Electronic Thesis and Dissertation Repository

9-7-2023 2:00 PM

Application of Wavelet Analysis and Random Field in Integrity Management of Pipelines Containing Dents and Corrosions

Junxiong Lin, *The University of Western Ontario*

Supervisor: Zhou, Wenxing, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Civil and Environmental Engineering
© Junxiong Lin 2023

Recommended Citation

Abstract

Metal loss corrosions and dents are two major threats to the integrity of oil and natural gas pipelines. In the pipeline industry, the Fitness-For-Service (FFS) assessment is commonly employed for pipelines containing these defects. However, FFS assessment usually assumes that a defect has a simple shape, and such a simplification may significantly affect the accuracy of the assessment. Therefore, retaining the actual shapes of defects and incorporating them into the FFS assessment can improve assessment accuracy. The main objective of the present thesis is to extract key information about the sizes, directions, and shapes of corrosions and dents from the measurement of in-service and excavated pipelines, and then improve the accuracy of FFS assessment based on the extracted information.

The first study develops a wavelet transform-based denoising method for the measured inner surface of in-service dented pipelines obtained from caliper tools. Since the inner surface is differently sampled along the longitudinal and circumferential directions, the commonly used denoising methods cannot sufficiently remove measurement errors from the signal. The proposed method is based on overcomplete expansion, and the overcomplete dictionary is constructed from the hyperbolic wavelet transform and stationary transform. The strain estimated from the signal denoised by the proposed method is closer to the actual strain than the other denoising method. An overcomplete dictionary that can effectively denoise the dent signal is then constructed based on the statistics of the measurement of in-service dented pipelines.

The second study explores the vital directional features and length scales of natural corrosion clusters that govern the burst capacity of corroded pipelines. The corrosion depths in a cluster are measured by high-resolution laser scans, and two-dimensional (2D) discrete wavelet transform (DWT) with a suitable wavelet function is employed to decompose the corrosion cluster. A methodology is proposed to determine level- and sub-band-dependent thresholds such that those wavelet coefficients below the thresholds have a negligible impact on the burst capacity predicted by the widely used RSTRENG model and can be ignored for the reconstruction of the cluster. The preserved wavelet coefficients show that longitudinally

orientated features with 4 – 32 mm in length have a greater influence on the remaining burst capacity than other features. This facilitates FFS assessment of corroded pipelines.

The third study aims to simulate the corrosion fields whose morphology and marginal distribution are close to the actual corrosion fields from limited information summarized from the ILI data. The corrosion field containing multiple corrosion anomalies is modelled as a nonhomogeneous non-Gaussian random field, where the spatial correlation and marginal distribution of anomalies are estimated from their sizes. The proposed methodology provides realizations of corrosion fields with the RSTRENG-predicted burst capacity closer to the actual burst capacity than the commonly used methodology that idealizes anomalies as cuboids.

The fourth study presents a framework to analyze and simulate nonhomogeneous non-Gaussian corrosion fields on the external surface of buried in-service pipelines by using continuous and discrete wavelet transforms. Continuous wavelet transform (CWT), dual-tree complex discrete wavelet transform (DT-CDWT), and dual-tree complex discrete wavelet with hyperbolic wavelet transform scheme (DT-CHWT) are incorporated into the iterative power and amplitude correction (IPAC) algorithm to extract the features of the natural corrosion field measured by a high-resolution laser scan and generate synthetic corrosion fields. The results indicate that the proposed framework can generate synthetic corrosion fields that effectively capture probabilistic characteristics of the measured corrosion field in terms of the scalogram, textural features, and burst capacity of the pipe segment containing the corrosion field.

Keywords

Summary for Lay Audience

Pipelines are widely considered the most efficient and safest way to transport large quantities of oil and gas products over long distances. However, the integrity of pipelines is often threatened by corrosion and mechanical damage such as dent. The fitness-for-service (FFS) assessment is generally employed in the pipeline industry to assess a pipeline segment containing corrosions or dents. The characteristics of naturally occurring corrosions and dents are essential to the accuracy of FFS assessment, and these characteristics can be summarized from the measurement of the in-service and excavated pipelines. This thesis aims at improving the accuracy of FFS assessment models for pipelines containing corrosions or dents using the wavelet transform and random field.

The pipeline industry usually runs inline inspection (ILI) tools to detect and size defects along the in-service pipelines. After denoising the ILI signals, the sizes and locations of detected defects are incorporated into some FFS assessment models to determine if the pipeline can work safely with these defects. The present study involves different stages of this process. Some modifications are introduced to the wavelet transform-based denoising to better distinguish the actual shape of dents from the noise caused by measurement errors. The size and direction of the corrosion anomalies that significantly affect the burst capacity are extracted and summarized by the discrete wavelet transform. Based on the statistics of corrosion anomalies, the corrosion defects summarized by ILI tools are modelled as random fields rather than simple deterministic shapes. Introducing uncertainties will make the modelled corrosion defects similar to the actual corrosion defects, thus improving the accuracy of the burst capacity of corroded pipes. If the naturally corroded pipeline surface is detailly measured, a methodology incorporating wavelet transforms and random field modelling can capture the main characteristics of the measured surface and generate synthetic surfaces with these characteristics. These improvements to the different stages provide more accurate assessment results than the commonly used methods in the pipeline industry.

Co-Authorship Statement

A version of Chapter 2, co-authored by Junxiong Lin and Wenxing Zhou, is going to be submitted to *Thin-Walled Structures* for potential publication.

A version of Chapter 3, co-authored by Junxiong Lin and Wenxing Zhou, is currently under review for possible publication in *International Journal of Pressure Vessels and Piping*

A version of Chapter 4, co-authored by Junxiong Lin and Wenxing Zhou, has been accepted in *Structure and Infrastructure Engineering* (accepted on June 2, 2023).

A version of Chapter 5, co-authored by Junxiong Lin, Wenxing Zhou, Xizhong Cui, and Han-Ping Hong, has been published in *Computers and Structures*, 276, 106957. https://doi.org/10.1016/j.compstruc.2022.106957.

Acknowledgments

It is my pleasure to take this opportunity to thank the following people, without whom I would not have been able to complete this research.

Foremost, I want to thank my supervisor Dr. Wenxing Zhou, who provided me with constant support and invaluable advice during my Ph.D. study. To me, exploring unfamiliar areas with newly learned knowledge is challenging. Without Prof. Zhou's valuable guidance, profound knowledge, and countless hours of revising my writing, completing this thesis would have been impossible. I am honored to be one of his students. I would also like to thank Dr. Han-Ping Hong, whose intelligent and precise thinking gave me practical advice and inspiration.

It was a difficult time that I was unable to return to China during my doctoral study due to the COVID-19 pandemic. Without my parents' endless love and unconditional support, I could not go through the hard times. My deepest gratitude always goes to my parents.

Thanks to Stella, the reason why I am here.

I would like to thank Yue Liu, Haotian Sun, Ziming He, Ji Bao, Yufei Shen, Wei Xiang, Shulong Zhang, Haoyi Zhang, Parnian Ghoraishi, and Rodrigo Silva in our research group. I would also like to thank Yongxu Liu, Xizhong Cui, Minyu Xiao, Chao Sheng, Ruifeng Duan, Xiaohan Jiang, Shaoqing Yu. As an inactive person, I would not be able to overcome the different difficulties of life and studies and have four colourful years in Canada without their help.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Abbreviations:

ACF	autocorrelation function
API	American Petroleum Institute
ASME	American Society of Mechanical Engineers
COV	coefficient of variation
CDF	cumulative distribution function
CER	Canada Energy Regulator
COV	coefficient of variation
CSA	Canadian Standards Association
CWT	continuous wavelet transform
DNV	Det Norske Veritas
DOST	discrete orthogonal S-transform
DT-CDWT	dual-tree complex discrete wavelet transform
DT-CHWT	dual-tree complex hyperbolic wavelet transform
DWT	discrete wavelet transform
EC	Eddy current
EGIG	European Gas Pipeline Incident Data Group
EC	Eddy current
EMAT	electromagnetic acoustic transducers

FEA	finite element analysis
FFS	fitness-for-service
GLCM	gray-level co-occurrence matrix
HWT	hyperbolic wavelet transform
ILI	in-line inspection
IPAC	iterative power and amplitude correction
MFL	magnetic flux leakage
PDF	probability density function
PIG	pipeline inspection gauge
POF	Pipeline Operators Forum
PHMSA	Pipeline and Hazardous Materials Safety Administration
PIG	pipeline inspection gauge
PSD	power spectral density
SMYS	specified minimum yield strength
SWT	stationary wavelet transform
UKOPA	United Kingdom Onshore Pipeline Operators' Association
UT	ultrasonic
1D	one-dimensional
2D	two-dimensional
3D	three-dimensional

Symbols

Chapter 1

wt	pipe wall thickness

Chapter 2

C_ψ	admissibility condition
d_{max}	maximum dent depth
E	elastic modulus of pipe steel
E_z	estimated energy of the noise-free signal
f	frequency (wavenumber)
H	detail coefficient of 1D DWT
HH	diagonal features of 2D DWT
HL	vertical features of 2D DWT
i	decomposition level
j	iteration step
K	strength coefficient of pipe steel
k	integer translation
L	approximation coefficient of 1D DWT
LH	horizontal features of 2D DWT

LL	approximation of 2D DWT
MAD	median absolute value of the 1st level wavelet detail coefficients
M_C	number of decomposition levels in the circumferential direction
M_L	number of decomposition levels in the longitudinal direction
N	vanishing moment
N_C	number of wavelet functions along the circumferential direction
N_L	number of wavelet functions along the longitudinal direction
n_h	strain hardening exponent
OD	pipe outside diameter
R	measured inner radius
R_0	noise-free inner radius
R_{df}	dent-free inner radius
R_d	denoised inner radius
R_n	noisy inner radius
$RMSE$	root mean square error
s	scaling factor
v	1D signal
v_w	wavelet coefficient of v
wt	pipe wall thickness
x	longitudinal coordinate

y	y-axis coordinate
z_r	reconstructed signal
$z_r^{(j)}$	component extracted at the jth iteration
$z_\psi^{(\alpha)}$	wavelet coefficient of 2D DWT in the α sub-band
α	sub-band indicator
β	number of data points in a signal
Δx	longitudinal sampling spacing
$\Delta \theta$	circumferential sampling spacing
ε	strain
$\varepsilon_c^{(b)}$	circumferential bending strain
$\varepsilon_c^{(m)}$	circumferential membrane strain
ε_{eff}	effective strain
$\varepsilon_{eff,\max}$	maximum effective strain
$\varepsilon_{eff,\max}^{(L)}$	maximum effective strain based on the large dictionary
$\varepsilon_{eff,\max}^{(S)}$	maximum effective strain based on the small dictionary
$\varepsilon_l^{(b)}$	longitudinal bending strain
$\varepsilon_l^{(m)}$	longitudinal membrane strain
ε_n	noise
ε_{pe}	plastic equivalent strain from FEA

$\varepsilon_{pe,\max}$	maximum plastic equivalent strain from FEA
η	threshold
γ	order of vanishing moment
κ_l	circumferential curvature
κ_l	longitudinal curvature
φ	scaling function
ψ	wavelet function
$\hat{\psi}$	Fourier transform of ψ
σ	stress
σ_n	standard deviation of the noise
σ_u	tensile strength of pipe steel
σ_y	yield strength of pipe steel
τ	translation
τ_x	translation along x direction
τ_y	translation along y direction
θ	circumferential coordinate
*	complex conjugate

Chapter 3

A_e	effective area

A_{inc}	maximum area change contributed to a wavelet coefficient
C_ψ	admissibility condition
$d_{\max}^{(c)}$	maximum corrosion depth of a cluster
E_0	energy of the natural corrosion cluster
E_r	energy of the reconstructed cluster after thresholding
f	frequency (wavenumber)
G_1	level parameter
G_2	level parameter
G_α	sub-band parameter
H	detail coefficient of 1D DWT
HH	diagonal features of 2D DWT
HL	vertical features of 2D DWT
i	decomposition level
k	integer translation
L	approximation coefficient of 1D DWT
LH	horizontal features of 2D DWT
LL	approximation of 2D DWT
l_e	effective length
M_e	Folias factor corresponding to the effective length
N	vanishing moment

OD	pipe outside diameter
p	burst capacity
p_0	burst capacity corresponding to a natural corrosion cluster
p_r	burst capacity corresponding to a reconstructed cluster
q_0	number of grid points within a corrosion cluster
q_w	number of non-zero wavelet coefficient after thresholding
s	scaling factor
v	1D signal
v_w	wavelet coefficient of v
wt	pipe wall thickness
x	longitudinal coordinate
y	circumferential coordinate
z	corrosion depth
$z^{(r)}$	corrosion depth in the reconstructed cluster
$z_\psi^{(\alpha,i)}$	wavelet coefficient of 2D DWT in the ith level α sub-band
α	sub-band indicator
η_d	corrosion depth value that essential for predicting p
$\eta_{i,\alpha}$	wavelet threshold corresponding to the ith level α sub-band
$\eta_{\alpha,i}^{(lb)}$	lower bound of normalized wavelet threshold
$\eta_{\alpha,i}^{(ub)}$	upper bound of normalized wavelet threshold

η_w	wavelet threshold
φ	scaling function
ψ	wavelet function
$\hat{\psi}$	Fourier transform of ψ
σ_y	tensile strength of pipe steel
τ	translation
τ_x	translation along x direction
τ_y	translation along y direction
*	complex conjugate

Chapter 4

A_e	effective area
a_C	circumferential decaying rate
a_L	longitudinal decaying rate
D	pipe outside diameter
E	elastic modulus of pipe steel
$\bar{d}$	mean corrosion depth of an anomaly
d_{max}	maximum corrosion depth of an anomaly
F_Z	target marginal distribution of corrosion depths in an anomaly
$F_{Z^{(B)}}$	empirical CDF of the non-zero value of $z^{(B)}$

K	strength coefficient of pipe steel
l	longitudinal length of an anomaly
l_C	circumferential correlation length
l_e	effective length
l_L	longitudinal correlation length
M	spatial modulator
M_e	Folias factor corresponding to the effective length
n	number of anomalies
n_h	strain hardening exponent
p_0	burst capacity corresponding to natural corrosion field
p_b	burst capacity
p_{ILI}	burst capacity corresponding to ILI idealization
p_s	burst capacity corresponding to synthetic field
$\overline{p}_s$	mean of p_s
T_x	transformed longitudinal coordinate for spatial modulator
T_y	transformed circumferential coordinate for spatial modulator
V_{ILI}	volume of metal loss of an ILI-idealized cuboid-shaped anomaly
V_{real}	actual volume of metal loss in an anomaly
w	circumferential width of an anomaly
wt	pipe wall thickness

x	longitudinal coordinate
x_c	longitudinal coordinate of anomaly centre
y	circumferential coordinate
y_c	circumferential coordinate of anomaly centre
z	corrosion field
$z^{(B)}$	nonhomogeneous homogeneous non-Gaussian field
$z^{(H)}$	homogeneous non-Gaussian field
ε	strain
ρ	correlation coefficient
σ_d	standard deviation of corrosion depths in an anomaly
σ_f	flow stress of pipe steel
σ_y	tensile strength of pipe steel
σ_y	yield strength of pipe steel
τ_x	longitudinal separation
τ_y	circumferential separation

Chapter 5

A_e	effective area
ASM	angular second moment

COR	correlation
$\boldsymbol{c}$	complex coefficients
$\boldsymbol{c}_{im}$	imaginary part of $\boldsymbol{c}$
$\boldsymbol{c}_{re}$	real part of $\boldsymbol{c}$
D	pipe outside diameter
f_0	central frequency (wavenumber)
G	gray level
HH	diagonal features
HL	vertical features
IDM	inverse difference moment
ITf($\bullet$)	inverse transform
i	decomposition level (2D DWT)
i_x	decomposition level along the *x* direction (HWT)
i_y	decomposition level along the *y* direction (HWT)
j	imaginary unit
LH	horizontal features
LL	approximation
l_e	effective length
M_e	Folias factor corresponding to the effective length
NCR	normalized contrast ratio

NSE	normalized Shannon entropy
P_b	burst capacity
r	rank of z_{pc}
s_x	scaling factor along x direction
s_y	scaling factor along y direction
$Tf(\bullet)$	forward transform
v	signal in 1D
w	white noise
w_f	transform of w
wt	pipe wall thickness
x	spatial index
y	spatial index
z	corrosion depth
z_{ac}	amplitude corrected field
z_{pc}	power corrected field
z_t	transform of z
z_w	wavelet coefficient of z
$z_\psi^{(\alpha)}$	α sub-band (2D DWT) wavelet coefficient of z
$z_\psi^{(\beta)}$	β sub-band (HWT) wavelet coefficient of z
Δx	sampling interval along the x-axis

Δy	sampling interval along the y-axis
ε_r	squared difference at the rth iteration
φ	scaling function
ψ	wavelet function
$\psi_{(\alpha)}$	wavelet function for the α sub-band (2D DWT)
$\psi_{A,(\alpha)}$	wavelet function for the α sub-band (HWT)
ψ_c	wavelet function in DT-CDWT
ψ_{im}	imaginary part of ψ_c
ψ_{re}	real part of ψ_c
σ_f	flow stress
σ_y	yield strength
τ_x	translation along x direction
τ_y	translation along y direction
θ	phase
ζ	ascending array that matches the target marginal distribution
$*$	complex conjugate

List of Appendices

Chapter 1

1 Introduction

1.1 Background

Pipelines are widely considered the most efficient and safest way to transport large quantities of oil and gas products over long distances. According to statistics provided by the Government of Canada (NRCan, 2020), there are more than 840,000 kilometres (km) of transmission, gathering and distribution pipelines in Canada. About 8% of the pipelines are regulated by the Canada Energy Regulator (CER, 2021), including 48,338 km of operating gas pipelines and 19,142 km of operating oil pipelines. With the increased pipeline service life, pipelines may be threatened by different damage mechanisms. Metal-loss corrosions and dents are two major threats to the integrity of oil and natural gas pipelines. Metal-loss corrosion is an electrochemical process caused by the interaction between the local environment and pipelines that results in a reduction in pipe wall thickness. The data collected by the Pipeline and Hazardous Materials Safety Administration (PHMSA) of the United States indicate that corrosion is a major damage mechanism, causing about 32% of the incidents on onshore gas transmission pipelines in the US between 2002 and 2013 (Lam and Zhou 2016). The report from the European Gas Pipeline Incident Data Group (EGIG 2020) has shown that 27% of the incidents on onshore gas transmission pipelines in Europe were caused by corrosion between 2010 and 2019. Dents are a common type of mechanical damage caused by a foreign object impacting the external surface of pipelines and may result from third-party excavation activities and rock impact. Dents are usually included in external interference in technical reports. It was the cause of 27.17% of onshore gas transmission pipeline incidents in Europe in the period from 2010 to 2019 (EGIG 2020). According to the report provided by the United Kingdom Onshore Pipeline Operators' Association (UKOPA), 21.5% of the product loss incidents between 1962 to 2019 were caused by external interference (Goodfellow 2021). Examples of indented and corroded pipelines are shown in Fig. 1 . To assess whether the pipeline containing defects is deemed

adequate for its purpose, the fitness-for-service (FFS) assessment is commonly carried out in the pipeline industry once the defects are detected.
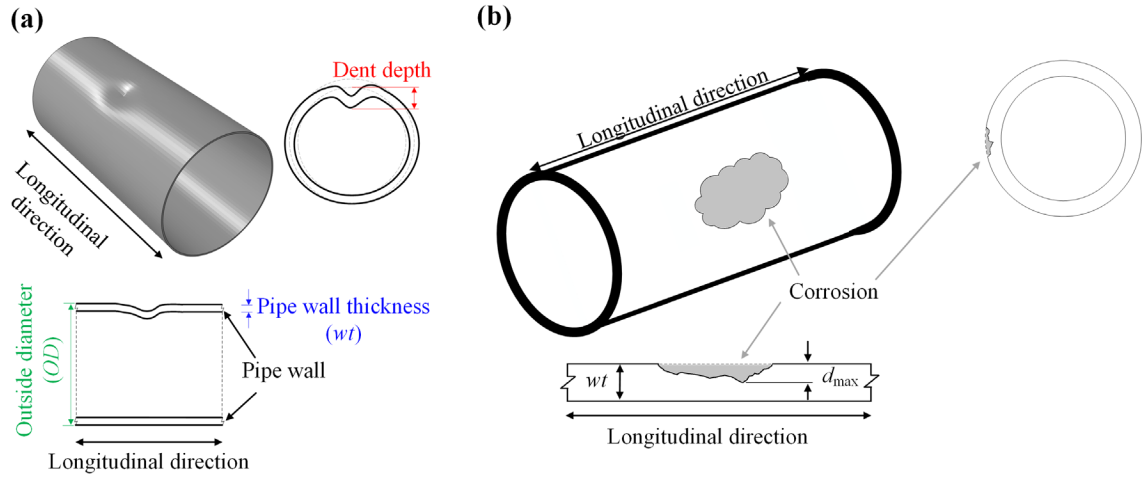
Figure 1.1 (a) indented pipeline; and (b) corroded pipeline

Defects on in-service pipelines are usually located and sized by the in-line inspection (ILI). ILI is a non-destructive evaluation technique that uses a pipeline inspection gauge (PIG) based on various inspection techniques to collect data of different defects. PIGs are propelled by the product in the pipeline, and the sensors on PIGs will collect different signals along the in-service pipelines. The commonly used technologies of ILI sensors (Xie and Tian 2018; Vanaei et al. 2017) for metal-loss defects include magnetic flux leakage (MFL), ultrasonic (UT), Electromagnetic acoustic transducers (EMAT), and Eddy current (EC). As the most widely used ILI technique for different types of defects (Xie and Tian 2018), MFL tools detect pipeline defects from the local change in a saturating magnetic field near the defects. Although the distribution and amount of flux leakage depend on the geometry of defects, reconstructing the detailed shapes of the defects from the MFL signals is a challenging work that requires heavy computation (Joshi et al. 2006). Therefore, the defects detected by MFL tools are usually summarized by their locations (longitudinal and circumferential coordinates) and sizes (longitudinal length, circumferential width, and maximum depth). In addition to MFL tools, UT is considered the most reliable ILI technology compared with lower measurement error

(Goedecke 2003) with the other technologies. UT tools measure the locations of the inner and outer surfaces by recording the time of reflected ultrasonic signal. Therefore, the geometry of defects can be obtained directly from the ultrasonic signal. The deformation of pipelines is usually measured by caliper tools, which measure the change in the pipeline's inner radii (Revie 2015). Some technologies, such as optical sensors (Sampath et al. 2019; Feng et al. 2016), are proposed to achieve better defect detection. However, these technologies have not yet been widely applied in the pipeline industry due to the operating environment and cost.
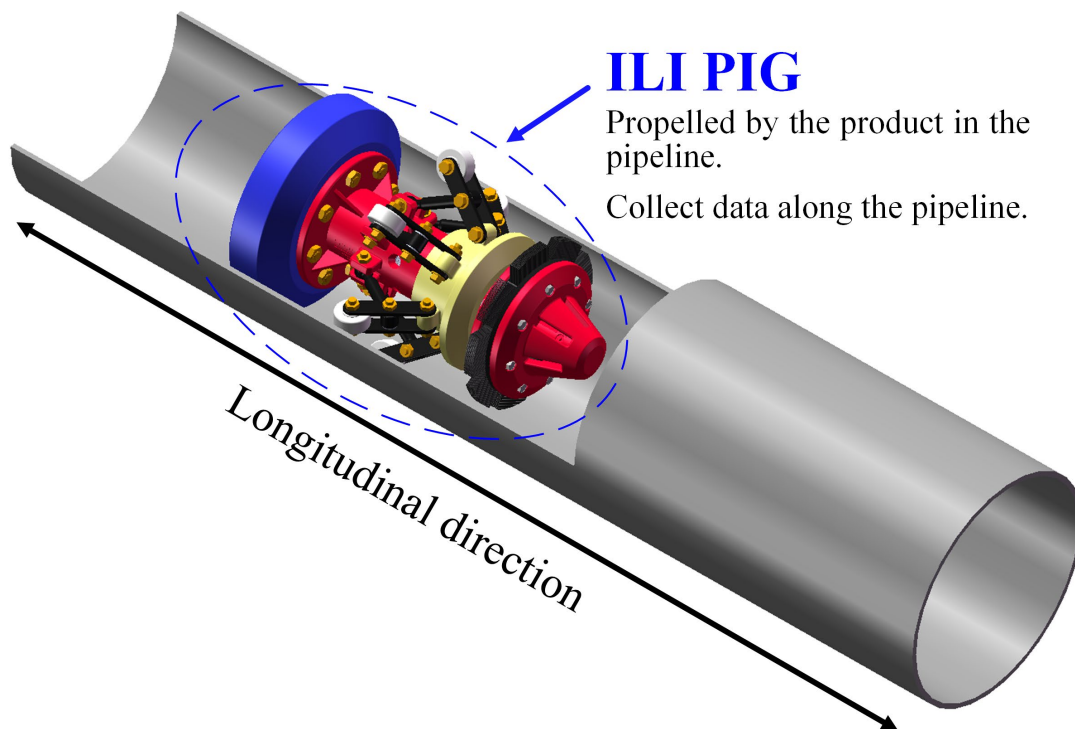
Figure 1.2 In-line inspection tool

In practice, the ILI signals contain non-negligible noise due to the accuracy of the sensors and the complex measurement environment, etc. Before extracting and summarizing information about defects from the ILI signal, an essential step is denoising the ILI signal. Most of the valuable information in the signal is included in the low-frequency (wavenumber) component, while the high-frequency components are mainly noise. It

follows that a simple denoising method is removing the high-frequency components from the measured signal, which can be achieved by the Fourier transform or convolving the signal by a low-pass filter (Lukasiewicz et al. 2006; Okoloekwe et al. 2018; Ergezinger et al. 2020; Zhang et al. 2022b). However, the Fourier transform is a global representation of the signal, which means that modifications to any frequency component (e.g. reducing the amplitude of the component) may change the values of all grid points in the signal. Therefore, some features in the signal corresponding to localized high-frequency components can be lost or distorted by removing high-frequency components. As a widely used tool in time-frequency analysis, the wavelet transform has been extensively employed to denoise a signal containing local features. A frequently used wavelet-based denoising method is thresholding (Afzal and Udpa 2002; Siqueira et al. 2004; Kathirmani et al. 2012; Sampath et al. 2019). In thresholding, the signal is decomposed to multiple components constructed from wavelet functions with different scaling factors and translations, and the low energy components are considered as noise and removed. In addition to thresholding, researchers have explored other wavelet-based denoising methods. Liang et al. (2008) constructed an overcomplete dictionary to suppress the noise of the ultrasonic echoes reflected from the inner surface of an offshore pipeline. Wu et al. (2022) incorporate wavelets and neural networks to denoise the pipeline defect detection signal. Although wavelet-based denoising has been commonly used in ILI signal processing, only a few studies (Belanger and Narayanan 2008; Liu et al. 2022) have employed wavelets to denoise the dent signals collected by caliper tools. The ILI signal can be considered as a two-dimensional (2D) signal whose sampling frequencies and lengths along two directions are significantly different: the longitudinal resolution of the ILI signal is governed by the sampling frequency of sensors, while the circumferential resolution is determined by the number of sensors arranged along the circumferential direction (Revie 2015). Therefore, the wavelet-based denoising methods applicable to other 2D signals may not have a good performance on the ILI signal.

After denoising, pipeline engineers need to extract defect information from the ILI signal. Dents are quantified by their types (e.g. kinked, plain and complex), maximum depths, circumferential widths, and longitudinal lengths (POF 2021). For corrosions, a pipe wall thickness (*wt*) reduction threshold, usually between 5% and 10% *wt* (Siraj and Zhou 2019;

Amaya-Gómez et al. 2022), is determined such that only if the metal loss at the measured point exceeds such a threshold, the point is considered a corrosion point. Adjacent corrosion points are grouped as isolated individual corrosion anomalies. Since it is challenging to obtain the detailed measurement of corrosion anomalies from the ILI data directly, an anomaly is usually characterized by its location and sizing parameters. Besides anomaly dimensions, researchers have also tried to obtain more information from the ILI signal through quantitative or qualitative analysis. Accurately classifying anomalies is crucial to the corrosion risk evaluation (Zhang et al. 2022a). Based on the longitudinal length and circumferential width of a detected corrosion anomaly, this anomaly can be included in one of the seven anomaly dimension classes: general, pitting, axial grooving, circumferential grooving, pin hold, axial slotting, and circumferential slotting (POF 2021).

Wavelet transform is a commonly used method for analysing signal components. A signal can be considered as a combination of scaling and wavelet functions with different scaling factors and translations. Figure 1.3 illustrates the scaling and wavelet functions of DB2 (Daubechies 1992), one of the Daubechies wavelets. Compared to the Fourier transform, another commonly used signal processing method that represents the signal in terms of global and periodic functions, the wavelet transform enables one to focus on the local information of the signal, which facilitates the localization and description of the defects. Tucker et al. (2003) and Kercel et al. (2003) decomposed the raw electromagnetic acoustic transducer data using the Daubechies wavelets and calculated the feature vector based on the Shannon entropy and energy of the wavelet coefficients for defect classification. Duong and Kim (2018) considered the entropy of the wavelet coefficients as the feature of an acoustic emission signal and employed neural networks to classify surface defects on pipelines. If the features in the ILI signal that have a significant impact on the FFS assessment can be effectively extracted at a low cost (low computational and storage cost) and incorporated into the FFS assessment, the risk of pipelines with defects can be more accurately evaluated.
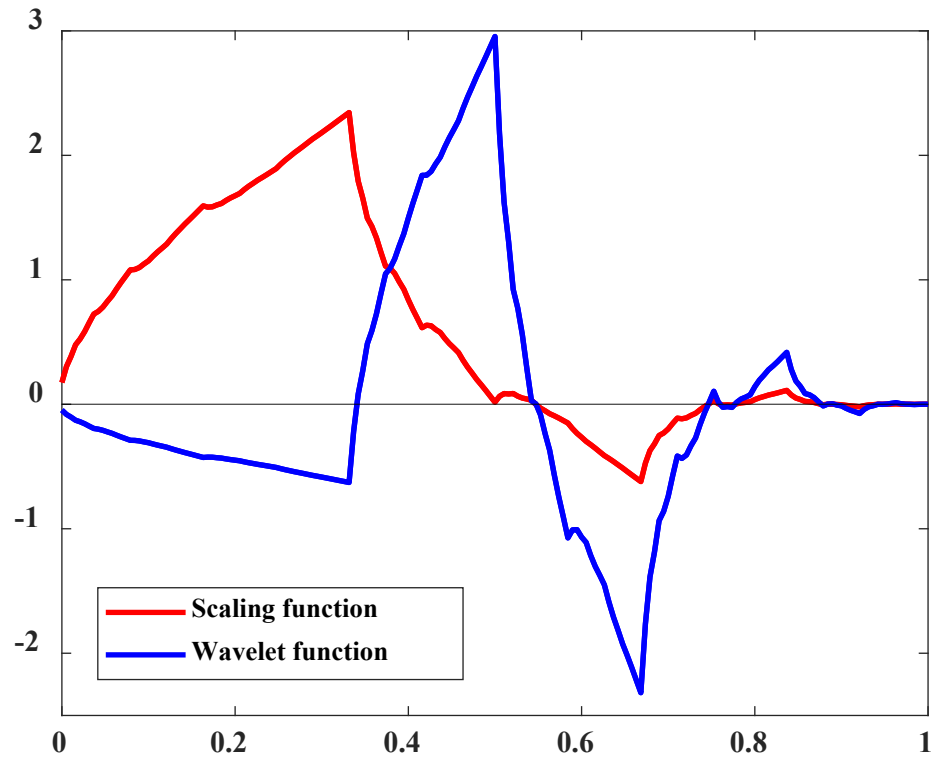
Figure 1.3 Scaling and wavelet functions of DB2 wavelet

Once the defect information in the ILI signal is extracted, many semi-empirical models are available to assess the pipelines with defects. For dents, the strain of pipe wall can be estimated from the shape of dents. As shown in Fig. 1.4 (Gao and Krishnamurthy 2015), if the inner radii of indented pipelines are available, we can calculate the membrane and bending strains of the pipe wall in both longitudinal and circumferential directions from the dent shapes. The effective strains, which will be used in some strain-based assessments, can then be calculated from these strain components. For corrosions, usually the closely spaced corrosion anomalies will be grouped as corrosion clusters based on some interaction rules (Lamontagne 2002; ASME 2019; CSA 2019), and the remaining burst capacity of a corroded pipeline is estimated based on the corrosion clusters. Besides the outside diameter and *wt* of the corroded pipeline, many well-known burst capacity models, including B31G (ASME 2012), B31G Modified (Kiefner and

Vieth 1989), DNV model (DNV 2017), SHELL92 (Ritchie and Last 1995), and PCORRC (Leis and Stephens 1997), only consider the maximum corrosion depth and longitudinal length of the corrosion cluster. Some semi-empirical models, such as RSTRENG model (Kiefner and Vieth 1989), CSA model (CSA 2019), and Psqr model (Zhang et al. 2018), consider the morphology of corrosion clusters and incorporate the detailed three-dimensional (3D) profiles into the burst capacity prediction. Figure 1.5 illustrates the river-bottom profile employed in the RSTRENG model. The prediction errors associated with these models are lower than other models (Zhou and Huang 2012; Zhang et al. 2018). However, it is very difficult to obtain the detailed profiles of corrosion anomalies from the ILI data, and the corrosion morphology is not included in the anomaly locations and sizes. Therefore, the detected corrosion anomalies are usually idealized as simple shapes (e.g. cuboids) if the detailed profiles are required in the burst capacity prediction. Such a simplification will overestimate the metal loss volume and inevitably introduces prediction errors.
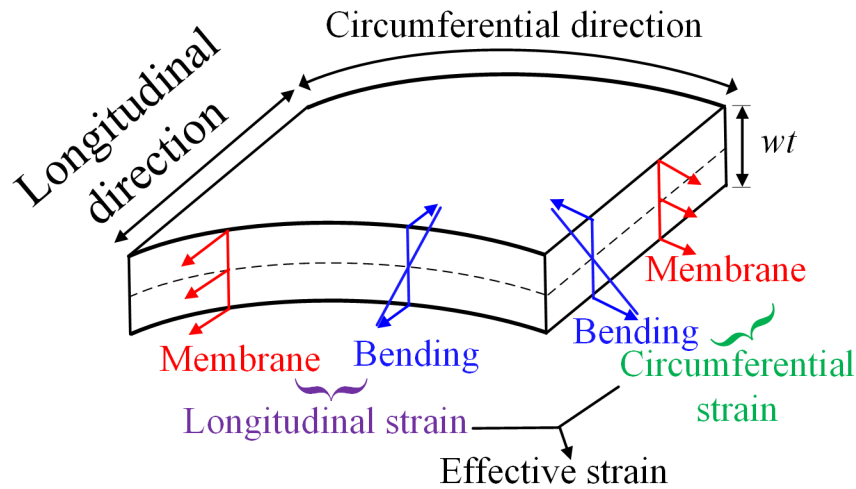
Figure 1.4 Strain components of indented pipelines
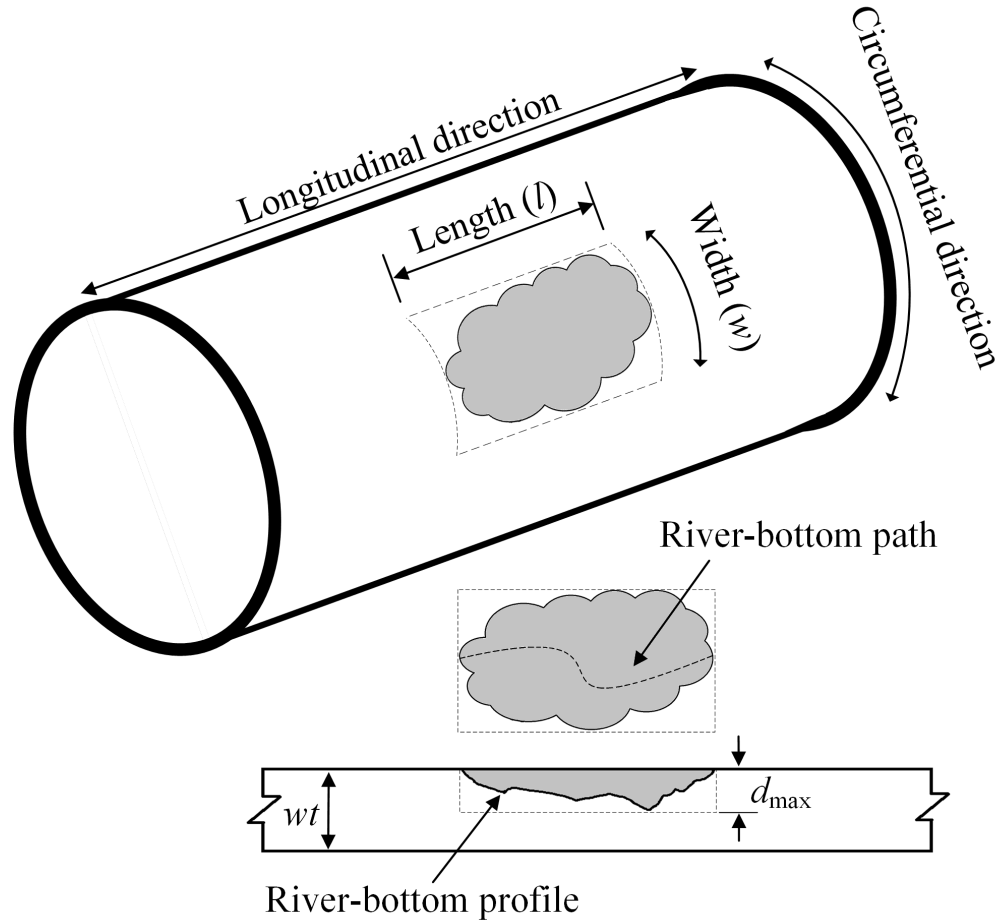
Figure 1.5 River-bottom profile of a corrosion defect

The development and validation of semi-empirical models employed in the FFS assessment are based on the full-scale burst tests of corroded pipe specimens, which are costly and time-consuming. In addition to the high costs, most of the burst tests involved to validate and develop these models employ artificially induced corrosion defects. Unlike the irregular shape of naturally occurring corrosion, artificially induced corrosion defects generally have a regular shape. Such a difference can cause considerable model errors. Since the finite element analysis (FEA) has a high accuracy in estimating the burst capacity of pipe specimens containing artificially induced or naturally occurring corrosion defects (Abdalla Filho et al. 2014; Bao et al. 2018; Mok et al. 1991; Zhang and Zhou 2020), using FEA to improve the existing semi-empirical models is cost-effective.

Some studies that model corrosion fields as random fields instead of using simplified anomalies have been reported in the literature. Aryai and Mahmoodian (2017) modeled the corrosion depths on cast iron water pipes as homogeneous Gaussian random fields. Garbatov and Soares (2019) characterized the corroded surface of steel bottom plates of ballast tanks in double-hull tankers as homogeneous lognormal random fields. Zhou et al. (2021) considered the corroded external surfaces of buried steel pipelines as nonhomogeneous non-Gaussian fields and employed the discrete orthogonal S-transform (DOST) (Stockwell et al. 1996) to characterize and simulate the corrosion field. Bao and Zhou (2021) considered the intermingling between corroded and corrosion-free areas on the external surface of buried pipelines and used a latent Gaussian field to characterize and simulate the corrosion field. Since naturally corroded external surfaces of buried pipelines are highly nonhomogeneous and non-Gaussian (Bao and Zhou 2021), it is essential to develop a method that can effectively extract the probabilistic characteristics of the natural corrosion fields and use them to generate nonhomogeneous non-Gaussian random fields. It will facilitate the development and validation of burst capacity models for corroded pipelines.

1.2   Objective

The support of the present research is financially provided by the Natural Sciences and Engineering Research Council of Canada (NSERC). The main objectives of this thesis are summarized as follows.

1) Propose a denoising methodology to sufficiently remove noises in the measured dent signals that significantly affect the strain assessment.

2) Propose a methodology to extract features that are crucial to predicting the remaining burst capacity of a corroded pipe segment, and investigate how different features in a natural corrosion cluster affect the burst capacity.

3) Propose a methodology to simulate nonhomogeneous non-Gaussian corrosion fields on the external surface of buried steel pipelines if only a few anomaly parameters obtained from the ILI data are available.

4) Presents a framework to analyze and simulate nonhomogeneous non-Gaussian corrosion fields on the external surface of buried in-service pipelines if detailed measurement is available.

This research will improve the FFS assessment employed in pipeline integrity management for pipelines containing naturally occurring corrosions and dents.

1.3   Scope of the study

This thesis consists of four main topics presented in Chapters 2 to 5, respectively. Chapter 2 proposes a denoising methodology based on the wavelet transform to remove noises in the measured dent signals that significantly affect the strain assessment. The proposed method is based on the overcomplete expansion, whereby the overcomplete dictionary is constructed by the stationary wavelet transform and hyperbolic wavelet transform. Finite element analysis is employed to obtain the noise-free inner radius signal and the corresponding strain of an indented pipe to validate the effectiveness of the proposed denoising method under different levels of noise. Suitable wavelet functions as well as the optimal decomposition levels for the dictionary are determined from real noisy dent signals measured by caliper tools.

Chapter 3 applies the two-dimensional discrete wavelet transform to analyze naturally occurring corrosion clusters on external surfaces of in-service gas transmission pipelines. A methodology is proposed to determine level- and sub-band-dependent thresholds such that those wavelet coefficients below the thresholds have a negligible impact on the burst capacity and can be ignored for the reconstruction of the cluster. The key directional features and length scales of a corrosion cluster that govern the burst capacity prediction are revealed by statistical analysis of the preserved wavelet coefficients.

Chapter 4 presents a methodology to simulate nonhomogeneous non-Gaussian corrosion fields on the external surface of buried steel pipelines by using inline inspection (ILI) data. The nonhomogeneous non-Gaussian corrosion field is assumed to consist of multiple homogeneous non-Gaussian anomalies that can be characterized by the marginal distribution and spatial autocorrelation function of the corresponding corrosion depth.

After generating realizations of corrosion anomalies, the generated anomalies are then merged into a single nonhomogeneous field by applying a spatial modulating function to each anomaly. Based on corrosion data collected from in-service pipelines, empirical relationships are developed to estimate parameters of the marginal distribution and autocorrelation function from the ILI information of corrosion anomalies. The advantages of the proposed methodology are investigated in comparison with the accuracy of the burst capacities of the corroded pipelines containing the synthetic corrosion field or the idealized corrosion field obtained from the ILI data.

Chapter 5 presents a framework to analyze and simulate nonhomogeneous non-Gaussian corrosion fields on the external surface of buried in-service pipelines by using continuous and discrete wavelet transforms. The considered transforms are the two-dimensional continuous wavelet transform (CWT) using the complex Morlet wavelets, dual-tree complex discrete wavelet transform (DT-CDWT), and dual-tree complex discrete wavelet with hyperbolic wavelet transform scheme (DT-CHWT); the natural corrosion field is measured using a high-resolution laser scan. Scalograms and marginal distribution of the measured corrosion field are incorporated into the iterative power and amplitude correction (IPAC) algorithm to generate synthetic corrosion fields. The scalogram, textural features, and burst capacity of the pipe segment containing the corrosion field are employed to validate whether the proposed framework can generate synthetic corrosion fields that effectively capture the probabilistic characteristics of the measured corrosion field.

1.4   Thesis format

This thesis is prepared in an Integrated-Article Format as specified by the School of Graduate and Postdoctoral Studies at Western University, London, Ontario, Canada. Six chapters are included in the thesis. The first chapter presents the introduction of the thesis, including the research background, objective of this thesis, scope of the study, and the thesis format. Chapters 2 to 5 are the main body of the thesis, where each chapter addresses one individual topic. Concluding remarks and recommendations for future research of these topics are given in Chapter 6.

References

Abdalla Filho, J. E., Machado, R. D., Bertin, R. J., & Valentini, M. D. (2014). On the failure pressure of pipelines containing wall reduction and isolated pit corrosion defects. *Computers & Structures*, 132, 22-33.

Afzal, M., & Udpa, S. (2002). Advanced signal processing of magnetic flux leakage data obtained from seamless gas pipeline. *NDT & E International*, *35*(7), 449-457.

Amaya-Gómez, R., Schoefs, F., Sánchez-Silva, M., Muñoz, F., & Bastidas-Arteaga, E. (2022). Matching of corroded defects in onshore pipelines based on In-Line Inspections and Voronoi partitions. *Reliability Engineering & System Safety*, *223*, 108520.

American National Standards Institute (ASME). Committee B31, Code for Pressure Piping. (2019). *Pipeline Transportation Systems for Liquids and Slurries: ASME Code for Pressure Piping, B31*. American Society of Mechanical Engineers.

American Society of Mechanical Engineers (ASME). (2012). *Manual for Determining the Remaining Strength of Corroded Pipelines: A Supplement to ASME B31 Code for Pressure Piping: an American National Standard*. American Society of Mechanical Engineers.

Aryai, V., & Mahmoodian, M. (2017). Spatial-temporal reliability analysis of corroding cast iron water pipes. *Engineering Failure Analysis*, *82*, 179-189.

Bao, J., Zhang, S., Zhou, W., & Zhang, S. (2018, September). Evaluation of burst pressure of corroded pipe segments using three-dimensional finite element analyses. In *International pipeline conference* (Vol. 51869, p. V001T03A043). American Society of Mechanical Engineers.

Bao, J., & Zhou, W. (2021). A random field model of external metal-loss corrosion on buried pipelines. *Structural safety*, *91*, 102095.

Belanger, A. A., & Narayanan, R. (2008, January). Direct strain calculation of pipe line dent from knot migration using a kinematic model free of material properties. In *International Pipeline Conference* (Vol. 48586, pp. 267-274).

Canada Energy Regulator (CER). (2021). *Canada's Pipeline System 2021: Economics of CER-Regulated Infrastructure.* https://www.cer-rec.gc.ca/en/data-analysis/facilities-we-regulate/canadas-pipeline-system/2021/index.html

Canadian Standards Association (CSA). (2019). *Oil and gas pipeline systems, CSA Standard Z662-19*. Canadian Standard Association. Mississauga, Ontario, Canada.

Daubechies, I. (1992). *Ten lectures on wavelets*. Society for industrial and applied mathematics.

Det Norske Veritas (DNV). (2017). *Recommended Practice DNV-RP-F101 Corroded pipelines*. Hovic, Norway.

Duong, B. P., & Kim, J. M. (2018). Pipeline fault diagnosis using wavelet entropy and ensemble deep neural technique. In *Image and Signal Processing: 8th International Conference*, ICISP 2018, Cherbourg, France, July 2-4, 2018, Proceedings 8 (pp. 292-300). Springer International Publishing.

EGIG. 11th Report of the European Gas pipeline Incident data Group, 2020. https://www.egig.eu/reports

Ergezinger, N., Virk, A. S., Woo, J., Kainat, M., & Adeeb, S. (2020, September). Application of Noise Filtering Techniques for the Quantification of Uncertainty in Dent Strain Calculations. In *International Pipeline Conference* (Vol. 84447, p. V001T03A026). American Society of Mechanical Engineers.

Feng, Q., Li, R., Nie, B., Liu, S., Zhao, L., & Zhang, H. (2016). Literature review: Theory and application of in-line inspection technologies for oil and gas pipeline girth weld defection. *Sensors*, *17*(1), 50.

Gao, M., & Krishnamurthy, R. (2015). Mechanical damage in pipelines: a review of the methods and improvements in characterization, evaluation, and mitigation. *Oil and Gas Pipelines*, 289-326.

Garbatov, Y., & Guedes Soares, C. (2019). Spatial corrosion wastage modeling of steel plates exposed to marine environments. *Journal of Offshore Mechanics and Arctic Engineering*, *141*(3), OMAE-18-1014.

Goedecke, H. (2003). Ultrasonic or MFL inspection: which technology is better for you. *Pipeline & gas journal*, *230*(10), 34-41.

Goodfellow, G. (2021). *UKOPA Pipeline Product Loss Incidents and Faults Report*. Report no. UKOPA/RP/21/001.

Joshi, A., Udpa, L., Udpa, S., & Tamburrino, A. (2006). Adaptive wavelets for characterizing magnetic flux leakage signals from pipeline inspection. *IEEE transactions on magnetics*, *42*(10), 3168-3170.

Kercel, S. W., Tucker Jr, R. W., & Varma, V. K. (2003, August). Pipeline flaw detection with wavelet packets and GAs. In *Intelligent Computing: Theory and Applications* (Vol. 5103, pp. 217-226). SPIE.

Kiefner, J. F., & Vieth, P. H. (1989). A modified criterion for evaluating the remaining strength of corroded pipe (No. PR-3-805). Columbus, OH: Battelle Columbus Div.

Lam, C., & Zhou, W. (2016). Statistical analyses of incidents on onshore gas transmission pipelines based on PHMSA database. *International Journal of Pressure Vessels and Piping*, *145*, 29-40.

Lamontagne, M. (2002, April). Interaction rules-An integral factor. In *CORROSION 2002*. OnePetro.

Leis, B. N., & Stephens, D. R. (1997, May). An alternative approach to assess the integrity of corroded line pipe-part I: current status. In *The seventh international offshore and polar engineering conference*. OnePetro.

Liang, W., Que, P. W., Lei, H. M., & Chen, L. (2008). Matching pursuit for decomposition and approximation of ultrasonic pulse-echo wavelet and its application in ultrasonic nondestructive evaluation. *Review of Scientific Instruments*, *79*(7), 075105.

Liu, S., Wang, H., Li, R., & Ji, B. (2022). A novel feature identification method of pipeline in-line inspected bending strain based on optimized deep belief network model. *Energies*, *15*(4), 1586.

Lukasiewicz, S. A., Czyz, J. A., Sun, C., & Adeeb, S. (2006, January). Calculation of strains in dents based on high resolution in-line caliper survey. In *International Pipeline Conference* (Vol. 42622, pp. 129-134).

Mok, D. H. B., Pick, R. J., Glover, A. G., & Hoff, R. (1991). Bursting of line pipe with long external corrosion. *International journal of pressure vessels and piping*, *46*(2), 195-216.

Natural Resources Canada (NRCan). (2020). *Pipelines Across Canada*. https://natural-resources.canada.ca/our-natural-resources/energy-sources-distribution/fossil-fuels/pipelines/pipelines-across-canada/18856

Okoloekwe, C., Kainat, M., Langer, D., Hassanien, S., & Adeeb, S. (2018, September). Algorithms for the Strain Based Analysis of Dented Pipelines. In *International Pipeline Conference* (Vol. 51869, p. V001T03A067). American Society of Mechanical Engineers.

Pipeline Operators Forum (POF). (2021). *Specifications and requirements for in-line inspection of pipelines*.

Revie, R. W. (Ed.). (2015). *Oil and gas pipelines: Integrity and safety handbook*. John Wiley & Sons.

Ritchie, D., & Last, S. (1995, April). Burst criteria of corroded pipelines-defect acceptance criteria. In *Proceedings of the EPRG/PRC 10th biennial joint technical meeting on line pipe research* (pp. 1-11).

Sampath, S., Bhattacharya, B., Aryan, P., & Sohn, H. (2019). A real-time, non-contact method for in-line inspection of oil and gas pipelines using optical sensor array. *Sensors*, *19*(16), 3615.

Siqueira, M. H. S., Gatts, C. E. N., Da Silva, R. R., & Rebello, J. M. A. (2004). The use of ultrasonic guided waves and wavelets analysis in pipe inspection. *Ultrasonics*, *41*(10), 785-797.

Siraj, T., & Zhou, W. (2019). Quantification of Measurement Errors in the Lengths of Metal-Loss Corrosion Defects Reported by Inline Inspection Tools. *Journal of Pressure Vessel Technology*, *141*(6), 061402.

Stockwell, R. G., Mansinha, L., & Lowe, R. P. (1996). Localization of the complex spectrum: the S transform. *IEEE transactions on signal processing*, *44*(4), 998-1001.

Tucker Jr, R. W., Kercel, S. W., & Varma, V. K. (2003, May). Characterization of gas pipeline flaws using wavelet analysis. In *Sixth International Conference on Quality Control by Artificial Vision* (Vol. 5132, pp. 485-493). SPIE.

Vanaei, H. R., Eslami, A., & Egbewande, A. (2017). A review on pipeline corrosion, in-line inspection (ILI), and corrosion growth rate models. *International Journal of Pressure Vessels and Piping*, *149*, 43-54.

Wu, L., Liang, W., & Sha, D. (2022). A novel adaptive noise reduction method for field natural gas pipeline defect detection signals. *International Journal of Pressure Vessels and Piping*, *199*, 104761.

