British Geological Survey

NATURAL ENVIRONMENT RESEARCH COUNCIL

Environmental factors influencing pipe failures

Base Products Programme

Open Report OR/17/09

BRITISH GEOLOGICAL SURVEY

Environmental factors influencing pipe failures

A.M. Tye, C. Kirkwood, R. Dearden, B.G. Rawlins, R.M. Lark, R.L. Lawley, D. Entwistle and K. Mee

Keywords

Ferrous pipeline failure, Plastic pipeline failure, Corrosion, Statistical Modelling.

Front cover

Cover picture details, delete if no cover picture.

Keyworth, Nottingham   British Geological Survey   2017

BRITISH GEOLOGICAL SURVEY

The full range of our publications is available from BGS shops at Nottingham, Edinburgh, London and Cardiff (Welsh publications only) see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications, including maps, for consultation.

We publish an annual catalogue of our maps and other publications; this catalogue is available online or from any of the BGS shops.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as basic research projects. It also undertakes programmes of technical aid in geology in developing countries.

The British Geological Survey is a component body of the Natural Environment Research Council.

British Geological Survey offices

BGS Central Enquiries Desk

Tel 0115 936 3143   Fax 0115 936 3276
email enquiries@bgs.ac.uk

Environmental Science Centre, Keyworth, Nottingham NG12 5GG

Tel 0115 936 3241   Fax 0115 936 3488
email sales@bgs.ac.uk

Murchison House, West Mains Road, Edinburgh EH9 3LA

Tel 0131 667 1000   Fax 0131 668 2683
email scotsales@bgs.ac.uk

Natural History Museum, Cromwell Road, London SW7 5BD

Tel 020 7589 4090   Fax 020 7584 8270
Tel 020 7942 5344/45   email bgslondon@bgs.ac.uk

Columbus House, Greenmeadow Springs, Tongwynlais, Cardiff CF15 7NE

Tel 029 2052 1962   Fax 029 2052 1963

Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB

Tel 01491 838800   Fax 01491 692345

Geological Survey of Northern Ireland, Colby House, Stranmillis Court, Belfast BT9 5BF

Tel 028 9038 8462   Fax 028 9038 8461

www.bgs.ac.uk/gsni/

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon SN2 1EU

Tel 01793 411500   Fax 01793 411501
www.nerc.ac.uk

Website www.bgs.ac.uk
Shop online at   www.geologyshop.com

Acknowledgements

Contents

FIGURES

TABLES

Summary

This report details work carried out under NERC grants NE/M008339/1 and NE/NO13026/1 which were collaborations between the British Geological Survey and Yorkshire Water, with an additional knowledge transfer component involving Scottish Water and Dŵr Cymru Welsh Water. The work examines whether models developed using environmental, topographical and geohazard information could complement existing management tools, and increase the understanding as to how pipe networks of different materials interact with their broader environment. This can be seen as a first step in identifying ways in which greater resilience could be built into pipe networks.

Broad groups of pipe materials were examined, these being the cast iron and plastic pipe networks for clean water and the concrete and clay networks for waste water. Modelling was undertaken using the spatial model package 'SPATSTAT' in the 'R' statistical platform. Initially Null Models were established to predict the 'expected number of bursts' per unit area (100 x 100 m cells) for each pipe type (construction material) and water type (clean or waste) of interest

based on the density (length) of pipe present in each cell. Single covariates, identified as being important in pipe failure obtained through an Expert Elicitation process with YW, were tested against the null model. Those significant covariates ($P<0.05$) were then included in a sequential model where covariates were added and kept. In the second part of the modelling exercise, a selection of additional environmental and geohazard information were added to the expert elicitation model and a similar modelling exercise undertaken.

Overall, the modelling exercise demonstrated that for the YW region, covariates such as roads, water source, and number of dwellings (as a likely proxy for pressure changes and use) were of greater influence on the pipe network than many geohazard factors. This may be due to the YW region being fairly benign to typical factors that are recognised as damaging pipe networks such as shrink swell clays and compressible ground. For both of these covariates the highest class of hazard wasn't found in the YW region. However, soluble (gypsum bearing rocks) ground were identified as a problem for the concrete waste water network and the potential presence of sulphide was important for the cast iron network. In addition, outputs showed that pipe networks associated with the coal measures and some areas of lacustrine clay appeared to have greater than expected pipe failures, which the covariates used in the model could not explain. Results were presented as heat maps, by combining the outputs from each 100 x 100 m cell cell using the model coefficients from the final sequential model for each pipe network. This allows a spatial assessment of the overall environmental, geological and topographical 'hostility' towards the pipe network. In addition, individual heat maps for each significant ($P<0.05$) covariate were created, placed on a single scale so that the intensity of each environmental, topographical and geohazard covariate could be compared across the YW region. End of project meetings were held between BGS and the water companies to disseminate results and discuss the benefits and possibilities of the modelling approach used.

1 Introduction

This report details the results from two NERC grants examining the impacts of geological, environmental and landscape factors on the resilience of underground pipe networks. Our major partner in these Knowledge Exchange grants was Yorkshire Water (YW); one of the largest water and sewerage companies in the UK supplying 1.24 billion litres of drinking water per day. YW manage 31,300 km of clean water mains and respond to 6000-7000 bursts per year. Pipeline failures can result in loss of supply to properties, damage and closure of the public highway, closures or other inconveniences to business and the general public. Leakage (estimated to be about 275 Ml/day) from the mains water network results in loss of pressure, increased demands on water demand and treatment works (increasing carbon emissions) and water-related ground instability. YW also operate ~53,000 km of sewage pipe and respond to 200-300 sewer collapses per year. Sewer failures result in the flooding of homes and businesses with foul water and potential increases in insurance premiums for those properties affected. Importantly, sewerage leaking from collapsed sewers, also contributes to diffuse pollution of rivers and groundwater. Yorkshire Water expect to invest £140m on clean water pipe renewal, repair and cleaning in the next 5-year asset management program (AMPs) and a similar amount is anticipated to be invested in maintaining and repairing the waste water network.

YW currently utilize a Below-Ground Asset Surveyor Predictor (BGASP) model to aid maintenance of their clean and waste water pipe networks. The BGASP model uses pipe age, material, diameter, previous failure locations, temperature and basic soil properties (i.e. type) to assess whether replacement pipe is needed. The overall model assessments are suitable for YW's planning of total investment value over the AMP cycle. This project aims to develop complementary models to those currently used by YW, with the aim of providing greater understanding of how pipe networks interact with their broader environments. Information used will include spatial soil characteristics (physical and chemical), landscape analysis, geological hazard properties and additional network or environmental factors. The project draws on BGS's wealth of 1:50 000 datasets on soil properties (e.g. texture, depth, and chemistry), groundwater depth and geohazards. Algorithms applied to high-resolution digital terrain models will be used to predict indices such as local soil wetness (related to local topographic position and changes in topography). The project will utilize comprehensive

datasets of clean and waste water failures supplied by YW (about 100 000 entries) that includes a range of pipe materials (cast iron, plastic, clay and concrete). These data are assessed using statistical methods to refine and quantify a conceptual model of the factors that control pipe failure. A key aim is to examine how the models developed may yield information relevant to water companies that will enhance the maintenance and resilience of pipe networks.

2 Materials and methodology

The work was based on a (i) GIS based data pre-processing package and (ii) a modelling package. The following sections document the GIS process and the general modelling outline.

2.1 GIS BASED DATA PRE-PROCESSING

2.1.1 2.1.1 Data sources

The following YW and BGS data sets were used in the project

Yorkshire Water provided the following data

- Clean water network – shapefile (ArcGIS format)
- Waste water network – shapefile (Excel)
- Clean water failures – shapefile (Excel)
- Waste water failures – shapefile (Excel)
- BGASP sewer failures – shapefile (Excel)
- Crossings – shapefile (Excel)
- AZNP – shapefile (Excel)
- Raw water temperatures – shapefile (Excel)
- Severe weather dashboard – shapefile (Excel)
- WTW-DMA connectivity – shapefile (Excel)
- CCTV data – shapefile (Excel)
- Drainage area zone – MID/MIF (MapInfo)
- Leakage control zone-DMA – MID/MIF (MapInfo)
- Clean water area – MID/MIF (MapInfo)
- Operational area – MID/MIF (MapInfo)
- Waste water area – MID/MIF (MapInfo)
- Distribution Management Area data pertaining to water source and pipe betwork pressure

The following BGS and OS data were used in the analysis:

- GeoSure – vector (ArcGIS)
    - Collapsible ground
    - Compressible ground
    - Landslides
    - Running sand
    - Shrink-swell
    - Soluble rocks
    - Sulphide / Sulphate

    - Parent material – vector (ArcGIS)
    - Corrosivity – vector (ArcGIS)
    - Digital Terrain Model derived topographic indices (e.g. Slope, aspect, CTI, Elevation) – raster (ESRI grids)
    - OS Strategic road/rail – vector (ArcGIS)

2.1.2 Expected outputs for geostatistical analysis

The objectives of the GIS processing package were to:

- Create a grid across the study area for the analysis. The resolution selected was a 100 m grid (i.e. each grid cell was 100 m x 100 m).

- The following statistics for each grid cell
    - Length of pipe material in each cell
    - Number of clean water failures for each pipe type
    - Number of waste water failures for each pipe type

- Summarise the following for BGS data in each cell
    - Area covered by each classification for each GeoSure layer
    - Area covered by each classification in soil Corrosivity layer
    - Area covered by selected attributes from the Parent Material layer
    - Summary information for each raster dataset

2.2 DATA ASSESSMENT AND FORMATTING

2.2.1 Pipe type material

The clean water failure data lists the different type of pipe material for the pipe that failed (NB: this doesn't apply for the waste water pipes as they are all made from either concrete or clay). The clean water pipe failure data was separated into different pipe types for the

analysis grouped into similar material categories. There were 30 different pipe types and some had no code associated with them. Some of the different pipe types were deemed to be of similar composition and were grouped to make analysis simpler (e.g. Plastic). Table 1 lists the 30 categories in the original data and how they were grouped. The groups listed as Cast iron, Plastic, Asbestos Cement and Clay were used for analysis

Table 1: Pipe type codes, descriptions and materials used in models

Code	Description	Type	Used in Model
2	Copper	Clean	
16	Dummy	Clean	
4	Galvanised Steel	Clean	
12	HDPE	Clean	Plastic Model
22	HEP30	Clean	
21	HPPE	Clean	
10	LDPE	Clean	
8	Lead	Clean	
11	MDPE	Clean	Plastic Model
25	MoPVC	Clean	
24	PE100	Clean	
23	PE80	Clean	
14	Pre-Stressed Concrete	Clean	
26	PVCa	Clean	
20	PVCu	Clean	
29	Stone	Clean	
6	uPVC	Clean	
7	Asbestos Cement	Clean/Waste	Waste pipe used in Concrete model
30	Brick	Clean/Waste	
1	Cast Iron	Clean/Waste	Cast Iron Model
28	Concrete	Clean/Waste	
3	Ductile Iron	Clean/Waste	
15	Glass Reinforced Concrete	Clean/Waste	
5	Steel	Clean/Waste	
AK	Alkathene	Waste	
BL	Bitumen	Waste	
CL	Cement	Waste	Concrete Model
CC	Concrete Box Culvert	Waste	
CSB	Concrete Segments Bolted	Waste	
CSU	Concrete Segments Unbolted	Waste	
GRP	Glass Reinforced Plastic	Waste	
IS	Insituform	Waste	
MAC	Masonary, coursed	Waste	
MAR	Masonary, random	Waste	
NA	Not Applicable	Waste	
PF	Pitch Fibre	Waste	
PL	Plastic	Waste	
PSC	Plastic/Steel Composite	Waste	
PE	Polyethylene	Waste	
PP	Polypropylene	Waste	
PVC	Polyvinyl Chloride	Waste	**Plastic Model**

4

RPM	Reinforced Plastic Matrix	Waste	
RL	Resin	Waste	
SI	Spun Iron	Waste	
U	Unknown	Waste	
VC	Vitrified Clay	Waste	Clay Model

2.2.2   Formatting

Prior to GIS analysis, the Yorkshire Water data (provided in Excel spreadsheet format) –
needed to be cleaned and reformatted so that it could be imported into the GIS (ArcGIS
version 10.0). This involved ensuring that the spreadsheets were in the correct format for
import into ArcGIS – i.e. column headings must be 10 characters or fewer (made up of
alphanumeric characters and underscores only).

2.2.3   Grid creation

The analysis was carried out in a custom-built grid covering the full extent of all the data.
Since the clean water and waste water areas vary very slightly, the full Yorkshire Water
Operational Area was used to define the full extent of the grid. This area was provided by
Yorkshire Water in shape file format. The first step was to create a rectangular mesh (grid)
covering the full extent of the YW operational area. Due to the shape of the YW operating
area (roughly heart-shaped), this meant that approximately half of the cells in the rectangular
grid contained no data and since the study area was very large (180 km E-W by 160 km N-S),
this meant creating an unnecessarily large grid. To reduce processing time, any cells that did
not overlap with the YW operating area were removed from the grid. A comma separated
values file (csv) was created containing the coordinates of the new grid corners. This was
used to produce a 100 m x 100 m grid in ArcGIS.

2.2.4   Length of pipe per grid square

To calculate the length of pipe per grid square, an IDENTITY tool performed a geometric
intersection between the pipe network data and the grid. This cuts each section of pipe
network line at the boundary of each grid square, so that only the portions of pipe line that
fall into each grid square are selected. The lengths of each section of line are then calculated
and summed for each grid square.

2.2.5   Pipe material per failure

Pipe material per failure is a simple statistical assessment (frequency analysis) of pipe
material as identified within the burst datasets supplied by Yorkshire Water. Grouping of the

various pipe materials broadly followed the following e.g. the pipe groups of Cast/Ductile Iron, Plastics, concrete, copper and clay etc.

2.2.6   Processing BGS/OS vector data

The aim of processing the BGS/OS data was to summarise, for each grid square, either the total area covered by each class (e.g. Class A of GeoSure Shrink-Swell) for each dataset or, in the case of the road data, the length of road of each type. The lengths of road of each type were calculated in the same way as calculating the length of pipeline per grid square, as outlined above.

2.2.7   Processing BGS/OS raster data

Raster datasets utilising terrain data were developed from NextMap Digital Terrain Model (DTM) data (50 m resolution) held by BGS and processed using ARCGIS and Spatial Analyst (elevation, slope, aspect, compound topographic index CTI). Standard ESRI-Spatial analysis tools and formulae were used to derive the terrain coefficients of slope and aspect. The CTI is a function of both slope and the upstream contributing area per unit width orthogonal to the direction of flow and is a steady state wetness index. It was assessed using the spatial analyst tool and follows the standard formulae of:

$$CTI = \ln(a/\tan B) \qquad\qquad \text{Eq. 1}$$

where a = Upstream contributing area (m^2) and is derived using the standard tools for Flow-accumulation and B = Slope (radians) .

Back interpolation of all grid data against the h vector asset data was performed using MapInfo and Vertical Mapper, providing standard analysis of minimum, maximum, range and cell count of terrain coefficients (per object).

2.2.8   Final GIS results

The final GIS datasets consisted of separate csv files summarising the area or length of each class/road type for each dataset, with the grid cell identifier (ID). These GIS datasets summarised the length of pipes per grid cell, number of failures per grid cell and the number of pipe failures of each pipe type per grid square. These were provided for geostatistical analysis with a file containing the YW grid cell ID's and the National Grid Easting/Northing of the centre point of each cell.

2.2.9 Topographic and categorical variables used

The continuous variables used in the modelling are shown in Table 2 along with an explanation of their relevance to pipe network failure. Similarly Table 3 reports on the categorical variables used. A continuous variable is a variable that has an infinite number of possible values, whereas a categorical value can only take on a certain number of values, and in this instance are defined as classes related to a geohazard.

Table 2: The continuous variables used in the model and an explanation as to impacts on pipe networks

Variable	Description	Explanation	Source
Compound Topographic Index (CTI) or Wetness Index	This is an index determining moisture in a 100 x 100 m cell as a function of slope, aspect and the upstream contributory area. It is a steady state wetness index and is commonly used to quantify topographic control on hydrological processes.	The CTI should identify areas of ground of different potential moisture contents by examining the paths surface water may follow across the landscape. Thus cells with a high CTI may have a greater potential for waterlogging and the establishment of corrosion cells.	Calculated using Terrain Analysis on the NEXTMAP 50 x 50m Digital Terrain Model
Slope	Average Slope within a 100 x 100 m cell.	Slope steepness may dictate (i) drainage rate, (ii) ground stability or (iii) pipe movement	Calculated using Terrain Analysis on the NEXTMAP 50 x 50m Digital Terrain Model
Elevation	Mean Elevation within a 100 x 100 m cell.	The effects of elevation might indicate the (i) positioning of the water table, (ii) effects of altitude and related changes in temperature.	Calculated using Terrain Analysis on the NEXTMAP 50 x 50m Digital Terrain Model
A-road	The density of A or major road within a 100 x 100 m cell.	A-class roads are likely to carry more and heavier traffic than other roads, however, they are generally built to better specifications than B- and C-class roads. Linked to vibrations affecting pipe integrity	Vector map Open OS maps at 1:25000 to 1:50000
B-road	The density of B road within a 100 x 100 m cell	Linked to vibrations affecting pipe integrity	Vector map Open OS maps at 1:25000 to 1:50000
C-road	The density of C or minor road within a 100 x 100 m cell.	Although C-class roads are unlikely to have the quantity of traffic of other roads the placement of the pipes might not be as for the other roads. Also linked to vibrations affecting pipe integrity	Vector map Open OS maps at 1:25000 to 1:50000

A-Resistivity	This data set identifies the likely resistivity values for the main lithological resistivity type for a geological unit. It includes different environmental situation such variations in porosity and water saturation	. Resistivity is a controlling factor for the corrosion of metal pipes and is linked particularly to clay content	Datasets used are DiGMapGB, the National Geotechnical Database, field resistivity values also the Berg algorithm and expert input. Entwisle et al. 2014
B-Resistivity	This data set identifies the likely resistivity values for the secondary lithological resistivity type for a geological unit. It includes different environmental situation such variations in porosity and water saturation	. Resistivity is a controlling factor for the corrosion of metal pipes and is linked particularly to clay content	Datasets used are DiGMapGB, the National Geotechnical Database, field resistivity values also the Berg algorithm and expert input. Entwisle et al. 2014
Aspect North	Northness was computed as the sine of aspect (compass direction of slope)	The direction a slope faces is important as it can affect ground thermal regimes and the moisture content of soil. A negative Aspect North is equivalent to Aspect South.	Calculated using Terrain Analysis on the NEXTMAP 50 x 50m Digital Terrain Model
Aspect East	Eastness was computed as the cosine of aspect (compass direction of slope)	The direction a slope faces is important as it can affect ground thermal regimes and the moisture content of soil. A negative Aspect East is equivalent to Aspect West.	Calculated using Terrain Analysis on the NEXTMAP 50 x 50m Digital Terrain Model.
Dwellings	Number of dwelling per 100 x 100m cell	This variable can act as a more local proxy for how use can cause pressure changes and stress in the system	Data obtained from Office of National Statistics from 2011 census

Table 3 The categorical co-variables used in the model and an explanation as to their impacts on pipe networks.

Variable	Description	Explanation	Source
Parent Material	Dominant soil parent material type, derived from the DiGMap50 surface geology (DiGMap50Plus), in the 100 x 100m cell.	Used as a generalised description of the geology of the parent material e.g. granite, sandstone and its possible influence on corrosivity or pipe failure.	BGS DiGMapPlus-Parent Material 1:50 000 scale (previously called the soil parent material map (Lawley, 2011)
Dominant mineralogy	Dominant bulk mineralogy, derived from the DiGMap50 surface geology (DiGMap50Plus) in the 100 x 100m cell (eg dominantly carbonate/ siliceous etc).	This is a very simplified classification of mineralogy and can be used to assess whether certain mineralogy (e.g. silica rich, carbonate rich, acid (igneous)) have an influence on corrosion or pipe failure.	BGS DiGMapPlus-Parent Material 1:50 000 scale (previously called the soil parent material map. (Lawley, 2011)
G-Grain	The typical grain size of soil parent materials as from the DiGMap50 surface geology (DiGMap50Plus), in the 100 x 100m cell.	This gives an indication of the dominant particle size (clay, silt, sand) of the soil parent material or subsoil. Will provide information regarding drainage.	BGS DiGMapPlus-Parent Material 1:50 000 scale (previously called the soil parent material map (Lawley, 2011)
Soil Group	The typical grain size of surface soils (as predicted from the DiGMap50 surface geology (DiGMap50Plus),), for the 100 x 100m cell.	This gives an indication of the dominant particle size (clay, silt, sand) of the surface soil. Will provide information regarding drainage.	BGS DiGMapPlus-Parent Material 1:50 000 scale (previously called the soil parent material map (Lawley, 2011)
Engineered-materials	Classification of the parent material units for use as engineering fill (partly based on The Highways Agency series 600). A description) of these materials expected within in the 100 x 100m cell.	This provides information regarding the behaviour of the soil with respect to it being used as a backfill material e.g. presence of sulphides or sulphates	BGS DiGMapPlus-Use as Engineered Fill 1:50 000 scale (Entwisle et al. 2013)
Collapsible ground	. Collapsible ground hazard from Geosure and applied to DiGMap-50Plus.	Collapsible ground occurs in certain deposits that consolidate very rapidly when loaded and then saturated. Ground resulting strain could affect pipework and potentially weakening corroded pipe leading to failure	Obtained from the BGS collapsible ground dataset consisting of 5 hazard categories uses the DiGMapGB-50, BGS documents on the geology, the BGS National Geotechnical Properties Database. It is based on known or likely behaviour of geological units. There is some input of expert knowledge.

			Aldiss, D, Diaz Doce & Northmore 2014, Booth et al. 2010, Lee and Diaz Doce (2010, 2014)
Compressible Ground	Compressible ground hazard from Geosure and applied to DiGMap-50Plus.	Compressible ground is highly deformable under load or water removal. They include very soft clay and peat. The differences in differential movement at the interface between compressible ground and less compressible ground and the variation in compressibility within compressible ground might affect the pipes.	Obtained from the BGS compressible ground dataset consisting of 5 hazard categories. The dataset was created using DiGMapGB-50, Superficial Thickness Model, the BGS National Geotechnical Properties Database and expert knowledge. Booth et al. 2010, Jones, L D, et al. 2015, Lee and Diaz Doce (2010, 2014)
Landslides	Landslide hazard from Geosure and applied to DiGMap-50Plus.	Ground movement due to landslides, could weaken or brake pipes including corroded pipes.	Obtained from the BGS landslide ground dataset consisting of 5 hazard categories based on the geology and their likely behaviour, mapped landslides and the slope angle from a digital terrain model (DTM) and expert judgement. Booth et al. 2010, Dashwood et al. 2014, Lee and Diaz Doce (2010, 2014)
Running Sand	Running sand hazard from Geosure and applied to DiGMap-50Plus	Running sand can occurs where saturated sand or coarse silt is intercepted by an excavation or borehole. The flow of sand into the excavation can cause ground movement affecting pipe stability, could weaken corroded pipe	Obtained from the BGS running sand dataset consisting of 5 hazard categories based on the known behaviour of geological units and expert knowledge. Booth et al. 2010, Lee and Diaz Doce (2010, 2014)
Shrink swell	Shrink-swell hazard from Geosure and applied to DiGMap-50Plus	This hazard is usually identified from .Increases in water content causes swelling and drying causes shrinkage. The ground movement can damage	Obtained from the BGS Shrink Swell dataset which consists of 5 hazard categories based primarily on the modified plasticity index, which is derived

		already corroded pipes.	from the liquid limit, plastic limit and the percentage of particles less than 0.425 mm. It uses the DiGMapGB-50, BGS superficial thickness model, BGS National Geotechnical Properties Database, a simplified glacial till layer and some input from expert opinion. Booth et al. 2010, Diaz Doce et al. 2015, Lee and Diaz Doce (2010, 2014) ; Jones and Terrington (2011
Soluble rocks	Soluble rocks (karst) hazard from Geosure and applied to DiGMap-50Plus	Ground movement due to dissolution of certain geological units can damage pipework including corroded pipes.	Obtained from the BGS soluble rocks dataset that consists of 5 categories of the likely occurrence of dissolution features. It is base on the geological units (DiGMapGB-50). A digital terrain model, superficial thickness model, Glacial limits dataset, BGS superficial permeability data, known occurrences of solution features. and expert opinion. Booth et al. 2010, Farrant et al. 2015, Lee and Diaz Doce 2014
Soil Corrosivity	This is taken from the BGS dataset and classifies soils based on their corrosive properties. It is based on the CIPRA Index and applied to DiGMap50Plus The corrosion classification occupying the greatest area in 100 x 100m cell is used.	Corrosive ground can, potentially, damage some types of pipe.	Obtained from the BGS Ferrous Corrosion dataset dataset which is based on 5 categories of soil properties. It is based on the CIPRA classification scheme. . (Tye et al. 2012)
Sulphide / Sulphate	This is the BGS sulphide / sulphate dataset	The presence of sulphide minerals can cause corrosion through their oxidation and the formation of H_2SO_4. The presence of elevated sulphate is often associated with the dissolution of gypsum deposits, thus causing subsidence.	

Distribution Management Area (data)	This is YW data and is data for water source	The source of water can have an effect on pipe networks. This can either be through the chemical nature of the water or the processing that is required before entering the system.	

2.3 STATISTICAL MODELLING - OUTLINE

The data available after the Pre-Processing (section 2.1) were of two kinds. The first were records of pipe bursts for a particular pipe material and for clean or waste water, each with a particular location in space. All pipe bursts were considered over the previous 10 years (2004 to 2014) as the target variable of interest. The second were the potential explanatory factors. These were mapped on 100×100 m cells and include those data listed previously (Table 3). The set of pipe bursts were considered for a particular pipe material and water type as a realization of a spatial point process. In particular it is assumed a non-homogeneous Poisson spatial point process (Diggle, 2013). Events of such a process are mutually independent, but the expected number of events per unit area (density of the process) might vary spatially. The 'spatstat' package was used (Baddeley and Turner, 2006) in R programming language for this purpose. This allows the estimation of non-homogeneous Poisson models by maximum likelihood. The models were fitted with the density of the Poisson process (i.e. the expected number of bursts per unit area) and modelled as a function of possible environmental explanatory factors.

