

CONTENTS PAGE

Sponsors Acknowledgement	Page 2
Conference Programme	Page 3
Poster Programme	Page 15
Oral Abstracts (in programme order)	Page 16
Software Demonstration Abstracts (in programme order)	Page 111
Poster Abstracts (in alphabetical order)	Page 122
Burlington House Fire Safety Information	Page 139
Ground Floor plan of The Geological Society	Page 140

We gratefully acknowledge the support of the sponsors for making this meeting possible.



CONFERENCE PROGRAMME

The following meeting rooms are available for ad-hoc demos and virtual fieldtrips on the following days:

DAY 1

Council Room & Arthur Holmes: 09.30 – 13.50/15.40 – 17.30

William Buckland: 09.30 – 12.50

DAY 2



Council Room & Arthur Holmes: 09.30 – 12.50/15.10 – 17.30


William Buckland: 09.30 – 12.30



DAY 3

Council Room & Arthur Holmes: 09.30 – 10.20/15.35 – 17.00

William Buckland: 09.30 – 10.20/14.20 – 17.00

Tuesday 27 February 2018	
08.30	Registration & tea, coffee & refreshments Sponsored by 
09.00	Welcome
09.10	KEYNOTE: Pixels in the cloud Ed Parsons (Google)
09.40	GeoSocial: Exploring the usefulness of social media mining in the applied natural geohazard sciences Emma Bee (British Geological Survey)
10.00	Big Data in the industry: A critical examination of modern data collection and use in engineering geosciences Oliver Dabson (CH2M)
10.20	Science and the digital revolution: data, standards and interdisciplinary integration Geoffrey Boulton (University of Edinburgh & CODATA)
10.40	Getting Legacy BGS Stratigraphic Data Ready for the Big Data Revolution Mike Howe (British Geological Survey)
11.00	Tea, coffee, refreshments and posters Sponsored by 
11.20	KEYNOTE: Data science for earth science – perspectives from industry Steve Garrett (Chevron)
11.50	Disparate E&P Big Data Impose Non-Desperate Machine Learning Methodologies Keith Holdaway (SAS Global O&G Domain)
12.10	Big data in the Geoscience: A portal to physical properties Andrew Kingdon (British Geological Survey)
12.30	Building Data Science Capability Ed Evans (NDB Upstream)
12.50	Lunch and posters
13.50	KEYNOTE: The Virtual Geoscience Revolution John Howell (University of Aberdeen)

14.20	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Zagros and Zechestein Richard Jones (University of Durham)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Seismic VR (Council Room) TBC (Halliburton)</p> <p>Osokey Stream - cloud-native seismic platform (Arthur Holmes Room) James Selvage (Osokey)</p>
14.35	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Zagros and Zechestein Richard Jones (University of Durham)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Seismic VR (Council Room) TBC (Halliburton)</p> <p>Osokey Stream - cloud-native seismic platform (Arthur Holmes Room) James Selvage (Osokey)</p>
14.50	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>A Geological tour of the Jurassic Coast, NE England Dave Hodgetts (University of Manchester)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Seismic VR (Council Room) TBC (Halliburton)</p> <p>Osokey Stream - cloud-native seismic platform (Arthur Holmes Room) James Selvage (Osokey)</p>
15.05	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>A Geological tour of the Jurassic Coast, NE England Dave Hodgetts (University of Manchester)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Seismic VR (Council Room) TBC (Halliburton)</p> <p>Osokey Stream - cloud-native seismic platform (Arthur Holmes Room) James Selvage (Osokey)</p>
15.20	<p>Tea, coffee, refreshments and posters</p> <p style="text-align: right;">Sponsored by </p>

15.40	KEYNOTE: Paradigm Shift Necessary for True Integrated Asset Value Maximization by Leveraging Big Data and Digital Transformation Satyam Priyadarshy (Halliburton)
16.10	Mining Geological Sentiment from Unstructured Text Paul Cleverley (Robert Gordon University)
16.30	Putting data at your fingertips: Utilising data analytics in E&P Information Management Christopher Frost (DataCo Global Limited)
16.50	Accessing Knowledge in Geoscience Text using Natural Language Processing Richard Jones (University of Durham)
17.10	Detecting Volcano Deformation in InSAR using Deep learning Pui Anantrasirichai (University of Bristol)
17.30 – 18.30	Wine reception
Wednesday 28 February 2018	
08.30	Registration and tea, coffee & refreshments Sponsored by 
09.00	Welcome
09.10	KEYNOTE: The Role of Data Regulating the UK Oil & Gas Industry Nick Richardson (OGA)
09.40	Over 120 years worth of hydrocarbon exploration, an example of how legacy data can address todays challenges. Mark Fellgett (British Geological Survey)
10.00	Digital Transformation in the North Sea Stephen Ashley (The Oil and Gas Technology Centre)
10.20	Two basic applications of machine learning for rock-physics and petro-physics Ehsan Naeini (IKON)
10.40	The Chronological Data Explosion: the Huge Potential for Source-to-Sink Insights within Exploration Graeme Nicoll (Halliburton)
11.00	Tea, coffee, refreshments and posters Sponsored by 
11.20	KEYNOTE: Digitalization and Data in Field Development Liz Wild (Shell)
11.50	Big data? The power is in the analysis Oliver Jordan (Statoil)
12.10	Virtual Glaciers and Glaciated Landscapes Derek McDougall (University of Worcester)
12.30	Improving access to UKCS Petrotechnical Data – the next step in a 20 year journey Daniel Brown (Common Data Access Limited)
12.50	Lunch and Posters

13.50	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Mars Robert Barnes (Imperial College London)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Facilitating the effective application of analogue databases to reservoir models (Council Room) Marijn Van Cappelle (Petrotechnical Data Systems)</p> <p>Seismic VR (Arthur Holmes Room) TBC (Halliburton)</p>
14.05	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Mars Robert Barnes (Imperial College London)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Facilitating the effective application of analogue databases to reservoir models (Council Room) Marijn Van Cappelle (Petrotechnical Data Systems)</p> <p>Seismic VR (Arthur Holmes Room) TBC (Halliburton)</p>
14.20	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Thompson Canyon, a multiscale fieltrip John Howell (University of Aberdeen)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Facilitating the effective application of analogue databases to reservoir models (Council Room) Marijn Van Cappelle (Petrotechnical Data Systems)</p> <p>Seismic VR (Arthur Holmes Room) TBC (Halliburton)</p>

14.35	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Thompson Canyon, a multiscale fieldtrip John Howell (University of Aberdeen)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Facilitating the effective application of analogue databases to reservoir models (Council Room) Marijn Van Cappelle (Petrotechnical Data Systems)</p> <p>Seismic VR (Arthur Holmes Room) TBC (Halliburton)</p>
14.50	<p>Tea, coffee, refreshments and posters Sponsored by</p> 
15.10	<p>KEYNOTE: Big Data and the British Geological Survey Garry Baker (British Geological Survey)</p>
15.40	<p>IGCP 648 Efforts to Compile Structured Data for Palaeogeographic Studies: some lessons learned Bruce Eglinton (University of Saskatchewan)</p>
16.00	<p>Palaeogeography and Big Data Paul Markwick (Knowing Earth Limited, University of Leeds & University of Bristol)</p>
16.20	<p>Data Mining and Visualization of Detrital Zircon Data: Assessment of Palaeogeographic and Geodynamic Setting Using Data from Laurentia Dean Meek (University of Saskatchewan)</p>
16.40	<p>The worldwide field course: use of 3D outcrop imagery in training Gary Nichols (RPS)</p>
17.00	<p>Panel Discussion</p>
18.00	<p>Close</p>
Thursday 1 March 2018	
08.30	<p>Registration and tea & coffee Sponsored by</p> 
09.00	<p>Welcome</p>
09.10	<p>KEYNOTE: When Failure (a lot of failure) Becomes an Option - Machine and Deep Learning on Seismic Data John Thurmond (Statoil)</p>
09.40	<p>Big data - A boundaryless future? Rhian Burrell (Osokey)</p>
10.00	<p>How machine learning systems can extract more qualified information from seismic acquisition and processing reports. Henri Blondelle (AgileDD)</p>
10.00-15.00	<p>Python and AI basics coding workshop for Geoscientists Ikon Science This workshop will be taking place at Ikon Science's offices at 1 The Crescent, Surbiton, London, KT64BN. Pre-registration is required.</p>
10.20	<p>Tea, coffee, refreshments and posters Sponsored by</p> 

10.40	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Seismic VR TBC (Halliburton)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>GBDB, an expanding open data system for geological information (Council Room) Junxuan Fan (Chinese Academy of Science)</p> <p>Facilitating the effective application of analogue databases to reservoir models (Arthur Holmes Room) Marijn Van Cappelle (Petrotechnical Data Systems)</p> <p>Geoscience data landscapes through integration with new cloud platforms (William Buckland Room) Interica</p>
10.55	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Seismic VR TBC (Halliburton)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>GBDB, an expanding open data system for geological information (Council Room) Junxuan Fan (Chinese Academy of Science)</p> <p>Facilitating the effective application of analogue databases to reservoir models (Arthur Holmes Room) Marijn Van Cappelle (Petrotechnical Data Systems)</p> <p>Geoscience data landscapes through integration with new cloud platforms (William Buckland Room) Interica</p>


11.10	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Structural Geology fieldtrip, Pembrokeshire Adam Cawood (University of Aberdeen)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>GBDB, an expanding open data system for geological information (Council Room) Junxuan Fan (Chinese Academy of Science)</p> <p>Facilitating the effective application of analogue databases to reservoir models (Arthur Holmes Room) Marijn Van Cappelle (Petrotechnical Data Systems)</p> <p>Geoscience data landscapes through integration with new cloud platforms (William Buckland Room) Interica</p>
11.25	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Structural Geology fieldtrip, Pembrokeshire Adam Cawood (University of Aberdeen)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>GBDB, an expanding open data system for geological information (Council Room) Junxuan Fan (Chinese Academy of Science)</p> <p>Facilitating the effective application of analogue databases to reservoir models (Arthur Holmes Room) Marijn Van Cappelle (Petrotechnical Data Systems)</p> <p>Geoscience data landscapes through integration with new cloud platforms (William Buckland Room) Interica</p>

11.40	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Virtual Volcanoes, Stromboli Dougal Jerram</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>QEMSCAN Big Data petrography: a tutorial (Council Room) Jenny Omma (Rocktype Ltd)</p> <p>Geoscience data landscapes through integration with new cloud platforms (Arthur Holmes Room) Interica</p> <p>GBDB, an expanding open data system for geological information (William Buckland Room) Junxuan Fan (Chinese Academy of Science)</p>
11.55	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Virtual Volcanoes, Stromboli Dougal Jerram</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>QEMSCAN Big Data petrography: a tutorial (Council Room) Jenny Omma (Rocktype Ltd)</p> <p>Geoscience data landscapes through integration with new cloud platforms (Arthur Holmes Room) Interica</p> <p>GBDB, an expanding open data system for geological information (William Buckland Room) Junxuan Fan (Chinese Academy of Science)</p>

12.10	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Geoscape and the Neuquen Basin Olivier Galland (University of Oslo)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>QEMSCAN Big Data petrography: a tutorial (Council Room) Jenny Omma (Rocktype Ltd)</p> <p>Geoscience data landscapes through integration with new cloud platforms (Arthur Holmes Room) Interica</p> <p>GBDB, an expanding open data system for geological information (William Buckland Room) Junxuan Fan (Chinese Academy of Science)</p>
12.25	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Geoscape and the Neuquen Basin Olivier Galland (University of Oslo)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>QEMSCAN Big Data petrography: a tutorial (Council Room) Jenny Omma (Rocktype Ltd)</p> <p>Geoscience data landscapes through integration with new cloud platforms (Arthur Holmes Room) Interica</p> <p>GBDB, an expanding open data system for geological information (William Buckland Room) Junxuan Fan (Chinese Academy of Science)</p>
12.40	<p>Lunch and Posters</p>
13.20	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Virtual Fieldtrip to seismic scale outcrops of the Triassic, Edgeøya, Svalbard Simon Buckley (Uni Research, Bergen)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Osokey Stream - cloud-native seismic platform (Council Room) James Selvage (Osokey)</p> <p>QEMSCAN Big Data petrography: a tutorial (Arthur Holmes Room) Jenny Omma (Rocktype Ltd)</p> <p>SAFARI (William Buckland Room)</p>

13.35	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Virtual Fieldtrip to seismic scale outcrops of the Triassic, Edgeøya, Svalbard Simon Buckley (Uni Research, Bergen)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Osokey Stream - cloud-native seismic platform (Council Room) James Selvage (Osokey)</p> <p>QEMSCAN Big Data petrography: a tutorial (Arthur Holmes Room) Jenny Omma (Rocktype Ltd)</p> <p>SAFARI (William Buckland Room)</p>
13.50	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Deltaic and shallow marine deposits on Eastern Utah James Mullins (University of Aberdeen)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Osokey Stream - cloud-native seismic platform (Council Room) James Selvage (Osokey)</p> <p>QEMSCAN Big Data petrography: a tutorial (Arthur Holmes Room) Jenny Omma (Rocktype Ltd)</p> <p>SAFARI (William Buckland Room)</p>
14.05	<p><u>VIRTUAL FIELDTRIP (Lecture Theatre)</u></p> <p>Deltaic and shallow marine deposits on Eastern Utah James Mullins (University of Aberdeen)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Osokey Stream - cloud-native seismic platform (Council Room) James Selvage (Osokey)</p> <p>QEMSCAN Big Data petrography: a tutorial (Arthur Holmes Room) Jenny Omma (Rocktype Ltd)</p> <p>SAFARI (William Buckland Room)</p>

14.20	<p><u>VIRTUAL FIELDTRIP</u> (Lecture Theatre)</p> <p>The use of Virtual Outcrop Models, digital geology and legacy data to reappraise Devonian basin evolution in NE Scotland and Shetland. Thomas Utley (Durham University)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Geoscience data landscapes through integration with new cloud platforms (Council Room) Interica</p> <p>SAFARI (Arthur Holmes Room)</p>
14.35	<p><u>VIRTUAL FIELDTRIP</u> (Lecture Theatre)</p> <p>The use of Virtual Outcrop Models, digital geology and legacy data to reappraise Devonian basin evolution in NE Scotland and Shetland. Thomas Utley (Durham University)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Geoscience data landscapes through integration with new cloud platforms (Council Room) Interica</p> <p>SAFARI (Arthur Holmes Room)</p>
14.50	<p><u>VIRTUAL FIELDTRIP</u> (Lecture Theatre)</p> <p>Facilitating the effective application of analogue databases to reservoir models Marijn Van Cappelle (Petrotechnical Data Systems)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Geoscience data landscapes through integration with new cloud platforms (Council Room) Interica</p> <p>SAFARI (Arthur Holmes Room)</p>
15.05	<p><u>VIRTUAL FIELDTRIP</u> (Lecture Theatre)</p> <p>Facilitating the effective application of analogue databases to reservoir models Marijn Van Cappelle (Petrotechnical Data Systems)</p> <p><u>SOFTWARE DEMONSTRATIONS</u></p> <p>Geoscience data landscapes through integration with new cloud platforms (Council Room) Interica</p> <p>SAFARI (Arthur Holmes Room)</p>

15.20	Tea, coffee, refreshments and posters Sponsored by 
15.35	KEYNOTE: Machine Learning Assisted Petroleum Geoscience: Can a computer learn to map stratigraphic architecture and reservoir quality by training on data? Eirik Larsen/Chris Jackson (Earth Science Analytics)
15.55	Generalized Classification of Lithology from Wireline Logs Using Machine Learning as Applied to the Permian Basin, USA and North West Shelf, Australia Jose Montero (Halliburton)
16.15	Making the case for Big Data petrography Jenny Omma (Rocktype Ltd)
16.35	Ensemble Learning Approach to Lithofacies Classification Using Well Logs Didi Sher Ooi (University of Bristol)
16.55	Discussion
17.15	Close

POSTER PROGRAMME

Day 1: 27 February 2018
Machine Learning for Mineral Prospectivity Mapping Lucille Ablett
Improved understanding of borehole instability mechanisms through development of an enhanced visualization-numerical modelling approach Abraham Audu (University of Exeter)
The application of imaging IR spectroscopy for mineralogical analysis of core and cuttings. Gavin Hunt (Spectra-Map Ltd)
Improved production decline analysis through the use of Machine Learning techniques Elliot Humphrey (University of Aberdeen & Toby Project Services)
Data Integration: Understanding the Importance of Data and Knowledge Management Helen McKenzie (Sellafield Ltd)
Day 2: 28 February 2018
Virtual Glaciers and Glaciated Landscapes Derek McDougall (University of Worcester)
Understanding and enabling data standards for large environmental data sets Francesca Laws (Sellafield Ltd)
Deciphering the sands of time: a zircon U–Pb age database for the Circum-Arctic region Michael Pointon (CASP)
Development of 3D geological models for the UK low level waste repository John Shevelan (LLW Repository Ltd)
Day 3: 1 March 2018
PCA as an error diagnostic at long-term resistivity monitoring sites Luke Sibbett (British Geological Survey & University of Nottingham)
A Statistical Model for Addressing Uncertainty in the Assessment of Performance of Ageing Retaining Structures to Groundwater Inundation Victoria Stephenson (University College London and The Alan Turing Institute) and Chris Oates (The Alan Turing Institute)
Facilitating the effective application of analogue databases to reservoir models Marijn Van Cappelle (Petrotechnical Data Systems)
A New Approach to Communicating the Science of Natural Hazard Processes in the Terrestrial and Marine Environment using Virtual Reality. An Erasmus Key Action 2 Funded Project. Malcolm Whitworth (University of Portsmouth)

**ORAL ABSTRACTS
(in programme order)**

Pixels in the cloud

Ed Parsons

Google

Google's mission is to "organise the world's information and make it universally accessible and useful", while much of this is achieved by building tools to meet the needs of the "mass market" Google Earth Engine manages geospatial information and makes it available for analysis for the geoscience community.

Earth Engine and other cloud based solutions offer a platform for petabyte-scale scientific analysis and visualization of geospatial datasets, both for public benefit and for business and government users. Earth Engine stores satellite imagery, organizes it, and makes it available for the first time for global-scale data mining. The public data archive includes historical earth imagery going back more than forty years, and new imagery is collected every day.

Cloud based platforms such as Earth Engine offer a new approach to geoscience research and teaching using web-based development tools for interactive algorithm development with instant access to petabytes of data from simple cost effective devices.

NOTES

GeoSocial: Exploring the usefulness of social media mining in the applied natural geohazard sciences

Emma J. Bee^a, Rosa Filgueira^b, and Jacob Poole^a

^a British Geological Survey, Environmental Science Centre, Nicker Hill, Keyworth, NG12 5GG, UK;

^b British Geological Survey, The Lyell Centre, Research Avenue South, Edinburgh, EH14 4AP, UK;

Correspondence: Emma Bee, British Geological Survey, Environmental Science Centre, Nicker Hill, Keyworth, Nottingham, NG12 5GG, UK.

Email: ebee@bgs.ac.uk

Tel: +44(0)115 9363044.

Obtaining real-time information about a geohazard event as it unfolds, such as a flood or earthquake, used to be largely limited to the professional media. Nowadays, obtaining news stories from social media (e.g. Facebook, Twitter, YouTube, Flickr etc.), directly as they unfold, is becoming the 'norm' for many in society. The Haitian Earthquake in January 2010 and the Great East Japan Earthquake in March 2011, provided some of the first natural hazard examples, to really demonstrate the power of social media over traditional news sources for obtaining, live information from which people and authorities could gain situational awareness.

From an applied geohazard scientists perspective, we are generally (initially at least) remote from the hazard event/ disaster zone but are often called upon to provide advice for government(s), responders and humanitarian response agencies. Social media has potential to enhance risk-based models, real-time monitoring of vulnerability and hazard-related impacts, as well as to provide insights into local resilience which can help disaster preparedness and recovery, all of which could inform scientific response and improve situational awareness.

GeoSocial is an openly available tool which allows users, including natural hazard scientists, to view geo referenced posts about landslides, aurora, flooding, volcanic eruptions and earthquakes from social media (i.e. Twitter) on a map in near to real time (see Figure 1). It was developed following the success of a challenge, called 'Hazard Map' jointly submitted to the International Space Apps 'hackathon' event held at the Met Office, UK in 2012. 'Hazard map' showed quite quickly that when a volume of 'tweets' located in a similar location and containing the same keywords, such as 'earthquake' were displayed as a heat map, your eye was drawn to the 'heat' and thus the 'hazard event and its location, enabling the user to be alerted to the situation on the ground and its impacts. The ability to automatically retrieve such 'User Generated Content' (UGC) for a number of hazards is beneficial to the 24/7 operator in the Met Office Hazard Centre, but also to an associated natural hazards science team e.g. working at the British Geological Survey. Such scientists could potentially make use of the wealth of information, publically available through such sites as Twitter, and thus help advance scientific understanding and/or provide better, or timelier, information and/or advice to government.

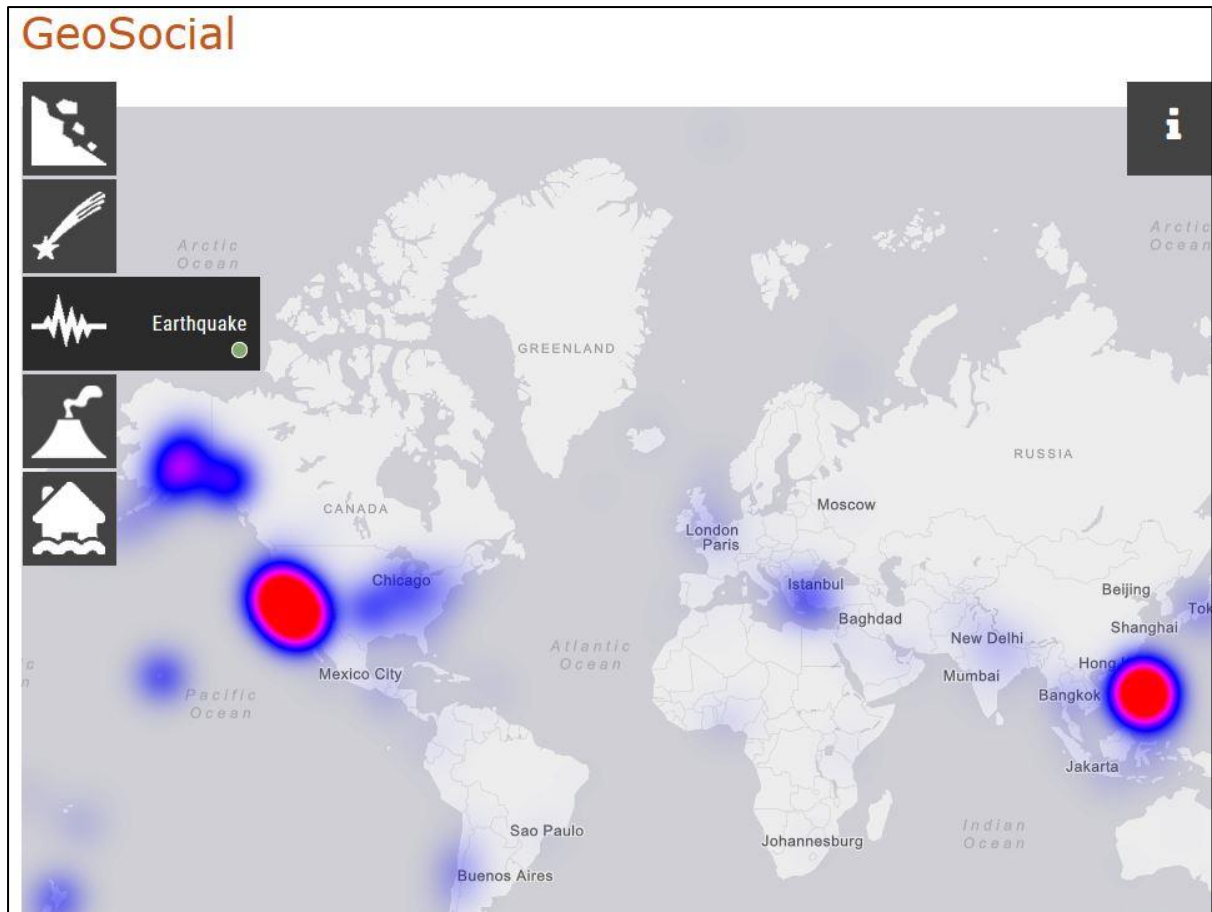


Figure 1: The BGS' 'Geosocial' heatmap showing people posting on Twitter tweets about earthquakes on 25th May 2017. There was a 5.3 magnitude earthquake in the Philippines at this time.

Whilst there are opportunities to obtain knowledge from social media, there are also many challenges. Obtaining reliable content that is accurately geo-located information is just one concern; ensuring that the information is not skewed by the sample demographic is another. There are also a number of ethical considerations around using this data that need to be explored. Whilst, the use of Big Data analytical tools such as machine learning can help resolve some of the challenges around usefulness of content, not all of the challenges are as easily resolved. This presentation explores the usefulness of such social media mining in the applied natural geosciences and discusses some of the opportunities and challenges face in more detail.

Acknowledgements

The authors would like to acknowledge and thank Joanne Robbins at the UKMet office for her joint responsibility in designing the 'hazard map' challenge for the International Space Apps Challenge 2012 and thus the concept behind BGS' GeoSocial application.

NOTES

Big Data in the industry: A critical examination of modern data collection and use in engineering geosciences

Oliver Dabson¹, Ross Fitzgerald¹

¹ CH2M, Elms House, 43 Brook Green, London W6 7EF

In recent years, the UK engineering geoscience industry has undergone something of a technological revolution. The introduction of the UK Government's Building Information Modelling (BIM) initiatives for public sector projects, combined with the more widespread availability of open-source datasets and advanced technology to practitioners, has resulted in the compilation of gigabytes of geospatial data for even the smallest engineering geoscience projects. This flood of information presents several perceived benefits, supplementing the user's knowledge to aid in the assessment of terrain, the calculation of risk, and the modelling of geological phenomena in three, sometimes four dimensions. However, the sheer volume of data, which in some cases is collected or accessed as a matter of procedure rather than to fulfil a specific objective, means that important datasets can be missed, lost or underutilised.

CH2M has been actively promoting the use of innovative technologies in terrestrial and submarine domains. It has recorded the feasibility and practicality of the methods used to collect data and assessed the quality of this data against its aim. Elevation datasets such as those produced from LiDAR, terrestrial laser scanning (TLS) and multibeam echosounders (MBES) and subsurface datasets produced from cone penetration tests and geophysical surveys have been employed on projects both within the UK and abroad to produce innovative solutions to client problems. We examine how these data and workflows have been used to fulfil project objectives in order to make comments on key obstacles to overcome for the future. We find that, although certain systems and datasets have inherent issues which can be ironed out with future updates to the various software packages, the limiting factor the uptake of effective and efficient Big Data solutions is human rather than technological. This challenge relates to improving the often-inadequate collaboration between the end-users of data, who can provide expert guidance on data requirements and purpose, and those that develop data specifications early in the project cycle.

NOTES

Science and the digital revolution: data, standards and interdisciplinary integration

Geoffrey Boulton

School of Geosciences, University of Edinburgh; and President of the Commission on Data for Science and Technology (CODATA)

The explosion of powerful digital technologies for data acquisition, storage and communication in the last two decades has unleashed a digital revolution that has profound opportunities for science to discover hitherto unsuspected patterns and relationships in nature and society, on scales from the molecular to the cosmic, and all in areas of human concern, from local health systems, to global sustainability. It has led to considerable advances in areas such as the search for new drugs and medical treatments, the forecasting of weather and climate, and the complex patterns of ocean circulation.

