GAN learning complex fluvial facies distribution from process-based modelling

Introduction

This paper describes our attempts to model complex fluvial architectures using generative adversarial networks (GAN), by training on realistic geological descriptions produced from sedimentary process models. We show that while previous work in this area has works well on relatively simple models, as the training dataset becomes more complex (with asymmetric channel geometries, variable along channel orientations and more complex facies associations) the GANs struggle to recreate the channel geometries and distribution of facies. In short, more work is required to make GANs successfully recreate complex geological systems.

To date, deep generative models have shown great promise in reservoir modelling. Recently work on deep generative learning show ways to build reservoir facies models while overcoming known limitation of geostatistics - stationarity, linearity, gaussianity, conditioning. Both variational autoencoder (VAE) and generative adversarial networks (GANs) achieve good performance on reproducing object-based model (Laloy et al. 2017; Zhang et al. 2019). Later, several conditioning methods are developed to force GANs to generate realizations matched well data (Zhang et al. 2019; Chan and Elsheikh 2019). Previous papers successfully reproduced pre-defined channel geometry. However, their datasets do not fully represent the geological realism due to the limitation of geostatistics. Especially for fluvial reservoir, process-based model is more suitable to provide training dataset because it better represent the real complexity of fluvial reservoir.

This work tackles two identified issues in fluvial facies modelling: (i) geological realism in modelling complex multi-facies distribution, their relationships, and facies geometry; (ii) extending GAN application to more than 3 facies modelling. Reproduction of geological complexity and facies placement rules by GAN is achieved through learning patterns from process-based models. We use a process-driven algorithm FLUMY™ to create a set of fluvial facies models as the training dataset. Because FLUMY can simulate various sedimentary architectures by tuning parameters. Thus, it allows us to assess uncertainty in depositional variation. After several trials, we find GANs have difficulty in learning this dataset. The generated realizations contain many unrealistic features. In this paper, we list two types of unrealistic features and solutions to counter them.

Methodology

GAN is a deep generative neural network model (Goodfellow et al. 2014). GANs are composed of two sets of neural networks called generator, \( G \) and discriminator, \( D \) (see Figure 1). Random noise is feed into the generator to forward simulate fake data. Discriminator takes real data and fake data as input to calculate the probability of the input is from training dataset. As GANs are very hard to train, we experimentally determine to use PatchGAN (Isola et al. 2017). Following Isola et al. (2017), we use hinge loss (Lim and Ye 2017) as the loss function.

![Figure 1 A schematic diagram of GANs model.](image-url)

FLUMY Fluvial Dataset
Figure 2 shows the model geometries produced by process-based (or proxies to sedimentary processes) models. Additional processes such as lateral accretion and avulsion (along the channel or more proximally outside of the model boundaries) creates much more complex channel geometries. The model also captures the complex relationships between channel fill and point bar deposits as well as lag, splay and levee facies that are hard or impossible to capture in geostatistical algorithms. As such most studies of machine learning based modelling do not use this kind of data as its less readily available and few benchmark case studies use these kinds of complex models.

We have therefore generated a training dataset using FLUMY to cover a range of fluvial systems with greater fidelity. The dataset reflects variability in the three geological processes: channel migration, aggradation, and avulsion. Nine facies in the simulation give a precise description of the modelled fluvial system. Although not all facies can be identified from wells, a comprehensive facies model is helpful to represent depositional processes and determine if it is realistic. For example, the left images of Figure 2 provide a better explanation of the sand distribution and geometry than the right images.

(a) High avulsion rate fluvial model

(b) Low avulsion rate fluvial model

*Figure 2 Examples of a FLUMY 2D realization.*

In this study, we choose a portion of the dataset containing different avulsion rates. Avulsion is a process of abandoning the old channel for a new channel (Heller and Paola 1996). Frequent avulsions cause shorter time for meanders development. Thus, high avulsion rate fluvial model contains more ribbon-like sand-body and has small channel sinuosity. Low avulsion rate fluvial model contains more sheet-like sand-body and chute cut-offs and has big channel sinuosity.