In order to estimate meaningful models it is necessary to define the domain in two-dimensional space in which the point process is defined. For this purpose the data on pipe distribution were used (density (length) of pipes by material and water type). A burst can only be recorded in a 100×100 m cell where the pipe of interest occurs. In the setting of the 'spatstat' package a mask could be defined from the pipe density data to define the domain within which events can possibly occur. It was also necessary to consider pipe density as an explanatory factor in the model. Pipe density varies across the Yorkshire Water area, and this inevitably induces variations in the density of the modelled spatial point process for bursts, even if external risk factors are spatially uniform. A "null" model for the density of the Poisson process, therefore, included the log density of the pipe type (construction material) and water type (clean or waste) of interest by default. The other explanatory factors could

then be considered, and assessment as to whether or not they provide additional information on the expected local density of pipe bursts.

Two alternative models for the density of pipe burst events could be compared because, in the 'spatstat' package, they are estimated by maximum likelihood. Two models are said to be nested if the simpler model can be regarded as a special case of the more complex. Thus, if model A contains only log pipe density as an explanatory factor, and model B contains log pipe density and compound topographic index (CTI), then model A is said to be nested in model B since model A is equivalent to model B with coefficients for CTI set to zero. The evaluation of the null hypothesis could then be evaluated; that CTI is unrelated to the density of pipe bursts by comparing the maximized log-likelihood for model A, l_A, with that for model B, l_B. Under the null hypothesis the statistic

$$L = 2(l_B - l_A) \qquad\qquad\qquad \text{(Eq. 2)}$$

is distributed asymptotically as chi-squared with degrees of freedom equal to the difference between the number of parameters estimated for the two models (1 here). Note that this would not be true if model A were equivalent to model B with the coefficient for CTI set at a boundary (Cox and Hinkley, 1990).

A more general comparison between models, not necessarily nested, can be made by computing Akaike's information criterion, *AIC* (Akaike, 1973), which is a measure of the relative quality of statistic models for a given set of data. If a model has *P* parameters, and the maximized likelihood for its fit is then

$$AIC = 2P - 2l. \qquad\qquad\qquad \text{(Eq. 3)}$$

It can be shown that selecting the model that minimizes AIC in some set of alternative models minimizes the expected information loss from the selection process. In this study a two-stage approach was taken to the selection of predictor variables for the non-homogeneous density of burst of pipes for a particular water type (clean or dirty) and made from a particular material. First, a list of potential explanatory factors elicited from experts at Yorkshire Water and presented in order of importance according to expert opinion. We then used this list to propose and fit test models as follows.

1. A "null" model, as described above, where the local density of bursts is a function of the log density of pipes of the target type and material in the local 100×100-m cell.

2. A series of models, each with log density of the pipe type and material and just one additional explanatory variable taken from the elicited list. Each model could be compared to the null model by means of the log-likelihood ratio.

3. A sequence of models in which each predictor was added to the set in turn, adding predictors in the order that they were presented in the elicited list. The improvement to the model achieved by adding each predictor could be tested by comparing it with the previous model in the sequence using the log-likelihood ratio statistic *L*.

On this basis the predictors, as identified by Yorkshire Water's experts, were examined, and an assessment made of the statistical evidence that they are informative about the expected density of bursts. Additional predictors not identified with factors identified by the experts were then considered.

Selecting candidate variables to add to those identified by elicitation avoided adding possible predictors correlated with variables already in the model. For this reason the correlations were examined between all available continuous predictor variables. We also examined the principal components of the correlation matrix. This allowed identification of additional predictor variables that were not correlated with predictors already in the model from the expert elicitation. Measurement of the degree of association between a categorical and a continuous predictor variable was undertaken by computing the coefficient of determination for a simple linear model in which observations corresponded to different levels of the categorical variable having different mean values of the continuous variable. The square root of the coefficient of determination is comparable to the correlations between continuous variables. Having identified a subset of additional predictors their potential value in non-homogeneous Poisson models for the density of bursts was tested in two ways:

1. As with variables identified in the expert elicitation, models were produced in which the candidate variable was the sole predictor and compared with the null model on the log-likelihood ratio.

2. Starting with the set of predictors identified from among those proposed from the elicitation, each of the additional predictors was added in turn, testing the improvement to the model on the log-likelihood ratio.

The 'spatstat' package provides a useful diagnostic for assessment of a model once it has been fitted. This is a "lurking variable" plot. The expected number of bursts within a local sub-region can be computed from the fitted model, and the difference between this number

and the actual number of bursts in the sub-region is a residual. The residuals were plotted as a map. The lurking variable plot shows the accumulated sum of residuals from south to north or west to east across the study region. The accumulated residuals over the whole region are zero, but the examination of the fluctuations of the accumulated residuals within the region with respect to an envelope shows the local deviation of accumulated residuals from zero, which is consistent with random fluctuation. This helps to identify variations in the apparent density of the process that the factors included within the model do not account for.

2.4 MODEL OUTPUTS

2.4.1 Model coefficients and maps

The model outputs come in two forms. Firstly the models are composed of a series of covariates that are ranked in order of their importance. The ranking is undertaken using log likelihood ratio (LLr) for the model with pipe density and the covariate, relative to the null model (pipe density the only covariate). Ranking on LLR is equivalent to ranking on the AIC. Continuous covariates used in the model will also have either a positive or negative sign. This indicates whether it has a positive or negative correlation to the number of pipe failures per unit length. Categorical variables are different in that a coefficient will be produced for each of the classes for the covariate. These can then be interpreted as to how they may be affecting pipe failure.

For interpretation of the coefficients of the continuous and categorical variables and their influence on failures the pipe network we use those produced when single covariates are added to the Null model. Coefficients are also produced when the covariates are added sequentially to the Null model. These coefficients are used when we assess the influence of each covariate compared to the others in the production of heat maps (see 2.4.2). These are used because in the sequential model there is a common intercept.

The second output comes in the form of maps of the modelled area along with data regarding the cumulative sum of raw residuals of the model for the X-Y co-ordinates of the spatial area being modelled. These graphs of the cumulative sum of raw residuals are known as Lurking Variable plots. Ideally the cumulative raw residuals should be within the limits imposed by the elipitical feature at zero which signifies the error in the model that can be considered random noise. Where it is beyond this ellipse, the graphs show the extent to which the model

is under or over-predicting the number of expected pipe failures per unit length (density) of pipe, indicated by the raw sum of residuals. A negative residual suggests that the model is over-predicting (blue colour in Figures) whilst a positive residual suggests that the model is under-predicting (red colour in Figures). Comparing how the different models are performing overall can be undertaken by comparing the sums of the raw residuals from the lurking variable plots, with lower values indicating improved model fits.

Heat maps

The basic model for the non-homogenous intensity of the pipe failure process takes the form

$$\lambda = \exp\{\beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots \dots\} \qquad \text{Eq. 4}$$

where β_0 is a constant intercept, β_1 is the coefficient for the first covariate, x_1 etc. For any cell in the map this will give an expected intensity of the process.

The total intensity maps were created by multiplying the individual intensity maps (each of the form $\exp(\beta_i * x_i)$). The resultant predictions of pipe failure intensity λ are extremely small numbers (e.g. 10^{-61}), so in order to make them user-friendly, the total intensity maps were scaled using the 'Scale' function in R, which for each cell subtracts the mean value of the output for all the cells and divides by the standard deviation. The scaled output values fall in the range of single to double digit numbers and are therefore more easily symbolised, labelled and interpreted using GIS software.

In addition, we can decompose equation 4 into multiplicative components $[\exp\{\beta_i x_i\}]$. These components can be used to show how any particular covariate contributes to the expected intensity of failures across the region in heat maps. Some caution must be exercised in the interpretation of these because of the possibility of correlations among the covariates. These individual heat maps have been plotted using a standardised scale using the lowest and highest model coefficients across all the significant covariates. This enables us to plot spatially the impact of each model covariate relative to each other and identify those covariates which contribute most to the overall intensity.

There is a difference between how the values for continuous and categorical covariates are obtained for each cell. For continuous covariates (e.g. slope, road type) the model coefficient obtained from the final sequential model for each covariate is multiplied by for example, the mean slope within a square or the length of a particular road type to give the intensity value.

16

For categorical covariates, the co-efficient for each class obtained from the final model is the value of the intensity given to that cell.

3 Results

3.1 EXPLORATORY DATA ANALYSIS OF DIFFERENT PIPE MATERIALS

After completion of the data pre-processing work package, a general analysis of pipe failure in the YW pipe network was undertaken. Fig 1 shows the relative frequency of clean water pipe length in the 100 x 100 m cells, where those cells in which cast iron pipe = 0 m have been omitted. The y-axis shows the probability density function. For each of the lines the sum of probabilities under each line will equal 1. So, for example, the probability that a 100 m or less of pipe in a cell is equal to the area under the line to the left of 100 m. Plotting the pipe data in this manner has the additional benefit in that it represents a check that the GIS manipulation of the pipe network was effectively handled. For example, with each cell being 100 x 100 m, it would be expected that if a cell contained only one cast iron clean water pipe and it crossed the whole cell, it's length would equal approximately 100m in length. Thus the peak in the probability distribution function at the 100 m mark suggests that this scenario existed for many of the 100 x 100 m cells.

There were also a large number of cells where no pipeline was recorded, these being more frequent for asbestos cement, ductile iron and plastic than for cast iron. This reflects the total pipe length across the YW region where Cast Iron accounts for 68% of total pipe length. Thus, it is likely that Cast Iron will have a fewer number of cells where it is not present. Other broad features can be discerned from the data. For example, the maximum length of a straight pipe in a cell is $(100 \times 100)^{0.5}$ which equals 141.2 m, so where the length of pipe in each cell is between 100 - 141 m it suggests it is crossing a cell at an angle or there are two parts of the pipe network present. When a single type of pipe has a length > 141 m within a cell it is likely that there are more than two parts of the pipe network within that cell.

Figure 1: Relative frequency of clean water pipe length (metres) in 100 x 100 m grid cells across the Yorkshire Water region for four types of pipe material
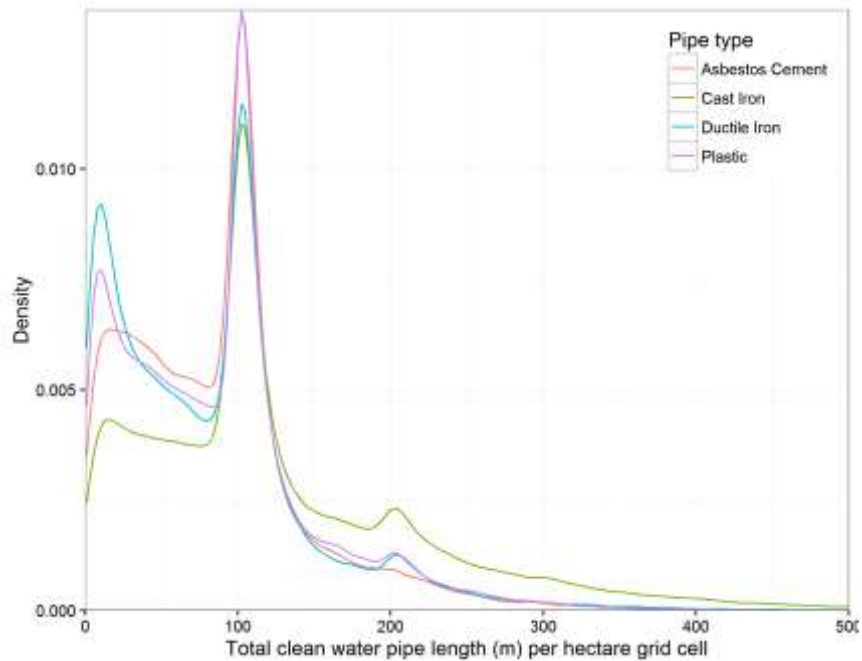
The original dataset on clean water pipe failure provided by Yorkshire Water had a total set of 89 687 failures, including a small number of failures in pipe materials comprising steel, copper, lead and glass resin. The failures for these pipe types were removed from the dataset as it was not possible to account for the proportion of the total pipe length in the 100 x 100 m grid cells (computed by BGS) in which they occurred. A small number of other pipe failures were also removed where the proportions of pipe types could not be accurately determined in each grid cells. After removing these entries there were a total of 87 162 pipe failures in 46 576 unique cells. The median and mean failure rate per cell are 2 and 3.1 respectively; the frequency distribution of failure rate per cell is positively skewed (skewness coefficient=2). The frequency distribution of total pipe length by material type per cell (Figure 2) shows that, with the exception of cast iron, the pipe types have similar frequency distributions (median length = 111-112 m); whilst in the case of cast iron, a larger proportion of cell pipe lengths per grid cell are substantially longer (median length = 193 m).

Figure 2 shows frequency of clean water pipe failures per km pipe for the 4 pipe materials. It is worth noting that there are a considerable number of variables that will contribute to pipe failure (e.g. corrosion, batches of pipe, type, dimensions) and the graphs give an overall

impression of the failures in the pipework. Cast Iron has a lower median failure rate (13.3/km of pipe) than the three other pipe types (between 17.8 and 20/km of pipe). However, cast iron is also the dominant material accounting for 68% of total pipe length; the other materials (plastic, asbestos cement and ductile iron) account for 17%, 9% and 6% of total pipe lengths (in cells with failures), respectively. This suggests that overall cast iron pipes are the most resilient material, considering that much of the network is likely to be of a greater age than more modern materials such as the plastics.

Figure 2: The frequency of clean water pipe failure rate (n failures per pipe kilometre) for four types of pipe material. Note the y axes have different scales.
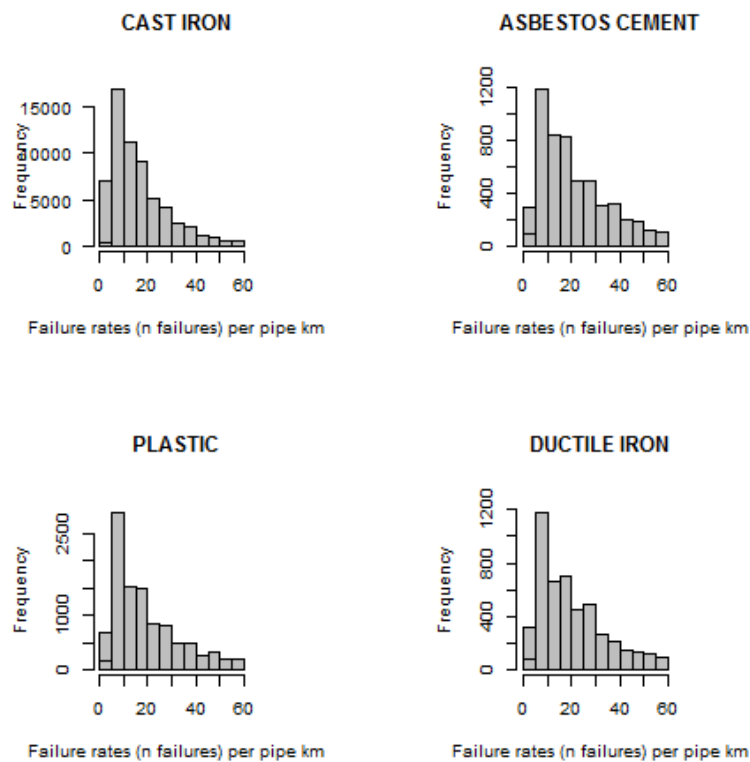
Figure 3 shows the distribution of ages when pipe failure occurs for 4 different pipe types. Again it is worth noting that cast iron is the dominant pipe type throughout the network. For Cast Iron there is a rapid increase in pipe failure after 40 years of installation. For Asbestos cement pipes there appears to be a large increase in failure 30 years after installation. For plastic it appears that, failure decreases after 10 years, although this might be because it is a more recent material (last ~40 yrs) However, the fact that >50% of failures occur in the first 10 years might reflect failures associated with the installation of plastic pipes and these will become apparent shortly after installation. There are generally mixed reports as to whether

Ductile or cast (gray) iron lasts longest. However, within this dataset ductile iron appears not to last as long, before leaks are reported. YW have identified a couple of causes for poor ductile iron performance. For example, In Bradfield (White Abbey Road) Ductile Iron pipes were installed but they contained no magnesium and so this led to failures between 1970 and 1990. It is also increasingly recognized that ductile iron corrodes in a different way to cast iron through both (i) graphitization and (ii) pitting, which means that the thinner pipe used is not as corrosion resistant as first considered. Although pitting occurs in a similar manner to cast iron, graphitization is a process where the metal constituents of the pipe degrade leaving the carbon shell structure of the pipe (Szeliga & Simpson, 2003). Graphitization is often overlooked as it may only appear as a subtle change in surface colour and can also occur under asphaltic paint pipe covering (Szeliga & Simpson, 2003). Failure often occurs after graphitisation through changes in water pressure, external loads or freezing and thawing (Szeliga & Simpson, 2003).

Figure 3: Age (years) of clean water pipe failure frequency (years) for four pipe types. Note that changes in pipe type installed with time exerts a strong influence on the age at failure; for example there are few plastic pipes older than 50 years whilst there are many cast iron pipes of ages greater than 100 years. Note the y axes have different scales.
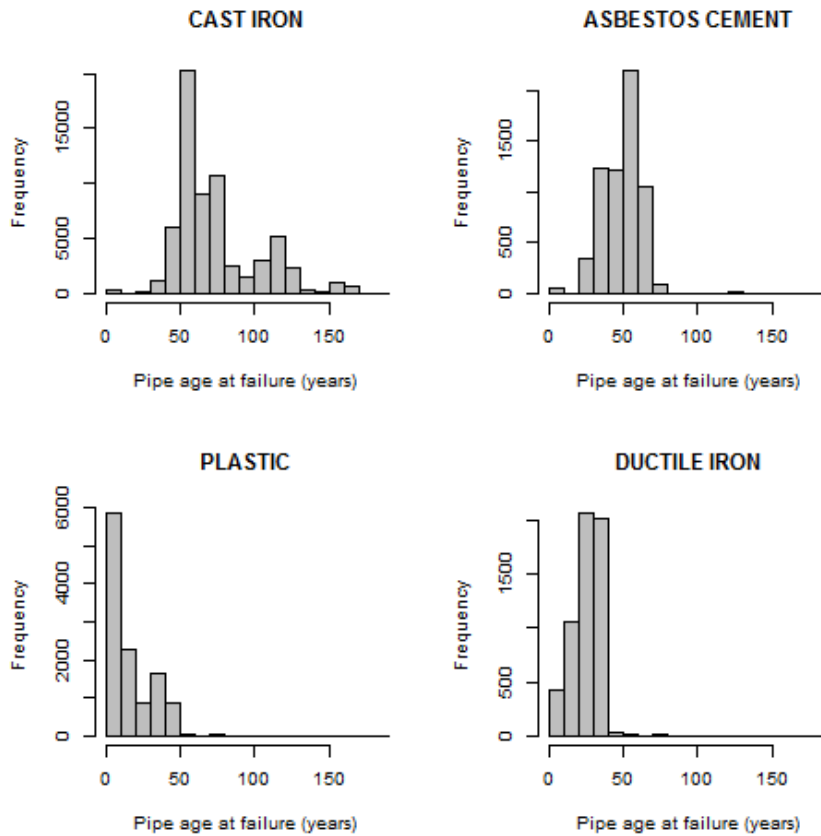
Figure 4 shows pipe failure rates (bursts per cell). All distributions show strong positively skewed distributions. For each material the greatest frequency of bursts of '0', that is the higher the number of bursts the less frequent they become, but for cast iron, for example, 8 bursts per cell was still found to occur in nearly 1000 cells.

Figure 4: Frequency of clean water pipe bursts (n bursts per cell) for the four pipe material types. Note the y axes have different scales.
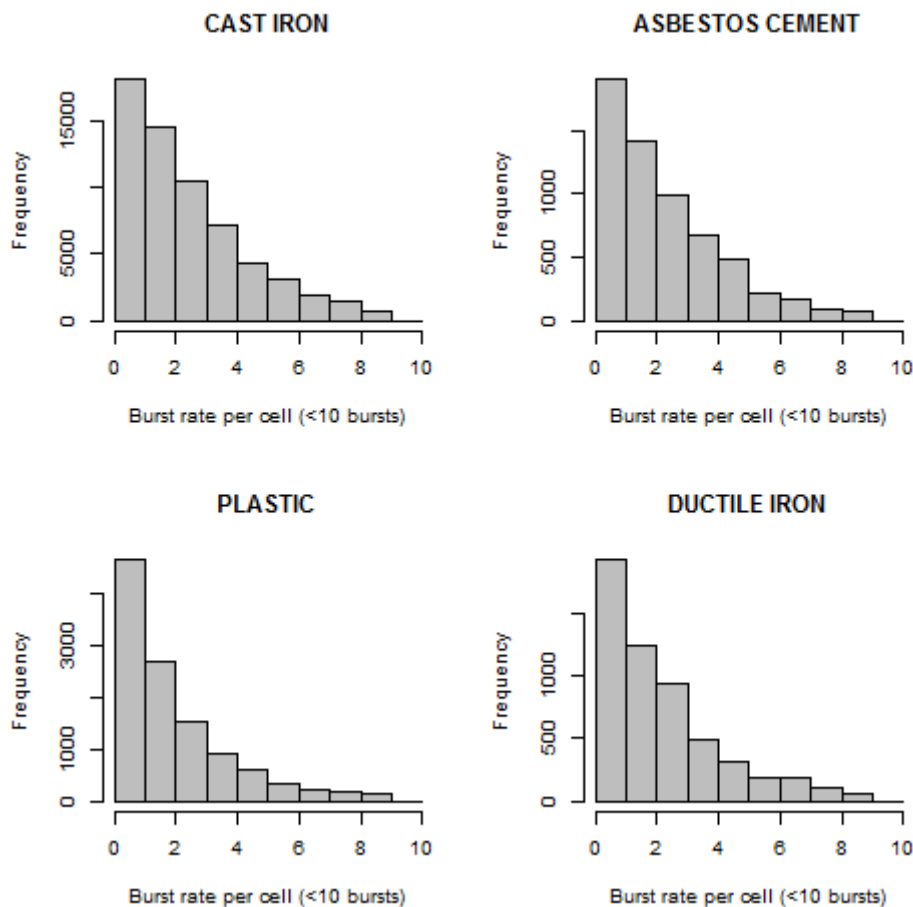
3.2   RANKING AND IDENTIFYING COVARIATES TO BE USED IN MODELS

The modelling process initially focused on clean water pipes (NERC NE/M008339/1) and those made only from cast iron and plastic. Cast iron was selected as it makes up the largest percentage of the YW pipe network whilst plastic was selected as it is now the most frequently used pipe for the clean water network. The first part of the modelling process involved an expert elicitation (EE) process with a group of YW employees responsible for maintenance and planning of the network. The aim was to identify the factors that they considered were most likely to cause pipe failure and was carried out in January 2015. For the second grant (NERC NE/NO13026/1) where additional data from the YW DMA relating to water source became available, the EE was repeated in Feb 2016. Those factors identified from the EE process would make up the covariates used in the initial EE models. The second

model involves the statistical selection of additional environmental and topographical factors. These would then be added to those covariates identified through the EE process.

3.2.1 Identifying explanatory variables through Expert Elicitation (EE) for predicting failures in the Cast Iron pipe network

Generally, YW believe that there are not many cast iron pipe failures within the actual pipe length. When these occur it is predominantly caused by corrosion and the creation of pin holes, which have a potential to blow out. Ground movement can cause circumferential fractures. Larger diameter cast iron pipes are made from very thick metal and the failures tend to occur at the joints or to the fittings (cast iron with lead joints). Small diameter pipes break more frequently; the metal is thinner and the pipes have more connections, pitting has more of an impact as the pipes tend to be of poorer quality. Expert elicitation with YW staff identified the factors considered most likely to produce failure in the cast iron pipework, and produced a ranking of these factors (Table 4). The identified environmental variables available, which could be regarded as proxies for some of the factors elicited from YW staff are also shown (Tables 4). There were no variables in the BGS dataset of geohazards, topographical or environmental indices that could describe soil moisture deficit or antecedent weather conditions. Water source is considered within the YW DMA data. The list of covariates used in the Expert Elicitation model and their rank are given in Table 5.

Table 4: Results of initial Expert Elicitation (EE) process and rank order of variables commonly associated with failure for cast iron pipes (1 = high correlation). Included are the co-variates included for each rank.

Rank	Variable	Notes
1	Corrosion	Particularly in 'damp' ground. Compound Topographic Index (CTI) used as the predictor as little variation in corrosion class across YW region.
2	Pipe Pressure	Pipe pressure could be considered within YW DMA data
3	Temperature in pipes	No BGS direct covariate available but could be considered within YW DMA data
4	Shrink-swell	Related to ground shrinkage, garden watering and increased weight of pipes. Use Shrink-swell classes as covariates
5	Soil-moisture deficit	No covariate available
6	Road vibration	Used A-road, B-Road and C-road length in each cell as covariates
7	Compressible deposits	Use compressible deposits classes as covariates

Table 5: Revised ranking list of variables to be used in Expert Elicitation (EE) models after YW DMA data became available.

Rank	Variable
1	DMA water source
2	Shrink Swell
3	CTI
4	A road
5	B road
6	C road
7	Compressible Ground

3.3 ADDITIONAL COVARIATE SELECTION

The second series of models examined whether improvements could be made to the EE models by including other geohazard, topographical and environmental factors. The additional covariates to be added to the EE model were identified after further statistical analysis. When choosing further covariates it is important to ensure that the model does not become over-parameterised, for example by including factors that might, to some extent, describe the same process. Thus, statistical relationships between selected covariates in the Yorkshire Water region were examined (Tables 6-8) to select further covariates for the model that were not correlated with covariates already selected through the EE and subsequent modelling.

Table 6 shows a correlation matrix between the six continuous covariates used in the EE model, along with four new continuous covariates (A-resistivity, B-Resistivity, 'Aspect North' and 'Aspect East'). The only strong correlation found was between A-resistivity and B-resistivity ($r=0.83$) which are the resistivity for the major and minor lithologies within a unit. The remaining covariates showed no strong correlation between each other, which suggested that were largely independent.

Table 7 provides information regarding how the explanatory variables might be related by reporting the correlations (r) from a principal component analysis. For example in Component 1, Av-slope and Av-Elevation, A-Resistivity and B-Resistivity show a reasonably strong negative correlation whilst the Compound topographic Index (CTI) has the opposite sign suggesting it is negatively correlated to these factors. In Component 2, the A and B resistivity are identified as being correlated to each other, demonstrating that a correlation existed between the resistivity of the major and minor lithologies within the parent material based. Component 3 suggests that the roads might all be important but component 4 identifies the B roads as being different from the A and C roads. Finally, component 5 identifies Aspect as being important. Additional data was obtained from the Office for National Statistics with respect to the number of people and the number of dwellings in each 100 x 100 m cell. A correlation of $r=0.93$ was found between these potential covariates, and we decided to use the number of dwellings within appropriate model formulations.

Table 6: The correlation (r) matrix for the seven continuous covariates assessed for use in the clean water cast iron pipe models for the YW region. 'Aspect East' was computed as cosine of aspect (compass direction of slope) and 'Aspect north' was computed as sine of aspect.

