Geophysical data integration and machine learning for multi-target leakage estimation in geologic carbon sequestration

Rafael Pires de Lima, Los Alamos National Laboratory, The University of Oklahoma; and Youzuo Lin, Los Alamos National Laboratory*

SUMMARY

Geologic carbon sequestration (GCS) can be considered essential for alleviating atmospheric level of anthropogenic carbon dioxide (CO_2). GCS is the process of capturing, injecting, and long-term storing CO_2 into a geologic formation. However, the injection of CO_2 into deep geologic layers can create pressure perturbations that can potentially lead to leakages from the storage formation to nearby aquifers or to the surface. CO_2 accumulations can cause harm to animals and humans, and brine or CO_2 leakage into other formations can disturb subsurface operations (e.g. oil production, groundwater). Therefore, an effective monitoring method is an essential procedure that needs to be taken to ensure CO_2 or brine leakage are detected. We use seismic and pressure data from hundreds of flow models generated with different hydrogeological parameters combined with convolutional neural networks to develop a processing pipeline that can estimate the amount of CO_2 and brine leaked. Our model has the potential to be used as a quick first inference tool, indicating whether further analysis and interventions need to take place.

INTRODUCTION

In geologic carbon sequestration (GCS), carbon dioxide (CO_2) is injected into a geologic formation (Gale, 2004; Holloway, 2005; Breunig et al., 2013), preventing anthropogenic CO_2 from entering the atmosphere (Breunig et al., 2013). CO_2 injection will give rise to a necessary CO_2 imaging and monitoring to ensure that the gas is being injected at the correct location, that it fills the storage reservoir as estimated, that it does not flow toward high-risk areas, and that the CO_2 remains sealed in the intended formation over time, not escaping to the surface or other geologic formations (Lumley, 2010). Different geologic formations, such as saline aquifers or depleted oil and gas reservoirs, have the potential to serve as storage units to capture CO_2. Due to their large storage capacities, saline aquifers within sedimentary basins in the United States (US) contain some of the most attractive geologic formations for GCS. However, as such saline aquifers are usually saturated with brine, the injection of large quantities of CO_2 can lead to a widespread and lasting pressure perturbation in the subsurface (Nicot, 2008; Birkholzer et al., 2012; Breunig et al., 2013). The elevated formation pressure might incur in caprock fracturing and fault reactivation, as well as pressure-driven leakage of CO_2 and brine (Rutqvist et al., 2008; Breunig et al., 2013). Extensive research has been conducted to improve fault imaging (e.g. Bahorich and Farmer, 1995; Gersztenkorn and Marfurt, 1999; Hale, 2013; Lima and Marfurt, 2017; Wu, 2017; Qi et al., 2018) thus, such high risk areas might be avoided due to a prior fault location and characteristics understanding. Another high risk factor are legacy wells. As legacy wells create artificial connections between geologic layers, they can also be pathways to potential leakages. Regardless of the proximity with beforementioned geohazards, as the increased pressure generates subsurface instabilities, GCS sites should be monitored as an early leakage detection can help mitigate potential significant pollutant spreading. We investigate a combination of time-lapse geophysical surveys and the implementation of machine learning techniques. Our objective is to help prototype a system with the capacity to automatically estimate the amount of CO_2 and brine leaked at a GCS site.

Most of time-lapse analysis monitoring techniques are based on the response of geophysical properties to changes in local geology, such as temperature, pressure, and fluid properties. Time-lapse surveys are commonly used in different plays (e.g. Swanston et al., 2003; Shabelansky et al., 2015; Gherasim et al., 2016) and even for GCS monitoring (e.g. Chadwick et al., 2005; Lumley, 2010; Bergmann et al., 2014). Jung et al. (2013) demonstrated how brine and CO_2 leakage through abandoned wells can be detected using pressure and surface-deformation time-lapse data. Geoscientists have been successfully using machine learning methodologies for roughly twenty years now. Applications of machine learning with geophysical data are vast and range from many different geophysical methods (e.g. Calderón-Maciás et al., 1998; Zhang et al., 2002; Schnetzler and Alumbaugh, 2017; Araya-Polo et al., 2017; de Lima and Marfurt, 2018; Sinha et al., 2018; Zhao et al., 2018; Zhou et al., 2018; Wu et al., 2019a,b; Lubo-Robles and Marfurt, 2019; Infante-Paez and Marfurt, 2019). In our investigation, we make use of both of these techniques, time-lapse surveys and machine learning, to address CO_2 and brine leakage from a synthetic dataset created based on a GCS site. We evaluate convolutional neural networks (CNN) as a improvement that can be coupled with a time-lapse CGS monitoring sensors. Specifically, we use vertical component seismic and pressure data as input to CNN to estimate the amount of CO_2 and brine leaked. Our approach integrates geophysical data (seismic and pressure data), and estimates the leakage of two different fluids (CO_2 and brine), which is an improvement on previously published applications (Zhou et al., 2018)

