Responses to Reviewer #1

We highly appreciate your time in reviewing our manuscript. The relevant comments have improved the quality of the manuscript. We now revised the manuscript accordingly. The detail responses are listed below. The modifications in the article are marked in the annotated manuscript.

Q1. This manuscript proposes a deep CNN with joint autoencoder and adversarial structures to predict the probability of subsurface palaeovalleys (derived from airborne electromagnetic data) using 2D land surface tomography. It has been claimed that the trained model "produces a square error < 0.10 across 93% of the validation areas". The prediction error contradicts the conclusion of a reliable model in reconstructing 3D palaeovalley patterns. This is consistent with the results in Figure 3. If we compare Figures 3d and 3f, it is quite clear that many structures that are present in the real 3D image are missing from the simulated image and indeed these two images are not similar. On the other hand, Figure 3c and 3e are very similar (Training set). This simple visual comparison reveals that the trained model is very overfitted and contradicts the claim of similar performance in training and validation areas (Abstract: "The trained neural network has a maximum square error < 0.10 and produces a square error < 0.10 across 93% of the validation areas"). I highly recommend the authors to provide more evidence on the performance of the proposed model in validation areas. A 3D map showing the spatial distribution of errors (for both validation and training areas) would be useful.

Reply: Partially agree and changes made.

(1) **Consistence.** We now carefully checked through the calculations and the resulting values, and make sure that the expression of "produces a square error < 0.10 across 93% of the validation areas" is consistent with the values used to draw Fig. 3. In the training domain, it was calculated that the 0.991 quantile of errors is equal to 0.1. The expression "The trained neural network has a maximum square error < 0.10" is now reformulated as "The trained neural network has a square error < 0.10 across 99% of training domain" in both the abstract and the context (e.g. Line 13, 202 and 213), to better express the findings.

(2) Overfitting. We now draw the 3D distribution of errors in the validation domain and also a plan view of errors averaged over ten layers (now Fig. 4). The error distribution is compared to the modern-day valley pattern suggested by the MrVBF in both validation and training domains, because the paleovalley geometry inherits the pattern of modern-day valley (comparing Fig. 3a and 3c, 3b and 3d). It is illustrated that the distribution of large errors in the validation domain is unrelated to the modern-day valley geometry in the training domain, but some concentrate on the boundaries of surface valley in the validation domain. The former confirmed that no overfitting problem occurs. The latter is induced by the convolution processes itself. This is now expressed in Lines 215-225.

A further comparison between the neural network without and with fully connected layers is now added in Lines 228-238. As

a result, using relatively small filter and removing the fully connected layer to under-parameter the neural network model helped reducing the overfitting risk.

Q2. The proposed model is used for subsurface structure mapping. Sub3DNet might be a better name for the model.

Re: Agree.

We now changed the model name as Sub3DNet.

Q3. It is good to discuss some of the limitations of the deep CNN models. For instance, too many structures are available. (e.g., number of convolutional and pooling layers) and it is not clear which structure is the best for the study presented in this manuscript.

Re: Agree.

We now added a Discussion section in **Lines 228-238**, to (1) compare the proposed CNN with the traditional structure, (2) to clarify the limitation of CNN model. We also put the details of the CNN structure optimization (e.g. depth, width, filter sizes of neural networks) in the support materials.

Responses to Reviewer #2

We highly appreciate your time in reading through this manuscript many times, and give us many constructive comments on the background presentation. We have revised the introduction and study area sections accordingly, to better articulate the research problems we aim to solve, among others. The detailed responses are listed below and the modifications are marked in the annotated manuscript.

Q1. My review of this paper is not favorable mainly because, despite repeated readings of it, I am unable to identify the specific research problem that the authors are seeking to solve, and because the case study used to demonstrate their method appears to be trivial in the context of subsurface characterization. Although it is possible my reactions stem from major misunderstanding of the descriptions of the objectives, methods and results, I have spent decades of my career mapping and modeling paleochannels, including application of AEM and other geophysical methods, yet I am unable to reconcile the separation between what the authors are writing and what I would consider to be understandable or obvious contributions.

Reply: We agree that the research problem was not well articulated and have made changes to support our claim the work is novel with many practical applications to better mapping of shallow subsurface features and their geometries.

We rewrote the Introduction to better articulate the novel contribution in the method development. The major changes included are:

(1) in the first paragraph (Lines 22-31), we now describe that big data sets on geology and geomorphology are globally available either as land surface observations (typically remote sensing and topographical data and their derivatives), or regionally available in a limited number of highly-developed mining and oil fields (e.g., downhole, surface and airborne geophysical interpretations). In Australia, the former are readily available at low cost, while the latter are often non-existing and expensive in remote desert areas where groundwater for town supply relies on access to shallow aquifers (Munday et al., 2020a). In their study, Munday et al. (2020) interpreted 17,000 line km of airborne electromagnetic (AEM) data covering an area of about 30,000 km², a fraction of the 422,000 km² Great Victoria Desert in central Australia. With an AEM line spacing of 2 km, with smaller infill areas where line spacing was reduced to 250 and 500 m to provide greater detail of the subsurface electrical conductivity, accurate mapping of palaeovaleys was achieved (Munday et al., 2020b). Application of such high-resolution data to much larger areas like the Victorian Desert would be cost prohibitive. Our goal is therefore to develop an efficient and generic tool to express the relationship between an easy-to-obtain dataset and a more costly dataset for the specific purpose of detecting palaeovalley features that would facilitate the discovery of new groundwater resources in arid and semi-arid regions. In other words, we seek to develop a novel method that uses AEM only for model development on a small training

area while the application (i.e. detection of palaeovalleys across large areas) uses readily available landsurface information that otherwise (i.e., without AEM coupling through a training procedure) would have had little value for palaeovalley detection.

(2) in the second paragraph (Lines 38-53), we describe the limitation of the existing methods. For example, the traditional geostatistical methods are skillful in interpolation but not in extrapolation. MPS is powerful in delineating complex subsurface structures, but its effect depends on the availability of the training data. These methods are developed and employed based on the single-support dataset, that is, the data types employed to define spatial relationship is presumed to be the same as those data types employed to predict the subsurface geo-body. They are often inefficient in capturing essential features and patterns from large and multiple-support datasets, or can do so only at a high computational cost. This more or less limits their application. The neural network model developed in this study, on the other hand, provides a framework with a flexible input data type (e.g. 2D land surface observations and others) and complex output datasets (e.g. 3D paleovalley pattern). It is capable to define nonlinear relationships among multiple-support datasets, and employ this relationship for prediction with merely easy-to-obtain input data.

