Gaussian fields for predicting the drift of oil&gas pipes for deep-water applications

Luca Pinciroli[1], Michele Compare[2], Enrico Zio[3], Gustavo Almeida[4], and Pedro Filgueiras[5]

[1]Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy.
E-mail: luca.pinciroli@polimi.it

[2]Aramis S.r.l., Via Bastioni di Porta Nuova 21, 20121, Milan, Italy.

[2]Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy.
E-mail: michele.compare@aramis3d.com

[3]Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156, Milan, Italy.

[3]Aramis S.r.l., Via Bastioni di Porta Nuova 21, 20121, Milan, Italy.

[3]MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France.

[3]Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea.
E-mail: enrico.zio@polimi.it

[4]Vallourec R&D, Rio de Janeiro, Brazil.
E-mail: g.almeida@vallourec.com

[5]Vallourec, 64 Rue de Leval, 59620, Aulnoye-Aymeries, France.
E-mail: pedro.filgueiras@vallourec.com

ABSTRACT

In the Oil & Gas industry, empirical models are used to estimate the drift of pipes for deep-water applications. These models encode pipe geometrical features, measured along the pipe length. If the estimated drift does not meet the operability requirements, then the pipe is rejected. This improves the quality of the purchased pipes, but strongly affects their production costs. We rely

on the Gaussian fields theoretical framework to address two issues: the a priori estimation of the probability of pipes rejection and the a posteriori estimation of the drift conformance probability, given the actual measured parameters. These are fundamental pieces of information for marketing decisions. A case study is considered to show the application of the theoretical framework. This is based on real pipe measurement data, which have been opportunely re-scaled to avoid the disclosure of relevant information.

1 INTRODUCTION

To meet the increasing Oil & Gas (O&G) demand of the last decades, exploration has been continuously increased, also in deep waters (Di Maio et al. 2017). Deep-water wells require structural components such as casing and tubing to work in extremely harsh operating conditions, characterized by very large pressure values. These challenge both the equipment operability and its capability of preventing catastrophic consequences for the environment, and heavy reputational and financial losses to the well operators. For this, different studies have been carried out in recent years, which resulted in the definition of empirical models guaranteeing very high quality of pipes for the specific deep-water conditions (API 2018; ISO/TR-10400 2018). These models relate fundamental quality factors such as collapse pressure, burst, etc., to manufacturing parameters such as pipe outer diameter, wall thickness, etc. (API 2018; ISO/TR-10400 2018), some of which are affected by epistemic uncertainty.

On the other side, the widespread digitalization process that is driving the fourth industrial revolution (Lasi et al. 2014) has enabled the O&G pipe industry to collect accurate measurements of the geometrical features entering the empirical models defining the pipe quality. In practice, then, a produced pipe is not purchased if the measured parameters lead to an estimate of the relevant pipe quality factors not meeting the corresponding design requirements, as these undermine the pipe reliability and operability.

The direct consequences of this measuring procedure are, on the one hand, the enhancement of the quality of the purchased pipes and, on the other hand, the increase of the pipe production costs, in an O&G industry context where currently there is a urgent need for cost reduction (Crooks 2016;

Hovem 2019).

Among the relevant pipe quality factors, we focus on pipe drift (ISO-11960:2014 2014). Roughly, this is a measure of the roundness of the inside wall of the pipe. When pipe drift is out of specification, tools, pumps, smaller pipes and other items are no longer guaranteed to pass through the pipe. Pipe drift, then, heavily impacts on the well operability and economic performances.

For pipe drift, we develop a novel framework based on the Gaussian field theory, which exploits the measurement data of pipe geometric parameters gathered from a mill to achieve a twofold objective:

- Estimation of the rejection probability, given the design of the pipe and its parameters. This is an a priori estimation, i.e., provided before pipe production, which relates to the capability of the mill of producing pipes with small tolerances. This issue is framed as an Excursion Probability Estimation Problem (EPEB, (Adler 2000)): we determine the probability that the drift random field gets out of specification. EPEBs are of great interest in many engineering fields such as aeronautics (Hoblit 1988) and hydrology (Garcia 2016), to cite a few.

- Estimation of the drift over a specific pipe. This is an a posteriori estimation, which maps the actual measurements of the drift model parameters onto the drift conformance probability (i.e, the probability of having drift values meeting the requirements over the entire pipe (for Standardization. International Electrotechnical Commission 2012)). The a posteriori estimation of the drift conformance probability is tackled through a Bayesian updating of the field, based on the gathered evidence of the measurements.

These a priori and a posteriori estimations are fundamental drivers for the decisions by the pipe manufacturer about the proper positioning in the market, as they can be used to relate the pipe production cost to the pipe quality, targeted at a specific application.

To the Authors' best knowledge, this is the first time that the Gaussian field theoretical framework is proposed for the estimation of pipes drift. Moreover, an additional contribution of this work lies in that we rely on numerical approaches to derive the a priori and a posteriori estimations, rather than

making assumptions on the shape of the Gaussian field kernel, which is expected to significantly change from one setting to another. This choice allows applying the methodological framework to classes of products, although it prevents from straightforwardly applying known techniques such as kriging (e.g., (Williams and Rasmussen 2006; O'Hagan 2006)). Finally, to strengthen the results, we also propose a sensitivity analysis to explore the economic effects of reducing the number of measurement points and their accuracy.

The remainder of this paper is organised as follows. In Section 2, the pipe drift model is sketched. In Section 3, the developed methodology for both the a priori and a posteriori estimations is presented. Section 4 reports the results obtained by applying the proposed methodology to a case study derived from a dataset of real pipe measurements. To protect intellectual property, the original data have all been modified by applying some corrective factors. In Section 5, the sensitivity analysis is proposed. Section 6 concludes the work.

2 PIPE DRIFT

Consider a pipe of length L mm. We assume the availability of a mathematical model that allows evaluating the pipe drift exploiting the measurements of geometrical parameters acquired by means of sensors at pipe sections $\boldsymbol{\chi} = \{\chi_1, ..., \chi_d\} \subset \Omega = [0, L] \subset \mathbb{R}$, equally spaced at a distance $\Delta = L/d$ mm. In this work, we assume that the available drift model is generally defined as:

$$Dr = f\left(\mathbb{E}_{l_{Dr}}\left[Wt_{av}\right], \mathbb{E}_{l_{Dr}}\left[Wt_{max}\right], \mathbb{E}_{l_{Dr}}\left[Od_{av}\right], \boldsymbol{\epsilon}_{Dr}\right) \tag{1}$$

where the used pipe geometrical parameters are:

- Average wall thickness, Wt_{av};
- Maximum wall thickness, Wt_{max};
- Average outer diameter, Od_{av};

and

- $\boldsymbol{\epsilon}_{Dr}$ is a vector of random variables representing the epistemic uncertainties in the drift model;
- $\mathbb{E}_{l_{Dr}}[Od_{av}], \mathbb{E}_{l_{Dr}}[Wt_{av}]$ and $\mathbb{E}_{l_{Dr}}[Wt_{max}]$ represent the average values over a critical length l_{Dr} mm of the average outer diameter, average wall thickness and maximum wall thickness, respectively. In particular, the estimations of these values at pipe section χ_i consider its first $N^{Dr} = \lceil l_{Dr}/\Delta \rceil$ subsequent sections, $i = 1, ..., d - N^{Dr} + 1$, where $\lceil \circ \rceil$ indicates the ceiling value of its argument. For confidentiality, the value of the critical length l_{Dr} is not reported.

Notice that in Eq. (1) it is assumed that the smaller the value of Dr, the worsen is the pipe quality.

3 METHODOLOGY

Assume that a dataset is available containing the geometrical parameters of N_T pipes, measured at points $\boldsymbol{\chi}$. We first rely on Eq. (1) to propagate the uncertain quantities $\boldsymbol{\epsilon}_{Dr}$ onto the drift values, through the Monte Carlo method. Namely, at every measurement location $\chi_i \in \boldsymbol{\chi}_D = \{\chi_1, ..., \chi_N\}$, where $N = d - N^{Dr} + 1$, we sample $N_{MC} >> 1$ (e.g., $N_{MC} = 10^7$) values from the distribution of $\boldsymbol{\epsilon}_{Dr}$. These samples enter the drift model to get the corresponding N_{MC} values of Dr, which determine the distribution of the drift assigned to the pipe section. Then, any quantile $p \in]0, 1[$ of this distribution can be considered, defining the random variable $X_p(\chi_i)$. This can be framed as the drift value at pipe section χ_i of the worst p-th portion of produced pipes.

