https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023

Natural Hazards and Earth System Sciences
Discussions

Brief communication: Small-scale geohazards cause significant and highly variable impacts on emotions

Evgenia Ilyinskaya [1], Vésteinn Snæbjarnarson [2,3], Hanne Krage Carlsen [4], Björn Oddsson [5]

[1]School of Earth and Environment, University of Leeds, Leeds, UK
[2]Miðeind ehf, Reykjavík, Iceland
[3]Department of Computer Science, University of Copenhagen
[4]Department of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
[5]Department of Civil Protection and Emergency Management, National Commissioner of the Icelandic Police, Reykjavik, Iceland

Correspondence to: Evgenia Ilyinskaya (e.ilyinskaya@leeds.ac.uk)

Abstract. The impact of geohazards on the mental health of the local populations, is well recognized but understudied. We used natural language processing (NLP) of Twitter posts to analyse the sentiments expressed in relation to a pre-eruptive seismic unrest, and a subsequent volcanic eruption in Iceland 2021. We show that despite the small size and negligible material damage, these geohazards were associated with a measurable change in expressed emotions in the local populations. The seismic unrest was associated with predominantly negative sentiments, but the eruption with predominantly positive. We demonstrate a cost-effective tool for gauging public discourse that could be used in risk management.

1 Introduction

Social media posts have been successfully used for assessing physical aspects of geohazards, for example, locating earthquakes (Earle, 2010; Steed et al., 2019). They have also been utilized for rapid assessment of material damage, and for aiding recovery efforts after several major geodisasters including hurricanes Harvey and Sandy, and the Great Tohoku earthquake and tsunami (e.g. (Earle, 2010; Chatfield and Brajawidagda, 2012; Guan and Chen, 2014)). The non-material impact of geohazards, including on the mental health of the local populations, is well recognized (Vo and Collier, 2013; Hlodversdottir et al., 2018; Becker et al., 2019; Gissurardóttir et al., 2019) but understudied. Studies are mostly done through interviews, and/or clinical assessments, where the results typically become available long after the event. Social media language and expressed sentiments can be used as indicators for public discourse, and have been shown to be predictive of individuals' mental health state and its deterioration (Eichstaedt et al., 2018; Oltmanns et al., 2021; Kelley and Gillan, 2022). Using artificial intelligence, such as natural language processing (NLP), it is possible to quickly process very large volumes of data for content and sentiment analysis (Park et al., 2015; Eichstaedt et al., 2018; Oltmanns et al., 2021). Here we use NLP on a dataset collected on the social

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023


Natural Hazards and Earth System Sciences Discussions

media platform Twitter to analyse the public views and expressed sentiments related to two types of globally common geohazards: a period of moderate seismic unrest and a small basaltic fissure eruption, using Iceland as the case study.

2 Methods

The Reykjanes peninsula in Iceland provided a highly suitable natural laboratory. Between 2019 and 2021 this densely populated area (~260,000 people within 40 km radius) experienced two distinct and prolonged periods of geohazards: 15 months of elevated seismicity (Sigmundsson et al., 2022) (from here on termed 'seismic unrest period') that abruptly subsided and was followed by a basaltic fissure eruption that lasted 6 months (Halldórsson et al., 2022) (termed 'eruption period'). The seismic unrest period took place between December 2019 and March 2021, including several intense earthquake swarms, with the largest event of magnitude 5.6 (Sigmundsson et al., 2022). The eruption took place between 19 March and 19 September 2021 at Mt Fagradalsfjall and effused relatively small lava flows within uninhabited valleys. The material damage caused by the seismicity and the eruption was negligible and no physical harm was reported. We were able to focus our study on local residents rather than tourists by analysing social media posts written in Icelandic as the language is spoken predominantly by people living in Iceland. In addition, due to covid-19 restrictions for most of our study period, the number of people traveling internationally was also at a record low in modern times.