Q2. My trouble with the objectives and problem definition can be best illustrated by first considering the geologic system the authors seek to better map in 3D. 'Paleochannels' can take on a number of different meanings depending on the geologic setting, but from what I can decipher from the introduction, methods and Fig. 2, by 'paleochannels' the authors are referring to incised valley fill deposits like those depicted in Fig. 2b, where the channels are bounded not by adjacent fluvial facies, but by granite. Setting aside for the moment that this looks more like a basin and range style of geologic structure than a paleochannel, based on the vague descriptions in the paper, I can only construe that the flattest portions of the DEM shown in Fig. 2c represent the Quaternary alluvial bottomlands representative of the top of the apparent paleochannels (i.e., top of sed facies in Fig. 2b). If that is true, the reader's reaction is inevitably: "Why is this even considered a challenging problem? From the topography it is already obvious where these so-called channel deposits locate."

Summarizing the case study, it appears that the DEM already nicely identifies locations of the paleochannels, which apparently have been further characterized using AEM, presumably to better identify their depths or depth to bedrock perhaps. This raises the question of what is the problem the authors are attempting to address? If the problem is to better identify x-y locations of the so-called paleochannels, that would appear moot because the DEM already shows them, which also raises the question of why you need DL. If the problem is to better identify paleochannel or incised valley-fill depths, that has apparently already been done with AEM; and furthermore, if the purpose is to use the DL algorithm to map the paleochannels depths so that AEM would not be needed, that also does not appear to make sense because the authors have not established a relationship between the DEM flatness metric and paleochannel depths.

It is possible that if the authors can be more specific about the geology of these 'paleochannel' features that they are trying to map and about what specifically they are trying to accomplish through the application of their ML methods, the above problems would be cleared up. As written, however, the manuscript lacks sufficient definition of the problem, description of their objectives, and description of how their research satisfies those objectives.

Reply: We agree that the problem should be better defined, with greater clarity of objectives and how those were achieved. The following changes have been made in response to the comments.

(1) The Introduction provides background geological information on the palaeovalley system of interest, and why ML is adopted to improve mapping of their location and their 2D/3D geometry (**now Lines 74-94**).

The case study area is a pre-Pliocene palaeovalley system in central Australia that has been postulated to contain significant groundwater resources (Dodds and Sampson, 2000). However, their geometry and extent remain largely hidden from view by a valley fill of Pliocene to Pleistocene sediments and overlying Quaternary sand dunes of the Great Victoria Desert (Lewis et al., 2010). Although the thicker valley fill sequences seem to be coincident with contemporary lows or valleys in the more subdued relief of the plains, the definition of the palaeovalley systems remains relatively poor (Munday et al., 2020a). This has been attributed to sandplain sediments forming a relatively continuous cover over much of the Musgrave Province down to 30-40 m depth; below this depth the definition of the palaeovalley systems becomes significantly clearer with a well-defined network of major alluvial channels and tributary systems. As is evident from an analysis of AEM images, the palaeovalley system has a highly irregular geometry with spatially varying depths to basement, and with heterogeneous infill resulting in lithologically controlled palaeovalley aquifers.

Our goal is therefore to develop an efficient and generic machine learning tool to express the relationship between an easy-toobtain dataset and a more costly dataset for the specific purpose of detecting palaeovalley features that would facilitate the discovery of new groundwater resources in arid and semi-arid regions. In other words, we seek to develop a novel method that uses AEM only for model development on a small training area while the application (i.e. detection of palaeovalleys across large areas) uses readily available landsurface information that otherwise (i.e., without AEM coupling through a training procedure) would have had little value for palaeovalley detection. Moreover, in addition to detection of palaeovalley location, the method should also derive the 3D palaeovalley geometry. Such methodology is premised on the existence of a mechanistic connection between landsurface features and subsurface distribution of palaeovaleys. To what degree such correlation exists (and can be cast in a predictive framework) between palaeovalley geometry and landsurface features derived from digital elevation data in the palaeovalley system of the Musgrave Province will be tested using a deep convolutional neural network methodology.

(2) The paleovalley pattern in this demonstration case is comparable to that of modern valley pattern. Thus, the MrVBF (a 2D

land surface observations) is related to the 3D paleovalley structure; but it cannot directly suggest the depth of paleovalley and width of the paleovalley at different depths. AEM-interpreted EC values is a direct index of 3D paleovalley structure (including both depth and width), but it is not available everywhere. We employed our method to define a relationship between MrVBF and AEM-interpreted EC in the data-rich area, and employed in those area where the AEM is not available to predict the 3D paleovalley pattern based merely on the MrVBF (**now Lines 167-176**).

(3) For the model verification, both the training and validations are conducted in those regions with AEM-interpreted EC. The weights in the neural network model is determined based on the data in the training area. The AEM data in the validation areas is just used to test the ability of the trained model in predicting 3D paleovalley structure, but do not participate in determining

the neural network model (now Lines 176-178).

 Gallant, J. C. and Dowling, T. I.: A multiresolution index of valley bottom flatness for mapping depositional areas, Water resources research, 39, 2003
 Ley-Cooper, A. and Munday, T.: Groundwater Assessment and Aquifer Characterization in the Musgrave Province, South Australia: Interpretation of SPECTREM Airborne Electromagnetic Data, Goyder Institute for Water Research Technical Report Series, 2013.

[3] Soerensen, C. C., Munday, T. J., Ibrahimi, T., Cahill, K., and Gilfedder, M.: Musgrave Province, South Australia: processing and inversion of airborne electromagnetic (AEM) data: Preliminary results. 1839-2725, Goyder Institute for Water Research Technical Report Series, 2016

Specific comments in the annotated PDF files

Line 22. Delete 'dramatically'

Reply: Change made.

Line 22-25. "data poor" contradicts "rich/big", and others

Reply: This sentence is now rephrased in Lines 22-24.

Line 29-30. This is most certainly not true of those method, although one might need to use them more expertly (e.g. through zoning of the model region) when non-stationarities are present; **Line 32.** "is still lacking" to "would be beneficial"; **Line 33.** "fill this gap". You have not identified as a gap, but rather a potential way of improving upon other methods. Rewrite to better describe the "gap" and what your method potentially does.

Reply: This part is rewritten to present the limitations of existing methods, and the major problem we wanted to solve with our developed neural network (Lines 38-50 and Line 74-94)

Line 39: Add 'e.g.'

Reply: Change made.

