Data analytics for open-pit mining: examining vehicle interactions, material movement and compositional uncertainty with bucket inference and Monte Carlo simulation

A PREPRINT

Raymond Leung*, Konstantin Seiler, Andrew Hill and Xin Zhao

Rio Tinto Sydney Innovation Hub
Faculty of Engineering
The University of Sydney
Sydney, NSW 2006
raymond.leung@sydney.edu.au

September 15, 2023

ABSTRACT

This article examines the use of data analytics in mining automation. It considers the benefits of integrating telemetry data in computational systems that model the state of a mine. Specifically, it focuses on interactions between load-haul vehicles and the information that may be harvested to facilitate material tracking in surface mining. Bucket dig positions are inferred from GPS data by analysing wheel-loader and excavator interactions with haul trucks. Evaluation shows this approach achieves higher precision and load-haul cycle recall, as extra buckets missing from the existing OEM source are discovered. Analysis reveals a consistent pattern of behaviour from equipment operators which justifies the use of a simple model that neglects the articulated motion of the excavator arm. A major contribution is the integration of two technical objectives: the ability to consistently locate excavation points within mining pits based on vehicle interactions, and describe compositional variation in the excavated material using accurate local grade models informed by geochemical assays. This allows material movement associated with individual load-dump events to be tracked and linked with the underlying geology via kriging estimates. The case study highlights the importance of having high-resolution data, as it enables transient ore dilution events (e.g. inadvertent transferral of waste to stockpiles) to be identified and factored into risk assessment. The information generated may be used for ore-properties tracking in a dynamic graphical model to capture correlations, uncertainties and compositional changes induced by material movement. Recent development and application opportunities including the role of machine learning are briefly discussed.

Keywords Load-haul-dump operations · mining excavation · material tracking · open-pit mining · compositional changes · grade risk

1 Introduction

This paper advocates the use of data analytics in open-pit mining. A specific objective is to combine load-haul vehicle telemetry data with orebody knowledge to improve risk perception. This risk principally refers to the likelihood of digging up material that deviates significantly from model prediction (e.g. the expected grade of iron within a small section of the pit), and inadvertently, transporting this material to the wrong destination. For instance, sending mine waste which has been mistaken as ore to a stockpile instead of a waste dump. Within surface mining, data analytics [1] may be broadly interpreted as computational intelligence or scientific analysis of information that leads to better insight or improvement in how a mine operates. Traditional information gathering includes geochemical analysis

*Corresponding author

of drill-hole samples and logging geophysical [2] or geomechanical [3] signals while drilling during the exploration and mine production phase. These standard practices provide a foundation for creating 3D resource models which describe mineral abundance and rock properties of an orebody. This knowledge can guide long-term mine planning, such as pit design and geotechnical investigation which enable mines to be developed and ores to be extracted in some optimal way. In contrast, short-term mine planning is responsible for the bulk of earthmoving activities and dynamic changes observed in an open-pit which continuously shape the mining terrain. Mobilisation of mining equipment and effective coordination are essential in ensuring load-haul operations are completed efficiently and on schedule. To satisfy productivity and quality constraints, vehicles are instructed on where to move, when to dig and where to off-load, how much material to transfer from a source to another location over a given time period, in order to achieve certain distribution, flow-rate, volume and grade targets [4]. In the process, these vehicles leave behind GPS footprints which allow material movements to be traced.

For excavation purpose, a mining bench is usually partitioned into grade-blocks. Grade-block boundaries are drawn to decompose a region into smaller mining units. Ideally, each grade-block maps to one interim/final destination and its composition is homogeneous—even though this is not generally true in practice. Although the average grade of the excavated material is computed for each grade-block, grade variation is commonly ignored as the grade value is assumed to be piecewise constant within the mining unit. Since grade-blocks are typically many tens to hundred times the size of a truck, these approximations make it difficult to capture the true variation (especially, large gradients and discontinuities) in the composition of an ore deposit, to properly inform or mitigate grade risk associated with digging in geologically complex areas. Furthermore, the dig bucket positions are often reported with low spatial fidelity. These deficiencies obfuscate the risk of grade and positional uncertainties, particularly near mineralisation boundaries where there is a high-to-low grade transition. An inability to resolve these differences, even as a continuous distribution of the expected grade, presents a very real risk of confusing ore and waste. This can lead to stockpile contamination and increase grade control risk, potentially causing ore blends to deviate from product specification. To minimise such impact, this study focuses on exploiting geospatial data to increase spatial awareness, through monitoring the interactions between earthmoving equipments; and being more attentive to local changes in composition, through fine-grained characterisation of excavated material rathering than modelling ore properties in bulk [5]. This vision is shown in its essence in Fig. 1.

First, dig location detection marks the start of a sequence of earthmoving events. In Fig. 1, E_1 and T_1 show a typical interaction between excavators and trucks. Second, the haul trucks T_2 and T_1 carry high-grade (red) and low-grade (blue) payload respectively. Their movements are tracked so that the source material can be mapped to destinations such as stockpiles and waste dumps. Third, the contours superimposed to the mining pit (in varying shades of red/blue) provide a probabilistic interpretation of the material. In the simplest case, they may represent the likelihood of ore grade exceeding some specified threshold. Alternatively, they may represent the risk of transporting certain material to the wrong destination. For instance, it is safe (respectively, risky) for E_2 to excavate the red (respectively, blue) material and load truck T_2 which is destined for a stockpile. This assessment conveys the risk of ore dilution and takes into account the compositional uncertainty of the material in the pit.

As a background note, geology modelling and material tracking are often treated as separate subjects in geostatistics and automation literature. For geology modelling, implicit surface modelling techniques have been used in [6] and [7] to capture intricate, large-scale geological structures of orebodies in iron ore deposits [8]. For grade prediction and uncertainty estimation, two prominent approaches are simulation of Gaussian random fields using Gibbs sampling [9] and Bayesian block estimation with Gaussian Processes [10, 11]. These works are concerned with characterising orebody composition. For material tracking, holistic approaches are scarce, although there are some promising dynamic probabilistic graphical models under development [12] which extend the discrete state-space formulation described in [13, 14] where digging and dumping events trigger volumetric and compositional changes (updates) in a voxelated 3D modelling space. Other aspects of stochasticity, such as the spatial uncertainty around material excavation have also been investigated in [15].

Within the realm of computer vision and machine learning, advanced data analytics have also been applied to excavation and mining equipment inference problems in interesting ways. In stockpile geometry modelling, laser scanners [16] and boundary segmentation techniques [17] have been used to track the shape of stockpile frontier. In automated load-haul cycle segmentation, hidden semi-Markov models and unsupervised learning [18] have been applied to truck GPS data to analyse vehicle behaviour, label transitions and infer the action (e.g., queue, load, move, dump) being performed at various times. In excavator activity recognition, deep neural network architectures are used to classify actions using video sequences captured by cameras mounted on excavators [19]. This is used to derive insight on the productivity of earthmoving equipment in the construction industry; however development or adaptation of similar technology is lagging in the mining industry [20]. This survey reveals a fragmented technology landscape within mining, and a general disconnect between orebody knowledge and material tracking in published literature. This paper aims to bridge this gap and provide examples of applications that may benefit from proper integration of these ideas.

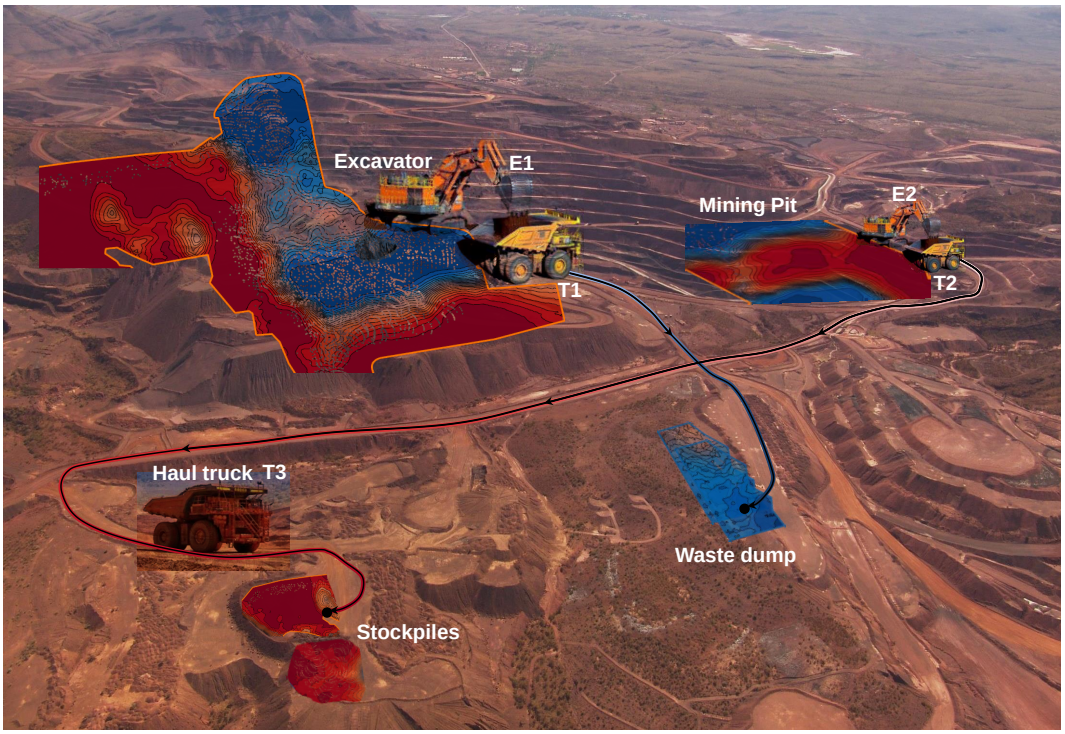

Figure 1: Conceptual representation of the key aspects in a material tracking system. (a) Monitoring interactions between earthmoving equipments such as excavators and trucks, see E_i and T_i. (b) Mapping source material and grade uncertainty to destinations such as stockpiles and waste dumps. (c) Probabilistic expression of material composition in the mining pits. Contours may be used to convey the risk of ore dilution associated with specific excavation and load-dump operations. Photo courtesy of Ulrich Hauser.

In terms of conceptual development, this paper demonstrates that incremental performance gain (in particular, increased recall of digging events in load-haul cycles) is possible by proactively monitoring the interactions between excavators and trucks using only GPS data. This minimises information loss which is important in itself as the performance of any tracking algorithm is dependent on the the quantity (completeness) and quality of the data it is served. The second contribution is adding a geology model to the excavator bucket position estimator, performing compositional and spatial analysis both at sufficiently high resolution. This is made possible by acknowledging and modelling grade variation within grade-blocks and estimating the digging location with greater precision than a mining unit which is represented by a polygonal enclosure that is at least an order of magnitude larger than a truck. Increasing spatial resolution is the key to improving grade risk perception and facilitating precision mining. This has not been demonstrated previously in literature, as far as the authors are aware. Apart from Innes [13, 14], Balamurali [15] and Bailey [12], few published works focus explicitly on load-haul cycles and attempt to use vehicle movements to track material stochastically through the mining chain. This work exploits the load-haul-dump (LHD) and geology link and uses vehicle interactions to infer material movement at higher spatial resolution and shorter time scales than the established norm. The contribution of individual buckets to the composition (grade) of stockpiles is examined through graphical means.

As an overview, this paper considers data analytics as an enabling tool that facilitates the achievement of long-term objectives in mining automation. A specific application is tracking material movement and compositional changes through spatial analysis, where dig positions are estimated by observing interactions between loaders and trucks. As motivation, Sec. 2 lays the foundation for understanding the key aspects of the problem and why it matters. Sec. 3 is devoted entirely to bucket position estimation. It describes the proposed method and interprets the experiment results. Sec. 4 considers the articulated motion of excavators and justifies the choice for a simple model. Sec. 5 focuses on application. This exposition demonstrates the steps involved in linking bucket position estimates with ore properties and how simple mappings from source to destination (e.g., pit to stockpile) transfer grade risk. Grade uncertainty and manifestation of grade risk are visualised using contour plots and Monte Carlo simulation.

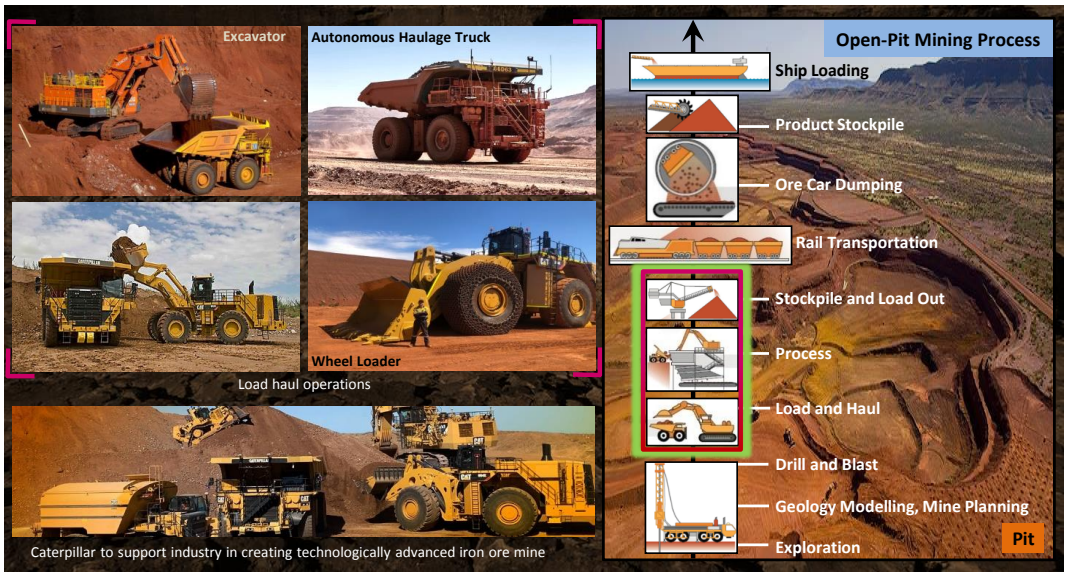

Figure 2: Material movement resulting from excavation and load-haul activities. For details, refer to [21].

