Smart Mining with an Intelligent Supervisory Agent through Automated Shovel Dig Allocation

Pedro Pablo Vasquez-Coronado, Solomon Acheampong, and Eugene Ben-Awuah
Mining Optimization Laboratory (MOL)
Laurentian University, Sudbury, Canada

ABSTRACT

In recent times, the mining industry uses a range of sophisticated technology including mine planning systems, autonomous equipment systems, and decision support systems to manage operational performance. Real-time intelligent decision-making in a mining environment with huge amounts of mine planning and mine operations data is the basis for smart mining. The goal of this research is to develop an intelligent autonomous supervisor to manage a continuous mining environment in real-time to achieve the required key performance indicators. Performance objectives are designed for each process and monitored accordingly during continuous mining. The research focuses on the application of deep reinforcement learning with Deep Q Networks algorithm for short-term mine planning. This approach uses a discrete event simulation model of the mining operation and an agent-based model to simulate equipment's behavior which interacts with an autonomous intelligent agent to manage the continuous mining environment by addressing random and dynamic processes during the mining operation. The intelligent supervisor identifies trends and shortfalls by observing huge amounts of mine planning and mine operations data and makes changes to improve the key performance indicators. The intelligent agent autonomously selects mining zones and allocates shovels and trucks to minimize real-time deviations from the set ore grade and ore tonnage targets for the processing plant.

1. Introduction

In an industry where improved efficiency and productivity are crucial to profitability, small improvements in throughput, speed, and efficiency can have an outstanding impact on production. Technological development in mining has been linked historically to the creation of equipment and processes to optimize operations. Autonomous equipment increases efficiency, reduces risk and cost. If we add to this digitization that increases capacity in managing analysis and integration of large amounts of data, then we will optimize processes and increase overall productivity in operations.

The use of state-of-the-art technology allows for higher production volumes. Australia already has a fully autonomous iron ore mine [3] and other mining countries such as Canada and Chile are on a similar path [25]. Global technological development is putting more pressure on mining companies to respond to the growing demand for metals. Production processes are integrated at several levels. Strategic issues include decisions made on the distribution of its resources to implement this approach and trade-offs. Tactical issues include productivity, mine design, and support systems. Finally, operational issues concern scheduling, ore control, and truck shovel allocation.

Operational decision-making is a challenge for companies. Experimenting with these systems in real life is risky, costly, and can lead to wrong decisions. Because of the complexity of the operation, the use of decision support modelling tools is necessary and useful for understanding

ISBN: 978-1-55195-486-8

interactions and improving system performance. While intelligent agent-based modelling is currently a trend, simulation is very useful to represent reality. One well-known system modelling technique is simulation, which mimics the operation of the real world on a computer [18]. By mimicking the operation of a real system, the simulation generates an artificial data of the system, and its observation allows inferences to be made concerning the system's operation/performance characteristics, including unexpected system outages, which could affect the entire system.

Machine learning is a very common Artificial Intelligence (AI) technique to automate repetitive systems tasks [21]. One of the companies using AI is Caterpillar. This helps to improve the identification of obstacles and increases the productivity of haulage trucks. The technique is based on a self-training process where thousands of identified images are required and used to adjust artificial intelligence to identify an object. This is achieved by a learning algorithm that identifies objects based on their experience. As larger amounts of data are processed and compared, performance is improved because objects can be identified more easily [35].

The objective of this paper is to present a simulation framework for an intelligent supervisory agent that supervises the operations of a short-term open pit mine. The intelligent supervision system identifies trends and shortfalls in operations based on a large amount of mine planning operations data.

To illustrate the strength of the model, this paper presents a case study and validates the model on a base case scenario.

2. Literature Review

Simulation is a representation of what happens in the real world in a virtual environment. The main purpose of simulation is to reduce risks that could cause a great loss in the real-world environment [20].

In the last few decades, the amount of data analyzed and the ability to perform calculations in a short time has increased. This has improved the accuracy and usability of simulation, which better matches reality. Companies simulate their systems with commercially available software. The software currently has a variety of tools or components that allow replicating and evaluating real-world systems. Electronic systems, process optimization, robotics, among others are examples [12].

Vasquez Coronado [31]used the discrete event simulation technique to develop and design a tool that notifies cycle times of equipment in an open pit operation that is below the required performance standards. On the other hand, an important challenge in supply chain simulation is modelling. Bodon [6] reported a method for modelling a supply chain from mine to shipping port, using simulation and optimization techniques. The approach was successfully applied in the Pt Kaltim Prima Coal in Indonesia. Askari-Nasab [2] designed an intelligent 3D simulator based on reinforcement learning to maximize the Net Present Value (NPV). The author demonstrated that his methodology optimized the discounted net pit value in comparison with Milawa algorithm. Shishvan and Benndorf [26] presented a multi-stage simulation-based optimization for continuous mining. This methodology integrates simulation and algorithms to find a solution to a workshop sequencing and transport problem. The work was carried out in a coal mine in Germany, resulting in improved mine dispatching decisions. Bissiri [5] used an agent-based simulation technique for shovel allocations. Agent-based modelling suggests a completely different view that the modeler should take when mapping a real-world system.

It should be emphasized that a simulation model's goal is to simulate the behaviour of the modelled system in order to determine what operating conditions and levels result in the best performance.In terms of decision theory, Shishvan and Benndorf [26] mentioned the parameters equivalent to the decision variables and the performance of the response variables of interest, also called Key

Performance Indicators (KPI),. In the case of supply chains, these variables are considered multiple attribute behaviors in decision making. To improve the results, intelligent agents are employed to interact with the simulation. Ozdemir and Kumral [24] proposed an agent-based Petri net simulation model to verify compliance with production targets, fuel consumption, and processing head grade control. The results showed that this approach could help in the design of processing plants, installation capacity, and monitoring of fuel consumption in mining operations.

Currently, there are expert systems capable of imitating human mental capacity, demonstrating independence when executing assigned tasks and responsibilities. Taking experience-based learning as a reference, each application has a degree of innovation that is framed in the field of computational systems and is linked to artificial intelligence. Intelligent agents perceive an environment through sensors and act on it using effectors, and they also can communicate with other agents. Sometimes, agents are developed as entities that constitute a system. This system is called Multiagent System, where agents interact and communicate with each other, to provide a solution to a problem. Reinforcement learning (RL) is what allows the algorithm to learn from its mistakes. In the beginning, the errors are numerous, but by providing a series of positive and negative signals associated with the successes and errors respectively, the algorithm will eventually learn on its own, until it becomes the most efficient of all. In this paper, a Discrete-Event Simulation modeling is used. With Anylogic Process Modeling Library, a real mining environment operation is modeled.

Jovanoski et al. [19] and Grigorye [15] examined modern simulation modeling using discrete event, agent-based, and system dynamics modeling. As shown in Figure 1, each method serves a specific range of abstraction levels and their respective application. System Dynamics (SD) is a very high abstraction and is often used for strategic modeling. The Discrete Event (DE) modelling method supports abstraction at the medium and lower-middle level, while Agent-Based (AB) simulation can range from detailed models in which agents represent physical objects to highly abstracted models where agents even represent companies.

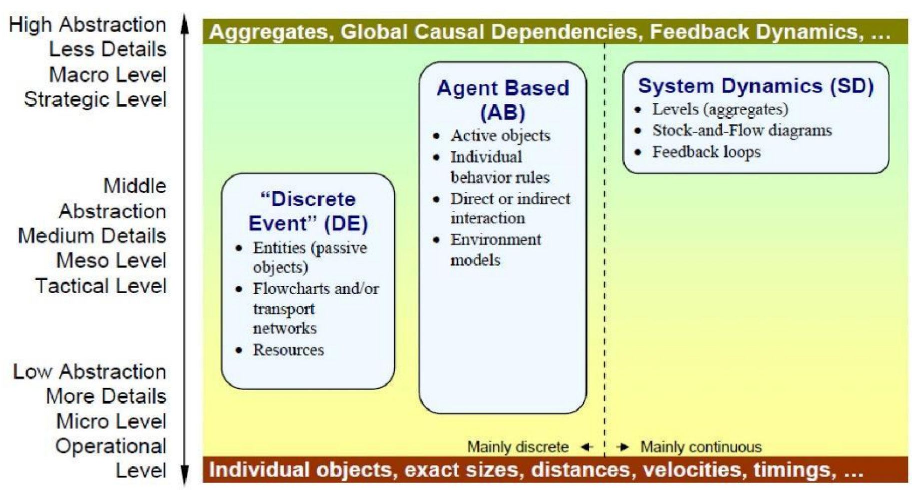

Figure 1. Method in simulation modelling [19].

Reinforcement learning is a type of machine learning technique that uses a specified agent to learn and interact with its environment by trial and error using feedback and experiences from its actions. Deep reinforcement learning on the other hand imports the use of deep neural networks which mimics the human brain functionality. Comparing reinforcement learning to other machine learning methods, like supervised and unsupervised, reinforcement learning improves on itself using rewards and punishment [28].

This paper introduces a Deep Q Networks (DQN) function-based RL algorithm, which is an improved version of q- learning algorithm with the use of deep learning and artificial neural networks.

The q-learning algorithm is a model free reinforcement learning algorithm. Model-free means independent from the transition probability distribution function and the reward function. Instead of learning the model, the agent is able to create or discover the best policy by monitoring the actions in several states and selecting the one that has the highest q-value in the present state.

2.1. Transition probability distribution function

This is the probability that causes a change in states (S). The process consists of a state space, transitional probability matrix describing all possible transitions and an initial state (S_1) from the state space. The transitional matrix is shown in Figure 2 below

$$P = \begin{bmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,j} & \dots & P_{1,S} \\ P_{2,1} & P_{2,2} & \dots & P_{2,j} & \dots & P_{2,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{i,1} & P_{i,2} & \dots & P_{i,j} & \dots & P_{i,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{S,1} & P_{S,2} & \dots & P_{S,j} & \dots & P_{S,S} \end{bmatrix}.$$

Figure 2. Transitional probability matrix / The Markov Matrix [11].

The Following illustrates the equation of transition probability distribution function

$$P_{ss'}^{a} = P[S_t + 1 = S' | S_t = S_1, a_t = a] \tag{1}$$

$P_{ss'}^{a}$ = Probability of action a, issued in state s, ending up in state s^1 with reward r

2.2. Reward function theory

Reward (R_t) is the feedback the environment provides after an agent takes an action and transitions from one state S_1 to another state S_2. The goal of the agent is to maximize the total rewards.

$$R_t = r_t + 1 + \gamma r_t + 2 + \gamma^2 r_t + 3 \ldots \gamma^n r_t + n \tag{2}$$

γ is the discount factor between 0 and 1. Discount factors are the additive that allows the agent to recognize how important future rewards are and always account for them. Setting the discount factor between 0 and 1 makes the agent look for immediate and future rewards accountability.

The following illustrates the equation of transition reward function

$$R_{ss^1}^{a} = \mathrm{E}[r_t + 1 | S_t = S, a_t = a, S_t + 1 = S^1] \tag{3}$$

2.3. Bellman equation

Bellman equation [4] is used to calculate the expected value of a given policy π. From Equation 3 we can see that a certain discount factor is applied for the agent to account for immediate and future rewards. According to Bellman equation [4] a long term reward R_{t+n} in a given action is equal to the sum of the reward of the current action R_t and the expected reward from the future actions R_{t+1}.

$$V(s) = \max_{\pi} (R(s,a) + \gamma V(s')) \quad (4)$$

From the Equation (4), V(s) is the value of the state. i.e., the numeric representation of a state which guides the agent to find its path.

2.3.1. Policy function

A Policy π is a function in charge of causing the agent to take an action (a) in a state (s). The sum of all the possible actions equals 1 (i.e., $\Sigma_a \pi(s,a) = 1$

2.3.2. State value function (V^{π})

The expected total rewards to be received starting from a state is presented in Figure 3.

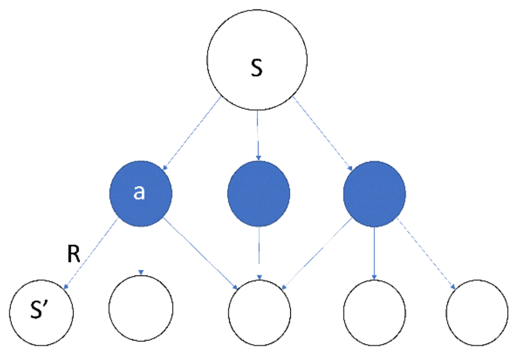

Figure 3. Diagram illustrating state value function [11].

Mathematical model behind state value function is presented in Eq.

$$V^{\pi}(s) = \sum_{\alpha} \pi(s,a) \sum_{s'} P^{\alpha}_{ss'} R^{a}_{ss'} + \lambda \sum_{\alpha} \pi(s,a) \sum_{s'} P^{a}_{ss'} V^{\pi}(s') \quad (5)$$

2.3.3. Action value function (Q^{π})

The action value is the value of an action taken at a state returning an expected total reward. It is shown in Figure 4.

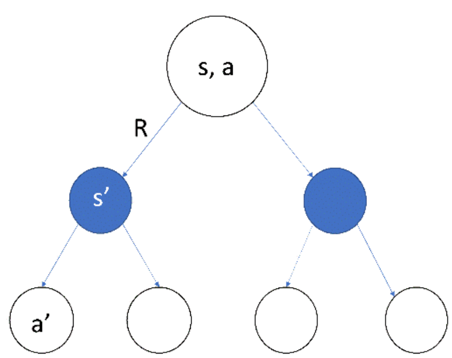

Figure 4. Diagram illustrating action value function [11].

Mathematical model behind action value function is illustrated in Equation 6.

$$Q^{\pi}(s,a)=\sum_{s'}P^{a}_{ss'}R^{a}_{ss'}+\gamma\sum_{s'}P^{a}_{ss'}V^{\pi}(s') \tag{6}$$

2.3.4. Optimal policy

The policy which maximizes the total cumulative reward is the optimal policy. Solving a Markov decision process means finding an optimal value function. From equation 4, $\max_a$ is the maximum reward for state(s) we can get from the system.

$$V\pi(s)=\max_{\pi} V_{\pi}(s) \tag{7}$$

2.3.5. Greedy policy

This function allows the agent to always prioritize the most optimal steps, hence the name greedy. The relationship between state value and action value can be valuated from Equation 8.

$$V(s_t)=\max_{\pi} Q(s_t,a) \tag{8}$$

From the above policies and functions, we can then define the bellman equation as

$$Q(s_t a_t)=r(s_t,a_t)+\gamma\frac{\max}{a_{t+1}}Q(s_{t+1},a_{t+1}) \tag{9}$$

2.4. Markov decision process

Markov Decision Process is a mathematical framework for modeling a decision control situation on a discrete-time stochastic environment [4]. It consists of states and actions with the actor as the agent. For each step, the process exists in any random state, the agent makes an action randomly from the available actions from a given state (s). A feedback or reward R_a (s, s') is provided by the environment as a form of motivation to the agent. The process is looped or continued for a number of epochs/episodes.

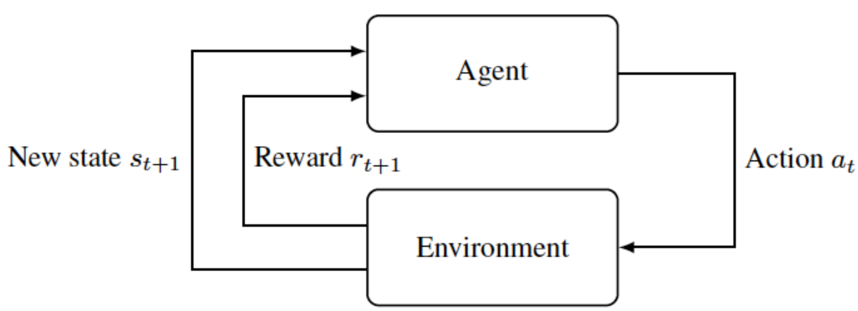

Figure 5. Markov decision process workflow of agent using DQN [14].

2.5. Deep Q Network

Deep learning is a type of machine learning approach which comprises of two or more layers, weights and biases, artificial neurons known as perceptron. Weights and biases play a very vital role in deep learning.The weight determines how much influence the input vector impacts on the output[34].