2.1 Natural language processing

Twitter is estimated to be used by 24% of Iceland's population (Gallup, 2021). We performed sentiment analysis using NLP on tweets posted between 9 December 2019 and 31 December 2021 (n = 10,341) containing a fixed set of earthquake- and eruption- related keywords in Icelandic. Appendix A contains further details about the methods, including the full list of keywords. A subset of 636 tweets was manually labelled as 'negative sentiment', 'positive sentiment', or 'neutral statement'. The rest of the dataset was labelled automatically into the same three categories by a language model that was fine-tuned for classification using the manually labelled data. The model, bilingual in English and Icelandic, was adapted for sentiment analysis using the English Stanford Sentiment Treebank (SST) dataset (Socher et al., 2013), as no explicit Icelandic sentiment analysis dataset exists, and this was the first time NLP sentiment analysis in Icelandic was attempted.

Initial results showed that the model found that earthquakes were negative and eruptions were positive in sentences that should have been labelled as neutral. To mitigate this, we masked out all of the earthquake- and eruption- related keywords, both during model training and full dataset analysis. Using a subset of tweets we manually verified that the model was labelling sentiments correctly as neutral when the keywords were masked out. The drop in model performance when masking was introduced was minor: there was a drop in accuracy from 71% to 69%, and F1 (the harmonic mean of the precision and recall) also droped from 71 to 69. Full valuation results are in Appendix A. The model performance reached the 'benchmark' for good performance in Twitter sentiment analysis proposed by Zimbra et al., (2018). This demonstrates the potential of our method

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023


Natural Hazards and Earth System Sciences Discussions

for broader use as sentiment analysis can be successfully achieved by NLP models bilingual in English and a local language without a need for a sentiment dataset in the local language.

The main potential limitation of our approach is that views expressed on Twitter may not fully represent views of people who chose to use different social media platforms, or none at all; this can be explored in future research by including more than one social media platform. This would demonstrate the applicability of the method to other countries, where popularity of different social media platforms may differ.

3 Results and Discussion

The two geohazards in our case study evoked a measurable emotional response as indicated by the content of Twitter posts. The majority of the tweets containing earthquake- and/or eruption-related keywords (62%) were evaluated as containing a sentiment (30% negative and 32% positive). The remaining 38% were neutral statements. Previous work has shown that very large and/or destructive earthquakes, such as Great Tohoku 2011 (Vo and Collier, 2013), Canterbury 2010 (Becker et al., 2019), and Ridgecrest 2019 (Ruan et al., 2022) cause distress in the affected populations (measured using social media analysis by Vo and Collier, (2013) and Ruan et al., (2022); and 'traditional' interview methods by Becker et al., (2019)). We show here that earthquakes which are orders of magnitude smaller and cause no physical harm and negligible material damage also evoke a significant emotional response. We also show that the level of public interest is dependent on both the intensity of the seismicity and its proximity to densely populated areas. The number of tweets containing earthquake-related keywords (from here on termed 'earthquake-tweets') correlated strongly ($r^2 = 0.66$, Figures 1a and 2a) with the number of earthquakes on the densely populated Reykjanes peninsula. The correlation disappeared when we considered the seismicity in the rest of Iceland, which is relatively sparsely populated (Figure 2b). The seismic unrest period was dominated by negative-sentiments with an average weekly positive/negative ratio 1 : 1.3. This finding agrees well with contemporary reports in the local media which described primarily negative experiences including anxiety (RÚV, 2021b) and disturbed sleep (RÚV, 2021a) caused by the earthquakes. The overall negative sentiments were likely caused by a combination of the physical discomfort of the ground shaking, and by the uncertainty about further development. It was already known that the seismicity was being caused by a magma intrusion but there was an uncertainty about whether, when, and especially where an eruption would happen; the areas considered to be at threat from lava flows included a town, and major infrastructure (power plants and tourism businesses). The seismic unrest was monitored closely and scientific interpretations were published continuously in the media. The constant reports will have been impossible to ignore and conflicting interpretations may have contributed to the anxiety and other negative emotions.