But one major potential is unrealised: to merge and integrate the data from different disciplines that will permit recognition of deep patterns in the multi-faceted complexity that underlies most of the major global challenges that humanity currently faces. This failing arises from the varying standards that have been used in the different disciplines of science to codify their data, and, in some disciplines, the inadequate definition of the vocabularies needed to categorise them, with the result that the integration of diverse data can generally only be achieved between closely allied fields. Developing efficient procedures for integration of data across all the disciplines of science, including the social sciences, could have a profound impact on the way that science is done in the 21st century and its capacity to contribute to solutions for major global challenges.

It is for these reasons that a community initiative is being launched, under the auspices of the International Council for Science (ICSU) that is designed to evolve into a decadal effort on the interoperability required to address major, interdisciplinary issues. The first, pilot phase, will be dedicated to three important global challenges of **disaster risk reduction, sustainable cities** and **infectious disease**, which will serve as proofs of concept for further efforts in improving data discovery, access, and reusability in support of interdisciplinary research. The geosciences make essential contributions to all three, which will depend upon data from and collaboration with the International Union of Geosciences (IUGS) and its Commission for Management and Application for Geoscience Information (CGI) and other related bodies. It is vital that the geosciences contribute their knowledge and capacities in engaging with such that are fundamental to achievement of the UN Sustainable Development Goals.

NOTES

Getting Legacy BGS Stratigraphic Data Ready for the Big Data Revolution

Mike Howe¹, Junxuan Fan² and Daniel Condon¹

1. *British Geological Survey, Keyworth, Nottinghamshire, NG12 5GG, UK; mhowe@bgs.ac.uk, dcondon@bgs.ac.uk*
2. *State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing 210008, China; fanjunxuan@gmail.com*

As part of its national role to understand the geology of the UK, the British Geological Survey has been carrying out systematic surveying of UK geology for over 180 years. These data are summarised in a series of memoirs and maps, however the primary data are mostly held in analogue form and therefore difficult to interrogate for new science. There are about 15,000 technical reports in BGS, most of which has not been digitized and is hard to use, even for the BGS staff. BGS wants them to be open, to be used by the geology community, development of a structured data system is required.

The Geobiodiversity Database (www.geobiodiversity.com) is the 'section-based' database that allows a range of stratigraphic data to be compiled for a given section located in space. The database then facilitates the integration of these data through a range of tools to understand spatial and temporal distribution of data, which is mostly suitable for the data source of BGS.

We have initiated a collaboration between the BGS and the Geobiodiversity Team that is aimed at extracting stratigraphic data from over 15,000 technical report for the geology of the UK. This involves extracting stratigraphic information from the reports, digitising that information within Geobiodiversity Database schemas. These data will become open and accessible primarily through the Geobiodiversity Database including the development of a BGS specific portal. We are also exploring the using of machine learning techniques within the work flow to expedite data ingestion, and incorporation of new functionality such as 'age modelling'. The overarching aims of this collaboration are to (1) facilitate the open accessibility of BGS's stratigraphic data for scientific research; (2) expand coverage of geographic coverage of data with the Geobiodiversity Database through the incorporation of a large, systematically collected dataset for the UK; and (3) develop the functionality of these data systems through sustained collaboration.

NOTES

Data science for earth science – perspectives from industry

Steve Garrett

UK Global Technology Centre, Chevron Energy Technology Company, Aberdeen, UK



Technology has played a key strategic role in the oil and gas industry over many decades. Outside of our industry, advances in analytics, data science, cloud computing, machine learning and the Internet of Things have allowed many human and machine activities to be transformed, particularly in terms of increased efficiency and lower cost.

Above ground, these technologies are ready to be adapted and adopted to transform oil and gas facilities design and operations, leveraging advances in cybersecurity, artificial intelligence, augmented reality, robotics (process automation, inspection, inventory management), additive manufacturing, unmanned aerial surveillance, autonomous underwater vehicles, sensors, facilities digital clones and project information management.

Below ground, the industry's subsurface work has been on a digital journey for 30 years as shown by advances in marine seismic (2D to 3D to 4D to OBN), interpretation (paper to digitisation to workstations to 3D models) and hardware (mainframes to workstations to clusters to data centres to the cloud). New digital technologies provide opportunities for the next step change in seismic (autonomous acquisition, Distributed Acoustic Sensors using fibre optics), interpretation (multi-variate pattern recognition using Artificial Intelligence), and infrastructure (cloud). However, the subsurface is usually under-sampled, and measurements are diverse and indirect – i.e. we measure properties such as acoustic impedance and resistivity, rather than direct measurements of oil and gas. This points to the continued importance of earth science domain knowledge, combined with computer science and statistical methods, to properly frame projects to apply data science to earth science.

To build a sufficiently large and resilient data set to apply such methods, there is potential for the industry to collaborate to address these challenges and opportunities, as shown by the current UK continental shelf project championed by the Technology Leadership Board and Exploration Task Force, to apply machine learning to seek missed pay using well logs from the Northern North Sea.

NOTES

Disparate E&P Big Data Impose Non-Desperate Machine Learning Methodologies

Keith R. Holdaway FGS

Advisory Industry Consultant, SAS Global O&G Domain

The oil and gas industry is overwhelmed by disparate data, both structured and unstructured, collated across multiple geoscientific and siloed disciplines. Moreover, the data are growing exponentially as digital oilfields are being implemented to manage conventional and unconventional reservoirs. Performing Exploratory Data Analysis (EDA) and generating data marts tailored to specific advanced analytical workflows are cornerstones to enable development and deployment of predictive models that are data-driven, both in real-time and across historical data sets.

The process to characterize the reservoirs of a mature field encapsulates the analysis of large datasets collected from well tests, historical production data and core analysis, enhanced by high-resolution mapping of seismic attributes to reservoir properties. It is imperative to capture the more subtle observations inherent in these data sets. Invariably, geostatistical methods can be implemented to accurately quantify heterogeneity, integrate scalable data and capture the scope of uncertainty. However, between 50 and 70 percent of the allotted time for any reservoir characterization study worth its investment should be concentrated on EDA. As an overture to spatial analysis, simulation and uncertainty quantification, EDA ensures consistent data integration, data aggregation, and data management, underpinned by univariate, bivariate and multivariate analysis.

To build data-driven models that can predict under uncertainty is essential to rapidly identify multi-dimensional parameters in a multivariate environment and thus surface hidden patterns and relationships in data that subsequently reduce time and resources in the critical decision-making cycles. With improved workflows and advances in High-Performance Computing, it is now possible to ascertain risk and quantify uncertainty for very large populations of data without sampling and losing knowledge. There is currently a rapid adoption of Machine Learning and Artificial Intelligence techniques driving significant innovations across the geoscience community. The massive data deluge observed over the past 15 years, combined with the recent price drop in the barrel of oil, is forcing Exploration and Production geoscientists to adopt new technologies. Hard data and soft data that tend to obviate from industry standards are the new oil. Geoscientists must harness the power of the potential knowledge hidden in the data.

Let predictive models be driven by the data and not by empirical petroleum engineering algorithms or deterministic methodologies. By marrying the stochastic with the interpretive school of thought, the upstream community can develop robust data-driven models that are kept current as new data are introduced.

NOTES

Big data in the Geoscience: A portal to physical properties

Andrew Kingdon, Mark Fellgett and Martin Nayembil

British Geological Survey, Keyworth, Nottingham, NG12 5GG, aki@bgs.ac.uk



Geosciences were early adopters of both computing and digital data; the precursors of the SEG-D and SEG-Y geophysical formats date from as far back as 1967. Data standards, for seismic (SEG-Y, SEG-D) or geophysical log (LAS, DLIS) data simultaneously make interpretation and visualisation of data practicable but also their binary nature makes applying analytical techniques unusually complex. Specialist software is often required to process and interpret different datatypes.

Such problems are exacerbated by historic poor data management practices. Datasets are rarely collated at the end of projects or stored with sufficient metadata to accurately describe them and many strategically useful datasets reach BGS incomplete, unusable or inaccessible. Whether this situation arose through a lack of foresight about the future value of data, poor practise or simply storage space restrictions these problems pose huge challenges to today's geoscientists.

Consequently, there are major problems with applying big data analytics to geoscience. For example, many techniques don't sample geology directly but use proxies needing further interpretation. The use of analytical techniques have commonly been limited by the high proportion of noise incorporated into the datasets with very significant interpretation skills required to identify the signal. Thus far successful applications of "big data" analytics have been limited to closed systems or analyses of very common digital data types.

Significant problems remain, including the lack of data that can be immediately interacted with and difficulties in bringing together multiple datasets about related phenomena. Also the lack of adequate metadata about the data available to understand its context and scope and how to apply and qualify results. Whilst geosciences datasets have all the attributes of big data – volume, veracity, velocity, value and variety – the last two controls are disproportionately significant. The first of these determines the usefulness of the data and the second is the biggest impediment to delivering on the promises that big data offers especially in Earth Sciences.

In order to deliver a standardised platform of data from which individual geological attributes can be identified BGS has invested in the creation of PropBase (Kingdon et al., 2016). This single portal facilitates the collation of datasets supplied in standardised formats. This allows all data from a single point feature (e.g. boreholes) or areas of interest) E.G. to be extracted together in a common format allowing all data to be immediately compared. The existence of PropBase portal allows a researcher to answer the question "What's available at a location?" It has already been used in site characterisation for the UK GeoEnergy Observatories project.

Such initiatives that allow collation of high volumes of data in a single extractable format are a critical step forward to allowing Big Data analytics. Combined with the increasing availability and ever lowering cost of high power computing and analytical routines, the opportunities for big data analytics are ever growing. However, substantial challenges remain and new and more interactions with computer scientists are needed to deliver on this promise.

References

Kingdon A., Nayembil, M.L., Richardson A.E., and Smith, A.G., 2016. A geodata warehouse: Using denormalisation techniques as a tool for delivering spatially enabled integrated geological information to geologists. *Computers and Geosciences*, 96, 87-97. DOI: 10.1016/j.cageo.2016.07.016.

NOTES

Building Data Science Capability

Ed Evans

NDB Upstream

FGS, FBCS, A committee member of the BCS (British Computer Society) Data Management Specialist Group.

A technologist from a very large E&P recently asked me ‘what software did you use for...?’ referring to a published case study on Data Analytics. The question suggests that there is an application that solves the problem. However, solutions are often more about process than technology. Data Science and Analytics is not a simple technology challenge. Like all business opportunities it is best addressed by a planned and coordinated approach with clear business goals, where the digital technology is an element of the solution.

Data, Data Science, Big Data and Analytics – whatever phrase is used - ought to be a key strategy in an industry coming out of recession and needing to maintain gains in efficiency and cost management. How does an oil company benefit from the approaches and tools from the world of Big Data without wasting millions on technology or consulting? How does an oil company, big or small, build the basic capability to be able to evaluate or select potential big data opportunities?

I would like to share the lessons learned from organising E&P Data for more 25 years and my recent big data and analytics experience in geoscience, drilling and production engineering.

My talk will refer to the following recent projects:

- National Oil Company: start with the business problem
- Large Independent: use experienced Subject Matter Experts
- Small Independent: prove the concept before scaling up
- National Archive: unstructured data must be structured

With reference to these recent examples the talk will aim to address some of the key questions around the development of Data Science capability.

Questions addressed:

- Technology Plays in Big Data
- Data Science vs. Data Management
- Data Science and the organization
- What are the lessons from previous examples?
- The Value of Big Data Strategy

NOTES

The Virtual Geoscience Revolution

John A. Howell¹, Magda Chmielewska¹, Simon J. Buckley²

¹ School of Geosciences, University of Aberdeen, UK

² Uni Research AS, Bergen, Norway

Since the start of the millennium, there has been a quiet revolution occurring in the Earth Sciences. For 200 years, from the time that William Smith published his ground breaking map, up until the turn of the millennium, field data were collected on paper maps using pencils and Rotring pens. Structural measurements were made with a compass clinometer, and lithology and mineralogy were manually interpreted using a hand lens and acid. Less than 20 years later, outcrops are routinely digitised using data collected by laser scanners and drones. Lithology can be remotely mapped using hyperspectral imagery and outcrop interpretation, and even fieldtrips can be undertaken from the desktop computer. We are already seeing advances into virtual reality and usable augmented reality is around the corner. The story of this digital outcrop revolution is the focus of this presentation and the six Virtual Fieldtrips being presented at this conference.

The virtual geoscience revolution resulted from parallel developments in geospatial positioning (GPS) and computing power, especially the ability to represent 3D surfaces and volumes in the workstation. The maturation of terrestrial lidar in the early 2000s provided a method to collect very large quantities of 3D point data and when combined with photographic data led to some of the first, photorealistic models or *virtual outcrops*. The cost and acquisition times associated with lidar surveying meant that virtual outcrops remained a fairly niche tool for geological research until 2011 when photogrammetry, a hundred year old method of generating spatially constrained points, directly from photos emerged into the mainstream, driven by new software algorithms and faster computers. This re-emergence of photogrammetry was coincident with the advent of cheap, high quality consumer drones that could collect hundreds of photos from outcrops in tens of minutes. Suddenly, virtual outcrops could be generated by anyone with a few hundred pounds to spend on a drone and a fast PC. Virtual outcrops have now entered the mainstream and are a revolutionising the way in which field data are collected and utilised.

A virtual outcrop is a geospatially constrained, photorealistic model of a cliff section within the computer. Virtual outcrops can be used to make structural measurements, measure bed thickness, and map geobodies and surfaces directly on the computer. Very large amounts of geological data can be generated far more quickly on the computer, than in the field. Cliff sections that are dangerous to reach can be accessed safely providing more statistically representative data coverage. The standard photorealistic model can be draped with other data such thermal or hyperspectral images to provide a unique insight into the geology. In addition to research opportunities, virtual outcrops can be used for teaching, either to supplement existing field courses or even to replace them for students who cannot get to the field. Libraries of virtual outcrops make it possible for students, researchers and industry personnel to see a set of the best examples of a given geological feature from around the world in a single session, helping to avoid the biases that arise from studying single data points. Virtual outcrops have been built from the pore (μm) to the basin scale (10s km). We even have virtual outcrops of Mars.

We are at the dawn of this new era. Developments in virtual reality already mean that the first immersive virtual outcrops have been developed. Faster data acquisition, automated interpretation by machine learning, vast user libraries of the world's best outcrops, and augmented reality for use in the field are all on the horizon. Welcome to the Virtual Geoscience Revolution.

NOTES

Paradigm Shift Necessary for True Integrated Asset Value Maximization by Leveraging Big Data and Digital Transformation

Dr. Satyam Priyadarshy

Halliburton

Integrated asset value maximization has been a topic of interest to upstream oil and gas industry for more than two decades with limited to no success. Traditionally, the industry has focused on optimization, rather than maximization. There are many challenges in achieving integrated asset value maximization, including, but not limited to, data integration, technology platform, and realizing value from data.

As an industry, the focus has been on geoscience software integration, subsurface uncertainty, and pressure/volume/temperature (PVT) consistency followed by optimization, thus providing incremental value by moving away from "silo" solutions. However, for true value maximization, a paradigm shift in the way the E&P lifecycle is viewed is necessary. As an industry, the big data and digital transformation need to be understood and developed.

The underlying principle for value maximization lies in the concept of an integrated, agile, and scalable platform providing the data, information, and actionable insight for near real-time or real-time value creation. This paper discusses challenges providing the integrated geosciences and related data sets for simulations, modeling, and insight generation on a continuous basis. It also discusses the specific, measurable, attainable, realistic, timely (SMART) approach for addressing these challenges and building a solution based on the global earth model (GEM).

NOTES

Mining Geological Sentiment from Unstructured Text

Dr Paul H Cleverley

Associate Lecturer, Robert Gordon University, Aberdeen, UK

Email: p.h.cleverley@rgu.ac.uk



Text and data mining is the use of automated analytical techniques to analyse text and data for patterns, trends and other useful information. Cognitive Computing combines this with Natural Language Processing (meaning) and Machine Learning (prediction) to 'mimic' human thought processes to augment decision making and provide actionable insights. This may be increasingly significant, given the exponentially growing 'information haystack' of geoscience literature and reports that exist both inside companies and within external journal/commercial repositories.

A collection of documents as a whole, may also be greater than the sum of its parts, yielding trends not discernible from any single document and too small to be picked up through manual means. There are also suggestions that the vast amount of unstructured information available could be used to challenge cognitive bias and/or organizational dogma. A comparative trend deemed 'surprising' to a geoscientist could stimulate creativity, leading to a learning event and ultimately, business value.

The dominant *text mining in Geosciences* published literature focuses on rules based extraction, such as 'counting' the frequency of specific concept occurrences within text and visualizing the results spatially and by Geological Time. Automated sentiment analysis has been widely applied to infer attitudes, opinions and emotions towards institutions, brands and topics for example. However, no known study has examined the context in which geoscience concepts are mentioned – *geoscience sentiment*.

Traditional 'out-of-the-box' sentiment algorithms can perform poorly without customization. For example, the terms '*fault*', '*thick*', '*old*' and '*expelled*' tend to generate 'negative' polarities as they have been trained on generic content which is not geoscientific in nature. For example, a "...*thick reservoir*" or "*Expelled hydrocarbons..*" is not generally bad news!

An ensemble machine learning approach was applied using OpenSource Python TextBlob to public domain petroleum system assessment reports. Over 1,000 sentence/semantic unit examples from public domain reports for the petroleum system elements '*source rock*', '*reservoir*', '*trap*' and '*seal*' were classified by geologists. These data were combined with a custom lexicon to train a Bayesian statistical classifier. Skip grams were used to incorporate word order effects in the model.

This machine learning classifier was applied to a test set of public domain documents in order to generate pie-charts per geological basin and by geological time. This visualization showed relative proportions of 'positive' and 'negative' sentiment towards each petroleum system element through time, simulating a *work task based search application* exploiting big data in the geosciences. This could conceptually be more widely generalizable to other geoscience work tasks given the necessary training. The results from the exploratory study including the visualizations, recall, precision and F1 scores will be presented along with limitations and areas identified for further research.

NOTES

Putting Data at our fingertips: Utilising data analytics in E&P Information Management

Christopher Frost, Neil Constantine
DataCo Global Limited,
Christopher.frost@dataco.co.uk

Improved technology accessibility, coupled with an E&P industry under pressure to reduce costs and shorten cycle times, have led to increased interest in data analytics and machine learning. These techniques have been successfully applied in other industries yet E&P has been slow to adopt, in part due to our large volumes of unstructured data. Estimates predict an exponential growth in unstructured data by the next decade¹ whilst data analytics has been highlighted as a top area of focus for oil and gas companies, with the potential to automate up to 30% of existing activities². This paper describes a real process to leverage information from unstructured data, which contain nuance and context not seen in traditional repositories, despite the many challenges.

Combining E&P experience with lessons from outside our industry, our process combined taxonomic analysis, optical character recognition (OCR) and natural language processing (NLP) to 'pull' information from End of Well reports and place this at users' fingertips. The value is twofold: high performance computing accelerates this process from weeks to hours, and the additional information *at an early stage* helps manage subsurface uncertainty and focus interpreter effort.

The process applied taxonomic analysis on file structures and metadata to rapidly identify what documents were relevant to the interpreter. These documents may be machine-readable, e.g. Word documents, or needed additional effort, e.g. scanned legacy reports, in which case an OCR process was applied to convert these images containing text into text files. Pre-processing techniques, including document straightening and image manipulation, were used to improve the OCR, whilst post-processing, including pattern-based substitution and spell-checking using domain-specific lexicons, minimised errors within the processed text. Additionally, these lexicons were expanded through the use of text analytics and iterative post-processing to further improve content readability.

NLP techniques were then applied to both OCR'd documents and other relevant machine-readable files to identify certain key themes, e.g. formation tops. These themes may be defined by an interpreter, or they may be revealed based on meaningful occurrence of particular text and were then fed into searches against external information stores to report additional information, such as analogues.

Finally, the process identified tabular data in source documents and extracted this to application-specific load sheets for review by the interpreter, after having applied logical QC to identify errors in the source document or the OCR process. It should be noted that the fidelity of these load sheets was based on a combination of source document quality and limitations of the OCR process, so we would not advocate use of these data without interpreter review.

This example has demonstrated a combination of E&P experience with automation and analytics using accessible technology to support the Geoscientist in better understanding of what is happening in a basin or well. It breaches historical limitations of time and capacity to

¹ Third International Workshop, HCI-KDD, 2013. Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data.

² Institute, McKinsey Global. 2016. THE AGE OF ANALYTICS: COMPETING IN A DATA DRIVEN WORLD.

allow the assimilation of large volumes of data on a shortened timeframe and – using an affordable architecture – delivers value to the E&P industry.

NOTES

Accessing Knowledge in Geoscience Text using Natural Language Processing

Richard Jones

*Geospatial Research Ltd, Department of Earth Sciences, University of Durham, DH1 3LE, UK
richard@geospatial-research.com*

Large quantities of geological, geophysical and geospatial data exist in both the public domain and in industry, particularly within hydrocarbon, mining and geotechnical companies. Enhanced availability and accessibility of geoscience data (enabled by improved processing power, storage and bandwidth, an increase in primary digital acquisition, and digitization of legacy analogue data), has helped to raise expectations that 'big data' will inevitably equate to 'new insights' that can readily be monetarized. Deriving additional value from large data repositories generally involves significant challenges, though many of these have already been identified and analysed by Artificial Intelligence and Knowledge Management initiatives spanning many decades.

A key realization is that different types of data can require radically different analytics, and strategies that work well in one industry or with one type of data are not necessarily easily transferable to other domains. This is illustrated by contrasting the types of analysis and insights gained from large geospatial datasets (typically highly structured, consistently formatted, with comprehensive metadata), compared with large archives of text documents such as scientific journals or reports on a corporate file server.

The potential value of 'mining' text repositories is demonstrated with new prototype software based on Natural Language Processing (NLP) to analyse and characterise geoscience documents. The underlying challenge in text mining is that most additional value within text documents is held within the semantic structure of the prose. Whereas keyword search strategies rely on the binary presence/absence of specific search terms or phrases, written text conveys meaning and relevance through the wider context of surrounding sentences and paragraphs. NLP is used to analyse the information content of text, and relate this to pre-defined ontologies that represent the semantic framework of knowledge within a specific domain.

NOTES

Detecting Volcano Deformation in InSAR using Deep learning

N. Anantrasirichai^{1*}, F. Albino², P. Hill¹, D. Bull¹, J. Biggs²

¹ Visual Information Laboratory, University of Bristol, UK,

² Volcanology, University of Bristol, UK

*n.anantrasirichai@bristol.ac.uk

Globally 800 million people live within 100 km of a volcano and currently 1500 volcanoes are considered active, but half of these have no ground-based monitoring. Alternatively, satellite radar (InSAR) can be employed to observe volcanic ground deformation, which has shown a significant statistical link to eruptions (Biggs, et al., 2014). Modern satellites provide large coverage with high resolution signals, leading to huge amounts of data. For example, the Sentinel-1 satellite allows us to acquire images of each of the world's volcanoes on a routine basis. It has a repeat cycle of 12 days and acquires data with a 250-km swath at a 5 m by 20 m spatial resolution (single look). This data is generated with greater than 10 TB per day or about 2 PB collected between 2014 and June 2017 (Fernández, et al., 2017). The explosion in data has brought major challenges associated with timely dissemination of information and distinguishing volcano deformation patterns from noise, which currently relies on manual inspection. Moreover, volcano observatories still lack expertise to exploit satellite datasets, particularly in developing countries.

Here, we present a novel approach to detect volcanic ground deformation automatically from InSAR images. This approach brings together satellite-based volcano geodesy and machine learning algorithms to develop new ways of automatically searching through large volumes of radar data to detect unusual patterns within the images. In this study, we use wrapped-phase InSAR images from NERC-COMET-LiCs (González, et al., 2016) covering volcanic regions in Ethiopia, Kenya, Italy and Turkey. Each image has an approximate size of 19500x19700 pixels (~1GB). A diagram of the proposed framework using deep learning with a convolutional neural network (CNN) is shown in Figure 1.

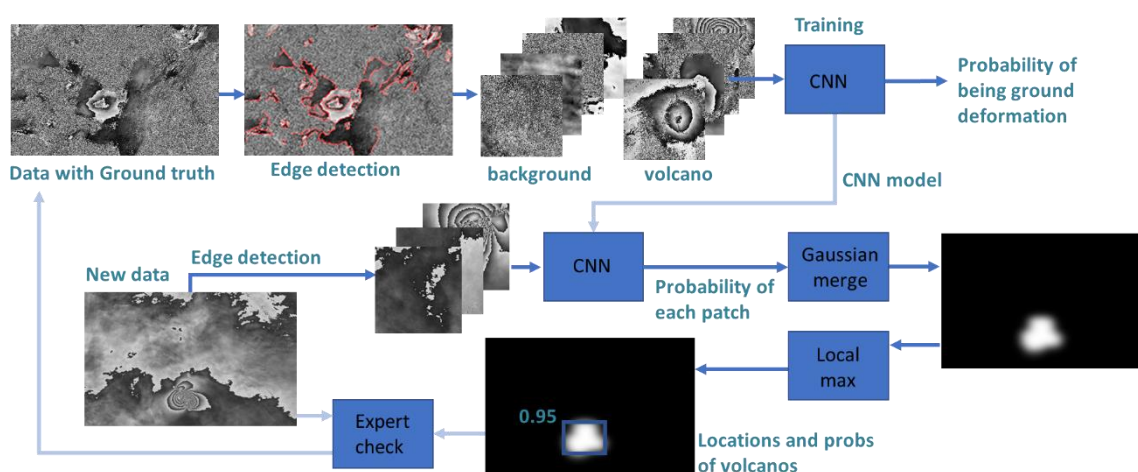


Figure 1: A proposed method to detect volcano deformation

Each training image is divided into patches equal to the size of input of the CNN (e.g. 224x224 pixels for AlexNet). They are overlapped by half of their size and the number of volcano patches are increased by shifting around the volcano area. As the number of background areas (negative samples) are significantly larger than those of the volcano patches (positive samples), only the background patches in which strong edges have been detected are used. To classify a potential signal as volcano deformation, patches with strong edges are tested and the probability results are merged with Gaussian weights. Finally, ground-truthing is

performed by an expert and if the detection is a false positive, it is incorporated in the ground truth to update the CNN model.

Figure 2 shows the Receiver Operating Curves (ROC) for 2-fold cross validation for the considered dataset. It compares several pretrained CNN architectures and texture features with a support vector machine (SVM) classifier as a baseline. The best true-positive rate and true-negative rate are 0.899 and 0.992, respectively, which are the results of AlexNet. Generally big data presents significant challenges to deep learning because of its large scale, heterogeneity, and non-stationary distribution (Chen & Lin, 2014). Our results demonstrate that deep learning with CNNs (Krizhevsky, et al., 2012) has significant potential to capture characteristics of volcano deformation present in InSAR data. The next step is to test whether these methods are capable of distinguishing between deformation signals and atmospheric artefacts in single images, or whether full time series are required.

