PROCEEDINGS, Thirty-Seventh Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, January 30 - February 1, 2012
SGP-TR-194

# FURTHER STATISTICAL ANALYSIS OF DIRECT SUBSOIL TEMPERATURE DATA (INCLUDING RESIDUALS ANALISIS) AT THE PAILAS GEOTHERMAL FIELD AND IN THE RINCÓN DE LA VIEJA NATIONAL PARK, GUANACASTE, COSTA RICA

Hakanson, Edward Charles

Instituto Costarricense de Electricidad
Centro de Servicio Recursos Geotérmicos
Guayabo de Bagaces
Guanacaste, Costa Rica
e-mail: edward_hakanson@yahoo.com; Echarles@ice.go.cr

## ABSTRACT

From June 2008 to September 2009, a field investigation for a Master's thesis in Natural Resource Management was carried out to identify the existence of subsoil thermal anomalies that may be correlated with anomalous heat flow along local faults at the Pailas Geothermal Field, located on the Pacific slope of the Quaternary Rincón de la Vieja volcanic complex in northwestern Costa Rica. A total of 240 direct subsoil temperature measurements were made at fifty-six one-inch diameter access holes equally spaced on a one-kilometer grid and drilled to depths ranging from 47 to 167 cm in iso-hyperthermal entisols and thermal to iso-hyperthermal inceptisols. This study covers the area that is directly impacted by the operation of Pailas 1, a 35 MWe binary rankine-cycle geothermal plant inaugurated in July 2011, part of the north bordering Rincón de la Vieja National Park and an area extending towards the northwest of the currently developed geothermal field. Mean annual subsoil temperatures have been reduced to their respective Z values for standard deviation analysis and residuals from multiple regression analyses on data collected at the end of the dry season ($\alpha$=0.05) are studied to try to pinpoint outliers that significantly deviate from the estimated mean value and that may represent a deep-seated geothermal anomaly. There appears to be a north-south trend of significantly higher residuals at higher elevations and an east-west trend at lower elevations.

## INTRODUCTION

The Pailas Geothermal Field is located on the southern slope of the Quaternary Rincón de la Vieja volcanic complex in northwestern Costa Rica and since July 2011 has continuosly produced sufficient heat flow to run 35 MWe Binary Rankine bottoming cycle geothermal plant as baseload power. In partial fulfillment of the requirements for a Master's of Science degree in Natural Resource Management with an environmental emphasis, field work was done at the Pailas Geothermal Field and in surrounding areas, including within the north-bounding Rincón de la Vieja National Park and a Non-Governmental Agency owned Property to the west called Mundo Nuevo to try to understand how shallow temperature measurements may be used in geothermal prospection at that field and minimize the overall environmental impact.

## LOCATION

Costa Rica is in Central America and was formed by the interaction of the subducting oceanic Cocos and overriding contintental Caribbean Plates which meet at the northwest trending mid-american trench off the Pacific coast. The field location is in northwestern Costa Rica approximately 250 km de la capital, San José Costa Rica on the southern slope of the Rincón de la Vieja andesitic volcanic complex which is composed of at least nine cones trending northwest and west. A lahar deposit caps the lithologic sequence locally covering the andesitic lavas near the Rincón de la Vieja volcanic complex and a Pleistocene to Quaternary beige pumiceous tuff further out. Thermal manifestations like hot springs, boiling mud pots, steaming ground and advanced argillic alteration (Las Pailas, Las Hornillas and Los Azufrales) are found locally at the base of the volcanic slope where the piedmont marks a change in the topography and may be related to local faults which have allowed the ascent of heat and hot water towards the surface. The mean annual ambient temperature (MAAT) oscilates between 24 and 30° year round and the mean annual subsoil temperature (MAST) oscilates between 19.2 and 27.7°C (Hakanson, 2011). The type of forest cover ranges between Basal to Wet Forest and Very Wet Pre-mountain Forest according to the classification by Holdridge.

## LOCAL GEOLOGY

The lithoestratigraphic column is composed of seven units: 1) Lahar (DARV); 2) Andesitic Lava (LARV); 3) Pitál Formation (PF); 4) Liberia Formation (LF); 5) Domes Unit (DU); 6) Bagaces Group (BG); and 7) Aguacate Group (AG). These units are only mentioned here and shown in Figure 1. For a more complete description of the local geology and regional geology the reader is referred to Chavarría et al (2010).

