The Robert M. Buchan Department of Mining

Predictive Geometallurgy and Geostatistics Lab

Annual Report 2021

Edited by: Julian M. Ortiz

Queen's University

Predictive Geometallurgy and Geostatistics Lab
Queen's University

Annual report 2021

This report summarizes the ongoing research of the Predictive Geometallurgy and Geostatistics Laboratory at Queen's University in Kingston, Ontario, Canada. 2021 was a challenging year, with continuous restrictions to meetings and limited group interactions. Despite these difficulties, the lab completed important and novel work. The delay in the release of this report is just another sign of the strain of the last couple of years, however, results are encouraging and the lab is doing important contributions to research and industry.

This year, two students graduated, one Master of Applied Science and one Doctor of Philosophy, two new students joined the group. The following two theses were completed in this period:

- Mehmet Altinpinar, M.A.Sc. (Sep. 2021), "***Synthetic high resolution block model for benchmarking mining technologies***"
- Sebastian Avalos, Ph.D. (Sep. 2021), "***Advanced predictive methods applied to geometallurgical modelling***"

The work in this annual report includes that of the six graduate students active in 2021. The research group is composed of:

- Mehmet Altinpinar, M.A.Sc. student
- Sebastian Avalos, Ph.D. student (continues as a Post-Doc)
- David Casson, Ph.D. student
- Kasimcan Koruk, M.A.Sc. student
- Paula Larronfo, Ph.D. student
- Alvaro Riquelme, Ph.D. student

Two new M.A.Sc. students started their programs in 2021

- Noble Potakey, M.A.Sc. student
- Alvaro Mariño, M.A.Sc. student

We continued collaboration with other faculty members and researchers, including:

- Willy Kracht, Adjunct Professor – The Robert M. Buchan Department of Mining (Queen's University) and Associate Professor – Department of Mining Engineering (U. de Chile). Dr. Kracht and Dr. Ortiz co-supervise Carlos Moraga in his Ph.D. in Mining Engineering at Universidad de Chile.

- Asli Sari, Assistant Professor – The Robert M. Buchan Department of Mining (Queen's University). Dr. Ortiz and Dr. Sari co-supervised Mehmet Altinpinar in his M.A.Sc.
- Raimon Tolosana-Delgado, Senior Scientist (Helmholtz-Zentrum Dresden-Rossendorf). Dr. Tolosana-Delgado hosted a research internship of Sebastian Avalos.
- Brian Frank, Professor – Electrical and Computer Engineering (Queen's University). Dr. Frank and Dr. Ortiz co-supervise Paula Larrondo in her Ph.D.

Eight contributions are available this year, totaling 92 pages, with very innovative topics, including causal inference, reinforcement learning, and topology of random fields, in addition to documentation of machine learning methods, planning and geostatistical methods. Industrial collaboration continues with SRK Consulting Canada, Natural Research Council (NRC) and ArcelorMittal Mining Canada G.P.

As always, we welcome industrial and academic collaboration. This provides opportunities to fund new graduate students and novel research, and directly benefits industrial partners. If interested, please send a note to julian.ortiz@queensu.ca.

Julian M. Ortiz
Associate Professor, The Robert M. Buchan Department of Mining
Director, Predictive Geometallurgy and Geostatistics Lab
Queen's University

December 2021

Table of contents

Ortiz JM (2021) ***Progress towards geometallurgical digital twins***, paper 2021-01 7

Avalos S, Ortiz JM (2020) ***Fundamentals of Deep Q-Learning***, paper 2021-02 14

Avalos S, Ortiz JM (2021) ***Understanding process performance with causal inference for continuous variables***, paper 2021-03 22

Riquelme A, Ortiz JM (2020) ***Notes on the topology of random fields, part I***, paper 2021-04 33

Casson D, Ortiz JM (2020) ***Application of Disjunctive Kriging in Sequential Simulation,*** paper 2021-05 46

Altinpinar M, Ortiz JM (2021) ***A quick guide to developing a mine plan***, paper 2020-06 61

Koruk K (2021) ***A simple implementation example of SVC,*** paper 2021-07 79

Koruk K (2021) ***Relevance Vector Machines: an introduction,*** paper 2021-08 88

Journal and Conference Publications and Presentations

Publications in book chapters, peer-reviewed journals and international conferences are listed below for 2021. These are not included in this report, since the copyright belongs to the corresponding publishers, but can be requested for personal use or research purposes directly to julian.ortiz@queensu.ca.

Book chapters

1. Sadeghi B, Ortiz JM (2021) ***Simulation***, in Encyclopedia of Mathematical Geosciences, Daya Sagar B, Cheng Q, McKinley J, Agterberg F (Eds.), Encyclopedia of Earth Sciences Series, Springer, Cham, 6 p. https://doi.org/10.1007/978-3-030-26050-7_292-1

2. Caers J, Mariethoz G, Ortiz JM (2021) ***Multiple Point Statistics***, in Encyclopedia of Mathematical Geosciences, Daya Sagar B, Cheng Q, McKinley J, Agterberg F (Eds.), Encyclopedia of Earth Sciences Series, Springer, Cham, 11 p. https://doi.org/10.1007/978-3-030-26050-7_24-1

Journal papers

1. Cevik IS, Leuangthong O, Cate A, Ortiz JM (2021) ***On the use of machine learning for mineral resource classification***, Mining, Metallurgy & Exploration, 38:2055-2073. https://doi.org/10.1007/s42461-021-00478-9

2. Faraj F, Ortiz JM (2021) ***A simple unsupervised classification workflow for defining geological domains using multivariate data***, Mining, Metallurgy & Exploration, 38: 1609-1623. https://doi.org/10.1007/s42461-021-00428-5

3. Riquelme AI, Ortiz JM (2021) ***Uncertainty assessment over any volume without simulation: revisiting multi-Gaussian kriging***, Mathematical Geosciences, 53:1375-1405. https://doi.org/10.1007/s11004-020-09907-9 (Correction: https://doi.org/10.1007/s11004-021-09927-z)

4. Cevik IS, Ortiz JM, Olivo GR (2021) ***A combined multivariate approach analyzing geochemical data for knowledge discovery: the Vazante-Paracatu Zinc district, Minas Gerais, Brazil***, Journal of Geochemical Exploration, Vol. 221, 106696. https://doi.org/10.1016/j.gexplo.2020.106696

Conference papers and presentations

1. Riquelme AI, Ortiz JM (2021) ***A non-stationary linear model of coregionalization***, in 11th International Geostatistical Congress, Toronto 2021, July 12-16, 2021.

2. Avalos A, Ortiz JM (2021) ***Geometallurgical modeling and deep Q-Learning to optimize mining decisions***, in 11th International Geostatistical Congress, Toronto 2021, July 12-16, 2021.

3. Cevik IS, Leuangthong O, Cate A, Machuca-Mory D, Ortiz JM (2021) ***Mineral resource classification using machine learning***, in 11th International Geostatistical Congress, Toronto 2021, July 12-16, 2021.

4. Larrondo P, Frank B, Ortiz JM (2021) ***State of the art in providing automated feedback to open-ended student work***, CEEA/ACEG 2021, Annual Conference of the Canadian Engineering Education Association, Charlottetown, PEI, June 20-23, 2021.

5. Riquelme AI, Ortiz JM (2021) ***An approach to characterize complex geological models based on higher-dimensional surfaces***, Geomin-Mineplanning 2021, 7th International Conference on Geology and Mine Planning, June 9-11, 2021.

6. Avalos S, Ortiz JM (2021) ***Open pit mine scheduling via deep Q-Learning***, Geomin Mineplanning 2021, 7th International Conference on Geology and Mine Planning, June 9-11, 2021.

7. Avalos S, Ortiz JM (2021) ***Heuristic risk-based policy to outline final pit in open mines***, CIM VTL 2021, Virtual Convention + Expo, May 3-6, 2021.

8. Riquelme AI, Ortiz JM (2021) ***A geostatistical approach to characterize complex geology***, CIM VTL 2021, Virtual Convention + Expo, May 3-6, 2021.

9. Ortiz JM (2021) ***Geometallurgical modeling to manage uncertainty in a mining system***, invited talk, 10 Years of Helmholtz Institute Freiberg for Resource Technology, September 9, 2021, Germany.

10. Ortiz JM (2021) ***Machine learning in mining***, invited webinar, Colegio de Ingenieros de Peru – Consejo Departamental de La Libertad, Apr 23, 2021. https://fb.watch/575-W1q_nB/

11. Ortiz JM (2021) ***Predictive models in geometallurgy***, invited seminar, Geoblast Chile, Dec 16, 2020.

12. Ortiz JM (2021) ***Multiple point geostatistics to model rock textures***, invited talk, NRC (National Research Council), Nov 5, 2020.

Funding

Research is possible thanks to the funding provided by Queen's University Research Initiation Grant, NSERC through funding reference nos. RGPIN-2017-04200 and RGPAS-2017-507956, Mitacs Globalink IT17457 in collaboration with TU Bergakademie Freiberg, Germany, NSERC-Alliance ALLRP 554627-20, in collaboration with ArcelorMittal Mining Canada, G.P., and National Research Council.

Progress towards geometallurgical digital twins[1]

Julian M. Ortiz (julian.ortiz@queensu.ca)

Abstract

A digital twin should capture the behavior of a process. However, most geometallurgical steps have uncertain inputs, and uncertain responses, since the ore properties are variable and not fully known, and the physics and chemistry involved in the processes may be too complex to fully understand them. Despite these challenges, uncertainty can be managed and a geometallurgical digital twin can be built, incorporating this uncertainty as a random variable. The response is therefore variable but can be optimized. In this note, some insights are provided about the steps taken in the research community to create the building blocks of what can become an integrated digital twin of the geometallurgical processing of the ore in a mining system.

1. Introduction

A digital twin can be defined as an ultra high-fidelity simulation of a real object or process, and that can be connected to this physical object or real process. Digital twins originated in the manufacturing industry and were later adopted by NASA for applications such as testing a vehicle under extreme conditions for space exploration and military applications [Glaessgen and Stargel, 2012]. In most applications, a concrete object or process is considered, thus limiting the modeling effort to a relatively limited number of physical laws or chemical reactions that are well understood.

Mining is moving towards Industry 4.0, incorporating concepts of interconnectivity, through Internet of Things (IoT) and smart automation [Loow et al., 2019]. The true digitalization of the mining industry will happen only when automated decision-making can be implemented. Currently there are many efforts in progress to implement integration, predictive modeling, and automation in the mining industry [Dominy et al. 2018], but the idea of a fully integrated and fully automated operation is far from practical at this point.

In the coming sections, approaches that are already available are discussed, that could be integrated into a geometallurgical framework, to provide the basis for automated decisions. Some of the missing components needed to achieve the idea of a geometallurgical digital twin of a mining operation are also identified. In particular, the focus is on uncertainty management and decision-making.

2. Geometallurgical framework

A geometallurgical digital twin (GDT) can be created by combining realistic models of the different stages of the mining value chain, where particular inputs lead to outputs that feed other processes downstream (Figure 1).

[1] Cite as: Ortiz JM (2021) Geometallurgical modeling framework, Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report 2021, paper 2021-01, 7-13.

.

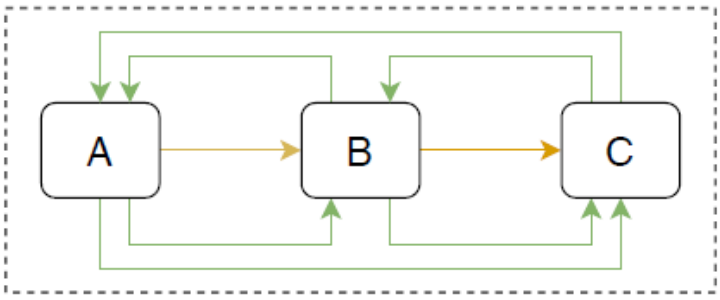


Figure 1: Illustration of connected processes [Avalos, 2021].

What is important, is that the outputs include all the relevant attributes that will influence the response of the process and of all subsequent processes (depicted as forward green arrows in the previous figure) [Ortiz et al., 2015; Avalos, 2021]. This means that the model must integrate many variables and if these variables cannot be measured, they must be incorporated through "soft sensors" or non-regressive predictive models (for example, stochastic models) that can realistically capture the associated uncertainty around an unbiased estimate of the true value. Furthermore, the components must be interrelated to account for the interactions between processes and feedback (backward green arrows in previous figure). Real time measurements can help maintain a stream of information used to calibrate and control the system [Benndorf and Jansen, 2017].

3. Building blocks

A mining system can be seen as a sequence of stages or processes. Broadly, Figure 2 shows a simplified depiction of the parallel between the actual ore deposit, mine and operation (at the top) and the modelled resources, reserves and extraction (bottom). Ideally, in a GDT the model (bottom) should be constantly fed with information from the real operation (top), and the stages and processes in the model should follow the systems approach presented in Figure 1.

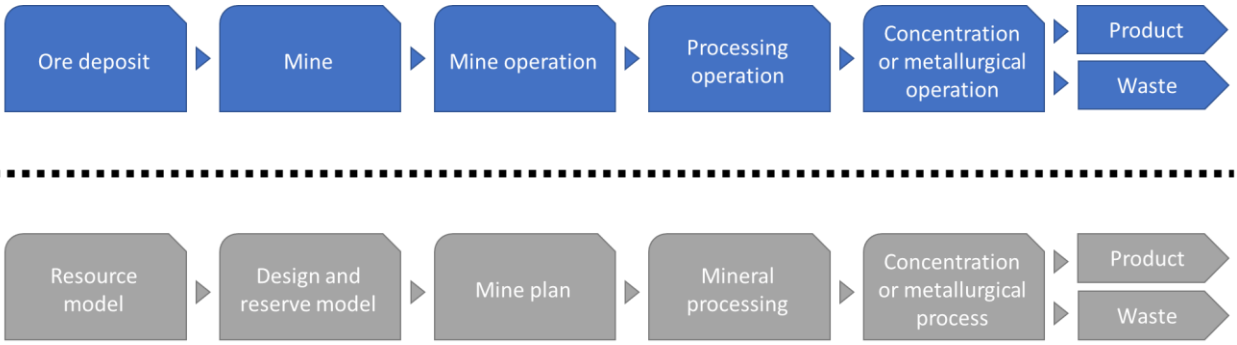


Figure 2: Mining system: actual (top) vs model (bottom).

Given the complexity of stages and processes in mining, the behavior of each component of the system can only be approximated. Real time sensors, composited measurements, and soft sensors can be used to update the status of the twin (model). This requires smart sensors and measurements, high speed communication to transfer this information in real time, and models that can predict the response and assess the potential variability linked to the uncertainty in rock properties, and in the process

performance. This predicted output must be compared to other measurements of the actual output, for the models to learn. Once properly calibrated and trained, a smart agent can take over the decision-making process, to: (1) Optimize each process and (2) Optimize the system. It should be emphasized that today, automatic control systems exist, but are limited to specific processes, particularly in processing plants.

4. Examples

Each stage in Figure 2 has seen significant progress with the use of geostatistics, machine learning, deep learning, and other statistical modeling techniques. Table 1 reviews some examples where these techniques are applied and references where these methods and models are developed.

Table 1: example applications of modeling into different stages of the mining value chain.

Stage	Step	Technique	References
Resource model	Domaining	Unsupervised geochemical classification for domaining	[Faraj and Ortiz, 2021]
		Machine learning to model alteration	[Berube et al., 2018]
		Geostatistical clustering	[Fouedjio at al., 2017]
	Geological modeling	Pluri-Gaussian simulation with local proportions	[Emery et al, 2008]
		Indicator simulation with locally varying directions	[Gutierrez and Ortiz, 2019]
		Deep learning for geological modeling	[Avalos and Ortiz, 2020]
	Attributes modeling	Multivariate modeling of geometallurgical attributes	[Deutsch et al., 2015]
		Compositional data modeling	[Tolosana-Delgado et al., 2019]
		Projection Pursuit multivariate transformation	[Barnett et al., 2014]
	Upscaling	Non-linear modeling of geometallurgical attributes	[Deutsch, 2015]
		Uncertainty assessment at any block support	[Riquelme and Ortiz, 2021]
		Change of support of non-additive variables	[Garrido et al., 2019]
Design and reserve model	Design optimization	Risk-based selection of ultimate pit limit	[Jelvez et al., 2022]
		Underground design optimization	[Sari and Kumral, 2020]
		Surface and underground optimization under uncertainty	[Montiel et al., 2015]
	Classification	Resource and reserve classification with machine learning	[Cevik et al. 2021]
Mine plan	Extraction sequence	Stochastic optimization for planning	[Dimitrakopoulos, 2011]
	Schedule	Stochastic integer programming accounting for uncertainty of geometallurgical attributes	[Morales et al., 2019]
	Demand-side management	Managing energy consumption via DSM for integration of renewable energy sources	[Diaz et al., 2016]
		Cutoff grade optimization based on stochastic resource models with stockpile for long-term planning	[Asad and Dimitrakopoulos, 2012]
Mineral processing	Crushing and grinding	Hardness prediction with deep learning	[Avalos et al, 2020a]
		Data-driven grinding processing modeling	[Lv et al., 2020]
	Particle size classification	AI driven air classification of particles	[Otwinowski et al., 2021]
Concentration or metallurgical process	Flotation	Deep learning to determine froth flotation performance	[Pu et al., 2020a]
		Deep learning froth flotation recovery prediction	[Pu et al., 2020b]
	Leaching	Recovery prediction in leaching using machine learning	[Flores and Leiva, 2021]

From this list, it is easy to see that, if these methods are wrapped appropriately to be connected into a larger and integrated system, the first steps towards automatic prediction, learning and automatic decision-making are possible. Encouraging results in the design of intelligent agents with reinforcement learning have already been developed [Avalos, 2021] and demonstrate that a data-driven approach that continuously learns and refines its results is possible, leading to a twin of the actual deposit and operation.

The uncertainty caused by the limited sampling, can be compensated with real-time feedback loops to ensure the model remains calibrated and close to the actual state of the operation.

Each one of the stages described above has a broad variety of problems, hence models are available for specific circumstances and need generalization. There are also many other aspects of the "towards full automation and integration" vision that can be considered, including: product control, environmental footprint control, energy and water use. See for example [Avalos et al, 2020b; Ortiz et al., 2020].

Finally, it is important to mention that these technologies can only be integrated if a proper high-speed and low latency communication protocol, such as the 6G technology [Boxall and Lacoma, 2021], and adequate computer power or algorithmic efficiency [Peredo et al., 2015; Peredo et al., 2018] is developed.

5. Conclusions

A digital twin of an operating mining operation is possible if advanced predictive technologies are used to put in place a model of each stage that accounts for the proper inputs and outputs that have an impact in the entire value chain. Tracking materials and measuring properties becomes essential to capture the system's behavior and learn from it, through data. Enabler technologies, such as high-speed data communication and high-performance computing are essential to achieve a fully automated and integrated model, that matches the operation state, while handling the forecasted uncertainty at every step and optimizing decisions, under these circumstances.

6. References

Asad MWA, Dimitrakopoulos R (2012) Optimal production scale of open pit mining operations with uncertain metal supply and long-term stockpiles, Resources Policy 37(1): 81-89. https://doi.org/10.1016/j.resourpol.2011.12.002

Avalos S (2021) Advanced predictive methods applied to geometallurgical modelling, Ph.D. thesis, Queen's University.

Avalos S, Ortiz JM (2020) Recursive Convolutional Neural Networks in a Multiple-Point Statistics framework, Computers & Geosciences, 141, 104522. https://doi.org/10.1016/j.cageo.2020.104522

Avalos SA, Kracht W, Ortiz JM (2020a) An LSTM approach to SAG mill operational relative-hardness prediction, Minerals, 10 (9), 734. https://doi.org/10.3390/min10090734

Avalos SA, Kracht W, Ortiz JM (2020b) Machine Learning and Deep Learning Methods in Mining Operations: a Data-Driven SAG Mill Energy Consumption Prediction Application, Mining, Metallurgy & Exploration, 37(4), 1197-1212. https://doi.org/10.1007/s42461-020-00238-1

Barnett RM, Manchuk JG, Deutsch CV (2014) Projection pursuit multivariate transform. Mathematical Geosciences, 46 (2), 337–359. https://doi.org/10.1007/s11004-013-9497-7

Benndorf J, Jansen JD (2017) Recent Developments in Closed-Loop Approaches for Real-Time Mining and Petroleum Extraction, Mathematical Geosciences, 49, 277-306. https://doi.org/10.1007/s11004-016-9664-8

Berube CL, Olivo GR, Chouteau M, Perrouty S, Shamsipour P, Enkin RJ, Morris WA, Feltrin L, Thiemonge R (2018) Predicting rock type and detecting hydrothermal alteration using machine learning and

petrophysical properties of the Canadian Malartic ore and host rocks, Pontiac Subprovince, Québec, Canada, Ore Geology Reviews, 96, 130-145.

Boxall A, Lacoma T (2021) What is 6G, how fast will it be, and when is it coming?, DigitalTrends. Retrieved December 10, 2021. https://www.digitaltrends.com/mobile/what-is-6g/

Cevik IS, Leuangthong O, Cate A, Ortiz JM (2021) On the use of machine learning for mineral resource classification, Mining, Metallurgy & Exploration, 38, 2055-2073. https://doi.org/10.1007/s42461-021-00478-9

Deutsch JL (2015) Multivariate spatial modelling of metallurgical rock properties, Ph. D. thesis, University of Alberta, 192 pp.

Deutsch JL, Palmer K, Deutsch CV, Szymanski J, Etsell TH (2015) Spatial Modeling of Geometallurgical Properties: Techniques and a Case Study, Natural Resources Research, 25(2), 161-181.

Dıaz S, Gil E, Palacios A (2016) Demand side management, an imperative asset for the mining industry, ENERMIN 2016: 3rd International Seminar on Energy Management in Mining, Santiago de Chile, 2016, p 10.

Dimitrakopoulos R (2011) Stochastic optimization for strategic mine planning: a decade of developments. Journal of Mining Science 84, 138–150. https://doi.org/10.1134/S1062739147020018

Dominy SC, O'Connor L, Parbhakar-Fox A, Glass HJ, Purevgerel S (2018) Geometallurgy—A Route to More Resilient Mine Operations, Minerals 8(12): 560. https://doi.org/10.3390/min8120560

Emery X, Ortiz JM, Cáceres AM (2008) Geostatistical modeling of rock type domains with spatially varying proportions: Application to a porphyry copper deposit, Journal of the South African Institute of Mining and Metallurgy, 108(5), 285-292. http://ref.scielo.org/xnnyxm

Faraj F, Ortiz JM (2021) A simple unsupervised classification workflow for defining geological domains using multivariate data, Mining, Metallurgy & Exploration, 38, 1609-1623. https://doi.org/10.1007/s42461-021-00428-5

Flores V, Leiva C (2021) A Comparative Study on Supervised Machine Learning Algorithms for Copper Recovery Quality Prediction in a Leaching Process. Sensors, 21(6): 2119. https://doi.org/10.3390/s21062119

Fouedjio F, Hill EJ, Laukamp C (2018) Geostatistical clustering as an aid for ore body domaining: Case study at the Rocklea Dome channel iron ore deposit, Western Australia. Applied Earth Science, 127(1), 15–29. https://doi.org/10.1080/03717453.2017.1415114

Glaessgen EH, Stargel DS (2012) The digital twin paradigm for future NASA and U.S. Air Force Vehicles, 53rd Structures, Structural Dynamics, and Materials Conference, American Institute of Aeronautics and Astronautics.

Gutiérrez R, Ortiz JM (2019) Sequential indicator simulation with locally varying anisotropy – Simulating mineralized units in a porphyry copper deposit, Journal of Mining Engineering and Research, 1(1), 1-7. https://doi.org/10.35624/jminer2019.01.01

Garrido M, Ortiz JM, Villaseca F, Kracht W, Townley B, Miranda R (2019) Change of support using non-additive variables with Gibbs Sampler: application to metallurgical recovery of sulphide ores, Computers & Geosciences, 122, 68-76. https://doi.org/10.1016/j.cageo.2018.10.002

Jelvez E, Morales N, Ortiz JM (2022) Stochastic final pit limits: an efficient frontier analysis under geological uncertainty in the open-pit mining industry, Mathematics, 10(1): 100. https://doi.org/10.3390/math10010100

Lööw J, Abrahamsson L, Johansson J (2019) Mining 4.0—the Impact of New Technology from a Work Place Perspective. Mining, Metallurgy & Exploration 36, 701–707. https://doi.org/10.1007/s42461-019-00104-9

Lv L, Deng Z, Liu T, Li Z, Liu W (2020) Intelligent technology in grinding process driven by data: A review. Journal of Manufacturing Processes, 2020 (58), 1039–1051. https://doi.org/10.1016/j.jmapro.2020.09.018

Montiel L, Dimitrakopoulos R, Kawahata K (2016) Globally optimising open-pit and underground mining operations under geological uncertainty. Mining Technology, 125, 2–14. https://doi.org/10.1179/1743286315Y.0000000027

Morales N, Seguel S, Cáceres A, Jélvez E, Alarcón M (2019) Incorporation of Geometallurgical Attributes and Geological Uncertainty into Long-Term Open-Pit Mine Planning. Minerals, 9(2): 108 https://doi.org/10.3390/min9020108

Ortiz JM, Kracht W, Townley B, Lois P, Cárdenas E, Miranda R, Álvarez M (2015) Workflows in Geometallurgical Prediction: Challenges and Outlook, in 17th Annual Conference IAMG 2015, Schaeben H, Tolosana-Delgado R, van den Boogaart KG, van den Boogaart R (Eds.), 6 p.

Ortiz JM, Kracht W, Pamparana G, Haas J (2020) Optimization of a SAG mill energy system: integrating rock hardness, solar irradiation, climate change and demand side management, Mathematical Geosciences, 52(3), 355-379. https://doi.org/10.1007/s11004-019-09816-6

Otwinowski H, Krzywanski J, Urbaniak D, Wylecial T, Sosnowski M (2022) Comprehensive Knowledge-Driven AI System for Air Classification Process. Materials, 15(1): 45. https://doi.org/10.3390/ma15010045

Peredo OF, Baeza D, Ortiz JM, Herrero JR (2018) A path-level exact parallelization strategy for sequential simulation, Computers & Geosciences, 110, 10-22. https://doi.org/10.1016/j.cageo.2017.09.011

Peredo O, Ortiz JM, Herrero JR (2015) Acceleration of the Geostatistical Software Library (GSLIB) by code optimization and hybrid parallel programming, Computers & Geosciences, 85(A), 210-233. https://doi.org/10.1016/j.cageo.2015.09.016

Pu Y, Szmigiel A, Apel DB (2020a) Purities prediction in a manufacturing froth flotation plant: The deep learning techniques. Neural Computing and Applications, 2020, 1–11. https://doi.org/10.1007/s00521-020-04773-2

Pu Y, Szmigiel A, Chen J, Apel DB (2020) FlotationNet: A hierarchical deep learning network for froth flotation recovery prediction. Powder Technology, 375, 317–326. https://doi.org/10.1016/j.powtec.2020.07.102

Riquelme AI, Ortiz JM (2021) Uncertainty assessment over any volume without simulation: revisiting multi-Gaussian kriging, Mathematical Geosciences, 53, 1375-1405. https://doi.org/10.1007/s11004-020-09907-9

Sari YA, Kumral M (2020) A Planning Approach for Polymetallic Mines Using a Sublevel Stoping Technique with Pillars and Ultimate Stope Limits, Engineering Optimization, 52(6), 932–944. https://doi.org/10.1080/0305215X.2019.1624739

Tolosana-Delgado R, Mueller U, van den Boogaart KG (2019) Geostatistics for Compositional Data: An Overview. Mathematical Geosciences 51, 485–526. https://doi.org/10.1007/s11004-018-9769-3

Annual Report 2021
Paper 2021-02

GEOMET
@QUEEN'S
UNIVERSITY

Fundamentals of deep Q-Learning [1]

Sebastian Avalos (sebastian.avalos@queensu.ca)
Julian M Ortiz (julian.ortiz@queensu.ca)

Abstract

Reinforcement learning has achieved remarkable performances on oriented decision making problems. The agent-environment framework provides the principles for mapping states-and-actions to expected value rewards, maximizing the long-term total reward. The mapping function can be retrieved from look-up tables when the space of states and actions are small enough to maintain computational efficiency, or parametrized as an approximation of the underlying mapping. In real life problems, the environment is often incomplete, and the space of states and actions are non trackable or computationally unmanageable. Recent advances on Deep Learning have led to implement deep neural networks to approximate the mapping function, referred as deep Q-Learning. In this brief article, we review the building blocks of reinforcement learning with a final focus on the principles of deep Q-Learning.

1. Introduction

The field of reinforcement learning (RL) has its roots and draws insights from neurosciense, psychology, and computer science (Ludvig et al., 2011). Collaborative efforts have helped to strengthen the RL framework, providing methods and models on how agents (animals, humans or robots) learn to make decisions from past experiences of agent-environment interaction (Sutton and Barto, 2018). From a computational perspective, reinforcement learning is the framework of machine learning in which an agent is trained to maximize the reward over time as a result of chosen actions in a sequence of interactions within a particular environment. The learning process follows the principles of sequential decision making, where actions influence the immediate reward, the environment state, and all subsequent environment states, feasible actions and possible states. Therefore, the agent must learn the evaluate the quality of taking an action based on the current environment state and according to the immediate reward and delayed rewards.

To introduce concepts, we explore the following reduced-learning setting: every time, an agent must make a choice between k different actions, receiving an immediate reward drawn from a stationary distribution, without perceiving and/or altering the environment. The aim is to maximize the total reward in a finite number of choices (time steps). Let a_t be the action selected at time step t, and r_t the corresponding reward. The value of selecting the action a corresponds to the expected reward, and can be expressed as:

$$q_*(a) \doteq \mathbb{E}[r_t | a_t = a] \tag{1}$$

[1] Cite as: Avalos S, Ortiz JM (2021) Fundamentals of deep Q-Learning, Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report 2021, paper 2021-02, 14-21.

Maximizing the total reward is trivial when $q_*(a)$ is known for all possible actions a: select the action with highest value. Naturally, we do not have access to the real value $q_*(a)$ but it can be estimated based on previous experiences. Let $q_t(a)$ be the estimated value of selecting action a at time step t. Without considering when an action was taken but only the number of times it was taken and each of the corresponding rewards, a simple way to estimate the value of an action is by the sample-average method as:

$$q_t(a) = \frac{\sum_{i=1}^{t-1} r_i \cdot \mathbb{1}_{a_i=a}}{\sum_{i=1}^{t-1} \mathbb{1}_{a_i=a}} \tag{2}$$

where $\mathbb{1}_{a_i=a}$ is 1 when $a_i = a$, and 0 otherwise. The action with highest value is drawn as:

$$a_t = \arg\max_a q_t(a) \tag{3}$$

The action(s) with highest estimated value is called a greedy action. We refer to the process of selecting a greedy action as *exploitation*, since the agent is exploiting the accumulated knowledge of previous experiences. We refer to the selection of non-greedy action as *exploration*. The latter allows the agent to update the estimate of non-greedy actions. The trade-off between exploration and exploitation is non trivial. Nevertheless, we can easily argue that exploration is fundamental in the early stages of a learning process, whereas exploitation is desired when certain stationarity is observed in the estimated values, in the latest stages of a learning process.

