
Exploratory Data Analysis of the Large Scale Gas Injection Test (Lasgit)

Dan Bennett

Geoenvironmental Research Centre (GRC)
Cardiff School of Engineering
Cardiff University

*Thesis submitted in candidature for the degree of
Doctor of Philosophy at Cardiff University*

August 2014



“Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise”

— John W Tukey, 1962.

Declarations

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed: (Dan Bennett)

Date: 2014-08-07

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed: (Dan Bennett)

Date: 2014-08-07

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed: (Dan Bennett)

Date: 2014-08-07

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed: (Dan Bennett)

Date: 2014-08-07

Acknowledgements

Firstly, I would like to thank my supervisors at Cardiff University, who I genuinely feel have worked tirelessly to accommodate me during this undertaking. Prof. Hywel Thomas, Dr. Phil Vardon, and Dr. Majid Sedighi, diolch i chi, thank you, and mamnoon, respectively. The effort they have made individually and collectively for my benefit leaves me indebted to them.

Secondly, I would like to thank Dr. Rob Cuss and Dr. Jon Harrington of the British Geological Survey, who have provided me with the opportunity to participate in their work, and the funding¹ to make it possible. I feel the time I have spent with them during this undertaking has matured me, and I can only hope that this is reflected in the thesis I have produced.

I am also extremely grateful to my Father, Jeff Bennett, and my emotional accomplice, Katie Quinn, for the love and support they have both provided me over these last few years. There is no doubt that without them I would have come apart at the seams by now. I apologise to them both in advance for the sudden and inevitable betrayal I shall commit when I force them to read this thesis.

The understanding of my family, and my friends, both within the Geoenvironmental Research Centre and without, has been invaluable to me also. They have stood by me during my increasing isolationism, which has been permeated only by my insistence on socialising via the ritual torture that is middle and long distance running. I thank you all, whether you chose to run with me or not, for not walking away from me.

Finally, I would like to dedicate this thesis to those who have become lost to me during its creation. You are all thought of, and all greatly missed.

¹The research leading to the results presented in this thesis has received funding from the European Atomic Energy Community's Seventh Framework Programme (FP7/2007–2011) under Grant Agreement No. 230357, the Fate Of Repository Gases (Forge) project.

Executive Summary

This thesis presents an Exploratory Data Analysis (EDA) performed on the dataset arising from the operation of the Large Scale Gas Injection Test (Lasgit). Lasgit is a field scale experiment located approximately 420 m underground at the Äspö Hard Rock Laboratory (HRL) in Sweden. The experiment is designed to study the impact of gas build-up and subsequent migration through the Engineered Barrier System (EBS) of a KBS-3 concept radioactive waste repository. Investigation of the smaller scale, or 'second order' features of the dataset are the focus of the EDA, with the study of such features intended to contribute to the understanding of the experiment.

In order to investigate Lasgit's substantial (26 million datum point) dataset, a bespoke computational toolkit, the Non-Uniform Data Analysis Toolkit (NUDAT), designed to expose and quantify difficult to observe phenomena in large, non-uniform datasets has been developed. NUDAT has been designed with capabilities including non-parametric trend detection, frequency domain analysis, and second order event candidate detection. The various analytical modules developed and presented in this thesis were verified against simulated data that possessed prescribed and quantified phenomena, before application to Lasgit's dataset.

The Exploratory Data Analysis of Lasgit's dataset presented in this thesis reveals and quantifies a number of phenomena, for example: the tendency for spiking to occur within groups of sensor records; estimates for the long term trends; the temperature profile of the experiment with depth and time along with the approximate seasonal variation in stress/pore-water pressure; and, in particular, the identification of second order event candidates as small as 0.1% of the macro-scale behaviours in which they reside.

A selection of the second order event candidates have been aggregated together into second order events using the event candidates' mutual synchronicities. Interpretation of these events suggests the possibility of small scale discrete gas flow pathways forming, possibly via a dilatant flow mechanism. The interpreted events typical behaviours, in addition to the observed spiking

tendency, also support the grouping of sensors by sensor type.

The developed toolkit, NUDAT, and its subsequent application to Lasgit's dataset have enabled an investigation into the small scale, or 'second order' features of the experiment's results. The analysis presented in this thesis provides insight into Lasgit's experimental behaviour, and as such, contributes to the understanding of the experiment.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 1.1 | Introduction | 1 |
| 1.2 | Radioactive waste disposal | 2 |
| 1.3 | The KBS-3 deep geological disposal concept | 5 |
| 1.4 | Gas generation in a KBS-3 repository | 9 |
| 1.5 | The Large Scale Gas Injection Test (Lasgit) | 9 |
| 1.5.1 | Experimental scenario | 10 |
| 1.6 | Analysis of large datasets | 12 |
| 1.7 | Objectives and thesis outline | 13 |
| 1.7.1 | Research objectives | 13 |
| 1.7.2 | Thesis outline | 13 |
| 2 | Literature review | 15 |
| 2.1 | Introduction | 15 |
| 2.2 | Compacted clay as an EBS | 16 |
| 2.2.1 | General properties of porous materials (soils) | 16 |
| 2.2.2 | General properties of clays | 16 |
| 2.2.2.1 | Montmorillonite/bentonite | 18 |
| 2.3 | Gas flow in compacted clays | 21 |
| 2.3.1 | Asserted mechanisms | 21 |
| 2.3.2 | Gas flow experimentation | 24 |
| 2.3.2.1 | Large scale gas injection testing | 27 |
| 2.3.3 | Gas flow modelling focusing on pathway formation | 28 |
| 2.4 | Field scale experimentation/activities | 29 |
| 2.5 | Data analysis | 30 |
| 2.5.1 | General analysis approaches and techniques | 30 |

| | | |
|----------|--|-----------|
| 2.5.1.1 | Time domain | 31 |
| 2.5.1.2 | Frequency domain | 32 |
| 2.5.1.3 | Stochastic methods | 33 |
| 2.5.2 | Applications of time series analysis techniques | 34 |
| 2.5.2.1 | Trend detection/determination | 34 |
| 2.5.2.2 | Spike/event detection | 35 |
| 2.5.2.3 | Frequency quantification | 39 |
| 2.5.3 | Data analysis in civil/environmental engineering | 39 |
| 2.5.4 | Non-uniformity in data analysis | 42 |
| 2.6 | Summary | 43 |
| 3 | The Lasgit experiment | 45 |
| 3.1 | Introduction | 45 |
| 3.2 | Experimental overview | 45 |
| 3.2.1 | General arrangement | 46 |
| 3.2.2 | Scenario and aims | 49 |
| 3.2.3 | Deposition hole DA3147G01 | 51 |
| 3.2.4 | Experimental apparatus | 51 |
| 3.2.4.1 | Stress sensor notations and locations | 54 |
| 3.2.4.2 | Pore-water pressure sensor notations and locations | 58 |
| 3.2.4.3 | Injection filter notations and locations | 58 |
| 3.2.4.4 | Filter mat notations and locations | 64 |
| 3.2.4.5 | PRH pore-water pressure sensors notations and locations | 64 |
| 3.2.4.6 | Displacement sensor notations and locations | 64 |
| 3.2.4.7 | Anchor cable load sensor notations and locations | 72 |
| 3.2.4.8 | Ancillary sensor notations and locations | 72 |
| 3.2.4.9 | Integrated temperature sensors | 72 |
| 3.3 | Experimental history | 72 |
| 3.4 | Experimental reliability | 79 |
| 3.5 | Data/Results | 80 |
| 3.5.1 | Data uniformity | 80 |
| 3.6 | Summary | 82 |

| | | |
|----------|--|------------|
| 4 | Data analysis toolkit development | 83 |
| 4.1 | Introduction | 83 |
| 4.2 | Non-uniform time series analysis approaches | 83 |
| 4.2.1 | Ignoring/disregarding non-uniformity | 84 |
| 4.2.2 | Down-sampling to uniformity | 85 |
| 4.2.3 | Interpolating to uniformity | 87 |
| 4.2.4 | Use of non-uniform analysis tools | 87 |
| 4.2.5 | Approach adopted for Lasgit analysis | 87 |
| 4.3 | Non-uniform data analysis toolkit | 89 |
| 4.3.1 | Capabilities and respective rationales | 90 |
| 4.3.1.1 | Second order event candidate detection | 90 |
| 4.3.1.2 | Frequency domain analysis | 90 |
| 4.3.1.3 | Long-term trend determination and removal | 91 |
| 4.3.1.4 | Aberrant point identification (spike detection) | 91 |
| 4.3.1.5 | Smoothing/averaging/down-sampling functions | 91 |
| 4.3.2 | Formulations of the analytical components of NUDAT | 91 |
| 4.3.2.1 | Time-scale windowing vs point-count windowing | 92 |
| 4.3.2.2 | Averaging/smoothing and down-sampling | 93 |
| 4.3.2.3 | Time series Quality Assurance (QA) | 97 |
| 4.3.2.4 | Event candidate detection | 100 |
| 4.3.2.5 | Trend detection and removal | 100 |
| 4.3.2.6 | Frequency domain analysis | 103 |
| 4.3.3 | Toolkit components summary | 104 |
| 4.4 | Analysis algorithms and application procedures | 104 |
| 4.5 | Implementation and application of NUDAT software | 106 |
| 4.5.1 | Coding specifics | 106 |
| 4.5.2 | Supported dataset format | 107 |
| 4.5.3 | Human Computer Interface (HCI) | 107 |
| 4.6 | Summary | 108 |
| 5 | Toolkit verification | 111 |
| 5.1 | Introduction | 111 |

| | | |
|----------|---|------------|
| 5.2 | Verification time series data | 111 |
| 5.2.1 | Non-uniformity in the verification dataset | 112 |
| 5.2.2 | Components of verification time series | 112 |
| 5.2.2.1 | Long term trends | 112 |
| 5.2.2.2 | Distortion | 115 |
| 5.2.2.3 | Smaller scale 'second order' features/events | 115 |
| 5.2.2.4 | Frequency content | 115 |
| 5.2.2.5 | Anomalous points | 118 |
| 5.2.2.6 | Noise | 118 |
| 5.2.3 | Aggregated verification time series | 118 |
| 5.3 | Toolkit interpretation of the verification dataset | 123 |
| 5.3.1 | Individual process evaluations | 123 |
| 5.3.1.1 | Spike detection | 123 |
| 5.3.1.2 | Trend detection (SSA) | 127 |
| 5.3.1.3 | Event candidate detection | 130 |
| 5.3.1.4 | Frequency domain inspection | 136 |
| 5.4 | Summary | 136 |
| 6 | Results of toolkit application | 141 |
| 6.1 | Introduction | 141 |
| 6.2 | EDA scope and limitations | 141 |
| 6.2.1 | Data standardisation and truncation | 141 |
| 6.2.1.1 | Standardisation of time stamps | 142 |
| 6.2.1.2 | Truncation of dataset | 142 |
| 6.3 | Toolkit application procedure | 143 |
| 6.4 | Results of toolkit application | 144 |
| 6.4.1 | Spike detection | 144 |
| 6.4.2 | NDFTs of de-spiked data | 147 |
| 6.4.3 | Event candidate detection of de-spiked data | 151 |
| 6.4.3.1 | Axial stress sensors | 153 |
| 6.4.3.2 | Pore-water pressure sensors within the bentonite buffer | 153 |
| 6.4.3.3 | Injection filters and hydration mats | 155 |

| | | |
|----------|---|------------|
| 6.4.3.4 | Lid displacements and retaining loads | 155 |
| 6.4.3.5 | Pore-water pressures in the pressure relief holes | 155 |
| 6.4.3.6 | Radial stress sensors | 160 |
| 6.4.3.7 | Pore-water pressures at rock wall/bentonite interface | 160 |
| 6.4.3.8 | Temperature sensors | 166 |
| 6.4.4 | Smoothing process of de-spiked data | 166 |
| 6.4.5 | Representative values, SSA, and trend removal | 170 |
| 6.4.5.1 | Representative values | 170 |
| 6.4.5.2 | Singular Spectrum Analysis | 170 |
| 6.4.5.3 | Trend removal | 172 |
| 6.4.6 | NDFTs of de-trended residuals | 174 |
| 6.4.7 | Event candidate detection of de-trended residual data | 179 |
| 6.4.8 | Smoothing process of de-trended residual data | 179 |
| 6.5 | Summary | 179 |
| 7 | Second order events | 181 |
| 7.1 | Introduction | 181 |
| 7.2 | Choice of event candidates investigated | 181 |
| 7.3 | Event at 2010-01-10 | 182 |
| 7.4 | Event at 2010-06-27 | 185 |
| 7.5 | Event at 2010-11-29 | 195 |
| 7.6 | Summary | 199 |
| 8 | Discussion of toolkit outputs and events | 203 |
| 8.1 | Introduction | 203 |
| 8.2 | Discussion of the toolkit application results | 203 |
| 8.2.1 | Spike detection | 204 |
| 8.2.2 | Initial NDFTs | 206 |
| 8.2.3 | Event candidate detection | 206 |
| 8.2.3.1 | Axial stress sensors | 206 |
| 8.2.3.2 | Pore-water pressure sensors within the bentonite buffer | 207 |
| 8.2.3.3 | Injection filters and hydration mats | 207 |
| 8.2.3.4 | Lid displacements and retaining loads | 207 |

| | | |
|----------|--|------------|
| 8.2.3.5 | Pore-water pressures in the pressure relief holes | 207 |
| 8.2.3.6 | Radial stress sensors | 208 |
| 8.2.3.7 | Pore-water pressures at rock wall/bentonite interface | 208 |
| 8.2.3.8 | Temperature sensors | 208 |
| 8.2.4 | Discussion of smoothing results | 209 |
| 8.2.5 | Discussion of the SSA derived trends | 209 |
| 8.2.6 | De-trended NDFTs | 210 |
| 8.2.7 | Comparison of toolkit outputs before and after de-trending | 214 |
| 8.3 | Discussion of the second order events detected | 218 |
| 8.3.1 | Event at 2010-01-10 | 218 |
| 8.3.2 | Event at 2010-06-27 | 220 |
| 8.3.3 | Event at 2010-11-29 | 222 |
| 8.4 | Summary | 224 |
| 9 | Conclusions and suggested further work | 227 |
| 9.1 | Introduction | 227 |
| 9.2 | Summary and conclusions of the research undertaken | 228 |
| 9.2.1 | Toolkit development | 228 |
| 9.2.2 | Verification of the developed toolkit | 229 |
| 9.2.3 | Application of NUDAT to Lasgit's dataset | 229 |
| 9.2.4 | Exploratory data analysis results | 230 |
| 9.2.5 | Micro/macro scale behaviour and observations | 231 |
| 9.3 | Overall conclusions | 232 |
| 9.4 | Potential further research | 232 |
| 9.4.1 | Further research on Lasgit's dataset | 233 |
| 9.4.2 | Further research on other datasets | 233 |
| 9.4.3 | Further characterisation and development of NUDAT analytical modules | 233 |
| | References | 235 |

List of Figures

| | | |
|------|--|----|
| 1.1 | Activity of spent nuclear fuel with time | 3 |
| 1.2 | UK radioactive waste inventory contributions by classification..... | 7 |
| 1.3 | KBS-3 concept schematic | 8 |
| 1.4 | Lasgit scenario time frame | 11 |
| 2.1 | Schematic depiction of idealised three phase material (soil) | 17 |
| 2.2 | Schematic representation clay mineral units | 19 |
| 2.3 | Scanning electron photomicrograph image of montmorillonite (bentonite) | 19 |
| 2.4 | Schematic representation of water in compacted bentonite | 20 |
| 2.5 | Classification of gas transport mechanisms in compacted clay..... | 22 |
| 2.6 | Gas flow rates into and out of a specimen | 26 |
| 3.1 | TBM assembly hall interior..... | 47 |
| 3.2 | Schematic depiction of Lasgit experimental arrangement | 48 |
| 3.3 | Filter array orientation on canister surface | 50 |
| 3.4 | Cross section of experimental arrangement at emplacement..... | 50 |
| 3.5 | DA3147G01 fracture map | 52 |
| 3.6 | Orientation of deposition hole, anchor cables, load cells, and fracture map grid | 53 |
| 3.7 | Schematic depiction of DA3147G01 experimental arrangement..... | 55 |
| 3.8 | Summary of axial stress evolution in DA3147G01 | 56 |
| 3.9 | Summary of radial stress evolution in DA3147G01..... | 57 |
| 3.10 | Summary of rockwall/bentonite interface pore-water pressure evolution in DA3147G01 | 60 |
| 3.11 | Summary of in-bentonite pore-water pressure evolution in DA3147G01 | 61 |
| 3.12 | Summary of in-bentonite pore-water pressure evolution in DA3147G01 (ex- cluding UB902)..... | 62 |
| 3.13 | Summary of pore-water pressures induced in canister filters | 65 |

| | | |
|------|---|-----|
| 3.14 | Summary of pore-water pressures induced in filter mats | 67 |
| 3.15 | Summary of evolution of pore-water pressure in pressure relief holes | 68 |
| 3.16 | Summary of evolution of vertical linear lid displacements | 70 |
| 3.17 | Summary of evolution of vertical linear pipe displacement | 71 |
| 3.18 | Summary of evolution of lateral linear lid displacements | 71 |
| 3.19 | Summary of anchor cable load evolution | 73 |
| 3.20 | Summary of air temperature evolution in Äspö HRL at experimental location . | 74 |
| 3.21 | Summary of temperature evolution in DA3147G01 | 75 |
| 3.22 | Summary of temperature evolution in UT911 | 76 |
| 3.23 | Lasgit sample rate with time | 81 |
| 4.1 | Example of consequences of disregarding the non-uniformity in Lasgit's dataset | 86 |
| 4.2 | Example of consequences of disregarding the non-uniformity in Lasgit's dataset | 86 |
| 4.3 | Illustrative example of consequences of small perturbation in sample rate | 88 |
| 4.4 | Example of potential loss of measured detail in a non-uniform dataset interpolated to a uniform sample rate..... | 88 |
| 4.5 | Comparison of windowing methods | 94 |
| 4.6 | Weighting functions used for time windowed moving averages | 96 |
| 4.7 | Spike identification threshold evolving with time | 99 |
| 4.8 | Summary of toolkit capabilities and resulting outputs | 105 |
| 4.9 | NUDAT's input file interface | 109 |
| 4.10 | NUDAT's tool selection menu | 109 |
| 4.11 | NUDAT's process settings interface..... | 110 |
| 5.1 | Non-uniform sample rate of the verification dataset | 113 |
| 5.2 | Verification dataset linear trend component..... | 113 |
| 5.3 | Verification dataset hyperbolic trend component | 114 |
| 5.4 | Verification dataset macro scale distortion component..... | 116 |
| 5.5 | Second order events within the verification dataset | 116 |
| 5.6 | Verification dataset frequency content component..... | 117 |
| 5.7 | NDFT of verification dataset frequency content..... | 117 |
| 5.8 | Spiking component of the verification dataset | 119 |
| 5.9 | Verification dataset uniformly distributed noise component | 119 |

| | | |
|------|---|-----|
| 5.10 | Verification dataset Gaussian distributed noise component | 120 |
| 5.11 | Distribution of verification dataset uniform noise component..... | 120 |
| 5.12 | Distribution of verification dataset Gaussian noise component | 121 |
| 5.13 | NDFT of verification dataset uniform noise component..... | 121 |
| 5.14 | NDFT of verification dataset Gaussian noise component..... | 122 |
| 5.15 | Verification dataset linear based time series (TSI) | 124 |
| 5.16 | Verification dataset hyperbolic based time series (TSII)..... | 124 |
| 5.17 | Toolkit processes as applied to the artificial verification dataset | 125 |
| 5.18 | Spike detection process applied to TSI..... | 126 |
| 5.19 | Spike detection process applied to TSII | 126 |
| 5.20 | The SSA derived trend of TSI | 128 |
| 5.21 | The SSA derived trend of TSII | 128 |
| 5.22 | The SSA derived trend compared to the linear and distortion components of TSI | 129 |
| 5.23 | The SSA derived trend compared to the hyperbolic and distortion components of TSII | 129 |
| 5.24 | TSI de-trended with event candidates..... | 131 |
| 5.25 | TSII de-trended with event candidates | 131 |
| 5.26 | Detailed view of the second order events in TSI | 133 |
| 5.27 | Detailed view of the second order events in TSII | 134 |
| 5.28 | Verification data event candidate expectations | 135 |
| 5.29 | Verification data event candidate heatmap | 135 |
| 5.30 | NDFT of de-trended TSI time series..... | 137 |
| 5.31 | NDFT of de-trended TSII time series | 137 |
| 5.32 | De-trended TSI time series with NDFT derived frequency content overlain | 138 |
| 5.33 | De-trended TSII time series with NDFT derived frequency content overlain ... | 138 |
| 6.1 | Toolkit processes as applied to Lasgit's dataset | 145 |
| 6.2 | Distribution of spikes detected per sensor record | 148 |
| 6.3 | Selection of temperature sensor records after spike detection | 149 |
| 6.4 | Example of low frequency ramping in PR908..... | 150 |
| 6.5 | HRL air temperature..... | 152 |
| 6.6 | NDFT of HRL air temperature | 152 |

| | | |
|------|--|-----|
| 6.7 | Intensity plot of the axial stress sensors' SDs..... | 154 |
| 6.8 | Intensity plot of the bentonite pore-water pressure sensors' SDs..... | 156 |
| 6.9 | Intensity plot of the filter pore-water pressure SDs | 157 |
| 6.10 | Intensity plot of the lid displacement sensors' SDs..... | 158 |
| 6.11 | Intensity plot of the pressure relief hole pore-water pressure sensors' SDs..... | 159 |
| 6.12 | UR931 pore-water pressure sensor record | 161 |
| 6.13 | Intensity plot of the radial stress sensors' SDs..... | 162 |
| 6.14 | Steps in radial stress sensors PR915 | 163 |
| 6.15 | Steps in radial stress sensors PC903 | 163 |
| 6.16 | Intensity plot of the pore-water pressure sensors' SDs located at the rock wall/bentonite interface | 164 |
| 6.17 | Steps in pore-water pressure sensors UR907..... | 165 |
| 6.18 | Steps in pore-water pressure sensors UR910..... | 165 |
| 6.19 | Intensity plot of the temperature sensors' SDs | 167 |
| 6.20 | Malfunction in sensor UT911..... | 168 |
| 6.21 | Perturbation in sensor PT908 | 168 |
| 6.22 | Detail of perturbation in sensor PT908 | 169 |
| 6.23 | Sensor UT901's repeating anomaly | 171 |
| 6.24 | Representative values sample rate | 171 |
| 6.25 | NDFTs of selected bentonite pore-water pressure sensor records de-trended residuals..... | 175 |
| 6.26 | Amplitude in °C of down-hole temperature sensors' annual cycles with depth .. | 177 |
| 6.27 | Phase offset in days of down-hole temperature sensors' annual cycles with depth | 177 |
| 6.28 | Amplitude in kPa of down-hole stress and pore-water pressure sensors' annual cycles with depth | 178 |
| 6.29 | Phase offset in days of down-hole stress and pore-water pressure sensors' annual cycles with depth | 178 |
| 7.1 | Pore-water pressure sensor record UB901 | 183 |
| 7.2 | Detailed view of UB901's sensor record from 2010-01-06 to 2010-01-14 | 183 |
| 7.3 | Intensity plot of in-bentonite pore-water pressure sensor records between 2010-01-06 and 2010-01-14..... | 184 |

| | | |
|------|--|-----|
| 7.4 | Detailed view of UB902's sensor record from 2010-01-06 to 2010-01-14 | 184 |
| 7.5 | Detailed view of UB926's sensor record from 2010-01-06 to 2010-01-14 | 186 |
| 7.6 | Intensity plot of injection filter and hydration mat sensor records between 2010-01-06 and 2010-01-14 | 186 |
| 7.7 | Detailed view of FL901's sensor record from 2010-01-06 to 2010-01-14 | 187 |
| 7.8 | Detailed view of FL903's sensor record from 2010-01-06 to 2010-01-14 | 187 |
| 7.9 | Intensity plot of axial stress sensor records between 2010-01-06 and 2010-01-14 | 188 |
| 7.10 | Detailed view of PC901's sensor record from 2010-01-06 to 2010-01-14 | 188 |
| 7.11 | Detailed view of PB901's sensor record from 2010-01-06 to 2010-01-14 | 189 |
| 7.12 | Pore-water pressure sensor record UB901 | 191 |
| 7.13 | Detailed view of UB901's sensor record from 2010-06-23 to 2010-07-01 | 191 |
| 7.14 | Composite intensity plot of sensor records between 2010-06-23 and 2010-07-01 | 192 |
| 7.15 | Detailed view of PC901, PC902, and PC903's sensor records from 2010-06-23 to 2010-07-01 | 193 |
| 7.16 | Detailed view of PB928's sensor records from 2010-06-23 to 2010-07-01 | 194 |
| 7.17 | Detailed view of UR911's sensor records from 2010-06-23 to 2010-07-01 | 194 |
| 7.18 | Pore-water pressure sensor record UB901 | 196 |
| 7.19 | Detailed view of UB901's sensor record from 2010-11-20 to 2010-12-08 | 196 |
| 7.20 | Intensity plot of in-bentonite pore-water pressure sensor records between 2010-11-20 and 2010-12-08 | 197 |
| 7.21 | Intensity plot of axial stress sensor records between 2010-11-20 and 2010-12-08 | 197 |
| 7.22 | Intensity plot of radial stress sensor records between 2010-11-20 and 2010-12-08 | 198 |
| 7.23 | Intensity plot of rock wall/bentonite interface pore-water pressure sensor records between 2010-11-20 and 2010-12-08 | 198 |
| 7.24 | Detailed view of PR917, PR914, PR912, and PR906's sensor records from 2010-11-20 to 2010-12-08 | 200 |
| 7.25 | Detailed view of PB927, PB924, and PB902's sensor records from 2010-11-20 to 2010-12-08 | 201 |
| 7.26 | Detailed view of UR922, UR912, and UB902's sensor records from 2010-11-20 to 2010-12-08 | 202 |

| | | |
|------|--|-----|
| 8.1 | Spike count with depth | 205 |
| 8.2 | Examples of the sensor records with good visual correlation to ambiguous annual frequency content | 211 |
| 8.3 | Examples of the sensor records with poor visual correlation to ambiguous annual frequency content | 212 |
| 8.4 | UR931's visual correlation to ambiguous annual frequency content | 213 |
| 8.5 | Pore-water pressure sensors with comparatively large annual cycles | 215 |
| 8.6 | Comparatively large annual cycles in pore-water pressure sensors at rock wall/bentonite interface | 215 |
| 8.7 | Locations of sensors with comparatively large annual cycles | 216 |
| 8.8 | Locations of sensors with comparatively large annual cycles | 217 |
| 8.9 | Spacial orientation of sensors identified as part of event on 2010-01-10 | 219 |
| 8.10 | Locations of sensors identified as part of event on 2010-11-29 | 223 |
| 8.11 | Remaining perturbation at 2010-11-29 | 225 |

List of Tables

| | | |
|-----|---|-----|
| 1.1 | Summary of radioactive waste disposal options..... | 4 |
| 1.2 | Summary of UK radioactive waste categories..... | 6 |
| 3.1 | Lasgit stress sensor notations and locations | 59 |
| 3.2 | Lasgit pore-water pressure sensor notations and locations..... | 63 |
| 3.3 | Canister injection filter notations and locations..... | 66 |
| 3.4 | Hydration filter mat notations and locations | 66 |
| 3.5 | Pore-water pressure sensors notations and locations in pressure relief holes | 69 |
| 3.6 | Linear displacement sensor notations and locations..... | 69 |
| 3.7 | Anchor cable load sensor notations and locations..... | 74 |
| 3.8 | Ancillary sensor notations and locations..... | 74 |
| 3.9 | Key phases of Lasgit's experimental history | 77 |
| 4.1 | Weighting functions for time windowed moving averages | 96 |
| 5.1 | Enumeration of spike detection results for TSI | 125 |
| 5.2 | Enumeration of spike detection results for TSII..... | 125 |
| 6.1 | Spikes detected in each sensor record | 146 |
| 6.2 | Number of SSA components summed to produce the trend, by sensor | 173 |

Nomenclature

| | | |
|----------------|-------|--|
| b | | Spreading parameter in Laplacian weighting function |
| e | | Exponential constant (≈ 2.71828183) |
| f | | Frequency (Hz) |
| i | | Square root of minus one ($\sqrt{-1}$) |
| j | | Iteration integer |
| k | | Autocorrelation function lag index |
| m | | Total mass of soil sample (geotechnical definition) |
| m_a | | Mass of gas phase (air) in soil sample (taken as 0) |
| m_s | | Mass of solids in soil sample |
| m_t | | Total mass of soil sample (non geotechnical definition) |
| m_w | | Mass of water in soil sample |
| n | | Time series position index |
| p | | Order of Autoregressive AR(p) function |
| $p(t_n)$ | | Point of a time series at n^{th} time index |
| p_n | | n^{th} point of a time series |
| $p_s(n)$ | | n^{th} smoothed point in time series |
| $p_s(t)$ | | Smoothed point in time series at time t |
| $p_s(t_n)$ | | Smoothed point in time series at n^{th} time index |
| q | | Order of Moving Average MA(q) function |
| r_k | | Correlation coefficient of time series with its lagged self at lag k |
| t_n | | Time series time index at n^{th} point |
| w | | (Gravimetric) water content of soil |
| $w_{(G)}(t_n)$ | | Moving average weight at time t_n for Laplacian weighting |
| $w_{(L)}(t_n)$ | | Moving average weight at time t_n for Gaussian weighting |
| $w_{(T)}(t_n)$ | | Moving average weight at time t_n for triangular weighting |
| $w_{(U)}(t_n)$ | | Moving average weight at time t_n for uniform weighting |
| w_L | | Liquid limit of soil |

| | |
|----------------------------|--|
| w_n | n^{th} weight in time series weighted moving average process |
| x | Arbitrary integer specifying summation upper bound |
| FN | Number of false negative spike detections/False negative ratio |
| FN_r | False negative ratio |
| FP | Number of false positive spike detections/False positive ratio |
| FP_r | False positive ratio |
| G_{ij} | Element of Givens rotation matrix |
| K | Defined in Equation (4.5) |
| L | Singular Spectrum Analysis window length (number of points) |
| M | Threshold of number of standard deviations from the mean to qualify as a spike |
| N | Number of points in a time series |
| N_w | Number of points in windowed subset of time series |
| P | Time series with values such that $P = (p_0, \dots, p_{N-1})$ or $P = (p(t_0), \dots, p(t_{N-1}))$ |
| T_0 | Time series time index at window centre |
| T_n | Time series time index at n^{th} point within window |
| TN | Number of true negative spike detections |
| TP | Number of true positive spike detections |
| U_n | Eigenvector of \mathbf{S} |
| V | Volume of soil sample |
| V_a | Volume of gas phase (air) in soil sample |
| V_n | Defined in Equation (4.9) |
| V_n^T | Transpose of V_n |
| V_s | Volume of solids in soil sample |
| V_v | Volume of voids in soil sample |
| V_w | Volume of water in soil sample |
| $W(\omega)$ | Power of time series at frequency defined by ω |
| $W(k)$ | Power of time series at frequency defined by k |
| Z | Purely random process $Z = (Z_0, \dots, Z_{N-1})$ with mean zero and variance σ_Z^2 |
| $\mathbf{G}(i, j, \theta)$ | Givens rotation matrix, defined in Equation (4.12) |
| \mathbf{S} | Defined in Equation (4.8) |
| \mathbf{X} | Trajectory matrix of a time series |
| \mathbf{X}_n | n^{th} component of a trajectory matrix, defined in Equation (4.7) |

| | | |
|-------------|-------|---|
| α | | Autoregressive function coefficients |
| β | | Moving Average function coefficients |
| ℓ | | Window length (time) |
| π | | Ratio of circle's circumference to its radius (≈ 3.14159265) |
| θ | | Angle (radians) |
| σ | | Standard deviation of a population / spreading parameter in Gaussian weighting function |
| μ | | Population mean |
| λ | | Eigenvalue of S |
| ω | | Rotational frequency ($2\pi f$) |
| $\theta(B)$ | | Polynomial of order q in B |
| Δt | | Time elapsed between two datum points |
| 2012–12–31 | | Standard ISO 8601 date format used within this document (ISO 2004) ¹ |

¹The date formats used in some figures in this document are presented in slightly different formats due to space reasons.

Acronyms

| | | |
|--------|-------|---|
| AEV | | Air Entry Value |
| AR | | Autoregressive |
| ARMA | | Autoregressive Moving Average |
| BGS | | British Geological Survey |
| BST | | British Summer Time |
| CEST | | Central European Summer Time |
| CET | | Central European Time |
| CSV | | Comma Separated Values |
| CVRF | | Constant Volume Radial Flow |
| CoRWM | | Committee on Radioactive Waste Management |
| DFT | | Discrete Fourier Transform |
| DST | | Daylight Saving Time |
| EBS | | Engineered Barrier System |
| EC | | European Commission |
| ECT | | Event Candidate Threshold |
| EDA | | Exploratory Data Analysis |
| EEG | | Electroencephalograph |
| EU | | European Union |
| FD | | Fractal Dimension |
| FEBEX | | Full-scale Engineered Barriers Experiment |
| FEM | | Finite Element Modelling |
| FFT | | Fast Fourier Transform |
| Forge | | Fate Of Repository Gases |
| GAMBIT | | Unknown acronym referring to a research coalition |
| GMT | | Greenwich Mean Time |
| GPR | | Ground Penetrating Radar |
| GPS | | Global Positioning System |
| HAW | | Higher Activity Waste |

| | |
|--------|---|
| HCI | Human Computer Interface |
| HLW | High Level Waste |
| HPC | High Performance Computing |
| HRL | Hard Rock Laboratory |
| ILW | Intermediate Level Waste |
| IQR | Interquartile Range |
| KBS | Kärnbränslesäkerhet (Nuclear Fuel Safety) |
| LLW | Low Level Waste |
| LLWR | Low Level Waste Repository |
| Lasgit | Large Scale Gas Injection Test |
| MA | Moving Average |
| MASH | Moving Average over Shifting Horizon |
| MEGAS | Modelling and Experiments on GAS migration in repository host rocks |
| NDFT | Non-uniform Discrete Fourier Transform |
| NEO | Non-linear Energy Operator |
| NF-Pro | Near Field Processes |
| NUDAT | Non-Uniform Data Analysis Toolkit |
| PCA | Principle Component Analysis |
| PRH | Pressure Relief Hole |
| QA | Quality Assurance |
| R&D | Research and Development |
| SD | Standard Deviation |
| SHM | Structural Health Monitoring |
| SKA | Square Kilometre Array |
| SKB | Svensk Kärnbränslehantering AB |
| SNF | Spent Nuclear Fuel |
| SONE | Supporters of Nuclear Energy |
| SSA | Singular Spectrum Analysis |
| SVD | Singular Value Decomposition |
| SVM | Support Vector Machine |
| TBM | Tunnel Boring Machine |
| THM | Thermo-Hydro-Mechanical |
| TSI | Time Series I |
| TSII | Time Series II |

| | | |
|------|-------|---------------------------------------|
| UCF | | Underground Characterisation Facility |
| UK | | United Kingdom |
| URL | | Underground Rock Laboratory |
| UTC | | Coordinated Universal Time |
| VLLW | | Very Low Level Waste |
| WMA | | Weighted Moving Average |

Chapter One

Introduction

1.1 Introduction

The use of nuclear technologies is a contentious issue. The debate surrounding their use often polarises opinion, particularly with respect to electricity generation. Organisations such as Greenpeace and Friends of the Earth explicitly state absolute opposition to nuclear power, citing the risk and the cost of operating a nuclear power plant (Greenpeace 2012; FoE 2012). Conversely, organisations such as Supporters of Nuclear Energy (SONE) actively promote the use of nuclear energy, citing the need to improve energy security and reduce CO₂ emissions relating to energy consumption (SONE 2012).

Within the European Union (EU) the cost and security of energy supply, the need for reduction of greenhouse gas emissions, and the safety of nuclear energy are all considered as complementary priorities (EC 2014). Currently, with respect to such priorities: nuclear generation provides 27% of the EU's electricity and over half of the EU's greenhouse gas-less electricity; the global market for the supply of uranium is considered stable and diversified; and an amended Nuclear Safety Directive is pending, intended to continue to improve the safety of nuclear technologies (EC 2014).

Nuclear energy, while considered acceptable by the European Commission (EC) as part of the energy security agenda, is however not promoted by it, as issues regarding resulting radioactive waste disposal remain unsolved. Additionally, radioactive wastes, i.e. waste products that contain radioactive material, can and do arise through a number of industrial activities other than nuclear power generation. These activities include, among others, nuclear weapons development programmes, and medical diagnoses and treatments (CoRWM 2006).

Owing to the potentially harmful nature of radioactive waste, safe disposal of that which already

exists is of paramount importance, regardless of support or opposition to any or all nuclear technologies. This was acknowledged in Flowers (1976). Flowers (1976) also recommended that the United Kingdom (UK) should make no commitment to nuclear power generation until a radioactive waste disposal method had been established. While progress has been made since the report regarding a disposal solution, research is still ongoing.

1.2 Radioactive waste disposal

The fundamental intention of radioactive waste disposal is to isolate the waste from the biosphere for a duration long enough to exploit the natural radioactive decay of the material, thus returning it to an activity level equivalent to that of naturally occurring uranium ore (Hedin 1997). In the case of long-lived Higher Activity Waste (HAW) this may require isolation periods in the order of 100,000 to 1,000,000 years or beyond. Figure 1.1 depicts this decay time scale for the activity of Spent Nuclear Fuel (SNF) relative to that of uranium ore.

Currently the UK's radioactive waste is in interim storage at either Sellafield, in Cumbria, or at one of over 30 locations around the UK (NDA 2010a). After a period of interim storage, a final disposal of the waste will be undertaken. A number of possible final disposal options exist for radioactive waste. Table 1.1 lists and briefly describes the options that were considered by the Committee on Radioactive Waste Management (CoRWM) (CoRWM 2006).

The report, entitled '*Managing our Radioactive Waste Safely: CoRWM's recommendations to Government*', recommended geological disposal as the best available approach to radioactive waste disposal in the UK, noting that deep geological disposal is the preferred approach for High Level Waste (HLW) following a preliminary interim storage period. This recommendation was accepted by the UK Government (UK Gov. 2006).

The UK has an existing radioactive waste legacy of approximately 95,920 m³. This waste volume is expected to rise to approximately 489,330 m³ (NDA 2010b). The larger 'committed' volume is due to the currently operating nuclear facilities' inevitable future decommissioning. The total activity of the UK (committed) radioactive waste inventory is approximately 78 × 10¹⁸ Bq (CoRWM 2006). The possibility of new build nuclear facilities leads to potentially larger volumes of still radioactive waste.

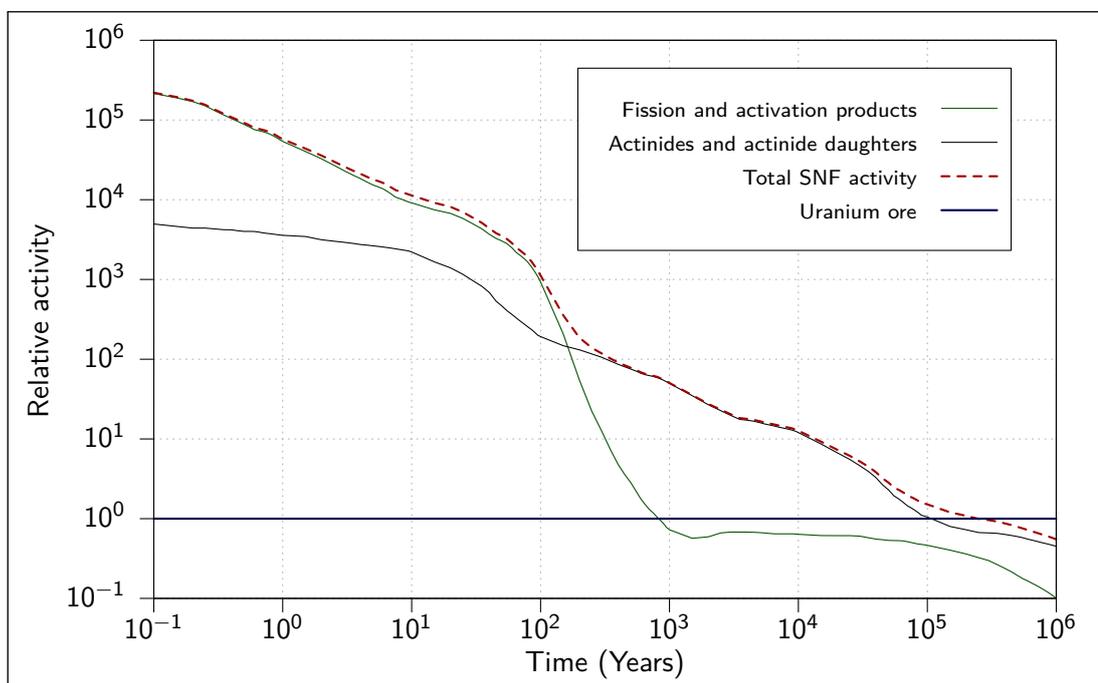


Figure 1.1: Activity of spent nuclear fuel with time relative to naturally occurring uranium ore (after Hedin 1997).

Table 1.1: Summary of radioactive waste disposal options considered by CoRWM (2006)

| Option | Description |
|----------------------------|---|
| Indefinite surface storage | In a sense, perpetual interim storage. Waste is stored at or just below surface level in managed and maintained facilities. This option also represents deferral of a decision to future generations. |
| Geological Disposal | Emplacement of waste within engineered barriers into either shallow (up to tens of metres) or deep geological hosts (500 to >1,000 metres). |
| Direct rock injection | Injection of liquid waste into rock strata. |
| Sea disposal | Emplacement of waste onto seabed. |
| Sub-seabed disposal | Emplacement of waste into geological hosts below the seabed. |
| Subduction zone disposal | Geological disposal of waste in a tectonic subduction zone. |
| Ice sheet disposal | Emplacement in permafrost such as glacial or polar ice regions. |
| Space disposal | Emplacement into high orbit or propulsion away from the planet, for example into the sun. |
| Dilution and dispersal | Disposal of waste over a large enough area that it does not pose a risk to the biosphere. |
| Transmutation | Reprocessing, recycling and reuse resulting in reduction of waste. |
| Volume reduction | Burning, incineration and melting of wastes to reduce volumes. Would be most effective when combined with other disposal options. |

A number of categories are used to classify radioactive waste by its nature and activity in the UK. Table 1.2 summarises these classifications. It should be noted that radioactive waste categorisation in the UK does not at present differentiate between short-lived and long-lived wastes (CoRWM 2006).

The UK legacy and committed waste volumes quoted above refer to the sum of the High Level Waste (HLW), Intermediate Level Waste (ILW), some Low Level Waste (LLW) that has not met the criteria for disposal at the Low Level Waste Repository (LLWR), and the plutonium and uranium stocks that may potentially be classified as radioactive waste. This inventory subset is collectively referred to as HAW (NDA 2011). An approximate breakdown of the contribution of each category to the UK radioactive waste inventory is presented in Figure 1.2. Notably, 92% of the activity associated with the UK's radioactive waste arises from the HLW and the SNF. This fraction of the waste inventory accounts for just 2% of the total waste volume.

1.3 The KBS-3 deep geological disposal concept

The final disposal concept for HLW within the UK is not yet decided. However, as deep geological disposal is the most likely disposal avenue for the UK, a technically developed deep geological disposal concept, the Swedish three layer Kärnbränslesäkerhet [Nuclear Fuel Safety] (KBS) disposal concept for HLW known as the KBS-3 concept, is used as a proxy here for a final design concept. The numeral '3' in KBS-3 refers to the number of isolating barriers in the design. The design is intended to be used in Sweden, its country of origin, and is also intended to be implemented in Finland.

In detail, the KBS-3 concept features vitrified radioactive waste stored in a cast iron insert surrounded by a copper canister. This canister is then placed in a bentonite Engineered Barrier System (EBS) lined borehole within a geological host at depth from the surface (SKB 2006). After the waste is emplaced the facility's access tunnels are backfilled to seal it from the biosphere. A schematic of the KBS-3 concept, depicting the layers multi-barrier approach, is shown in Figure 1.3.

The first barrier, the canister, is designed to contain the waste for several thousand years, achieved by way of being sealed within a strong and corrosion resistant material (SKB 2006). Should the canister fail, the EBS is designed to retard water flow and hence dissolved radionuclide

Table 1.2: Summary of UK radioactive waste categories (Vankerckhoven 1998; CoRWM 2006)

| Category | Description |
|--------------------------------|--|
| High Level Waste (HLW) | Very radioactive, mainly comprised of fission products and generating a significant amount of heat. If deemed waste, SNF would fall under this description. |
| Intermediate Level Waste (ILW) | Less radioactive than HLW. ILW does not generate a heat to a level that affects the considerations when designing management facilities but can still require significant shielding. |
| Low Level Waste (LLW) | Less radioactive still. Unsuitable for standard landfill. In the UK LLW is primarily disposed of at the Low Level Waste Repository (LLWR) in Cumbria. |
| Very Low Level Waste (VLLW) | Can be disposed of in small concentrations at conventional landfill sites. |

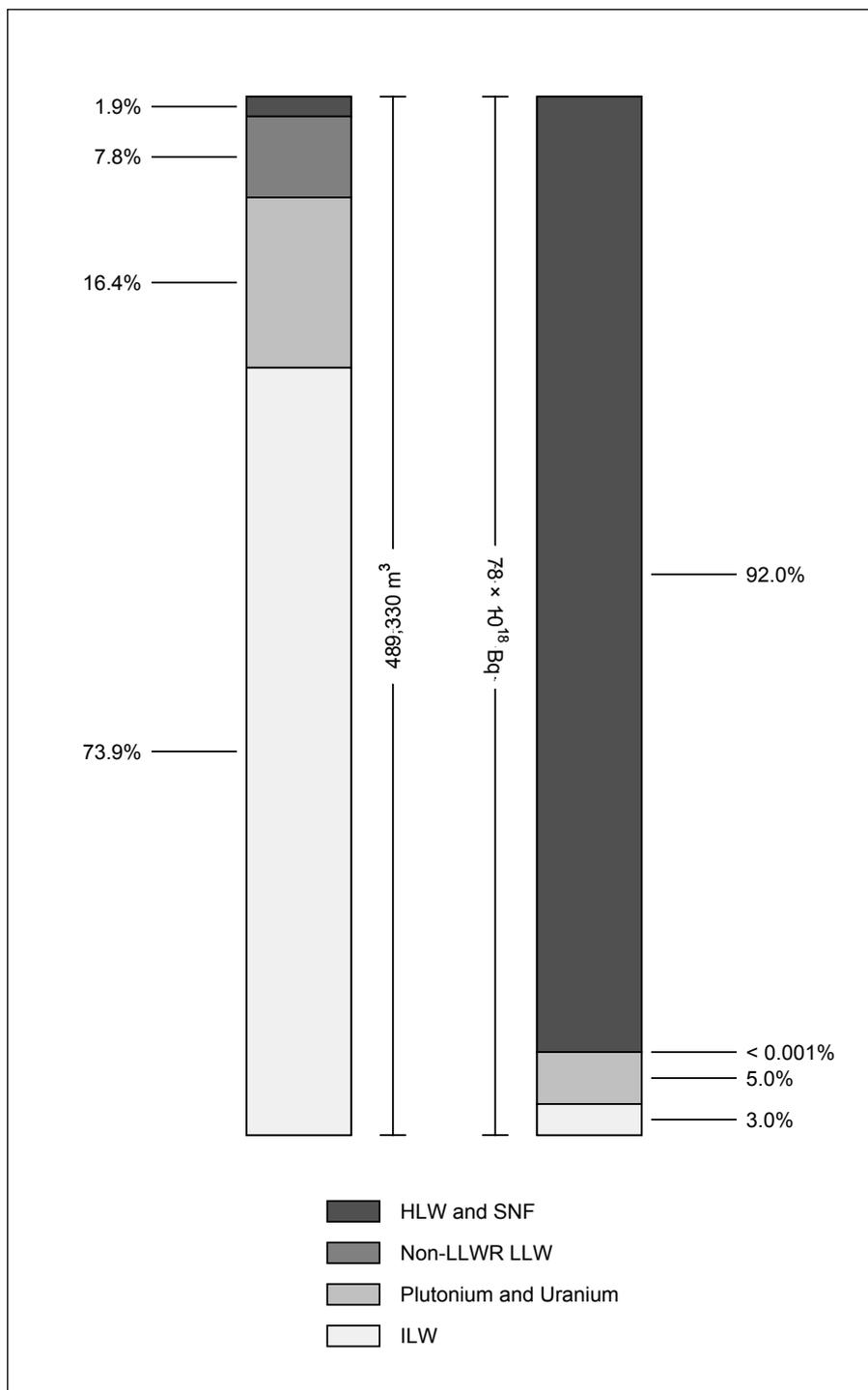


Figure 1.2: UK radioactive waste inventory volume and radioactivity contributions by classification using percentages reported by CoRWM (2006).

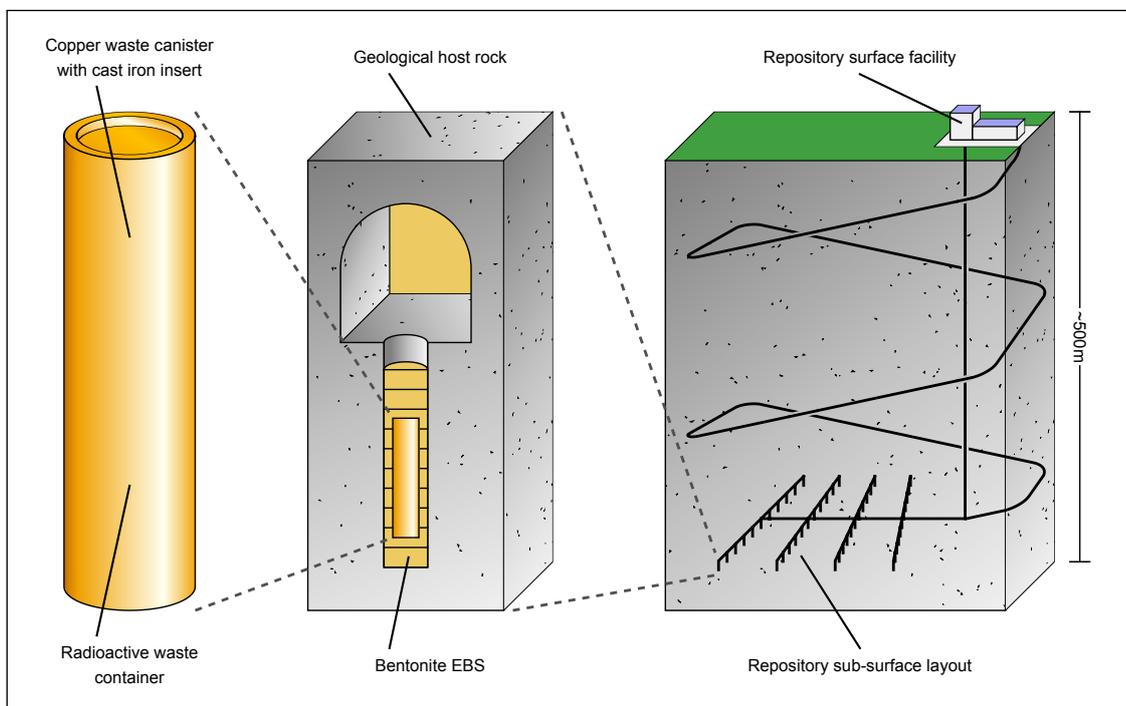
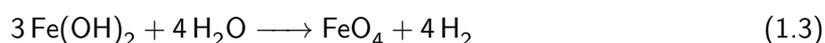
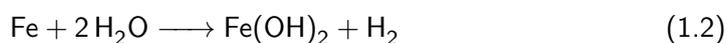


Figure 1.3: KBS-3 concept schematic (after SKB 2003).

migration out of the deposit. This is achieved through use of an extremely low permeability material such as compacted bentonite. The sorptive, swelling, and sealing properties of bentonite also contribute to the retardation of radionuclide migration. Thirdly, the geological host has both the properties of being a low flow environment, and at great distance from the biosphere, providing a third barrier (SKB 2006).

1.4 Gas generation in a KBS-3 repository

A number of processes in such a radioactive waste repository, including radioactive decay of the waste, breakdown of cellulose, or radiolysis of water may produce gas. In the specific case of a KBS-3 concept HLW repository however, the most potentially significant source of gas is the anaerobic corrosion of the cast iron canister insert (Mallants et al. 2007). The chemical processes for such an occurrence are presented in Equations (1.1), (1.2) and (1.3) (Mallants et al. 2007; Reardon 1995):



The build up and subsequent migration of gas generated by this process is of significance to the safety of such a radioactive waste repository. Migration of a hydrogen gas phase may impact the performance of one or more of the barriers, e.g. through chemical interaction, or by creation of preferential flow pathways that facilitate increased radionuclide release.

1.5 The Large Scale Gas Injection Test (Lasgit)

The Large Scale Gas Injection Test (Lasgit) is a field scale experiment, designed to simulate the build up, and to study the subsequent migration of gas through the EBS, in a KBS-3 repository environment. The experiment has been in almost continuous operation since 2005–02–01, and at the time of writing is still operating. As the study of the data arising from Lasgit's operation is the main focus of this report a detailed overview of the experiment is given in Chapter Three,

however, a brief outline of the experimental set-up is given here.

Lasgit consists of a full scale copper waste canister emplaced in a bentonite lined borehole, as per the KBS-3 design. The emplacement borehole measures approximately 9.1 m in length and approximately 1.75 m in diameter, while the copper canister is approximately 4.84 m in length and approximately 1.05 m in diameter. The resting position of the canister inside the borehole is axially central, and its base sits approximately 500 mm from the borehole floor, supported by a bentonite block.

The canister used in Lasgit has been modified with sintered filters on its surface that are used to inject gas remotely via a control system near the emplacement borehole. Experimental instrumentation consists of number of temperature, total stress, and pore-water pressure sensors at various points within and around the bentonite, and a number of hydration mats are used to accelerate the hydration of the bentonite. The borehole is sealed with a concrete plug and steel lid that is anchored into the surrounding rock mass, simulating backfill.

1.5.1 Experimental scenario

As is evident from Equations (1.1), (1.2) and (1.3), water is required for gas generation to take place. This necessitates a repository scenario in which water has entered into a waste canister and accessed the iron insert in anaerobic conditions. Due to the expected substantial life of the outer copper of the canister, a failure leading to such an ingress would likely be at a time such that the emplaced waste had cooled to an effectively isothermal equilibrium with its surroundings, and the EBS had achieved a high degree of (re)saturation via the uptake of groundwater in the host rock. Figure 1.4 depicts approximately where Lasgit's scenario lies in a generalised repository timeline.

As such, Lasgit is intended to simulate a gas build up (and subsequent flow) occurring during this period. The general experimental procedure has involved hydrating the bentonite EBS both naturally and artificially via the host rock and hydration mats, and subsequently injecting gas via the sintered filters on the canister surface into the clay over a number of campaigns. The study of the dataset arising from the operation of Lasgit is the primary focus of this research.

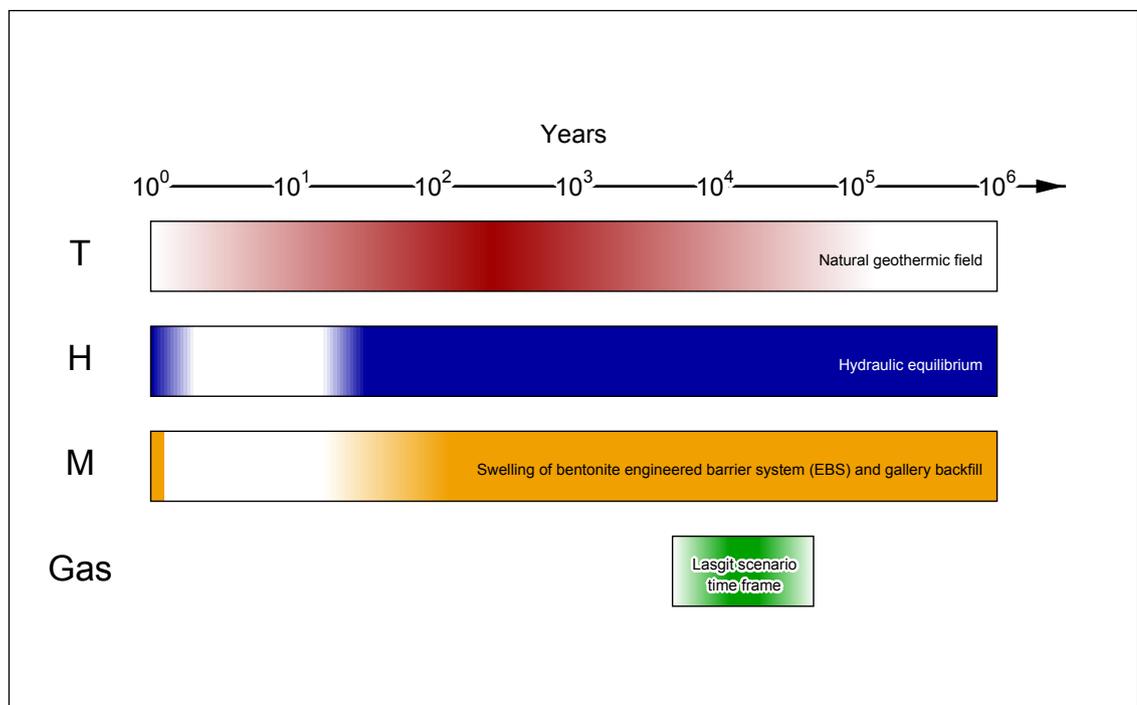


Figure 1.4: Lasgit scenario time frame within a Thermo-Hydro-Mechanical (THM) context (after Andra 2005).

1.6 Analysis of large datasets

A natural consequence of an undertaking such as Lasgit, a long term, field scale experiment, is a correspondingly large scale data acquisition/resulting dataset. Both the nature of radioactive waste disposal activities, i.e. their typically large scale, and the availability of monitoring and logging technology that facilitates automated data collection, naturally lead to the production of 'big data'.

This phenomenon is not limited to radioactive waste disposal activities, for example Mattmann (2013) describes a near future in which the Square Kilometre Array (SKA) is anticipated to collect 22,000,000,000 TB of data per year. Likewise, Xu et al. (2007) describes a data collection processes spanning 24 hours sampled at several hundred Hz, within which features lasting 20–70 ms are of interest. Activities such as computer simulations of technical or scientific problems also produce predictive data that must be analysed and interpreted. The expansion of such data producing and gathering activities has been described as an era of 'data-centric science' (Committee on the Analysis of Massive Data et al. 2013).

Challenges associated with the analysis of such datasets are frequently associated with computational memory capacity, algorithmic efficiency, seek times for data access, certainty regarding the security and provenance of the data, and management, collection and distribution of the data (Committee on the Analysis of Massive Data et al. 2013). Often overlooked is the logistic challenge of collecting uniformly sampled data over a long period. Changes in data collection procedures or hardware malfunctions/failures during a long term data collection operation can cause gaps or changes in sampling rate, leading to non-uniformity within a dataset. As length of data collection operation increases, so does the probability of encountering such issues, to the point of inevitability.

While the data arising from Lasgit's operation has not reached the amounts described above, it is substantial and beyond the scope of a hand analysis. As such, there is a focus on computational analysis of the dataset throughout this body of work, focusing particularly on statistical and time series analysis, and the ability to accommodate non-uniformly sampled data in such an analysis.

1.7 Objectives and thesis outline

1.7.1 Research objectives

The objective of this research is to robustly analyse the dataset arising from Lasgit's operation, and in doing so contribute to the understanding of the experimental, and in turn to the safety of deep geological radioactive waste disposal. To achieve this, a number of tasks have been identified:

1. Develop, in a manner tailored to Lasgit's analysis needs, a methodology for such an analysis, i.e. an Exploratory Data Analysis (EDA).
2. In development of such a methodology, seek a computational solution to its implementation that:
 - (a) Can accommodate non-uniformly sampled data.
 - (b) Is sensitive to and capable of analysing small scale, or 'second order' phenomena in the dataset.
3. Suitably verify any methodology and implementation developed to analyse the data in order to be able to properly interpret the results.
4. Apply the developed EDA methodology to Lasgit's dataset and present the results.
5. Discuss, interpret, and present the physical implications of any results arising from such an application.

1.7.2 Thesis outline

Lasgit is a large scale experimental project focusing on gas migration in the context of radioactive waste disposal. This PhD thesis presents the research undertaken while studying the experiment and its arising dataset. Described within is a robust data analysis of the experiment, including the development of a new data analysis toolkit, and an initial analysis of the underlying processes represented in the experimental dataset.

Within Chapter One, Introduction, a brief background to the research undertaken is presented, along with an outline of the thesis and the objectives of the work undertaken. Chapter Two, Literature review, presents a review of material published on gas flow mechanisms in EBS materials focusing on a particular flow phenomenon known as dilatant flow. The material reviewed include theoretical formulations, gas injection experimentation at the laboratory scale, and gas flow modelling work. Ultimately, the information gap being investigated by Lasgit is identified. Other field scale undertakings in the radioactive waste arena, and the concept of EDAs and their accompanying data analysis techniques are also reviewed. Chapter Three, The Lasgit experiment, provides a detailed review of the Large Scale Gas Injection Test (Lasgit) experiment, aggregating the technical information relied upon during subsequent analysis.

In Chapter Four, Data analysis toolkit development, the development of a time series data analysis toolkit, the Non-Uniform Data Analysis Toolkit (NUDAT), intended for use on Lasgit's dataset is described. Chapter Five, Toolkit verification, presents the work undertaken to verify the toolkit developed in the previous chapter, using an artificially created dataset as a benchmark.

Subsequently, Chapter Six, Results of toolkit application, presents the results arising from the application of the toolkit to Lasgit's dataset. Chapter Seven, Second order events, utilises the results derived in the previous chapter, aggregating them into sets of suspected small scale, or 'second order' events. Three such events are presented. Chapter Eight, Discussion of toolkit outputs and events, references both the results from Chapter Six and the derived events from Chapter Seven in a discussion of the physical implications of EDA performed.

Chapter Nine, Conclusions and suggested further work, seeks to summarise the proceeding chapters by presenting the main conclusions of the work undertaken, and subsequently looks forward at possible future work.

Chapter Two

Literature review

2.1 Introduction

The scope of the investigation presented in this research is defined largely as the study of an existing piece of work, the Large Scale Gas Injection Test (Lasgit). As such, a literature review has been undertaken to obtain an understanding of the relevant aspects of gas flow in compacted clays, and to identify the appropriate tools and approaches for an investigation into the arising experimental dataset.

While a vast amount of literature exists pertaining to the behaviour of materials expected to be used in a radioactive waste repository, a review of gas flow through an Engineered Barrier System (EBS) material *as a free phase* is focused on here in order to provide a framework within which the outcomes of this research can be interpreted. Included in this chapter is a review of the key properties and behaviours of soil, specifically bentonite clay, and its transport mechanisms is presented, along with a summary of the gas flow experimentation and computational modelling that has taken place in bentonite and similar materials. Consideration of other large-scale and long-term experimental or monitoring activities has also been undertaken, including the likelihood of such activities producing non-uniform data.

With such Exploratory Data Analysis (EDA) requirements in mind, a section on some of the available texts pertaining to EDA and time series analysis in general is also presented. Specifically (where available) texts that consider the analysis of unevenly sampled, or 'non-uniform' time series data are included. The juxtaposition of the sections on generation of potentially non-uniform data and the sparsity of the texts available on its analysis is intended to highlight the potential need to accommodate non-uniform data in an analysis and the difficulty that may be encountered in attempting to do so.

2.2 Compacted clay as an EBS

2.2.1 General properties of porous materials (soils)

The standard unsaturated soil model, commonly presented in soil focused textbooks (e.g. Barnes 2000; Mitchell 1993; Craig 1992), asserts a multi-phase material, typically consisting of three phases: solids, water, and air. An idealised schematic of these phases is presented in Figure 2.1. Fredlund and Rahardjo (1993) describe a fourth phase of this model, the contractile skin at the air-water interface. The surface tension this phase possesses gives rise to the matric suction in a soil, which is inversely proportional to pore radius. It is described as having a thickness of only a few molecular layers, and as such can be considered to be part of the water phase in the soil model without significant error.

A note on (gravimetric) water content: Water content by mass (w) is often quoted as the fraction $\frac{m_w}{m_t}$, where m_w is the mass of water in a sample of soil and m_t is the total sample mass. In geotechnical and soil science contexts (gravimetric) water content refers to the fraction $\frac{m_w}{m_s}$, where m_s is the mass of dry solids (Barnes 2000; Fredlund and Rahardjo 1993; Craig 1992). It should be noted that the latter definition is used in this body of work.

2.2.2 General properties of clays

Given the relative uniformity of air and water phases in soils in comparison to the solids, the variations between the configurations of the solid phases exert a comparatively large influence on the properties and characteristics, and as such soils are typically defined according to the solid phase. BS 5930 (BSI 1999) specifies particle size ranges for the solids in different soil classifications, with clays typically having particle sizes ≤ 0.002 mm (≤ 2 μ m). Additionally, most clay solids consist of some stacked arrangement of sheets of either silica tetrahedrons or alumina octahedrons (Grim 1962). It is the orientations and order of these stacked sheets that differentiate the clay classifications.

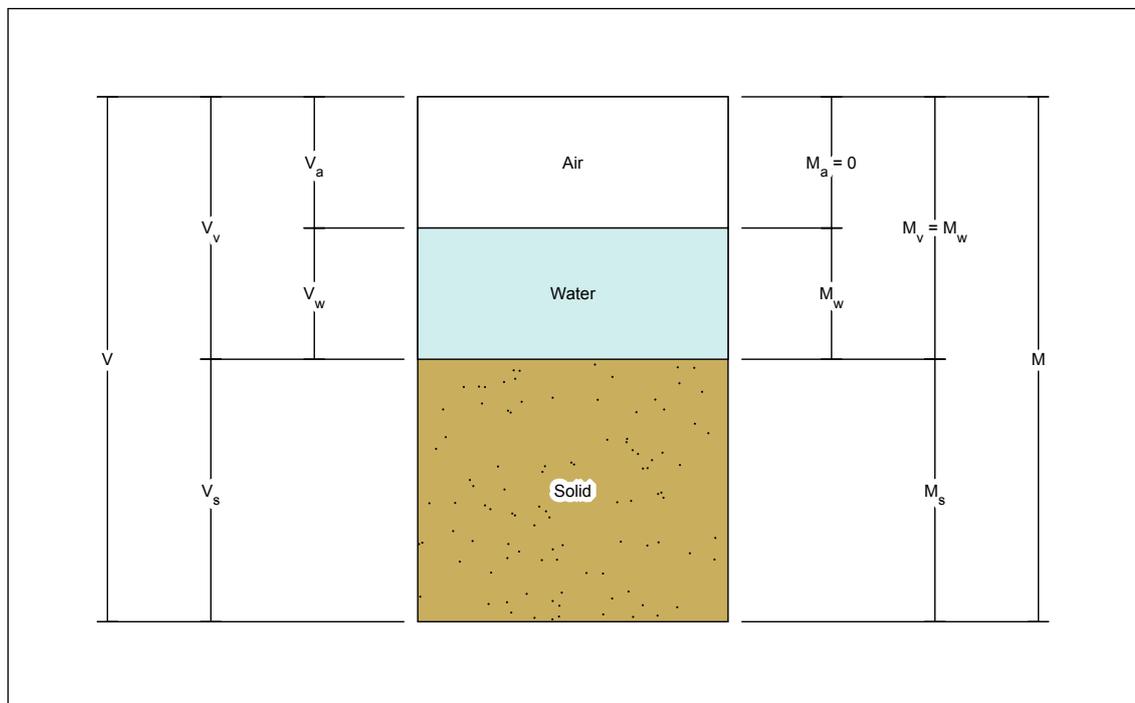


Figure 2.1: Schematic depiction of idealised three phase material (after Barnes 2000; Mitchell 1993; Craig 1992).

2.2.2.1 Montmorillonite/bentonite

A unit of montmorillonite solid (the primary ingredient of bentonite) consists of two tetrahedral sheets sandwiching a central octahedral sheet. The tips of the tetrahedrons are oriented to point towards the central octahedral sheet. Figure 2.2 depicts the tetrahedral and octahedral units, and the arrangement of the sheets in montmorillonite, while Figure 2.3 shows a scanning electron photomicrograph image of montmorillonite. A layered rather than granular structure is visible.

Bradbury and Baeyens (2003) summarises a number of interactions between the minerals and water within montmorillonite/bentonite. Water taken up into bentonite can reside in the space between clay mineral units (interlayer water) as shown with the H_2O notation in Figure 2.2, in the electric double layer on the outer surfaces of the clay stacks, and as free water in larger pore spaces. The interlayer water is described as having a 'more structured nature' than the non-interlayer water. Figure 2.4 depicts these various water classifications.

Bradbury and Baeyens (2003) also detail the generation of swelling pressure resulting from the expansion of the interlayer spacing upon water ingress, and latterly electrical double layer repulsion. The potential for swelling is stated to be sufficient to cause initial air filled porosity to be reduced to that of the interlayer spacing.

Texts such as Pusch and Karnland (1986), and Pusch and Carlsson (1985) are more explicit in their description of interlayer water, describing it as 'non-mobile' or 'ice like'. Fredlund (1991), while considering negative pore-water pressures, also suggests that water in soils when acting in tension is like a solid. Villar and Lloret (2004) suggests that the density of the non-free water in a bentonite increases with a decrease in pore volume. A gaseous analogue for non-mobile water, in which inaccessible pore space containing entrapped air is described in Faybishenko (1995)¹. Consequently, the description of a clay as 'fully saturated' may require an explicit statement regarding the inaccessible pore space.

Both the permeability (Mesri and Olson 1971) and the swelling pressure (Börgesson et al. 1996) have been related to the void ratio in bentonite and other clays. As noted above, the hydration and subsequent swelling of bentonite can lead to the closure of void space, and therefore can subsequently affect the permeability of the material. This concept is elaborated upon in Thomas

¹Faybishenko (1995) focuses on loams rather than bentonite.

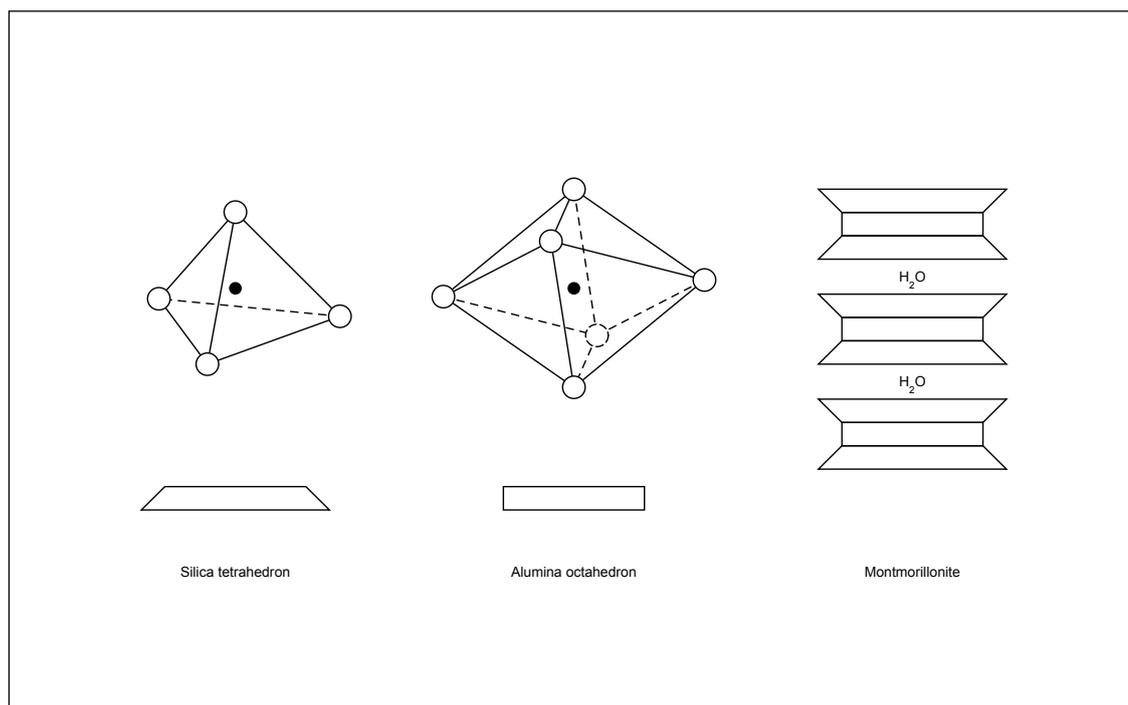


Figure 2.2: Schematic representation clay mineral units in montmorillonite (after Grim 1962).

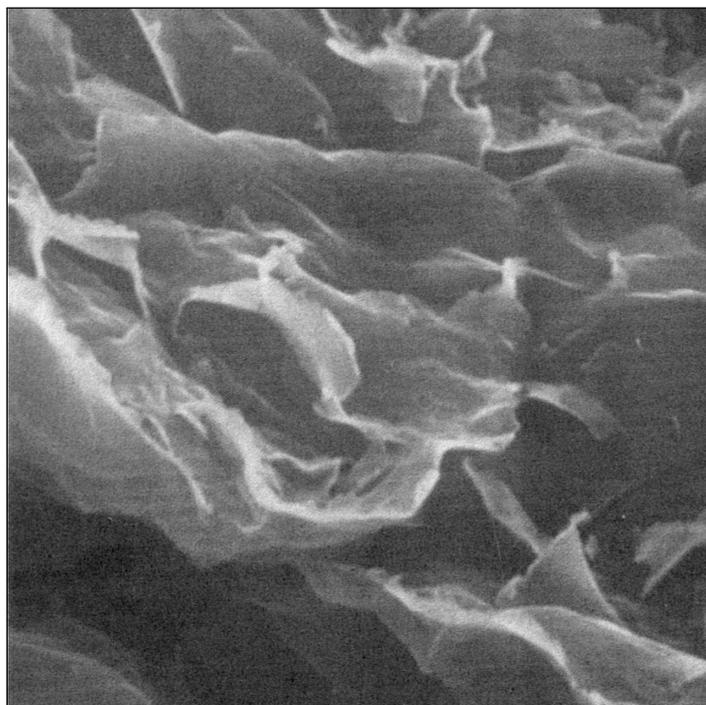


Figure 2.3: Scanning electron photomicrograph image of montmorillonite (bentonite). Image width is $7.5\ \mu\text{m}$ (Mitchell 1993).

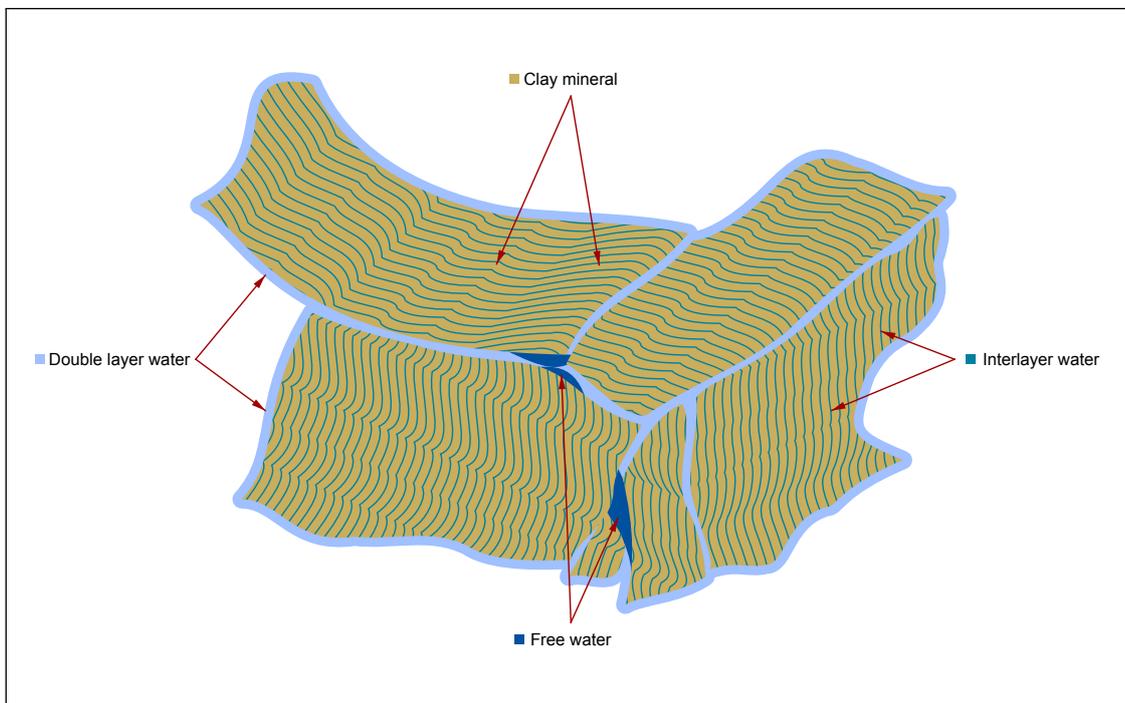


Figure 2.4: Schematic representation of interlayer water, double layer water, and “free water” in compacted bentonite (after Bradbury and Baeyens 2003).

et al. (2014, 2009, 2003). It is proposed that in addition to hydration leading to swelling and consequently a reduction in water flow, the water adsorbed onto the micro-structure of the clay becomes immobile and unavailable for further flow.

2.3 Gas flow in compacted clays

Due to the possibility of gas generation occurring in a radioactive waste repository, gas flow through compacted clays has been of interest to various researchers since the early 1980s. An array of publications on the mechanisms by which gas migrates through the such materials exist, along with a collection of experimental works performed. Consensus, particularly in the earlier material, is somewhat elusive. In particular, Western (1994) described the need to balance radionuclide containment against the potential need for gas dissipation in a repository as an 'intractable' problem.

2.3.1 Asserted mechanisms

There are four distinct mechanisms by which gas may be transported through a porous medium such as a compacted clay that are considered here (Norris 2009):

1. Single phase diffusion and advection of dissolved gasses
2. Two phase visco-capillary flow in which gas displaces water
3. Dilatant flow in which gas pressure dilates pore space locally to accommodate flow
4. Fracture flow in which macro scale rupturing of the material occurs.

Marschall et al. (2005) also provides a summary of the mechanisms, which are depicted in Figure 2.5.

The first mechanism, single phase diffusion, involves gas dissolution into pore-water (a process governed by Henry's law). Subsequent movement of the dissolved gas is by either diffusion from higher dissolved concentrations to lower (Fick's law) and/or by movement of the pore-water within the material from a point of higher pressure to lower pressure (Darcy's law). Such a flow mechanism has been characterised and numerically simulated by Vardon et al. (2014).

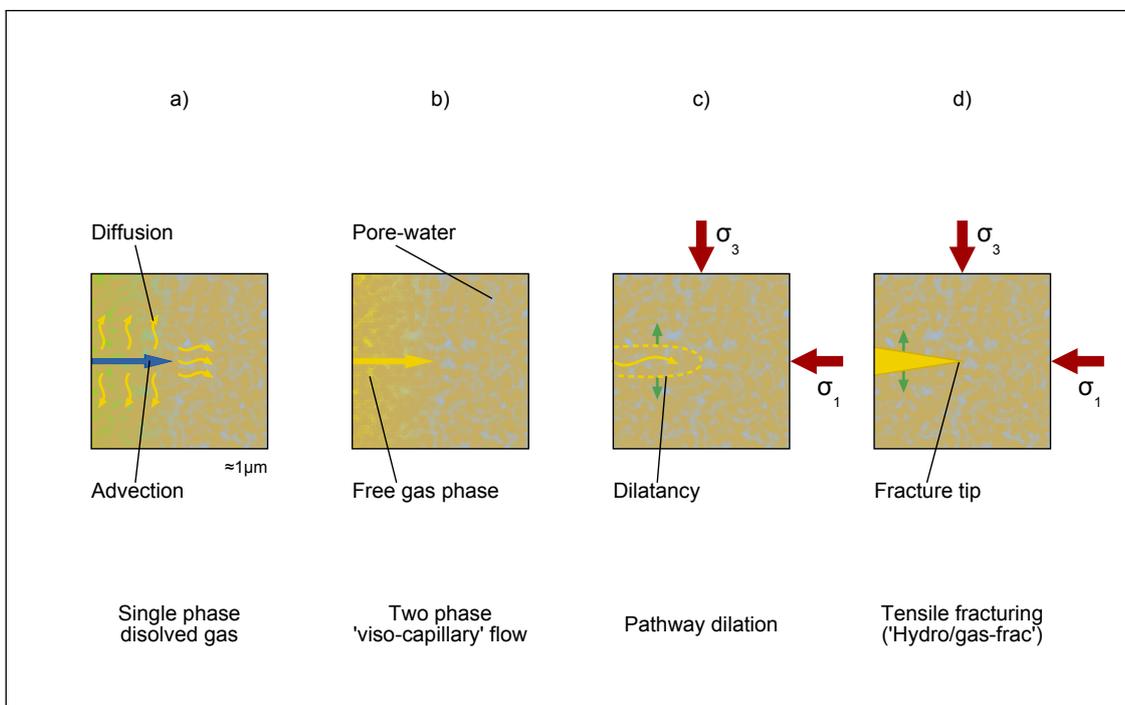


Figure 2.5: Classification of gas transport mechanisms/processes in compacted clay (after Marschall et al. 2005).

The second mechanism, two phase visco-capillary flow, can occur when the rate of gas generation exceeds the rate at which gas can dissolve into the available pore-water. In this situation a free gas phase forms. Should the pressure of the gas exceed the Air Entry Value (AEV) of the material it can displace the pore-water and migrate through the pore space of the material under a Darcian flow regime, again, characterised and numerically analysed by Vardon et al. (2014). The gas pressure required to obtain gas entry in this way was correlated to the liquid limit of clays with 60% or more particles smaller than $80\ \mu\text{m}$ by Fleureau et al. (1993). Equation (2.1) presents the relationship derived by Fleureau et al. (1993) between liquid limit (w_L) and AEV (in kPa), where w_L is expressed as a percentage:

$$\text{AEV} = 32.4 \times w_L - 466.7 \quad (2.1)$$

Taking the average of the liquid limits presented in Grim (1962) for Na^+ montmorillonite, approximately 442%, produces an AEV of approximately 13.8 MPa.

When pore space in the porous material is small the matric suction caused by the pore-water's contractile skin can grow large. Correspondingly, the gas pressure required to displace the clay fabric and cause dilation in the pore space may be lower than that required to displace the water at a sufficient rate to prevent further gas pressure build up, or even at all (Norris 2009; Olivella and Alonso 2008). In this case the third mechanism, dilatant flow, can occur. An alternative view, presented in Birgersson et al. (2008), posits that the immobility of the water adsorbed onto the clay mineral with the pore space requires the gas to consolidate the clay water system to create the pore volume required to transport it.

Horseman and Harrington (1994) formulated such a framework for gas pressure induced dilation at lower pressures than that which would be required to displace water from a small capillary. A number of features of this type of flow are noted in the report, including the instability of the gas flux, the possibility/probability of such flow mechanisms occurring between sections of visco-capillary flow, and the minimal levels of desaturation caused by the gas flow. The report also notes the similarity to fracture formation and propagation.

Rodwell et al. (1999) presented a review covering a vast array of gas flow literature, and considered all expected gas flow situations within the scope of a number of radioactive waste repository concepts. The conclusions relating to a clay based EBS indicate that the dilatant flow

mechanism is likely to be dominant in a repository condition. The review also acknowledges the lack of consensus at the time regarding such a flow mechanism.

The fourth mechanism, fracture propagation, involves macro scale rupturing of the material induced by the gas pressure. The likelihood of this situation arising in a repository condition is dismissed by Horseman and Harrington (1994), stating that previous flow mechanisms would dissipate a gas pressure build up through higher gas fluxes.

While apparently attempting to reconcile the expected high gas pressure required to displace water in a visco-capillary flow regime with experimental observations, Pusch et al. (1987) asserted gas flow in smectite occurred along a “small number of passages when a certain critical gas pressure has been reached”, specifically “wide passages”. It is suggested that in artificial montmorillonite the pore space is filled with a clay gel that is displaced by gas flow. Somewhat similarly, Impey et al. (1997) seeks to reconcile the continuum model of Darcy flow with the discrete pathway observations. A capillary network model is adopted in which an effective path of least resistance is found through a domain modelled with a given capillary size distribution. Gas is modelled as displacing water in the largest interconnected set of capillaries through a visco-capillary action.

Tanai et al. (1997) conducted experiments to determine the gas permeability of bentonite at different levels of saturation. It was concluded that at approximately 72% saturation the continuous gas phase within the clay stops. Beyond this point visco-capillary flow that displaces pore-water, or dilatant flow must take place for gas to migrate as a free phase.

Certain observations have also been made regarding the influence of scale and boundary condition on the dilatant flow mechanism. Rodwell et al. (1999) highlighted the potential sensitivity of EBS dilatancy to the confining conditions applied to it. Cool et al. (2004), in addition to stating that visco-capillary flow mechanism is inadequate for gas flow in a bentonite EBS, calls for larger scale experimentation to be conducted.

2.3.2 Gas flow experimentation

Gas injection tests into bentonite were performed by Pusch and Forsberg (1983) using nitrogen. An absence of desaturation of the test samples was noted. Assuming the gas flow mechanism to be Darcian visco-capillary flow, the results were interpreted to be evidence for flow through

only the largest capillaries.

Building on the previously noted dilatant flow framework, Horseman and Harrington (1994), Horseman et al. (1996), Horseman and Harrington (1997), Horseman et al. (1997, 1999), and Harrington and Horseman (1999) collectively present experimental work on characterisation of dilatant flow behaviour in compacted clays. Sample materials used were saturated bentonite and Boom clay, and helium was used as the injected gas. Notably, gas flows into and out of the samples were measured along with gas pressures. A particular gas pressure and flow rate profile is identified and associated with the dilatant flow mechanism. Figure 2.6 depicts this typical profile.

The profile consists of gas flow into the system and a corresponding increase in gas pressure. The peak in gas pressure corresponds to an increase in flow out of the sample. This is termed 'breakthrough'. Both the gas pressure and the flow out of the system then begin to decline and approach an asymptote. The flow out of the system is described as approaching a quasi-steady state in the literature. When flow into the system was halted a further set of declines occurred in pressure and flow out of the sample, with flow out of the sample approaching zero, and the excess gas pressure approaching an asymptote described as the 'shut-in' pressure.

Ortiz et al. (1996) reports on a collection of gas injection tests as part of the Modelling and Experiments on GAS migration in repository host rocks (MEGAS) project (including some of the experimental work presented above). The report concludes that there is evidence for both classical two-phase visco-capillary flow and dilatant flow, although impurities were commonly noted in samples where visco-capillary flow was observed.

Experimental work was performed by Kirkham (1995), and Hume (1999) involving a battery of gas injection tests at various soil properties and injection pressures. Hume (1999) suggests the possibility of a time dependency for gas breakthrough in relation to the gas pressure applied, and observed behaviour constituting evidence against classical two-phase flow, but not necessarily excluding its possibility.

Hokari et al. (1997) conducted experiments on varying size samples of sand/bentonite mixtures. The possibility of scale effects with regard to gas breakthrough into the sample is suggested, however the transmissivity of gas after breakthrough was not found to be affected by scale.

