Earth and Space Science

RESEARCH LETTER
10.1029/2025EA004227

Identification of Natural Hydrogen Seeps: Leveraging AI for Automated Classification of Sub-Circular Depressions

N. Ginzburg[1], J. Daynac[2], S. Hesni[2], U. Geymond[1,3], and V. Roche[2]

[1]Institut de Physique du Globe de Paris, Université Paris Cité, CNRS, Paris, France, [2]Laboratoire de Planétologie et Géosciences, LPG UMR 6112, CNRS, Le Mans Université, University Angers, Nantes Université, Le Mans, France, [3]IFP Énergies Nouvelles, Rueil-Malmaison, France

Key Points:

- Deep learning image classification improves the accuracy of identifying subcircular depressions using satellite and high-resolution imagery
- High-resolution imagery outperforms spectral indices in classifying hydrogen-related depressions, aiding efficient exploration
- The method tested in Brazil confirms promising hydrogen emission zones and could be extended worldwide

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

V. Roche,
Vincent.roche@univ-lemans.fr

Citation:

Ginzburg, N., Daynac, J., Hesni, S., Geymond, U., & Roche, V. (2025). Identification of natural hydrogen seeps: Leveraging AI for automated classification of sub-circular depressions. *Earth and Space Science*, *12*, e2025EA004227. https://doi.org/10.1029/2025EA004227

Received 16 JAN 2025
Accepted 28 APR 2025

Author Contributions:

Conceptualization: U. Geymond, V. Roche
Data curation: N. Ginzburg, J. Daynac, S. Hesni
Formal analysis: N. Ginzburg, J. Daynac, S. Hesni
Funding acquisition: V. Roche
Investigation: N. Ginzburg, J. Daynac, S. Hesni, U. Geymond, V. Roche
Methodology: U. Geymond, V. Roche
Software: N. Ginzburg, J. Daynac, S. Hesni
Supervision: U. Geymond, V. Roche
Validation: J. Daynac, S. Hesni, U. Geymond, V. Roche
Visualization: N. Ginzburg, J. Daynac

Abstract Hydrogen has long been used as an energy vector, but the recent discovery of natural hydrogen (H_2) opens the door for its use as a direct energy source. Identifying H_2 seepages is therefore crucial to advance exploration. Although the scientific community does not yet fully understand the parameters controlling H_2 leaks from underground, sub-circular depressions (SCDs) appear to be key indicators associated with these emissions. However, distinguishing SCDs from similar landforms remains a challenge. This study leverages open-source multispectral and high-resolution imagery to train a deep learning model (YOLOv8) for classifying rounded landforms and detecting H_2-related structures (i.e., SCDs). The model achieved 90% accuracy with Google Maps© imagery, outperforming Sentinel-2 multispectral data. Applied to a pre-existing data set from Brazil, the model allowed a large-scale screening, discarding 52% of the structures as non-H_2 emitting ones and pinpointing high-potential areas for field validation. Future enhancements, including, for example, higher-resolution input data and morphometric analysis, would aim to reduce false positives and boost predictive accuracy. This approach significantly improves H_2 exploration efficiency, with global applicability including some region-specific adjustments during post-processing analyses.

Plain Language Summary This study explores a new way to identify natural hydrogen (H_2) emissions, which can originate from deep accumulation leakages of a possible new clean and sustainable energy. H_2 seeps, often linked to sub-circular depressions in the landscape, are promising but challenging to find and discriminate from other rounded landforms. Using artificial intelligence (AI) combined with satellite images, we developed a method to pinpoint these features automatically. By comparing images with different levels of resolution, we found that high-resolution images from sources like Google Maps© were the most effective, identifying H_2-related features with 90% accuracy. We tested this approach in Brazil's São Francisco Basin, analyzing approximately two thousand structures. This technique can be applied worldwide, offering a faster and more reliable way to find H_2 emissions and supporting the exploration of cleaner energy. Future improvements could make it even more accurate by including more precise input data and refining how features are classified.

1. Introduction

Driven by increasing demands from the steel and fertilizer industries, and the push to meet net-zero emissions targets, global hydrogen consumption has increased in recent years, rising from 60 Mt in 2000 to 100 Mt in 2024, with further acceleration expected (IEA, 2024). However, current hydrogen production methods - such as energy-intensive electrolysis and CO_2-emitting steam methane reforming - raise concerns about their sustainability (Lapi et al., 2022). Encouragingly, recent studies suggest that the potential of natural hydrogen (hereafter referred to as H_2) has been significantly underestimated (Worman et al., 2020; Zgonnik, 2020), opening the door to more cost-effective and environmentally friendly production options.

H_2 seeps are getting more and more documented in continental regions (Lefeuvre et al., 2021; Lévy, Boka-Mene, et al., 2023; Prinzhofer et al., 2019; Truche et al., 2024), and are thought to result from deep subsurface H_2 generation that migrates upwards through a wide range of processes, including mainly diffusion and advection (Donzé et al., 2024; Lefeuvre et al., 2022; Strauch et al., 2023). In parallel, recent findings in Mali (Maiga et al., 2023a, 2023b; Prinzhofer et al., 2018) and central Australia (Leila et al., 2022) support the presence of H_2 accumulations beneath the Earth's surface. They suggest the existence of H_2 systems similar to conventional petroleum plays (Doust, 2010; Magoon, 1995), typically involving a source rock, a reservoir, and a cap rock (Gaucher et al., 2023; Prinzhofer & Cacas-Stentz, 2023). In that sense, exploring H_2 seeps at the surface could prove crucial for

Writing – original draft: N. Ginzburg, J. Daynac, S. Hesni, U. Geymond, V. Roche
Writing – review & editing: U. Geymond, V. Roche

exploration, as they may indicate underlying H_2 accumulations and reservoir leakage, much like pockmarks that correspond to surface signatures of gas leakage in petroleum systems (Hovland & Judd, 1988; Pinet et al., 2008).

