Les Cahiers du GERAD ISSN: 0711–2440

A simultaneous stochastic optimization framework for selecting additional infill drilling locations

Z. Levinson, R. Dimitrakopoulos

G–2022–60

December 2022

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2022
– Bibliothèque et Archives Canada, 2022

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2022
– Library and Archives Canada, 2022

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

A simultaneous stochastic optimization framework for selecting additional infill drilling locations

Zachary Levinson [a]

Roussos Dimitrakopoulos [a, b, c]

[a] *COSMO – Stochastic Mine Planning Laboratory, McGill University, Montréal (Qc), Canada, H3A 2A7*

[b] *Department of Mining and Materials Engineering, McGill University, Montréal (Qc), Canada, H3A 2A7*

[c] *GERAD, Montréal (Qc), Canada, H3T 1J4*

zachary.levinson@mcgill.ca
roussos.dimitrakopoulos@mcgill.ca

December 2022
Les Cahiers du GERAD
G–2022–60

Abstract : An innovative simultaneous stochastic optimization approach is proposed that combines actor-critic reinforcement learning and stochastic mathematical programming techniques to determine infill drilling locations in a mining complex. In strategic mine planning under uncertainty, the long-term production schedule is designed to define the extraction sequence, destination policy, and processing stream that maximize net present value and minimize deviations from production targets. The optimized decisions are driven by the inputs; a set of stochastic orebody simulations that describe the materials in the ground; and the design of the engineering system used to transform extracted materials into valuable products. The following work investigates the value of information by determining if additional infill drilling would lead to changes in the long-term production schedule by adapting the schedule using simultaneous stochastic optimization. Additional data is collected by infill drilling and the samples are used to update the stochastic simulations of the material attributes. A case study at a copper mining complex is completed to test the simultaneous stochastic approach for selecting infill drilling locations and demonstrates a 5.7% increase in net present value given a $1 M budget.

Keywords : Additional drilling, simultaneous stochastic optimization, reinforcement learning, mining complex, value of information

1 Introduction

An industrial mining complex is a multifaceted engineering system where raw materials are extracted from several mines, processed, and transported to be sold to customers and the market. Recent advancements in stochastic mathematical programming have demonstrated the ability to simultaneously optimize the long-term production schedule of a mining complex under uncertainty (Goodfellow & Dimitrakopoulos, 2016, 2017, 2015; Montiel & Dimitrakopoulos, 2015, 2018). The production schedule defines the long-term extraction sequence, destination policy, and processing stream decisions using a single mathematical model that maximizes net present value and manages risk. The production schedule that is obtained using simultaneous stochastic optimization requires two significant components: first, a mathematical model of a mining complex that transforms raw materials into saleable products and second, the input stochastic orebody simulations that provide a geostatistical representation of the materials in the ground. The input stochastic orebody simulations are a critical aspect of the optimization as the location, uncertainty, and quality of the material in the ground is the primary driver of the optimization.

A major challenge in the mining industry is determining whether it is worthwhile to drill in certain areas to gather additional information and update the input stochastic orebody simulations prior to optimization. Collecting information about a mineral deposit is commonly accomplished by infill drilling. Data collected from unsampled locations is used to update the stochastic orebody simulations. However, determining the value of different drilling locations and configurations is challenging. This is because in long-term strategic mine planning additional drilling information is only valuable if it results in changing the schedule to improve financial forecasts and overcomes the cost of collecting data by adapting the schedule with new information. This work investigates an extension of the simultaneous stochastic optimization framework that combines actor-critic reinforcement learning and stochastic mathematical programming to select infill drilling locations that will lead to changes in the long-term production schedule and increase the value of a mining complex.

An actor-critic reinforcement learning approach with deep neural networks (Lillicrap et al., 2016; Sutton & Barto, 2018) is used to find a policy for selecting infill drilling locations given a set of features obtained directly from the long-term production schedule. The neural network policy, an input-output mapping, guides the optimization process to relevant drilling locations with a stochastic search of the solution space, learning through continuous trial-and-error. Each time a drillhole is selected, the long-term production schedule is simultaneously optimized using the additional drilling information collected by sampling a single simulation of the mineral deposit. The process is repeated using different simulations to ensure that stable drilling areas are found independent of the simulations selected. A reward is computed by calculating the objective function of the stochastic mathematical program used for simultaneously optimizing the production schedule given additional information. Then, the objective of the original schedule is recalculated using the additional information and the difference is found. The difference in the objective function is the numerical reward. The reward is used to update the neural network parameters by strengthening those actions that improve performance. Several recent works have demonstrated the successful performance of deep neural networks for a variety of problems by providing a rich representation of the input features for action selection (Kumar et al., 2020; Lillicrap et al., 2016; Mnih et al., 2015; Silver et al., 2017). This is the primary reason for investigating the selection of infill drilling locations using a deep neural network.

