An investigation into improved ways of analysing the incidence of infrastructure damage along local road networks

Alan Yates

Cranfield UNIVERSITY

1. INTRODUCTION

- Two million road openings (“street works”) carried out every year by utility companies
- Local authorities estimate £218 million per annum spent on premature maintenance due to “trenching”
- All parties looking for ways to reduce costs and disruption through better planning and coordination

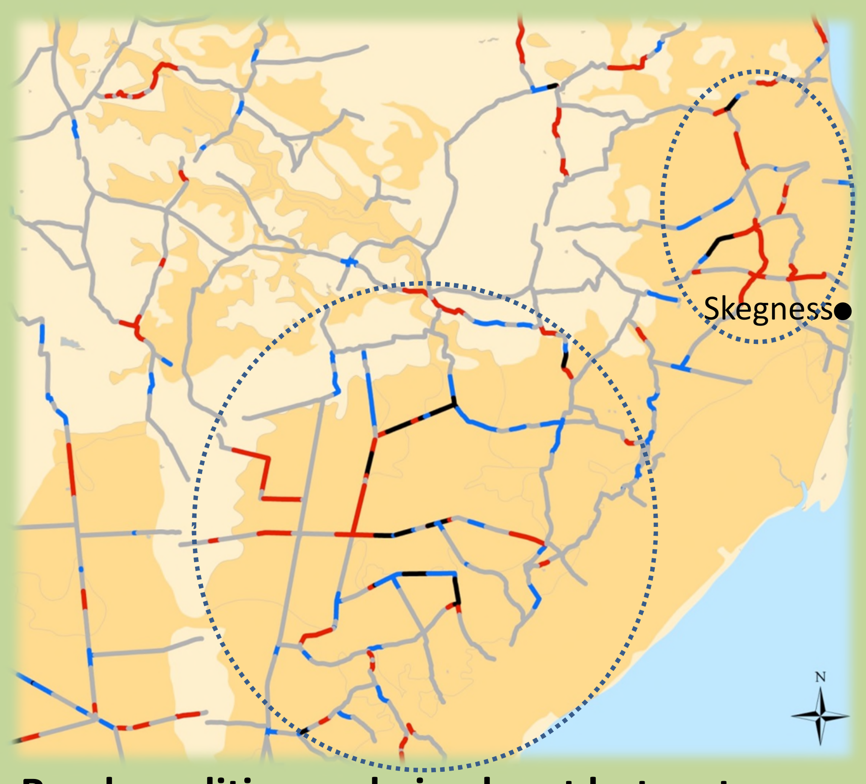
2. OBJECTIVES

Aim: An investigation into novel data mining techniques to determine the distribution and causal factors of infrastructure damage

Objectives:

- Identify a suitable approach for identifying common infrastructure failure “hotspots”
- Assess CART and Random Forest data mining techniques to identify causal factors, predict failures and assist in prioritising maintenance and repair activity
- Test techniques using road damage and water burst data for Lincolnshire Study Area (East Lindsey)

3. WHERE ARE THE HOTSPOTS IN THE EAST LINDSEY STUDY AREA?



Road condition and pipe burst hotspots (west of Skegness)

High RCI, Low Burst — High Burst, High RCI

High Burst, Low RCI — Low Burst, Low RCI

Area of common interest – consider joint planning and occupancy

High soil corrosivity — Low corrosivity

Method

- Input road network, locations of pipe bursts and locations with Road Condition Index (RCI) > 100 (poor road condition)
- Run SANET network kernel density algorithm (Okabe, 2012) (various bandwidths from 50 – 800m)

Results

- Kernel density (KD) functions across the road network
- High > 0.5 SD from mean
- Low <0.5 SD from mean
- Identified areas with high KD for RCI and bursts – Candidate for joint action
- Hotspots coincide with areas of high soil corrosivity and shrink swell potential

A Network kernel density method was used to identify shared problem locations between organisations

