Machine learning based prediction of channelisation during dissolution of carbonate rocks

Florent Brondolo[1], P A Cilli[1], I B Butler[1], A P Fraser-Harris[1], Katriona Edlmann[1], and Christopher Mcdermott[1]

[1]University of Edinburgh

November 24, 2022

Abstract

Evolving preferential dissolution channels are common features formed during reactive fluid flow in carbonate rocks. Understanding these is of particular importance in applications involving subsurface engineered reservoirs but predicting their progression is currently challenging and poorly understood. Here, we propose a new approach to predict both the spatial distribution and extent of dissolution using a combination of experimental work, X-ray microtomography (μCT) and machine learning. We have conducted experiments, under reservoir conditions of temperature and pressure, involving pre- and post-flooding μCT characterisations, and coupled the outputs with a neural network to predict locations where carbonate was most likely to be dissolved. Our simulations demonstrate that our new solution can identify the key geometrical features that are important during dissolution, and can accurately predict the location and spread of dissolution. An important benefit of this approach is that it can accurately predict dissolution channels through forward prediction, while it does not require further chemical parameters, using instead common and accessible variables.

Machine learning based prediction of channelisation during dissolution of carbonate rocks

F, Brondolo[1,2,3], P. A. Cilli[1,2,4], I. B. Butler[1,2], A. Fraser-Harris[1,2],K. Edlmann[1,2]and C. McDermott[1,2]

[1]School of Geosciences, University of Edinburgh, Edinburgh, UK

[2]International Centre for Carbonate Reservoirs, Edinburgh, UK

[3]Now at Stanford University - SLAC, Menlo Park, USA

[4]Now at Department of Earth Sciences, University of Oxford, Oxford, UK

Key Points:

- Artificial neural network
- Machine learning
- µCT
- Channels
- Carbonates

Corresponding author: F. Brondolo, `brondolo@slac.stanford.edu`

Abstract

Evolving preferential dissolution channels are common features formed during reactive fluid flow in carbonate rocks. Understanding these is of particular importance in applications involving subsurface engineered reservoirs but predicting their progression is currently challenging and poorly understood. Here, we propose a new approach to predict both the spatial distribution and extent of dissolution using a combination of experimental work, X-ray microtomography (µCT) and machine learning. We have conducted experiments, under reservoir conditions of temperature and pressure, involving pre- and post-flooding µCT characterisations, and coupled the outputs with a neural network to predict locations where carbonate was most likely to be dissolved. Our simulations demonstrate that our new solution can identify the key geometrical features that are important during dissolution, and can accurately predict the location and spread of dissolution. An important benefit of this approach is that it can accurately predict dissolution channels through forward prediction, while it does not require further chemical parameters, using instead common and accessible variables.

1 Introduction

Injection of fluid into carbonate reservoir rocks is a widely used process involved in subsurface engineered reservoirs to manage permeability and fluid flow (geothermal, groundwater management, carbon sequestration, enhanced oil recovery, etc.). The injected fluid creates changes in the fluid dynamic and stress state, leading to dissolution where the pore network, chemistry, temperature, fluid composition and pressures all influence the location, degree, and spread of the preferential channelling (Hoefner & Fogler, 1988; C. N. Fredd & Fogler, 1998; Golfier et al., 2002; Menke et al., 2017). The reactive and heterogeneous nature of carbonates make predictions of fluid behaviour challenging, and much work has been done on channelisation and classification in a variety of fluid-mineral systems as a function of the fluid flow rate and the fluid properties (Hoefner & Fogler, 1988; Steefel & Lasaga, 1990; Frick et al., 1994; Bazin et al., 1995; C. Fredd et al., 1996; C. N. Fredd & Fogler, 1998; Golfier et al., 2002; Walle et al., 2015), where fluid properties have been identified as largely controlling dissolution and channelisation in carbonates (Golfier et al., 2002). Numerical modelling studies have attempted to recreate these

dissolution processes by including variables influencing the general shape and spread of the dissolution footprint, such as system pressure, permeability, velocity of the fluid, or diffusion rate across boundary layers. Models have been tested, from a conceptual approach that considered a pre-existing cylindrical wormhole (Hung et al., 1989; Wang et al., 1993; Buijse et al., 1997; Huang et al., 1997, 1999), to more complex approaches focusing on the grain scale (Hoefner & Fogler, 1988; Daccord et al., 1989), on the fluid mechanics (Daccord, Lenormand, & Lietard, 1993; Daccord, Lietard, & Lenormand, 1993), or the mass and flow transfer (Liu et al., 1997; Chen et al., 1997). Most of these approaches displayed reasonable qualitative results of channel geometry and were backed by experimental outputs, against computationally expensive treatments operating over millimetre scale volumes. Here, we have coupled experimental work and Artificial Neural Networks (ANNs). The benefit of ANNs stems from the non-linear aspect of the solving algorithms coupled with their ability to learn and recognise patterns (Basheer & Hajmeer, 2000). Although studies have joined µCT imaging and machine learning as a segmentation tool for 3D volumes (Cortina-Januchs et al., 2011; Chauhan et al., 2016) and for rock modulus estimations (Sonmez et al., 2006), no work has been published on predictions of the spatial distribution of carbonate dissolution, purely relying on µCT images. Our approach has the advantage that it works as a predictive tool for channel spatial distribution, spread, and magnitude, over centimetre large volumes, in relatively short computational times. We have combined experimental data with an ANN to develop a predictive tool for preferential flow-path development.

The model presented in this study used datasets that were generated during experimental investigations of reactive fluid flow in carbonate samples. We investigated channels development through carbonate samples of heterogeneous nature by comparing the ANN computed solutions to 4 experimental results. For the experimental fluid flows, we used a range of flow rates and these have been named High Flow Rate 1 and 2 (HFR 1; HFR 2), Medium Flow Rate (MFR), and Low Flow Rate (LFR). Our coupled numerical work included a pre-processing of pre-experimentally tested core samples µCT-scans followed by a training of the ANN against the post-experimental channels data. The pre-experimental data - referred to as Input data - were based on 18 + 1 variables describing the geometrical attributes of the pore network (steps A and B in Figure 1). The differential result between pre- and post-

experimental µCT-scans allowed to pinpoint dissolution channels, leading to the generation of the Signature dataset (steps C and D in Figure 1). We trained multiple hidden layer ANNs on six datasets (including Input and Signature datasets) and blindly predicted on two (including the Input dataset only), corresponding to the four experimental regimes (further explained in Section 2.1). By doing so, we have been able to favourably predict the occurrence, shape, and magnitude of the dissolution pathways evolution in heterogeneous carbonate rocks using only attributes extracted from µCT scans, before flooding, on representative volumes. Moreover, the processing times of our solution were significantly smaller than the various computationally expensive systems models (Budek & Szymczak, 2012), with the non-negligible advantage of using larger cuboids inputs (Blunt et al., 2013; Bijeljic et al., 2004).

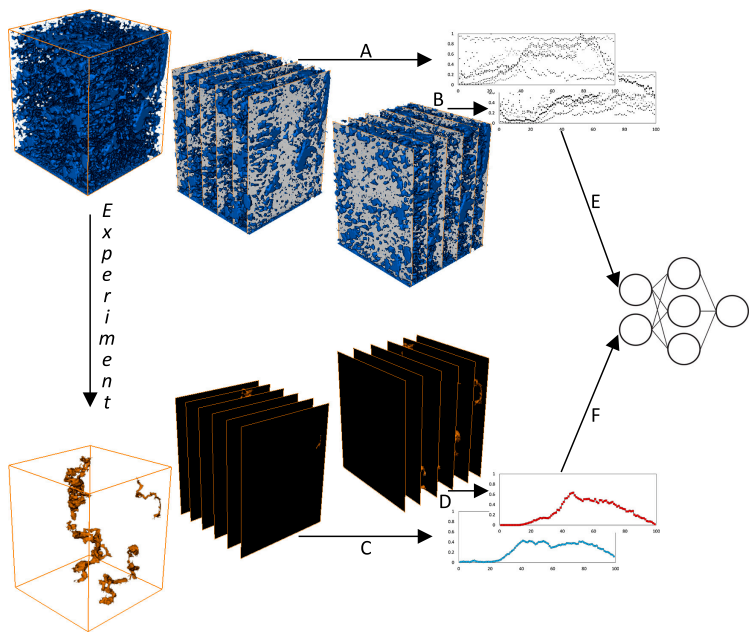

Figure 1: Data acquisition workflow. The two sub-sampled cuboids (pre-flooding and channels) are of the same size. Both stacks have been re-sliced a hundred times in two orthogonal directions. A & B: input data acquisition; C & D: signature data (or true solution) acquisition; E & F: input and signature data as attributes for the ANN.

2 Materials and Methods

2.1 Experimental matrix and dissolution regimes

The experimental dataset used to develop our methodology, and train and test the ANN, comprised a set of four experiments on highly heterogeneous - in porop-ermeability - travertine samples. Each core was 3.8 cm in diameter and differing in length (6.8 cm $< L <$ 8.1 cm). The experimental procedure started with a pre-experimental preparation and conditioning of the core samples, followed by µCT acquisitions of the clean cores. The post-experimental process consisted of sonicat-ing the samples in distilled water, before drying them for a week at 65 °C for one week, followed by post-experimental µCT acquisitions. The experimental flooding consisted in injecting an artificially made seawater of known pH (cf. supplemen-tary information). The four experiments were carried under realistic geo-reservoir conditions of pressure and temperature (temperature $T = 60$ °C). The effective stress used in this study refers to the work of Terzaghi (1951), while the pore volume rate (PV_{rate}) used in this study is described by: $PV_{rate} = Q(t)/V_p$, with $Q(t)$ the amount of fluid injected per minute logged (m^3) and V_p the volume of pore of the rock sample (m^3). The porosity is calculated before the experiments, using the triple weighing technique (Luquot et al., 2016), and displays an average value of ∼11 % (from ∼5 % to ∼14 %). Table 1 presents the four experimental scenarios.

Experiment	Flow rate (cm^3/min)	PV_{rate} (-)	Eff. stress (MPa)	Conf. pressure (MPa)
HFR 1	15.58	2.6	10	50
HFR 2	14.25	2.5	40	50
MFR	6.24	1	10	50
LFR	1	0.2	40	50

Table 1: Flow rate, pore volume rate, effective pressure and confining pressure used for the four experimental floodings. Further petrophysical and chemical data on the rock samples are given in supplementary information.

2.2 µCT Processing

2.2.1 Data Acquisition

For each core, pre- and post-experimental flooding, tomographic data were acquired at 130 kV, and 25 W target power loading. Each dataset consists of 2,000 projections; each of 2 s duration, during a 360° revolution. Reconstruction by filtered back-projection used Octopus v8.7 software (Dierick et al., 2004), while post-processing data analysis and registrations of the pre- and post-flooding dataset of each rock, followed by the processing of the differences between both stacks were done using Fiji (Schindelin et al., 2012) and Avizo®9 functions.

