Phenomena exposure from the large scale gas injection Test (Lasgit) dataset using a bespoke data analysis toolkit.

Bennett DP[1], Cuss RJ[2], Vardon PJ[1,3], Harrington JF[2], Thomas HR[1]

1. Geoenvironmental Research Centre (GRC), Cardiff School of Engineering, Cardiff University, Queen's Buildings, The Parade, Cardiff, CF24 3AA, United Kingdom. BennettDP@cardiff.ac.uk.

2. Transport Properties Research Laboratory, British Geological Survey (BGS), Keyworth, Nottingham, NG12 5GG, United Kingdom.

3. Geo-Engineering Section, Department of Geoscience and Engineering, Delft University of Technology, PO Box 5048, 2600 GA Delft, The Netherlands.

Abstract

The Large Scale Gas Injection Test (Lasgit) is a field scale experiment designed to study the impact of gas build up and subsequent migration through an Engineered Barrier System (EBS). Lasgit has a substantial experimental dataset containing in excess of 21 million datum points. The dataset is anticipated to contain a wealth of information, ranging from long-term trends and system behaviours to small scale or 'second order' features. In order to interrogate the Lasgit dataset a bespoke computational toolkit, designed to expose difficult to observe phenomena, has been developed and applied to the dataset. The preliminary application of the toolkit, presented here, has resulted in a large number of phenomena being indicated/quantified, including: highlighting of second order events (small gas flows, perturbations in stress/pore-water sensors etc.); and quantification of temperature record frequency content. Localised system behaviour has been shown to occur along with systematic aberrant behaviours that remain unexplained.

1. Introduction

The Large Scale Gas Injection Test (Lasgit) is a field scale experiment operated by the British Geological Survey (BGS) located at approximately 420m depth in Svensk Kärnbränslehantering AB's (SKB's) Äspö Hard Rock Laboratory (HRL) in Sweden. Lasgit has been designed to study the impact of gas build up and subsequent migration through the Engineered Barrier System (EBS) of the Swedish KBS-3 disposal concept for high level radioactive waste (Cuss et al. 2011; SKB 2006). Specific experimental aims include improving understanding of the gas entry and flow mechanism into and through the EBS, and its effect on buffer saturation levels. Modifications to the KBS-3 concept for the Lasgit experiment include: sintered filters varying in size and location on the canister surface for the purpose of gas injection, simulating a canister defect and point source of gas; filter mats in and around the bentonite buffer material, used to accelerate the hydration of the system; a retained lid to simulate backfill of the repository; and a high number of sensors within the bentonite

buffer material and at interfaces between the buffer and the canister/host rock (Cuss et al. 2010). Figure 1 depicts Lasgit's experimental arrangement, illustrating a canister emplaced in a deposition hole, the retaining structure simulating repository backfill, the hydration mats, and the approximate distribution of sensors within the deposition hole.

Lasgit has been in continuous operation since February 2005, and is considered in the analysis presented here from 2007 to mid 2012.[1] Each of the circa 175 instruments have been recorded with a high frequency in comparison to the physical processes being measured. Since 2007, there exists approximately 120,000 records per sensor, averaging 2.5 logs/hour. This has led to a substantial dataset containing in excess of 21 million datum points, approximately 18 million of which originate from sensors within and around the deposition hole, and the remainder being associated with the experimental control systems. The dataset is anticipated to contain a wealth of information, ranging from long-term trends and system behaviours to small scale or 'second order' features.

Instrumentation within the deposition hole consists of 32 total stress sensors (oriented either radially or axially), 26 pore-water pressure sensors, and 56 temperature sensors. Additionally there are 16 pore-water pressure transducers associated with the hydration mats and injection filters that act as sensors. Attached to the deposition hole retaining set-up there are 10 displacement sensors. Nine pore-water pressure sensors are located in boreholes adjacent to the deposition hole. The ambient air temperature of the facility at the experimental location is also monitored.

The experimental history has consisted of an initial period of natural and artificial hydration of the EBS lasting approximately two years, followed by the first gas injection test phase (during which major gas breakthrough did not occur). Hydration of the EBS was then continued for approximately one more year. Two more gas injection test phases, each lasting in excess of a year, were then performed back to back. Major gas breakthrough occurred during both of these gas test phases. Overviews of these major phases of the experimental history and the prominent events that occurred during each phase are reported in Harrington et al. (2008); Cuss et al. (2010); and Cuss et al. (2011). In addition to the small scale phenomena associated with changes in flow regime that are considered in the analyses presented in these reports a large number of second order perturbations remain uninvestigated.

