Intelligent ILI Alignment, Growth & Interaction Zone Analysis

1. Problem Statement

Inline Inspection (ILI) runs from different years cannot be directly compared due to:

- Odometer drift (wheel slippage accumulates over distance)
- Vendor differences (Rosen vs Baker Hughes)
- Renamed or missing pipeline features
- Sensor uncertainty (ID/OD swaps, sizing tolerance)

Impact: Simple row-by-row matching produces false growth, missed corrosion, and unsafe conclusions.

2. Data Overview

Year	Vendor	Tool	Pipeline Length
2015	Baker Hughes	MFL-A/XT	~57,340 ft
2022	Baker Hughes	C-MFL	~57,445 ft

Time Interval:
2015-05-06 → 2022-02-23 = **6.80 years**

3. Groundwork & Standardization

3.1 Column Standardization

Different vendors used different names for the same physical quantity.

Examples:

- `Log Dist. [ft]` → `distance`
- `ILI Wheel Count [ft]` → `distance`
- `Wt [in]` vs `WT [in]` → `wall_thickness`

Why: Alignment and matching cannot work unless physical meaning is consistent.

4. Phase 1 — Alignment ("Common Ruler")

4.1 Why Joint Numbers Fail

If a pipe section is replaced:

- All downstream joint numbers change
- Same joint number ≠ same steel

Conclusion: Joint numbers are not physical references

4.2 The Core Idea (Rubber-Sheeting)

- Odometer error grows gradually
- Error at 10,000 ft < error at 50,000 ft
- Therefore correction must be **scaled**, not shifted

Analogy: Two stretched tape measures measuring the same hallway.

4.3 Anchor Selection (Skeleton)

We aligned datasets using **rare, physically fixed features**:

Included:

- Valves
- Tees (including Stopple / Area Start via elevation)
- Taps

Excluded:

- Girth welds (too frequent)
- Field bends (ambiguous start/end)

Why: Rare features act like "exit signs", not telephone poles.

4.4 Elevation Fingerprint Matching

Even if names differ, **elevation does not change**.

Example match:

- 2015: "Area Start Tee", Elevation 175.59 ft
- 2022: "Stopple Tee", Elevation 175.30 ft

Result: Correctly matched despite naming differences.

4.5 Alignment Results

- **87 anchor points** matched
- Drift increased smoothly from:
  - ~0 ft near start
  - ~27 ft at ~10,000 ft
  - ~130 ft near end

This confirmed **linear slippage**, validating interpolation.

Final check:
2022 corrected end ≈ 2015 end (~57,340 ft)

5. Phase 2 — Anomaly Matching (Growth Engine)

5.1 Why We Focused on Metal Loss

Anomaly Type	Grows?	Used
Metal Loss	Yes	✅
Cluster	Yes	✅
Dent	No	❌
Manufacturing	No	❌

Reason: Growth prediction only makes sense for corrosion.

5.2 Matching Logic (Multi-Factor)

Each 2015 anomaly searched 2022 anomalies using:

1. Distance (±10 ft, corrected)
2. Clock position (circular math)
3. ID/OD consistency
4. Dimensional similarity (length sanity check)

5.3 Confidence Scoring

Start at **100%**, subtract penalties:

Factor	Penalty
Distance	−2 pts / ft
Clock difference	−0.5 pts / degree
ID/OD mismatch	−30 pts
Length inconsistency	−15 pts

Why: The algorithm must know when it might be wrong.

5.4 Matching Results

Category	Count
Verified Matches (>80%)	**~530–700**
Uncertain Matches	Present (flagged)
New in 2022	~1,800
Missing	Set to zero growth

This separation is **intentional**, not a failure.

6. Growth Calculations

6.1 Why Annualized Growth

Total growth is misleading without time normalization.

Formula:

Depth_Rate = (Depth_2022 − Depth_2015) / 6.80 years

6.2 Observed Growth Statistics

- **Average depth growth:** ~0.6 % / year
- **Maximum observed growth:** ~49% total
- **High-risk tail:** >10% growth beyond tool tolerance

Insight: Averages look safe, outliers drive failure.

7. Graph Analysis

7.1 Bell Curve (Growth Distribution)

What we saw:

- Clear Gaussian shape
- Slight right shift
- Long positive tail

Interpretation:

- Center = measurement noise
- Negative growth = sensor tolerance
- Right tail = active corrosion

Conclusion: Matching is physically real, not random.

7.2 Unity Plot (Depth 2015 vs 2022)

- Diagonal line = no growth
- Gray band (±10%) = tool tolerance
- Points above gray band = real growth

Why judges care:
This visually proves algorithm accuracy.

8. Phase 3 — Interaction Zones (DBSCAN)

8.1 Why Clustering Is Needed

A single pit rarely fails a pipe.
Multiple nearby pits **interact** and act as one defect.

8.2 Why DBSCAN

Reason	Explanation
No K needed	Clusters unknown in advance

Noise handling	Isolated pits ignored
Density-based	Matches corrosion physics

8.3 Parameters Used

- `eps = 1.0 ft`
- `min_samples = 2`

Meaning:
Two or more defects within 1 ft → interaction zone.

8.4 Clustering Results

- **252 interaction zones**
- Some clusters:
  - 200 defects
  - 60% max depth
  - High combined risk

Value: Reduced 3,600 rows into 252 dig locations.

9. Dashboard Analysis (Usability)

9.1 KPI Insights

- **Active defects:** ~3,600
- **Verified matches:** ~530
- **Interaction zones:** 252
- **Max growth:** 49%

Story:
Many defects → few true threats → very few critical zones.

9.2 Key Visuals

View	Purpose
Unity plot	Validation
Risk profile	Where along pipeline
Cluster table	Where to dig
Histogram	Corrosion bloom evidence

Judges can understand risk in **seconds**, not spreadsheets.

10. Engineering Assumptions (Explicit)

Assumption	Justification
Linear slippage	Wheel wear accumulates gradually
Missing = 0 growth	Likely repaired or static
New = NaN	Unknown prior state (honest)
Elevation stable	Terrain doesn't move
Tool tolerance ±10%	Industry standard

11. Business Value Delivered

- Converts raw ILI data into **actionable dig plans**
- Reduces false alarms
- Highlights worst-case threats
- Supports integrity regulations
- Saves inspection & excavation cost

12. Future Industry Roadmap

Short Term

- Circumferential clustering (2D DBSCAN)
- Auto-generated dig sheets
- Integration with GIS

Medium Term

- Bayesian growth modeling
- Pressure re-rating simulation
- Probability-of-failure scoring

Long Term

- Real-time integrity twins
- ML-assisted re-inspection planning
- Fleet-wide corrosion benchmarking

13. Final Conclusion

This project is **not a data exercise**.
It is a **physics-aware, safety-driven integrity system**.

- Alignment solved the hardest problem
- Matching respected uncertainty
- Growth metrics were honest
- Clustering delivered real engineering value
- Dashboard made it usable