Supporting Information

Unlocking Solutions: Innovative Approaches to Identifying and Mitigating the Environmental Impacts of Undocumented Orphan Wells in the United States

Daniel O'Malley[1], Andrew A. Delorey[1]*#, Eric J. Guiltinan[1], Zhiwei Ma[1], Teeratorn Kadeethum[2], Greg Lackey[3], James Lee[1], Javier E. Santos[1], Emily Follansbee[1], Manoj C. Nair[4,5], Natalie J. Pekney[3], Ismot Jahan[1], Mohamed Mehana[1], Priya Hora[2], J. William Carey[1], Andrew Govert[6], Charuleka Varadharajan[7], Fabio Ciulla[7], Sebastien C. Biraud[7], Preston Jordan[7], Mohit Dubey[7], Andre Santos[7], Yuxin Wu[7], Timothy J. Kneafsey[7], Manvendra K. Dubey[1], Chester J. Weiss[2], Christine Downs[2], Jade Boutot[8], Mary Kang[8], and Hari Viswanathan[1]*

[1]Los Alamos National Laboratory, Los Alamos NM 87545
[2]Sandia National Laboratory, Albuquerque NM 87123
[3]National Energy Technology Laboratory, Pittsburgh PA 15236
[4]National Oceanic and Atomospheric Administration, Washington DC 20230
[5]Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder 80309
[6]Department of Energy, Washington DC 20585
[7]Lawrence Berkeley Laboratory, Berkeley CA 94720
[8]McGill University, Montreal QC H3A 0G4

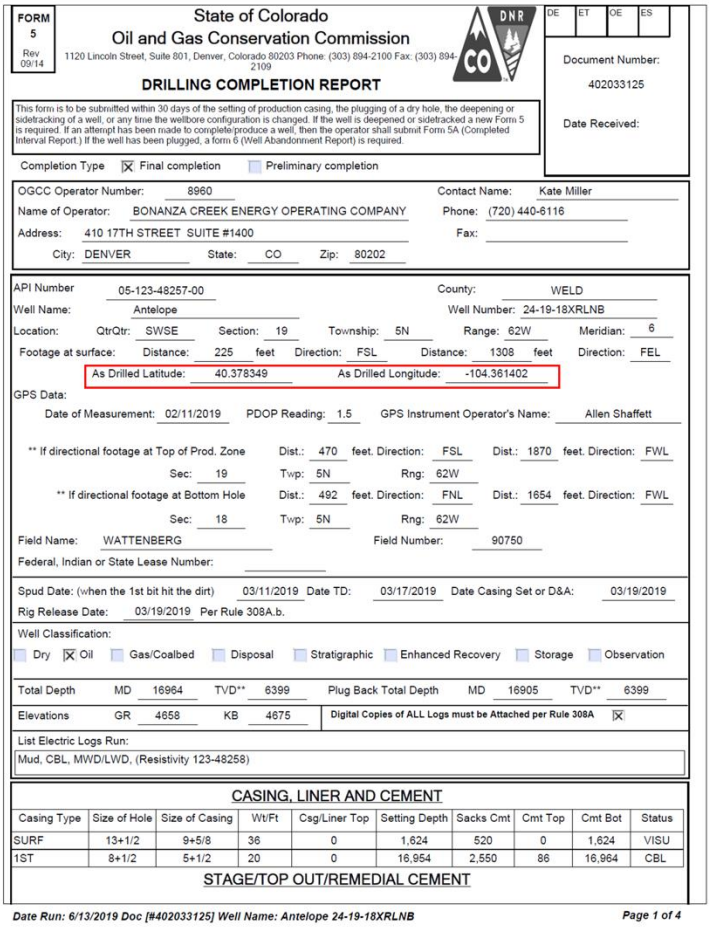
** Equal contributions,[#]Corresponding author*

Supporting Information consists of 7 sections labeled A-G, which includes 7 figures and 1 table.

There are 20 pages in total including references.

A. Details of large language modeling

In this section, we provide a summary of how to extract important information, i.e., well locations (longitude and latitude) from historical well documents by utilizing large language models (LLMs).