There was a statistically significant change ($p < 5e\text{-}05$) to predominantly positive sentiments associated with the start of the eruption (Figure 1b and 1c). The average weekly positive : negative ratio during the eruption was 1.4 : 1 compared to 1 : 1.3 prior. It is possible that the increase in positivity was even larger because the NLP evaluation consistently reported a smaller positive : negative ratio during the eruption period compared to the manual analysis (Figure 1c). To the best of our knowledge,

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023


Natural Hazards and Earth System Sciences Discussions

this is the first time that an increase in positive attitudes associated with a start of an eruption is recorded among the local populations. We propose that the positive attitude is best explained by a combination of geophysical and societal factors. Many of the positive sentiments were expressing relief that an eruption would bring an end to the near-constant earthquakes, and to the associated uncertainty (examples of tweets: "*I want it to erupt so that the earthquakes stop*"; "*There is something absurd about being able to sleep soundly through the night now that there is an eruption in one's backyard*"). This agrees with previous studies on highly destructive events where positive statements were found associated with relief that the event is over (e.g. Great Tohoku 2011 earthquake; Vo and Collier, 2013), or messages of hope or pray for rescue and recovery (e.g. Hurricane Harvey; Zou et al., 2019). We also found that the Fagradalsfjall eruption directly evoked positive emotions in the local populations, including joy and pleasure. Our findings suggest that the close proximity of the Fagradalsfjall eruption site to populated areas may have been an important factor that enhanced the positive attitudes, as it allowed more people to experience it first-hand (examples of tweets "*The eruption is so fantastic, even on my second visit <…> I adore seeing all kinds of people there enjoying the nature and being outdoors*"). The site was open to the general public and was within one hour's drive from the capital city followed by one hour hike. The public started arriving in large numbers within hours of the eruption starting and the total number of visits via the hiking trails equated 310,000 (Icelandic tourism dashboard, 2021). It is possible that the social and physical isolation caused by the covid-19 pandemic further enhanced the positive experience, as a visit to the eruption site provided both a distraction and an opportunity to socialize and exercise in a relatively covid-safe outdoor setting. Furthermore, the eruption had an unprecedented live coverage through multiple high-quality webcams, and social media feeds (Wadsworth et al., 2022), allowing participation and enjoyment of people who could not travel in person, and thereby potentially kicking off a new chapter of 'remote' volcano tourism. "*It is midday and I have done nothing except watching the eruption* [via live streams]. *So beautiful!*"). Previous studies focusing on tourists in volcanic areas e.g. (Benediktsson et al., 2011; Davis et al., 2013; Donovan, 2018) have reported overwhelmingly positive attitudes, but given their very limited participant number and type, it was unknown how the findings related to the general public in local communities. Our results show that eruptions which do not directly endanger lives and infrastructure may cause a measurable increase in positive attitudes on population-wide level, and furthermore suggest that given the opportunity, people across different ages and physical abilities are keen to experience volcanic activity up close.

4 Conclusions

Our findings are important for risk assessment and management because we show that even small-sized geohazards without significant material damage can cause a measurable change in expressed sentiments in the local populations, which in turn, may indicate an impact on people's mental health. Incorporating such analysis into local risk management has the potential for immediate and longer-term benefits. For example, knowing that the public views an eruption (or another natural hazard) as a generally enjoyable event may allow the risk managers to adopt a suitable approach to achieve compliance should they need to restrict the site access. A longer-term benefit may come from a better understanding of the potential mental health

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023


Natural Hazards and Earth System Sciences Discussions

125 burden on the local populations. While our method does not provide direct measures of the mental health state, and is not intended to replace more formal investigations, it may be used to quickly gauge whether communities are under stress and may require additional surveying and/or resources. We show that pre-eruptive events (here, the seismic unrest) can potentially be more detrimental to mental well-being than the actual eruption and should be considered in studies of impacts. Finally, in our quest to reduce the risk posed by geohazards – in this case, a fissure eruption - we should not dismiss the potential mental
130 health benefits from allowing people to experience them where possible, even though the benefits will be difficult to quantify and weigh up against the (often more obvious) risks.

Data and code availability

Historical tweets can be obtained from Twitter API through an academic research access. The code used here for downloading the tweets is a python package TwitterAPI (source code available at https://github.com/geduldig/TwitterAPI). The datasets
135 and code generated in this study are available in an open-access repository https://github.com/vesteinn/fagradalsfjall_eruption_sentiment.

Author contributions

EI and HKC conceived the study idea. VS obtained and processed the Twitter data and led the NLP analysis with input from
140 EI. EI led the manual Twitter data analysis. EI led the overall manuscript writing with input from all authors.