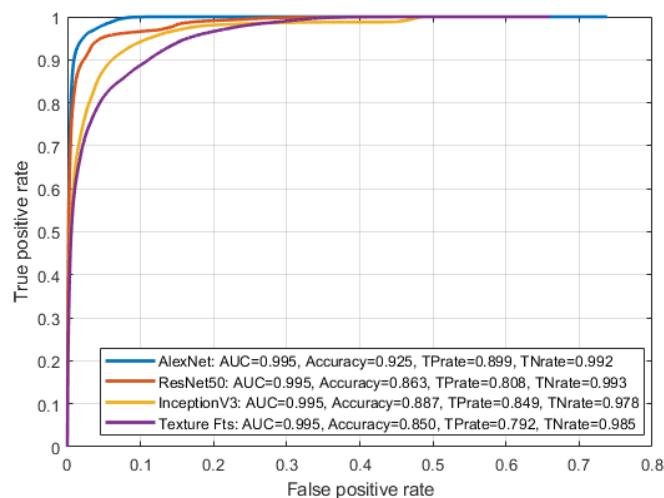


Figure 2: ROC curves for the 2 folds of cross validation

References

- Biggs, J. et al., 2014. Global link between deformation and volcanic eruption quantified by satellite imagery. *Nature communications* 5.
- Chen, X. W. & Lin, X., 2014. Big Data Deep Learning: Challenges and Perspectives. *IEEE Access*, Volume 2, pp. 514-525.
- Fernández, J., Pepe, A., Poland, M. & Sigmundsson, F., 2017. Volcano Geodesy: Recent developments and future challenges. *Journal of Volcanology and Geothermal Research*.
- González, P. et al., 2016. *LiCSAR: Tools for automated generation of Sentinel-1 frame interferograms*. s.l., AGU Fall Meeting.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, Volume 25, pp. 1097-1105.

NOTES

The Role of Data in Regulating the UK Oil & Gas Industry

Nick Richardson

OGA

The United Kingdom Oil and Gas Authority (OGA) was created as a result of one of the key recommendations of the 2014 Wood Review. The OGA is a government company with the role of regulating, influencing and promoting the UK oil and gas industry. A series of strategies and associated delivery programmes set out how the OGA, UK government and industry should work together to achieve the principal objective of maximising economic recovery (MER) from the UK Continental Shelf. Of fundamental importance to the attainment of this objective is the provision of high-quality data in order to support industry activity.

The OGA has been working in a number of areas to enhance data provision to industry, through proposed changes to regulations, the establishment of a National Data Repository, enhancement of external portal applications, publication of Open Data and web services, the acquisition of modern broadband seismic data, and an annual survey of industry activity. The OGA has published, and will continue to publish insights into the datasets that it holds, with the latest examples including a new set of regional geoscientific maps across the UKCS and technical summaries of subsurface targets. Changes to the OGA's data architecture will increasingly enable the application of data analytics and machine learning techniques to identify and exploit exploration and production opportunities.

NOTES

Over 120 years worth of hydrocarbon exploration, an example of how legacy data can address today's challenges.

Mark W. Fellgett^{1*}, Andrew Kingdon¹ and Anthony Myers²

¹British Geological Survey, Environmental Science Centre, Nicker Hill, Keyworth, Nottingham, NG12

² Retired

Unconventional hydrocarbon exploration has been a contentious issue in the UK since tremors were associated with a hydraulic fracturing operation at the Preese Hall 1 well in Lancashire (Green et al., 2012). Following this a review of hydraulic fracturing concluded that more information was needed to characterise faults, fractures and stresses in UK shales (Mair et al., 2012).

One of the key requirements for planning a Shale Gas operation is an understanding of the in-situ stress field. The stress field is important as it predicts both the orientation and the plane in which hydraulic fractures propagate. However as of 2016 stress field information across the UK was limited with only 24 sites having information to characterise the stress field (Heidbach et al., 2016).

To investigate whether legacy data could be used to help characterise the stress field required an examination of the BGS archives. As a national data repository the British Geological Survey holds records for over 2000 onshore oil and gas wells drilled across the UK landmass which date from 1896 to present day. The data types associated with these wells vary from formats such as TIFF, bitmaps, text files and pdfs to industry specific non binary formats such as LIS and DLIS. Effective use of this archive requires a mixture of data mining, expert interpretation and innovate techniques to convert the files to more accessible formats.

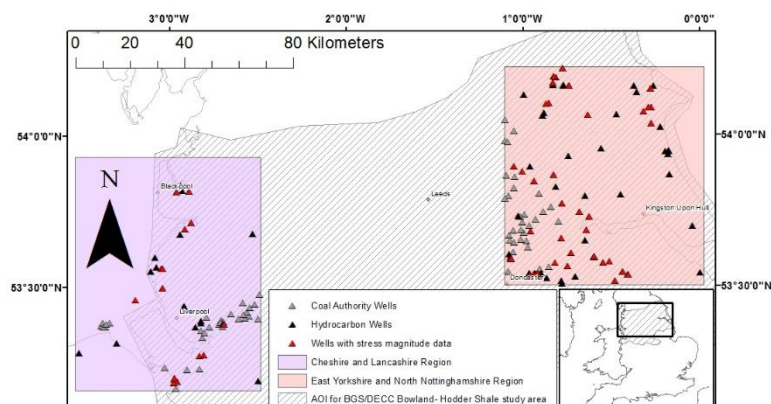


Figure 3: Map showing two UK regions where stress magnitudes have been investigated. Deep Boreholes are the most common source of stress field information and are shown as triangles. Boreholes in red have information to characterise stress magnitude data. Hatched zone corresponds to the area of interest from the BGS/DECC Bowland-Hodder Shale study, Andrews et al. (2013).

provide information to operators and regulators two areas were selected to see if any hydrocarbon boreholes had relevant information.

Utilising the legacy data found stress field information available for around 40% of the hydrocarbon boreholes across the regions. In total legacy information is now available for 75 sites across the UK and is now being used in the planning of well operations. However more

Much of the relevant information was contained in a series of scanned reports held as multi page files in TIFF format. Interrogating these files required the conversion to pdfs and then the use of optical character recognition which was performed on mass using FME and Adobe Acrobat Pro.

Currently there is active unconventional exploration interest in the UK, particularly in two areas underlain by the Bowland Shale Group (Figure 1). To identify relevant data to

work is needed to investigate these issues and better techniques are required to automate the process of collecting stress field information.

References

1. Andrews, I.J. 2013. The Carboniferous Bowland Shale gas study: geology and resource estimation. British Geological Survey for Department of Energy and Climate Change, London, UK.
2. Green, C.A., Styles, P., and Baptie, B.J. 2012. Preese Hall Shale Gas Fracturing Review and Recommendations for Induced Seismic Mitigation. Report for DECC
3. Mair, R., Bickle, M., Goodman, D., Roberts, R., Selley, R.C., Shipton, Z., Thomas, H. & Younger, P. 2012. Shale gas extraction in the UK: a review of hydraulic fracturing. Royal Society & Royal Academy of Engineering 105pp.royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/shale-gas/2012-06-28-Shale-gas.pdf

NOTES

Digital Transformation in the North Sea

Stephen Ashley, Steve Roberts

Manager, Digital Transformation Solution Centre at The OGTC. Stephen.Ashley@TheOGTC.com

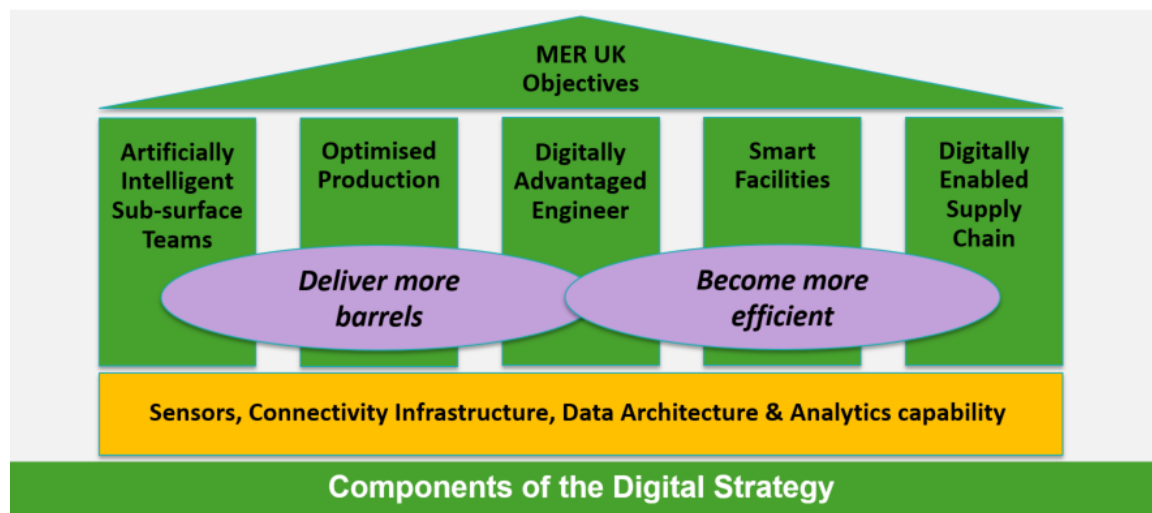
Advisor, Digital Transformation Solution Centre at The OGTC. Steve.Roberts@TheOGTC.com

The Oil and Gas Technology Centre officially opened in February 2017 and quickly establish a Digital Transformation Solution Centre. This Solution Centre invests in developing and testing solutions that collect, manage and analyse large data sets in a collaborative way for the benefit of the basin. Digitally based solutions have the power to not only enable an existing business or technical process but also to transform the way work is done and ultimately, how entire industries operate. Digital solutions that make greater use of typically underutilized data sets to gain insights have enormous potential to contribute to MER UK objectives.

A number of business facing themes have been established, underpinned by the key fundamental enabling themes of digital measurement, connectivity systems, data architecture and analytics capability. These are:-

- Artificially Intelligent Subsurface Teams
- Optimised Hydrocarbon Production
- Digitally Advantaged Offshore Worker
- Smart Facilities
- Digitally Enabled Supply Chain

Six Digital Themes in service of MER UK Objectives

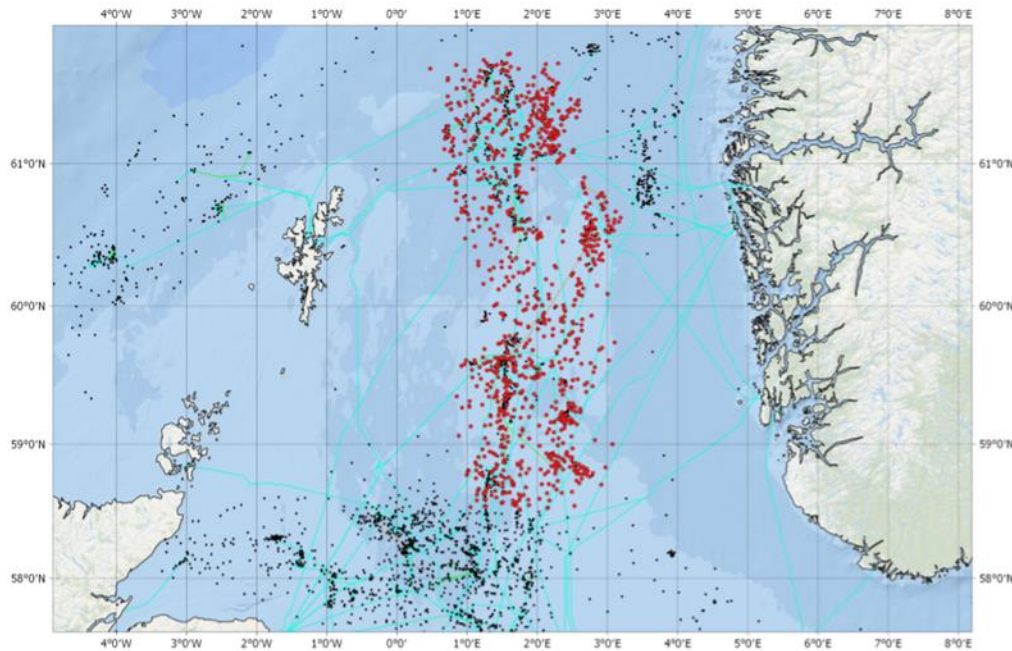


The 'Artificially Intelligent Subsurface Team' vision is focused on finding and producing more barrels. Subsurface teams working on exploration and reservoir development activities regularly compile and analyse large, complex data sets from multiple sources, such as 2D and 3D seismic, well logs and geological reports. These teams create sophisticated subsurface models which integrate as much of the available structured and un-structured data as possible, enabling them to predict the potential for new hydrocarbon discoveries and manage the performance of producing reservoirs. However, these analyses are prone to the limitations of human ability to compile information, objectively assess and interpret the information without bias and carry forward the associated uncertainties.

The vision for the future is to use the latest artificial intelligence (AI) techniques to augment a sub-surface team's ability to integrate and interpret these complex datasets and improve the objectivity and predictability of the assessments. It is envisaged that AI approaches can

potentially automate some of the routine, base data compilation and screening tasks and then augment the capability of the team to interpret and integrate the conditioned data, freeing individuals to focus on the more complex decisions.

The OGTC has recently issued a 'Call for Ideas' focused on using machine learning, or other sophisticated analytical techniques, to identify missed pay opportunities in the North Sea. The selected area of interest is the Northern North Sea where the initial focus is on using all publicly available well data within this province. It is envisaged that future projects would also use available seismic data.



Northern North Sea Area

The intent of the initiative is to quickly and consistently identify overlooked potential exploration opportunities from existing well data. Then, identify new targets with the potential to have an impact on deferring upcoming cessation of production and decommissioning decisions in the Northern North Sea.

Over 175,000 items of data associated with wells are being made available to the selected projects for analysis. There has been great collaboration between the Oil & Gas Authority, Common Data Access Ltd and The OGTC on this initiative.

NOTES

Two basic applications of machine learning for rock-physics and petro-physics

Ehsan Zabihi Naeini

Ikon Science

Machine learning (or artificial intelligence) has been around for decades. The recent popularity and emergence in the oil and gas industry is likely due to the abundance of data and the need to squeeze out more information as well as the availability of computing power, advances in various flavours of Neural Networks, and user-friendly coding and library packages. In this paper two basic applications of using such libraries are shown, for facies classification, and for volume of shale and porosity prediction. There is perhaps no novelty in these applications but rather the focus is on how easy it is to implement such tools.

1- Introduction

Machine learning has been around for decades. For example, one of the most useful and familiar applications which was invented in the 1990s is the spam filter that has helped millions of people in their day to day life. Another useful tool, perhaps more familiar for Geoscientists, is Support Vector Machines (SVM) which algorithmically was developed as early as 1960s and became an industrial tool for classification and regression in 1990s (Smola and Scholkopf, 2004). One could argue SVMs are the ancestors for Neural Networks which by itself have undergone a revolutionary progress with the invention of Convolutional Neural Networks more recently. There are many other examples and algorithms (e.g. Naïve Bayes, Decision Trees, Ensemble Methods etc.). It is not the intention to explain these algorithms in details here. Instead, the focus is on demonstrating the application of two branches of machine learning, classification and regression, for facies classification and petrophysical log prediction. These two applications are essential parts of rock physics and petrophysics analysis workflows and therefore a machine learning approach can save us time (and therefore money) when dealing with a big dataset (e.g. many wells). Furthermore, this paper shows how easily one can implement these applications using publicly available library packages.

2- Facies Classification

Facies classification is important as it enables us to distinctly define rocks of interest and to build a better understanding of the depositional environments encountered in the wellbore. Typical reservoir properties considered for lithology prediction are mineral composition, especially volume of shale, porosity, fluid saturations, acoustic and elastic properties. Identifying facies becomes even more important when one attempts to find their 3D distribution using sophisticated seismic inversion algorithms (Zabihi Naeini and Exley, 2017) in which per facies depth-trended rock physics models are at the heart of such algorithms.

As an example, 5 wells in the Forties field were used to train a SVM for facies classification using P-wave and S-wave sonic, density, porosity and volume of shale and sand logs as inputs. Figure 1 shows one of the wells used for training where as expected there is a good resemblance between the machine predicted and human interpreted facies (this is to be expected as this well is one of the wells used for training). A more useful QC for data scientists is to obtain the confusion matrix from the test dataset (30% of the data was used for testing and 70% for training). Roughly speaking, one would expect that the diagonal elements of the confusion matrix would have higher values than the off-diagonal elements when the classification is performing satisfactorily. Another QC value is to compute the classification rate which is the percentage of how accurate we classified the facies compared to the interpreted facies. Here the classification rate was 92%.

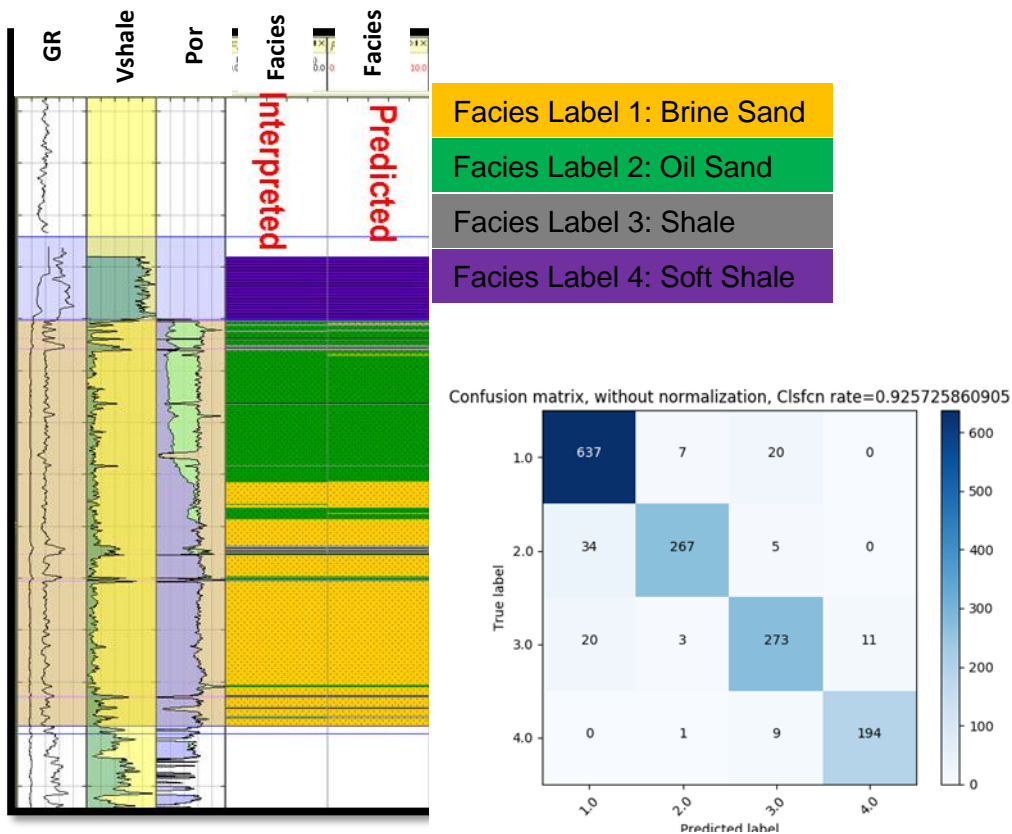


Figure 1: (Left) Well log panels and the interpreted/predicted facies. (Right) Confusion matrix.

The most interesting part is to examine the performance of the classifier on blind wells. This is shown in Figure 2 where one can observe an excellent performance of SVM in predicting the facies which is almost identical to what we interpreted manually. As can be observed later it is rather easy to use other methods of classification and compare the performance of each.

Blind well 1

Blind well 2

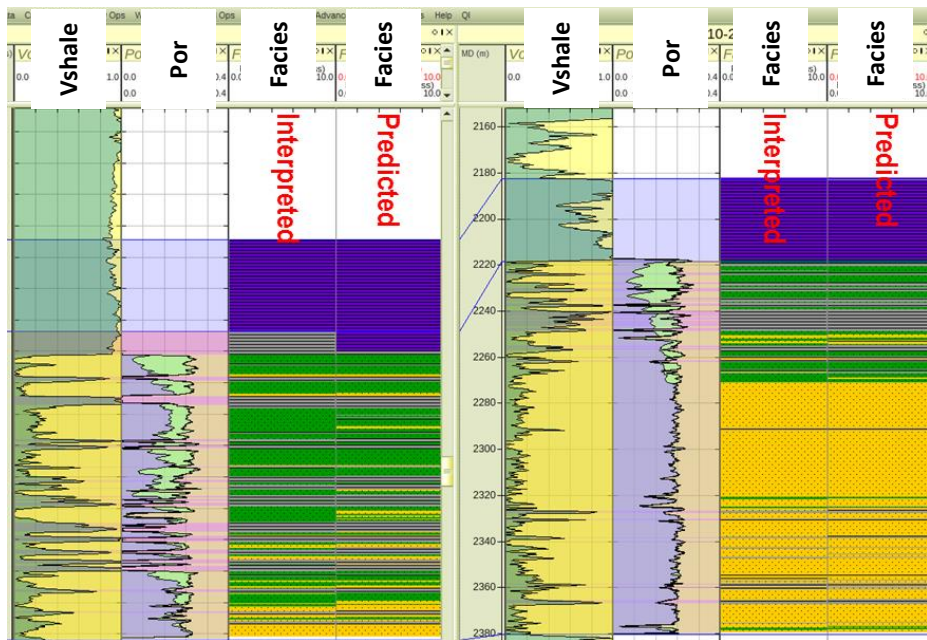


Figure 2: Predicted facies for two blind wells.

3- Petrophysical Property Prediction

Broadly speaking, the inputs to petro-physics workflows are Gamma ray, deep resistivity, density, Neutron and sonic logs and the main outputs are volume of shale, effective porosity and water saturation. Hopefully it is clear for the reader that the outputs here are continuous properties which means the problem is suited for regression. In machine learning, classification and regression algorithms are pretty much the same except for the output (target) variables i.e. whilst in classification the target is a discrete property (e.g. as in the previous section a set of pre-defined facies identified with integer numbers), in petrophysics prediction we aim to model a continuous log using regression. In other words, we can also use a SVM for regression, and thus predict the volume of shale as well as porosity.

It is important that the input data is conditioned to remove any instances of erroneous log response, for example logs responding to rugose hole conditions or logged through casing. For this experiment, data from the Central North Sea area was used, comprising of logs over a range of different rock types and geological ages. In each well, the petrophysics logs (required for training and testing) have been derived using standard deterministic petrophysics methodology.

Four wells were used as a training dataset and one well was used as a blind well test and the objective is to compute the volume of shale and porosity. Hence, Gamma ray, density, resistivity and Neutron logs from the 4 wells were used as a training (and testing with the same 70% and 30% split as before) dataset to predict volume of shale. Then given the volume of shale the same process was repeated to predict porosity. Once the training is completed, the SMV can be used to predict volume of shale and porosity on the blind well – as shown in Figure 3. Overall, it can be observed that the prediction is very good for both volume of shale and porosity logs (total RMS error is 0.02). However, a closer inspection (Figure 3, right) of the sand dominated zone (yellow area in density log) indicates some error between the interpreted (black) and predicted porosity (red) logs. This is due to the fact that in the training dataset from four other wells did not have sandstone with such high porosity values.

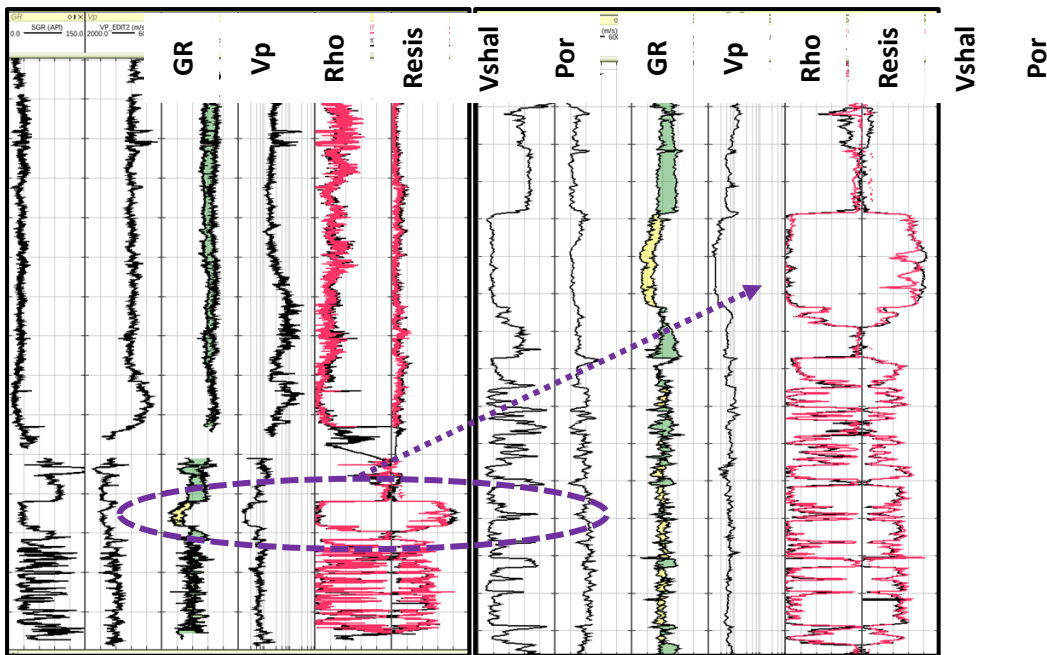


Figure 3: Volume of shale and porosity prediction for a blind well. (Left) Full well log panels and (Right) zoom over the sandy section highlighted with an ellipse. The interpreted volume of shale and porosity logs are shown in black and the predicted ones are in red.

4- Implementation Tips

In this work, Python was used for programming plus RokDoc external interface to access all the well data and scikit-learn (<http://scikit-learn.org>), which is an open source library for simple and efficient data mining. With that in mind, the pseudo code below shows how easy one can implement the methods described in this paper:

```
import numpy as np
from sklearn.svm import SVC

df = get_dataframe(vp,vs,rho,por,facies,volume_fraction_set)
dfdata = df[df['is_null']==False]
clf = SVC()
clf.fit(Xtrain[XfeatureNames],Xtrain['Facies'])

# Apply the classifier to all the log data and pass a facies log back to RokDoc.

PredictedFacies=clf.predict(scaler.transform(dfdata[XfeatureNames]))
return PredictedFacies
```

It is also fairly straightforward to change the classifier algorithm and reuse the rest of the code. However, often in machine learning applications, most of the time taken is to prepare and condition the data which is not discussed here.

5- Discussions and Conclusions

Two simple applications of machine learning were shown for facies classification and petro-physical parameter prediction. The aim here is to encourage the readers to discover these conventional machine learning tools which not only can be helpful in practice but also as was shown they are very efficient and user friendly. Of course the more advanced machine learning

applications, for example convolutional neural networks, would require more work but would also come with bigger rewards.

6- Acknowledgment

I would like to thank Neil Whitfield, Iestyn Russell-Hughes and Michel Kemper for useful discussions during the course of this work.

7- References

Smola, A., and Scholkopf, B., 2004. A tutorial on support vector regression: Statistics and Computing, **14**, 199-222.

Zabihi Naeini, E., and Exley, R., 2017. Quantitative interpretation using facies-based seismic inversion: Interpretation, **5**(3), SL1-SL8.

NOTES

The Chronological Data Explosion: the Huge Potential for Source-to-Sink Insights within Exploration

Graeme Nicoll, Fabian Kohlmann, and Benjamin Saunders
Halliburton
Graeme.Nicoll@halliburton.com

Recent technology advances and the reduction of analytical costs have triggered the availability of a vast wealth of geochronological and thermochronological data within the public domain. Within the last 20 years, for example, the number of published detrital zircon studies has increased at a near-exponential rate. As a result, this expanding volume of new chronology data is becoming increasingly difficult to manage. Significant challenges also remain to standardize and geographically locate these datasets from multiple sources, as well as to interpret them in a consistent manner and to integrate them into a regional understanding in a timely fashion. However, the development of a Big Data-style approach to analyzing these large datasets after they are compiled can provide significant advantages, and new quantifiable approaches provide a major opportunity to aid the explorationist in the continuing search for natural resources.

As demonstrated in this paper, the creation of a globally consistent and standardized database that contains, for example, geochronology, detrital geochronology, and thermochronology information yields significant dividends for understanding the Earth's 4D evolution. Even at the simplistic level of being able to view geological data in a consistent temporal and spatial framework enables a considerably enhanced regional understanding. This standardized dataset can then be used to aid a wide array of derivative interpretations, ranging from gross depositional environment maps to plate tectonic models.

