Article information

Article title

GAN River-I: A Process-based Low NTG Meandering Reservoir Model Dataset for Machine Learning Studies

Authors

Chao Sun*, Vasily Demyanov, Daniel Arnold

Affiliations

Institute of Geoenergy Engineering, Heriot-Watt University, Edinburgh, EH14 4AS, Scotland, U.K.
www.geodatascience.hw.ac.uk

Corresponding author's email address and Twitter handle

cs157@hw.ac.uk

Keywords

generative adversarial networks, facies modelling, meandering river, conceptual geological model

Abstract

The GAN River-I data set is designed to provide a stern test for machine learning and geostatistical tools that wish to recreate the complex geometries of realistic facies distributions in subsurface reservoirs. It provides more complex, non-stationary facies distributions than previous open data sets, some of which have modelled channels but do not include the number and complex association of facies types of this data set.

GAN River-I is a dataset of 2D layers of 3D facies models produced from a process-based simulator of a meandering fluvial system. The data set contains 25 simulated 3D cubes, converted into three datasets consisting of 16,000 2D models/images, each representing the increasing complexity of the modelled facies. The number of facies decreases between the three data sets, with nine facies, seven facies and three facies, respectively. The facies reduction is carried out by amalgamating similar facies in terms of their likely permeability to represent flowing units in a subsurface reservoir. The data is therefore provided to allow users to increase the model complexity in a manageable and comparable way between groups using the data.

GAN River-I covers a range of low NTG meandering patterns with varied avulsion rates. Each dataset comprises an ensemble of meandering models representing various plausible patterns and, therefore, can be used as a geologically plausible benchmark for testing generative models' performance. We provide three data file formats, including image, Ndarray and GSLIB, to adapt to different researchers' preferences.

Specifications table

Subject	Computers in Earth Sciences
Specific subject area	Meandering system, facies modelling
Type of data	GSLIB Ndarray Image
How the data were acquired	By computer simulation using FLUMY$^{\text{TM}}$ 5.912 Process-based channelized reservoir models. Copyright © MINES PARIS-PSL / ARMINES. Free download from https://flumy.minesparis.psl.eu
Data format	Raw
Description of data collection	We collect all horizontal slices from all 25 3D simulations from FLUMY$^{\text{TM}}$ to compose this dataset and save them in three formats to make it convenient for researchers with various backgrounds. Besides this original dataset, we provide two additional sets with fewer facies to meet the demands of data complexity in different tasks.
Data source location	· Institution: Heriot-Watt University · City/Town/Region: Edinburgh · Country: United Kingdom
Data accessibility	Repository name: Github Direct URL to data: https://github.com/GeoDataScienceUQ/GANRiverI
Related research article	C. Sun, V. Demyanov, D. Arnold, Geological Realism in Fluvial Facies Modelling with GAN under Variable Depositional Conditions. Computational Geosciences. In Review.

Value of the data

- GAN River-I can work as a benchmark dataset to train and test deep generative models' performance in learning complex geological patterns.
- Researchers working in geomodelling, computer vision and deep learning can benefit from GAN River-I as it introduces a complex model generation task to the research community.

- GAN River-I can also directly work as an ensemble of fluvial facies models to investigate geological uncertainty in geomodelling, history matching and forecasting in various industries, such as oil and gas, carbon capture storage, geothermal and hydrology etc.

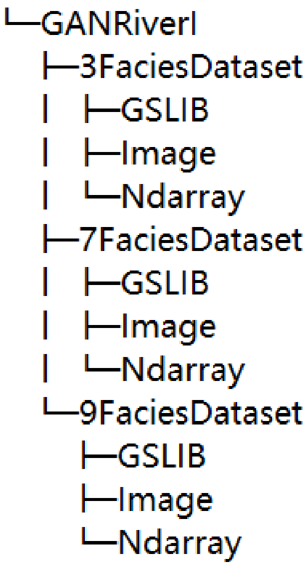
1. Objective

We have created a data set to provide a fiendishly tough challenge for the growing community of researchers developing machine learning and geostatistical methods to model subsurface reservoir geology accurately. To this end, we created the GAN River-I fluvial meandering dataset, which has complex facies geometries and associations. We used a stochastic process-based model, FLUMY^{TM} [1].