When looking at a single action that has been selected n times, we can compute the current estimated value q_n as:

$$\begin{aligned}
q_{n+1} &= \frac{r_1 + r_2 + \cdots r_{n-1} + r_n}{n} \\
q_{n+1} &= \frac{1}{n}\left(r_n + \frac{n-1}{n-1}\sum_{i=1}^{n-1} r_i\right) \\
q_{n+1} &= \frac{1}{n}\Big(r_n + (n-1)\cdot q_n\Big) \\
q_{n+1} &= q_n + \frac{1}{n}\Big(r_n - q_n\Big)
\end{aligned} \tag{4}$$

The previous update representation, from q_n to q_{n+1} knowing the last reward r_n and the number of times that the action has been taken n, has the structure:

$$\text{NewEstimation} \longleftarrow \text{OldEstimation} + \alpha \cdot [\text{Target} - \text{OldEstimation}] \tag{5}$$

The expression $[\text{Target} - \text{OldEstimation}]$ denotes the error between the desired value and the old estimation. The parameter α controls the rate in which the estimation value is updated, and is often expressed as a function of the time step and the corresponding action, $\alpha_t(a)$.

The distribution of reward probabilities has been assumed constant over time. In this scenario, an equal weight of previous experience is reasonable, such as $\alpha_t(a) = \frac{1}{n}$, which changes over time. For non-stationary situations, the intuition suggests to increase the weights to recent experiences and decrease the weights of old ones. Using a constant value $\alpha \in (0, 1]$ satisfies the desired property, transforming Eq. 4 into a weighted sum of the past reward and past estimation :

$$q_{n+1} = \alpha \cdot r_n + (1 - \alpha) \cdot q_n \tag{6}$$

By recursion, Equation 6 can be rewritten as a function of past rewards and the initial estimation as:

$$q_{n+1} = (1 - \alpha)^n \cdot q_1 + \sum_{i=1}^{n} \alpha \cdot (1 - \alpha)^{n-i} \cdot r_i \tag{7}$$

To guarantee convergence over time, we need α to satisfy both:

$$\sum_{t=1}^{\infty} \alpha_t(a) = \infty \qquad , \qquad \sum_{t=1}^{\infty} \alpha_t^2(a) < \infty \tag{8}$$

where the former expression implies enough steps to overcome initial conditions, while the latter expression implies a decrease in the step-size during learning to secure convergence. Note that the latter expression is not met when *alpha* is set constant, a desired property in non-stationary situations.

Until now, the learning process has focused on the estimation of the action values to maximize a total reward. Either stationary or non-stationary, the value of each action has been assumed unrelated to the context of learning. When the context is considered, the agent must learn how to evaluate actions conditioned to different situations. The dynamic of learning in a agent-environment framework is described in the following section.

2. The agent-environment framework

The interaction agent-environment is often discretized in time steps. Time steps are not required to represent the formal time dimension but rather sequential decisions steps. Formally, at each time step $t = 0, 1, 2, 3, ..., T$, the agent perceives the partial or complete state of the environment $s_t \in \mathcal{S}$ and must take an action $a_t \in \mathcal{A} = \{1, ..., |\mathcal{A}|\}$, receiving a single reward $r_{t+1} \in \mathcal{R} \subset \mathbb{R}$ and modifying the environment into its next state s_{t+1}, as shown in Figure 1.

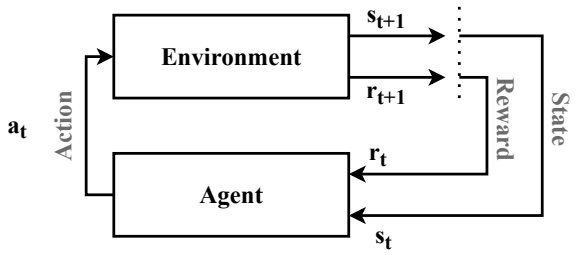


Figure 1: Reinforcement learning, agent-environment interaction scheme.

An agent-environment interaction results into a sequence of events (trajectory) in the form:

$$s_0, a_0, r_1, s_1, a_1, r_2, s_2, a_2, ... \tag{9}$$

Whenever T, $\mathcal{A}$, and $\mathcal{R}$ are finite subsets, the learning framework can be formally represented and described as a finite Markov Decisions Process (MDP). In MDP, the entire system is characterized by the mapping function from the pair [state, action] into [next state, reward]. Formally, let $s', s \in \mathcal{S}$, $r \in \mathcal{R}$, and $a \in \mathcal{A}$, the probability p of transition from state s into s' by taking action a and receiving the reward r is written as:

$$p(s', r|s, a) \doteq P(S_{t+1} = s', R_{t+1} = r|S_t = s, A_t = a) \qquad , \qquad \sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{R}} p(s', r|s, a) = 1 \tag{10}$$

where p represents the dynamic of the entire MDP system. Therefore, the probability of choosing an arbitrary action a depends only on the current state s and not on previous states.

Similar to the value estimation of an arbitrary action a in the non-associate task of Equation 7, we need to estimate the value of taking an arbitrary action a at any state s, denoted as $q(s, a)$. The cumulative reward G_t, at time step t, can be expressed from Equation 9 into Equation 11 as:

$$G_t = r_{t+1} + r_{t+2} + \cdots + r_T \tag{11}$$

with T as the final step. The previous formula works on finite oriented tasks in which the order of rewards is irrelevant and the agent-environment interaction sequence is finite. In order to make Equation 11 suitable for continuous oriented tasks or when the order of rewards matters, a discounted factor $\gamma \in [0, 1[$ is introduced, such that the cumulative discounted reward is computed as:

$$G_t \approx r_{t+1} + \gamma \cdot r_{t+2} + \gamma^2 \cdot r_{t+3} + \gamma^3 \cdot r_{t+4} + ... + \gamma^{T-t-1} \cdot r_T = \sum_{k=0}^{T} \gamma^k \cdot r_{t+k+1} \tag{12}$$

From now on, we assume $T \to \infty$ without loss of generality. As the sequence of rewards depends on the sequence of actions taken over the sequence of states, we look for an estimator of the pair state-action, in terms of future rewards, to guide the agent. The agent acting behaviour on the environment is referred as the agent's *policy*.

Let $\pi(a|s)$ be the probability of chosen action a at the state s under the agent's policy π. The state-value function, $v_\pi(s)$, represents the expected total reward of the state s for policy π, and is formally expressed as:

$$v_\pi(s) = \mathbb{E}_\pi\big[G_t \big| s_t = s\big] \approx \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k \cdot r_{t+k+1} \Big| s_t = s\right] \tag{13}$$

We define $q_\pi(s, a)$ as the action-value function, corresponding to the expected return of taking action a at state s under the policy π at time step t, and then following the same policy. It is computed as:

$$q_\pi(s, a) = \mathbb{E}_\pi\big[G_t \big| a_t = a, s_t = s\big] \approx \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k \cdot r_{t+k+1} \Big| a_t = a, s_t = s\right] \tag{14}$$

Both, Equation 13 and Equation 14 can be estimated by previous experiences, similar to the simple average-method described earlier, when the number of states and actions are small enough to be stored and retrieved. When the space of action and/or state makes the store-and-retrieve

process inefficient, v_π and q_π can be parametrized, reducing the number of parameters describing the functions. When the space of actions and/or space become unmanageable or incomplete during the process of learning, the use of deep neural network architectures serves to map State-Actions with NextState-Rewards. The latter is referred as deep Q-Learning, and we elaborate on this concepts in the following section. Before that, we introduce the concept of Bellman equations by rewriting Equation 13 as:

$$\begin{aligned} v_\pi(s) &= \mathbb{E}_\pi\big[r_{t+1} + \gamma \cdot G_{t+1} \big| s_t = s\big] \\ v_\pi(s) &= \sum_{a\in\mathcal{A}} \pi(a|s) \sum_{s'\in\mathcal{S}} \sum_{r\in\mathcal{R}} p(s', r|s, a)\Big[r + \gamma \cdot \mathbb{E}_\pi\big[G_{t+1}\big|s_{t+1} = s'\big]\Big] \\ v_\pi(s) &= \sum_{a\in\mathcal{A}} \pi(a|s) \sum_{s'\in\mathcal{S}} \sum_{r\in\mathcal{R}} p(s', r|s, a)\Big[r + \gamma \cdot v_\pi(s')\Big] \end{aligned} \tag{15}$$

The last expression translates into weighting each possible future response $[r + \gamma \cdot v_\pi(s')]$ by their probabilities of occurrence $\pi(a|s)p(s', r|s, a)$. In other words, it represents the value of a state as a function of the possible immediate rewards and value states.

Similarly, Equation 14 can be rewritten as:

$$q_\pi(s, a) = \mathbb{E}_\pi\big[r_{t+1} + \gamma \cdot G_{t+1} \big| a_t = a, s_t = s\big] \approx \mathbb{E}_\pi\left[r_{t+1} + \gamma \cdot \sum_{k=0}^{\infty} \gamma^k \cdot r_{(t+1)+k+1} \Big| a_t = a, s_t = s\right] \tag{16}$$

and by the principles of the Bellman equation, restated as:

$$q_\pi(s, a) = \mathbb{E}_\pi\big[r_{t+1} + \gamma \cdot q_\pi(s_{t+1}, a_{t+1}) \big| a_t = a, s_t = s\big] \tag{17}$$

We have from Equation 17 that the state-action value $q_\pi(s, a)$ can be decomposed into the immediate reward r_{t+1} of taking action a on the state s at time step t plus the discounted state-action value function $q_\pi(s_{t+1}, a_{t+1})$ at the next time step $t + 1$, recursively.

3. Deep Q-Learning

The Q-Learning technique was proposed by Watkins and Dayan (1992) as a simple approach for learning by successively improving the assessment of particular actions at particular states. The action-value function in Q-Learning is updated according to the expression:

$$q(s_t, a_t) \leftarrow (1 - \varepsilon) \cdot q(s_t, a_t) + \varepsilon \cdot \big[r_{t+1} + \gamma \cdot \max_{a_{t+1}} q(s_{t+1}, a_{t+1})\big] \tag{18}$$

where $\gamma \in [0, 1[$ and $\varepsilon \in [0, 1]$ are the discount factor and learning coefficient. When $\varepsilon : 1$ the action-value function is updated according to the received reward and discounted maximum action-value at the next state. When $\varepsilon : 0$ the action-value is not updated. This resembles the trade-off between exploration and exploitation. Indeed, we define an iteration as the moment when the agent has interacted with the environment through the entire time period or until the interaction has ended. Then, let ε_i be the epsilon value at the ith iteration, the ε_{decay} parameter controls the rate between exploration and exploitation as the training progresses, as $\varepsilon_{i+1} \leftarrow \varepsilon_i \cdot \varepsilon_{decay}$. Figure 2 illustrates the effects on ε_i by using $\varepsilon_{decay} : 0.99$

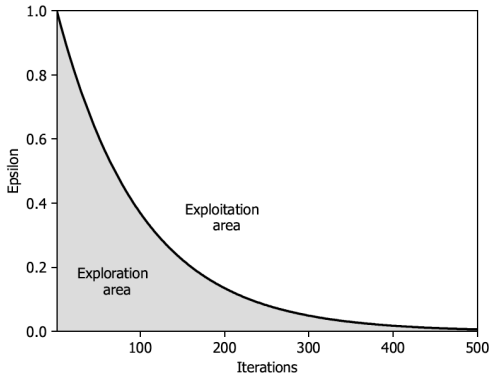


Figure 2: Epsilon greedy method. Exploration - exploitation dilemma.

The previous formulation requires to build a look-up table of size $\mathcal{S} \times \mathcal{A} \times T$ for all possible states (discrete), actions and time steps. Thus, the applicability of Q-Learning techniques has been constrained by computational power and limited to low-dimensional state and action spaces. In response, Mnih et al. (2015) proposed the Deep Q-Learning framework in which a deep neural network is trained to approximate the function action-value function $q(s_t, a_t)$. The latter directly extends the state space from a discrete to a continuous space. An extended theoretical and statistical analysis can be found at Fan et al. (2020). In the following, we focus on the main principles of Deep Q-Learning. We rewrite Equation 17 as function of states-and-actions as:

$$q_\pi(s_t, a_t) = \mathbb{E}_\pi\Big[r(s_t, a_t) + \gamma \cdot \sum_{s_{t+1} \in \mathcal{S}} p(s_{t+1}|s_t, a_t) \max_{a_{t+1}} q_\pi(s_{t+1}, a_{t+1})\Big] \tag{19}$$

Solving Equation 19, the optimal policy π corresponds to:

$$\pi(s) = \arg\max_{a \in \mathcal{A}} q(s, a) \tag{20}$$

The action-value function $q(s, a)$ is approximated by a deep neural network (DQN) that outputs a set of action-values of the form $q_\theta(s, \cdot)$ where θ corresponds to the set of neural network parameters. The use of a DQN allows the implementation of two tricks that accelerate the learning stage: replay memory and target network.

Let $\mathcal{M}$ be the set of experiences (memory set) stored in the form $(s_t, a_t, r_{t+1}, s_{t+1})$. Let $q_{\theta^*}(s, a)$ be a target network. The learning process starts with an empty memory set $\mathcal{M} = \emptyset$, random weights on the network parameter θ, and initial state s_0. The weights of the target network are initialized as $\theta^* = \theta$. At each iteration ith, we start from $t = 0$ until the interaction agent-environment ends. At each time t, the following steps are carried out:

1. With probability ε_i a random action is selected, and with probability $(1 - \varepsilon_i)$ the action is selected by $\arg\max_{a_t \in \mathcal{A}} q_\theta(s_t, a_t)$.
2. Once the action is executed, the immediate reward and the new state are stored in $\mathcal{M}$.
3. Randomly draw n transition samples from $\mathcal{M}$: $\{(s_j, a_j, r_j, s'_j)\}_{j \in [n]}$
4. For each sample, compute the target value $y_j = r_j + \gamma \cdot \arg\max_{a \in \mathcal{A}} q_{\theta^*}(s'_j, a)$

5. Update the network parameters using an optimization algorithm with a temporal learning rate α_t. For instance, using the gradient descent method:

$$\theta \leftarrow \theta - \alpha_t \cdot \frac{1}{n} \sum_{j=1}^{n} [y_j - q_\theta(s_j, a_j)] \cdot \nabla_\theta q_\theta \tag{21}$$

6. Every τ time steps, update the target network parameters as $\theta^* \leftarrow \theta$.

As a result, after training, the optimal policy $\pi_\theta(s)$ with respect to $q_\theta(s, a)$ is obtained as:

$$\pi_\theta(s) = \arg\max_{a \in \mathcal{A}} q_\theta(s, a) \tag{22}$$

4. Final remarks

We have covered the building blocks of reinforcement learning but many aspect have been left aside for simplicity, such as temporal difference learning, On-policy and Off-policy, SARSA, Monte Carlo tree search, exhaustive search, among others methods and principles. From the revised fundamentals, special attention is suggested on the following aspects when applying deep Q-Learning:

Environment representation To obtained valid, realistic and/or functional state-value and action-value functions, the environmental must be adequately represented in such a way that the agent is capable to interpret the differences between different states.

Reward The reward drives and guides the agent learning process. The reward must be in line with the long-term goal and must avoid pitfalls in which the agent would maximize the total reward without necessarily achieving the long-term goal.

Deep neural network The architecture of the network must be in accordance with the nature of the environment (spatial-temporal). Ensemble architectures would improve the capacity of approximating the action-value function in deep Q-Learning.

5. Acknowledgements

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2017-04200 and RGPAS-2017-507956.

6. Bibliography

Fan, J., Wang, Z., Xie, Y., Yang, Z., 2020. A theoretical analysis of deep Q-learning, in: Learning for Dynamics and Control, PMLR. pp. 486–489.

Ludvig, E.A., Bellemare, M.G., Pearson, K.G., 2011. A primer on reinforcement learning in the brain: Psychological, computational, and neural perspectives. Computational neuroscience for advancing artificial intelligence: Models, methods and applications , 111–144.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al., 2015. . nature 518, 529–533.

Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: An introduction. MIT press.

Watkins, C.J., Dayan, P., 1992. Q-learning. Machine learning 8, 279–292.

GEOMET
@QUEEN'S
UNIVERSITY

Understanding process performance with causal inference for continuous variables [1]

Sebastian Avalos (sebastian.avalos@queensu.ca)
Miguel Cuba (miguel.cuba@arcelormittal.com)
Melanie Bolduc (melanie.bolduc@arcelormittal.com)
Julian M Ortiz (julian.ortiz@queensu.ca)

Abstract

In mining operations, relations of cause and effect are not always clear and the presence of spurious correlations further confuses the analysis of causal inferences. In this article, we briefly review the principles of counterfactual analysis and the study of causal inference for continuous variables using the Kolmogorov-Smirnov test. The methodology is applied to study the impact of grade values and source proportions on a SAG mill energy consumption. The approach can be applied in any context, to understand the performance of a specific process as a function of the input variables, and draw causal relationships that can be validated with domain expertise.

1. Introduction

In mining operations, we often develop predictive models between a set of predictors to a single response or a set of responses. The quality of such models relies on the quality of the measured values of predictors and responses, and on the underlying relationships between them. It is well known that correlation does not necessarily mean causation, and the presence of spurious correlations further confuses the analysis of causal inferences. Thus, we can not rely on the predictive models to answer *what-if* questions related to causal effects, since the former exploit linear and non-linear correlations while the latter looks for counterfactual conditions.

The fields of counterfactual analysis and causal inference provide frameworks and principles to formulate causal problem from a statistical perspective. They have been applied in several other disciplines, such as pharmaceutical industry, sociological studies and epidemiology. It seems natural to transfer the previous frameworks into the mining context, extending the current tools of predictive modeling practitioners to deal with what-if questions.

In section 2 we present the fundamentals of counterfactual analysis. We focus the analysis to continuous variables. In section 3 we define a simple methodology to determine the presence or not of a cause-effect relation between the possible states of a response attribute conditioned to the state of a predictor attribute. The method is applied in a mining context in section 4 studying the impact on a SAG mill energy consumption of two grades and the proportion of three ore sources. A final discussion is presented in section 5.

[1] Cite as: Avalos S, Ortiz JM (2021) Understanding process performance with causal inference for continuous variables, Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report 2021, paper 2021-03, 22-32.

2. Background

Counterfactual theories of causation study conditional statements of the form *if A were true, then B would be true*, and any variations, such as *If A had not occurred, C would not have occurred* (Menzies and Beebee, 2020). The theoretical analysis of causation started with David Lewis's theory (Lewis, 1974). Since then, the field has expanded, refined and matured. Causal inference, or causal modeling, is the state-of-the-art branch of counterfactual theory, providing mathematical models and causal representations (Pearl et al., 2016).

The causal inference principle is that data alone is not enough to explain causality, and the story behind the data elements is required. The story is formally conveyed into a graph representation, with nodes representing the data elements, and edges the connections between them. The edges and their directions form a direct graph that represents the underlying story of the data. A detailed analysis on the graph structure and node's connections can be found at Pearl et al. (2016).

The data and its associated direct graph provide the framework to deal with what-if questions by means of *interventions* and *conditioning*. We intervene a variable when we fix its value. The edges flowing in the nodes are removed. This modifies the original graph, and often changes the value of subsequent nodes (Altinpinar and Ortiz, 2020). When conditioning on a variable, the graph does not change but we rather focus on a subset of the original data, satisfying the node condition. To draw reliable causal conclusions, the graph must be a valid representation of the studied phenomenon.

The effects of interventions are analyzed by means of probability theory and statistical metrics. Formulas such as Controlled Direct Effect (CDE), Average Causal Effect (ACE), and test of goodness-of-fit are some of the most used tools (Maldonado and Greenland, 2002). In the next section, we describe how to build the direct graph and a particular test of goodness-of-fit, in the context of mining and continuous variables.

3. Methodology

3.1. Causal model representation

The mining operational context and the expert knowledge on the expected relationship between variables must be reflected in the direct graph. First, all variables in the data must be created as nodes. If one or more unmeasured variables must be considered, additional nodes to the graph should be added. Then, connections must be drawn between nodes to represent the reality based on expert knowledge, process flowcharts, and/or feasible cause-effect relations. The graph representation and the applied interventions are the critical elements during causal analysis.

Assuming an adequate graph representation and that all variables have continuous values, we proceed to *conditioning* on the nodes and not *intervening* on the graph by removing them. The conditioning could be on a single variable, $\mathbf{x}$, or a set of them, $\mathbf{x}_1, \mathbf{x}_2, ...$, by means of inequalities over a specific value. It is reflected into conditional probabilities in the graph, such as $\mathcal{P}(\mathbf{x}_2 \geq 0.7)$, that translates into selecting a subset of the original data where $\mathbf{x}_2$ meets the condition. We study the impact of a variable $\mathbf{x}_1$ on a variable $\mathbf{x}_3$ when $\mathbf{x}_1$ meets a certain criteria conditioning to an additional variable $\mathbf{x}_2$. In other words, we study the difference between $\mathcal{P}(\mathbf{x}_3|\mathbf{x}_1)$ and $\mathcal{P}(\mathbf{x}_3|\mathbf{x}_1, \mathbf{x}_2)$. The principle can be extended to more than two variables at a time.

3.2. Two-sample Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test for goodness-of-fit (Massey Jr, 1951) is applied to the conditional cumulative distributions functions on $\mathcal{P}(\mathbf{x}_3|\mathbf{x}_1)$ and $\mathcal{P}(\mathbf{x}_3|\mathbf{x}_1, \mathbf{x}_2)$, represented as empirical distributions F_1, n and F_2, m, respectively, with n and m being the amount of samples in $\mathcal{P}(\mathbf{x}_3|\mathbf{x}_1)$ and $\mathcal{P}(\mathbf{x}_3|\mathbf{x}_1, \mathbf{x}_2)$, respectively. The Kolmogorov–Smirnov statistic is defined as:

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)| \tag{1}$$

The null hypothesis states that samples on F_1, n and F_2, m are drawn from the same global distribution. The null hypothesis is rejected at a level α when:

$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{n \cdot m}} \tag{2}$$

where $c\,(\alpha)$ is computed as:

$$c\,(\alpha) = \sqrt{-0.5 \cdot \ln\left(\frac{\alpha}{2}\right)} \cdot \tag{3}$$

Back into the causal analysis, whenever the null hypothesis is rejected between $\mathcal{P}(\mathbf{x}_3|\mathbf{x}_1)$ and $\mathcal{P}(\mathbf{x}_3|\mathbf{x}_1, \mathbf{x}_2)$, we say that $\mathbf{x}_2$ has a cause-effect on x_3, conditioned to x_1.

4. Case of study

4.1. Context

The previous methodology is applied in a simple open pit mining scenario. The run-of-mine ore is fed to the processing plant from three different sources: A (sA), B (sB), and C (sC). The ore is sent to a primary crusher, resulting in a blended material. This crushed blend is moved into a SAG mill, through a conveyor belt, where the ore is further blended and reduced in particle size. A contextual scheme is shown in Figure 1.

The SAG mill consumes a high amount of energy. As a driver for better short-term ore scheduling, the decision makers are interested in measuring the impact that grade 1 (g1) and grade 2 (g2), along with the proportion of sources in the blended ore, have on the SAG mill energy consumption (EC).

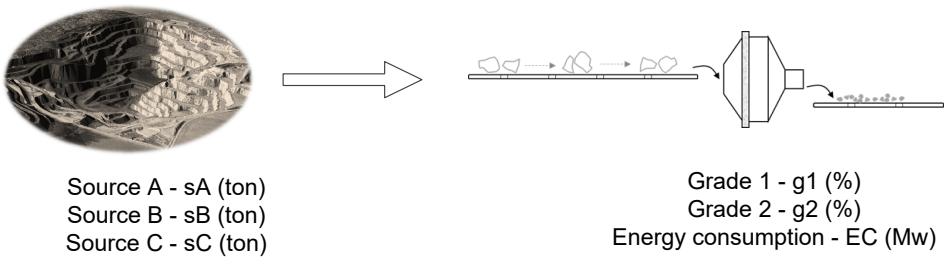


Figure 1: Contextual scheme and the location of measured data.

4.2. Database

A total of 250 daily measurements of the source tonnages (ton), grades (%), and energy consumption (Mw) have been collected. Table 1 displays the database main statistics. A visualization of the entire database time series is displayed in Figure 2. As we are interested in understanding the influence of source proportions and grades on the SAG mill energy consumption, scatter plots including their kernel density maps are shown in Figure 3. Note that we are not displaying the EC against the source proportion. No clear correlation is shown on the tonnage per source, but a slight positive correlation between EC and sC.

Figure 2: Time series visualization of sources, grades, and energy consumption.

Table 1: Database main statistics.

	sA (ton)	sB (ton)	sC (ton)	g1 (%)	g2 (%)	Energy consumption (Mw)
Min	2,479	0	0	0.26	0.24	6,000
Mean	25,968	33,744	33,048	6.98	10.59	34,161
Max	53,857	93,219	84,228	14.31	20.57	47,423
St Dev	10,792	16,429	20,160	2.55	3.56	6,940
Count	250	250	250	250	250	250

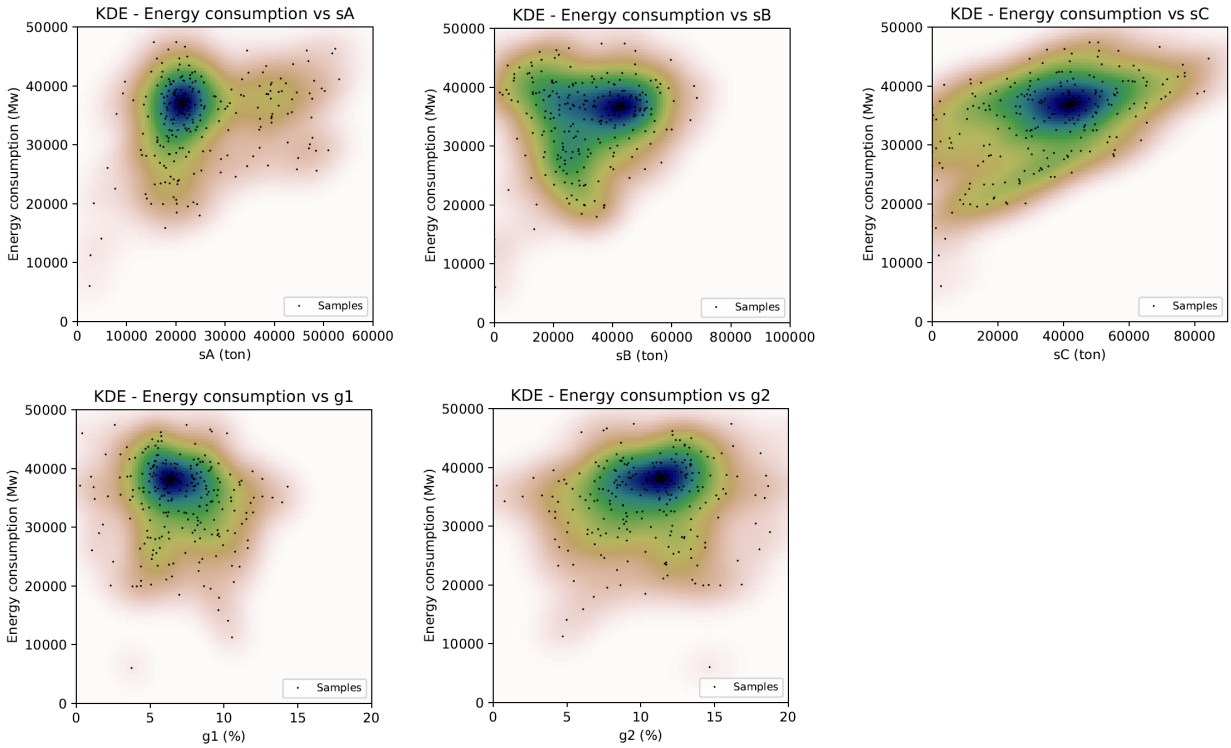


Figure 3: Scatter plots and kernel density maps of energy consumption against source tonnages and grades percentages.

4.3. Analysis

The contextual setting of the problem is translated into a causal model representation (Figure 4). Here, we interrogate the causal model about the influence on both grades and source proportion on the energy consumption.

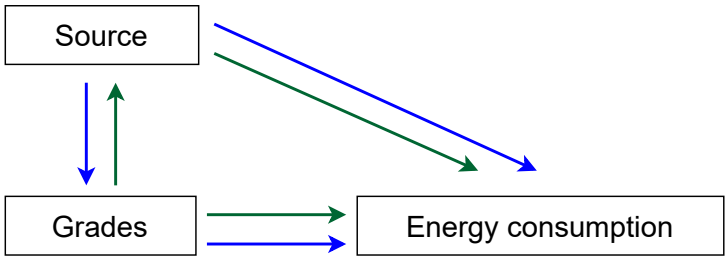


Figure 4: Causal model representation. Blue: impact of the grade on the energy consumption. Green: impact of the source proportion on the energy consumption.

Let pX be the proportion of a source in the blended material. For instance $p30$ of sA means that source A represents 30 % of the blended ore. Let pV be the V-percentile of a grade. For instance $p20$ of Grade 1 refers to the 20-percentile value of the Grade 1 distribution. Using the scheme of Figure 4 we have two possible pathways

Blue pathway Impact of grade on the energy consumption.

First, we select the subset of measurements in which a source is above a certain percentage. For instance, in vector representation we write - Source A above an X %, as $EC\,[sA > pX]$. We compute the conditional cumulative distribution function (ccdf) of the resulting subset, named ccdf_1.

Secondly, from the previous subset of measurements, we select an smaller subset of data with grade values above a certain percentile. For instance, in vector representation we write - Source A above an X % AND all Grade 1 above a percentile pV, as $EC\,[sA > pX, g1 > pV]$. We compute the ccdf of the resulting subset, named ccdf_2.

Green pathway Impact of the source proportion on the energy consumption.

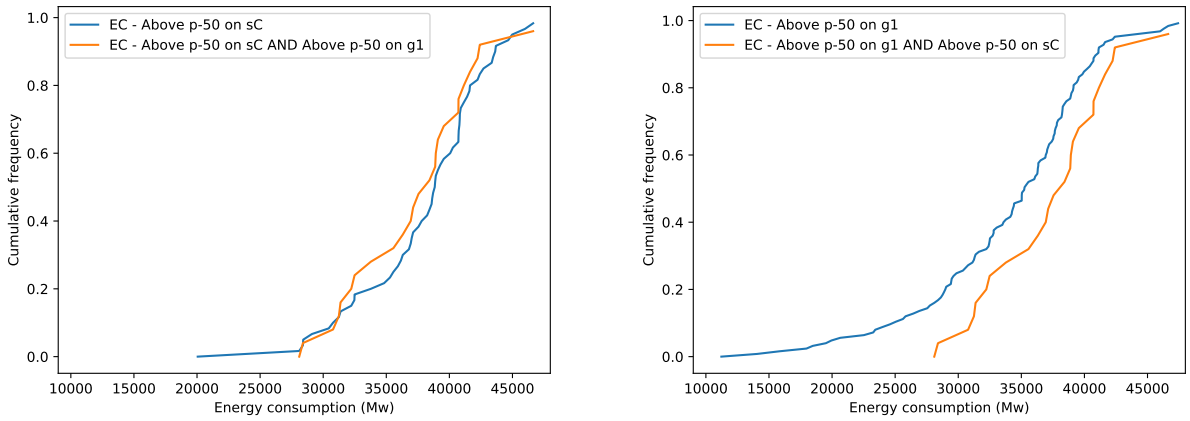
We start by selecting the measurements with grade values above a certain percentile. For instance, in vector representation we write - All Grade 1 above a percentile pV, as $EC\,[g1 > pV]$. We compute the ccdf of the resulting subset, named ccdf_1.