DATA

Buscheck et al. (2017) generated thousands of groundwater flow models to better understand effects caused by $C0_2$ injection at GCS sites. The model is based on a hypothetical CO_2 storage reservoir in the Vedder Formation at the Kimberlina site in the southern San Joaquin Basin, California. Each one of the models varies according to the amount of CO_2 injected and the physical properties of the aquifer, such as porosity, solid density, and permeability. We use 500 different reservoir and groundwater flow scenarios (static model) extracted from

10.1190/segam2019-3215405.1

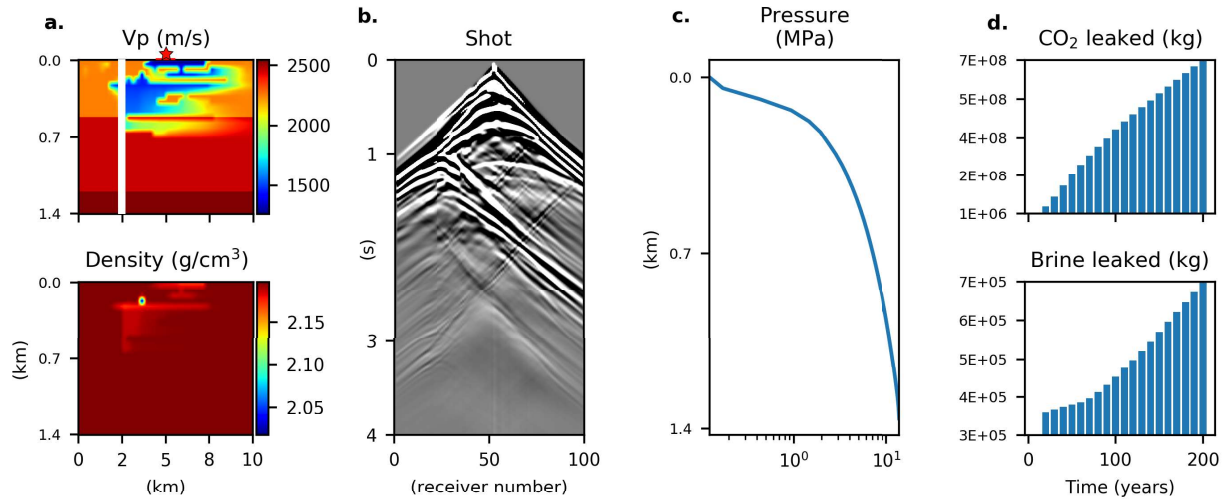

Figure 1: (a) One of the configurations of the geoogic model after 200 years of CO_2 injection and leakage. The white vertical line in the Vp image indicates the position of the legacy well containing the sensors measuring pressure. The injection well is not visible in this image (3 km to the left). (b) A a seismic shot taken at position 5 km (star in panel (a)). (c) The measured pressure in the legacy well. And (d) the CO_2 and brine leakage acumulated in the area.

Buscheck et al. (2017) dataset. The static model is composed of three geological layers, an injecting well, and a leaking well. Each one of the 500 static models were used to compute a flow simulation of 20 time steps, each time step representing a time difference of 10 years (simulation). Therefore, for each one of the simulations, we have a geological model indicating the differences in the geological properties as well as the amount of CO_2 leaked through the leaking well.

For each time step, we used a finite difference forward model to generate three shot gathers with 100 receivers. Each receiver contains 4000 samples, with a sampling rate of 1 ms. Figure 1 shows one simulation example containing the geological model, one of the three simulated shots, as well as the step-by-step CO_2 and brine leaked. In our simulations, a leakage is any amount of CO_2 or brine that enters the model, i.e., that somehow breaks the bottom-most layer at 1.4 km (the saline aquifer being injected with CO_2).