A series of hydrogen gas injection tests into Fo-Ca clay were performed and reported by Gallé

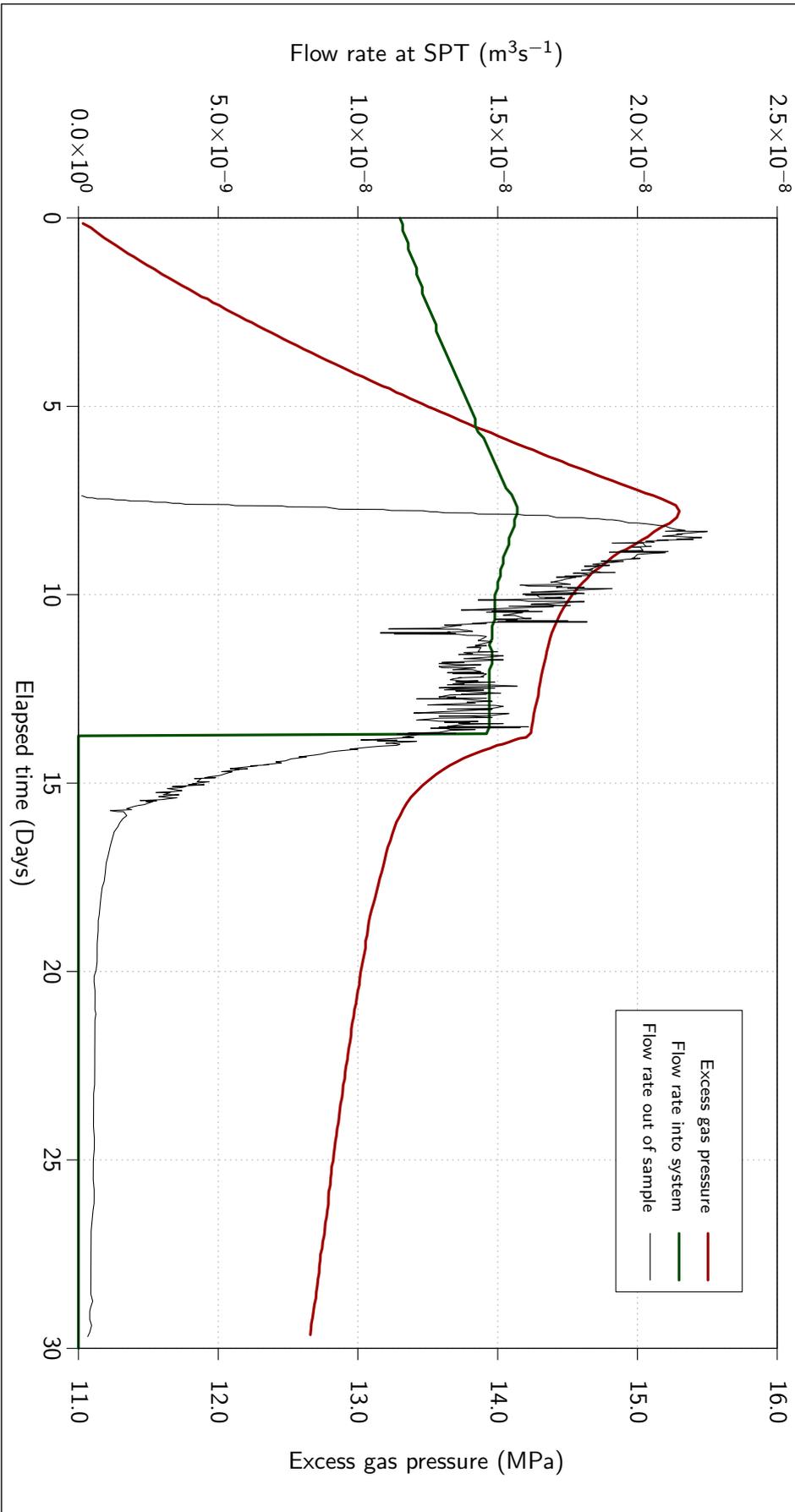


Figure 2.6: Gas flow rates into and out of a specimen of bentonite along with excess gas pressure (after Harrington and Horseman 1999).

(2000), Gallé (1998), and Gallé and Tanai (1998). Results indicated that clays with low saturations tended towards capillary flow, and clay with high saturations led to dilatant flow. The non-Darcian nature of the dilatant flow regime was noted in the reports, and a possible time/scale dependency highlighted.

In contrast to the experimentation performed on samples with some form of confinement, Donohew et al. (2000) performed a programme of 143 gas injection experiments into unconfined clay pastes. In every case, gas movement was found to be entirely through pressure induced pathways.

Experiments performed by Graham et al. (2002) sought to determine a relationship between the time taken for breakthrough to occur and the gas pressure applied in a number of bentonite/sand mixtures, and the degree of saturation required for a significant resistance to gas flow to occur. A sudden increase in gas pressure required for breakthrough was reported at 93%, similar to the behaviour noted in Tanai et al. (1997). Notably, breakthrough was found to occur at lower gas pressures during slow tests with prolonged applications of a constant gas pressure.

Harrington and Horseman (2003), and Horseman et al. (2004) report on gas injection tests performed into volumetrically confined, termed Constant Volume Radial Flow (CVRF), and axially volumetrically confined clay samples. The volumetrically confining apparatus was designed to elicit a radial flow by injecting into the centre of the sample and placing multiple sinks around the sample, while the radially confining apparatus investigated axial flow. In the radial flow experiment a transient flow pathway, with flow switching spontaneously from one sink to another, was observed. Sensitivity to boundary conditions was also noted, with large swelling pressures measured in place of sample dilation and a higher gas entry pressure attributed to the confinement. It was also inferred from the results that gas penetration would not compromise the integrity of the buffer function as evidenced by the 'resealing' of flow paths seen in the transient flow phenomenon.

2.3.2.1 Large scale gas injection testing

To investigate the flow of gas in an EBS material at large scale, an experiment called Lasgit was undertaken (NF-Pro 2006). The objective was to investigate i) the potential effects of EBS scale on gas flow; ii) the effect of the in-situ boundary conditions provided by the host

rock with respect to how volumetrically confining it may be; and iii) gas flow initiation from its anticipated formation point, at the canister/EBS interface. This allowed gas flow to be investigated in the context of the expected repository design. As the main focus of this work, Lasgit is reviewed in detail in Chapter Three.

Key Lasgit publications include the technical reports and journal publications pertaining to the experimental environment, set-up, operation, and analysis: Bäck (2003), Cuss et al. (2010, 2011, 2012), Hardenby (2004), Harrington et al. (2008), Johannesson (2003, 2004), Nowak et al. (2007), and Sandén (2003). Additionally, there are a small number of publications relating specifically to the work presented in this thesis: Bennett et al. (2011), Bennett et al. (2012a,b, 2013), and Bennett et al. (2014).

2.3.3 Gas flow modelling focusing on pathway formation

Gens and Alonso (1992) extended a conceptual model for partially saturated expansive clays that incorporated suction (Alonso et al. 1990) with an ability to take into account a dual scale effect of the clay material (macrostructure/microstructure). While the conceptual model in question was mainly focused on reproducing the behaviour of soils qualitatively the quantitative aspects of the scale effects suggest that compacted clays such as those intended for use as an EBS would be dominated by the microstructure effects.

The GAMBIT group's efforts to model gas flow through compacted bentonite is reported in Hoch et al. (2004). The report highlights the ability to successfully model the gas entry into the material, but the inability to represent the (unknown) pathway formation with the material. Rodwell (2005) elaborates on the group's efforts to establish a suitable model.

An review of alternative approaches undertaken was presented by Alkan and Müller (2008). The report provides a comparison of modelling approaches that have been attempted in order to account for the non-visco capillary gas flow mechanisms observed in compacted clays. Largely, standard visco-capillary or fracture flow models have been reported to be modified conceptually or coupled to a mechanical models, often with pre-defined discontinuities (e.g. Alonso et al. 2006). The validation of a number of the approaches described is reported to indicate a "sufficiently good" match to experimental data.

In Olivella and Alonso (2008) a coupled model for gas flow in clay in which the gas pressure

controls the aperture of a fracture modelled in a finite element is presented. The model is shown to be capable of reproducing the breakthrough flow profile observed in dilatant gas flow experiments described above.

In parallel to Lasgit's experimental operation, modelling work attempting to recreate the experimental behaviour observed was undertaken (Arnedo et al. 2010, e.g. Calder et al. 2013). Attempts to recreate the discrete flow pathways observed during gas injection testing have been made through modifications to continuum parameters such as the permeability on a discrete basis.

Vardon et al. (2014), and Vardon et al. (2010) investigated the effect of gas generation and diffusion rates in repository conditions in order to assess the likelihood of a free gas phase occurring, and to assess the likelihood of a pressure high enough for visco-capillary or dilatant gas flows occurring. The conditions under which such pressures occurred during modelling were considered overly conservative and unlikely to occur in real repository conditions.

2.4 Field scale experimentation/activities

In addition to Lasgit, a number of large/field scale activities relating to radioactive waste disposal research have been undertaken or are currently under way at both Äspö Hard Rock Laboratory (HRL), and other research facilities. A common target for investigation is the EBS. Examples, all of which are highly instrumented and long lived, include:

- The Prototype Repository project (Svemar and Pusch 2000; Johannesson et al. 2007) seeks to investigate holistically, a series of full size waste canister emplacements in a Kärnbränslesäkerhet [Nuclear Fuel Safety] (KBS)-3 concept repository. The project is intended to be operational for a number of decades, and will provide an ability to verify numerical models along with aiding the development and verification of the engineering techniques used.
- Dixon et al. (2002) describes two large scale experiments conducted at Atomic Energy of Canada's Underground Rock Laboratory (URL). The buffer-container experiment recorded the water uptake process of a large mass of buffer emplaced around a heater for approximately two and a half years, while the isothermal test monitored water uptake without a central heater for six and a half years.

- The Full-scale Engineered Barriers Experiment (FEBEX) I and II project is a full scale feasibility demonstration for EBS manufacture and assembly, and an intended source for Thermo-Hydro-Mechanical (THM) model development (GTS 2014). The experimental arrangement, described as the Spanish spent fuel disposal concept, is similar to the KBS-3 concept in a horizontal arrangement.

Additionally, examples of large-scale monitoring programs at intended implementation sites, such as the rock mechanics temperature monitoring program undertaken at the Olkiluoto Underground Characterisation Facility (UCF) by Posiva Oy (Sedighi et al. 2014; Johansson and Siren 2014), also exist.

As a consequence of longevity of such undertakings, non-uniform data acquisition may arise due to periodic maintenance, instrument failure, power interruptions, extreme natural phenomena, or simply removal of erroneous measurements (Posenato et al. 2010). The probability of experiencing an event that disrupts otherwise uniform data collection increases with the length of the experimental undertaking (O'Connor et al. 1995; Halpern 1978; Kapur and Lamberson 1977; Bazovsky 1961).

2.5 Data analysis

Data analysis is a broad field that can be approached mathematically, computationally, stochastically, from a time or frequency domain perspective, or simply intuitively. Often two or more of these approaches are combined. Publications in the field range from texts that present a variety of data analysis aspects (e.g. Tukey 1977) to those more specifically focused on a single technique (e.g. Golyandina et al. 2001). For example, Tukey (1977) presents a generalised approach to EDA, while in comparison to this, texts such as Golyandina et al. (2001) explore one specific analytical process in detail.

2.5.1 General analysis approaches and techniques

At the more general end of texts on EDA, Tukey (1977) focuses on classification of the appearance of data, and hence insight that can be drawn from such activities, as opposed to statistical confirmation and measurement. Inspection of the distribution of data is advocated,

along with smoothing of the data, and identification of 'outliers'. Smoothing is broached, crudely, in the text through the use of moving median and eye fitting. Outlier detection is achieved using thresholds defined by the Interquartile Range (IQR) (i.e. the difference between the midpoints that lie between the median value and the extreme values in a dataset).

Texts such as Diggle (1990), Chatfield (1989), Box and Jenkins (1976), and Anderson (1971) that focus on time series analysis all provide mathematical formulations and procedures that include Weighted Moving Averages (WMAs), differencing, autocorrelation, and stochastic modelling processes such as the Moving Average (MA), Autoregressive (AR), and the combination Autoregressive Moving Average (ARMA). Investigative techniques in the frequency domain, such as the Discrete Fourier Transform (DFT), are also presented.

2.5.1.1 Time domain

Activities such as smoothing via a WMA, or parametric and non-parametric trend determination (achieved through regression or component decomposition respectively) are typically conducted in the time domain. The formulation for a WMA is defined in Chatfield (1989) and others as:

$$p_s(n) = \sum_{n=1}^{N_w} w_n p_n \quad (2.2)$$

where p_n is the n^{th} point in a window of N_w points, w_n is the weighting factor associated with the n^{th} point, $p_s(n)$ is the n^{th} smoothed point, and $\sum w_n = 1$. Diggle (1990) expands on this formulation by suggesting that it is desirable for the weights to fall off symmetrically about a maximum at the central point. Typically, the formulation is presented as applied to a window with a fixed number of points either side of a central point, and pre-defined weights associated with each point, e.g.:

$$p_s(n) = \frac{p_{n-2} + 2p_{n-1} + 3p_n + 2p_{n+1} + p_{n+2}}{9} \quad (2.3)$$

The formulation for parametric trend determination via linear or polynomial regression is given in Diggle (1990), and Anderson (1971). To fit a curve of order x to data, the constants c_j of Equation (2.4) are systematically modified to minimise the sum of the squares of the residuals. Other equation forms, such as exponential or logarithmic, may also be fitted to data in this

way.

$$p(t_n)_{Trend} = \sum_{j=0}^x c_j t_n^j \quad (2.4)$$

where $p(t_n)_{Trend}$ is the fitted trend at time index t_n (the n^{th} time index).

Non-parametric trend determination techniques, such as Singular Spectrum Analysis (SSA) (Golyandina et al. 2001), involve decomposing a time series via a Singular Value Decomposition (SVD) into a set of components that sum to the original input. Typically among the components are trends with forms derived from the data rather than prescribed by the particular form of equation fitted.

2.5.1.2 Frequency domain

Texts such as Oppenheim and Schaffer (1975) focus almost entirely on time series analysis in the frequency domain. Typically this takes the form of application of the DFT to establish the frequency content of a time series. Oppenheim and Schaffer (1975) gives the DFT as:

$$W(k) = \sum_{n=0}^{N-1} p_n \cdot e^{-i2\pi \frac{k}{N} n} \quad (2.5)$$

where $W(k)$ is the power of the frequency corresponding to k , $k = 0, \dots, N - 1$, N is the number of datum points in the time series, and p_n is the n^{th} point in the time series.

The Fourier Transform as a methodology to expose the frequency content of a time series is also presented in Bloomfield (2000), Diggle (1990), and Chatfield (1989). Diggle (1990) briefly discusses the strengths and weaknesses of spectral analysis using a DFT. It highlights the usefulness of the interpretability of such a power spectrum when considering physical cycles in the data. Caveats such as the potential for harmonic frequencies to be indicated, as well as 'leakage' of measured amplitude to nearby frequencies resulting from the discretisation process are also highlighted. Additionally, the requirement for a stationary input time series is noted, however it is noted that the criteria can essentially be met via de-trending in practice.

Other points of consideration regarding the DFT include the assertion by Velasco (2003) that states that in spectral analysis, a zero frequency peak is associated with non-stationarity and/or trending behaviours, implying the DFT can be used as an informal test for such occurrences. Other possible restrictions on use include the note by Hook (2014) that states that when using

a frequency domain transformation to filter out higher frequencies the process is required to be invertible so as to be able to restore the time domain.

2.5.1.3 Stochastic methods

Stochastic methods, such as the autocorrelation function of a time series, or the MA, AR, and ARMA functions, are presented in texts such as Diggle (1990), Chatfield (1989), Hannan et al. (1985), Box and Jenkins (1976), and Anderson (1971). In the case of MA, AR, and ARMA models, the fitting of such functions to time series data is intended to switch the burden of analysis from the analysis of the data to the more 'parsimonious' model parameters (Box and Jenkins 1976).

The autocorrelation function seeks to define how a time series correlates with itself at various lags. The correlation coefficient r_k at a lag of k time steps is approximated, for a uniformly spaced time series, by:

$$r_k = \frac{\sum_{n=1}^{N-k} (p_n - \bar{p})(p_{n+k} - \bar{p})}{\sum_{n=1}^N (p_n - \bar{p})^2} \quad (2.6)$$

where p_n is the n^{th} point in the time series, \bar{p} is the mean value of the time series, and N is the number of datum points in the time series.

The change in correlation coefficient r_k with the change in lag k can be used as an indicator of periodicity, trend, or the absence of randomness in the data. In the case of uniformly sampled data the displacement in time defined by a displacement k will be uniform. However a non-uniformly sampled data may provide an inconsistent representation of the correlation in time for any particular lag k .

The MA, AR, and ARMA models are models that can be used to represent stationary processes in which consecutive values are not independent of one another, but dependent on previous values. The MA function of order q is defined as:

$$p_n = \beta_0 Z_n + \dots + \beta_q Z_{n-q} \quad (2.7)$$

in which p_n is the n^{th} point in the time series, the set of β s are constants, and the Z s a discrete

and purely random process scaled such that $\beta_0 = 1$. The AR function of order p is defined as:

$$p_n = \alpha_1 p_{n-1} + \cdots + \alpha_p p_{n-p} + Z_n \quad (2.8)$$

In which the α s are constants. The combination of the two models into an ARMA function of order (p, q) is given by:

$$p_n = \alpha_1 p_{n-1} + \cdots + \alpha_p p_{n-p} + Z_n + \beta_1 Z_{n-1} + \cdots + \beta_q Z_{n-q} \quad (2.9)$$

These models may be fitted to the time series they are intended to represent via systematically altering the α and β coefficients to minimise the residuals when compared with the measured processes.

2.5.2 Applications of time series analysis techniques

In a review on data mining time series, Fu (2011) provides a general summary of approaches that include techniques such as: down-sampling the data to reduce the number of points in a time series; the use of a 'shape description alphabet' to characterise each section of a time series; transformation of the time series into the frequency domain via a DFT; and Principle Component Analysis (PCA) of a time series via a SVD.

The stated focus of Fu (2011) is on techniques intended to aid the discovery of information. Often in the literature applications of time series analysis techniques are focused towards uncovering or quantifying a more specific information aim. Examples of which are detailed in the following sections.

2.5.2.1 Trend detection/determination

In order to extend the trivial case of linear trend detection, Visser (2004) addresses the *detection* of flexible trends in environmental time series through the extension of linear trend estimation. Confidence intervals for the trend, for the rate of change of the trend, and for the difference between the final trend point and the previous points are derived. This approach allows for the significance of the trend to be tested with time, the significance of the apparent changes in trend to be tested, and how significantly the present deviates from the past. This allows

environmental policy impact to be broadly characterised in comparison to past behaviour.

Another investigation into the *detection* of trends, in this case in temperature records for India as a whole and seven sub-regions within it, was performed by Sonali and Nagesh Kumar (2013). The investigation concentrated on the detection of a trend with a statistical significance, rather than the determination of a quantified trend. A variety of trend detection techniques are presented, notably focusing on linear and/or monotonously increasing trends. Techniques referenced include:

- Linear regression
- Mann-Kendall testing
- Spearman's rank correlation test
- Sen's slope
- A novel/innovative approach involving plotting the first half of the data against the second half and interpreting the plot.

Examples of trend *determination* include the use of SSA (Golyandina et al. 2001) by Rocco S (2013) on data pertaining to the failure of mechanical components. The derived trends were subsequently utilised to forecast the failure rate of components after longer use.

2.5.2.2 Spike/event detection

Wilson and Emerson (2002) provides a review of development between c1975 and c2002 on spike detection in Electroencephalograph (EEG) analysis. The approaches/procedures reported in the review generally involve a description of what a spike can be considered to be, and a comparison of the data parametrised in some way to the description. The main conclusions of the review highlight the poor performance of algorithmic approaches (in comparison to human experts) in cases where spiking is indistinct from background noise or other phenomena, and the tendency towards false positives. The ambiguity associated with the initial classification of what constitutes a spike is suggested to be conducive to this. The context within which these assertions are made lead to the further conclusion that the problem should be approached from an interpretive viewpoint (in a sense event detection), rather than via the single issue of spike detection.

Similarly, Acir and Güzeliş (2004) states that false positives 'plague' all automatic detection systems. The spike detection method detailed in this paper involves a pre-classification phase to select spike candidates, and a subsequent post-classification phase to refine the selection. Local minima/maxima criteria are used as part of the pre-classification process, along with an AR model to identify non-stationary portions of the data. The second phase, involving the local minima and maxima that are found to be non-stationary, are classified into spike and non-spike sets using a Support Vector Machine (SVM). The spike duration and amplitude are the parameters of choice for each classification phase.

Ekanadham et al. (2014) states that there is no de facto standard for spike sorting. The lack of standard is asserted "despite the ubiquity and succinct formulation of the problem". The process of spike detection is subsequently defined with three broad steps: Candidate identification, feature determination (parametrisation), and subsequent classification.

A paper by Mukhopadhyay and Ray (1998) describes spike detection as a two stage process: pre-emphasis of the spike, followed by classification via threshold criteria. Arguably, the pre-emphasis step could be considered to be a convolution of candidate identification and parametrisation steps. The primary technique described in the paper is the use of a Non-linear Energy Operator (NEO) with smoothing, used to estimate the instantaneous energy at various points in the sensor record. The use of smoothing with the NEO is asserted to reduce false positive detection, but increases false negative detection.

Esteller et al. (2001) reviews a number of Fractal Dimension (FD) estimation techniques for detection of transients in a time series. It is suggested that the exponential sensitivity of Katz's algorithm in providing an estimate of the actual FD of time series data emphasises higher FDs and hence leads to clearer differentiation in transients of different types.

Parameter threshold as a criteria for spike detection in EEG study is discussed in Brown et al. (2007). The paper describes the classification of spikes by a large amplitude in comparison to background levels and a characteristic duration of such an increase, or by an apparent peak to trough amplitude combined with a peak to trough time threshold.

Liu et al. (2012) states that "threshold crossing is the most widely used for spike detection because of its simplicity and low computational complexity". NEO and other similar energy operators are suggested as popular for EEG analysis, as spikes possess greater energy within

a given time interval as background behaviour. Other criteria are also considered, such as localised amplitude or Standard Deviation (SD) of the signal, with a deviation criteria of 3–5 SDs for SD thresholding. In addition to simple thresholds, spike morphology (structure and shape) is considered. Four spike constituent elements are noted: erosion, dilation, opening and closing.

With respect to anomaly detection in a dataset, Orlando (2012) presents work on the trade off between detection and resolution in Ground Penetrating Radar (GPR) surveying. The paper seeks to address the issues associated with sensing of phenomena within a medium possessing similar sized heterogeneities to those features of interest. The process utilises lagged measurements to align the estimated derivatives with the peak of the spike.

Gaspard et al. (2014) developed an algorithm for detection of EEG spikes using a instantaneous power coupled with the rate of change in the 'upslope' and 'downslope' portions of the spike and the temporal relationship between the three measures. The three measures are then reduced to a univariate measure and statistically checked for significance.

Multiple spike detection techniques are presented in Azami and Sanei (2014). Threshold criteria include the signal amplitude, SD, FD, and use of the NEO. These parametrisations are coupled with various smoothing techniques, including SSA. The paper concludes that SSA while requiring a window length twice that of the frequency content of interest, was useful for smoothing when combined with FD and NEO parametrisation. Additionally, SD is suggested as potentially useful in generalised event detection.

Pimentel et al. (2014) presents work on novelty detection in data through characterisation of the non-novel data and detecting an absence of match. This approach may be useful when the novelty sought in a dataset is too sparse, rare or unknown for adequate training of a selection process. Tukey's outlier test is given as a specific example, along with a number of statistical and difference methods for detecting aberrance, neural nets/SVMs that are trained to reject input not similar to the training, subspace methods such as PCA in which anomalies manifest as large variations in the residual, and use of information theory. The paper cites a number of possible applications including: electronic security; medical diagnostics and monitoring; engineering monitoring and damage detection; image processing and surveillance; text mining; and sensor networks.

A general framework for event detection in time series is presented in Lara et al. (2014). The proposed generalised methodology involves defining the salient features of events of interest in order to create criteria for their detection. Comparison of multiple time series to one another based on the similarity of defined events contained within them can then be quantified.

A note on spike detection accuracy measurement: A number of different metrics that measure the accuracy of spike detection algorithms are present throughout the literature. Counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) detections are typically used to enumerate them. Brown et al. (2007) defines the ‘sensitivity’ and ‘accuracy’ of spike detection processes as:

$$Sensitivity = \frac{TP}{TP + FN} \quad (2.10)$$

$$Accuracy = \frac{FP}{TP + FP + FN} \quad (2.11)$$

whereas Liu et al. (2012) defines a ‘hit rate’ in the same manner as the ‘sensitivity’ as above, but defines ‘precision’ in place of ‘accuracy’ as:

$$Precision = \frac{TP}{TP + FP} \quad (2.12)$$

Likewise, Gaspard et al. (2014) uses the same ‘sensitivity’ and ‘accuracy’ as above, but also defines the ‘false detection ratio’ FP_r as:

$$FP_r = \frac{FP}{TP + FP} \quad (2.13)$$

whereas Mukhopadhyay and Ray (1998), and Azami and Sanei (2014) define the false negative ratio FN_r and the false positive ratio FP_r as:

$$FN_r = \frac{FN}{TP + FN} \quad (2.14)$$

$$FP_r = \frac{FP}{TP + FN} \quad (2.15)$$

Additionally, Azami and Sanei (2014) also utilises the 'sensitivity' defined above. Kim and McNames (2007) uses a 'total error' (TE) as a performance indicator, defined as:

$$TE = \frac{FP + FN}{TP + FN} \quad (2.16)$$

2.5.2.3 Frequency quantification

Examples that use the application of frequency domain quantification as a primary investigative technique include Reikard (2012). The work presented contrasts the use of an AR model, frequency domain inspection, and neural nets to forecast SO_2 and SO_4 concentrations (associated with volcanic smog, or 'vog') over a short time horizon. The frequency domain forecasting was found to be best due to a strong diurnal cycle in the concentration data.

Similarly, Koirala et al. (2010) makes use of the autocorrelation function, the ARMA, and frequency domain analysis using the Fast Fourier Transform (FFT) to study persistence in chloride concentrations in hydraulic systems and their connectivity to rainfall and discharge. The frequency domain analysis performed revealed periodicities in both the short term and long term concentrations.

Gillberg and Ljung (2010b) presents a methodology for determining a continuous function that has been measured discretely and non-uniformly in time. In the companion paper that focused on uniformly sampled data (Gillberg and Ljung 2010a), a good estimate was produced by calculating the DFT of the time series, and using the DFT to approximate the continuous Fourier Transform, which can be inverted to determine a continuous function. The non-uniformly sampled data problem is accommodated by fitting a spline via polynomial interpolation to the data, and utilising a continuous Fourier Transform on that spline to determine a continuous function.

2.5.3 Data analysis in civil/environmental engineering

Data analysis in a civil and/or environmental engineering settings requires the analytical procedures and the quantities they measure to be relatable to physical processes and effects. Structural Health Monitoring (SHM) is one example in which a change in measured quantities

can be related directly to a change in structural performance. For example, Maalej et al. (1999) presents the spectral analysis of strain sensors attached to the Salmon River bridge in Canada. The monitoring process involved calculating the frequency domain information using short windows in order to accommodate the non-stationary nature of vibration as a reaction to transient loading. Detected changes in frequency content that occurred under loading were interpreted as indicating alterations in the stiffness of structural members.

Owen et al. (2001) presents work on the monitoring of damage to a cable stayed bridge and to reinforced concrete beams in a similar conceptual way to Maalej et al. (1999). The decaying/damped frequencies of interest in Owen et al. (2001) necessitate very short windows to overcome the DFT limitations pertaining to stationarity, but simultaneously requires sufficient window length to detect lower frequencies. To overcome this limitation ARMA modelling of the data is used to encompass the frequency content and allow its quantification.

In a similar approach again, Carden and Brownjohn (2008) presents a SHM program in which an ARMA model is used to parametrise structural responses to loading of a reference structure. The ARMA parameters are then classified and statistically tested to test for significant structural change. Forced excitations of the monitored structures are used in this case to ensure the validity of the statistical test for change.

Erdoğan and Güral (2009) describes a systematic data analysis approach applied to the SHM of the Istanbul Bosphorus bridge. Lateral, longitudinal, and vertical movements are monitored. The data analysis procedure begins by acknowledging the component nature of time series data:

$$P = P_{Trend} + P_{Periodic} + P_{Stochastic} \quad (2.17)$$

in which P is a time series with values such that $P = (p_0, \dots, p_{N-1})$. Thus the corollary in which $P = (p(t_0), \dots, p(t_{N-1}))$:

$$p(t_n) = p(t_n)_{Trend} + p(t_n)_{Periodic} + p(t_n)_{Stochastic} \quad (2.18)$$

The stochastic component of the measured movements are then partially eliminated through use of a WMA:

$$p_s(t_n) = \frac{p(t_n) + 4p(t_{n+1}) + 6p(t_{n+2}) + 4p(t_{n+3}) + p(t_{n+4})}{16} \quad (2.19)$$

Notably, the WMA appears to have a forward bias in time. The trending components of the measurements are then estimated utilising a polynomial fitting via a least squares method:

$$\rho(t_n)_{Trend} = \sum_{j=0}^x c_j t_n^j \quad (2.20)$$

The trending component is then removed, and a FFT applied to quantify the remaining periodic components. With the trending and periodic components estimated, the residual stochastic component of the original data is then able to be studied and parametrised using the autoregressive function and the ARMA model.

A novel tool for EDA of datasets with seasonal variation and interannual variability is presented in Anghileri et al. (2014), the 'Moving Average over Shifting Horizon (MASH)' tool. The MASH tool accommodates datasets where processes are implicitly dependent on season and also significantly variable between years. The tool was applied to hydro-climatic records to quantify trends occurring across 35 years. In this case a moving average across a number of years was generated from a set of moving averages across the same period within each of those years. This allows observation of a smoothed annualised behaviour to be observed, and the changes in that annualised behaviour through the years to be observed as well. This process can effectively bypass the seasonal behaviour and interannual variability, allowing linear regression or Sen's slope trend detection methods to be used across comparative seasonal points.

In a similar vein to the monitoring reported in Maalej et al. (1999), Owen et al. (2001), and Carden and Brownjohn (2008), but undertaken using smaller reference structures, de Lautour and Omenzetter (2010) uses an AR model to represent and parametrise the structural behaviour. Notably, instead of utilising frequency domain analysis or statistical tests to detect changes in the AR model of the measured data, a neural network is used to detect and interpret changes in the stationarity modelled by the AR function.

Sirca and Adeli (2012) identifies a number of approaches to SHM in a review of journal papers published post 1995. Techniques reported include: i) Finite Element Modelling (FEM) used to determine the expected engineering parameters; ii) neural networks and genetic algorithms used to parametrise the structures; iii) frequency domain and other signal processing techniques; and iv) use of chaos theory. Consideration is also given to so called multi-paradigm approaches, where combinations of these techniques are used.

In a paper on the complexities of analysis and interpretation of trend in ozone data, Hogrefe et al. (1998) mentions non-uniformity (temporally and spatially) of the data as an analysis constraint. It is suggested that a typical approach for such analysis is to use a regularly spaced aggregated average. Additionally, parametric trend, seasonal, and AR approaches have been used. Use of a Kolmogorov-Zurbenko filter is presented, which accommodates this non-uniformity by varying the number of values used to calculate a specific datum point, but maintaining the time scale over which physical phenomena are investigated. A variation on this time scale focused approach is detailed in Chapter Four.

Displacement of structural elements in a coal mine, as recorded by Global Positioning System (GPS) data, are analysed in Xiao and Zhang (2011). Analysis techniques utilised include noise filtering and subsequently polynomial, exponential, and periodic fitting, enabling a clearer understanding of the deformation occurring.

2.5.4 Non-uniformity in data analysis

A tendency for time series analysis literature to implicitly focus on the analysis of uniformly spaced datum points has become evident during this review. For example, neither the autocorrelation nor the ARMA functions that stochastically model a time series would maintain their physical meaning if the lags and steps incorporated into their formulations were to dynamically change. A number of texts go so far as to explicitly exclude non-uniformity from their consideration. Bloomfield (2000), Chatfield (1989), Box and Jenkins (1976), Oppenheim and Schaffer (1975), and Anderson (1971) all possess a single line or paragraph of text limiting the scope of the material they present to uniformly spaced data.

In some cases more attention is paid to the issue of non-uniformity, for example Hannan et al. (1985) devotes a chapter to time series analysis with unequally spaced data. However the focus of the chapter is towards the representation of the data using a space-state representation, rather than accommodation of the non-uniformity in an un-modelled methodology. Diggle (1990) explicitly acknowledges the possibility of non-uniform time series while defining the notation used throughout the book, however the analysis techniques described within are presented almost exclusively in uniformly spaced notation.

Formulations focused specifically on the generalities of unevenly spaced time series analysis

are presented in the unpublished papers Eckner (2013, 2012). The papers collectively present definitions and algorithms that pertain to an unevenly sampled time series' inter-observation nature, i.e. whether the spaced between observations are considered constant or linearly varying, and the options regarding bounding of windows passed across such a time series.

Texts on specific techniques relied upon later in this document that include reference to non-uniform data include Golyandina et al. (2001), and Bagchi and Mitra (1999). The SSA process described in Golyandina et al. (2001) is stated not to require uniformly sampled input data. There are however no mechanisms for accounting for a varying sample rate in the procedures described. Similarly, Bagchi and Mitra (1999) describes a DFT process that is not uniformly spaced in the frequency domain, but does not accommodate non-uniformity in the input time series. Hogrefe et al. (1998), as mentioned above, accommodates non-uniform data (spatially and temporally) by way of a regular average derived from an aggregate of proximate datum points. A variation on this general approach is described in Chapter Four.

2.6 Summary

The review of literature presented in Sections 2.2 to 2.3 of this chapter have focused on the basic properties of montmorillonite/bentonite (or where appropriate similar materials) and its expected fluid transport mechanisms, specifically gas transport. Four possible gas transport mechanisms have been noted, gas diffusion, visco-capillary flow, dilatant flow, and fracture flow. The dilatant flow mechanism has been anticipated to be the dominant flow mechanism for a free gas phase in the conditions expected within a radioactive waste repository. The modelling work on dilatant gas flow that was reviewed does not yet predict the pathway formation observed experimentally. This may be in part due to competing hypotheses regarding the dilational mechanism and subsequent flow.

Gas flow experimentation identified a number of behaviours associated with dilatant flow, including a breakthrough and shut-in phenomena, instability of flow pathways, sensitivity to boundary conditions, and potential sensitivity to scale. The unknowns regarding the rigidity of the specific boundary conditions of an in-situ disposal emplacement that would be provided by a host rock, and the availability of only laboratory scale gas flow experimentation constitute a significant information gap in this regard.

Within Section 2.4, note was made of long-term, large-scale undertakings and the implicit volume and (non-)uniformity of the expected data acquisition associated with them. A cursory overview of generalised data analysis texts reviewed in Section 2.5 suggests that accommodation of such non-uniformity during an EDA may be complex, with techniques that directly accommodate the non-uniformity absent in favour of techniques that aggregate uniform data from a non-uniform source.

Chapter Three

The Lasgit experiment

3.1 Introduction

In order to investigate the information gap identified in Chapter Two, regarding the effects of scale and boundary conditions on gas flow in compacted clay materials, the Large Scale Gas Injection Test (Lasgit) was undertaken.

As a necessary prelude to the data analysis and scientific interpretation of the dataset arising from Lasgit, a detailed description of the experimental set-up, along with a summary of the operational history and experimental results of Lasgit is presented in this chapter.

The information aggregated in this chapter is gathered from a selection of journal articles, summary reports, and technical reports. It is intended to provide a single point of reference that enables contextualisation of the results and discussion of the analysis of the arising dataset that are presented in latter chapters.

3.2 Experimental overview

Lasgit is a field-scale experiment located at approximately 420 m depth in Svensk Kärnbränslehantering AB (SKB)'s Äspö Hard Rock Laboratory (HRL) in Sweden. It aims to investigate the potential impact of gas build up and flow on the Engineered Barrier System (EBS) in a deep geological radioactive waste repository environment, specifically the Swedish three layer Kärnbränslesäkerhet [Nuclear Fuel Safety] (KBS) disposal concept for high level radioactive waste, known as the KBS-3 concept (Cuss et al. 2011; SKB 2006).

During its operational life Lasgit has formed part of the Near Field Processes (NF-Pro) and Fate Of Repository Gases (Forge) European Research and Development (R&D) projects. Assembly,

maintenance, and operation of Lasgit is performed jointly by SKB, Sweden and the British Geological Survey (BGS), United Kingdom (UK).

Within the SKB HRL facility, Lasgit is located in deposition hole DA3147G01, a borehole drilled in 1999 approximately 8.5 m in length and 1.75 m in diameter. DA3147G01, is in turn, located in the floor of Tunnel Boring Machine (TBM) assembly hall. Lasgit's data acquisition and gas injection/hydration control systems are located in a gas laboratory adjacent to DA3147G01 in the TBM assembly hall. Figure 3.1 depicts the interior of the TBM assembly hall, with the (blue) gas laboratory on the left, and the retaining lid of Lasgit on the right.

The gas laboratory is mechanically automated and accessible via the internet, allowing remote control of the experimental set-up and access to data as it is recorded. Day-to-day operation of Lasgit is performed by the BGS using the described method of access.

3.2.1 General arrangement

Lasgit's experimental set-up, based on the KBS-3 repository concept, includes a full scale copper waste canister with a compacted bentonite EBS emplaced in a vertical borehole, in a granite host rock, at a depth of several hundred metres from the surface. A schematic overview of the experimental arrangement is shown in Figure 3.2. The copper canister is shown within DA3147G01 along with retaining arrangement, filter mats at the outer edge, and various sensor/filter locations indicated by the coloured shapes.

Repository backfill conditions are simulated via the retaining arrangement, comprising of a concrete plug retained by a steel lid anchored into the host rock (Bäck 2003). Filters were installed onto the surface of the canister to provide point-like gas source locations, mimicking potential points of canister breach that could lead to water infiltration and therefore anaerobic iron corrosion and gas generation occurring. Figure 3.3 depicts the locations of injection filters fitted to the canister surface. Some 150 sensors were also installed in and around the experimental emplacement. Section 3.2.4 details the location, type, and notation of each of these sensors.

At emplacement, pre-fabricated compacted bentonite blocks, rings and pellets were used to form the EBS. The manufacture of these components is described in detail in Johannesson et al. (1995), and Johannesson (2003). Figure 3.4 depicts a cross section of the initial experimental



Figure 3.1: TBM assembly hall interior. Gas laboratory (left) and Lasgit's retaining lid with anchor cables (right) (Cuss et al. 2010).

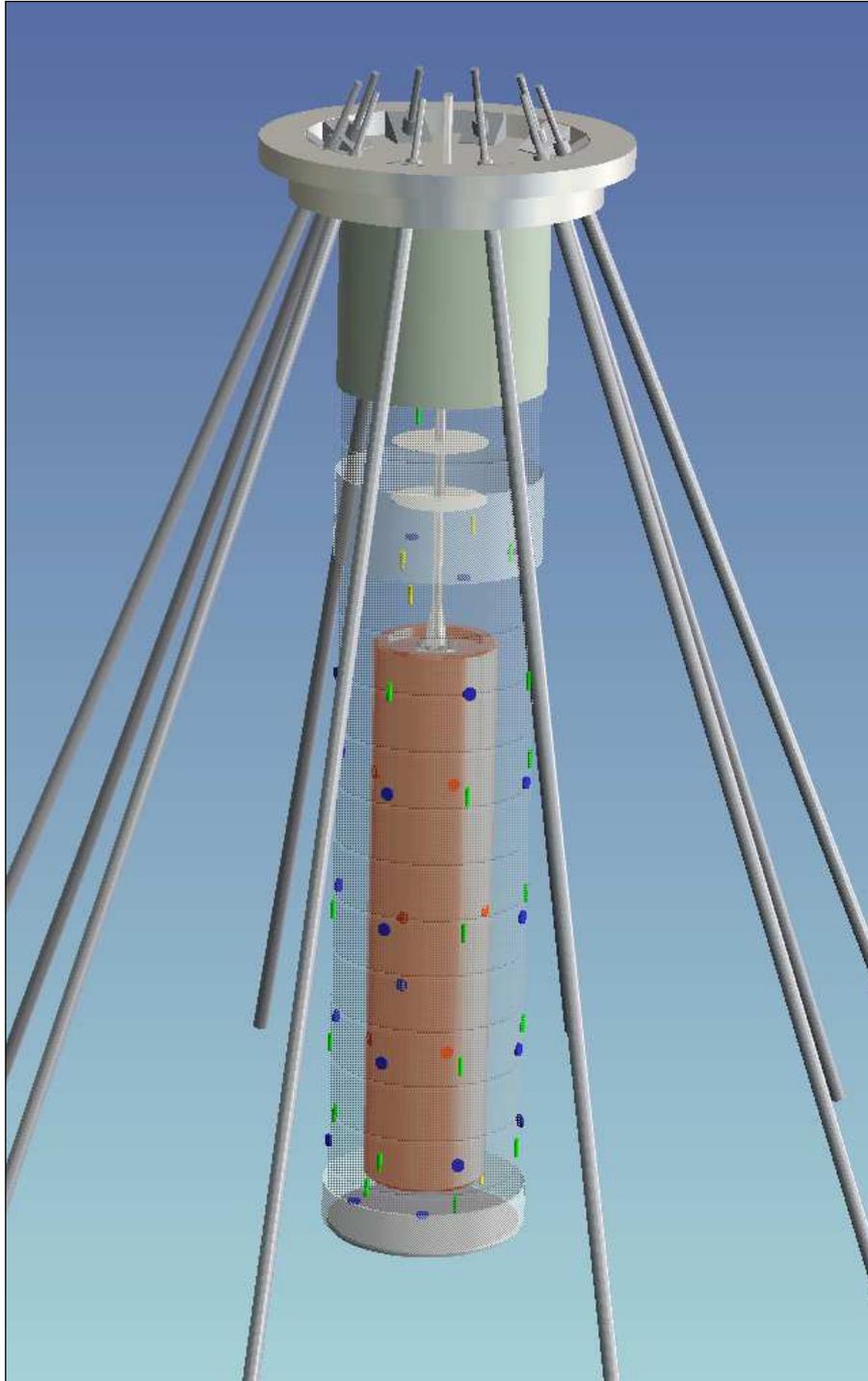


Figure 3.2: Schematic depiction of Lasgit experimental arrangement (SKB 2008).

emplacement, from the central axis of the emplacement outwards. Tolerances for emplacement of the apparatus were accommodated by a 50 mm void between the host rock and the EBS, and a 10 mm void between the EBS and the canister. The outer (50 mm) void was then filled with bentonite pellets. The inner (10 mm) void was left to close due to swelling of the bentonite upon (re)hydration. In the case of Lasgit, hydration took place both naturally and artificially before each gas injection campaign over the (to date) eight year operational life of the experiment (Cuss et al. 2010).

3.2.2 Scenario and aims

Specifically, Lasgit's experimental scenario replicates the post-thermal phase (circa 10,000 years or more post waste emplacement) of a KBS-3 concept repository. This scenario implies a fully (re)saturated EBS under isothermal conditions, and satisfies the criteria for gas generation occurring: i.e. sufficient availability of water and enough time for canister failure to have occurred. As such, the scenario represents a potential repository situation in which the copper canister has failed in the presence of available water (via the saturated EBS), allowing corrosion of the iron insert and subsequent gas generation and build up to occur (SKB 2006).

Specific experimental aims include the answering of questions regarding i) the sensitivity of gas flow within a bentonite based EBS to boundary conditions, ii) possible scale dependencies of the gas flow mechanism within the EBS, and iii) to provide high quality field scale data regarding (re)hydration of the EBS (Cuss et al. 2011).

Specific questions regarding gas flow in a compacted bentonite EBS when considering the boundary condition/scale dependant framework which have been attempted to be answered by Lasgit include:

- What is the minimum pressure gas will enter the clay?
- What mechanisms control gas entry?
- How much pore-water is displaced by the gas?
- Is flow along preferential pathways or through the material matrix?
- What mechanisms control the direction of flow?

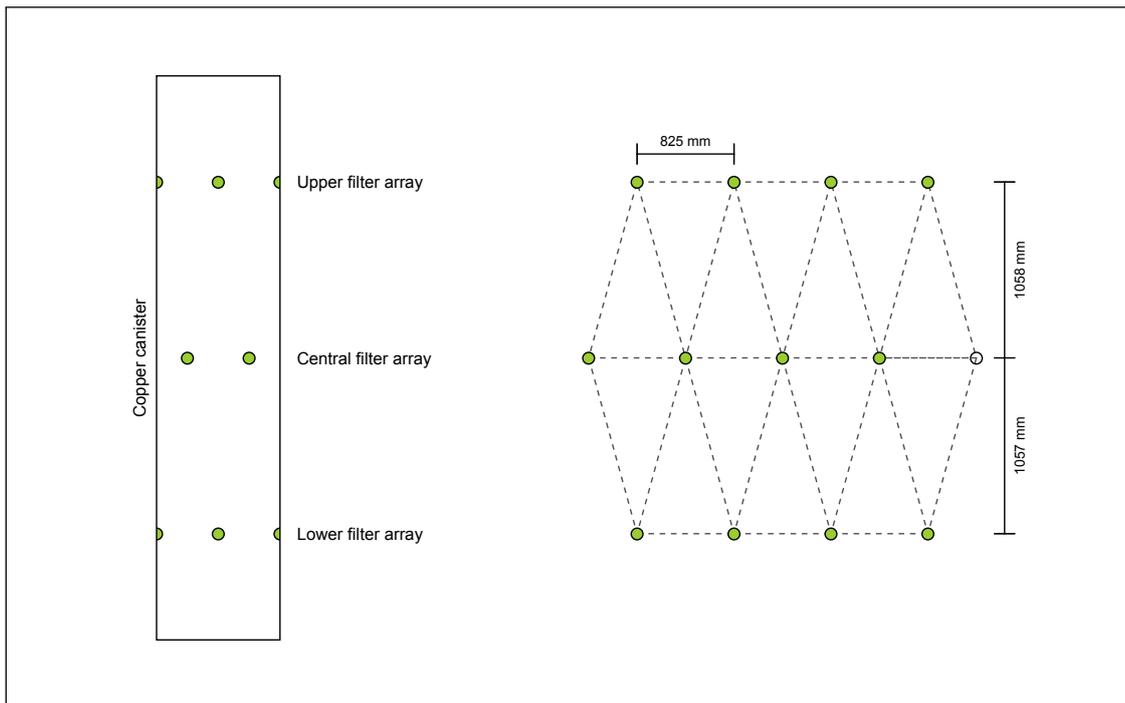


Figure 3.3: Filter array orientation on canister surface (after Harrington et al. 2008). The central filter array is rotated by 45° on the canister surface relative to the upper and lower filter arrays, producing a triangular mesh when unwrapped.

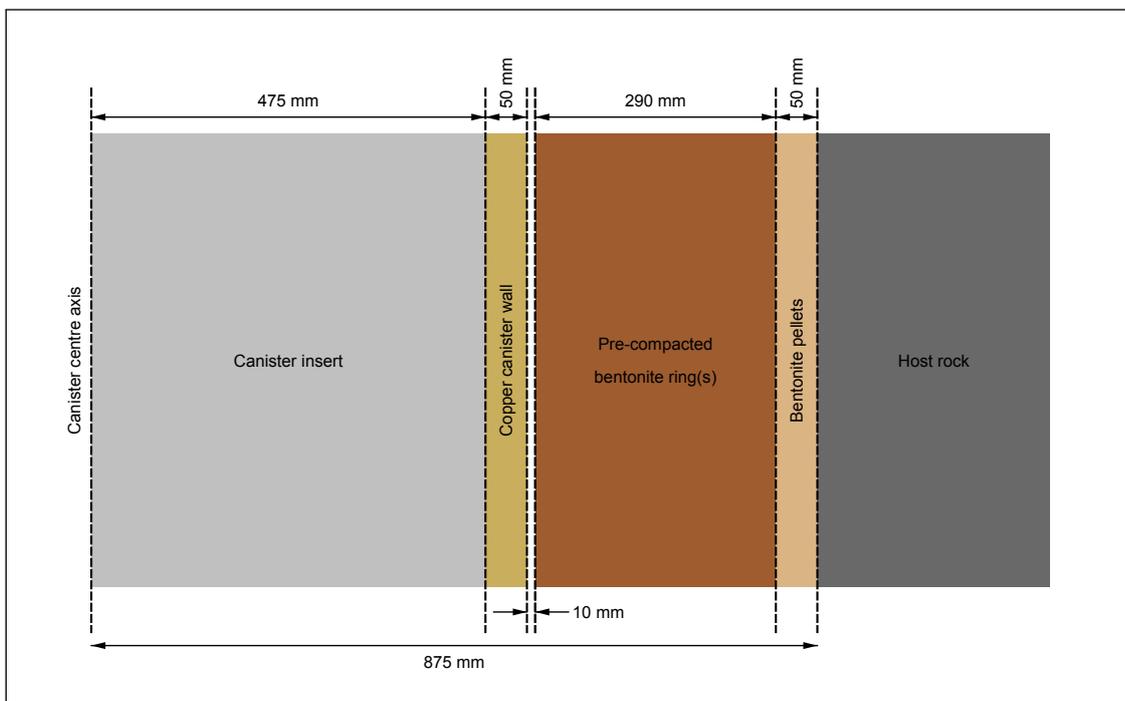


Figure 3.4: Cross section of experimental arrangement at emplacement (after Cuss et al. 2010).

- Is gas flow episodic?
- Will gas migration affect barrier performance?
- Is the gas pressure transmitted to the host rock?
- What is the maximum gas pressure possible in the EBS?

3.2.3 Deposition hole DA3147G01

The host rock, that contains DA3147G01, is predominantly Äspö diorite (Hardenby and Lundin 2003). DA3147G01 was drilled in 1999 with a diameter of 1.75 m and to a depth of approximately 8.5 m (Hardenby and Lundin 2003), although subsequent concrete castings at the top of the hole have increased its effective depth to 9.1 m (Sandén 2003).

Extensive geological mapping and analysis of the hydrology of both the TBM assembly hall and DA3147G01 have been conducted (Hardenby and Lundin 2003). The hydrological study performed prior to the experiment estimated inflow of water from the host rock into DA3147G01 to be approximately 240–260 litres/day (Hardenby 2004). Conductivity of some of the fractures residing on the wall of DA3147G01 was also confirmed through gas leakage testing (Hardenby 2004). Figure 3.5 depicts the mapped fractures on the wall and base of DA3147G01. Figure 3.6 provides orientation for the fracture map with respect to the experimental apparatus emplaced within DA3147G01.

3.2.4 Experimental apparatus

Lasgit is highly instrumented, possessing approximately 150 sensors in and around DA3147G01. The specifics pertaining to the location, type, and the associated notation of each of these sensors is presented in this section.

Instrumentation within the deposition hole consists of 32 total stress sensors (oriented either radially or axially), 26 pore-water pressure sensors, and 56 temperature sensors¹. Additionally there are 16 pore-water pressure transducers associated with the hydration mats and injection filters that act as sensors. Additionally, attached to the retaining lid, there are seven displacement

¹There are also seven relative humidity sensors located within DA3147G01, however the data arising from them is not available to this study.

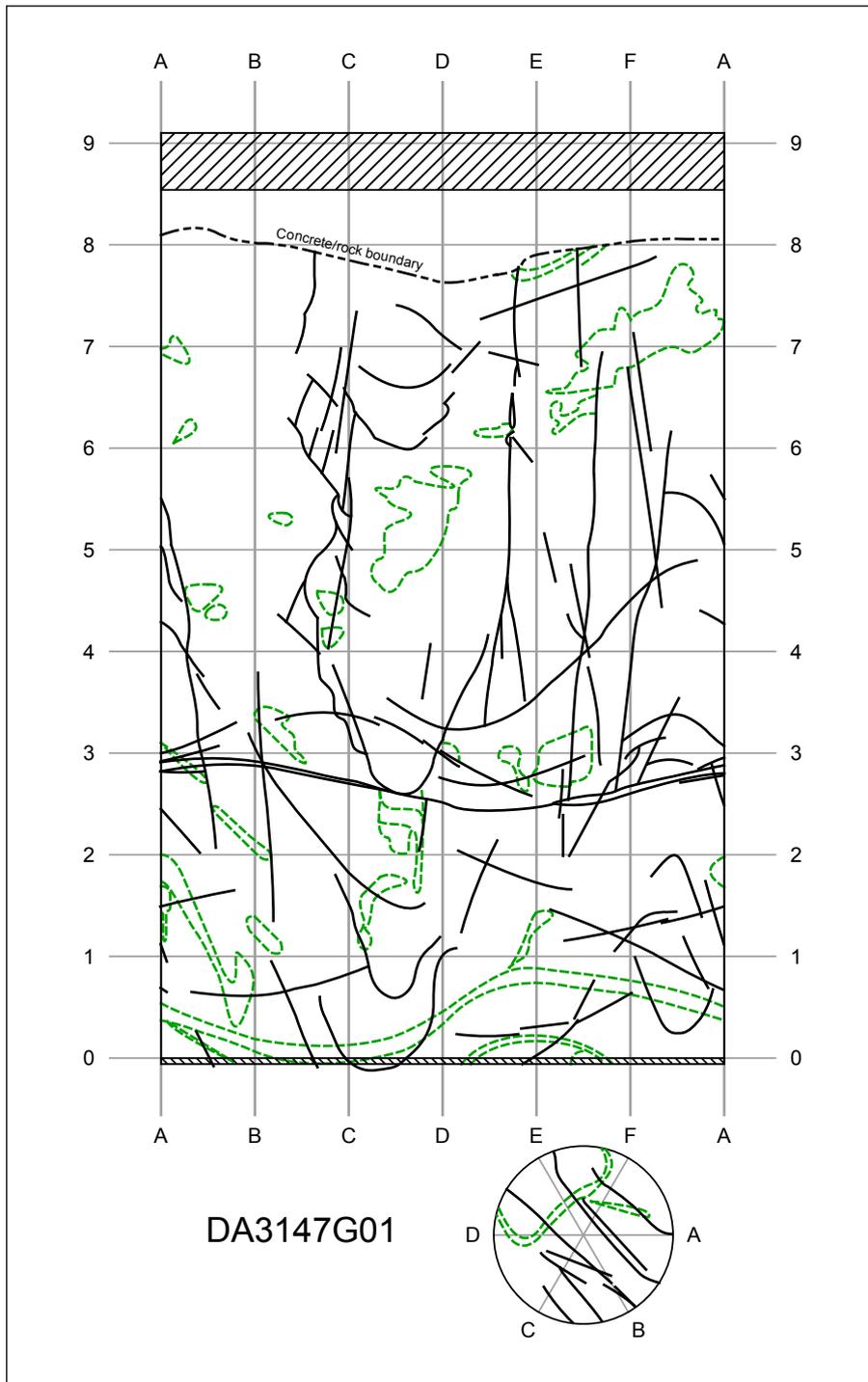


Figure 3.5: DA3147G01 fracture map (after Hardenby and Lundin 2003). Depiction of a screed application at the bottom of the hole (mentioned in a hand written note in Sandén (2003)) is shown via the lower hatching. A concrete frame surrounding the hole at the top (increasing the effective depth) is shown via the upper hatching. Vertical measurements are taken upwards from the screed in metres.

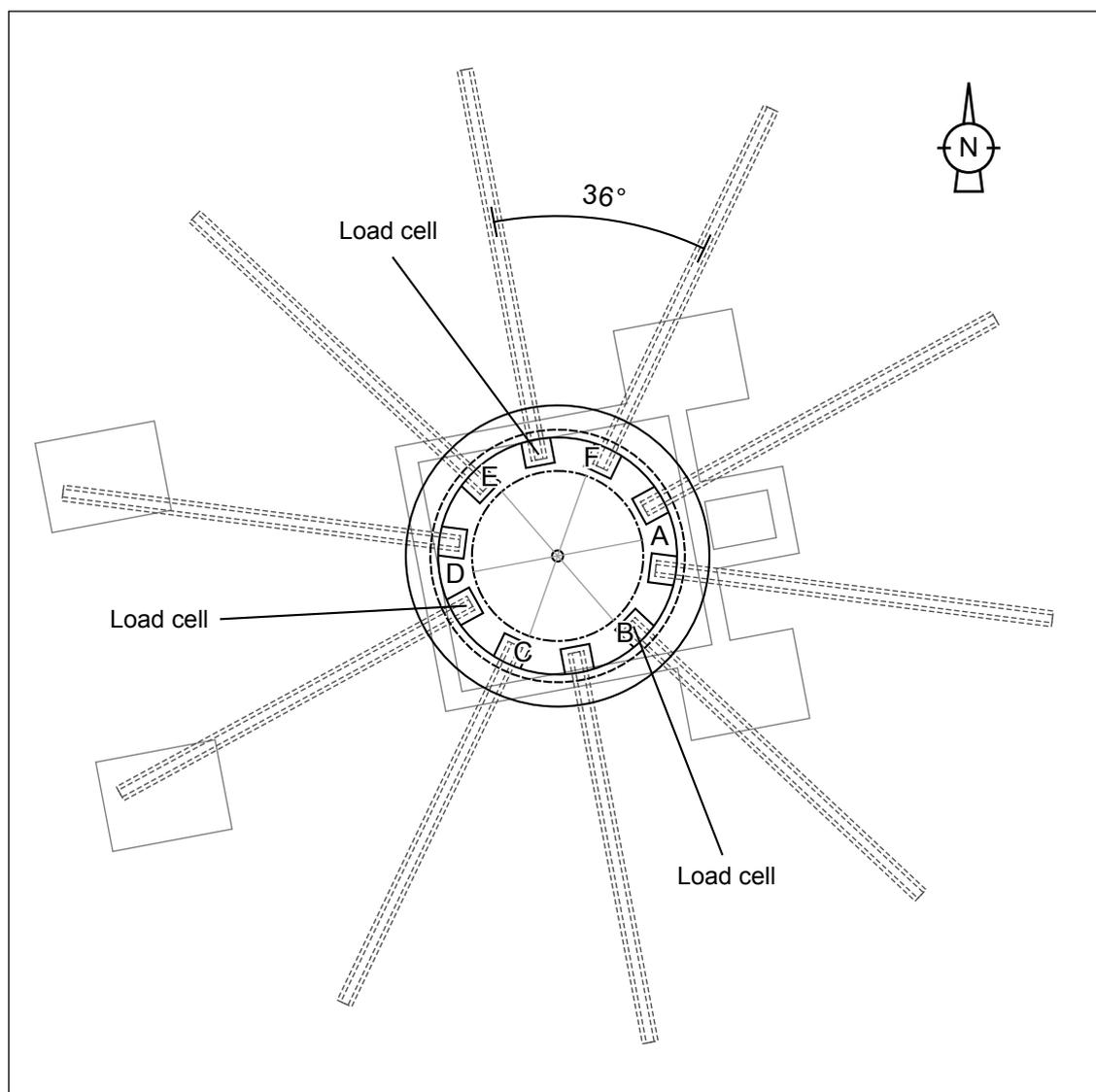


Figure 3.6: Orientation of deposition hole, anchor cables, load cells, and fracture map grid (depicted by the lettering A-F) (after Bäck 2003).

sensors (five vertical and two lateral) and three load sensors. Nine pore-water pressure sensors are located in 'pressure relief boreholes' adjacent to the deposition hole¹. The ambient air temperature of the facility at the experimental location is also monitored. Figure 3.7 schematically shows the experimental arrangement in cross section with measuring sections for each set of sensors/filters/major component.

A full list of sensor and filter notations and locations is presented in Tables 3.1 to 3.8 (summarised from Sandén 2003). For sensors within DA3147G01 heights are measured upwards from the screed at the base of the hole, radii are measured outwards from the vertical central axis of DA3147G01, and angles are measured in degrees anti-clockwise in plan view from 'A' about the centre of DA3147G01 (Figure 3.6). Other sensor locations are given descriptively in their respective tables.

A number of parameters pertaining to the gas laboratory are also recorded (e.g. pressures, volumes, and flow rates in control systems), details of which are available in Harrington et al. (2008), and Cuss et al. (2010). These are typically used for diagnostic purposes and the time series derived from them are not referenced in this work.

The presentation of sensor information is grouped in the following subsections by sensor type, namely: i) stress sensors; ii) pore-water pressure sensors; iii) canister injection filters; iv) hydration filter maps; v) pressure relief hole pore-water pressure sensors; vi) lid displacement sensors; vii) anchor cable load sensors; and viii) ancillary temperature sensors.

3.2.4.1 Stress sensor notations and locations

Table 3.1 details the stress sensor notations and locations within DA3147G01. The prefix PB signifies sensor emplacement within bentonite measuring stress axially, PR prefix signifies sensor emplacement at bentonite/rock wall interface measuring stress radially (tangential to the rock wall), PC prefix signifies sensor emplacement at canister surface/bentonite interface (varying radially and vertically in orientation). Figure 3.8 presents the sensor records arising from the axially oriented stress sensors, and Figure 3.9 presents the sensor records arising from the radial stress sensors.

¹The pressure relief holes were drilled after the installation of the experimental equipment in DA3147G01. Section 3.3 details their drilling and subsequent closure with packers.

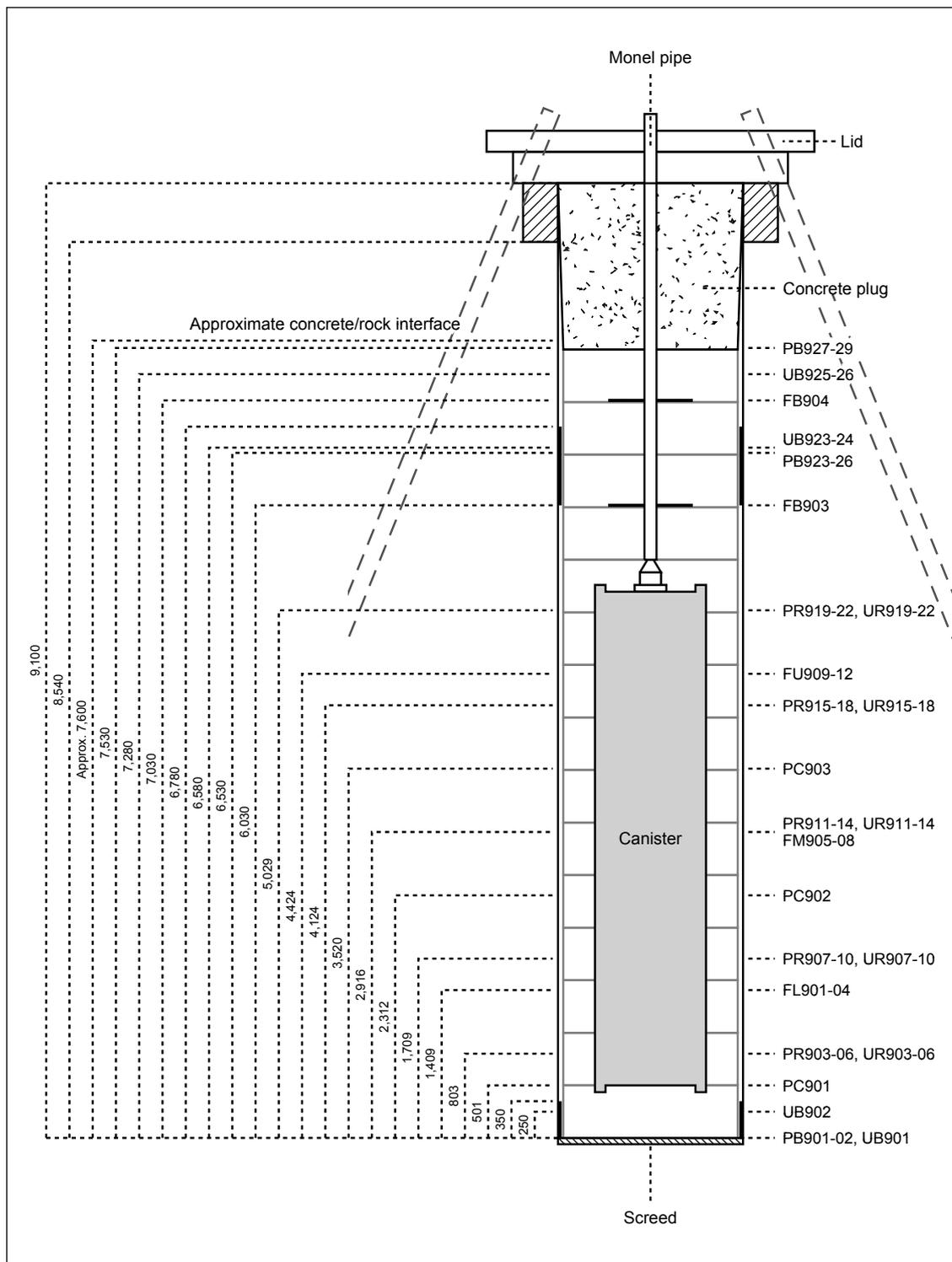


Figure 3.7: Schematic depiction of DA3147G01 experimental arrangement (after Bäck 2003; Sandén 2003).

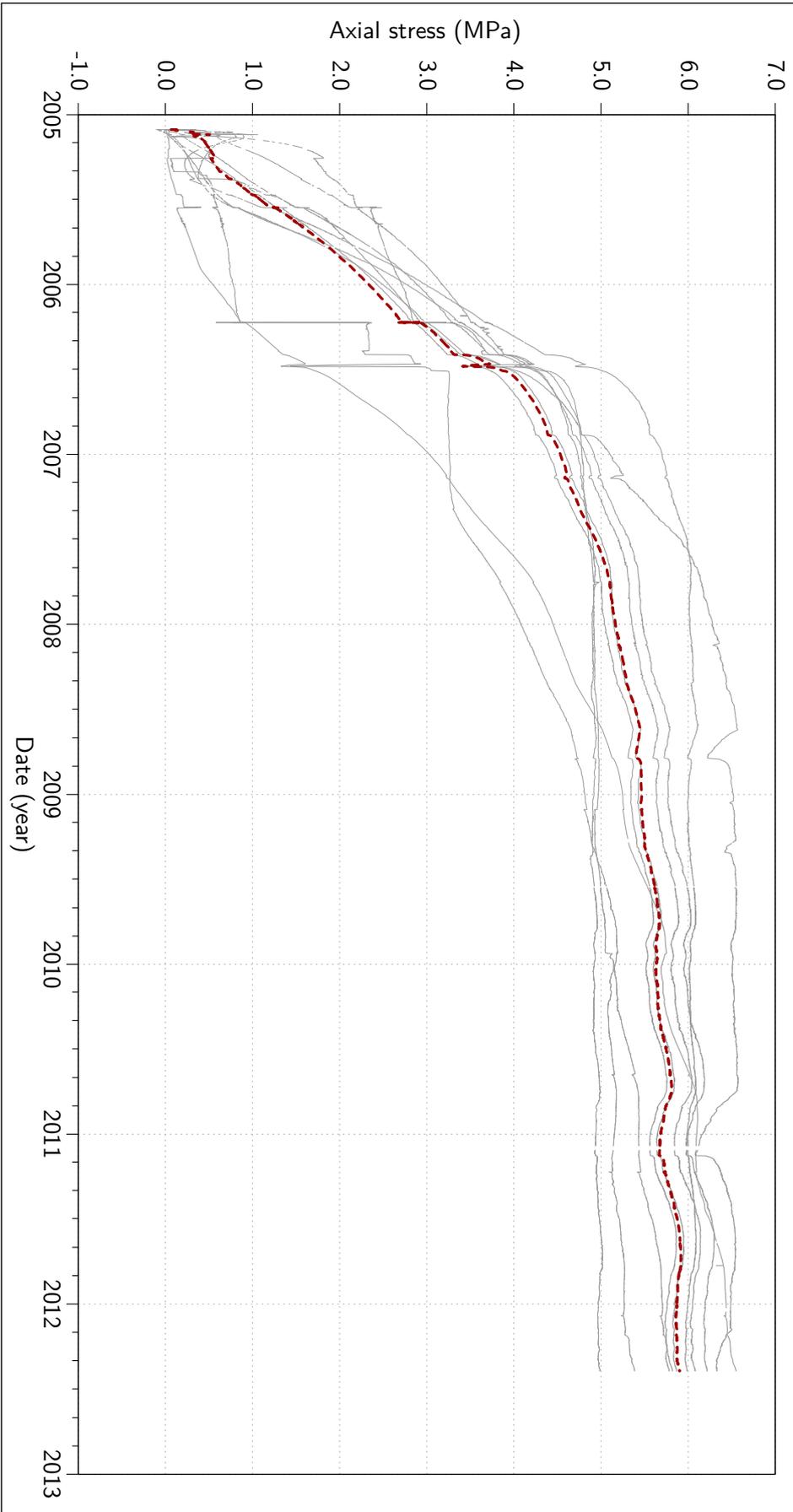


Figure 3.8: Summary of axial stress evolution in DA3147G01. Individual axial stress sensors in grey, average value in red.

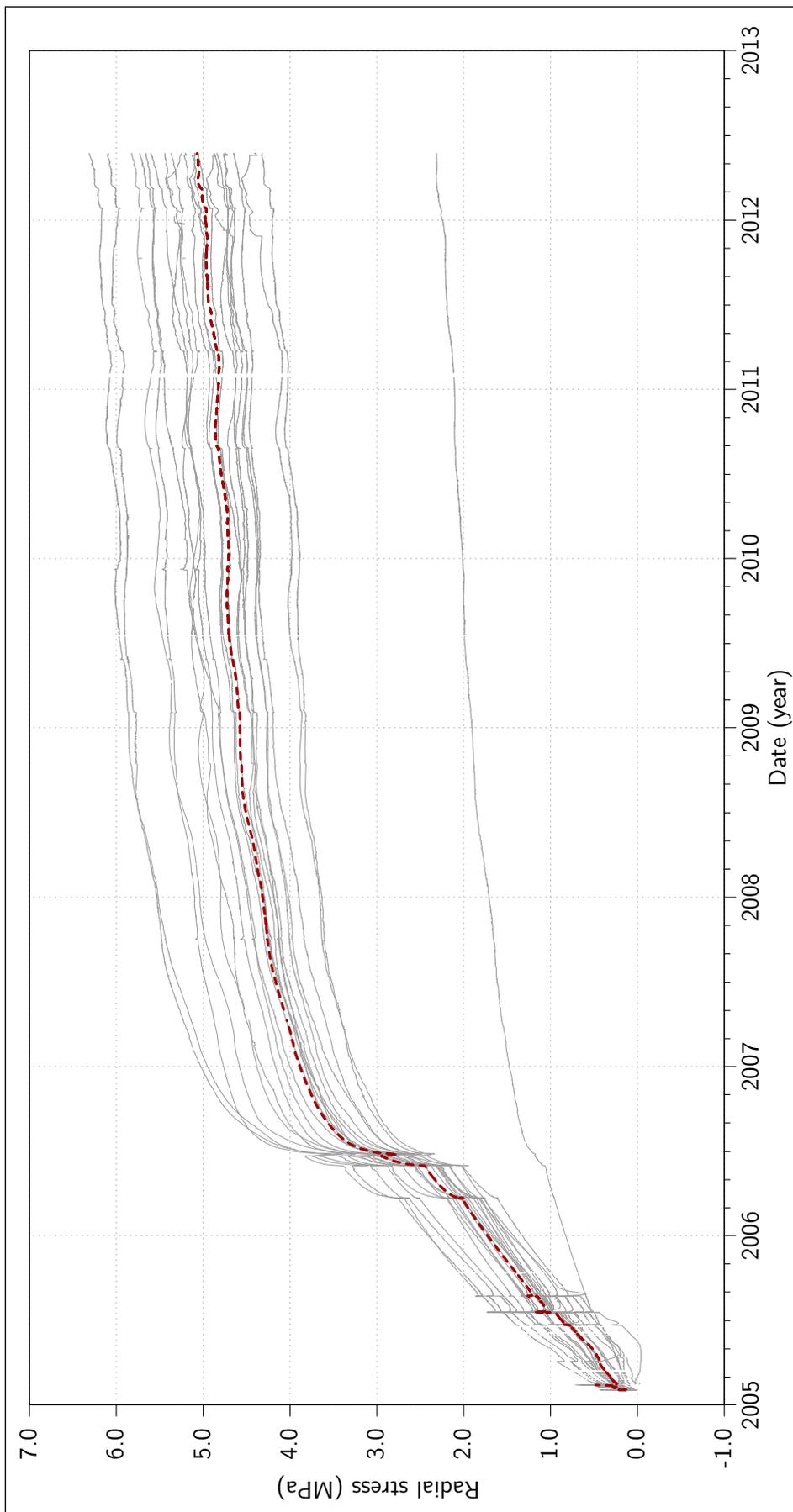


Figure 3.9: Summary of radial stress evolution in DA3147G01. Individual radial stress sensors in grey, average value in red.

The PB and PR stress sensors are manufactured by Geokon (model 4800-1X), and have an operating range of 0–20 MPa with an accuracy of 0.1% of full scale. Temperature sensors (thermocouples) are also incorporated into each of the PB and PR sensors, accounting for 29 of the 56 temperature sensors within DA3147G01. The temperature sensors are referred to using the prefix PT with the corresponding stress sensor number, and implicitly have the corresponding locations. The PC sensors, which are manufactured by Druck, possess an operating range of approximately 0–13 MPa and an accuracy similar to the Geokon stress sensors of 0.075–0.2% of full scale. The PC sensors, which do not have an incorporated temperature sensor, are marked with an asterisk in Table 3.1.

3.2.4.2 Pore-water pressure sensor notations and locations

Table 3.2 details the pore-water pressure sensor notations and locations within DA3147G01. The prefix UB signifies sensor emplacement within bentonite, UR prefix signifies sensor emplacement at the bentonite/rock wall interface. Figure 3.10 presents the sensor records arising from the pore-water pressure sensors at the bentonite/rock-wall interface and Figure 3.11 presents the sensor records arising from the in-bentonite pore-water pressure sensors. A large increase in reported pressure is present in UB902 at approximately 2010, skewing the scale of the remaining in-bentonite pore-water pressure sensors. For clarity, an additional figure, Figure 3.12, reproduces the data presented in Figure 3.11 *omitting sensor UB902*.

The pore-water pressure sensors are manufactured by Geokon (model 4500S), and have an operating range of 0–20 MPa with an accuracy of 0.1% of full scale. Temperature sensors (thermocouples) are also incorporated into each of the UB and UR sensors, accounting for a further 26 of the 56 temperature sensors within DA3147G01. The temperature sensors are referred to using the prefix UT with the corresponding pore-water pressure sensor number, and implicitly have the corresponding locations.

3.2.4.3 Injection filter notations and locations

Table 3.3 details the canister injection filter notations and locations. The prefix FU signifies canister filters in the upper array, FM prefix signifies canister filters in the middle array, FL prefix signifies canister filters in the lower array (see Figure 3.3). The radius 525 mm corresponds

Table 3.1: Lasgit stress sensor notations and locations. Sensors without an incorporated temperature sensor are marked with an asterisk.

| Sensor ID | Orientation | Sensor Position | | |
|-----------|-------------|-----------------|-------------|-------------|
| | | Angle (°) | Radius (mm) | Height (mm) |
| PB901 | Axial | 0 | 0 | 0 |
| PB902 | Axial | 225 | 675 | 0 |
| PC901* | Axial | 0 | 0 | 501 |
| PR903 | Radial | 0 | 875 | 803 |
| PR904 | Radial | 90 | 875 | 803 |
| PR905 | Radial | 180 | 875 | 803 |
| PR906 | Radial | 270 | 875 | 803 |
| PR907 | Radial | 45 | 875 | 1709 |
| PR908 | Radial | 135 | 875 | 1709 |
| PR909 | Radial | 225 | 875 | 1709 |
| PR910 | Radial | 315 | 875 | 1709 |
| PC902* | Radial | 315 | 525 | 2312 |
| PR911 | Radial | 45 | 875 | 2916 |
| PR912 | Radial | 135 | 875 | 2916 |
| PR913 | Radial | 225 | 875 | 2916 |
| PR914 | Radial | 315 | 875 | 2916 |
| PC903* | Radial | 135 | 525 | 3520 |
| PR915 | Radial | 45 | 875 | 4124 |
| PR916 | Radial | 135 | 875 | 4124 |
| PR917 | Radial | 225 | 875 | 4124 |
| PR918 | Radial | 315 | 875 | 4124 |
| PR919 | Radial | 0 | 875 | 5029 |
| PR920 | Radial | 90 | 875 | 5029 |
| PR921 | Radial | 180 | 875 | 5029 |
| PR922 | Radial | 270 | 875 | 5029 |
| PB923 | Axial | 0 | 675 | 6530 |
| PB924 | Axial | 90 | 675 | 6530 |
| PB925 | Axial | 180 | 675 | 6530 |
| PB926 | Axial | 270 | 675 | 6530 |
| PB927 | Axial | 90 | 250 | 7530 |
| PB928 | Axial | 135 | 675 | 7530 |
| PB929 | Axial | 315 | 675 | 7530 |

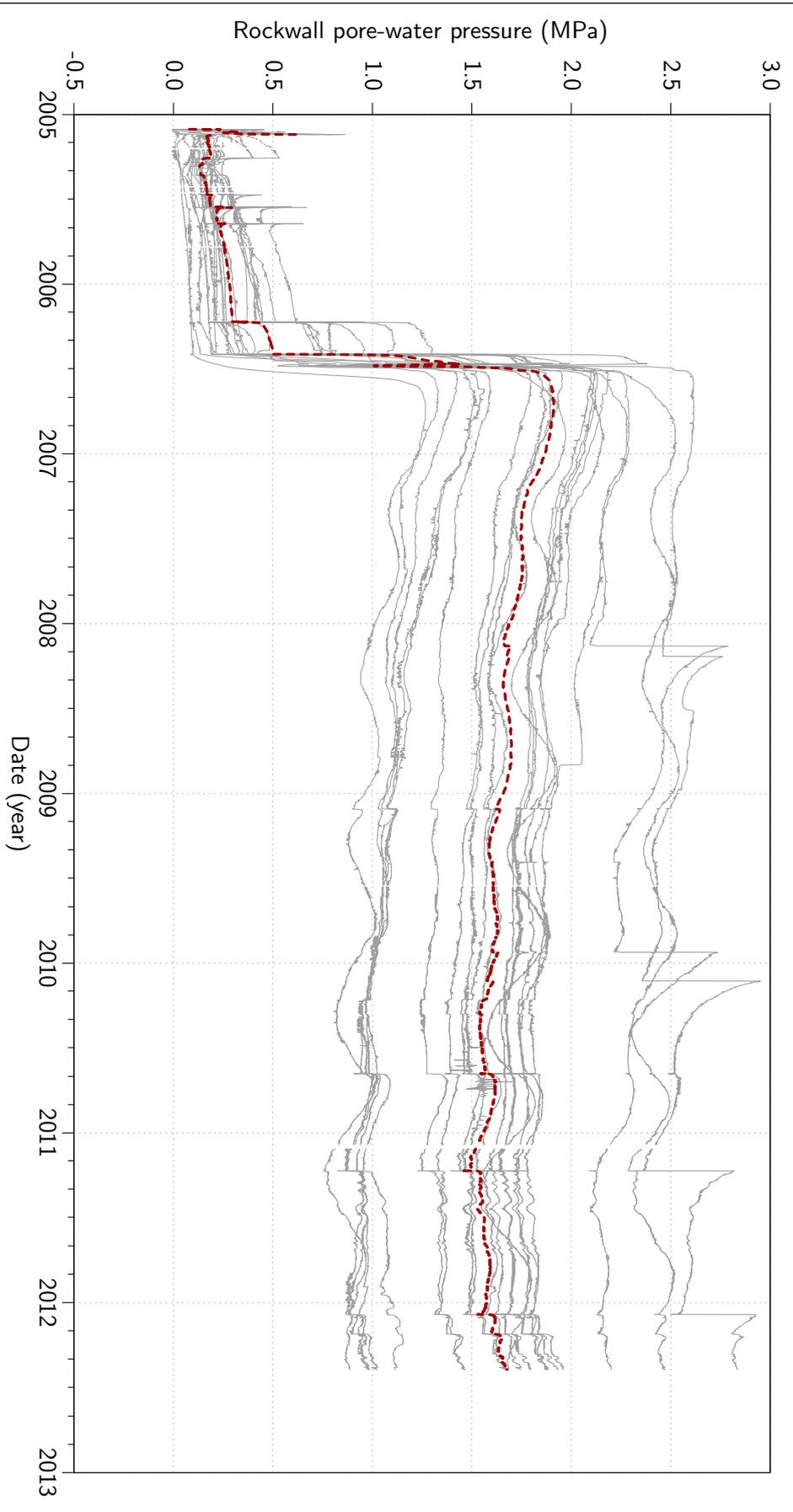


Figure 3.10: Summary of rockwall/bentonite interface pore-water pressure evolution in DA3147G01. Individual UR sensors in grey, average value in red.

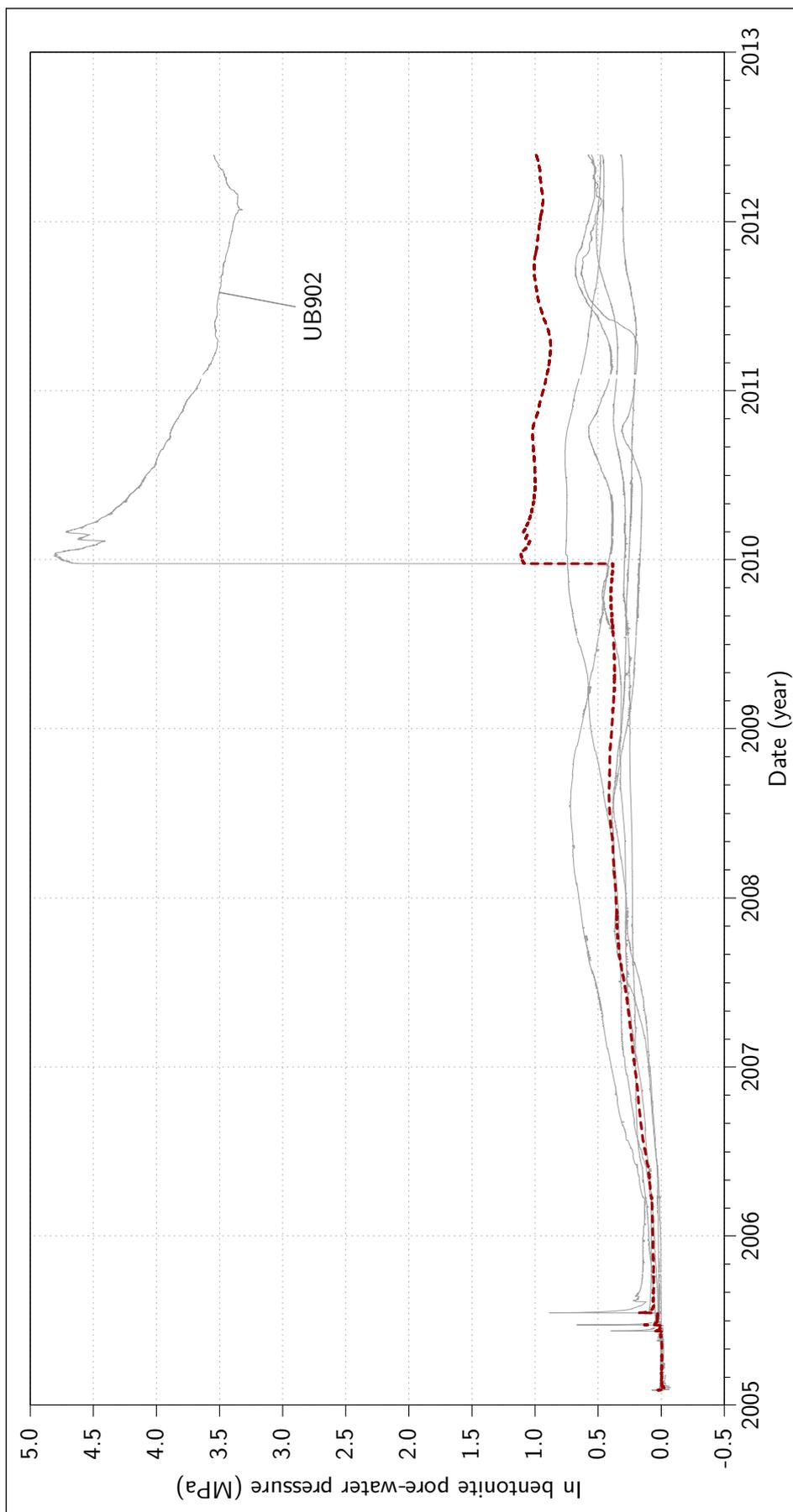


Figure 3.11: Summary of in-bentonite pore-water pressure evolution in DA3147G01. Individual UB sensors in grey, average value in red.

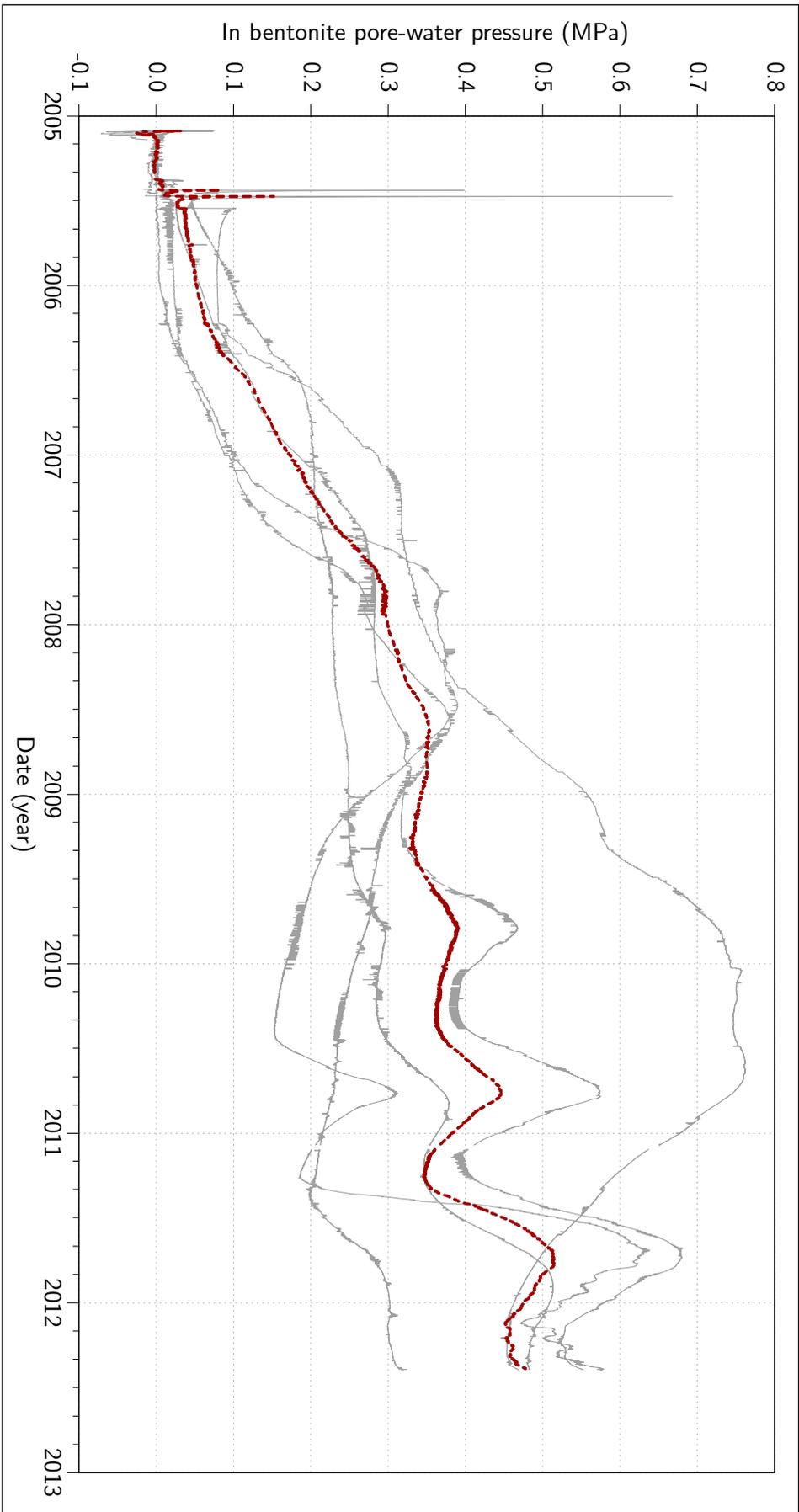


Figure 3.12: Summary of in-bentonite pore-water pressure evolution in DA3147G01. Individual UB sensors in grey, average value in red. Summary excludes UB902.

