# 1 Introduction

This document describes database and file layouts for the Seismic Hazard estimates used in the Appalachian Basin Geothermal Play Fairway Analysis (GPFA-AB; Jordan et al., 2015). A detailed description of those techniques may be found in Horowitz and Appalachian Basin GPFA Team (2016) which is an updated version of Horowitz (2015).

# 2 Details of Worms Stored in a GIS

Of the techniques used in our analysis, probably the least familiar to the reader are the Poisson Wavelet Multiscale Edge Detection ('worm' for brevity; Hornby et al., 1999) results, and how they are calculated. This section describes the essence of the worm edge detection, and how those geometric objects are stored in a PostGIS database as well as reproduced here for external use.

Mathematically, the worms are calculated as points on each level of a potential fields upward continuation (e.g. Blakely, 1996, pp 313-314) via an edge detection procedure. Loci of points in a function $f(x, y; z)$ that satisfy

$$\frac{\nabla_2(\|\nabla_2 f\|) \cdot \nabla_2 f}{\|\nabla_2 f\|} = 0 \tag{1}$$

are local maxima in horizontal gradients and are marked as belonging to an edge. Here,

$$\nabla_2 f \equiv (\partial/\partial x, \partial/\partial y) f$$

which is the horizontal 2D gradient vector of $f(x, y; z)$, and

$$\|\nabla_2 f\| \equiv \sqrt{(\partial f/\partial x)^2 + (\partial f/\partial y)^2}$$

which is the magnitude of that gradient vector. Strictly, points with zero horizontal gradient also satisfy equation (1) but they do not commonly occur in real data. Intuitively, equation (1) can be thought of as describing where the gradient of the (normalized) magnitude of the gradient of $f$ changes sign when projected along a gradient "streamline" of $f$.

Obviously, the gradients described above (and elsewhere in the processing) must be calculated in Cartesian coordinates, not in latitudes and longitudes. For the purposes of this project we performed all these calculations in UTM Zone 18N (EPSG code 32618) even though some of the data actually lie nearby in an adjacent UTM zone. We simply accepted the relatively minor projection errors resulting from approximating one UTM zone with another for those nearby locations.

The function $f$ in our situation is the upward-continued gravity or pseudo-gravity field from our region. Because $f$ is approximated via a raster representation, we actually mark points at zero-crossings of equation (1) along linearly interpolated pixel boundaries – yielding a super-resolved (sub-pixel) precision of locations for edge points.

Using freely available mathematical network/graph software (NetworkX; Hagberg et al., 2008), we construct a graph using the zero-crossing points identified above as graph-nodes, and connecting them via graph-edges to their nearest neighbours (identified via a spatial proximity query using kD-trees, and limited to being no further away than the diagonal length of a pixel). We then use NetworkX and custom code to construct a complete set of minimum-spanning-trees from the previously partly-organised graph data. Those minimum-spanning-trees organise our worms into distinct segments. Each segment is composed of ordered nodes that are zero-crossings of equation (1) and straight-line connecting edges of length no greater than the diagonal of a pixel.

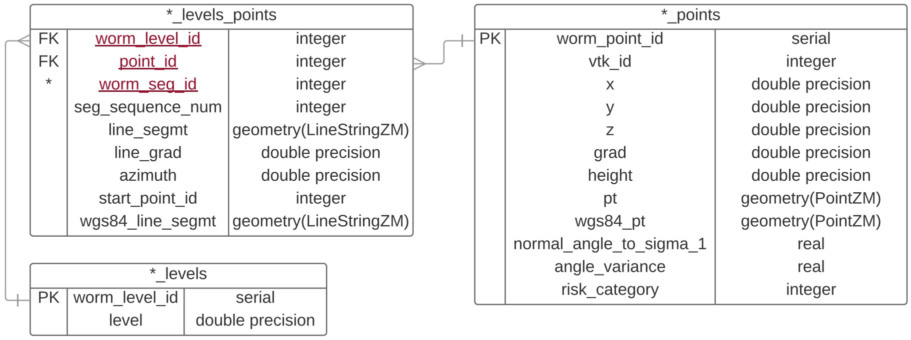


Figure 1: Entity Relational Diagram for storing worms in a PostGIS database. The table labelled *_levels holds the set of heights. The table labelled *_points holds the points and various auxiliary attributes. The table labelled *_levels_points – in addition to providing the many-to-many relationship between the other two tables – also stores the edge's geometry as well as some attributes. This latter database table has a composite primary key (shown underlined in red) made up of pointers into the other two tables and an index 'worm_seg_id' that enumerates the minimum-spanning-tree segments computed previously by NetworkX – these segment-enumerating indexes are only unique within each layer, not within the database table. Within ArcGIS or QGIS, displaying either geometry field in the *_points table shows the worm points, while displaying either geometry field in the *_levels_points table shows the connected worm edges.