On the other hand, the biases, which are typically independent and constant with a value of 1, are added to the following layer.It serves as an assurance that even when an input vector contains zero values, there will still be an activation in the neurons [34]. Equation 10 represent mathematical equation of a neuron

$$Y = \sum (weight * input) + bias \tag{10}$$

2.5.1. Stochastic gradient descent

When training a deep neural net on a dataset, the gradient descent functions as an optimization process to minimize the loss, normally called the loss function.

The loss function serves as a check as to how the model performs given the parameter of weights and biases. The gradient descent helps in finding the precise parameter to use in order to achieve a good model.[34].

2.5.2. Back-propagation

Back propagation simply means calculating the gradient of a loss function. During the training process of a neural network, each weight within the network is calculated automatically with a deferential algorithm, the calculated gradients are then used to update the weights [7].

2.5.3. Learning rate

This is the measure of the extent to which newer information acquired overrides old information. During training process of a neural network, the learning determines the step size for which the gradients are calculated[33]. Learning rate ranges from 0-1. The larger the rate, the more the network will overshoot the minimum. The lower the range, the longer the training process.

2.5.4. Deep Q learning network

DQN is a type of Reinforcement learning method usually implemented on a Markov decision process stage to train agents to take decisive and good actions for the purpose of maximizing a defined reward function. The policy trains and improves itself from experience by performing a series of actions and receiving rewards (negative or positive) from the environment [8]. The decision process workflow is shown in Figure 6.

Deep Q Network composes of mainly of these techniques: Experience Replay: This is a technique introduced into the network to make the network updates stable. At each data collection, there are a set of transitions added to the buffer called replay buffer. The whole idea of experience replay is to compute loss and its gradient using a mini-batch of transitions from the replay buffer [12].

Target Network: Target network acts as an error measure. Considering every action the agent takes in an environment, the target network ensures the agent replays the whole actions again and again as if it was starting afresh every time it takes an action [9].

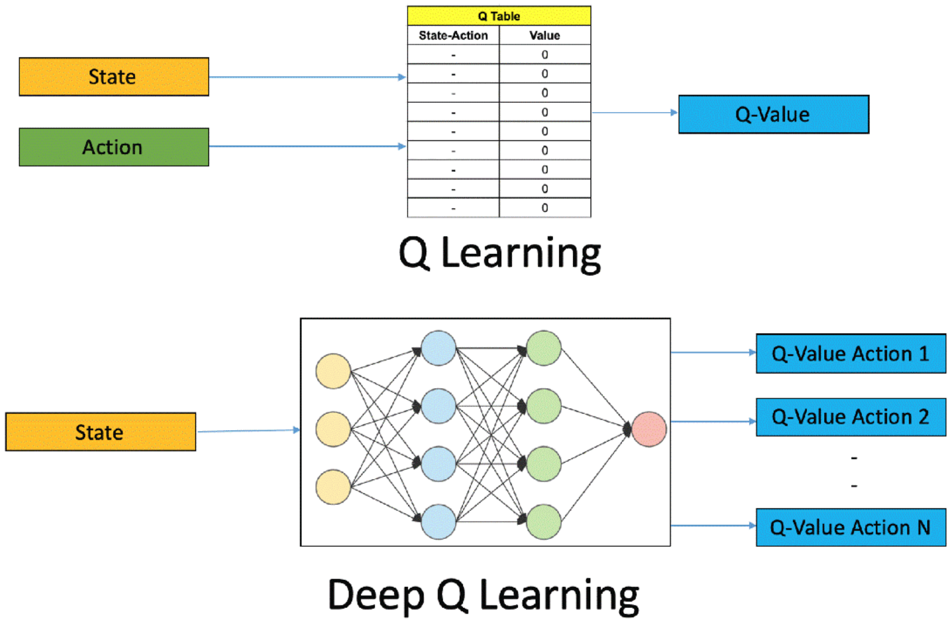

Figure 6. Graphical examples of Deep Q learning [9].

The algorithm below [10] describes the pseudocode of DQN. Q (s, a) is for storing Q values for every time an action(a) is triggered usually by the policy/agent in a state (s). The idea of the agent is to maximize the Q value function.It learns from the Q value which returns an expected total from a given state-action pair. Anytime the agent takes a move, a reward is returned thus R (s, a, s') becomes the new target the agent pursues from Q (s, a). γ max is a discount factor. Discount factors cause reward functions to lose more of their values as more as the agent keeps on acting on the environment. This causes the agent to seek more rewards [10].

Pseudo Algorithm:

Start with Q_o (s, a) for all s, a.

Get initial state (s)

For k=1, 2.... n

Sample action(a), get next state s'

If s' is terminal:

target = $R(s,a,s')$

Sample new initial states s'

Else:

target = $R(s,a,s')+\gamma \max_{a'} Q_k(s',a')$

$Q_{k+1}(s,a) \leftarrow (1-a)Q_k(s,a)+a[t\arg et]$

$s \leftarrow s'$

3. Problem Statement

Large-scale surface mines are those operations that can produce hundreds of millions of tons per year. These mines require various components to ensure a continuous flow of material to meet daily demand. Each time material is processed, it must have acceptable grade ranges. Proper blending is accomplished with a distribution of trucks, varying shovel mining rates or approximating the proportions of material in the zones to be mined.

Each mine generally has major elements or processes. Each of these processes has sub-processes with thousands of variables and scenarios, which are impossible for the human eye and skills to consider, much less control them to make decisions with the speed, accuracy and timeliness that the process requires. That is, man does not have the ability to react to the events that occur every day in the mining processes, however, human can create the tools on which to rely on such as decision support systems.

The critical aspects of this new situation may require a novel approach to supervision. Stable production is a goal, reducing variability with the implementation of autonomous technologies that are capable of handling thousands of data. Systems must be integrated, and near-term plans must be practical to predict the results to be obtained [27].

There is the development of decision support systems that allow recording and hosting data from the mining exploration process, in addition to having the ability to handle different geostatistical mathematical analysis models and obtain a solution from the model, presenting the opportunity to optimize the process helping in decision making.

The development of intelligent agents on the existing fleet management systems using new autonomous collection and visualization technologies will help to monitor individual aspects of production in real time. Production problems, such as mix optimization, can be addressed by creating self-tuning algorithms within a machine learning tool that will, in turn, feed back to the intelligent supervisor with knowledge-based decision options. Reconciliation can be performed at the end of the production process, comparing them with the expected values obtained once the metallurgical process is completed.

4. Methodology

The proposed methodology aims at creating a virtual mining operational experimentation environment using a decision model supported by discrete event simulation and agent-based simulation. It is expected that the decision-maker initially establishes operational values (decision variables) of the supply chain, then simulates such operation and obtains some performance levels, which should be considered in implementing in the real world or testing other sets of values of the decision variables until obtaining results that lead to making better decisions. The following describes the development required to obtain the virtual operating experimentation environment. [26]. The reserarch methodology used is presented in Figure 7. One type of data is needed which is the long-term production schedule. The production schedule is exported from GEOVIA Whittle [29]. The production schedule is an Excel file containing coordinates of each block to be mined, tonnage of material to be mined, mining period, pushback, grade, metal content and recovery. Mining, haulage and plant stages are simulated. The intelligent supervisor recognizes and corrects deviations based on the defined Key Performance Indicators (KPIs). The output is an operational short-term plan for all stages.

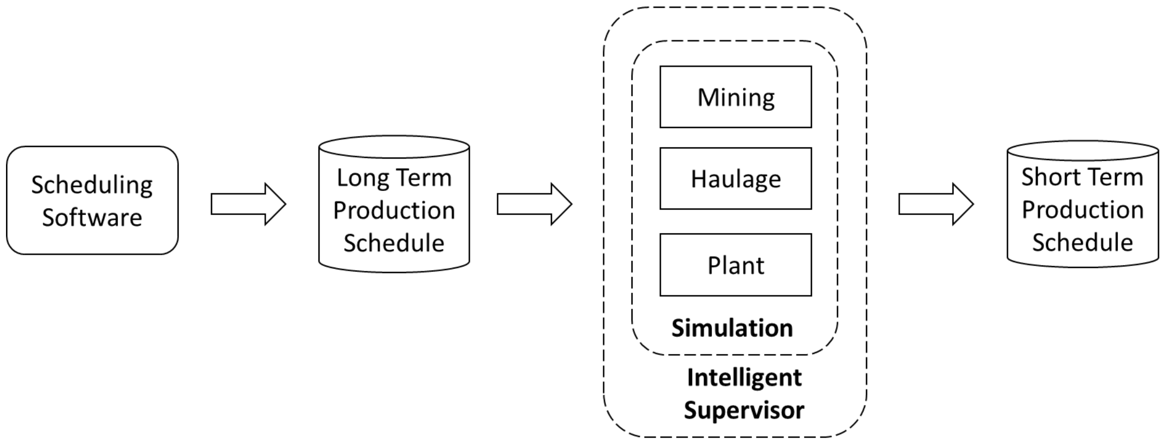

Figure 7. Research methodology.

The model simulates the mining operation involved in the production and transportation of material to the processing plant. The chain consists of an open pit, a fleet of trucks available for transport, shovels for loading, crusher, stockpile, and processing plant. Since the selected simulation package AnyLogic [1] offers the functionality to develop hybrid models, it has been selected as the model development platform.

Agent-based simulation can capture the change of objects after an event occurs, and discrete event simulation can be modeled by accounting for the flow of sequential events. In addition, system dynamics simulation can capture the overall model results without a detailed configuration of objects [32].

The Reinforced intelligent agent interacts with the simulation environment to find an optimal path in solving a discrete complex problem. In our case, subordinate agents, in the AnyLogic process modeling library are programmed to gather the necessary data needed for our reinforced agent to interact.

Our agent follows the Markov Decision Process (MDP) principle in getting best possible solutions. An MDP Agent learns from a problem and aims to achieve a certain goal. The agent interacts with the simulation environment continuously (called epochs/ episodes). The agent selects a set of actions (A) and the environment responds to these actions and presents new situation of States (S') to the agent in a form of a reward (R)/ a score [32]. Markov decision process is shown in Figure 5

4.1. Phases of research

This section shows three consecutive experiments that present a result in each phase, which becomes a fundamental element for the next one. This achieves a common thread and coherence that guarantees the rigorousness of the research and its results. Presented in Table 1 is the research experiment and phases.

Table 1. Research Phases.

	Experiment 1		Experiment 2		Experiment 3
Activity	**Phase 1**	**Phase 2**	**Phase 3**	**Phase 4**	**Phase 5**
1	Testing model with 9 zones per period	Testing model with 27 zones per period	Model and allocate trucks no uncertainties	Introduce more uncertainties for plant.	Increase mining and processing rate in order to reduce stockpile inventory.

2	One stockpile	Multiple Stockpiles. Split base on cut-off. Upgrade grade on accumulative average	Model haulage profile no uncertainties	Introduce more uncertainties in trucks.	Spatial Shovel diggability data based on kriging/machine learning
3	Two shovels working at the same time	3 shovels working at the same time	Availability of shovels, and trucks. Add maintenance, repairs, and breaks		Crusher productivity based on rock properties diggability data / drilling ratio
4	Using operational delays and uncertities		Availability of plant. Add maintenance, repairs, and breaks		

4.2. Simulation

This section covers the simulation sequence of the original design of the mining process (Figure 8). The discrete event simulator simulates the production process, which begins with the extraction of the material and concludes with the transportation of the material to all destinations (plant, stockpile, waste rock dump). Material movement is performed by a system of truck shovels.

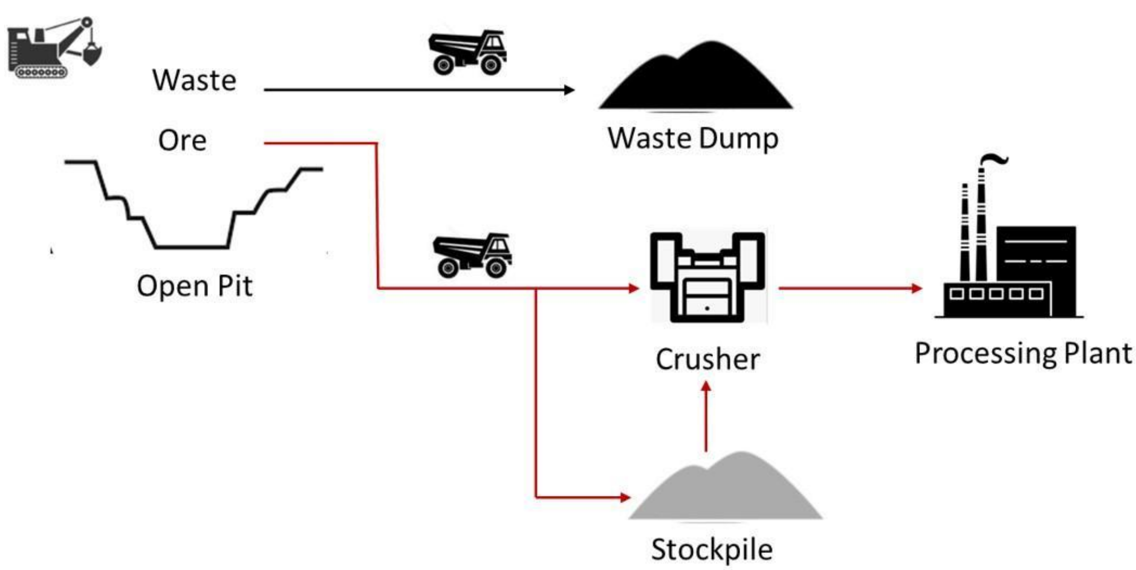

Figure 8. Mining simulation model.

The next section explains the processes used in the simulation of reading, mining, haulage, dumping and plant. The logical sequence, indicating the processes and decision points for Phase 1is shown in Figure 9 and Phase 2 in Figure 10.

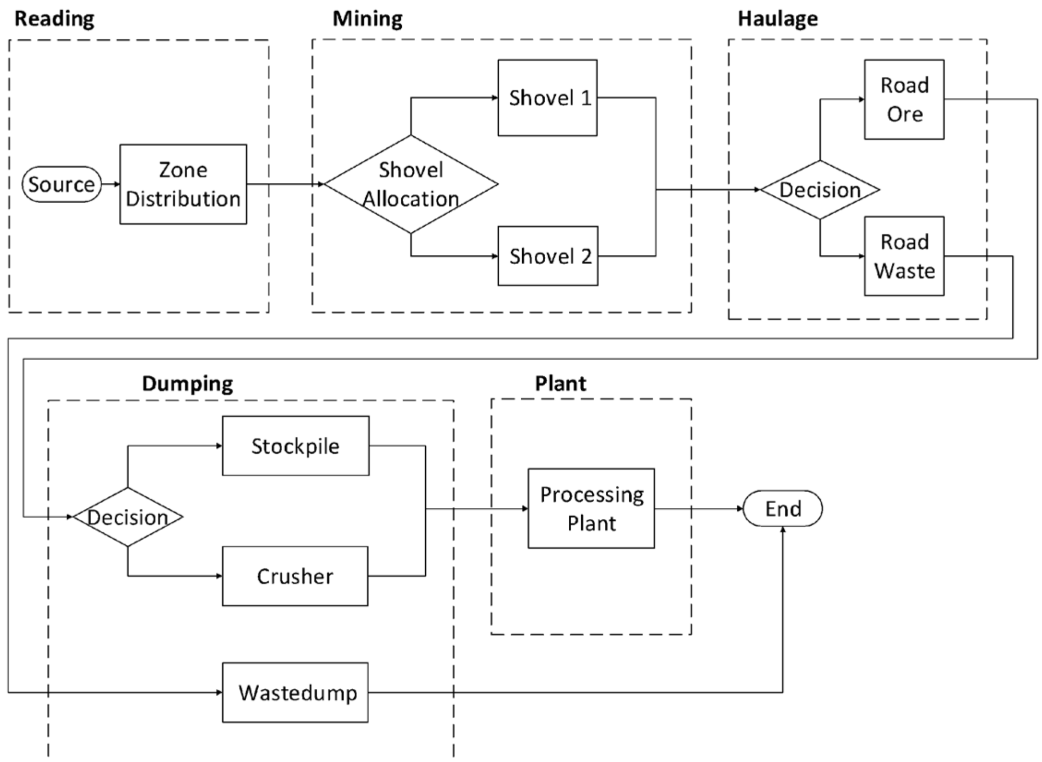

Figure 9. Logical sequence for simulation model experiment 1.

4.3. Reading

Each agent that enters the system represents a block of the production program and they read the data one at a time as agents enter the system. The production model contains the following information:

Table 2. Block model parameters.