Secondly, from the previous subset of measurements, we select the smaller subset of data in which a source is above a certain percentage. For instance, in vector representation we write - All Grade 1 above a percentile pV AND Source A above an X %, as $EC\,[g1 > pV, sA > pX]$. We compute the ccdf of the resulting subset, named ccdf_2.

Then, regardless of the pathway, we compute the pValue of the two-sample Kolmogorov-Smirnov test for goodness-of-fit between ccdf_1 and ccdf_2. If the obtained pValue is above the critical value $\alpha : 0.05$, the null hypothesis is rejected, and therefore, the element (grade or source proportion) has an impact on the energy consumption. We won't compute the pValue if either the ccdf_1 or ccdf_2 have less than 10 samples.

4.4. Results

We begin by presenting the resulting ccdf_1 and ccdf_2 when analyzing both pathways (see Figure 5). The distributions are conditioned on sC above $p50$ (50% in blended ore) and g1 above $p50$ (grade values above the 50-percentile). When applying the K-S test, the null hypothesis is accepted on Figure 5a and rejected on Figure 5b. This means that the test indicates that in the former case, the values are drawn from the same distribution, while in the latter case, they come from different distributions. In other words, the grade does not have a significant effect on the energy consumption, while the source does.



(a) Impact of grade on the EC. *(b) Impact of the source proportion on the EC.*

Figure 5: Conditional cumulative distribution functions. Conditioning on sC above p50 and g1 above p50.

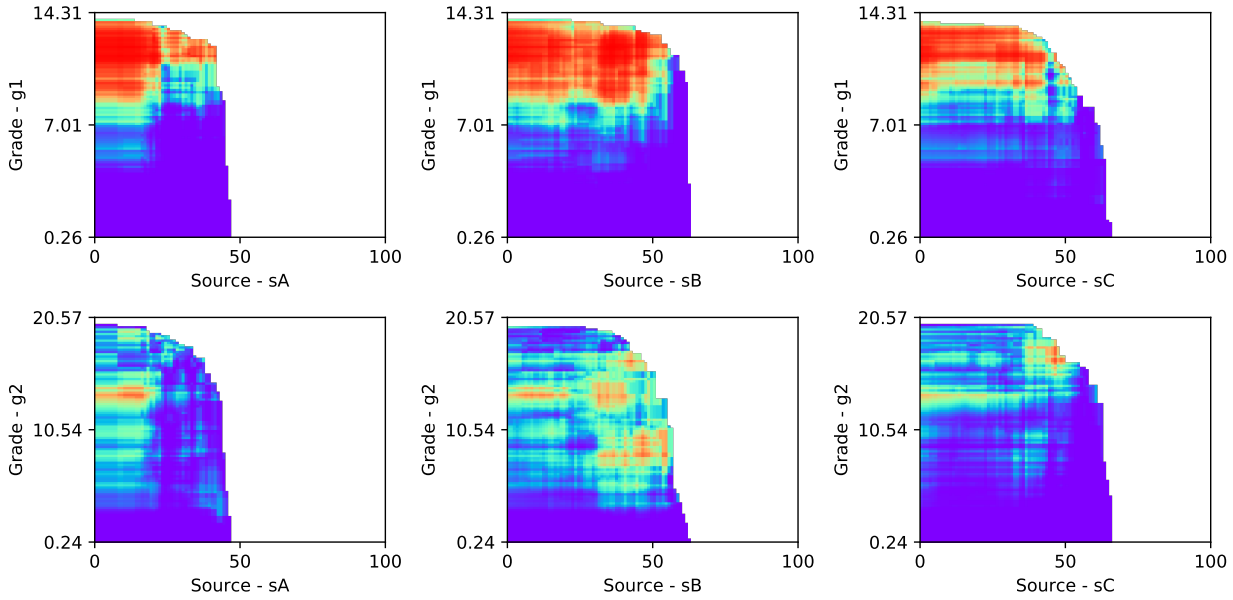


Figure 6: Impact of grade on the energy consumption. Grade 1 (top) and Grade 2 (bottom). Displaying the pValue of the two-sample Kolmogorov-Smirnov test.

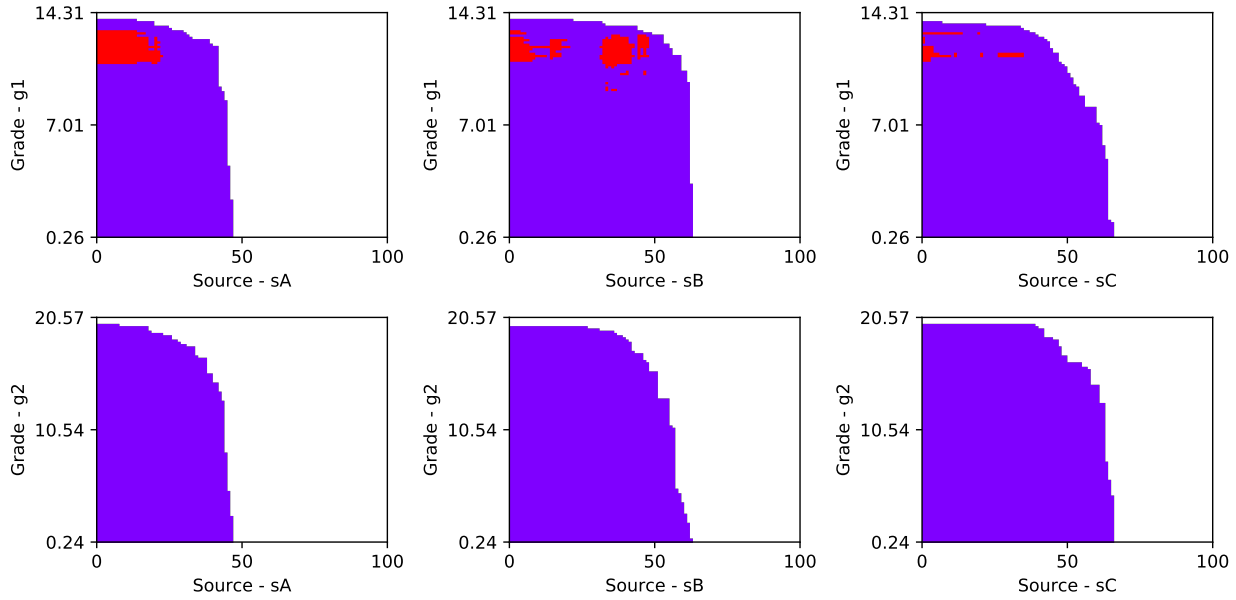


Figure 7: Impact of grade on the energy consumption. Grade 1 (top) and Grade 2 (bottom). Thresholding the pValue of the two-sample Kolmogorov-Smirnov test. Purple zone: accepting the null hypothesis. Red zone: rejecting the null hypothesis.

We extend the analysis to each grade and source. Grades are conditioned by percentiles while the source proportions goes from $[0, 1, ..., 99, 100]\%$. Figure 6 shows the pValue of the two-sample K-S test of the grade impact on the energy consumption for g1 (top) and g2 (bottom). The white areas correspond to insufficient amount of samples to compute the ccdf. By thresholding the maps with $\alpha : 0.05$ wherever the pValue is greater or equal to α, the Figure 7 is obtained.

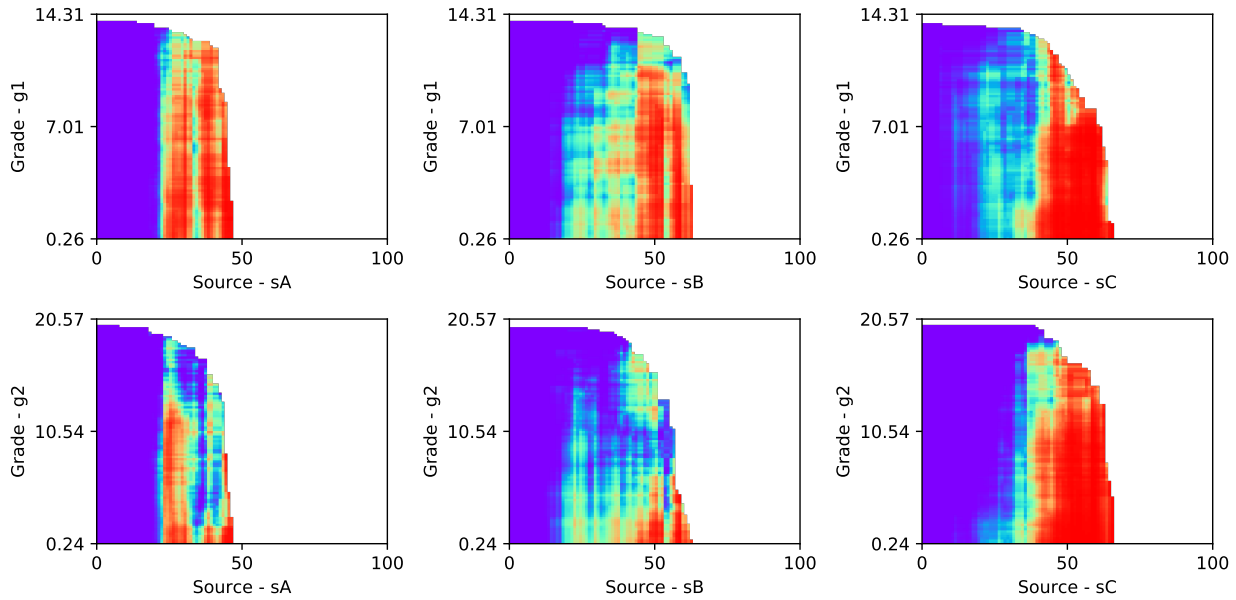


Figure 8: Impact of the source proportion on the energy consumption. Grade 1 (top) and Grade 2 (bottom). Displaying the pValue of the two-sample Kolmogorov-Smirnov test.

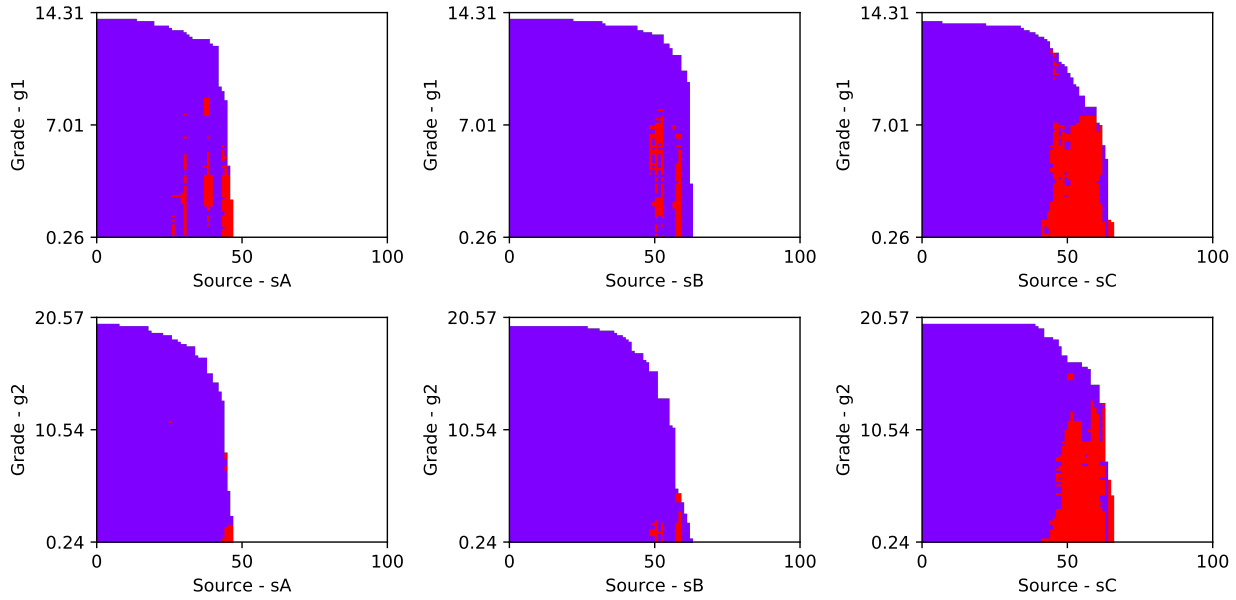


Figure 9: Impact of the source proportion on the energy consumption. Grade 1 (top) and Grade 2 (bottom). Thresholding the pValue of the two-sample Kolmogorov-Smirnov test. Purple zone: accepting the null hypothesis. Red zone: rejecting the null hypothesis.

The analysis is repeated on the impact of source proportion on the EC. Figure 8 shows the pValue of the corresponding two-sample K-S test for sA (left), sB (middle), and sC (right). The white areas correspond to insufficient amount of samples to compute the ccdf. By thresholding the maps with $\alpha : 0.05$ wherever the pValue is greater or equal to α, the Figure 9 is obtained.

4.5. Final takeaways

The maps of Figure 7 and Figure 9 summarize the *causalities* of grades and sources proportions, respectively. From them, the following key results are derived:

1. The Grade 1 (g1) **influences** the SAG mill energy consumption *if and only if* the value of g1 is above the 80-percentile (10.9 %) AND the proportion of sA or sB are below 24 % and 22 %, respectively, regardless of the proportion of sC.
2. The Grade 2 (g2) **has no influence** on the SAG mill energy consumption, regardless of the proportion of sources in the fed ore.
3. The proportion of Source A (sA) **has no influence** on the SAG mill energy consumption, regardless of the grade values of g1 and g2.
4. The proportion of Source B (sB) **has no influence** on the SAG mill energy consumption, regardless of the grade values of g1 and g2.
5. The proportion of Source C (sC) **influences** the SAG mill energy consumption *if and only if* the proportion is above 46 %, regardless of the grade values of g1 and g2.

The previous key results can be transferred into operational decisions either to avoid falling into areas where g1 or sC influence on the EC, or to expect EC variations when the conditions of points 1 and 5 are met.

5. Conclusions

Causal inference analysis has been widely applied in other disciplines, such as pharmaceutical industry and sociological studies. Transferring the theoretical background and acquired knowledge into mining operations is a fruitful area for applied research.

Data alone is not enough to explain causality. The story behind the variables is fundamental for the causal analysis. It translates into a graph representation. The graph representation and the interventions and/or conditioning must meet the real phenomenon and the what-if question being asked.

The graph representation can be as simple as the case study shown in the article or much more complex, when several processes and/or variables are considered. In addition, when testing the null hypothesis for goodness-of-fit, we have used the Kolmogorov–Smirnov test but other methods can be applied as well, such as Chi-squared test.

In the case study, and extrapolated to any causal inference analysis, causal models may indicate the impact on the energy consumption of a certain setting between grade value and source proportion but they do not describe if the change is positive or negative, meaning an increase or decrease in the consumed energy.

6. Acknowledgements

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2017-04200 and RGPAS-2017-507956. We also thank the support of ArcelorMittal Mining Canada G.P.

7. Bibliography

Altinpinar, M., Ortiz, J.M., 2020. Review of causal inference and modeling. Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report 2020, 130–146.

Lewis, D., 1974. Causation. The Journal of Philosophy 70, 556–567.

Maldonado, G., Greenland, S., 2002. Estimating causal effects. International Journal of Epidemiology 31, 422–429.

Massey Jr, F.J., 1951. The kolmogorov-smirnov test for goodness of fit. Journal of the American Statistical Association 46, 68–78.

Menzies, P., Beebee, H., 2020. Counterfactual Theories of Causation, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Winter 2020 ed.. Metaphysics Research Lab, Stanford University.

Pearl, J., Glymour, M., Jewell, N.P., 2016. Causal inference in statistics: A primer. John Wiley & Sons.

Queen's University

Annual Report 2021
Paper 2021-04

GEOMET
@QUEEN'S
UNIVERSITY

Notes on the Topology of Random Fields[1]

Alvaro I. Riquelme (alvaro.riquelme@queensu.ca)
Julian M. Ortiz (julian.ortiz@queensu.ca)

Abstract

The purpose of these notes is to give an understandable introduction to the topology of random fields. To this end we detail simple proofs when their understanding is deemed essential for geostatisticians, and we omit complex proofs that are too technical. Our principal effort will be in showing the main steps in the proof for the expression that gives us the expected value of the Euler characteristic, χ, of the excursion sets $Z^{-1}[u, +\infty)$, of a smooth and isotropic random field Z on $\mathbb{R}^N$ restricted to a sub-manifold M, for the case in which Z is a Gaussian field

$$\mathbb{E}\big[\chi\big(M \cap Z^{-1}[u, +\infty)\big)\big]$$

The principal motivation for this is the application of the concepts involved behind the study of the Euler characteristic, in particular, to applications such as the reconstruction of geological bodies. Figures will be employed in order to illustrate some of the concepts.

1. Introduction

1.1. Road Map

We begin by giving a very rough description of the path that we are going to follow. All the notions introduced here will be formalized in the later sections. Let us consider $Z = \{Z(x) : x \in D \subseteq \mathbb{R}^N, N \geq 1\}$ a Random Field (RF) defined on a fixed continuous domain of interest D of the Euclidean space $\mathbb{R}^N$. Here $\mathbb{R}$ denotes the set of all real numbers. The random field itself might be real or vector valued, that is, it can take values in $\mathbb{R}^k$, for any $k \geq 1$. We find this type of situations, for example, when we analyze any multi-element databases (Copper, Gold, Iron and more, in different locations of the space). We will analyze, however, the case $k = 1$. The domain of interest D that we want to tackle, as an introductory example, is a *connected* subset of $\mathbb{R}^2$. Let us take a square domain $D = [0, 20]^2$. In summary, the RF may be defined in all $\mathbb{R}^2$, but will be focused in the RF on $\mathbb{R}^2 \cap [0, 20]^2$ (Fig. 1, top), and the notation for this map is:

$$\begin{aligned} Z\colon [0,20]^2 \subset \mathbb{R}^2 &\to \mathbb{R} \\ x &\mapsto Z(x) \end{aligned}$$

[1]Cite as: Alvaro I. Riquelme , Julian M. Ortiz (2021) Notes on the Topology of Random Fields, Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report 2021, paper 2021-04, 33-45.

1.1.1. *Modeling Geology from RFs*

One of the common tricks used in geological modeling is to use the RF Z to obtain geological bodies (or lithofacies). This is done by looking at the *preimage* of a given target subset $T \subset \mathbb{R}$ of values in the image of the map given by Z, $\text{preim}_Z(T)$, where in this case

$$\text{preim}_Z(T) := \{x \in D \subseteq \mathbb{R}^N \mid Z(x) \in T\}.$$

Note that it is not accurate to define geological bodies by looking at the *inverse image* of a given set, since Z is not a bijection. However, since we are already aware of this subtlety, we are going to do some abuse of notation and call $Z^{-1}(T)$ to the $\text{preim}_Z(T)$.

We let T take different forms. If the interval is of the type $Z^{-1}[u, \infty)$, we call it the *level set* u, which is one of the most common image sets used (Fig. 1, bottom). Also T can take the bit more general form of an interval, $[a, b]$, with $a \leq b$, or in other cases, union (or finite intersection) $\bigcup_{j\in J}[a_j, b_j]$ ($\bigcap_{j\in J}[a_j, b_j]$) of ranges. Thus, we can build different geological bodies just by getting some $Z^{-1}[a, b)$.

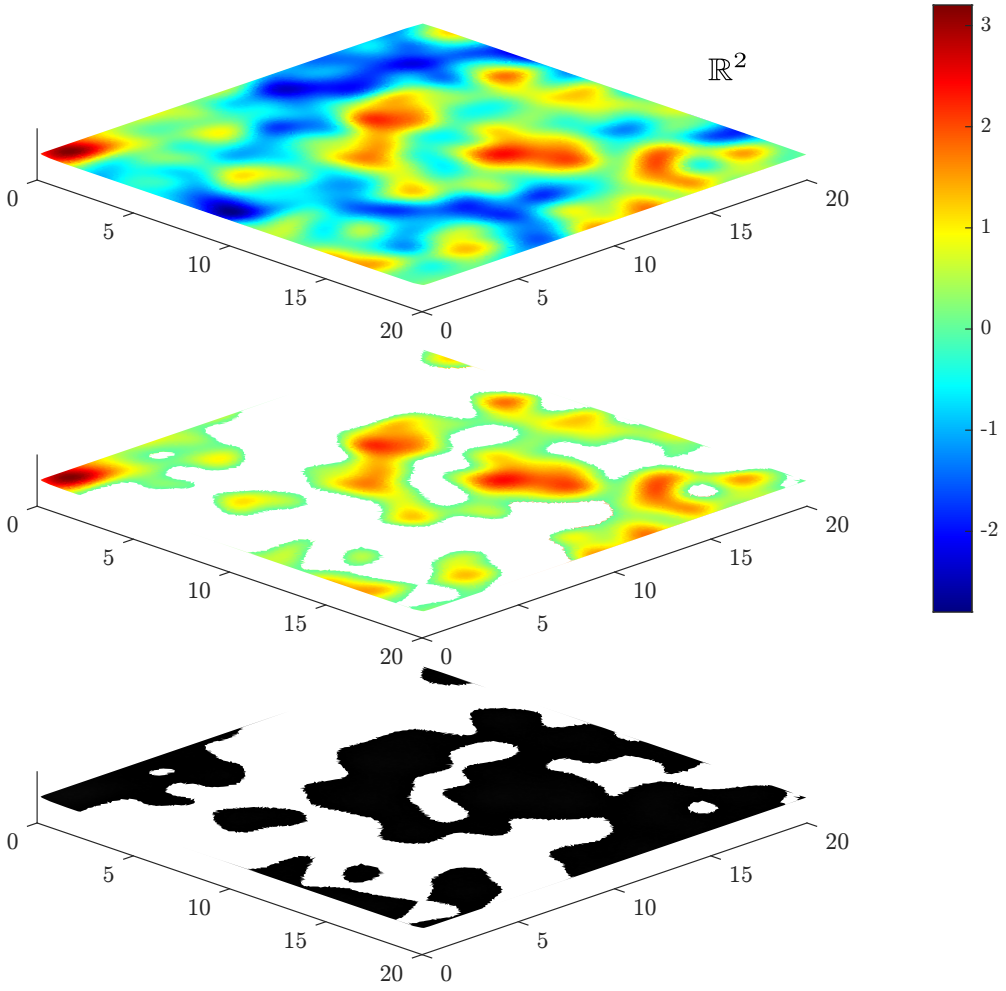


Figure 1: Example of a RF defined on $\mathbb{R}^2 \cap [0, 20]^2$ (top), and the process of defining a geological body by looking the set given by $\mathbb{R}^2 \cap \text{preim}_Z([0, +\infty))$.

1.1.2. Triangulations and the Euler Characteristic

We notice, in particular, that $Z^{-1}[u, \infty)$ is a planar surface, or more specifically, it is formed by the disconnected union of different planar surfaces, which may or may not possess holes inside. By changing the threshold from u to v, components of $Z^{-1}[u, \infty)$ may merge and new components may be born, and possibly later merge with another of the components of $Z^{-1}[u, \infty)$, changing as a result the geometry of the level set. We are interested in following these changes in the *topology* of these sets, as a function of u. In order to achieve this objective, we need a measure that gives us direct or indirect information about the number of disjoint component and holes in the set. For this, the next step will be to define a *triangulation* on the level set u.

In the mathematical language triangulations of a set are referred as *simplicial complex*, and they come with the notion of dimension within them, with each triangle called a *simplex* of dimension k. They can be open sets, or closed. Very roughly speaking, the linearly independent points of the simplex are called *vertices* and the simplices spanned by subsets vertices called the *faces*. A 0-dimensional face of a simplex is also called a *vertex*. A 1-dimensional face of a simplex is also called an *edge*.

Definition 1. *The standard* 0*-simplex is the point* $1 \in \mathbb{R}$ *This is also the standard open* 0*-simplex.*

Definition 2. *The standard 1-simplex is homeomorphic (a continuous bijection whose inverse is also continuous) to an interval. See Figs. 2 and 3.*

Definition 3. *The standard 2-simplex is a "triangle" with vertices at* $(1, 0, 0)$*,* $(0, 1, 0)$*, and* $(0, 0, 1)$*. Figure 4 depicts a* 2*-simplex.*

Definition 4. *The standard 3-simplex is a tetrahedron with vertices at* $(1, 0, 0, 0)$*,* $(0, 1, 0, 0)$*,* $(0, 0, 1, 0)$*, and* $(0, 0, 0, 1)$*. Figure 5 depicts a* 3*-simplex.*

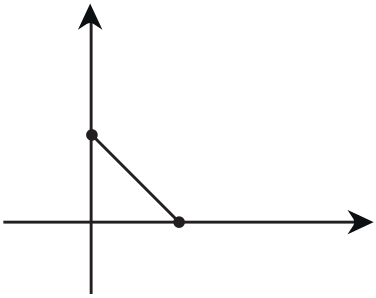


Figure 2: The standard 1*-simplex.*

Figure 3: A 1*-simplex.*

Definition 5. *A 0-simplex has only itself as a face.*

Definition 6. *The standard* 1*-simplex* $[v_0, v_1]$ *(and hence all* 1*-simplices) has itself as a 1-dimensional face. It also has two* 0*-dimensional faces,* $[v_0]$ *and* $[v_1]$*.*

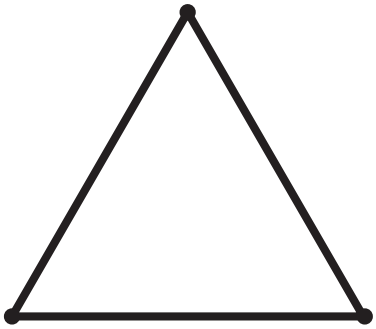


Figure 4: A 2-simplex.

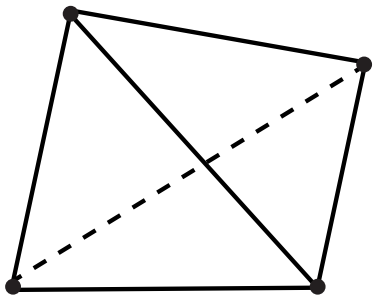


Figure 5: A 3-simplex.

Definition 7. *A 2-simplex has one 2-dimensional, three 1-dimensional, and three 0-dimensional faces.*

Definition 8. *A 3-simplex has one 3-dimensional, four 2-dimensional, six 1-dimensional, and four 0-dimensional faces.*

The notion of triangulation is formalized by defining the concept of *simplicial complex*, which is basically a map from the space of simplices to the topological target space, in our case the set $Z^{-1}[u, \infty)$. For examples of simplicial complexes, see Figs 6, 7, and 8.

The idea now is to cover the space $Z^{-1}[u, \infty)$ with a triangulation that approximately covers this set (Fig. 9). Then, one can proceed to compute the Euler Characteristic of our given set.

Definition 9. *The Euler characteristic of a finite simplicial complex K of dimension k is computed via the following formula:*

$$\chi(K) = \sum_{i=0}^{k} (-1)^i \#\{\text{simplices of dimension } i \text{ in K}\}$$

In the case of a surface or a 2-dimensional simplicial complex K, the Euler characteristic (EC) is given by $\chi(K) = V - E + F$, with vertices (V), edges (E), and faces (F). We notice that the EC takes values on the integers, i.e, $\chi(K) \in \mathbb{Z}$.

We will come back to this topics to cover more details and more general notions for Euler characteristic in the following sections, for instance, to show that the Euler characteristic is independent

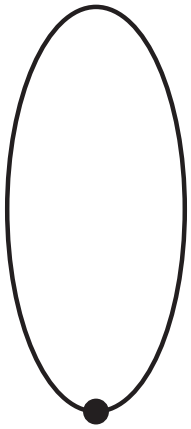


Figure 6: A 1-dimensional simplicial complex.

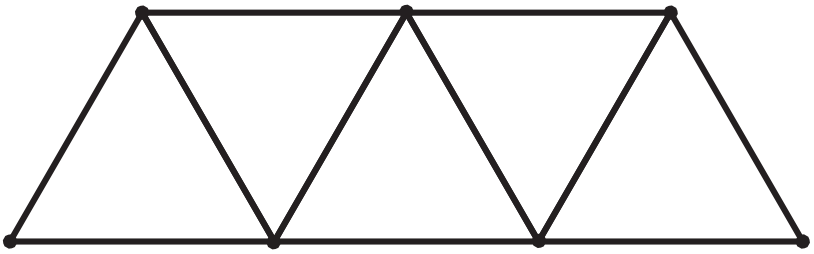


Figure 7: A 2-dimensional simplicial complex.

on the triangulation imposed to the set K, and also invariant under continuous deformations of the set.

Finally, taking several realizations of RFs Z, now we are able to compute

$$\mathbb{E}\left[\chi\left(D \cap Z^{-1}[u, +\infty)\right)\right]$$

1.1.3. Enters Morse Theory

Theoretical computations by counting the number of vertices, edges, and faces, although very intuitive and easy to follow, can not lead us too far. Indeed, it is a very primitive approach to the problem and we will have to leave it rapidly, specially since it is not clear yet how to relate directly the concept of triangulation and RFs. We will take in stead an indirect approach to the problem.

Morse Theory investigates how functions defined on a surface M (and *manifolds*, which are a generalization of the concept of surface, for higher dimensions) are related to geometric aspects of the surfaces. Surfaces are easy to visualize, and all the essential points of the theory readily appear in the case of surfaces.

1.1.3.1. Critical points of functions.

Let us consider a function $z = f(x)$ in one variable. We assume that both x and y are real numbers. A point x_0 which satisfies

$$f'(x_0) = 0$$

is called a *critical point* of the function f. The points at which f takes its maximum or minimum values, and the inflection point of $y = x^3$, are examples of critical points. The critical points of

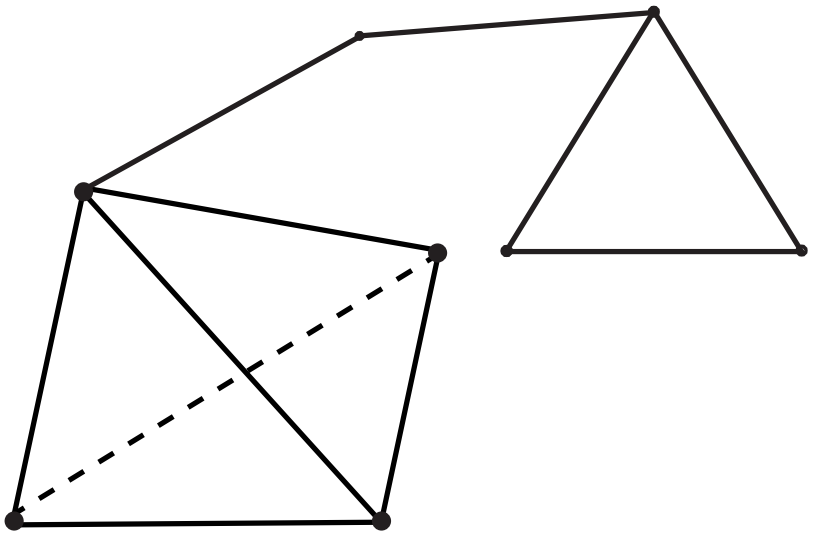


Figure 8: A 3-dimensional simplicial complex.

f fall in two classes according to the values of the second derivative of f, $f''(x_0)$. We call x_0 a *non-degenerate critical point* if $f''(x_0) \neq 0$, and a *degenerate critical point* if $f''(x_0) = 0$.

