

# Pipeline Inspection Analytics

*Learning from weak signals to predict the degradation risk level*

*Report by L. Querella, PhD, MSc, Data Scientist, Mar 2021*

Author: Laurent Querella, PhD, MSc  
Senior Data Scientist (freelance) at TotalEnergies

Note: The details of this work were presented in applied mathematics conference at TotalEnergies and at two other international conferences (Paris, Frankfurt)

Disclaimer: Confidential information are not included in this report

Version: Mar 2021

## Table of Contents

1	Abstract .....	3
2	Introduction.....	4
2.1	Context .....	4
2.2	Objectives .....	7
2.3	Inspection data – Techniques.....	8
3	Challenges .....	10
4	Data science approach .....	11
4.1	Audit and descriptive statistics.....	11
4.2	Data fusion .....	12
4.3	Data enrichment.....	13
4.4	Geospatial visualisation of results .....	15
4.5	Machine learning modelling.....	16
4.5.1	Segments and singular tubes.....	16
4.5.2	Unsupervised learning.....	16
4.5.3	Classification of anomaly types .....	18
4.5.4	Predictive model of the corrosion rate .....	19
4.6	Degradation risk level model.....	19
4.6.1	Likelihood of leak occurrence.....	20
4.6.2	Baseline weights .....	20
4.6.3	Computation of risk level .....	22
4.6.4	Domain expert analysis .....	23
4.7	Key Success Factors .....	24
5	Decision-support tool for inspection.....	25

## 1 Abstract

Inline inspection and external inspection of oil and gas pipelines are crucial to detect various types of anomalies and degradation which can possibly yield failures with hazardous consequences, economic losses, and irreversible human and environmental damage. So far it is not possible to predict when and where a failure would occur.

Most of the existing models consider only a limited number of factors such as corrosion or fatigue wear and only a few attempts toward a comprehensive modeling of pipeline condition have been made in the past decade.

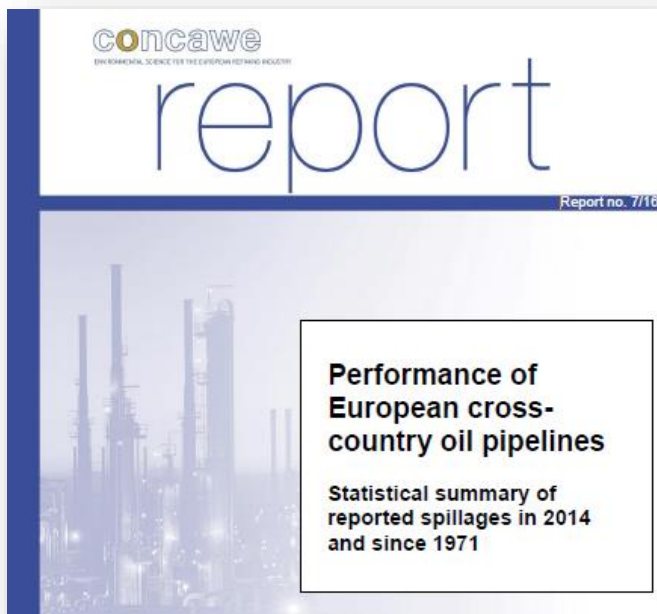
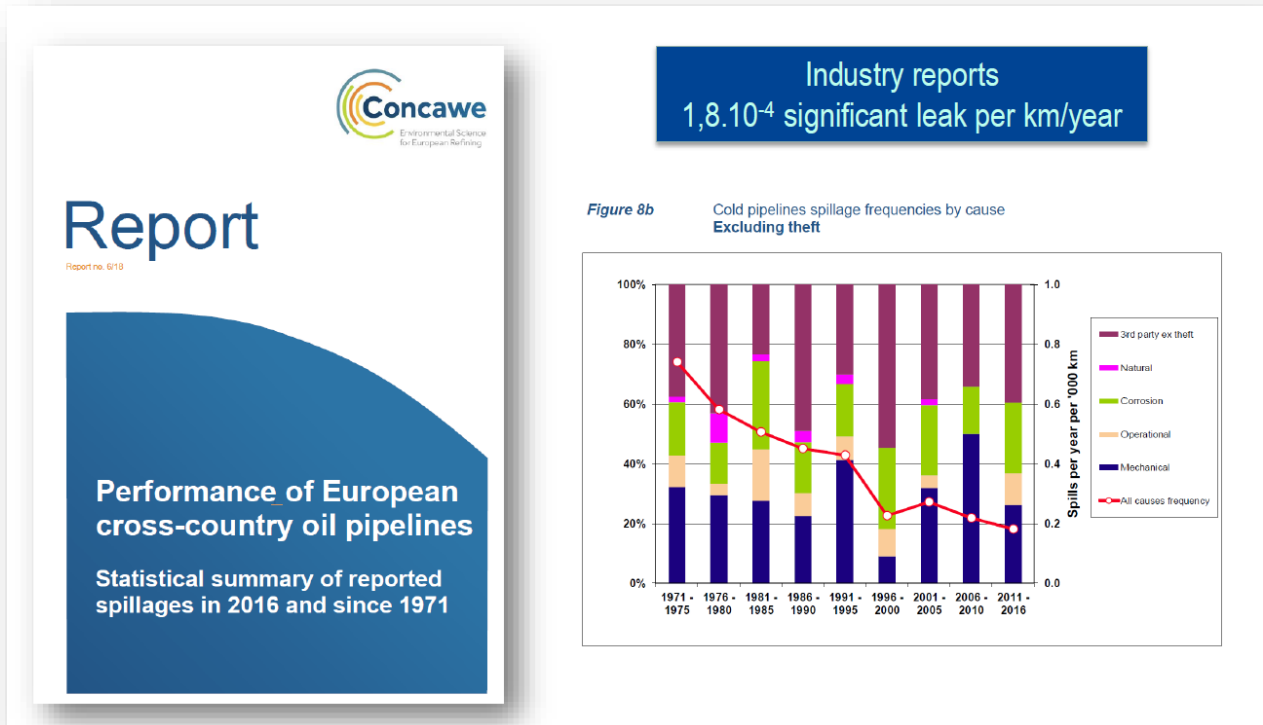
We tackle this problem with the data science methodology and tools applied to a major French oil pipeline. More specifically,