Machine learning and assisted interpretation techniques can potentially provide unparalleled insights into complex geological systems. However, geological data is often associated with significant interpretations. For Big Data technologies to be applicable to traditional geoscience information, clean, consistent datasets that can be machine-read in a meaningful manner are essential.

The paper also describes how the use of strict data control techniques for chronological data and the combination with associated metadata enables the rapid construction of fundamental first-pass Source-to-Sink relationships. Source-to-Sink studies, an often-laborious undertaking, can require weeks to complete because of a large array of unstructured data from specialized disciplines. Consequently, these studies are not routinely performed within the typical hydrocarbon exploration workflow; they are primarily now within the domain of consultants and academics. Examples from the Gulf of Mexico and the Atlantic margins demonstrate how effective approaches can be integrated to reconstruct first-pass Source-to-Sink relationships within minutes, rather than weeks, while providing a robust and intuitive predictive framework that can be readily used by petroleum geologists to help inform exploration and reduce risk.

NOTES

Digitalization and Data in Field Development

Liz Wild

Shell

Digitalisation might be a hot topic but it certainly is not new; the energy industry has been relying on digital technologies for decades, from subsea robotics in the 1970s, to helping in the development of computer memory (Random Access Memory) to process seismic data in 1980s. Digital technologies have developed rapidly in recent years and the pace of change is ever increasing: decreasing costs, increasingly sophisticated analysis, increasing processing power, more power internet worldwide. This is disrupting and affecting both business (and personal lives) on a global scale.

Digitalisation in field development can offer a competitive edge in hydrocarbon recovery. It starts with employing advanced seismic data acquisition technologies like ocean bottom, broadband and 4D/timelapse seismic to capture information about the subsurface. Advanced processing and imaging capability is then needed to harness value from the vast amounts of information.

But in addition to good data, selecting the right field development concept also relies on adequate understanding and quantification of the ever-present subsurface uncertainties. Ultimately, the foundation of effective field development is the quality of joint decision making by the asset team, and sharing information across subsurface disciplines easily and effectively is essential. New digital platforms can facilitate this sharing of data and uncertainties across the modelling process, helping the team to select the best development plans and optimise recovery.

We can and should anticipate continued rapid advancement in digitalisation. Some trends we may see include advanced analytics like autonomous imaging and interpretation and/or modelling optimisation; and artificial Intelligence computing that can more quickly explore seismic data with the aim of making bigger discoveries.

Today, more than ever, technical and economic successes in exploration and production projects depend on safe, competitive, effective and robust field development and management. Through improved data quality and monitoring, we can build effective, integrated models across multiple teams, helping them to make better decisions, increase recovery rates and reduce field development costs.

NOTES

Big data? The power is in the analysis

Oliver Jordan¹ Adam McCarthy² and Paul Roberts¹

¹ Statoil UK, Paddington, London, olijo@statoil.com

² Statoil ASA, Stavanger Norway admc@statoil.com

It is clear that new approaches and technologies need to be leveraged to make further discoveries and unlock the potential of mature basins. As such, the analysis and integration of so called 'big data' is a central ingredient to future success. This approach would enable the explorer to increase the opportunities available, effectively screen and analyse these opportunities, and ultimately broaden the scope of the portfolio.

Whilst increasingly large datasets provide clear opportunities, there remain significant challenges for utilising these data effectively for both scientific research and industry application. Software limitations, data silos and even culture can cause difficulties in exposing and consuming data in a holistic manner. While these can sometimes be mitigated, the process can be laborious, and results in earth scientists spending more time developing work-arounds rather than focusing on solving subsurface problems. With data preparation absorbing the greatest percentage of a projects run-time, this issue must be addressed.

Currently, even modest advancements in computer power mean that large-scale regional evaluations of giant seismic and well datasets are now possible. Vintage and new data can be integrated, offsetting patch-work data coverage, encouraging cross-border access and challenging established views of play models. This approach has been undertaken by Statoil Exploration where efforts have been made to improve access and broaden the integration of numerous data types under a unified platform. This allowed prospect-scale interpretation within a broader 'One North Sea'-scale framework, facilitating a 'zoom-in, zoom-out' process to improve conceptual understanding of hydrocarbon plays.

However, the next generation of explorers require faster, smarter tools at their disposal, relying less on existing software functionality and more on intuitive, automated workflows to examine and identify areas of interest. To be successful, the link between subsurface domain expertise and data science must be significantly strengthened, but it also requires cultural change, embracing new ways of working. While ongoing activities in Statoil are exploring new technology (such as artificial intelligence), it is crucial to recognise that data alone is not the constraining factor, it's also the analytical tools available and the knowledge of our earth scientists to interpret the results. To develop and access these tools, we must welcome the approach of the data science community, empowering development through collaborative, cross-network solutions. Embracing these digital communities could prove vital for the earth scientist and give rise to the next generation of exploration.

NOTES

Virtual Glaciers and Glaciated Landscapes

Derek A McDougall

Institute of Science and the Environment, University of Worcester, Henwick Grove, Worcester UK – WR2 6AJ

d.mcdougall@worc.ac.uk

The topic of glaciers and glaciation is sometimes perceived by students and teachers as being more difficult than others associated with more familiar environments. Nevertheless, well-designed fieldwork in areas with glaciers and/or glaciated landscapes can make all the difference to their understanding and enjoyment of it. Unfortunately, fieldwork to these environments rarely takes place at the same time as students learn about them in class. Sometimes, fieldwork does not take place at all. In providing on-demand, simulated fieldwork to glaciers and glaciated landscapes, this resource provides a solution.

Virtual Glaciers and Glaciated Landscapes is based around high quality, 360° interactive panoramas (panospheres), which allow users to look all around and zoom in on features of interest. The panoramas are linked, so it is possible to proceed through the landscape, going from one panorama to the next. This provides a sense of scale and orientation, and helps users understand how the landscape changes from one location to the next. In due course, the panoramas will include sound (for contemporary glacial environments), further contributing to the immersive experience, as well as links to other resources, such as 3D models and video.

The locations for Phase 1 of the project are the Helvellyn Range, Keskadale and Mosedale (all in the English Lake District) and the Moiry, Ferpècle and Arolla valleys (Swiss Alps). Phase 2 will see additional locations, including Snowdonia (north Wales) and at least one more alpine location. All the panoramas are marked in *Google Maps*, and downloadable geolocation data allows integration with *Google Earth*, *ArcGIS* and other location-aware software.

The core virtual fieldwork resources are provided without interpretation, which means that they can be incorporated into the curriculum at any level. In Phase 2 of the project, there will be a separate, password-protected resource area with guidance on interpretation and use in the classroom, and a forum for teachers and academics. The virtual fieldwork approach is intrinsically a more interactive and engaging approach than simply presenting static images, and provides the basis for a range of learner activities and challenges. These include: landform and landscape interpretation exercises (applying and extending what has just been covered in a lecture, seminar or through reading); field sketching; 'field-checking' of geomorphological maps (produced using orthoimagery and elevation datasets in a GIS); student-produced geomorphological field guides (as a form of assessment); preparation for real fieldwork (whether or not to these specific locations); follow-up to fieldwork in these locations or when real fieldwork does not take place for some reason (e.g. poor weather, illness, budget constraints).

The resources can be viewed on any modern device with an internet connection, including phones and tablets, but they are best appreciated on a large monitor or by using virtual reality headsets. There is no requirement for any browser plug-ins or other proprietary software, although the browser must be up-to-date (HTML 5 compliant). In other words, the resources are easy to use and there is no need for expensive equipment.

This project remains a work in progress. Whilst the core resources are available to use now, development work will continue over the next 24 months.

ACKNOWLEDGEMENTS

I am most grateful to the following organisations, which have supported the development of this project: the Institute of Science and the Environment Learning and Teaching Fund (University of Worcester), the Quaternary Research Association, and the British Society for Geomorphology.

NOTES

Improving access to UKCS Petrotechnical Data – the next step in a 20 year journey

Daniel Brown

Manager, CDA Projects, Common Data Access Limited

dbrown@cda.com

To perform big data studies in the geosciences you need lots and lots of data. But where can you find enough, open geoscience data to enable problems to be tackled with the big data tool box?

For the last twenty years, CDA has been operating an industry-funded collaborative data store for offshore well and seismic data. As a not-for-profit company, CDA's focus has historically been to enable cost effective data sharing between its members, and to support their compliance in reporting that data to government.

But this world is changing.

The Wood Review identified access to data as a barrier to successful exploitation of the UK's oil and gas reserves, and since its formation, the Oil and Gas Authority has been pursuing an Information Management Strategy to facilitate access for everyone to oil and gas data – including the well and seismic data collected by petroleum licensees. Their recent consultation on the creation of a formal UK National Data Repository for petrotechnical data – built on the foundations laid by CDA – marks a major step in the realisation of this strategy.

Here, we share a brief history of collection of well and seismic data from the UK Continental Shelf, and the steps taken since the Wood Review to improve access to that data for academia and industry. These include the passage of the Energy Act 2016, which dedicates an entire chapter to improving petrotechnical data access; publication by the OGA through CDA of nearly 100Tb of government-funded seismic data under open conditions of use; and most recently, work to establish a formal UK National Data Repository dedicated to the stewardship and dissemination of petrotechnical information.

We also consider the next step in this journey – how the requirements of today's data scientists may be incorporated into new government requirements for the preservation, reporting, and publication of geotechnical information – enabling the most relevant types of geotechnical data to be stored, published, and accessed in a manner well suited to big data studies – and welcome contributions from meeting participants on how best to achieve this.

NOTES

IGCP 648 Efforts to Compile Structured Data for Palaeogeographic Studies: some lessons learned

Bruce Eglinton (*Department of Geological Sciences, University of Saskatchewan, Canada; bruce.eglington@usask.ca*)

Sally Pehrsson (*Geological Survey of Canada*)

David Huston (*Geoscience Australia*)

Dave Evans (*Yale University, USA*)



IGCP 648, an international project focused on better understanding the global supercontinent cycle and its links to mantle geodynamics, has developed several databases to facilitate the standardized compilation of relevant geoscience data to facilitate regional to global investigations. DateView provides the mechanism to compile geochronology and isotope geochemistry data while StratDB is focused on lithostratigraphy, ore deposits and large igneous provinces. Other databases relevant to the project include the GPMDB palaeomagnetic and PaleoBioDB systems. Lithogeochemical data are also used, drawn from project compilations and from online databases such as NAVDAT and GEOROC. Several palaeogeographic reconstruction models are also being developed; here we focus on the PalaeoPlates model which runs back to 1800 Ma.

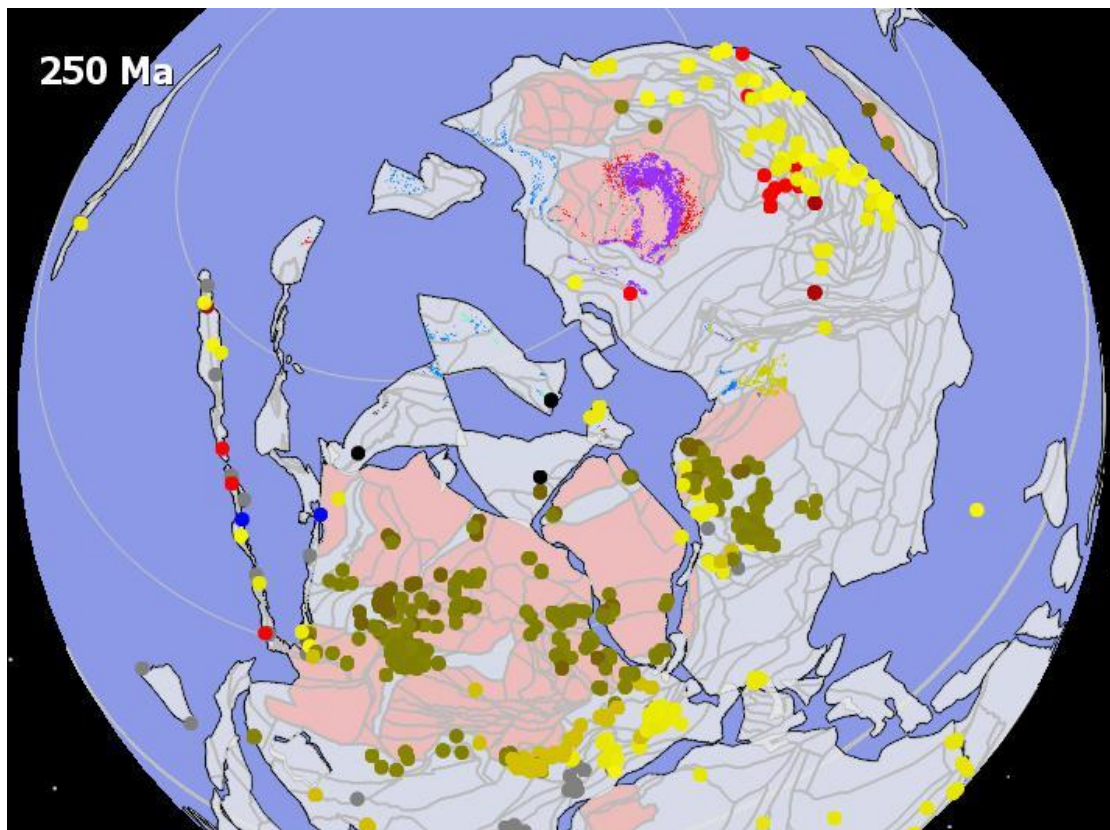
DateView currently holds more than 120 000 summary age and initial isotope composition records plus about 70 000 detrital zircon ages and Pb isotope composition data for more than 15 000 ore deposits. All data are geo-located so as to be used in GIS and palaeogeographic context. StratDB stores essential lithostratigraphic information such as unit name, rank, minimum and maximum numeric age, depositional setting for sediments, geodynamic setting and rock association. Ore deposits and large igneous provinces are similarly attributed with respect to age, geo-location and appropriate attributes. Many GIS geological maps are proprietary so the focus has been on ensuring that linking unique ID's exist between the IGCP 648 attribute databases and the maps rather than providing a map server interface.

Working with these datasets illustrates many of the limitations of routine big-data applications when applied without appropriate geoscience expertise or where data are not well organized and attributed. Several aspects of current approaches to geological map production severely limit rapid, semi-automatic processing of information and linking of point attribute data to polygon or polyline information. One of the unique aspects of most geological assessments is that, in addition to multi-dimensional attribute data, one also needs to cater for changes in age (time) and place (geolocation), often also while taking into account significant imprecision. Many geological maps, especially in the Precambrian, attribute polygons very broadly with respect to age and lump mapped areas together based on lithology (e.g. Mesoproterozoic granitoids), rather than by lithostratigraphy (Suite A, Suite B, etc), even though some form of systematic categorization is normally applied by the original geological mapping teams.

Effective investigation of regional to global crustal evolution and its association with mineralization requires changes to our approach to compiling and releasing geological information. We need more numerical attributes, which can be more precisely quantified, rather than broad-brush categorization. The ability to drill down into more detailed (smaller scale) data sets needs to be developed. Targeting for mineral exploration also needs to recognize that not all information is relevant at all scales. Requirements for greenfields exploration is often very different to that for detailed prospect evaluation and workflow solutions need to address these different paradigms.

Ultimately, while better processing speeds, improved storage mechanism for large datasets and easier access to some data will benefit some industry and researchers, the current paradigm in which most companies and individuals still store data in file and project folders is not going to help us find more deeply buried deposits. Nor is a system in which many academic

publications and company submissions to geological surveys are provided as PDF documents. In many cases, geoscience data are still published without numeric geographic locality information. Even more fundamentally, if we do not marry good geological expertise with good IT strategies, then the promise of, and the ability to selectively utilize information relevant to earth science investigations, will not be successful. Several examples illustrate some of the advantages of structuring data to meet the requirements of geoscientists and of pitfalls associated with inappropriate use of machine-learning without proper geoscience insight.



Merging of multiple geoscience datasets such as geological maps, geochronology, geophysics in palaeogeographic context, based on attribute data from the IGCP 648 database systems. When data are effectively structured, such merging need take no more than a few minutes.

NOTES

Palaeogeography and Big Data

Paul Markwick^{1,2,3}

1. *Knowing Earth Limited.*

2. *Visiting Lecturer, Basin Structure Group, University of Leeds*

3. *Visiting Research Fellow, School of Geographical Sciences, University of Bristol*

paul.markwick@knowing.earth

'Big Data' in geology, especially in frontier exploration, can be powerful. But, the value of compiling, managing and interrogating large, diverse datasets is only as great as the ability to access that data easily, the quality of the data and, most importantly, the questions asked of that data. This requires a means of representing and visualizing the diversity and breadth of information available to better understand it. This is not a new problem.

In the 19th century databases were physical. Collections of specimens from around the world formed the first museums and zoological gardens, and then, in book form the world's libraries. Even in the 19th century, the wealth of knowledge meant that no one individual could know everything and expertise was increasingly restricted to fields of study. This was all the more problematic for geology, which brought together elements from all the sciences and then added the complication of time. Printed volumes such as Humboldt's *Cosmos*, Lyell's *Principles*, or Lapparent's *Traite*, brought together all that was known within geology, and by the 1870s Reclus, in France, was showing global maps of the distribution of mountains and volcanoes and other geomorphological patterns that today we would readily understand. This was "big data" 19th century style. But, by the 1870s it was clear that there was too much information to be able to fully interrogate and understand the significance of what geologists were seeing.

One solution to the vastness of the growing geological record, was found in palaeogeography, the representation of the past geography of the Earth. Within palaeogeography, geologists could place all their observations, from tectonics to deposition to fossils and palaeoclimate, and through this see spatially and temporally how these phenomena changed and interacted. Land-sea maps had been around since at least Lyell, although the term "paleogeography" did not appear until Hunt in 1873. The first global maps were those of Lapparent in 1900. But, it was not until Schuchert in 1910, and his extensive North American palaeogeography atlas, that the full power of palaeogeographies, including applications in oil and gas exploration, became more widely promoted. It was also Schuchert who expanded the basic workflow of Hunt for building maps, which incorporated the full breadth and diversity of geological information, from the structural framework to depositional environments, palaeodrainage and palaeotopography.

Palaeogeography is still a crucial, and largely under-utilized tool for geologists and explorationists today. It provides a means of representing, visualizing and interrogating the much larger and more diverse geological datasets of the 21st century, with direct applications to exploration, climate change, biodiversity and environmental investigations.

This talk provides an overview of the workflow used for building and utilizing palaeogeography maps to better understand the Earth system using large and diverse geological datasets. This is largely through Geographic Information Systems (GIS). As part of this we will also look at the issues of data confidence and provenance, data management, as well the inherent problems of the geological record itself.

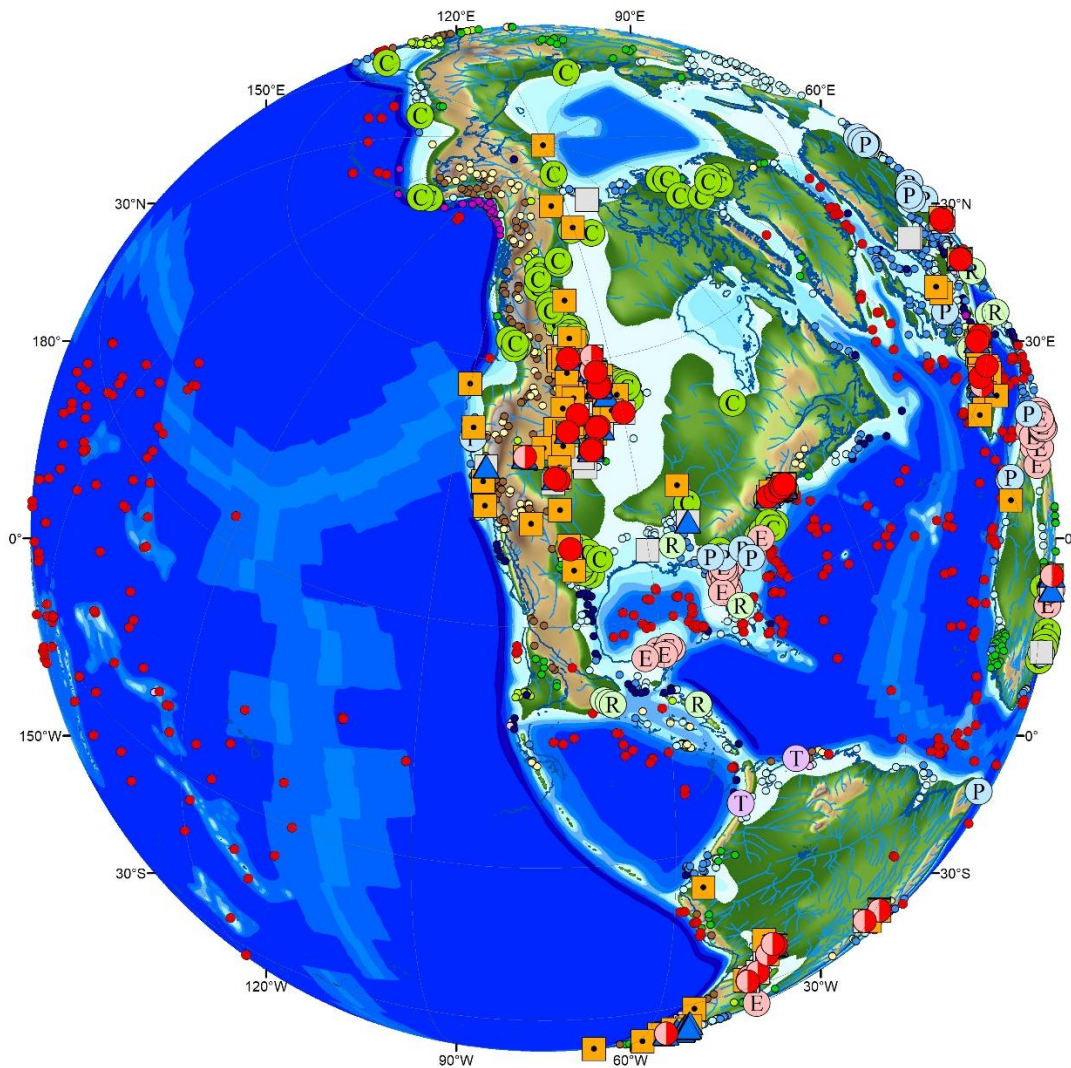


Figure 1. Maastrichtian palaeogeography built through a workflow that takes and integrates data from a diversity of fields from mantle processes and crustal architecture (structural framework and crustal type), through plate modelling to the reconstruction of depositional environments and stratigraphy, to drainage reconstruction (for source to sink analysis) and palaeotopography and palaeobathymetry. The results then form the boundary conditions of Earth Systems models and through these models of lithofacies prediction (retrodiction). The data shown in this example includes climate proxies, fossil vertebrates and well and outcrop data. Compiling and managing this data is a huge task made tractable by the visual simplicity of palaeogeographic maps, which can then be used to ask further, more advanced questions of the Earth.

NOTES

Data Mining and Visualization of Detrital Zircon Data: Assessment of Palaeogeographic and Geodynamic Setting Using Data from Laurentia

Dean Meek (*Department of Geological Sciences, University of Saskatchewan;*
dmm447@mail.usask.ca)

Bruce Eglington (*Department of Geological Sciences, University of Saskatchewan*)

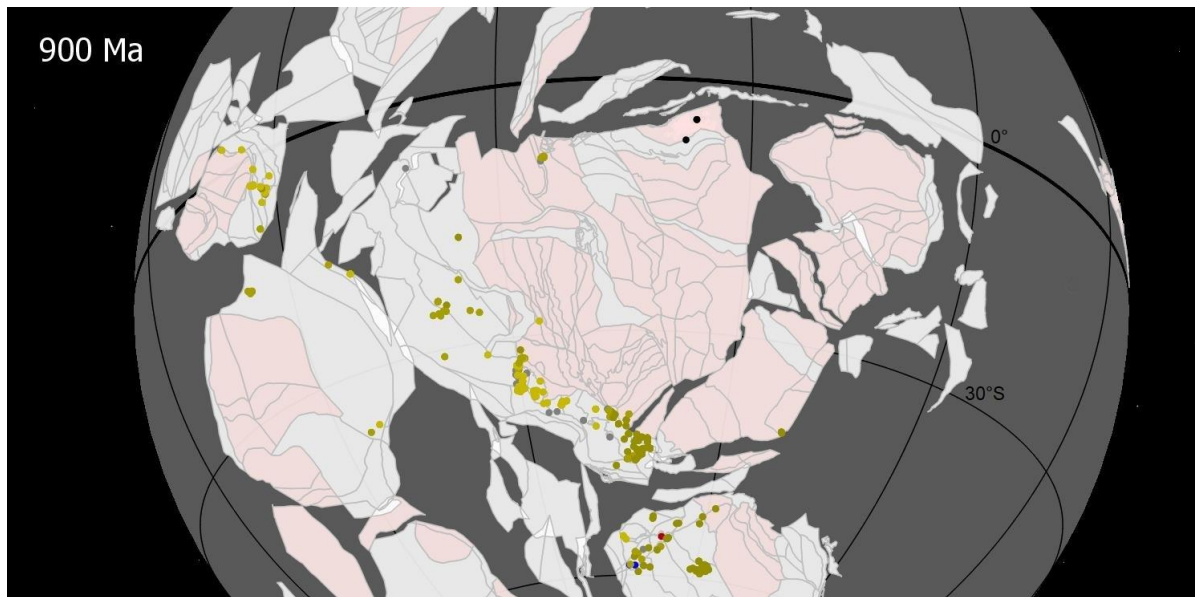
Compilation of published detrital zircon data for the western and northern margins of Laurentia are combined with GIS map information, palaeogeographic models and igneous and metamorphic geochronology to facilitate a better understanding of the crustal evolution of the region. More than 33,000 U-Th-Pb analyses and over 3,000 Lu-Hf analyses of detrital zircon grains were compiled from 36 publications. More than 69,000 igneous, metamorphic and cooling ages, captured in the online DateView database, were also considered, together with initial Nd isotope compositions for a variety of North American whole-rock samples. GIS information from maps published by 7 geological surveys have been attributed using standardized values. All data are presented in age-appropriate palaeogeographic context using the PalaeoPlates reconstruction model and GPLates or Paleogis software as a data mining exercise.

Extensive use of databases has been an essential aspect of this exercise, both as a source of data and as a tool to facilitate conversions and addition of attribute information. Geological map information was modified to express age in terms of maximum and minimum numeric age rather than timescale periods. Default values from the geological timescale have been used unless more precise values are available from more detailed studies. Sedimentary depositional environment was simplified to a maximum of 12 categories.

