Geothermal Play Fairway Analysis Codes Documentation

Calvin A. Whealton (caw324@cornell.edu)

March 18, 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 codes written for the project.

Bottom-hole Temperature Corrections

These scripts, codes, 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 a 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=modifed 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 ΔT are given below. Depth (z_0) is in meters and temperature correction (ΔT) 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.

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

$$\Delta T_{RomeTr.} = 0 \tag{2}$$

$$\Delta T_{WVa.} = \max\{15, -3.562 + 0.00763z_0\} \tag{3}$$

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

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 based 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 algo=1
outlier_loc_rad	Local outlier identification for algo=2
outlier_loc_grid	Local outlier identification for algo=3
outlier_glob	Global outlier identification

Variable	Description
X	R data frame with variables named x_coord: Longitude coordinate in km y_coord: Latitude coordinate in km test: Variable to be tested for being an outlier
algo	algorithm for determining local outlier 1: (Default) Finds nearest points 2: Finding all points within a given radius 3: Gridding data
outcri	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
pt_eval	Number of points used in when algo=1 (default = 25)
rad_eval	Radius (in km) at which to take points when algo=2 (default = 16)
box_size	Size of spacing (in km) to form grids when algo=3 (default = 32)
pt_min	Minimum number of points required to perform local test for algo=2 or 3 (default = 25)
rad_max	Maximum radius (in km) at which to take points when algo=1 (default = 16)
k_glob	Constant multiplied by the upper- and lower-half quartile ranges in global analysis (default = 3)
k_log	Constant multiplied by the upper- and lower-half quartile ranges in local analysis (default = 3)
type	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 rad_max when algo=1) 2: Fewer than pt_min points in region when algo=2 or 3)
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.

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

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

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).

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

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.