Line 119-121. This would appear to represent existing fluvial drainage characteristics. For this to be useful for 'training' the DL model there would have to be a mechanistic connection between these surface features and the subsurface distribution of paleochannels. There is a big problem with this approach: modern geomorphic surface characteristics seldom represent or

Reply: Agree.

While the occurrence of palaeovalleys is correlated to the modern-day valley pattern (Jiang et al., 2019), their exact location and geometry in the case study area cannot simply be inferred from modern geometric surface features such as the 2D Multipleresolution Valley Bottom Flatness (MrVBF) index (calculated from the digital elevation model) (Gallant and Dowling, 2003). The correlation is complicated by the presence of relatively continuous sandplain sediments that cover the palaeovalleys. On the other hand, the vertical structure of a palaeovalley can be interpreted from an airborne electromagnetic (AEM) survey (Ley-Cooper and Munday, 2013; Soerensen et al., 2016). The MrVBF index exists across the entire Australia continent, while AEM data of sufficient spatial granularity only exists in a limited number of prospective mining fields. Our neural network model establishes a relationship between the MrVBF index (high values are indicative of locations with a high probability of deposition of alluvial sediments) and the AEM-interpreted 3D palaeovalley structure. This relationship is then used to predict the 3D palaeovalley structure in those areas with only MrVBF data but without the AEM dataset (**now Lines 167-176**).

An area 80 km west of the training area is first used to validate the trained neural network in generating 3D PAI. The statistics of squared errors between the simulated 3D PAI and real PAI are calculated at all 200×200×10 voxels. As shown in Fig. 3, the squared error in the training dataset is below 0.1 for 99% of the training domain and with a mean value of about 0.03, and the squared error of the predicted 3D PAI is well below <0.1 for 93% of the validation domain, with a mean squared error of about 0.04. The patterns of the generated palaeovalley in both horizontal and vertical directions align with those inferred from the AEM-derived PAI. This indicates that the deep-learning neural network structure developed in this work is capable of incorporating the relationships between the MrVBF and the buried palaeovalley patterns, and allowing for reliable predictions beyond the training area (Lines 200-207).

Figure 2. No clear. Do the valley bottoms in Fig. 2c correspond each to the type of channel and facies depicted in 2b? If yes, does that mean these are all incised into granite? In that case, the predictive geologic problem would appear to be trivial.

Reply: The valley bottom flatness data from Fig.2c represents the input data for the neural network model, noting that the modern-day valley pattern is correlated with the occurrence of palaeovalleys, however their exact location and geometry in the case study area cannot simply be inferred from the 2D Multiple-resolution Valley Bottom Flatness (MrVBF) index alone. The 2D conceptual model of a palaeovalley (Fig. 2b) is a very simplified representation of the heterogeneous structure of the palaeovalleys in the Musgrave Province. The valley bottoms of Fig 2c have a high likelihood to contain palaeovalley features, incised in a more or less unweathered (resistive) basement rock. This does not make the geologic problem trivial: however, it does provide the basis for delineating the palaeovalley base using a cut-off resistivity boundary. Without such resistivity contrast between basement rock and conductive infill the AEM method would have difficulty in delineating any palaeovalley

accurately (now Lines 155-166).

Line 121-123. Is the point here to use AEM results as a ground truth and demonstrate that you could do as good, or almost as good, without the AEM and just using your DL approach based on surficial information? Not clear.

Reply: Yes. The MrVBF index exists across the entire Australia continent, while AEM data of sufficient spatial granularity only exists in a limited number of prospective mining fields. Our neural network model establishes a relationship between the MrVBF index (high values are indicative of locations with a high probability of deposition of alluvial sediments) and the AEM-interpreted 3D palaeovalley structure. This relationship is then used to predict the 3D palaeovalley structure in those areas with only MrVBF data but without the AEM dataset. For the method verification, both the training and prediction are conducted in the area where AEM data is available. Note that the weights in the neural network are determined based on the training area. The AEM data in the other areas are only used to test the predictive capability of the trained neural network (now Lines 176-178).

Sub3DNet1.0: A deep learning model for regional-scale 3D subsurface structure mapping

Zhenjiao Jiang ^{1,2}, Dirk Mallants ², Lei Gao ², Gregoire Mariethoz ³, Luk Peeters ⁴

1 Key Laboratory of Groundwater Resources and Environment, Ministry of Education, College of Environment and Resources, Jilin University, Changchun, 130021, China

2 CSIRO Land & Water, Locked Bag 2, Glen Osmond, SA 5064, Australia
3 University of Lausanne, Faculty of Geosciences and Environment, Institute of Earth Surface Dynamics, Lausanne, Switzerland
4 CSIRO Mineral Resources, Locked Bag 2, Glen Osmond, SA 5064, Australia

Corresponding author: Zhenjiao Jiang (jiangzhenjiao@hotmail.com)

Abstract. This study introduces an efficient deep learning approach based on convolutional neural networks with joint autoencoder and adversarial structures for 3D subsurface mapping from <u>2D</u> surface observations. The method was applied to delineate palaeovalleys in an Australian desert landscape. The neural network was trained on a 6,400 km² domain by using a land surface tomography as 2D input and an airborne electromagnetic (AEM)-derived probability map of palaeovalley presence as 3D output. The trained neural network has a squared error < 0.10 across 99% of the training domain and produces a squared error < 0.10 across 93% of the validation domain, demonstrating that it is reliable in reconstructing 3D palaeovalley

15 patterns beyond the training area. Due to its generic structure, the neural network structure designed in this study and the training algorithm have broad application potential to construct 3D geological features (ore bodies, aquifer) from 2D land surface observations.

1 Introduction

5

- Imaging the Earth's subsurface is crucial for the exploration and management of mineral, energy and groundwater resources, their reliability depending on the availability and quality of geological data. Although the amount and quality of geological data obtained from borehole logs, geophysical prospecting and remote sensing has increased dramatically over the past decades, their spatial distribution is highly uneven. Big data sets on geology and geomorphology are globally available either as land surface observations (typically remote sensing and topographical data and their derivatives), or only regionally available in a limited number of highly-developed mining and oil fields (e.g., downhole, surface and airborne geophysical interpretations).
- In Australia, the former are readily available at low cost, while the latter are often non-existing and expensive in remote desert areas where a key challenge is to secure groundwater for town supply, often available only from shallow aquifers (Munday, 2020a, b). In their study, Munday et al. (2020a, b) interpreted 17,000 line km of airborne electromagnetic (AEM) data covering an area of about 30,000 km² within the much larger Great Victoria Desert in central Australia (422,000 km²). With a AEM line spacing of 2 km, with smaller infill areas where line spacing was reduced to 250 and 500 m to provide greater detail of the subsurface electrical conductivity, accurate mapping of palaeovaley aquifers was achieved (Munday, 2020a). Application of such high-resolution data to much larger areas like the entire Victorian Desert would be cost prohibitive. Most data exist as

land surface observations in a limited number of highly-developed areas such as mining and oil fields. Predictive models are much needed for extracting hidden information from local scale rich/big datasets for predicting 3D regional scale subsurface structures in data poor areas.