Table 3.2: Lasgit pore-water pressure sensor notations and locations.

| Sensor ID | Location | Sensor Position | | |
|-----------|-----------|-----------------|-------------|-------------|
| | | Angle (°) | Radius (mm) | Height (mm) |
| UB901 | Bentonite | 90 | 300 | 0 |
| UB902 | Bentonite | 270 | 437.5 | 250 |
| UR903 | Rock wall | 45 | 875 | 803 |
| UR904 | Rock wall | 135 | 875 | 803 |
| UR905 | Rock wall | 225 | 875 | 803 |
| UR906 | Rock wall | 315 | 875 | 803 |
| UR907 | Rock wall | 0 | 875 | 1709 |
| UR908 | Rock wall | 90 | 875 | 1709 |
| UR909 | Rock wall | 180 | 875 | 1709 |
| UR910 | Rock wall | 270 | 875 | 1709 |
| UR911 | Rock wall | 0 | 875 | 2916 |
| UR912 | Rock wall | 90 | 875 | 2916 |
| UR913 | Rock wall | 180 | 875 | 2916 |
| UR914 | Rock wall | 270 | 875 | 2916 |
| UR915 | Rock wall | 0 | 875 | 4124 |
| UR916 | Rock wall | 90 | 875 | 4124 |
| UR917 | Rock wall | 180 | 875 | 4124 |
| UR918 | Rock wall | 270 | 875 | 4124 |
| UR919 | Rock wall | 45 | 875 | 5029 |
| UR920 | Rock wall | 135 | 875 | 5029 |
| UR921 | Rock wall | 225 | 875 | 5029 |
| UR922 | Rock wall | 315 | 875 | 5029 |
| UB923 | Bentonite | 45 | 675 | 6580 |
| UB924 | Bentonite | 225 | 675 | 6580 |
| UB925 | Bentonite | 135 | 440 | 7280 |
| UB926 | Bentonite | 315 | 440 | 7280 |

to the radius of the copper canister. Figure 3.13 presents the pore-water pressures induced in the canister filters. The pressure transducers connected to the injection filters are Hi-Pro Model 1000 Series transducers, and possess an operating range of 0–20 MPa with an accuracy of 0.25% of full scale.

3.2.4.4 Filter mat notations and locations

Table 3.4 details the EBS hydration filter mat notations and locations. The prefix FB signifies circular hydration mats placed horizontally within the bentonite EBS, and FR prefix signifies radial hydration mats placed vertically around the deposition hole at the bentonite/rock wall interface. Figure 3.14 presents the pore-water pressures induced in the hydration filter mats. The pressure transducers connected to the filter mats are Hi-Pro Model 1000 Series transducers, and possess an operating range of 0–20 MPa with an accuracy of 0.25% of full scale. The filter mats have a nominal thickness of 2 mm.

3.2.4.5 PRH pore-water pressure sensors notations and locations

Table 3.5 details the pore-water pressure sensor notations and locations in the pressure relief holes. Pore-water pressure sensors located within the pressure relief holes adjacent to DA3147G01 are located in the intervals between the packer installations. Precise depths are not available, however depth within the pressure relief holes is increasing in the order presented. Figure 3.15 presents the pore-water pressure sensor records arising from the sensors within the pressure relief holes PRH-1 and PRH-2.

3.2.4.6 Displacement sensor notations and locations

Table 3.6 describes the linear displacement sensor notations and locations. Displacements are measured relative to the TBM assembly hall floor and ceiling. Figures 3.16 and 3.17 present the vertical linear displacement measurements of the retaining lid and monel pipe. Figure 3.18 presents the lateral linear displacements of the retaining lid. The displacement sensors are manufactured by Solartron, and have been set-up with an operating range of -3.6 mm to 3.6 mm. The resolution of the displacement sensors is asserted to be $< 0.1 \mu\text{m}$.

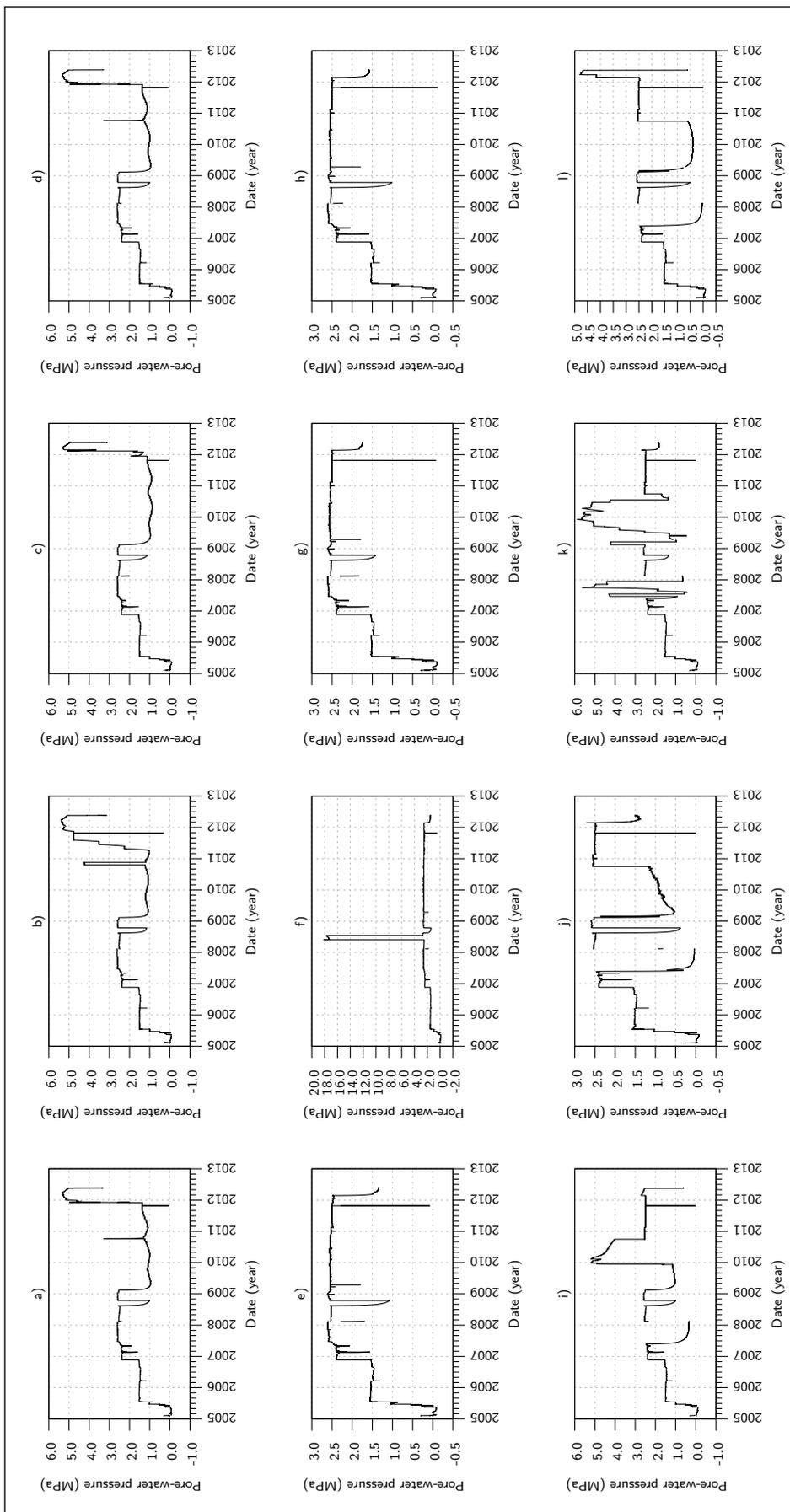


Figure 3.13: Summary of pore-water pressures induced in canister filters.

Table 3.3: Canister injection filter notations and locations.

| Filter ID | Filter Position | | |
|-----------|-----------------|-------------|-------------|
| | Angle (°) | Radius (mm) | Height (mm) |
| FL901 | 0 | 525 | 1409 |
| FL902 | 90 | 525 | 1409 |
| FL903 | 180 | 525 | 1409 |
| FL904 | 270 | 525 | 1409 |
| FM905 | 45 | 525 | 2916 |
| FM906 | 135 | 525 | 2916 |
| FM907 | 225 | 525 | 2916 |
| FM908 | 315 | 525 | 2916 |
| FU909 | 0 | 525 | 4424 |
| FU910 | 90 | 525 | 4424 |
| FU911 | 180 | 525 | 4424 |
| FU912 | 270 | 525 | 4424 |

Table 3.4: Hydration filter mat notations and locations.

| Filter ID | Description |
|-----------|--|
| FR901 | Radial filter mat at bentonite/rock wall interface encircling deposition hole from base (0.00 m) to 0.35 m height (total height 0.35 m). |
| FR902 | Radial filter mat at bentonite/rock wall interface encircling deposition hole from 6.08 m height to 6.83 m height (total height 0.75 m). |
| FB903 | Horizontal circular filter mat in bentonite, 0.4 m radius at 6.03 m height, centred in DA3147G01 in plan. |
| FB904 | Horizontal circular filter mat in bentonite, 0.4 m radius at 7.03 m height, centred in DA3147G01 in plan. |

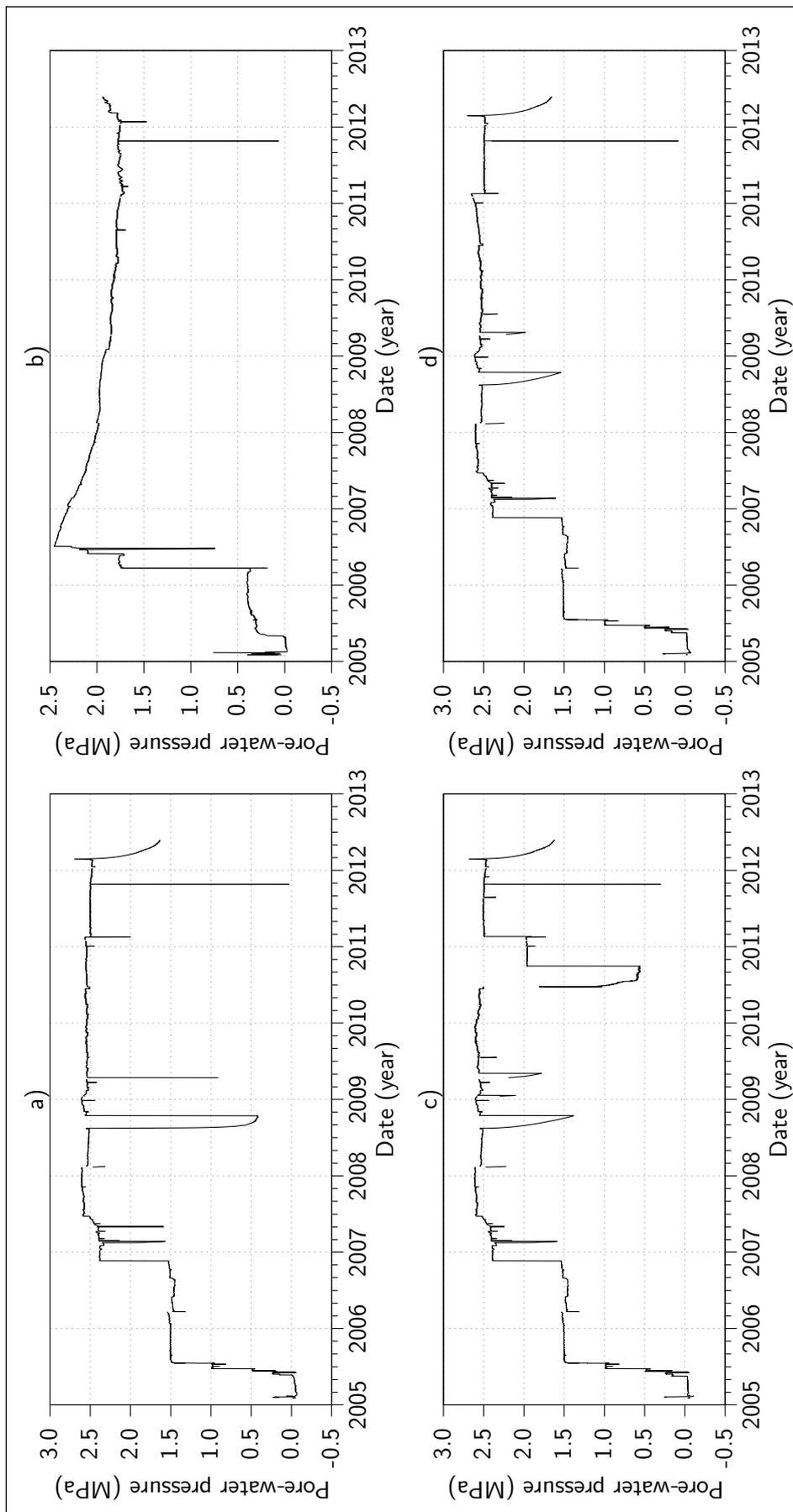


Figure 3.14: Summary of pore-water pressures induced in filter mats.

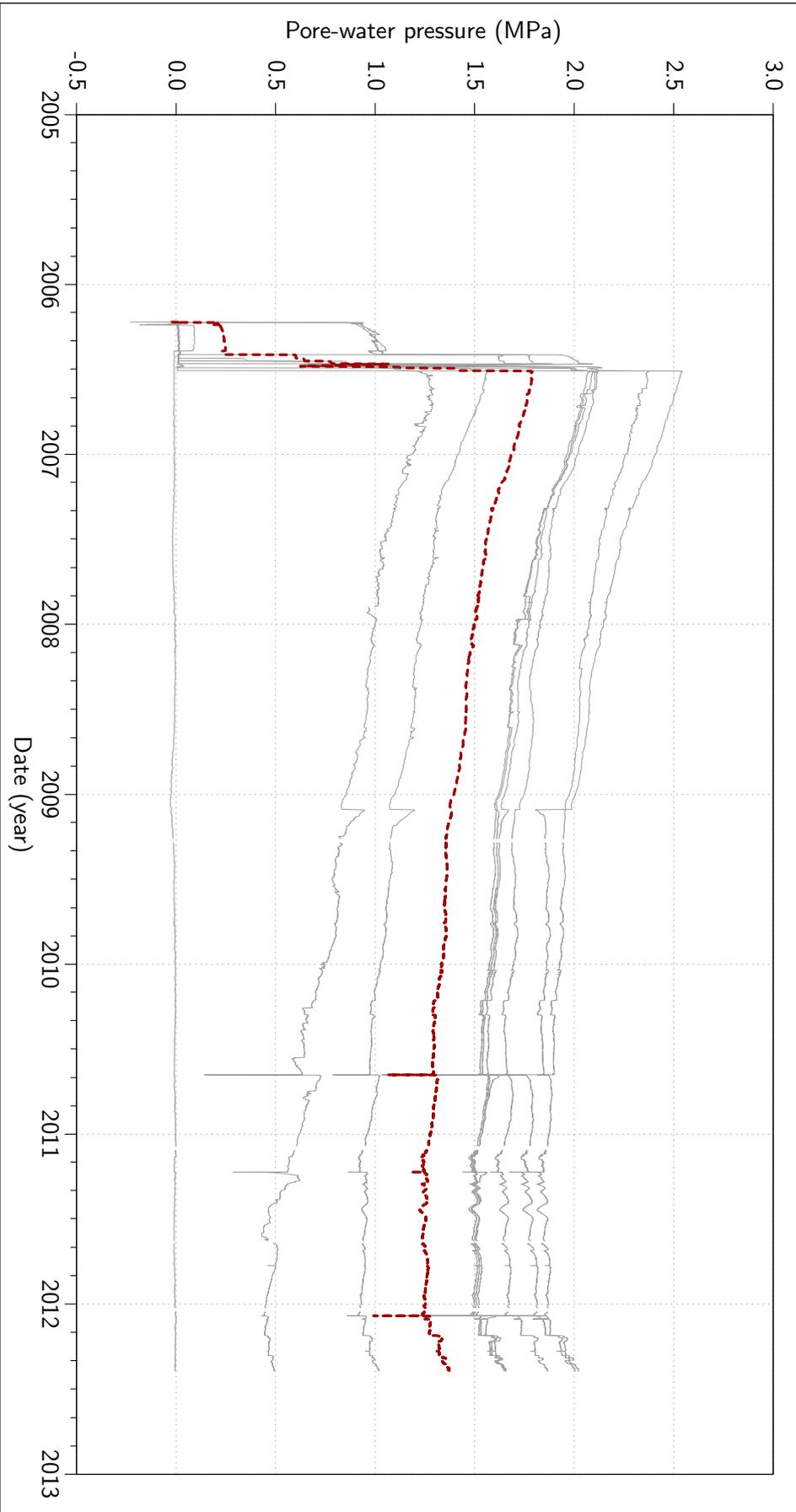


Figure 3.15: Summary of evolution of pore-water pressure in pressure relief holes. Individual axial stress sensors in grey; average value in red.

Table 3.5: Pore-water pressure sensors notations and locations in pressure relief holes.

| Sensor ID | Pressure relief hole |
|-----------|----------------------|
| UR927 | PRH-1 |
| UR926 | PRH-1 |
| UR925 | PRH-1 |
| UR924 | PRH-1 |
| UR923 | PRH-1 |
| UR931 | PRH-2 |
| UR930 | PRH-2 |
| UR929 | PRH-2 |
| UR928 | PRH-2 |

Table 3.6: Linear displacement sensor notations and locations.

| Sensor ID | Description |
|-------------------------|--|
| DP901 | Movement of monel pipe relative to lid in vertical direction. |
| DP902 DP903 DP904 | Movement of lid relative to TBM assembly hall floor in vertical direction (at three locations on lid). |
| DP905 | Movement of lid relative to TBM assembly hall ceiling in vertical direction. |
| DP906 DP907 | Lateral (horizontal) movement of lid. |

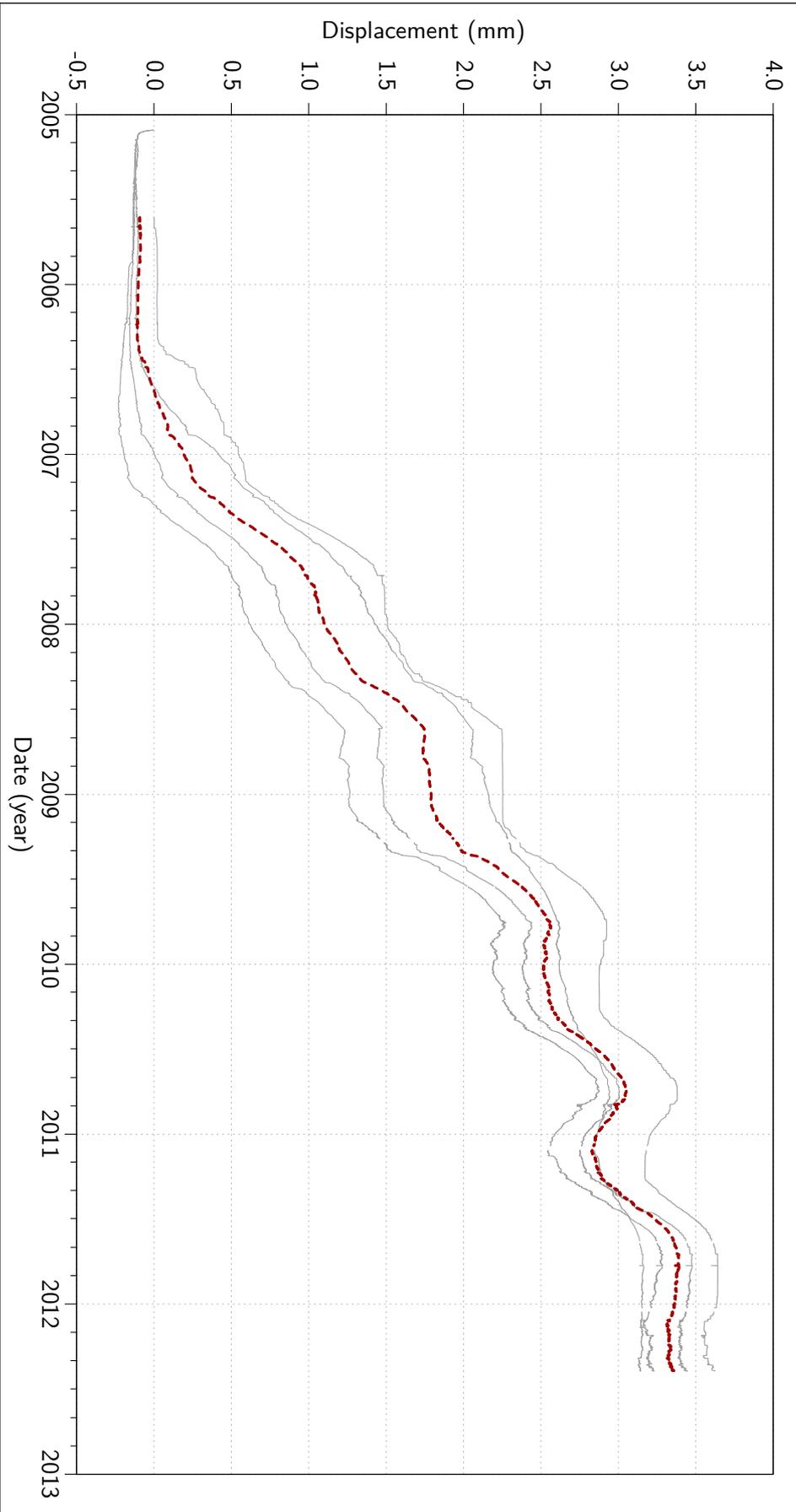


Figure 3.16: Summary of evolution of vertical linear lid displacements. Individual axial stress sensors in grey, average value in red.

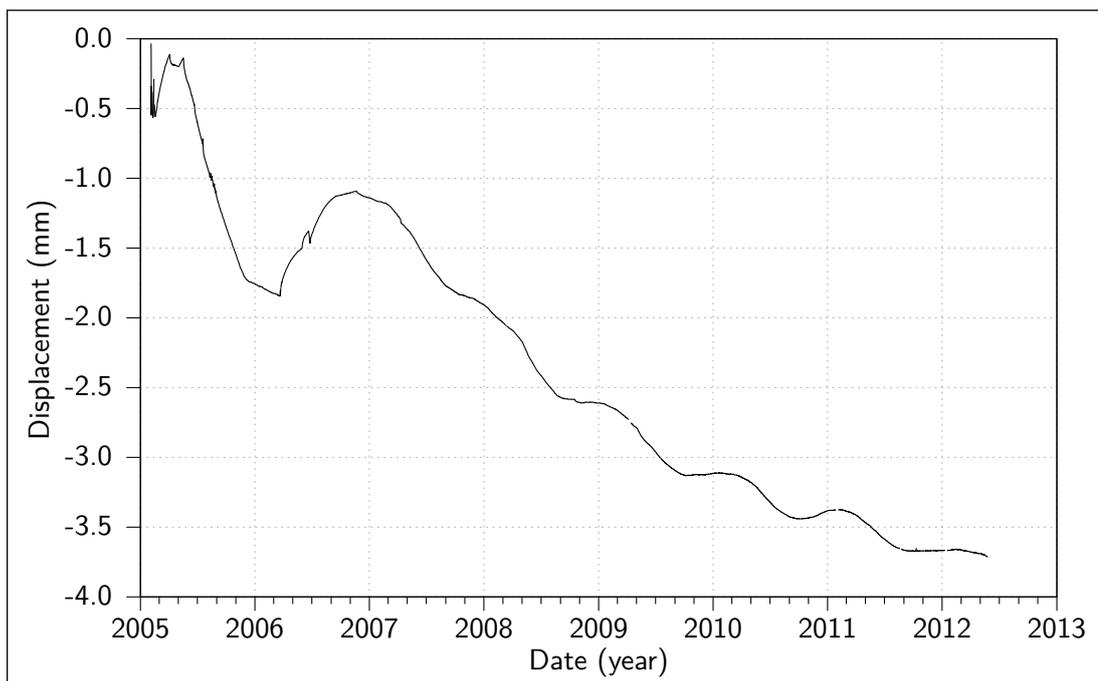


Figure 3.17: Summary of evolution of vertical linear pipe displacement.

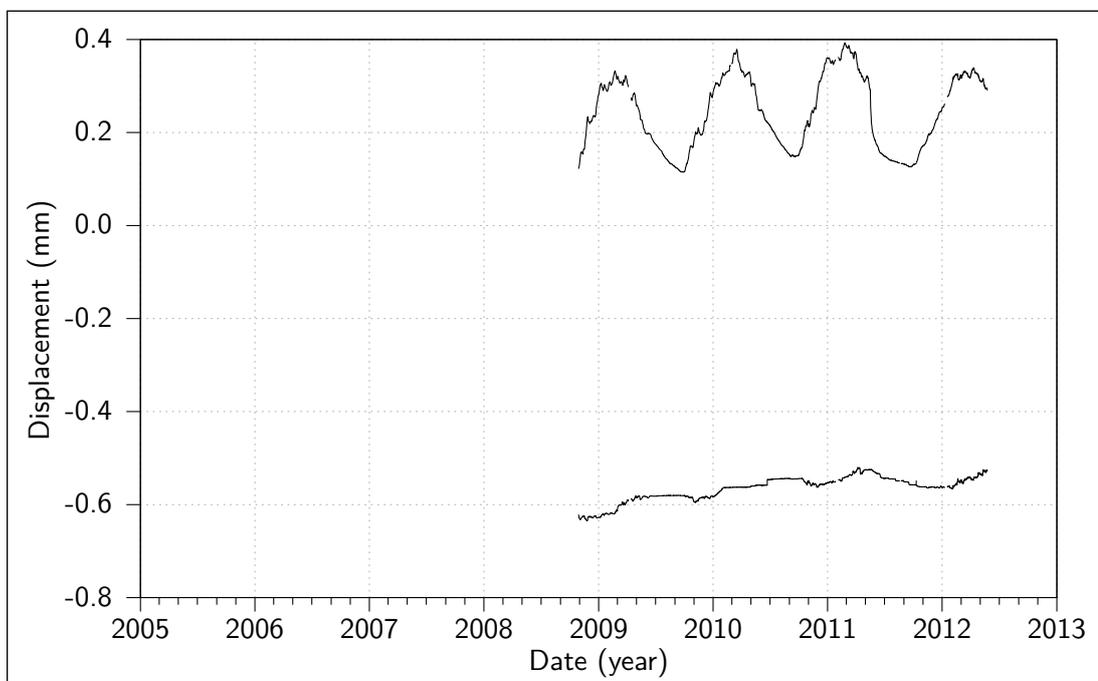


Figure 3.18: Summary of evolution of lateral linear lid displacements.

3.2.4.7 Anchor cable load sensor notations and locations

Table 3.7 details the anchor cable load sensor notations and locations. Three of the ten anchor cables (spaced approximately 120° apart and angled at approximately 21.8° to the vertical axis of DA3147G01) used to retain the concrete plug and lid possess load sensors. The locations of the load sensors is shown in Figure 3.6. Figure 3.19 presents the anchor cable load sensor records. The anchor load cells are manufactured by Glötzl, (model KN 5000 A 160). The load cells possess an operating range of 5,000 kN with an accuracy of ± 10 kN.

3.2.4.8 Ancillary sensor notations and locations

Table 3.8 details the notations and locations of the ancillary (temperature) sensors located within and around DA3147G01 that are not incorporated into a stress or pore-water pressure sensor. Figure 3.20 presents the sensor records arising from the air temperature sensor, TA905, at the experimental location. The thermocouple located inside the canister is a type 'K' manufactured by Pentronic.

3.2.4.9 Integrated temperature sensors

An aggregated summary of the temperature sensors within DA3147G01, i.e. the sensors integrated into the stress and pore-water pressure sensors (Sections 3.2.4.1 and 3.2.4.2), and the temperature sensor located within the canister (Section 3.2.4.8) is presented here. Figure 3.21 presents the sensor records arising from the temperature sensors within DA3147G01 *ommitting* PT901, UT924, and UT911. Figure 3.22 presents the sensor records arising from the temperature sensor UT911

3.3 Experimental history

A summary of the major phases (hydration and gas injection campaigns) of Lasgit's experimental history is presented in Table 3.9. Both elapsed days from initiation and date ranges are provided, with the latter format used throughout the remainder of this document.

Lasgit's official start occurred on 2005-02-01 at 09:39:54 Coordinated Universal Time (UTC),

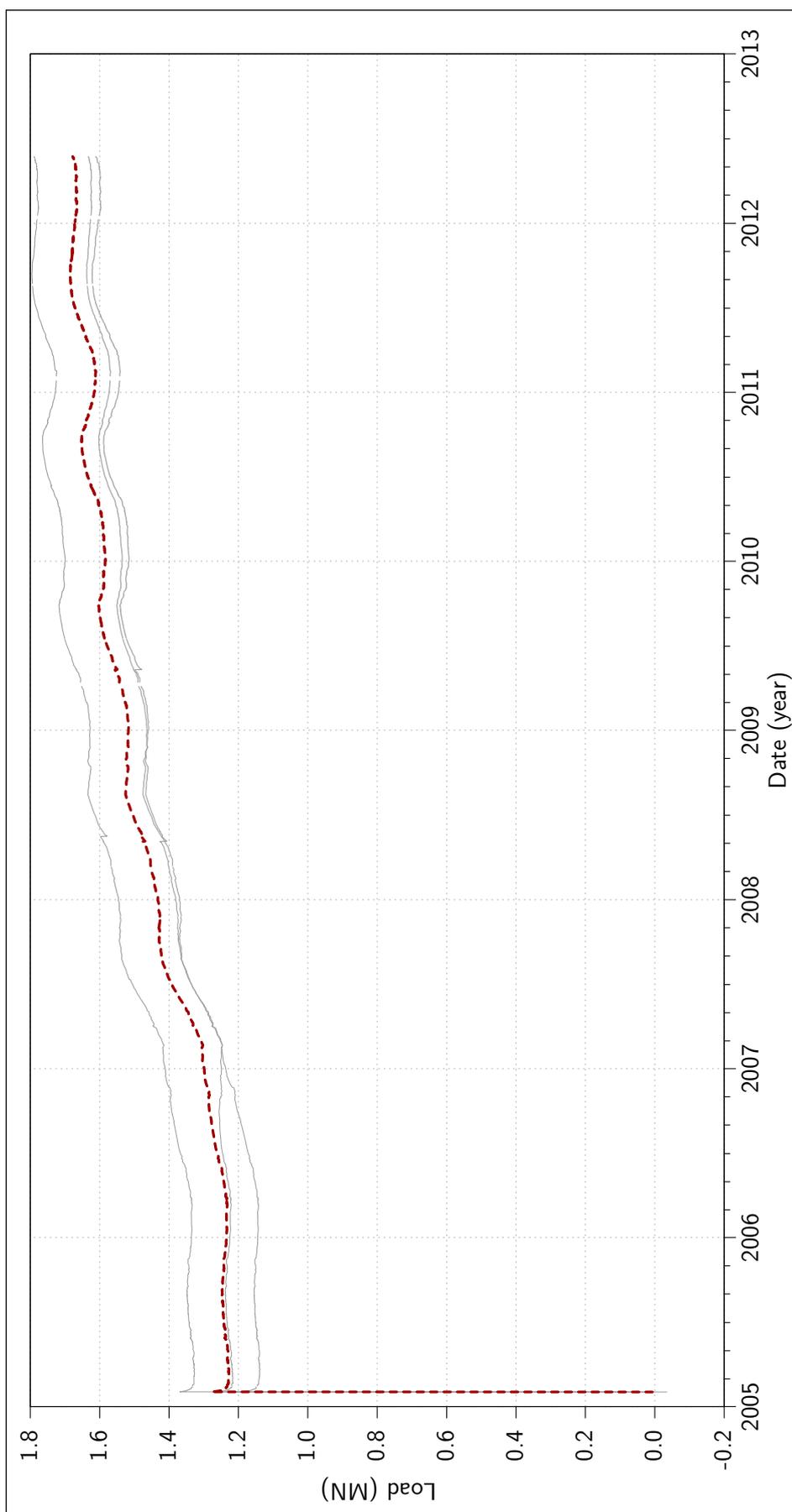


Figure 3.19: Summary of anchor cable load evolution. Individual anchor cable loads in grey, average value in red.

Table 3.7: Anchor cable load sensor notations and locations.

| Sensor ID | Description |
|-----------|---|
| LP901 | Anchor cable load sensors are positioned approximately 120° apart on three of the ten anchor cables. Cables were pre-tensioned to 1,300 kN initially. |
| LP902 | |
| LP903 | |

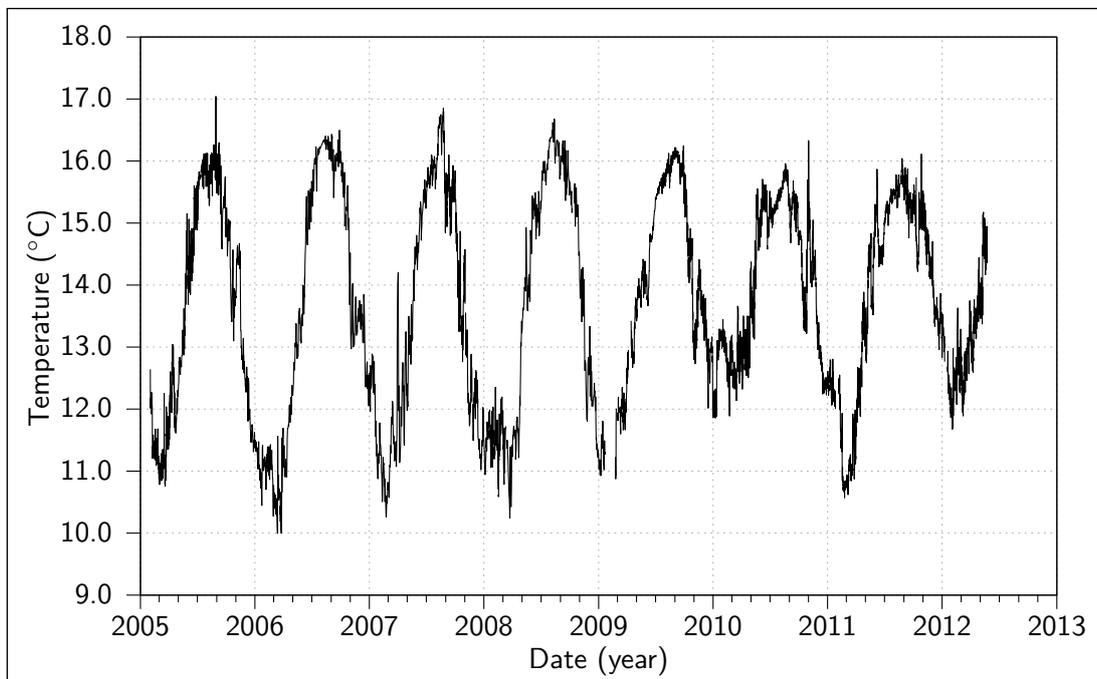


Figure 3.20: Summary of air temperature evolution in Äspö HRL at experimental location

Table 3.8: Ancillary sensor notations and locations.

| Sensor ID | Description |
|-----------|---|
| TA905 | Air temperature in TBM assembly hall (°C) |
| TC901 | Temperature inside the canister (depth indeterminate, °C) |

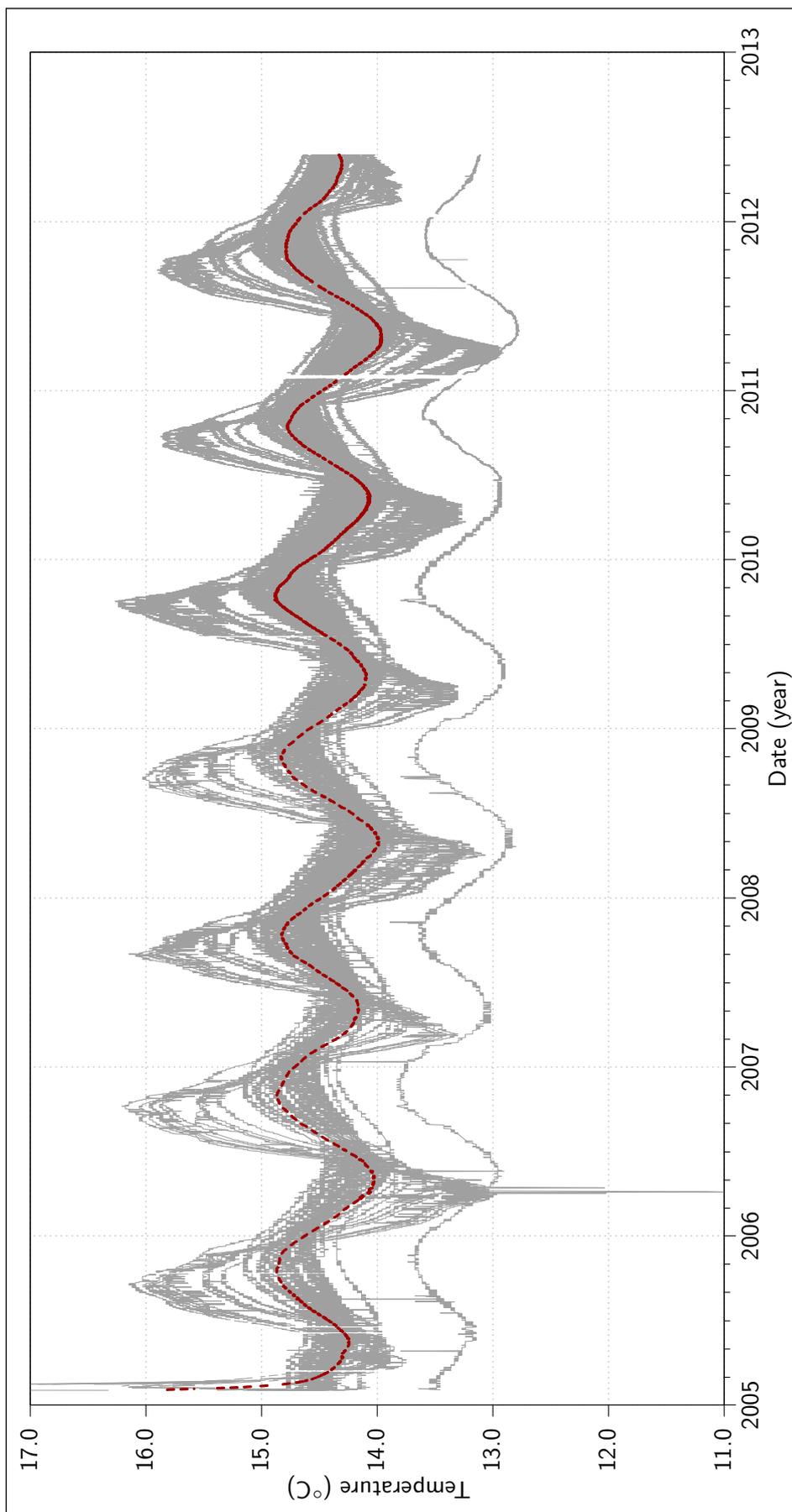


Figure 3.21: Summary of temperature evolution in DA3147G01. Individual PT/UT sensors in grey, average value in red.

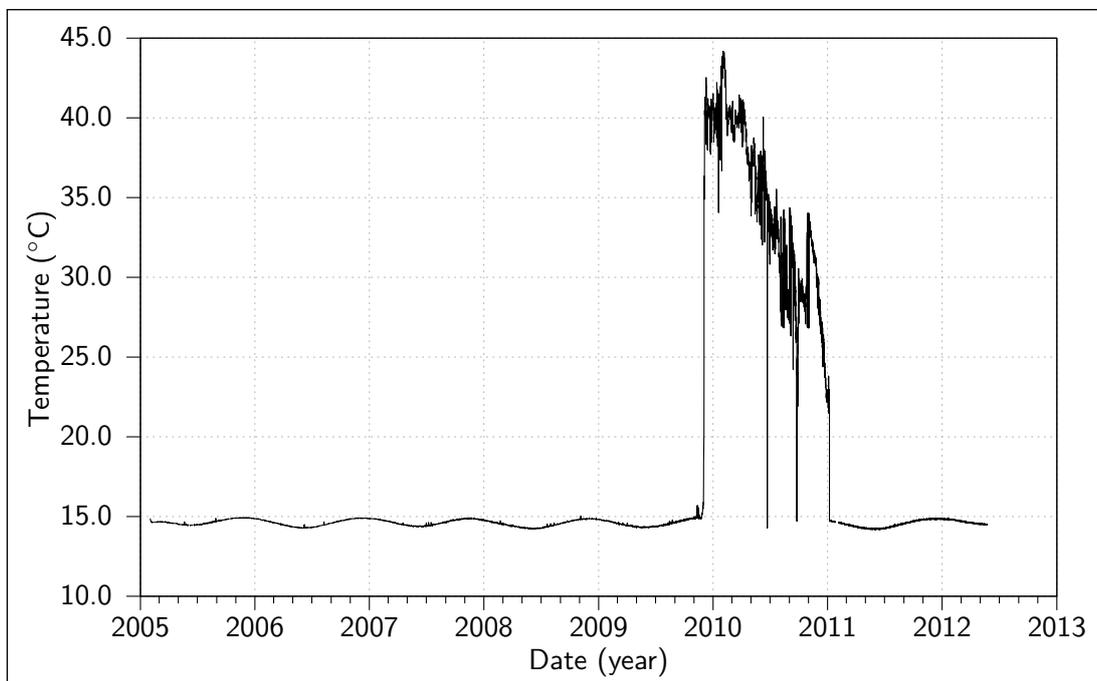


Figure 3.22: Summary of temperature evolution in UT911.

Table 3.9: Key phases of Lasgit's experimental history and expected ongoing phases, summarised from Harrington et al. (2008), and Cuss et al. (2010).

| Event | Elapsed days | Dates |
|---|--------------|--------------------------|
| Lasgit started | 0 | 2005-02-01 |
| System testing and natural hydration phase | 0-106 | 2005-02-01 to 2005-05-18 |
| Pressure relief holes drilled and pumps installed | 62-77 | 2005-04-04 to 2005-04-19 |
| Artificial hydration phase one | 106-843 | 2005-05-18 to 2007-05-25 |
| Closure of pressure relief holes | 415-519 | 2006-03-23 to 2006-07-05 |
| Gas Test One in filter FL903 | 843-1110 | 2007-05-25 to 2008-02-16 |
| Hydraulic test | 843-917 | 2007-05-25 to 2007-08-07 |
| Gas injection test | 917-1010 | 2007-08-07 to 2007-11-08 |
| Hydraulic test | 1010-1110 | 2007-11-08 to 2008-02-16 |
| Artificial hydration phase two | 1110-1430 | 2008-02-16 to 2009-01-01 |
| Gas Test Two in filter in FL903 | 1430-2019 | 2009-01-01 to 2010-08-13 |
| Hydraulic test | 1473-1577 | 2009-02-13 to 2009-05-28 |
| Gas injection test | 1577-1964 | 2009-05-28 to 2010-06-19 |
| Hydraulic test | 1964-2019 | 2010-06-19 to 2010-08-13 |
| Gas Test Three in filter in FU912 | 2019-2072 | 2010-08-13 to 2010-10-05 |
| Test abandoned | 2072 | 2010-10-05 |
| Gas Test Three in filter in FU910 | 2072-2725 | 2010-10-05 to 2012-07-19 |
| Hydraulic test | 2085-2141 | 2010-10-18 to 2010-12-13 |
| Leak off test | 2141-2257 | 2010-12-13 to 2011-04-08 |
| Gas injection test | 2257-2725 | 2011-04-08 to 2012-07-19 |
| Gas Test Four in filter in FL903 | 2725- | 2012-07-19 to present |

when closure of DA3147G01 occurred. Initially, a phase of system testing and natural hydration was begun. During this phase, the natural pore-water pressure was found to be high enough to cause bentonite to be extruded from the deposition hole (Cuss et al. 2010; Harrington et al. 2008). To counteract this phenomenon two Pressure Relief Holes (PRHs), PRH-1 and PRH-2, were drilled beside DA3147G01. Pumps were installed in the pressure relief holes to reduce pore-water pressure local to the deposition hole, which successfully halted the bentonite extrusion.

After testing of the experimental equipment an artificial hydration phase began. During this phase pore-water pressure in the hydration mats and filters was systematically raised, leading occasionally to piping of bentonite, but ultimately resulting in stable increases in pore-water pressure and swelling pressures within DA3147G01. Closure of the pressure relief holes also occurred during this phase, leading to significant increases in pore-water pressures and stresses, however the swelling pressure of the bentonite by this point was sufficient to resist further piping or extrusion.

The first gas injection campaign, 'Gas Test One', was undertaken after just over two years of natural and artificial hydration had taken place. Gas pressure was raised in an injection filter (FL903) in the lower filter array by means of a constant rate of injection. Major gas breakthrough and flow into the system was observed during the test, exhibiting the qualitative form observed in laboratory tests, such as those performed by Harrington and Horseman (1999). However the induced gas flow did not intercept another sensor.

After the gas injection was completed, natural and artificial hydration of the system was then resumed, and continued for just under one year. The maturation of the bentonite between hydraulic tests before and after this phase, indicated by continued increase in swelling pressure, was not found to significantly influence the hydraulic properties of the material (Cuss et al. 2010, 2011).

The second gas injection campaign, 'Gas Test Two', was then undertaken. Gas pressure was raised in the same injection filter used in Gas Test One (FL903) utilising steps of constant flow rate injection. Major gas breakthrough was observed in a similar qualitative fashion to Gas Test One. Additionally, large increases in pressure were recorded shortly after in injection filter FL901 and pore-water pressure sensor UB902. The lack of detection at sensors located at

positions between the source and receptors suggests a tortuous (non-trivial) and discrete flow path.

The third gas injection campaign, 'Gas Test Three', was undertaken immediately after Gas Test Two, with no designated period of hydration occurring between them. The first attempt to perform Gas Test Three (utilising injection filter FU912) was abandoned shortly after its initiation due to an apparent direct hydraulic connection between the injection filter and injection filter FU909. This was illustrated by an immediate increase in detected pressure in FU909 matching that of FU912. Injection filter FU910 was substituted. Gas injection in filter FU910 was performed in a similar manner to Gas Test Two. Despite the unavailability of data for the final portion of Gas Test Three¹, gas breakthrough and flow was observed. Large increases in pressure are seen in sensors FU909/FU912, FU911, and then FL904 after gas breakthrough and flow.

The forth gas injection campaign, 'Gas Test Four', is scheduled to end approximately February 2014 and is intended to be the final test performed.

3.4 Experimental reliability

Among the 150 sensors installed in Lasgit, two temperature sensors (PT901 and UT924) have failed entirely. Another temperature sensor (UT911) has reported temperatures considerably higher (approximately 10°C to 25°C) than the consensus of the other temperature sensors within DA3147G01, for a period of approximately one year. Additionally, a pore-water pressure sensor located in PRH-2 (UR931) has consistently reported pressures just below zero.

The impact of the noted sensor failures with respect to the scientific aims and overall scientific interpretation of the experiment can be considered minimal. This is due to the small number of failures, their type, and the sensor redundancy within the system. Temperature sensors do not monitor any of the discrete flow processes of interest within the experiment, and hence the unavailability of temperature information at three locations of the 56 installed does not prevent an overall temperature profile being developed. Additionally, the anomalous pore-water pressure sensor located in PRH-2 is redundant within the borehole as well as being outside of

¹Experimental data available for analysis spans from 2005-02-01 to 2012-05-24 (days 0 to 2669), with release of the data by the BGS occurring in a staged manner up to the latter date, after an initial data delivery early in 2010.

the compacted bentonite EBS that is the focus of the experiment.

3.5 Data/Results

Macro-scale analysis of the experimental outputs have been (and continue to be) analysed by the BGS. Informal observations during such analyses have noted a potential wealth of smaller scale, or 'second order' information in the dataset, however a detailed analysis of the smaller scale features is yet to be performed.

The dataset arising from Lasgit thus far (from 2005-02-01 to 2012-05-24) consists of circa 27.5 million datum points spread across approximately 175 time series records (150 of which are sensor records, with the remainder pertaining to logs of the control system parameters), and sampled non-uniformly at an average of around 2.5 logs/hour. This equates to 157,362 logging cycles recorded for each sensor by 2012-05-24.

3.5.1 Data uniformity

Non-uniformity of sample rate within Lasgit's dataset occurs for two primary reasons: intentional changes to the sensor logging rate by the experiment operators (BGS); and, on a smaller scale, fluctuations or malfunctions in the data acquisition systems used. The latter of these causes also has the effect of offsetting one uniform section from another by an arbitrary amount. This has the effect of disrupting the synchronisation between two differently sampled, but otherwise uniform sections, and hence makes non-uniform any lower sample rate common to the two sections that would otherwise exist. Figure 3.23 depicts the sample rate (in logs/hour) of Lasgit throughout its experimental life. The short term perturbations in the sample rate (spikes and dips) are typically as a result of hardware and/or software fluctuations and other stability issues, while the longer term changes in the sample rate baseline are a result of changes to the experimental settings. Shaded areas correspond to gas injection campaigns which, as the main focus of the experiment, were considered to warrant higher resolution study (temporally). The extent of non-uniformity in the experimental sample rate highlighted in Figure 3.23 impacts the ability to perform traditional time series analysis, as typically a uniformly sampled time series is assumed (Chatfield 1989).

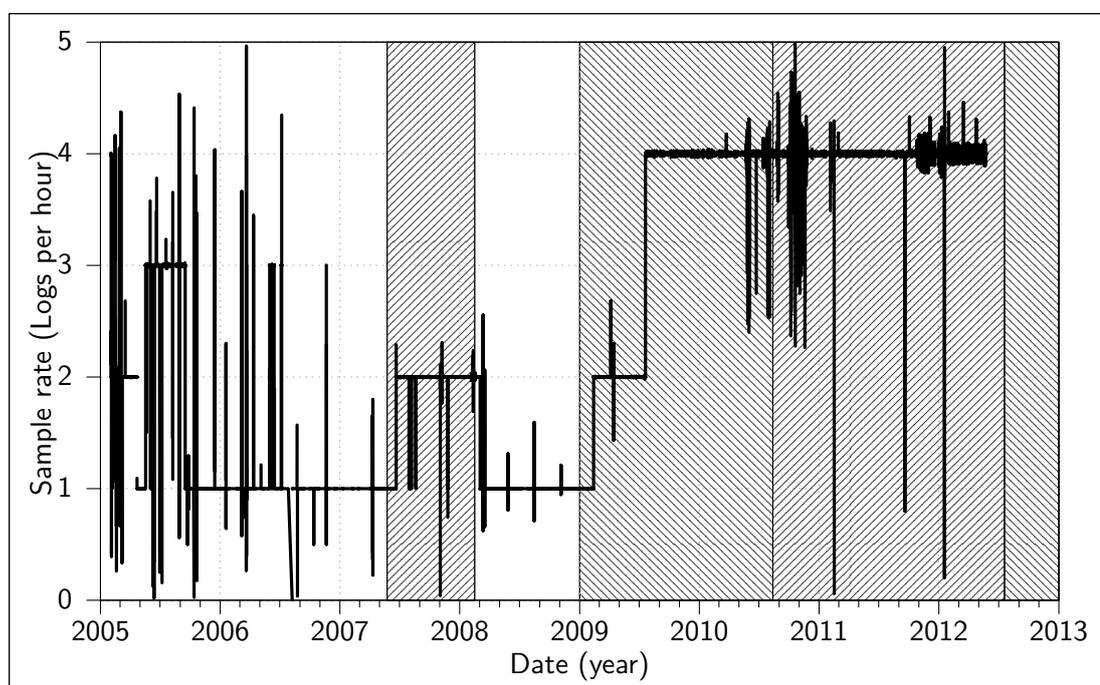


Figure 3.23: Lasgit sample rate with time. Hatched areas correspond to gas injection campaigns.

3.6 Summary

Lasgit is a highly complex experiment, both in terms of experimental set-up, and operational history. While macro scale analyses of the experimental results have been performed (e.g. Cuss et al. 2011), little attention has been paid to the smaller scale phenomena present in the dataset. The details of the experimental set-up and history presented within this chapter provides the necessary context to perform and interpret such an analysis.

The high level of instrumentation and longevity of operation leads to a large dataset which contains a large degree of non-uniformity. The approach adopted to accommodate such a dataset in a second order analysis is presented in Chapter Four.

Chapter Four

Data analysis toolkit development

4.1 Introduction

As highlighted in Chapter Three, the incompatibility between the Large Scale Gas Injection Test's (Lasgit's) non-uniform dataset and traditional techniques for time series analysis necessitates the investigation of appropriate solutions available to perform an Exploratory Data Analysis (EDA) for this research. Discussion of the general approaches available to non-uniform time series analysis are presented in this chapter along with the advantages and disadvantages of each approach. Specific consideration is given to performing a small scale, or 'second order' analysis. Subsequently, the selection and development of a series of analytical processes is described. The incorporation of the analytical processes into a data analysis toolkit, the Non-Uniform Data Analysis Toolkit (NUDAT), is presented.

4.2 Non-uniform time series analysis approaches

The general term 'time series analysis' covers a broad range of activities that may range from simple inferential inspection of a line graph to an advanced statistical analysis of a dataset used to assess a hypothesis about the dataset. A fundamental in time series analysis is the measurement of the extent to which, or the exploitation of, the fact that proximate datum points are interconnected or non-independent of each other (Box and Jenkins 1976). In the context of this study, time series analysis will constitute an activity or procedure that exposes or quantifies interpretable information contained within a time series based dataset.

As stated in Chapter Three, Lasgit has a non-uniform sample rate. The experimental sample rate averages approximately 2.5 logs/hour, but varies throughout the experimental history. As is also stated in Chapter Three, this is in part due to intentional changes to the sample rate by

the experiment operators, i.e. the British Geological Survey (BGS), and also due to fluctuations in the hardware used for data acquisition.

Non-uniformity in a dataset presents a significant impediment to applying time series analysis techniques as a uniform sample rate is usually assumed. Indeed, instructive texts on the topic of time series analysis by Box and Jenkins (1976), and Chatfield (1989) both focus entirely on uniform time series for their subject matter.

With respect to the task of analysing non-uniform time series a number of approaches to circumvent this impediment are reviewed here, which include:

1. Ignoring/disregarding the non-uniformity in the dataset and applying suitable techniques that assume uniformity.
2. Down-sampling the dataset (e.g. utilising every other datum point) to obtain a uniform sample rate common to the whole dataset.
3. Choosing a convenient uniform sample rate and interpolating data between the non-uniformly gathered points at appropriate timestamps.
4. Applying analysis algorithms/tools that do not require or rely on uniformly spaced input to produce valid and meaningful results.

The advantages and disadvantages of these approaches are discussed in detail in Sections 4.2.1 to 4.2.4, and the consideration of the appropriate approach to analyse Lasgit's dataset is presented in Section 4.2.5.

4.2.1 Ignoring/disregarding non-uniformity

Ignoring or disregarding non-uniformity in a dataset may be a suitable option when the non-uniformity present comprises only a small portion of an otherwise uniform dataset, or in situations where the analytical processes used are not particularly sensitive to non-uniformity. The suitability criteria however is highly subjective. It is dependent on the nature of the non-uniformity (e.g. missing data, skewing of sample rate), and the nature and accuracy of the inferences being sought from the analytical method applied.

Time series analysis techniques that assume uniformity in the input data effectively enforce this assumption. As such, the results of such analytical tools may therefore be considered

questionable or inaccurate if there is significant non-uniformity in the input data. This is due to such a time series analysis technique effectively interpreting improperly time indexed measurements. The extent of impact of the non-uniformity on the reliability of results obtained from such an analysis should be considered against the (in)convenience of mitigating the erroneous assumptions through the use of one of the other approaches listed.

A visual example of the consequence of disregarding non-uniformity in Lasgit's dataset is shown in Figures 4.1 and 4.2. In this example a measured down-hole temperature profile from the dataset arising from Lasgit is compared with the effective profile interpreted when data are assumed to be uniformly spaced. Figure 4.1 depicts a down-hole temperature measurement from Lasgit's dataset. Figure 4.2 depicts the same measurements spaced evenly in time (uniformly). The changes in Lasgit's sample rate (see Figure 3.23) causes the distortions visible in the latter temperature profile (Figure 4.2).

4.2.2 Down-sampling to uniformity

Down-sampling a dataset, by utilising only selected datum points in order to achieve a lower sample rate that is common to the whole dataset, has the advantage of allowing more standard and established techniques for analysis to be performed. However, this involves a reduction in data resolution, as it is typically achieved by discarding all data intermediate to those data with timestamps that coincide with the new (lower) sample rate. This approach requires a common sample rate that is useful to the intended analysis, and that is present throughout the dataset.

Figure 4.3 illustrates how even a small additional non-uniformity in an otherwise uniform dataset may disrupt the presence of a lower common sample rate throughout the dataset. An example time series is shown that possesses a varying sample rate. The elapsed time between each datum point in various uniform sections of the time series is denoted by Δt . Black datum points depict uniformly spaced data at the lower ($\Delta t = 20$) sample rate and correspond with the vertical grid line spacing. Grey datum points are those points in the time series that do not correspond to the lower common sample rate. As depicted, a small perturbation in sample rate ($\Delta t = 13$) may prevent subsequent points from coinciding with integer multiples of the initial sample rate, hence prohibiting selection of those points when down-sampling by selecting a

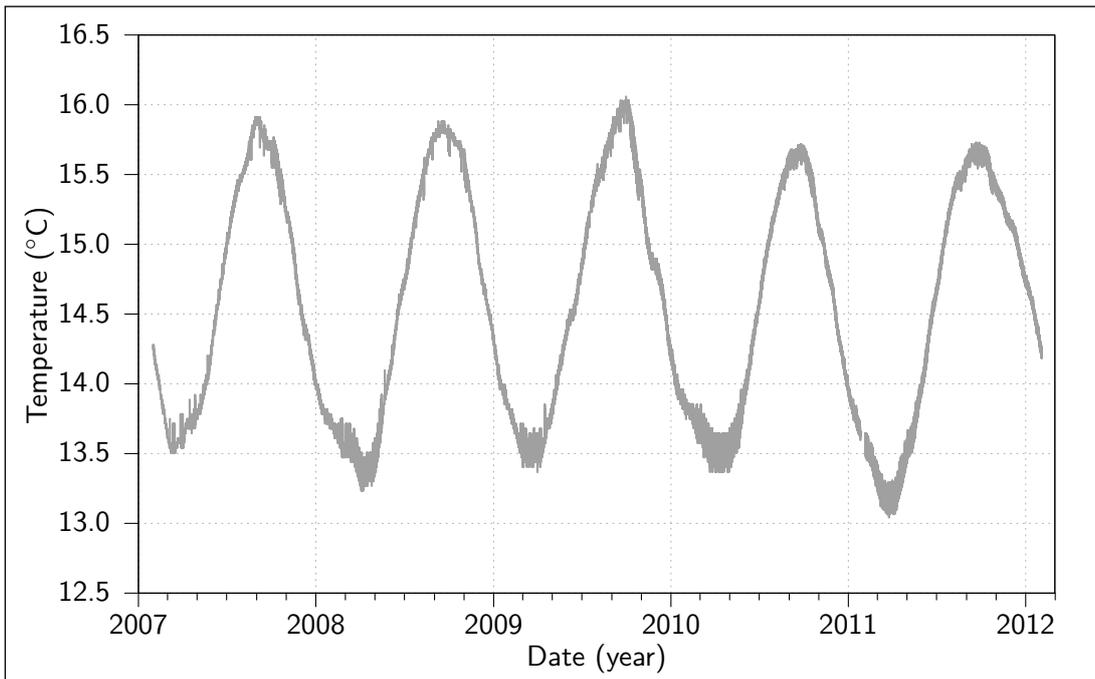


Figure 4.1: Example of consequences of disregarding the non-uniformity in Lasgit's dataset. This figure depicts a measured (non-uniformly sampled) down-hole temperature sensor (real data), while Figure 4.2 depicts the hypothetical consequences of assuming a uniform step between each datum point.

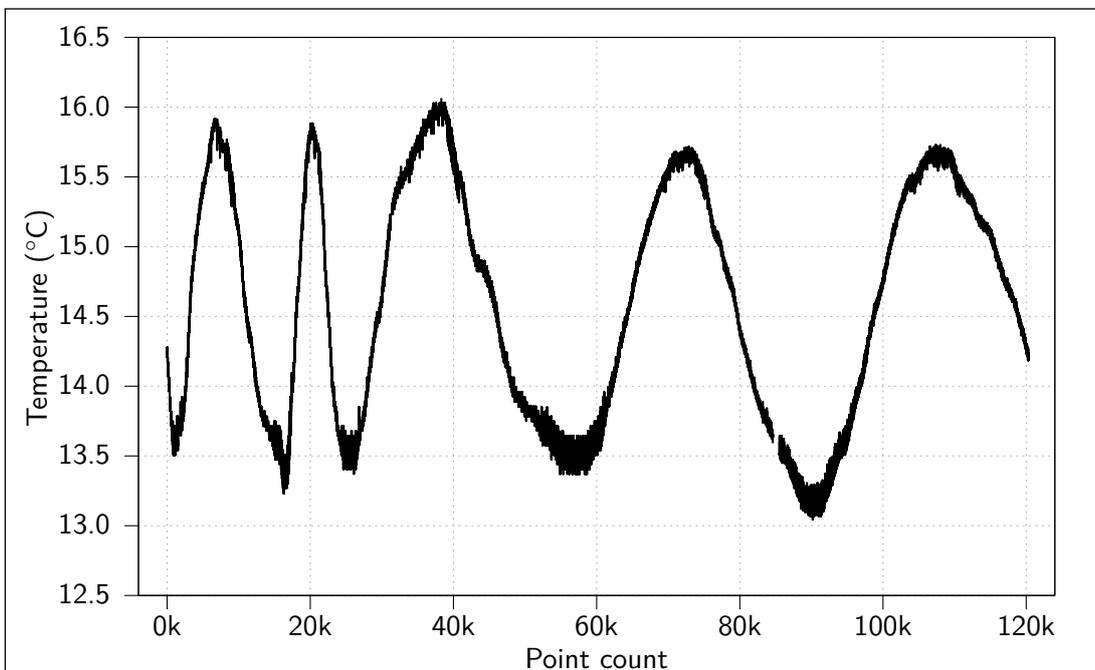


Figure 4.2: Example of consequences of disregarding the non-uniformity in Lasgit's dataset. Figure 4.1 depicts a measured (non-uniformly sampled) down-hole temperature sensor (real data), while this figure depicts the hypothetical consequences of assuming a uniform step between each datum point.

lower common sample rate.

4.2.3 Interpolating to uniformity

Defining a desired sample rate and generating a dataset that conforms to it using interpolation of the original data is likely to generate a dataset that is similar to the original and suitable for uniform time series analysis. However, as a significant proportion of the measured data may be discarded by this process in favour of the new 'generated' data, there is a risk that the resulting dataset would not accurately represent the original measured data. This lack of representation is likely to be most pronounced at local minima and maxima in the original dataset. Except where the position of the interpolated points coincide with the minima/maxima location, the interpolative process is likely to truncate the measured magnitudes. Figure 4.4 illustrates this truncation process at a local maximum point in a hypothetical time series. Consideration of the validity of the use of representative enumeration over original measurement should be made, and balanced against the convenience of analysis of that data.

4.2.4 Use of non-uniform analysis tools

The application of time series analysis tools and algorithms that are capable of accommodating non-uniform input to a non-uniform dataset negates the need for the steps discussed in the preceding sections (i.e. mitigating the need to make potentially poor assumptions, to down-sample, or re-sample of the dataset). However, as stated above, typical time series analysis techniques may lack the ability to accommodate non-uniform input. The selection and development of appropriate techniques that accommodate non-uniform input, or adaptation of existing techniques is therefore required in order for this to be a viable option. Such an approach would allow the use of measured data at full resolution in an analysis, at the cost of increased preparatory work preceding the analysis pertaining to the development of the analysis techniques.

4.2.5 Approach adopted for Lasgit analysis

The appropriate approach to analyse Lasgit's (or any other) dataset is determined by i) the type and magnitudes of non-uniformity present in the dataset, ii) the (subjective) acceptability

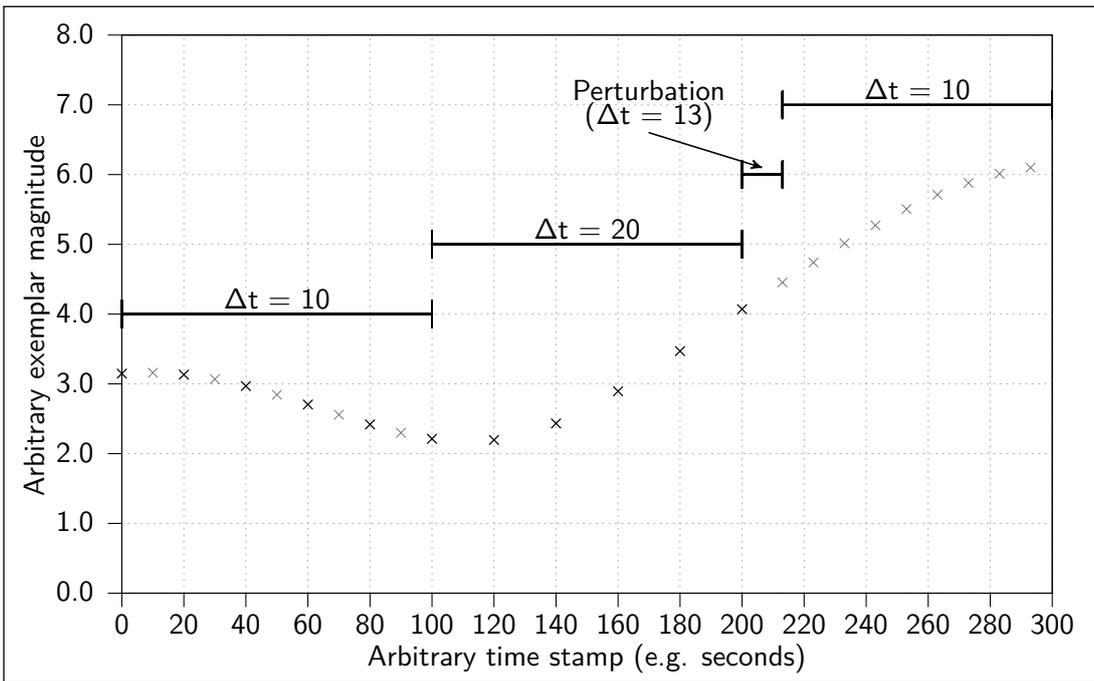


Figure 4.3: Illustrative example of how a small perturbation in sample rate can cause a lack of synchronisation between sections either side of the perturbation, preventing sample rate down-sampling.

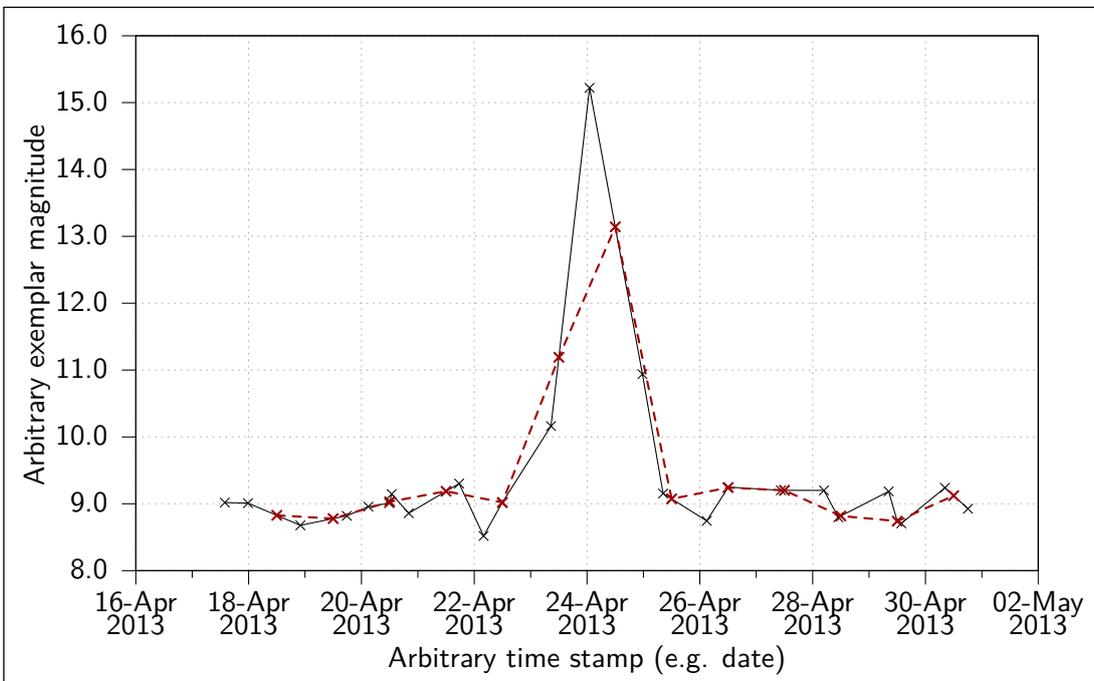


Figure 4.4: Example of potential loss of measured detail in a non-uniform dataset interpolated to a uniform sample rate. The truncation of the magnitude of a measured local maximum is illustrated by the dashed line.

of the implications of the approach taken, and iii) the convenience of performing the analysis in terms of pre-processing requirements or preparatory work.

In the case of Lasgit's dataset large variations in sample rate throughout the experimental history effectively preclude disregarding the non-uniformity, as the non-uniformity is not isolated to small areas but rather makes significant proportions of the dataset incompatible with a single analytical approach. Additionally, the small perturbations in sample rate produce a lack of background synchronisation between uniform areas, therefore preventing precise down-sampling to a fully uniform dataset (as exemplified in Figure 4.3). In this case a significant number of small disregarded non-uniformities would likely remain, as the retained data would likely be assigned on a 'closest fit' basis to the down sampled grid.

Disregarding such small and highly localised non-uniformities after down-sampling may produce sufficient analytical accuracy in the case of Lasgit's dataset. The amount of data discarded, however, in order to achieve a sample rate common to the entire dataset (1 log/hour) would require the elimination of up to 75% of the recorded data in some sections. Furthermore, the interpolative approach to generating a fully uniform dataset would consist of vastly more 'created' data than measured data in the resulting dataset, due to the lack of underlying synchronisation between otherwise uniform sections.

In the case of Lasgit's dataset, and in light of the sensitivity and detail required to perform a second order analysis, the preferred option is therefore to utilise a non-uniform time series analysis. Adopting this approach allows an investigation to preserve the integrity and detail of the actual measured data. Additionally, this approach can avoid potentially invalid analytical assumptions pertaining to the assumed uniformity of the dataset.

4.3 Non-uniform data analysis toolkit

The choice of particular analytical techniques are necessarily guided by the information sought from the dataset and the particular nature of the dataset. The techniques developed for NUDAT are intended to solve the particular information exposure aims of the study of Lasgit's dataset, as well as being intended to be as general as possible in their applicability.

4.3.1 Capabilities and respective rationales

Considering the 'second order' nature of the investigation intended (as introduced in Sections 3.5 and 3.6) and the size of the dataset, the following analysis features were considered and incorporated into NUDAT:

- Second order event candidate detection and event synchronisation indication
- Frequency domain analysis
- Long-term trend determination and subsequent removal
- Aberrant point identification (spike detection)¹
- Smoothing/averaging/down-sampling functions²

It is noted that the specific usefulness of each capability when performing an EDA would be dependant on the dataset analysed. However, the general applicability of the features listed above collectively lend themselves to an EDA type investigation. Each capability is applicable to any time series based dataset, and a specific 'Lasgit analysis centric' rationale exists for each, which supports its inclusion in NUDAT. Sections 4.3.1.1 to 4.3.1.5 outline the rationales that support the inclusion of the capabilities listed above into NUDAT.

4.3.1.1 Second order event candidate detection

The rationale for implementing second order event candidate detection is to provide a quick and consistent identification of potential small scale/second order events for further observational interpretation, where the highlighted events may otherwise be obscured by longer-term/larger-scale phenomena. This feature is of particular use in a large dataset where small scale observation is impractical due to the impracticality of manually searching large datasets. The ability to quickly determine the synchronisations of these events across multiple time series between such events is the rationale behind the event coupling indication feature.

4.3.1.2 Frequency domain analysis

Frequency domain analysis is useful to deterministically calculate the amplitudes and phase offsets of cyclic components in the dataset. This includes the ability to distinguish multiple

¹Quality Assurance (QA) processes are likely to be highly specific to a particular form of phenomenon.

²Highly general functionality that is likely to be easy to implement. Included for completeness.

cyclic components with distinct frequencies. The quantitative measurement of frequency information in a time series is preferable to an estimation derived from observation such as measuring cyclic components peak-to-peak.

4.3.1.3 Long-term trend determination and removal

The rationale behind long-term trend detection and removal is to provide the ability to determine a long term macro-scale behaviour of the dataset and subsequently remove it. Preferably, this is achieved non-parametrically, i.e. the trend is derived without specification of an assumed form. This process leaves a residual time series containing smaller scale, and shorter term phenomena (along with frequency components) which facilitates an easier investigation.

4.3.1.4 Aberrant point identification (spike detection)

Within any experimentally derived dataset aberrant measurements (spikes) are likely to exist. The manifestation or classification of spikes may differ between datasets, and have different scientific implications, however, within Lasgit's dataset, spikes manifest as individual aberrant datum points. As such, the spike detection implemented reflects this. The detection of spikes within Lasgit's dataset enables the removal of such points if deemed necessary.

4.3.1.5 Smoothing/averaging/down-sampling functions

General smoothing, averaging, and/or down-sampling can be useful as a pre-process to reduce noise before simple visualisations, or to produce a lower resolution dataset for coarser preliminary analysis and investigation. The rationale for inclusion is one of providing completeness of data analysis capability.

4.3.2 Formulations of the analytical components of NUDAT

Detailed descriptions of the general and specific approaches taken to accommodate non-uniform input, and the specifics pertaining to the implementation of each of the capabilities are presented in Sections 4.3.2.1 to 4.3.2.6. Section 4.3.2.1 presents a windowing technique that underpins a

number of the components implemented in NUDAT, while the remaining sections address the specifics pertaining to the components described above.

4.3.2.1 Time-scale windowing vs point-count windowing

Advancing a window along a time series and applying an analytical process to the data captured by that window at each step is a common activity in time series analysis. The concept of a moving window of this nature underpins a number of analytical process. Examples of which include Weighted Moving Averages (WMAs) for noise reduction and parametrisations for statistical inferences. The size of the window and the process applied to the captured data at each iteration are typically chosen with consideration of the time scales over which the processes represented by the data are evolving and the information that is sought from the data about the processes.

When a window is centred on a datum point, a predefined number of datum points immediately preceding and succeeding the centred point are defined as being within the window (point-count windowing). When the time series sample rate is uniform, the window length (ℓ , in time) is equal to $(n - 1) \times \Delta t$, where n is the number of datum points and Δt is the elapsed time between two datum points. The fixed and predefined number of datum points captured by a window of this nature is also convenient for subsequent implementation of the desired analytical processes that are intended to be applied.

In the instances where Δt is not constant across a time series (i.e. a non-uniform dataset) the window length in time will vary as the window progresses along the time series. Additionally, the temporal centre of the window will no longer *necessarily* coincide with the centred datum point. Evolution of the data with time would not be measured consistently between two differently sampled regions by the point-count method due to the differences in data density with time. To account for this temporal variability a time-scale windowing method was employed. This involves a window of length ℓ (in time) which captures a potentially varying number of datum points that may possess arbitrary positions within the window as it moves across a time series. Implementation of an analytical processes that is subsequently applied to data gathered in this way must account for this potential variability.

Figure 4.5 depicts a comparison of the two windowing methods described when applied to

a time series with a varying sample rate. An arbitrary non-uniformly sampled time series is shown transitioning from a comparatively higher sample rate to a lower sample rate. The grey shading indicates the points forward and behind the central point that would be encompassed by a time-scale window, while the bar above the time series indicates the points that would be selected using a point-count method. It can be seen that the point-count method places a bias in time towards the lower sample rate region when the window is transitioning between the two locally uniform regions. Additionally, the time over which the windowed points were gathered would not be consistent between the two respective regions of different sample rate. Furthermore, if a large temporal gap exists in the data, the point-count windowing method may group vastly disparate data together as the window transitions the gap.

Time-scale windowing, by definition, collates data by temporal proximity. In the case of a uniform time series, time-scale windowing and point-count windowing are equivalent and will produce identical sets of collected data given equal window lengths. Time-scale windowing however, while it potentially places a burden of increased complexity on the implementation of the subsequently applied processes, provides a methodology for valid analysis of a non-uniform dataset that considers a consistent period of data acquisition. This can better reflect the reality of temporally evolving processes that are represented in the data studied. Time-scale windowing method has been implemented in NUDAT in all instances that utilise a moving window.

4.3.2.2 Averaging/smoothing and down-sampling

Utilising the windowing method described above to perform a WMA (also referred to as smoothing) invalidates the classical approach of predefined weighting factors (Chatfield 1989, e.g.). Accommodating non-uniform time indexes, therefore, requires the use of a continuous function to define weighting factors as a function of each points position within the window. Weighted points are then summed and the total is divided by the sum of the weights, as would be the case with predefined weighting factors:

$$p_s(t) = \frac{\sum_{n=1}^{N_w} w_n p_n}{\sum_{n=1}^{N_w} w_n} \quad (4.1)$$

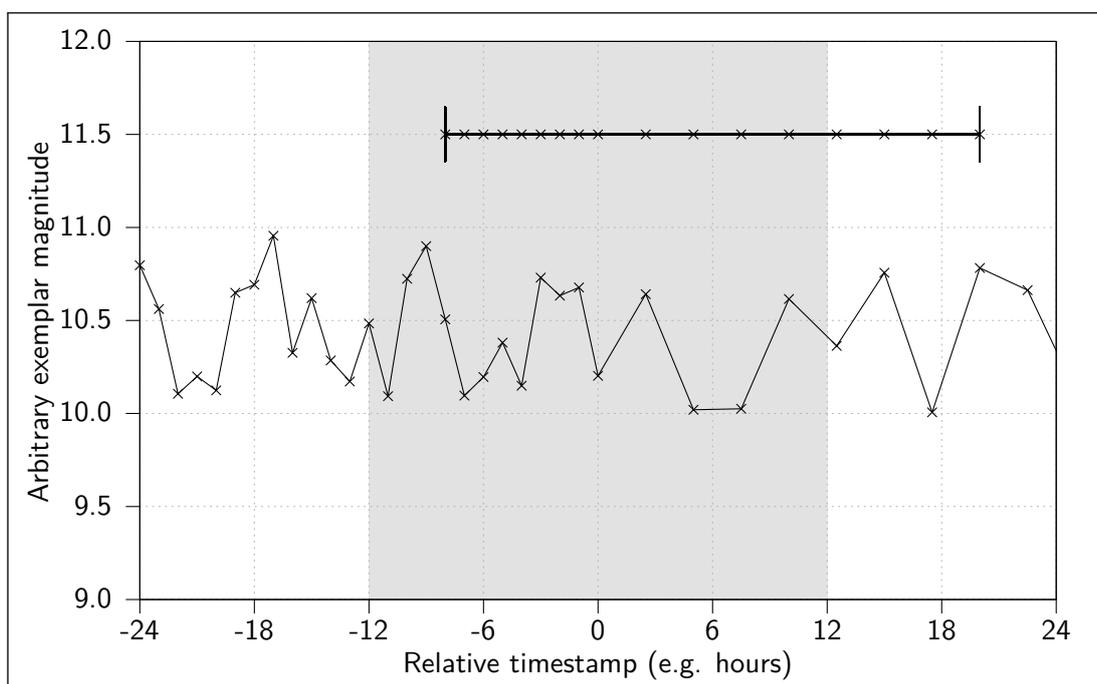


Figure 4.5: Comparison of windowing methods. Time-scale windowing (grey shading) collects points based on temporal proximity while point-count windowing (raised bar) collects points sequentially with no consideration of sample rate uniformity.

where $p_s(t)$ is an averaged or smoothed datum point with the time index t in the centre of the window, N_w is the number of points contained within the window, w_n is the weight applied to the n^{th} point in the window, and p_n is the value of the n^{th} point in the window. The values for t at each window location are assigned by iterating through the input time series and assigning the associated time indices as each window centre (i.e. one window centre for each averaged point in the input/output time series). This constraint can be ignored, and arbitrary time indices can be assigned however, for example when down-sampling the input time series (discussed below).

A series of continuous weighting functions have been implemented to provide the weighting values at the time indices of the n^{th} point (T_n) within a window (of length ℓ) centred at a time index designated as T_0 . Table 4.1 presents the WMA weighting methods and the associated functions which have been implemented in NUDAT, and Figure 4.6 depicts the form of these functions across a window, normalised to unity at T_0 . It should be noted that the implementation of the weighting functions in NUDAT omits this normalisation step for optimisation purposes, as a proportional scaling of the weighting functions does not affect the relative weighting applied to the averaged points.

Choosing $\sigma = \frac{\ell}{6}$ for Gaussian weighting ensures that 99.7% of the area under the curve falls within the function window (effectively equating the window length to three Standard Deviations (SDs) in each direction from T_0). The parameter choice for Laplacian weighting is chosen for convenience such that the weights at the limits of the window are of similarly negligible magnitudes as the Gaussian weighting function at the limits of the window. Additionally, NUDAT allows user specification of the window length ℓ to suite the desired refinement of the smoothing performed.

Down-sampling of a dataset can be achieved by utilising a uniformly weighted moving average and iterating the window across the dataset with a step and window size (ℓ) set to the length required to achieve the desired lower sample rate. Additionally, the extremities of the data captured by each window may be truncated to remove extreme upper and lower values, decreasing the influence of aberrant values on the process. NUDAT implements an optional truncation of values that lie outside the threshold of $1.5 \times$ Interquartile Range (IQR) from the lower and upper quartiles of the windowed data when performing down-sampling. Values excluded by this criteria are referred to as 'outliers' (Tukey 1977). The average values calculated

Table 4.1: Weighting functions for time windowed moving averages. (Standard Gaussian and Laplacian functions Weisstein 2013a,b).

| Weighting type | Weighting function | Parameter choice |
|----------------|--|---------------------------|
| Uniform | $w_{(U)}(T_n) = 1$ | |
| Triangular | $w_{(T)}(T_n) = \frac{-2 T_n - T_0 }{\ell} + 1$ | |
| Gaussian | $w_{(G)}(T_n) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\left(\frac{(T_n - T_0)^2}{2\sigma^2}\right)}$ | $\sigma = \frac{\ell}{6}$ |
| Laplacian | $w_{(L)}(T_n) = \frac{1}{2b} \cdot e^{-\left(\frac{ T_n - T_0 }{b}\right)}$ | $b = \frac{\ell}{10}$ |

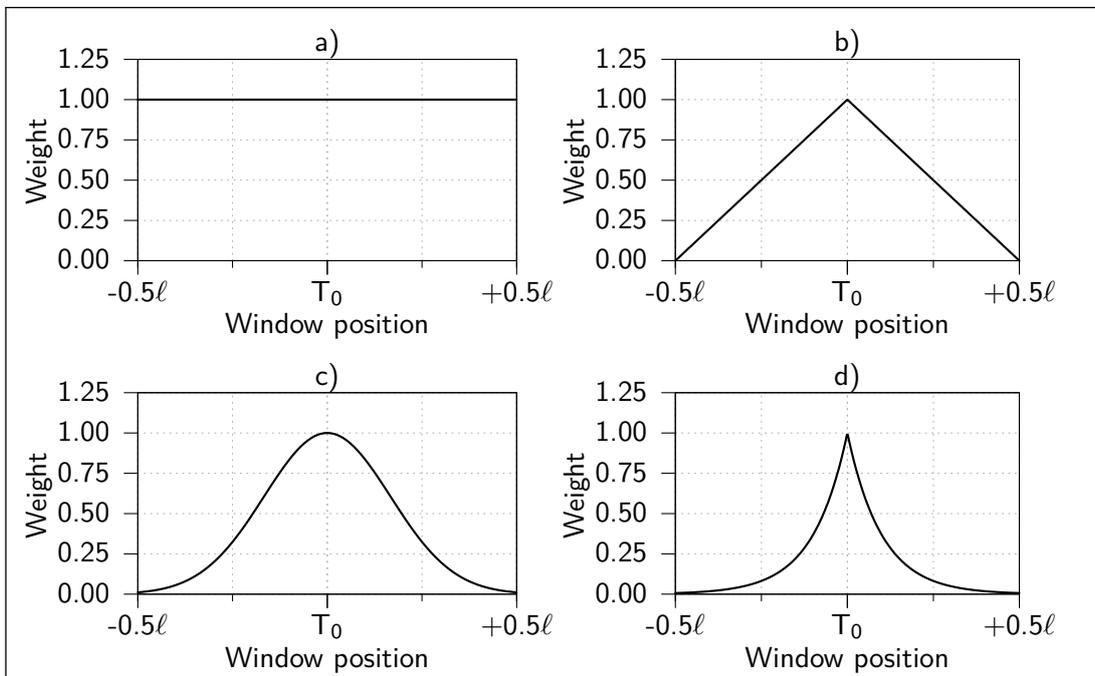


Figure 4.6: Weighting functions used for time windowed moving averages (normalised to unity at T_0). a) Uniform weighting. b) Triangular weighting. c) Gaussian weighting. d) Laplacian weighting. Depicted window length is designated ℓ and is centred on T_0 .

constitute a lower resolution representation of the original dataset, with each datum point in the down-sampled dataset being representative of the data captured by its corresponding window and temporally located at the centre of that window.

4.3.2.3 Time series Quality Assurance (QA)

The QA requirements of an experimentally collected dataset are specific to the quality requirements of the analysis intended to be applied to the dataset. Quality issues, e.g. spiking, sensor drift etc., are likely to have different definitions between different studies. In the case of Lasgit's dataset, the primary quality issue is the presence of aberrant datum points, or spikes, which are characterised here as a single point significantly disparate from the consensus of the surrounding data. The inferred implication of this characterisation is that such datum points are not representative of the physical reality being measured and quantified. The identification of such datum points provides an option to exclude them from an analytical processes subsequently applied to the dataset.