Recent research shows that machine learning approaches, like GAN and VAE, can reproduce quite complex fluvial channel patterns [2,3]. However, the datasets used in these papers lack the full complexity of a real fluvial system as they miss critical facies types like point bars and downstream variations in channel fill. Often the training datasets are simulated by object-based models, like TiGenerator [4], which cannot, for instance, capture the complex arrangement of lateral and downstream accretion that occurs in real rivers. Facies models from process models like FLUMY^{TM} capture the realistic facies geometries by mimicking geological processes and the resulting deposits but can’t condition to known data from the subsurface, so they have limited direct applicability to this problem. Therefore, the GAN River-I is timely and well suited for the next phase of GAN research into geological modelling, providing the pre-canned and comparable data for training machine learning-based algorithms. No benchmark dataset of this complexity for geological reservoir modelling is yet available in the research community. The data may also have applications in river characterization or hydrology.

2. Data description

GAN River-I has three training datasets, and each consists of 16,000 2D data/images converted from 25 3D low NTG meandering simulations with five different avulsion rates. Each dataset contains three folders because we store data in three formats, including GSLIB (.dat), Ndarray (.npy) and image (.png) file (see Figure 1). Each 2D data/image is 256X256 and named in the form of ‘Group_Sample_Layer_Facies.format’. For example, ‘2_1_0_Facies.npy’ means this data comes from the base (0th) layer of sample one in group 2 and saves in Ndarray format.



```
└─GANRiverI
   ├─3FaciesDataset
   |  ├─GSLIB
   |  ├─Image
   |  └─Ndarray
   ├─7FaciesDataset
   |  ├─GSLIB
   |  ├─Image
   |  └─Ndarray
   └─9FaciesDataset
      ├─GSLIB
      ├─Image
      └─Ndarray
```

Figure 1. The dataset structure of data in the repository.

The three training datasets have a different number of facies suitable for various needs of benchmarking applications to model meandering systems of different levels of detail. Every facie has a unique code in these training datasets (see Table 1).

Table 1. Relationship between FLUMY^{TM} Facies and Dataset Codes.

Facies	9-Facies Dataset	7-Facies Dataset	3-Facies dataset
Channel Lag	1	0 (Channel Lag)	0 (Point Bar)
Point Bar	2	1 (Point Bar)	0 (Point Bar)
Sand Plug	3	2 (Sand Plug)	1 (Channel)
Crevasse Splay I	4	3 (Crevasse Splay)	2 (Background)
Crevasse Splay II Channel	5	3 (Crevasse Splay)	2 (Background)
Crevasse Splay II	6	3 (Crevasse Splay)	2 (Background)
Levee	7	4 (Levee)	2 (Background)
Overbank	8	5 (Overbank)	2 (Background)
Mud Plug	9	6 (Mud Plug)	1 (Channel)

The 9-Facies dataset uses the raw facies code exported from FLUMY, which preserves all details produced by modelled processes. Lateral accretion packages composed of point bar (sand) and channel lag (coarse residual) deposit at the inner bank with channel migrating. Abandoned channels fill with sand (sand plug) and mud (mud plug) after avulsion or meander cut-off. Every overbank flooding places deposits on the flooding plain from proximal to distal, transiting from silt (levee) to shale (overbank). Erosive deposits (Crevasse splay I) deposit aside the meander when levee breaches occur. Sediments may evolve from erosive to non-erosive (Crevasse splay II), and FLUMY may add channels (Crevasse splay Channels) to the non-erosive deposits. This training dataset suits testing algorithms' power in learning processes and multi-facies distribution [5] because the data preserves all facies transitions, regardless of their frequency.

The 7-Facies dataset keeps facies with contrasting rock properties, while the facies model is simplified by reducing the crevasse splay-associated facies. We know crevasse splay deposits can form a significant part of fluvial systems and preserve complex facies distribution [6]. In contrast, crevasse splay deposits only take up about 0.4% in the simulated FLUMY^{TM} models. The 7-Facies dataset involves precise facies distribution within the geo-bodies that impact the subsurface flow response and can describe the complex sand distribution, such as the connected point-bars scenario [7]. This dataset can be the benchmark for testing advanced machine learning algorithms in learning multi-facies distribution.

The 3-Facies dataset focuses on the meandering channel shape and the placement of lateral accretion packages. This dataset adapts to evaluating ML models in learning geometries [8] and can work as an ensemble of facies models to describe isolated sand bodies, like the isolated point bars scenario [7].

