

1 QuakeMigrate: Earthquake Onset Detection and
2 Location Method

3

4 August 6, 2019

5 Jonathan D. Smith 12, Tom Winder 1, Thomas S. Hudson 13, Conor Bacon

6 1, Tim Greenfield 1, Julian Drew 4, Robert S. White 1

7

8 1 Department of Earth Sciences, University of Cambridge, United Kingdom.

9 2 California Institute of Technology, Pasadena, California

10 3 NERC British Antarctic Survey, Cambridge, UK.

11 4 SIGMA3, Houston, Texas, United States of America.

12 jds70@alumni.cam.ac.uk - Seismological Research Letters

13

Abstract

14 Traditionally, detecting microseismic events has been undertaken by

15 using event triggers individually at each station within a seismic network.

16 This generally results in many false triggers due to the presence of in-
17 coherent noise sources, such as anthropogenic noise. Here, we present

18 a new modular based approach to detect microseismicity using a coherence
19 based method, called QuakeMigrate. The method involves calculating
20 onset functions for P- and S-waves that mark the arrival of seismic
21 phases, before back-propagating these through time and space to find a
22 global maximum in the coalescence function. This method differs from
23 previous methods as it is open source using standard python libraries,
24 provides a realistic estimate of the spatial uncertainty of a given event,
25 estimates the arrival time uncertainty at each station, and provides a cus-
26 tomisable modular structure to allow for future development. Throughout
27 this report we outline the inversion procedure before demonstrating the
28 software's usage for three differing spatial scale examples: Antarctic ice-
29 quakes at the Rutford Ice Stream, volcano-tectonic earthquakes associated
30 with the Bardarbunga-Holuhraun dyke intrusion, and induced seismicity
31 in the Groningen region. These examples illustrate the effectiveness of the
32 QuakeMigrate for microseismic detection over a range of spatial scales and
33 with different background noise sources.

34 1 Introduction

35 Traditional event detection procedures commonly consist of three stages: pick-
36 ing and classifying seismic phase arrivals at individual seismic stations; asso-
37 ciating these time picks with distinct events; and finally inverting for origin
38 time and location. However, these methods are often challenging when applied
39 to microseismic datasets. Picking a satisfactory fraction of low signal-to-noise
40 ratio phase arrivals at individual stations without introducing a large num-

ber of false triggers is challenging due to local noise sources (anthropogenic, meteorological etc.), and short inter-event times introduce issues with event association. This can lead to false detections (artifacts) and mis-location of events if mis-picks are allowed to contribute to the final location. An alternative approach to these traditional detection methods instead exploits the coherence of energy detected across the whole network of sensors to detect and locate earthquakes. This was first implemented in the Source Scanning Algorithm (SSA) [*Kao, H. and Shan, S. (2004)*]. Subsequent studies have demonstrated the advantages in robustness to noise, and in removing the reliance on phase picking. Applications include monitoring anthropogenically induced microseismicity ([*Grigoli et al. (2018)*]), volcano seismology ([*Drew et al. (2013)*]) and landslides ([*Hibert et al. (2014)*]).

We introduce QuakeMigrate, a new, robust, computationally efficient and modular approach to earthquake detection and location using a coherence method. The current framework is suitable for implementation into a research workflow to process archived waveform data and produce a catalogue of microearthquakes, along with a comprehensive output of statistics for further event filtering and other processing. The open source and modular nature of the software provides scope for continuous improvement and addition of further pre-processing and post-processing steps. These adaptations can be made for personal use or an individual can make overall contributions to the QuakeMigrate software, benefiting the whole user base.

64 **2 Theory**

65 A seismic inversion procedure is formulated for the detection and location of seis-
66 mic events using a continuous network seismic dataset. This procedure expands
67 on the prior work of [*Drew et al. (2013)*], with the formulation of a waveform
68 onset function stacking technique. Outlined below is the theory behind the for-
69 mulation of the inversion procedure.

70

71 We determine onset functions for each seismic station by using a short-term-
72 average (STA) to long-term average (LTA) function, representing a form of noise
73 normalization of the original seismic signal. The onset function for the P-wave
74 is determined from the vertical component of each seismic station and the S-
75 wave onset function determined from the combined horizontal components. In
76 source scanning procedures the onset function is understood to represent the
77 probability density function for the timing of an arrival. The arrival time picks
78 are determined by the first derivative of the onset function, with timing uncer-
79 tainty determined as the second derivative of the logarithm of the onset function.
80 [*Drew et al. (2013)*] demonstrated that the second derivative of the logarithm
81 can be approximated by a gaussian distribution. Using this approximation, we
82 set the logarithm of the onset function to represent the seismic arrival error for
83 each seismic station. To find the coherence of signals between seismic stations
84 a ‘coalescence’ function is computed as the sum of the logarithm of the onset

85 functions for the seismic stations. The coalescence function f_c is given by:

$$f_c(t, \vec{s}) = \prod f_{r_i}(t + t_{g_i}(\vec{s})) = \exp\left(\sum_{i=1}^n L_{R_i}(t + t_{g_i}(\vec{s}))\right) \quad (1)$$

86 where t and \vec{s} are the time and space variables, f_{r_i} the onset function for station
87 i , t_{g_i} the travel-time from a spatial location to station i and L_{R_i} is the natu-
88 ral logarithm of f_{r_i} . The coalescence function is a four-dimensional function
89 dependent on spatial and temporal terms. As such it can be reduced to an
90 instantaneous spatial map, $\vec{s}(t)$, and maximum value at each time step, $\hat{f}_c(t)$.
91 An additional term can be computed that represents the maximum coalescence
92 value at each time step divided by the mean of the instantaneous spatial map,
93 this represent the normalised maximum coalescence value $\hat{f}_{Nc}(t)$. The nor-
94 malised coalescence value is insensitive to sudden changes in the global energy of
95 the system, these are typically from sudden changes in the number of stations
96 or the propagation of the energy through the volume after the earthquake.

97 A event is triggered when the maximum coalescence value or normalised
98 maximum coalescence value exceeds some predefined detection threshold level,
99 with a time window defined for the period for the exceedence of this threshold
100 value (shown by the yellow region in Figure 1). A marginal minimum repeat pe-
101 riod is applied to these triggered events to reduce multiple triggers on the same
102 event. This minimum repeat windows are represented by the greyed regions in
103 (Figure 1). If the minimum repeat windows overlap between events, then the
104 events are then combined into a single event. As the exact time of the event
105 is unknown a marginal time window equal to the expected time uncertainty
106 is defined about the maximum coalescence value, as is represented by the red

107 region in Figure 1.