The method proposed considers the effect of local grade variability and uncertainty on the long-term production schedule using stochastic orebody simulations. Several approaches for optimizing infill drilling locations have used stochastic simulations of the material grades, each with a different objective. Dirkx and Dimitrakopoulos (2017) define an optimal infill drilling pattern in a stockpile based on changes to material classification using a multi-armed bandit approach. Boucher et al. (2005) use simulations of correlated variables to observe the impact of drilling on the profits obtained and evaluate different stockpiling options. Alternate approaches, without the use of geostatistical simulations, suggest using a random search to convert inferred resources into indicated resources or indicated into measured based on a geometric criterion (Jewbali et al., 2017). The common limitation of each of these methods is that they do not consider the relationship between the new information collected and the resulting optimized long-term production schedule. Instead, conversion indicators, material misclassification, and resource/reserve impacts are considered that do not address potential changes to the long-term schedule and related financial forecasts. Froyland et al. (2018) discuss this relationship and highlight that the value of information can only be realized if changes are made to the long-term schedule. The simultaneous stochastic optimization framework for drillhole selection must consider how it changes the schedule to ensure future value can be obtained, as misclassification and other indicators may not lead to differences in the schedule leading to unnecessary drilling expenses.

This work presents a simultaneous stochastic optimization approach that combines actor-critic reinforcement learning with stochastic mathematical programming and extends the simultaneous stochastic optimization to consider the selection of infill drilling locations in a mining complex. In the following sections, the actor-critic reinforcement learning framework and combined simultaneous stochastic optimization approach is described. This is followed by a case study at a copper mining complex that demonstrates the impact of selecting a set of infill drilling locations using this innovative approach. Lastly, conclusions and recommendations for future work are provided.

2 Method

The extended simultaneous stochastic optimization approach for selecting infill drilling locations is presented in Figure 1. The method is described in a sequence of steps. First, a set of stochastic orebody simulations are used as input into a simultaneous stochastic optimization framework to represent the uncertain supply of material (Boucher & Dimitrakopoulos, 2009; Godoy, 2002). An optimized long-term production schedule is obtained including the extraction sequence, destination policy, and processing stream decisions that maximize net-present value and manage technical risk. For further details on the simultaneous stochastic optimization of a mining complex interested readers are referred to the following reference (Goodfellow & Dimitrakopoulos, 2016).

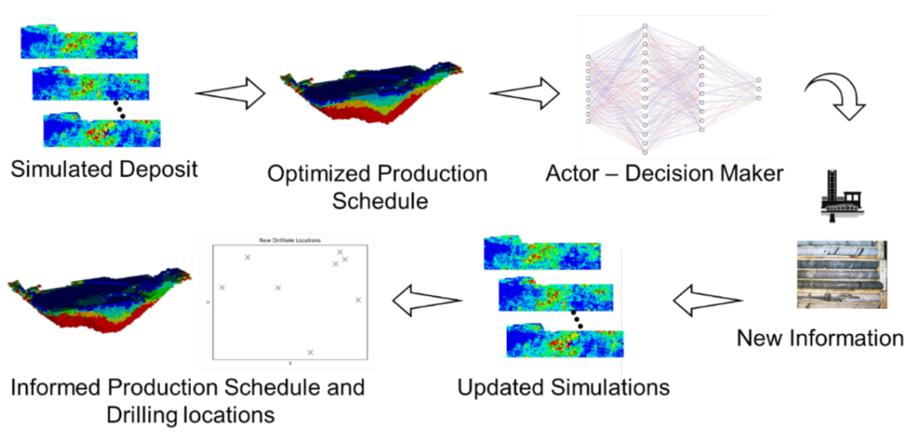


Figure 1: Primary workflow (i) generate a set of equiprobable stochastic simulations; (ii) simultaneous stochastic optimization of a mining complex; (iii) sample drillhole from stochastic simulation (iv) ensemble Kalman filter update; (v) informed drilling locations and resulting schedule

After producing the simulations and optimal schedule given the current information, the long-term production schedule is used to obtain a set of criteria for action selection. Further information can be taken from operational data and the simulated orebody realizations. The information is presented to a trained neural network (an actor) that decides which additional drilling locations are likely to increase the performance of the long-term production schedule using a learned input-output mapping. Given the drilling location selected, samples are obtained from a stochastic simulation of the mineral deposit. These samples are used to update the simulations by applying ensemble Kalman filter (EnKF) as additional information becomes available (Benndorf, 2020; Evensen, 1994, 2003; Kumar & Srinivasan, 2019). The simultaneous stochastic optimization approach is then used to generate a new informed production schedule and provides the final output; a set of drillhole locations that maximize the value of additional information under the constraint of a budget. Further details of each step in the optimization process are discussed subsequently. In addition, it should be noted that several different stochastic simulations are used to represent the drilling data to ensure similar locations are drilled in a series of separate runs of the framework.

2.1 Criteria for action selection

The criteria for selecting infill drilling locations are based on features derived directly from the optimized production schedule, orebody simulations, and readily available information on site. The reinforcement learning framework is provided a numerical representation of previously drilled locations, the period of extraction that the drillhole will intersect, and areas of higher or lower grade variability. These features are mapped to an output drillhole location using a neural network function approximator. Through continuous trial and error and policy updates, the most beneficial drilling locations can be learned by optimizing the parameters using a stochastic gradient method and a performance function (Kingma & Ba, 2015). During training, the actor learns which aspects from the input criteria affect the position of the selected drillhole. The input criteria for drillhole selection are generalized to determine relevant areas to drill within the mining complex based on the material uncertainty not in the blocks themselves but on the distribution of the metal generated downstream at the processing facilities during each period of production. This incorporates scheduling aspects that may not be specific to a block but, also relate to the spatial distribution of grades within the deposits to be mined in a single production period, which may indicate areas to target for infill drilling. For example, years with higher material variability in terms of metal production may be targeted more or less heavily depending on how this impacts the production schedule performance.