Field	Description
BlockID	Block identification
Coordinate-X	Location in X
Coordinate-Y	Location in Y
Coordinate-Z	Location in Z
Rocktype:	0 Waste, 1 Ore
Tonnage	Amount of material
MetalContent	Contained metal
Recovery	Percentage of ore recovery
Al_2O_3 Grade	Aluminide grade
Al_2O_3 (tonnage)	Aluminide Tonnage
SiO_2 Grade	Silica grade
SiO_2 (tonnage)	Silica tonnage
Pushback	Mining phase
Period	Mining Period

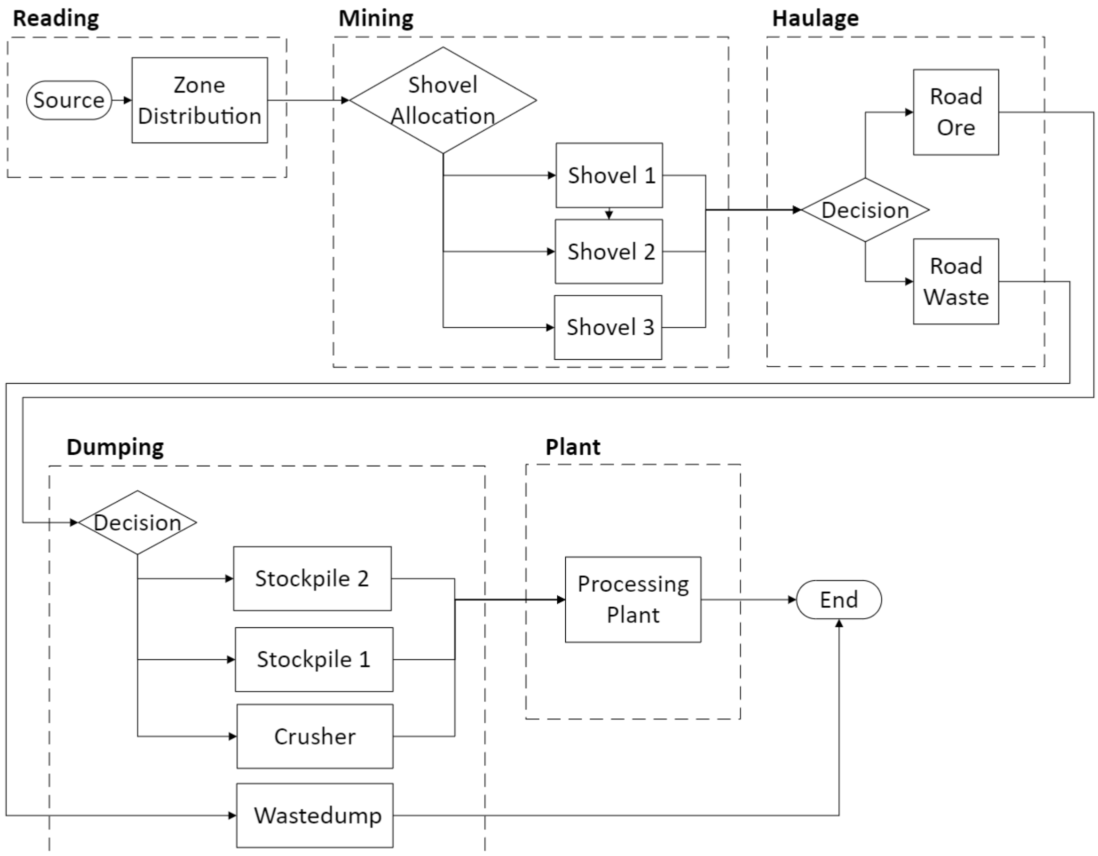

Figure 10. Logical sequence for simulation model experiment 2.

4.4. Mining

There is a logical decision to mine the blocks. The truck goes directly to the available shovel, if it is occupied it goes to another shovel. If both are occupied, the truck is moved to the shovel with the shortest queue. The shovel that will load the truck is assigned, and the maneuvering and parking time are simulated. Afterward, the loading time of the truck is simulated, and finally, the shovel is released to load the next truck.

4.5. Haulage

In this sub-model, the decision logic is found if the truckload is ore or waste. If the load is identified as ore, it is sent to the route that goes to the plant, while if it is detected as waste, it will be sent to the waste dump. The routes would simulate the travel time of the trucks to their destination.

4.6. Dumping

The truck with ore is sent directly to the plant for crushing and processing. If there is no queue in the unloading area, the material is unloaded, and dumper returns to the loading area to be assigned to a shovel. If the plant is occupied, a truck queue will form. If the queue is greater than 5 dumpers, the ore is stored in the stockpiles. The stockpiles will send the ore to the plant when required by a front-end loader and two dump trucks dedicated exclusively for this operation.

If the material in the dump truck is waste rock, it is taken to the dump, beginning the maneuvers of accommodation, backing up, and parking in the area specified for unloading. Once the material has been unloaded, the dumper returns to the loading front to start a new cycle.

4.7. Plant

The plant receives the material sent directly from the mine or stockpiles. Feeding must be continuous to avoid plant stoppages due to material shortage. The resource that will perform the processing of the block is assigned. Shutdown paralyzes the plant activities for 15 days for maintenance, but this will only happen annually. The Plant block simulates the time it takes for the plant to process the material received, once the processing is finished, the plant is available to receive the next load of ore.

4.8. Replication

By running the model, the obtained values for the variables and parameters at the completion of the simulation will be significantly different if the model is executed multiple times [13]. The number of simulations of an individual test should be justified by the goals and required simulation accuracy [16]. Each replicate simulation is a feasible solution within the probability distribution of all possible combinations. Holford et al [16] state that the number of simulation replicates should be based on the objectives and accuracy of the study.

A 95% confidence interval was used to validate the number of replicates, establishing a minimum range of 2 and a maximum of 10 replicates with which to run the simulation. Different numbers of replicates were used for each case due to the type of data and variations in the model.

4.9. Intelligent mining agent (IMA)

The proposed model presumes a total of 27 zones from which two shovels and 17 trucks are defined. It has been suggested that the two shovels work simultaneously, alternating between different areas according to objective needs. Available trucks are to schedule and are allocated to shovels as soon as they are in need. The model illustrates 27 mining zones from which the agent is required to choose from which to mine first. The first 5 zones are in the superior level and zone 6-27 are in the inferior level. Zones 1-5 are viewed in Figure 11.

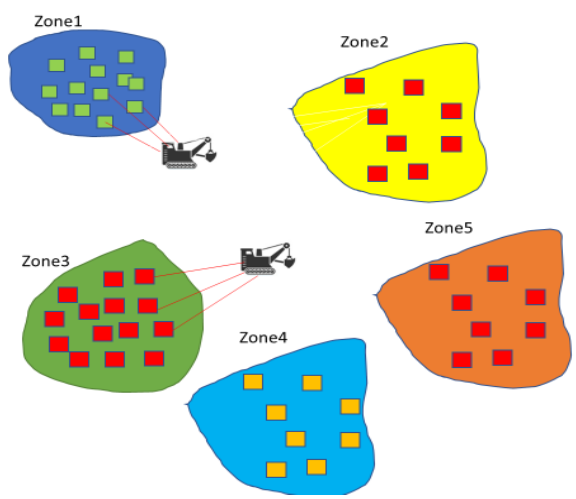

Figure 11. Graphical representation of the zone's distribution.

4.9.1. Training implementation architecture

The grades and tonnages (rectangular boxes) contained in each of the zones are assumed to be known. However, these metrics are not released to the intelligent agent. The intelligent agent must uncover these values by training and learning from the environment. It first extracts the top level zones to continue with the lower zones.

- Observation state

The state of observation is the experience sample given to the agent by the environment. Here the observation state chosen were the zones, shovels, and trucks availability. A total observation space of 9 is fed into the DQ network.

- Sample space action

The sample action is the function in which the agent chooses to act on the environment. As a sample action is performed, a new state (S') is produced and a reward is given, causing the agent to consider the next action.

In this research, a sample action space of 4 is implemented for the agent to choose from. The action spaces range from 0-3 where each action does the following:

0: Take no Action

1: Choose zones with good minimum deviation

2: Allocate available shovels for selected zones

3: Allocate available trucks for the active shovels

hovel movements between zones is an objective; after all, the agent needs to make sure that a shovel is relocated depending on the location of these blocks. A shovel may be assigned a different zone if the agent feels that lower plant production is being achieved.

It is an objective of the intelligent agent to make sure that the two shovels are always working together at the same time and trucks are distributed among shovels working at a higher pressure (that is enough material content to mine).

4.9.2. Selection of zones

A two-set criteria has been given to the agent to aid in the zone selection. i.e., ore grade and ore tonnage. A target has been set for the 2 Constraints. The agent makes a random restricted action, restricted means; the agent is only allowed to mine from the top-level zones first before it proceeds to mine the lower level. Considering at any instance the agent makes the wrong random action, which is by choosing to mine the lower level instead of starting from the upper level, penalty feedback is rewarded to the agent. When the agent selects from the zones, each block in the zone it selects contains available ore grade that gets subtracted from the target grade and tonnage per week. This is called a minimum deviation reward. The accumulation of these rewards is cumulated for blocks in each zone to extract the highest possible rewards available for individual zones. Since each block contains different grades, it generates a variation between the rewards that are either high or low.

- Reward function

The DQN runs an algorithm called an epsilon greedy which makes the agent hungry for more rewards. The policy seeks to maximize their rewards, therefore anything that will reduce their total reward is omitted. Once this happens, the agent will only extract selected zones (i.e., zones with higher rewards leaving the rest unmined). Another reward function is implemented to check that instance as soon as the agent starts performing this activity. This function deducts rewards from its total gained, allowing the agent to seek out available areas to mine. A certain weekly target is initialized for the agent with each block containing the grade and ore tonnage. A requirement for the agent is not to exceed that scale.

A weekly target parameter set for the three zones is as follows: Metal content processed 19,642,990 %mass, Al_2O_3 grade is 49%, and total tonnage (ore and waste) equals 1,354,166.6 tonnes/week.

Before the training process, these targets are reset. Anytime the agent runs an epoch, for each block, the agent mines in a week. The grade and tonnage are compared and subtracted from their respective weekly targets. The results are scaled and normalized for each added. The total normalized result constitutes a reward for one epoch in a week.

- Training stage

A custom deep reinforcement learning algorithm is implemented to train the agent in the simulation. Training means running an agent through a simulation environment for a number of times. During this process, the agent learns all possible actions to take because it begins to learn from experience.

An input vector of 6 observations is fed into the network (Figure 12). They are then passed through a hidden layer of 300 nodes and four output outcomes depicting our four actions. The hidden dense layer contains a series of DQN algorithms including back propagation and a learning rate.

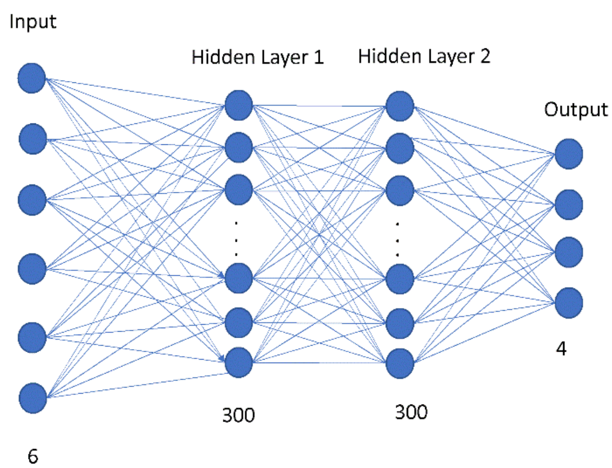

Figure 12. Graphical representation of the training phase.

5. Conceptual Layout

A bauxite deposit which has 2 types of rocks that will be mined by an open pit is been considered. The mining will be done by truck and shovel system. The material extracted from the mine will be sent to the waste dump if it is waste rock or it will be sent to plant or stockpile for processing if it is an ore.

GEOVIA GEMS [30] generated the topography, the block model and the pit design. The pit optimization and its production sequencing were performed in GEOVIA Whittle [29] Milawa algorithm to determine the NPV.

The optimal pit design structure and mine layout were designed using GEOVIA GEMS [30] [17], and the optimization was performed in Minkah [23]. The pit shell design has a total tonnage of 3,000 Mt of which 1,566 Mt is ore with grades of 51% Al_2O_3 and 5% SiO_2. The available material generated by GEOVIA Whittle [29], the designed pit boundary, and pushbacks are presented in Table 3.

Table 3. Summary of material tonnages [17].

Description	Total tonnage (Mt)	Ore tonnage (Mt)
Whittle optimum pit shell	2763.0	1610.0
Designed pit shell	3003.0	1566.0
Pushback 1	822.0	402.0

Pushback 2	1260.0	587.0
Pushback 3	912.0	544.0

In order to have a higher cash flow and to obtain higher grade zones, the pit was divided into three phases, allowing more individual control of the construction phases. The pushback design was similar in the three zones, each with its own access ramp to simplify its sequential exploitation. The pushbacks are shown in Figure 13. Phased mining allows each pushback to be fully mined before moving on to the next. After the first phase is exhausted, the area could be used as a waste dump while mining continues with the next phase.

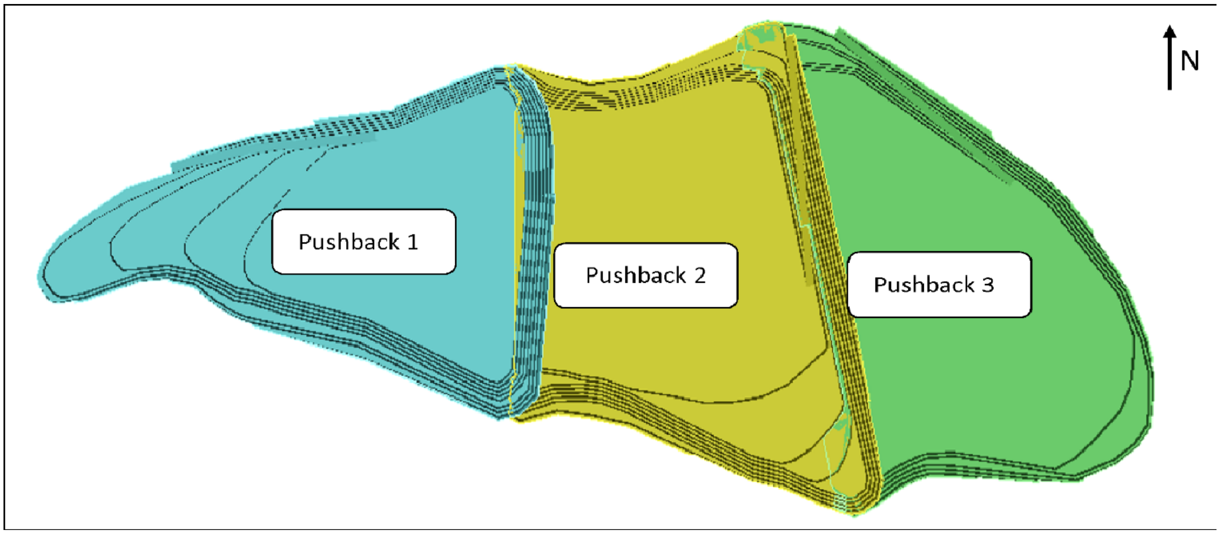

Figure 13. Pushback designs [17].

Economic parameters considered for GEOVIA Whittle [29] to generate the final pit and pushbacks are as follows in Table 4.

Table 4. Economic parameters [23].

Description	Value (units)
Reference mining cost	\$3.16 /tonne
Reference processing cost	\$9.6 /tonne
Reference stockpile cost	\$0.5 /tonne
Selling price	\$0.76 /%mass
Discount rate	10 %

The mine design has been designed for truck mining. The truck haulage system in the area of study has two loading points and a crushing site. The layout includes switchbacks, road, reclaimer, process plant, tailings and waste dump and tailings. These settings will be used throughout the basic simulation designs. This conceptual layout is shown in Figure 14.