2 Background

Open-pit mining entails many complex operations, these range from resource development, mine planning, fleet scheduling to ore shipment. The relevant aspects of this work are the digging and loading activites performed by excavators and wheel-loaders, including the transportation of material by haul trucks. These activities are depicted in Fig. 2(left) and the scope of these operations is highlighted in Fig. 2(right). For readers who are perhaps less familiar with operations in an open-pit iron ore mine, the processes described in [21] provide the background needed to contextualise this work. Interested readers are also referred to [21] for a brief overview of the problems and technologies encountered in the mining supply chain.

The reasons for focusing on digging and load-haul activities are two-fold. First, a detailed account of material movement and material composition matters—uncertainty about either can compromise grade control and processing downstream. Second, these activities amount to major logistical problems over time. To put this into perspective, approximately 1 billion tonnes of rock is moved each year. This is enough to fill a giant stadium like the Melbourne Cricket Ground every two days. Its sheer scale translates to significant cost and meaningful opportunities for improvement.

The term 'load-haul cycle' plays a pivotal role in subsequent discussion. A load-haul cycle commences when a digger transfers a bucket of excavated material into the back of a haul truck. Such interaction typically occurs multiple times until the truck becomes fully loaded. The truck then leaves the loading site to off-load this material at a dump destination—this journey constitutes a single load-haul cycle.

To facilitate a proper understanding of these movements, Fig. 3 shows the GPS traces of vehicles that participate in these load-haul events. In Fig. 3, the blue/orange/green/red lines show the routes travelled by several haul trucks. In these examples, the trajectories extend from a source in the mining pit (e.g., PIT_AREA_1) to various destinations (such as the crusher and STOCKPILE_Y3). For data analysis, it is the digging location that holds the key to what and where the material is sourced.

Delving deeper, Fig. 4 shows the data available for analysis. It explains the current limitations and the proposed plan. The existing system uses only data from the *lh_cycles* and *lh_buckets* tables to model material movement. In *lh_cycles*, the `digging equipment` and `loaded vehicle` fields allow the interacting vehicles (wheel-loader/excavator and truck) to be identified. Although the `dig location` and `dump location` are given, they represent only the area code for the source and destination rather than 3D coordinates. To obtain dig and dump position coordinates, the *lh_cycles* table must be joined with *lh_buckets*. Unfortunately, the coverage resulting from this sql-merge is incomplete. Typically, 9-13% of `lh_cycles` have no corresponding `lh_buckets`. In other words, only 87-91% of the load-dump intervals are recalled. Furthermore, the `dig position` and `dump position` estimates from the *lh_buckets* table also tend to be inconsistent and less accurate. This calls for improvement in spatial localisation and data completeness.

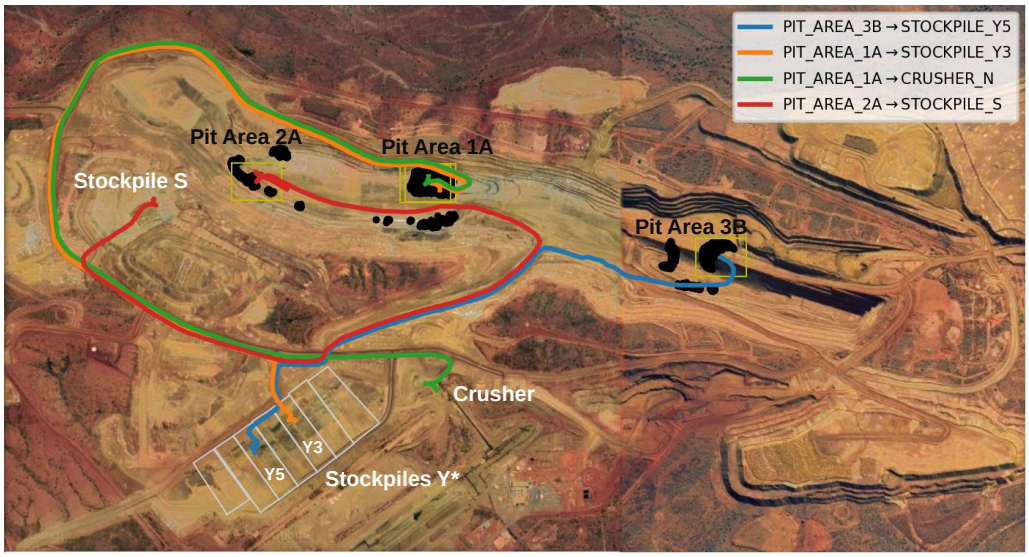

Figure 3: Trajectories of trucks travelling between a source and destination during a load-haul cycle (Aerial photo courtesy of Google Earth)

Load-haul tables in PostgreSQL database

lh_cycles

lh_cycle_id [primary key]

Vehicles ★
- Digging equipment (wheel-loader/excavator ID)
- Loaded vehicle (truck ID)

Locations
- Dig location (source area code)
- Dump location (destination area code)

Timing ★
- Load timestamps (arrival, load-start, full)
- Dump timestamps (arrival, dump-start, empty)
- Durations (haul, queue etc.)

Geology and material properties
- Grade-block name (only for in-pit excavation) ✠
- Material description (high grade, waste…)
- Payload

lh_buckets

lh_bucket_id [primary key]
lh_cycle_id [foreign key]

Vehicle
- Digging equipment ID

Timing
- Digging duration
- Load timestamp

Bucket positions
- Dig position (approx. coordinates)
- Dump position (into back of truck)

LINKS
★ Relate load-haul cycle to equipment position using vehicle ID and timestamps.
✠ Relate excavated material to grade-block properties (deduce average composition)
✪ Infer grade properties of individual buckets using relevant blasthole assays.

Tables with spatial and geochemical data

equipment_positions

equipment_position_id [primary key]
- Equipment type (wheel-loader, excavator or truck)
- Vehicle ID ★
- Timestamp ★
- Position (coordinates) ✪
- Orientation

gradeblock_properties
- Grade-block name ✠
- Grade-block geometry (polygon) ✪
- Grade-block average composition

hole_properties

blasthole_id [primary key]
- Blasthole coordinates+dims ✪
- Blasthole assays (% Fe, SiO_2, Al_2O_3…)

Figure 4: Overview of available data. The existing system uses only information in the *lh_cycles* and *lh_buckets* tables to model material movement. The proposed system analyses vehicle interactions using GPS data from the *equipment_positions* table, as well as geometry and chemical assay data from the *gradeblock_properties* and *hole_properties* tables to infer and track the composition of each bucket at much finer scale.

The proposed system incorporates GPS data from the *equipment_positions* table, as well as geometry and chemical assay data from the *gradeblock_properties* and *hole_properties* tables (see Fig. 4). For instance, using the vehicle IDs and loading time interval from *lh_cycles*, it is possible to analyse the interactions between the excavator and truck using GPS data from the *equipment_positions* table. The bucket dig position estimation approach will be described in Sec.3. Using the `grade-block name` in *lh_cycles* allows the average composition of the excavated material to be deduced from the *gradeblock_properties* table. This too suffers from inadequate resolution due to the piecewise constant grade restriction. Using `blasthole assays` from relevant grade-blocks in the excavated region, a local geology model may be constructed to permit spatial variation and computation of the grade uncertainty. This makes it possible to assess the level of safety (or risk) associated with digging in different areas at the resolution of individual buckets. An extended example will be provided in Sec.5.

In terms of contribution, the first objective of this work is to investigate the benefits of estimating the dig locations by analysing spatial interactions between the loaders/excavators and trucks—to establish whether the recall rate and accuracy may be improved. The second objective considers the effects of neglecting the articulated motion of the excavator arm. Evidence is presented to justify the simplifications made in estimating the dig positions. The third and final objective considers how material movement and orebody chemistry can be combined to monitor compositional changes and identify risks associated with grade uncertainty. This is accomplished using visualisation, local grade variance estimation and Monte Carlo simulation.

3 Estimating dig positions based on loader–truck interactions

Procedure 1 Estimating bucket dig positions

Require: Access to tables *lh_cycles* and *equipment_positions* in database

Definitions: Loading vehicle (henceforth referred as "loader") represents either an excavator or wheel-loader.
For a given load-haul cycle k, the loading timestamps L_k comprises $[t_k^{\text{arrival}}, t_k^{\text{load-start}}, t_k^{\text{full}}]$.
Vehicle orientation at time t is denoted $\boldsymbol{\theta}_t = [\cos\theta_t, \sin\theta_t]^T$.

Input: Processing time interval $\mathcal{T} \equiv [T_{\text{start}}, T_{\text{finish}}]$.
GPS x-y coordinates for loaders ($i \in \mathcal{I}$) and trucks ($j \in \mathcal{J}$), viz. $\{\mathbf{p}_{i,t}^{\text{loader}}\}_{t\in\mathcal{T}}$ and $\{\mathbf{p}_{j,t}^{\text{truck}}\}_{t\in\mathcal{T}}$
Load-haul cycle records: $[i_k, j_k, L_k]$ denotes [loaderID, truckID, loading times].
Location constraint: indicate whether to look for digging events *inside* or *outside* stockpiles.
Bucket range r, the distance from the reported GPS vehicle coordinates to the bucket.
Tolerance for spatial proximity, ϵ. Default: 0m.

1: **for** each load-haul cycle k inside processing interval $\mathcal{T}$ **do**
2: **[Identification step]** Identify the interacting vehicles: loader i_k and truck j_k.
3: **[Matching step]** Find the loader's trajectory $\mathcal{P} \equiv \{\mathbf{p}_{i_k,t}^{\text{loader}}\}_{t\in[\min(T_k),\max(T_k)+\Delta)}$ during which the truck remains stationary.
4: **if** $\mathcal{P}$ is not null and satisfies the location constraint **then**
5: **[Filtering step]** Remove noise (spikes and burst errors) from the loader's GPS trace [see ⑬, ⑭ in Fig. 5]
6: **[Segmentation step]** Find the set of close-contact points, $\mathcal{D} \equiv \{\mathbf{p}_{i_k,\tau}^{\text{loader}}\}_\tau$, where material is dumped by the loader into the truck at time τ [see ⑦ and ⑯ in Fig. 5]. Redundant points are discarded.
[Validation step] This requires the bucket position $\mathbf{b}_{i_k,\tau}$ to be inside the truck or at least located within a distance of ϵ from the truck's bounding box [refer to crosses in Fig. 5].
The bucket dump position is computed as $\mathbf{b}_{i_k,\tau} = \mathbf{p}_{i_k,\tau}^{\text{loader}} + r\boldsymbol{\theta}_{i_k,\tau}^{\text{loader}}$
7: **for** each truck loading event (n) $\star$ **do**
8: **[Detection step]** Find the point $\mathbf{p}_{i_k,\tau*}^{\text{loader}}$ in the loader GPS sequence, say from $[\mathbf{p}_{i_k,\tau_n}^{\text{loader}}, \mathbf{p}_{i_k,\tau_n+1}^{\text{loader}}, \mathbf{p}_{i_k,\tau_n+2}^{\text{loader}}, ..., \mathbf{p}_{i_k,\tau_{n+1}}^{\text{loader}}]$, that is furthest away from the truck [see ④ and ⑪ in Fig. 5]
9: **[Estimation step]** Predict the bucket dig position: compute $\hat{\mathbf{d}}_{i_k,n} = \mathbf{p}_{i_k,\tau*}^{\text{loader}} + r\boldsymbol{\theta}_{i_k,\tau*}^{\text{loader}}$.
10: Reject the solution if $\hat{\mathbf{d}}_{i_k,n}$ infringes the space occupied by the truck. Otherwise, let $t_{i_k,n} = \tau^*$.
11: **end for**
12: **end if**
13: **end for**

Output: Collection of $[k, i_k, j_k, t_{i_k,n}, \hat{\mathbf{d}}_{i_k,n}]$

$\star$ Individual loading events are demarcated by successive close-contact points in $\mathcal{D}$, see ⓪–⑦ and ⑧–⑯ in Fig. 5

The proposed method for loader dig position estimation utilises information from the tables *lh_cycles* and *equipment_positions* shown in Fig. 4. By definition, a loading vehicle refers to either a wheel-loader or excavator; this algorithm applies equally to both. The procedure described on page 6 consists of seven steps. In the identification step, attention is turned to individual load-haul cycles (each assigned a unique `lh_cycle_id`) and GPS measurements for the relevant truck and loading vehicle. In the matching step, the timestamps in the loader's GPS sequence are checked to identify the portion of its trajectory that overlaps with the loading interval specified in the load-haul cycle. This also includes a nominal digging period before the loader starts loading the truck. In the filtering step, anomalous measurements (such as spikes and burst errors) are removed from the GPS trace before further analysis. The segmentation step finds all the "close-contact" points observed in the loader's trajectory where the loading vehicle repeatedly approaches the haul truck. In Fig. 5, the circled numbers represent the loader's GPS coordinates which are sampled on average every 5 seconds. The close-contact points correspond to samples 7 and 16. The validation step performs

a feasibility test—it checks that the estimated dump location is within the bounding box of the truck to ensure the loader can actually transfer material into the truck. The segmentation step splits the trajectory into multiple load-dump segments where each interval (see ⓪-⑦ or ⑧-⑯ in Fig. 5) contains a single loading event (bucket transfer). To infer the dig location, the detection step first finds the points on the loader's trajectory that are furthest away from the truck in each loading segment. These positions are denoted $\mathbf{p}_t$ (see samples 4 and 11 in Fig. 5). Then, the estimation step computes the bucket dig position as $\hat{\mathbf{d}}_t = \mathbf{p}_t + r\boldsymbol{\theta}_t$ where r denotes the bucket range (distance from the reported loader position to the bucket) and the orientation vector $\boldsymbol{\theta}_t = [\cos\theta_t, \sin\theta_t]^T$ specifies the heading of the loading vehicle.