The quantification of i) the consensus of the data surrounding a suspect datum point, and ii) a 'significant' disparation from that consensus, is achieved in NUDAT through statistical parametrisation. Within a time scale window, centred on the suspect point, the deviation of the suspect point from the windowed data mean is compared with the SD of the windowed data. This provides a measure of the discrepancy of the suspect point from the expected value within the window. A threshold to test for disparance in terms of the number (M) of SDs from the mean is established in NUDAT, with points exceeding the threshold designated as spikes:

$$|p_n - \mu| > M\sigma \quad (4.2)$$

where p_n is the suspect datum point, μ is the mean of the windowed data excluding the suspect datum point, M is defined by the operator, and σ is the SD of the windowed data (Weisstein 2013c), excluding the suspect datum point:

$$\sigma = \sqrt{\frac{\sum_{n=1}^{N_w} (p_n - \mu)^2}{N_w - 1}} \quad (4.3)$$

where $n \neq$ the index of the point at the window centre.

The choice of window length for spike detection is based on: i) the sample rate(s) within the dataset; ii) the time scales of which macro scale phenomena and second order scale phenomena occur; and iii) the frequency of occurrence of the aberrant points intended to be detected. As such, the choice of window length should respectively: i) be sufficient enough to ensure a large enough number of data within each window to provide a statistically valid SD calculation; ii) be short enough to ensure that the influence of macro scale behaviour within the window is negligible; and iii) be short enough to ensure that the frequency of occurrence of aberrant points within the window does not significantly alter the quantified consensus of the surrounding points.

The choice of the threshold parameter M is defined by how discriminating the desired spike detection should be, with a lower threshold corresponding to an increase in spike classification. The default value of M , as implemented in NUDAT, is three. This is chosen to coincide with the upper bound of the 68–95–99.7 rule for normally distributed data. In the case where the noise component of a time series is normally distributed this threshold choice will limit false positive spike detection to an average of 0.3%. Adjustment of the threshold parameter can also be used to compensate for a sub-optimal window length selection in some circumstances. More aggressive spike classification, for example, may be used to compensate when frequent spike occurrence influences the measured consensus (SD).

Each datum point in a dataset can be tested in the way described above by iterating the centre of the test window over each time index within the dataset. A pass/fail condition is recorded for each comparison between the suspect point's deviation and the defined threshold of the window. An example of the threshold evolving with time is presented in Figure 4.7. The grey background depicts the calculated threshold as the window is iterated over the hypothetical time series. The dashed sections of the time series depict connections to datum points that fall outside of this threshold (deemed spikes). As the window progresses and encompasses more aberrant points the threshold increases locally. The false negative detection (spike present by not detected) depicted in this figure (the fifth spike, undashed) demonstrates that, in regions where spiking is common, the spiking phenomenon itself may be considered non-aberrant system behaviour by the threshold criteria, but rather an change in inherent variance of the time series. This illustrates the potential impact of poor choice of window length and threshold.

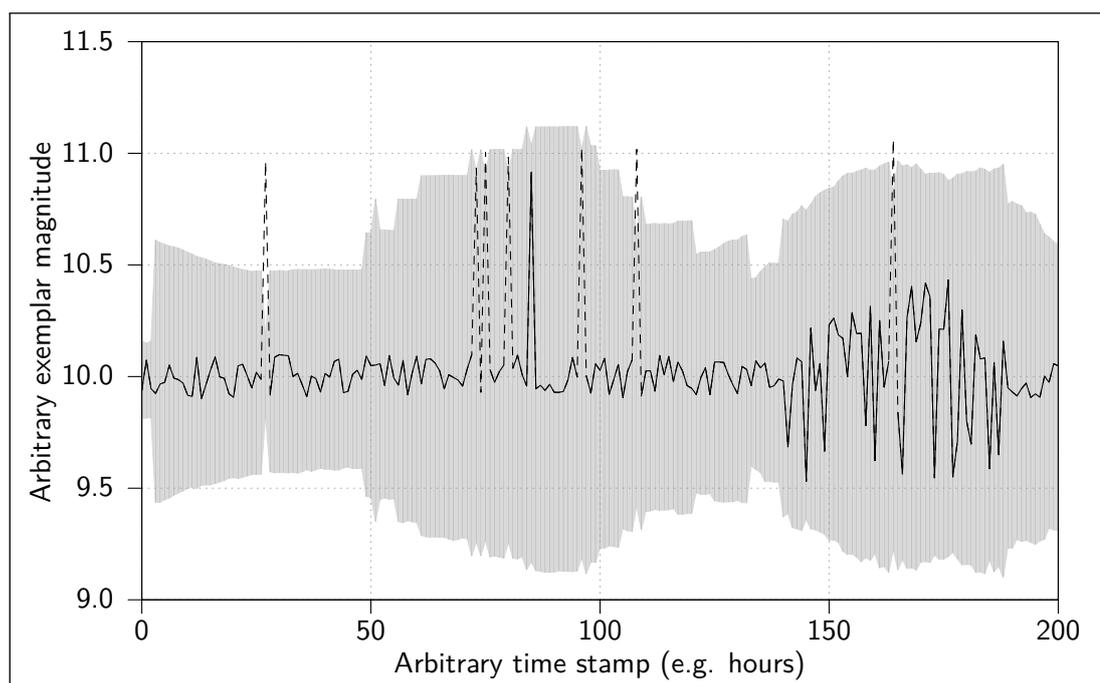


Figure 4.7: Example of spike identification threshold evolving with time and reflecting the variance in the moving windowed data. The threshold (grey shading) dips when centred on a spike due to the exclusion of that point when calculating the threshold. Points outside the calculated threshold (spikes) are depicted by dashed lines, with one undetected spike depicted with a solid line.

4.3.2.4 Event candidate detection

The study of a dataset as large as that arising from Lasgit also requires investigation of the second order features that may be present within it. This may necessitate a process for establishing guidance towards features of potential interest that warrant further investigation, i.e. an event candidate detection capability. A broad definition of an event candidate in this context is simply a perturbation in the data, e.g. a sharp or effectively instantaneous change in magnitude, a change in inherent noise level, a change in gradient etc.

Parametrisation, such as the SD of a moving window (described above) of a time series can be used to identify such perturbations. SD has been implemented as the parametrisation technique, as changes in the SD of a collection of datum points possessing a perturbation has the advantage of being independent of perturbation form. The SD parameter will also be sensitive to second order events if the window scale is shorter than the rate of macro scale evolution. Event candidates detected by this method are identified by relative prominence of the magnitude of the local SD, with larger, shorter lived, and more isolated magnitudes corresponding to larger and more isolated/independent perturbations.

Synchronisation of event candidates across multiple time series within a dataset can also be identified by observation of temporally coincident peaks in SD magnitude. Coincident (or synchronised) event candidates in multiple time series within a dataset can lend weight to the interest of the event candidates, and can be used to quickly identify connections between time series, thus potentially refining cross-referencing of events between, for example, multiple sensor records.

4.3.2.5 Trend detection and removal

Singular Spectrum Analysis (SSA) is a non-parametric technique for decomposing a time series into components that are typically categorised as trending, oscillatory, and noise. While all types of components derived by SSA may be useful to time series analysis, the trending components (or their sum), and their subsequent removal from an input time series, are the focus of the implementation of the technique in NUDAT.

The procedure for SSA, as described in Golyandina et al. (2001), involves converting a time

series into a matrix of lagged vectors, decomposing the lagged vector matrix into components, and performing averaging of the decomposed matrices to convert them back to time series components. The length of the lagged vectors used in the decomposed matrix is defined by user input and, as with window length, pertains to the time scale over which the process operates. Larger window lengths produce more general trending behaviour, while shorter window lengths retain more shorter term variation in the derived trends. Golyandina et al. (2001) states that a uniform sample rate of the input time series is not a necessary requirement. However as the lagged vectors are of fixed number of points (effectively a point-count window) the same issue regarding inconsistent analysis of temporal evolution between regions of uniform sampling rate are present as those discussed above. A methodology has been adopted to deal with non-uniformity in the analysis of Lasgit's dataset, which is discussed in detail in Chapter Six.

The matrix of lagged vectors (\mathbf{X}) from a time series P (where $P = (p_0, \dots, p_{N-1})$) is formed as such (Golyandina et al. 2001):

$$\mathbf{X} = \begin{bmatrix} p_0 & p_1 & p_2 & \cdots & p_{K-1} \\ p_1 & p_2 & p_3 & \cdots & p_K \\ p_2 & p_3 & p_4 & \cdots & p_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{L-1} & p_L & p_{L+1} & \cdots & p_{N-1} \end{bmatrix} \quad (4.4)$$

where L is the lagged vector length, N is the number of points in the time series, and:

$$K = N - L + 1 \quad (4.5)$$

The decomposition of \mathbf{X} consists of separating the matrix into L components that sum to \mathbf{X} (Golyandina et al. 2001):

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_L \quad (4.6)$$

where each component (\mathbf{X}_n) is derived as such (Golyandina et al. 2001):

$$\mathbf{X}_n = \sqrt{\lambda_n} U_n V_n^T \quad (4.7)$$

in which λ and U are the eigenvalues ordered in descending absolute magnitude and their corresponding eigenvectors of the matrix \mathbf{S} , where:

$$\mathbf{S} = \mathbf{X}\mathbf{X}^T \quad (4.8)$$

and V_n^T is defined as the transpose of the vector V_n , which in turn is defined as (Golyandina et al. 2001):

$$V_n = \frac{\mathbf{X}^T U_n}{\sqrt{\lambda_n}} \quad (4.9)$$

The calculation of the eigenvalues and corresponding eigenvectors of \mathbf{S} is achieved by application of the Jacobi eigenvalue algorithm (Golub and van der Vorst 2000; Golub and Loan 1996). The Jacobi eigenvalue algorithm involves iteratively pre and post multiplying \mathbf{S} by a Givens rotation matrix (\mathbf{G}) (Golub and Loan 1996):

$$\mathbf{S}' = \mathbf{G}^T \mathbf{S} \mathbf{G} \quad (4.10)$$

The Givens rotation matrix (Equation (4.12)) is formed at each step by assigning i and j to indices corresponding to an element in the upper triangle of \mathbf{S} and setting θ as such (Golub and Loan 1996):

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{-2S_{ij}}{S_{jj} - S_{ii}} \right) \quad (4.11)$$

$$\mathbf{G}(i, j, \theta) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos(\theta) & \cdots & \sin(\theta) & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -\sin(\theta) & \cdots & \cos(\theta) & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \quad (4.12)$$

With the Givens rotation matrix defined as above, the multiplication described in Equation (4.10) reduces the element S_{ij} to zero. When all elements in the upper triangle have converged sufficiently close to zero, the main diagonal of \mathbf{S} will constitute a close approximation of the eigenvalues sought. The corresponding eigenvectors of \mathbf{S} are determined by iterative

pre-multiplication of an identity matrix by each Givens rotation matrix determined to calculate the eigenvalues.

Once the eigen information has been used to decompose the lagged vector matrix (\mathbf{X}) each component matrix (\mathbf{X}_n) is averaged across the diagonals defined by $i + j = \text{constant}$. The result of this averaging converts the component matrices to component time series. Just as the decomposed matrix components (Equation (4.6)) sum to the original lagged vector matrix, so do the component time series derived from them sum to the original time series.

Typically, but not necessarily, the trending components are associated with the highest magnitude eigenvalues, the noise components with the lowest magnitude eigenvalues, and the oscillatory components with the intermediate (Golyandina et al. 2001). Selection, therefore, of the trending components for subtraction from a time series based dataset, requires visual inspection of the form of the trends, and hence manual decision once the components have been determined by this process.

4.3.2.6 Frequency domain analysis

Quantitative identification of the frequency content of a time series can be determined mathematically by applying a Discrete Fourier Transform (DFT) (e.g. Oppenheim and Schaffer 1975). The application of a DFT to a time series determines the amplitude and phase offset of an oscillatory component of the specified frequency in that time series. The standard DFT procedure utilises the length and sample rate of the input time series to define the inspected frequency points, namely at wavelengths that are integer divisions of the time series length up to the Nyquist limit (a maximum frequency of half the sample rate, Chatfield (1989)). The power, W expressed as a complex number, of a particular frequency component in a time series is calculated as follows (Oppenheim and Schaffer 1975):

$$W(k) = \sum_{n=0}^{N-1} p_n \cdot e^{-i2\pi \frac{k}{N} n} \quad (4.13)$$

where $k = 0, \dots, N - 1$.

In the case of a non-uniform time series, such as Lasgit's dataset, the Nyquist limit is undefined and the component n in the exponential function will not necessarily be synchronised with the

temporal location of p_n in the time series. This may potentially lead to the DFT interpreting the effective distortions depicted in Figures 4.1 and 4.2. As such a modification of the standard DFT have been made to accommodate a non-time series derived definition of frequency locations to inspect, and to apply non-uniform temporal weighting of the data in the time series. The modification of a DFT to a Non-uniform Discrete Fourier Transform (NDFT) is achieved by substituting ω for $2\pi\frac{k}{N}$ where $\omega = 2\pi f$ (in which f is the inspected frequency), and n is substituted with t_n :

$$W(\omega) = \sum_{n=0}^{N-1} p(t_n) \cdot e^{-i\omega t_n} \quad (4.14)$$

The modification substituting ω for $2\pi\frac{k}{N}$ necessitates manual specification of the frequencies investigated, however it allows a non-uniform range of frequencies to be inspected. This is the definition of an NDFT given in Bagchi and Mitra (1999). Additionally, substituting n with t_n allows a non-uniform input time series to be accommodated. The term NDFT refers to the combined modification when used in this study.

The implementation of the NDFT capability in NUDAT provides the option to automatically remove the zero frequency component (sometimes referred to as the 'DC component') from the input time series. This zero-normalisation of the input reduces leakage into the low frequency components calculated.

4.3.3 Toolkit components summary

The desirable capabilities of the toolkit listed in Section 4.3.1 have been developed into functioning toolkit components, and have been described in Sections 4.3.2.2 to 4.3.2.6. Figure 4.8 summarises these components and associates them with their respective outputs. The individual toolkit components can be used either in isolation or in sequence (dependant on analysis requirements) with the exception of frequency domain analysis, which does not output a time series suitable for use as input into another capability.

4.4 Analysis algorithms and application procedures

The toolkit processes are intended to be applied sequentially and, if necessary, recursively. An example chain of application may be comprised of the following steps:

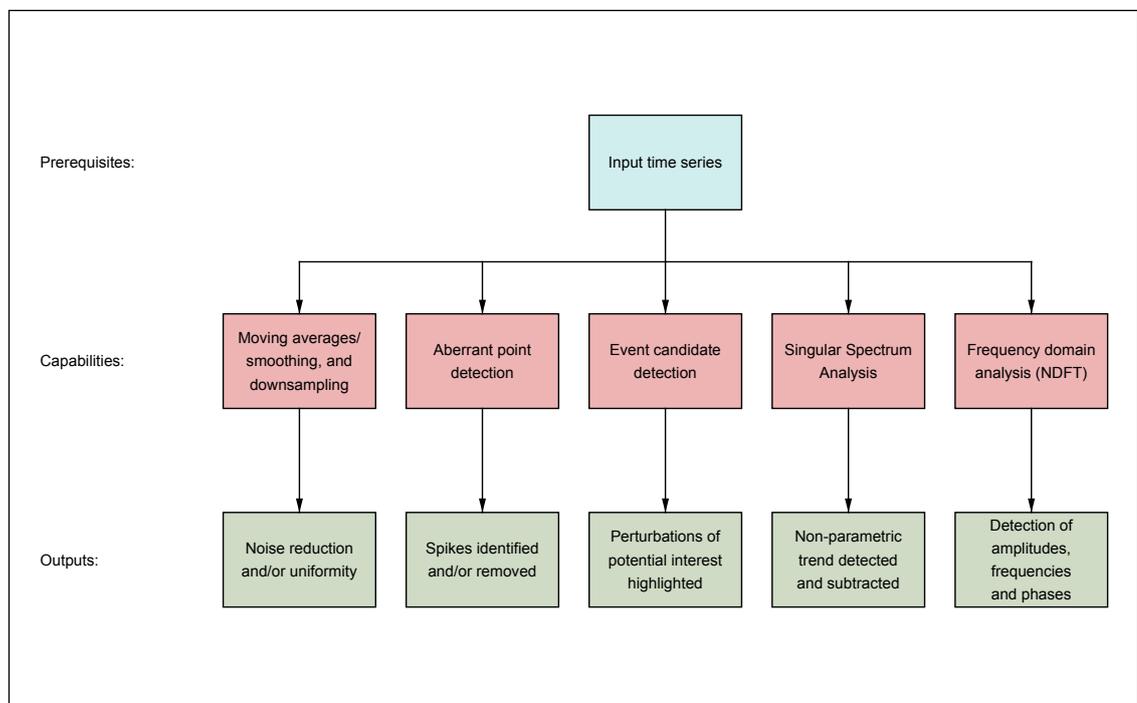


Figure 4.8: Summary of toolkit capabilities and resulting outputs.

- Apply spike detection process to raw time series data
- Down-sample the de-spiked time series using averaging functions to representative values occurring daily
- Apply SSA to down-sampled time series, to determine the long term trends
- Subtract long term trends from full resolution de-spiked time series
- Calculate NDFT for a range of potentially relevant frequency contents for de-trended time series
- Apply event candidate detection process to de-trended time series
- Investigate indicated event candidates by classical inspection means

This example sequence of analytical processes (and other sequences) can be performed over multiple time series within a dataset, with the frequency and event information aggregated between the time series for comparison or cross referencing. Additionally, smoothing of a process output can be added as a stage where considered necessary (e.g. pre-processing a time series possessing a high level of noise to improve visualisations).

4.5 Implementation and application of NUDAT software

This section summarises the technical aspects of the toolkit implementation. The supported computational platforms are briefly summarised, as well as the data file format prerequisites, and the known limitations and omissions of the implementation with respect to data time stamps. A brief description of the Human Computer Interface (HCI) is also given.

4.5.1 Coding specifics

The NUDAT software has been developed using the C# (C-Sharp) programming language and requires the Microsoft .NET Framework for a Windows desktop environment. The majority of the code is implemented as a single threaded linear process. Exceptions to this include the native C# parallel processing functionality utilised in the algorithm that implements the calculation of the NDFTs, and the pre and post multiplication of the **S** matrix by the Givens

rotation matrix in the Jacobi eigenvalue algorithm as part of the SSA functionality. The source code can be compiled into either a 32 or 64 bit executable. Compilation and execution has been tested successfully using both Windows XP and Windows 7 operating systems, in 32 and 64 bit versions respectively.

4.5.2 Supported dataset format

NUDAT supports Comma Separated Values (CSV) text data files for both input and output. Output from NUDAT is routinely generated in CSV format for ease of application of subsequent processes, allowing a chain of analysis to be developed and the results of each step utilised or built upon. This output format is also compatible with a wide range of other software.

The time stamp format required for a time series must match a format that can be successfully parsed by the C# 'DateTime.Parse()' function. The ISO 8601 extended format is recommended for time stamp format in NUDAT input files, and as such, all time stamp outputs from NUDAT are in the ISO 8601 extended format: 'YYYY-MM-DD hh:mm:ss'. Most processes require that time stamps (and therefore corresponding time series measurements) are arranged in ascending date-time order. Time zones have not been given any consideration within NUDAT's implementation, and as such their use within input time stamps is discouraged.

4.5.3 Human Computer Interface (HCI)

NUDAT's HCI consists of a collection of standard interactive Windows forms. Input files are listed using the main window for analysis by either 'dragging and dropping' or selection via the File → Open menu. The user then selects the analytical process to apply to the input data via the Tool menu, and subsequently configures any necessary settings (e.g. window length, thresholds, or frequency ranges etc.). The input file list is then iterated over and the selected process applied in turn to the data in each file. The output of each process application is stored in a subdirectory of the input location, named by the process applied. Sequential application is achieved by loading the output files from one process as input.

Figures 4.9 to 4.11 present NUDAT's HCI. Depicted are the main window (Figure 4.9), in which input files of CSV format are loaded and the highlighted file previewed; the tool selection menu (Figure 4.10), in which the process to be applied to the input is chosen; and a process

configuration window (Figure 4.11), in this case the SSA window, in which the parameters of the analytical process are specified and the process executed.

4.6 Summary

The general approaches to the investigation (e.g. an EDA) of a time series based dataset that is sampled non-uniformly have been considered (Section 4.2). Consideration to the size of Lasgit's dataset, coupled with the second order detail required to be investigated, a non-uniform accommodating approach was considered most appropriate. As such the development of a time series analysis toolkit capable of performing such an analysis has been undertaken and presented.

The capabilities of the toolkit, summarised in Figure 4.8, have been chosen for their general applicability to a generic time series, their ability to be modified to accommodate non-uniformity where applicable, and their information exposing qualities. The software implementation of the toolkit, NUDAT, facilitates the application of each analytical process individually and in sequence, allowing an analytical chain to be built that is capable of quantifying information about the dataset and quickly exposing and identifying potential second order events of interest for further classical observational analysis. The characterisation and verification process for the developed components is presented in Chapter Five.

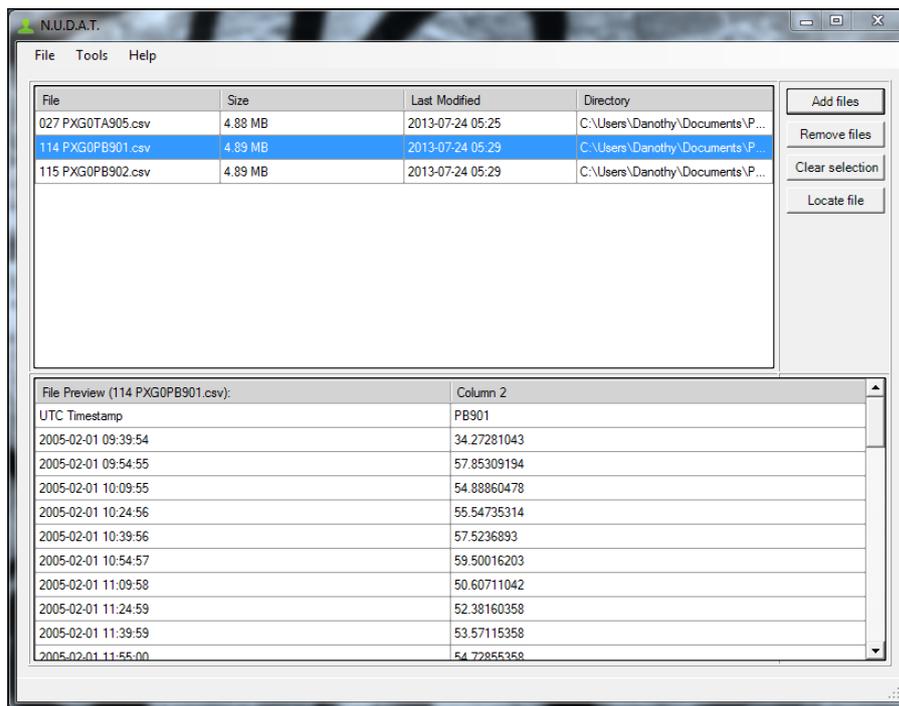


Figure 4.9: NUDAT's input file interface. Input files are listed along with a preview of the highlighted file contents.

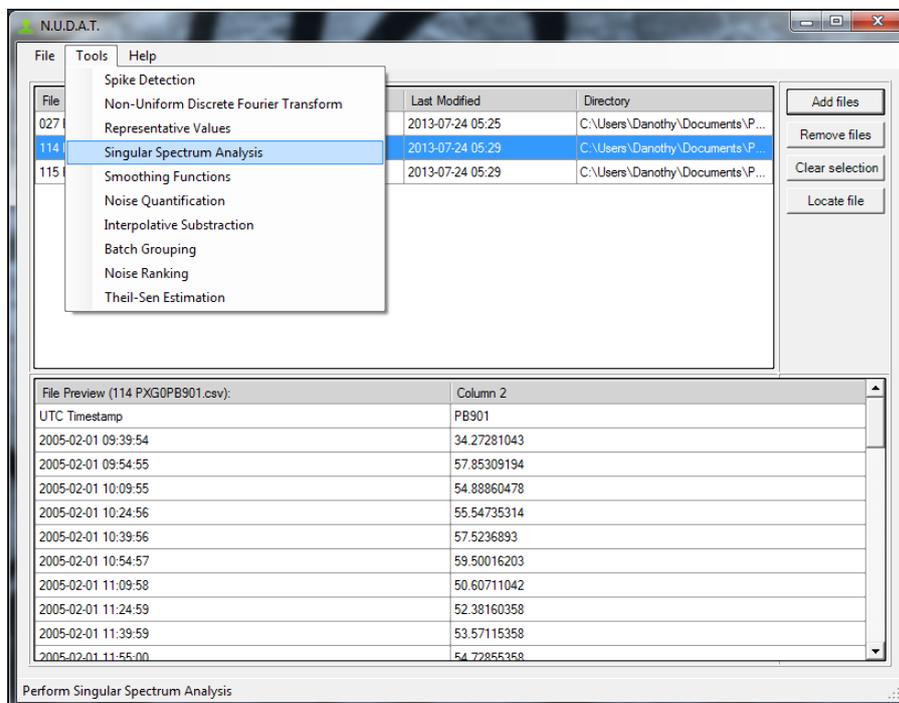


Figure 4.10: NUDAT's tool selection menu. Various processes can be applied to the input files.

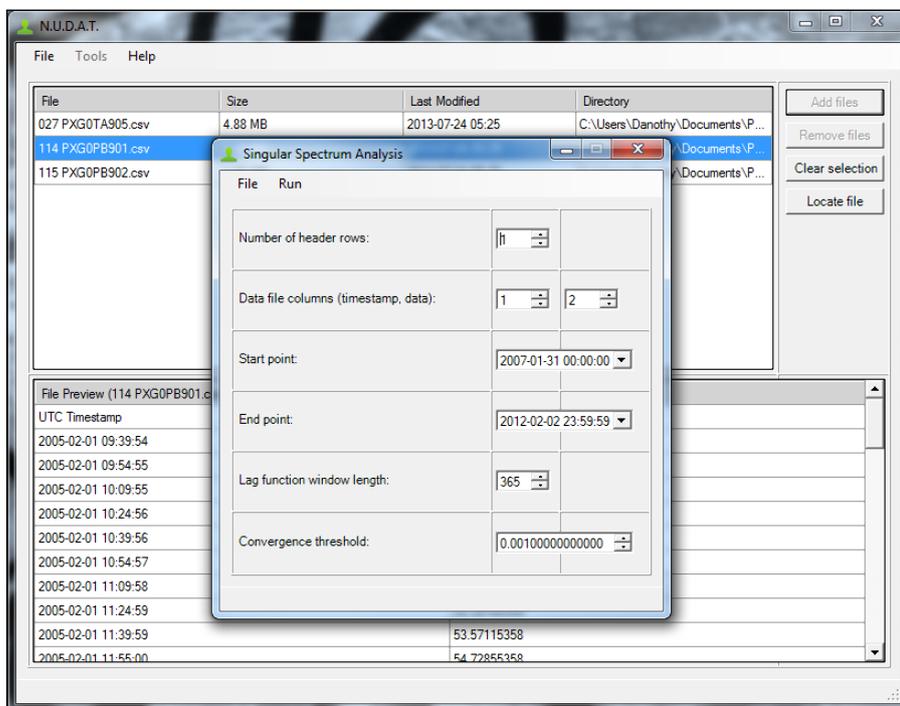


Figure 4.11: NUDAT's process settings interface. The SSA process settings are displayed.

Chapter Five

Toolkit verification

5.1 Introduction

In order to verify the analytical components of the Non-Uniform Data Analysis Toolkit (NUDAT) described and developed in Chapter Four, a series of test applications have been carried out. This chapter describes the development of an artificial dataset, containing two time series, each possessing a mixture of shared and unique features that are explicitly known and understood. The development of the features and their combination into two similar, but distinct, time series is described in Section 5.2. These time series are intended to possess similar features to those expected to be encountered in the Large Scale Gas Injection Test's (Lasgit's) dataset.

The artificially created dataset is then subjected to an Exploratory Data Analysis (EDA) via the application of NUDAT, with the toolkit performance measured by comparing its outputs to those expected from the known inputs. This process, described in Section 5.3, while similar to the intended application of NUDAT to Lasgit's dataset, also applies individual toolkit processes to the components of the artificial dataset in isolation as part of the verification process.

5.2 Verification time series data

In order to verify the collective data analysis capabilities of NUDAT, a dataset with characteristics representative of the dataset, but also with known content has been developed. Therefore an artificial time series based dataset was created, enabling the results of the application of NUDAT to be compared with the known inputs to the dataset. This comparison enables a characterisation of each of the toolkit capabilities, and a basis by which to assess the accuracy of implementation. The following sections describe the individual components of the created dataset and their incorporations into the two time series (referred to in this context as Time

Series I (TSI) and Time Series II (TSII)) that are ultimately used as verification data.

5.2.1 Non-uniformity in the verification dataset

Artificial non-uniformity has been included in the verification dataset in order to test the non-uniform capabilities of NUDAT. To achieve, this the verification dataset has been generated with a varying 'sample' rate. The artificial sample rate varies five times between 0.1 logs/hour and 0.4 logs/hour, leading to six periods of uniform sample rate with an overall average of 0.25 logs/hour across the dataset. Figure 5.1 shows the sample rate with time of the dataset developed for the verification process. There is a total of 16,381 datum points (i.e. 16,381 time indices) in each time series in the dataset. This is approximately 10% of the size of each time series in Lasgit's dataset. The average sample rate of the verification dataset is intended to reflect a similar proportion of Lasgit's dataset sample rate. The time indices generated by this sample rate profile will be common to all time series components (and hence the resulting verification time series).

5.2.2 Components of verification time series

The artificial dataset has been created in an additive fashion, with components intended to mimic phenomena observed in Lasgit's dataset. Specifically, the verification dataset consists of two artificially created time series, each possessing a unique trending component and noise component, and each sharing the same spiking, second order event, frequency content, and distortion components. A detailed description of each component and their combination is provided below.

5.2.2.1 Long term trends

Each verification time series possesses a trending component that acts as a base form of the time series, to which other components are added. The first time series (TSI) is based on a linear increase (Figure 5.2), and the second time series (TSII) based on a hyperbolic tangent function that spans the same magnitudes (Figure 5.3).

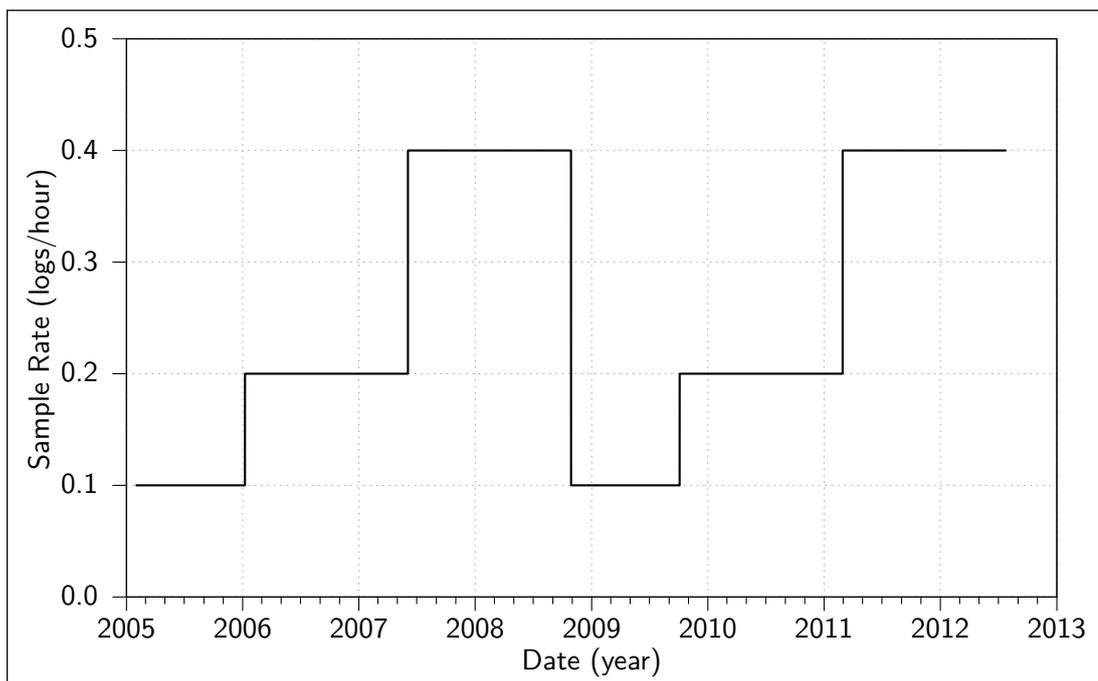


Figure 5.1: Non-uniform sample rate of the verification dataset.

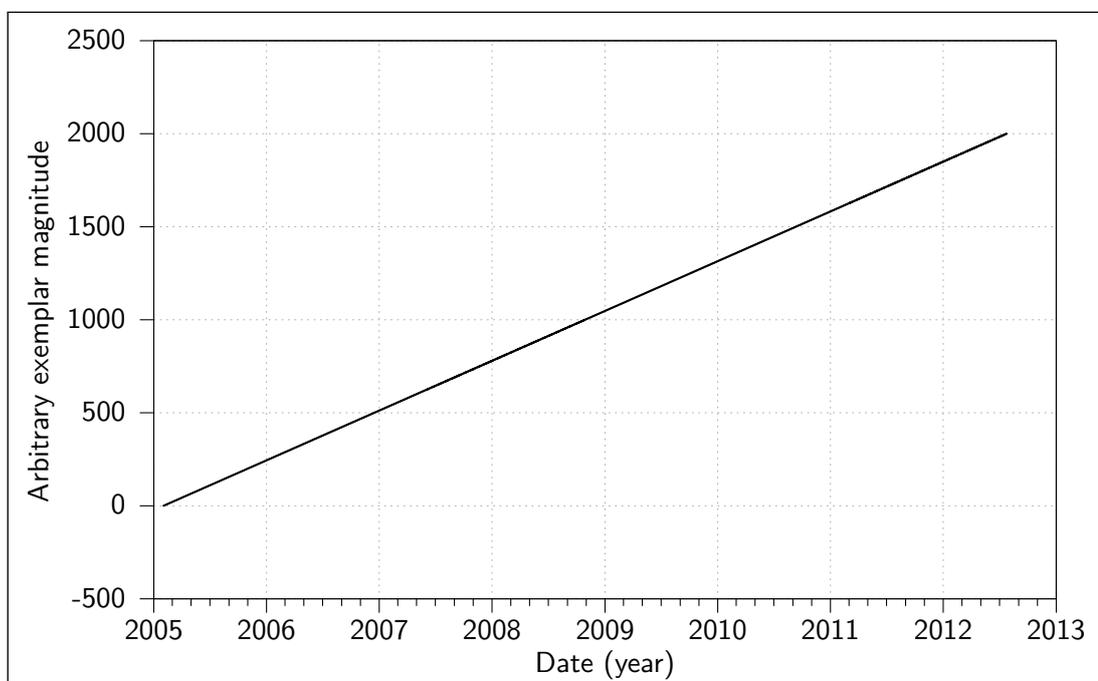


Figure 5.2: Verification dataset linear trend component. The linear trend varies constantly from zero magnitude to 2000 between 2005-02-01 and 2012-07-24.

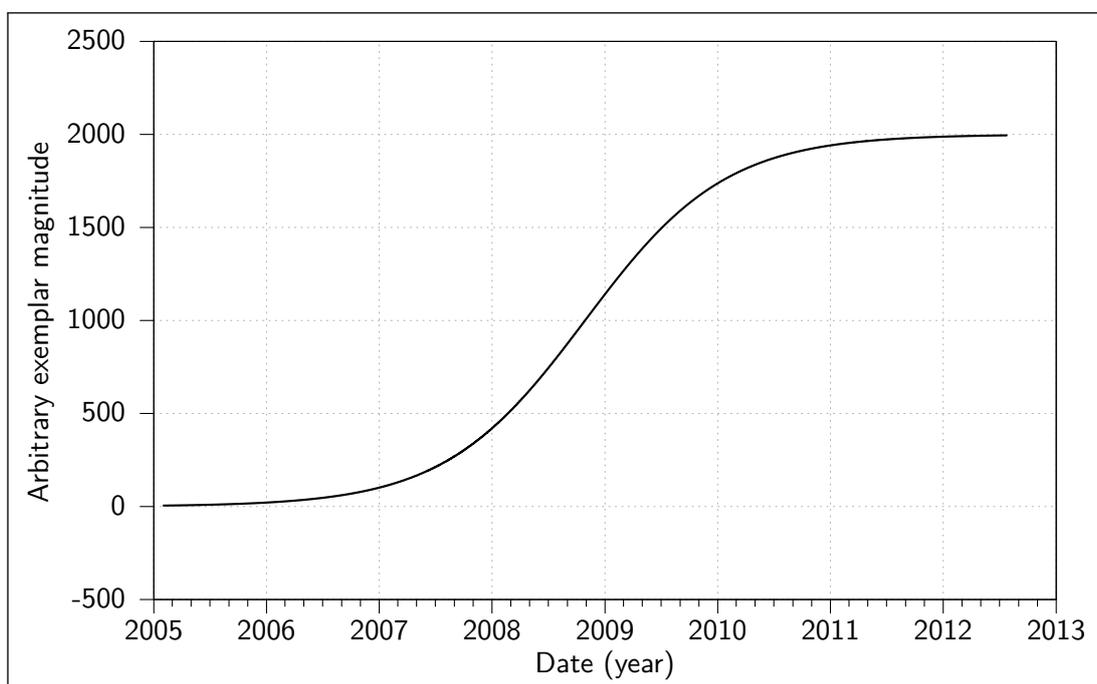


Figure 5.3: Verification dataset hyperbolic trend component. The hyperbolic trend varies according to the $\tanh(x) + 1$ function (where x varies linearly from -3 to 3) from approximately zero magnitude to approximately 2000 between 2005-02-01 and 2012-07-24.

5.2.2.2 Distortion

In addition to the long term trend components, a large distortion component (Figure 5.4), representative of an influence on a physical system that increases and then diminishes over a macro scale (i.e. non second order) length of time. The inclusion of such a distortion component is intended to provide insight into the ability of the Singular Spectrum Analysis (SSA) detection to identify macro scale behaviours with mixed or multiple rates of evolution inherent within them. The distortion component is included in both TSI and TSII.

5.2.2.3 Smaller scale 'second order' features/events

Of particular interest to the study of Lasgit's dataset is the sensitivity of NUDAT to the smaller scale or 'second order' features/events that may be located within and occluded by macro scale features of a time series. To test the ability to identify this content a small number of artificially generated 'events', in which a perturbation from the trend is induced and subsequently exponentially decayed away, are added to the trends. Figure 5.5 depicts the form, magnitude and location of the (fifteen) events generated for the verification dataset. The set of fifteen artificially created second order events are included in both TSI and TSII.

5.2.2.4 Frequency content

Assessment of the ability to quantify the frequency content of a time series has been achieved through the inclusion of a time series component containing the sum of the two sinusoids shown in Figure 5.6, each typical of natural cyclic variations (annual, daily). The first sinusoid possesses a wavelength of 365 days (a frequency of approximately 3.17×10^{-8} Hz) and an amplitude of 100. The second sinusoid possesses a wavelength of one day (a frequency of approximately 1.16×10^{-5} Hz) and an amplitude of 25. The frequency content component is included in both TSI and TSII.

The Non-uniform Discrete Fourier Transform (NDFT) of this frequency content, in isolation from the other components of the verification dataset, is depicted in Figure 5.7. Peaks at wavelengths of 365 days and one day can be seen with magnitudes of approximately 100 and 25 respectively, corresponding to the input frequency content.

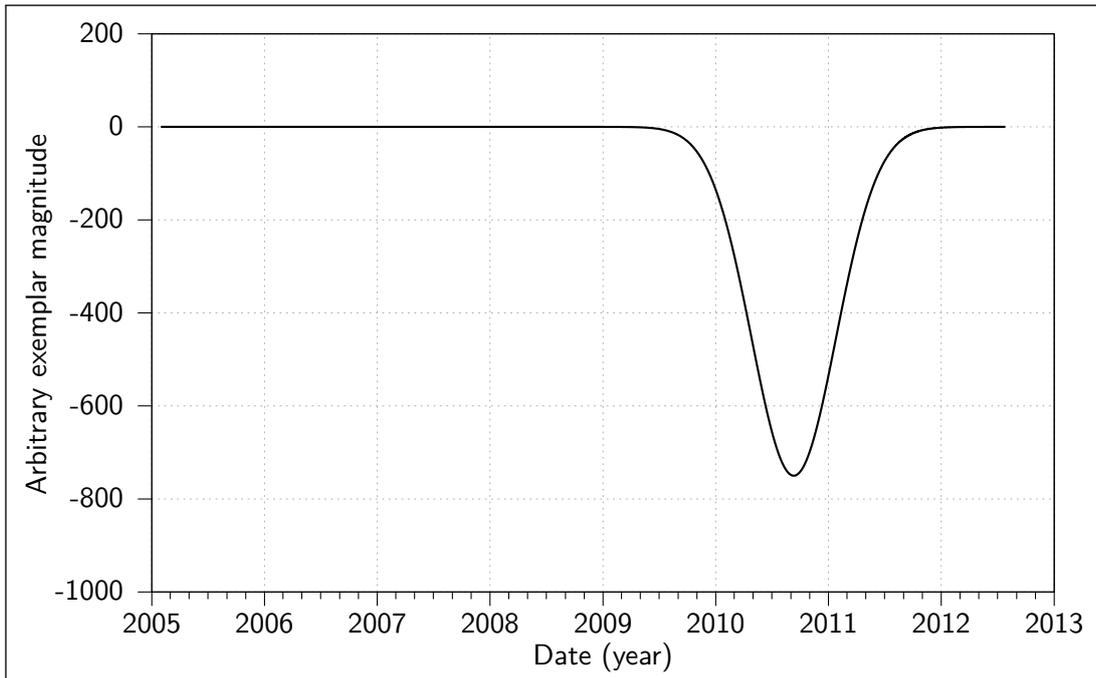


Figure 5.4: Verification dataset macro scale distortion component.

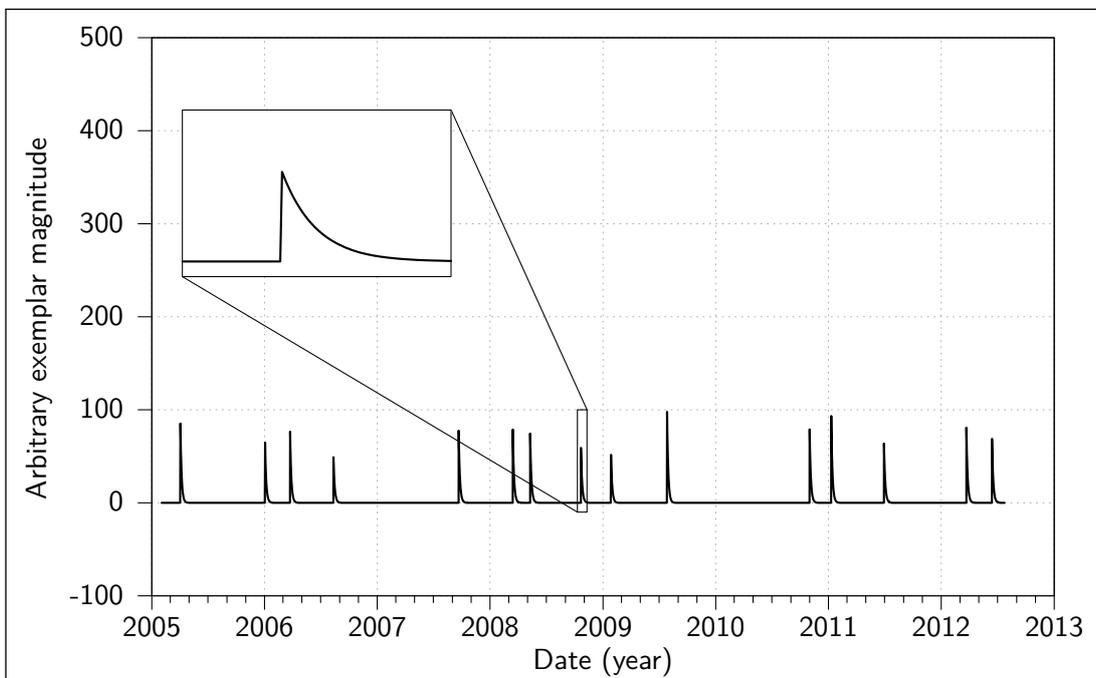


Figure 5.5: Second order events within the verification dataset. Fifteen randomly occurring events in which a small scale perturbation occurs and then decays in an exponential fashion.

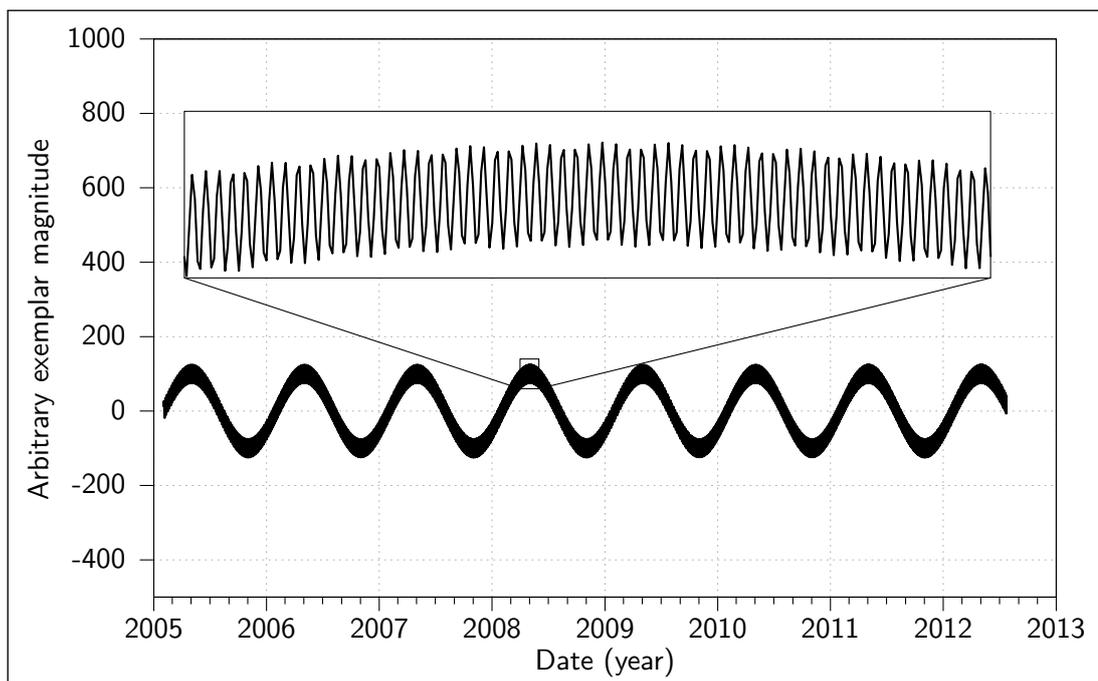


Figure 5.6: Verification dataset frequency content component. Two frequency components combined, one with a 365 day wavelength and magnitude 100, the other with a daily wavelength and magnitude 25.

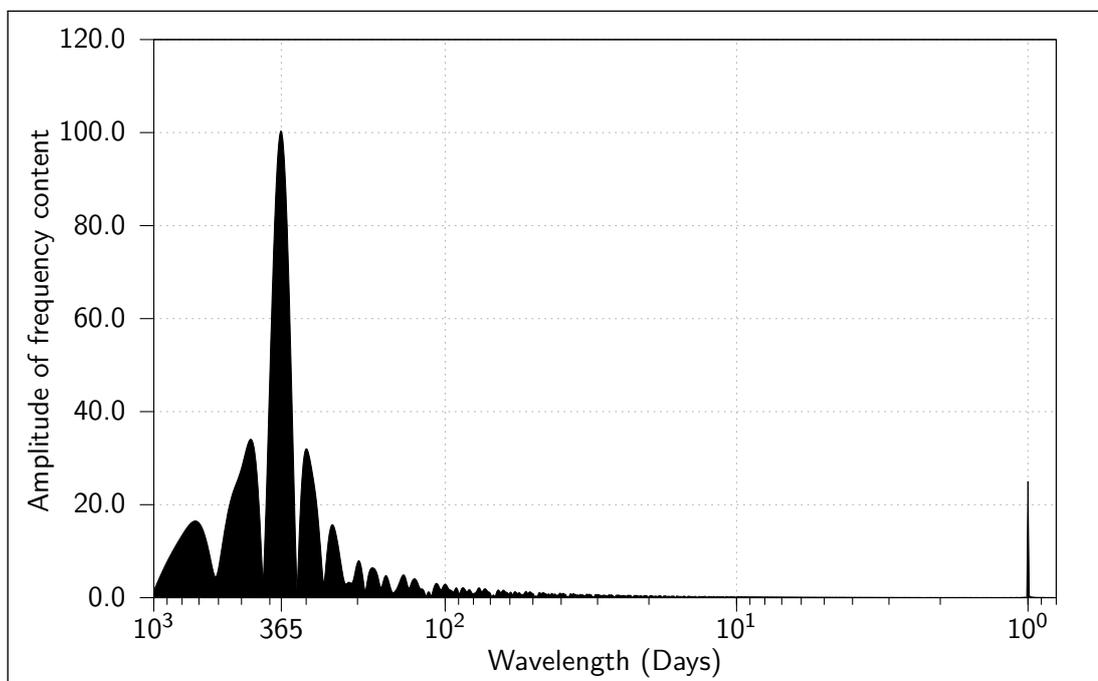


Figure 5.7: NDFT of verification dataset frequency content. Peaks of magnitudes 100 and 25 can be seen clearly at the 365 day and one day wavelengths respectively.

5.2.2.5 Anomalous points

A set of isolated anomalous (aberrant) points (also known as spikes), that are characterised by their individual and significant aberrance from the consensus of the surrounding data, have been randomly generated. The resulting artificial time series component is depicted in Figure 5.8. A total of 84 spikes have been generated with uniform probability of occurrence in time. This equates to approximately 0.5% of the datum points in the artificial time series. The selection of a positive magnitude for the artificial spikes qualitatively reflects the predominance of positive spiking in Lasgit's dataset. The sign convention of the spikes, however, has no mathematical significance to the analysis performed by NUDAT, as detailed in Chapter Four. The spiking component is incorporated into both TSI and TSII.

5.2.2.6 Noise

Two noise profiles have been generated for incorporation into the verification dataset. The first (Figure 5.9) is uniformly distributed about zero in the range $-25 \leq x \leq 25$. The uniform noise component was incorporated into TSI (the linearly trending artificial time series). The second (Figure 5.10) is normally (Gaussian) distributed about zero with a Standard Deviation (SD) of $\frac{25}{3}$ (leading to 99.7% of points lying in the range $-25 \leq x \leq 25$). The Gaussian noise component was incorporated into TSII (the hyperbolically trending artificial time series).

The distributions and NDFTs of the uniform and Gaussian noise components in isolation are shown (respectively) in Figures 5.11, 5.12, 5.13 and 5.14. Inspection of these figures suggests that the pseudo-random process used to generate the noise components has approximated the expected distributions well and not introduced or excluded any specific frequency components.

5.2.3 Aggregated verification time series

The components described in the preceding sections were combined into two similar but distinct verification time series. Figures 5.15 and 5.16 show the time series (TSI and TSII) that are the aggregate (summation) of the artificial time series components. As stated above TSI, the linear trending time series, utilises the uniformly distributed noise component, while TSII, the

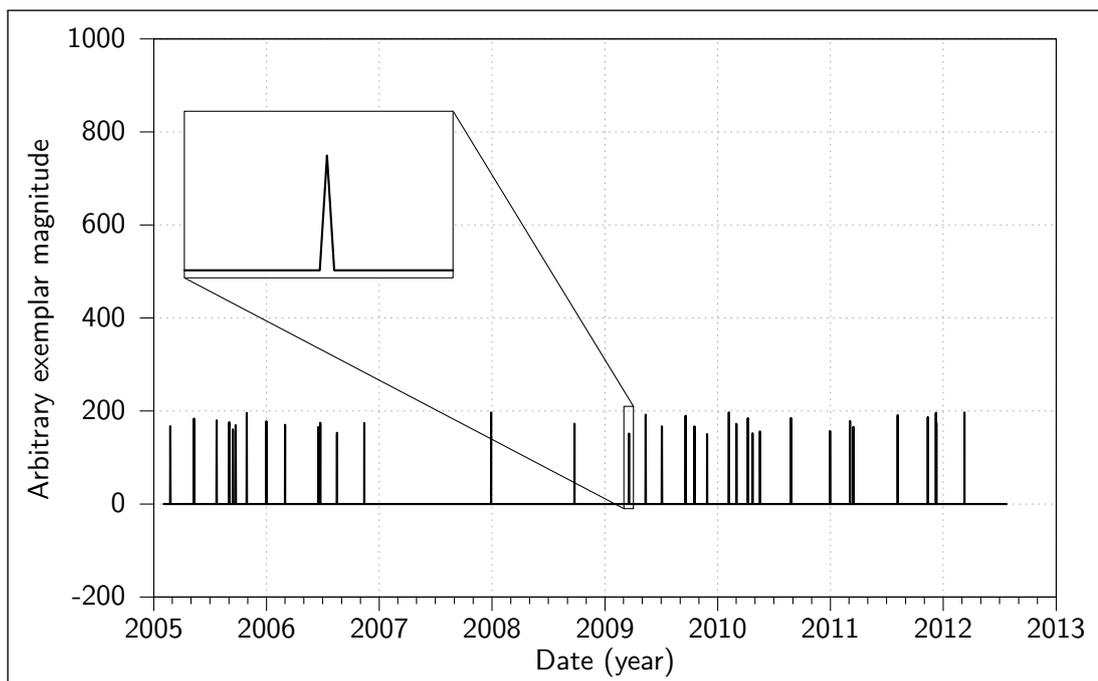


Figure 5.8: Spiking component of the verification dataset. 84 randomly generated single points aberrant from the consensus of their surrounding points.

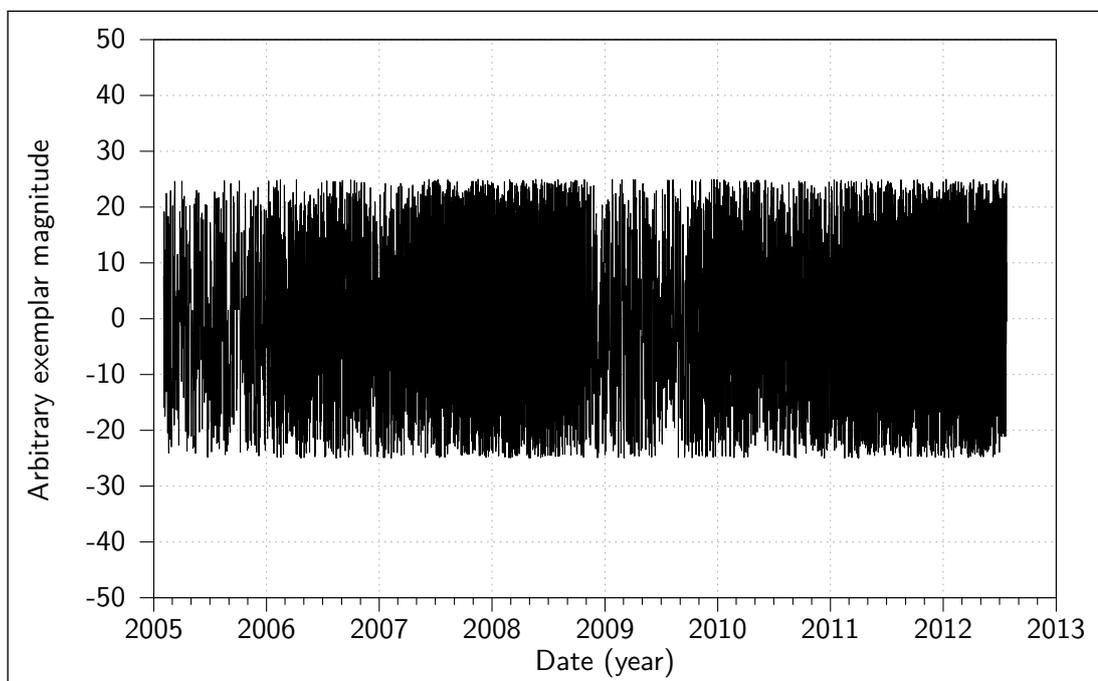


Figure 5.9: Verification dataset uniformly distributed noise component. The noise ranges between -25 and 25 in magnitude with equal frequency of occurrence at intermediate values.

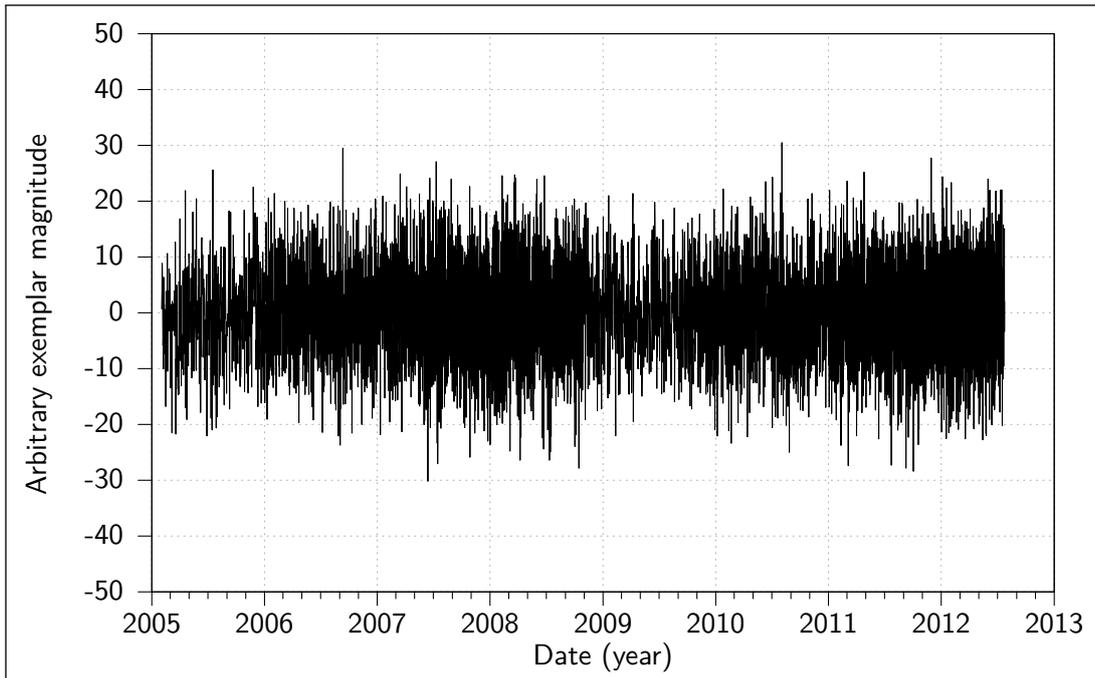


Figure 5.10: Verification dataset Gaussian distributed noise component. The noise has a mean of zero and a SD of $\frac{25}{3}$ (leading to 99.7% of points lying between -25 and 25).

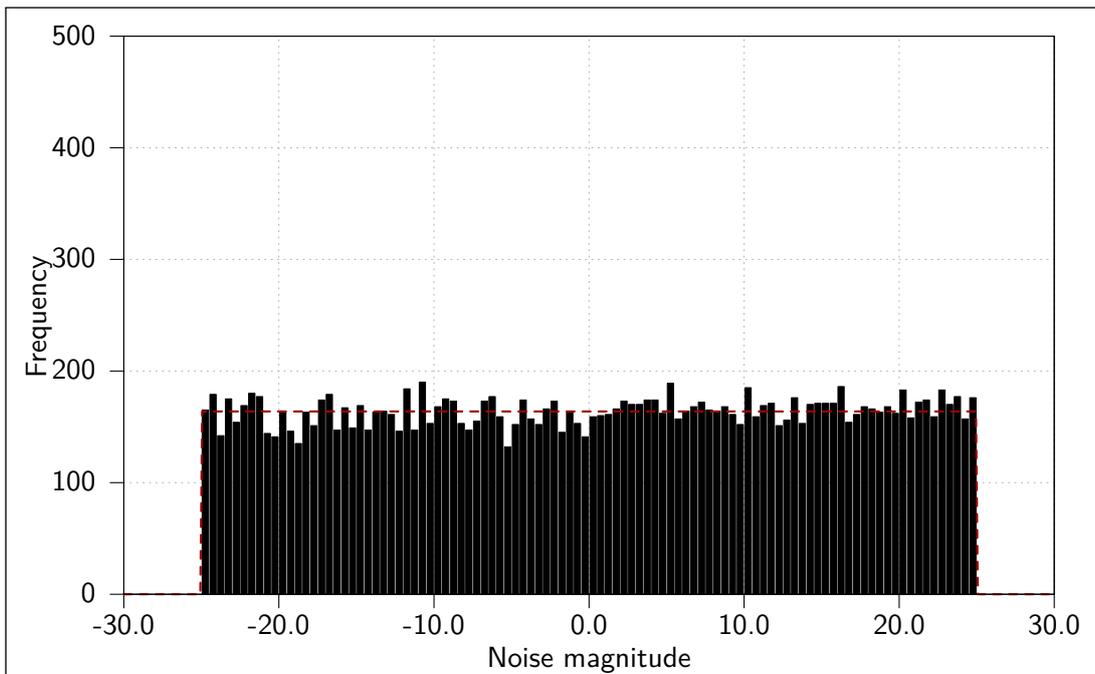


Figure 5.11: Distribution of verification dataset uniform noise component.

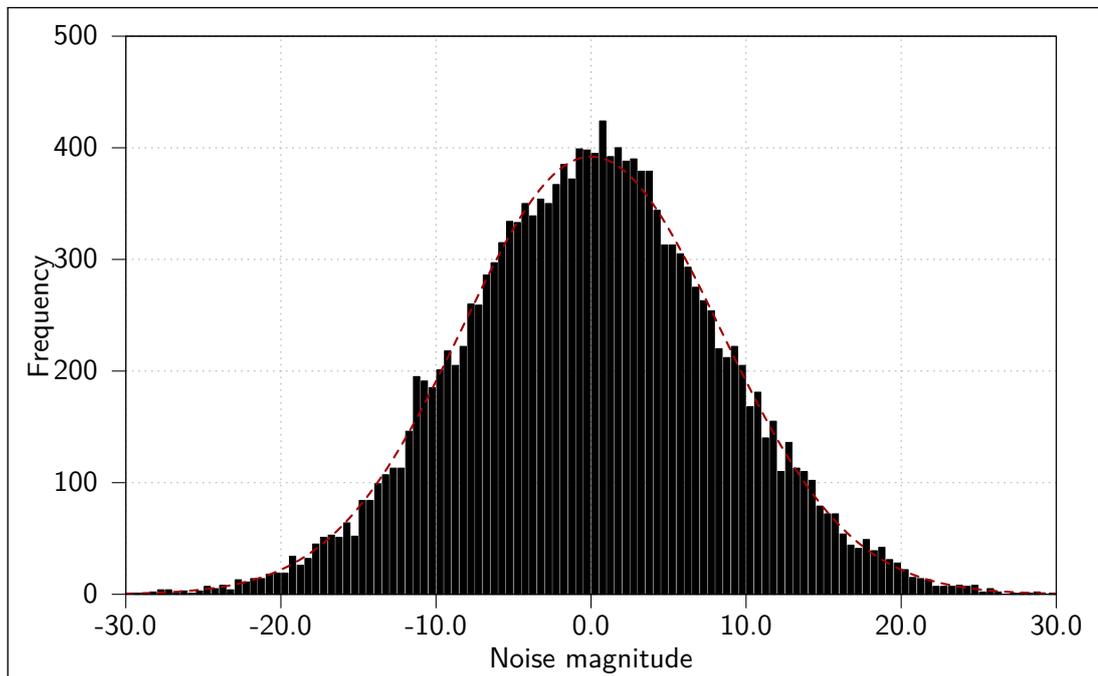


Figure 5.12: Distribution of verification dataset Gaussian noise component.

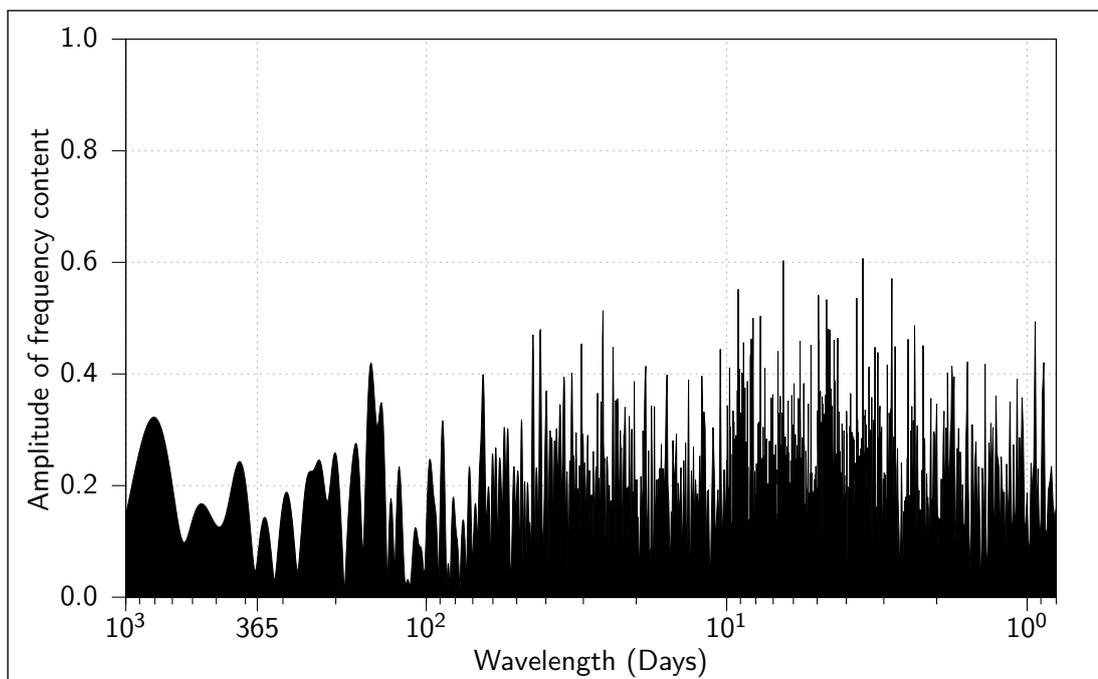


Figure 5.13: NDFT of verification dataset uniform noise component.

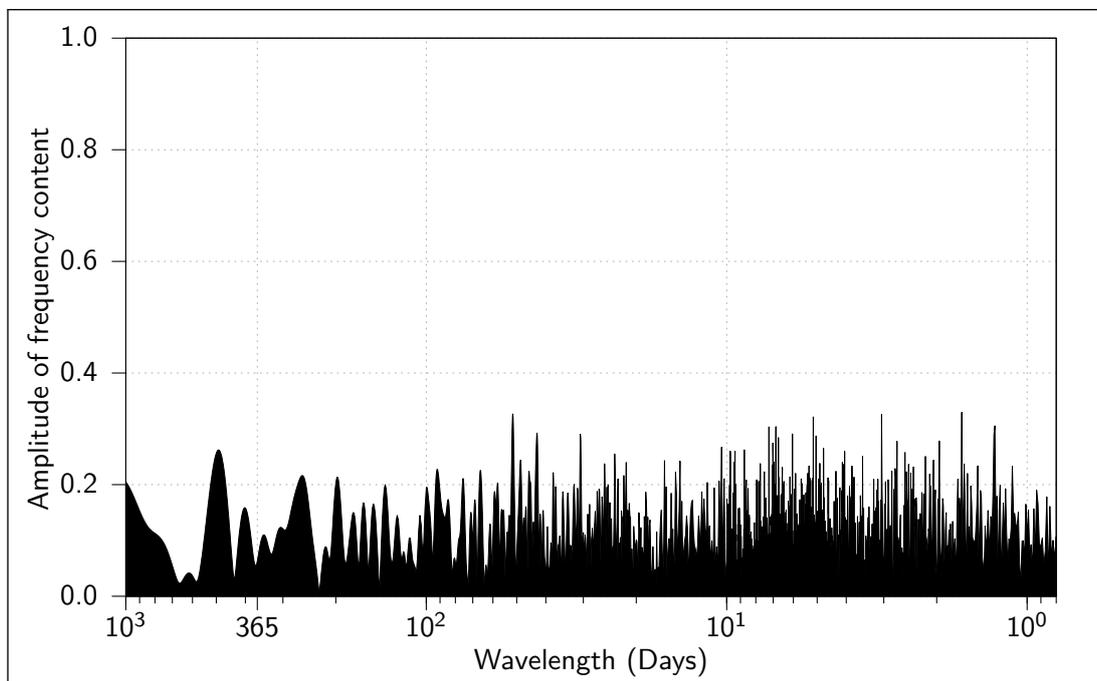


Figure 5.14: NDFT of verification dataset Gaussian noise component.

hyperbolic trending time series, utilises the Gaussian distributed noise. Additionally, each time series aggregates all the remaining components: the distortion component, the second order event component, the frequency content component, and the spiking component.

5.3 Toolkit interpretation of the verification dataset

The application of NUDAT to the artificial verification dataset (depicted in Figure 5.17) consisted of a sequential application of the following processes:

1. Spike detection (and subsequent removal)
2. Down-sampling to daily representative values
3. SSA with a 365 day lagged vector window to detect macro-scale trends
4. Removal (subtraction) of the SSA derived trends from full resolution, de-spiked data
5. Analysis of de-trended data using the NDFT process
6. Event candidate detection of residual data

5.3.1 Individual process evaluations

5.3.1.1 Spike detection

Tables 5.1 and 5.2 quantify the number of points correctly identified as spikes in each time series, along with the number of type I (false positive) and type II (false negative) errors made by the process (Devore and Peck 2005). Figures 5.18 and 5.19 visually depict the spike identification results for each time series. The datum points identified as spikes are shown in black, with the remaining data overlain in grey.

The single undetected spike (the false negative) in TSI is due to a suppression in the magnitude of the spike relative to the baseline signal and a comparatively higher noise level due to the uniform distribution of artificially created noise. The spike in question was randomly generated with a magnitude close to the lower bound specified and was combined with a noise component that possesses a negative value. Additionally, the baseline noise level, and therefore threshold for spike detection was higher in the case of the linear based artificially created time series. The

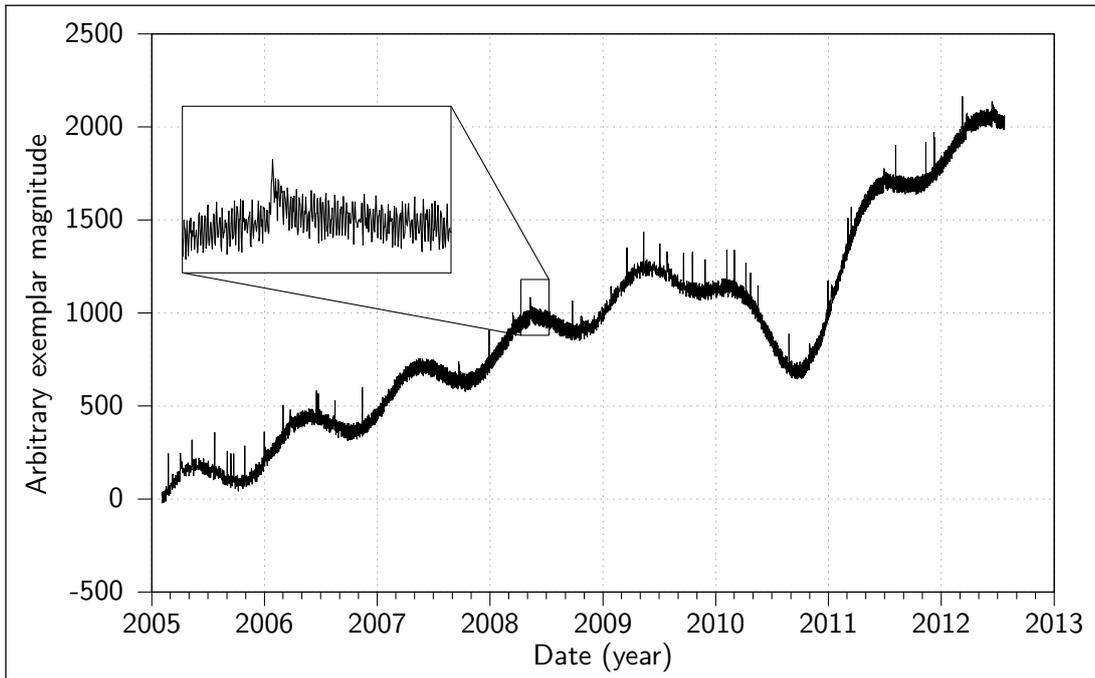


Figure 5.15: Verification dataset linear based time series (TSI). The common components aggregated with the linear trend and uniformly distributed noise components.

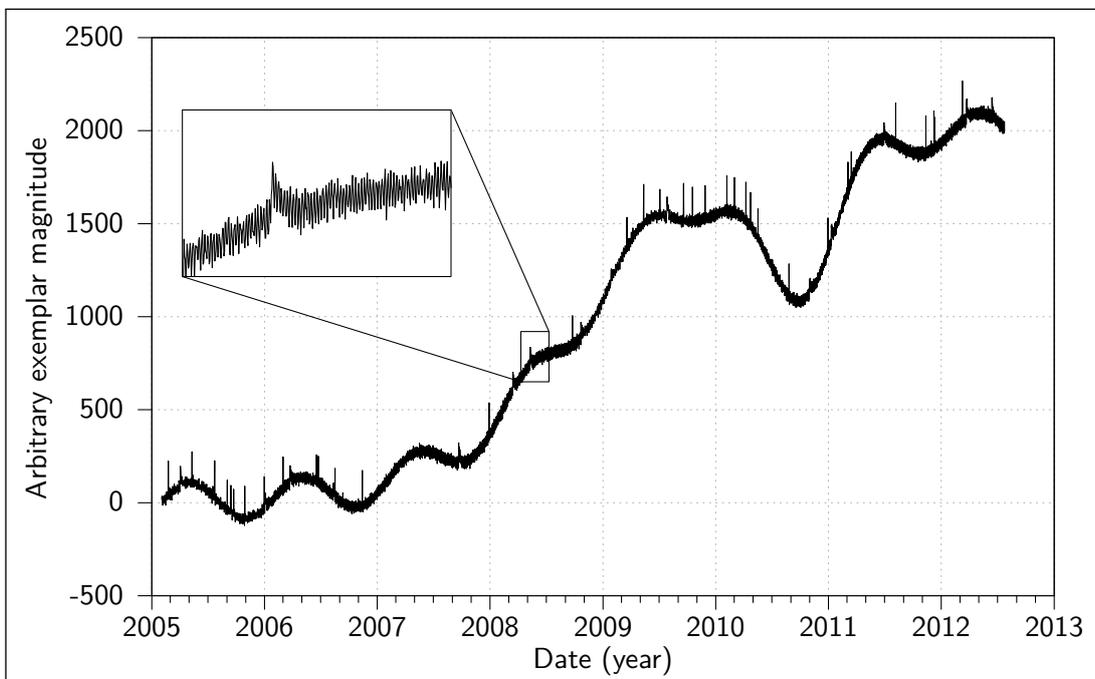


Figure 5.16: Verification dataset hyperbolic based time series (TSII). The common components aggregated with the hyperbolic trend and Gaussian distributed noise components.

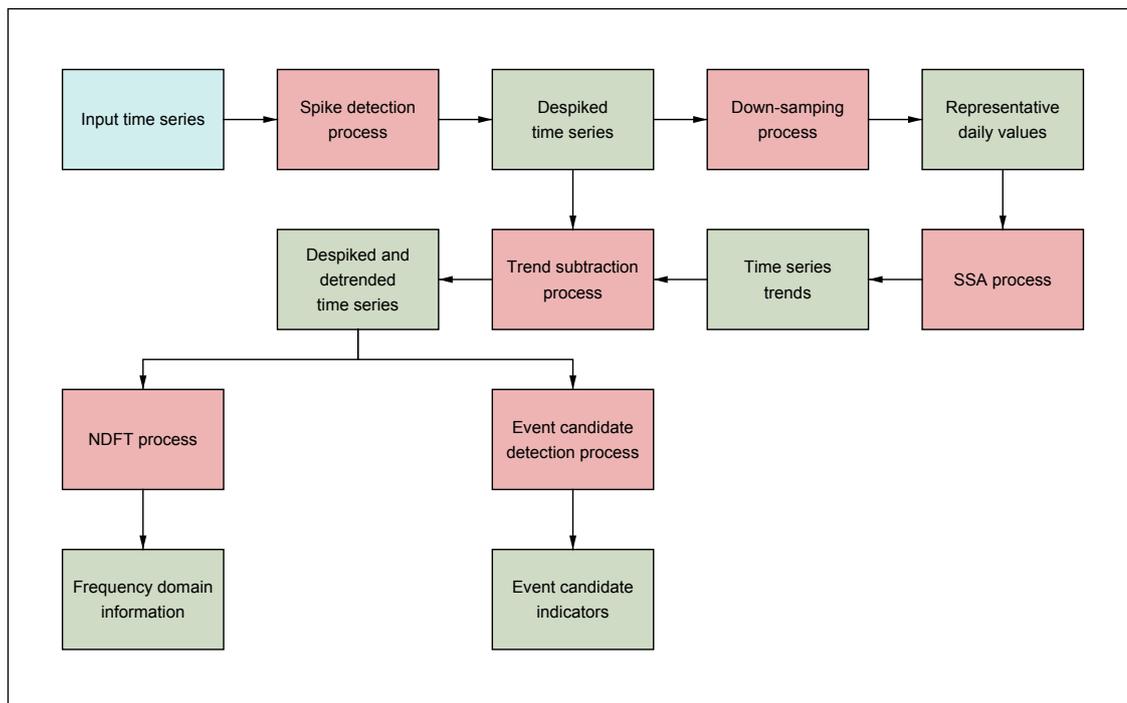


Figure 5.17: Toolkit processes as applied to the artificial verification dataset, and their subsequent outputs as inputs to the next process.

Table 5.1: Enumeration of spikes detected by NUDAT in TSI along with enumeration of type I and type II detection errors.

| | Detected | Not detected | Actual value |
|----------|----------|--------------|--------------|
| Spike | 83 | 1 | 84 |
| No spike | 2 | 16,295 | 16,297 |

Table 5.2: Enumeration of spikes detected by NUDAT in TSII along with enumeration of type I and type II detection errors.

| | Detected | Not detected | Actual value |
|----------|----------|--------------|--------------|
| Spike | 84 | 0 | 84 |
| No spike | 6 | 16,291 | 16,297 |

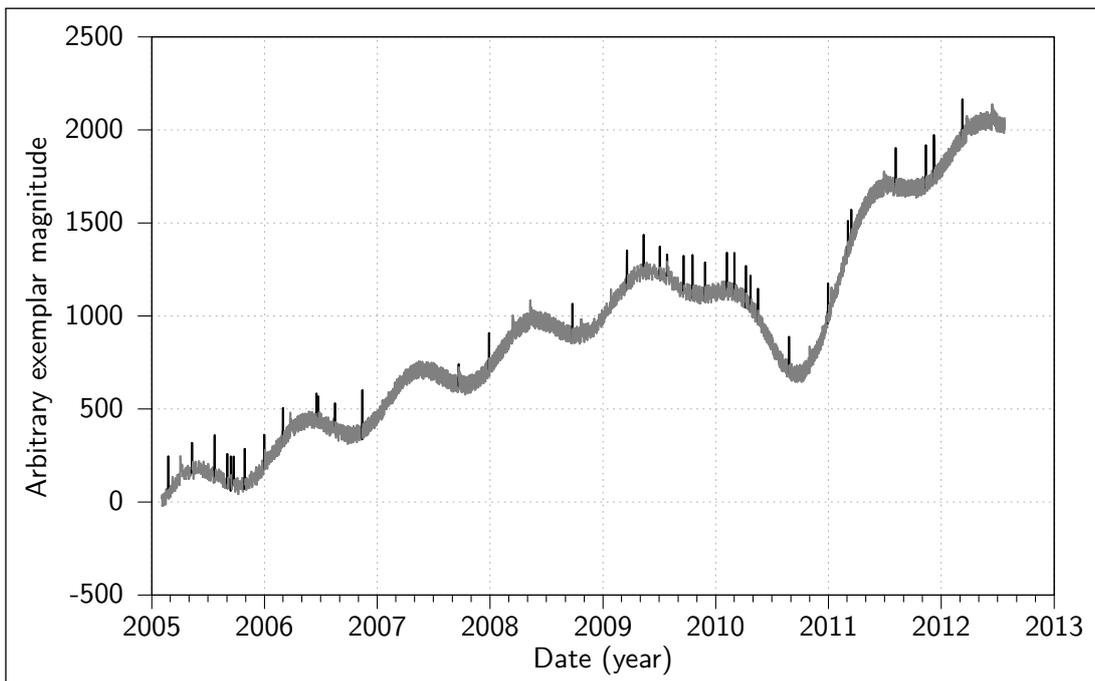


Figure 5.18: Spike detection process applied to TSI. Datum points identified as spikes (statistically anomalous points) in black, with non-spiking in grey.

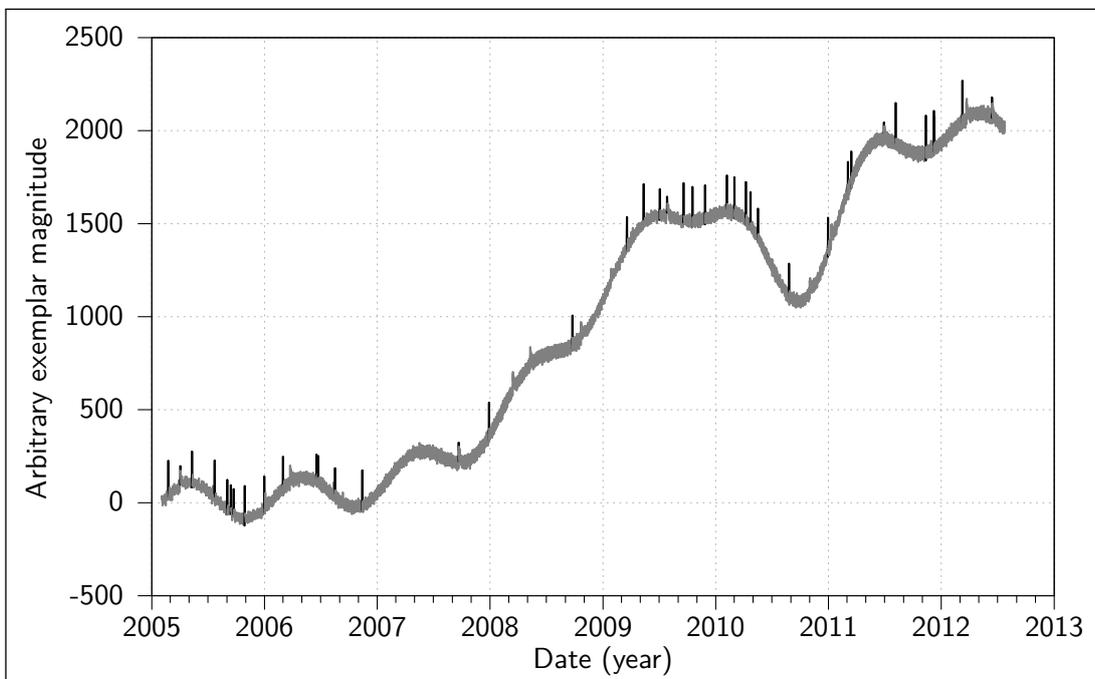


Figure 5.19: Spike detection process applied to TSII. Datum points identified as spikes (statistically anomalous points) in black, with non-spiking in grey.

spike in question was approximately 2.42 SDs from the expected value. The marginally higher number of occurrences of non-spikes identified as spikes (false positive) in the hyperbolic based artificially created time series is attributed to the application of Gaussian noise that has no absolute bounding, leading to the infrequent inclusion of outlying points.

The anomalous point (spike) detection process was found to be successful in identifying the artificially created anomalous points in the artificial dataset. The detection process identified all but one spike out of a total of 168 (84 per time series), and mis-identified a total of eight non-spikes as spikes from a total of 32,594 (16,297 per time series).

5.3.1.2 Trend detection (SSA)

The application of the SSA process generated one trending component for each time series in the artificially created dataset. In each case, the trend component was associated with the eigenvalue with the largest magnitude at the decomposition stage. Figures 5.20 and 5.21 show the SSA derived trends for TSI and TSII respectively.

Visual inspection of Figures 5.20 and 5.21 shows that the SSA derived trends reflect qualitatively the nature of the artificially created time series. The derived trending components manifest approximately as an increase at an appropriate rate with a downward distortion starting at the beginning of 2010 and returning to normal before the end of 2012 for both TSI and TSII.

The ability of the SSA process to derive the underlying trend can be directly evaluated due to the artificial nature of the dataset. To that end, Figures 5.22 and 5.23 compare the SSA derived trends to the linear and hyperbolic trend components combined with the distortion component.

No oscillatory behaviour is apparent in either of the SSA derived trends despite the frequency content included in the artificially created dataset. This is attributed to the length of the lagged vector window (365 days), which allows the decomposition process window to encompass a full cycle of the frequency content, and hence derive a trending component that averages that cycle.

Large sections of the SSA derived trends closely match the magnitude and form of the linear and hyperbolic input components. The discrepancies between the input trend components and

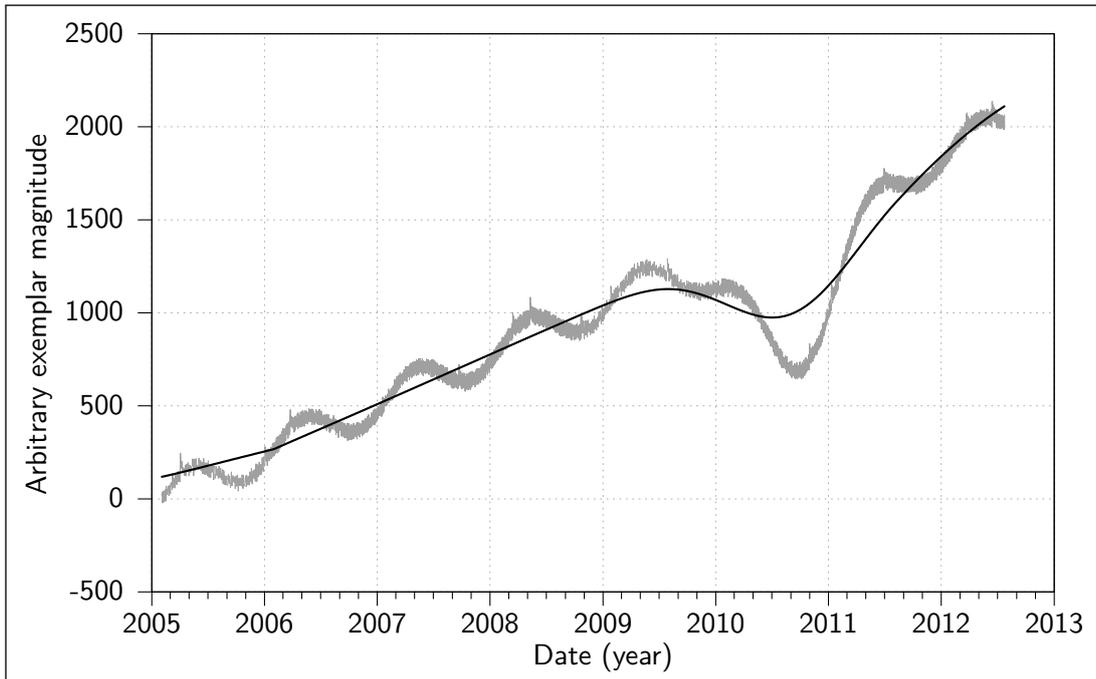


Figure 5.20: The SSA derived trend (black) of TSI after spike removal (grey).