108

109 The coalescence procedure is re-implemented to determine the integration of
110 the instantaneous spatial map ($\vec{s}(t)$) across the marginal window for each event
111 to define the event, to represent the spatial map as

$$f_s(\vec{s}) = \int_{t_D-ta}^{t_D+ta} f_c(\tau, \vec{s}) d\tau \quad (2)$$

112 where t_D is the proposed event time, t_a is the marginal time window and f_s is the
113 spatial function. The hypocentral location is formulated in three methods: max-
114 imum coalescence value is determine from the spatial map interpolated using a
115 spline function to remove the effect from gridding; a local Gaussian determine by
116 fitting a three-dimensional Gaussian function about the maximum coalescence,
117 with the centre of the Gaussian ellipse defining the location and one standard
118 deviation errors returned for the X-, Y- and Z- direction; and, a global Gaussian
119 calculated using a weight covariance matrix method for the entire spatial map.

120

121 Uncertainties associated with the phase arrival time triggers at individual
122 stations are estimated using Gaussian fitting. The Gaussian is fitted about the
123 maximum of the logarithm of the onset function, L_{R_i} , i.e. the picked arrival
124 time, for each station. This is performed within a time window about the
125 modelled arrival time for the earthquake at the station. The time window is
126 defined by t_w , which is the combination of the marginal time window t_a and a
127 percentage error of the total travel-time $t_v b$. The Gaussian is only returned if
128 the peak value is above a percentage of the variance of the onset function. The

¹²⁹ uncertainty in the arrival time of a particular station i , for phase j is given by,

$$\delta t_{i,j} = (t - t_{a,i,j}) \left(\frac{1}{2 \ln(\frac{A}{L_{R_i}(t)})} \right)^{\frac{1}{2}} \quad (3)$$

¹³⁰ where t is time, $L_{R_i}(t)$ is the logarithm of the onset function at time t , $t_{a,i,j}$ is
¹³¹ the arrival time at station i , for phase j and A is a fitting parameter. $\delta t_{i,j}$, $t_{a,i,j}$
¹³² and A are allowed to vary, minimizing the misfit between this equation and the
¹³³ data, in order to find the best fitting value of $\delta t_{i,j}$, the arrival time uncertainty.

¹³⁴

¹³⁵ 3 Software Functionality

¹³⁶ 3.1 Processing Procedures

¹³⁷ QuakeMigrate uses a modular process for the detection and location of seismic
¹³⁸ events. An outline of this procedure is displayed in Figure 2. QuakeMigrate can
¹³⁹ be broadly separated into five modules: Travel Time - Determination or Loading
¹⁴⁰ of travel-time look-up tables for each seismic station within a predefined velocity
¹⁴¹ structure; Seismic Data - Definition of the continuous seismic data structure to
¹⁴² load; Detect - Determination of the maximum coalescence value for a reduced
¹⁴³ size model; Trigger - Determination of time periods which exceed a given thresh-
¹⁴⁴ old value; and Location - re-analysis of triggered events to refine event location.
¹⁴⁵ The main processing stages are the Detect and Location stages, interfacing with
¹⁴⁶ a series of C modules for the efficient calculation of energy coalescence over a
¹⁴⁷ given time and space domain. The module structure of the software and the
¹⁴⁸ extensive output information allows the user to provide different onset function,

149 travel-time grids and picking methods to allow for continuous development of
150 the software package and help improve the user experience.

151 **3.2 Travel Time - Model Setup and travel-time tables**

152 Travel-time look-up tables for P- and S-wave arrivals are constructed for each
153 seismic station for some user defined search grid. These travel-time look-up
154 tables can be formulated using homogeneous or 1D velocity structures, by using
155 the QuakeMigrate inbuilt functions taking advantage of the NonLinLoc Eikonal
156 procedure (*[Lomax, A.(2005)]* or *skfmm* Python Fast Marching Method (CITE,
157 *[Rawlinson, N. and Sambridge, M. (2005)]*) travel-time formulation. For more
158 complex problems this procedure can load external three-dimensional travel-
159 time tables of the NonLinLoc format. The travel-time look-up tables are saved
160 into a single file which only has to be formulated once, with the allowance to
161 transfer between different systems.

162 **3.3 Seismic Data - Defining continuous data information**

163 QuakeMigrate uses the ObsPy [*Beyreuther et al.(2010)*] python package for load-
164 ing continuous seismic data. Since archive definition may be different for each
165 seismic processing group, the user defines the archive structure using the func-
166 tion 'path_structure'. This section of the software then defines a method for
167 the conversion of the archive into a standard ObsPy structure to be used by
168 QuakeMigrate. Any pre-processing of data can be undertaken prior to this
169 stage, by the addition of user specific functions. However, we suggest conducting

170 quality-control (QC) on the original data before processing using QuakeMigrate.

171

172 QuakeMigrate requires the sampling rate for all stations to be identical
173 an encompasses a resampling procedure dependent on the user defined 'sam-
174 pling_rate'. If the sampling rate of each seismic stream is the same as the user
175 defined 'sampling_rate' then no resampling is applied. A band-pass filtering
176 procedure is implemented within QuakeMigrate through the definitions of user
177 variables 'p_bp_filter' and 's_bp_filter' variables, representing the bandpass val-
178 ues for the P- and S-filters respectively, in the form LowPass, HighPass and
179 Corners respectively.

180 **3.4 Detect - Detection of coherent signals**

181 An initial coalescence detection procedure is implemented on decimated travel-
182 time grids to reduce the computational demand for the detection of events from
183 the continuous seismic data. This procedure relies on the parameters defined in
184 prior sections in addition to a user defined 'decimate', describing the decimation
185 factor of the grid in the X-, Y-, and Z-direction respectively. This stage is imple-
186 mented across a series of time-steps with size defined by the variable 'time-step'.

187

188 The output generated from the detection stage include a mini-seed output
189 'scancmseed' with traces of maximum coalescence values, normalised coalescence
190 values, and the changing location of the maximum coalescence value in X, Y
191 and Z.

192 **3.5 Trigger - Event Triggering above threshold value**

193 Events are triggered from the 'scancseed' output by using the class 'Trigger'.
194 An event is triggered when the coalescence value exceeds some user defined
195 detection threshold, defined by the variable 'detection_threshold', with the ori-
196 gin time uncertainty represented by the user defined marginal window, defined
197 by the variable 'marginal_window'. To minimise the multiple triggers from the
198 same event a user defined minimum repeat is specified using the variable 'min-
199 imum_repeat'. The theory about each of these variables is outlined in Section
200 2.

201

202 A list of detected events and their threshold values are outputted in a 'TriggeredEvents.csv'
203 file format. This file gives an output of EventNumber, Maximum Coalescence
204 Time, Maximum Coalescence Value, Maximum Coalescence Location XYZ,
205 Marginal Minimum Time, Marginal Maximum Time and Event ID. In addi-
206 tion, a output '.pdf' file is generated giving the detected time-windows for each
207 of the events, corresponding decimated location of the maximum coalescence
208 value for each of these detected events, and the changing station geometry across
209 the time-period of investigation (Figure 3). This stage is computationally effi-
210 cient and can be re-run multiple times to identify the effects of varying detection
211 threshold, marginal window, and minimum repeat.