Systematic computational analysis of the Lasgit dataset implicitly reduces the subjectivity of any observations made. Additionally, computational analysis is required due to the size of the dataset and the scale of the details of interest. Moreover, second order phenomena are able to be exposed. However, due to the length and complexity of the experiment, the Lasgit dataset is not typically suited to 'out of the box' computational analysis. Specifically, non-uniformity in the sampling rate (arising both intentionally and unavoidably in the case of the Lasgit dataset, see Figure 2) is incompatible with

---

[1] This is due to the first two years of data possessing significant distortions and anomalies resulting from necessary operator interventions, such as the drilling and subsequent sealing of pressure relief holes (Harrington et al. 2008; Cuss et al. 2010).

any time series analysis algorithm or analysis technique that assumes a uniform sample rate. Such assumptions, while not ubiquitous, are typical in time series analysis (e.g. Box and Jenkins 1976). Applying such an algorithm to non-uniform data may result in a distortion of the physical meaning of the input. This is illustrated in Figure 3, where non-uniformly collected data is shown with a uniform spacing as significantly distorted. Solutions such as 'down-sampling' the dataset (i.e. using only the points coinciding with a lower common sample rate in the dataset) or interpolating between measured datum points to produce a uniform sample rate are possible. However, both potentially risk the loss of detail in the dataset (particularly at local maxima and minima), or, in the case of down-sampling, may not be possible if a common sample sub-rate is not present across the dataset.

2. Bespoke toolkit

In order to accommodate the non-uniform nature of the Lasgit dataset when performing an Exploratory Data Analysis (EDA), a bespoke computational toolkit has been developed (Bennett et al. 2012). The toolkit implementation, in addition to avoiding the requirement for uniform input data, makes minimal assumptions and requires minimal knowledge about the nature of the input. This ensures that the applicability of the toolkit to long-term, large-scale experimental datasets is as broad as possible. Key toolkit capabilities (depicted in Figure 4 and further elaborated in Bennett et al. 2012) include:

- Non-parametric trend identification using Singular Spectrum Analysis (SSA) (e.g. Golyandina et al. 2001).
- Statistical spike identification for isolated/individual aberrant/anomalous datum points.
- Event candidate identification, sensitive to second order events.
- Frequency domain analysis utilising a Discrete Fourier Transform (DFT) (Bagchi and Mitra 1999) modified to accommodate non-uniform input data.
- Smoothing and averaging functions that utilise moving windows defined as periods of time rather than a fixed number of points around a focus (time windowing).
- Visualisation technique for identifying where and when event candidates occur across multiple sensors/time series simultaneously.

The capabilities included in the toolkit were selected for their abilities to expose phenomena, i.e. extract useful information from a series of measurements that would otherwise be obscured or occluded when observed. For example: trend identification and subsequent removal can expose second order events that may otherwise be subsumed within macro-scale system behaviours. This would also facilitate quantification of observable seasonal effects using frequency domain analysis; spike detection, smoothing and averaging functions allow noise to be reduced, improving the visibility and measurability of observable phenomenon; and visualisation of simultaneous event candidates

allows quick aggregation of the relevant information surrounding a second order event in a large and complex system.

Implementation of each tool limits operator defined parameters to statistical thresholds and time window length scales, and where appropriate generalises the data analysis algorithms to accommodate non-uniform input. An example of such a generalisation (the modification to the DFT) is shown in equations (1a) and (1b). The DFT is a common and well understood method for accessing frequency domain information, and is simple to implement computationally. The modification described here pertains to the specification of frequency range inspected, with the term $2\pi kn/N$ being replaced by ωt_n allowing arbitrary frequencies to be investigated and accurate time weighting of the input signal to be achieved. The modified Non-uniform Discrete Fourier Transform (NDFT) no longer relies on the input time series possessing a uniform sample rate. The terms of equations (1a) and (1b) are defined in Table 1. Other toolkit generalisations are described in Bennett et al. (2012).

$$P(k) = \sum_{n=0}^{N-1} F_n \cdot e^{-i2\pi \frac{k}{N} n} \tag{1a}$$

Becomes:

$$P(\omega) = \sum_{n=0}^{N-1} F(t_n) \cdot e^{-i\omega t_n} \tag{1b}$$

3. Results

The procedural application of the toolkit to the Lasgit dataset consisted of: initial spike detection, utilising a 48 hour window and a spike threshold of three standard deviations from the mean, followed by exclusion of the detected spikes from subsequent analysis; trend identification and removal utilising a 365 day SSA window to define trend scale; second order event detection and frequency domain analysis on the residual; visualisation of the second order event information across sensor types to highlight any periods of synchronised behaviour; and a smoothing step to improve clarity of visualisations. A number of quantifications and observations were established as a result of the toolkit applications. A selection of salient points are presented here.

Approximately 130,000 datum points out of the circa 18 million originating from the deposition hole (0.7%) were identified as spikes, with a significant concentration (approximately 56% of all the spikes) occurring in the temperature sensor records (which account for 37% of the down-hole record). Additionally, temperature sensor spiking appears concentrated in the peaks and troughs of the records. Several hundred second order event candidates were also highlighted, examples of which are shown in Figure 5 for sensor UB901. Strong seasonal (annual) cycles were found in every temperature sensor record, and in most stress and pore-water pressure sensors. The exceptions to this being those sensors containing large steps or discontinuities in their records, resulting from gas/water flow

interception, or in some cases malfunction. Correlation between the seasonal effects (i.e. the change in cycle amplitude and time-offset) and depth were strong in the temperature sensors. While annual cycles are present in stress and pore-water pressure sensors, the strong correlation of amplitude and phase offset with depth is not present. Figure 6 shows the correlation between the seasonal effects and depth in the temperature sensors in the upper Subfigures, and the lack of such a correlation in the stress and pore-water pressure sensors within the deposition hole in the lower Subfigures.

