

Overview: Spatial Quality Anomalies Diagnosis (SQUAD) Tool for QGIS

1. Introduction and Background

Knowledge about health facility locations is important in addressing HIV, maternal and child mortality, and other issues. As a result, there has been a rapid growth in large geospatial data sets, such as master facility lists (MFLs). An MFL and other similar lists typically contain locations of health facilities as well as attributes of the facilities—including name, address, or which administrative unit the facility is located in.

Assessing the quality of these data sets can be challenging, because there are two types of possible errors: **spatial errors** and **attribute errors**. Assessing spatial errors involves looking at such things as the presence of a coordinate, whether it is properly recorded, and the accuracy of its location. Assessing attribute errors involves determining whether attributes such as site name or site ID are correct.

When you work with large spatial data sets, manually reviewing each record to validate both location and attribute information can be prohibitively time-consuming. A more effective approach would be to look for anomalies in the data that may indicate data quality issues. MEASURE Evaluation's **Spatial Quality**

Anomalies Diagnosis (SQUAD) Tool—<https://www.measureevaluation.org/resources/tools/geographic-information-systems/squad-tool>—identifies six types of anomalies in spatial data:

Anomaly 1 Missing coordinates	Anomaly 4 Duplicate facility name
Anomaly 2 Coordinates truncated; possible typos	Anomaly 5 Site falls outside of expected district, but near a district boundary (< 2 kilometers)
Anomaly 3 Duplicate coordinates	Anomaly 6 Site falls far outside of expected district (< 2 kilometers)

2. Prerequisites to Use the Tool

Two inputs are required to use the SQUAD Tool. Both should be stored as shapefiles. (Properly formatted sample files are provided with the tool.) The tool works with QGIS 2.18 or QGIS 3.0.

Here are the details about each of the two required shapefiles:

Sites. Should be a point file of sites (i.e., facilities) that contains separate fields for:

- Unique identifier for each site
- Name of each site
- Name of administrative unit in which the site is supposedly located
- Separate fields for X and Y coordinates (longitude and latitude) of each site

Administrative units. Should be a polygon file for administrative units in the country or region of interest. The file should contain a field for the name of each administrative unit (e.g., “district”). These names should come from the same standard list of administrative unit names that the sites use.

If the administrative unit file contains especially detailed polygons, it may be necessary to use the “Simplify Geometries” command to eliminate some of the finer detail, so that the tool will not run excessively slowly. (The recommended administrative unit file size is 2 megabytes or less.)

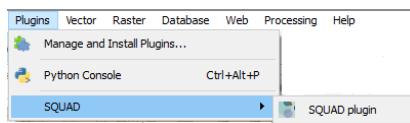
3. How to Use the Tool

1. Create a new project file in QGIS.
 (An example project file, SQUAD_Demo.qgs, has also been provided to permit users to test the tool.)

2. Load in the two required shapefiles.

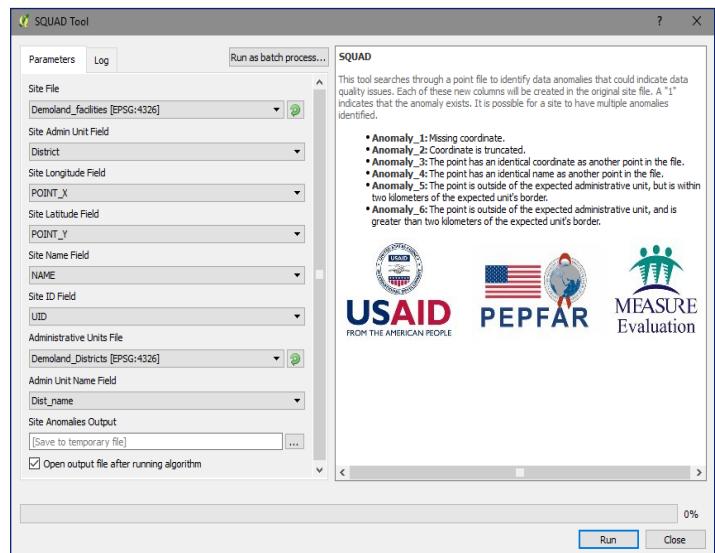
- Site locations
- Administrative units

3. Copy the “SQUAD Plugin” to the “Plugins” folder and activate it.



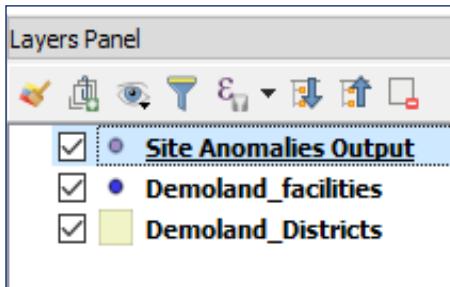
4. Click on the “SQUAD” tool to run it.

A dialogue box will open. You must identify the inputs to the tool, as illustrated here:



4. Results from the Tool

While the tool is running, a progress bar will appear. When the script has run, a new layer called “Site Anomalies Output” will be added to the Layers Panel. You can choose to save this as another shapefile or as an Excel spreadsheet.



Fields for each anomaly code will be added to the attribute table of the new layer. A value of “1” means the anomaly exists for the record.

Records that have an anomaly can then be investigated to determine the quality of the data. For instance, it could be all right for sites to be in the same location or to have the same name. If an Anomaly_5 is found, the error could be inaccuracy of the administrative unit boundary, or it could mean that there is inaccuracy in the site location.

Site Anomalies Output :: Features total: 80, filtered: 80, selected: 0										
	NAME	Region	Country	POINT_X	POINT_Y	District	UID	Anomaly_1	Anomaly_2	Anomaly_3
1	Robin	Animal	Demoland	-44.89533611730	30.86118355340	Bird	dc6cba704964			
2	Hawk	Animal	Demoland	-45.01422905090	30.81785074000	Bird	437ef5aa687a			
3	Pigeon	Animal	Demoland	-44.70000000000	31.13618070930	Bird	8e52d1290c31		1	
4	Woodpecker	Animal	Demoland	-44.74283078150	31.00000000000	Bird	4c2c62e984ee		1	
5	Crow	Animal	Demoland	-45.04645066750	30.90451645490		7c55d0a02f16			
6	Parrot	Animal	Demoland	-44.81005638280	30.75396183680	Bird	60bd9e0f0e34			
7	Ostrich	Animal	Demoland	-45.05256077720	31.00701527390	Bird	c3023ba971c7			