2.2.2 Channel Resolution

The channels formed during our experimental fluid flooding can be detected through image processing by processing the difference between the pre- and post-experimental µCT volumes, while taking into account the initial porosity. The two types of datasets were generated. The first one represented the 3D volumes of pre-experimental scans - referred to as input data. These datasets were used for training the ANN and/or predicting the preferential pathway(s) location and magnitude. The second one were 3D volumes of dissolution channels - referred to as signature data. These datasets represent the true solution of channel(s) formation and were used for training the ANN.

As a way to account for the difference in samples sizes, we sub-sampled the 3D stacks into cuboids of variable side lengths (550 px to 650 px large) and constant axial length of 710 slices (∼2.5 cm by ∼2.9 cm). Both input and signature volumes of a single core sample are sub-sampled at the same location: To save further computational time, the 3D sub-sampled volumes were re-sliced a hundred times in two orthogonal directions - each sampled direction creating a dataset which we treated as independent - with respect to the axial axis and the original orientation of the sample within the sub-sampled core. Figure 1 presents the workflow for µCT data acquisition. In total, we have scanned four samples, translated into eight datasets, which were later divided into training & validation data (six datasets) and blind test data (two datasets) when developing the ANN.

2.3 Modelling

2.3.1 Input Data

The input data were a set of calculated geometric, physical, and simple statistical variables for predicting material loss during experimental floodings. Extraction of the information involved a conversion from the 16-bit grayscale 2D slices stacks to normalized 1D variables that can be evaluated by the ANN.

The formating of the input data has been done via a Visual Basic for Applications (VBA) batch coupled with Corel® X7 suite has been used as a quick way to apply the same formatting to each 2D slice. The batch automatically and sequentially thresholded, smoothed, vectorized, and resized each image to its original size.

For each formatted 2D slice, a set of nineteen relevant variables were collected through an in-depth image analysis. Most variables can easily be explained through image interpretation and simple mathematics (1, 6, 7, 9, 10, 13, 14, 15, 17), as well as variable 18 (PV_{rate}, cf. Section 2.1). The remaining variables (2, 3, 4, 5, 8, 11, 12, 16, 19) have been calculated using a bespoke pre-processor which performed calculations on the equivalent elliptical shapes of each pore and the pore network (cf. supporting information). Figure 2 presents a simplified workflow for pore network generation. The rationale behind the use of a 2D network of pores rather than a 3D pore network skeletonization enables the network to operate on a desktop, where our networking software could extract a simplified set of 2D attributes which highlighted the key characteristics encountered during a 3D analysis. Studies have simplified the complex structure of the pores by the ellipse equivalent shape of a pore (Fournier et al., 2011), while Tsukrov (Tsukrov & Kachanov, 1993) demonstrated that elongated pores could be replaced by their ellipse-equivalent shape for DEM modelling. This network, generated for each scanned slice, was based on the arrangement and overlapping state of the 2D porosity: Each pore of a 2D slice was replaced by an ellipse of equivalent area, shape, and orientation. These ellipses were then enlarged by a constant factor. This enlarging factor, called the *area of influence*, was a computational way of representing the hydrogeological influence of a pore around its neighbourhood; or the numerical way of imaging the 3D influence of a pore on a 2D slice. The 2D arrangement of a set of links symbolized a pore network. A link set between two pores suggested the potential existence of a pathway between these two

pores in the rock. The analysis of this pore network allowed the calculation of the variables 2, 3, 4, 5, and 19, while the analysis of the ellipses defined the variables 8, 11, 12, and 16.

2.3.1.1 Pore network analysis (variables 2, 3, 4, 5, 19)

Variables 2, 3, 4, 5, and 19 are determined as follows: the area of connected pores (2) is the sum of the 2D area of the pores which are part of a connected network. The total (3), median (4) and mean length connection (5) are basic mathematical calculations using the length of every link from a 2D slice. Finally, the I/O connection (19) is a variable which is not part of the variables processed by the ANN, but rather an independent measurement used in determining the potential breakthrough location. A recursive function analyses the 2D network of links, and detects if at least one path between bottom to top of the image is found.

2.3.1.2 Ellipse analysis (variables 8, 11, 12, 16)

The ellipse shape of a pore can resolve the following variables: the ratio of pore area (8) represents the ratio between the area of the largest pore over the mean pore area of a 2D slice. This variable is used for excluding large outliers. Both the small (11) & the large (12) ellipse perimeter are calculations of both the ellipse shape of a pore and its enlarged version. The mean aspect ratio (16) is represented by the ratio of the minor axis b over the major axis a of an ellipse.

2.3.2 Signature Data Pre-processing

The signature data refers to the estimated channels magnitudes and locations. This dataset was computed from the differential result between the pre- and post-experimental scans. Our methodology involved a registration of both unaltered and altered datasets into the same 3D space, allowing us to further subtract both stacks in order to account for potential differences. Isolating and computing the dissolution channels and sub-sampling has been dine under Avizo ®9. The cuboids were re-sliced and thresholded using the Fiji *AutoTresholding* function (Schindelin et al., 2012; Ridler et al., 1978). The percentage of black and white area was calculated for each 2D slice using a batch code based on the *Measure* function under Fiji (Schindelin et al., 2012), and was used as the true solution of the channel shape and size for a set of slices of the 3D stack. The signature data were finally normalized

so that the maximum percentage area of white equals 1 (presence of dissolution channel) while the minimum was equal or close to 0 (no dissolution detected).

2.3.3 Regression and Neural Network Modelling

In this work, we trained both linear regressions and multiple hidden layer ANNs on six datasets and predicted on two, which corresponded to the four experimental regimes as explained in Section 2.1. With this network trained, we then predicted the remaining two, blind datasets' spatial channel signatures, corresponding to the remaining experimental regimes. For the four combinations of three training experimental regimes (six datasets) and one blind test experimental regime (two datasets), we performed a linear regression where eighteen normalized features were input and fitted to minimize the least-squares misfit when compared to the measured spatial channel signature extracted from before and after µCT scans, as described in Section 2.3. We did not perform any regression or model training using the I/O variable (19), which was held aside for comparison, as seen in Section 2.3.1. After the linear regression was parametrized, we performed modelling using an ANN based on the MATLAB's Deep Learning Toolbox (Hudson Beale et al., 2018), with three hidden layers, consisting of 11, 8, and 5 neurons respectively. All eighteen variables were normalized, as discussed in Section 2.3.1, before inputting into the network. All transfer functions between the input and all hidden layers in the ANN were hyperbolic tangent functions. The transfer function between the last hidden layer and the output layer was linear. Our experimental aim was to train the ANN on three experimental regimes (six datasets; three rocks), and predict channel formation on a fourth experimental regime (two datasets; one rock).

We have randomly partitioned the data from six training datasets into 74% training data and 26% validation data. We trained an ANN given this random partitioning of training and validation data and forward-modelled the spatial channel signature on the remaining two blind test datasets. This workflow has been repeated 3,000 times independently, each time training a new network given a different random partitioning of training and validation data from the same six datasets, and produced 3,000 predictions of the two blind datasets' spatial channel signatures. This amount of iteration allowed to obtain large enough outputs in reasonably short computing times (half a day for ANN training over 3,000 iterations, while predicting processing

was done within minutes). These 3,000 predictions have been made into a density plot which shows the most likely spatial channel signature, as well as the sensitivity of the network to the partitioning of input data into test and validation datasets. We performed this workflow for all four combinations of three training experimental regimes (six datasets) and one prediction experimental regime (two datasets), allowing us to simulate four, independent experiments.

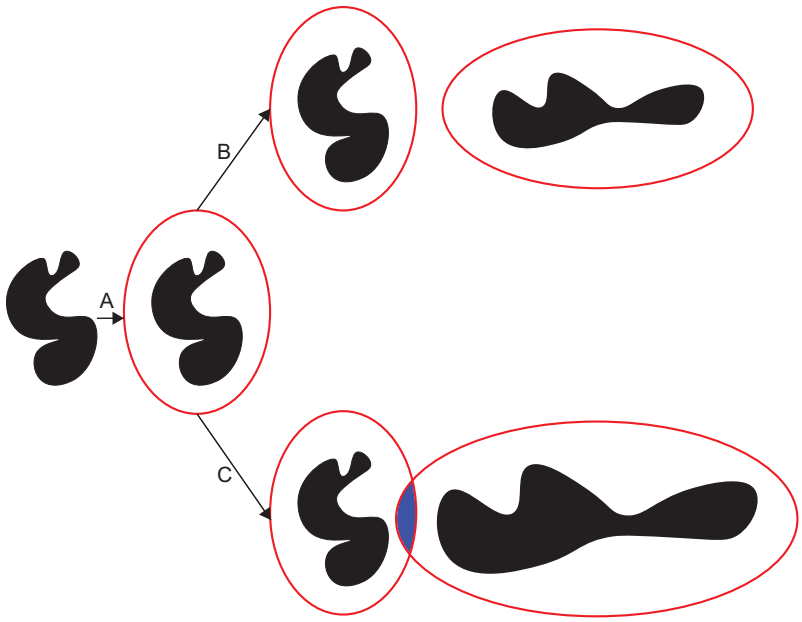

Figure 2: Steps for the detection of a link between two pores (black shapes), virtually representing a suspected connection between two pores in a rock. The step (A) represents the area of influence applied around a pore through a multiplier of the original pore area. Case (B) shows two non-overlapping pores. Case (C) displays a case of overlapping ellipses.

3 Results

3.1 Example of Post-experimental µCT Results

Figure 3 presents an example of signature from two datasets (blue curve; cf. steps C and D in Figure 1) plotted under their corresponding channel. The background greyscale images are the last image from the image stack in the Y-axis (left) and X-axis (right), and are displayed in a way to contextualize the channels in their

volumes. The cross-plotting of the thresholded percentage area of black to white of a channel offers a good insight into the location, the spread, and the magnitude of the created pore space. The values of the signature data were cross-normalized between datasets.