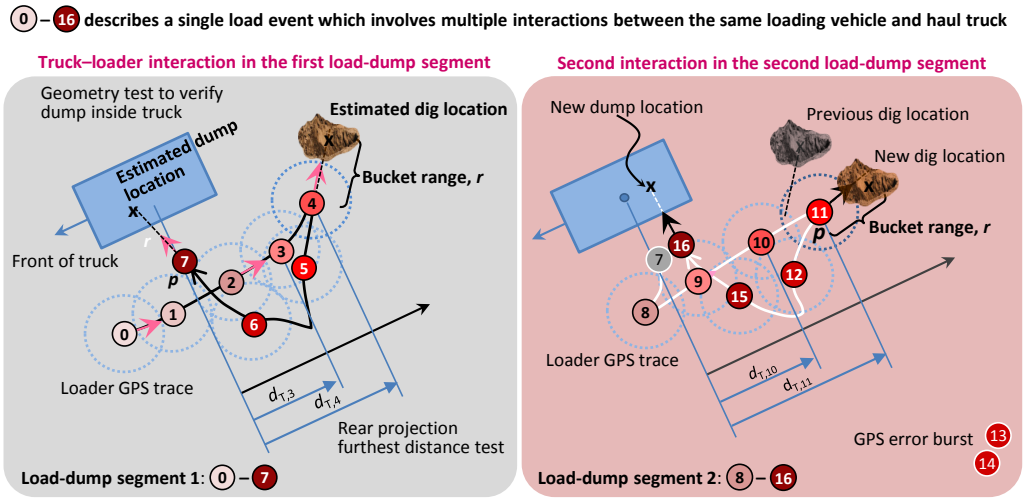

Figure 5: Estimating dig position based on loader-truck interaction (explained in main text)

3.1 Experiments

The above procedure was applied to GPS data collected at a mine site to estimate dig bucket positions for multiple vehicles over a four-week period. The analysis deals separately with wheel-loaders that dig inside known boundaries of stockpiles, and vehicles (both wheel-loaders and excavators) that operate outside stockpiles. Table 1 provides a statistical summary of this data. The number of truck-proximity and geometry validated dig bucket position estimates is given by `bucket_dig_estimates`. This equals the number of `load_dump_segment` candidates minus the number of infringements (`dig_within_loader` and `dig_within_truck`) where the proposed dig location is found inside the truck's bounding box or the loader's path. The `load_dump_segments` count excludes `redundant_close_contact_points` which appear when multiple GPS measurements are transmitted while the loader is stationary as it transfers material into the truck. The next two rows emphasize the importance of noise filtering. They show the GPS data contains spikes and error bursts. Thus, it is essential to remove these outliers (spatial discontinuities) to avoid erroneous interpretation of the data. The last three rows indicate the number of load-events. Each load event pertains to a series of interactions between a specific loader and truck within a load-haul cycle; this usually lasts no more than a few minutes.

The purpose of these experiments is to assess the benefits of estimating dig locations based on loader-truck interactions. The following analysis examines digging patterns, highlights data integrity issues with the existing *lh_buckets* estimates, and demonstrates performance gain in terms of precision and recall.

3.2 Dig patterns

Figure 6 shows the dig pattern of wheel-loaders and excavators at several locations. Each dot represents a dig position estimated by the algorithm. These pictures reveal something intuitive. Colouring the dig positions by their timestamp shows the excavation progresses from front to back and follows a choreographed sequence rather than a random path.

3.3 Dig position estimates: comparison with lh_buckets

Figure 7 provides a spatial comparison of the dig estimates with those in the *lh_buckets* table. Their differences, $\mathbf{d}_t^{\text{proposed}} - \mathbf{d}_t^{\text{lh\_buckets}}$, are displayed as a vector field in quiver plots. These differences (magnitude and orientation) can be seen more easily using polar plots and kernel density estimation. The graphs show general agreement. However,

	Week 1-2 (year 1)		Week 3-4 (year 1)	
Trips _____ stockpiles	inside	outside	inside	outside
load_dump_segments (number of candidates)	10285	31094	17923	44606
- dig_within_loader_violation	646	758	1005	890
- dig_within_truck_violation	4423	7910	7592	11624
bucket_dig_estimates	**5216**	**22426**	**9326**	**32092**
redundant_close_contact_points	1355	9003	2429	12874
gps_error_bursts	93	269	164	409
gps_spikes	320	915	563	1188
load_events (specifically involving excavators)	0	6884	0	7544
load_events (specifically involving wheel-loaders)	2292	2221	3167	2082
load_events (total)	2292	9176	3168	9640

Table 1: Data profile: load events and dig bucket statistics for trips inside and outside stockpiles

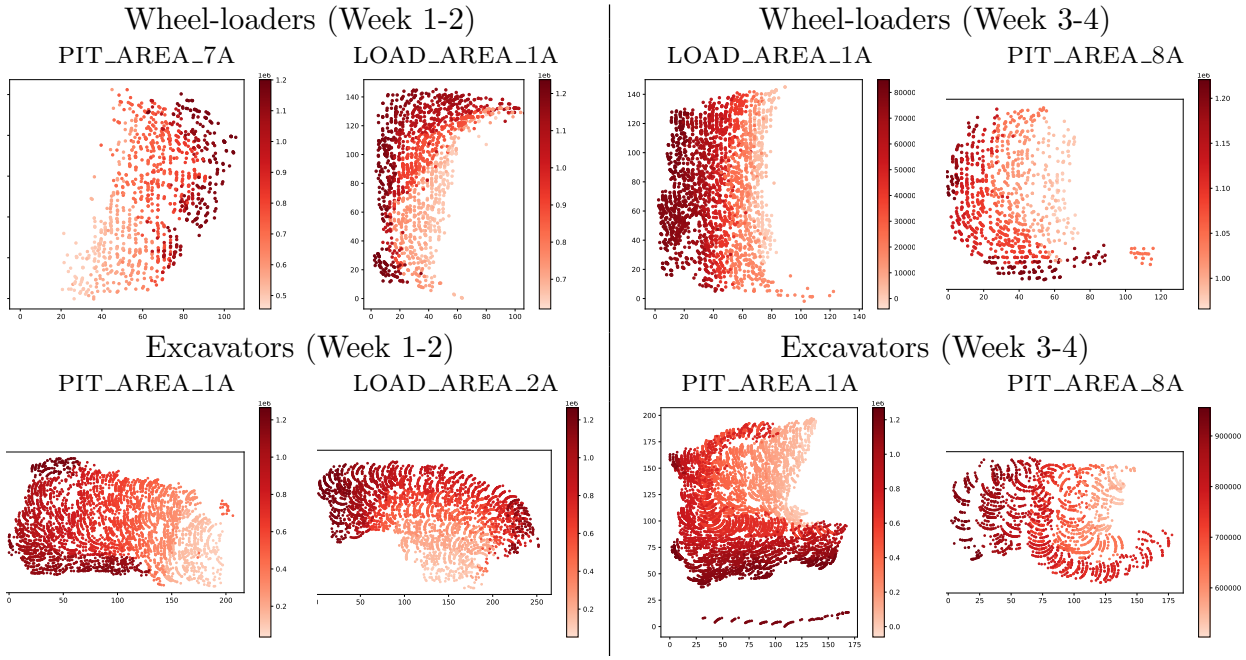

Figure 6: Dig patterns coloured by bucket timestamp at several load locations

the dig positions obtained using the proposed method tend to be more tightly clustered. The observed outliers are found almost exclusively in the *lh_buckets* table, as they coincide with the tail of large arrows in the vector field.

Figure 8 highlights some inconsistencies in the existing *lh_buckets* estimates. The histograms show the distance between a loader's dig and dump locations. Some of the wheel-loader bucket positions provided by *lh_buckets* are clearly problematic as the dig-dump separation is close to zero which is unrealistic.

3.4 Validation procedure: precision and lh_cycles recall rate

A key indicator of performance is the *lh_cycles* recall rate. Knowing the dig position of at least one bucket in a given load-haul cycle allows the material in the truck to be characterised. The main questions relate to the number of load-dump intervals that contain only dig estimates from the *lh_buckets* table, versus dig estimates from the *proposed* method, and how many contain at least one from both? To facilitate analysis, let B and P denote dig estimates from the *lh_buckets* table and those obtained with the *proposed* method, respectively. Further, let C denote load-dump events

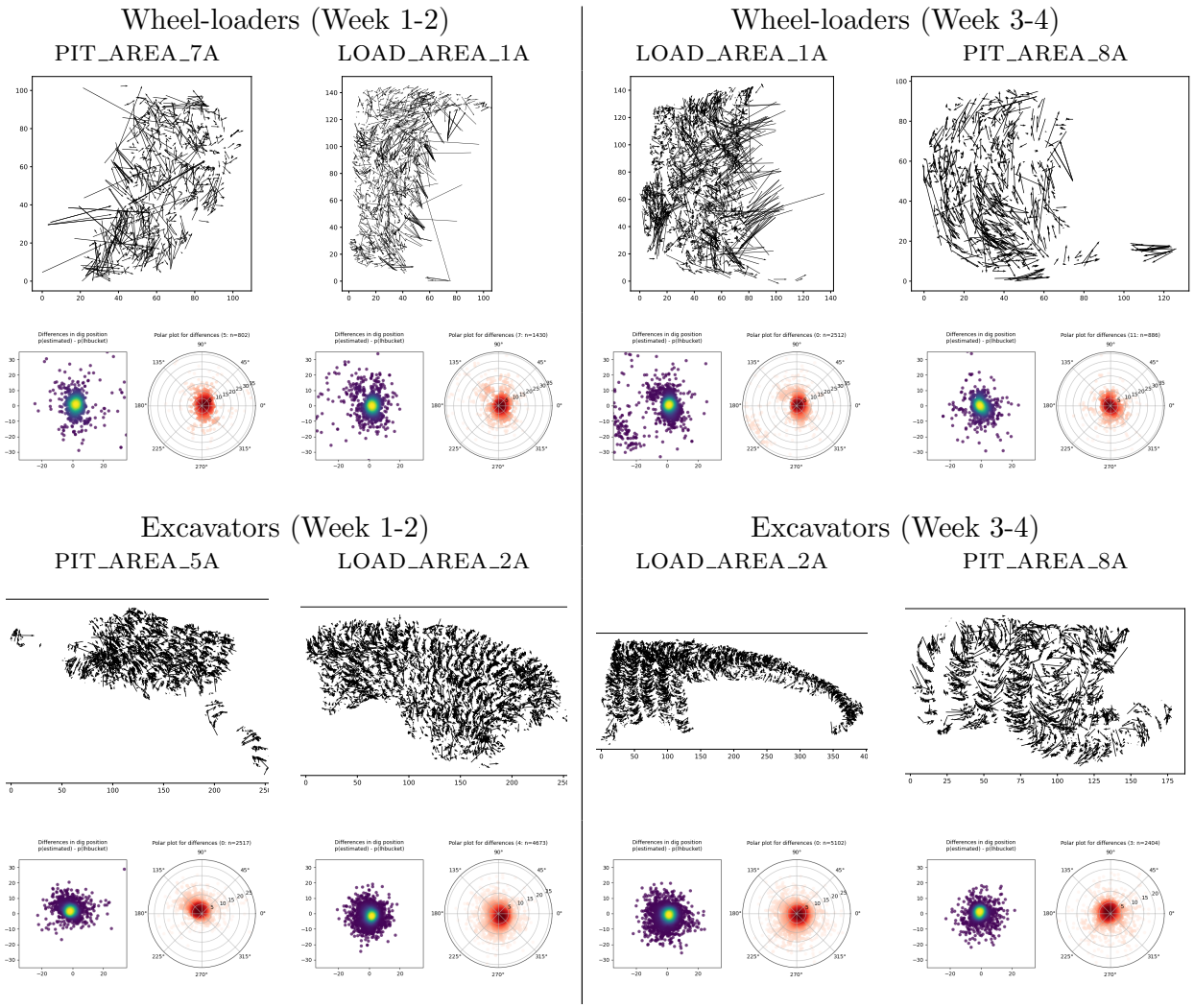

Figure 7: Differences between the proposed and *lh_buckets* dig position estimates

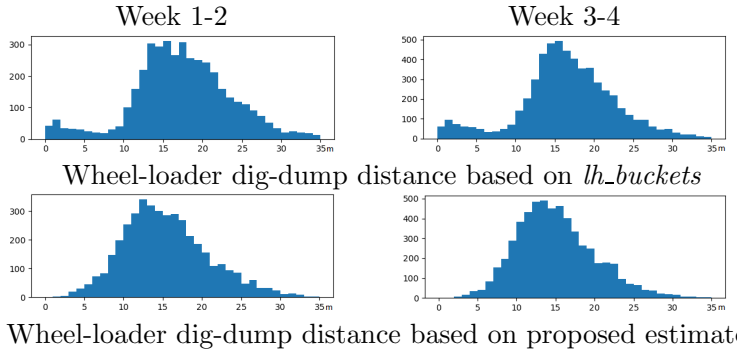

Figure 8: Dig-to-dump distance distribution reveals data integrity issues in *lh_buckets* estimates

reported in the *lh_cycles*. At its core, the experiments involve counting if one or more estimates—from *lh_buckets* (B) and/or the *proposed* method (P)—is contained in each time interval [`load_time_start`, `load_time_full`]. The

digger's load-dump interval may be written as $\mathcal{I}_k \equiv [t_k^{(\text{start})} - \Delta^{\text{-}}, t_k^{(\text{finish})}] \in C$ where $\Delta^{\text{-}}$ represents a small tolerance for a possible delay in the onset of `load_time_start` as triggered by the strain measurement system [22].[2]

Using set notations, the number of *lh_cycles* $\mathcal{I}_k$ that contain dig estimates from (i) *lh_buckets*, (ii) the *proposed* method and (iii) both are denoted by $|C \cap B|$, $|C \cap P|$ and $|C \cap (B \cap P)|$, respectively. Similarly, $|C \cap \bar{B} \cap P|$ refers to *lh_cycles* that contain dig(s) from the *proposed* estimates but not *lh_buckets*, and $|C \cap B \cap \bar{P}|$ represents the opposite. Of particular interest are *lh_cycles*

- $|C \cap (B \cup P)|$ that contain dig(s) from **either** *lh_buckets* **or** the *proposed* method;
- $|C \cap \bar{B} \cap \bar{P}|$ that contain **neither** dig(s) from *lh_buckets* **nor** the *proposed* method.