- We first perform a geospatial and temporal data fusion of relevant heterogeneous sources (pipe tally, inspection, weather, environment, ...) yielding a multidimensional view of the pipeline hosted in a geographical information system.
- Then, we develop machine learning models to classify anomaly types and predict the degradation risk level at each pipeline segment or at singular tubes.

The outcome will help the pipeline stakeholders (inspection and operators) to prioritize their inspection efforts by highlighting significant likelihood of failure stemming from the correlation of weak signals that would otherwise remain unnoticed.

## 2 Introduction

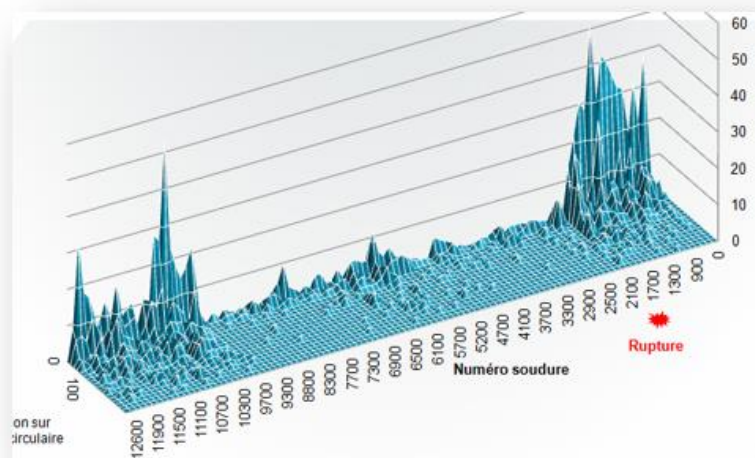
### 2.1 Context



Industry reports  
 $2.10^{-4}$  significant leak per km/year  
due to corrosion

## INSPECTION

- 3rd party inline inspection reports missed anomalies



## ANOMALIES

- Small undetected anomalies/defects can cause major impacting events



## HAZARDS

- Safety
- Financial
- Environmental
- Reputation

## 2.2 Objectives

### Improve pipelines integrity using advanced analytics

#### Average leak rate reduction



50% reduction → ~1M€/year savings & reputation (all RC pipelines)

#### Decision support tool to improve inspection plan



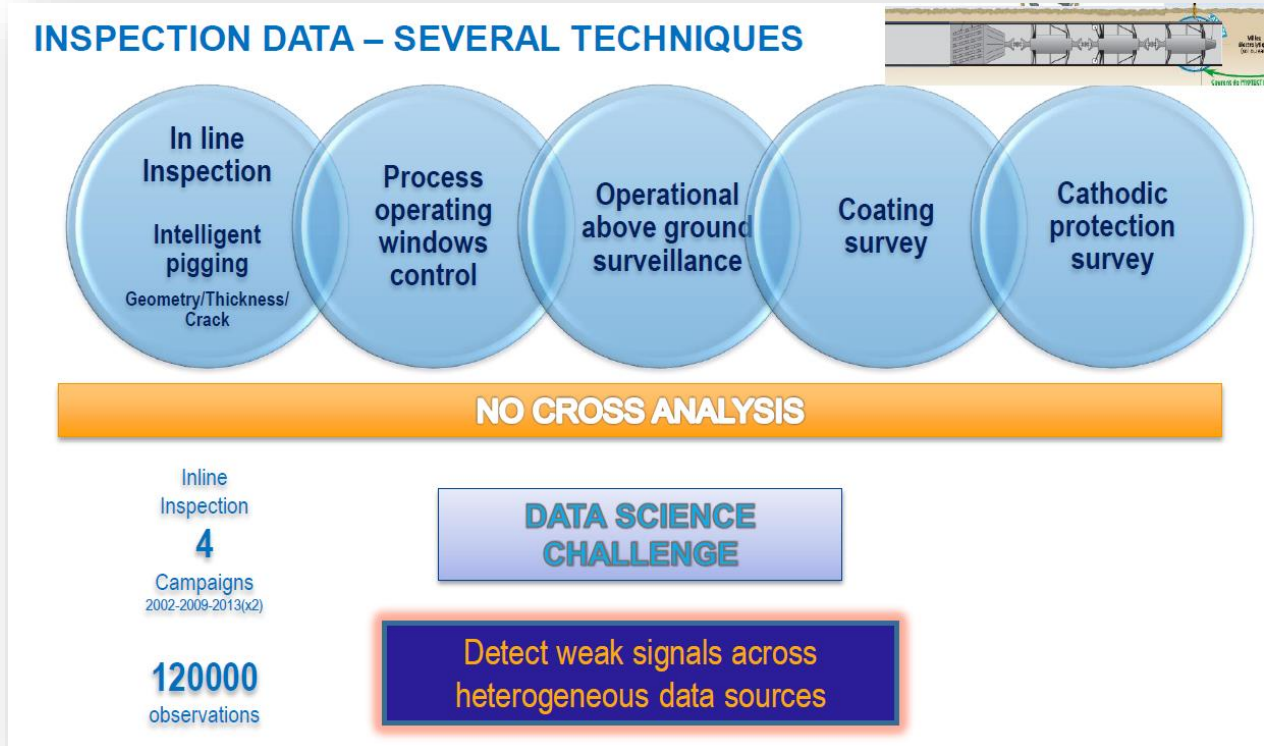
Prioritise dig out based on degradation risk model results  
→ top risk segments/tubes to investigate (~<30k€ cost per dig out)



