

# Geothermal Play Fairway Analysis Code Documentation

Calvin A. Whealton (caw324@cornell.edu)

May 4, 2015

The following document provides supporting information for codes written by Calvin Whealton for the Geothermal Play Fairway Analysis grant. This is not a comprehensive list of all code written for the project. Section 1 is for bottom-hole temperature corrections, section 2 is for outlier identification, and section 3 is for reservoir ideality.

## Bottom-hole Temperature Corrections

These scripts, code, and documents pertain to the bottom-hole temperature correction calculations.

### **func\_BHT\_NY\_PA\_WV\_corr.R**

**Description:** Script with a function to compute bottom-hole temperature corrections for NY, PA, and WV based on the region. Function accepts an R data frame and returns the data frame with two additional columns for the corrected bottom-hole temperature and the error.

Variable	Description
X	R data frame with variables named bht_c: Recorded bottom-hole temperature in Celsius calc_depth_m: Calculated depth of well in meters reg: Region for that point, 0=Rome Trough and areas south east in PA, 1=Allegheny Plateau (NY and north east PA), 2=West VA, and 3=modified Harrison Correction (used in previous studies of the area)

The output is a data frame with the following columns added.

Output	Description
corr_bht_c	Corrected bottom-hole temperature in Celsius
corr_error	Error from calculation of correcting the bottom-hole temperature 0: no errors 20: depth outside of normal range 21: depth is negative 22: depth is missing for Allegheny Plateau or West Virginia data 30: categorical variable not 0,1, 2, or 3 32: categorical variable missing 42: bottom-hole missing

**Equations:** The equations used to calculate the temperature correction  $\Delta T$  are given below. Depth ( $z_0$ ) is in meters and temperature correction ( $\Delta T$ ) is in Celsius. See Frone and Blackwell (2010) for details on the modified Harrison correction. If no region is specified no temperature correction is computed and an error is printed in the output.

$$\Delta T_{Alle.Plat.} = \begin{cases} \max\{0, -23.48 + (1.791 \times 10^{-2})z_0\}, & (z_0 < 4000\text{m}) \\ 48.16, & (z_0 > 4000\text{m}) \end{cases} \quad (1)$$

$$\Delta T_{RomeTr.} = 0 \quad (2)$$

$$\Delta T_{WVa.} = \min\{15, -3.562 + 0.00763z_0\} \quad (3)$$

$$\Delta T_{Mod.Harr.} = \begin{cases} 0, & (z_0 < 1000\text{m}) \\ -16.51 + (1.83 \times 10^{-2})z_0 - (2.34 \times 10^{-6})z_0^2, & (1000\text{m} < z_0 < 3860\text{m}) \\ 19.07, & (z_0 > 3860\text{m}) \end{cases} \quad (4)$$

## example\_bht\_corr.R

**Description:** Script to run the function `func_BHT_NY_PA_WV_corr.R` with synthetic data in `bht_test_data.csv`. The output should be corrected BHTs and cases that generate all errors for missing values or values outside typical ranges.

### **bht\_test\_data.csv**

**Description:** Synthetic data to test func\_BHT\_NY\_PA\_WV\_corr.R for proper corrections and generating all possible errors.

# Outliers

These codes, scripts, and files pertain to the outlier identification procedures.

## **outlier\_identification.R**

**Description:** Script with several functions to compute outliers using one of the specified algorithms with the specified inputs. Most algorithms have separate functions described below.

Function	Description
outlier_iden	General function to call other functions and perform outlier identification
outlier_loc_pts	Local outlier identification for <b>algo=1</b>
outlier_loc_rad	Local outlier identification for <b>algo=2</b>
outlier_loc_grid	Local outlier identification for <b>algo=3</b>
outlier_glob	Global outlier identification

Variable	Description
<b>X</b>	R data frame with variables named <b>x.coord</b> : Longitude coordinate in km <b>y.coord</b> : Latitude coordinate in km <b>test</b> : Variable to be tested for being an outlier
<b>algo</b>	algorithm for determining local outlier 1: (Default) Finds nearest points 2: Finding all points within a given radius 3: Gridding data
<b>outcri</b>	Outlier criteria 1: (Default) Only local outliers flagged as outliers 2: Only local and global outliers flagged as outliers 3: Only global outliers flagged as outliers
<b>pt_eval</b>	Number of points used in when <b>algo</b> =1 (default = 25)
<b>rad_eval</b>	Radius (in km) at which to take points when <b>algo</b> =2 (default = 16)
<b>box_size</b>	Size of spacing (in km) to form grids when <b>algo</b> =3 (default = 32)
<b>pt_min</b>	Minimum number of points required to perform local test for <b>algo</b> =2 or 3 (default = 25)
<b>rad_max</b>	Maximum radius (in km) at which to take points when <b>algo</b> =1 (default = 16)
<b>k_glob</b>	Constant multiplied by the upper- and lower-half quartile ranges in global analysis (default = 3)
<b>k_loc</b>	Constant multiplied by the upper- and lower-half quartile ranges in local analysis (default = 3)
<b>type</b>	Type of quantile estimation (default = 7, see R documentation)

The output is a data frame with the following columns added. The local outlier columns will only be added when **outcri**=1 or 2. The global outlier columns will only be added when **outcri**=2 or 3. The **outs** column will be present in all output.

