



Zoning: Technical Guidance Note 4

The Dividing Process

Prepared by: Roy Ferguson Consultancy

Date: May 2018



EUROPE & SCOTLAND
European Regional Development Fund
Investing in a Smart, Sustainable and Inclusive Future

Table of Contents

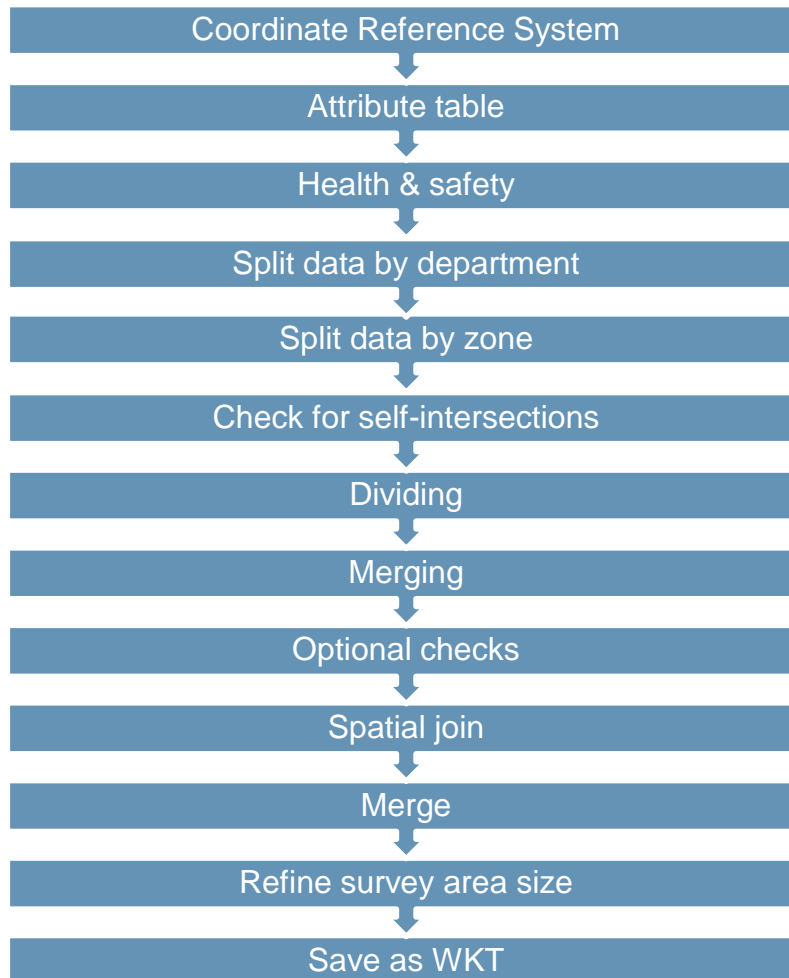
1.	Introduction	3
2.	The Process	3
3.	Preparing the Data	4
3.1.	Coordinate Reference System	4
3.2.	Attribute Table	4
3.3.	Health & Safety	4
3.4.	Split Data by Department	5
3.5.	Split Data By Zone	6
3.6.	Self-Intersections	7
4.	Dividing	9
4.1.	Dividing Simple Data	11
4.2.	Dividing A Network of Connected Polygons	11
4.3.	Dividing Layers Containing Complex Polygons	14
4.4.	Using the 'Polygon Divider'	20
5.	Getting Data into Final Form	21
5.1.	Merge Shapefiles	21
5.2.	Optional Checks	22
5.3.	Spatial Join	22
5.4.	Attribute Table	25
5.5.	Merge shapefiles	25
5.6.	Refining Survey Area Size	25
5.7.	Save as WKT	26

1. Introduction

This Guidance Note provides a step-by-step guide to preparing your zoned data for use with the 'Polygon Divider' software. This is followed by a description of how to use the 'Polygon Divider' and the steps required to get your data into its final form. Lastly, an overview of some of the main methods for identifying and resolving data quality issues is provided.

2. The Process

The diagram below provides an overview of the processes:



It is important that a clear filing and naming convention is followed, as a large number of shapefiles will be produced during the dividing process. It is also worthwhile keeping notes during the process of what has been divided, what is still to be divided and what files have been used when a merge has taken place. This reduces the chance of human error and makes it easier to go back and make changes should any issues be identified later in the process.

3. Preparing the Data

3.1. Coordinate Reference System

The coordinate reference system (CRS) should be set to OSGB 1936/British National Grid (EPSG:27700):






- Right click on the layer in the 'Map Legend' and click on 'Set Layer CRS'; and
- Type 27700 into the 'Filter' box and click 'OK'.

3.2. Attribute Table

The 'Attribute Table' should only contain the following fields:

- Org;
- Department;
- Zone;
- H_and_S; and
- Reason.


Any additional fields should be removed via the following process:

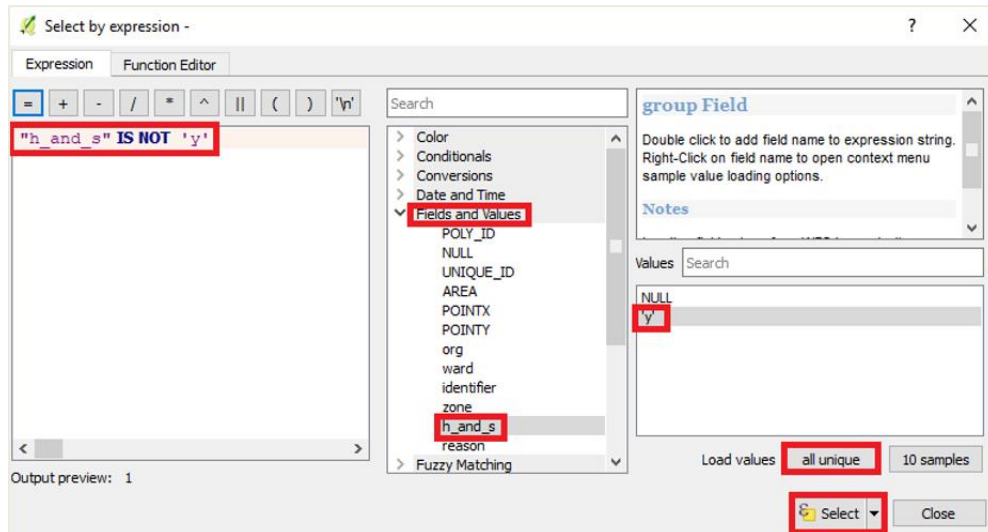
- Left click on the divided layer in the 'Map Legend'.
- Click on the 'Attribute Table' button  on the top toolbar.
- Make the layer editable .
- Click on the delete fields button .
- Select the layers you want to delete from the dialog box and click 'OK'.
 - It may take a few minutes if you are deleting a number of fields from a large database.
- Save the changes  and stop the layer from being editable .

Note: It is important that this step of the process is undertaken prior to the data being split by department and by zone. This is to ensure that the step is only undertaken once.

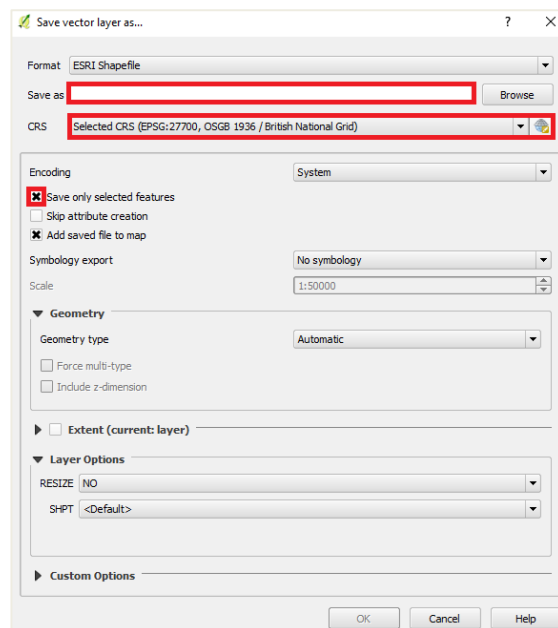
3.3. Health & Safety

Surveys will not be undertaken on land that have health and safety or access issues. Therefore, all areas identified with health and safety risks should be removed from the dataset. This is achieved via the following process:

- Left click on your divided layer in the 'Map Legend';
- Click on the 'Select features using an expression'  button in the top toolbar;
- In the dialog box, expand the 'Fields and Values' list and double click on 'h_and_s';
- Type IS NOT;
- Click on the 'all unique' button and double click on 'Y' in the box above;
- The dialog box should look like this:



- Click on the 'Select' button on the bottom right of the dialog box;
- Close the dialog box;
- Right click on your layer in the 'Map Legend' and select 'Save As...';
- Browse to where you want the new layer to be saved;
- Make sure the CRS is British National Grid;
- **Important:** Check the 'Save only selected features' checkbox;
- Click 'OK';