	CTI	Slope	Elevation	A-Road	B-Road	C-Road	A Res	B Res	Aspect North	Aspect East
CTI	1	0.557	0.478	0.005	0.007	0.021	-0.212	-0.207	-0.004	-0.001
Slope	0.557	1	0.524	0.029	0.012	0.047	0.244	0.243	0.001	-0.003
Elevation	0.478	0.524	1	0.053	0.026	0.098	0.315	0.300	0.001	-0.001
A-Road	0.005	0.029	0.053	1	0.004	0.052	-0.003	0.001	0.001	-0.001
B-Road	0.007	0.012	0.026	0.004	1	0.028	0.001	0.003	0.001	-0.001
C-Road	0.021	0.047	0.098	0.052	0.028	1	0.023	0.025	-0.001	0.001
A Resistivity	0.212	0.244	0.315	0.003	0.001	0.023	1	0.832	0.001	-0.001
B Resistivity	0.207	0.243	0.300	0.001	0.003	0.025	0.832	1	0.001	-0.001
Aspect North	0.004	0.001	0.001	0.001	0.001	0.001	0.001	0.001	1	-0.001
Aspect East	0.002	0.003	0.001	0.001	0.001	0.001	-0.001	-0.001	-0.001	1

Table 7: Correlations between selected covariates and their principal component scores

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
CTI	0.0005	-0.0001	-0.0062	-0.0030	-0.0020	0.0030	0.2528	-0.9674	-0.0034	-0.0006
Slope	-0.0012	0.00016	0.01380	0.00056	-0.0007	-0.0000	-0.9673	-0.2528	-0.0004	-0.0007
Elevation	-0.054	0.02254	0.99703	0.04582	0.0059	-0.0020	0.0150	-0.0027	-0.0000	0.0000
A-Road	0.00003	-0.00042	-0.00675	0.01728	0.9998	0.0017	-0.0003	-0.0021	0.0000	-0.0000
B-Road	-0.00003	-0.0001	-0.0024	0.00618	0.0016	-0.9999	0.0008	-0.0029	0.0000	-0.0000
C-Road	-0.0013	-0.0009	-0.04591	0.9987	-0.0175	0.0062	0.0006	-0.0025	-0.0000	0.0000
A Resistivity	-0.6864	0.72515	-0.05413	-0.0027	0.00001	0.0005	0.0002	-0.0000	0.0000	-0.0000
B Resistivity	-0.7250	-0.6882	-0.02431	-0.0027	-0.0003	0.00014	0.0004	-0.0000	-0.0000	0.0000
Aspect North	-0.0000	0.0000	0.000002	-0.000003	0.00003	-0.00007	-0.0001	0.0034	-0.9063	-0.4225
Aspect East	0.000025	-0.00006	-0.000005	0.000008	-0.00002	0.0000	0.0006	-0.0006	0.4225	-0.9063

In Table 8 the correlation values (r) between the 10 continuous covariates and twelve categorical covariates that make up the BGS geohazard datasets and properties of the soil such as soil type from DiGMap50Plus (PM_Class), dominant mineralogy, grain size and the likely fill properties of soil. In this context, the correlation is the square root of the coefficient of determination for a linear model in which the continuous covariate is the dependent variable with a different mean value for each level of the categorical independent variable. It is evident that the only significant correlations were between resistivity and the soil

properties. As soil properties (e.g particle size) are key factors in determining resistivity and the calculation of resistivity contains the likely variation of clay percentage, these positive correlations are expected. Improved correlations were obtained using A-Resistivity. From this statistical analysis, eleven covariates were chosen for including in the models along with those covariates selected through EE. The major choice was the selection of A-Resistivity instead of both A and B resistivity, as they were strongly correlated. In addition, as CTI was related to PM-Code, Soil Group and Fill code it could be seen as a factor which accounted for the soil textural properties.

Additional data was obtained from the Office for National Statistics with respect to the number of people and the number of dwellings in each 100 x 100m cell. A correlation of r = 0.90 was found between these potential covariates, and we decided to use the number of dwellings within appropriate model formulations

Table 8: Absolute correlation values (r) between twelve categorical covariates and ten continuous covariates (see Tables 2 & 3 for covariate descriptions).

	Collapsible Ground	Compressible Ground	Soluble Ground	Shrink Swell	Corrosive Sand	Running Sand	Landslide	Parent Material	Dominant Mineralogy	G - Grain	Soil Group	Engineered Materials
CTI	0.3000	0.4941	0.0898	0.3377	0.2463	0.4073	0.1415	0.6195	0.3332	0.2084	0.4655	0.5858
Slope	0.1162	0.3046	0.0593	0.3016	0.2323	0.3077	0.2034	0.5112	0.2249	0.1883	0.4272	0.4808
Elevation	0.0486	0.4794	0.0845	0.3676	0.2255	0.2738	0.1676	0.6822	0.5045	0.4652	0.5018	0.6005
A-Road	0.0016	0.0406	0.0166	0.0129	0.0260	0.0118	0.0267	0.0806	0.0581	0.0476	0.0444	0.0574
B-Road	0.0168	0.0323	0.0123	0.0129	0.0253	0.0136	0.0053	0.0592	0.0325	0.0344	0.0392	0.0452
C-Road	0.0652	0.0797	0.0541	0.0289	0.0558	0.0695	0.0538	0.2015	0.1447	0.1235	0.1356	0.1497
A Resistivity	0.1374	0.1793	0.1885	0.3237	0.1815	0.2806	0.0246	0.7862	0.5382	0.3674	0.6417	0.7363
B Resistivity	0.1448	0.1897	0.1649	0.2634	0.1814	0.2588	0.0380	0.7712	0.5301	0.3215	0.5309	0.6411
Aspect North	0.0022	0.0029	0.0021	0.0021	0.0032	0.0095	0.0013	0.0132	0.0050	0.0020	0.0064	0.0057
Aspect East	0.0006	0.0140	0.0138	0.0139	0.0146	0.0153	0.0140	0.0090	0.0032	0.0021	0.0040	0.0044

3.4   THE CLEAN WATER CAST IRON NETWORK

3.4.1   The Null model (Model 1)

A null model was created for the whole of the Yorkshire region where the density of bursts is a function of the log density (length) of cast iron pipes in each 100 x 100m cell (Figure 5). The red colours indicate an under prediction (positive residual) of the density of failures whilst the blue colours represent where over prediction (negative residual) occurs. The cumulative raw residuals on the X and Y axis indicate the total residual on the x or y axis. Thus the null model suggests that in particular, an under-prediction of pipe failure per unit area occurs in the central part of the YW region moving in a SW-NE direction and an small over-prediction in the SW of the region.

Figure 5: The null model (model 1) for the whole of the Yorkshire Water region where the density of bursts is a function of the log density of cast iron pipe in each 100 x 100m cell. Red indicates model under prediction and blue over-prediction.
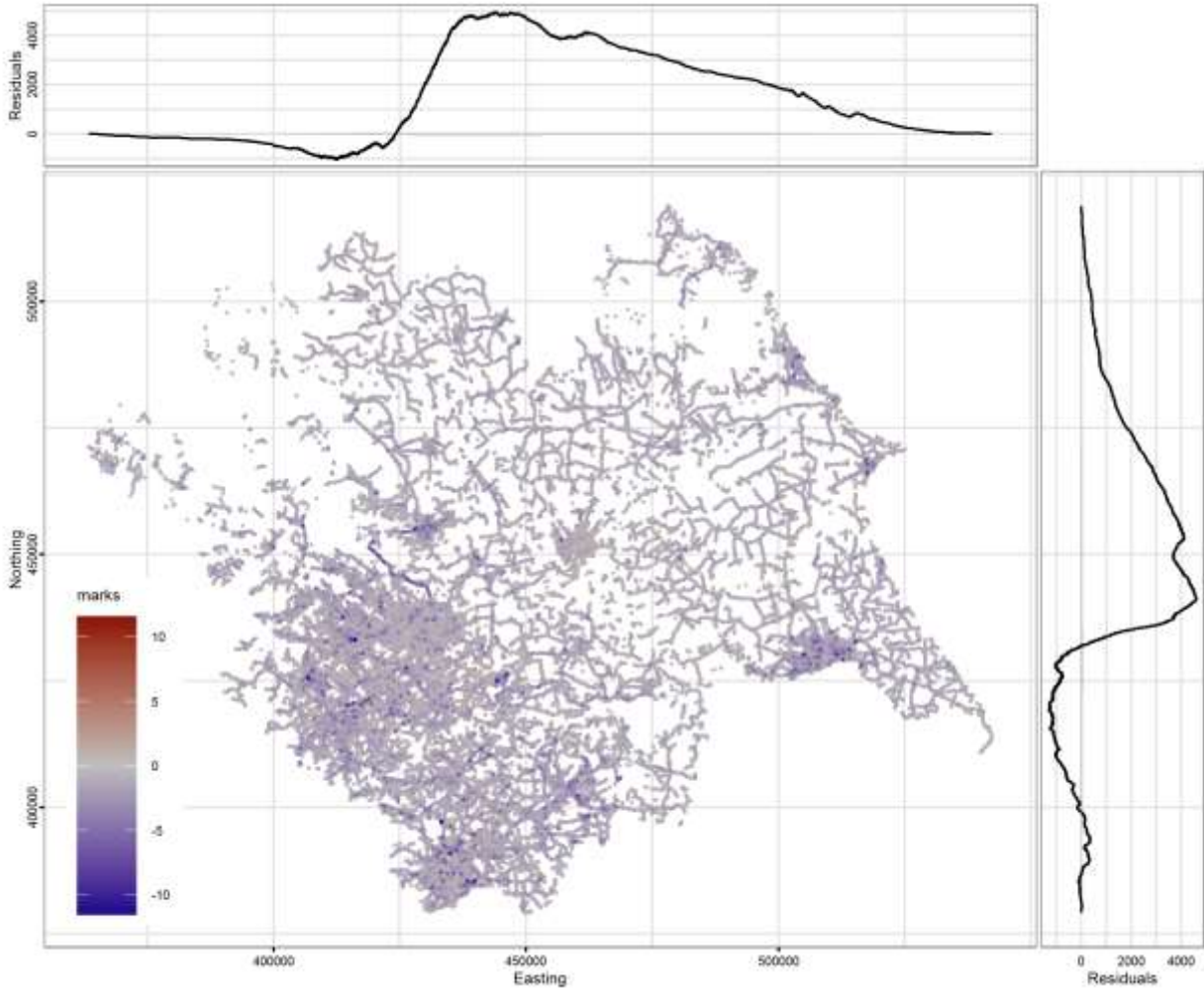
3.4.2 Fitting models by addition of single explanatory variables identified from the expert elicitation (Model 2)

A series of models were parametrised each with log density of the pipe type and material and just one additional explanatory variable taken from the elicited list. Each model could be compared to the null model by means of the log-likelihood ratio (Table 9). The covariates were ranked and the density of C road within the 100 x 100 m cells was found to be the most important variable. The LLr of the covariates suggested an order of importance of C-Road > water source > shrink swell > A Roads > Compressible deposits > CTI > slope > B roads. On this occasion B-roads were found not to be a significant factor. Table 9 shows the coefficients and the sign of the correlation for the continuous variables whilst Tables 10-12 show the coefficients of the categorical variables. In Table 11 it can be seen that there is no Class E for the shrink swell ground as none exists in the YW region, whilst in Table 12, the very low coefficient (-25.79) found for Class E (generally considered to be peat) is because although there is pipe in this category of Compressible ground, no pipe failures have been recorded.

Table 9: Output from spatial point process model fitting with a series of single covariates, added to a null model in which cast iron length is included as a covariate (Model 2)

Order added	Model	AIC	diffAIC	logLIK	LLr	pval	coef	Rank
6	C Road	1411950	-5847.93	-705971.9	5849.93	0.001	0.00367	1
1	Water source	1414721	-3077.01	-707353.4	3087.01	0.001	N/A	2
2	Shrink swell clay	1417458	-339.67	-708724.0	345.67	0.001	N/A	3
4	A Road	1417635	-162.97	-708814.4	164.97	0.001	-0.00134	4
7	Compressible	1417758	-40.03	-708872.9	48.03	0.001	N/A	5
3	CTI	1417771	-26.43	-708882.7	28.43	0.001	-0.01290	6
5	B Road	1417799	1.61	-708896.7	0.38	0.537	-0.00008	7

AIC = Akaike's information criterion; diffAIC = difference in AIC between the null and new model

Table 10 Coefficients for the water source categorical variables when added to the null model as a single variable (Model 2)

Class	Coefficient
Ground Water	-14.36613
Ground waters & Upland IRE	-14.19618
Impounding Reservoir	-14.07417
River Abstraction	-14.02787
Upland IRE & River Abstraction	-13.80554

Table 11 Coefficients of shrink swell clay categorical variables when added to the null model as a single variable (Model 2)

class	coefficient	Class Description
A	-14.28562	Ground conditions predominantly non plastic; No action
B	-14.15875	Ground conditions predominantly low plastic; No action
C	-14.34979	Medium plasticity; action required
D	-14.65175	High Plasticity

Table 12 Coefficients of the compressible ground categorical variables when added to the null model as a single variable (Model 2)

class	coefficient	Class Description
A	-14.25815	No indicators of compressible ground – No action
B	-14.28803	Very slight potential of compressible deposits
C	-14.16804	Slight possibility of compressibility problems
D	-14.30661	Significant potential for compressibility problems
E	-25.79313	Very significant potential of compressibility problems

3.4.3 Fitting models by sequential addition of explanatory variables identified from the expert elicitation (Model 3)

Each statistically significant predictor added as a single predictor to the null model was then fitted in turn, in the order that they were ranked in the elicited list to give a final EE sequential model with seven covariates (Table 13).

Table 13: P-value from tests for sequential addition of statistically significant covariates identified from the expert elicitation added to the null model. LLr is the log likelihood ratio statistic expressing how many times more likely the data are based on addition of this covariate in comparison to the previous model.

	Model	pval	LLr
1	Water Source	0.001	1543.50
2	Shrink Swell Clay	0.001	86.35
3	CTI	0.208	0.78
4	A Road	0.001	83.57
5	B Road	0.014	3.00
6	C Road	0.001	2612.37
7	Compressible	0.001	43.43

Figure 6 shows the output of this model. By comparing the cumulative sum of raw residuals with the null model (Figure 5) it can be seen that including all the covariates determined from the expert elicitation produces a model that improves the description of the data. The positive sum of residuals as shown in both the x and y axis appears to suggest that the model continues to under predict in the central area of the YW region.

Figure 6: Final lurking variable plot for the best fit model (Model 3) based on the expert elicitation process where covariates are added in sequential order. The red areas indicate where the model under predicts the number of expected pipe bursts per cell, whilst the blue over-predicts per 100 x 100m cell.
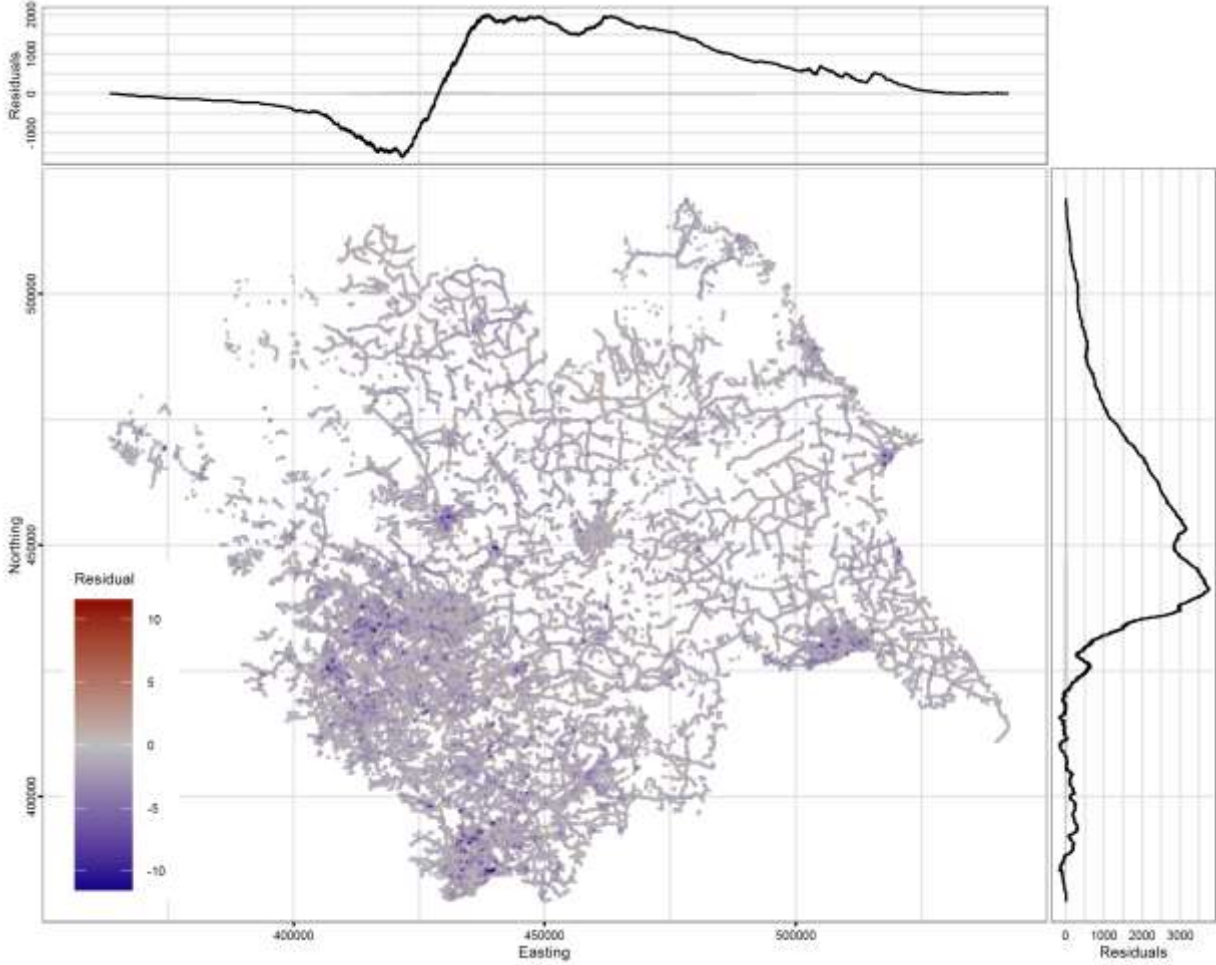
3.4.4   Fitting models by addition of single explanatory variables identified from the expert elicitation and other topographic and environmental indices (Model 4)

The number of variables used in the model was increased following the selection criteria outlined in Section 3.3. These were then added to the variables selected through the Expert Elicitation procedure. The 14 variables selected are shown in Table 14, which also reports the results of this analysis. The major explanatory variables, those with the greatest diffAIC and LLr

values, are C roads > number of Dwellings > Water Source > Sulphide/Sulphate > Solubility > shrink swell clay > corrosivity > A roads > Compressible ground > A- resistivity > CTI. Three of the covariates were found not to be significant at P <0.05, these being Aspect North, Aspect East, B-roads and solubility. Each continuous variable also has a sign (+/-) attached to it and these represent whether there is a negative or positive correlation to the density of bursts expected in a 100 x 100 m cell. The coefficients for the categorical variables, not previously reported (Table 10-12) are shown in Tables 15 to 17.

Table 14: Full region output from spatial point process model fitting with a series of single covariates, added to a null model (Model 4).

	Model	AIC	diffAIC	logLIK	LLr	pval	coef	rank
6	C Road	1411950	-5847.93	-705971.9	5849.93	0.001	0.00367	1
13	Dwellings	1413710	-4087.88	-706851.9	4089.88	0.001	0.01700	2
1	Water source	1414721	-3077.015	-707353.4	3087.01	0.001	N/A	3
14	Sulphide/Sulphate	1416125	-1672.95	-708058.4	1676.95	0.001	N/A	4
11	Solubility	1416797	-1000.47	-708392.6	1008.47	0.001	N/A	5
2	Shrink swell clay	1417458	-339.67	-708724.0	345.67	0.001	N/A	6
12	Corrosivity	1417583	-214.87	-708787.4	218.87	0.001	N/A	7
4	A Road	1417635	-162.97	-708814.4	164.97	0.001	-0.00134	8
7	Compressible	1417758	-40.03	-708872.9	48.03	0.001	N/A	9
8	A Resistivity	1417759	-39.21	-708876.3	41.21	0.001	-0.00005	10
3	CTI	1417771	-26.43	-708882.7	28.43	0.001	-0.01290	11
9	Aspect East	1417796	-1.52	-708895.1	3.52	0.060	-0.01047	12
10	Aspect North	1417799	1.29	-708896.5	0.70	0.400	-0.00465	13
5	B-Road	1417799	1.61	-708896.7	0.38	0.537	-0.00008	14

Table 15 Coefficients of the soluble ground categorical variables when added to the null model as a single variable (Model 4)

class	coefficient	Ground Classification
A	-14.11959	Soluble rocks not thought to be present
B	-14.65934	Soluble rocks are present but unlikely to cause problems
C	-14.35564	Significant Soluble rocks are present with low possibility of localised subsidence or dissolution related degradation of bedrock
D	-14.17672	Very significant soluble rocks are present with a moderate possibility of localised natural subsidence or dissolution related degradation of bedrock
E	-14.19376	Very significant soluble rocks are present with a high possibility of localised subsidence or dissolution of bedrock

33

Table 16 Coefficients of the soil corrosivity categorical variables when added to the null model as a single variable (Model 4)

class	Coefficient	Ground Classification
class 1	-14.22632	Unlikely to cause corrosion
class 2	-14.48361	May cause corrosion
class 3	-14.29592	Likely to cause corrosion

Table 17 Coefficients of the sulphate / sulphide categorical variables when added to the null model as a single variable (Model 4)

class	Coefficient	Ground Classification
HIGH	-14.43124	Presence of Sulphate
LOW	-13.93590	Presence of Sulphide
NONE	-14.27611	Background concentrations

3.4.5 Fitting models by sequential addition of explanatory variables identified from the expert elicitation and other topographic and environmental indices (Model 5)

Following on from the fitting of Model 4 where the null model was fitted with individual covariates, a full sequential model was fitted where each previously identified significant ($P<0.05$) covariate was added to the null model. The differences between LLr values in Table 18 indicate the importance of the covariate. All models showed a significant ($P<0.05$) improvement from the previous model by adding additional covariates. Coefficients for the continuous covariates can be found in Table 18. In addition, coefficients for the categorical covariates for model 5 can be found in (Tables 19 - 24). These are slightly different numerically to the coefficients for the categorical variables obtained when individual categorical variables were added to the null model as they all share a common intercept value of the null model. The full model based on the sequential model is presented in Figure 7. When examining the sum of raw residuals in the lurking variable plot, it can be seen that (i) an area still exists in the middle of the YW region where the model under predicts which is still present and (ii) a slight model over-prediction occurs in the SW region which is heavily urbanised. The sum of the raw residuals is again lower than the EE sequential model (Figure 6), demonstrating that the inclusion of other environmental factors improves the model parameterisation.

Table 18: Full region P-values based on the log likelihood ratios tested using the Chi-squared distribution (testing model 5 with added covariate against the previous model in the sequence in which covariates are retained where P<0.001). Aspect was not included because it was not a statistically significant predictor across the full region.