METHODOLOGY

The leakage estimation task can be modeled as a function $\mathbf{F}^\star$ that takes seismic and pressure data and returns the estimated amount of CO_2 and brine leaked. Traditionally, geoscientists use model based methods, i.e., using mathematical approximations to model the underlying physical process, to estimate the fluid or rock physics of the studied region. Such model based approach commonly relies on reservoir simulator coupled with a rock physics model. The model based approach is also heavily dependent on the expert domain experience, a combination of geophysics, geology, and petroleum engineering. In contrast, we adopt a data-driven approach, wherein we estimate the function $\mathbf{F}^\star$ from the statistics within the data. Though this statistical approach requires less domain knowledge, our results indicate that such approach can be efficient and help quickly estimate the amount of leakage. In that manner, such statistical based approach can be used to trigger further more intense and domain knowledge-based investigations, such as traditional fluid flow modeling.

Out of the 500 simulations we used, we selected 50 to be part of the test set. The remaining 450 are used for training. As our objective is to find a function $\mathbf{F}^\star$ that receives as input seismic and pressure data, and outputs the estimated CO_2 and brine leaked, we concatenate the results obtained using the Inception V3 (Szegedy et al., 2015) on the seismic path, and a simple CNN model on the pressure path. Figure 3 shows a visualization of the final deep learning model we applied. We compare the results we obtained using a single data input (seismic only) and a combination of the available geophysical data (seismic and pressure).

To reduce the seismic dimensions, originally 4000 samples by 100 receivers (4000 x 100) for each shot, we apply an anti-alias filter and decimate the data. After the decimation, each one of the shots has 400 samples and 100 receivers (400 x 100). That reduces the Nyquist frequency from 500 Hz to 50 Hz, however we observe that the important reflections and information are still well represented in the data. As order of mass leaked CO_2 and brine leakage mass in the simulations varies from 0 to 10^{10} (kg), we standardized the leaked mass using the natural logarithmic function:

$$\hat{Y} = ln(Y + 1) \tag{1}$$

where Y represents the original value of leakage mass (CO_2 or brine), and $\hat{Y}$ stands for standardized leakage mass. We add 1 to avoid taking the logarithm of 0. Therefore, our results are reported based on the natural logarithm of the leaked CO_2 and brine.

10.1190/segam2019-3215405.1

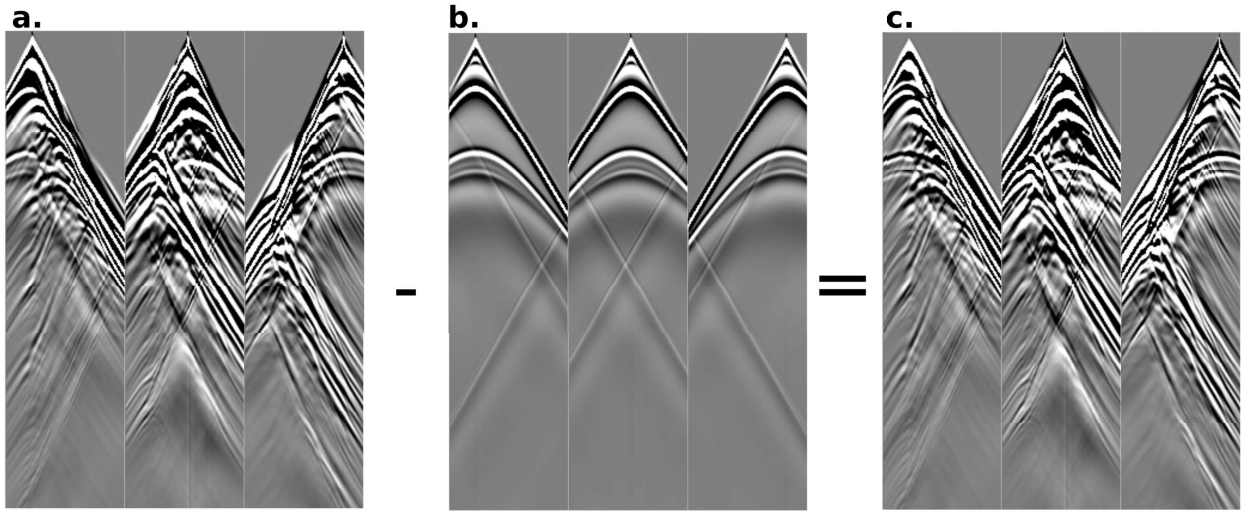

Figure 2: A simple ilustration of data preparation. We subtract (a) three shots taken after 200 years of CO_2 injection from (b)the same three baseline shots. The resulting difference (c) shows the geological changes. (c) Is the data we used in our experiments. We apply the same process (computing the difference) with the pressure data.

