

Machine learning for paleo-geology and mineral exploration: a spatiotemporal odyssey

Rohitash Chandra
School of Mathematics and Statistics
UNSW Sydney

ARC ITTC Data Analytics for Resources and Environment
Seminar Series
14th December, 2021



Overview

1. Coupling deep learning with Plate Tectonics models for mineral exploration
2. Paleoclimatology and paleo-geology with Bayesian machine learning
3. Convolutional Neural Networks for Lithological Mapping via Remote Sensing
4. Data Augmentation with Generative Adversarial Learning
5. Uncertainty quantification with Bayesian deep learning



Prof. Dietmar Muller
Former ARC Laureate Fellow,
EarthByte Group, University of
Sydney



Julian Diaz-Rodriguez
Master of Science, University of Sydney

Coupling deep learning with Plate Tectonics models for mineral exploration

J Diaz-Rodriguez, RD Muller, **R Chandra**.
Predicting the emplacement of Cordilleran porphyry copper systems using a spatio-temporal machine learning model. *Ore Geology Reviews* 137, 2021: 104300



EarthBYTE
Building a Virtual Earth



UNSW
THE UNIVERSITY OF NEW SOUTH WALES

Data Analytics for Resources and Environments, Australian Research Council - Industrial Transformation Training Centre, Australia



THE UNIVERSITY OF
SYDNEY

Introduction

Porphyry copper (Cu) systems occur along magmatic belts derived in subduction zones. Our current understanding of their formation is restricted to observations from the overriding plate, resulting in a knowledge gap in terms of processes occurring in convergence zones through time.

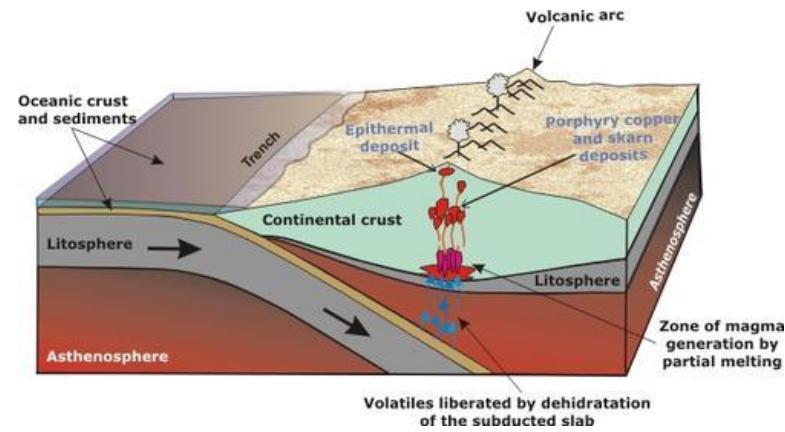
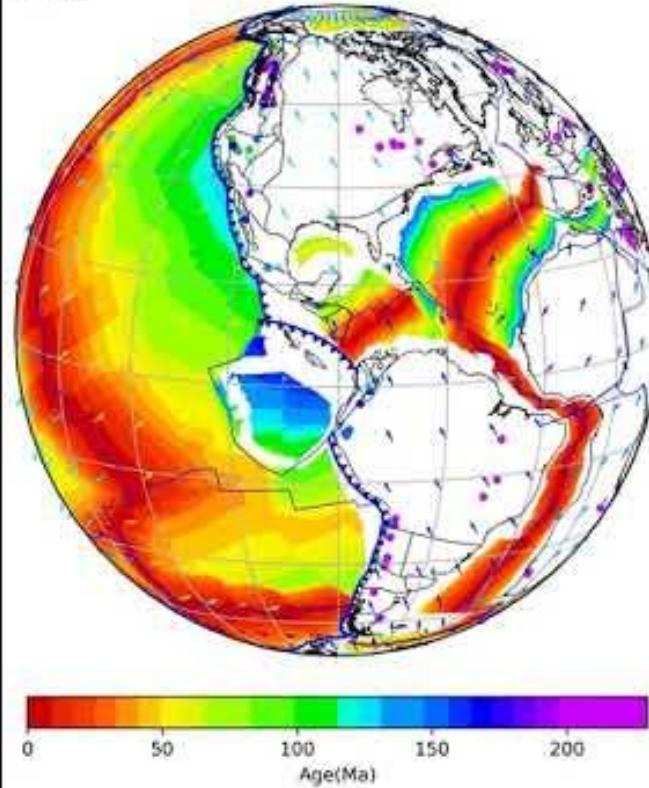
An association between key tectonic processes and the timing and location of porphyry Cu systems requires linking geological observations to plate tectonic subduction models.

We connect the evolution of subduction zones and downgoing slab properties with the history of porphyry ore deposition across the Americas by using a spatio-temporal machine learning approach.

We use these spatio-temporal properties to apply a wide range of prominent machine learning methods and show the results in terms of accuracy of predictions on the test dataset.

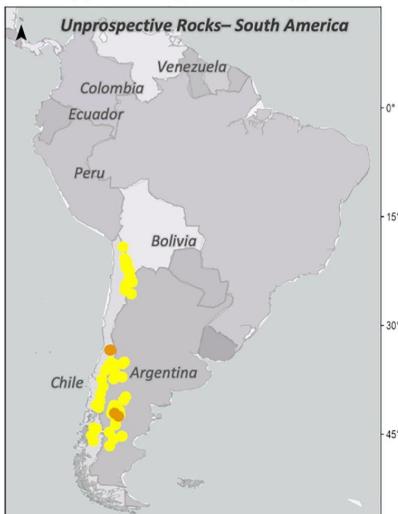
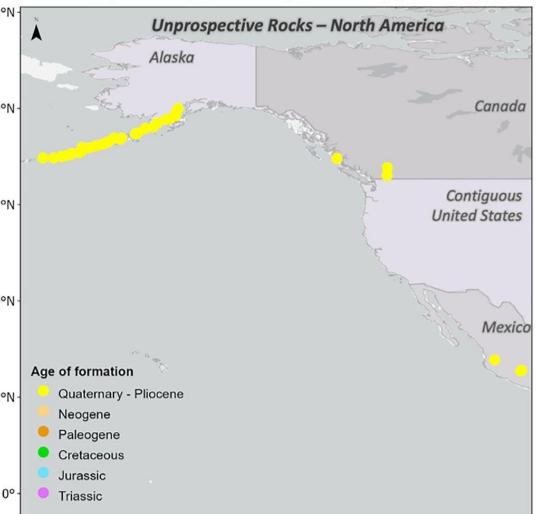
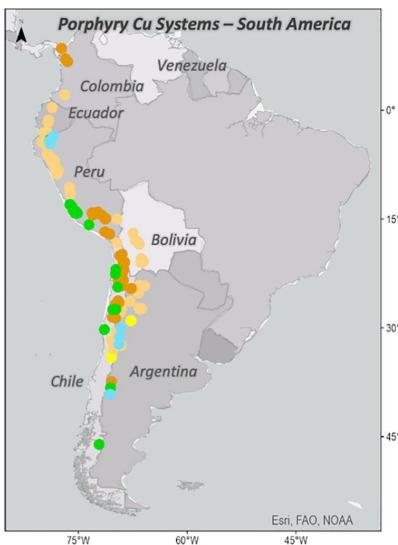
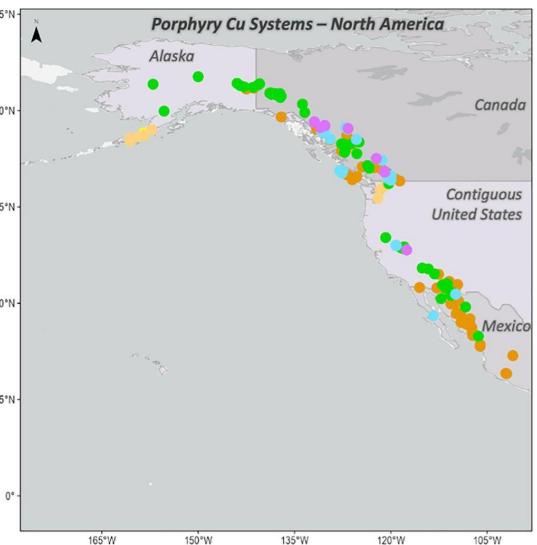
We incorporate plate tectonic subduction models with machine learning models for the generation of copper mineralization prediction maps.

96 Ma



Schematic diagram of the tectonic setting of porphyry, epithermal and skarn deposits. Source:

https://www.geo.fu-berlin.de/en/v/geolearning/mountain_building/resources/po...



The GPlates homepage features a large, semi-transparent world map in the background. Overlaid on the map is a 3D globe showing geological reconstructions, with various plates colored in shades of blue, red, and yellow, and movement arrows indicating plate motion. In the upper left, the GPlates logo is displayed. The main title "GPlates" is prominently shown in a large, bold, dark font. Below the title is a detailed description of the software's capabilities. A row of four small globe icons represents different reconstruction stages. A green button labeled "Download Now" is located at the bottom left. The overall design is clean and professional, emphasizing the software's focus on geological data visualization.

GPlates

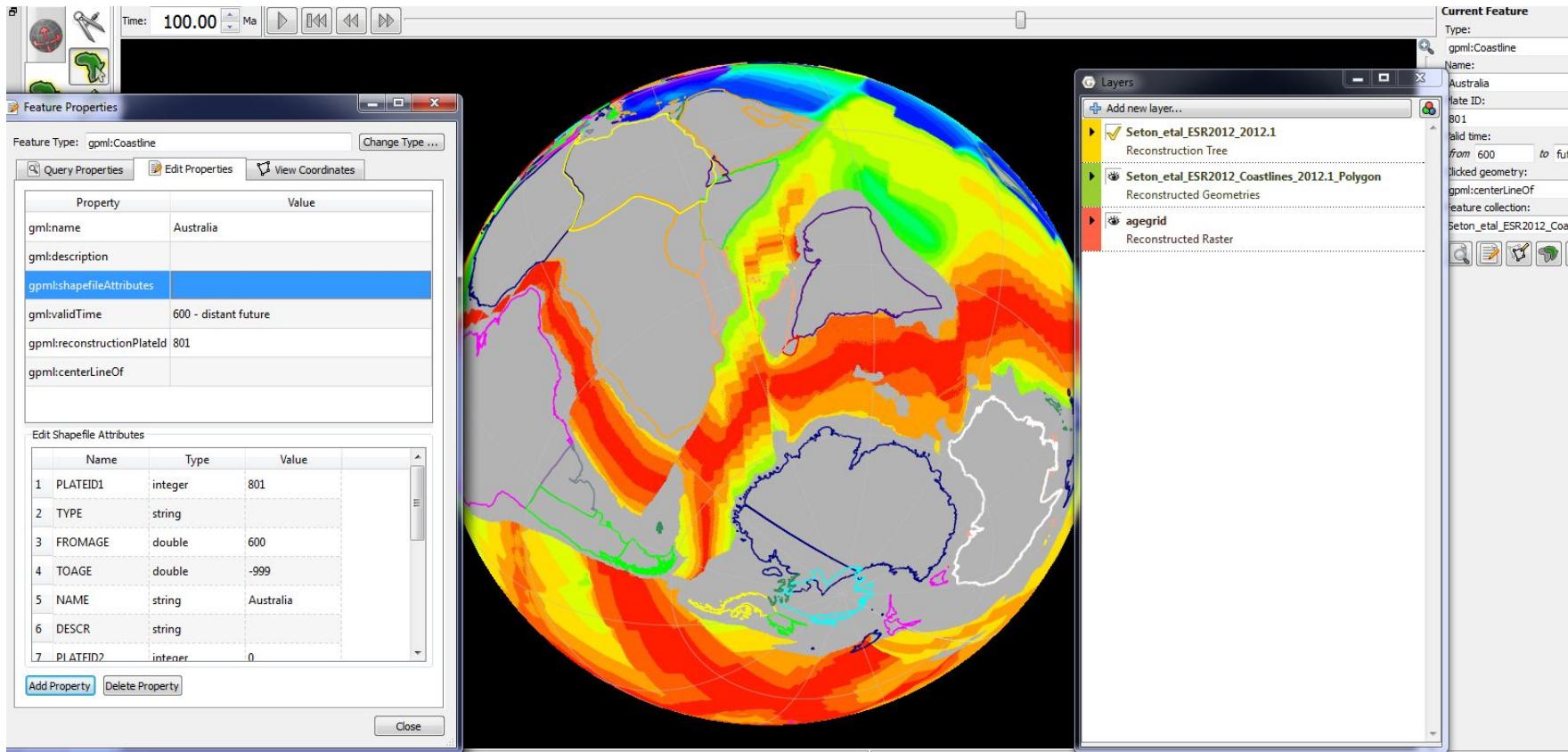
GPlates is a plate tectonics program. Manipulate reconstructions of geological and paleogeographic features through geological time. Interactively visualize vector, raster and volume data. PyGPlates is the GPlates Python library. Get fine-grained access to GPlates functionality in your Python scripts.

Latest release v2.2

[Download Now](#)

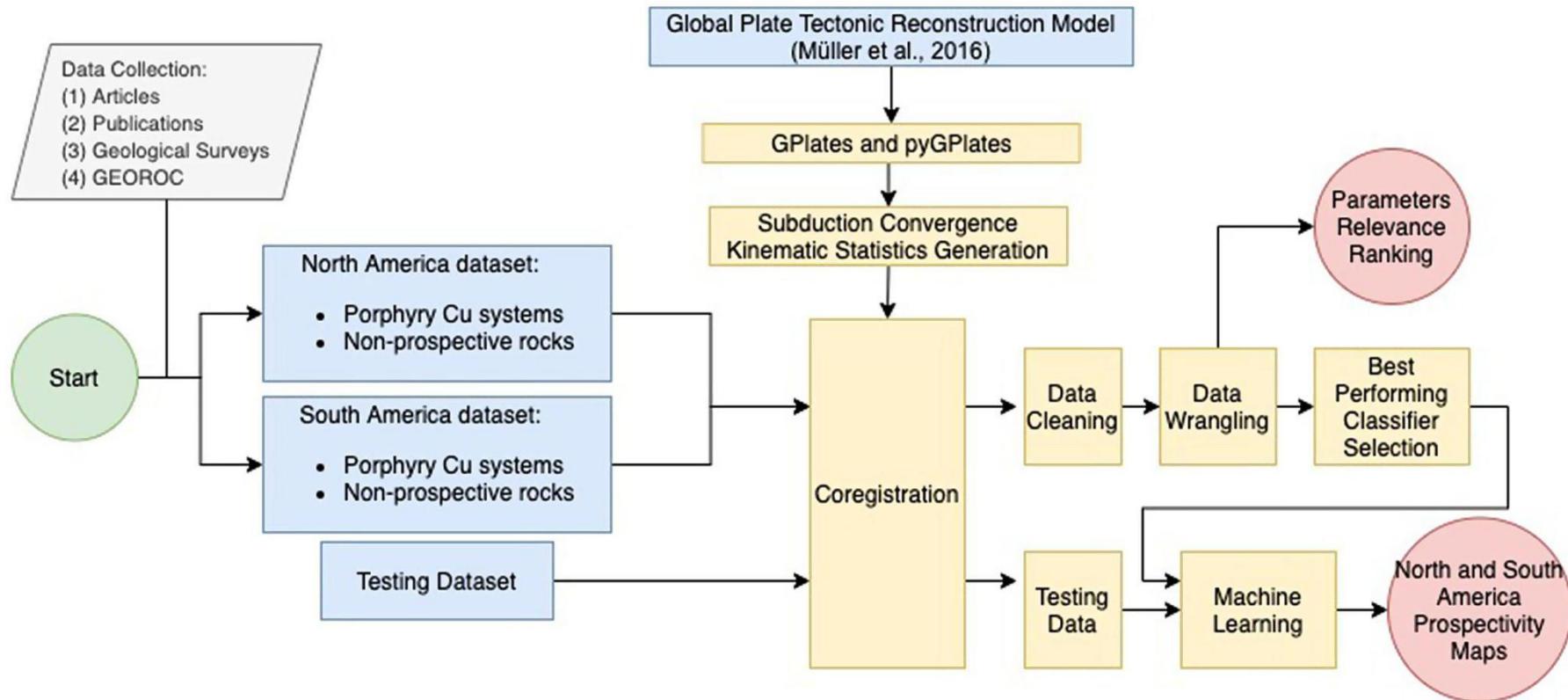


GPlates Software

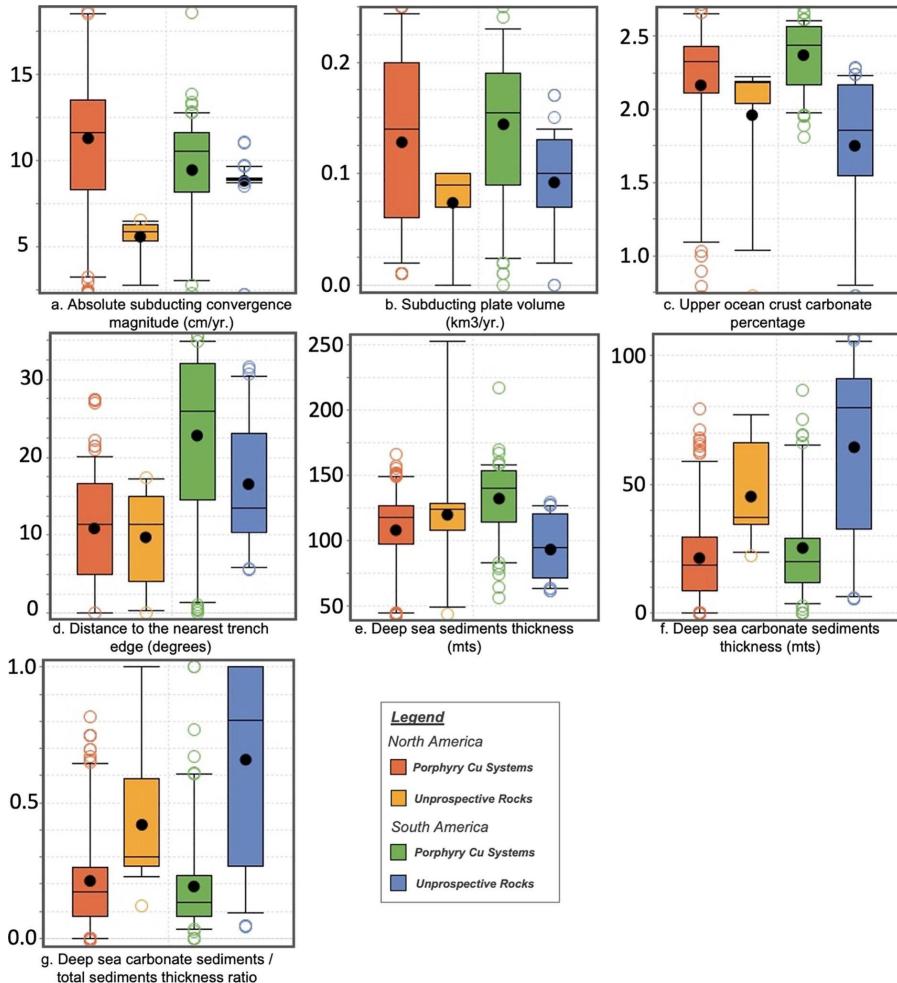


Source: <https://en.wikipedia.org/wiki/GPlates>

Framework: Coupling Plate Tectonics with Machine Learning



Data Visualization



Mean values and standard deviation of selected parameters related to the dynamic evolution of subduction zones at the time of formation of porphyry Cu systems and non-prospective rocks in North America and South America. * Kilometres calculated using a great circle distance formula for comparison purposes (in B&W).

Parameters	North America			South America				
	Porphyry Cu systems		Unprospective Rocks	Porphyry Cu systems		Unprospective Rocks	Mean	SD
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Subducting convergence magnitude – Absolute vector (cm/yr.)	11.30	4.87	5.59	1.03	9.46	3.16	8.84	1.08
Subducting plate volume (km ³ /yr.)	0.13	0.07	0.07	0.03	0.14	0.06	0.09	0.04
Upper ocean crust carbonate percentage	2.17	0.47	1.96	0.42	2.37	0.22	1.75	0.42
Distance to the nearest trench edge (degrees) (kilometres*)	10.88 (1210)	7.01 (780)	9.74 (1083)	5.72 (636)	22.79 (2534)	10.65 (1184)	16.56 (1841)	8.04 (894)
Deep sea total sediments thickness (m)	108.30	31.25	120.15	42.92	132.43	26.09	93.49	23.00
Deep sea carbonate sediments thickness (m)	21.23	17.06	45.24	16.63	25.21	30.38	64.39	32.95
Deep sea carbonate sediment / total sediment thickness ratio	0.21	0.19	0.42	0.22	0.19	0.18	0.66	0.36

Correlation matrix using Spearman's correlation score for the dynamic evolution parameters related to the formation of porphyry Cu systems for South America. Strong correlations are defined as Spearman's correlation scores above 0.75. Grey colours are for scores below 0.75. Red shades of colour range between 0.75 and 1.0. Blue colours are for the main diagonal of the matrix. (in colour).

South America	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Subd. Conv. Magn. - Absolute Vect.	1.00	0.91	-0.01	0.48	-0.38	-0.15	-0.88	0.83	-0.06	0.17	-0.11	0.06	0.80	0.11	0.12
2. Subd. Conv. Magn. - Orthogonal Vect.	0.91	1.00	-0.14	0.37	-0.48	-0.01	-0.83	0.91	-0.18	0.18	-0.10	0.04	0.85	0.03	0.03
3. Subd. Conv. Magn. - Parallel Vect.	-0.01	-0.14	1.00	0.26	0.16	-0.69	0.09	-0.10	0.96	0.03	0.02	0.12	-0.06	0.20	0.18
4. Trench Plate Magn. - Absolute Vect.	0.48	0.37	0.26	1.00	-0.75	-0.23	-0.12	0.11	0.29	0.42	-0.12	0.44	0.49	0.56	0.58
5. Trench Plate Magn. - Orthogonal Vect.	-0.38	-0.48	0.16	-0.75	1.00	-0.19	0.10	-0.17	0.10	-0.48	0.09	-0.34	-0.54	-0.43	-0.46
6. Trench Plate Magn. - Parallel Vect.	-0.15	-0.01	-0.69	-0.23	-0.19	1.00	0.13	-0.12	-0.48	-0.03	0.07	-0.09	-0.07	-0.15	-0.10
7. Down-going Plate Magn. - Absolute Vect.	-0.88	-0.83	0.09	-0.12	0.10	0.13	1.00	-0.92	0.16	-0.01	0.15	0.08	-0.70	0.08	0.08
8. Down-going Plate Magn. - Orthogonal Vect.	0.83	0.91	-0.10	0.11	-0.17	-0.12	-0.92	1.00	-0.18	0.04	-0.16	-0.07	0.77	-0.10	-0.13
9. Down-going Plate Magn. - Parallel Vect.	-0.06	-0.18	0.96	0.29	0.10	-0.48	0.16	-0.18	1.00	0.06	0.06	0.17	-0.08	0.23	0.24
10. Distance to the nearest trench edge	0.17	0.18	0.03	0.42	-0.48	-0.03	-0.01	0.04	0.06	1.00	-0.24	0.58	0.45	0.62	0.61
11. Deep-sea carbonate sediment thickness	-0.11	-0.10	0.02	-0.12	0.09	0.07	0.15	-0.16	0.06	-0.24	1.00	-0.31	-0.23	-0.19	-0.17
12. Upper-ocean crust carbonate percentage	0.06	0.04	0.12	0.44	-0.34	-0.09	0.08	-0.07	0.17	0.58	-0.31	1.00	0.36	0.82	0.81
13. Subducting plate volume	0.80	0.85	-0.06	0.49	-0.54	-0.07	-0.70	0.77	-0.08	0.45	-0.23	0.36	1.00	0.40	0.39
14. Sea floor age	0.11	0.03	0.20	0.56	-0.43	-0.15	0.08	-0.10	0.23	0.62	-0.19	0.82	0.40	1.00	0.99
15. Deep-sea sediment total thickness	0.12	0.03	0.18	0.58	-0.46	-0.10	0.08	-0.13	0.24	0.61	-0.17	0.81	0.39	0.99	1.00

Results

Cross validation results for different supervised machine learning classifiers. Cross validation calculated using a ten-fold with a shuffle split on 80/20 training/ testing ratio. Range of colours just for indicative purposes from Gray (up to 85%) and shades of yellow (85%) to red (100%). The complete data used for cross validation is included in supplementary data. (in colour).

