Greek Statistical Institute
Proceedings of the 36th Panhellenic & 2nd International Stat Conf (2024), pp.000-000

INTEGRATION OF BIG DATA ANALYTICS IN DIGITAL TWINNING OF MINERAL DEPOSITS

Agni Patra and Ioannis Kapageridis
Laboratory of Mining Informatics and Machine Learning Applications,
Department of Mineral Resources Engineering, University of Western Macedonia,
Greece
dmre00004@uowm.gr, ikapageridis@uowm.gr

ABSTRACT

Having access to in-depth data is necessary for the mining industry to support accurate decision-making, as well as meet a growing list of sustainability and compliance requirements. Digital twin technology is rapidly gaining ground in this data-driven industry to supply this crucial data. A digital twin is a virtual copy of a physical system or object that is constantly updated with data from its real-world physical counterpart, allowing accurate and in-depth simulation, analysis, and monitoring. The application of digital twinning in the case of mineral deposits has been considered recently [Hodgkinson and Elmouttie, (2020); Kumar and Dimitrakopoulos, (2022); Patra et al. (2022)]. Mineral deposits and the associated resources are not dynamic systems or entities. However, our perception of a mineral deposit is dynamic and changes through time during various stages of mineral exploration and later during mining with the inflow of new information and data from multiple sources. The application of big data analytical procedures is crucial in the digital twinning of mineral deposits, and it opens the door to a new age of mineral resource modelling and mine planning tools. This paper presents the way big data analytics can be applied for this purpose and the associated challenges.

Keywords: big data, analytics, digital twin, mineral deposits, integration

1. INTRODUCTION

There are various misconceptions about the definition of a digital twin, which is not just a model of an entity or a process simulation. There are diverse definitions in the literature, but at a minimum, a digital twin is a synchronised, real-time pairing of a virtual and a physical domain. The keywords here are synchronised and real-time, as the digital twin needs to be constantly updated with information from its physical counterpart, resulting in the prediction of its own behaviour and informing decision-makers with sufficient precision to ensure adequate productivity and safety in real-time. A model of a physical system or a simulation of its operation and behaviour is not a digital twin [Wright and Davidson, (2020)]. According to Glaessgen and Stargel

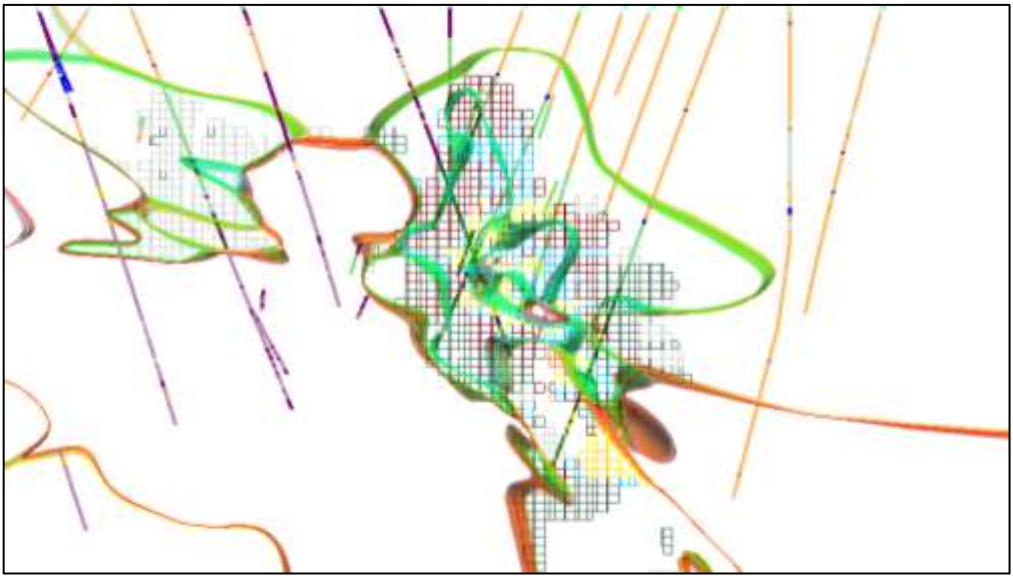
(2012), *"a digital twin means an integrated multiphysics, multiscale, probabilistic simulation of a complex product, which functions to mirror the life of its corresponding twin"*, while Grieves and Vickers (2017), mentioned that *"a digital twin is a set of virtual information constructs that fully describes a potential or actual physical manufactured product from the micro atomic level to the macro geometrical level"*. At its optimum, any information that could be obtained from inspecting a physically manufactured product can be obtained from its digital twin.