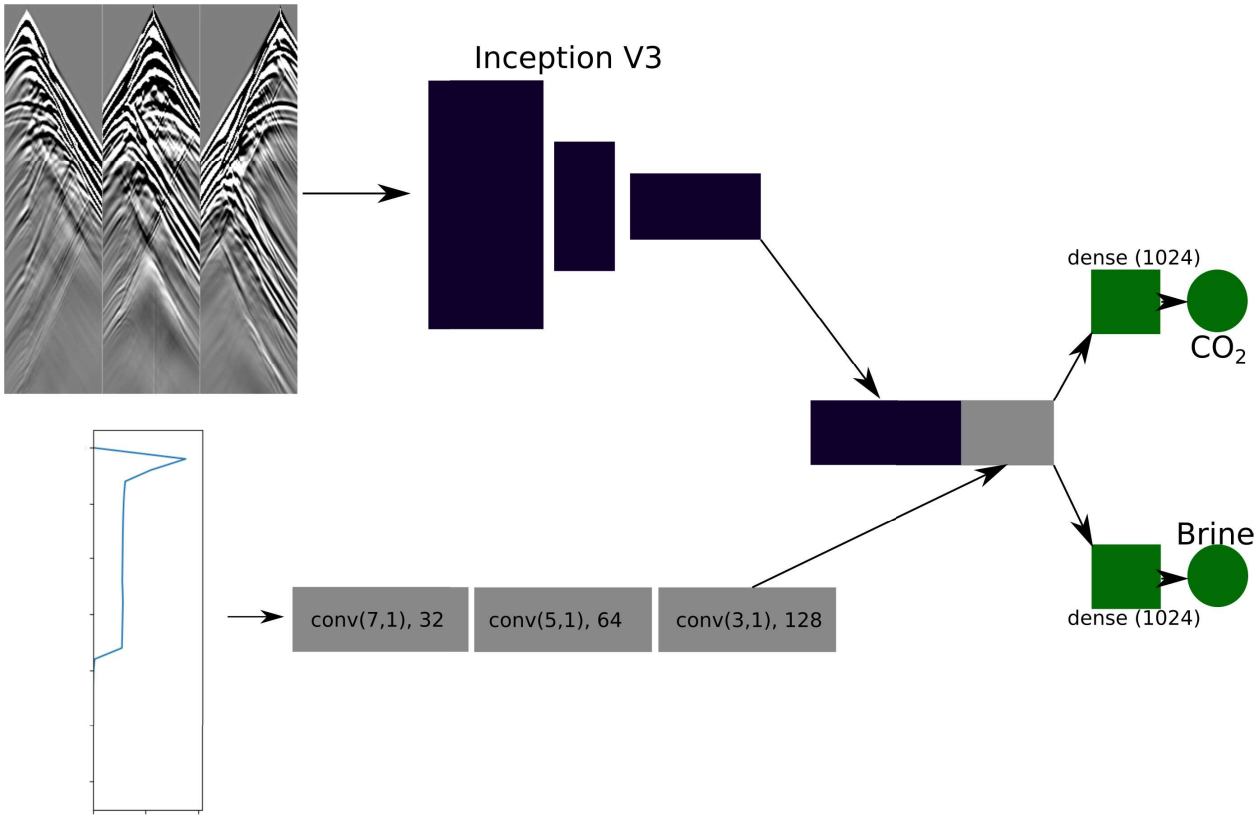

Figure 3: Visual representation of the deep neural network we used. The seismic data goes through Inception V3 CNN model. The pressure data goes through two convolutional layers (conv(3,1),128 denotes using 128 (3x1) convolution kernels with valid padding). The resulting seismic and pressure features are concatenated. We then split the model in two paths, both with one layer of 1024 densely connected neurons and a final regression neuron.