High risk resulting from weak signals interactions  
→ unnoticed with conventional approaches



## 2.3 Inspection data – Techniques





# INSPECTION & DETECTION OF ANOMALIES

## ANOMALIES

Anomalies detected on the PLIF pipeline are mainly linked to

- ▶ **Deformation**

Undetected → damage to coating / corrosion / cracks / ...



- ▶ **Thickness** (corrosion)



- ▶ **Crack**



## INSPECTIONS

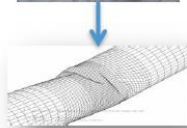
- ▶ **External inspection**

- Above ground survey (walkers, airplanes)
- Work permits
- Leak of electrical flow...

- ▶ **Internal inspection**

- **PIG** (Pipeline Intelligent Gauge ?) with three kinds of scrapers/crawler-based robotic tools:

Geometry tool



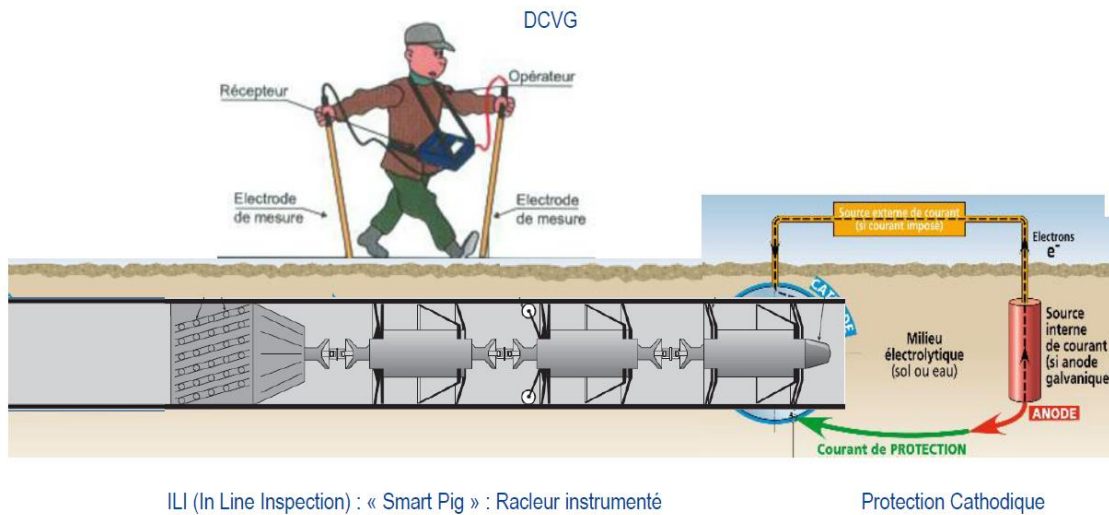
UT Thickness measurement tool



UT Crack detection tool



## INSPECTION – additional information



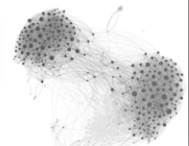
## 3 Challenges

### Improve pipelines integrity using advanced analytics

How to combine data from multiple sources to get a multi-dimensional inspection view?