2. THE CHALLENGE OF TWINNING A MINERAL DEPOSIT

2.1 Mineral deposits and mineral resource models

A mineral deposit is a concentration of solid material of economic interest in or on the Earth's crust with reasonable prospects for eventual economic extraction. Most of the rock mass of a mineral deposit cannot be seen prior to its excavation and statistical methods and expert judgements are typically used to provide an estimate of its properties and behaviour. Mineral deposits are commonly approached as simplified mineral resource models based on mineable units called blocks. These resource models are built using all necessary available information and they are regularly updated during the life of a mine. We rely on various types of sampling which is normally limited, and we still need expert judgement to reach an estimate of its properties and behaviour. Figure 1 shows a cross section through a resource model, including drillhole traces, a geometrical model of geological boundaries, and resource blocks.

Figure 1. *A cross-section of a resource model.*



2.2 The Dynamic Concept of a Mineral Deposit

Mineral deposits and the contained mineral resources are not dynamic systems or entities. At first sight, they don't seem to be entities that would benefit from a digital twin. However, there is a dynamic concept and our perception of them changes through time during various stages of mineral exploration and later during mining. There is a continuous inflow of new data from various sources and new information during the life of a mining project. The difficult issue here is that current approaches to building models of mineral deposits are work-intensive, repetitive, and time-consuming and lead to a static representation, consisting of multiple disconnected components, such as samples and composites databases, models of spatial variability, geological, structural and topographical models, block models and others, and human intervention is required to update with new data and information. Digital twinning of mineral deposits can be the answer to these issues and provide a more holistic approach for mine planning purposes.

If we integrate all necessary components in a digital twin with adaptive modelling and estimation methods that can reduce human intervention and overall time requirements, that might be the answer to these issues, and that is the subject of our research work.

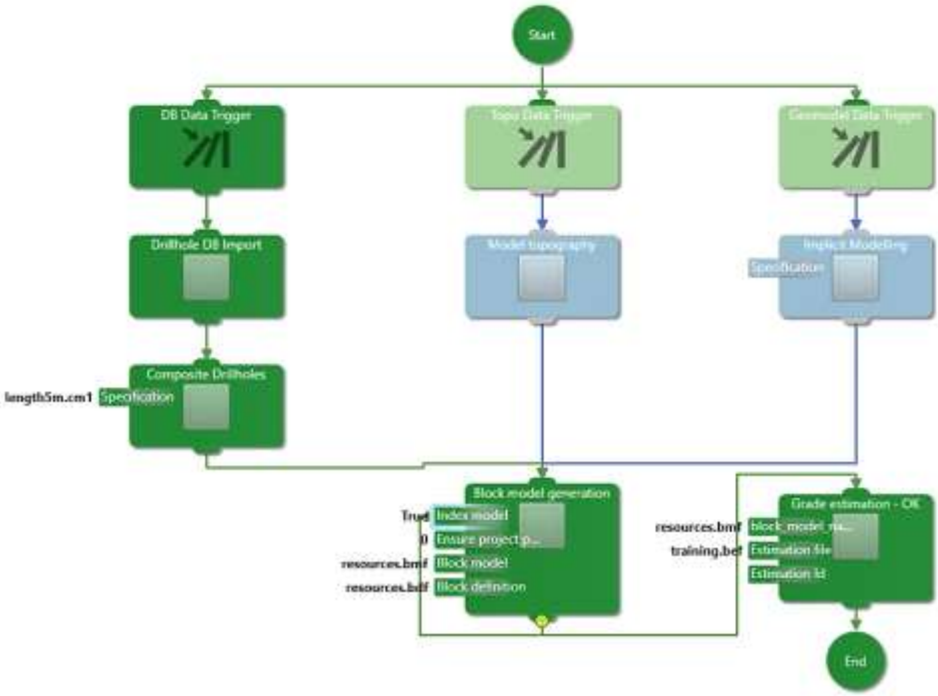
2.3 Mineral Resource Digital Twin Framework