FORM 5 Rev 09/14

State of Colorado
Oil and Gas Conservation Commission
1120 Lincoln Street, Suite 801, Denver, Colorado 80203 Phone: (303) 894-2100 Fax: (303) 894-2109

DRILLING COMPLETION REPORT

DE ET OE ES

Document Number: 402033125

Date Received:

This form is to be submitted within 30 days of the setting of production casing, the plugging of a dry hole, the deepening or sidetracking of a well, or any time the wellbore configuration is changed. If the well is deepened or sidetracked a new Form 5 is required. If an attempt has been made to complete/produce a well, then the operator shall submit Form 5A (Completed Interval Report.) If the well has been plugged, a form 6 (Well Abandonment Report) is required.

Completion Type [X] Final completion [ ] Preliminary completion

OGCC Operator Number: 8960 Contact Name: Kate Miller
Name of Operator: BONANZA CREEK ENERGY OPERATING COMPANY Phone: (720) 440-6116
Address: 410 17TH STREET SUITE #1400 Fax:
City: DENVER State: CO Zip: 80202

API Number 05-123-48257-00 County: WELD
Well Name: Antelope Well Number: 24-19-18XRLNB
Location: QtrQtr: SWSE Section: 19 Township: 5N Range: 62W Meridian: 6
Footage at surface: Distance: 225 feet Direction: FSL Distance: 1308 feet Direction: FEL
As Drilled Latitude: 40.378349 As Drilled Longitude: -104.361402
GPS Data:
Date of Measurement: 02/11/2019 PDOP Reading: 1.5 GPS Instrument Operator's Name: Allen Shaffett

** If directional footage at Top of Prod. Zone Dist.: 470 feet. Direction: FSL Dist.: 1870 feet. Direction: FWL
Sec: 19 Twp: 5N Rng: 62W
** If directional footage at Bottom Hole Dist.: 492 feet. Direction: FNL Dist.: 1654 feet. Direction: FWL
Sec: 18 Twp: 5N Rng: 62W
Field Name: WATTENBERG Field Number: 90750
Federal, Indian or State Lease Number:

Spud Date: (when the 1st bit hit the dirt) 03/11/2019 Date TD: 03/17/2019 Date Casing Set or D&A: 03/19/2019
Rig Release Date: 03/19/2019 Per Rule 308A.b.

Well Classification:
[ ] Dry [X] Oil [ ] Gas/Coalbed [ ] Disposal [ ] Stratigraphic [ ] Enhanced Recovery [ ] Storage [ ] Observation

Total Depth MD 16964 TVD** 6399 Plug Back Total Depth MD 16905 TVD** 6399
Elevations GR 4658 KB 4675 Digital Copies of ALL Logs must be Attached per Rule 308A [X]

List Electric Logs Run:
Mud, CBL, MWD/LWD, (Resistivity 123-48258)

CASING, LINER AND CEMENT

Casing Type	Size of Hole	Size of Casing	Wt/Ft	Csg/Liner Top	Setting Depth	Sacks Cmt	Cmt Top	Cmt Bot	Status
SURF	13+1/2	9+5/8	36	0	1,624	520	0	1,624	VISU
1ST	8+1/2	5+1/2	20	0	16,954	2,550	86	16,964	CBL

STAGE/TOP OUT/REMEDIAL CEMENT

Date Run: 6/13/2019 Doc [#402033125] Well Name: Antelope 24-19-18XRLNB Page 1 of 4

Figure A1 – An example of a historical well completion report. The location of the well is represented as a latitude and longitude of 40.378349 and -104.361402, respectively.

To date, numerous LLMs have been developed for various natural language processing tasks, such as conversation, text generation, document analysis, translation, and question answering, e.g., LayoutLM[1], GPT-3[2], and BERT[3]. Our focus here is on using LLMs to answer questions based

on historical documents. The information extraction workflow of this work has two directions: one for LLM preparation and the other for document processing.