2.2 Updating simulations using Ensemble Kalman Filter (EnKF)

An EnKF updating framework is applied. Simulations are updated using new information collected from drilling and the prior stochastic orebody simulations (Figure 2). The magnitude of the update is based on the error between the simulated realizations and observed drillhole data. Benndorf (2020) describes the EnKF updating framework for mineral resource applications in detail. Drilling information is generated by sampling a randomly selected stochastic simulation of the mineral deposit at the drillhole location selected; the simulated data is used to represent the drilling sample data in a real deposit. A different randomly sampled simulation of the mineral deposit is used to represent the drillhole data over several runs ensuring that within each run of the optimization the drillholes selected are within similar areas.

The EnKF framework provides an efficient process for updating the simulations of the deposit without re-simulating the entire deposit. A spatial random field of the attribute of interest is denoted $\mathbf{Z}(\mathbf{x})$ where each element $\mathbf{Z}(\mathbf{x}_j)$ is a random variable associated with a block position at location $\mathbf{x}_j$ for $j = \{1, \ldots, N\}$. Each realization of the spatial random field and the corresponding material grade is denoted $\mathbf{z}^* (\mathbf{x})_{t,s}$. This provides the predicted value at location $\mathbf{x}_j$ given the information gathered up to timestep t and the simulated realization s, where $t = \{1, \ldots, T\}$. A timestep represents each time new information becomes available by selecting a new drillhole location. The spatial random field is limited to a neighbourhood of blocks in the deposit as it is only desirable to update parts of the realization neighbouring the newly collected additional information. A T by N matrix $\mathbf{A}$ represents the contribution of each new drillhole sample up to the present timestep t. The rows of matrix $\mathbf{A}$ represent the contribution of each drillhole selected to the posterior simulated orebody realizations, which is a matrix of zeros and ones. When one of the N potential drilling locations are drilled all the

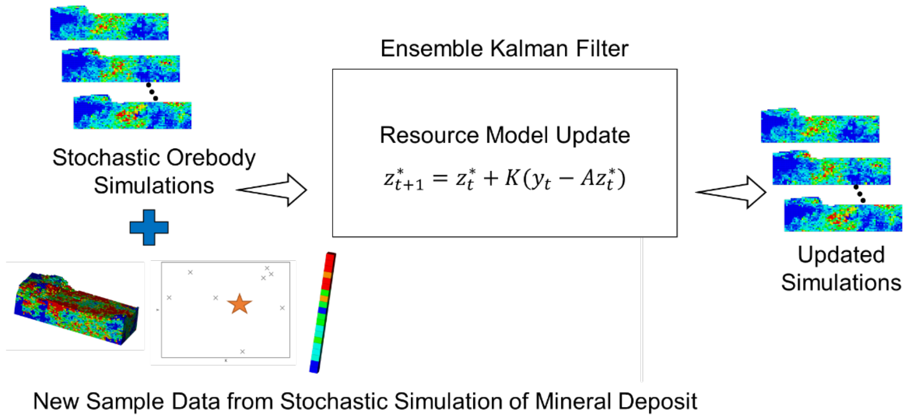


Figure 2: Ensemble Kalman Filter approach for updating stochastic simulations with additional drilling information

locations intersected by drilling vertically contribute to the update and are included as a one in the matrix $\mathbf{A}$. Each block location $\mathbf{x}_j$ in the neighbourhood is then updated using the following update:

$$\mathbf{z}^*(\mathbf{x})_{t+1,s} = \mathbf{z}^*(\mathbf{x})_{t,s} + \underbrace{\mathbf{C}_{t,t}\mathbf{A}^T\left(\mathbf{A}\mathbf{C}_{t,t}\mathbf{A}^T + \mathbf{C}_{v,v}\right)^{-1}}_{\text{Kalman Gain}}\underbrace{\left(\mathbf{y}_{t,s} - A\mathbf{z}^*(\mathbf{x})_{t,s}\right)}_{\text{Innovation}} \tag{1}$$

where $\mathbf{z}^*(\mathbf{x})_{t+1,s}$ is the updated attribute for each realization s after additional information is collected in timestep t. The covariance matrix $\mathbf{C}_{t,t}$ is a N *by* N matrix approximating the auto- and cross covariance of the random field between each element $\mathbf{Z}(\mathbf{x}_j)$. The Kalman gain matrix depicted in Equation 1 represents the weighting factor that determines the contribution of the update to those neighboring locations considered in the update. This includes the covariance matrix $\mathbf{C}_{v,v}$ of the measured error in the sample. The second component in the update is the innovation, which represents the error between the predicted outcome of the prior model and the observed drillhole value $\mathbf{y}_{t,s} = \mathbf{y}_t + \epsilon_s$. The observation $\mathbf{y}_t$ is the data value collected through drilling and a small noise term is added to consider measurement error. The ensemble Kalman filter approach uses the set of simulated scenarios to approximate the covariance matrix and increase computational efficiency. This updating framework is applied to the set of stochastic simulations following each new drillhole selection.