Examples of specific details revealed by the toolkit are presented in Figures 7 and 8. Three pore-water pressure sensors have been identified as having anomalously large amplitude variations, UR912, UR914 and UR920. The locations of these sensors are presented in Figure 7 with respect to the fracture intersections with the emplacement borehole in order to consider the possible effect of the experimental boundary condition. Figure 8 shows a systematic aberration in a down-hole temperature sensor (UT901) exposed by (time window) smoothing of the signal. The aberration exposed is a sharp (but small) increase in the average temperature reported by the sensor followed by an equivalent decrease approximately three months later. This phenomenon occurs approximately 1.5 months after each annual minimum.

Second order event candidates highlighted by the toolkit require detailed manual investigation on an individual basis. An initial survey of the highlighted events indicates that a mixture of sudden steps in stress, pore-water pressure, and gas injection pressure sensor records along with brief changes in sensor noise levels are represented. Manual investigation of the second order event candidates highlighted by the toolkit application is ongoing, and beyond the scope of the initial results presentation.

4. Discussion

A possible statistical explanation for the concentration of spikes detected around the peaks and troughs in the temperature sensor records is the relatively low temperature variation recorded during those periods, leading to a local lowering of the threshold beyond which a point is considered a spike. The magnitudes of the spikes detected, however, suggest that this is not the case. Additionally, the 48 hour window applied to the spike detection is too short for annual processes to exert large effects the variation within the window. This phenomenon of concentration of spikes is qualitatively similar therefore to other observable changes in the inherent 'noise' level in sensor records, as these changes are seemingly not linked to hydrological or geomechanical processes within the system. The cause therefore is unknown, and therefore warrants further exploration.

The strong correlation between depth and the temperature attributes (amplitude and phase, Figure 6) implies that the driver for the down-hole temperature profile is the HRL air temperature (which is ventilated from the surface of the facility). The modified NDFT procedure, however, appears to be finely sensitive to experimental variations/distortions in the input data. The more complexly driven physical processes, such as those measured by the stress and pore-water pressure

sensors, display detectable annual cycles (a prominent peak at a wavelength of approximately 365 days in almost every de-trended sensor record) but do not possess the tight correlation with depth that the more simply driven processes such as those measured by the temperature sensors do.

Analysis of the example phenomenon exposed by the application of the toolkit can be undertaken. . The three pore-water pressure sensors with high annual amplitudes (UR912, UR914 and UR920), cannot be correlated with the down-hole temperature profile, and are also not correlated spatially or obviously linked to boundary conditions. While two of these sensors are in close proximity to major fractures in the deposition hole rock wall (see Figure 7) the upper sensor (UR920) does not appear to intersect such a feature. Additionally, sensors UR911 and UR913 (not shown on Figure 7) lie on the same plane as UR912 and UR914 but do not have similar higher annual amplitudes. This does not exclude fracture influence on the relevant sensors as a possibility however. It is conceivable that preferential flow of water to these locations occurs from a location that is more influenced by HRL air temperature, or that other material heterogeneities exist in the system.

The absence of distortions in the residual records (after trend removal) of the high amplitude sensors indicates that they are driven by processes less complex than other nearby sensors; however the processes influencing the different observed behaviours in these particular locations is not apparent from the boundary conditions of the experiment, such as the deposition hole fracture map. It is also possible that the sensors exhibiting high amplitudes are hydraulically connected by some other mechanism than the host rock fracture network.

The systematic aberration shown in the deepest temperature sensor (UT901, shown in Figure 8), in the absence of other phenomena that coincide or synchronise with the steps, suggests a sensor or logging equipment error. The step changes (in the order of 0.05°C) are subsumed almost entirely by the sensor noise for the majority of the record. As seasonal temperature changes of approximately ±1°C appear to be driving stress and pore-water pressure changes in the order of tens of kPa, the error introduced by malfunctions of this magnitude are likely to be insignificant.

5. Conclusions

A bespoke analytical toolkit has been developed (Bennett et al. 2012) and applied to the Lasgit dataset in order to investigate 'second order' phenomenon present in the dataset. The developed toolkit has been shown to be accommodating of non-uniform time series input and sensitive to second order events, exposing phenomena that would otherwise be occluded during an Exploratory Data Analysis (EDA). The toolkit components implemented have been chosen for broad applicability to a variety of physical processes and inputs.