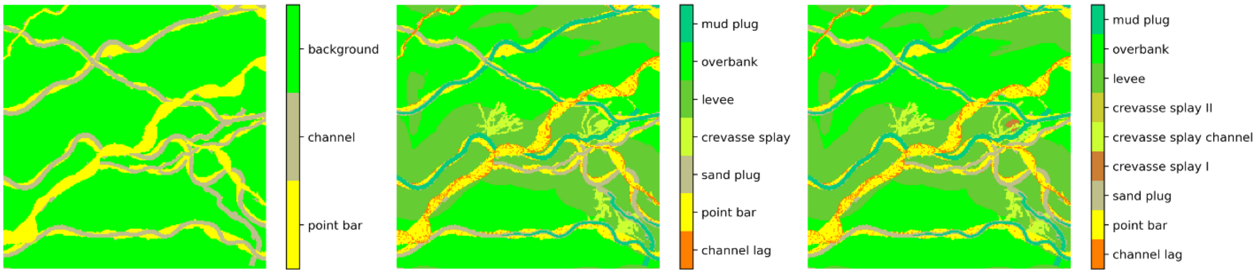


Figure 2. An example of FLUMY^{TM} data in the three training datasets.

3. Experimental design, materials and methods

3.1 FLUMY Simulation

FLUMY^{TM} is a stochastic and process-based model which creates facies models by mimicking channel evolution in temporal and deposits associated facies along the channel centrelines. Figure 3 illustrates the modelled geological processes in FLUMY^{TM}, including aggradation, channel migration, avulsion, meander cut-off and levee breaches. The aggradation process deposits sedimentary on the modelled domain when overbank flooding occurs. Channel migration refers to the channel lateral migrating on the modelled field, commonly estimated by the bend theory, which is the linearized hydraulic equation [9]. Meander cut-off and avulsion are two types of channel abandonment. The meander cut-off forms a new channel by connecting two sides of a meander loop and abandoning the old channel [10]. Avulsion, including regional and local avulsion, is a process of the meandering river abandoning the old channel for a new one [11]. We use FLUMY^{TM} v.5.912 as the simulator to generate an ensemble of meandering models reflecting various sedimentary settings.

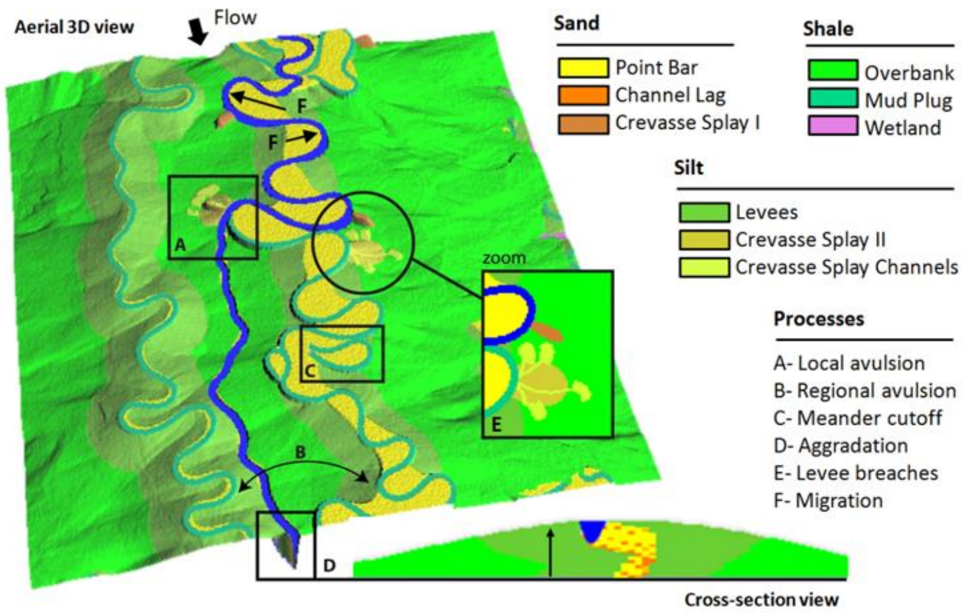


Figure 3. Processes of the meandering system in FLUMY^{TM}. From FLUMY^{TM} v.5.912 user guide.