2.3 Training the actor-critic reinforcement learning agent

An off-policy actor-critic reinforcement learning agent with deep neural network function approximators is applied to optimize the drillhole selection. This framework has been selected to determine actions in a high dimensional continuous action space. The method is based off a reinforcement learning agent designed for continuous control (Lillicrap et al., 2016). A typical reinforcement learning problem is constructed by allowing an agent or decision maker to interact with a mining complex (environment) E over a discrete set of timesteps. During each timestep t the agent receives an observation from the mining complex $s_t = \phi(\text{sched}_t)$, given the current long-term production schedule (sched_t), and a continuous action $a_t \in \mathbb{R}^3$ is chosen to determine the coordinate of the drillhole (X, Y) and if drilling should occur at that location within the mining complex. The agent then receives a numerical reward r_{t+1} from the mining complex based on the action taken. The learning framework is illustrated in Figure 3.

Actor-critic reinforcement learning uses an agent composed of two components: an actor and a critic. The actor learns a parametrized policy that maps the current state of the mining complex s_t to an action $a_t = \mu(s_t|\theta^\mu)$ and controls the decision-making process where θ^μ denotes the parameters

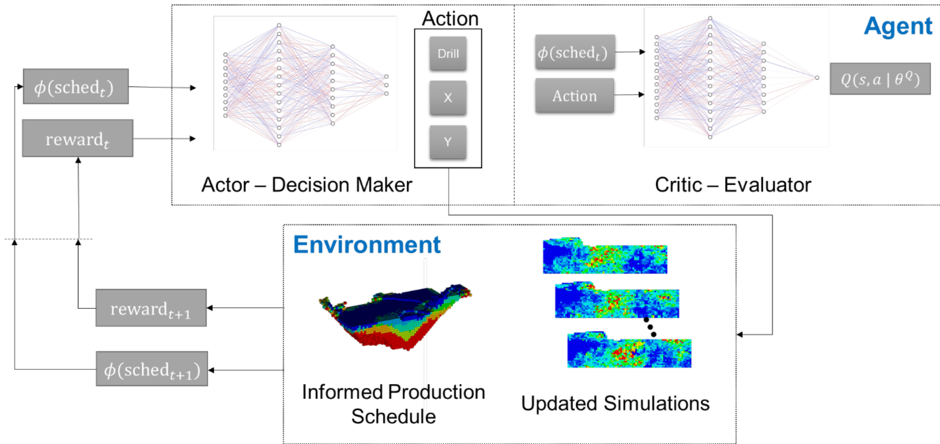


Figure 3: Actor-critic reinforcement learning for infill drillhole selection

of the policy network. Similarly, the critic learns the action-value function $Q\left(s_t, a_t|\theta^Q\right)$ through a second set of parameters θ^Q. The action value-function is used to approximate the expected value of taking an action a_t in state s_t parametrized by θ^Q. Upon initialization, the actor $\mu\left(s_t|\theta^\mu\right)$ and critic network $Q\left(s_t, a_t|\theta^Q\right)$ are randomly initialized with weights θ^μ and θ^Q, respectively. A duplicate of each of these networks are created as a target network to stabilize learning denoted Q' and μ' with weights such that $\theta^{\mu'} \leftarrow \theta^\mu$ *and* $\theta^{Q'} \leftarrow \theta^Q$. Lastly, a replay buffer denoted $\mathcal{R}$ is initialized. The replay buffer is used to train the network with an off-policy learning method. As the agent interacts with the mining complex, each experience it obtains is stored as a tuple (s_t, a_t, r_t, s_{t+1}) in the buffer that contains the information from the most recent interaction with the mining complex. These experiences or transitions are randomly sampled after enough experience is collected to minimize the correlations between samples and are used to update the neural network parameters.

Each episode of learning is a sequence of states, actions, and rewards that end in a terminal state. In the mining complex, the actor decides which locations to infill drill based on the input features and proceeds until the terminal state is reached. The terminal state can occur when the agent decides to no longer continue drilling or a budgetary constraint is exceeded. A random process $\mathcal{N}$ introduces noise to each action to provide adequate exploration of the solution space. This is defined by the exploratory policy $\mu'\left(s_t\right) = \mu\left(s_t|\theta^\mu\right) + \mathcal{N}$. After each episode, the random process $\mathcal{N}$ is re-initialized for exploration. In addition, the initial state of the environment is received by calculating a set of features $s_t = \phi(sched_t)$ derived from the optimized production schedule $sched_t$ and considering the information available prior to additional drilling.