	North America		South America	
	Mean	SD	Mean	SD
Support Vector (RBF k.)	98.3%	1.7%	97.9%	2.0%
Random Forest	95.4%	4.3%	97.4%	1.7%
Multi-layer Perceptron	96.5%	3.9%	96.0%	1.6%
Gaussian Process	92.8%	3.7%	94.7%	3.3%

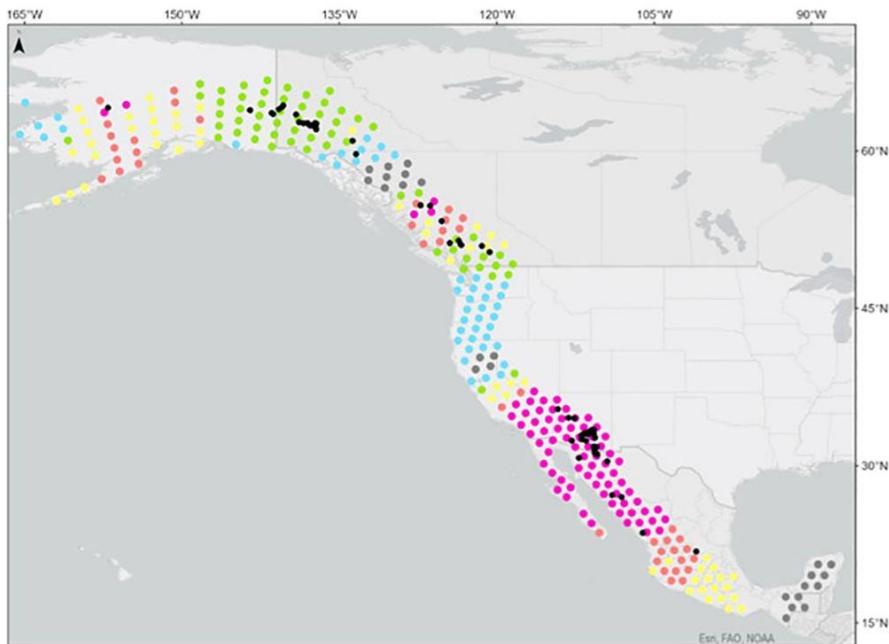
Confusion matrix performance metrics for different supervised machine learning classifiers for North America. (in colour).

Classifier	North America					Recall	Precision	Accuracy	F1	AUC score
	True Negative	False Positive	False Negative	True Positive						
Support Vector (RBF k.)	9	1	0	47		100.0%	97.9%	98.2%	98.9%	100.0%
Multi-layer Perceptron	9	1	0	47		100.0%	97.9%	98.2%	98.9%	100.0%
Gaussian Process	7	3	1	46		97.9%	93.9%	93.0%	95.8%	99.4%
Random Forest	9	1	2	45		95.7%	97.8%	94.7%	96.8%	98.7%

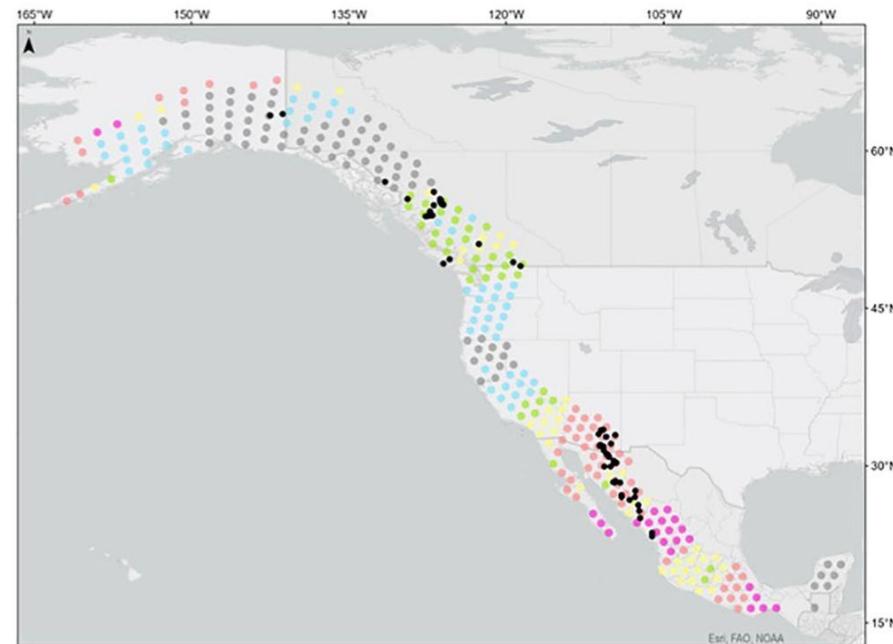
Confusion matrix performance metrics for different supervised machine learning classifiers for South America. (in colour).

Classifier	South America					Recall	Precision	Accuracy	F1	AUC score
	True Negative	False Positive	False Negative	True Positive						
Support Vector (RBF k.)	25	0	0	29		100.0%	100.0%	100.0%	100.0%	100.0%
Random Forest	24	1	1	28		96.6%	96.6%	96.3%	96.6%	99.2%
Multi-layer Perceptron	23	2	1	28		96.6%	93.3%	94.4%	94.9%	97.9%
Gaussian Process	23	2	2	27		93.1%	93.1%	92.6%	93.1%	95.7%

a. 80 – 60 Ma.



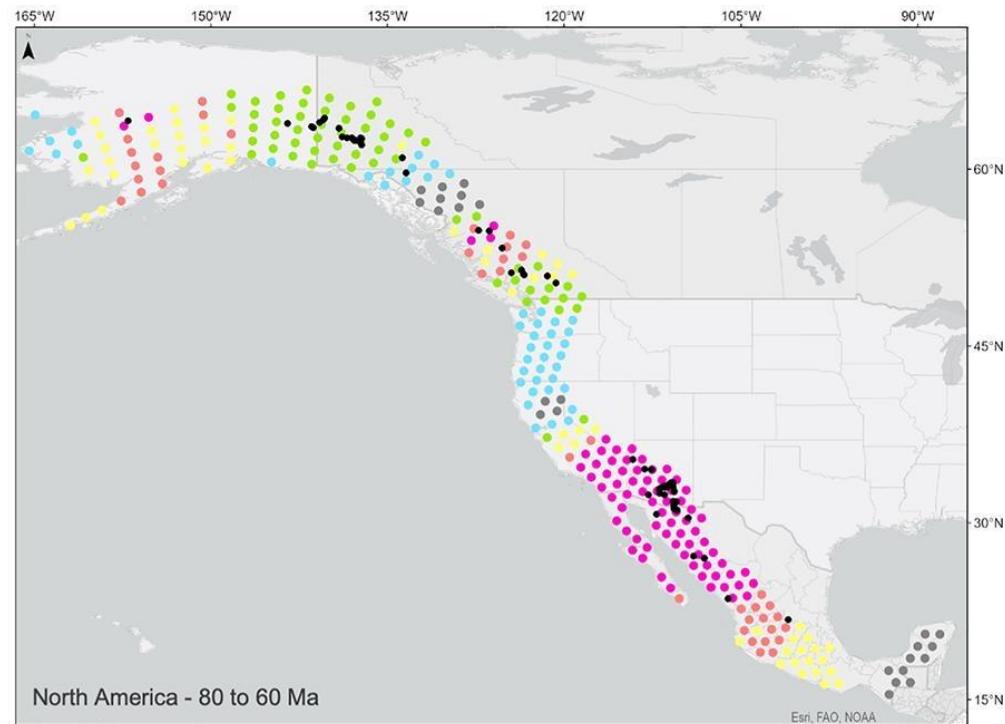
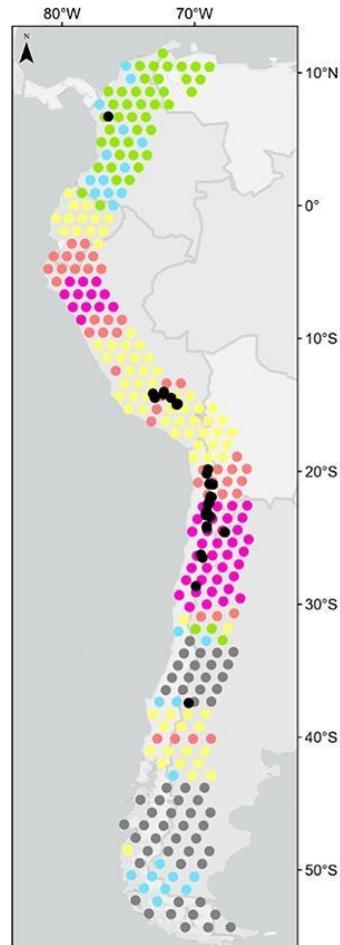
b. 60 – 40 Ma.



Legend

Score Prediction – Support Vector Classifier

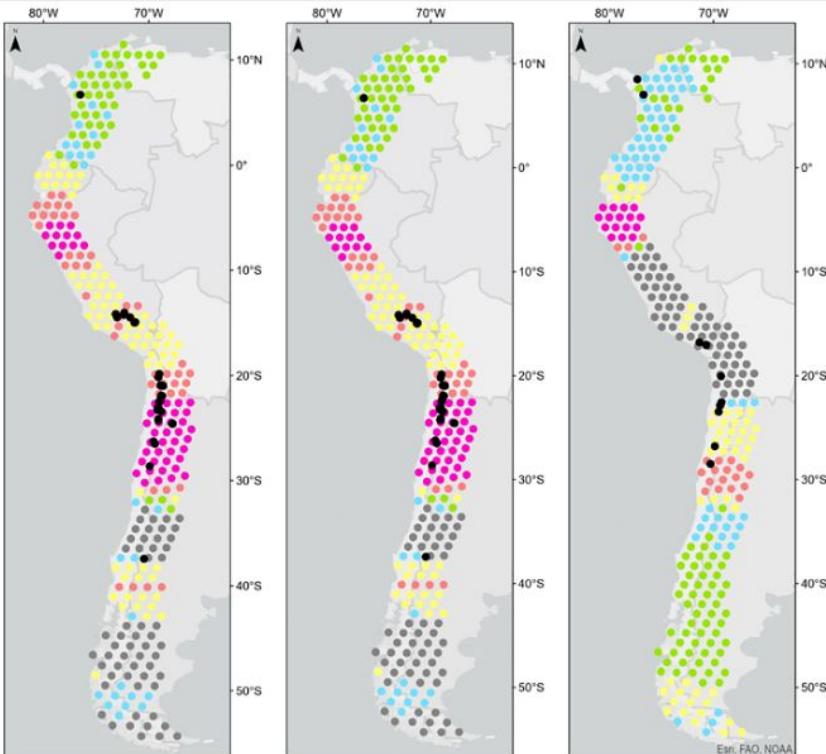
- **Very High Prospectivity**
- **High Prospectivity**
- **High – Moderate Prospectivity**
- **Low – Moderate Prospectivity**
- **Low Prospectivity**
- **Very Low Prospectivity**
- **Porphyry Cu Systems**



Legend

Score Prediction – Support Vector Classifier

- | | |
|-------------------------------|------------------------|
| • Very High Prospective | • Low Prospective |
| • High Prospective | • Very Low Prospective |
| • High – Moderate Prospective | • Porphyry Cu Systems |
| • Low – Moderate Prospective | |



a. 23 to 3 Ma

b. 47 to 27 Ma

c. 66 – 47 Ma

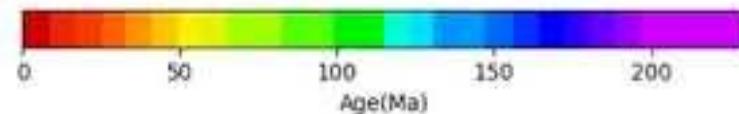
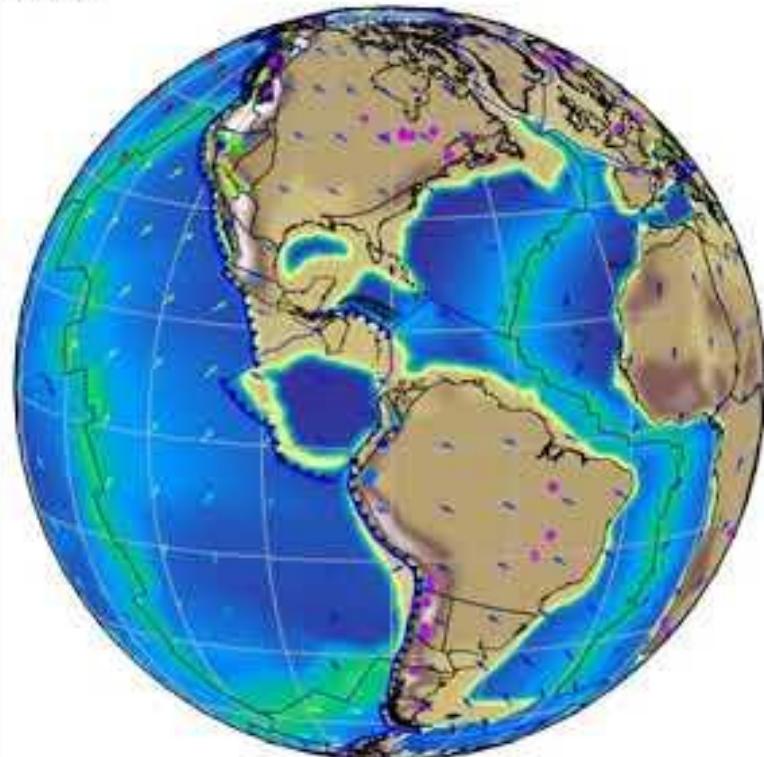
Legend

Score Prediction – Support Vector Classifier

- | | |
|---------------------------------|--------------------------|
| • Very High Prospectivity | • Low Prospectivity |
| • High Prospectivity | • Very Low Prospectivity |
| • High – Moderate Prospectivity | • Porphyry Cu Systems |
| • Low – Moderate Prospectivity | |

Fig. 6. Prospectivity map for porphyry Cu systems in South America using a support vector classifier with RBF kernel classifiers. (a) 23 to 3 Ma. (b) 47 to 27 Ma. (c) 66 to 47 Ma. (in colour).

68 Ma



Discussion

We propose a deep time spatio-temporal machine learning model to identify highly prospective areas and endowment ages for porphyry Cu mineralisation along the western Cordillera of North America and the Andes in South America.

This open-source workflow is based on pyGPlates which allows us to reconstruct in time and space the kinematic processes occurring at the boundary of convergent plates, including features of the downgoing plate such as deep sea sediment thickness and oceanic crustal or lithospheric properties. These parameters constrain the necessary geological processes to generate porphyry Cu systems under adequate overriding plate conditions.

We conclude that the most important of those parameters linked to the formation of porphyry Cu systems across North and South America is the absolute magnitude of convergence velocity. This magnitude is on average faster at the time when these systems formed as opposed to the rates related to the emplacement of non-prospective intrusions.

Paleoclimatology and paleo-geology with Bayesian machine learning

Chandra, R., Cripps, S., Butterworth, N., & Muller, R. D. (2021). Precipitation reconstruction from climate-sensitive lithologies using Bayesian machine learning. *Environmental Modelling & Software*, 139, 105002.



Prof. Dietmar Muller
Former ARC Laureate
Fellow, EarthByte Group,
University of Sydney

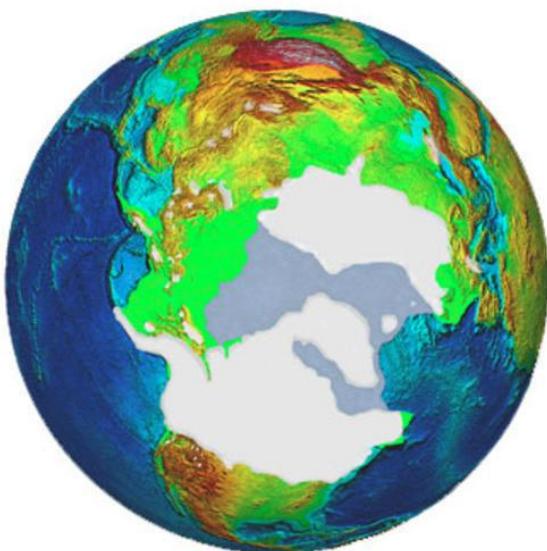


Prof. Sally Cripps
Former Director, Centre for
Translational Data Science, and
ITTC DARE, University of Sydney
Research Director, Data61

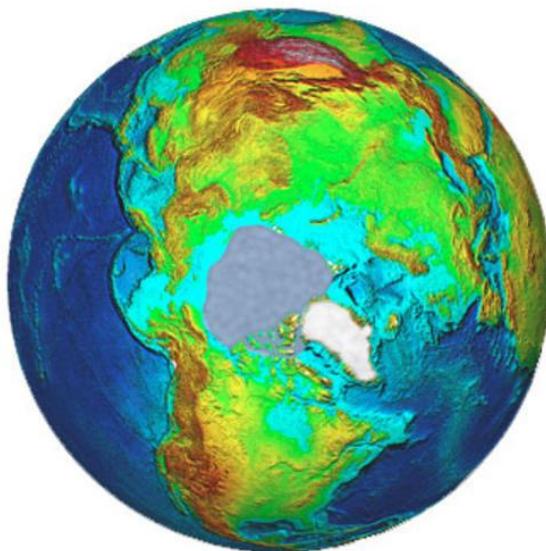


Dr. Nathaniel Butterworth
Senior Informatics Engineer
Sydney Informatics Hub
University of Sydney

18,000 Years Ago



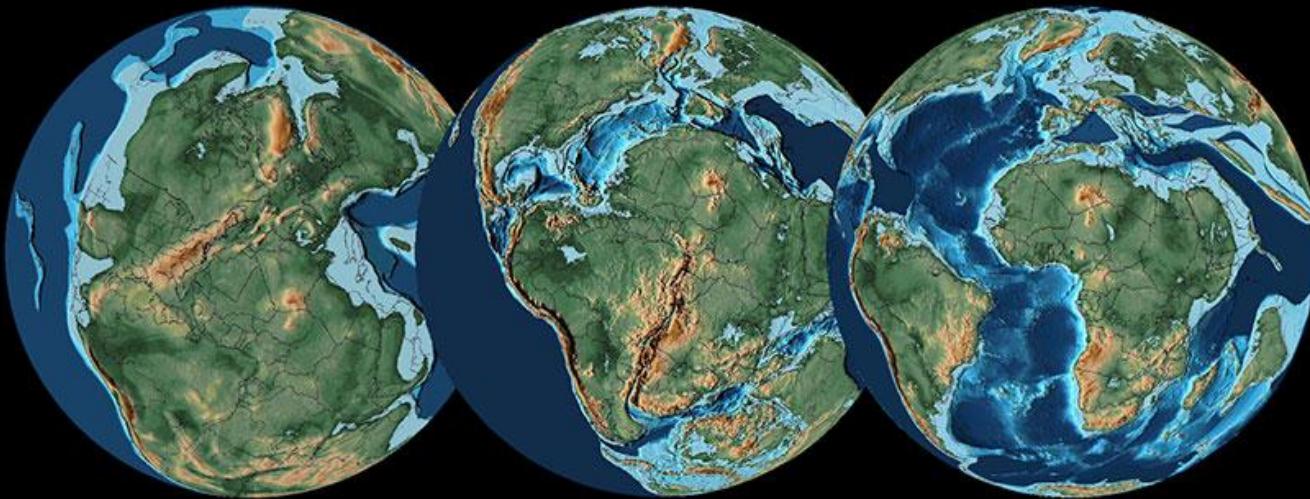
Modern Day



Glacial Ice Sea Ice

Northern Hemisphere Ice Coverage

NOAA National Centers for Environmental Information



During the Mesozoic, the fragmentation of Pangaea contributed to long-term climate trends, such as weaker seasonality and warmer global mean temperatures. Shown are paleogeographic reconstructions for (left to right) 250 million, 150 million, and 70 million years ago. Credit: Jan Landwehrs; data provided by Christopher Scotese

Lee, J. (2021), Simulating 195 million years of global climate in the Mesozoic, *Eos*, 102, <https://doi.org/10.1029/2021EO161439>. Published on 30 July 2021.

Background

Palaeoclimatology is the study of climates for which direct measurements were not taken and uses a variety of proxy methods from Earth and life sciences to obtain data , such as those preserved within rocks, sediments, boreholes, and ice sheets.

Paleogeology uses the principles and methods of geology to reconstruct the geological history of Earth and examines the vastness of geologic time, measured in billions of years. It covers changes in the Earth, gradual and sudden, over this deep time.