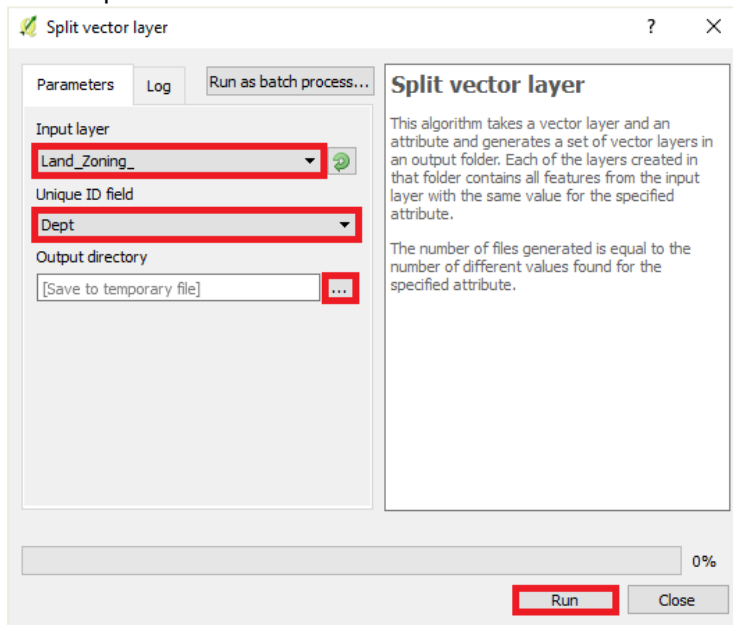
3.4. Split Data by Department

It is anticipated that surveying responsibilities will fall on the individual Local Authority department. Consequently, it is important that the data is separated into the various departments prior to being divided.

The process is as follows:

- Navigate to Vector > Data Management Tools > Split Vector Layer
 - 'Input layer' is the layer you want to split.

- 'Unique ID field' is the layer you want to split by (i.e. Department).
- Click on the three dots next to the 'Output directory' box to navigate to the folder you want to save the layers into. Please note the files will not be loaded to QGIS on completion.



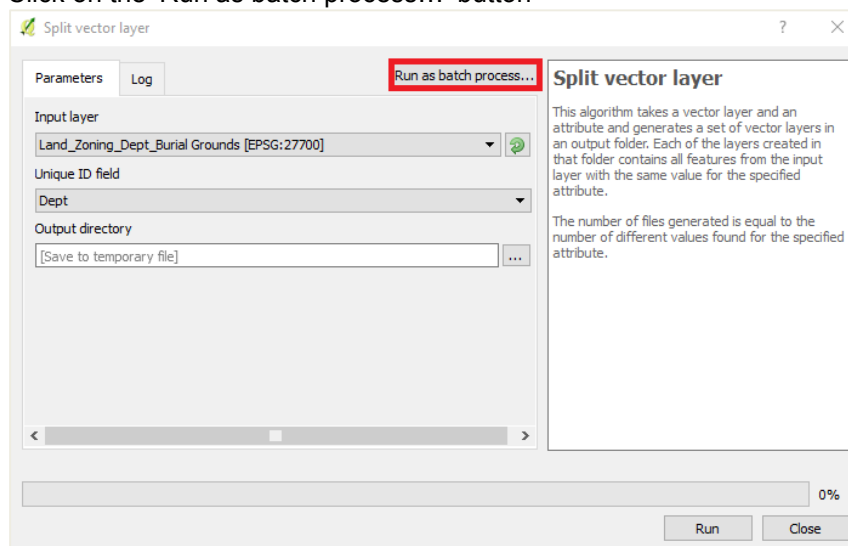
It is worth checking that no departments have been dropped during this process. This will occur if the field name in the original table had an asterisk at the end of it (which happens when QGIS shortens a field name because of its length).

3.5. Split Data By Zone

A survey area should not represent more than one 'Zone'. Consequently, each department should be split by zones prior to being divided.

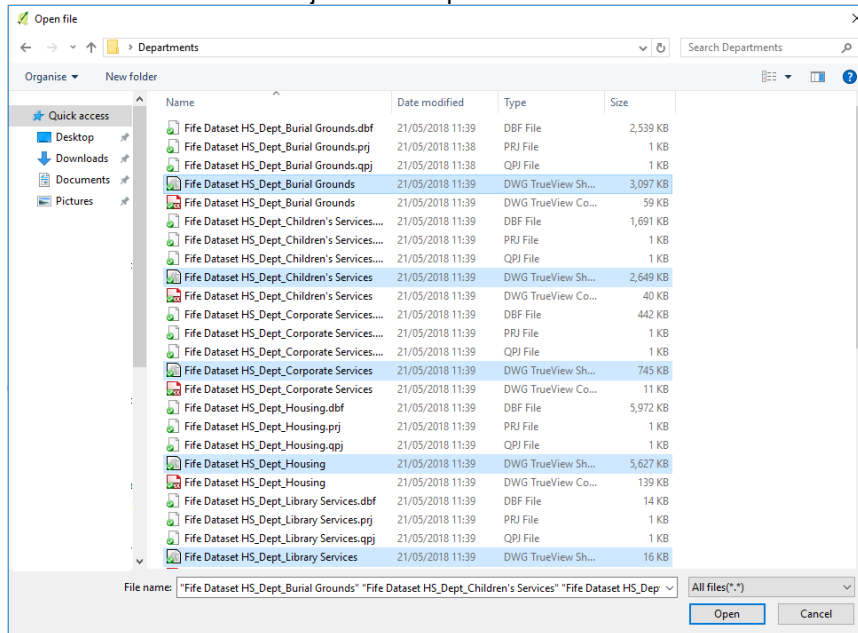
The process is as follows:

- Navigate to Vector > Data Management Tools > Split Vector Layer
 - Click on the 'Run as batch process...' button

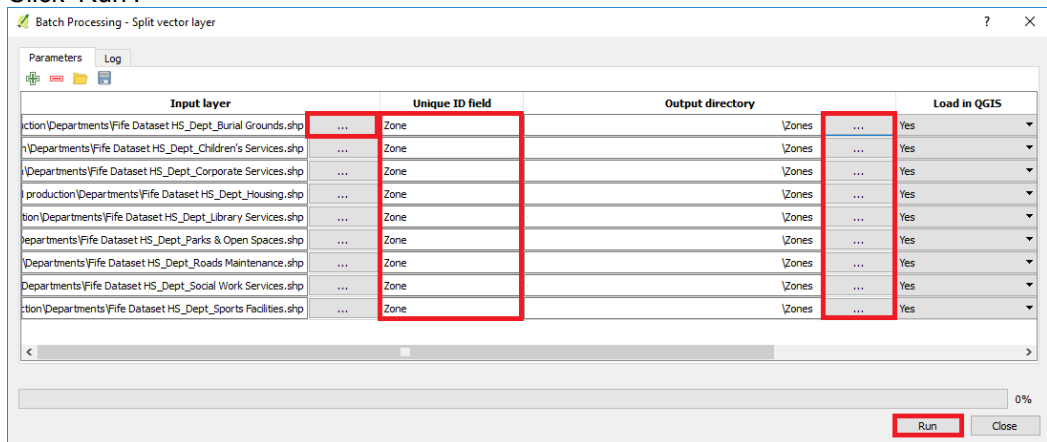


- Click on the three dots next to the top 'Input layer' box and click on 'Select from filesystem'.

- Navigate to where you saved your departments and select all of the shapefiles.
 - **Note:** Sorting by Type in the Explorer (i.e. clicking on the Type header) makes it easier to select just the shapefiles.



- Enter Zone as the 'Unique ID field'.
- Click on the three dots next to the 'Output directory' box to navigate to the folder you want to save the layers into.
- Click 'Run'.



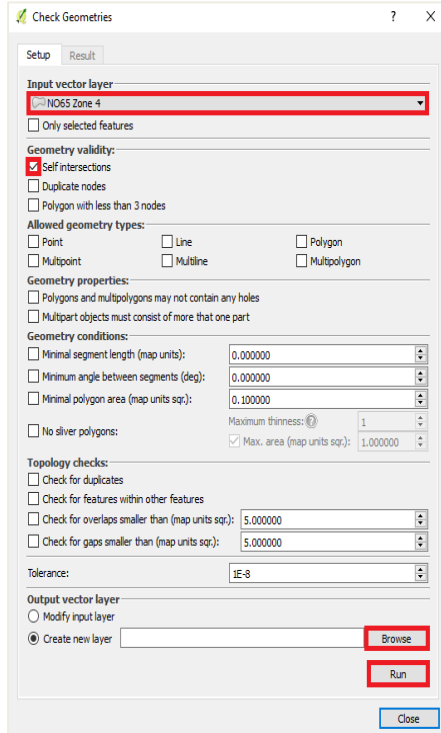
Note: This step assumes that QGIS 2.18.19 is being utilised. Batch processing is not available in some earlier versions of QGIS. If you are running an earlier version of QGIS you will need to either update to version 2.18.19 or split each department individually.