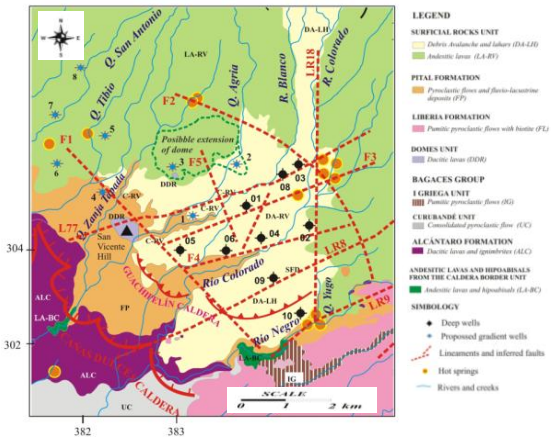

*Figure 1: Local geologic map (Taken from Chavarría, et al. 2010).*

## SOIL TYPES

The soils in the field area are classified within the Orders of Entisols and Inceptisols, specifically in the Great Groups Ustorthents and Dystrandepts, respectively (Figure 2). The temperature regime of these soils as determined in this investigation ranges from thermal to iso-hyperthermal. The Ustorthents are found at lower elevations overlying the pyroclastic flows of the Pital Formation, the Liberia Formation, dacitic ignimbrites of the Alcántaro Formation and the Curubandé Unit while Inceptisols are found elsewhere in the field area overlying more recent deposits.

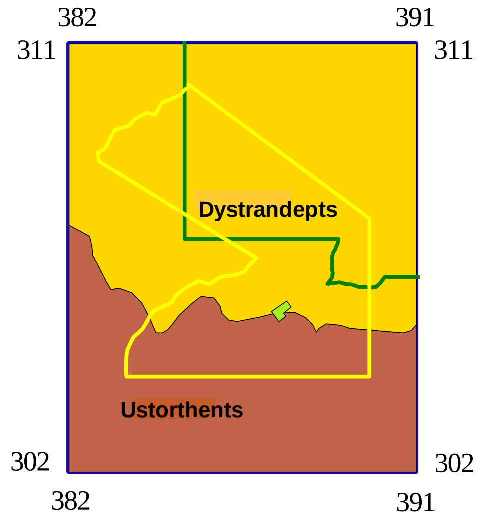


*Figure 2: Soil types found in the field area. Source: Atlas Costa Rica 2008.*

## FIELD DATA COLLECTED

In all 234 subsoil temperature measurements were taken over a 15 month period at 56 observation holes drilled to depths ranging from 26 to 163 cm in open pastures and forested areas spread over an area of approximately 30 Ha and equally distanced one kilometer between the elevations of 420 to 1290 meters above sea level. Measured subsoil temperatures ranged from 15 to 29°C and the mean average was 23.5°C.

## STATISTICAL ANALYSIS OF FIELD DATA

Statistical analysis of the field data was done to test the null and alternate hypotheses that anomalously high subsoil temperatures may be detectable with shallow temperature measurements at the Pailas Geothermal Field where the alternative hypothesis is that anomalously high subsoil temperatures in the area (with a confidence level of 95%) are detectable with shallow temperature measurements while the null hypothesis is that anomalously high subsoil temperatures are not detectable with shallow temperature measurements.

$$H_0\colon \mu_1 \leq \mu_0 + \sigma_{1.96} \qquad \text{Eq. 1}$$
$$H_1\colon \mu_1 \leq \mu_0 + \sigma_{1.96} \qquad \text{Eq. 2}$$

Where:
$H_0$: Null hypothesis
$H_1$: Alternative hypothesis
$\mu_1$: Mean annual subsoil temperature at a specific access hole
$\mu_0$: Mean annual subsoil temperature for the entire population of measurements
$\sigma_{1.96}$: 1.96 standard deviations from the population mean ($\mu_0$).

Any mean values that are greater than 1.96 standard deviations from the average mean would be considered as anomalous and possibly related to a deep heat source.

Microsoft Excel was used to do the statistical analysis of subsoil temperature data and is shown in the following sections. First a descriptive analysis of the data was done both on the entire population of the data and then on the MAST determined at each of the 56 access holes (Table 1).