Geological timeframes

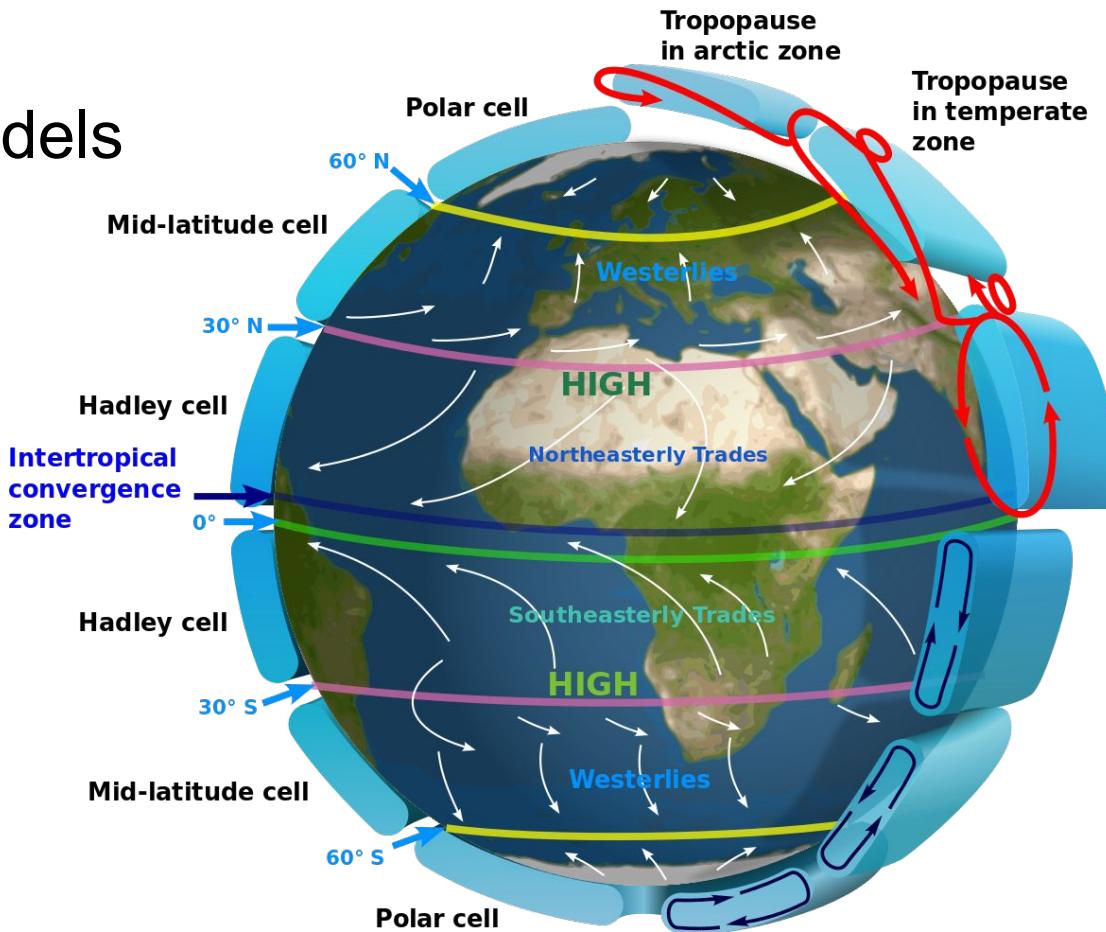
United States Geological Survey - Graham, Joseph, Newman, William, and Stacy, John, 2008, The geologic time spiral—A path to the past (ver. 1.1): U.S. Geological Survey General Information Product 58, poster, 1 sheet.

Available online at

<http://pubs.usgs.gov/gip/2008/58/>



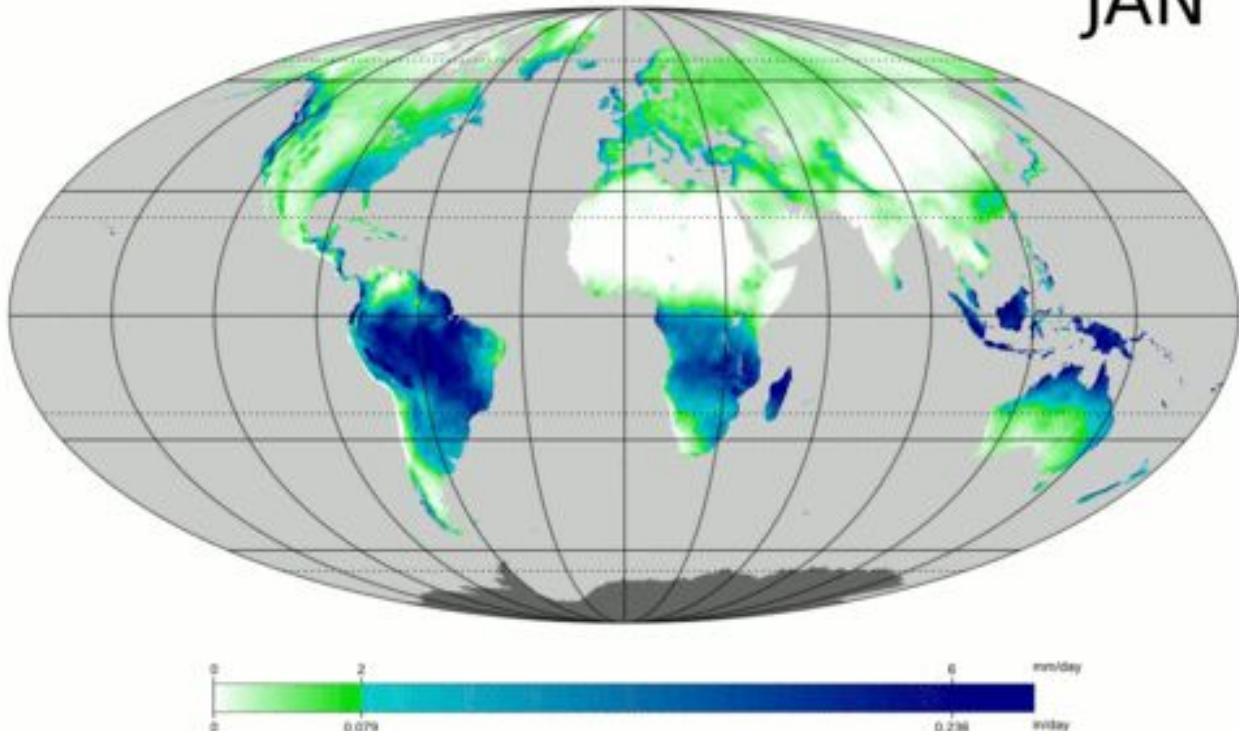
Global Circulation Models



Source: National Geographic: <https://www.nationalgeographic.org/photo/pervailing-winds/>

JAN

New, M., Lister, D., Hulme, M. and
Makin, I., 2002: A high-resolution
data set of surface climate over
global land areas. *Climate
Research* 21: 1–25



Overview

Although global circulation models (GCMs) have been used for the reconstruction of precipitation for selected geological time slices, there is a lack of a coherent set of precipitation models for the Mesozoic-Cenozoic period (the last 250 million years).

There has been dramatic climate change during this time period capturing a super-continent hothouse climate, and continental breakup and dispersal associated with successive greenhouse and ice-house climate periods.

We present an approach that links climate-sensitive sedimentary deposits such as coal, evaporites and glacial deposits to a global plate model, reconstructed paleo-elevation maps and high-resolution GCMs via Bayesian machine learning. We model the joint distribution of climate-sensitive sediments and annual precipitation through geological time, and use the dependency between sediments and precipitation to improve the models predictive accuracy.

Our approach provides a set of 13 data-driven global paleo-precipitation maps between 14 and 249 Ma, capturing major changes in long-term annual rainfall patterns as a function of plate tectonics, paleo-elevation and climate change at a low computational cost.

Simulated Data by GCMs

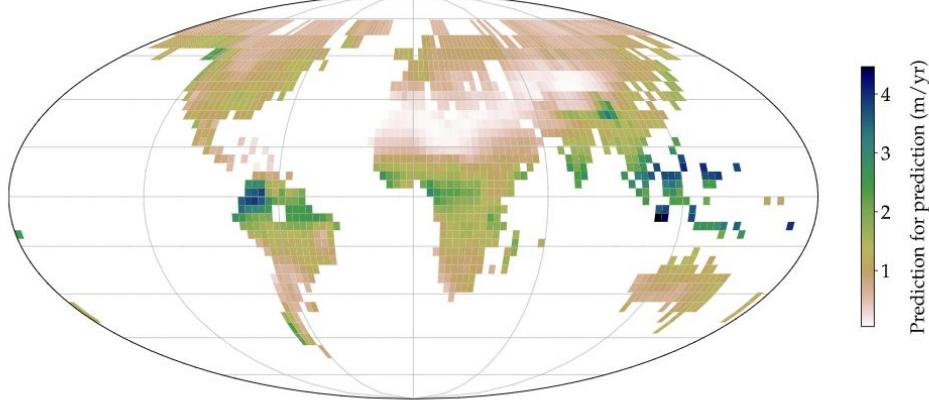


Fig. 1. Mid-Miocene precipitation model estimation ([Herold et al., 2011](#)).

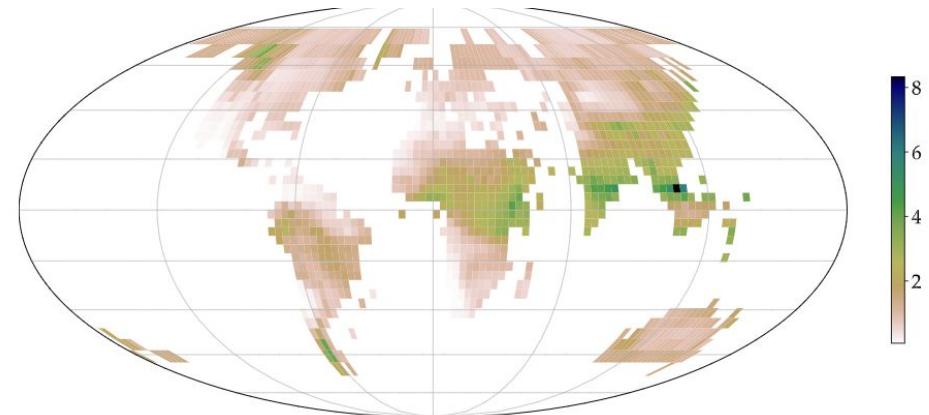


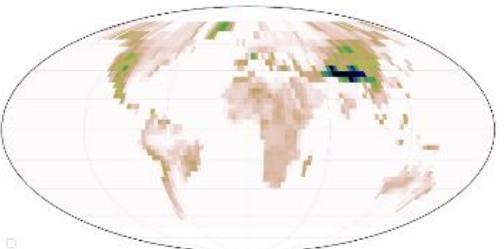
Fig. 2. Late Eocene precipitation model estimation ([Hutchinson et al., 2018](#)).

Lithology data

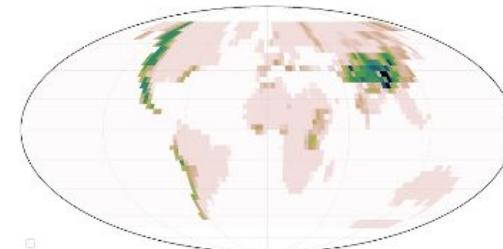
Data description showing reconstruction timeslices (Ma) showing given precipitation (precip.) simulated data (number of grid samples) for Miocene and Eocene ([Cao et al., 2019](#)). The unavailable precipitation data is shown as not applicable (n/a) which our model will estimate. The number of locations for available deposits for coal, evaporites (eva.) and glacial (gla.) is shown which adds up to the present (N_p) number of deposit grids.

Era	Period	Epoch/Age	Timespan (Ma)	Timeslice (Ma)	Total (N_t)	Present (N_{D_t})	Precip.	Coal	Eva.	Gla.
Cenozoic	Neogene	Miocene	23.0–5.3	14	1763	335	1763	241	86	8
		Oligocene	33.9–23.0	28	1761	282	n/a	229	52	1
		Eocene	47.8–33.9	38	1766	200	1766	146	50	4
	Palaeogene	Early Eocene (Ypresian)	56.0–47.8	51	1748	278	n/a	219	58	1
		Palaeocene	66.0–56.0	61	1653	163	n/a	120	42	1
Mesozoic	Cretaceous	Late Cretaceous (Coniacian–Maastrichtian)	89.8–66.0	77	1490	236	n/a	151	85	0
		Late Cretaceous (Albian–Turonian)	113.0–89.8	101	1628	252	n/a	170	82	0
		Early Cretaceous (Berriasian–Aptian)	145.0–113.0	129	1650	292	n/a	185	102	5
	Jurassic	Late Jurassic	164.0–145.0	154	1630	180	n/a	85	95	0
		Early and Middle Jurassic	201.0–164.0	182	1675	330	n/a	249	81	0
Triassic	Triassic	Late Triassic	237.0–201.0	219	1731	217	n/a	142	75	0
		Middle Triassic	247.0–237.0	242	1594	84	n/a	21	63	0
		Early Triassic	252.0–247.0	249	1548	73	n/a	24	49	0

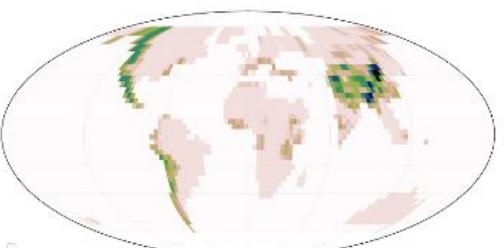
Paleo-elevation data



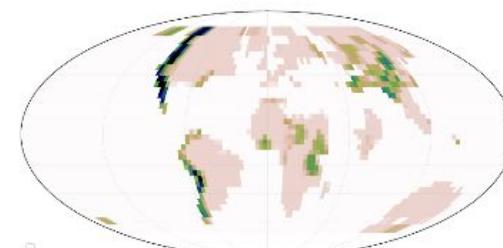
(a) 14 Ma



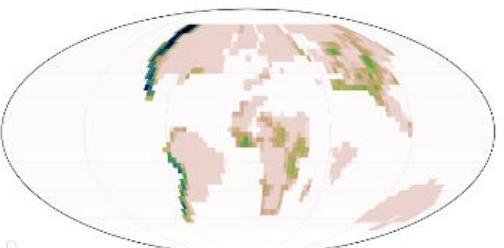
(b) 28 Ma



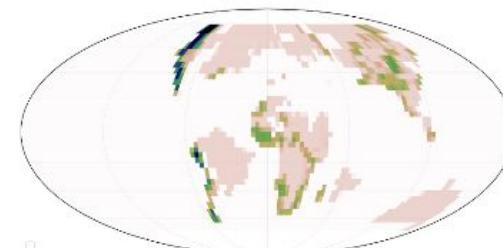
(c) 38 Ma



(d) 51 Ma



(e) 61 Ma



(f) 77 Ma

Data visualisation

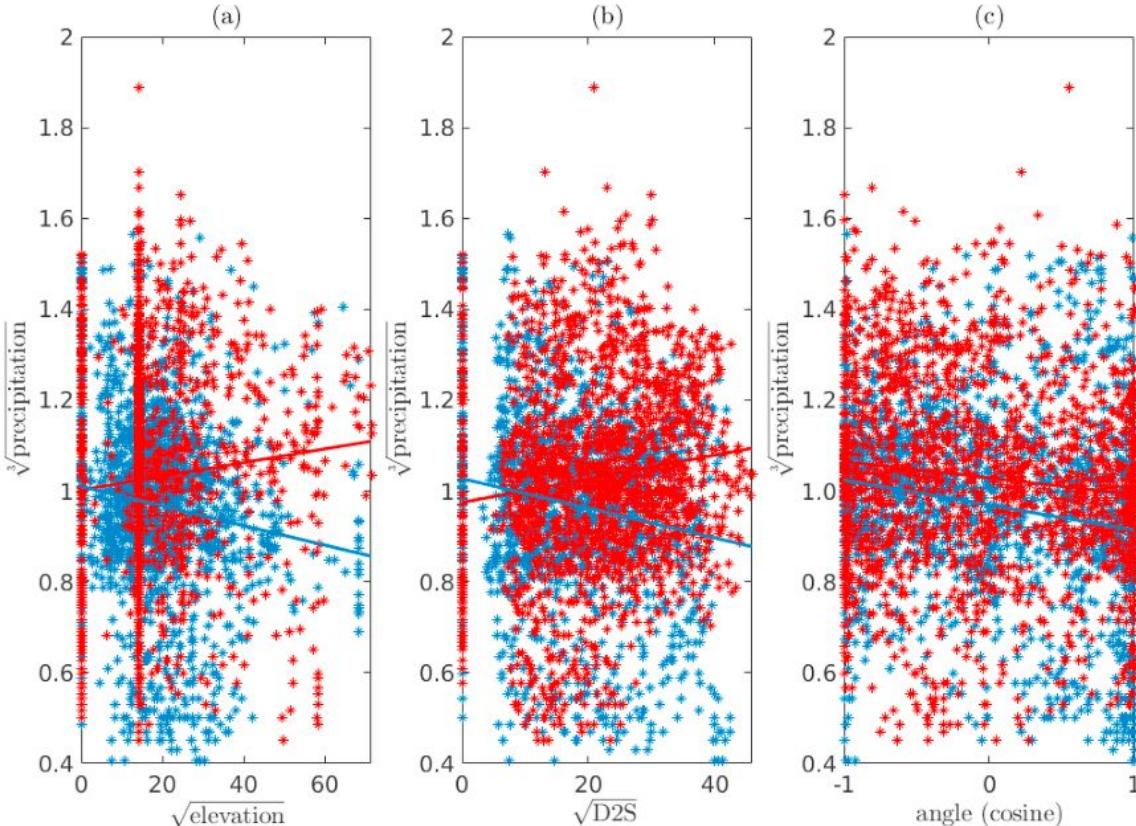
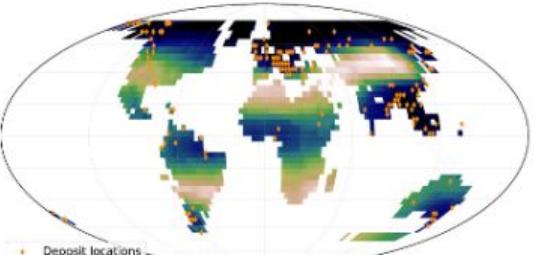


Fig. 5. Scatter plots for establishing the marginal relationship of the spatial covariates with respect to the precipitation in Miocene (blue) and Eocene (red) combined. The assumption of a linear relationship between $\sqrt[3]{\text{precipitation}}$ and the variables (a) ($\sqrt{\text{elevation}}$, (b) $\sqrt{\text{D2S}}$, and (c) angle (cosine) is reasonable.

