Predicting CO\textsubscript{2} Buoyant Flow Saturation in Heterogeneous Geologic Formations with Machine Learning

HAILUN NI\textsuperscript{1}, SAHAR BAKHSHIAN\textsuperscript{1}, ANTHONY ZUNIGA\textsuperscript{2}, TIP MECKEL\textsuperscript{1}

\textsuperscript{1}GCC, BEG, The University of Texas at Austin
\textsuperscript{2}The Department of Computer Science and Engineering, University of North Texas
Sub-meter scale barriers can determine migration pathways, speed of plume movement, and CO₂ storage capacity

CO₂ retained per grid block: 3% -> 48%
CO₂ retained: 10% -> 21%
Sub-meter scale barriers can determine migration pathways, speed of plume movement, and CO$_2$ storage capacity

Increasing grain size contrast between matrix and laminae

(Increasing degree of heterogeneity)

\[ S_{NWP} = 0.27\% \]
\[ S_{NWP} = 3.24\% \]
\[ S_{NWP} = 31.5\% \]
\[ S_{NWP} = 36.8\% \]

CO$_2$ retained per grid block: 3% → 48%
CO$_2$ retained: 10% → 21%

Krishnamurthy, 2020
Why do modified invasion percolation simulations?

**3D representation of the monitored CO$_2$ plume**

**Modified invasion percolation simulation result**

Cavanagh & Haszeldine, 2014
Heterogeneity in natural geologic formations is affected by two major factors:

- **Grain size**: Trevisan et al., 2017
  - Coarse: Upper, Lower
  - Medium: Upper, Lower
  - Fine: Upper, Lower
  - Very fine: Upper, Lower

- **Bedform architecture**: Rubin & Carter, 2005; Meckel et al., 2017
  - Laminae = 1
  - Matrix = 0
Data generation: modified invasion percolation simulations

- 59 bedform architectures
- 40 grain size contrast cases
- 50 stochastic realizations
- = 118,000 simulations run

Rubin & Carter, 2005; Meckel et al., 2017; Trevisan et al., 2017
Data: first look

![Graph showing contrast between matrix and laminae]

Average CO₂ saturation (%) vs. Contrast between matrix and laminae, δ (-)

Bedform:
- # 3
- # 5
- # 13
- # 19
- # 27
- # 36
- # 42a
- # 63
- # 4
- # 15
- # 16
- # 17
- # 18
- # 21
- # 22a
- # 22b
- # 25
- # 29
- # 32a
- # 32b
- # 32c
- # 34a
- # 34b
- # 34c
- # 38
- # 40
- # 42b
- # 43a
- # 43b
- # 45
- # 46a
- # 46b
- # 46c
- # 46d
- # 46e
- # 46f
- # 46g
- # 46h
- # 46i
- # 46j
- # 46k
- # 46l
- # 46m
- # 46n
- # 55
- # 56
- # 58
- # 59
- # 65
- # 66
- # 67
- # 69
- # 71
- # 72
- # 73
- # 74
- # 77
- # 78
- # 79
Model training: training and test set

All data

59 Bedforms

Training set

51 Bedforms

Test set

8 Bedforms

Each bedform architecture model has all of its 40 grain size contrast cases included.
Model results: first model

\[ R^2 = 0.86 \]
\[ RMSE = 8.7 \]
Model building: with machine learning

- Add more features
  - Grain sorting
  - Geological entropy
  - Bedform descriptors
    - Planform shape
    - Shape and behavior through time
    - Crest orientation
    - Lamination type and shape

- Try different machine learning regression models
  - K nearest neighbors
  - Linear regression
  - Tree-based ensemble models
    - Random forest
    - Gradient-boosted trees
  - Artificial neural networks
Model building: feature selection
Model results: second model

• Random forest model

\[ R^2 = 0.95 \]
\[ RMSE = 5.3 \]
Model results: feature importance

Feature importance bars for various features, with 'Grain size contrast' being the most important.
Important features: grain sorting and laminae grain size

Symbol size represents absolute laminae grain size
Important features: lamination type and shape

• High CO\textsubscript{2} saturation:
  • Continuous ripple lamination

• Low CO\textsubscript{2} saturation:
  • Discontinuous cross-lamination
Validation: experiments

(a) Exp. A  Exp. B

Exp. C  Exp. D

End of Drainage

CO₂ saturation

High

Low

\[\left(\langle S_{CO₂}\rangle \right) \text{%}\]

\[0.00 \quad 10.00 \quad 20.00 \quad 30.00 \quad 40.00 \quad 50.00 \quad 60.00\]

 Experimental  Predicted  Simulated

\[R^2 = 0.80\]

\[RMSE = 7.2\]

Krishnamurthy, 2020
Conclusions

Single-feature model

$R^2 = 0.86$
$RMSE = 8.7$

Multi-feature model

$R^2 = 0.95$
$RMSE = 5.3$

Feature importance:
- Grain size contrast
- Lamina median grain size
- Bottom filled volume
- Lamination type_ripple
- Entropic scale
- Grain sorting_Extremely well sorted
- Global entropy
- Planform shape_3D-superimposed
- Grain sorting_Very well sorted
- Grain sorting_Moderately sorted
- Crest orientation_oblique
- Grain sorting_Well sorted
- Shape and behavior through time_variable
- Crest orientation_transverse
- Crest orientation_Longitudinal
- Planform shape_2D

Graph showing predicted vs. actual $S_{CO2}$ values.
Important features: geological entropy and entrograms

Bianchi & Pedretti, 2017; 2018
Potential model use case: upscaling critical CO₂ saturation for heterogeneous domains

Threshold capillary pressure

P1

P2

P3

P4

CO₂ saturation

Irreducible water saturation

Water saturation

1 − critical CO₂ saturation