We create five low NTG meandering simulations with the same FLUMY^{TM} parameters for each avulsion rate to describe the aleatory uncertainty of the natural system by inherent stochasticity. The stochastic process allows FLUMY^{TM} to create different models using the same settings while only changing the seed value that we use: ‘165426111’, ‘165426222’, ‘165426333’, ‘165426444’ and ‘165426555’ in this study. We fix the FLUMY^{TM} parameters to keep a low NTG (about 20%) and moderate sedimentation rate (0.4 cm/year), except for the avulsion rate. We assign different avulsion rates to build a range of geologically realistic meandering patterns that cover a possible range of natural variability (see Figure 4).

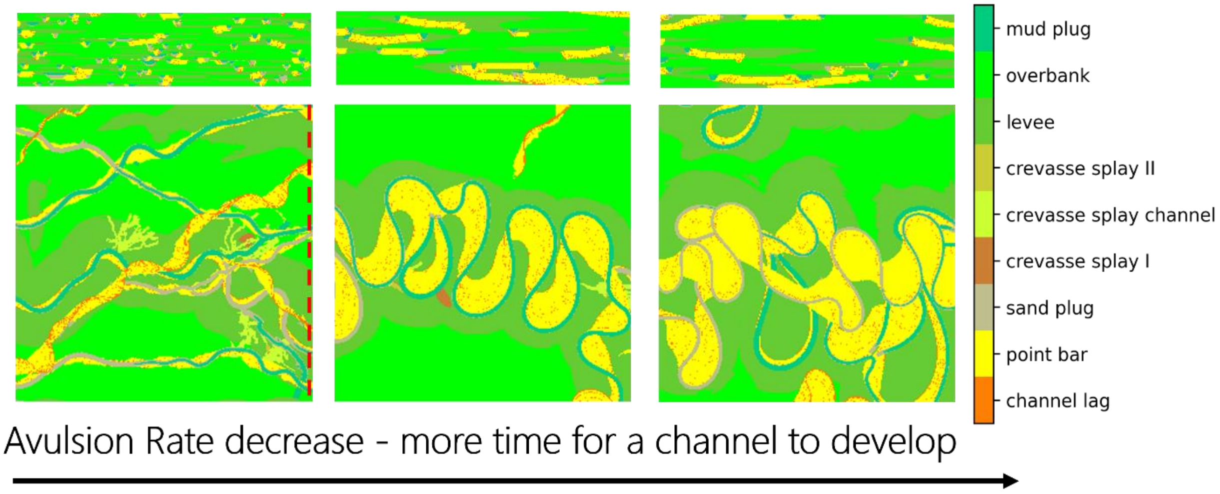


Figure 4. FLUMY^{TM} simulations with different avulsion rates. Vertical exaggerated.

FLUMY^{TM} parameters selection relies on the non-expert user calculator (Nexus) [12]. The input of Nexus includes channel maximum depth, sandbodies extension index (SEI) and net to gross

(NTG). We set the maximum depth to 5 meters in Nexus to determine the channel geometry parameters in FLUMY^{TM}. After a trial and error, we set Nexus' NTG to 10%, which results in the NTG of the meandering models reaching around 0.2 in our study. We use the SEI in Nexus to determine the values of different avulsion rates and refer to the modern river records to support the selected avulsion period range. We choose the minimum (20) and maximum values (160) of the SEI and three typical values (50, 80, 110) indicating ribbon, standard and sheet scenarios, respectively. Table 2 summarises the five SEI values and corresponding FLUMY^{TM} avulsion parameters. According to the recording of modern rivers, the avulsion period ranges from 28 to 1400 years [13]. This range works as extra support for checking the avulsion period range instead of a constraint on the avulsion period because Slingerland suggested that a bigger range may exist, and no theoretical limits exist.

Table 2. Avulsion parameters used in FLUMY^{TM}.

Group Number	1	2	3	4	5
Sandbodies Extension Index	20	50	80	110	160
Regional Avulsion Period, T_R	190	480	800	1000	1500
Local Avulsion Period, T_L	105	270	435	600	885
Total Avulsion Period	68	173	282	375	557

* Avulsion includes regional avulsion and local avulsion; $\frac{1}{Avulsion\ Period} = \frac{1}{T_R} + \frac{1}{T_L}$

3.2 Datasets Production

We convert every FLUMY^{TM} 3D simulation cube into 2D data/images by slicing every 0.1 meters vertically. We export the 3D sequence between 0 to 64 meters to avoid undefined values in FLUMY^{TM} and sample the discretized cubes every 0.1 meters, which produces 640 2D data/images. The same process operates on all 25 simulations. Thus, the dataset consists of 16,000 data/images, named the 9-Facies dataset in GAN River-I.