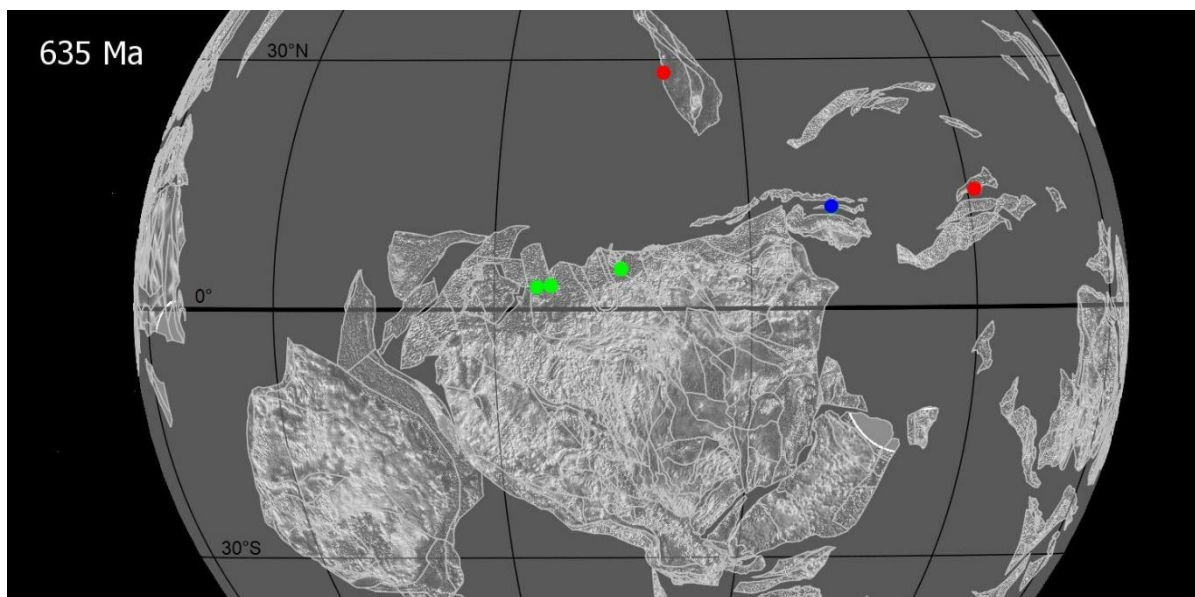
Detrital zircon data were captured with estimated deposition age, detrital grain age and, where available, Th/U ratio and initial Hf isotope composition. At this stage, most detrital zircon studies lack Hf isotope data (fewer than 10% have data) and the situation for igneous rocks which might be considered as provenance sources is much worse. Sm-Nd isotope data are available and we use the good correlation of Nd and Hf initial compositions for terrestrial rocks (Vervoort et al, 2011) to facilitate regional comparisons and assessment.

The detrital zircon data are presented on plots of grain age relative to deposition age with probability distributions represented by heat colours, a new technique to visualize large detrital datasets. Matching igneous and metamorphic ages were extracted from DateView for all important peaks in the probability distributions for each detrital sample and combined in plate reconstruction software to visualize the potential provenance regions in their palaeogeographic context. These potential provenance sites were further refined based on initial hafnium isotope composition. Each detrital sample was also assigned an inferred geodynamic setting based on the concepts proposed by Cawood et al. (2012). Sedimentary depositional environments, based on geological map polygon information, were added to the palaeogeographic context to further illustrate the regional geological history.

Earth science data and their visualization frequently need to capture changes in both time and space and increasingly need to process larger datasets. This exercise demonstrates many of the advantages of using structured data systems for working with these larger earth science data compilations while also making data more easily available for other researchers in future. It has been limited to the visualization of data but future investigations will draw on multivariate and machine learning techniques to achieve more quantitative assessments.



Potential provenance localities (yellow to brown symbols) for two samples (black symbols) deposited in the north west of Laurentia at about 900 Ma. Palaeogeography and provenance constraints require sediment transport across Laurentia, as previously suggested by Rainbird et al. (1997).



Inferred geodynamic setting for sediments deposited in various localities along and adjacent to western Laurentia at about 635 Ma. Green = passive margin; red = accretionary and blue = collisional. Classification based on the methodology of Cawood et al. (2012).

NOTES

The worldwide field course: use of 3D outcrop imagery in training

Gary Nichols

RPS Training, Goldvale House, Church Street West, Woking , GU21 6DH

Outcrop images that can be viewed in 3D can be seen as a way of creating a 'virtual field course' allowing people to view geological features in an area without going there. Providing training for petroleum geoscientists in this way may appeal on grounds of cost, but there may be alternative approaches. Professional training courses are generally focussed on a particular topic and not on the geology of an area: locations are chosen as the best that area can offer as a means of illustrating geological features. To see all the best examples in outcrop, locations from a variety of places would ideally be chosen, but by their nature field courses are geographically restricted.

The most effective use of 3D outcrop images in professional training is to integrate them into both classroom and field courses. In the classroom, 3D images that best illustrate the subject matter may be selected from around the world to provide course participants with something better than a simple photograph of geological features. On field courses the imagery can be used to provide a different perspective of the locations visited, and bring in examples from other areas which complement the examples seen in the field. The 3D data is very effective at adding value to training and providing additional case studies as analogues.

A 'virtual field trip' to a specific location would be limited by the scope of outcrops in that area, but would not offer the benefits of actually being in the field – for example the opportunity to view the geology at all scales, from the single grain up to the overall basin setting. Outcrop images in 3D provide visualisation only of the intermediate scale, and these are most effectively used to enhance a training course and to provide a range of examples from around the world as illustration and examples.

NOTES

When Failure (a lot of failure) Becomes an Option - Machine and Deep Learning on Seismic Data

John Thurmond¹, Jens Grimsgaard², Daniel Barker³, Eskil Dahl², Line Eikvil³, Guy Robert Hall⁴, Ragnar Hauge³, Marit Holden³, Vegard Sangolt⁵, Arnt-Børre Salberg³, Eirik Time⁶, David Wade⁷, Anders Waldeland³

1. Statoil Research & Technology, Houston, TX, USA
2. Statoil Research & Technology, Bergen, Norway
3. Norsk Regnesentral, Oslo, Norway
4. Statoil Development and Production Norway, Bergen, Norway
5. Statoil Research & Technology, Oslo, Norway
6. Statoil Research & Technology
7. Statoil Software Innovation Stavanger, Norway

Automated recognition of patterns in data (Machine Learning) and automated classification and segmentation (Deep Learning) have recently become an area of significant research advances; this is partially due to improved algorithms, and partially due to increased available computing power. Specifically, Deep Learning has recently achieved super-human levels at suitable problems such as the classification of images. Despite the presently over-hyped state of these technologies, they are certain to have significant impact on many aspects of work in the geosciences. The interpretation of seismic data has long been approached from both the direction of quantitative analysis of attributes (e.g. acoustic impedance inversions), and via strong reliance on seismic interpreters who are skilled and have great experience in recognizing patterns they have seen before. The best seismic interpretations include both elements, and both approaches can be used to research ways to accelerate seismic interpretation – a goal which would save industry many thousands of man-hours of work and could relieve much of the tedium of seismic interpretation. Machine Learning techniques are highly suitable for analysis of de-spatialized attributes generated from seismic, also benefitting from advances in attribute generation from various sources. Once a clustering is rendered back into its spatial context, it still requires an interpreter to make sense of the patterns revealed. In contrast, Deep Learning allows patterns in seismic data to be recognized, but still requires an interpreter to put these patterns into context. Deep Learning also requires an enormous library of examples to draw from (much like a good human interpreter), and our catalog of interpretations are not necessarily fully suitable for training these algorithms. As such, we are far away from anything resembling automated interpretation, but these tools should allow us to make human interpretation both more efficient and more effective. We have applied these approaches on a wide array of seismic challenges, from *ab initio* interpretation of regional seismic volumes for exploration through to extremely detailed analysis of mature fields in production. While our work is in an early phase, it is already yielding results; it is possible for Deep Learning algorithms to make seismic interpretations, and Machine Learning can quickly reveal interpretable patterns that previously were elusive or labor-intensive to generate. As these techniques advance, it will be necessary to determine how we can best utilize our vast arrays of previously collected seismic data and interpretations to continuously improve algorithms that will aid us in interpreting seismic – a very large “big data” challenge.

NOTES

Big data - A boundaryless future?

Rhian Burrell, Charles Jones, Joseph Nicholson and James Selvage (Osokey)
rhian@osokey.com



To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science. ~ Albert Einstein

This is a fantastic idea but how do we achieve it in reality? Big data may be a key piece of the puzzle.

As an industry, we have huge volumes of data (both live on our computer systems and archived on tapes in warehouses). But how do we access it all effectively and efficiently?

Imagine if we could access all our global data in one place, new connections could be made between data sets and observations (Figure 1). Advances in technology now mean that this could be a reality. All data could be digitally stored whilst maintaining easy access. By applying cloud-native technology, we can remove the challenge of handling big data, instead enabling access and analysis to drive further scientific advancement and innovation.

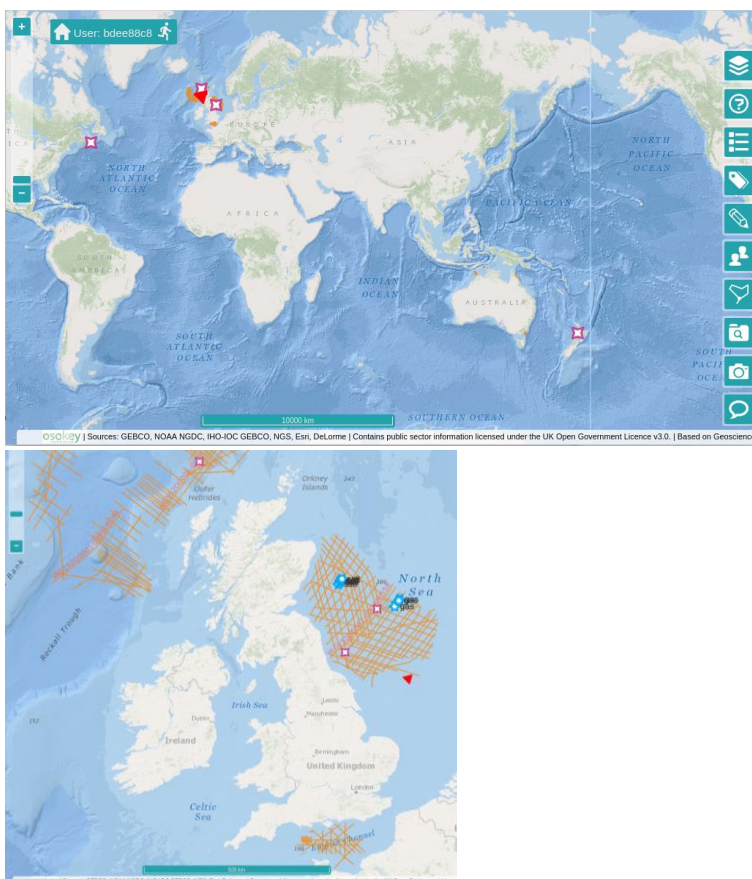


Figure 1 - Example of accessing a global data repository in one place using cloud-native technology [1]

The other key piece of the puzzle is the geoscientists. They are at the heart of the issue, turning data into value. As individuals we provide analysis and insights into the data, yet

imagine what we could do as a global community. By breaking down siloes and removing barriers, we would have the potential to share and build on each others' ideas providing alternative perspectives. Old problems can be viewed from a new angle or entirely new concepts may be developed. By sharing these ideas, observations and knowledge, a global digital analogue database could be created. Observations, questions and interpretations could be digitally captured and directly tagged to the underlying data creating a global industry knowledge base. Building connections between data sets and analogues will unlock the hidden potential within the data. Not only would this add significant business value but it would also generate an exceptional educational resource.

Further connections can be built between industry and academia with a centralised industry repository acting as a test environment for algorithm development and consistent basin-wide computations. Academic research can be deployed quickly and effectively into the industry strengthening the connection between R&D and industry challenges.

This paper will explore the challenges described and share examples of where digital technology could address these problems. The emphasis of this paper is on starting afresh with big data, developing cloud-native solutions rather than migrating legacy systems and processes. The future roles academia, industry and national data repositories could take in a future big data ecosystem will also be discussed.

References

[1] Contains public sector information licensed under the UK Open Government Licence v3.0.

NOTES

How machine learning systems can extract more qualified information from seismic acquisition and processing reports.

Henri Blondelle

AgileDD

Background and purpose of the lecture:

This lecture is derived from a pilot done at TOTAL E&P to apply machine learning systems for extracting qualified meta-data from unstructured documents related to seismic acquisition surveys.

Lecture summary

Major IOCs store some large volume of seismic data and may access additional large volumes from seismic multi-client vendors or National Data Repositories. Accessing the seismic information using geographical criteria is no more a problem thanks to the implementation of GIS tools. But, accessing quickly the right seismic data with specific sub-surface objectives as criteria remains an important challenge. The reason why? Mainly because not enough metadata have been associated with the seismic data in the past and the quality of the existing metadata is not well estimated. Current corporate seismic databases suffer several limitations:

- Stored information and information sources (reports) are disconnected. The available information is rarely sourced.

- The information quality is difficult to measure, not stored with the information and sometimes not at the expected level.
- It is impossible for a user to select a seismic using some criteria not taken into account in the seismic database design. An adaptation of the data-model will force to re-access a lot of raw documents to populate the new attribute.
- The shortage of qualified E&P data managers makes difficult new database population or database QC campaigns.

Since machine learning systems have proved recently their ability to extract information from large volumes of unstructured data in various part of our industry, a pilot has been implemented at TOTAL E&P with the support of AgileDD to evaluate the capacity of a machine learning to extract information automatically from seismic acquisition related documents. The long-term ambition is to automatize the E&P document indexing in order to create more index and to qualify more rapidly the existing data.

During the pilot study, a particular attention has been paid to the machine learning curve in order to check the possibility to establish effective learning models with only a small amount of seed information. The learning curve has been observed for a collection of 40 metadata of various types. This paper related the lessons learned from this experiment and opens perspectives for a more efficient use of E&P legacy data.

Lesson learned: What would the lecturer want the audience to be left with and take with them from this speech:

- It is possible to automatize the indexing of very technical unstructured documents such as acquisition surveys reports, seismic QAQC reports or navigation reports,
- In addition of populating rapidly some data-bases ,the machine learning systems are able to source the extracted information and measure its confidence level.
- The same systems can be applied to any technical documents (eg: well reports, well logs), only the learning models will differ.

At a time where it is required to extract more from legacy data, machine learning systems are becoming an important thumbs in the hands of E&P data-managers.

NOTES

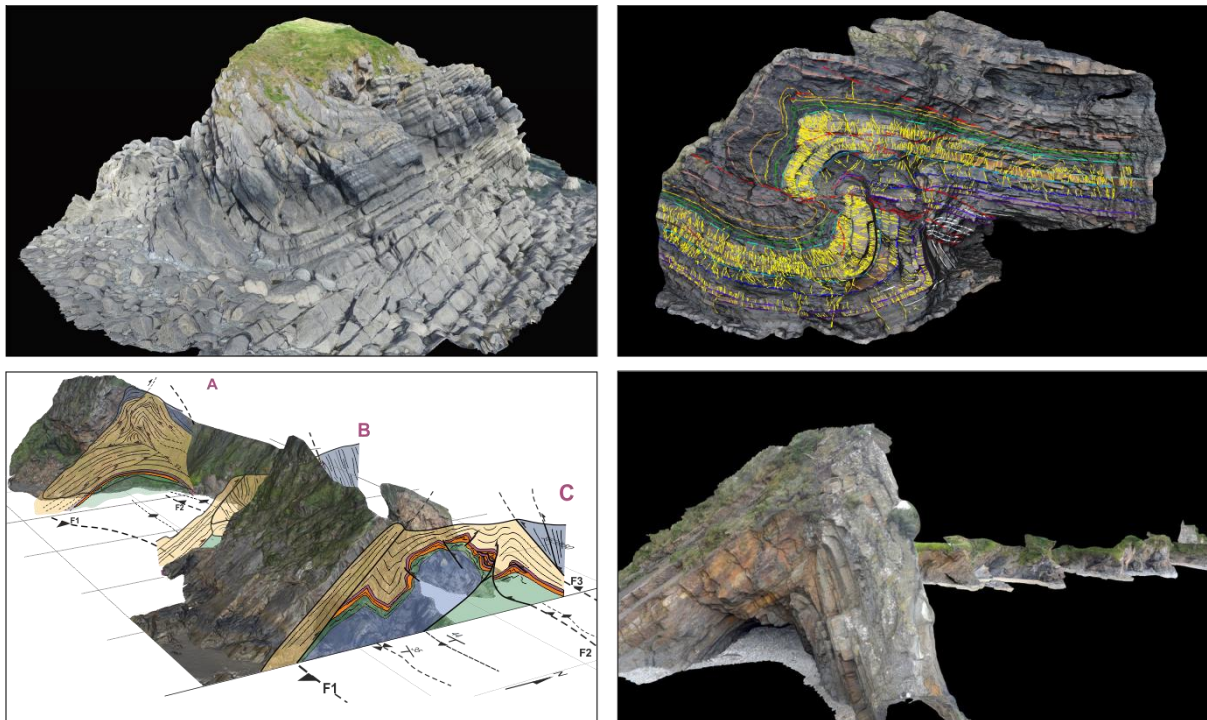
VIRTUAL FIELDTRIP: Structural style through the stratigraphy of Variscan Pembrokeshire: a virtual field trip

Adam J. Cawood

Department of Geology and Petroleum Geology, School of Geosciences, University of Aberdeen, Meston Building, Kings College, Aberdeen, AB24 3UE

The coastal sections of Pembrokeshire provide excellent exposure of various fold-thrust structures of Variscan origin. Deformed units are of Lower Palaeozoic (pre-rift) to Carboniferous (post-rift) age, and record a variety of distinct structural styles and deformation mechanisms. Structural styles record variable mechanical properties through the stratigraphy, imparted by a range of factors (e.g. depth of burial, diagenetic history, thickness/composition and fault-controlled facies distribution). The classic outcrops of West Angle, St Ann's head, Stackpole Quay, Broadhaven, Monkstone Point and Saundersfoot expose a spectacular series of examples which highlight these distinct structural styles in units deposited over c. 250 Mya.

These six outcrops are presented in digital form as part of a Variscan Pembrokeshire virtual field trip. Using these virtual outcrops as a series of case studies, we highlight changes in structural style through the stratigraphy of Pembrokeshire. We outline a number of approaches



for structural data collection and interpretation using virtual outcrops and present workflows which allow for the quantification of variability in deformation mechanisms and structural style.

NOTES

VIRTUAL FIELDTRIP: A virtual field trip to seismic-scale outcrops of the Triassic, Edgeøya, Svalbard

Simon J. Buckley^{1,2}, Kari Ringdal¹, Benjamin Dolva¹, Isabelle Lecomte², Ingrid M. Anell³, Alvar Braathen³

¹ Uni Research CIPR, Bergen, Norway

² Department of Earth Science, University of Bergen, Norway

³ Department of Geosciences, University of Oslo, Norway

Large-scale cliff sections on Edgeøya, an island in the eastern part of the Svalbard Archipelago, comprise mostly Triassic sedimentary successions, and are onshore outcrops of formations found on the offshore Barents Shelf. These outcrops are of high interest for studying the subsurface geology found offshore on the Barents Shelf, using seismic-scale features such as small-scale growth faults, paralic deposits and low-angle clinoforms (Anell et al., 2016). At Kvalpynten, an outcrop at the southern end of Edgeøya, these features can be observed, providing the opportunity to compare offshore seismic data with the outcrop expression using seismic modelling. Synthetic seismic modelling provides a means of aiding interpretation of seismic data by quantitatively and qualitatively assessing resolution, scale and level of detail of modelled onshore structures (Lecomte et al., 2015). This can give valuable insights into potential resolution and survey parameter issues, as well as limitations with seismic data.

Developments in digital outcrop modelling techniques, such as lidar and photogrammetry, now allow the accurate and high-resolution capture of large and inaccessible outcrop sections, at scales comparable with seismic vertical resolution. This opens up new possibilities for generating georeferenced input to synthetic seismic modelling, which allows generation of highly visual end products for educational purposes. This contribution presents a virtual field trip to the Triassic outcrops at Kvalpynten. The dataset is based on virtual outcrop models collected using boat-based photogrammetry, covering over 10km of continuous cliff sections. The 3D outcrop model is combined with geological interpretations, logs, field photos and information panels, as well as regional elevation data to set the context of the trip. In addition, synthetic seismic data are fused with the virtual outcrop models to highlight the scale of the exposures compared to offshore seismic data. The combined dataset is explored in LIME, a high performance viewer designed for displaying disparate geospatial and field datasets (<http://virtualoutcrop.com/lime>).

Acknowledgements: The authors acknowledge an internal grant award from Uni Research CIPR for developments made to the LIME software during this research project. Aspects of this work have been funded by the Research Council of Norway and Tullow Oil Norge, Lundin Norway, Statoil Petroleum, Edison Norge and Dea Norge through the Petromaks 2 programme (Trias North, project number 234152).

References:

- Anell, I., Lecomte, I., Braathen, A., and Buckley, S.J., 2016. Synthetic seismic illumination of small-scale growth faults, paralic deposits and low-angle clinoforms: A case study of the Triassic successions on Edgeøya, NW Barents Shelf. *Marine and Petroleum Geology*, 77: 625-639. [doi:10.1016/j.marpetgeo.2016.07.005](https://doi.org/10.1016/j.marpetgeo.2016.07.005).
- Lecomte, I., Lavadera, P.L., Anell, I.M., Buckley, S.J., Schmid, D.W. & Heeremans, M., 2015. Ray-based seismic modeling of geologic models: Understanding and analyzing seismic images efficiently. *Interpretation*, 3(4): SAC71-SAC89.

NOTES

VIRTUAL FIELDTRIP: The use of Virtual Outcrop Models, digital geology and legacy data to reappraise Devonian basin evolution in NE Scotland and Shetland.

Thomas A.G. Utley¹, Robert E. Holdsworth^{1,3}, Ken J.W. McCaffrey^{1,3}, Edward D. Dempsey², Rob A. Strachan³ and Richard R. Jones⁴.

thomas.a.utley@durham.ac.uk

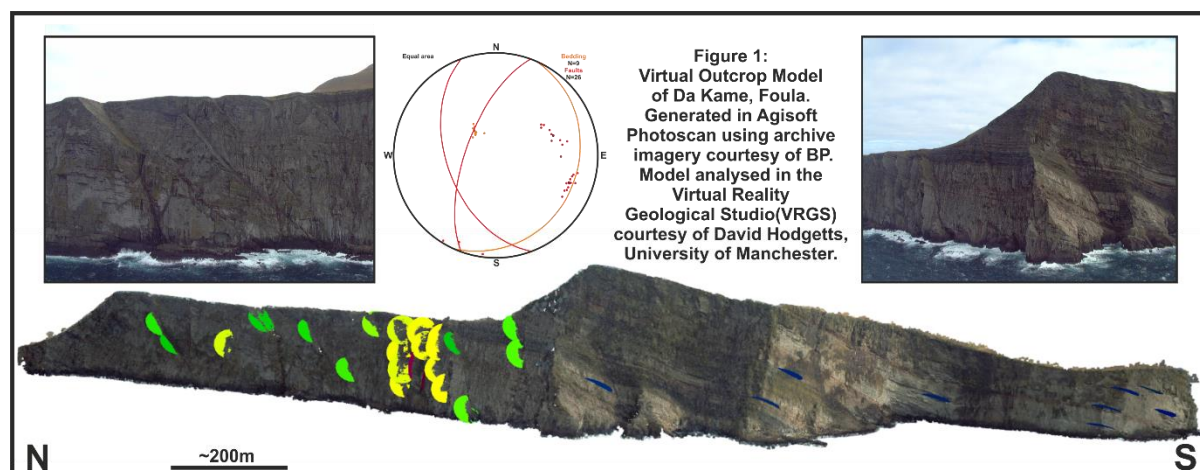
¹ Department of Earth Sciences, Durham University, Science Labs, Durham, DH1 3LE, UK

² School of Environmental Sciences, University of Hull, Hull, HU6 7RX

³ School of Earth and Environmental Sciences, Burnaby Building, Burnaby Road, University of Portsmouth, PO1 3QL, UK

⁴ Geospatial Research Limited, Office Suite 7, Harrison House, 1 Hawthorn Terrace, Durham, DH1 4EL, UK

The extraction of 3D geological information from virtual outcrop models created using drone-based photogrammetry allows for the collection of useful and valuable geological data, quickly and relatively easily, from areas previously inaccessible, overlooked and considered to be 'missing' or 'lost' (See Fig.1). Outcrops can be now re-visited at any time and further geological analysis carried out, producing data and knowledge which can be now be easily shared and viewed in 3D. These new digital geological methodologies are not a replacement for traditional fieldwork, but should be seen as the 'norm' and used to supplement and complement traditional fieldwork, desk-based studies and geological data analysis.



Use of these models, in conjunction with the re-interpretation and re-examination of large quantities of onshore and offshore legacy data generated by industry and academia is a more cost effective and time efficient approach, avoiding the duplication of datasets. Vast quantities of data can now be collated, interrogated and analysed in new ways, deriving new value, and highlighting trends and features hidden away within solitary datasets and in archives.

We illustrate this using a reappraisal of the stratigraphy, structure and tectonic evolution of onshore analogues for the Devonian-Carboniferous Clair Basin that includes detailed analysis of basement/cover contacts, and the structure of the overlying Devonian sequences. This is achieved through mixture of desk-based study, reappraisal and analysis of legacy and new onshore and offshore datasets, fieldwork, structural analysis and the production of Virtual Outcrop Models.

Our new synthesis extends from Shetland, Orkney and Caithness and consistently reveals synchronous faulting and synformal growth folding in all areas. These observations are

consistent with models of constrictional extension during regional sinistral transtentional Devonian basin development. This unusual basin architecture together with the diversity of basement/cover relationships apparent in the Orcadian Basin could lead to difficulties in exploration and appraisal of resources offshore. Thus established models for Devonian basin development used offshore may require revision.

NOTES

Facilitating the effective application of analogue databases to reservoir models

Marijn van Cappelle^{1,*}, Simon Lomas², Luca Colombera³, Marco Patacci³, Alwin ten Hove¹, Sarah Cobain², Ben Meyer² & Viki O'Connor²

¹ *Petrotechnical Data Systems B.V., Lange Kleiweg 10, 2288 GK, Rijswijk, The Netherlands*

² *Petrotechnical Data Systems Ltd, Chester House, 81-83 Fulham High Street, London SW6 3JA, United Kingdom*

³ *School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, United Kingdom*

*contact: marijn.vancappelle@pds.nl

In the petroleum industry, reservoir models of oil and gas fields are built in order to estimate hydrocarbon volumes in place and forecast hydrocarbon production. Subsurface datasets used for conditioning models consist mainly of 1D well data: inferring the three-dimensional architecture of the inter-well volume involves significant uncertainty. It is widely acknowledged that applying rock outcrop analogues to reservoir models helps reducing uncertainty through the inclusion of geological realism.

In recent years the compilation of many analogue databases has been undertaken. Often these databases are constructed, maintained and used within petroleum companies. In other cases, the databases are vehicles for transferring data obtained in the context of academic research to sponsoring industry consortia. However, integrating data from these databases into workflows in development and production teams has proven difficult. Hurdles preventing geoscientists querying and applying data include: 1) the wide range in training and experience of staff, 2) time constraints, 3) constraints in the number of reservoir models which can be built and run, 4) practical limitations in the detail/resolution of reservoir models, and 5) the lack of quantification for all the parameters used in reservoir modelling.

We present a method (Ava Clastics) designed to overcome these hurdles of effectively integrating outcrop analogue data in reservoir modelling workflows. An online portal has been developed that makes it possible to quickly filter relevant outcrop analogues contained in relational databases of fluvial (FAKTS), shallow-marine (SMAKS) and deep-marine (DMAKS) clastic depositional systems. These relational databases store data on sedimentary units at multiple scales, describing their geometries, their relations with surrounding elements, their hierarchical organization, and their lithological heterogeneity. All the analogues contained in the databases are classified on the geological controls of the depositional systems, contextual information, and metadata, allowing users to select analogues that are most suitable in application to the reservoir at hand. The filtered analogue data is visualised graphically in real time, to enable the user to review database outputs and optimize the selection of inputs to a reservoir model.

Subsequently, the filtered analogue data are automatically parameterized to match the requirements of different stochastic facies modelling algorithms, thanks to workflows for translation of analogue data into model inputs. Where available, the analogue data is directly used as parameters in the modelling algorithm. When parameters are not directly available from suitable analogues, empirical relationships that are established in the scientific literature are used to calculate input parameters. After review, the parameterized composite analogue can be directly imported into the reservoir modelling package.