*Table 1: Descriptive statistical analysis on subsoil temperature data obtained on the southern slope of the Rincón de la Vieja volcanic complex.*

| $T_{subsoil}$ | | *MAST* | |
|---|---|---|---|
| Average | 23,59017094 | Average | 23,51267857 |
| Standard Error | 0,156144547 | Standard Error | 0,237735796 |
| Median | 23,8 | Median | 23,5375 |
| Mode | 24,6 | Mode | 21,8 |
| Standard Deviation | 2,38855227 | Standard Deviation | 1,779051798 |
| Sample Variance | 5,705181945 | Sample Variance | 3,165025299 |
| Kurtosis | 2,531876333 | Kurtosis | 0,052191388 |
| Skewness | - | Skewness | 0,076343567 |
| Range | 15,6 | Range | 8,475 |
| Minimum | 13,7 | Minimum | 19,225 |
| Maximum | 29,3 | Maximum | 27,7 |
| Sum | 5520,1 | Sum | 1316,71 |
| Count | 234 | Count | 56 |
| Confidence Level(95,0%) | 0,307635602 | Confidence Level(95,0%) | 0,476433179 |

To normalize the data to units of standard deviation, Z values were calculated by subtracting the mean value from the measured value and dividing by the standard deviation as in Equation 3.

$$Z= (x-\mu)/\sigma \qquad \text{Eq. 3}$$

Where:
Z: Z value
x: $\mu_1$
μ: $\mu_0$
σ: Standard deviation

Z values of the entire population of temperature measurements (n = 234) ranged from -4.14 to 2.39 while those of the MAST (n = 56) ranged from -2.41 to 2.35 (three of which are higher than 1.96). The lower minimum from the entire population is due to low temperatures measured specifically at the end of the rainy season and the onset of the dry season. Since the analysis was done on a site specific basis and not a season specific basis, only the MAST values are considered here.

A linear regression of the data (MAST) was done using elevation as the independent variable (x axis) and temperature as the dependent variable (y axis) which showed a 0.6 to 0.7 °C decrease in temperature with each 100 meter rise in elevation for a moderate to considerable negative correlation with a correlation factor of R = -0.52 to -0.79, $R^2$ of 0.24 a 0.62 and p values between 4.14 x 10-13 and 3.63 x 10-5 which shows that this negative correlation is significant at a 95% confidence level (α = 0.05). Figure 3 shows a the linear regression done on calculated Z values.

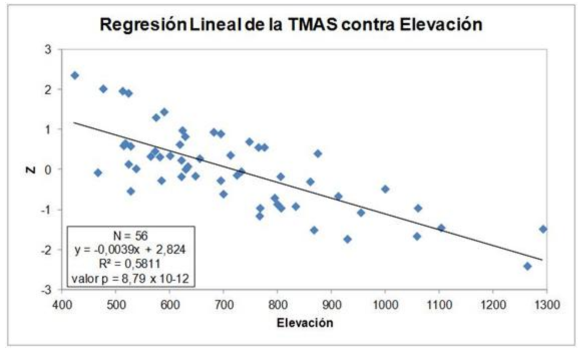


*Figure 3: Linear regression of Z values of the mean annual subsoil temperatures calculated from field data at 56 access holes. Notice that three values are greater than 1.96.*

To account for the annual heat wave which may penetrate to a depth of 20 meters a time series analysis of the data was done which shows the coolest subsoil temperatures at the end of the rainy season, increasing during the dry season until they reach maximum values at onset of the rainy season (Figure 4).

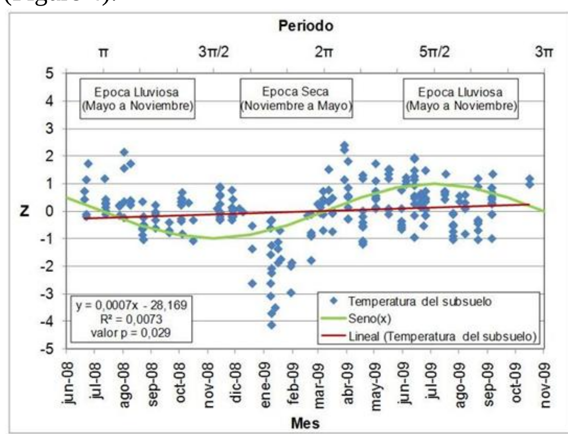


*Figure 4: Time series plot of Z values calculated from data using average daily subsoil temperatures during the sampling period. Note the low temperatures at the end of the rainy season and the high temperatures at the end of the dry season.*

Parametric analyses including covariance of nominal variables (geographical coordinates, elevation, depth, MAST, MAAT, thermal gradient (dT/dz) from 20 cm, and topographic gradient) as well as variance of ordinal variables (elevation, land use, geological substrate, topographic aspect and subsoil texture) were done (Tables 2 and 3).