- 35 Commonly used methods for modelling complex geological structures include geostatistical approaches such as sequential Gaussian or indicator simulation (Lee et al., 2007), transition probability simulation (Felletti et al., 2006; Weissmann and Fogg, 1999), or multiple-point simulation (MPS) methods (Hu and Chugunova, 2008; Mariethoz and Caers, 2014; Strebelle, 2002). Most geostatistical approaches are suitable for "interpolation", which performs well in predicting 3D subsurface structures within the data-rich region (Kitanidis, 1997). However, their ability to "extrapolate" a 3D subsurface structure is limited.
- 40 <u>Alternatively, MPS is an advanced method to quantify the complex spatial structure based on training images. It transfers the quantified structures to the data-scarce region for stochastic predictions; however, a realistic 3D training image is difficult to obtain. Overall, most existing subsurface structure modelling approaches are developed to analyse a single-support dataset, that is, the data types employed to define spatial structures are presumed to be identical to as those employed for predictive purposes (de Marsily et al., 2005). Better defining and utilizing the relationship among multiple-support datasets allows</u>
- 45 regional-scale subsurface structure imaging based on easy-to-obtain dataset. However, these existing methods are often inefficient in capturing essential features and patterns from large and multiple-support datasets, or can do so only at a high computational cost (Laloy et al., 2018). The analysis of multiple support datasets, e.g. downhole geophysical logs and 3D seismic with lithofacies, is still based on subjective expert knowledge. However, they often present drawbacks such as not being able to represent elongated and connected geological structures such as palaeovalleys, being inefficient in capturing
- 50 essential features and patterns from very large training datasets, or presenting a high computational cost. A fast and reliable tool for high resolution 3D subsurface imaging based on multiple support big dataset is still lacking. A fast and reliable tool capable of deriving a robust relationship among multiple-support big datasets is much-needed for high-resolution imaging of 3D subsurface structures.

Deep learning approaches specialised in big data mining have the potential to fill this gap (Gu et al., 2017; Hinton and Salakhutdinov, 2006; Marcais and de Dreuzy, 2017). Applications in the geosciences include earthquake detection based on seismic monitoring (Mousavi and Beroza, 2019; Perol et al., 2018), or disaster recognition from remote sensing data (Amit et al., 2016; Längkvist et al., 2016), among others. A recent breakthrough in deep learning is the 2D to 3D image processing (Niu et al., 2018; Sinha et al., 2017; Wu et al., 2016; Yi et al., 2017). Such approaches point out a novel way to rapidly and automatically identify covered 3D subsurface structures directly from readily-available 2D surface observations (e.g. digital elevation models, land cover maps, signals captured by airborne geophysical surveys). <u>A neural network framework that reliably transforms 2D input data into 3D output data is required that has the flexibility to fuse multiple types of geology and geophysical input data (e.g. 1D downhole logs, 2D surface and cross-sectional profiles, and 3D seismic interpretations) for complex 3D geological subsurface structure imaging.
</u>

To this end, we designed a deep convolutional neural network (CNN) with joint autoencoder (Kingma and Welling, 2013) and

- 65 adversarial structures (Goodfellow et al., 2014). The autoencoder component features large input and output images connected by a small latent space. This structure is advantageous for the fusion of complex input data and 3D image reconstruction. Its training involves direct back-propagation according to a voxel-wise independent heuristic criterion, and thus often needs a large training dataset to constrain the model and avoid overfitting (Laloy et al., 2018). The generative adversarial learning tries to generate multiple images inheriting the probability structure of one real image, which relaxes the need for very large training
- 70 dataset. For method verification, the proposed approach approach is applied to an Australian desert landscape to generate regional-scale 3D palaeovalley patterns from 2D digital terrain information. -in an Australian desert landscape_We aim to demonstrate that the interplay between autoencoder and adversarial components provides a generic tool to exploit more effectively geophysical, land surface and other data to generate realistic regional-scale 3D geological structures.

The case study area is a pre-Pliocene palaeovalley system in central Australia that has been postulated to contain significant

- 75 groundwater resources (Dodds and Sampson, 2000). However, their geometry and extent remain largely hidden from view by a valley fill of Pliocene to Pleistocene sediments and overlying Quaternary sand dunes of the Great Victoria Desert (Lewis et al., 2010). Although the thicker valley fill sequences seem to be coincident with contemporary lows or valleys in the more subdued relief of the plains, the definition of the palaeovalley systems remains relatively poor (Munday, 2020a). This has been attributed to sandplain sediments forming a relatively continuous cover over much of the Musgrave Province down to 30-40
- 80 m depth; below this depth the definition of the palaeovalley systems becomes significantly clearer with a well-defined network of major alluvial channels and tributary systems. As is evident from an analysis of AEM images, the palaeovalley system has a highly irregular geometry with spatially varying depths to basement, and with heterogeneous infill resulting in lithologically controlled palaeovalley aquifers (Munday, 2020a).

Our goal is therefore to develop an efficient and generic machine learning tool to express the relationship between an easy-toobtain dataset and a more costly dataset for the specific purpose of detecting palaeovalley features that would facilitate the discovery of new groundwater resources in arid and semi-arid regions. In other words, we seek to develop a novel method that uses AEM only for model development on a small training area while the application (i.e. detection of palaeovalleys across large areas) uses readily available landsurface information that otherwise (i.e., without AEM coupling through a training procedure) would have had little value for palaeovalley detection. Moreover, in addition to detection of palaeovalley location,

90 the method should also derive the 3D palaeovalley geometry. Such methodology is premised on the existence of a mechanistic connection between landsurface features and subsurface distribution of palaeovaleys. To what degree such correlation exists (and can be cast in a predictive framework) between palaeovalley geometry and landsurface features derived from digital elevation data in the palaeovalley system of the Musgrave Province will be tested using a deep convolutional neural network methodology.