A large number of second order event candidates have been identified by the application of the toolkit, providing guidance towards perturbations in the dataset of potential interest for further investigation. They include changes in gas flow, stress and pore-water pressure, and changes in system noise. This ability to locate small areas of potential interest is of major use given the amount

of data and the scale of the perturbations in comparison to the overall sensor magnitudes. The identification of these event candidates is also not dependent on the separability of the seasonal components from the non-seasonal components in the dataset, nor on the macro scale behaviours of the experiment.

Otherwise unobservable phenomenon, e.g. small systematic behaviours and localised differences in behaviours, have also been exposed through application of the toolkit. In the cases where there is an inability to firmly identify any obscured phenomena that synchronises with observable phenomena there is an improved confidence in the conclusion of the absence of such additional phenomena. The small recurring steps in UT901's sensor record are not measurable or observable elsewhere in the dataset, suggesting they are an isolated but unexplained phenomenon.

Expected system behaviours, such as a highly ordered and depth correlated down-hole temperature profile, have been quantitatively confirmed. However, the stress and pore-water pressure sensors, while possessing similar quantifiable seasonal behaviours, lack this highly correlated profile. The additional non-seasonal aspects of the stress and pore-water pressure sensor records in comparison to the temperature sensor records appear to prevent the measurement of this correlation with depth, should it exist underlying the distortions, while not preventing the measurement of the presence of the seasonal components.

The development of the toolkit has therefore enabled the quantification of observable phenomena, and improved the efficiency of the manual investigation of small scale or 'second order' analysis of a large scale, non-uniform dataset.

6. Acknowledgements

7. References

Bagchi, S. and Mitra, S.K. (1999) The nonuniform discrete Fourier transform and its applications in signal processing. Kluwer Academic Publishers. ISBN: 0-7923-8281-1.

Bennett, D.P., Cuss, R.J., Vardon, P.J., Harrington, J.F., Philp, R.N., Thomas, H.R. (2012) Data analysis toolkit for long-term, large-scale experiments. Mineralogical Magazine 76 (8) p. 3355-3364.

Box, G.E.P. and Jenkins, G.M. (1976) Time series analysis: forecasting and control. Revised Edition. Holden-Day.

Cuss, R.J., Harrington, J.F. and Noy, D.J. (2010) Large scale gas injection test (Lasgit) performed at the Äspö Hard Rock Laboratory: Summary report 2008. Technical Report TR-10-38, Svensk Kärnbränslehantering AB.

Cuss, R.J., Harrington, J.F. and Noy, D.J., Wikman, A. and Selin, P. (2011) Large scale gas injection test (Lasgit): Results from two gas injection tests. Physics and Chemistry of the Earth, 36, 1729-1742.

Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001) Analysis of Time Series Structure: SSA and Related Techniques. Chapman & Hall/CRC.

Hardenby, C. and Lundin J. (2003) Äspö Hard Rock Laboratory - TBM assembly hall: Geological mapping of the assembly hall and deposition hole. International Progress Report IPR-03-28. Svensk Kärnbränslehantering AB.

Harrington, J.F., Birchall, D.J., Noy, D.J., and Cuss, R.J. (2008) Large scale gas injection test (Lasgit) performed at the Äspö Hard Rock Laboratory: Summary report 2007. Technical Report CR/07/211, British Geological Survey.

SKB (2006) Long-term safety for KBS-3 repositories at Forsmark and Laxemar – a first evaluation: Main Report of the SR-Can project. Technical Report TR-06-09, Svensk Kärnbränslehantering AB.

SKB (2008) Äspö Hard Rock Laboratory. Annual report 2007. Technical Report TR-08-10, Svensk Kärnbränslehantering AB

Table 1 – Terms used in equations (1a) and (1b)