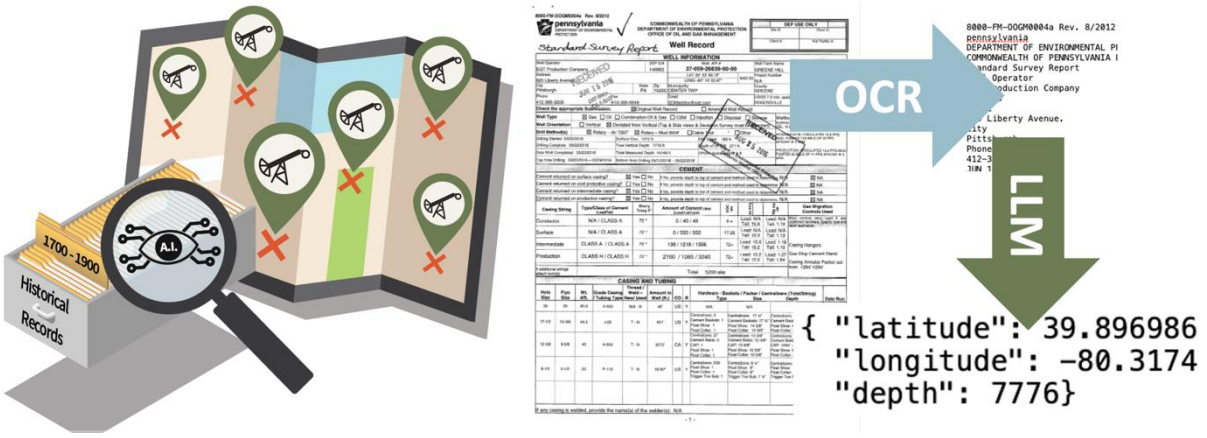


Figure A2: Information about inadequately documented oil & gas wells can be extracted from historical records using large language models. Here, we show information about the location and depth of a well extracted from a scanned document with complicating factors like stamps and handwritten notes. These forms sometimes have errors (such as the longitude having the wrong sign – indicating the wrong hemisphere of the globe) that large language models can automatically correct.

The detailed steps are illustrated here: At the beginning of LLM preparation, we need to select an appropriate question answering LLM from the public domain. A few tests are required to determine whether this model meets the requirements of this task. Model training or fine-tuning might be required if the initial model is not feasible for this task. On the other hand, we must process the historical documents during the data processing procedure once the historical data was collected. Because of the nature of historical documents, various pre-processing including data cleaning, denoising, and digitizing might be required. Next, we combine the LLM and historical document dataset for the question and answering tasks, during which customized questions can be prepared and subjected to the LLM (Figure A2). After processing/analyzing the

documents, it is anticipated that the LLM can provide answers to the questions based on the documents.

In this study, the DocQuery model was used to extract well location information from 150 historical documents from the Colorado Oil and Gas Conservation Commission as shown in Figure A1. The DocQuery model was trained based on Microsoft's LayoutLM model using SQuAD2.0 and DocVQA dataset, and it can analyze semi-structured and unstructured documents (PDFs, scanned images, etc.). We did not perform a retrain or fine-tuning procedure to DocQuery as it yields excellent results with 100% accuracy for the 150 documents. A data pre-processing process is not required here as the 150 PDF documents are text-based and do not have noisy information. To utilize the DocQuery for well location extraction, we asked two questions as shown here:

What is the latitude of the well?

What is the longitude of the well?

After processing the document and the questions, the DocQuery provides the correct well location in terms of latitude and longitude of 40.378349 and -104.361402, respectively, for this testing example. The location extraction time is around 1 to 2 seconds per document. It should be noted that the time for downloading the DocQuery model is not considered here, which depends on the speed of the internet. Once downloaded, the DocQuery model can be stored locally.

It is worth mentioning that the current DocQuery model struggles with a more complex dataset, e.g., hand-written documents. The future directions of this work include (1) fine-tuning existing LLM models with a customized dataset of questions and answers based on historical documents;

(2) implementing advanced LLMs to improve the extraction accuracy on the complex and challenging dataset.