Several studies have reported H_2 concentrations in sub-circular depressions (called here SCDs) in continental regions. These features have been documented in various countries such as Russia (Larin et al., 2015), the USA (Zgonnik et al., 2015), Brazil (Moretti, Prinzhofer, et al., 2021; Prinzhofer et al., 2019), Australia (Davies et al., 2024; Frery et al., 2021), and Namibia (Moretti et al., 2022; Roche et al., 2024). Typically ranging from 10 to several hundred meters in diameter, SCDs often coincide with gaps in vegetation and topographic depressions of up to a few meters. Their occurrence is concomitant with high H_2 concentrations detected during soil measurements, reaching up to several 1,000 ppm H_2 in long-term monitoring (Moretti, Prinzhofer, et al., 2021; Prinzhofer et al., 2019). While their formation dynamics remain unclear, they are generally believed to result from gas migrating upward through soft sedimentary layers like pockmarks (Lévy, Boka-Mene, et al., 2023), even though the measured H_2 signals do not allow for a clear identification of their origin (e.g., Etiope et al., 2024). The fact that these structures create visible surface expressions, identifiable in satellite imagery, makes SCDs promising targets for H_2 exploration. However, distinguishing them from other similar rounded landforms, such as salt lakes, karstic dolines, or impact craters, remains challenging (Lévy, Roche, et al., 2023). This is partly explained by the challenge of defining H_2 seeps, as the processes governing H_2 migration and accumulation are still not fully understood (e.g., Lodhia et al., 2024). Some studies have focused on the morphological parameters of SCDs (e.g., width, depth, slope, as proposed by Moretti, Brouilly, et al., 2021; Malvoisin & Brunet, 2023), while others have applied multispectral indices, such as vegetation index (Aimar et al., 2023; Carrillo Ramirez et al., 2023; Lévy, Roche, et al., 2023; Moretti et al., 2022), to distinguish SCDs from other landforms. Despite these efforts, there is still no consensus on a standardized method to classify these rounded landforms efficiently. Developing a robust methodology will enhance the likelihood of discovering H_2 in the future.

Recent advances in artificial intelligence (AI) and its increasing application in various geoscience disciplines offer unique, automated solutions to the challenges of mapping terrestrial and extraterrestrial landforms (Barrett et al., 2022; Daynac et al., 2024; Marochov et al., 2021). Such use of AI could, over time, pave the way for SCDs detection and thus for H_2 exploration, unlocking a better understanding of the relationship between the spatial distribution of H_2 seepages and subsurface generation and migration processes. In this study, we combine our fieldwork and AI expertise to present an automated method for classifying rounded landforms, and more specially the SCDs. Our approach involves satellite data sets for training the DL model. It evaluates the most effective input data—whether multispectral or visible imagery—to achieve the highest true positive recognition rate. By using only open-access data (Sentinel-2 and Google Maps© images) and the YOLOv8 algorithm we demonstrate that AI-driven H_2 exploration is effective and accessible to all.

2. Materials and Methods

2.1. Selection of the Rounded Landforms

To develop an automated classification tool, a diverse array of rounded landforms from various regions worldwide were selected (Figure 1). All identified features were systematically categorized into two distinct groups. Group 1 corresponds to SCDs (i.e., rounded landforms where elevated H_2 concentrations have been measured in the field), selected from various countries (Russia, the USA, Brazil, and Namibia, see Figure 1). In addition, structures presenting the same characteristics as those where H_2 concentrations have been measured, and located nearby (within a few kilometers), were also used for classification. Group 2 corresponds to other rounded landforms that do not emit H_2. It includes eight categories: (a) Fairy circles—circular barren patches found in arid grasslands, notably in the Namib Desert, southern African continent (Getzin et al., 2021); (b) Fairy forts—remains of stone circles, ringforts, hillforts, or other circular prehistoric dwellings in Ireland; (c) Farm circles—circular farm fields; (d) Flooded dunes—dunes that have been filled with water; (e) Impact craters - depressions in the Earth's surface formed by the hypervelocity impact of smaller objects; (f) Karstic features—structures formed by karst processes; (g) Salt lakes—lakes with salt and mineral concentrations significantly higher than typical lakes; (h) Thermokarst lakes—shallow depressions formed by the melting of permafrost. For Group 1 and each category of Group 2, 60 structures were mapped to ensure a balanced representation across all types, resulting in a total of 540 structures to train the model (Table 1). Coordinates of the selected structures may be found in Table S1 in Supporting Information S1.