Completed

Uncover patterns not revealed by conventional analytics



## Improve pipelines integrity using advanced analytics



Merge data from highly heterogeneous sources to get a multi-dimensional inspection view

COMPLETED



Geospatial & temporal fusion of all available data



I. Classify anomaly types reported by inline inspection tools (scrappers) ✓



Create a reliable predictive model of degradation risk level from available data

COMPLETED

II. Modeling the corrosion rate ✗ - modeling the severity of corrosion defects ✓

III. Modeling the degradation risk level ✓

IV. Leverage the risk model as a decision-support tool for inspection



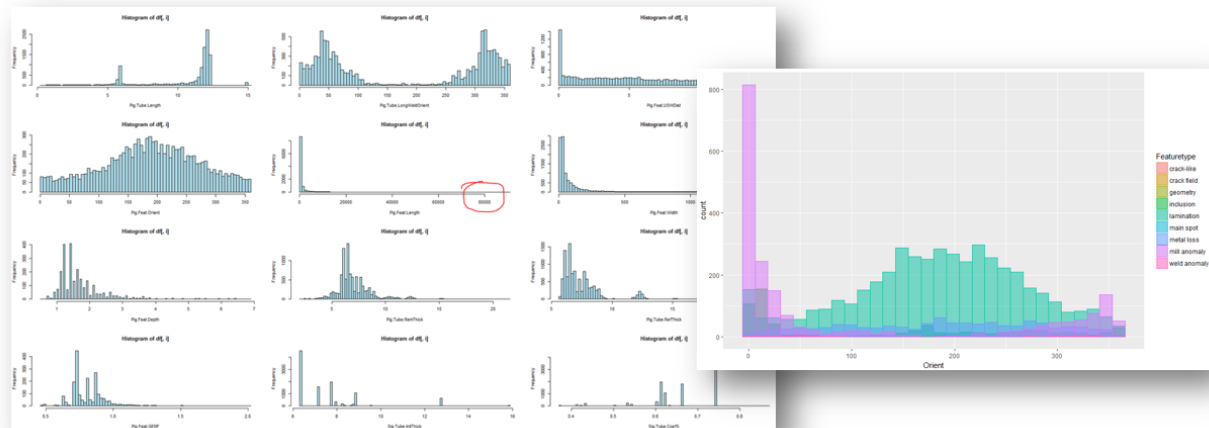
## 4 Data science approach

### 4.1 Audit and descriptive statistics

#### Inline inspection

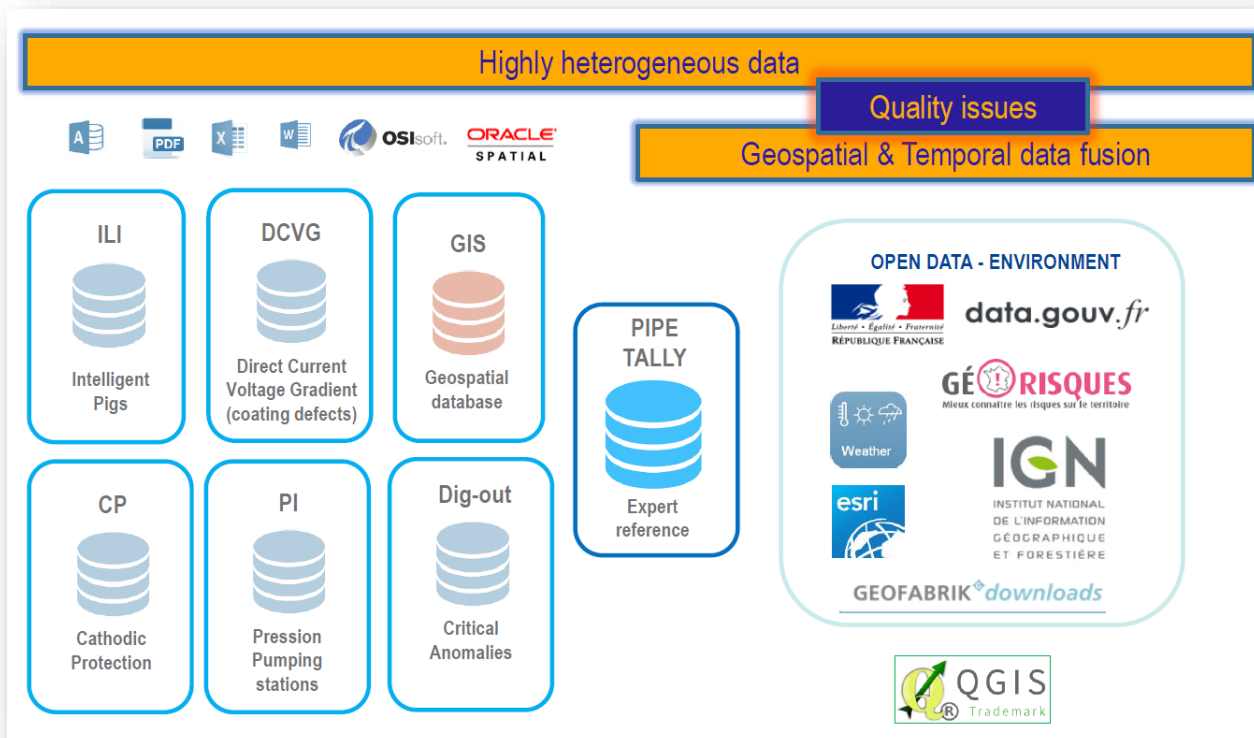
Uncover patterns & correlations – Outliers

## Multi-dimensional view of the pipeline @Tube level



Evolution of individual anomalies and analysis with domain experts.

## 4.2 Data fusion



### 4.3 Data enrichment

How to combine data from multiple sources to get a multi-dimensional inspection view?