4. WHAT FACTORS CHARACTERISE WEEKS WITH HIGH NUMBERS OF BURSTS?

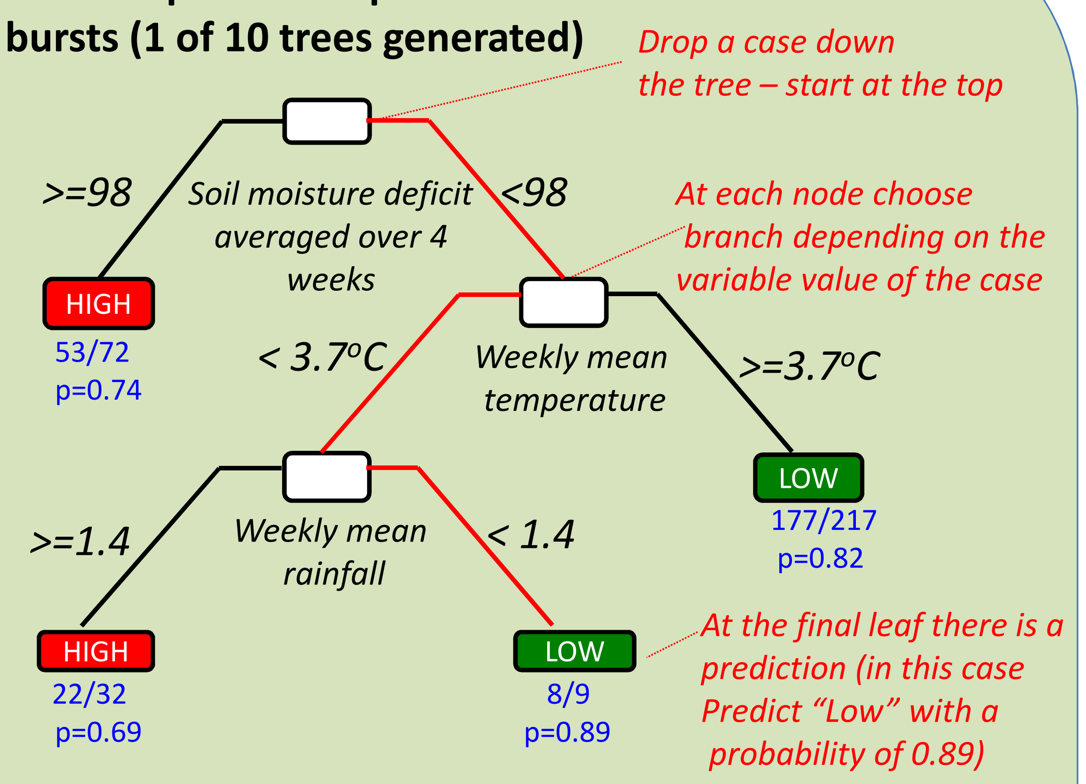
Method

- Use 9 years of weekly data recordings of soil moisture deficit, temperature and rainfall
- Train a CART decision tree to predict if number of bursts across the network in a week is >5 (High) or <=5 (Low)
- Train on 70% of available data and use 30% to validate performance

Results

- Moderate performance (accuracy, area under ROC curve) but balanced sensitivity/specificity)
- Both test and training accuracies similar (so not over-fitted)

An example tree to predict weeks with >5 bursts (1 of 10 trees generated)



	Prediction	Actual H	Actual L	Total	Accuracy	Sensitivity	Specificity	Area under ROC curve
Train	H	75	29	104	0.79	0.65	0.86	0.72
	L	41	185	226				
	Total	116	214	330				
Test	H	34	22	56	0.74	0.69	0.76	0.71
	L	15	69	84				
	Total	49	91	140				

Training and test results for the tree displayed above

A classification tree suggested causal factors and is an effective tool for exploring findings with stakeholders as it shows how it arrives at its predictions

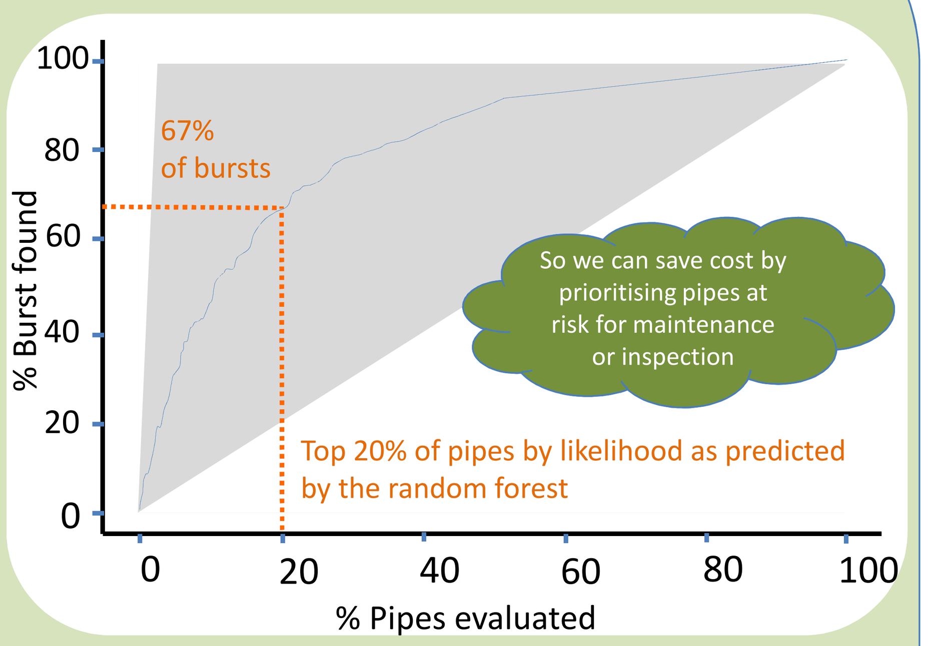
5. WHICH INDIVIDUAL PIPES ARE AT RISK OF FAILURE?

Method

- Use a “random forest” of decision trees to predict failure of individual pipes based on two years of burst records
- Use “static” factors to predict burst likelihood (e.g. pipe length, material, age, soil shrink swell potential in area) (following Harvey, 2014)

Results

- 72% accuracy overall (78% for bursts, 72% for non bursts)
- Had to optimise criteria for classifying as a burst because of class imbalance (only 2.5% of the entire population of pipes burst over 2 years)
- Ordering the pipes by probability of burst would allow selection of high risk pipes for maintenance or inspection given a constrained budget.



Prediction	Confusion Y	Confusion N	Total	Accuracy	Sensitivity	Specificity	Area Under ROC curve
Y	101	1401	1502	0.72	0.78	0.72	0.81
N	28	3589	3617				
Total	129	4990	5119				

Test results for Random Forest (optimised cut-off)

A “random forest” predicted priority pipes at high likelihood of failure

6. CONCLUSIONS

- The network kernel density method is straightforward to apply and provides an effective visual representation of “hotspots”
- Trials on predictive techniques produced reasonable results at an aggregate level, but predicting failure at individual pipe level with time series inputs is hampered by severe class imbalance
- Further research with alternative class imbalance remedies would allow definitive assessment of limits of predictive techniques
- Techniques could be used in combination to assist in co-ordinating more efficient street works

www.cranfield.ac.uk/courses/masters/environmental-informatics.html

Cranfield University, Cranfield, Bedfordshire,MK43 0AL
alankyates@googlemail.com
Supervisor: Dr Steve Hallett