the highest weight; then, in this research, the selected point is the closest point with rounded coordinates, which is more convenient for simulation.

A simple example is given in Fig. 4 taking into account nine points, which are given in Table 2. The function  $M$  is the list of the measurements and  $E$  is the list of errors of interpolation when discarding the point.

The method divides the space in eight triangles of equal area. The maximum error is found at the points (0,0) and (2,0) equal to 1.75. In this case, the two-bottom triangles will most likely be chosen for new simulations (blue crosses), this is where more accuracy is needed. The method detects the most problematic point, here the only point not in the plan  $z = y + 1$ , and refines around it.

To summarize, the Delaunay triangulation method and a greedy algorithm are used here to select appropriate new simulation points. With such a formulation, the gain of the selected point is maximized.

### 2.4 1-D law

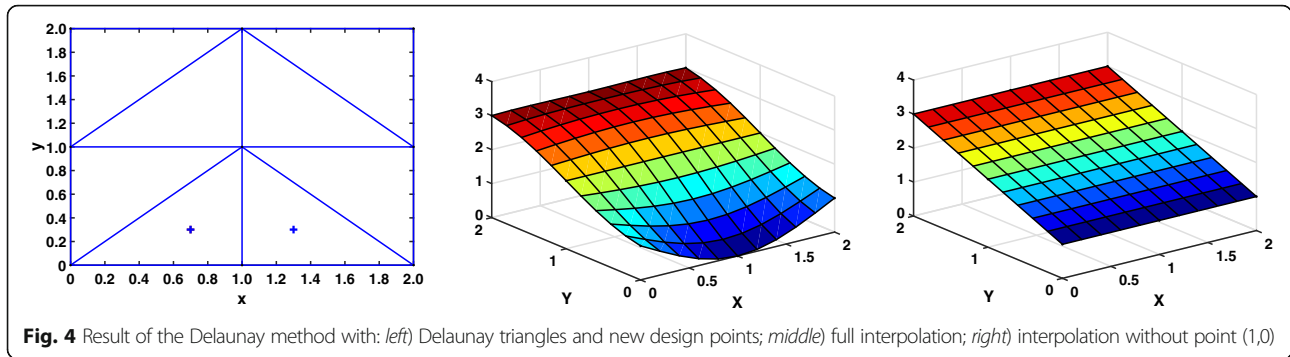
Once the appropriate simulation points are selected and computed (the lookup table is filled), a 1-D law can be developed to couple it with current transport models which are all one dimensional. Simulations are time consuming and therefore can only be done in a very small network, but it is impossible in a real one. So, a 1-D model should be designed in order to predict the transport in an operating network that can be used in real time. It uses the lookup table of the simulations results and determines the proportion of mixing with the help of the Kriging interpolation. The simulations are executed based on the following scenario: pure water coming from In1 and contaminated water coming from In2. The law first needs to be generalized in case of any water concentration at each inlet.

Let's consider  $C_1$ ,  $C_2$ ,  $C_3$  and  $C_4$  the mean concentrations on the cross sections for the following pipes In1, In2, Out1 and Out2 (cf. Fig. 2). In the CFD simulation,  $C_1 = 0$  mg/L,  $C_2 = 1$  mg/L have been chosen. We calculate the ratio  $\theta$  of the average mass flux  $Q_3 C_3$  going out of the output 1 divided by the introduced mass flux:

$$\theta = \frac{Q_3 C_3}{Q_1 C_1 + Q_2 C_2}$$

And in the case  $C_1 = 0$  mg/L,  $C_2 = 1$  mg/L:





$$\theta = \frac{Q_3 C_3}{Q_2 C_2} = \frac{Q_3 C_3}{Q_2}$$

We then define  $\theta^*$  as the deviation from the perfect mixing ratio:

$$\theta^* = \theta - \frac{Q_3 \times \frac{(Q_1 C_1 + Q_2 C_2)}{Q_1 + Q_2}}{Q_1 C_1 + Q_2 C_2} = \theta - \frac{Q_3}{Q_1 + Q_2} = \theta - \frac{Q_3}{Q_3 + Q_4} \tag{14}$$

$\theta^*$  is the coefficient saved in the lookup table.  $\theta^*$  is bounded from below by  $-1$  and from above by  $+1$ . In practice  $\theta^*$  was found in the range  $[-0.056, 0.361]$ .

From  $\theta^*(R_1^{in}, R_1^{out}, Re, \frac{L}{D})$ , that is interpolated at values which are not in the table by the Kriging method,  $\theta(R_1^{in}, R_1^{out}, Re, \frac{L}{D})$  is calculated from Eq. (14).

Therefore, if  $C_1 = 0$  mg/L,  $C_2 = 1$  mg/L:

$$Q_3 C_3 = \theta(R_1^{in}, R_1^{out}, Re, L/D) \times Q_2$$

This can be generalized for any  $C_1$  and  $C_2$ . The mass flux leaving by outlet Out1 is composed of a part coming from In1 and another one from In2, the same for Out2.

$$\begin{cases} Q_3 C_3 = \theta_{13} \times Q_1 C_1 + \theta_{23} \times Q_2 C_2 \\ Q_4 C_4 = \theta_{14} \times Q_1 C_1 + \theta_{24} \times Q_2 C_2 \\ \theta_{23} = \theta \end{cases}$$

We thus have a system of three equations and six unknowns ( $C_3, C_4, \theta_{13}, \theta_{23}, \theta_{14}, \theta_{24}$ ), since we know the flows,  $C_1$  and  $C_2$ . To resolve the system, we need three more constrains, they can be found as follows:

When  $C_1 = 0$ , the mass-balance between inlets and outlets must be satisfied:

**Table 2** Example of the Delaunay method with associated errors

M(x, y)	x=0	x=1	x=2	E(x, y)	x=0	x=1	x=2
y=2	3	3	3	y=2	0.70	0.48	0.70
y=1	2	2	2	y=1	0.57	0.53	0.57
y=0	1	0	1	y=0	1.75	1	1.75

$$(Q_1 C_1 + Q_2 C_2) = Q_2 C_2 = (Q_3 C_3 + Q_4 C_4)$$

So

$$Q_2 C_2 = \theta_{23} \times Q_2 C_2 + \theta_{24} \times Q_2 C_2$$

Assuming that  $Q_2 C_2$  is not zero leads to:

$$\theta_{24} = (1 - \theta_{23})$$

Similarly, when  $C_1 = 0$  and  $Q_1 C_1$  is positive, we get:

$$\theta_{14} = (1 - \theta_{13})$$

Finally, for  $C_1 = C_2 > 0$ , the mixing of fluids of similar concentration should give a fluid of same concentration,  $C_3 = C_4 = C_1 = C_2 > 0$ :

$$\begin{cases} Q_3 = \theta_{13} \times Q_1 + \theta_{23} \times Q_2 \\ Q_4 = \theta_{14} \times Q_1 + \theta_{24} \times Q_2 \end{cases}$$

then

$$\begin{cases} \theta_{13} = \frac{Q_3 - \theta_{23} \times Q_2}{Q_1} \\ \theta_{14} = \frac{Q_4 - \theta_{24} \times Q_2}{Q_1} \end{cases}$$

This gives four more equations from which three are independent, for example:

$$\left[ \theta_{13} = \frac{Q_3 - \theta_{23} \times Q_2}{Q_1} ; \theta_{24} = (1 - \theta_{23}) \text{ and } \theta_{14} = (1 - \theta_{13}) \right]$$

$$\Rightarrow \theta_{14} = \frac{Q_4 - \theta_{24} \times Q_2}{Q_1}$$

We have six equations for six unknowns that can be reduced to four equations with four unknowns:

$$\begin{cases} Q_3 C_3 = \theta_{13} \times Q_1 C_1 + \theta_{23} \times Q_2 C_2 \\ Q_4 C_4 = (1 - \theta_{13}) \times Q_1 C_1 + (1 - \theta_{23}) \times Q_2 C_2 \\ \theta_{23} = \theta \\ \theta_{13} = \frac{Q_3 - \theta_{23} \times Q_2}{Q_1} \end{cases} \tag{15}$$

$C_3$  and  $C_4$  can be calculated from  $\theta, C_1, C_2$  and the flows rates.

These equations allow the definition of a law that can be used for every double T-junction (with equal

diameters). CFD simulation results have helped to fill a lookup table with a non-dimensional indicator of the deviation from the perfect mixing. A system of equations that generalizes the case of an introduction of a contaminant only at inlet 2 to both inlets is derived. Firstly the Kriging interpolation is used to determine the coefficient  $\theta$  and then is put as input in Eq. (21) to get the concentration at the outlets.