To simplify the notation, we indicate by $X_{p,i}$ the variable $X_p(\chi_i)$, $i = 1, ..., N$.

The vector of random variables $X_{p,i}$, $i = 1, ..., N$, and the corresponding measured values on pipe j, $j = 1, ..., N_T$, $p \in]0, 1[$, are indicated by, respectively:

$$\mathbf{X}_p = \left[X_{p,1}, ..., X_{p,N}\right] \tag{2}$$

$$\mathbf{x}_p^j = \left[x_{p,1}^j, ..., x_{p,N}^j\right] \tag{3}$$

For every pipe section $\chi_i \in \boldsymbol{\chi}_D$, we can estimate the average value and standard deviation of

$X_{p,i}$ over the N_T pipes:

$$\mathbb{E}[X_{p,i}] \approx \frac{\sum_{j=1}^{N_T} x_{p,i}^j}{N_T} \tag{4}$$

$$\sigma_{X_{p,i}} \approx \sqrt{\frac{\sum_{j=1}^{N_T} \left(x_{p,i}^j - \mathbb{E}[X_{p,i}]\right)^2}{N_T - 1}} \tag{5}$$

On this basis, we can derive the non-negative definite covariance matrix $\mathbf{K}_p(\mathbf{X}_p, \mathbf{X}_p)$, which can be also referred to as kernel, whose (i, k) entry reads:

$$\begin{aligned} K_p(i,k) &= \mathbb{E}\left[\left(X_{p,i} - \mathbb{E}[X_{p,i}]\right) \cdot \left(X_{p,k} - \mathbb{E}[X_{p,k})\right)\right] \\ &\approx \frac{\sum_{j=1}^{N_T} \left(x_{p,i}^j - \mathbb{E}[X_{p,i}]\right) \cdot \left(x_{p,k}^j - \mathbb{E}[X_{p,k}]\right)}{N_T - 1} \end{aligned} \tag{6}$$

Finally, we define the correlation coefficient

$$\rho_p(i,k) = \frac{K_p(i,k)}{\sigma_{X_{p,i}} \cdot \sigma_{X_{p,k}}} \tag{7}$$

To check that the Gaussian field theoretical framework is applicable to address both issues in Section 1, we have to verify that:

- The process is wide-sense stationary. This means that $\mathbb{E}[X_{p,i}]$ is constant for any $\chi_i \in \boldsymbol{\chi}_D$ and the covariance function is invariant to translations, i.e., it depends only on $\tau = \chi_i - \chi_j$ (Papoulis and Pillai 2002; Williams and Rasmussen 2006). We verify this through the Augmented Dickey-Fuller (Dickey and Fuller 1981) and the Phillips-Perron (Phillips and Perron 1988) tests, whose null hypothesis is that a unit root is present in a time series sample, whereas the alternative hypothesis is trend-stationarity. The constant mean is indicated by $\mathbb{E}[X_p(\chi)] = \mu_p, \forall \chi \in \Omega, p \in]0, 1[$.
- The distribution of $X_{p,i}$ is normal for $i = 1, ..., N$. To check this, we can rely on normality tests such as Shapiro-Wilk (Plackett 1983a), Chi-squared (Plackett 1983b) or Lilliefors tests

(Lilliefors 1967). We can state that the process is Gaussian when the number n_p of sections in which the test is positive is $n_p \approx N$.

If these two conditions are verified, we have a Gaussian random field in Ω, such that $\mathbf{X}_p \sim \mathcal{N}\left(\mathbf{M}_p, \mathbf{K}_p\right)$ is multivariate normal for any set $\chi_D = \{\chi_1, ..., \chi_N\} \subset \Omega$ (Adler and Taylor 2009), for any $N \in \mathbb{N}$, where $\mathbf{M}_p \in \mathbb{R}^N$ is the vector containing N times the mean value μ_p. The Probability Density Function (pdf) of $\mathbf{X}_p$ reads

$$f_{\mathbf{X}_p}\left(\mathbf{x}_p\right) = \frac{1}{\sqrt{(2\pi)^N det\mathbf{K}_p}} e^{-\frac{1}{2}(\mathbf{x}_p - \mathbf{M}_p)^T \mathbf{K}_p^{-1}(\mathbf{x}_p - \mathbf{M}_p)} \tag{8}$$

where $\mathbf{x}_p = \left(x_{p,1}, ..., x_{p,N}\right) \in \mathbb{R}^N$, $det\mathbf{K}_p$ is the determinant of $\mathbf{K}_p(\mathbf{X}_p, \mathbf{X}_p)$, whereas $\mathbf{K}_p^{-1}$ is its inverse matrix.

With respect to the estimation of $\mathbf{K}_p$, $p \in]0, 1[$, one way to proceed is to check whether one of the analytical kernel models available in the literature (e.g., square exponential kernel, rational quadratic kernel, periodic kernel, combinations of them, etc. (Williams and Rasmussen 2006; Duvenaud et al. 2013)) is suitable to represent the covariance function $K_p(\tau)$ between the measurement points. However, we avoid making this assumption and, then, derive the covariance matrix numerically, only. This allows applying the same procedure to different classes of products, which are expected to have different kernel shapes. On the other hand, this choice prevents us from straightforwardly applying known techniques such as kriging (e.g., (Williams and Rasmussen 2006; O'Hagan 2006)) and arises numercal issues related to the positive definiteness of the covariance matrix (Williams and Rasmussen 2006).

Consider two points χ_i and $\chi_k \in \chi_D$ such that their distance is $\tau = |i - k|$. Then, we estimate the covariance as a function of the distance τ by evaluating the average value of the covariance between every pair (χ_i, χ_k), such that their distance is equal to τ, for any $\tau \leq N$ (Papoulis and Pillai 2002):

$$K_p(\tau) \approx \frac{\sum_{i=1}^{N-\tau} K_p(i, i+\tau)}{N - \tau} + \frac{\sum_{i=1}^{N-\tau} \left(\mathbb{E}[X_{p,i}] - \mu_p\right) \cdot \left(\mathbb{E}[X_{p,i+\tau}] - \mu_p\right)}{N - \tau} \tag{9}$$

where the second addend can be disregarded if it is small. For wide-sense stationary Gaussian processes, we can estimate the power spectral density as:

$$F_p(\omega) = \int_{-\infty}^{\infty} K_p(\tau) e^{-j\omega\tau} d\tau - 2\pi\mu_p\delta(\omega) = \int_0^{\infty} K_p(\tau) cos\omega\tau d\tau - 2\pi\mu_p\delta(\omega) \tag{10}$$

where the second equality holds because $K_p(\tau)$ is real, whereas $\delta(\omega)$ indicates the Dirac's delta.

3.1 Pipe rejection probability estimation

For any quantile $p \in]0, 1[$, the pipe rejection probability can be estimated as the probability that the drift of the worst p-th portion of the pipe population goes under the drift conformity requirement threshold $T_D \in \mathbb{R}$ in at least one point $\chi \in \Omega$. This is the excursion probability (Adler and Taylor 2009):

$$P\left\{\inf_{\chi\in\Omega} X_p(\chi) \leq T_D\right\} \tag{11}$$

Notice that we are now considering the entire pipe length, rather than a discrete set of its points. According to Adler (Adler 2000), there is no explicit formula for estimating this probability value in the general Gaussian case, despite the fact that it appears in a variety of problems. Nevertheless, the following approximation holds:

$$\begin{aligned} P\left\{\inf_{\chi\in\Omega} X_p(\chi) \leq T_D\right\} &= P\left\{U_p^{T_D}(\Omega) \geq 1 \cup X_{p,1} \leq T_D\right\} \\ &\leq P\left[U_p^{T_D}(\Omega) \geq 1\right] + P\left\{X_{p,1} \leq T_D\right\} \\ &\leq \mathbb{E}\left[U_p^{T_D}(\Omega)\right] + P\left\{X_{p,1} \leq T_D\right\} \end{aligned} \tag{12}$$

where $U_p^{T_D}(\Omega)$ is the number of downcrossing by X_p of the conformity requirement threshold T_D in Ω, whereas $P\left\{X_{p,1} \leq T_D\right\} = P_p^0$ is needed since $\mathbb{E}\left[U_p^{T_D}(\Omega)\right]$ does not consider that the initial point of X_p can be positioned under T_D.