212

213 **3.6 Locate - Refining event location**

214 Using the ‘_TriggeredEvents.csv’ the events are relocated for the triggered marginal
215 window. These locations use the parameters defined for the detection stage, to
216 allow the user to locate using differetn filters, onset windows and decimation
217 factors. A ‘.event’ file is returned for each event, containing columns of Date-
218 Time, Maximum Coalescence, Maximum Coalescence Location X,Y,Z, Local
219 Gaussian location X,Y,Z, Local Gaussian fit errors X,Y,Z, Global Gaussian lo-
220 cation X,Y,Z and Global Gaussian location Error X,Y,Z. In addition, a ‘.stn’
221 file is returned containing modelled travel times, phase arrival picks and pick
222 uncertainties for each of the seismic stations. Additional user defined variables
223 can be used to save coalescence grids, coalescence videos coalescence location,
224 and arrival plots for each seismic instrument. (‘plot_coal_grid’,‘plot_coal_video’,
225 ‘plot_coal_picture’ and ‘plot_coal_trace’ respectively).

226

227 **4 Detection and Location Examples**

228 To demonstrate the usage of QuakeMigrate for the detection and location of
229 seismic events, we outline the processing procedure for three examples: ice-
230 quakes at the Rutford Ice Stream, Antarctica; volcano-tectonic earthquakes as-
231 sociated with the Bardarbunga-Holuhraun dyke intrusion, Iceland; and induced
232 seismicity in the Groningen region, the Netherlands. These applications repre-
233 sent variations in station distribution, deployment duration of seismic networks,

234 magnitude ranges, rate of seismicity, amplitude variation of noise sources, and
235 spread of location depths. Each case has its own challenges. Here we demon-
236 strate how QuakeMigrate can be adapted to efficiently and robustly detect and
237 locate earthquakes in each scenario. A summary of the parameters used in each
238 analysis are presented in Table 1, along with the computation costs and memory
239 requirements shown in Table 2

240 4.1 Antarctic Basal Icequakes

241 Icequakes are the release of seismic energy due to the movement or fracture of
242 ice. Basal icequakes occur at the interface between a sliding glacier and the un-
243 derlying till or bedrock. Detecting these events can be challenging due to noise
244 sources such as surface crevassing ([*Hudson et al. (2019)*]). QuakeMigrate can
245 provide an effective method for detecting basal icequakes because other sources
246 of glacial seismic noise are likely to be filtered out during the coalescence stage
247 of detection and efficiently tuning initial search parameters. The case study
248 we use to illustrate this is based on the work detailed in [*Hudson et al. (2019)*],
249 here just summarizing some of their results.

250
251 Data was recorded on 10 geophones deployed on the Rutford Ice Stream,
252 West Antarctica, in 2009 ([*Smith et al. (2015)*]). The Rutford Ice Stream moves
253 at $\sim 400 \text{ m a}^{-1}$ ([*Murray et al. (2007)*]), with basal icequakes associated with
254 glacial sliding occurring at a rate of thousands per day ([*Smith et al. (2015)*]).
255 Figure 5 shows icequakes detected using QuakeMigrate for the 20th to 21st Jan-

256 uary 2009 ([*Hudson et al. (2019)*]). All events detected using QuakeMigrate are
257 relocated using NonLinLoc ([*Lomax, A.(2005)*]).

258

259 The parameters used to detect these icequakes are given in Table 1. The
260 critical icequake detection parameters are the frequency filter parameters and
261 the STA/LTA ratios. We use a frequency range of 20 to 200 Hz since basal
262 icequakes typically have higher peak frequencies in their spectra than microseis-
263 micity observed in other environments. The choice of an STA/LTA ratio of 0.01
264 s to 0.25 s also reflects this. We use a coalescence detection threshold of 1.75,
265 although this value is network and site dependent. The icequakes are detected
266 using a homogeneous velocity model with $v_P = 3841 \text{ m s}^{-1}$ and $v_S = 1970 \text{ m}$
267 s^{-1} , as in [*Smith et al. (2015)*].

268

269 The location of icequake seismicity near the ice-bed interface (shown by the
270 grey line in Figure 5) using a homogeneous velocity model indicates how effective
271 QuakeMigrate is at detecting real events and their associated phase arrivals.

272 QuakeMigrate is particularly effective for detecting basal icequakes because
273 the software allows for quickly testing many different sets of band-pass filter
274 and STA/LTA parameters. This is important as icequake properties, such
275 as source frequencies and the attenuation characteristics of medium, are often
276 poorly known at the start of an investigation, and so trial and error to obtain
277 an initial catalogue of icequakes would otherwise be challenging and tedious.
278 Future development of QuakeMigrate could also allow for the automation of

279 such a parameter testing process. This could be implemented simply by au-
280 tomatically running QuakeMigrate over a number of different frequency bands
281 and *STA/LTA* values and selecting the most effective set of parameters that
282 produced a catalogue maximising the number of detected events while minimis-
283 ing the number of false triggers. It is also conceivable that machine learning
284 techniques could be implemented, to allow QuakeMigrate to automatically learn
285 the optimal search parameters, saving investigators time while also producing
286 the best icequake location catalogue for regions where little prior knowledge of
287 potential seismic sources is known.

288 **4.2 Bardarbunga Dyke Earthquakes**

289 Plate spreading and volcanism in Iceland is focussed within three neo-volcanic
290 zones that lie along the axis of the Mid-Atlantic ridge. The University of Cam-
291 bridge has operated a dense local seismic network in the Northern Volcanic Zone
292 continuously since 2009 to study the extensive microseismicity in the neighbour-
293 ing Askja and Bardarbunga volcanic systems. Earthquakes caused by geother-
294 mal activity and tectonic deformation in the shallow crust, and magma and fluid
295 movement both shallow and deep have been detected, ranging from -1*ML*5, at
296 a background rate of 4000 events per year Greenfield2018. Network cover-
297 age varies both spatially and through time, due to shifting research focusses,
298 the difficulty in installing instruments on Vatnajokull glacier and challenging
299 field conditions through the Icelandic winter. The network comprises \sim 70 3-
300 component broadband seismometers at its peak. An automatic algorithm is

301 essential to process such a large volume of seismic data containing such a high
302 frequency of events.