B. Unlocking Insights from Satellite Imagery: Image Classification and Data Fusion Approaches for UW Detection

Our focus revolves around two main aspects. The first aspect pertains to image classification, where machine learning algorithms classify satellite images into distinct categories. For instance, these algorithms can differentiate between images containing undocumented orphan well (UOW) features and those that do not, as depicted in Figure B1(a). This approach facilitates the automated analysis of extensive satellite imagery volumes, yielding valuable insights for identifying UOWs. The underlying machine learning problem we tackle here is a binary classification problem[4,5]. Binary classification refers to a machine learning task that assigns input data points to one of two classes: either 0, indicating the absence of UOWs, or 1, denoting the presence of at least one UOW. The primary objective is to construct a model capable of accurately predicting the class label of new and unseen instances, leveraging the patterns and relationships it learns from the training data.

Here, our chosen classifier is RegNet_Y_400MF[6]. RegNet_Y_400MF is a specific architecture variant of the Regularized Network that has been specifically designed to be both efficient and scalable for a range of computer vision tasks. Its large parameter count, and computational efficiency characterize this model, making it particularly suitable for scenarios with limited resources. RegNet_Y_400MF can be effectively utilized in image classification, object detection, and semantic segmentation tasks. It offers a well-balanced trade-off between model size, computational efficiency, and performance, allowing optimal results in various applications.

The second aspect we delve into is data fusion, a crucial step in satellite data analysis. Satellite data often originates from various sources and sensors, each providing unique types of information. Using machine learning algorithms, data fusion facilitates the merging and integration of data from multiple input streams. This process leads to a more comprehensive and precise representation within analysis systems. In our context, we combine two distinct data streams: RGB-formatted satellite images[7] and persistent homology (PH)[8,9], as depicted in Figure B1(b). Persistent homology extracts features from a topological space by utilizing a defined function. Regarding remote sensing data, we can envision the data as a two-dimensional rough surface, where the pixel values describe its morphology. We employ digital elevation models (DEMs) and digital surface models (DSMs) to calculate PH[8]. We hypothesize that by incorporating

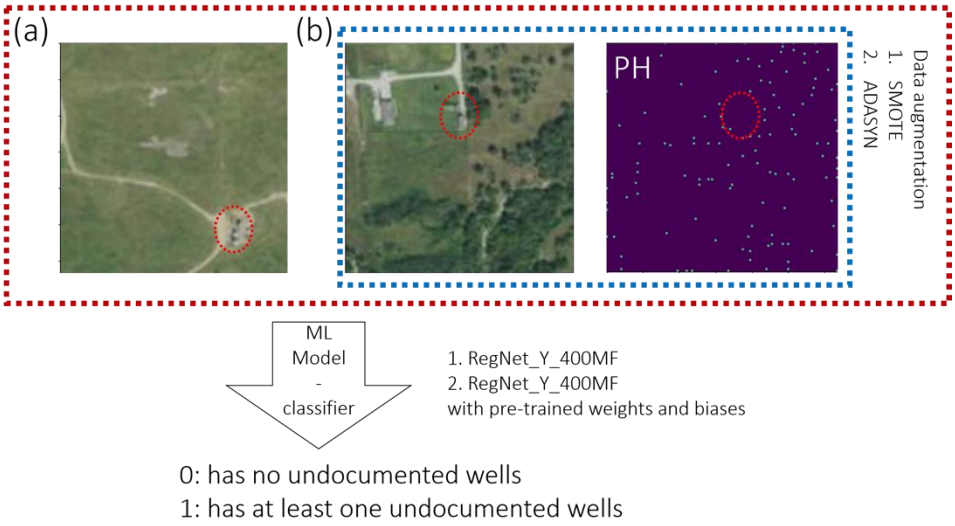


Figure B1. SNL approach: (a) using ML model with only satellite imagery and (b) using ML model with satellite imagery and PH as an input. Here, we mitigate a data imbalance problem (in which there are many more samples with no UOWs than ones with UOWs) by using data augmentation techniques of SMOTE and ADASYN). We also utilize a pre-trained model to improve the generalizability of our framework.