Order Added	Model	pval	LLr	coef
1	Water Source	0.001	1543.50799	N/A
2	Shrink swell Clays	0.001	86.35445	N/A
3	A Road	0.001	82.16877	0.3825773
4	C Road	0.001	2443.65776	0.2662982
5	Compressible Ground	0.001	43.55864	N/A
6	A Resistivity	0.001	12.63903	-0.534795
7	Soluble ground	0.001	213.36582	N/A
8	Soil Corrosivity	0.001	20.02504	N/A
9	Dwellings	0.001	417.94804	-4095012
10	Sulphide / sulphate	0.001	531.43914	N/A

Table 19 Coefficients for water source from the cast iron clean water network obtained using Model 5

class	coefficient
	-13.97587
Ground water	-13.67287
Ground water & upland IRE	-13.58593
Impounding reservoir	-13.53901
River abstraction	-13.39990
Upland IRE and River abstraction	-13.27143

Table 20 Coefficients for shrink swell clays from the cast iron clean water network obtained using Model 5

class	coefficient	Class Description
A	-13.97587	Ground conditions predominantly non plastic; No action
B	-13.91232	Ground conditions predominantly low plastic; No action
C	-13.92539	Medium plasticity; action required
D	-14.20621	High Plasticity

Table 21 Coefficients for compressible ground from the cast iron clean water network using Model 5

class	coefficient	Class Description
A	-13.97587	No indicators of compressible ground – No action
B	-13.77658	Very slight potential of compressible deposits
C	-13.52180	Slight possibility of compressibility problems
D	-13.85492	Significant potential for compressibility problems
E	-24.93983	Very significant potential of compressibility problems

35

Table 22 Coefficients for soil corrosivity from the cast iron clean water network using Model 5

class	coefficient	Ground Classification
class 1	-13.97587	Unlikely to cause .
class 2	-14.03686	May cause corrosion
class 3	-13.88253	Likely to cause corrosion

Table 23 Coefficients for soluble ground conditions for the cast iron clean network using Model 5

class	Coefficient	Ground Classification
A	-13.97587	Soluble rocks not thought to be present
B	-14.27221	Soluble rocks are present but unlikely to cause problems
C	-14.16169	Significant Soluble rocks are present with low possibility of localised subsidence or dissolution related degradation of bedrock
D	-13.91201	Very significant soluble rocks are present with a moderate possibility of localised natural subsidence or dissolution related degradation of bedrock
E	-14.12601	Very significant soluble rocks are present with a high possibility of localised subsidence or dissolution of bedrock

Table 24 Coefficients for sulphide/ sulphate in soils from the cast iron clean water network using Model 5

class	coefficient	Ground Classification
HIGH	-13.97587	Presence of Sulphate
LOW	-13.52185	Presence of Sulphide
NONE	-13.81354	Background concentrations

Figure 7: Result of full model (Model 5) or the YW region using sequential addition of covariates. Examination of the combined X and Y axis residuals suggest that overall the model is under predicting the number of pipe failures per unit length of pipe, with the red colours indicating where this is happening to the greatest extent and the blue the least.
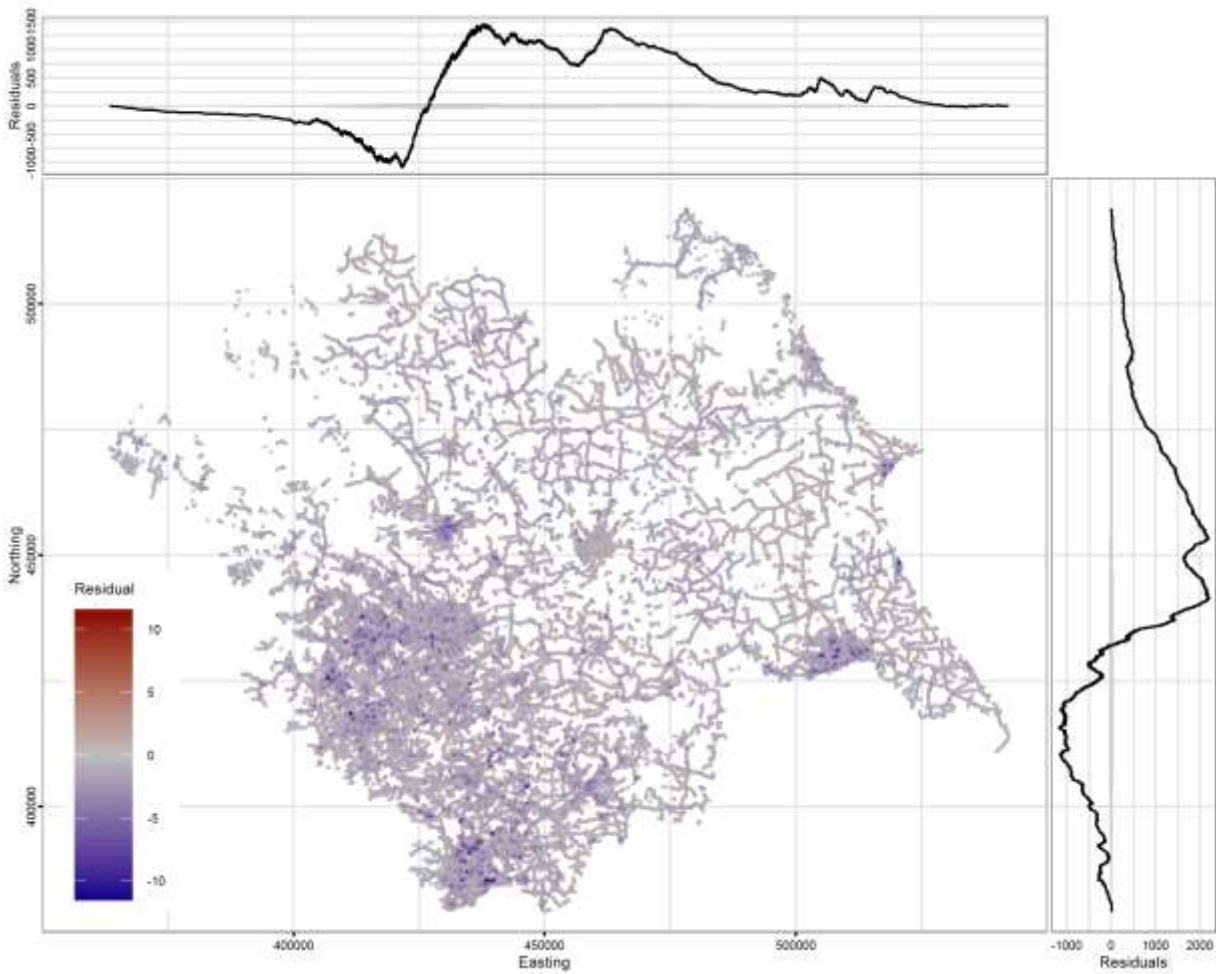
3.4.6   Discussion

3.4.7   Model Performance

For the Cast Iron clean water pipe network, several models have been produced. The first model was the Null model that predicts the number of expected bursts associated with the density (length) of pipe in each 100 x 100m cell. This progressed to a sequential model based on covariates obtained from an expert elicitation process (Model 3) and a final sequential model where other additional environmental and geological factors were included (Model 5). Both Model 3 and Model 5 delivered large decreases in total raw residual compared to the Null model (Model 1) as demonstrated by the total sum of residuals in the lurking variable plots across the YW area (Figures 5, 6 and 7). The modelling process was initially based on the Expert Elicitation exercise undertaken with the YW employees. Results demonstrated that the covariates

that YW identified were all highly significant with the exception of B roads. In particular, issues relating to C roads (e.g traffic volume and vibration, other utilities digging up the road) were the strongest predictors. Shrink swell clays and compressible deposits were found to be significant geological based predictors, but their LLr values were much lower. In the subsequent models, the addition of other environmental predictors such as water source, the number of dwellings per cell and the sulphide/sulphate layer were significant (P<0.05) factors in decreasing the model residual in the final EE+ model (Model 5). Water source is important as chemicals used to reduce the turbidity of water through flocculation (e.g. aluminium sulphate) can increase internal pipe corrosion, thereby possibly enhancing the external effects that may contribute to pipe failure.

By examining the Lurking Variable Plot in the final sequential model (Model 5) it can be seen that there is an area in the centre of the YW region, moving in a roughly SW-NE direction where the model under-predicts pipe failure. This area was obvious in each of the models presented including the Null model and represents an area where the model has failed to account for a process or environmental factor which impacts on pipe failure. Maps of geology and geohazards were examined for possible explanation. The first explanation is part of this area lies on the Lower Coal Measures. It was considered that the inclusion of the Sulphide / Sulphate layer may account for this as there may have been increased sulphide minerals in the soil which when oxidised would create H_2SO_4. Whilst this data proved to be one of the models major covariates, some model under-prediction remained in this area, suggesting that this pipe failure may be related to issues of ground re-settlement after the removal of coal (Marino, 2000). The second area of model under prediction is an area of lacustrine clays deposits from the Glacial Lake Humber. Lacustrine clays are typically poor at bearing weight and this may be an influence. Thus, for both these areas greater than expected pipe failures may occur because of geological type or related properties.

The second output from the model analysis that can be used for improving our understanding of the cast iron pipe network, are the model coefficients obtained from models 2 and 4. Table 25 below, examines the possible reasons for the sign of the correlation for continuous covariates and how we may interpret the meaning of the coefficients of the different classes of the categorical covariates. Whereas the continuous covariate coefficients are quite explanatory, describing a positive or negative correlation between the covariate and the expected number of pipe bursts in each cell, greater knowledge of geology, geohazard data and environmental factors is required to

understand the categorical variables. For the categorical variables we are comparing the numerical value of the coefficients against the different classes of the covariate, with the coefficients with the greatest numerical values being of greater influence than lower values.

Initially it was considered, that the model output would offer a relatively simplistic interpretation of the categorical variables along the lines of 'an increasing number of pipe failures would occur as the class of each geohazard increased in severity' (i.e. a linear response). However, this was not the case and an understanding of how the dataset for each geohazard was derived (primarily for the insurance industry to assess risk to buildings) was required. For example, the low class of the Sulphide / sulphate dataset actually represents the sulphide containing soils, whilst the 'high' class represents the sulphate bearing rocks which when they collapse cause much greater damage to buildings, through subsidence. As the geohazard datasets were produced for their effects on buildings, it was necessary to understand how the pipe network interacts with the soil in what has been described as soil-structure-pipe interactions for settlement and deflection (Olliff et al. 2001). This reflects how different soil types interact with the pipe type and the load that it may be subjected so that the right balance between flexibility and rigidity is achieved. As some of these geohazards are connected with clay (shrink-swell, compressible ground), soils will then be expected to behave differently according to clay content and type which is why we suggest the categorical coefficients do not behave in a linear way in the descriptions below. Thus, if no obvious trend in coefficients is seen with the classes of the categorical co-variable than it is likely that the categorical coefficients are reflecting the ability of the soil in the categorical class to provide improved settlement and deflection, for the pipeline.

Table 25: Interpretation of the outputs from adding individual covariates to the null model for the YW region (Model 2 & 4)

Rank	Covariate	+/- coefficient	Notes
1	C Road	+	Positive correlation between the density of C roads in a cell and pipe failure per unit length. Pipe failure could be a result of lower quality road construction designed for lower frequency and load of vehicles causing greater vibration. Potential for poor drainage in the sub-grade of the road. This may also be due to construction activity and third party damage.
2	Dwellings	+	Indicates that increased pipe failure occurs as the number of dwellings in a cell increases, suggesting increased pipe failure could be associated with pressure changes within the system and use on the system.
3	Water Source	N/A	There is evidence that water source can play a key role in pipe failure and this could be through source in the pipe. In the YW region, failure in pipes where water is supplied from upland river abstraction or impounding reservoirs is the greatest suggesting that some internal corrosion may be taking place as a result of water treatment.

39

4	Sulphate / Sulphide	N/A	The highest coefficient is found in the Low class and the lowest in the High class. Consulting the BGS geohazard map the Low class is dominantly on the coal measures and Oxford clay formations thus representing the possible presence of sulphide. The High Class is associated with gypsum bearing rocks where pipes would be buried in soils which are likely to have lost SO_4 from gypsum via leaching and are unlikely to cause increased failure, unless substantial subsidence occurs. This result suggests that the presence of increased sulphide is having an effect on the pipe network.
5	Soluble Ground	N/A	The results show that the lowest coefficient is found for Class B suggesting that there is less pipework failure on the chalk and limestone soluble rocks types. The indication is that the soils may be shallow and the pipe may rest on rock thus maintaining greater support. Class C, D and E are based on soluble rocks which are likely to have gypsum deposits (Permian mudstones). In these rocks the solubility is a lot deeper, so the pipes would exist in normal soils and this is reflected by the coefficients being similar to Class A (no soluble rocks considered present).
6	Shrink Swell	N/A	Results suggest that the coefficient for Class D shrink swell was the smallest, whilst the values of the coefficient for Class A-C were similar and were slightly larger. It is possible that water leaks in Class D may expand the clays creating a self-sealing effect. However, the top Class of shrink swell is not present in the YW region so that the potential effects of shrink swell have not been fully tested.
7	Corrosivity	N/A	The lowest coefficient was found in the soil corrosivity Class 2 (May cause corrosion). By examining coefficient maps, it was found that Class 2 consisted largely of slowly permeable chalky till soils, some well drained calcareous soils associated with the Chalk Downs in the YW region. The presence of carbonate and high pH is known to prevent corrosion. A small area of corrosivity Class 2 soils consisted of a lacustrine clays and is perhaps wrongly classified and should be in Class 3, as they are predominantly clay and have poor drainage. This however demonstrates the complexity of the CIPRA classification in terms of weighting and how the final score is calculated. Overall the results are suggesting that pipes in a high pH, high carbonate environment appear more resistant to pipe failure. This may also tie in with the soluble ground results.
8	A road	-	Negative correlation between pipe failure per unit length and the density of A roads in a cell. This may be related to improved road construction associated with high vehicle numbers and heavier vehicles, better sub grade drainage, with particular reference to water table and pipe installation, or pipes being sited next to the road.
9	Compressible Ground hazard	N/A	There weren't large differences between the Class A-D in the size of the coefficient. However, Class E had a much smaller coefficient, and this was because although pipe is sited within areas of Class E, no failures were recorded. As Class E generally represents peat like deposits, leakages may be hard to spot.
10	A resistivity	-	Resistivity is the most heavily weighted factor in the corrosion dataset and so to a degree resistivity may already have been included. A negative correlation between A-resistivity and pipe failure was found suggesting that greater pipe corrosion occurred at low resistivity which is expected. Could also indicate clay and moisture factors.
11	Compound Topographic Index	-	A negative correlation between CTI and pipe failure indicating greater frequency of pipe failure when a soil is potentially dryer. This may suggest that soils that dry out maybe slightly more prone to differential ground movement.

3.4.8   Using coefficients from the sequential model to produce heat maps

Heat maps were produced according to section 2.4.2. The Overall coefficient intensity maps, based on equation 4, were produced by combining the coefficients for each 100 x 100m cell (Section 2.4). This provides an indication, based on the coefficients derived from the final sequential model, of the intensity (hostility) of the overall environmental against the pipe

network in each cell. This is provided for the whole YW region area is shown in Figure 8 and a smaller section is shown in Figure 9.

Figure 8 Total Intensity map of YW region for the cast iron clean water network showing areas which are most hostile to pipe networks produced using significant variables obtained using model 5.
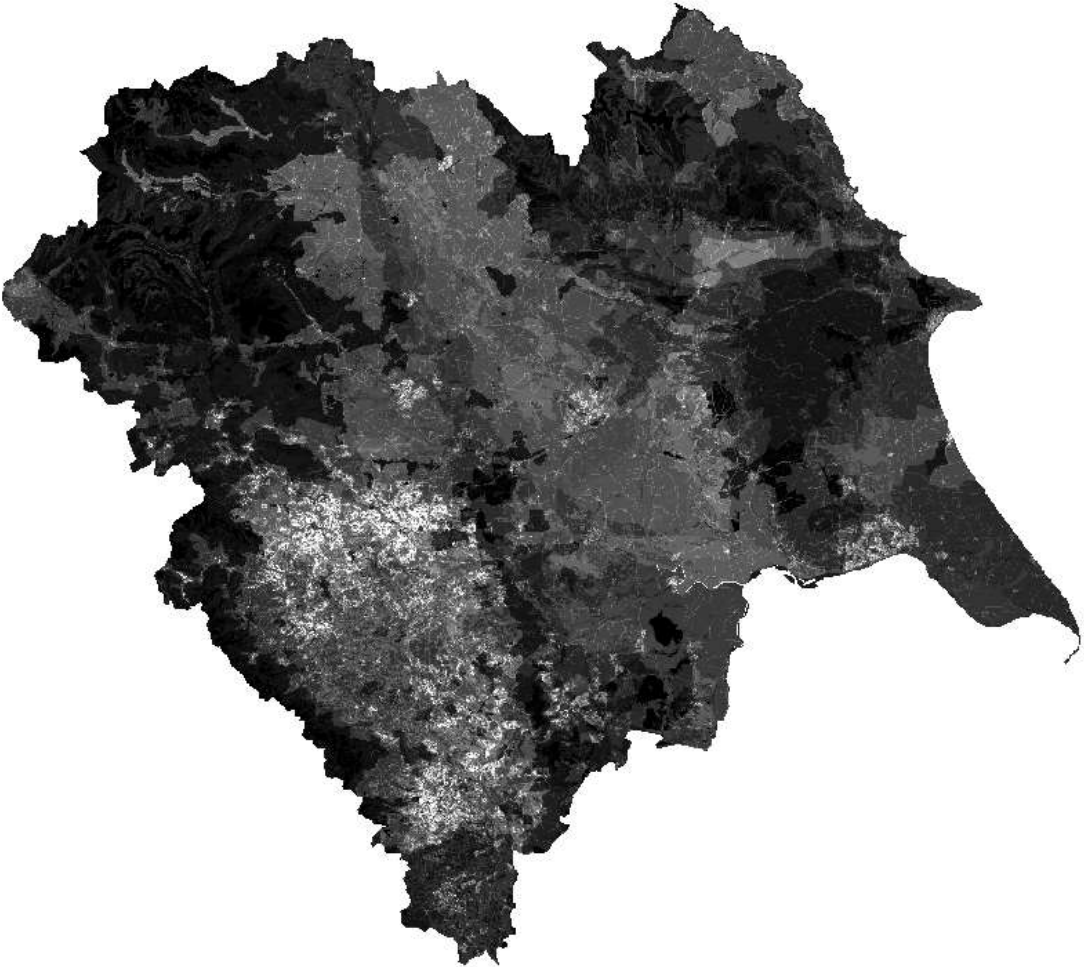
Figure 9 Close up of section of Total Intensity map (Figure 9) for the YW cast iron clean water network based on significant outputs from Model 5
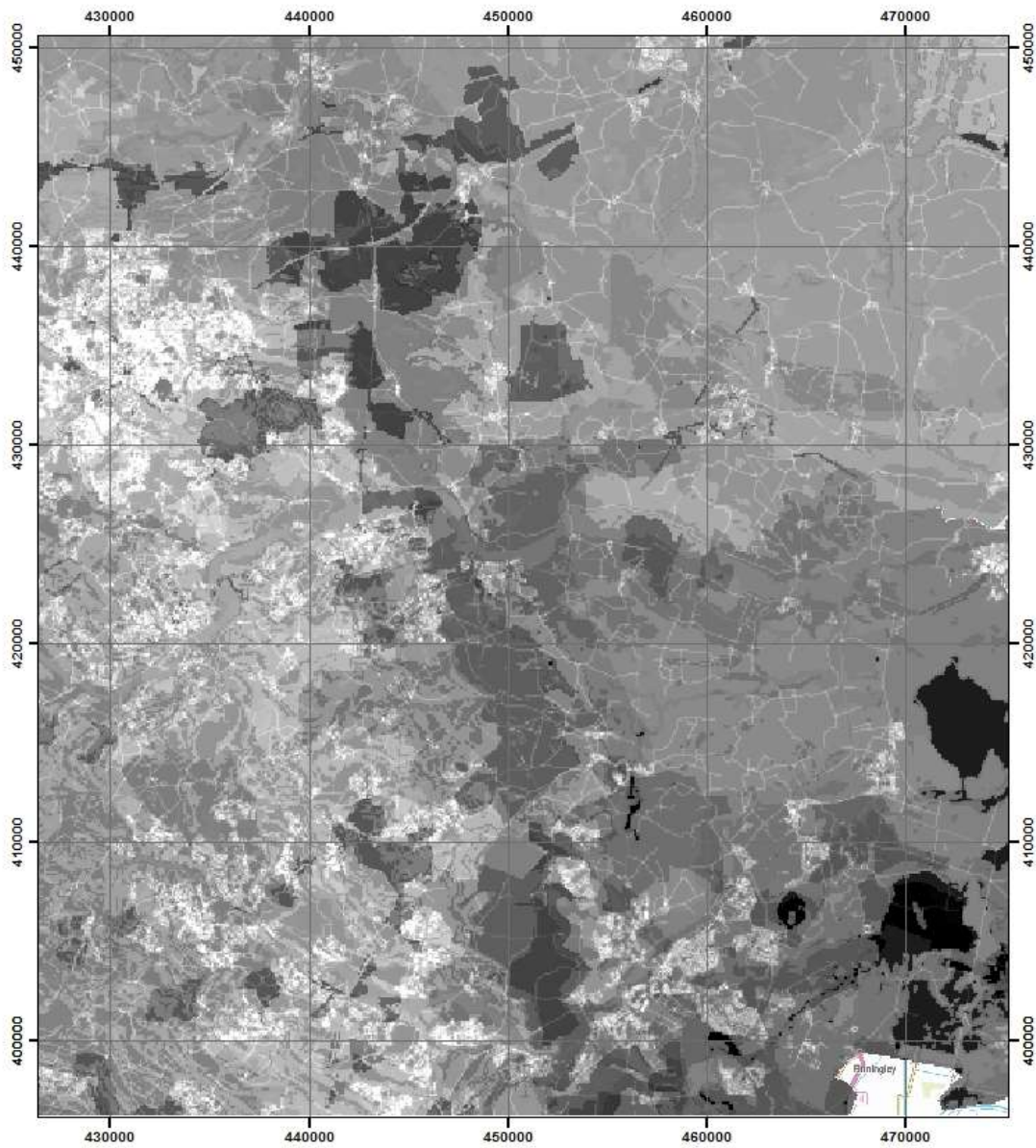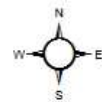
Examination of the individual factors (see Section 2.4) that contribute to the overall Heat map (Figures 8 & 9) can be examined by producing heat maps of each covariate where all the coefficients are placed on the same numerical scale (i.e. from the lowest coefficient value to the highest across all the categorical variables) and colour scale. Thus the maps show the spatial intensity of each covariate across the YW area, and with the benefit of allowing us to directly compare the effect of each coefficient.

Figure 10 shows the individual heat maps for the significant covariates from Model 5. For cells in which A road is present these are mostly in the yellow and red colour range, indicating that potentially traffic on these roads may damage nearby pipework. When used as a single covariate in Models 2 and 4, A-roads were found to have a negative coefficients (see Table 25). However, within the final sequential model (Model 5) A roads have a positive coefficient indicating that the inclusion of other variables had an effect on the model residual, and that A-road traffic had an effect on pipe network failures. For C roads the highest values (red colours) are found in urban areas and crossroads, demonstrating the effect high densities of C-roads can have on the pipe network. Outside these urban areas much of the rest of the network is pale yellow, indicating the less dense C road network. The C roads in urban areas in particular appear to have the most impact on pipe failure, possibly suggesting that increased traffic or other urban activities (digging up roads by other utilities) may be the cause. Mapping the water source identified large areas where the values were high, with large areas being red or orange. This suggests that for much of the YW region, there is a possible contribution to failure caused by water source (i.e. treatment of water). The positive coefficient for the number of dwellings suggests that higher pipe failure occurs with increased number of dwellings per 100 x 100 m, as this is likely to contribute to pressure changes through the network, thus causing corroded pipes to fail (e.g. creating pin holes). Whilst much of the area had low values, and impact, being green, the highest values (red) were found to identify certain urban areas including parts of Leeds, Bradford, Hull and Halifax as well as smaller towns such as Harrogate and Rippon. The other major covariate that has a large impact on pipe failure is sulphate and sulphide. In particular the areas in red represent the Coal Measures and Oxford clay formations which are likely to have sulphide present, which may oxidise and produce H_2SO_4. Compared to the human influenced factors (water source, roads, dwellings), the remaining geohazards (compressible ground, soluble ground, soil corrosivity, shrink swell clays) generally had considerably less impact (low values) and variability across the area, reflecting the similarity of the categorical coefficients for each geohazard.

Figure 10 Heat maps produced for the significant variables for the cast iron clean water network obtained using results from model 5
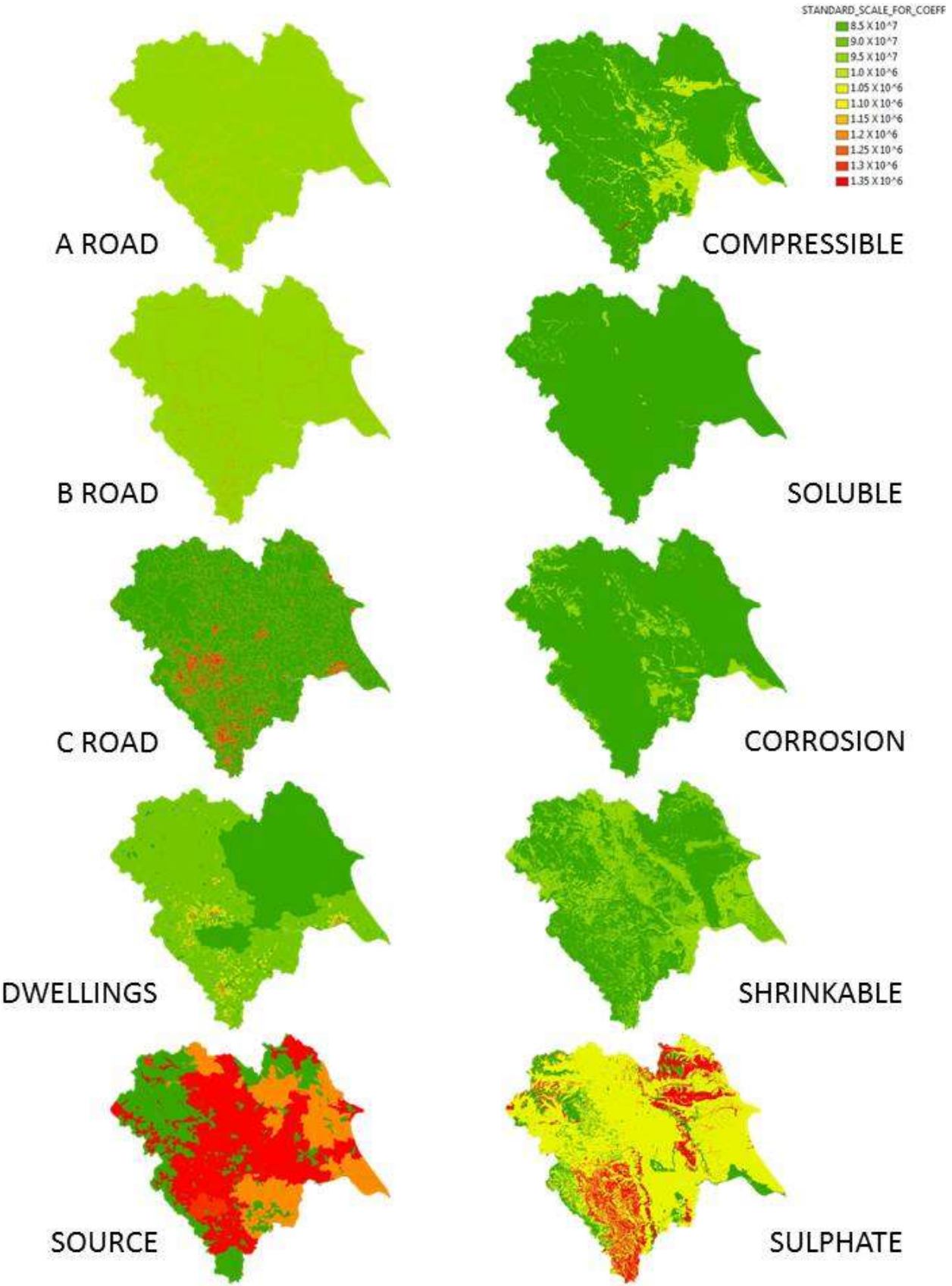
3.5 THE CLEAN WATER PLASTIC PIPE NETWORK

3.5.1 Introduction

Plastic pipe make up ~8 % of YW clean water pipes. The expert elicitation did not provide a definitive ranking list for pipe failure mechanisms. However, contaminated ground was identified as a major problem, but mechanisms are not understood. Failure mechanisms included (i) poor construction methods resulting in joint failure, (ii) PVC may become brittle, but insufficient quantities of this pipe type have been installed to identify such issues and (iii) poor bedding of the pipes. The joints are created by electro-fusion or are mechanical couples or butt-fused joints (jointed above ground). When MDPE pipes were first introduced, the electro-fusion fittings had a high failure rate failing when the pipes were uncurled. A similar modelling procedure was undertaken as for the clean water cast iron pipe network, whereby a null model is produced, followed by the addition of factors from an EE exercise and then the inclusion of other environmental factors.

3.5.2 The Null Model (Model 1)

A null model was created for the whole of the Yorkshire region where the density of bursts is a function of the log density of plastic pipes in each 100 x 100 m cell (Figure 11). The red colours indicate an under prediction (positive residual) of the density of failures whilst the blue colours represent where over prediction (negative residual) occurs. The cumulative raw residuals on the X and Y axis indicate that an under-prediction of pipe failure per unit area occurs in the central part of the YW region as both the total residual on the X and Y axis are both positive.