*Table 2: Covariance of nominal variables. Note the strong positive relationship between elevation and topographic gradient and the negative correlation between elevation and MAST.*

| | Easting | Northing | Elevation | Depth (cm) | MAST | MAAT | dT/dz (from 20 cm) | Topographic Gradient |
|---|---|---|---|---|---|---|---|---|
| Easting | 4,55E+06 | | | | | | | |
| Northing | -1,48E+06 | 6,17E+06 | | | | | | |
| Elevation | 4,94E+04 | 4,03E+05 | 3,69E+04 | | | | | |
| Depth (cm) | -2,73E+03 | 1,77E+04 | 1,67E+03 | 1,86E+03 | | | | |
| MAST | -1,02E+03 | -2,67E+03 | -2,58E+02 | -8,17E+00 | 3,11E+00 | | | |
| MAAT | -8,30E+02 | -2,52E+03 | -2,16E+02 | -2,00E+00 | 2,49E+00 | 4,12E+00 | | |
| dT/dz (from 20 cm) | -3,94E+00 | -6,10E+00 | -6,62E-01 | -8,10E-02 | 8,77E-03 | 1,09E-02 | 2,18E-04 | |
| Topographic gradient | 2,70E+03 | 6,16E+03 | 7,44E+02 | 2,92E+01 | -8,41E+00 | -9,23E+00 | 1,12E-02 | 2,51E+02 |

*Table 3: ANOVA analysis of subsoil temperatures within 100 meter elevation belts.*

| Groups | Count | Sum | Average | Variance |
|---|---|---|---|---|
| 400 - 500 m | 3 | 78,18 | 26,06 | 5,49 |
| 500 - 600 m | 15 | 369,10 | 24,61 | 1,72 |
| 600 - 700 m | 13 | 312,10 | 24,01 | 0,88 |
| 700 - 800 m | 10 | 232,06 | 23,21 | 1,57 |
| 800 - 900 m | 6 | 134,90 | 22,48 | 1,48 |
| 900 - 1000 m | 4 | 87,00 | 21,75 | 0,97 |
| 1000 - 1100 m | 2 | 42,35 | 21,18 | 0,78 |
| 1100 - 1200 m | 1 | 20,93 | 20,93 | - |
| 1200 - 1300 m | 2 | 40,10 | 20,05 | 1,36 |

| Origin of Variance | Sum of Squares | Degrees of Freedom | Root Mean Square | F | p value | Critical value for F |
|---|---|---|---|---|---|---|
| Entre grupos | 101,91 | 8 | 12,74 | 8,30 | 6,24E-07 | 2,14 |
| Dentro de los grupos | 72,17 | 47 | 1,54 | | | |
| Total | 174,08 | 55 | | | | |

The ANOVA analysis shows that the ordinal environmental variables which have a significant control on subsoil temperature (p-value < 0.05 < critical value for F < calculated value for F) in descending order are: 1) Topographic gradient; 2) Land use; 3) Elevation; and 4) Geologic substrate. However, as can be seen in the covariance, the topographic gradient has a strong positive correlation with elevation.

To complement the ANOVA analysis, a residuals analysis was carried out from a multiple regression done on the data by plotting measured subsoil temperatures according the ordinal variables: land use, geological substrate, topographic aspect, topographic gradient and soil texture using elevation as the independent variable (x axis) and temperature as the dependant variable (y axis). The data used is from the measurement period corresponding to the end of the dry season since it is expected that these temperatures were less influenced by the percolation of rain waters (Figure 5). The corresponding linear regression of estimated subsoil temperatures at the end of the dry season is y = -0.0067x + 29.035 with an $R^2$ of 0.4525 and a p-value of $1.35 \times 10^{-5}$, which exemplifies the significance of the data.

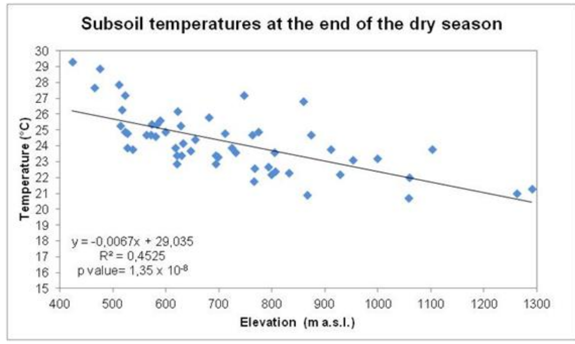


*Figure 5: Linear regression of subsoil temperatures obtained at the end of the dry season. The low p value shows the high significance of the data (>95% confidence level).*

Using Equation 3, Z values were calculate for the subsoil temperatures obtained at the end of the dry season (Figure 6) which were in turn analyzed by standardized residuals (Figure 7) with respect to the linear regression of Z values (y = 0.0037x + 3.1043).