Table 1: Description of *_level table fields.

| Key? | Field Name | Storage | Description |
|---|---|---|---|
| Primary Key | worm_level_id | integer | A unique integer: level 1, level 2, etc. |
| | level | double precision | Height/depth in kilometers. |

Once the above data structures are computed, we store them in a working PostGIS database for retrieval by GIS software. Three interconnected tables are involved: one for the set of upward continuation levels (equivalent to underground depth via the inverse wavelet transform arguments in Boschetti et al., 2001; Hornby et al., 2002); one for the node points; and one for the edges. Figure 1 displays the database tables' structure and their interconnections. Tables 1, 2, and 3 show an alternate view of the same data structures.

Open source (BSD licensed) software to perform these computations are described by Horowitz and Gaede (2014). The *git* source code management repository is located at `https://bitbucket.org/fghorow/bsdwormer`. By far, the easiest way to get the code running locally is to follow the directions at `https://bitbucket.org/fghorow/bsdwormer/wiki/Installation%20via%20a%20virtual%20machine%20using%20Vagrant`.

PostGIS databases require an appropriately configured Postgres server to be useful. Also, the interrelations between the GIS information sources shown in figure 1 are not preserved upon export to the ArcGIS shapefile format commonly used for GIS data interchange – due to limitations apparently inherent in the design of shapefiles. For these reasons, we chose to use a spatialite file (a spatial extension to the sqlite single-file database storage format; Furieri, 2015) as an interchange format that preserves all of the interrelations between our database tables. Modern versions of ArcGIS and QGIS (and hopefully other GIS systems) can read, display, and manipulate spatialite database files

Table 2: Description of *_points table fields.

| Key? | Field Name | Storage | Description |
|---|---|---|---|
| Primary Key | worm_point_id | serial | A unique integer. |
| | vtk_id | integer | VTK visualization point index. |
| | x | double precision | The UTM Eastings in meters. |
| | y | double precision | The UTM Northings in meters. |
| | z | double precision | Height above/below field raster in meters. |
| | grad | double precision | The value of M in (pseudo)milligals/meter. |
| | height | double precision | Corresponds to worm_level_id in *_levels table. |
| | pt | geometry(PointZM) | PostGIS PointZM geometry in UTM18N coordinates. |
| | wgs84_pt | geometry(PointZM) | PostGIS PointZM geometry in WGS84 coordinates. Stored for convenience of downstream processing. |
| | normal_angle_to_sigma_1 | real | Angle between the normal to a secant of edges at this point and the World Stress Map direction of principal compressive stress interpolated to this location. See Horowitz and Appalachian Basin GPFA Team (2016) for |
| | angle_variance | real | The variance estimate for normal_angle_to_sigma_1. See FIXME for details. |
| | risk_category | integer | An arbitrary risk metric determined by the magnitude of normal_angle_to_sigma_1. See text for discussion. |

Table 3: Description of *_points table fields.

| Key? | Field Name | Storage | Description |
|---|---|---|---|
| Foreign Key | worm_level_id | integer | Pointer to worm_level_id in the *_levels table. |
| Foreign Key | point_id | integer | Pointer to worm_point_id for the end point of the graph edge in the *_points table. |
| Composite | worm_seg_id | integer | Identifier for connected graph edges, as determined by NetworkX's minimum spanning tree algorithm. |
| | seg_sequence_num | integer | Index that orders graph edges within the same worm_seg_id. |
| | line_segmt | geometry(LineStringZM) | PostGIS LineStringZM geometry for each graph edge in UTM18N coordinates. |
| | line_grad | double precision | The value of M averaged between the values for point_id and start_point_id in (pseudo)milligals/meter. |
| | Azimuth | double precision | Orientation of the graph edge in degrees East of North. |
| (Foreign Key) | start_point_id | integer | Pointer to worm_point_id for the start point of the graph edge in the *_points table. |
| | wgs84_line_segmt | geometry(LineStringZM) | PostGIS LineStringZM geometry for each graph edge in WGS84 coordinates. Stored for convenience of downstream processing. |

on an equal footing with other spatial databases, and so by this strategy we gain portability for our worm databases.