Figure 8 The Null model for the clean water plastic pipe network across the YW region.
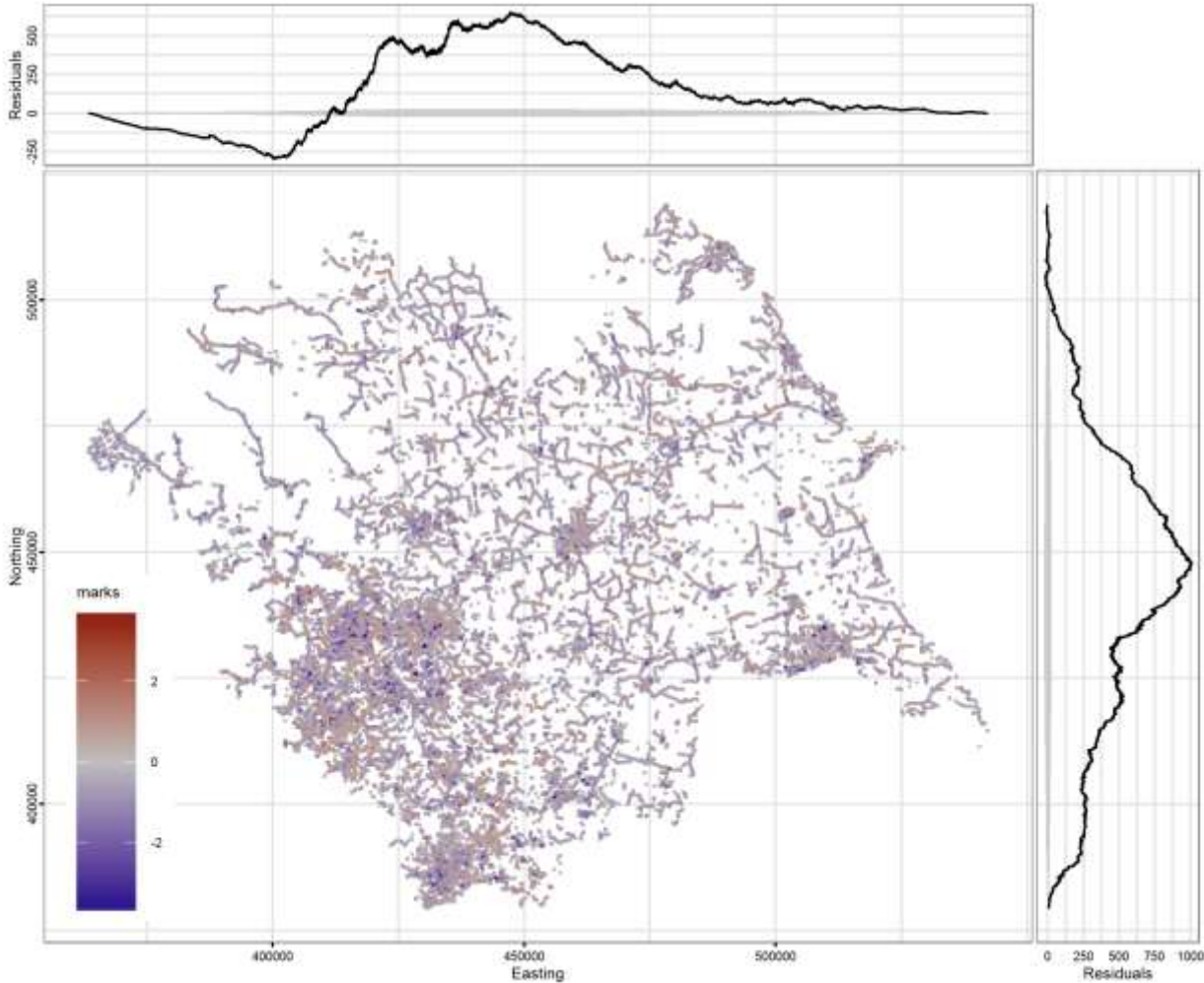
3.5.3 Fitting the single variables – expert elicitation (Model 2)

The 8 variables from the expert elicitation (identified by the Order Added column in Table 26) were then added to the null model (Table 26). Only the CTI was found not to be significant at $P<0.05$ and this was omitted from further modelling. The most important parameters were found to be C-Roads > Compressible ground > Shrink swell clay > slope > B roads > A roads > elevation. Coefficients for the categorical variables are shown in Tables 27 - 28.

Table 26 Metrics of models consisting of individual predictor variables added to the null model independently of each other (Model 2)

Order Added	Model	AIC	diffAIC	logLIK	LLr	pval	coef	Rank
7	C Road	290918.8	-619.60	-145456.4	621.60	0.001	0.0027	1
8	Compressible	291455.4	-82.94	-145721.7	90.94	0.001	N/A	2
4	Shrink swell clay	291490.1	-48.24	-145740.1	54.24	0.001	N/A	3
2	Slope	291512.9	-25.52	-145753.4	27.52	0.001	0.01584	4
6	B Road	291514.7	-23.67	-145754.3	25.67	0.001	0.00164	5
5	A Road	291525.7	-12.71	-145759.8	14.71	0.001	0.00098	6
3	Elevation	291535.2	-3.18	-145764.6	5.18	0.022	-0.00026	7
1	CTI	291540.1	1.72	-145767.1	0.27	0.602	0.00290	8

46

Table 27 Coefficients of shrink swell clay categorical variables for the plastic clean water network obtained using Model 2

class	coefficient	Ground Classification
A	-13.65130	Ground conditions predominantly non plastic; No action
B	-13.61929	Ground conditions predominantly low plastic; No action
C	-13.77499	Medium plasticity; action required
D	-13.71920	High Plasticity

Table 28 Coefficients of Compressible Ground categorical variables for the plastic clean water network obtained using Model 2

class	coefficient	Ground Classification
A	-13.63858	No indicators of compressible ground – No action
B	-13.70570	Very slight potential of compressible deposits
C	-12.78644	Slight possibility of compressibility problems
D	-13.77871	Significant potential for compressibility problems
E	-15.12142	Very significant potential of compressibility problems

3.5.4 Fitting a sequential model using the covariates from the Expert Elicitation (Model 3)

The next step involved fitting the EE covariates in the form of a sequential model. The covariates were added in the order of the EE exercise. Results are reported in Table 29. From the LLr values it can be seen that the order of importance changes slightly from the Model 2 so that C-roads > shrink swell clay > Slope > Compressible > elevation > B-roads > A-roads, suggesting that some of the residual is being accounted for by different factors. The model output is presented in Figure 12 and it can be seen that the model residual is greatly reduced by the inclusion of the factors from the Expert Elicitation exercise.

Table 29 Metrics of sequential addition of expert elicited predictor variables to the null model (Model 3)

Order added	Model	pval	LLr
1	CTI	0.602	0.13
2	Slope	0.001	20.35
3	Elevation	0.001	14.52
4	Shrink Swell Clay	0.001	30.48
5	A Road	0.001	6.57
6	B Road	0.001	13.85
7	C Road	0.001	392.95
8	Compressible	0.001	19.73

Figure 9: Lurking variable plot for the best fit model based on the expert elicitation process where covariates are added in sequential order (Model 3) for the plastic clean water model. Red indicates under prediction (+) whilst blue indicates over prediction (-) in the number of expected pipe bursts per cell.
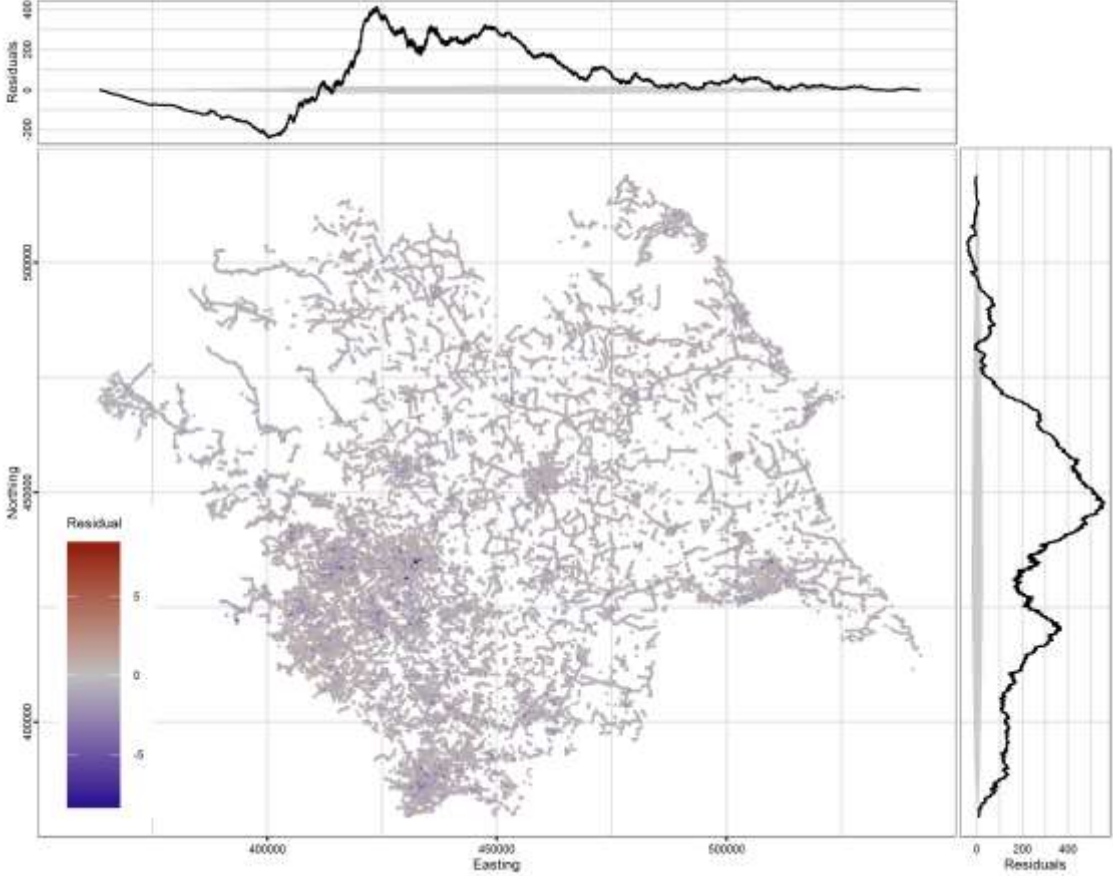
3.5.5 Addition of other environmental parameters to the Null model (Model 4)

Table 30 reports on the addition of the full range of geohazard and environmental factors to the null model of the plastic clean water network. In particular the addition of the number of Dwellings per unit area, the sulphur/sulphide geohazard dataset and the solubility dataset were found to improve the null model compared to other factors from the Expert Elicitation exercise. The DMA data was not included as plastic is considered resistant to internal corrosion after water treatment. Tables 31 - 33 provide information on the coefficients of the categorical variables that were significant in this exercise, that have not already been reported (Table 27 - 28). It was found that Aspect North and East, ground resistivity and CTI were not significant ($P < 0.05$).

Table 30: Full region output from spatial point process model fitting with a series of single covariates, added to a null model in which plastic pipe length is included as a covariate.

Order Added	Model	AIC	diffAIC	logLIK	LLr	pval	coef	Rank
7	C Road	290918.8	-619.60	-145456.4	621.60	0.001	0.00270	1
14	Dwellings	291065.5	-472.86	-145529.8	474.86	0.001	0.01388	2
15	Sulphur / Sulphide	291390.4	-147.95	-145691.2	151.95	0.001	N/A	3
12	Solubility	291430.9	-107.47	-145709.4	115.47	0.001	N/A	4
8	Compressible	291455.4	-82.94	-145721.7	90.94	0.001	N/A	5
4	Shrink Swell Clay	291490.1	-48.24	-145740.1	54.24	0.001	N/A	6
13	Corrosivity	291494.8	-43.53	-145743.4	47.53	0.001	N/A	7
2	Slope	291512.9	-25.52	-145753.4	27.52	0.001	0.01584	8
6	B Road	291514.7	-23.67	-145754.3	25.67	0.001	0.00164	9
5	A Road	291525.7	-12.71	-145759.8	14.71	0.001	0.00098	10
3	Elevation	291535.2	-3.18	-145764.6	5.18	0.022	-0.00026	11
11	Aspect North	291536.8	-1.58	-145765.4	3.58	0.058	-0.02462	12
9	A-Resisitivity	291537.3	-1.12	-145765.6	3.12	0.077	-0.00003	13
1	CTI	291540.1	1.72	-145767.1	0.27	0.602	0.00290	14
10	Aspect East	291540.2	1.84	-145767.1	0.15	0.698	-0.00504	15

Table 31 Coefficients of Soluble Ground categorical variables for the plastic clean water network using Model 4

class	coefficient	Ground Classification
A	-13.62546	Soluble rocks not thought to be present
B	-13.92987	Soluble rocks are present but unlikely to cause problems
C	-13.94777	Significant Soluble rocks are present with low possibility of localised subsidence or dissolution related degradation of bedrock
D	-13.95281	Very significant soluble rocks are present with a moderate possibility of localised natural subsidence or dissolution related degradation of bedrock
E	-13.67453	Very significant soluble rocks are present with a high possibility of localised subsidence or dissolution of bedrock

Table 32 Coefficients of Soil Corrosivity categorical variables for the plastic clean water network using Model 4

class	coefficient	Ground Classification
class 1	-13.64076	Unlikely to cause corrosion
class 2	-13.74753	May cause corrosion
class 3	-13.76599	Likely to cause corrosion

Table 33 Coefficients of Sulphur/Sulphide categorical variables for the plastic clean water network using Model 4

class	coefficient	Ground Classification
HIGH	-13.67361	Sulphate containing ground
LOW	-13.42907	Sulphide containing ground
NONE	-13.71043	Background concentrations of sulphate / sulphide

3.5.6 Fitting models by sequential addition of explanatory variables identified from the expert elicitation and other topographic and environmental indices (Model 5)

The plastic pipe model was then run as a sequential model with all the significant ($P<0.05$) covariates identified from Table 30. Results can be seen in Table 34. All covariates were significant after they were added sequentially with the order being C-Road >> Dwellings > Shrink swell clays > Sulphate / sulphide > Soluble ground ~ Compressible ground ~ Elevation ~ B roads ~Slope > A-roads. This again was slightly different to the order obtained in Table 17, suggesting that some factors were accounting for different parts of the residuals. Coefficients for the continuous covariates for the final sequential models can be found in Table 34, whilst coefficients for the categorical covariates can be found in tables 35 - 38. The coefficients for the continuous and categorical covariates will be used in the heat maps as they all share the same intercept. Figure 13 shows the result of Model 5 to predict the spatial density of plastic pipe failures across the YW region, and again the lurking variable plot demonstrates that the raw residual has decreased compared to Model 3.

Table 34: Results of sequential model (Model 5) for the plastic pipe network across the YW region

Order added	Model	pval	LLr	coef
1	Slope	0.001	13.76181	0.01584
2	Elevation	0.001	18.28425	0.01324
3	Shrink swell clay	0.001	30.04497	N/A
4	A Road	0.001	6.76801	-0.02339
5	B Road	0.001	13.99134	-0.01912
6	C Road	0.001	388.21845	0.01605
7	Compressible Ground	0.001	18.48142	N/A
8	Soluble ground	0.001	18.80902	N/A
9	Dwellings	0.001	36.38143	-0.04924
10	Sulphate / Sulphide	0.001	27.49916	N/A

Table 35 Coefficients of shrink swell categorical variables for the plastic clean water network using Model 5

Class	coefficient	Ground Classification
A	-13.56305	Ground conditions predominantly non plastic; No action
B	-13.50838	Ground conditions predominantly low plastic; No action
C	-13.55372	Medium plasticity; action required
D	-13.48841	High Plasticity

Table 36 Coefficients of Compressible ground categorical variables for the plastic clean water network using Model 5

class	coefficient	Ground Classification
A	-13.56305	No indicators of compressible ground – No action
B	-13.60241	Very slight potential of compressible deposits
C	-12.70039	Slight possibility of compressibility problems
D	-13.63478	Significant potential for compressibility problems
E	-14.65151	Very significant potential of compressibility problems

Table 37 Coefficients of Soluble Ground categorical variables for the plastic clean water network using Model 5

class	coefficient	Ground Classification
A	-13.56305	Soluble rocks not thought to be present
B	-13.72095	Soluble rocks are present but unlikely to cause problems
C	-13.79242	Significant Soluble rocks are present with low possibility of localised subsidence or dissolution related degradation of bedrock
D	-13.68829	Very significant soluble rocks are present with a moderate possibility of localised natural subsidence or dissolution related degradation of bedrock
E	-13.59609	Very significant soluble rocks are present with a high possibility of localised subsidence or dissolution of bedrock

Table 38 Coefficients of sulphate and sulphide categorical variables for the plastic clean water network using Model 5

class	coefficient	Ground Classification
HIGH	-13.56305	Sulphate containing ground
LOW	-13.39823	Sulphide containing ground
NONE	-13.57022	Background concentrations of sulphate / sulphide

Figure 10 Result of Model 5 for the YW region using sequential addition of covariates for plastic pipe failures. Note the decrease in the cumulative sum of raw residuals compared to the null model in Figure 9. Examination of the combined X and Y axis residuals suggest that overall the model is under predicting the number of pipe failures per unit length of pipe, with the red colours indicating where this is happening to the greatest extent and the blue the least.
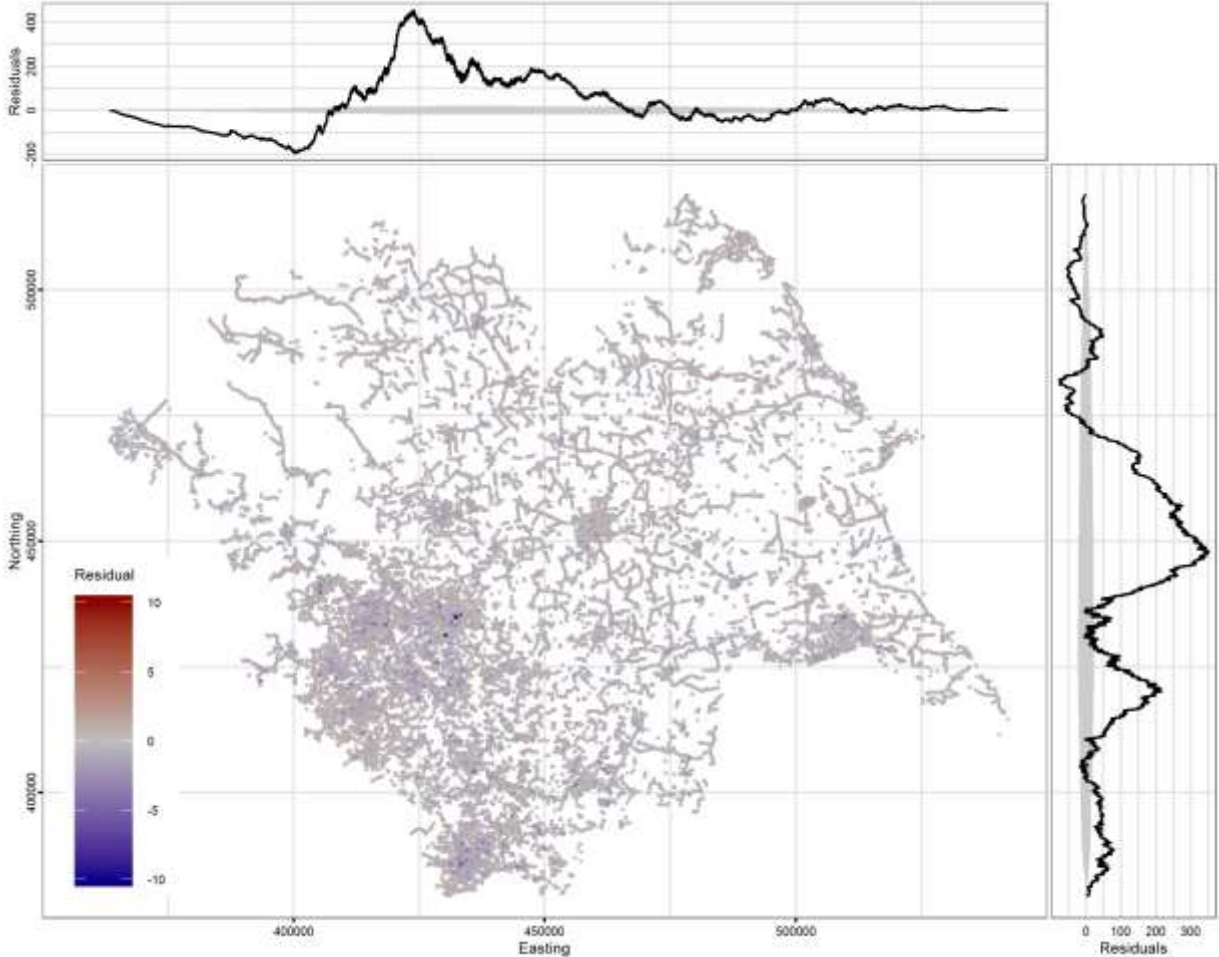
3.5.7 Discussion of the Plastic pipe network

3.5.8 Model Performance

The addition of the co-variables from the Expert Elicitation process (Model 3) to the Null model reduced the total raw residual of the model as demonstrated in the Lurking variable plots. Model 3 was then improved further by the addition of extra environmental co-variables (Model 5). The greatest areas of model under-prediction appear to be associated with the Coal Measures and the Millstone Grit group, suggesting that subsidence or faulting may have a detrimental effect on pipeline stability. Considered explanations for the effects the variable and categorical variables may have on the plastic pipe network are shown in Table 39 below. For the categorical variables it was found that for shrink swell and compressible ground, the coefficients do not follow a linear trend, possibly reflecting the potential for soil-structure-pipe interactions in different soil

types. Although, plastic pipe is considered to be more resistant to ground movements than rigid pipes because of their flexibility (Olliff et al. 2001), an important part of their installation is how much deflection the soil structure-pipe interaction enables the plastic to undergo before it is damaged. It is possibly that this property, which is likely to change with soil type, is determining the coefficients in these categorical classes, where there does not appear to be a major effect of the geohazard. However, it is possible to link some categorical coefficients to specific geological units. For instance, for the compressible ground, the one area which has the highest coefficient can be identified as the alluvium of the River Don near Sheffield, suggesting that in this area ground conditions may promote plastic pipe failure. As the alluvium is considered to be reasonably homogenous, chemical pollution from the steel industry that interacts with the plastic could be considered.

Table 39 Interpretation of the outputs from adding individual covariates to the Null model for the plastic pipe clean water network in the YW region.

Rank	Covariate	+/- coefficient	Notes
1	C Road	+	Positive correlation between the density of C roads in a cell and pipe failure per unit length. Could be a result of lower quality road construction designed for lower frequency and load of vehicles. Vibration and resulting friction with the sub grade would appear to be key processes
2	Number of Dwellings	+	Positive correlation again suggesting a link between possible pipe pressure and use.
3	Sulphur/Sulphide	N/A	The Low Class within the BGS dataset had the highest coefficient and this represents the soils that are likely containing sulphide. Thus this largely represents the area related to the coal measures and may represent a proxy for old mining subsidence, as the acidity produced via sulphide oxidation is not considered a major impact on plastic pipes. However, this effect also appears to be high on the Kimmeridge clay near Scarborough. The coefficients for the High Sulphates / sulphides class which covers the sulphate bearing soils and the Background class are similar.
4	Soluble ground conditions	N/A	The lowest coefficient values are found for Class B and Class C which represent the chalk and limestone bedrocks. These soils typically have thin soils so may have something to do with being based on the pipes being sited on hard rocks. Class D, E and Class A have similar coefficients and may represent deeper soils that behave similarly
5	Compressible Ground	N/A	The behaviour of plastic pipe in compressible ground can result in either higher or lower stability according to the combination of pipe and ground conditions. The highest coefficient (Class C) represents only the alluvium of the river Don near Sheffield, suggesting pollution effects on the plastic. Class E the lowest coefficient represents peat deposits where failures may be difficult to detect. The other classes have similar coefficients suggesting that there is no great difference between ground conditions.
6	Shrink - swell	N/A	Class C & D have lower coefficients than Class A & B suggesting that the plastic pipe is achieving greater support within some deposits capable of ground movements, or possibly self sealing if leaks are present. This is recognised as occurring with plastic pipes and some deformable soils. No Class E is present in the YW region.

7	Soil Corrosivity	N/A	The lowest coefficients are found for the Class 2 and Class 3 which are both categories suggesting enhanced corrosion for ferrous iron. However, for the plastic pipe network these classes may indicate those classes with greater clay content, thus possibly acting as a proxy for ground movements which may help accommodate pipe stability. This is a processes indicated by the Compressible ground and shrink swell classes.
8	Slope	+	The role of slope may indicate an effect of sideways pressure on causing pipe movement. This is likely to be pressure on joints.
9	B road	+	Positive correlation between the density of B roads in a cell and pipe failure per unit length. Could be a result of lower quality road construction designed for lower frequency and load of vehicles. Vibration and resulting friction with sub grade would appear to be key processes.
10	A road	+	Positive correlation between the density of A roads in a cell and pipe failure per unit length. Could be a result of lower quality road construction designed for lower frequency and load of vehicles. Vibration and resulting friction with sub grade would appear to be key processes. However, A-road only accounts for a very small diffAIC.
11	Elevation	-	Negative correlation between pipe failure per unit length and elevation suggesting that more failures occur at low elevation. However, this only accounts for a small very diffAIC. This could relate to differences in the thermal regime. PVC pipe has thermal expansion up to 5x that of ductile iron which may affect pipes.

3.5.9 Using coefficients from the sequential model to produce heat maps

As the coefficients from the sequential model share a common intercept they can be used to directly compare their influence within the model through heat maps. Although the heat maps all appear largely green there are isolated red values (C-roads) which produce the range that the co-variable coefficients are standardised on. Figure 14 shows the heat maps for the significant variables from Model 5.