During each timestep within an episode:

1. An action $a_t = \mu\left(s_t|\theta^\mu\right) + \mathcal{N}_t$ is selected based on the current policy network and exploration noise.
2. The agent selects the nearest drillhole location to the output coordinates given by the agent's action a_t. If drilling occurs, the drillhole is sampled and used to update a set of simulated realizations $\mathbb{S}_t$ using EnKF. This results in a new set of simulated realizations $\mathbb{S}_{t+1}$ updated to account for the information collected through infill drilling. (Note, if the agent decides not to drill part three is skipped and the reward $r_t = 0$ for the transition as drilling is not completed. The episode is then terminated and given the original state of the mining complex).
3. After drilling, a reward r_t is generated by simultaneously optimizing the mining complex given new information. The objective function used for optimizing the long-term production schedule maximizes net present value (NPV) and minimizes deviations from production targets by applying a set of penalties (PEN), see Appendix A. Considering this objective, the reward is

computed by determining the difference in the objective function between the optimized long-term production schedule given new information (Sched_{t+1}) and the schedule obtained prior to collecting new information (Sched_t):

$$r_t = \mathbb{E}_{\mathbb{S}_{t+1}}\left[NPV - PEN|\text{Sched}_{t+1}\right] - \mathbb{E}_{\mathbb{S}_{t+1}}\left[NPV - PEN|\text{Sched}_{t}\right] - \text{drilling cost} \tag{2}$$

which accounts for the additional cost of drilling and considers the most up-to-date information available ($\mathbb{S}_{t+1}$). The state s_{t+1} can then be reconstructed using the optimized schedule that considers new information and the updated stochastic simulations. Notice, the optimized schedule given new information will likely change due to the difference in input and a positive reward can only be realized if the improvement in the objective function overcomes the cost of additional drilling.

4. Each transition in the environment is stored in a replay buffer $\mathcal{R}$ as a tuple (s_t, a_t, r_t, s_{t+1}) and a small batch of P transitions are sampled $(s_i, a_i, r_i, s_{i+1}) \sim \mathcal{R}$ to facilitate learning. The bootstrapped return is calculated using the target network that combines the reward received in timestep i and the predicted action-value provided by the critic target network:

$$y_i = r_i + Q^{'}\left(s_{i+1}, \mu^{'}\left(s_{i+1}\middle|\theta^{\mu^{'}}\right)\middle|\theta^{Q^{'}}\right) \tag{3}$$

5. The critic and policy network are updated by minimizing the loss L and using the sampled policy gradient $\nabla_{\theta^\mu} J$ to improve the policy, respectively.

$$L = \frac{1}{P}\sum_i \left(y_i - Q\left(s_i, a_i|\theta^Q\right)\right)^2 \tag{4}$$

$$\nabla_{\theta^\mu} J \approx \frac{1}{P}\sum_i \nabla_a Q\left(s, a|\theta^Q\right)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu}\mu\left(s|\theta^\mu\right)|_{s=s_i} \tag{5}$$

6. Lastly, the target networks are then updated using the updated policy and critic network parameters:

$$\theta^{Q^{'}} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q^{'}} \tag{6}$$

$$\theta^{\mu^{'}} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu^{'}} \tag{7}$$

where $\tau \in (0, 1]$ is the weighting factor that gradually updates each target network to be closer to current network parameters. This process is repeated until the agent reaches convergence, or the maximum number of episodes are reached.

2.4 Testing the agent

After learning is completed, the trained actor is used to select additional infill drilling locations in a mining complex. The stochastic orebody simulations of the deposit prior to additional information are provided as input along with the stochastically optimized production schedule prior to additional information. The random process noise used for action exploration is eliminated, and drillholes are selected until the budget constraint is reached or the trained actor decides it is no longer valuable to continue drilling. The resulting output is the drillhole coordinates that lead to the largest improvement in net present value and mitigated risk based on the agent's previous learnings. The learning process is repeated several times with different stochastic simulations of the mineral deposit for sampling drillholes to ensure stability of the results and that similar locations are found independent of the stochastic realization that is sampled.

3 Case study in a copper mining complex

The extended simultaneous stochastic optimization framework for selecting additional infill drilling locations is tested in an operating copper mining complex with a single open pit mine, several processing streams, stockpiles, and waste dump destinations. The objective of the optimization framework is to determine a set of locations to drill that maximize net present value and minimize the risk of deviating from production targets given a budgetary constraint. Figure 4 shows the configuration of the mining complex that the stochastic mathematical model considers and the allowable flow of materials from the mine to the customers and the market. A set of stochastic orebody simulations are used to represent the uncertain material source. The simulated attributes include total copper and soluble copper. There are two processing streams; a process plant and heap leach destination that produce copper concentrate and copper cathode products, respectively. In addition, a stockpile facility is included in the mining complex to defer material to be processed in later periods. The extraction sequence, destination policy, and process stream decisions are optimized using simultaneous stochastic optimization and the resulting production schedule is the initial input to the actor-critic reinforcement learning framework.