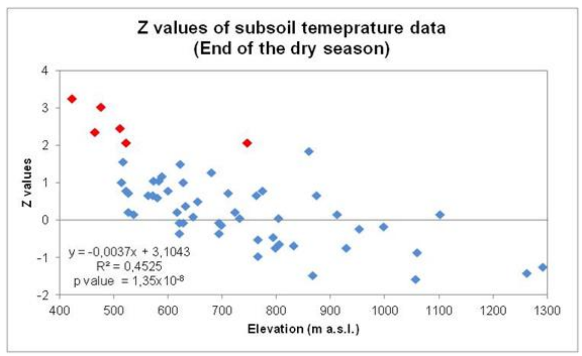


*Figure 6: Z values of the subsoil temperatures obtained at the end of the dry seson (n = 56). Points in red are greater than 1.96.*

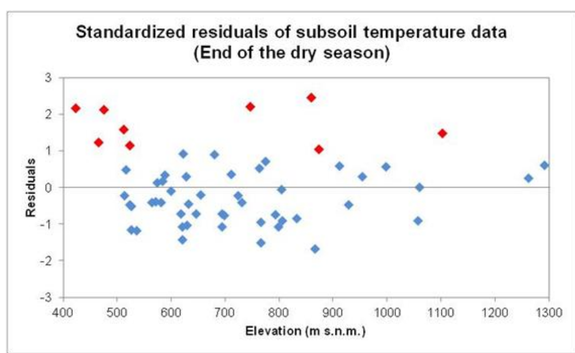


*Figure 7: Standardized residuals of the subsoil temperatures obtained at the end of the dry seson (n = 56). Points in red are greater than 1.*

The corresponding standardized residuals of Z values ranked greater than 1.96 are all above 1 standard residual. For that reason all other standardized residuals that exceed 1 residual are also considered as being possibly anomalous. These high standardized residuals (>1) were found between 400-500 meters, 750-900 and at just over 1100 meters above sea level, some of which coincide with Z values above 1.96.

Furthermore, plotting these data as per ordinal variable shows that in accordance with the ANOVA analysis land use (vegetative cover) and geologic substrate may have a significant effect on the subsoil temperature, however it is still necessary to convert these ordinal variables into their nominal equivalents (% of land cover) in order to do the respective covariance analysis with elevation.

There are still three measurement points that have standardized residuals greater than 1 (at 859, 873 and 1102 m a.s.l.) and although they have insignificant Z values, with further investigation they may prove to actually be hot spots. For this reason, these points should not be excluded from future deeper drilling.

## CONCLUSIONS

The mean annual subsoil temperatures on the southern slope of the Rincón de la Vieja volcanic complex decrease approximately 0.6 to 0.7 °C with each 100 meter rise in elevation.

The subsoil temperatures in the field area follow a sinusoidal pattern that corresponds to the annual heat wave.

The null hypothesis of this investigation must be rejected since it was possible to highlight areas with significantly anomalous subsoil temperatures by way of Z value and residuals analysis. These areas correspond to iso-hyperthermal Dystrandepts at lower elevations overlying the pumitic pyroclastic flows of the Pital Formation and dacitic ignimbrites of the Alcántaro Formation.

The main controls on subsoil temperature at the Pailas Geothermal Field and in adjacent properties are elevation, land use and the geologic substrate.

Three points with higher residuals but insignificant Z values are located higher up on the andesitic lava flow covered slopes of Rincón de la Vieja where the possible presence of a shallow cold aquifer may partially mask the thermal anomaly.

Similar statistical analysis of geoscientific data (Descriptive statistics, Z value calculation, Time series data, Covariance, ANOVA and Residuals analysis) is encouraged in other geothermal fields.

Although the expression of significantly high temperature Z values in test holes of 1.5 meters was not evident at higher elevations in the Pailas Geothermal Field and over Quaternary andesitic lava flows, the higher residuals suggest that possibly the humid climatic conditions and the probable existence of cold shallow aquifers in these areas inhibit high heat flow from reaching the surface except for at known thermal manifestations like Las Pailas, Las Hornillas and Los Azufrales for which reason deeper holes may be necessary to achieve significantly high Z values in these areas.

## REFERENCES


Atlas Costa Rica 2008

Chavarría L., Mora O., Hakanson E., Gálvez M., Rojas M., Molina F. & Murillo A. (2010) Geologic Model of the Pailas Geothermal Field, Guanacaste, Costa Rica. Proceedings World Geothermal Congress 2010. Bali, Indonesia, 25-29 April 2010.

Hakanson E. 2011. Statistical Analysis Of Direct Subsoil Temperature Data Obtained At And Around The Las Pailas Geothermal Project And In Rincón De La Vieja National Park, Guanacaste, Costa Rica. PROCEEDINGS, Thirty-Sixth Workshop on Geothermal Reservoir Engineering. Stanford University, Stanford, California, January 31 - February 2, 2011.