Figure 11 Heat Maps produced from the coefficients of significant co-variables using Model 5 for the YW clean water plastic pipe network
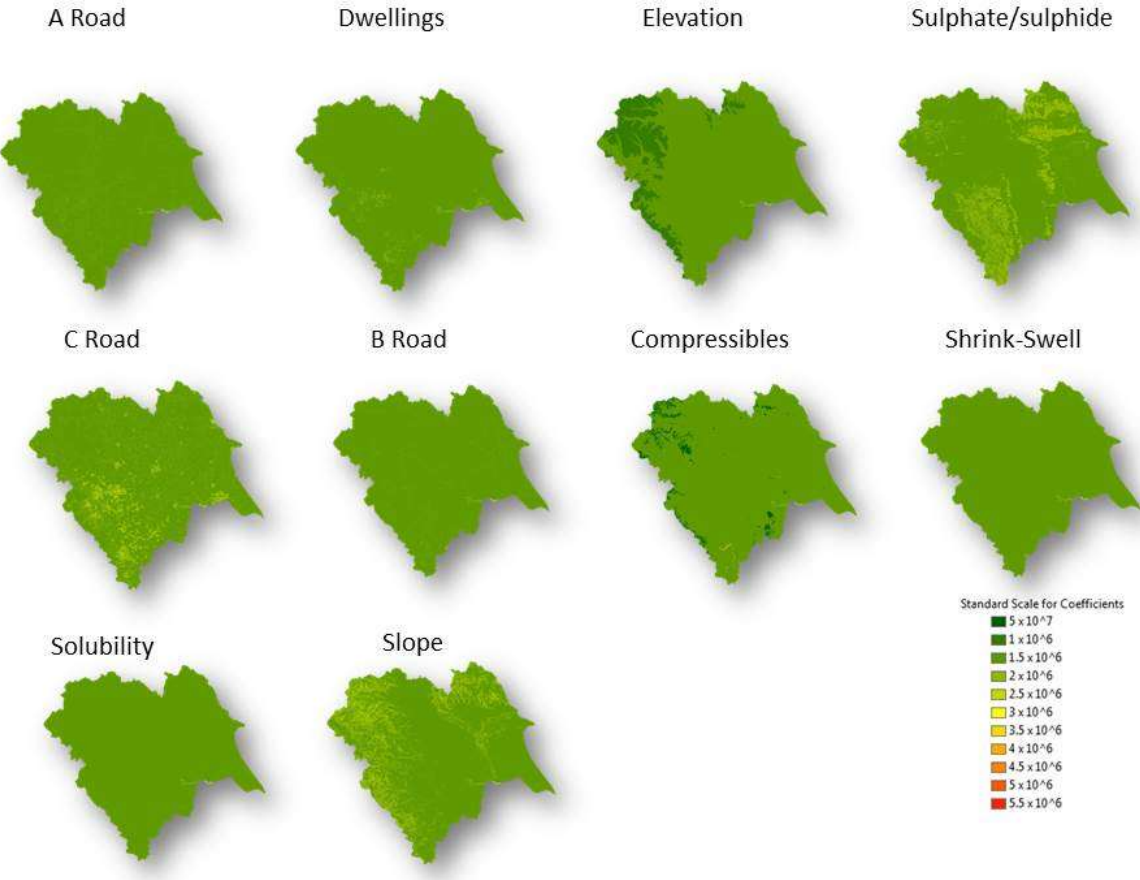
The heat maps demonstrate the spatial effect of the covariates. Thus the greatest effects on the plastic pipe network can be seen in the urban areas of Leeds and Bradford as demonstrated by the higher coefficients of the C-roads and the spatial distribution of the dwellings. Geological based hazards can be seen in the sulphate / sulphide heat map which identifies the coal measure areas. Slope appears to be important in the north east and north west of the YW region. However, most of the geological coefficients are low (green) and can be considered to have a minor influence compared to the anthropogenic influenced factors (roads). The overall coefficient intensity maps were also produced by multiplying together the coefficients for each 100 x 100m cell. This provides an indication where the most hostile environments for the plastic pipe network are for the whole YW region area is shown in Figure 15.

Figure 12 Overall heat map showing intensities of hostile environments to plastic pipe network across the YW region using the coefficients produced from significant co-variables using Model 5
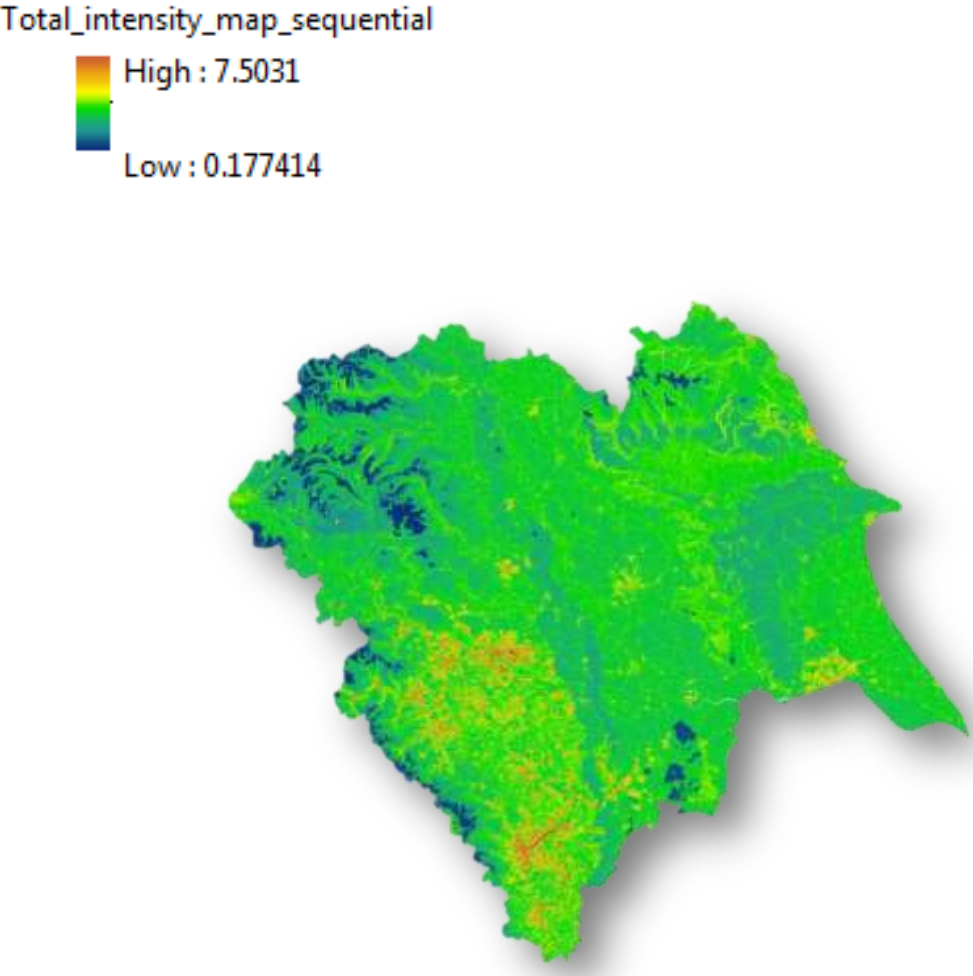
3.6   THE WASTE WATER CONCRETE PIPE NETWORK

Having developed the methodology for the clean water cast iron and plastic pipe networks, a similar approach was applied to the waste water networks where the major materials are concrete and clay. The change in materials and quality of water produces a different set of factors that may influence pipe failure. One of the most important factors in the failure of concrete pipes is that internally the pipes can be corroded through the production of H_2S gas from bacterial decomposition of sewage, leading to the production of H_2SO_4. Thus, differential ground movement can then act on the internally corroded concrete pipes leading to failure. A key factor influencing H_2S production is the slope of the pipes because it determines the speed at which sewage is moved along. A major factor externally comes from the presence of sulphate in soils because this can lead to concrete attack through the formation of the mineral thaumasite in the concrete, which helps break the concrete apart. Thus the first step was to undertake an Expert Elicitation for the concrete pipe network.

3.6.1 Expert elicitation for concrete pipework

An initial EE was undertaken in January 2015 for the original NERC grant. However prior to undertaking the modelling exercise it was repeated by e-mail for the current project, where we have the potential to include information from the DMA and number of dwelling datasets. The expert elicitation produced the following order of factors that the covariates should be tested for in the EE model.

1. External Sulphate / sulphide
2. Slope
3. Road vibration (this could be a proxy for depth of pipe)
4. Differential ground movement – Shrink swell, Compressible ground

Other contributory factors considered problematic but for which data was not available or reasonable proxies could be used were for mining collapse, water depth and the removal of external support by water removal of soil.

3.6.2 The Null model (Model 1)

For the modelling procedure developed, a null model was produced for the expected number of failures per unit area based on the density of pipework (Figure 16).

Figure 13 The null model for the concrete waste water network for the Yorkshire Water region where the density of bursts is a function of the log density of concrete pipe in each 100 x 100m cell. Red indicates model under prediction (positive residuals) and blue over prediction (negative residuals).
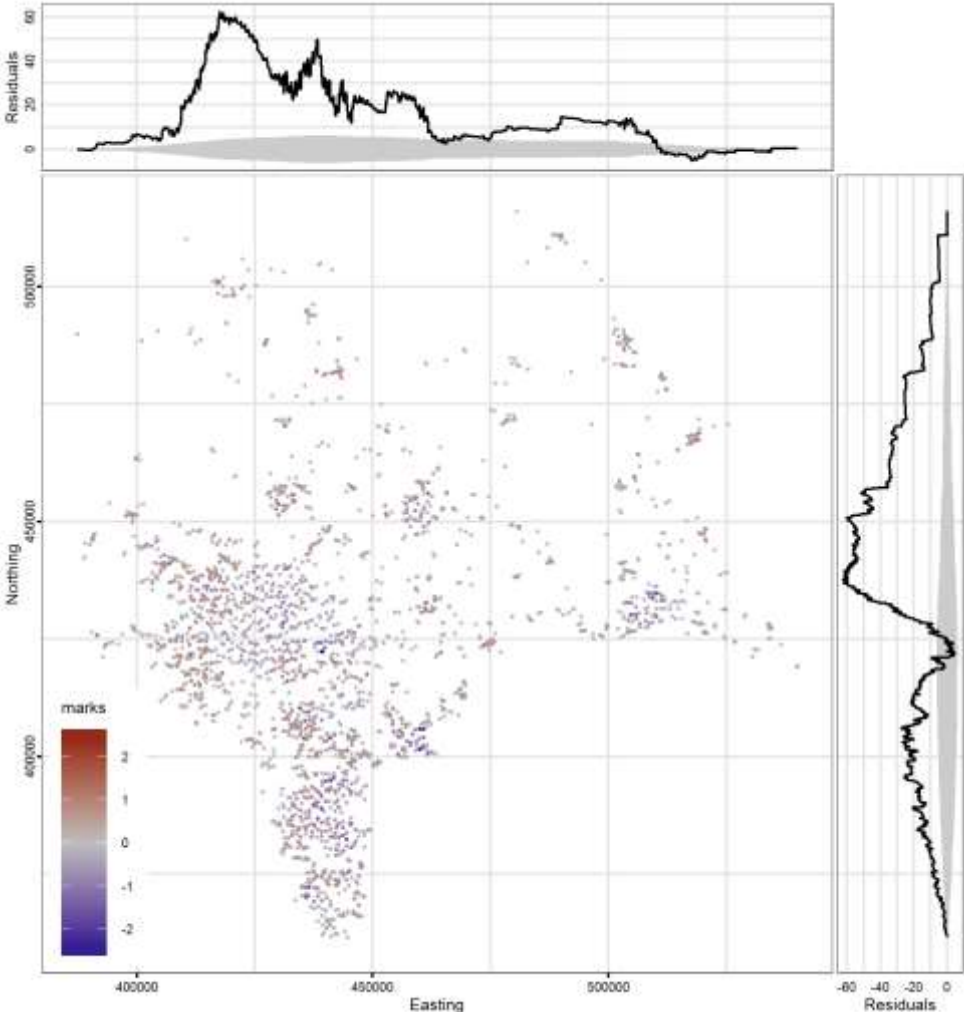
3.6.3 Adding single factors from the Expert Elicitation exercise to the Null Model (Model 2)

The next stage of the modelling procedure requires the covariates identified through the Expert Elicitation process to be added to the null model one at a time to assess the contribution that they make to the pipe failure process. Results can be found in Table 40. The table shows the order in which they were added and their eventual rank. Results showed that in order of importance the factors were Slope > B Roads > Compressible deposits > Shrink swell clay > A Roads > C roads. Surprisingly, the sulphate and sulphide dataset was found not to be significant at $P < 0.05$. However, the results showed that the remaining factors picked out in the Expert Elicitation process were all found to be highly significant. The continuous covariates all had positive

correlations with concrete pipe failure. The coefficients for the categorical covariates can be seen in Tables 41 and 42. The very low coefficient in for Class 4 in the shrink swell model is because no pipe failures were recorded in this class, although pipe was present.

Table 40 Outputs from running the Null model with individual predictor variables using Model 2

Order Added	Model	AIC	diffAIC	logLIK	LLr	pval	coef	rank
2	Slope	17890.80	-28.52	-8942.3	30.52	0.001	0.08033	1
4	B Road	17900.41	-18.90	-8947.2	20.90	0.001	0.00577	2
7	Compressible	17911.88	-7.43	-8950.9	13.43	0.003	N/A	3
6	Shrink swell clay	17915.36	-3.95	-8952.6	9.95	0.018	N/A	4
3	A Road	17912.54	-6.77	-8953.2	8.77	0.003	0.00288	5
5	C Road	17913.38	-5.93	-8953.6	7.93	0.004	0.00125	6
1	Sulphate / Sulphide	17917.77	-1.54	-8954.8	5.54	0.062	N/A	7

Table 41 Coefficients of shrink swell clay categorical variables for the concrete waste water network using Model 2

class	coefficient	Ground Classification
A	-14.92473	Ground conditions predominantly non plastic; No action
B	-15.16688	Ground conditions predominantly low plastic; No action
C	-15.14786	Medium plasticity; action required
D	-26.78188	High Plasticity

Table 42 Coefficients of the compressible ground categorical variables for the concrete waste water network using Model 2

class	coefficient	Ground Classification
A	-14.96005	No indicators of compressible ground – No action
B	-15.38384	Very slight potential of compressible deposits
C	-16.08610	Slight possibility of compressibility problems
D	-15.36875	Significant potential for compressibility problems

3.6.4   Adding the EE covariates sequentially to the Null model for the concrete waste water network (Model 3)

The next stage of the modelling process was to add the significant (P < 0.05) variables to the null model in sequential order. This produced a slightly different order of importance in the co-variables where Slope > C road > B- road > Compressible > A road. Shrink swell clay was no longer significant (P<0.05). The results can be seen in Table 43, with the completed model in Figure 17. Decreases in the residuals can be seen compared the null model.

Table 43 P-value from tests for sequential addition of statistically significant covariates identified from the expert elicitation added to the null model (Model 3). LLr is the log likelihood ratio statistic expressing how many times more likely the data are based on addition of this covariate in comparison to the previous model

	Model	pval	LLr
1	Slope	0.001	16.70
2	A Road	0.007	3.61
3	B Road	0.001	9.76
4	C Road	0.001	11.37
5	Shrink swell clay	0.467	1.27
6	Compressible	0.035	4.29

Figure 14 Final lurking variable plot for the best fit model based on the expert elicitation process where covariates are added in sequential order (Model 3) for the concrete waste water network. The red areas indicate where the model under predicts the number of expected pipe bursts per cell, whilst the blue over-predicts per 100 x 100m cell
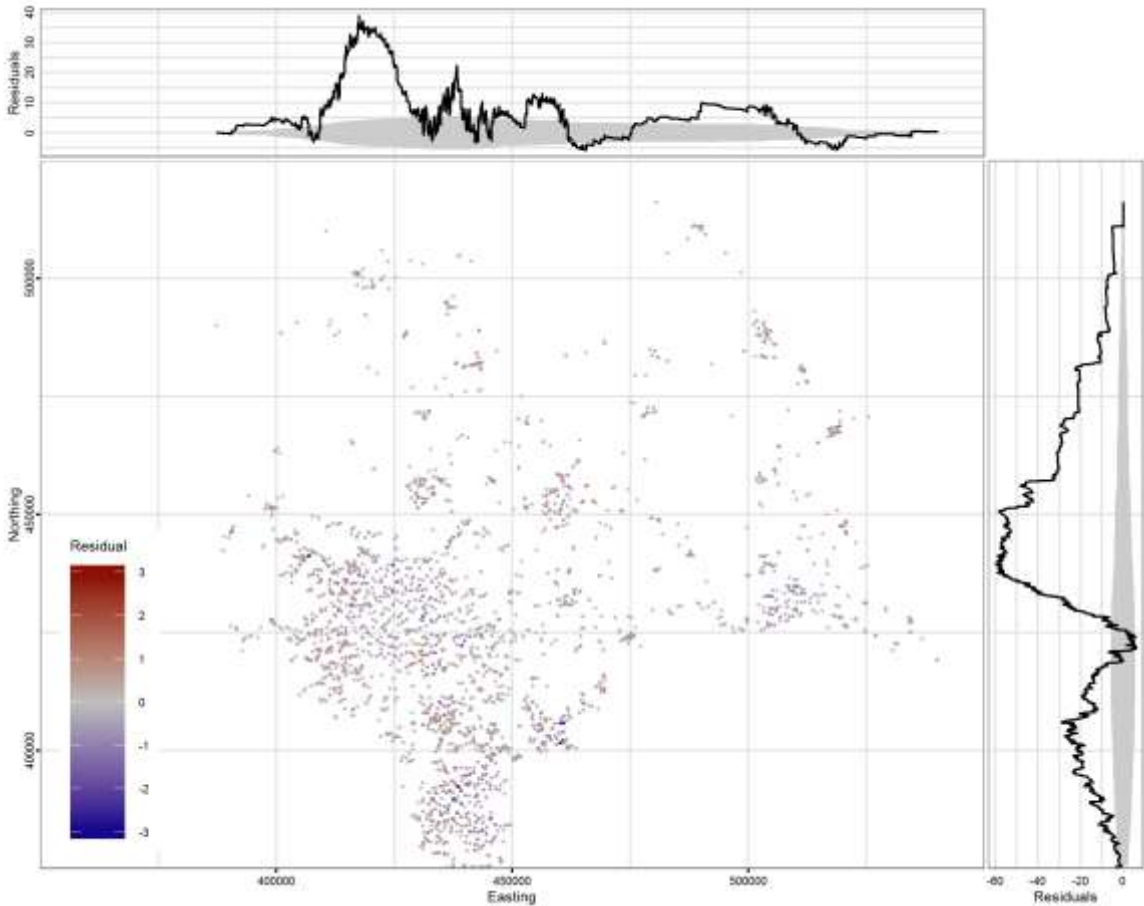
3.6.5 Adding other environmental factors to the EE model (Model 4)

The other environmental factors were then added onto the null model one at a time to assess whether they are significant and their importance, as demonstrated by the LLr (Table 44). It was found that the new order of ranking was Number of Dwellings > Slope > Solubility > B Road > Compressible deposits > Shrink swell > A road > C road. The remaining covariates shown in Table 44 were not found to be significant at P < 0.05. There were positive correlations between the expected pipe failures and the significant (P < 0.05) continuous covariates. The coefficients for the new categorical covariates tested can be seen in Tables 45.

Table 44 Full region output from spatial point process model fitting with a series of single covariates, added to a null model in which plastic pipe length is included as a covariate (Model 4)

	Model	AIC	diffAIC	logLIK	LLr	pval	coef	rank
14	Dwellings	17886.76	-32.55	-8940.3	34.55	0.001	0.01755	1
2	Slope	17890.80	-28.52	-8942.3	30.52	0.001	0.08033	2
12	Solubility	17904.86	-14.45	-8946.4	22.45	0.001	N/A	3
4	B Road	17900.41	-18.90	-8947.2	20.90	0.001	0.00577	4
7	Compressible	17911.88	-7.43	-8950.9	13.43	0.003	N/A	5
6	Shrink Swell	17915.36	-3.95	-8952.6	9.95	0.018	N/A	6
3	A Road	17912.54	-6.77	-8953.2	8.77	0.003	0.00288	7
5	C Road	17913.38	-5.93	-8953.6	7.93	0.004	0.00125	8
1	Sulphate / Sulphide	17917.77	-1.54	-8954.8	5.54	0.062	N/A	9
13	Corrosivity	17919.12	-0.19	-8955.5	4.19	0.122	N/A	10
8	CTI	17918.38	-0.93	-8956.1	2.93	0.086	-0.03869	11
9	A Resistivity	17920.57	1.25	-8957.2	0.74	0.387	0.00007	12
10	Aspect East	17921.25	1.92	-8957.6	0.07	0.790	0.01516	13
11	Aspect North	17921.31	1.99	-8957.6	0.00	0.925	0.00520	14

Table 45 Coefficients of the Soluble Ground categorical variables for the concrete waste water network obtained from using Model 4

class	coefficient	Ground Classification
A	-15.06431	Soluble rocks not thought to be present
B	-15.03381	Soluble rocks are present but unlikely to cause problems
C	-14.98741	Significant Soluble rocks are present with low possibility of localised subsidence or dissolution related degradation of bedrock
D	-13.99384	Very significant soluble rocks are present with a moderate possibility of localised natural subsidence or dissolution related degradation of bedrock
E	-12.32065	Very significant soluble rocks are present with a high possibility of localised subsidence or dissolution of bedrock

3.6.6 Adding variables sequentially for the concrete waste water network (Model 5)

The significant environmental covariates were added in a sequential model and the improvement in the model output caused by the inclusion of these additional factors is noted. This process is reported in Table 46. Within this new model, the order of importance was Slope > soluble ground > C-road > Dwellings > B-road > A-road whilst compressible ground and shrink swell clay were not significant at $P<0.05$. The continuous coefficients are found in Table 46, whilst the categorical variables are shown in Tables 47.   Figure 18 shows Model 5.

Table 46 Metrics of sequential addition (Model 5) of expert elicited predictor variables to sequential model, starting from null model

Order added	Model	pval	LLr
1	Slope	0.001	15.26
2	A Road	0.004	4.00
3	B Road	0.001	9.53
4	C Road	0.001	11.37
5	Shrink swell clay	0.259	2.00
6	Compressible	0.051	3.87
7	Solubility	0.001	12.21
8	Dwellings	0.001	9.78

Table 47 Coefficients of the Soluble Ground categorical variables for the concrete waste water network using model 5

class	coefficient	Ground Classification
A	-15.70013	Soluble rocks not thought to be present
B	-15.56259	Soluble rocks are present but unlikely to cause problems
C	-15.26850	Significant Soluble rocks are present with low possibility of localised subsidence or dissolution related degradation of bedrock
D	-14.22220	Very significant soluble rocks are present with a moderate possibility of localised natural subsidence or dissolution related degradation of bedrock
E	-12.97769	Very significant soluble rocks are present with a high possibility of localised subsidence or dissolution of bedrock

Figure 15. Result of full model for the YW region using sequential addition of covariates for the concrete waste water network (Model 5). Note the decrease in the cumulative sum of raw residuals compared to the null model in Figure 9. Examination of the combined X and Y axis residuals suggest that overall the model is under predicting the number of pipe failures per unit length of pipe, with the red colours indicating where this is happening to the greatest extent and the blue the least.
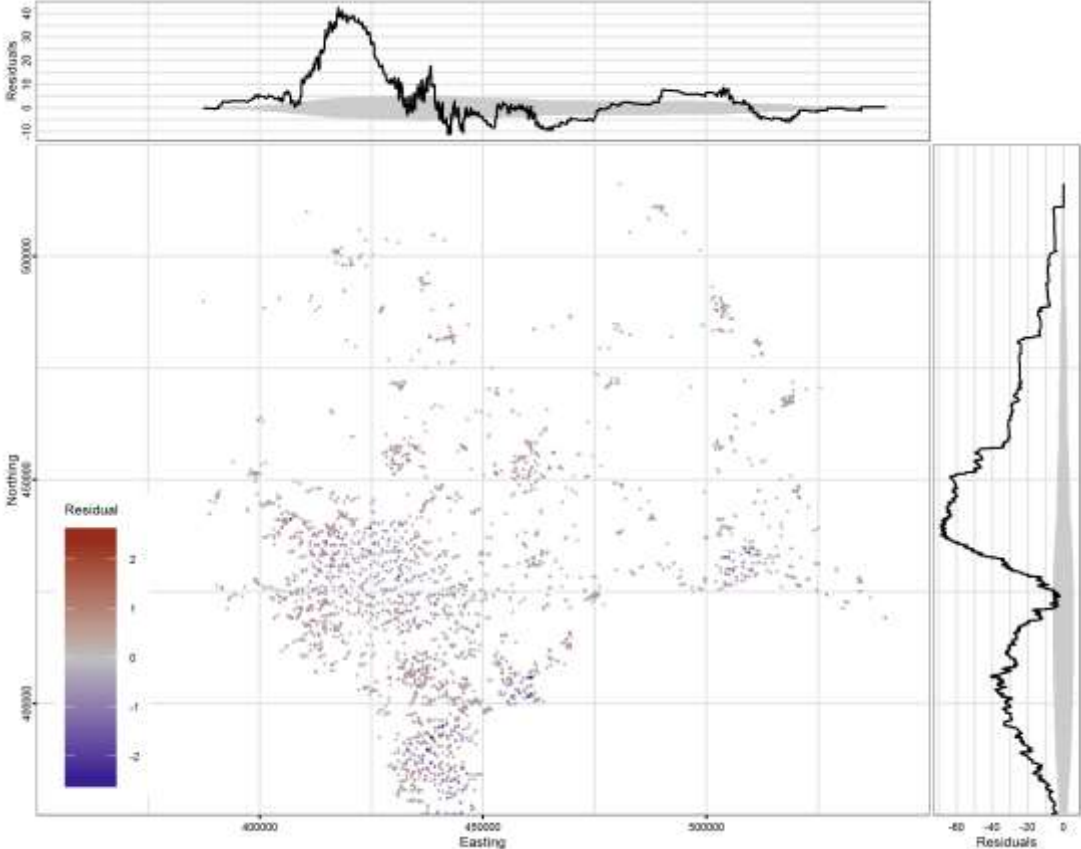
3.6.7   Discussion of the Concrete pipe network

The addition of the covariates from the Expert Elicitation process (Model 3) to the Null model reduced the total raw residual of the model as demonstrated in the lurking variable plots. The EE model was then improved further by the addition of extra environmental co-variables (Model 5). Analysis of the over and under prediction produced by the model is complicated to untangle as both appear in the same south western region of the model. This is the area of greatest urban density, along with the coal measures that may induce subsidence. Considered explanations for the continuous and categorical coefficients obtained when added individually to the Null model are shown in Table 48. Positive correlations were found for the continuous variables. In particular increasing slope and the possible effects of gravity on heavy pipes, especially when full, along with problems associated with pipelines in or close to roads are the major issues. The

categorical variables associated with compressible ground may therefore be more related to settlement and deflection in different soil types.

Table 48 What the coefficients mean for the concrete waste water network models.