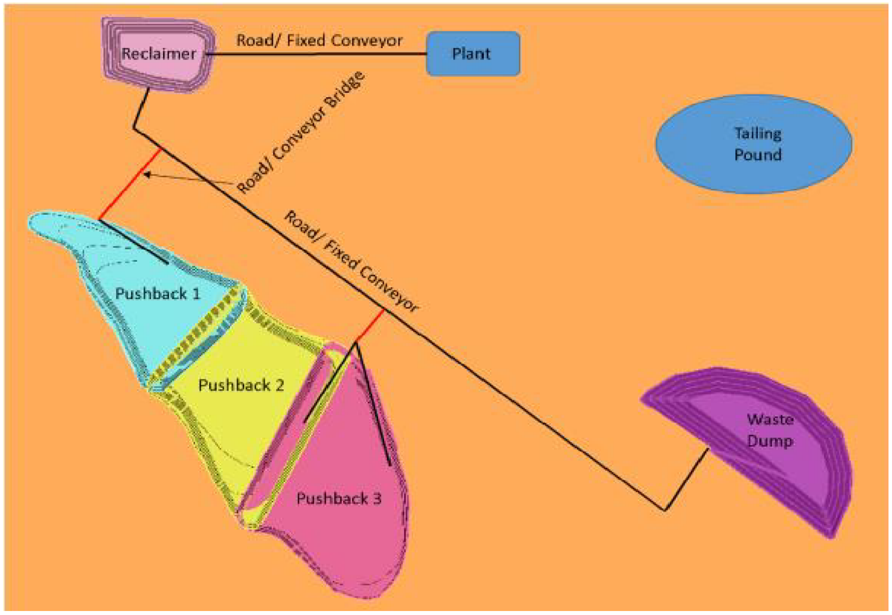

Figure 14. The conceptual design of the operating system [17].

A strategic plan was generated using GEOVIA Whittle [29], displaying the mining, processing and grade capacities. The mining capacity and processing capacity schedule are presented in Figure 15 and Figure 16 respectively.

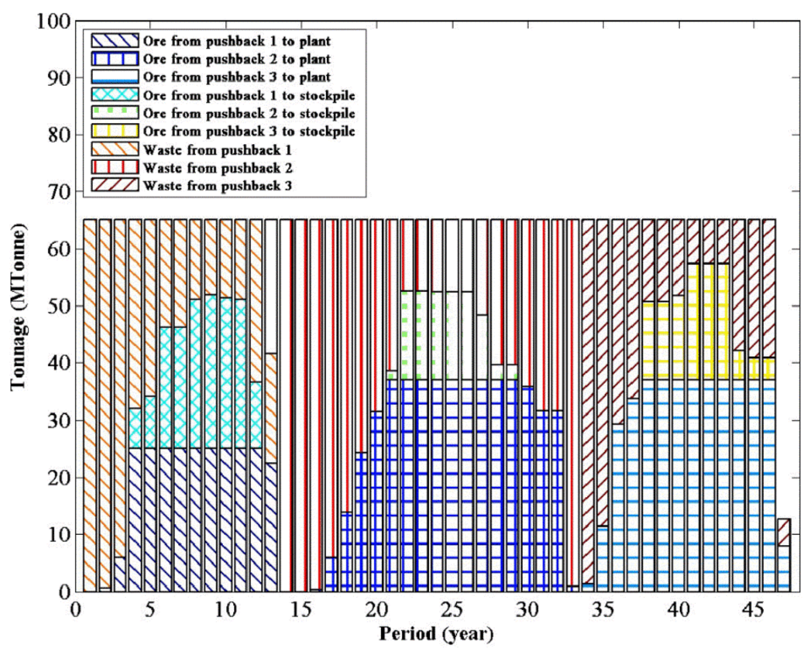

Figure 15. Mining activity in each pushback [17].

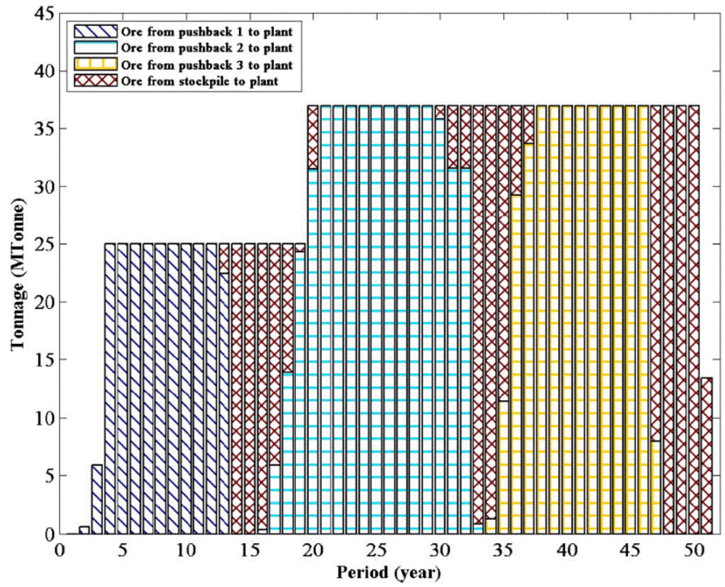

Figure 16. Processing plant material tonnage schedule [17].

The following are the parameters used by GEOVIA Whittle [29] in the optimization of mine.

- Life of mine is 47 years. Mining rate is 65 Mtpa.
- Processing Rate from years 1 to 19 is 25 Mtpa, and 37 Mtpa from year 20 onwards.
- Processing finished in year 51.
- Stockpile is necessary to meet processing requirements.

6. Experimental Designs

Design of experiments is a method that aims to perform a series of tests to induce deliberate changes in order to find out if the model works as intended. The test model is similar to a product prototype that must be tested in different ways [20]. The importance of knowing the results of each test allows us to know the effects on the model [22]. Once the model is completed, verification and validation are important. Verification is the process of ensuring that the model behaves as intended and validation is the process of testing that the model behaves as the real system does [20].

Model verification was performed with a single entity that followed the block flow of the system and behaved as expected. Validation was done by changing the proportions of mining and processing in the model. It was done with a push production scheduling database. The variables compared were mine* life. The ratios mentioned above were decreased and increased. The main effects and possible interactions between the input factors and the results were measured, concluding that the results were as expected. Following these experiments, the model is ready to allow the addition of new configurations in the future.

7. Cases to Evaluate

7.1. Experiment 1

The methodology was evaluated in 6 cases which were described above. For all cases, only pushback 1 data was considered. The simulation was set to hours and the parameters used for the scenarios of application are the following:

7.1.1. Case 1 – Life-of-mine production

This scenario is known as the base case. The goal is to reproduce the mining and processing activities in Anylogic software [1]. All scenarios assume the mine works 365 days a year and 24 hours a day. That case replicates the realistic system being the strategic production schedule generated by the GEOVIA Whittle [29].

The following configuration was setup in the simulation model:

- Recreate mining and processing activities.
- The mining ratio is 7,420 tonnes/h.
- The ratio for processing plant was 2,850 tonnes/h.
- Based on the production rates the proposed mining fleet are:
  - 17 trucks (200 t per truck)
  - 2 shovels (34 m3 per shovel)

7.1.2. Case 2 – Short term production.

The case was established on a short-term (annual) production schedule with weekly tonnage and grade results from the long-term production schedule. This case corresponds to the following characteristics:

- Focused on Pushback 1 Period 8.
- Same mining rates for previous case.
- 1 stockpile

7.1.3. Case 3 – Short term production with operational delays.

In this case, the operating delays are added as deterministic parameters. This case corresponds to the following characteristics:

- Same mining rates for previous case.
- Delay in spot time for trucks in shovel, plant, stockpile, and waste dump.
- Delay in road to transport material for the plant and waste dump.
- 1 stockpile

7.1.4. Case 4 – Short term production with operational delays and uncertainties.

Operational uncertainties were added using statistical distributions as stochastic parameters.

- Same mining rates for previous case.
- Spotting time was set normal distribution 5%
- Shovel production rate was set triangular distribution (3339 t/h, 3710 t/h, 4081 t/h)
- The processing plant feed rate was set at normal distribution (2853 t/h, 285 t/h)
- Haulage time was set at triangular distribution +/- 10%
- 1 stockpile

7.1.5. Case 5 – Short term production with operational delays and intelligent agent

The scenario is built on Case 3. The smart agent is now integrated with the mining simulation model.

- Same mining rates and delays for previous case

- The intelligent agent is trained on the mining simulation model based on a number of iterations (Epochs).
- During simulation, the intelligent agent monitors real time variations in ore tonnes, ore grade compared to target KPIs.
- A reward system is set for the intelligent agent if it achieves a required target.
- Intelligent agents will make necessary changes in shovel location, truck speed, and increase processing plant ratios.
  - Weakly Goals:
    - o Production >= 480,000 tonnes
    - o Grade 49% Al_2O_3
- The smart agent will replicate the decision making of a mine supervisor. It will perform the minimum tonnage required in the production schedule, while meeting the head grade and finally meeting the quantity of metal processed by the processing plant.
- Actions maximizing deviation reward include those performed by the smart agent through zone selection, shovel assignment.
- 1 stockpile

7.1.6. Case 6 – Short term production with operations delays, uncertainties and intelligent agent.

This case is based on Case 4. The smart agent is integrated into the simulation with delays and operational uncertainty.

- Same mining rates and delays for previous case
- Smart agents apply their training in this case to monitor the virtual mine simulation
- 1 stockpile

7.2. Experiment 2

Only pushback 1 data was considered. The simulation was set to hours and the parameters used for the scenarios of application are the following:

7.2.1. Case 7 – Short term production with operational delays, stockpiling and an intelligent agent.

The scenario is built on Case 3. The smart agent is now integrated with the mining simulation model.

- The mining ratio is 7,420 tonnes/h.
- The ratio for processing plant was 2,850 tonnes/h.
- Based on the production rates the proposed mining fleet are:
  - o 17 trucks (200 t per truck)
  - o 3 shovels (34 m3 per shovel)
- Delay in spot time for trucks in shovel, plant, stockpile, and waste dump.
- Delay in road to transport material for the plant and waste dump.
- Testing model with 27 zones per period
- Two stockpiles separated on the basis of cut-off grade.

- Update grade on accumulative average
- 3 shovels working at the same time
- Intelligent agents will make necessary changes in shovel location, truck speed, and increase processing plant ratios.
  - Weakly Goals:
    - Production >= 480,000 tonnes
    - Grade 49% Al_2O_3

8. Results

The section presents the results of the model by replicating GEOVIA Whittle's production schedule and model results using simulation, delays, and smart agents in short term. The results are generated in a database and graphs are developed and excel spreadsheets are used. From the replications, the following results are obtained as explained in module 4.2. Case 2 was submitted to 4 replications to reach a 95% confidence interval, Case 3 and Case 5 were respectively submitted to 6 and 7 replications, and finally Cases 4 and 6 were subjected to 8 replications.

8.1. Case 1 – Full production

The results of the full production case are shown below. The operating schedule and the plant schedule are presented in Figure 17 and Figure 18. From the results, the following information is provided.

The mining duration is 13 years, and the processing of the ore is 20 years. The processing plant continues to run the ore from the stockpile up to period 20. The mining is 821 Mt (ore and waste) and the plant process is 429 Mt. Based on the results, the model is consistent in replicating the production schedule generated by GEOVIA Whittle [29], which was the objective of this case. Mining production schedule in Figure 17 graphically matched with Figure 15(Whittle's results). The plant feed production in Figure 18 graphically matched with Figure 16 (Whittle's results). Instead, the simulation detects the tonnage of the block and processes it to completion, then starts processing another block. Only after all the tonnage of the block has been processed does the system recognize the tonnage of each block that arrives at its destination. The system ignores a partial block that is in process at the end of the period. Outputs are as expected, and the model is prepared to include delays, uncertainties and smart agents in the short term.

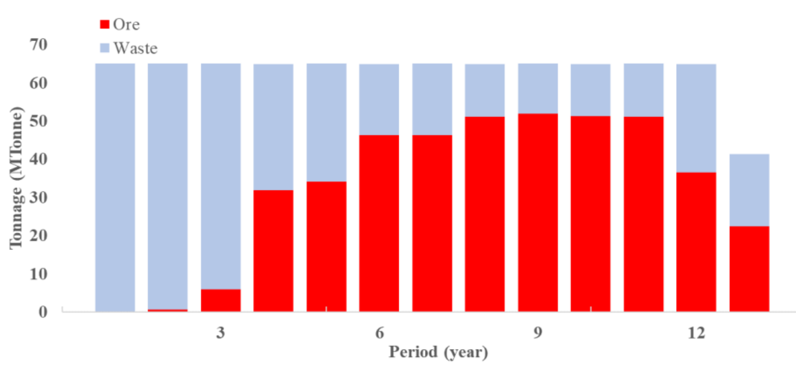

Figure 17. Annual mining production schedule.

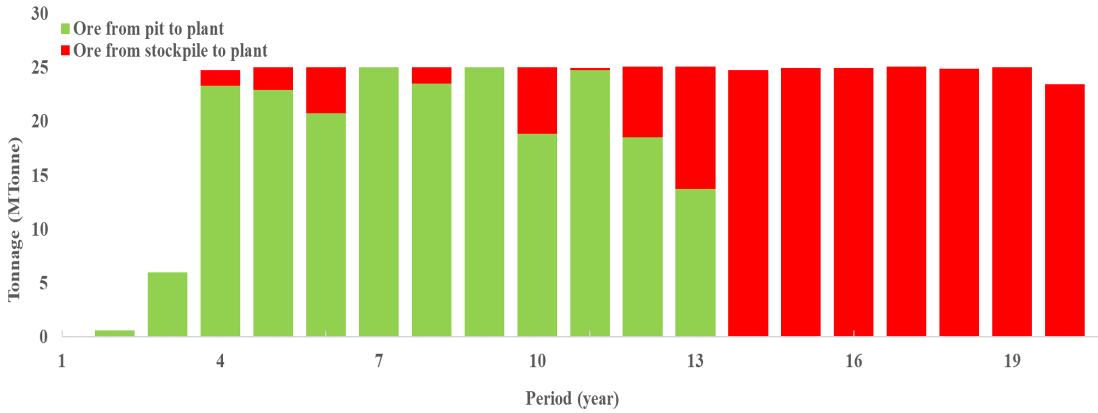

Figure 18. Annual plant feed schedule.

8.2. Case 2 – Short term production.

The evaluation for this scenario was done on a short-term basis for Pushback 1 and Period 8. According to the mining plan, 65,00,066 tons needed to be mined. Figure 19 shows the mining production schedule. The simulation finished in week 53. The mining production is steady throughout all periods; however, the ore drops in week 9 and 10. Figure 20 presents the mining distribution schedule by zones.

Figure 21 shows a section view of the Mining zones. The numbers in each block denote the zone number. Take note that the mining sequence starts in order (Figure 19), that is because the simulation continues the mining sequence given by GEOVIA Whittle [29], mining the ore and waste blocks progressively without prioritizing.

Figure 22 illustrates the feed production schedule for the plant. Plant total period in the plant is 107 with 51,043,472 tons and 2.03 billion mass % in metal content. Figure 23 represents the stock inventory. Peak storage capacity is achieved in week 52 with 26 Mt. Beyond this period the inventory decreases because it is feeding the plant.

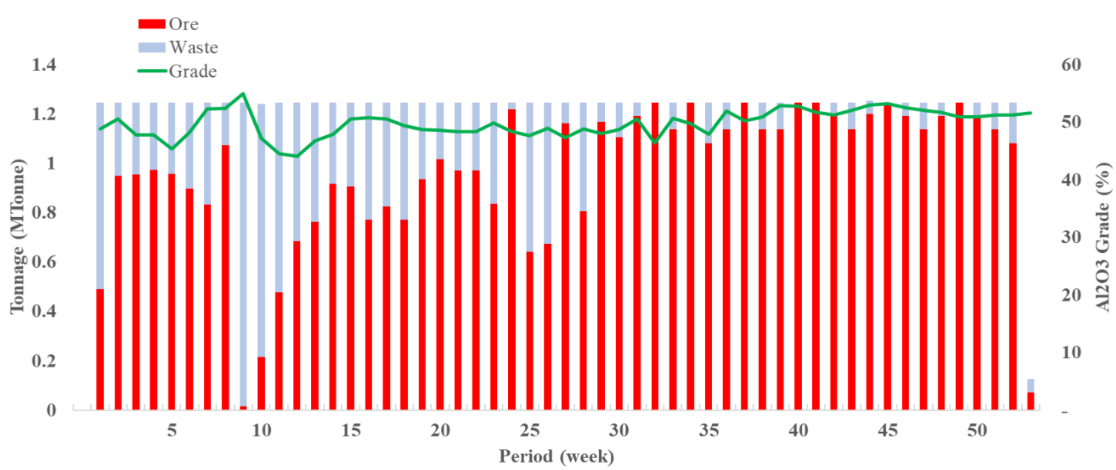

Figure 19. Weekly mining production schedule showing material types.