| Term	Definition
$P(k)$ and $P(\omega)$	Power (amplitude) of frequency in input time series at the frequency corresponding to k or ω
$F(n)$ and $F(t_n)$	Value of input time series at point n or time t_n
n	Index of record in input time series (n = 0 … N - 1)
N	Number of datum points in input time series
ω	Defined as $2\pi f$
k	DFT iterator to define frequency interrogation points (k = 0 … N - 1)
i	Square root of minus one
f	Frequency at which to inspect time series (user defined)
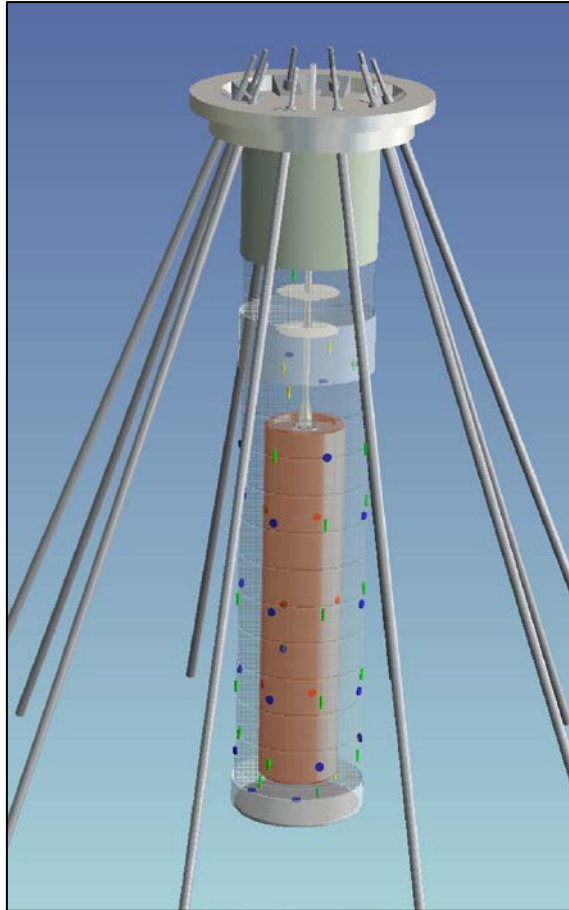
Figure 1 – Schematic depiction of Lasgit's experimental arrangement. A copper canister is shown in an emplacement borehole with hydration mats and retention simulating repository backfill. The approximate distribution of experimental instrumentation is shown by the coloured shapes (SKB 2008).
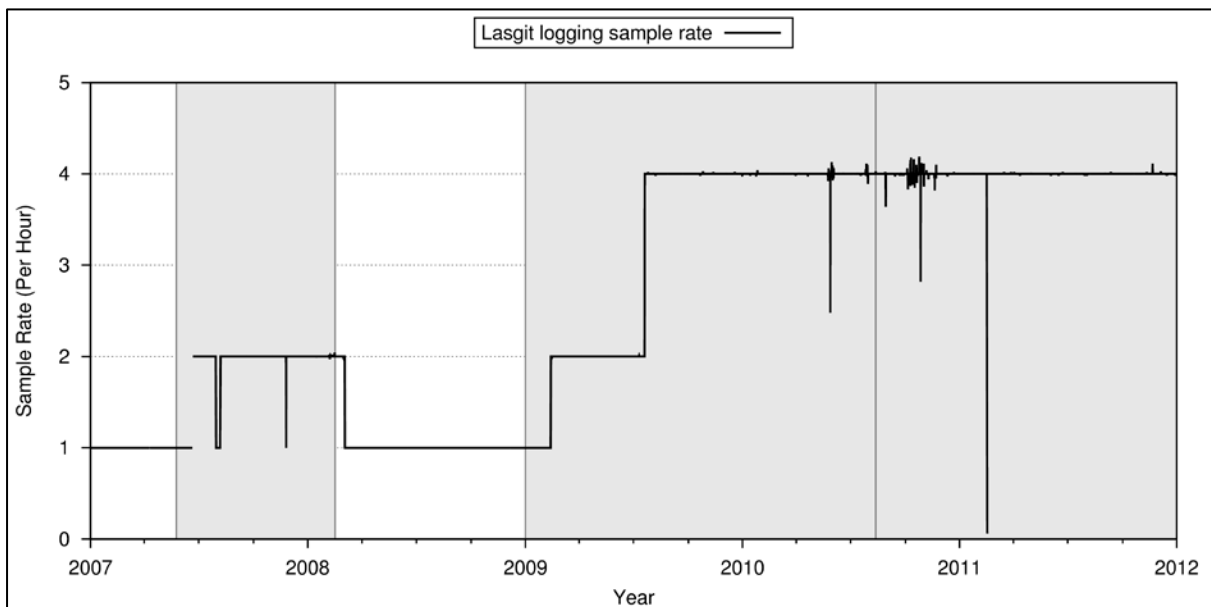