Highly heterogeneous data

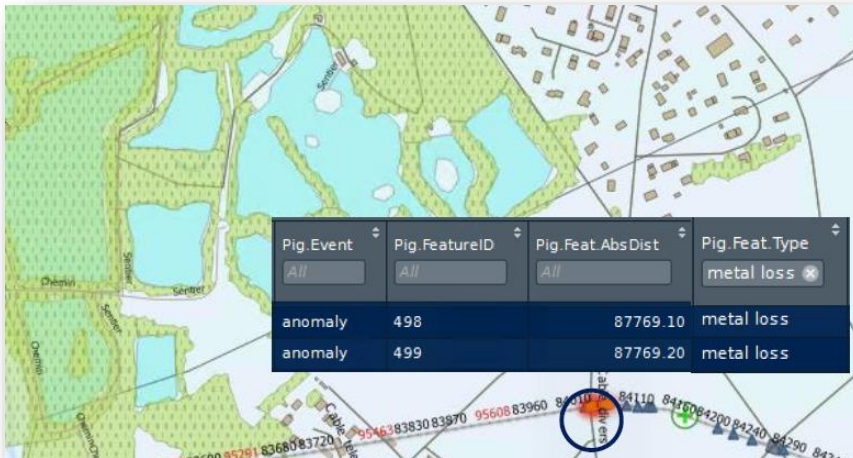


Sig.Tube.Length	Sig.Tube.Nuance	Sig.Tube.WeldUS	Sig.Tube.WeldDS	Sig.Tube.Ref
All	All	All	All	All
11.67	X60	4875	4876	1603
11.67	X60	4875	4876	1603

Sig.Tube.Crossing	Sig.Tube.Compens
All	All
rue des tritons	Dalle BA Longueur : 14 2m*2m Epaisseur : 15cm .

## Geospatial & Temporal data fusion



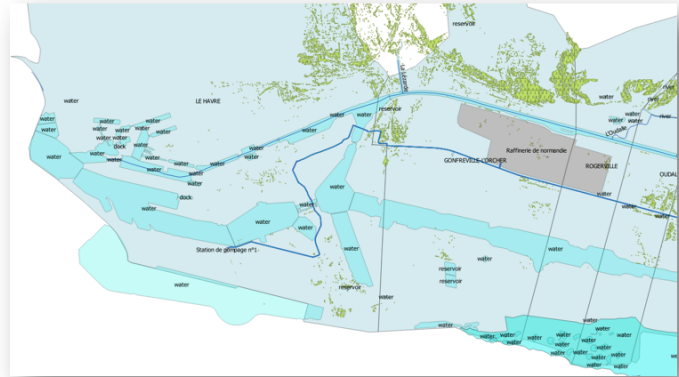
## GEOSPATIAL DATA FUSION

### + Data enrichment

#### Geospatial merge of heterogeneous data

Pipe Tally + GIS + Inspection (Pig, DCVG, PC)

- + geographic data (roads, waterways, ...)
- + weather conditions (local stations)
- + soil condition (clay, sand, ...)
- + pressure



### + Feature engineering

#### Density and evolution of anomalies

Short- and long-term corrosion rates – Dent – Cathodic protection – Coating defects - ...

### Result = input dataset for machine learning modeling

Pipeline/Tube characteristics + Inspection data + Environment data



Each tube is geolocalised & any characteristic (inspection data, environment, ...) is easily visualised on GIS (e.g. fatigue, soil nature, repairs, defects, railways, roads, waterways, ...)

→ enriched data sets → @tube & @pig anomaly → machine learning

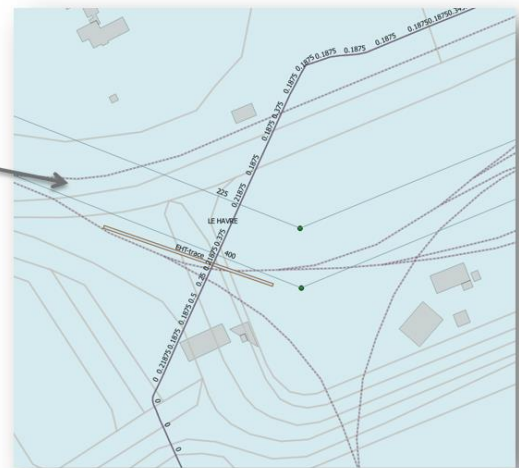
~200  
variables

25334  
tubes

120000  
Pig records

#### 4.4 Geospatial visualisation of results

##### Quantum GIS



All relevant pipeline and environmental characteristics are assigned at each point on the pipeline  
**Risk level prediction** can be visualized and most critical points identified on the map (color code)

## 4.5 Machine learning modelling

### 4.5.1 Segments and singular tubes

#### Definition of granularity for the degradation risk level

##### Focus on segment

Homogeneous segmentation based on statistical clustering combined with domain expert knowledge or arbitrary split every 100m

##### Focus on singular point

Tube with at least one “crossing” of influencing environmental element (river, electric line, railways, ...) ~20% tubes