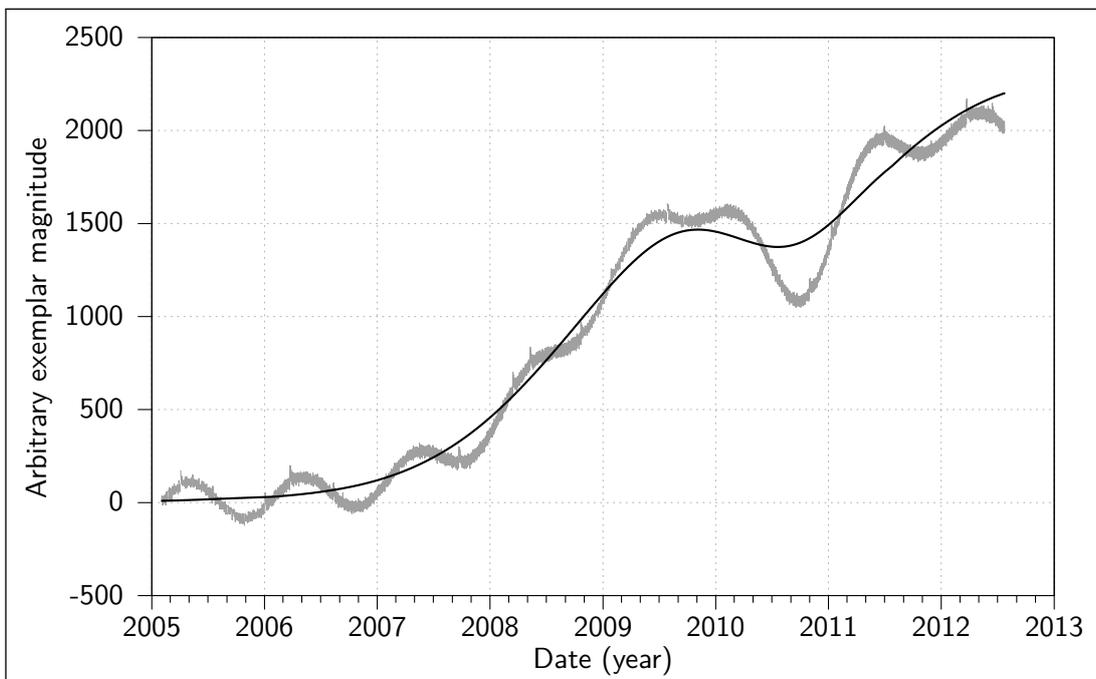


Figure 5.21: The SSA derived trend (black) of TSII after spike removal (grey).

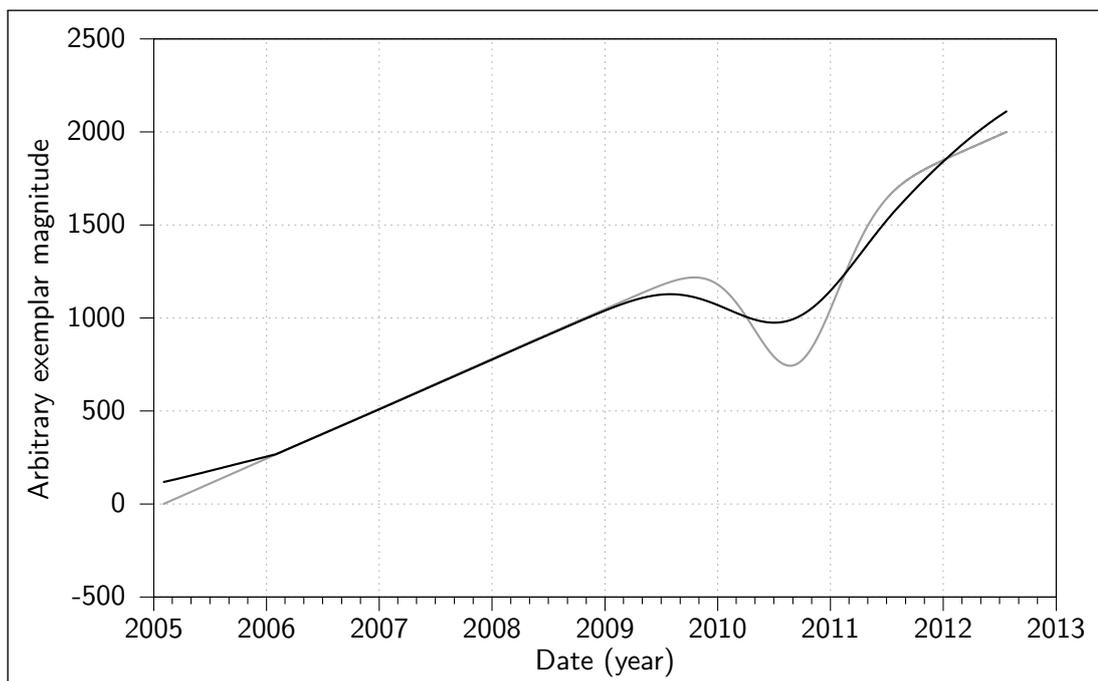


Figure 5.22: The SSA derived trend (black) compared to the linear and distortion components of TSI (grey).

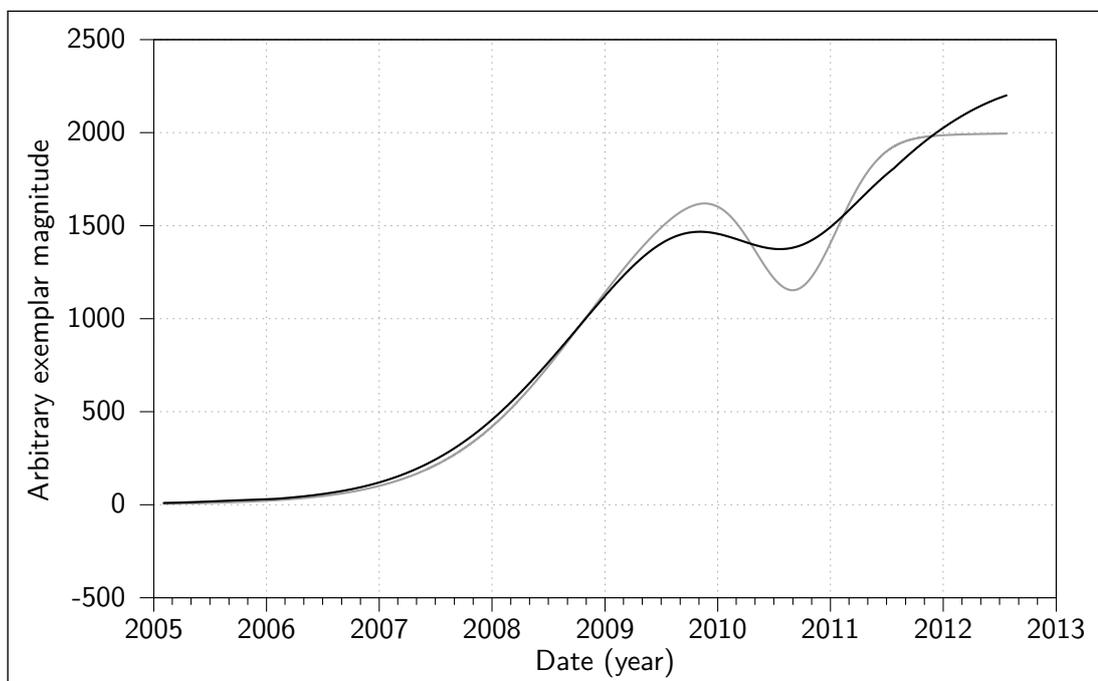


Figure 5.23: the SSA derived trend (black) compared to the hyperbolic and distortion components of TSII (grey).

the SSA derived trends are however predominantly concentrated around the distortion in both the linear and hyperbolic based time series. In each case the distortion is under represented in magnitude and over represented in duration. This is likely due to the length of the lagged vector window being large enough to partially smooth out the distortion.

Additionally, both SSA derived trends have a positive discrepancy at the end of the time series that is attributed to the proximity of the distortion to the end of the time series. The discrepancy between the SSA derived trend and the linear based time series in the first year is not accounted for.

The ability of the SSA process to derive a close approximation of the underlying trends in TSI and TSII non-parametrically (i.e. derived from the input data, rather than fitting an operator prescribed form) equips the toolkit with the ability to perform trend identification/removal on arbitrary time series. As minimal foreknowledge of the nature of the input data is required subjectivity associated with manual trend fitting is reduced.

5.3.1.3 Event candidate detection

Subtraction of an SSA derived trend from a time series produces a residual in which the smaller scale or 'second order' features in the dataset are, in terms of relative magnitudes, less occluded than in the original. Measurement of the local SD for the purposes of second order event candidate detection therefore will likely be more sensitive when applied to a de-trended time series.

It should be noted that TSI has a marginally higher baseline SD than TSII. This is due to the use of the uniform based noise component rather than the Gaussian based noise component. While the SD of the Gaussian distributed noise component is explicitly set to $\sigma = \frac{25}{3} \approx 8.3$ (Section 5.2.2.6), the SD of a uniformly distributed noise component that spans the equivalent of 3σ from a mean of zero (i.e. -25 to 25) is $\sigma = \frac{50}{\sqrt{12}} \approx 14.4$.

Figures 5.24 and 5.25 respectively show the de-trended linear and hyperbolic time series, along with the magnitude of the local SD with time. The isolated and brief increases in SD indicate a local perturbation from the local baseline. However they do not provide any information relating to the nature of the perturbation, and hence only highlight a location as an event candidate.

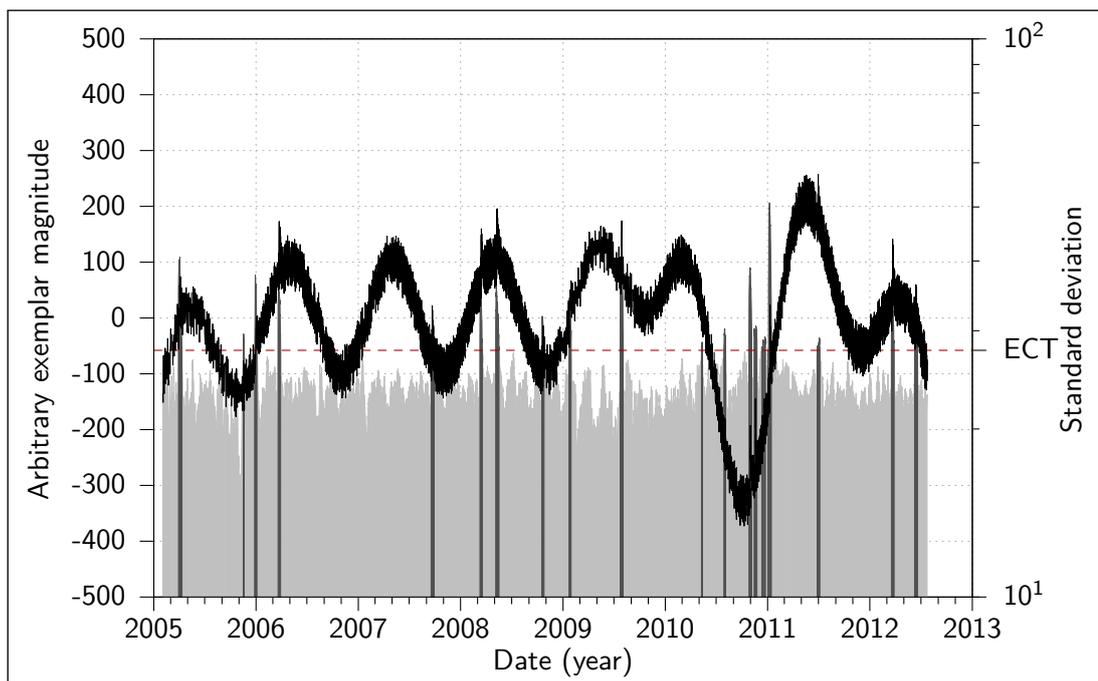


Figure 5.24: TSI de-trended (black), with local standard deviation with time (light grey), and event candidates (dark grey).

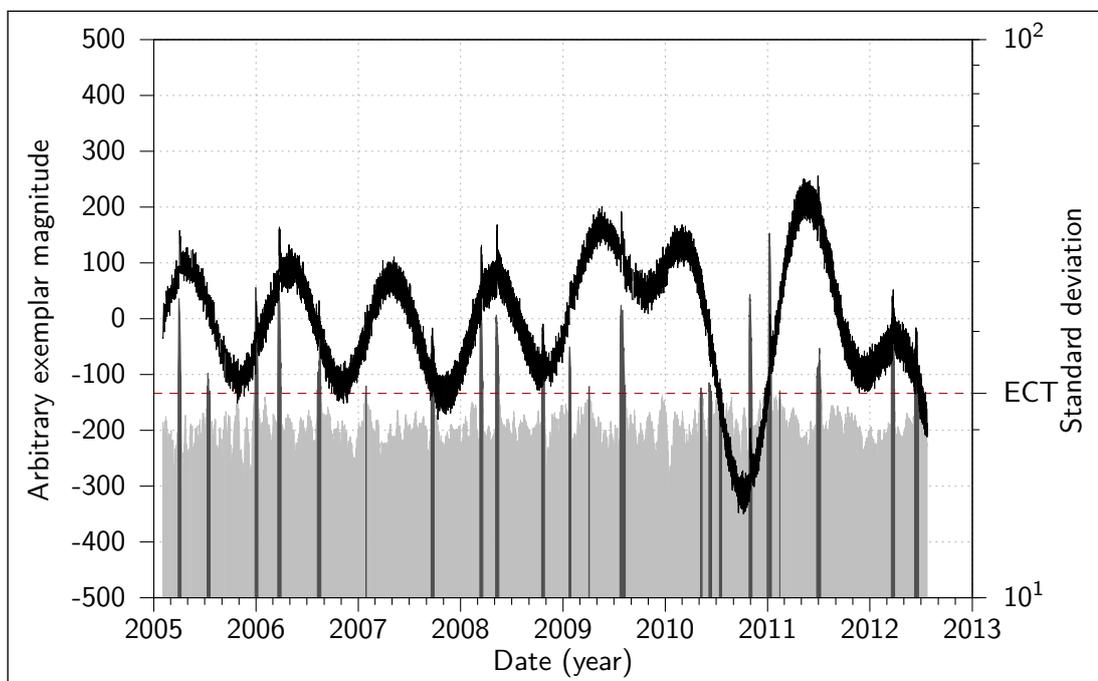


Figure 5.25: TSII de-trended (black), with local standard deviation with time (light grey), and event candidates (dark grey).

A classification of what constitutes the identification of an event candidate, in terms of relative prominence (i.e. brief, isolated increases) of the local SD, is hampered by the subjective nature of what relative increase in magnitudes may be considered significant, and how isolated any increase may be. The dark grey event candidates depicted in Figures 5.24 and 5.25 consist of the local SD values that are upper outliers, i.e. points with magnitudes beyond $1.5 \times$ Interquartile Range (IQR) above the upper quartile of the noise level (labelled as Event Candidate Threshold (ECT) and depicted with a red dashed line). Furthermore, the multiplier applied to the IQR can be set high and then incrementally reduced, thus identifying event candidates in descending order of significance.

Visual inspection of the event candidate detection plots suggests 21 event candidates for the linear based artificially created time series, and 22 event candidates for the hyperbolic based artificially created time series, based on the number of separate ECT breaking peaks. A detailed view of the 15 regions containing the second order events depicted in Figure 5.5 are shown in Figures 5.26 and 5.27 for TSI and TSII respectively.

With one exception, the artificial second order events are correctly highlighted by the event candidate detection process in both TSI and TSII. The event location inspected in Subfigure 5.26d shows an absence of event detection, but also an absence of the signal perturbation that the applied second order event should have created. This is due to a convolution of the event and the randomly generated noise at that point, resulting in a suppression of the intended perturbation. Additionally, the false negative spike (recorded in Tables 5.1 and 5.2) that remains in the artificially created verification dataset after de-spiking is shown in Subfigure 5.26k.

The intensities of the event candidate detection process can also be aggregated and compared in 'heatmap' plots. This allows for a visual inspection of the synchronisation of event candidates between multiple time series within a dataset. An example of such an aggregation is shown in Figures 5.28 and 5.29. Figure 5.28 shows the event candidate intensities for each time series with time along with markers that indicate the location of the artificially created second order events. Figure 5.29 shows the same information without the second order event markers. Additionally, heatmaps with more selective time windows may be used to quickly identify all other time series in a dataset that have correlating event candidates with an specific event observed in any particular individual time series.

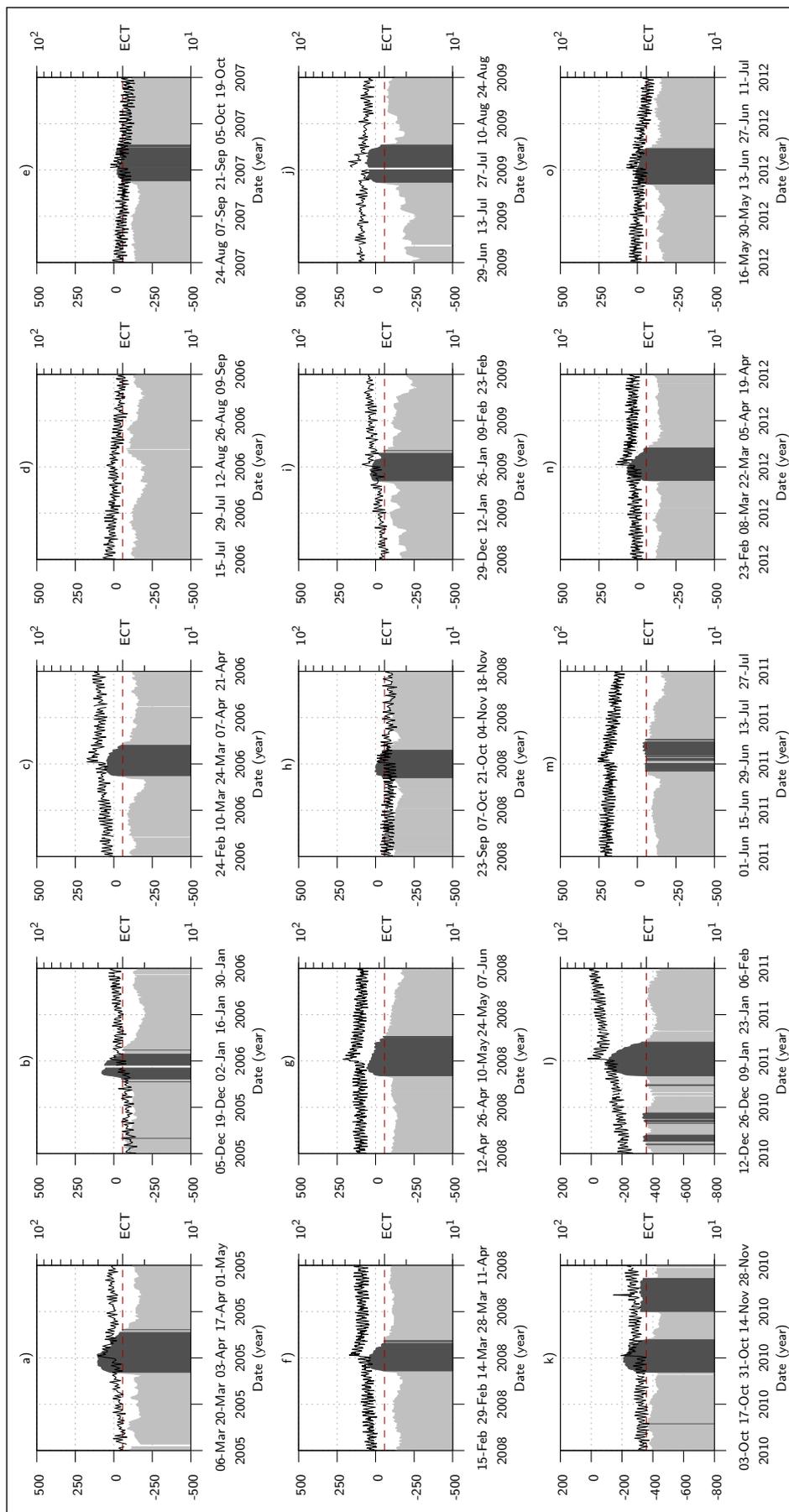


Figure 5.26: Detailed view of the second order events in TSI. Event detection is absent in Subfigure d, while a spike is also detected in Subfigure k.

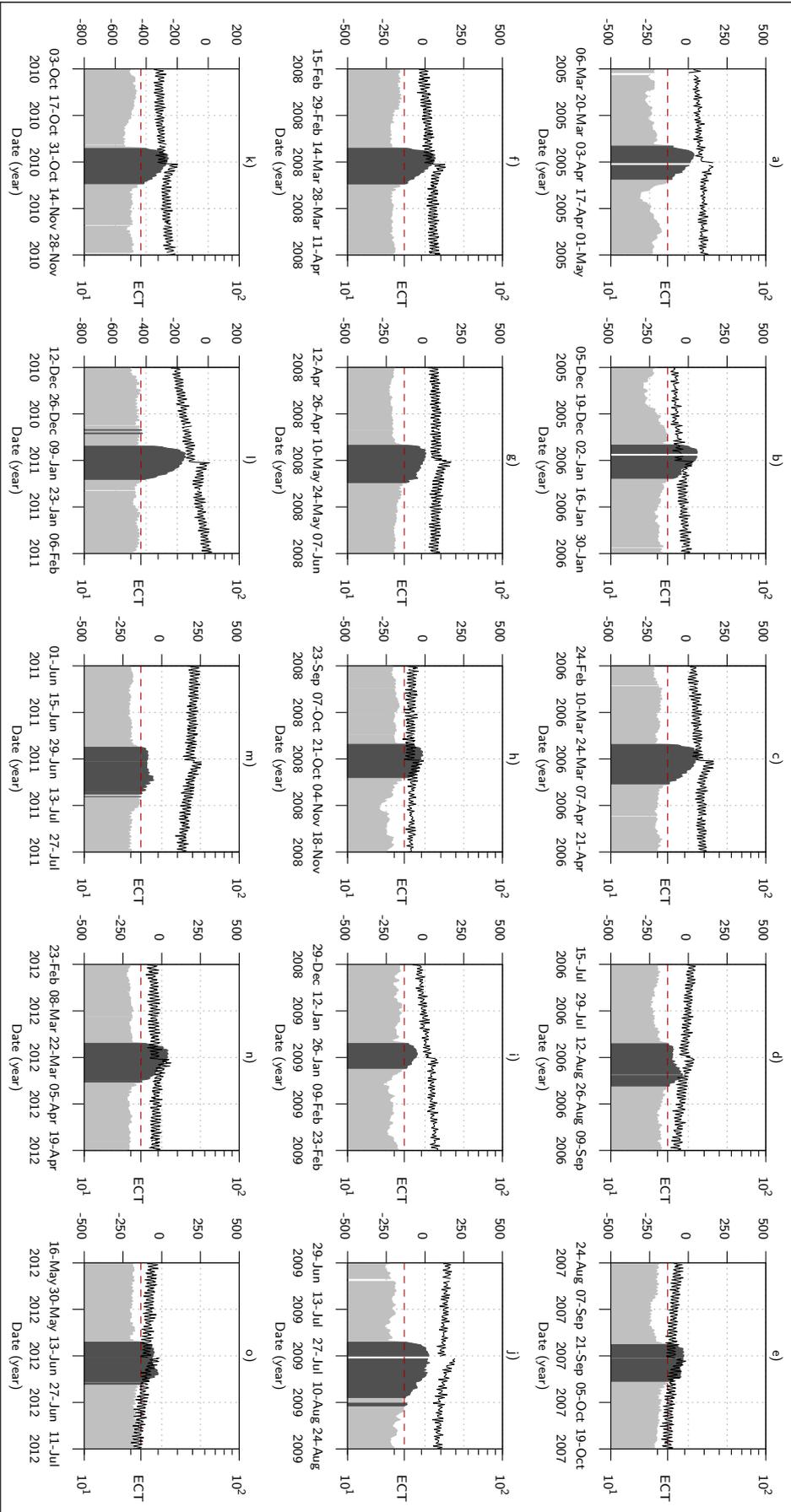


Figure 5.27: Detailed view of the second order events in TSII. Event detection is as expected.

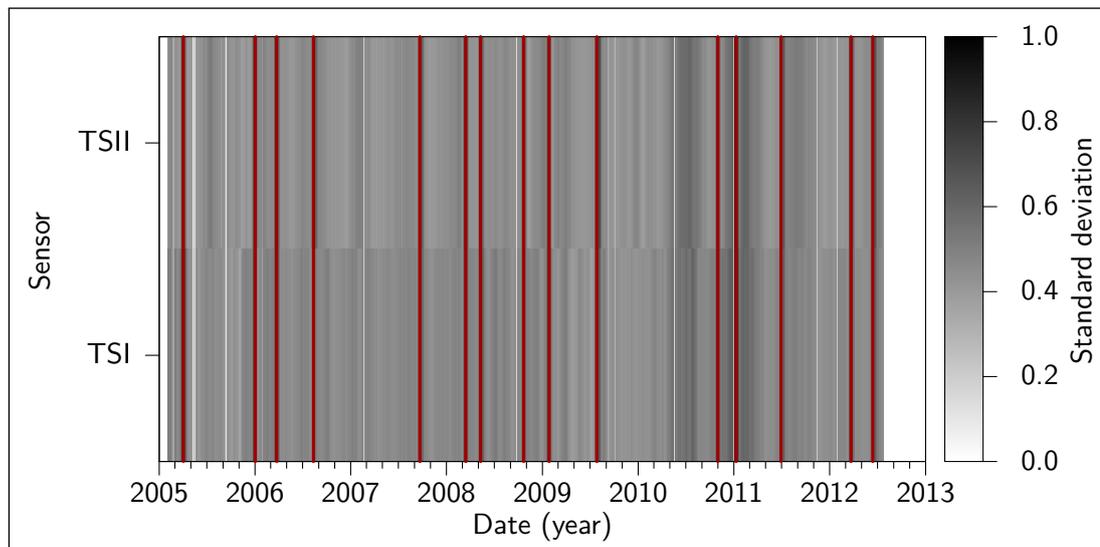


Figure 5.28: Verification data event candidate expectations. The intensity of the local SD is depicted in colour scale, with multiple time series aligned for comparison. Red lines indicate the locations of artificial events in the verification dataset.

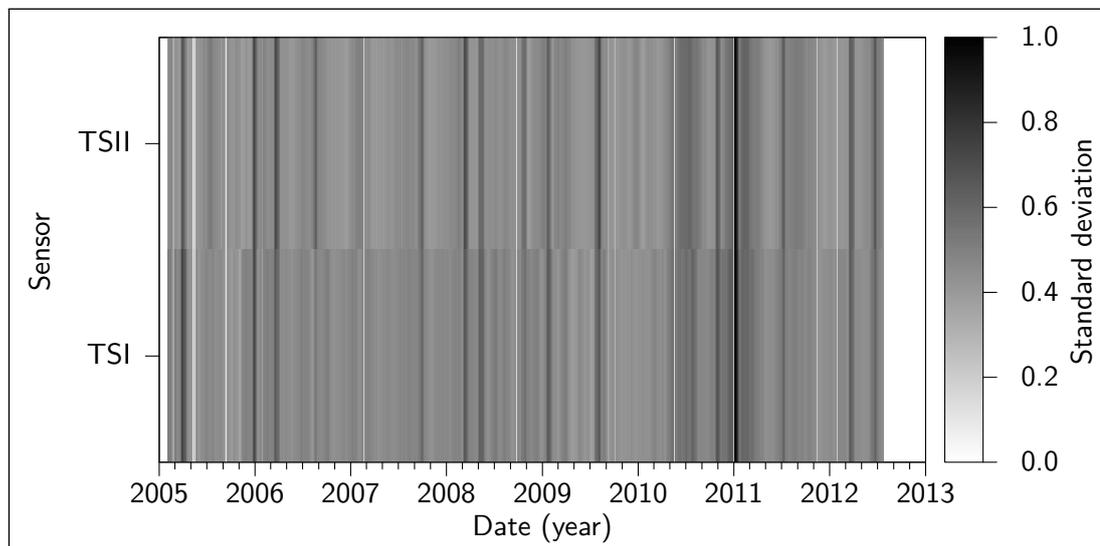


Figure 5.29: Verification data event candidate heatmap. The intensity of the local SD is depicted in colour scale, with multiple time series aligned for comparison. Most expected events can be easily observed across the dataset at global scale.

The event candidate detection process, as demonstrated in Figures 5.26 and 5.27, has successfully highlighted all but one of the events included in TSI and TSII, along with a remaining spike in TSI. This successful highlighting, coupled with the small number of additional event candidates that do not correspond to an artificially created event, suggests that attention would be correctly drawn to areas of potential interest in an experimental dataset, significantly improving the efficiency of investigation. The ability to observe synchronisation of small scale events also provides the ability to ensure that the relevant time series in a dataset are considered in an analysis of such an event, while providing a discriminating factor to justify the discarding of others.

5.3.1.4 Frequency domain inspection

The frequency domain inspection of the de-trended dataset (an NDFT performed on the residuals depicted in Figures 5.24 and 5.25) revealed a strong peak at wavelengths of 365 days and one day, with magnitudes that corresponded closely to the input magnitudes of the frequency content (Figure 5.6, magnitudes 100 and 25 respectively). Figures 5.30 and 5.31 depict respectively the NDFT for the linear based and hyperbolic based artificially created time series. The form of each NDFT is similar to that of the initial NDFT performed on the input frequency content (depicted in Figure 5.7). The additional magnitude in the low frequency (high wavelength) portions of the latter NDFTs performed are likely due to the distortions present in the de-trended time series from 2010 to 2012.

Utilising the phase, amplitude, and wavelength information generated by the NDFT a visual comparison of the frequency content detected to the de-trended residual can be made. The annual frequency content detected by the NDFT process is shown overlain onto the linear based and hyperbolic based artificially created time series in Figures 5.32 and 5.33 respectively. The daily frequency content, while detected by the NDFT process, is not shown here due to its small magnitude.

5.4 Summary

The application of NUDAT to the artificially created dataset shows that both the macro-scale, and smaller scale 'second order' features in time series based data can be detected and

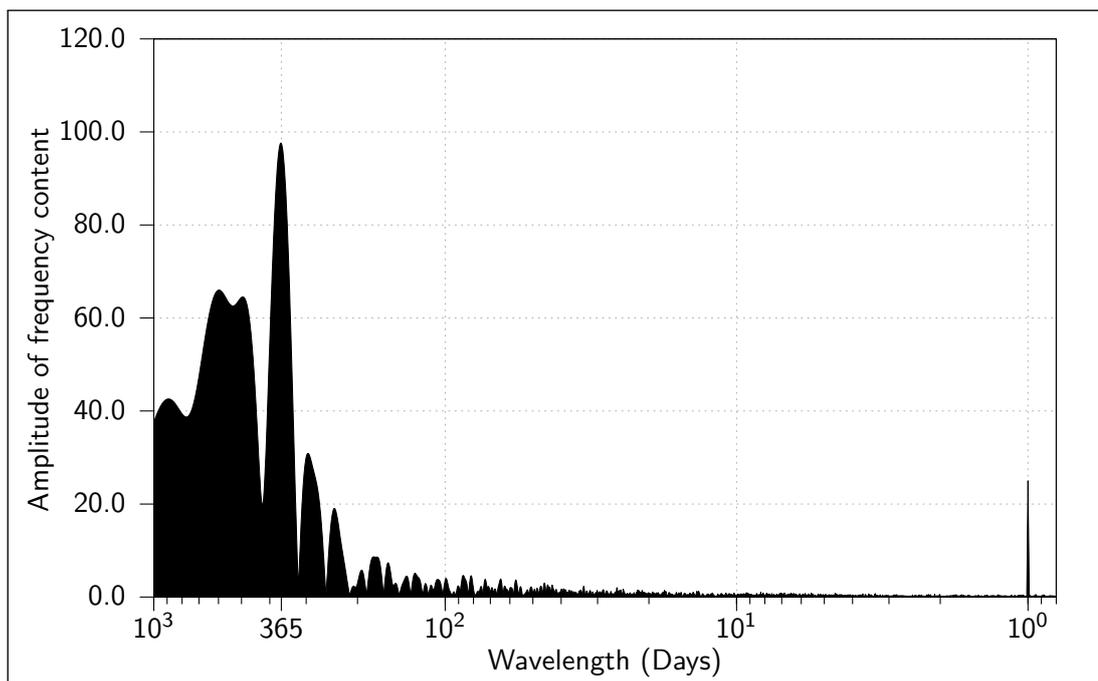


Figure 5.30: NDFT of de-trended TSI time series. Yearly and daily peaks are clearly observable, corresponding to the input frequency content.

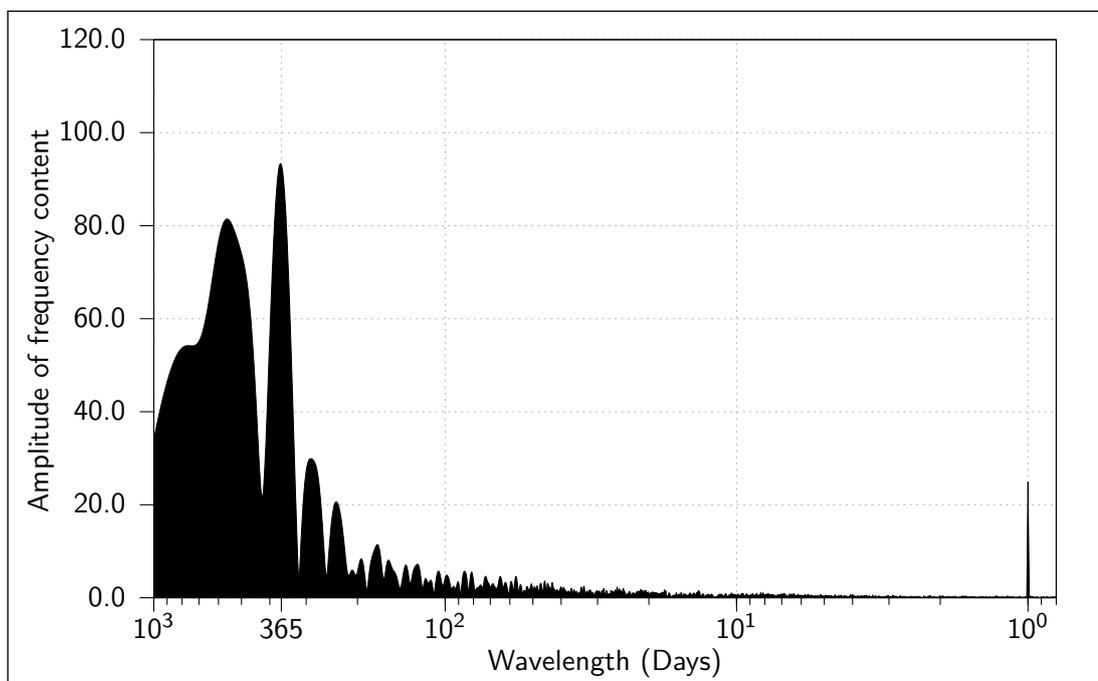


Figure 5.31: NDFT of de-trended TSII time series. Yearly and daily peaks are clearly observable, corresponding to the input frequency content.

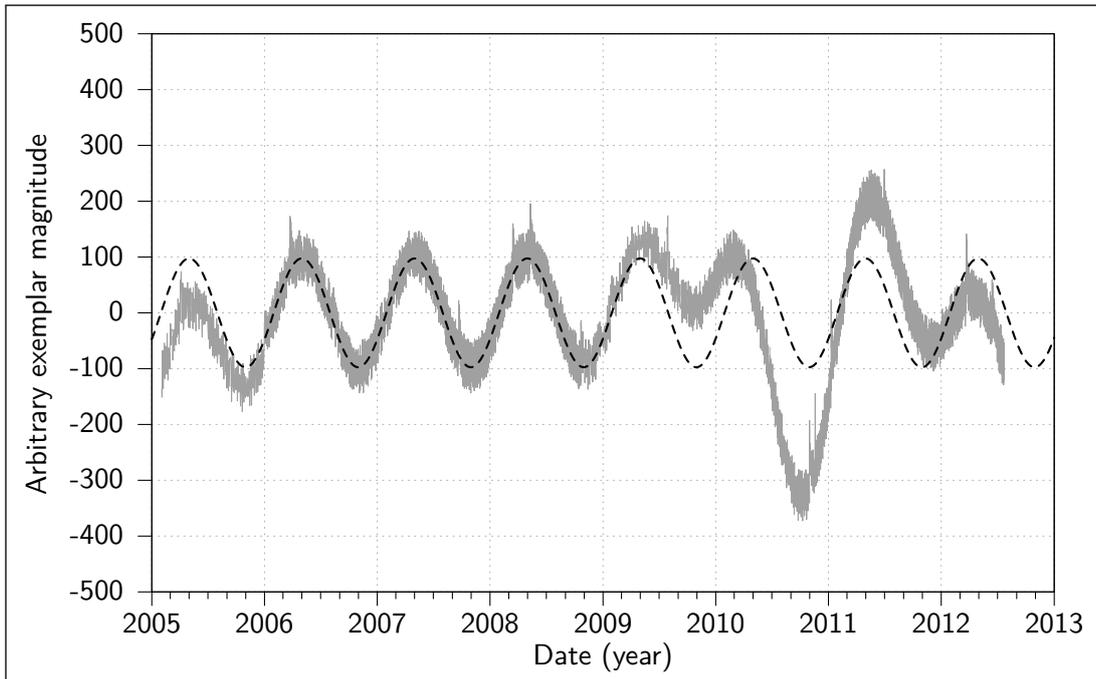


Figure 5.32: De-trended TSI time series with NDFT derived frequency content overlain.

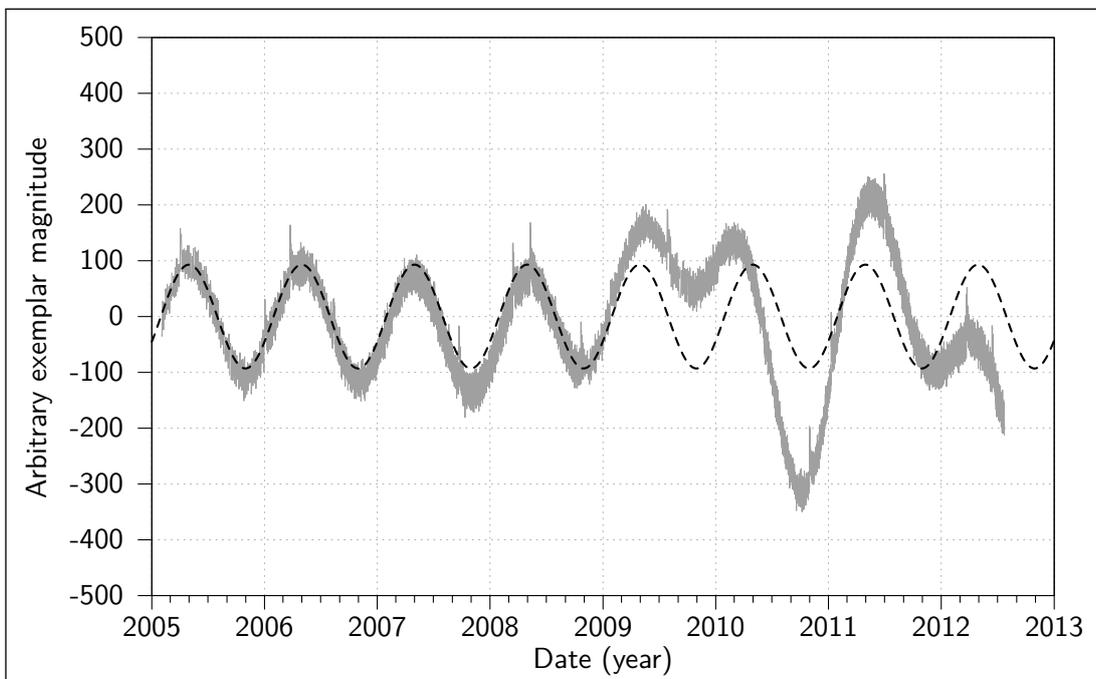


Figure 5.33: De-trended TSII time series with NDFT derived frequency content overlain.

highlighted effectively. The data processes developed in Chapter Four are shown to be able to perform well when applied to control input that approximates the features of experimental data. Non-parametrically derived trend detection, frequency content determination, second order events candidate detection, and spike detection have all been demonstrated to work as intended in NUDAT by comparison of EDA outputs to the known make-up of the input data. Furthermore, the visualisations built around NUDAT outputs facilitate classical, observational analysis by highlighting areas of potential interest and probable correlations. Additionally, the use of non-uniform input data during the successful verification process confirms that NUDAT is capable of accommodating such data.

The ability of NUDAT to accommodate non-uniformly sampled data, coupled with the successful verification of the processes developed in Chapter Four, suggest that NUDAT should facilitate an EDA of Lasgit's dataset. The measured accuracy of the toolkit outputs improves confidence in such an application within the limits of the similarity of the verification data to the experimental data intended for investigation. The subsequent application of NUDAT to Lasgit's dataset and the interpretation of the results are presented in the following Chapters.

Chapter Six

Results of toolkit application

6.1 Introduction

Chapters Six and Seven collectively present the results of the application of the Non-Uniform Data Analysis Toolkit (NUDAT) to the Large Scale Gas Injection Test's (Lasgit's) dataset. Chapter Six presents the results of the application of NUDAT to Lasgit's dataset from a macro-behaviour perspective, while Chapter Seven presents the detailed investigation of three second order events highlighted by the toolkit.

The Exploratory Data Analysis (EDA) enabled by NUDAT, and presented in Chapters Six and Seven, is intended to build on the analysis and reporting already performed by the British Geological Survey (BGS) that has been summarised/referenced in Chapter Three. The phenomena represented in Lasgit's dataset that is exposed and quantified by the application of NUDAT will be the focus of the reporting.

Within this chapter an outline of the necessary pre-processing of Lasgit's dataset, and the toolkit application procedure is presented in Sections 6.2 and 6.3, respectively. Subsequently, an overview of the various toolkit outputs from a macro-behaviour perspective is presented in Section 6.4. The discussion and inferences made from the results presented in Chapters Six and Seven are subsequently presented in Chapter Eight.

6.2 EDA scope and limitations

6.2.1 Data standardisation and truncation

Some minor pre-processing of Lasgit's dataset was required in order to prepare for the EDA. This included i) converting the data from the proprietary Microsoft Excel format to the

non-proprietary text based Comma Separated Values (CSV) format; ii) converting the time stamps in the data files to conform to the ISO 8601 standard date format; and iii) addressing inconsistencies and removing anomalies within the dataset that may hinder analysis. The details of the actions undertaken for items ii and iii are described in Sections 6.2.1.1 and 6.2.1.2.

6.2.1.1 Standardisation of time stamps

During Lasgit's experimental life a number of Daylight Saving Time (DST) shifts occur. Each year, during March and October, a DST shift of one hour occurred in the time stamp record, forwards and backwards respectively. The specific local times at which these DST shifts have taken place indicates that the time zone of the logging computer was altered at some point between 2010-11-01 and 2011-03-26. Before this period the DST shifts occurred 1 AM \rightleftharpoons 2 AM, corresponding to the Greenwich Mean Time (GMT)/British Summer Time (BST) time zone, whereas after this period the DST shifts occurred 2 AM \rightleftharpoons 3 AM, corresponding to the Central European Time (CET)/Central European Summer Time (CEST) time zone. Between these dates there is only one time step larger than the 60 minute difference between the inferred time zones: a gap of approximately 17 hours from 2011-02-15 14:48:52 to 2011-02-16 07:53:50¹. It is during this gap that the change from GMT/BST to CET/CEST is assumed to have occurred.

To prepare the Lasgit dataset for analysis the time stamps were standardised to Coordinated Universal Time (UTC) in order to provide an unambiguous and always increasing time stamp. Additionally, records with time stamps that occurred less than one minute from a previous record were removed for the dataset. Given the typical logging rates identified in Chapter Three, such records were considered likely to have arisen during malfunction of the logging system. This action eliminated 691 records from the dataset.

6.2.1.2 Truncation of dataset

As detailed in Section 3.3, there were a number of operator interventions taken in the initial stages of the experiment involving the control of the pore-water pressure in and around DA3147G01 (see Table 3.9). This intervention has resulted in stress and pore-water pressure

¹When expressed as UTC

time series records that vary rapidly in the early stages of the experiment, juxtaposed with the more slowly varying latter portion of the experiment. This induced rapid variation in stress and pore-water pressure could impact on a computationally performed EDA of the Lasgit dataset in two main ways: i) the combined analysis of a rapidly varying period alongside a more typically and slowly varying period may negatively impact the accuracy of aspects of the analysis such as trend detection; and ii) the physical processes represented in the dataset may not be sufficiently representative of the natural and simulated natural physical processes that are the experimental investigation's main focus, due to the extent of the interventions.

In order to mitigate the influence of the initial stage of the experiment, the EDA performed therefore has only considered a subsection of the available dataset (see Section 3.5). The first two years of data have thus been truncated, leading to a start date for consideration in the EDA of 2007-02-01. For convenience an integer number of years has been considered in the EDA, resulting in a truncation of the final few months of data and an end date of 2012-01-31. Additionally, a number of the ancillary sensors/logged parameters are not considered in the EDA (e.g. logging equipment battery voltages), reducing the number of time series considered to 150. The particular sensors considered are those listed in Tables 3.1 to 3.8.

The truncation process and selection of sensors to consider in the EDA effectively reduces the number of datum points in the dataset from the previously stated figure of circa 27.5 million to a figure of circa 18 million, as each of the 150 time series considered has undergone 120,113 logging cycles between the selected start and end dates of the EDA.

6.3 Toolkit application procedure

The application of NUDAT to Lasgit's truncated dataset (depicted in Figure 6.1) consisted of a sequential application of the following processes:

1. Spike detection (and subsequent removal) utilising a 48 hour window and a spike threshold of three standard deviations (3σ) from the mean.
 - (a) Non-uniform Discrete Fourier Transform (NDFT) of de-spiked data (Section 6.4.2).
 - (b) Event candidate detection of de-spiked data by utilising a 6, 12, 24, and 48 hour windows (Section 6.4.3).

- (c) Smoothing of de-spiked data by utilising a 24 hour window with uniform, triangular, Gaussian, and Laplacian weighting functions (Section 6.4.4).
- 2. Down-sampling of data to daily representative values which are centred on noon of each day.
- 3. Singular Spectrum Analysis (SSA) with a 365 day lagged vector window to detect potential long term trends.
- 4. Removal (subtraction) of the SSA derived trends from full resolution, de-spiked data.
 - (a) NDFT of residual (de-trended) data (Section 6.4.6).
 - (b) Event candidate detection of residual data utilising a 6, 12, 24, and 48 hour windows (Section 6.4.7).
 - (c) Smoothing of residual data by utilising a 24 hour window with uniform, triangular, Gaussian, and Laplacian weighting functions (Section 6.4.8).

6.4 Results of toolkit application

6.4.1 Spike detection

The spike detection process highlighted 129,630 datum points (approximately 0.7% of the considered dataset) as beyond the defined spike threshold. The number of spikes detected in each sensor record is presented in Table 6.1. A significant proportion, approximately 55.8%, of all spikes detected were located in the temperature sensor records. The temperature sensor records account for approximately 37.3% of the considered dataset.

Excluding the records with no data, the sensor record with the greatest number of detected spikes is PT927 (2403 spikes), while the sensor with the least number of detected spikes is UR931 (50 spikes). These sensors are the temperature sensors associated with the axial stress sensor PB927, and the pressure in the packers in PRH-2, respectively. A distribution of the number of spikes detected per sensor record is shown in Figure 6.2. Subfigure 6.2a shows the distribution of detected spikes in all considered sensor records, while Subfigures 6.2b, 6.2c, and 6.2d show the distributions for the stress and pore-water pressure sensors within DA3147G01 (excluding

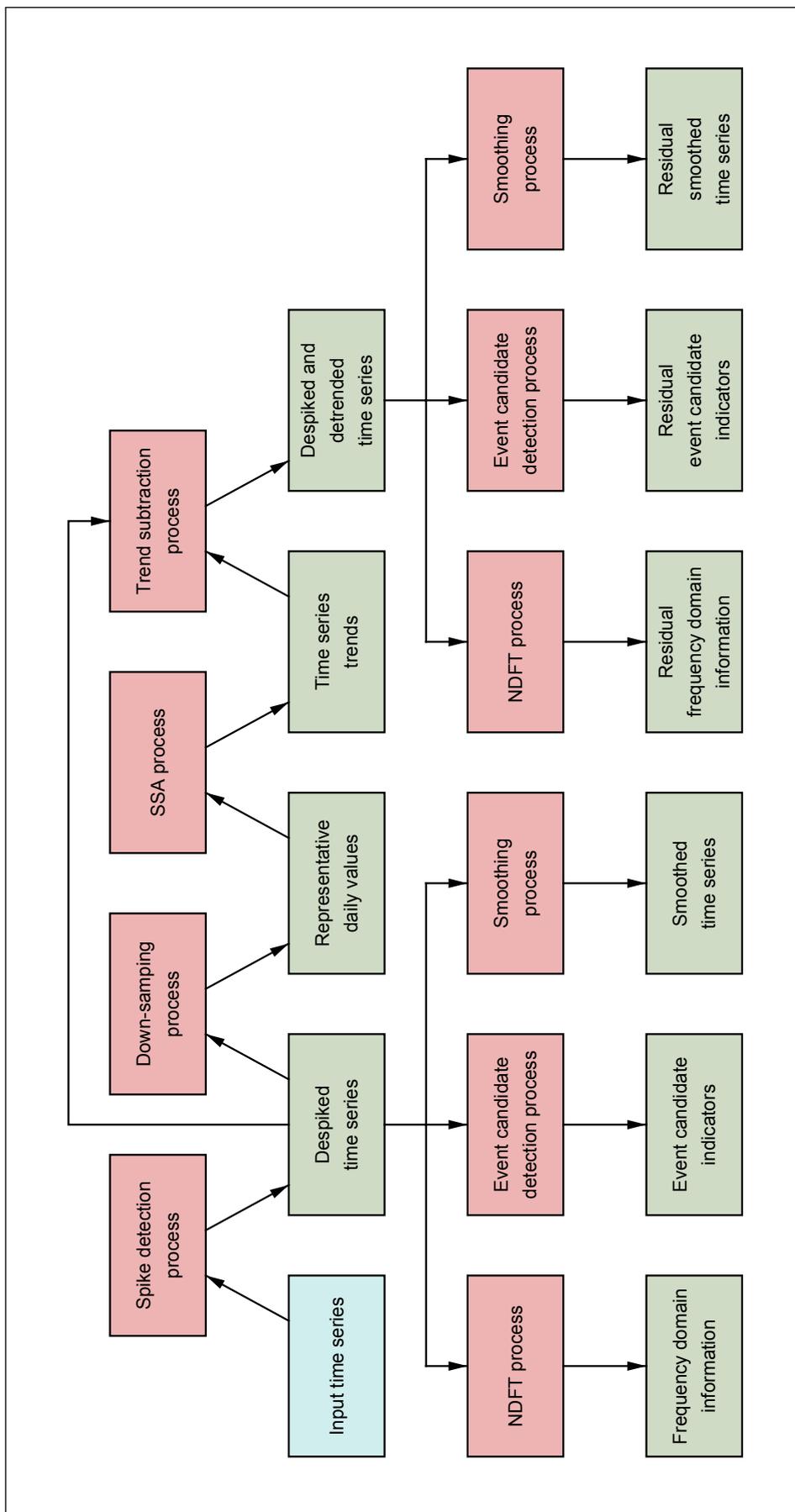


Figure 6.1: Toolkit processes and outputs as applied to Lasgiti's dataset.

Table 6.1: Number of spikes detected in each sensor record considered by the EDA.

| Sensor ID | Spike count | Sensor ID | Spike count | Sensor ID | Spike count |
|-----------|-------------|-----------|-------------|-----------|-------------|
| DP901 | 589 | PR915 | 773 | UR908 | 908 |
| DP902 | 237 | PR916 | 893 | UR909 | 765 |
| DP903 | 244 | PR917 | 915 | UR910 | 938 |
| DP904 | 215 | PR918 | 902 | UR911 | 1299 |
| DP905 | 291 | PR919 | 973 | UR912 | 815 |
| DP906 | 398 | PR920 | 869 | UR913 | 999 |
| DP907 | 55 | PR921 | 1189 | UR914 | 761 |
| FB903 | 220 | PR922 | 1379 | UR915 | 825 |
| FB904 | 311 | PT901 | No Data | UR916 | 857 |
| FL901 | 207 | PT902 | 1099 | UR917 | 739 |
| FL902 | 174 | PT903 | 1051 | UR918 | 778 |
| FL903 | 236 | PT904 | 1167 | UR919 | 857 |
| FL904 | 209 | PT905 | 1091 | UR920 | 683 |
| FM905 | 275 | PT906 | 1075 | UR921 | 936 |
| FM906 | 270 | PT907 | 1442 | UR922 | 810 |
| FM907 | 272 | PT908 | 1141 | UR923 | 106 |
| FM908 | 282 | PT909 | 1218 | UR924 | 105 |
| FR901 | 317 | PT910 | 1102 | UR925 | 122 |
| FR902 | 280 | PT911 | 1247 | UR926 | 131 |
| FU909 | 230 | PT912 | 1208 | UR927 | 120 |
| FU910 | 216 | PT913 | 1238 | UR928 | 118 |
| FU911 | 214 | PT914 | 1047 | UR929 | 115 |
| FU912 | 241 | PT915 | 1166 | UR930 | 126 |
| LP901 | 150 | PT916 | 1154 | UR931 | 50 |
| LP902 | 137 | PT917 | 1280 | UT901 | 923 |
| LP903 | 107 | PT918 | 1322 | UT902 | 1126 |
| PB901 | 439 | PT919 | 1499 | UT903 | 1053 |
| PB902 | 680 | PT920 | 1267 | UT904 | 1064 |
| PB923 | 1069 | PT921 | 1252 | UT905 | 863 |
| PB924 | 856 | PT922 | 1855 | UT906 | 1125 |
| PB925 | 906 | PT923 | 1731 | UT907 | 1175 |
| PB926 | 934 | PT924 | 1531 | UT908 | 880 |
| PB927 | 1190 | PT925 | 1366 | UT909 | 1268 |
| PB928 | 1589 | PT926 | 1512 | UT910 | 1196 |
| PB929 | 1464 | PT927 | 2403 | UT911 | 1180 |
| PC901 | 73 | PT928 | 2173 | UT912 | 1254 |
| PC902 | 138 | PT929 | 2104 | UT913 | 1231 |
| PC903 | 121 | TA905 | 342 | UT914 | 1445 |
| PR903 | 669 | TC901 | 596 | UT915 | 1265 |
| PR904 | 937 | UB901 | 739 | UT916 | 1127 |
| PR905 | 642 | UB902 | 505 | UT917 | 1416 |
| PR906 | 652 | UB923 | 1153 | UT918 | 1330 |
| PR907 | 973 | UB924 | 1016 | UT919 | 1252 |
| PR908 | 658 | UB925 | 989 | UT920 | 1477 |
| PR909 | 843 | UB926 | 801 | UT921 | 1386 |
| PR910 | 732 | UR903 | 958 | UT922 | 1482 |
| PR911 | 739 | UR904 | 923 | UT923 | 1932 |
| PR912 | 907 | UR905 | 863 | UT924 | No Data |
| PR913 | 1111 | UR906 | 736 | UT925 | 2266 |
| PR914 | 808 | UR907 | 960 | UT926 | 2229 |

PCs and filters/filter mats), their incorporated temperature sensors, and the remaining sensors, respectively. Inspection of the spike occurrence distributions for these sensor groupings reveals a clustering of spike occurrence within each grouping, and a notable difference in the locus of spike clustering between groupings.

A selection of records from the highest spiking classification above, the temperature sensor records, are shown in Figure 6.3. The datum points that have met the threshold for spike classification are highlighted in black. In general, a lesser density of spikes can be observed during the periods of peak temperature in the temperature cycles.

All subsequent toolkit outputs reported here exclude the datum points highlighted as spikes from their input, as described in Section 6.3 and depicted in Figure 6.1.

6.4.2 NDFTs of de-spiked data

An NDFT analysis of each of the sensor records was taken after the de-spiking process was completed. In addition to the zero frequency component, the NDFT process inspected a dense set of frequencies corresponding to wavelengths in the range 0.1 days to 10,000 days ($\approx 1.16 \times 10^{-4}$ Hz to $\approx 1.16 \times 10^{-9}$ Hz). The zero frequency component derived for each sensor record was subtracted from the corresponding input data for each subsequent inspection of the non-zero frequencies. It should be noted that while the inspection of the frequency domain at wavelength greater than the length of record analysed has been performed, the results obtained at wavelengths longer than circa 2,500 days ($\approx 4.63 \times 10^{-9}$ Hz) cannot be considered as quantitatively valid. As such they are considered in this work only as a qualitative indicator of the presence of longer term behaviours/trends in the data.

The frequency content of the non-temperature sensor records (e.g. stress, pore-water pressure) predominantly express increasing amplitudes at lower frequencies, an example of which is shown in Figure 6.4. Non-temperature sensor records with distinct amplitudes at annual wavelengths include PB928, PB929, UR908, UR912, UR914, UR915, UR920, UR921, and UR922. PB928 and UR914 are distinct in comparison with others as they possess relatively lower amplitudes at larger wavelengths other than the peak at annual. The sensor record arising from UR931 possesses a small but prominent daily frequency component of approximately 0.1 kPa.

The temperature sensor records predominantly exhibit prominent annual frequency content.

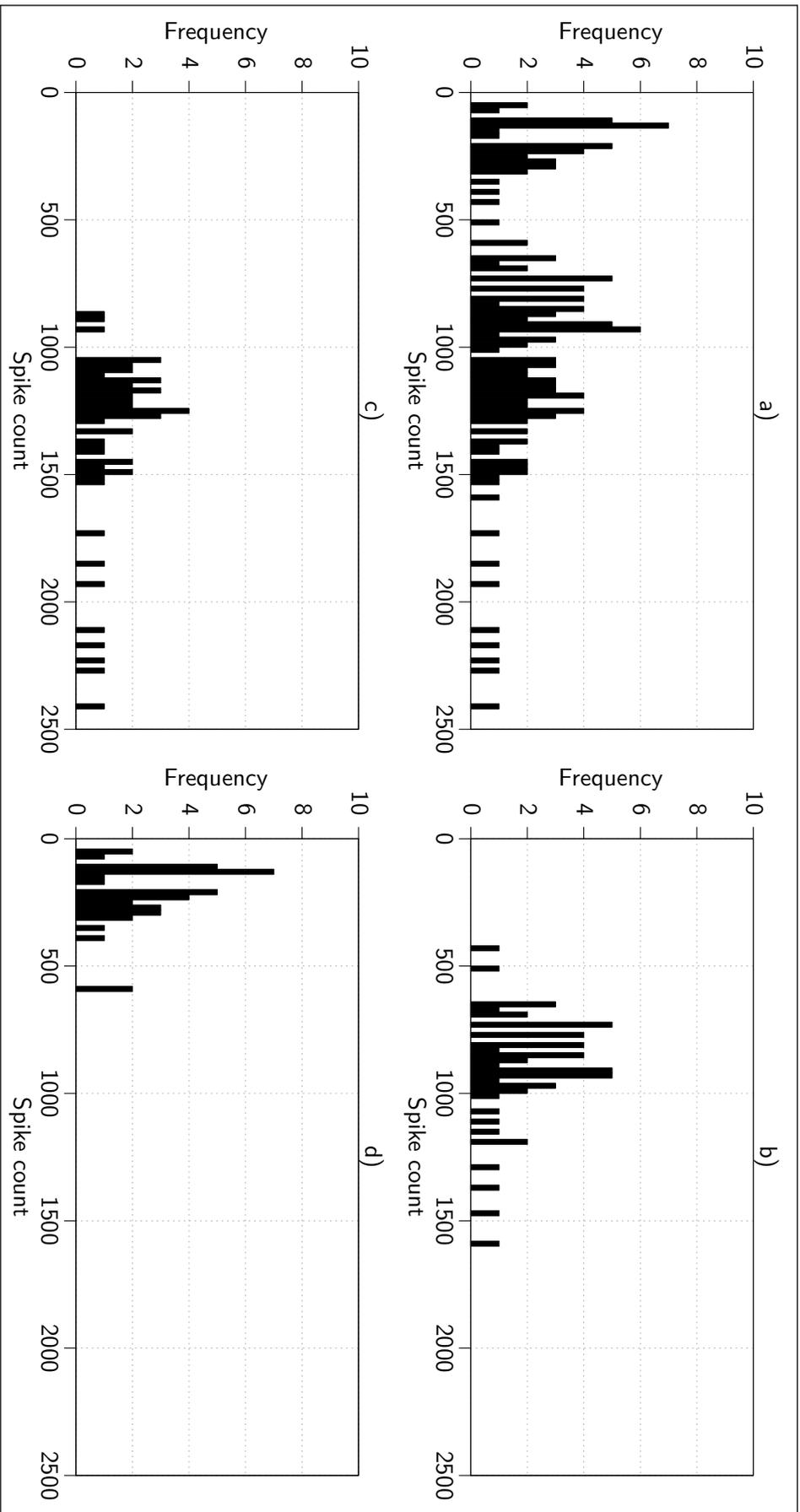


Figure 6.2: Distribution of spikes detected per sensor record. a) All considered sensors. b) Stress and pore-water pressure sensors in DA3147G01, excluding PCs (i.e. PBs, PRs, UBs, and URs not in pressure relief holes). c) Temperature sensors associated with those sensors identified as present in Subfigure b) (PTs and UTs). d) All remaining sensors (PCs, TC, TA, DPs, LPs, FBs, FMs, FLs, FUs, FRs, and URs in pressure relief holes).

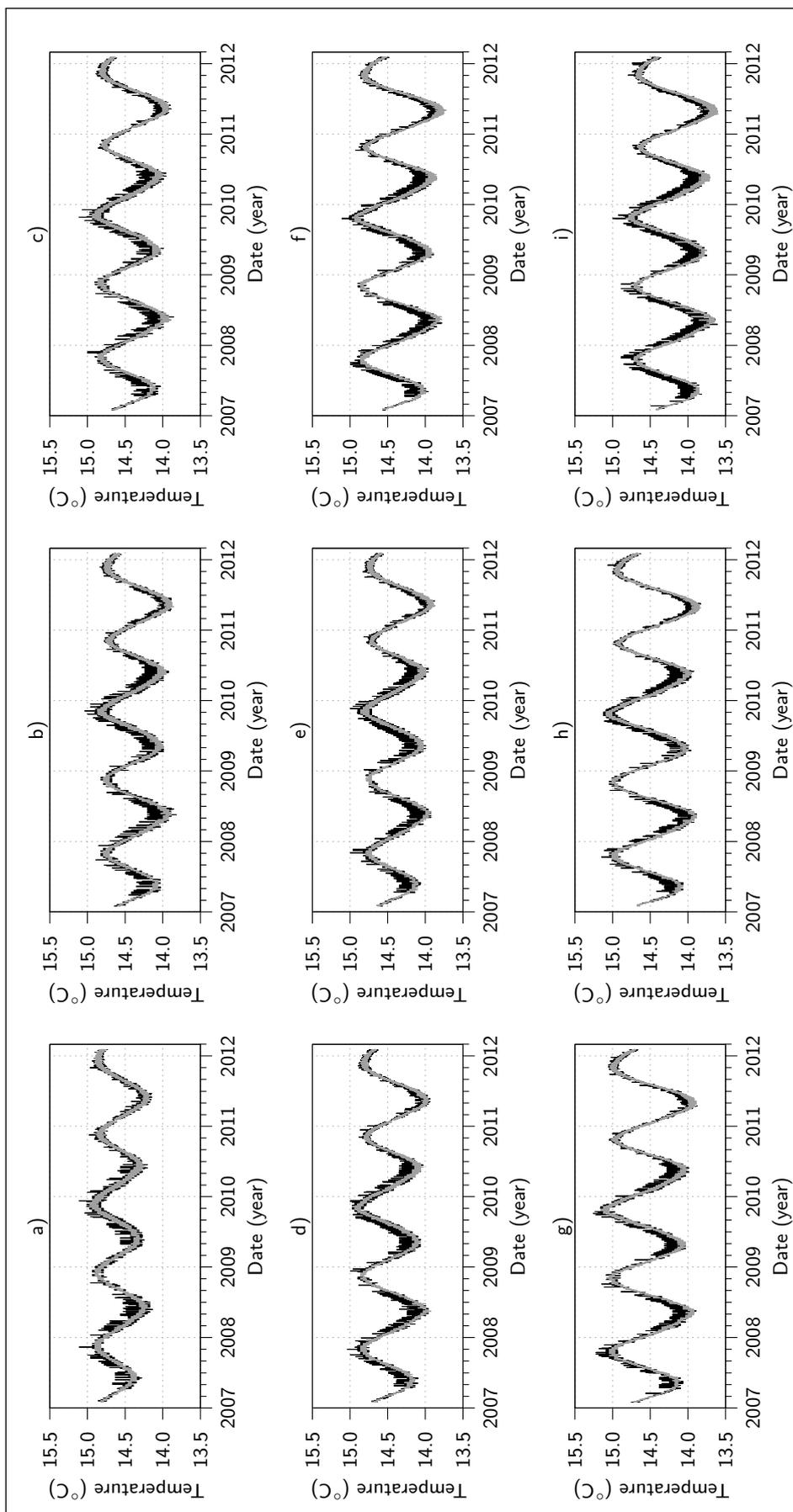


Figure 6.3: Selection of temperature sensor records after spike detection. Spikes are highlighted in black with de-spiked signal overlain in grey.

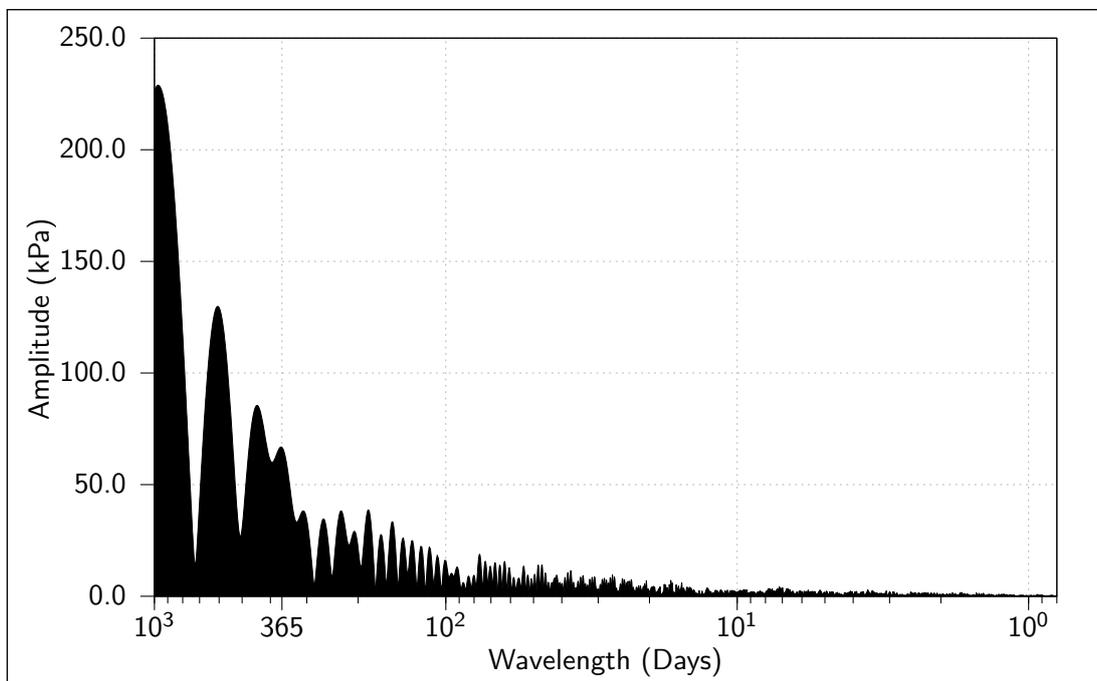


Figure 6.4: Example of low frequency ramping in PR908.

The amplitudes of the annual frequency content present in the temperature sensor records, arising from sensors located within DA3147G01, range from approximately 0.1 °C to approximately 1.3 °C. Exceptions include sensor records arising from UT911, PT901, and UT924. UT911 possesses a large departure from the expected range of reported temperatures from late 2009 to early 2011, while PT901, and UT924 have both failed (Section 3.4).

Specific note is made of the Hard Rock Laboratory (HRL) air temperature record (TA905, Figure 6.5), which is the only temperature record to possess a prominent (in comparison to surrounding noise levels) daily frequency component in its NDFT. Figure 6.6 depicts the NDFT for TA905's sensor record, with the annual and daily frequency peaks highlighted. These peaks correspond to an annual temperature cycle of approximately 1.95 °C and a daily temperature cycle of 0.06 °C.

6.4.3 Event candidate detection of de-spiked data

An overview of the results of the event candidate detection process is presented in this section as a series of 'intensity' plots in Figures 6.7 to 6.19. Individual second order events are not discernible at the scale that the event candidate information is visualised in these Figures. However, a qualitative impression of the extent to which the groups of sensor records have synchronous perturbations can be obtained.

The event candidate information is grouped and visualised by sensor type. The groups visualised in Figures 6.7 to 6.19 are as follows:

- Axial stress sensors.
- Pore-water pressure sensors within the bentonite buffer.
- Pore-water pressure transducers within the injection filters and hydration mats.
- Lid vertical and lateral displacement sensors.
- Pore-water and packer pressures in the pressure relief holes.
- Radial stress sensors on the canister/bentonite interface and rock wall/bentonite interface.
- Pore-water pressure sensors at the rock wall/bentonite interface.

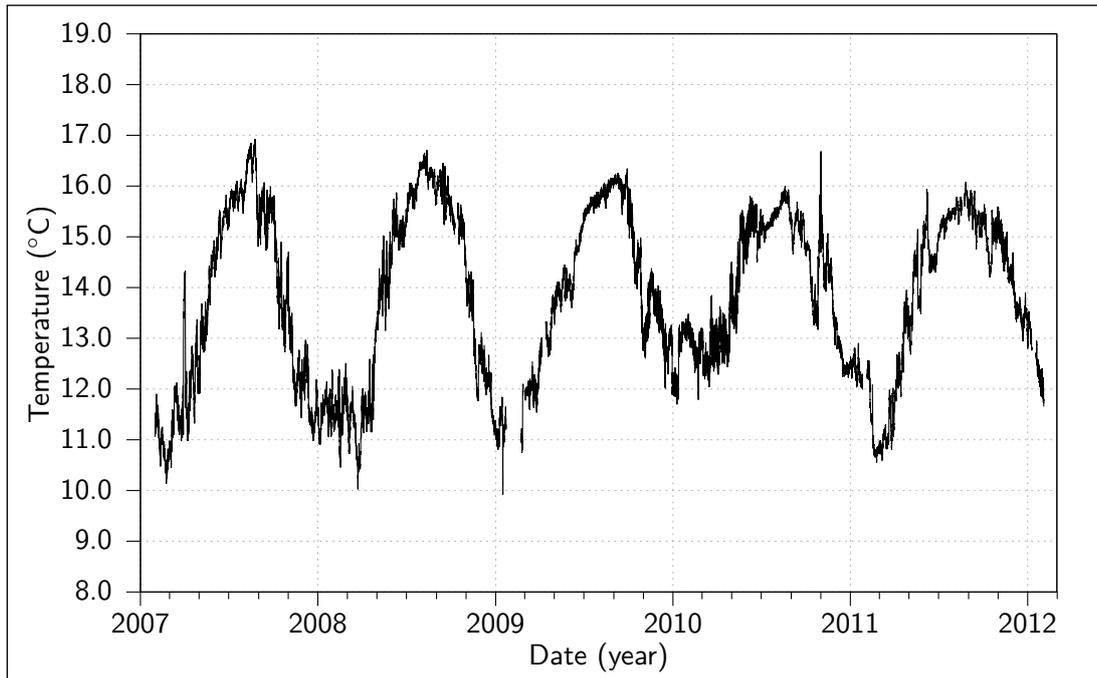


Figure 6.5: Äspö Hard Rock Laboratory air temperature in TBM assembly hall.

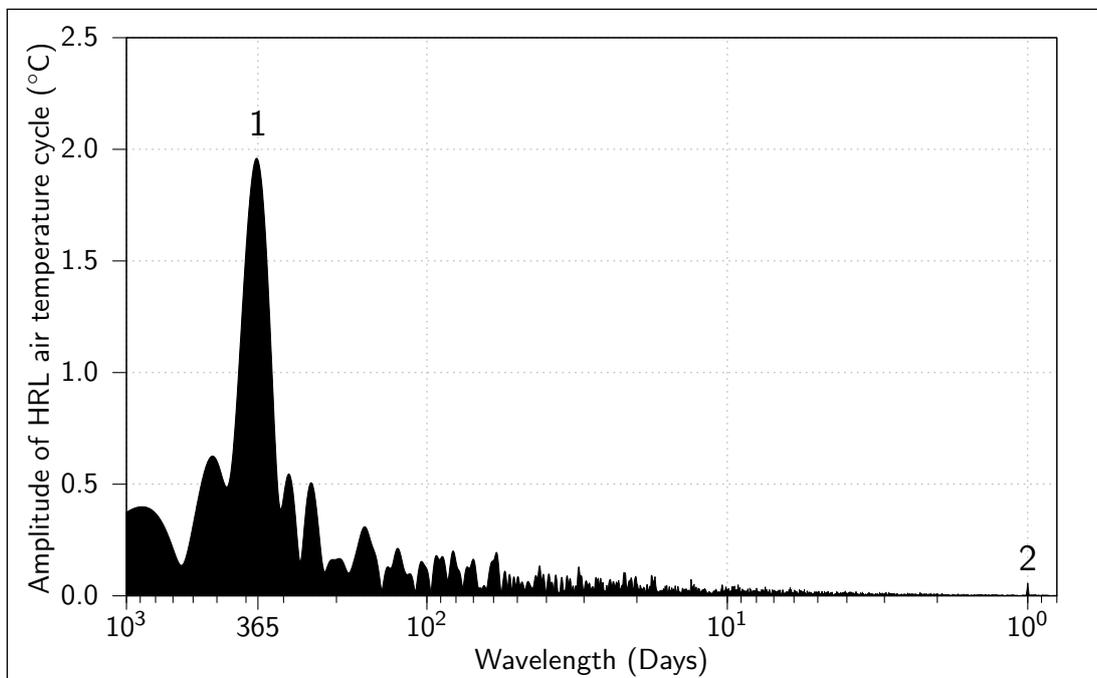


Figure 6.6: NDFT of the Äspö Hard Rock Laboratory air temperature in TBM assembly hall.
1) Annual frequency peak, 2) daily frequency peak.

- Temperature sensors (incorporated into the down-hole stress and pore-water pressure sensors, as well as the internal canister temperature sensor, and HRL air temperature sensor).

All the intensity plots presented in Section 6.4.3 are formed using a 48 hour detection window. In the overview visualisations each sensor record is scaled to range between 0.0 and 1.0 (normalised). Sensors IDs for the sensors located within DA3147G01 (or one of the pressure relief holes) are listed on the vertical axis sorted by depth from the lid (or top of pressure relief hole where appropriate), and then alphabetically/numerically when located at the same depth. The sensors not located within DA3147G01 or one of the pressure relief holes are listed alphabetically/numerically. Vertical lineation across multiple sensor records can be observed in most of the intensity plots, with the exceptions being the in-bentonite pore-water pressure sensors.

6.4.3.1 Axial stress sensors

Figure 6.7 aggregates the event candidate results for the axial stress sensors in DA3147G01. PB923, PB924, PB925, PB926, PB927, PB928, and PB929 are located above the canister, while PB901, PB902, and PC901 are located below it. Vertical lineation across sensor records is observable in both the subset of sensor above the canister as well as the subset of sensors below it. Examples include an above canister lineation during February of 2011, and below canister lineations at the start of February 2009, during December 2009, during August 2010, and during March 2011. Additionally, a single vertical lineation is also observable during October 2008 across all the axial stress sensors except PB902. This latter lineation appears to have the greatest intensity.

6.4.3.2 Pore-water pressure sensors within the bentonite buffer

Figure 6.8 presents the event candidate results for the pore-water pressure sensors emplaced within the bentonite buffer within DA3147G01. UB923, UB924, UB925, and UB926 are located above the canister, while UB901 and UB902 are located below. UB902's sensor record is notable within the visualisation due to the apparent lack of intensity, except for a short time during December 2009. The aberrant intensity apparent in UB902 is due to a large change in

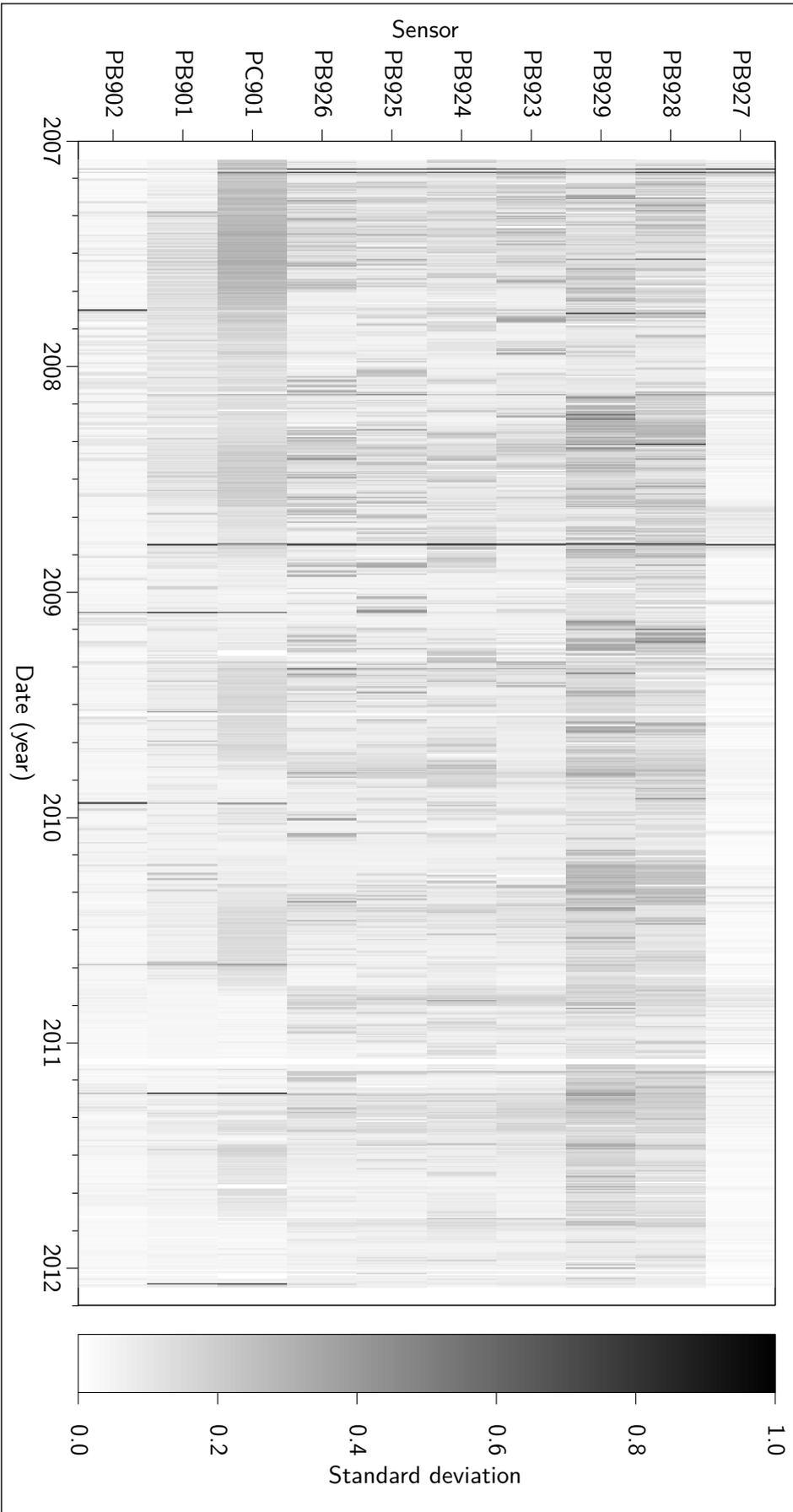


Figure 6.7: Intensity plot of the axial stress sensors' SDs from 2007-02-01 to 2012-02-01.

sensor magnitude resulting from a macro scale gas flow intercepting the sensor (as summarised in Section 3.3).

6.4.3.3 Injection filters and hydration mats

The aggregation of the results of the event candidate detection process for the injection filters and hydration mats (i.e. the pressure transducers) are shown in Figure 6.9. The sensor records depicted in Figure 6.9 are characterised by an apparent lack of intensity with a number of short term higher intensity lines, a number of which spanning all or most of the sensor records. The sensor with the least synchronisation of intensity lines is FM906. Additionally, FR901 possesses only two major point of synchronisation at this scale.

6.4.3.4 Lid displacements and retaining loads

The event candidate results associated with the displacement of the deposition hole lid (vertically and laterally) and the load sensors on the lid restraining anchors are aggregated in Figure 6.10. As detailed in Tables 3.7 and 3.6, LP901, LP902, and LP903 are anchor cable loads, DP901, DP902, DP903, DP904, and DP905 are various vertical displacements, and DP906 and DP907 are lateral displacements. The three categories of sensor are visually discernible in Figure 6.10 by their typical intensity profiles. The vertical displacement sensors are vertically lineated in wide (approximately six month) bands of marginally higher and lower intensities. The higher intensity regions appear within the central portions of each year (during the summer months). The lateral displacement sensors have an apparent lack of intensity with lineated lines of high intensity (as seen in UB902 and the filter/hydration mat sensor records). The anchor load sensor records show a marginally more uniform intensity profile.

6.4.3.5 Pore-water pressures in the pressure relief holes

The event candidate results arising from the pore-water pressures sensor records in Pressure Relief Holes (PRHs) PRH-1 and PRH-2 are shown in Figure 6.11. As highlighted in Table 3.5, UR923, UR924, UR926, UR926, and UR927 are located in PRH-1, while UR928, UR929, UR930, and UR931 are located in PRH-2.

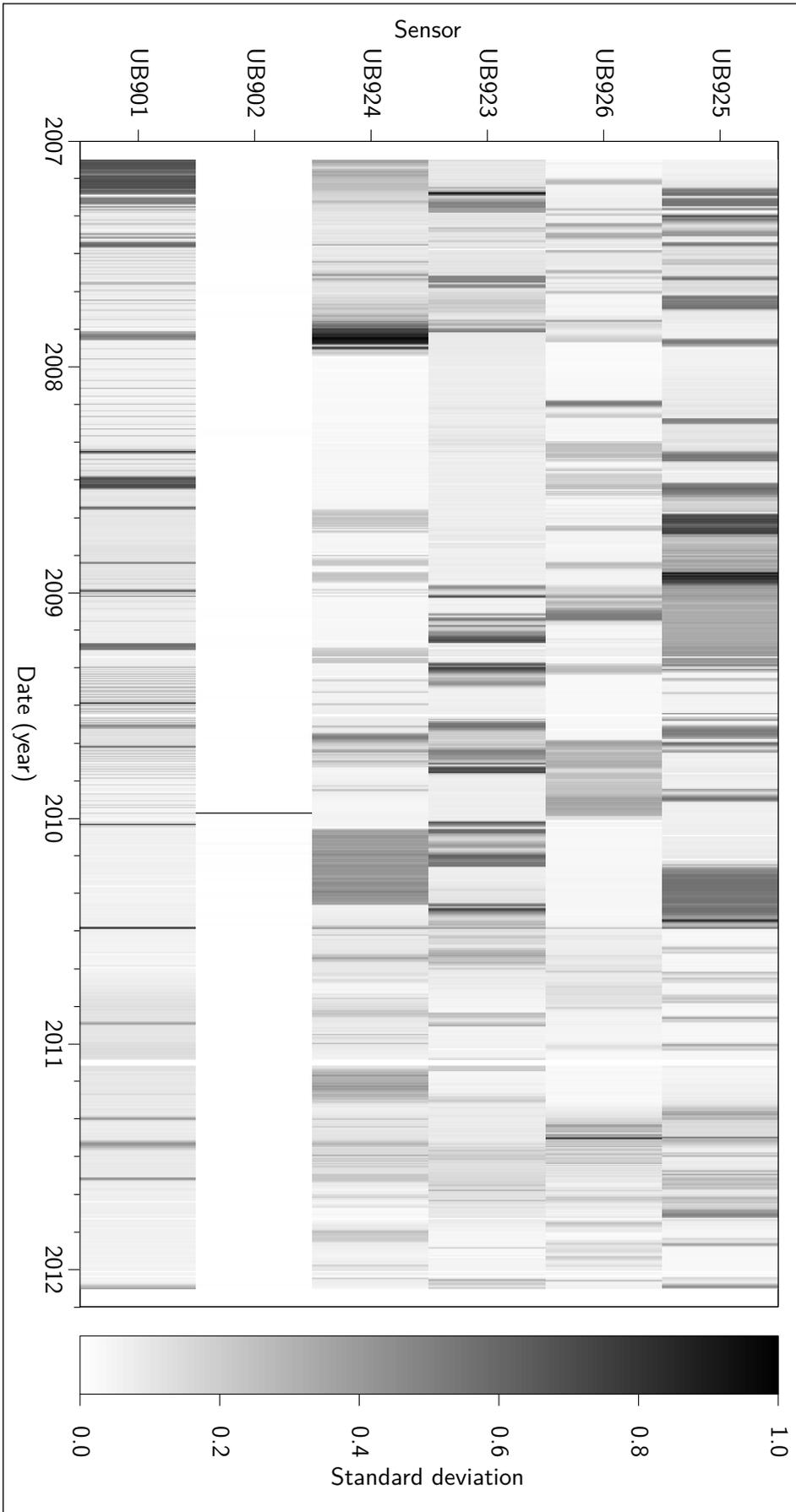


Figure 6.8: Intensity plot of the bentonite pore-water pressure sensors' SDs from 2007-02-01 to 2012-02-01.