Rank	Covariate	+/- coefficient	Notes
1	Number of dwellings	+	Positive correlation between number of dwellings and pipe failure suggests that increased use of the pipe network has a detrimental effect. For example greater H_2S gas production causing internal corrosion.
2	Slope	+	A positive correlation between expected pipe failure and slope suggests that the weight of waste in the pipe on slopes may cause greater failure. This is not the slope at which the pipes may be laid to increase the rate of flow, thus decreasing H_2S production, but is indicative of heavy weight causing sideways movement.
3	Solubility	N/A	There was an increase in the size of the coefficients suggesting that as solubility of rock conditions increased there was greater pipe failure. Whilst the coefficient values were very similar for Classes A, B, and C, they increased considerably for Classes D and E. These two classes take into account the gypsum bearing rocks around Rippon, suggesting that subsidence may occur, causing failure but also that the presence of sulphate from gypsum may contribute to failure through concrete rot.
4	B Road	+	A positive correlation suggesting that vibration and road effects may increase failure
5	Compressible Ground	N/A	The coefficients indicate that the background ground conditions has the larger coefficient than for areas where there is an increasing compressible ground problem. The suggestion is that the coefficients from classes B-D may be relating to the soil structure – pipe interactions which vary between soils and affect pipe rigidity. No concrete pipe was found in Class 5 which is why it is missing.
6	Shrink Swell	N/A	The suggestion from the coefficients is that the background ground condition has the largest coefficient compared to areas where there is an increasing shrink swell ground problem. The suggestion is that the coefficients from classes B-D may relate more to the soil structure – pipe interactions which vary between soils and affect pipe rigidity. There is a very significant decrease in coefficient size for the highest shrink swell class. Some concrete pipe is found in Class 5 but no failures have been recorded which explains the very ow coefficient.
7	A Road	+	A positive correlation suggesting that vibration and road effects may increase failure
8	B Road	+	A positive correlation suggesting that vibration and road effects may increase failure

3.6.8 Using coefficients from the sequential model to produce heat maps

Coefficients from Model 5 were used to produce individual heat maps (Figure 19). One of the reasons why the maps appear largely green, is that there is only one cell which is red (on the slope map in the NW corner), thus representing the high coefficient, meaning that the coefficients still need to be scaled to this. Whilst most of the heat maps are green, suggesting little difference in the low impact of the co-variables, yellow colours indicating greater impact can be seen in the slope map and particularly on the solubility maps. The area of greatest impact

is associated with solubility on the gypsum bearing rocks around Rippon where ground subsidence is likely. This again is highlighted in the total intensity map (Figure 20).

Figure 16 Heat Maps for the YW concrete waste water network where coefficients from the significant co-variables from Model 5 are plotted on a standardised colour scale
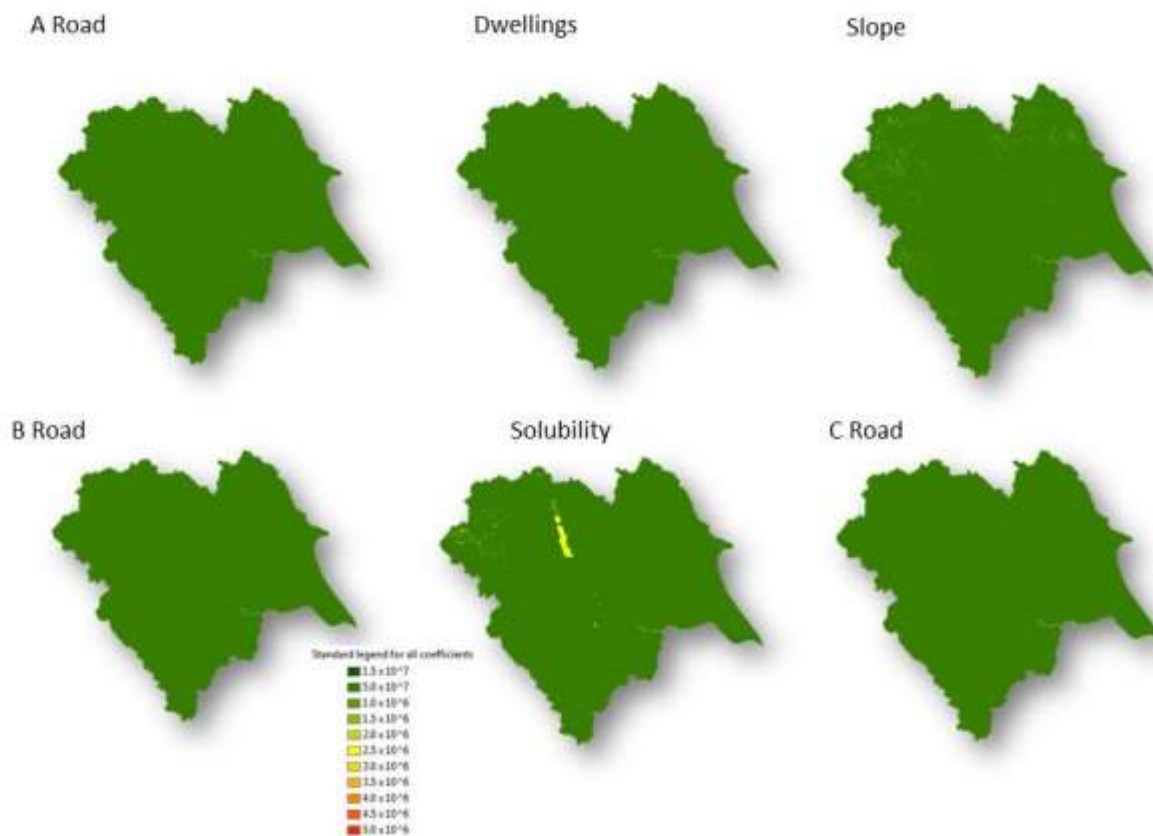
Figure 17 Total Intensity map of YW region for the concrete waste water network showing areas which are most hostile to pipe networks produced using significant variables obtained using model 5.

Total_intensity_map_sequential
High : 12.7542
Low : 0.190778
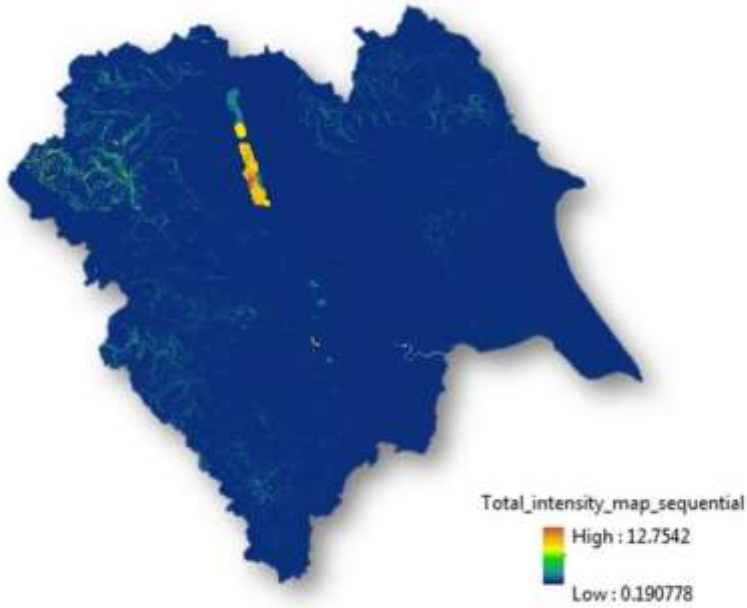
3.7   THE WASTE WATER CLAY PIPE NETWORK

3.7.1   Expert Elicitation

The Expert Elicitation process was similar to that for the concrete waste water network, with the exception that the presence of sulphide and sulphate are not considered an issue with clay as it is with concrete. Thus, after the EE exercise the order that variables should be introduced into the model was

1. Slope or pipe fall (this provides an indication of how quickly sewage will be transported thus reducing H_2S production). This is also important because of the weight of the full pipes
2. Road vibration (this could be a proxy for depth of pipe)
3. Differential ground movement – Shrink swell, Compressible ground

3.7.2   The Null model (Model 1)

The null model is shown below in Figure 21. The null model shows, that based purely on the density of pipes per 100 x 100 m cell the greatest model under performance is in an area in the

SW of the YW region, typically coinciding with Carboniferous rocks types such as the Millstone Grit and Lower Coal Measures.

Figure 18 The null model for the whole of the Yorkshire Water region where the density of bursts is a function of the log density of clay waste water pipe in each 100 x 100m cell. Red indicates model under prediction (positive residuals) and blue over prediction (negative residuals).
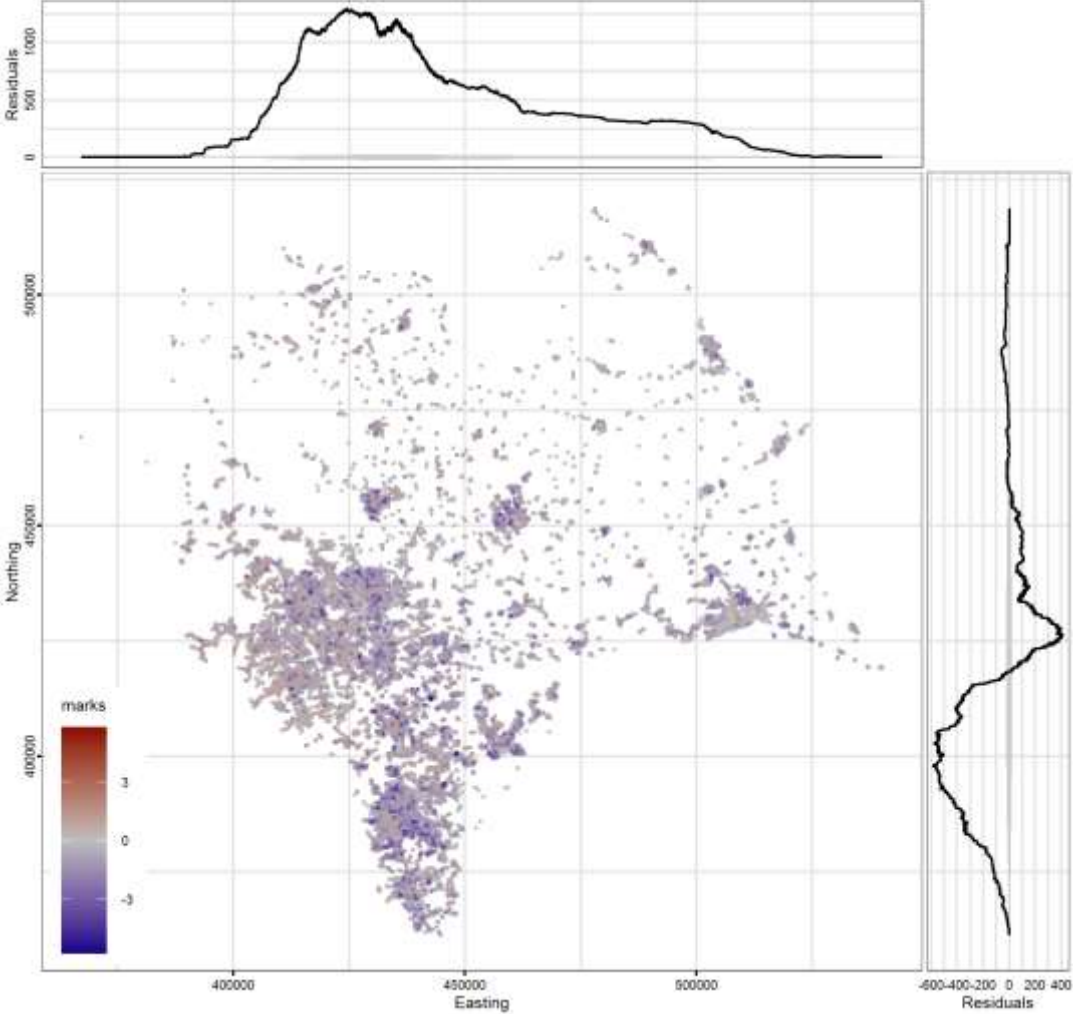
3.7.3 Adding single covariates from the EE to the null model (Model 2)

The next stage was to add single covariates identified in the Expert Elicitation process to the Null model (Table 49). Results show that the importance of covariates were in the order Slope >> C road >> A road > Compressible ground > Shrink swell clay > B Road. Positive coefficients indicating a positive correlation between the continuous covariates (Slope and Road types) were found. The coefficients for the categorical covariates are presented in Tables 50 and 51.

Table 49 Outputs from running the Null model with individual predictor variables for the clay waste water network (Model 2)

Order added	Model	AIC	diffAIC	logLIK	LLr	pval	coef	rank
1	Slope	243370.5	-528.37	-121682.3	530.37	0.001	0.07749	1
4	C Road	243695.3	-203.54	-121844.7	205.54	0.001	0.00177	2
2	A Road	243793.8	-105.07	-121893.9	107.07	0.001	0.00279	3
6	Compressible	243807.3	-91.62	-121898.6	97.62	0.001	N/A	4
5	Shrink Swell Clay	243815.6	-83.24	-121902.8	89.24	0.001	N/A	5
3	B Road	243853.3	-45.56	-121923.7	47.56	0.001	0.00255	6

Table 50 Model coefficients for Shrink swell clays for the clay pipe waste water network (No Class E present in YW region) obtained using Model 2

class	coefficient	Ground Classification
A	-15.75950	Ground conditions predominantly non plastic; No action
B	-15.74784	Ground conditions predominantly low plastic; No action
C	-16.24323	Medium plasticity; action required
D	-16.69923	High Plasticity

Table 51 Model coefficients for compressible ground conditions for the clay pipe waste water network (No pipework in Class E through YW region) obtained using Model 2

class	coefficient	Ground Classification
A	-15.75234	No indicators of compressible ground – No action
B	-16.65175	Very slight potential of compressible deposits
C	-16.98420	Slight possibility of compressibility problems
D	-16.10494	Significant potential for compressibility problems

3.7.4   Adding covariates sequentially to the Expert Elicitation model (Model 3)

The next stage was to add the covariates identified from the EE exercise to the null model in sequential order. Table 52 reports the extent to which each variable improves the model, with slope and C roads being the most important. The order of the co-variables with respect to their impact is slightly changed in the sequential EE model compared to adding the co-variables individually in that the order is Slope > C-road > A-road > B-road > shrink swell > compressible ground. The output of the model can be seen in Figure 22 and it can be seen that adding the EE variables has reduced the model residuals.

Table 52 Metrics of sequential addition of expert elicited predictor variables to sequential model, starting from null model (Model 3)

	Model	pval	LLr
1	Slope	0.001	265.18
2	A Road	0.001	45.64
3	B Road	0.001	26.90
4	C Road	0.001	187.57
5	Shrink swell clay	0.001	25.35
6	Compressible	0.001	17.60

Figure 19 Final lurking variable plot for the best fit model based on the expert elicitation process where covariates are added in sequential order (Model 3). The red areas (positive residual) indicate where the model under predicts the number of expected pipe bursts per cell, whilst the blue (negative residual) over-predicts per 100 x 100m cell
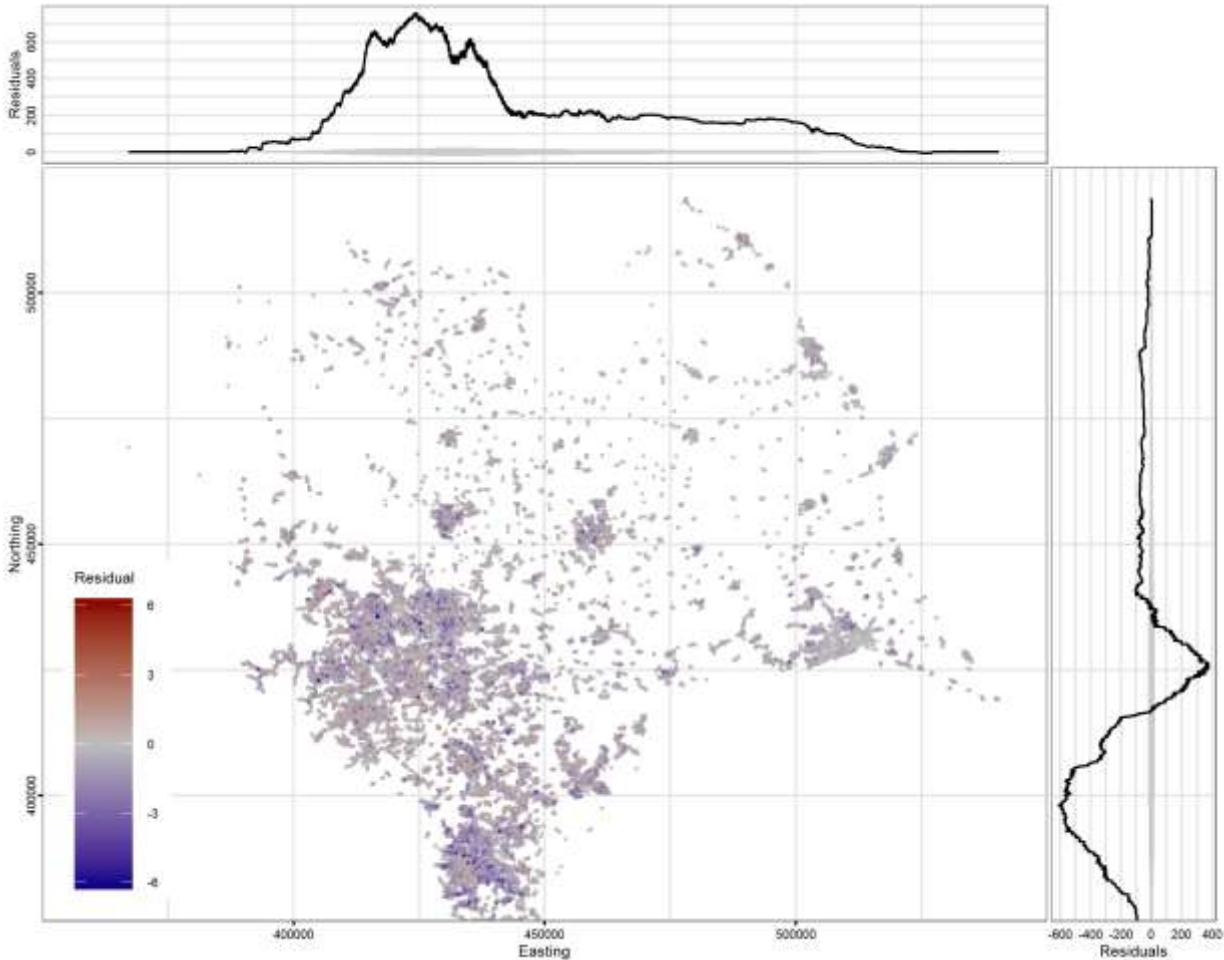
3.7.5 Adding other environmental factors individually to the Null model (Model 4)

To assess whether other environmental factors may contribute to the density of pipe failure we then added each as a single covariate to the Null model. Out of the new variables added the

Number of Dwellings, corrosivity, CTI, solubility and A-Resistivity were found to be highly significant (Table 53). The number of Dwellings showed a positive correlation, whilst the CTI showed a negative correlation. Coefficients for the categorical variables added to the null model individually are shown in Table 54 and 55. Aspect north and East were found not to be significant at $P<0.05$.

Table 53 Metrics of Model 4 where individual predictor co-variables are added to the null model independently of each other

Order added	Model	AIC	diffAIC	logLIK	LLr	pval	coef	rank
1	Slope	243370.5	-528.37	-121682.3	530.37	0.001	0.07749	1
13	Dwellings	243452.6	-446.28	-121723.3	448.28	0.001	0.01306	2
4	C Road	243695.3	-203.54	-121844.7	205.54	0.001	0.00177	3
12	Corrosivity	243731.8	-167.08	-121861.9	171.08	0.001	N/A	4
7	CTI	243768.7	-130.24	-121881.3	132.24	0.001	-0.08053	5
2	A Road	243793.8	-105.07	-121893.9	107.07	0.001	0.00279	6
6	Compressible	243807.3	-91.62	-121898.6	97.62	0.001	N/A	7
5	Shrink Swell Clay	243815.6	-83.24	-121902.8	89.24	0.001	N/A	8
3	B Road	243853.3	-45.56	-121923.7	47.56	0.001	0.00255	9
11	Solubility	243862.4	-36.47	-121925.2	44.47	0.001	N/A	10
8	A Resistivity	243885.7	-13.22	-121939.8	15.22	0.001	0.00007	11
10	Aspect North	243900.1	1.22	-121947.1	0.77	0.378	0.01256	12
9	Aspect East	243900.8	1.94	-121947.4	0.06	0.821	-0.00322	13

Table 54 Model coefficients obtained from Model 4 for Soluble ground conditions for the clay pipe waste water network

class	coefficient	Ground Classification
A	-15.79007	Soluble rocks not thought to be present
B	-16.09131	Soluble rocks are present but unlikely to cause problems
C	-16.10745	Significant Soluble rocks are present with low possibility of localised subsidence or dissolution related degradation of bedrock
D	-15.55886	Very significant soluble rocks are present with a moderate possibility of localised natural subsidence or dissolution related degradation of bedrock
E	-15.35128	Very significant soluble rocks are present with a high possibility of localised subsidence or dissolution of bedrock

Table 55 Model coefficients obtained from Model 4 for corrosive ground conditions for the clay pipe waste water network

class	coefficient	Ground Classification
class 1	-15.71212	Unlikely to cause corrosion
class 2	-16.38022	May cause corrosion
class 3	-16.37824	Likely to cause corrosion

3.7.6 Adding other environmental factors sequentially to the null model (Model 5)

The final stage for the waste water clay network is to add the significant variables from Table 53 sequentially to the Null model to produce a final model. Table 56 reports how each factor improves the model, with slope and C roads improving the model by the greatest extent. The final model output can be seen in Figure 23, which again shows that the model residuals are reduced compared to the EE model (Figure 21). Categorical coefficients are shown in Tables 57 - 60.

Table 56 P-value from sequential addition of statistically significant co-variables added to the null model (Model 5). LLr is the log likelihood ratio statistic expressing how many times more likely the data are based on addition of this covariate in comparison to the previous model.

	Model	pval	LLr	coef
1	Slope	0.001	265.18	0.07749
2	A Road	0.001	45.64	0.03933
3	B Road	0.001	26.90	0.02714
4	C Road	0.001	187.57	0.02177
5	Shrink swell clays	0.001	25.35	N/A
6	Compressible Ground	0.001	17.60	N/A
7	Solubility	0.001	11.62	N/A
8	Soil Corrosivity	0.001	22.51	N/A
9	Dwellings	0.001	96.382	-0.13615

Table 57 Coefficients of Shrink swell clays obtained from Model 5 for the clay waste water network.

class	coefficient	Ground Classification
A	-15.80660	Ground conditions predominantly non plastic; No action
B	-15.71318	Ground conditions predominantly low plastic; No action
C	-15.83429	Medium plasticity; action required
D	-16.40793	High Plasticity

Table 58 Coefficients of Compressible Ground obtained from Model 5 for the clay waste water network.

class	coefficient	Ground Classification
A	-15.80660	No indicators of compressible ground – No action
B	-16.49140	Very slight potential of compressible deposits
C	-16.91715	Slight possibility of compressibility problems
D	-15.71720	Significant potential for compressibility problems

Table 59 Coefficients of Soluble Ground obtained from Model 5 for the clay waste water network

class	coefficient	Ground Classification
A	-15.80660	Soluble rocks not thought to be present
B	-15.82645	Soluble rocks are present but unlikely to cause problems
C	-15.88758	Significant Soluble rocks are present with low possibility of localised subsidence or dissolution related degradation of bedrock
D	-15.49445	Very significant soluble rocks are present with a moderate possibility of localised natural subsidence or dissolution related degradation of bedrock
E	-15.26674	Very significant soluble rocks are present with a high possibility of localised subsidence or dissolution of bedrock

Table 60 Coefficients of soil Corrosivity obtained from Model 5 model for the clay waste water network.

class	coefficient	Ground Classification
class 1	-15.80660	Unlikely to cause corrosion
class 2	-16.29156	May cause corrosion
class 3	-16.23575	Likely to cause corrosion

Figure 20 Result of full model for the YW region using sequential addition of covariates for the clay waste water network (Model 5). Note the decrease in the cumulative sum of raw residuals compared to the null model in Figure 9. Examination of the combined X and Y axis residuals suggest that overall the model is under predicting the number of pipe failures per unit length of pipe, with the red colours indicating where this is happening to the greatest extent and the blue the least.
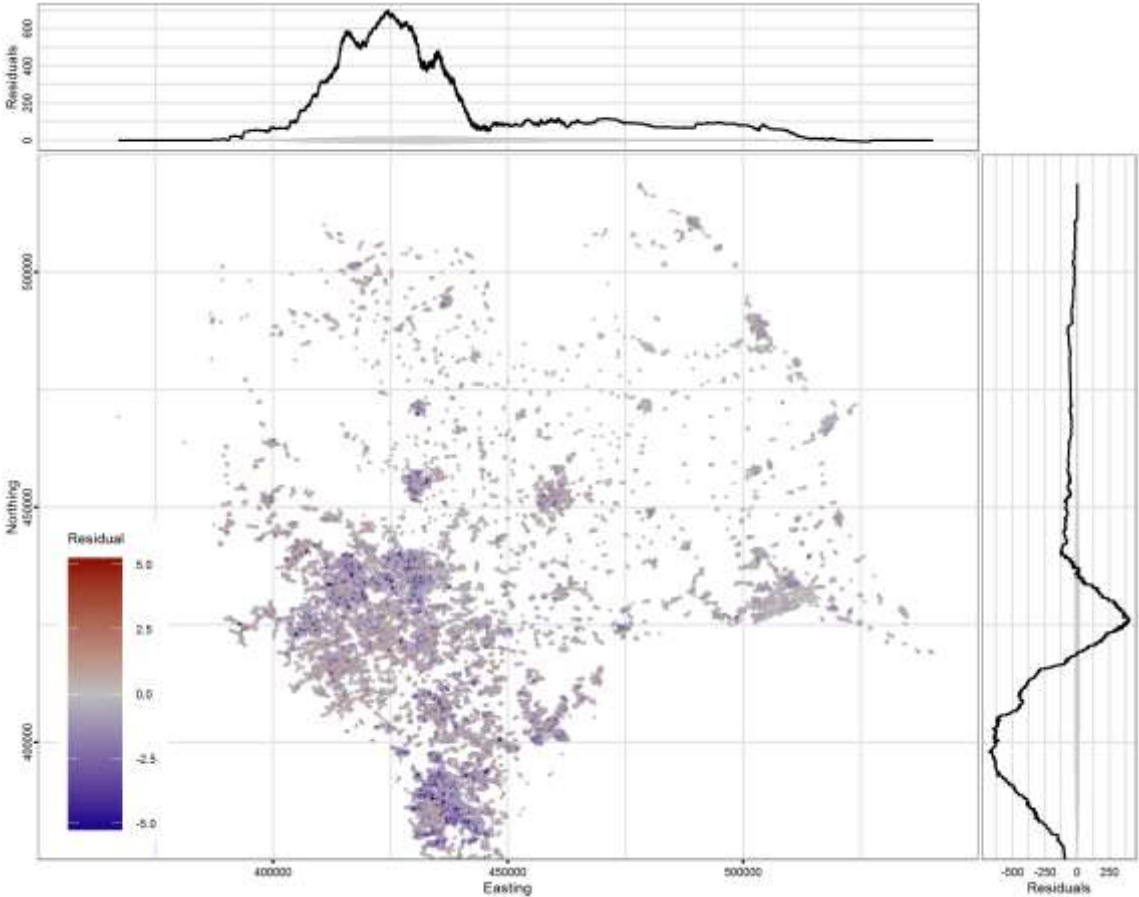
3.7.7 Discussion of waste water Clay network

Improvements to the Null model were obtained with the addition of co-variables from the Expert Elicitation and the later inclusion of other environmental co-variables, as demonstrated in the reduction of the Total raw residuals presented in the lurking variable plots (Figures 21-23). Interpretation of the lurking variable plot suggests that there may be an area of model under - prediction in the Leeds – Bradford area, possibly associated with subsidence from the coal measures. An area of over- prediction also appears to be associated with the Sheffield urban area, which is harder to suggest possible reasons for as one or more of the coefficients may be over-estimating a response. The possible influence of significant covariates added individually to the Null model on the pipe network are explained in Table 61.