1.1.3.2. Hessian.
We now move to a real-valued function

$$z = f(x, y)$$

of two variables, where x and y are both real numbers. We may think of a pair (x, y) of real numbers as a point in the xy-plane. In this way f becomes a function defined on the plane, which assigns a real number to each point in the plane. We can visualize the graph of this function in the 3-dimensional space with three orthogonal axes x, y, $z = f(x, y)$.

Definition 10. *(Critical points of functions of two variables). We say that a point $p_0 = (x_0, y_0)$ in the xy-plane is a critical point of a function $z = f(x, y)$ if the following conditions hold:*

$$\frac{\partial f}{\partial x}(p_0) = 0, \qquad \frac{\partial f}{\partial y}(p_0) = 0.$$

We assume in this definition that the function $f(x, y)$ is of class C^∞ (differentiable to any desired degree). Such a function is also called a C^∞-function or a smooth function.

Example 1. *The origin $0 = (0, 0)$ is a critical point of each of the following three functions:*

$$z = x^2 + y^2, \qquad z = x^2 - y^2, \qquad z = x^2 - y^2$$

(see Fig. 11)

Now we need to define non-degenerate and degenerate critical points for functions of two variables. The reader may be tempted to define a critical point p_0 to be non-degenerate if it satisfies

$$\frac{\partial^2 f}{\partial x^2}(p_0) \neq 0, \qquad \frac{\partial^2 f}{\partial y^2}(p_0) \neq 0. \tag{1}$$

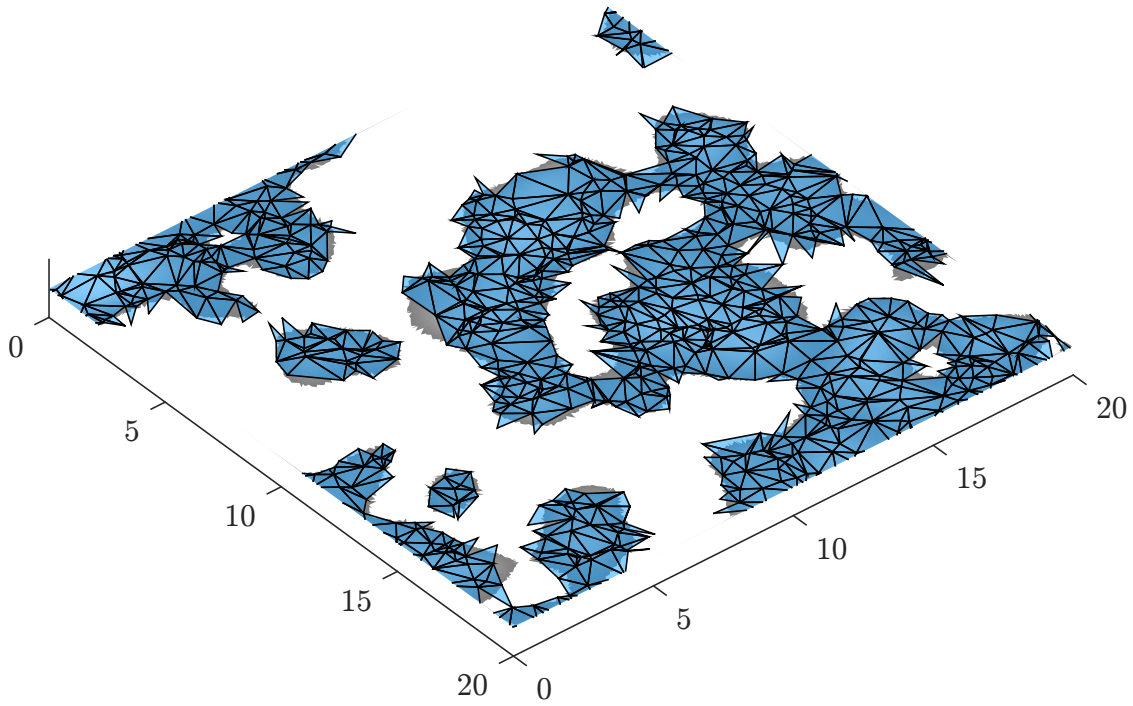


Figure 9: Example of triangulation defined on $\mathbb{R}^2 \cap \mathrm{preim}_Z([0,+\infty))$ *for a given RF* Z.

This is, in fact, a "bad definition", since after some coordinate changes, the condition (1) would no longer hold for the same f and p_0 in general. We want the concept of *non-degenerate critical points* or that of *degenerate critical points* to be independent of choice of coordinates. The following definition satisfies this requirement.

Definition 11. *(i) Suppose that* $p_0 = (x_0, y_0)$ *is a critical point of a function* $z = f(x, y)$. *We call the matrix*

$$\begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(p_0) & \dfrac{\partial^2 f}{\partial x \partial y}(p_0) \\ \dfrac{\partial^2 f}{\partial y \partial x}(p_0) & \dfrac{\partial^2 f}{\partial y^2}(p_0) \end{pmatrix},$$

of second derivatives evaluated at p_0, *the Hessian of the function* $z = f(x, y)$ *at a critical point* p_0, *and denote it by* $H_f(p_0)$.

(ii) A critical point p_0 *of a function* $z = f(x, y)$ *is non-degenerate if the determinant of the Hessian of* f *at* p_0 *is not zero; that is,* p_0 *is non-degenerate if we have the following:*

$$det\, H_f(p_0) = \frac{\partial^2 f}{\partial x^2}(p_0)\frac{\partial^2 f}{\partial y^2}(p_0) - \left(\frac{\partial^2 f}{\partial x \partial y}(p_0)\right)^2 \neq 0.$$

On the other hand, if $det\, H_f(p_0) = 0$, *we say that* p_0 *is a degenerate critical point.*

Notice that the matrix $H_f(p_0)$ *is a symmetrix matrix, since* $\dfrac{\partial^2 f}{\partial x \partial y}(p_0) = \dfrac{\partial^2 f}{\partial y \partial x}(p_0)$.

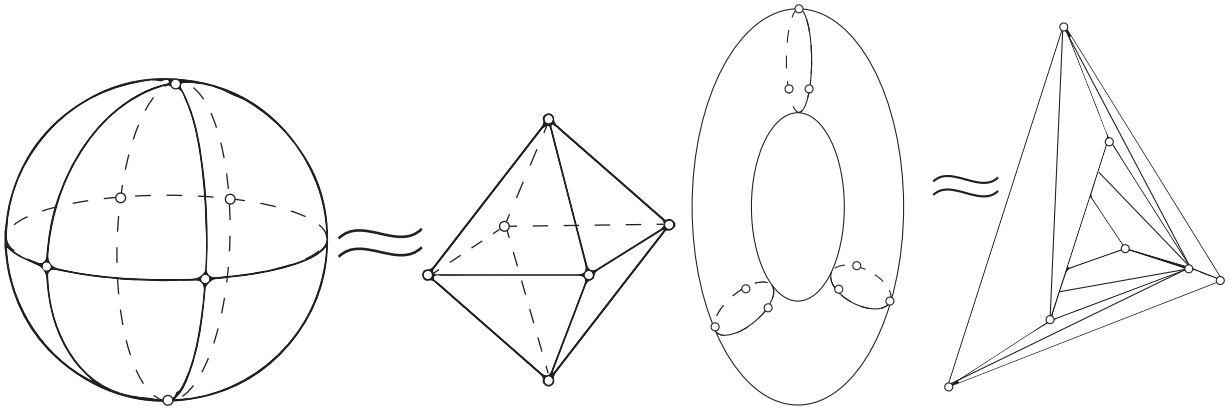


Figure 10: The EC of the two-dimensional sphere is $\chi(\mathbb{S}^2) = V - E + F = 6 - 12 + 8 = 2$ *and of the the two-dimensional torus* $\chi(\mathbb{T}^2) = V - E + F = 9 - 27 + 18 = 0$.

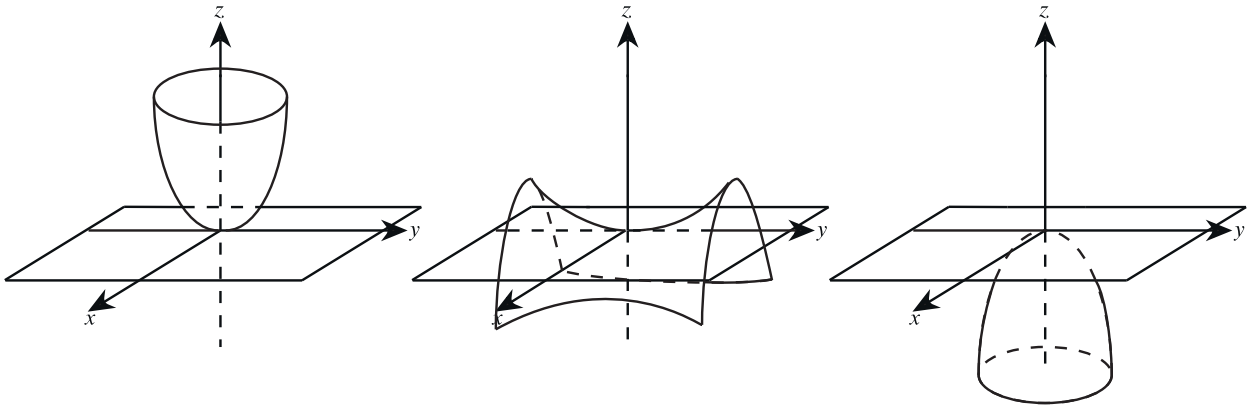


Figure 11: The graphs of $z = x^2 + y^2$, $z = x^2 - y^2$ *and* $z = x^2 - y^2$, *respectively from the left.*

Example 2. *Let us compute the Hessian for each of the three functions in Example 1 evaluated at the origin* 0.
(i) For $z = x^2 + y^2$, *the Hessian at the origin is*

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

(ii) For $z = x^2 - y^2$, *the Hessian at the origin is*

$$\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

(iii) For $z = -x^2 - y^2$, *the Hessian at the origin is*

$$\begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}.$$

The determinant of each of these matrices is not zero, and hence the origin 0 *is a non-degenerate*

critical point for each of the three functions.

Example 3. *Consider the function $z = xy$. The origin 0 is its critical point. The Hessian at 0 is*

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and its determinant is not zero; hence, the origin 0 is a non-degener.ate critical point. In fact, the function $z = xy$ is obtained from $z = x^2 + y^2$ by a coordinate change.

Example 4. *The origin 0 is a critical point of the function $z = x^2 + y^3$, but the Hessian of this function at 0 is*

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix},$$

whose determinant is zero. Thus 0 is a degenerate critical point of $z = x^2 + y^3$.

Now we proceed to state our first theorem.

Theorem 1. *(The Morse lemma) Let p_0 be a non-degenerate critical point of a function f of two variables. Then we can choose appropriate local coordinates (x', y') in such a way that the function f expressed with respect to (x', y') takes one of the following three standard forms:*

$$(i) \quad f = x'^2 + y'^2 + c \tag{2}$$

$$(ii) \quad f = x'^2 - y'^2 + c \tag{3}$$

$$(iii) \quad f = -x'^2 - y'^2 + c \tag{4}$$

where c is a constant $\big(c = f(p_0)\big)$ and p_0 is the origin $\big(p_0 = (0, 0)\big)$ in the new coordinates.

This theorem says that a function looks extremely simple near a non-degenerate critical point: for a function of two variables, a suitable coordinate change will make it one of the three simple functions we saw in Example 1.

Definition 12. *(Index of a non-degenerate critical point). Let p_0 be a non-degenerate critical point of a function f of two variables. We choose a suitable coordinate system (x, y) in some neighborhood of the point p_0 so that the function f has a standard form given by Theorem 1. Then we define the index of the non-degenerate critical point p_0 of f to be 0, 1 and 2, respectively for $f = x^2 + y^2 + c$, $f = x^2 - y^2 + c$ and $f = -x^2 - y^2 + c$. In other words, the number of minus signs in the standard form is the index of p_0.*

We see immediately from the respective graphs (Fig 11) of the functions $f = x^2 + y^2$, $f = x^2 - y^2$ and $f = -x^2 - y^2$ that if the point p_0 has index 0, then f takes a minimum value at p_0. If the index of p_0 is 1, then in some neighborhood of p_0, f may take values strictly larger than $f(p_0)$ or it may take values strictly smaller than $f(p_0)$. If the index of p_0 is 2, then f takes a maximum value at p_0. Thus the index of a non-degenerate critical point p_0 is determined by the behavior of f near p_0.

1.1.3.3. Morse functions on surfaces.
In the previous sections we limited ourselves to "local" investigation of critical points in their neighborhoods. We now turn to a "global" investigation which involves the shape of a space as a whole. In this section we consider two-dimensional spaces; that is, *surfaces*.

Some examples of closed surfaces were depicted in Fig. 10 and Fig. 12: a sphere and a torus in Fig. 10, and closed surfaces of genus two and three in Fig. 12. By the *genus* of a closed surface we mean the number of "holes" in it. The genus of a torus is one and that of a sphere is zero. We consider a closed surface of genus g for any natural number g. If one thinks of a torus as a "float", then one might think of a surface of genus two as a "float for two persons". Similarly we may think of a surface of genus g as a "float for g persons".

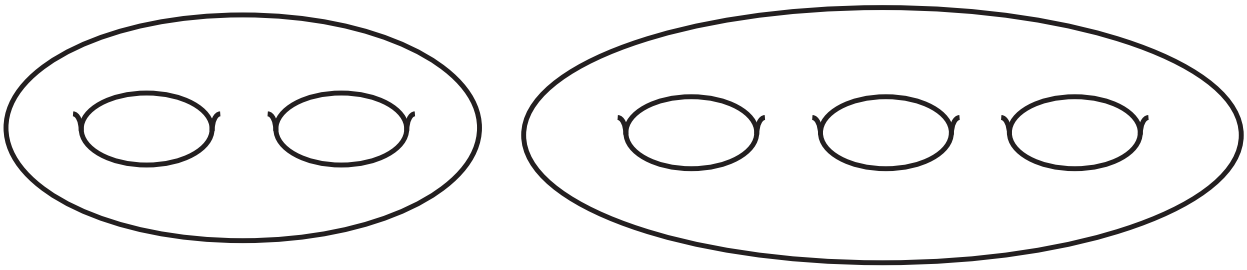


Figure 12: Closed surfaces of genus 2 and genus 3.

We denote the sphere by $\mathbb{S}^2$. The superscript 2 represents the dimension of the sphere. We denote a torus by $\mathbb{T}^2$. We often denote by Σ_g the closed surface of genus g, and in this case Σ_0 and Σ_1 are nothing but a sphere $\mathbb{S}^2$ and a torus $\mathbb{T}^2$, respectively.

Let M be a surface. We call a map

$$f : M \to \mathbb{R},$$

which assigns a real number to each point p of M, a function on M. Notice that a surface is curved, so that local coordinates on it are also curved in general (cf. Fig. 13).

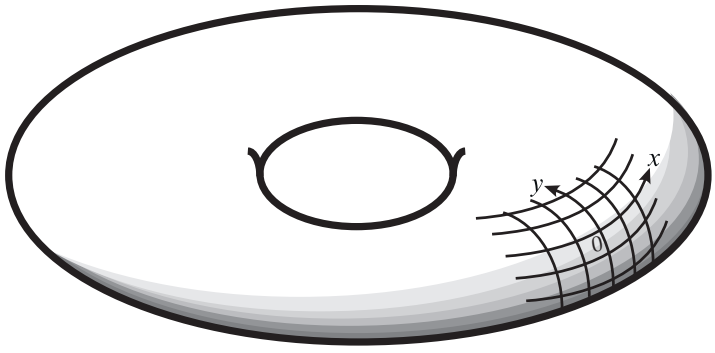


Figure 13: A local coordinate system on a surface.

We say that a function $f : M \to \mathbb{R}$ defined on a surface M is of class C^∞ (or smooth) if it is of class C^∞ with respect to any smooth local coordinates at each point of M. The concept of a "critical point" we saw in the previous section carries over to a function $f : M \to \mathbb{R}$ defined on a surface M with the aid of local coordinates. More precisely we say that a point p_0 of a surface M

is a *critical point* of a function $f : M \to \mathbb{R}$ if

$$\frac{\partial f}{\partial x}(p_0) = 0, \qquad \frac{\partial f}{\partial y}(p_0) = 0. \tag{5}$$

with respect to local coordinates in some neighborhood of p_0. We saw in the first section that non-degenerate critical points are stable and have some convenient properties in contrast to degenerate critical points. Therefore the functions on a surface with only non-degenerate critical points would be *nice* ones. Based on this consideration, we define

Definition 13. *(Morse functions). Suppose that every critical point of a function $f : M \to \mathbb{R}$ on M is non-degenerate. Then we say that f is a Morse function.*

1.1.3.4. Critical points and the EC.

Let us come back to the example on which the surface M is the sphere or the torus. We consider the unit sphere $\mathbb{S}^2$ with the orthogonal coordinates (x, y, z) in three-dimensional space $\mathbb{R}^3$; that is, $\mathbb{S}^2$ is defined by the equation

$$x^2 + y^2 + z^2 = 1.$$

Let

$$\begin{aligned} f \colon M \subset \mathbb{R}^3 &\to \mathbb{R} \\ (x, y, z) &\mapsto z \end{aligned}$$

be the nice function on $\mathbb{S}^2$ which assigns to each point $p = (x, y, z)$ on $\mathbb{S}^2$ its third coordinate z. One might say that f is the "height function". Then f is a Morse function, with two critical points; the north pole $p_0 = (0, 0, 1)$ and the south pole $q_0 = (0, 0, 1)$ (Fig. 14). One easily sees that f has no other critical points, and in fact, it is a lemma that for a Morse function defined on a closed surface has only a finite number of critical points.

Now we are ready to state one of the most striking results, and which relates the number of critical points and the EC.

Theorem 2. *(The Morse theorem) Let $f : M \to \mathbb{R}$ be a Morse function on a compact manifold M of dimension n, and denote by k_λ the number of critical points of f with index λ. Then the Euler characteristic of M is given by:*

$$\chi(M) = \sum_{\lambda=0}^{n} (-1)^\lambda k_\lambda \tag{6}$$

The intuition behind achieving this theorem, in a very rough and informal fashion, is that we can build our compact surface M by *attaching* our nice graphs function of Fig. 11 one by one, and using continuous deformations, until we get the surface M, in what is called a "handle decomposition". The graphs function of Fig. 11 are particular cases of "handles". During this process of construction, the important geometrical changes (the topological ones) appear when the continuous deformations are close to some critical point (Fig. 15).

In order to use this theorem, we will need to study under which conditions the RFs that we will making use of have the property of being Morse functions *with probability one*. This will be one

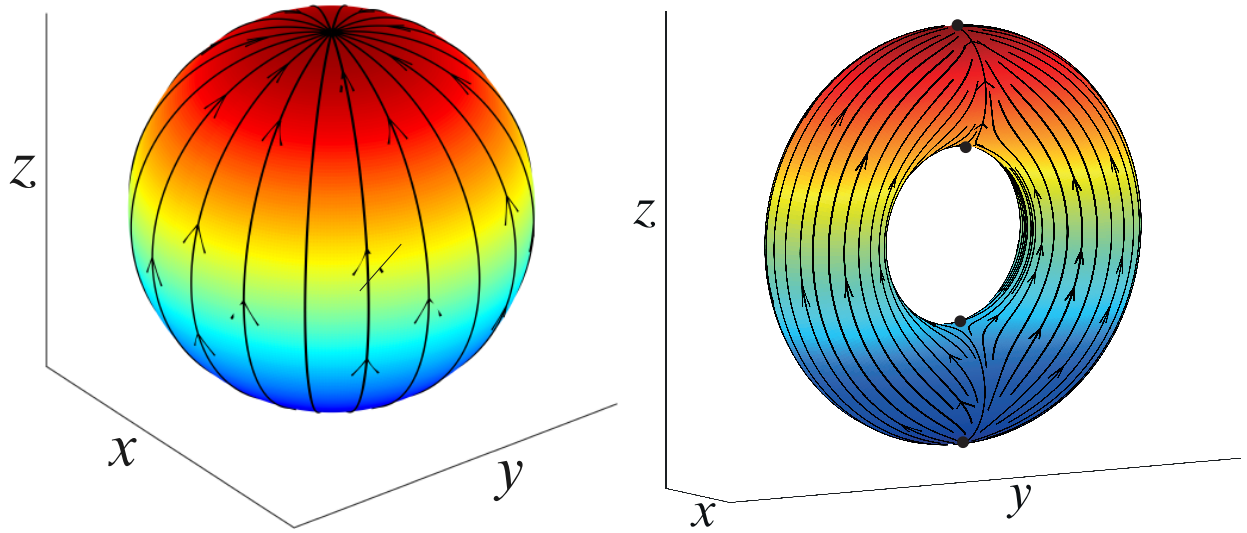


Figure 14: The height function and their gradient of the surface of $\mathbb{S}^2$ and $\mathbb{T}^2$. The sphere has 2 critical points (north and south pole): one of index 0 and another one of index 2. Therefore, $\chi(\mathbb{S}^2) = 1 \cdot 1 - 1 \cdot 0 + 1 \cdot 1 = 2$. The torus has 4 critical points: one of index 0, two of index 1, and another one of index 2. Therefore, $\chi(\mathbb{T}^2) = 1 \cdot 1 - 1 \cdot 2 + 1 \cdot 1 = 0$.

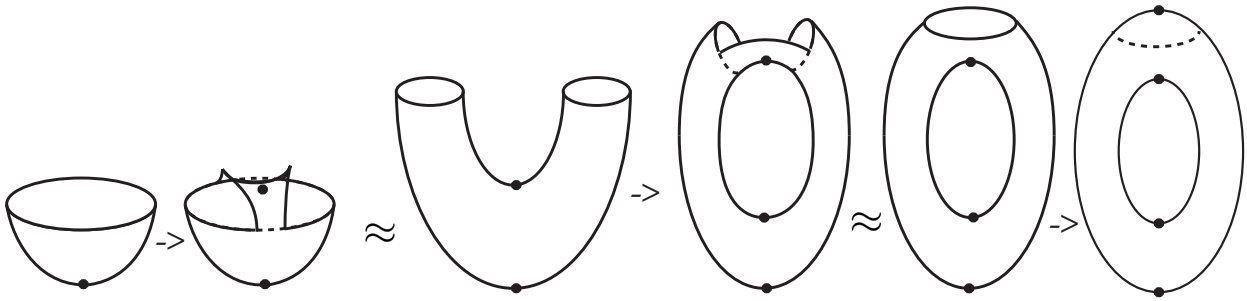


Figure 15: Building a torus by attaching "handles" to the previous ones and using continuous deformations.

of the technical issues that we must address on the second part these notes. Once this has been established, we will proceed with the following trick. We are going to use the value of the RF in our planar surface of Fig. 1 as both height and as a Morse function, i.e., $f = Z$.

After establishing the mentioned technical issues and others, the problem of finding the EC of the RF Z is basically reduced to find the number k_λ of points $p \in D$ for which:

(i) the RF Z at p is above the desired threshold u, i.e., $Z(p) \geq u$;

(ii) the RF Z at p is also a critical point, i.e.,

$$\frac{\partial Z}{\partial x}(p) = 0, \qquad \frac{\partial Z}{\partial y}(p) = 0;$$

(we summarize this condition by writing $dZ = 0$)

(iii) the index of the non-degenerate critical point p, for which we are going to see that is a function of the Hessian of Z (more precisely, equal to the number of negative eigenvalues of the Hessian of Z), i.e., the index$\big(H_Z(p)\big)$ is equal to λ.

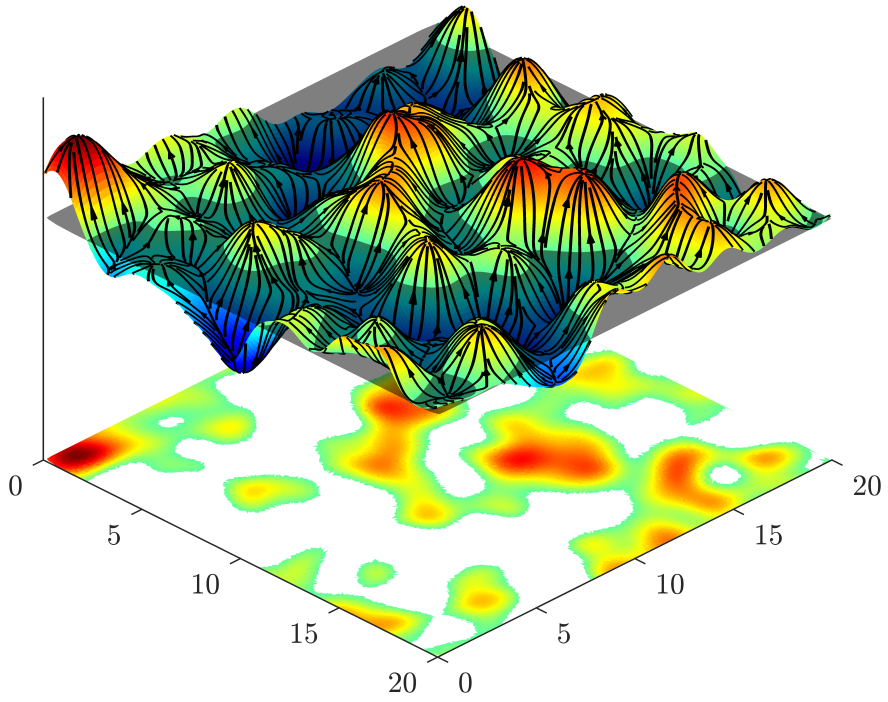


Figure 16: Schematic showing the trick used to find $\chi(Z^{-1}[u,+\infty))$: to use the RF Z as the height function in order to find critical points above the threshold u.

Then, we would like to use the Morse theorem for one realization of Z to find the value of the EC

$$\chi\big(Z^{-1}[u,+\infty)\big) = \sum_{\lambda=0}^{n}(-1)^{\lambda}\#\{Z > u, dZ = 0, \text{index}\big(H_Z\big) = \lambda\}.$$

Sadly, this will not be possible, but taking the expected value on both sides will be, in order to obtain first $\mathbb{E}\big[\chi\big(Z^{-1}[u,+\infty)\big)\big]$. The continuation of these notes is devoted to finding the expected values on the right hand side (RHS). By working a bit more, we are going to be able to find $\mathbb{E}\big[D \cap \chi\big(Z^{-1}[u,+\infty)\big)\big]$.

2. Acknowledgements

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2017-04200 and RGPAS-2017-507956.

Application of Disjunctive Kriging in Sequential Simulation[1]

David Casson (3drc1@queensu.ca)
Julian M. Ortiz (julian.ortiz@queensu.ca)

Abstract

Quantifying uncertainty in ore body modelling improves economic outcomes from mine development and operation. Simulation is a frequently utilized technique in geostatistics providing information on the level of uncertainty associated with estimated ore body properties at unsampled locations, both individually and in aggregate. Available simulation techniques do not allow for informed application of a random function model relying instead on: i) the assumption that the random function is MultiGaussian (as in sequential Gaussian simulation), ii) no assumption on the random function model (as in sequential indicator simulation), or iii) brute force inference of higher order statistics (multipoint simulation with machine learning). Disjunctive Kriging utilizes polynomial expansions of random function families to estimate parameters at unsampled locations. An algorithm has been developed to apply Disjunctive Kriging in sequential simulation using the Hermitian expansion of a Gaussian random function. Using the same conditioning information, simulation results closely match those from sequential Gaussian simulation. This suggests applying the algorithm with non Gaussian random function families and associated polynomial expansions will provide a valuable sequential simulation tool.

1. Introduction

1.1. Context in Mineral Resource Estimation

Traditionally, the qualified person completing a Mineral Resource estimate applied kriging to arrive at a best estimate for mineral resource grade and tonnage. Kriging is a deterministic method unable to provide a relevant measure of uncertainty associated with its deterministic estimates. The qualified person would classify the mineral resource as measured, indicated, or inferred based on personal experience and industry association guidance related to drill hole spacing and deposit type. For example, in a copper porphyry setting drills holes of a certain spacing might result in indicated resources, while drill holes at slightly greater spacing would be inferred. These specific spacings differ depending on the type of ore body and geological setting.

Simulation provides a quantification of uncertainty in resource modelling. This allows an estimator to arrive at a quantitative view of how probable it is that a given block has a given value (e.g., probability that grade of block is above cut off). This probabilistic estimate of uncertainty can be carried through the

[1] Cite as: Casson D., Ortiz J. M. (2021) Application of Disjunctive Kriging in Sequential Simulation, Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report 2021, paper 2021-05, 46-60.

mine design and operation to make better decisions, understanding the range of possible operational and financial outcomes and associated likelihood.

Simulation infers the properties of the regionalized variable under study based on the available data samples. The accuracy of the resulting simulations as a predictor of outcomes and probabilities depends entirely on how well the inferred model matches the true properties of the regionalized variable. Said simply, the appropriateness of the chosen model for the variable under study will dictate the accuracy and usefulness of simulation results.

1.2. Traditional Simulation Techniques

Commonly applied simulation techniques include Sequential Gaussian Simulation and Sequential Indicator Simulation. As is described further below, both these sequential simulation methods require a random draw from an estimated Conditional Cumulative Distribution Function ("CCDF") at unsampled locations. The CCDF, which is based on information in a neighborhood deemed to be relevant, reflects the probability that the unknown value at the location in question is below any value in its possible range. Simulation will not be successful if the CCDF is not accurate. Sequential Gaussian Simulation relies on the assumption that the regionalized variable under study is Multigaussian in nature. This is often not true. Simulation will also not be successful if it is not practical to implement. Sequential Indicator Simulation makes no assumption on the model of the variable under study but requires computation and modelling of many indicator variograms which makes it a cumbersome approach and challenging to implement in practice.

The assumption of MultiGaussianity provides the theoretical basis for many methods in geostatistics. The specific indicator properties of a multipoint (multivariate) Gaussian variable (i.e., indicator variogram) can be used to define a Gaussian function for the local CCDF. These Gaussian local distribution functions determine the probabilities of potential values at the unsampled locations, providing a measure of "local" uncertainty. Sequential Gaussian Simulation ("SGS") extends this approach to estimate global uncertainty amongst a population of unsampled locations, applying Monte-Carlo simulation along with the inferred local Gaussian CCDF's at unsampled locations to arrive at a simulated point value. Sequential simulation takes a random path through unsampled locations and once a simulated value is determined at a given location, it is assumed fixed and incorporated into the CCDF's for subsequently simulated unsampled locations. Repeating this sequential simulation provides multiple potential realizations of the unsampled locations creating, in aggregate, a measure of uncertainty of the entire unsampled set of locations. Such an approach is applied to create a measure of uncertainty related to grade and tonnage estimates, e.g., the probability across 1000 simulations that a given block is ore or waste.