The issue of estimating the mean of $U_p^{T_D}(\Omega)$ has been addressed in (Adler 2000; Williams and Rasmussen 2006; Adler 2010; Lindgren 2006). In particular, Rice (Rice 1944), proposed a formula

for stationary and almost surely continuous Gaussian processes on $\Omega = [0, 1]$, which has been extended and generalized (e.g., (Bulinskaya 1961; Ito et al. 1963; Ylvisaker 1965)):

$$\mathbb{E}\left[U_p^{T_D}(1)\right] = \frac{1}{2\pi}\sqrt{\frac{M_p^2}{M_p^0}}e^{-\frac{(T_D-\mu_p)^2}{2M_p^0}} \tag{13}$$

where μ_p is the mean of the Gaussian process and M_p^0 and M_p^2 are the spectral moments ((Adler and Taylor 2009; Lindgren 2006))

$$\begin{aligned} M_p^0 &= \int_0^\infty F_p(\omega)d\omega = K_p(0) = \mathbf{K}_p(i,i) \quad i = 1, ..., N \\ M_p^2 &= \int_0^\infty \omega^2 F_p(\omega)d\omega \end{aligned} \tag{14}$$

The lack of knowledge of the analytical formula of the kernel requires applying a numerical approach for the estimation of M_p^2: we use the Fast Fourier Transform (FFT) algorithm (Sorensen et al. 1987) to numerically evaluate the power spectral density, $F_p(\omega)$ in Eq. (10), which is, then, integrated through the trapezoidal rule for approximating the definite integral in Eq. (14) (Atkinson 1989).

For many types of random processes, including Gaussian, the number of downcrossings of T_D converges to a Poisson process as T_D decreases (Leadbetter et al. 2012). The interarrival "times" over $[0, \chi]$ of this Poisson process are exponentially distributed with rate equal to $\mathbb{E}\left[U_p^{T_D}(1)\right] + \frac{P_p^0}{\chi}$ (Leadbetter et al. 2012). If the number of downcrossings of T_D obeys a Poisson process, then the probability $P_p^{T_D}(\chi)$ that T_D has been crossed by X_p at least once in $[0, \chi] \subseteq \Omega$ is (Papoulis and Pillai 2002; Leadbetter et al. 2012):

$$P_p^{T_D}(\chi) = 1 - e^{-\left(\mathbb{E}\left[U_p^{T_D}(1)\right]\chi + P_p^0\right)} \tag{15}$$

If $\mathbb{E}\left[U_p^{T_D}(1)\right]\chi + P_p^0$ is small, then we can write (Papoulis and Pillai 2002; Leadbetter et al. 2012):

$$P_p^{T_D}(\chi) \approx \mathbb{E}\left[U_p^{T_D}(1)\right]\chi + P_p^0 \tag{16}$$

3.2 Drift conformance probability estimation for a specific pipe

Once the Gaussian field is characterized by its mean and covariance, it can be used to estimate the drift conformance probability of the specific pipe, based on the acquired measurements. For generality, we assume that the sensors for measuring the pipe characteristrics and the positions they are located on are not necessarily the same of those used to estimate the field. Namely, we indicate by $\mathbf{Y}_p = \left[Y_p(\chi_{i_1}), ..., Y_p(\chi_{i_{n_{Y_p}}})\right]$ the measured values of drift available at a subset $\mathbf{\Phi}_p \subseteq \chi_D = \{\chi_1, ..., \chi_N\}$ of n_{Y_p} pipe sections, $p \in]0, 1[$. We assume

$$\mathbf{Y}_p = \mathbf{A}_p \mathbf{X}_p + \boldsymbol{\epsilon} \tag{17}$$

where $\mathbf{A}_p$ is a $[n_{Y_p} \times N]$ matrix defining the elements of $\mathbf{X}_p$ that are observed, whereas $\boldsymbol{\epsilon}$ is a zero-mean Gaussian random noise vector in $\mathbb{R}^{n_{Y_p}}$, with covariance matrix $\mathbf{\Sigma}_\epsilon$, i.e., $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{\Sigma}_\epsilon)$. Although more complex relationships can be modeled by Eq. (17), for simplicity we assume that the entries of $\mathbf{A}_p$ are binaries $(i, k) \in \{0, 1\}$ such that $\sum_{k=1}^{N} \mathbf{A}_p(i, k) = 1$ and $\mathbf{\Sigma}_\epsilon = \sigma_\epsilon^2 \mathbf{I}_{n_{Y_p}}$.

The probability distribution $f_{\mathbf{Y}_p|\mathbf{X}_p}$ of $\mathbf{Y}_p$ conditional to the given random field is still normal, being $\mathbf{Y}_p$ a sum of normal random variables:

$$f_{\mathbf{Y}_p|\mathbf{X}_p} \sim \mathcal{N}\left(\mathbf{A}_p \mathbf{M}_p, \mathbf{\Sigma}_{\mathbf{Y}_p}\right) \tag{18}$$

where $\mathbf{\Sigma}_{\mathbf{Y}_p} = \mathbf{A}_p \mathbf{K}_p \mathbf{A}_p^T + \mathbf{\Sigma}_\epsilon$ (Malings 2017; O'Hagan 2006).

Bayesian inference allows defining the posterior pdf of the random field conditional to these observations (O'Hagan 2006):

$$f_{\mathbf{X}_p|\mathbf{Y}_p} \propto f_{\mathbf{Y}_p|\mathbf{X}_p} \cdot f_{\mathbf{X}_p} \tag{19}$$

Namely, based on evidence $\mathbf{Y}_p$, the joint probability of the values of $\boldsymbol{X}_p$ in points $\boldsymbol{\chi}_D$ reads (Malings 2017)

$$\mathbf{X}_p|\mathbf{Y}_p \sim \mathcal{N}\left(\mathbf{M}_{p|\mathbf{Y}_p}, \mathbf{\Sigma}_{p|\mathbf{Y}_p}\right) \tag{20}$$

where

$$\mathbf{M}_{p|\mathbf{Y}_p} = \mathbf{M}_p + \mathbf{K}_p \mathbf{A}_p^T \boldsymbol{\Sigma}_{\mathbf{Y}_p}^{-1} \left(\mathbf{Y}_p - \mathbf{A}_p \mathbf{M}_p \right) \tag{21}$$

$$\boldsymbol{\Sigma}_{p|\mathbf{Y}_p} = \mathbf{K}_p - \mathbf{K}_p \mathbf{A}_p^T \boldsymbol{\Sigma}_{\mathbf{Y}_p}^{-1} \mathbf{A}_p \mathbf{K}_p^T \tag{22}$$

In Figure 1-(a), an example is shown of field updating considering one measurament section χ_i with no uncertainty, i.e., $\mathbf{A}_p \in [1 \times N]$ such that $\mathbf{A}_p(1,i) = 1$ and $\sigma_\epsilon = 0$. Notice that, in this case, the field variability reduces to 0 at the measured point. In fact, it can be derived from Eq. (22) that $\boldsymbol{\Sigma}_{p|\mathbf{Y}_p}(i,i) = 0$. The field variability also shrinks at the other points, due to their dependece on the measured value, as defined by the covariance function.

Once the posterior distribution has been estimated, we can proceed with the evaluation of the drift conformance probability of a specific pipe. To do this, we apply Monte Carlo sampling from the posterior distribution of $\mathbf{X}_p|\mathbf{Y}_p$ and estimate the portion of these samples overcoming the conformity requirement threshold T_D (Figure 1-(b)). This sampling from the multivariate normal distribution $f_{\mathbf{X}_p|\mathbf{Y}_p}$ is relatively easy, as we can rely on the Cholesky decomposition (Press et al. 1993; Murphy 2012; Golub and Van Loan 2012).