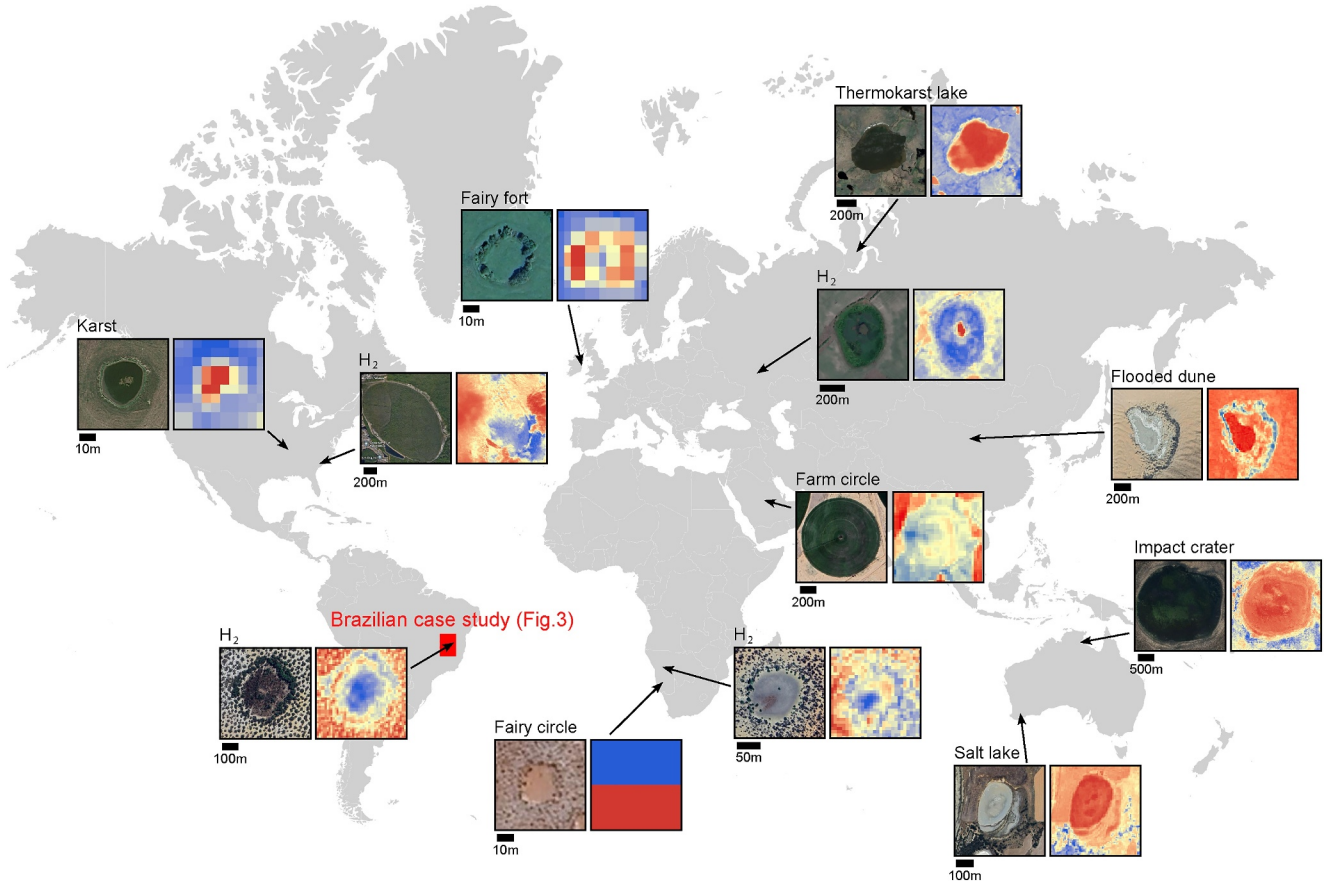

Figure 1. Examples of structures used in the study. Google Maps© images are displayed in natural colors, while red-and-blue images correspond to multispectral indices (here Normalized Difference Vegetation Index). They constitute the two data sets used to train the model (see Sections 2.2 and 2.3). The red rectangle indicates the location of the Brazilian case study, on which the efficiency of the methodology was tested (see Section 3.2).

2.2. Building the Input Data Sets

Following the selection of the various rounded landforms, two distinct data sets were created to train the model and assess their relevance for AI-based classification. The first data set comprised multispectral index rasters of the landforms, as recent studies have highlighted their efficiency in identifying SCDs (Aimar et al., 2023; Lévy, Roche, et al., 2023; Moretti et al., 2022). While these previous researches employed multiple indices, their respective effectiveness had not been systematically evaluated. To address this, nine indices were tested to identify the most relevant features for SCDs classification (Text S1, Table S2 in Supporting Information S1). A

Table 1
Selection of Rounded Landforms Used to Train the Artificial Intelligence Classification Model

	Group 1		Group 2—(not H_2-related landforms)						
	SCD	Fairy circle	Fairy fort	Farm circle	Flooded dune	Impact crater	Karst	Salt lake	Thermokarst
Brazil	24			10		3			
Namibia	18	60		10		2			
Russia	12								30
USA	6			20			45	4	10
Australia						13		56	
Other			60	20	60	42	5		20

Note. The whole data set is available in Table S1 in Supporting Information S1.

cross-correlation analysis (Figure S1c in Supporting Information S1) revealed strong interdependencies among them, limiting their individual contribution to classification. Ultimately, the Normalized Difference Vegetation Index (NDVI) was selected due to its proven effectiveness in highlighting vegetation anomalies at SCD locations, while the Brightness Index (BI) was retained for its low correlation with NDVI and its role in distinguishing SCDs from salt-related formations. Further details on the indices selection process can be found in Supporting Information S1. Additionally, as proposed by Lévy, Roche, et al. (2023), the seasonal timing of multispectral image acquisition was considered because it affects indices clarity due to changing weather conditions (e.g., snow cover). After investigation, dry-season multispectral data were preferred to wet-season data, since they better highlight indices signatures of SCDs (see Figure S1 and Text S2 in Supporting Information S1).

One of the challenges in analyzing multispectral index signatures of rounded landforms is the spatial resolution of the satellite imagery, given the small size of some structures under investigation (e.g., as small as 10 m in diameter, as shown in Figure 1). Because the goal of this study was to focus exclusively on freely available data, Sentinel-2 satellite imagery was selected. It provides extensive coverage of the Earth's surface at a 10-m resolution, with some bands limited to 20- or 60-m resolution (https://sentiwiki.copernicus.eu/web/s2-mission). To overcome this limited resolution of open-access multispectral data, high-resolution (<10 m) Google Maps© visible imagery was also investigated. However, it is important to note that Google Maps© does not provide images with a precise acquisition date, and the platform aggregates data from different periods, potentially introducing variability in surface conditions. Unlike Sentinel-2, where seasonal variations can be accounted for, Google Maps© images may reflect a mixture of seasonal and temporal data without clear metadata. This limitation should be considered in future refinements of the model, particularly for regions where surface conditions vary significantly over time. The results of AI training involving the two data sets were compared to each other.

Both types of data feeding the two data sets were downloaded by developing a Python script (Text S3 in Supporting Information S1) and running it using the open-access Google Engine©. This script enables the opening of a location in Google Maps© using specified coordinates and then capturing a screenshot containing a visible image of the rounded landform of interest. All Google Maps© screenshots were taken with a single structure per image, adjusting the zoom level to align the image boundaries with the structure's limits.

2.3. Training the AI Model

The two data sets built previously, based respectively on multispectral indices and visible images, were then used to train the DL algorithm. All structures were classified using supervised learning of the pre-trained Ultralytics YOLOv8 model (https://github.com/ultralytics/ultralytics). YOLO is an open-source DL model for computer vision that supports tasks such as detection, segmentation, pose estimation, tracking and clssification. YOLO utilizes a Convolutional Neural Network (CNN) architecture for image classification. When using a CNN, the input image undergoes convolutional operations where filters slide over the image to extract features such as edges, textures, and patterns (LeCun et al., 2015). These features are then passed through pooling layers to reduce dimensionality while retaining essential information. The extracted features are processed through fully connected layers during training, where the model learns to identify patterns and adjust its weights. Once trained, these weights are used to predict class probabilities on new data, indicating the likelihood that the image belongs to each predefined category (LeCun et al., 2015). To expedite the training process, code was executed in Google Collaboratory (Bisong, 2019), leveraging Google's cloud servers and their graphics processing units (e.g., T4 GPU). 450 images were used, and the ratio of the training data set to the validation data set was 80% and 20%, with 10 images per category (90 in total) used for the validation.

3. Results and Discussion

3.1. Training Results

The performance of both AI models trained was evaluated using a confusion matrix, which highlights the ratio of correctly and incorrectly predicted images for each category (Figure 2). In this study, the confusion matrix includes predictions across the nine categories. Higher values along the main diagonal indicate better model performance. While particular attention is given to the classification of SCDs, the performance across other categories also provides valuable insights, especially for recognizing different types of structures.