For instance, let's consider a Double T-junction with a diameter of 0.1 m and an interpipe of length  $L = 5D$ . The flow rates are for each input and output such that Reynolds numbers are  $Re_1 = Re_2 = Re_3 = Re_4 = 2500$ . The four parameters are respectively,  $Re = 5000$ ,  $L = 5$ ,  $R_1^{in} = 50$ ,  $R_1^{out} = 50$ . The interpolation function returns  $\theta = 0.6$ , then  $\theta_2 = 0.6$  and  $\theta_1 = 0.4$ . Let's say that a contaminated water comes from inlet1 with  $C_1 = 1$  mg/L and another water at inlet 2 with  $C_2 = 4$  mg/L then the concentration on both outputs are:  $C_3 = 2.8$  mg/L and  $C_4 = 2.2$  mg/L.

### 3 Results

In this part the results are given and interpreted. First the Kriging calibration and the explanation for the parameter choice are described. Then, it is shown how the Delaunay triangulation helps to choose the design points and reduces the computational overall cost. Finally, the CFD simulation results, for different Reynolds numbers and contaminant inlet-exit ratios (Fig. 14), are given and interpreted to explain the behavior of the imperfect

mixing in Double T-junctions depending on the different entry parameters.

#### 3.1 Kriging calibration

This part follows the theory on the calibration of the Kriging parameters. It has been shown that after simplification, the Eq. (15) can be used to determine the parameters ( $\sigma$ ,  $\theta_k$  and  $\gamma_k$ ).

The problem is not convex, therefore to understand the behavior of the function  $f(\tilde{\theta}, \alpha)$ , it has been plotted on

Fig. 5 with the generalized correlation parameter  $\tilde{\theta} \in [0, 10]$ , the regularization term  $\alpha \in [0, 1]$  and the polynomial degree  $k$  between 0 and 3 as well as no polynomial,  $k$  being the degree of the polynomial. With  $\sigma$ ,  $\beta$  fixed as optimal values from Eqs. 11 and 12). When there is not a polynomial term it is called simple Kriging (SK), when  $k = 0$  its name is ordinary Kriging, for  $k > 0$  it is universal Kriging.

For  $k$  between 0 and 3 and SK, there are 3 regions for  $\alpha$ . If  $\alpha$  is too small ( $< 0.45$ ), the R matrix may not be invertible (because it is not regularized enough), therefore its determinant is zero and the function  $f$  equals infinity. On the other hand, if  $\alpha$  is too large ( $> 0.6$ ), the correlation matrix may be too close to  $(1 + \alpha)I_n$ , therefore its determinant is approximately  $(1 + \alpha)^n$ , the matrix is invertible but its determinant in floating-point arithmetic is +infinity.

To get the optimized parameters, the system presented in the Appendix (Eq. 24) has been resolved by choosing  $\tilde{\theta} = 4$  at first iteration (because visually near the minimum) and  $\alpha = 0.5$ , for  $k = 0, 1, 2, 3$ :

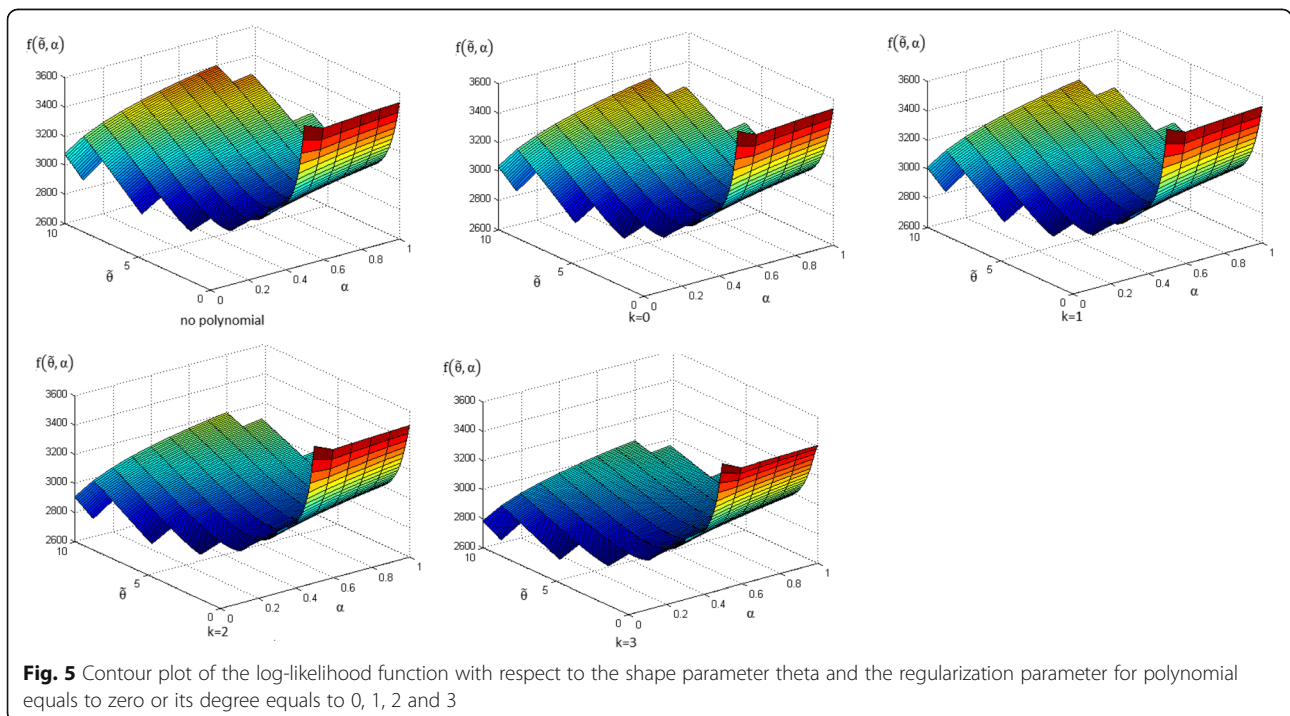


Table 3 sums up the result found by resolving Eq (21), final value of  $\tilde{\theta}$  increases with  $k$ . As  $\sigma$ ,  $\beta$  are different for each  $k$ , the initial value of  $f$  gives different initial values. In this case the initial value of  $f$  is further away from the optimal value for lower values of  $k$ . The optimal value of  $f$  decreases with  $k$ . The average absolute error is the lowest for  $k = 0$ . Except for that, for average error and standard deviation, their values decrease with  $k$  increasing. Minimum errors are close to zero for all and maximum errors decrease with  $k$  increasing.

Finally, we have chosen  $k = 3$ ,  $\hat{\sigma}$ ,  $\hat{\beta}$ ,  $\tilde{\theta} = 3.63$  and  $\alpha = 0.5$  with  $\gamma_k$  components all being fixed to 1 that correspond to a minimal standard deviation in this case.

### 3.2 Delaunay triangulation

In the Theoretical part it was shown that the Delaunay triangulation can be used for the design of simulation points. Here, it is described how it can be also used to reduce the computational cost overall compared to a classical Structured design algorithm.

In Fig. 6, the behavior of the Delaunay method is highlighted with an application of the case  $L = 5D$  and Reynolds number = 1000. On the left can be seen the Delaunay triangles partitioning the domain space, the vertices being the point of simulation or the boundaries. The blue cross defines the new point of simulation to perform. On the right is plotted the interpolation result, initially and after multiple simulations chosen by the Delaunay method. The repartition of the simulation chosen points is not structured, most points are chosen on the right part where the straight inlet is dominating,  $R_1^{in} > 50$ . It is also where the simulations give imperfect mixing results and the derivatives are bigger.