these additional input streams, we can enhance the ability of our machine-learning models to discern meaningful patterns. Consequently, this integration aims to improve the accuracy of our models in identifying UOWs.

satellite imagery	augmentation	F1 score (mean ± std)			
		without pre-trained model		with pre-trained model	
		in-distribution	out-of-distribution	in-distribution	out-of-distribution
without PH	none	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	SMOTE	0.93 ± 0.005	0.24 ± 0.060	0.78±0.004	0.71±0.010
	ADASYN	0.94 ± 0.005	0.26 ± 0.014	0.78±0.003	0.71±0.024
with PH	none	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	SMOTE	0.98±0.001	0.01±0.018	0.94±0.002	0.18±0.020

Table B1. results (mean ± std) of RegNet_Y_400MF model with different input and model configurations

The detection of UOWs often encounters a class imbalance issue in the available data. This means that the number of positive examples (UOWs) is considerably smaller than that of negative examples (non-UOWs). Such an imbalance can result in biased models that perform poorly on the minority class. To mitigate this issue, we employ various techniques, including oversampling the minority class using Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic (ADASYN)[10–12]. In brief, SMOTE generates synthetic samples for the minority class by interpolating between neighboring instances[11]. On the other hand, ADASYN focuses on the challenging-to-

learn instances of the minority class by assigning them higher weights during the generation process[12].

We present our results in Table B1, using the F1 score as our evaluation metric. The F1 score is a balanced measure that combines precision (the model's ability to identify positive predictions correctly) and recall (the model's ability to find all positive instances). It ranges from 0 to 1, with 1 being the best possible F1 score. In the absence of data augmentation techniques, our framework, with no oversampling, suffers from the data imbalance problem, resulting in an F1 score of 0.0. However, as we incorporate data augmentation techniques like SMOTE or ADASYN, we observe a significant improvement in the performance of our models.

In the absence of data fusion, where only satellite images are used as input but leveraging pre-trained models, the generalization ability of the models can be enhanced. Pre-trained models are initially trained on large-scale datasets (ImageNet in this case) before being employed for a specific task. This pre-training enables the models to learn from diverse data and extract useful features, which can aid in generalizing to unseen examples.

We also consider the performance on both in-distribution and out of-distribution data. Here, in-distribution refers to testing on examples from the same region (e.g., county or state) as the model was trained on. Out of-distribution testing refers to testing on examples from a different region. Specifically, when evaluating the models on in-distribution testing data, the F1 score may experience a decrease. However, when testing on out-of-distribution data, the F1 score shows a significant improvement, nearly tripling in value. This transferability of pre-trained models is advantageous as it allows us to utilize models trained on one dataset, such as Oklahoma, for areas like New York.

On the other hand, incorporating PH as an additional input (i.e., using both satellite images and PH) can enhance performance for in-distribution testing. However, it may compromise the model's generalization ability when dealing with out-of-distribution testing scenarios. Therefore, careful consideration should be given to the inclusion of PH, weighing its impact on the model's overall performance and its ability to generalize across different testing scenarios.

C. Location extraction from historic maps

Our source dataset is the Historical Topographic Maps Collection (HTMC)[13], which is a digital archive of approximately 190,000 georeferenced topographic maps published by the USGS from 1884 to 2006 covering the US. Within the HTMC is a series of maps, which are referred to as quadrangles, that have consistent colors and symbols for features. The quadrangle maps include almost 200 symbols, while the background colors determine the type of surface features (e.g., forests and water bodies). The symbols generally consist of simple geometric shapes, with oil and gas wells represented by black circles. Because of their relatively simple and consistent shape, the identification of well symbols via traditional computer vision techniques such as color clustering, edge detection and template matching is a viable option. More modern methods that leverage neural networks for computer vision have been proven to be efficient in segmenting images, the action of partitioning an image into areas containing objects of interest, and robust against input variations.