95 2 Method

The adversarial neural network for 3D subsurface imaging involves three steps: (1) patch extraction and representation, (2) nonlinear mapping and reconstruction, and (3) statistical expression of the generated image (Fig. 1). The first step is referred to as 'encoder' (Fig. 1a), which is employed to fuse the information contained in the 2D land surface observation images (input data) into a low-dimension layer by successive convolutions (Hinton and Salakhutdinov, 2006):

$$h(\mathbf{x}) = f(\mathbf{W} \cdot \mathbf{x} + \boldsymbol{b}),\tag{1}$$

100 where f is a nonlinear function referred to as "activation function", **W** is a matrix of weights and **b** is a bias vector in the encoder.

The encoder can be designed to contain multiple layers, where the number of layers is defined as 'depth'. Each layer can contain multiple images, with the number of images defined as 'width'. The images in one layer are convoluted to generate the elements in the image of the next layer by weight filters, and the elements in the low-dimension layer of the encoder (the

105 output) are called 'code'. The process of convolution is illustrated in Fig. 1b, which shows that with a filter size of 2×2 (for a 2D image convolution for example), one element in the output layer is related to 4 elements in the input layer. Thus, the spatial correlation scale addressed by the convolutional neural network can be controlled by the filter size in both vertical and horizontal directions.

The weight and bias in the encoder are trained to ensure that the code follows a standard normal distribution, by minimizing 110 the Kullback–Leibler divergence (L1), defined as (Kullback and Leibler, 1951):

$$L1 = \frac{1}{2N} \sum_{i=1}^{N} (\mu^2 + \sigma^2 - \log\sigma^2 - 1)_i, \tag{2}$$

where *N* is number of codes in the final output layer of the encoder, μ and σ are the mean and standard deviation of the codes, respectively.

(**a**)



Figure 1. (a) Adversarial convolution neural network composed of (1) encoder for input images patch extraction and representation, (2) decoder for nonlinear mapping and 3D image reconstruction, (3) discriminator for distinguishing the generated 3D image and real image after statistical expression; and features of the (b) convolution and (c) deconvolution processes with the colour representing the origin of the deconvoluted values. For mapping palaeovalley patterns in an Australian desert landscape, the input data uses the 2D MrVBF (an index calculated from globally available digital elevation model); the output is a 3D probability map of palaeovalley presence. For convenience of 3D convolution, the 2D input image (800×800×1) is simply repeated in 10 layers to form a 3D input dataset (800×800×10). Following a structure optimization by trial-and-error, the encoder is designed to contain 4 layers, with a width of 64, 32, 32 and 1 in each layer, respectively; the decoder contains 6 layers, with a width of 1, 16, 32, 32, 64, and 128, respectively; the discriminator contains 4 layers with a width of 128, 64, 32, 1, respectively.

In the second step, the codes are converted into a 3D output image by deconvolution (referred to as 'decoder', Fig. 1a), which

125 is a process involving a zero-padding before the convolution (Fig. 1c). The combination of decoder and encoder forms a 'generator', linking input and output images. The generated 3D image is referred to as 'simulated image'.

To ensure that the simulated image is comparable to a real image, a voxel-wise independent heuristic criterion is minimized. The mean squared error (L2) between simulated and real images at all voxels is used as criterion to update the weight and bias in the decoder, which is expressed as:

$$L2 = \frac{1}{M} \|G(\mathbf{z}) - \mathbf{Y}\|^2,$$
(3)

130 where *M* is the number of voxels in the output 3D image, **Y** is the real image, **z** is the code generated from the encoder, and $G(\cdot)$ represents the convolutional calculations in the decoder (in the same form as Eq. 1).

However, if only a limited number of real 3D images are available to train the network, the use of a voxel-wise independent criterion often leads to an overfitting problem. Goodfellow (2014) proposed a generative adversarial network structure, which adds a 'discriminator' to convert simulated and real images to a vector, respectively, by an identical convolution process (Fig. 1a). Adversarial criteria are proposed, typically expressed by binary cross entropy functions as:

$$L3 = -\frac{1}{v} \log[D(G(\mathbf{z}))], \tag{4}$$

and

135

140

$$L4 = -\frac{1}{v}\log[D(\mathbf{Y})] - \frac{1}{v}\log[1 - D(G(\mathbf{z}))],$$
(5)

where *V* is the size of the output vector via the discriminator, and $D(\cdot)$ represents the calculations (Eq. 1) in the discriminator. The weights in the discriminator are trained to minimize *L*4, which attempts to distinguish the vectors generated from the real and simulated 3D images. The weights in the generator are trained to minimize *L*3, which attempts to fool the discriminator to be unable to identify the vector generated from the simulated 3D image. In such a way, the generator can produce images aligned with the real image in terms of probability structure (Goodfellow et al., 2014).

Finally, while the loss function L4 is minimized to optimize the weights in the discriminator, a comprehensive loss function combining L1, L2 and L3 is employed to optimize the weights in the generator, which is expressed as (Wu et al., 2016):

$$L_g = \mathbf{a} \cdot L\mathbf{1} + \mathbf{b} \cdot L\mathbf{2} + \mathbf{c} \cdot L\mathbf{3},\tag{6}$$

where a, b, c are the coefficients on each loss function. This loss function makes it convenient to vary the neural network structure between semi-supervised learning with additional adversarial neural network by defining coefficient c as non-zero value, and supervised learning with merely autoencoder neural network with c as zero.

The hyperparameters (including the width, depth, filter size and the coefficients in generator loss functions, etc.) defining the neural network structure, are determined by trial-and-error tests (Supplementary materials). Weight and bias in generator and discriminator are trained to minimize L_g and L4 using the stochastic gradient descent algorithm, referred to as adaptive moment

150 estimation (ADAM) (Kingma and Ba, 2014). We implemented the above convolution neural network using the Tensorflow Python library (Abadi et al., 2016). Once the neural network is trained, the 'generator' in the network (Fig. 1a) is used independently to generate 3D subsurface structures from the 2D land surface observations.