Developing a digital twin of a mineral resource requires the formulation and automation of various analytical and modelling steps that lead to the generation of the necessary model components such as geological and topographical boundaries, structural surfaces, and grade estimates. Furthermore, it needs the usage of artificial intelligence algorithms in combination with incoming new information and sensor data for real-time monitoring [Kumar, A., Dimitrakopoulos, R. (2022)]. A sensible approach to accomplish digital twinning is to rely on an existing mine planning software package with scripting and workflow modelling capabilities. For our research, we use a graphical environment called Workflow Editor which, is part of the Maptek Vulcan software package [Patra, A., *et al.* (2022)]. Workflow Editor is used to develop automated procedures that access a wide range of existing geological modelling, and grade estimation tools and additional data processing and modelling components written in Python. Some of the additional components use machine learning algorithms to speed up the digital twin operation and reduce the requirements for human intervention.

Workflow Editor allows the integration of both existing modelling functionality and new data processing and modelling components. Figure 2 shows a simple example of how Workflow Editor can be used to develop a mineral resource digital twin. There are the Start and End buttons and all the commands are connected via connection lines, giving automatically the outcome of a procedure. The procedure starts with Drillhole import and ends with the Grade estimation of a mineral resource, as represented in

Figure 2. The intermediate stages are Composite Database, Model Topography, Implicit Modelling, and Block Model generation, which are all based on existing Vulcan functionality. The result of the whole procedure can be the resource model of Figure 1. Any mineral deposit could be approached as a digital twin by a similar arrangement in Workflow Editor.

Figure 2. *A flow chart of a resource model digital twin in Workflow Editor.*



2.4 Digital Twins of Mineral Deposits – the Challenges

While engineered systems can be digitally twinned, natural systems containing inherent uncertainties, such as mineral deposits, introduce challenges, especially where human-intensive procedures are required [Hodgkinson, J.H., Elmouttie, M. (2020)]. Traditional mineral resource modelling methods tend to require repetition of the entire procedure to account for changes in the available data and information, even when the scope of these changes is spatially limited. These methods also tend to be less suitable for automation, making them more difficult to integrate into a digital twin of a mineral deposit. On top of these, there are many different types of mineral deposits, each potentially requiring a different combination of analytical and modelling processes, the choice of which remains the result of expert judgment. All these challenges must be addressed, and big data analytics can provide the means.

3. BIG DATA ANALYTICS

Some of the challenges mentioned above can be answered using big data analytics. Big data is a term used to describe large volumes of structured and unstructured data produced by a process on a day-to-day basis. This data can come from various sources with different characteristics, which are commonly referred to as the 3Vs or 5Vs, for volume, velocity, variety, value and veracity. Big data analytics enables the integration, analysis and interpretation of large and complex datasets from diverse sources. It encompasses data from a variety of sources, at the planning and operation stage of an industrial or other processes. The mineral resource modelling process can include this kind of dataset and can significantly be enhanced by enabling the integration, analysis, and interpretation of complex datasets.