For the purpose of identifying UOWs, we trained a U-Net convolutional neural network[14] with hand-labeled data from 50 different maps for the task of wells symbol extraction. The value of the intersection-over-union, a metric that measures the performance of a segmentation algorithm, after training is equal to 0.8 in the validation set. When applied to quadrangle maps, the algorithm detects the presence of oil and gas well symbols with a precision of 0.98 on the test set. Because the maps are georeferenced, the pixel location in the image can be translated into geographical coordinates. The position of the detected wells is compared to the ones from official state databases in California and Oklahoma. A fixed spatial buffer is used to label a well as a potential UOW or IOW. We additionally performed a visual assessment using current satellite

imagery or historical aerial photographs. Once the potential well locations are identified, they are provided to the field teams for further investigation to verify the existence of the wells.

D. Time domain reflectometry

Characterization of an undocumented orphan well is performed after discovering the well. There are two goals: 1) determine whether the well is leaking pollutants to the atmosphere, groundwater, soil and assess the extent and severity of any environmental hazard; 2) determine the physical characteristics of the well (depth, casing, cement, condition, etc.) that are needed for plugging and abandonment.

Leakage to the atmosphere is relatively straightforward to detect, and we have already described our approach to quantify the amount of methane leakage. Other gases of concern include hydrogen sulfide[15] and various volatile hydrocarbons[16]. Detection of any of these gases raises the priority of remediating the well and plugging and abandoning it.

Leakage to soil can often be detected by visual examination where the primary concerns are high salinity brines and hydrocarbons. If substantial, this leakage is likely to impact vegetation or surface soil which can be readily identified either visually or utilize available sensors, e.g., x-ray fluorescence to identify elements characteristic to leakage. It is also possible to use remote sensing techniques to observe discoloration of soil and vegetation. In this case, soil impacts can be used to identify the location of the UOW. Once the problem has been identified, conventional soil-sampling techniques can be used to assess the extent of pollution.

Leakage to groundwater may be more difficult to detect, and often requires geophysical technologies for characterization and mapping. An array of ground, handheld and airborne electrical and electromagnetic (EM) technologies are available, including those that have been widely used for hydro-geophysical research. By identifying conductivity anomalies caused by brine invasion and associated mineral precipitation, these geophysical tools have been

demonstrated for groundwater contaminant plume identification and delineation. Generally, direct sampling or wellbore confirmation is desired for geophysical data validation and calibration. Thus, our recommended practice is to first determine if nearby drinking water wells or surface waters show evidence of abnormal salinity or hydrocarbons[17].

As cost is always a factor, we first seek to learn as much about the well as possible without resorting to expensive down-well techniques. We are testing the use of acoustic and electromagnetic time-domain-reflectometry (TDR) methods to characterize well depth (Figure D1) and access general casing conditions in terms of presence of major damages. The TDR methods rely on impedance contrast caused by anomalies on the casing, e.g., due to damage, and are non-invasive, quick, and easy to deploy without downhole wireline or equipment deployment. These novel technologies are aimed as screening tools for rapid borehole condition assessments to guide further characterization needs. While these techniques are conceptually straightforward for a simple tubular pipe, wells are typically more complex with multiple casing layers, with a production casing string consisting of individual casing lengths (about 40 ft/13 m each) with threaded connections; the presence of cement outside each casing layer; the presence of devices such as packers inside the casing; the presence of cement plug and other materials inside the casing; and the possible presence of casing damage such as corrosion or deformation. We will test these techniques on known wells to determine the efficacy of this approach.