Table 61 Possible explanations for the nature of model coefficients for the waste water clay network where single covariables are added to the Null model.

Rank	Covariate	+/- coefficient	Notes
1	Slope	+	A positive correlation between expected pipe failure and slope suggests that the weight of waste in the pipe on slopes may cause greater failure
2	Number of Dwellings	+	Positive correlation between number of dwellings and pipe failure suggests that increased use of the pipe network has a detrimental effect.
3	C Road	+	A positive correlation suggesting some interactions with traffic volume and vibration
4	Corrosivity	N/A	The coefficients for Class 2 and Class 3 are similar, whilst both are higher than Class 1 where the soils are not thought to be corrosive. With resistivity (Clay) being such a dominant part of the CIPRA corriosion classification, these results suggest that the corrosivity index is possibly identifying soils with clay contents that promote good stability and soil – structure - pipe interactions.
5	CTI	-	A negative correlation exists between CTI and pipe failure. A possible explanation is that the soils with low CTI may have greater variations in their thermal and moisture regimes, potentially leading to greater differential ground movement.
6	A Road	+	A positive correlation suggesting some interactions with traffic volume and vibration
7	Compressible	N/A	The coefficients for Class B and Class C are similar and are both lower than Class 1 where the soils are not thought to be susceptible to ground movements caused by compressible deposits. This suggests that soils in the compressible classes are defining their stability and soil – structure - pipe interactions. No pipeline was present in Class E
8	Shrink Swell Clay	N/A	Coefficients for classes A and B are relatively similar, with class C and D having lower coefficients. This could suggest greater stability and improved soil – structure - pipe interactions in class C and D because of the presence of clay or that there is a self-sealing occurring if pipes do break. There is no Class E in the YW region so the co-variable is not fully tested.
9	B Road	+	A positive correlation suggesting some interactions with traffic volume and vibration
10	Solubility	N/A	Class D & E had the highest coefficients suggesting that the soluble rocks they were identifying (e.g. gypsum bearing rocks near Ripon) had an influence on increasing pipe failure, possibly through subsidence. Class B and C had the lowest coefficients and these areas are more related to chalk and limestone suggest that these can provide greater stability to the pipe network.
11	A resistivity	+	The positive correlation suggests that there was increased failure with higher resistivity. Higher resistivity is found in soils with lower clay contents, which confirms the suggestion from other covariables such as shrink swell and compressible deposits that soils that promote better stability and soil – structure - pipe interactions are being identified as having lower failure rates.

3.7.8   Individual heat maps

Individual heat maps for the significant variables from Model 5 for the clay network are presented in Figure 24. Based on the pipe failures recorded, these heat maps show spatially where the individual co-variables may present a danger to the clay pipe network across the YW area. Whilst the road networks are generally a pale yellow, the greatest areas of yellow and red

can be seen in the upland areas of the Yorkshire Dales, Peak District and the North York Moors, where the slopes are greatest. The effects of solubility (dissolution) are greatest around Ripon.

Figure 21 Individual heat maps for the significant co-variables from Model 5 for the clay waste water network placed on a standardized scale
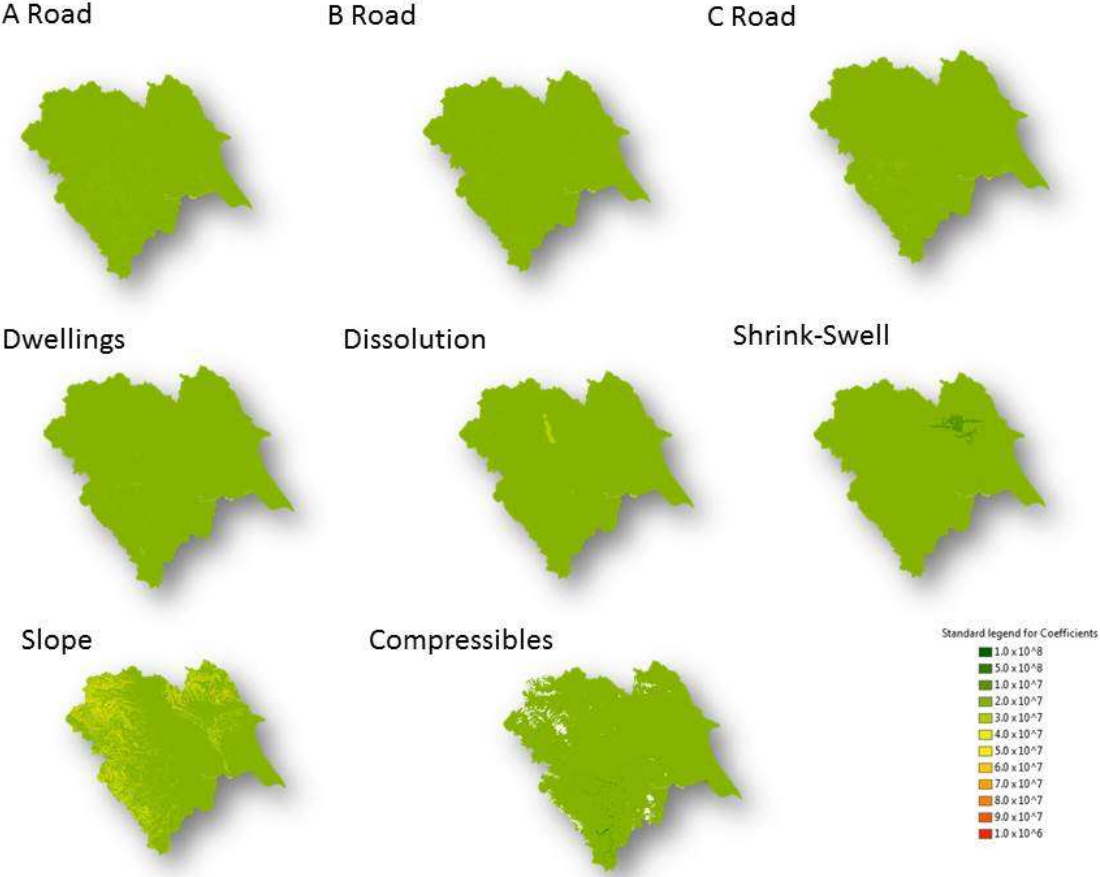
3.7.9 Total Coefficient heat map

The Total intensity heat map is formed by combining the coefficients from the significant co-variables from Model 5 are shown in Figure 25. This shows clearly the areas which exist in areas of greatest hostility to the pipe network across the YW region. As well as picking out the potential for slope to cause failure, the urban areas are also highlighted.

Figure 22 Total Intensity heat map for the clay waste water network obtained by combining significant co-variable coefficients from Model 5.

Total Intensity

Total_intensity_map_sequential
High : 12.7542

Low : 0.190778
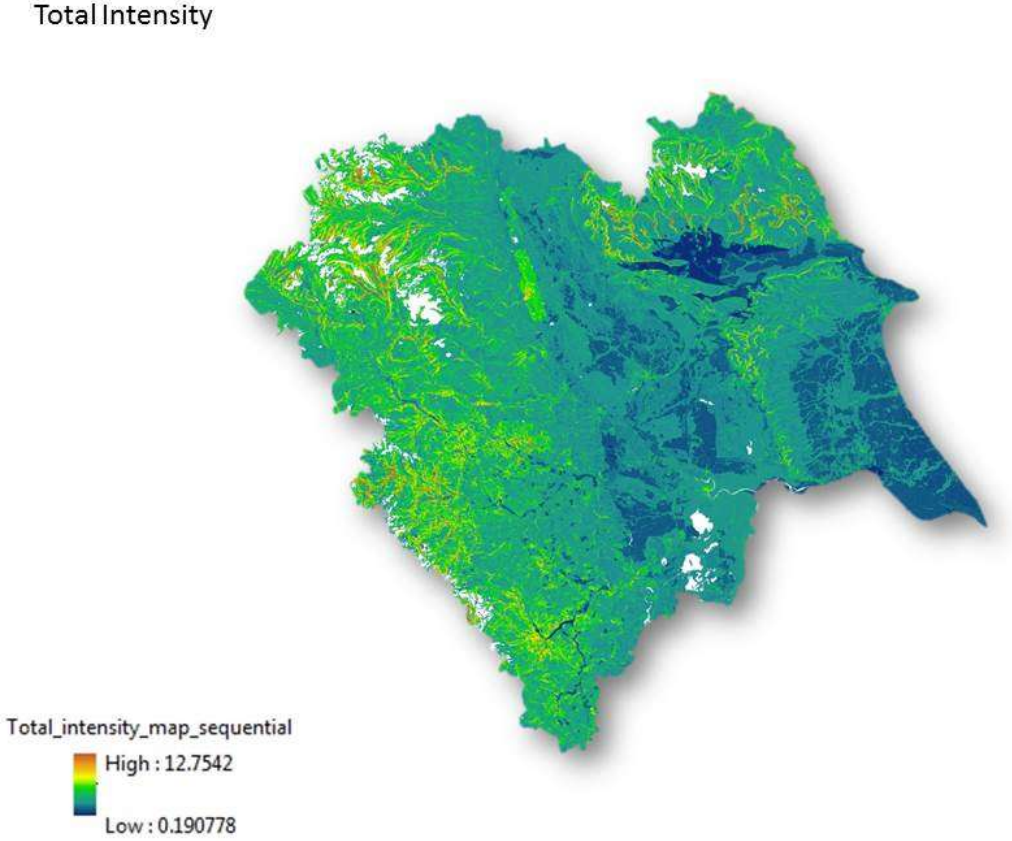
4 General Discussion

4.1 THE VALUE OF THE MODEL OUTPUTS

The major aim of the project was to assess whether incorporating geological and environmental factors into models of pipe failure, water companies could develop greater understanding of their pipe networks. This may enable them to consider ways through which greater resilience can be built in, particularly with respect to a changing climate and increasing population. Typically water companies assess the current condition of their pipe assets by looking at age and increasingly internal camera assessments. Our approach is complementary and looks spatially at the distribution of pipe failure with respect to the density of pipe, and links these to geological,

topographical and environmental factors. Thus the model produces spatial information of where these factors may have the greatest impacts on the pipe network. This is achieved through:

1. Interpretation of lurking variable plots allows an assessment of areas of the YW region where the model underperforms, allowing exploration of possible other factors that are causing damage to the network. Where under-prediction coincides on both the X and Y axes, reasons can be more easily identified. For example, in this work the models for different pipe materials consistently under-predicted pipe failure in areas associated with the coal measures and for cast iron, one area associated with lacustrine deposits. There are valid reasons (subsidence, poor load bearing strength) why these areas may have been identified.

2. The identification of significant model co-variables allows us to understand those factors that are having an effect on the network. Whereas the continuous variables produce a ± coefficient, greater interpretation is required for the categorical variables as it was not always a linear response.

3. A combined heat map can be produced by combining all the coefficients for each 100 x 100 m cell to show where the pipe network is at greatest risk.

4. The coefficients from the significant model co-variables from the final sequential model can be used in the production of individual heat maps which can help explain the factors contributing to the combined heat map. Thus by combining the coefficient with the categorical class or the continuous variable number for each 100 x 100 m cell, the areas in which individual covariates could impact the pipe network can be assessed spatially across the YW region. In addition, by taking the highest and lowest coefficient from all the covariates and standardising the colour scheme we can also compare the impacts of the covariates on the pipe network.

4.2 WHAT WE HAVE LEARNT

Using the approaches outlined above, the following are the key points from analyses of the 4 pipe networks for YW:

1. For the YW region, non-geological factors generally had the greatest impact on pipe network failure including factors associated with road networks, water source and the number of dwellings.

2. The co-variables identified in the expert elicitation were usually found to be significant at P<0.05, demonstrating that YW had good knowledge regarding reasons for failure within their pipe network. The inclusion of factors identified through the Expert Elicitation exercise always improved the Null model. However the inclusion of further environmental and geohazard factors (e.g. dwellings, Sulphide / sulphate) resulted in improvements to the Expert Elicitation models.

3. Where geological and topographic factors were important these included slope for the concrete and clay networks which with gravity and weight obviously produces stress on the network. Solubility was important for the concrete pipe, suggesting subsidence in the Rippon area was a major source of failure. Sulphate and sulphide was important for cast iron, identifying partly the coal measures.

4. Some geological units appeared to cause problems for the pipe network beyond those accounted for in the list of co-variables. In particular these included the coal measures where subsidence may occur and one area of the lacustrine clays associated with the Glacial Lake Humber. Lacustrine clays typically have poor loading capacity.

5. The continuous variables were relatively easy to interpret as to their role in pipe network failure, whereas the use of the geohazard categorical variables did not always provide linear responses.

6. Shrink swell and compressible ground are the two geohazards often cited as having major impacts on pipe network failure and both were found to be significant (P<0.05), but the categorical coefficients obtained were non-linear. In addition, both datasets were not tested sufficiently. The YW region did not have a Class 5 region (Highly plastic soils) for shrink swell clays so no estimation of the most extreme shrink-swell clay soils could be made. For compressible ground conditions where pipe networks did pass through the highest class of risk, no known failures were found, which resulted in very low coefficient values. The interpretation of the categorical variables was therefore difficult. However, we suggest that the non-linearity of the coefficients obtained for the classes within these datasets may indicate broad ranges of different soil types and their specific properties which determine the settlement and deflection of different pipe materials in the soil. The amount of clay in the soil and it's type are fundamental to shrink swell and compressibility but are also fundamental to processes that enable support to the pipeline. This needs to be examined further.

7. Some of the geohazard datasets needed interpretation because of the way they were created (e.g solubility, sulphide / sulphate). For example the solubility dataset could be split up into soft rocks that may dissolve (e.g. chalk and limestone) and those that may

have soluble horizons causing subsidence (e.g gypsum containing rocks). For the sulphate and sulphide they could be split again. Individual datasets could be more appropriate and easier to use in some circumstances.

4.3 REVIEW OF WORK WITH YORKSHIRE WATER, SCOTTISH WATER AND WELSH WATER

Presentations of results were made to Yorkshire Water during the course of these two grants. The meetings are reviewed here.

4.3.1 Meeting with Yorkshire Water – 22nd June 2015

A meeting was held with representatives of YW on June 22nd 2015 to gain feedback from the initial model results. Discussions focused on the possible causes of pipe failures caused by factors that were not included within the model and this particularly applied to the under prediction of the model in the SW of the YW region (Leeds-Bradford). These included

- Surge demand
- Water pressure changes
- Water temperature & temperature change
- Source of water
- Drainage
- Climate

The area of model under prediction (Leeds –Bradford) is the largest urban area and this is likely to be where surge demand will most regularly occur. In addition, it is also a hilly region within the YW region and this may also cause greater changes in water pressure within the pipe network to occur. Both these factors are recognised as causing the blow out of pinhole corrosion to occur in cast iron pipes. Future data used in the model could include calculated 'change of slope' within a 100 x 100 m cell as well as mean slope to consider these pressure changes. A further possibility for the Leeds –Bradford area is that the source of water is different to the rest of the YW region (DWI, 2014). It was suggested that Leeds-Bradford may be served by reservoirs whilst much of the rest of the region by groundwater, and YW confirmed. These different water sources will have different temperatures and hydro-chemical variations. Ground water temperatures should be constant, whereas reservoir water will vary depending on the season and weather. In addition the different chemistries (e.g. pH, SO_4^{2-}, Cl^-) of the water may have an effect on the internal corrosion of pipes. Data on temperature and pressure is held by

YW for the Distribution Management Areas (DMA). There are 2300 DMA's each serving between 800 and 900 properties and these could be included within the model.

The major covariate in the models for pipe failure was linked to 'C-roads'. YW cited the following as possible contributory factors. The major weaknesses within a pipe network are the join between lengths of pipes. The pipe network associated with C roads is usually dominated by a greater frequency of connections between the 'water main' and the domestic pipe. In addition, smaller diameter and thinner pipes may be used in much of this part of the network. Within the C-roads we suggested that poor drainage in the sub grade may encourage anaerobic conditions associated with ponding of water can lead to corrosion. YW state that the infill of trenches is generally limestone gravel from about 1970 onwards. However, no comments were made about drainage. This suggests that in future modelling different drainage factors such as change of slope angle and drainage x geology may be appropriate. Different data and information sets were discussed. BGS could use derived data from NextMap to calculate the change of slope and the SUDS dataset for drainage get away.

Climate features were also considered. Cast iron and ductile pipes have peak bursts during the winter – December to February relating to low temperatures. Plastic pipes tend to fail more often during the summer. If the failure data is dated then an assessment of climate on bursts could be done within the model. Soil Moisture Deficit (SMD) was also mentioned and this should be feasible.

The causes of failure associated with the plastic pipe network appear to be related to vibration, slope and clay. All three road types were identified, with C roads > B roads > A roads in the ranking, suggesting that the larger better constructed roads have a lower effect. Slope may play an effect through gravity distorting the pipes. Interestingly, clay was identified as a proxy of A-resistivity. The hardness of and rigidity of dry clay or its contributions towards ground movements may be significant. Hardness may cause chaffing of the pipeline with vibrations. Interestingly there was a negative correlation with compressible ground suggesting that a pipe in slightly giving material may be slightly protected. This may also be why there was a negative correlation between plastic pipe failure and shrink-swell clays.

4.3.2 Meeting with Yorkshire Water – 14th July 2016

A further meeting was held with YW where improvements in the cast iron and plastic models created in Grant NE/NO13026/1 were presented as well as results from the waste water clay and concrete models. Discussions were had regarding explanations of results. Main points regarding results of pipe networks were:

1. C-roads more likely to be in road whilst A- and B-roads are in the pavement if possible. In addition because of the type of road there will be differences in size and number of connections. In C-roads it's likely to be a 'distribution' network whilst larger road we have the principal mains.
2. Unlikely for pipes put in shrink swell soils to be differently engineered – Current YW models do not have this differentiation in their models built a basic soil type factor
3. Extra protection would be given to pipes if peat is present
4. Pipe depth is a factor that hasn't been included within the model because of access to suitable data. Standard depth for cost and temperature. More variability for concrete waste water. There is scope for shallower depth with plastic pipes. It will help to reduce carbon footprint (digging) being able to reduce depth to which pipes are laid. Waste pipes often installed using micro-tunnelling technologies these days rather than open cut.

4.3.2   Meeting with Scottish Water – 27th June 2016

A lot of decisions are currently made according to beliefs rather than evidence. There is a drive in Scottish Water to become more data driven / evidence based. They are aware of a correlation between pressure and pipe failure – e.g. 10% reduction in pressure results in 14% reduction in failure rate (but, it may just prolong inevitable failure by corrosion). In their efforts to reduce pipe failure, 50 % of the reduction they have been able to achieve has been as a result of pressure management. The rest is mains rehab and operations management. They do not currently engineer to account for geological conditions such as shrink swell. Pipe systems are 'off the shelf', not specific to ground conditions.

4.3.3   Expert elicitation

An expert elicitation exercise was undertaken with Scottish Water so that discussion of their results would not bias their opinions on what they considered were the principal reasons for failure within their pipelines. Compared to YW, Scottish Water there was more slightly more focus on weather effects (Scotland generally having longer and colder winters). These are the results of the Expert Elicitation process.

Cast Iron:

1. Age (exacerbates all other factors – an interaction effect?)
2. Pressure (pressure transience rather than constant pressure)
3. Ground temperature
4. Weather (seasonality, cold and wet winter vs drying out of ground in summer).
5. Ground heave (failure by ring splits, ring fractures)
6. Water source (ground, surface, chemical treatment – internal corrosion > external corrosion)
7. Road vibration (construction sites, building and piling – in theory 600mm of cover makes this negligible but they are suspicious)
8. Contaminated ground (e.g. High rate of corrosion/pitting at Innerleithen due to copper contamination?).

Plastic:

1. Installation error
2. Pressure transience
3. Presence of hydrocarbons possibly (e.g. peat)

Asbestos Cement:

1. External factors (e.g. pH, water chemistry, soil types).
2. Mechanical joint failure – installation problems or corrosion of nuts and bolts (see cast iron).

Pre-stressed Concrete:

1. Catastrophic join failure

Clay pipes:

1. Root infiltration
2. Ground distortion / disturbance (rubber seals will pop out).

General additional hazards:

1. Ground water infiltration may be an issue – they mentioned mining areas and red ochre – acid mine drainage?
2. Peat - pipe buoyancy and mobility, and hydrocarbons.
3. Running sands an issue for sewers.
4. Mine collapse an issue but rare.

4.3.4 Suggestions from Scottish Water during BGS presentation:

1. They were interested if we had included a flooding layer as a predictor variable in our models, with reference being to ground water infiltration of sewer system.
2. Saline infiltration was an issue on the East Coast, where rising mains are metallic.
3. They were interested our use of Number of dwellings per cell and mentioned work that they had done looking at social demographic /class and sewer blockages.

4. C road influence may be due to construction activity and third party damage. Possible pipe failures as a result of contrasting ground conditions between made ground under the road and natural ground beyond.
5. We could perhaps use CEH's land use classification as a predictor (but land use may effectively already be explained in what we have used).
6. With regard to dwellings – is it possible that our burst data includes bursts in minor house-feeding pipes, which have been mistakenly appended to the mains?
7. It was suggested that the model could be validated by splitting the data into blocks of different age and comparing the resultant models.

4.3.5   Meeting with Welsh Water – 22nd July 2016

An expert elicitation process was carried out with the main comments for cast iron being similar to those from Scottish Water and Yorkshire Water. Again weather, particularly the autumn period when a greater number of failures are reported was mentioned.

Cast Iron

For the cast iron network the results are shown below

1. Climate – winter freezing
2. Soil moisture deficit
3. Corrosive soils
4. Age
5. Pressure trabnsience
6. Joints

References

British Geological Survey holds most of the references listed below, and copies may be obtained via the library service subject to copyright legislation (contact libuser@bgs.ac.uk for details). The library catalogue is available at: http://geolib.bgs.ac.uk.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle, in: Second International Symposium on Information Theory, edited by: Petov, B. N. and Csaki, F., Akademia Kiado, Budapest, 267–281.

Aldiss, Diaz Doce & Northmore. 2014. GeoSure Version 7 methodology: Collapsible Deposits. *British Geological Survey Internal Report IR/14/016*. British Geological Survey, Keyworth, Nottingham, UK. 31pp.

Baddeley, A., Turner, R. 2006. Modelling spatial point patterns in R. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, D. Stoyan (Eds). *Case Studies in Spatial Point Process Modeling*, Lecture Notes in Statistics Volume 185, pp 23–74

Booth, K A, Diaz Doce, D, Harrison, M & Wildman G (editors). 2010. User guide for the British Geological Survey GeoSure dataset. *British Geological Survey Open report OR/10/006.* British Geological Survey, Keyworth, Nottingham, UK. http://nora.nerc.ac.uk/13840/

Cox, D.R., Hinkley, D.V. 1990. Theoretical Statistics. Chapman & Hall, London.

Diaz Doce. D, Jones, L D & Lee K A. 2015. GeoSure Version 7 Methodology: Shrink Swell. British Geological Survey Internal report, IR/14/017. British Geological Survey, Keyworth, Nottingham, UK. 28pp.

Dashwood, C, Diaz Doce, D & Lee, K A. 2014. Geosure Version 7 methodology@ Landslides (Slope Instabiity). British Geological Survey Internal report, IR/14/014. British Geological Survey, Keyworth, Nottingham, UK. 36pp.

Dearden, R.; Tye, A.M., Marchant, A. 2014 User guide for the Corroded Asset Failure: Ferrous map. Nottingham, UK, British Geological Survey. (OR/14/001).

Dearden, R.; Tye, A.; Marchant, A. 2014 User guide for the Pipe Leakage Impacts map. Nottingham, UK, British Geological Survey, 19pp. (OR/14/010).

Diggle, P.J. 2013. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns (third edition) Boca Raton:Chapman and Hall/CRC Press.

Entwisle, D C, Lawley, R S and Lee K A. 2012. User Guide DiGMapPlus+ Engineering Properties: Use for Engineered Fill dataset (version 1). British Geological Survey Open Report, OR/12/091. 18pp.

Entwisle, D C, White, J C, Busby, J P, Lawley, R S & Cooke, I L. 2014. Electrical resistivity model of Great Britain: User Guide. *British Geological Survey Open report, OR/14/030*. British Geological Survey, Keyworth, Nottingham, UK. 15pp.

Farrant, A R, Cooper, A H & Diaz Doce. 2015. GeoSure Version 7 methodology: Soluble rocks (Dissolution). *British Geological Survey Open report, OR/14/012*. British Geological Survey, Keyworth, Nottingham, UK. 63pp.

Jones, L D & Terrington, R. 2011. Modelling volume change potential in the London Clay. *Quaternary Journal of Engineering Geology and Hydrogeology*, 44, 109-122. http://nora.nerc.ac.uk/13629/

Jones, L D, Diaz Doce, Lee, K A & Entwisle, D C 2015. GeoSure Version 7 Methodology: Compressible Ground. *British Geological Survey Internal report IR/14/015*. British Geological Survey, Keyworth, Nottingham, UK. 30pp.

Lee, K A and Diaz Doce. 2014. User guide for the British Geological Survey GeoSure dataset. *British Geological Survey Open report OR/14/012*. British Geological Survey, Keyworth, Nottingham, UK. 17pp.

Lawley, R. 2011. The Soil Parent Material Database: A User Guide. British Geological Survey Open Report, OR/08/034. 53pp

Marino, G.G. 2000. Pipelines exposed to coal mine subsidence face risk of serious damage. Pipeline and Gas Journal, 227 (11), 37.

Olliff, J.L., Rolfe, S.J., Wijeyesekera, D.C. & Reginold, J.T. 2001. Soil-Structure-pipe interaction with particular reference to ground movement induced failures. Proceedings of Plastic Pipes XI, Munich, 3td-6th September 2001.