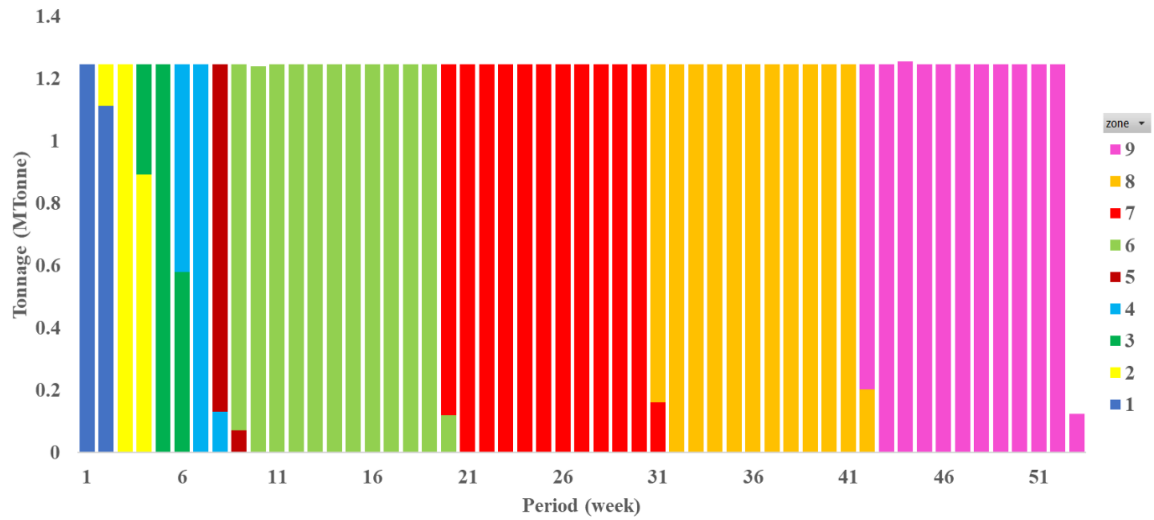

Figure 20. Weekly mining distribution schedule showing zones.

Elevation	Zones				
1482	1	2	3	4	5
1457	6	7	8	9	

Figure 21. Mining zones for shovel allocation.

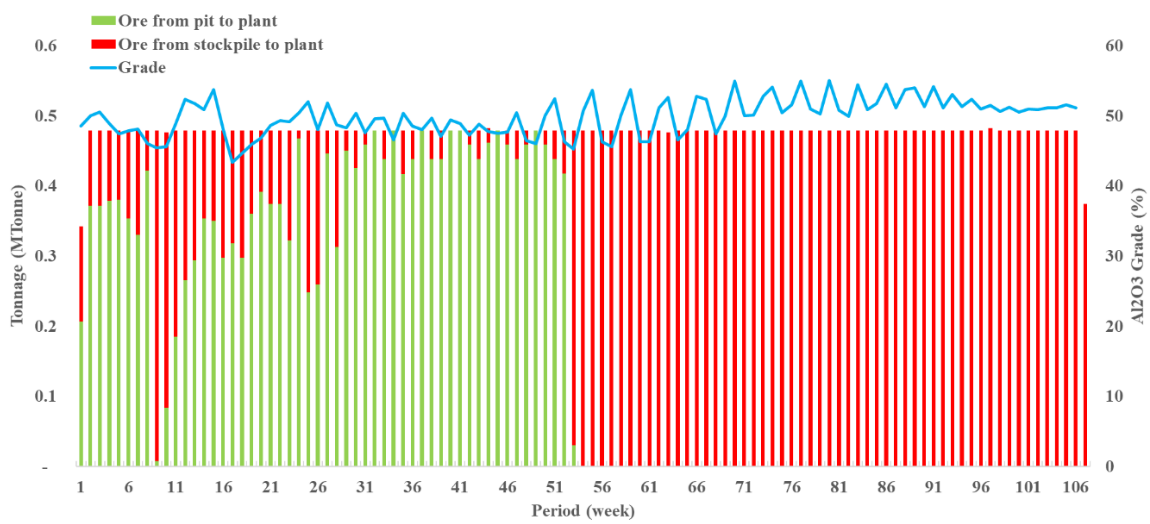

Figure 22. Weekly plant feed schedule.

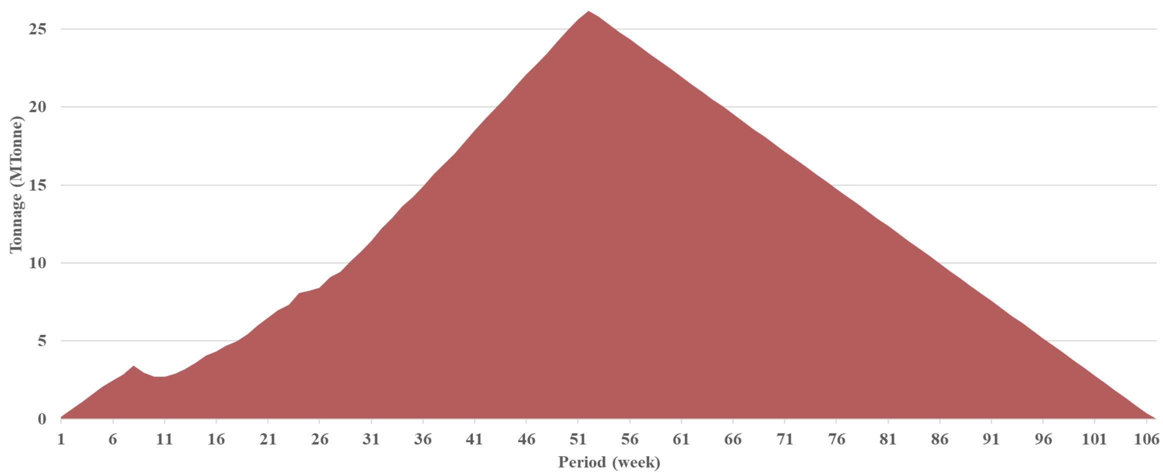

Figure 23. Stock inventory.

8.3. Case 3 – Full production with operations delays.

There is operational delay in the simulation in this scenario. The simulation is reading the blocks in sequence following the GEOVIA Whittle schedule [29], without selecting the blocks with better grades. The mine production displayed in Figure 24 is steady for all periods and delays are visible. For periods 10 and 11 the production declines, this happens because the mining sequencing follows the strategic plan generated by GEOVIA Whittle [29]. The distribution schedule by zones is presented in Figure 25. The plant's feed production schedule is presented in Figure 26 and Figure 27 shows the stock inventory.

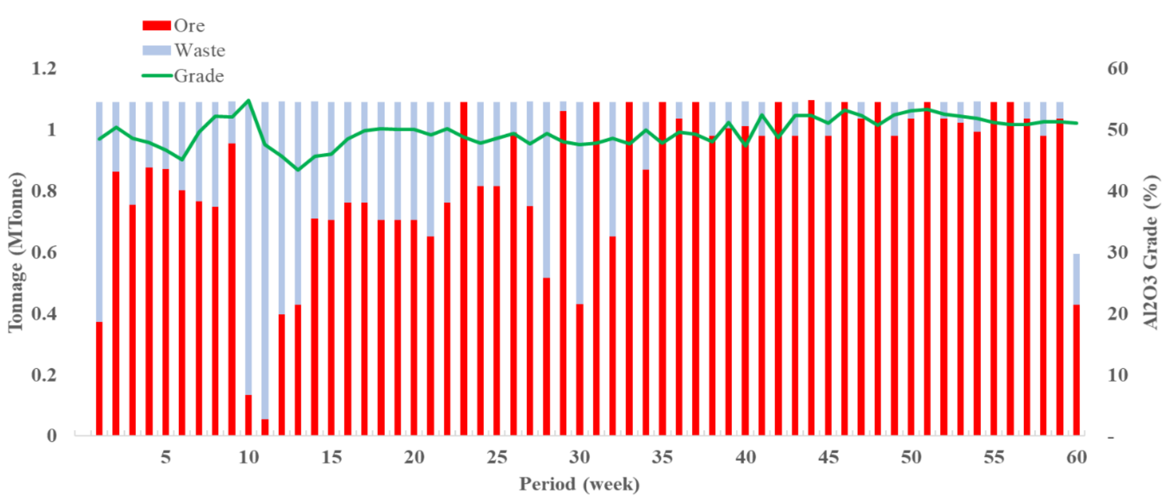

Figure 24. Weekly mining production schedule showing material types.

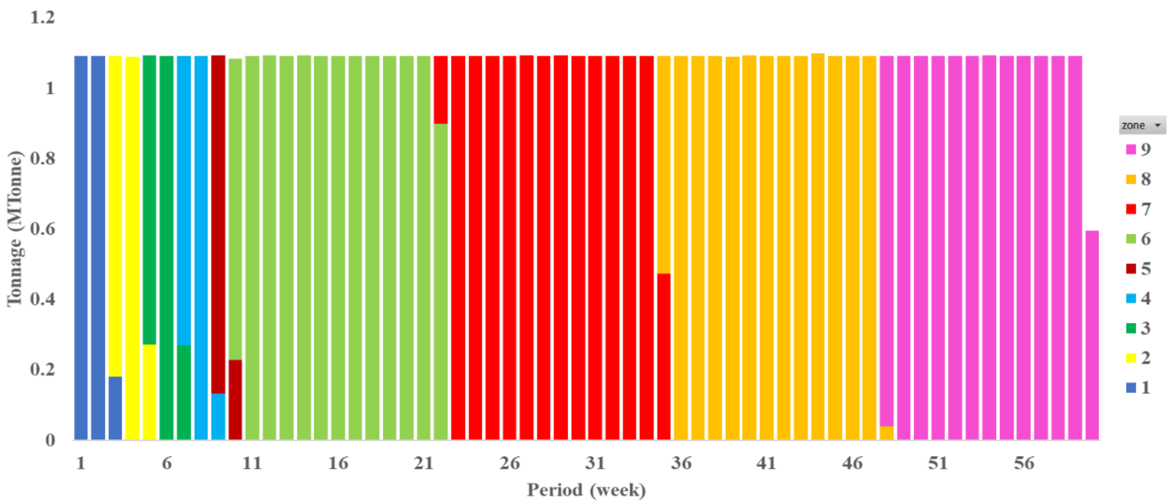

Figure 25. Weekly mining distribution schedule showing zones.

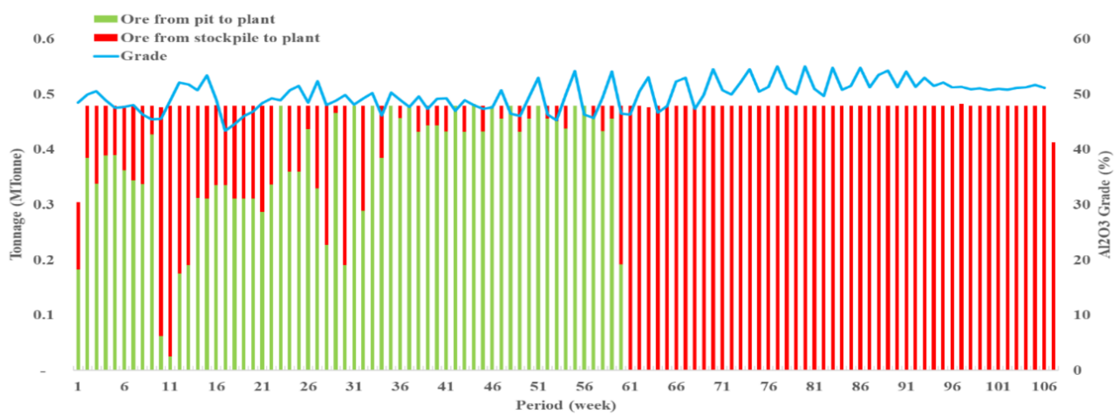

Figure 26 Weekly plant feed schedule.

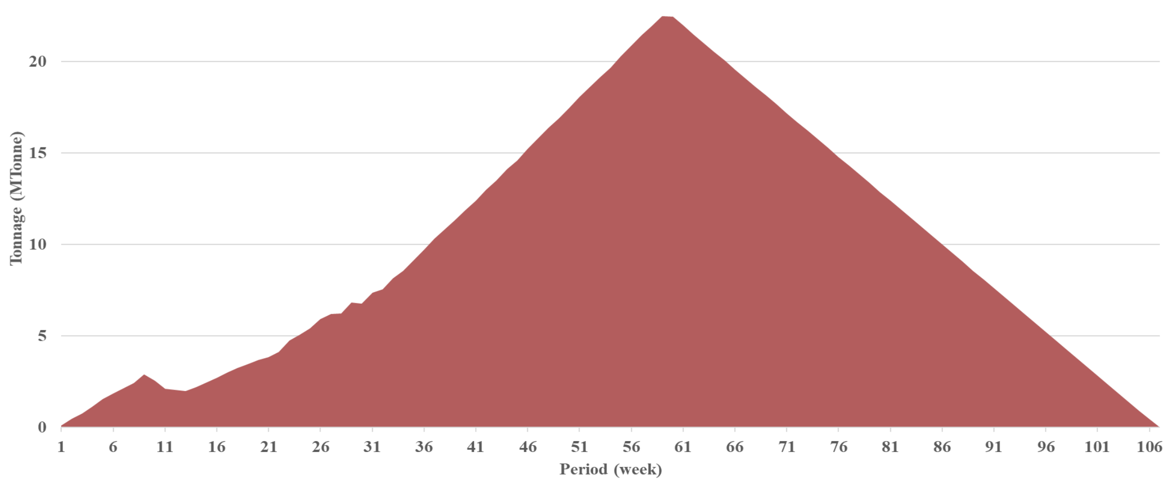

Figure 27. Stock Inventory.

8.4. Case 4 – Short term production with operational delays and uncertainties.

Uncertainty in operations is present, demonstrating variations in mining schedules. The following figures present the mining simulation results with delays and uncertainties. Figure 28 illustrates the weekly mining production schedule. Mining ends in period 59. The weekly mining distribution schedule is shown in Figure 29 showing the zones. Figure 30 shows the tonnages of ore sent to the processing plant. The grade behavior at the processing plant is similar to the other cases. For the first 9 periods in Case 3 the ore is 7,009,860 tons and for the same periods for Case 4 it moves 7,142,300 tons. That means that there is no ore to be mined in Period 10. Material is received by the plant at a constant feed rate from the stockpile starting in Period 62, with the stockpile at maximum capacity (Figure 31).

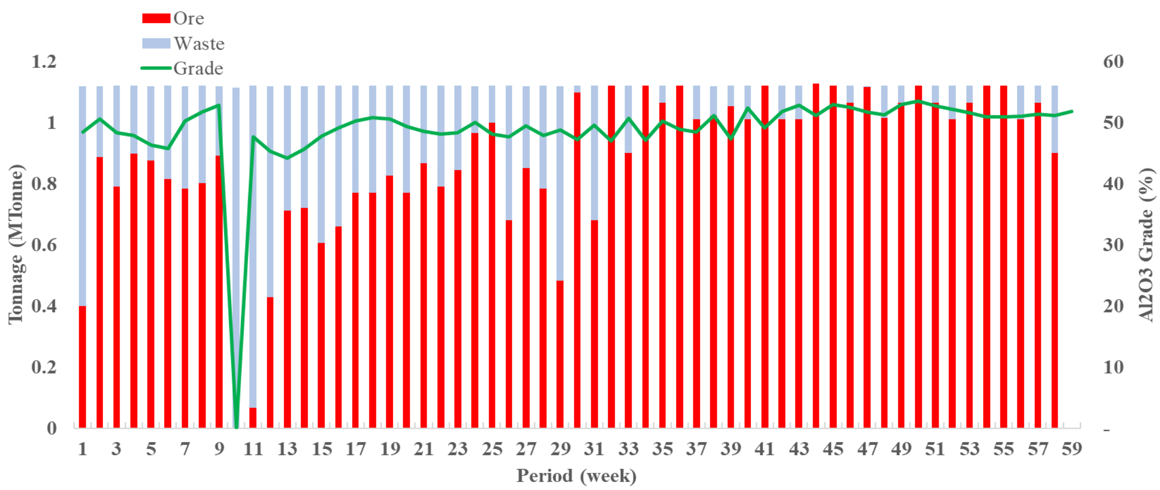

Figure 28. Weekly mining production schedule showing material types.