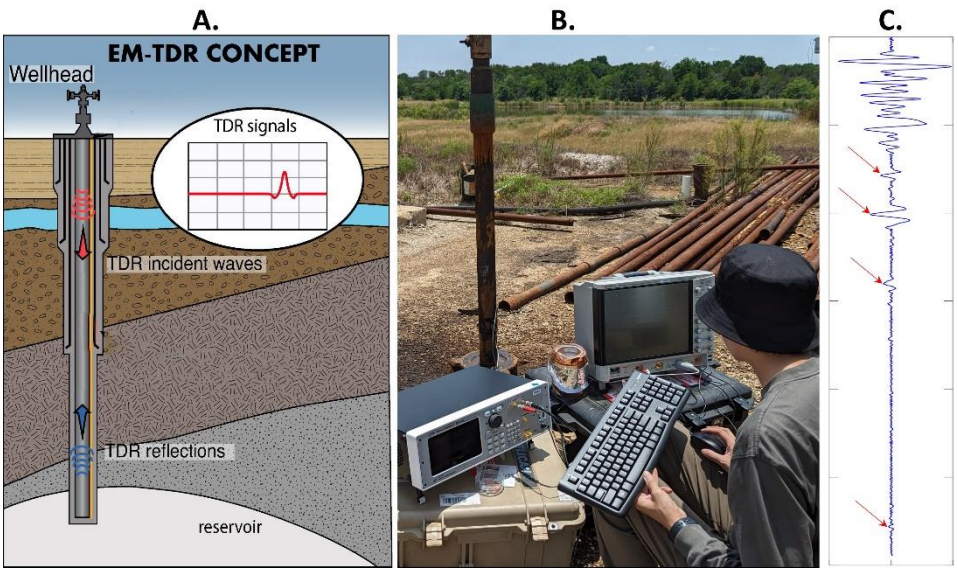


Figure D1. Electromagnetic Time Domain Reflectometry (EM-TDR) for non-invasive borehole condition screening. A. concept of borehole condition monitoring using guided wave EM-TDR; B. Testing of prototype EM-TDR system at an O&G well in Oklahoma; C. An example of EM-TDR data showing identified features, e.g., damages or bottom, indicated with the red arrows. (photo credit: Yuxin Wu)

E. Magnetometry

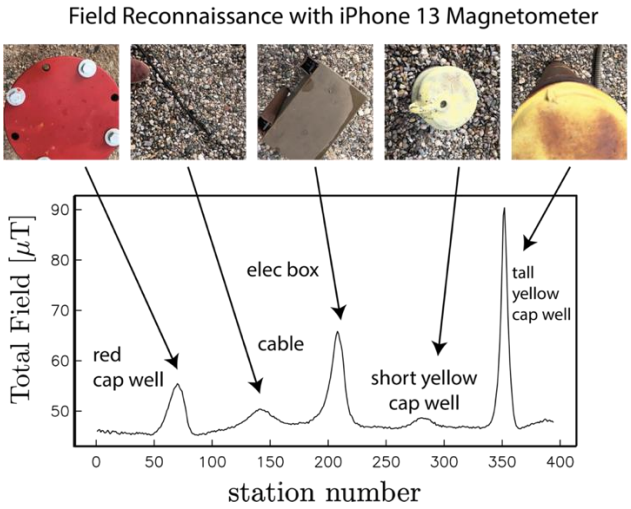


Figure E1. Rapid iPhone13 magnetic field reconnaissance test along a roughly 30m profile 1.5 m over a few magnetically permeable targets. Data are uncorrected for drift or static shift but are stacked over a 0.01 s window to reduce noise seen in the 100 Hz sample rate.

F. Machine learning using multi-sensor drone data

Figure F1. Left: E400 Fixed-wing drone by Event 38 equipped with a Sensys R3 magnetometer. Middle: HX8 by Harris Aerial with Geometrics MagArrow in the foreground. Right: Four HX8 flight paths were reconstructed compared to 1 flight utilizing half the battery of the E400. (photo credit: Eric Guiltinan)

G. Acoustic Methods

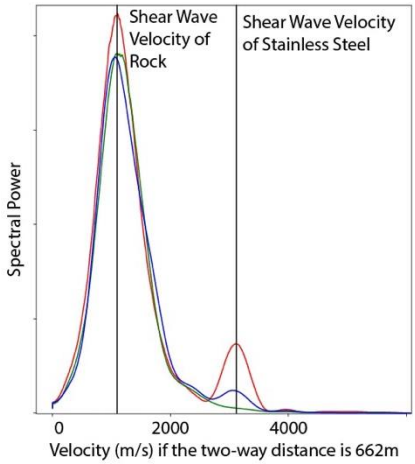


Figure G1. Spectra for autocorrelation with horizontal axis converted to velocity if the travel distance is the two-way distance through the well casing. Red, green, blue indicate east, north, vertical, respectively.