Acknowledgements

Twitter data was made available by academic API access of Twitter. EI acknowledges funding from NERC COMET.

References

Becker, J. S., Potter, S. H., McBride, S. K., Wein, A., Doyle, E. E. H., and Paton, D.: When the earth doesn't stop shaking:
145 How experiences over time influenced information needs, communication, and interpretation of aftershock information during the Canterbury Earthquake Sequence, New Zealand, Int. J. Disaster Risk Reduct., 34, 397–411, https://doi.org/10.1016/j.ijdrr.2018.12.009, 2019.

Benediktsson, K., Lund, K. A., and Huijbens, E.: Inspired by Eruptions? Eyjafjallajökull and Icelandic Tourism, Mobilities, 6, 77–84, https://doi.org/10.1080/17450101.2011.532654, 2011.

150 Chatfield, A. and Brajawidagda, U.: Twitter tsunami early warning network: a social network analysis of Twitter information flows, Fac. Eng. Inf. Sci. - Pap. Part A, 2012.

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023


Natural Hazards and Earth System Sciences Discussions

Davis, S., A., A., Aoki, K., Cook, M., Duley, S., Huff, M., and Logan, C.: Managing risk and allure at volcanoes: In Hawaii: How close is too close?, Geoj. Tour. Geosites, 12, 85–93, 2013.

Donovan, A.: Sublime encounters: Commodifying the experience of the geos, Geo Geogr. Environ., 5, e00067, https://doi.org/10.1002/geo2.67, 2018.

Earle, P.: Earthquake Twitter, Nat. Geosci., 3, 221–222, https://doi.org/10.1038/ngeo832, 2010.

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D. A., and Schwartz, H. A.: Facebook language predicts depression in medical records, Proc. Natl. Acad. Sci., 115, 11203–11208, https://doi.org/10.1073/pnas.1802331115, 2018.

Gallup: Samfélagsmiðlamæling Gallup, Iceland, 2021.

Gissurardóttir, Ó. S., Hlodversdóttir, H., Thordardóttir, E. B., Pétursdóttir, G., and Hauksdóttir, A.: Mental health effects following the eruption in Eyjafjallajökull volcano in Iceland: A population-based study, Scand. J. Public Health, 47, 251–259, https://doi.org/10.1177/1403494817751327, 2019.

Guan, X. and Chen, C.: Using social media data to understand and assess disasters, Nat. Hazards, 74, 837–850, https://doi.org/10.1007/s11069-014-1217-1, 2014.

Halldórsson, S. A., Marshall, E. W., Caracciolo, A., Matthews, S., Bali, E., Rasmussen, M. B., Ranta, E., Robin, J. G., Guðfinnsson, G. H., Sigmarsson, O., Maclennan, J., Jackson, M. G., Whitehouse, M. J., Jeon, H., van der Meer, Q. H. A., Mibei, G. K., Kalliokoski, M. H., Repczynska, M. M., Rúnarsdóttir, R. H., Sigurðsson, G., Pfeffer, M. A., Scott, S. W., Kjartansdóttir, R., Kleine, B. I., Oppenheimer, C., Aiuppa, A., Ilyinskaya, E., Bitetto, M., Giudice, G., and Stefánsson, A.: Rapid shifting of a deep magmatic source at Fagradalsfjall volcano, Iceland, Nature, 609, 529–534, https://doi.org/10.1038/s41586-022-04981-x, 2022.

Hlodversdottir, H., Thorsteinsdottir, H., Thordardottir, E. B., Njardvik, U., Petursdottir, G., and Hauksdottir, A.: Long-term health of children following the Eyjafjallajökull volcanic eruption: a prospective cohort study, Eur. J. Psychotraumatology, 9, 1442601, https://doi.org/10.1080/20008198.2018.1442601, 2018.

Volcanic eruption in Geldingadalir: https://www.maelabordferdathjonustunnar.is/en/volcanic-eruption-in-geldingadalir, last access: 30 September 2021.

Kelley, S. W. and Gillan, C. M.: Using language in social media posts to study the network dynamics of depression longitudinally, Nat. Commun., 13, 870, https://doi.org/10.1038/s41467-022-28513-3, 2022.