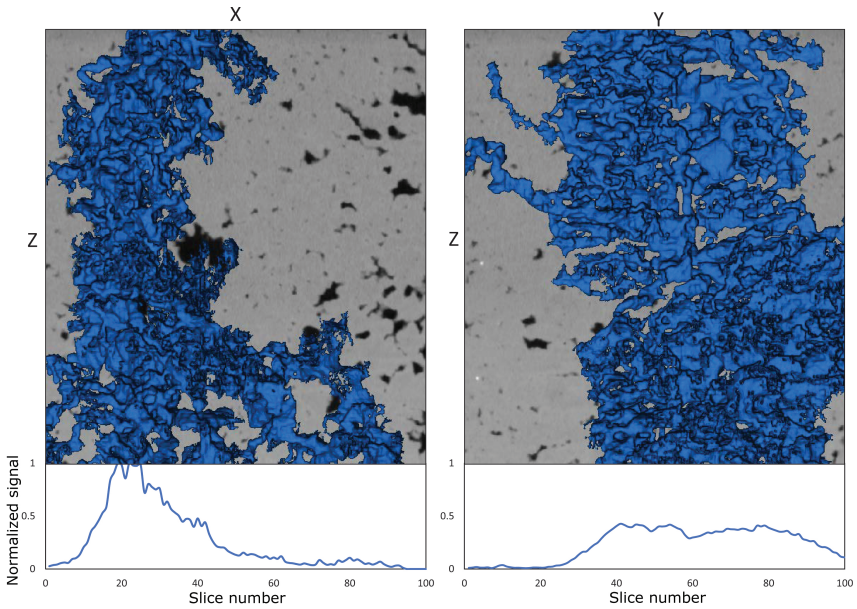

Figure 3: Example of signatures plotted under their respective cross-sectional direction for the HFR 1 experiment. Left: XZ direction; Right: YZ direction. The blue volumes represent the material removed after fluid flow through the core sample, while the plot underneath each graph represents the intensity of this material removal, per direction.

3.2 ANN Outputs

The predicted spatial channel signatures from the fitted linear regression models, as seen in Figure 5, are displayed as white curves, with the signatures measured from µCT scans displayed as red curves. Table 3 shows the percentage decrease in RMS error when predicting channel location and magnitude with the ANN over a linear regression, with values ranging from 26.5% to over 90% decrease in error. Moreover, the linear regressions model was generally ineffective at predicting spatial channel signatures. A ranking of the linear regression weights for all features is given in Table 2 and Figure 4 over all four training scenarios. In Figure 4, we show the

linear regression weights for all features. The last feature shown (19) is the size of the constant term or bias in the linear regression.

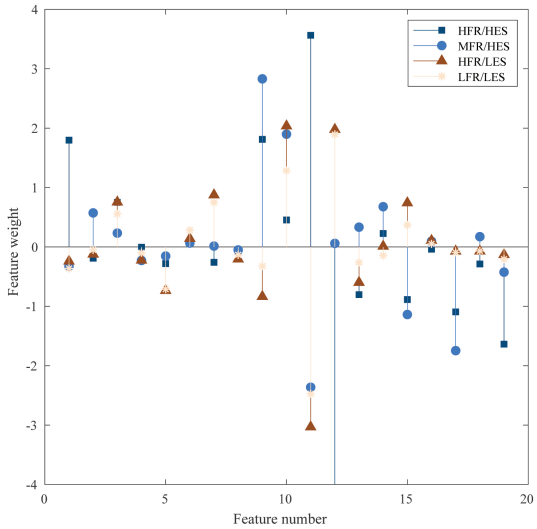

Figure 4: Linear regression weights for all features. The outlying weight on feature 12 is -6.22.

As all features were normalized before performing linear regression, we propose that features that ended with small weights were related to physical attributes which had little effect on a channel's formation. By this reasoning, we interpret that features 2, 4, 6, 8, 16, and 18 all correspond to physical properties which had little influence. These features are the area of connected pores, the median length of connections, the median pore area, the ratio of pore area, the mean ellipse aspect ratio, and the mean distance between pores. By the same reasoning as above, we interpret features with larger weights as proxies for the rocks' physical attributes which broadly exerted a stronger influence on channel formation. These features include numbers 9 to 12, which correspond to the number of pores, the mean pore perimeter, the small ellipse perimeter, and the large ellipse perimeter respectively. Of course, this reasoning of inferring feature influence on enhanced permeability of pre-existing pathway from linear regression weights is flawed as normalized features with large outliers may require large scaling to minimize their fitting residuals. Also, as we see in many cases, linear regression is not an effective predictor of channel formation on blind test data. Nevertheless, this analysis offers a crude, qualitative first estimate

of which features may or may not be important in channel formation prediction processes within heterogeneous carbonate rocks featuring macropores.

#	Key parameters	Influence
1	Total pore area	Medium
2	Area of connected pores	Low
3	Total length connection	Low
4	Median length connection	Low
5	Mean length connection	Low
6	Median pore area	Low
7	Mean pore area	Medium
8	Ratio pore area	Low
9	Number of pores	High
10	Mean pore perimeter	High
11	Small ellipse perimeter	High
12	Large ellipse perimeter	High
13	Porosity	Medium
14	Number of pore greater than mean size	Low
15	Number of pore greater than median size	Medium
16	Mean ellipse aspect ratio	Low
17	Mean distance between pores	Medium
18	Pore volume rate	Low
19	I/O connection	-

Table 2: Summary of the key parameters used in this study and their apparent relative influence on preferential channel formation.

4 Discussion

The background density plots (Figure 5) show the distribution of blind predictions generated by the 3,000 neural network simulations, given each combination of three training experimental regimes (six datasets) and one blind test experimental regime (two datasets). Intensity ranges from low (blue) to high (yellow) number of

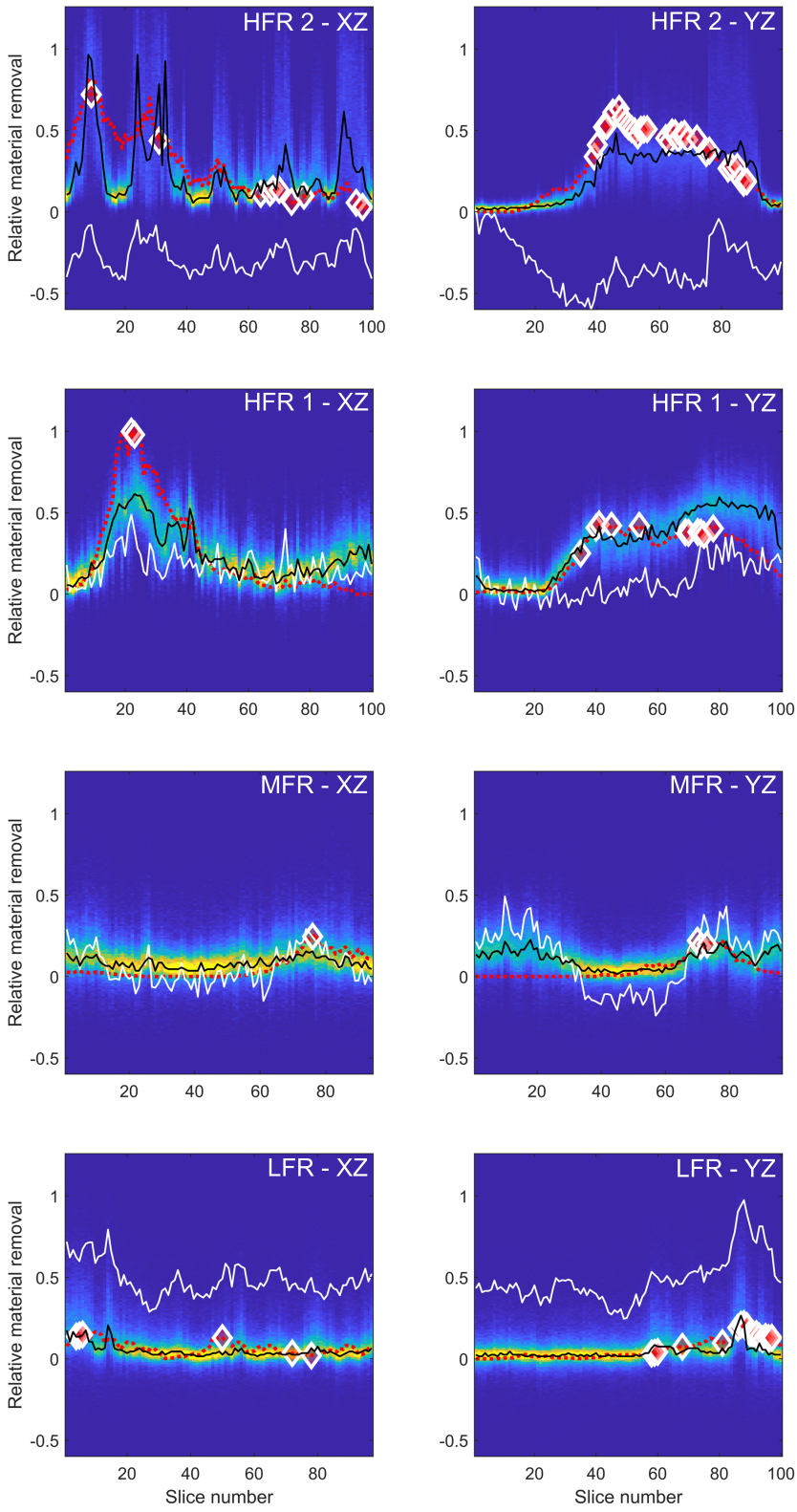

Figure 5: Measured data (red) overlain on density plot of 3,000 trained neural networks' blind test predictions for the four experiments. The white lines show the fitted linear regression models. The black lines represent the best fit for the network blind test predictions. The diamonds are the non-neural network solutions for regions of likelihood for breakthrough independently computed using variable 19 (not included in the neural network predictions).

Core name	ANN RMSE	Linear Reg. RMSE	Difference in RMSE (%)
LFR XZ	0.04	0.45	-90.6
LFR YZ	0.04	0.43	-90.2
HFR 2 YZ	0.13	0.23	-42.5
HFR 2 XZ	0.18	0.27	-35.9
HFR 1 YZ	0.11	0.66	-83.4
HFR 1 XZ	0.20	0.59	-65.7
MFR YZ	0.09	0.19	-51.3
MFR XZ	0.07	0.09	-26.5

Table 3: Percentage of the decrease in root-mean-square-error between linear regression and our ANN solution (respectively white and black lines in Figure 5). Our solution shows an increase in prediction quality of up to ∼90%.

solutions in the bins histogram. The black lines represent the averaged best solutions from our predictions. The "true" solutions (signatures) have been plotted in red on top of each result. We observed a generally effective prediction of dissolution channel location and magnitude by our approach for all four training and prediction experiments. There was a notable improvement of accuracy over the linear regression modelling. The spread in the density plot at locations with a large spatial channel signature indicated the sensitivity of neural network training to the particular, random segmentation of validation and training data. The red diamonds on Figure 5 indicate where the feature 19 has found an existing pore space connection from sample input to output in the pre-flooded rock. This single feature was an effective predictor for breakthrough and principal channel(s) location. This implied that dissolution channels are likely to occur where there is a pre-existing input/output connection in the rock before flooding. We note this feature only predicted the location, rather than the magnitude. For this reason, the use of our neural network method was beneficial over using only feature 19.