Figure 2 – Sample (logging) rate of Lasgit experiment varying with time. Grey shaded regions correspond to gas injection test phases of the experimental history.
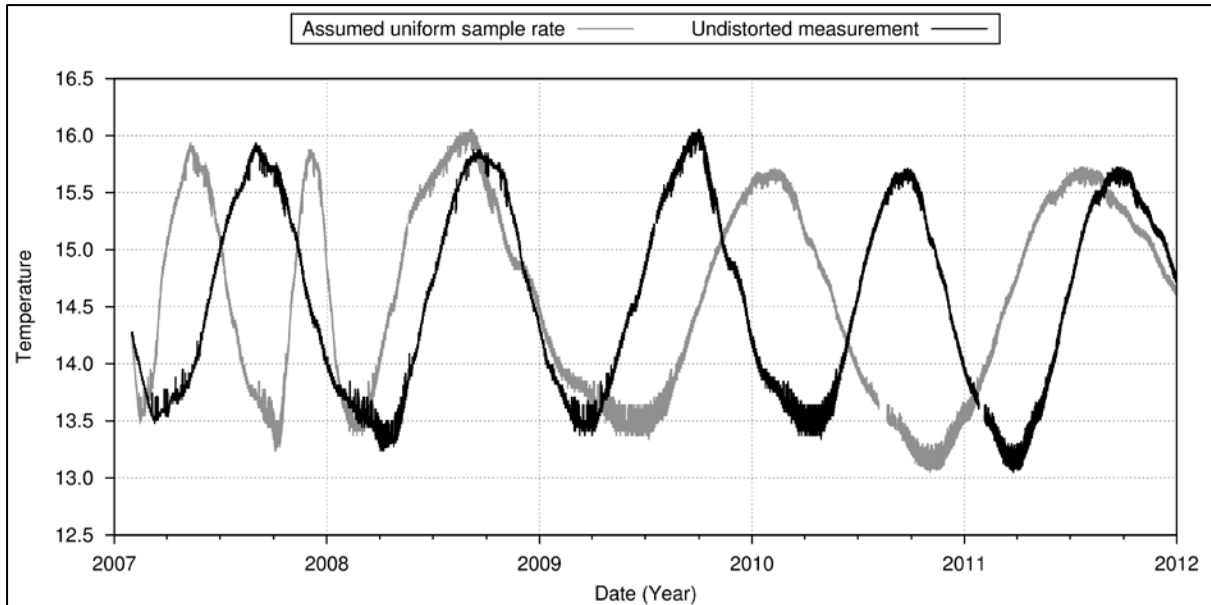
Figure 3 – Down hole temperature record shown undistorted (black) temperature record and distorted (grey) temperature record resulting from assuming a uniform sample rate when analysing non-uniform data.
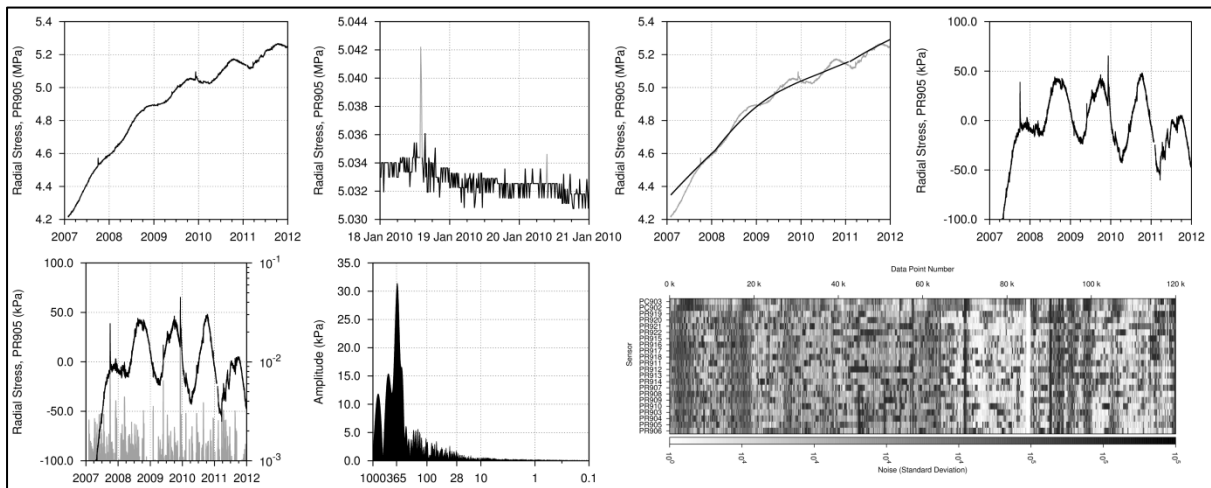
Figure 4 – Toolkit capabilities. Top row, from left to right: original signal; detected spikes in grey with remaining signal in black; SSA derived trend in