These counters are used to compute the lh_cycle **recall rate** given the timing of the *lh_buckets* and digs inferred by the *proposed* method. Spatial **precision** represents the second aspect that is important for performance evaluation. Since the actual dig positions are unknown, a backward search interval, $[t_k^{(\text{finish})} - \Delta, t_k^{(\text{finish})}]$ with $\Delta = 30$s, is used to find the equipment position and orientation of the truck closest to `load_time_full`. The signed distance of the dump position from the truck bounding box at this very moment is computed for both sets B and P. This analysis is approximate since it does not consider all vehicle poses and GPS coordinates during the entire loading interval, nonetheless it is adequate for our purpose. Although a negative signed distance d strictly indicates the dumping point is inside the truck, we have relaxed this requirement to $d \leq 3$m to make up for measurement noise / localisation error. Precision is defined as (number of dumps inside truck)/(number of dumps reported). The $\mathbf{F_1}$ **score** computes the harmonic mean of the precision and recall rate[3] by giving equal weight of importance to both. The F_1 score provides an overall measure of effectiveness of the solution in terms of **spatial precision** and **temporal coverage**.

3.5 Interpretations

Table 2 presents validation results against *lh_cycles* for wheel-loaders digging inside and outside the boundary of known stockpiles. Comparing $|C \cap B|$ with $|C \cap P|$, *lh_buckets* recall more load intervals. However, it has lower precision and efficiency. The precision for *lh_buckets* is 81.8%–85.4% while the precision for the *proposed* method is $\gtrsim 99\%$. In terms of lh_cycles per bucket, the ratio is higher for the *proposed* method than *lh_buckets*. For instance, $|C \cap P|/|P| = 1822/6266 \approx 29.1\%$ whereas $|C \cap B|/|B| = 1985/10323 \approx 19.2\%$ based on the Week 3-4 figures.

The most significant benefit however is that combining the *lh_buckets* and *proposed* dig estimates provide a 97.3%–97.9% coverage of the load-haul cycles. Under the status quo, using only data from *lh_buckets* recall only 88.6%–89.5% of *lh_cycles* for wheel-loader trips inside stockpiles, and 91.4%–91.8% of *lh_cycles* for trips outside stockpiles. These represent gains of 7.8%–8.9% and 6.0% respectively. The proposed method (analysis of loader-truck interactions) provides additional dig locations that were missed. Based on earlier observations from Figs. 7 and 8, for the 73.0%–78.3% of digs that the *lh_buckets* and our estimates have in common, the *proposed* method will likely provide more consistent and higher quality dig estimates.

Validation results for excavators are similarly shown in Table 3. Although the statistics reported for the corresponding period (from the year 1 column) appear different, they have nonetheless been verified to be correct. The fact that data for 2 of 5 excavators were missing from the *lh_buckets* table accounts for such differences, viz. a reduction in the *lh_buckets* and combined recall rates, $|C \cap B|$ and $|C \cap (B \cup P)|$. Repeated experiments using more recent data (see year 2 column) produce results broadly consistent with the baseline for wheel-loaders. A gain of 5.7%–7.9% is observed which increases the combined recall rate to 98.0%–98.4%.

To reinforce the meaning of load interval recall, Fig. 9 shows the temporal correspondence of *lh_buckets* and the proposed dig estimates with *lh_cycles* in an arbitrary 8-hour window. In this tripartite graph, each panel represents a vehicle (either a wheel-loader or an excavator) and the time axis (from left to right) is wrapped around every 2 hours. B and P are represented by orange and blue nodes, with edges connected to the set C (horizontal bars) when there is a match. Specifically, information gain is conveyed by light green intervals (circled in red); these dig estimates are only available courtesy of the proposed method.

This section has exploited the first link shown in Fig. 4 where equipment positions are related to load-haul cycles via vehicle IDs and timestamps. It shows that proactive analysis of loader-truck interactions can produce consistent bucket position estimates, recover missing dig buckets and conditionally improve the load-haul cycle recall rate.

[2] This turns out to have negligible effect on the analysis. However, one should ensure this adjustment does not infringe any load-haul cycle that immediately precedes the current interval. So, technically, this should perhaps be written as $\mathcal{I}_k \equiv [t_k^{(\text{start})} - \min\{\Delta^{\text{-}}, \max\{t_k^{(\text{start})} - t_{k-1}^{(\text{finish})}, 0\}\}, t_k^{(\text{finish})}]$.

[3] $F_1 = (\text{precision} \times \text{recall})/[0.5 \times (\text{precision} + \text{recall})]$

	Week 1-2 (year 1)		Week 3-4 (year 1)	
Loading vehicle	wheel-loader		wheel-loader	
Trips ___ stockpiles	inside	outside	inside	outside
effective number of lh_buckets $\|B\|$	10097	9762	13401	10323
effective number of proposed buckets $\|P\|$	4677	4593	8349	6266
number of lh_cycles $\|C\|$	2600	2293	3462	2160
recall(lh_buckets) $\|C \cap B\|$	2327 (**89.5%**)	2097 (**91.4%**)	3068 (**88.6%**)	1985 (**91.8%**)
recall(proposed) $\|C \cap P\|$	2103 (80.9%)	1842 (80.3%)	2963 (85.6%)	1822 (84.3%)
recall(both) $\|C \cap B \cap P\|$	1899 (73.0%)	1707 (74.4%)	2653 (76.6%)	1692 (78.3%)
recall(either) $\|C \cap (B \cup P)\|$	2531 (**97.3%**)	2232 (**97.3%**)	3378 (**97.6%**)	2115 (**97.9%**)
recall(lh_buckets, not(proposed)) $\|C \cap B \cap \bar{P}\|$	428 (16.4%)	390 (17.0%)	415 (12.0%)	293 (13.5%)
recall(proposed, not(lh_buckets)) $\|C \cap \bar{B} \cap P\|$	204 (7.8%)	135 (5.9%)	310 (8.9%)	130 (6.0%)
precision(lh_buckets)	8263 (81.8%)	8284 (84.8%)	11449 (85.4%)	8634 (83.6%)
precision(proposed)	4669 (99.82%)	4544 (98.9%)	8348 (99.98%)	6231 (99.4%)
F_1(lh_buckets)	85.5%	88.0%	87.0%	87.5%
F_1(proposed)	**89.4%**	**88.7%**	**92.2%**	**91.3%**

Table 2: Performance of lh_buckets and proposed dig estimator for wheel-loaders ($\Delta^- = 0$s)

	Week 1-2		Week 3-4	
Year	1	2	1	2
Loading vehicle	excavator		excavator	
Trips ___ stockpiles	outside		outside	
effective number of lh_buckets $\|B\|$	15919	29287	18607	39959
effective number of proposed buckets $\|P\|$	11300	12080	18507	15306
number of lh_cycles $\|C\|$	6779	8671	7642	10146
recall(lh_buckets) $\|C \cap B\|$	3615 (**53.3%**)†	7812 (**90.1%**)	4064 (**53.1%**)†	9401 (**92.7%**)
recall(proposed) $\|C \cap P\|$	4905 (72.3%)	6869 (79.2%)	5972 (78.1%)	8155 (80.4%)
recall(both) $\|C \cap B \cap P\|$	2539 (37.4%)	6180 (71.3%)	3073 (40.2%)	7572 (74.6%)
recall(either) $\|C \cap (B \cup P)\|$	5981 (**88.2%**)	8501 (**98.0%**)	6963 (**91.1%**)	9984 (**98.4%**)
recall(lh_buckets, not(proposed)) $\|C \cap B \cap \bar{P}\|$	1076 (15.8%)	1632 (18.8%)	991 (12.9%)	1829 (18.0%)
recall(proposed, not(lh_buckets)) $\|C \cap \bar{B} \cap P\|$	2366 (34.9%)	689 (7.9%)	2899 (37.9%)	583 (5.7%)
precision(lh_buckets)	14787 (92.8%)	18943 (64.7%)	16894 (90.7%)	28579 (71.5%)
precision(proposed)	11191 (99.0%)	19348 (98.9%)	18367 (99.2%)	24221 (98.9%)
F_1(lh_buckets)	67.7%	75.3%	67.0%	80.7%
F_1(proposed)	**83.6%**	**88.0%**	**87.4%**	**88.7%**

† Significant reduction relative to Table 2 mainly due to missing data in the *lh_buckets* table

Table 3: Performance of lh_buckets and proposed dig estimator for excavators ($\Delta^- = 0$s)

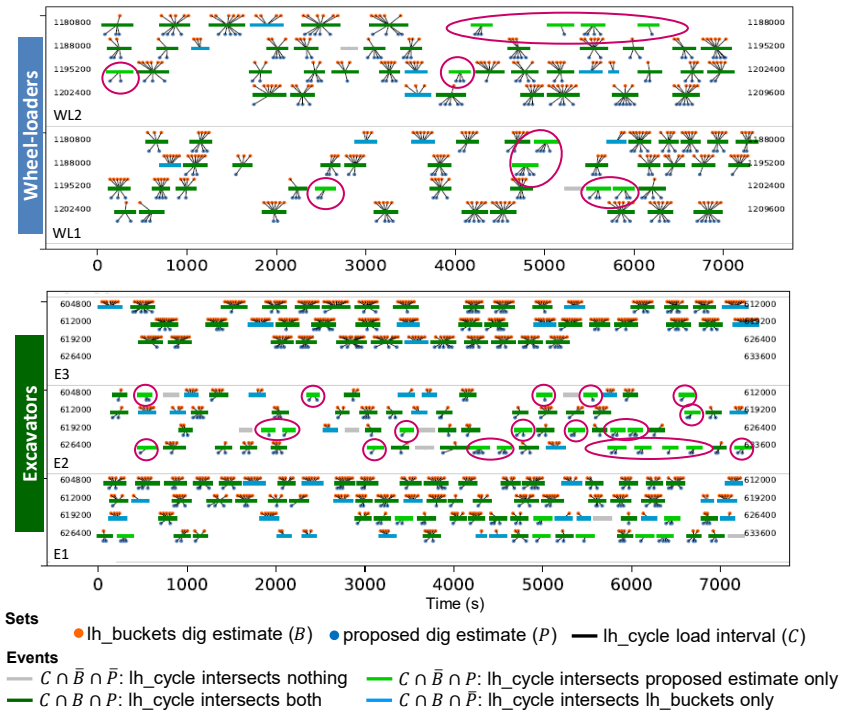

Figure 9: Temporal association of *lh_buckets* and the proposed dig estimates with *lh_cycles*

4 Effects of articulated motion in the excavator's arm

The experiments thus far have assumed the bucket range (see r in Fig. 5) is fixed. The aim here is to quantify any performance penalty associated with using a constant bucket range in $\mathbf{d}_t = \mathbf{p}_t + r\boldsymbol{\theta}_t$ to estimate the bucket position of excavators. The null hypothesis is that using a rigid 2D motion model is sufficient for estimating the bucket position of excavators. However, opinion is divided as to the actual utility of incorporating joint angles in the calculation. Looking at Fig. 10, there is clearly articulated motion involved in the arm movement of excavators. The central question is whether neglecting the hydraulic cylinders which control the joint angles for the boom, stick and bucket results in a reasonable approximation? In practice, how much performance do we sacrifice by adopting a simple model, does it matter much? The ensuing experiments will provide clarity on this question and measure any increase in false negatives (FN) in terms of lh-cycle recall when the bucket range is fixed.

Let's begin with a thought experiment and consider what happens if (i) the bucket range parameter, r, is allowed to vary with each bucket and (ii) the search for the bucket position is conducted in a narrow range, $r_t \in [12-\epsilon, 12+\epsilon]$, where $r_t = \|\mathbf{p}_t - \mathbf{d}_t\|$ for each load-dump interaction.[4] One may discover that on some occasions, dump events are misdetected by fixing $r = 12$ because the actual distance is shorter (overshooting occurs) and at other times, the actual distance is longer than $r = 12$. Both cases may result in the dismissal of a proposed location as the dump lies outside the truck. Expanding this aperture, as ϵ gets reasonably large, the number of misdetected cases should converge given the physical constraints on the excavator's arm. By increasing ϵ (from 0 to 4m, say, in steps of 0.5m) and allowing the bracket $[12-\epsilon, 12+\epsilon]$ to expand accordingly, it is possible to estimate the loss that results from neglecting variations in the bucket range. In this scenario, the bucket position estimation problem is posed in 2D, so the equivalent distance in the x-y plane is of interest and the joint angles (e.g., the pitch of the boom and stick) are neglected.

This allows the same code (existing implementation) to be used which already supports statistics gathering. The script is only slightly modified to permit override of the default bucket range parameter, r. Value substitution is done uniformly for all excavator models, in each instance the default (12m) is replaced with the value configured for each experiment, $8 + 0.5\lambda$, where $0 \leq \lambda \in \mathbb{Z} \leq 16$. The analysis again focuses on the *lh_cycles* recall rate. Table 4 reports the number of distinct `lh_cycle_id` discovered for the excavators as a function of the bracket size, ϵ. Recall that each bracket encompasses a set of r values in the interval $[12-\epsilon, 12+\epsilon]$.

[4]The optimal bucket range (best constant value) was determined to be 12m following a parameter tuning procedure which maximises the `bucket_dump_interval_count` and `bucket_dig_proper_count`.