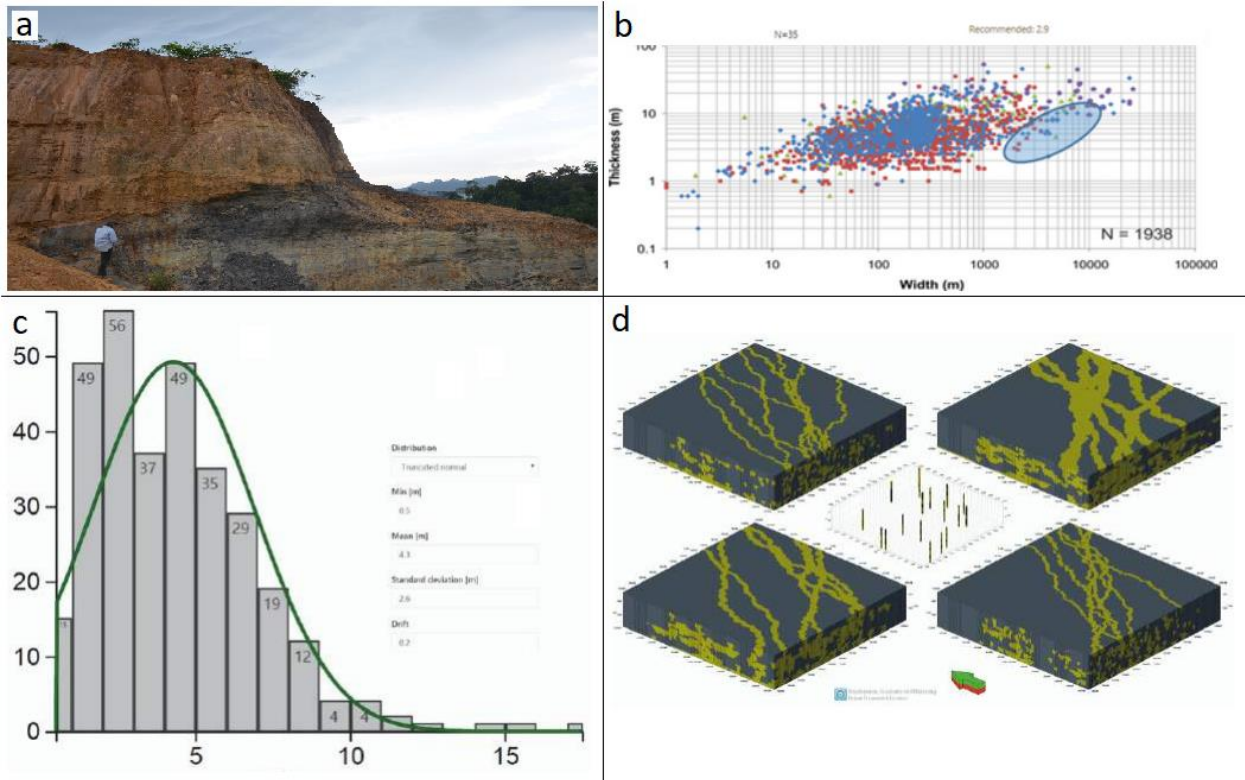


Figure. a) Rock outcrop analogues help in reducing uncertainty in inferring rock properties in between well. b) Analogue data is stored in databases. c) Parametrization of analogue data for use in stochastic facies modelling. d) Application of the parametrised analogue data to reservoir models.

NOTES

Machine Learning Assisted Petroleum Geoscience: Can a computer learn to map stratigraphic architecture and reservoir quality by training on data?

Eirik Larsen^{1*}, Behzad Alaei¹, Dimitrios Oikonomou¹, Christopher A-L. Jackson^{1,2}, Idar A. Kjølraug³, Kristian Helle³, Ryo Sakamoto³

¹Earth Science Analytics AS, Prof. Olav Hanssensvei 7A, 4021 Stavanger, Norway, ²Basins Research Group (BRG), Department of Earth Science and Engineering, Imperial College, London, SW7 2BP, UK, ³Moeco Oil & Gas Norge AS, Haakon VII's gate 1, 0161 Oslo.

*eirik.larsen@earthanalytics.no



Petroleum geoscience is hard, particularly when it comes to predicting key physical properties away from known measurements. It is hard because it is so complex. It is hard because there are no simple rules, like Newton's laws of motion, that can help us predict the spatial distribution of, for instance, reservoir properties, or where to look for the next big commercial discovery. It is basically hard to "codify" and "formalize" what we do routinely as petroleum geoscientists.

Until today we have attacked these kinds of "hard-to-codify" problems by assigning teams of human experts to solve them. The combined experience of these experts, who typically come from multiple disciplines, helps us extract knowledge and insights from the available data. What if we could replicate this method with computers? Can we have the computer learn relationships directly from the data, from all relevant sources? This is exactly what machine learning is for. In this talk we will explore examples of how increasing numbers of researchers and industry practitioners are currently applying artificial intelligence to solve petroleum geoscience problems.

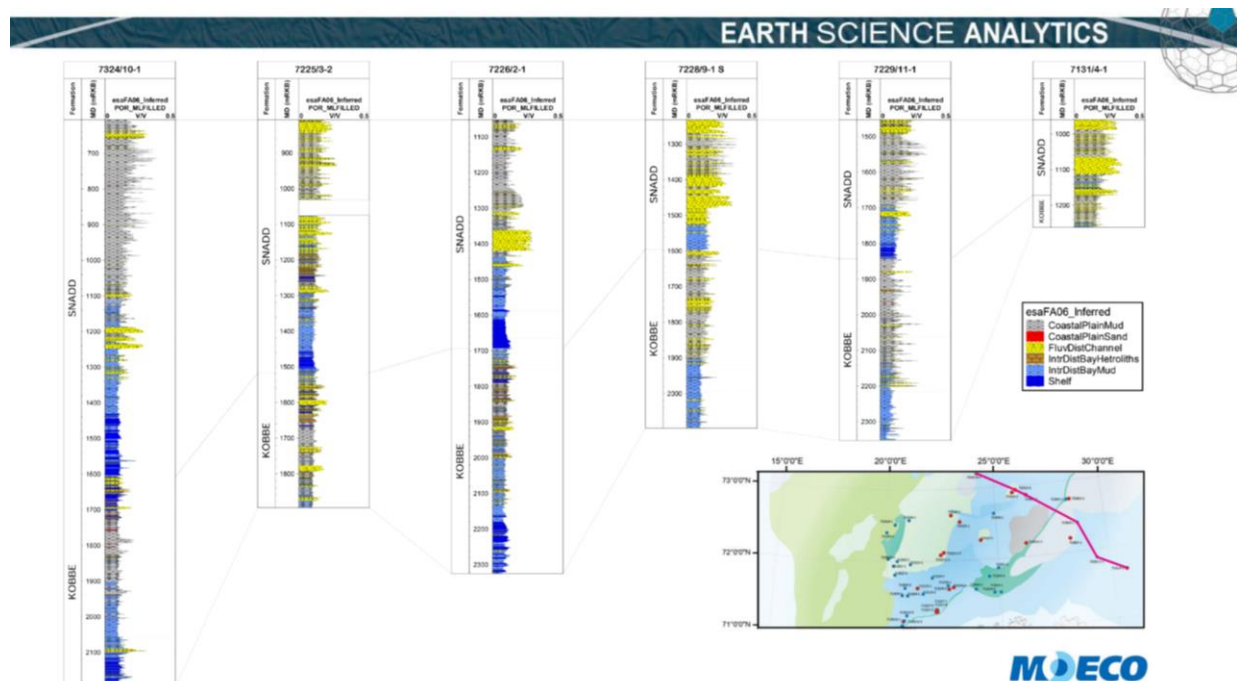


Fig. 1. Machine-learning derived porosity logs coloured by machine-learning derived facies associations.

The incredibly rich subsurface data and metadata available in national data repositories such as CDA and Diskos are perfect resources for training machine learning models. Machine-learning models work remarkably well when large, structured and labelled data sets are available for training. We will soon be able to use this technology to build incredibly detailed,

high-dimensional models using all our data. When machine-learning models are trained on smaller data sets they enable petroleum geoscientists to better understand the spatial distribution of reservoir properties and hydrocarbons. This technology is available; it is being used today and it is not solely a technology of the future, and because of this, workflow efficiency is being improved by orders of magnitude. Prediction accuracy is exceeding that of traditional "best practice", today. Imagine what it will be like tomorrow when really large data sets are available for training models.

We present a series of case studies illustrating how machine learning is being applied in exploration and production geoscience. We discuss how machine learning can be applied to both reservoir characterization and exploration at both regional scale and prospect level. Machine Learning technology and data science is exposing to geoscientists hidden relationships in measured data; it removes biases and provides metrics for predictions and estimations. We discuss the potential of value creation by applying machine learning on very large data sets, and the value to society that can be created by making data sources openly accessible.

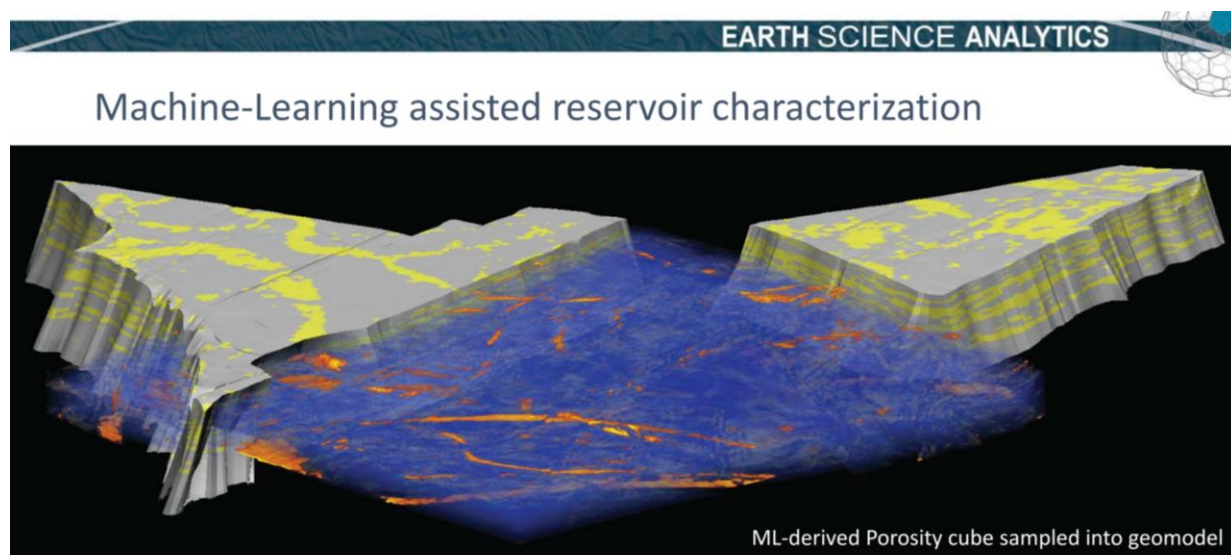


Fig. 2. Machine-learning derived porosity cube sampled into geomodel.

The currently applied machine-learning technology shows that interdisciplinary approaches lead to a deeper understanding of our prediction problems, providing a framework for creative solutions and improved decision-making. The future for decision-making technology for exploration and production is here, today, and we should integrate this technology into our workflows to enable data driven and cost-effective decisions. Does machine learning have the power to transform petroleum geoscience today, like Newton's calculus transformed physics more than 300 years ago?

NOTES

Generalized Classification of Lithology from Wireline Logs Using Machine Learning as Applied to the Permian Basin, USA and North West Shelf, Australia

Jose M Montero^{1*}, Jiazuo Zhang¹, Graham Baines¹, Thomas O'Toole¹

¹ Halliburton. * Corresponding Author: jose.montero@halliburton.com

Subsurface databases can include many hundreds of thousands of uninterpreted wells that would require hundreds of person-years to interpret manually. Consequently, methods for automated well interpretation and correlation are necessary to use these data in exploration and production workflows. Here, a supervised machine learning (ML) pipeline is demonstrated that quickly delivers lithology interpretations from wireline logs.

The capacity of a supervised ML algorithm to perform robust classifications depends on the availability of high-quality, accurate, and, ideally, large training datasets. For this study, a training dataset was built using wells from the North West Shelf, Australia and the Permian Basin, USA. Petrophysical interpretations were performed on wireline logs from these wells using advanced multiminerals methods to determine lithology. These interpretations were ground-truthed using available observations from core and mud logs.

Interpreted data were then used to train and test a supervised ML pipeline for lithology classification. Of the ML methods tested, the eXtreme gradient boosted tree (XGBoost) algorithm yielded the best results, in agreement with previous studies (e.g., Hall and Hall 2017). Several feature engineering steps were implemented that improved classification performance by mimicking the methods and thought processes used by petrophysicists. Posterior probabilities on the classifications were also calculated to assess the confidence in the automated interpretations.

The final pipeline was tested for a range of scenarios. Specifically, tests included minimum data requirements to obtain a high-quality prediction and the ability of a model trained on data from one basin to reliably predict lithology in wells from different geological settings. In all scenarios, the method yielded good results, suggesting it can be applied to rapidly provide first pass lithology interpretations, enabling companies to begin leveraging their backlog of uninterpreted datasets.

NOTES

Making the case for Big Data petrography

Jenny Omma*, Henrik Omma, Martin Rittner, Aukje Benedictus, Mar Cortés-López
Rocktype Ltd, 87 Divinity Road, Oxford, OX4 1LN
**jenny.omma@rocktype.com*

Rocks are multivariate systems with properties of interest (e.g. porosity, brittleness etc) responding to tens or hundreds of rock characteristics. These independent and dependent variables include, but are not limited to, grain size, sorting, mineralogy, presence of organic material, pressure, temperature and fluid chemistry, all of which change through time. A rock's detailed characteristics is determined by the interplay between its provenance, transport conditions, depositional and diagenetic history.

Reservoir quality studies aim to predict rock properties away from sample control. Predictions are often based on small sample sets coupled with larger proxy datasets (e.g. petrophysical logs) or with rock property equations (e.g. quartz cement-temperature models).

Significant advances in the speed of petrographic (term used generally, meaning “the study of rock”) data acquisition and interpretation techniques allows for the generation of Big Data petrographic datasets, based on sample numbers, N, orders of magnitude larger than previously feasible.

In order to robustly establish correlation, or preferably, causation, between rock characteristics and rock properties, and between physical processes and rock properties, we need sample numbers far in excess of what we are familiar with today.

Experimental design theory calls for several repeat tests where only one variable is different and all others constant. In our case that means systematically sampling, where possible, several instances of every variable rock characteristic, which quickly adds up to thousands to hundreds of thousands of samples per study. Advanced data analytics techniques should be applied to the datasets created from these samples, in line with techniques applied in other industries tasked with prediction (such as insurance, marketing or finance industries).

We believe we are entering a hugely exciting period for our field, with Big Data petrography poised to revolutionise our ability to predict rock properties. We expect to see frequent breakthroughs in capability and understanding, similar to what we currently see in the related field of genetics. Competitive advantage will be created for companies harnessing this opportunity.

At Rocktype, we create Big Data petrographic datasets using automated SEM-EDS technology, QEMSCAN, a product of Thermo Fisher Scientific. For this talk, aspects of our QEMSCAN Big Data petrography workflow for reservoir quality analysis will be illustrated using a 1000 sample dataset from Barents Sea wells, collected, scanned and interpreted over a two month period. Our aim in the near future is for this type of analysis to be possible in a week or less.

We also feel that is important to acknowledge that change to an industry, be it incremental or disruptive, leads to rational and emotive pushback. We must seriously engage with and address concerns, for example regarding overreliance on black box solutions, concerns about side-lining current, proven capabilities in pursuit of the next great thing or fear of being excluded from our own profession due to technological advances we don't understand. Conversely, we must challenge ourselves not to reject new techniques due to its unfamiliarity, often expressed as a range of concerns centred on “geological thinking being taken out of the workflow”. Science advances and we must advance with it. Creating more statistically robust datasets and retraceable interpretations, enabled by the digital revolution, is the only

scientifically and economically viable course. We call for open discussions within companies and research groups to enable these improvements, challenging emotive resistance to change by properly addressing real concerns shared by many.

NOTES

Ensemble Learning Approach to Lithofacies Classification Using Well Logs

Didi-Ooi, S.M.^{1*}, Zhang, L.C.², Cheng Zhan³

¹*School of Earth Sciences, University of Bristol*

²*CGG*

³*Anadarko Petroleum*

* *email: didi.ooi@bristol.ac.uk*

This paper presents a geologic facies classification using a supervised machine learning gradient boosting classifier, implemented in scikit-learn. With over 4000 training examples based on 7 predictor variables from logs in ten wells and over 800 test data in two wells to predict 9 different facies classes from a carbonate gas reservoir, the classifier algorithm resulted in F1 score of 0.630.

Geoscientists have been analyzing extensive amounts of data over the years. The proliferation of computing power, open-source tools and robust algorithms have made the process more robust and efficient. Facies classification consists of assigning a rock type based on measured properties from indirect measurement (i.e. wireline logs) and/or direct study of the cores. However, core sampling is often not ideal budget-wise and conventional lithofacies classification by human interpreters is very time-consuming. Alternative methods using machine learning with neural networks have been proposed by Busch et al. (1987), but this technique is yet to be exploited due to constraints in computing power.

We propose the use of a meta-algorithm method called ensemble learning which combines several machine learning techniques to enhance the model's predictive accuracy. This algorithm was applied through participation of the SEG Machine Learning Contest. The training data set consists of seven features: gamma ray, resistivity, photoelectric effect, density/neutron porosity difference, average density/neutron porosity, nonmarine/marine indicator, and relative position. The association of the features result in rock facies hand-labelled at half-foot depth intervals.

This small supervised multi-class classification problem employs an *extreme gradient boosted trees* model which is an ensemble of *decision trees*. During each training layer, a decision tree classifier learns a set of thresholds to separate features belonging to different classes. This method combines prediction of large numbers of near random decision trees, but they have the tendency to overfit the training data. To avoid this, gradient boosting classifiers such as *XGBoost* algorithm can help by splitting training data observations into different subsets (Friedman, 2000). A limited number of features are then selected from each subset to train a separate decision tree.

The importance of feature engineering, which requires domain knowledge expertise to generate additional features, is crucial to improve the model performance. A set of augmented features based on neighbouring intervals, which exploits the spatial correlation of the data, were adopted, capturing the greatest gradient of the feature vectors for each variable between two adjacent samples. This has improved the accuracy of our lithofacies classification model by 47%, as compared to the benchmark score (Hall, 2016), providing a welcoming reassurance that human experts still have a valuable role to play.

Due to the proprietary nature of geoscience data, the data size is limited to what was provided, but this provides a stepping stone for geoscientists to further explore and apply machine learning techniques to lithofacies prediction, while employing their domain knowledge to further enhance the accuracy of predictive models. Potential applications can be used in many ways including validating velocity models for seismic data, fault interpretation and well top interpretation.

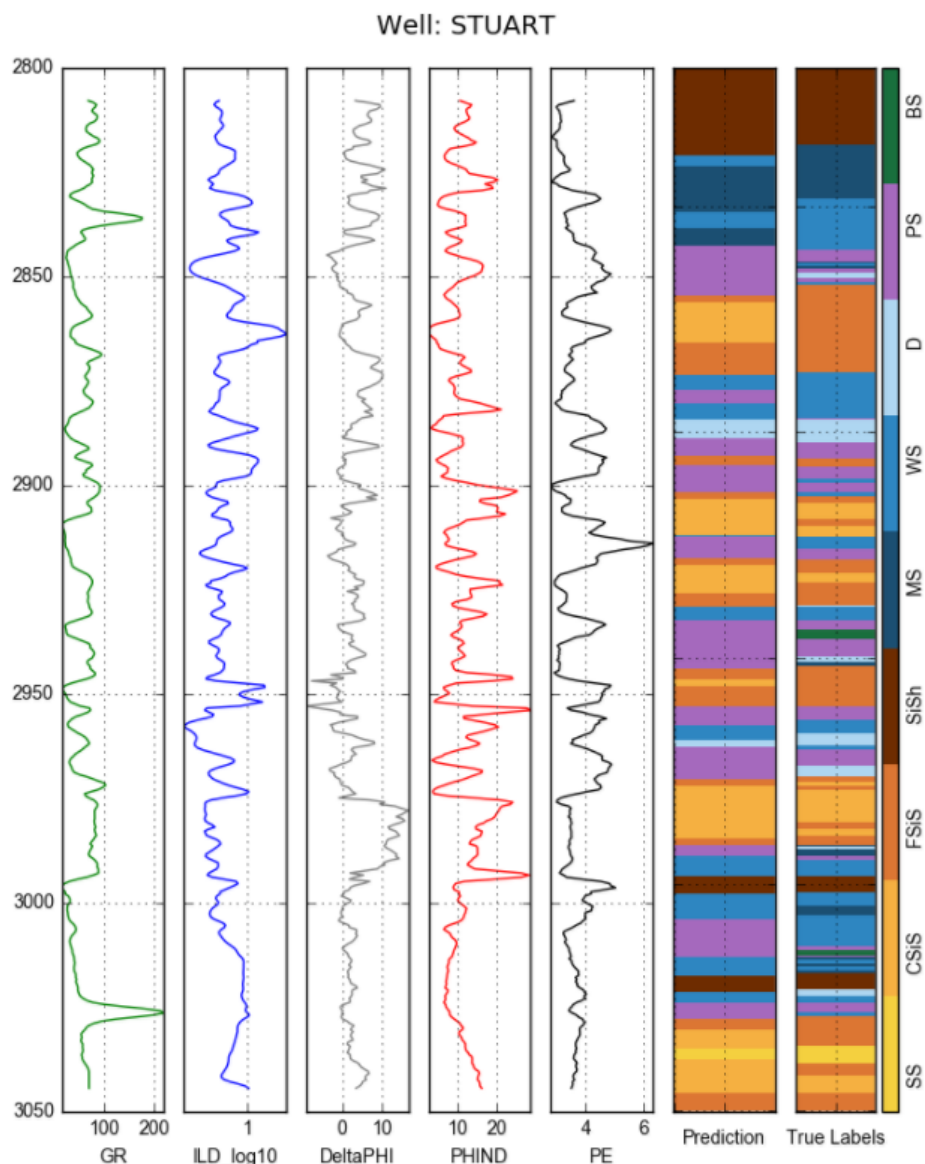


Figure 1. Left to right: A series of well logs - Gamma ray (GR), resistivity (ILD_log10), photoelectric effect (PE) neutron-density porosity difference (DeltaPHI), average neutron-density porosity (PHIND); followed by facies classification solution showing the predicted facies (Prediction) and true interpreted facies (True Labels). The facies are, from top to bottom: bafflestone (BS), packestone (PS), dolomite (D), wackestone (WS), mudstone (MS), siltstone and shale (SiSh), nonmarine fine siltstone (FSiS), nonmarine coarse siltstone (CSiS), nonmarine sandstone (SS).

Acknowledgements

Thanks to Matt Hall (Agile) and Brendon Hall (Enthought) for organizing the SEG Machine Learning challenge.

References

Hall, B. (2016). Facies classification using machine learning. *The Leading Edge*, 35(10), 906-909.

Friedman, J. H. (2000) Greedy function approximation: A gradient boosting machine: *Annals of Statistics*, 29, 1189–1232.

NOTES

SOFTWARE DEMONSTRATION ABSTRACTS (in programme order)

Osokey Stream - cloud-native seismic platform

Rhian Burrell, Charles Jones, Joseph Nicholson and **James Selvage**

Osokey

james@osokey.com

Osokey Stream is a cloud-native platform seismic platform running on Amazon Web Services (AWS). The platform shows a number of innovations that are made possible by taking a cloud-native approach to development:

- Automated and scalable ingestion of SEG-Y data
- Encryption of SEG-Y data at rest and in transit
- Access 2D/3D post- and pre-stack seismic on any HTML5 web browser (Figure 1)
- Real-time global collaboration (Figure 2)
- On-the-fly attribute generation
- Storage optimisation using analytics
- SEG-Y files can be shared securely for a limited time period
- A platform for future innovations, e.g. machine learning (Figure 2)

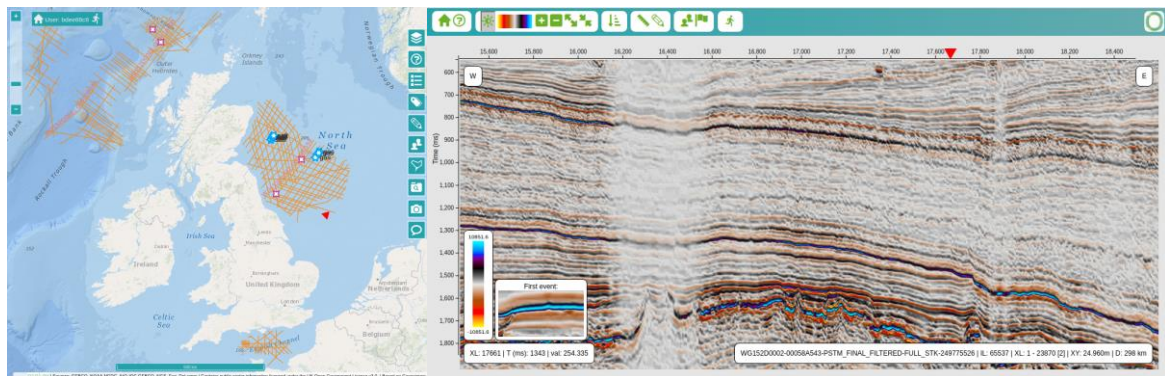


Figure 1 - 2D seismic data from UK [1] being made available through Osokey Stream.

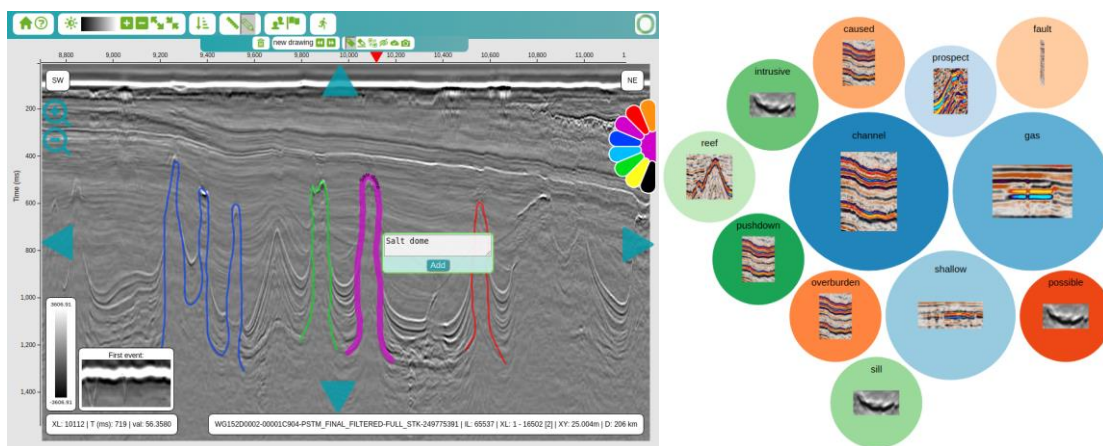


Figure 2 - Left: real-time collaboration is made possible. Right: By linking human observations as metadata associated with the SEG-Y data it becomes a valuable resource for machine learning.

Reference

[1] Contains public sector information licensed under the UK Open Government Licence v3.0.