The Gaussian approach described above is in contrast to the indicator approach applied in Sequential Indicator Simulation ("SIS") which makes no assumption regarding model function or parameters. While the indicator approach can account for non-symmetric distributions and/or non-diffusive distributions, it has several limitations described below (Machuca-Mory et al 2008):

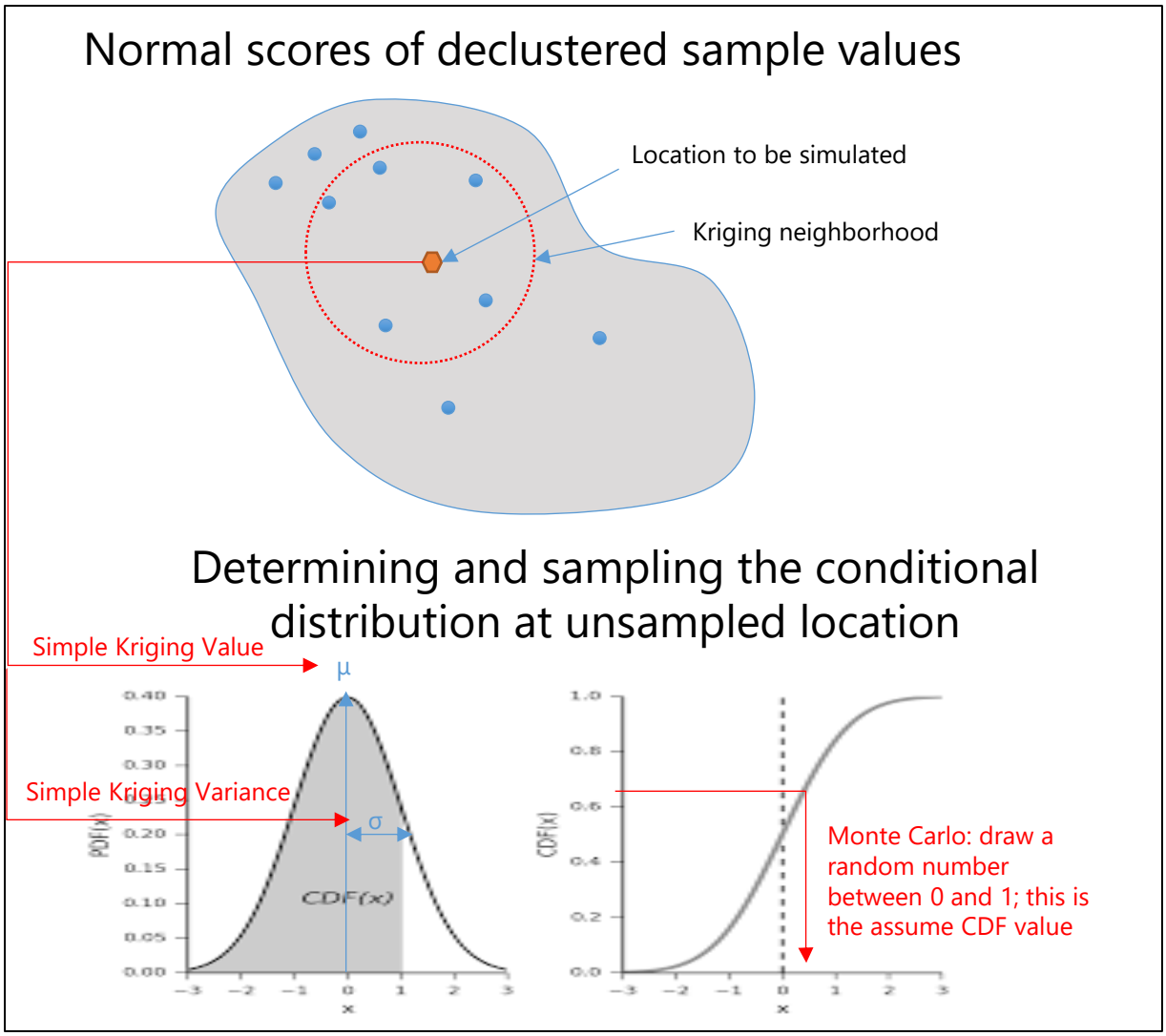
- Discretization and interpolation of the conditional CDF creates error.

- Cokriging of indicator thresholds (which is theoretically required) is not practical with a linear model of co-regionalization (challenging to limit cross and direct indicator variograms to linear combinations of specified families of models).
- Tail behavior is difficult to model given sparse relevant data points.

The steps to sequential Gaussian simulation are described below.

1. Decluster data to get representative distribution;
2. Transform data to normal scores, based on the representative distribution;
3. Visit nodes in a random path;
4. At every node, search for nearby samples or previously simulated nodes;
5. Krige the normal scores using samples and nodes (this requires the variogram model of the normal score transform of the original variable);
6. Draw a value from the conditional distribution;
7. Back transform the simulated value.

The figure below adapted from Ortiz (2019) illustrates the key steps in SGS simulation.

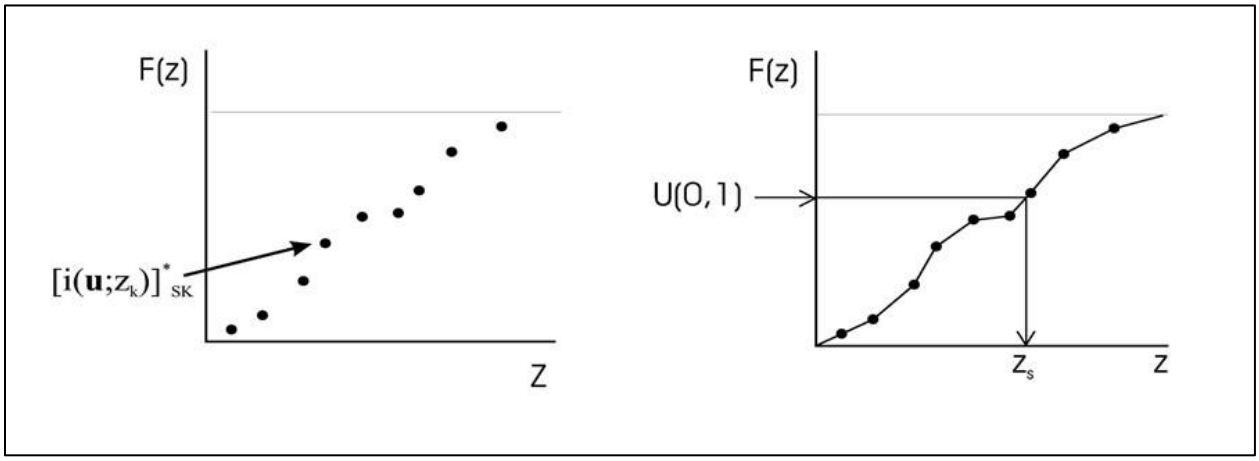


Sequential indicator simulation is non-parametric in nature. No assumptions are made regarding the bi-variate distribution.

The steps to sequential indicator simulation are described below.

1. Visit nodes in a random path;
2. At every node, search for nearby samples or previously simulated nodes;
3. Select a number of threshold values for the variable in question;
4. For each threshold:
    a. Code nearby known or simulated data points as "1" or "0" if they are below of above the threshold respectively;
    b. Apply simple kriging to these values (requires a variogram fit to the indicator data points);
    c. Result is probability that value at unsampled location is below the threshold;
5. Interpolate the threshold-probability data points into a CDF curve;
6. Draw a value from the conditional distribution.

The figure below adapted from Ortiz (2019) illustrates the interpolation approach to modelling the CCDF using indicator kriging.



1.3. Polynomial Expansions for Modelling Probability Density Functions and Cumulative Density Functions

Determination of local conditional Probability Density Functions ("PDF's") is an important component of modern geostatistics. Conditional PDF's (and corresponding Cumulative Distribution Functions or "CDF's") allow for simulation of an attribute at an unsampled location conditioned to the available and relevant data. The sequential simulation technique relies on a random draw from the CCDF at the unsampled location.

Asymptotic expansions are a mathematical technique for approximating a function with finite variance by a weighted average of a family of functions related to a "developing distribution function" or random function family. This type of mathematical formulation can be applied to model CDF's in situations where suitable developing distribution functions are selected.

In the first part of the twentieth century, significant research was completed on approximating empirical distributions with theoretical functions. In many cases, this work focused on the use of asymptotic expansions, in which the error of the approximation approaches zero as a parameter (the order) of the expansion approaches infinite (Wallace, 1958). The premise upon which these expansions are based is found in the Charlier Differential Series.

A random function is completely described by its moments. For any random function, the moments can be defined with the moment generating function shown below.

$$M_X(t) = E[e^{tX}] \quad (1)$$

In a similar manner to the Moment Generating Function, a random variable and its distribution can be defined by its Characteristic Function *"f(t)"* or sometimes *"φ(t)"*. The Characteristic Function is similar to the Moment Generating Function. It resembles a Fourier Transform (but in complex conjugate) of the underlying random variable's PDF. The expression is shown below.

$$\varphi_X(t) = E[ei^{tX}] \quad (2)$$

Cumulants are similar but not identical to moments. A random variable can be completely described by its cumulants or its moments. The cumulants are defined by the Cumulant Generating Function, which is the natural logarithm of the characteristic function. As a result, the characteristic functions can be written as the natural exponent of the cumulant generating function. This is shown below for a characteristic function "*f(t)*".

$$f(t) = e^{\left[\sum_{r=1}^{\infty} \kappa_r \frac{(it)^r}{r!}\right]} \quad (3)$$

Using the above expression for both an empirical characteristic function *f(t)* and a chosen theoretical characteristic function *ψ(t)*, an expression may be defined to relate the theoretical and empirical characteristic functions in terms of the difference between cumulants of each (κ_r and γ_r respectively) as shown below.

$$f(t) = e^{\left[\sum_{r=1}^{\infty} (\kappa_r - \gamma_r) \frac{(it)^r}{r!}\right]} \psi(t) \quad (4)$$

If the PDF of the chosen theoretical distribution (random function family) referred to here as *Ψ(x)* and all of its derivatives are continuous and have finite variance (i.e., vanish in the extremes but are continuous otherwise), integration by parts is possible and therefore expression of the two PDF's (theoretical and empirical) based on the difference between the cumulants is possible. This is shown below where "D" is the differential operator and comes from the integral relationship between probability distribution functions and associated characteristic functions.

$$F(x) = e^{\left[\sum_{r=1}^{\infty} (\kappa_r - \gamma_r) \frac{(-D)^r}{r!}\right]} \Psi(x) \quad (5)$$

By selecting a family of theoretical distributions with finite variance, we can create an expansion of the probability distribution function to be approximated based on the derivatives of the theoretical PDF (and associated CDF) selected and the differences between observed and theoretical cumulants. As noted by Mustapha and Dimitrakopolous (2010) "it is relatively easy in many statistical situations to determine moments, but it is extremely hard or impossible to determine the distributions themselves".

1.4. Review of Hermitian Polynomial Expansion

The Edgeworth Approximation is a specific case of an asymptotic expansion, where the theoretical distribution function is selected to be the normal distribution (which satisfies the criteria of finite variance). In choosing the normal distribution we are not assuming that the empirical distribution to be modelled is truly Gaussian, we are simply choosing the normal distribution to base our expansion on. If the empirical were perfectly Gaussian then no expansion would be needed to write the empirical PDF in terms of the theoretical Gaussian PDF. Intuitively the closer the empirical distribution is to the theoretical the "better" the expansion should be (i.e., how close the expansion approximates the theoretical distribution for any given truncation of the expansion series). Once the choice of the Gaussian distribution is made for the theoretical developing function, the relationship between the cumulants can be further defined based on the theoretical cumulants of the normal distribution. The mean and variance of the empirical and theoretical distribution functions are set equal resulting in an expansion based on the difference of the higher order cumulants. The higher order cumulants of the theoretical normal distribution are zero which further simplifies the equation to depend only on higher order cumulants of the empirical distribution. The empirical cumulative distribution function to be approximated can be written as an expansion of a polynomial and the derivatives of the normal distribution as shown in the equation below based on the observed cumulants of the empirical distribution. Here the λ values are derived from the observed cumulant values of each order in the empirical data set.

$$F_n(x) = \phi(x) - \frac{\lambda_3 \phi^3(x)}{-6\sqrt{n}} + \frac{1}{n}\left[\frac{\lambda_4 \phi^4(x)}{24} + \frac{\lambda_3^2 \phi^6(x)}{72}\right] + \cdots \qquad (6)$$

Given the unique relationship between the normal distribution and its derivatives, and its well-known representation in Hermite Polynomials, the Edgeworth Approximation can be rewritten as a function of Hermite Polynomials, $h_{(n)}(x)$, which is shown below for an expansion up to order 6.

$$e_4(x) = \phi(x)\left[1 + \frac{\kappa_3 h^3(x)}{6\sqrt{n}} + \frac{\kappa_4 h^4(x)}{24n} + \frac{\kappa_3^2 h^6(x)}{72n}\right] \qquad (7)$$

The Hermite Polynomial of order n, $h_{(n)}(y)$, is defined in the below equation where $g(y)$ is the standard gaussian pdf function.

$$H_n(y) = \frac{1}{\sqrt{n!}\, g(y)} \frac{d^n g(y)}{dy^n} \qquad \forall n \geq 0 \qquad (8)$$

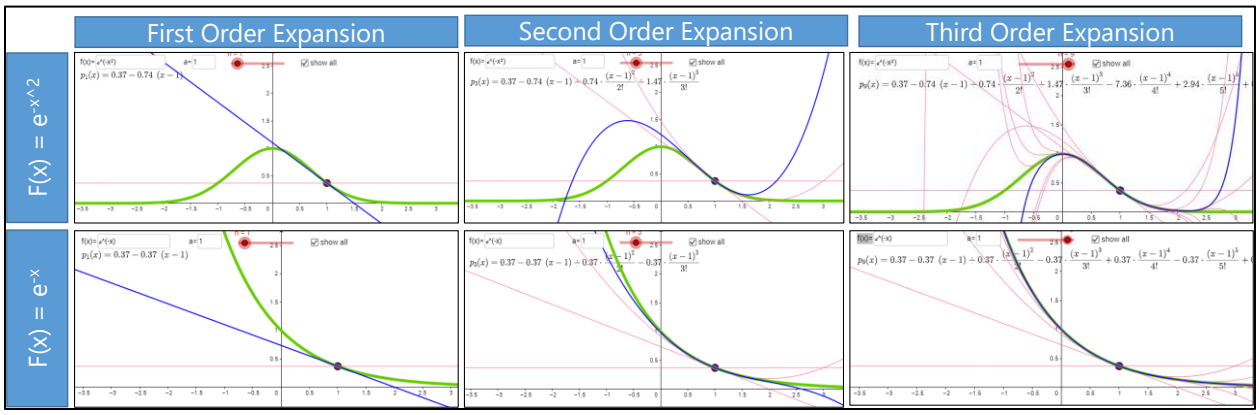
The expression for calculating Hermite Polynomials can be simplified to a recursive formula as shown below, called Rodrigues Formula, where $H_0 = 1$ and $H_1 = -y$.

$$H_{n+1}(y) = -\frac{1}{\sqrt{n+1}} y H_n(y) - \sqrt{\frac{n}{n+1}} H_{n-1}(y) \qquad \forall n \geq 0 \qquad (9)$$

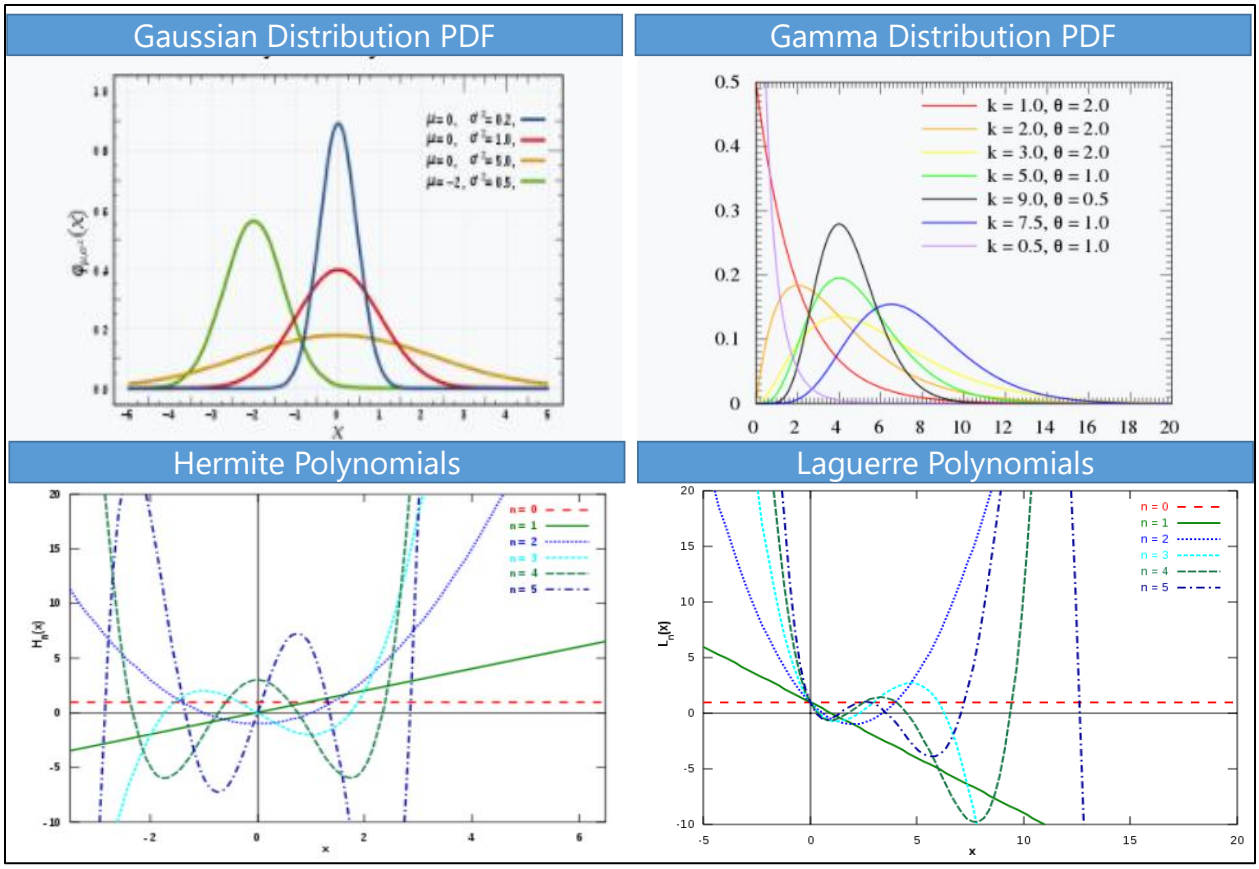
$$f(y(u)) = \sum_{n=0}^{\infty} f_n H_n(y(u)) \qquad (10)$$

1.5. Practical Illustration of Modelling a Finite Function with an Asymptotic Polynomial Expansion

An image of a Taylor polynomial approximation for two functions is shown in the figure below to illustrate the concept of polynomial approximation for a finite function.



Extending the concept to Gaussian and Gamma functions, the images below illustrate how Hermitian and Laguerrian polynomials are suited to approximating Gaussian and Gamma random functions.



1.6. Disjunctive Kriging

Disjunctive Kriging involves the use of polynomial asymptotic expansions to estimate values at unsampled locations. Asymptotic polynomial expansions are used to define the global anamorphosis function based on the sample data and respective transformed values. This can be a Gaussian transform / Hermitian expansion or a Gamma transform / Laguerrian expansion depending on which distribution is deemed most appropriate for the attribute being measured. The polynomial expansion values calculated for sample values/locations can be used in conjunction with simple kriging to determine the transformed (e.g., normal score or "gamma score") value at an unsampled location. The global anamorphosis function is used to determine the "raw" back transformed value of the attribute at the unsampled location.

Disjunctive kriging is simple co-kriging of the polynomial expansion of all orders (so informed by relative location and polynomial values at neighboring sampled locations). The polynomial expansions considered are orthogonal basis; the covariance between polynomials of different orders is zero and the covariance of polynomials of different orders is linked to a single variogram model; as a result, disjunctive kriging becomes simple kriging of the polynomial values for the transformed data values and then a linear sum of the resulting polynomial values across orders. This approach allows for the calculation of an expected value at an unsampled location based on statistical distance of nearby data points and reflecting the local CCDF as embedded in the polynomial values across orders at nearby sampled locations.

The polynomial expansion can be chosen to correspond to a given random function selection. While disjunctive kriging using Gaussian / Hermitian expansions is well documented as a substitute for simple MultiGaussian kriging, its application with other families of bivariate function families (such as MultiGamma) has not been as thoroughly described. Work has been done testing the relative appropriateness of a given expansion for modeling conditional PDF's by comparing a Hermitian expansion for the PDF and a Laguerrian expansion for the PDF, with the actual PDF from the exhaustive data set.

Application in Sequential Simulation

The polynomial approximation technique combined with disjunctive kriging can be used to define an expression for the local CCDF. Such polynomial expansions reflect the choice of bi-variate random function family. Ortiz (2004) provides a good overview of fitting a finite function (any finite function is acceptable) with an expansion of Hermite Polynomials. The approach is to set the function equal to a weighted sum of the Hermite Polynomial values as shown below. The coefficient value for a given order n is solved by calculation of the expected value of the function and the Hermite Polynomial of a given order n. The same author provides a method of determining the CDF function (indicator function $I_Y(u;y_c) = prob\ y \leq y_c$) of a given distribution based on Hermite Polynomial values as shown in the equations below. (note: $G(y)$ is the standard normal CDF)

$$I_Y(u; y_c) = \sum_{p=0}^{P} \psi_p H_p\big(Y(u)\big) \qquad (11)$$

$$\psi_0 = G(y_c) \qquad (12)$$

$$\psi_p = \frac{1}{\sqrt{p}} H_{p-1}(y_c) g(y_c) \qquad (13)$$

These equations let us choose (or observe) our experimental moments (mean and variance of Gaussian PDF $g(y_c)$ above) which inform the coefficients of the Hermitian expansion. The approximation of the CDF is a sum of products at each order of a coefficient (based on experimental moments) and the Hermitian polynomial. At each unsampled location we apply disjunctive kriging to determine the polynomial values based on the polynomial values calculated at the sample locations. We then apply the above equations (derived from the observed moments) to model the local CCDF function. A random draw can then be taken from this CCDF to arrive at a simulated value for the unsampled location.

2. Simulation Methodology

2.1. Prior Work

Emery (2006) also completed work on the use of various bivariate random function families to estimate ore body attributes at unsampled locations with Disjunctive Kriging. Work has also been completed (Emery 2002) using isofactorial representation of the bivariate random function family and disjunctive kriging to simulate ore body parameters ("sequential isofactorial simulation"). In this instance, while the theory is well described, the practical implementation including the approach sampling the local conditional cumulative distribution function is not.

2.2. Simulation Approach

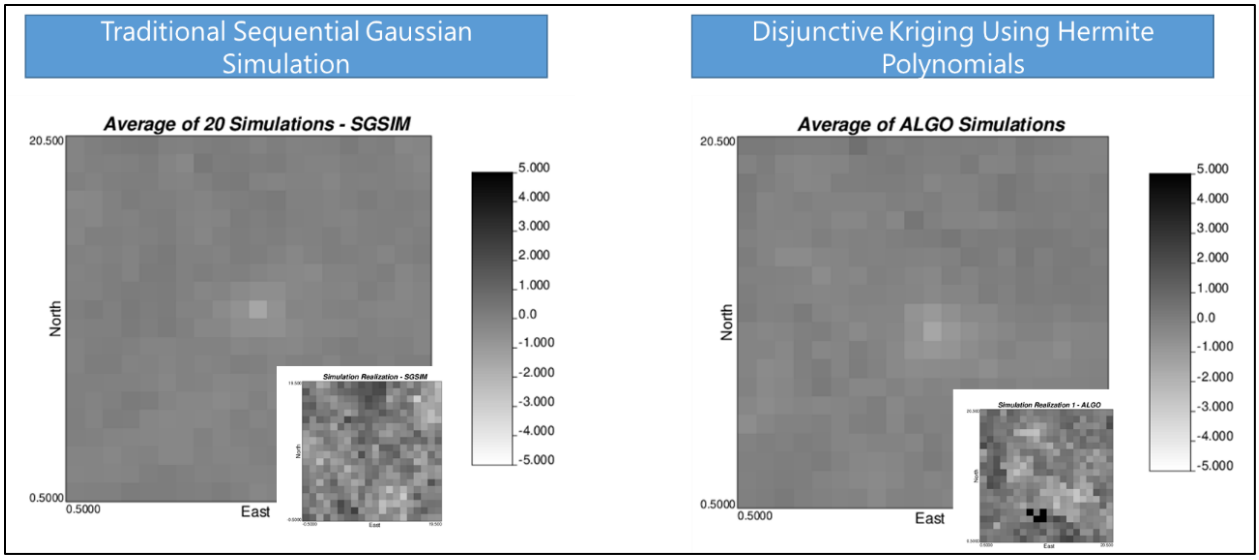
A sequential simulation algorithm was developed for sequential simulation using disjunctive kriging of Hermitian polynomials. The Hermitian polynomials up to a selected order "N" are calculated at all sampled locations based on normal score values. A variogram model is fitted to normal score values. Simple kriging is completed to solve for the Hermitian values of each order at the unsampled location. The covariance between polynomials of an order N is based on the normal score value variogram with the resulting correlogram value raised to the power "N".

Once the Hermitian values at the unsampled location are determined, the CCDF function at that location, (as a weighted sum of the Hermitian values) is used for the random draw. The weightings for each order Hermitian polynomial in the CCDF are derived based on equations 11, 12 and 13 shown in section 1.6 above. The result is an equation for the CCDF value at an unsampled location as a function of the actual (unknown) variable value at that location.

Sequential simulation draws a random number between zero and one and assumes this to be the CCDF value. Because the CCDF (by definition) is a monotonic function, a guess and check bounding algorithm was designed to iteratively determine the corresponding actual variable value that the CCDF random draw corresponds to (within a specified tolerance). The algorithm evaluates the CCDF value for the endpoints and midpoint of a range of the underlying variable; this range is then halved based on the relative location of the randomly drawn simulated CCDF value. The process is repeated until the drawn CCDF value is deemed close enough to one of the calculated CCDF values.

3. Results

It is possible to compare the results of the Hermitian disjunctive kriging simulation algorithm against frequently utilized traditional sequential Gaussian simulation programs. For such a comparison a 200x200 empty array was initialized and a seed value of -1.5 is placed in the center of the array. The seeded array is run through both the Hermitian disjunctive kriging algorithm, and GSLIB's Sequential Gaussian Simulation ("SGS") program. The results of both are shown below and indicate that simulation using Hermitian polynomials appears to accurately simulate a multi-Gaussian distribution.



4. Extensions of Methodology

4.1. MultiGamma Random Function

Wilson and Wragg (1973) provide three methods "…for the reconstruction of a continuous probability density function f(x) from given values of the moments of the distribution." One of these methods involves use of an asymptotic expansion, specifically an expansion of Laguerre Polynomials. This approach is suited to distributions that are expected or observed to be "Gamma like". This contrasts with the Hermite Polynomial expansions described by Edgeworth that are appropriate for distributions observed or expected to be "Gaussian like".

When dealing with the Hermitian expansion, the Gaussian distribution function is characterized by its mean and variance, which may be determined in a straightforward manner from the values of the sample data. In the Laguerrian case, the Gamma distribution function is characterized by a shape parameter α and a rate parameter β that are not as easily determined from the sample data. The shape and rate parameter are key inputs to the PDF and CDF functions for Gamma distributions. In order to utilize the Laguerrian expansion, we require a value for the shape and rate parameters. Using certain integral conditions, these parameters can be determined according to the below equations based on the first (mean) and second (variance) moments of the transformed sample data (Mustapha and Dimitrakopoulos 2010).

$$\alpha = \frac{2m_1^2 - m_2}{m_2 - m_1^2} \qquad (14)$$

$$\beta = \frac{m_1}{m_2 - m_1^2} \qquad (15)$$

With these parameters calculated, the equations of the Laguerrian expansion and polynomials are set out below.

$$P(z) \approx P_{\infty}^{\alpha}(z) = \sum_{n=0}^{\infty} r_n L_n^{(\alpha)}(z) \phi_{GAM}(z) \quad (16)$$

$$\phi_{GAM}(z) = \frac{\beta}{\Gamma(\alpha+1)} z^{\alpha} e^{-z} \quad (17)$$

$$L_n^{(\alpha)}(z) = \sum_{i=0}^{n} \frac{(-1)^i}{i!} \binom{n+\alpha}{n-i} z^i \quad (18)$$

$$r_n = \frac{n!\,\Gamma(\alpha+1)}{\Gamma(n+\alpha+1)} \sum_{i=0}^{n} \frac{(-1)^i}{i!} \binom{n+\alpha}{n-i} \beta^i m_i \quad (19)$$

It can be seen above that the PDF value as determined by the Laguerrian expansion is fully defined by the observed moments of order 1, 2, …, i of the available (and relevant) sample data. In the Hermitian case, as previously described, the PDF is determined based on the observed cumulants of the empirical distribution where such cumulants can be directly related to observed moments. These equations let us choose (or observe) our experimental moments which inform the coefficients and thus define the global PDF distribution function. The approximation of the local PDF (or CDF) is a sum of products across orders of a coefficient (based on experimental moments) and a Laguerre or Hermite polynomial. The Laguerrian case equations are more complicated than the Hermitian case; requiring both higher orders of cumulants and also incorporation of theoretical moments.

Mustapha and Dimitrakopoulos (2010) noted that a Laguerrian expansion around a Gamma distribution "…is suited for simulating high complex natural phenomena that deviate from Gaussianity". They examined certain data sets comparing a Hermitian expansion for the PDF and a Laguerrian expansion for the PDF, with the actual PDF from the exhaustive data set. When applying the Hermitian expansion, they utilized certain practical corrections to the Hermitian expansion such as the Saddle Point Approximation. The results showed better performance of a Laguerrian expansion relative to a Hermitian expansion of the same order (which had negative probabilities) or a Hermitian expansion with Saddlepoint Approximation (which was undefined in certain areas).

The choice of theoretical distribution used to develop the PDF function to approximate the empirical CCDF is very important under truncation. While in the case of an infinite expansion, any choice of finite developing distribution function will allow for convergence of the approximation, selection of a theoretical distribution that more closely fits the empirical data should reduce the order of the expansion required to achieve a certain quality of fit. This in turn allows for more efficient computational implementation. Application of our mapping tool for determination of the "best" suited random function family (and associated orthogonal polynomial expansion) has potential to improve computational efficiency and accuracy of ore body parametric simulation.

4.2. Destructuration of Grade

Emery (2008) identified that the destructuration of grade can be modelled by "randomizing" the correlation between the orders of the polynomials (i.e., the correlations itself becomes a random function). This correlation random function can range between a regular BiGaussian (or BiGgamma) model with no destructuration or a full mosaic model representing complete destructuration. This approach takes the pure BiGaussian (and BiGamma) model and extends them to more generalized "Hermitian" and "Laguerrian" models respectively. Without this adjustment, the BiGaussian and BiGamma models would be described as diffusive. The diffusive property mathematically requires that the correlogram of the polynomials of order "p" are equal to the correlogram of the variable raised to the power p. Practically, the diffusive property means that non-uniform connectivity of values (i.e. connectivity of extreme values) cannot be reflected in the model. This is due to the higher order correlograms trending to zero as the power p increases (pure nugget).