Xie, M., & Tian, Z. (2018). A review on pipeline integrity management utilizing in-line inspection data. *Engineering Failure Analysis*, *92*, 222-239.

Zhang, S., Yan, J., Kariyawasam, S., Huang, T., & Al-Amin, M. (2018, September). A more accurate and precise method for large metal loss corrosion assessment. In *International Pipeline Conference* (Vol. 51869, p. V001T03A059). American Society of Mechanical Engineers.

Zhang, H., Cao, J., Liang, H., & Cheng, G. (2022a). Research on corrosion defect identification and risk assessment of well control equipment based on a machine learning algorithm. Petroleum.

Zhang, S., Zhang, B., & Wang, R. (2022b, September). A Transparent ASME B31. 8-Based Strain Assessment Method Using 3D Measurement of Dent Morphology. In *International Pipeline Conference* (Vol. 86571, p. V002T03A057). American Society of Mechanical Engineers.

Zhang, S., & Zhou, W. (2020). Assessment of effects of idealized defect shape and width on the burst capacity of corroded pipelines. *Thin-Walled Structures*, 154, 106806.

Zhou, W., Bao, J., Cui, X. Z., & Hong, H. P. (2021). Modeling and simulating nonhomogeneous non-Gaussian corrosion fields on buried pipelines and its use in predicting burst capacities of corroded pipelines. *Engineering Structures*, *245*, 112957.

Zhou, W., & Huang, G. X. (2012). Model error assessments of burst capacity models for corroded pipelines. *International Journal of Pressure Vessels and Piping*, *99*, 1-8.

Chapter 2

2 A Wavelet-based Denoising Methodology for Pipeline Dent Assessments

2.1 Introduction

Dents are a common type of mechanical damage on buried steel pipelines and may result from excavation activities near the pipeline right of way and rock impact. Since dents cause stress and strain concentrations and facilitate the initiation and growth of surface cracks (He & Zhou 2021), many codes and standards such as ASME B31.8 (ASME 2018), CSA Z662 (CSA 2019) and API RP 1183 (API 2020) have specified the strain-based or depth-based assessment and acceptance criteria for dents.

The strain-based criterion is considered more accurate than the depth-based criterion for the dent assessment (Okoloekwe et al. 2020; Zhao et al. 2022) and has been widely used in the pipeline industry. The widely used ASME B31.8 standard recommends a methodology to evaluate the equivalent strain at the dent apex based on the pipe wall thickness, and the dent depth, length, and curvature. A dent with the equivalent strain exceeding a critical strain of 6% is considered unacceptable and requires mitigation. Lukasiewicz et al. (2006) and Czyz et al. (2008) developed an algorithm based on the finite element analysis (FEA) and large deformation thin shell theory to estimate the tangential displacements of the indented pipe inner surface based on measurements obtained from the in-line caliper tools. This algorithm allows engineers to estimate the bending and membrane strains at the dent with complex geometry. Gao et al. (2008) introduced three improvements to the ASME B31.8 methodology to reduce the non-conservatism in the strain assessment. Noronha et al. (2010) considered the effects due to the number of sensors and resolution of the in-line inspection (ILI) tool on the strain estimation, and improved the accuracy of the strain estimation by applying the fourth-order B-spline curves to interpolate the dent contour. Okoloekwe et al. (2018a) introduced some modifications to the ASME B31.8 formulations to estimate the strain of the entire field rather than the dent apex. The accuracy of the modifications is validated by FEA. Considering the noise in the ILI data, Zhang et al. (2022) proposed a Fourier

transform-based denoising method to estimate the strains within the entire dented area from the ILI data. Zhao and Cheng (2022) employed FEA to determine the equivalent strain and introduced a ductile damage failure indictor for the dent assessment.

Accurately estimating strains of an indented pipe requires the detailed deformation profile of the dent. In practice, due to the vibration and measurement error of the caliper tools used for the dent measurement, the obtained signal contains non-negligible noises. The dent sizing tolerance estimated from the tool vendor's data varies from $\pm 0.51\% OD$ to $\pm 1.10\% OD$, where OD denotes the pipe outside diameter, at a confidence level of 95% (Gao & Krishnamurthy 2015). To eliminate or minimize the influence of the noise, denoising the measured signal is necessary. The Fourier transform-based denoising and filtering are viable options: by removing the high frequency (frequency is used interchangeably with wavenumber in the following) components from the measured signal or convolving the signal by a filter, engineers can easily obtain a smoothed signal that retains the dent shape (Lukasiewicz et al. 2006; Okoloekwe et al. 2018b; Ergezinger et al. 2020; Zhang et al. 2022). However, there are some potential issues associated with the Fourier transform-based filtering. Since the noises associated with different caliper tools vary, there is a lack of a rigorous, general method to determine the cut-off frequency or filter parameters. Second, the Fourier transform is a global representation of the signal, which means that modifications to any wavenumber component (e.g. reducing the amplitude of the component) may change the values of all data points in the reconstructed signal. Therefore, some features in the signal corresponding to localized high-wavenumber components (e.g. girth welds) can be lost or distorted by the removal of high-wavenumber components.

The wavelet transform is commonly employed to denoise signals containing local features. The wavelet transform-based thresholding method has been widely used to denoise ILI signals (Afzal & Udpa 2002; Siqueira et al 2004; Kathirmani et al. 2012; Sampath et al. 2019). In addition to thresholding, researchers have explored other wavelet-based denoising methods. Liang et al. (2008) constructed an overcomplete dictionary to suppress the noise of the ultrasonic echoes reflected from the inner surface of an offshore pipeline. Wu et al. (2022) incorporated wavelets and neural networks to

denoise the pipeline defect detection signal. Although the wavelet-based denoising has been commonly used to process the ILI signals obtained using magnetic flux leakage and ultrasonic tools, only a few studies (Belanger & Narayanan 2008; Liu et al. 2022) have employed wavelets to denoise the dent signals collected by caliper tools. Therefore, it is valuable to explore the use of the wavelet transform to better denoise the dent signals such that the dent strains can be more accurately estimated.

In the present study, we propose a denoising method based on the wavelet transform to remove noises in the measured dent signals that significantly affect the strain assessment. The proposed method is based on the overcomplete expansion (Goodwin & Vetterli, 1999), whereby the overcomplete dictionary is constructed by the stationary and hyperbolic wavelet transforms (Nason & Silverman 1995; DeVore et al. 1998). Suitable wavelet functions as well as the optimal decomposition levels for the dictionary are determined based on real dent signals measured by caliper tools. FEA is further carried out to numerically simulate both noise-free and noisy dent signals, which allow us to compare the effectiveness of different denoising methods and demonstrate that the proposed method results in better denoising results for the strain assessment than the commonly used wavelet transform-based thresholding.

The rest of the paper is organized as follows. In Section 2.2, we first introduce noisy signals for indented in-service pipelines collected from caliper tools, and then present the finite element model for extracting the noise-free dent signal and methodologies for evaluating strains of an indented pipe segment. Section 2.3 presents fundamentals and different variations of the wavelet transform, and the denoising method based on the overcomplete dictionary constructed from the stationary and hyperbolic wavelet transforms. The effectiveness of the denoising using the proposed overcomplete dictionary-based method and thresholding method commonly reported in the literature is compared in Section 2.4. A small overcomplete dictionary suitable for the dent signal denoising is constructed based on the statistics of the real dent signal decomposition. Concluding remarks are presented in Section 2.5.

2.2   Dent morphology and strain assessment

2.2.1   Noisy dent signals measured from in-service pipelines

In practice, the inner radii of in-service pipelines are measured by the in-line caliper tool to detect and size dents (Zhang et al. 2022). The caliper tool measures the distance between the inner surface of the pipeline and centre of the tool (it is usually assumed that the centre coincides with that of the pipeline before deformation and does not change during the measuring process). Sensors on the caliper tool are arranged along the circumferential direction, and each sensor measures the inner radii along the pipelines at a specific orientation angle (assuming that the tool does not rotate during measuring process) as the tool is moving along the in-service pipeline (Zhang et al. 2022).

Forty-two noisy signals corresponding to measured inner radii of indented pipeline segments are provided by a Canadian pipeline operator to the present study. The outside diameters (OD) and wall thicknesses (wt) of these pipelines are between 609.6 and 1066.8 mm, and 7.80 and 9.88 mm, respectively. All the dents are located on the base metal and away from girth and seam welds. The noise-free measurements corresponding to these noisy signals are however unavailable. The data of each of the signals are organized in a matrix format. Each row of the matrix represents the inner radii measured by a particular sensor, i.e. the inner radius at a specific orientation angle along the pipeline segment. Each column of the matrix represents the inner radii of the pipeline segment along the circumferential direction at a particular longitudinal coordinate. Let $R(x, \theta)$ (mm) denote the measured inner radius at the longitudinal coordinate x (in mm) and the circumferential coordinate θ (in rad). The dent-free inner radius R_{df} is given by $R_{df} = OD/2 - wt$, and the maximum dent depth $d_{\max}$ is defined as R_{df} minus the minimum value of the measured inner radii, i.e. $d_{\max} = R_{df} - \min\{R\}$. The values of $d_{\max}$ of the signals are between 0.6% and 5.7%OD. The longitudinal sampling spacing Δx, i.e. the distance between two longitudinally neighbouring measurement points, in these signals is between 1 and 3 mm. The circumferential sampling spacing $\Delta\theta$ is determined by the number of sensors on the caliper tool. The arc length of the circumferential sampling spacing ($R_{df}\Delta\theta$) of the inner radii measured by the caliper tools are between 30 and 35 mm as governed by the number of sensors in the tool.

Figure 2.1 illustrates three of the collected noisy signals from pipe segments with different values of *OD* and *wt*. The number of sensors corresponding to the three signals are 68, 60, and 100, respectively. Some characteristics of the dent data employed in the present study are revealed by Fig. 2.1. First, the measured inner radii are contaminated by the noise. Although the inner surface of a pipeline can be inherently rough, the roughness is generally too small to be picked up by the caliper tool. Therefore, the roughness in the measured inner radii is considered as the noise in the signal. Second, a dent may only be captured by a very small number of sensors. If an indented area is defined, somewhat arbitrarily, to correspond to $(R_{df} - R)/R_{df} > 1\%$, then the circumferential width of a dent is generally less than one rad. The majority of the 42 dents are captured by only three or four sensors, and in the most extreme case only one sensor captured the dent. As a comparison, most of the 42 dents have a longitudinal length greater than 100 mm, i.e. over 30 grid points, with the minimum length of 50 mm (25 grid points). Therefore, a dent signal may have a dent over 30 grid points along the longitudinal direction, but only one or two grid points along the circumferential direction. The significant differences in the lengths and resolutions along the longitudinal and circumferential directions of the dent signals represent a significant challenge to the signal denoising.
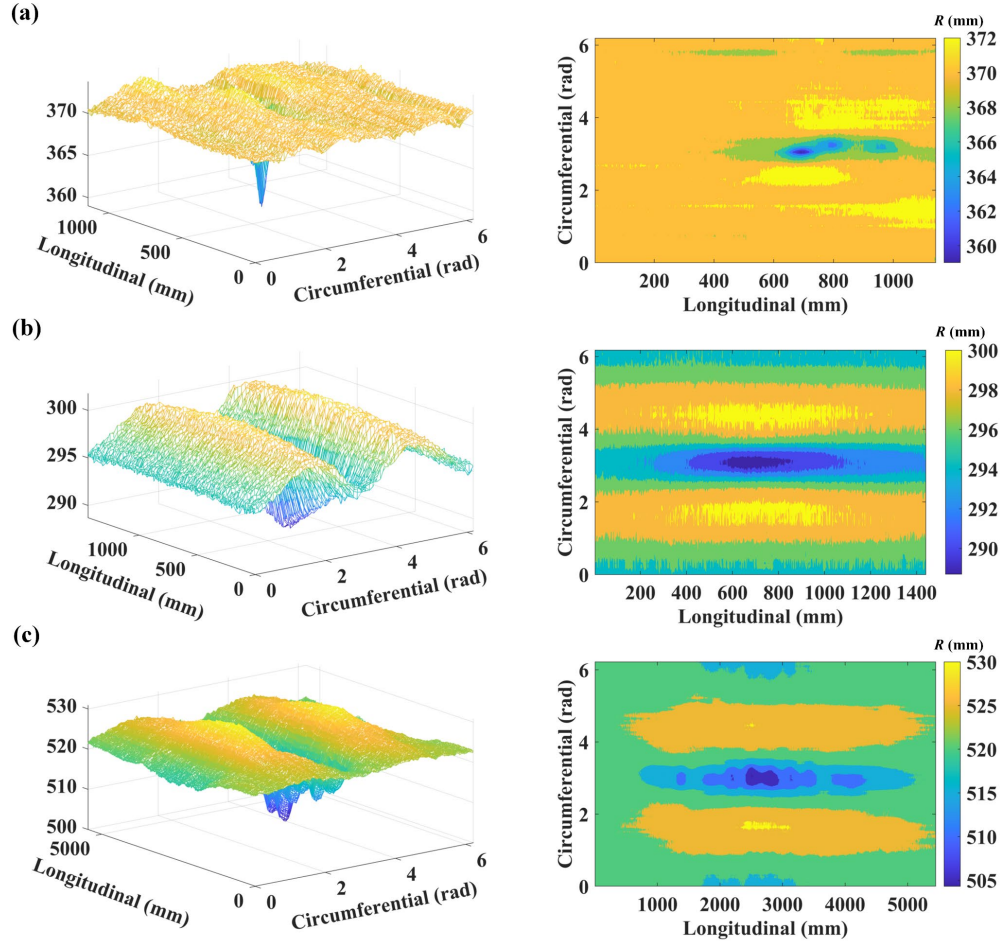
Figure 2.1 Three noisy signals of the inner radii measured from in-service pipelines: (a) $OD = 762$ **mm,** $wt = 9.53$ **mm,** $d_{max} = 1.63\%$ OD**; (b)** $OD = 609.6$ **mm,** $wt = 7.8$ **mm,** $d_{max} = 1.36\%$ OD**; and (c)** $OD = 1066.8$ **mm,** $wt = 9.88$ **mm,** $d_{max} = 1.80\%$ OD

2.2.2   Noise-free dent morphology and strain in FEA

The three-dimensional (3D) elasto-plastic FEA has proven to be an effective tool for the strain and fatigue analyses of a pipe segment subjected to indenters of different shapes (Bao et al. 2022). To obtain the noise-free dent morphology and corresponding strain of an indented pipeline segment, the 3D FEA is employed in this study to evaluate the equivalent strain of a pipe segment ($OD = 609.6$ mm, $wt = 7.6$ mm, length = 1200 mm)

subjected to a rigid hemispherical indenter (diameter = 25 mm = 4.1%*OD*). The diameter and displacement of the indenter are selected based on, respectively, the minimum length and maximum depth of the 42 dent signals described in Section 2.2.1 such that the maximum strain calculated from the FEA is higher than the strains of all the measured dents. The finite element model is developed and implemented using the commercial package ABAQUS 2018. The eight-node hexahedral (C3D8) element with full integration is employed in the model. The radial and circumferential mesh resolutions are 1.5 and 16 mm, respectively. Five layers of elements are used along the pipe wall thickness (radial) direction. The longitudinal mesh resolution in the dented area is 1.5 mm, and coarse meshes with the longitudinal resolution of 10 mm are employed at the dent-free area at both ends. The total number of nodes and elements are 251,099 and 208,576, respectively. One end and the bottom of the pipe segment are considered as a fixed support, and the longitudinal displacement is constrained at the other end. A vertical displacement of 35 mm (about 5.7%*OD*) is assigned to the top surface of the hemispherical indenter, and the pipe is unpressurized. The finite element model is illustrated in Fig. 2.2.
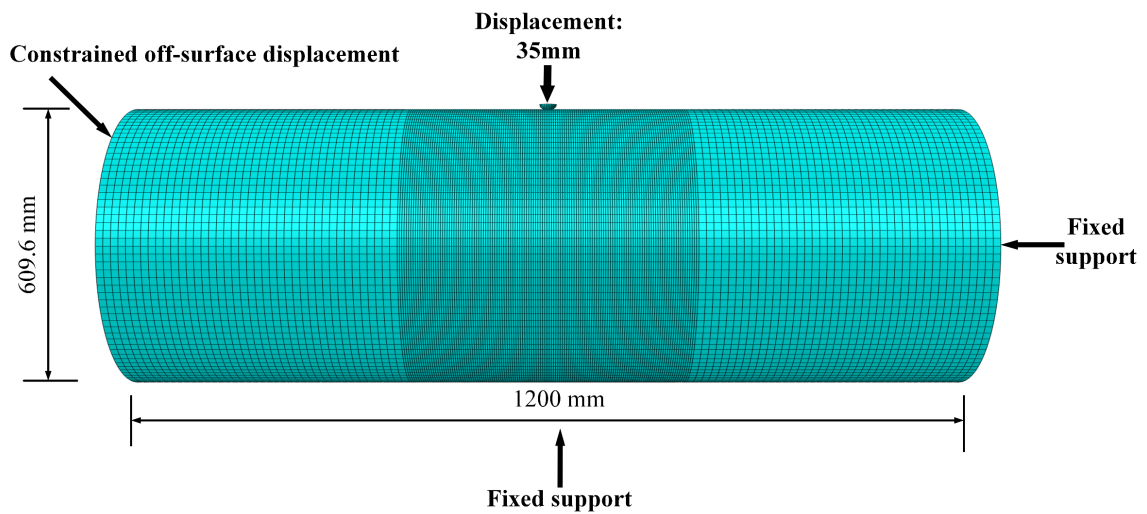
Figure 2.2 The full 3D FEA model of a pipeline segment subjected to dent

The elastic modulus E = 200 GPa, yield strength σ_y = 369 MPa and tensile strength σ_u = 540 MPa are assigned to the pipe steel. The von Mises yield criterion, isotropic strain hardening and finite-strain formulation for the large deformation are adopted in the model. The following true stress - true strain (σ - ε) relationship is adopted in FEA:

$$\sigma = \begin{cases} E\varepsilon, & \sigma < \sigma_y \\ K\varepsilon^{n_h}, & \sigma \geq \sigma_y \end{cases} \tag{2.1}$$

where $K = \frac{\sigma_u e^{n_h}}{(n_h)^{n_h}}$ is the strength coefficient, and n_h is the strain hardening exponent. The value of n_h is estimated using the following empirical equation (Zhu and Leis, 2005):

$$n_h = 0.244 \left(\frac{\sigma_u}{\sigma_y} - 1\right)^{0.604} \tag{2.2}$$

The cylindrical coordinate and plastic equivalent strain ε_{pe} of the inner pipe surface are extracted after computation. The stress analysis is not considered in the present study.

2.2.3 Strain assessment based on dent morphology

In ASME B31.8, the equivalent strain in a dent is calculated from the circumferential bending strain, longitudinal bending strain, and longitudinal extensional (membrane) strain. In the present study, only plain dents (i.e. dents without co-existing damages such as gouges and corrosion) are considered, and the equivalent strain is calculated from the bending and membrane strains along both the longitudinal and circumferential directions. The information required for the strain assessment is extracted from the denoised inner radius signals (Lukasiewicz et al. 2006; Okoloekwe et al. 2018a; Zhang et al. 2022).

The bending strain at a given point on the external or internal surface of the pipe wall is estimated from the curvature of the pipe surface at this point. For a point in a cylindrical coordinate system $R(x, \theta)$, the longitudinal and circumferential curvatures, $\kappa_l(x,\theta)$ and $\kappa_c(x,\theta)$, are calculated by (Okoloekwe et al. 2018b; Zhang et al. 2022):

$$\kappa_l(x,\theta) = \frac{\frac{d^2 R(x,\theta)}{dx^2}}{\left(1 + \left(\frac{dR(x,\theta)}{dx}\right)^2\right)^{3/2}} \tag{2.3}$$

$$\kappa_c(x,\theta) = \frac{R^2(x,\theta)+2\left(\frac{dR(x,\theta)}{d\theta}\right)^2 - R(x,\theta)\frac{d^2R(x,\theta)}{d\theta^2}}{\left(R^2(x,\theta)+\left(\frac{dR(x,\theta)}{d\theta}\right)^2\right)^{3/2}} \tag{2.4}$$

The cubic spline is employed in this study so that the partial derivatives in the above equations can be calculated from the piecewise polynomial functions (Okoloekwe et al. 2018a). Once the curvatures are calculated by Eqs. (2.3) and (2.4), the bending strains along the longitudinal and circumferential directions, $\varepsilon_l^{(b)}(x,\theta)$ and $\varepsilon_c^{(b)}(x,\theta)$, can be calculated by (Zhang et al. 2022):

$$\varepsilon_l^{(b)}(x,\theta) = \frac{wt}{2}\kappa_l(x,\theta) \tag{2.5}$$

$$\varepsilon_c^{(b)}(x,\theta) = \frac{wt}{2}\left(\frac{1}{R_{df}} - \kappa_c(x,\theta)\right) \tag{2.6}$$

Note that the plane section assumption is implied in the bending strain calculation. Since the calculation of bending strains involves the first- and second-order derivatives of the inner radii, the noise, especially the high-wavenumber components of the noise, in the dent signals will significantly affect the strain calculation. If the noise is not effectively removed, the bending strains can be greatly overestimated.

The membrane strain quantifies the extension in the mid-plane of the pipe wall due to the dent deformation. In the cylindrical coordinate system, the membrane strains along the longitudinal and circumferential directions, $\varepsilon_l^{(m)}(x,\theta)$ and $\varepsilon_c^{(m)}(x,\theta)$, can be calculated as follows (Zhang et al. 2022):

$$\varepsilon_l^{(m)}(x,\theta) = \sqrt{1+\left(\frac{dR(x,\theta)}{dx}\right)^2} - 1 \tag{2.7}$$

$$\varepsilon_c^{(m)}(x,\theta) = \frac{\sqrt{R^2(x,\theta)+\left(\frac{dR(x,\theta)}{d\theta}\right)^2} - R_{df}}{R_{df}} \tag{2.8}$$

Thin-walled pipes are usually assumed to be in a plane stress condition such that the radial strain (i.e. change of the pipe wall thickness) is neglected. It follows that the effective strain ε_{eff} can be calculated by (Zhang et al. 2022; Lukasiewicz et al. 2006):

$$\varepsilon_{eff} = \frac{2}{\sqrt{3}} \sqrt{\varepsilon_l^2 + \varepsilon_l \varepsilon_c + \varepsilon_c^2} \tag{2.9}$$

where $\varepsilon_l = \varepsilon_l^{(m)} + \varepsilon_l^{(b)}$ and $\varepsilon_c = \varepsilon_c^{(m)} + \varepsilon_c^{(b)}$. Equation (2.9) is similar to the methodology used in ASME B31.8 to estimate the dent strain, but the circumferential membrane strain is ignored and the longitudinal membrane strain is calculated from the dent length and depth in ASME B31.8. Compared with the plastic equivalent strain calculated in the FEA, the effective strain calculated using Eq. (2.9) takes into account the elastic strain as well as the plastic strain. However, since the plastic strain at the dent apex is much higher than the elastic strain, the maximum effective strain calculated using Eq. (2.9) is close to the plastic equivalent strain obtained in FEA. This will be demonstrated in Section 2.4.

2.3 Wavelet transforms and denoising

2.3.1 Continuous and discrete wavelet transform

Let $v(x)$ denote a one-dimensional (1D) signal. The coefficients of continuous wavelet transform (CWT) of $v(x)$, $v_w(s, \tau)$, are given by (Daubechies 1992):

$$v_w(s, \tau) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} v(x) \psi^* \left(\frac{x-\tau}{s} \right) dx \tag{2.10}$$

where s is the scaling factor; τ is the translation; $\psi(x)$ is the so-called mother wavelet function, and * denotes the complex conjugate. If $\psi(x)$ satisfies the admissibility condition, i.e. $C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(f)|^2}{|f|} df < \infty$, where $\hat{\psi}(f)$ denotes the Fourier transform of $\psi(x)$ with f (Hz) being the frequency, the function $v(x)$ can be reconstructed from $v_w(s, \tau)$ by the inverse CWT:

$$v(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{v_w(s,\tau)}{\sqrt{|s|}} \psi \left(\frac{x-\tau}{s} \right) \frac{1}{s^2} d\tau ds \tag{2.11}$$

In CWT, s and τ are continuous, and both the forward and inverse transforms need to calculate the integral. The high computational cost limits the value of CWT for engineering applications. The non-redundant DWT is more commonly used in practice when processing a discrete signal. Let $v(x)$ ($x = 0, 1, 2, \ldots, 2^n - 1$) denote a 1D discrete

signal, where n is a positive integer. In 1D DWT, $v(x)$ can be decomposed into components constructed with wavelets and scaling functions as follows (Daubechies 1992):

$$v(x) = \sum_k 2^{-\frac{n}{2}} L_n(k)\varphi(2^{-n}x - k) + \sum_{i=1}^{n} \sum_k 2^{-\frac{i}{2}} H_i(k)\psi(2^{-i}x - k) \tag{2.12}$$

where $\varphi(x)$ denotes the scaling function (i.e. the low-pass filter); $\psi(x)$ is the mother wavelet function (i.e. the high-pass filter); k denotes translation; i is the decomposition level, and L and H are the approximation and detail coefficients, respectively. The square of the absolute value of a coefficient is considered the energy of the component corresponding to this coefficient (Mallat 1989). Usually $\varphi(x)$ and $\psi(x)$ are well designed so that $2^{-\frac{n}{2}}\varphi(2^{-n}x - k)$ and $2^{-\frac{i}{2}}\psi(2^{-i}x - k)$ in all terms are orthonormal bases with compact support and fast frequency decay. The approximation coefficients $L_n(k)$ and detail coefficients $H_i(k)$ are obtained by convolving $v(x)$ and the corresponding analysis filters. The approximation coefficients $L_n(k)$ and detail coefficients $H_i(k)$ can be simply considered as the high- and low-frequency components of the signal, respectively.

In 1D DWT, the signal is decomposed to two sub-bands (high-frequency and low-frequency) at each level by the high-pass and low-pass filters, and the coefficients at higher levels are calculated by decomposing the approximation coefficients at lower levels. Given a 2D discrete signal $z(x, y)$ ($x = 0, 1, 2, \ldots, 2^n - 1$; $y = 0, 1, 2, \ldots, 2^m - 1$), the 2D DWT is similar to 1D DWT but with four sub-bands at each level. The wavelet coefficients in 2D DWT, $z_\psi^{(\alpha,i)}(\tau_x, \tau_y)$, is given by (Daubechies 1992):

$$z_\psi^{(\alpha,i)}(\tau_x, \tau_y) = \frac{1}{2^i} \sum_x \sum_y z(x, y)\psi_{(\alpha)}\left(\frac{x-\tau_x}{2^i}, \frac{y-\tau_y}{2^i}\right) \tag{2.13}$$

$$\psi_{(\alpha)}(x, y) = \begin{cases} \varphi(x)\varphi(y), & \alpha = LL \\ \varphi(x)\psi(y), & \alpha = LH \\ \psi(x)\varphi(y), & \alpha = HL \\ \psi(x)\psi(y), & \alpha = HH \end{cases} \tag{2.14}$$

where $\psi_{(\alpha)}(x, y)$ denotes the wavelet function corresponding to the α sub-band; *LL*, *LH*, *HL*, and *HH* denote the one approximation sub-band and three detail sub-bands in 2D DWT: i.e. approximation (low-pass and low-pass), horizontal detail (low-pass and high-pass), vertical detail (high-pass and low-pass) and diagonal detail (high-pass and high-pass), respectively. These sub-bands characterize different properties of the signal: *LL* can be considered as the 2D signal in a lower resolution; *LH* , *HL* and *HH* represent horizontal, vertical, and diagonal patterns of the 2D signal, respectively. Let β denote the number of data points of a signal. DWT can provide a good approximation with sparse wavelet coefficients, and the numerical complexity is only $O(\beta)$ (Mallat 2008). Therefore, it is widely used in data compression and feature detection. Details of the forward and inverse DWT are given in (Daubechies 1992; Mallat 2008; Torrence and Compo 1998).

Many well-developed wavelet functions are available for DWT, such as the Daubechies wavelets (denoted as DBγ, where γ is the order of vanishing moment of the wavelet function), Daubechies symlets (denoted as Symlets γ), coiflets, and biorthogonal wavelets (Mallat 2008). In the present study, we select wavelet functions from the Daubechies wavelets and symlets (γ = 1, 2, …, 10) for denoising because these wavelets have the minimum support given the vanishing moment.

2.3.2   Modifications to the discrete wavelet transform

Although DWT has been widely employed in signal denoising and data compression due to the low numerical complexity, it has some disadvantages (Coifman and Donoho 1995; Chen et al. 2013). One of the disadvantages is that DWT with dyadic decimation is shift-variant due to its down-sampling: all the wavelet coefficients may change significantly if a signal is slightly shifted. To address the shift variance issue, some modifications have been introduced to DWT, one of which is the stationary wavelet transform (SWT) (Nason and Silverman 1995). SWT can be considered as the undecimated DWT with continuous translation at each sub-band, i.e. the τ_x and τ_y in Eq. (2.13) are continuous. In DWT, the higher is the decomposition level, the fewer are the wavelet coefficients. However, the number of wavelet coefficients in all sub-bands in SWT is identical to the number of data

points in the original signal. SWT addresses the shift variance issue and improves the spatial resolution of locating features of the signal, albeit at a higher computational cost.

Another disadvantage of DWT is that the decomposition strategy of 2D DWT in two dimensions is inflexible. In 2D DWT, the bases $\psi_{(\alpha)}(x, y)$ are constructed from the scaling and wavelet functions with identical scales in both dimensions. Therefore, the maximum 2D DWT decomposition level of a 2D signal is constrained by the shorter dimension of the signal. If a signal has dozens of rows but thousands of columns, the maximum decomposition level is constrained by its row number. This may result in that the decomposition along the column direction is insufficient to highlight the characteristics of the signal. This inflexibility can be addressed by using independent scales along two dimensions to generate anisotropic wavelet bases, which is known as the hyperbolic wavelet transform (HWT) (DeVore et al. 1998; Roux et al. 2013). The wavelet coefficients in HWT are calculated by (Farouj et al. 2016):

$$z_\psi^{(\alpha)}\left(i_x, i_y, \tau_x, \tau_y\right) = \frac{1}{2^{\frac{i_x+i_y}{2}}}\sum_x \sum_y z(x,y)\psi_{(\alpha)}\left(\frac{x-\tau_x}{2^{i_x}}, \frac{y-\tau_y}{2^{i_y}}\right) \tag{2.15}$$

where i_x and i_y are the decomposition level along the x and y directions, respectively. Similar to the 2D DWT, usually the translations τ_x and τ_y in HWT are discrete, and the total number of HWT coefficients is identical to the number of data points in the signal. Compared to 2D DWT, HWT has more sub-bands, but the number of translations is limited in each sub-band. Therefore, HWT has a higher wavenumber resolution but a poorer spatial resolution than DWT. As modifications to the 2D DWT, SWT and HWT are theoretically compatible. One can combine SWT and HWT by simply assigning continuous τ_x and τ_y to Eq. (2.15).

2.3.3   Denoising and overcomplete dictionary

Let $z_n(x, y) = z(x, y) + \varepsilon_n(x, y)$ denote a noisy 2D signal, where ε_n represents a Gaussian white noise with variance σ_n^2, i.e. $\varepsilon_n \sim N(0, \sigma_n^2)$. It can be inferred from Eqs. (2.13) and (2.15) that the wavelet transform of z_n is the sum of the wavelet transforms of z and ε_n. Since ε at different locations are independent identically distributed (i.i.d.)

standard Gaussian random variables and $\iint |\psi_{(\alpha)}|^2 dx dy = 1$, the wavelet transform of ε is a quasi-white noise with the corresponding wavelet coefficients also following $N(0, \sigma_n^2)$ (Donoho 1995). Therefore, the basic idea of the wavelet-based denoising is transforming the noisy signal from the time or space domain to the wavelet domain where main components of the noise-free signal are included in a small portion of the coefficients with high energy, and then discarding most of the wavelet coefficients with low energy. If the noise-free signal can be sparsely represented in the wavelet domain, most of the removed wavelet coefficients are calculated from the noise, and the variance of the noise can be estimated from the removed wavelet coefficients (Donoho 1995; Berkner & Wells 1998; Mallat 2008). Typically the fewer wavelet coefficients are needed to represent a noise-free signal, the more effective is the denoising. The commonly used DWT-based denoising methods are soft- and hard-thresholding. A universal threshold η for the thresholding is estimated by (Donoho 1995):

$$\eta = \frac{MAD}{0.6745} \sqrt{2 \ln(\beta)} \qquad\qquad (2.16)$$

where *MAD* is the median absolute value of the wavelet coefficients in the first level detail sub-bands. In hard-thresholding, all the wavelet coefficients with absolute values below η are set to zeros. In soft-thresholding, also known as the wavelet shrinkage, the coefficients with the absolute value below the threshold are set to zeros, while the other coefficients are adjusted toward zero by subtracting η from their modes (Donoho 1995; Mallat 2008).

A signal can be represented as a linear combination of multiple components, which are called atoms. In signal processing, the set of all the atoms employed in the decomposition or reconstruction is called a dictionary (Rubinstein et al. 2010). Given a signal with β data points, if a dictionary includes β orthogonal atoms, this dictionary is complete. Regardless of the value of any point in the original signal, a complete dictionary can always use its atoms to uniquely represent the signal. The process of performing 2D DWT on a 2D signal can be considered as decomposing this signal by a complete dictionary, whose atoms are calculated from the wavelet and scaling functions with different translations and scaling factors. The advantage of using a complete

dictionary is any signal can be uniquely decomposed and perfectly reconstructed, which makes the computation and interpretation convenient. However, the unique decomposition does not imply that it is the sparsest decomposition (i.e. the signal is represented by the minimum number of atoms). It is possible that a signal can be represented with only one atom in one complete dictionary, but requires dozens of atoms in another complete dictionary.

A dictionary including more than β atoms is overcomplete if it can perfectly reconstruct any signal with β points. Since an overcomplete dictionary has more atoms than the minimum number of atoms required for the perfect reconstruction, there are infinite ways to decompose a signal by an overcomplete dictionary. An overcomplete dictionary can be constructed by adding extra atoms to a complete dictionary with the extra atoms obtained by translating, rotating, or stretching the atoms in the complete dictionary, or from another dictionary (Rubinstein et al. 2010). Suppose that all the atoms of a complete dictionary are included in an overcomplete dictionary. In the overcomplete dictionary-based signal decomposition, the number of used atoms can be less than that in the complete dictionary-based decomposition, because the decomposition based on the complete dictionary is only one of the cases of the decomposition based on the overcomplete dictionary. In other words, using an overcomplete dictionary allows a signal to be represented by fewer atoms, which facilitates operations such as signal denoising and compression. However, since the representation is not unique, seeking a sparse representation of the signal becomes a challenging task. Although searching the global optimal solution for the overcomplete dictionary is very complicated, the local optimal solution obtained using some greedy algorithm (e.g. matching pursuit) is satisfactory in many cases (Goodwin and Vetterli, 1999). Unlike thresholding, the overcomplete dictionary-based denoising can be achieved by extracting the major components from the noisy signal one by one until the residual part is considered as pure noise.

As will be shown in Section 2.4, the large differences in the number of rows and columns and the grid spacing in both directions of the measured dent signals lead to poor denoising results and severely overestimated dent strains if the commonly used DWT

thresholding methods are employed. Since a dent signal is not morphologically complicated and may be represented by a very small number of wavelet bases, properly constructing an overcomplete dictionary is a possible solution to the abovementioned problem. As described in Section 2.2.2, the measured inner radii of in-service pipelines are noisy and have different spacing along two directions. This results in that the DWT-based denoising cannot effectively remove the noise. An overcomplete dictionary constructed from HWT and SWT can overcome the deficiencies of DWT. To achieve a sparser representation, different wavelet functions can be selected in the two directions. The following procedure is proposed to extract the dent morphology from the noisy signal z_n:

(1) Select the wavelet functions for the signals. The wavelet functions along both directions can be different. Suppose that N_L and N_C wavelet functions are selected in the longitudinal and circumferential directions, respectively.

(2) Estimate the variance of the noise. Performing one level 2D DWT on the noisy signal, where the wavelet function in each direction is selected as the wavelet with the shortest support length among all the wavelet functions selected in that direction in Step (1). For example, suppose that DB5 and DB7 are selected for the longitudinal direction, and DB4 and DB6 are selected for the circumferential direction. In this case, DB5 and DB4 are selected as the wavelet functions for the longitudinal and circumferential directions, respectively, in the noise variance estimation. The variance is estimated as $\sigma_n^2 = \left(\frac{MAD}{0.6745}\right)^2$ (Donoho 1995), where MAD is the median absolute value of the wavelet coefficients in the LH, HL, and HH sub-bands.

(3) Estimate the energy of the noise-free signal z: $E_z = \sum_x \sum_y \left(z(x,y)\right)^2 \approx \sum_x \sum_y \left(z_n(x,y)\right)^2 - \beta \sigma_n^2$, where β is the number of data points in z_n.

(4) Determine the maximum decomposition levels along two directions. Suppose that M_L and M_C decomposition levels are assigned to the longitudinal and circumferential directions, respectively.

(5) Calculate all the 2D analysis and synthesis filters for different sub-bands. The total number of 2D analysis filters is $4N_LN_CM_LM_C$.

(6) By convolving the noisy signal and all the analysis filters, all the wavelet coefficients can be calculated. For each analysis filter, β wavelet coefficients are calculated. To ensure the number of the wavelet coefficients corresponding to each analysis filter is the same as the number of data points of the signal, the signal will be periodically extended before convolution.

(7) Select the wavelet coefficient with the greatest absolute value. Perform the inverse wavelet transform on the selected wavelet coefficient to calculate the corresponding component $z_r^{(j)}$, where $z_r^{(j)}$ denotes the component calculated at the jth ($j = 1, 2, \ldots$) iteration. Subtract $z_r^{(j)}$ from the noisy signal. The residual signal will be used to calculate the wavelet coefficients in the next iteration.

(8) Let j_c denote the number of components that have been subtracted in the current iteration. Let $z_r = \sum_{j=1}^{j_c} z_r^{(j)}$ denote the reconstructed signal based on all the calculated components in the current iteration. If $\sum_x \sum_y \left(z_r(x,y) \right)^2 < E_z$, repeat Steps (6)-(7) on the residual signal.

(9) Once $\sum_x \sum_y \left(z_r(x,y) \right)^2 > E_z$, z_r is considered as the noise-free dent signal and can be used to estimate the maximum effective strain.

The aforementioned procedure extracts the main components from the noisy signal based on an overcomplete dictionary constructed by HWT, SWT, and specific wavelet functions. Most of the noise are suppressed, as the total number of extracted components is typically less than 0.01β. However, to extract one component, $4N_LN_CM_LM_C\beta$ wavelet coefficients are calculated, which is computationally expensive. To reduce the computational cost, it is necessary to reduce the size of the overcomplete dictionary. To this end, we first construct a large overcomplete dictionary by assigning multiple wavelet functions and high decomposition levels along two directions. The large overcomplete dictionary is then used to decompose 32 of the 42 real dent signals described in Section

2.2.1. The statistics of the extracted components (i.e. wavelet coefficients) for the 32 signals provide insights into the actual dent morphology, whereas the remaining ten signals are used for validation. Based on such insights, we then construct a smaller overcomplete dictionary by limiting the decomposition levels and selecting the suitable wavelet functions while ensuring that the dent signals can still be sparsely represented. The small overcomplete dictionary is then employed to denoise all 42 noisy dent signals to investigate its adequacy.

2.4   Results and discussion

2.4.1   Results based on FEA

The FEA and denoising are carried out by using a server with two Intel(R) Xeon Gold 5218R CPU @2.10 GHz (20 cores and 40 threads) and 768 GB RAM. The cylindrical coordinates and the plastic equivalent strains of the inner surface of the pipe model shown in Fig. 2.2 after indentation are illustrated in Fig. 2.3. The dent apex is located at $x = 600$ mm and $\theta = \pi$. The maximum plastic equivalent strain $\varepsilon_{pe,\max} = 28.70\%$ appears at the dent apex. Although the dent is small, it leads to a high curvature around the dent apex, and the high bending strain results in a high plastic equivalent strain. Therefore, the strain at the dent apex is much greater than the critical strain of 6% specified in AMSE B31.8.
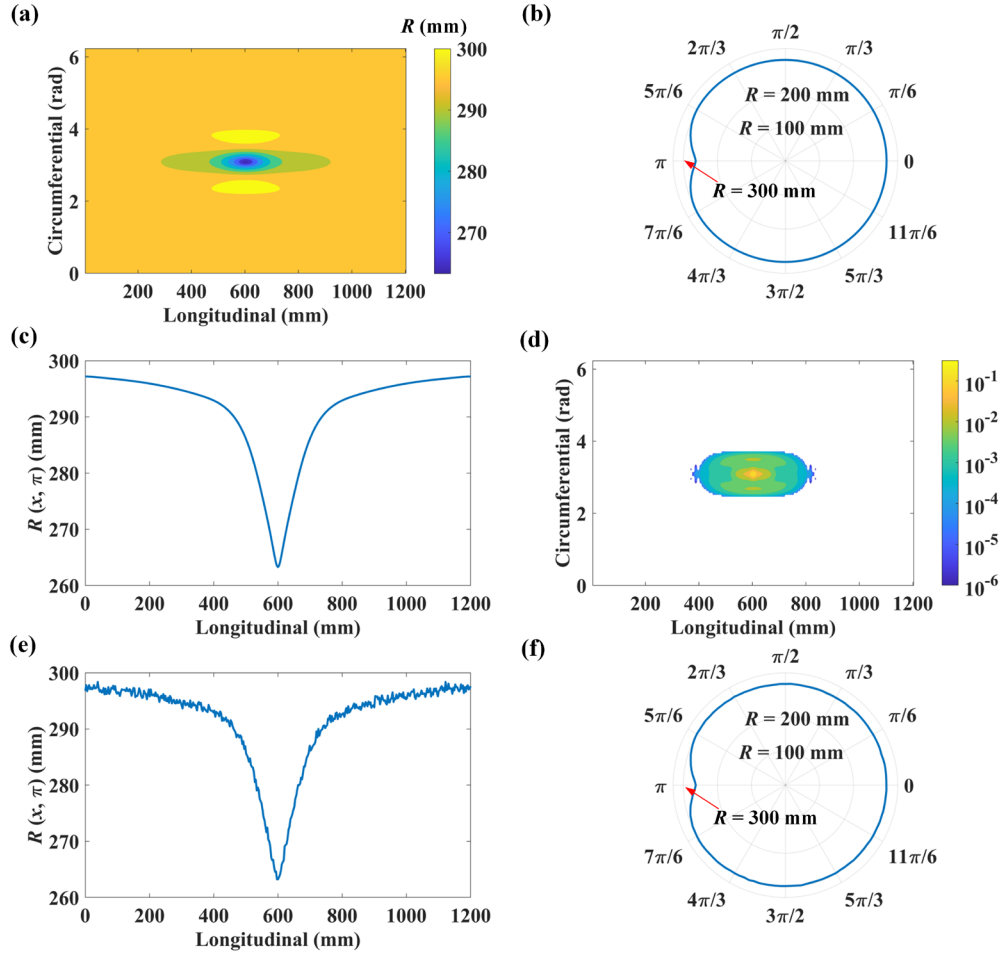
Figure 2.3 The cylindrical coordinate and the plastic equivalent strain of the pipe inner surface after denting: (a) 2D plot of $R(x, \theta)$; (b) inner radii at $x = 600$ mm (crossing the dent apex); (c) inner radii at $\theta = \pi$ (crossing the dent apex); (d) plastic equivalent strain; (e) noisy inner radii ($\sigma_n = 0.5$ mm) at $\theta = \pi$ (crossing the dent apex); and (f) noisy inner radii ($\sigma_n = 0.5$ mm) at $x = 600$ mm (crossing the dent apex)

To simulate the inner radii measured by caliper tools in practice, the inner surface from the FEA result is sampled with a longitudinal sampling spacing $\Delta x = 3$ mm and a circumferential sampling spacing $\Delta \theta = 2\pi/60$. Let $R_0(x, \theta)$ denote the noise-free dent signal. The effective strains of this indented pipe segment estimated by the method

described in Section 2.2.3 are shown in Fig. 2.4. The maximum effective strain $\varepsilon_{eff,max}$ = 26.97% also appears at the dent apex. The strain is slightly lower than $\varepsilon_{pe,max}$ = 28.70% calculated by FEA. This difference is due likely to the sampling spacing used to simulate the measurements of the caliper tool.   In FEA, the longitudinal and circumferential spacing of the nodes near the dent are 1.5 and 16 mm, respectively, which is about half of the corresponding sampling spacing used to simulate the caliper measurements. Although the values of $R_0(x, \theta)$ are accurate at the sampled grid points in the simulated caliper measurements, the curvature at the grid points may be underestimated due to the larger spacing, leading to slightly underestimated bending strains.   However, $\varepsilon_{eff,max}$ obtained using Eq. (2.9) is only marginally lower than $\varepsilon_{pe,max}$ calculated in FEA, which demonstrates the validity and accuracy of Eq. (2.9).
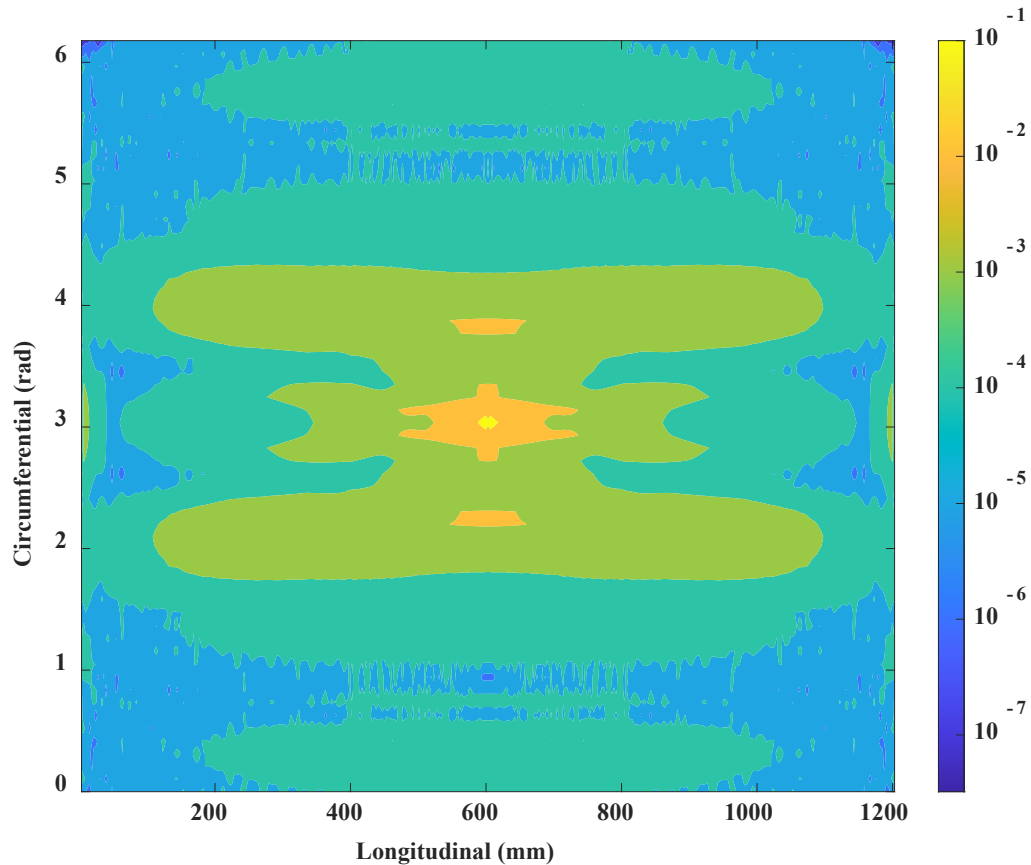
Figure 2.4 The effective strains estimated from the measured inner surface

The effective strains shown in Fig. 2.4 are calculated based on the noise-free $R_0(x, \theta)$, which is extremely difficult, if not impossible, to obtain from in-service pipelines. Let $R_n(x, \theta) = R_0(x, \theta) + \varepsilon(x, \theta)$ and $R_d(x, \theta)$ denote the noisy and denoised signals, respectively, where $\varepsilon \sim$ i.i.d. $N(0, \sigma_n^2)$. The DWT-based hard- and soft-thresholding are employed to denoise R_n with σ_n assumed to be 0.1, 0.3, and 0.5 mm. The root mean square error $RMSE = \sqrt{\dfrac{\Sigma_x \Sigma_\theta [R_d(x,\theta) - R_0(x,\theta)]^2}{\beta}}$, and the strain of the indented pipe estimated from the denoised signal are employed to quantify the effectiveness of denoising. The denoising results and estimated strains are compared in Fig. 2.5. The selected wavelet function in Fig. 2.5 for each scenario has the lowest $RMSE$ among all the wavelet functions mentioned in Section 2.3.1. However, neither the hard-thresholding nor soft-thresholding leads to satisfactory denoising results. Although $RMSE$ for all the denoised signals are low, the effective strains estimated from the denoised signals are much higher than the actual strain; for example, the maximum effective strains $\varepsilon_{eff,max}$ calculated from signals denoised using the hard- and soft-thresholding ($\sigma_n = 0.5$ mm) equal 135.40% and 44.38%, respectively, compared with $\varepsilon_{pe,max} = 28.70\%$ obtained in FEA.
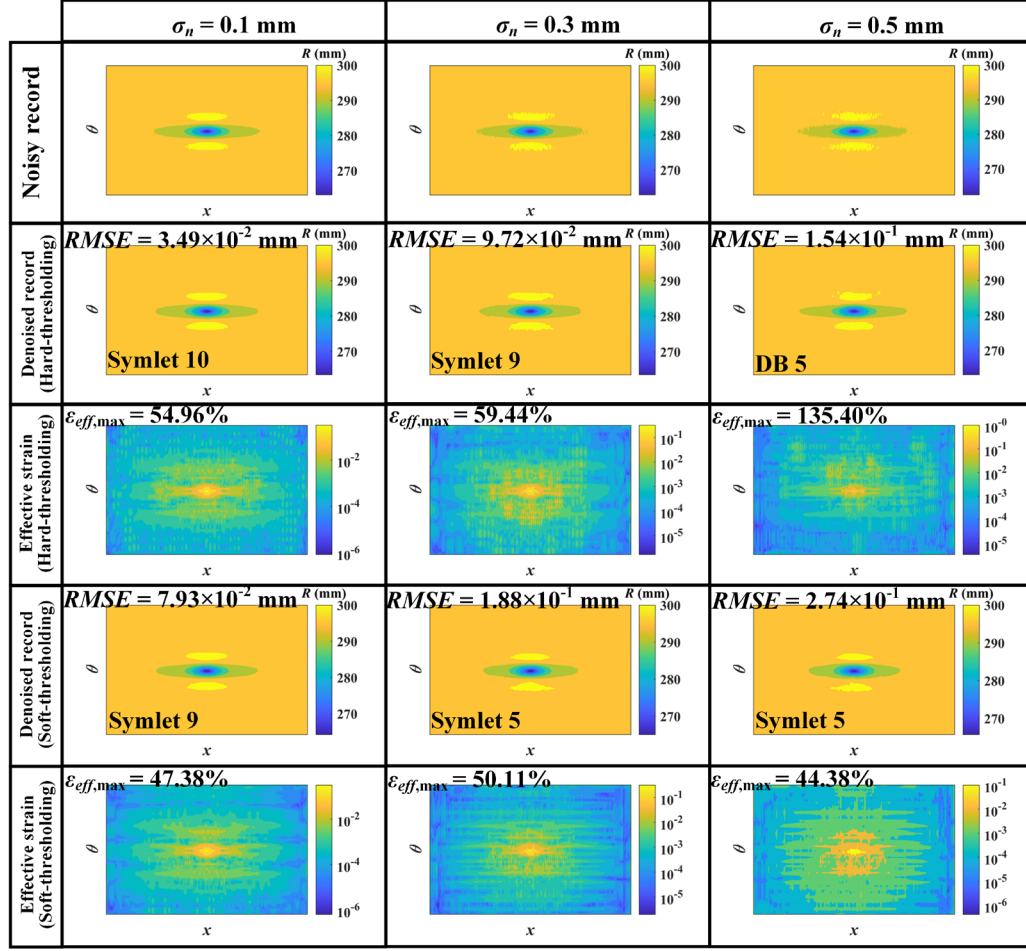
Figure 2.5 Denoising based on the hard- and soft-thresholding and the effective strain estimation based on the denoised signals

Figure 2.6 depicts results of the proposed overcomplete dictionary-based denoising method. The wavelet functions for the longitudinal and circumferential directions are DB10 and Symlet 4, respectively. The maximum decomposition levels in the longitudinal and circumferential directions are 8 and 3, respectively. Table 2.1 compares the $RMSE$ and $\varepsilon_{eff,\max}$ corresponding to the proposed denoising method, hard- and soft-thresholding. The $RMSE$ of the proposed denoising method is higher than those of the hard- and soft-thresholding for $\sigma_n = 0.1$ mm but lower for $\sigma_n = 0.3$ mm and 0.5 mm.