If the samples $\tilde{\mathbf{x}} = (\tilde{x}_1, ..., \tilde{x}_N)_\eta$, $\eta = 1, ..., H >> 1$, from the updated field $\mathbf{X}_p$ are all larger than T_D over the entire pipe length, we can state that the pipe will fulfill the conformity requirement threshold T_D with probability at least $1 - p \in]0, 1[$. The larger the value of H, the larger the confidence in this statement.

Formally, for every pipe $j \in \{1, ..., N_T\}$, based on the actual measurements $\mathbf{y}_p^j$ we can define the binary variable

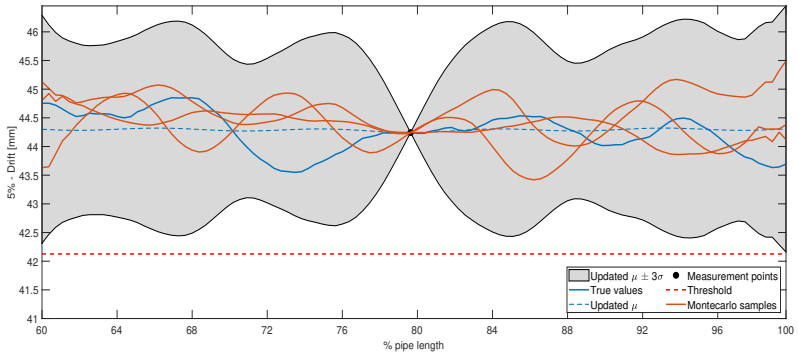
$$\Gamma_p^j = \begin{cases} 1 & \text{if } \min_{\eta=1,...,H} min_{\chi_i \in \chi_D} \left(\tilde{\mathbf{x}}_\eta \longleftarrow \mathbf{X}_p | \mathbf{y}_p^j \right) \geq T_D \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

243 On this basis, the drift conformance probability of pipe j can be defined as:

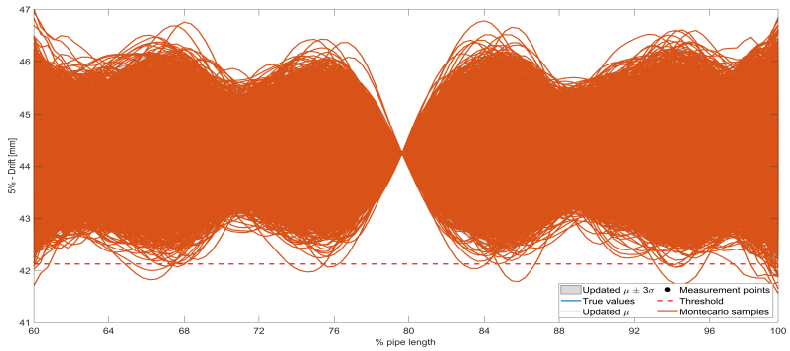
244 $$R^j = 1 - \min\{p \in]0,1[\,|\, \Gamma_p^j = 1\} \tag{24}$$

245 A different perspective is to consider for every quantile $p \in]0,1[$ the portion $\hat{R}_p^j$ of samples $\tilde{\mathbf{x}}$ that

246 are above the threshold T_D. This way, we give the pipe drift conformance probability to requirement



(a)



(b)

Fig. 1. (a) Random field updating by means of noise-free measurements using one measurement point. (b) Representation of all the Monte Carlo samples. For confidentilaity, only the final part of the pipe is shown.

threshold T_D with probability at least $1 - p$. Formally:

$$\Gamma_{p,\eta}^{j} = \begin{cases} 1 & \text{if } min_{\chi_i \in \chi_D} \left(\tilde{\mathbf{x}}_\eta \longleftarrow \mathbf{X}_p | \mathbf{y}_p^j \right) \geq T_D \\ \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

$$\hat{R}_p^j = \frac{\sum_{\eta=1}^{H} \Gamma_{p,\eta}^{j}}{H} \tag{26}$$

Although this definition is more rigorous than that in Eq. (24), nonetheless it is more difficult to manage for decision making.

Notice that to estimate the downcrossing probability, we assume that the values of the drift between two successive sampled values are on the linear interpolation. This assumption is justified by the very large correlation we expect between $X_{p,i}$ and $X_{p,i\pm1}$, $i = 2, .., N$, given that their values share the expectation of the geometrical parameters on $N^{Dr} - 1$ pipe sections. Roughly, this large correlation tells us that if we know the value of the drift in one point, we also know its value in the neighbors.

Finally, notice that this Monte Carlo procedure and the related assumptions can be replaced by easier procedures (e.g., (Williams and Rasmussen 2006; O'Hagan 2006)), if we knew the analytical expression of the kernel.

4 CASE STUDY

The available dataset is composed of the measurements of minimum, average and maximum values of Wt_{av}, Wt_{max} and Od_{av} along the stretch of $N_T = 169$ similar pipes. The pipe length L and the measurement distance Δ are not reported for confidentiality. They lead to have a number of measuement points $d > 100$. Notice that to protect intellectual property, the original data have all been arbitrarily re-scaled, together with the conformity requirement threshold T_D, set to 42.13.

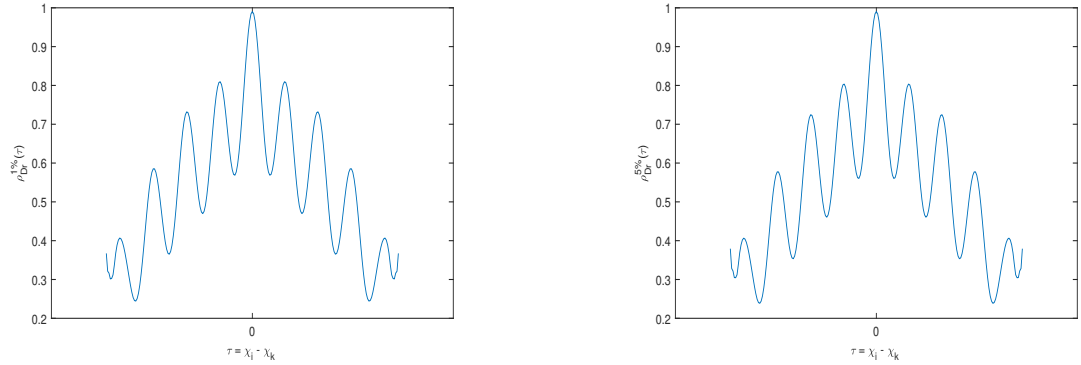
In Table 1, we report the results of the application of the normality tests to the $N_T = 169$ available pipes. The first two columns show the numbers of measurement sections in which

quantiles $p = 0.01$ and $p = 0.05$, respectively, do not present a normal behavior. From these results, we can derive the portion $\frac{n_p}{N}$ of pipe sections in which the behavior can be considered normal (last two columns). We have also verified that the pipe sections where distributions are not proved to be Gaussian are between pipe sections where they are. The closeness of the ratios $\frac{n_p}{N}$ to 1 and the low credibility of having pipe sections with distributions that are different from those of the neighbors lead to conclude that distributions can be considered Gaussian everywhere.

$Test$	$p = 0.01$	$p = 0.05$	$n_{0.01}/N$	$n_{0.05}/N$
$Shapiro - Wilk$	7	8	0.94	0.93
$Chi - squared$	13	6	0.88	0.95
$Lilliefors$	14	8	0.87	0.93

TABLE 1. Results of the normality tests for the 1st and 5th percentiles of the drift (Section 2) for the 169 pipes.

Figure 2 shows the behaviour of the correlation coefficients for $X_{0.01}$ and $X_{0.05}$, which are estimated through the procedure presented in Section 3.



(a) Correlation coefficient of $X_{0.01}$

(b) Correlation coefficient of $X_{0.05}$.

Fig. 2. Correlation coefficient behaviour for $X_{0.01}$ and $X_{0.05}$. Abscissa is not detailed for confidentiality.