Oltmanns, J. R., Schwartz, H. A., Ruggero, C., Son, Y., Miao, J., Waszczuk, M., Clouston, S. A. P., Bromet, E. J., Luft, B. J., and Kotov, R.: Artificial intelligence language predictors of two-year trauma-related outcomes, J. Psychiatr. Res., 143, 239–245, https://doi.org/10.1016/j.jpsychires.2021.09.015, 2021.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., and Seligman, M. E. P.: Automatic personality assessment through social media language, J. Pers. Soc. Psychol., 108, 934–952, https://doi.org/10.1037/pspp0000020, 2015.

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023

185 Ruan, T., Kong, Q., McBride, S. K., Sethjiwala, A., and Lv, Q.: Cross-platform analysis of public responses to the 2019 Ridgecrest earthquake sequence on Twitter and Reddit, Sci. Rep., 12, 1634, https://doi.org/10.1038/s41598-022-05359-9, 2022.

Grindvíkingar geti sett sig í spor fólks með kæfisvefn: https://www.ruv.is/frett/2021/03/16/grindvikingar-geti-sett-sig-i-spor-folks-med-kaefisvefn, last access: 30 September 2021.

190 Skjálftakvíði vegna aðstæðna sem maður ræður ekki við: https://www.ruv.is/frett/2021/03/01/skjalftakvidi-vegna-adstaedna-sem-madur-raedur-ekki-vid, last access: 30 September 2021.

Sigmundsson, F., Parks, M., Hooper, A., Geirsson, H., Vogfjörd, K. S., Drouin, V., Ófeigsson, B. G., Hreinsdóttir, S., Hjaltadóttir, S., Jónsdóttir, K., Einarsson, P., Barsotti, S., Horálek, J., and Ágústsdóttir, T.: Deformation and seismicity decline before the 2021 Fagradalsfjall eruption, Nature, 609, 523–528, https://doi.org/10.1038/s41586-022-05083-4, 2022.

195 Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., and Potts, C.: Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, Seattle, Washington, USA, 1631–1642, 2013.

Steed, R. J., Fuenzalida, A., Bossu, R., Bondár, I., Heinloo, A., Dupont, A., Saul, J., and Strollo, A.: Crowdsourcing triggers rapid, reliable earthquake locations, Sci. Adv., 5, eaau9824, https://doi.org/10.1126/sciadv.aau9824, 2019.

200 Vo, B.-K. H. and Collier, N.: Twitter Emotion Analysis in Earthquake Situations, IJCLA, 4, 15, 2013.

Wadsworth, F. B., Llewellin, E. W., Farquharson, J. I., Gillies, J. K., Loisel, A., Frey, L., Ilyinskaya, E., Thordarson, T., Tramontano, S., Lev, E., Pankhurst, M. J., Rull, A. G., Asensio-Ramos, M., Pérez, N. M., Hernández, P. A., Calvo, D., Solana, M. C., Kueppers, U., and Santabárbara, A. P.: Crowd-sourcing observations of volcanic eruptions during the 2021 Fagradalsfjall and Cumbre Vieja events, Nat. Commun., 13, 2611, https://doi.org/10.1038/s41467-022-30333-4, 2022.

205 Zimbra, D., Abbasi, A., Zeng, D., and Chen, H.: The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation, ACM Trans. Manag. Inf. Syst., 9, 1–29, https://doi.org/10.1145/3185045, 2018.

Zou, L., Lam, N. S. N., Shams, S., Cai, H., Meyer, M. A., Yang, S., Lee, K., Park, S.-J., and Reams, M. A.: Social and geographical disparities in Twitter use during Hurricane Harvey, Int. J. Digit. Earth, 12, 1300–1318, https://doi.org/10.1080/17538947.2018.1545878, 2019.

