**Leakage acoustic data**

**Summary information**

This document contains a summary of the data contained in the spreadsheet, with further details of the exact column contents in the second half of the document.

* **Acoustic logger data (columns A to J)**

“Gutermann Zonescan 820” acoustic loggers are deployed by Leakage Technicians in water distribution zones with potential leakage.  Loggers are collected in the morning and deployed in a different location in the afternoon (known as “Lift and Shift”).  Once a logger is collected, the data is sent to the Gutermann Cloud service, and then downloaded to our systems using an API web request. Broadly, this includes data on when the logger is deployed, collected and makes a measurement, a leak score calculated from the acoustic file and other acoustic information.

* **Data collated as part of current acoustic leak detection process (columns K to N)**

Leakage Analysts use the Gutermann data to assess the likelihood of a leak using a manual analytical process, generating a spreadsheet listing possible leaks called points of interest (POIs). POIs located *between* multiple loggers that are detecting the same noise at the same time are called “correlations” and are given a higher weighting, as there is increased likelihood of the leak noise being at that location. Note that POIs will be located *at* the logger location, while correlations are located *between* multiple loggers i.e. not at the same location as a logger. The data produced during this analysis is found in columns K to L. For each POI, the analyst would also provide its location and a mapping system picture showing the surroundings (see next bullet point). Each POI is then followed up by a Leakage Inspector. Feedback is sent back to the analyst to record success and failure on the spreadsheet against the original POI – this is found in column M, with the Leakage Inspector ID in column N.

* **Asset information and Geographic Information System (GIS) data associated with the logger’s location (columns O to AH)**

The Gutermann download would ordinarily include the logger’s latitude and longitude.

The location is entered manually into a spreadsheet for further analysis using GIS systems for information such as mains type, size, age etc. – this is published in columns O to X.

For security reasons, we have not published the grid reference locations, which would allow extrapolation of the pipe location and a network map. Instead we have replaced the grid references with distances from relevant environmental factors such as rivers, railways, different road classes, control valves etc. (columns Y to AH). This has been collected by using the logger location to interrogate our GIS systems.

* **Information about jobs raised nearby with the leakage team (columns AI to AS)**

For any leakage job, as BAU we enter the details (called “job card data”) onto our work management system, and they are then imported into our GIS system. The logger’s location is used to interrogate our GIS system, to provide the date, reference number and description for any nearby leakage jobs, occurring within 30 days of logger recording.  Columns AI to AS contain this information categorised as jobs within 5m, 20m and 100m from the logger, followed by the number of days after collection that the geographically closest, soonest leakage job was raised, and the distance between the logger and that job.

* **Leakage savings, where they could be estimated (columns AT to AU)**

This is manually calculated from the information held for a leakage job, in a leakage reporting software called WATERNET. The calculation compares a 3-day average of estimated leakage before and after the leak has been repaired – the metric used for this is “minimum night flow”, which is the lowest recorded daily flow through a meter, usually between 3 and 6 am. The values are imported to the original POI spreadsheet. In the published dataset, this information is found in columns AT to AU.

We have extracted the data from the systems described above and compiled it into this single data set for your ease.

As a guide, although all of this data is available for building your solution, only columns B to J and O to AH would be available as input to the model – for clarity these are shaded blue in the published dataset.

**Leakage acoustic data – column descriptions**

Columns A to J – Gutermann logger data

|  |  |  |
| --- | --- | --- |
| **Col.** | **Column Heading** | **Description** |
| A | Lookup | The unique identifier for a specific recording by a specific logger |
| B | collect\_DA | The date that the acoustic logger was collected |
| C | deploy\_DAT | The date that the acoustic logger was deployed |
| D | Measure\_Da | The date that the acoustic logger made its first measurement |
| E | logger | The serial number of the acoustic logger that made the recording |
| F | leakScore | This is the leak score that the logger recorded on the date in column D ("Measure\_Da"). The leak score is an algorithm calculated from the acoustic file recorded by the logger. |
| G | dbMin | The minimum noise that the logger recorded on the date in column D ("Measure\_Da"). |
| H | Spread | This is the difference between the minimum recorded noise and the most frequently occurring noise and is a measure of consistency. The larger the spread, the less consistent the recorded noise. If the spread is too small the noise is too consistent to be a leak. |
| I | projectNam | The ID for the Leakage Technician that deployed the logger |
| J | measTime | The time at which the acoustic logger made its first measurement |