3.6. Self-Intersections

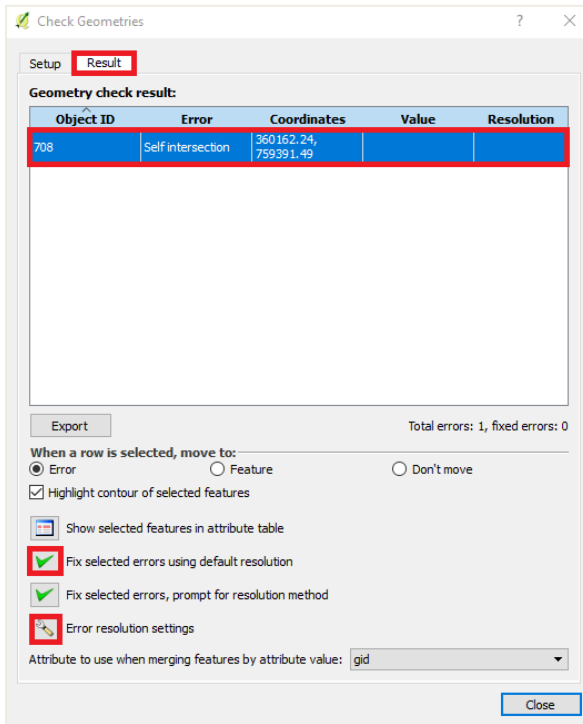
Self-intersecting polygons cause problems for various tools within QGIS. For example, the buffering tool will reach the self-intersection point and mistake it for the end of the polygon (i.e. the section at the other side of the self-intersection will not be buffered). With respect to this process, this would result in a loss of data, in other cases it will cause QGIS tools to fail. It is therefore very important that self-intersecting polygons are identified and fixed prior to undertaking the subsequent steps in the process.


The process for identifying and fixing self-intersecting polygons is as follows:

- Vector > Geometry Tools > Check Geometries.
 - **Note:** If the 'Geometry Checker' is not showing in the Vector pulldown menu, you will need to search for 'Geometry Checker' in the plugin manager (Plugins > Manage and Install Plugins...) and check the checkbox.
- Select the 'Input vector layer'.
- Check 'Self intersections'.
- 'Browse' to where you want the new layer to be saved.
- Click 'Run'.
- **Note:** This tool remembers the settings from previous runs. Therefore, you may have to deselect some of the other options if you have used the tool previously.



- Any self-intersecting polygons that are identified will be listed in the results tab:

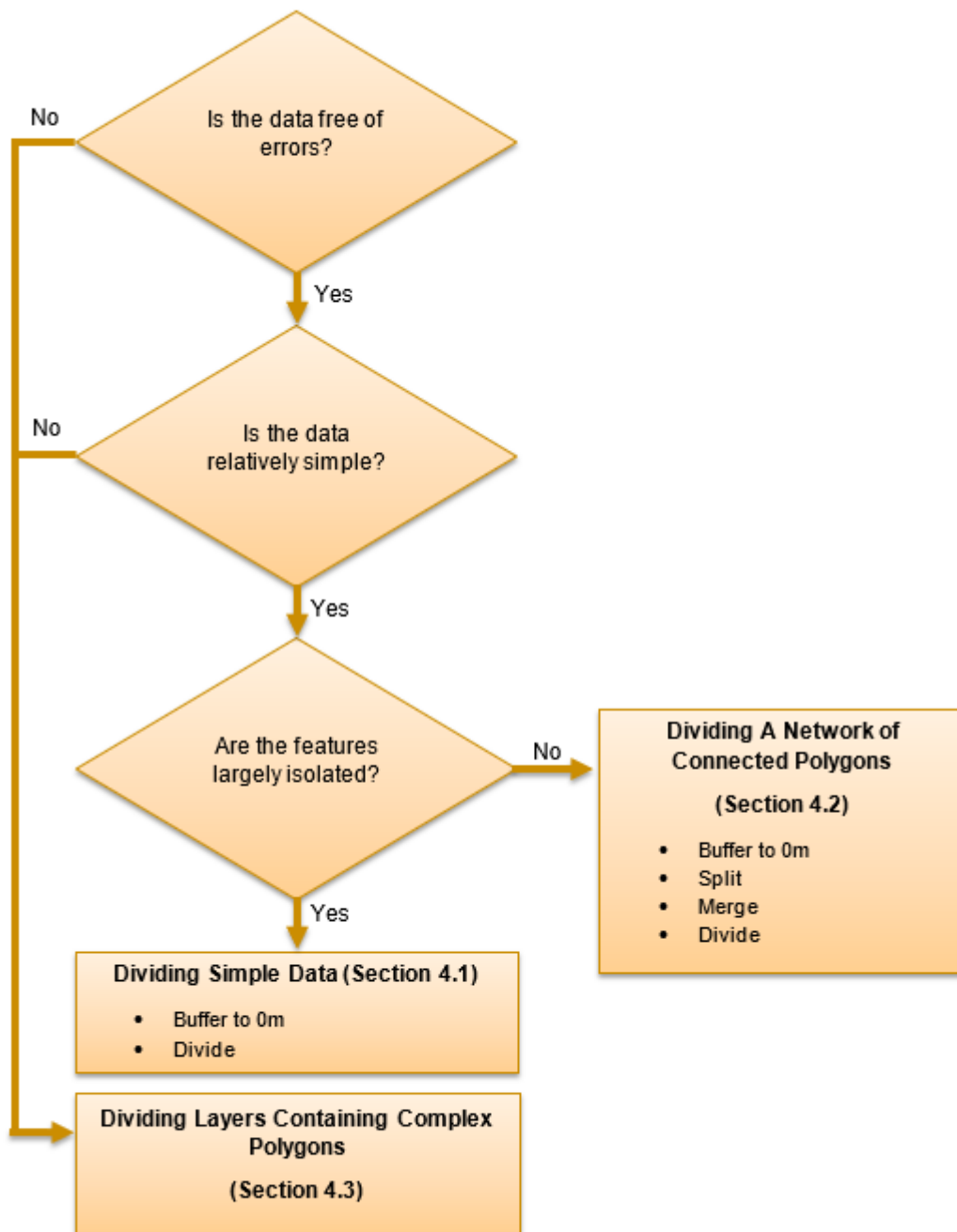


- Clicking on the 'Error resolution settings'  allows you to choose the default resolution. This should be set to 'Split features into a multi-object feature'.
- Click on the 'Fix selected errors using default resolution' to fix the errors.

Note: In cases where the self-intersections are extreme, the tool may fail to identify all instances within the data or they may cause the tool to crash. Further details on error identification and resolution is provided in **Appendix 1**.

4. Dividing

There are a number of factors that need to be considered when deciding on the optimum approach to dividing data. The diagram below can be useful when choosing which approach to take:

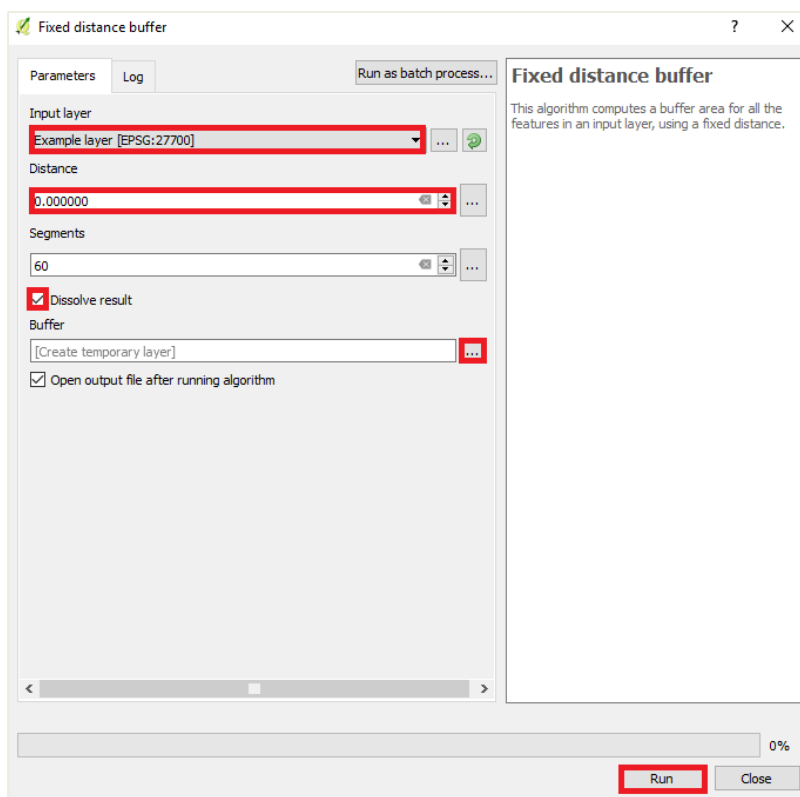


4.1. Dividing Simple Data

A by-product of separating the data by department and zone is that in some instances it will reduce the complexity and size of the data to the extent that the entire layer can be buffered and divided without being broken up any further. **Appendix 2** provides examples of data where this approach has been utilised.

The process is as follows:

- Buffer the entire layer to 0m (the reasons for buffering to 0m is provided in **Appendix 3**). The process is as follows:
 - Navigate to Vector > Geoprocessing Tools > Fixed distance buffer



- Follow the steps detailed in Section 4.4: Using the 'Polygon Divider'.

Note: Larger layers can be buffered more quickly using OpenJUMP (**Appendix 4**).

4.2. Dividing A Network of Connected Polygons

In some cases the polygons would not be considered complex in isolation, but become complex when they are part of a large connected network (Figure 1).