**Table 3** Result of optimization of the Kriging parameters by resolving the log-likelihood function with respect to the shape parameter theta and the regularization parameter for no polynomial and polynomial degree equals to 0, 1, 2 and 3

Polynomial degree	No polynomial	$k = 0$	$k = 1$	$k = 2$	$k = 3$
$\tilde{\theta}$	2.76	2.81	2.85	3.13	3.63
$\alpha$	0.5	0.5	0.5	0.5	0.5
$f(\tilde{\theta}, \alpha)$ initial	2902.92	2895.64	2891.3	2855.29	2806.2
$f(\tilde{\theta}, \alpha)$ optimal	2866.93	2864.77	2863.32	2841.27	2804.62
average error		1.75	1.94	1.86	1.82
standard deviation		3.40	3.24	2.83	2.37
min		0.02	0	0	0
max		35.19	34.09	31.4	26.86

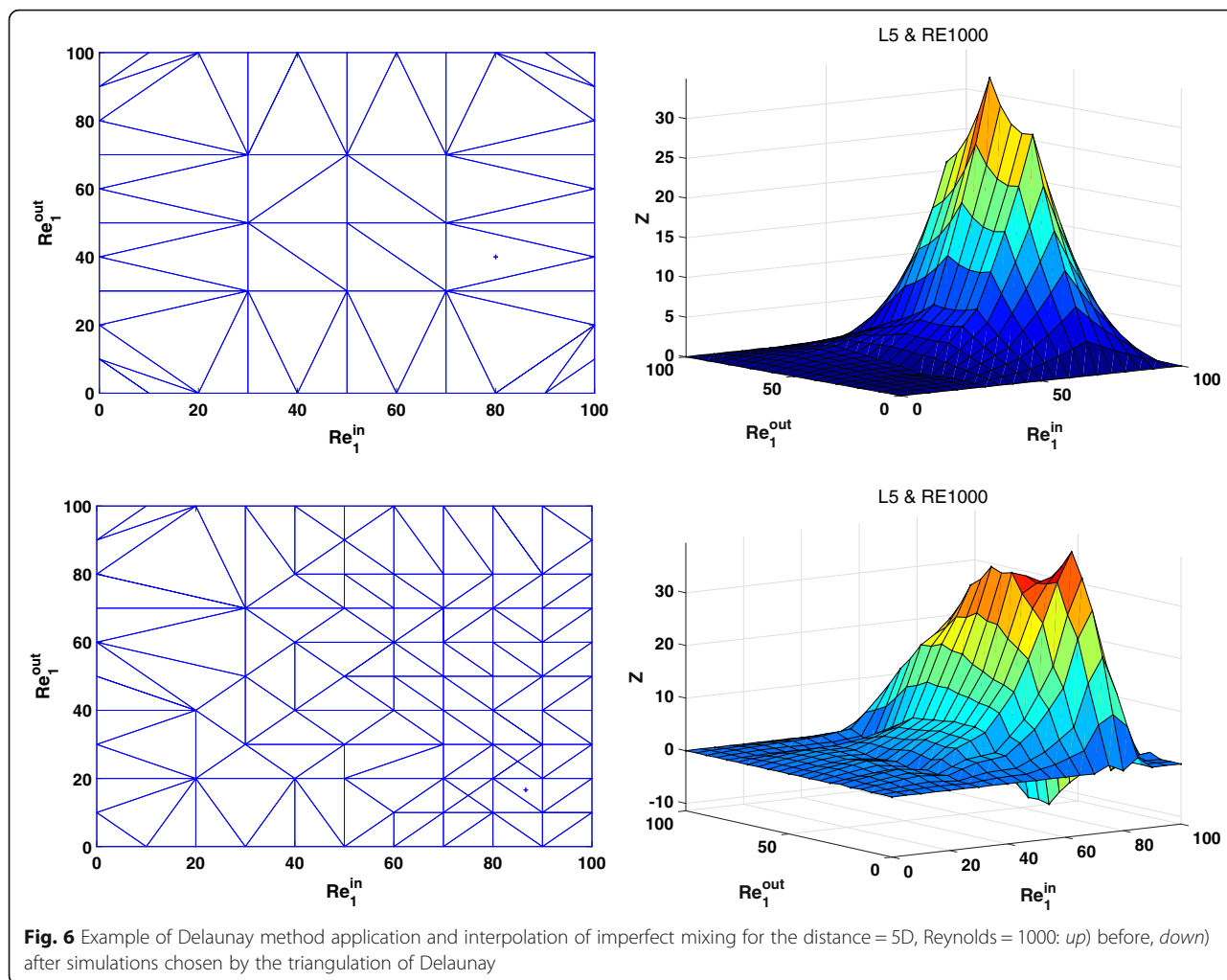
Table 4 sums up the different new simulation points to consider that were found for the case where distance =  $5D$  and  $Re = 1000$  are fixed. The first column refers to the order of the simulations made given the Delaunay method and the number 0 is for the initial simulations. The second and third columns give the percentage of inlet 1 Reynolds number and outlet 1 Reynolds number to the averaged Reynolds number 1000. Finally the fourth and fifth columns are the sum of the errors, either averaged (divided by current number of points) or the maximum, of all the points simulated at each stage.

In Table 4 and Fig. 7, it can be observed that for both types of error, it is globally decreasing. The error can increase when a particular point is found, in this example (80, 30), but then the Delaunay method will search for points around that will decrease the global error. Figure 6 on the left shows the points of Delaunay locations (vertices of the Delaunay triangle). At the end the method concentrates the points where there is a need for more information, here in the right part, when the straight inlet is dominating.

The Delaunay method has been used for space partitioning for selecting new CFD simulations to perform. At each step, it calculates the absolute interpolation errors, which can be used as a stop criterion. For instance, if we have fixed the mean error at 3 as the criteria of convergence, there is no need to continue the simulations after the 19th simulation. It gives an effective way to consume less computational time, by favoring simulations that give the most information.

In order to show that the Delaunay method performs better than the Structured method in terms of design point selection, they are compared. Figure 7 shows the average and maximum error of interpolation (as explained before) when adding points of simulation. To compare both methods there is the need to define how the Structured algorithm was performed. The Structured selections have been made using the following sets of simulation points in Fig. 8: the groups go from 0 to 8 (each having its own color). Each number represents a set of points, whose number is between 8 and 12 points. For instance, at first, both methods begin with the set 0 composed of 9 points corresponding to Reynolds number combination ratios of 30/50/70. Then, 10/50/90 (set 1) combination is added to the precedent points. Henceforth it goes until set 8, when the domain is filled with a fully-structured repartition of points.

For both average and maximum interpolation error, the Delaunay method gives better results with a steady average behavior of error decreasing. For the average error, both methods give similar results at first, until a point when the Delaunay method error stays smaller compared to the Structured error. For maximum error, the Structured method doesn't have a well-



**Fig. 6** Example of Delaunay method application and interpolation of imperfect mixing for the distance = 5D, Reynolds = 1000: up) before, down) after simulations chosen by the triangulation of Delaunay

defined behavior compared to average steady decrease of the Delaunay method. Each new peak of increase can be explained by new regions of space not well modeled. The Delaunay triangulation is then selecting points in that region until the error decreases again. This justifies the use of the Delaunay method over the Structured method.

The Delaunay method has been used for space partitioning for selecting new CFD simulations to perform. At each step, it calculates the absolute interpolation errors, which can be used as a stop criterion. For instance, if we have fixed the mean error at 3 as the criteria of convergence, there is no need to continue the simulations after the 19th simulation. It gives an effective way to consume less computational time, by favoring simulations that give the most information.

### 3.3 Simulation results and interpretations

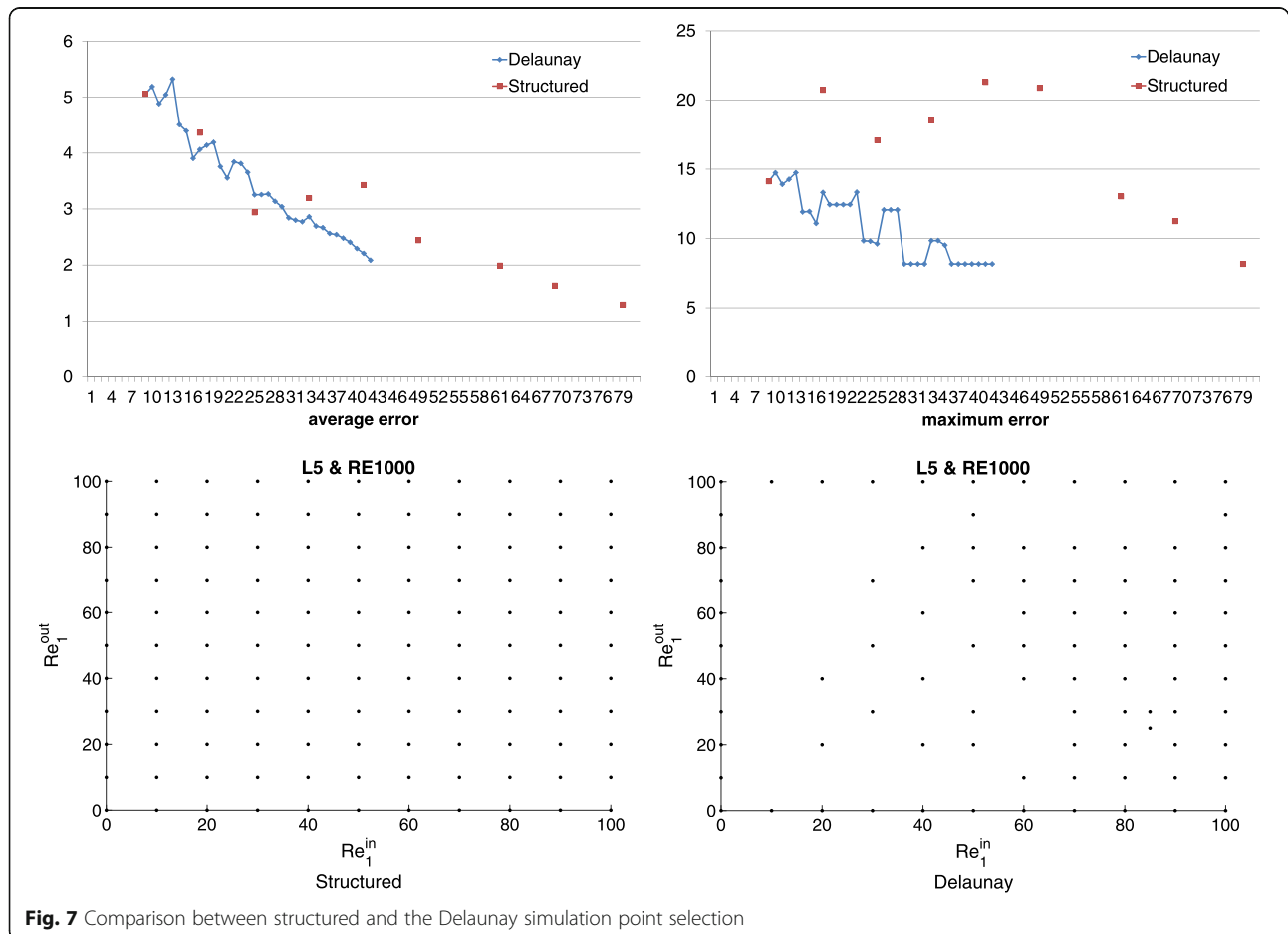
In this part, the CFD simulation results and the parameters influence on the behavior of the mixing in the Double T-junctions, are given and interpreted.