210

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023

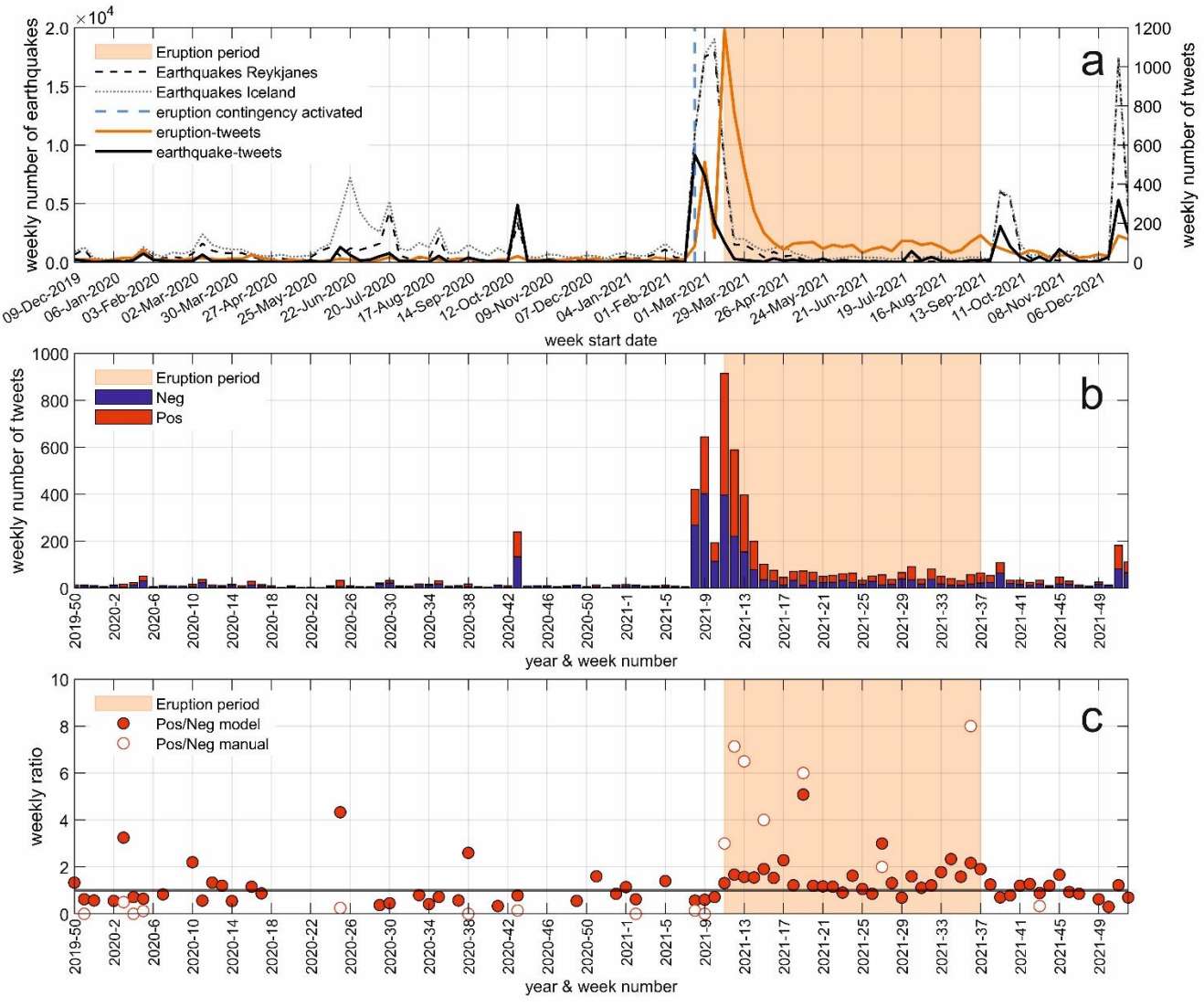
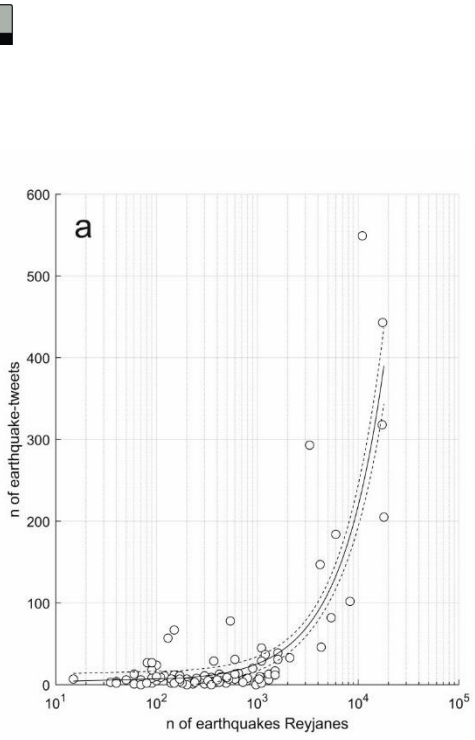