Columns K to N – Data collated as part of current acoustic leak detection process

|  |  |  |
| --- | --- | --- |
| **Col.** | **Column Heading** | **Description** |
| K | Validated | This shows whether the measurement date matches the date deployed and collected; if the dates are not valid the job data should not be trusted. |
| L | Type\_of\_PO | This is a description of the type of POI, whether it is:   * a point identified by a single logger – this is called a "POI" * or a point between multiple loggers identified by cross-referencing their acoustic data – this is called a "correlation"   NB: POIs will be located *at* the logger location, while correlations are located *between* multiple loggers i.e. not at the same location as a logger. The job distance (column AS) measures from the logger to the leakage job, *not* from the point of correlation to the leakage job. Hence the job distance may appear greater for a correlation than for a POI. |
| M | Result | A description of the result based on a Leakage Inspector’s interpretation of their visit. A Leakage Inspector may not report correctly on every instance, and it may depend on the competency of the inspector i.e. this is an opinion-based field. |
| N | Leakage\_Inspector | The Leakage Inspector who carried out the work on the POI |

Columns O to AH – Asset information and GIS-related calculations

|  |  |  |
| --- | --- | --- |
| **Col.** | **Column Heading** | **Description** |
| O | pipeType | The classification of the pipe at which the logger was deployed |
| P | Diameter | The diameter of the pipe at which the logger was deployed |
| Q | Units | The unit of the diameter |
| R | Date\_Laid | The date that the pipe was known to be laid |
| S | Age\_of\_Pipe | The age of the pipe |
| T | Material | The material that the pipe is made of |
| U | Bedrock | The type of bedrock in which the pipe is laid |
| V | WIS Zone | The zone on our supply network in which the logger was deployed |
| W | DMA\_Number | A number indicating the specific zone on our supply network in which the logger was deployed. DMA stands for District Metered Area and is a smaller unit than a WIS Zone i.e. a WIS zone supplies water to multiple DMAs. |
| X | DMA Ref | A reference indicating the specific zone on our supply network in which the logger was deployed |
| Y | River\_Distance(m) | The distance to the nearest river |
| Z | Stream\_Distance(m) | The distance to the nearest stream |
| AA | Railway\_Distance(m) | The distance to the nearest railway |
| AB | A\_Road\_Distance(m) | The distance to the nearest A-road |
| AC | B\_Road\_Distance(m) | The distance to the nearest B-Road |
| AD | Motorway\_Distance(m) | The distance to the nearest motorway |
| AE | Crossroad\_Distance(m) | The distance to the nearest crossroads |
| AF | Woodland\_Distance(m) | The distance to the nearest woodland/forest |
| AG | Footpath\_Distance(m) | The distance to the nearest footpath |
| AH | distance\_to\_control\_Valve | The distance in metres (down the network pipe) to the nearest control valve. A control valve creates a restriction in the network and therefore turbulent water; this in turn makes a noise similar to a leak noise.  *Note that this column includes pumps and boosters as control valves and sources of noise.* |

Columns AI to AS – Job card data

|  |  |  |
| --- | --- | --- |
| **Col.** | **Column Heading** | **Description** |
| AI | Jobs5m | If a job was raised within 5 metres of the logger location within 30 days of the recording, the date of that job |
| AJ | JobRef5m | If a job was raised within 5 metres of the logger location within 30 days of the recording, the reference of that job |
| AK | JobDesc5m | If a job was raised within 5 metres of the logger location within 30 days of the recording, the description of that job |
| AL | Jobs20m | If a job was raised within 20 metres of the logger location within 30 days of the recording, the date of that job |
| AM | JobRef20m | If a job was raised within 20 metres of the logger location within 30 days of the recording, the reference of that job |
| AN | JobDesc20m | If a job was raised within 20 metres of the logger location within 30 days of the recording, the description of that job |
| AO | Jobs100m | If a job was raised within 100 metres of the logger location within 30 days of the recording, the date of that job |
| AP | JobRef100m | If a job was raised within 100 metres of the logger location within 30 days of the recording, the reference of that job |
| AQ | JobDesc100m | If a job was raised within 100 metres of the logger location within 30 days of the recording, the description of that job |
| AR | Job raised (Days) after logger collected | The number of days after collection that the geographically closest, soonest, leakage job was raised.  This column also includes negative numbers, because jobs may be opened before the logger makes its recording, and repaired afterwards.  A single job number may be associated with several nearby loggers. |
| AS | j100mDist | For the geographically closest, soonest, leakage job that was raised, the distance between the logger and that job. |

Columns AT to AU – Leakage savings data

|  |  |  |
| --- | --- | --- |
| **Col.** | **Column Heading** | **Description** |
| AT | Leakage Saving | A leakage saving value (in m3/hr) that has been associated with the DMA (see column W) on the day of repair.  NB: If multiple jobs were repaired on the same day in the same DMA, the saving would appear against each of these multiple jobs (as the whole saving, not a proportion of it). |
| AU | Not a Leak Job | This indicates a job raised that was not a leak job, and was associated to the network in some other way. |