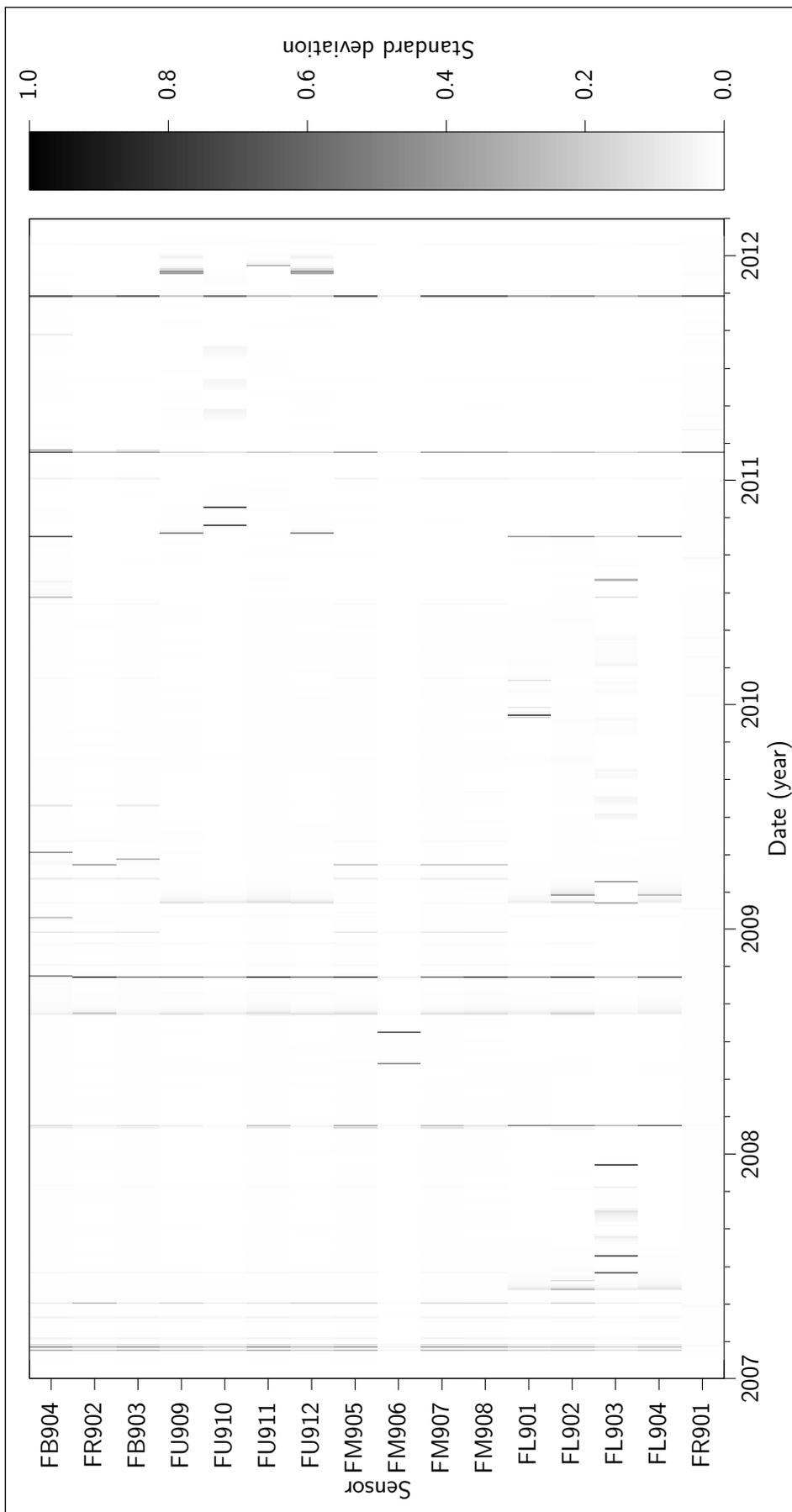


Figure 6.9: Intensity plot of the filter pore-water pressure SDs from 2007-02-01 to 2012-02-01.

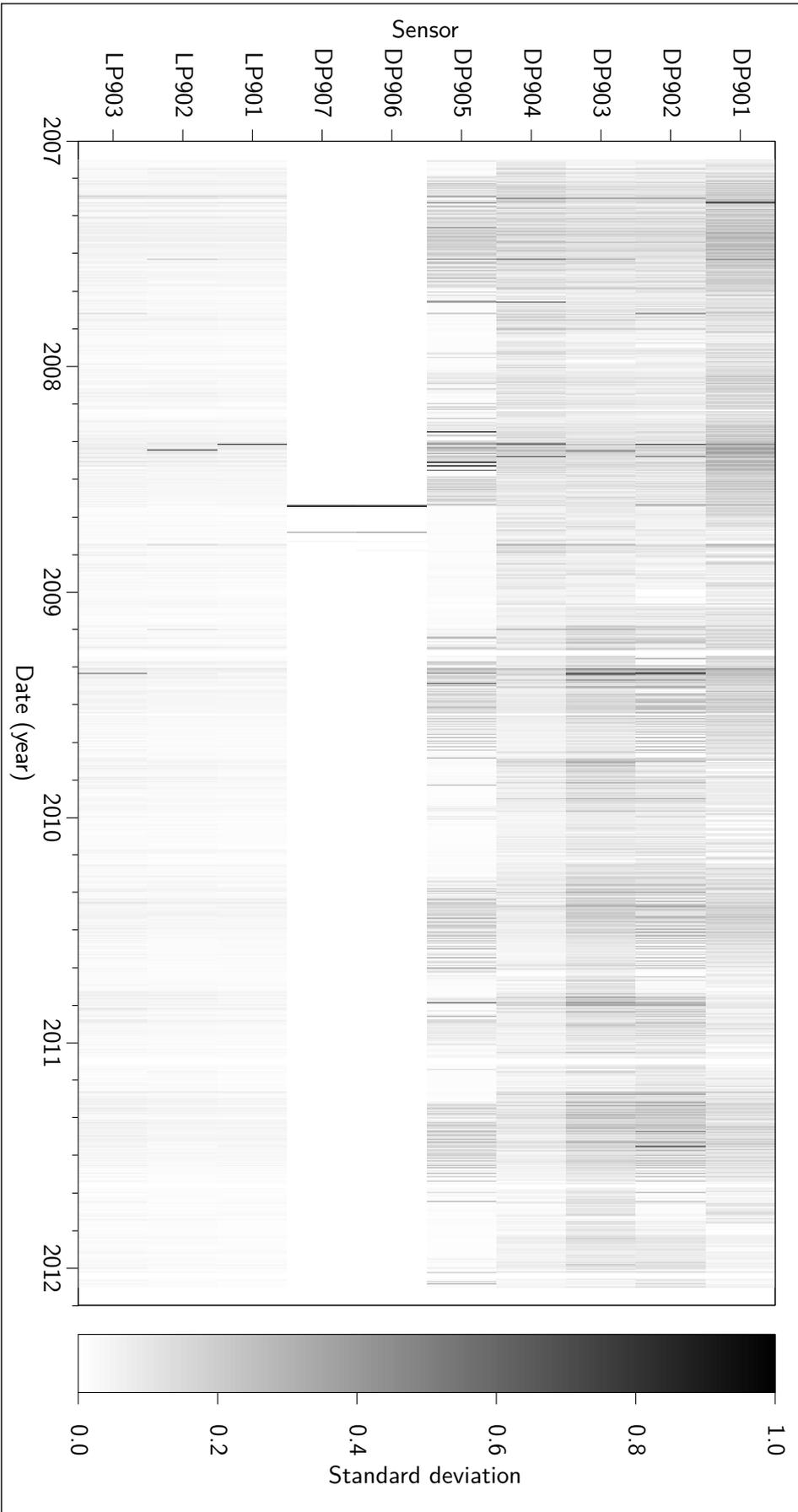


Figure 6.10: Intensity plot of the lid displacement and lid retaining load sensors' SDs from 2007-02-01 to 2012-02-01.

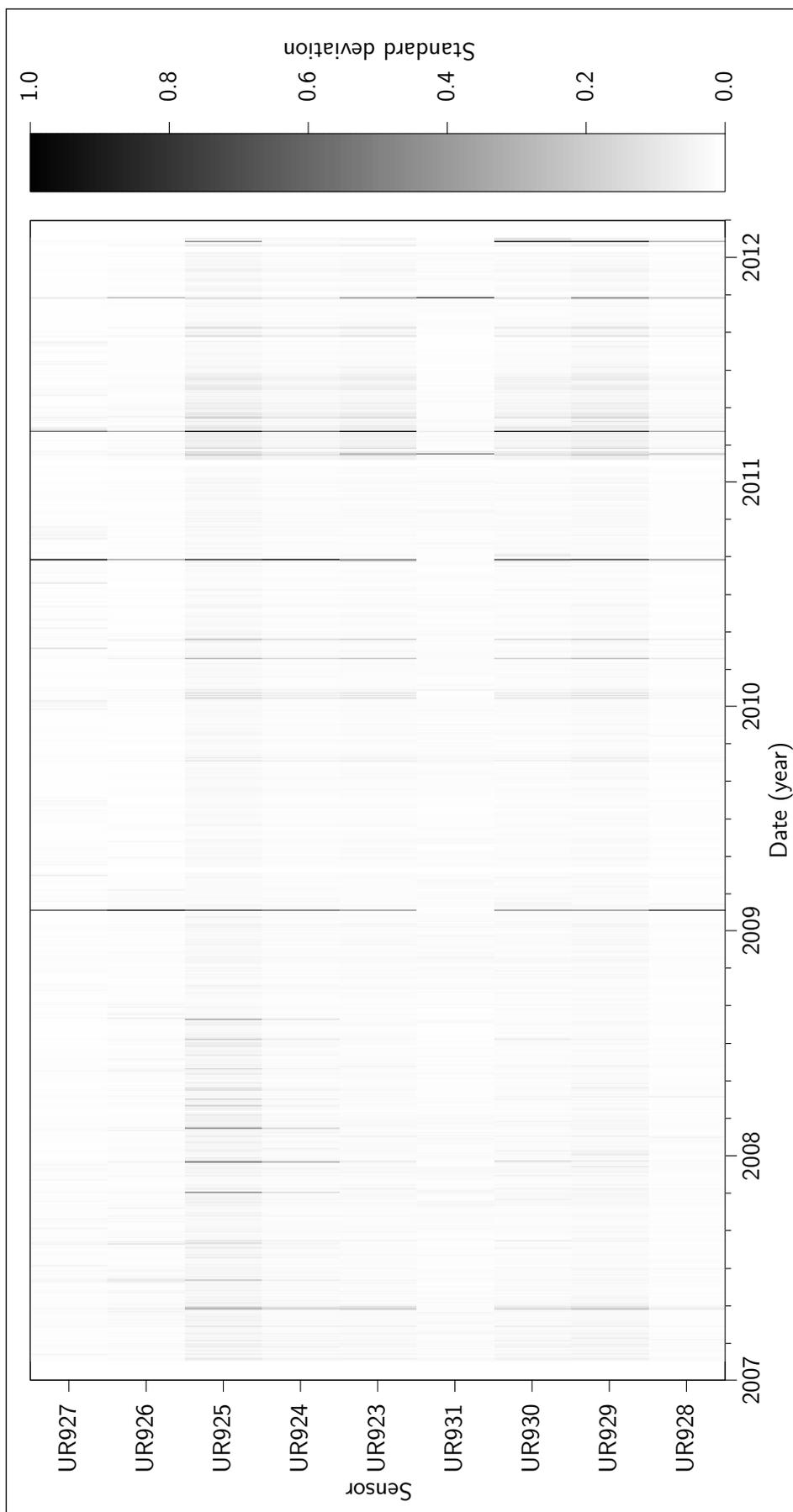


Figure 6.11: Intensity plot of the pore-water pressures in pressure relief holes' SDs from 2007-02-01 to 2012-02-01.

Lineation is apparent across most of the sensors in the set. An apparent lack of intensity for most of the record durations with lines of higher intensity at February 2009, August 2010, and March 2011 are observable. UR931 appears to lack the bands of high intensity that are synchronised with, and hence form, the lineation across the other sensors. Inspection of the sensor record for UR931, presented in Figure 6.12, shows that the sensor is reporting a slightly negative pore-water pressure with only minor exception.

6.4.3.6 Radial stress sensors

The event candidate results for the radial stress sensors on both the rock wall/bentonite interface (PRs) and the canister/bentonite interface (PCs) are aggregated in Figure 6.13. Lineation is observable across a number of combinations of sensor records at various points in the experimental history, with some lines of high intensity spanning most of the sensor records.

PR915 and PC903 are noteworthy for their apparent lack of intensity across much of their respective sensor records, each both possessing a short period of high intensity near the beginning of 2012. The records of these two sensors are depicted in Figures 6.14 and 6.15, where the steps that cause the high intensity lines seen in Figure 6.13 are highlighted.

6.4.3.7 Pore-water pressures at rock wall/bentonite interface

Figure 6.16 presents the event candidate results for the pore-water pressure sensor records at the rock wall/bentonite interface. Vertical lineation is present synchronously across most sensors in a number of locations, most notably at August 2010 and March 2011. UR907 and UR910 (both at the same depth within DA3147G01) are notable for their apparent lack of intensity and sparse lines of high intensity.

Of the two highlighted sensor records (UR907 and UR910), UR907 is most aberrant as it does not conform to the synchronisation of the vertical lineation of the group, while UR910 has a line of high intensity in synchronisation with the other sensors in the set. The lack of apparent intensity for most of the sensor record and the lines of high intensity are caused by 'steps' in UR907 and UR910, depicted in Figures 6.17 and 6.18.

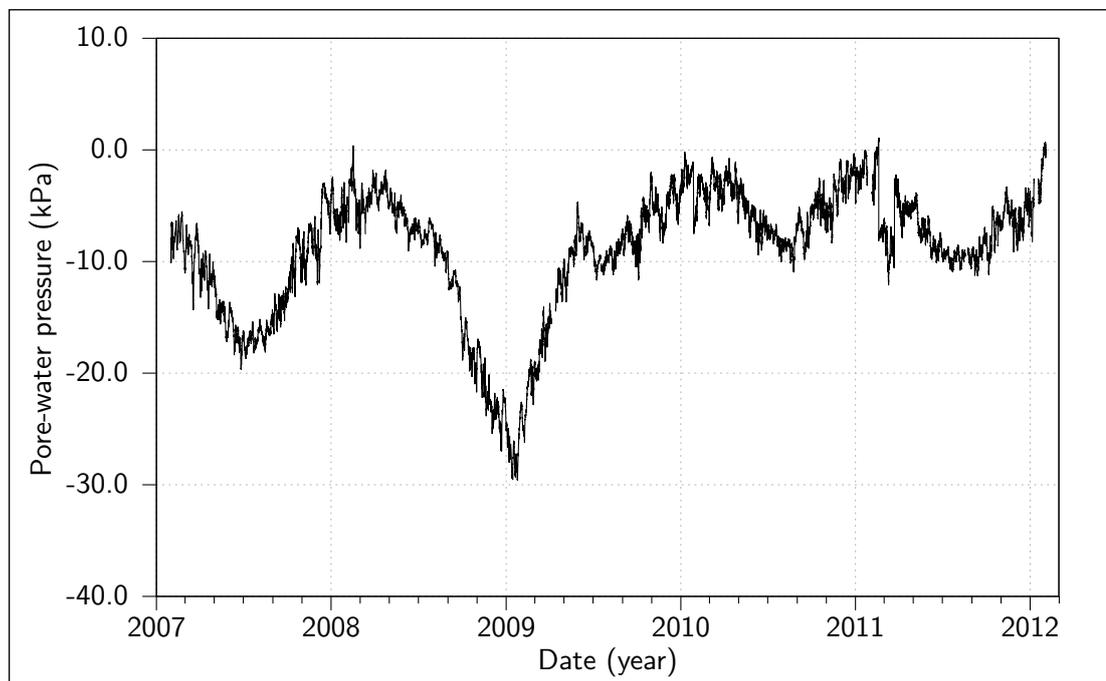


Figure 6.12: UR931 pore-water pressure sensor record. Pore-water pressures are reported to be slightly negative for the majority of the record duration.

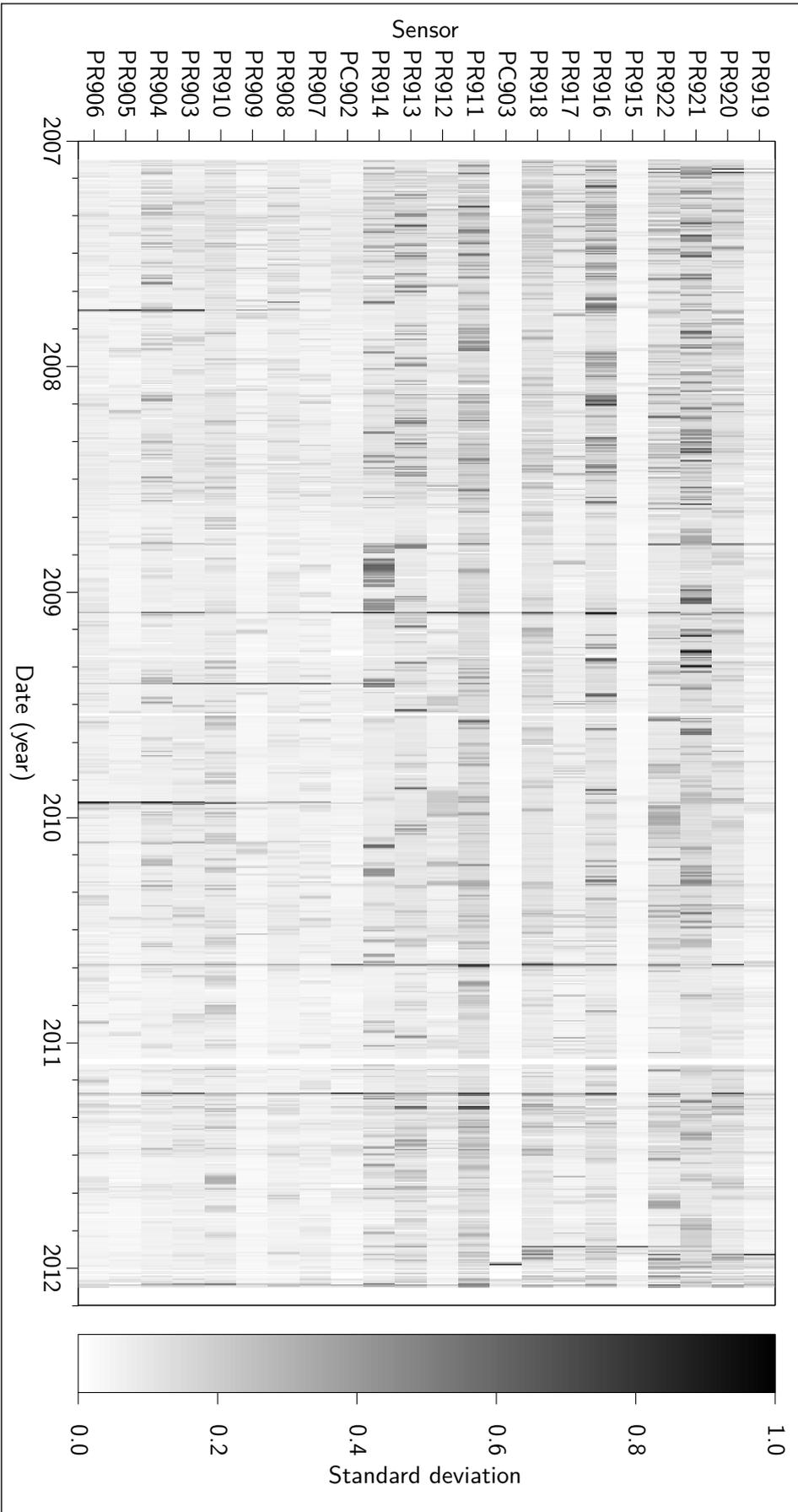


Figure 6.13: Intensity plot of the radial stress sensors' SDs from 2007-02-01 to 2012-02-01.

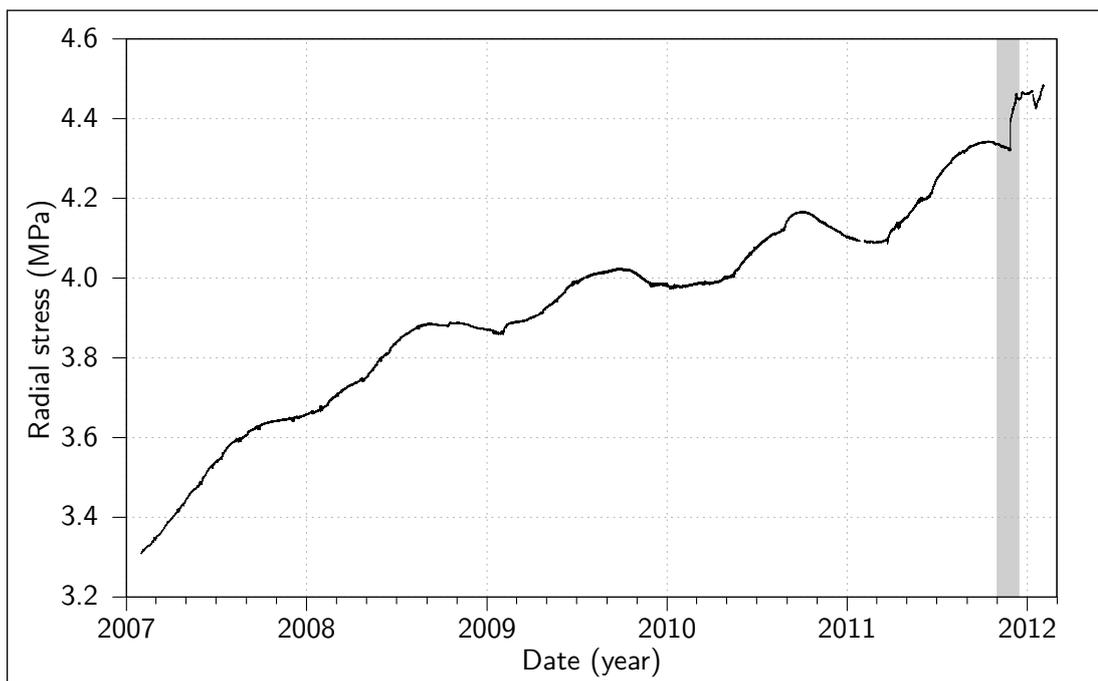


Figure 6.14: Steps (highlighted in grey) in radial stress sensors PR915, causing the high intensity lines depicted in Figure 6.13.

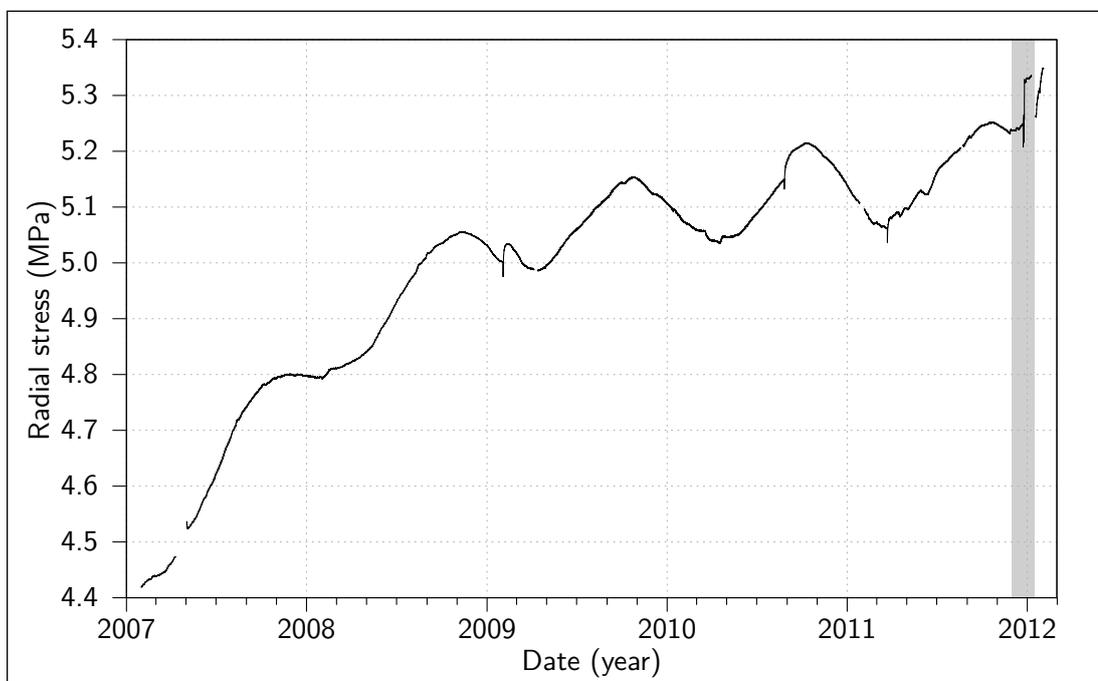


Figure 6.15: Steps (highlighted in grey) in radial stress sensors PC903, causing the high intensity lines depicted in Figure 6.13.

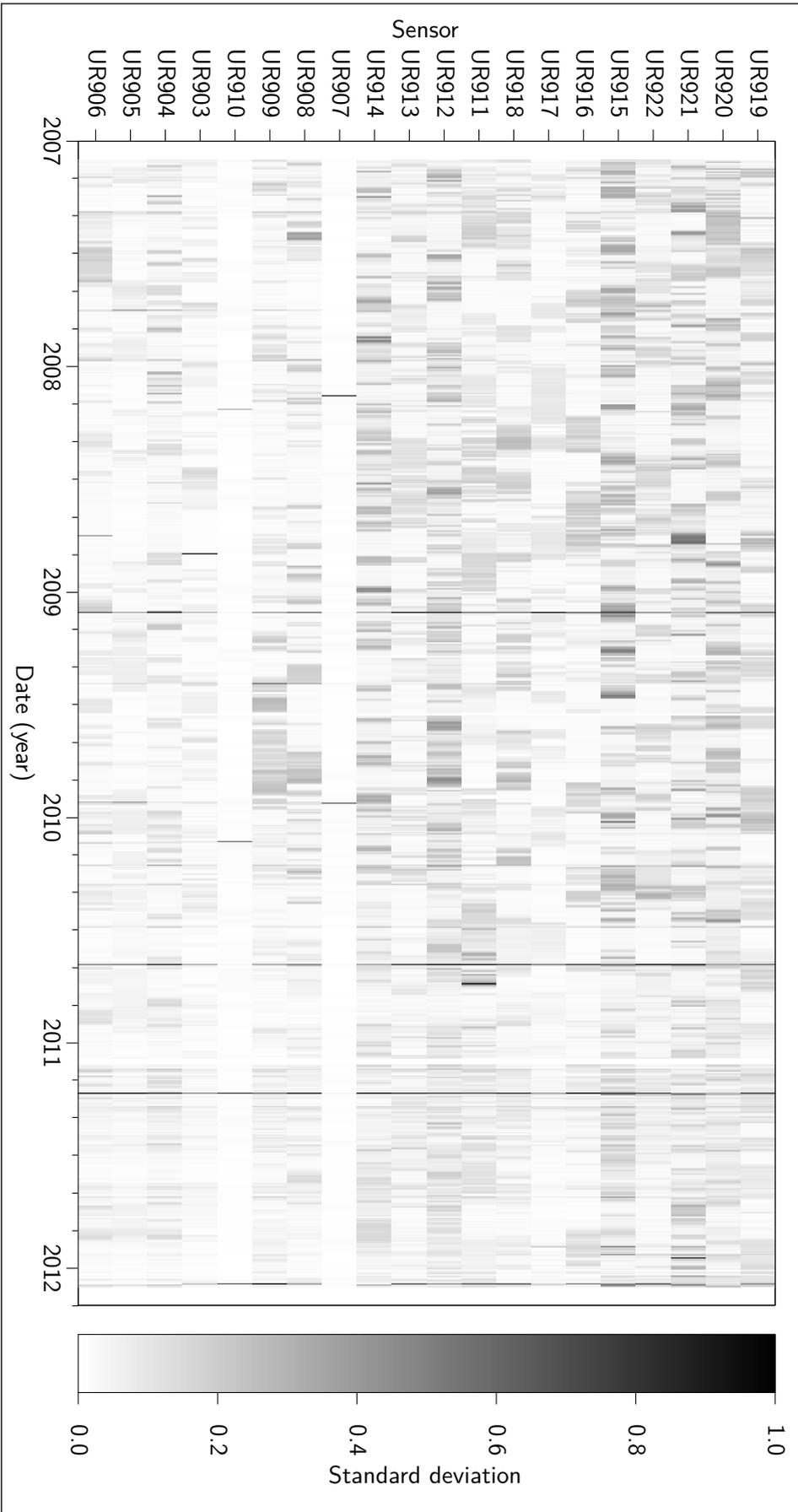


Figure 6.16: Intensity plot of the pore-water pressure sensors' SDs located at the rock wall/bentonite interface from 2007-02-01 to 2012-02-01.

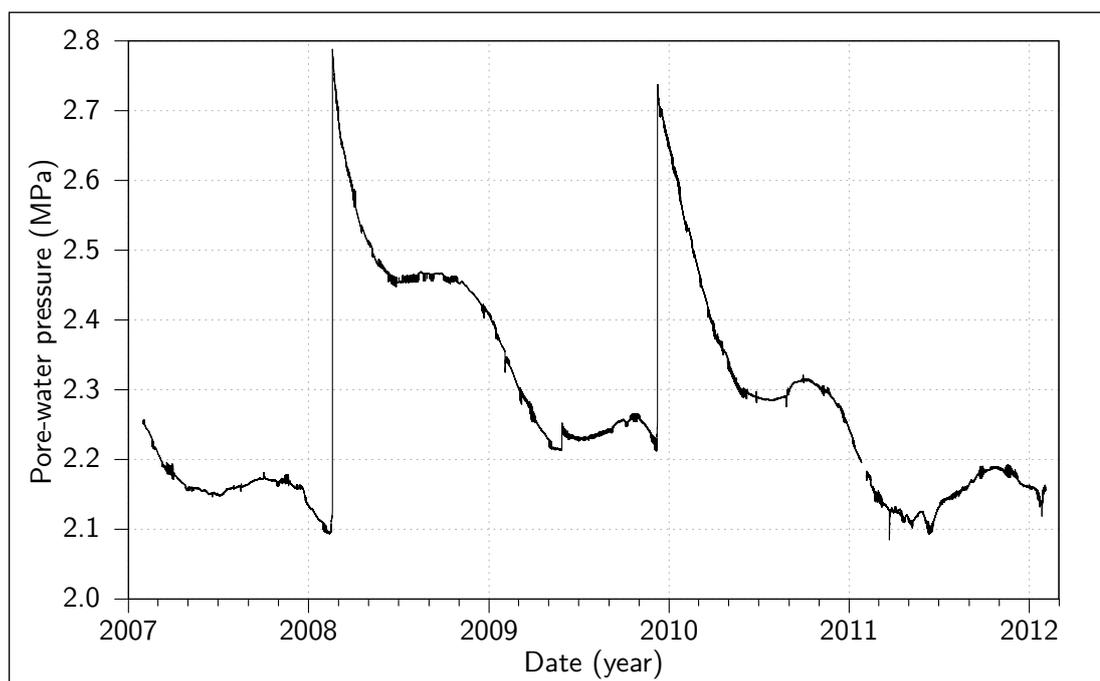


Figure 6.17: 'Steps' in pore-water pressure sensors UR907, causing the high intensity lines depicted in Figure 6.16.

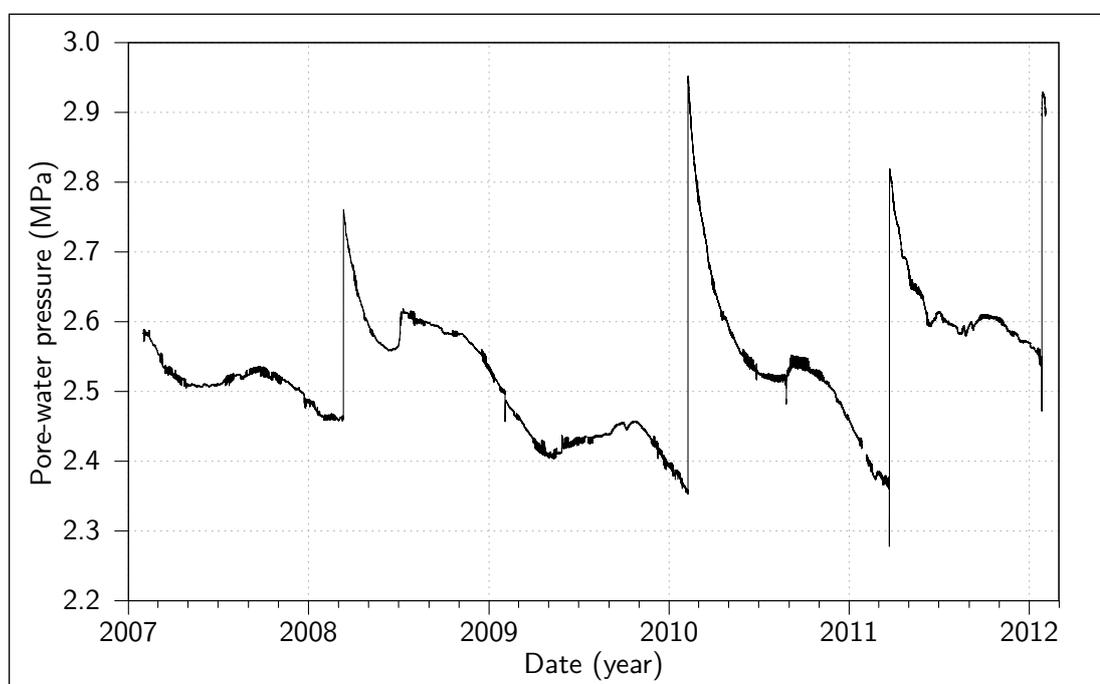


Figure 6.18: 'Steps' in pore-water pressure sensors UR910, causing the high intensity lines depicted in Figure 6.16.

6.4.3.8 Temperature sensors

The event candidate results for the temperature sensors associated with the stress and pore-water pressure sensors within DA3147G01 are shown in Figure 6.19. Vertical lineation is strongly present across the figure, with clear delineation between the sensors above the canister and surrounding or below the canister (PT927 through PT926 and PT919 through UT901 on Figure 6.19 respectively). TA905 and TC901 are grouped separately as they represent the HRL air temperature and the internal temperature of the canister at an indeterminate depth.

Further inspection of Figure 6.19 reveals that four sensors are notable for their disparity, i.e. UT924, UT911, PT908, and PT901. Each of these sensors exhibits an apparent lack of intensity of Standard Deviation (SD) for all or most of the time series. In the cases of PT901 and UT924 this is due to an absence of records as a result of sensor malfunction. UT911 however expresses an apparent sensor malfunction (Figure 6.20), resulting in an section of the time series presenting a SD many times larger than elsewhere. This causes the majority of the sensor record to appear low valued when viewed on a relative (normalised) scale.

PT908 (Figure 6.21) possesses a short term downwards perturbation from approximately 2007–09–18 to 2007–09–20 (Figure 6.22). This perturbation, while smaller in magnitude and length than that in UT911, causes the SD to appear low valued elsewhere in the sensor's row in Figure 6.19.

6.4.4 Smoothing process of de-spiked data

In general, the smoothing processes that were applied to each time series within the dataset resulted in a reduction of noise in the resulting outputs. In all but one case the reduction in noise did not reveal any noteworthy or remarkable phenomena. The noteworthy result revealed by the application of smoothing is a recurring anomaly in UT901's temperature record.

The revealed anomaly occurs approximately 1.5 months after each annual minimum in the considered timespan. It consists of a small (in the order of 0.05°C) and abrupt change in the reported temperature, and is followed by a similar downwards shift approximately three months later. From the 2010–2011 annual minimum onwards the step change is visible in the raw data, but is completely subsumed by the variation in the sensor record earlier than this point.

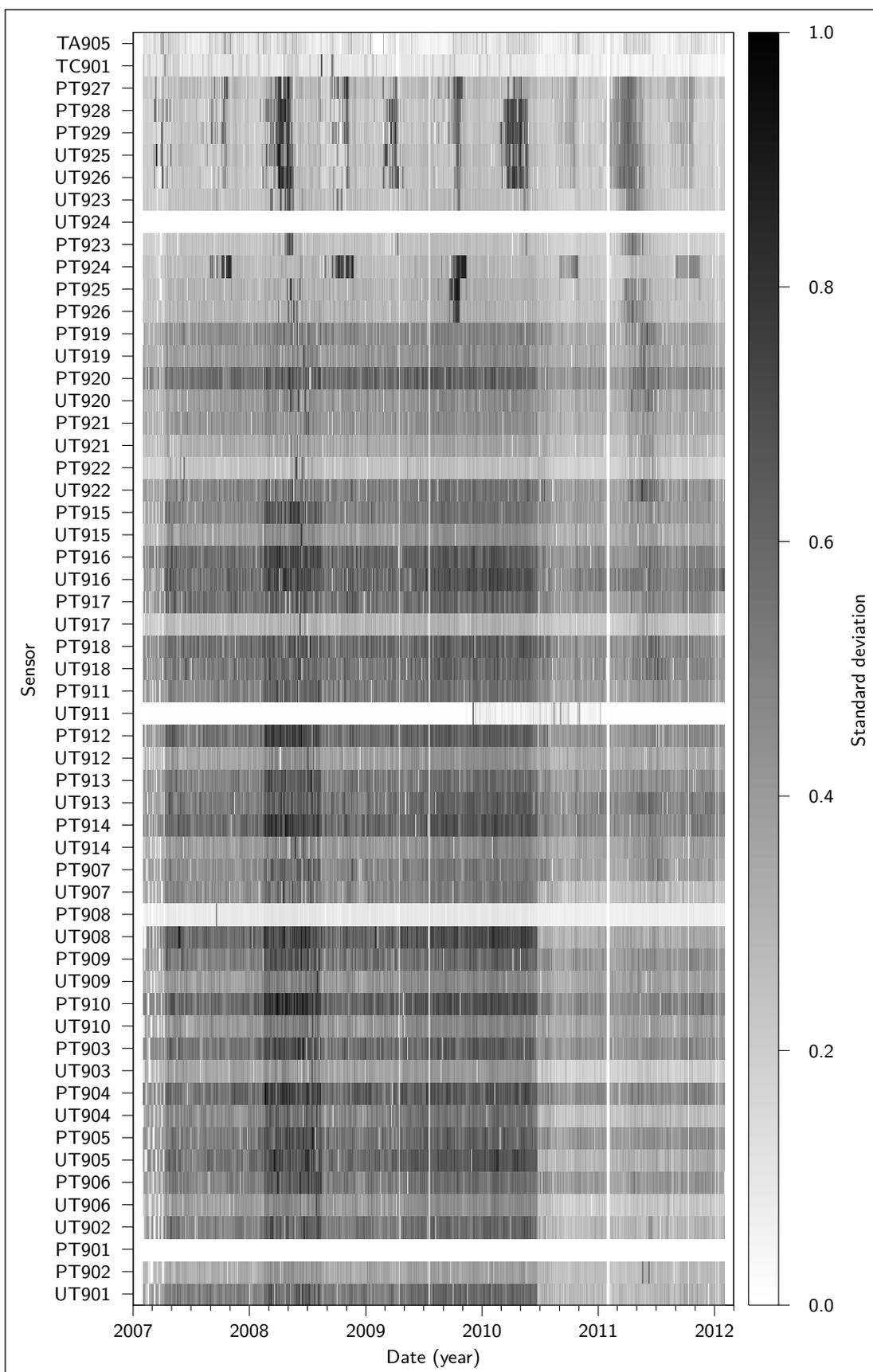


Figure 6.19: Intensity plot of the temperature sensors' SDs from 2007-02-01 to 2012-02-01.

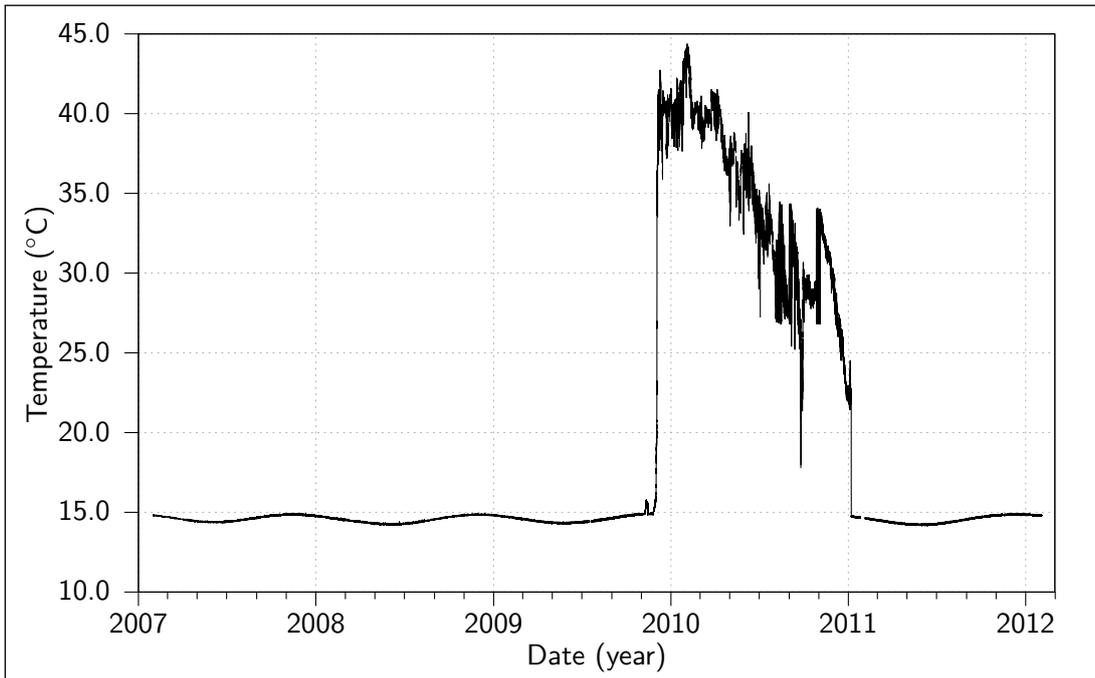


Figure 6.20: Malfunction in sensor UT911. Temperature reporting becomes highly aberrant between approximately 2009-12-01 and 2011-01-07.

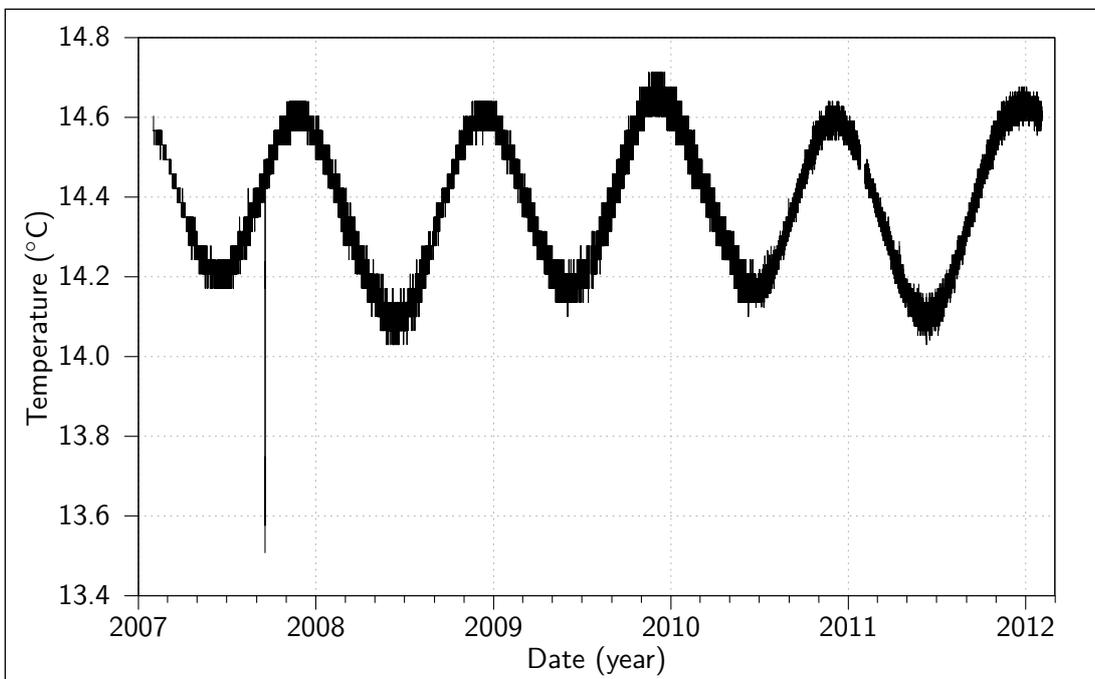


Figure 6.21: Perturbation in sensor PT908. At approximately 2007-09-17 an abrupt and short lived variation in reported temperature is recorded.

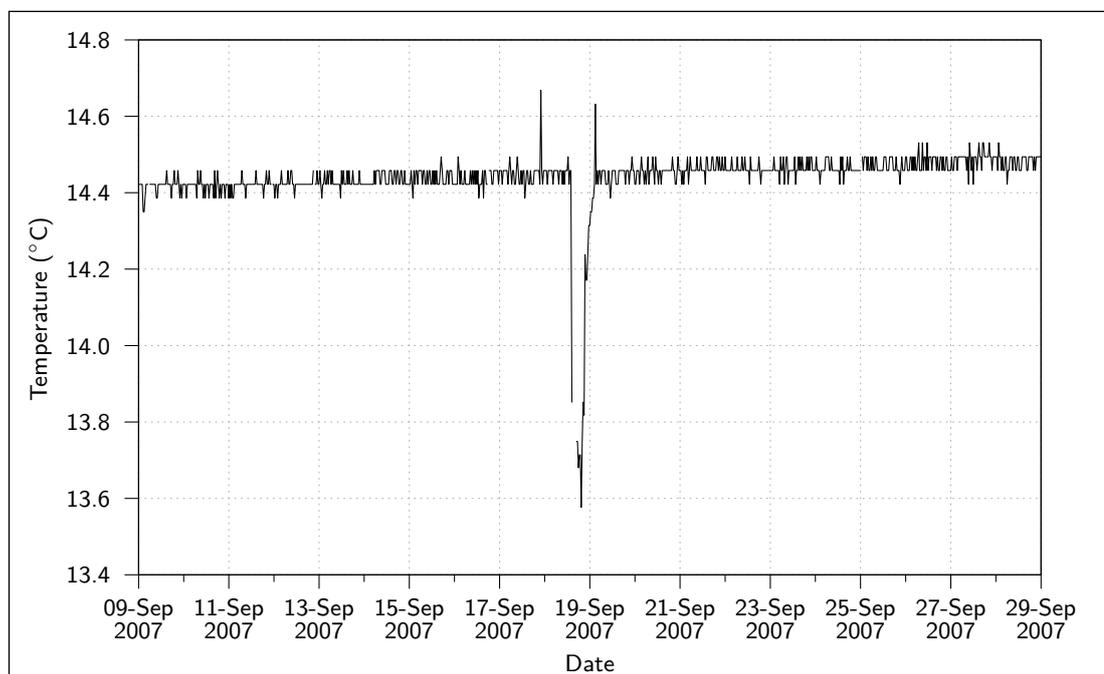


Figure 6.22: Detail of perturbation in sensor PT908. A sudden drop from and subsequent return to the norm in reported temperature can be seen.

Figure 6.23 shows UT901's sensor record in grey, with the smoothed results overlain in black. Arrows highlight the positive and negative shifts.

6.4.5 Representative values, SSA, and trend removal

6.4.5.1 Representative values

The SSA process, as described in Chapter Four, does not strictly require a uniform sample rate, but does make use of a moving window that is a fixed number of points in length rather than a fixed period of time in length. Additionally, the computational effort required to decompose an input time series into components increases rapidly with the length of the window used. As the identification and removal of potential 'long term' trends is of interest, requiring a 'long' window, a compromise regarding the input data's uniformity and resolution has been reached.

The down sampling of data to daily representative values generates an input with a resolution suitable for 'long term' trends to be identified. Additionally, the resulting sample rate possesses enough uniformity to enable the use of a 365 point window to sufficiently equate to a year of elapsed time. The down-sampling process calculated the uniformly weighted average of the points in a 24 hour window, centred on 12:00:00 (noon) of each day in the specified date range of interest. Where the number of points averaged was greater than 10 the outliers were removed from the daily subset before averaging (see Section 4.3.2.2).

The sample rate of the resulting dataset is depicted in Figure 6.24. The non-uniformity remaining in the resulting sample rate is only present where gaps in the data previously spanned more than one full day. The remaining non-uniformity in the dataset has been considered isolated enough to patch with linearly interpolated data. As such, the SSA analysis was performed on a uniform daily representative dataset.

6.4.5.2 Singular Spectrum Analysis

The SSA process, using a 365 point window, was applied to each down-sampled sensor record, decomposing it into 365 separate components that sum together to the original (down-sampled) input. The average time to complete each decomposition with the stated parameters and input

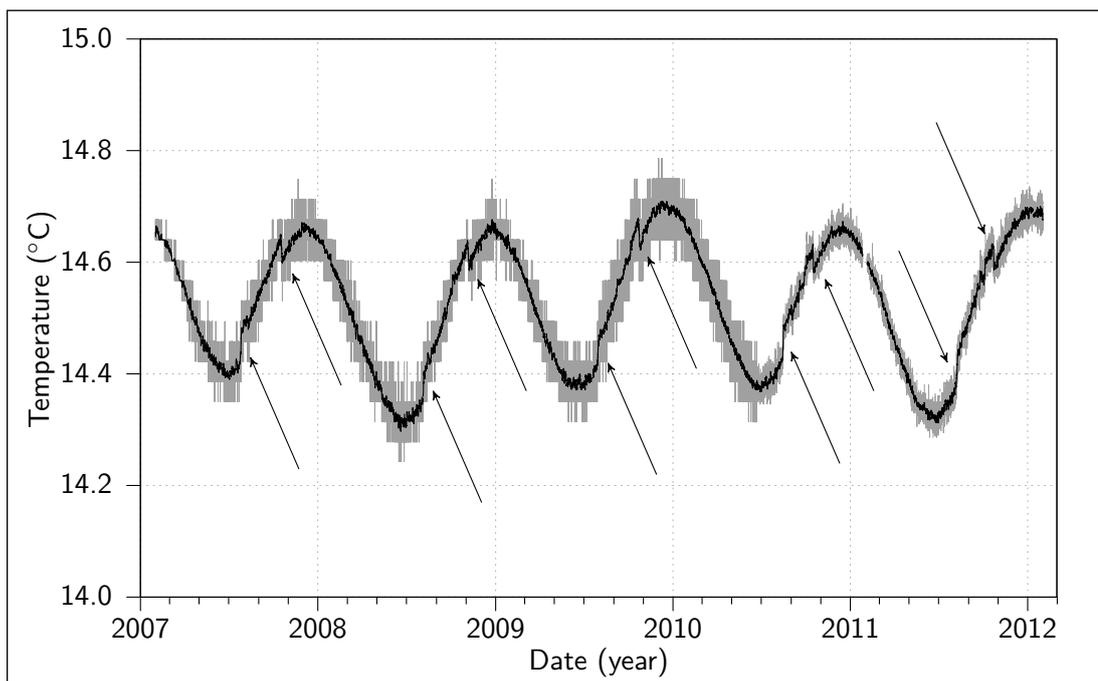


Figure 6.23: Sensor UT901's repeating anomaly. Original (de-spiked) signal in grey, with smoothed signal (uniformly weighted moving average, 24 hour window) overlain in black. Arrows highlight the systematic steps up and down.

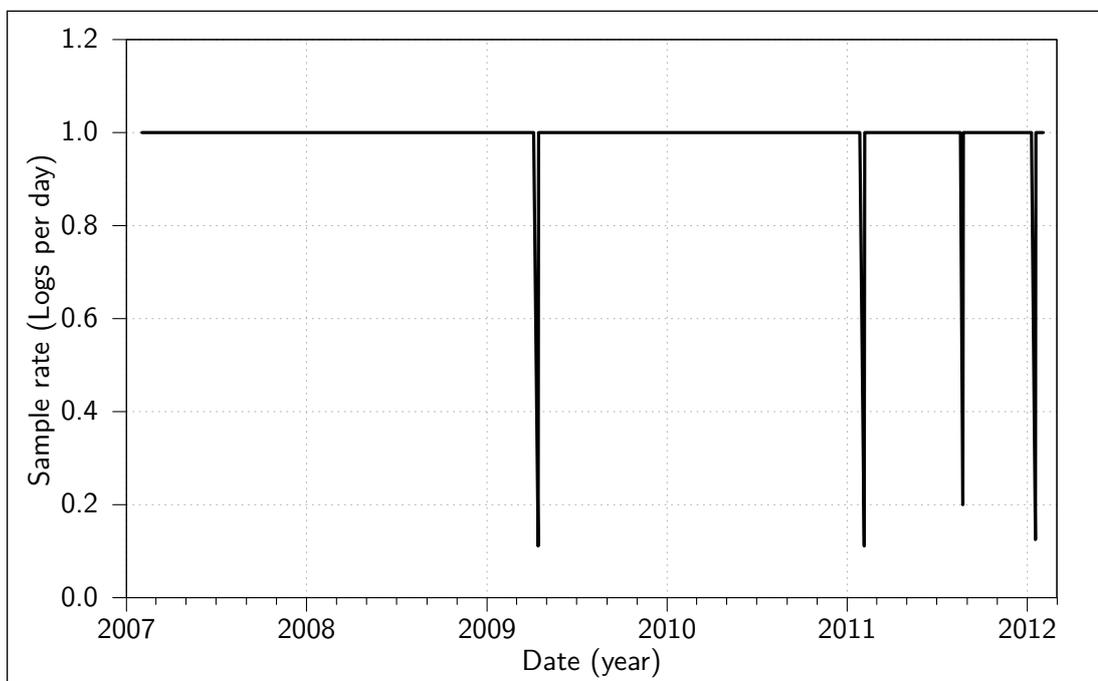


Figure 6.24: Sample rate resulting from the down-sampling of the full resolution data to daily representative values.

was 31 minutes and 52 seconds. The displacement sensor records and the temperature sensor records originating from sensors within DA3147G01 were typically quicker to decompose. The average time taken to complete the decomposition process for these sensor records was 18 minutes and 50 seconds.

The definition of a trend given in Golyandina et al. (2001) was used to evaluate whether a component resulting from the decomposition process can be qualified as a trend or not, i.e. if it could be described as a non-stationary and 'slowly' varying component of a time series across the duration of the record. The precise definition of 'slowly' is left largely to the context of the data analysis and operator discretion, but is elaborated on further by the description of trend scales that 'describe the general tendency of the series' and the 'detailed trend'. The former of the two sub-definitions is favoured for removal in order to expose second order detail.

The time series components arising from the decomposition process were sorted by their associated eigenvalue magnitude from the absolute greatest in magnitude to the absolute least in magnitude. This resulted in a grouping of the trending components together at the start of the sorted list obtained. A visual inspection was used to determine which of the SSA derived components fitted the definition of a long term trend given above. The number of components derived from each time series that were classified as long term trends is shown in Table 6.2. These trending components were therefore subsequently removed from their respective time series.

The use of 'zero' components for the trend removal on those sensor records listed in Table 6.2, specifically the column headed 'Zero', is due to abrupt changes in the recorded magnitudes that are not well represented by the decomposition process. Falling into this category are most of the operator controlled injection filters, the lid's lateral displacement sensors, UB902 which experienced a large change in magnitude due to a macro scale gas flow intercepting it, and UT911 which has a period of apparent malfunction with step changes at the beginning and end of the anomalous period.

6.4.5.3 Trend removal

The SSA derived components possess the same sample rate as the down-sampled sensor records used for input (i.e. the daily representative values). The subtraction of the selected components

Table 6.2: Number of SSA components summed to produce the trend, by sensor. PT901 and UT924 are not included due to the absence of data in their records.

| Zero | One | Two |
|-------|------------------------------|-------|
| DP906 | All un-listed sensor records | UB901 |
| DP907 | | UB925 |
| FB903 | | UB926 |
| FB904 | | |
| FL901 | | |
| FL902 | | |
| FL903 | | |
| FL904 | | |
| FM905 | | |
| FM906 | | |
| FM907 | | |
| FM908 | | |
| FR902 | | |
| FU909 | | |
| FU910 | | |
| FU911 | | |
| FU912 | | |
| UB902 | | |
| UT911 | | |

(trends) from the full resolution sensor records therefore requires a value for each full resolution datum point to be interpolated from the points available in the SSA derived trend. As the SSA derived trends possess a slowly evolving nature, interpolating between the daily data to a higher resolution is unlikely to be subject to the interpolation issues highlighted in Chapter Four.

The resultant sensor records after the trend subtraction are referred to as the 'de-trended residuals'. All subsequent toolkit outputs reported here utilise the de-trended residual sensor records as input, as described in Section 6.3 and depicted in Figure 6.1.

6.4.6 NDFTs of de-trended residuals

NDFTs of each of the sensor records were taken after the de-trending process, with the same frequency domain inspected as that described in Section 6.4.2. The zero frequency components were not subtracted from the corresponding input data in this case however, as the de-trending process leaves the residual data with an average value close to zero.

A greater number of frequency domain forms are present in the de-trended residual NDFTs when compared with the NDFTs of the de-spiked time series, described in Section 6.4.2. The NDFTs of the de-trended residual sensor records, noted as having had no trend component removed (the column headed 'Zero' in Table 6.2), possess the low frequency ramping indicative of long term trends that is described in Section 6.4.2. The NDFTs of the sensor records noted as having two components contributing to the removed trend (the column headed 'Two' in Table 6.2) possess an amplitude peak at or around the 365 day wavelength. However in the cases of UB925 and UB926 there is lack a single prominent peak at the annual cycle frequency. For comparison, UB901 possesses a prominent peak at the 365 day wavelength. Figure 6.25 depicts the frequency domains for UB901's, UB925's, and UB926's sensor records.

Of the sensor records de-trended with one SSA derived component, the temperature sensors (including the HRL air temperature and internal canister temperature) possess a prominent peak at the 365 day wavelength. Specific note is made of the HRL air temperature which possesses an annual temperature cycle of approximately 1.98 °C and a daily temperature cycle of 0.06 °C. A summary of the amplitudes and phase offsets of the temperature cycles with depth of the sensor records originating from sensors within DA3147G01 (excluding UT911) are presented in Figures 6.26 and 6.27, respectively. The amplitude and phase offset values for the

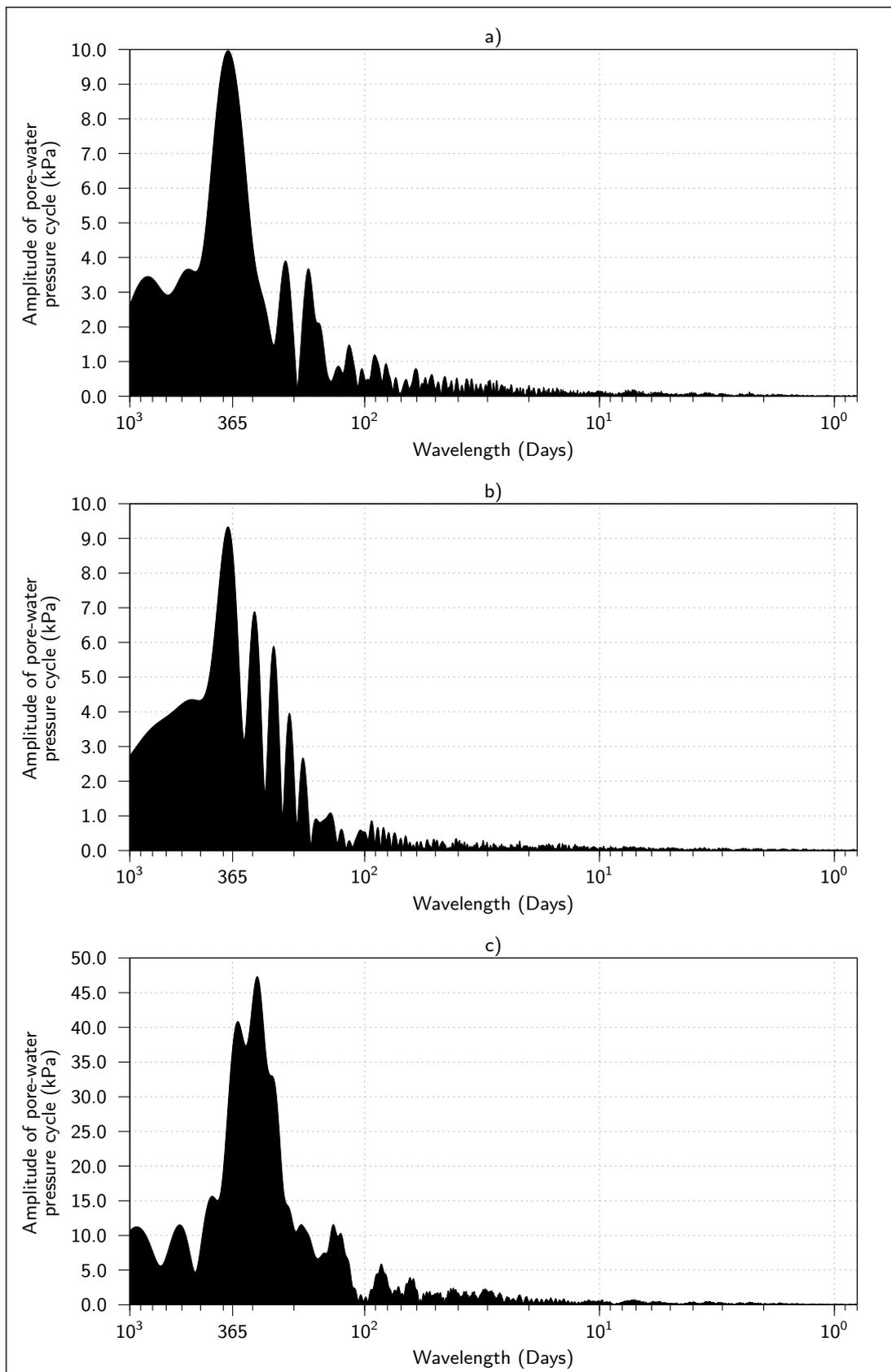


Figure 6.25: NDFTs of selected bentonite pore-water pressure sensor records de-trended residuals. a) UB901; b) UB925; c) UB926

internal canister temperature at the 365 day wavelength are 0.37°C and approximately 88 days from the HRL air temperature.

The NDFT of FR901, which is the only sensor record from the set of hydration mats and injection filters to have an SSA derived trend removed, possesses a series of high wavelength peaks from approximately 365 days to approximately 2,000 days. The NDFTs of the vertical lid displacement sensor records possess a prominent annual peak, with some moderate peaks in the lower frequency (high wavelength, > 365 days) region. NDFTs of the lid anchor cable load sensor records possess a prominent annual peaks with minor low frequency amplitude in comparison to the NDFTs of the lid displacement sensor records.

Of the remaining 64 stress and pore-water pressure sensor records, nine are located in the pressure relief holes and the remaining 55 are located within DA3147G01. The NDFTs of the pore-water pressure sensor records from the pressure relief holes typically possess a peak or peaks around the 365 day wavelength. However they lack a single prominent peak at the 365 days wavelength. The exception to this is UR929, located in PRH-2, which has a single prominent peak corresponding to an annual cycle. Additionally, a small daily frequency component is discernible in the PRH NDFTs after de-trending. The daily frequency component of UR931 retains a magnitude of approximately 0.1 kPa, with the magnitudes of the others sensors within the PRHs ranging from approximately 0.2 kPa to approximately 0.5 kPa and averaging approximately 0.3 kPa.

Of the 55 stress and pore-water pressure sensor records originating from the sensors located within DA3147G01, 48 possess a prominent annual cycle peak in the NDFT results. The remaining seven sensors (PC901, PR906, PB927, PB901, UR916, UR907, and UR910) all possess peaks at (or very near to) *and* around the 365 day wavelength, but lack a single prominent peak at the 365 days wavelength.

Figures 6.28 and 6.29 present a summary of the amplitudes and relative phase offsets of the de-trended residual sensor records from stress and pore-water pressure cycles within DA3147G01. The phase offsets and amplitudes are presented with respect to depth of the sensor in DA3147G01. The correlations between sensor depth and annual cycle amplitudes/phase offsets observed in the temperature sensor record (Figures 6.26 and 6.27) are not present in the corresponding stress and pore-water pressure sensor records.

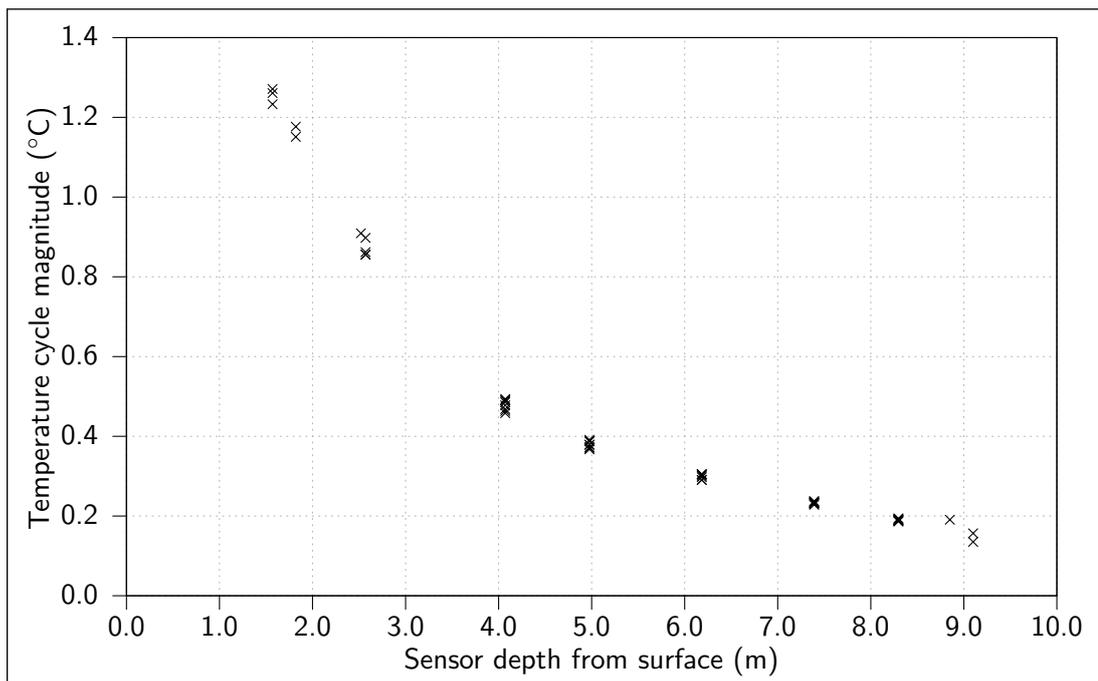


Figure 6.26: Amplitude in °C of down-hole temperature sensors' annual cycles with depth.

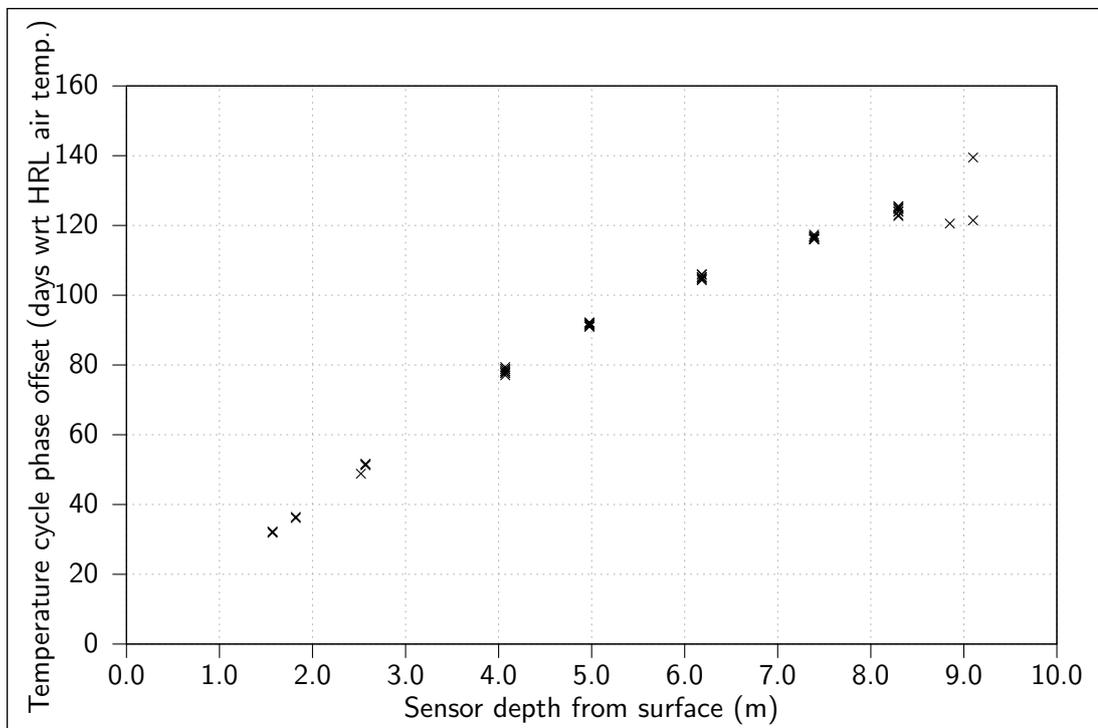


Figure 6.27: Phase offset in days (with respect to HRL air temperature) of down-hole temperature sensors' annual cycles with depth.

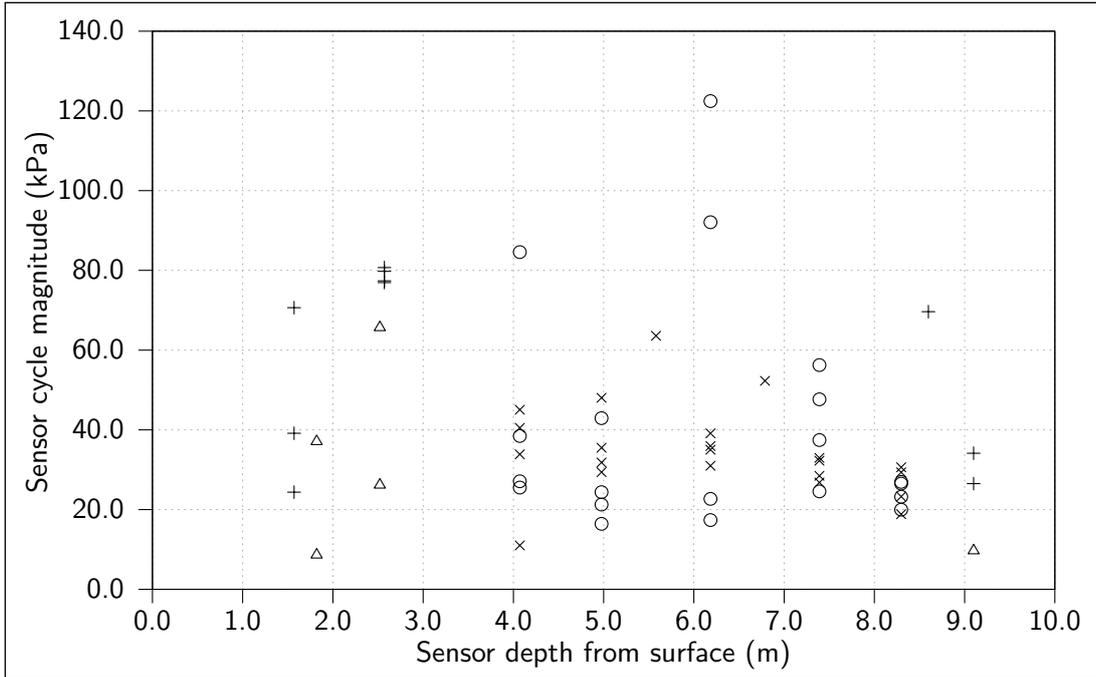


Figure 6.28: Amplitude in kPa of down-hole stress and pore-water pressure sensors' annual cycles with depth. 'Greek crosses' signify axial stress sensors, 'saltires' signify radial stress sensors, 'circles' signify rock wall pore-water pressure sensors, and 'triangles' signify bentonite pore-water pressure sensors.

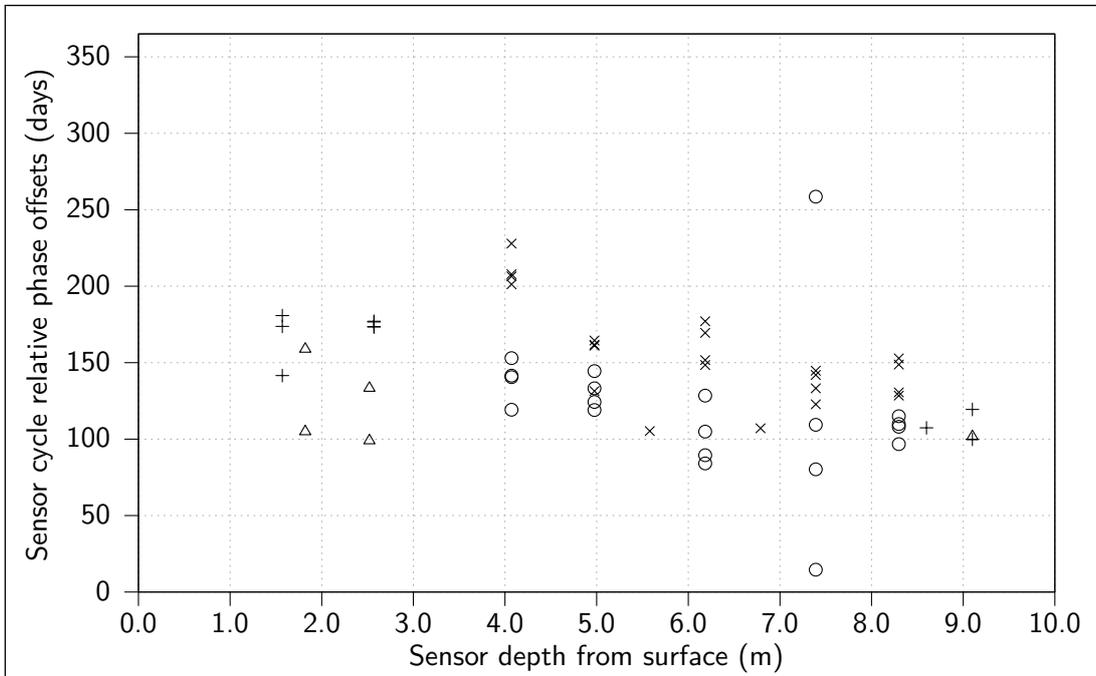


Figure 6.29: Phase offset in days (from an arbitrary epoch) of down-hole stress and pore-water pressure sensors' annual cycles with depth. 'Greek crosses' signify axial stress sensors, 'saltires' signify radial stress sensors, 'circles' signify rock wall pore-water pressure sensors, and 'triangles' signify bentonite pore-water pressure sensors.

6.4.7 Event candidate detection of de-trended residual data

When viewed at the scale used in Section 6.4.3, the event candidate intensity plots are relatively unchanged after the de-trending process. The primary differences are slight alterations in contrast between the brief high intensity areas and the lower intensity areas between them in a small number of sensor records. In most cases however, the differences between the initial and de-trended residual intensity plots are not perceptible at the macro scale.

6.4.8 Smoothing process of de-trended residual data

The results of the smoothing process applied to the de-trended residual sensor records remain unchanged from those reported in Section 6.4.4, i.e. the smoothing of the sensor records yielded a reduction in noise, but in all but one case did not reveal any noteworthy or remarkable phenomena. The noteworthy result revealed by the application of smoothing was a recurring anomaly in UT901's temperature record (previously noted in Section 6.4.4).

6.5 Summary

The application of NUDAT to Lasgit's dataset, and the quantitative and qualitative results arising from the application have been presented in detail in the sections above. Using NUDAT, Lasgit's dataset has been systematically analysed in a way that has not previously been attempted, i.e. on a second order scale, and while taking account of the non-uniformity in the sampling rate of the sensor records.

A summary and overview of each toolkit output, before and after the de-trending process, has been presented in the sections above. Such toolkit outputs have been successful in identifying: i) the number of detected spikes in the dataset, and their distribution throughout the sensor records; ii) the annual frequency content present in the dataset, and in the case of the temperature sensor records the relationship of such frequency content to depth within DA3147G01; and iii) the extent to which synchronisation between sensors of a similar type occurs on a macro-scale.

The detailed discussion of the results presented in this chapter and Chapter Seven, focusing on the inferences that can be made from the phenomena identified in the sensor records and the

interpretation of second order events, is presented in Chapter Eight.

Chapter Seven

Second order events

7.1 Introduction

As stated in Chapter Six, Chapters Six and Seven collectively present the results of the application of the Non-Uniform Data Analysis Toolkit (NUDAT) to the Large Scale Gas Injection Test's (Lasgit's) dataset. Chapter Six presents the results of the application of NUDAT to Lasgit's dataset from a macro-behaviour perspective, while Chapter Seven presents the detailed investigation of three second order events highlighted by the toolkit.

The Exploratory Data Analysis (EDA) enabled by NUDAT, and presented in Chapters Six and Seven, is intended to build on the analysis and reporting already performed by the British Geological Survey (BGS) that has been summarised/referenced in Chapter Three. The phenomena represented in Lasgit's dataset that is exposed and quantified by the application of NUDAT will be the focus of the reporting.

Within this chapter a brief description pertaining to the choice of events investigated is presented in Section 7.2, followed by the detailed investigation of three second order events highlighted by the toolkit in Sections 7.3 to 7.5. The discussion and inferences made from the results presented in Chapters Six and Seven are subsequently presented in Chapter Eight.

7.2 Choice of event candidates investigated

The selection of the specific second order event candidates to be investigated was influenced by the relative isolation and prominence of an event candidate impulse, and by the proximity (both temporally and spatially) to known macro scale gas flow events that have occurred during gas injection tests. The macro scale gas flow occurring from FL903 to FL901 and onto UB902 at the end of 2009 (Section 3.3) encourages attention towards data arising from sensors

located at or near the bottom of DA3147G01, and to data recorded during late 2009 and on into 2010. The investigations of three events meeting this criteria are presented in the following sections.

7.3 Event at 2010–01–10

Figure 7.1 shows the pore-water pressure sensor UB901 record with the event candidate results of the corresponding de-trended residual sensor record. The absolute values of the event candidate intensity are used here, derived from a 48 hour window length. The impulses above the Event Candidate Threshold (ECT) are highlighted in darker grey. The time period corresponding to the event candidate highlighted by the arrow is approximately between 2010–01–06 and 2010–01–14. This time period is presented in detail for a number of sensor records.

Figure 7.2 shows UB901's sensor record between 2010–01–06 and 2010–01–14 in detail. The event candidate intensity shown was derived using a 24 hour window. Within the time period 2010–01–06 to 2010–01–14, six sharp increases, or 'steps', in UB901's reported pore-water pressure are observed, ranging from 2 kPa to 5 kPa in magnitude. The reported pore-water pressure does not have an appreciable change in magnitude between the steps.

Figure 7.3 shows the event candidate intensity in the specified time period for all in-bentonite pore-water pressure sensor records. In this case, the event candidate intensities derived from each sensor record are shown ranked by magnitude.

UB902 has a period of high intensity that synchronises closely with the intensity peak highlighted in Figure 7.2. UB926 also has a high intensity event candidate near to the highlighted period, but slightly offset. Figure 7.4 shows UB902's sensor record between 2010–01–06 and 2010–01–14 in detail. The event candidate intensity shown was derived using a 24 hour window.

The reported pore-water pressure in UB902's sensor record possesses a sharp drop of approximately 30 to 40 kPa that is closely synchronised with the largest positive step in Figure 7.2. Three less well defined and smaller drops in pore-water pressure are also observed after the synchronised event in the considered time period. Initially there is a positive trend within the period observed, which appears to partially level off after the main drop.

Figure 7.5 shows UB926's sensor record between 2010–01–06 and 2010–01–14 in detail. The

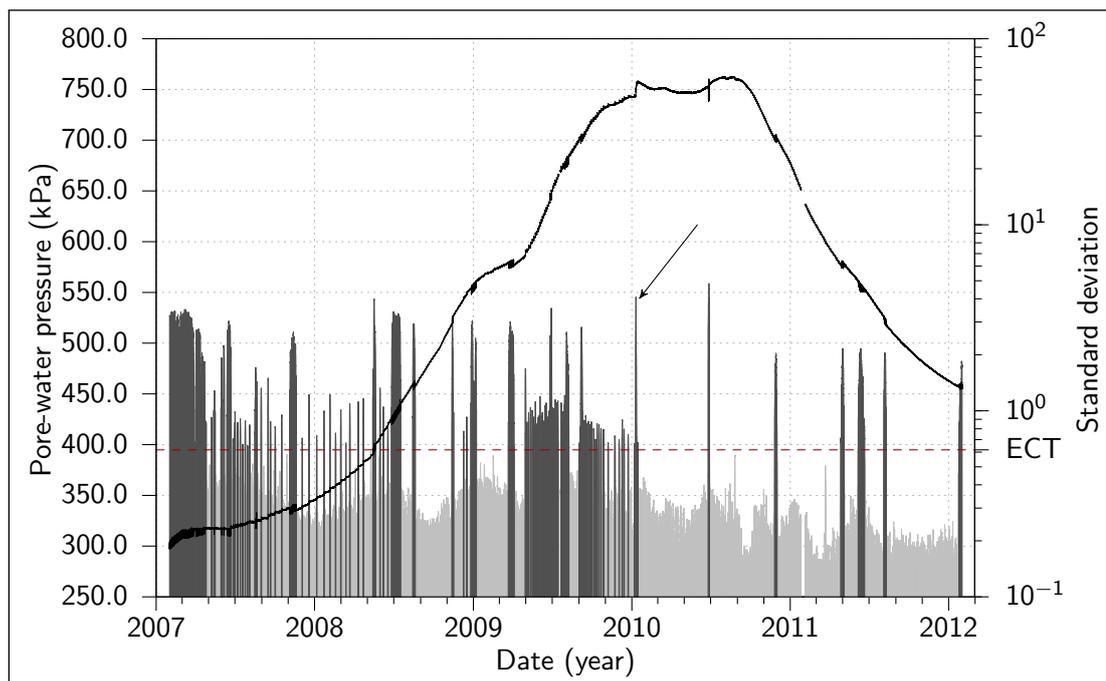


Figure 7.1: Pore-water pressure sensor record UB901 (black line), with the de-trended residual event candidate impulses (grey). Event candidates are highlighted in dark grey where they exceed the ECT.

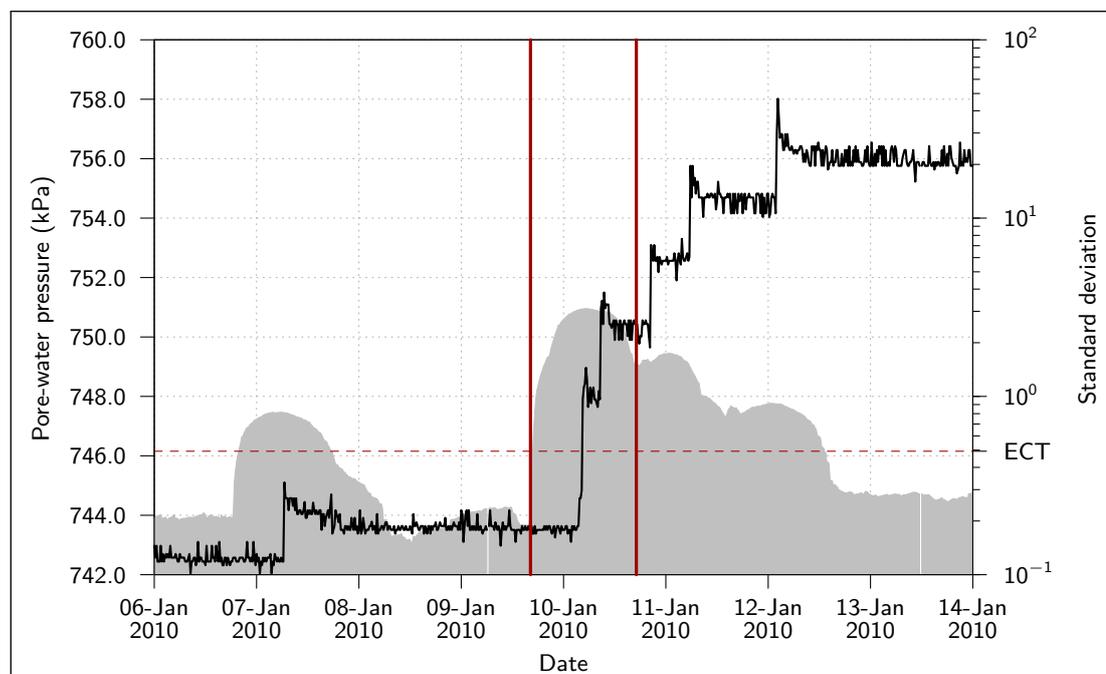


Figure 7.2: Detailed view of UB901's sensor record from 2010-01-06 to 2010-01-14.

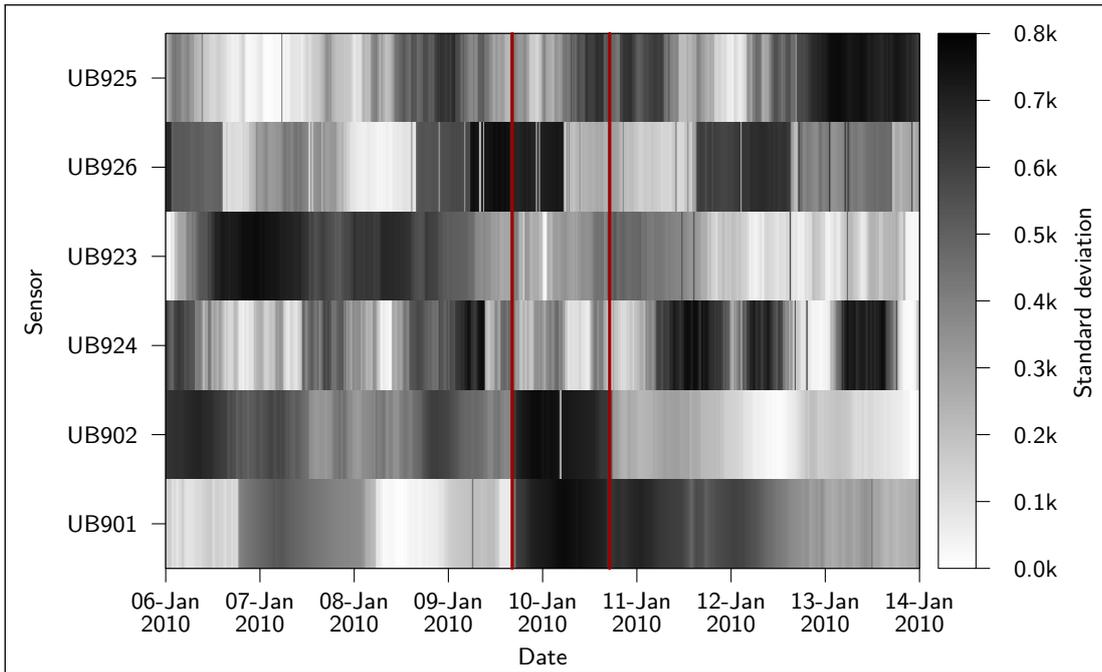


Figure 7.3: Intensity plot of in-bentonite pore-water pressure sensor records between 2010-01-06 and 2010-01-14 (ranked values).