The influence of the key parameters has been assessed through an analysis of the linear regression weights for all the features processed by the ANN (Figure 4). We attributed a rank to a feature by summing, per feature, the weights of each of the

four experiments. A feature was ranked "Low" if the sum S was ≤ 1; "Medium" if $1 \leq S \leq 2$, and "High" if $S < 2$. The parameters ranked as "High" in Table 2 were assumed important as they reflected how physical (eg. fluid dynamic) and chemical variables (such as reactive transport, the acidity of the fluid, chemical interaction, etc.) could have been influenced by the surface of the pores so that the larger a pore perimeter was, the more important the wall surface in contact with the flooding fluid should have been. These effects were positively impacted by the amount of pores present. This also validated the ellipse shape assigned to each pore as a correct simplification of the general shape of a pore (Tsukrov & Kachanov, 1993; Fournier et al., 2011). We noted that the variables linked to the area of the pores were not ranking higher than "Medium" (and most often "Low"). We explained this by the relatively small importance of the area of the pores. While large pores should have driven more fluid, the large perimeter (likely associated with a large area) guaranteed more wall-fluid interaction that could have been associated with a higher degree of alteration. This was consistent with Darcy's law, where the flow will tend to be slower and residence time longer. This was possibly accentuated in the case of travertine rocks by the initial high surface roughness caused by calcite overgrowth in the pores, leading to the very high initial perimeter (although not measurable at the scale of our scans). We also believe that the ANN has made a clear distinction between porosity and number of pores, for reasons similar to that which have been explained above: a large effective porosity could have been associated with large pore areas, while the number of pores remained largely uncorrelated to the area of the pores.

5 Conclusions

This study offers a new way to accurately predict the location and shape of channels formed during water flooding in carbonates, by coupling Artificial Neural Networks (ANNs) and µCT images. A limited number of studies have already successfully linked these two tools as a segmentation method, and for rock modulus estimation, but none used ANNs for dissolution prediction. While it is commonly stated that the velocity of flow at the inlet of a core sample is the main factor for rock dissolution and/or material removal, this is only part of the story. The findings of our work showed that spatial distribution of the porosity evolution can be

predicted using only the pore network information held by the rock sample; where the micro- or macro-heterogeneities of the porous medium drive the flow instabilities to direct the fluid flow and, as such, chemical removal, towards zones of highest permeability leading to material loss. Our results showed that specific variables stand-out of the ANN analysis, and validate that geometric factors linked to the porosity and pore shape of a rock contain, most of the time, the necessary data for predicting material loss during rock-water flooding. If a linear combination of these µCT-extracted attributes can successfully predict a rock's spatial channel signature, the weights from the linear regression could be considered indicators of the influence of µCT-extracted feature in channel formation. While this statement is valid in heterogeneous travertine rocks, we remain careful with other types of carbonate, or even other types of lithology which have not been tested in this study.

Acknowledgments

The authors would like to thank Petrobras and Shell for their sponsorship of the International Centre for Carbonate Reservoirs (ICCR), and for permission to publish this work from the GeoMeChem project.

The data used in this work are available at the following address: `https://doi.org/10.6084/m9.figshare.c.5335454` (Brondolo et al., 2021).

References

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, *43*(1), 3–31.

Bazin, B., Roque, C., Bouteca, M., et al. (1995). A laboratory evaluation of acid propagation in relation to acid fracturing: Results and interpretation. In *Spe european formation damage conference.* Society of Petroleum Engineers.

Bijeljic, B., Muggeridge, A. H., & Blunt, M. J. (2004). Pore-scale modeling of longitudinal dispersion. *Water Resources Research*, *40*(11).

Blunt, M. J., Bijeljic, B., Dong, H., Gharbi, O., Iglauer, S., Mostaghimi, P., ... Pentland, C. (2013). Pore-scale imaging and modelling. *Advances in Water Resources*, *51*, 197–216.

Brondolo, F., Cilli, P., Fraser-Harris, A., Butler, I., Edlmann, K., & McDer-

mott, C. (2021, Mar). *Machine learning based prediction of channelisation during dissolution of carbonate rocks.* figshare. Retrieved from https://figshare.com/collections/Machine_learning_based_prediction_of_channelisation_during_dissolution_of_carbonate_rocks/5335454/1 doi: 10.6084/m9.figshare.c.5335454

Budek, A., & Szymczak, P. (2012). Network models of dissolution of porous media. *Physical Review E*, *86*(5), 056318.

Buijse, M., et al. (1997). Mechanisms of wormholing in carbonate acidizing. In *International symposium on oilfield chemistry.* Society of Petroleum Engineers.

Chauhan, S., Rühaak, W., Khan, F., Enzmann, F., Mielke, P., Kersten, M., & Sass, I. (2016). Processing of rock core microtomography images: Using seven different machine learning algorithms. *Computers & Geosciences*, *86*, 120–128.

Chen, Y., Fambrough, J., Bartko, K., Li, Y., Montgomery, C., Ortoleva, P., et al. (1997). Reaction-transport simulation of matrix acidizing and optimal acidizing strategies. In *International symposium on oilfield chemistry.* Society of Petroleum Engineers.

Cortina-Januchs, M., Quintanilla-Dominguez, J., Vega-Corona, A., Tarquis, A., & Andina, D. (2011). Detection of pore space in ct soil images using artificial neural networks. *Biogeosciences*, *8*(2), 279–288.

Daccord, G., Lenormand, R., & Lietard, O. (1993). Chemical dissolution of a porous medium by a reactive fluid—i. model for the "wormholing" phenomenon. *Chemical Engineering Science*, *48*(1), 169–178.

Daccord, G., Lietard, O., & Lenormand, R. (1993). Chemical dissolution of a porous medium by a reactive fluid—ii. convection vs reaction, behavior diagram. *Chemical engineering science*, *48*(1), 179–186.

Daccord, G., Touboul, E., Lenormand, R., et al. (1989). Carbonate acidizing: toward a quantitative model of the wormholing phenomenon. *SPE production engineering*, *4*(01), 63–68.

Dierick, M., Masschaele, B., & Van Hoorebeke, L. (2004). Octopus, a fast and user-friendly tomographic reconstruction package developed in labview®. *Measurement Science and Technology*, *15*(7), 1366.

Fournier, F., Leonide, P., Biscarrat, K., Gallois, A., Borgomano, J., & Foubert, A. (2011). Elastic properties of microporous cemented grainstones. *Geophysics*,

76(6), E211–E226.

Fredd, C., Fogler, H. S., et al. (1996). Alternative stimulation fluids and their impact on carbonate acidizing. In *Spe formation damage control symposium.* Society of Petroleum Engineers.

Fredd, C. N., & Fogler, H. S. (1998). Influence of transport and reaction on wormhole formation in porous media. *AIChE journal*, *44*(9), 1933–1949.

Frick, T., Mostofizadeh, B., Economides, M., et al. (1994). Analysis of radial core experiments for hydrochloric acid interaction with limestones. In *Spe formation damage control symposium.* Society of Petroleum Engineers.

Golfier, F., Zarconne, C., Bazin, B., Lenormand, R., Lasseux, D., & QUINTARD, M. (2002). On the ability of a darcy-scale model to capture wormhole formation during the dissolution of a porous medium. *Journal of fluid Mechanics*, *457*, 213–254.

Hoefner, M., & Fogler, H. S. (1988). Pore evolution and channel formation during flow and reaction in porous media. *AIChE Journal*, *34*(1), 45–54.

Huang, T., Hill, A., Schechter, R., et al. (1997). Reaction rate and fluid loss: the keys to wormhole initiation and propagation in carbonate acidizing. In *International symposium on oilfield chemistry.* Society of Petroleum Engineers.

Huang, T., Zhu, D., & Hil, A. (1999). Prediction of wormhole population density in carbonate matrix acidizing. In *Spe european formation damage conference.* Society of Petroleum Engineers.

Hudson Beale, M., Hagan, M. T., & Demuth, H. B. (2018). *Impress - supporting education across europe.* Retrieved from `https://uk.mathworks.com/help/pdf_doc/deeplearning/nnet_ref.pdf`

Hung, K., Hill, A., Sepehrnoori, K., et al. (1989). A mechanistic model of wormhole growth in carbonate matrix acidizing and acid fracturing. *Journal of petroleum technology*, *41*(01), 59–66.

Liu, X., Ormond, A., Bartko, K., Ying, L., & Ortoleva, P. (1997). A geochemical reaction-transport simulator for matrix acidizing analysis and design. *Journal of Petroleum Science and Engineering*, *17*(1-2), 181–196.

Luquot, L., Hebert, V., & Rodriguez, O. (2016). Calculating structural and geometrical parameters by laboratory measurements and x-ray microtomography: a comparative study applied to a limestone sample before and after a dissolution

experiment. *Solid Earth*, *7*(2).

Menke, H., Bijeljic, B., & Blunt, M. (2017). Dynamic reservoir-condition microtomography of reactive transport in complex carbonates: Effect of initial pore structure and initial brine ph. *Geochimica et Cosmochimica Acta*, *204*, 267–285.

Ridler, T., Calvard, S., et al. (1978). Picture thresholding using an iterative selection method. *IEEE transactions on Systems, Man and Cybernetics*, *8*(8), 630–632.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... others (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, *9*(7), 676.

Sonmez, H., Gokceoglu, C., Nefeslioglu, H., & Kayabasi, A. (2006). Estimation of rock modulus: for intact rocks with an artificial neural network and for rock masses with a new empirical equation. *International Journal of Rock Mechanics and Mining Sciences*, *43*(2), 224–235.

Steefel, C. I., & Lasaga, A. C. (1990). Evolution of dissolution patterns. In *Chemical modeling of aqueous systems ii* (p. 212-225). American Chemical Society. Retrieved from `https://pubs.acs.org/doi/abs/10.1021/bk-1990-0416.ch016` doi: 10.1021/bk-1990-0416.ch016

Terzaghi, K. (1951). *Theoretical soil mechanics.* Chapman And Hall, Limited.; London.

Tsukrov, I., & Kachanov, M. (1993). Solids with holes of irregular shapes: effective moduli and anisotropy. *International Journal of Fracture*, *64*(1), R9–R12.

Walle, L., Papamichos, E., et al. (2015). Acidizing of hollow cylinder chalk specimens and its impact on rock strength and wormhole network structure. In *49th us rock mechanics/geomechanics symposium.* American Rock Mechanics Association.

Wang, Y., Hill, A., Schechter, R., et al. (1993). The optimum injection rate for matrix acidizing of carbonate formations. In *Spe annual technical conference and exhibition.* Society of Petroleum Engineers.