To verify the stationarity of the random fields, we use the augmented Dickey-Fuller and the Phillips-Perron tests, which both reject the null hypothesis for both X_p, $p = 0.01$ and $p = 0.05$, with a significance level $\alpha = 0.05$. This confirms the stationarity of the random fields and allows

us using the Rice's formula (Eq. (13)) to evaluate the pipe rejection probability (Table 2).

	$p = 0.01$	$p = 0.05$
$P_p^{T_D}(L)$	0.98	0.04

TABLE 2. Pipe rejection probability

The rejection probability is very large when considering the first percentile, $P_{0.01}^{T_D}(L) \leq 0.98$, whereas it is very small when considering the fifth percentile, $P_{0.05}^{T_D}(L) \leq 0.04$. In words, we are almost sure to reject the worst 1% of the pipes due to drift non-conformity and we are confident that 95% of the pipes almost surely fulfill the drift requirement T_D. It is important to keep in mind that since the original data have been arbitrarily re-scaled, the obtained results are not indicative of the actual production performance. Figures 3 and 4, show the behavior of $P_p^{T_D}(L)$, with $p = 0.01$ and $p = 0.05$, respectively, as a function of the difference between threshold T_D and mean value μ_p. The larger the distance, the smaller the probability of downcrossing. Figures 3 and 4 also show the mean number of downcrossings, $\mathbb{E}\left[U_p^{T_D}(1)\right]$ as obtained from Eq. (13), and the value $P_p^0 = P\left\{X_{p,1} \geq T_D\right\}$, which is, obviously, standard normally distributed.

Finally, Figure 5 and Figure 6 show, respectively, the boxplots of $\mathbf{X}_{0.01}$ and $\mathbf{X}_{0.05}$ along the final portion of the pipe length (the remaning part is not shown for confidentiality). In agreement with the obtained results, $\mathbf{X}_{0.01}$ overlaps $T_D = 42.13$ almost everywhere, whereas $\mathbf{X}_{0.05}$ is almost everywhere far above that same threshold.

With respect to the conformance probability of a specific pipe, we have to update the Gaussian field with the measurements collected from the specific pipe, as described in Section 3.2. Figure 7 shows the posterior distribution along the final portion of the pipe length on one out of the N_T pipes, say j', obtained considering all its available measurements: $\mathbf{A}_p = \mathbf{I}_N$, $\boldsymbol{\epsilon} = 0$. From Eq. (22), we can see that this entails $\boldsymbol{\Sigma}_{p|\mathbf{x}_p^{j'}} = 0$. For this specific pipe j', $\Gamma_{0.05}^{j'} = 1$, whereas $R^{j'} = 0.98$.

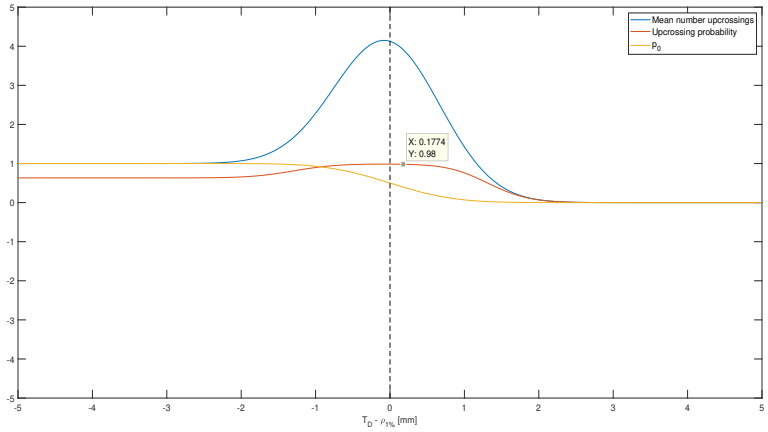


Fig. 3. Rejection probability with respect to $X_{0.01}$ as function of $T_D - \mu_{0.01}$.

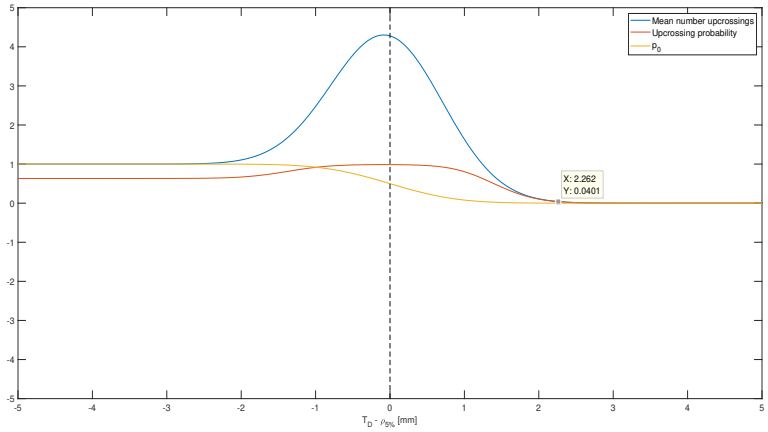


Fig. 4. Rejection probability with respect to $X_{0.05}$ as function of $T_D - \mu_{0.05}$.

5 SENSITIVITY ANALYSIS

To strengthen the results of the case study, we perform a sensitivity analysis on two factors relevant for the proposed methodology:

- The number of measurement points, n_{Y_p}, collected for each pipe;
- The uncertainty in the gathered measurements.

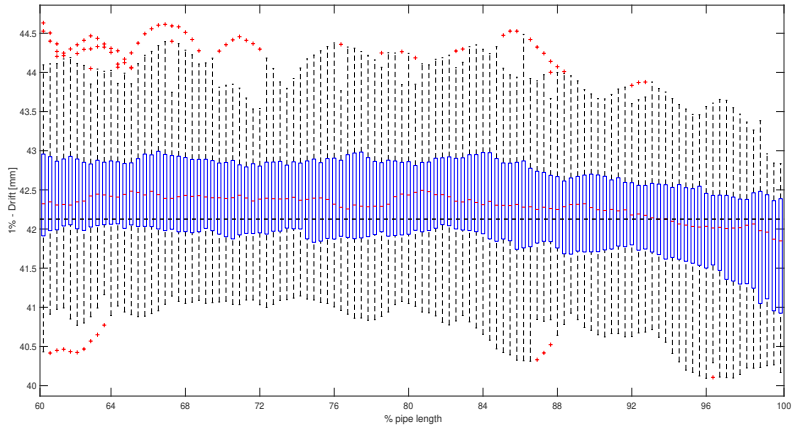


Fig. 5. Boxplot of $X_{0.01}$ along the pipe length.

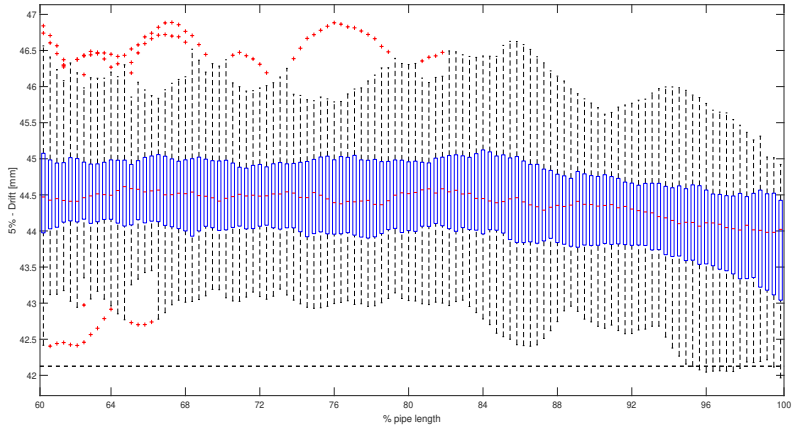


Fig. 6. Boxplot of $X_{0.05}$ along the pipe length.