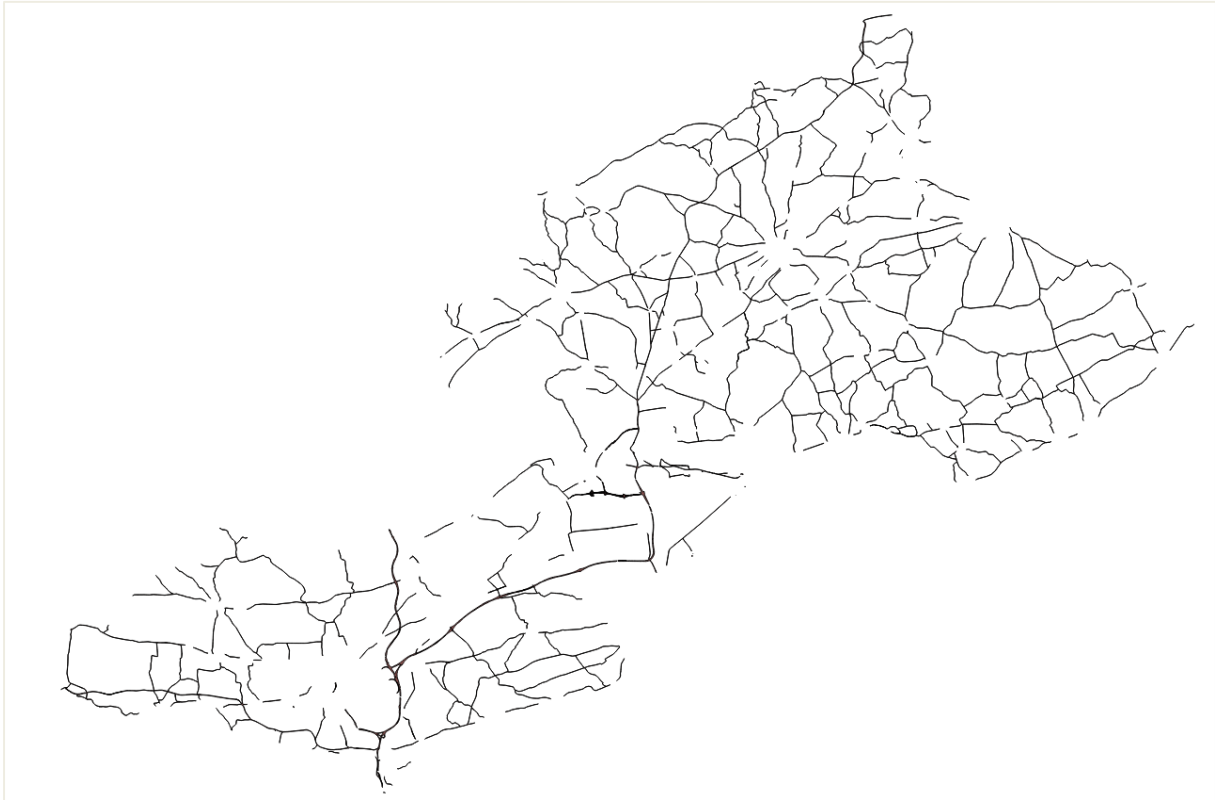
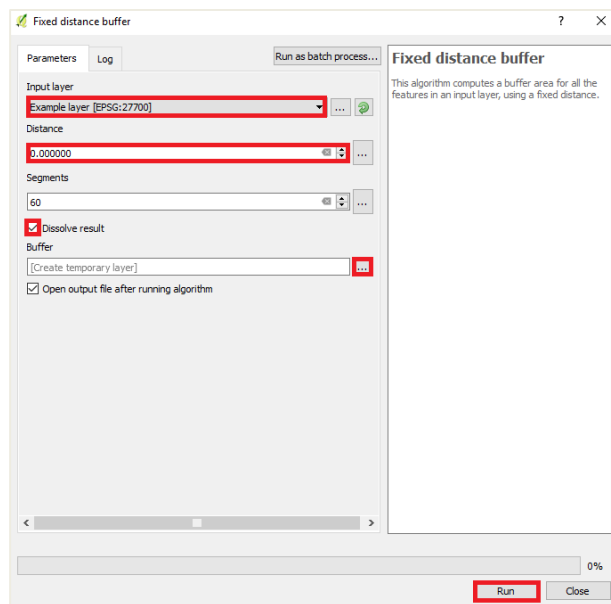



Figure 1: Example (Roads Maintenance: Zone 6) where simple polygons become complex due to being part of a large connected network.

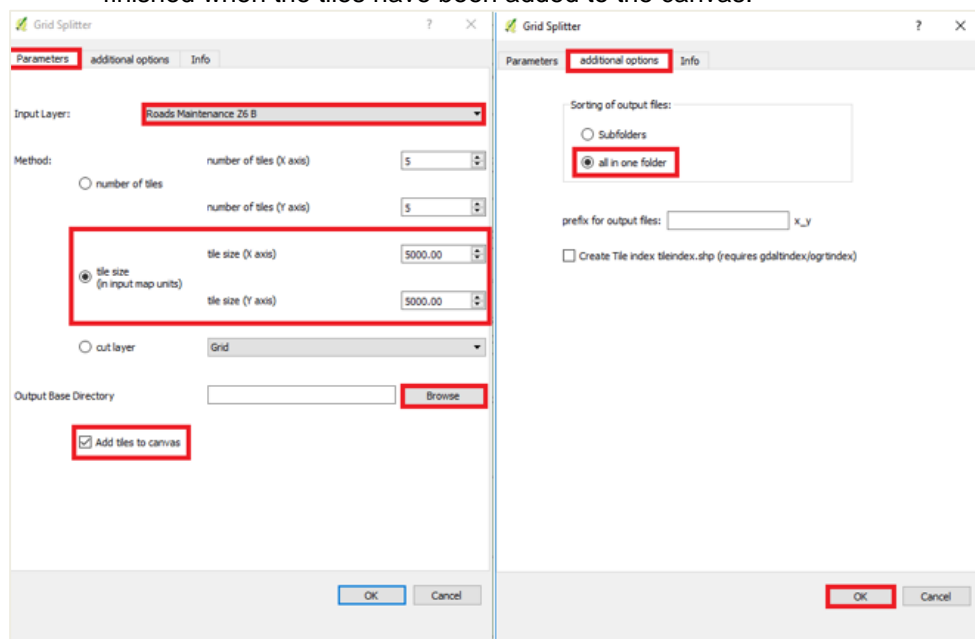
Where this is the case it is recommended that you create splits in the data prior to dividing. The process is as follows:

- Buffer the entire layer to 0m:
 - Navigate to Vector > Geoprocessing Tools > Fixed distance buffer



- In some instances buffering in QGIS will take a long time. Where this is the case, OpenJUMP is a quicker alternative (see **Appendix 4**).
- Split the layer into sections:

- Open the 'gridSplitter'  plugin (can be installed via Plugins > Manage and Install Plugins...).
 - **Note:** This tool does not work on Mac versions of QGIS.
- The 'Input Layer' is the layer you have just buffered to 0m.
- Under 'Method' select 'tile size' and enter the size of grid you want to split the layer by. This example uses 5000(m), but you may wish to experiment with different sizes, as the optimum size of grid will be data dependant.
 - **Note:**
 - If the grid size is small and the area large, thousands of tiles will be generated. In these instances it would be more efficient to either break the data into smaller sections prior to undertaking the pre-split, or to use a larger grid size.
 - Small grid squares take longer to split (as they generate more tiles) but will generally be quicker to divide (as the tiles generated will be less complex). Conversely large grid squares will split more quickly, but will generally take longer to divide.
 - A poorly chosen grid size can have a negative impact on the number of survey areas generated by the dividing process.
- Navigate to the 'Output Base Directory' where the files will be saved.
- Check the 'Add tiles to canvas' box.
- Click on the 'additional options' tab and select the 'all in one folder' option.
- Click 'OK'
 - **Note:** There is no progress bar with this tool and it can take a while to run (particularly when large numbers of tiles are being generated). The tool is finished when the tiles have been added to the canvas.



- Merge the split layers into a single shapefile:
 - Vector > Data Management Tools > Merge Vector Layers
 - Click on the three dots to the right of the 'Layers to merge' box and check the split layers (it is likely to be quicker to click 'Select all' and manually uncheck the layers not relating to the split).
 - Stipulate where you want to save it and click 'Run'.
- Follow the steps detailed in Section 4.4: Using the 'Polygon Divider'.

4.3. Dividing Layers Containing Complex Polygons

Not all layers will be uniformly simple or complex. For example, some Roads Maintenance layers are likely to be simple in rural areas, but highly complex in urban areas. In these instances it is inefficient to treat the whole layer as complex and ineffective to treat the whole layer as simple. It is therefore recommended that the layer is broken down into sections, so that the approach can be tailored to the complexity of the data.

Figure 2 provides an example of a dataset that contains a mixture of simple and complex polygons that cover a large area.



Figure 2: Example dataset

Given the complex areas within the dataset, treating the entire layer as simple would be ineffective. To demonstrate this, the simple approach was run on this dataset for 4 hours 45 minutes before being abandoned having reached only 24% completion.