### 4.5.2 Unsupervised learning

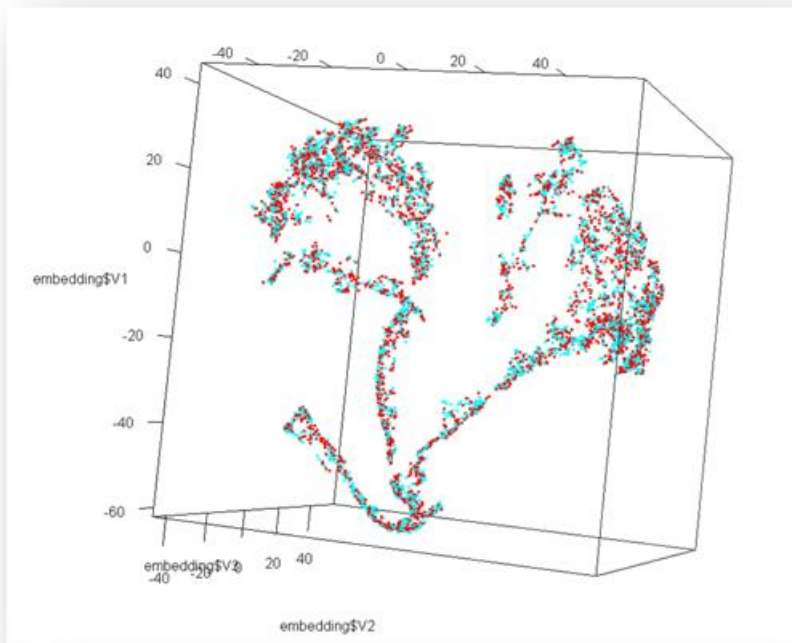
Unsupervised learning – Clustering tubes in homogeneous families (*segments*)

#### Inspection data – Unsupervised learning

##### Exploratory

PCA – t-SNE



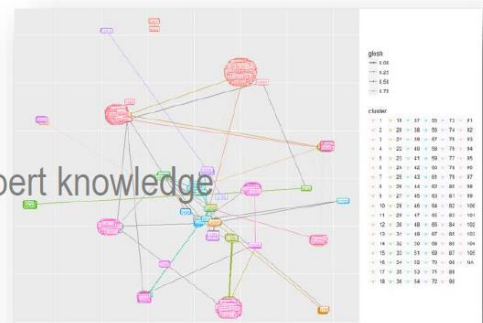


## Segment

Homogeneous segmentation based either on

- dynamic cut
- statistical clustering combined with domain expert knowledge

→ segment as extra tube attribute



LargeVis & HDBScan algorithms

#### 4.5.3 Classification of anomaly types

##### Classify anomaly types reported by ILI

##### Multi-class classification on 2013 campaign

*Extreme gradient boosting (trees) – xgboost*

##### Almost perfect classification of anomaly types

- Consistency of anomaly reporting (ILI owner) based on interpreted data above reporting threshold
- Possible improvement (real added value): classification model based on **raw ILI data**
  - challenging the results of ILI reports

Confusion Matrix and Statistics

Prediction \ Reference	dent	geometry	inclusion	lamination	main.spot	metal.loss
dent	45	0	0	0	0	0
geometry	0	68	0	0	0	0
inclusion	0	0	86	1	0	0
lamination	0	0	0	1480	1	0
main.spot	0	0	0	0	7	0
metal.loss	0	0	0	0	0	480

Overall Statistics

Accuracy : 0.9991  
95% CI : (0.9967, 0.9999)  
No Information Rate : 0.6831  
P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.9981  
McNemar's Test P-value : NA

##### Influencing factors

Tube/anomaly characteristics (thickness, radial position, ...)

No external variables

#### 4.5.4 Predictive model of the corrosion rate

Create a reliable predictive model of degradation risk level from available data

##### II. Modeling the corrosion rate & severity of corrosion defects

Elasticnet and xgboost algorithms  
with regularization

##### II-1 Predictive model of the **corrosion rate @tube level** ❌

Evolution of thickness measured by inline inspection:

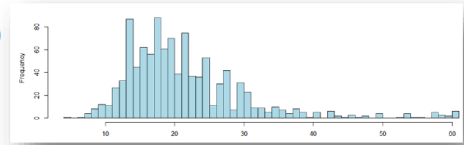
→ ILI measurement errors, initial thickness not reliable, 2 campaigns (2009/2013)

Noise >> Signal

##### II-2 Predictive model of the **depth percentage @metal loss defect** ✓

Target variable = **depth percentage %** (= 1 – remaining thickness %)

→ Regression coefficients and top influencing factors (pos & neg)  
e.g. top soil cation exchange, forest, natural zone znieff, ...