Since a noise-free signal can be decomposed into several components with different energies, there are two parts of *RMSE* during denoising: (1) the energies of the misestimated components, and (2) some components are excluded due to their low energies. For the proposed denoising method, the numbers of extracted components and associated atoms corresponding to the three assumed values of the noise variance are identical, which suggests that part (2) of *RMSE* is unchanged for the above example.
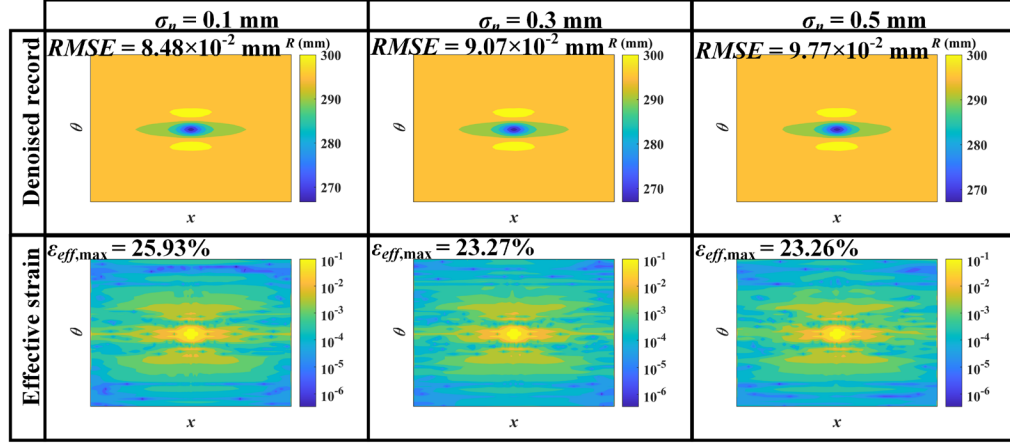
Figure 2.6 Denoising and effective strain assessment based on the overcomplete dictionary constructed from HWT and SWT

Table 2.1 *RMSE* and $\varepsilon_{eff,max}$ of the three denoising method ($\varepsilon_{pe,max}$ = 28.70%)

	σ_n = 0.1 mm		σ_n = 0.3 mm		σ_n = 0.5 mm	
	RMSE (mm)	$\varepsilon_{eff,max}$	RMSE (mm)	$\varepsilon_{eff,max}$	RMSE (mm)	$\varepsilon_{eff,max}$
Proposed method	8.48×10^{-2}	25.93%	9.07×10^{-2}	23.27%	9.77×10^{-2}	23.26%
Hard-thresholding	3.49×10^{-2}	54.96%	9.72×10^{-2}	59.44%	1.54×10^{-1}	135.40%
Soft-thresholding	7.93×10^{-2}	47.38%	1.88×10^{-1}	50.11%	2.74×10^{-1}	44.38%

To quantify part (2) of *RMSE* associated with the proposed method, we calculate the noise-free coefficients corresponding to these extracted components, where the noise-free coefficients are obtained from the noise-free signal. These noise-free coefficients are then used to reconstruct a new signal. By calculating the *RMSE* between the new signal and the noise-free signal, part (2) of *RMSE* can be quantified. For the above example, it is estimated to be 8.41×10^{-2} mm, which is higher than the *RMSE* of the thresholding-based denoising for $\sigma_n = 0.1$ mm. It follows that the denoising performance of the proposed method is not as good as the thresholding-based denoising methods in terms of *RMSE* if the variance of the noise is low. However, *RMSE*s of the thresholding-based denoising methods are more sensitive to the variance of the noise than that of the proposed method. The variance of the noise mainly affects part (2) of *RMSE* associated with the proposed method, while it affects both parts of *RMSE* for the thresholding-based methods. For the thresholding-based denoising methods, as the variance of the noise increases, fewer wavelet coefficients are preserved, and the differences between the values of preserved coefficients and those of noise-free coefficients increase. Therefore, for $\sigma_n \geq 0.3$ mm, the proposed method leads to a better denoising performance in terms of *RMSE* than the thresholding method.

The maximum effective strain estimated from the signal denoised by the proposed method is in excellent agreement with the actual effective strain. Since the maximum effective strain is usually governed by the longitudinal bending strain, this implies that the proposed method can better suppress the noise that causes high bending strains. Note that the maximum effective strain corresponding to the proposed method is consistently slightly lower than the actual equivalent strain. This can be explained by noting that the proposed method extracts the high energy components from the noisy signal one by one until the total energy of the reconstructed signal reaches the target energy level, as opposed to removing the low energy components in the thresholding-based method. Most of the high-energy components of the noise-free dent signal are mainly low-wavenumber components, which means that some low-energy high-wavenumber components of the noise-free signal that can cause high strain may be excluded.

2.4.2   Results based on real dent signals

The 42 collected noisy dent signals are denoised by the proposed method. The estimated values of σ_n for these signals are between 0.09 and 0.53 mm. The wavelet functions for both directions include all the Daubechies wavelets and Daubechies symlets with no more than 10 vanishing moments. The maximum decomposition levels are determined by the number of grid points of the signals in both directions. The size of the overcomplete dictionary depends on the number of grid points of the signal, number of wavelet functions involved and the maximum decomposition levels in both directions. It follows that each component extracted from the noisy signal requires the calculation of tens of millions of wavelet coefficients. Although using such a large overcomplete dictionary ensures that the signal can be represented by as few components as possible, the computational cost is high: it takes several hours to denoise a 120×4800 signal using the server described in Section 2.4.1.

To improve the computational efficiency of the proposed denoising method, we seek to reduce the size of the overcomplete dictionary. To this end, we first analyze the statistics of the components extracted from 32 of the 42 real dent signals to gain a better understanding of the effectiveness of the considered different wavelet functions and decomposition levels. Figure 2.7 compares the statistics of the components extracted from the 32 signals, including the numbers and energy of the components corresponding to different wavelet functions and decomposition levels along both directions. In terms of the numbers of extracted wavelet coefficients along the longitudinal direction, the mostly used wavelet function is DB10, followed by Symlet 5, Symlet 7, and Symlet 9. In the circumferential direction, Symlet 4 and DB4 are the mostly used wavelet functions. Most of the extracted components are constructed by the wavelet filters with a longitudinal decomposition level between 5 and 7, and a circumferential decomposition level of 2 or 3.

(a)
Longitudinal wavelet (Energy proportion)

(b)
Longitudinal wavelet (coefficient number)

(c)
Circumferential wavelet (Energy proportion)

(d)
Circumferential wavelet (coefficient number)

(e)
Longitudinal decomposition level

(f)
Circumferential decomposition level
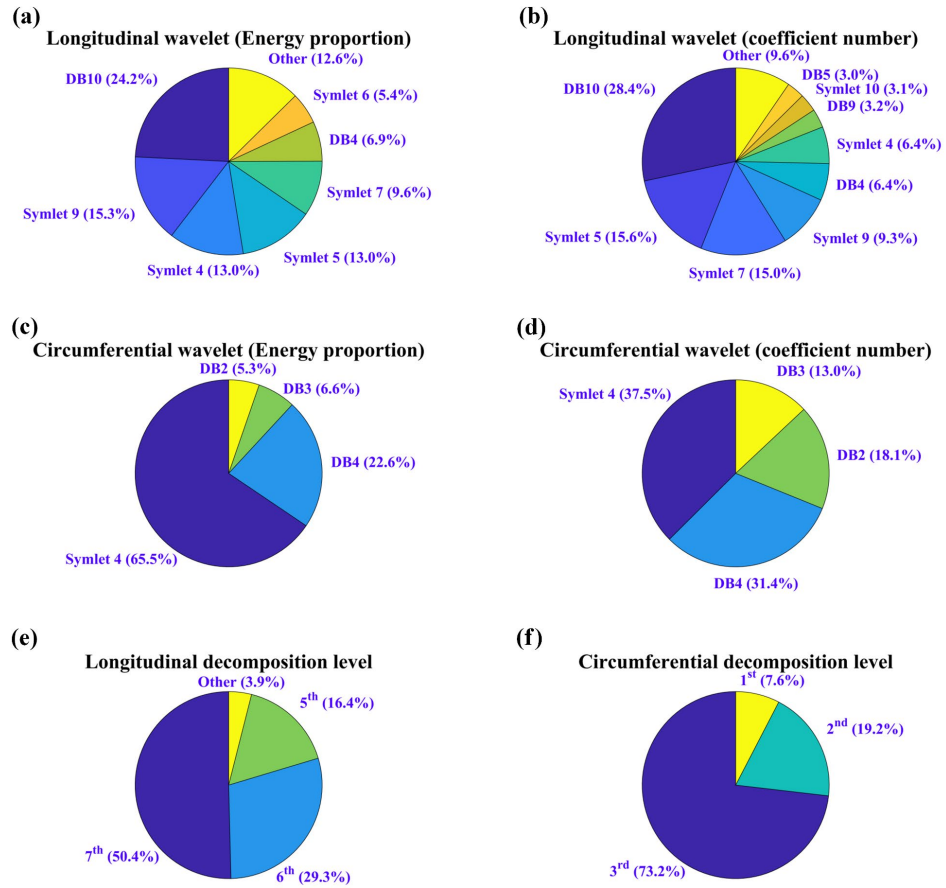
Figure 2.7 Statistics of the components extracted from the noisy dent signals: (a) the energy of components (longitudinal wavelet functions); (b) the number of components (longitudinal wavelet functions); (c) the energy of components (circumferential wavelet functions); (d) the number of components (circumferential wavelet functions); (e) the number of components (longitudinal decomposition level); and (f) the number of components (circumferential decomposition level)

The above-described statistics of the extracted components provide the basis to construct a smaller dictionary. Based on the results in Fig. 2.7, we can consider employing fewer wavelet functions and limiting the maximum decomposition level along two directions to

construct an overcomplete dictionary. DB10 is selected as the wavelet function for the longitudinal direction, and a maximum decomposition level of 7 is considered adequate. For the circumferential direction, Symlet 4 is a suitable choice, and a maximum decomposition levels of 3 is considered adequate. The number of atoms in a small dictionary constructed by using one wavelet function in each of the two directions, and 7 and 3 decomposition levels in the longitudinal and circumferential directions, respectively, is on average only 1/550 of that in the large dictionary employed to generate the results shown in Figs. 6 and 7. The denoising results based on this dictionary for the three dent signals shown in Fig. 1 are illustrated in Fig. 8 below. Compared to Fig. 2.1, a significant change in Fig. 2.8 is that the denoised signals are smoother than the noisy signals. The differences between the longitudinal locations of the dent apexes before and after denoising are between 2 and 6 mm. The differences between the dent depths of the denoised and noisy signals are 0.33 to 1.09 mm, which are within the caliper tools' measurement tolerance of 0.51%OD. Results in Fig. 2.8 suggest that the proposed denoising method based on the small dictionary can adequately preserve the depth, location, and shape of the dent.
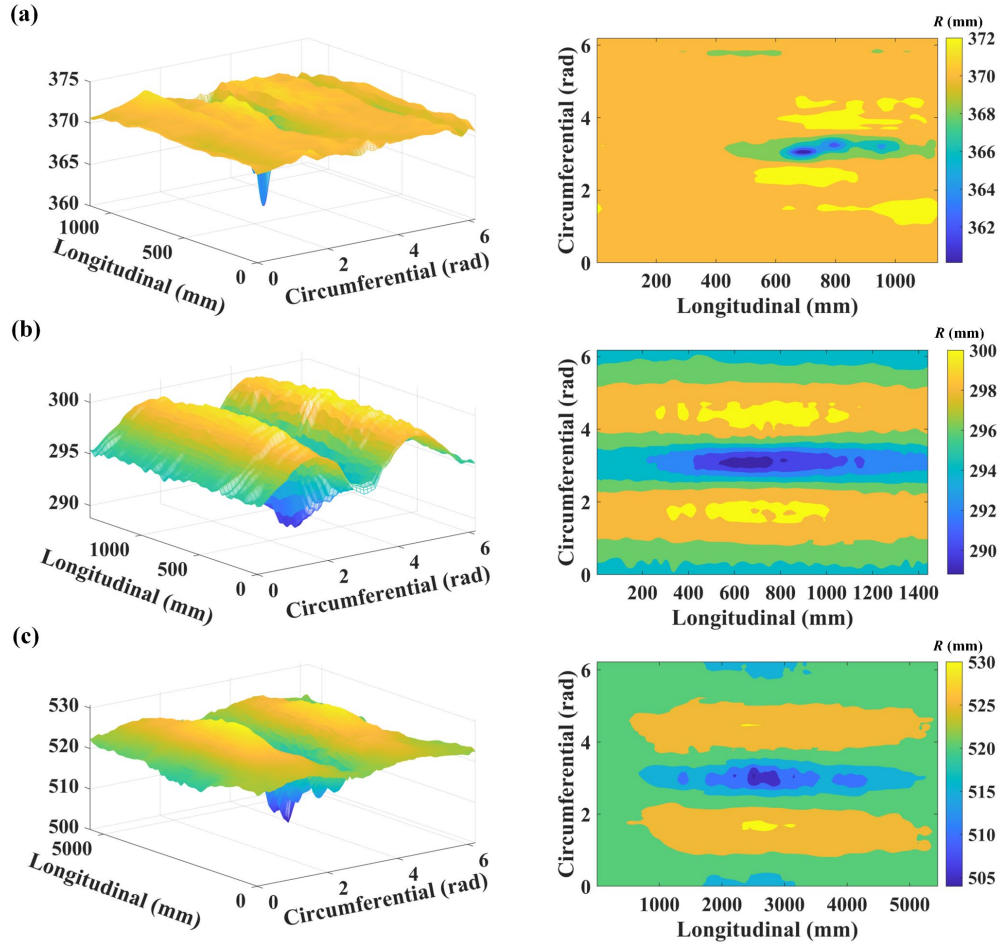
Figure 2.8 Using the small dictionary to denoise the three noisy dent signals shown in Figure 1: (a) $\varepsilon_{eff,max}$ = 1.80%, dent apex longitudinal location difference = 2 mm, dent apex depth difference = 1.09 mm; (b) $\varepsilon_{eff,max}$ = 1.37%, dent apex longitudinal location difference = 3 mm, dent apex depth difference = 0.11 mm; and (c) $\varepsilon_{eff,max}$ = 3.11%, dent apex longitudinal location difference = 6 mm, dent apex depth difference = 0.33 mm

If both the large and small dictionaries are applied to all the 42 collected noisy dent signals, the number of extracted components based on the small dictionary is on average 1.56 times that based on the large dictionary. If we assume the energy of the residual

noise in the denoised signals is proportional to the number of extracted components, the residual noise energy based on the small dictionary is 56% higher than that based on the large dictionary. However, the time required to denoise a signal using the small dictionary is on average 1/48 of that using the large dictionary. The number of the extracted components using the small dictionary is on average 13% of the number of wavelet coefficients preserved by the hard- and soft-thresholding. Therefore, it is feasible to denoise the noisy dent signals and estimate the strain using the proposed method based on the small dictionary. Let $\varepsilon_{eff,\max}^{(L)}$ and $\varepsilon_{eff,\max}^{(S)}$ denote the maximum effective strains calculated from the signal denoised by the large and small dictionaries, respectively. Figure 2.9 compares $\varepsilon_{eff,\max}^{(L)}$ and $\varepsilon_{eff,\max}^{(S)}$ of the 42 dent signals, and the mean and standard deviation of $\varepsilon_{eff,\max}^{(S)} - \varepsilon_{eff,\max}^{(L)}$ are -0.24% and 2.52%, respectively. For the 10 dent signals used for validation purpose, the mean and standard deviation of $\varepsilon_{eff,\max}^{(S)} - \varepsilon_{eff,\max}^{(L)}$ are -0.95% and 1.53%, respectively. These results indicate that the effective strains estimated from the signal denoised using the small dictionary are on average in a good agreement with those from the signals denoised using the large dictionary.
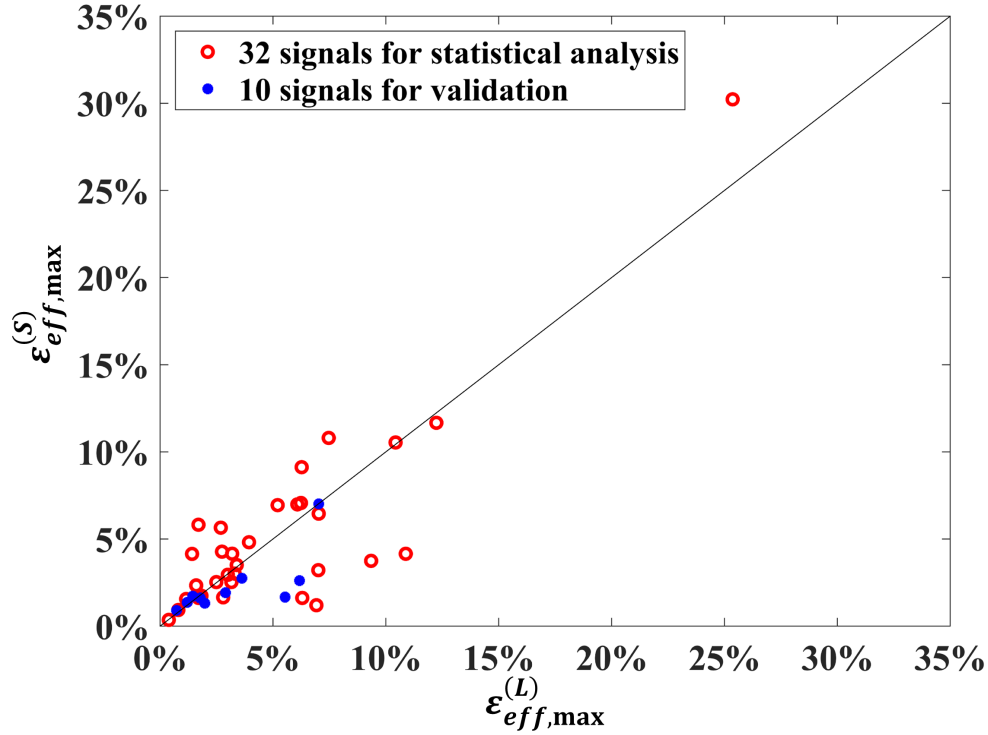
Figure 2.9 The maximum effective strains calculated from 42 dent signals denoised by the proposed method based on the small and large dictionaries

For some dent signals, there is a large difference in the effective strains obtained from the two dictionaries. One reason for the large difference in the effective strains is that the locations corresponding to $\varepsilon_{eff,\max}^{(S)}$ and $\varepsilon_{eff,\max}^{(L)}$ are different. Among the sixteen dent signals with an estimated maximum effective strain difference greater than 1% (i.e. $\left|\varepsilon_{eff,\max}^{(L)} - \varepsilon_{eff,\max}^{(S)}\right| > 1\%$), ten signals have $\varepsilon_{eff,\max}^{(S)}$ and $\varepsilon_{eff,\max}^{(L)}$ at different locations (longitudinal or circumferential location difference > 50 mm). Figure 2.10 compares the denoised signals based on the two dictionaries for one dent signal with $OD = 609.6$ mm, $wt = 7.8$ mm and $d_{\max} = 1.1\%OD$. The maximum effective strains estimated from the two denoised signals occur at different locations: $\varepsilon_{eff,\max}^{(S)}$ occurs near the dent apex whereas $\varepsilon_{eff,\max}^{(L)}$ occurs outside of the dented area. If we only focus on the dented area, $\varepsilon_{eff,\max}^{(L)}$

$= 5.58\%$, which is close to $\varepsilon^{(S)}_{eff,\mathrm{max}} = 4.16\%$. For the denoised signal based on the large dictionary, one of the extracted components at the position of $\varepsilon^{(L)}_{eff,\mathrm{max}}$ is constructed by the 1st level wavelet function of Symlet 4, which may be the reason for the high strain. Since the noise-free signal is not available, we cannot determine whether this component should be considered as noise. It is possible that some atoms in the large dictionary are very unlikely to be used in the noise-free signals. However, these atoms may construct some noise components with high energy, leading to the noise components being extracted during denoising and inaccurately estimated strains.
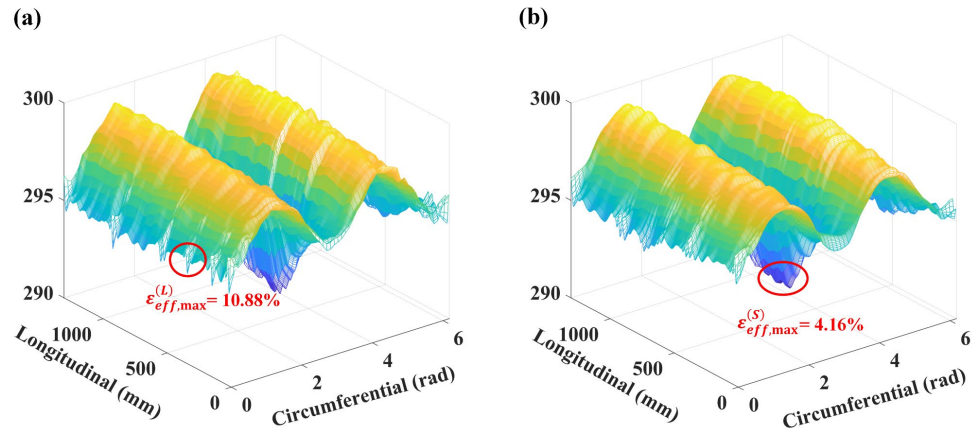
Figure 2.10 Denoised signals based on: (a) the large dictionary; and (b) the small dictionary

The proposed denoising method based on the overcomplete dictionary constructed by SWT and HWT is shown to be more effective than the commonly used thresholding-based methods; however, there are certain limitations associated with the proposed method. First, the proposed method is in general computationally expensive. If we assume that the number of extracted components is proportional to the size of the signal, the numerical complexity of the proposed method based on either the large or small dictionary is $O(\beta^2 \log\beta)$. Therefore, even with a small dictionary, denoising a given signal can be a lengthy process if the signal is large. Second, the proposed method mainly retains low-wavenumber components, while the majority of high-wavenumber

components, whether they are noise or actual features of the dent, are removed during the denoising process. It follows that the effectiveness of the proposed method may be reduced if there are narrow features, e.g. the girth and seam welds on the pipeline, overlapping with the dent. Therefore, further studies are needed to improve the computational efficiency of the proposed method and effectiveness of the method for denoising dent signals containing high-wavenumber features.

2.5 Conclusion

Denoising the signal of dents in steel pipelines obtained from the caliper inspection tool is critical to the strain-based fitness-for-service pipeline dent assessment. The present study investigates the wavelet transform-based denoising method for dent signals . It is demonstrated that the commonly used wavelet transform-based hard- and soft-thresholding denoising methods are ineffective due to the characteristics of dent signals measured from the caliper tool, i.e. the lengths and sampling spacings of the signal along the longitudinal and circumferential directions differ markedly. To overcome the deficiencies of the thresholding-based denoising method, we propose to denoise the dent signals by the overcomplete expansion, whereby the overcomplete dictionary is constructed by SWT and HWT. The effectiveness of the proposed denoising method is validated by using FEA to simulate a dent on a pipe segment and generating simulated dent signals based on results of FEA. Under all the considered noise levels, the maximum effective strain in the dent estimated from the signals denoised using the proposed method are markedly closer to the actual strain than that based on the thresholding method.

To reduce the size of the overcomplete dictionary employed in the proposed method to improve its computational efficiency, we denoise 42 real dent signals collected from in-service pipelines using a large overcomplete dictionary constructed by all the Daubechies wavelets and Daubechies symlets with no more than 10 vanishing moments. The statistics of components of the actual dent morphology extracted from 32 of 42 signals provide the basis to construct a small dictionary using DB10 and Symlet 4 along the longitudinal and circumferential directions, respectively. The computing time to denoise a dent signal using the small dictionary is on average only 1/48 of that using the large

dictionary. The maximum effective strains of the indented pipe segments estimated from the signal denoised by using the small dictionary are on average in good agreement with those estimated from the signals denoised using the large dictionary for the 42 real dent signals. Future studies are recommended to further improve the computational efficiency of the proposed denoising method and effectiveness of the method for denoising dent signals containing high-wavenumber features such as girth and seam welds.

References

Afzal, M., & Udpa, S. (2002). Advanced signal processing of magnetic flux leakage data obtained from seamless gas pipeline. *NDT & E International*, *35*(7), 449-457.

API (2020). *API Recommended Practice (RP) 1183, Assessment and management of dents in pipelines*. American Petroleum Institute, Washington, DC (2020).

ASME (2018). *Gas transmission and distribution piping systems - ASME Code for Pressure Piping, B31.8*. The American Society of Mechanical Engineers, New York (2018).

Bao, J., Zhang, S., Zhang, B., Wang, R., & Zhang, K. (2022, September). An Automatic Dent Assessment Tool Using Finite Element Method. In *International Pipeline Conference* (Vol. 86571, p. V002T03A028).

Belanger, A. A., & Narayanan, R. (2008, January). Direct strain calculation of pipe line dent from knot migration using a kinematic model free of material properties. In *International Pipeline Conference* (Vol. 48586, pp. 267-274).

Berkner, K., & Wells, R. O. (1998, November). Wavelet transforms and denoising algorithms. In *Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers* (Cat. No. 98CH36284) (Vol. 2, pp. 1639-1643). IEEE.

Chen, G., Xie, W., & Zhao, Y. (2013, June). Wavelet-based denoising: A brief review. In *2013 fourth international conference on intelligent control and information processing (ICICIP)* (pp. 570-574). IEEE.

Coifman, R. R., & Donoho, D. L. (1995). *Translation-invariant de-noising* (pp. 125-150). Springer New York.

CSA (2019). *Oil and gas pipeline systems. CSA standard Z662-19.* Canadian Standard Association, Mississauga, Ontario (2019).

Czyz, J. A., Lukasiewicz, S. A., Sun, C., & Adeeb, S. (2008). Calculating dent strain. *Pipeline and Gas Technology*, *2*, 38-45.

Daubechies, I. (1992). *Ten lectures on wavelets*. Society for industrial and applied mathematics.

DeVore, R. A., Konyagin, S. V., & Temlyakov, V. N. (1998). Hyperbolic wavelet approximation. *Constructive Approximation*, *14*(1), 1-26.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE transactions on information theory*, *41*(3), 613-627.

Ergezinger, N., Virk, A. S., Woo, J., Kainat, M., & Adeeb, S. (2020, September). Application of Noise Filtering Techniques for the Quantification of Uncertainty in

Dent Strain Calculations. In *International Pipeline Conference* (Vol. 84447, p. V001T03A026). American Society of Mechanical Engineers.

Farouj, Y., Freyermuth, J. M., Navarro, L., Clausel, M., & Delachartre, P. (2016). Hyperbolic Wavelet-Fisz denoising for a model arising in Ultrasound Imaging. *IEEE Transactions on Computational Imaging*, *3*(1), 1-10.

Gao, M., & Krishnamurthy, R. (2015). Mechanical damage in pipelines: a review of the methods and improvements in characterization, evaluation, and mitigation. *Oil and Gas Pipelines*, 289-326.

Gao, M., McNealy, R., Krishnamurthy, R., & Colquhoun, I. (2008, January). Strain-based models for dent assessment: a review. In *International Pipeline Conference* (Vol. 48586, pp. 823-830).

Goodwin, M. M., & Vetterli, M. (1999). Matching pursuit and atomic signal models based on recursive filter banks. *IEEE Transactions on Signal Processing*, *47*(7), 1890-1902.

He, Z., & Zhou, W. (2021). Fatigue reliability analysis of dented pipelines. *Journal of Pipeline Science and Engineering*, *1*(3), 290-297.

Kathirmani, S., Tangirala, A. K., Saha, S., & Mukhopadhyay, S. (2012). Online data compression of MFL signals for pipeline inspection. *NDT & E International*, *50*, 1-9.

Liang, W., Que, P. W., Lei, H. M., & Chen, L. (2008). Matching pursuit for decomposition and approximation of ultrasonic pulse-echo wavelet and its

application in ultrasonic nondestructive evaluation. *Review of Scientific Instruments*, *79*(7), 075105.

Liu, S., Wang, H., Li, R., & Ji, B. (2022). A novel feature identification method of pipeline in-line inspected bending strain based on optimized deep belief network model. *Energies*, *15*(4), 1586.

Lukasiewicz, S. A., Czyz, J. A., Sun, C., & Adeeb, S. (2006, January). Calculation of strains in dents based on high resolution in-line caliper survey. In *International Pipeline Conference* (Vol. 42622, pp. 129-134).

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, *11*(7), 674-693.

Mallat, S. (2008). *A Wavelet Tour of Signal Processing: The Sparse Way* 3rd ed (New York: Academic).

Nason, G. P., & Silverman, B. W. (1995). The stationary wavelet transform and some statistical applications. *Wavelets and statistics*, 281-299.

Noronha Jr, D. B., Martins, R. R., Jacob, B. P., & de Souza, E. (2010). Procedures for the strain based assessment of pipeline dents. *International Journal of Pressure Vessels and Piping*, *87*(5), 254-265.

Okoloekwe, C., Aranas, N., Kainat, M., Langer, D., Hassanien, S., Roger Cheng, J. J., & Adeeb, S. (2018a). Improvements to the ASME B31. 8 dent strain equations. Journal of Pressure Vessel Technology, 140(4).

Okoloekwe, C., Kainat, M., Langer, D., Hassanien, S., & Adeeb, S. (2018b, September). Algorithms for the Strain Based Analysis of Dented Pipelines. In International Pipeline Conference (Vol. 51869, p. V001T03A067). American Society of Mechanical Engineers.

Okoloekwe, C., Fowler, M., Virk, A., Yoosef-Ghodsi, N., & Kainat, M. (2020, September). Reliability-based assessment of safe excavation pressure for dented pipelines. In *International Pipeline Conference* (Vol. 84447, p. V001T03A013). American Society of Mechanical Engineers.

Roux, S. G., Clausel, M., Vedel, B., Jaffard, S., & Abry, P. (2013). Self-similar anisotropic texture analysis: The hyperbolic wavelet transform contribution. *IEEE Transactions on Image Processing*, *22*(11), 4353-4363.

Rubinstein, R., Bruckstein, A. M., & Elad, M. (2010). Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, *98*(6), 1045-1057.

Sampath, S., Bhattacharya, B., Aryan, P., & Sohn, H. (2019). A real-time, non-contact method for in-line inspection of oil and gas pipelines using optical sensor array. *Sensors*, *19*(16), 3615.

Siqueira, M. H. S., Gatts, C. E. N., Da Silva, R. R., & Rebello, J. M. A. (2004). The use of ultrasonic guided waves and wavelets analysis in pipe inspection. *Ultrasonics*, *41*(10), 785-797.

Torrence, C., & Compo, G. P. (1998). *A practical guide to wavelet analysis*. Bulletin of the American Meteorological society, *79*(1), 61-78.

Wu, L., Liang, W., & Sha, D. (2022). A novel adaptive noise reduction method for field natural gas pipeline defect detection signals. *International Journal of Pressure Vessels and Piping*, *199*, 104761.

Zhang, S., Zhang, B., & Wang, R. (2022, September). A Transparent ASME B31. 8-Based Strain Assessment Method Using 3D Measurement of Dent Morphology. In *International Pipeline Conference* (Vol. 86571, p. V002T03A057). American Society of Mechanical Engineers.

Zhao, J., & Cheng, Y. F. (2022). A new criterion based on strain determination for dent assessment of pipelines. *Journal of Pipeline Science and Engineering*, *2*(1), 18-28.

Zhao, J., Lv, Y. R., & Cheng, Y. F. (2022). Standards and methods for dent assessment and failure prediction of pipelines: A critical review. *Petroleum Science*, *19*(6), 3029-3045.

Zhu, X. K., & Leis, B. N. (2005). Influence of yield-to-tensile strength ratio on failure assessment of corroded pipelines. *The Journal of Pressure Vessel Technology*, *127*(4), 436-44

Chapter 3

3 Discrete Wavelet Analysis of External Corrosion Clusters on Pipelines for Burst Capacity Prediction

3.1 Introduction

Pipelines are a safe and economical way to transport large quantities of oil and gas products over long distances. With the increased pipeline service life, pipelines may be threatened by different damage mechanisms. The data collected by the Pipeline and Hazardous Materials Safety Administration (PHMSA) of the United States indicate that external corrosion is a major damage mechanism, causing about 24% of the incidents on onshore gas transmission pipelines in the US between 2002 and 2013 (Lam & Zhou 2016). The report from the European Gas Pipeline Incident Data Group (EGIG 2020) has shown that 27% of the incidents on onshore gas transmission pipelines in Europe were caused by corrosion between 2010 and 2019. Corrosion reduces the pipe wall thickness and consequently the pressure containment capacity, i.e. burst capacity, of the pipeline. Therefore, corrosion assessment and mitigation are essential components of the pipeline integrity management program.

As will be explained in Section 3.2.1, a corrosion cluster can be considered to consist of multiple anomalies with irregular shapes (Vanaei et al. 2017). Not all the anomalies in a cluster affect the burst capacity of the corroded pipeline segment. When assessing a pipeline segment containing a corrosion cluster, ignoring some of the anomalies in the cluster may introduce negligible errors in the predicted burst capacity because the stress concentration caused by shallow anomalies is likely to be less severe than those caused by deep anomalies. However, simply ignoring corrosion anomalies with depths below a specific threshold depth may not be adequate for the burst capacity evaluation. For example, a shallow anomaly with large longitudinal length and circumferential width may have a non-negligible effect on the burst capacity. Furthermore, anomalies with the same size and shape but oriented in different directions may not have the same impact on the burst capacity. Therefore, it is valuable to analyze the directional features and length scales of the corrosion clusters to gain insights into their effects on the burst capacity of

corroded pipelines. The analysis of directional features and length scales can be achieved by using the wavelet transform to perform the so-called multi-resolution analysis (Daubechies 1992; Li et al. 2003; Mallat 2008; Walker 2008; Walnut 2002) of corrosion clusters.

The wavelet transform has been applied to signals from inline inspections of corroded pipelines. Siqueira et al. (2004) improved the signal-to-noise ratio of the pipe inspection signals by first thresholding the discrete wavelet transform (DWT) coefficients of the noisy signal and then using the thresholded coefficients to reconstruct the signal. Sampath et al. (2019) employed the wavelet packet and different wavelets to analyze the corrosion signal measured by non-contact optical sensor array for noise cancelation and feature extraction. Afzal and Udpa (2002) proposed an algorithm for detecting corrosion pits and mechanical damage in pipelines based on signals obtained from inspection tools using the magnetic flux leakage technique. This algorithm employs soft-thresholding to remove the additive white noise from the signal. Qi et al. (2010) used DWT to extract the peak shape and position of the signal reported by the ultrasonic pipeline inspection tool to accurately locate the corrosion defects on the pipe wall. Zajam et al. (2019) developed a pipeline health monitoring technique by measuring the vibration response of the pipeline and then using the wavelet transform of the measured response to locate the corrosion defects on the pipe wall. Tucker et al. (2003) and Kercel et al. (2003) decomposed the raw electromagnetic acoustic transducer data using the Daubechies wavelets and calculated the feature vector based on the Shannon entropy and energy of the wavelet coefficients for defect classification. Duong and Kim (2018) considered the entropy of the wavelet coefficients as the feature of an acoustic emission signal and employed the neural networks to classify surface defects on pipelines. However, the application of the wavelet transform to extract the information of directional features and length scales of the corrosion clusters in the context of their impact on the pipeline burst capacity has, to our best knowledge, not been reported in the literature.

Employing DWT with different thresholds for different sub-bands allows one to extract the directional features and length scales of corrosion clusters that have a large impact on the burst capacity prediction. In the DWT-based denoising analysis, the thresholds are

usually developed by minimizing the sum of the squares of the difference between the noise-free and denoised signals. Since the purpose of DWT-based denoising is to remove as much measurement error from the noisy signal as possible, the signal reconstructed based on the remaining wavelet coefficients after thresholding may still include many features that have little influence on the burst capacity prediction. In the present study, we propose a method to estimate the threshold values for different sub-bands of the 2D DWT of the natural corrosion clusters. The proposed method enables one to retain a small number of wavelet coefficients while ensuring that the difference in the predicted burst capacities of the original and reconstructed corrosion clusters is negligible. To this end, high-resolution laser scan data of natural corrosion clusters on the external surfaces of in-service buried pipe segments are collected and analyzed using the wavelet transform. Based on the preserved wavelet coefficients of the corrosion clusters, we gain further insights into the key directional features and length scales that govern the burst capacity prediction.

The rest of the paper is organized as follows. In Section 3.2, the corrosion cluster measurements obtained from the high-resolution laser scan are presented. The RSTRENG model for predicting the burst capacity of a pipe segment containing a corrosion cluster is also described in Section 3.2. Section 3.3 presents the fundamentals of the wavelet transform and selection of the wavelet function for the analysis in the present study. A methodology is then proposed in this section to estimate the thresholds for the wavelet coefficients in different sub-bands. In Section 3.4, we determine the optimal values of some empirical parameters in the proposed methodology from the wavelet analysis of a relatively large number of natural corrosion clusters. By analyzing the number, energy, and impact on the burst capacity prediction of the wavelet coefficients preserved by the thresholds estimated from the proposed methodology, we explore the key directional features and length scales of natural corrosion clusters that govern the burst capacity prediction in this section. Concluding remarks are presented in Section 3.5.

3.2 Corrosion clusters and burst capacity prediction

3.2.1 Scanned data of corrosion clusters

The corrosion measurement data employed in the present study are provided by a Canadian pipeline operator. Zhang et al. (2018) reported the burst tests of 14 full-scale corroded pipe specimens removed from in-service transmission pipelines. The outside diameters (*OD*) of these pipe specimens are between 407.4 and 764.1 mm, and wall thicknesses (*wt*) are between 5.9 and 8.5 mm. The dimensions and material properties of the pipe steel are summarized in Bao et al. (2018). These specimens have extensive, naturally-occurring corrosions on their external surfaces. After excavation, each pipe segment was scanned by a high-resolution laser-scanning device. The device measures corrosion depths on the pipe external surface at uniformly spaced points (grid points), with a grid spacing of 1 mm (arc length) along the pipe circumferential direction and either 1 mm or 2 mm along the longitudinal direction, i.e. a 1 × 1 mm or 1 × 2 mm grid. In the present study, 106 corrosion clusters are collected from the laser scan data. These corrosion clusters are randomly divided into two groups: 95 corrosion clusters are used to estimate the thresholds for different sub-bands, while the remaining 11 clusters are used to validate the adequacy of the thresholds. The collected corrosion clusters have at least 40 grid points in both the longitudinal and circumferential directions, and the maximum corrosion depths of different clusters are between 12.3% and 78.7%*wt*.

Let $z(x, y)$ denote the corrosion depth (in the unit of %*wt*) at the grid point (x, y) of a corrosion cluster, where x and y are the longitudinal and circumferential coordinates, respectively, and $d_{\max}^{(c)} = \max\{z\}$ denote the maximum corrosion depth of the corrosion cluster, respectively. Figure 3.1 shows the two- and three-dimensional (2D and 3D) plots of three collected corrosion clusters. There are some common features of the clusters: the majority of the measured corrosion depths within a given cluster are quite small, and grid points with large corrosion depths are non-uniformly distributed within the cluster. To have more insights into the composition of the corrosion cluster, we apply different threshold depths to the corrosion cluster shown in Fig. 3.1(a) such that any corrosion depths less than the threshold are considered corrosion-free, i.e. zero corrosion depth. Threshold depths of 5%*wt*, 10%*wt*, and 15%*wt* are applied. The corrosion cluster shown

in Fig. 3.1(a) after being subjected to such threshold depths is depicted in Fig. 3.2. It can be observed from this figure that the corrosion cluster now consists of a series of corrosion "islands" - referred to corrosion anomalies - with varying sizes. There are some drawbacks in applying threshold depths directly to identify anomaly features within a corrosion cluster. Figure 3.2 indicates that the number of corrosion anomalies and their sizes are sensitive to the threshold depth applied. A single threshold depth may not be adequate to identify all the anomalies that greatly influence the burst capacity of the pipeline at the corrosion cluster. On the other hand, applying different thresholds to the DWT coefficients associated with different sub-bands can more effectively extract the main directional features and length scales of the corrosion cluster and identify anomalies that significantly influence the burst capacity by retaining a small number of coefficients.
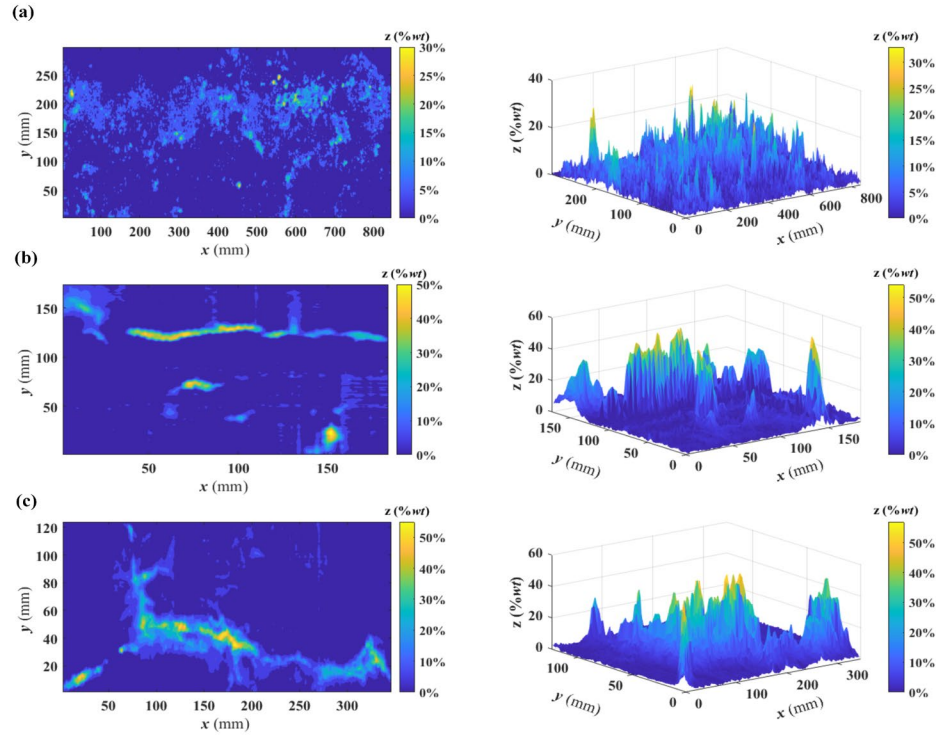
Figure 3.1 Corrosion depths measured by the laser scan on the external surfaces of in-service pipelines: (a) OD = 408.2 mm, wt = 6.2 mm, $d_{\text{max}}^{(c)}$ = 33.0%wt; (b) OD = 407.7 mm, wt = 5.9 mm, $d_{\text{max}}^{(c)}$ = 54.2%wt; and (c) OD = 407.4 mm, wt = 5.9 mm, $d_{\text{max}}^{(c)}$ = 56.9%wt
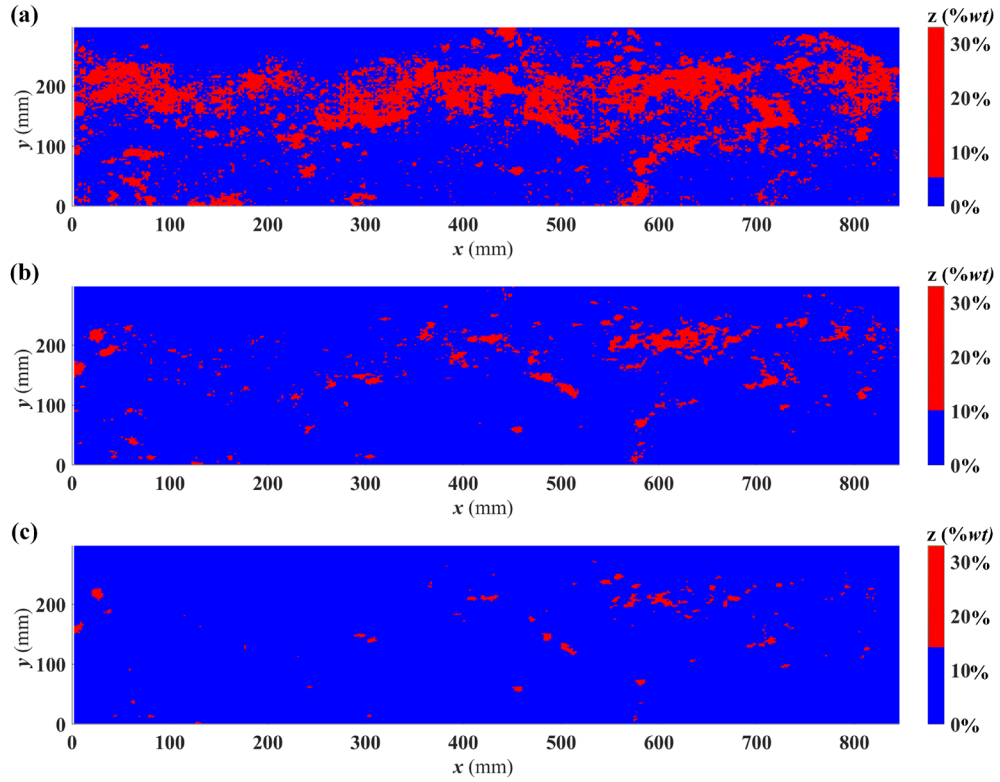
Figure 3.2 Corroded and corrosion-free areas in a given corrosion cluster by applying different threshold depths: (1) 5%*wt*; (2) 10%*wt*; and (3) 15%*wt*

3.2.2   RSTRENG model

The RSTRENG model (Kiefner and Vieth 1989) is a well-known semi-empirical model for predicting the burst capacities of corroded pipelines considering the corrosion morphology. The model employs the river-bottom profile of the corrosion cluster, which is determined by projecting the deepest points (i.e. the "river bottom") within the corrosion cluster onto a longitudinal plane that is perpendicular to the pipe wall.

The model evaluates the burst capacity *p* at a corrosion cluster as follows:

$$p_b = \frac{2 \cdot wt \cdot (\sigma_y + 69)}{OD} \frac{1 - \frac{A_e}{l_e \cdot wt}}{1 - \frac{A_e}{M_e \cdot l_e \cdot wt}} \tag{3.1}$$

$$M_e = \begin{cases} \sqrt{1 + 0.6275\frac{l_e^2}{OD \cdot wt} - 0.003375\left(\frac{l_e^2}{OD \cdot wt}\right)^2}, & \frac{l_e^2}{OD \cdot wt} \leq 50 \\ 3.3 + 0.032\frac{l_e^2}{OD \cdot wt}, & \frac{l_e^2}{OD \cdot wt} > 50 \end{cases} \tag{3.2}$$

where σ_y is the yield strength of the pipe steel in MPa; A_e and l_e are the effective area and length, respectively, of the river bottom profile of the corrosion cluster, and M_e is the so-called Folias factor corresponding to the effective length of the corrosion cluster. The determination of A_e and l_e involves an iterative process to identify the particular contiguous sub-portion of the river-bottom profile that leads to the lowest predicted burst capacity. Details of the RSTRENG model have been well described in many references (e.g. Kiefner and Vieth 1989; Zhang et al. 2018).

3.3   Discrete wavelet transform-based thresholding

3.3.1   Wavelet transform

Let $v(x)$ denote a one-dimensional (1D) signal. The coefficients of continuous wavelet transform (CWT) $v(x)$, $v_w(s, \tau)$, are given by (Daubechies 1992):

$$v_w(s, \tau) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} v(x)\psi^*\left(\frac{x-\tau}{s}\right) dx \tag{3.3}$$

where s is the scaling factor; τ is the translation; $\psi(x)$ is the so-called mother wavelet function, and * denotes the complex conjugate. If $\psi(x)$ satisfies the admissibility condition, i.e. $C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(f)|^2}{|f|} df < \infty$, where $\hat{\psi}(f)$ denotes the Fourier transform of $\psi(x)$ with f (Hz) being the frequency, the function $v(x)$ can be reconstructed from $v_w(s, \tau)$ by the inverse CWT:

$$v(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{v_w(s,\tau)}{\sqrt{|s|}} \psi\left(\frac{x-\tau}{s}\right) \frac{1}{s^2} d\tau ds \tag{3.4}$$

In CWT, s and τ are continuous, and both the forward and inverse transform need to calculate the integral. The high computational cost limits the value of CWT for engineering applications. The non-redundant DWT is more commonly used in practice

when processing a discrete signal. Let $v(x)$ ($x = 0, 1, 2, \ldots, 2^n - 1$) denotes a 1D discrete signal, where n is a positive integer. In 1D DWT, $v(x)$ can be decomposed in terms of wavelet and scaling functions as follows (Daubechies 1992):

$$v(x) = \sum_k 2^{-\frac{n}{2}} L_n(k)\varphi(2^{-n}x - k) + \sum_{i=1}^{n} \sum_k 2^{-\frac{i}{2}} H_i(k)\psi(2^{-i}x - k) \tag{3.5}$$

where $\varphi(x)$ denotes the scaling function (i.e. the low-pass filter); $\psi(x)$ is the mother wavelet function (i.e. the high-pass filter); k denotes translation; i is the decomposition level, and L and H are the approximation and detail coefficients, respectively. Usually $\varphi(x)$ and $\psi(x)$ are well designed so that $2^{-\frac{n}{2}}\varphi(2^{-n}x - k)$ and $2^{-\frac{i}{2}}\psi(2^{-i}x - k)$ in all terms are orthonormal bases with compact support and fast frequency decay. The approximation coefficients $L_n(k)$ and detail coefficients $H_i(k)$ are obtained by convolving $v(x)$ and the corresponding analysis filters.

In 1D DWT, the signal is decomposed to two sub-bands (high-frequency and low-frequency) at each level by the high-pass and low-pass filters, and the coefficients at higher levels are calculated by decomposing the approximation coefficients at lower levels. Given a 2D discrete signal $z(x, y)$ ($x = 0, 1, 2, \ldots, 2^n - 1$; $y = 0, 1, 2, \ldots, 2^m - 1$), the 2D DWT is similar to 1D DWT but with four sub-bands at each level. The wavelet coefficients in 2D DWT, $z_\psi^{(\alpha,i)}(\tau_x, \tau_y)$, is given by (Daubechies 1992):

$$z_\psi^{(\alpha,i)}(\tau_x, \tau_y) = \frac{1}{2^i} \sum_x \sum_y z(x, y)\psi_{(\alpha)}\left(\frac{x-\tau_x}{2^i}, \frac{y-\tau_y}{2^i}\right) \tag{3.6}$$

$$\psi_{(\alpha)}(x, y) = \begin{cases} \varphi(x)\varphi(y), & \alpha = LL \\ \varphi(x)\psi(y), & \alpha = LH \\ \psi(x)\varphi(y), & \alpha = HL \\ \psi(x)\psi(y), & \alpha = HH \end{cases} \tag{3.7}$$

where $\psi_{(\alpha)}(x, y)$ denotes the wavelet function corresponding to the α sub-band; *LL*, *LH*, *HL*, and *HH* denote the one approximation sub-band and three detail sub-bands in 2D DWT: i.e. approximation (low-pass and low-pass), horizontal detail (low-pass and high-pass), vertical detail (high-pass and low-pass) and diagonal detail (high-pass and high-pass), respectively. These sub-bands characterize the properties of the signal: *LL* can be

considered as the 2D signal in a poorer resolution; *LH*, *HL*, and *HH* can represent horizontal, vertical, and diagonal patterns of the 2D signal, respectively. Different sub-bands may have different influences on the burst capacity prediction. DWT can provide a good approximation with sparse wavelet coefficients, and the numerical complexity is only $O(N)$ (Mallat 2008). Therefore, it is widely used in data compression and feature detection. Details of the forward and inverse DWT are given in (Daubechies 1992; Mallat 2008; Torrence&Compo 1998).