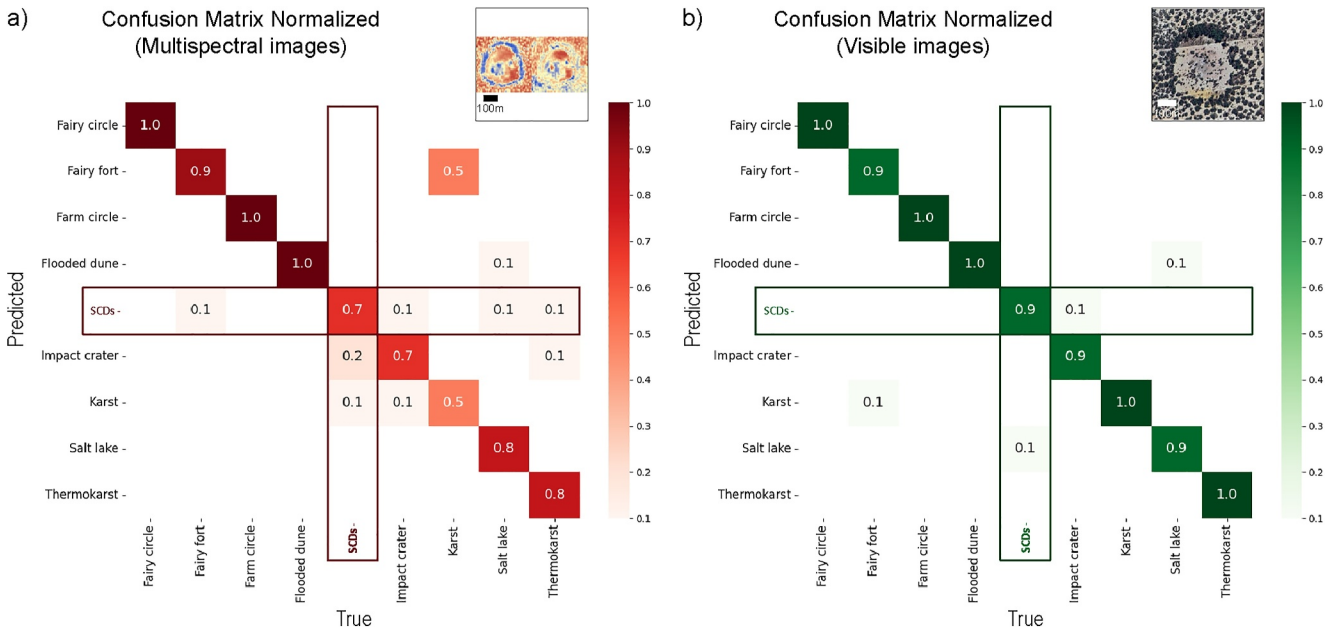

Figure 2. Confusion matrices obtained from the classification models: (a) Model trained using Normalized Difference Vegetation Index (left) and Brightness Index (right) multispectral indices. (b) Model trained using Google Maps© images. Highlighted columns and rows correspond to the sub-circular depressions category. Note that for both models 50 epochs were sufficient to train the model (Figure S2 in Supporting Information S1).

The confusion matrix associated with the NDVI and BI indices model (Figure 2a) correctly identified 70% of the instances as true positives for the SCDs, but misclassified 10% as karst and 20% as impact crater categories. Additionally, 10% of instances from the fairy fort, 14% from impact crater, 10% from salt lake, and 10% from thermokarst were incorrectly classified as SCDs (false positive). In contrast, the Google Maps© model (Figure 2b) achieved a 90% true positive rate for SCDs, with 10% misclassified as salt lake and 14% from impact crater incorrectly classified as SCDs (false positive).

The comparison of the two confusion matrices indicates good results for both model trainings, suggesting the effectiveness of AI for rounded landforms classification. The model trained on Google Maps© imagery performs even better, likely due to the higher resolution of the input data and its ability to capture more complex morphological patterns, including surface texture variations, shading effects, and contextual relationships between theses morphological features, which enhance its ability to distinguish SCDs from other rounded landforms. However, irrespective of the input data set used, the misclassifications may arise from (a) data quality, which for Google Maps© images can vary from one region to another, leading to inconsistencies in classification; and (b) the confusion of SCDs with impact craters and salt lakes due to their similar morphological characteristics, thus highlighting the importance of post-processing using additional criteria to improve differentiation.

3.2. Application of the Classification: The São Francisco Case Study

After training, the Google Maps-based model was applied to classifying a preexisting data set of suspected SCDs from the Brazilian São Francisco Basin (see location in Figure 1). In this basin, the presence of H_2 in some depression has been validated in the past by punctual soil measurements and long-term monitoring (Moretti, Prinzhofer, et al., 2021; Prinzhofer et al., 2019). Lately, De Freitas et al. (2024) mapped thousands of additional rounded landforms in the basin, considering all of them as SCDs, without further fieldwork validation. Thus, using this extended data set from the São Francisco Basin (a) constitutes an effective tool to validate the applicability of our AI classification method, and (b) provides the possibility to investigate whether the rounded landforms present in the basin all correspond to SCDs.

The Brazilian data set was divided into 11 clusters based on the location of the rounded landforms (Figure 3a), and each cluster was then processed separately through the model. Each landform was then categorized into two groups: H_2 structures (i.e., structures predicted to belong to the SCDs category with a confidence level >0.5) and