10.1190/segam2019-3215405.1

RESULTS AND DISCUSSION

To evaluate the performance of our approach, we present the resulting metrics obtained with the test set. Table 1 shows the summary of the errors in the test data when using as input seismic only, and seismic and pressure data combined. Figure 4 shows the results graphically. We observe that the addition of pressure data provided a small improvement in the prediction of small leaks for both CO_2 and brine predictions (blue circles/lines in Figure 4). Such small improvement is not reflected in the CO_2 metrics, however it is present in the brine metrics (Table 1). One of the reasons for better metrics of seismic only CO_2 predictions is that such results are more focused for higher amount of CO_2 leakage (Figure 4a., focused orange dots for leakage above $\sim$10).

		log(CO_2)	log(brine)
sesmic only	RMSE	0.87	1.66
	MAE	0.50	1.22
seismic and pressure	RMSE	0.91	1.53
	MAE	0.58	1.20

Table 1: Summary results of the two approaches: seismic data as input, and a combination of seismic and pressure data. We present the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) as our metrics. For both metrics, 0 means a perfect score. The values of ln(CO_2) range from 0 to 20 while ln(brine) values range from 6 to 16.

CONCLUSIONS AND FUTURE WORK

We demonstrated that a machine learning approach is a methodology that can provide highly accurate results for CO_2 and brine leakage in CGS sites. Our results indicate that a CGS monitoring system, with permanently installed sensors, perhaps also permanently installed sources, could be coupled with a machine learning model. Such monitoring system could have the capability to quickly estimate the amount of leakage and help geoscientists with early warnings. We observed that combining seismic and pressure data provides marginally better results for brine estimation when compared with a monitoring system based solemnly in seismic data. Further testing with increased number of examples of small leaks ($CO_2 < 10^6$ kg and brine $< 10^4$ kg) might help us understand the benefits of geophysical integrated monitoring. We believe the implementation we described needs to be tested against presence of noisy data. The following necessary step is to test the methodology we described using real seismic and pressure data. The results we encountered using machine learning techniques also should be compared with estimates obtained with traditional fluid flow reservoir modeling methodologies.

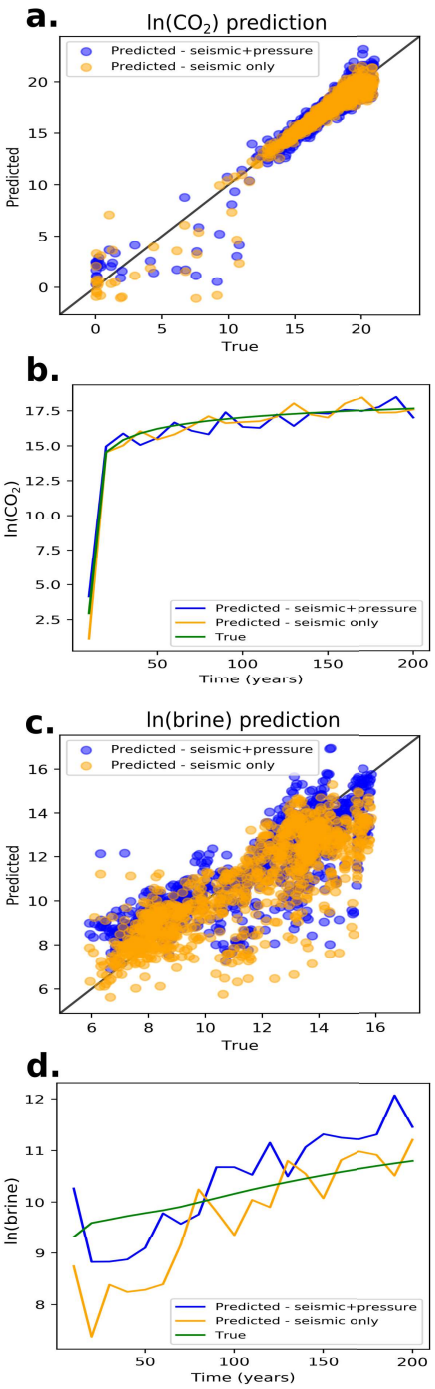

Figure 4: Summary and examples of the predicted and true values of leaked CO_2 and brine. (a) Shows values of predicted vs true CO_2 leaked and (b) shows CO_2 leaked for one selected simulation example. (c) Shows values of predicted vs true brine leaked and (d) shows brine leaked for one selected simulation example.

ACKNOWLEDGMENTS

This work was funded by the U.S. DOE Office of Fossil Energy Carbon Storage program. The experiment was performed using supercomputers of LANL's Institutional Computing Program. Rafael acknowledges CNPq (grant 203589/2014-9) for the financial support allowing the pursuit of his Ph.D. studies.