To analyze the effects of these variables, we consider the utility function C:

$$C = TP \cdot s_{TP} + TN \cdot s_{TN} + FP \cdot s_{FP} + FN \cdot s_{FN} \tag{27}$$

where TP, TN, FP and FN represent the probabilities of true positive (i.e., purchasing of a non-drifting pipe), true negative (i.e., scrapping of an drifting pipe), false positive (i.e., scrapping of a non-drifting pipe) and false negative (i.e., purchasing of a drifting pipe), respectively, whereas s_{TP}, s_{TN}, s_{FP} and s_{FN} represent the corresponding costs. The value of s_{TP} has been arbitrarly set

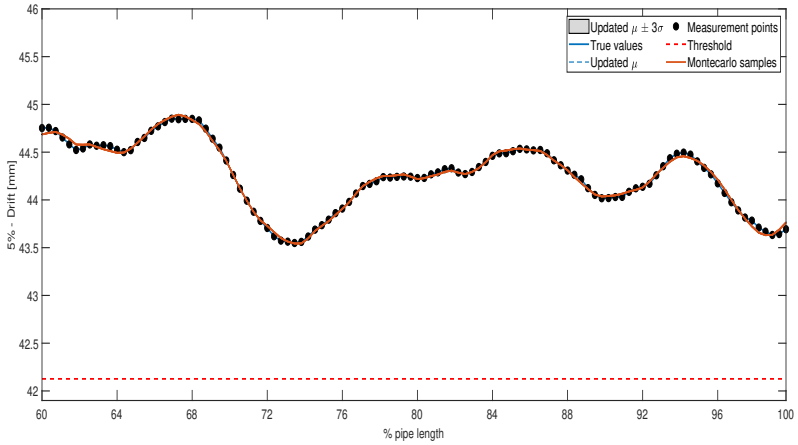


Fig. 7. Random field updating when $\mathbf{Y}_p = \mathbf{x}_p^{j'}$, with $p = 0.05$.

equal to 100\$, whereas the values of s_{TN}, s_{FP} and s_{FN} have been scaled on s_{TP}: $s_{TN} = -0.5\ s_{TP}$, $s_{FP} = -1.5\ s_{TP}$ and $s_{FN} = -5.5\ s_{TP}$.

5.1 Number of measurement points

We investigare the impacts of the number of measurement points on the estimation of:

- The downcrossing probability.
- The drift conformance probability of the specific pipe.

Influence on the a priori estimation of the downcrossing probability

We are now considering the a priori probability. Then, we can assume that $FP = 0$ and $FN = 0$ for any estimated kernel and that s_{TP} and s_{TN} consider the effects of wrong a priori estimates. The reference expected utility $\mathbb{E}[C_\Delta]$ relates to the case in which the set of measurements is complete and with no uncertainty. For example, in case of $p = 0.05$, we have $TN = 0.04$ and $TP = 1 - 0.04 = 0.96$ (Table 2), whereby $\mathbb{E}[C_\Delta] = 0.96 \cdot 100 - 0.04 \cdot 50 = 94$\$.
We estimate the utility $\mathbb{E}[C_\alpha]$ as a function of the increasing distance $\alpha\Delta$, $\alpha > 1$, between two consecutive measurement points. To do this, we re-estimate the covariance function through Eqs. (7 - 9) for each value of α, and, on this basis, re-estimate TP and TN. Figure 8 shows the covariance function for different values of α: the estimations are accurate up to $\alpha \leq 5$.

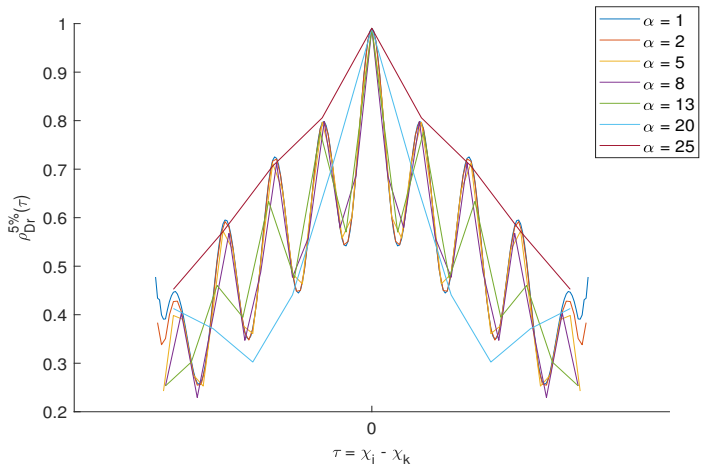


Fig. 8. Correlation coefficient vs distance $\alpha\Delta$ between two consecutive measurement points.

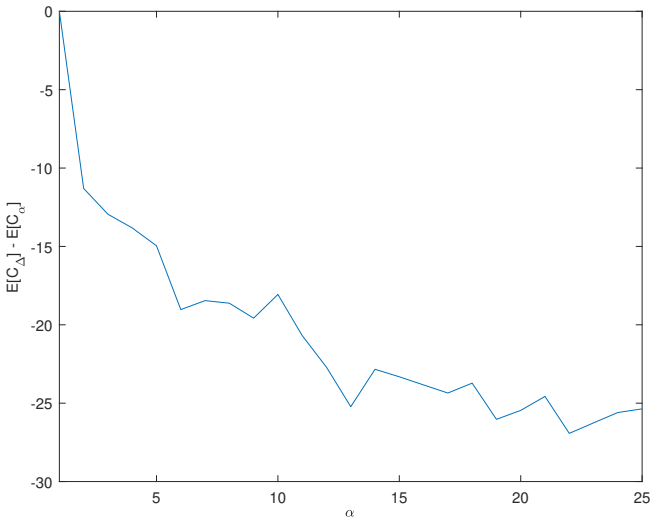
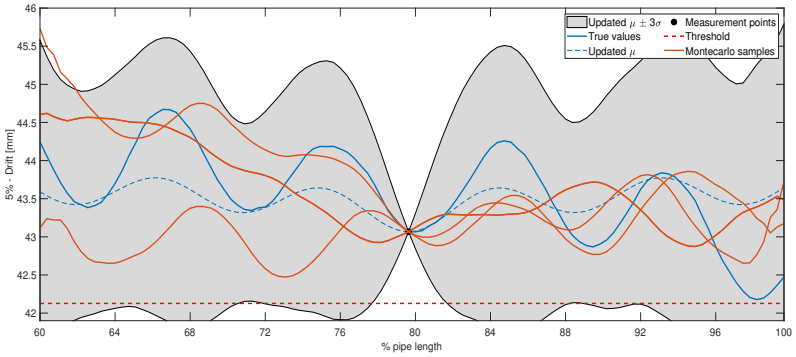
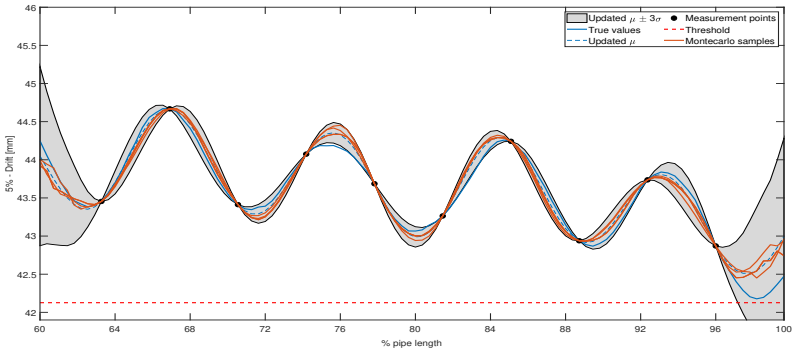


Fig. 9. Cost function as function of distance $\alpha\Delta$ between two consecutive measurement points.

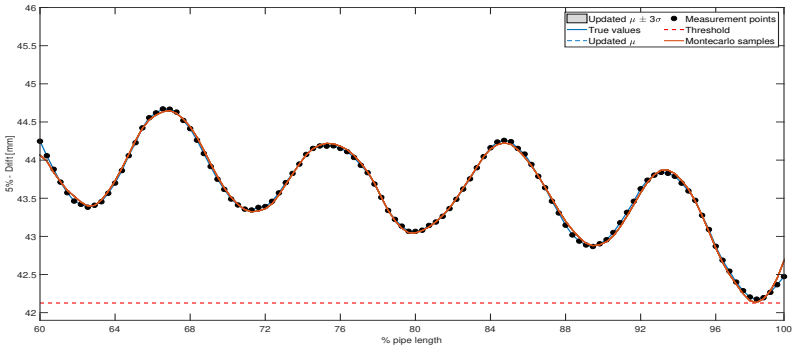
Figure 9 shows the behavior of $\mathbb{E}[C_\Delta] - \mathbb{E}[C_\alpha]$: the larger the value of α, the larger the loss of information about the field; this leads to an increment of the negative term of the cost function, i.e., FN. The non-monotonic behavior in Figure 9 is explained by Figure 8. Although in general larger distances $\alpha\Delta$ lead to less accurate estimations, however this is not true everywhere: slightly more distant points can provide slightly more accurate estimations of the covariance and, thus, of the downcrossing probability (Figure 8).