To prevent neural networks overfitted, we adopt data augmentation to extend the volume of training dataset. One sample is extended to eight samples. The data augmentation operations include (1) anti-clockwise rotate 90°, (2) anti-clockwise rotate 180°, (3) clockwise rotate 90°, (4) flip along X-axis, (5) flip along Y-axis, (6) flip along principal diagonal, (7) flip along counter diagonal.

**Experiments: outline the research Questions**

1) Can GAN reproduce the complexity of fluvial deposit captured by process-based model
2) How to reduce the amount of geologically unrealistic features generated in GAN realisations

**The pros and cons of GAN Simulation**
In general, GANs capture the highly curvilinear channel shape and placement of facies along channel centreline (see Figure 4a and Figure 6a). However, we find a couple of unrealistic features in GANs simulated realizations.

The first unrealistic feature is GANs tend to generate facies in the place of other facies whose value is close to it. We call this ‘mislabelling’ issue in this study. Figure 4a shows a type of ‘mislabelling’. GANs generate sand plug in place of the point bar in the red circle.

This is caused by the linear mapping pre-processing in GANs model. Facies is categorical data while GAN is a numerical algorithm. Thus, facies variables are coded to integers before feed into GANs. Linear mapping facies to a continuous range results in an incorrect numerical relationship between facies. To tackle this issue, we use one-hot encoder (Niculescu-Mizil et al. 2009) to do the data pre-processing and replace the last activation of generator with softmax function. The generator output size in spectral dimension is equal to the total number of facies. To get the facies model, we just need to run argmax function to convert the generator output to a single layer of facies model. Results indicate the ‘mislabelling’ issue is largely reduced with the help of one-hot encoder (see Figure 4b).

![Figure 4 Example of generation with 'mislabelling' (a) and without 'mislabelling' (b).](image)

The second unrealistic feature is an incorrect channel-levee relationship (see Figure 6a). The facies relationships in FLUMY are complex because FLUMY discretize the fluvial simulation based on elevation. Geo-bodies may exist across different layers vertically (they have a thickness), thus, there are some channels containing discontinuous levee in 2D. Different channels may deposit in the same elevation, which adds further channel-levee relation complexity to the 2D realization.

To help GAN learn better, we trained our GAN with an additional discriminator, which has same architecture as discriminator but takes regrouped data as input (see Figure 5).

![Figure 5 Input data to discriminator (a) and regrouped data to additional discriminator (b).](image)

During training, two discriminators are used together to discern whether the input is real. The mean value of the calculated losses is used to update the GAN model. Compared to the baseline, the levee placement gets more reasonable with the help of the additional discriminator (see Figure 6b).
Figure 6 Example of generation with incorrect channel-levee relation from baseline (a) and without incorrect channel-levee relation from multi-discriminator GAN (b).

Conclusions

PatchGAN demonstrated the capability of modelling complex fluvial deposits with large number of facies and a high geological realism. Instead of using object-based model to create training dataset, we use FLUMY to generate a multi-facies fluvial dataset which is closer to the real complexity of fluvial reservoir. The result indicates PatchGAN reproduce the curvilinear geometry and placement rules of fluvial facies honouring the limitation of PatchGAN as a form of texture.

We proposed two methods to tackle two geologically unrealistic features in GAN generations, respectively. Embedding one-hot encoder in GAN can largely reduce the ‘mislabelling’ issue. Implementation of multi-discriminators can efficiently help GAN learn better in the aspect of spatial relations among different facies.

Acknowledgements

This project is sponsored by Heriot-Watt University via their James Watt Scholarship Scheme whose support is gratefully acknowledged. We would like to thank MINES ParisTech for their software FLUMY. Free download from http://cg.ensmp.fr/flumy.

References

FLUMY™ Process-based channelized reservoir models Copyright © MINES ParisTech / ARMINES Free download from http://cg.ensmp.fr/flumy