303

304 It is also essential that this algorithm produces a robust catalogue with
305 comprehensive statistics for filtering and further post-processing, as manually
306 checking for artefacts and mislocations is prohibitively time consuming. This is
307 exacerbated by the irregular pattern of seismicity through time, which makes
308 it hard to distinguish artefacts from new areas of activity. An example is the
309 surge in seismicity during the 2014 Bardarbunga-Holuhraun dyke intrusion -
310 c.30,000 earthquakes were induced within the subsiding caldera and along the
311 previously aseismic dyke pathway as magma propagated 48km laterally over 13
312 days, before erupting in the Holuhraun plain.

313

314 The results from processing 1 hour of data between midnight and 1am on
315 the 24th August 2014, midway through the intrusion period, are presented here
316 to illustrate how QuakeMigrate can be utilised to maximise the scientific dis-
317 coveries from such a dataset.

318

319 A travel-time look-up table (LUT) was generated using NonLonLoc Lo-
320 max2000 from the 1D velocity model used by CITATION on a 3D grid sur-
321 rounding Bardarbunga spanning 140 x 130 x 45 km with node spacing of 0.5
322 km in all dimensions. using NonLinLoc, and imported as a QuakeMigrate LUT.
323 MiniSEED data read in directly from a bespoke data archive was resampled to

324 50Hz and bandpass filtered between 2-16 Hz and 2-12 Hz for P and S phases
325 respectively, before generating an onset function based on an STA/LTA func-
326 tion with short and long time windows of 0.2 s and 1 s for both P and S phases.
327 For the Detect stage, the LUT grid was decimated by a factor of [5,5,2] and
328 processed in 120 second time chunks, resulting in a runtime of x hours running
329 in parallel on 8 processors.

330

331 Detecting and locating microseismic events recorded in this dataset presents
332 three main challenges: (1) the large range of earthquake magnitudes (and hence
333 wavetrain amplitudes and durations) combined with (2) the extremely high fre-
334 quency of events, with inter-event times commonly shorter than event-station
335 travel times; (3) the large variation in station spatial density, and hence sensi-
336 tivity, across the search grid.

337

338 Issues (1) and (2) present a trade-off: it is desirable to miss as few events
339 as possible by setting a small minimum inter-event time, but necessary to im-
340 plement this condition to prevent multiple events being triggered by a single
341 large-magnitude, long-wavetrain event. The dynamic triggering window im-
342 plemented in QuakeMigrate successfully avoids false detections in the period
343 around a ML 4.01 earthquake in Bardarbunga caldera while also detecting two
344 events of ML 1 with origin times separated by just 8 seconds FIGURES. P
345 and S phase arrivals for these two events, and thus peaks in the onset function,
346 will be very closely spaced in time at some stations, but will be migrated away

347 from each respective hypocentre at their respective origin times, so will not in-
348 troduce bias. In contrast, a detection approach based on first reducing the data
349 to phase-arrival picks would have to make the correct event association before
350 inverting for location, and would therefore struggle with this scenario.

351

352 The highly uneven spatial sampling provided by the Cambridge network over
353 the region surrounding Bardarbunga inevitably leads to some false detections in
354 regions of poor coverage. CAUSE DETERMINED BY RUNNING SUBSET OF
355 DATA AND INSPECTING OPTIONAL PLOTS. This most commonly occurs
356 when high amplitude arrivals from large events located close to the dense part
357 of the network are back-propagated earlier than the event origin time. Far away
358 from the dense part of the network, these back-propagated signals once again
359 begin to overlap as the wavefronts become sub-parallel, and in quiet periods this
360 signal coalesce sufficiently to exceed the trigger threshold. However, the statis-
361 tics output by QuakeMigrate make these artefacts straightforward to identify
362 and remove. 00:08:50; 00:28:36; 00:32:44; 00:38:24; 00:40:30; 00:47:05; 00:49:14;
363 00:53:38; 00:58:02 - covariance of entire grid - peak-fitting error - time-picks only
364 made at far stations, and associated with real event $-i$ relocation with NLLOC
365 will identify this event as a multiple. NEED TO CHECK

366 4.3 Induced Seismicity in the Groningen Region

367 The Groningen gas field, situated in the north-east of the Netherlands has been
368 in production since 1963, with 70% of the estimated 2800 billion cubic meters of

³⁶⁹ gas now extracted ([*Bourne et al.(2014) Bourne, Oates, Van Elk, and Doornhof*]).

³⁷⁰ Prior to gas extraction the region was considered aseismic with no recorded his-
³⁷¹ torical events. However, since the 1990s small magnitude earthquakes have been
³⁷² detected, with these shallow events causing non-structural building damage and
³⁷³ public concern ([*Dost et al.(2012) Dost, Goutbeek, Eck, and Kraaijpoel*]).

³⁷⁴

³⁷⁵ This industry dataset is located in a region of high anthropogenic noise,
³⁷⁶ which would normally cause a catalogue to have a large number of false trig-
³⁷⁷ gers. QuakeMigrate stacks across many stations, limiting detected events to
³⁷⁸ those that are coherent across multiple stations, thus reducing the number of
³⁷⁹ false triggers on detection. Detection and location of coherent events in in-
³⁸⁰ duced seismicity is important as other techniques such as template matching
³⁸¹ act to spatially bias seismicity. QuakeMigrate allows the detection threshold to
³⁸² be altered quickly to search for a maximum coalescence value that recovers an
³⁸³ earthquake catalogue with the lowest false trigger number, without bias of the
³⁸⁴ initially detected seismicity.

³⁸⁵

³⁸⁶ QuakeMigrate computationally efficiency to calculate hypocentral locations
³⁸⁷ outlines a key usage for hazard assessment in induced seismicity scenarios. Stan-
³⁸⁸ dard induced seismicity processing procedures incorporate a full-waveform inver-
³⁸⁹ sion procedure, such a method is computationally expensive limiting such a tech-
³⁹⁰ niques usage to areas with sparse seismic catalogues. We demonstrate QuakeMi-
³⁹¹ grate's usage for the Groningen case study by formulating travel-time look-up for

392 a complex 3-D velocity model of the Groningen region ([*Nederlandse Aardolie Maatschappij*(2016)]).
393 We decimate the travel-time tables by a factor of 10 in the horizontal directions
394 to reduce the initial computational cost, with the final tables having a dimen-
395 sion size of $500m \times 500m \times 50m$, in the x-,y- and z-dimensions. QuakeMigrate is
396 implemented to detect and locate events across the period of Jan-2018 to March-
397 2018. A example of coalescence for the earthquake at 14:00:52 8/Jan/2018 is
398 shown in Figure 4; demonstrating a coherent hypocentral location (Figure 4a),
399 and P- and S-waves first arrivals that are consistent with the waveforms (Figure
400 4b).