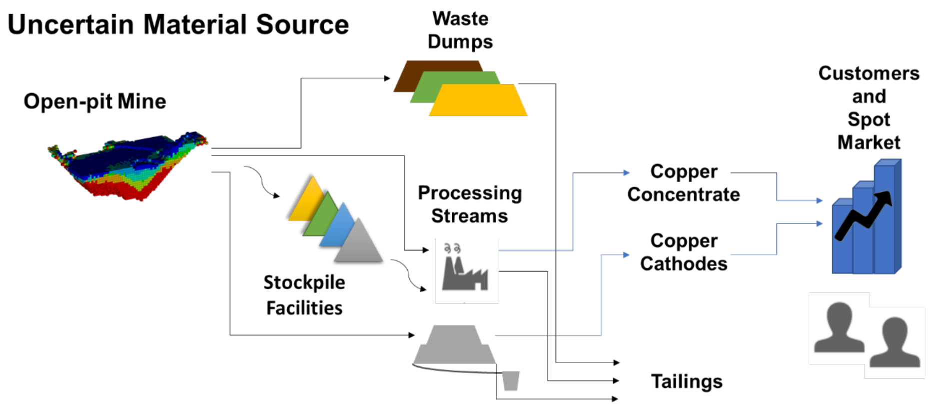
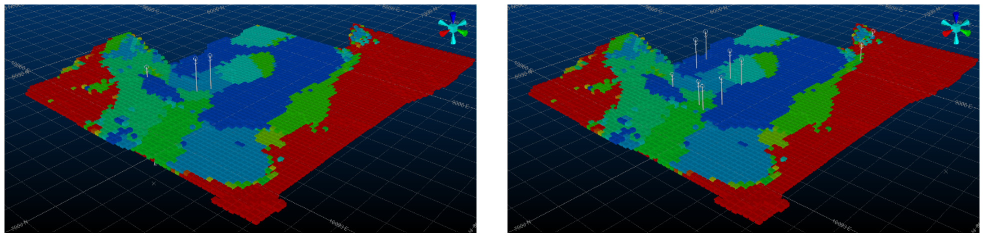


Figure 4: Copper mining complex

The case study considers a set of 10 stochastically simulated orebody realizations with a block size of 20 x 20 x 15 m^3 as input and 108,000 blocks. The long-term production schedule has a 10-year mine life. To account for the nature of the stochastic simulations of the mineral deposit, several different simulations are used to represent the drillhole samples in a real deposit. These simulations are different than the ones used for stochastically optimizing the mining complex. Lastly, there is a budgetary constraint of \$1 M and using this budget the algorithm can select from 3640 potential drilling locations. The objective is to maximize the expected net-present value while reducing the risk of deviating from production targets.

During training, the actor-critic reinforcement learning agent is left to interact with its environment to learn a set of infill drilling locations that provide the largest improvement in the long-term production schedule. Two example episodes of training the reinforcement learning approach are shown in Figure 5. The background image shows the initial production schedule generated using the information prior to additional infill drilling and the periods of extraction are indicated by the colors of each block. Marked using a white circular survey location are the different drilling locations determined by the actor during each episode in training. The total reward quantifies the improvement in the objective function obtained through collecting additional information. Additionally, the number of holes drilled in each episode are listed. It is important to notice that the number of drillholes changes between episode as the reinforcement learning agent also decides when to stop drilling. The cumulative reward obtained varies significantly depending on where drilling is commenced, and the potential value added by adapting the production schedule to new information. Finally, the trained actor is used to select the number of drillholes and appropriate locations. The stochastic orebody simulations are updated using EnKF and a new production schedule is created using simultaneous stochastic optimization.

In Figure 6, the resulting production schedule prior to new information (left) is compared to the production schedule with additional information (right). The primary findings reveal the extents of the final pit are similar between the two production schedules, however, the extraction sequence over



Episode	1	Episode	40
Total Reward	-3.5 x 10^6	Total Reward	155.9 x 10^6
Number of Drillholes	3	Number of Drillholes	10

Figure 5: Training examples

the 10-year schedule changes significantly in early periods. In the main areas of drilling, denoted with a light blue circle, the schedule changes due to additional drilling information adapting to an area of lower risk to generate further value for the mining complex. As a result of extracting this area earlier in the mine life other areas are pushed back to later periods. This is noticeable in the central part of the deposit as the simultaneous stochastic optimization approach adapts the schedule to manage production targets and increases NPV. The adapted schedule generates higher early cashflows accounting for the time value and increasing the NPV from \$3.5 B to \$3.7 B by adapting to the schedule to new information, a 5.7% or \$201 M improvement (Figure 7). Furthermore, the schedule impacts the total copper production at the processing plant over the life-of-mine, reducing the recovery of copper concentrate by 33 kt (Figure 8). However, significant value is obtained by adapting the production schedule to mine higher grade material earlier in the schedule resulting in higher metal production during the first two periods.

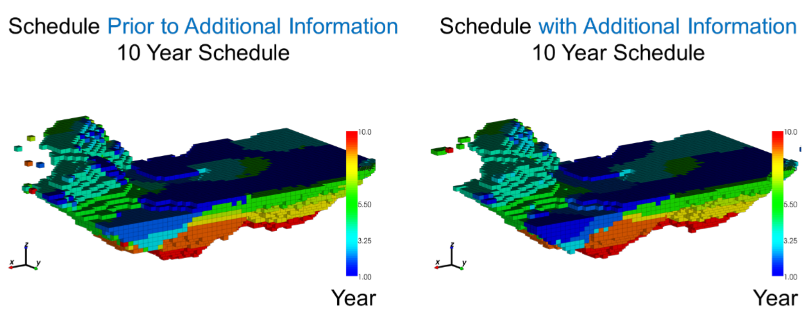


Figure 6: Production schedule: (left) prior to additional information; (right) after additional drilling

Lastly, five stochastic simulations of the mineral deposit are used for sampling the drillhole information in separate runs of the drillhole selection framework. The tests show that the method was insensitive to the simulation used to represent the real deposit as similar areas are drilled during each run of the optimization process (Figure 9). The drillholes selected are localized in the northeast portion of the pit and there is considerable overlap when using different simulated realizations to represent the sampled drillhole data. The drilling area is contained within a 250 m x 500 m area in the north-east corner of the open pit mine across all realizations used. Therefore, the drilling optimization method identifies the area to drill in a stable way that is not affected by using different or additional realizations.