In order to create non-diffusive more generalized models, the correlation coefficient is randomized. The correlation coefficient of higher orders is similarly the randomized correlation coefficient raised to that specific order p. A Beta distribution for the randomized variable is often selected with parameters βρ(h) and β(1-ρ(h)) where the scalar parameter β takes a value between zero and one. Under this condition, it can be shown that the correlation coefficient for order "p" is described by the equation below rather than the diffusive case where it is $\rho(h)^p$ (Chiles and Delfiner 1999)

$$\forall p \in N^* \quad T_p(h) = \frac{\Gamma(\beta)\Gamma(\beta\rho(h)+p)}{\Gamma\big(\beta\rho(h)\big)\Gamma(\beta+p)} \qquad (20)$$

The above equation allows us to calculate the correlation coefficient for our Hermitian and Laguerrian polynomial expansions to create non-diffusive models. The question of how to choose the best β for a given data set is not readily apparent. Emery (2005) incorporates the randomizing distribution as a function of β into the equation for variograms of order ω, such that using the equations previously outlining a selection criterion can be calculated; the clear drawback of this approach is the inability to separate the choice of bi-variate random function family and choice of destructuration (β). An alternate approach described by Emery in 2002 involves the use of the relation between the observed variogram and madogram as shown in the equation below for the Laguerrian case. A similar expression can be derived for the Hermitian case.

$$\gamma_1(h) = \frac{\Gamma(\alpha+\frac{1}{2})\Gamma(\beta)}{\sqrt{\pi}\Gamma(\alpha)\Gamma(\beta+\frac{1}{2})}\frac{\Gamma(\frac{\beta\gamma(h)}{\alpha}+\frac{1}{2})}{\Gamma(\frac{\beta\gamma(h)}{\alpha})} \qquad (21)$$

These equations will be used to create a tool that optimizes the choice of β to the available data. In practical terms the equation above can be thought of as a method to assess the amount of destructuration present in a data set.

5. Conclusions

A sequential simulation algorithm has been developed with the ability to model non-Gaussian random function based simulations. Polynomial expansions are used to encode information content in the available sample data. Disjunctive kriging is used to model the CCDF function at unsampled locations. Unsampled locations are visited in a random order and at each location a random draw is made from the local CCDF function. The simulation algorithm is applied with a Gaussian random function model and associated Hermitian polynomial expansion. The results under specific conditioning data are in line with the results from a traditional sequential Gaussian simulation verifying the accuracy of the approach. Incorporation of a Gamma random function model (and associated Laguerre Expansion) is outlined. Consideration on using a coefficient of destructuration to allow simulation of non-diffusive random functions is also described.

6. Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2017-04200 and RGPAS-2017-507956.

7. References

Bhattacharya R., (2010) Normal Approximation and Asymptotic Expansions Society for Industrial and Applied Mathematics

Chiles J.P., Delfiner P., (1999) Geostatistics Modelling Spatial Uncertainty Wiley, New York NY

Dimitrakopoulos, M. (2009). High-order Statistics of Spatial Random Fields: Exploring Spatial Cumulants for Modeling Complex Non-Gaussian and Non-linear Phenomena. Mathematical Geosciences, 42(1), 65–99

Edgeworth F., (1907) On the Representation of Statistical Frequency by a Series. Journal of the Royal Statistical Society, 70 (1), 102–106

Emery X., (2002) Conditional simulation of nonGaussian random functions, Mathematical Geology, 34, p 79-100

Emery X., (2005a) Variograms of Order ω: a tool to validate a bivariate distribution model, Mathematical Geology, 37, p. 163-181

Emery X., (2005b) Geostatistical Simulation of Random Fields with Bivariate Isofactorial Distributions by Adding Mosaic Models. Stochastic environmental research and risk assessment, 19 (5), 348–360

Emery X., (2005c) Conditional Simulation of Random Fields with Bivariate Gamma Isofactorial Distributions. Mathematical geology, 37 (4), 419–445

Emery X., (2006) A Disjunctive Kriging Program for Assessment Point-Support Conditional Distributions Computers and Geosciences 32 965-983

Emery X., (2007) Using the Gibbs Sampler for Conditional Simulation of Gaussian-Based Random Fields. Computers & geosciences, 33 (4), 522–537

Emery X., (2008) Substitution Random Fields with Gaussian and Gamma Distributions: Theory and Application to a Pollution Data Set. Mathematical geosciences, 40 (1), 83–99

Emery X., Kremer F., (2008) A Survey of Random Field Models for Simulating Mineral Grades, in VII International Geostatistics Congress.

Goovaerts, P. (1997) Geostatistics for Natural Resources Evaluation, Oxford University Press, New York

Machuca-Mory D. F., Ortiz J. M., and Deutsch, C. V., (2008) On the challenge of using sequential indicator simulation for the estimation of recoverable reserves, International Journal of Mining, Reclamation and Environment, 22(4): 285-299

Mustapha D., Dimitrakopoulos R., (2010) Generalized Laguerre Expansions of Multivariate Probability Densities with Moments. Computers & mathematics with applications 60 (7), 2178–2189

Mustapha D., Dimitrakopoulos R., (2010b) High-order Stochastic Simulation of Complex Spatially Distributed Natural Phenomena. Mathematical geosciences. [Online] 42 (5), 457–485.

Ortiz J., (2004) A Step by Step Guide to Bi-Gaussian Disjunctive Kriging. In Geostatistics Banff; Springer Netherlands: Dordrecht; Vol. 14, pp 1097–1102

Ortiz J., (2019) MINE 834 Lectures. Queens University

Skovgaard M., (1986) On Multivariate Edgeworth Expansions International Statistical Review / Revue Internationale De Statistique, vol. 54, no. 2, pp. 169–186

Taylor Polynomial Expansions Website https://mathinsight.org/applet/taylor_polynomial accessed January 2021

Wallace D.L., (1958) Asymptotic Approximations to Distributions, The Annals of Mathematical Statistics, vol. 29, no. 3, pp. 635–654

Wilson G. A., Wragg A., (1973) Numerical methods for approximating continuous probability density functions using moments IMA J. Appl Math. 12 165173

Withers N., (2014) The Dual Multivariate Charlier and Edgeworth Expansions Statistics & probability letters 87, 76–85

Yao D., Dimitrakopoulos R., (2018) A New Computational Model of High-Order Stochastic Simulation Based on Spatial Legendre Moments, Mathematical geosciences 50 (8), 929–960

.

A Quick Guide to Developing a Mine Plan[1]

Mehmet Altinpinar (19ma57@queensu.ca)
Julian M. Ortiz (julian.ortiz@queensu.ca)

Abstract

Mining value chain is a series of interdisciplinary processes including prospecting, exploration, development, exploitation, mineral processing and reclamation. Mine planning plays an important role in the mining value chain for both open pit and underground mines starting from the end of the exploration stage and it continues toward the development and exploitation stages. In this paper, a quick guide for a life of mine planning process of open pit mining is provided and a case study is presented using a resource block model that is developed by implementing a synthetic drillhole campaign on a synthetic mineral deposit.

1. Introduction

In general terms, mine planning for an open pit operation starts with a block model and it involves determination of i) Whether a given block in the model should be mined or not; ii) If it is to be mined, when it should be mined; and iii) Once it is mined then how it should be processed (Dagdelen, 2001). Such block model used to be simply a geological model created by including the grades of the minerals to be exploited and processed. The current practices, on the other hand, involve more advanced models that incorporate expert knowledge from different areas like geology, mining, mineral processing, extractive metallurgy, mathematical modeling and computing to improve the use of resources such as ore, water, energy, equipment and labor (Ortiz, 2019).

After a resource block model is developed through exploration studies, three progressive stages of study are performed within the scope of planning (Hustrulid et al., 2013), namely:

1. Conceptual study (Scoping): This is the stage where a project idea is extensively transformed into an investment proposition by using historical data as a reference when making estimations regarding the capital and operation costs and highlighting the major aspects of a possible mining project.
2. Pre-feasibility study: This is an intermediate level study between a relatively inexpensive conceptual study and a relatively expensive feasibility study. It has a higher level of confidence compared to scoping, yet, still not suitable for an investment decision. Objective of the project, ore tonnage and grade, production schedule, capital cost, operating cost and revenue estimates, taxes and financing and cash flow tables are the important sections of a report to be generated at the end of a pre-feasibility study.

[1] Cite as: Altinpinar M, Ortiz JM (2021) A Quick Guide to Developing a Mine Plan, Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report 2021, paper 2021-06, 61-78.

3. Feasibility study: This is the stage in which a definitive technical, environmental and commercial base for an investment decision is provided. In addition to the important sections to be included in a pre-feasibility report, a feasibility report is prepared by including general information regarding the project area, such as topography, climate, population, services, etc., as well as the information about geology, mining, metallurgy, and environmental effects.

Besides resource estimation studies, these three stages of study should be carried out within the framework of a codified set of rules and guidelines such as NI 43-101 (National Instrument for the Standards of Disclosure for Mineral Projects within Canada), JORC (Joint Ore Reserves Committee) Code, SAMREC (South African Code for the Reporting of Mineral Resources and Mineral Reserves), etc., when reporting and disclosing information related to mineral properties. These guidelines do not only constitute instruments for reporting and disclosing purposes but also provide the mining companies with a vision to perform their technical studies in the most efficient and proper way. Some definitions and interpretations used in the following sections of this paper are taken from NI 43-101 (2016) and CIM (2014).

2. Methodology

In this study, a generalized methodology for mine planning is given for an open pit mine for a single commodity (e.g. copper). Different methodologies can be adopted based on project, type and number of commodities, mining method (i.e. open pit or underground), etc. Figure 1 shows the methodology followed in this study.

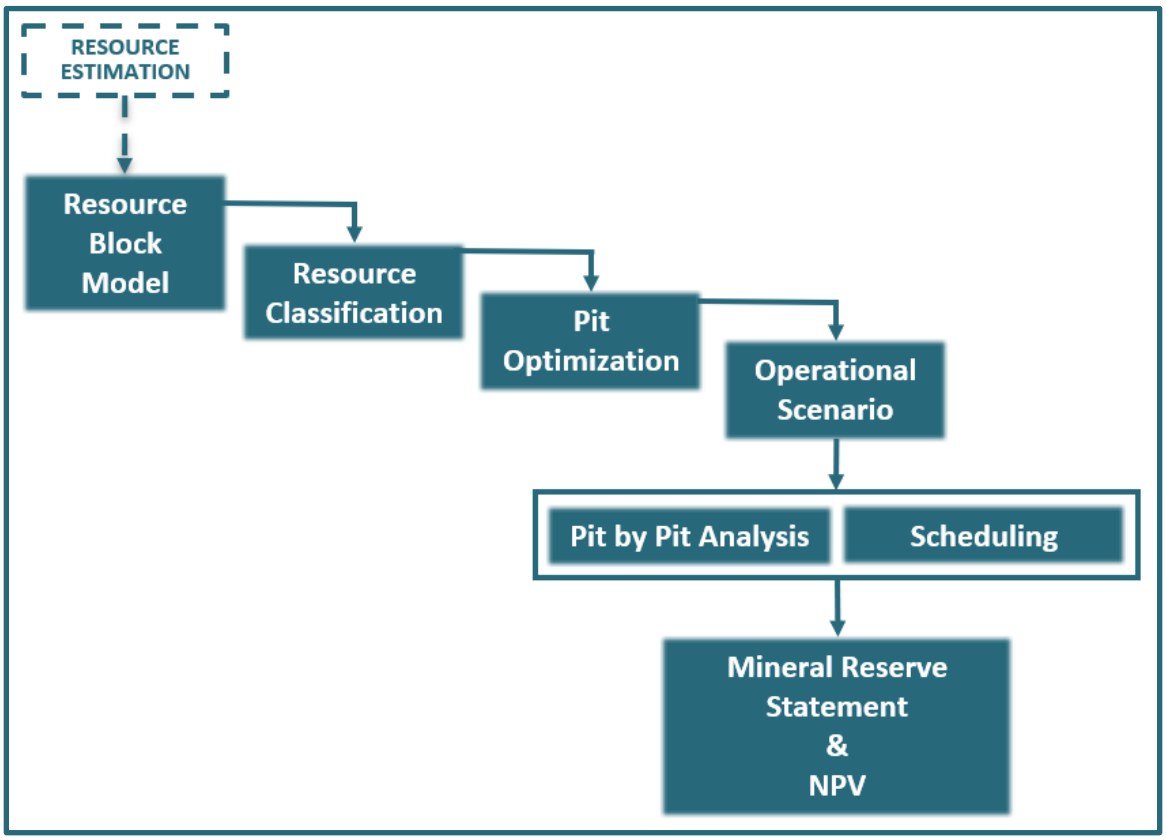


Figure 1 Methodology

2.1. Resource Block Model

A resource block model is obtained by performing resource estimation studies, and it should be validated before continuing with mine planning studies. After it is created and validated accordingly, the block model is now ready for the planning stage and it should, at least, include the parameters shown in Table 1 for the purpose of mine planning studies.

Table 1 Block Model Parameters for Mine Planning Studies

No	Variable	Remarks
1	xCentre	x coordinate of the centre of the block
2	yCentre	y coordinate of the centre of the block
3	zCentre	z coordinate of the centre of the block
4	xDimension	Dimension of the block in the x direction
5	yDimension	Dimension of the block in the y direction
6	zDimension	Dimension of the block in the z direction
7	Volume	Volume of the block
8	Domain	Domain of the block
9	Density	Block density for reporting tonnage.
10	RockCode	To determine if the block is ore, waste or airblock
11	NP	Number of estimation passes to categorize the block for resource classification
12	Grade	Grade of the element to be reported

The block model is typically comprised of several domains, which are basically identified according to the geological setting of the mineral deposit and based on the similarities and differences between the geological features, such as lithology, alteration and mineralization. Different types of rock materials present in a mineral deposit have different types of production (exploitation and excavation) and processing or treatment behaviours. Therefore, the domains in the resource model are categorized, in general, as:

1. Mineralized (ore) domains: The materials coming from these domains are sent to the mineral processing plant to extract the elements of interest.
2. Transition domains: The materials from these domains are either processed in processing plant or they are treated as waste materials. Further metallurgical tests are required to figure out how to treat these domains.
3. Waste domains: The materials coming from these domains are dumped in the waste dump areas within the project site.

2.2. Mineral Resource Classification

The resource block model is categorized based on the geological confidence as well as on the confidence level of the estimation process. There are three categories defined in NI 43-101 (2016) and CIM (2014), namely:

1. **Measured Mineral Resource:** A Measured Mineral Resource is that part of a Mineral Resource for which quantity, grade or quality, densities, shape, and physical characteristics are estimated with confidence sufficient to allow the application of Modifying Factors to support detailed mine planning and final evaluation of the economic viability of the deposit.

Geological evidence is derived from detailed and reliable exploration, sampling and testing and is sufficient to confirm geological and grade or quality continuity between points of observation.

A Measured Mineral Resource has a higher level of confidence than that applying to either an Indicated Mineral Resource or an Inferred Mineral Resource. It may be converted to a Proven Mineral Reserve or to a Probable Mineral Reserve.

Mineralization or other natural material of economic interest may be classified as a Measured Mineral Resource by the Qualified Person when the nature, quality, quantity and distribution of data are such that the tonnage and grade or quality of the mineralization can be estimated to within close limits and that variation from the estimate would not significantly affect potential economic viability of the deposit. This category requires a high level of confidence in, and understanding of, the geology and controls of the mineral deposit.

2. **Indicated Mineral Resource:** An Indicated Mineral Resource is that part of a Mineral Resource for which quantity, grade or quality, densities, shape and physical characteristics are estimated with sufficient confidence to allow the application of Modifying Factors in sufficient detail to support mine planning and evaluation of the economic viability of the deposit.

   Geological evidence is derived from adequately detailed and reliable exploration, sampling and testing and is sufficient to assume geological and grade or quality continuity between points of observation.

   An Indicated Mineral Resource has a lower level of confidence than that applying to a Measured Mineral Resource and may only be converted to a Probable Mineral Reserve.

   Mineralization may be classified as an Indicated Mineral Resource by the Qualified Person when the nature, quality, quantity and distribution of data are such as to allow confident interpretation of the geological framework and to reasonably assume the continuity of mineralization. The Qualified Person must recognize the importance of the Indicated Mineral Resource category to the advancement of the feasibility of the project. An Indicated Mineral Resource estimate is of sufficient quality to support a Pre-Feasibility Study which can serve as the basis for major development decisions.

3. **Inferred Mineral Resource:** An Inferred Mineral Resource is that part of a Mineral Resource for which quantity and grade or quality are estimated on the basis of limited geological evidence and sampling. Geological evidence is sufficient to imply but not verify geological and grade or quality continuity.

   An Inferred Mineral Resource has a lower level of confidence than that applying to an Indicated Mineral Resource and must not be converted to a Mineral Reserve. It is reasonably expected that the majority of Inferred Mineral Resources could be upgraded to Indicated Mineral Resources with continued exploration.

   An Inferred Mineral Resource is based on limited information and sampling gathered through appropriate sampling techniques from locations such as outcrops, trenches, pits, workings and drill holes. Inferred Mineral Resources must not be included in the economic analysis, production schedules, or estimated mine life in publicly disclosed pre-feasibility or feasibility studies, or in the life of mine plans and cash flow models of developed mines.

There are several parameters set during the estimation process and these parameters determine the level of confidence of each estimation pass. Such parameters include search ellipsoid radii, which are basically determined by performing variography analysis on element grade, the angles (i.e. bearing, plunge and dip) determining the positioning of ore body within mineral deposit, number of minimum and maximum samples to be used during each estimation pass, etc. Each estimation pass has a lower level of confidence than the previous pass. The estimation process is typically completed in 3 or 4 passes. Categorizing the estimation results based on the number of estimation pass is one of the ways for resource classification found in technical reports. However, there are better ways to categorize mineral resources according to the geological and geostatistical confidence, for example, the number of samples used to estimate a block grade, the distance between those samples and the estimated block, etc.

2.3. Pit Optimization

After resource classification, the resource block model is transferred to an optimization module or software to perform the pit optimization process. In this study, Geovia WhittleTM (Whittle) was used as the optimization software and the parameters defined in this study are the ones used in Whittle. The optimization is an implementation of the Lersch-Grossmann algorithm.

There are some adjustments made before transferring the resource model to Whittle in order to obtain proper results from the pit optimization process. These adjustments are explained below:

1. **Setting the grades of the blocks in the inferred mineral resource category to zero (0)** in order to ensure that they are excluded during the pit optimization and from mineral reserve statement. Only the measured and indicated mineral resource categories are taken into consideration when developing a production plan.

   This step includes all of the inferred blocks in the waste, transition and ore domains:
   a. Replacing the grades of all blocks in waste domains and transition domains (if they will be treated as waste material based on the results of metallurgical tests) with zero.
   b. Replacing the grades of all non-measured and non-indicated category blocks in ore domains with zero.

2. **Replacing all RockCode values with "6666" for the blocks with a grade value of zero (0)** in order to mark them as waste blocks in Whittle.
3. **Replacing all RockCode values with "5555" for the air blocks*** in the block model in order to mark them as air blocks in Whittle.

* air blocks are the ones above topography

Having completed these adjustments, the block model is now ready to be imported to Whittle for pit optimization. First, the optimization process is carried out by setting the parameters shown in Table 5 in order to generate nested pit shells for several revenue factors ranging from a low value (for example 0.1) to a high value (for example 2.0) with a reasonable step, such as 0.02. Here 0.1 and 2.0 are the coefficients to multiply the long term reference price of the commodity in question. So, it is expected to obtain a pit shell even with the lowest revenue factor, 0.1, and such pit shell would be the starter pit, i.e. where the excavator will be located in the field to start production. The pit shell to be obtained using the revenue factor 1.0 is the one that would be attained with the long term reference price of the commodity. All other

pit shells to be created using different revenue factors are generated to account for possible fluctuations in the commodity price in the future and understand the best sequence of extraction, so that the risks and/or benefits associated with the project could be addressed and interpreted.

2.4. Operational Scenario

After completing the optimization process as explained in Section 2.3, a new operational scenario is added in Whittle by using the parameters shown in Table 7 and the pit by pit analysis and scheduling steps are performed.

2.4.1. Pit by Pit Analysis

A pit by pit graph is obtained as a result of the pit by pit analysis in Whittle. It is a bar chart where each bar represents a pit shell corresponding to a revenue factor. Starter pit, pushbacks and ultimate pit shell can be determined by looking at this graph. Ideally, each "jump" observed in this graph is a candidate of a pushback and should be used in different combinations with other pushback candidates when trying to obtain the optimum scheduling scenario using the "best approach" option in Whittle. In some cases, however, it may not be possible to schedule by using the best approach option due to the exposure of significant amount of ore and waste material at a certain point during the mine life. In such cases, the said significant amount of material should be evenly split into a number of pushbacks, which have approximately equal production lives that are reasonable according to the mining and processing capacity of the project.

Ultimate pit limit is another factor when conducting pit optimization process. According to Mwangi et al. (2020) the main idea behind ultimate pit limit (UPL) optimization is the maximization of the total difference between the total cost of mining the valuable minerals and the overlying waste and the value that is obtained from the valuable mineral that will be mined with respect to satisfying all the pit slope stability and operational constraints. Therefore, the ultimate pit limit should be determined by comparing the risk associated with extending the life of mine and thereby increasing the amount of material (both valuable minerals and the overlying waste) to be removed (excavated) with the value to be obtained from the operation. It should be emphasized that the determination of the ultimate pit limit does not account for the time value of money, that is, the undiscounted value of the pit is considered in the optimization. Therefore, the added value of an expansion in the future will not reflect its true present value, considering the discount rate.

2.4.2. Scheduling

After being determined as explained in Section 2.4.1, starter pit, pushbacks and UPL are used as parameters to run the scheduling module of Whittle. The resultant graph and the corresponding spreadsheet show the amount of ore and waste materials to be excavated in each period during the life of mine. Also, open pit cash flow, discounted open pit cash flow and discounted cumulative open pit cash flow are reported in the scheduling output spreadsheet.

2.5. Mineral Reserve Statement and NPV

The starter pit and the pushbacks as well as the ultimate pit shell are transferred back to the mining software (Maptek Vulcan was used in this study) for visualization and reserve calculation studies. The measured and indicated blocks inside the UPL are now labelled as proven and probable reserves,

respectively. This is the simplified case of converting resources to reserves. In practice, however, all modifying factors including, but not limited to, mining, processing, metallurgical, infrastructure, economic, marketing, legal, environmental, social and governmental factors, which are set forth in the CIM Definition Standards for Mineral Resources & Mineral Reserves, should be taken into consideration (Altinpinar, 2021; CIM, 2014).

The net present value of a project can be obtained from the scheduling output spreadsheet generated at the end of scheduling module of Whittle. One could also calculate a project's NPV manually (in MS Excel for example) by using the same parameters shown in Table 7.

3. Case Study

3.1. Resource Block Model

The resource block model used in this study was obtained by performing a synthetic drillhole campaign on a synthetic mineral deposit, which was developed with a high resolution for a porphyry copper deposit (Altinpinar, 2021). The block model parameters are summarized in Table 2 and the domain classification is given in Table 3.

Table 2 Resource Block Model Parameters

No	Parameter	Value
1	# of blocks in the X (easting) direction	200
2	# of blocks in the Y (northing) direction	200
3	# of blocks in the Z (elevation) direction	100
4	Total # of blocks	4,000,000
5	Block dimensions in the X (easting) direction	10
6	Block dimensions in the Y (northing)direction	10
7	Block dimensions in the Z (elevation) direction	10
8	# of domains defined	16
9	# of ore domains	9
10	# of transition domains	2
11	# of waste domains	5
12	Density (fixed density value was used)	2.7 t/m3

Table 3 Domains in the Resource Block Model

Domain	Description	Category
111	Leached (LIX) - All Low Grade	Waste
222	Partially Leached (PLIX) - All Low Grade	Waste
333	Oxide (OX) - All Low Grade	Waste
444	Mixed (MIX) in Quartz (Qz)	Transition
499	Mixed (MIX) – Other	Transition
520	Primary in Sediments	Ore
530	Primary in Porphyries	Ore
532	Primary in Low Grade Porphyry	Ore
540	Primary in Dacite and Andesite	Ore
560	Primary in Hornfel, Skarn, Anhydrite	Ore
566	Primary in Hornfel	Ore
577	Primary in Skarn	Ore
580	Primary in Breccias	Ore
588	Primary Not Altered	Waste
599	Primary in Anhydrite	Ore
888	Host Rock	Waste

3.2. Mineral Resource Classification

The resource block model was categorized based on the estimation passes. In other words, the blocks estimated during the 1st pass were labelled as "Measured" mineral resource, the blocks estimated during the 2nd pass were labelled as "Indicated" mineral resource and the blocks estimated during the 3rd and 4th passes were labelled as "Inferred" mineral resource. The resource categories are summarized in Table 4.

Table 4 Resource classification

Parameter	Unit	Measured	Indicated	Inferred from the Mineralized Domains		Inferred from the Waste & Transition Domains	Inferred Overall
		NP_1	NP_2	NP_3	NP_4		
# blocks	count	306,487	148,450	123,298	264,537	2,683,821	3,071,656
Total Volume	m^3	306,487,000	148,450,000	123,298,000	264,537,000	2,683,821,000	3,071,656,000
Total Tonnage	t	827,514,900	400,815,000	332,904,600	714,249,900	7,246,316,700	8,293,471,200
Average Cu Grade	%	0.495	0.491	0.479		0.073	0.124
Total Cu	t	4,096,199	1,968,002	5,015,870		5,289,811	10,283,904

3.3. Pit Optimization

After resource classification was completed and the adjustments listed in Section 2.3 were made, the block model was imported to Whittle and pit optimization was performed by using the parameters shown in Table 5. Before the pit optimization process, the block model was reblocked in Whittle from 4 million blocks to 500 thousand block for ease of processing.

For the purpose of this case study, several combinations were tried for the range of revenue factor values and the corresponding pit by pit analyses were checked accordingly. Finally, the range from 0.1 to 0.52 with a step size of 0.005 was adopted as it generated the best alternative for the nested pit shells and also it was the best range option addressing the significant amount of ore and waste material exposed in year 4.2. The results of pit optimization process are shown in Table 6.

Table 5 Parameters for the pit optimization

Parameter	Unit	Value	Remarks
Slope	deg	45	Overall pit slope
Mining cost	$/t	2.75	Based on the industrial references
Mining recovery fraction	%	90	10% of the ore material will not be recovered during the production
Dilution	%	10	10% of the waste material will be sent to the processing plant
Processing cost	$/t	8	Based on the industrial references
Process recovery	%	85	Copper froth flotation recovery
Selling price	$/t	9750	LME – simply the current price was used
Selling cost	$/t	100	Including selling cost, insurance, freight
Revenue factors	-	0.1 to 0.52	Using 85 fixed factors with a step size of 0.005

Table 6 Generated pit shells

Pit	Minimum Revenue Factor	Maximum Revenue Factor	Rock Tones	Ore Tones	Strip Ratio	Max Bench	Min Bench	Cu Units	Cu Grade
1	0.34	0.345	10800	5346	1.02	44	43	2673	0.4999
2	0.35	0.355	27000	10692	1.53	44	43	5811	0.5435
3	0.36	0.37	129600	45441	1.85	45	42	25368	0.5583
4	0.375	0.375	276936300	91662517	2.02	48	25	49985506	0.5453
5	0.38	0.38	277068600	91943182	2.01	48	25	50071160	0.5446
6	0.385	0.385	291816000	99056035	1.95	48	24	53248638	0.5376
7	0.39	0.39	292607100	99459658	1.94	48	24	53421829	0.5371
8	0.395	0.395	310702500	107232742	1.9	48	24	56958453	0.5312
9	0.4	0.4	1280952900	463439398	1.76	48	10	229248172	0.4947
10	0.405	0.405	1294520400	468408506	1.76	48	10	231615624	0.4945
11	0.41	0.41	1321128900	478830533	1.76	48	10	236372364	0.4936
12	0.415	0.415	1345221000	488071094	1.76	48	10	240573258	0.4929
13	0.42	0.42	1468627200	520919591	1.82	49	9	258444841	0.4961
14	0.425	0.425	1496461500	530507642	1.82	49	9	262926543	0.4956
15	0.43	0.43	1534204800	542463971	1.83	49	8	268707117	0.4953
16	0.435	0.435	1547426700	548168153	1.82	49	8	271051757	0.4945
17	0.44	0.44	1564852500	555823625	1.82	49	8	274140416	0.4932
18	0.445	0.445	1578066300	561498404	1.81	49	8	276422454	0.4923
19	0.45	0.45	1630413900	580492743	1.81	49	8	284581087	0.4902
20	0.455	0.455	1645091100	586234347	1.81	49	8	286934404	0.4895
21	0.46	0.46	1654865100	590799831	1.8	49	8	288640431	0.4886
22	0.465	0.465	1665365400	595472235	1.8	49	8	290405259	0.4877
23	0.47	0.47	1701502200	607027614	1.8	49	8	295468453	0.4867
24	0.475	0.475	1718069400	612892176	1.8	49	8	297881050	0.486
25	0.48	0.48	1725075900	615575868	1.8	49	8	298935655	0.4856
26	0.485	0.485	1739553300	621207879	1.8	49	8	301104903	0.4847
27	0.49	0.49	1774890900	630993732	1.81	49	7	305540167	0.4842
28	0.495	0.495	1819268100	643201323	1.83	49	7	311039542	0.4836
29	0.5	0.5	1832814000	647753442	1.83	49	7	312859124	0.483
30	0.505	0.505	1883150100	663836883	1.84	49	7	319400841	0.4811
31	0.51	0.51	1902482100	670372368	1.84	49	7	321965508	0.4803
32	0.515	0.515	1917691200	675357514	1.84	49	7	323931481	0.4796
33	0.52	0.52	1931585400	679949728	1.84	49	7	325714262	0.479

3.4. Operational Scenario

After completing the pit optimization step, a new operational scenario was added using the parameters shown in Table 7 and pit by pit analysis and scheduling steps were run.

Table 7 Parameters for operational scenario

Initial capital cost	M$	500		Initial investment for the project
Terminal value	M$	50		Assets to be sold at the end of the project
Discount rate per period	%	10		The rate used in Discounted Cash Flow Analysis
Mining limit	Mt	Year 1	30	Annual open pit production capacity was assumed to be low during the first three years of the operation with an increasing rate and eventually reaching the full capacity at the fourth year.
		Year 2	60	
		Year 3	90	
		Year 4-18	120	
Processing limit	Mt	80		Annual processing plant capacity (based on the stripping ratio)
Mining cost adjustment factor	-	None		Adjustment factor was not applied
Processing cost adjustment factor	-	None		Adjustment factor was not applied

3.4.1. Pit by Pit Analysis

The results of the pit by pit analysis are shown in Table 8 and in Figure 2 as a graph.