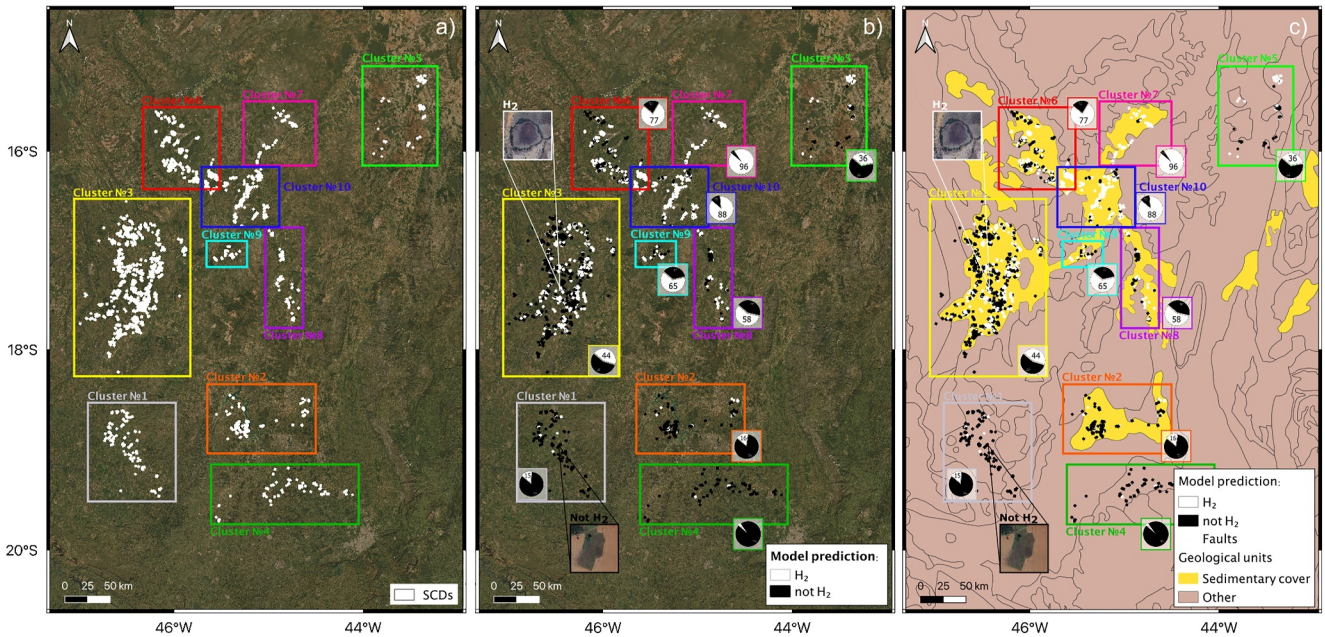

Figure 3. (a) Cluster map of the Brazilian data set from De Freitas et al. (2024). The location of the area is shown as a gray rectangle in Figure 1. (b) Model prediction map after applying our classification methodology. Pie charts show the ratio of predicted "H_2"and "not H_2" categories in each cluster. The small images below show examples of classified structures, constituting the starting data set. ESRI satellite imagery was used as the base map. (c) Example of a post-processing tool to improve the exploration strategy. The geological background is used here to show that most sub-circular depressions are located in sedimentary basins.

non-H_2 structures (i.e., all other types of rounded landforms). At the basin scale, 52% of the landforms were discarded during processing (i.e., +1,000 structures), and were therefore considered as not H_2-related structures (Figure 3b). At the local scale, a strong variability is visible between clusters, with some of them reaching a close-to-100% value of expected H_2-structures (e.g., Cluster n°6) while other clusters reaching a close-to- 0% value of expected H_2-structures (e.g., Cluster n°4). It is self-evident that the way of clustering may affect the percentage of positive responses attributed to each cluster. In any case, according to the model predictions, the majority of SCDs are concentrated in the northern part of the region, where some of the rounded landforms have already been confirmed to emit H_2 by fieldwork (Prinzhofer et al., 2019). Interestingly, a simple overlay of the modeling results with a first-order geological map showing the distribution of soft Cenozoic sedimentary cover (Figure 3c), reveals a correlation between the less promising southern clusters (clusters n°1 and n°4) and the absence of sediments. This correlation highlights the consistency between the regional geology and the expected absence of SCDs. Although it was not further investigated here, it provides evidence that using additional data may help further refine the classification, either as direct input data for the predictive AI classification model processing or as post-processing to help filter the results. For example, these additional data may consist of geological-hydrogeological information (e.g., shape of the drainage basin, depth of the water table, presence of regional fault to support deep H_2 migration) or rounded landform morphometric information (e.g., size, diameter, depth, circularity). Furthermore, although our results align with the field observations reported by De Freitas et al. (2024), no independent field validation was carried out in this study. Future work should incorporate independent in situ gas measurements to confirm AI-based predictions and refine the model's suitability.

3.3. General Workflow and Perspectives for Future H_2 Exploration

The methodology presented in this study offers a novel approach for quickly targeting potential H_2 seep areas, giving field researchers more precise indications for evaluating emission viability. It enables an automated re-evaluation of entire data sets, a task that would have been challenging to accomplish manually (e.g., the Brazilian São Francisco basin case). Ultimately, this approach aims to quantify the likelihood of detecting the presence of H_2 seeps in a given region, for which predictions are deemed promising enough to justify fieldwork for validating H_2 emissions. To summarize, we propose a three-step workflow for automatically classifying rounded landforms and identifying which of them are the most likely related to H_2 emissions (Figure 4):