Supporting Information for Machine learning based prediction of channelisation during dissolution of carbonate rocks

F, Brondolo[1,2,3], P. A. Cilli[1,2,4], I. B. Butler[1,2], A. Fraser-Harris[1,2],K. Edlmann[1,2]and C. McDermott[1,2]

[1]School of Geosciences, University of Edinburgh, Edinburgh, UK

[2]International Centre for Carbonate Reservoirs, Edinburgh, UK

[3]Now at Stanford University - SLAC, Menlo Park, USA

[4]Now at Department of Earth Sciences, University of Oxford, Oxford, UK

Contents of this file

1. Text
2. Figures S1 to S9
3. Tables S1 to S2

Introduction

The text provides additional details on the preparation process for the rock samples, the µCT data and their processing, the software used to process the µCT data. Additional information are given regarding the variables processed by the artificial neural network solution, and their computation by our bespoke software.

Sample Characteristics

We used travertine rocks, whose formation and morphology were first introduced in the 1960' (Kitano, 1963), and further studied by various authors (Marques Erthal, 2018; Chafetz & Folk, 1984; Pentecost, 1990; Guo & Riding, 1992, 1994, 1998; Chafetz & Guidry, 1999; Kele et al., 2008; Jones & Renaut, 2010; Wright, 2012; Chafetz, 2013; Chitale et al., 2015; Wright & Barnett, 2015; Claes et al., 2017; Erthal et al., 2017; Fouke et al., 2000; Soete et al., 2015; Yagiz, 2009). Travertine rocks are considered good proxies for the upper Pre-salt reservoir rock (shrub framestone) due to petromorphological resemblance between both formations (Marques Erthal, 2018; Claes et al., 2017; Virgone et al., 2013; Borghi et al., 2013; Schröder et al., 2016; Rezende & Pope, 2015; Soete et al., 2015; Chitale et al., 2015; Boyd et al., 2015). Travertines are carbonate formations deposited in lacustrine environments. They are structured in millimetres to centimetres layers of $CaCO_3$ expending upwards. They are highly heterogeneous and can be brittle depending on the burial history of each sample. All four samples' geometrical and physical properties are presented in Table S1. For this study, the core samples have been sourced from large blocs of travertines (shrub layer, sampled from Saturnia Italy).

Travertines are mainly made of calcium; usually $> 90\%CaCO_3$ (Pentecost, 2005). The secondary principal element is magnesium which substitutes Ca and can be found under the form of calcite magnesian, dolomite, or montmorillonite ($(Ca,Mg)CO_3$) in low quantities (Chafetz & Folk, 1984; Pentecost, 2005; Kele et al., 2008; Erthal et al., 2017). The results of an XRD analysis ran on our block of travertines are shown in Table S2.

Figure S1 shows µCT slices taken from each core sample used in this work. The porous media is composed of heterogeneously distributed pores, whose size and shape appeared

to widely vary throughout a slice. A ($LFRLES$ sample) and B ($MFRHES$ sample) displayed a mixture of shrubby porosity coupled with coated bubbles which have been formed due to degassing during the precipitation of the calcite. Both samples evolved towards the shrubby structures seen in C ($HFRHES$ sample) and D ($HFRLES$ sample), which displayed elongated pores and dendritic calcite crystals (not seen under µCT imaging) (Ronchi & Cruciani, 2015).

µCT Imaging Acquisition

Fluids injected through carbonate rocks generate dissolution that physically relates to the formation of wormholes. They are preferential flow-paths that increase fluid conductivity and allow the injection fluid to flow more easily from one end of the sample to another (Hoefner & Fogler, 1988; Fredd & Fogler, 1998; Golfier et al., 2002). These changes in the rock matrix can be detected using µCT scanning acquisition and post-processing of the images, by arithmetically comparing the pre- and post-experimental datasets. The µCT facility of the University of Edinburgh has been used for qualitative and quantitative analyses of these post-flooding damages. A reconstructed volume provides us with a 3D map of absorbed light due to the difference in density of the matrix of a rock (Landis & Keane, 2010). For each core, pre- and post-experimental flooding tomographic data were acquired at ∼130 kV, and ∼25 W target power loading, with a copper energy filter. A rotary stage is placed on top of an adjustable screw, which allowed the scans of different parts of a sample by adjusting its elevation. The instrument had a conventional cone-beam configuration with a Feinfocus 10-160 kV dual transmission/reflection source, a MICOS UPR-160 air

bearing rotary stage and a Perkin Elmer XRD 0822 amorphous silicon 1 MP flat-panel camera with a terbium-doped gadolinium oxysulfide scintillator. The geometry of the cone-beam window was approximately 3 cm wide, which can be less than the length of the cores, therefore each scan of a whole sample was split into two to three sections. Scans consisted of 2,000 projections; each of 2 s duration, during a 360° revolution. They were first recorded as 2D TIFF of projected images, followed by tomographic reconstructions into 3D stacks. Further details about acquisition and reconstruction processes along with the underlying physics of µCT scanning can be found in various studies (Landis & Keane, 2010; Salvo et al., 2010). For each section, reconstruction by filtered back-projection was done by using Octopus v8.7 software (Dierick et al., 2004), while post-processing data analysis was done through Fiji (Schindelin et al., 2012) and Avizo® 9 functions (`https://www.thermofisher.com/uk/en/home/industrial/electron-microscopy/electron-microscopy-instruments-workflow-solutions/3d-visualization-analysis-software.html`).

Additional Information on µCT Imaging Processing

The reconstruction stage of the projected images was done using Octopus v8.7 (`https://octopusimaging.eu/`). Back processing, registration, and analysis were done via Avizo® 9, using the Image Registration Wizard and the Merge tool. Image Registration Wizard assists the user into registering several 3D volumes by mathematically enclosing the overlap between them while the Merge function combines the stacks into one set of data. Our methodology involved the registration of both unaltered (pre-experimental) and altered (post-experimental) datasets into the same 3D space,

allowing us to further subtract both stacks to account for potential differences. As such, we used several tools in Avizo® 9: the Arithmetic operator, Interactive Thresholding, Label Analysis, and the Analysis Filter function. The Arithmetic operator applies a mathematical condition on the input 16-bit datasets; here we wanted to have the absolute difference between the pre- and post-experimental dataset. The Interactive Thresholding module differentiates the volumes based on the intensity values of the raw images. Thresholding was applied over the entire volume, and a voxel was set to black if its intensity I_{ij} is less than a fixed value T set by the user, so that $I_{ij} > T = 1$ and $I_{ij} < T = 0$. The Label Analysis function creates a connected network of voxels depending on their 3D arrangement, i.e. whether they are connected by their sides, corners, or faces. Finally, we used the Analysis Filter which applies a user-defined filter over any geometrical parameter of the 3D volume. We filtered our dataset to isolate the main parts of the channels by removing 3D volumes smaller than a certain amount of voxels. The drawback of imaging 38 mm diameter large samples relies on acquiring the images at a lower resolution, against the high energy needed to scan through the entire sample. As a result, our volumes displayed a voxel side length of 40 µm, which was enough to resolve most of the features that could be expected in these kinds of studies, but too large to obtain a detailed picture of the microporosity.

µCT Imaging Output Data

µCT images have been obtained on dry samples, before and after each experimental flooding. The core samples have been cleaned for ∼15 min in a sonic bath, after the coring/cutting phases and after the flooding stage, to drive the remaining seawater out

the sample. These extra steps have allowed us to obtain clearer outputs from the µCT scanning. The reconstruction process produced ∼2.5 GB 16-bit greyscale images.

The shape and spread of a wormhole have been calculated by evaluating the density of nodes per region of ∼0.01 cm^2. A wormhole formation expands and branches as the injection rate is increased (Siddiqui et al., 2006) implying that the thickness of a wormhole should then be primarily linked to the forcing of the fluid. In this work, we have isolated the main wormholes and branches.

µCT Images: Pre-processing

The pre-processor used in this work is an evolving solution developed over the entire course of this study and is usually referred to as the "ICCR-Macropore". It has been used in two ways:

- As a processor of macro structures of 2D slices. It performed geometrical calculations based on the images and generated spreadsheets which were used for further analysis.
- As a data pre-processor of the OGS software (Kolditz et al., 2012). It generated 2D mesh files of the slices for further HMC modelling. Figure S4 shows an example of input file for OGS used.

The solution uses 2D slices taken from 3D volumes of the image stacks. The volumes were rendered by the µCT scan of a rock sample while the reconstruction process of the 3D rendering was done under Fiji (Schindelin et al., 2012) and Avizo® 9.

Pre-processor Software: Numerical Structure of a Pore

The interpretation of an SVG file (converted TIFF files) by the ICCR-Macropore creates instances or objects. Figure S2 is a simplified Unified Modelling Language (UML) diagram displaying the relationship between the different classes. At this stage, a pore is interpreted as a polygon; a polygon is comprised of nodes, referred to as points; each pair of point forms a line. These classes belong to a single project. The relationship between classes can be summarized as follows:

- A project contains 1 to n Polygon(s).
- A polygon belongs to 1 Project only and is made of n Point(s).
- A point belongs to 1 Polygon only.
- A line is made of 2 Points, and can only belong to 1 Polygon.

Pre-processor Software: Numerical Analysis Workflow

Figure S3 presents the general workflow for using the tool. Four types of input file are necessary (further developed in Section "Pre-processor software: Input files"). The tool offers three options: option 1) and 2) process one or more orthogonal slices, while option 3) observes the evolution of porosity and area of the pores of a 2D image stack in the axial direction. Along with the above-mentioned geometrical analysis, the ICCR-Macropore also creates .GEO and .GLI files of the scanned slices, which are inputs for OGS.

Pre-processor Software: Input Files

The input files are described as follows:

- CNT file: this file is the control file. It contains interpretive and boundary variables. This user-defined file is read by the executable before analysing each image and defines how finely the images should be interpreted.

• SVG file(s): an SVG file is an XML-based vectorized image. The well-documented encoding system allowed us to create reading functions capable of extracting the geometrical data of a 2D slice such as the contour of the pores, their positions, the size of the picture, and the differentiation between pores and background rock matrix. The input picture could take the form of a picture of a clean-cut of a rock or a slice of a 3D volume of a core sample. Each picture was further binarized and converted into an SVG file.

• Comm file: the comm file is a communication file used by the Slice Picker (further described in Section "Slice Picker Software").

• Slice text files: they contain the names of the SVG files that should be processed by the ICCR-Macropore, and take the form of two text files for the two sampled orthogonal direction (X_input.txt and Y_input.text).

Pre-processor Software: Output Files

The tool can extract data from an SVG file format according to the user-defined settings. It produces the following files:

• data.geo is used for meshing a slice (Figure S4). A GEO file can be further edited using Gmesh (Geuzaine & Remacle, 2009), before producing the mesh for modelling purposes. The data.geo file is further used as a mechanical mesh (eg. in Section "Example of 2D analysis of fluid flow").