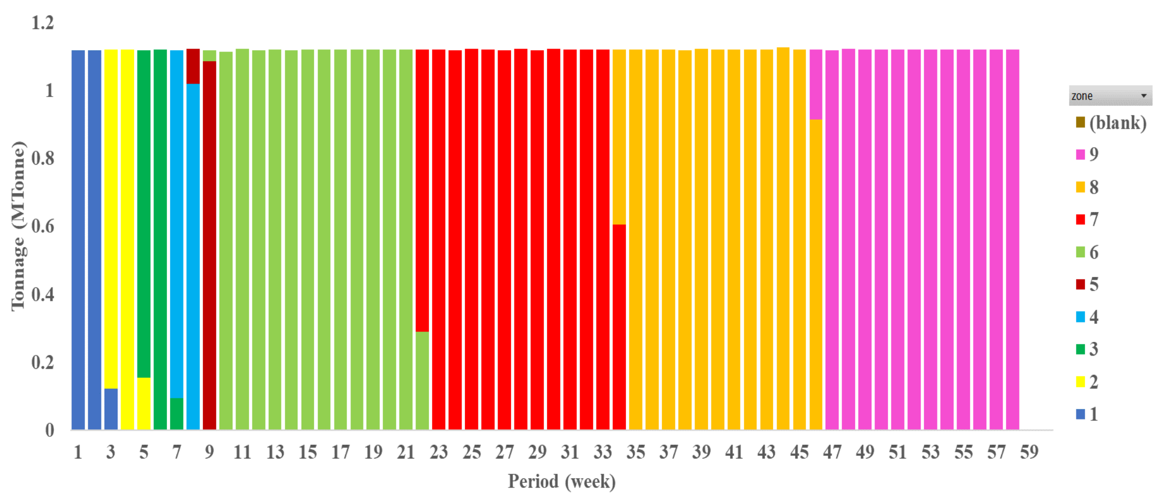

Figure 29. Weekly mining distribution schedule showing zones.

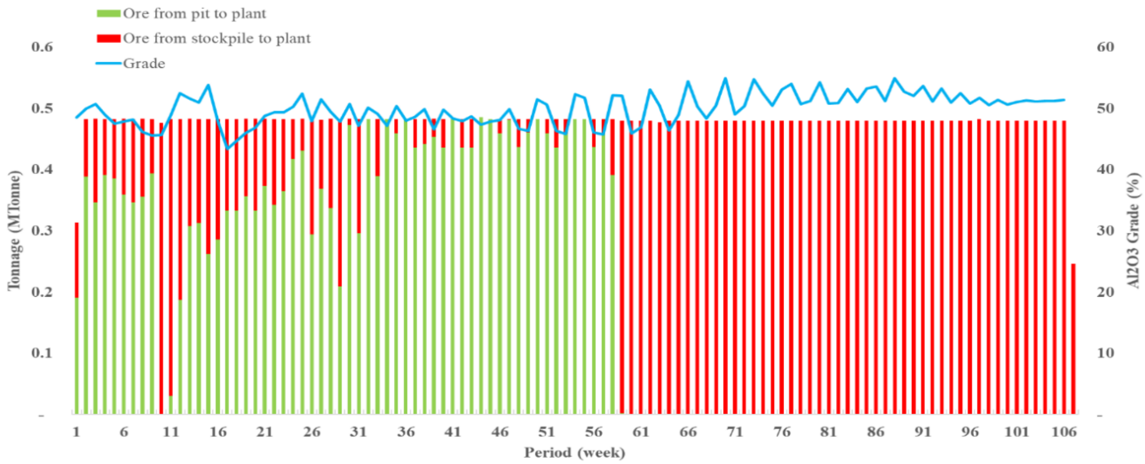

Figure 30 Weekly plant feed schedule.

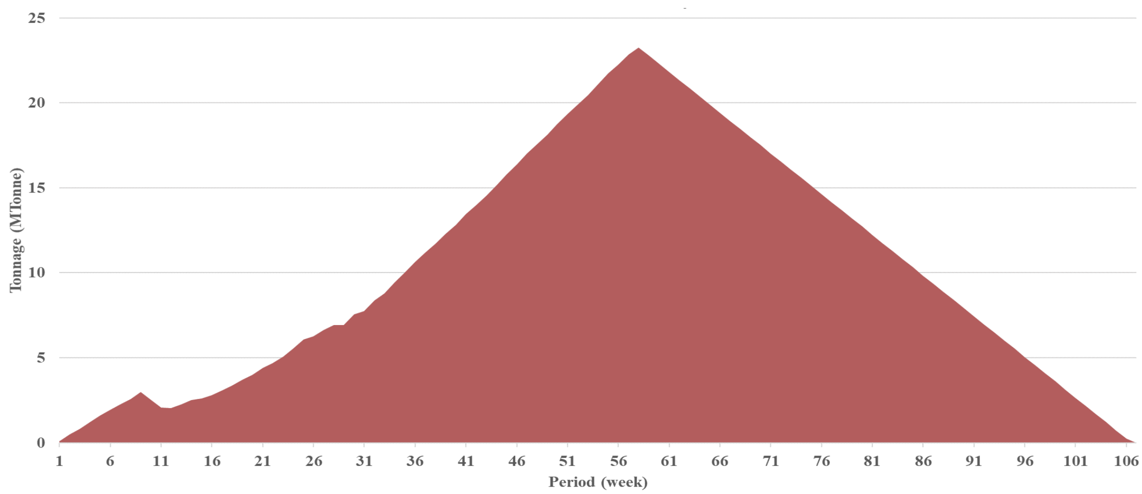

Figure 31 Stock inventory.

8.5. Case 5 – Short term production with operational delays and intelligent agent.

Results of smart agent interaction with mining simulation for this scenario are presented. The agent takes the decision* which zones are starting to be mined. Material has reached the selected destinations, and this shows how realistic the short-term production schedules are according to the objectives set for the intelligent mining agent. Figure 28 is a weekly mining production schedule showing material types. Weekly mining distribution schedule showing zones is shown in Figure 33 and the production of the shovels can be seen in the Figure 34. Two shovels work at the same time and change location in the zones according to the mining target. A zone is not mined until the zone above it has been mined. Figure 28 and Figure 33 present a fall in production in the last periods, this is due to the reason that only one shovel is working in zone 6 and the other one has moved to another work zone after finishing zone 7. The plant receives most material directly from the mine up to Period 62 (Figure 29), after this Period it receives material directly from the Stockpile. The stockpile reaches their maximum capacity in Period 50 with 23.6 Mt (Figure 36).

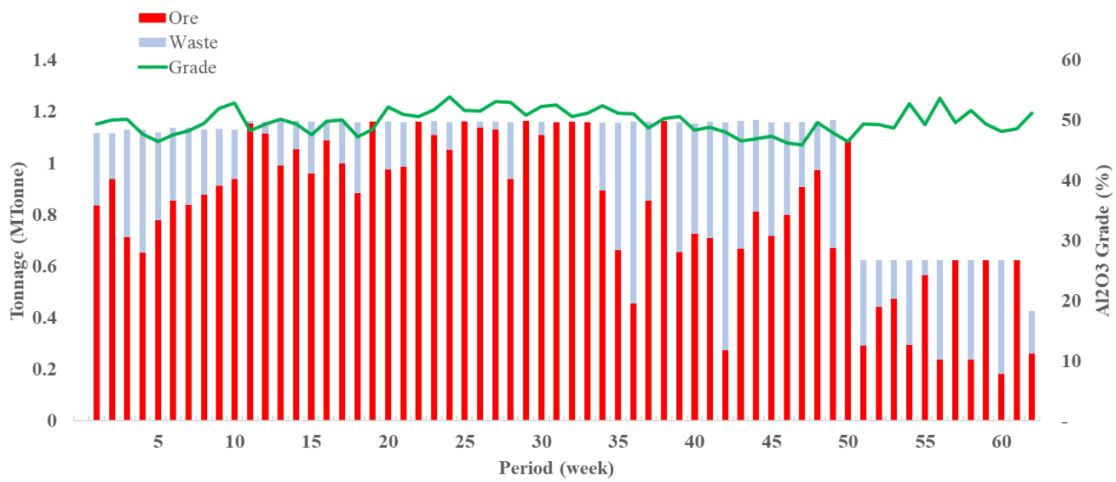

Figure 32. Weekly mining production schedule showing material types.

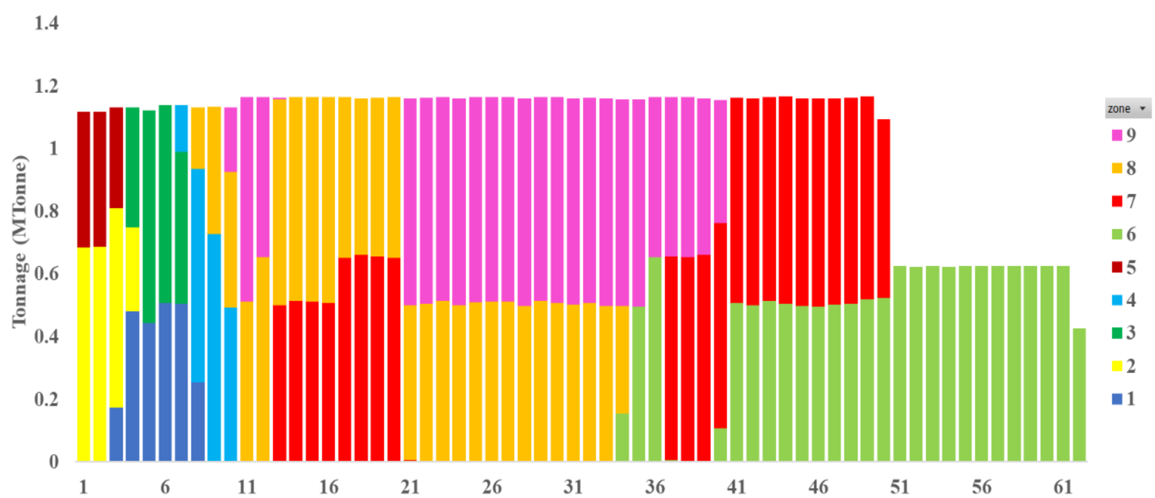

Figure 33. Weekly mining distribution schedule showing zones.

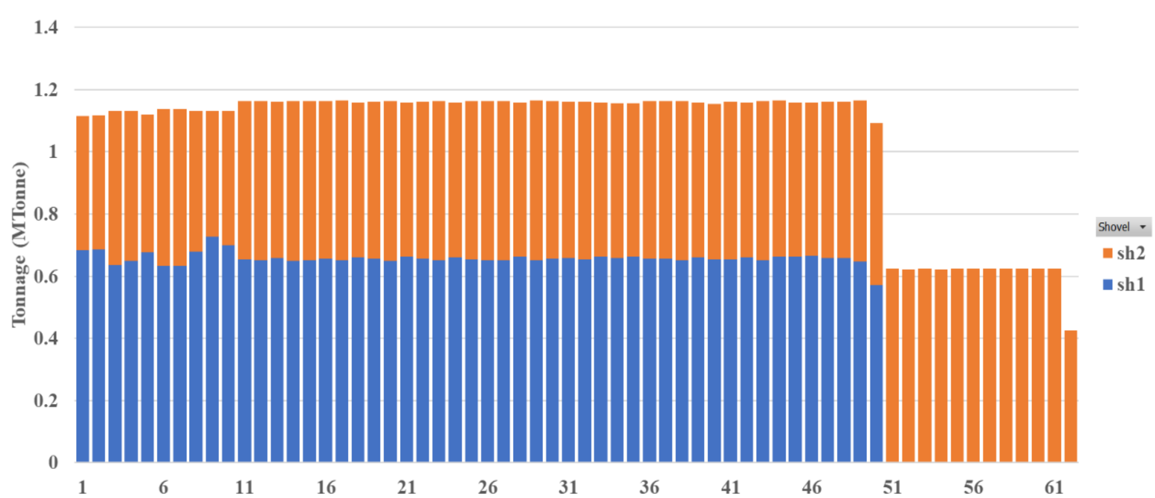

Figure 34. Weekly mining production schedule showing shovels.

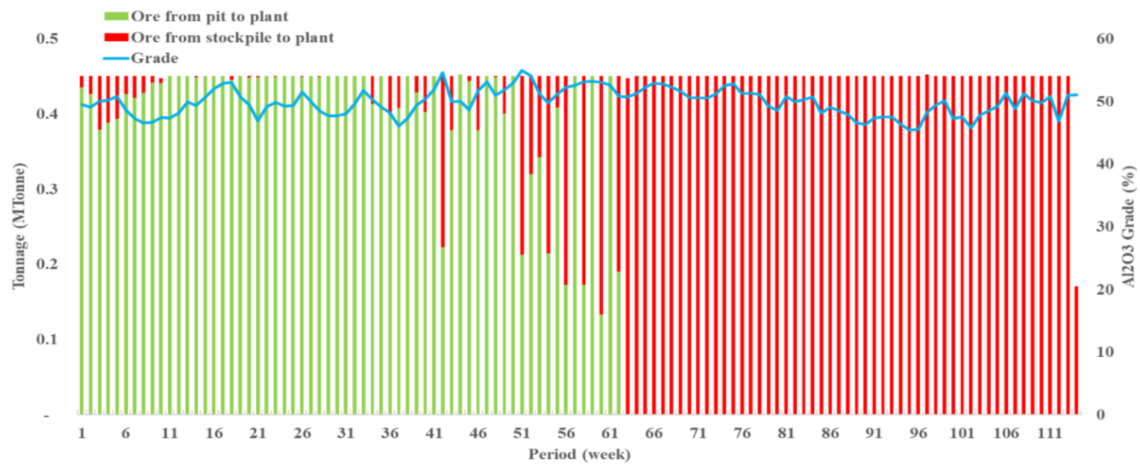

Figure 35. Weekly plant feed schedule.

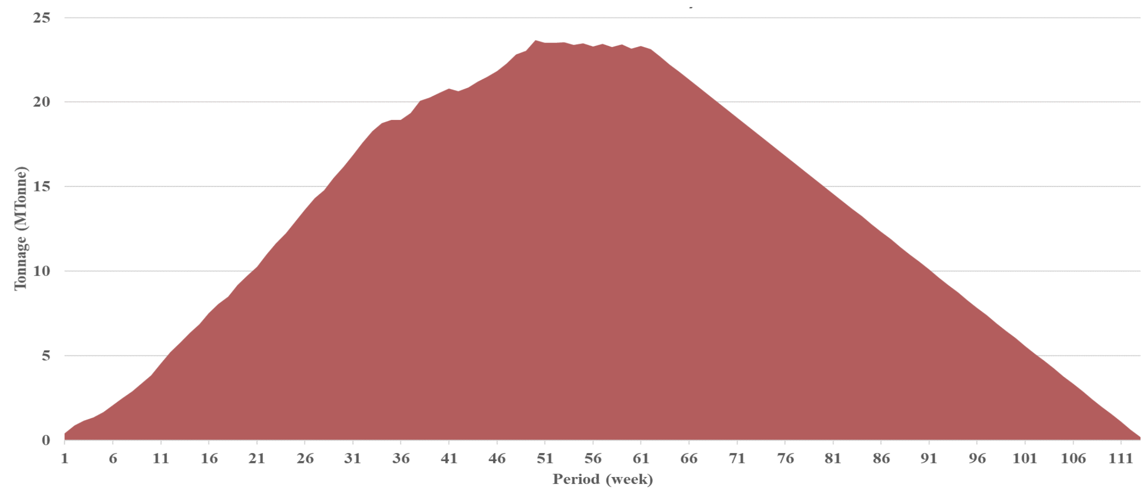

Figure 36. Stock inventory.

8.6. Case 6 – Short term production with operations delays, uncertainties with an intelligent agent.

This case presents the results of the smart agent with delays and operational uncertainty in the simulation. Mining simulation ended on week 61 for mining production and week 106 for plant (Figure 40). There are production variations in the mining sequence as presented in Figure 37 and Figure 38.