3 Results

We use a CSIRO dataset to test the effectiveness of our deep-learning approach in predicting 3D palaeovalley patterns in the

- 155 Anangu Pitjantjatjara Yankunytjatjara (APY) lands of South Australia (Fig. 2a and 2b). As demonstrated in Fig. 2b, the palaeovalley networks in the APY lands are remnants of the Early Cenozoic inset valleys with Tertiary coarse-to-fine grained sands infill and a thin and variable Quaternary eolian sediments cover (Magee, 2009). The 2D conceptual model of a palaeovalley (Fig. 2b) is a very simplified representation of the heterogeneous structure of the palaeovalleys in the Musgrave Province (Munday, 2020a). The valley bottom flatness data from Fig.2c represents the input data for the neural network model,
- 160 noting that the modern-day valley pattern is correlated with the occurrence of palaeovalleys, however their exact location and geometry in the case study area cannot simply be inferred from the 2D Multiple-resolution Valley Bottom Flatness (MrVBF) index alone (Gallant and Dowling, 2003). The valley bottoms of Fig. 2c have a high likelihood to contain palaeovalley features, incised in a more or less unweathered (resistive) basement rock. This does not make the geologic problem trivial: however, it does provide the basis for delineating the palaeovalley base using a cut-off resistivity boundary. Without such resistivity
- 165 <u>contrast between basement rock and conductive infill the AEM method would have difficulty in delineating any palaeovalley</u> <u>accurately.</u>

While the occurrence of palaeovalleys is correlated to the modern-day valley pattern (Jiang et al., 2019), their exact location and geometry in the case study area cannot simply be inferred from modern geometric surface features such as the 2D Multiple-resolution Valley Bottom Flatness (MrVBF) index (calculated from the digital elevation model). The correlation is complicated

- 170 by the presence of relatively continuous sandplain sediments that cover the palaeovalleys. On the other hand, the vertical structure of a palaeovalley can be interpreted from an airborne electromagnetic (AEM) survey (Ley-Cooper and Munday, 2013; Soerensen et al., 2016). The MrVBF index exists across the entire Australia continent, while AEM data of sufficient spatial granularity only exists in a limited number of prospective mining fields. Our neural network model establishes a relationship between the MrVBF index (high values are indicative of locations with a high probability of deposition of alluvial sediments).
- 175 and the AEM-interpreted 3D palaeovalley structure. This relationship is then used to predict the 3D palaeovalley structure in those areas with only MrVBF data but without the AEM dataset. For the method verification, both the training and prediction are conducted in the area where AEM data is available. Note that the weights in the neural network are determined based on the training area. The AEM data in the other areas are only used to test the predictive capability of the trained neural network.

The dataset includes a 100-m, 2D MrVBF_index across the entire model domain (Gallant and Dowling, 2003), and a 3D electrical conductivity dataset (400-m horizontal and 10-m vertical resolution) interpreted from AEM survey in the APY Lands (Ley-Cooper and Munday, 2013; Soerensen et al., 2016). Previous hydrogeological characterization indicates that high bulk electrical conductivity values (EC) are a proxy for palaeovalley presence (Jiang et al., 2019; Munday et al., 2013; Taylor et al., 2015). Thus, a palaeovalley aquifer index (PAI) is defined as:

$$PAI = \frac{\log_{10}(EC) - \log_{10}(EC)_{min}}{\log_{10}(EC)_{max} - \log_{10}(EC)_{min}}$$

(7)

where *max* and *min* represent the maximum and minimum logarithm of EC values over the entire dataset, respectively. PAI
ranges from 0.0 to 1.0 and is calculated in the first 100 m depth at the AEM-surveyed area, which is considered as a ground-truth 3D probability map of palaeovalley occurrences with a spatial resolution of 400 m×400 m×10 m.



Figure 2. Datasets for delineating 3D palaeovalley in the Anangu Pitjantjatjara Yankunytjatjara (APY) Lands of South Australia: (a) location of the largest deserts in Australia and (b) general conceptual model of palaeovalley sedimentary facies revealed by over 90% borehole logs, (c) multiple resolution valley bottom flatness index, and (d) electrical conductivity (at depths of 30 to 40 m with a horizontal resolution of 400 m) inferred by airborne electromagnetic surveys in the APY Lands, forming an indicator of palaeovalley occurrence.

- A neural network simulator is established and trained to relate the AEM-derived PAI (output image) with 2D MrVBF data (input image). The training dataset covers part of the APY Lands (6,400 km²) (hereafter referred to as 'training area'). Both loss functions for discriminator and generator were monitored when training the model to verify the network being trained sufficiently (Supplementary materials). Training of the network under 10,000 iterations on a high-performance computer (Tesla P-100-SXM2-M-16GB) required 100 to 150 minutes of computation time. Once trained, generating of 3D image from 2D MrVBF required less than five seconds on a desktop computer.
- 200 An area 80 km west of the training area is first used to validate the trained neural network in generating 3D PAI. The statistics of squared errors between the simulated 3D PAI and real PAI are calculated at all 200×200×10 voxels. As shown in Fig. 3, the

squared error in the training dataset is below 0.1 for 99% of the training domain and with a mean value of about 0.03, and the squared error of the predicted 3D PAI is well below <0.1 for 93% of the validation domain, with a mean squared error of about 0.04. The patterns of the generated palaeovalley in both horizontal and vertical directions align with those inferred from the

205 AEM-derived PAI. This indicates that the deep-learning neural network structure developed in this work is capable of incorporating the relationships between the MrVBF and the buried palaeovalley patterns, and allowing for reliable predictions beyond the training area.



- 210 Figure 3. Multiple resolution valley bottom flatness (MrVBF) (a and b) converted to the 3D palaeovalley aquifer index (PAI) in the training area (c) and validation area (80 km west to the train area) (d) by the neural network simulator, compared with AEM-derived PAI (ground truth data) (e) and (f) generated from airborne electromagnetic surveys. The trained neural network with the squared error < 0.10 across 99% of the training zone (a total of 200 ×200 ×10 voxels), results in a PAI error < 0.10 across 93% of this validation zone having errors exceeding 0.20.
- 215 Furthermore, in both the training and validation domains, the palaeovalley geometry in each layer is generally comparable to the surface valley geometry indicated by the MrVBF index at land surface (compare Fig. 3a and 3c, Fig. 3b and 3d), with varying width at different depths. A comparison of the PAI error with the surface valley pattern in the validation domain (Fig. 4) shows that the spatial distribution of the largest prediction errors is rather random, with some concentration at the boundaries

of modern-day valleys. This is related to the convolution processes itself (see further). The error distribution in the validation domain is independent from the modern-day valley geometry in the training area, suggesting that no overfitting problem occurs.



Figure 4. (a) 3D distribution of squared errors between simulated PAI and real PAI in the validation domain, and (b) plan view of the mean of squared errors from ten layers, overlain by the surface (modern-day) valley (validation and training domains). The large errors, to some extent, focus on the edge of modern-day valley in the validation domain, but are unrelated to the modern-day valley in the training domain, suggesting that the overfitting does not occur.