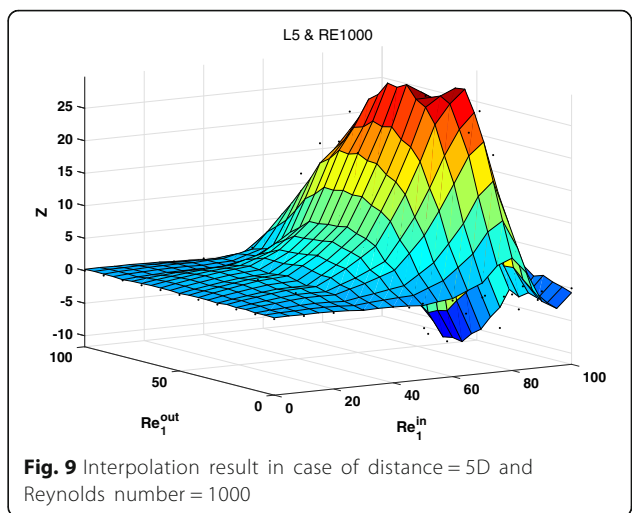
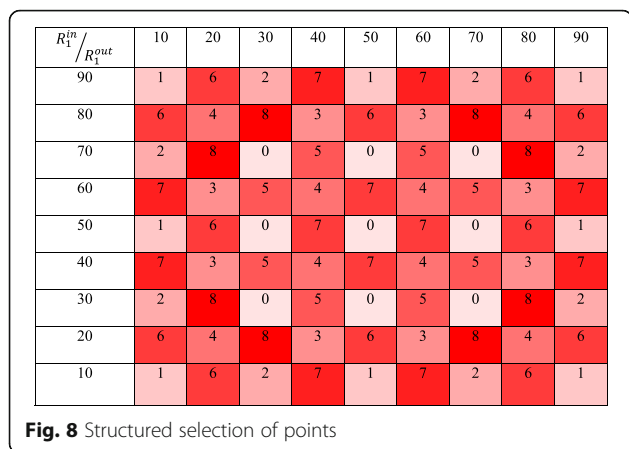
Table 5 gives a non-exhaustive list of the simulations performed. The first column is the distance between Ts (in diameters), the second is the average Reynolds number (in the center of the pipe), the third column is the Reynolds percentage of inlet 1 (straight inlet see Fig. 1) from the previous one, the fourth is for the outlet 1 ( $R_2^{in}$  and  $R_2^{out}$  can be found from  $100 - R_1^{in}$  and  $100 - R_1^{out}$  respectively). The last two columns give the repartition of contaminant for outlet 1 and 2 averaged in time.

For the interpretation of results, it is worth recalling that 100% of the contaminant comes from inlet 2 and 0% from inlet 1. One way of viewing Table 5 is by fixing two parameters and plotting the result with the two others varying. In Fig. 9 the distance is fixed to 5D and the Reynolds number is 1000. It shows Z, the difference between the result found in the simulations and the complete mixing plan ( $z = R_1^{out}$ ). Therefore the values are inside the boundary space  $[0 - R_1^{out}; 100 - R_1^{out}]$ . The difference is up to 30%, and therefore cannot be neglected. It can also be seen that for  $R_1^{in} < 50$ , that is when the straight inlet is dominated, the mixing law is

**Table 4** Delaunay points for distance = 5D and Re = 1000

Number	$R_1^{in}$	$R_1^{out}$	Mean. error	Max error	Number	$R_1^{in}$	$R_1^{out}$	Mean. error	Max error
0	30	30	5.1	14.1	6	60	80	4.4	11.1
0	30	70	5.1	14.1	7	80	30	3.9	13.3
0	50	50	5.1	14.1	8	80	20	4.1	12.4
0	70	30	5.1	14.1	9	70	20	4.1	12.4
0	70	70	5.1	14.1	10	80	80	4.2	12.4
0	50	30	5.1	14.1	11	70	80	3.7	12.4
0	50	70	5.1	14.1	12	70	10	3.6	13.3
0	30	50	5.1	14.1	13	90	10	3.8	9.8
0	70	50	5.1	14.1	14	80	10	3.8	9.8
1	80	40	5.2	14.7	15	40	60	3.7	9.6
2	80	60	4.9	13.9	16	70	60	3.3	12
3	60	60	5.1	14.3	17	90	20	3.3	12
4	60	40	5.3	11.9	18	50	20	3.3	12
5	80	50	4.5	11.9	19	50	90	3.0	8.2





almost complete. The mixing is inverted when  $R_1^{in} > 85$ , more concentration is detected at Outlet 2 than in perfect mixing condition. A peak is visible with its top around  $R_1^{in} = 70$  found in every configuration.

Figure 10 shows the results when the Reynolds number is fixed at 5000 and the distance is increased from 5D to 10D. It can be observed that there is a peak situated in the right part (when the straight inlet is dominating). The effect of increasing the distance between the two T-junctions looks to be a reduction of the peak in height and width. More simulations are needed for the case 5D as well as tests for grid convergence when the straight inlet is dominating ( $R_1^{in} > 80$ ).

Now, CFD simulations are plotted and interpreted. First is given the repartition of simulations that have been carried out for the four parameters (Re, L,  $R_1^{in}$  and  $R_1^{out}$ ). Then, two ways of interpreting the results are presented. One is by fixing the Reynolds number and the distance, the other by fixing the inlet and outlet Reynolds ratios.

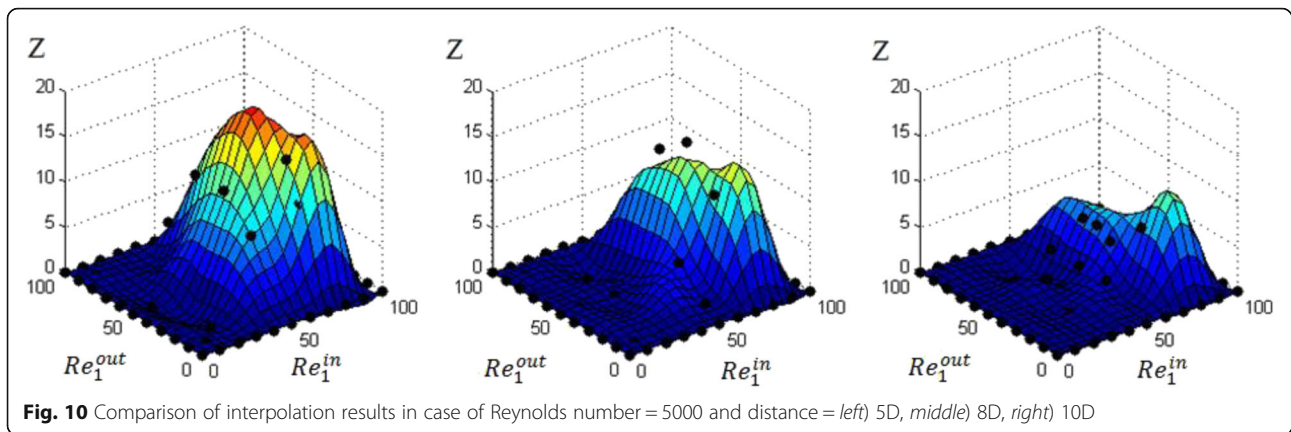
All simulation points have been gathered on

Figure 11 that shows the list of simulations that have been realized, here represented by points. Each point is defined by its Reynolds number and its intermediate pipe length as well as its value of  $R_1^{in}$  and  $R_1^{out}$  respectively in x-axis and y-axis. The Delaunay method has been mostly used on the space  $Re = 1000/L = 5D$  and  $Re = 5000/L = 5D$  to ensure the validity of the interpolation.