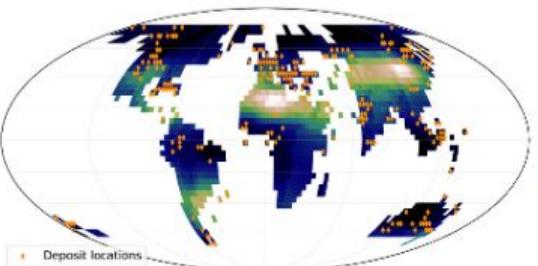
Results: coal



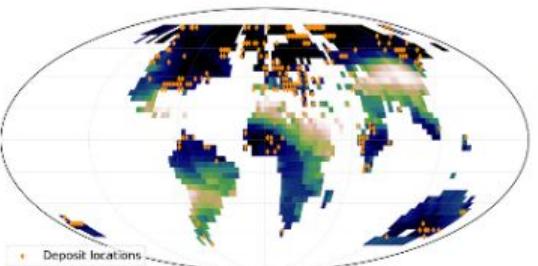
(a) 14 Ma



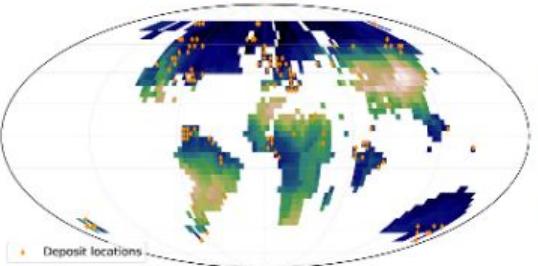
(b) 28 Ma



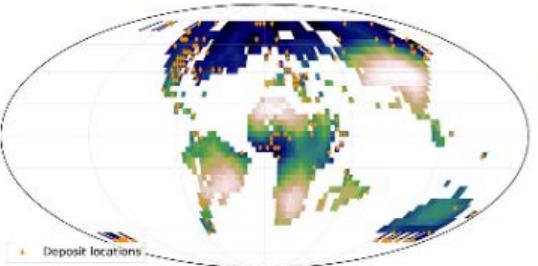
(c) 38 Ma



(d) 51 Ma



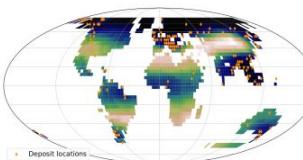
(e) 61 Ma



(f) 77 Ma



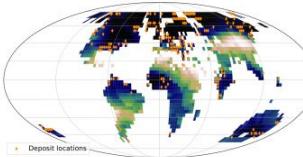
(a) 14 Ma



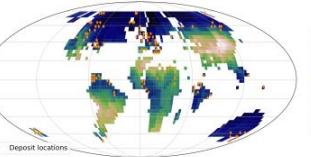
(b) 28 Ma



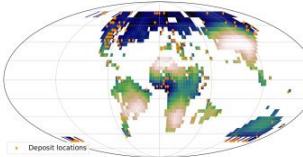
(c) 38 Ma



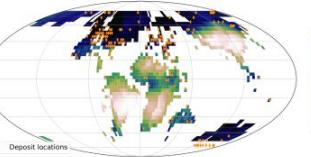
(d) 51 Ma



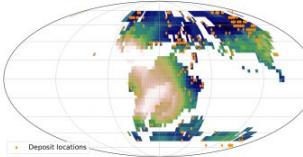
(e) 61 Ma



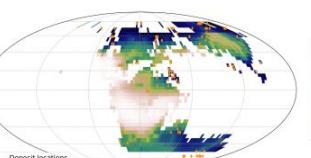
(f) 77 Ma



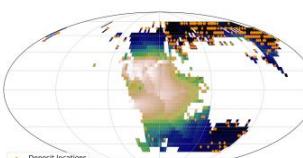
(g) 101 Ma



(h) 129 Ma

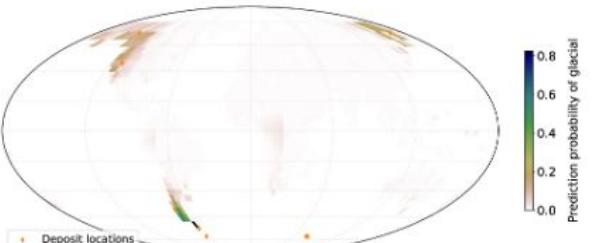


(i) 154 Ma

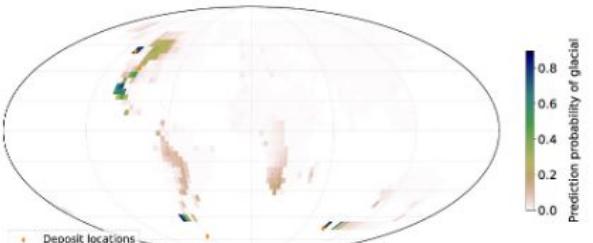


(j) 182 Ma

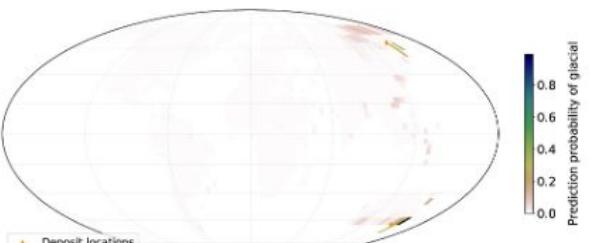
Results - glacial



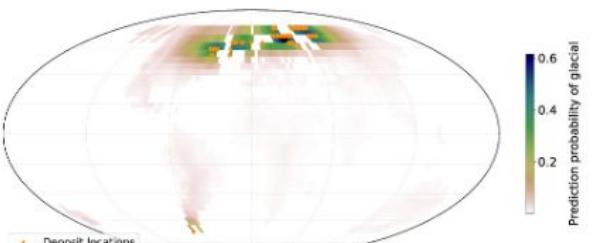
(a) 14 Ma



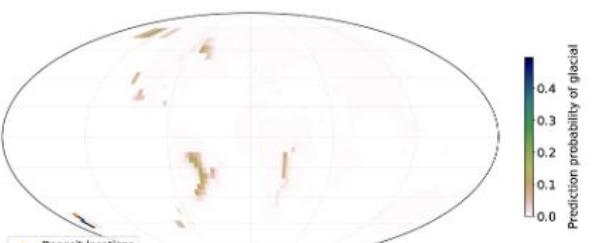
(b) 28 Ma



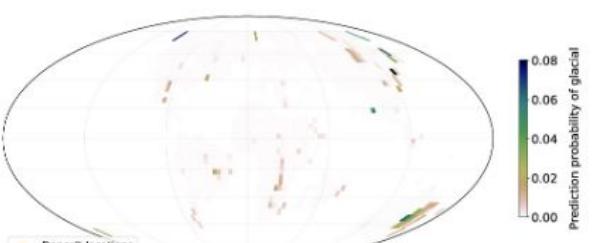
(c) 38 Ma



(d) 51 Ma

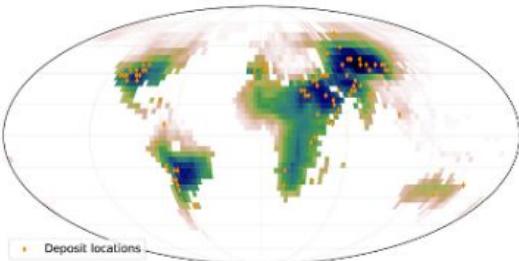


(e) 61 Ma

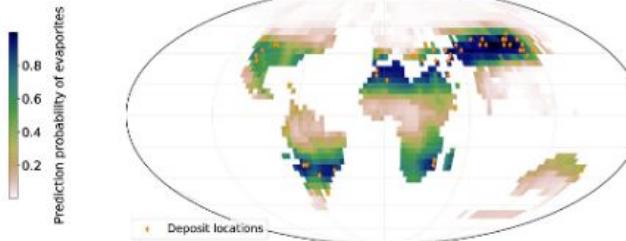


(f) 77 Ma

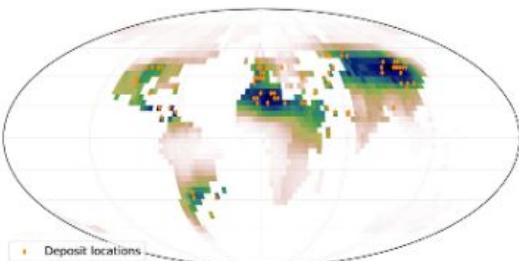
Results - evaporates



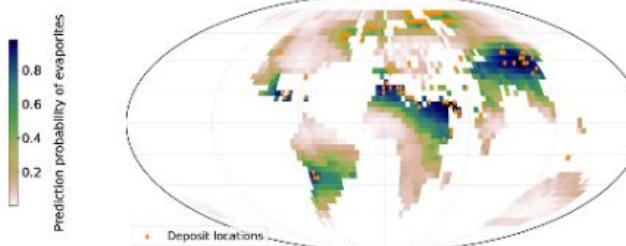
(a) 14 Ma



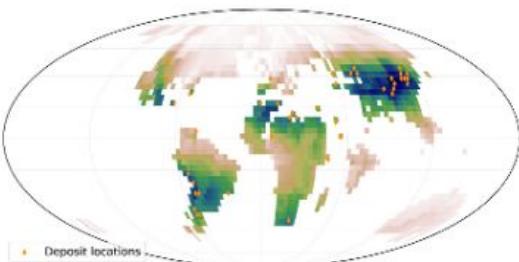
(b) 28 Ma



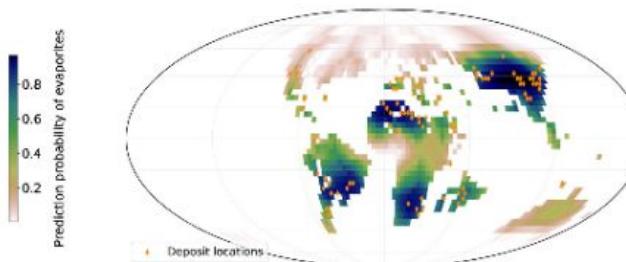
(c) 38 Ma



(d) 51 Ma

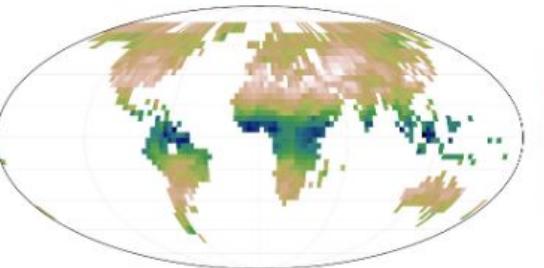


(e) 61 Ma

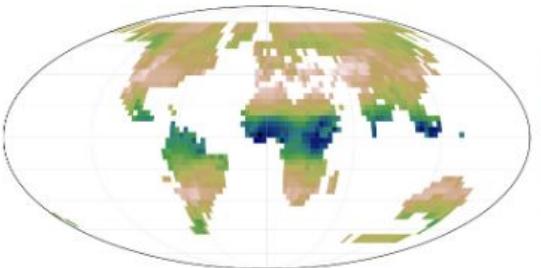


(f) 77 Ma

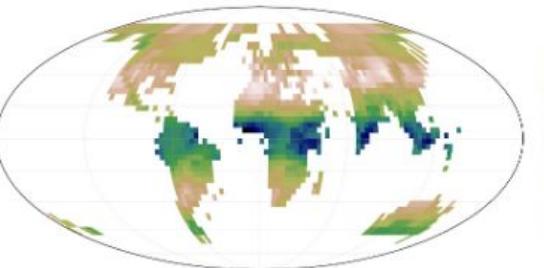
Results - precipitation



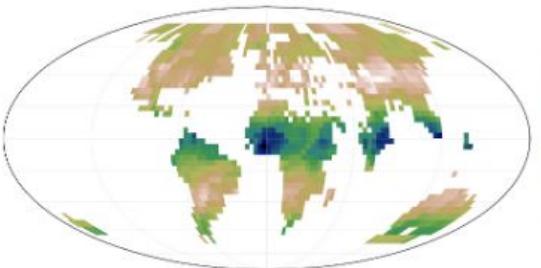
(a) 14 Ma



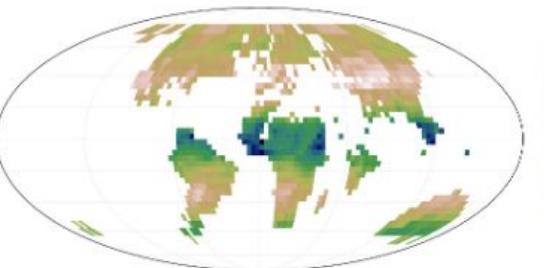
(b) 28 Ma



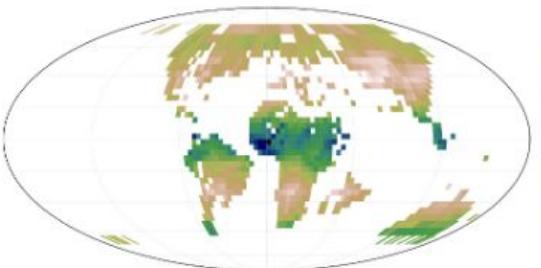
(c) 38 Ma



(d) 51 Ma



(e) 61 Ma



(f) 77 Ma

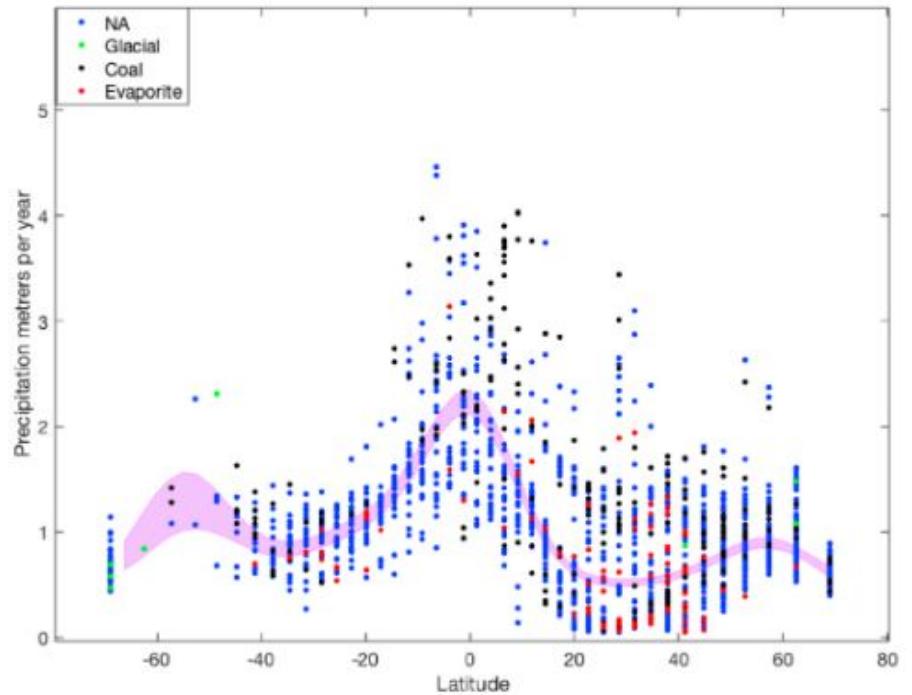


Fig. 12. We plot 95% credible intervals (shown in pink) for the annual median precipitation by the model as a function of latitude for the Miocene. The observations are colour coded according to deposit information; black for coal, red for evaporites, green for glacial and blue for not available (NA).

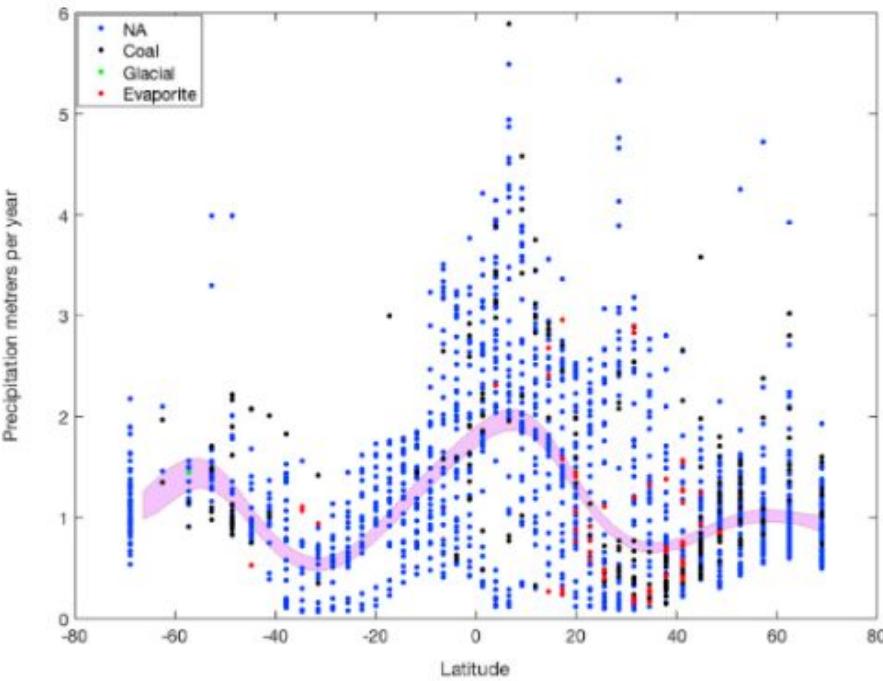


Fig. 13. We plot 95% credible intervals (shown in pink) for the annual median precipitation by the model as a function of latitude for the Eocene. The observations are colour coded according to deposit information; black for coal, red for evaporites, green for glacial, and blue for not available (NA).

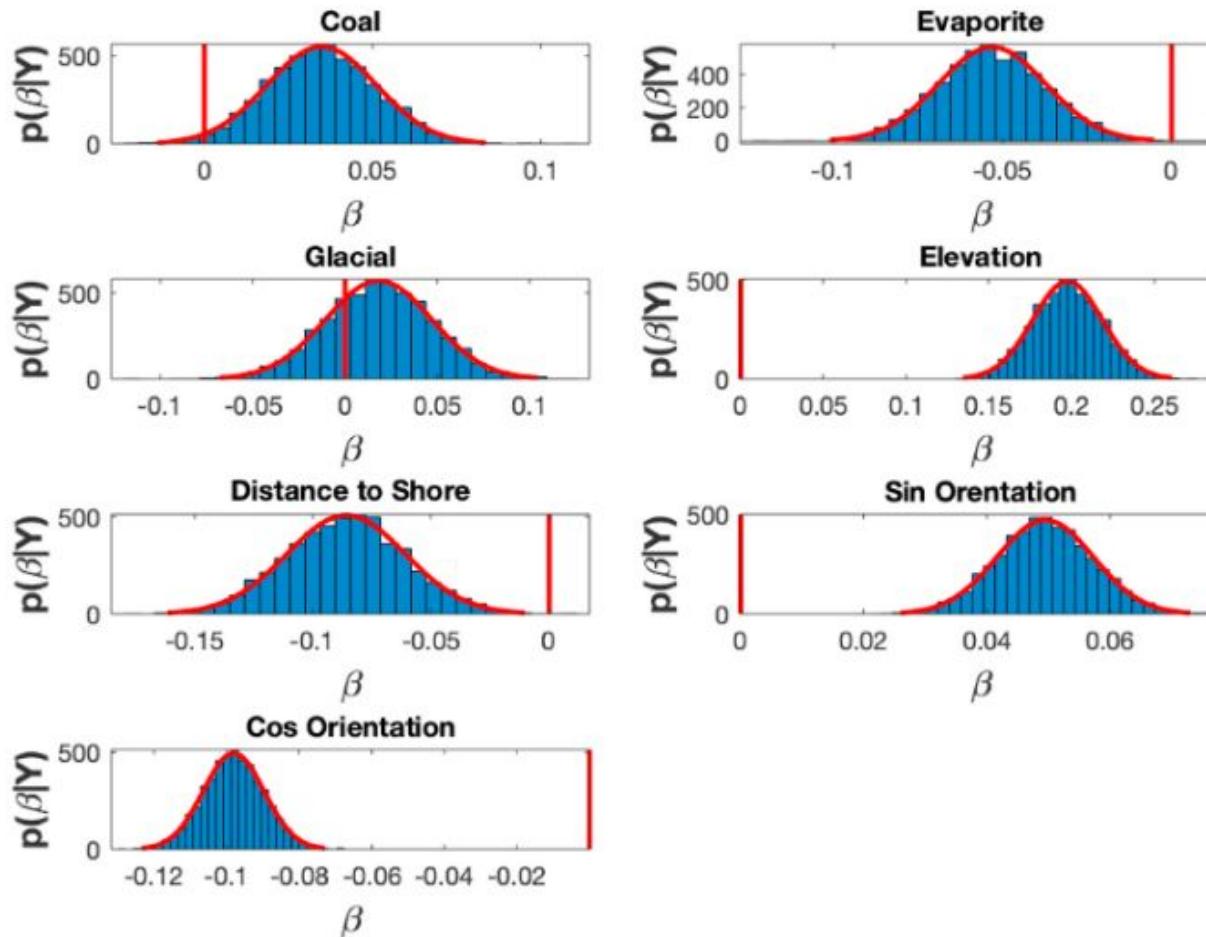
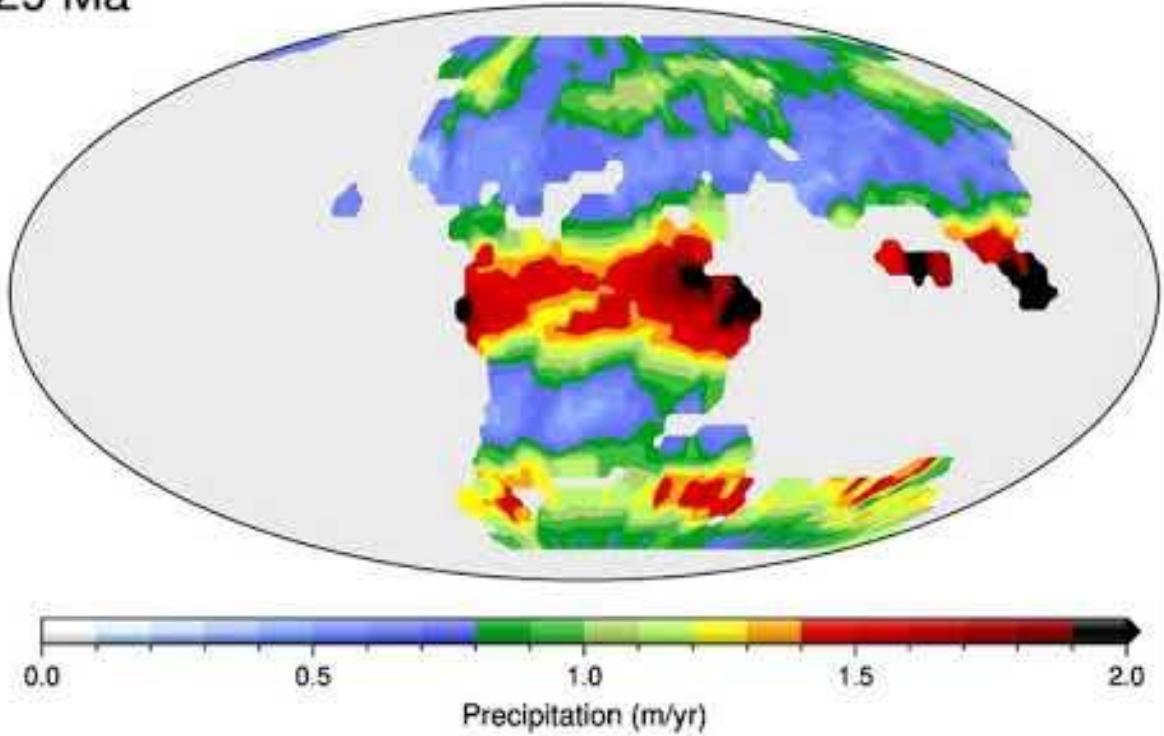


Fig. 18. Regression coefficients of the respective input features used for predicting precipitation using data from the Miocene and Eocene as training. These regression coefficient represent the marginal impact of that variable on precipitation after controlling for the other variables.

129 Ma



Discussion

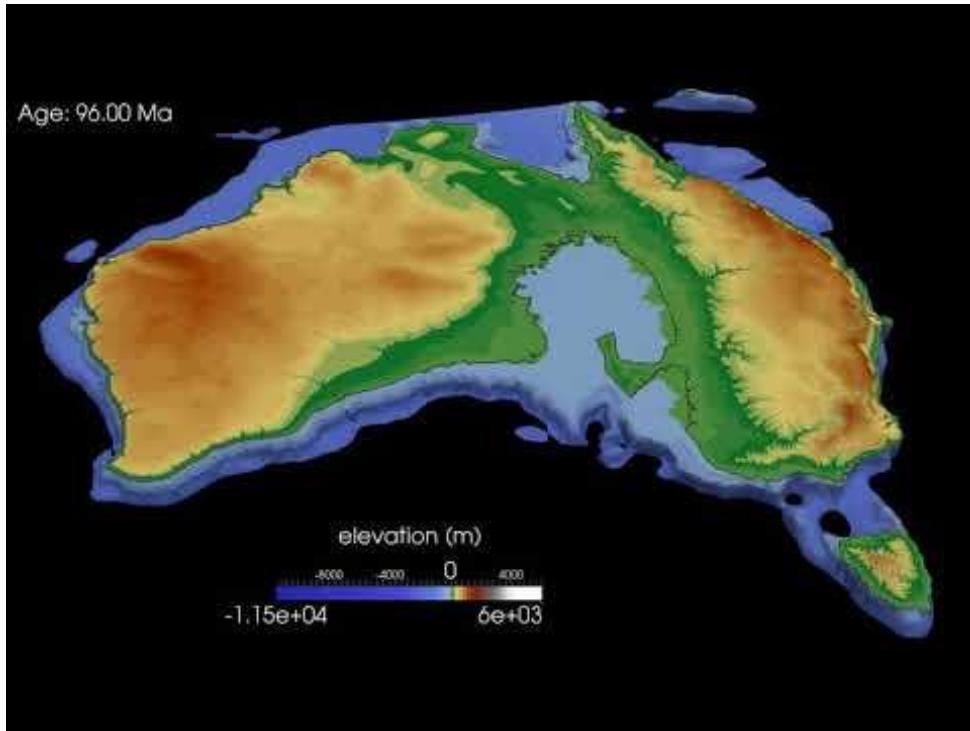
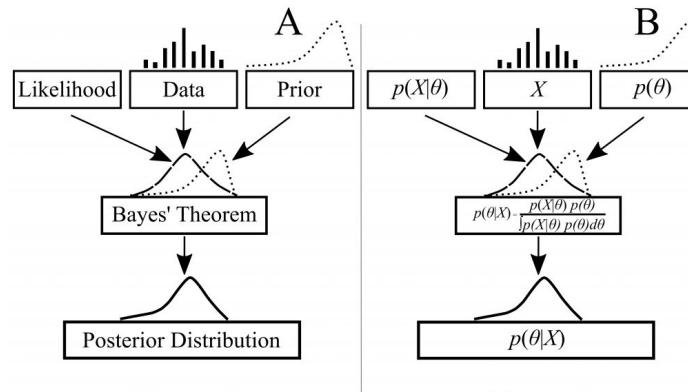
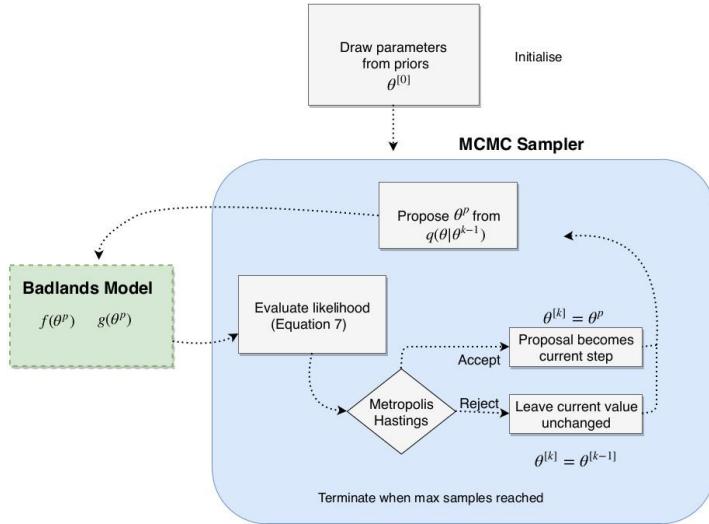
The geological and climate history of the planet is a complex puzzle to solve given sparse, limited and noisy datasets. The complexity increases when we go further back in time in millions of years ago covering supercontinents such as Pangea and plate tectonics that shaped the Earth we know today.