• data1D2D.geo is used for meshing the links between pores (Figure S7). Similar to the "data.geo" file, the links between pores can be edited and adjusted under Gmesh (Geuzaine & Remacle, 2009) before being meshed. The theory behind link generation is described in Section "Pre-processor software: Ellipse simplification".

• svg.gli is used in OGS (Kolditz et al., 2012) to connect the meshes with the desired system of equations that calculates a solution. The .gli file represents the geometry of the mesh (under the form of points and polylines). These geometric objects connect the mesh to the rest of the OGS code, calculating the solution. An example of such work is given in Section "Example of 2D analysis of fluid flow"

• Geometrical attributes: they are the results of mathematical interpretation of the polygons analysed by the ICCR-Macropore functions. Each of the attributes of interest is described in Section "Pre-processor software: Geometrical attributes".

Pre-processor Software: Geometrical Attributes

Area of each pore: The area of each pore is calculated from any irregularly shaped polygon. Each polygon's perimeter being formed by points of known coordinates (x, y), we use the following formula for Area calculation:

$$a = \left| \frac{(x_1y_2 - y_1x_2) + (x_2y_3 - y_2x_3)... + (x_ny_1 - y_nx_1)}{2} \right| \tag{1}$$

Pore perimeter: The perimeter of irregularly shaped pores is calculated using the distances between each pair of point (x_1, y_1) and (x_2, y_2), so that:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{2}$$

Coefficient A (major axis): The coefficient A represents the largest axis of an ellipse shape of a pore (Figure S5).

Coefficient B (minor axis): The coefficient B represents the smallest axis of the ellipse shape of a pore and is perpendicular to the major axis (Figure S5).

D and l: D and l parameters are the average width and height of each pore, plus the average width and height of the entire porosity of slice.

Distances: This file contains a summary of the distance between each pore. Each pore is defined by a centre point of coordinate (x, y), while the distance between two pores is computed using the same function as for the pore perimeter.

Ellipse perimeter: The perimeter of an ellipse is calculated using Ramanujan's approximation (Ramanujan, 1914), and is expressed as follows:

$$\begin{aligned} h &= \frac{(a-b)^2}{(a+b)^2} \\ p &\approx \pi(a+b)\left(1+\frac{3h}{10+\sqrt{4-3h}}\right) \end{aligned} \tag{3}$$

with a and b being the major and minor axis of the ellipse. This formula has been chosen due to its very good approximation of near-circular ellipses while being very simple to implement.

Ratio: Per 2D slice scanned, the ICCR-Macropore generates an excel file which contains all aspect ratios (a/b) of all ellipses (pores) present on that slice.

Pre-processor Software: Ellipse Simplification

The ICCR-Macropore has been developed as a 2D interpreter of pore network. This network, generated for each scanned slice and saved under the "data1D2D.geo" file, relies on the arrangement and overlap state of the 2D porosity: each pore of a 2D slice is virtually replaced by an ellipse of equivalent area, shape, and orientation. Studies have simplified the complex structure of the pores by the ellipse equivalent shape of a pore (Fournier et al., 2011), while (Tsukrov & Kachanov, 1993) demonstrated that elongated pores could be replaced by their ellipse-equivalent shape for DEM modelling. Although we have not done any DEM modelling, we assume that the ellipse shape has a similar impact on the surrounding pores. The ellipses generated by the ICCR-Macropore are

enlarged by a factor controlled by the user and defined as the *area of influence* and set constant for the entire sample analysis. This factor is a 2D computational way of representing the 3D hydro-geological influence of a pore over its surrounding area. Figure S5 presents a simplified workflow for pore connection detection. The overlap-detection function of ICCR-Macropore uses the geometrical attributes of an ellipse: knowing that the sum of the distances between the foci $F_{1,2}$ and any point of the ellipse E_1 is a constant C_1, we assume that if the distance between foci of E_1 and any point of a second ellipse E_2 is less than C_1, then E_2 overlaps E_1, leading to the creation of a link, referred to as "Line" in Figure S2, between those two pores. A link is a 2D representation of a potential connection in 3D, while the 2D arrangement of a set of links symbolizes the pore network of a 2D slice (Figure S7). The ellipse simplification has the advantage of avoiding corner effects that can be encountered by a rectangular simplification against more resource-consuming processing. Figure S6 shows an example of artefact produced during link detection between two pores: we observe that replacing the pores by rectangular shapes (case A) would have developed a connection (represented by a blue area) where case B does not detect any.

The materialized links are displayed in Figure S7. For this example, a slice has been analysed three times, using three different *area of influence*. As observed in Figure S7-Left, a small factor of 5 only connected the larger pores, which had a large influence on their nearest neighbours. In Figure S7-Middle, a factor of 10 was applied on the same slice. In this case, most of the larger vugs were connected, while the microporosity began to be part of the global hydraulic network. No connection between the top and bottom of the slice is detected yet. With Figure S7-Right, we saw that almost all pores were part

of a network. Only the ones sitting where the rock matrix is most present are avoided, highlighting the poor influence of rock matrix and microporosity during water flooding. The larger vugs have created a bottom-to-top connection.

Slice Picker Software

The Slice Picker is a software written in C++ and specifically developed as a pre-processor for the Artificial Neural Network study, by generating a pre-formatted dataset of variables. The software is based on the ICCR-Macropore main functions.

The following files are needed to run the Slice Picker (Figure S8):

- The executable of the ICCR-Macropore, along with the input files needed for a correct use of the software.
- The Comm file. The "Status" variable should be set to 0 before running the Slice Picker.
- The "Slices" text file, which should contain the names of the files to be analysed.

The Slice Picker generates a formatted excel spreadsheet containing 18 variables, further used by the neural network solution, plus 1 variable (In/Out connection). The In/Out connection is a Boolean offering a rapid and alternative way to control whether a breakthrough might occur on a 2D slice of a pre-experimental scan of a rock. The function, compiled within the executable of the ICCR-Macropore, can only be triggered through the use of the Slice Picker and is based on the recursive analysis of the connected network of pores, based on the *area of influence*. The variable is set to 1 if at least one route connects the bottom of the slice to the top, and 0 otherwise.

Slice Picker & ICCR-Macropore Interactions

The "Status" variable of the Comm file guides both software into splitting tasks (Figure S9). A Comm status of 0 indicates a running Slice Picker, while a Comm status of 1 puts the Slice Picker in a stand-by mode during which the ICCR-Macropore processes the current 2D slice through a set of available and hidden functions only triggered through the use of the Slice Picker. Once done with the slice processing, the ICCR-Macropore updates the Comm file accordingly and closes itself, while the Slice Picker saves the set of 18 + 1 variables in a vector. The process continues until each slice has been analysed, after which an excel spreadsheet containing all the data gathered by the coupled solution is generated.

Acknowledgments. The authors would like to thank Petrobras and Shell for their sponsorship of the International Centre for Carbonate Reservoirs (ICCR), and for permission to publish this work from the GeoMeChem project.

The data used in this work are available at the following address: `https://doi.org/10.6084/m9.figshare.c.5335454` (Brondolo et al., 2021).

References

Borghi, L., Corbett, P. W., et al. (2013). Lacustrine carbonates-for the purpose of reservoir characterization are they different? In *Otc brasil.*

Boyd, A., Souza, A., Carneiro, G., Machado, V., Trevizan, W., Santos, B., ... others (2015). Presalt carbonate evaluation for santos basin, offshore brazil. *Petrophysics*, *56*(06), 577–591.

Brondolo, F., Cilli, P., Fraser-Harris, A., Butler, I., Edlmann, K., & Mc-

Dermott, C. (2021, Mar). *Machine learning based prediction of channelisation during dissolution of carbonate rocks.* figshare. Retrieved from https://figshare.com/collections/Machine_learning_based_prediction_of_channelisation_during_dissolution_of_carbonate_rocks/5335454/1 doi: 10.6084/m9.figshare.c.5335454

Chafetz, H. S. (2013). Porosity in bacterially induced carbonates: Focus on microporesmicroporosity after bacterial decay. *AAPG bulletin*, *97*(11), 2103–2111.

Chafetz, H. S., & Folk, R. L. (1984). Travertines; depositional morphology and the bacterially constructed constituents. *Journal of Sedimentary Research*, *54*(1), 289–316.

Chafetz, H. S., & Guidry, S. A. (1999). Bacterial shrubs, crystal shrubs, and ray-crystal shrubs: bacterial vs. abiotic precipitation. *Sedimentary Geology*, *126*(1-4), 57–74.

Chitale, V. D., Alabi, G., Gramin, P., Lepley, S., Piccoli, L., et al. (2015). Reservoir characterization challenges due to the multiscale spatial heterogeneity in the presalt carbonate sag formation, north campos basin, brazil. *Petrophysics*, *56*(06), 552–576.

Claes, H., Erthal, M. M., Soete, J., Özkul, M., & Swennen, R. (2017). Shrub and pore type classification: Petrography of travertine shrubs from the ballık-belevi area (denizli, sw turkey). *Quaternary International*, *437*, 147–163.

Dierick, M., Masschaele, B., & Van Hoorebeke, L. (2004). Octopus, a fast and user-friendly tomographic reconstruction package developed in labview®. *Measurement Science and Technology*, *15*(7), 1366.

Erthal, M. M., Capezzuoli, E., Mancini, A., Claes, H., Soete, J., & Swennen, R. (2017). Shrub morpho-types as indicator for the water flow energy-tivoli travertine case (cen-

tral italy). *Sedimentary geology*, *347*, 79–99.

Fouke, B. W., Farmer, J. D., Des Marais, D. J., Pratt, L., Sturchio, N. C., Burns, P. C., & Discipulo, M. K. (2000). Depositional facies and aqueous-solid geochemistry of travertine-depositing hot springs (angel terrace, mammoth hot springs, yellowstone national park, usa). *Journal of Sedimentary Research*, *70*(3), 565–585.

Fournier, F., Leonide, P., Biscarrat, K., Gallois, A., Borgomano, J., & Foubert, A. (2011). Elastic properties of microporous cemented grainstones. *Geophysics*, *76*(6), E211–E226.

Fredd, C. N., & Fogler, H. S. (1998). Influence of transport and reaction on wormhole formation in porous media. *AIChE journal*, *44*(9), 1933–1949.

Geuzaine, C., & Remacle, J.-F. (2009). Gmsh: A 3-d finite element mesh generator with built-in pre-and post-processing facilities. *International journal for numerical methods in engineering*, *79*(11), 1309–1331.

Golfier, F., Zarconne, C., Bazin, B., Lenormand, R., Lasseux, D., & QUINTARD, M. (2002). On the ability of a darcy-scale model to capture wormhole formation during the dissolution of a porous medium. *Journal of fluid Mechanics*, *457*, 213–254.