The results of the four parameter interpolation have been drawn on Figs. 12 and 13. The first figure gives a lot of information on the behavior of this law. We may observe that when the second input (see Fig. 2) is dominating ( $R_1^{in} < 50\%$ ), the mixing is almost perfect. On the other hand, when it is the first input that is dominating, the mixing is not perfect. And it converges toward the perfect mixing with the increase of the distance and the Reynolds number. The case  $Re = 30,000, L = 5D, R_1^{in} = 70$  was additionally computed and is showing that the function converges to a stable equilibrium for high Reynolds number, which may not be the perfect mixing depending on the distance. In case 5D, 70/50 it tends to  $Z = 10$ , that

**Table 5** Results of the CFD simulations for different parameters (non-exhaustive)

Distance (in diameter)	Reynolds number	$R_1^{in}$	$R_1^{out}$	$Z = ave R_1^{out}$	$ave R_2^{out}$	Distance (in diameter)	Reynolds number	$R_1^{in}$	$R_1^{out}$	$Z = ave R_1^{out}$	$ave R_2^{out}$
5	1000	20	20	21.6	78.4	5	10,000	50	50	62.5	37.5
5	1000	20	80	81.1	18.9	5	10,000	50	70	82	18
5	1000	30	30	30.1	69.9	5	10,000	70	50	69	31
5	1000	30	70	71.7	28.3	8	5000	30	30	31	69
5	1000	70	30	58.8	41.2	8	5000	30	50	50	50
5	1000	70	70	93.4	6.6	8	5000	30	70	70	30
5	1000	80	20	54.8	45.2	10	1000	20	20	21.5	78.5
5	1000	80	80	86.3	13.7	10	1000	20	80	81.5	18.5
5	5000	20	20	20	80	10	1000	30	30	32.5	67.5
5	5000	20	80	79	21	10	1000	30	70	70.5	29.5

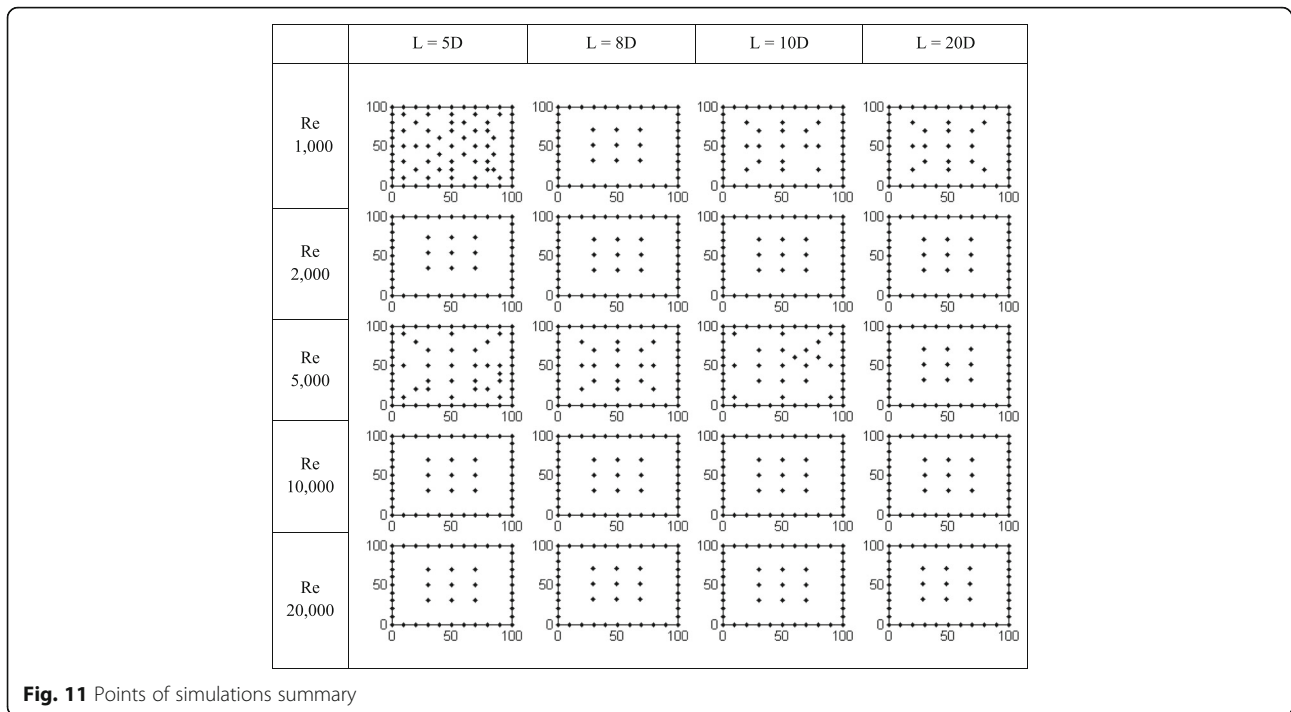


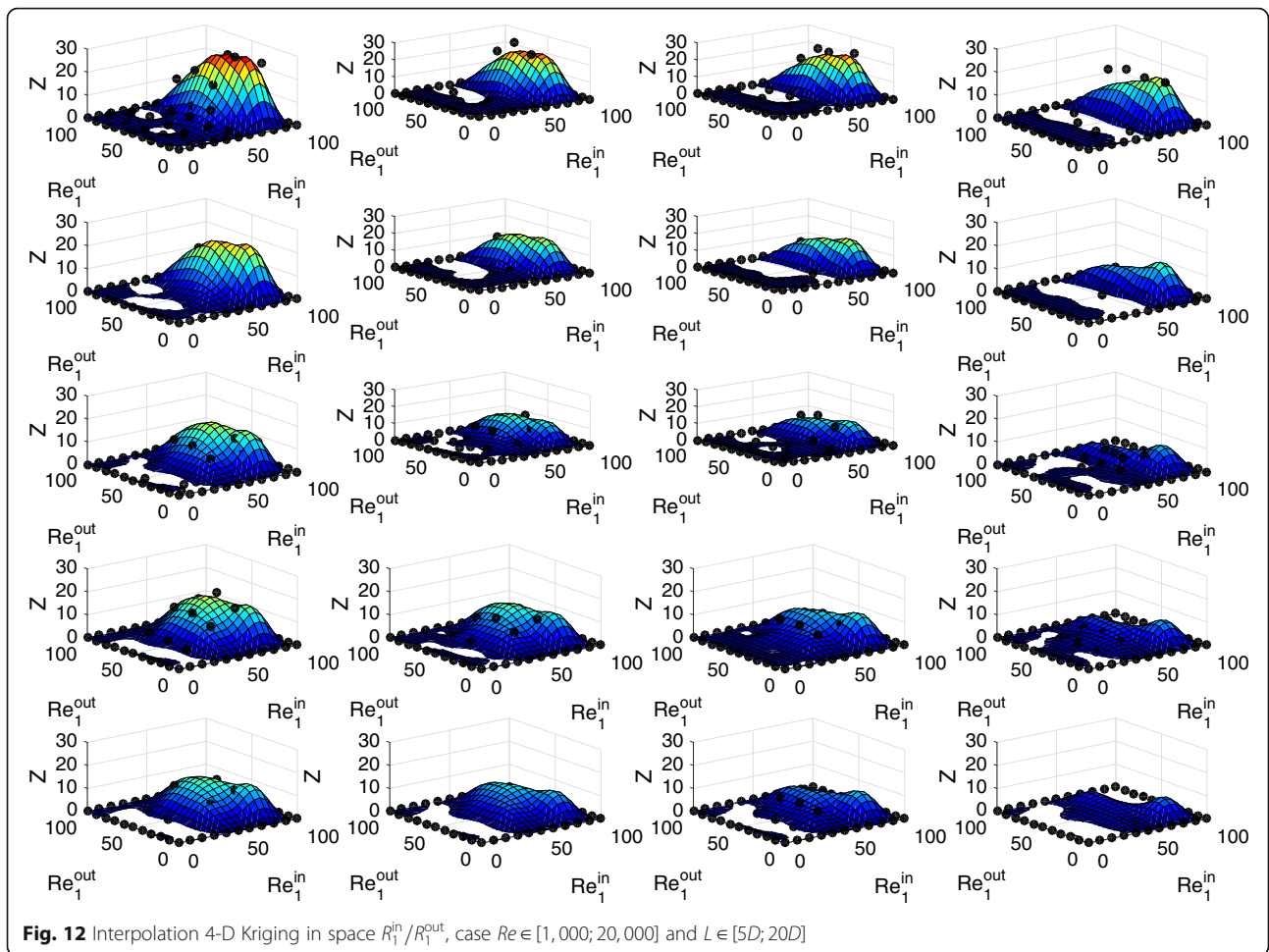
means that in this case where the ratio at the output is 60% of mass introduced goes into Out1 (instead of 50%) and 40% in Out2. When D is increased, a slow converging to perfect mixing in the laminar case can be seen. Figure 13 shows that for different configurations of Reynolds fractions in input and output, the behavior law is very different. As seen in Fig. 12, when  $R_1^{in}$  is lower than 50 the mixing is mostly perfect. In the laminar case the mixing has a clear behavior, under  $R_1^{in} = 50$  it is perfect mixing and above 70 it is imperfect mixing. In the turbulent case it is more progressive.