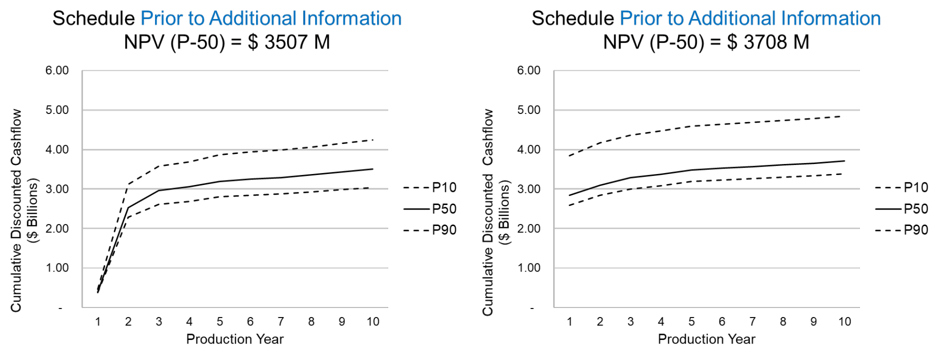


Figure 7: Cumulative discounted cashflows given the schedule: (left) prior to additional information; (right) after additional drilling

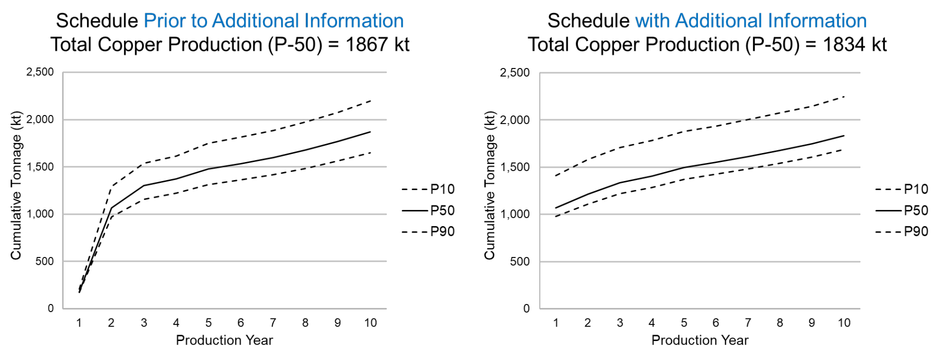


Figure 8: Total copper production given the schedule: (left) prior to additional information; (right) after additional drilling

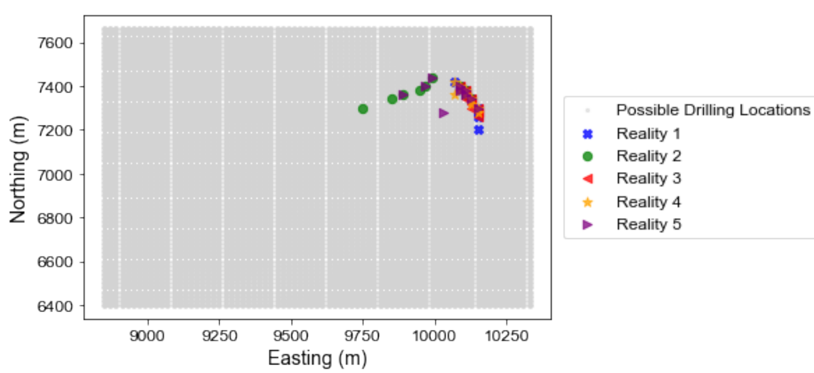


Figure 9: The selected infill drilling location using different simulated realizations of the mineral deposit

4 Conclusions

An extension to the simultaneous stochastic optimization approach is proposed that determines additional infill-drilling locations in a mining complex. The combined reinforcement learning and stochastic mathematical programming framework is tested in an operating copper mining complex. The results demonstrate a $201 M or 5.7% increase in net present value given a $1 M investment in additional drilling. Using the infill drilling selection framework and updating the simulations to account for additional information, the value of the mining complex is increasing significantly by adapting the production schedule to new information using simultaneous stochastic optimization. Future work may consider the possibility of using drillholes that can have different azimuth and dips to allow more flexibility in the drilling direction. In addition, supplementary information related to the deposit, production schedule and technical information could be used in the selection process to further inform the reinforcement learning approach on the most valuable locations to drill.