Figure 1: Timeseries plot of Twitter data during the pre-eruptive seismic unrest period (December 2019 – March 2021) and the eruption period (March – September 2021, highlighted with yellow). a) The weekly number of earthquake- and eruption- related tweets together with the weekly number of earthquakes. Number of weekly earthquakes is shown separately for Reykjanes peninsula and Iceland as a whole. b) The weekly number of tweets evaluated as expressing positive ('pos') or negative ('neg') sentiments by the NLP model. c) The average positive/negative ('pos/neg') weekly ratio, as evaluated by the NLP model in the whole dataset, and manually in the data subset. The data shown in c) include only weeks where the total number of tweets was > 10 to avoid bias introduced by very low numbers.

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023



Natural Hazards and Earth System Sciences
Discussions

Figure 2: Scatter plot showing the number of earthquake-related tweets as a function of the number of earthquakes on a) the Reykjanes peninsula, $r^2 = 0.66$ and b) the rest of Iceland (excluding Reykjanes peninsula), no statistically significant relationship. The solid line is a linear regression model fit and the dotted lines are the 5% significance level

225

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023


Natural Hazards and Earth System Sciences
Discussions

Appendix A

Theory and previous work on tweet-sentiment analysis

Neural language models using the Transformer architecture (Vaswani et al., 2017) have in recent years had a great impact in the field of natural language processing (NLP). These models are trained using representation learning to embed words (create vector representations for the words and their parts) for use in tasks such as machine translation, question answering and sentiment analysis. Such models are first trained in an unsupervised manner on raw text to maximize embedding quality across all available contexts. The models, e.g. BERT (Devlin et al., 2019) and T5 (Kale and Rastogi, 2021), can then be adapted using fine-tuning on much smaller labeled datasets for a variety of classification tasks that depend on the context of text in natural language.

Sentiment analysis is a common area of research in NLP with connections to hate speech analysis and stance detection(Maas et al., 2019; Socher et al., 2013; Mohammad et al., 2016) Here we use it in a traditional positive, negative and neutral classification setting but with regards to the geohazards surrounding the volcanic eruption in Fagradalsfjall in 2021.

Extended methods

Having gained academic access to historical tweets, we use the python package TwitterAPI (source code available at https://github.com/geduldig/TwitterAPI) to download tweets posted between 9 December 2019 and 31 December 2021. The tweets are filtered for a fixed set of earthquake- and eruption-related keywords in Icelandic (details on keywords below). Duplicates, including retweets, were removed from the data. In total 10,341 individual tweets matched our criteria and were used for further analysis.

For keywords, tweets were filtered with the morphological variations of eruption and earthquake in Icelandic, along with the emoji for volcano. The full list used is:

Eruption-related: *' 🌋 ', 'gos', 'gosið', 'gosinu', 'gossins', 'eldgos', 'eldgosið', 'eldgosi', 'eldgoss', 'gjósa', 'gaus'.*

Earthquake-related: *'skjálfti', 'skjálfta', 'skjálftar', 'skjálftum', 'skjálftarnir', 'skjálftana', 'skjálftanum', 'skjálftann', 'skjálftan', 'skjálftans', 'skjálftinn'.*

A subset of this dataset was manually annotated by sentiment into negative, positive or neutral. 224 tweets were labeled as negative, 241 as positive and 171 as neutral. Tweets were selected for manual annotation semi-randomly; outside of the eruption period we randomly selected a subset of tweets in periods with either highly elevated seismicity, or relatively low seismicity levels. During the 6-months long eruption period we randomly selected a subset of tweets from each of the 3 first weeks of the eruption when eruption-related tweet frequency was at the highest level. We then randomly selected tweets in eruption weeks 5, 9 17 and 26 to cover different time periods of the eruption.