We used machine learning and statistical methods to provide inference by estimating the unknown data, which consist of climate-sensitive sedimentary deposits such as coal and evaporites like salt and gypsum, deposited over the last 250 million years.

The presence of coal implies the presence of particular types of **forests back in time**. The distribution of these deposits is in turn linked to the distribution of rainfall back in time that ranges in millions of years.

Our model estimates the missing deposit data and uses it to estimate precipitation for 11 time-slices over the last 250 million years. Our model has been trained using precipitation data generated from global circulation models for two time periods in the Miocene and Eocene, 15 and 34 million years ago.

Bayesian Geoscientific models - Landscape evolution models



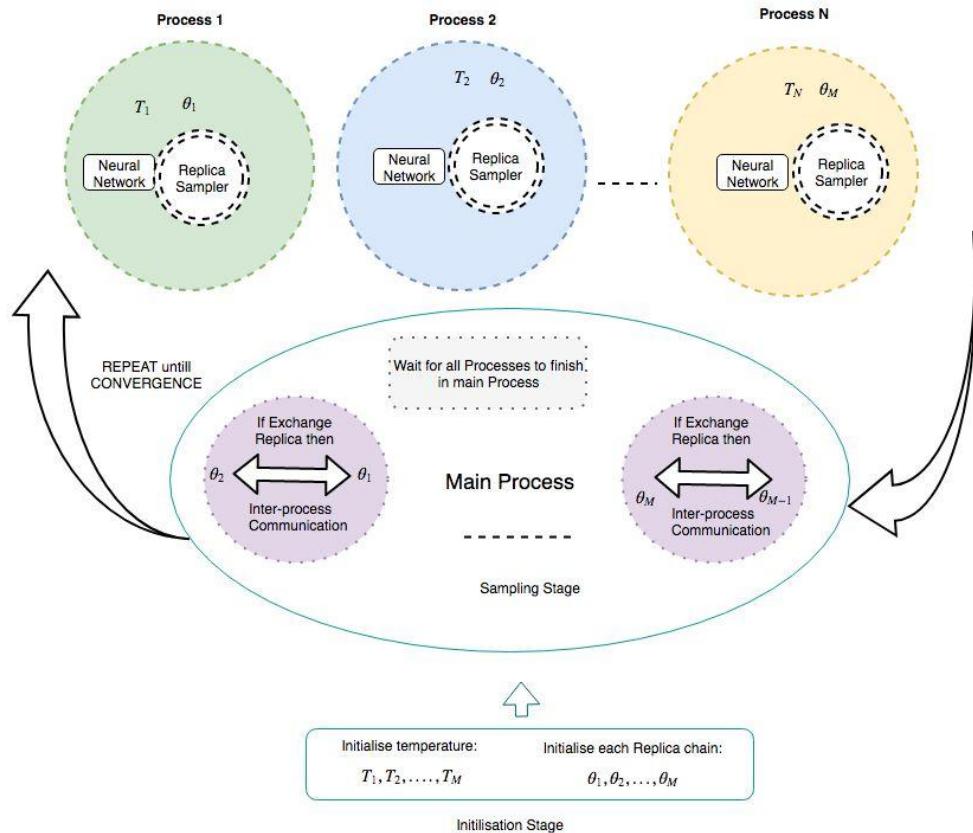
Bayeslands Framework with parallel computing



Prof. Dietmar Muller
University of Sydney



Dr. Tristan Salles
University of Sydney

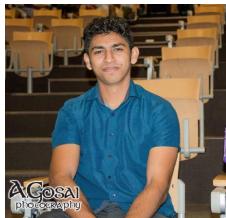




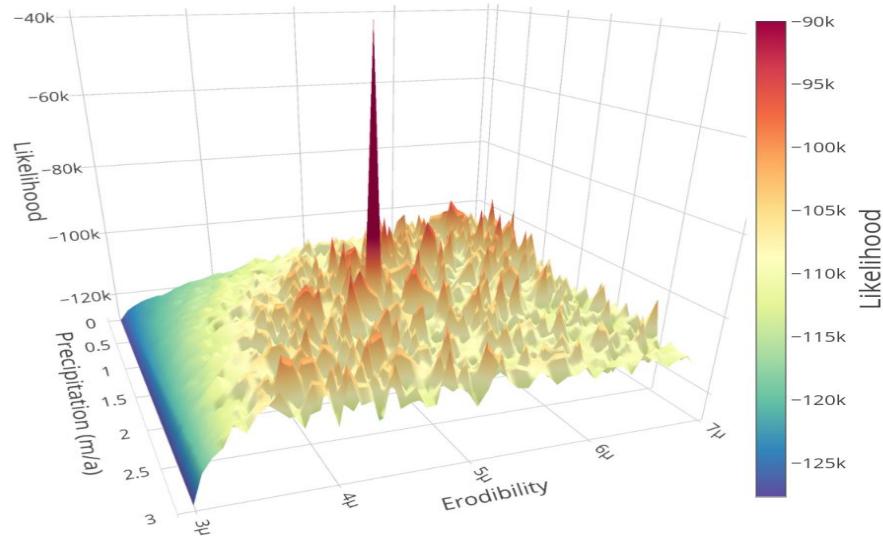
Danial Azam
Research
Engineer
EarthByte,
University of
Sydney



Dr. Nathaniel
Butterworth
Senior
Informatics
Engineer
Sydney
Informatics Hub
University of
Sydney



Ratneel Deo



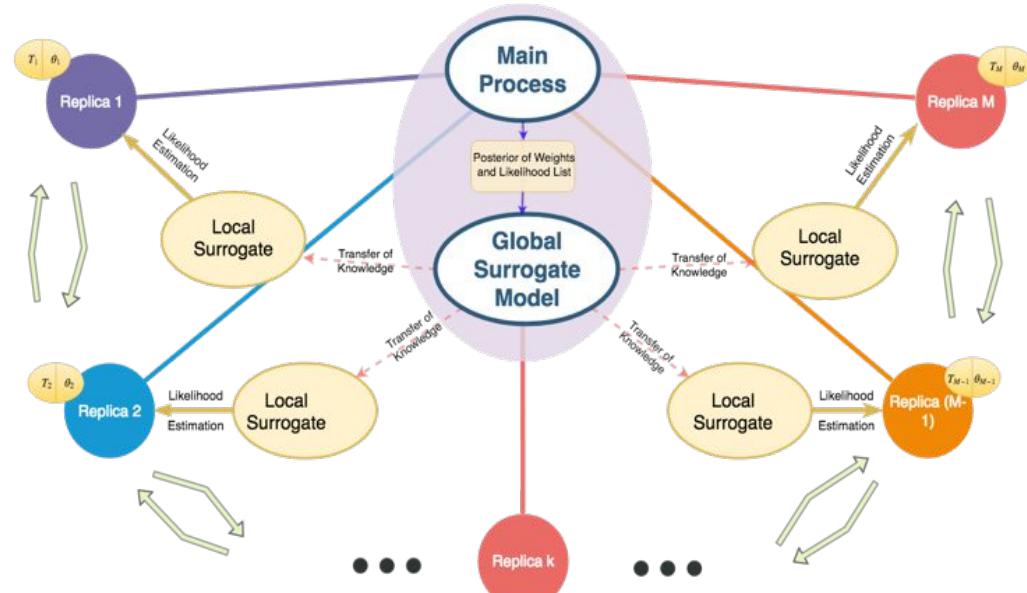
Log-likelihood surface
plot (Precip vs Erod)

Bayesian optimisation: Surrogate-assisted Bayeslands

Computationally expensive models?

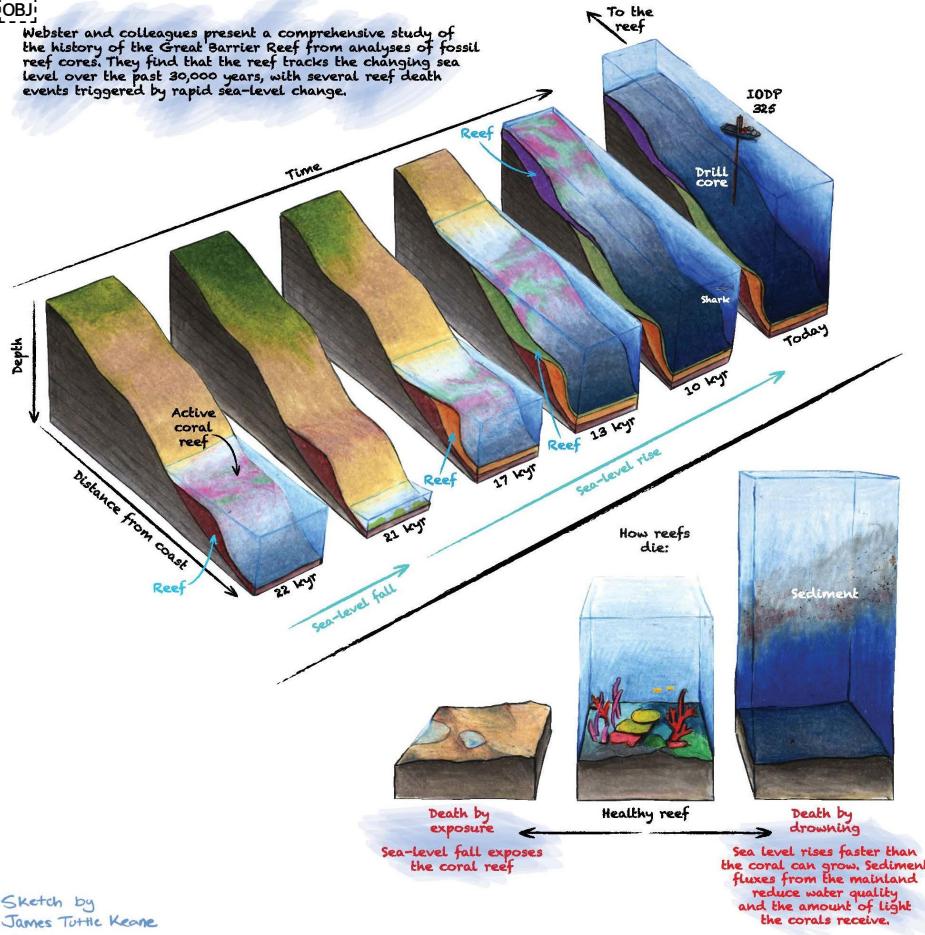
Surrogate assisted optimisation

Bayesian inference - Parallel tempering MCMCs^[OB]

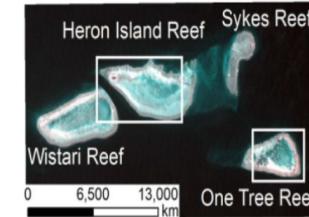
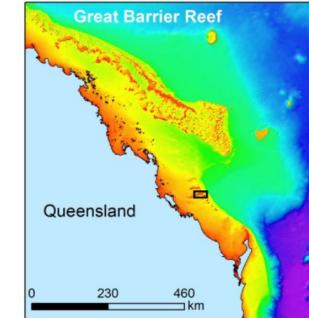
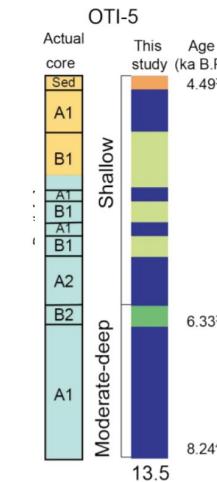


OBJ

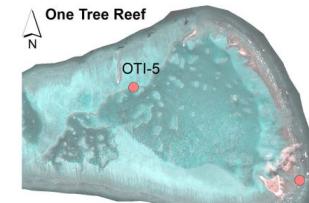
Webster and colleagues present a comprehensive study of the history of the Great Barrier Reef from analyses of fossil reef cores. They find that the reef tracks the changing sea level over the past 30,000 years, with several reef death events triggered by rapid sea-level change.



Geological reef modelling

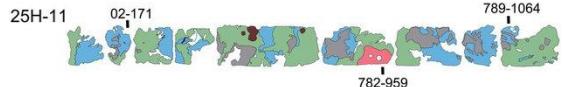


- Legend**
- Carbonate sediment
 - Shallow, exposed assemblage
 - Deep, exposed assemblage
 - Moderate-deep, protected assemblage
 - Deep, protected assemblage
 - Shallow palaeo-water depths (0-4 m)
 - Moderate palaeo-water depths (4-8 m)
 - Deep palaeo-water depths (>8 m)

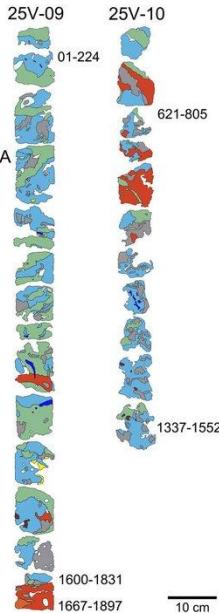


Data Fusion and Bayesian inference in BayesReef

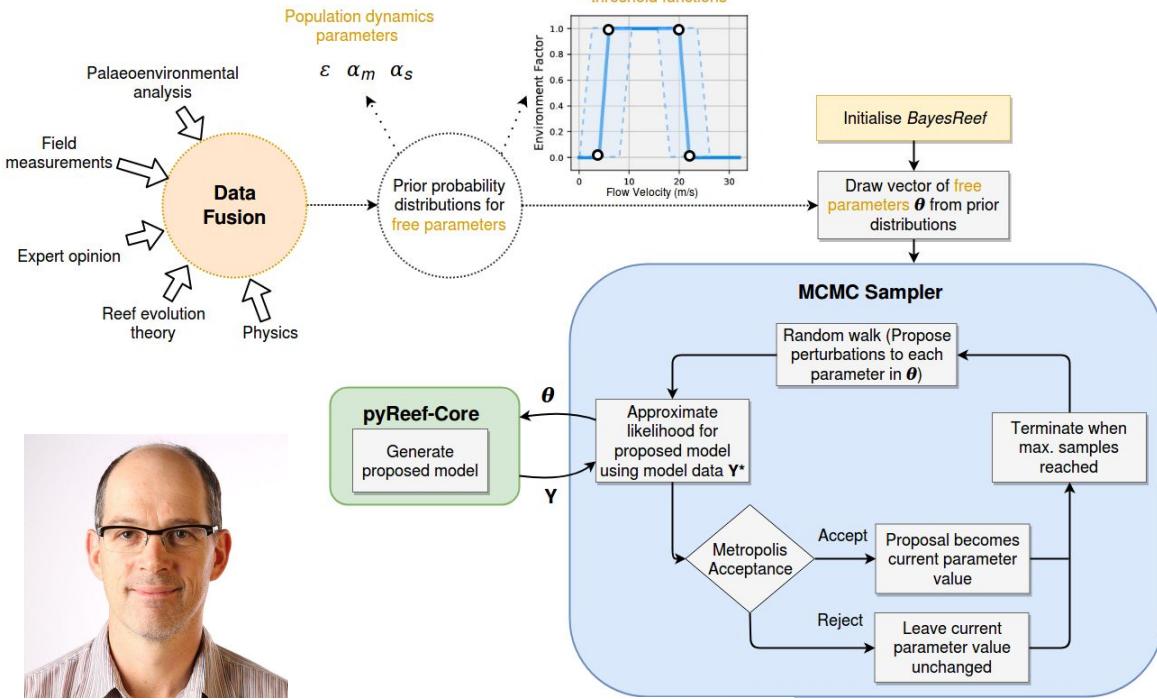
HORIZONTAL CORE



VERTICAL CORES



A
T
A
T



Prof. Jody Webster
University of Sydney

Pall J; Chandra R; Azam D; Salles T; Webster JM; Scalzo R; Cripps S, 2020, 'Bayesreef: A Bayesian inference framework for modelling reef growth in response to environmental change and biological dynamics', *Environmental Modelling and Software*, vol. 125, pp. 104610 - 104610



Jodie Pall, USyd
Medal
(Honours)
University of Sydney

Machine learning for drill-core analysis - mining



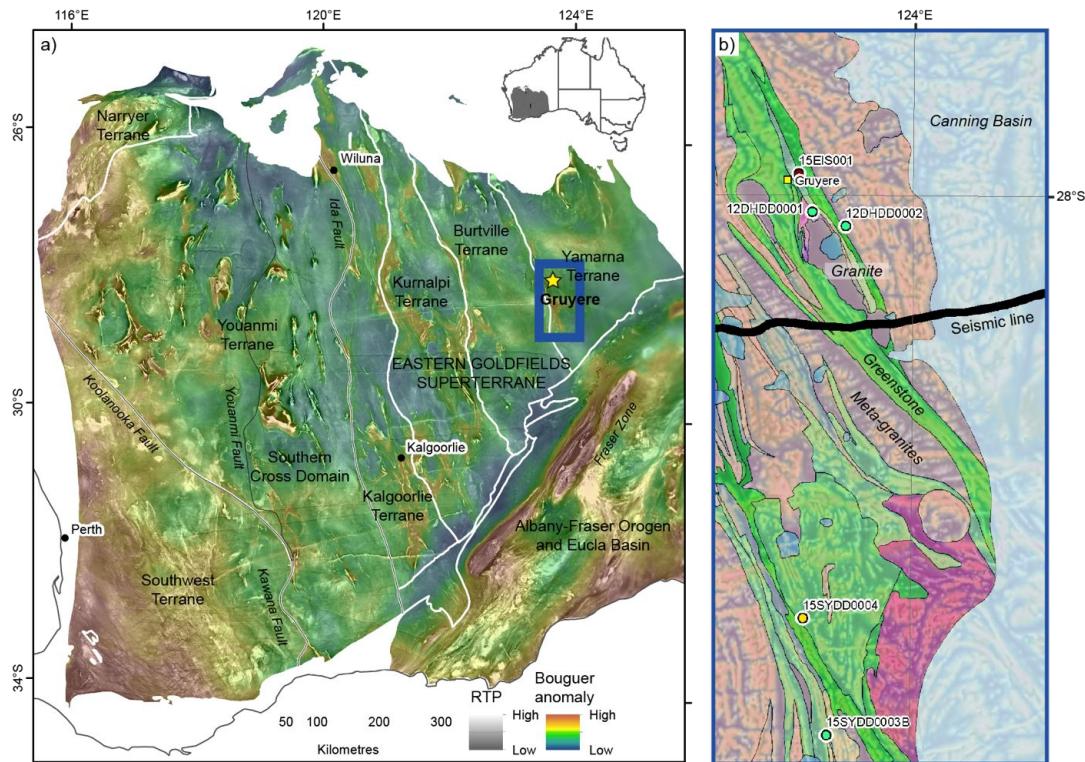
A/Prof. Stuart Clark
UNSW



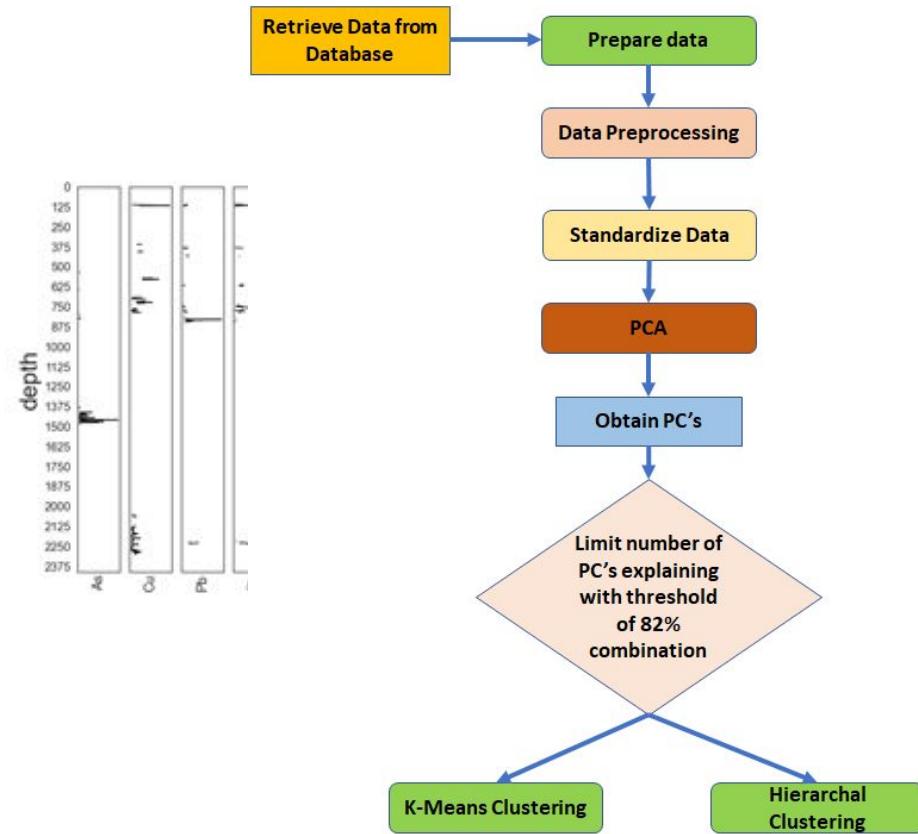
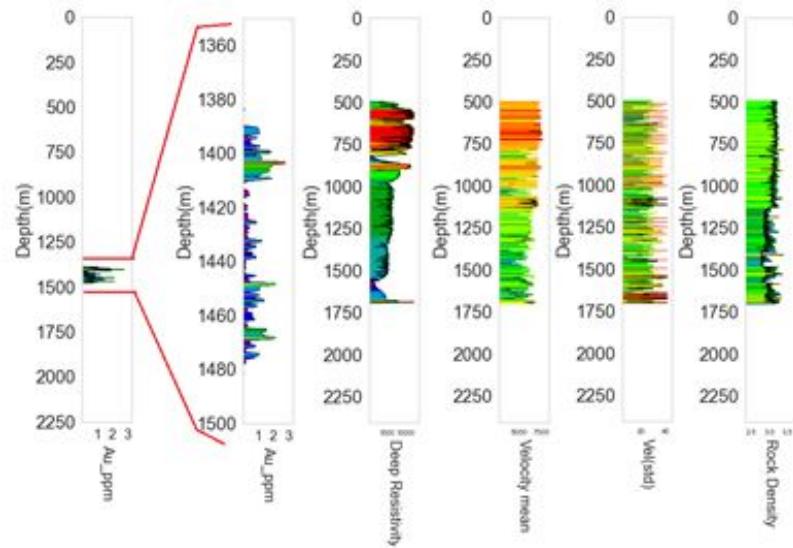
Subhash Chandra
PhD student, UNSW



Dr Mark Lindsey
CSIRO Science Lead
-Mining



S. Chandra, M. Lindsay, S. Clark, and R. Chandra, Drill-core analysis for mineral composition with an unsupervised machine learning framework, Ore Geology Reviews, 2022 (In review)



Convolutional Neural Networks and Remote Sensing Lithological Mapping

H. Shirmard, E. Farahbakhsh, E. Heidari, A. B. Pour, B. Pradhan, R. D. Müller and **R. Chandra**. A comparative study of convolutional neural networks and conventional machine learning models for lithological mapping using remote sensing data. *Remote Sensing* (to appear in January 2022)