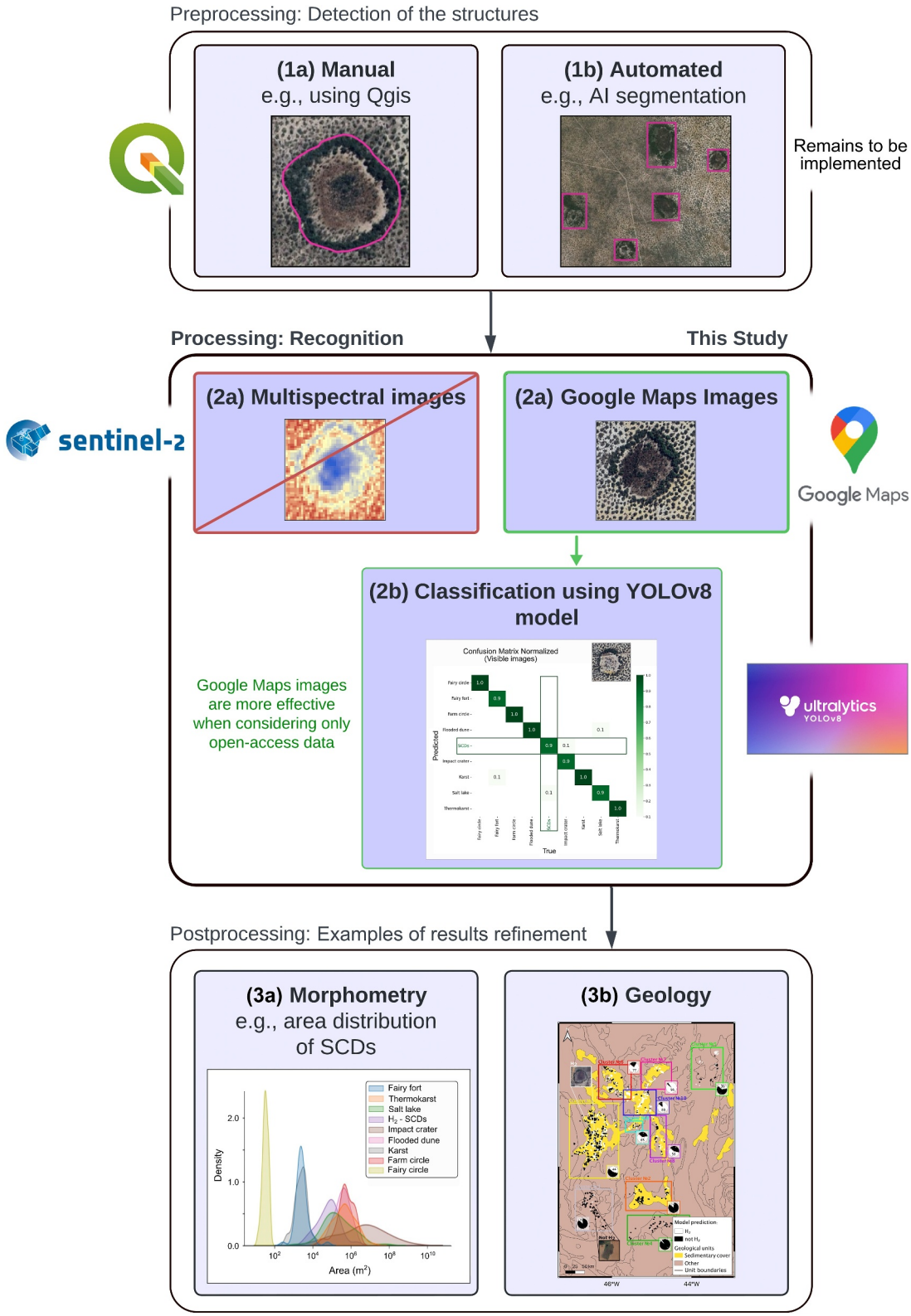

Figure 4. Workflow for improving the targeting of H_2-related rounded landforms. It consists of three key steps: (a) detecting rounded landforms, (b) classifying the identified rounded landforms, and (c) refining the results. The importance of incorporating morphometric parameters and statistical analyses in the post-processing step is detailed (Figure S3 and Text S4 in Supporting Information S1).

1. Mapping of rounded landforms in a chosen region. This step can be achieved by time-consuming manual work as in this study, or automatically through future developments;
2. Classification and recognition of SCDs. This step constituted the core of the present study. We propose an AI classification model using the open-access YOLOv8 algorithm. Considering only open-access data sets, we demonstrate that high-resolution visible images from Google Maps© are more effective than multispectral index data sets from Sentinel-2 satellite for model training;
3. Refinement of the model results. This step, which remains to be implemented in future work, may, for instance, involve morphometric information on SCDs (e.g., landform area, see Figure S3 and Text S4 in Supporting Information S1), spatial statistical analyses (e.g., cluster vs. isolated landforms), and regional geological information. As mentioned previously, one of the main challenges is differentiating between impact craters and SCDs. In this specific case, we can leverage studies that use a similar approach (i.e., based on satellite imagery) to analyze craters (e.g., Baratoux & Folco, 2024; Navee et al., 2024) and compare them with our data. For example, their sizes differ: craters often have diameters greater than 1 km (Navee et al., 2024), whereas SCDs typically have smaller diameters. Additionally, SCDs tend to be located in sedimentary basins, whereas craters are found across all geological contexts (both basement and sedimentary basins). Finally, analyzing the spatial distribution of SCDs, which often form clusters (Larin et al., 2015; Roche et al., 2024; Zgonnik et al., 2015) rather than appearing as isolated structures, could also help refine SCD detection. There are therefore various approaches that may differ depending on the study area. This post-processing is expected to provide additional insights to further filter the model results.

A significant room for improvement remains in the current workflow methodology, notably in terms of input data. While multispectral indices show theoretically high potential for SCDs classification (e.g., Moretti et al., 2022; Lévy, Roche, et al., 2023), their effectiveness in the present methodology is low due to the limited resolution of open-access data. Improved resolution would enhance SCD identification, reducing false positives and further improving workflow outcomes. Moreover, although the model performs optimally in trained areas, its generalization to new regions may be limited by local geological and climatic variability. To enhance its adaptability, an iterative approach could be considered, involving retraining the model with a small number of additional samples for each new study area. This strategy would progressively refine predictions, improve the recognition of SCDs, and ultimately significantly strengthen the model's robustness in diverse geological contexts. Finally, the AI model is based on training performed using extensive data sets. Currently, the number of known SCDs (i.e., validated by field H_2 measurements) is still limited, precluding the models to get trained optimally. This limitation should be addressed in the years to come with the growing number of H_2 exploration campaigns.

4. Conclusion

We used an open-source data set combined with machine learning models, specifically image classification, to recognize SCDs. The model was trained on various rounded landforms, including the combination of indices (BI and NDVI) and the Google Maps© images. The best performance is achieved with high-resolution Google Maps© images (90% true positive rate), suggesting the method's effectiveness in distinguishing potential H_2 seeps. The model was then applied to the data set São Francisco Basin in Brazil. We show that 48% of structures previously mapped are likely associated with H_2 seeps, based on available observations. However, as direct field measurements have not been performed, this relationship should be interpreted with caution. Further, the refinement of morphometric parameters combined to statistical analyses on spatial distribution, and information on regional geology can also be used to enhance the accuracy of SCD identification. It would also be interesting to study their evolution over time, as this would provide constraints on migration pathways and possible relationships with reservoirs. Overall, these results not only carry significant implications for H_2 exploration in Brazil, but also provide a rapid screening methodological framework that can be applied on a global scale, improving the efficiency of exploration and identification of new regions with high H_2 potential.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The whole dataset of the SCDs used for training is available in Zenodo platform (https://doi.org/10.5281/zenodo.15376870, Geymond, 2025). Sentinel-2 multispectral imagery was obtained from the European Space Agency's Copernicus Open Access Hub (https://scihub.copernicus.eu). Google Maps© images are accessible to anyone through the Google Maps© platform. The YOLOv8 model, an open- source deep learning framework for image classification and detection, was employed in this study. The model and its documentation are freely available at https://docs.ultralytics.com. The dataset from the São Francisco Basin, referenced in the case study, was obtained from De Freitas et al. (2024) and can be accessed through their publication or by contacting the authors.

Acknowledgments
Funding was provided by PULSAR (Pays de La Loire). We would like to express our sincere gratitude to I. Martinez for supporting this project at IPGP at the beginning, and to the authors of the dataset used in the Brazilian case study, particularly V. Azor de Freitas and I. Moretti, for providing valuable data from the São Francisco Basin. We warmly thank the three anonymous reviewers and the editor D. Baratou for their constructive feedback and valuable support throughout the review process.