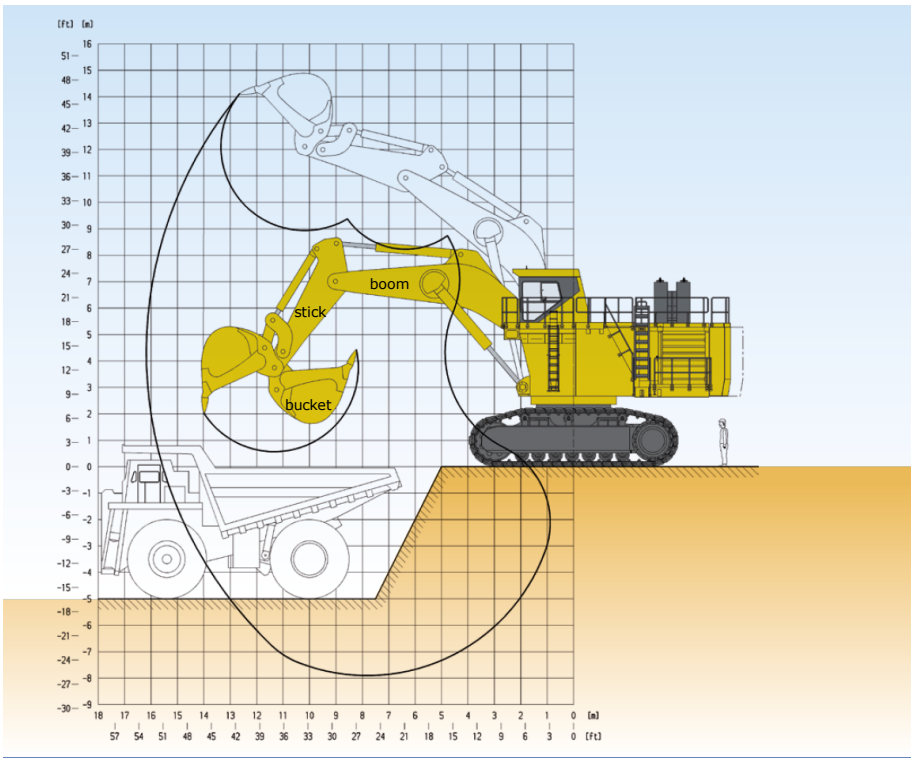

Figure 10: A Komatsu PC3000-6 excavator has boom and stick lengths of 6.8m and 4m and a bucket capacity of approximately 16.5m^3. The average bucket range is about 12m.

	Bucket range bracket size, ϵ								
Excavator	**0**	0.5	1	1.5	2	2.5	3	3.5	4
E1	535	536	538	539	539	539	541	541	541
E3	424	425	427	430	430	430	431	432	432
E4	926	927	930	931	932	932	934	934	935
E5	989	1002	1007	1014	1017	1017	1018	1019	1019
E6	190	191	191	191	191	191	191	191	191
combined	3064	3081	3093	3105	3109	3109	3115	3117	3118

Table 4: Week 1 statistics: number of distinct lh_cycles discovered (n_{LHC}) vs bucket range bracket size (ϵ)

4.1 Results

The results are easier to see when counts are converted to recall rates as

$$R_{\text{LHC}}(\epsilon) = \frac{n_{\text{LHC}}(0)}{n_{\text{LHC}}(\epsilon)} \tag{1}$$

For any positive ϵ, we are comparing the number of distinct lh_cycles discovered under two conditions: the numerator corresponds to using a default bucket range of $r = 12$m (with $\epsilon = 0$) and the denominator mirrors the experience of using a finite set of bucket range values between $12 - \epsilon$ and $12 + \epsilon$ with 0.5m spacing.

Table 5 shows the recall rate converges as ϵ grows. The minimum recall (0.9706) is observed for excavator E5 and the effective recall rate for all excavators combined is 0.9827 during Week 1.

Instead of recall, the False Negative Rate (FNR) provides an equivalent measure in terms of the loss (or shortfall in matching lh_cycles) which converges to zero as the bracket increases beyond a certain point. FNR statistics are presented in Table 6. Finally, the combined results over a four-week period are reported in Table 7. To interpret these results, recall ratios $R_{\text{LHC}}(\epsilon)$ and false negative rates $\text{FNR}_{\text{LHC}}(\epsilon)$ are plotted in Figs. 11 and 12.

	Bucket range bracket size, ϵ								
Excavator	0	0.5	1	1.5	2	2.5	3	3.5	4
E1	1	0.9981	0.9944	0.9926	0.9926	0.9926	**0.9889**	0.9889	0.9889
E3	1	0.9976	0.9930	0.9860	0.9860	0.9860	0.9838	**0.9815**	0.9815
E4	1	0.9989	0.9957	0.9946	0.9946	0.9936	0.9914	0.9914	**0.9904**
E5	1	0.9870	0.9821	0.9753	0.9753	0.9725	0.9715	**0.9706**	0.9706
E6	1	**0.9948**	0.9948	0.9948	0.9948	0.9948	0.9948	0.9948	0.9948
combined	1	0.9945	0.9906	0.9868	0.9868	0.9855	0.9836	0.9830	**0.9827**

Table 5: Week 1 lh_cycle recall rate, $R_{\text{LHC}}(\epsilon)$, as a function of bucket range bracket size, ϵ.

		Bucket range bracket size, ϵ								
Excavator	$n_{\text{LHC}}(\infty)$	0	0.5	1	1.5	2	2.5	3	3.5	4
E1	541	1.11%	0.92%	0.55%	0.37%	0.37%	0.37%	**0.00%**	0.00%	0.00%
E3	432	1.85%	1.62%	1.16%	0.46%	0.46%	0.46%	0.23%	**0.00%**	0.00%
E4	935	0.96%	0.86%	0.53%	0.43%	0.32%	0.32%	0.11%	0.11%	**0.00%**
E5	1019	2.94%	1.67%	1.18%	0.49%	0.20%	0.20%	0.10%	**0.00%**	0.00%
E6	191	0.52%	**0.00%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
combined	3118	1.73%	1.19%	0.80%	0.42%	0.29%	0.29%	0.10%	0.03%	**0.00%**

Table 6: Week 1 lh_cycle false negative rate, $\text{FNR}_{\text{LHC}}(\epsilon)$, as a function of bucket range bracket size, ϵ.

		Bucket range bracket size, ϵ								
Excavators	Statistic	0	0.5	1	1.5	2	2.5	3	3.5	4
combined	FNR	1.85%	1.22%	0.89%	0.60%	0.43%	0.32%	0.18%	0.05%	0.00%
combined	conditional gain†	0.61%	0.45%	0.35%	0.26%	0.19%	0.13%	0.06%	0.02%	**0.00%**

† The conditional gain refers to $[n_{\text{LHC}}(\epsilon) - n_{\text{LHC}}(0)]/n_{\text{LHC}}(0) \mid C \cap \bar{B} \cap P$, the percentage improvement in `lh_cycles` recall given these cycles are missing in *lh_buckets*.

Table 7: Four week aggregates: combined statistics for all excavators as a function of bucket range bracket size, ϵ.

From the graphs, the small amplitude of the peak FNR suggests there is a sweet spot and operators tend to concentrate on digging at an equivalent distance of around 12m with at most 1.5m variation covering about 99% of cases. There is not much variation in terms of reach between successive buckets. The maximum performance penalty in the lh_cycle recall rate is about 3% for one particular excavator (E5), but the average maximum is around 1.8% with r fixed at 12m.[5] The conditional information gain (minor improvement in *lh_cycles* recall from increasing ϵ, given they are missing from *lh_buckets*) is about 0.6%. On this basis, it is reasonable to conclude the joint angles and arm movement of excavators are secondary factors that can largely be ignored for the purpose of estimating 2D bucket positions.

5 Application: Linking bucket position estimates with ore properties to inform grade risk

This section considers some of the steps involved in tracking volumetric and compositional changes associated with load-dump events. The objective is to convey concepts fundamental to material tracking using various forms of visualisation, setting out clearly what can be achieved by integrating bucket position estimates with orebody knowledge. A specific goal is to illustrate how grade risks are transferred by linking dig buckets to geochemistry and combining with truck dump position estimates.

The Sankey diagram in Fig. 13 describes material movement between load and dump locations as reported in the *lh_cycles* table. It registers the volumetric flow from load locations (such as a pit or blast area) to various dump locations (such as crusher, stockpile, or waste dump). This complex coupling between sources and destinations is already apparent even through a high-level abstraction of material transfer for a real mine. Behind each bucket, every dig position in the pit is contained in a mining unit called a *grade-block* which partitions the excavated area according

[5] This can be seen from the brown curve in Fig. 12 or from Table 7.

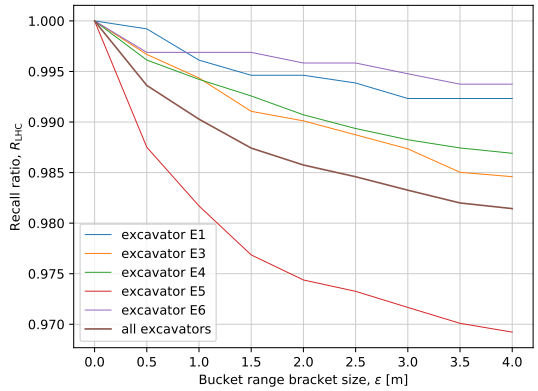

Figure 11: Convergence in aggregate recall ratios

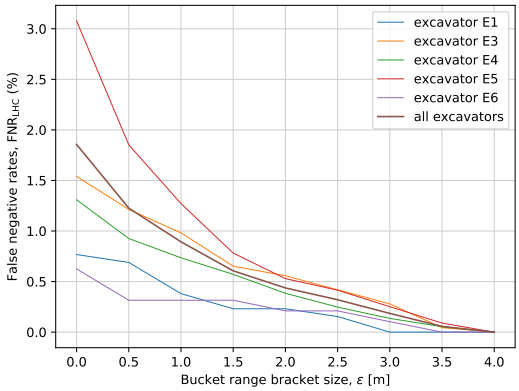

Figure 12: False negative rates

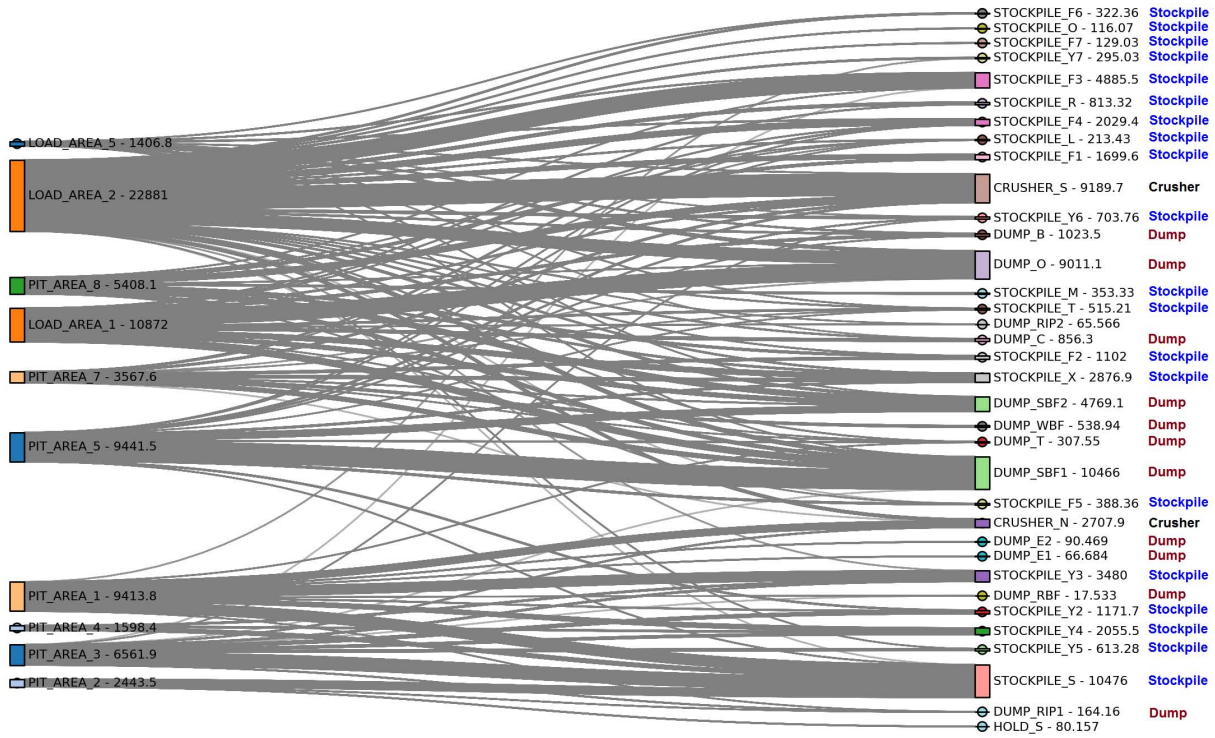

Figure 13: Sankey diagram: material movement between load and dump locations as reported in lh_cycles

to the expected grade (e.g., based on the concentration of Fe, SiO_2 and Al_2O_3). To untangle the source-recipient relationships, a subset of these connections (e.g., single-source to multiple-destination, or multiple-source to single-destination mappings) can be plotted using stacked histograms to describe both the volume and chemistry associated with these load-dump transactions. Examples are shown in Fig. 14. The left plot shows predominantly high-grade iron minerals being transferred from PIT_AREA_3 mainly to stockpiles. However, based on the information available, occasional mistakes were made. Low-grade materials (Fe $\leq 52\%$ shown in yellow) were inadvertently transferred to STOCKPILE_Y2/3 and STOCKPILE_S. The right plot shows mostly low-grade material at PIT_AREA_5 destined for waste dumps. Efficiency improvement was possible if operators had realised mid-to-high grade ore (Fe$\geq 58\%$ shown in red) were present and being sent to SBF1—a waste dump.

