EARTHSCIENCES

Using machine learning to predict the source and abundance of geological hydrogen

Primary supervisor:	Claire Nichols https://www.earth.ox.ac.uk/people/claire-nichols
Co supervisor(s):	Mark van der Wilk https://www.cs.ox.ac.uk/people/mark.vanderwilk/ Samira Barzin https://www.eci.ox.ac.uk/person/dr-samira-barzin
Key words:	Geological hydrogen, machine learning, Net Zero, sustainability, energy security
Research theme(s):	• Earth Resources
Eligible courses for this project:	• MSc by Research in Earth Sciences (2-3 years) • DPhil in Earth Sciences (3-4 years) • Interdisciplinary Life and Environmental Science Landscape Award (ILESLA) • Intelligent Earth (UKRI CDT)

Overview

Today the value of the annual global hydrogen market is estimated to be ca $225 billion with estimates up to $1.4 trillion by 2050 if hydrogen, in addition to current use, replaces natural gas as an energy vector. Hydrogen plays a critical role in processes such as ammonia and methanol production (50%), petroleum refinement (45%) and steel manufacturing (5%). Almost all hydrogen to meet these demands is produced from fossil fuels, contributing around 2.5% of global carbon dioxide emissions. It has been projected that clean hydrogen could contribute 11% of the cumulative emission reductions required to stay within 2°C of global warming. Achieving Net Zero by 2050, which is a primary aim of the United Nations Paris Agreement, requires radical changes in energy usage. Clean hydrogen is now being considered as a Net Zero solution in numerous energy sectors including for power stations, residential heating and power, personal vehicles and long-haul shipping and air travel.

Numerous sources of clean hydrogen have been explored in recent years. Hydrogen generated from fossil fuels can be coupled to carbon sequestration and storage (CCS) in order to significantly reduce its carbon footprint, although some calculations suggest this could result in even higher emissions due to inefficient storage (Howarth & Jacobson, 2021) . Since 2018, green hydrogen, which is generated via water electrolysis from renewable electricity gained significant attention and investment as a solution to global hydrogen demand, but operational costs rapidly became far too high to be feasible (see figure). To meet Net Zero demands in 2050 would require 63% of the world's electricity just for

hydrogen production (Osselin et al., 2022) , double the global amount of electricity currently produced by renewables. By 2024, numerous green hydrogen initiatives had crumbled.

In this project we will explore the feasibility of geological hydrogen as a long-term, sustainable energy source. There are many outstanding questions regarding where geological hydrogen is formed, how quickly it is generated, and whether most seeps are from trapped or actively-generated sources. We will use a combination of global geophysical datasets and geospatial data to build a model that can assess where geological hydrogen is mostly likely to be produced at sufficiently high rate to be sustainable, and where it is accessible for use by society.

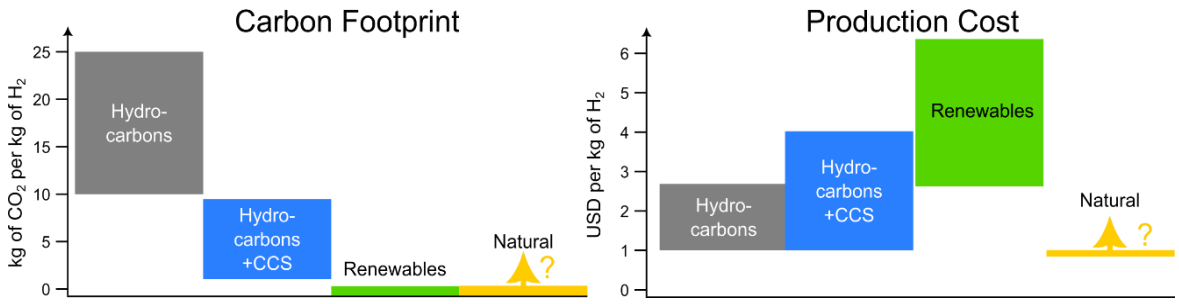


Figure showing the production costs and carbon footprints for different sources of hydrogen: grey = steam methane reformation of coal/gas without carbon capture and storage (CCS); blue = steam methane reformation of coal/gas with CCS; green = electrolysis using renewable electricity; yellow = natural/stimulated natural. Adapted from Ballentine et al., (2025)

Methodology

We will train a model to detect patterns linking hydrogen occurrences to lithology, topography, crustal thickness, lithospheric thickness, gravity, aeromagnetics and heat flow and assess the systemic drivers of hydrogen demand by integrating multiple global datasets, with a particular emphasis on remote sensing products such as the AlphaEarth Foundations model.

The model will be ground-truthed via fieldwork to specific locations to verify the results and refine model outputs.

Timeline

Year 1: Training in how to deal with geospatial data and how to approach machine learning on spatially/temporally variable datasets. Literature review on the geological controls on hydrogen generation.

Years 2 and 3: First model results and fieldwork to verify them. Refinements to model and submit publication on key findings.

Year 4: Presentation at relevant international conference, finish thesis.

Training & Skills

- Making and presenting a scientific talk/poster
- Presenting to policy makers/industry
- Writing a scientific paper
- Good practices in data management
- Computational skills (e.g., Python, GitHub)

- Geospatial data analysis
- Machine learning
- Teaching/mentoring undergraduates and Masters students

References & Further Reading

Ballentine, C. J., Karolytė, R., Cheng, A., Sherwood Lollar, B., Gluyas, J. G., & Daly, M. C. (2025). Natural hydrogen resource accumulation in the continental crust. In *Nature Reviews Earth and Environment* (Vol. 6, Number 5, pp. 342–356). Springer Nature. https://doi.org/10.1038/s43017-025-00670-1

Howarth, R. W., & Jacobson, M. Z. (2021). How green is blue hydrogen? *Energy Science and Engineering*, *9*(10), 1676–1687. https://doi.org/10.1002/ese3.956

Osselin, F., Soulaine, C., Fauguerolles, C., Gaucher, E. C., Scaillet, B., & Pichavant, M. (2022). Orange hydrogen is the new green. *Nature Geoscience*, *15*(10), 765–769. https://doi.org/10.1038/s41561-022-01043-9

Further Information

Contact: Claire Nichols (Claire.nichols@earth.ox.ac.uk)