Table 8 Results of the pit by pit analysis

Final pit	Open pit cashflow best disc ($)	Open pit cashflow specified disc ($)	Open pit cashflow worst disc ($)	Ore Input best (tonne)	Waste best (tonne)	Mine life years best	Mine life years specified	Mine life years worst	IRR best %	IRR specified %	IRR worst %
1	-449,854,976	-449,854,976	-449,854,976	5,346	5,454	0.0	0.0	0.0	0.0	0.0	0.0
2	-449,687,427	-449,687,427	-449,687,427	10,692	16,308	0.0	0.0	0.0	0.0	0.0	0.0
3	-448,660,277	-448,660,277	-448,660,277	45,441	84,159	0.0	0.0	0.0	0.0	0.0	0.0
4	1,321,835,117	1,321,723,265	1,321,723,265	93,343,833	183,592,467	3.9	3.9	3.9	48.9	48.8	48.8
5	1,322,944,917	1,322,958,987	1,322,958,987	93,440,061	183,628,539	3.9	3.9	3.9	48.9	48.9	48.9
6	1,419,722,843	1,415,707,788	1,415,707,788	100,574,298	191,241,702	4.0	4.0	4.0	50.8	50.0	50.0
7	1,424,545,976	1,420,245,209	1,420,245,209	100,846,944	191,760,156	4.0	4.0	4.0	50.8	50.0	50.0
8	1,544,174,588	1,532,655,196	1,532,655,196	108,721,601	201,980,899	4.2	4.2	4.2	52.4	50.1	50.1
9	4,372,395,169	3,466,108,146	3,466,108,146	495,357,685	785,595,215	12.9	13.0	13.0	53.6	28.8	28.8
10	4,399,537,802	3,462,636,363	3,462,636,363	499,303,033	795,217,367	12.9	13.1	13.1	53.6	28.6	28.6
11	4,459,663,124	3,472,658,482	3,472,658,482	508,693,282	812,435,618	13.1	13.3	13.3	53.6	28.2	28.2
12	4,513,496,779	3,481,083,449	3,481,083,449	517,033,042	828,187,958	13.2	13.5	13.5	53.7	27.9	27.9
13	4,726,216,184	3,407,245,975	3,407,245,975	549,306,843	919,320,357	14.3	14.4	14.4	53.7	26.1	26.1
14	4,773,594,194	3,387,226,300	3,387,226,300	557,986,074	938,475,426	14.5	14.7	14.7	53.8	25.7	25.7
15	4,832,006,744	3,364,501,391	3,364,501,391	569,298,210	964,906,590	14.8	15.0	15.0	53.8	25.3	25.3
16	4,853,321,420	3,362,941,916	3,362,941,916	574,109,610	973,317,090	14.9	15.1	15.1	53.8	25.2	25.2
17	4,883,526,547	3,360,953,090	3,360,953,090	581,021,988	983,830,512	15.1	15.2	15.2	53.8	25.0	25.0
18	4,905,460,918	3,360,465,255	3,360,465,255	585,710,430	992,355,870	15.2	15.3	15.3	53.8	24.9	24.9
19	4,985,845,794	3,328,612,470	3,328,612,470	604,325,202	1,026,088,698	15.6	15.8	15.8	53.8	24.3	24.3
20	5,005,432,691	3,318,632,797	3,318,632,797	608,954,838	1,036,136,262	15.7	15.9	15.9	53.8	24.2	24.2
21	5,018,816,787	3,306,180,874	3,306,180,874	612,475,179	1,042,389,921	15.8	16.0	16.0	53.8	24.1	24.1

Table 8 (continued)

22	5,032,512,917	3,297,718,479	3,297,718,479	616,201,341	1,049,164,059	15.9	16.1	16.1	53.8	24.0	24.0
23	5,076,523,492	3,265,071,050	3,265,071,050	626,917,398	1,074,584,802	16.2	16.4	16.4	53.8	23.6	23.6
24	5,096,475,507	3,249,369,884	3,249,369,884	632,049,558	1,086,019,842	16.4	16.5	16.5	53.8	23.4	23.4
25	5,104,039,026	3,239,385,058	3,239,385,058	633,901,947	1,091,173,953	16.4	16.5	16.5	53.8	23.3	23.3
26	5,121,227,528	3,225,145,789	3,225,145,789	638,972,628	1,100,580,672	16.5	16.7	16.7	53.9	23.2	23.2
27	5,154,872,808	3,182,157,563	3,182,157,563	648,023,406	1,126,867,494	16.8	16.9	16.9	53.9	22.8	22.8
28	5,196,279,800	3,122,062,345	3,122,062,345	659,707,089	1,159,561,011	17.2	17.3	17.3	53.9	22.3	22.3
29	5,209,276,496	3,104,521,979	3,104,521,979	663,708,570	1,169,105,430	17.3	17.4	17.4	53.9	22.2	22.2
30	5,255,794,518	3,045,152,821	3,045,152,821	679,628,958	1,203,521,142	17.7	17.8	17.8	53.9	21.8	21.8
31	5,272,292,903	3,015,841,770	3,015,841,770	685,651,227	1,216,830,873	17.9	18.0	18.0	53.9	21.6	21.6
32	5,284,296,033	2,996,149,103	2,996,149,103	690,042,966	1,227,648,234	18.0	18.1	18.1	53.9	21.4	21.4
33	5,296,073,784	2,977,502,943	2,977,502,943	694,119,291	1,237,466,109	18.1	18.2	18.2	53.9	21.3	21.3

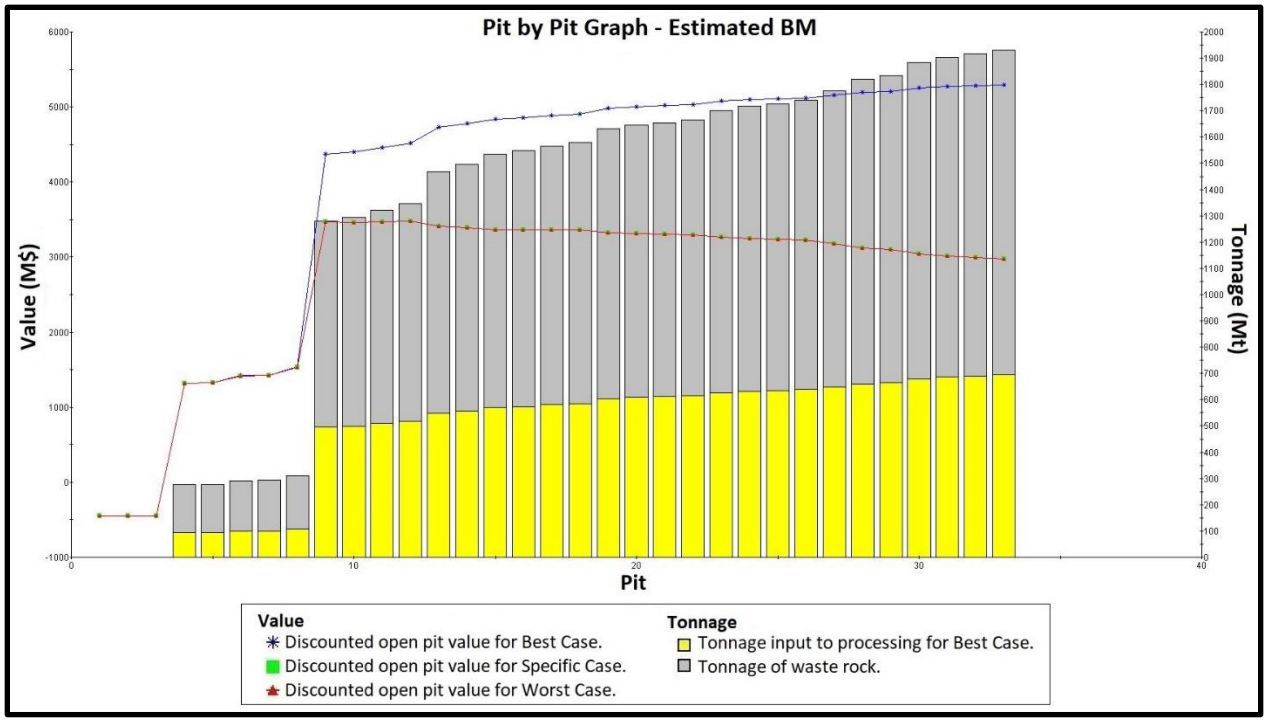


Figure 2 Pit by pit graph

As seen in Figure 2, a big jump in reserve at the 9th pit shell, which corresponds approximately to the fourth year of the operation, is obtained even with a slight increase in revenue factor and a significant amount of material (ore and waste) is exposed. This means that the scheduling after that period is not sensitive to the direction of mining and basically, similar materials are being accessed for a long period of time. Therefore, the life of mine was manually divided into periods that are close to each other as shown in Table 9 and the 33rd pit shell was selected as the ultimate pit limit.

Table 9 Pushbacks

Pushback	Period (Year)	Pit Shell
1	The first 4 years	Pit 8
2	The next 5 years	Halfway through Pit 9
3	The next 4 years	Until the end of Pit 9
4	Until the end of LOM (18 years)	Pit 33 (UPL)

3.4.2. Scheduling

The scheduling step was run using Whittle's Milawa NPV option, which aims at maximizing the net present value of the project. The scheduling graph and the scheduling outputs are given in Figure 3 and Table 10, respectively.

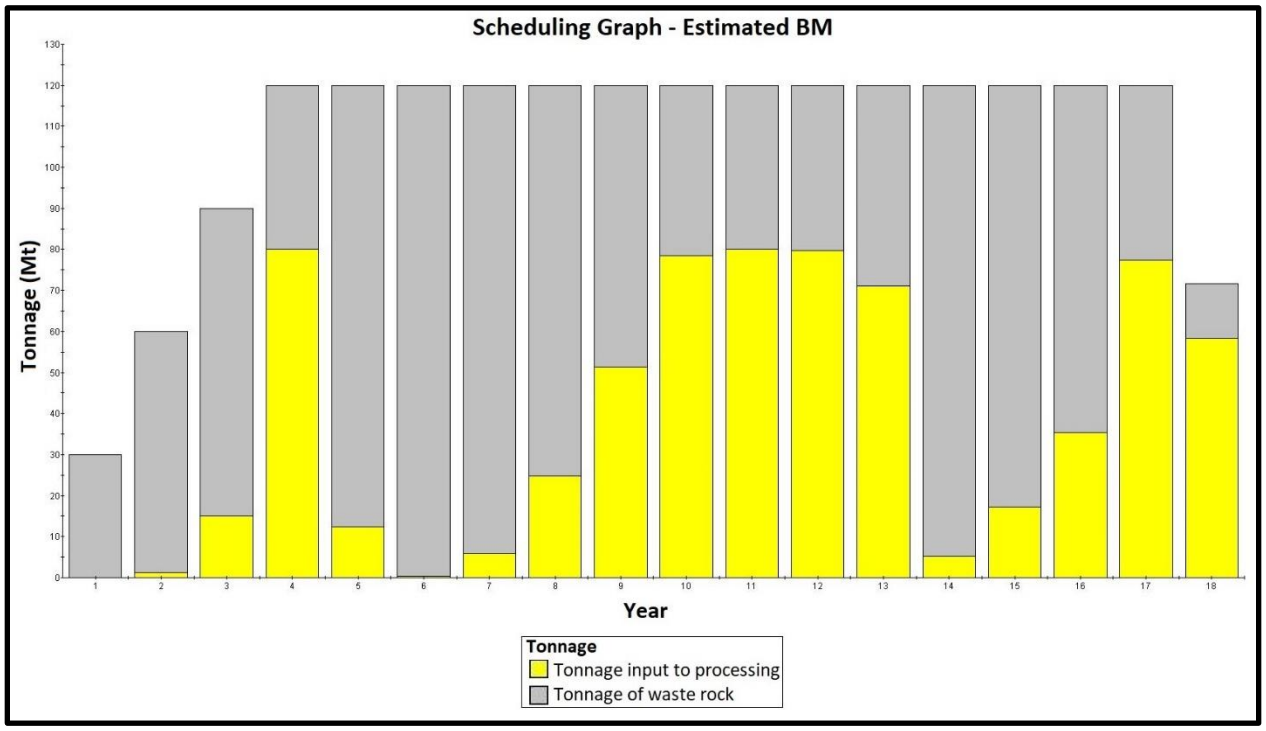


Figure 3 Scheduling graph

The pushbacks as well as the UPL were then transferred back to Maptek Vulcan to visualize the progress of the open pit operation as shown in Figure 4.

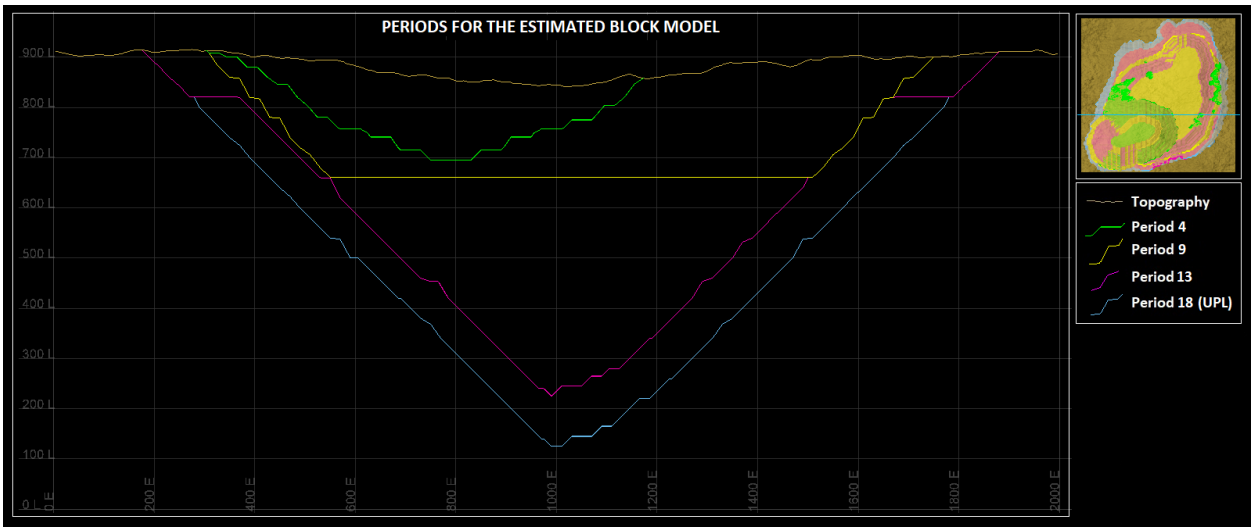


Figure 4 Progress of the open pit operation

Table 10 Scheduling outputs and net present value of the project

Period (year)	Tonne Input (kt)	Waste Tonne (kt)	Strip Ratio	Grade Input Cu (%)	Open Pit Cashflow (M$)	Open Pit Cashflow Discounted (M$)	Open Pit Cumulative Cashflow Discounted (M$)
1	8	29,992	999.99	0.5496	-82	-75	-75
2	1,209	58,791	48.64	0.4697	-128	-106	-181
3	15,085	74,915	4.97	0.4039	132	99	-82
4	80,000	40,000	0.5	0.5284	2,497	1,706	1,624
5	12,420	107,580	8.66	0.6732	256	159	1,783
6	345	119,655	347	0.4476	-320	-181	1,603
7	5,895	114,105	19.36	0.3713	-198	-101	1,501
8	24,847	95,153	3.83	0.3831	252	118	1,619
9	51,338	68,662	1.34	0.4085	980	415	2,034
10	78,515	41,485	0.53	0.4252	1,780	686	2,721
11	79,996	40,004	0.5	0.4221	1,799	631	3,351
12	79,768	40,232	0.5	0.5578	2,682	854	4,206
13	71,090	48,910	0.69	0.506	2,052	594	4,800
14	5,239	114,761	21.91	0.3258	-232	-61	4,739
15	17,305	102,695	5.93	0.3873	81	19	4,758
16	35,360	84,640	2.39	0.5208	898	195	4,954
17	77,410	42,590	0.55	0.4324	1,797	355	5,309
18	58,291	13,295	0.23	0.512	1,835	339	5,648
			Total		16,081	5,648	5,648

3.5. Mineral Reserve Statement and NPV

As shown in Table 10, the discounted net present value of the project was calculated by Whittle as US$ 5,648M. In Whittle, different options (e.g. Milawa Balanced, which aims at establishing a balance

between the open pit production rate and the capacity of the mineral processing plant) can be selected based on the mine planning strategy of the project and different NPV values can be obtained.

After completing the pit optimization and scheduling process in Whittle, the UPL was imported to Maptek Vulcan to carry out mineral reserve calculations. The measured and indicated blocks inside the UPL were labelled as proven and probable reserves, respectively. The results are given in Table 11 for the case without a threshold and in Table 12 for the case where a threshold of Cu grade greater than and equal to 0.15% was applied. Figure 5 shows the proven and probable blocks within the UPL for the case with a threshold of Cu ≥ 0.15%.

Table 11 Mineral reserve statement (without a threshold)

Domain	PROVEN RESERVE				PROBABLE RESERVE				TOTAL PROVEN & PROBABLE			
	# of Blocks	Ore (kt)	Mean Grade Cu %	Metal Content (kt)	# of Blocks	Ore (kt)	Mean Grade Cu %	Metal Content (kt)	# of Blocks	Ore (kt)	Mean Grade Cu %	Metal Content (kt)
520	172,347	465,337	0.488	2,270.84	32,645	88,142	0.514	453.05	204,992	553,478	0.492	2,723.89
530	5,700	15,390	0.454	69.87	1,023	2,762	0.415	11.46	6,723	18,152	0.448	81.33
532	1,918	5,179	0.759	39.31	1,901	5,133	0.709	36.39	3,819	10,311	0.734	75.70
540	8,172	22,064	0.408	90.02	3,478	9,391	0.461	43.29	11,650	31,455	0.424	133.31
560	728	1,966	0.648	12.74	561	1,515	0.647	9.80	1,289	3,480	0.648	22.54
566	16,128	43,546	0.888	386.68	6,384	17,237	0.749	129.10	22,512	60,782	0.849	515.79
577	1,315	3,551	0.318	11.29	28	76	0.317	0.24	1,343	3,626	0.318	11.53
580	723	1,952	0.365	7.13	176	475	0.366	1.74	899	2,427	0.365	8.86
599	4,614	12,458	0.345	42.98	93	251	0.307	0.77	4,707	12,709	0.344	43.75
Total	211,645	571,442	0.513	2,930.86	46,289	124,980	0.549	685.85	257,934	696,422	0.519	3,616.71

Table 12 Mineral reserve statement with a threshold (Cu grade ≥ 0.15%)

Domain	PROVEN RESERVE				PROBABLE RESERVE				TOTAL PROVEN & PROBABLE			
	# of Blocks	Ore (kt)	Mean Grade Cu %	Metal Content (kt)	# of Blocks	Ore (kt)	Mean Grade Cu %	Metal Content (kt)	# of Blocks	Ore (kt)	Mean Grade Cu %	Metal Content (kt)
520	170,785	461,120	0.491	2,264.10	32,492	87,728	0.516	452.68	203,277	548,848	0.495	2,716.78
530	5,700	15,390	0.454	69.87	1,023	2,762	0.415	11.46	6,723	18,152	0.448	81.33
532	1,918	5,179	0.759	39.31	1,899	5,127	0.710	36.40	3,817	10,306	0.735	75.71
540	8,172	22,064	0.408	90.02	3,478	9,391	0.461	43.29	11,650	31,455	0.424	133.31
560	728	1,966	0.648	12.74	561	1,515	0.647	9.80	1,289	3,480	0.648	22.54
566	16,127	43,543	0.888	386.66	6,384	17,237	0.749	129.10	22,511	60,780	0.849	515.76
577	1,315	3,551	0.318	11.29	28	76	0.317	0.24	1,343	3,626	0.318	11.53
580	723	1,952	0.365	7.13	176	475	0.366	1.74	899	2,427	0.365	8.86
599	4,613	12,455	0.345	42.97	93	251	0.307	0.77	4,706	12,706	0.344	43.74
Total	210,081	567,219	0.516	2,924.08	46,134	124,562	0.550	685.49	256,215	691,781	0.522	3,609.57

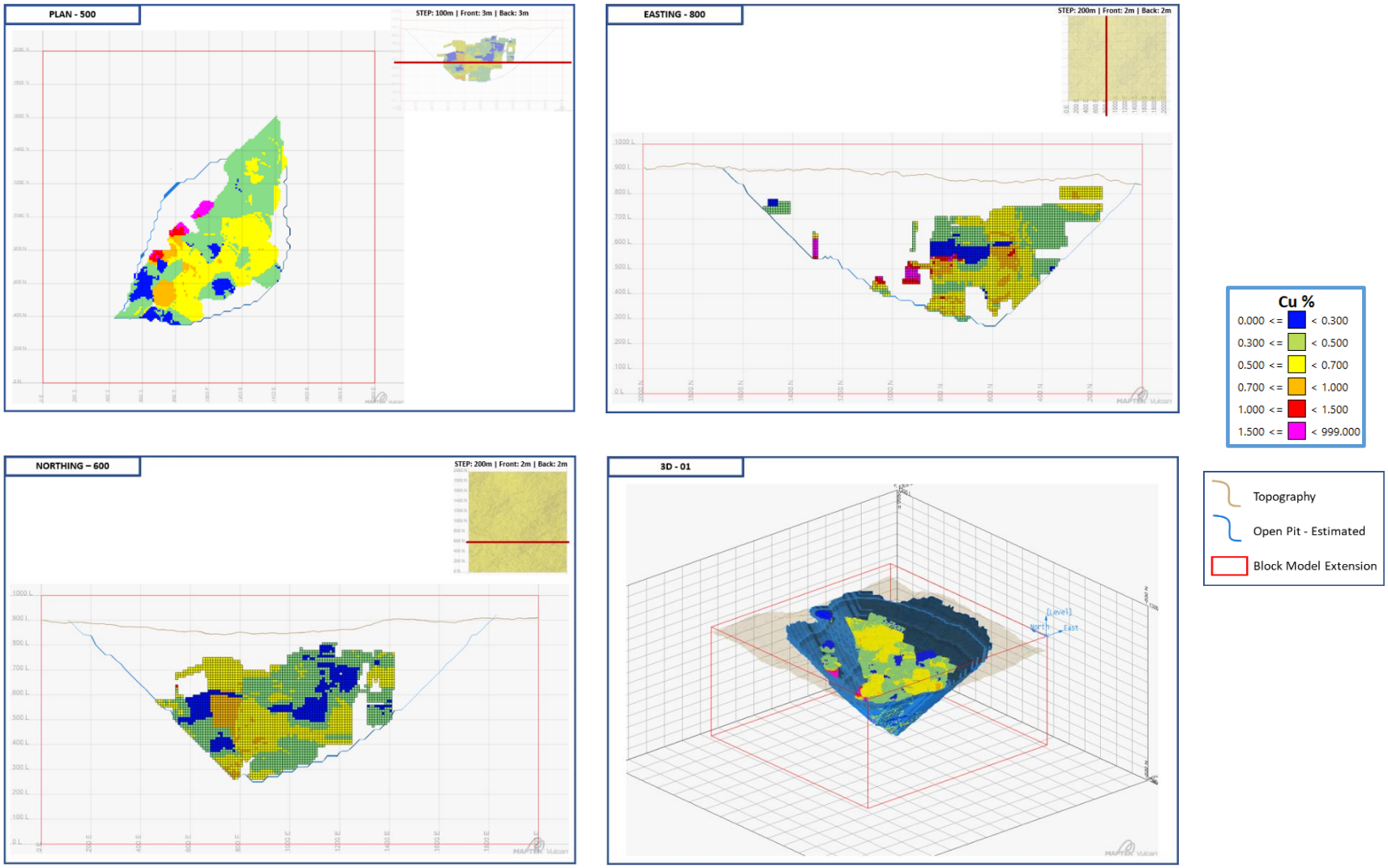


Figure 5 Proven and probable blocks within the open pit (Cu grade ≥ 0.15%)

4. Conclusions

Mine planning is one of the most important operational steps in achieving strategic objectives of a mining project. The entire mine planning process should be intently conducted with a very high level of detail and accuracy as there are already many uncertainties about the mining value chain. In this paper, the process of mine planning was outlined with a quick guide, which includes the steps for pit optimization, scheduling and mineral reserve statement. With the help of a case study, the methodology was applied on a block model, which was obtained through resource estimation studies performed by using a synthetic mineral deposit, and it was explained how the involved parameters can be used during the performance of mine planning.

The stages presented include the selection of the ultimate pit limit, which results from the application of the Lersch-Grossmann algorithm and uses undiscounted value. The selection of the ultimate pit limit should consider the risk of extending the life of the mine additional years, versus the value added by the additional ore. The time value of money reduces the present value of future resources, therefore, typically the ultimate pit limit is not the one that maximizes the undiscounted value. Once the UPL is selected, the pushbacks need to be determined, which provide the direction in which the pit evolves over time. These pushbacks should be large enough to contain several years of production and should be large so the pit can be expanded from one phase to the next with enough space to create working benches safely. Finally, the yearly schedule must account for a sustained ore feed to the processing plant, while maintaining a relatively steady stripping ratio, to facilitate the operation. Different optimization approaches are available for this purpose.

5. References

Altinpinar M (2021). Synthetic High Resolution Block Model For Benchmarking Mining Technologies. M.A.Sc. Thesis, Queens's University, Kingston.

CIM (2014). Definition Standards for Mineral Resources & Mineral Reserves.

Collins SE, Holbek P, Strickland D (2018). NI 43-101 Technical Report for the Copper Mountain Mine.

Dagdelen K (2001). Open Pit Optimization - Strategies for Improving Economics of Mining Projects Through Mine Planning, 17th International Mining Congress and Exhibition of Turkey – IMCET 2001, ISBN 975-395-417-4.

Hustrulid W, Kuchta M, Martin R (2013). Open Pit Mine Planning and Design, 3rd Edition.

Moore CM, Miranda HM, Hampton AP, Ritchie DG (2019). Lundin Mining Corporation. Technical Report on the Chapada Mine, Goiás State, Brazil.

Mwangi AD, Jianhua Z, Gang H, Kasomo RM, Innocent MM (2020). Ultimate Pit Limit Optimization Methods in Open Pit Mines: A Review. J Min Sci 56, 588–602. https://doi.org/10.1134/S1062739120046885

National Instrument 43-101 (2016). Standards of Disclosure for Mineral Projects.

Ortiz JM (2019). Geometallurgical modeling framework, Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report 2019, paper 2019-01, 6-16.

SRK Consulting (2020). NI 43-101 Technical Report, Feasibility Study for the Josemaria Copper-Gold Project, San Juan Province, Argentina.

A Simple Implementation Example of SVC[1]

Koruk, Kasimcan (kasimcan.koruk@queensu.ca)

Abstract

Support Vector Machines (SVM) is a powerful method in the world of machine learning, because it can handle complex regression and classification problems. Today, many machine learning libraries offer robust modules for several methods, and Scikit-learn's Support Vector Classifier (SVC) is one of them. This article's objective is to determine the boundaries of a binary-class problem using SVC. In this context, the article shows a simple example of SVC on a 3-D drillhole data. To see the potential of SVC, it is applied on 3D pseudo-data which are comprised of composite samples of vertical drillholes distributed homogenously in space. Cross-validation is applied on the data to determine the optimum parameters of the method. Posterior probabilities are obtained thanks to Platt's method embedded in SVC module. The data is split into train and test datasets to assess the accuracy of the model. The article shows that SVC can provide complex and accurate models at a cost of possible misclassifications of outliers in the multi-dimensional space.

1. Introduction

Support Vector Machine (SVM) is a powerful method for machine learning because it can handle complex regression and classification problems (Bishop, 2006) and it has been a popular method for many years (Géron, 2017). The name of SVM comes from the subset of training samples, known as support vectors, utilized in the decision function. Support vectors let computer use less amount of training subsets, and thus the model becomes memory efficient (Pedregosa et al., 2011). In the case of a classification problem, support vectors optimize class margins to a maximum possible span. To do this, SVM utilizes a complex mathematical algorithm. In the following subsections, first the mathematics behind SVM are explained in a simple manner, then an implementation of SVC on a simple drillhole data is shown with the details of programming in Python. SVC is a strong classification module offered by Scikit-learn library in Python. An important notice for the implementation is that posterior probability results are obtained from SVC, which is not possible normally. Although there are some earlier approximations to compute posterior probability of SVM, Platt (1999) succeeded to obtain a better posterior probability using a sigmoid model. The method of Platt (1999) to obtain posterior probability is embedded in SVC of Scikit-learn, and it is activated in the implementation of SVC.

[1] Cite as: Koruk K (2021) A Simple Implementation Example of SVC, Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report 2021, paper 2021-07, 79-87.

2. Support Vector Machines

2.1. Theoretical Review of SVM

The mathematics behind SVM can be challenging for a learner. The best approach to understand the logic of SVM is to start from a simple two-class and two-parameter case. For this simple case, SVM model takes the linear form

$$y(x_n) = w^T \Phi(x_n) + b \tag{1}$$

Where x_n is training input from 1 to N number of inputs, $y(x_n)$ is target classes, $\Phi(x_n)$ is a fixed transformation function of x, w^T is a coefficient vector which is maximizing the distance between two classes and b is the parameter to control the bias (Bishop, 2006). The equation is quite similar to logistic regression technique, however the main difference between logistic regression and SVM is that SVM maximizes the distance between classes with the help of margins at each side of the class boundary (Figure 1). The samples located on these margins are called support vectors.

Going back to the equation, SVM tries to approximate best values to the unknowns, w^T and b, to maximize the margins of the boundary. Optimum condition is provided when $1/\|w\|$ is maximized. If classification can be performed without letting any misclassification, the classification is called hard margin classification like in Figure 1. In its simplest manner, the problem takes the form

$$t_n(w^T \Phi(x_n) + b) = 1 \tag{2}$$

where t_n is target for samples from 1 to N number of samples.

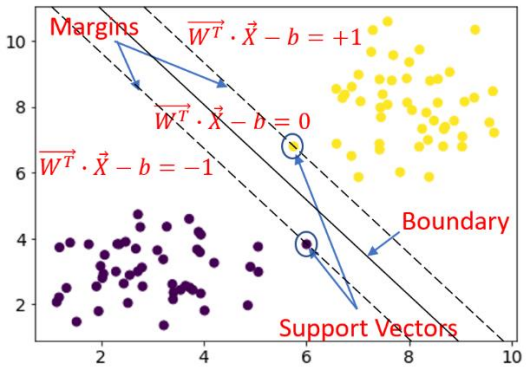


Figure 1 Hard margin linear SVM classification

However, in most cases hard margin classification is not possible. At this point, soft margin classification comes as a solution. Soft margin classification lets the model misclassify some samples to reach the optimum condition (Figure 2). To make misclassification possible, slack variables, ξ_n, are introduced into the equation (Bishop, 2006). $\xi_n = 0$ when data points are correctly classified, and the rest is $\xi_n = |t_n - y(x_n)|$, meaning that samples located on boundary or samples passing boundary are penalized up to 1. After the introduction of slack variables, the classification problem takes the form:

$$t_n y(x_n) \geq 1 - \xi_n \tag{3}$$

Theoretically, minimizing the summation of condition (3) for n from 1 to N number of samples can provide the best classification:

$$min \sum_{n=1}^{N} \xi_n + \frac{1}{2}\|w\|^2 \tag{4}$$

The equation (4) is scaled by a parameter C, and C, with constraint of being higher than 0, regularizes the complexity of the model.