Fig. 15 shows the proportion and grade of material transferred from various sources to a single destination. The left plot shows the low-to-mid grade material (Fe $\leq 55\%$) sent to the crusher mainly derives from PIT_AREA_8. Feeding

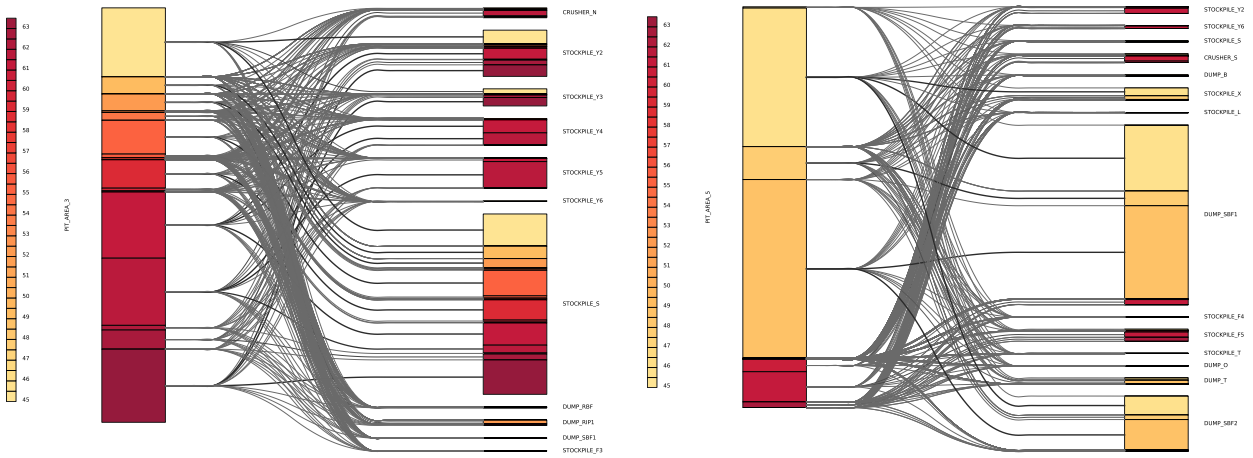

Figure 14: Composition and material transfer: stacked histograms for single-source multiple-destinations. Sources shown: (left) PIT_AREA_3, (right) PIT_AREA_5. Grades are estimated using grade-block averages.

low grade material to a crusher is quite likely unintentional. The right plot shows the dump location SBF1 receiving an unexpected payload (Fe $\geq 58\%$) from LOAD_AREA_2.

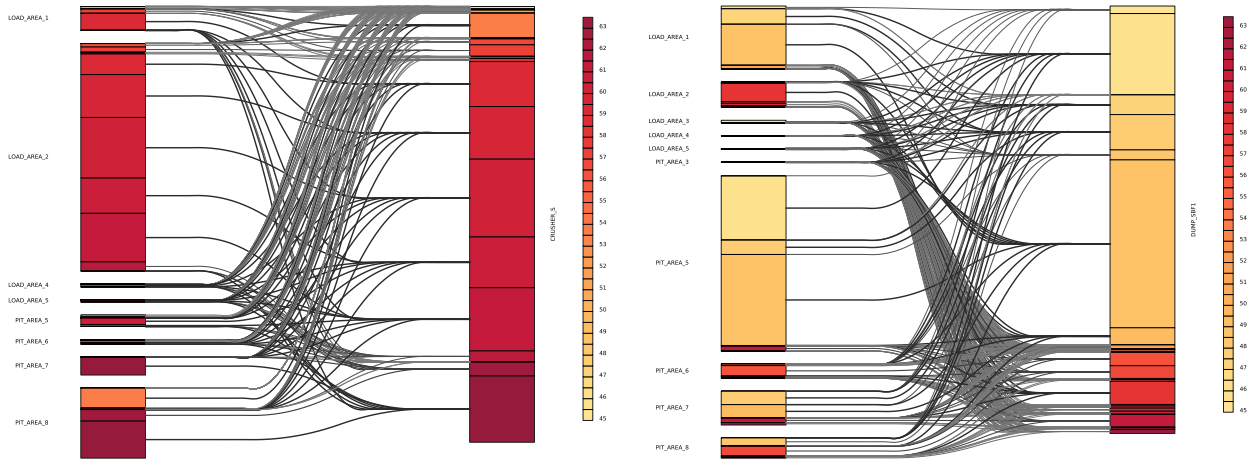

Figure 15: Composition and material transfer: stacked histograms for multiple-sources single-destination. Destinations shown: (left) S_CRUSHER, (right) DUMP_SBF1. Grades are estimated using grade-block averages.

The illustrations thus far have focused on load/dump locations without pinpointing the bucket coordinates. Likewise, geochemical properties (Fe concentration values used in the stacked histograms) are inherited from grade-block averages. This provides a pragmatic, piecewise-constant approximation rather than a spatially continuous geology model. The benefits of having the dig bucket estimates are the high degree of spatial localisation that it affords, viz. the ability to *resolve* differences and be quite precise about these source-target connections. If a fine-grained grade model is constructed from the blastholes within the grade-blocks, using ordinary kriging for instance, one can readily identify problematic areas and understand how grade risks are transferred. This approach fully exploits the second and third links in Fig. 4 which confer load-dump events with grade-block and assay properties. To reinforce this point, Figs. 16 and 17 show the grade-blocks and blasthole samples at two source locations: LOAD_AREA_2 and PIT_AREA_1. Blastholes are coloured by (top) Fe and (bottom) Al_2O_3 concentration.

Typically, excavation near the edges of a high-grade (white) and waste (grey) block presents challenges in terms of maintaining proper grade control. In transition regions, extra attention is needed to prevent waste (W) from mingling with high-grade (HG) material and being sent to the crusher or interim stockpile, the reverse situation which is less a risk but more an efficiency issue is the disposal of HG/LG ore at a waste dump.

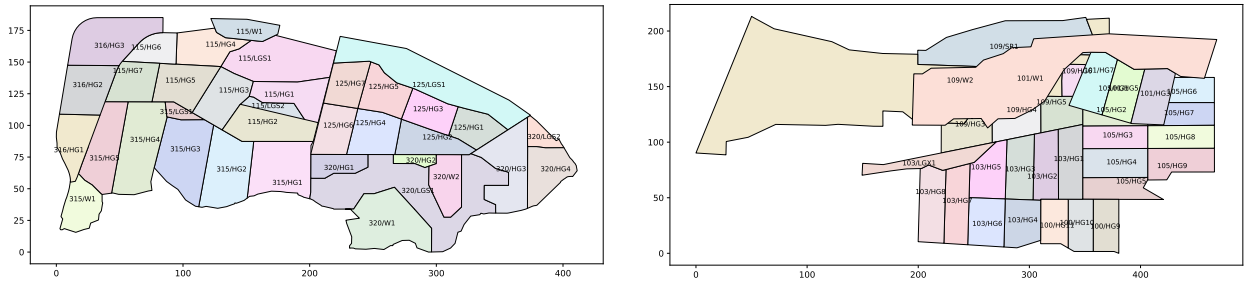

Figure 16: Grade-blocks at (left) LOAD_AREA_2, (right) PIT_AREA_1

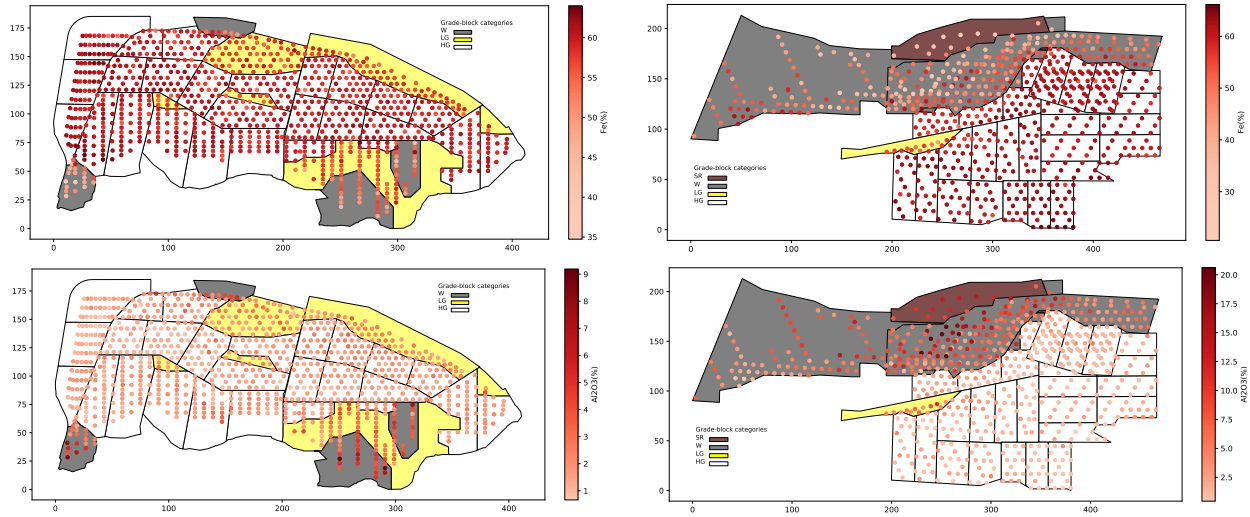

Figure 17: Blasthole samples within grade-blocks at (left) LOAD_AREA_2, (right) PIT_AREA_1

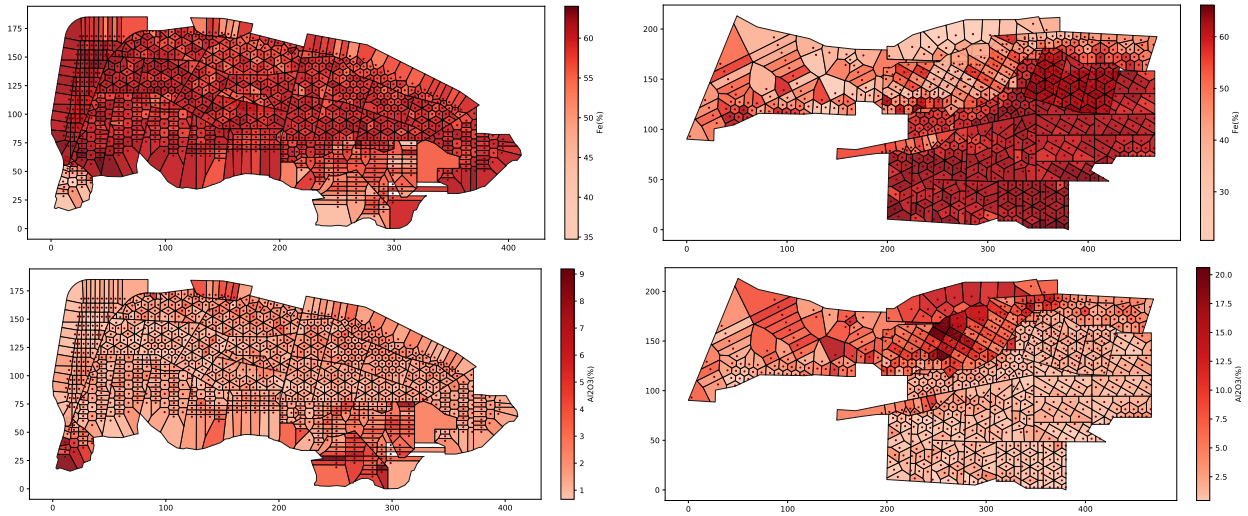

Figure 18: Blasthole-induced Voronoi tessellation of grade-blocks — (left) LOAD_AREA_2, (right) PIT_AREA_1

Fig. 18 uses Voronoi tessellation[6] to show geochemical variation at the grade-block level. Chemical concentration is by no means the only relevant attribute for mine operators, material type (e.g., percentages of goethite or haematite vs

[6]The Voronoi diagram may be produced using the Lloyd-Max algorithm. The cell centres coincide with the position of blastholes

shale) [23] and geometallurgical properties are often modelled as well. Rock hardness may be inferred using measurements obtained while drilling. The Voronoi approach provides an avenue for characterising the spatial distribution of blastholes. Two measures that reflect sampling fairness and geochemical consensus among the assay samples within a grade-block have been proposed in [24]. These measures can indicate the reliability of average grade estimates computed for selective mining units.

If the relevant properties in each grade-block are characterised by a Normal distribution, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the variance for a linear combination of correlated random variables (buckets or truck dumps) at a destination may be computed using

$$\text{var}\left(\sum_{i=1}^{n} w_i X_i\right) = \sum_{i=1}^{n} w_i^2 \text{var}(X_i) + 2\sum_{i=1}^{n}\sum_{j>i}^{n} w_i w_j \Sigma_{i,j} \tag{2}$$

where w_i describes the bucket payload (or dumps) associated with grade-block i and $\Sigma_{i,j} = \text{cov}(X_i, X_j)$. A key point about Fig. 18 is that it shows contrasting composition in different parts of the excavated area which reflects natural variation in the pit.