To understand the behavior of the mixing, Fig. 14 shows the longitudinal section of the double T-junction

in the case  $Re = 1000$ ,  $L = 5D$  and  $R_1^{in}$  and  $R_1^{out} = 30/50/70$  and plots the scalar. It can be seen that when the percentage of inlets at the first input is lower than  $R_1^{in} = 30\%$ , the contaminated water coming from In2 hits the opposite wall, creating turbulence. When  $R_1^{in} = 50\%$ , the turbulence is caused by the corner of the second T-junction. The more fluid going into the second output, the less stable is the flow. For  $R_1^{in} = 70$ , most of the contaminated water goes into the first output regardless of the output velocity repartition.

More than 250 simulations have been made to describe the phenomenon in four directions (Reynolds, inner T distance,  $R_1^{in}$  and  $R_1^{out}$ ). The results have been





interpolated and projected on two different spaces ( $R_1^{in}, R_1^{out}$ ) and ( $Re, L$ ). When the first input is dominated ( $R_1^{in} < 50$ ) the mixing is perfect. The mixing becomes imperfect when  $R_1^{in}$  increases. This behavior shift is more pronounced for laminar cases whatever the inter T-junction distance, and progressive in turbulent cases. There is a maximum of imperfect mixing around  $R_1^{in} = 70$ . The mixing tends to perfect mixing when the inter T-junction distance increases, it is almost perfect for  $L = 20D$  in the turbulent case. It also decreases when the Reynolds number is increased but tends to equilibrium that may not be perfect mixing (tends to 10 when  $L = 5D$ , %  $R_1^{in} = 70$ ,  $R_1^{out} = 50$ ). The behavior described can be explained by looking into the longitudinal section of the double T-junction. The mixing is created by hitting the walls, which depends on the ratios of the flow rates in input and output.

**4 Conclusions**

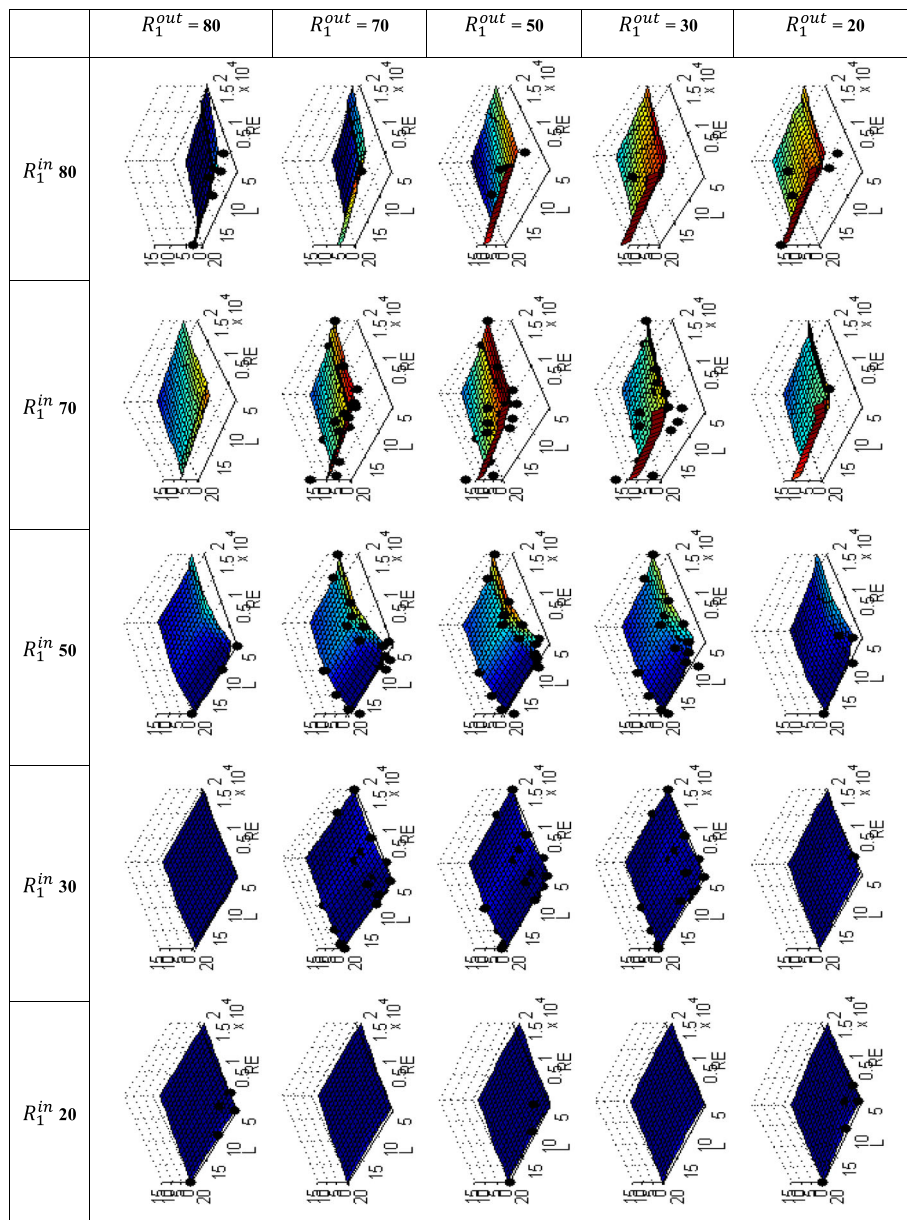
In this paper, the objective was to improve water distribution quality transport modeling for security management in case of contamination events. The focus was

made on mixing conditions at Double T-junctions. Numerous 3-D CFD simulations of different configurations of Double T-junction mixing were performed and a 1-D law for imperfect mixing was derived that is added to the full network transport model. The mixing was explored using four input parameters: Reynolds number in the interpipe, the length between the two T-junctions and the ratio of Reynolds number at the inlets and at the outlets. As it is impossible to perform computations at every point, a problem-specific Kriging method was developed for interpolating mixing percentages at non-simulated configurations. For sampling design we adapted a Delaunay triangulation method to determine new configurations to simulate.

To complement this study, two CFD simulation tools were mainly used, Code Saturne for direct numerical simulations (DNSs) and Fluent for Large Eddy Simulation (LESs). The results of the simulations were included in a lookup table and were interpolated with the interpolation Kriging method.