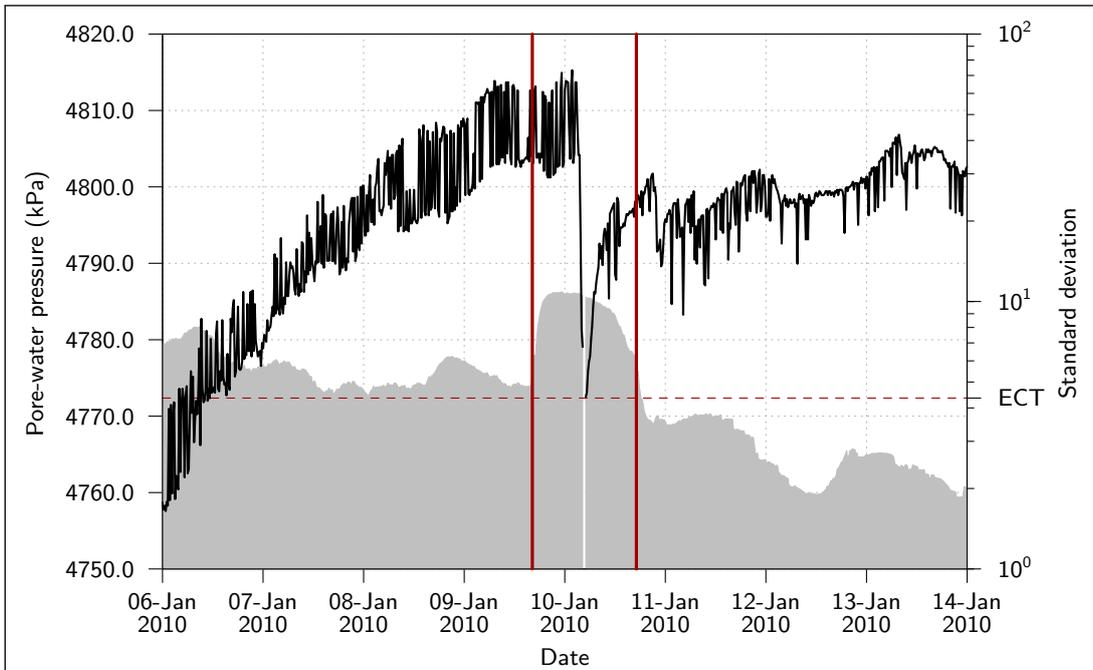


Figure 7.4: Detailed view of UB902's sensor record from 2010-01-06 to 2010-01-14.

event candidate intensity was derived using a 24 hour window and does not reach a sufficient magnitude to cross the ECT line. The increased event candidate intensity that occurs slightly before the highlighted period in UB926's sensor record corresponds to a spike with a magnitude of approximately 7 kPa. There is a slight downward trend within the period observed.

Figure 7.6 shows the event candidate intensity in the specified time period for all injection filter and hydration mat sensor records. In this case the intensities are presented as absolute values and not scaled. FL901 possesses a period of high intensity synchronised with with the period highlighted in Figure 7.2. The detailed view of FL901's sensor record between 2010-01-06 and 2010-01-14 is shown in Figure 7.7. During the highlighted period FL901's reported pore-water pressure drops sharply by approximately 17 kPa. There is a general downward trend within the period observed.

As described in Section 3.3, FL903 is the gas injection source for Gas Test Two, which resulted in major gas flow to FL901 and UB902. Figure 7.8 shows a detailed view of FL903's sensor record from 2010-01-06 to 2010-01-14. Figure 7.6 does not indicate an event synchronised with the specified time period, an observation that is confirmed in Figure 7.8.

Figure 7.9 shows the event candidate intensity in the specified time period for all axial stress sensor records. The intensities are presented as normalised values and derived from a 24 hour window.

Sensor records PC901 and PB901 have high intensity regions that are synchronised with the specified time period. Figures 7.10 and 7.11 show detailed views of PC901's and PB901's sensor records respectively. Both PC901 and PB901 possess sudden increases of approximately 5 kPa during the highlighted time period. These steps occur in the context of the sensors reporting over 5 MPa.

7.4 Event at 2010-06-27

Figure 7.12 shows the pore-water pressure sensor UB901 record with the event candidate results of the corresponding de-trended residual sensor record. Absolute values of the event candidate intensity are used here, derived from a 48 hour window length. Impulses above the ECT are highlighted in darker grey. The time period corresponding to the event candidate highlighted

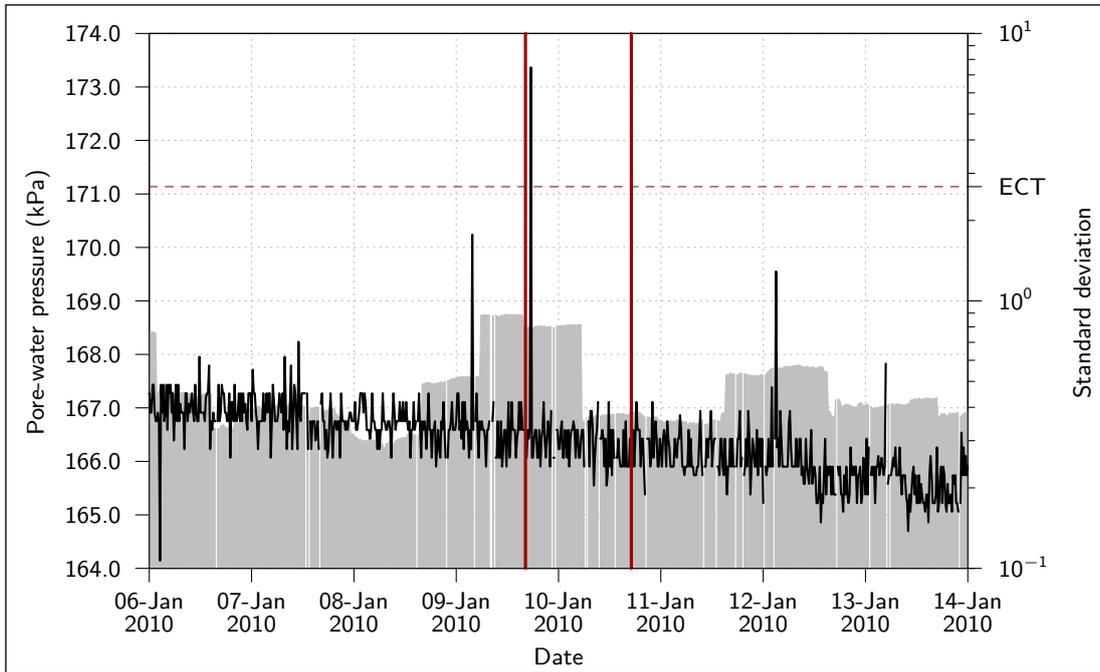


Figure 7.5: Detailed view of UB926's sensor record from 2010-01-06 to 2010-01-14.

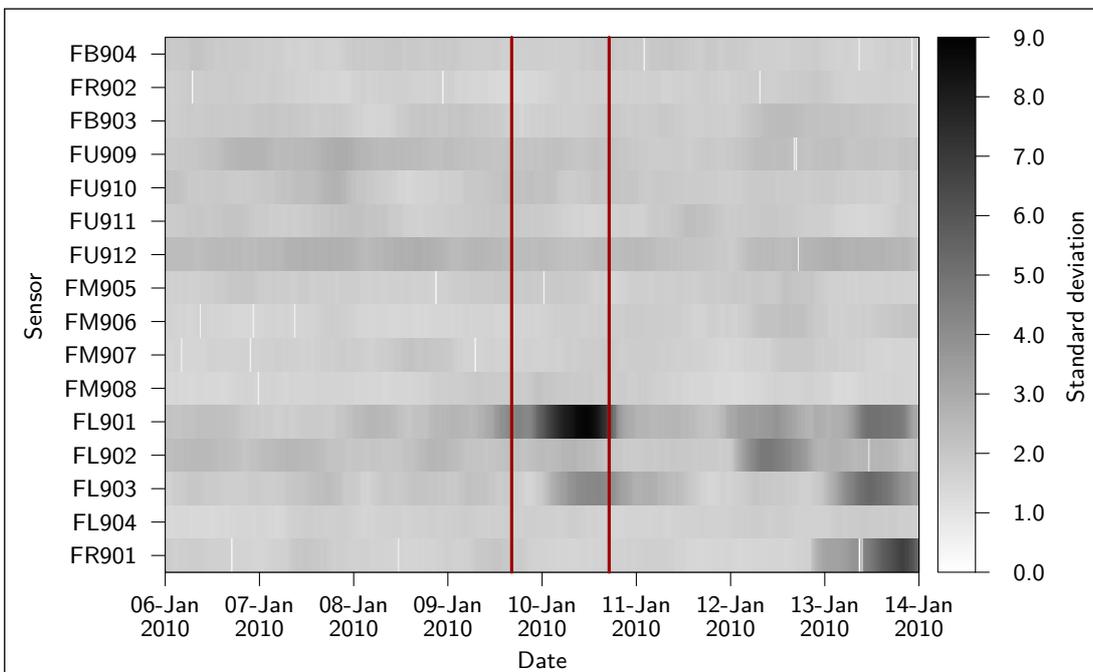


Figure 7.6: Intensity plot of injection filter and hydration mat sensor records between 2010-01-06 and 2010-01-14 (absolute SD values).

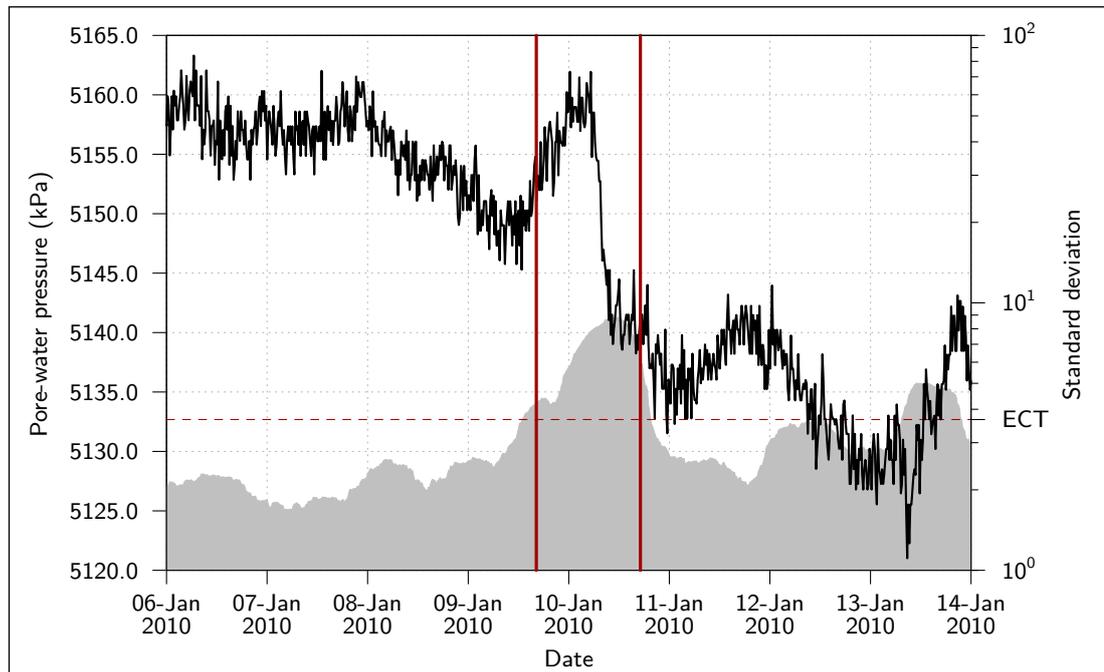


Figure 7.7: Detailed view of FL901's sensor record from 2010-01-06 to 2010-01-14.

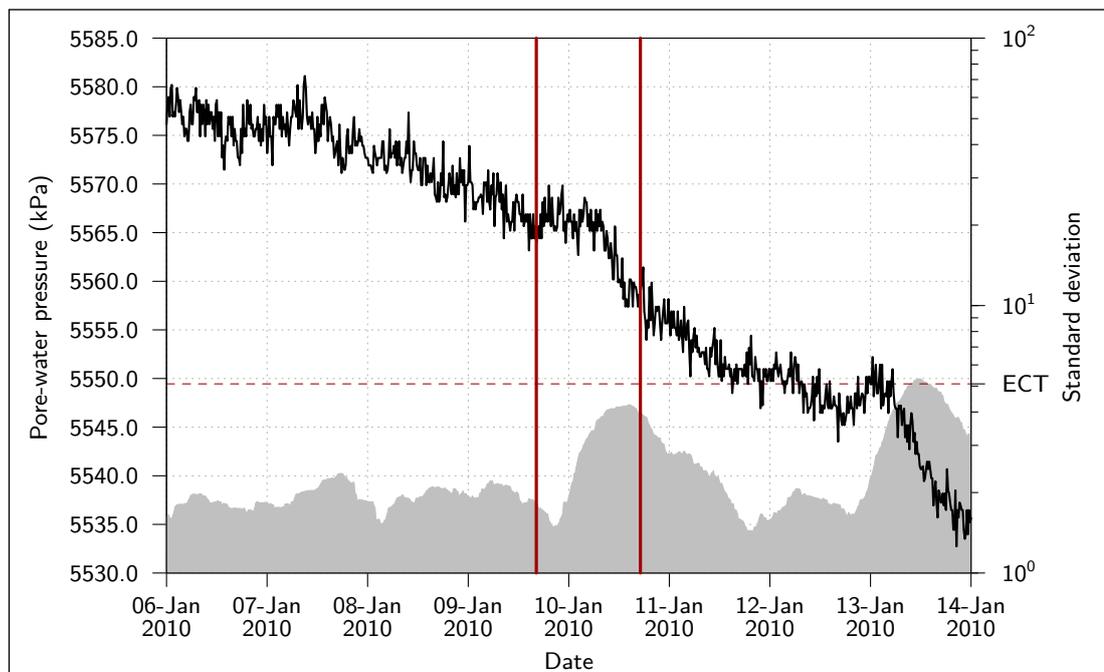


Figure 7.8: Detailed view of FL903's sensor record from 2010-01-06 to 2010-01-14.

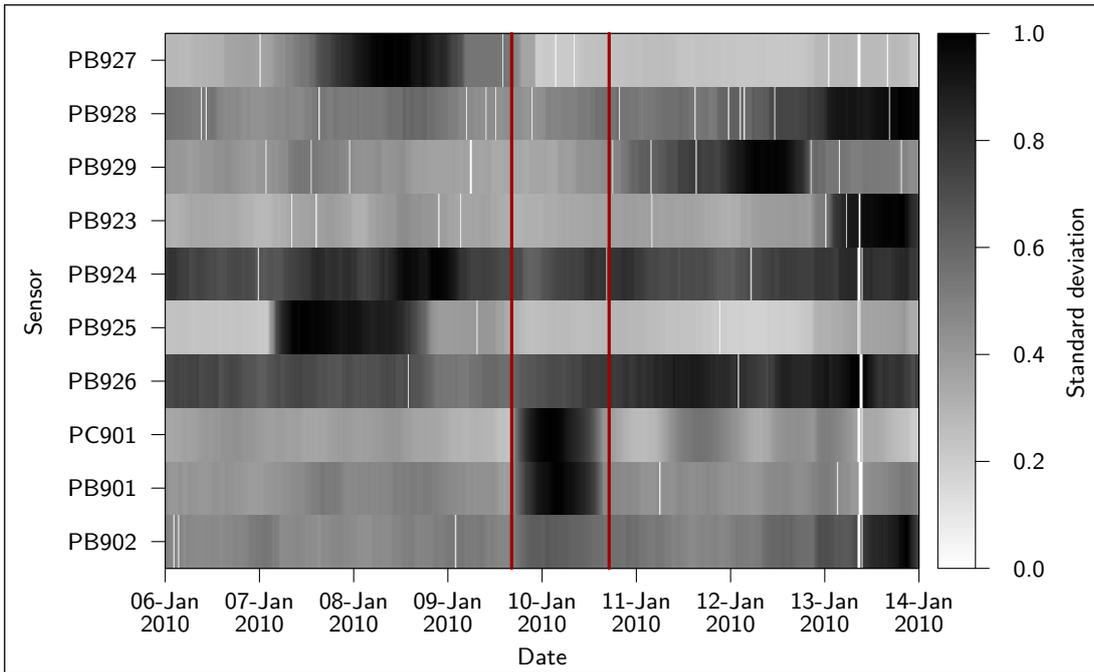


Figure 7.9: Intensity plot of axial stress sensor records between 2010-01-06 and 2010-01-14 (normalised SD values, 24 hour window).

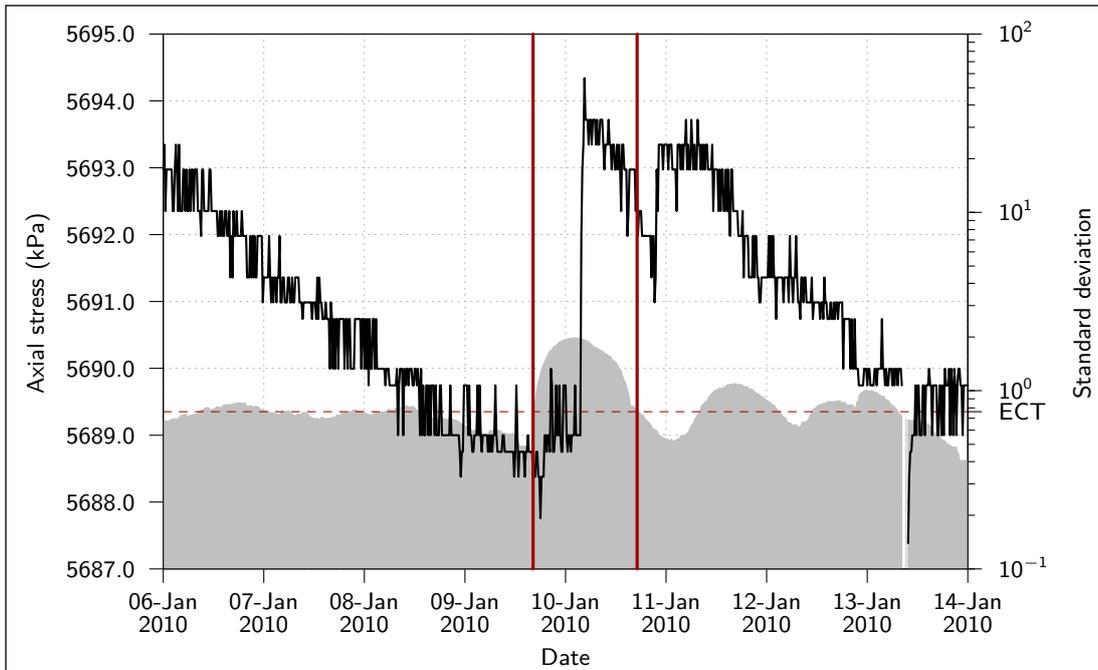


Figure 7.10: Detailed view of PC901's sensor record from 2010-01-06 to 2010-01-14.

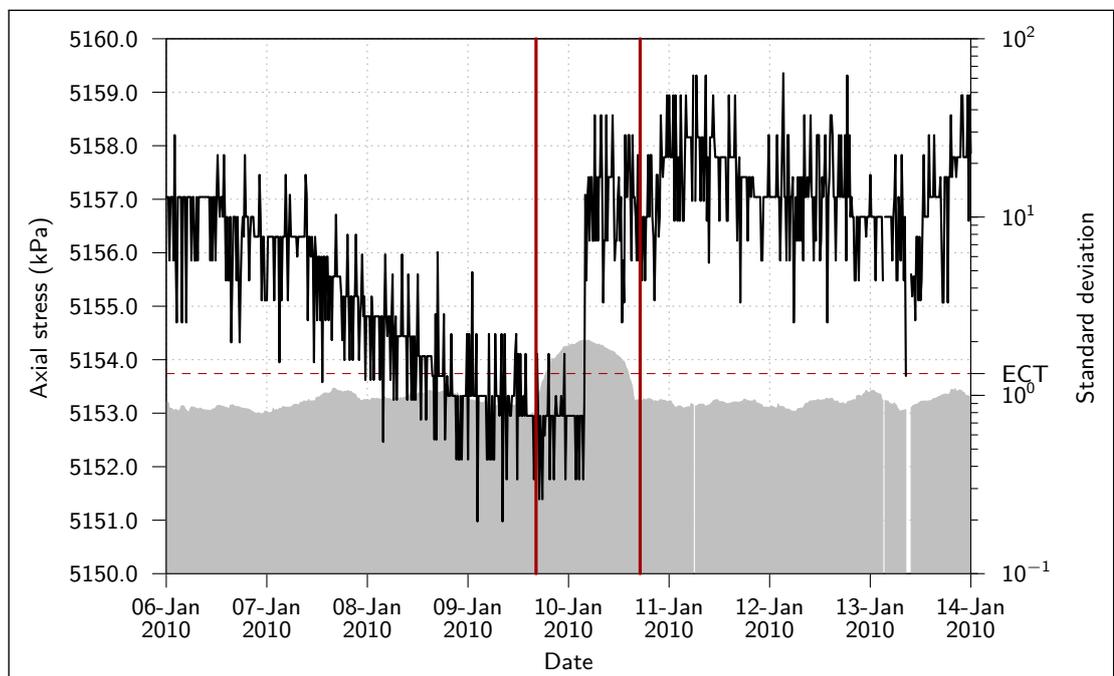


Figure 7.11: Detailed view of PB901's sensor record from 2010-01-06 to 2010-01-14.

by the arrow is approximately 2010-06-23 to 2010-07-01. This time period is presented in detail for a number of sensor records.

Figure 7.13 shows UB901's sensor record between 2010-06-23 and 2010-07-01 in detail. The event candidate intensity shown is derived using a 24 hour window. Within the highlighted time period the sensor record switches abruptly from possessing a low noise content to a comparatively high noise content for approximately 28 hours, and then returns abruptly to the preceding level of noise content.

A composite of the event candidate intensity in the specified time period for all axial stress sensor records, in-bentonite pore-water pressure sensor records, radial stress sensor records, and rock wall-water pressure sensor records is shown in Figure 7.14. An increase in the event candidate intensity is observable across most of the range of sensor records during the time period highlighted.

Notably the intensities of sensor records PC901, PC902, and PC903 (shown in Figure 7.15), sensor records PB928 (shown in Figure 7.16), and sensor record UR911 (shown in Figure 7.17) are exceptions to the otherwise ubiquitous increase in intensity seen in the highlighted period in Figure 7.14.

PC901, PC902, and PC903's sensor records have no observable disturbance present in the highlighted time period. PB928 and UR911's sensor records however do have a local increase in noise magnitude during the highlighted time period, however similarly sized perturbations exist nearby that reduce the comparative event candidate intensity measure. With the exception of PC901, PC902, and PC90, all sensors listed in Figure 7.14 possess a phenomenon similar to the originally identified event, each having a slightly varying magnitude and duration.

It is noted that the relative levels of noise before and after the increase during the identified events are not necessarily similar in each sensor record. Specifically, UB925 possesses significantly less noise immediately after the identified event in comparison to immediately before. Additionally, events or event indicators such as those described above are not present in the temperature sensor records arising from within DA3147G01, any of the filters or filter mat sensor records, or in any of the sensor records originating outside of DA3147G01.

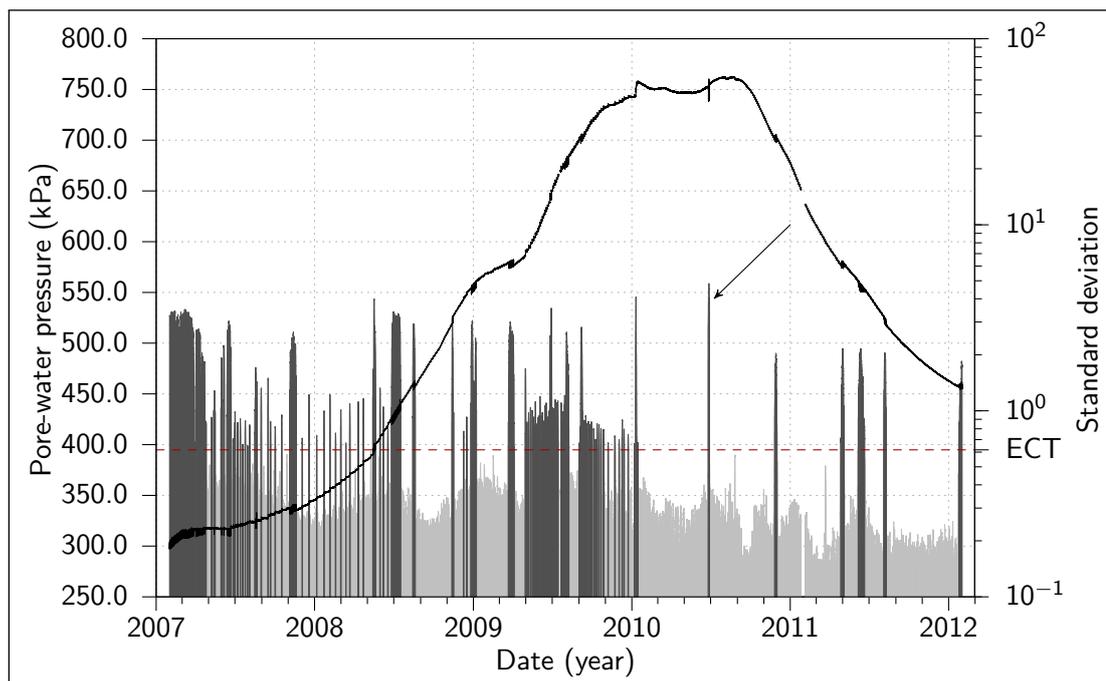


Figure 7.12: Pore-water pressure sensor record UB901 (black line), with the de-trended residual event candidate impulses (grey). Event candidates are highlighted in dark grey where they exceed the ECT.

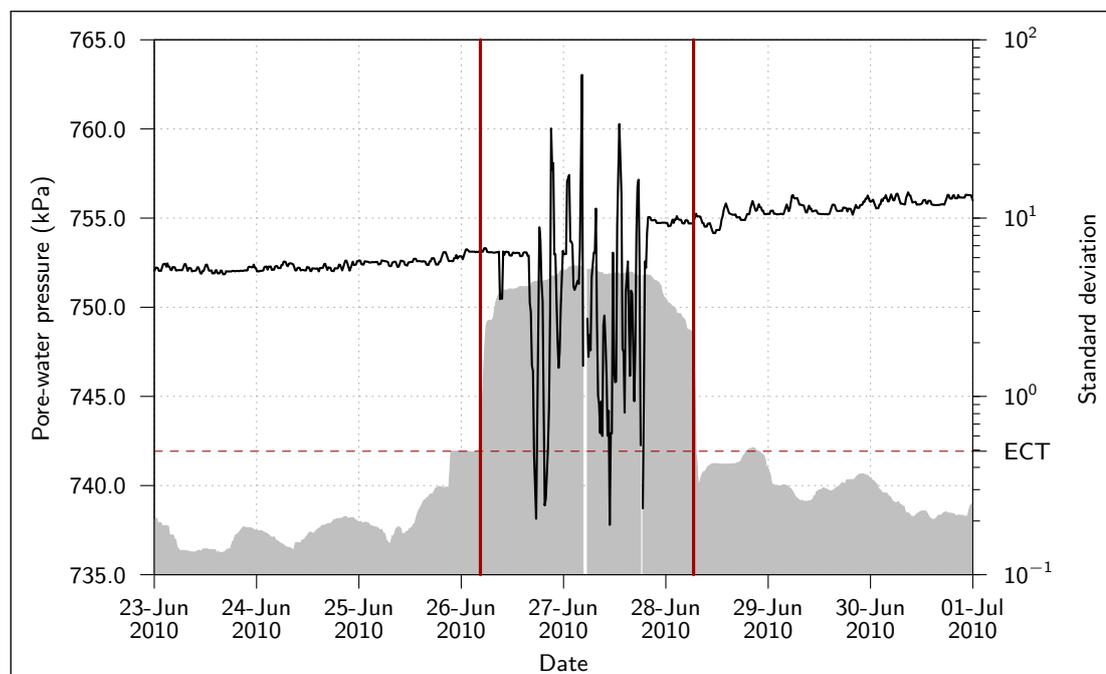


Figure 7.13: Detailed view of UB901's sensor record from 2010-06-23 to 2010-07-01.

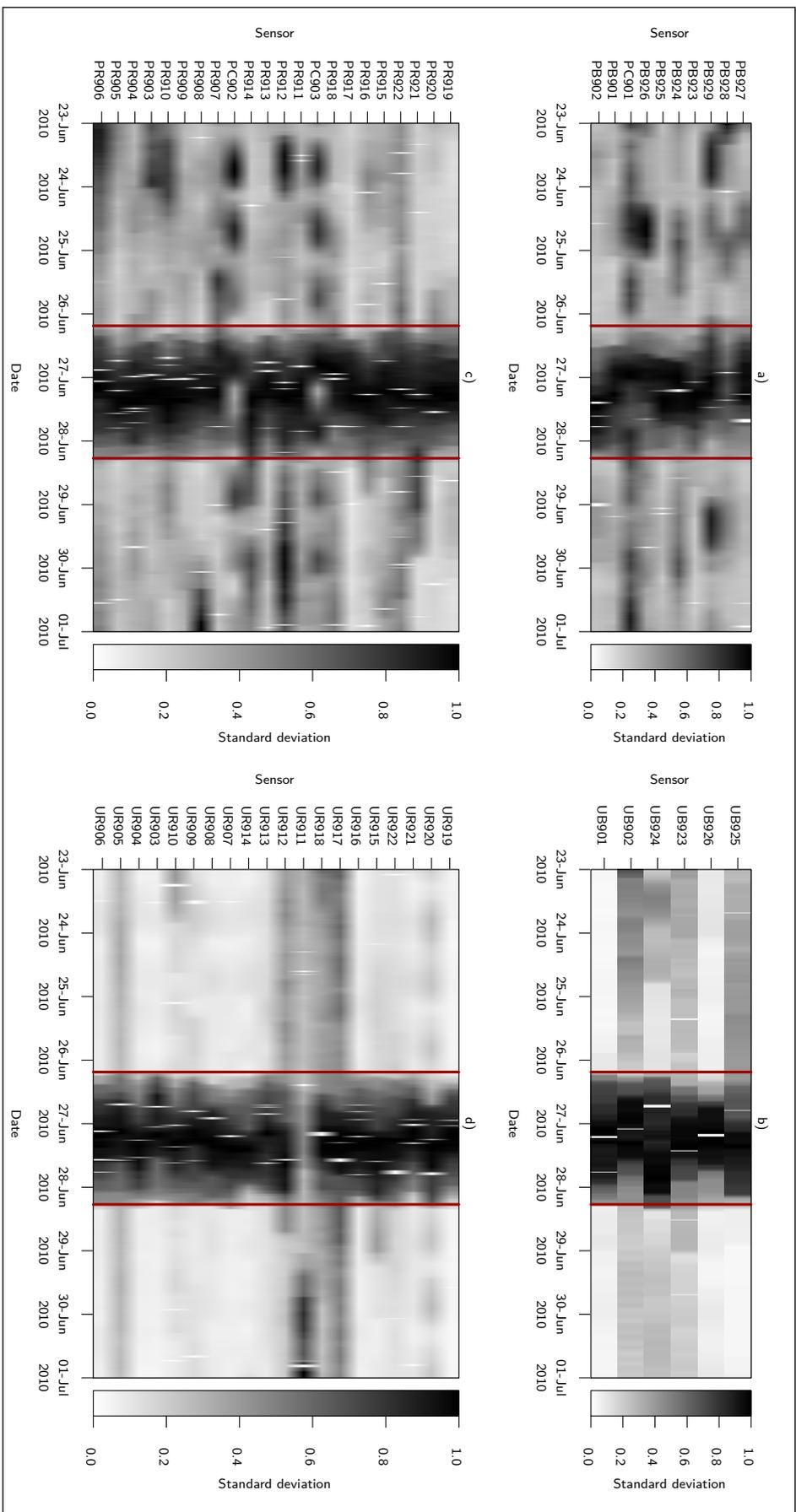


Figure 7.14: Intensity plot of a) axial stress sensor records, b) in-bentonite pore-water pressure sensor records, c) radial stress sensor records, and d) rock wall pore-water pressure sensor records between 2010-06-23 and 2010-07-01 (normalised SD values, 24 hour window).

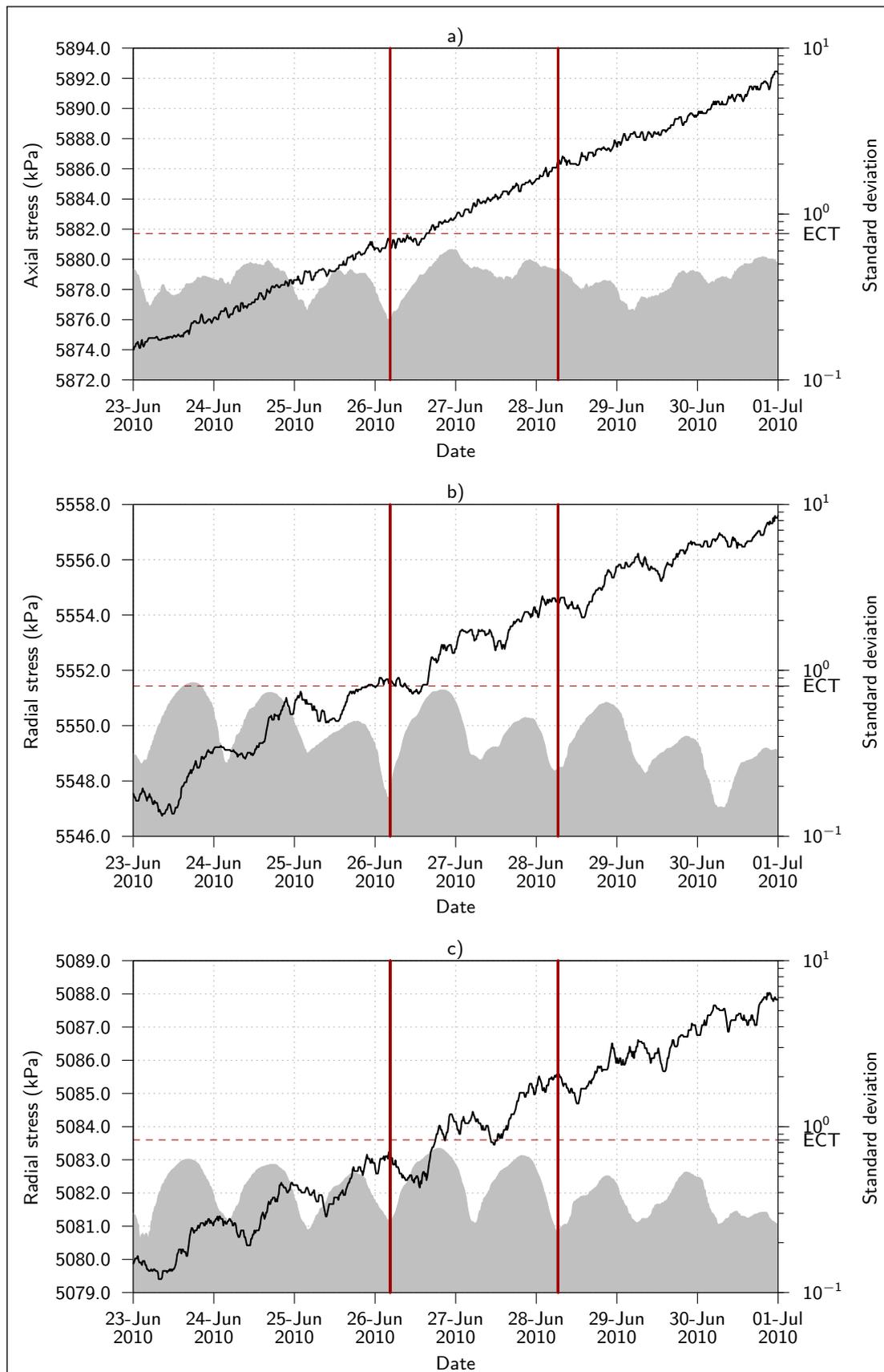


Figure 7.15: Detailed view of PC901, PC902, and PC903's sensor records from 2010-06-23 to 2010-07-01.

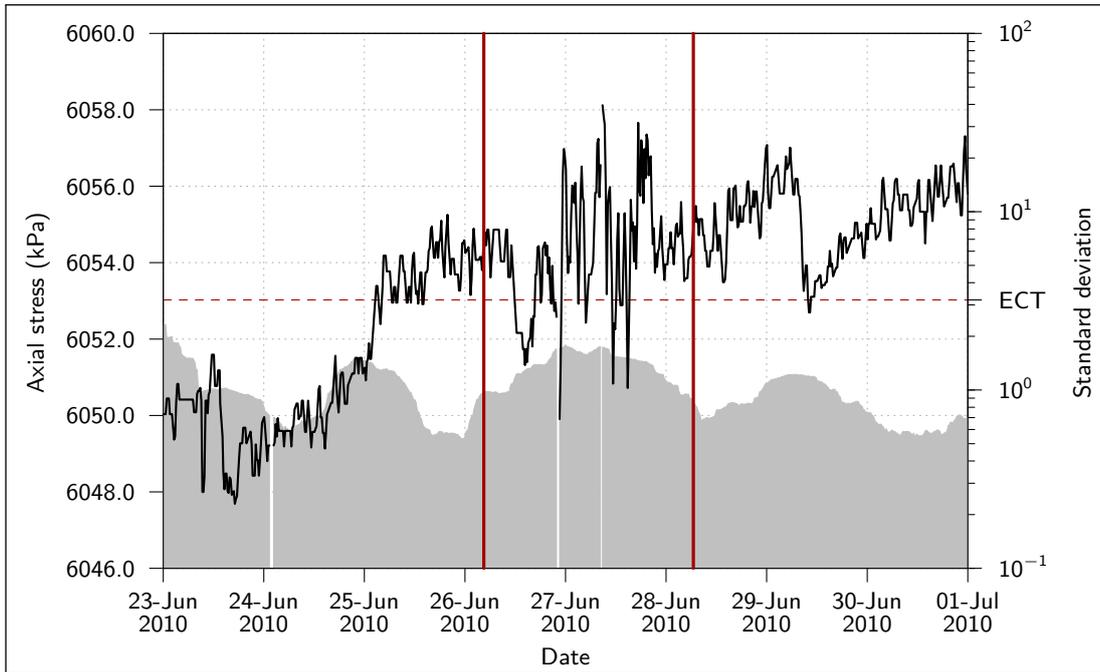


Figure 7.16: Detailed view of PB928's sensor records from 2010-06-23 to 2010-07-01.

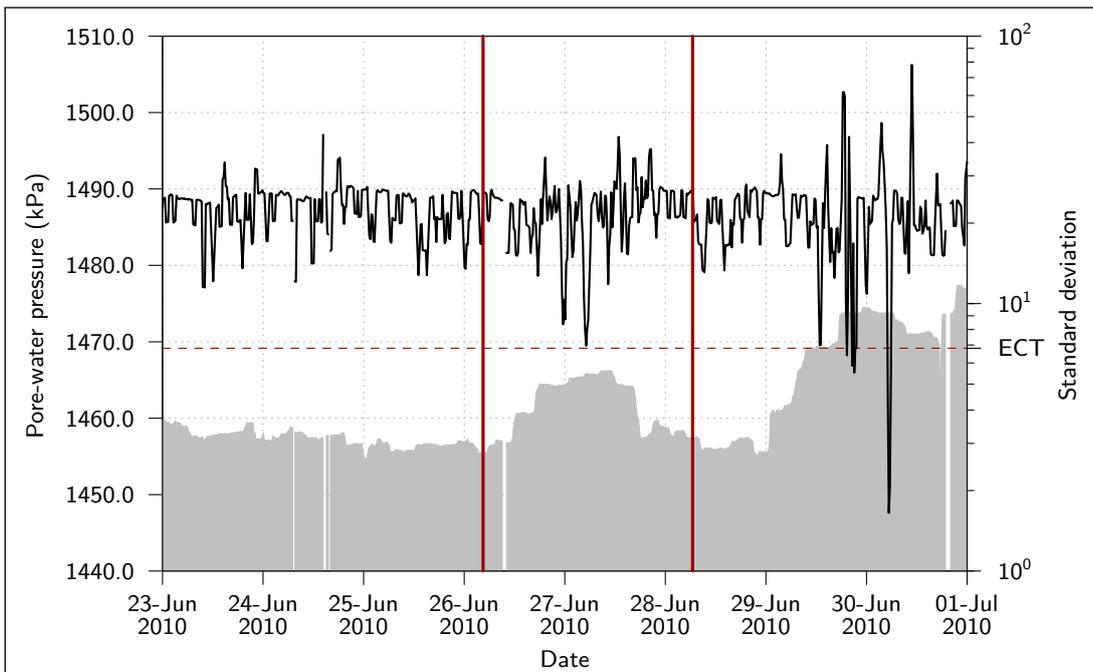


Figure 7.17: Detailed view of UR911's sensor records from 2010-06-23 to 2010-07-01.

7.5 Event at 2010–11–29

Figure 7.18 shows the pore-water pressure sensor UB901 record with the event candidate results of the corresponding de-trended residual sensor record. The absolute values of the event candidate intensity are used here, derived from a 48 hour window length. Impulses above the ECT are highlighted in darker grey. The time period corresponding to the event candidate highlighted by the arrow is approximately 2010–11–20 to 2010–12–08. This time period is presented in detail for a number of sensor records.

Figure 7.19 shows UB901's sensor record between 2010–11–20 and 2010–12–08 in detail. The event candidate intensity shown was derived using a 24 hour window. During this time, over a period of approximately six days, an increase of approximately 6 kPa occurred. Immediately before and after this transition phase, the sensor record possessed a constant downward trend of approximately 0.9 kPa day^{-1} , and a 'low' noise level (Standard Deviation (SD) of approximately 0.2 kPa).

Figures 7.20, 7.21, 7.22, and 7.23 shows the event candidate intensity in the specified time period for all in-bentonite pore-water pressure, axial stress, radial stress, and rock wall/bentonite interface pore-water pressure sensor records respectively. Figures 7.20 and 7.22 depict the absolute values of the SD, while Figures 7.21 and 7.23 depict the normalised values.

In addition to the seed event in UB901, an increase in intensity is present in UB902 (Figure 7.20). Figure 7.21 possesses localised increases in intensity in PB927, PB924, and PB902. Most prominently, in Figure 7.22, the sensor record of PR906 possesses a well synchronised and prominent event candidate. Other radial stress sensors with notable event candidates include PR917, and PR914. A marginal increase in intensity is present in PR912. In Figure 7.23 sensor records UR922 and UR912 possess a marginal increase in intensity at the beginning and end of the highlighted period respectively.

Figure 7.24 presents a detailed view of the sensor records arising from PR917, PR914, PR912, and PR906 (Subfigures a) to d) respectively). With the exception of UR912, the increase in SD noted in these sensor records all exceed the respective ECTs associated with each sensor record. A drop of approximately 6 kPa occurs at the beginning of the highlighted period in PR917. In contrast, an increase in noise during the highlighted period is present in PR914,

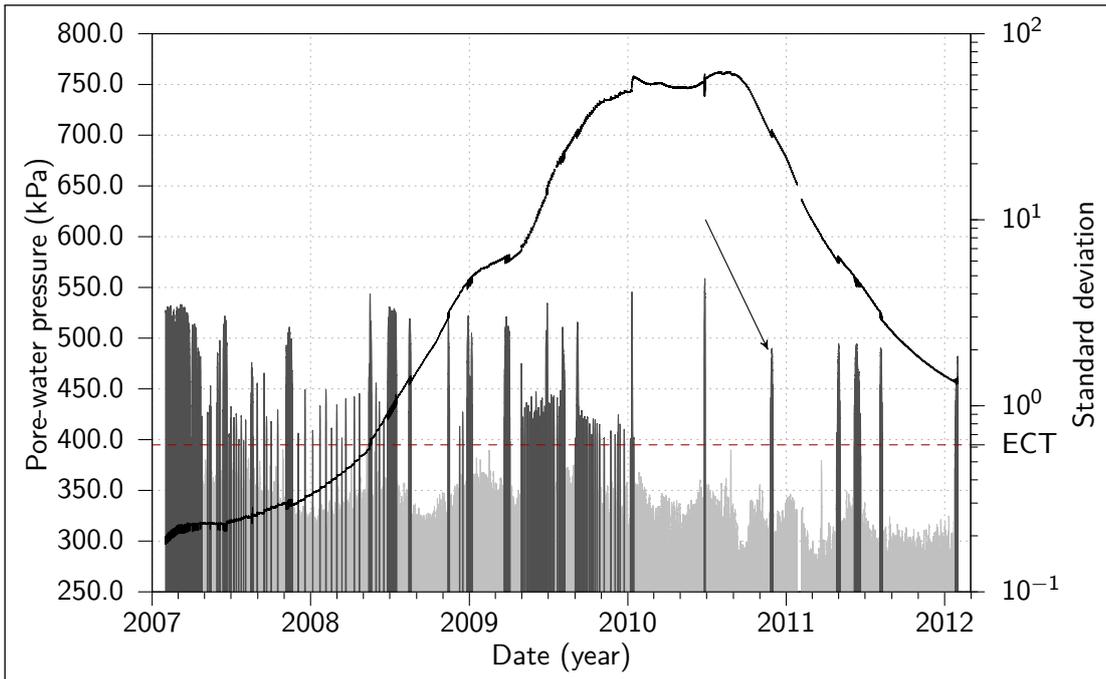


Figure 7.18: Pore-water pressure sensor record UB901 (black line), with the de-trended residual event candidate impulses (grey). Event candidates are highlighted in dark grey where they exceed the ECT.

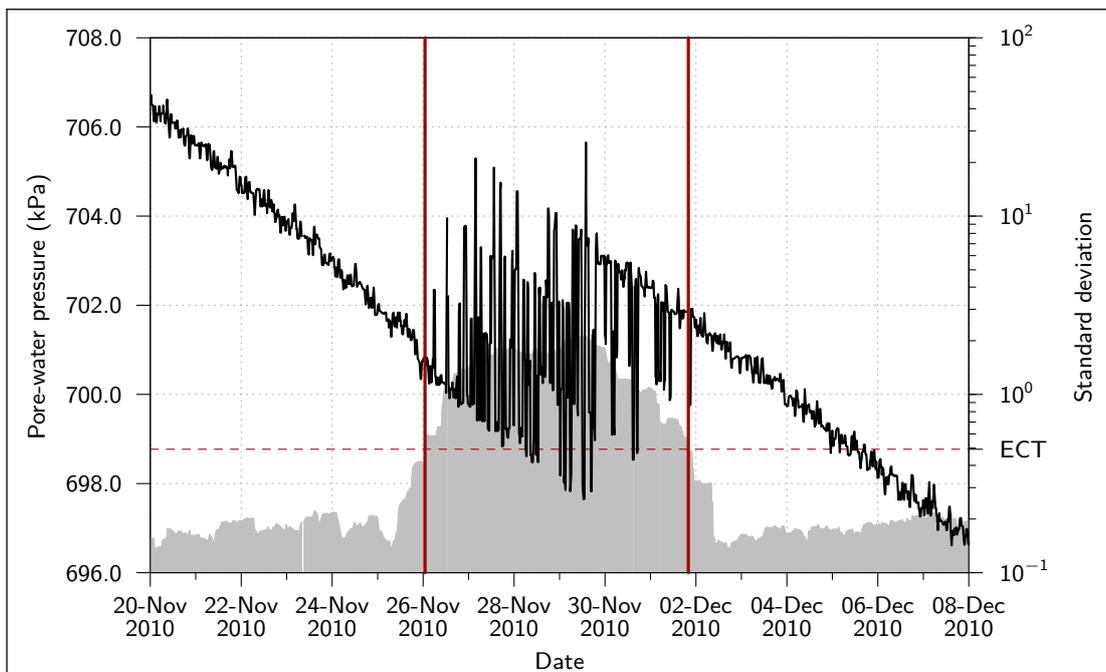


Figure 7.19: Detailed view of UB901's sensor record from 2010-11-20 to 2010-12-08.

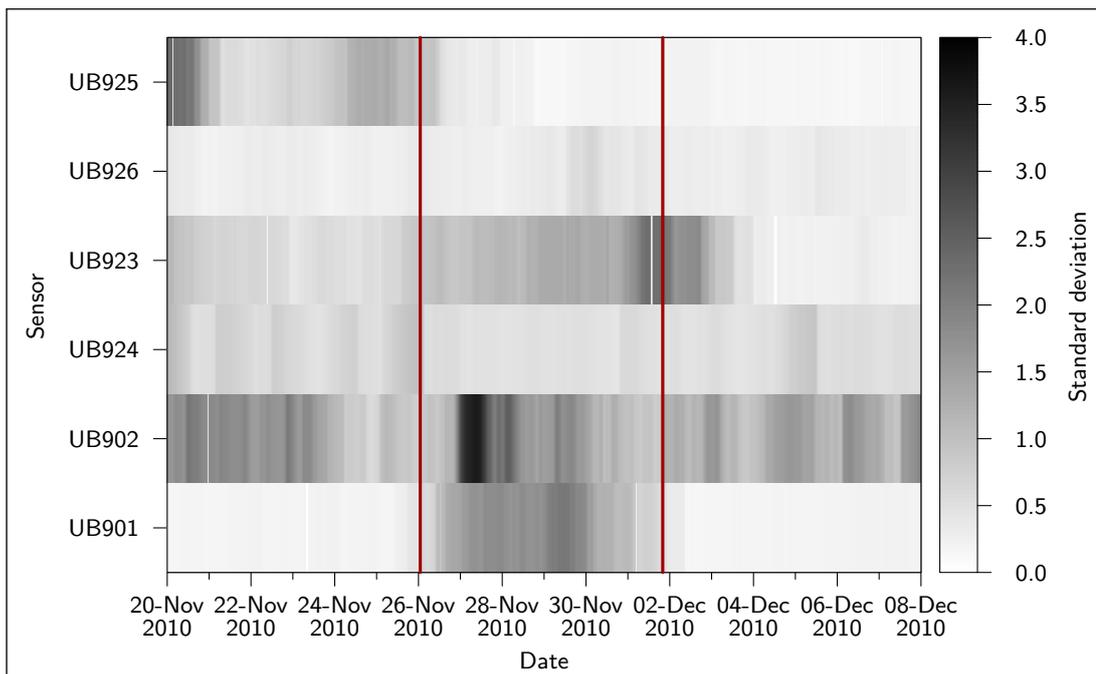


Figure 7.20: Intensity plot of in-bentonite pore-water pressure sensor records between 2010-11-20 and 2010-12-08 (absolute SD values, 24 hour window).

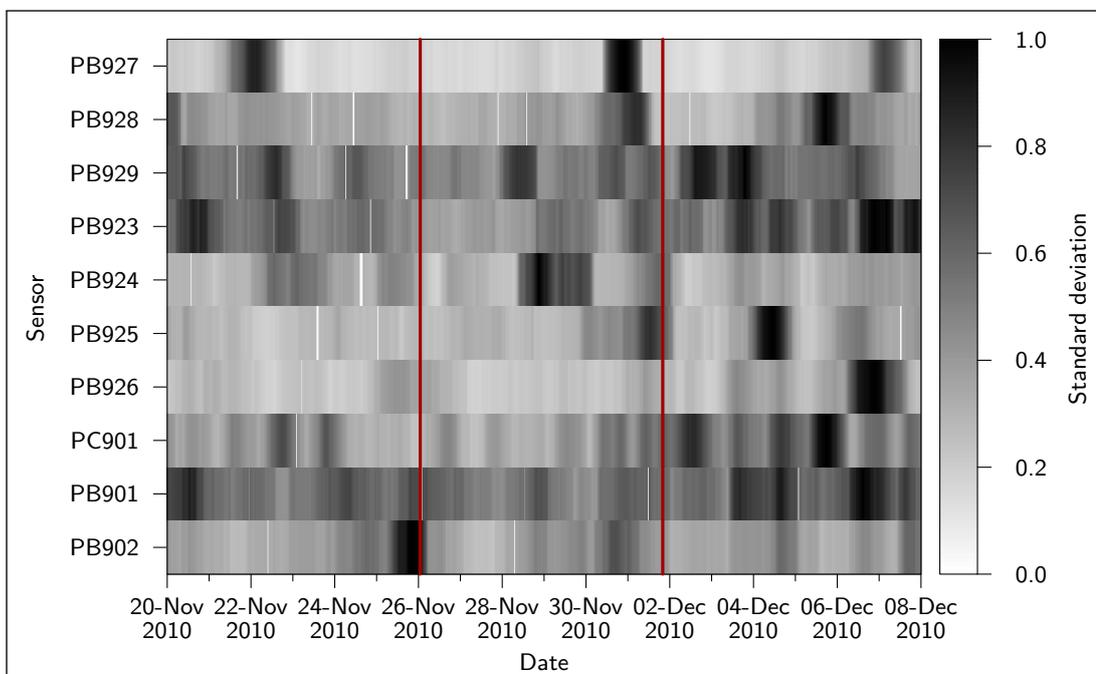


Figure 7.21: Intensity plot of axial stress sensor records between 2010-11-20 and 2010-12-08 (normalised SD values, 24 hour window).

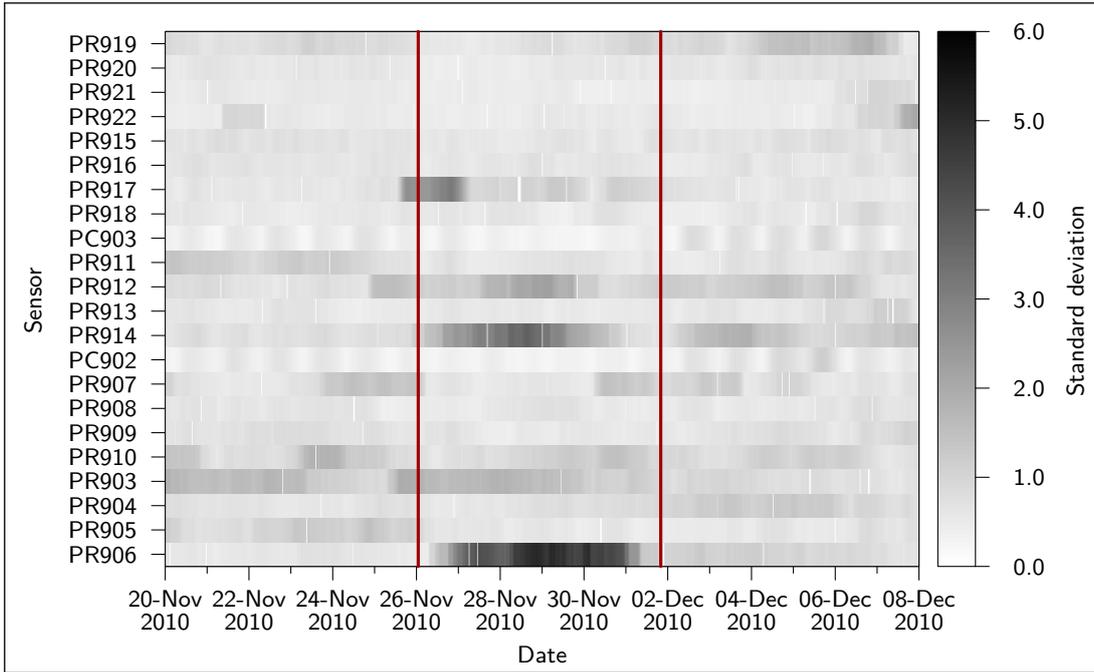


Figure 7.22: Intensity plot of radial stress sensor records between 2010–11–20 and 2010–12–08 (absolute SD values, 24 hour window).

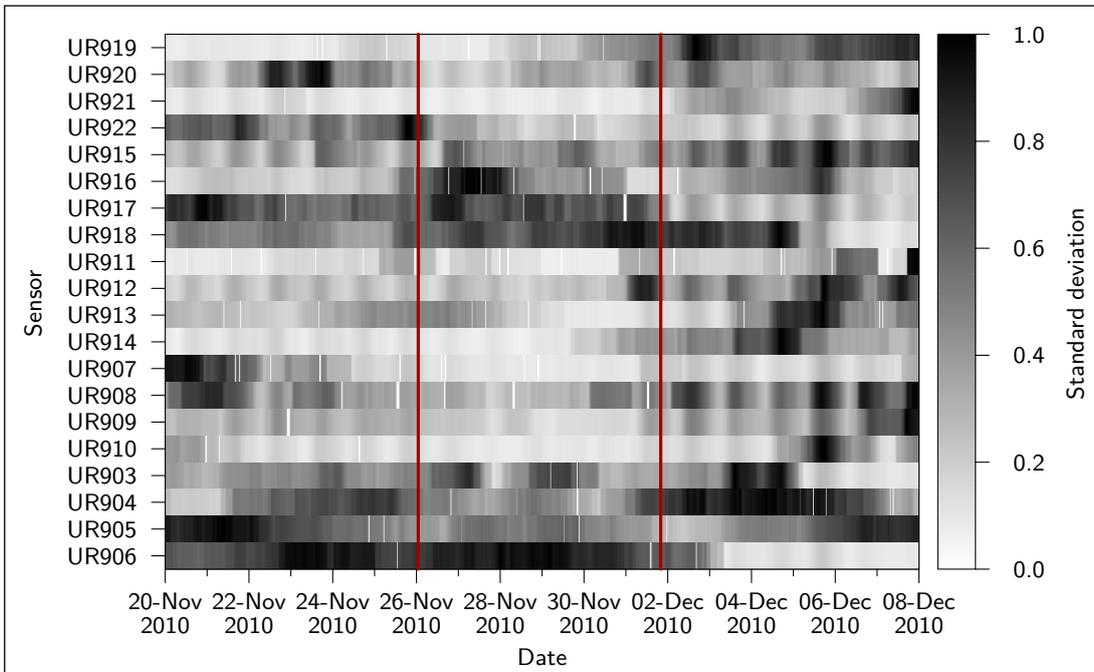


Figure 7.23: Intensity plot of rock wall/bentonite interface pore-water pressure sensor records between 2010–11–20 and 2010–12–08 (normalised SD values, 24 hour window).

PR912, and PR906. The increase in noise has positive influence on the magnitude in PR912, but a negative influence on the magnitude of PR914 and PR906.

Figure 7.25 presents a detailed view of the sensor records arising from PB927, PB924, and PB902's sensor records. The increase in SDs noted in these sensor records all exceed the respective ECTs associated with each sensor record. PB927 possesses an abrupt positive step in magnitude of approximately 10 kPa. It is noted that other similar changes in magnitude occur in PB927's sensor record in close proximity to, but not within, the highlighted period. PB924's sensor record possesses a drop of approximately 10 kPa during a short period of noise during the highlighted period. PB902's sensor record however, possesses an abrupt drop of approximately 4 kPa at the beginning of the highlighted time period.

Figure 7.25 presents a detailed view of the sensor records arising from UR922, UR912, UB902's sensor records. None of the sensor records depicted have an increase in SD that exceeds the ECT. An abrupt drop of approximately 5 kPa is present at the start of the highlighted period in UR922's sensor record, while a similarly sized increase is present in UR912's sensor record at the end of the highlighted period. UB902's sensor record possesses a perturbation of approximately 10 kPa in the positive that then dissipates to the original trend during the highlighted period.

7.6 Summary

Using NUDAT, Lasgit's dataset has been systematically analysed in a way that has not previously been attempted, i.e. on a second order scale, and while taking account of the non-uniformity in the sampling rate of the sensor records. NUDAT's event candidate detection process has enabled the rapid identification of sets of sensor records that possess synchronously occurring second order perturbations. In turn, this enables the specific sensor records relevant to an identified second order event to be aggregated, and the identified event to be investigated in detail. A detailed account of three such second order events identified by the application of NUDAT to Lasgit's dataset has been presented in the sections above.

The detailed discussion of the results presented in this chapter and Chapter Six, focusing on the inferences that can be made from the phenomena identified in the sensor records and the interpretation of second order events, is presented in Chapter Eight.

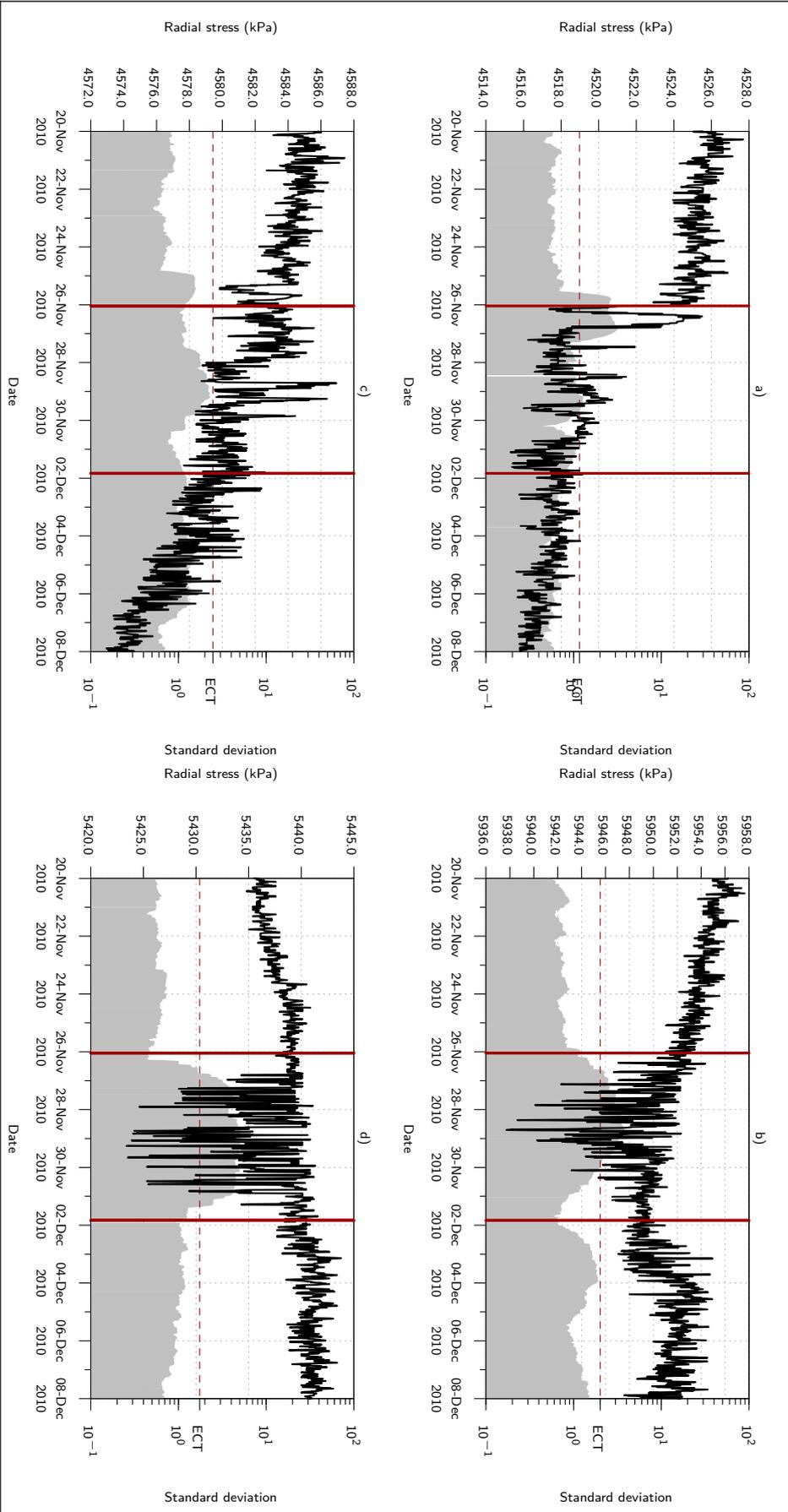


Figure 7.24: Detailed view of a) PR917, b) PR914, c) PR912, and d) PR906's sensor records from 2010-11-20 to 2010-12-08.

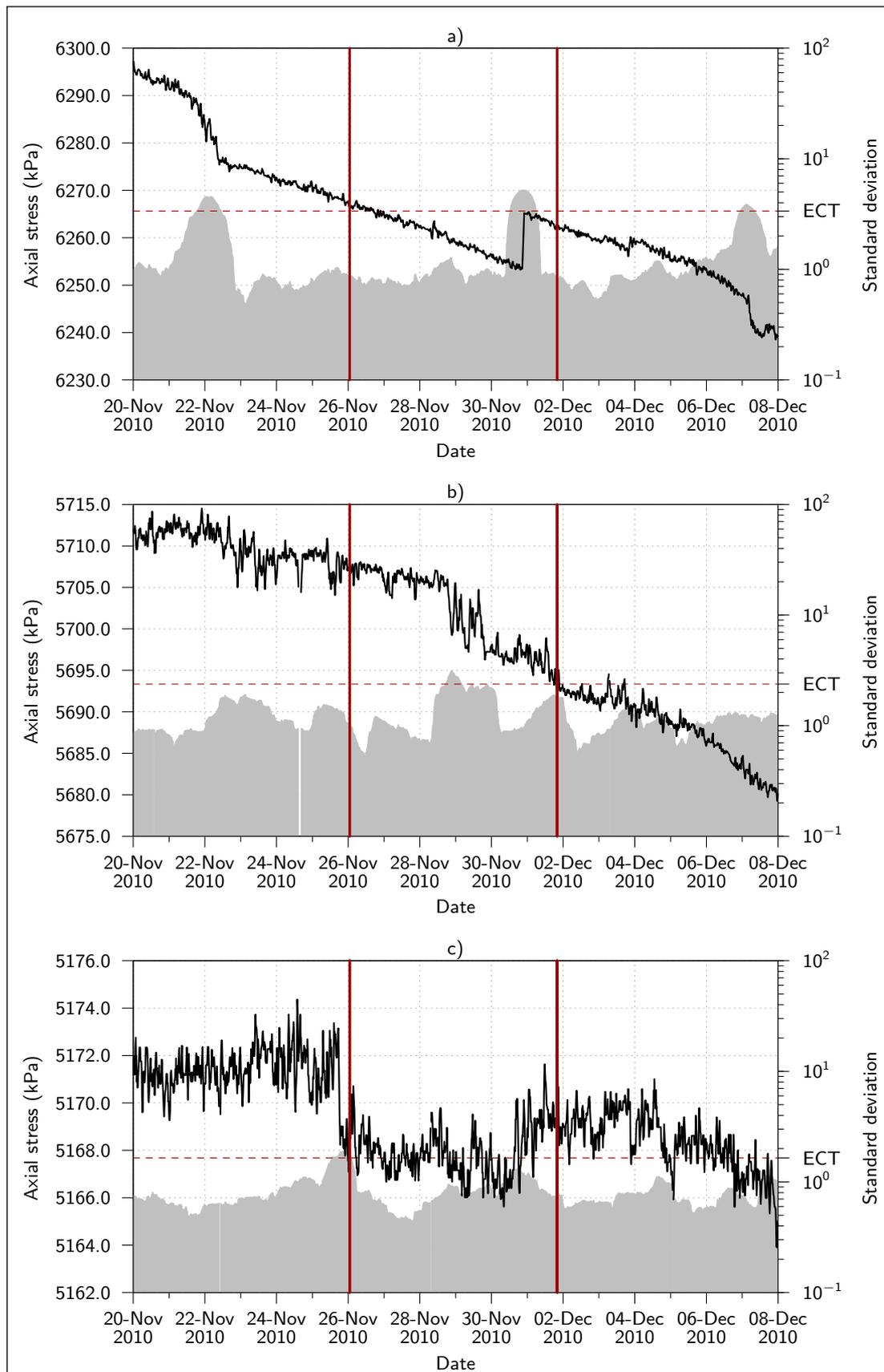


Figure 7.25: Detailed view of a) PB927, b) PB924, and c) PB902's sensor records from 2010-11-20 to 2010-12-08.

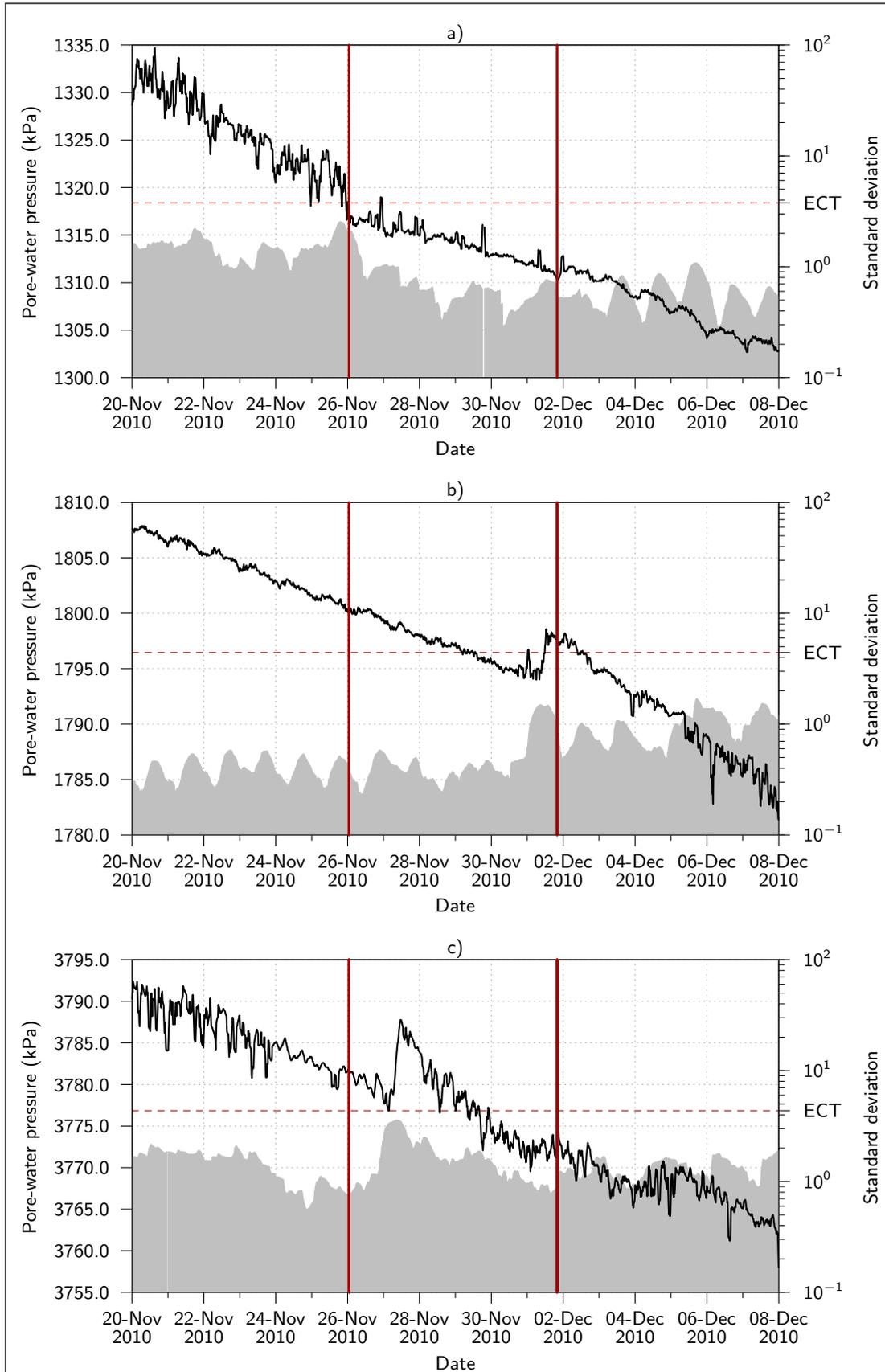


Figure 7.26: Detailed view of a) UR922, b) UR912, and c) UB902's sensor records from 2010-11-20 to 2010-12-08.

Chapter Eight

Discussion of toolkit outputs and events

8.1 Introduction

A discussion of the results obtained from the application of the Non-Uniform Data Analysis Toolkit (NUDAT) to the Large Scale Gas Injection Test's (Lasgit's) dataset as reported in Chapters Six and Seven is presented within this chapter. Sections 8.2 and 8.3 explore and interpret the physical implications of the toolkit outputs of each analytical process applied, and the behaviours observed in the detected second order events respectively.

The toolkit outputs discussed in Section 8.2 include the spike detection results, the Non-uniform Discrete Fourier Transform (NDFT) results, the aggregates of the event candidates, and the smoothing/moving average results. Where possible, speculation on the cause of detected/measured phenomena is made. Section 8.3 focuses on the potential physical interpretations relating to each set of perturbations identified as a second order event. The conclusions drawn from the interpretations made are intended to contribute to the overall understanding of the experimental program by providing example cases of behaviours of potential interest within a high level radioactive waste repository.

8.2 Discussion of the toolkit application results

As noted in the introduction, the following subsections discuss the macro-scale results obtained through application of NUDAT to Lasgit's dataset. This corresponds to the results that are presented in Chapter Six. While the toolkit application and results are geared towards second order information, the macro-scale implications of the toolkit outputs are considered in the following subsections by interpretation of the aggregated results within the context of the dataset as a whole, rather than focusing on specific second order occurrences.

8.2.1 Spike detection

The detection of a spike by NUDAT, on average, less than once in every one hundred datum points can be argued to be an indication of the infrequency of spike occurrence within the dataset, and is therefore of little consequence to remove. Alternatively, comparing the total number of spikes detected within the dataset (129,630) to the number of datum points arising from a single sensor record (120,113) implies that approximately one sensor's 'worth' of data has been lost to spiking. However, useful experimental results have been obtained despite the absence of data from a number of other sensors. Additionally, the equivalent loss of a sensor's worth of data in the situation where the losses are spread throughout the dataset rather than concentrated in one place is likely to impose a lower impact on the quality of the data. It can be reasoned therefore that the impact of the spike removal process on the quality of the data is minimal, while the impact on the quality of any process subsequently applied to that data is likely to be positive.

Consideration of the sensor records with the greatest and fewest number of spikes are notable for being associated with a sensor that is reporting unexpectedly low values (UR931), and a sensor with a relatively high magnitude range for its class (PT927), respectively. It is not possible though to correlate propensity to spike with magnitude range, due to the varying nature of the sensor records (trending, cyclic etc.). However, within DA3147G01, a slight negative correlation between spike occurrence and sensor depth exists, as shown in Figure 8.1. PC901, PC902, and PC903 are a visibly distinct locus in this figure.

It is unclear whether the clustering of spike occurrence distributions within broad sensor classifications (Figure 6.2) is an indicator of anything significant. Most prominently, the distinctions between the distributions suggest that the cause of spiking within the stress and pore-water pressure sensor records in DA3147G01 is independent from the cause in their corresponding integrated temperature sensor records. In turn, the sensor records arising from the 'canister associated' sensors (i.e. those sensors in physical contact with the canister), filter mats, or sensors not within DA3147G01 are clustered together, suggesting a further separation.

Of particular interest in the latter grouping are the stress sensors in contact with the canister (PC901, PC902, and PC903). These particular sensors are differentiated from the other stress

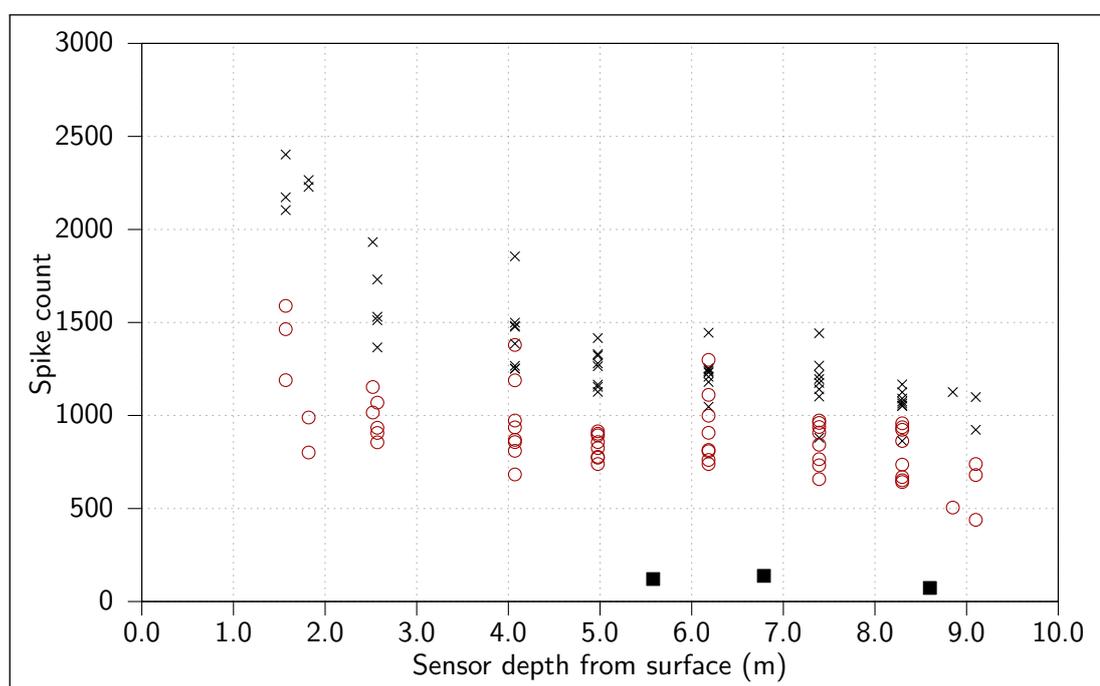


Figure 8.1: Sensor record spike count with depth within DA3147G01. 'Saltires' signify temperature sensors, 'circles' signify pore-water pressure and stress sensors, and squares signify canister stress sensors (PC sensors).

sensors within DA3147G01 by being in contact with the canister, and from the other sensors in the low spiking cluster by not being operator specified or for occurring outside of DA3147G01. This implies that the canister may have some influence over sensor behaviour.

8.2.2 Initial NDFTs

The initial NDFT results (before trend removal) essentially serve to confirm that the process is not of any quantitative use when there are significant trending components in the input dataset. However, departures from the typical profile of the frequency domain seen within any of the individual sensor types in Lasgit's dataset typically correspond to large anomalies in the particular sensor record in question. Notably, UT911 is an example of this feature. Inspection of the frequency domain for a dataset that contains trending components may therefore be of limited use in some cases to provide a basis for categorisation of behaviour.

Where there is no significant trending component to the input data the process provides quantitative results that align very closely with de-trended input after removal of the zero frequency component. This further serves to isolate the trending components as the complicating issue with respect to the frequency domain.

8.2.3 Event candidate detection

8.2.3.1 Axial stress sensors

The overview of the event candidate information for the axial stress sensors within DA3147G01 (Figure 6.7) shows that there are two sets of sensors that have notable synchronisation of responses, those located above the canister, and those below. Grouping of such synchronisations by proximity is unsurprising, however the presence of one major synchronised event across all sensors demonstrates that axial stress events may still occur globally to the experiment, although less frequently in comparison to the occurrences in the subsets of axial sensors above and below the canister.

8.2.3.2 Pore-water pressure sensors within the bentonite buffer

The lack of an obvious macro scale synchronisation in the in-bentonite pore-water pressure sensors (Figure 6.8) implies that perturbations in the pore-water pressure state are more localised than changes in the axial stress state in the system. This is likely due to the low permeability medium effectively isolating or partitioning the sensors from each other and from the boundary conditions over short time scales. The aberrant intensity noted in UB902 caused by a macro scale gas flow event demonstrates the effectiveness of this partitioning.

8.2.3.3 Injection filters and hydration mats

The macro scale synchronisation apparent in the injection filters and filter mats (Figure 6.9) is attributed to their nature, i.e. operator controlled points of pore-water or gas pressure. The gas injection test procedures (see Chapter Three, Harrington et al. (2008), and Cuss et al. (2010)) typically involved synchronous control of all or most injection filters and hydration mats, which would lead to prominent event candidates at those moments across the set of sensors. Conversely, the pressures in the transducers would be stably maintained by mechanical means (the pumps in the gas laboratory), leading to a notable absence of events between these moments.

8.2.3.4 Lid displacements and retaining loads

The sub-setting of the different classes of displacement sensors and retaining load sensors visible in Figure 6.10 are likely a result of the different measuring intent and orientations of the sensors. While the evolution of various measured displacements is not discernible from the event candidate information, a seasonality in the perturbations is implied. Greater event candidate intensity in the summer months may imply temperature influenced displacement.

8.2.3.5 Pore-water pressures in the pressure relief holes

The Pressure Relief Holes (PRHs) are, once sealed with packers, largely isolated from the experiment evolution, and hence more closely reflect the near field pore-water pressure conditions. There are fewer physical processes occurring within the PRHs, with the measured parameter

(the pore-water pressure) being influenced from outside the PRHs. As such the synchronisation between event candidates across the pore-water pressure sensor records is not unexpected. The lack of synchronisation between UR931 and the remaining PRH pore-water pressure sensors appears to be as a result of the sensor malfunctioning, or reporting an ambient pressure in a non-sealed environment, possibly exposed to atmosphere.

8.2.3.6 Radial stress sensors

The radial stress sensors in DA3147G01, with the exception of PR921, are the group of sensors with the greatest visual similarity to one another throughout the lifetime of the experiment (see Figure 3.9). As with the axial stress sensors, there are a number of localised synchronisations, along with more global synchronisations between event candidates. Unlike the axial stress sensors however, none of the sensors in question are located above the top or below the bottom of the canister. This proximity may contribute to the slightly more uniform nature of Figure 6.13.

8.2.3.7 Pore-water pressures at rock wall/bentonite interface

The pore-water pressure sensors at the bentonite/rock wall interface can be interpreted in the same way as the radial stress sensors. The similar elevations and uniformity again point to the proximity of the sensors producing a more uniform event candidate intensity plot. However, in contrast with the in-bentonite pore-water pressure sensors, both a localised and a more global event candidate synchronisation exist, suggesting that the sensor contact with the rock wall bypasses the isolation of the bentonite. This assertion is in agreement with the observations made regarding the sensors in the PRHs. The lack of synchronisation across the full sensor set at the two most prominent global candidates suggests the possibility of a complexity in the boundary condition of the experiment.

8.2.3.8 Temperature sensors

The result of the temperature sensor event candidate detection process (depicted in the event candidate intensity plot in Figure 6.19) demonstrates a high degree of synchronisation between almost all of the 57 sensor records depicted. As noted in Section 6.4.3, there is a sub-grouping

of the sensors located above, and adjacent to and below the canister. A number of the lineations depicting synchronisation are common to these sub-groupings however. Examples include the abrupt increases and reductions in noise levels. Additionally, after the reduction in noise levels mid 2010 there appears to be a slightly deeper seasonality depicted, with an appreciable lag increasing with depth.

8.2.4 Discussion of smoothing results

There is no attributable cause apparent in the rest of the dataset for the temperature shifts recorded in UT901 (Figure 6.23). Such a systematic aberration, in the absence of other phenomena that coincide or synchronise with the steps, suggests a sensor or logging equipment error. As seasonal temperature changes of between approximately $\pm 0.5^{\circ}\text{C}$ and $\pm 1.0^{\circ}\text{C}$ are typical within the down-hole temperature records, the error introduced by suspected malfunctions of this magnitude are likely to be insignificant. The reduction in temperature sensor records noise that occurred mid 2010 is also visible in the sensor record, shown by the original data (grey) matching more closely the overlain smoothed data (black). As with the systematic aberration revealed by the smoothing process, there is no attributable cause that is apparent in the rest of the dataset.

8.2.5 Discussion of the SSA derived trends

The majority of sensor records were determined to possess only one macro scale Singular Spectrum Analysis (SSA) trending component after decomposition. This lack of detection of more than one long term trend in many of the sensor records may be considered an absence of evidence for multiple long term influences being present in the system. However, it is also possible that the SSA process has subsumed one or more existing trends into a single trend estimation in the majority of the sensor records. Since absence of evidence cannot be considered to be evidence of absence, no assertion about the number of influences on the long term/macro-scale trends can be made. As such the SSA derived trends are considered solely as representative estimates of any aggregated trends within the dataset.

8.2.6 De-trended NDFTs

The majority of the de-trended sensor records present a significant and well defined annual peak in the frequency domain. Of those that have been highlighted in Chapter Six as more ambiguous, many still possess an annual phase and amplitude measurement that has a strong visual correlation to the de-trended signal. Examples include sensors PC901, PR906, PB901, and FR901, all of which are depicted in Figure 8.2 with their respective calculated annual cycles overlain.

Exceptions to this looser criteria for annual frequency determination include sensors PB927, UR916, UR907, and UR910. Figure 8.3 depicts the derived annual frequency content along with the input data. The increased presence of distortions and steps in the input data is cited as the cause for the absence of useful frequency domain information.

Marginal correlations between the NDFT derived annual frequency content and the input data are achieved in the majority of the PRH sensors. Figure 8.4 provides a visual example for UR931. Additionally, the emergence of a measurable daily frequency component in the de-trended PRH sensor records indicates that trend removal is a useful and necessary process where trending exists.

The correlation in temperature sensor frequency content (amplitude and phase) with depth is very strong within DA3147G01. The more complex 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. The modified NDFT procedure therefore, appears to be finely sensitive to the experimental variations/distortions in the input data that are introduced by non-cyclic processes.

Estimating the height within DA3147G01 of the canister temperature sensor (TC901) from its measured annual amplitude and phase offset in relation to the correlations depicted in Figures 6.26 and 6.27 theoretically places it at approximately 5 m depth from surface.

Notably, three pore-water pressure sensors located at the rock wall/bentonite interface express relatively larger annual cycle amplitudes than other similar sensors. These sensors are highlighted

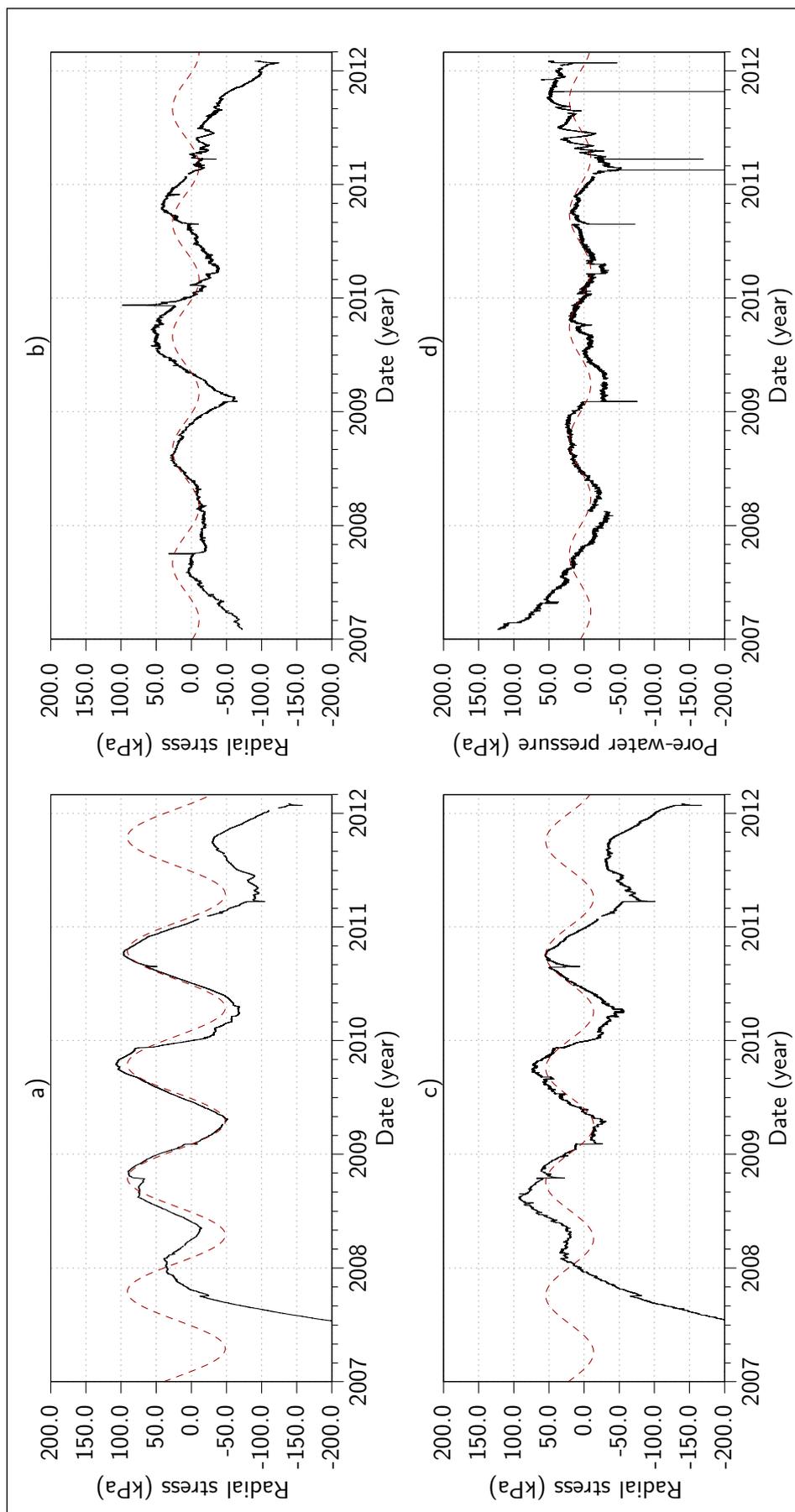


Figure 8.2: Examples of the sensor records with good visual correlation to ambiguous annual frequency content. a) PC901. b) PR906. c) PB901. d) FR901.