401

402 Remaining uncertainty on the hypocentral location can be attributed to
403 limitations in a P- and S-wave ray-tracing procedure, with seismic energy dis-
404 tributed along multiple ray paths. QuakeMigrate currently incorporates the P-
405 and S-wave travel-times determined by the shortest ray path. If multipathing is
406 present earthquake locations can have large spatial uncertainty due to the lack
407 of coalescence between seismic arrivals. Wagner et al (2013) demonstrated a
408 method of computing multiple travel-time paths for each source location formu-
409 lating separate travel-time tables for each independent phase. Such an approach
410 could be incorporated into QuakeMigrate, with coalescence determined across
411 multiple paths, reducing hypocentral error even further.

412

413 **5 Conclusions**

414 We have introduced QuakeMigrate, a Python and C based modular structured
415 software, for microseismic earthquake detection and location. The open source
416 and modular nature of QuakeMigrate provides scope for continuous improve-
417 ment and addition of further pre- and post-processing steps, enabling contribu-
418 tions from other users that benefit the entire user community. QuakeMigrate
419 is suitable for implementation into a research framework to process continuous
420 seismic data, generating a catalogue of earthquakes with comprehensive physi-
421 cally justified and statistically rigorous outputs. We demonstrate the usage of
422 this software for three differing scenarios: Antarctic icequakes, volcano-tectonic
423 earthquakes associated with the Bardarbunga-Holuhraun dyke intrusion and
424 induced seismicity in the Groningen region. The scenarios demonstrate how
425 QuakeMigrate can effectively detect and locate microseismicity for a range of
426 spatial scales, varying network geometries, a range of earthquake magnitudes,
427 differing rates of seismicity, different seismic noise levels, and a range of hypocen-
428 tral depths.

429 **6 Data and resources**

430 The QuakeMigrate software is available from <https://github.com/Ulvetanna/QuakeMigrate>
431 for research and teaching purposes only, under a MIT licensing agreement. For
432 additional information please contact jds70@alumni.cam.ac.uk. Scientific mod-
433 ules included within this software NumPy, SciPy and Obspy are available at

434 <https://www.numpy.org>, <https://www.scipy.org>, and <https://www.obspy.org>.
435 Alterations and adaptations to the source code are regularly made to include
436 advances and exciting new features, for changes and adaptations please visit
437 <https://github.com/Ulvetanna/QuakeMigrate> for more info.
438 J.D.S. is supported is supported by the Engineering and Physical Sciences
439 Research Council with CASE support from Shell Global Solutions International
440 B.V. Groningen waveform data were provided by Koninklijk Nederlands Me-
441 teorologisch Instituut. The Iceland data was collection under the NERC Geo-
442 physical Equipment Facility Loan 868 and 1022. The Rutherford data collection
443 was funded by NERC grant NE/B502287/1 and NERC Geophysical Equipment
444 Facility Loan 852. University of Cambridge Earth Sciences contribution number
445 esc.XXXX.

446 References

447 [Beyreuther *et al.*(2010)] Beyreuther, M., Barscher, R., Krischer, L., Megies,
448 T., Behr, Y., and Wassermann, J. (2010), ObsPy: A Python Toolbox for
449 Seismology, *Seismological Research Letters*, 81(3), 530–533, doi:10.1785/
450 gssrl.81.3.530.
451 [Bourne *et al.*(2014)] Bourne, S. J.,
452 S. J. Oates, J. Van Elk, and D. Doornhof (2014), A seismological model
453 for earthquakes induced by fluid extraction from a subsurface reservoir,

- 454 *Journal of Geophysical Research: Solid Earth*, **119**(12), 8991–9015, doi:
455 10.1002/2014JB011663.
- 456 [Dost *et al.*(2012)Dost, Goutbeek, Eck, and Kraaijpoel] Dost, B., F. Goutbeek,
457 T. V. Eck, and D. Kraaijpoel (2012), Monitoring induced seismicity in the
458 North of the Netherlands, *Scientific status report*, <http://www.sciamachy-validation.org/bibliotheek/knmipubWR/WR2012-03.pdf>.
- 459
- 460 [Drew *et al.* (2013)] Drew, J., White, R.S., Tilmann, F. and J. Taraseqicz
461 (2013), Coalescence microseismic mapping, *Geophys. J. Int.*, doi:10.1093/
462 gji/ggt331.
- 463 [Greenfield *et-al.* (2018)] Greenfield, T., White, R.S., Winder, T., gstsdtir, T.
464 (2018), Seismicity of the Askja and Brdarbunga volcanic systems of Iceland,
465 20092015, *JVGR* **000**, doi:10.1016/j.jvolgeores.2018.08.010.
- 466 [Hudson *et al.* (2019)] Hudson, T. S., Smith J.D., Brisbourne A.M., and White
467 R.S. (????), Automated detection of basal icequakes and discrimination
468 from surface crevassing, *Annals Glaciol.*, **60**(79), 1–11.
- 469 [Kao, H. and Shan, S. (2004)] Kao, H. and Shan, S., (2004), The source-
470 scanning algorithm: mapping the distribution of seismic sources in time
471 and space, *Geophysc. J. Int.*, **157**, 589–594.
- 472 [Lomax, A.(2000)] Lomax, A., J. Virieux, P. Volant and C. Berge (2000), Prob-
473 abilistic earthquake location in 3D and layered models: Introduction of a
474 Metropolis-Gibbs method and comparison with linear locations, *Advances
475 in Seismic Event Location*, 101–134, doi:10.1785/0120040141.

476 [Lomax, A.(2005)] Lomax, A. (2005), A reanalysis of the hypocentral location
477 and related observations for the great 1906 California earthquake, *Bul-*
478 *letin of the Seismological Society of America*, 95, 861–877, doi:10.1785/
479 0120040141.

480 [Murray et-al (2007)] Murray, T., Smith A.M., King M.A., and Weedon G.P.
481 (2007), Ice flow modulated by tides at up to annual periods at Rutford Ice
482 Stream, West Antarctica, *Geophys. Res. Lett.*, **34**(18), 6–11.

483 [Nederlandse Aardolie Maatschappij(2016)] Nederlandse Aardolie Maatschap-
484 pij (2016), Technical Addendum to the Winningsplan Groningen 2016,
485 *NAM*.

486 [Tarantola(2005)] Tarantola, A. (2005), *Inverse Problem Theory and Methods*
487 *for Model Parameter Estimation*, Society for Industrial and Applied Math-
488 ematics.

489 [Rawlinson, N. and Sambridge, M. (2005)] Rawlinson, N. and Sambridge, M.,
490 (2005), The fast marching method: and effective tool for tomographic imag-
491 ing and tracking multiple phases in complex layered media, *Exploration*
492 *Geophysics*, **36**, 341–350.

493 [Smith et al. (2015)] Smith, E.C., Smith, A.M., White, R.S., Brisbourne, A.M.,
494 and H.D. Pritchard (2015), Mapping the ice-bed interface characteristics
495 of Rutford Ice Stream, West Antarctica, using microseismicity, *J. geophys.*
496 *Res.*, **120**, 1881–1894.