→ When a metal loss defect is detected we can predict  
its “severity” (depth%) with 4-5% error

#### 4.6 Degradation risk level model

Data-driven + Domain expertise + Heuristics

#### 4.6.1 Likelihood of leak occurrence

##### III. Modeling the degradation risk level

Predict likelihood of occurrence of leakage @Segment → @Tube

Target variable = degradation risk level (i.e. likelihood of critical defect/leakage)

$$\text{Risk level} = \sum \beta_j X_j$$

$X_j$  = Risk attributes (tube characteristics, inspection, environment factors)

$\beta_j$  = Weights

Most critical ILI defects → dig out campaigns → only 110 tubes stamped with a label (74 confirmed defects)

→ Baseline model from domain expert knowledge & heuristics (no machine learning here)

→ Top risk tubes identified on map: known spots & new zones to examine

→ Validated by domain experts based on current knowledge

Next ILI (Nov'18): opportunity to confirm newly predicted high risk zones



#### 4.6.2 Baseline weights

##### III. Modeling the degradation risk level

$$\text{Risk level} = \sum \beta_j X_j$$

$X_j$  = Risk attributes (tube characteristics, inspection, environment factors)

$\beta_j$  = Weights

→ How to set the baseline weights?

##### A. Qualitative information

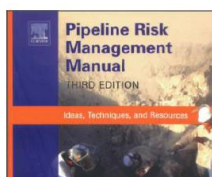
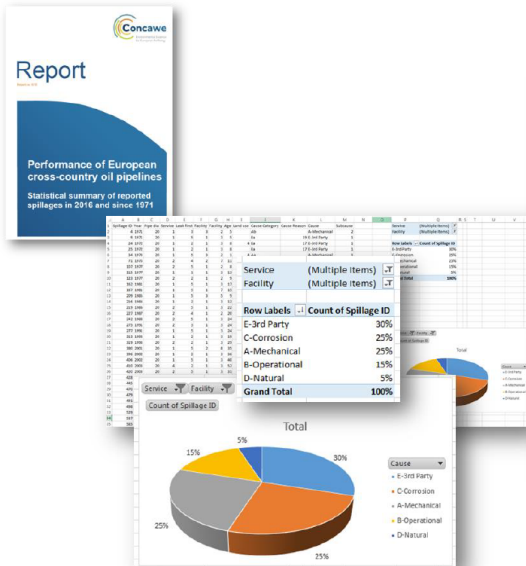


Table 1.1 Cause categorisation tree

Primary	Secondary	Reason
A Mechanical	Ab Design and Materials	1 Incorrect design
		2 Faulty material
	Aa Construction	3 Incorrect material specification
		4 Age or fatigue
		5 Faulty weld
		6 Construction damage
		7 Incorrect installation
B Operational	Ba System	8 Equipment
	Bb Human	9 Instrument & control systems
		10 Not depressurised or drained
		11 Incorrect operation
		12 Incorrect maintenance or construction
C Corrosion	Ca External	13 Incorrect procedure
		14 Coating failure

## B. Statistics aggregation



## C. Risk matrix consolidation

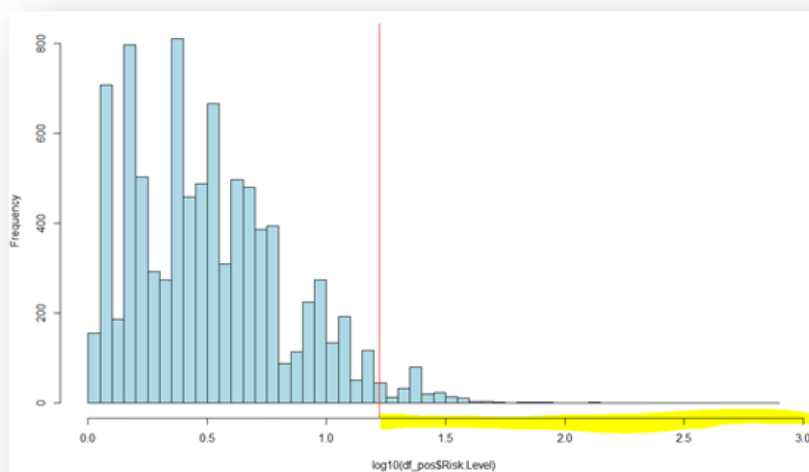
Workshops with Métier → relative importance and sign of weights

Risk Attribute	Risk category	Sub category	3rd Party	Corrosion	Mechanical	Operational	Natural
Sig.Tube.Compens	A	Aa	Y strong	Y weak < 0 / Check defect gain			
Sig.Tube.ZoneEDS	E	Ea	Urban > 0	Rural > 0			
Sig.Tube.Wear	A	Ba			levels	levels	
Sig.Tube.AgeCut	A	Ab		SD - levels			
Geo.ResTyp	E	Eb	SD - levels	complex			
Geo.ZnatTyp	D	Db	Rural < 0	Marais - type SZNEFF			levels
Geo.LandTyp	E	Ea	multiple	multiple - forest > 0			levels
Geo.Volt	D	Db		> 0			
Geo.VyID	D	Db		< 0			
Geo.VegTyp	D	Db	< 0 arboré	< 0 arboré			forest > 0
Geo.Repair	A	Aa		< 0 mitigation fouilles	< 0		
Geo.WaterCross	D	Db		> 0			
Geo.WaterProxi	D	Da		> 0			
Geo.RoadCross	E	Eb	Y > 0				
Geo.Bridge	E	Eb	Y > 0				
Geo.Tunnel	E	Eb	Y > 0				
Geo.RailCross	C	Cb	Y > 0	> 0			
Geo.RailProxi	E	Eb	Y > 0	> 0			
Geo.DefSol	D	Db		Argile > 0			
Geo.DefCl	C	Ca		levels			
Geo.PathDefect_2013	C	Ca		Complex			
Feature	B	Ba	SD combined rural	SD levels	SD levels	SD levels	
Dent_2013	B	Ba	SD		SD		
Dent.Feat.Subtype.dent_2013	B	Ba					
Dent.Feat.Subtype.dent_with_metal_loss_2013	B	Ba		SD			
Dent.Feat.Subtype.ripple_2013	B	Ba					
Dent.Feat.Orient_Type_2013	B	Ba		SD			
Dent.Feat.Area_max_2013	B	Ba		SD if metal loss			
Anomaly_2013	B	Ba		SD			
Anom.Feat.Depth_max_2013	B	Ba		SD			