Associate Professor
Universiti Malaysia
Terengganu, Malaysia



Prof. Biswajeet
Pradhan, UTS Sydney



Hodjat Srimard, Exploration
Geologist, Iran



Ehsan Farahbakhsh,
Postdoctoral
Research Associate,
EarthByte Group,
University of Sydney



Prof. Dietmar Muller
Former ARC Laureate
Fellow, EarthByte
Group, University of
Sydney

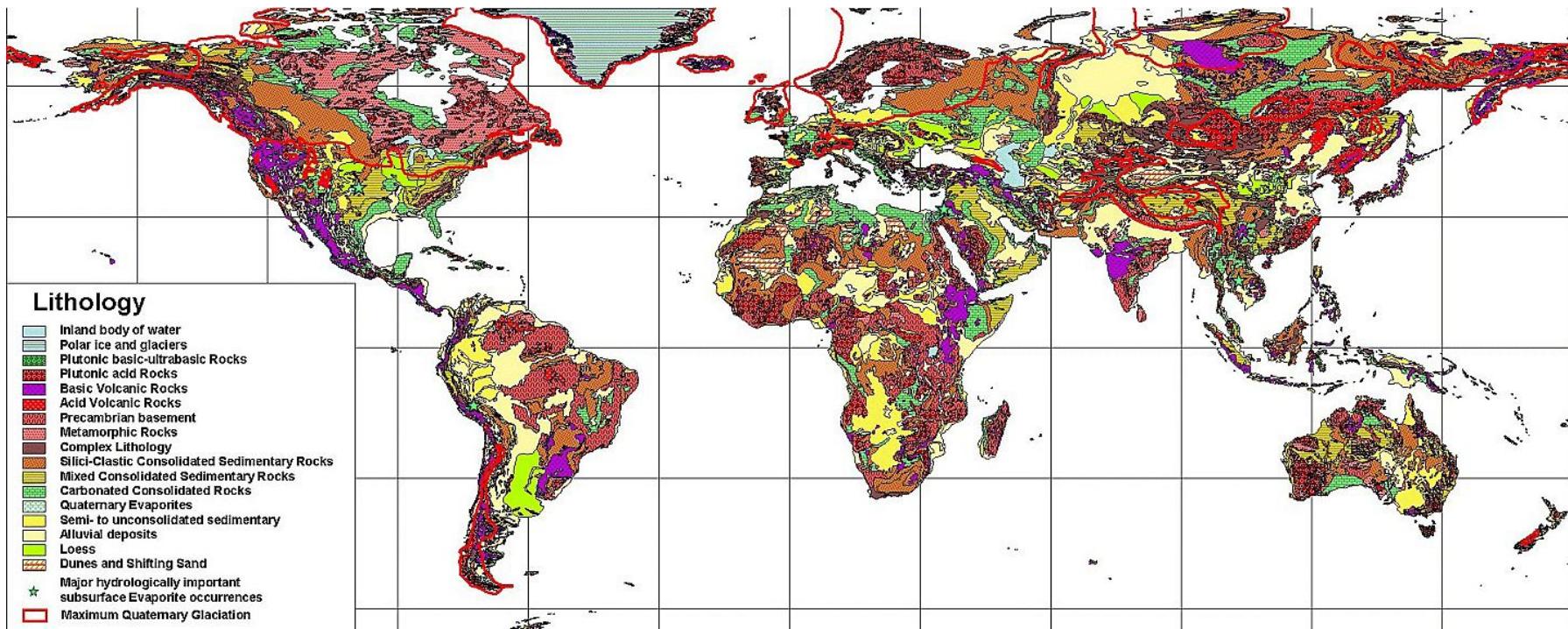
Introduction

Lithological mapping is a critical aspect of geological mapping that can be useful in studying the mineralization potential of a region and has implications for mineral prospectivity mapping. This is a challenging task if done manually, particularly in highly remote areas that necessitates a large number of participants and resources.

The combination of machine learning methods and remote sensing data can provide an easy, low-cost, and accurate approach for mapping lithological units. In this study, we use deep learning via convolutional neural networks (CNNs) and conventional machine learning methods involving support vector machines and multilayer perceptron to map lithological units of a mineral-rich area in the southeast of Iran.

Moreover, we use and compare the efficiency of three different types of multispectral remote sensing data including operational land imager (OLI), advanced spaceborne thermal emission and reflection radiometer (ASTER), and Sentinel-2.

Global lithology map



Dürr, H. H., Meybeck, M., & Dürr, S. H. (2005). Lithologic composition of the Earth's continental surfaces derived from a new digital map emphasizing riverine material transfer. *Global Biogeochemical Cycles*, 19(4).

Figure: Simplified tectonic map of Iran on the left and a true color image of the study area obtained by Sentinel-2 data on the right. The red points shown on the satellite image refer to the samples collected from the study area.

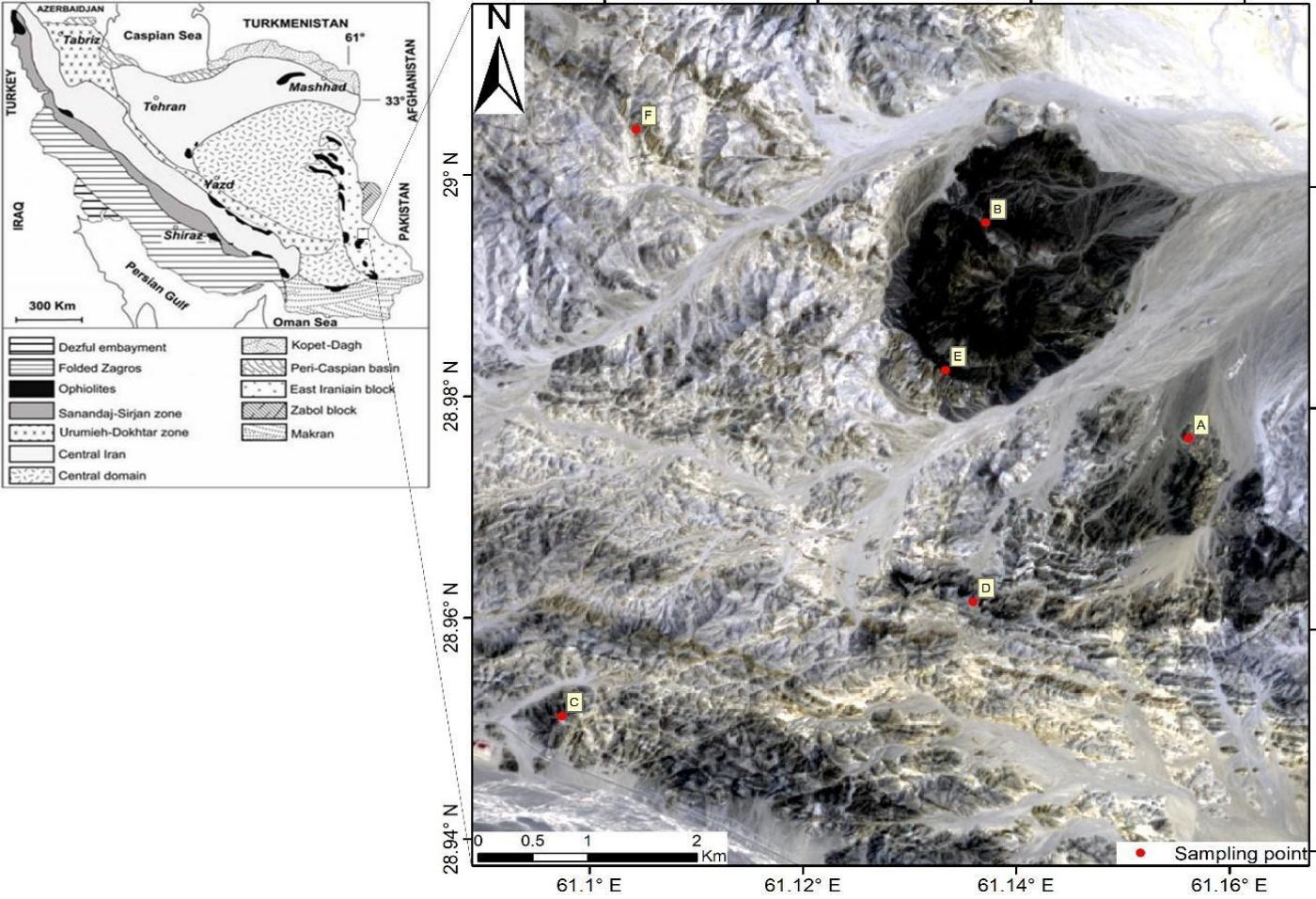




Figure. An overview of exposed quartz monzonite rock units within the study area. The photo has been taken close to the sampling location of sample D.

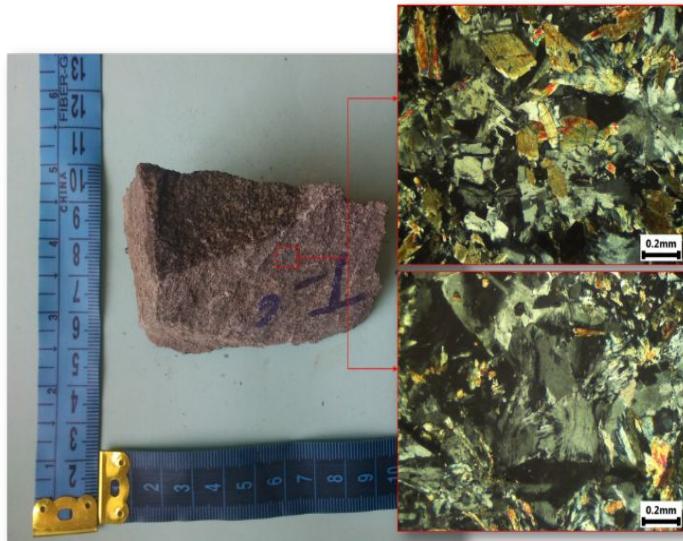
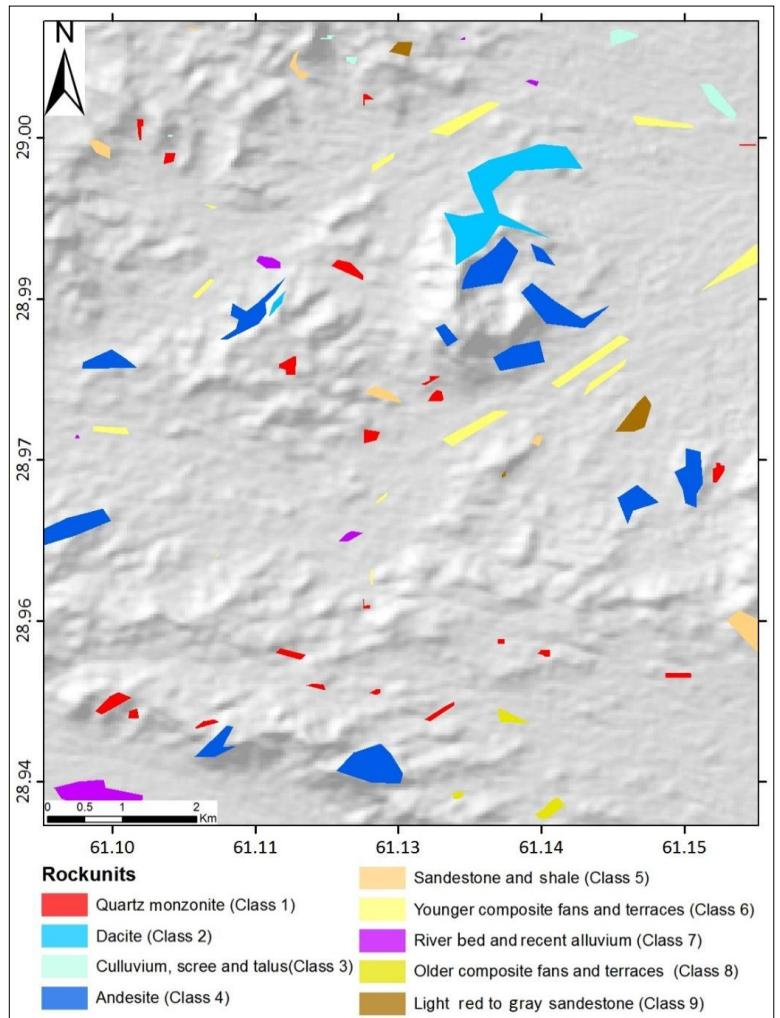
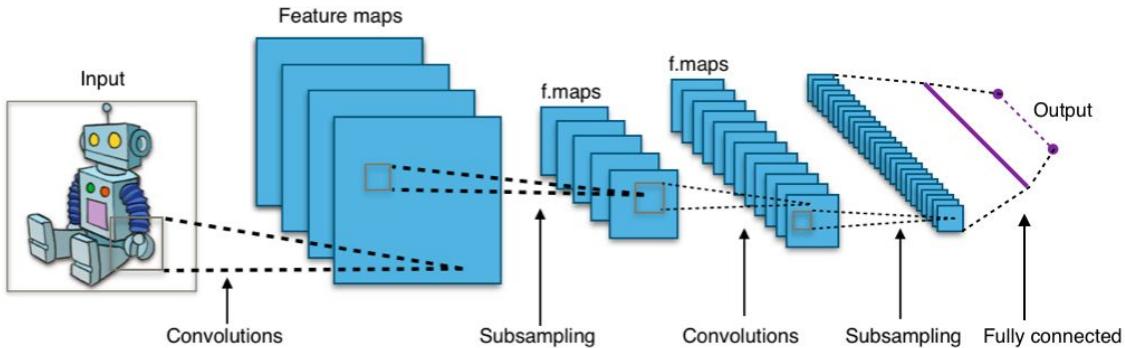
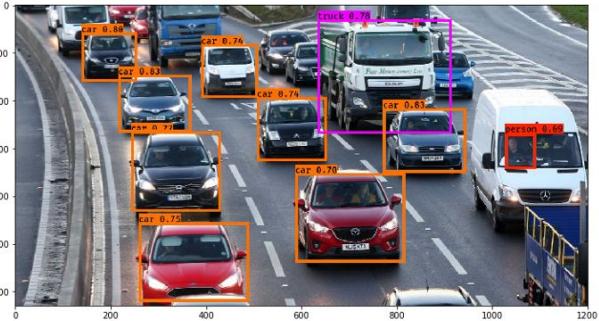


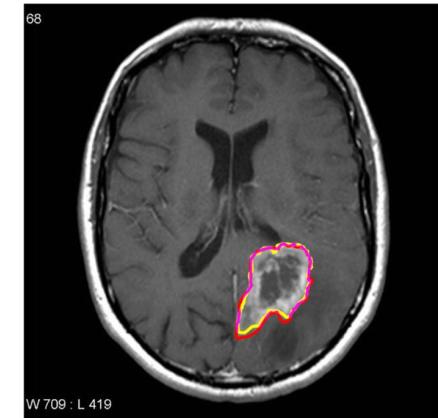
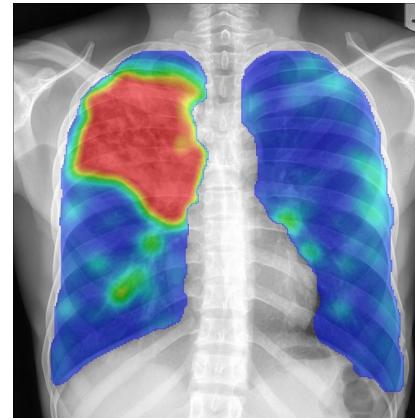
Figure. Sample C taken from the quartz monzonite unit and the microscopic sections. In the upper section, plagioclase, hornblende, and epidote phenocrysts are obvious and in the lower section, quartz phenocrysts accompanied by the radial secondary growth of quartz can be seen.



Deep learning via Convolutional Neural Networks



Source: https://en.wikipedia.org/wiki/Convolutional_neural_network



Multispectral and hyperspectral data

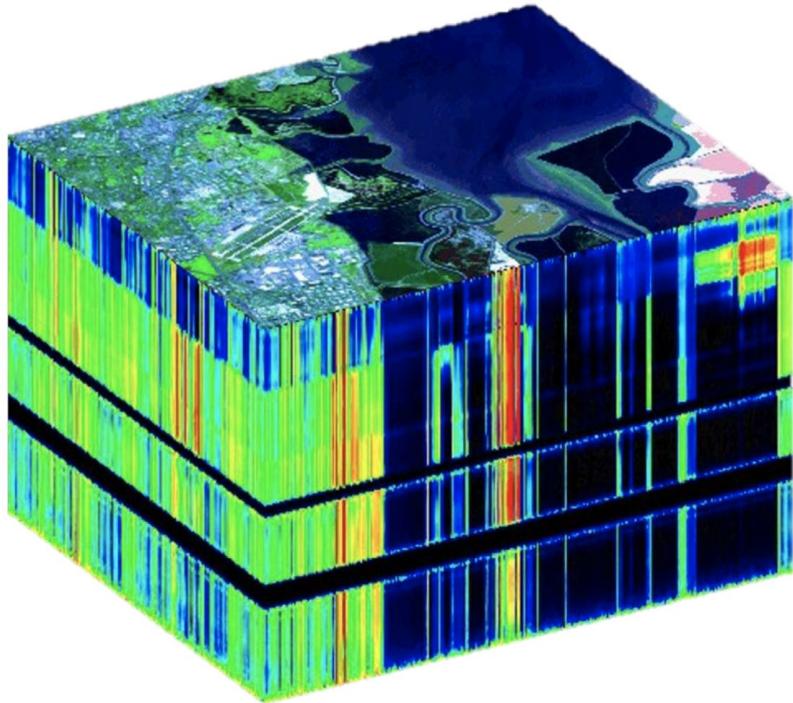
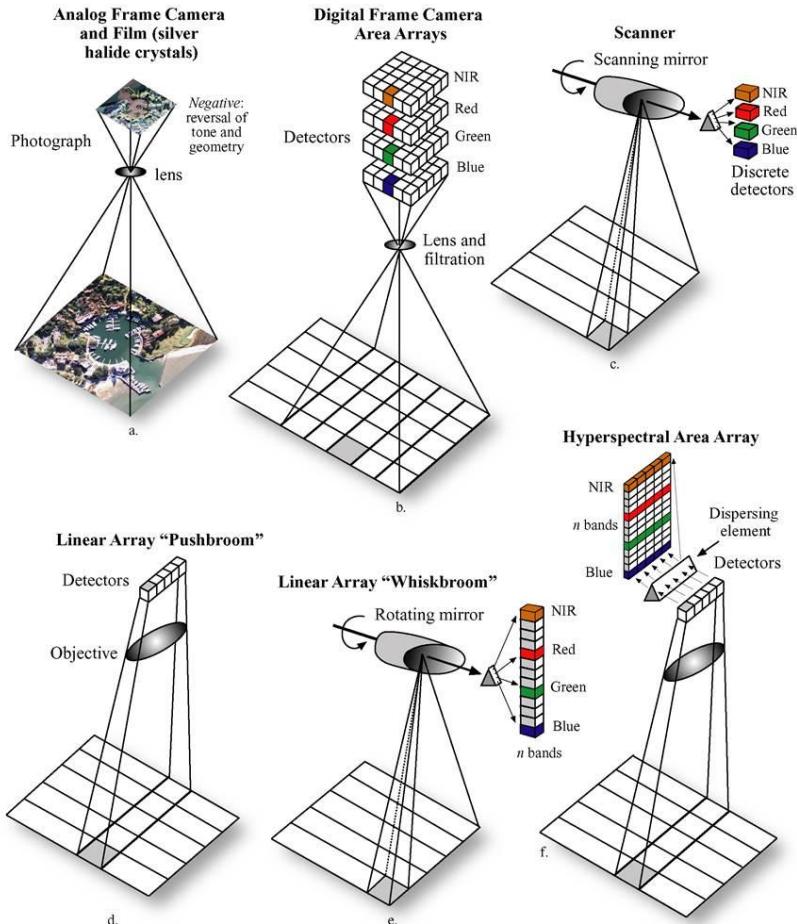


Image Credit: [NASA JPL](#)

Remote Sensing Systems Used to Collect Aerial Photography, Multispectral and Hyperspectral Imagery



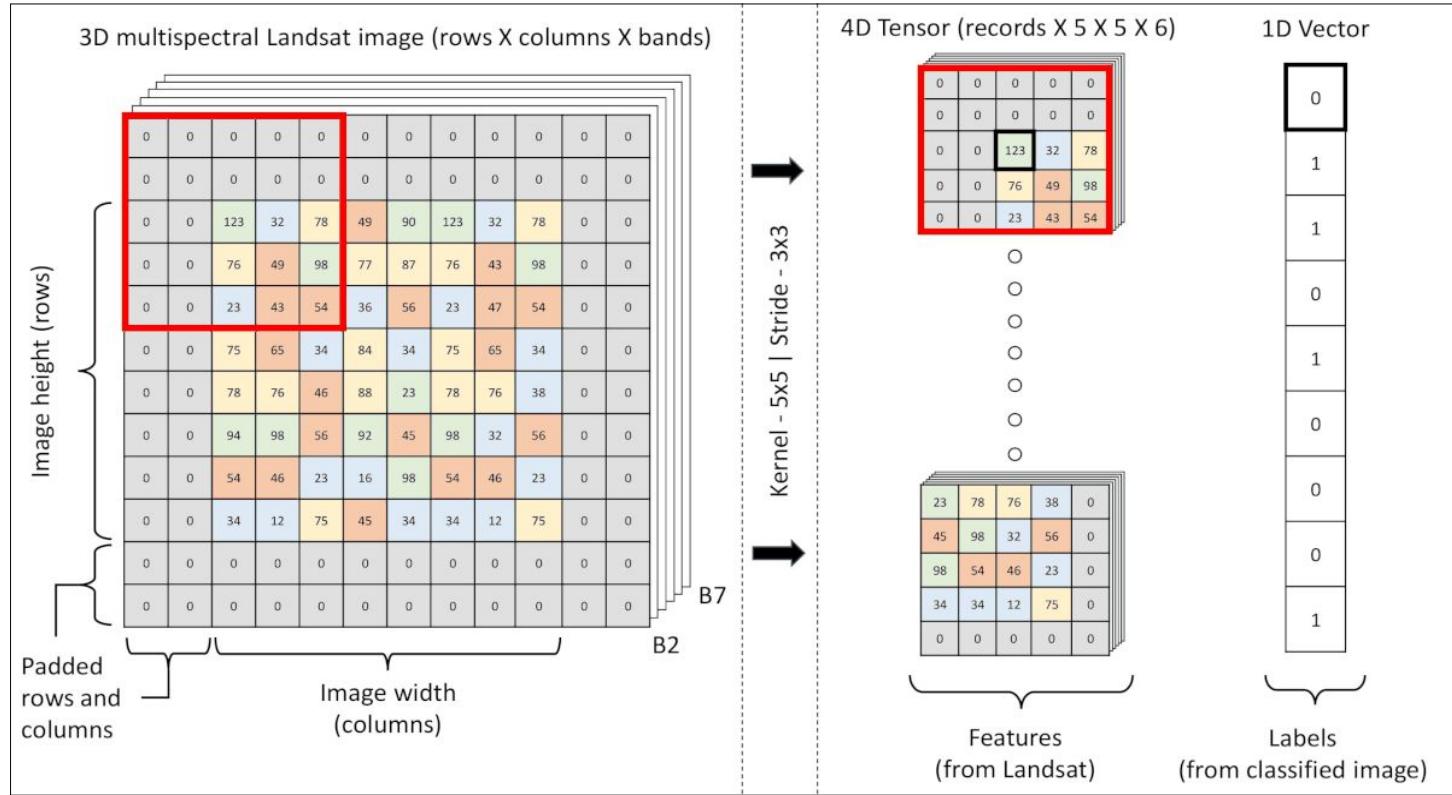


Figure. An illustration of generating training chips for a CNN model using the selected bands of Landsat 8 with a 5x5 kernel and a 3x3 stride.