3.1 Mineral Deposits as Sources of Big Data

As mentioned above, big data is characterized by its massive volume, high velocity, and diverse variety of data types. In addition to these three Vs, characteristics such as veracity, value, variability, visibility, validity, and volatility further define the complexity and challenges of big data analytics. The common characteristics of big data that also characterize mineral deposits are:

- **Volume:** The amount of data generated periodically. It includes both structured data (e.g., databases) and unstructured data (e.g., social media posts, emails). Large volumes of data can be generated at various stages of mineral exploration and mining, including actual measurements and expert views and interpretations.
- **Velocity:** The speed at which data is generated and collected. With the advent of real-time data streaming and internet-connected devices, data can be generated at an unprecedented pace. The flow of data and information is mostly regular during mineral exploration stages, but quite irregular in between and once mining commences. The digitalisation of exploration and mining gradually reduces the associated time delays.
- **Variety:** The different types of data available. Big data can come in various forms, including text, images, videos, sensor data, and more. Data of mineral deposits can also come in different formats such as spreadsheets, 3D point clouds, survey lines, images, text, and more.

Mineral deposits are indeed sources of big data, as large volumes of data can be generated at various stages of mineral exploration and mining, with the flow of data being mostly regular during these stages, and their format varying quite a lot. The digitalisation of exploration and mining gradually reduces the time delays associated with the availability of important data and information. Mining groups also make efforts to standardize data formats, something that will allow easier flow of data between different processes. Table 1 shows major sources of data.

__Table 1.__ Major sources big data for mineral deposits.

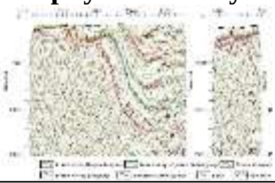
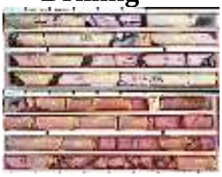
Sources of data / Features	Satellite images 	Geophysical surveys 	Drilling 
Stage	Mineral exploration	Mineral exploration	Mineral exploration and production
Volume	Gigabytes per image	Gigabytes to terabytes per survey	Megabytes
Velocity	Depends on the time resolution of the satellite system	Can be high during the survey	Low as drilling is a non-continuous process
Variety	Low as most images produced are of standard formats	Low as most surveys are recorded in standard file formats	Medium. Drillhole data is recorded in spreadsheet format, but columns/variables will vary from one deposit to another
Scope	Identify rock types, geological structures, and alteration zones, detect surface minerals	Geological modelling including rock types and structures	Geological modelling including domaining, structural modelling, and grade estimation
Automated analysis and interpreta-tion	Possible using supervised and unsupervised methods of image processing, feature extraction, classification, etc.	Possible using supervised and unsupervised methods of analysis and interpretation	Requires (geo)statistics that can be automated to some extent - possible use of machine learning methods

Table 1. *Major Sources of data as Big Data for mineral deposits(continued).*

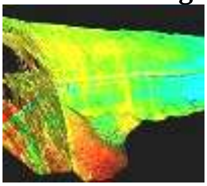
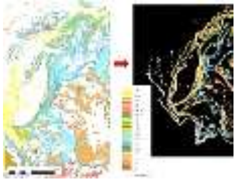
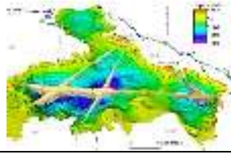
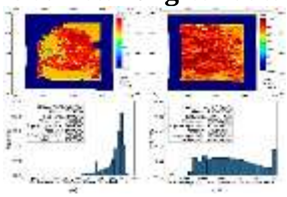
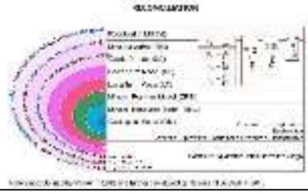
Sources of data / Features	**Laser scanning** 	**Geological maps** 	**Hydrogeological measurements** 
Stage	Mineral exploration and production	Mineral exploration	Mineral exploration and production
Volume	Megabytes to gigabytes	Megabytes up to gigabytes depending on scan resolution	Megabytes
Velocity	Can be high during the survey	Low	Low
Variety	Low as most scan files produced are of standard formats	Low as most images produced are of standard formats	Medium - piezometer readings, ground penetrating radar, water table data
Scope	Geological modelling, topographical modelling, geotechnical analysis	Geological modelling	Geological modelling
Automated analysis and interpretation	Possible using supervised and unsupervised methods of analysis and interpretation	Possible using supervised and unsupervised methods of image processing, feature extraction, classification, etc	Requires statistics and geostatistics that can be automated to some extent - possible use of machine learning methods