Zone 7 starts producing in period 22 (Figure 38), then it stops, according to the priorities of the intelligent agent and continues in Period 30 until the end of its production. The extraction of this zone is interrupted several times because the agent prioritizes other zones according to the objectives of the mine. It is inferred that only one shovel is mining zone 6 in the last periods (Figure 39) and the other shovel moved to another zone to start another mining period. The inventory of ore in Figure 41 reaches its peak in Period 51, at Period 62, the plant receives exclusively stockpile material.

At the beginning of the mining operation, the stockpile begins to receive material. This happens because if the queue of trucks waiting for the crusher is greater than 5, then the material is sent to the stockpile in order to make the trucks available.

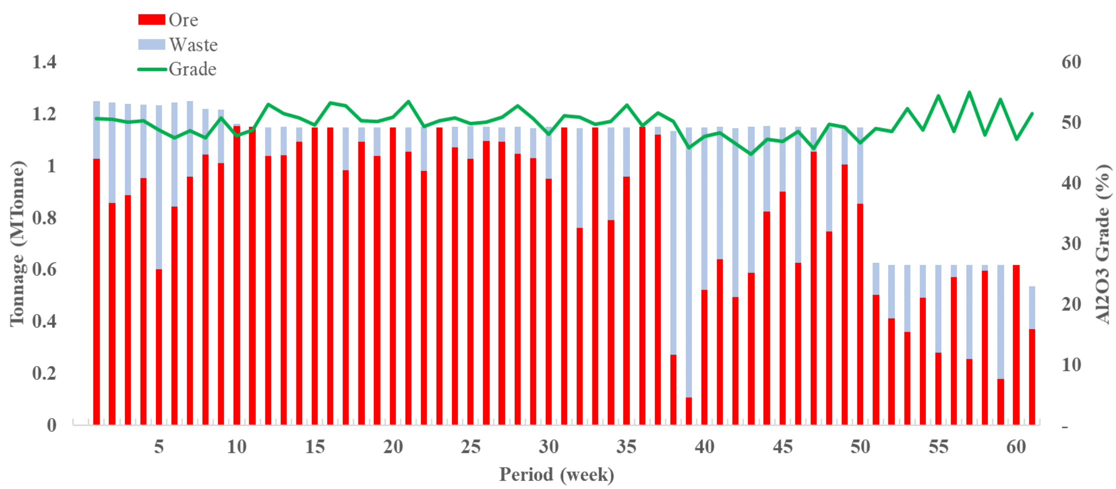

Figure 37. Weekly mining production schedule showing material types.

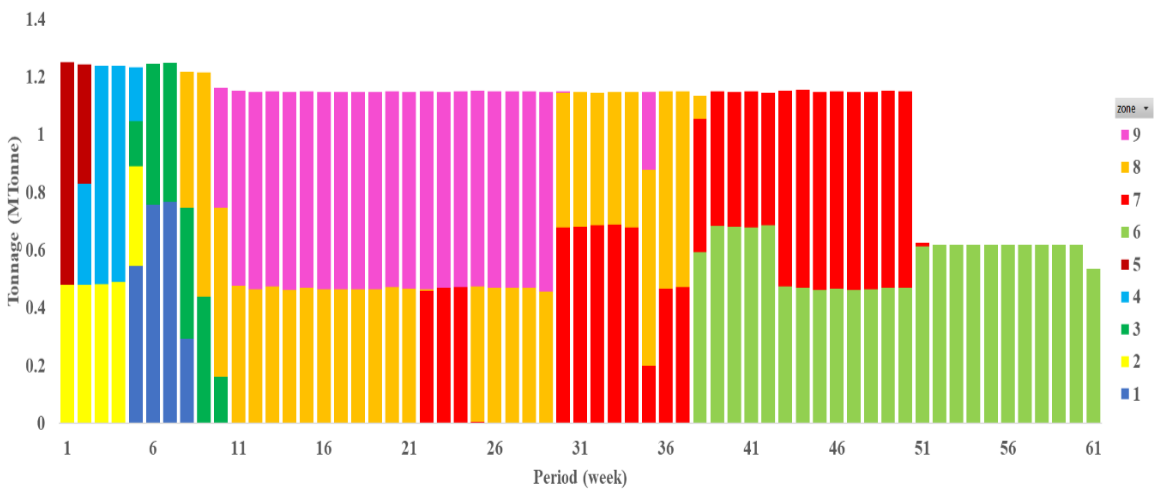

Figure 38. Weekly mining distribution schedule showing zones.

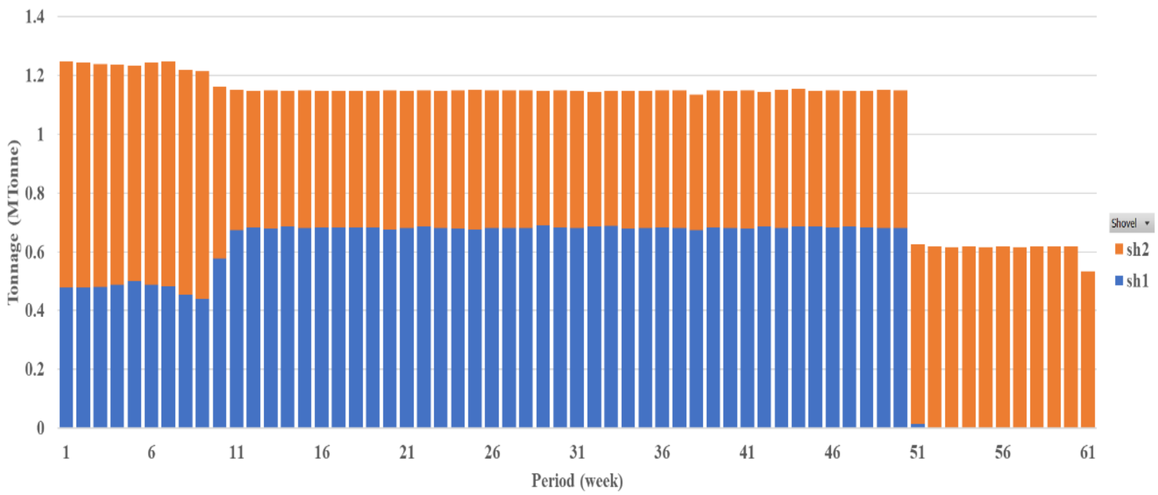

Figure 39. Weekly mining production schedule showing shovels.

Figure 40. Weekly plant feed schedule.

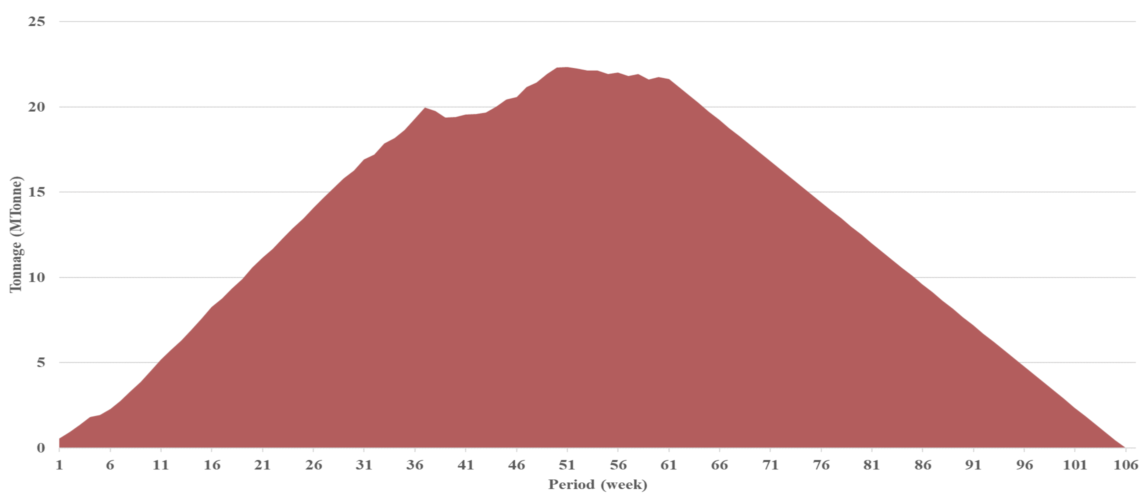

Figure 41. Stock inventory.

8.7. Case 7 – Short term production with operational delays, stockpiling with an intelligent agent.

The results of the integration of the intelligent agent with operation delays defined in the simulation are presented. The weekly mining production schedule (Figure 42) shows that less ore is mined after period 40. Figure 43 and Figure 44 shows 3 zones are being mined at the same time because 3 shovels are working at the same time. Figure 44 shows the weekly production per shovel. In the last periods only two shovels are working because another shovel is in another zone. In the weekly plant feed schedule (Figure 45) the first periods ore from the stockpile are processed. This is because in some periods, the direct shipment of mining does not cover the plant's requirements and it is necessary to send ore from the stockpiles. The plant grade from period 76 onwards is acceptable because at that same time the plant receives low grade ore from the stockpile. Figure 46 shows the ore inventory in the high- and low-grade stockpiles. The maximum capacity of the stockpiles is 22 MTonnes. The constant shipment of ore to the plant starts from period 61, time in which the mine has stopped sending ore to be processed.

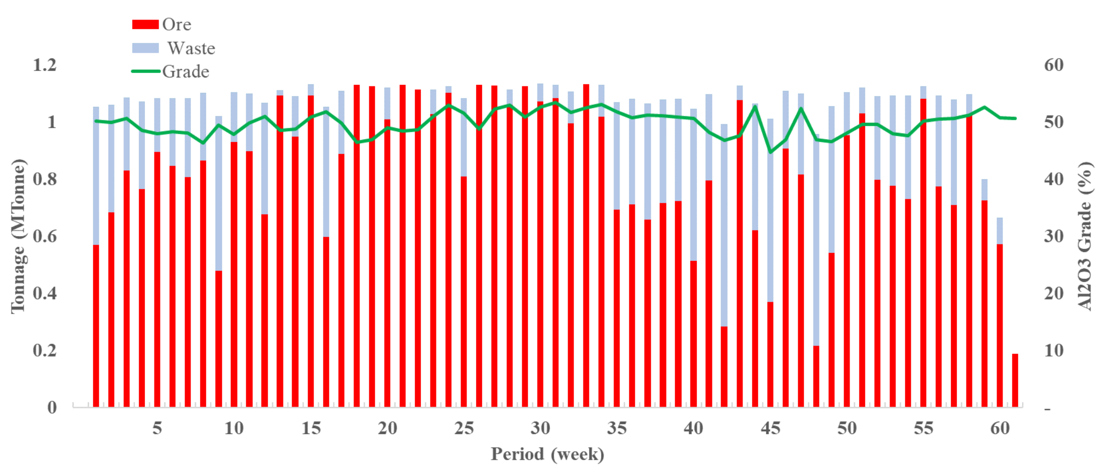

Figure 42. Weekly mining production schedule showing material types.

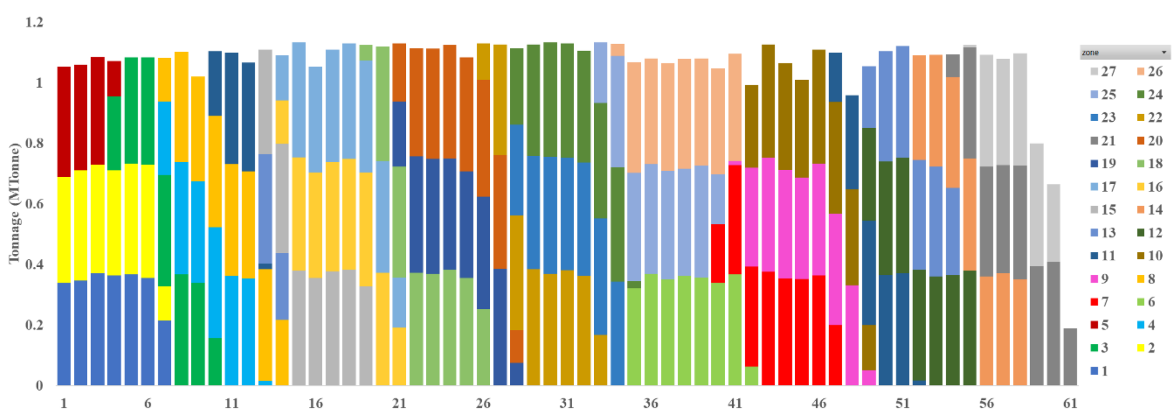

Figure 43. Weekly mining distribution schedule showing zones.

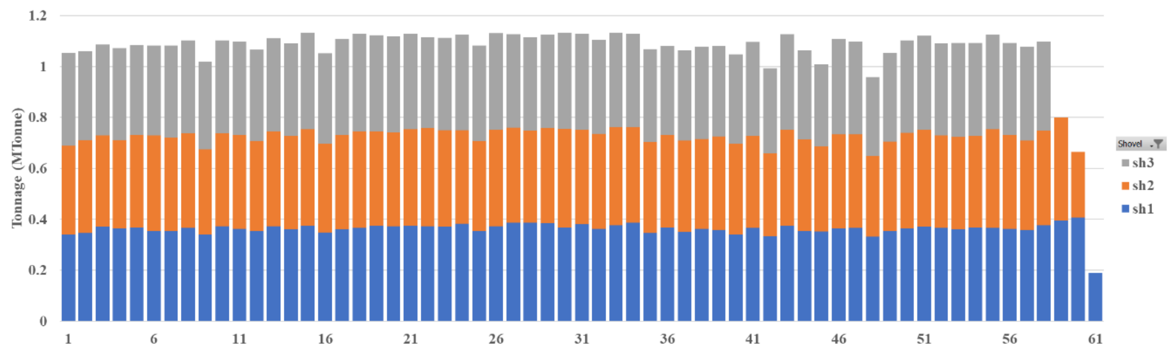

Figure 44. Weekly mining production schedule showing shovels.

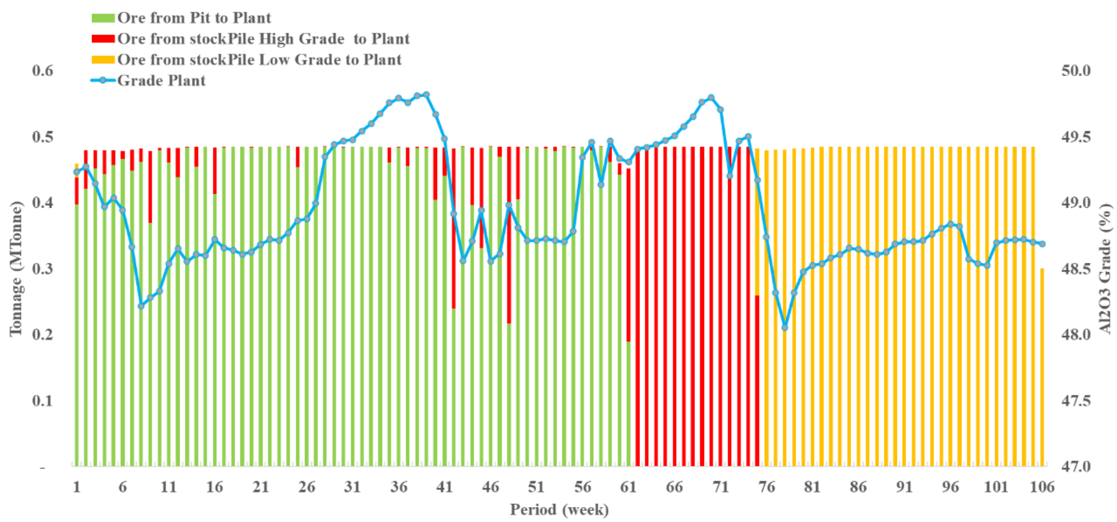

Figure 45. Weekly plant feed schedule.