References

(1) Xu, Y.; Li, M.; Cui, L.; Huang, S.; Wei, F.; Zhou, M. LayoutLM: Pre-Training of Text and Layout for Document Image Understanding. **2019**. https://doi.org/10.48550/ARXIV.1912.13318.

(2) Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; Amodei, D. Language Models Are Few-Shot Learners. **2020**. https://doi.org/10.48550/ARXIV.2005.14165.

(3) Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. **2018**. https://doi.org/10.48550/ARXIV.1810.04805.

(4) Lorena, A. C.; De Siqueira, M. F.; De Giovanni, R.; De Carvalho, A. C. P. L. F.; Prati, R. C. Potential Distribution Modelling Using Machine Learning. In *New Frontiers in Applied Artificial Intelligence*; Nguyen, N. T., Borzemski, L., Grzech, A., Ali, M., Eds.; Lecture Notes in Computer Science; Springer Berlin Heidelberg: Berlin, Heidelberg, 2008; Vol. 5027, pp 255–264. https://doi.org/10.1007/978-3-540-69052-8_27.

(5) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521* (7553), 436–444. https://doi.org/10.1038/nature14539.

(6) Radosavovic, I.; Kosaraju, R. P.; Girshick, R.; He, K.; Dollár, P. Designing Network Design Spaces. arXiv 2020. https://doi.org/10.48550/ARXIV.2003.13678.

(7) U.S. Department of Agriculture. National Agriculture Imagery Program Aerial Imagery, 2021. https://naip-image-dates-usdaonline.hub.arcgis.com/ (accessed 2022-06-27).

(8) Bauer, U. Ripser: Efficient Computation of Vietoris–Rips Persistence Barcodes. *J Appl. and Comput. Topology* **2021**, *5* (3), 391–423. https://doi.org/10.1007/s41468-021-00071-5.

(9) Natural Resources Conservation Service. 2-Meter LiDAR Digital Elevation Model, 2010. https://okmaps.org/ogi/search.aspx (accessed 2022-07-07).

(10) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(11) Blagus, R.; Lusa, L. SMOTE for High-Dimensional Class-Imbalanced Data. *BMC Bioinformatics* **2013**, *14* (1), 106. https://doi.org/10.1186/1471-2105-14-106.

(12) Haibo He; Yang Bai; Garcia, E. A.; Shutao Li. ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*; IEEE: Hong Kong, China, 2008; pp 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969.

(13) USGS. Historical Topographic Maps - Preserving the Past. https://www.usgs.gov/programs/national-geospatial-program/historical-topographic-maps-preserving-past.

(14) Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*; Navab, N., Hornegger, J., Wells, W. M., Frangi, A. F., Eds.; Lecture Notes in Computer

Science; Springer International Publishing: Cham, 2015; Vol. 9351, pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
(15) El Hachem, K.; Kang, M. Methane and Hydrogen Sulfide Emissions from Abandoned, Active, and Marginally Producing Oil and Gas Wells in Ontario, Canada. *Science of The Total Environment* **2022**, *823*, 153491. https://doi.org/10.1016/j.scitotenv.2022.153491.
(16) DiGiulio, D. C.; Rossi, R. J.; Lebel, E. D.; Bilsback, K. R.; Michanowicz, D. R.; Shonkoff, S. B. C. Chemical Characterization of Natural Gas Leaking from Abandoned Oil and Gas Wells in Western Pennsylvania. *ACS Omega* **2023**, *8* (22), 19443–19454. https://doi.org/10.1021/acsomega.3c00676.
(17) Jackson, R. B.; Vengosh, A.; Darrah, T. H.; Warner, N. R.; Down, A.; Poreda, R. J.; Osborn, S. G.; Zhao, K.; Karr, J. D. Increased Stray Gas Abundance in a Subset of Drinking Water Wells near Marcellus Shale Gas Extraction. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (28), 11250–11255. https://doi.org/10.1073/pnas.1221635110.