__Table 1.__ Major Sources of data as Big Data for mineral deposits(continued).

Sources of data / Features	**Geometallurgical data** 	**Operational data** 
Stage	Mineral exploration and production	Production
Volume	Megabytes	Megabytes
Velocity	Medium	Low
Variety	Low	High as the data can differ depending on the mining method, equipment, life of mine, and even ESG factors
Scope	Assessment of ore recovery, cutoff grades, processing costs, etc.	Resource model orientation and resolution (size of mining units), mining and processing recovery, dilution, operational costs
Automated analysis and interpretation	Principal component analysis (PCA), cluster analysis, and geostatistical methods	Not possible for now – human input is required

According to Table 1, some of the major sources of data for mineral deposits are examined under the prism of big data and within the scope of developing a digital twin. It is quite evident that there are many different sources with fairly different characteristics, related to various stages of mineral exploration and mining production. Data from some of these sources are easier to integrate in an automated process and enable the development of a digital twin, such as satellite images, and LIDAR data, while others traditionally require human intervention to analyse and interpret such as geophysical surveys and drillhole data. The geological maps on one side can be easily analysed with automated procedures, but other sources such as operational data that require human input, are difficult to replace with some machine learning or artificial intelligence algorithms.

3.2 Applications of big data in mineral resource digital twin development

According to Qi and Tao (2018), digital twins as well as big data significantly contribute to the promotion of smart manufacturing. Having the usage of accurate analysis and prediction capabilities of big data, digital twin-driven smart manufacturing will be made more responsive and predictive. Referring to the mining industry, big data analytics can also aid the development of a mineral resource digital twin in the following stages or ways:

1. **Data Integration**: They can facilitate the integration of datasets from various sources, ensuring that the digital twin captures a comprehensive and accurate representation of the mining asset and its surrounding environment. The integration of heterogeneous datasets includes geological, geochemical, geophysical, drilling, metallurgical, and operational data.

2. **Data Preprocessing**: They can automate the preprocessing of raw data, ensuring that data are standardized, consistent, and suitable for analysis, improving the accuracy and reliability of mineral resource models. The preprocessing of raw data includes data cleaning, normalization, transformation, and quality control.

3. **Feature Extraction**: They can enable the extraction of relevant features from large datasets. Feature extraction techniques, including machine learning algorithms and statistical methods, identify key variables that influence mineralization and resource distribution. Examples of relevant features that are extracted from large datasets are mineralogical signatures, geochemical anomalies, geological structures, or spatial patterns.

4. **Predictive Modelling**: Big data analytics can support the development of predictive models that estimate mineral resources and reserves based on geological, geochemical, and geophysical data. Machine learning algorithms, such as random forests, support vector machines, and neural networks, are used to build predictive models that capture complex relationships between input variables and resource outcomes.

5. **Spatial Analysis**: They can also enable spatial analysis of geological and geospatial data to identify spatial trends, patterns, and correlations within the exploration area. Spatial analysis techniques, including geostatistics, spatial interpolation, and spatial clustering, help identify mineralized zones, define exploration targets, and optimize drill hole placements. enable spatial analysis of geological and geospatial data to identify spatial trends, patterns, and correlations within the exploration area.

6. **Uncertainty Analysis**: Uncertainty analysis and risk assessment in mineral resource modelling by quantifying uncertainties associated with data inputs, model parameters, and resource estimates is also an area of application for big data analytics. Monte Carlo simulation, sensitivity analysis, and probabilistic modelling techniques evaluate the impact of uncertainty on resource outcomes and guide decision-making under uncertainty.