The parameters of this method were optimized for the problem by minimizing a log-likelihood function. As the



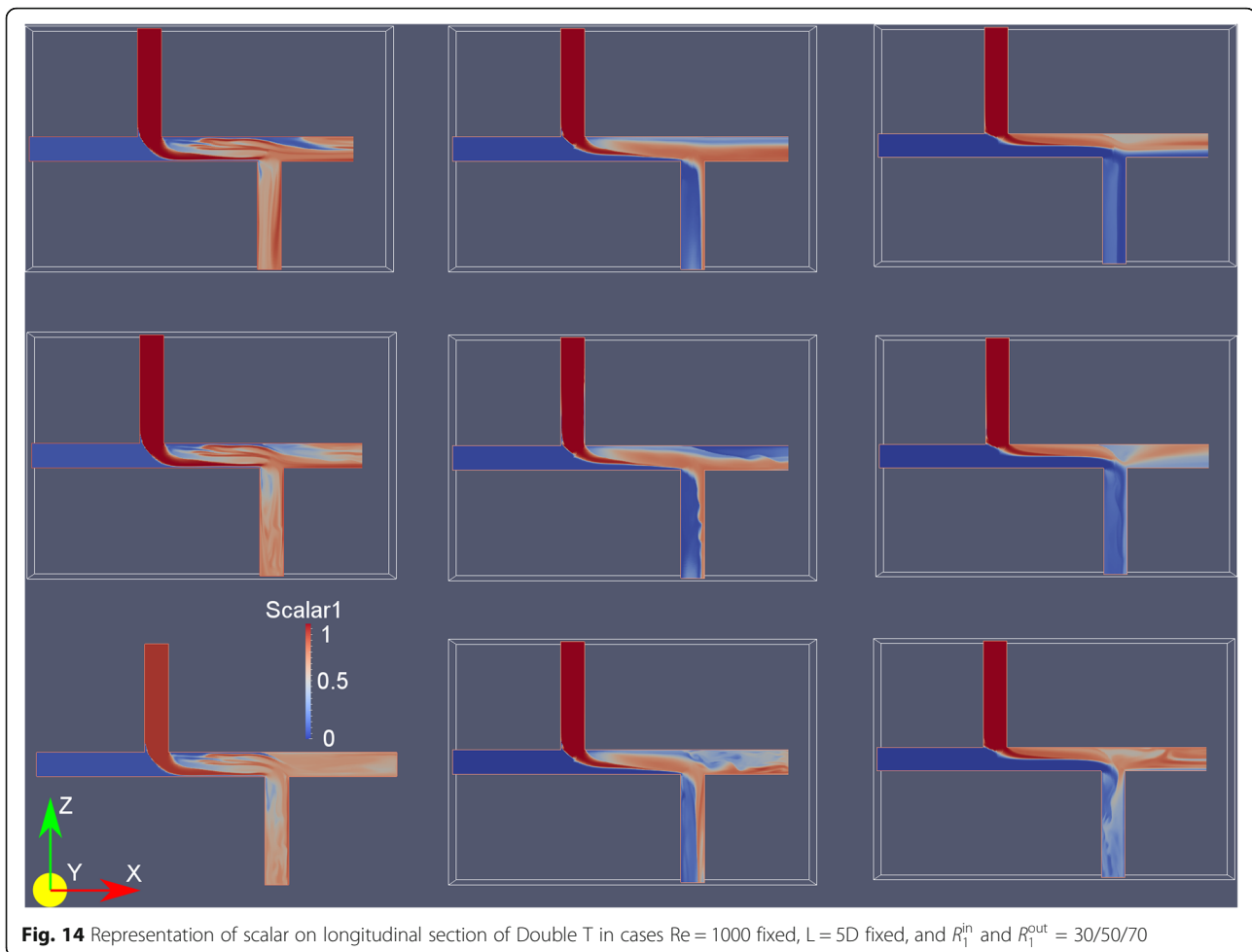


**Fig. 13** Interpolation 4-D Kriging in space L/Re, case  $R_1^{in}$  and  $R_1^{out} = 20/30/50/70/80$

problem is not strictly convex and not defined everywhere, choices were made to simplify the problem allowing a linear system of equations to be solved.

Moreover, a Delaunay method was used to optimize the repartition of simulation points and therefore the sampling design. The Kriging interpolation was then coupled with the Delaunay triangulation method and was compared to a Structured algorithm and performed better with fewer points. It shows that the Delaunay triangulation is a simple and efficient method to find new points of simulation reducing the number of calculations to be made.

The results were interpreted by fixing both the Reynolds number and the length of the interpipe, or both the Reynolds number ratio at the inputs and at the outputs. The contamination was introduced through one of the two inlets, namely inlet In2. The results show that perfect mixing still occurs when inlet In2 is dominating. It is due to the fluid coming from the input 2 going straight to the wall, creating a lot of turbulence. In the other cases, we observe imperfect mixing with a peak at around  $R_2^{in} = 30$ . It can also be seen that when the length increases the



**Fig. 14** Representation of scalar on longitudinal section of Double T in cases  $Re = 1000$  fixed,  $L = 5D$  fixed, and  $R_1^{in}$  and  $R_1^{out} = 30/50/70$

mixing becomes perfect for all configurations, except in laminar cases where imperfect mixing can still be observed for a length of 20 diameters. In the same way, the increase of the Reynolds number decreases the imperfect mixing impact, but at some point, it converges to a value that may not be the complete mixing.

Finally, a 1-D transport model was created and implemented. It considers advection/reaction along pipes, imperfect mixing for Double T-junctions, and perfect mixing for simple junction nodes. For imperfect mixing the mean cross-sectional concentration at the two outlets was calculated. The result depends on the Reynolds number at half distance, on the two inlet and outlet flow rate ratios and on the interpipe length. A lookup table was deduced from the CFD simulations, and the Kriging method was used for points that are not in the table. In cases of imperfect mixing, the interpipe length is most likely small and therefore the transport inside such pipe is neglected. The concentrations at the beginning of the outlet pipes depend directly on the flow rates and concentration of the

end of the inlet pipes based on the 1-D law defined in this paper.

This article gives a procedure to create a reduced order model for any problem with high computing demanding simulations and multiple parameters. The first step is to define a basic design of experiment, for instance a coarse design of points. From those simulations, the Kriging parameters are calibrated solving an optimization problem. Then the Delaunay triangulation method is used to get new simulation points that reduce the overall error of interpolation. Afterward, a lookup table is filled for values with smallest numerical error. Finally, this table is used with the Kriging interpolation to define a 1-D law that defines the reduced order model. This model can be used in real time as contrary to the simulations.

This study was performed for the general case of Double T-junctions, composed of pipes with the same diameter, and no slip-wall condition in CFD simulations. Future research will consist of generalization of the mixing law to pipes of unequal diameters and for pipes with

large roughness that influences the turbulence inside the Double T-junctions.

**Appendix**

**Regression model and maximum-likelihood parameter estimation**

In the section Kriging Interpolation solution, we have seen how to solve the Kriging interpolation. In this Appendix the calibration of different parameters  $\sigma$ ,  $\beta$ ,  $\theta_k$  is clarified. Some elements about such a procedure have been given in the literature [11], but details are missing. We show here how the calibration of the Kriging interpolation parameters is applied in our case.

The Kriging parameters ( $\sigma$ ,  $\theta_l$  and  $\gamma_k$ ) need to be calibrated. For that, it is supposed that the probability density function associated to  $Y$  on the sample points  $s_1, s_2, \dots, s_n$  is a multidimensional Gaussian:

$$\frac{1}{(2\pi)^{\frac{n}{2}} \det(K)^{\frac{1}{2}}} e^{-\frac{1}{2}(Z-F\beta)^T K^{-1} (Z-F\beta)} \tag{16}$$

With

$$K = (K_{ij})_{i,j \in [1,m]}, \quad Z = (Z(s_i))_{i \in [1,n]}, \quad \text{et } F = (F_l(s_i))_{i \in [1,n], l \in [1,p]}$$

The  $Y$  probability density function can be rewritten as:

$$\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}} \det(R)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}(Z-F\beta)^T R^{-1} (Z-F\beta)}$$

With

$$R = (R(Y(s_i), Y(s_j)))_{i,j \in [1,n]}$$

$\sigma$ ,  $\beta$ ,  $\theta_k$ ,  $\gamma_k$  ( $k \in [1, m]$ ) are estimated with the help of a maximum likelihood optimisation. The function log-likelihood to minimise is:

$$f(\sigma, \beta, \theta_k, \gamma_k) = \frac{n}{2} \log(2\pi\sigma^2) + \frac{1}{2} \log(\det(R)) + \frac{1}{2\sigma^2} (Z-F\beta)^T R^{-1} (Z-F\beta) \tag{17}$$

The function  $f$  is differentiated with respect to  $\sigma$  and  $\beta$  to calculate optimality condition by cancelling the gradient function:

$$\frac{\partial f(\sigma, \beta, \theta_k, \gamma_k)}{\partial \sigma} = \frac{n}{\sigma} - \frac{1}{\sigma^3} (Z-F\beta)^T R^{-1} (Z-F\beta)$$

Therefore,

$$\frac{\partial f(\sigma, \beta, \theta_k, \gamma_k)}{\partial \sigma} = 0 \Rightarrow \hat{\sigma} = \frac{1}{n} (Z-F\beta)^T R^{-1} (Z-F\beta) > 0 \tag{18}$$

And

$$\nabla_{\beta} f(\sigma, \beta, \theta_k, \gamma_k) = \frac{1}{2\sigma^2} 2F^T R^{-1} (Z-F\beta) \quad \text{then} \quad \frac{\partial f(\sigma, \beta, \theta_k, \gamma_k)}{\partial \beta} = 0 \Rightarrow \hat{\beta} = (F^T R^{-1} F)^{-1} F^T R^{-1} Z \tag{19}$$