10.1190/segam2019-3215405.1

REFERENCES

Araya-Polo, M., T. Dahlke1, C. Frogner, C. Zhang, T. Poggio, and D. Hohl, 2017, Automated fault detection without seismic processing: The Leading Edge, **36**, 208–214, doi: https://doi.org/10.1190/tle36030208.1.

Bahorich, M., and S. Farmer, 1995, 3-D seismic discontinuity for faults and stratigraphic features: The coherence cube: The Leading Edge, **14**, 1053–1058, doi:https://doi.org/10.1190/1.1437077.

Bergmann, P., M. Ivandic, B. Norden, C. Rücker, D. Kiessling, S. Lüth, C. Schmidt-Hattenberger, and C. Juhlin, 2014, Combination of seismic reflection and constrained resistivity inversion with an application to 4D imaging of the CO_2 storage site, Ketzin, Germany: Geophysics, **79**, no. 2, B37–B50, doi: https://doi.org/10.1190/geo2013-0131.1.

Birkholzer, J. T., A. Cihan, and Q. Zhou, 2012, Impact-driven pressure management via targeted brine extraction — Conceptual studies of CO_2 storage in saline formations: International Journal of Greenhouse Gas Control, **7**, 168–180, doi: https://doi.org/10.1016/j.ijggc.2012.01.001.

Breunig, H. M., J. T. Birkholzer, A. Borgia, C. M. Oldenburg, P. N. Price, and T. E. McKone, 2013, Regional evaluation of brine management for geologic carbon sequestration: International Journal of Greenhouse Gas Control, **14**, 39–48, doi: https://doi.org/10.1016/j.ijggc.2013.01.003.

Buscheck, T. A., K. Mansoor, X. Yang, and S. A. Carroll, 2017, Simulated data for testing monitoring techniques to detect leakage in groundwater resources: Kimberlina model with wellbore leakage: Technical report, Lawrence Livermore National Laboratory.

Calderón-Maciás, C., M. K. Sen, and P. L. Stoffa, 1998, Automatic NMO correction and velocity estimation by a feedforward neural network: Geophysics, **63**, 1696–1707, doi: https://doi.org/10.1190/1.1444465.

Chadwick, R. A., R. Arts, and O. Eiken, 2005, 4D seismic quantification of a growing CO_2 plume at Sleipner, North Sea: Geological Society, London, Petroleum Geology Conference series, **6**, 1385–1399, doi:https://doi.org/10.1144/0061385.

Gale, J., 2004, Geological storage of CO_2: What do we know, where are the gaps and what more needs to be done? Energy, **29**, 1329–1338, doi: https://doi.org/10.1016/j.energy.2004.03.068.

Gersztenkorn, A., and K. J. Marfurt, 1999, Eigenstructure-based coherence computations as an aid to 3-D structural and stratigraphic mapping: Geophysics, **64**, 1468–1479, doi:https://doi.org/10.1190/1.1444651.

Gherasim, M., E. L'Heureux, and Z. Lorenzo, 2016, Evaluation and correction of sources of 4D noise using modeled ocean-bottom-node data: The Leading Edge, **35**, 880–886, doi: https://doi.org/10.1190/tle35100880.1.

Hale, D., 2013, Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3D seismic images: Geophysics, **78**, no. 2, O33–O43, doi: https://doi.org/10.1190/geo2012-0331.1.

Holloway, S., 2005, Underground sequestration of carbon dioxide-a viable greenhouse gas mitigation option: Energy, **30**, 2318–2333, doi: https://doi.org/10.1016/j.energy.2003.10.023.

Infante-Paez, L., and K. J. Marfurt, 2019, Using machine learning as an aid to seismic geomorphology, which attributes are the best input? Interpretation, 1–60, doi: https://doi.org/10.1190/int-2018-0096.1.

Jung, Y., Q. Zhou, and J. T. Birkholzer, 2013, Early detection of brine and CO_2 leakage through abandoned wells using pressure and surface-deformation monitoring data: Concept and demonstration: Advances in Water Resources, **62**, 555–569, doi: https://doi.org/10.1016/j.advwatres.2013.06.008.

De Lima, R. A. P., and K. Marfurt, 2017, Quantifying fault connectivity drilling hazards through simple flow computations: 87th Annual International Meeting, SEG, Expanded Abstracts, 2345–2349, doi: https://doi.org/10.1190/segam2017-17786445.1.