DWT-based thresholding is achieved by setting the low-energy wavelet coefficients to zeros, where the energy associated with a coefficient is defined as the square of the absolute value of this coefficient. The commonly used wavelet thresholding includes hard- and soft-thresholding. In hard-thresholding, all the coefficients with an absolute value below the threshold are set to zeros. In soft-thresholding, also known as the wavelet shrinkage, coefficients with an absolute value below the threshold are set to zeros, while the other coefficients are adjusted toward zero by subtracting the threshold from their mode (Chang et al. 2000; Donoho 1995; Mallat 2008). Usually the soft-thresholding is employed to denoise a signal containing additive Gaussian white noises. Since the corrosion clusters in this study are measured by the laser scan device, which can be considered noise-free, hard-thresholding is employed in the present study. In the following sections, $z^{(r)}$ denotes the corrosion depth in the reconstructed corrosion cluster by applying the hard-thresholding to the wavelet coefficients. Since the corrosion depth must be non-negative, negative values of $z^{(r)}$ in the reconstructed cluster are set to zeros.

3.3.2　Wavelet functions for pipeline corrosion clusters

Many well-developed wavelet functions are available for DWT, such as the Daubechies wavelets (denoted as DBN, where N is the vanishing moment of the wavelet function), Daubechies symlets (denoted as Symlets N), coiflets, and biorthogonal wavelets (Mallat 2008). We select a wavelet function from the Daubechies wavelets and symlets for feature extraction because these wavelets have the minimum support given a vanishing moment. Let q_0 denote the number of grid points within a corrosion cluster, and q_w denote the number of non-zero wavelet coefficients after hard-thresholding. Define $E_0 = \Sigma_x \Sigma_y |z(x, y)|^2$ and $E_r = \Sigma_x \Sigma_y |z^{(r)}(x, y)|^2$ as the total energies of a corrosion cluster before and

after processing (i.e. thresholding), respectively. Let p_0 and p_r denote the burst capacities corresponding to the cluster before and after processing, respectively. Both p_0 and p_r are predicted by using the RSTRENG model. If the maximum decomposition level, the wavelet function, and the wavelet threshold η_w are specified, $z^{(r)}$ can be determined from z for a given cluster, and the value of q_w, E_r, and p_r can also be determined. The maximum decomposition level is constrained by the size of the signal. In the present study, the measured corrosion clusters are decomposed up to the 5th level, since some clusters are relatively small and therefore cannot be decomposed to higher levels. Bao and Zhou (2021) have shown that ignoring corrosion depths that are less than 10%wt has a negligible impact on the burst capacity evaluation. We therefore consider for now the corrosion depths deeper than $\eta_d = 10\%wt$ to be essential for the burst capacity prediction. Table 3.1 compares the mean values and the coefficient of variation (COV) of q_w/q_0, E_r/E_0, and p_r/p_0 under the maximum decomposition level of five and a decomposition level-independent threshold $\eta_w = \eta_d = 10\%wt$ for the 106 collected corrosion clusters described in Section 3.2.1.

Table 3.1 Statistics of q_w/q_0, E_r/E_0, and p_r/p_0 in the case of maximum decomposition level of five, and a constant threshold $\eta_w = 10\%wt$

Wavelet function	Mean of q_w/q_0 (%)	Mean of E_r/E_0 (%)	Mean of p_r/p_0 (%)	COV of p_r/p_0 (%)
DB1 (Haar)	1.80	89.11	101.19	0.78
DB2	1.63	89.78	100.38	0.64
DB6	1.68	89.90	101.01	0.71
DB10	1.69	89.62	101.14	0.78
DB20	1.76	88.94	101.34	0.89
Symlets 6	1.61	90.24	100.86	0.68
Symlets 10	1.61	90.29	100.93	0.70
Symlets 20	1.62	90.06	100.99	0.72

Table 3.1 indicates that the eight different wavelets considered all result in an excellent compression performance on the raw data. In addition to the values of the wavelet coefficients, one also needs to store the information about the wavelet filters corresponding to these coefficients. Therefore, the storage requirement of these wavelet coefficients will be $2q_w$. On average, about 90% of the total energy of the raw data is preserved by the wavelet coefficients that require only about 1/30 of the storage requirements of the raw data. Besides the low storage requirement, the RSTRENG-predicted burst capacities based on the corrosion cluster reconstructed from the preserved coefficients are very close to those based on the actual corrosion cluster. Since the mean value of p_r/p_0 corresponding to DB2 is the closest to unity of all the wavelets included in Table 3.1, DB2 appears to perform better in retaining the defect profile information in the context of the burst capacity prediction. Although Symlets 6, Symlets 10, and Symlets 20 are slightly better than DB2 in terms of the data compression (i.e. lower values of q_w/q_0 and higher values of E_r/E_0), the means of p_r/p_0 corresponding to these wavelets are not as close to unity as that corresponding to DB2. In general, identifying corrosion anomalies requires a high resolution in the spatial domain. However, symlets 6, symlets 10, and symlets 20 have higher vanishing moments than DB2, which means that they have longer filter lengths that may not be able to separate and capture closely spaced corrosion features. Therefore, DB2 is selected for the subsequent analyses.

3.3.3 Thresholds for the burst capacity prediction

The previous section considers a constant, sub-band-independent threshold to preserve the wavelet coefficients for reconstructing the corrosion cluster. However, such a constant threshold may not be optimal as it does not consider the different effects on the burst capacity due to wavelet coefficients at different decomposition levels and in different sub-bands. We therefore propose a threshold $\eta_{i,\alpha}$ for the ith level wavelet coefficients of the α sub-band:

$$\eta_{i,\alpha} = G_\alpha \cdot (G_1)^i \cdot (G_2 \eta_d) \tag{3.8}$$

where G_α is the sub-band parameter; G_1 is the level parameter, and $G_2\eta_d$ can be considered as the threshold depth such that corrosion depths below the threshold are considered to have a negligible impact on the burst capacity prediction. G_α, G_1, and G_2 are to be determined to optimize $\eta_{i,\alpha}$. If $G_\alpha = G_1 = 1$, $\eta_{i,\alpha}$ is a constant threshold that applies to all wavelet coefficients. The thresholds estimated by $G_\alpha = 1$ and $G_1 \neq 1$ are level-dependent, while they are sub-band-dependent if G_α depends on the sub-band of the wavelet coefficients. Different thresholds provide different characteristics of the natural corrosion clusters. The level-dependent thresholds account for the influence of the length scales of the features in the corrosion clusters. With the same corrosion depth, features of larger sizes have greater influence on the burst capacity than features of smaller sizes, and such an influence of the length scales is reflected by the level-dependent thresholds. In addition to the size, the impact of the directional features on the burst capacity is accounted for in the sub-band-dependent thresholds.

The sub-band parameter G_α is determined based on the impact of different sub-bands on the burst capacity prediction. Wavelet coefficients from different sub-bands can be interpreted as patterns with different orientations and sizes in the corrosion cluster. In the RSTRENG model, the burst capacity is governed by the effective area of the river-bottom profile. Since the river-bottom profile is the projection of the deepest point on a longitudinal plane, the influence of an elongated longitudinally orientated corrosion anomaly on the burst capacity is much stronger than other orientations under the same volume of metal loss and the maximum corrosion depth. Therefore, coefficients from the *LH* sub-band have a greater impact on the burst capacity than the other sub-bands.

The influence of a wavelet coefficient on the burst capacity prediction can be quantified by the increase in the area of the river-bottom profile caused by this coefficient. Consider that a 2D signal reconstructed by a wavelet coefficient with a value of unity. By projecting the maximum values of such a 2D reconstructed signal to the longitudinal plane, the area A_{inc} enclosed by the projection is the maximum area change contributed by this wavelet coefficient to the river-bottom profile area of the corrosion clusters. Figure 3.3 compares the 2D reconstructed signals based on the DB2 wavelet coefficients with the same values and decomposition levels but in different sub-bands, and the

corresponding projections on the longitudinal plane. The projected area depends on the selected wavelet function. The signal reconstructed by the *LH* coefficient has the greatest projected area, followed by *HH* and *LL*, while the *HL* coefficient leads to the smallest projected area. The ratios of the projected areas of the *LL*, *LH*, *HL*, *HH* coefficients are approximately 1.5:2:1:1.8, which suggests that the *LH* coefficient has twice the influence on the river-bottom profile as the *HL* coefficient with the same values. It should be noted that this is the ratio of the maximum area change wavelet coefficients in different sub-bands may contribute to the river bottom profile. The projected area ratio does not represent the actual change of the effective area of the river-bottom profile, as the effective area is determined by multiple wavelet coefficients. Nevertheless, the projected area ratio can be employed to quantify the influence of the wavelet coefficients and therefore determine the threshold ratio for different sub-bands. If the thresholds are dependent on the sub-band, the threshold ratios for *LL*, *LH*, *HL*, *HH* can be taken as 4/3:1:2:10/9, i.e. the inverse of the projected area ratios. Therefore, the values of G_α for *LL*, *LH*, *HL*, and *HH* are recommended to be 4/3, 1, 2, and 10/9, respectively.

(a)

(b)
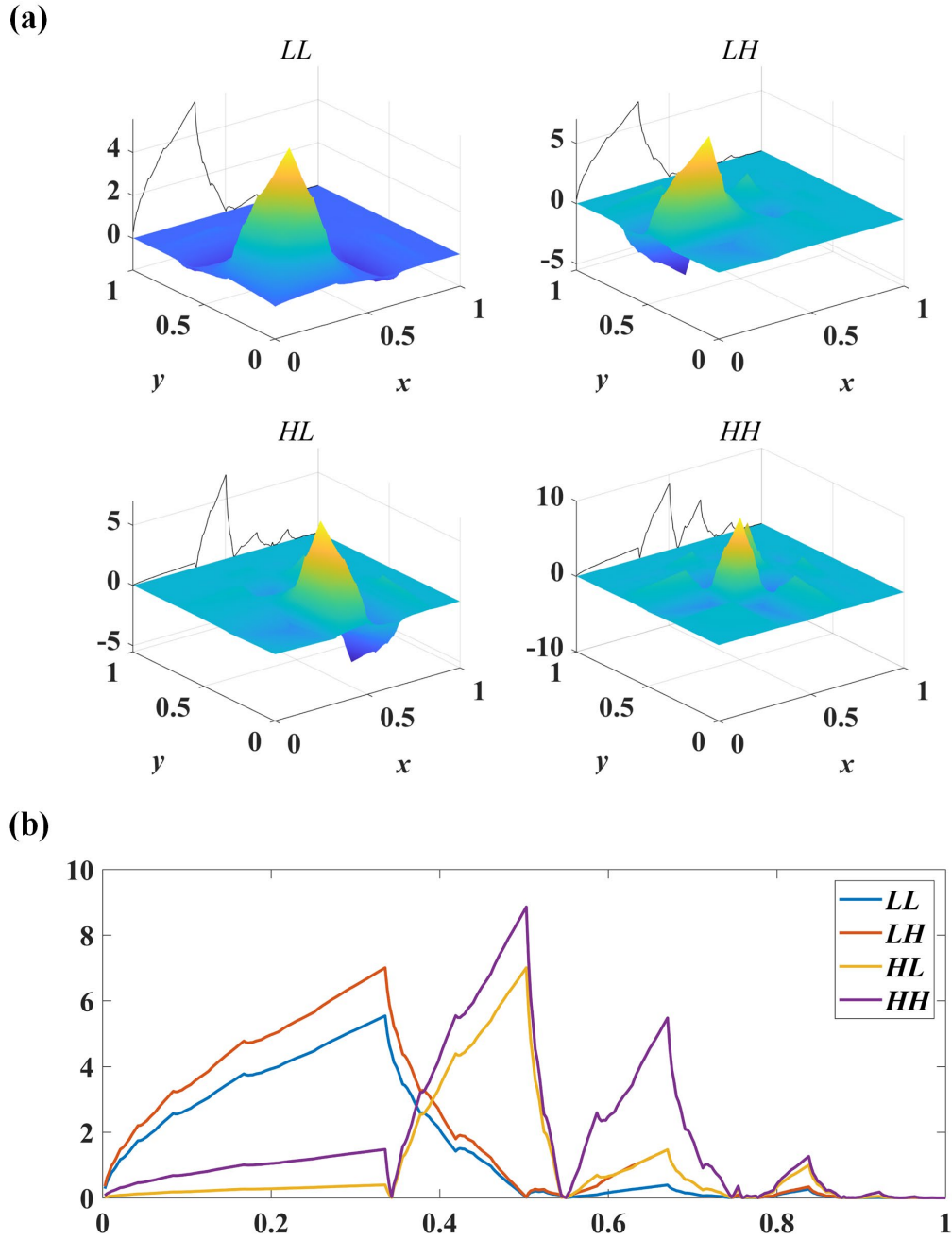
Figure 3.3 Signals reconstructed based on wavelet coefficients in different sub-bands: (a) 2D signals of *LL* (approximation), *LH* (longitudinal feature), *HL* (circumferential feature), and *HH* (diagonal feature); and (b) signals' projections on the longitudinal plane

Suppose that there are two features corresponding to two wavelet coefficients of the same value in the same detail sub-band (e.g. both belong to the *HL* sub-bands) but different

decomposition levels (the level difference is Δi). The feature corresponding to the high-level wavelet coefficient is flatter than the feature corresponding to the low-level wavelet coefficient: the longitudinal length and circumferential width of the former are $2^{\Delta i}$ times as long as those of the latter, but the corrosion depth of the former is only $1/2^{\Delta i}$ that of the latter. Since the decomposition level i is reflected in the form of an exponent on the sizes of the features, the decomposition level i is introduced in the form of an exponent in Eq. (3.8).

It is necessary to estimate the upper bound of the threshold for each sub-band to determine a reasonable range of values for G_1 and G_2. Suppose that corrosion depths deeper than η_d in a corrosion cluster have non-negligible influences on the burst capacity prediction. To ensure that all the grid points with depths greater than η_d are retained in the reconstructed corrosion cluster, the wavelet coefficients calculated from these grid points should be preserved. Therefore, the upper bound of the threshold can be defined as follows. If the amplitude of a wavelet coefficient is above the upper bound, then among all the grid points involved in the calculation of such a wavelet coefficient, there must be at least one grid point whose corrosion depth is greater than η_d. Let $\eta_{\alpha,i}^{(ub)}$ denote, respectively, the upper bounds of the normalized wavelet coefficients $z_\psi^{(\alpha,i)}(\tau_x, \tau_y)/\eta_d$ in the ith level and α sub-band. Since the corrosion clusters are non-negative, and the analysis filters used to calculate the wavelet coefficients for different sub-bands are known for a given wavelet function, one can compute the values of $\eta_{\alpha,i}^{(ub)}$ based on the taps of the analysis filters (see Appendix A). The numerical values of $\eta_{\alpha,i}^{(ub)}$ for the first five decomposition levels based on DB2 are summarized in Table 3.2. Note that directly employing $\eta_{\alpha,i}^{(ub)}\eta_d$ as the threshold for the ith level α sub-band will lose considerable important features. Based on the values of $\eta_{\alpha,i}^{(ub)}$ given in Table 3.2, the values of G_1 and G_2 are limited to $G_1 < 2$ and $G_2 < 1$, respectively.

Table 3.2 $\eta_{\alpha,i}^{(ub)}$ **for** $z_\psi^{(\alpha,i)}(\tau_x,\tau_y)/\eta_d$ **of different sub-bands based on DB2**

Decomposition level	Sub-bands α			
	LL	LH	HL	HH
1	2.34	1.4	1.4	1.4
2	4.78	2.72	2.72	2.66
3	9.58	5.39	5.39	5.21
4	19.15	10.39	10.39	9.68
5	38.32	20	20	17.92

3.4 Results and discussion

To determine the appropriate values of G_1 and G_2, parametric analyses are carried out whereby ranges of G_1 and G_2 values are substituted into Eq. (3.8) to determine $\eta_{i,\alpha}$, which is then employed as the threshold to retain the wavelet coefficients for reconstruction. For the constant threshold (i.e. both G_α and G_1 fixed at unity), the mean and COV of p_r/p_0 computed for the 95 corrosion clusters described in Section 3.2.1 corresponding to different values of G_2 are summarized in Fig. 3.4(a) and (b), respectively. Figure 3.4 indicates that the error in the burst capacity prediction of the reconstructed corrosion clusters increases linearly with the increase of G_2. A lower value of G_2 results in a higher number of wavelet coefficients being retained, the mean value of p_r/p_0 being closer to unity, and the COV of p_r/p_0 being closer to zero, i.e. more effective reconstruction of the corrosion cluster in terms of the burst capacity prediction. The mean values of q_w/q_0 and E_r/E_0 for the 95 clusters corresponding to different values of G_1 and G_2 are shown in Figs. 3.4(c) and 3.4(d). The value of q_w/q_0 rapidly decreases as G_2 increases from 0.5 to 1.0, and then decreases more slowly as G_2 increases beyond 1.0. This means that increasing the constant threshold is efficient up to a certain point in terms of removing wavelet coefficients. The optimal values of G_2 can be selected by limiting, somewhat arbitrarily,

the mean value of p_r/p_0 to be within 1% of unity, COV of p_r/p_0 to be less than 1%, and having the lowest mean of q_w/q_0. Based on these criteria, the optimal value of G_2 for the constant threshold is determined to be 1.8. In other words, if one aims to quickly identify the features in the corrosion clusters that influence the burst capacity prediction by a constant threshold, $\eta_{i,\alpha} = 18\%wt$ is a good choice.
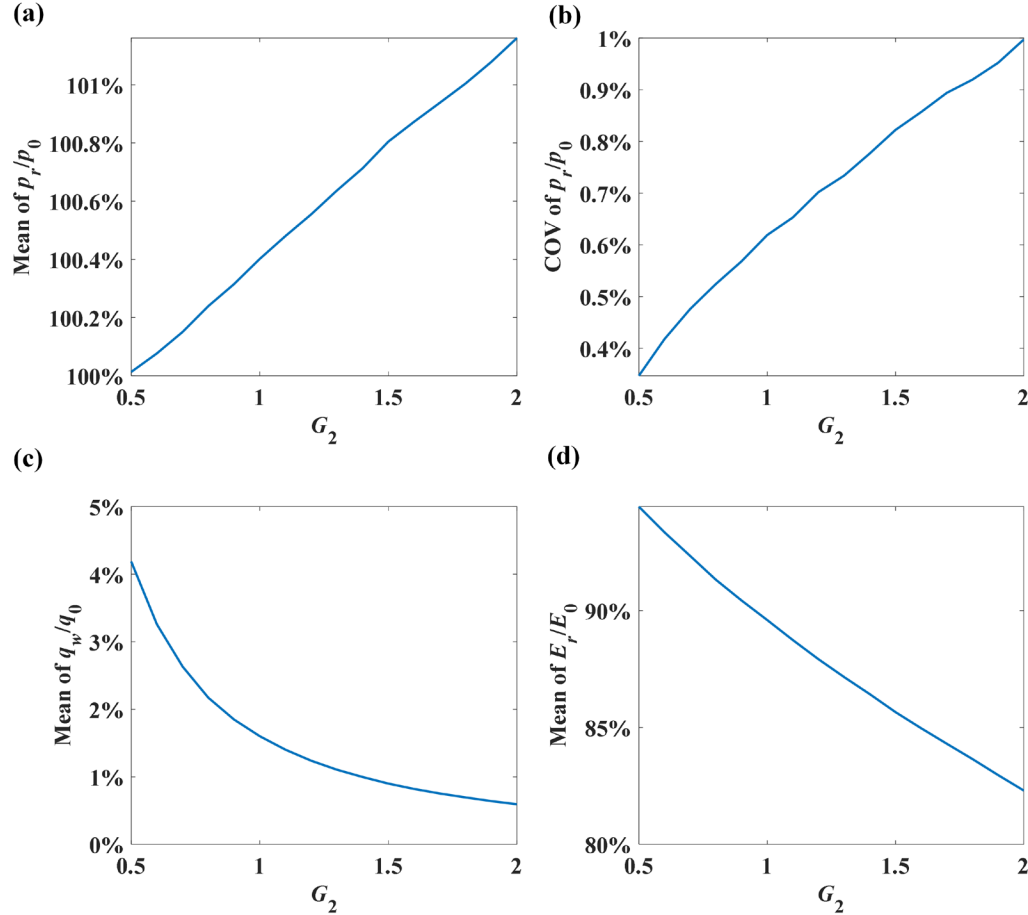
Figure 3.4 Statistics of applying a constant threshold on the corrosion clusters: (a) mean of p_r/p_0; (b) COV of p_r/p_0; (c) mean of q_w/q_0; and (d) mean of E_r/E_0

Figures 3.5 and 3.6 compare the statistics of p_r/p_0, q_w/q_0, and E_r/E_0 resulting from applying level-dependent (with G_α fixed at unity) and sub-band-dependent (G_α is assumed to be 4/3, 1, 2, and 10/9 for LL, LH, HL, and HH, respectively) thresholds on the

corrosion clusters, respectively. Since the thresholds are exponentially affected by G_1, these statistical indicators exponentially increase or decrease with G_1. As shown in Figs. 3.5 and 3.6, larger values of G_1 and G_2 result in fewer wavelet coefficients being retained for reconstruction, i.e. more efficient data compression. For the level-dependent and sub-band-dependent thresholds, the optimal values of G_1 and G_2 can be obtained following the same procedure as described above. For the level-dependent thresholds, the optimal values of G_1 and G_2 are selected to be 1.2 and 1.0, respectively, i.e. $\eta_{i,\alpha} = 1.2^i \cdot (10\%wt)$. For the sub-band-dependent thresholds (G_α is 4/3, 1, 2, and 10/9 for LL, LH, HL, and HH, respectively), the optimal values of G_1 and G_2 are selected to be 1.2 and 0.8, respectively, i.e. $\eta_{i,\alpha} = G_\alpha \cdot 1.2^i \cdot (8\%wt)$.
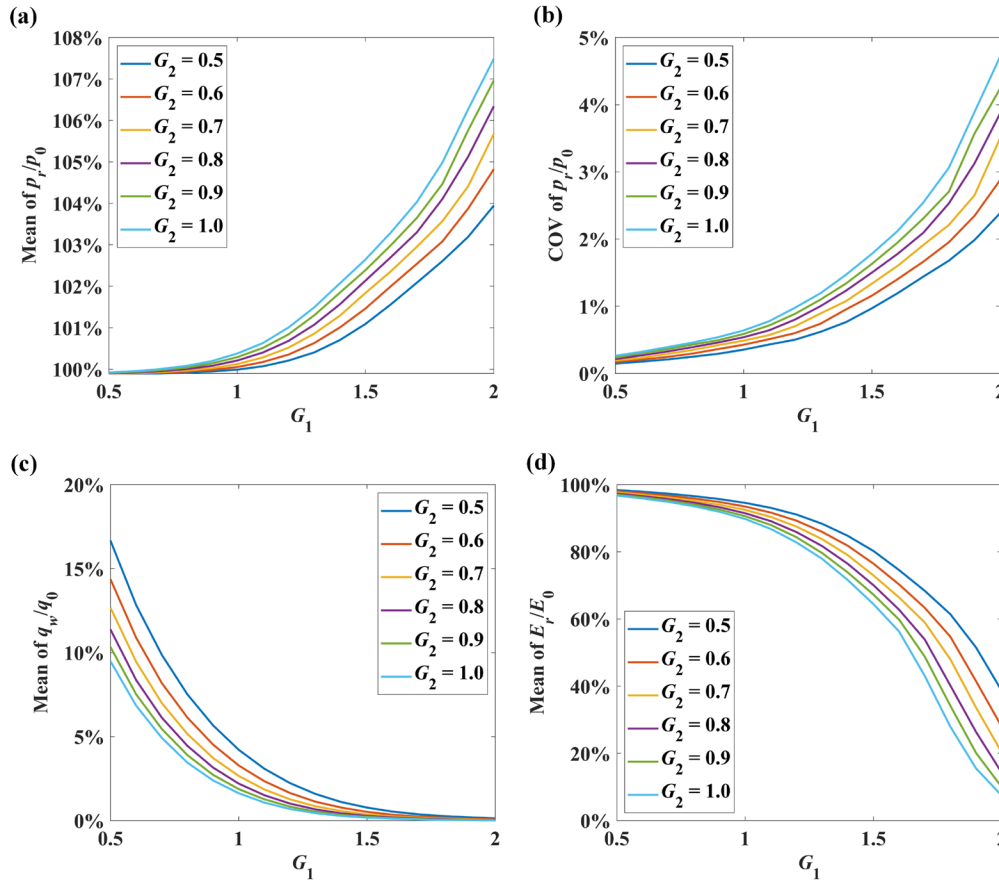
Figure 3.5 Statistics of applying the level-dependent thresholds on the corrosion clusters: (a) mean of p_r/p_0; (b) COV of p_r/p_0; (c) mean of q_w/q_0; and (d) mean of E_r/E_0
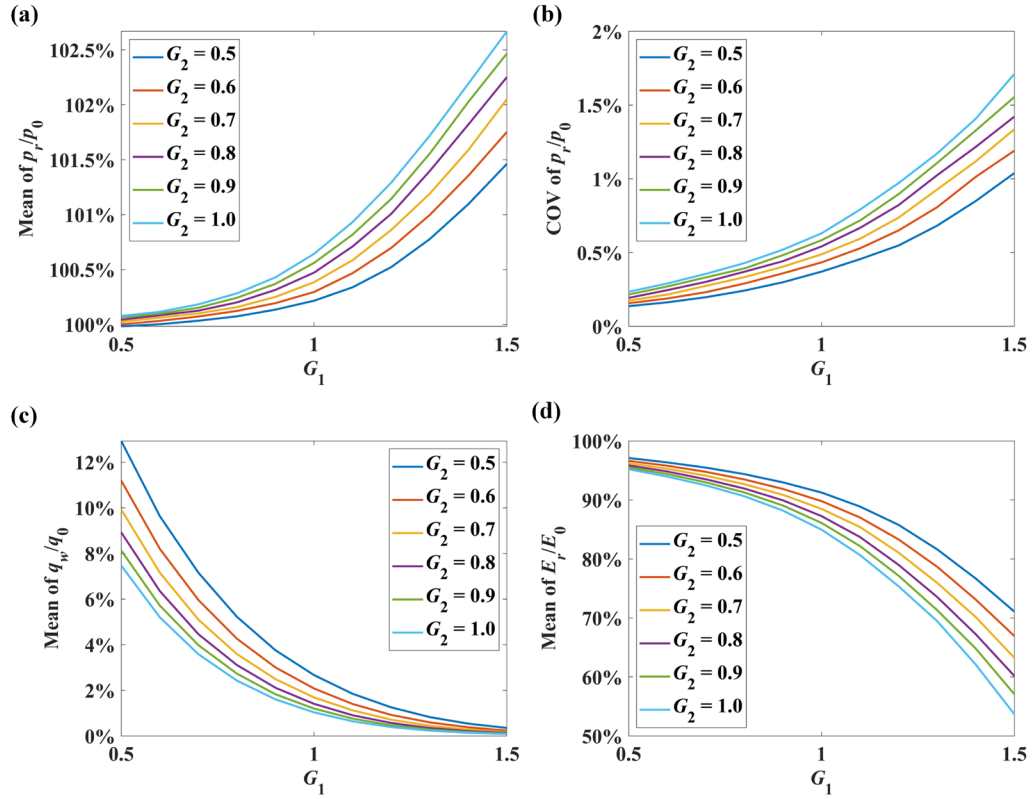
Figure 3.6 Statistics of applying the sub-band-dependent thresholds ($G_0 = 4/3$, 1, 2, and 10/9 for *LL*, *LH*, *HL*, and *HH*) on the corrosion clusters: (a) mean of p_r/p_0; (b) COV of p_r/p_0; (c) mean of q_w/q_0; and (d) mean of E_r/E_0

Table 3.3 summarizes the optimal expressions of the constant, level-dependent, and sub-band-dependent thresholds. The effectiveness of selected G_1 and G_2 in terms of the mean and COV of p_r/p_0 as well as the mean value of q_w/q_0 for the 95 corrosion clusters are also summarized in Table 3.3. Furthermore, the mean value and COV of p_r/p_0 as well as the mean value of q_w/q_0 for the 11 corrosion clusters excluded from the parametric analysis are summarized in Table 3.3 to further demonstrate the effectiveness of the selected values of G_1 and G_2.

Table 3.3 Optimal expressions and performance of the constant, level-dependent, and sub-band-dependent thresholds

Threshold	Optimal expression	95 clusters (parametric analysis)			11 clusters (validation)		
		mean of p_r/p_0	COV of p_r/p_0	mean of q_w/q_0	mean of p_r/p_0	COV of p_r/p_0	mean of q_w/q_0
Constant	$18\%wt$	101.0%	0.96%	0.70%	101.0%	0.92%	0.70%
Level-dependent	$1.2^i(10\%wt)$	101.0%	1.00%	0.70%	100.8%	0.92%	0.69%
Sub-band-dependent	$G_\alpha 1.2^i(8\%wt)$ $G_\alpha = 4/3, 1, 2,$ or $10/9$	101.0%	0.85%	0.58%	100.9%	0.79%	0.57%

From the optimal expressions of the three thresholds, one can gain further insights into some properties of the features that greatly affect the burst capacity prediction. First, the optimal value of G_1 in both level-dependent and sub-band-dependent thresholds is 1.2, which implies that the influence of the longitudinal length of a feature on the burst capacity prediction should not be ignored. In 2D DWT, if the length and width of a feature are doubled (i.e. the decomposition level of the corresponding wavelet coefficient increased by one) while the depth remain the same, the value of the wavelet coefficient corresponding to this feature becomes twice the original. Therefore, if the burst capacity prediction only depends on the depth of the features, the optimal value of G_1 should be two. However, the optimal value of 1.2 means that the length of the features is important to the burst capacity prediction. This also supports the hypothesis mentioned in Section 3.2.1, i.e. using a constant threshold depth to extract anomalies as the main features of the corrosion cluster may not be adequate. Second, the direction of the anomaly features has a non-negligible impact on the remaining burst capacity. The number of wavelet coefficients retained based on the sub-band-dependent thresholds is on average only 83% of those retained based on the constant ($18\%wt$) and level-dependent thresholds while still resulting in a lower COV of p_r/p_0. This suggests that the coefficients retained by the

sub-band-dependent thresholds can more effectively capture the part of the corrosion cluster that has a great impact on the burst capacity prediction.

Figure 3.7 compares the reconstructed corrosion clusters of the three clusters shown in Fig. 3.1 based on four different thresholds. Employing a constant threshold of 18%*wt* mainly retains some large size features of the natural corrosion clusters, while many minor anomalies are lost. By contrast, more small-sized anomalies are preserved in the reconstruction using level-dependent thresholds. The reconstructions based on the sub-band-dependent thresholds are the most visually different from the natural corrosion clusters among the four thresholds, as many circumferential patterns in the natural corrosion clusters are removed. However, the RSTRENG-predicted burst capacities of these reconstructions are almost identical to those of the corresponding natural corrosion clusters. It follows that although they may introduce a large visual difference, the discarded details have little effect on the burst capacity prediction.
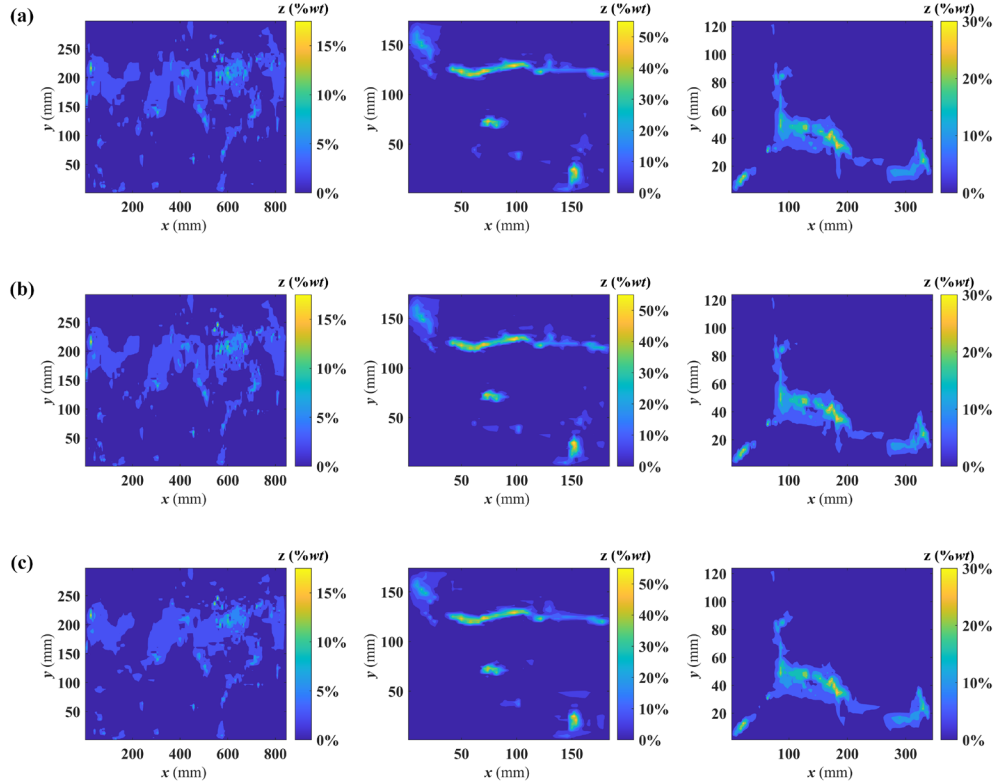
Figure 3.7 Reconstructed corrosion clusters based on different thresholds: (a) constant 18%*wt*; (b) level-dependent; and (c) sub-band-dependent

The numbers and energy proportions of wavelet coefficients in different sub-bands reveal the statistics of the directional features and length scales of the natural corrosion clusters. Figure 3.8 illustrates the numbers of preserved wavelet coefficients in each sub-band, which reflects the proportion of directional features of different length scales in the natural corrosion clusters that have a significant impact on the burst capacity prediction. Results from the constant (18%wt) and level-dependent thresholds show that about half of the preserved wavelet coefficients are HL coefficients. In comparison, the number of LH coefficients (corresponding to longitudinal features) is approximately half that of HL coefficients, and the proportion of HH coefficients is even lower. This suggests that the features in both longitudinal and diagonal directions are in general fewer than the circumferential features in natural corrosion clusters. The number of HL coefficients preserved by the sub-band-dependent thresholds is halved compared to those preserved by the constant and level-dependent thresholds. However, the accuracy of the predicted burst capacity is unaffected. This further suggests the low importance of the circumferential features for the burst capacity prediction. In terms of the numbers of preserved wavelet coefficients at decomposition levels, coefficients for the 3rd and 4th levels account for 50~60% of the preserved coefficients if level-dependent or sub-band-dependent thresholds are employed. This proportion suggests that in natural corrosion clusters, corrosion anomalies that have an impact on the burst capacity prediction are approximately 4 – 16 mm in width and 4 – 32 mm in length. Figure 3.9 compares the energy proportion of preserved coefficients in each sub-band. The energy of LL coefficients usually occupies the majority (68 - 76%) of the energy of preserved wavelet coefficients. This is expected, since the corrosion clusters are non-negative, and for each directional feature, at least one LL coefficient with energy no lower than this feature's energy is required to ensure that the reconstructed corrosion cluster is also non-negative. One may propose to preserve the LL coefficients at the 5th decomposition level only so that over 70% of the total energy of the corrosion cluster can be retained by about 0.1% of the total wavelet coefficients. However, if only the 5th level LL coefficients are employed to reconstruct the corrosion cluster, the mean of p_r/p_0 is 105.3%. This

indicates that the information included in the other sub-bands is essential for the burst capacity prediction.
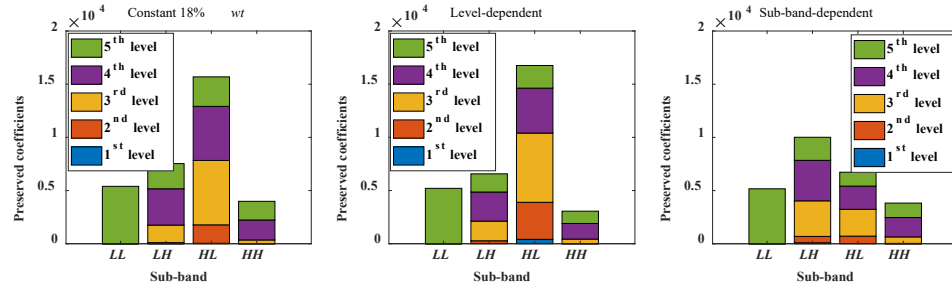
Figure 3.8 The number of preserved wavelet coefficients in each sub-band
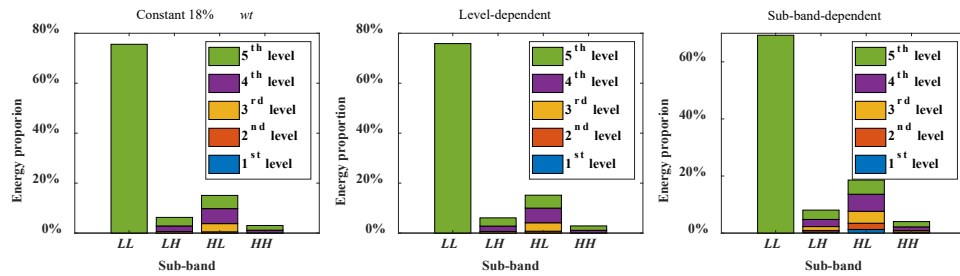
Figure 3.9 The energy proportion of preserved wavelet coefficients in each sub-band

After employing the constant, level-dependent, and sub-band-dependent thresholds, the influence of the sub-bands on the burst capacity prediction can be quantified by further removing the corresponding wavelet coefficients. Table 3.4 compares the errors in the predicted burst capacities of the reconstructions after removing all the wavelet coefficients from some detail sub-bands. It can be inferred from Table 3.4 that most of the coefficients preserved by the optimal sub-band-dependent thresholds are important to the burst capacity prediction: discarding the coefficients may result in significant errors. Furthermore, the error in the predicted burst capacity introduced by removing the coefficients in the 5th decomposition level is lower than that by removing the 4th decomposition level. This suggests that a higher level of decomposition of the corrosion

clusters may not be necessary as much of the information critical to the burst capacity prediction is included in the 4th and the lower decomposition levels. Since the longitudinal resolution of the corrosion clusters involved in the present study is either 1 mm or 2 mm, the 4th decomposition level represents a length scale between 16 and 32 mm. This suggests that it is unnecessary to decompose the natural corrosion cluster to a scale greater than 32 mm. Furthermore, longitudinal and circumferential features have greater influence on the burst capacity prediction than diagonal features. The results in Table 4 suggest that accurately identifying longitudinally orientated corrosion anomalies with lengths between 4 and 16 mm is crucial for accurately predicting the remaining burst capacity of a corroded pipeline segment.

Table 3.4 Mean of ($p_r/p_0 - 100\%$) after removing some sub-bands

Threshold	Optimal expression	Remove decomposition level (%)					Remove detail sub-bands (%)		
		1st	2nd	3rd	4th	5th	LH	HL	HH
Constant	18%wt	1.60	1.91	2.80	3.20	3.08	3.49	3.53	2.42
Level-dependent	$1.2^i(10\%wt)$	1.67	2.00	2.78	3.07	3.00	3.41	3.63	2.34
Sub-band-dependent	$G_\alpha 1.2^i(8\%wt)$, $G_\alpha = 4/3, 1, 2,$ or $10/9$	1.60	2.22	3.05	3.38	3.37	3.96	3.39	2.76

The above discussion explores the directional features and length scales essential for the burst capacity prediction. The corrosion clusters employed in the present study is measured by a high-resolution laser-scanning device; such data may not be widely available in practice. In the pipeline industry, the commonly used in-line inspection (ILI) for the in-service pipeline corrosion detection may have a comparable scanning

resolution along the longitudinal direction (e.g. ILI tools employing the ultrasonic technology), but the scanning resolution along the circumferential direction is limited by the number of sensors in the inspection tool. The circumferential scanning spacing may be dozens of times longer than the longitudinal scanning spacing, resulting in measurements with more longitudinal patterns rather than circumferential patterns. This is one of the limitations of the present study as the proposed method to estimate the threshold cannot be directly applied to the ILI data. However, this limitation can be simply addressed by adjusting G_α and G_2 based on the resolutions of the ILI data in both directions. Some ILI tools, e.g. those employing the magnetic flux leakage (MFL) technique, indirectly measure the corrosion depth. In this case, further studies are needed to investigate if the observations and insights obtained in the present study are also applicable to the MFL ILI data. The empirical parameters G_1 and G_2 in Eq. (3.8) are determined by minimizing the number of the preserved wavelet coefficients under the constraint of limiting the expected error of the burst capacity prediction. The thresholds estimated by these empirical parameters can preserve most of the wavelet coefficients that have an impact on the burst capacity prediction, but some minor corrosion anomalies are excluded in the reconstructions. Although the influence of these small anomalies on the burst capacity is negligible, ignoring them may lead to errors in estimating the corrosion growth rate. Finally, the measurement errors associated with the ILI data should also be considered if the wavelet analysis is applied to the ILI data.

3.5 Conclusion

High-resolution laser scan data for 106 natural corrosion clusters identified on the external surfaces of buried in-service natural gas transmission pipelines in Canada are collected. The 2D DWT using DB2 is employed to decompose the collected natural corrosion clusters and reconstruct the clusters after removing some of the wavelet coefficients. The RSTRENG model, which is widely used in the pipeline industry, is employed to evaluate the burst capacities of the pipe segment containing the natural and reconstructed corrosion clusters. A thresholding methodology is proposed to retain a small number of wavelet coefficients essential to the burst capacity prediction. The

proposed method allows one to predict the burst capacity of the corrosion cluster with an error of no more than 1% while retaining less than 1% of the wavelet coefficients.

The statistics of the wavelet coefficients preserved by the thresholds estimated from the proposed methodology provide insights into the properties of the natural corrosion cluster in the context of the burst capacity prediction. Features with different directions and length scales have different influences on the burst capacity prediction. In the natural corrosion clusters, longitudinal features are observed to have lower energy than circumferential and diagonal features, but have a larger impact on the burst capacity prediction. The features having the largest impact on the burst capacity are approximately 4 – 16 mm in size. Decomposing the corrosion cluster by more than five levels is unnecessary as the wavelet coefficients at the 5th and lower levels contribute most to the burst capacity prediction. Further studies are recommended to investigate the potential application of the proposed methodology to ILI data.

References

Afzal, M., & Udpa, S. (2002). Advanced signal processing of magnetic flux leakage data obtained from seamless gas pipeline. *NDT & E International*, *35*(7), 449-457.

Bao, J., Zhang, S., Zhou, W., & Zhang, S. (2018, September). Evaluation of burst pressure of corroded pipe segments using three-dimensional finite element analyses. In *International pipeline conference* (Vol. 51869, p. V001T03A043). American Society of Mechanical Engineers.

Bao, J., & Zhou, W. (2021). Influence of depth thresholds and interaction rules on the burst capacity evaluation of naturally corroded pipelines. *Journal of Pipeline Science and Engineering*, *1*(1), 148-165.

Chang, S. G., Yu, B., & Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. *IEEE transactions on image processing*, *9*(9), 1532-1546.

CSA. (2019). Oil and gas pipeline systems. *CSA Standard Z662-19*. Canadian Standard Association. Mississauga, Ontario, Canada.

Daubechies, I. (1992). *Ten lectures on wavelets*. Society for industrial and applied mathematics.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE transactions on information theory*, *41*(3), 613-627.

Duong, B. P., & Kim, J. M. (2018). Pipeline fault diagnosis using wavelet entropy and ensemble deep neural technique. In *Image and Signal Processing: 8th International Conference*, ICISP 2018, Cherbourg, France, July 2-4, 2018, Proceedings 8 (pp. 292-300). Springer International Publishing.

EGIG. 11th Report of the European Gas pipeline Incident data Group, 2020. https://www.egig.eu/reports

Kercel, S. W., Tucker Jr, R. W., & Varma, V. K. (2003, August). Pipeline flaw detection with wavelet packets and GAs. In *Intelligent Computing: Theory and Applications* (Vol. 5103, pp. 217-226). SPIE.

Kiefner, J. F., & Vieth, P. H. (1989). *A modified criterion for evaluating the remaining strength of corroded pipe* (No. PR-3-805). Battelle Columbus Div., OH (USA).

Lam, C., & Zhou, W. (2016). Statistical analyses of incidents on onshore gas transmission pipelines based on PHMSA database. *International Journal of Pressure Vessels and Piping*, *145*, 29-40.

Li, J. P., Tang, Y. Y., Wickerhauser, V., Daugman, J., Peng, L., & Zhao, J. (Eds.). (2003). *Wavelet Analysis And Its Applications (In 2 Vols), Proceedings Of The Third International Conference On Waa*. World Scientific.

Mallat, S. (2008). *A Wavelet Tour of Signal Processing: The Sparse Way* 3rd ed (New York: Academic).

Qi, S., Liu, J., & Jia, G. (2010, October). Study of submarine pipeline corrosion based on ultrasonic detection and wavelet analysis. In *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)* (Vol. 12, pp. V12-440). IEEE.

Sampath, S., Bhattacharya, B., Aryan, P., & Sohn, H. (2019). A real-time, non-contact method for in-line inspection of oil and gas pipelines using optical sensor array. *Sensors*, *19*(16), 3615.

Siqueira, M. H. S., Gatts, C. E. N., Da Silva, R. R., & Rebello, J. M. A. (2004). The use of ultrasonic guided waves and wavelets analysis in pipe inspection. *Ultrasonics*, *41*(10), 785-797.

Torrence, C., & Compo, G. P. (1998). *A practical guide to wavelet analysis*. Bulletin of the American Meteorological society, *79*(1), 61-78.

Tucker Jr, R. W., Kercel, S. W., & Varma, V. K. (2003, May). Characterization of gas pipeline flaws using wavelet analysis. In *Sixth International Conference on Quality Control by Artificial Vision* (Vol. 5132, pp. 485-493). SPIE.

Vanaei, H. R., Eslami, A., & Egbewande, A. (2017). A review on pipeline corrosion, in-line inspection (ILI), and corrosion growth rate models. *International Journal of Pressure Vessels and Piping*, *149*, 43-54.

Walker, J. S. (2008). *A primer on wavelets and their scientific applications*. CRC press.

Walnut, D. F. (2002). *An introduction to wavelet analysis*. Springer Science & Business Media.

Zajam, S., Joshi, T., & Bhattacharya, B. (2019). Application of wavelet analysis and machine learning on vibration data from gas pipelines for structural health monitoring. *Procedia Structural Integrity*, *14*, 712-719.

Zhang, S., Yan, J., Kariyawasam, S., Huang, T., & Al-Amin, M. (2018, September). A more accurate and precise method for large metal loss corrosion assessment. In

International Pipeline Conference (Vol. 51869, p. V001T03A059). American Society of Mechanical Engineers.

Chapter 4

4 Simulating Nonhomogeneous Non-Gaussian Corrosion Fields on Pipelines Based on In-line Inspection Data

4.1 Introduction

Metal-loss corrosion is one of the main threats to the structural integrity of buried oil and gas pipelines. According to the report from the European Gas Pipeline Incident Data Group (EGIG 2020), 27% of the incidents on onshore gas transmission pipelines in Europe were caused by corrosion between 2010 and 2019. The data collected by the Pipeline and Hazardous Materials Safety Administration (PHMSA) of the United States indicate that about 24% of the incidents on onshore gas transmission pipelines in the US were caused by external corrosion between 2002 and 2013 (Lam and Zhou 2016). Therefore, corrosion assessment and mitigation are essential components of the pipeline integrity management program.

Corrosion is a stochastic process (Zhang and Zhou, 2014; Zhou et al., 2017; Zhou et al., 2021), and a corrosion defect on a pipeline is a field of random metal loss. The spatial modeling of corrosion on steel pipelines or steel plates on ships has been reported in the literature. Garbatov and Soares (2019) proposed two models to simulate the corroded surface of steel bottom plates of ballast tanks in double-hull tankers. One model considers that corrosion pits with uncorrelated corrosion depth are uniformly distributed on the steel plate. The other employs randomly-located elliptic paraboloids with uncorrelated shape coefficients to characterize the corroded surface. Aryai and Mahmoodian (2017) consider the corrosion depths on cast iron water pipes as homogeneous Gaussian random fields. Bao and Zhou (2021) proposed a model that captures the spatial intermittency between corroded and corrosion-free regions to characterize and simulate the nonhomogeneous non-Gaussian corrosion field on the external surface of buried pipelines based on high-resolution data obtained from the laser scan device. Hong et al. (2021a; 2021b) proposed an iterative power and amplitude correction (IPAC) algorithm to simulate nonstationary and non-Gaussian processes. This algorithm was employed by Zhou et al. (2021) to simulate the corroded external surfaces

of buried steel pipelines based on the discrete orthogonal S-transform (Stockwell et al. 1996). However, it may be difficult in practice to collect from in-service pipelines the high-resolution laser scan data needed in the models proposed by Bao and Zhou (2020) and Zhou et al. (2021).

In the pipeline industry, a widely used method to collect the corrosion data of an in-service pipeline is the in-line inspection (ILI). ILI is a non-destructive evaluation technique that uses a pipeline inspection gauge (PIG) to detect and size corrosion defects on in-service pipelines. The commonly used technologies for ILI include magnetic flux leakage, ultrasonic, and electromagnetic acoustic transducers (Xie and Tian, 2018; Vanaei et al., 2017). Isolated individual corrosion anomalies reported by ILI are characterized by three sizing parameters and two location parameters. Anomalies shallower than a detection threshold, typically 5–10% of the pipe wall thickness (Siraj and Zhou 2019), are not reported. ILI has the capability of detecting relatively large anomalies and estimating the maximum depth (d_{max}, in the through-wall thickness direction) length (l, in the pipe longitudinal direction) and width (w, in the pipe circumferential direction) of each anomaly. These data greatly facilitate various tasks involved in the pipeline corrosion management practice such as the evaluation of the remaining burst capacities of a pipeline at given corrosion anomalies and estimating the corrosion growth over time (Al-Amin and Zhou, 2014; Zelmati et al., 2022). The majority of engineering critical assessment models used in practice to evaluate the remaining burst capacity of corroded pipelines require the length and maximum depth of the corrosion anomaly as part of the model input (Zhou and Huang, 2012); such information of the corrosion anomaly can be readily obtained from the ILI data. However, the model error associated with such burst capacity models can be high because they do not take into account detailed three-dimensional profiles of the corrosion anomaly (Zhou and Huang, 2012). While assessment models that incorporate detailed corrosion profiles have been shown to have a higher predictive accuracy (Zhou and Huang, 2012; Zhang et al., 2018), it is very difficult, if not impossible, to obtain the detailed corrosion profile from the ILI data directly. This suggests that a methodology to generate corrosion profiles based on the ILI data will be highly beneficial to the corrosion

assessment practice. However, to our best knowledge, such a methodology has not been reported in the literature.

In the present study, we propose a methodology to simulate corrosion fields on the external surface of buried pipelines based on the corresponding ILI data. The corrosion field is assumed to consist of multiple homogeneous non-Gaussian corrosion anomalies whose marginal distributions and spatial correlations are estimated from the ILI data. We apply a non-Gaussian simulation algorithm proposed by Masters and Gurley (2003) in conjunction with the ILI data to generate multiple realizations of corrosion anomalies and then merge them into a synthetic field after modulation. The local extreme values of the corrosion depth in the synthetic field match the ILI data. The advantages of the proposed methodology is demonstrated by comparing probabilistic characteristics and burst capacities of the synthetic corrosion fields with those of the actual corrosion fields and fields idealized based on the ILI data. It is worth noting that Brennan et al. (2002) employed a similar approach to study corrosion fields on pipelines by modelling the corrosion flaw (i.e. cluster) to be a non-homogeneous random field that consists of a series of homogenous Gaussian random fields. The mean, variance and correlation of each homogeneous Gaussian field are then estimated based on the corrosion inspection data.

In Section 4.2, the corrosion fields measurement obtained from the high-resolution laser scan are presented. The ILI data, obtained from a virtual ILI process, corresponding to these corrosion fields are also presented in this section to demonstrate the disadvantages of estimating the burst capacity based directly on the ILI data. In Section 4.3, we first describe the proposed methodology to simulate the nonhomogeneous non-Gaussian corrosion fields based on the ILI data. We then introduce the procedures to develop the marginal distribution, autocorrelation function (ACF), and spatial modulator corresponding to corrosion anomalies based on the ILI data. Section 4.4 presents synthetic corrosion fields generated from the proposed methodology, and compares the burst capacities of natural corrosion fields, ILI idealizations, and synthetic fields. Concluding remarks are given in Section 4.5.