#### 4.6.3 Computation of risk level

→ Computation of risk level @Tube

High risk level = top 1% percentile of all 25K tubes  
= 254 tubes



→ Out of which 90% are NOT associated with critical ILI defect



#### 4.6.4 Domain expert analysis

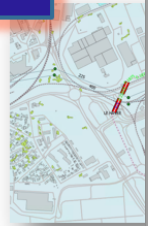
Degradation risk predictive model → top risk segments identified on map

Validation with domain experts

→ Discriminate top-risk segments from *known* risk zones

Prescriptive analytics

→ Highlight *new* high risk segments -otherwise unnoticed



Top risk zones where explanations can be given by domain experts

Examples

Well-known critical anomalies

Top risk zones not selected by conventional analyses

Examples

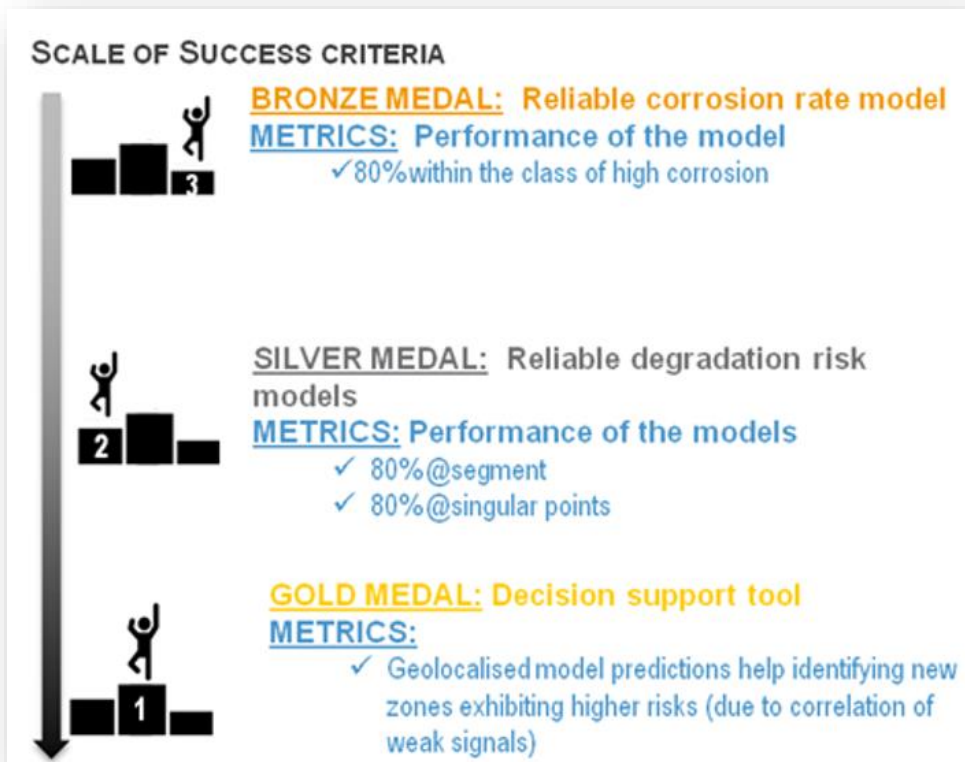


Example of risk factors:

- Metal loss defects
- Weaker cathodic protection
- Singular tube
- gaine/pipe contact

The result details are confidential and can't be disclosed in this report.

#### 4.7 Key Success Factors



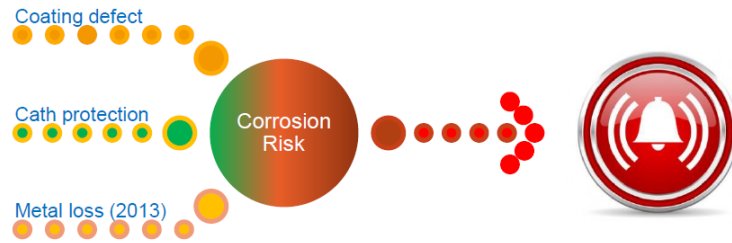


## 5 Decision-support tool for inspection

IV. Leverage the risk model as a decision-support tool for inspection

→ Top risk tubes identified on map: new zones to examine

Additive/multiplicative effect of weak signals → high risk alert (illustration for corrosion)



Decision criteria to conduct dig out

- Singular tubes
- Impact in case of hazard
  - Urban, protected natural zone, ...
- Accessibility & cost of dig out

QGIS available