Given the large areas of relatively simple data, it would be inefficient to treat the entire layer as complex. This is primarily down to the number of tiles that would need to be generated to create the pre-splits in the dataset (Figure 3). To demonstrate this, a pre-split using a 1km grid was run on the dataset. It took 1 hour 15 minutes just to pre-split the data.

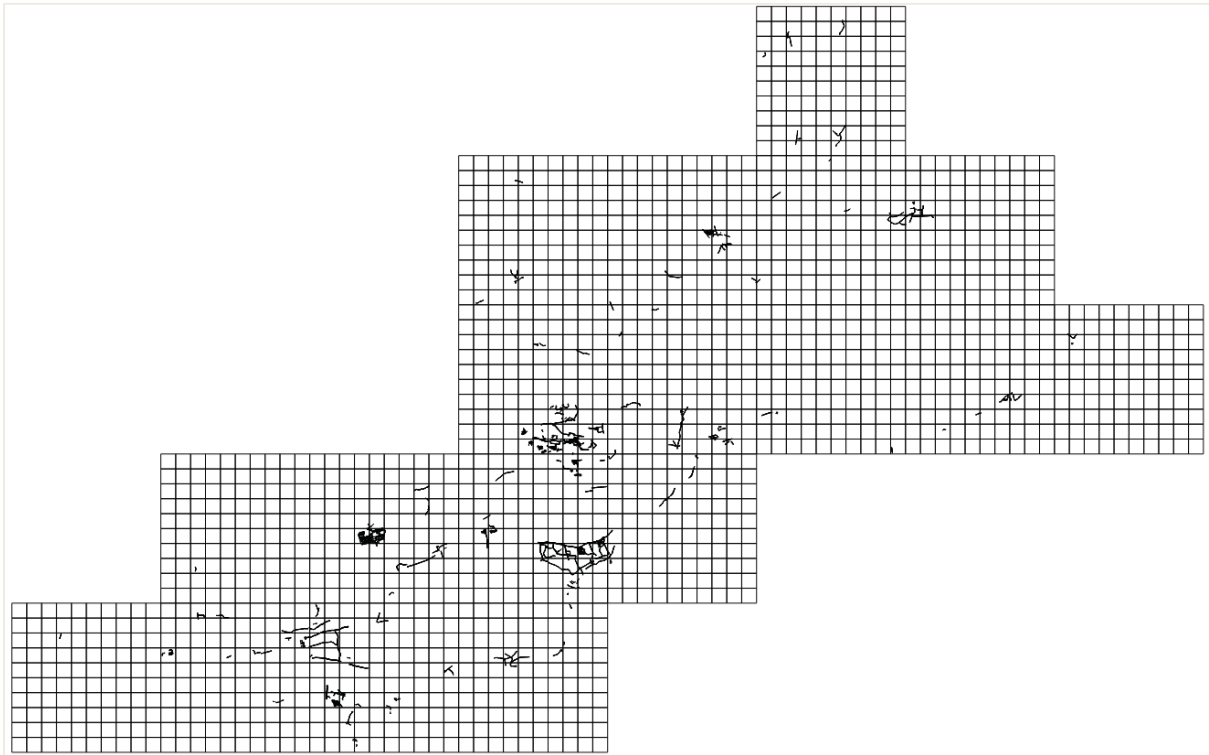


Figure 3: Illustration of the number of tiles that would need to be generated to undertake a useful pre-split on the dataset

The solution is to separate the data into areas of simple and complex data. This allows you to tailor the dividing approach to best suit the data. Figure 4 illustrates how the example dataset was separated into simple and complex areas. The two simple areas (green and brown) were divided using the simple method in 9 and 34 minutes respectively. The complex area (red) was divide using the complex method (1k pre-split) and was divided in 31 minutes. Therefore, by separating the data into simple and complex layers, the whole layer was divided in 1 hour 15 minutes.

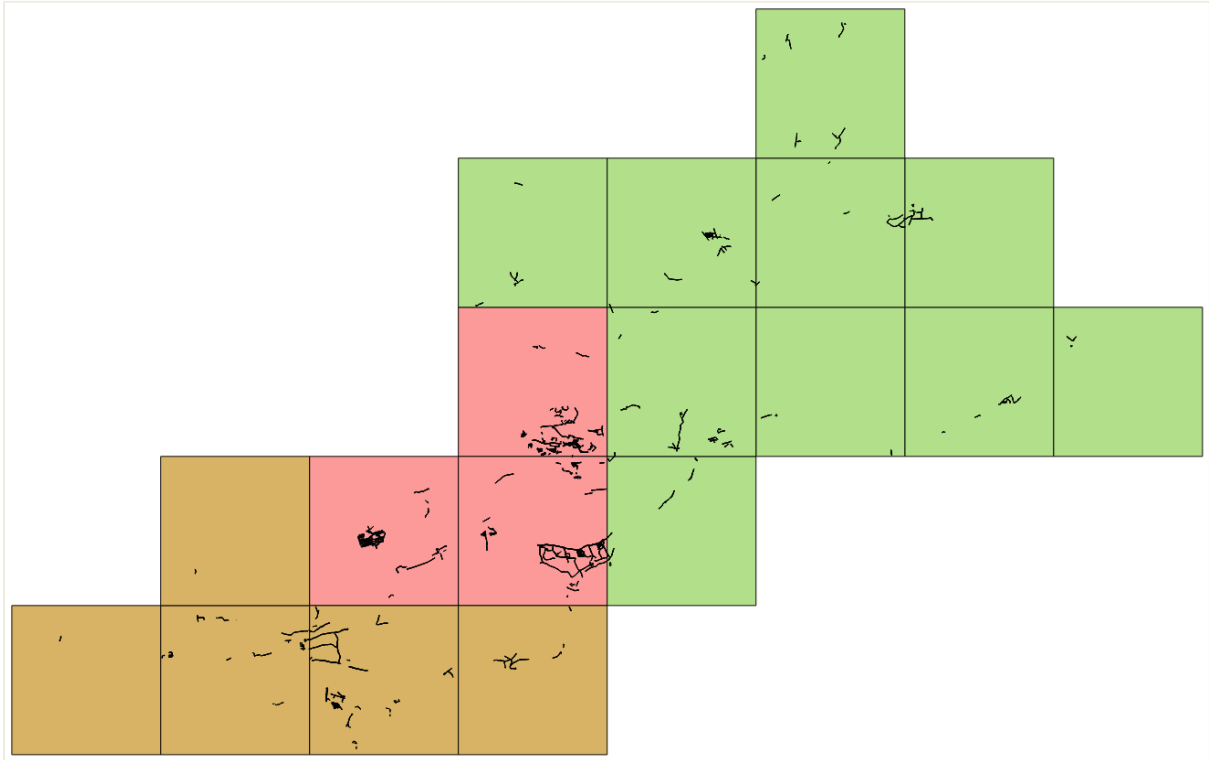


Figure 4: Illustration of how the dataset was separated into two simple areas (green & brown) and one complex area (red)

Where the data is highly complex (Figure 5) and/or complex across a large proportion of the area, it is important that the data is broken down into small sections. By way of example, the area shown in Figure 6 was broken down into 17 sections and utilised both the simple and complex processes.