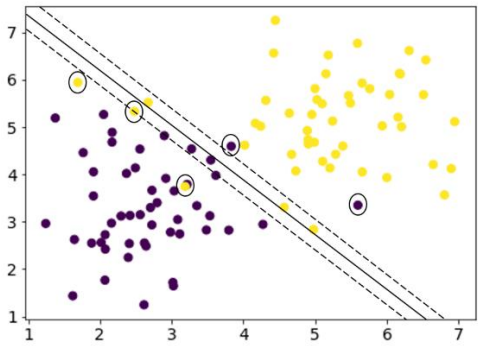


Figure 2 Soft margin linear SVM classification; encircled samples are misclassified samples

So far, the mathematical explanation of SVM is done as simple as possible with focus on linearly separable class problems. However, most of the real scenarios require non-linear solutions with quadratic function problems. Detailed explanation for quadratic function problems can be found in Bishop (2006). For the sake of an easier comprehension, the problem is kept simple. However, it is worth to mention kernel functions to better understand the non-linear solutions. The transformation functions we called earlier $\Phi(X_n)$ are kernel functions. Technically, Kernel functions works as if more features are added to sample spaces to make classes separable (Géron, 2017). Most popular kernel functions are tabulated in Table 1 below:

Table 1 Types of kernel functions

Kernel Types	*Function*
Linear Kernel	$k(x, x') = x^T x'$
Polynomial Kernel	$k(x, x') = (x^T x' + r)^d$
Radial Basis Kernel	$k(x, x') = e^{(-\gamma\|x-x'\|^2)}$
Sigmoid Kernel	$k(x, x') = \tanh(\gamma(x^T x') + r)$

Linear kernel, polynomial and radial basis kernel are derived from same the equation. While d equals 1 in polynomial kernel, polynomial kernel becomes linear kernel. Increasing d can be considered as increasing the amount of feature space in the training process. If d approximates to ∞, polynomial kernel approximates to radial basis function. The approximation is performed by making use of Taylor Expansion Series. Fundamentally, employing radial basis function can yield similar but much faster performance compared to polynomial kernel when d is ∞. Therefore, radial basis function (RBF) is preferred on the implementation of SVC.

Classifiers involving posterior probability are very useful in target recognition (Platt, 1999). Platt (1999) offered a modification to a previous multinomial likelihood non-sparse machine method. Platt offers sigmoid function for probabilities:

$$P(y = 1|y(x_n)) = \frac{1}{1 + \exp\,(A\,y(x_n) + B)} \tag{5}$$

where $y(x_n)$ is the equation (1), and A and B are the parameters found by minimizing a cross-entropy error function:

$$\min - \sum_n t_n \log(p_n) + (1 - t_n)\log\,(1 - p_n) \tag{6}$$

where p_i is equation (5). The method of Platt (1999) to obtain posterior probability is embedded in SVC of Scikit-learn, and it is activated in the implementation of SVC.

2.2. Implementation of SVC on a Simple Data

Jupyter notebook is employed to implement SVC. SVC is a fully developed module, and it offers solutions to some problems like unbalanced class problems. The critical parameters shaping models are regularization parameter C and gamma. The parameter C practically controls overfitting of the data. The parameter C with low values makes the model smooth, and C with high values may cause overfitting. Gamma defines how far a sample can have influence on the model. A low gamma makes the model general, and a large gamma can cause individuality of samples.

For the implementation of SVC, a 3-D pseudo drillhole data is created on which SVC is applied. The data is comprised of 56 vertical drillholes. Number of drillholes along East and North axes are 8 and 7 respectively. Each drillhole has 25 composite samples with 2 meter-length with total of 1400 samples in the data, and samples are classified under 2 classes: host rocks as -1 and ores as 1 (Figure 3). Summary statistics of the data can be seen in Table 2.

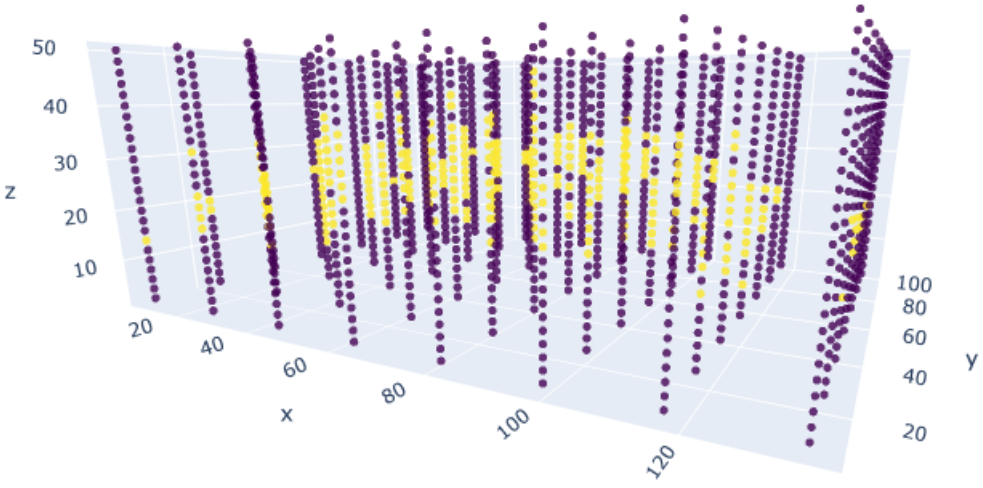


Figure 3 Distribution of pseudo data in the 3-D space. Data is illustrated using Plotly. Class 1 and -1 are shown with yellow and blue colors respectively.

Table 2 Summary statistics of the data

	Count	Mean	St. Dev.	Min.	25%	50%	75%	Max.
East	1400	72.00	41.258	9.0	40.5	72.0	103.5	135.0
North	1400	53.00	30.011	8.0	23.0	53.0	83.0	98.0
Elevation	1400	26.00	14.427	2.0	14.0	26.0	38.0	50.0
Value	1400	-0.55	0.833	-1	-1	-1	-1	1

The variables East, North and Elevation are utilized for training, and the variable Class is employed for target. The samples are split into train and test datasets to check accuracy of the model. Test ratio of the dataframe is set as 0.25. At the preprocessing stage, standardization is applied on the training data before fitting the data to the model. Class weight of SVC is set to balanced to consider the imbalanced ratio of classes while modelling. Normally, SVM is not capable of predicting probability, however SVC offers probability prediction using Platt's method (Platt, 1999). During modelling, prediction probability is also activated. To apply SVC, cross validation is applied first to determine ideal parameters. Cross validation is performed for the values and conditions expressed in Table 3. The optimum parameters determined for C and gamma are 0.1 and 5 respectively. However, C is increased slightly, and it was chosen as 1 to avoid high regularity effect.

Table 3 Cross-validation parameters of SVC

Parameters	***Cross-Validation***						
C	0.1	0.5	1	10	100	1000	
Gamma	5	2	1	0.1	0.01	0.001	0.0001

Time spent for the modelling was less than 1 second, meaning that training time is not a concern for SVC especially for such a simple dataset. The result of the model is both successful visually (Figure 4a) and accuracy according to confusion matrix (Figure 4b). Figure 4a shows distribution of probabilities and ore samples of drillholes (blue dots). As seen in Figure 4a high probability zones correspond to the ore samples. Both accuracy of class -1 and 1 are over 85% (Figure 4b), and overall accuracy is 88% which is quite promising.

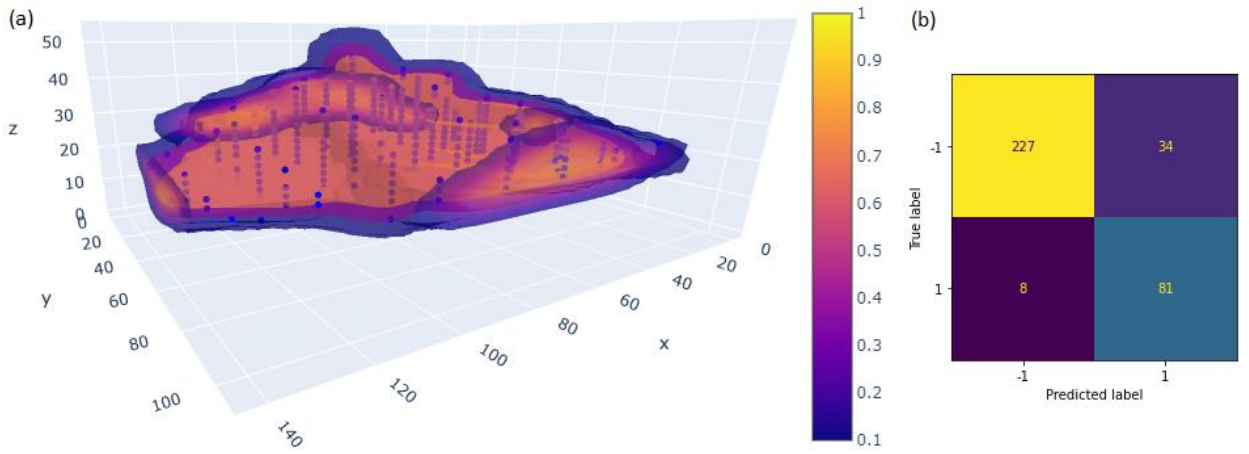


Figure 4 (a) 3-D visualization of probabilities and (b) confusion matrix of prediction results

To better investigate the model, cross-sections are created along North axis at each line of mesh grids. Two of the cross-sections overlain by drillholes are illustrated in Figure 5. Background color shows the map of probability. Blue and yellow crosses are host rock and ore samples of drillholes, respectively, and green circles show support vectors. For most of the areas, the model represents drillhole samples well. However, notice that there are some areas intended to be misclassified in terms of probability, which is desirable.

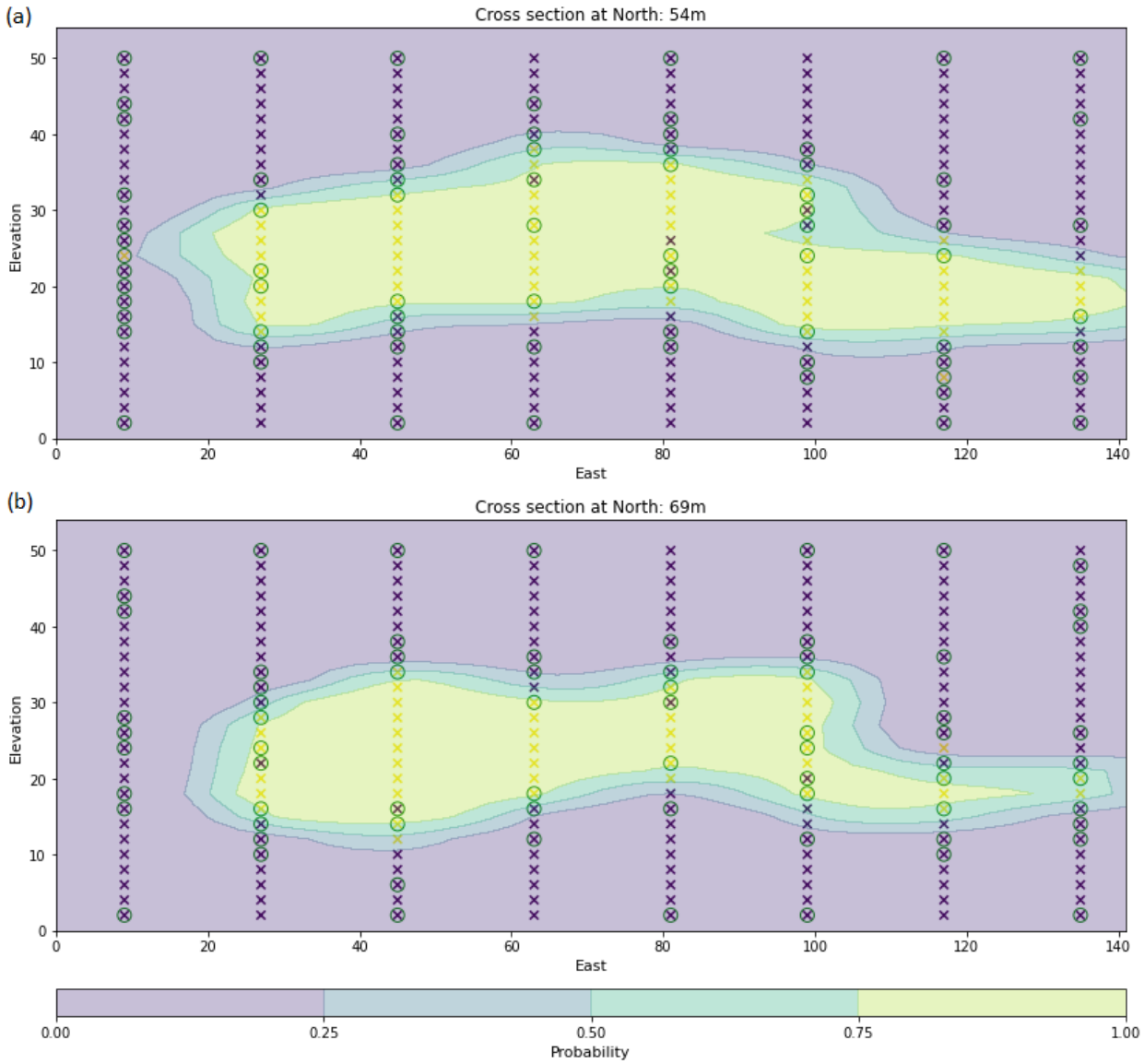


Figure 5 Cross-sections at North 54m (a) and 69m (b).

2.3. Scripts for Cross-sections

For illustration purposes, the library Plotly and Pyplot module of matplotlib are employed. Plotly is utilized for dynamic 3-D illustrations, and Pyplot is utilized for 2-D cross-sections. An automated script is written to create dataframes of a series of cross-sections (Figure 6). The dataframes created after the script shown in Figure 6 are utilized in a semi-automated script (Figure 7) which is written to plot cross-sections shown in Figure 5.

```
from math import modf
#Each mesh line has 3 meters distance to the neighbor mesh. Min is 0. Max is 105.
created_data=[]
for mesh in np.arange(0,105, 3):
    #Run1 and Run2 would be required if the mesh spacing was  represented by a fractional number.
    run1="intnum=modf({num})".format(num=mesh)
    exec(run1)
    run2="titleint=int(intnum[1])"
    exec(run2)
    #df_x is the dataframe which contains x, y and z coordinates of mesh grids and the corresponding probabilities.
    # Run 3, 4, 5 and 6 are required for the mapping of probabilities.
    #reshape(48,19) comes from number of mesh points created along the corresponding axis
    run3="df_N{titleint}=df_x.loc[(df_x['North']>=({num}-1)) & (df_x['North']<=({num}+1))]".format(num=mesh, titleint=titleint)
    exec(run3)
    run4="East_N{titleint} = df_N{titleint}['East'].to_numpy().reshape(48,19)".format(titleint=titleint)
    exec(run4)
    run5="Elevation_N{titleint} = df_N{titleint}['Elevation'].to_numpy().reshape(48,19)".format(titleint=titleint)
    exec(run5)
    run6="proba_N{titleint} = df_N{titleint}['1-prob'].to_numpy().reshape(48,19)".format(titleint=titleint)
    exec(run6)
    # Run7 creates the dataframe of samples located in the corresponding mesh along the North Axis
    run7="df_smple_{titleint}=df.loc[(df['North']>=({num}-1)) & (df['North']<=({num}+1))]".format(num=mesh, titleint=titleint)
    exec(run7)
    # Run8 creates the dataframe of support vectors located in the corresponding mesh along the North Axis
    # df_sv is the dataframe which contains x, y and z coordinates of support vectors.
    run8="sv_N{titleint}= df_sv.loc[(df_sv['North']>=({num}-1)) & (df_sv['North']<=({num}+1))]".format(num=mesh, titleint=titleint)
    exec(run8)
    # Run9 creates a list of dataframes created after running this script
    run9="created_data.append('df_N{titleint}')".format(titleint=titleint)
    exec(run9)
```

Figure 6 Script written to create dataframes of a series of cross-sections

```
#Assign an appropriate value to 'cross_section' to see an available Cross-section
#Drillhole samples are only available at cross-sections 9, 24, 39, 54, 69, 84 and 99.
cross_section=int(84)

plot_cs="""
plt.figure(figsize=(15,6))
plt.scatter(df_smple_{cross_section}['East'], df_smple_{cross_section}['Elevation'], s=40, c=df_smple_{cross_section}['Value'], marker="x")
plt.scatter(sv_N{cross_section}['East'], sv_N{cross_section}['Elevation'], s=100, facecolor="none", edgecolor="g")
plt.contourf(East_N{cross_section}, Elevation_N{cross_section}, proba_N{cross_section}.reshape(48, 19), np.linspace(0, 1, 5), alpha=0.3)
cbar=plt.colorbar()
cbar.set_label('Probability', fontsize=11)
plt.title('Cross section at North: {cross_section}m', fontsize=12)
plt.xlabel('East', fontsize=11)
plt.ylabel('Elevation', fontsize=11)
plt.gca().set_aspect("auto", adjustable="box")
""".format(cross_section=cross_section)
exec(plot_cs)
```

Figure 7 Script written to plot cross-sections

3. Discussion and Conclusion

Sci-kit learn offers a strong module for SVM. The documentation of SVC is very strong; therefore, all the answers can be found to any technical questions that arise during the application step. Moreover, the options like imposing the imbalance of classes, and posterior probabilities makes SVC valuable. The time spent on training was less than 1 second for a data set with 1400 samples, which is perfect for quick decision-making. Yet, considering training time increases exponentially, SVC still can be time-demanding for huge drillhole datasets. Again, the simple scenario was based on a binary class problem. When the number of targets is higher than 2, training becomes demanding for SVC since modelling scheme is not traditional like some other classifiers, e.g., decisiontreeclassifier. Therefore, multiclass problems can be demanding for SVC (Pedregosa et al., 2011). Although probabilities are accurate for the example, it is also important to notice that Sci-kit learn addresses that probability prediction has some issues (Pedregosa et al., 2011), therefore Platt's method can need verification and validation in real and more complex scenarios.

SVC is successfully applied to simple data. However, there are some points worth to mention about choosing the correct parameters and functions. Regularity of the model is an important concern. An overfitted model can be deceptive. Therefore, the regularity parameter was kept slightly higher than the optimum value obtained by cross-validation to prevent overfitting. Moreover, how the accuracy is determined for cross-validation is also important. A balanced accuracy can be more informative in terms of determining the optimum parameters. Another point, RBF as a kernel function is robust, however other kernel functions can also be effective under different scenarios. Lastly, analyzing and observing the results are at least as important as the accuracy of the model. To better observe the model, cross-sections are created along North axis. However, cross-sections created with different angles and directions can be required for some other cases. Therefore, either more developed cross-section modules can be utilized for the complex scenarios, or the scripts shared in the articles can be improved.

Overall, SVC proved that it can model a simple dataset. However, it is also important to notice that the application of SVC on a real drillhole data can be much more demanding.

4. References

Bishop, C.M., 2006. Kernel Methods & Sparse Kernel Machines. Pattern Recognition and Machine Learning. Springer Science+Business Media, LLC, Singapore.

Géron, A., 2017. Support Vector Machines. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow., 2nd Edition, O'Reilly Media, Inc.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, vol. 12, pp. 2825-2830.

Platt, J.C., 1999. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. MIT Press.

Relevance Vector Machines: An Introduction[1]

Koruk, Kasimcan (kasimcan.koruk@queensu.ca)

Abstract

Support Vector Machines (SVM) is a well-known machine learning method that has been available for many years. Although SVM offers a strong solution for machine learning problems combining generalization properties with sparse kernel technique, it normally does not provide posterior probabilities. As an alternative, Relevance Vector Machine (RVM) offers a Bayesian formulation to classification and regression problems. RVM is a promising machine learning method and open to new developments. This article reviews some basic principles of RVM, and it summarizes advantages and disadvantages of the method in comparison to SVM. The article also compares RVM and SVM according to the results of applications on a real drillhole dataset. The applications shows that the biggest challenge of RVM to be overcome is training time for huge datasets.

1. Introduction

Over-fitting is generally the main challenge for classification problems. SVM is one of the valid supervised learning methods which can handle over-fitting with minimum misclassification and maximum possible margins thanks to its generalization properties with sparse kernel technique (Tipping, 2001). SVM has become a popular method with several application examples in the literature (Tipping, 2000; Géron, 2017). With the help of support vectors utilized in decision functions, SVM provides sparsity to the solutions of machine learning problems.

Although SVM is a strong decision machine, it does not output posterior probabilities and the sparsity of SVM is limited, because the number of support vectors can increase linearly as the number of training data increases (Tipping 2000). As an alternative, Relevance Vector Machine (RVM) offers sparser solutions, and more importantly it offers a Bayesian formulation to classification problems (Bishop, 2006). RVM principally possesses the structure of SVM with some modifications. The article underlines the modifications and summarizes the differences between RVM and SVM in terms of formulation. The article also compares RVM and SVM according to the results of applications on a real drillhole dataset.

2. Relevance Vector Machines

2.1. Theoretical Review of RVM

Fundamentals of RVM are presented by Tipping (2000). RVM is fundamentally the same as SVM considering the functional form:

[1] Cite as: Koruk K (2021) Relevance Vector Machines: An Introduction, Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report 2021, paper 2021-08, 88-92.

$$y(x) = \sum_{n=1}^{N} w_n k(x, x_n) + b \quad (1)$$

The main difference is the introduction of a new hyper-parameter α which is assigned to each weight vector w_n. Posterior probability of the targets, t or y(x), is given by

$$p(t|X, w, \beta) = \prod_{n=1}^{N} p(t_n|x_n, w, \beta^{-1}) \quad (2)$$

where $\beta=\sigma^{-2}$. With these newly assigned hyper-parameters α, the posterior probability of the weight takes the form

$$p(w|\alpha) = \prod_{n=1}^{M} N(w_i|0, {\alpha_i}^{-1}) \quad (3)$$

After an initial value to hyper-parameters α and β, RVM predicts the probability iteratively. With each iteration, α and β are aimed to be maximized. As the hyper-parameters approximate to maximum values, the weights approximate to zero mean and covariance, and thus become redundant on the probability prediction. The rest of the vectors with non-zero weights control the model, and they are called 'relevance vector'. Unlike SVM, relevance vectors are not necessarily located on the boundary (Figure 1).

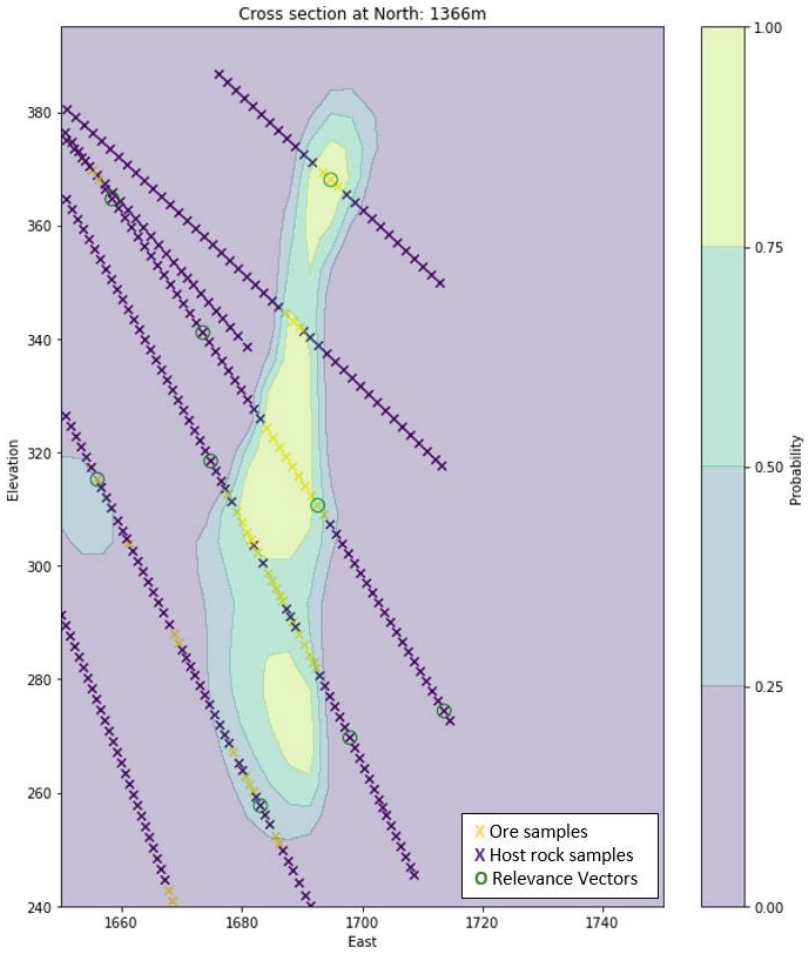


Figure 1 A cross-section of RVM model showing Bayesian probability overlain by training samples (composite drillhole samples)

While RVM looks more promising compared to SVM, there are some theoretical drawbacks of RVM worth to mention. All advantages and disadvantages of the RVM are listed in table below.

Table 1 Comparison of RVM and SVM

Advantages
RVM can make probability prediction.
Better than SVM when number of classes is more than two.
No need for cross-validation because there is no regularization parameter C.
Fewer decision functions because most of the weights of samples approximate to zero.
Disadvantages
More time for the training step.
Computation cost increases exponentially as number of classes increases.

2.2. Application of RVM and SVM

RVM and SVM are applied on a real drillhole data to make a tangible comparison. The comparison was based on the time spent on the training and accuracy of prediction results. Accuracy of the probabilities are left aside to further studies. The information about drillhole data is kept simple for the

sake of the privacy of the project. The data is huge with more than 10000 binary-class samples. Ore samples are classified as 1, and the rest of the samples are classified as -1. When RVM is applied on the data directly, the training step lasts for hours and even days. To decrease the time spent on the training step, RVM and SVM are applied on a specific zone of the data narrowed to 1616 samples. The ratio of ore samples over the total samples is 12.19%. The data are split into training and test groups to assess the accuracy with 25% ratio.

SVM is applied on the data using SVC module of Scikit-learn (Pedregosa et al., 2011). The critical parameter shaping a model is gamma and regularization parameter C. Gamma defines how far a sample can have influence on the model. A low gamma makes the model general, and a large gamma can cause individuality of samples. The parameter C with low values makes the model smooth, and C with high values may cause overfitting. To apply SVC, cross validation is applied first to determine ideal parameters. Cross validation is performed to determine optimum parameters, then C and gamma are determined as 1000 and 5, respectively. Moreover, imbalance between the classes is taken into account using an option embedded in the module. RVM trials are done using the module EMRVC of the library sklearn-rvm 0.1.1, compiled by a university research group of King's College London. The project group claims that the library is compiled according to the implementation of Tipping (2000), and they adapted the API of scikit-learn to the module. As stated in the previous section, there is no parameter C in RVM. Gamma is determined equal to 5 as the final decision.

The results are illustrated in confusion matrices (Figure 2). The result of SVC (Figure 2b) is relatively better than that of EMRVC (Figure 2a), considering accuracy of ore samples. Predictions on ore samples increases up to 76% (Figure 2c) when the imbalance is imposed on the SVC model, at the cost of slightly losing accuracy of host rock prediction. On the other hand, SVC showed clearly better performance compared to EMRVC in terms of training time. Even when the cross-validation is considered, the time spent on the training was only 5 seconds for the case of SVC. However, training took 6 minutes to finalize the training using EMRVC.

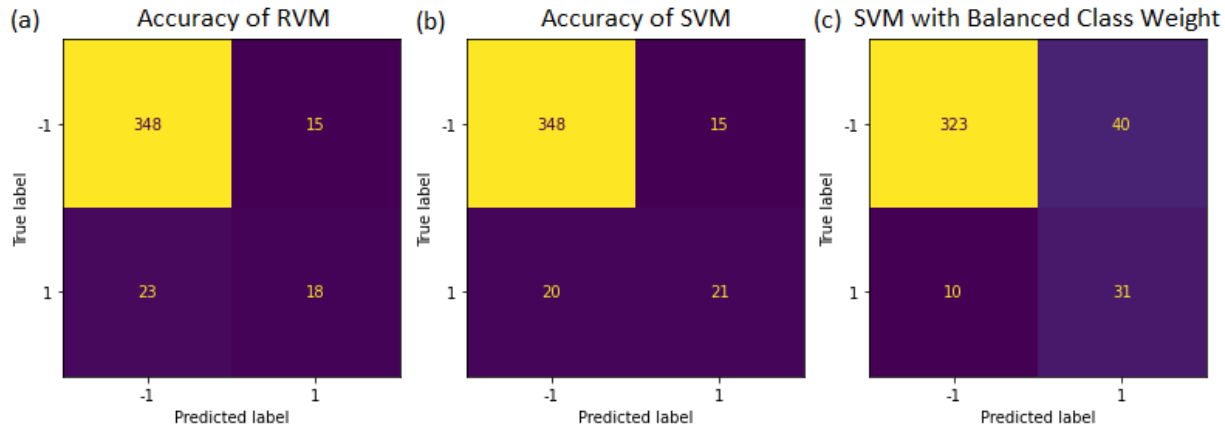


Figure 2 Accuracy of the results of (a) RVM, (b) SVM and (c) SVM with balanced class weight shown on confusion matrices

3. Discussion and Conclusion

RVM is theoretically a promising technique since it offers substantial developments to SVM. The sparsity of RVM is noteworthy considering the sparsity of SVM is limited. The absence of regularization parameter frees RVM from cross-validation step, which can be a time-consuming step for machine

learning methods. Probabilistic outputs make RVM again valuable in the world of machine learning. Although RVM offers the advantages specified above, RVM has some practical obstacles to overcome against SVM. The application showed that training time spent by RVM is roughly 50 times more than the time spent by SVM. This problem is one of the most important obstacles to overcome for RVM. While the regularization parameter C requires cross-validation, it also offers higher control to the SVM which is not the case for RVM. The necessary changes on C made SVM model slightly more accurate compared to RVM. Lastly, while the SVC of Sci-kit learn is a fully developed module with detailed documentation, EMRVC of sklearn-rvc is a module developed by a university project group and requires some improvements. Because SVC offers solution for imposing the imbalance of classes, much higher accuracy is obtained in terms of predicting ore samples. Developing the option of balanced class prediction is especially necessary for the real-world datasets.

In conclusion, RVM has serious obstacles to overcome, and it requires some serious improvements before it displaces the position of SVM in the world of machine learning.

4. References

Bishop, CM, 2006. Kernel Methods & Sparse Kernel Machines. Pattern Recognition and Machine Learning. Springer Science+Business Media, LLC, Singapore.

Géron, A., 2017. Support Vector Machines. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow., 2nd Edition, O'Reilly Media, Inc.

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, vol. 12, pp. 2825-2830.

Platt, JC., 1999. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. MIT Press.

Tipping, Michael E., 2000. The relevance vector machine. Advances in neural information processing systems, pp. 652 – 658.

Tipping, Michael E., 2001. Sparse Bayesian Learning and the Relevance Vector Machine. Journal of Machine Learning Research 1, pp. 211-244.