(a)



(b)



(c)

Fig. 10. Random field updating by means of noise-free measurements using (a) one measurement point, (b) ten measurement points and (c) all the available measurement points.

Influence on the estimation of the specific pipe drift conformance probability

Consider the covariance $\mathbf{K}_p$ estimated on points $x_{p,i}^j$, $p \in]0,1[$, $i = 1,...,N$, $j = 1...,N_T$. We now investigate the effect on the a posteriori estimation of pipe drift conformance probability

of changing the number of available measurement points used to update the Gaussian field. For example, Figure 10 shows the field updating obtained with different numbers of points for one out of the N_T pipes of the available dataset, say j''. Specifically, Figure 10-(a) shows the case in which a single measurement is collected along pipe j''. In this case, the uncertainty of the field is zero at the measured point and is narrower than before along the pipe because of the covariance function. Increasing the number of points to 10 (Figure 10-(b)), the uncertainty decreases significantly. Finally, in Figure 10-(c) we can see that the uncertainty is zero when $\mathbf{A}_p = \mathbf{I}_N$. In this respect, notice that in the case in which the entire set of measurements is used to update the field, the knowledge of the correct kernel has no effect on the final drift conformance probability estimation: as mentioned before, $\mathbf{\Sigma}_{p|\mathbf{Y}_p}$ in Eq. (22) always reduces to 0, whatever the kernel is.

The results of the estimation of the drift conformance probability for the specific pipe j'' obtained by means of the Monte Carlo procedure with different numbers of measurement points are reported in Table 3. In this case, we apply both Eq. (24) and Eq. (26). The non-monotonic behavior is due to the set of measurement points selected for this specific case.

To estimate the economic effects of the reduction of the measurement points, we rely on the value of information $\text{VoI}(\mathbf{Y_p})$, which is defined as the difference between the expected utility achieved when we have a full knowledge of the field $\mathbf{X}_p$ and when the available set is $\mathbf{Y_p}$:

$$VoI(\mathbf{Y_p}) = \mathbb{E}[C(\mathbf{X}_p)] - \mathbb{E}[C(\mathbf{Y_p})] \tag{28}$$

where we set:

$$\mathbb{E}[C(\mathbf{X}_p)] = \sum_{j=1}^{N_T} \left(R^j \cdot s_{TP} + (1 - R^j) \cdot s_{TN}\right) \tag{29}$$

$$\mathbb{E}[C(\mathbf{Y}_p)] = \begin{cases} \sum_{j=1}^{N_T} \left(R^j \cdot s_{TP} + (1 - R^j) \cdot s_{TN} + \Delta r^j \cdot s_{FP}\right) & if\ \Delta r^j > 0 \\ \sum_{j=1}^{N_T} \left(R^j \cdot s_{TP} + (1 - R^j) \cdot s_{TN} + |\Delta r^j| \cdot s_{FN}\right) & if\ \Delta r^j < 0 \end{cases} \tag{30}$$

where $\Delta r^j = R^j|\mathbf{x}_p^j - R^j|\mathbf{y}_p^j$.

In words, VoI represents the maximum amount of money that is worth to pay for getting accurate measurements.

Table 4 reports the VoI as a function of the number of measurement points. Considering noise-free measurements, the loss is significant both when 1 and 10 measurement points are considered.

	1 *point*	10 *points*	*Every point*
$R^{j''}$	0.86	0.89	0.95
$\hat{R}^{j''}_{0.05}$	0.9189	0.9128	1

TABLE 3. Drift conformance probability of the specific pipe as function of the number of measurement points.

	1 *point*	10 *points*	*Every point*
VoI	-26.4	-9.04	0

TABLE 4. Value of Information as function of the number of measurement points.

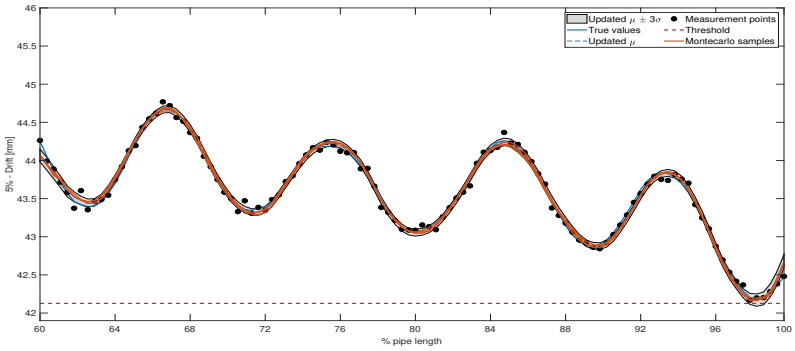
5.2 Measurements uncertainty

We now consider the uncertainty in the drift estimates, which is modeled as a Gaussian noise, with respect to the true measured value (Eq. (17)).

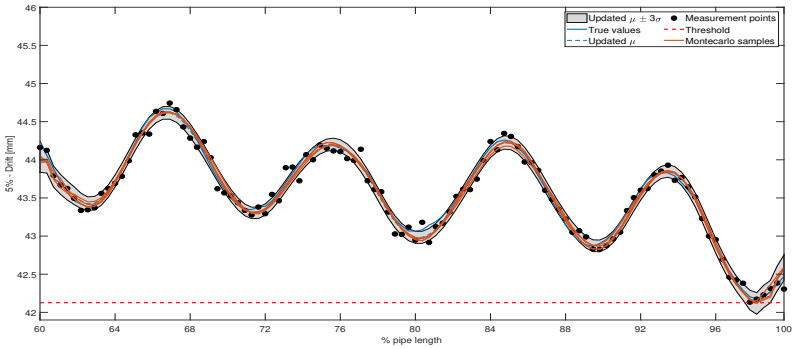
Figure 11 and Figure 12 show the random field updating of the specific pipe j'', in case of different values of the noise standard deviation, σ_ϵ. The uncertainty is no longer zero in the measurement points. In these cases, Monte Carlo samples do not overlap on a single line.

Table 5 shows the drift conformance probability estimation for the specific pipe j'', considering all the measurement points. From the first column, it can be noticed that the drift conformance probability estimation is very different from the corresponding one in Table 3.

Table 6 reports the VoI as a function of the number of measurement points. The results reported in Table 5 and Table 6 are obtained considering the average over one hundred Monte Carlo sample for each pipe in order to take into account the variability in the measurements induced by the noise. Considering noisy measurements, the lack of precise knowledge on the measurements leads



(a)



(b)

Fig. 11. Random field updating by means of noisy measurements using (a) $\sigma_\epsilon = 0.05$, (b) $\sigma_\epsilon = 0.1$.

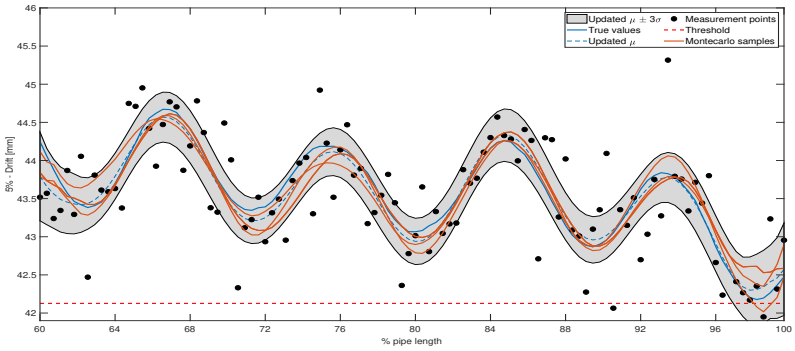
382 to significant losses. This entails that the investment in technology that increases the accuracy is
383 valuable.