7. **Real-time Monitoring**: A major key in developing a digital twin of a mineral resource is real-time monitoring and analysis of exploration and mining operations, and big data analytics can provide the necessary tools to allow it, such as providing timely insights into ore grade, production performance, and process efficiency.
8. **Data Visualisation and Reporting**: Big data analytics supports data visualization and reporting tools that communicate modelling results, insights, and trends to stakeholders and all involved parties effectively. Interactive dashboards, 3D visualization, graphical reports can enhance data interpretation, decision-making, as well as collaboration among multidisciplinary teams.

Some big data analysis methods which can be applicable to mineral resource digital twin development and operation include regression analysis, principal component analysis (PCA), clustering methods, neural networks, random forests, support vector machines, and deep learning (particularly convolutional neural networks). There is a wide range of machine learning algorithms available and, finding, the appropriate one for the analysis of data, from specific sources, is crucial in the development of the digital twin.

4. CONCLUDING REMARKS AND FUTURE WORK

Having access to in-depth data is necessary for the mining industry to support accurate decision-making, as well as meet a growing list of sustainability and compliance requirements. Digital twin technology is rapidly gaining ground in this data-driven industry to supply this crucial data. The application of big data analytical procedures is crucial in the digital twinning of mineral deposits and opens the door to a new age of mineral resource modelling and mine planning tools. By leveraging big data analytics, mining companies can improve the accuracy, efficiency, and effectiveness of mineral resource modelling, leading to better-informed decisions, optimised resource utilization, and enhanced operational performance. Furthermore, by leveraging big data analysis methods, mineral resource digital twins can provide mining companies with a holistic view of their operations, improve decision-making processes, optimize resource utilization, and enhance operational performance across the entire mining lifecycle.

However, more future research work should be done to implement the digital twinning of mineral resources via big data analytics integration. Multiple factors should be considered, such as the virtual-reality interaction environment of the digital twin, numerous sensor data for real-time monitoring, the complexity and diversity of data, the variety and adversity of statistical and artificial intelligence techniques being still difficult to detect the appropriate algorithm, the geological uncertainty and variety of deposits, the digital infrastructure with plenty of technical issues, and human knowledge and perspective that will always be needed to stipulate the value of a digital twin.

ΠΕΡΙΛΗΨΗ

Η πρόσβαση σε εμπεριστατωμένα δεδομένα είναι απαραίτητη για την εξορυκτική βιομηχανία προκειμένου να υποστηριχθεί η ακριβής λήψη αποφάσεων, καθώς και για την εκπλήρωση ενός αυξανόμενου καταλόγου απαιτήσεων βιωσιμότητας και συμμόρφωσης. Η τεχνολογία των ψηφιακών διδύμων κερδίζει γρήγορα έδαφος σε αυτή τη βιομηχανία που βασίζεται στα δεδομένα για την παροχή αυτών των κρίσιμων δεδομένων. Ένας ψηφιακός δίδυμος είναι ένα εικονικό αντίγραφο ενός φυσικού συστήματος ή αντικειμένου που ενημερώνεται συνεχώς με δεδομένα από το πραγματικό φυσικό αντίστοιχό του, επιτρέποντας την ακριβή και σε βάθος προσομοίωση, ανάλυση και παρακολούθηση. Η εφαρμογή της ψηφιακής διδυμοποίησης στην περίπτωση των ορυκτών κοιτασμάτων έχει εξεταστεί πρόσφατα [Hodjkinson and Elmouttie (2020); Kumar and Dimitrakopoulos (2022); Patra *et al.* (2022)]. Τα ορυκτά κοιτάσματα και οι σχετικοί πόροι δεν είναι δυναμικά συστήματα ή οντότητες. Ωστόσο, η αντίληψή μας για ένα ορυκτό κοίτασμα είναι δυναμική και αλλάζει με την πάροδο του χρόνου κατά τη διάρκεια των διαφόρων σταδίων της εξερεύνησης ορυκτών πόρων και αργότερα κατά τη διάρκεια της εξόρυξης με την εισροή νέων πληροφοριών και δεδομένων από διάφορες πηγές. Η εφαρμογή αναλυτικών διαδικασιών μεγάλων δεδομένων είναι ζωτικής σημασίας για την ψηφιακή αδελφοποίηση των κοιτασμάτων ορυκτών πόρων και ανοίγει την πόρτα σε μια νέα εποχή εργαλείων μοντελοποίησης ορυκτών πόρων και σχεδιασμού ορυχείων. Το παρόν έγγραφο παρουσιάζει τον τρόπο με τον οποίο μπορούν να εφαρμοστούν οι αναλυτικές μέθοδοι μεγάλων δεδομένων για τον σκοπό αυτό και τις σχετικές προκλήσεις.