4.2   Practical background

A corrosion defect on a buried pipeline is a random field of metal loss. In the following sections, the term "corrosion defect" refers to a corrosion field consisting of two or more anomalies, where the concept of corrosion anomaly is introduced in conjunction with the ILI data. Figure 1 shows the three-and two-dimensional (3D and 2D) plots of a corrosion field (defect) on the external surface of a buried in-service natural gas pipeline in Canada with an outside diameter (D) of 407.7 mm and a wall thickness (wt) of 6.2 mm. These plots are generated from data obtained from a laser scanning device that measures corrosion depths at uniformly spaced points (grid points) on the pipe surface, with a circumferential (arc length) spacing of 1 mm and a longitudinal spacing of 2 mm, i.e. 1 × 2 mm. Given the laser scan data, the so-called river-bottom profile (Kiefner & Vieth, 1989) of the defect can be obtained, as depicted in Fig. 4.1(a) and 4.1(c). The river-bottom profile is commonly employed in the practical fitness-for-service (FFS) assessment of corrosion defects and generated by projecting the deepest points (i.e. the "river bottom") within a corrosion defect onto a longitudinal plane that is perpendicular to the pipe wall, thus creating a two-dimensional profile of the three-dimensional corrosion field (Kiefner and Vieth, 1989). Given the river-bottom profile, the RSTRENG model (Kiefner and Vieth, 1989), considered one of the most accurate FFS assessment models for corroded pipelines, is typically employed to estimate the burst capacity of the pipeline at the defect. However, laser scan data for corrosion defects are not commonly available in practice because one can only perform the laser scan after a corroded pipe segment has been excavated for repair or replacement. The FFS assessment is therefore most often performed based on the ILI data.
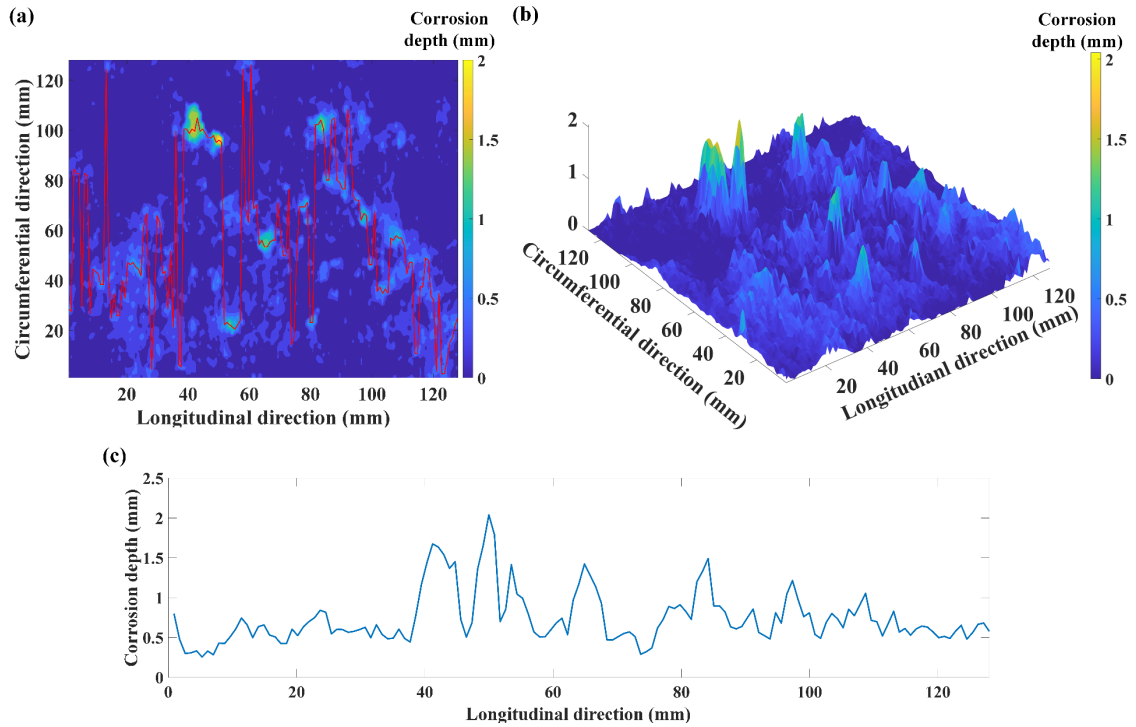
Figure 4.1 Corrosion depths measured by the laser scan on the external surface of an in-service pipeline with *wt* = 6.2 mm: (a) two-dimensional plot with highlighted river-bottom profile, (b) three-dimensional plot, and (c) river-bottom profile

Consider again the corrosion defect shown in Fig. 4.1. The pipe segment containing the defect has been subjected to an ILI employing the magnetic flux leakage technology; however, the corresponding ILI data were not provided to the present study. To illustrate the FFS assessment based on the ILI data, a virtual ILI is performed in this study to derive the ILI data from the laser scan data. The virtual ILI process consists of the following steps.

1) A threshold depth is applied to the laser scan data such that any grid points with the corrosion depths less than the threshold are considered corrosion-free, i.e. the corresponding corrosion depths set to zero (Fig. 4.2(a)). The grid points (highlighted in red) with corrosion depths greater than the threshold are identified as corrosion points. The threshold is equivalent to the detection limit of the ILI tool (Siraj and

Zhou, 2019; Smart et al., 2018). In this study, the threshold is set at 5%*wt*, consistent with typical practice.

2) By applying the threshold depth (detection limit), the corrosion field is now divided into multiple "islands" of corrosion points, as depicted in Fig. 4.2(b); each island is referred to as an individual corrosion anomaly. Two corrosion points are considered part of the same anomaly if they are longitudinally or circumferentially connected (i.e. two points are adjacent to each other longitudinally or circumferentially).

3) Let *x* and *y* denote the longitudinal and circumferential coordinate, respectively, of a given point. Each anomaly is characterized in ILI by two location parameters x_c and y_c, i.e. longitudinal and circumferential coordinates of the centre of the anomaly, and three sizing parameters, i.e. the length *l*, width *w*, and maximum depth d_{max}, as depicted in Fig. 4.2(b).
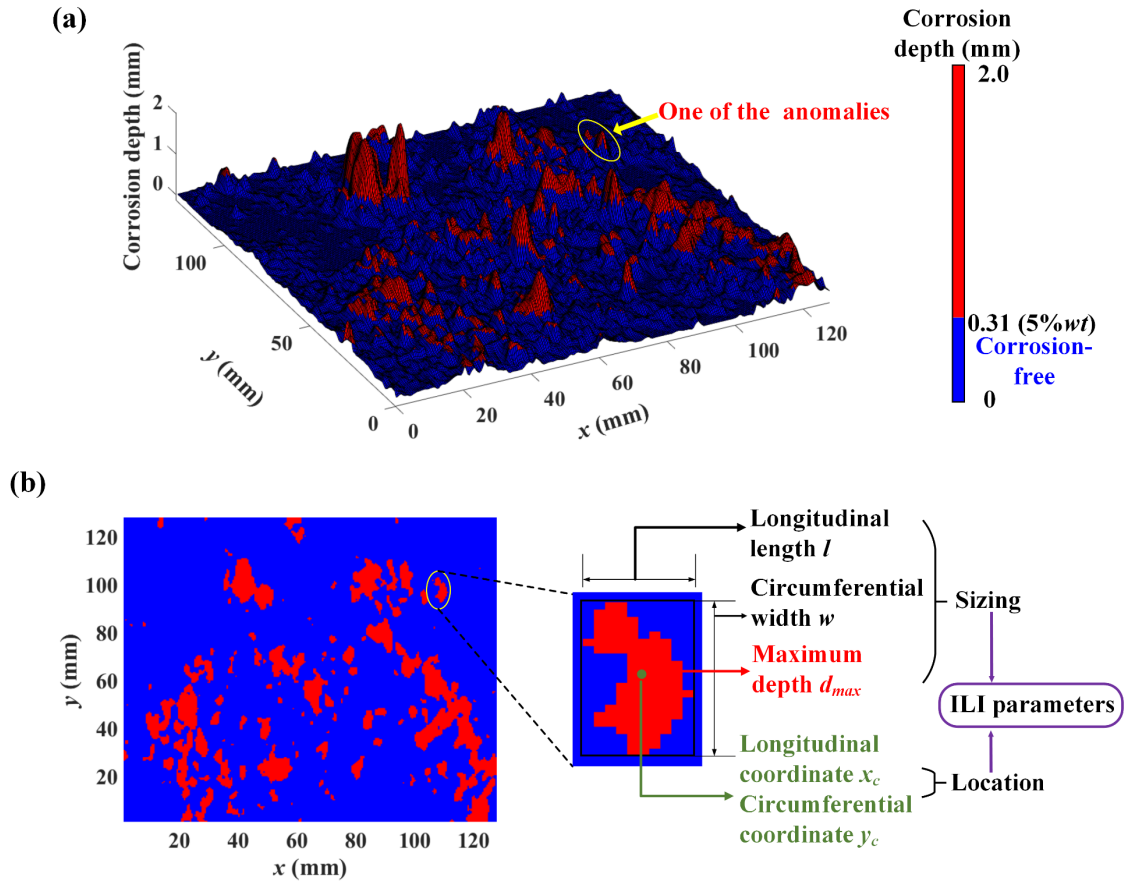
Figure 4.2 A corrosion field on the external surface of an in-service pipeline with *wt* = 6.2 mm: (a) three-dimensional plot; and (b) two-dimensional plot and the definition of ILI parameters for an anomaly

Since ILI does not provide detailed 3D profiles of individual anomalies, each anomaly is typically idealized as a cuboid in ILI. It follows that a corrosion defect consists of one or more cuboids in ILI. The corrosion defect shown in Fig. 42(a) consists of 115 detected anomalies after the virtual ILI process. In this study, an anomaly is ignored if either its length or width is less than five grid points. Neglecting small anomalies with the length or width less than five grid points slightly reduces the estimated volume of metal loss, but the influence is negligible as such anomalies are in general shallow with d_{max} marginally greater than 5%*wt*. Bao and Zhou (2021b) have shown that ignoring corrosion depths that are less than 10%*wt* has a negligible impact on the burst capacity evaluation. For the

105 corrosion clusters included in the present study, the RSTRENG-predicted burst capacities are on average 91.4% of the actual burst capacities if all the anomalies are preserved and idealized as cuboids. This ratio increases only slightly to 91.5% if the anomalies with the length or width less than five grid points are neglected.

After removing small anomalies, only 24 anomalies are preserved. These preserved anomalies and the corresponding river-bottom profile based on the cuboid idealization are illustrated in Fig. 4.3. The cuboid idealization of individual corrosion anomalies leads to the conservative characterization of the corrosion defect: the total volumes of metal loss of the defect are estimated to be 3295.0 and 6380.2 mm^3, respectively, based on the laser scan data and corresponding (virtual) ILI idealizations, respectively. The ILI-based river-bottom profile of the defect is depicted in Fig. 4.3(c). For comparison, the river-bottom profile of the defect based on the laser scan data is also shown in the figure. The area enclosed by the ILI-based profile is 44.20% greater than that enclosed by laser scan-based profile. If the RSTRENG model is applied to evaluate the burst capacity of the pipeline at the corrosion defect (assuming the yield strength (σ_y) of the pipe steel to equal 369 MPa), the evaluated burst capacities equal 11.25 and 12.03 MPa corresponding to the ILI-based and actual river-bottom profiles, respectively: a difference of 6.5%. The above illustrates the deficiency of carrying out FFS assessment based on the ILI idealization of the corrosion defect. While a difference of 6.5% does not seem particularly large, it is important to note that failures of corroded pipelines can have severe consequences and that repair of corroded pipelines is costly (it costs more than $200,000 in Canada to excavate and repair a single corroded pipe joint that is about 12 m long based on information provided by industry experts). It follows that even a relatively small improvement in the accuracy of the FFS assessment could have significant implications for improving the pipeline safety and reducing the cost of maintenance. A methodology that provides more accurate burst capacity prediction based on ILI data is therefore of great practical value.
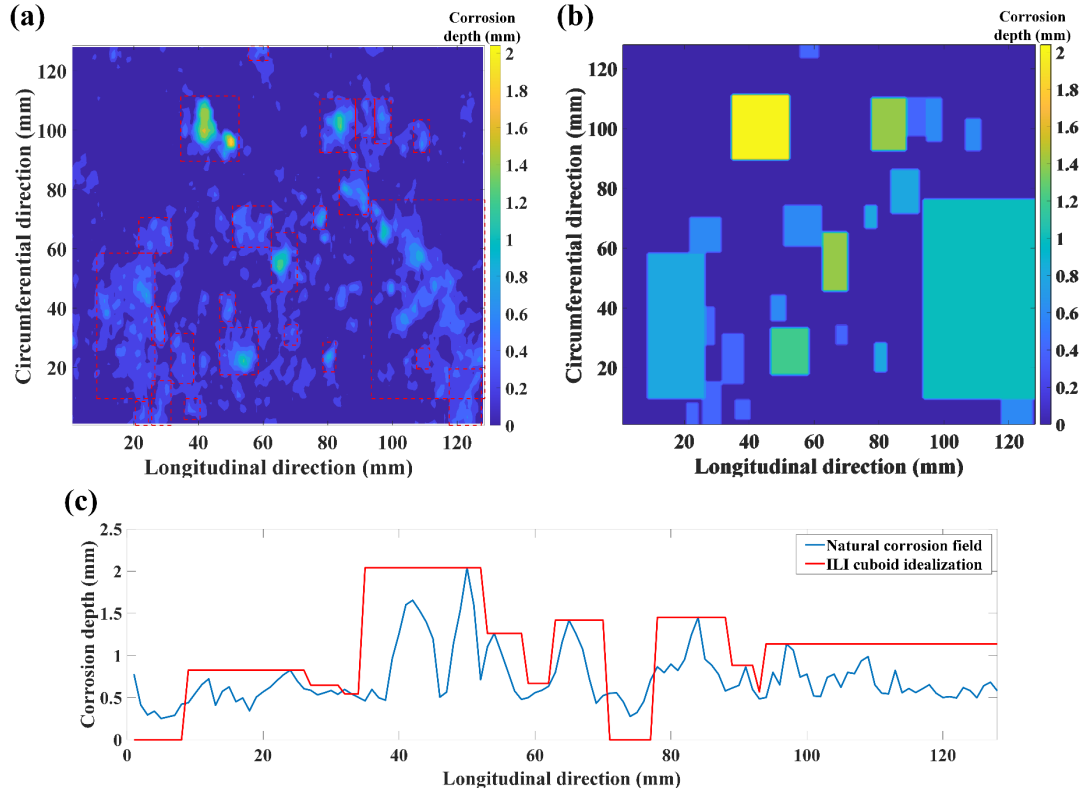
Figure 4.3 Natural corrosion field and the corresponding ILI cuboid idealization: (a) the natural corrosion field, where the preserved ILI-detected anomalies are highlighted by the red dashed line; (b) two-dimensional plot of the ILI-idealized corrosion field; and (c) the river-bottom profiles of the natural corrosion field and its ILI cuboid idealization

4.3 Simulation of corrosion fields based on ILI data

4.3.1 General methodology

As explained in Section 4.2, a corrosion anomaly reported by ILI is represented by two location parameters and three sizing parameters. Suppose that n ($n = 1, 2, \ldots$) corrosion anomalies are identified in a given corrosion defect in the ILI data. The field is represented by $5n$ parameters, and they are the only information available to pipeline integrity engineers in addition to the basic pipeline attributes (e.g. D, wt, and σ_y). The

methodology proposed in the present study is aimed at simulating corrosion fields based on these available parameters.

The main idea behind the proposed methodology is that a corrosion defect can be considered as a nonhomogeneous non-Gaussian random field consisting of multiple anomalies. Each anomaly is assumed to be a homogeneous non-Gaussian random field with the marginal probability distribution and ACF estimated based on the sizing parameters of the anomaly obtained from ILI. These homogeneous non-Gaussian anomalies are then spatially modulated based on the location parameters of the anomalies obtained from ILI and merged to form the nonhomogeneous non-Gaussian field that characterizes the corrosion defect.

4.3.2 Marginal distribution, ACF, and spatial modulator of corrosion anomalies

The marginal distribution and ACF of the homogeneous non-Gaussian random field that characterizes the individual corrosion anomaly are evaluated based on the laser scan data of 105 corrosion defects on 14 naturally corroded pipe specimens removed from in-service natural gas pipelines in Canada (Zhang et al., 2018; Bao and Zhou, 2021a). These corrosion defects have the maximum depths ranging from 10 and 80% of the pipe wall thickness. Note that the laser scanner applied to 6 of the 14 specimens has a scanning resolution of 1×2 mm, and the laser scanner applied to the remaining 8 specimens has a resolution of 1×1 mm. A total of 3408 corrosion anomalies are then identified from the 105 corrosion fields through the virtual ILI process described in Section 2 (the actual ILI data for the 14 pipe segments are not provided to the present study). The 3408 anomalies range in sizes from $w \times l = 5 \times 6$ mm to $w \times l = 349 \times 642$ mm.

The empirical probability distributions of the corrosion depths in 3408 anomalies are developed. Ideally the empirical distribution for a given anomaly should be estimated using corrosion depths at well separated grid points within the anomaly such that the samples can be considered approximately independent. For small anomalies that contain

less than 100 grid points, this will markedly reduce the number of samples available for evaluating the empirical distribution. However, such anomalies are generally shallow with d_{max} less than 14%wt and therefore have a small impact on the burst capacity evaluation. For the 2158 anomalies that contain more than 100 grid points, it is observed that the mean and standard deviation of the corrosion depth estimated using well-separated samples (with the longitudinal separation greater than 0.3l or the circumferential separation greater than 0.3w) are very close to those estimated using all the samples. Therefore, the marginal distribution is estimated using all the sample points within the anomaly in the following.

The beta distribution with the lower and upper bounds equal to zero and maximum corrosion depth (d_{max}) of the anomaly, respectively, is found to fit the empirical distribution well. As an illustration, the empirical probability density function (PDF) and corresponding fitted beta distributions for four anomalies with d_{max}/wt ranging from 20% to 50% are illustrated in Fig. 4.4. The empirical PDF is estimated by the kernel density estimation (Li and Ng, 2010). One hundred equally spaced points in the range of depth values are assigned, and the probability of corrosion depths within each interval are calculated. The empirical probability values are then smoothed by a Gaussian kernel with details given in (Li & Ng, 2010). This method has been coded in the MATLAB build-in function "ksdensity".
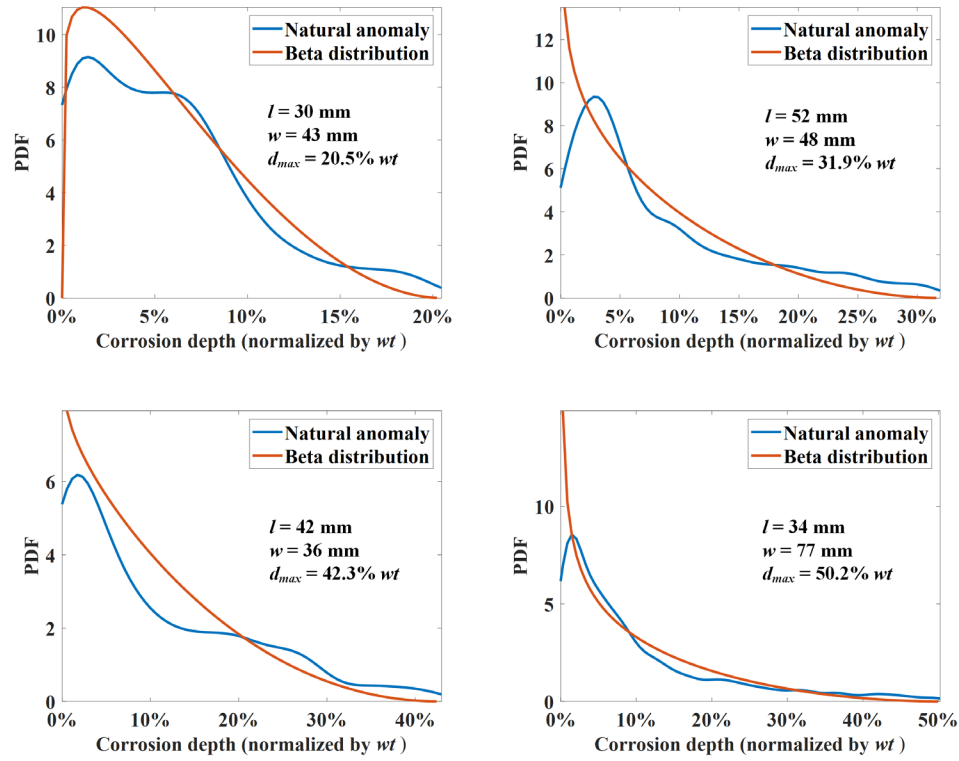
Figure 4.4 The empirical PDF and fitted beta distributions for four corrosion anomalies

As described in Section 4.3.1, the goal of the proposed methodology is to simulate corrosion fields based only on the ILI data associated with the corrosion anomalies. It follows that one needs to estimate the parameters of the marginal distribution (i.e. the fitted beta distribution) of the corrosion depths in the anomaly from the three ILI sizing parameters (i.e. l, w and d_{max}). Let σ_d and $\bar{d}$ denote the standard deviation and mean value, respectively, of corrosion depths in a given anomaly. The quantity $V_{real} = l \cdot w \cdot \bar{d}$ (mm³) quantifies the actual volume of the metal loss in the anomaly. It is observed from the data corresponding to the 3408 anomalies (Fig. 4.5) that there is a moderately strong linear relationship between σ_d and d_{max}, and that there is a strong linear relationship between $\ln(V_{real})$ and $\ln(V_{ILI})$, where $V_{ILI} = l \cdot w \cdot d_{max}$ (mm³) is the volume

of the ILI-idealized cuboid-shaped anomaly. Therefore, the linear regression analysis is employed to estimate the moments of the marginal distribution from the ILI sizing parameters. The following two equations are then proposed based on Fig. 4.5

$$\sigma_d = 0.179 d_{max} + 0.027 \tag{4.1}$$

$$\bar{d} = 1.130 \frac{(l \cdot w \cdot d_{max})^{0.811}}{l \cdot w} \tag{4.2}$$

where l, w, d_{max}, σ_d and $\bar{d}$ are all in the unit of mm. The coefficients of determination for Eqs. (4.1) and (4.2) are 0.758 and 0.977, respectively. Once the lower bound (zero), upper bound (d_{max}), σ_d and $\bar{d}$ are determined, the parameters of the marginal distribution (i.e. beta distribution) of the corrosion depths within the anomaly can be estimated by the method of moments (Ang and Tang, 2006).
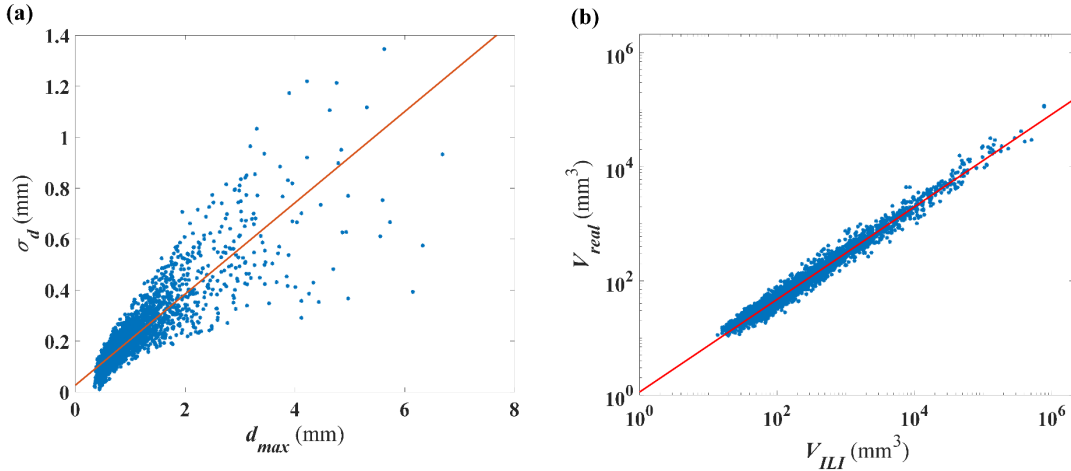
Figure 4.5 (a) Anomaly depth versus the standard deviation of actual corrosion depth; and (b) the cuboidal volume loss of anomalies versus the real volume loss (in logarithm)

In addition to the marginal distribution, the ACF or equivalently the power spectral density (PSD) function of the corrosion anomaly needs to be evaluated. Since the corrosion anomaly is characterized as a homogeneous random field in the present study, its PSD and ACF are connected through the Wiener-Khinchin theorem (Ponomarev,

2020). The empirical ACF $\rho(\tau_x,\ \tau_y)$ for a given set of longitudinal and circumferential distances (τ_x and τ_y) is calculated by evaluating the correlation coefficient between the corrosion depths at two sets of grid points separated by τ_x and τ_y. In the present study, the empirical ACF for each of the 3408 anomalies are calculated from the laser scan data, and the following exponential function is used to fit the ACF:

$$\rho(\tau_x, \tau_y) = \exp\left(-\left(\frac{\tau_x}{l_L}\right)^{a_L} - \left(\frac{\tau_y}{l_C}\right)^{a_C}\right) \tag{4.3}$$

where l_L and l_C are the longitudinal and circumferential correlation lengths (mm), respectively; a_L and a_C are the longitudinal and circumferential decaying rates, respectively. For each reported anomaly, l_L, l_C, a_L, and a_C are fitted using the least squares method. Equation (4.3) is justified in that the exponential-type autocorrelation function is widely employed in the literature (Schulz et al., 2018; Williams and Rasmussen, 1995). Furthermore, the anisotropic correlation lengths and decaying rates in Eq. (4.3) provide flexibility for fitting the empirical ACF. The histogram of fitted parameters of ACF of 3408 anomalies are shown in Fig. 4.6. As an illustration, the empirically calculated and corresponding fitted ACF for three anomalies with different ILI sizing parameters are shown in Fig. 4.7.
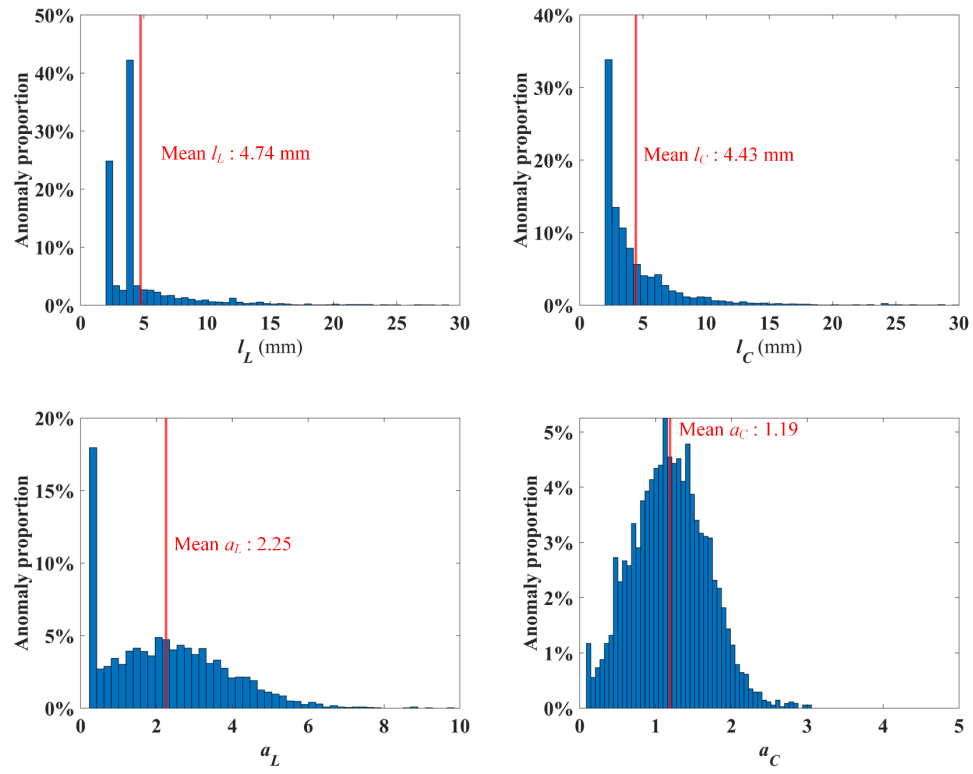
Figure 4.6 Histogram of fitted parameters of ACF of 3408 anomalies
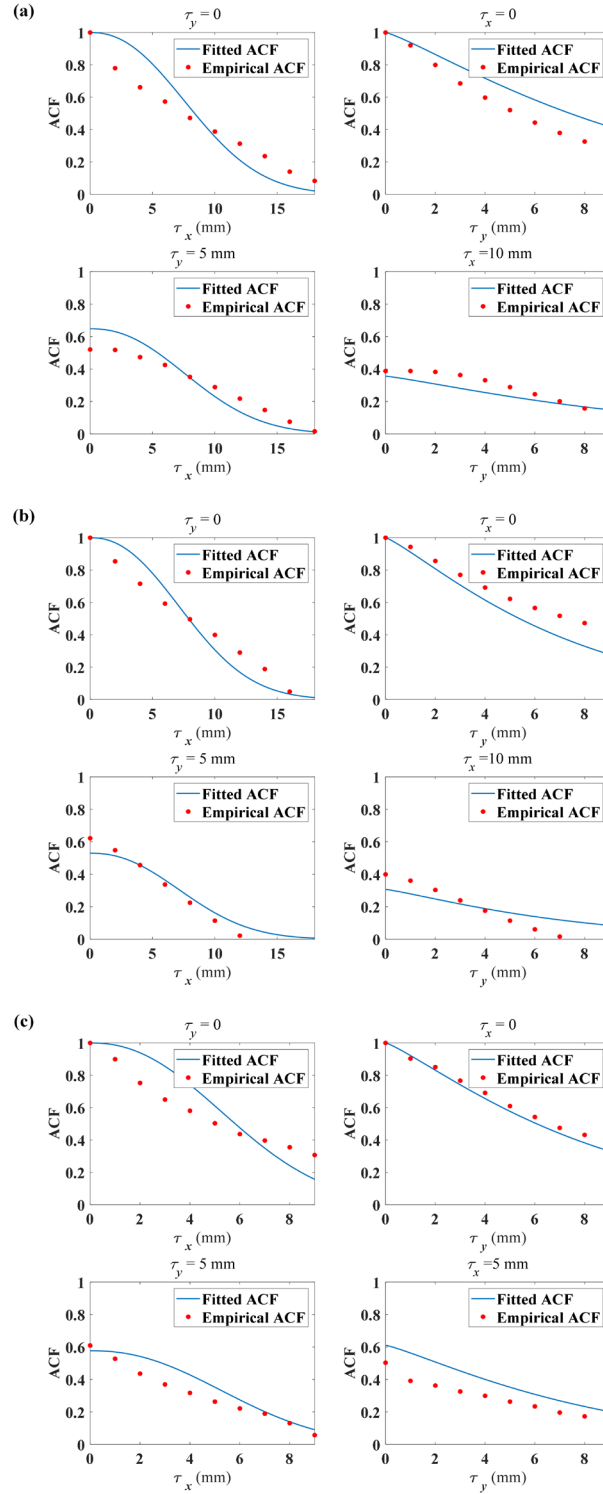
Figure 4.7 Empirical (red points) and corresponding fitted ACF for three anomalies with different ILI sizing parameters ($l \times w \times d_{max}$): (a) 66 mm$\times$80 mm$\times$1.055 mm; (b) 60 mm$\times$49 mm$\times$1.293 mm; and (c) 37 mm$\times$59 mm$\times$1.086 mm

Similar to the marginal distribution of the corrosion depth, the goal is to estimate l_L, l_C, a_L, and a_C from the three ILI sizing parameters (i.e. l, w, and d_{max}). To this end, the correlation coefficients between the ILI sizing parameters and parameters of ACF for the 3408 anomalies are summarized in Tables 4.1 and 4.2. The correlation coefficients shown in Tables 4.1 and 4.2 suggest that one can use $\ln(l)$ and $\ln(w)$ to estimate $\ln(l_L)$ and $\ln(l_C)$, respectively. The values of $\ln(l_L)$ ($\ln(l_C)$) are plotted versus $\ln(l)$ ($\ln(w)$) for the 3408 anomalies in Fig. 4.8(a) (4.8(b)). Based on Fig. 4.8, the following linear regression equations, represented by the red lines in the figure, are then proposed:

$$\ln(l_L) = 0.630 \ln(l) - 0.352 \tag{4.4}$$

$$\ln(l_C) = 0.654 \ln(w) - 0.555 \tag{4.5}$$

where l, w, l_L, and l_C are all in mm. Since a_L and a_C do not have sufficiently strong correlations with any of the three ILI sizing parameters, fixed values of a_L and a_C that equal corresponding mean values of the 3408 anomalies are proposed, i.e. $a_L = 2.25$ and $a_C = 1.19$. After substituting Eqs. (4.4) and (4.5) into Eq. (4.3), the ACF employed in the present study is given below

$$\rho(\tau_x, \tau_y) = \exp\left(-\left(\frac{1.421\tau_x}{l^{0.630}}\right)^{2.25} - \left(\frac{1.741\tau_y}{w^{0.654}}\right)^{1.19}\right) \tag{4.6}$$

where l, w, τ_x, and τ_y are all in mm.

Table 4.1 The correlation coefficients of the ILI sizing parameters and the fitted ACF parameters

ILI sizing parameters	Parameters of ACF			
(mm)	l_L (mm)	l_C (mm)	a_L	a_C
l	0.545	0.420	0.030	-0.024
w	0.639	0.591	0.172	-0.092
d_{max}	0.368	0.354	0.322	-0.177

Table 4.2 The correlation coefficients of the logarithm of the ILI sizing parameters and the fitted ACF parameters (l, w, d_{max}, l_L, and l_C are in mm)

Logarithm of ILI sizing parameters	Logarithm of parameters of ACF			
	$\ln(l_L)$	$\ln(l_C)$	$\ln(a_L)$	$\ln(a_C)$
$\ln(l)$	0.759	0.502	0.150	-0.007
$\ln(w)$	0.519	0.734	0.347	-0.015
$\ln(d_{max})$	0.328	0.562	0.421	0.169
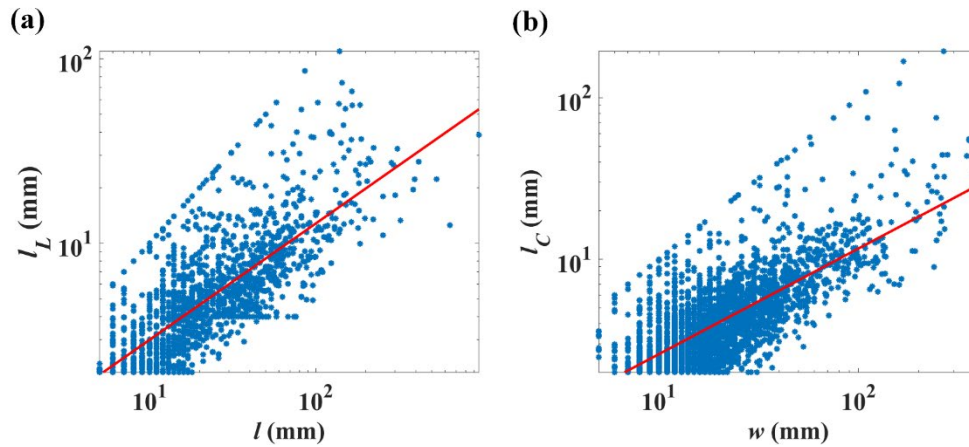
Figure 4.8 (a) Anomaly length l versus the fitted correlation length l_L (in logarithm); and (b) anomaly width w versus the fitted correlation length l_C (in logarithm)

Once the ACF and marginal distribution are estimated, they can be employed to generate non-Gaussian homogeneous fields corresponding to individual corrosion anomalies. To this end, a spatial modulator $M(x, y)$ is employed to ensure that the non-zero corrosion depths are localized in the anomaly's region. The spatial modulator is essentially the window function that is commonly used in the signal processing to focus on a specific segment of the signal such as analyzing the localized frequency of a nonstationary signal (Sejdić et al., 2009). A rectangular window function $M(x, y)$, whose value is unity if (x, y)

is inside the anomaly and zero if (x, y) is outside the anomaly, can be selected as $M(x, y)$. However, this leads to discontinuities at the anomaly's edges, which is incompatible with real anomalies. Moreover, the discontinuity will lead to significant stress concentrations if the finite element analysis is employed to evaluate the burst capacity of the synthetic corrosion field. Therefore, a Gaussian tapered rectangular window function is selected in this study to allow a smooth transition of the corrosion depth from the inside to outside of a corrosion anomaly. Such a function has been employed in signal processing studies to extract the local information and calculate the correlation between different portions of the signal (Allen et al., 2014; Yang et al., 2014). The width of the smooth transition is approximately three times the standard deviation of the Gaussian function, which is selected to be 3 mm. This allows the transition width to be slightly shorter than the length of the smallest anomaly considered, which is 10 mm. The boundaries of an anomaly with location parameters (x_c and y_c) and sizing parameters (l and w) are defined as: $x_{min} = x_c - l/2$, $x_{max} = x_c + l/2$, $y_{min} = y_c - w/2$, and $y_{max} = y_c + w/2$. Based on the above, the spatial modulator $M(x, y)$ for the anomaly is defined by

$$M(x, y) = \exp\left(-\left(\frac{T_x}{3}\right)^2 - \left(\frac{T_y}{3}\right)^2\right) \tag{4.7}$$

$$T_x = \max\{0, \ x_{min} - x, \ x - x_{max}\} \tag{4.8}$$

$$T_y = \max\{0, \ y_{min} - y, \ y - y_{max}\} \tag{4.9}$$

where T_x and T_y are in mm. The 2D plot and the profile $M(x, y_c)$ of the spatial modulator for an anomaly with $x_c = y_c = 50$ mm and $l = w = 20$ mm are illustrated in Fig. 4.9.
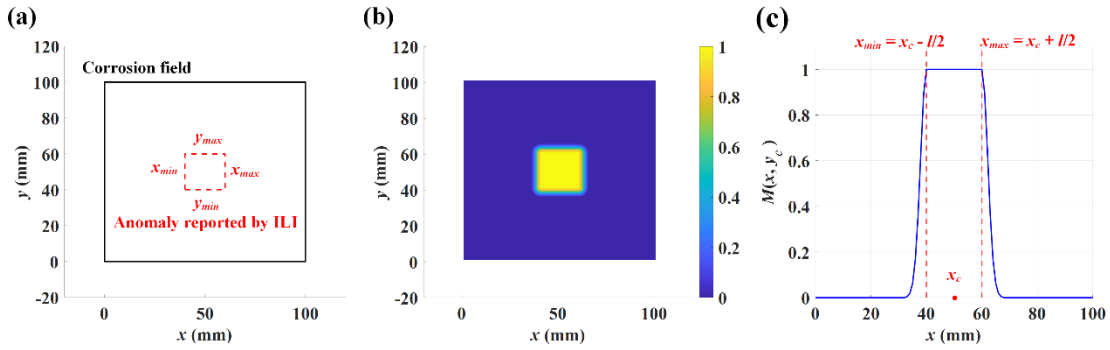
Figure 4.9 The spatial modulator $M(x, y)$ constructed from an ILI reported anomaly: (a) the boundaries and centre of the anomaly; (b) the 2D modulator $M(x, y)$ based on the location and sizing parameters (x_c, y_c, l, and w) of this anomaly; and (c) the profile $M(x, y_c)$ of this modulator

4.3.3   Simulation procedure

Consider a corrosion defect consisting of n anomalies. The following procedure is employed to simulate the corrosion depths within the ith (i = 1, 2, …, n) anomaly based on the two location parameters ($x_{c,i}$ and $y_{c,i}$) as well as three sizing parameters (l_i, w_i and $d_{max,i}$).

(1) Estimate the mean $\bar{d}_i$ and standard deviation $\sigma_{d,i}$ of the corrosion depths from l_i, w_i and $d_{max,i}$ using Eqs. (4.1) and (4.2). Define the marginal distribution of the corrosion depths to be a beta distribution with the above-determined $\bar{d}_i$ and $\sigma_{d,i}$, a lower bound of zero and an upper bound of $d_{max,i}$.

(2) Define the ACF using Eq. (4.6) and evaluate the corresponding PSD function using the Wiener-Khinchin theorem.

(3) Based on the marginal distribution and the PSD function obtained from previous steps, use the algorithm proposed by Masters and Gurley (2003) to generate a homogeneous non-Gaussian field $z_i^{(H)}(x, y)$. The details of Masters and Gurley's algorithm are summarized in Appendix B.

(4) Define the spatial modulator $M_i(x, y)$ from $x_{c,i}$, $y_{c,i}$, l_i and w_i using Eqs. (4.7) – (4.9). The nonhomogeneous non-Gaussian field $z_i^{(B)}(x, y) = z_i^{(H)}(x, y) \cdot M_i(x, y)$ is obtained from $z_i^{(H)}(x, y)$ modulated by $M_i(x, y)$. The corrosion depth on any grid point in $z_i^{(B)}(x, y)$ is set to zero if it is less than 0.001 mm.

(5) Estimate the empirical cumulative distribution function $F_{Z_i^{(B)}}(z)$ of the non-zero value of $z_i^{(B)}(x, y)$. Let $F_{Z_i}(z)$ denote the target marginal distribution for this anomaly. The non-zero value of $z_i^{(B)}(x, y)$ is then replaced by $F_{Z_i}^{-1}\left(F_{Z_i^{(B)}}\left(z_i^{(B)}(x, y) \right) \right)$, where $F_{Z_i}^{-1}(\bullet)$ denote the inverse of $F_{Z_i}(\bullet)$. Define the random field after matching the marginal distribution as a spatially modulated field $z_i(x, y)$.

Although Masters and Gurley's algorithm ensures that the marginal distribution of the homogeneous field $z_i^{(H)}(x, y)$ matches the target distribution obtained in Step (1), the marginal distribution of non-zero corrosion depth of $z_i^{(B)}(x, y)$ may deviate from the target distribution after Step (4), especially for small anomalies that do not have many grid points. Therefore, Step (5) is introduced to ensure that the spatially modulated field $z_i(x, y)$ has non-zero corrosion depths consistent with the target marginal distribution. The entire corrosion field is then obtained by $z^{(B)}(x, y) = \max\{z_1(x, y), z_2(x, y), \dots, z_n(x, y)\}$. That is, the corrosion depth at a given grid point of the synthetic corrosion field $z_i^{(B)}(x, y)$ is the maximum value of the n spatially modulated fields $z_i(x, y)$ ($i = 1, 2, \dots, n$) at this point. The procedure of simulating the corrosion field $z^{(B)}(x, y)$ based on ILI parameters is illustrated in Fig. 4.10 for a field consisting of three anomalies.
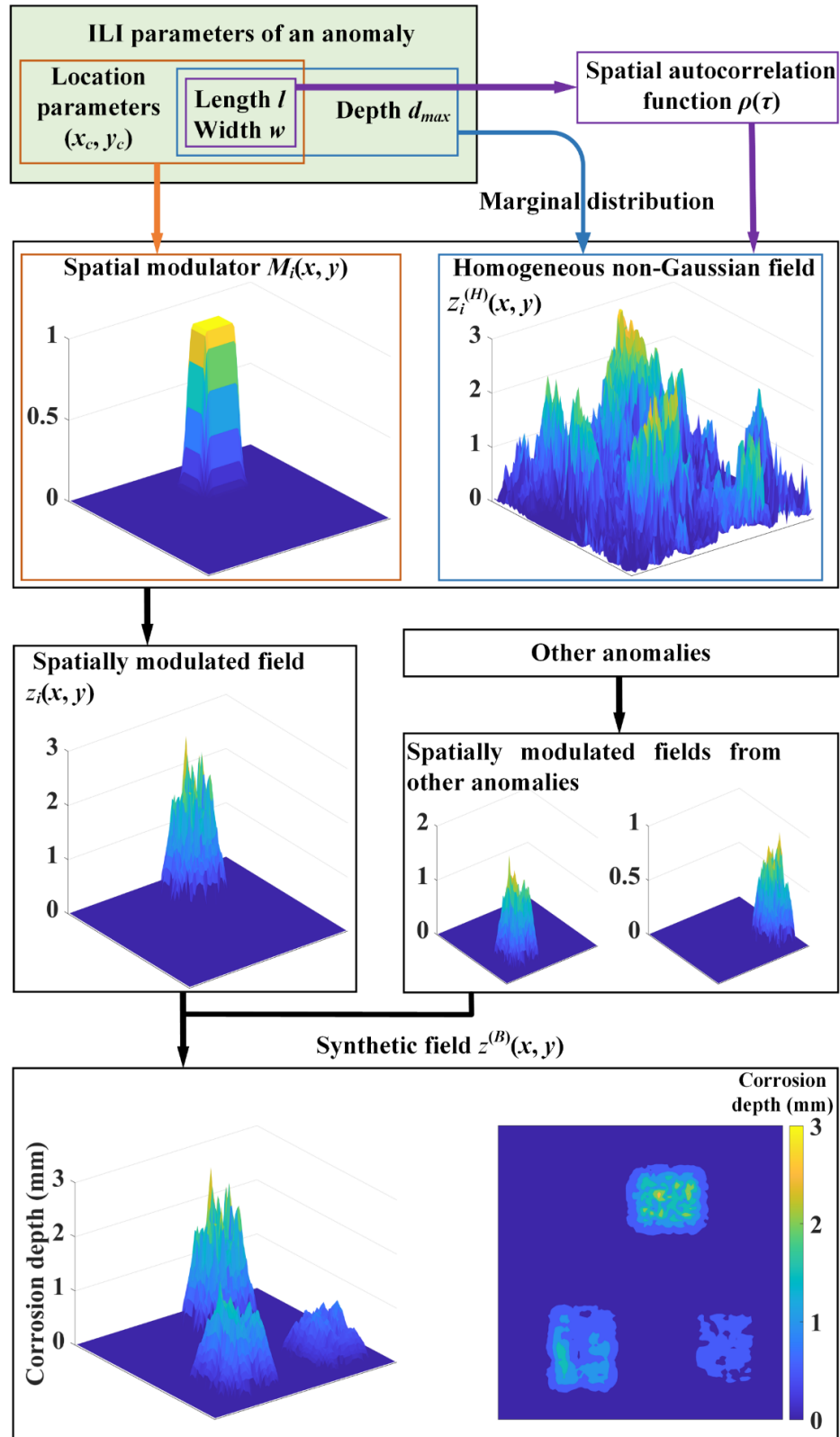
Figure 4.10 Procedures of simulating corrosion fields based on ILI parameters

4.3.4   Burst capacities of pipe segments containing corrosion field

To illustrate the practical value of the simulation methodology described in Section 4.3.3, the burst capacity of the synthetic corrosion field based on the ILI information is evaluated and compared with those of the actual and ILI-idealized cuboid-shaped corrosion fields. The burst capacities of the 105 natural corrosion defects described in Section 4.3.2 and the corresponding synthetic fields are computed using the well-known RSTRENG model (Kiefner and Vieth, 1989). The RSTRENG model is a semi-empirical model for predicting the burst capacities considering the corrosion morphology. The model evaluates the burst capacity p_b at a corrosion defect as follows:

$$p_b = \frac{2 \cdot wt \cdot \sigma_f}{D} \frac{1 - \frac{A_e}{l_e \cdot wt}}{1 - \frac{A_e}{M_e \cdot l_e \cdot wt}} \tag{4.10}$$

$$M_e = \begin{cases} \sqrt{1 + 0.6275 \frac{l_e^2}{D \cdot wt} - 0.003375 \left(\frac{l_e^2}{D \cdot wt}\right)^2}, & \frac{l_e^2}{D \cdot wt} \leq 50 \\ 3.3 + 0.032 \frac{l_e^2}{D \cdot wt}, & \frac{l_e^2}{D \cdot wt} > 50 \end{cases} \tag{4.11}$$

where σ_f is the flow stress of the pipe steel and defined as $\sigma_y + 69$ (MPa); A_e and l_e are the effective area and length, respectively, of the river bottom profile of the 3D corrosion field, and M_e is the so-called Folias factor corresponding to the effective length of the corrosion field. The procedure to determine A_e and l_e is not presented for the sake of brevity as it has been well described in many references (e.g. Kiefner and Vieth, 1989; Zhang et al., 2018). As the 105 corrosion defects are identified on 14 different pipe segments (Zhang et al., 2018), the values of D, wt and σ_y employed in RSTRENG for a given defect are those of the pipe segment that contains the defect as reported in Zhang et al. (2018).