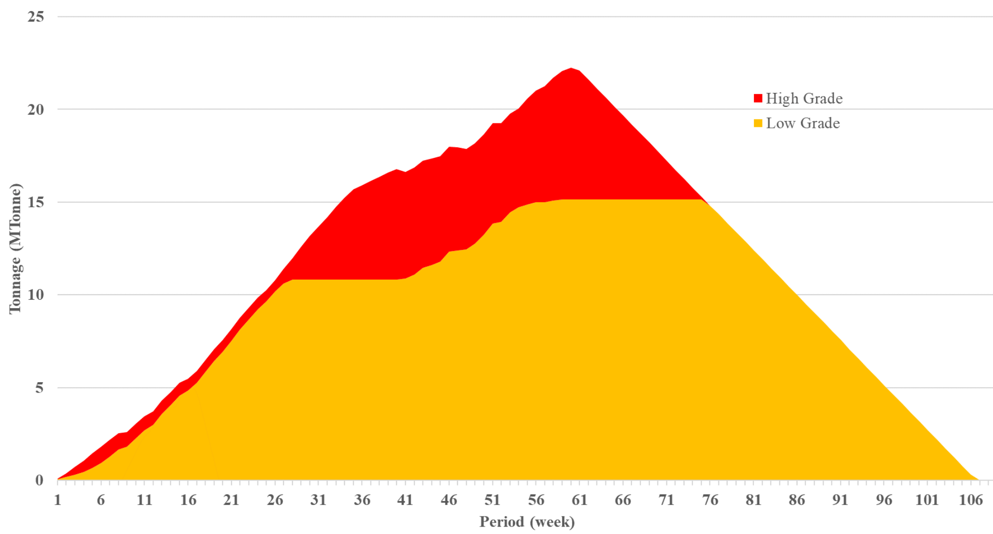

Figure 46. Stock inventory.

Intelligent agent act as supervisor, assigns equipment to zones to obtain better ore extraction with better grades. In this way, metal content is more stable and better and there is less material stored in the stockpile which decreases storage costs.

An analysis of the economic implications of the differences between the cases is provided. Results of the cumulative cash flow obtained in the different scenarios over the life of the mine are shown in Figure 47 for Cases 2, 3 and 5 and 7 and in Figure 48 for Cases 2, 4 and 6. These cases were grouped together to evaluate the impact of the smart agent on the cash flow.

After Period 66, it is observed that the cumulative cash flow for Case 5 starts to deviate from Case 2 and 3, as a result of lower metal content.

Cash flow for Case 5 starts from $6.1M in Period 1. Cumulatively it rises to $840.5 M in Period 114 once the life of mine is reached. Given the operating constraints of Case 3, cash flow at the end of the mine life is $839.3 M versus the $837.9 M obtained in Case 2. Finally, Case 7 has the best cash flow, with a value of $ 841.6 M.

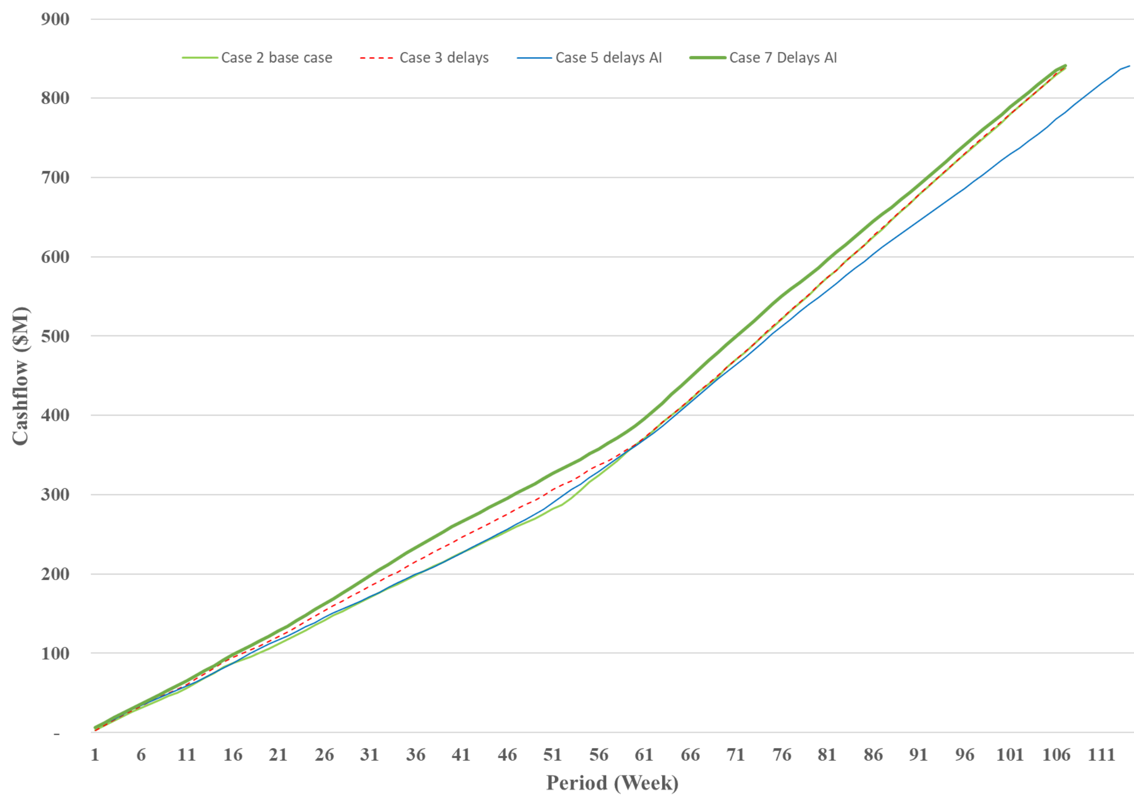

Figure 47. Cumulative cash flow Cases 2-3-5-7.

The current value curves show similar behaviour (Figure 47). Both Cases 2 and 4 are almost the same after Period 60 compared to Case 6. Following Period 26 until the end, the cumulative cash flow for Case 6 is seen to be higher in comparison to the other cases. The result is higher metal content during the early periods due to the higher ore extraction from the mine. This is shown in Figure 48.

Cumulative cash flow for Case 6 begins with $6.1 M in Period 1 and rises to $842.4 M in Period 106, when the life of mine is achieved. With the operating constraints and uncertainty of Case 4, the total cash flow at the end of the mine life is $839.0M, versus $837.9M for Case 2.

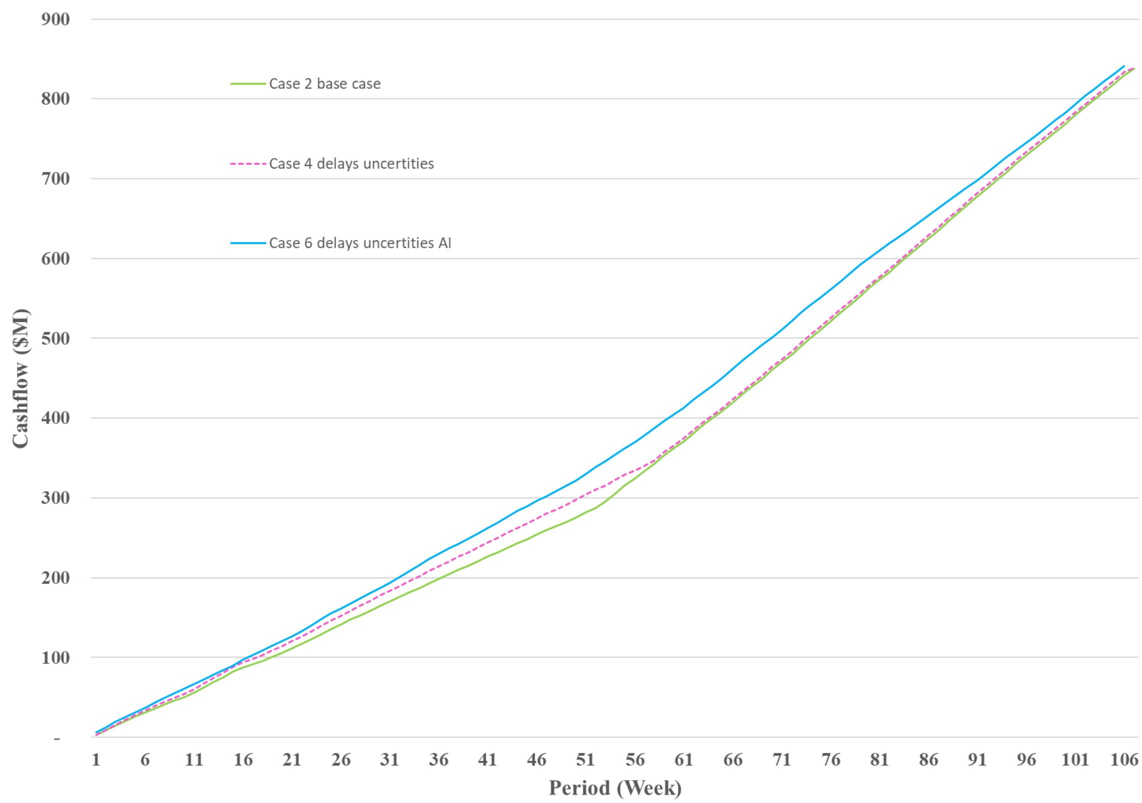

Figure 48 Cumulative cash flow Cases 2-4-6.

9. Conclusions

A case study was conducted for a bauxite deposit using a simulation model for a truck shovel mining. This paper analyzes the integration of an intelligent agent in a mining simulation to handle a real-time continuous mining environment. Research on the application of smart agents in mining using deep reinforcement learning is scarce, which opens a range of potential research opportunities. This research is innovative in the effort to provide a holistic approach to the use of artificial intelligence integrated with mine planning, operational delays, and uncertainty.

This research was performed for a bauxite deposit using a simulation model for a truck shovel mining system using AnyLogic [1]. The first 6 cases were carried out with 9 extraction zones, 2 shovels and 17 trucks (Phase 1). 27 zones, 3 shovels, and 14 for Case 7 (Phase 2). Case 1 replicated the mining plan generated by GEOVIA Whittle [29]. Case 2 was conducted in a short time frame for pushback 1, considering Period 8. Case 3 introduced operational delays, Case 4 added uncertainties, Case 5 admitted the supervision of an intelligent agent to the simulation with delays, and Case 6 incorporated uncertainties to Case 5. Case 7 included multiple stockpiles base on cut-off grade.

The last three cases visualize the working situation of the open pit mine truck haulage system using the simulation model supervised by the intelligent agent in decision making.

The present study developed an agent-based approach combined with discrete events to understand the impact of delays and uncertainties on decision-making related to truck speed, processing system capacity, and shovel placement made by an intelligent supervisor to optimize mine process operations.

The findings demonstrate the agent chooses zones with sufficient grade for mining. In addition, the agent actively works with two shovels at the same time, increasing the speed of the operation.

Furthermore, it was seen that the least active shovels are quickly given trucks to load. Uncertainties make any task more realistic; an intelligent agent can take these uncertainties into account when managing a mining operation and achieving production targets.

10. References

[1] AnyLogic (2021). Anylogic 8.7 Personal Learning Edition. Retrieved January 05,2022, from: https://www.anylogic.com/

[2] Askari-Nasab, H. (2006). Intelligent 3D interactive open pit mine planning and optimization. Ph.D Thesis, University of Alberta, Edmonton, Alberta, Pages 167.

[3] Bellamy, D. and Pravica, L. (2011). Assessing the impact of driverless haul trucks in Australian surface mining. *Resources policy*, *36* (2), 149-158.

[4] Bellman, R. E. and Dreyfus, S. E. (1962). *Applied dynamic programming.* Princeton University Press, Princeton, NJ, Pages 363.

[5] Bissiri, Y. (2002). Application of agent-based modeling to truck-shovel dispatching systems in open pit mines. Ph.D Thesis, The University of British Columbia, Pages 119.

[6] Bodon, P., Fricke, C., Sandeman, T., and Stanford, C. (2011). Modeling the mining supply chain from mine to port: A combined optimization and simulation approach. *Journal of Mining Science volume 47* (2), 202-211.

[7] Brownlee, J. (2021). Deep Learning. Retrieved from: https://machinelearningmastery.com/

[8] Chaturvedi, M. (2021). What are dqn reinforcement learning models. Analytics India Magazine.

[9] Choudhary, A. (2019). A hands-on introduction to deep q-learning using openai gym in python. Analytics Vidhya. .

[10] DeepMind (2021). What if solving One problem could UNLOCK solutions to thousands more? Deepmind. . Retrieved from: https://deepmind.com/

[11] Gagniuc, P. (2017). *Markov Chains: From Theory to Implementation and Experimentation.* John Wiley & Sons, New Delhi, India, 1st Edición ed, Pages 256.

[12] Glockner, C. (2018). Simulators: The KEY training environment for applied deep reinforcement learning. Medium. . Retrieved from: https://towardsdatascience.com/simulators-the-key-training-environment-for-applied-deep-reinforcement-learning-9a54353f494f.

[13] Gomez, J. O. and José Alfredo Jiménez García, A. V. B. A. a. A. J. R. L. (2018). Analysis of the impact of the increase of production volume on the material supply system and work in process through simulation scenarios. *International Journal of Supply Chain and Operations Resilience*, *3* (3), 260-273.

[14] Gou, S. (2019). DQN with Model-Based Exploration: Efficient Learning on Environments with Sparse Rewards.

[15] Grigoryev, I. (2019). AnyLogic in three days a quick course in simulation modeling, 5th ed.

[16] Holford, N. H., Hale, M., Ko, H., Steimer, J., Sheiner, L., Peck, C., Bonate, P., Gillespie, W., Ludden, T., and Rubin, D. (1999). Simulation in drug development: good practices. *Center for Drug Development Science (CDDS)*, *1* 23.

[17] Hosseini, E. B.-A. a. N. S. (2017). An economic evaluation of a primary haulage system for a bauxite mine: load and haul versus in-pit crushing and conveying. . *Report Eight. MOL* 14.

[18] Huberman, B. A. and Glance, N. S. (1993). Evolutionary Games and Computer Simulations. *Proceedings of the National Academy of Sciences - PNAS*, *90* (16), 7716-7718.

[19] Jovanoski, B., Polenakovik, R., Gecevska, V. and Minovski, R., (2014). *Applying a suitable simulation approach for processes on different management levels*. in Proceedings of XVI INTERNATIONAL SCIENTIFIC CONFERENCE ON INDUSTRIAL SYSTEMS, University of Novi Sad - Faculty of Technical Sciences Department of Industrial Engineering and Management, Novi Sad, Serbia., pp. 327-332.

[20] Kelton, W. D., Sadowski, R. P., & Zupick, N. B. (2015). *Simulation with arena.* McGraw-Hill Education, New York, N.Y., Sixth edition ed, Pages 635.

[21] Langley, P. and Simon, H. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, *38* (11), 54-64.

[22] Lazic, L. and Velasevic, D. (2004). Applying simulation and design of experiments to the embedded software testing process. *Software testing, verification & reliability*, *14* (4), 257-282.

[23] Minkah, E. A. (2014). Determination of optimized Inchiniso pit, Ghana Bauxite Company Limited. MSc Thesis, University of Mines and Technology, Tarkwa, Pages 86.

[24] Ozdemir, B. and Kumral, M. (2018). Appraising production targets through agent-based Petri net simulation of material handling systems in open pit mines. *87* 138-154.

[25] Sánchez, F. and Hartlieb, P. (2020). Innovation in the Mining Industry: Technological Trends and a Case Study of the Challenges of Disruptive Innovation. *Mining, metallurgy & exploration*, *37* (5), 1385-1399.

[26] Shishvan, M. S. and Benndorf, J. (2019). Simulation-based optimization approach for material dispatching in continuous mining systems. *European Journal of Operational Research*, *275* (3), 1108-1125.

[27] Shiv Prakash, U. and Hooman, A.-N. (2018). Simulation and optimization approach for uncertainty-based short-term planning in open pit mines. *International journal of mining science and technology*, *28* (2), 153-166.

[28] Silver, D. (2009). Reinforcement learning and simulation-based search in computer Go.Thesis, University of Alberta,

[29] Systèmes, D. (2019). Geovia Whittle. Strategic mine planning software. Ver. 4.7.3, Vancouver. Ver.

[30] Systèmes, D. (2020). Geovia GEMS. Strategic mine planning software. Ver. 6.8.3, Vancouver. Ver.

[31] Vasquez-Coronado, P. P. (2014). Optimization of the haulage cycle model for open pit mining using a discrete-event simulator and a context-based alert system. MSc Thesis, The University of Arizona, Tucson, Pages 156.

[32] Vidhya, A. (2020). Reinforcement Learning via Markov Decision Process. .

[33] Wiki, A. (2021). Gradient Descent. Retrieved from: https://docs.paperspace.com/machine-learning/wiki/gradient-descent

[34] Wiki, A. (2021). Weights and Biases. Retrieved from: https://docs.wandb.ai/

[35] Zhou, V. (2019). How to take advantage of artificial intelligence in the mining industry. Australian Mining. . Retrieved Australian Mining from: https://www.australianmining.com.au/features/how-to-take-advantage-of-artificial-intelligence-in-the-mining-industry/