NOTES

Facilitating the effective application of analogue databases to reservoir models

Marijn van Cappelle^{1,*}, Simon Lomas², Luca Colombera³, Marco Patacci³, Alwin ten Hove¹, Sarah Cobain², Ben Meyer² & Viki O'Connor²

¹ *Petrotechnical Data Systems B.V., Lange Kleiweg 10, 2288 GK, Rijswijk, The Netherlands*

² *Petrotechnical Data Systems Ltd, Chester House, 81-83 Fulham High Street, London SW6 3JA, United Kingdom*

³ *School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, United Kingdom*

*contact: marijn.vancappelle@pds.nl

In the petroleum industry, reservoir models of oil and gas fields are built in order to estimate hydrocarbon volumes in place and forecast hydrocarbon production. Subsurface datasets used for conditioning models consist mainly of 1D well data: inferring the three-dimensional architecture of the inter-well volume involves significant uncertainty. It is widely acknowledged that applying rock outcrop analogues to reservoir models helps reducing uncertainty through the inclusion of geological realism.

In recent years the compilation of many analogue databases has been undertaken. Often these databases are constructed, maintained and used within petroleum companies. In other cases, the databases are vehicles for transferring data obtained in the context of academic research to sponsoring industry consortia. However, integrating data from these databases into workflows in development and production teams has proven difficult. Hurdles preventing geoscientists querying and applying data include: 1) the wide range in training and experience of staff, 2) time constraints, 3) constraints in the number of reservoir models which can be built and run, 4) practical limitations in the detail/resolution of reservoir models, and 5) the lack of quantification for all the parameters used in reservoir modelling.

We present a method (Ava Clastics) designed to overcome these hurdles of effectively integrating outcrop analogue data in reservoir modelling workflows. An online portal has been developed that makes it possible to quickly filter relevant outcrop analogues contained in relational databases of fluvial (FAKTS), shallow-marine (SMAKS) and deep-marine (DMAKS) clastic depositional systems. These relational databases store data on sedimentary units at multiple scales, describing their geometries, their relations with surrounding elements, their hierarchical organization, and their lithological heterogeneity. All the analogues contained in the databases are classified on the geological controls of the depositional systems, contextual information, and metadata, allowing users to select analogues that are most suitable in application to the reservoir at hand. The filtered analogue data is visualised graphically in real time, to enable the user to review database outputs and optimize the selection of inputs to a reservoir model.

Subsequently, the filtered analogue data are automatically parameterized to match the requirements of different stochastic facies modelling algorithms, thanks to workflows for translation of analogue data into model inputs. Where available, the analogue data is directly used as parameters in the modelling algorithm. When parameters are not directly available from suitable analogues, empirical relationships that are established in the scientific literature are used to calculate input parameters. After review, the parameterized composite analogue can be directly imported into the reservoir modelling package.

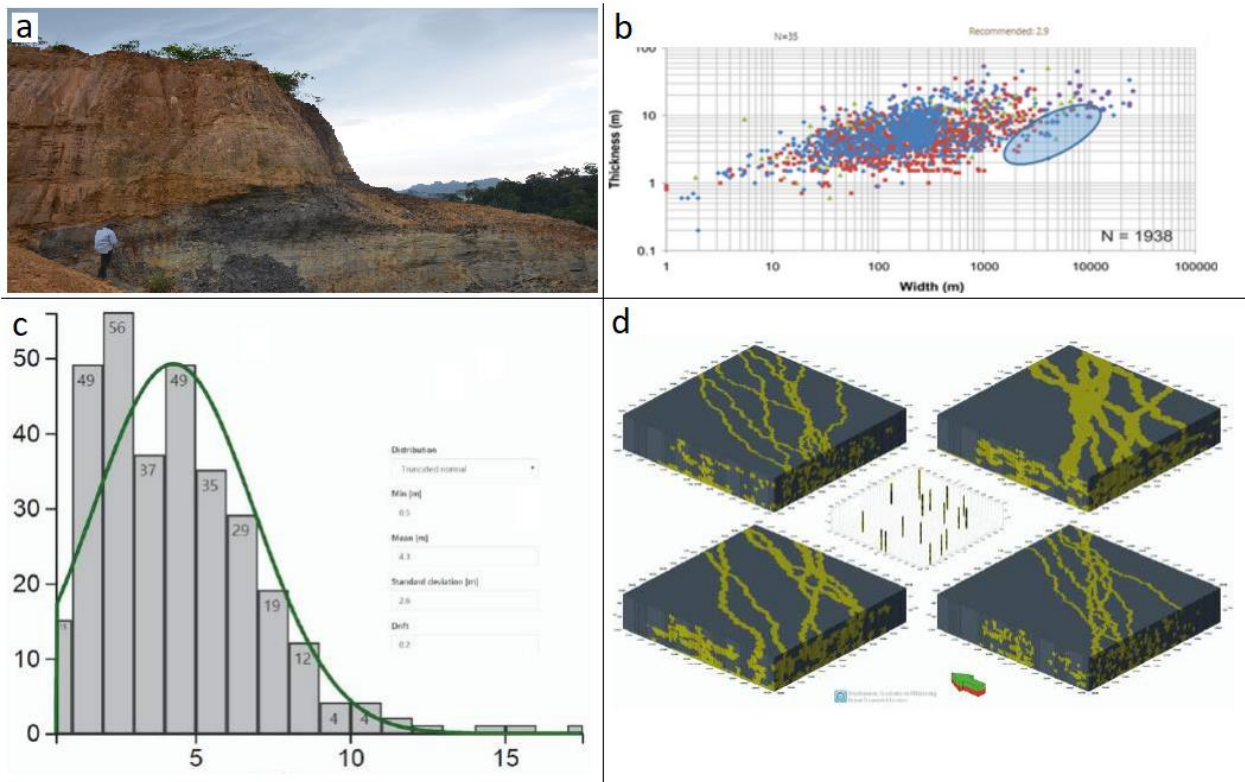


Figure. a) Rock outcrop analogues help in reducing uncertainty in inferring rock properties in between well. b) Analogue data is stored in databases. c) Parametrization of analogue data for use in stochastic facies modelling. d) Application of the parametrised analogue data to reservoir models.

NOTES

GBDB, an expanding open data system for geological information

Junxuan Fan¹, Xudong Hou², Jiao Yang¹

1 State Key laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Science, Nanjing 210008, China; fanjunxuan@gmail.com

2 Key Laboratory of Economic Stratigraphy and Palaeogeography, Chinese Academy of Sciences, Nanjing 210008, China

The Geobiodiversity Database (GBDB, <http://www.geobiodiversity.com>) is an open-access database which was initially designed for the management and analysis of stratigraphic and paleontological information. Its goal is to facilitate regional and global collaborations focused on regional and global correlation, quantitative stratigraphy, systematics, paleogeography and paleoecology. It became the formal database of the International Commission on Stratigraphy (ICS) in August 2012 and that of the International Palaeontological Association (IPA) in February 2015. As of November 2017, an abundance of over 17,000 sections, 86,000 collections, 450,000 fossil occurrences and 16,000 taxonomic opinions from the world has been digitized into the GBDB system.

Many visualization and analytical tools have been developed to make the database more useful as scientific and educational tools. A newly designed tool is the parallel computing program for quantitative stratigraphy – CONOP.SAGA, which is designed for the big data analysis of the stratigraphic and palaeontological data. After it was successfully developed in early 2017, the program has been run on the second largest supercomputer in the world - “Tianhe II” for over 6 million CPU core-hours in order to reconstruct the high-resolution Palaeozoic marine biodiversity. Another powerful function/app is the GeoPano which is designed for the panoramic demonstration of mm-scale digitization of outcrops. The integrated database supports the permanent digitization and preservation of the outcrop information and can be used for virtual fieldtrips and virtual outcrops.

After ten years' development, the GBDB is now being expanded to cover a wide range of disciplines of geological science, such as palaeogeography, geochemistry, geochronology, and sedimentology. Some new techniques are tested and integrated into the GBDB system, such as 3D palaeogeographic visualization, machine learning and virtual reality.

NOTES

Geoscience data landscapes through integration with new cloud platforms

Jamie Hisee

Interica

Interica will present and demonstrate how their technology is enabling clients to gain deeper insight into geoscience data landscapes through integration with new cloud platforms. As Oil and Gas companies look to leverage advances in cloud compute, storage and analytics Interica's technology provides unique data discovery, analysis and action that integrates with these new cloud and hybrid environments, platforms and technologies to deliver E&P organisation with the tools to control the huge growth in their subsurface related datasets.

Interica will demonstrate their forward vision for their PRM and PARS solutions, running in a hybrid cloud environment, and integrating with a range of interpretation to simulation applications and E&P datasets. Furthermore, Interica will highlight how integration through open API's will enable improved visualisation, delivering E&P organisation with the capability to discover, analyse and act of their subsurface data in a modern and integrated environment.

NOTES

QEMSCAN Big Data petrography: a tutorial

Jenny Omma*¹, Aukje Benedictus¹, Jon Gluyas², Mar Cortés-López¹, Henrik Omma¹, Martin Rittner¹

¹Rocktype Ltd, 87 Divinity Road, Oxford, OX4 1LN

²University of Durham

*jenny.omma@rocktype.com

At Rocktype, we create Big Data petrographic datasets with using a QEMSCAN instrument, an x-ray energy dispersive spectroscopic (EDS) scanning electron microscope (SEM) based tool that provides mineralogy, porosity, major element chemistry, rock texture and calculated rock properties through automated data collection and a mixture of automated and manual data analysis.

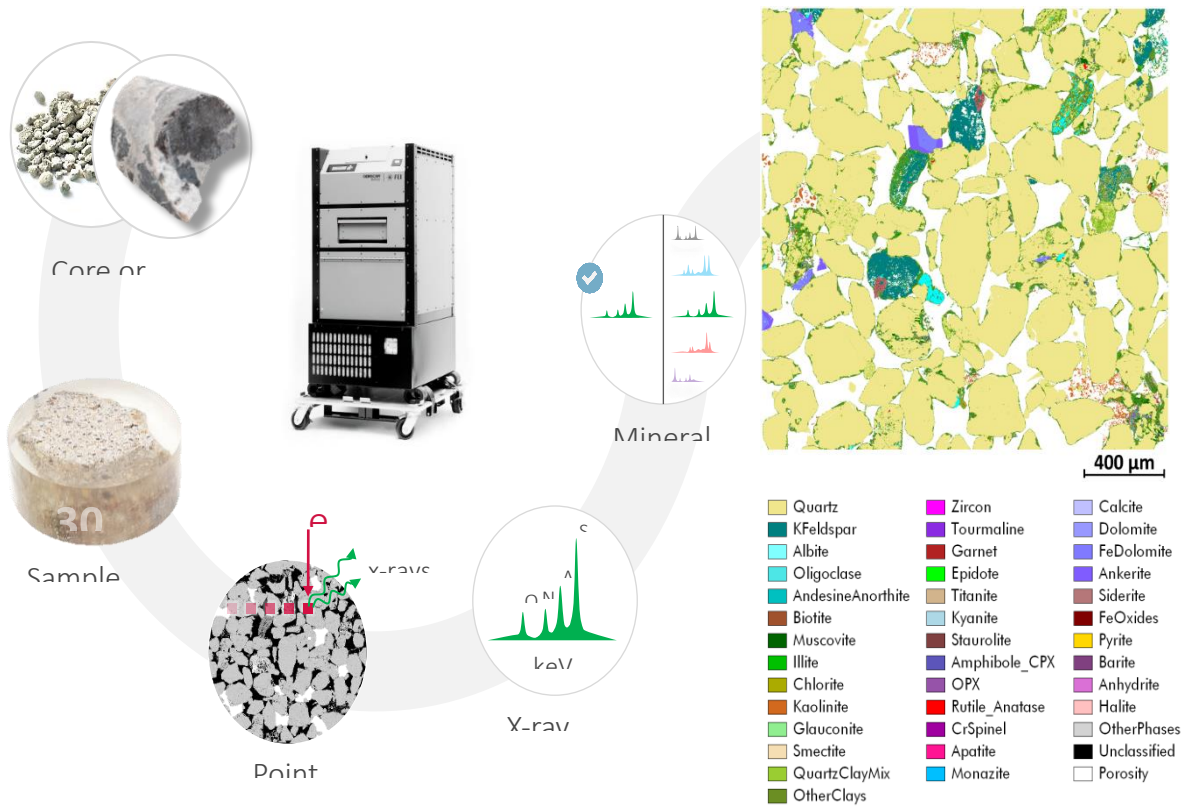
A sample is mounted in resin and polished to reveal a flat surface. The SEM electron beam is programmed to hit the sample surface in a pre-determined, but operator selected, x-y raster spacing. Electrons in the sample are excited by the incident electron beam and when they relax, release a packet of X-ray energy, which is collected by detectors within the SEM (the basis of EDS).

The combination of X-ray energies (spectra) for a given analysis point are diagnostics of one or more mineral phases. Non-automated EDS analysis returns the spectra and elemental proportions to the operator, whereas the automated QEMSCAN technique runs the signal through an element-to-mineral database, assigns the pixel to a mineral phase (or Unknown) and moves the beam to the next analysis point.

For a sample, data are collected from 50K to 10MM analysis points and combined to create a very rich digital dataset. Using the QEMSCAN iDiscover software the data can be interrogated as a bulk dataset, on a sample by sample basis, on an object by object basis (e.g. one grain or one pore is an object) or on a pixel by pixel basis. One single QEMSCAN analysis represents a Big Data petrographic dataset in its own right.

The contribution aims to provide a tutorial on the QEMSCAN technology, given that most workers do not have direct access to the hardware or software. It will illustrate key QEMSCAN data collection and data analysis techniques by description of two sandstones from the Rotliegend Group, Southern North Sea (one sample is shown below). We will present details of the data acquisition, mineral list decision tree process, and present details of the petrographic results including QEMSCAN mineral maps, BSE maps, elemental maps, pixel neighbor maps (pore lining phases etc), object based data such as pore and grain size and bulk sample data.

QEMSCAN is a registered trademark of Thermo Fisher Scientific via its subsidiary FEI Inc.



NOTES

POSTER ABSTRACTS

(in alphabetical order)

Machine Learning for Mineral Prospectivity Mapping

Lucille Ablett

Maps of surficial geochemical composition are essential to our understanding of the natural environment, with geochemical knowledge directly influencing our interactions with it. Geochemical maps are critical for mineral exploration, often providing the initial indication of a potentially economic resource. Historically, geochemistry has been modelled based on spatial autocorrelation. These models fail to incorporate auxiliary information that could provide greater insight into the geochemical composition of the near-surface. This study investigates the effectiveness of machine learning algorithms for predictive modelling of near-surface geochemistry in SW England using a range of high resolution, remotely-sensed, geoscientific information. It also provides a critical assessment of the suitability of each method in the context of mineral exploration.

The algorithms considered are support vector regression, random forests, and extreme gradient boosted trees. Using 10-fold cross-validation, the accuracy of these methods in predicting the concentration of 10 elements measured from soil samples is assessed, including investigation of the susceptibility of these algorithms to changes to the size of the sample data. Whilst random forests were found to generate the most accurate models when using a large training dataset, models generated using smaller datasets failed to predict the target variables when tested against an unseen dataset. Extreme gradient boosted trees produced the most accurate models when trained using small datasets, whilst support vector regression was most successful at modelling the extremes in the data. Maps of predicted mineralisation were also produced. Although the predicted values of these maps vary between models, the spatial distribution of element concentration was found to be correlated between the methods used.

Improved understanding of borehole instability mechanisms through development of an enhanced visualization-numerical modelling approach

Abraham Audu*, Professor John Coggan, Dr Matthew Eyre

*Camborne School of Mines, University of Exeter, Cornwall Campus, Cornwall TR10 9FE, UK (*Email: aa804@exeter.ac.uk)*



This research will utilize digital reconstruction of borehole data and incorporation of three-dimensional numerical modelling of borehole instability mechanisms to demonstrate how innovative use of both virtual and augmented reality (VR and AR) can provide improved visualization and enhanced understanding of complex phenomena controlling deformation and failure of rock surrounding a borehole at depth.

Wellbore interaction with stress continues to evolve throughout the life of the production/injection well until abandonment. Developing fit-for-purpose models to be used for stability predictions will help to mitigate problems like borehole break out, casing shear, unwanted fluid loss and cement sheath failure. The project, through analysis of case study data, will provide new approaches for documentation and communication of modelled results through a virtual environment, highlighting spatial and temporal changes resulting from evolution of stress induced wellbore failure.

The visualization aspect of the research provides a medium for explanation of data to aid better understanding of borehole instability mechanisms through exploration of large data sets and modelling techniques. Patterns in visualizations will be generated using the Gestalt Principles incorporating the four preattentive attributes of visual perception namely colour, form, movement and spatial positioning. The aim here is to convert the mass quantities of geological and engineering data into a format usable for active decision making. AR and VR platforms provide easily distributed file formats that can be used by a wide range of stakeholders exploring the data in virtual space, from any angle, in a dynamic environment.

The research project will demonstrate the integrated visualization-modelling is an important part of successful stakeholder engagement and dissemination/management of geoscientific data and how this approach can be used as a platform for data evaluation, monitoring and optimization of hydrocarbon exploitation

The application of imaging IR spectroscopy for mineralogical analysis of core and cuttings.

Dr Gavin Hunt

Spectra-Map Ltd, Unit 33 Wheatley Business Centre, Old London Road, Wheatley, Oxford, OX33 1XW, U.K.

email: ghunt@spectra-map.co.uk

Infrared imaging spectroscopy is a Big Data petrographic technique that has been used in the mining industry for several decades and is now becoming well-established in the oil sector. It is complementary to other mineral analytical techniques, such as XRD or point counting. However, a major drawback of these techniques, including the older point sampling IR spectrometers, is that they require sample preparation – an expense in both time and money. More importantly they cannot provide continuous quantitative mineralogical data, which otherwise could be used to aid reservoir characterisation and identify trends that may be useful in calibrating log interpretation models.

The most recent advance in SWIR (Short Wave Infra-Red) spectroscopy is the use of real-time imaging reflectance spectrometers. These are non-contact and non-destructive, and acquire continuous mineral data in a detailed sub-mm pixel image format. The first of these introduced to the UK oil industry in 2008 is the portable SpecCam imaging spectrometer. It has been designed for detailed and automated mineral logging from cuttings and core and plugs over the mm to metre scale.

The current spectrometer design can characterise many of the key molecular vibrations associated with amorphous and crystalline minerals that are important to understanding the performance of tight and unconventional reservoirs. The IR image data can be used to accurately discriminate and quantify different polytypes of the swelling and non-swelling clays, carbonates and sulphates. It also uniquely provides hydrocarbon information (presence or absence) and whether the oil is 'light', 'heavy' or an invasive OBM.

Its ability to image and map subtle compositional and crystallinity changes in a variety of important reservoir-influencing minerals in a continuous Big Data format, can help identify unconformities and aid well to well correlations. In addition, the impact and interrelationship of clay and carbonate mineralogies on hydrocarbon distribution and permeabilities can be studied.

SpecCam image data can also be used early on in the sampling workflow. For example, SWIR data allows optimal selection of discrete sample points for scanning electron microscope-energy dispersive spectroscopy (SEM-EDS; QEMSCAN®), especially where sampling in heterogeneous reservoirs is difficult with the naked eye. Data from the 2 techniques can be integrated to provide more robust determination of petrophysical calibration parameters, improving the log calibration workflow.

SWIR imaging technology addresses 3 major limitations of point sampling techniques; low productivity, inability to show detailed spatial distribution of minerals and low data density. Imaging IR spectrometers, such as the SpecCam, overcome these limitations and can provide a unique, spatially-detailed and continuous dataset, which provides a link between point sampling methods and continuous logs.

Improved production decline analysis through the use of Machine Learning techniques

Humphrey, E^(a, b)., Elleuch, R^(c).

(a) School of Geosciences, King's College, University of Aberdeen, AB24 3UE Aberdeen, United Kingdom

(b) Toby Project Services, MK45 2US Maulden, United Kingdom

(c) Incubit Inc., 3F Shirokanedai Manaka Building, 3-17-5 Shirokanedai, Minato-ku, Tokyo-to, Japan

Predicting the performance of petroleum wells is vital for estimating future reserves and for efficient field development planning. Conventional methodologies are focused around production decline analysis (PDA) which extrapolates early production data using graphical curves. The key challenges with PDA are that graphical functions are unable to capture near-term heterogeneity in production rates, and require curves to be screened and adjusted to try to represent reservoir conditions; these issues ultimately increase reservoir estimate uncertainty and planning-based risk.

Recent advances in computational abilities has rapidly advanced the field of Artificial Intelligence, particularly associated with Machine Learning, and allows us to maximise the value of existing data at no additional cost. Whilst machine learning techniques are commonly used in the Technology & Banking sectors, the Petroleum industry has been slow to adopt these techniques despite having a wealth of historical data.

This study investigates the use of deep learning neural networks to forecast petroleum well production rates from time-series data. Monthly production data was taken from conventional wells in South Dakota (USA) and initially reprocessed to remove the influences of human interaction and used to build a Gated Recurrent Unit Neural Network model (GRU-NN). Produced models were classified as either univariate or multivariate; using oil production rates against time or a combination of oil, gas and water production rates. Models were subsequently used to create forecasts which estimated production rates for a given well during a 24-month period.

Production forecasts generated from GRU-NN models were compared against Arps PDA techniques to evaluate the accuracy of forecast results against real data. GRU-NN models demonstrated the ability to create forecasts which displayed near-term heterogeneity in production rates, similar to patterns seen in real production data, which could not be achieved by Arps techniques. Forecasts generated for a variety of well production profiles led to an improved forecast accuracy between 18-34% compared to conventional Arps techniques, performing particularly well at forecasting late-stage production rates. When comparing the accuracy between GRU-NN model forecasts, multivariate models were found to outperform univariate wells by utilising patterns associated with the interaction between oil, gas and water production rates.

Machine learning techniques have been found to outperform conventional Arps decline methods in our sample dataset, ultimately leading to more accurate reserve estimates and the ability to identify low-productivity wells which subsequently improves workover and resource management strategies during oilfield production. Machine Learning forecasts have been created rapidly using historical data and without the need for time-consuming reservoir modelling, highlighting the significance of this technique in small-scale production settings which are economically sensitive.

Understanding and enabling data standards for large environmental data sets

Francesca Laws (*Sellafield Ltd, Land Quality*), Helen M^cKenzie (*Sellafield Ltd, Land Quality*), Stuart Cooper (*Sellafield Ltd, Land Quality*).

Over the last five years Land Quality, Sellafield Ltd, have been working to deliver a project to combine all their data regarding groundwater and contaminated land in to a single source. In 2012 Land Quality took delivery of the Land Quality Data Management System (LQDMS), a product from Informed Solutions, configured for use for Sellafield data.

The configuration took place as a mutual piece of work to develop an Entity-Relationship model for the data required, building the connections that link groundwater monitoring wells, boreholes, sampling programmes, analytes, field data, sample data, etc. This resulted in a data model that conveyed the complicated web of relationships that previously had to be reconsidered every time data aggregation was needed for analysis.

The database underwent a programme of data migration, targeting a series of historical data sources. The data migration plan recognised that sources could hold duplicate data, data of varying quality, different aspects of that data and so on. The plan was designed to migrate in the poorer data first so that should any duplication or lower quality be encountered, it would be overwritten by the later more complete datasets. However all actions in the database are recorded as part of an audit history and these can be examined for each record to ensure no data are lost or provenance can be traced back to original sources.

The completion of this migration lead on to a programme of quality analysis which is still underway, given that all data are related to specific sample points, every effort has been made to ensure that the Sample Point base data are correct, leading to comparing the migrated data to original logs and reports. Once this process is complete, a review of the analytical data will take place. It is envisaged that this will be a continual process of refinement, users will find gaps or errors that need resolution, data trawls will find further data sources to migrate, data fields may be created or retired, etc.

This refined and quality assured source of information can then start to be tied in to our other applications, through the use of GIS. All the data in the system have a spatial reference, their sample location, and as such can be visualised in GIS. GIS is tied in to our records management system, based on documents related to our technical committees which are held in our records management system, these documents are all tagged with metadata and have their own spatial references. GIS itself contains a number of spatial datasets, underpinned with metadata. Eventually you get to a point where you are able to produce a layer cake of information that is quality assured, updated and accessible through multiple systems to interrogate and assess areas, either with current data, historical data or a combination. Users will have access to the quality marking of the data and be able to produce auditable studies that support decision making going forward.

Virtual Glaciers and Glaciated Landscapes

Derek A McDougall, *Institute of Science and the Environment, University of Worcester, Henwick Grove, Worcester UK – WR2 6AJ*
d.mcdougall@worc.ac.uk

The topic of glaciers and glaciation is sometimes perceived by students and teachers as being more difficult than others associated with more familiar environments. Nevertheless, well-designed fieldwork in areas with glaciers and/or glaciated landscapes can make all the difference to their understanding and enjoyment of it. Unfortunately, fieldwork to these environments rarely takes place at the same time as students learn about them in class. Sometimes, fieldwork does not take place at all. In providing on-demand, simulated fieldwork to glaciers and glaciated landscapes, this resource provides a solution.

Virtual Glaciers and Glaciated Landscapes is based around high quality, 360° interactive panoramas (panospheres), which allow users to look all around and zoom in on features of interest. The panoramas are linked, so it is possible to proceed through the landscape, going from one panorama to the next. This provides a sense of scale and orientation, and helps users understand how the landscape changes from one location to the next. In due course, the panoramas will include sound (for contemporary glacial environments), further contributing to the immersive experience, as well as links to other resources, such as 3D models and video.

The locations for Phase 1 of the project are the Helvellyn Range, Keskadale and Mosedale (all in the English Lake District) and the Moiry, Ferpècle and Arolla valleys (Swiss Alps). Phase 2 will see additional locations, including Snowdonia (north Wales) and at least one more alpine location. All the panoramas are marked in *Google Maps*, and downloadable geolocation data allows integration with *Google Earth*, *ArcGIS* and other location-aware software.

The core virtual fieldwork resources are provided without interpretation, which means that they can be incorporated into the curriculum at any level. In Phase 2 of the project, there will be a separate, password-protected resource area with guidance on interpretation and use in the classroom, and a forum for teachers and academics. The virtual fieldwork approach is intrinsically a more interactive and engaging approach than simply presenting static images, and provides the basis for a range of learner activities and challenges. These include: landform and landscape interpretation exercises (applying and extending what has just been covered in a lecture, seminar or through reading); field sketching; 'field-checking' of geomorphological maps (produced using orthoimagery and elevation datasets in a GIS); student-produced geomorphological field guides (as a form of assessment); preparation for real fieldwork (whether or not to these specific locations); follow-up to fieldwork in these locations or when real fieldwork does not take place for some reason (e.g. poor weather, illness, budget constraints).

The resources can be viewed on any modern device with an internet connection, including phones and tablets, but they are best appreciated on a large monitor or by using virtual reality headsets. There is no requirement for any browser plug-ins or other proprietary software, although the browser must be up-to-date (HTML 5 compliant). In other words, the resources are easy to use and there is no need for expensive equipment.

This project remains a work in progress. Whilst the core resources are available to use now, development work will continue over the next 24 months.

ACKNOWLEDGEMENTS

I am most grateful to the following organisations, which have supported the development of this project: the Institute of Science and the Environment Learning and Teaching Fund (University of Worcester), the Quaternary Research Association, and the British Society for Geomorphology.

Data Integration: Understanding the Importance of Data and Knowledge Management

H. McKenzie¹, F. Laws¹, S. Cooper¹

¹*Land Quality & End States, Sellafield Ltd.*

Corresponding author: helen.m.mckenzie@sellafieldsites.com

Geological and hydrogeological information has been collected over several decades at the Sellafield site. This information has been stored and used in systems relevant to the objectives of the data capture at the time of acquisition. This has led to a large dataset stored in silos that have differing structures and data quality. In 2011/12 Sellafield Ltd. commissioned a new data management platform configured to the business needs and incorporating suitable flexibility in the system to enable improved data management as the data requirements and needs evolve as the site moves into the remediation phase.