4. Discussion

4.1 Neural network with and without fully connected layer

The traditional convolution neural network is often ended by a fully connected layer in the encoder (e.g. Wu et al., 2016), to adequately fuse the input information for prediction. In this study, a 3D image with size of $25 \times 25 \times 5$ is employed for the final

230 output layer of the encoder (Fig. 1), without a fully connected layer. For comparison, a fully connected layer with a vector of 3125 (25×25×5) elements is employed as well. As shown in Fig. 5, both models can be trained to generate the paleovalley in

the training domain successfully (Fig. 5a to Fig. 5c and 5b, respectively). However, with a fully connected layer, the trained model fails to generate paleovalleys in the validation domain. Under an alternative MrVBF as input (Fig. 5d), the predicted paleovalley has a geometry very similar to that of the training domain (compare Fig. 5e and 5d). This suggests an apparent overfitting, caused by the fully connected operation fusing the input MrVBF globally.

235



Figure 5. Input MrVBF in (a) training area and (d) validation area, and (b) and (e) the generated PAI at depth of 30-40 m with fully connected operation in the encoder, and (c) and (f) without fully connected operation.

- Alternatively, the model without a fully connected layer can predict well the paleovalley following the MrVBF pattern. Without the fully connected layer, the convolution processes with 3D filter addressed the local relationship of MrVBF and PAI. The correlation scale is determined by the size of the filter; the lager the filter, the larger the correlation scale addressed. The filter size can be determined by a trial-and-error test, according to the misfit between the predicted geological variable and the ground truth data in both training and validation domains. In this study, a filter with a size of 4×4×2 is employed for the encoder and discriminator, while a filter with size of 5×5×2 is employed for the decoder (details in the supplementary materials).
- 245 Training and validation suggest that using relatively small filter and removing the fully connected layer to under-parameter the neural network model helped reducing the overfitting risk. Although the performance of the neural network model with this given structure is acceptable, relatively large errors still occur at the boundaries of the paleovalley where the MrVBF

values vary sharply. This is because local convolution potentially broadens the influence of large MrVBFs; adaptive optimization of filter size in each convolution layer potentially solves this problem.

250 4.2 Adversarial neural network versus autoencoder neural network

Furthermore, another 19 validation areas west to the training domain (Fig. 2d) are used to monitor the decay in the accuracy of predicted palaeovalley patterns. This is done using two different models: semi-supervised learning with additional adversarial neural network and supervised learning with only autoencoder neural network (controlled by coefficient c in Eq.



255

Figure 6. Squared errors between the true 3D palaeovalley aquifer index (PAI) directly calculated from AEM-derived electrical conductivity, and PAI predicted by (a) autoencoder neural network using supervised learning and (b) adversarial neural network using semi-supervised learning in the areas west to the training area, with separation distance varying from 0 to 80 km; and (c) an overfitting test with a random 2D MrVBF as input to predict PAI at depth of 30-40 m following adversarial learning and only autoencoder.

260 aut

As shown in Fig. 6, an extremely small error (<0.01) can be achieved when constructing palaeovalleys in the training area by supervised learning using only the autoencoder neural network. The mean error resulting from the semi-supervised learning

with additional adversarial network is higher, i.e. about 0.03. Within the 19 validation areas, the mean squared errors in predicting palaeovalley patterns by both neural networks are well below 0.04.

- 265 While the autoencoder learning generally performs better than the adversarial learning in terms of mean squared errors, its prediction errors (especially the 95% quantile) increase much faster with the separation distance between validation and training areas. This indicates that using the autoencoder only can potentially lead to very large errors (or poor predictions) in case of discrepancies between training and prediction areas. We hypothesize that these errors are due to model overfitting in the case of using only the autoencoder learning. To confirm this, we conduct an overfitting test based on a synthetic ground
- 270 being a random MrVBF input following a uniform distribution (i.e. non-informative) ranging from 0 to 1, which should result in a uniform PAI distribution. As shown in Fig. 6c, a uniform PAI can be generated by adversarial learning, while the predicted PAI by using only the autoencoder learning results in structured patterns. This means that the weights trained by purely supervised learning inherit too much information hidden in the training dataset, which is inflexible in predicting 3D palaeovalley patterns with strong variations from the input image. On the other hand, adversarial learning is much more robust
- 275 to discrepancies and the accuracy decays only slightly in predicting 3D structures in areas further away from the training area, which is a highly desired property in real world applications.

4 Conclusions

This study developed an efficient and reliable adversarial convolutional neural network simulator for generating 3D subsurface structures directly from 2D land-surface data. The proposed generic structure of the convolutional neural network was
composed of an 'encoder' to fuse 2D input data into low-dimension codes following a normal distribution, a 'decoder' to nonlinearly map the low-dimension codes into 3D subsurface images, and a 'discriminator' to statistically express the generated and real subsurface image into a vector for adversarial semi-supervised learning based on a single training image.

The neural network was successfully tested in mapping the 3D palaeovalley systems in the northeast Great Victoria Desert, Australia. Training and validation involved using the multiple resolution valley bottom flatness (MrVBF) (input) and 3D
palaeovalley aquifer index (PAI) (output) on an area of 80×80 km². The neural network trained to a maximum error <0.1 can predict 3D PAI with errors <0.1 at over 90% of the validation zones.

The outstanding performance of the deep-learning neural network for 3D subsurface structure imaging has applications as a generic novel tool for making better use of existing multiple-support geophysical, land surface, and remote sensing data for better management of limited resources such as groundwater.

290 Codes/Data Availability

The original data of MrVBF and AEM were provided by John Gallant and Tim Munday, respectively, and are available freely from CSIRO Data Access Portal https://doi.org/10.4225/08/5701C885AB4FE and https://doi.org/10.25919/5d0868d48591e. The codes for neural network developed in Tensorflow are now provided in https://doi.org/10.7910/DVN/DDEIUV.

Acknowledgement

295 Funding support for this study was provided by the National Key R&D Program of China (2018YFB1501803), and the International Postdoctoral Exchange Fellowship Program (2017) from China Postdoctoral Council in combination with CSIRO funding through the Land and Water Business Unit and the Future Science Platform Deep Earth Imaging. We thank Guillaume Rongier and Hoel Seille for the internal review and constructive suggestions on this article.

References

- 300 Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., and Isard, M.: Tensorflow: a system for large-scale machine learning, 2016, 265-283.
 - Amit, S. N. K. B., Shiraishi, S., Inoshita, T., and Aoki, Y.: Analysis of satellite images for disaster detection, 2016, 5189-5192.
 de Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., and Violette, S.: Dealing with spatial heterogeneity, Hydrogeology Journal, 13, 161-183, 2005.
- 305 Dodds, S. and Sampson, L.: The Sustainability of Water Resources in the Anangu Pitjantjatjara Lands, South Australia., Department for Water Resources, Adelaide, 2000.
 Felletti, F., Bersezio, R., and Giudici, M.: Geostatistical simulation and numerical upscaling, to model ground-water flow in a sandy-gravel, braided river, aquifer analogue, Journal of Sedimentary Research, 76, 1215-1229, 2006.
 Gallant, J. C. and Dowling, T. I.: A multiresolution index of valley bottom flatness for mapping depositional areas, Water resources research, 39, 2003.
 - Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative adversarial nets, 2014, 2672-2680.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., and Cai, J.: Recent advances in convolutional neural networks, Pattern Recognition, 2017.