To label the other tweets automatically an already adapted language model was fine-tuned on the manually annotated data for use in labeling as follows: The data was split randomly into a train and validation set of 493 and 121 tweets respectively. An pre-trained Icelandic-English language model (Snæbjarnarson and Einarsson, 2022) was first adapted for binary sentiment

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023


Natural Hazards and Earth System Sciences Discussions

analysis using the English Stanford Sentiment Treebank (SST) dataset (Socher et al., 2013). The original bilingual model was trained using Fairseq (Ott et al., 2019), from Facebook AI research, but then ported to the Transformers (Wolf et al., 2020) library and adapted for sentiment analysis. While no explicit Icelandic sentiment analysis dataset exists the model seems to behave well in general sentiment analysis for Icelandic due to its bilingual pre-training, this behavior has been documented well before and is referred to as transfer-learning (Ruder et al., 2019). The model was then further fine-tuned on the training data for five full iterations (epochs) using a learning rate of 2e-5 and a batch size of 8. As the initial SST fine-tuned model was only trained to classify positive or negative sentiment the output layer of the neural network was modified to provide three classes with the new classification node initialized by averaging over the prior two. Initial results showed that the model found that earthquakes were negative and were eruptions positive in sentences that should have been labeled as neutral. To mitigate this, commonly seen short-cut to labeling (Geirhos et al., 2020), we masked out all of the earthquake- and eruption- related keywords, both during model training and actual evaluation. Using a subset of tweets we manually verified that the model was labeling sentiments correctly as neutral when the keywords were masked out. The drop in performance when masking was minor: there was a drop in accuracy from 71% to 69%, and F1 (the harmonic mean of the precision and recall) also droped from 71 to 69. Evaluation results are shown in Table A1.

Keywords masked	Accuracy	F1	Recall	Precision
No	71.4%	71.3%	72.9%	71.6%
Yes	69.0%	68.8%	70.5%	68.9%

Table A1: Evaluation results for models trained on the collected tweets

The datasets and code are available in an open-access repository https://github.com/vesteinn/fagradalsfjall_eruption_sentiment for reproducibility and further use by the community.

The key NLP contributions are; a new dataset for Icelandic sentiment analysis of tweets about geological events, a fine-tuned Icelandic sentiment model by means of transfer learning and manually curated data available as open-access, and an evaluation of masking to prevent shortcut learning in sentiment analysis.

Appendix A references

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. N.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Languages Technologies, Minneapolis, Minnesota, 2019.

Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., and Wichmann, F. A.: Shortcut learning in deep neural networks, Nat. Mach. Intell., 2, 665–673, https://doi.org/10.1038/s42256-020-00257-z, 2020.

https://doi.org/10.5194/nhess-2023-6
Preprint. Discussion started: 17 April 2023

Kale, M. and Rastogi, A.: Text-to-Text Pre-Training for Data-to-Text Tasks, https://doi.org/10.48550/arXiv.2005.10433, 8 July 2021.

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and Potts, C.: Learning Word Vectors for Sentiment Analysis, ACL, 2019.

Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., and Cherry, C.: SemEval-2016 Task 6: Detecting Stance in Tweets, in: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), SemEval 2016, San Diego, California, 31–41, https://doi.org/10.18653/v1/S16-1003, 2016.

Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., Grangier, D., and Auli, M.: fairseq: A Fast, Extensible Toolkit for Sequence Modeling, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations), Minneapolis, Minnesota, 48–53, https://doi.org/10.18653/v1/N19-4009, 2019.

Ruder, S., Peters, M. E., Swayamdipta, S., and Wolf, T.: Transfer Learning in Natural Language Processing, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials, Minneapolis, Minnesota, 15–18, https://doi.org/10.18653/v1/N19-5004, 2019.

Snæbjarnarson, V. and Einarsson, H.: Cross-Lingual QA as a Stepping Stone for Monolingual Open QA in Icelandic, https://doi.org/10.48550/arXiv.2207.01918, 5 July 2022.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., and Potts, C.: Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, Seattle, Washington, USA, 1631–1642, 2013.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is All you Need, in: Advances in Neural Information Processing Systems, 2017.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., and Rush, A.: Transformers: State-of-the-Art Natural Language Processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Online, 38–45, https://doi.org/10.18653/v1/2020.emnlp-demos.6, 2020.