Appendix A

The simultaneous stochastic optimization framework outlined in Goodfellow and Dimitrakopoulos (2016) uses the following objective function:

$$\underbrace{\max \frac{1}{|\mathbb{S}|} \sum_{i \in \mathcal{S} \cup \mathcal{P} \cup \mathbb{M}} \sum_{t \in \mathbb{T}} \sum_{(h \in \mathbb{H})} \sum_{s \in \mathbb{S}} p_{h,i,t} v_{h,i,t,s}}_{\text{Discounted revenues and costs}} - \underbrace{\frac{1}{|\mathbb{S}|} \sum_{i \in \mathcal{S} \cup \mathcal{P} \cup \mathbb{M}} \sum_{t \in \mathbb{T}} \sum_{(h \in \mathbb{H})} \sum_{s \in \mathbb{S}} \left(c^{+}_{h,i,t} d^{+}_{h,i,t,s} + c^{-}_{h,i,t} d^{-}_{h,i,t,s} \right)}_{\text{Risk discounted penaties for deviations}} \quad (8)$$

The optimization objective maximizes the profit accounting for costs and revenues that are required to generate the products in the first term and minimizes the penalties from deviations from production targets using the second term. Further details regarding the notation can be found in the reference.

References

Benndorf, J. (2020) Closed Loop Management in Mineral Resource Extraction (S. Cham (ed.); 1st ed.). Springer Nature

Boucher, A. and Dimitrakopoulos, R. (2009) Block simulation of multiple correlated variables. Mathematical Geosciences, 41(2), 215–237 https://doi.org/10.1007/s11004-008-9178-0

Boucher, A., Dimitrakopoulos, R. and Vargas-Guzmán, J. A. (2005) Joint simulations, optimal drillhole spacing and the role of the stockpile. In O. Leuangthong and C. V. Deutsch (Eds.), Geostatistics Banff 2004 (pp. 35–44). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-3610-1_4

Dirkx, R. and Dimitrakopoulos, R. (2017) Optimizing infill drilling decisions using multi-armed bandits: Application in a long-term, multi-element stockpile. Mathematical Geosciences, 50(1), 35–52. https://doi.org/10.1007/s11004-017-9695-9

Evensen, G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research, 99(C5), 10–143,10–162. https://doi.org/10.1029/94jc00572

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. Ocean Dynamics, 53(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Froyland, G., Menabde, M., Stone, P. and Hodson, D. (2018) The value of additional drilling to open pit mining projects. In: R. Dimitrakopoulos (Ed.), Advances in Applied Strategic Mine Planning (pp. 119–138). Springer, Cham. https://doi.org/10.1007/978-3-319-69320-0_10

Godoy, M. (2002) The effective management of geological risk in long-term -production scheduling of open pit mines. The University of Queensland, Brisbane

Goodfellow, R. and Dimitrakopoulos, R. (2016) Global optimization of open pit mining complexes with uncertainty. Applied Soft Computing Journal, 40(3), 292–304. https://doi.org/10.1016/j.asoc.2015.11.038

Goodfellow, R. and Dimitrakopoulos, R. (2017) Simultaneous stochastic optimization of mining complexes and mineral value chains. Mathematical Geosciences, 49(3), 341–360. https://doi.org/10.1007/s11004-017-9680-3

Goodfellow, R. and Dimitrakopoulos, R. (2015) Stochastic optimization of open-pit mining complexes with capital expenditures: application at a copper mining complex. Proceedings of the Application of Computer and Operations Research in the Mineral Industry (APCOM), 657–667

Jewbali, A., Silva, D., Inglis, R. and Allen, L. (2017) Developing an Optimization Framework for Drill Hole Planning. Application of Computers and Operations Research in the Mineral Industry, 2–17 to 2–28

Kingma, D. P. and Ba, J. L. (2015) Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–15

Kumar, A., Dimitrakopoulos, R. and Maulen, M. (2020) Adaptive self-learning mechanisms for updating short-term production decisions in an industrial mining complex. Journal of Intelligent Manufacturing, 31(7), 1795–1811. https://doi.org/10.1007/s10845-020-01562-5

Kumar, D. and Srinivasan, S. (2019) Ensemble-Based Assimilation of Nonlinearly Related Dynamic Data in Reservoir Models Exhibiting Non-Gaussian Characteristics. Mathematical Geosciences, 51(1), 75–107. https://doi.org/10.1007/s11004-018-9762-x

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. and Wierstra, D. (2016, September). Continuous control with deep reinforcement learning. 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings. http://arxiv.org/abs/1509.02971

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. and Hassabis, D. (2015) Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533. https://doi.org/10.1038/nature14236

Montiel, L. and Dimitrakopoulos, R. (2015). Optimizing mining complexes with multiple processing and transportation alternatives: An uncertainty-based approach. European Journal of Operational Research, 247(1), 166–178. https://doi.org/10.1016/j.ejor.2015.05.002

Montiel, L. and Dimitrakopoulos, R. (2018). Simultaneous stochastic optimization of production scheduling at Twin Creeks Mining Complex, Nevada. Mining Engineering, 70(12), 12–20

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., Van Den Driessche, G., Graepel, T. and Hassabis, D. (2017). Mastering the game of Go without human knowledge. Nature, 550(7676), 354–359. https://doi.org/10.1038/nature24270

Sutton, R. S. and Barto, A. G. (2018). Reinforcement Learning: An Introduction (second). The MIT Press.