- 497 [Grigoli *et al.* (2016)] Grigoli, F., Cesca, S., Krieger, L., Kriegerowski, M.,
498 Gammaldi, S., Horalek, J., Priolo, E., Dahm, T. (2016), Automated micro-
499 seismic event location using Master-Event Waveform Stacking, *Scientific*
500 *Reports.*, **6**, 10.1038/srep25744.
- 501 [Poiata *et al.* (2016)] Poiata, N., Satriano, C., Villette, JP., Bernard, P., and
502 Obara, K. (2016), Multiband array detection and location of seismic sources
503 recorded by dense seismic networks, *Geophysical Journal International*,
504 **205**, 10.1093/gji/ggw071.
- 505 [Grigoli *et al.* (2018)] Grigoli, F., Cesca, S., Vassallo, M., and Dahm, T. (2013),
506 Automated Seismic Event Location by Travel-Time Stacking: An Applica-
507 tion to Mining Induced Seismicity, **84**, 10.1126/sciadv.1700578.
- 508 [Perol *et al.* (2018)] Perol, T., Gharbi, M., and Denolle, M. (2018), Convolu-
509 tional neural network for earthquake detection and location, *Science Ad-*
510 *vances*, **4**.
- 511 [Hibert *et al.* (2014)] Hibert, C., Mangeney, G., Grandjean, G., Baillard, C.,
512 Rivet, D., Shapiro, N.M., Satriano, C., Maggi, A., Boissier, P., Ferrazzini,
513 V., and Crawford, W. (2014), Automated identification, location, and vol-
514 ume estimation of rockfalls at Piton de la Fournaise volcano, *J. Geophys*
515 **119**, 10.1002/2013JF002970.
- 516 [Ross *et al.* (2018)] Ross, Z.E., Yue, Y., Meier, M.A., Hauksson, E., and
517 Heaton, T.H. (2014), PhaseLink: A Deep Learning Approach to Seismic
518 Phase Association, *CoRR* **119**, abs/1809.02880.

Table 1: Table of specific QuakeMigrate parameters for each of the separate studies.

Parameters	Induced Seismicity in Groningen	Antarctic Basal Icequakes	Volcanic dyke injection
Sampling Rate	50	1000	50
Frequency filter P-wave - [Low,High,Poles]	[4Hz,16Hz, 4]	[10Hz,125Hz, 4]	[4Hz,16Hz, 4]
Frequency filter S-wave - [Low,High,Poles]	[4Hz,16Hz, 4]	[10Hz,125Hz, 4]	[4Hz,16Hz, 4]
Onset Windows P-wave - [STA,LTA]	Detect = [0.2s,1s], Locate = [0.1s,0.5s]	[0.01s,0.25s]	[0.2s,1s]
Onset Windows S-wave - [STA,LTA]	Detect = [0.2s,1s], Locate = [0.1s,0.5s]	[0.05s,0.5s]	[0.2s,1s]
Timestep	30s	0.75s	120s
Deimation Factor - [X,Y,Z]	[10,10,1]	[1,1,1]	[5,5,1]
Detection Threshold	1.3	1.8	1.2
Minimum Repeat Window	5.0	1.75	1.2
Marginal Window	0.5	2.75	4

Table 2: Table of computational cost for each scenario.

Example	Computational Cost (Data Hours/CPU Hours)	Memory Requirements (GB)
Induced Seismicity in Groningen	25	4.8GB
Antarctic Bassal Icequakes	240	6GB
Volcanic Earthquakes in Iceland	—	—

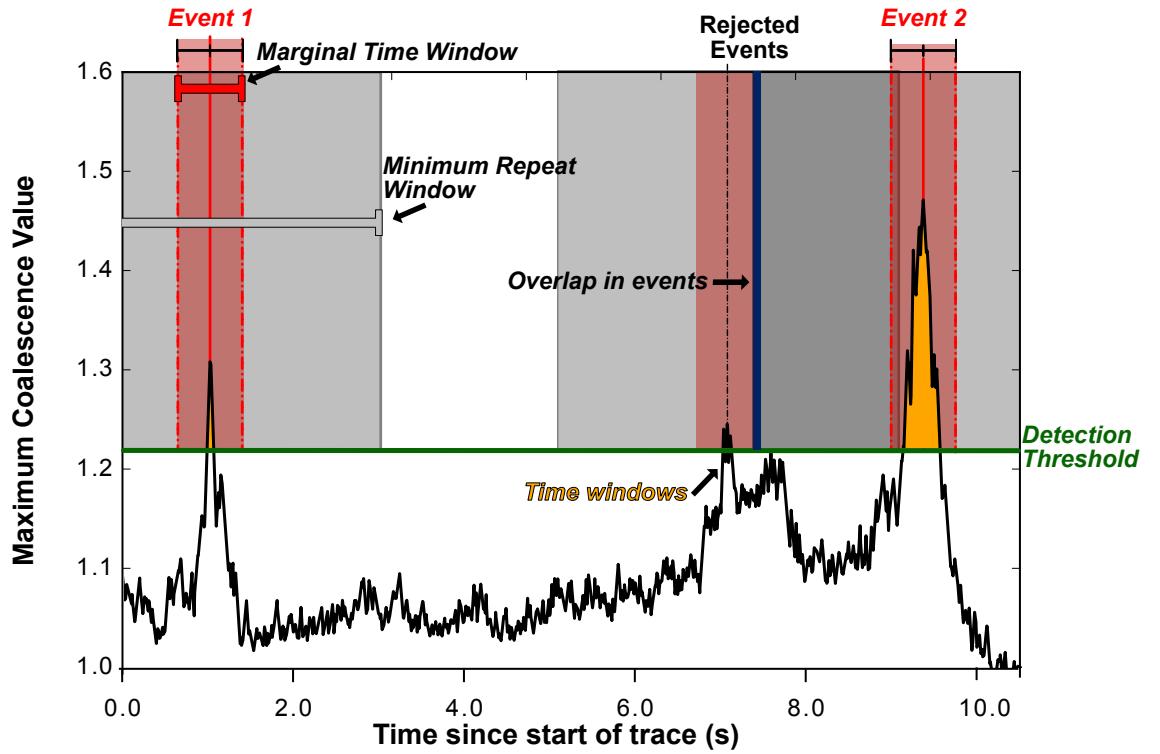


Figure 1: Triggering of events from the maximum coalescence value. Time windows are defined for periods exceeding the detection threshold. A minimum repeat window is applied to each time window to determine overlaps. Events time uncertainty is defined by the marginal time window about the maximum coalescence value.