Figure 5: Example of 1km grid square containing highly complex polygons

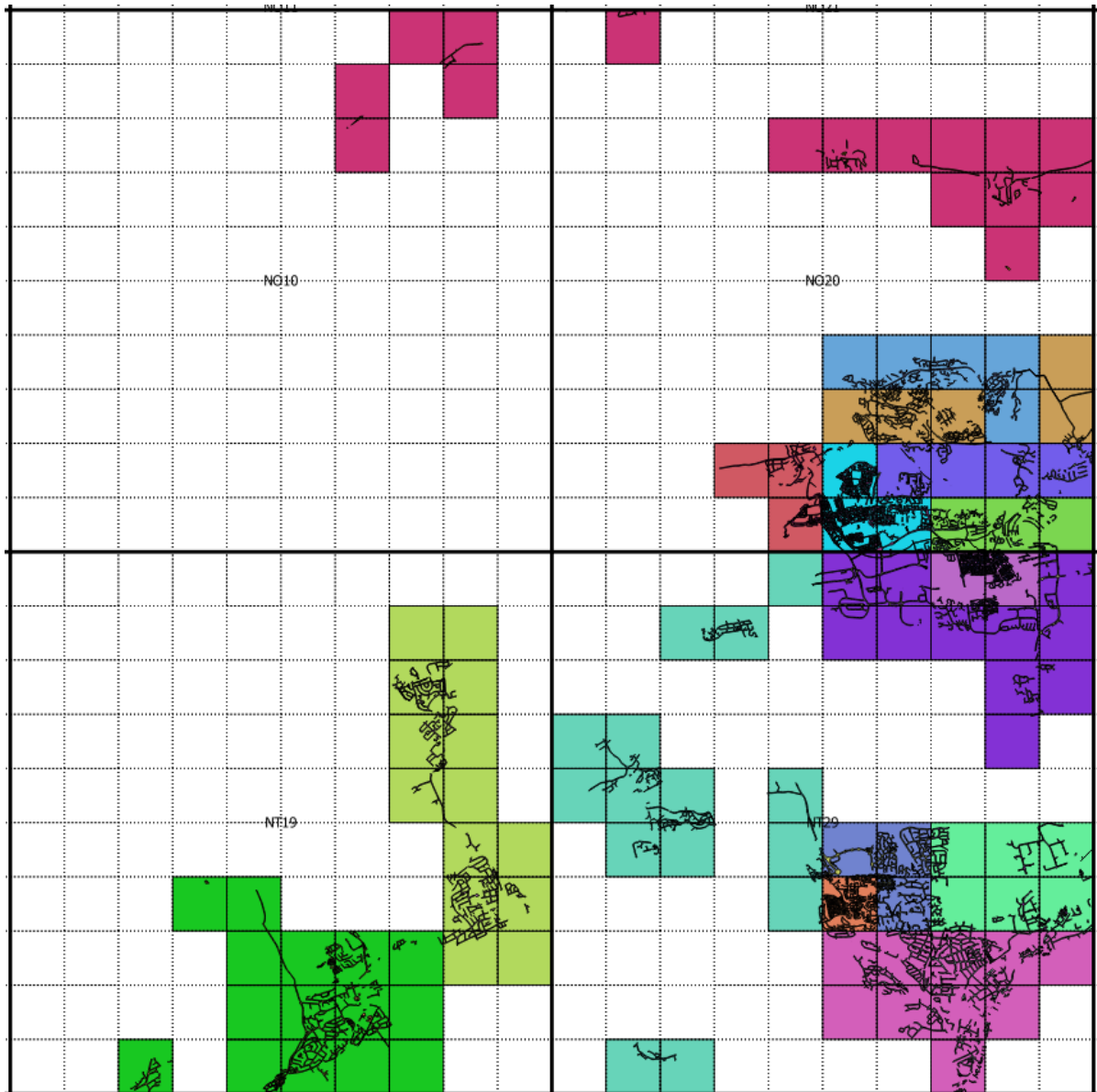


Figure 6: Example where the data is complex across a large proportion of the area (the coloured squares represent how the data was broken down prior to being divided)

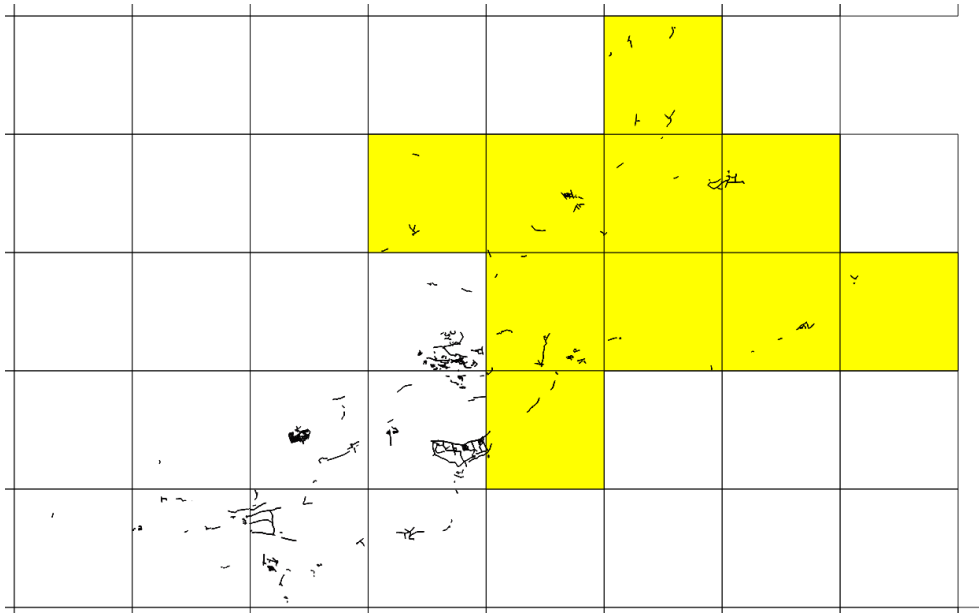
Always err on the side of caution when deciding on what size of section to split complex data into. The 'gridSplitter' tool used to pre-split the data becomes very inefficient if asked to generate too many tiles. Likewise, the 'Polygon Divider' will take a significant amount of time if too large a complex dataset is used. The quality of the data being divided should also be considered when deciding on how large an area to divide. This is because the time lost when an error in the data causes the 'Polygon Divider' to fail is minimal if the 'Polygon Divider' has only been running for a few minutes (i.e. a small area) but can be significant if it has been running for a few hours (i.e. a large complex area).

Numerous layers may be generated during this process, so it is essential that a robust file naming and organisation system is used to keep track of what areas have been divided.

The process for dividing layers that contain complex polygons is as follows:

- Separate the data into simple and complex layers:
 - Load the OS 10km Grid Square layer (this can be sourced from https://github.com/charlesroper/OSGB_Grids).

- The advantage of using OS Grid squares is that the sections are then referenceable and easily tracked.
- Select the grid squares you want to use to clip the layer.
 - For simple data select an area you are confident can be divided in one go.
 - Complex data will be broken down into further smaller sections later in the process and as such how you group the complex data together is not as important at this stage.
 - Any 10km grid square that has complex data in it should be considered complex.

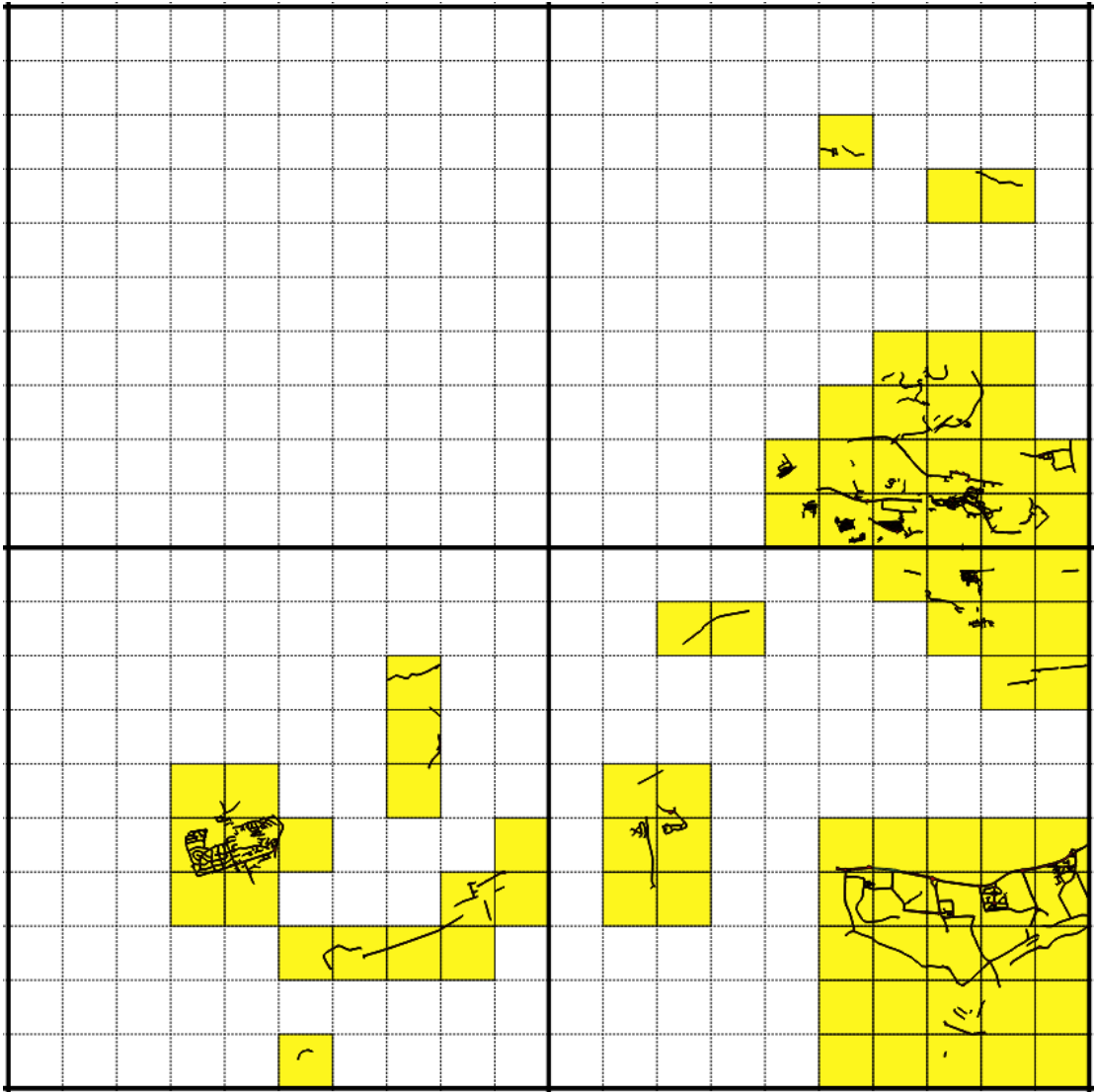


- With the area selected, navigate to Vector > Geoprocessing Tools > Clip:
 - The 'Input vector layer' is your department zone layer.
 - The 'Clip layer' is your grid layer.
- Browse to where you want the 'Output shapefile' to be saved.
- Repeat the process until all of the original layer has been clipped into layers representing simple or complex data.