Hinton, G. E. and Salakhutdinov, R. R. J. s.: Reducing the dimensionality of data with neural networks, 313, 504-507, 2006.
 Hu, L. and Chugunova, T.: Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review,
 Water Resources Research, 44, 2008.

Jiang, Z., Mallants, D., Peeters, L., Gao, L., Soerensen, C., and Mariethoz, G.: High-resolution paleovalley classification from airborne electromagnetic imaging and deep neural network training using digital elevation model data, Hydrology and Earth

- System Sciences, 23, 2561-2580, 2019.
 Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.
 Kingma, D. P. and Welling, M.: Auto-encoding variational bayes, International Conference on Learning Representations, 2013.
 Kitanidis, P. K.: Introduction to Geostatistics: Applications in Hydrogeology, Cambridge University Press, 1997.
 Kullback, S. and Leibler, R. A.: On information and sufficiency, The annals of mathematical statistics, 22, 79-86, 1951.
- Laloy, E., Hérault, R., Jacques, D., and Linde, N.: Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network, Water Resources Research, 54, 381-406, 2018.
 Längkvist, M., Kiselev, A., Alirezaie, M., and Loutfi, A.: Classification and segmentation of satellite orthoimagery using convolutional neural networks, Remote Sensing, 8, 329, 2016.
 Lee, S.-Y., Carle, S. F., and Fogg, G. E.: Geologic heterogeneity and a comparison of two geostatistical models: Sequential
- 330 Gaussian and transition probability-based geostatistical simulation, Advances in water resources, 30, 1914-1932, 2007. Lewis, S., English, P., Wischusen, J., Woodgate, M., Gow, L., Hanna, A., and Kilgour, P.: The palaeovalley groundwater project: Operational update on demonstration study sites July 2009 to April 2010, Geoscience Australia 302 pp., 2010. Ley-Cooper, A. and Munday, T.: Groundwater Assessment and Aquifer Characterization in the Musgrave Province, South Australia: Interpretation of SPECTREM Airborne Electromagnetic Data, Goyder Institute for Water Research Technical
- **335** Report Series, 2013. Magee L W · Palaec

Magee, J. W.: Palaeovalley groundwater resources in arid and semi-arid Australia: A literature review, Geoscience Australia, 2009.

Marcais, J. and de Dreuzy, J. R.: Prospective Interest of Deep Learning for Hydrological Inference, Groundwater, 55, 688-692, 2017.

340 Mariethoz, G. and Caers, J.: Multiple-point geostatistics: stochastic modeling with training images, John Wiley & Sons, 2014. Mousavi, S. M. and Beroza, G. C.: A Machine-Learning Approach for Earthquake Magnitude Estimation, Geophysical Research Letters, 2019.

Munday, T., Abdat, T., Ley-Cooper, Y., and Gilfedder, M.: Facilitating Long-term Outback Water Solutions (G-FLOWS Stage-1: Hydrogeological Framework, Goyder Institute for Water Research Technical Report Series, 2013.

345 Munday, T., Gilfedder, M., Costar, A., Blaikie, T., Cahill, K., Cui, T., Davis, A., Deng, Z., Flinchum, B., Gao, L., Gogoll, M., Gordon, G., Ibrahimi, T., Inverarity, K., Irvine, J., Janardhanan, S., Jiang, Z., Keppel, M., Krapf, C., Lane, T., Love, A., Macnae, J., Mallants, D., Mariethoz, G., Martinez, J., Pagendam, D., Peeters, L., Pickett, T., Raiber, M., Ren, X., Robinson, N., Siade, A., Smolanko, N., Soerensen, C., Stoian, L., Taylor, A., Visser, G., Wallis, I., and Xie, Y.: Facilitating Long-term Outback Water Solutions (G-FLOWS Stage 3): Final Summary Report, Goyder Institute for Water Research, 2020a. 350 Munday, T., Taylor, A., Raiber, M., Soerensen, C., Peeters, L., Krapf, C., Cui, T., Cahill, K., Flinchum, B., Smolanko, N., Martinez, J., Ibrahimi, T. and Gilfedder, M: Integrated regional hydrogeophysical conceptualisation of the Musgrave Province, South Australia, Goyder Institute for Water Research Technical Report 2020b. Niu, C., Li, J., and Xu, K.: Im2Struct: Recovering 3D Shape Structure from a Single RGB Image, structure, 4096, 80, 2018.

Perol, T., Gharbi, M., and Denolle, M.: Convolutional neural network for earthquake detection and location, Science Advances, 4, e1700578, 2018.

Sinha, A., Unmesh, A., Huang, Q., and Ramani, K.: SurfNet: Generating 3D shape surfaces using deep residual networks, 2017.

Soerensen, C. C., Munday, T. J., Ibrahimi, T., Cahill, K., and Gilfedder, M.: Musgrave Province, South Australia: processing and inversion of airborne electromagnetic (AEM) data: Preliminary results. 1839-2725, Goyder Institute for Water Research

360 Technical Report Series, 2016.

Taylor, A., Pichler, M., Olifent, V., Thompson, J., Bestland, E., Davies, P., Lamontagne, S., Suckow, A., Robinson, N., and Love, A.: Groundwater Flow Systems of North-eastern Eyre Peninsula (G-FLOWS Stage-2): Hydrogeology, geophysics and

environmental tracers, Goyder Institute for Water Research Technical Report Series, 2015.
 Weissmann, G. S. and Fogg, G. E.: Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework, Journal of Hydrology, 226, 48-65, 1999.

Wu, J., Zhang, C., Xue, T., Freeman, B., and Tenenbaum, J.: Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling, 2016, 82-90.

370 Yi, L., Shao, L., Savva, M., Huang, H., Zhou, Y., Wang, Q., Graham, B., Engelcke, M., Klokov, R., and Lempitsky, V.: Large-scale 3d shape reconstruction and segmentation from shapenet core55, arXiv preprint arXiv:1710.06104, 2017.

Strebelle, S.: Conditional simulation of complex geological structures using multiple-point statistics, Mathematical Geology, 34, 1-21, 2002.