References

Aimar, L., Frery, E., Strand, J., Heath, C., Khan, S., Moretti, I., & Ong, C. (2023). Natural hydrogen seeps or salt lakes: How to make a difference? Grass patch example, Western Australia. *Frontiers of Earth Science*, *11*, 1236673. https://doi.org/10.3389/feart.2023.1236673

Baratoux, D., & Folco, L. (2024). Impact structures and meteorites in North Africa. *The Geology of North Africa*, 591–630. https://doi.org/10.1007/978-3-031-48299-1_20

Barrett, A. M., Balme, M. R., Woods, M., Karachalios, S., Petrocelli, D., Joudrier, L., & Sefton-Nash, E. (2022). NOAH-H, a deep-learning, terrain classification system for Mars: Results for the ExoMars Rover candidate landing sites. *Icarus*, *371*, 114701. https://doi.org/10.1016/j.icarus.2021.114701

Bisong, E. (2019). *Building machine learning and deep learning models on Google cloud platform: A comprehensive guide for beginners*. Apress. https://doi.org/10.1007/978-1-4842-4470-8

Carrillo Ramirez, A., Gonzalez Penagos, F., Rodriguez, G., & Moretti, I. (2023). Natural H2 emissions in Colombian ophiolites: First findings. *Geosciences*, *13*(12), 358. https://doi.org/10.3390/geosciences13120358

Davies, K., Frery, E., Giwelli, A., Esteban, L., Keshavarz, A., & Iglauer, S. (2024). A natural hydrogen seep in Western Australia: Observed characteristics and controls. *Science and Technology for Energy Transition*, *79*, 48. https://doi.org/10.2516/stet/2024043

Daynac, J., Bessin, P., Pochat, S., Mourgues, R., & Shumack, S. (2024). A new workflow for mapping dune features (outline, Crestline and defects) combining deep learning and skeletonization from DEM-derived data. *Geomorphology*, *463*, 109369. https://doi.org/10.1016/j.geomorph.2024.109369

De Freitas, V. A., Prinzhofer, A., Françolin, J. B., Ferreira, F. J. F., & Moretti, I. (2024). Natural hydrogen system evaluation in the São Francisco Basin (Brazil). *Science and Technology for Energy Transition*, *79*, 95. https://doi.org/10.2516/stet/2024091

Donzé, F. V., Bourdet, L., Truche, L., Dusséaux, C., & Huyghe, P. (2024). Modeling deep control pulsing flux of native H2 throughout tectonic fault-valve systems. *International Journal of Hydrogen Energy*, *58*, 1443–1456. https://doi.org/10.1016/j.ijhydene.2024.01.178

Doust, H. (2010). The exploration play: What do we mean by it? *Bulletin*, *94*(11), 1657–1672. https://doi.org/10.1306/06301009168

Etiope, G., Ciotoli, G., Benà, E., Mazzoli, C., Röckmann, T., Sivan, M., et al. (2024). Surprising concentrations of hydrogen and non-geological methane and carbon dioxide in the soil. *Science of the Total Environment*, *948*, 174890. https://doi.org/10.1016/j.scitotenv.2024.174890

Frery, E., Langhi, L., Maison, M., & Moretti, I. (2021). Natural hydrogen seeps identified in the north Perth basin, Western Australia. *International Journal of Hydrogen Energy*, *46*(61), 31158–31173. https://doi.org/10.1016/j.ijhydene.2021.07.023

Gaucher, E. C., Moretti, I., Pélissier, N., Burridge, G., & Gonthier, N. (2023). The place of natural hydrogen in the energy transition: A position paper. https://doi.org/10.5281/ZENODO.8108239

Getzin, S., Yizhaq, H., & Tschinkel, W. R. (2021). Definition of "fairy circles" and how they differ from other common vegetation gaps and plant rings. *Journal of Vegetation Science*, *32*(6), e13092. https://doi.org/10.1111/jvs.13092

Geymond, U. (2025). Identification of natural hydrogen seeps: Leveraging AI for automated classification of sub-circular depressions [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.15376871

Hovland, M., & Judd, A. G. (1988). *Seabed pockmarks and seepages: Impact on geology, biology and the marine environment*. Graham&Trotman.

IEA. (2024). Global hydrogen review 2024. Retrieved from https://www.iea.org/reports/global-hydrogen-review-2024

Lapi, T., Chatzimpiros, P., Raineau, L., & Prinzhofer, A. (2022). System approach to natural versus manufactured hydrogen: An interdisciplinary perspective on a new primary energy source. *International Journal of Hydrogen Energy*, *47*(51), 21701–21712. https://doi.org/10.1016/j.ijhydene.2022.05.039

Larin, N., Zgonnik, V., Rodina, S., Deville, E., Prinzhofer, A., & Larin, V. N. (2015). Natural molecular hydrogen seepage associated with surficial, rounded depressions on the European Craton in Russia. *Natural Resources Research*, *24*(3), 369–383. https://doi.org/10.1007/s11053-014-9257-5

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

Lefeuvre, N., Truche, L., Donzé, F., Ducoux, M., Barré, G., Fakoury, R., et al. (2021). Native H_2 exploration in the Western Pyrenean Foothills. *Geochemistry, Geophysics, Geosystems*, *22*(8). https://doi.org/10.1029/2021GC009917

Lefeuvre, N., Truche, L., Donzé, F.-V., Gal, F., Tremosa, J., Fakoury, R.-A., et al. (2022). Natural hydrogen migration along thrust faults in foothill basins: The North Pyrenean Frontal Thrust case study. *Applied Geochemistry*, *145*, 105396. https://doi.org/10.1016/j.apgeochem.2022.105396

Leila, M., Loiseau, K., & Moretti, I. (2022). Controls on generation and accumulation of blended gases (CH4/H2/He) in the Neoproterozoic Amadeus Basin, Australia. *Marine and Petroleum Geology*, *140*, 105643. https://doi.org/10.1016/j.marpetgeo.2022.105643

Lévy, D., Boka-Mene, M., Meshi, A., Fejza, I., Guermont, T., Hauville, B., & Pelissier, N. (2023). Looking for natural hydrogen in Albania and Kosova. *Frontiers of Earth Science*, *11*, 1167634. https://doi.org/10.3389/feart.2023.1167634

Lévy, D., Roche, V., Pasquet, G., Combaudon, V., Geymond, U., Loiseau, K., & Moretti, I. (2023). Natural H_2 exploration: Tools and workflows to characterize a play. *Science and Technology for Energy Transition*, *78*, 27. https://doi.org/10.2516/stet/2023021

Lodhia, B. H., Peeters, L., & Frery, E. (2024). A review of the migration of hydrogen from the planetary to basin scale. *Journal of Geophysical Research: Solid Earth*, *129*(6), e2024JB028715. https://doi.org/10.1029/2024JB028715

Magoon, L. B. (1995). The play that complements the petroleum system—A new exploration equation. *Oil & Gas Journal*, *73*.