Output	Description
outs	Binary variable for points being an outlier (1=outlier)
out_loc_lo	Binary variable for point being a local low outlier (1=outlier))
out_loc_hi	Binary variable for point being a local high outlier (1=outlier)
out_loc_lq	Lower quartile for local outlier test (NA if not tested)
out_loc_mq	Median for local outlier test (NA if not tested)
out_loc_uq	Upper for local outlier test (NA if not tested)
out_loc_lb	Lower bound for local outlier test (NA if not tested)
out_loc_ub	Upper bound for local outlier test (NA if not tested)
out_loc_rad	Maximum distance to point (only for algo=1)
out_loc_pts	Number of points in local area (only for algo=2 and 3)
out_loc_error	Error in local outlier calculation 0: No errors 1: Some points outside <code>rad_max</code> when <code>algo=1</code> 2: Fewer than <code>pt_min</code> points in region when <code>algo=2</code> or <code>3</code> )
out_glob_lo	Binary variable for points being an global low outlier (1=outlier)
out_glob_hi	Binary variable for points being an global high outlier (1=outlier)

**Equations:** The equations used to calculate low and high outlier bounds are given below. In these equations  $q$  is a variable of interest (`test` in the data) with subscripts denoting quantiles and  $k$  is the constant (`k_loc` or `k_glob`). The equations can be applied locally or globally. Aguirre (2014) uses an outlier test that is similar, but the version implemented in this code is more flexible.

$$B_{lower} = q_{0.25} - k(q_{0.5} - q_{0.25}) \quad (5)$$

$$B_{upper} = q_{0.75} + k(q_{0.75} - q_{0.5}) \quad (6)$$

## example\_outlier\_code.R

**Description:** Script to run outlier identification functions with test data and the Cornell dataset. Later portions do not need to be run because they were testing sensitivity of the algorithm to the input parameters. They are kept in the code for potential future analysis.

## out\_test\_grid.csv

**Description:** Synthetic data to test the local outlier identification algorithm that uses gridding.

**out\_test\_rad.csv**

**Description:** Synthetic data to test the local outlier identification algorithm that uses maximum radius.

**out\_test\_pt.csv**

**Description:** Synthetic data to test the local outlier identification algorithm that uses number of points.

**cornell\_data.csv**

**Description:** Cornell heat flow database with 8,919 points used to test sensitivity of the outlier identification algorithm. See Cornell University (2014).

# Reservoir Ideality

## MainIdeality.m

**Description:** Script that runs the reservoir ideality uncertainty analysis. Subsidiary functions are called from this script. This script also imports an example dataset and uncertainty mapping from `TestFormationData.csv` and `TestUncertaintyLevels.csv`, respectively. This script can be modified to include more graphs and statistical analysis.

## GenRandNums.m

**Description:** Function to generate random numbers from uniform, triangular, normal, or lognormal distributions.

Variable	Description
mean	Mean value of the distribution (real-space)
unc	uncertainty (spread) of the distribution as a percentage of the mean uniform: bounds are defined from uncertainty triangular: bounds are defined from uncertainty normal: 95% central region of distribution defined by uncertainty lognormal: real-space coefficient of variation defined by uncertainty
dist	Distribution selected 1: Non-standard uniform distribution 2: Triangular distribution, symmetric 3: Normal distribution 4: Lognormal distribution
reps	Number of replicates to generate

The output is a column vector with of numbers for the specified distribution.

## MonteCarloApprox.m

**Description:** Function to calculate Monte Carlo Approximation of the distribution of reservoir ideality.



Variable	Description
mat	Matrix of random values for reservoir ideality calculation column 1: $k$ , permeability/conductivity column 2: $H$ , thickness column 3: $P_o$ , pressure column 4: $P_b$ , pressure column 5: $R_o$ , radius
mu	Viscosity
Rb	Casing radius
ideality	ideality metric 1: $(2\pi/\mu)kH(P_o - P_b)/(\ln(R_b) - \ln(R_o))$ 2: not currently used
reps	Number of replicates to generate

Note that in the code  $P_{diff} = -|P_o - P_b|$  is used. The difference between  $P_o$  and  $P_b$  should be large so that this is generally not necessary, but it ensures that there are no problems. The output is a column vector with the distribution of the reservoir ideality.

### **TestFormationData.csv**

**Description:** Example file for input of formation data. The version of MATLAB used to develop this could not handle strings and numbers, so all input converted to numbers. The header line is dropped when reading-in the file.

### **TestUncertaintyLevels.csv**

**Description:** Example file for mapping uncertainty levels in **TestFormationData.csv** to percentage uncertainty. Column 1 is the uncertainty mapping (1-5). Columns 2-6 are the percentage uncertainty associated with that level for  $k$ ,  $H$ ,  $P_o$ ,  $P_b$ , and  $R_o$ , respectively.

## References

Aguirre, G. A. (2014). *Geothermal Resource Assessment: A Case Study of Spatial Variability and Uncertainty Analysis for the States of New York and Pennsylvania*. Master's Thesis, Environmental and Water Resources Systems Engineering, School of Civil and Environmental Engineering, Cornell University.

Cornell University (2014). Cornell University Heat Flow Database (NY and PA). Southern Methodist University Geothermal Laboratory. (Accessed 16 June 2014) [geothermal.smu.edu/static/DownloadFilesButtonPage.htm](http://geothermal.smu.edu/static/DownloadFilesButtonPage.htm)

Frone, Z. and Blackwell, D. (2010). Geothermal Map of the Northeastern United States and the West Virginia Thermal Anomaly. *Geothermal Resources Council Transactions* 34:339-344.