Challenges:

Historical data were stored in a number of formats, from structured relational databases to spreadsheets and proprietary modelling software, e.g. gINT, AGS etc. These 'silos' of complex datasets have led to challenges in determining duplicate records and a range of data quality issues had been identified. The solution that was developed to manage this data effectively was a SQL server relational database delivered by Informed Solutions, based on their InformedINSIGHT © platform, that was configured to align data workflows with Sellafield Ltd. working practices. A key step was defining the data model and ensuring it was suitable for the legacy data as well as new data that will be collected over the lifetime of the site. The database included a data migration toolkit to enable successful migration of data from the historical systems into a central data management system that could be queried. In addition the system was designed to understand the spatial context of the data from the geographical location of the sampling points to their three dimensional construction. This has allowed the data to be easily integrated with GIS software both through exports containing spatial data that can be used in GIS software and directly to ArcGIS using SQL queries. The benefits of such a system mean that there is a single authoritative data source to support management and planning towards achieving the site End State.

Opportunities:

An opportunity at the beginning of the project was realised to understand how the historical data landscape was impacting the ability to analyse, monitor and report the data and how this could be refined to meet the needs of future data management and analytics requirements. Methods to integrate data were developed to maintain the understanding of land quality at the Sellafield site and its future management. The improved data management capability has allowed further development of the knowledge capture and management which in turn will underpin key decisions as the site moves towards its end state in a changing regulatory environment.

Deciphering the sands of time: a zircon U–Pb age database for the Circum-Arctic region

Pointon, M.A.¹, Flowerdew, M.J.¹, Sinclair, S.^{1†} and Lipp, A¹.

¹CASP, West Building, Madingley Rise, Madingley Road, Cambridge, CB3 0UD, United Kingdom.

[†]Deceased 08/08/17

Email: michael.pointon@caspcam.ac.uk

How and when the Amerasia Basin opened are poorly understood but have significant implications for the geological histories and petroleum systems of sedimentary basins within the circum-Arctic region. This lack of understanding is in part due to the paucity of direct evidence and subsequent overprinting by the High Arctic Large Igneous Province. Uranium-lead detrital zircon geochronology, employed alone or in conjunction with Lu–Hf isotopes, is a widely employed tool for reconstructing sedimentary provenance. It can provide insights into the opening of the Amerasia Basin by constraining the pre-rift configuration of Arctic tectonic terranes and sediment transportation pathways.

Despite being comparatively remote, a large volume of published U–Pb age data exists from the onshore margins and offshore continental shelves of the Arctic Ocean. Rigorous comparison of these data is, however, seldom straightforward. This is because of variations in the treatment and presentation of data used by different studies. Recalculation of data, using a common set of parameters and criteria, is often necessary; for example, using the same age system ($^{206}\text{Pb}/^{238}\text{U}$ versus $^{207}\text{Pb}/^{206}\text{Pb}$) and employing thresholds on analytical precision and U–Pb age discordance to remove imprecise and unreliable analyses. Such recalculation is a laborious endeavour, particularly when substantial volumes of data are involved.

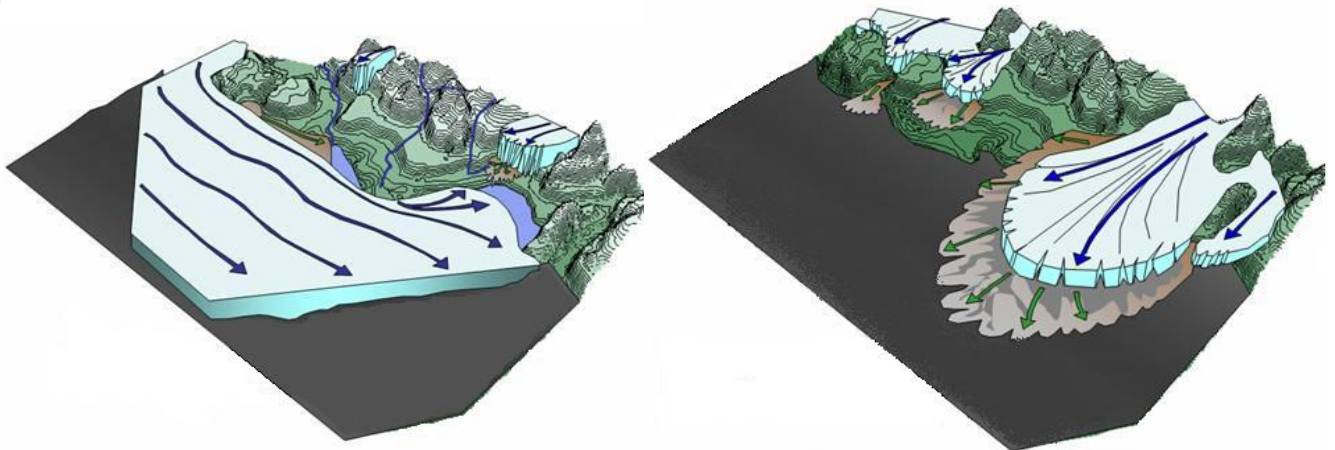
In an attempt to elucidate the opening of the Amerasia Basin, a standardised dataset of published zircon U–Pb ages and Lu–Hf isotope data from the circum-Arctic region is being developed as a geographical information system (GIS) database using ArcGIS®. The database serves as both a repository for large volumes of U–Pb age and Lu–Hf isotope data, and as a platform from which to interrogate the dataset. Custom database tools have been developed within ArcGIS® using Microsoft Visual Studio®. These facilitate the searching of the database and visualisation of U–Pb age and Lu–Hf isotope data within the GIS environment as epsilon hafnium plots, conventional probability density and cumulative density plots, as well as novel probability density heat map barcodes. Furthermore, similarity measures, using multidimensional scaling, are being developed to enable data to be compared with statistical rigour.

Compilation of such a database is a considerable undertaking. So far we have focused predominantly on compiling zircon U–Pb age data from igneous and metamorphic rocks within the circum-Arctic region, although compilation of detrital zircon U–Pb age data is ongoing. The database presently contains over 2,500 individual samples, >60,000 U–Pb zircon analyses and >7,400 Lu–Hf zircon analyses.

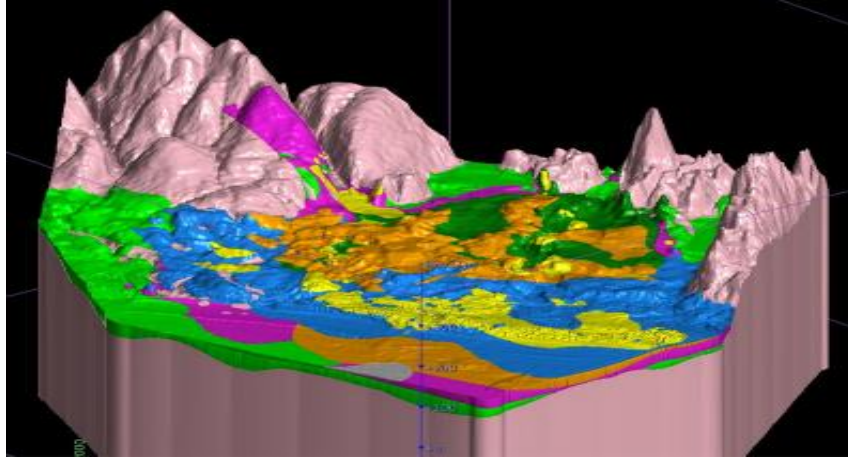
Development of 3D geological models for the UK low level waste repository

John Shevelan *LLW Repository Ltd* john.shevelan@llwrsite.com, Hannah Woollard *Wood plc*, Chris Coleman *Fugro*

The Quaternary geology beneath the LLWR site, in West Cumbria, is highly complex and likely reflects a history of deposition and erosional events dating back through the Quaternary Period. The complexity has been revealed as a consequence of the extensive investigations that have been undertaken at the LLWR. The site has been subject to a series of intrusive and non-intrusive investigations dating back to 1939 when the site was being developed as a Royal Ordnance Facility. Since then approximately 650 boreholes have been drilled, along with geophysical investigations, within and in close proximity to the LLWR site. Regional information has been provided from the extensive site investigation programme undertaken in the 1990s by Nirex but also information gathered from investigation programmes at Sellafield and more recently to support new build at Moorside. However, whilst the lateral coverage of boreholes on the site is extensive, the vertical penetration to the deeper Quaternary deposits and bedrock is more limited. The descriptive quality due to the age of many of the borehole records has caused additional problems.



A number of interpretations and accounts have been produced to assess the formation and likely variability of the Quaternary deposits. Initial interpretations based on a limited number of boreholes have been superseded by the development of 3D geological models that incorporate all the available data. Early models were constrained by conceptual models whereas later models have returned to the original borehole data to reduce anomalies and be more realistic. However, there is scope to improve the geological model by fuller use of available borehole data with new consideration of the lateral extrapolation of material types between boreholes based on a comparison with modern analogues of glacial deposits and other depositional processes and reconsideration of the link with the regional stratigraphic model, and in the long term; possibly reconsideration of that regional model itself.



PCA as an error diagnostic at long-term resistivity monitoring sites

Luke Sibbett^[1,2,3], Paul Wilkinson^[2], Jonathan E. Chambers^[2], Li Bai^[1], O. Kuras^[1] and P.I. Meldrum^[1]

[1] – *British Geological Survey, UK*

[2] – *University of Nottingham, UK*

[3] – *psxlr@nottingham.ac.uk*

Electrical resistivity tomography (ERT) is increasingly being used in long term remote monitoring, across applications such as slope stability, permafrost, CO₂ sequestration and contaminant tracking. Often, data are collected daily over a long period, each survey consisting of $\sim 10^3 - 10^4$ electrode configurations, alongside ancillary information from point sensors. There is an increasing need for automation techniques in the processing, analysis and interpretation of the data. One key part of this process is diagnostics and data quality control. It is vital to detect faults in the acquisition process as early as possible so that remedial work can be carried out before too much data has been lost or contaminated.

Currently, reciprocal measurements (where the current and potential electrodes are exchanged) are often used as an estimate of data quality, as many sources of systematic error violate the principle of reciprocity. An empirical threshold is then used to identify bad configurations which are then filtered out. The percentage of data filtered out can be used as an indicator of data quality.

Not all data errors are captured by the reciprocal measurements. Our motivating case study is a two-year long monitoring experiment at the Sellafield nuclear site. During the monitoring period, an electrical connector to borehole 1 of the ERT sensor array suffered mechanical damage, which is thought to have caused degradation of the sensor connections through moisture ingress and corrosion. However, this did not violate ERT reciprocity and, as the site was inaccessible for inspection, remained unnoticed until the post processing stage months later (and confirmed by visual checks).

Here, we present a principal component analysis (PCA) control chart analysis which has proved successful in identifying and highlighting the onset of the degradation from a change in the correlation structure of the measured errors. PCA can be used to generate a reduced dimensional representation of data, by exploiting correlations between the variables. The first three months of reciprocal error data (during which the connector was intact) were used to train a PCA model which explained 90% of the variance in the errors. The ability of the PCA model to capture the behavior of the errors was measured by a Q statistic, the mean squared residual between the data and the PCA model. While Q remained stable elsewhere, there is a clear jump in the residuals for configurations involving borehole 1 (figure 2) on 21st July. This is consistent with the appearance of artifacts in the inverted images, which originally were unexplained for some time due to the reliance on reciprocal error data quality assessment. This method has the potential to be used to detect other such cases as they occur, allowing remedial works to be carried out and preventing loss of data.

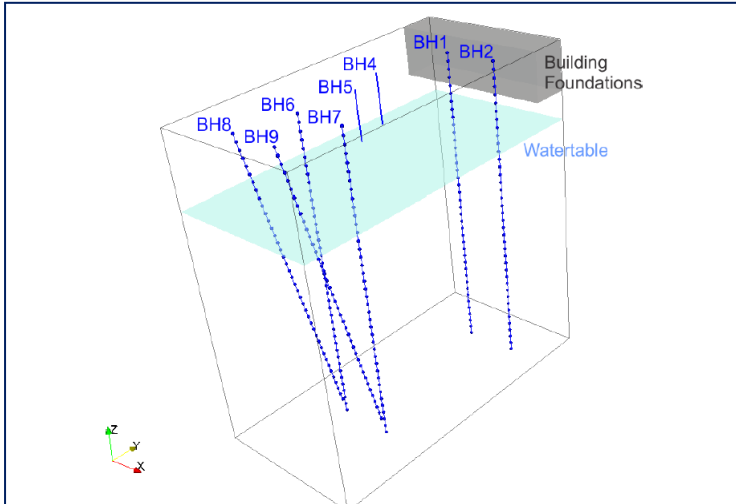


Figure 1: Borehole layout for the Sellafield simulated contaminant plume monitoring experiment.

Bipole-bipole measurements were made between the four

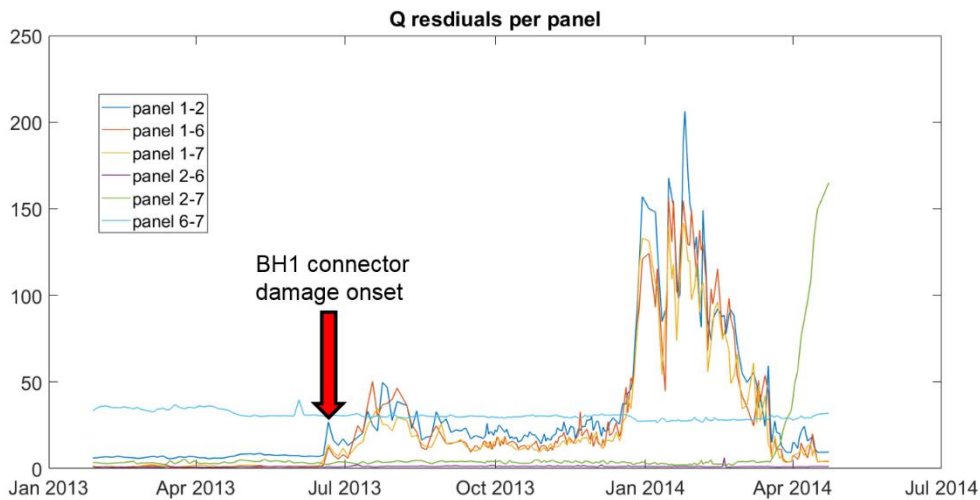


Figure 2: Q residuals for panels between numbered pairs of boreholes. The Q residuals describe the misfit between the reciprocal errors at each time step and the PCA model. The PCA training period lasted until 28th April.

A Statistical Model for Addressing Uncertainty in the Assessment of Performance of Ageing Retaining Structures to Groundwater Inundation

V.Stephenson^{1,3}, C.Oates^{1,2}

¹ University College London, ² Newcastle University, ³ The Alan Turing Institute

Increased rainfall amounts in the UK are leading to a growing recurrence of groundwater flooding events. These cause problems for embedded and retaining structures through heightened and fluctuating pore water pressures, which lead to ground deformation and subsequent movement and damage to construction. These issues are especially significant in relation to ageing structures, where decay in the engineered system and subsequent loss of structural integrity exacerbates damage caused by groundwater effects. Uncertainty associated with the various components of the problem further compounds the risk. As a result, understanding and managing the risks for engineering structures in this context is complex, often requiring estimation of geotechnical parameters that lead to assumptive measures of performance.

Presented here is an analysis of groundwater conditions surrounding a tunnel retaining structure in central London. Various sources have been considered which offer groundwater level information, including direct borehole readings, piezometer monitoring and secondary data including rainfall accumulations. These are appraised in order to produce a site-specific probabilistic groundwater fluctuation profile. The method considers traditional groundwater modelling approaches including Gaussian-distribution assessment, and compares this with monitored groundwater levels to produce a Bayesian distribution derived from site-based prior knowledge. Sources of uncertainty in the output are discussed in relation to the physical context of the site, and the geotechnics of the problem.

The groundwater model is then used for the assessment of an historic, ageing retaining wall to failure by sliding and overturning. A statistical approach is also taken to the definition of the key engineering parameters in the mechanical models defining the sliding and overturning failure mechanisms, including those of the wall and surrounding ground. Application of random fields is first used to define the problem and this is later improved using a Bayesian approach to predict and describe the wall construction and behaviour, considering its degraded state.

The final output of the work is a statistically derived model for failure of an idealised retaining structure in the context of uncertain geotechnical conditions. The work is derived from both anecdotal data and accepted statistical techniques, and as a result produces a solution which can be both robustly computed and referenced against a specific geographical location. The objective of the work is to describe and evidence a simple technique for overcoming uncertainty in groundwater flood risks to retaining structures, and through this promote the use of larger datasets for the improvement of site specific groundwater fluctuation prediction in the future.

Facilitating the effective application of analogue databases to reservoir models

Marijn van Cappelle^{1,*}, Simon Lomas², Luca Colombera³, Marco Patacci³, Alwin ten Hove¹, Sarah Cobain², Ben Meyer² & Viki O'Connor²

¹ *Petrotechnical Data Systems B.V., Lange Kleiweg 10, 2288 GK, Rijswijk, The Netherlands*

² *Petrotechnical Data Systems Ltd, Chester House, 81-83 Fulham High Street, London SW6 3JA, United Kingdom*

³ *School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, United Kingdom*

*contact: marijn.vancappelle@pds.nl

In the petroleum industry, reservoir models of oil and gas fields are built in order to estimate hydrocarbon volumes in place and forecast hydrocarbon production. Subsurface datasets used for conditioning models consist mainly of 1D well data: inferring the three-dimensional architecture of the inter-well volume involves significant uncertainty. It is widely acknowledged that applying rock outcrop analogues to reservoir models helps reducing uncertainty through the inclusion of geological realism.

In recent years the compilation of many analogue databases has been undertaken. Often these databases are constructed, maintained and used within petroleum companies. In other cases, the databases are vehicles for transferring data obtained in the context of academic research to sponsoring industry consortia. However, integrating data from these databases into workflows in development and production teams has proven difficult. Hurdles preventing geoscientists querying and applying data include: 1) the wide range in training and experience of staff, 2) time constraints, 3) constraints in the number of reservoir models which can be built and run, 4) practical limitations in the detail/resolution of reservoir models, and 5) the lack of quantification for all the parameters used in reservoir modelling.

We present a method (Ava Clastics) designed to overcome these hurdles of effectively integrating outcrop analogue data in reservoir modelling workflows. An online portal has been developed that makes it possible to quickly filter relevant outcrop analogues contained in relational databases of fluvial (FAKTS), shallow-marine (SMAKS) and deep-marine (DMAKS) clastic depositional systems. These relational databases store data on sedimentary units at multiple scales, describing their geometries, their relations with surrounding elements, their hierarchical organization, and their lithological heterogeneity. All the analogues contained in the databases are classified on the geological controls of the depositional systems, contextual information, and metadata, allowing users to select analogues that are most suitable in application to the reservoir at hand. The filtered analogue data is visualised graphically in real time, to enable the user to review database outputs and optimize the selection of inputs to a reservoir model.

Subsequently, the filtered analogue data are automatically parameterized to match the requirements of different stochastic facies modelling algorithms, thanks to workflows for translation of analogue data into model inputs. Where available, the analogue data is directly used as parameters in the modelling algorithm. When parameters are not directly available from suitable analogues, empirical relationships that are established in the scientific literature are used to calculate input parameters. After review, the parameterized composite analogue can be directly imported into the reservoir modelling package.

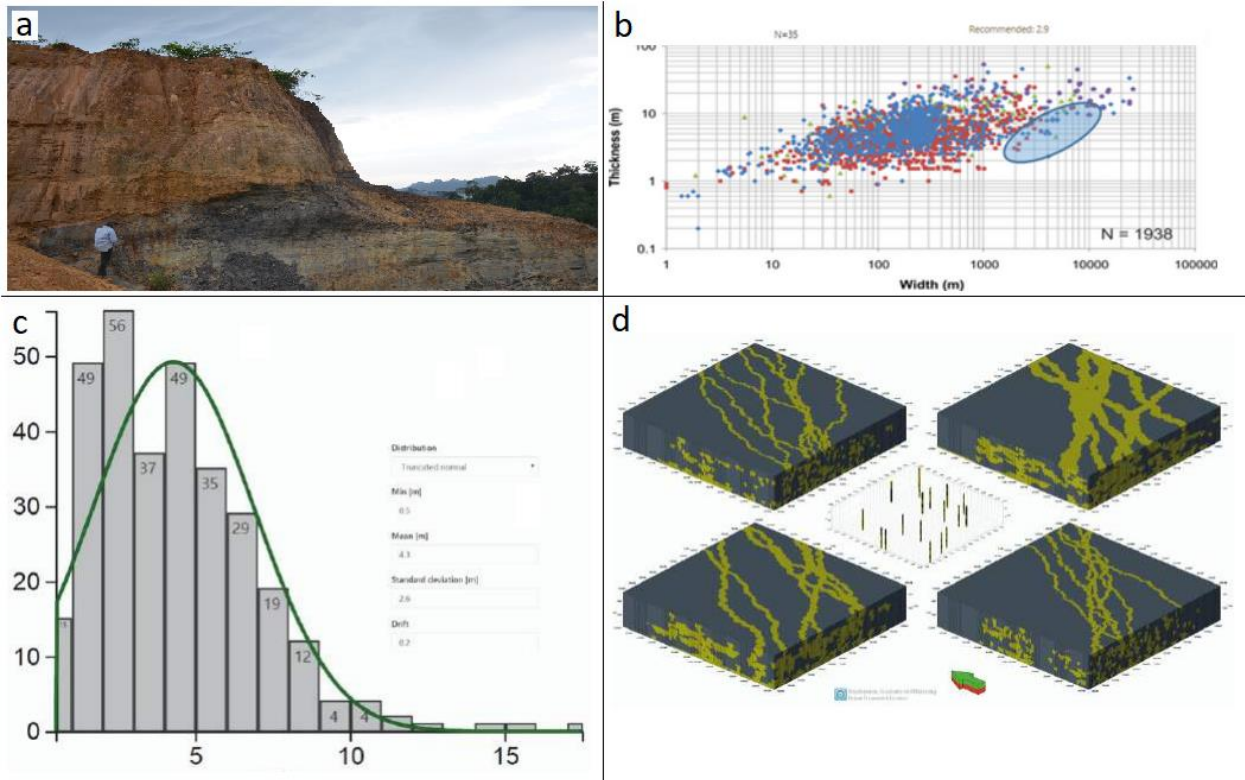


Figure. a) Rock outcrop analogues help in reducing uncertainty in inferring rock properties in between well. b) Analogue data is stored in databases. c) Parametrization of analogue data for use in stochastic facies modelling. d) Application of the parametrised analogue data to reservoir models.

A New Approach to Communicating the Science of Natural Hazard Processes in the Terrestrial and Marine Environment using Virtual Reality. An Erasmus Key Action 2 Funded Project.



Malcolm Whitworth¹, Mel Krokos², Derek Rust¹, Paraskevi Nomikou³, Alessandro Tibaldi⁴, Ugo Becciani⁵, Benjamin van Wyk de Vries⁶, Danilo Reitano⁷, Efthimios Lekkas⁸, Nicholas Bilham⁹, James Brown¹⁰, Robert Lovas¹¹, Frederique Metzeldard¹², Serafim Poulos¹³, Lee Jones¹⁴.

¹ School of Earth and Environmental Sciences, University of Portsmouth, UK.

² School of Creative and Cultural Industries, University of Portsmouth, UK.

³ Geology and Geoenvironment, National and Kapodistrian University of Athens, Greece.

⁴ Department of Earth and Environmental Sciences, University of Milano-Bicocca, Italy.

⁵ Istituto Nazionale Di Astrofisica (INAF), Sicily, Italy.

⁶ Universite Clermont Auvergne, Clermont-Ferrand, France.

⁷ Istituto Nazionale Di Geofisica E Vulcanologia (INGV), Italy.

⁸ Earthquake Planning and Protection Organization, Greece.

⁹ The Geological Society of London, UK.

¹⁰ Fugro Limited, UK.

¹¹ Obuda University, Budapest, Hungary.

¹² Institut de Physique du Globe de Paris, France.

¹³ Hellenic Oceanographers Association, Greece.

¹⁴ British Geological Survey (BGS), UK.

Contact: malcolm.whitworth@port.ac.uk

Project website: <https://sites.google.com/port.ac.uk/3dtelc/>

Recent surveys have shown that there is a significant demand for highly-skilled professionals in the environmental and geoscience sectors, who can be innovative, creative and have particular skills in advanced 3D spatial analysis, multi-disciplinarily, data management and informatics, mathematical and numerical skills, fieldwork and observations skills, earth observation and image analysis. The University of Portsmouth is the lead institution on a three-year trans-European Erasmus+ Key Action 2 project (1 September 2017 – 31 August 2020), comprising 13 academic, research and industrial institutions from across Europe including the UK, Greece, Italy, Hungary and France. It will focus on the combined study of natural hazards in terrestrial and marine environments, through a programme of data sharing, scientific and technical collaboration and ultimately curriculum development at postgraduate level to help students develop advanced skills in these areas.

The aim of the project is to develop a series of very high resolution 3D virtual reality models of different natural hazards observed in the onshore and offshore environments using a mixture of data acquired from airborne drone, high resolution imagery, LiDAR and submersible platforms; and develop a teaching toolkit that can be used in classrooms to teach about onshore and offshore environments as a continuum. Initially, the project is focused on developing datasets and teaching materials for postgraduate level university courses, but it will also have other applications in industry training settings, undergraduate courses and for use in outreach activities in both school and youth environments.

To facilitate this, the project aims to develop a series of freely available toolkits, datasets and teaching materials to allow students to navigate these environments using virtual reality headsets, map and measure features on the ground surface and seabed to simulate real field mapping activities; and then export features for further analysis. The project outputs will support terrestrial field courses and subsea terrain evaluation by providing ready to use virtual reality models, navigation and mapping tools for the virtual environment, waypoints and feature location data, videos and training materials, curriculum framework documentation and

associated teaching materials. These outputs will be open and freely available to all, including industry, academic, school and youth organisations, if you would like to know more about this project, or would like invitations to meetings and disseminations events, then please contact malcolm.whitworth@port.ac.uk.

Burlington House Fire Safety Information

If you hear the Alarm

Alarm Bells are situated throughout the building and will ring continuously for an evacuation. Do not stop to collect your personal belongings.

Leave the building via the nearest and safest exit or the exit that you are advised to by the Fire Marshal on that floor.

Fire Exits from the Geological Society Conference Rooms

Lower Library:

Exit via main reception onto Piccadilly, or via staff entrance onto the courtyard.

Lecture Theatre

Exit at front of theatre (by screen) onto Courtyard or via side door out to Piccadilly entrance or via the doors that link to the Lower Library and to the staff entrance.

Main Piccadilly Entrance

Straight out door and walk around to the Courtyard.

Close the doors when leaving a room. **DO NOT SWITCH OFF THE LIGHTS.**

Assemble in the Courtyard in front of the Royal Academy, outside the Royal Astronomical Society. Event organizers should report as soon as possible to the nearest Fire Marshal on whether all event participants have been safely evacuated.

Please do not re-enter the building except when you are advised that it is safe to do so by the Fire Brigade.

First Aid

All accidents should be reported to Reception and First Aid assistance will be provided if necessary.

Facilities

The ladies toilets are situated in the basement at the bottom of the staircase outside the Lecture Theatre.

The Gents toilets are situated on the ground floor in the corridor leading to the Arthur Holmes Room.

The cloakroom is located along the corridor to the Arthur Holmes Room.

Ground Floor Plan of the Geological Society, Burlington House, Piccadilly