σ_ϵ	0.05	0.1	0.5	1
$R^{j''}$	0.94	0.93	0.91	0.89
$\hat{R}^{j''}_{0.05}$	0.5266	0.5011	0.6579	0.7435

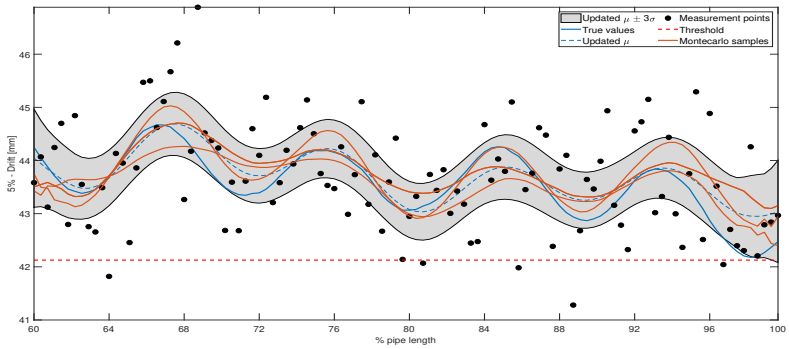
TABLE 5. Drift conformance probability for the specific pipe as a function of the noise standard deviation σ_ϵ.

σ_ϵ	0.05	0.1	0.5	1
VoI	-1.10	-2.04	-9.22	-15.71

TABLE 6. Value of Information as a function of the noise standard deviation σ_ϵ.



(a)



(b)

Fig. 12. Random field updating by means of noisy measurements using (a) $\sigma_\epsilon = 0.5$ and (b) $\sigma_\epsilon = 1$.

6 CONCLUSIONS

In this work, we have developed a methodology to a priori estimate the rejection probability of pipes for deep sea application and a posteriori estimate the drift conformance probability of a specific pipe, based on the actual measurements.

The method is based on the Gaussian fields theory, which rely on a kernel function. Instead of using an analytical kernel, we have used its numerical approximation. Then, we have applied Rice's formula for the estimation of the rejection probability and a Monte Carlo sampling approach for the estimation of the specific pipe drift conformance probability. The methodology has been applied to a dataset of pipe geometrical parameters derived from a real case study, collected at discrete points along the pipes length. The method has been shown to provide results in accordance with experts' estimations.

Finally, a sensitivity analysis with respect to the number of measurement points and the related uncertainty has been proposed. It has been shown that these two aspects can strongly affect both the a priori and the a posteriori estimations.

Future research work will focus on addressing the issues considered in this work, but relaxing the assumptions on field stationarity and normality. Moreover, efforts will be devoted to investigating how analytical kernels can be used even if their shapes are expected to strongly vary based on pipe characteristics. Moreover, the theoretical framework proposed will be extended to the other pipe quality factors.

Data Availability Statement

Some or all data, models, or code generated or used during the study are proprietary or confidential in nature and may only be provided with restrictions.

REFERENCES

Adler, R. and Taylor, J. (2009). *Random Fields and Geometry*. Springer Monographs in Mathematics. Springer New York, <https://books.google.it/books?id=R5BGvQ3ejloC>.

Adler, R. J. (2000). "On excursion sets, tube formulas and maxima of random fields." *Ann. Appl. Probab.*, 10(1), 1–74.

Adler, R. J. (2010). *The Geometry of Random Fields*. Society for Industrial and Applied Mathematics, <https://epubs.siam.org/doi/abs/10.1137/1.9780898718980>.

API (2018). "Bulletin on formulas and calculations for casing, tubing, drill pipe and line properties." *API Bulletin 5C3, 7th ed.*

Atkinson, K. E. (1989). *An Introduction to Numerical Analysis (2nd Edition)*. John Wiley & Sons, New York.

Bulinskaya, E. V. (1961). "On the mean number of crossings of a level by a stationary gaussian process." *Theory of Probability & Its Applications*, 6(4), 435–438.

Crooks, E. (2016). "How can the oil industry cut costs." *Financial Times*.

Di Maio, F., Baraldi, P., Brivio, L., Zio, E., and Magno, C. (2017). "Resistance-based probabilistic design by order statistics for an oil and gas deep-water well casing string affected by wear during kick load." 119.

Dickey, D. and Fuller, W. (1981). "Likelihood ratio statistics for autoregressive time series with a unit root." *Econometrica*, 49, 1057–1072.

Duvenaud, D., Lloyd, J., Grosse, R., Tenenbaum, J., and Zoubin, G. (2013). "Structure discovery in nonparametric regression through compositional kernel search." *Proceedings of the 30th International Conference on Machine Learning*, S. Dasgupta and D. McAllester, eds., Vol. 28 of *Proceedings of Machine Learning Research*, Atlanta, Georgia, USA, PMLR, 1166–1174, <http://proceedings.mlr.press/v28/duvenaud13.html> (17–19 Jun).

for Standardization. International Electrotechnical Commission, I. O. (2012). *Uncertainty of Measurement - Part 4: ISO IEC GUIDE 98-4. Role of measurement uncertainty in conformity assessment*. Number pt. 4 in ISO/IEC guide. International Organization for Standardization.

Garcia, R. (2016). *Hydraulic Design Manual*. Texas Departement of Transportation.

Golub, G. H. and Van Loan, C. F. (2012). *Matrix computations*, Vol. 3. JHU press.

Hoblit, F. M. (1988). *Gust loads on aircraft: concepts and applications*. American Institute of Aeronautics and Astronautics.

Hovem, L. (2019). "The outlook for the oil and gas industry in 2019." *DNV GL's Industry Outlook report*.

ISO-11960:2014 (2014). "Petroleum and natural gas industries — steel pipes for use as casing or tubing for wells.

ISO/TR-10400 (2018). "Petroleum and natural gas industries-formula and calculation for casing, tubing, drill pipe and line pipe properties.

Ito, K. et al. (1963). "The expected number of zeros of continuous stationary gaussian processes." *Journal of Mathematics of Kyoto University*, 3(2), 207–216.

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., and Hoffmann, M. (2014). "Industry 4.0." *Business and Information Systems Engineering*, 6(4), 239–242.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (2012). *Extremes and related properties of random sequences and processes*. Springer Science & Business Media.

Lilliefors, H. W. (1967). "On the kolmogorov-smirnov test for normality with mean and variance unknown." *Journal of the American Statistical Association 62*, 399–402.

Lindgren, G. (2006). "Lectures on stationary stochastic processes." *PhD course of Lund's University*.

Malings, C. A. (2017). "Optimal sensor placement for infrastructure system monitoring using probabilistic graphical models and value of information." Ph.D. thesis, Carnegie Mellon University, Carnegie Mellon University.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

O'Hagan, A. (2006). "Bayesian analysis of computer code outputs: a tutorial." *Reliability Engineering and System Safety*, 91, 1290–1300.

Papoulis, A. and Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.

Phillips, P. and Perron, P. (1988). "Testing for a unit root in time series regression." *Biometrika*, 75(2), 335–346.

Plackett, R. L. (1983a). "Karl pearson and the chi-squared test." *International Statistical Review / Revue Internationale de Statistique*, 51(1), 59–72.

Plackett, R. L. (1983b). "Karl pearson and the chi-squared test." *International Statistical Review 51*, 59–72.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1993). *Numerical Recipes in FORTRAN; The Art of Scientific Computing*. Cambridge University Press, USA, 2nd edition.

Rice, S. O. (1944). "Mathematical analysis of random noise." *Bell System Technical Journal*, 23(3), 282–332.

Sorensen, H. V., Douglas, L. J., T., H. M., and S., B. C. (1987). "Real-valued fast fourier transform algorithms." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-35(6), 849–863.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*, Vol. 2. MIT press Cambridge, MA.

Ylvisaker, N. D. (1965). "The expected number of zeros of a stationary gaussian process." *The Annals of Mathematical Statistics*, 36(3), 1043–1046.