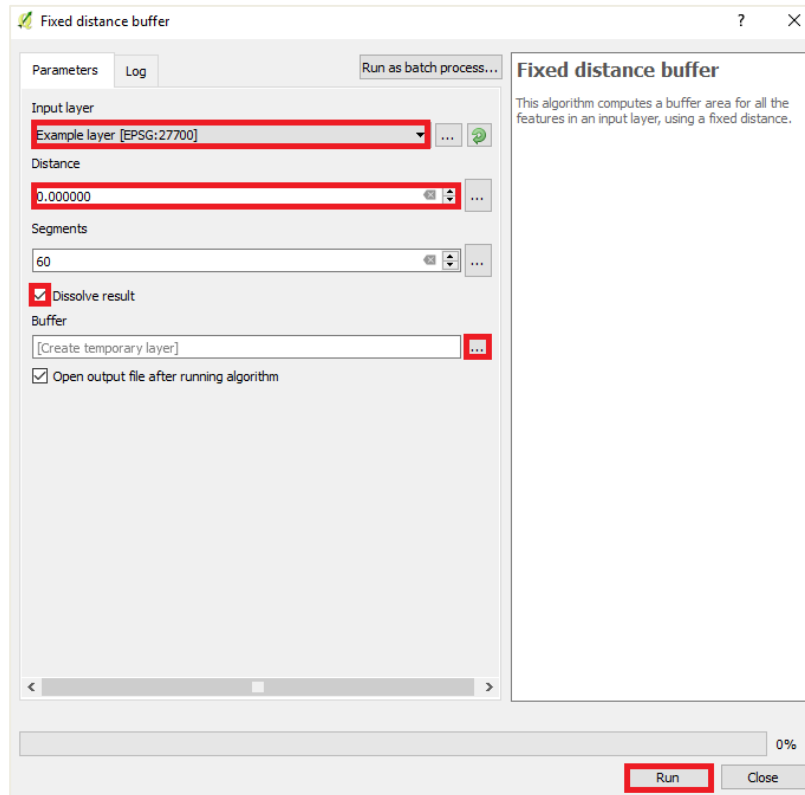
For simple data follow the process detailed in Section 4.1.


For complex data follow the process below:

- Load the 1km OS Grid Squares (this can be sourced from https://github.com/charlesroper/OSGB_Grids).
- Select the grid squares that contain your data.
 - Avoid selecting too many grid squares at once, as the tool used for the pre-split becomes inefficient when asked to generate a large number of tiles (example provided in Figure 3)
 - Areas of highly complex data should be broken into smaller sections (example provided in Figure 6).



- Save the selected grid squares as a new layer:
 - Right click on the 1km OS Grid layer in the 'Map Legend' and select 'Save As...'
 - Browse to where you want the output file to be saved.
 - Make sure CRS is British National Grid.
 - **Important:** Check 'Save only selected features' box.
 - Click 'OK'.
- Use the newly saved grid layer to clip the data.
 - Vector > Geoprocessing Tools > Clip.
 - 'Input layer' is your complex layer.
 - 'Clip layer' is your newly created grid layer.
 - Browse to where you want the output file to be saved.
- Buffer the clipped layer to 0m:
 - Navigate to Vector > Geoprocessing tools > Fixed distance buffer




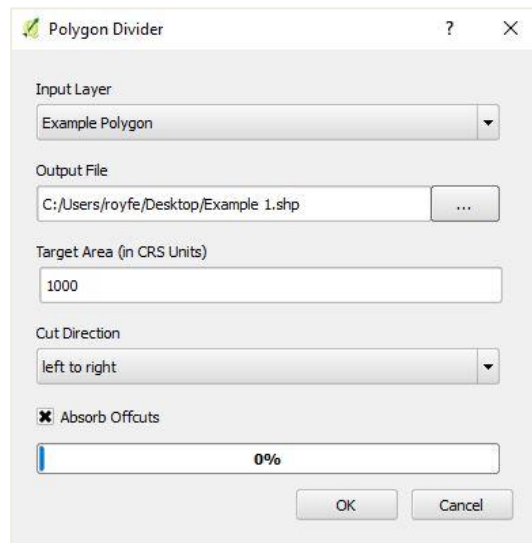
- In some instances buffering in QGIS will take a long time. Where this is the case, OpenJUMP is a quicker alternative (see **Appendix 4**).
- Pre-split the data:
 - Open the 'gridSplitter'  plugin (can be installed via Plugins > Manage and Install Plugins...).
 - **Note:** This tool does not work on Mac versions of QGIS.
 - The 'Input Layer' is the layer you have just buffered.
 - Under 'Method' select 'cut layer' option and chose your grid layer from the pulldown menu.
 - Navigate to the 'Output Base Directory' where the files will be saved.
 - Check the 'Add tiles to canvas' box.
 - Click on the 'additional options' tab and select the 'all in one folder' option.
 - Click 'OK'
 - **Note:** The tool is finished when the tiles have been added to the canvas.
- Merge the split layers into a single shapefile:
 - Vector > Data Management Tools > Merge Vector Layers
 - Click on the three dots to the right of the 'Layers to merge' box and check the split layers (it is likely to be quicker to click 'Select all' and manually uncheck the layers not relating to the split).
 - Stipulate where you want to save it and click 'Run'.
- Follow the steps detailed in Section 4.4: Using the 'Polygon Divider' and repeat for all complex data.

4.4. Using the 'Polygon Divider'

The process for using the 'Polygon Divider' is as follows :

- Check the layer to be divided has its CRS set to British National Grid to ensure the 'Target Area' is in metres;

- Click on the 'Polygon Divider' button  in the top toolbar (or select from the 'Plugins' pulldown menu).



- Select the 'Input Layer' to be divided;
- Select the save location for the divided layer:
 - Consider the naming convention used. You may want to reference the department, grid square, zone, buffered and divided (e.g. Roads_Maintenance_NO21_Z5_B_Divided).
- Set the 'Target Area' to 1000;
- Select the 'Cut Direction' of the plugin.
 - **Note:** the direction in which a polygon is cut up, has an impact on the output. In some instances, changing the 'Cut Direction' will improve the overall output, in others, it may reduce the number of offcuts. In some cases, the algorithm will be forced to adjust the cut direction in order to divide certain geometries;
- It is envisaged that you will want to keep the 'Absorb Offcuts' checkbox checked. This minimises the number of offcuts where possible; and
- The file will be added to the Map Legend as 'Divided Polygon'. It is recommended that you resave the layer (right click, save as...).
 - **Important:** For each department zone, save all of the divided layers in a single folder as this makes it simpler when merging into a single shapefile.

5. Getting Data into Final Form



5.1. Merge Shapefiles

Department zone layers that contained complex polygons (i.e. have been divided in multiple sections) should be merged back into a single layer. For example, if you had broken your Roads Maintenance Zone 4 layer into 3 sections during the divide process, you should merge the 3 divided layers into a single shapefile (i.e. Roads_Maintenance_Z4_B_D). The process is as follows:

- Load into QGIS the layers you want to merge.
- Navigate to Vector > Data Management Tools > Merge Vector Layers
- Click on the three dots to the right of the 'Layers to merge' box and check the layers you want to merge.
- Stipulate where you want to save it and click 'Run'.

5.2. Optional Checks

Data quality (in particular self-intersections) along with human error can result in data being lost during the buffering and dividing process. Consequently you may choose to undertake a basic check for data loss at this point. The process is as follows:

- Navigate to Vector > Geoprocessing Tools > Difference
 - The 'Input layer' is the original department zone layer.
 - The 'Difference layer' is the divided department zone layer.
 - Check 'Ignore invalid input features [optional]'.
 - Click 'Run'.
- Left click on the 'Difference' layer in the 'Map Legend'.
- Click on the 'Attribute Table' button  on the top toolbar;
 - Click on the 'Open field calculator'  button on the toolbar of the 'Attribute Table' window.
 - Type 'Area' in the 'Output field name'.
 - Select 'Decimal number (real)' from the 'Output field type' menu.
 - Type \$area in the expression box.
 - Click 'OK'.
 - Click twice on the newly created field heading ('Area') in the 'Attribute Table'. This will sort the features by area size (largest to smallest).
 - If none of the features have a significant area, no further action is required.
 - If there is a feature with a large area:
 - Left click on the row number to select the row.
 - Then right click anywhere on the row and click on 'Zoom to feature'.
 - Decide whether the missing feature is significant (i.e. needs to be included as a potential survey area). If it does, you will need to create a layer that captures the missing polygon(s), run through the dividing process and merge with the relevant divided department zone layer.






This process would need to be repeated for each department zone.

It is a time consuming step, so you may only want to undertake it if you have concerns over data losses.

5.3. Spatial Join

Before performing the 'Spatial Join', it is **important** that any fields relating to the original undissolved layer are removed. This ensures the joined layer does not have duplicates that would affect the prescribed field naming conventions.