Guo, L., & Riding, R. (1992). Aragonite laminae in hot water travertine crusts, rapolano terme, italy. *Sedimentology*, *39*(6), 1067-1079.

Guo, L., & Riding, R. (1994). Origin and diagenesis of quaternary travertine shrub fabrics, rapolano terme, central italy. *Sedimentology*, *41*(3), 499–520.

Guo, L., & Riding, R. (1998). Hot-spring travertine facies and sequences, late pleistocene, rapolano terme, italy. *Sedimentology*, *45*(1), 163–180.

Hoefner, M., & Fogler, H. S. (1988). Pore evolution and channel formation during flow

and reaction in porous media. *AIChE Journal*, *34*(1), 45–54.

Jones, B., & Renaut, R. W. (2010). Calcareous spring deposits in continental settings. *Developments in Sedimentology*, *61*, 177–224.

Kele, S., Demény, A., Siklósy, Z., Németh, T., Tóth, M., & Kovács, M. B. (2008). Chemical and stable isotope composition of recent hot-water travertines and associated thermal waters, from egerszalók, hungary: Depositional facies and non-equilibrium fractionation. *Sedimentary Geology*, *211*(3-4), 53–72.

Kitano, Y. (1963). Geochemistry of calcareous deposits found in hot springs. *The Journal of earth sciences, Nagoya University*, *11*(1), 68–100.

Kolditz, O., Bauer, S., Bilke, L., Bottcher, N., Delfs, J., Fischer, T., ... Zehner, B. (2012, 9 1). Opengeosys: an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (thm/c) processes in porous media. *Environmental Earth Sciences*, *67*(2), 589–599. doi: 10.1007/s12665-012-1546-x

Landis, E. N., & Keane, D. T. (2010). X-ray microtomography. *Materials characterization*, *61*(12), 1305–1316.

Marques Erthal, M. (2018). *Travertine shrub structures: origin, diagenetic modifications and petrophysical characteristics - tivoli case (central italy).* Retrieved from `https://lirias.kuleuven.be/retrieve/487062$$DMarcelleMarques_Thesis.pdf`

Pentecost, A. (1990). The formation of travertine shrubs: mammoth hot springs, wyoming. *Geological Magazine*, *127*(2), 159–168.

Pentecost, A. (2005). *Travertine.* Springer Science & Business Media.

Ramanujan, S. (1914). Modular equations and approximations to π. *Quart. J. Math*, *45*, 350–372.

Rezende, M., & Pope, M. (2015). Importance of depositional texture in pore characterization of subsalt microbialite carbonates, offshore brazil. *Geological Society, London, Special Publications*, *418*, SP418–2.

Ronchi, P., & Cruciani, F. (2015). Continental carbonates as a hydrocarbon reservoir, an analog case study from the travertine of saturnia, italy. *AAPG Bulletin*, *99*(4), 711–734.

Salvo, L., Suéry, M., Marmottant, A., Limodin, N., & Bernard, D. (2010). 3d imaging in material science: Application of x-ray tomography. *Comptes Rendus Physique*, *11*(9-10), 641–649.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... others (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, *9*(7), 676.

Schröder, S., Ibekwe, A., Saunders, M., Dixon, R., & Fisher, A. (2016). Algal–microbial carbonates of the namibe basin (albian, angola): implications for microbial carbonate mound development in the south atlantic. *Petroleum Geoscience*, *22*(1), 71–90.

Siddiqui, S., Nasr-El-Din, H. A., & Khamees, A. A. (2006). Wormhole initiation and propagation of emulsified acid in carbonate cores using computerized tomography. *Journal of Petroleum Science and Engineering*, *54*(3-4), 93–111.

Soete, J., Kleipool, L. M., Claes, H., Claes, S., Hamaekers, H., Kele, S., ... Swennen, R. (2015). Acoustic properties in travertines and their relation to porosity and pore types. *Marine and Petroleum Geology*, *59*, 320–335.

Tsukrov, I., & Kachanov, M. (1993). Solids with holes of irregular shapes: effective moduli and anisotropy. *International Journal of Fracture*, *64*(1), R9–R12.

Virgone, A., Broucke, O., Held, A.-E., Lopez, B., Seard, C., Camoin, G., ... others (2013). Continental carbonates reservoirs: the importance of analogues to understand presalt discoveries. In *Iptc 2013: International petroleum technology conference.*

Wright, V. P. (2012). Lacustrine carbonates in rift settings: the interaction of volcanic and microbial processes on carbonate deposition. *Geological Society, London, Special Publications*, *370*, SP370–2.

Wright, V. P., & Barnett, A. J. (2015). An abiotic model for the development of textures in some south atlantic early cretaceous lacustrine carbonates. *Geological Society, London, Special Publications*, *418*, SP418–3.

Yagiz, S. (2009). Predicting uniaxial compressive strength, modulus of elasticity and index properties of rocks using the schmidt hammer. *Bulletin of Engineering Geology and the Environment*, *68*(1), 55–63.

Sample	Length (mm)	Diam. (mm)	ϕ (%)	V_p (cm^3)	V_s (cm^3)
$HFRHES$	71.58	38	7.57	6.20	75.71
$HFRLES$	68.68	38	7.65	5.96	72.01
$MFRHES$	73.39	38	7.11	5.89	76.94
$LFRLES$	81.2	38	5.64	5.2	86.99

Table S1. Petrological characteristics of the samples. The volume of solid was calculated from the difference between the volume of the cylindrical core and the volume of pore: V - V_p.

Compound	Sample 1 (%)	Sample 2 (%)
Calcite	95.15	95.02
Calcite magnesian	4.41	4.02
Dolomite	0.24	0.47
Quartz	0.1	0.1
Montmorillonite	0.1	0.4

Table S2. Chemical composition of a block of travertine (Saturnia Italy). Data from a in-house XRD analysis.

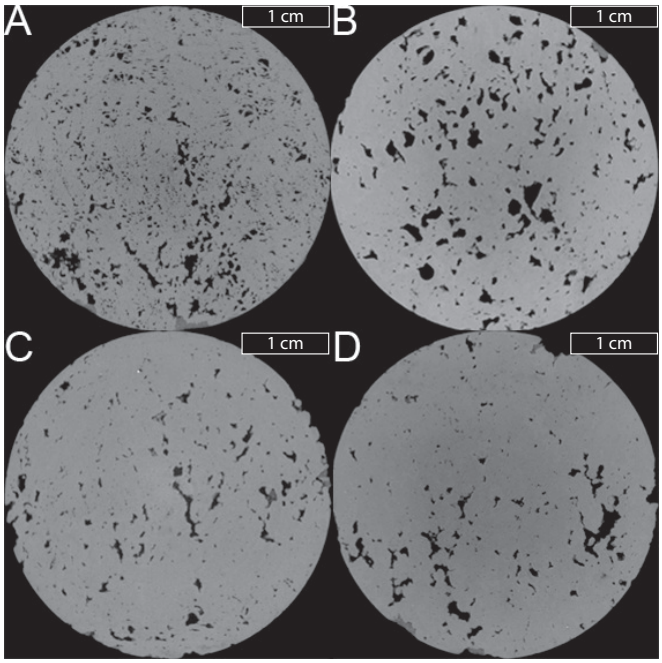

Figure S1. µCT slices of each travertine core sample used in this study. A: $LFRLES$; B: $MFRHES$; C: $HFRHES$; D: $HFRLES$. Image size: each slice is 3.8 cm by 3.8 cm.

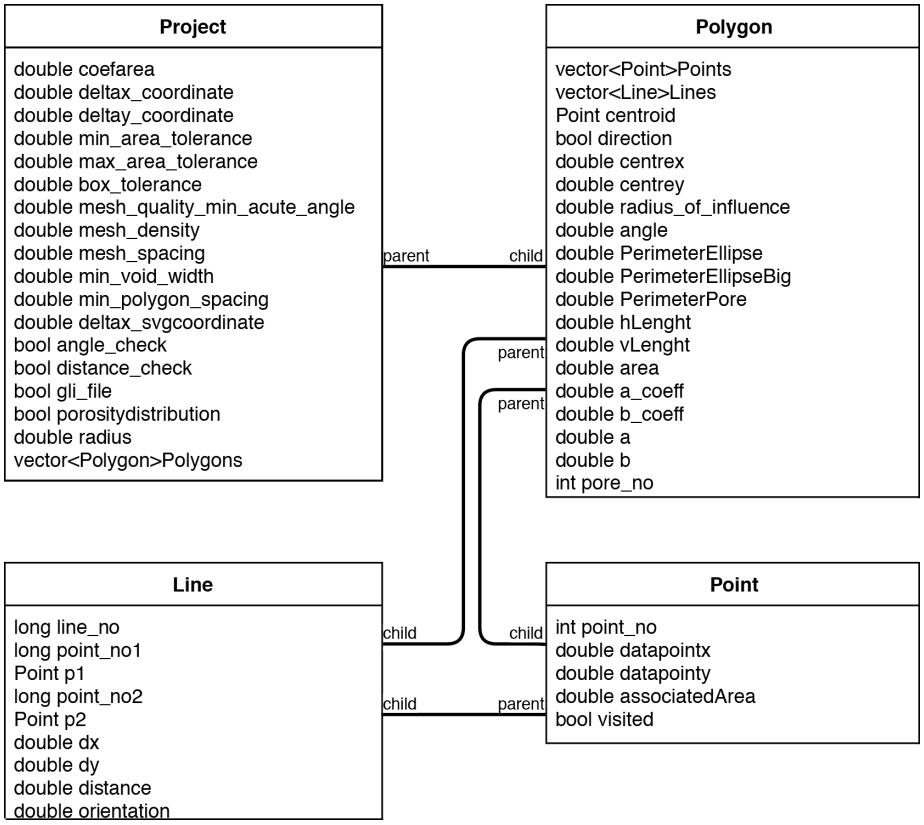

Figure S2. Simplified UML diagram presenting the main classes of the ICCR-Macropore, as well as the relationship between them.

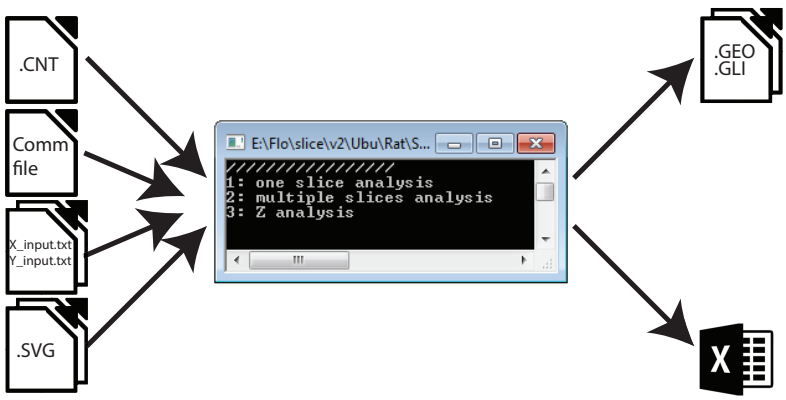

Figure S3. General workflow for the ICCR-Macropore. Three input types are necessary (a CNT file, a comm file, a text file, and the SVG pictures). The tool generates three types of output files (mesh file, geo files, and geometrical data).