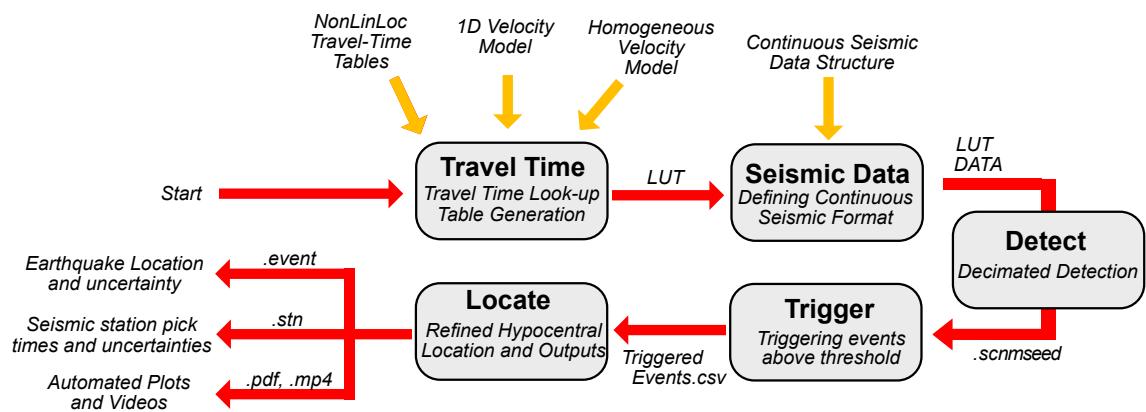


Figure 2: Flow diagram outlining the stages of QuakeMigrate.

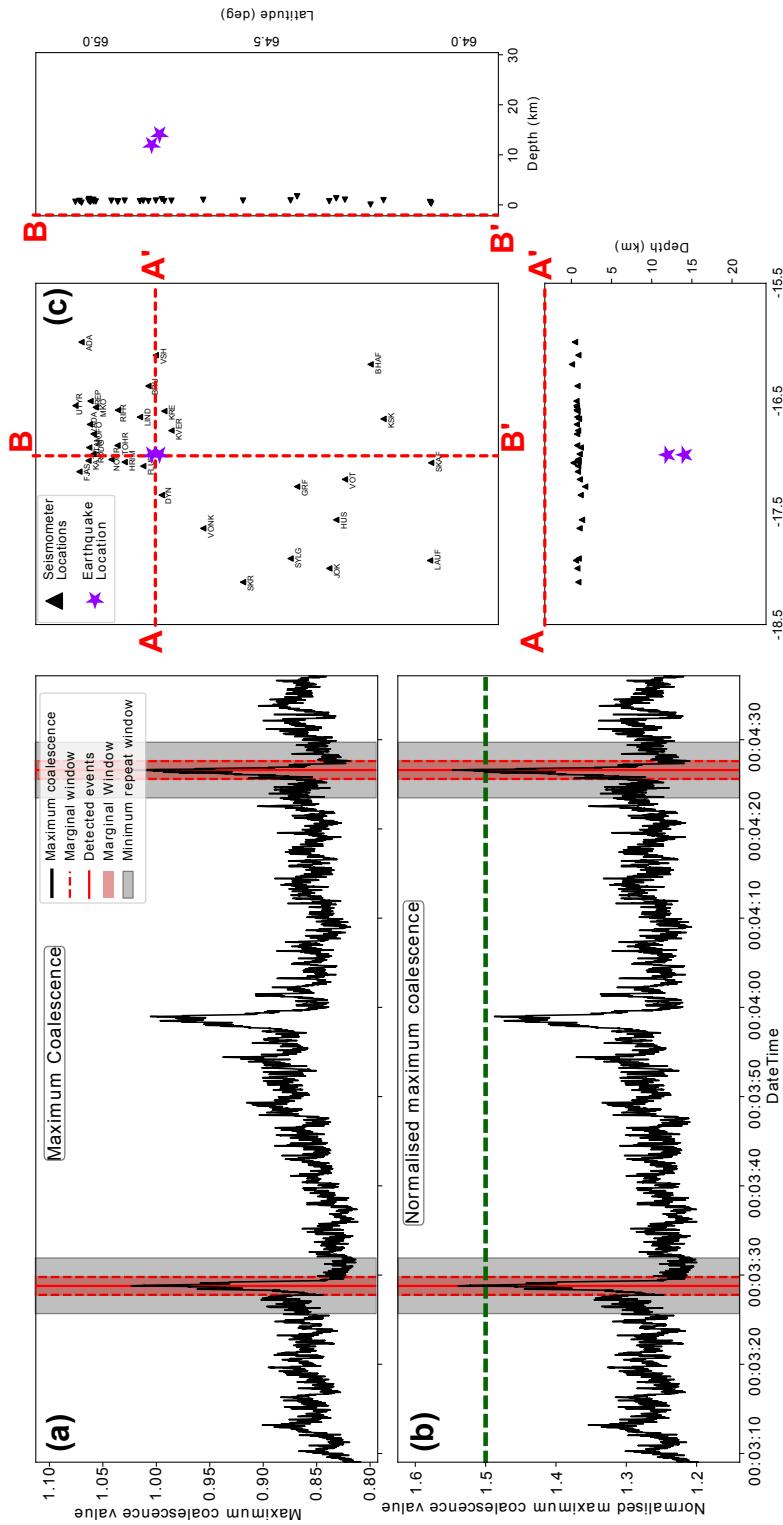


Figure 3: Triggering events output .pdf generated using a normalised detection threshold procedure. (a) Maximum coalescence variation through time. (b) Normalised maximum coalescence value through time. (c) Initial triggered earthquake locations and their relationship to station locations.