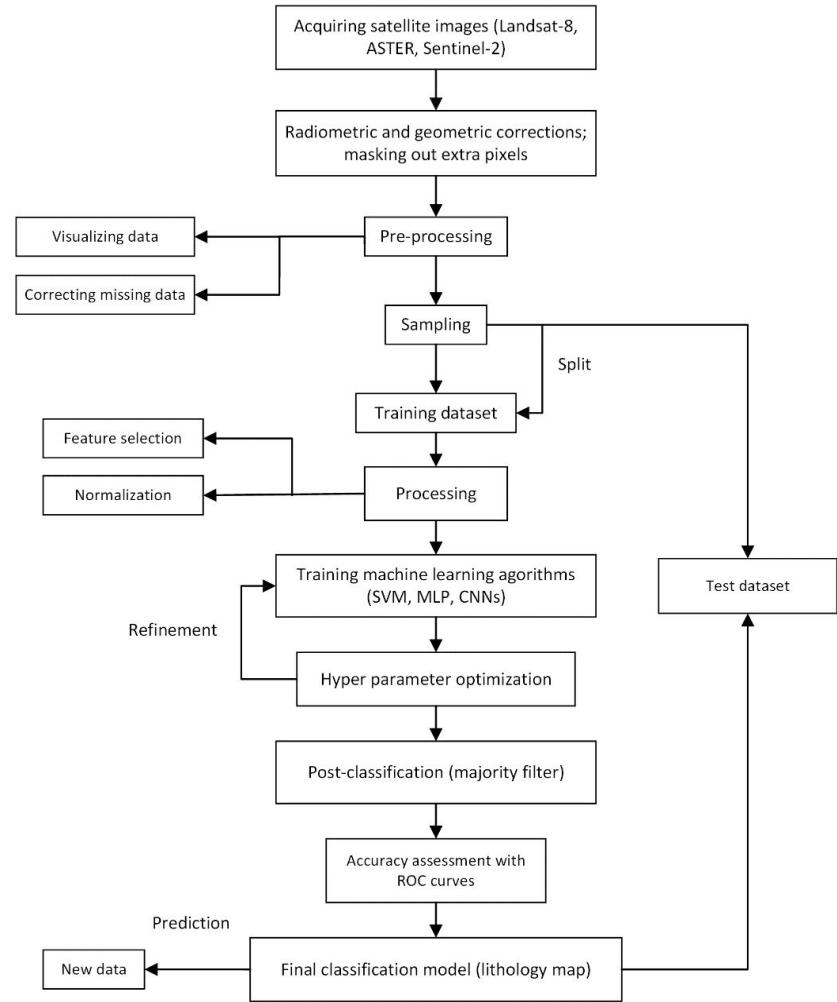


Figure. Proposed framework for applying SVM, MLP, and CNN on remote sensing data and mapping lithological units.

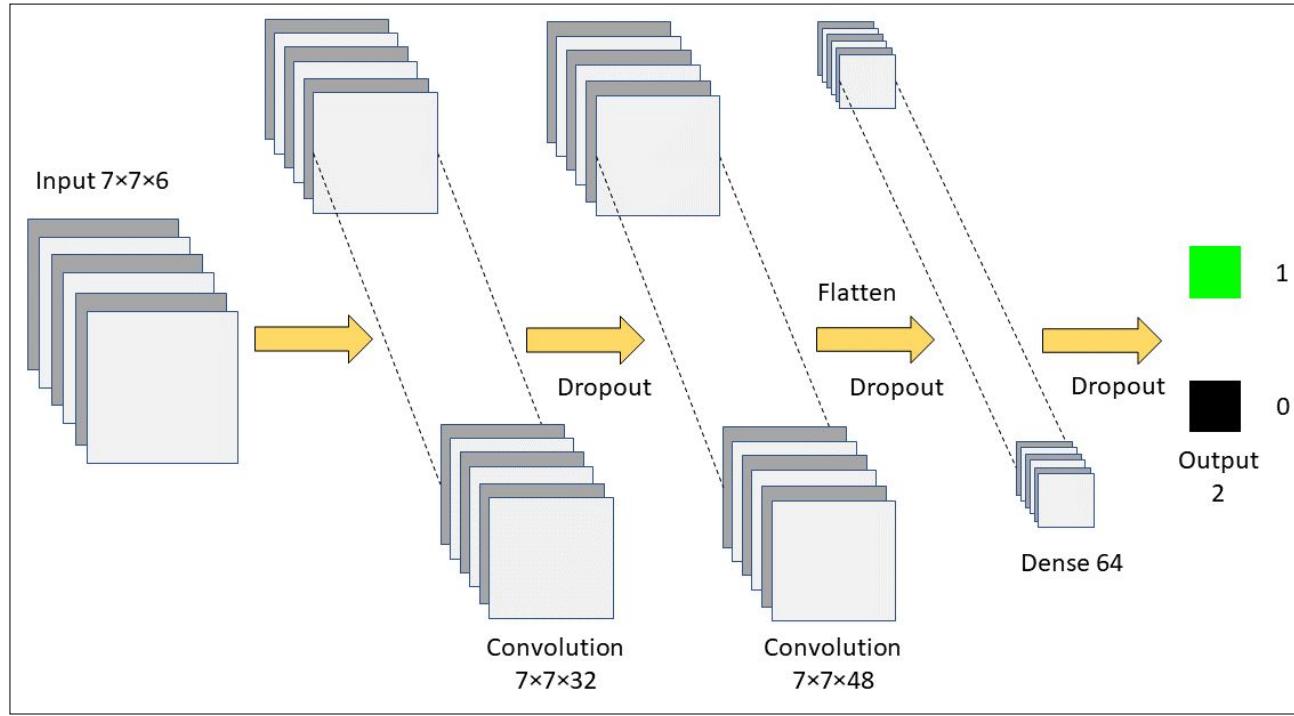
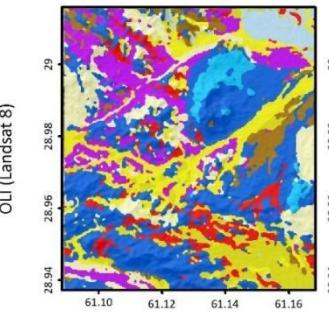
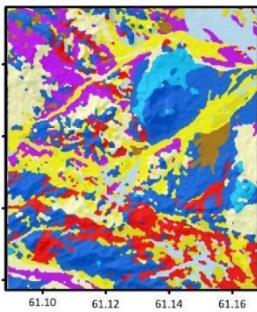


Figure. CNN model architecture adopted for classifying the selected bands of Landsat 8 data.

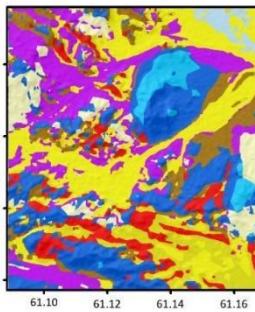
Support Vector Machine



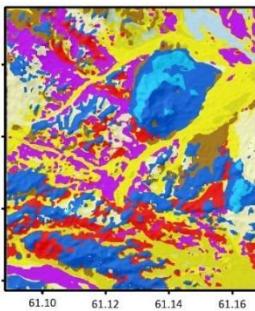
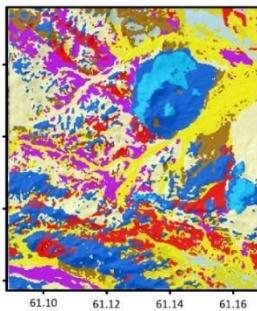
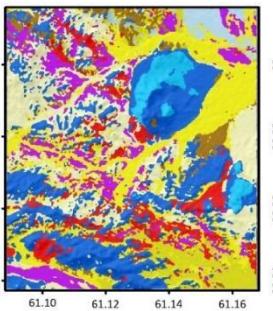
Multilayer Perceptron



Convolutional Neural Networks



ASTER



Sentinel 2A

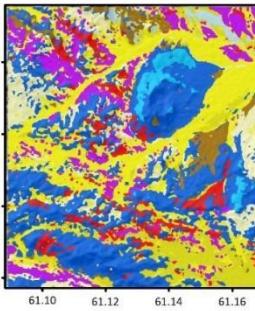
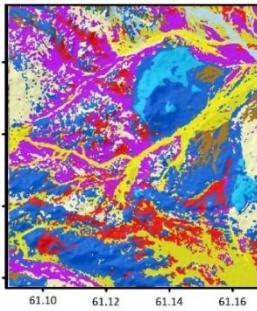
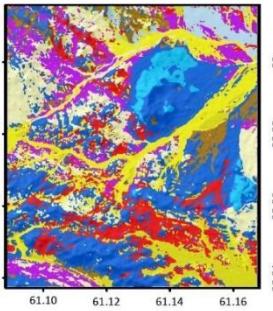


Figure. Classified maps obtained by applying support vector machine, multilayer perceptron, and convolutional neural networks on OLI (Landsat-8), ASTER, and Sentinel-2 satellite images. The legend is the same as the map provided in Figure 2.

Lithology type	Class number	OLI (Landsat 8)			ASTER			Sentinel-2		
		SVM	MLP	CNN	SVM	MLP	CNN	SVM	MLP	CNN
Quartz Monzonite	1	0.98	0.98	1	0.97	0.99	1	0.88	0.98	0.99
Dacite	2	0.99	0.99	0.99	0.88	0.99	1	0.89	0.99	0.99
Colluvium Scree and Talus	3	0.97	0.97	0.99	1.00	0.99	0.99	0.99	0.99	0.99
Andesite	4	0.91	0.91	0.99	0.90	0.95	0.99	0.80	0.91	0.97
Sandstone and Shale	5	0.97	0.97	0.99	0.89	0.97	1	0.94	0.96	0.98
Younger Composite Fans and Terraces	6	0.92	0.92	0.99	0.94	0.94	0.99	0.95	0.94	0.99
River Bed and Recent Alluvium	7	0.95	0.95	0.99	0.96	0.99	1	0.92	0.92	0.99
Older Composite Fans and Terraces	8	0.90	0.90	1	0.95	0.90	0.99	0.95	0.91	0.99
Light Red to Gray Sandstone	9	0.95	0.95	0.99	0.97	0.98	1	0.96	0.94	0.99

Table.. The accuracy of each class (lithological unit) obtained by applying SVM, MLP, and CNN on OLI (Landsat-8), ASTER, and Sentinel-2 satellite images.

Discussion

We showed the efficiency of applying machine and deep learning techniques on satellite data.

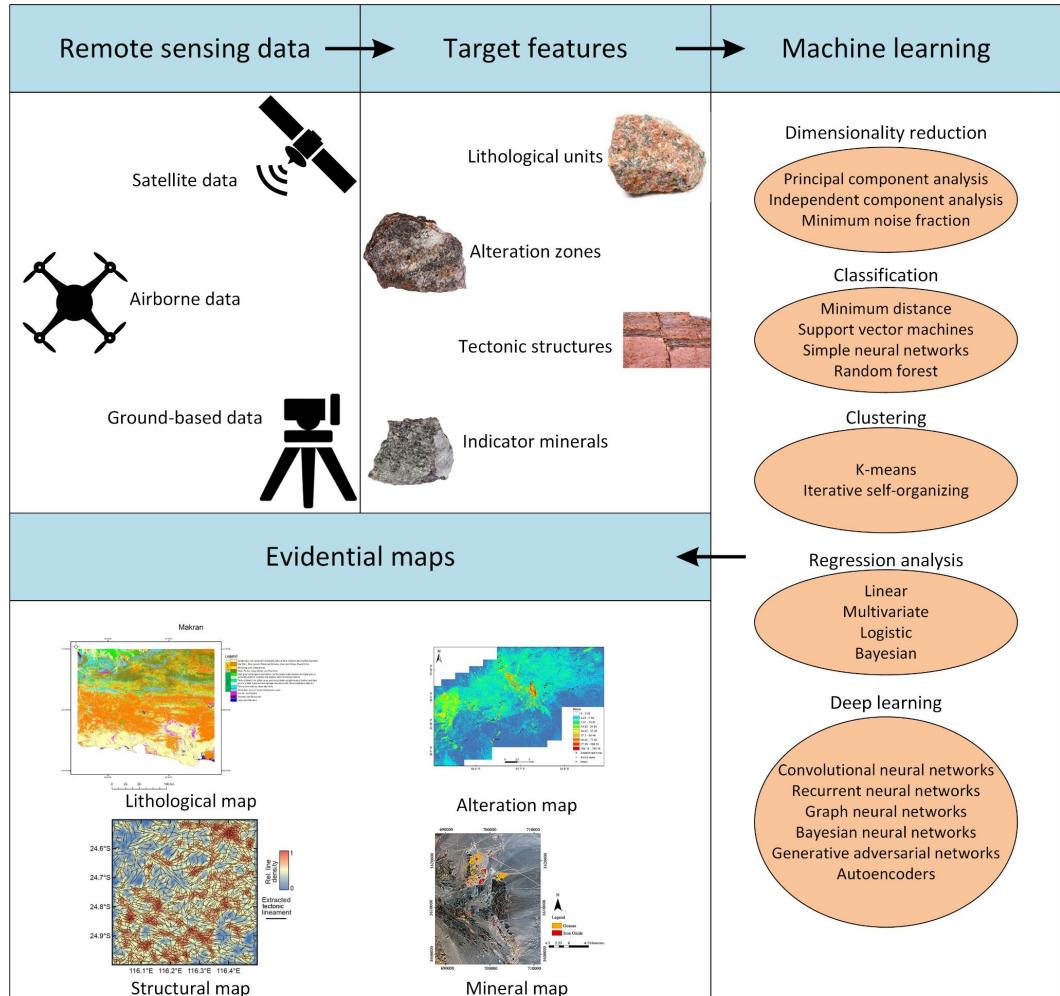
We observed that the combination of CNN and ASTER data provides the most accurate lithological map of the study area based on the ROC curves. Such maps can be considered as base maps for further geological fieldworks and a reliable factor aiding in making decision for mineral exploration operations.

The results show that CNNs and conventional machine learning methods are effective using the respective satellite data sources in generating an accurate lithological map of the study area. However, the combination of CNNs and ASTER satellite data provides the best performance and the highest accuracy and adaptability with field observations and laboratory analysis results.

Our framework presented in a Jupyter notebook is an open-source community tool for mapping lithological units using any multi- or hyper-spectral data. This notebook can significantly enhance the ability of exploration geologists to map lithological units.

Review of progress

Shirmard, H., Farahbakhsh, E., Müller, R. D., & Chandra, R. (2022). A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sensing of Environment*, 268, 112750.



The road ahead

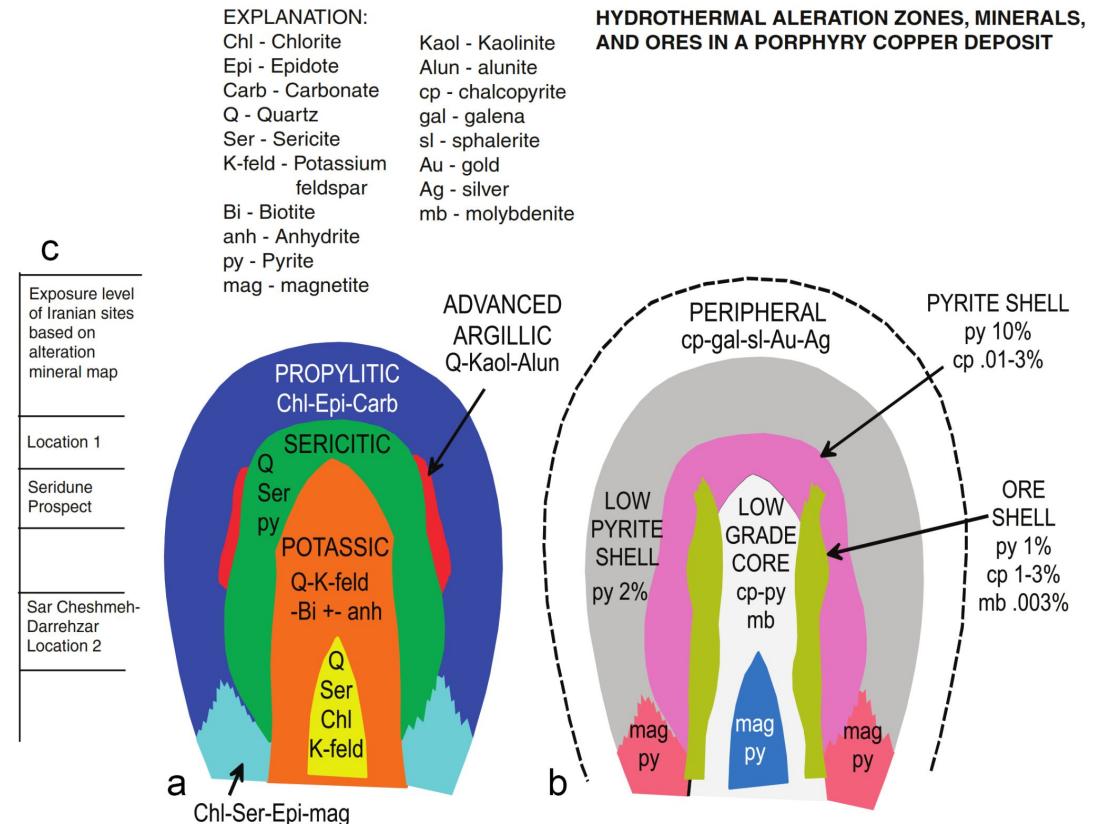


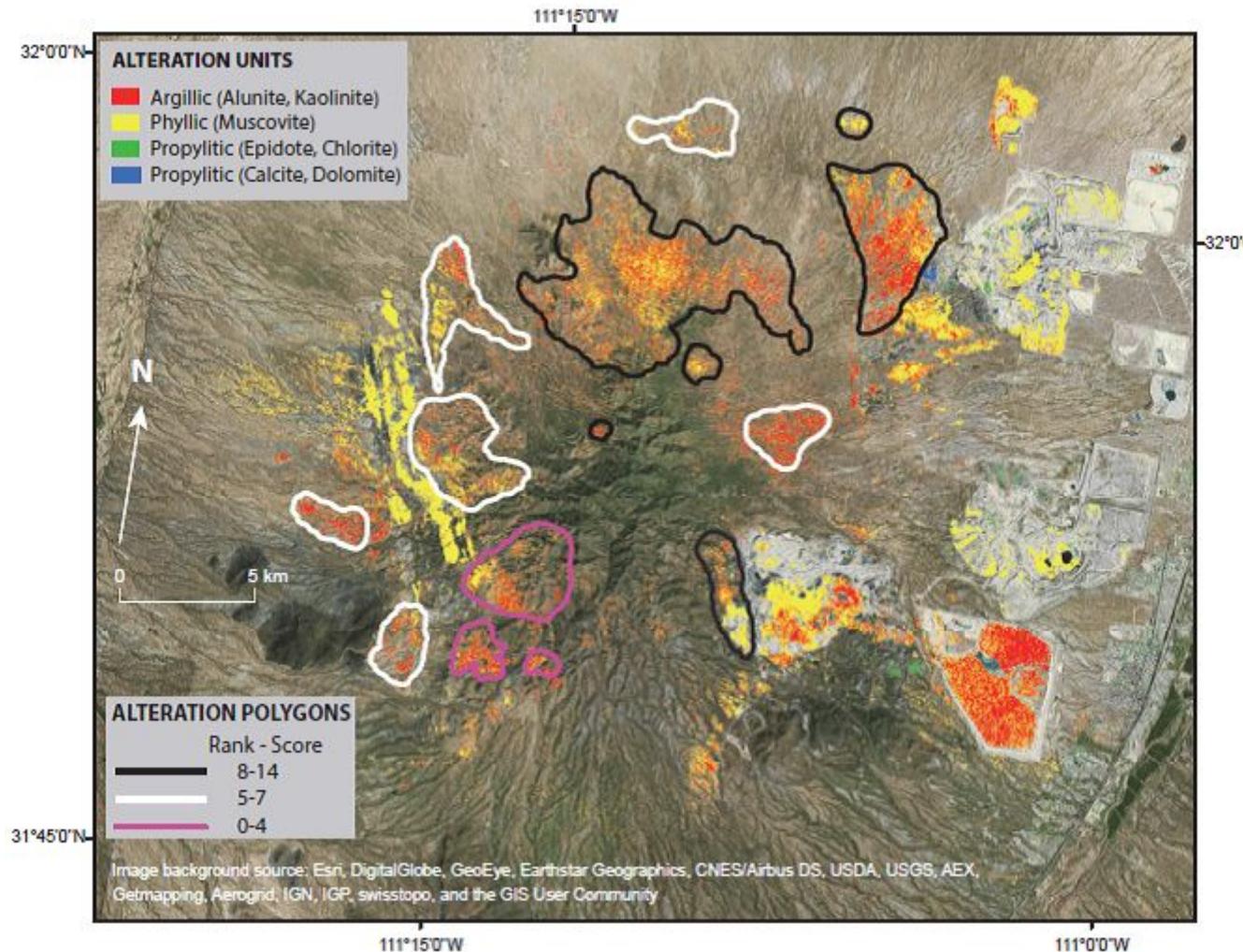
Convolutional Neural Networks and Remote Sensing for Alteration Zones

Dhiraj Pimparkar (Indian Institute of Technology - Jammu)

Ehsan Farahbakhsh (University of Sydney)

Anurag Sharma (University of the South Pacific)