# 3 Files

These files contain the final worm results, and some of their critical upstream intermediate processing results from Jordan et al. (2015) (as detailed in Horowitz and Appalachian Basin GPFA Team, 2016; Horowitz, 2015).

**Files in the Gravity Folder** These are the gravity grids and worms for the region.

- **GravStationsMerged.{shp,dbf,prj,shx}** A collection of shapefile parts containing the locations and Bouguer anomaly estimates for the measurements underlying the above interpolated grids. The Bouguer readings are in milligals, and the positions are in units of meters in the UTM zone 18N coordinate system.
- **AppBasinMergedBGA2500.tif** This is a geotiff (see, e.g. GDAL Development Team, 2016) of floating point values of the Bouguer anomaly interpolated for our study area. Pixel size is 2.5 km.
- **AppBasinMergedBGA2500Padded.tif** This is a geotiff of the file immediately above, padded and apodized for 2D spatial Fourier transform operations.
- **bga_worms.sqlite** This is a Spatialite (Furieri, 2015) file containing the Bouguer gravity worms in the format described in the previous section.
- **BGAWorms.sql** This is a file of sql commands for the spatialite system (Furieri, 2015) that retrieves all of the interconnected Bouguer worm tables from a PostGIS database and creates the .sqlite file above. Uses the "mod_virtualpg" sqlite extension (found at `https://www.gaia-gis.it/fossil/virtualpg/index`). Included here for reference.

**Files in the Magnetic Folder** These are the magnetic grids and worms for the region.

- **ravat_NURE-NAMAM2008_UTM18N_AppBasin.tif** A total magnetic intensity map of the region; extracted from Ravat et al. (2009). In geotiff format.
- **ravat_NURE-NAMAM2008_UTM18N_AppBasin_PSG.tif** A pseudogravity transform ( see, eg. Blakely, 1996, pp. 343ff) of the above file. Computed using the commercial software package Oasis montaj `http://www.geosoft.com/products/oasis-montaj/overview`.
- **ravat_NURE-NAMAM2008_UTM18N_AppBasin_padded_PSG.tif** This is a geotiff of the file immediately above, padded and apodized for 2D spatial Fourier transform operations.
- **ravat_mag_worms.sqlite** This is a Spatialite (Furieri, 2015) file containing the pseudogravity worms in the format described in the previous section.
- **PSGWorms.sql** This is a file of sql commands for the spatialite system (Furieri, 2015) that retrieves all of the interconnected pseudogravity worm tables from a PostGIS database and creates the .sqlite file above. Uses the "mod_virtualpg" sqlite extension (found at `https://www.gaia-gis.it/fossil/virtualpg/index`). Included here for reference.

**Files in the Earthquakes folder** These are the recorded seismic events occurring in our study region. They are presented here in multiple forms.

**earthquakes.{shp,dbf,prj,shx}** A collection of shapefile parts containing the locations and other attributes of the earthquake hypocenters used in this work.

**earthquakes.{csv,xlsx}** The same data in both Excel spreadsheet and csv formats.

**SeismicHypocenters.xls** The same information as above in the spreadsheet format requested by the GDR and NGDS.

**Files in the WorldStressMap folder** These are the observations and smoothed results from Heidbach et al. (2008, 2010).

**wsm2008.xls** The authors (Heidbach et al., 2008) request that this primary spreadsheet is only to be distributed from their website. Please obtain the raw spreadsheet from there.

**wsm2008.sql** Data from the raw spreadsheet above have been translated into a PostGIS database for use in this project. This file contains the PostGIS SQL statements to recreate that database.

**wsm2008_smoothed.{csv,zip}** The smoothed $\sigma_1$ ($SH_{max}$) directions from the algorithm described in Heidbach et al. (2010). The authors request that these data too only be downloaded from the Heidbach et al. (2008) website.

**DOEOrientationInStressField.ipynb** A Jupyter (formerly IPython: Pérez and Granger, 2007) notebook. The code for evaluating the smoothing algorithm of Heidbach et al. (2010) at each node of a worm is contained in this notebook. Some standard Python 2.x libraries are required, as well as the interface library between Python and the GIS representation of the worms (found in the git repository: Horowitz and Carpenter, 2016).

**Files in the top-level folder** . These are the LaTeXfiles and their intermediates for generating this README.pdf file.

# References

R. J. Blakely. *Potential Theory in Gravity and Magnetic Applications*. Cambridge University Press, Cambridge, 1996. doi: 10.1017/CBO9780511549816. URL `http://dx.doi.org/10.1017/CBO9780511549816`.