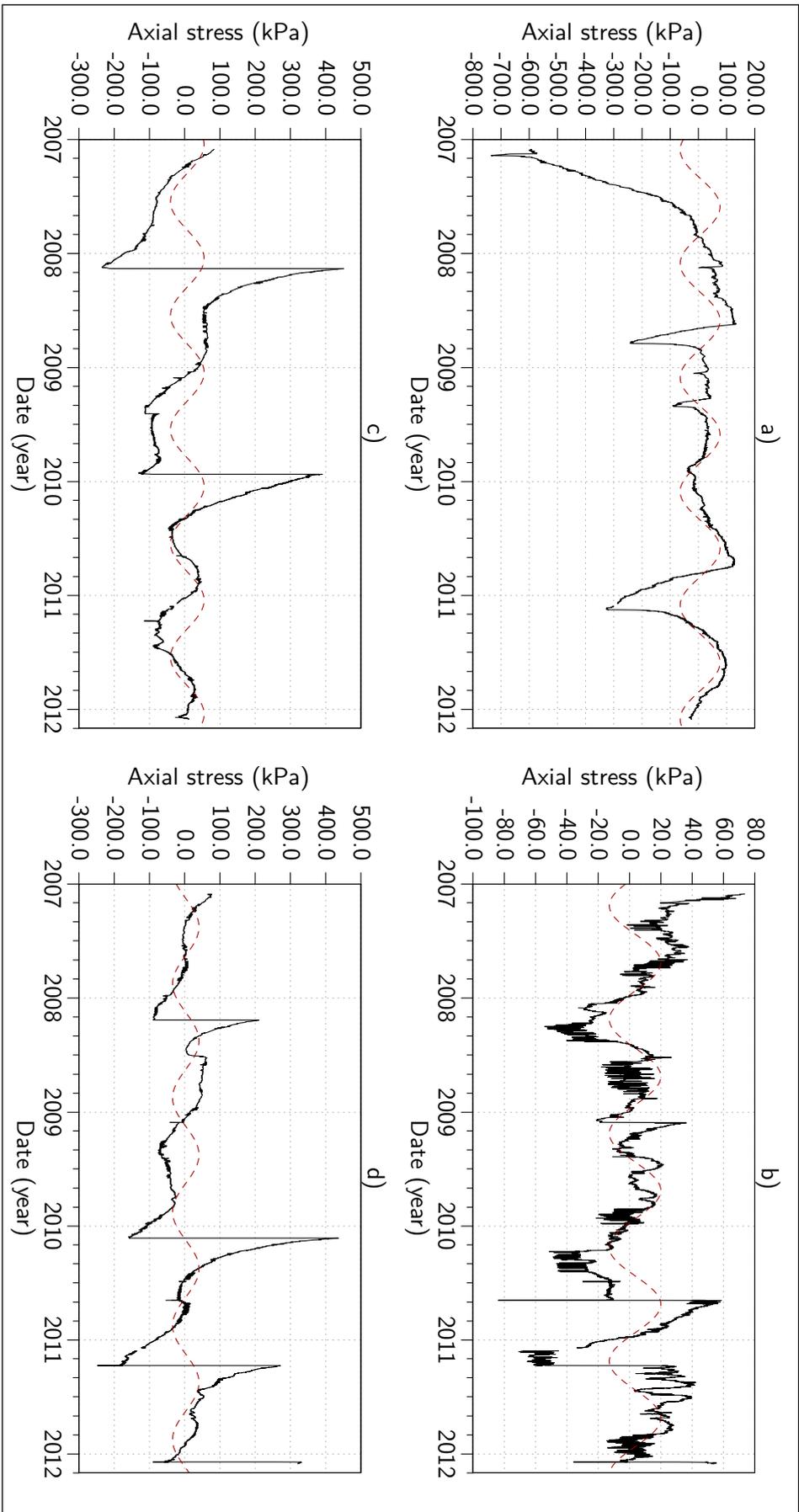


Figure 8.3: Examples of the sensor records with good visual correlation to ambiguous annual frequency content. a) PB927. b) UR916. c) UR907. d) UR910.

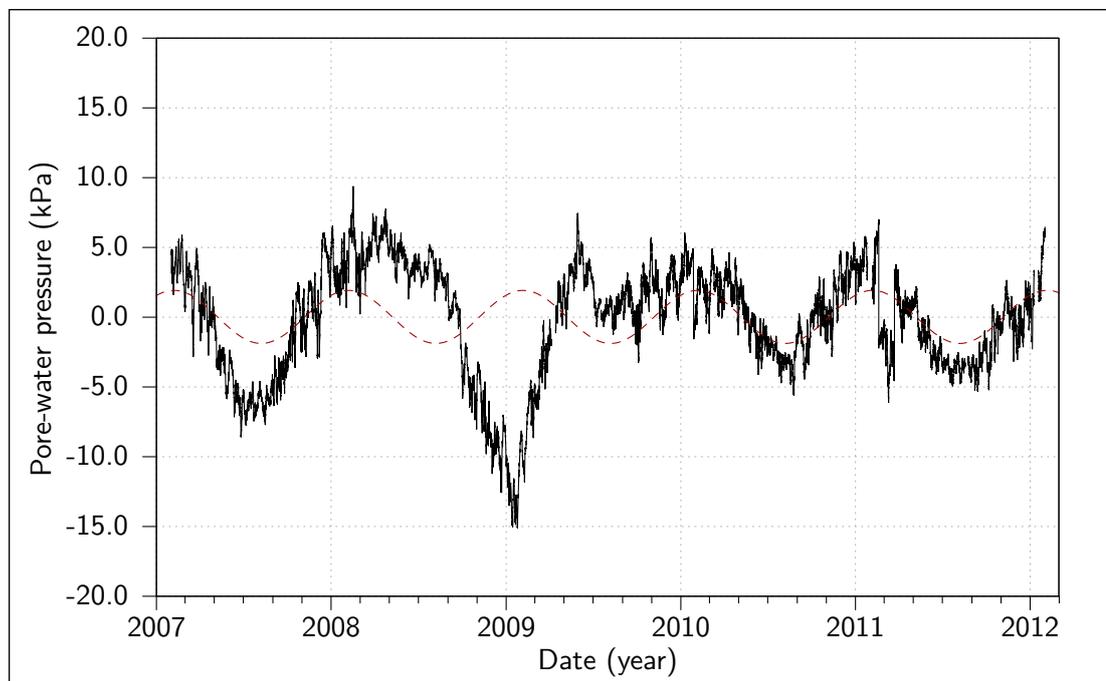


Figure 8.4: UR931's visual correlation to ambiguous annual frequency content.

in Figure 8.5. While annual frequency content has been measured in most of the rock wall/bentonite interface pore-water pressure sensors (excluding those noted in Figure 8.3), the seasonality is more visible in UR920, UR914, and UR912, even before de-trending, as shown in Figure 8.6.

Figure 8.7 depicts the approximate locations of sensors UR912, UR914, and UR920 with respect to the fractures in the rock wall, while Figure 8.8 depicts their relative positions within DA3147G01. These sensors are either located on or in close proximity to recorded fractures, however they are not unique in their intersections with such boundary conditions. For example, sensor UR911 is located at the same height as UR912 and UR914, lies on the vertical 'A', and hence intersects a major fracture feature. UR911 does not however possess the comparatively large (in relation to other rock wall pore-water pressure sensors) annual cycles of UR912, UR914, and UR920.

The modified NDFT process appears to be able to detect and enumerate accurately the frequency content in the input time series data when distortions in the data are minimal, but also provide a useful approximate enumeration of major frequency components in most cases when minor distortions exist. Such enumerations have permitted a speculation on an unknown parameter (the elevation of TC901) within the experimental set-up that can be confirmed when decommissioning takes place, along with highlighting similarities in anomalous behaviour that cannot be firmly correlated to boundary conditions.

8.2.7 Comparison of toolkit outputs before and after de-trending

The most notable difference when comparing toolkit output between trend possessing and de-trended input is the improvement in frequency domain quantification. In the case of Lasgit, a large number of annual cycles, and a small number of daily cycles have been effectively quantified as a result of removing the trend from the input data. In contrast, the macro scale observable effects of trend removal are limited however. As stated in Section 6.4.7, in most cases the difference between the event candidate intensity plots before and after trend removal is negligible when observed at the macro scale. Additionally, the removal of the trending components appeared to have no influence on the smoothing/moving average function.

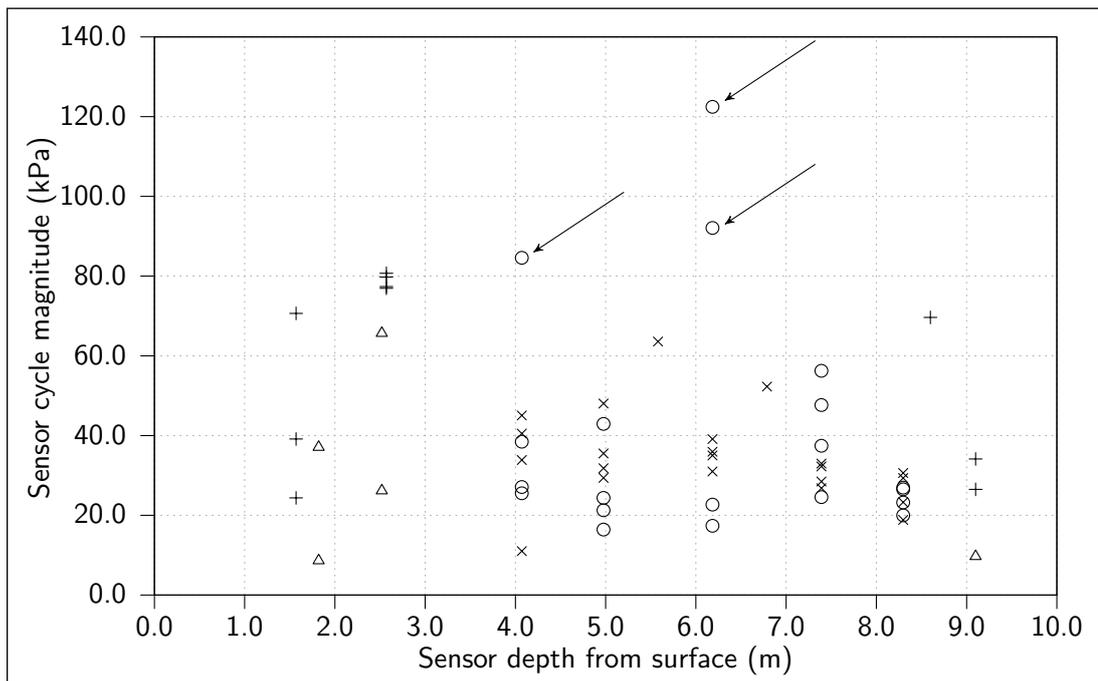


Figure 8.5: Pore-water pressure sensors with comparatively large annual cycles.

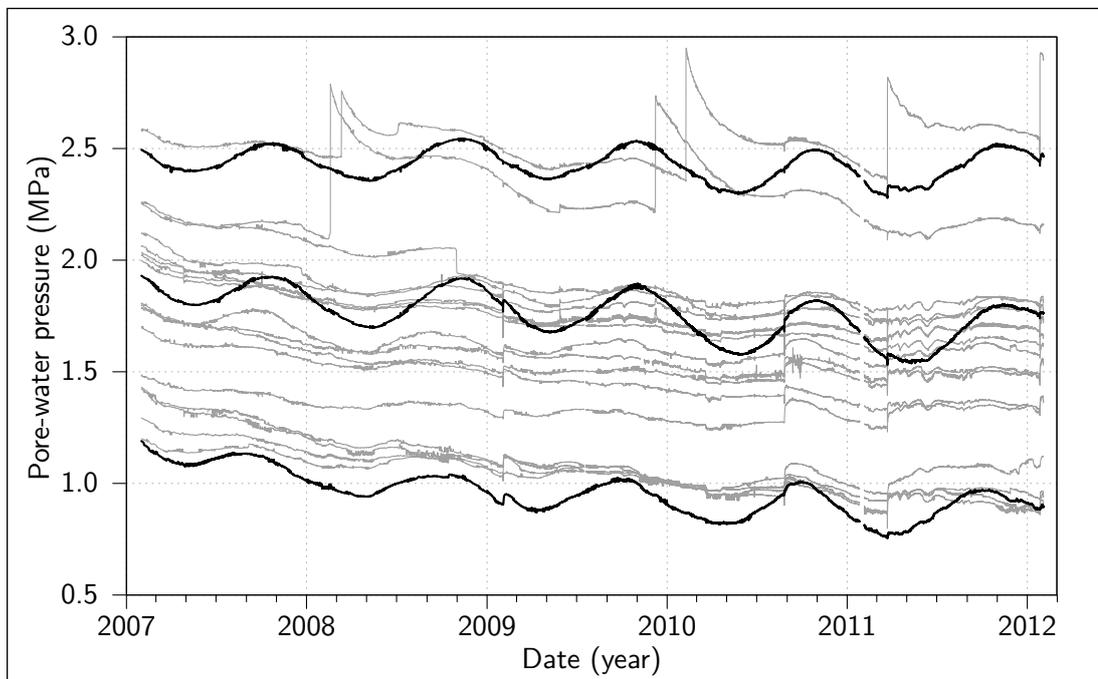


Figure 8.6: Comparatively large annual cycles in pore-water pressure sensors at rock wall/bentonite interface in DA3147G01. Individual UR sensors in grey, UR920, UR914, and UR912 are highlighted in black.

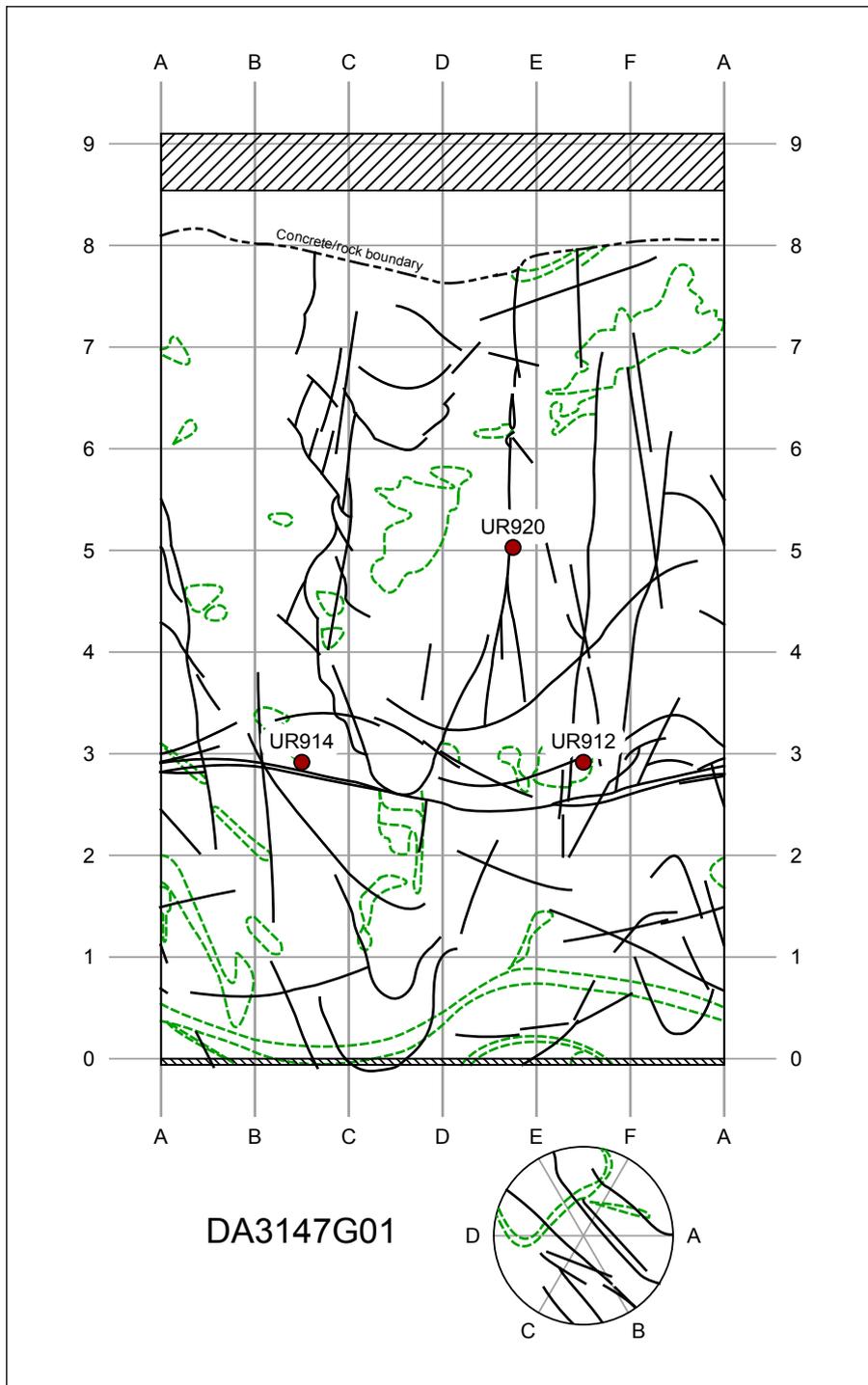


Figure 8.7: Locations of sensors with comparatively large annual cycles in relation to deposition hole fracture map.

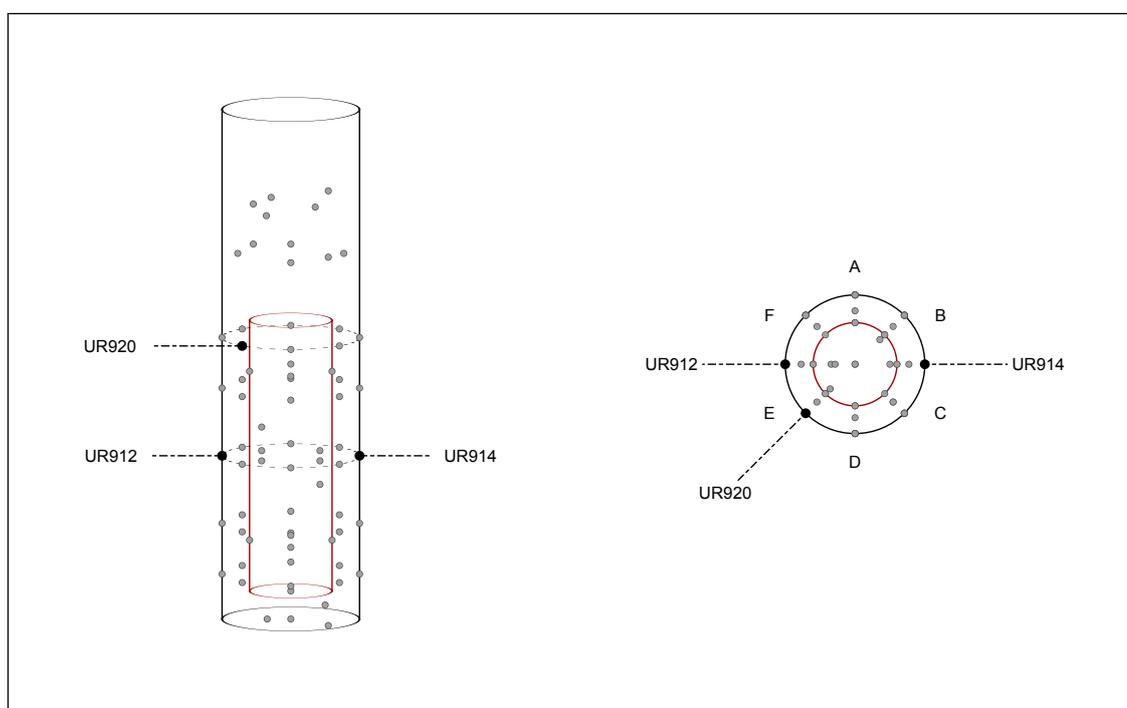


Figure 8.8: Locations of sensors with comparatively large annual cycles in relation to deposition hole arrangement.

8.3 Discussion of the second order events detected

8.3.1 Event at 2010–01–10

The relative physical locations of the sensors relevant to the event identified on 2010–01–10 are shown schematically in Figure 8.9. Arrows ① and ② in Subfigure 8.9a indicate the direction¹ of the preceding macro scale gas flow from the injection point FL903 to FL901 and on to UB902, as noted in Section 3.3. The ⊕ and ⊖ symbols in Subfigure 8.9b indicate the signs of the steps identified in the sensor records depicted in Section 7.3.

Interpreting the collection of events within the paradigm of a dilatant gas flow mechanism (i.e. where the gas pressure itself causes flow pathways to mechanically form, or 'dilate', through the bentonite Engineered Barrier System (EBS)) would suggest a short lived gas flow pathway formed between sensors UB902 and UB901, and a small flow occurred, originating from the volume of gas contained in a previously formed flow pathway between FL901 and UB902, subsequently arriving at UB901. The increases in stress present in PC901 and PB901 suggests that this pathway was mechanically formed and gas passed between them. The lack of second order event in filter FL903 suggests that this hypothetical flow pathway between FL901 and UB902 was isolated from FL903.

Dilatant flow mechanisms such as that described above are typically identified (in part) in laboratory test conditions by a characteristic gas flow rate into a clay system during pathway formation (e.g. Cuss et al. 2011; Harrington and Horseman 2003, 1999). As the apparent gas source for the potential second order gas flow event of interest is the reservoir of gas contained within the previous flow path from FL901 to UB902, and hence not connected to a flow measurement device, this aspect of flow mechanism/pathway formation conformation is not present.

It is also possible that flow from the gas reservoir between FL901 and UB902 migrated to PC901 and PB901 in addition to UB901. The absence of a detected second order flow migration to other proximate sensors, such as PB902, would indicate localised flow pathway(s). This interpretation does not consider or dismiss the possibility of a mechanical induced pathway.

¹Flow paths between each sensor are indeterminate. Arrows are indicative only of start and end points of flow paths.

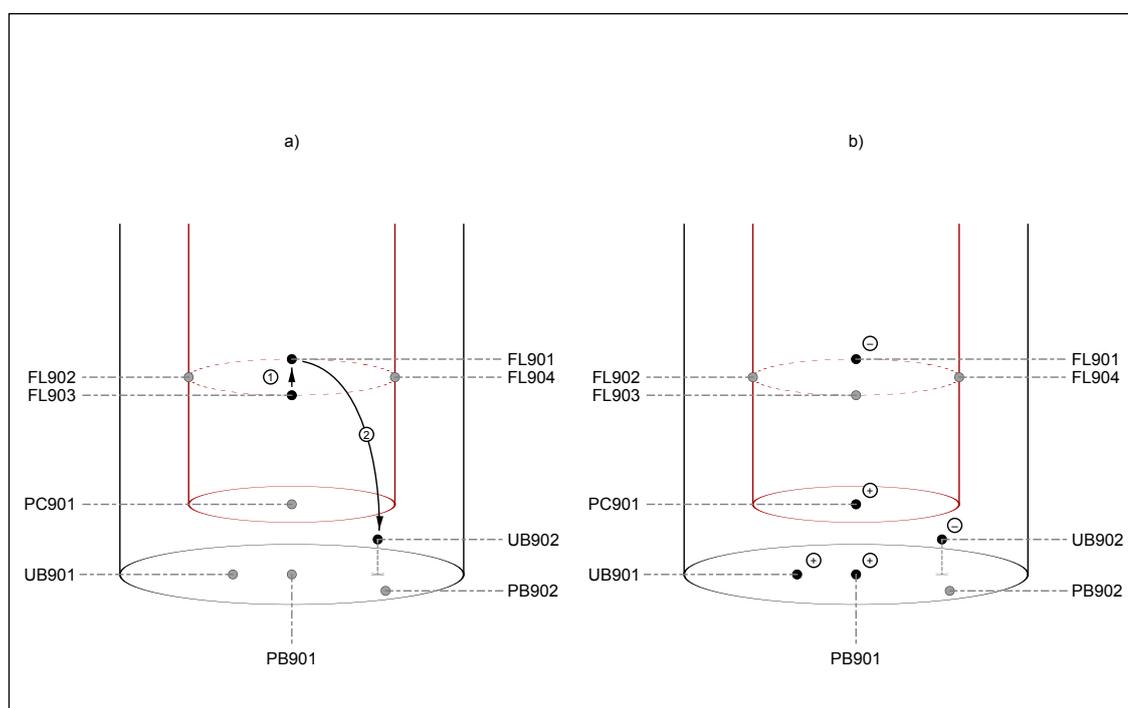


Figure 8.9: Spatial orientation of sensors identified as part of event on 2010-01-10. a) Macro scale gas flow from FL903 (injection filter) to FL901 and on to UB902. b) Locations and signs of steps in sensor records.

By further exploring the isolated nature of the flow, it is noted that a number of preferential flow pathways potentially exist within DA3147G01 due to the pipework serving the instrumentation. However, according to the information in Johannesson (2004) there is no such potential pathway that could provide a complete route between UB902 and the sensors that reported increases in measured stress/pore-water pressure. Similarly, there is no single common interface between the canister/bentonite/rock wall that unites the relevant potential engineered preferential flow pathways while also excluding the potential engineered pathway to PB902. This implies that either: i) the apparent flow event exploited two separate interfaces simultaneously; ii) that a single exploited interface, in this case the bentonite/rock wall interface, provided tortuous flow routes to multiple destinations, but that the engineered pathway to PB902 was bypassed or not viable; or iii) that a localised flow through the bentonite occurred.

Additionally, it has thus far been assumed that the flowing fluid in these events is a discrete gaseous phase. Pore-water flow in a porous media, such as the bentonite EBS in question, will move approximately uniformly through the pore space of an approximately homogeneous material. The localisation of the apparent flows would require preferential pathways of significantly lower permeability to enable such non-uniformity in pore-water flows. The swelling pressure developed in the bentonite EBS suggests that significant water infiltration has occurred (Cuss et al. 2010), which will have induced sealing and reduced the effectiveness for water exploitation of preferential pathways within the bentonite EBS, and hence made pore-water flow of this nature unlikely. Flow events of the immediacy observed during this event are therefore more likely to consist of the injected gas than pore-water.

With respect to the scale of the events investigated here, while the second order event that comprised the initial observation was visible in the macro scale plot of the sensor record (UB901), the events identified as synchronous were not visually discernible at macro scale visualisations. The identified second order events possessed magnitudes as many as 10^3 times smaller than the macro scale magnitudes of the sensor records they were contained within.

8.3.2 Event at 2010–06–27

Comparison of the sensors that express the phenomenon noted in the time period highlighted in Figures 7.13 to 7.17 with those that do not, re-enforces a number of the potential sub-grouping

of sensors previously inferred. The event is notable for only occurring within DA3147G01, i.e. the phenomenon is absent in the PRH sensor records, displacement sensor records, and load sensor records. Additionally, within DA3147G01, the phenomenon is absent in the temperature sensor records and the injection filters/filter mat sensor records. While casual observation suggests that the event is ubiquitously present in the remaining sensor records arising from DA3147G01, the absence of any perturbations in the stress sensors on the canister surface refutes this, and again segregates those sensors from the other stress sensors.

The event in question happened approximately seven days after Gas Test Two's shut in phase was stopped and a post gas injection hydraulic test was initiated. Coupled with the presence of the phenomenon in both the total stress and pore-water pressure sensor records, this suggests that the event may be a pore-water pressure perturbation. The potential source of such a perturbation is uncertain however. There is an absence of a causative impulse recorded the injection filter or filter mat sensor records, discounting a surge or similar from the gas laboratory. Additionally, while an external influence on pore-water pressure in the facility may not be present in the injection filter and hydration mat sensor records due to the gas laboratory pumps ability to provide active pressure feedback, there is no perturbation in the PRH sensor records. Additionally, it would be difficult to reconcile the evidence for the in bentonite pore-water pressure sensors being largely isolated from one another with an apparent ability for a second order pore-water event to permeate the system as a whole.

Alternatively, a perturbation in the stress state of the system may have been transmitted to the pore-water pressure sensors through the stress bearing capacity of pore-water in a medium such as compacted bentonite. Such a perturbation could plausibly be confined to DA3147G01. The omni-directional response in the stress sensors (i.e. radially and axially) may point towards the perturbation being transmitted through the non-structural 'free' water within the EBS in this case, as opposed to the interlayer water.

Such a system wide perturbation in stress, pore-water pressure, or both simultaneously, could potentially be caused by changes in the position of the retaining lid, or a change in the retaining load effecting mechanical changes in the bentonite below. No atypical perturbation in the displacement and load sensor records exist during the event window, however.

It is also possible that the recorded perturbations arise somehow from an unknown aspect of

the logging system. This possibility is aligned with the concept of the 'classes' of sensor that were implied in the spike occurrence clustering data, as only specific sub-groups of sensors are affected. Other than the absence of other causation and the re-observation of sensor groupings however, there is no direct evidence for this event being caused by the experimental set-up.

The cause and source of the perturbations highlighted within the time window in question remain open to speculation. While the temporary increase in noise reported may be a real physical effect, or the result of an experimental artefact, the observation of which sensor records they occur within reinforce the previous observations of sub-groups of sensors possessing similar second order behaviours.

8.3.3 Event at 2010–11–29

In contrast to the two previous events discussed, in which the behaviours detected were similar across the sensor records they were detected in (i.e. steps in magnitude at 2010–01–10 and noise increases at 2010–06–27), the behaviours detected during this event window are significantly varied. Additionally, the locations of the perturbations identified in Section 7.5 are spread throughout the deposition hole without any apparent spatial pattern, as shown in Figure 8.10. This casts an ambiguity over the analysis, and presents the possibility that the detected perturbations have occurred synchronously at random, rather than causatively.

With this possibility in mind, PB927 is effectively discounted from the analysis of the event. As noted in Section 7.5, other similar changes in magnitude occur in the sensor record in close proximity to, but not within the event time window. This implies that the behaviour within the window is not atypical for the sensor during that period, and therefore the synchronisation with other perturbations in the experimental system is more likely to be coincidental.

The seed event in UB901's sensor record is unique with respect to its form throughout the dataset during the relevant time window. The drops in magnitude at the start of the event window in PR917, PB902, and UR922, as a subset, are not correlated spatially or by sensor type. The gains in magnitude that subsequently dissipate back to the initial sensor record trend in UR912 and UB902 are also not correlated spatially or by sensor type. The sensors that express increases in noise (PR914, PR912, and PR906) are correlated by sensor type, but not

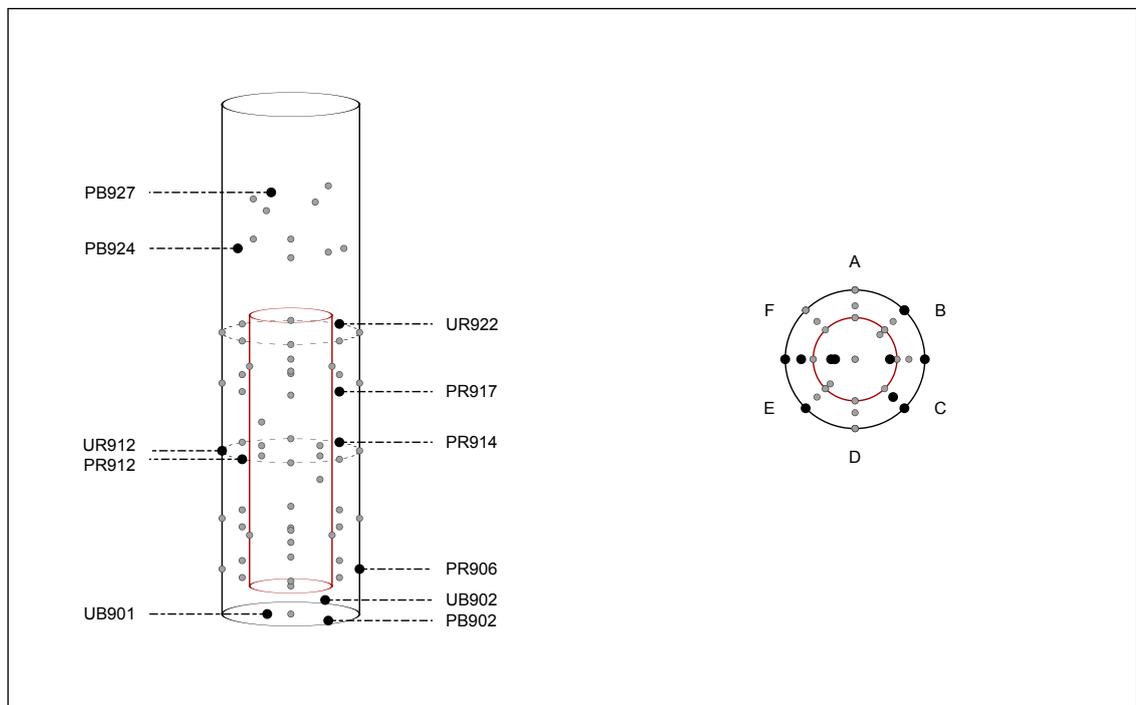


Figure 8.10: Locations of sensors identified as part of event on 2010-11-29 in relation to deposition hole arrangement.

spatially.

A possible connection between the sensors identified in this possible event is the fracture network within the host rock surrounding DA3147G01. However there are too many unknowns regarding preferential flow paths within the fracture network to speculate on the nature of flows between the relevant sensors. Additionally, there are sensors identified within this event that are not located on the deposition hole wall. The same is true regarding speculation on preferential flow paths within the bentonite EBS, or possible points in the system that are significant but are not instrumented.

Further discounting of perturbations that are not spatially correlated to the seed event allow a simplification of the speculative process. It is noted however that this does not allow firm conclusions to be drawn. Figure 8.11 depicts UB901, UB902, PB902, and PR906. These sensors are located within the bottom of DA3147G01.

Within the narrower scope of the four sensors depicted in Figure 8.11 the possibility emerges that a flow event took place from PB902 to UB901, although the suggested transmission of pressure was not instantaneous as seen during the event at 2010-01-10, nor has there been a macro scale gas flow event noted as intercepting that sensor. The added complexity in such a transmission though may result from an interaction with other adjacent measurement points, namely UB902 and PR906. The direction of the noise in PR906 implies a tentative source, rather than a tentative receptor however.

8.4 Summary

The development of NUDAT and its subsequent application to Lasgit's dataset has been instrumental in highlighting and subsequently analysing a complex set of second order aspects and events occurring during the experiment. Macro-scale interpretations of the aggregated small scale, or 'second order' features revealed have been conducted along with in depth interpretations of synchronised perturbations. To that end, the developed toolkit has been shown to be accommodating of non-uniform time series input of a large scale, and sensitive to second order events as many as 10^{-3} times smaller in magnitude than the macro-scale behaviour they are present in, exposing phenomena that would otherwise be occluded during an Exploratory Data Analysis (EDA).

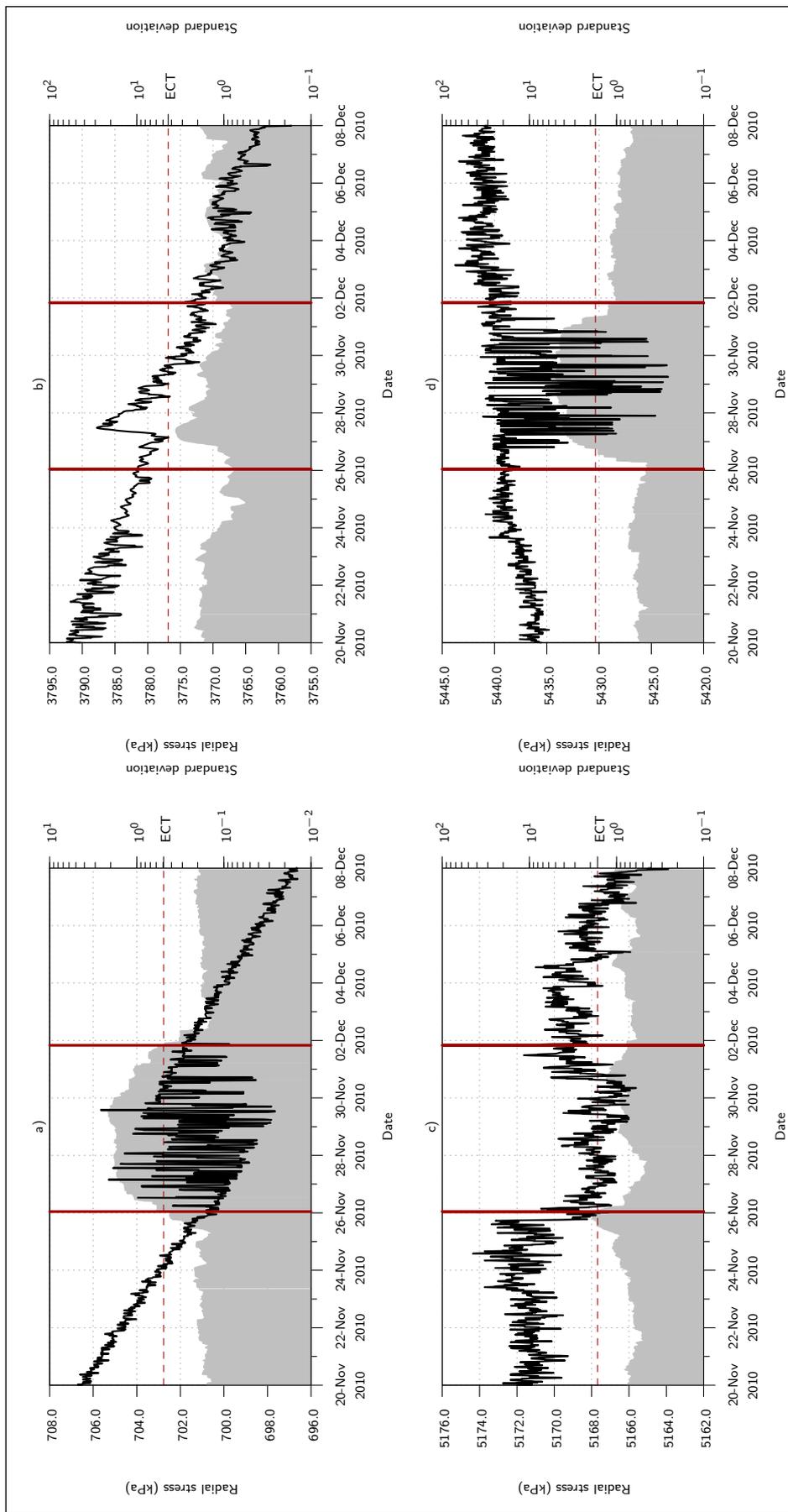


Figure 8.11: Remaining perturbation at 2010–11–29 after discounting non-spatially correlated sensor records. a) UB901. b) UB902. c) PB902. d) PR906.

Specifically, review of the toolkit outputs has highlighted evidence for localised flow pathway(s) existing, with the possibility of gas pressure induced formation of pathways facilitating such flow (i.e. a dilatant flow event). Additionally, there is evidence for sub-groups of sensors behaving similarly over a variety of types of phenomena, for example, sensors PC901, PC902, and PC903 can be shown to be distinct from other stress sensors within DA3147G01, but similar to each other with respect to spike propensity and response to global influences. This in turn implies that the canister, as a common element to these sensors, is a significant influence on their behaviours.

Furthermore, de-trending and frequency domain inspection reveals an almost ubiquitous annual cycle in the dataset that is quantitatively measurable, and in the case of less distorted sensor records, highly correlated with depth within DA3147G01. Additionally, a number of daily cycles are detectable in various sensor records, highlighting the sensitivity of the toolkit.

Some of the information revealed is ambiguous, contradictory or unexplained however. The reduced synchronisation between the in-bentonite pore-water pressure sensors, for example, is not reflected in the 2010-06-27 event. Additionally, there are no apparent causative influences for that event or the temperature anomaly revealed in UT901.

However, despite some inconclusive information, or in some cases partially contradictory interpretations appearing possible, a number of clearly defined behaviours and events have been identified. The interpretation of such aspects, such as the second order event occurring at 2010-01-10, contributes to the body of evidence used to understand gas flow in saturated, compacted clays. Additionally, the complexity, heterogeneity, and ambiguous sensitivity to the boundary conditions of such an emplacement is highlighted by these interpretations.

Conclusions and suggested further work

9.1 Introduction

In response to the need to safely dispose of the existing radioactive waste inventory resulting from nuclear technologies, a programme of research has been undertaken to develop a deep geological disposal facility capable of isolating the waste for 100,000 to 1,000,000 years. While various conceptual models, such as the Swedish Kärnbränslesäkerhet [Nuclear Fuel Safety] (KBS) concept, are well developed, a number of technical issues remain. Within the compacted bentonite Engineered Barrier System (EBS) component of the KBS-3 model, significant uncertainties associated with the formation of gas pathways and migration of gas exists. In order to study the process of gas migration in a repository context the Large Scale Gas Injection Test (Lasgit) was undertaken.

This PhD thesis has presented research undertaken in a detailed and systematic computational analysis of the results produced through the operation of Lasgit. The objectives of this research have been achieved through development and verification of a computational Exploratory Data Analysis (EDA) methodology, and its application to Lasgit's dataset. The implementation of the EDA methodology resulted in the Non-Uniform Data Analysis Toolkit (NUDAT), a computational toolkit capable of accommodating non-uniform input data and performing a range of time series analysis techniques.

The discussion and interpretation of the outcome of such an analysis were intended to contribute to the overall understanding of the experiment. Presented in this chapter is a summary of the work done in achieving these objectives, a summary of the main conclusions drawn from the interpretation, and a brief consideration on the possibilities for the continuation of this research.

9.2 Summary and conclusions of the research undertaken

9.2.1 Toolkit development

Challenges addressed during the development of NUDAT were the long-term nature and large quantity of data to be analysed, the need for accommodation of non-uniform data, and the requirement for exposure of small scale, or 'second order' information. The computational solution developed in order to perform the desired EDA on Lasgit's dataset was named NUDAT. During the development of NUDAT, the choice of analytical techniques incorporated was guided by the nature of the information sought from the dataset and the characteristics of the dataset.

Generally required capabilities, i.e. those useful for the majority of datasets, such as smoothing/moving average functions and trend detection and subsequent removal, were incorporated into the toolkit. Other more specifically focused techniques such as frequency domain analysis, spike detection for the specific type of spiking observed in the dataset, and second order event detection were also incorporated. These techniques, when used in sequence/combination, allowed for the exposure of second order features in the dataset that were of interest to the analysis of the experiment.

The options pertaining to the handling of the non-uniformity present in the dataset were explored during the development of the analytical software. Accommodation of non-uniform input at the algorithmic level was decided to be the most appropriate course of action, in favour of a preprocessor standardisation step. To this end, a collection of data analysis techniques were developed or modified in order to accommodate such input.

In particular, non-uniformity was accommodated through: i) time scale windowing, as opposed to point count windowing that was defined and adopted for any processes in the toolkit that utilises a moving window; and ii) modifying the standard Discrete Fourier Transform (DFT) procedure to decouple it from the time series sample rate and length and producing the Non-uniform Discrete Fourier Transform (NDFT) allowed frequency content to be measured. The processes implemented in NUDAT that utilise a moving window include the smoothing/averaging functions, the spike detection, and the event candidate detection.

In summary, the capabilities/processes implemented in NUDAT were:

- Moving averages/smoothing/down-sampling functions with multiple weighting profiles
- Spike (aberrant point) detection
- Trend detection and removal (SSA)
- Frequency domain analysis (NDFT)
- Event candidate detection, with event intensity ranking and cross time series synchronisation visualisation

Additionally, the capabilities incorporated into NUDAT, including those developed or modified to accommodate non-uniform input, were chosen for a general applicability to time series data, while remaining specifically relevant to the phenomena observed in Lasgit's dataset.

9.2.2 Verification of the developed toolkit

In order to verify the capabilities and implementation of NUDAT, an application to a dataset with known phenomena was undertaken. For this purpose an artificial dataset was created with macro scale and second order features. The phenomena included in the created dataset included long-term trends with distortions, seasonal (cyclic) behaviour, noise, spiking, and small perturbations. The application of NUDAT to the artificial dataset demonstrated the successful detection of trends with enough accuracy to render the residual effectively stationary, making it suitable for further analysis. The frequency content was detected with high accuracy, and a high sensitivity/accuracy/precision was found for the spike and event detection processes. The success of the verification application in identifying correctly the known phenomena in the data provided confidence in the validity of the subsequent analysis of Lasgit's dataset.

9.2.3 Application of NUDAT to Lasgit's dataset

Lasgit is a highly complex experiment, both in terms of experimental set-up, and operational history. The dataset arising from Lasgit available to this research consisted of circa 27.5 million datum points spread across approximately 175 time series records. The data are sampled non-uniformly at an average of around 2.5 logs/hour. This equates to 157,362 logging cycles

recorded for each sensor. While macro scale analyses of the experimental results have been performed previously, little attention has been paid to the smaller scale phenomena present in the dataset.

Application of NUDAT to Lasgit's dataset consisted of spike detection and removal, followed by frequency domain analysis, event candidate detection, and noise reduction. Additionally, the de-spiked data underwent trend detection and removal, and the frequency domain analysis, event candidate detection, and noise reduction subsequently repeated on the de-trended data. The application process yielded information about spike propensity, frequency content, and small scale perturbations in the dataset that are of potential interest to an analyst for further investigation.

9.2.4 Exploratory data analysis results

An overall level of approximately 0.7% of the data were identified as spikes and subsequently removed from further analysis. This equates to approximately one sensor records 'worth' of data being removed, spread across the dataset. Distinct clusters of tendency to spike were also observed in various sets/classes of sensors within DA3147G01.

Trend detection and removal exposed the seasonal variation present in the dataset. This allowed more accurate measurement and quantification via the NDFT process. The annual frequency content within the de-trended dataset was prominent, and in the case of the temperature sensor records, highly ordered and correlated with depth of the sensors within DA3147G01. While annual frequency content was present in most of the non-temperature sensors it was not measurably correlated with depth.

A number of second order perturbations were detected by the event candidate detection process, and in one case revealed by the noise reduction process. Perturbations as small as 10^3 times smaller than the macro scale behaviour are clearly highlighted by this process. The synchronicity of such perturbations within the experimental system was used as an indicator for the likelihood of concurrent being interconnected. Three sets of NUDAT highlighted perturbations were investigated and presented as second order events. The events were chosen for investigation for having a prominent perturbation proximate both spatially and temporally to a previously identified macro scale gas flow event.

The events studied in detail occurred at approximately 2010-01-10, 2010-06-27, and 2010-11-29. The first event studied, 2010-01-10, occurs across five sensors, all of which are located towards the bottom of DA3147G01. Small drops in pressure are reported at sensor locations that have previously been observed to have intercepted a macro scale gas flow event. Synchronously, small increases are noted in nearby sensors. The second event studied, 2010-06-27, appeared to have a global presence within DA3147G01. Closer inspection of the data revealed that while the perturbation detected was almost ubiquitous, sensors attached to the canister did not report the disturbance observed. The third event studied, 2010-11-29, consisted of a number of perturbations more varied in nature than the previous two events, and with less overall synchronicity.

9.2.5 Micro/macro scale behaviour and observations

Interpretation of many of the NUDAT's results imply similarity of behaviour in the experimental system across sensor classes. Propensity to spike distinctly divides sensors within DA3147G01 into temperature, non-temperature sensors, and canister attached sensors. This is compounded by the correlation of spike occurrence with depth, which possesses three distinct locus of points that have very little intersection with one another: i) the temperature sensors; ii) the pore-water pressure and stress sensors; and iii) the canister attached sensors.

The exposure and quantification of an annual cycle almost ubiquitously throughout the dataset, and in some cases a daily frequency cycle, implies a sensitivity to external influences like the air temperature in Äspö. Variance within such measurements however are not adequately explained by consideration of the knowledge of the boundary conditions.

Evidence of discrete flow pathways is present in the first event studied, with the possibility of spontaneous pathway formation having taken place, potentially via a dilatant flow mechanism. The second event studied further compounds the apparent classification of sensors into subsets. While suggesting that the experiment may be sensitive to external influences, the exact nature and cause of the second event studied remains unclear. The third event studied serves to demonstrate the complexity of the system that remains unmonitored despite the high levels of instrumentation as determination of the physical processes that occurred during the third event is not possible.

9.3 Overall conclusions

As a result of this research it is asserted that:

- A new toolkit, NUDAT, has been developed, capable of analysing the small scale, or ‘second order’ features within a time series based dataset that possesses non-uniformity, such as Lasgit’s dataset.
- The processes implemented in the toolkit have been subjected to a series of verification processes that provide confidence in the accuracy of the analytical components of the toolkit for further applications that seek to quantify phenomena that may be of interest to an analyst in such a dataset.
- Application of the toolkit to Lasgit’s dataset constitutes a robust and methodical exploration of the data which has facilitated the identification and analysis of small scale, or ‘second order’ features in the dataset up to 10^3 times smaller than the macro scale behaviour.
- The analysis interpretation identifies both global and localised events occurring, along with apparent groupings of behaviours between classes of sensors, and a prominent seasonal variation that is quantifiable across almost all of the dataset.
- The collections/sets of second order perturbations studied imply the possibility of experimental sensitivity to outside influences and also to the canister, and suggest the possibility of dilatant flow like events occurring on a small scale.

9.4 Potential further research

A number of avenues for continuation of the research undertaken and presented in this thesis are available. Given the size of Lasgit’s dataset, and the ongoing nature of the experiment, further study of the experiment is warranted. Additionally, broadening of the scope of NUDAT through development and addition of other analytical modules may also be of benefit. Application of NUDAT to datasets derived from other experimental or monitoring programmes may also be beneficial to both the ongoing characterisation and development of NUDAT, and to the assisted programmes through analytical insight.

9.4.1 Further research on Lasgit's dataset

An immediately accessible continuation of the research presented in this thesis is the continuation of the study of the second order events identified in Lasgit's dataset. Three such events have been detected, presented, and analysed in the previous chapters, however a number of other prominent event candidate impulses, particularly in proximity to other gas injection campaigns, are yet to be investigated in detail. Additionally, analysis of more recent data produced by the operation of Lasgit may be of interest for analysis. In particular, analysis of the data produced during the future of the experiment, in the period before decommissioning, may be beneficial. Interpretations made from an analysis of such an experimental stage can be attempted to be verified by direct observation during decommissioning. Further permutations of the toolkit's analytical parameters used during the analysis, may also contribute to the understanding of the experiment by revealing or highlighting phenomena of interest that occur on a different temporal scale to the scales thus far investigated.

9.4.2 Further research on other datasets

The dataset arising from the monitoring programme at Olkiluoto, a Underground Characterisation Facility (UCF) intended to become a radioactive waste repository, is likely to be suited to analysis via application of NUDAT. Rock and air temperature records at a number of locations on the site are recorded with a non-uniform sample rate. The use of NUDAT on datasets derived by other experimental or monitoring programmes is not necessarily limited to datasets derived from radioactive waste disposal activities, however. Other long term geotechnical, environmental, economic etc. monitoring activities that generate time series may be suited to analysis with NUDAT, or useful for characterisation of the analytical processes.

9.4.3 Further characterisation and development of NUDAT analytical modules

Possible useful future developments of NUDAT include:

1. Characterisation of the behaviour of NUDAT when provided with input data containing features not present in the validation dataset used in Chapter Five. Examples include the

effects of:

- (a) Modulated frequency content in a time series
- (b) Trend or variability in seasonality
- (c) Absence of a specifically targeted phenomenon

2. Development of and addition of analytical processes. Examples include:

- (a) Modifying the whole of time series processes, e.g. the NDFT, to operate within a moving window
- (b) Implementation of a non-uniform accommodating autocorrelation function
- (c) Trend significance detection and Moving Average over Shifting Horizon (MASH)

3. Optimisation and further parallelisation of the code allowing use in High Performance Computing (HPC) environments.

References

- Acir, N and Güzeliş, C (2004). "Automatic spike detection in EEG by a two-stage procedure based on support vector machines". In: *Computers in Biology and Medicine* 34 (2004), pp. 561–575. DOI: 10.1016/j.combiomed.2003.08.003.
- Alkan, H and Müller, W (2008). "Approaches for modelling gas flow in clay formations as repository systems". In: *Physics and Chemistry of the Earth* 33 (2008), S260–S268.
- Alonso, EE, Gens, A, and Josa, A (1990). "A constitutive model for partially saturated soils". In: *Géotechnique* 40 (1990), pp. 405–430.
- Alonso, EE, Olivella, S, and Arnedo, D (2006). "Mechanisms of gas transport in clay barriers". In: *Journal of Iberian Geology* 32 (2006), pp. 175–196.
- Anderson, TW (1971). *The statistical analysis of time series*. John Wiley and Sons, 1971.
- Andra (2005). *Dossier 2005 granite: Phenomenological evolution of a geological repository*. Tech. rep. Agence nationale pour la gestion des déchets radioactifs (Andra), 2005.
- Anghileri, D, Pianosi, F, and Soncini-Sessa, R (2014). "Trend detection in seasonal data: from hydrology to water resources". In: *Journal of Hydrology* 511 (2014), pp. 171–179. DOI: 10.1016/j.jhydrol.2014.01.022.
- Arnedo, D, Olivella, S, and Alonso, EE (2010). "Large scale injection test (Lasgit) modelling". In: *Clays in natural & engineered barriers for radioactive waste confinement 4th international meeting*. 2010.
- Azami, H and Sanei, S (2014). "Spike detection approaches for noisy neuronal data: Assessment and comparison". In: *Neurocomputing* 133 (2014), pp. 491–506. DOI: 10.1016/j.neucom.2013.12.006.

-
- Bäck, R (2003). *Djupförvarsteknik: Large Scale Gas Injection Test - Lid report*. Tech. rep. TD-04-07. Svensk Kärnbränslehantering AB, 2003.
- Bagchi, S and Mitra, SK (1999). *The nonuniform discrete Fourier transform and its applications in signal processing*. Kluwer Academic Publishers, 1999.
- Barnes, G (2000). *Soil Mechanics: Principles and Practice*. Second Edition. Palgrave, 2000.
- Bazovsky, I (1961). *Reliability theory and practice*. Prentice-Hall, 1961.
- Bennett, DP, Cuss, RJ, Vardon, PJ, Harrington, JF, and Thomas, HR (2011). "Data Analysis of a Large Scale Gas Injection Test". In: *Geological Disposal of Radioactive Waste: Underpinning Science and Technology — 18th – 20th October 2011, Loughborough University*. Abstract 59. Nuclear Decommissioning Authority. 2011, p. 101.
- Bennett, DP, Cuss, RJ, Vardon, PJ, Harrington, JF, and Thomas, HR (2014). "Phenomena exposure from the Large Scale Gas Injection Test (Lasgit) dataset using a bespoke data analysis toolkit". In: *Clays in Natural and Engineered Barriers for Radioactive Waste Confinement, Geological Society, London, Special Publications 400* (2014).
- Bennett, DP, Cuss, RJ, Vardon, PJ, Harrington, JF, Philp, RN, and Thomas, HR (2012a). "Data analysis toolkit for long-term, large-scale experiments". In: *Mineralogical Magazine 76* (2012), pp. 3355–3364.
- Bennett, DP, Cuss, RJ, Vardon, PJ, Harrington, JF, and Thomas, HR (2012b). "Information contained within the Large Scale Gas Injection Test (Lasgit) dataset exposed using a bespoke data analysis toolkit". In: *Clays in Natural and Engineered Barriers for Radioactive Waste Confinement, 5th International meeting — 22nd – 25th October 2012, Montpellier, France*. Abstract GPTP/11. Andra. 2012, pp. 716–717.
- Bennett, DP, Cuss, RJ, Vardon, PJ, Harrington, JF, Sedighi, M, Shaw, RP, and Thomas, HR (2013). "'Second order' exploratory data analysis of the Large Scale Gas Injection Test (Lasgit) dataset, focused around known gas migration events". In: *Fate of repository gases project; Gas generation and migration international symposium and workshop 5th-7th February 2013, Luxembourg*. Abstract PSA3. Nuclear Energy Agency. 2013, p. 50.

-
- Birgersson, M, Åkesson, M, and Hökmark, H (2008). "Gas intrusion in saturated bentonite - A thermodynamic approach". In: *Physics and Chemistry of the Earth* 33 (2008), S248–S251.
- Bloomfield, P (2000). *Fourier analysis of time series: An introduction*. Second Edition. Probability and Statistics. John Wiley and Sons, 2000.
- Börgesson, L, Karnland, O, and Johannesson, LE (1996). "Modelling of the physical behaviour of clay barriers close to water saturation". In: *Engineering Geology* 41 (1996), pp. 127–144.
- Box, GEP and Jenkins, GM (1976). *Time Series Analysis: Forecasting and control*. Ed. by E Robinson. Revised Edition. Time series analysis and digital processing. Holden-Day Inc, 1976.
- Bradbury, MH and Baeyens, B (2003). "Porewater chemistry in compacted re-saturated MX-80 bentonite". In: *Journal of Contaminant Hydrology* 61 (2003), pp. 329–338.
- Brown III, MW, Porter, BE, Dlugos, DJ, Keating, J, Gardner, AB, Storm Jr, PB, and Marsh, ED (2007). "Comparison of novel computer detectors and human performance for spike detection in intracranial EEG". In: *Clinical Neurophysiology* 118 (2007), pp. 1744–1752. DOI: 10.1016/j.clinph.2007.04.017.
- BSI (1999). *Code of practice for site investigations*. BS 5930. British Standards Institution. 1999.
- Calder, N, Walsh, R, and Avis, J (2013). "3D gas transport model of Lasgit gas injection test". In: *Fate of repository gases project; Gas generation and migration international symposium and workshop 5th-7th February 2013, Luxembourg*. Abstract PS1A3. Nuclear Energy Agency. 2013, p. 11.
- Carden, EP and Brownjohn, JMW (2008). "ARMA modelled time-series classification for structural health monitoring of civil infrastructure". In: *Mechanical Systems and Signal Processing* 22 (2008), pp. 295–314. DOI: 10.1016/j.ymsp.2007.07.003.
- Chatfield, C (1989). *The analysis of time series: An introduction*. Fourth Edition. Chapman and Hall, 1989.
-

- Committee on the Analysis of Massive Data, Committee on Applied and Theoretical Statistics, Board on Mathematical Sciences and Their Applications, Division on Engineering and Physical Sciences, and National Research Council (2013). *Frontiers in Massive Data Analysis*. The National Academies Press, 2013.
- Cool, W, Cuñado, M, Johnson, L, Mäntynen, M, Müller, W, Norris, S, Rodwell, W, Sellin, P, Snellman, M, Talandier, J, Vieno, T, and Vines, S (2004). "EC GASNET project: Gas issues in the safety assessment of deep repositories for nuclear waste". In: *Scientific Basis for Nuclear Waste Management XXVII*. Ed. by VM Oversby and LO Werme. Vol. 807. MRS Symposium Proceedings. Materials Research Society. 2004, pp. 703–708.
- CoRWM (2006). *Managing our Radioactive Waste Safely: CoRWM's recommendations to Government*. Tech. rep. Committee on Radioactive Waste Management (CoRWM), 2006.
- Craig, RF (1992). *Soil Mechanics*. Fifth Edition. Chapman and Hall, 1992.
- Cuss, RJ, Harrington, JF, and Noy, DJ (2010). *Large scale gas injection test (Lasgit) performed at the Äspö Hard Rock Laboratory: Summary report 2008*. Tech. rep. TR-10-38. Svensk Kärnbränslehantering AB, 2010.
- Cuss, RJ, Harrington, JF, Noy, DJ, Wikman, A, and Sellin, P (2011). "Large scale gas injection test (Lasgit): Results from two gas injection tests". In: *Physics and Chemistry of the Earth* 36 (2011), pp. 1729–1742.
- Cuss, RJ, Harrington, JF, Noy, DJ, Bennett, DP, and Sellin, P (2012). "The natural and artificial hydration of a bentonite engineered barrier system in a full scale KBS-3V mock-up; Results from the first 7 years of the Large Scale Gas Injection Test (Lasgit)". In: *Clays in Natural and Engineered Barriers for Radioactive Waste Confinement, 5th International meeting — 22nd – 25th October 2012, Montpellier, France*. Abstract GM/ISE/2. Andra. 2012, pp. 605–606.
- de Lautour, OR and Omenzetter, P (2010). "Damage classification and estimation in experimental structures using time series analysis and pattern recognition". In: *Mechanical Systems and Signal Processing* 24 (2010), pp. 1556–1569. DOI: 10.1016/j.ymssp.2009.12.008.

-
- Devore, JL and Peck, R (2005). *Statistics : the exploration and analysis of data*. Fifth Edition. Thomson Brooks/Cole, 2005.
- Diggle, PJ (1990). *Time series a biostatistical introduction*. Clarendon press, 1990.
- Dixon, D, Chandler, N, Graham, J, and Gray, MN (2002). "Two large-scale sealing tests conducted at Atomic Energy of Canada's underground research laboratory: the buffer-container experiment and the isothermal test". In: *Canadian Geotechnical Journal* 39 (2002), pp. 503–518.
- Donohew, AT, Horseman, ST, and Harrington, JF (2000). "Gas entry into unconfined clay pastes at water contents between the liquid and plastic limits". In: *Environmental Mineralogy: microbial interactions, anthropogenic influences, contaminated land and waste management. Mineralogical Society Series 9* (2000), pp. 369–394.
- EC (2014). *Communication from the commission to the European Parliament and the council: European energy security strategy*. Tech. rep. SWD(2014) 330 final. European Commission, 2014.
- Eckner, A (2012). "A framework for the analysis of unevenly spaced time series data". Self published. 2012. URL: http://eckner.com/papers/unevenly_spaced_time_series_analysis.pdf.
- Eckner, A (2013). "Algorithms for Unevenly-Spaced Time Series: Moving Averages and Other Rolling Operators". Self published. 2013. URL: http://eckner.com/papers/ts_alg.pdf.
- Ekanadham, C, Tranchina, D, and Simoncelli, EP (2014). "A unified framework and method for automatic neural spike identification". In: *Journal of Neuroscience Methods* 222 (2014), pp. 47–55. DOI: 10.1016/j.jneumeth.2013.10.001.
- Erdoğan, H and Güla, E (2009). "The application of time series analysis to describe the dynamic movements of suspension bridges". In: *Nonlinear Analysis: Real World Applications* 10 (2009), pp. 910–927. DOI: 10.1016/j.nonrwa.2007.11.013.
-

-
- Esteller, R, Vachtsevanos, G, Echauz, J, and Litt, B (2001). "A Comparison of Waveform Fractal Dimension Algorithms". In: *IEEE Transactions on Circuits and Systems* 48 (2001), pp. 177–183. DOI: 10.1109/81.904882.
- Faybishenko, BA (1995). "Hydraulic behaviour of quasi-saturated soils in the presence of entrapped air: Laboratory experiments". In: *Water Resources Research* 31 (1995), pp. 2421–2435.
- Fleureau, JM, Kheirbek-Saoud, S, Soemitro, R, and Taibi, S (1993). "Behavior of clayey soils on drying-wetting paths". In: *Canadian Geotechnical Journal* 30 (1993), pp. 287–296.
- Flowers, B (1976). *Nuclear power and the environment*. Tech. rep. Cmmd. 6618. Royal Commission on Environmental Pollution, 1976.
- FoE (2012). *Five need-to-know facts about nuclear power*. Website. Accessed: 2012–10–02. Friends of the Earth. 2012. URL: http://www.foe.co.uk/campaigns/climate/issues/nuclear_index.html.
- Fredlund, DG (1991). "How negative can pore-water pressure get?" In: *Geotechnical News* September (1991), pp. 44–46.
- Fredlund, DG and Rahardjo, H (1993). *Soil mechanics for unsaturated soils*. John Wiley and Sons Inc., 1993.
- Fu, T (2011). "A review on time series data mining". In: *Engineering Applications of Artificial Intelligence* 24 (2011), pp. 164–1814. DOI: 10.1016/j.engappai.2010.09.007.
- Gallé, C (2000). "Gas breakthrough pressure in compacted Fo-Ca clay and interfacial gas overpressure in waste disposal context". In: *Applied Clay Science* 17 (2000), pp. 85–97.
- Gallé, C (1998). "Migration des gaz et pression de rupture dans une argile compactee destinee a la barriere ouvree d'un stockage profond". In: *Bulletin de la Societe Geologique de France* 169 (1998). Translation of abstract from French to English sourced, original article cited unseen, pp. 675–680.

-
- Gallé, C and Tanai, K (1998). "Evaluation of gas transport properties of backfill materials for waste disposal: H₂ migration experiments in compacted Fo-Ca clay". In: *Clays and Clay Minerals* 46 (1998), pp. 498–508.
- Gaspard, N, Alkawadri, R, Goncharova, II, and Zaveri, HP (2014). "Automatic detection of prominent interictal spikes in intracranial EEG: Validation of an algorithm and relationship to the seizure onset zone". In: *Clinical Neurophysiology* 125 (2014), pp. 1095–1103. DOI: 10.1016/j.clinph.2013.10.021.
- Gens, A and Alonso, EE (1992). "A framework for the behaviour of unsaturated expansive clays". In: *Canadian Geotechnical Journal* 29 (1992), pp. 1013–1032.
- Gillberg, J and Ljung, L (2010a). "Frequency domain identification of continuous-time output error models, Part I: Uniformly sampled data and frequency function approximation". In: *Automatica* 46 (2010), pp. 1–10. DOI: 10.1016/j.automatica.2009.10.004.
- Gillberg, J and Ljung, L (2010b). "Frequency domain identification of continuous-time output error models, Part II: Non-uniformly sampled data and B-spline output approximation". In: *Automatica* 46 (2010), pp. 11–18. DOI: 10.1016/j.automatica.2009.10.005.
- Golub, GH and Loan, CFV (1996). *Matrix computations*. Third Edition. Johns Hopkins University Press, 1996.
- Golub, GH and van der Vorst, HA (2000). "Eigenvalue computation in the 20th century". In: *Computational and Applied Mathematics* 123 (2000), pp. 35–65.
- Golyandina, N, Nekrutkin, V, and Zhigljavsky, A (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Monographs on Statistics and Applied Probability 90. Chapman and Hall/CRC, 2001.
- Graham, J, Halayko, KG, Hume, HB, Kirkham, T, Gray, M, and Oscarson, D (2002). "A capillarity-advective model for gas break-through in clays". In: *Engineering Geology* 64 (2002), pp. 273–286.

-
- Greenpeace (2012). *Nuclear*. Website. Accessed: 2012–10–02. Greenpeace. 2012. URL: <http://www.greenpeace.org/international/en/campaigns/nuclear/>.
- Grim, RE (1962). *Applied clay mineralogy*. McGraw-Hill Publishing Company Limited, 1962.
- GTS (2014). *Full-scale Engineered Barriers Experiment (FEBEX)*. Website. Accessed: 2014–06–03. Grimsel Test Site. 2014. URL: <http://www.grimsel.com/gts-phase-v/febex/febex-introduction->.
- Halpern, S (1978). *The Assurance Sciences*. Prentice-Hall, 1978.
- Hannan, EJ, Krishnaiah, PR, and Rao, MM (1985). *Time series in the time domain*. Elsevier science publishers, 1985.
- Hardenby, C (2004). *The Lasgit hole DA3147G01 Hydrogeology*. Tech. rep. IPR-04-52. SwedPower AB, 2004.
- Hardenby, C and Lundin, J (2003). *Geological mapping of the assembly hall and deposition hole*. Tech. rep. IPR-03-28. SwedPower AB, 2003.
- Harrington, JF and Horseman, ST (1999). "Gas transport properties of clays and mudrocks". In: *Muds and Mudstones: Physical and Fluid Flow Properties Geological Society of London, Special Publication 158* (1999), pp. 107–124.
- Harrington, JF and Horseman, ST (2003). *Gas migration in KBS-3 buffer bentonite. Sensitivity of test parameters to experimental boundary conditions*. Tech. rep. TR-03-02. British Geological Survey, 2003.
- Harrington, JF, Birchall, DJ, Noy, DJ, and Cuss, RJ (2008). *Large scale gas injection test (Lasgit) performed at the Äspö Hard Rock Laboratory: Summary report 2007*. Tech. rep. CR/07/211. British Geological Survey, 2008.
- Hedin, A (1997). *Spent nuclear fuel — how dangerous is it? A report from the project "Description of risk"*. Tech. rep. TR-97-13. Svensk Kärnbränslehantering AB, 1997.

-
- Hoch, AR, Cliffe, KA, Swift, BT, and Rodwell, WR (2004). *Modelling Gas Migration in Compacted Bentonite: GAMBIT Club Phase 3 Final Report*. Tech. rep. POSIVA 2004–02. Posiva, 2004.
- Hogrefe, C, Rao, ST, and Zurbenko, IG (1998). “Detecting trends and biases in time series of ozonesonde data”. In: *Atmospheric Environment* 32 (1998), pp. 2569–2586. DOI: 10.1016/S1352-2310(97)00512-8.
- Hokari, T, Okihara, M, Ishii, T, Ishii, T, and Ikuse, H (1997). “Experimental study on scale effects of bentonite sand mixtures on gas migration properties”. In: *Scientific Basis for Nuclear Waste Management XX*. Ed. by WJ Gray and IR Triay. Vol. 465. MRS Symposium Proceedings. Materials Research Society. 1997, pp. 1019–1026.
- Hook, J (2014). “Smoothing non-smooth systems with low-pass filters”. In: *Physica D: Nonlinear Phenomena* 269 (2014), pp. 76–85. DOI: 10.1016/j.physd.2013.11.016.
- Horseman, ST and Harrington, JF (1994). *Migration of repository gases in an overconsolidated clay*. Tech. rep. WE/94/7. British Geological Survey, 1994.
- Horseman, ST and Harrington, JF (1997). *Study of Gas Migration in Mx80 Buffer Bentonite*. Tech. rep. WE/97/7. British Geological Survey, 1997.
- Horseman, ST, Higgo, JJW, Alexander, J, and Harrington, JF (1996). *Water, gas and solute movement through argillaceous media*. Tech. rep. CC-96/1. Nuclear Energy Agency, 1996.
- Horseman, ST, Harrington, JF, and Sellin, P (1997). “Gas migration in Mx80 buffer bentonite”. In: *Scientific Basis for Nuclear Waste Management XX*. Ed. by WJ Gray and IR Triay. Vol. 465. MRS Symposium Proceedings. Materials Research Society. 1997, pp. 1003–1010.
- Horseman, ST, Harrington, JF, and Sellin, P (1999). “Gas migration in clay barriers”. In: *Engineering Geology* 54 (1999), pp. 139–149.
-

-
- Horseman, ST, Harrington, JF, and Sellin, P (2004). "Water and gas movement in Mx80 bentonite buffer clay". In: *Scientific Basis for Nuclear Waste Management XXVII*. Ed. by VM Oversby and LO Werme. Vol. 807. MRS Symposium Proceedings. Materials Research Society. 2004, pp. 715–720.
- Hume, HB (1999). "Gas Breakthrough in Compacted Avonlea Bentonite". MSc. Department of Soil Science, University of Manitoba, Whnipeg, 1999.
- Impey, MD, Grindrod, P, Takase, H, and Worgan, KJ (1997). "A capillary network model for gas migration in low-permeability media". In: *SIAM Journal on Applied Mathematics* 57 (1997), pp. 597–608.
- ISO (2004). *Data elements and interchange formats – Information interchange – Representation of dates and times*. ISO 8601. International Organization for Standardization. 2004.
- Johannesson, LE (2003). *Djupförvarsteknik: Large Scale Gas Injection Test - Manufacturing of buffer for the large scale gas injection test*. Tech. rep. TD-04-01. Clay Technology AB, 2003.
- Johannesson, LE (2004). *Project Lasgit: Buffer emplacement and disposal of canister*. AP TD F122-03-035. Svensk Kärnbränslehantering AB. 2004.
- Johannesson, LE, Börgesson, L, and Sandén, T (1995). *Compaction of bentonite blocks: Development of technique for industrial production of blocks which are manageable by man*. Tech. rep. TR-95-19. Clay Technology AB, 1995.
- Johannesson, LE, Börgesson, L, Goudarzi, R, Sandén, T, Gunnarsson, D, and Svemar, C (2007). "Prototype repository: A full scale experiment at Äspö HRL". In: *Physics and Chemistry of the Earth* 32 (2007), pp. 58–76.
- Johansson, E and Siren, T (2014). *Results of Monitoring at Olkiluoto in 2012*. Tech. rep. 2013-47. Posiva O, 2014.
- Kapur, KC and Lamberson, LR (1977). *Reliability in Engineering Design*. John Wiley and Sons, 1977.

-
- Kim, S and McNames, J (2007). "Automatic spike detection based on adaptive template matching for extracellular neural recordings". In: *Journal of Neuroscience Methods* 165 (2007), pp. 165–174. DOI: 10.1016/j.jneumeth.2007.05.033.
- Kirkham, T (1995). "Development of test equipment and procedures for determination of the gas-breakthrough pressure of compacted clay materials with preliminary results". MSc. Department of Civil and Geological Engineering, University of Manitoba, Whnipeg, 1995.
- Koirala, SR, Gentry, RW, Mulholland, PJ, Perfect, E, and Schwartz, JS (2010). "Time and frequency domain analyses of high-frequency hydrologic and chloride data in an east Tennessee watershed". In: *Journal of Hydrology* 387 (2010), pp. 256–264. DOI: 10.1016/j.jhydrol.2010.04.014.
- Lara, JA, Lizcano, D, Pérez, A, and Valente, JP (2014). "A General Framework for Time Series Data Mining based on Event Analysis: Application to the Medical Domains of Electroencephalography and Stabilometry". In: *Journal of Biomedical Informatics* (2014). In Press. DOI: 10.1016/j.jbi.2014.06.003.
- Liu, X, Yang, X, and Zheng, N (2012). "Automatic extracellular spike detection with piecewise optimal morphological filter". In: *Neurocomputing* 79 (2012), pp. 132–139. DOI: 10.1016/j.neucom.2011.10.016.
- Maalej, M, Karasaridis, A, Hatzinakos, D, and Pantazopoulou, SJ (1999). "Spectral analysis of sensor data in civil engineering structures". In: *Computers and Structures* 70 (1999), pp. 675–689. DOI: 10.1016/S0045-7949(98)00211-9.
- Mallants, D, Jacques, D, and Perko, J (2007). "Modelling multi-phase flow phenomena in concrete barriers used for geological disposal of radioactive waste". In: *Proceedings of the 11th International Conference on Environmental Remediation and Radioactive Waste Management ICEM2007*. ICEM07-7203. 2007.
- Marschall, P, Horseman, S, and Gimmi, T (2005). "Characterisation of Gas Transport Properties of the Opalinus Clay, a Potential Host Rock Formation for Radioactive Waste Disposal". In: *Oil & Gas Science and Technology* 60 (2005), pp. 121–139.
-

-
- Mattmann, CA (2013). "A vision for data science". In: *Nature* 493 (2013), pp. 473–475.
- Mesri, G and Olson, RE (1971). "Mechanism controlling the permeability of clays". In: *Clays and Clay Minerals* 19 (1971), pp. 151–158.
- Mitchell, JK (1993). *Fundamentals of soil behaviour*. Second Edition. John Wiley and Sons Inc., 1993.
- Mukhopadhyay, S and Ray, GC (1998). "A New Interpretation of Nonlinear Energy Operator and Its Efficacy in Spike Detection". In: *IEEE Transactions on Biomedical Engineering* 45 (1998), pp. 180–187. DOI: 10.1109/10.661266.
- NDA (2010a). *Geological Disposal: An introduction to the generic Disposal System Safety Case*. 2010.
- NDA (2010b). *Radioactive waste in the United Kingdom*. Website. Accessed: 2012–11–26. Nuclear Decommissioning Authority. 2010. URL: <http://www.nda.gov.uk/ukinventory/summaries/index.cfm>.
- NDA (2011). *Interim Storage of Higher Activity Waste Packages - Integrated Approach: Industry Guidance*. Tech. rep. Nuclear Decommissioning Authority, 2011.
- NF-Pro (2006). *NF PRO News*. Newsletter. 2006.
- Norris, S (2009). *Summary of Gas Generation and Migration – Current State-of-the-Art*. Tech. rep. D1.2R. FORGE, 2009.
- Nowak, T, Köster, H, Flentje, R, Sanchez-Herrero, S, and Lege, C (2007). *Lasgit Large Scale Gas Injection Test: Hydraulic tests with surface packer systems*. Tech. rep. IPR-07-14. Federal Institute for Geosciences and Natural Resources, 2007.
- O'Connor, PDT, Newton, D, and Bromley, R (1995). *Practical Reliability Engineering*. Third Edition Revised. John Wiley and Sons, 1995.
- Olivella, S and Alonso, EE (2008). "Gas flow through clay barriers". In: *Géotechnique* 58 (2008), pp. 157–176.

-
- Oppenheim, AV and Schaffer, RW (1975). *Digital Signal Processing*. Prentice-Hall, 1975.
- Orlando, L (2012). "Detecting steel rods and micro-piles: A case history in a civil engineering application". In: *Journal of Applied Geophysics* 81 (2012), pp. 130–138. DOI: 10.1016/j.jappgeo.2012.01.009.
- Ortiz, L, Volckaert, G, Cannière, D, Horseman, ST, Harrington, JF, Impey, M, and Einchcoomb, S (1996). "MEGAS: Modelling and experiments on gas migration in repository host rocks". In: *Project on Effects of Gas in Underground Storage facilities for radioactive waste (PEGASUS project) - Proceedings of a progress meeting held in Rapolano Terme, Italy on 14 and 15 June 1995*. Ed. by B Haijink. EUR 16746 EN. 1996.
- Owen, JS, Eccles, BJ, Choo, BS, and Woodings, MA (2001). "The application of auto-regressive time series modelling for the time-frequency analysis of civil engineering structures". In: *Engineering Structures* 23 (2001), pp. 521–536. DOI: 10.1016/S0141-0296(00)00059-6.
- Pimentel, MAF, Clifton, DA, Clifton, L, and Tarassenko, L (2014). "A review of novelty detection". In: *Signal Processing* 99 (2014), pp. 215–249. DOI: 10.1016/j.sigpro.2013.12.026.
- Posenato, D, Kirpakaran, P, Inaudi, D, and Smith, IFC (2010). "Methodologies for model-free data interpretation of civil engineering structures". In: *Computers & Structures* 88 (2010), pp. 467–482. DOI: 10.1016/j.compstruc.2010.01.001.
- Pusch, R and Carlsson, T (1985). "The physical state of pore water of Na smectite used as a barrier component". In: *Engineering Geology* 21 (1985), pp. 257–265.
- Pusch, R and Forsberg, T (1983). *Gas migration through bentonite clay*. Tech. rep. TR-83-71. University of Luleå, 1983.
- Pusch, R and Karnland, O (1986). *Aspects of the Physical State of Smectite-adsorbed Water*. Tech. rep. TR-86-25. Swedish Geological Co, 1986.
- Pusch, R, Hökmark, H, and Börgesson, L (1987). *Outline of models of water and gas flow through smectite clay buffers*. Tech. rep. TR-87-10. Swedish Geological Co, 1987.
-

-
- Reardon, EJ (1995). "Anaerobic Corrosion of Granular Iron: Measurement and Interpretation of Hydrogen Evolution Rates". In: *Environmental science and technology* 29 (1995), pp. 2936–2945.
- Reikard, G (2012). "Forecasting volcanic air pollution in Hawaii: Tests of time series models". In: *Atmospheric Environment* 60 (2012), pp. 593–600. DOI: 10.1016/j.atmosenv.2012.06.040.
- Rocco S, CM (2013). "Singular spectrum analysis and forecasting of failure time series". In: *Reliability Engineering & System Safety* 114 (2013), pp. 126–136. DOI: 10.1016/j.ress.2013.01.007.
- Rodwell, WR (2005). *Summary of a GAMBIT Club Workshop on Gas Migration in Bentonite, Madrid 29–30 October, 2003*. Tech. rep. TR-05-13. Svensk Kärnbränslehantering AB, 2005.
- Rodwell, WR, Harris, AW, Horseman, ST, Lalieux, P, Müller, W, Ortiz Amaya, L, and Pruess, K (1999). *Gas Migration and Two-Phase Flow through Engineered and Geological Barriers for a Deep Repository for Radioactive Waste*. Status Report EUR 19122 EN. EC/NEA, 1999.
- Sandén, T (2003). *Project Lasgit: Deposition Hole Instrumentation*. AP TD F122-03-024. SKB, 2003.
- Sedighi, M, Bennett, DP, Masum, SA, Thomas, HR, and Johansson, E (2014). *Analysis of temperature data at the Olkiluoto*. Tech. rep. 2013-58. Posiva O, 2014.
- Sirca Jr, GF and Adeli, H (2012). "System identification in structural engineering". In: *Scientia Iranica* 19 (2012), pp. 1355–1364. DOI: 10.1016/j.scient.2012.09.002.
- SKB (2003). *Deep repository for spent nuclear fuel*. Svensk Kärnbränslehantering AB, 2003.
- SKB (2006). *Long-term safety for KBS-3 repositories at Forsmark and Laxemar — a first evaluation: Main Report of the SR-Can project*. Tech. rep. TR-06-09. Svensk Kärnbränslehantering AB, 2006.
- SKB (2008). *Äspö Hard Rock Laboratory: Annual report 2007*. Tech. rep. TR-08-10. Svensk Kärnbränslehantering AB, 2008.

-
- Sonali, P and Nagesh Kumar, D (2013). "Review of trend detection methods and their application to detect temperature changes in India". In: *Journal of Hydrology* 476 (2013), pp. 212–227. DOI: 10.1016/j.jhydrol.2012.10.034.
- SONE (2012). *SONE - Supporters of Nuclear Energy*. Website. Accessed: 2012–10–02. Supporters of Nuclear Energy. 2012. URL: <http://www.sone.org.uk/>.
- Svemar, C and Pusch, R (2000). *Äspö Hard Rock Laboratory: Prototype Repository*. Tech. rep. IPR–00–30. Svensk Kärnbränslehantering AB, 2000.
- Tanai, K, Kanno, T, and Gallé, C (1997). "Experimental study of gas permeabilities and breakthrough pressures in clays". In: *Scientific Basis for Nuclear Waste Management XX*. Ed. by WJ Gray and IR Triay. Vol. 465. MRS Symposium Proceedings. Materials Research Society. 1997, pp. 995–1002.
- Thomas, HR, Cleall, PJ, Chandler, N, Dixon, D, and Mitchell, HP (2003). "Water infiltration into a large-scale in-situ experiment in an underground research laboratory". In: *Géotechnique* 53 (2003), pp. 207–224. DOI: 10.1680/geot.2003.53.2.207.
- Thomas, HR, Cleall, PJ, Dixon, D, and Mitchell, HP (2009). "The coupled thermal-hydraulic-mechanical behaviour of a large-scale in situ heating experiment". In: *Géotechnique* 59 (2009), pp. 401–413. DOI: 10.1680/geot.2009.59.4.401.
- Thomas, HR, Vardon, PJ, and Cleall, PJ (2014). "Three-dimensional behaviour of a prototype radioactive waste repository in fractured granitic rock". In: *Canadian Geotechnical Journal* 51 (2014), pp. 246–259. DOI: 10.1139/cgj-2013-0094.
- Tukey, JW (1962). "The Future of Data Analysis". In: *The Annals of Mathematical Statistics* 33 (1962), pp. 1–67.
- Tukey, JW (1977). *Exploratory Data Analysis*. Ed. by F Mosteller. Addison-Wesley, 1977.
- UK Gov. (2006). *Response to the Report and Recommendations from the Committee on Radioactive Waste Management (CoRWM)*. Tech. rep. The UK Government and the devolved administrations, 2006.

-
- Vankerckhoven, P (1998). *Radioactive waste categories - current position (1998) in the EU Member States and in the Baltic and Central European countries*. Tech. rep. EUR 18324 EN. European Commission, 1998.
- Vardon, PJ, Thomas, HR, Masum, SA, Chen, Q, and Nicholson, D (2014). "Simulation of repository gas migration in a bentonite buffer". In: *Engineering and Computational Mechanics* (2014), pp. 1–10. DOI: 10.1680/eacm.12.00018.
- Vardon, PJ, Masum, SA, and Thomas, HR (2010). *Initial investigation of gas pressure data ranges for use in the post closure Disposal System Safety Case*. Tech. rep. Geoenvironmental Research Centre, 2010.
- Velasco, C (2003). "Nonparametric frequency domain analysis of nonstationary multivariate time series". In: *Journal of Statistical Planning and Inference* 116 (2003), pp. 209–247. DOI: 10.1016/S0378-3758(02)00235-5.
- Villar, MV and Lloret, A (2004). "Influence of temperature on the hydro-mechanical behaviour of a compacted bentonite". In: *Applied Clay Science* 26 (2004), pp. 337–350. DOI: 10.1016/j.clay.2003.12.026.
- Visser, H (2004). "Estimation and detection of flexible trends". In: *Atmospheric Environment* 38 (2004), pp. 4135–4145. DOI: 10.1016/j.atmosenv.2004.04.014.
- Weisstein, EW (2013a). *Gaussian Function*. Website. Accessed: 2013–12–12. MathWorld. 2013. URL: <http://mathworld.wolfram.com/GaussianFunction.html>.
- Weisstein, EW (2013b). *Laplace Distribution*. Website. Accessed: 2013–12–12. MathWorld. 2013. URL: <http://mathworld.wolfram.com/LaplaceDistribution.html>.
- Weisstein, EW (2013c). *Standard Deviation*. Website. Accessed: 2013–12–12. MathWorld. 2013. URL: <http://mathworld.wolfram.com/StandardDeviation.html>.
- Western, R (1994). *Out of their depth: The inadequacies of Nirex research on the safety of Nuclear Waste Disposal*. Tech. rep. Friends of the Earth, 1994.

- Wilson, SB and Emerson, R (2002). "Spike detection: a review and comparison of algorithms". In: *Clinical Neurophysiology* 113 (2002), pp. 1873–1881. DOI: 10.1016/S1388-2457(02)00297-3.
- Xiao, J and Zhang, J (2011). "Analysis and application of automatic deformation monitoring data for buildings and structures of mining area". In: *Transactions of Nonferrous Metals Society of China* 21 (2011), S516–S522.
- Xu, G, Wang, J, Zhang, Q, Zhang, S, and Zhu, J (2007). "A spike detection method in EEG based on improved morphological filter". In: *Computers in Biology and Medicine* 37 (2007), pp. 1647–1652. DOI: 10.1016/j.compbimed.2007.03.005.