Uncertainty quantification with Bayesian deep learning

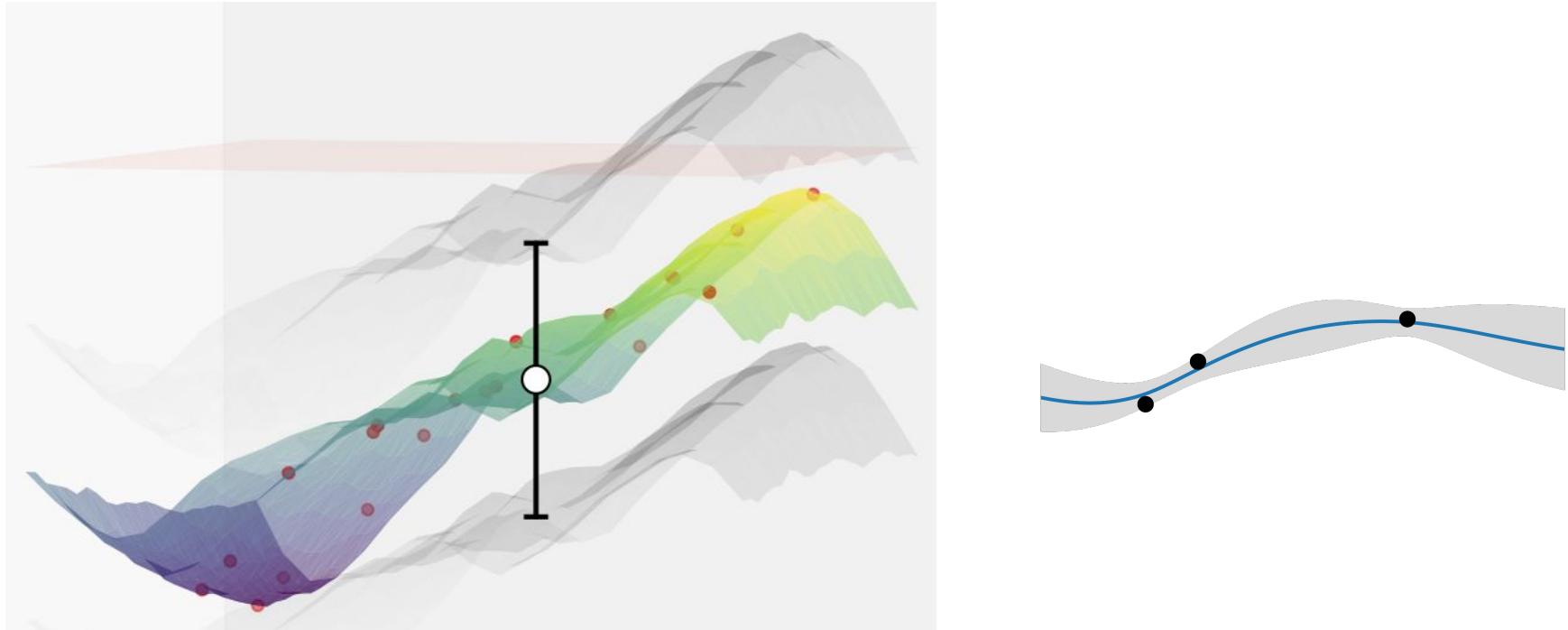
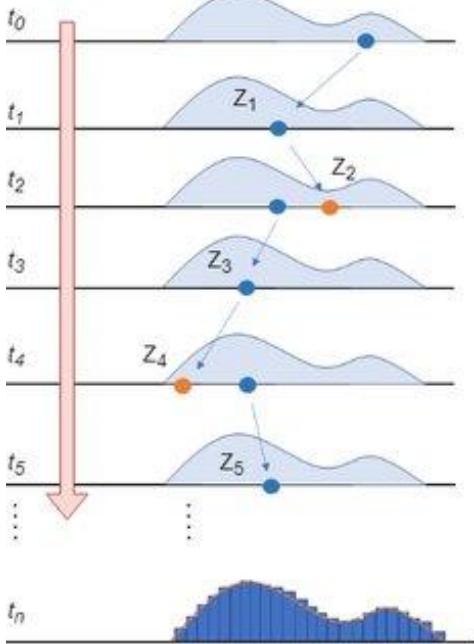


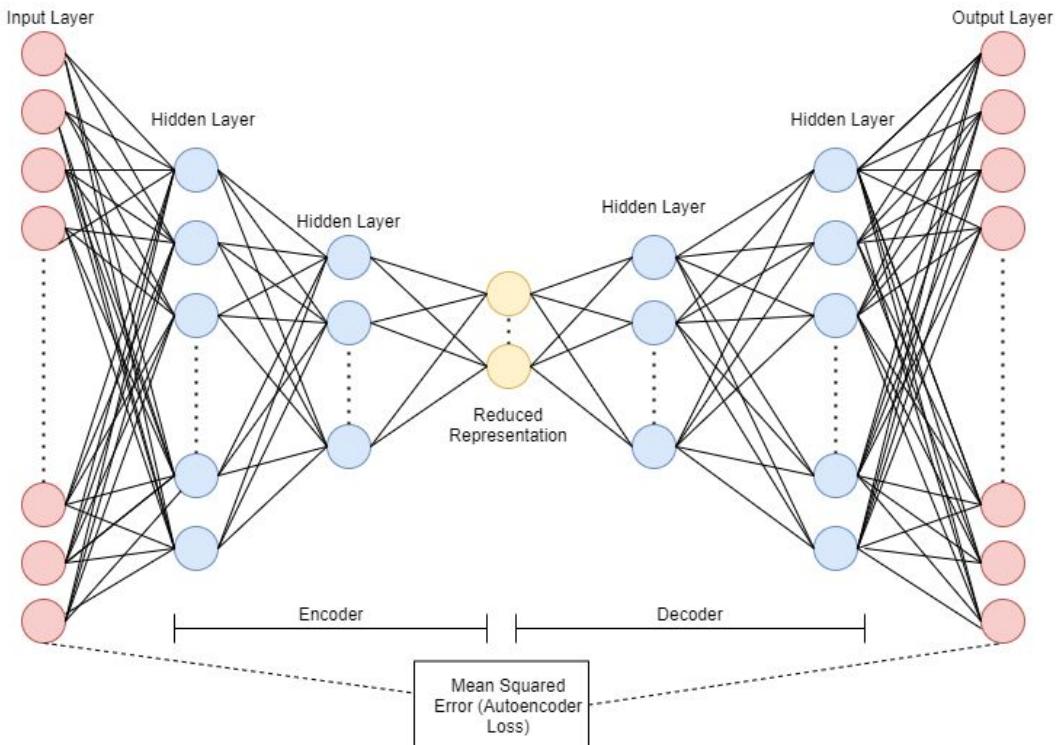
Figure Source: <https://cerfacs.fr/description-s poc-uq/>

Bayesian Autoencoders via MCMC



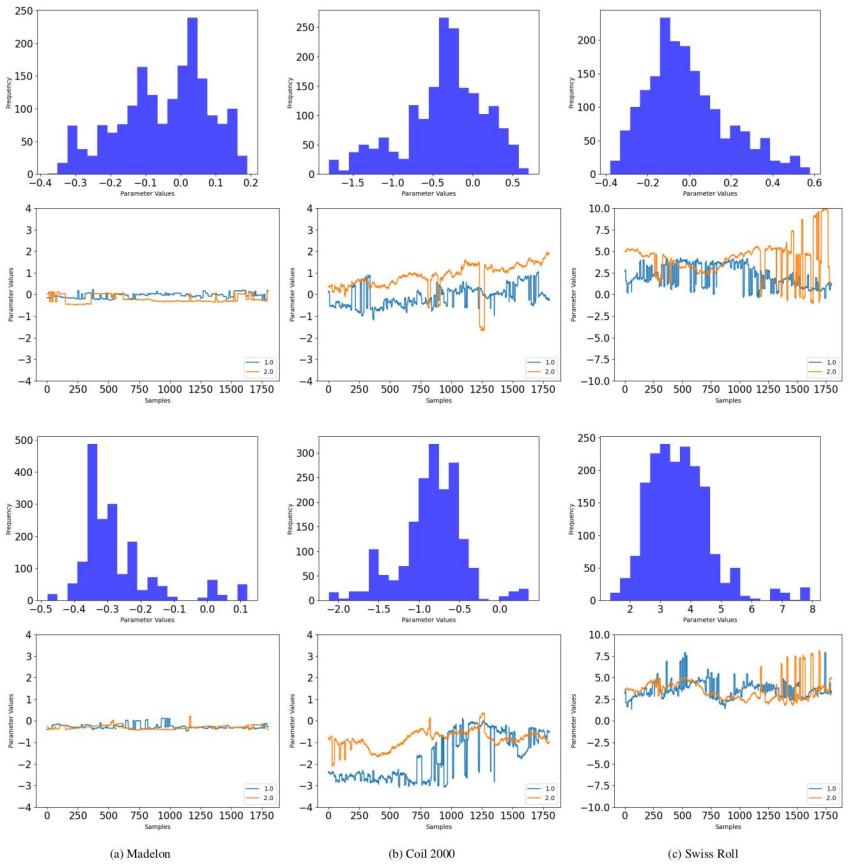
MCMC Sampler

θ_0
Accept $\theta_1 = z$
Reject $\theta_2 = \theta$
Accept $\theta_3 = z$
Reject $\theta_4 = \theta$
Accept $\theta_5 = z$

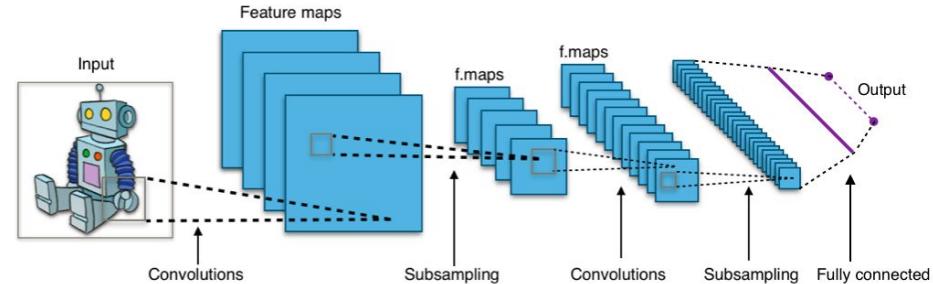
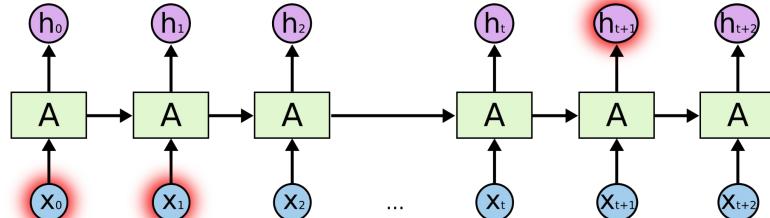
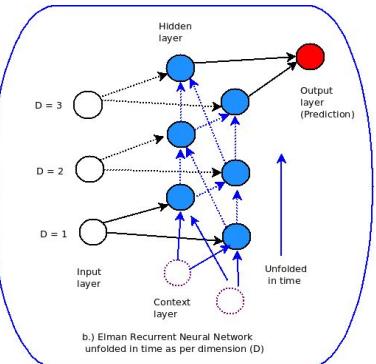
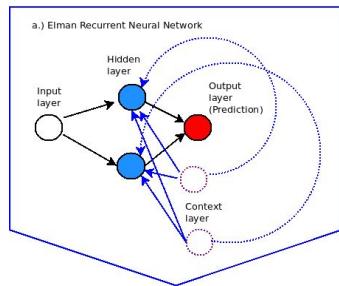


Deep Autoencoder

Chandra, R., Jain, M., Maharana, M., & Krivitsky, P. N. (2021). Revisiting Bayesian Autoencoders with MCMC. *arXiv preprint arXiv:2104.05915*. <https://arxiv.org/abs/2104.05915>



Bayesian Convolutional Neural Networks and Recurrent Neural Networks via MCMC

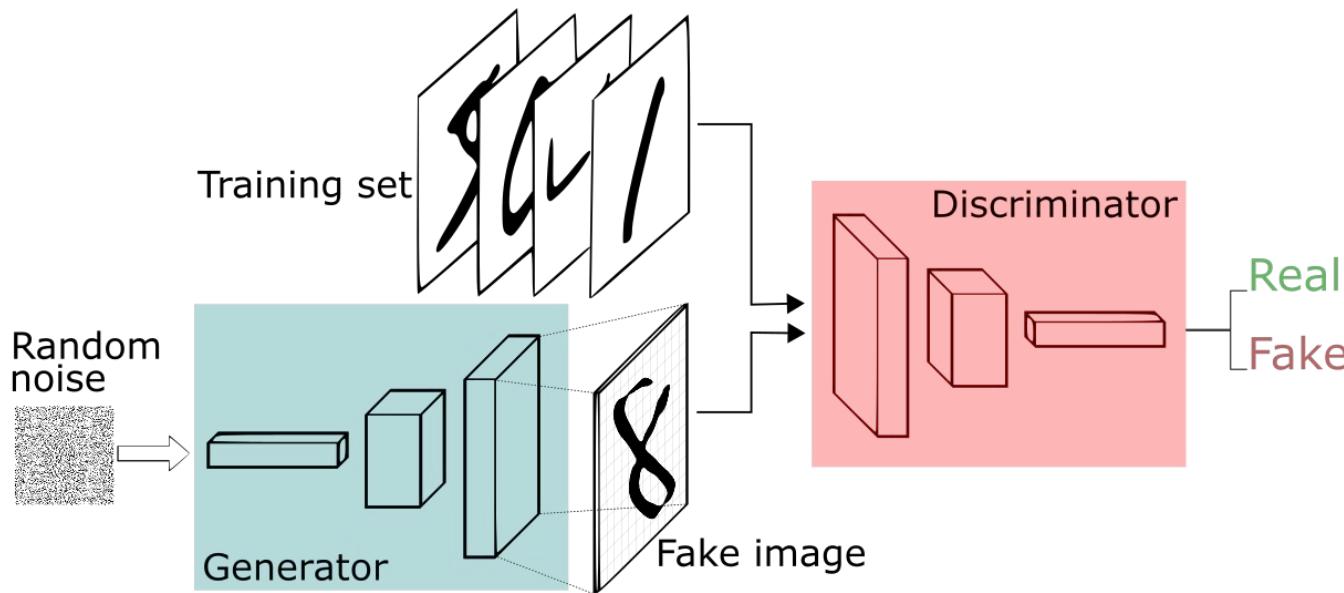


Pulkit Mahajan
IIT-Jammu
India

Shravan Raviraj
Manipal Institute
of Technology
India

Prof. Scott Sisson
UNSW Data Science Hub

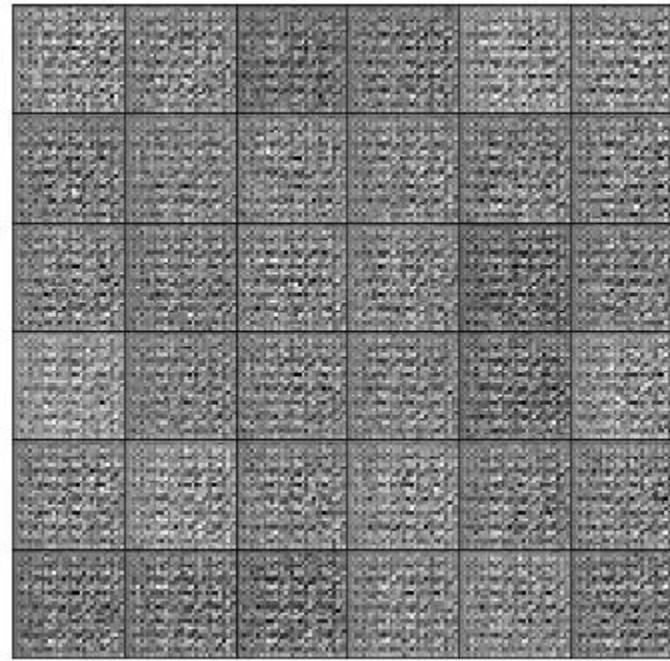
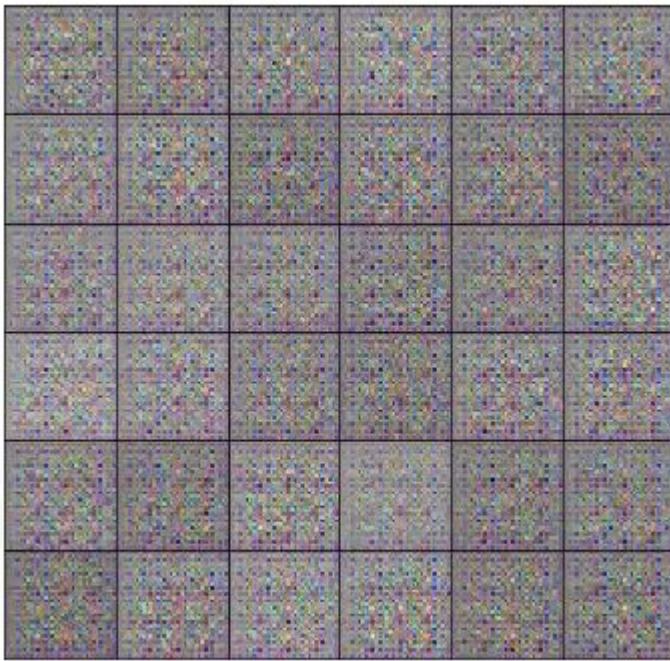
Data Augmentation: Generative Adversarial Networks (GANs)



In the GAN framework:

- The generator maximizes the probability of making the discriminator mistakes its inputs as real.
- The discriminator guiding the generator to produce more realistic images.

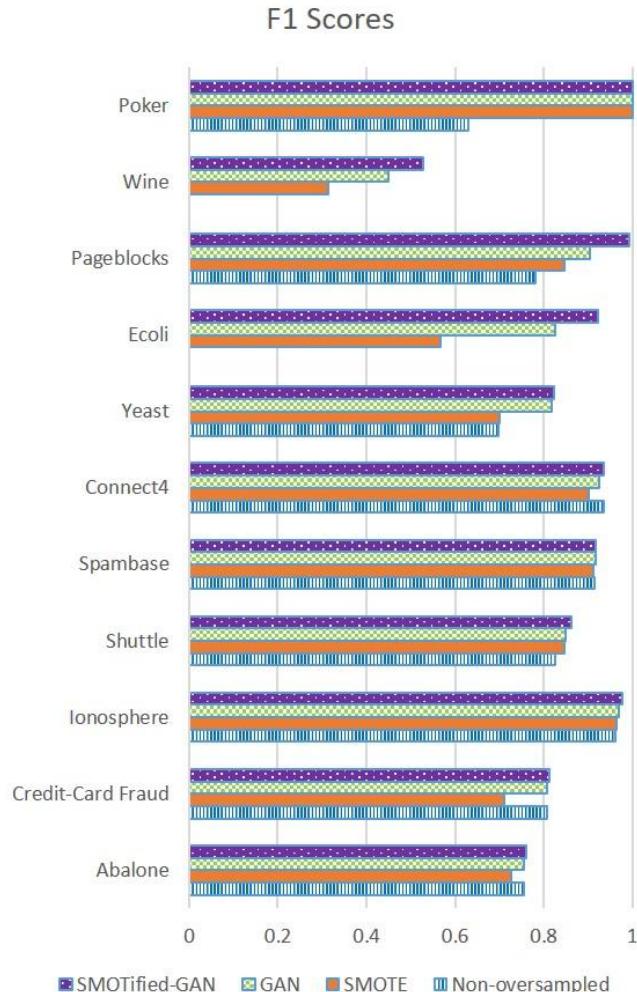
In the perfect equilibrium, the generator would capture the general training data distribution. As a result, the discriminator is always unsure of whether its inputs are real or not.



“We used a 4 layer convolution network for (both discriminator and generator) with batch normalization to teach a model to generate SVHNs and MNIST images. Above, are the SVHN’s (left) and MNIST (right) generator samples during training.” Source:
<https://sthalles.github.io/intro-to-gans/>

Pattern classification (non-image datasets)

Can be extended
to non-Image
data?



Anuraganand Sharma, Prabhat Kumar Singh, Rohitash Chandra,
“SMOTified-GAN for class imbalanced pattern classification problems”.
<https://arxiv.org/submit/3871315/view>

Related projects (2022)

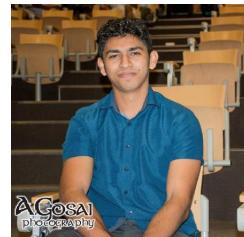
Coupling deep learning with hydrological models (incoming PhD student Arpit Kapoor) with Prof. Lucy Marshall



Arpit Kapoor
UNSW



Prof. Lucy
Marshall
UNSW



Ratneel Deo
USyd

Deep learning for analysis of geological reef evolution data (PhD student Ratneel Deo) with Dr. Tristan Salles and Prof. Jody Webster

Collaboration

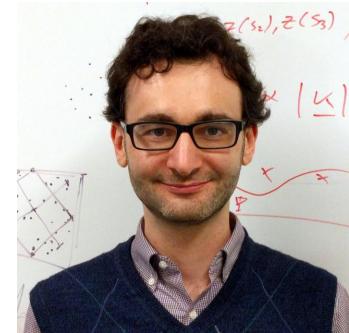
A/Prof Willem Vervoort
Director, DARE
University of Sydney



A/Prof. Martin Anderson
UNSW



Prof. Robert Kohn
UNSW



Dr. Pavel Krivitsky - UNSW

Applications of Bayesian Deep Learning for DARE problems (looking for further collaboration)

References

1. Chandra, R., Cripps, S., Butterworth, N., & Muller, R. D. (2021). Precipitation reconstruction from climate-sensitive lithologies using Bayesian machine learning. *Environmental Modelling & Software*, 139, 105002.
2. Shirmard, H., Farahbakhsh, E., Müller, R. D., & Chandra, R. (2022). A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sensing of Environment*, 268, 112750.
3. H. Shirmard, E. Farahbakhsh, E. Heidari, A. B. Pour, B. Pradhan, R. D. Müller and R. Chandra. A comparative study of convolutional neural networks and conventional machine learning models for lithological mapping using remote sensing data. *Remote Sensing* (to appear in January 2022)
4. J Diaz-Rodriguez, RD Muller, R Chandra. Predicting the emplacement of Cordilleran porphyry copper systems using a spatio-temporal machine learning model. *Ore Geology Reviews* 137, 2021: 104300
5. Chandra, R., Jain, M., Maharan, M., & Krivitsky, P. N. (2021). Revisiting Bayesian Autoencoders with MCMC. *arXiv preprint arXiv:2104.05915*. <https://arxiv.org/abs/2104.05915>
6. A Sharma, Prabhat Kumar Singh, R Chandra, "SMOTified-GAN for class imbalanced pattern classification problems". <https://arxiv.org/submit/3871315/view>
7. S. Chandra, M Lindsay, S Clark, and R. Chandra, Drill-core analysis for mineral composition with an unsupervised machine learning framework, *Ore Geology Reviews*, 2022 (In review)
8. Chandra R; Azam D; Kapoor A; Dietmar Müller R, 2020, 'Surrogate-assisted Bayesian inversion for landscape and basin evolution models', *Geoscientific Model Development*, vol. 13, pp. 2959 - 2979
9. Pall J; Chandra R; Azam D; Salles T; Webster JM; Scalzo R; Cripps S, 2020, 'Bayesreef: A Bayesian inference framework for modelling reef growth in response to environmental change and biological dynamics', *Environmental Modelling and Software*, vol. 125, pp. 104610 - 104610
10. Chandra R; Müller RD; Azam D; Deo R; Butterworth N; Salles T; Cripps S, 2019, 'Multicore Parallel Tempering Bayeslands for Basin and Landscape Evolution', *Geochemistry, Geophysics, Geosystems*, vol. 20, pp. 5082 - 5104

Open Source Software

All papers provide open source software and data via Github repository

Many thanks to everyone for attending and special thanks to everyone behind the scenes.