F. Boschetti, P. Hornby, and F. G. Horowitz. Wavelet Based Inversion of Gravity Data. *Exploration Geophysics*, 32,(1):48–55, 2001. doi: 10.1071/EG01048. URL `http://dx.doi.org/10.1071/EG01048`.

Alessandro Furieri. *SpatiaLite website*, 2015. URL `https://www.gaia-gis.it/fossil/libspatialite/index`.

GDAL Development Team. *GDAL - Geospatial Data Abstraction Library, Version 1.11*. Open Source Geospatial Foundation, 2016. URL `http://www.gdal.org`.

Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using NetworkX. In Gäel Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, August 2008. URL `https://conference.scipy.org/proceedings/scipy2008/paper_2/`.

O. Heidbach, M. Tingay, A. Barth, J. Reinecker, D. Kurfeß, and B. Müller. The World Stress Map database release 2008. Online Resource, 2008. URL `http://dx.doi.org/10.1594/GFZ.WSM.Rel2008`.

Oliver Heidbach, Mark Tingay, Andreas Barth, John Reinecker, Daniel Kurfeß, and Birgit Müller. Global crustal stress pattern based on the World Stress Map database release 2008. *Tectonophysics*, 482(1-4):3–15, February 2010. ISSN 00401951. doi: 10.1016/j.tecto.2009.07.023. URL `http://dx.doi.org/10.1016/j.tecto.2009.07.023`.

P. Hornby, F. Boschetti, and F. G. Horowitz. Analysis of potential field data in the wavelet domain. *Geophysical Journal International*, 137(1):175–196, 1999. doi: 10.1046/j.1365-246x.1999.00788.x. URL `http://dx.doi.org/10.1046/j.1365-246x.1999.00788.x`.

P. Hornby, F. G. Horowitz, and F. Boschetti. A physical interpretation of the Poisson wavelet transform of potential fields. In *Proceedings, EGS XXVII General Assembly*, Munich, 2002. European Geophysical Society, European Geophysical Society. URL `http://www.cosis.net/abstracts/EGS02/05568/EGS02-A-05568.pdf`.

Frank Horowitz. Identifying potentially activatable faults for the Appalachian Basin Geothermal Play Fairway Analysis. Internal memo for DOE GPFA project, September 2015. URL `https://gdr.openei.org/files/682/13_GPFA-AB_IdentifyingPotentiallyActivatableFaults.pdf`.

Franklin G. Horowitz and Appalachian Basin GPFA Team. Risk of seismicity from potential direct-use operations in the Appalachian Basin geothermal play fairway project. In *41st Annual Stanford Geothermal Workshop*. Stanford University, February 2016. URL `https://pangea.stanford.edu/ERE/db/GeoConf/papers/SGW/2016/Horowitz.pdf`.

Franklin G. Horowitz and Korin Carpenter. WormDBStuff: 0.1.1. Online Resource, July 2016. URL `http://dx.doi.org/10.5281/zenodo.58309`.

Franklin G. Horowitz and Oliver Gaede. BSDWormer; an open source implementation of a Poisson wavelet multiscale analysis for potential fields. In *2014 Fall Meeting*, number T43C-4743. American Geophysical Union, December 2014. URL `http://abstractsearch.agu.org/meetings/2014/FM/T43C-4743.html`.

Teresa Jordan, Frank Horowitz, Jery Stedinger, Jefferson Tester, Erin Camp, Calvin Whealton, Jared Smith, Brian Anderson, Kelydra Welcker, Xiaoning He, Maria Richards, Cathy C. Pace, Matt Hornbach, Zachary Frone, Christine Ferguson, Rahmi Bolat, and Maria B. Magnani. Low temperature Geothermal Play Fairway Analysis for the Appalachian Basin: Phase 1 report, October 16, 2015. Technical report, US Dept. of Energy; Geothermal Technologies Office, October 2015. URL `http://gdr.openei.org/submissions/682`.

Fernando Pérez and Brian E. Granger. IPython: a system for interactive scientific computing. *Comput. Sci. Eng.*, 9(3):21–29, May 2007. URL `http://ipython.scipy.org`.

D. Ravat, C. Finn, P. Hill, R. Kucks, J. Phillips, R. Blakely, C. Bouligand, T. Sabaka, A. Elshayat, A. Aref, and E. Elawadi. *A Preliminary, Full Spectrum, Magnetic Anomaly Grid of the United States with Improved Long Wavelengths for Studying Continental Dynamics: A Website for Distribution of Data*. Number Open-File Report 20091258. USGS, 2009. URL `http://pubs.usgs.gov/of/2009/1258/`.