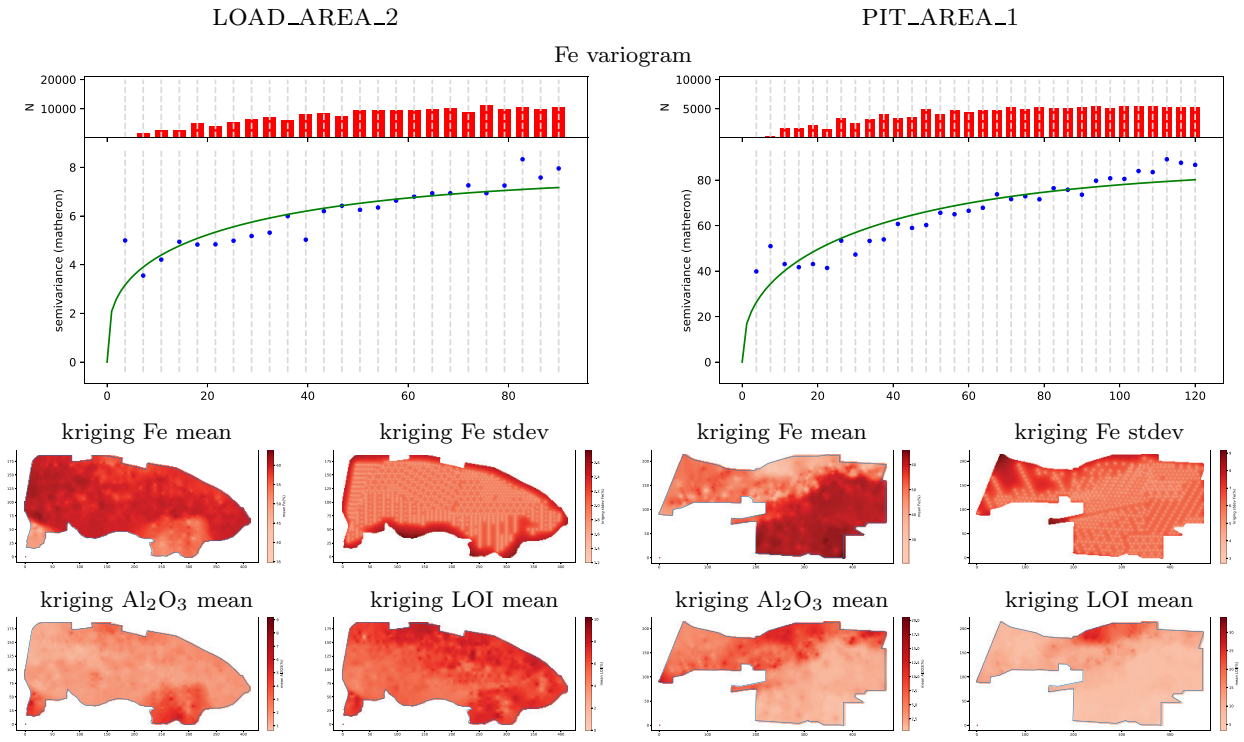

Figure 19: Variogram and kriging estimates for Fe, Al_2O_3 and LOI (loss on ignition) mean and Fe standard deviation — (left) LOAD_AREA_2, (right) PIT_AREA_1

To estimate the composition of the bucket at a dig location, a 3D grade model may be constructed using conditional simulation [25] or Gaussian Processes [10] for instance. Ideally, the grade prediction block model is faithful to the geological structure of the mine [26] and has the ability to rectify its structure when updated geological boundaries (e.g., stratigraphic or mineralisation surfaces) are given [27]. For demonstration purpose, ordinary kriging [28, 29] is used to interpolate the grade using blasthole samples in the excavated region. Fig. 19 shows the variograms computed for Fe using a Matérn kernel and the Mathéron (linear unbiased) estimator [30]. The fitted variogram expresses how spatial dependence in observations changes with separating distance. Kriging determines the optimal weights for interpolation from the spatial covariance structure of the data [31]. These models facilitate interpolation of the Fe grade at precisely the location where digging occurs. This inference procedure produces the bucket Fe grade estimates shown in Fig. 20. It links a bucket to geochemical properties at each digging spot. An immediate consequence is the increased fidelity of the stacked histograms in Fig. 21 compared with the same plots based on grade-block averages in Fig. 14.

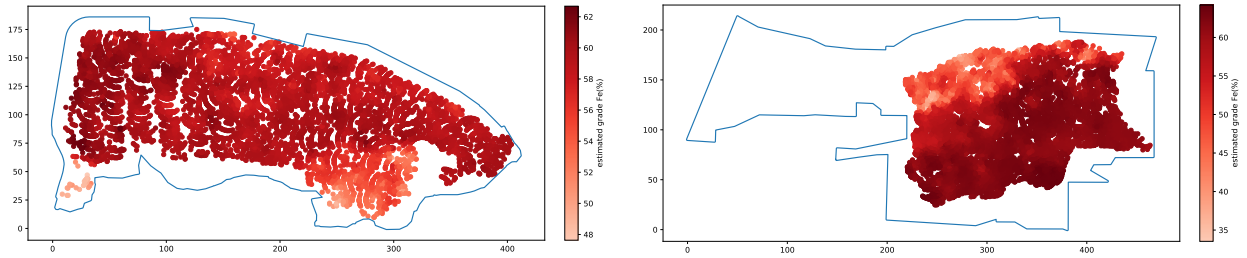

Figure 20: Kriging mean grade estimates for the dig buckets — (left) LOAD_AREA_2, (right) PIT_AREA_1

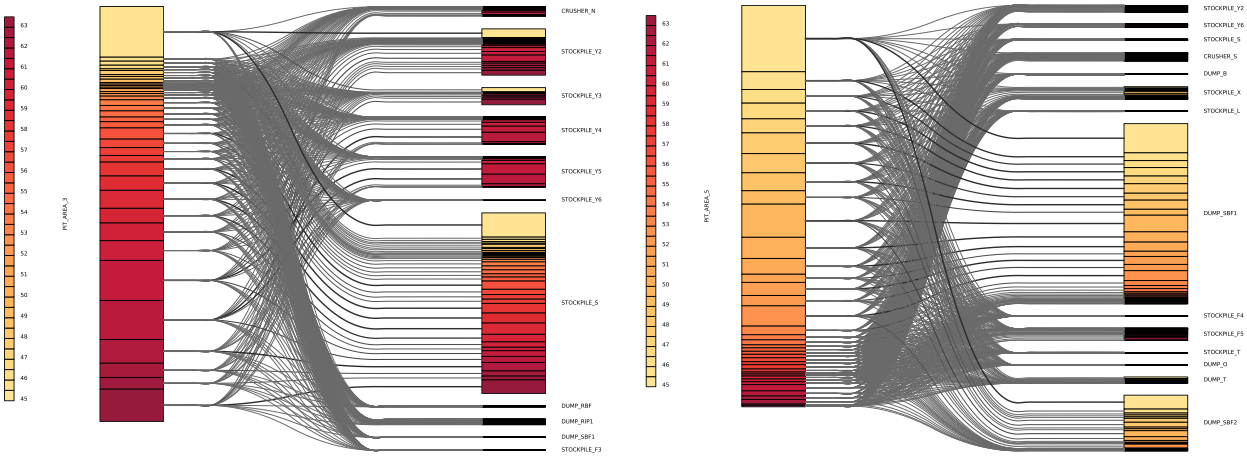

Figure 21: Composition and material transfer plots similar to Fig. 14. Grades estimated with kriging interpolation of blasthole assays have higher fidelity. Loading locations are (left) PIT_AREA_3, (right) PIT_AREA_5

Finally, we integrate all these steps and illustrate how grade risks are transferred by mapping dig buckets (within a loading zone) to truck dump positions (within a stockpile).[7] Fig. 22 shows the grade-blocks, blastholes etc. in a problematic area. The kriging estimated Fe grade at the dig locations reveal waste material is being excavated in an otherwise HG zone—see northeast pile in the dig pattern. This relates to the example in Fig. 14(left) where it was pointed out this low-grade (Fe $\leq 52\%$) material was moved unexpectedly from PIT_AREA_3 to STOCKPILE_Y2 and STOCKPILE_Y3. To confirm this, the sources contributing to the composition of the truck dumps at STOCKPILE_Y2 and STOCKPILE_Y3 are shown in Fig. 23. In the "source identification" panel, the blue dots corresponding to PIT_AREA_3 have been identified as the root cause of the contamination and elevated uncertainty. In the case of STOCKPILE_Y2, the waste dumps (with Fe $\leq 50\%$) are more localised whereas they tend to be more scattered in STOCKPILE_Y3.

5.1 Summary

This example demonstrates that linking bucket position estimates with ore properties is crucial for understanding the risks posed by material movement. The key to success is fine-grained localisation, which requires both dig events and material composition to be modelled at sufficiently high resolution. This is made possible by monitoring interactions between excavators and haul trucks using GPS data, and capturing geochemical variations with higher precision than grade-blocks. As noted in Sec. 3, the bucket inference procedure improves data completeness and provides consistent dig location estimates at the source. The dump locations at the destinations (e.g. stockpiles) are similarly obtained by analysing vehicle GPS patterns to track material movement. Knowledge about the composition and movement of material allows adverse events such as ore dilution to be identified. In the remainder of this paper, we will explore several issues and consider how the system might evolve to handle complex movements in a more sophisticated way. First, the notions of uncertainty and grade-risk are presented using Monte Carlo simulation. Then, we consider the arguments for extending the deterministic load-dump mappings (demonstrated in this paper) to stochastic material

[7] The locations where trucks unload are found by detecting specific manoeuvres such as three-point turn, then matching vehicle gps coordinates with dump time intervals reported in lh_cycles.

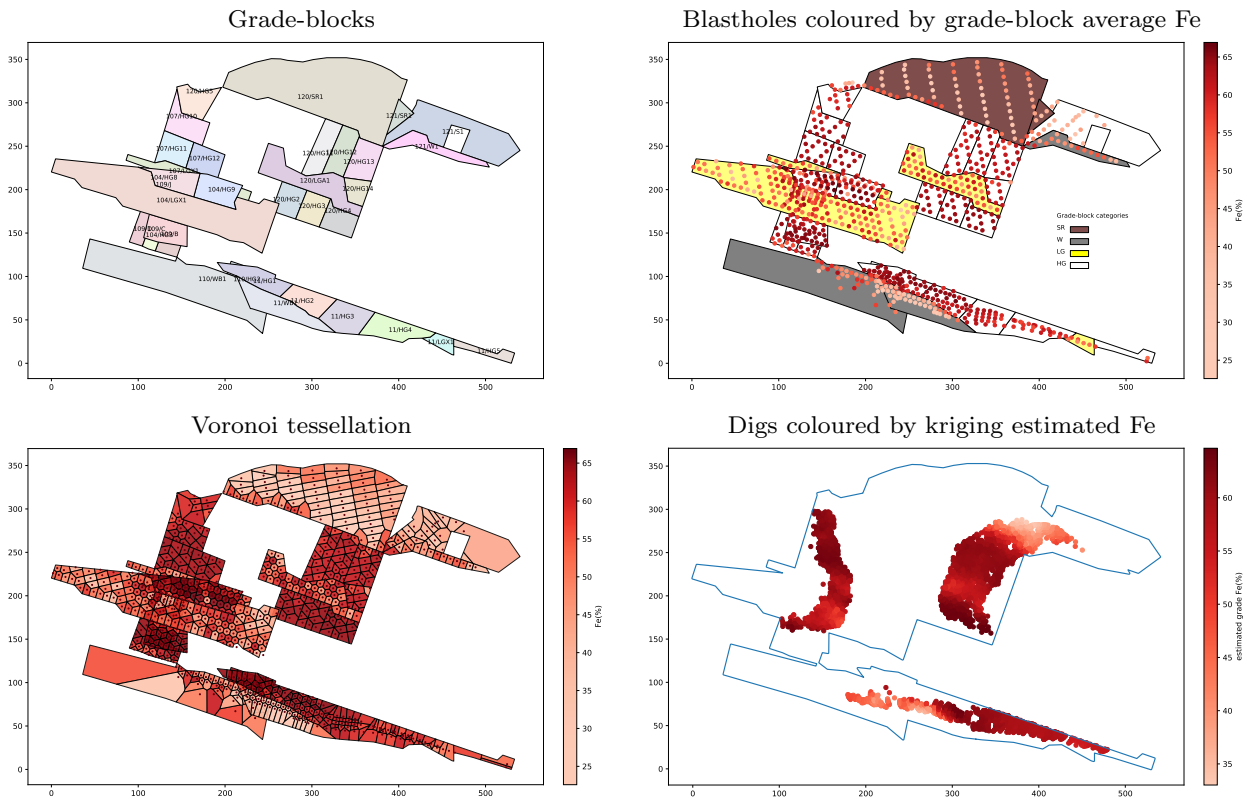

Figure 22: Analysis of a problematic area. Load location: PIT_AREA_3

tracking using a dynamic probabilistic graphical model. Finally, we consider future possibilities in light of recent developments and studies published in the literature.

6 Discussion

Section 5 has shown the value of being able to pinpoint the locations where load-dump events take place. Linking material movement to orebody chemistry at sufficiently high resolution, it is possible to identify problematic areas at specific sites. For instance, when out-of-spec material is transferred inadvertently to a stockpile. This was one of the stated objectives in the introduction. Apart from establishing this link, high fidelity region-based grade models were constructed using kriging and blasthole assays near the area of excavation. Beyond deterministic mappings, the ensuing discussion considers three further questions to broaden our current perspective. First, how to incorporate uncertainties or use probabilities to convey grade-risk? Second, what compelling reasons are there for tracking material stochastically using a dynamic probabilistic graphical model? Third, looking at the bigger picture, what synergy is created by data analytics with respect to other open-pit mining processes?

6.1 Uncertainties and perception of grade-risk

Monte Carlo (MC) simulation can provide an effective way to synthesize the effects of positional and grade uncertainties. For illustration, MC simulation was performed in PIT_AREA_3 to emulate the effects of bucket and compositional uncertainties.[8] Fig. 24(a) shows the predicted mean of Fe when the bucket position is perturbed. This mimics the effects of material dispersion which is similar to convolution. Fig. 24(b) highlights grade differences with respect to Fig. 22(lower right) which assumed perfect localisation (zero interaction). Fig. 24(c) simulates the grade uncertainty and shows the standard deviation peaks at locations where the ore transitions from high-grade to low-grade or waste. Having computed $\hat{\mu}_{\mathbf{x}}$ and $\hat{\sigma}_{\mathbf{x}}$, the grade at any modelled location $\mathbf{x}$ may be viewed through the lens of a probability distribution, $\mathcal{N}(\hat{\mu}_{\mathbf{x}}, \hat{\sigma}_{\mathbf{x}})$. Suppose an excavator has been tasked to move high-grade material to a stockpile. The risk of digging up low-grade material and transferring this accidentally to a stockpile might be expressed as

[8]The procedure is given in Appendix A. Results are obtained using the parameters $n_1 = 30, n_2 = 500, r = 15$ and $\rho = 25$.