Maiga, O., Deville, E., Laval, J., Prinzhofer, A., & Diallo, A. B. (2023a). Characterization of the spontaneously recharging natural hydrogen reservoirs of Bourakebougou in Mali. *Scientific Reports*, *13*(1), 11876. https://doi.org/10.1038/s41598-023-38977-y

Maiga, O., Deville, E., Laval, J., Prinzhofer, A., & Diallo, A. B. (2023b). Trapping processes of large volumes of natural hydrogen in the subsurface: The emblematic case of the Bourakebougou H2 field in Mali. *International Journal of Hydrogen Energy*, *S036031992305214X*, 640–647. https://doi.org/10.1016/j.ijhydene.2023.10.131

Malvoisin, B., & Brunet, F. (2023). Barren ground depressions, natural H2 and orogenic gold deposits: Spatial link and geochemical model. *Science of the Total Environment*, *856*, 158969. https://doi.org/10.1016/j.scitotenv.2022.158969

Marochov, M., Stokes, C. R., & Carbonneau, P. E. (2021). Image classification of marine-terminating outlet glaciers in Greenland using deep learning methods. *The Cryosphere*, *15*(11), 5041–5059. https://doi.org/10.5194/tc-15-5041-2021

Moretti, I., Brouilly, E., Loiseau, K., Prinzhofer, A., & Deville, E. (2021). Hydrogen emanations in intracratonic areas: New guide lines for early exploration basin screening. *Geosciences*, *11*(3), 145. https://doi.org/10.3390/geosciences11030145

Moretti, I., Geymond, U., Pasquet, G., Aimar, L., & Rabaute, A. (2022). Natural hydrogen emanations in Namibia: Field acquisition and vegetation indexes from multispectral satellite image analysis. *International Journal of Hydrogen Energy*, *S0360319922036631*(84), 35588–35607. https://doi.org/10.1016/j.ijhydene.2022.08.135

Moretti, I., Prinzhofer, A., Françolin, J., Pacheco, C., Rosanne, M., Rupin, F., & Mertens, J. (2021). Long-term monitoring of natural hydrogen superficial emissions in a Brazilian cratonic environment. Sporadic large pulses versus daily periodic emissions. *International Journal of Hydrogen Energy*, *46*(5), 3615–3628. https://doi.org/10.1016/j.ijhydene.2020.11.026

Navee, E. O. M., Baratoux, D., Aoudjehane, H. C., Mhamdi, H. S., & Raji, M. (2024). Systematic search of circular structures using satellite imagery to identify potential new impact structures in Mauritania. *Journal of African Earth Sciences*, *216*, 105303. https://doi.org/10.1016/j.jafrearsci.2024.105303

Pinet, N., Duchesne, M., Lavoie, D., Bolduc, A., & Long, B. (2008). Surface and subsurface signatures of gas seepage in the St. Lawrence Estuary (Canada): Significance to hydrocarbon exploration. *Marine and Petroleum Geology*, *25*(3), 271–288. https://doi.org/10.1016/j.marpetgeo.2007.07.011

Prinzhofer, A., & Cacas-Stentz, M.-C. (2023). Natural hydrogen and blend gas: A dynamic model of accumulation. *International Journal of Hydrogen Energy*, *48*(57), 21610–21623. https://doi.org/10.1016/j.ijhydene.2023.03.060

Prinzhofer, A., Cissé, C. S. T., & Diallo, A. B. (2018). Discovery of a large accumulation of natural hydrogen in Bourakebougou (Mali). *International Journal of Hydrogen Energy*, *43*(42), 19315–19326. https://doi.org/10.1016/j.ijhydene.2018.08.193

Prinzhofer, A., Moretti, I., Françolin, J., Pacheco, C., D'Agostino, A., Werly, J., & Rupin, F. (2019). Natural hydrogen continuous emission from sedimentary basins: The example of a Brazilian H2-emitting structure. *International Journal of Hydrogen Energy*, *44*(12), 5676–5685. https://doi.org/10.1016/j.ijhydene.2019.01.119

Roche, V., Geymond, U., Boka-Mene, M., Delcourt, N., Portier, E., Revillon, S., & Moretti, I. (2024). A new continental hydrogen play in Damara Belt (Namibia). *Scientific Reports*, *14*(1), 11655. https://doi.org/10.1038/s41598-024-62538-6

Strauch, B., Pilz, P., Hierold, J., & Zimmer, M. (2023). Experimental simulations of hydrogen migration through potential storage rocks. *International Journal of Hydrogen Energy*, *48*(66), 25808–25820. https://doi.org/10.1016/j.ijhydene.2023.03.115

Truche, L., Donzé, F.-V., Goskolli, E., Muceku, B., Loisy, C., Monnin, C., et al. (2024). A deep reservoir for hydrogen drives intense degassing in the Bulqizë ophiolite. *Science*, *383*, 618–621. https://doi.org/10.1126/science.adk9099

Worman, S. L., Pratson, L. F., Karson, J. A., & Schlesinger, W. H. (2020). Abiotic hydrogen (H_2) sources and sinks near the Mid-Ocean Ridge (MOR) with implications for the subseafloor biosphere. *Proceedings of the national academy of sciences of the United States of America* (Vol. 117(24), 13283–13293). https://doi.org/10.1073/pnas.2002619117

Zgonnik, V. (2020). The occurrence and geoscience of natural hydrogen: A comprehensive review. *Earth-Science Reviews*, *203*, 103140. https://doi.org/10.1016/j.earscirev.2020.103140

Zgonnik, V., Beaumont, V., Deville, E., Larin, N., Pillot, D., & Farrell, K. M. (2015). Evidence for natural molecular hydrogen seepage associated with Carolina bays (surficial, ovoid depressions on the Atlantic Coastal Plain, Province of the USA). *Progress in Earth and Planetary Science*, *2*(1), 31. https://doi.org/10.1186/s40645-015-0062-5

References From the Supporting Information

Dogan, H. M. (2009). Mineral composite assessment of Kelkit River Basin in Turkey by means of remote sensing. *Journal of Earth System Science*, *118*(6), 701–710. https://doi.org/10.1007/s12040-009-0059-9

Huete, A., Justice, C., & Liu, H. (1994). Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment*, *49*(3), 224–234. https://doi.org/10.1016/0034-4257(94)90018-3

Ji, L., Zhang, L., Wylie, B. K., & Rover, J. (2011). On the terminology of the spectral vegetation index (NIR − SWIR)/(NIR + SWIR). *International Journal of Remote Sensing*, *32*(21), 6901–6909. https://doi.org/10.1080/01431161.2010.510811

Khan, N. M., Rastoskuev, V. V., Sato, Y., & Shiozawa, S. (2005). Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agricultural Water Management*, *77*(1–3), 96–109. https://doi.org/10.1016/j.agwat.2004.09.038

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

Peñuelas, J., & Inoue, Y. (1999). Reflectance indices indicative of changes in water and pigment contents of peanut and wheat leaves. *Photosynt*, *36*(3), 355–360. https://doi.org/10.1023/A:1007033503276

Segal, D. (1982). Theoretical basis for differentiation of ferric-iron bearing minerals, using landsat mss data. *Proceedings of symposium for remote sensing of environment, 2nd thematic conference on remote sensing for exploratory geology* (Vol. 949, p. 951).

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0