The process for deleting fields from the 'Attribute Table' is:

- Left click on the divided layer in the 'Map Legend'.
- Click on the 'Attribute Table' button  on the top toolbar.
- Make the layer editable .
- Click on the delete fields button .
- Select the layers you want to delete from the dialog box and click 'OK'.
- It may take a few minutes if you are deleting a number of fields from a large database.
- Save the changes  and stop the layer from being editable .

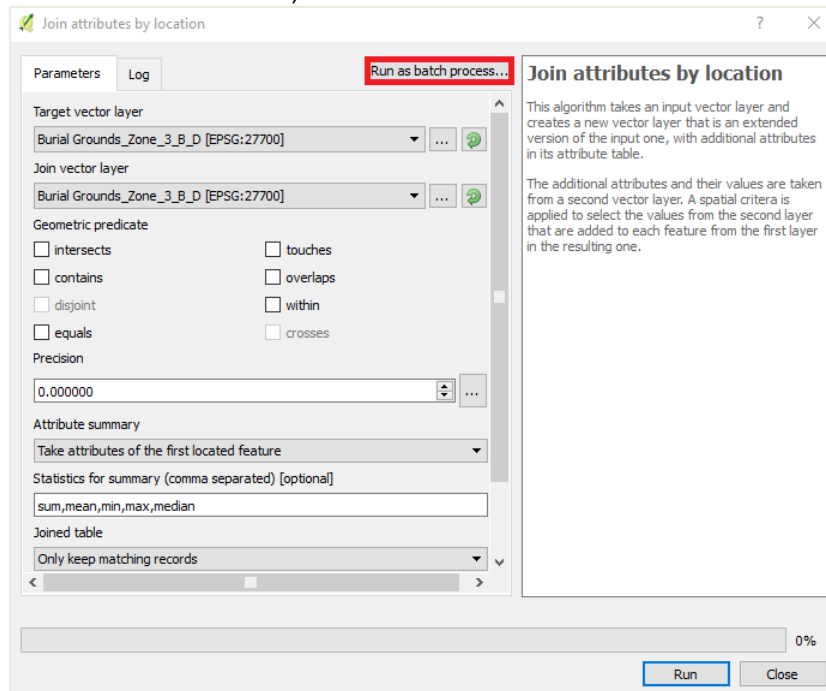
The only fields remaining after the above process should be:

- Poly_ID;

- Unique_ID;
- Area;
- PointX; and
- PointY.


The 'Spatial Join' process is:

- Navigate to Vector > Data Management Tools > Join attributes by location
- Click on the 'Run as batch process...' button (this may not be available if you are running an earlier versions of QGIS).



- The 'Target vector layer' is the divided department zone.
- The 'Join vector layer' is the original department zone.
- The 'Geometric predicate' should be set to 'intersects'.
- 0 should be entered for precision.
- 'Attribute summary' should be left at the default setting (i.e. 'Take attributes of the first located feature').
- 'Statistics for summary (comma separated)' should be left at default (i.e. 'sum,mean,min,max,median')
- 'Joined table' should be set to 'Keep all records (including non-matching target records)'.
- 'Joined layer' navigate to the folder where you want to save your spatial joins.
 - Enter SJ as the file name and click 'Save'.
 - A box will pop up:
 - Select 'Fill with parameter values' from the 'Autofill mode' menu.
 - Select 'Target vector layer' from the 'Parameter to use' menu.
- 'Load in QGIS' should be left at the default setting (i.e. 'Yes').
- Click 'Run' (it may take a while given the file sizes).

5.4. Attribute Table

Any unwanted or duplicate fields should be deleted  from the 'Attribute Table'. The exact fields that should remain are:

- Org;
- Department;
- Zone;
- H_and_S;
- Reason;
- Poly_ID;
- Unique_ID;
- Area;
- PointX; and
- PointY.

5.5. Merge shapefiles

It is recommended that you merge the spatially joined shapefiles (the file size of the Roads Maintenance layer may be quite large and as such you may wish to keep it as a separate layer).


The process for merging is as follows:

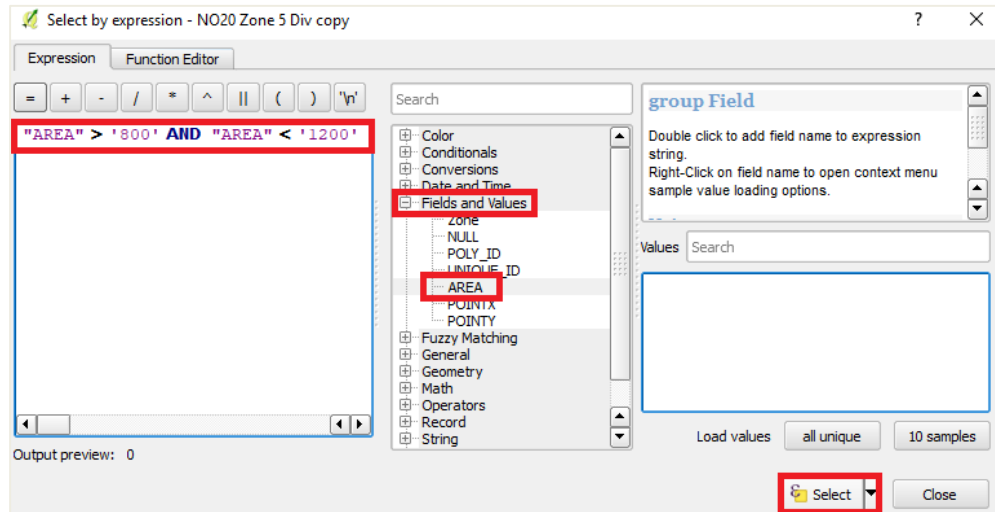
- Navigate to Vector > Data Management Tools > Merge Vector Layers
- Click on the three dots to the right of the 'Layers to merge' box and check the layers you want to merge.
- Stipulate where you want to save the merged shapefile and click 'Run'.

5.6. Refining Survey Area Size

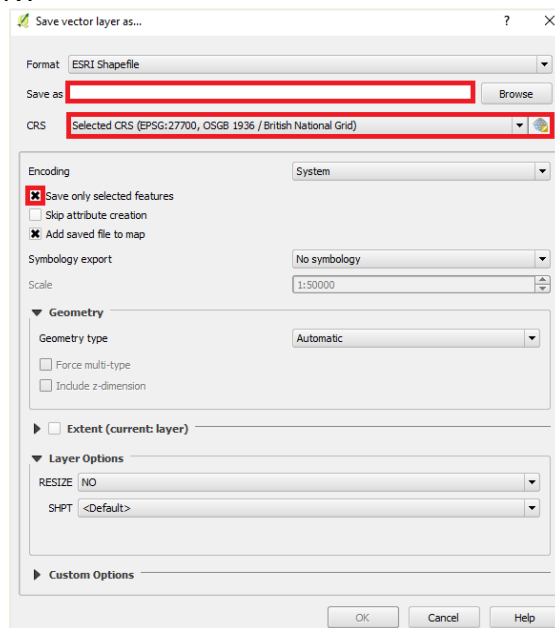
The next step is to refine the size of the survey areas to ensure all survey areas are between 800m² and 1,200m².

The process for refining the survey areas is:

- Left click on your merged layer in the 'Map Legend'.
- Click on the 'Select features using an expression'  button in the top toolbar;
- In the dialog box, expand the '**Fields and Values**' list and double click on 'Area';
- Type:
 - '800' AND "AREA" < '1200'
- Click on the 'Select' button on the bottom right of the dialog box:



- Close the dialog box.
- Right click on your divided layer in the 'Map Legend' and select 'Save As...'
- Browse to where you want the new layer to be saved.
- Make sure the CRS is British National Grid.
- **Important:** Check the 'Save only selected features' checkbox.
- Click 'OK'.



- You now have a divided layer with survey areas ranging between 800 and 1200m².

5.7. Save as WKT

The software developers have requested that the files be provided as WKT files. Therefore, the final step is to resave the shapefiles as WKT.

The process is as follows:

- Right click on the layer in the Map Legend and select 'Save As...'
- Change the format to 'Comma Separated Value [CSV]'
- 'Browse' to where you want to save the output file.
- Make sure 'CRS' is set to British National Grid.
- Change the 'GEOMETRY' to 'AS_WKT'

- Click 'OK'.

Save vector layer as...

Format: Comma Separated Value [CSV]

File name: [] Browse

Layer name: []

CRS: Selected CRS (EPSG:27700, OSGB 1936 / British National Grid)

Encoding: System

Save only selected features

▶ Select fields to export and their export options

Add saved file to map

Symbology export: No symbology

Scale: 1:50000

▼ Geometry

Geometry type: Automatic

Force multi-type

Include z-dimension

▶ Extent (current: layer)

▼ Layer Options

CREATE_CSVT: NO

GEOMETRY: AS_WKT

LINEFORMAT: <Default>

SEPARATOR: COMMA

WRITE_BOM: NO

OK Cancel Help

 01786 433930

 zerowastescotland.org.uk

 @ZeroWasteScot

Zero Waste Scotland is a registered company in Scotland [SC436030]. Registered office: Ground Floor, Moray House, Forthside Way, Stirling FK8 1QZ