We group some facies to make the 7-Facies dataset and 3-Facies dataset. For the 7-Facies dataset, we name three facies, 'crevasse splay I', 'crevasse splay channel' and 'crevasse splay II', as 'crevasse splay', and keep the rest unchanged. Then, we use one unique code to represent each facies. We take similar operations to make the 3-Facies dataset. 'Channel lag' and 'point bar' compose the new 'point bar' in 3-Facies dataset because they both belong to lateral accretion deposits in the meandering system. 'Sand plug' and 'mud plug' comprise the new 'channel' of the 3-Facies dataset as they deposit in the abandoned channel. Table 1 summarises the code of every facies in each dataset.

Ethics statements

We claim this dataset doesn't involve human subjects, animal experiments or data collected from social media platforms.

CRediT author statement

Chao Sun: Conceptualization, Methodology, Software, Validation, Writing-Original draft, Visualization. **Vasily Demyanov:** Conceptualization, Supervision, Writing-Reviewing and Editing. **Daniel Arnold:** Conceptualization, Supervision, Writing-Reviewing and Editing.

Acknowledgments

The work contained in this paper contains work conducted during a PhD study undertaken as part of the Natural Environment Research Council (NERC) Centre for Doctoral Training (CDT) in Oil & Gas [grant number NEM00578X/1]. 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™, Process-based channelized reservoir models. Copyright © MINES PARIS-PSL / ARMINES. Free download from http://cg.ensmp.fr/flumy.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.L. Grimaud, F. Ors, M. Lemay, I. Cojan, J. Rivoirard, Preservation and Completeness of Fluvial Meandering Deposits Influenced by Channel Motions and Overbank Sedimentation, J Geophys Res Earth Surf. 127 (2022). https://doi.org/10.1029/2021JF006435.

[2] E. Laloy, R. Hérault, J. Lee, D. Jacques, N. Linde, Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network, Adv Water Resour. 110 (2017) 387–405. https://doi.org/10.1016/j.advwatres.2017.09.029.

[3] S. Chan, A.H. Elsheikh, Parametric generation of conditional geological realizations using generative neural networks, Comput Geosci. 23 (2019) 925–952. https://doi.org/10.1007/s10596-019-09850-7.

[4] A. Maharaja, TiGenerator: Object-based training image generator, Comput Geosci. 34 (2008) 1753–1761. https://doi.org/10.1016/j.cageo.2007.08.012.

[5] C. Sun, V. Demyanov, D. Arnold, GAN learning complex fluvial facies distribution from process-based modelling, 82nd EAGE Annual Conference & Exhibition. 2021 (2021) 1–5.

[6] C.E. Burns, N.P. Mountney, D.M. Hodgson, L. Colombera, Anatomy and dimensions of fluvial crevasse-splay deposits: Examples from the Cretaceous Castlegate Sandstone and Neslen Formation, Utah, U.S.A., Sediment Geol. 351 (2017) 21–35. https://doi.org/10.1016/j.sedgeo.2017.02.003.

[7] M.E. Donselaar, I. Overeem, Connectivity of fluvial point-bar deposit: An example from the Miocene Huesca fluvial fan, Ebro Basin, Spain, American Association of Petroleum Geologists Bulletin. 92 (2008) 1109–1129. https://doi.org/10.1306/04180807079.

[8] C. Sun, V. Demyanov, D. Arnold, Comparison of popular Generative Adversarial Network flavours for fluvial reservoir modelling, 82nd EAGE Annual Conference & Exhibition. 2021 (2021) 1–5.

[9] S. Ikeda, G. Parker, K. Sawai, Bend theory of river meanders. Part 1. Linear development, J Fluid Mech. 112 (1981) 363–377. https://doi.org/10.1017/S0022112081000451.

[10] C.E. Russell, Prediction of sedimentary architecture and lithological heterogeneity in fluvial point-bar deposits, University of Leeds, 2017.

[11] P.L. Heller, C. Paola, Downstream changes in alluvial architecture; an exploration of controls on channel-stacking patterns, Journal of Sedimentary Research. 66 (1996) 297–306.

[12] A. Bubnova, On the conditioning of process-based channelized meandering reservoir models on well data, Université Paris Sciences & Lettres, 2018.

[13] R. Slingerland, N.D. Smith, River avulsions and their deposits, Annu Rev Earth Planet Sci. 32 (2004) 257–285. https://doi.org/10.1146/annurev.earth.32.101802.120201.