The three-dimensional (3D) elasto-plastic finite element analysis (FEA) has proven to be highly accurate in evaluating the burst capacity of corroded pipelines (Bao et al., 2018; Zhang and Zhou, 2020). To provide a benchmark for the burst capacity evaluation and validate the accuracy of the RSTRENG model, the 3D FEA is also employed to evaluate

the burst capacities of the pipeline segment (D = 407.7 mm and wt = 6.2 mm) containing one of the 105 natural corrosion fields, the corresponding ILI idealization, and synthetic fields. The edges of anomalies in ILI idealizations are smoothed by a Gaussian function with a standard deviation of 3 mm to avoid severe stress concentrations in FEA. The finite element model is developed and implemented using the commercial package ABAQUS 2018. The elastic modulus E = 167 GPa, σ_y = 369 MPa and the tensile strength σ_u = 540 MPa are obtained from the coupon test results. The von Mises yield criterion, isotropic strain hardening and finite-strain formulation for the large deformation near burst are adopted in the model. The following true stress - true strain (i.e., σ - ε) relationship is adopted in FEA:

$$\sigma = \begin{cases} E\varepsilon, & \sigma < \sigma_y \\ K\varepsilon^{n_h}, & \sigma \geq \sigma_y \end{cases} \tag{4.12}$$

where $K = \frac{\sigma_u e^{n_h}}{(n_h)^{n_h}}$ is the strength coefficient, and n_h is the strain hardening exponent. The value of n_h is estimated using (Zhu & Leis, 2005):

$$n_h = 0.244 \left(\frac{\sigma_u}{\sigma_y} - 1 \right)^{0.604} \tag{4.13}$$

The eight-node hexahedral (C3D8) element with full integration is used in this study. The pipeline segment is extended by three times of outside diameter corrosion-free at both ends to avoid the end effect. The mesh resolution at corroded area is identical to the resolution of the corrosion field. Coarse meshes are employed at the transition area and corrosion-free area outside the corrosion field. Four layers of elements are used along the pipe wall thickness direction. An example of an FEA model used in this study is illustrated in Fig. 4.11. The mesh resolution at the corroded area is 2 mm in the longitudinal direction and 1 mm in the circumferential direction. The mesh resolution at the transition area is 4 × 2 mm, while it is 8 ×4 mm in the corrosion-free area. The total number of nodes and elements are 183,120 and 146,208, respectively. One end of the pipe segment is considered as a fixed support, and longitudinal displacement is constrained at the other end. The burst capacity is defined as the internal pressure at which the maximum nodal von Mises stress within the corrosion field reaches the true

stress corresponding to the ultimate tensile strength (Zhang and Zhou, 2020). The simulation of corrosion fields and FEA are carried out on a desktop with Intel(R) Core(TM) i7-8700 CPU @3.20 GHz and 16 GB RAM
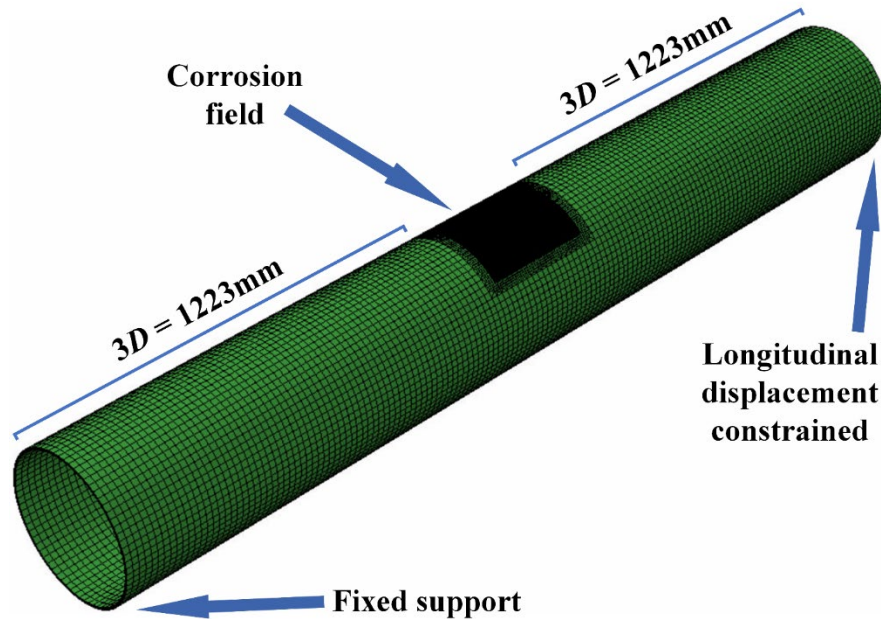
Figure 4.11 FEA model of the corroded pipe segment

4.4 Results and discussion

The computation time to generate a sample of the synthetic corrosion field depends on the size of the field and the number of anomalies in the field, and it usually takes less than ten seconds for one realization. The time to simulate the corrosion fields based on the (virtual) ILI information of the natural corrosion field shown in Fig. 1 is 0.28 second per realization. Figure 4.12 compares the natural corrosion field, ILI idealization, and two synthetic fields.
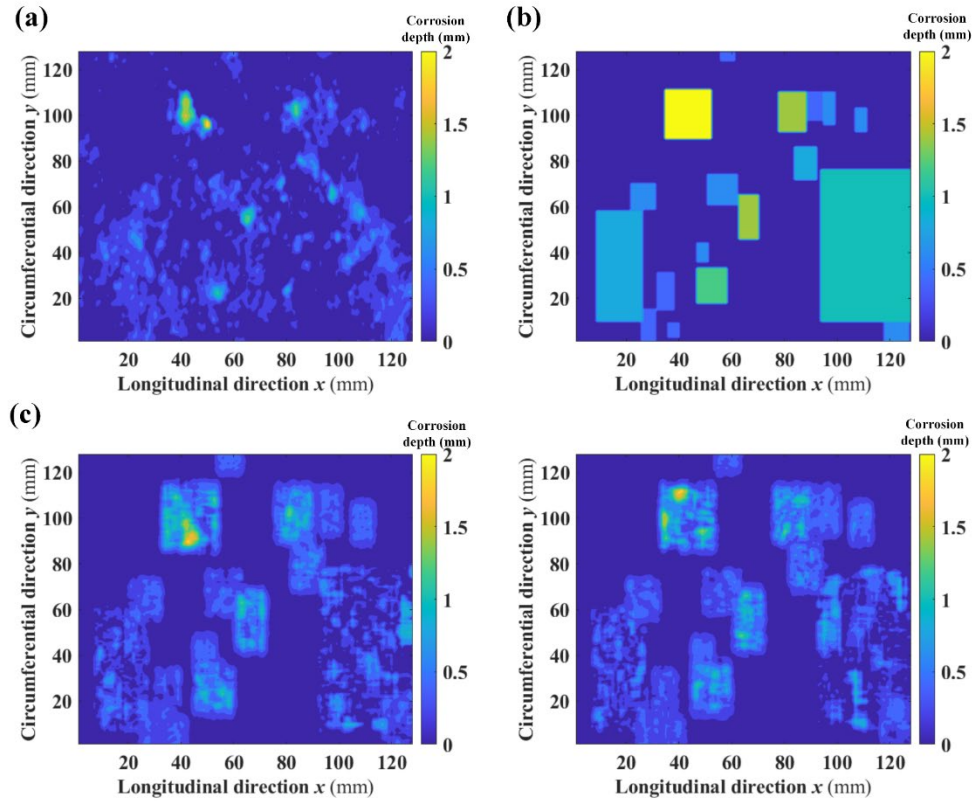
Figure 4.12 Two-dimensional plot of corrosion fields:(a) natural corrosion field; (b) ILI idealization; and (c) two synthetic fields

As shown in Fig. 4.13, the synthetic fields match the reported ILI data: grid points in the corrosion-free area (i.e. not reported by ILI) have corrosion depths below 5%*wt*, and the corrosion depth of each grid point does not exceed the corresponding ILI idealization. The ILI idealization can be considered as the envelope of synthetic fields. Since the textural information of the corrosion fields is not included in the ILI parameters, the generated synthetic fields are not visually similar to the natural corrosion field. If the detailed profile (such as the laser scan data) of the entire corrosion cluster is available, one can generate the synthetic samples of the cluster that are visually similar to the actual cluster, as demonstrated in previous studies by Bao and Zhou (2021) and Zhou et al. (2021). However, detailed corrosion profiles are usually unavailable in practice. We assume in the present study that the only known information about an individual

corrosion anomaly is its length, width and maximum depth, and two location parameters, as reported by ILI. The profile of the synthetic corrosion anomaly is generated based on such information. Therefore, it is not unexpected that the synthetic corrosion clusters are visually different from the corresponding actual cluster.
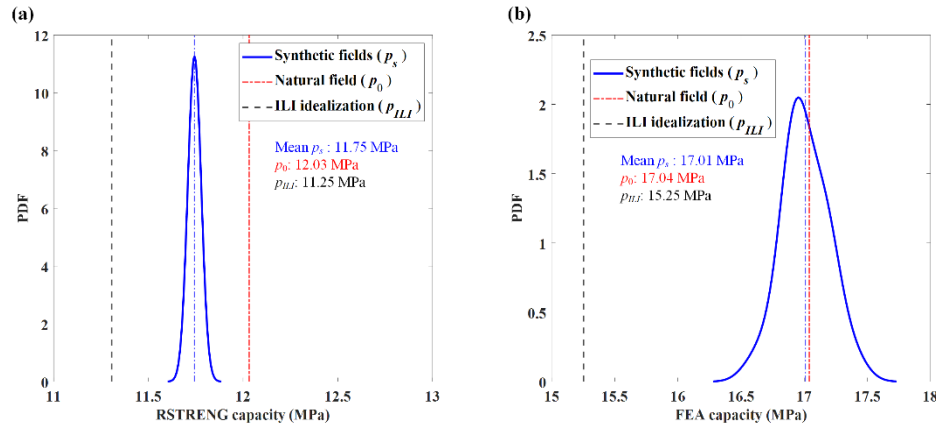
Figure 4.13 Empirical PDF of the burst capacity of pipe segment containing the synthetic fields, natural fields, and ILI idealization: (a) predicted by the RSTRENG model; and (b) predicted by FEA

Let p_0, p_{ILI}, and p_s denotes the burst capacities corresponding to the natural corrosion field, ILI idealization, and synthetic field, respectively. The values of p_0 and p_{ILI}, and PDF of the p_s of 100 realizations of the synthetic field corresponding to the natural corrosion field shown in Fig. 4.1 are compared in Fig. 4.13. Note that both the RSTRENG model and FEA are employed to predict p_0, p_{ILI} and p_s. The figure indicates that for this particular corrosion field, the RSTRENG-predicted burst capacities for the synthetic fields are 1.6% - 2.9% lower compared with p_0. The burst capacities of the synthetic fields are closer to p_0 than p_{ILI} regardless if the RSTRENG model or FEA is employed: the mean of RSTRENG (FEA)-predicted p_s is 2.4% lower than p_0, whereas the RSTRENG (FEA)-predicted p_{ILI} is 6.5% lower than p_0. Furthermore, the variability of p_s is generally small: the coefficients of variation (COV) of RSTRENG- and FEA-predicted p_s are 0.3% and 1.0%, respectively.

The RSTRENG-predicted burst capacities of 105 natural corrosion fields, as well as the corresponding ILI idealizations and synthetic fields are compared in Fig. 4.14. For each corrosion field, 100 synthetic fields are generated. The error bar in Fig. 4.14 corresponds to the lower and upper bounds of burst capacities of the 100 realizations. Since the ILI idealizations typically overestimate the volume loss of the pipe wall due to corrosion, p_{ILI} are conservative compared with p_0 for most of the 105 corrosion fields. Only two corrosion fields have p_{ILI} slightly higher than the corresponding p_0. The corrosion depth of most grid points in these two natural corrosion fields are around 5%wt. Since ILI does not record the corrosion depth below 5%wt, the effective area of the river-bottom profile of the ILI idealization is slightly smaller than that of the natural corrosion field, resulting in the non-conservative prediction of the burst capacity. Let $\bar{p}_s$ denote the mean of p_s for a specific natural corrosion field. The mean value of p_{ILI}/p_0 of the 105 corrosion fields is 91.5%, while it is 97.4% for $\bar{p}_s/p_0$. This demonstrates the effectiveness of the proposed methodology for simulating synthetic fields.
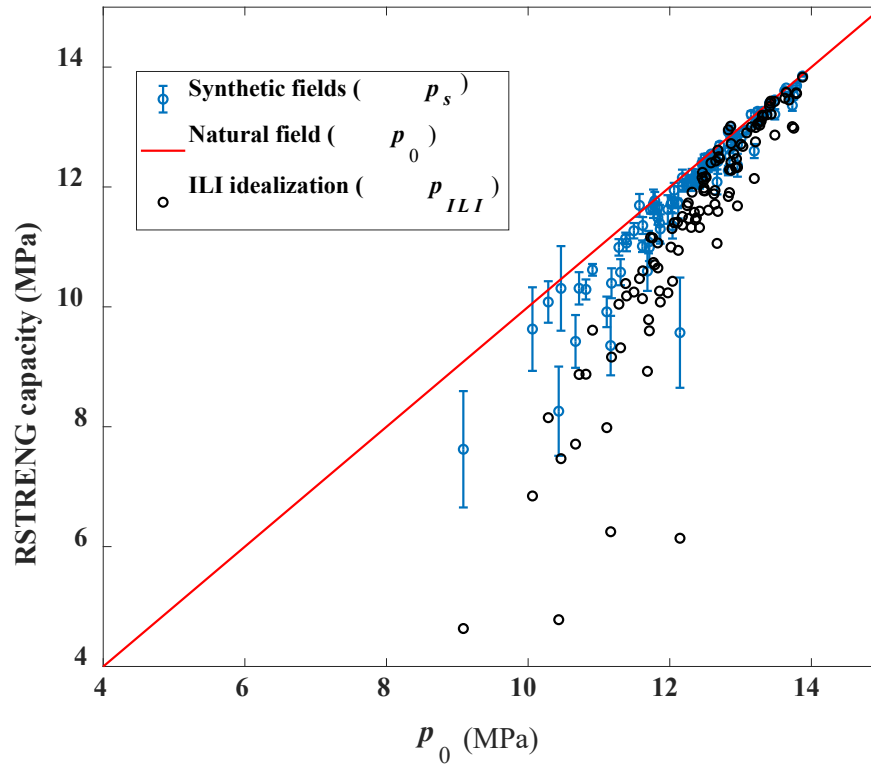
Figure 4.14 RSTRENG-predicted burst capacity of pipe segment containing the synthetic corrosion fields, natural corrosion fields, and ILI idealizations

Although the synthetic and natural fields are visually different, the statistics of the corrosion depths in both fields are similar. This is critical for the burst capacity prediction. We observe that the marginal distribution of the corrosion depth has a larger effect on the burst capacity than the spatial correlation of the corrosion depth, while the visual similarity between the synthetic and actual corrosion fields depends largely on the spatial correlation. Since the marginal distribution of corrosion depths is effectively captured through the empirical models (i.e. Eqs. (4.1) and (4.2)) proposed in this study, the burst capacity of the synthetic field is similar to that of the actual corrosion field.

It is worth noting that p_{ILI} is highly conservative compared with p_0 for a few corrosion fields. For example, one of such fields is depicted in Fig. 4.15. The corrosion pits are

densely distributed in this field. Since the grid points between corrosion pits have a corrosion depth slightly deeper than 5%*wt*, these pits are grouped and identified as a few large anomalies, resulting an extremely conservative prediction of the burst capacity based on the ILI idealization. Although the predictions based on the synthetic fields are also conservative, they are much more accurate than the ILI idealization: $p_{ILI}/p_0 = 50.9\%$, whereas $\bar{p}_s/p_0$ is 82.5%.
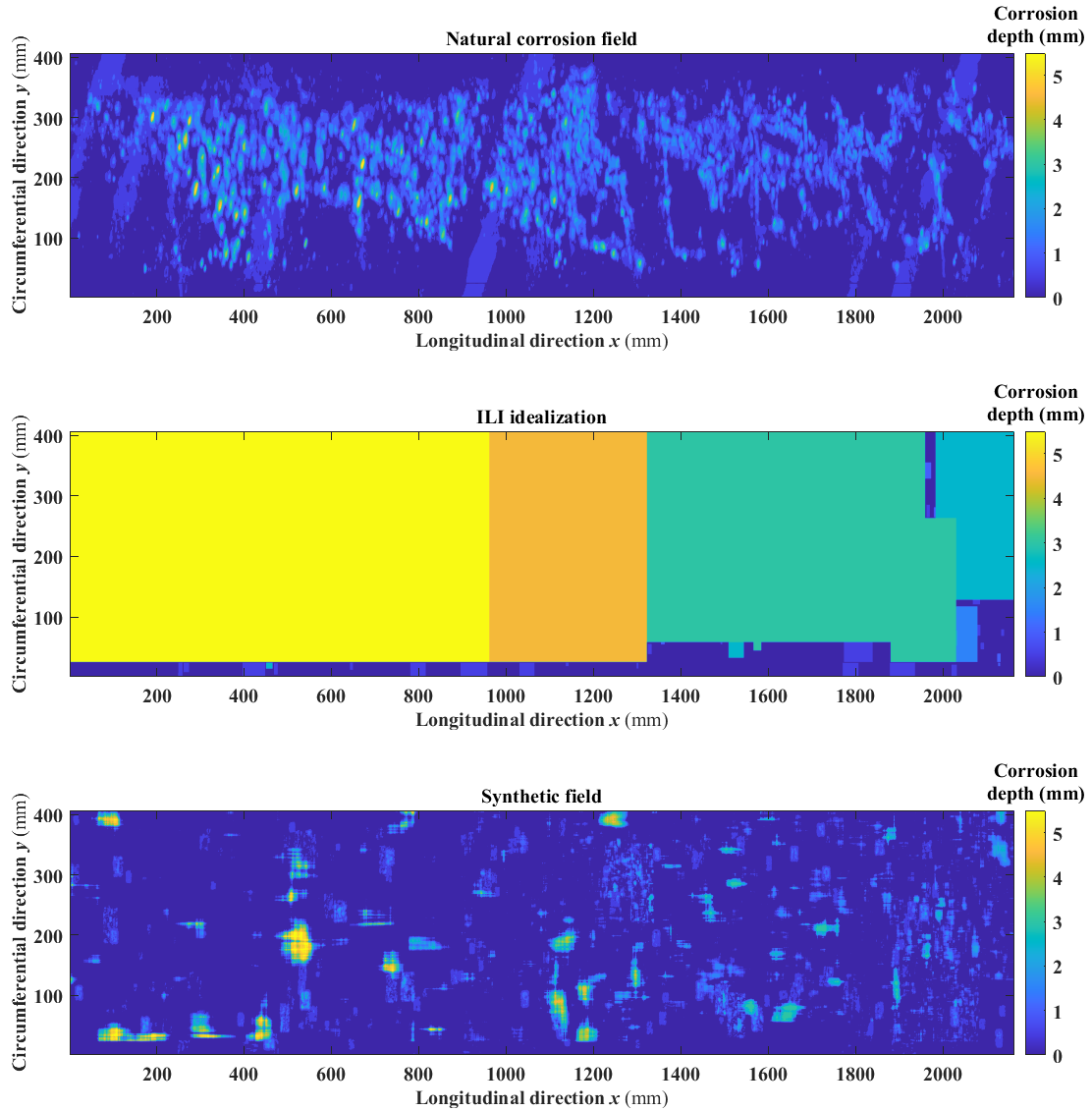
Figure 4.15 Two-dimensional plot of corrosion fields:(a) natural corrosion field; (b) ILI idealization; and (c) synthetic field

In Section 4.3.2, fixed values of a_L and a_C are suggested, which equal the respective mean values obtained from the fitted values of the ACFs of 3408 anomalies, because the correlation coefficients between these two parameters and the ILI sizing parameters are low. However, the histograms in Fig. 4.6 indicate that the fitted a_L values range mostly between 0 and 6, and a_C values range mostly between 0 and 2.5. If different values of a_L and a_C are employed in Eq. (4.6), the burst capacity corresponding to the synthetic fields may change. Table 4.3 summarizes the mean ratio $\overline{p_s}/p_0$ under different values of a_L and a_C. Each of the mean ratio is calculated from 3150 synthetic fields (i.e. 30 realizations of the synthetic field for each of the 105 natural corrosion fields) using the RSTRENG model. As the values of a_L and a_C vary, the mean ratio varies between 0.968 and 0.985. If both a_L and a_C are greater than one, the mean ratio is almost invariant. The insensitivity of the mean ratio to a_L and a_C suggests that the use of fixed values of a_L and a_C does not have a significant influence on the prediction of the burst capacity of the synthetic field.

Table 4.3 The mean ratio $\overline{p_s}/p_0$ for 3150 synthetic fields under different values of a_L and a_C

a_C	a_L										
	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5
0.5	0.978	0.983	0.984	0.984	0.985	0.985	0.985	0.985	0.985	0.985	0.985
1.0	0.969	0.974	0.974	0.975	0.975	0.975	0.975	0.975	0.975	0.975	0.975
1.5	0.968	0.972	0.973	0.973	0.973	0.973	0.973	0.973	0.974	0.974	0.974
2.0	0.968	0.972	0.973	0.973	0.973	0.973	0.973	0.973	0.973	0.973	0.973
2.5	0.968	0.972	0.973	0.973	0.973	0.973	0.973	0.973	0.973	0.973	0.973

Since we do not have the actual ILI data, the ILI parameters are obtained by performing a virtual ILI process on the laser scan data without considering the measurement

uncertainties associated with ILI in practice. Therefore, the ILI parameters used in the present study are noise free. This is a limitation of the proposed method as ILI data in practice inevitably contain measurement errors. Further studies are therefore needed to investigate to what extent the measurement error in ILI influences the accuracy of the burst capacity predicted using the proposed method. It will be also interesting to compare the degrees to which the measurement error affects the accuracy of the assessments based on the simple cuboid idealization of the corrosion cluster and proposed method. Another limitation of the proposed method is that although the empirical equations proposed in this study (i.e. Eqs. (4.1) – (4.6)) quantify the general trend in the data, they may not adequately capture the characteristics of corrosion clusters with complex morphologies. Therefore, more detailed profiles of corrosion clusters and real ILI data corresponding to the clusters need to be collected in the future to address these limitations.

4.5 Conclusion

We propose a methodology to simulate corrosion fields on the external surface of buried pipelines based on the ILI data. The methodology treats natural corrosion fields as nonhomogeneous non-Gaussian random fields and individual anomalies within the corrosion field as homogeneous non-Gaussian random fields. High-resolution laser scan data for 3408 anomalies within 105 natural corrosion fields identified on buried natural gas transmission pipelines in Canada are used to evaluate the marginal probability distributions and autocorrelation functions of corrosion depths within the anomalies. A virtual ILI process is also employed to generate the ILI data corresponding to the anomalies. Empirical equations are then developed to predict the parameters of the marginal distribution and autocorrelation function of the corrosion depths from the ILI-reported corrosion depth, length and width for a given anomaly.

To simulate the corrosion field given ILI data, the Masters and Gurley's algorithm is employed to generate realizations of homogeneous non-Gaussian fields representing corrosion anomalies. A Gaussian spatial modulating function is then applied to the simulated corrosion anomalies such that each anomaly is localized at the corresponding position reported by ILI. Finally, all the simulated anomalies within the corrosion field

are combined to generate the synthetic corrosion field. The effectiveness of the proposed methodology is demonstrated by comparing the burst capacities of the pipeline corresponding to the actual corrosion field, synthetic corrosion field and ILI-idealized field. The RSTRENG model, which is widely used in the pipeline industry, and elastoplastic FEA are employed to evaluate the burst capacity.

Compared with the commonly used ILI idealization with cuboidal anomalies, the burst capacities of synthetic fields generated using the proposed methodology are less conservative. For the 105 natural corrosion fields considered in the present study, the burst capacity of the synthetic field is on average 97.4% of that of the actual corrosion field, whereas the burst capacity of the ILI-idealized field is on average 91.5% of that of the actual field. For corrosion fields with densely distributed corrosion pits, the burst capacity predicted based on the ILI idealization can be substantially lower than the actual burst capacity; such an issue is largely resolved by using the synthetic field to predict the burst capacity. The proposed methodology provides a practical tool for practitioners to more effectively utilize the ILI data to improve the pipeline corrosion management practice.

References

Al-Amin, M., & Zhou, W. (2014). Evaluating the system reliability of corroding pipelines based on inspection data. *Structure and Infrastructure Engineering*, *10*(9), 1161-1175.

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, *24*(3), 663-676.

Ang, A. H., & Tang, W. H. (2006). Probability Concepts in Engineering Planning: Emphasis on Applications to Civil and Environmental Engineering, John Wiley and Sons.

Aryai, V., & Mahmoodian, M. (2017). Spatial-temporal reliability analysis of corroding cast iron water pipes. *Engineering Failure Analysis*, *82*, 179-189.

Bao, J., & Zhou, W. (2021a). A random field model of external metal-loss corrosion on buried pipelines. *Structural safety*, *91*, 102095.

Bao, J., & Zhou, W. (2021b). Influence of depth thresholds and interaction rules on the burst capacity evaluation of naturally corroded pipelines. Journal of Pipeline Science and Engineering, 1(1), 148-165.

Bao, J., Zhang, S., Zhou, W., & Zhang, S. (2018, September). Evaluation of burst pressure of corroded pipe segments using three-dimensional finite element analyses. In *International pipeline conference* (Vol. 51869, p. V001T03A043). American Society of Mechanical Engineers.

Brennan, D. P., Akpan, U. O., Orisamolu, I. R., Glover, A., & Konuk, I. (2002, January). Probabilistic modeling of corroded pipeline structures. In *International Pipeline Conference* (Vol. 36207, pp. 1811-1820).

EGIG. (2020). *11th Report of the European Gas pipeline Incident data Group*. Retrieved from EGIG website: https://www.egig.eu/reports/$60/$61

Garbatov, Y., & Guedes Soares, C. (2019). Spatial corrosion wastage modeling of steel plates exposed to marine environments. *Journal of Offshore Mechanics and Arctic Engineering*, *141*(3), 031602

Hong, H. P., Cui, X. Z., & Qiao, D. (2021a). An algorithm to simulate nonstationary and non-Gaussian stochastic processes. *Journal of Infrastructure Preservation and Resilience*, *2*(1), 17.

Hong, H. P., Cui, X. Z., & Zhou, W. X. (2021b). A model to simulate multidimensional nonstationary and non-Gaussian fields based on S-transform. *Mechanical Systems and Signal Processing*, *159*, 107789.

Kiefner, J. F., & Vieth, P. H. (1989). *A modified criterion for evaluating the remaining strength of corroded pipe* (No. PR-3-805). Battelle Columbus Div., OH (USA).

Lam, C., & Zhou, W. (2016). Statistical analyses of incidents on onshore gas transmission pipelines based on PHMSA database. *International Journal of Pressure Vessels and Piping*, *145*, 29-40.

Li, F., & Ng, M. K. (2010). Kernel density estimation based multiphase fuzzy region competition method for texture image segmentation. *Communications in Computational Physics*, *8*(3), 623.

Masters, F., & Gurley, K. R. (2003). Non-Gaussian simulation: cumulative distribution function map-based spectral correction. *Journal of engineering mechanics*, *129*(12), 1418-1428.

Ponomarev, A. V. (2020). Systems Analysis of Discrete Two-Dimensional Signal Processing in Fourier Bases. *Advances in Signal Processing: Theories, Algorithms, and System Control*, 87-96.

Sejdić, E., Djurović, I., & Jiang, J. (2009). Time–frequency feature representation using energy concentration: An overview of recent advances. *Digital signal processing*, *19*(1), 153-183.

Siraj, T., & Zhou, W. (2019). Quantification of Measurement Errors in the Lengths of Metal-Loss Corrosion Defects Reported by Inline Inspection Tools. *Journal of Pressure Vessel Technology*, *141*(6), 061402.

Smart, L., Li, Y., Nestleroth, J. B., & Ward, S. (2018, September). Interaction Rule Guidance for Corrosion Features Reported by ILI. In *International Pipeline Conference* (Vol. 51869, p. V001T03A028). American Society of Mechanical Engineers.

Stockwell, R. G., Mansinha, L., & Lowe, R. P. (1996). Localization of the complex spectrum: the S transform. *IEEE transactions on signal processing*, *44*(4), 998-1001.

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, *85*, 1-16.

Vanaei, H. R., Eslami, A., & Egbewande, A. (2017). A review on pipeline corrosion, in-line inspection (ILI), and corrosion growth rate models. *International Journal of Pressure Vessels and Piping*, *149*, 43-54.

Williams, C., & Rasmussen, C. (1995). Gaussian processes for regression. *Advances in neural information processing systems*, 8.

Xie, M., & Tian, Z. (2018). A review on pipeline integrity management utilizing in-line inspection data. *Engineering Failure Analysis*, *92*, 222-239.

Yang, Z., Craddock, R. C., Margulies, D. S., Yan, C. G., & Milham, M. P. (2014). Common intrinsic connectivity states among posteromedial cortex subdivisions: Insights from analysis of temporal dynamics. *Neuroimage*, *93*, 124-137.

Zelmati, D., Bouledroua, O., Ghelloudj, O., Amirat, A., & Djukic, M. B. (2022). A probabilistic approach to estimate the remaining life and reliability of corroded pipelines. *Journal of Natural Gas Science and Engineering*, *99*, 104387.

Zhang, F., Rosenfeld, M., & Gustafson, J. (2018, September). Fitness for service analysis of the circumferential extent of corrosion in pipelines. In *International Pipeline Conference* (Vol. 51869, p. V001T03A030). American Society of Mechanical Engineers.

Zhang, S., Yan, J., Kariyawasam, S., Huang, T., & Al-Amin, M. (2018, September). A more accurate and precise method for large metal loss corrosion assessment. In *International Pipeline Conference* (Vol. 51869, p. V001T03A059). American Society of Mechanical Engineers.

Zhang, S., & Zhou, W. (2020). Assessment of effects of idealized defect shape and width on the burst capacity of corroded pipelines. *Thin-Walled Structures*, *154*, 106806.

Zhang, S., & Zhou, W. (2014). Cost-based optimal maintenance decisions for corroding natural gas pipelines based on stochastic degradation models. *Engineering Structures*, *74*, 74-85.

Zhou, W., Bao, J., Cui, X. Z., & Hong, H. P. (2021). Modeling and simulating nonhomogeneous non-Gaussian corrosion fields on buried pipelines and its use in predicting burst capacities of corroded pipelines. *Engineering Structures*, *245*, 112957.

Zhou, W., & Huang, G. X. (2012). Model error assessments of burst capacity models for corroded pipelines. *International Journal of Pressure Vessels and Piping*, *99*, 1-8.

Zhou, W., Xiang, W., & Hong, H. P. (2017). Sensitivity of system reliability of corroding pipelines to modeling of stochastic growth of corrosion defects. *Reliability Engineering & System Safety*, *167*, 428-438.

Zhu, X. K., & Leis, B. N. (2005). Influence of yield-to-tensile strength ratio on failure assessment of corroded pipelines. *The Journal of Pressure Vessel Technology*, *127*(4), 436-44

Chapter 5

5   Application of Wavelet Transforms to the Simulation of Corrosion Fields on Buried Pipelines

5.1   Introduction

Onshore and offshore pipelines are safe and effective means to transport large quantities of hydrocarbons and are vital to the economic well-being and security of modern society. Historical incident data reveal that metal-loss corrosion is one of the most significant threats to the structural integrity of pipelines. Twenty-seven percent of the incidents on onshore gas transmission pipelines in Europe were caused by corrosion between 2010 and 2019 (EGIG 2020). According to the data collected by the Pipeline and Hazardous Materials Safety Administration (PHMSA) of the United States, about 24% of the incidents on onshore gas transmission pipelines in the US were caused by external corrosion between 2002 and 2013 (Lam and Zhou 2016). Therefore, corrosion assessment and mitigation are essential components of the pipeline integrity management program.

Corrosion causes localized pipe wall thickness loss and thus reduces the burst capacity of the pipeline. Simple-to-use semi-empirical models, such as the well-known ASME B31G (ASME 1991) and RSTRENG models (Kiefner and Vieth 1989), are widely employed in the pipeline industry to evaluate the burst capacity of corroded pipelines. The development and validation of such models are based on the full-scale burst tests of corroded pipe specimens, which are costly and time-consuming to carry out. Many studies have demonstrated the high accuracy of elasto-plastic finite element analysis (FEA) to evaluate the burst capacity of pipe specimens containing artificially induced or naturally occurring corrosion defects (Abdalla et al. 2014; Bao et al. 2018; Cronin 2000; Mok et al. 1991; Zhang and Zhou 2020). It follows that FEA can be employed to carry out full-scale burst tests of corroded pipe specimens numerically, resulting in significant cost savings and improved efficiency. In this regard, it is desirable to simulate synthetic corrosion defects to be incorporated into FEA that capture the inherent probabilistic

properties of naturally-occurring corrosion features observed from the corroded pipe surfaces.

Probabilistic models for corrosion fields have been reported in the literature. Aryai and Mahmoodian (2017) modeled the corrosion depths on cast iron water pipes as homogeneous Gaussian random fields. Garbatov and Soares (2019) characterized the corroded surface of steel bottom plates of ballast tanks in double-hull tankers as homogeneous lognormal random fields. Zhou et al. (2021) considered the corroded external surfaces of buried steel pipelines as nonhomogeneous non-Gaussian fields and employed the discrete orthogonal S-transform (DOST) (Stockwell et al. 1996) to characterize and simulate the corrosion field. Bao and Zhou (2021) considered the intermingling between corroded and corrosion-free areas on the external surface of buried pipelines and used a latent Gaussian field to characterize and simulate the corrosion field. Laser scan data of naturally corroded external surfaces of buried pipelines (Bao and Zhou 2021) indicate that the corresponding corrosion field is nonhomogeneous. This combined with the non-negative nature of the corrosion depth suggests that the corrosion field on buried pipelines is a nonhomogeneous non-Gaussian field.

The wavelet transform is a suitable tool to analyze signals and has been widely used in the civil engineering field. Casavola et al. (2018) detected the corrosion area on bronze plates by performing a one-dimensional (1D) continuous wavelet transform (CWT) of the scan data using the Morlet wavelet and thresholding the fringe intensity calculated from the resulting wavelet coefficients. Ferreira et al. (2021) employed the discrete wavelet transform (DWT) to analyze corrosion fields on buried pipelines and incorporate the coefficients from DWT into a deep neural network to predict the burst capacity of corroded pipelines. Song et al. (2002) and Xu et al. (2009) performed two-dimensional (2D) DWT of scans of corroded metal plates to classify and predict corrosion by calculating the energy and entropy of different components. Roux et al. (2013) proposed a framework to study images (or fields) and estimate self-similarity with rotation and anisotropy parameters based on the hyperbolic wavelet transform (HWT). After screening for the useful information from the magnetic flux leakage signals obtained from the in-line inspection of in-service pipelines, Kathirmani et al. (2012) performed

DWT using the Daubechies wavelets for data compression. Based on the spectra information of the ultrasonic signal obtained from the in-line inspection of pipelines, Song and Que (2006) used the Gaussian wavelet to construct a bank of band-pass filters to suppress the noise in the collected ultrasonic data. Combining features extracted from the statistical techniques and the wavelet transform, Zadkarami et al. (2016) developed a multi-layer perceptron neural network classifier to detect pipeline leakages. Saadatmorad et al. (2022) performed the 1D DWT on the quasi-Pearson-based correlation signal between the mode shapes of intact and damaged steel beams to improve the accuracy and robustness of damage detection.

The wavelet transforms have also been used to simulate random processes and fields. A practical approach for simulation is the inverse DWT using modified or generated wavelet coefficients. Gurley and Kareem (1999) suggested multiplying coefficients from the DWT of nonstationary records by a Gaussian white noise of unit variance to generate signals with time-frequency characteristics similar to those of the parent signals. Otsuka et al. (2018; 2020) applied 2D DWT based on several real-valued wavelet families to the scan records of surfaces machined by the turning or milling process and generated 2D records by replacing nonessential wavelet coefficients with random numbers. Based on the orthonormality of wavelet functions, some researchers (Dijkerman and Mazumdar 1994; Nychka et al. 2002; Spanos and Rao 2001; Zeldin and Spanos 1996) defined the expectation and covariance of wavelet coefficients and generated wavelet coefficients for realizations of nonhomogeneous processes. Chavez and Cazelles (2019) performed CWT using the complex Morlet wavelet to evaluate the time-frequency spectrum, also known as the scalogram, in the wavelet transform of nonstationary series and randomized the phases of wavelet coefficients to simulate surrogates aimed at hypothesis testing. Hong et al. (2021a) proposed an iterative power and amplitude correction (IPAC) algorithm to simulate nonstationary non-Gaussian stochastic processes. The IPAC algorithm incorporates the amplitude or spectrum obtained from a suitable time-frequency or time-scale transform and has been extended to multi-dimensional fields (Zhou et al. 2021; Hong et al. 2021b). However, the use of the IPAC algorithm with highly redundant transforms such as CWT and S-transform can be computationally intensive for multi-dimensional fields.

While DWT is computationally more efficient than CWT, real-valued critically sampled DWT has some drawbacks such as the shift variance (i.e. a small shift in the signal markedly perturbs the wavelet coefficients around singularities) and susceptibility to aliasing (Kingsbury 2001; Selesnick et al. 2005). To overcome the drawbacks of real-valued critically sampled DWT and without introducing too much redundancy, Kingsbury (2021) (see also Selesnick et al. (2005)) proposed the dual-tree complex discrete wavelet transform (DT-CDWT). DT-CDWT employs separable filter banks that are designed to result in the transform with near shift-invariance, perfect reconstruction, and reduced aliasing, while maintaining a relatively low degree of redundancy (the redundancy is four for two-dimensional decomposition as explained in Section 2.2.2) and thus high computational efficiency. In addition to DWT, these DT-CDWT filters could also be extended and used in other transform schemes, such as wavelet packet transform (Bayram and Selesnick 2008) and HWT (Roux et al. 2013; DeVore et al. 1998; Oulhaj et al. 2017). A few developments (Selesnick 2001; Selesnick 2004) closely related to DT-CDWT are also noteworthy. Motivated by the undecimated DWT, Selesnick proposed the double-density DWT (2001) to reduce the shift sensitivity. The double-density DWT uses one scaling function and two distinct wavelet functions to approximate the CWT and is nearly shift invariant. Selesnick (2004) further proposed the so-called double-density dual-tree CDWT by combining the double-density DWT with DT-CDWT. Such a transform has been shown to lead to excellent performance in imaging processing (Chitchian et al. 2012; Adam et al. 2018), although at a markedly higher computational cost than that of DT-CDWT (the redundancy of the former is four times that of the latter). To our best knowledge, the use of redundant transforms such as the CWT and DT-CDWT to simulate 2D nonhomogeneous non-Gaussian random fields has not been reported in the literature.

In the present study, we propose a framework to analyze and simulate nonhomogeneous non-Gaussian corrosion fields on the external surface of buried in-service pipelines by using the discrete and continuous complex wavelet transforms. More specifically, we employ three transforms, namely CWT with the complex Morlet wavelets, DT-CDWT, and the decomposition scheme that is consistent with HWT but using the same filter as that used in DT-CDWT (denoted as DT-CHWT for simplicity), to analyze high-

resolution corrosion measurements obtained from the external surface of buried in-service pipelines. By treating the corrosion depth on the pipe surface as a 2D nonhomogeneous non-Gaussian random field, we apply the IPAC algorithm in conjunction with CWT, DT-CDWT, and DT-CHWT to generate realizations of the corrosion field. The probabilistic characteristics of the simulated corrosion fields are compared with those of measured fields to investigate the suitability of wavelet-based IPAC techniques for simulating corrosion on pipelines. While the present study and Zhou et al.'s study (2021) both employ the IPAC algorithm to simulate corrosion fields on buried pipelines, there are key differences between these two studies. The DOST in conjunction with IPAC is adopted in Zhou et al.'s study (2021). The DOST is a non-redundant transform and computationally efficient; however, its resolution of the space-wavenumber characteristics of the corrosion field is lower than that of the redundant transforms. Therefore, certain features of the natural corrosion field might not be captured in the synthetic fields generated based on the DOST. The present study employs CWT and DT-CDWT, both being redundant transforms, to better capture the characteristics of the corrosion field. In particular, we focus on DT-CDWT (and DT-CHWT) as it is a moderately redundant transform and is designed to capture directional features in two and higher dimensions without a heavy computational burden. While the double-density DT-CDWT is also a viable transform for the purpose of the present study, it will be considered in the future.

In Section 5.2, an example of corrosion field measurement that is used to aid the exposition of the framework is described. The fundamentals of CWT, DT-CDWT, and DT-CHWT, as well as methodologies to incorporate them into IPAC to simulate corrosion fields, are also presented in Section 5.2. Section 5.3 presents corrosion fields generated from the wavelet-based IPAC algorithm, and analyses of the probabilistic characteristics of the simulated fields in comparison with those of the measured field. The performances of IPAC algorithms based on CWT, DT-CDWT, and DT-CHWT are further assessed in terms of the errors in iterations and computational cost in Section 5.3. Concluding remarks are presented in Section 5.4.

5.2   Data and methodology

5.2.1   Scanned corrosion fields

Figure 5.1 depicts the laser-scanned corrosion field with a length (in the pipe longitudinal direction) of 512 mm and a width (in the pipe circumferential direction) of 256 mm on the external surface of a naturally corroded in-service buried pipeline in Canada with an outside diameter (D) of 763.4 mm and a wall thickness (wt) of 8.5 mm.   The measurement is performed on-site using the HandyScan 3D laser scanner after the corroded pipe segment has been excavated, de-coated, and cleaned. The laser scanner measures the corrosion depth on the pipe surface with a resolution of 1 mm in both the longitudinal and circumferential directions, i.e. a grid of 1×1 mm on the pipe surface. The corrosion depth (z) in the unit of mm at a given grid point represents the loss of the pipe wall thickness at this point. Following Bao and Zhou (2021) and Zhou et al. (2021), by considering the measurement error of the laser scan device and the negligible effect of shallow corrosion depths on the pipe burst capacity (Bao and Zhou 2021), grid points with corrosion depths below a threshold depth of 5%wt are assumed to be corrosion free (i.e. corrosion depths set to zero at such points). The empirical cumulative distribution function (CDF) of the corrosion depths and the empirical probability density function (PDF) of the non-zero corrosion depths are shown in Fig. 5.2. Only 23.3% of the grid points in the scanned area have corrosion depths greater than 5%wt; therefore, the area of PDF shown in Fig. 5.2b is 0.233. Figures 5.1 and 5.2 illustrate the nonhomogeneous non-Gaussian characteristics of the corrosion field
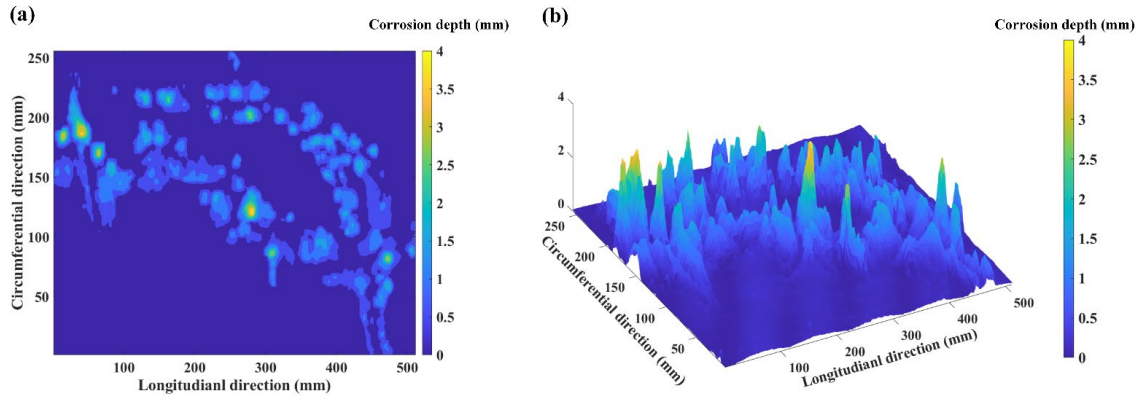
Figure 5.1 Corrosion depths measured by the laser scan on the external surface of an in-service pipeline: (a) two-dimensional plot, and (b) three-dimensional plot
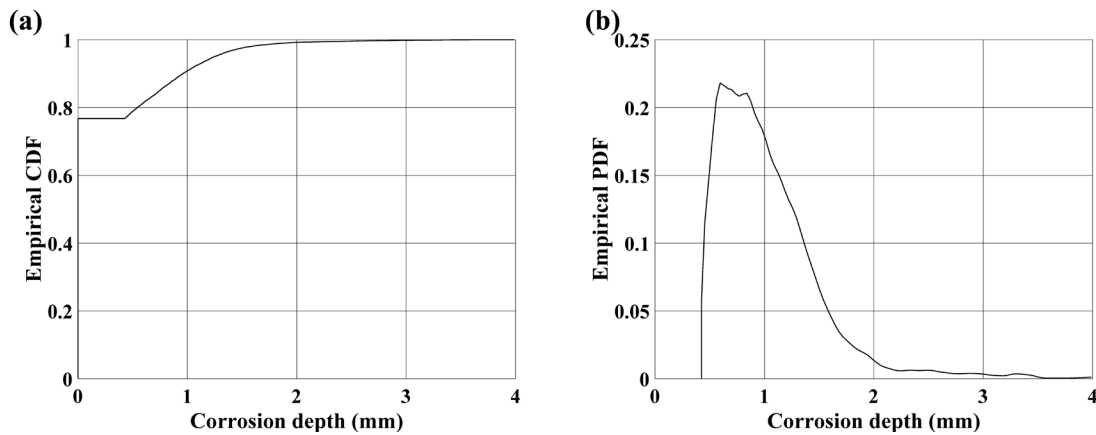
Figure 5.2 Marginal distribution of corrosion depths on the external surface of a naturally corroded buried pipe segment: (a) empirical CDF; and (b) empirical PDF of non-zero corrosion depths

5.2.2   Wavelet analysis

5.2.2.1   Continuous and discrete wavelet transforms for signals in 2D

Let $z(x, y)$ denote a real-valued 2D function indexed on x and y. The coefficient of CWT of $z(x, y)$, $z_w(s_x, \tau_x; s_y, \tau_y)$, is given by (Daubechies 1992):

$$z_w\left(s_x,\tau_x;s_y,\tau_y\right)=\frac{1}{\sqrt{|s_xs_y|}}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}z(x,y)\psi^*\left(\frac{x-\tau_x}{s_x},\frac{y-\tau_y}{s_y}\right)dxdy \tag{5.1}$$

where s_x and s_y are the scaling factors along the x and y directions, respectively; τ_x and τ_y are the translations along the x and y directions, respectively; $\psi(x, y)$ is the so-called mother wavelet function, and * denotes the complex conjugate. If $\psi(x, y)$ satisfies the admissibility condition, the function $z(x, y)$ can be reconstructed from $z_w(s_x, \tau_x; s_y, \tau_y)$ by the inverse CWT.

Many different mother wavelet functions, which can be real- or complex-valued, are available for CWT (Kankar et al. 2011). The Morlet wavelet has been extensively used in feature extraction, textual analysis, pattern recognition, anomaly detection, and location (Lin and Qu 2000; El Hassani et al. 2012; Lin and Lu 2010; Huang and Hsieh 1999). The complex Morlet wavelet function is selected in the present study for its robustness, sensitivity to directional features of the 2D signal, and computational efficiency achieved by incorporating the fast Fourier transform in coding (Chopra and Marfurt 2015; Antoine et al. 1996; Cohen 2019). Specifically, the 2D complex Morlet function with the central wavenumber $3/\pi$ and smallest scale $s_{min} = 2.6$ is employed. Details of the 2D complex Morlet function are given in (Lilly and Olhede 2009; Lee 1996; Farge 1992; Torrence and Compo 1998).

It follows from Eq. (5.1) that the coefficient of the (2D) CWT is defined by four variables: two translations (τ_x, τ_y) and two scales (s_x, s_y). In CWT, all these variables are continuous, and the quantity $\left|z_w(s_x,\tau_x;s_y,\tau_y)\right|^2$ is related to the energy density of the field $z(x, y)$ at the wavenumber $(f_0/s_x, f_0/s_y)$ and translation (τ_x, τ_y) (Mallat 1989). Although CWT has an excellent resolution for data analysis, its computational cost could be very high especially if the repeated use of CWT and inverse CWT for the 2D field is required.

DWT (Daubechies 1992; Mallat 1989) markedly reduces the computational burden as compared to CWT. DWT that uses Mallat's tree algorithm is often referred to as the standard DWT with a good balance between resolution and computational cost. If the standard DWT employs real-valued filters, it is non-redundant. Given a signal in 2D, the

2D DWT produces different sub-records that emphasize different directional features. The standard 2D DWT of a signal in 2D, $z(x, y)$ ($x = 0, 1, 2, \ldots, 2^n - 1$; $y = 0, 1, 2, \ldots, 2^m - 1$), $z_{\psi}^{(\alpha)}(i, \tau_x, \tau_y)$, is given by (Daubechies 1992):

$$z_{\psi}^{(\alpha)}(i, \tau_x, \tau_y) = \frac{1}{2^i} \sum_x \sum_y z(x, y) \psi_{(\alpha)} \left(\frac{x - \tau_x}{2^i}, \frac{y - \tau_y}{2^i} \right) \tag{5.2}$$

$$\psi_{(\alpha)}(x, y) = \begin{cases} \varphi(x)\varphi(y), \alpha = LL \\ \varphi(x)\psi(y), \alpha = LH \\ \psi(x)\varphi(y), \alpha = HL \\ \psi(x)\psi(y), \alpha = HH \end{cases} \tag{5.3}$$

where i is the level of 2D DWT; $\psi_{(\alpha)}(x, y)$ denotes the wavelet function constructed from the tensor product of different filter combinations; $\varphi(\cdot)$ and $\psi(\cdot)$ are the low-pass and high-pass filters, respectively; LL, LH, HL, and HH denote the combinations of filters as illustrated in Fig. 5.3: i.e. approximation (low-pass and low-pass), horizontal features (low-pass and high-pass), vertical features (high-pass and low-pass) and diagonal features (high-pass and high-pass), respectively. The DWT coefficients can be arranged in a matrix of the same size as the original 2D record. $z(x, y)$ can be reconstructed based on $z_{\psi}^{(\alpha)}(i, \tau_x, \tau_y)$ by the inverse DWT if the wavelet filters satisfy the perfect reconstruction. Details of the forward and inverse DWT are given in (Daubechies 1992; Mallat 1989; Mallat 1999).
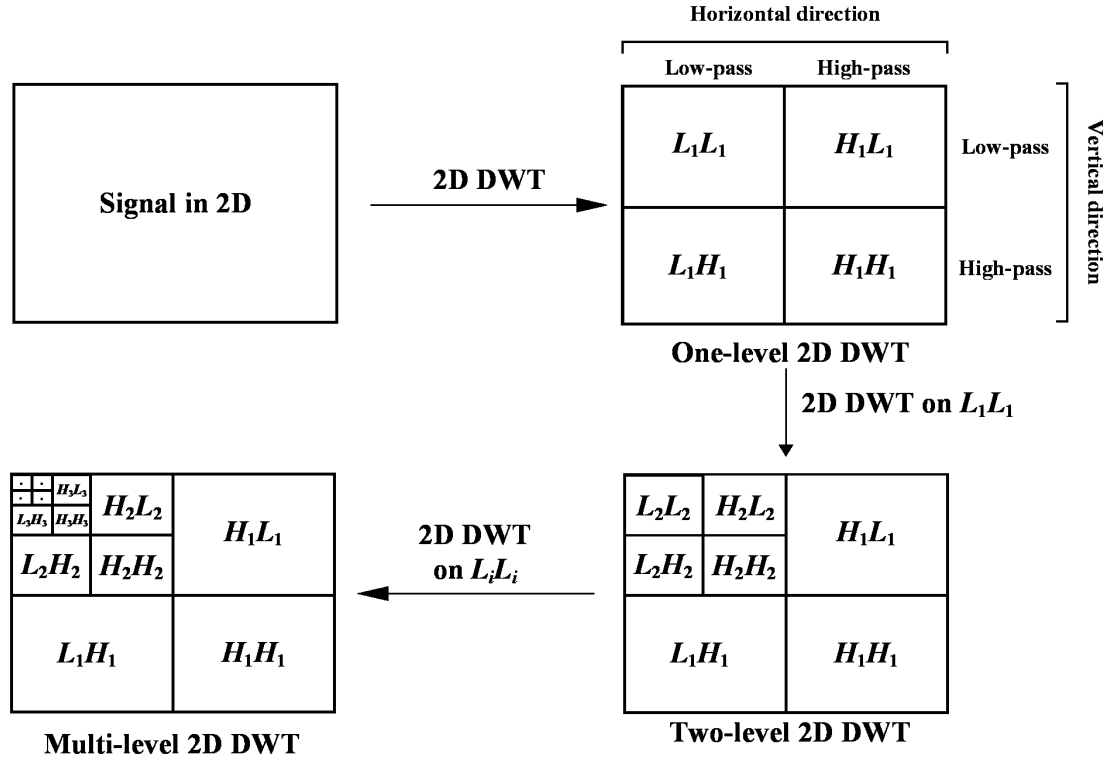
Figure 5.3 Standard 2D DWT of a signal in 2D

A signal in multi-dimension can be decomposed in different ways based on different definitions of the wavelet function and decomposition schemes used in the multi-dimensional DWT. In the standard 2D DWT, the 2D DWT bases $\psi_{(\alpha)}(x,y)$ are constructed from the scaling and wavelet functions with identical scales at both dimensions. If multi-dimensional wavelet bases, such as the bases used in HWT, are constructed by functions with independent scales or dilation factors along different dimensions, this forms the anisotropic wavelet bases that are more flexible in texture analysis (Roux et al. 2013; DeVore et al. 1998). The coefficients of HWT of a signal in 2D are given by (Farouj et al. 2016):

$$z_\psi^{(\beta)}(i_x, i_y, \tau_x, \tau_y) = \frac{1}{2^{\frac{i_x+i_y}{2}}} \sum_x \sum_y z(x,y)\psi_{A,(\alpha)}\left(\frac{x-\tau_x}{2^{i_x}}, \frac{y-\tau_y}{2^{i_y}}\right) \tag{5.4}$$

$$\psi_{A,(\alpha)}(x,y) = \begin{cases} \varphi(x)\varphi(y), \alpha = LL \\ \varphi(x)\psi(y), \alpha = LH \\ \psi(x)\varphi(y), \alpha = HL \\ \psi(x)\psi(y), \alpha = HH \end{cases}$$

(5.5)

where i_x and i_y are the decomposition level along the x and y directions, respectively, and $\psi_{A,(\alpha)}(x,y)$ denotes the wavelet function for different combinations of filters but possibly include the anisotropic effect. Because the HWT coefficients can be calculated from H_iL_i or L_iH_i from 2D DWT, HWT can also be considered as an extension of DWT that provides anisotropic analysis. HWT has higher resolutions in the wavenumber domain but lower resolutions in the spatial domain than the standard 2D DWT. $z(x, y)$ can be reconstructed based on $z_\psi^{(\beta)}(i_x, i_y, \tau_x, \tau_y)$ using the inverse HWT if the wavelet filters satisfy the perfect reconstruction.