Lima, R. P. de, and K. J. Marfurt, 2018, Principal component analysis and K-means analysis of airborne gamma-ray spectrometry surveys: doi: https://doi.org/10.1190/segam2018-2996506.1.

Lubo-Robles, D., and K. J. Marfurt, 2019, Independent Component Analysis for reservoir geomorphology and unsupervised seismic facies classification in the Taranaki Basin, New Zealand: Interpretation, 1–76, doi:https://doi.org/10.1190/int-–https://doi.org/2018–0109.1.

Lumley, D., 2010, 4D seismic monitoring of CO_2 sequestration: The Leading Edge, **29**, 150–155, doi: https://doi.org/10.1190/1.3304817.

Qi, J., B. Lyu, A. AlAli, G. Machado, Y. Hu, and K. Marfurt, 2018, Image processing of seismic attributes for automatic fault extraction: Geophysics, **84**, no. 1, O25–O37, doi: https://doi.org/10.1190/geo2018-0369.1.

Rutqvist, J., J. T. Birkholzer, and C.-F. Tsang, 2008, Coupled reservoir-geomechanical analysis of the potential for tensile and shear failure associated with CO_2 injection in multilayered reservoir–caprock systems: International Journal of Rock Mechanics and Mining Sciences, **45**, 132–143, doi: https://doi.org/10.1016/j.ijrmms.2007.04.006.

Schnetzler, E. T., and D. L. Alumbaugh, 2017, The use of predictive analytics for hydrocarbon exploration in the Denver-Julesburg Basin: The Leading Edge, **36**, 227–233, doi: https://doi.org/10.1190/tle36030227.1.

Shabelansky, A. H., A. Malcolm, and M. Fehler, 2015, Monitoring viscosity changes from time-lapse seismic attenuation: case study from a heavy oil reservoir: Geophysical Prospecting, **63**, 1070–1085, doi:https://doi.org/10.1111/1365-https://doi.org/2478.12229.

Sinha, S., Y. Wen, R. P. De Lima, and K. Marfurt, 2018, Statistical Controls on Induced Seismicity: Proceedings of the 6th Unconventional Resources Technology Conference: doi:https://doi.org/10.15530/urtec-2018-2897507.

Swanston, A. M., P. B. Flemings, J. T. Comisky, and K. D. Best, 2003, Time-lapse imaging at Bullwinkle field, Green Canyon 65, offshore Gulf of Mexico: Geophysics, **68**, 1470–1484, doi: https://doi.org/10.1190/1.1620620.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, 2015, Rethinking the inception architecture for computer vision: arXiv e-prints, arXiv:1512.00567.

Wu, X., 2017, Directional structure-tensor-based coherence to detect seismic faults and channels: Geophysics, **82**, no. 2, A13–A17, doi: https://doi.org/10.1190/geo2016-0473.1.

Wu, X., L. Liang, Y. Shi, and S. Fomel, 2019, FaultSeg3D: using synthetic datasets to train an end-to-end convolutional neural network for 3D seismic fault segmentation: Geophysics, 1–36, doi:https://doi.org/10.1190/geo2018-0646.1.

Wu, H., B. Zhang, F. Li, and N. Liu, 2019, Semiautomatic first-arrival picking of microseismic events by using the pixel- wise convolutional image segmentation method: Geophysics, **84**, no. 3, V143–V155, doi: https://doi.org/10.1190/geo2018-0389.1.

Zhang, L., M. M. Poulton, and T. Wang, 2002, Borehole electrical resistivity modeling using neural networks: Geophysics, **67**, no. 6, 1790–1797, doi: https://doi.org/10.1190/1.1527079.

Zhao, T., F. Li, and K. J. Marfurt, 2018, Seismic attribute selection for unsupervised seismic facies analysis using user-guided data-adaptive weights: Geophysics, **83**, no. 2, O31–O44, doi: https://doi.org/10.1190/geo2017-0192.1.

Zhou, Z., Y. Lin, Y. Wu, Z. Wang, R. Dilmore, and G. Guthrie, 2018, Spatial-temporal densely connected convolutional networks: An application to CO_2 leakage detection: 88th Annual International Meeting, SEG, Expanded Abstracts, doi: https://doi.org/10.1190/segam2018-2998454.1.