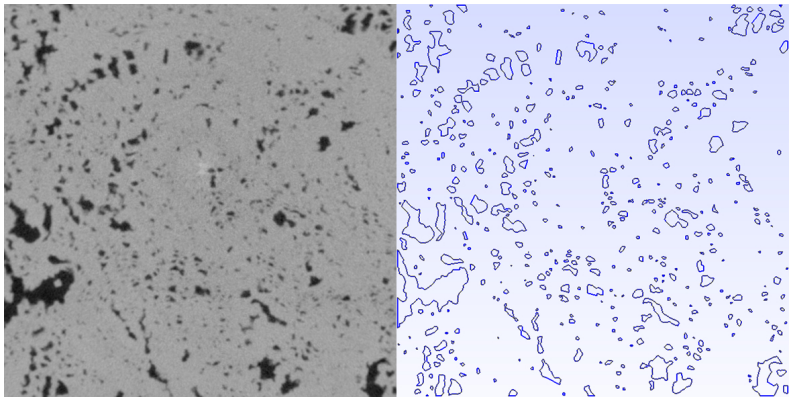

Figure S4. Left) Slice from a travertine sample used in this study ($LFRLES$); Right) Isolated porosity after ICCR-Macropore processing.

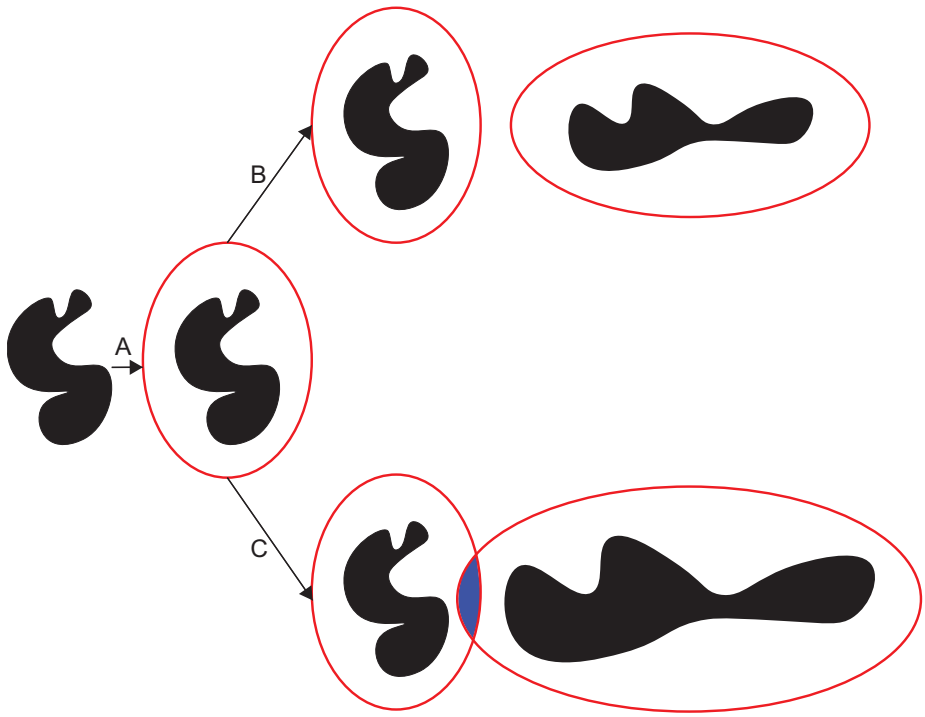

Figure S5. Steps for the detection of a link between two pores (black shapes). Step (A) represents the area of influence applied around a pore through a multiplier of the original pore area. The case (B) shows two non-overlapping pores; Case (C) displays a case of overlapping ellipses.

Figure S8. Input files of the Slice Picker. The three input units are the ICCR-Macropore, the Slice.txt file, and the Comm file.

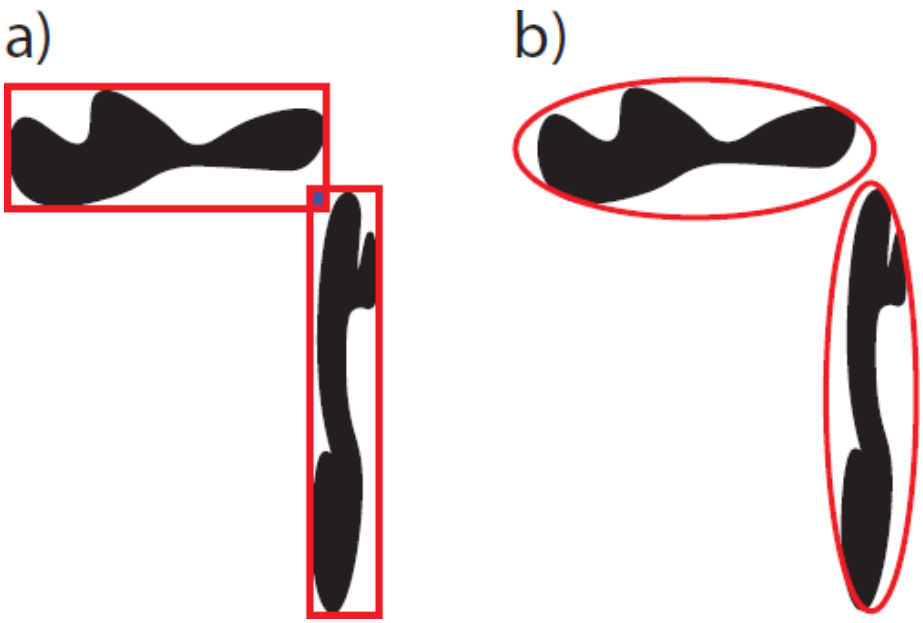

Figure S6. Sketch of overlap detection (blue square) between two pores (black) defined by area equivalent of a) rectangles and b) ellipses.

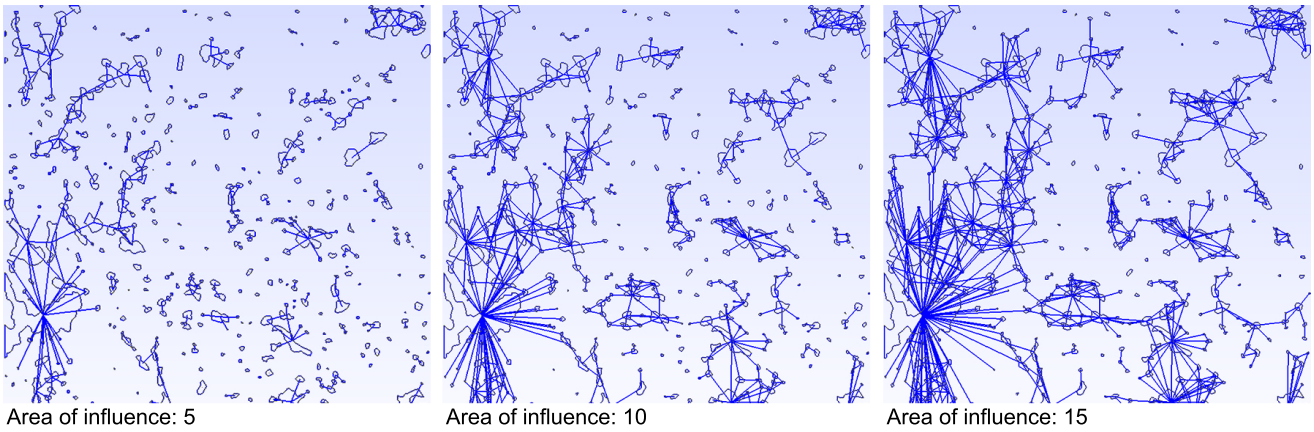

Figure S7. Visual rendering of the 2D connections between pores forming a network of interconnected vugs. Three different *area of influence* have been used: 5, 10, and 15.

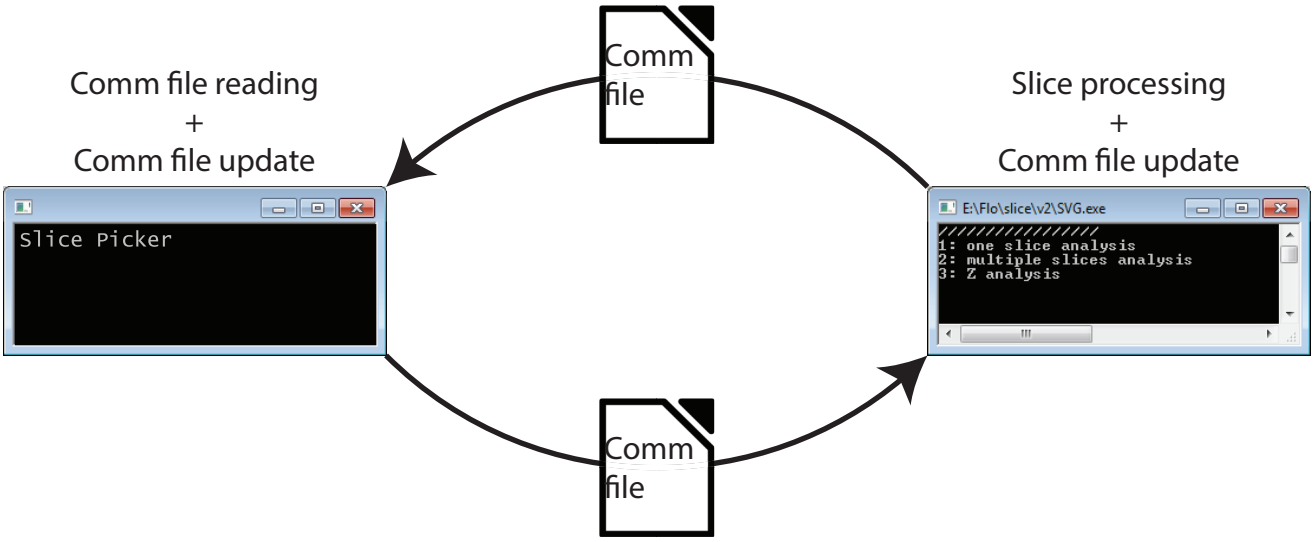

Figure S9. Programming logic and communication between the Slice Picker and the ICCR-Macropore. The Comm file is used as an intermediate file.