In this study, the HWT scheme is performed as illustrated in Fig. 5.4. Similar to the standard 2D DWT, the HWT coefficients can also be arranged in a matrix of the same size as the original 2D record. In Fig. 4, $H_{i_x}H_{i_y}$ represents the coefficients from the 2D wavelet function $\psi_{i_x,i_y}(x,y) = 2^{-\frac{i_x+i_y}{2}} \psi(2^{-i_x}x)\psi(2^{-i_y}y)$ constructed by the i_x^{th} ($i_x = 1, 2, \ldots,$ n) level horizontal wavelet function and the i_y^{th} ($i_y = 1, 2, \ldots, m$) level vertical wavelet function, and L_n denotes that the scaling function $\varphi(\cdot)$ at the nth level is employed instead of the wavelet function $\psi(\cdot)$. Details of the forward and inverse HWT are given in (Roux et al. 2013; DeVore et al. 1998).
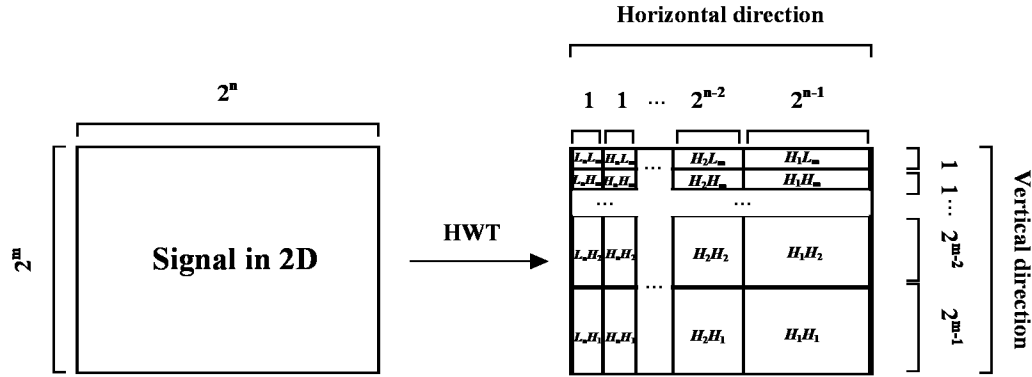
Figure 5.4 The HWT scheme used in this study

5.2.2.2   Dual-tree complex discrete wavelet transform

As mentioned in the introduction, the real-valued DWT has some drawbacks such as the shift variance and susceptibility to aliasing. To overcome these drawbacks, Kingsbury (2001) proposed DT-CDWT. In DT-CDWT, a real-valued 1D record $v(x)$ is represented by complex wavelet coefficients obtained from two wavelet trees (i.e. the real and imaginary trees). The wavelet function $\psi_c(x)$ used in the DT-CDWT can be expressed as follows:

$$\psi_c(x) = \psi_{re}(x) + j\psi_{im}(x) \tag{5.6}$$

where $j = \sqrt{-1}$ is the imaginary unit; $\psi_{re}(x)$ is an even function for the real tree; $\psi_{im}(x)$ is an odd function for the imaginary tree, and $\psi_{re}(x)$ and $\psi_{im}(x)$ are an approximate Hilbert transform pair (Selesnick et al. 2005). Both $\psi_{re}(x)$ and $\psi_{im}(x)$ are real-valued wavelet functions. When applying DT-CDWT to a real-valued 1D record $v(x)$ with length 2^n, one can obtain 2^n real-valued coefficients c_{re} from the real tree using $\psi_{re}(x)$, and 2^n real-valued coefficients c_{im} from the imaginary tree using $\psi_{im}(x)$. By assigning $c = c_{re} + j \cdot c_{im}$, $v(x)$ can be represented by 2^n complex coefficients c. It follows that the redundancy of DT-CDWT for 1D records is two. Since DT-CDWT filters satisfy the prefect reconstruction condition, $v(x)$ can be reconstructed from c.

If DT-CDWT is applied to a 2D field, two complex matrices, c_1 and c_2, of the same size as the original 2D record are obtained. The real and imaginary parts of c_1 consist of coefficients obtained using $\psi_{1,re} = \psi_{re}(x)\psi_{re}(y) - \psi_{im}(x)\psi_{im}(y)$ and $\psi_{1,im} = \psi_{re}(x)\psi_{im}(y) + \psi_{im}(x)\psi_{re}(y)$, respectively. The real and imaginary parts of c_2 consist of coefficients obtained using $\psi_{2,re} = \psi_{re}(x)\psi_{re}(y) + \psi_{im}(x)\psi_{im}(y)$ and $\psi_{2,im} = \psi_{im}(x)\psi_{re}(y) - \psi_{re}(x)\psi_{im}(y)$, respectively. These 2D dual-tree wavelet functions are designed to better distinguish positive and negative spectral features in the wavenumber domain, thus avoiding ambiguity in the spatial domain. Since wavelet coefficients of the 2D decomposition using DT-CDWT can be considered a combination of wavelet coefficients using four sets of real-valued wavelet functions, the 2D record is obtained by summing reconstructions from four sets of real-valued coefficients. The redundancy of 2D decomposition using DT-CDWT is four. The decomposition scheme used for DT-CDWT is generally carried out using that shown in Fig. 5.3. More implementation details of DT-CDWT can be found in Selesnick et al. (2005).

In the present study, the use of the decomposition scheme shown in Fig. 4 but with the same filters is also considered. For simplicity of reference, in such a case, it is referred to as DT-CHWT. Filters used in DT-CDWT and DT-CHWT are summarized in Appendix C.

5.2.2.3   IPAC-based simulation of corrosion fields

To generate realizations of a corrosion field $z(x, y)$ with the prescribed amplitude of the transform coefficient or energy distribution (i.e., power spectral density function) and marginal CDF, the IPAC algorithm proposed by Hong et al. (2021a) is employed in this study. This algorithm has been extended to sample random fields (Zhou et al. 2021) based on the S-transform (Stockwell et al. 1996). In the present study, the amplitude (or the scalogram) obtained from the three wavelet transforms mentioned in the previous section is employed as the prescribed condition in the IPAC algorithm to simulate corrosion fields.

Applying the CWT using the complex Morlet wavelet, DT-CDWT or DT-CHWT to a 2D real-valued record produces complex coefficients. Let $Tf(\bullet)$ denote a forward transform, and $ITf(\bullet)$ denote the corresponding inverse transform. In this study, the transform pair $(Tf(\bullet), ITf(\bullet))$ could be (CWT, inverse CWT), (DT-CDWT, inverse DT-CDWT) and (DT-CHWT, inverse DT-CHWT). The wavelet scalogram based on the selected transform pair is defined as $|z_t|^2 = |Tf(z(x, y))|^2$ (Mallat 1999), which represents the intensity of local waves defined by the corresponding wavelet functions, and $|z_t| = |Tf(z(x, y))|$ denotes the amplitude of the transform coefficient.

To simulate a field with a size $N \times M$ (i.e., a field with N point and a sampling interval Δx along the x-axis and with M point and a sampling interval Δy along the y-axis), where N and M are powers of two, the wavelet-based IPAC procedure consists of the following steps.

1) Generate a sample with size $N \times M$ from the prescribed target marginal distribution and rearrange the sample into a sequence ζ of ascending order.

2) Select a transform pair $(Tf(\bullet), ITf(\bullet))$ from the above-indicated three options.

3) Sample an $N \times M$ white noise, $w(x, y)$. Apply the selected forward transform on $w(x, y)$ and obtained complex wavelet coefficients $w_f = Tf(w(x, y))$. The phase information $\theta = \arg(w_f)$.

4) Calculate the power corrected field $z_{pc}(x, y) = ITf(|z_t| \cdot e^{j\theta})$ based on the prescribed amplitude of transform coefficient $|z_t|$ (or scalogram $|z_t|^2$). Find the rank of $z_{pc}(x, y)$ in the ascending order, denoted as $r(x, y)$, for all x and y.

5) Assign the amplitude corrected field (i.e. distribution matching) $z_{ac}(x, y) = \zeta(r(x, y))$, and calculate the phase information $\theta = \arg(Tf(z_{ac}))$.

6) Repeat steps 4) and 5) until the convergence criterion is satisfied.

We first consider the sequence ζ to be obtained from the measured field directly, which means all the simulated fields are surrogates with identical probability levels. The

convergence is tracked based on $\varepsilon_r = \sum_x \sum_y \left(z_{ac}^{(r)} - z_{ac}^{(r-1)} \right)^2 / \sum_x \sum_y \left(z_{ac}^{(r-1)} \right)^2$, where $z_{ac}^{(r)}$ denotes z_{ac} of the rth ($r \geq 2$) iteration. For the numerical analysis carried out in the present study, $\varepsilon < 0.1\%$ is employed as the convergence criterion. Note that the above IPAC procedure ensures that the generated samples have the same scalogram that is identical to the target scalogram. Alternatively, one may argue that each sample should have its own scalogram with the average scalogram of all the samples matching the target scalogram (Dolan and Spano 2001). In this case, the IPAC procedure can be slightly revised by including a digital filter in Steps 1) and 4). That is, in Step 1), a homogeneous Gaussian noise field with the power spectral density equal to unity, $w_G(x, y)$, is sampled. The simulation of the $w_G(x, y)$ is carried out by sampling independent Gaussian distributed random variables at all grid points with zero mean and variance equal to $1/(\Delta x \Delta y)$, where Δx and Δy are the sampling intervals. In Step 4), $|z_t|$ is then replaced by the amplitude that includes the effect of the digital filter through $w_G(x, y)$. The amplitude of $w_G(x,y)$ is $z_{GTf} = Tf(w_G(x, y))$. The equation $z_{pc}(x, y) = ITf(|z_t| \cdot e^{j\theta})$ in step 4) is then replaced by $z_{pc}(x, y) = ITf(|z_t| \cdot |z_{GTf}| \cdot e^{j\theta})$. Figure 5.5 presents a flowchart of the steps of the IPAC algorithm.
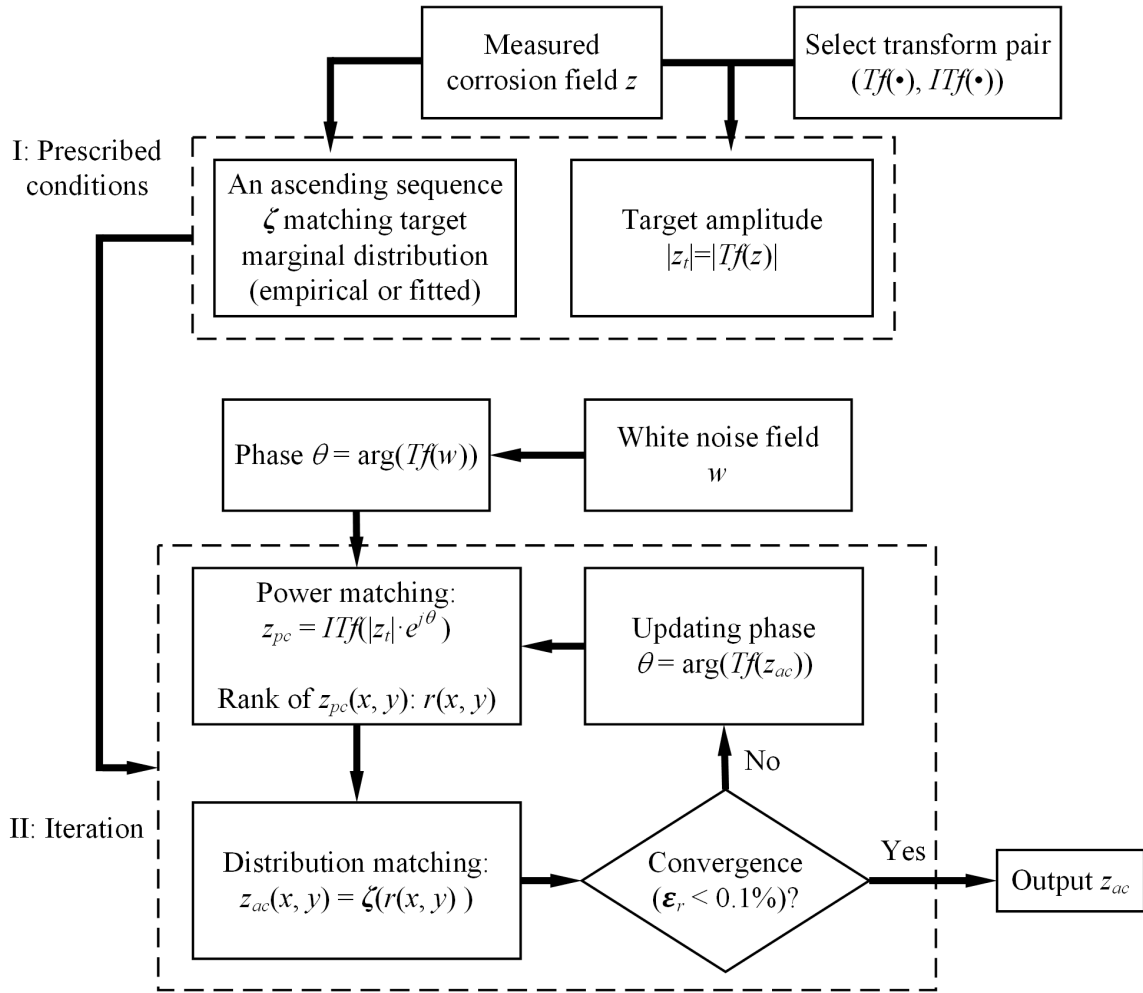
Figure 5.5 Flowchart for the IPAC algorithm to simulate corrosion fields

5.2.3 Metrics adopted for comparing simulated and measured fields

Since the IPAC algorithm shown in Step 5) uses the inverse CDF to assign samples $z_{ac}(x, y)$, the marginal distribution of the simulated corrosion field given by $z_{ac}(x, y)$ matches the prescribed CDF that is assigned based on the measured field. Therefore, the simulated and measured fields are compared in terms of the amplitude or scalogram obtained from the selected transform pair, texture features extracted from the gray level co-occurrence matrix (GLCM) (Hall-Beyer 2017), and the burst capacity (P_b) of the pipeline containing the corrosion field. Let $\varepsilon_{r(i)}$ denotes ε_r of the ith realization,

performances of different transform pairs are also assessed by the mean error $\overline{\varepsilon}_r = \sum_{i=1}^{100} \varepsilon_{r(i)} / 100$ at each iteration step and the total computation time.

In CWT, the scalogram $|z_t|^2 = |z_w(s_x, \tau_x; s_y, \tau_y)|^2$ is a four-dimensional function used in both the IPAC procedure and result comparison. In DT-CDWT and DT-CHWT, the wavelet coefficients can be arranged into two complex matrices c_1 and c_2, and $|z_t|^2 = |c_1|^2 + |c_2|^2$ is used for the scalogram comparison. After specifying the transform pair, 100 realizations are generated based on the selected transform pair and the corresponding target scalogram. The mean scalogram $|\overline{z}_t|^2$ and the standard deviation of the scalograms from the 100 realizations, and the correlation between the target scalogram $|z_t|^2$ and the mean scalogram $|\overline{z}_t|^2$ are used to assess the performance of the selected transform pair.

The GLCM of a 2D field is a square matrix with the number of rows (columns) equal to the number of gray levels in the 2D record. The matrix element $P(q, l|\Delta x, \Delta y)$ ($q, l = 1, 2, \ldots$) is the joint probability function of the gray levels of a pair of grid points separated by the distance $(\Delta x, \Delta y)$, with q and l denoting the gray levels of the two grid points, respectively. One hundred gray levels are considered in this study: the gray level G at the grid point (x, y) is defined as $G(x, y) = 100\lceil z(x, y)/wt \rceil$, where $\lceil \bullet \rceil$ represents rounding up to the nearest positive integer that is greater than $\bullet$. Three grid separations are considered, i.e. $(\Delta x, \Delta y) = (1, 0)$, $(0, 1)$, and $(1, 1)$, which correspond to horizontally (longitudinally), vertically (circumferentially), and diagonally neighbouring pairs of grid points, respectively. Five textural features, including the normalized Shannon entropy (*NSE*), angular second moment (*ASM*), normalized contrast ratio (*NCR*), correlation (*COR*), and inverse difference moment (*IDM*), are extracted from GLCM [46]. *NSE* quantifies the disorder of the record, while *ASM* represents the orderliness. *NCR* quantifies the gray level difference between the separated grid points, while *IDM* quantifies the homogeneity of the record. *COR* denotes the predictable and linear relationship between the two grid points. If a 2D record is highly irregular, this record has high values of *NSE* and *NCR*, and low values of *ASM, COR*, and *IDM*. Details of the evaluation of these features are given in Appendix D.

The well-known RSTRENG model is used to evaluate P_b as follows (Kiefner and Vieth 1989):

$$P_b = \frac{2 \cdot wt \cdot \sigma_f}{D} \frac{1 - \dfrac{A_e}{l_e \cdot wt}}{1 - \dfrac{A_e}{M_e \cdot l_e \cdot wt}} \tag{5.7}$$

$$M_e = \begin{cases} \sqrt{1 + 0.6275 \dfrac{l_e^2}{D \cdot wt} - 0.003375 \left(\dfrac{l_e^2}{D \cdot wt} \right)^2}, & \dfrac{l_e^2}{D \cdot wt} \le 50 \\ 3.3 + 0.032 \dfrac{l_e^2}{D \cdot wt}, & \dfrac{l_e^2}{D \cdot wt} > 50 \end{cases} \tag{5.8}$$

where σ_f is the flow stress of the pipe steel and defined as $\sigma_y + 69$ (MPa) with σ_y being the yield strength of the pipe steel; A_e and l_e are the effective area and length, respectively, of the 2D longitudinal profile of the 3D corrosion field, which is generated by projecting the corrosion field onto a longitudinal plane that is perpendicular to the pipe wall thickness, and M_e is the so-called Folias factor corresponding to the effective length of the corrosion field. The procedure to determine A_e and l_e is not presented for the sake of brevity as it has been well described in many references (Kiefner and Vieth 1989; Zhang et al. 2018).

5.3   Results and discussion

The wavelet coefficients are calculated by applying the CWT, DT-CDWT, and DT-CHWT to the measured corrosion field shown in Fig. 5.1. The scalograms constructed from these coefficients are illustrated in Figs. 5.6 and 5.7. Due to the difficulty in visualizing the four-dimensional CWT scalogram, Fig. 5.6 only includes the CWT scalogram for specific scaling factors. In CWT, the quantity $\left| z_w(s_x, \tau_x; s_y, \tau_y) \right|^2$ is related to the energy density of the field $z(x, y)$ at the wavenumber (f_0/s_x, f_0/s_y) and translation (τ_x, τ_y) domain. In DT-CDWT and DT-CHWT, the energy is defined as the square of the amplitude of the transform coefficient as discussed in the previous section. As observed from Figs. 5.6 and 5.7, most of the energy is concentrated at large scales (low

wavenumbers), and the energy is not homogeneously distributed in the spatial domain. The results shown in the figure indicate that over 90% of the total energy is concentrated in 0.4% of the amplitude square of the transform coefficients at long scales.

	$s_x = s_{x,min}$	$s_x/s_{x,min} = 32$
$s_y = s_{y,min}$		
$\dfrac{s_y}{s_{y,min}} = 32$		
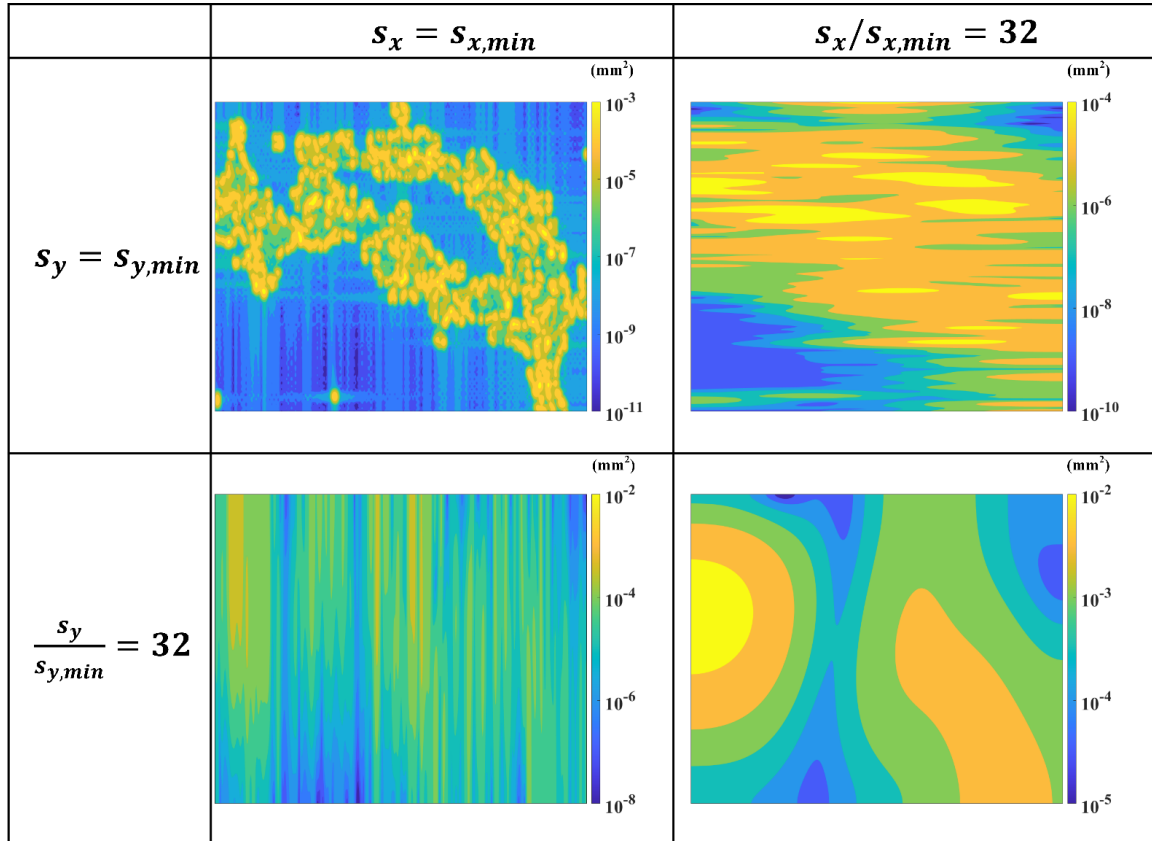
Figure 5.6 CWT scalogram of the measured corrosion field
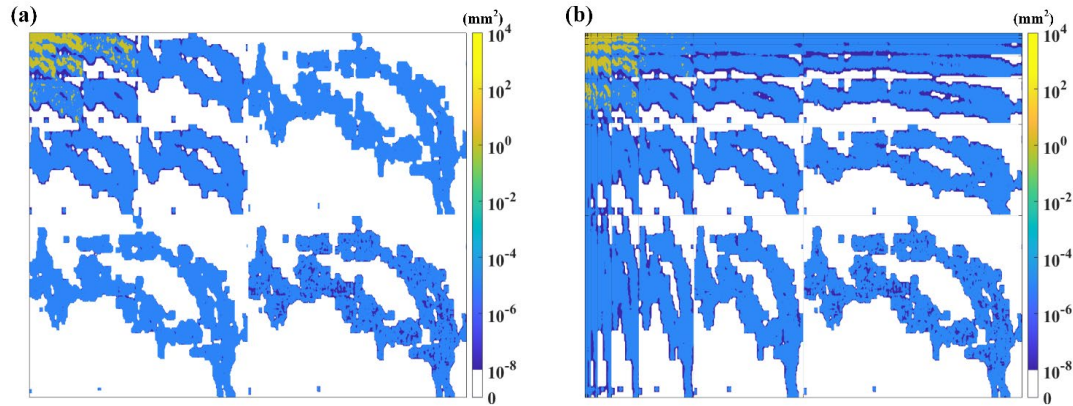
Figure 5.7 Scalogram of the measured corrosion field defined according to different transforms: (a) DT-CDWT; and (b) DT-CHWT

The prescribed marginal distribution includes a 76.7% probability mass of zero corrosion depth (see Section 5.2.1) and a 23.3% weight of the non-zero depth distribution as depicted in Fig. 5.2b, and the square root of the scalograms shown in Figs. 5.6 and 5.7 are employed as the prescribed amplitude of the transform in the IPAC algorithm. The simulation is carried out by using a server with 2 Intel(R) Xeon Gold 5218R CPU @2.10 GHz (20 core and 40 threads) and 768 GB RAM, where the computing time will be reported shortly. Figure 5.8 depicts four samples of the simulated corrosion field using each of the three transform pairs. The mean and standard deviation of the scalograms of the 100 simulated fields based on CWT, DT-CDWT, and DT-CHWT are shown in Figs. 5.9, 5.10, and 5.11, respectively.
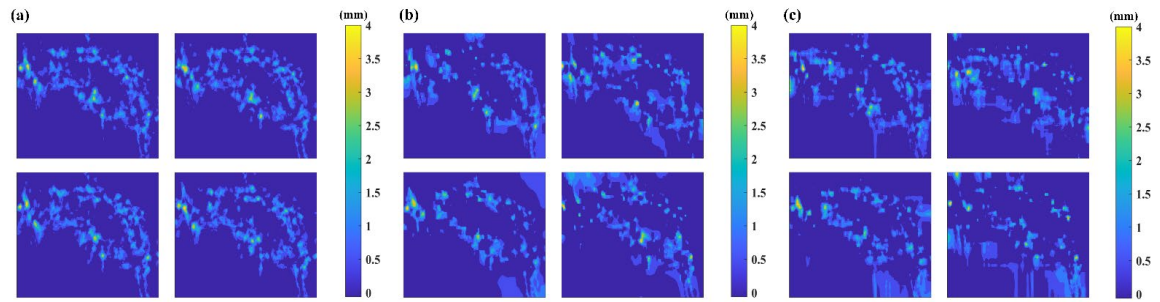
Figure 5.8 Simulated corrosion fields based on: (a) CWT; (b) DT-CDWT; and (c) DT-CHWT
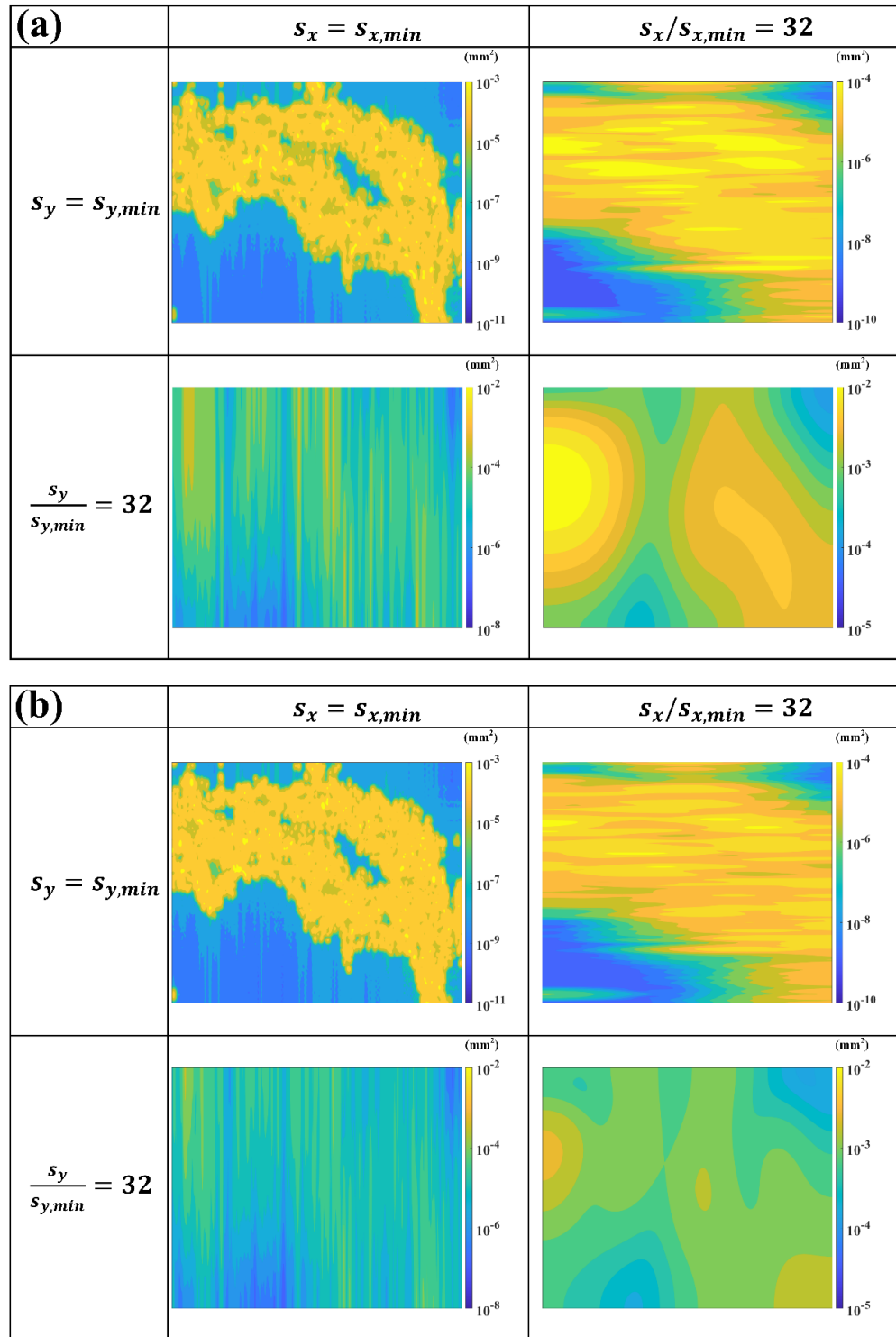
Figure 5.9 Scalograms of simulated corrosion fields based on CWT: (a) mean scalogram of 100 realizations; (b) standard deviation of scalograms of 100 realizations
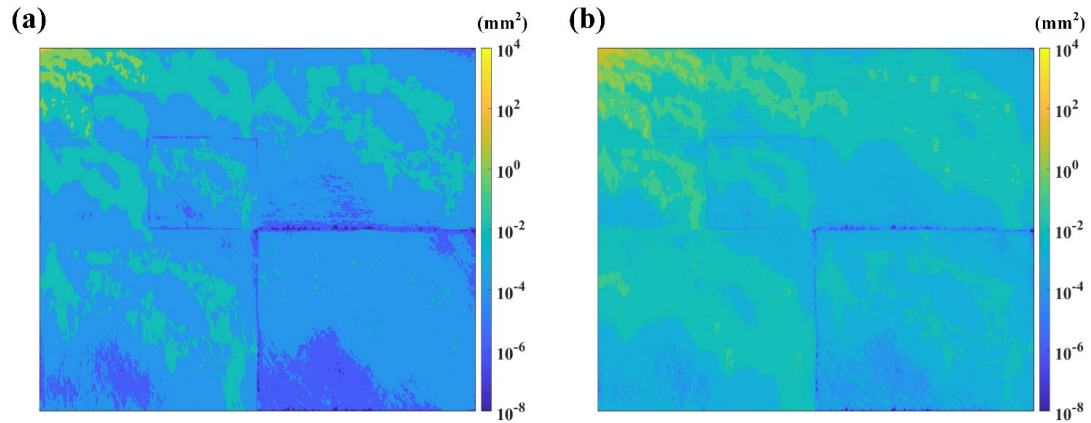
Figure 5.10 Scalograms of simulated corrosion fields based on DT-CDWT: (a) mean scalogram of 100 realizations; (b) standard deviation of scalograms of 100 realizations
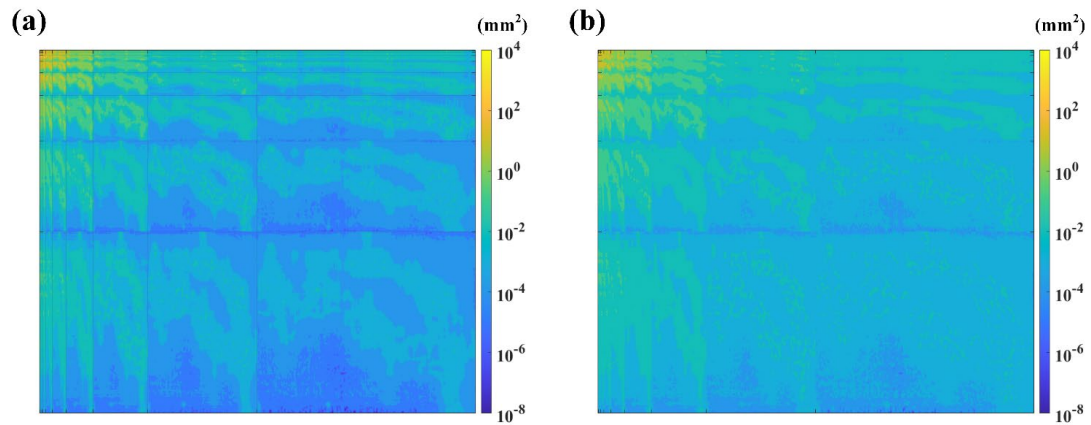
Figure 5.11 Scalograms of simulated corrosion fields based on DT-CHWT: (a) mean scalogram of 100 realizations; (b) standard deviation of scalograms of 100 realizations

Inspection of the results presented in Figs. 5.8 to 5.11 indicates that simulated corrosion fields are similar to the measured corrosion field. The spatial information, such as shallow corrosion areas and locations of deep corrosion pits, is well preserved in the realizations and their scalograms. Compared with DT-CHWT, realizations from DT-

CDWT appear more similar to the measured field. The wavelet bases used in DT-CDWT have a higher spatial resolution in the spatial domain than DT-CHWT, resulting in DT-CDWT better representing the spatial features of the signal.

The similarity between the mean of the scalograms and target scalograms is reflected by the high correlation coefficient between them, which is greater than 0.99 corresponding to the CWT, DT-CDWT, and DT-CHWT. This indicates that the simulated field matches the target scalogram on average. Although some standard deviations of scalograms of the realizations are as high as 300 mm^2, these high standard deviations are usually found in part of the scalograms with high mean values. As mentioned before, over 90% of the total energy is concentrated in 0.4% of the amplitude square of the transform coefficients at long scales in DT-CDWT and DT-CHWT. The coefficient of variations (COV) of the scalograms at the positions and scales corresponding to these coefficients is less than 50%.

The results of the texture analysis based on GLCM are summarized in Table 5.1. The texture characteristics of the realizations from DT-CDWT are closest to those of the measured field, but the texture characteristics associated with the realizations from different methods are only marginally different. The mean values of *NCR* corresponding to the simulated fields are about 35% higher than that of the measured field, indicating that simulated fields tend to have higher energy at high wavenumbers (short scales in the wavelet transform). However, such a discrepancy has little impact on the similarity between the simulated and measured fields because the energy of the field is predominantly concentrated in wavelet coefficients corresponding to long scales (i.e. low wavenumbers).

Table 5.1 Mean texture features of simulated and measured corrosion fields

$(\Delta x, \Delta y)$	Corrosion samples	NSE $(\times 10^{-1})$	ASM $(\times 10^{-1})$	NCR $(\times 10^{-4})$	COR $(\times 10^{-1})$	IDM $(\times 10^{-1})$
(1, 0)	Measured field	1.792	5.718	1.318	9.778	9.098
	CWT	1.864	5.663	1.574	9.734	8.773
	DT-CDWT	1.796	5.750	1.757	9.703	9.066
	DT-CHWT	1.852	5.701	1.756	9.703	8.870
(0, 1)	Measured field	1.840	5.676	1.636	9.724	8.978
	CWT	1.908	5.621	1.913	9.678	8.645
	DT-CDWT	1.837	5.718	2.189	9.631	8.967
	DT-CHWT	1.890	5.666	2.173	9.633	8.782
(1, 1)	Measured field	1.965	5.570	2.717	9.543	8.597
	CWT	2.003	5.526	3.032	9.490	8.381
	DT-CDWT	1.930	5.632	3.570	9.398	8.738
	DT-CHWT	1.992	5.562	3.502	9.408	8.491

RSTRENG is employed to evaluate the burst capacity of the pipeline corresponding to the measured corrosion field as well as each of the simulated corrosion fields by considering $D = 763.4$ mm, $wt = 8.5$ mm, and the yield strength of the pipe steel ξ_y equal to 535 MPa. The PDFs of the capacities corresponding to the corrosion fields generated using different wavelet transforms are compared in Fig. 5.12. The mean burst capacities of the 100 realizations corresponding to the CWT, DT-CDWT, and DT-CHWT are 11.17, 11.10, and 11.11 MPa, respectively, all very close to the burst capacity corresponding to the measured corrosion field (11.25 MPa). The values of COV of all three distributions are less than 0.6%, indicating that the variability of the burst capacities corresponding to the simulated corrosion fields is low. The close mean values and small variability imply

that the scalograms considered in the present study capture the specific characteristics of the measured corrosion field that govern the burst capacity.
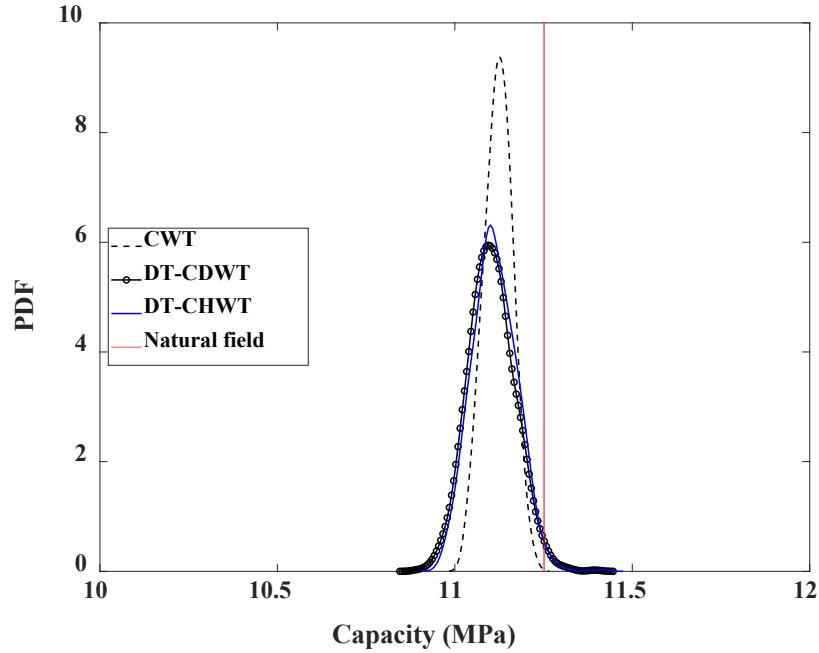
Figure 5.12 Empirical PDF of the burst capacities of the pipe segment containing realizations of the simulated corrosion fields

The convergence rates in the IPAC iterations corresponding to different transform pairs are compared in Table 5.2 in terms of the mean error $\bar{\varepsilon}_r$ at the rth iteration step. As shown in Table 5.2, the difference between subsequent iterations is negligible ($< 1\%$) after only five iterations. The average numbers of iteration steps required to achieve $\varepsilon <$ 0.1% are 10, 18, and 17 corresponding to CWT, DT-CDWT, and DT-CHWT, respectively. Although it can be inferred from Table 5.2 that the application of the IPAC algorithm with CWT converges within fewer iterations than that with DT-CDWT and DT-CHWT, the computing time needed for CWT is much longer than the other two transforms. Table 5.3 compares the CPU time of the IPAC iterations corresponding to different transform pairs. The table clearly demonstrates the computational efficiency of DT-CDWT and DT-CHWT as compared with CWT.

Table 5.2 Mean error at different iteration steps corresponding to different transform pairs

Transform	$\overline{\varepsilon}_5$	$\overline{\varepsilon}_{10}$	$\overline{\varepsilon}_{20}$	$\overline{\varepsilon}_{30}$
CWT	0.33%	0.10%	0.04%	0.03%
DT-CDWT	0.77%	0.23%	0.07%	0.04%
DT-CHWT	0.66%	0.18%	0.06%	0.03%

Table 5.3 CPU time of simulation corresponding to different transform pairs

Transform	5 iterative steps		30 iterative steps	
	1 sample	100 samples	1 sample	100 samples
CWT	2144 s	≈18 h	12475 s	>150 h
DT-CDWT	0.6 s	6 s	3 s	31 s
DT-CHWT	1 s	11 s	5 s	77 s

Besides the computation time, the high computational cost of CWT is also reflected in the high requirement of the computer hardware. The minimum RAM required to generate samples of 256×512 grid points with CWT is about 130 GB. The required RAM will increase exponentially to simulate a larger field. Although CWT has better spatial and wavenumber resolution than DT-CDWT and DT-CHWT, the extremely high redundancy and computational cost severely could limit its application. Discussions in the previous sections suggest that high space-wavenumber resolution is generally unnecessary for the simulation of corrosion fields on pipelines. It follows that DT-CDWT and DT-CHWT are well suited to simulate synthetic corrosion fields on pipelines whereas CWT is not a practically viable option. For completeness, additional analysis

was carried out by using three additional scanned corrosion fields as the target fields. The obtained results are summarized in Appendix E. It is also worth mentioning that although the burst capacities corresponding to the simulated fields are estimated by RSTRENG in the present study for computational efficiency, FEA can be employed to evaluate the burst capacity with higher accuracy, as shown in Zhou et al (2021).

Modifications of the input to the IPAC algorithm can be introduced for specific purposes. For example, to reduce the amount of IPAC input data, we can generate the sequence ζ based on a fitted distribution of corrosion depths. Besides, to increase the variability of the sampled field, the amplitude used in the IPAC algorithm can be replaced by the amplitude that includes a digital filter. The beta distribution with a lower bound of 5%*wt* and an upper bound of 95%*wt* is employed to fit the non-zero depth distribution depicted in Fig. 5.2b. To illustrate the effect of using the fitted distribution and digital filtering, we repeat the analysis but use DT-CDWT and DT-CHWT for the example considered with the target scalogram of the measured corrosion field shown in Fig. 5.7. Typical sampled fields, in this case, are shown in Fig. 5.13a. The mean and standard deviation of the scalogram estimated based on 100 samples generated by using DT-CDWT or DT-CHWT are calculated and shown in Figs. 5.13b and 5.13c, respectively. The CDF of the fitted beta distribution is illustrated in Fig. 5.14, and the distribution of the calculated burst capacity by using the simulated samples is shown in Fig. 5.15. A comparison of the results presented in Fig. 5.13 with those shown in Figs. 5.10 and 5.11 indicates that the mean of the scalogram is not affected by considering the fitted distribution and digital filter, whereas the standard deviation of the scalogram increases as compared to that without considering the digital filter. The latter is expected and has the desired effect of increasing the variability of the sampled field. A comparison of Figs. 5.15 and 5.12 indicates that the consideration of the fitted distribution and digital filter can maintain the mean value and increases the variability of the burst capacity corresponding to the sampled fields. The COVs of the burst capacity under different conditions are summarized in Table 5.4. Although the flexibility of selecting the marginal distribution and amplitude of the scalogram in this framework enables us to study how some statistical characteristics of the corrosion field affect the accuracy of semi-empirical burst capacity models, the parameter of the marginal distribution and amplitude of the

scalogram are related. For example, the variance of the marginal distribution can be estimated from the scalogram; therefore, the consistency in assigning the marginal PDF and scalogram should be considered when introducing modifications and carrying out the parametric assessment.
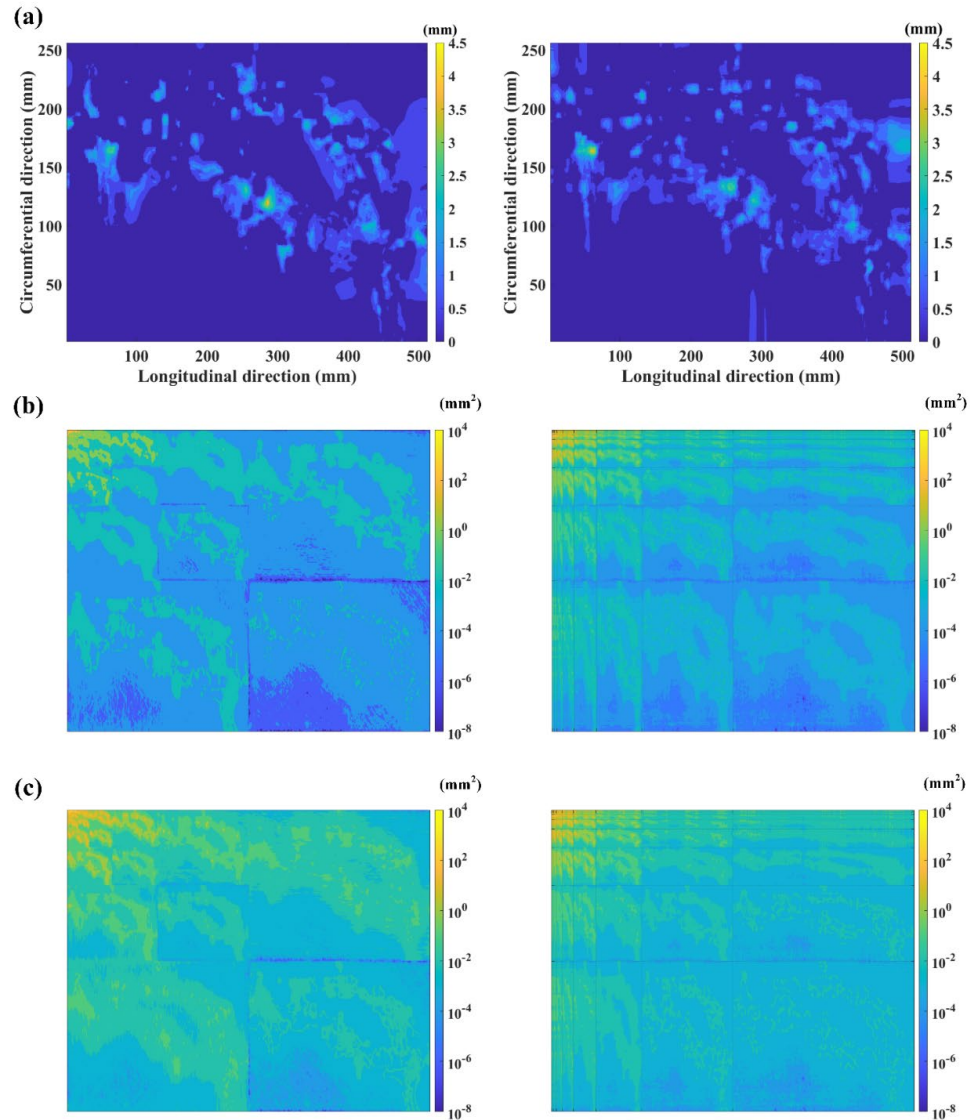
Figure 5.13 Simulated corrosion fields based on DT-CDWT (left column) and DT-CHWT (right column), considering the fitted corrosion depth distribution and digital filter: (a) typical sampled corrosion fields; (b) mean of the scalogram estimated from 100 sampled fields; and (c) standard deviation of the scalogram estimated from 100 sampled fields
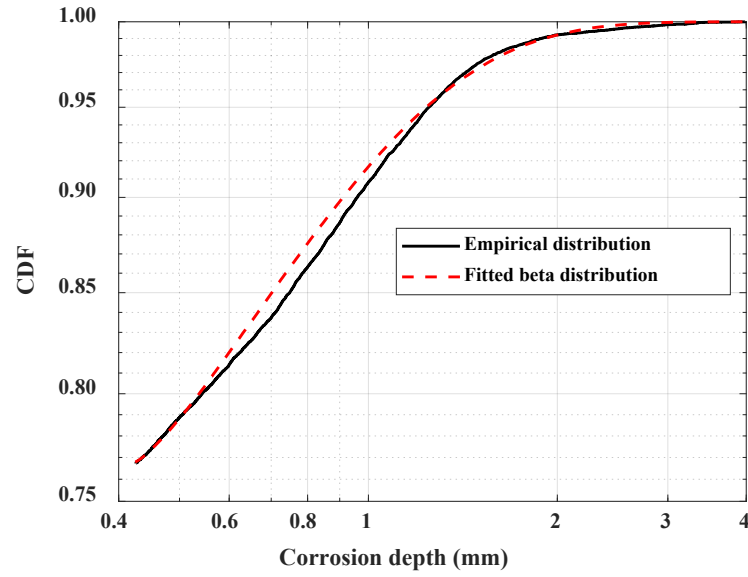
Figure 5.14 CDF of the empirical distribution and fitted beta distribution of corrosion depth
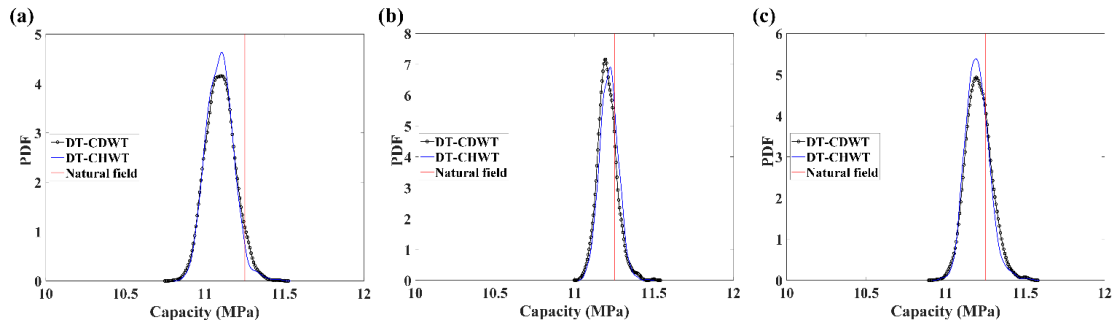
Figure 5.15 Empirical PDF of the burst capacity of pipe segment obtained from realizations of the simulated field considering: (a) surrogate and amplitude with digital filter; (b) fitted beta distribution and amplitude without digital filter; and (c) fitted beta distribution and amplitude with the digital filter

Table 5.4 COVs of the burst capacity under different conditions

Marginal distribution	Scalogram modification	Transform	
		DT-CDWT	DT-CHWT
Surrogate	No digital filter	0.59%	0.57%
	With digital filter	0.86%	0.77%
Beta distribution	No digital filter	0.58%	0.55%
	With digital filter	0.81%	0.80%

5.4 Conclusion

In the present study, we present a framework to analyze and simulate nonhomogeneous non-Gaussian corrosion fields on the external surface of buried in-service pipelines by using the discrete and continuous wavelet transforms. The considered wavelet transforms include the CWT using the complex Morlet wavelet, DT-CDWT, and DT-CHWT. The natural corrosion field, which is measured by a high-resolution high-accuracy laser scanner and is used as the target field, is on the external surface of an in-service buried pipeline. The IPAC algorithm is employed to simulate the corrosion fields with the prescribed amplitude of the transform defined based on the considered wavelet transforms and the empirical marginal distribution of the corrosion depth. The simulated fields are assessed by comparing scalograms of the selected transform pair, textural characteristics extracted from the GLCM, and burst capacities of the pipeline containing the corrosion field estimated by the RSTRENG model. Scalograms of the measured corrosion field indicate that most of the energy of the field is conserved by components at long scales (low wavenumber). In DT-CDWT and DT-CHWT, over 90% of the total energy is concentrated in 0.4% of the amplitude square of the transform coefficients at long scales. The energy is not homogeneously distributed in the spatial domain. Simulation results indicate that the corrosion fields generated by using the IPAC algorithm together with the CWT, DT-CDWT, and DT-CHWT capture probabilistic characteristics of the natural corrosion field. The burst capacities of the simulated

corrosion fields are almost the same as that of the natural corrosion field, and the variability of the burst capacities corresponding to the simulated corrosion fields is low. Introducing a digital filter to the amplitude information can slightly increase the variability of the simulated fields, but the COVs of the burst capacities are less than 1%. Although the resolution of DT-CDWT and DT-CHWT in the spatial-wavenumber domain is lower than that of CWT, these moderately redundant transforms can capture most of the textural information of the natural corrosion fields and preserve such information in the synthetic fields with markedly lower computational costs than CWT. The burst capacities corresponding to the synthetic fields generated based on DT-CDWT and DT-CHWT differ by less than 0.5% from those of the synthetic fields corresponding to CWT. The application of this framework can be used to generate synthetic corrosion fields based on the target measured corrosion field with empirical distribution or fitted distribution and with or without a digital filter. Since the use of DT-CDWT and DT-CHWT is associated with significantly reduced computing time as compared to that associated with CWT, they are excellent candidates for simulating synthetic corrosion fields that can be used to facilitate the development and validation of burst capacity models for corroded pipelines.

References

Abdalla Filho, J. E., Machado, R. D., Bertin, R. J., & Valentini, M. D. (2014). On the failure pressure of pipelines containing wall reduction and isolated pit corrosion defects. *Computers & Structures*, *132*, 22-33.

Adam, M., Ng, E. Y., Oh, S. L., Heng, M. L., Hagiwara, Y., Tan, J. H., ... & Acharya, U. R. (2018). Automated detection of diabetic foot with and without neuropathy using double density-dual tree-complex wavelet transform on foot thermograms. *Infrared Physics & Technology*, *92*, 270-279.

Antoine, J. P., Vandergheynst, P., & Murenzi, R. (1996). Two‐dimensional directional wavelets in image processing. *International journal of imaging systems and technology*, *7*(3), 152-165.

ASME. (1991). Manual for determining the remaining strength of corroded pipelines. *ASME B31G*, 1991.

Aryai, V., & Mahmoodian, M. (2017). Spatial-temporal reliability analysis of corroding cast iron water pipes. *Engineering Failure Analysis*, *82*, 179-189.

Bao, J., Zhang, S., Zhou, W., & Zhang, S. (2018, September). Evaluation of burst pressure of corroded pipe segments using three-dimensional finite element analyses. In *International pipeline conference* (Vol. 51869, p. V001T03A043). American Society of Mechanical Engineers.

Bao, J., & Zhou, W. (2021). A random field model of external metal-loss corrosion on buried pipelines. *Structural safety*, *91*, 102095.

Casavola, C., Pappalardi, P., Pappalettera, G., & Renna, G. (2018). A fringe projection based approach for corrosion monitoring in metals. *Experimental Techniques*, *42*, 291-297.

Chavez, M., & Cazelles, B. (2019). Detecting dynamic spatial correlation patterns with generalized wavelet coherence and non-stationary surrogate data. *Scientific reports*, *9*(1), 7389.

Chitchian, S., Mayer, M. A., Boretsky, A. R., Van Kuijk, F. J., & Motamedi, M. (2012). Retinal optical coherence tomography image enhancement via shrinkage denoising using double-density dual-tree complex wavelet transform. *Journal of biomedical optics*, *17*(11), 116009-116009.