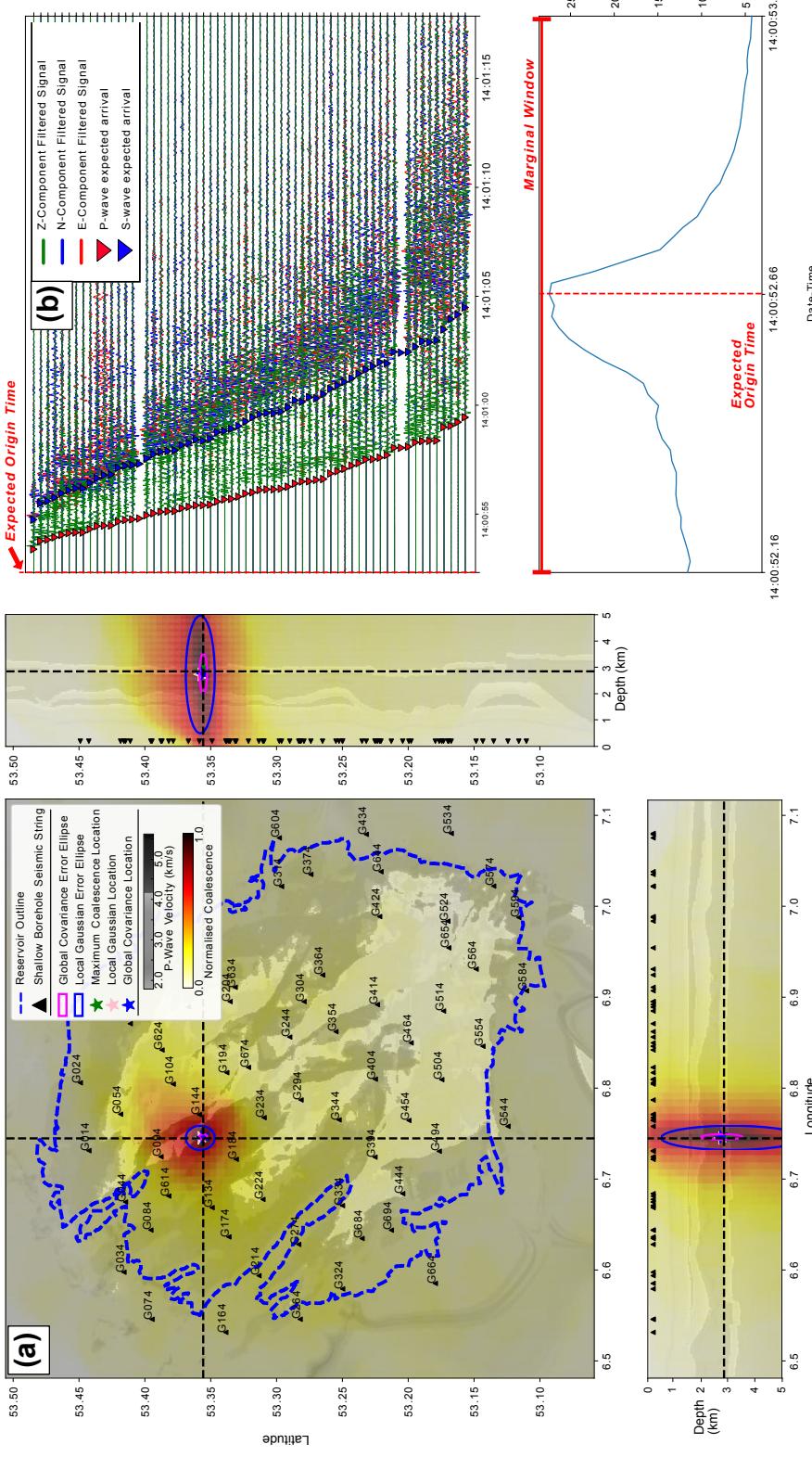


Figure 4: Normalised coalescence map of the Groningen 14:00:52 8th Jan 2018 earthquake overlaid on the 3D P-wave velocity model. (a) Represents a depth slice at the point of maximum coalescence, with triangles representing stations used in the coalescence. Blue and Pink error ellipses represent the Local and Global Gaussian error for the earthquake location. Blue dashed line represents the reservoir outline at depth (b) Represent the filtered seismic traces for each of the seismic stations. Red dashed line show the optimal origin time for the earthquake. Red triangles the P-wave pick and Blue triangles the S-wave pick. (c) Represents the coalescence value variation across the marginal window. Red dashed line represents the maximum coalescence value across this period, representing the expected origin time. Location uncertainty is derived as the integration of the spatial map over this marginal window.

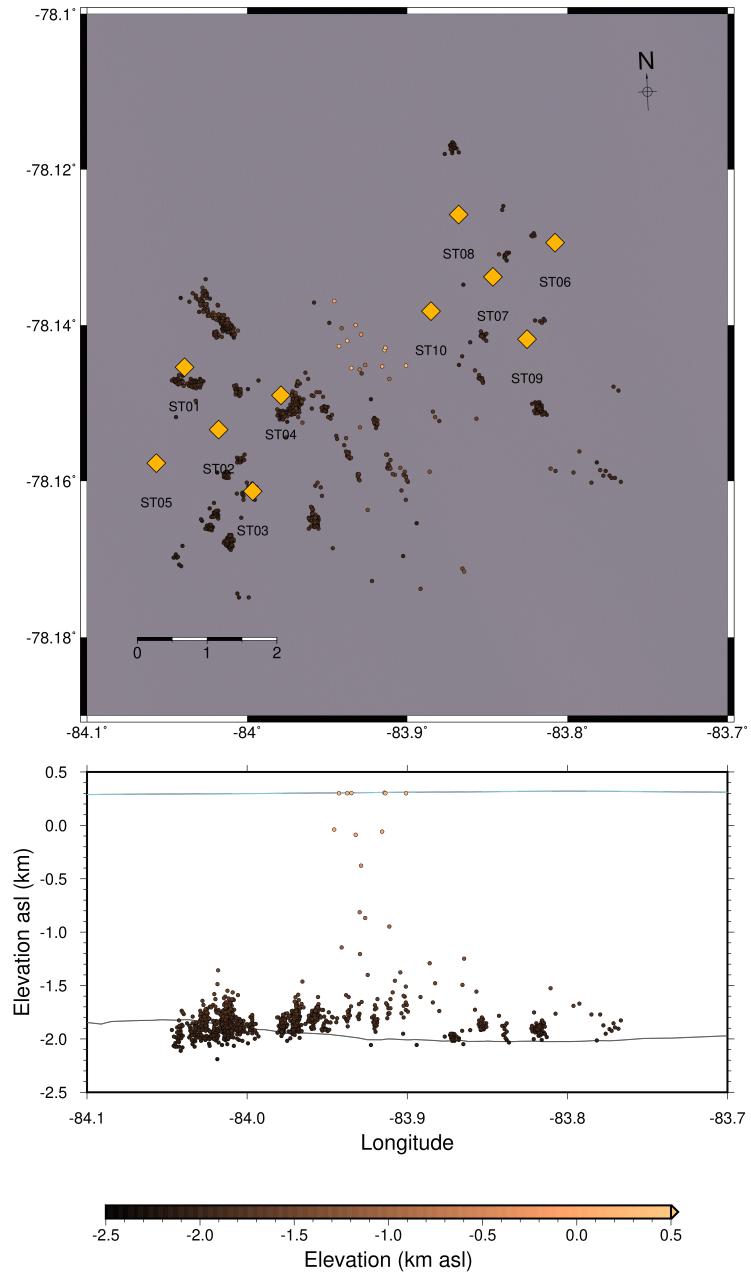


Figure 5: Plot of icequake hypocentres at the Rutford Ice Stream, Antarctica, for the 20th to 21st January 2009. All events detected are relocated using NonLinLoc ([*Lomax, A.*(2005)]). The ice surface and ice-bed interface are indicated by the blue and grey lines, respectively. This figure is originally from [Hudson *et al.* (2019)].