Acknowledgments: The authors gratefully acknowledge the financial support of the University of Western Macedonia and funding of this research.

REFERENCES

Barnewold, L. and Lottermoser, B.G. (2020), *«Identification of digital technologies and digitalisation trends in the mining industry»*, International Journal of Mining Science and Technology, 30, 747-757, Elsevier.

Coetzee, B.J., Sonnendecker, P.W. (2022), *Fully automated coal quality control using digital twin material tracking and statistical model predictive control for yield optimization during production of semi soft coking-and power station coal*, J South Afr Inst Min Metall. 122(8):1–7. doi:10.17159/2411 -9717/ 2002/ 2022.

Fu, T., Zhang, T., Lv Y., Song, X., Li, G., Yue, H. (2023), *Digital twin-based excavation trajectory generation of uncrewed excavators for autonomous mining*. Autom Constr. 151:104855. doi 10.1016/ j.autcon. 2023. 104855.

Glaessgen, E., Stargel, D. (2012). The digital twin paradigm for future NASA and U.S. air force vehicles. 10.2514/6.2012-1818.

Grieves, M., Vickers, J. (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. 10.1007/978-3-319-38756-7_4.

Hodgkinson, J.H., Elmouttie, M. (2020). *Cousins, Siblings and Twins: A Review of the Geological Model's Place in the Digital Mine*. Resources 2020, 9, **24**; DOI:10.3390/resources9030024.

Kumar, A., Dimitrakopoulos, R. (2022). *Updating geostatistically simulated models of mineral deposits in real-time with incoming new information using actor-critic reinforcement learning.* Computers & Geosciences **158** (2022) 104962, DOI: 10.1016/j.cageo.2021.104962.

Nagovitsyn, O.V., Stepacheva, A.V. (2021), *Digital twin of solid mineral deposit*. J Min Sci. 57(6): 1033–1040. doi:10.1134 /S10 62739 12106 0168.

Nogueira de Andrade, M.A., Lepikson, H. A., Machado, C.A. (2021), *A New Framework and Methodology for Digital Twin Development*, Conference Paper DOI: 10.1109/ INDUSCON 51756. 2021. 9529701. https://www.researchgate.net/publication/354514967.

Patra, A., Kapageridis, I., Asvesta, A., Sinatkas, I. (2022). *Framework for the Development of a Mineral Resource Digital Twin.* 16th International Congress of the Geological Society of Greece, Bulletin of the Geological Society of Greece, Sp. Publ. 10.

Plavšić, J. and Mišković, I. (2023), *Industrial applications of digital twin technology in the mining sector: An overview*, CIM Journal, 14:2, 97-106, DOI: 10.1080/19236026.2022.2145431.

Qi, Q., and Tao, F. (2018), *Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison*, Volume 6, p:35385-3593, DOI: 10.1109/ IEEE Journal ACCESS.2018.2793265.

Wright, L. and Davidson, S. (2020), *How to tell the difference between a model and a digital twin*, Advanced Modelling and Simulation in Engineering Sciences, volume 7, article number 13.