black over original signal in grey; and residual signal after trend removal. Bottom row, from left to right: second order event impulses in grey with residual signal in black; frequency content (amplitude with wavelength) of residual; and intensity plot of second order event impulses showing events occurring simultaneously across multiple sensors.
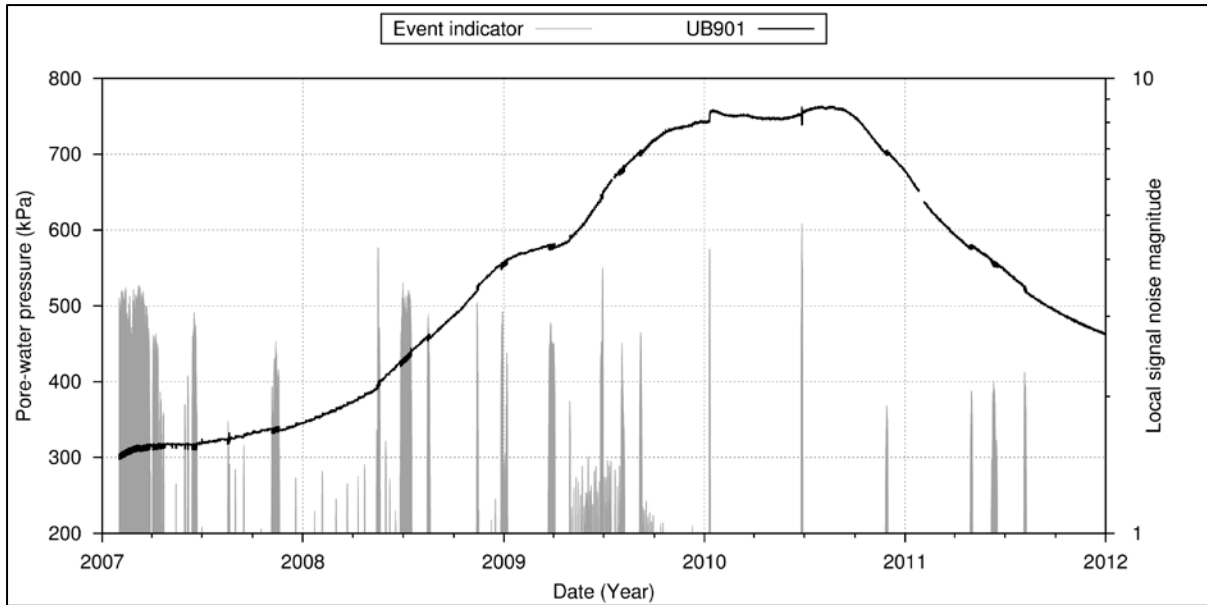
Figure 5 – Examples of second order events in sensor UB901. Prominence of noise impulses indicates small scale changes in sensor behaviour (i.e. second order event candidates).
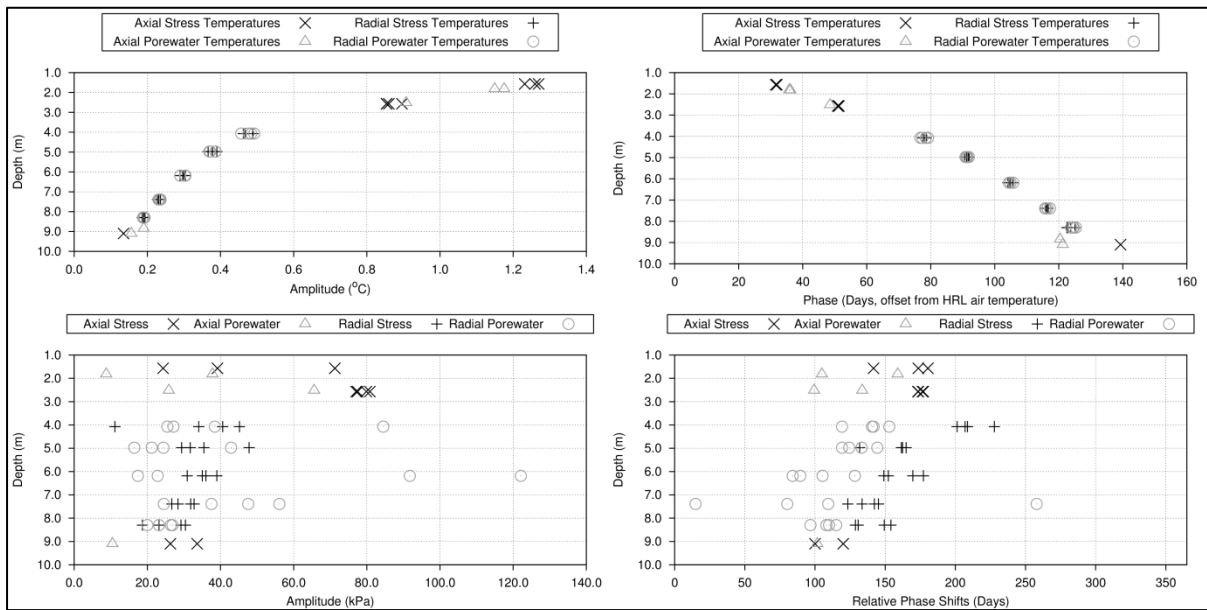