Chopra, S., & Marfurt, K. J. (2015, October). Choice of mother wavelets in CWT spectral decomposition. In *SEG International Exposition and Annual Meeting* (pp. SEG-2015). SEG.

Cohen, M. X. (2019). A better way to define and describe Morlet wavelets for time-frequency analysis. *NeuroImage*, *199*, 81-86.

Cronin, D. S. (2000). *Assessment of corrosion defects in pipelines*.

Daubechies, I. (1992). *Ten lectures on wavelets*. Society for industrial and applied mathematics.

DeVore, R. A., Konyagin, S. V., & Temlyakov, V. N. (1998). Hyperbolic wavelet approximation. *Constructive Approximation*, *14*(1), 1-26.

Dijkerman, R. W., & Mazumdar, R. R. (1994). Wavelet representations of stochastic processes and multiresolution stochastic models. *IEEE Transactions on Signal Processing*, *42*(7), 1640-1652.

Dolan, K. T., & Spano, M. L. (2001). Surrogate for nonlinear time series analysis. *Physical Review E*, *64*(4), 046128.

EGIG. 11th Report of the European Gas pipeline Incident data Group, 2020. https://www.egig.eu/reports

El Hassani, A. S. E. B., El Hassouni, M., Jennane, R., Rziza, M., & Lespessailles, E. (2012). Texture analysis for trabecular bone X-ray images using anisotropic morlet wavelet and Rényi entropy. *In Image and Signal Processing: 5th International Conference, ICISP 2012, Agadir, Morocco, June 28-30, 2012. Proceedings 5* (pp. 290-297). Springer Berlin Heidelberg.

Farge, M. (1992). Wavelet transforms and their applications to turbulence. *Annual review of fluid mechanics*, *24*(1), 395-458.

Farouj, Y., Freyermuth, J. M., Navarro, L., Clausel, M., & Delachartre, P. (2016). Hyperbolic Wavelet-Fisz denoising for a model arising in Ultrasound Imaging. *IEEE Transactions on Computational Imaging*, *3*(1), 1-10.

Ferreira, A. D. M., Afonso, S. M., Willmersdorf, R. B., & Lyra, P. R. (2021). Multiresolution analysis and deep learning for corroded pipeline failure assessment. *Advances in Engineering Software*, *162*, 103066.

Garbatov, Y., & Guedes Soares, C. (2019). Spatial corrosion wastage modeling of steel plates exposed to marine environments. *Journal of Offshore Mechanics and Arctic Engineering*, *141*(3), 031602.

Gurley, K., & Kareem, A. (1999). Applications of wavelet transforms in earthquake, wind and ocean engineering. *Engineering structures*, *21*(2), 149-167.

Hall-Beyer, M. (2017). *GLCM texture: A tutorial* v. 3.0 March 2017.

Hong, H. P., Cui, X. Z., & Qiao, D. (2021a). An algorithm to simulate nonstationary and non-Gaussian stochastic processes. *Journal of Infrastructure Preservation and Resilience*, *2*(1), 17.

Hong, H. P., Cui, X. Z., & Zhou, W. X. (2021b). A model to simulate multidimensional nonstationary and non-Gaussian fields based on S-transform. *Mechanical Systems and Signal Processing*, 159, 107789.

Huang, S. J., & Hsieh, C. T. (1999). High-impedance fault detection utilizing a Morlet wavelet transform approach. *IEEE Transactions on Power Delivery*, *14*(4), 1401-1410.

Kankar, P. K., Sharma, S. C., & Harsha, S. P. (2011). Fault diagnosis of ball bearings using continuous wavelet transform. *Applied Soft Computing*, *11*(2), 2300-2312.

Kathirmani, S., Tangirala, A. K., Saha, S., & Mukhopadhyay, S. (2012). Online data compression of MFL signals for pipeline inspection. *NDT & E International*, *50*, 1-9.

Kiefner, J. F., & Vieth, P. H. (1989). *A modified criterion for evaluating the remaining strength of corroded pipe* (No. PR-3-805). Battelle Columbus Div., OH (USA).

Kingsbury, N. (2001). Complex wavelets for shift invariant analysis and filtering of signals. *Applied and computational harmonic analysis*, *10*(3), 234-253.

Lam, C., & Zhou, W. (2016). Statistical analyses of incidents on onshore gas transmission pipelines based on PHMSA database. *International Journal of Pressure Vessels and Piping*, *145*, 29-40.

Lee, T. S. (1996). Image representation using 2D Gabor wavelets. *IEEE Transactions on pattern analysis and machine intelligence*, *18*(10), 959-971.

Lilly, J. M., & Olhede, S. C. (2008). Higher-order properties of analytic wavelets. *IEEE Transactions on Signal Processing*, *57*(1), 146-160.

Lin, J., & Qu, L. (2000). Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. *Journal of sound and vibration*, *234*(1), 135-148.

Lin, Z., & Lu, B. (2010, August). Iris recognition method based on the imaginary coefficients of Morlet wavelet transform. In *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery* (Vol. 2, pp. 573-577). IEEE.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, *11*(7), 674-693.

Mallat, S. G. (1999). *A wavelet tour of signal processing: the sparse way*. Elsevier.

Masters, F., & Gurley, K. R. (2003). Non-Gaussian simulation: cumulative distribution function map-based spectral correction. *Journal of engineering mechanics*, *129*(12), 1418-1428.

Mok, D. H. B., Pick, R. J., Glover, A. G., & Hoff, R. (1991). Bursting of line pipe with long external corrosion. *International journal of pressure vessels and piping*, *46*(2), 195-216.

Nychka, D., Wikle, C., & Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, *2*(4), 315-331.

Otsuka, A., Muramatsu, K., & Nagata, F. (2020). Data-driven generation of random skin model shapes by using wavelet transformation. *Computers & Industrial Engineering*, *150*, 106860.

Otsuka, A., Muramatsu, K., Sugi, M., & Nagata, F. (2018). Characteristic extraction of machined surface using wavelet transformation. *Artificial Life and Robotics*, *23*, 316-320.

Oulhaj, H., Rziza, M., Amine, A., Toumi, H., Lespessailles, E., El Hassouni, M., & Jennane, R. (2017). Anisotropic discrete dual-tree wavelet transform for improved classification of trabecular bone. *IEEE transactions on medical imaging*, *36*(10), 2077-2086.

Roux, S. G., Clausel, M., Vedel, B., Jaffard, S., & Abry, P. (2013). Self-similar anisotropic texture analysis: The hyperbolic wavelet transform contribution. *IEEE Transactions on Image Processing*, *22*(11), 4353-4363.

Saadatmorad, M., Talookolaei, R. A. J., Pashaei, M. H., Khatir, S., & Wahab, M. A. (2022). Pearson correlation and discrete wavelet transform for crack identification in steel beams. *Mathematics*, *10*(15), 2689.

Selesnick, I. W. (2001). The double density DWT. In *Wavelets in Signal and Image Analysis: From Theory to Practice* (pp. 39-66). Dordrecht: Springer Netherlands.

Selesnick, I. W. (2004). The double-density dual-tree DWT. *IEEE Transactions on signal processing*, 52(5), 1304-1314.

Selesnick, I. W. (2008). On the dual-tree complex wavelet packet and M-band transforms. *IEEE Transactions on Signal Processing*, *56*(6), 2298-2310.

Selesnick, I. W., Baraniuk, R. G., & Kingsbury, N. C. (2005). The dual-tree complex wavelet transform. *IEEE signal processing magazine*, *22*(6), 123-151.

Song, S. P., & Que, P. W. (2006). Wavelet based noise suppression technique and its application to ultrasonic flaw detection. *Ultrasonics*, *44*(2), 188-193.

Song, S. Z., Wang, S. Y., Gao, Z. M., Zhang, X. Y., Ma, Y. J., Li, J. Z. (2002) Atmospheric forepart corrosion behaviours of nonferrous metal based on image recognition. *Acta Metallurgica Sinica*, *30*(8), 893-896.

Spanos, P. D., & Rao, V. R. S. (2001). Random field representation in a biorthogonal wavelet basis. *Journal of engineering mechanics*, *127*(2), 194-205.

Stockwell, R. G., Mansinha, L., & Lowe, R. P. (1996). Localization of the complex spectrum: the S transform. *IEEE transactions on signal processing*, *44*(4), 998-1001.

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, *79*(1), 61-78.

Xu, S. J., Weng, Y. J., & Liu, X. H. (2009). Correlations between wavelet image features and corrosion parameters for carbon steel corroded in soil. *Corrosion Science and Protetion Technology*, *21*(5), 429-432.

Zadkarami, M., Shahbazian, M., & Salahshoor, K. (2016). Pipeline leakage detection and isolation: An integrated approach of statistical and wavelet feature extraction with multi-layer perceptron neural network (MLPNN). *Journal of Loss Prevention in the Process Industries*, *43*, 479-487.

Zeldin, B. A., & Spanos, P. D. (1996). Random field representation and synthesis using wavelet bases. *Journal of Applied Mechanics*, *63*, 946-952.

Zhang, S., Yan, J., Kariyawasam, S., Huang, T., & Al-Amin, M. (2018, September). A more accurate and precise method for large metal loss corrosion assessment. In *International Pipeline Conference* (Vol. 51869, p. V001T03A059). American Society of Mechanical Engineers.

Zhang, S., & Zhou, W. (2020). Assessment of effects of idealized defect shape and width on the burst capacity of corroded pipelines. *Thin-Walled Structures*, *154*, 106806.

Zhou, W., Bao, J., Cui, X. Z., & Hong, H. P. (2021). Modeling and simulating nonhomogeneous non-Gaussian corrosion fields on buried pipelines and its use in

predicting burst capacities of corroded pipelines. *Engineering Structures*, *245*, 112957.

Chapter 6

6 Summary, Conclusions and Recommendations for Future Study

6.1 General

From measuring in-service pipelines to predicting pipeline status, pipeline assessment includes multiple stages. The accuracy of pipeline assessment is affected by the accuracy of the respective assessments at these stages. Small improvements at multiple stages may be more effective in improving the accuracy of the final assessment results than optimizing the assessment at one single stage. A pipeline containing defects can therefore be assessed more effectively and maintained in a timely and cost-effective manner. Thus, the present study attempts to improve the methods commonly used in different stages of pipeline assessment by introducing a few practical assumptions, given the limited available information.

This research employs wavelet transforms and random field analysis to address four major issues in the context of the fitness-for-service (FFS) assessment of integrity management of buried steel pipelines containing dents and corrosions. Conclusion drawn from the four individual studies are summarized as follows.

6.2 A Wavelet-based Denoising Methodology for Pipeline Dent Assessments

Chapter 2 proposed a denoising method for the dent signals of indented pipelines measured using caliper tools. The proposed method is based on the overcomplete expansion, whereby the overcomplete dictionary is constructed by the stationary wavelet transform (SWT) and the hyperbolic wavelet transform (HWT). Finite element analysis (FEA) is employed to obtain the noise-free morphology and equivalent strain of the inner surface of an indented pipe segment. By artificially adding different levels of Gaussian white noise to the noise-free dent signal and denoising such a noisy signal, we compare the denoising performance of the commonly used wavelet transform-based thresholding methods and the proposed method. The proposed method achieves lower root mean

square error (*RMSE*) if the standard deviation of the noise is above 0.3 mm. The maximum effective strains in the dent estimated from the signal denoised using the proposed method are markedly closer to the actual strain than that based on the commonly used thresholding methods.

To improve the computational efficiency of the proposed method, we construct a large overcomplete dictionary by all the Daubechies wavelets and Daubechies symlets with no more than ten vanishing moments and then denoise 42 noisy dent signals collected from in-service pipelines using such a dictionary. The statistics of extracted components indicate that most extracted components of the natural dent signals are constructed by DB10 along the longitudinal direction and Symlet 4 along the circumferential direction. The decomposition levels along the longitudinal and circumferential directions of most extracted components are 7 and 3, respectively. Differences in the wavelet functions and decomposition levels along two directions further support that the commonly used thresholding is inadequate for the dent signals. Based on the statistics of extracted components, we construct a small overcomplete dictionary that uses DB10 and Symlet 4 along the longitudinal and circumferential directions, respectively. On average, employing the small dictionary will extract 56% more components than employing the large dictionary, but the computation time is reduced to 1/48. Most of the estimated maximum effective strains based on the large and small dictionaries are close. This research will improve the fitness-for-service assessment of pipelines containing dents.

6.3 Discrete Wavelet Analysis of External Corrosion Clusters on Pipelines for Burst Capacity Prediction

Chapter 3 applies the two-dimensional discrete wavelet transform (2D DWT) to analyze 106 naturally occurring corrosion clusters on external surfaces of in-service gas transmission pipelines. The Daubechies wavelet with two vanishing moments (DB2) is adopted in the analysis, and the widely used RSTRENG model is employed to compute the burst capacity of pipelines with corrosion clusters. A methodology is proposed to determine constant, level- and sub-band-dependent thresholds such that those wavelet coefficients below the thresholds have a negligible impact on the burst capacity and can be ignored for the cluster reconstruction. The proposed methodology enables one to

predict the burst capacity of the corrosion cluster with an error of no more than 1% while retaining less than 1% of the wavelet coefficients. Although the corrosion clusters reconstructed by the retained wavelet coefficients may have a large visual difference compared to the original corrosion clusters, the discarded details have little effect on the burst capacity prediction.

Statistical analysis of the preserved wavelet coefficients shows that most of the features in the natural corrosion clusters are circumferential features, but their influence on the burst capacity prediction is not as significant as longitudinal features. Features that greatly impact the burst capacity prediction are approximately 4 – 16 mm in width and 4 – 32 mm in length. It is unnecessary to decompose the natural corrosion cluster to a scale greater than 32 mm because the wavelet coefficients at higher levels have a negligible impact on the burst capacity prediction. This research allows one to quickly extract information that has a large impact on the burst capacity prediction from the measured corrosion clusters.

6.4 Simulating Nonhomogeneous Non-Gaussian Corrosion Fields on Pipelines Based on In-line Inspection Data

Chapter 4 presents a methodology to simulate nonhomogeneous non-Gaussian corrosion fields on the external surface of buried steel pipelines by using inline inspection (ILI) data. We assume that the nonhomogeneous non-Gaussian corrosion field consists of multiple homogeneous non-Gaussian anomalies that can be characterized by the marginal distribution and spatial autocorrelation function of the corresponding corrosion depth. High-resolution laser scan data for 3408 anomalies within 105 natural corrosion fields identified on buried natural gas transmission pipelines in Canada are used to evaluate the marginal probability distributions and autocorrelation functions of corrosion depths within the anomalies, and the ILI parameters of anomalies are obtained from a virtual ILI process. Empirical relationships between the sizing parameters of these anomalies and the marginal probability distributions as well as the autocorrelation functions of corrosion depths within the anomalies are developed. An algorithm proposed by Masters and Gurley is employed to generate homogeneous non-Gaussian corrosion fields representing corrosion anomalies. Spatial modulators constructed from the ILI location parameters are

then employed to localized the synthetic corrosion anomalies. The synthetic corrosion fields are generated by combining these simulated anomalies. For the 105 natural corrosion fields considered in this chapter, the burst capacity of the synthetic field is, on average, 97.4% of that of the actual corrosion field, whereas the burst capacity of the ILI-idealized field is, on average, 91.5% of that of the actual field. Compared with the commonly used ILI idealization with cuboidal anomalies, the burst capacities predicted by the synthetic corrosion fields are less conservative. The proposed methodology will improve the accuracy of the fitness-for-service assessment of corroded pipelines in practice and reduce the cost of maintenance.

6.5 Application of Wavelet Transforms to the Simulation of Corrosion Fields on Buried Pipelines

Chapter 5 presents a framework to analyze and simulate nonhomogeneous non-Gaussian corrosion fields on the external surface of buried in-service pipelines by using continuous and discrete wavelet transforms. The considered transforms are the two-dimensional continuous wavelet transform (CWT) using the complex Morlet wavelets, dual-tree complex discrete wavelet transform (DT-CDWT), and dual-tree complex discrete wavelet with hyperbolic wavelet transform scheme (DT-CHWT); the natural corrosion field is measured using a high-resolution laser scan. Scalograms and marginal distribution of the measured corrosion field are incorporated into the iterative power and amplitude correction (IPAC) algorithm to generate synthetic corrosion fields. The framework is explained and illustrated using a numerical example.

The simulated fields are assessed by comparing scalograms of the selected transform pair, textural characteristics extracted from the gray-level co-occurrence matrix (GLCM), and burst capacities of the pipeline containing the corrosion field estimated by the RSTRENG model. Scalograms of the measured corrosion field show that most energy of the field is conserved by components at long scales. Results of textual features and burst capacity indicate that the proposed method can capture the probabilistic characteristics of the natural corrosion field, and these characteristics are well-preserved in the synthetic fields. The differences in these probabilistic characteristics between the CWT-based synthetic fields and the synthetic fields based on the other two transforms are small, while the

computation time differs significantly. It follows that DT-CDWT and DT-CHWT, which require less computation time, are recommended for simulating synthetic corrosion fields that can be used to facilitate the development and validation of burst capacity models for corroded pipelines.

6.6 Limitations of Current Studies and Recommendations for Future Study

The limitations of current studies and recommendations for future study are summarized as follows.

1. The denoising performance of the proposed method in Chapter 2 depends on the selected overcomplete dictionary. Although employing a large dictionary with many atoms enables one to represent a signal with as few components as possible, some of the components from such a large dictionary may be very unlikely to be found in the natural dent morphology. Therefore, to achieve a better denoising performance, it is necessary to construct an overcomplete dictionary consisting of atoms that are commonly seen in the natural dent morphology. To this end, noise-free measurements of the inner radii of in-service indented pipelines are required. Besides, the computation time needs to be further shortened to achieve higher practicality. Reducing iterations in the matching pursuit or employing other algorithms to seek an optimal solution for the overcomplete expansion can be involved in future studies.

2. The study presented in Chapter 3 is based on the RSTRENG model, as FEA is too time-consuming. Therefore, the model error of the RSTRENG model may affect the empirical parameters (i.e. G_α, G1, and G_2) obtained in the present study. Besides, the proposed method is developed based on detailed corrosion depth measurements, which may not be available in some ILI technologies such as the magnetic flux leakage. FEA can be involved in future studies to obtain better empirical parameters, and different information extraction methods should be developed for different ILI signals.

3. In Chapter 4, since the ILI data corresponding to the corrosion clusters is unavailable, a virtual ILI process is employed to obtain ILI parameters for anomalies. However, the virtual ILI process is not the real ILI process, which means the ILI parameters obtained from the virtual process may significantly differ from the actual ILI parameters (e.g. uncertainty of parameters due to measurement error). Besides, the beta distribution and exponential autocorrelation function are employed, while they significantly differ from the empirical marginal distribution and autocorrelation function of some anomalies. Therefore, analysis based on the real ILI parameters should be considered in future studies to improve the practicality of the proposed method. Additionally, selecting different forms of autocorrelation functions and distributions based on the value range of ILI parameters may better simulate the corrosion field.

4. The variability of synthetic fields generated by the framework presented in Chapter 5 may be too small. In other words, the synthetic fields are too similar to the natural corrosion fields. Besides, to generate synthetic fields by the proposed framework, detailed scalograms are required. These issues may be solved by extracting and summarizing essential information in the scalograms. By keeping only the parts of the scalogram with high energy or summarizing the scalogram as several or dozens of quantities, we can introduce more uncertainties into the simulation. Furthermore, it is promising to introduce or develop a transform with an arbitrary choice of redundancy and can be incorporated into IPAC, so that the relation between the redundancy of the transform and the resolution of synthetic fields in the time (spatial) and frequency (wavenumber) domains can be further investigated. This can be involved in the future study so that one can choose a more suitable transform according to the requirements (e.g. variability, spatial or wavenumber resolution, computational cost) of generating synthetic fields.

Appendices

Appendix A: Determination of the upper bounds of wavelet thresholds

Let $F_{i,\alpha}(j_1, j_2)$ ($j_1, j_2 = 1, 2, \ldots, A_{max,i}$ are the integer index of filters, and $A_{max,i}$ is the length of the 1D filter at the ith level) denote the 2D analysis filter used to calculate the ith level wavelet coefficients of the α sub-band. A wavelet coefficient is calculated by convolving the analysis filters and the original signal:

$$z_\psi^{(\alpha,i)}(\tau_x, \tau_y) = \sum_x \sum_y z(x,y) F_{i,\alpha}(\tau_x - x, \tau_y - y) \tag{A.1}$$

Equation (3.6) and Eq. (A.1) are equivalent, because $F_{i,\alpha}$ is the non-zero part of $\frac{1}{2^i}\psi_{(\alpha)}$ arranged in reverse order in both directions. In Chapter 3, a corrosion cluster z is periodically extended at four bounds when calculating the wavelet coefficients. Let $z^{(PE)}$ denotes the periodically extended z. Let $B_{i,\alpha,\tau_x,\tau_y}(x,y) = z^{(PE)}(x,y) \geq 0$, where $\tau_x - A_{max,i} \leq x \leq \tau_x - 1$ and $\tau_y - A_{max,i} \leq y \leq \tau_y - 1$. $B_{i,\alpha,\tau_x,\tau_y}$ can be considered as a part of z near (τ_x, τ_y) for calculating $z_\psi^{(\alpha,i)}(\tau_x, \tau_y)$. Since $F_{i,\alpha}$ is determined if a wavelet function is selected, the wavelet coefficient $z_\psi^{(\alpha,i)}(\tau_x, \tau_y)$ only depends on the grid point values in $B_{i,\alpha,\tau_x,\tau_y}$.

Based on the definition of $\eta_{\alpha,i}^{(ub)}$, if $\left|z_\psi^{(\alpha,i)}\right| > \eta_{\alpha,i}^{(ub)}\eta_d$, at least one gird point in $B_{i,\alpha,\tau_x,\tau_y}$ have a corrosion depth greater than η_d. Let $\mathbf{J}^{(+)}$ and $\mathbf{J}^{(-)}$ denote sets of points that $F_{i,\alpha}(j_1, j_2) > 0$ if $(j_1, j_2) \in \mathbf{J}^{(+)}$ and $F_{i,\alpha}(j_1, j_2) < 0$ if $(j_1, j_2) \in \mathbf{J}^{(-)}$, respectively. Consider the positive and negative taps of $F_{i,\alpha}$, respectively denoted as $F_{i,\alpha}^{(+)}$ and $F_{i,\alpha}^{(-)}$, as follows:

$$F_{i,\alpha}^{(+)}(j_1, j_2) = \begin{cases} F_{i,\alpha}(j_1, j_2), & (j_1, j_2) \in \mathbf{J}^{(+)} \\ 0, & other \end{cases} \tag{A.2}$$

$$F_{i,\alpha}^{(-)}(j_1, j_2) = \begin{cases} F_{i,\alpha}(j_1, j_2), & (j_1, j_2) \in \mathbf{J}^{(-)} \\ 0, & other \end{cases} \tag{A.3}$$

For the *LL* sub-band, $\left|\Sigma_{j_1}\Sigma_{j_2}F_{i,\alpha}^{(+)}(j_1,j_2)\right| > \left|\Sigma_{j_1}\Sigma_{j_2}F_{i,\alpha}^{(-)}(j_1,j_2)\right|$. For *LH*, *HL*, and *HH* sub-bands, $\left|\Sigma_{j_1}\Sigma_{j_2}F_{i,\alpha}^{(+)}(j_1,j_2)\right| = \left|\Sigma_{j_1}\Sigma_{j_2}F_{i,\alpha}^{(-)}(j_1,j_2)\right|$. It follows that if all the grid points in $B_{i,\alpha,\tau_x,\tau_y}$ corresponding to $\mathbf{J}^{(+)}$ have a corrosion depth η_d while other grid points are zeros, the wavelet coefficient $z_\psi^{(\alpha,i)}(\tau_x,\tau_y) = \eta_{\alpha,i}^{(ub)}\eta_d$. Therefore, $\eta_{\alpha,i}^{(ub)}$ can be calculated by:

$$B_{i,\alpha,\tau_x,\tau_y}(x,y) = \begin{cases} \eta_d, & x = \tau_x - j_1, y = \tau_y - j_2, (j_1,j_2) \in \mathbf{J}^{(+)} \\ 0, & other \end{cases} \tag{A.4}$$

$$\eta_{\alpha,i}^{(ub)} = \left|\Sigma_x\Sigma_y B_{i,\alpha,\tau_x,\tau_y}(x,y)F_{i,\alpha}(\tau_x - x, \tau_y - y)\right|/\eta_d = \Sigma_{j_1}\Sigma_{j_2}F_{i,\alpha}^{(+)}(j_1,j_2) \tag{A.5}$$

If the wavelet function for 2D DWT is selected, $F_{i,\alpha}(j_1,j_2)$ is determined. The values in Table 3.2 are calculated by Eqs. (A.2), (A.4), and (A.5) with $F_{i,\alpha}(j_1,j_2)$ determined by DB2.

Appendix B: Master and Gurley's algorithm for non-Gaussian simulation

Masters and Gurley (2003) proposed an algorithm for non-Gaussian simulation. This method enable researchers to generate synthetic record $x(t)$ with target power spectral density function S_{xx}^T and cumulative distribution function F_X^T. The algorithm is presented as follows:

1) Generate a Gaussian record $x(j\Delta t)$ using the following equation:

$$x(j\Delta t) = 2 \sum_{k=0}^{M-1} \sqrt{S_{xx}^T(k\Delta\omega)\Delta\omega} \cdot e^{i\phi_k} \cdot e^{i(k\Delta\omega)(j\Delta t)} \tag{B.1}$$

where Δt is the time increment; M is the index of the highest contributing frequency; k is the index of contributing frequency; $\Delta\omega$ is the frequency increment; e is the base of the natural logarithms; $i = \sqrt{-1}$ is the imaginary unit; and ϕ_k are phase angles that uniformly and independently distributed over $[0 \; 2\pi]$. Denote $x(j\Delta t)$ as $x_{pc}(j\Delta t)$. $x_{pc}(j\Delta t)$ matches S_{xx}^T but the marginal distribution will deviate from F_X^T.

2) Denote the cumulative distribution function (CDF) of $x_{pc}(j\Delta t)$ as $F_{\hat{X}}$. Let $x_{ac}(j\Delta_t) = (F_X^T)^{-1}\left(F_{\hat{X}}\left(x_{pc}(j\Delta t)\right)\right)$. $x_{ac}(j\Delta_t)$ matches F_X^T but deviate from S_{xx}^T.

3) Apply Fourier transform on $x_{ac}(j\Delta_t)$ and obtain the phase ϕ_k for each frequency component. $x_{pc}(j\Delta t)$ is generated using Eq. (A1), where the phase ϕ_k are replaced by the phases calculated from $x_{ac}(j\Delta_t)$.

4) Repeat step (b) and (c) until the error $\varepsilon = \dfrac{\sum_j\left(x_{ac}(j\Delta_t)-x_{pc}(j\Delta_t)\right)^2}{\sum_j\left(x_{ac}(j\Delta_t)\right)^2}$ reaches the user-determined tolerance. In this study, $\varepsilon = 0.1\%$ is employed.

Usually $x_{ac}(j\Delta t)$ will be used as the output. This algorithm can be used to generate multi-dimensional signals, if the multi-dimensional power spectral density function is given.

Appendix C: DT-CDWT filters used in this study

In 1D DT-CDWT, a 1D signal is decomposed by two separable trees (i.e. the real tree $\psi_{re}(\cdot)$ and the imaginary tree $\psi_{im}(\cdot)$), and the complex coefficients are calculated by assigning wavelet coefficients from the real tree as the real part, and wavelet coefficients from the imaginary tree as the imaginary part. The first level decomposition is undecimated, and coefficients at odd and even positions are considered as coefficients from the real and imaginary trees, respectively. The filters used at the first level can be any filters with the perfect reconstruction. In this study, the (5,7)-tap near-orthogonal filters used at the first level are listed in Table C.1 (Antonini et al. 1992). The number of taps is the number of coefficients in a filter.

Table C.1 The (5,7)-tap near-orthogonal filters

Analysis filter		Synthesis filter	
Low-pass	High-pass	Low-pass	High-pass
	3/280	-3/280	
-1/20	-3/56	-3/56	-1/20
1/4	-73/280	73/280	-1/4
3/5	17/28	17/28	3/5
1/4	-73/280	73/280	-1/4
-1/20	-3/56	-3/56	-1/20
	3/280	-3/280	

For the second and higher level of 1D DT-CDWT, the shift invariance is achieved by the Q-shift filters proposed by Kingsbury (2001). The 10-tap Q-shift filters for the real and imaginary trees are listed below:

Table C.2 The 10-tap Q-shift filters for the real tree

Analysis filter		Synthesis filter	
Low-pass	High-pass	Low-pass	High-pass
0.05113041	-0.00618188	-0.00618188	-0.05113041
-0.01397537	0.00168968	-0.00168968	-0.01397537
-0.10983605	-0.10023122	-0.10023122	0.10983605
0.26383956	-0.00087362	0.00087362	0.26383956
0.76662847	0.56365571	0.56365571	-0.76662847
0.56365571	-0.76662847	0.76662847	0.56365571
0.00087362	0.26383956	0.26383956	-0.00087362
-0.10023122	0.10983605	-0.10983605	-0.10023122
-0.00168968	-0.01397537	-0.01397537	0.00168968
-0.00618188	-0.05113041	0.05113041	-0.00618188

Table C.3 The 10-tap Q-shift filters for the imaginary tree

Analysis filter		Synthesis filter	
Low-pass	High-pass	Low-pass	High-pass
-0.00618188	-0.05113041	0.05113041	-0.00618188
-0.00168968	-0.01397537	-0.01397537	0.00168968
-0.10023122	0.10983605	-0.10983605	-0.10023122
0.00087362	0.26383956	0.26383956	-0.00087362
0.56365571	-0.76662847	0.76662847	0.56365571
0.76662847	0.56365571	0.56365571	-0.76662847
0.26383956	-0.00087362	0.00087362	0.26383956
-0.10983605	-0.10023122	-0.10023122	0.10983605
-0.01397537	0.00168968	-0.00168968	-0.01397537
0.05113041	-0.00618188	-0.00618188	-0.05113041

Denoting the complex filters in 1D DT-CDWT as $\psi_c(t) = \psi_{re}(t) + j\psi_{im}(t)$, filters of 2D DT-CDWT are calculated from tensor product:

$$\psi_{c1}(x,y) = \psi_c(x) \otimes \psi_c(y) = \left[\psi_{re}(x)\psi_{re}(y) - \psi_{im}(x)\psi_{im}(y)\right] + j\left[\psi_{im}(x)\psi_{re}(y) + \psi_{re}(x)\psi_{im}(y)\right]$$

$$\text{(C.1)}$$

$$\psi_{c2}(x,y) = \psi_c(x) \otimes \psi_c^*(y) = \left[\psi_{re}(x)\psi_{re}(y) + \psi_{im}(x)\psi_{im}(y)\right] + j\left[\psi_{im}(x)\psi_{re}(y) - \psi_{re}(x)\psi_{im}(y)\right]$$

$$\text{(C.2)}$$

Appendix D: GLCM features

Let k denote the number of gray levels of a 2D record. The GLCM P_{ij}, $0 \leq i, j \leq k\text{-}1$, can be calculated if the pixel separation $(\Delta x, \Delta y)$ is specified. The following five textural features can be obtained from P_{ij} (Hall-Beyer 2017):

1) Shannon entropy: $-\sum_i \sum_j P_{ij} \log_2 P_{ij}$. Define $P_{ij} \log_2 P_{ij} = 0$ if $P_{ij} = 0$.

   Normalized Shannon entropy (**NSE**): $NSE = -\dfrac{\sum_i \sum_j P_{ij} \log_2 P_{ij}}{2 \log_2 k}$

2) Angular second moment (**ASM**): $ASM = \sum_i \sum_j P_{ij}^2$

3) Contrast ratio: $-\sum_i \sum_j P_{ij} \cdot (i-j)^2$

   Normalized contrast ratio (**NCR**): $NCR = -\dfrac{\sum_i \sum_j P_{ij} \cdot (i-j)^2}{(k-1)^2}$

4) Correlation (**COR**): $COR = \sum_i \sum_j \left[\dfrac{(i \cdot j) \cdot P_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y} \right]$

5) Inverse difference moment (**IDM**, local homogeneity): $IDM = \sum_i \sum_j \dfrac{P_{ij}}{1 + (i-j)^2}$

These textural features quantify the irregularity of the record: a record with high values of *NSE* and *NCR*, and low values of *ASM*, *COR* and *IDM* is usually irregular.

Appendix E: Additional corrosion fields modelling

Additional examples for simulating the corrosion fields are presented based on the target scanned corrosion features. In all cases, the conclusions drawn from the example shown in the main text are equally applicable to the results for the examples presented in this appendix.
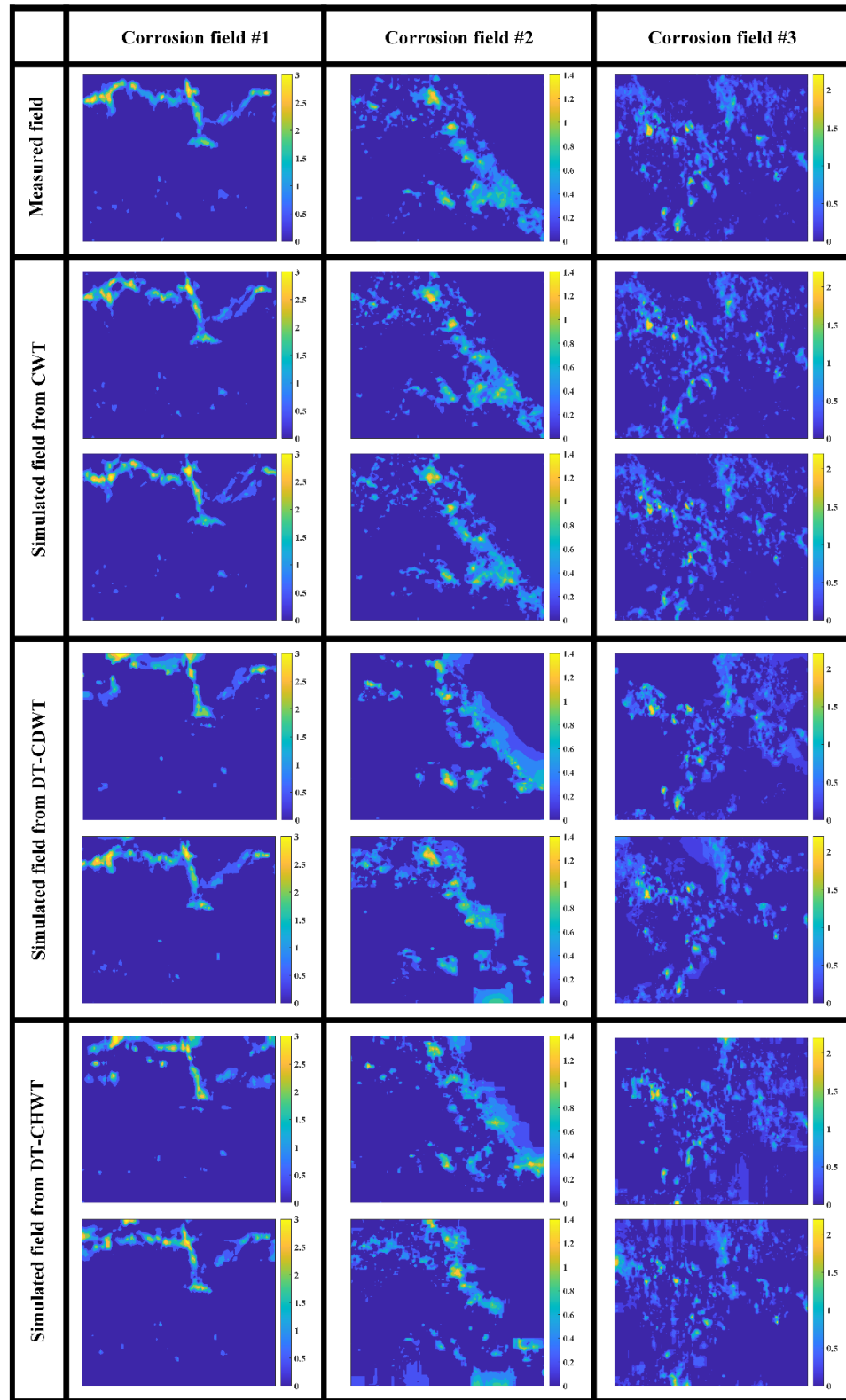
Figure E. 1 Measured and simulated corrosion fields based on three transforms
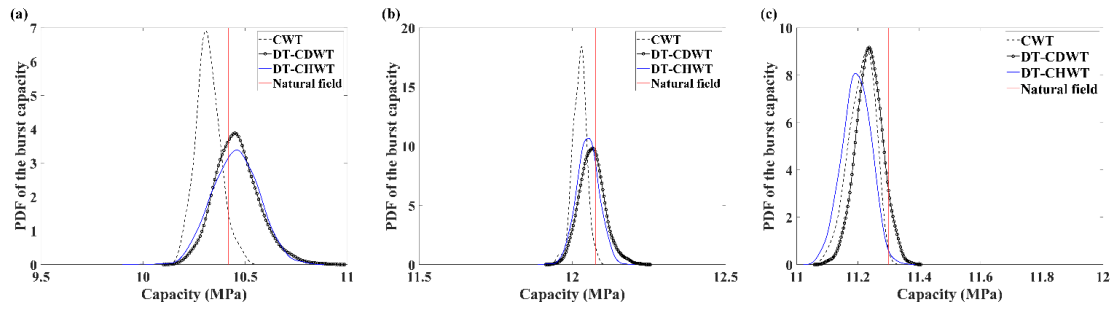
(unit: mm)

Figure E. 2 Empirical PDF of the burst capacity of pipe segment obtained from realizations of the simulated field: (a) Corrosion field #1; (b) Corrosion field #2; and (c) Corrosion field #3
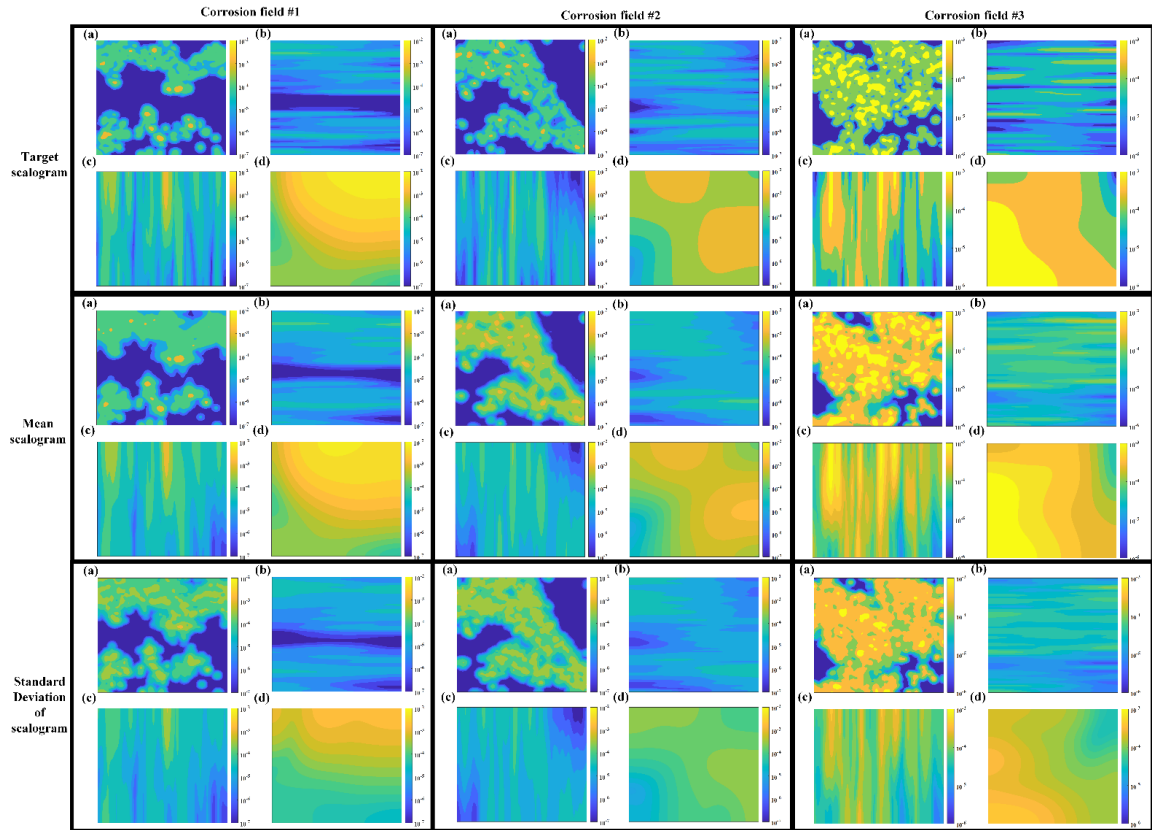
Figure E. 3 Target, mean, and standard deviation of CWT scalograms of 100 simulated corrosion fields (unit: mm^2): (a) $s_x/s_{x,min} = 1$, $s_y/s_{y,min} = 1$**; (b)** $s_x/s_{x,min} = 16$, $s_y/s_{y,min} = 1$**; (c)** $s_x/s_{x,min} = 1$, $s_y/s_{y,min} = 16$**; and (d)** $s_x/s_{x,min} = 16$, $s_y/s_{y,min} = 16$
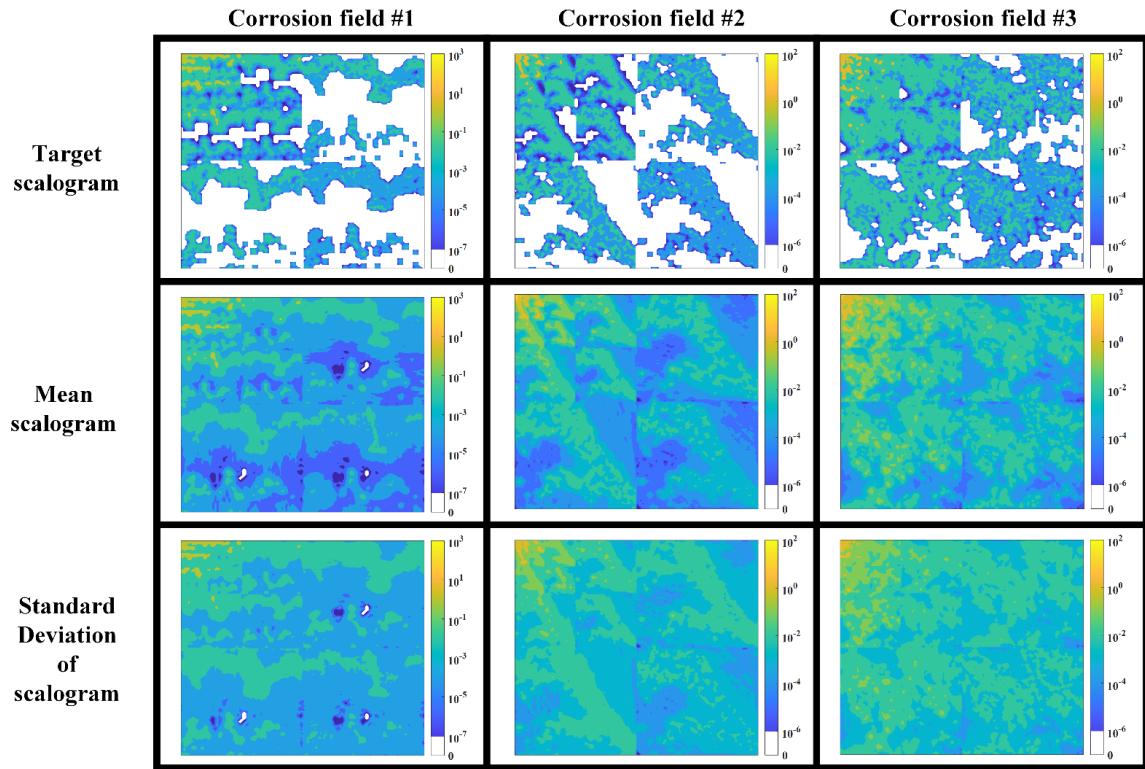
Figure E. 4 Target, mean, and standard deviation of DT-CDWT scalograms of 100 simulated corrosion fields (unit: mm^2)
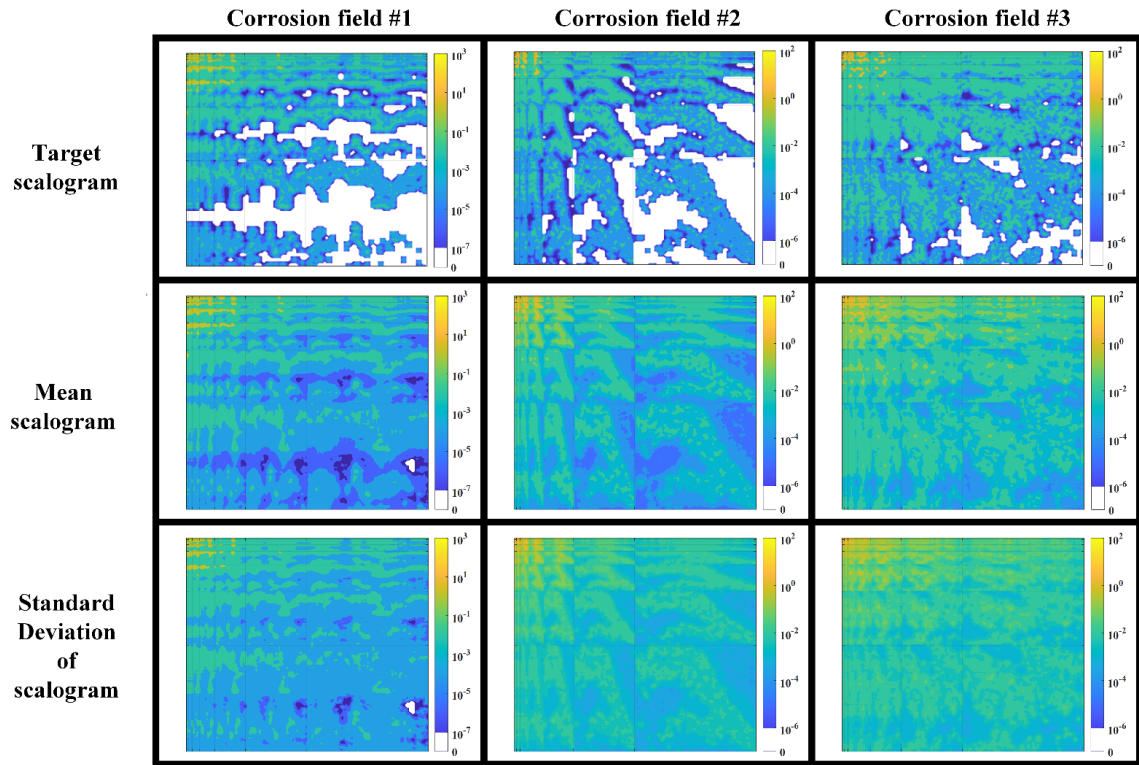
Figure E. 5 Target, mean, and standard deviation of DT-CHWT scalograms of the 100 simulated corrosion fields (unit: mm^2)

The COVs of the burst capacities of simulated fields obtained from corrosion field #1 using DT-CDWT and DT-CHWT are about 1.3%, while the COV of other cases are below 1% (the same pipe geometric and material properties as described in Section 5.3 are employed in RSTRENG to evaluate the burst capacity).

Table E.1 Mean texture features of simulated and measured corrosion fields. (Δx, Δy) = (1, 0)

| Corrosion samples		NSE ($\times 10^{-1}$)	ASM ($\times 10^{-1}$)	NCR ($\times 10^{-3}$)	COR ($\times 10^{-1}$)	IDM ($\times 10_{-1}$)
Corrosion field #1	Measured field	1.731	6.367	0.901	9.095	8.433
	CWT	1.759	6.313	1.012	8.981	8.345
	DT-CDWT	1.650	6.533	0.873	9.116	8.735
	DT-CHWT	1.711	6.418	0.947	9.040	8.538
Corrosion field #2	Measured field	1.617	5.868	0.319	8.686	8.462
	CWT	1.657	5.837	0.374	8.455	8.367
	DT-CDWT	1.567	6.058	0.316	8.694	8.770
	DT-CHWT	1.620	5.913	0.381	8.421	8.587
Corrosion field #3	Measured field	2.284	4.432	1.141	7.069	7.334
	CWT	2.303	4.399	1.213	6.882	7.250
	DT-CDWT	2.212	4.665	1.149	7.037	7.771
	DT-CHWT	2.282	4.440	1.287	6.675	7.435

Table E.2 Mean texture features of simulated and measured corrosion fields. (Δx, Δy) = (0, 1)

Corrosion samples		NSE ($\times 10^{-1}$)	ASM ($\times 10^{-1}$)	NCR ($\times 10^{-3}$)	COR ($\times 10^{-1}$)	IDM ($\times 10_{-1}$)
Corrosion field #1	Measured field	1.667	6.479	0.589	9.408	8.670
	CWT	1.700	6.427	0.679	9.317	8.500
	DT-CDWT	1.628	6.559	0.686	9.304	8.764
	DT-CHWT	1.670	6.485	0.670	9.318	8.609
Corrosion field #2	Measured field	1.568	5.975	0.262	8.921	8.695
	CWT	1.625	5.902	0.322	8.672	8.475
	DT-CDWT	1.542	6.097	0.282	8.831	8.839
	DT-CHWT	1.584	5.980	0.329	8.634	8.696
Corrosion field #3	Measured field	2.125	4.774	0.594	8.475	7.858
	CWT	2.166	4.711	0.682	8.246	7.691
	DT-CDWT	2.093	4.898	0.665	8.287	8.103
	DT-CHWT	2.126	4.792	0.713	8.158	7.955

Table E.3 Mean texture features of simulated and measured corrosion fields. (Δx, Δy) = (1, 1)

Corrosion samples		NSE ($\times10^{-1}$)	ASM ($\times10^{-1}$)	NCR ($\times10^{-3}$)	COR ($\times10^{-1}$)	IDM ($\times10_{-1}$)
Corrosion field #1	Measured field	1.788	6.253	1.325	8.677	8.321
	CWT	1.812	6.205	1.476	8.522	8.229
	DT-CDWT	1.717	6.418	1.370	8.607	8.560
	DT-CHWT	1.774	6.291	1.420	8.547	8.346
Corrosion field #2	Measured field	1.701	5.707	0.455	8.131	8.232
	CWT	1.735	5.662	0.526	7.841	8.130
	DT-CDWT	1.645	5.909	0.461	8.090	8.539
	DT-CHWT	1.700	5.736	0.545	7.738	8.325
Corrosion field #3	Measured field	2.337	4.301	1.391	6.446	7.173
	CWT	2.358	4.242	1.491	6.183	7.059
	DT-CDWT	2.265	4.536	1.385	6.437	7.583
	DT-CHWT	2.330	4.298	1.554	5.979	7.240

References

Antonini, M., Barlaud, M., Mathieu, P., & Daubechies, I. (1992). Image coding using wavelet transform. *IEEE Trans. Image Processing*, *1*, 20-5.

Hall-Beyer, M. (2017). *GLCM texture: A tutorial* v. 3.0 March 2017.

Masters, F., & Gurley, K. R. (2003). Non-Gaussian simulation: cumulative distribution function map-based spectral correction. *Journal of engineering mechanics*, *129*(12), 1418-1428.

Selesnick, I. W., Baraniuk, R. G., & Kingsbury, N. C. (2005). The dual-tree complex wavelet transform. *IEEE signal processing magazine*, *22*(6), 123-151.

Curriculum Vitae

Name:	Junxiong Lin
Post-secondary Education and Degrees:	South China University of Technology Guangzhou, Guangdong, China 2012-2016 B.Eng. South China University of Technology Guangzhou, Guangdong, China 2016-2018 M.ESc. Western University London, Ontario, Canada 2019-2023 Ph.D.
Honours and Awards:	Western Graduate Research Scholarship 2019-2023
Related Work Experience	Teaching Assistant and Research Assistant Western University 2019-2023

Publications:

Articles in Refereed Journals

[1] **Lin, J**., & Zhou, W. (2023). Simulating nonhomogeneous non-Gaussian corrosion fields on pipelines based on in-line inspection data. *Structure and Infrastructure Engineering* (accepted).

[2] **Lin, J**., Zhou, W., Cui, X. Z., & Hong, H. P. (2023). Application of wavelet transforms to the simulation of corrosion fields on buried pipelines. *Computers & Structures*, *276*, 106957.

Conferences Papers

Lin, J., & Zhou, W. (2022, September). Fractal analyses of metal loss corrosion on buried pipelines. In *13th International Conference on Structural Safety and Reliability (ICOSSAR 2021-2022)*.