The gradients of function  $f$  with respect to  $\theta_k$  and  $\gamma_k$  are given by:

$$\frac{\partial f(\sigma, \beta, \theta_k, \gamma_k)}{\partial \theta_k} = \frac{1}{2} \text{tr}(R^{-1} [R \cdot D_k]) - \frac{1}{2\sigma^2} (Z-F\beta)^T R^{-1} [R \cdot D_k] \times (Z-F\beta) \tag{20}$$

Where  $\text{tr}$  is the matrix operator trace;

$$[R \cdot D_k]_{ij} = R_{ij} \times (D_k)_{ij}$$

with  $\circ$  the product of Hadamard; and

$$(D_k)_{ij} = -|s_{ik} - s_{jk}|^{\gamma_k} \frac{\partial f(\sigma, \beta, \theta_k, \gamma_k)}{\partial \gamma_k} = \frac{1}{2} \text{tr}(R^{-1} [R \cdot E_k]) - \frac{1}{2\sigma^2} (Z-F\beta)^T R^{-1} [R \cdot E_k] \times (Z-F\beta) \tag{21}$$

Where

$$(E_k)_{ij} = -\theta_k \log(|s_{ik} - s_{jk}|) |s_{ik} - s_{jk}|^{\gamma_k}$$

Finally,  $\sigma$ ,  $\beta$ ,  $\theta_k$ ,  $\gamma_k$  are estimated by solving the following minimization problem:

$$\begin{cases} \min f(\sigma, \beta, \theta_k, \gamma_k) = \frac{n}{2} \log(2\pi\sigma^2) + \frac{1}{2} \log(\det(R(\theta, \gamma))) \\ + \frac{1}{2\sigma^2} (Z-F\beta)^T R(\theta, \gamma)^{-1} (Z-F\beta) \end{cases} \sigma \in \mathbb{R}^+, \beta \in \mathbb{R}^{+p}, \theta \in \mathbb{R}^{+m}, \gamma \in \mathbb{R}^{+m} \tag{22}$$

It is solved using the Trust Region Optimization [8] by giving it the gradients which permits finding the minimum of an unconstrained multivariable function. From a first trust region it then expands or contracts the region by comparing the predicted and actual improvement of the objective function realization. As the problem is not strictly convex and not defined everywhere as seen in the Result parts, choices are needed to simplify the problem.  $\sigma$  and  $\beta$  are taken as optimal, values of  $\gamma_k$  components have all been fixed arbitrary to 1, as no significant change has been seen around this value, and the  $\theta_k$  have been

expressed with regards to a non-dimensional parameter scaled by the maximum distance in all four directions.

$$\theta_k = \tilde{\theta} / \max(|s_{ik} - s_{jk}|)^{Y_k} \quad (23)$$

With  $\tilde{\theta}$  the non-dimensional parameter that needs to be optimized. This last equation is chosen arbitrary and implies a scale relationship between the  $\theta_k$  that reduce by four the problem dimension.

A second one is the regularization term defined thereafter. In this case, and considering the observations collected, the objective function may be undefined, due to the correlation matrix  $R$  for small  $\theta$ , which is not an invertible matrix. Therefore, a regularization term is added to the diagonal of the correlation matrix  $R$ :

$$R_\alpha = R + \alpha I_n$$

With  $\alpha$  the regularization term. This is equivalent to making a ridge regression or using a Tikhonov regularization technique, it is also called nugget effect.

The new problem to solve becomes:

$$\left\{ \begin{array}{l} \min_f(\hat{\theta}) = \frac{n}{2} \log(2\pi\hat{\sigma}^2) + \frac{1}{2} \log(\det(R_\alpha(\theta))) \\ + \frac{1}{2\hat{\sigma}^2} (Z - F\hat{\beta})^T R_\alpha(\theta)^{-1} (Z - F\hat{\beta}) \hat{\theta} > 0, \\ \theta = \tilde{\theta} / \max(|s_{ik} - s_{jk}|)^{Y_k} \end{array} \right. \quad (24)$$

#### Acknowledgments

We wish to acknowledge the three following institutes for letting us use their computation cluster to carry out the simulations performed in this project:

- Mésocentre de Calcul Intensif Aquitain (MCIA), Avakas supercomputer.
- Centre Informatique National de l'Enseignement Supérieur (CINES), Jade supercomputer.
- Irstea Clermont computation grid.

This work has been performed in the framework of the SMaRT-OnlineWDM project. The SMaRT-OnlineWDM project is supported by the German Federal Ministry of Education and Research (BMBF; project: 13 N12180) and by the French Agence Nationale de la Recherche (ANR; project: ANR-11-SECU-006).

#### Authors' contributions

All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Irstea, 50, avenue de Verdun, F-33612 Cestas, France. <sup>2</sup>M2N, IMATH, CNAM, 292 rue Saint-Martin, F-75141 Paris Cedex 03, France.

Received: 21 September 2015 Revised: 24 November 2016

Accepted: 2 December 2016 Published online: 20 January 2017

#### References

1. Abraham, W.: On the efficient design of statistical investigations. *Ann. Math. Stat.* **14**(2), 134–140 (1943)

2. Archambeau, F., Méchitoua, N., Sazik, M. Code Saturne: A Finite Volume code for the computation of turbulent incompressible flows – Industrial Applications. *International Journal on Finite Volumes* (2004)
3. Boulos, P.F., Altman, T., Jarrige, P.A., Collevati, F.: An event-driven method for modeling contaminant propagation in water networks. *Appl. Math. Modelling*. **18**, 84–92 (1994)
4. Braun, M., Bernard, T., Ung, H., Piller, O., Gilbert, D.: Model based investigation of transport phenomena in water distribution networks for contamination scenarios. *Procedia Engineering*. **70**, 191–200 (2014)
5. Chauvet, P., Galli, A.: Universal kriging, Centre de Géostatistique - Ecole des Mines de Paris (France). (1982)
6. Yang, C.-S., Kao, S.-P., Lee, F.-B., Hung, P.-S. Twelve different interpolation methods: a case study of surfer 8.0. National Chung Hsing University, Taichung, Taiwan, ROC and Feng Chia University, Taichung, Taiwan, ROC (2004)
7. Choi, C.Y., Shen, J.Y., Austin, R.G.: Development of a Comprehensive Solute Mixing Model (AZRED) for Double-Tees, Cross, and Wye Junctions. *Water Distribution Systems Analysis* **2008**, 1–10 (2008)
8. Conn, A.R., Gould, N.I.M., Toint, P.L.: Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM J. Numer. Anal.* **25**(2), 433–460 (1998)
9. Constans, S., Bremond, B., Morel, P.: Simulation and control of chlorine levels in water distribution networks. *J. Water Resour. Plann. Manag.* **129**(2), 135–145 (2003)
10. Costa, J.-P., Pronzato, L., Thierry, E. A comparison between Kriging and Radial Basis Function networks for nonlinear prediction. Conference: Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP'99), Antalya, Turkey, June 20–23, 1999.
11. Cressie: *Statistics for spatial data*. Wiley, New York (1993)
12. Fabrie, P., Gancel, G., Mortazavi, I., Piller, O.: Quality modeling of water distribution systems using sensitivity equations. *J. Hydraul. Eng.* **136**(1), 34–44 (2010)
13. Ho, C.K., Khalsa, S.S.: EPANET-BAM: water quality modeling with incomplete mixing in pipe junctions. In: Proceedings of the Water Distribution System Analysis Conference, Kruger National Park, South Africa. (SAND2008-3065C), p. 11. (2008)
14. Yin, J., Ng, S.H., Ng, K.M. A study on the effects of parameter estimation on kriging model's prediction error in stochastic simulations. Proceedings of the Winter Simulation Conference (2009)
15. Kay, M.D., Beckman, R.J., Conover, W.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**(2), 239–245 (1979)
16. Porteau, <http://porteur.irstea.fr/>. Accessed on 22 Nov 2016.
17. Rossman, L.A. EPANET users' manual. U.S. Environmental Protection Agency, National Risk Management Research Laboratory, Office of Research and Development, Cincinnati (2000)
18. Ung, H., Piller, O., Gilbert, D.: Quasi-real time modeling for security of a drinking water distribution network. *Procedia Engineering*. **70**, 800–809 (2014)
19. Waeytens, J., Chatellier, P., Bourquin, F.: Inverse computational fluid dynamics: influence of discretisation and model errors on flows in water network including junctions. *J. Fluids Eng.* **137**(9), 17 (2015)
20. Wim C.M. van Beers, Jack P.C. Kleijnen. Kriging interpolation in simulation; a survey. Proceedings of the Winter Simulation Conference (2004)

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)