Figure 6 – Lasgit down-hole frequency analysis information. Top left: temperature sensor annual amplitudes with depth. Top right: temperature sensor annual cycle offsets (from HRL air temperature) with depth. Bottom left: stress and pore-water pressure sensor annual amplitudes with depth. Bottom right: stress and pore-water pressure sensors annual cycle offsets with depth.
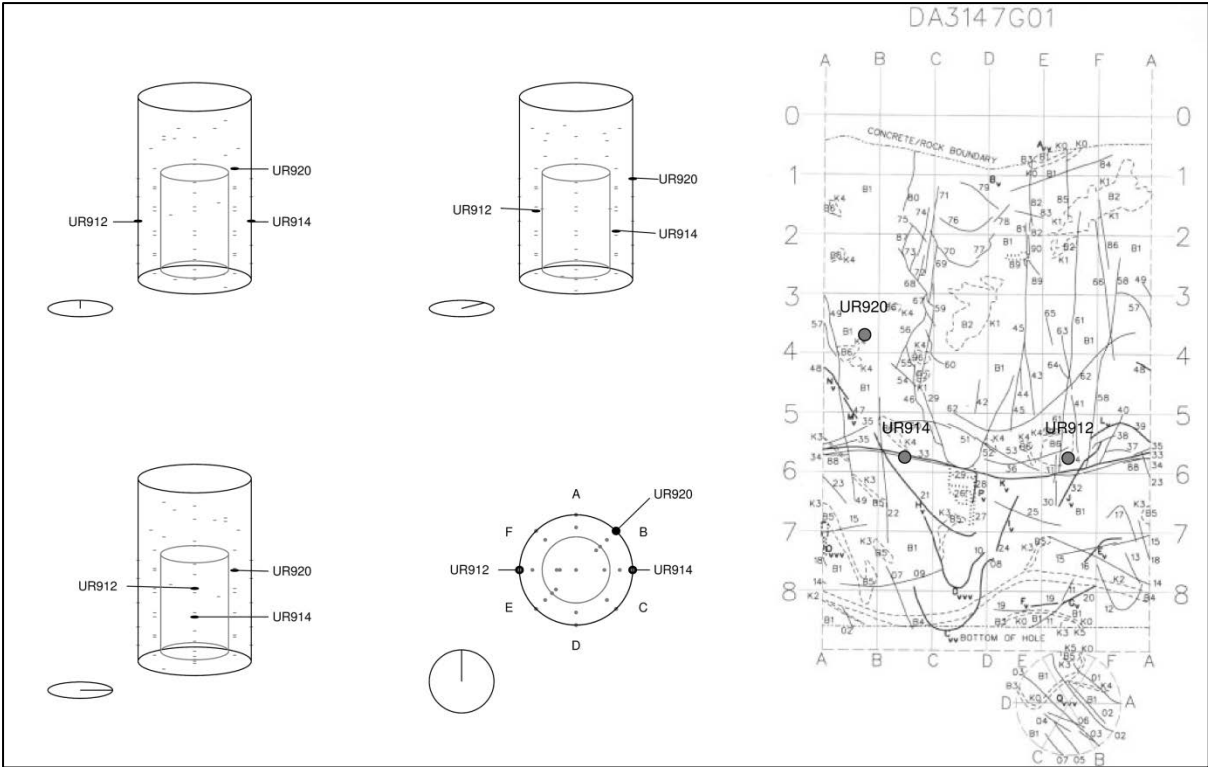
Figure 7 – High amplitude pore-water pressure sensor locations spatially, highlighted in black (left section) and with respect to borehole fractures (right section). Fracture map from Hardenby and Lundin (2003).
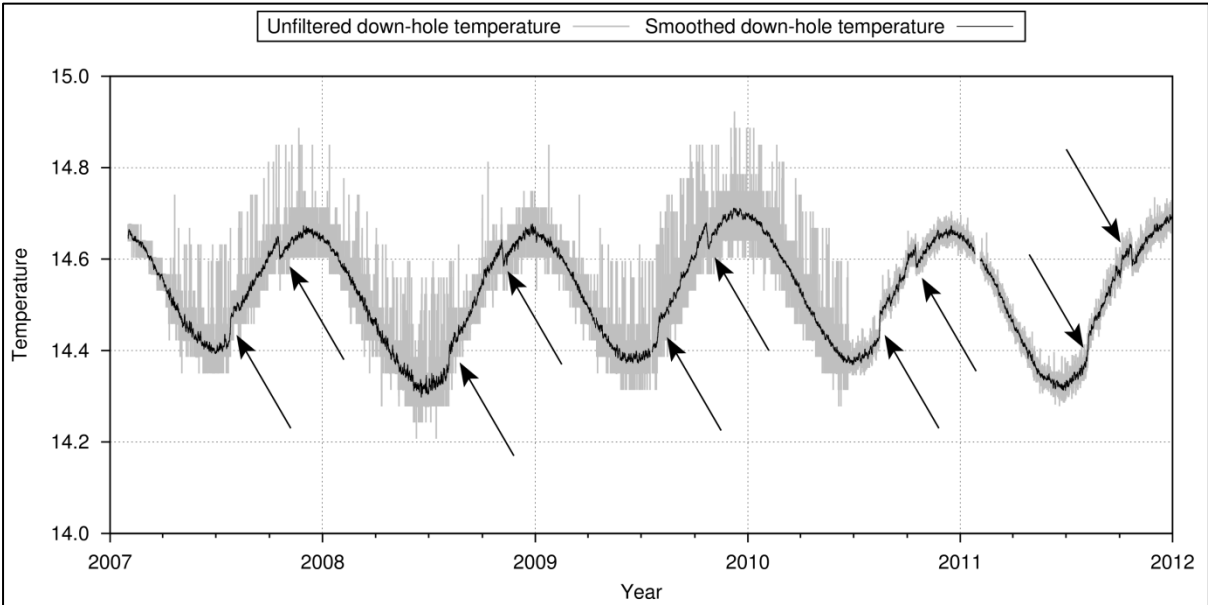

Figure 8 – Systematic aberration in down-hole temperature sensor record UT901. Original signal in grey, time windowed smoothed signal in black, showing an upward step followed by downward step each year indicated by arrows.