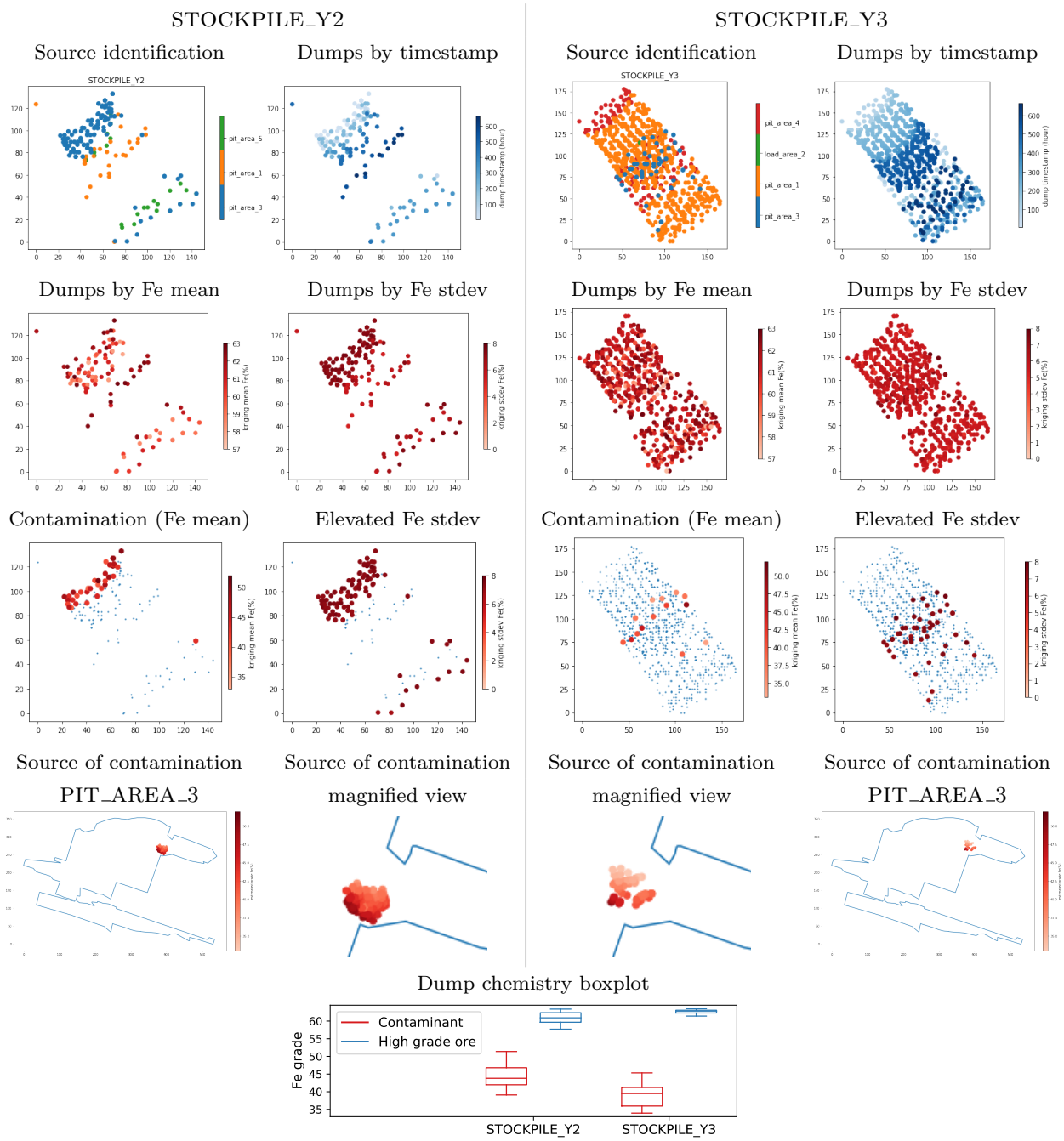

Figure 23: Linking material movement with ore properties, connecting loading sites with truck dump locations in the stockpiles

$p(\text{Fe} \leq \tau) = \Phi_{\hat{\mu}_{\mathbf{x}},\hat{\sigma}_{\mathbf{x}}}(\tau)$ for some arbitrary threshold τ where

$$\Phi_{\mu,\sigma}(\tau) \equiv \int_{-\infty}^{\tau} \frac{\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} dy \tag{3}$$

Concretely, Fig. 24(d) shows the risk, $p(\text{Fe} \leq 52)$, taking into account grade variability. Level-sets may be used to convey the level of risk (or safety) associated with digging in different areas. The contour plot in Fig. 25 shows the corresponding risk profile using level-sets. Blue regions are deemed safe digging areas. Red regions are classified

as high risk where $p(\text{Fe} \leq 52)$ exceeds 50%. When dig positions (the white dots) are superimposed on this map, it can provide guidance on truck destination selection. For instance, a conflicting destination (e.g. stockpile Y_2) may be overriden to ensure mine waste excavated from the dark red regions are transferred correctly to waste dumps.

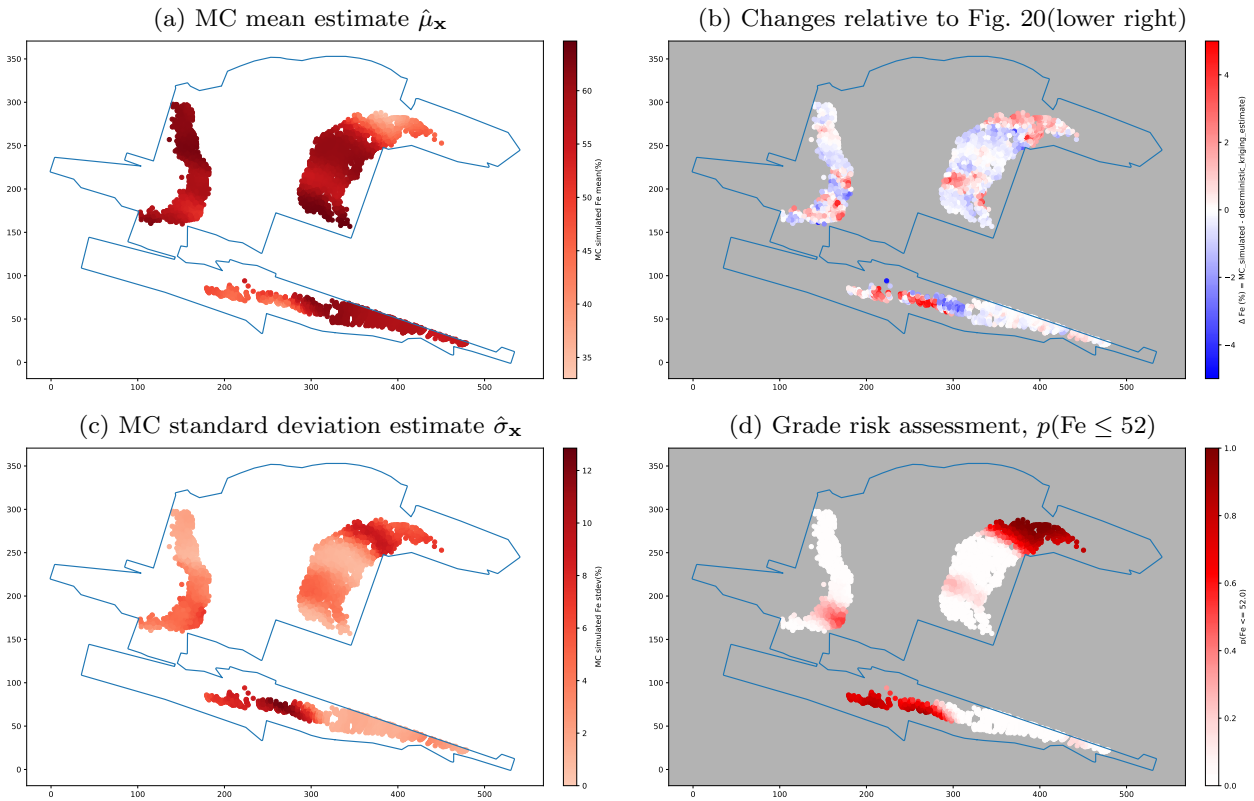

Figure 24: Monte Carlo simulation allowing for positional and grade uncertainties. Load location: PIT_AREA_3

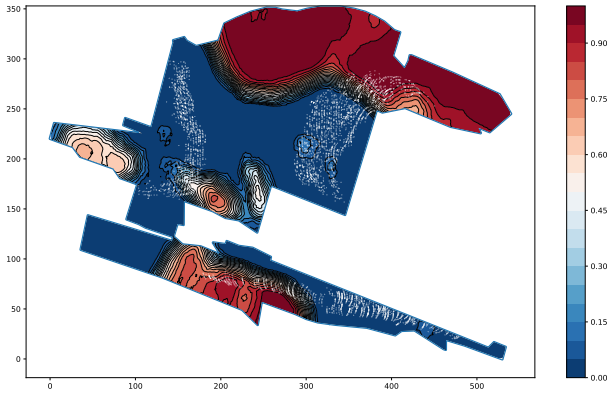

Figure 25: Monte Carlo simulation: level-set representation of risk profile $p(\text{Fe} \leq \tau)$ with $\tau = 52$

Table 8 outlines some of the opportunities presented by probabilistic grade modelling using MC simulation and contour plots. Amongst these, the 'strategic pit selection' idea is further explored in Fig. 26. This figure illustrates what is known as *spatial diversity*—that it is possible to source the same material (material with the closest match to the specification $58 \leq \text{Fe} < 61$ are rendered in red) from multiple locations. These bounds can be set freely, depending on blending requirements and operational objectives; potentially across multiple chemical components. The significance is that such picture presents a mine planner with multiple options and a risk-based view of the pit. Compositional likelihood, vehicle proximity and travel distance are factors that can be factored into source-target selection and dispatch decisions [32] to optimise fuel use and material delivery to the crusher.

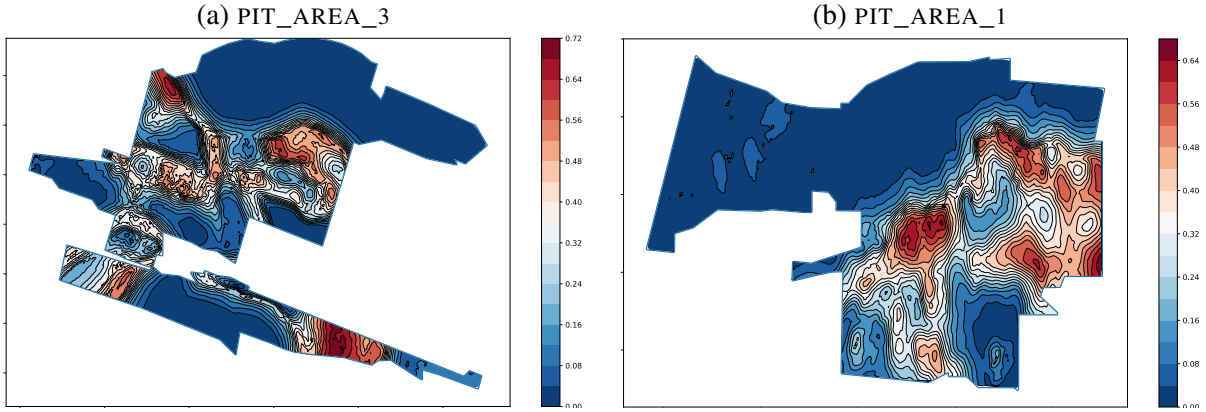

Figure 26: Seeking material within a targeted band: $p(58 < \text{Fe} \leq 61)$ at PIT_AREA_3 and PIT_AREA_1

Opportunities	Benefits
Geologically-aware, risk-factors informed excavation and load-dump operations	• Communicate level of safety (or risk) associated with digging at different locations across a mine site. • Ensure mine waste are transferred to waste dumps. Reduce risk of stockpile contamination. Potential improvement in ore recovery.
Precision mining	• Ability to target material within a particular band (e.g. $58 < \text{Fe} \leq 61$)
Strategic pit selection and equipment dispatch	• Present multiple sourcing options at high-resolution for short-term mine planning and asset optimisation. • Ability to take risk-adverse actions, e.g., re-routing truck to new destination on-demand to dispose low-quality (or low-confidence) material. • Assign specific load-haul instructions to fleet considering where and how many sources are available, vehicle proximity, fuel usage etc.
Pervasive adoption and vindication of stochastic frameworks (a paradigm shift) in the mining industry	• Recognise the limitations of using hard thresholds for grade-control. • More sophisticated understanding of the inherent risk when a grade-block is not homogeneous. Its average grade might be marginally higher than the waste threshold when it contains half ore and half waste. For a block with mean LOI concentration of 6.5 and s.d. of 1, if the acceptance condition is given by LOI ≤ 7, it still runs a 30% risk of exceeding the threshold since $p(\text{LOI} > 7) = 1 - \Phi_{\mu,\sigma}(7) = 0.3085$.

Table 8: Opportunities presented by probabilistic grade modelling using MC simulation and contour plots

6.2 Material tracking using a dynamic probabilistic graphical model

The arguments for employing a dynamic probabilistic graphical model (DPGM) for material tracking stem from the emergence of complex serial dependencies, and the need to quantify uncertainties properly when material are mixed. Specifically, the grade variance may increase or decrease in non-trivial ways when material are pooled together from different locations depending on their correlation. In mining excavation, the risk associated with waste contamination in a stockpile may be categorised as endogenous (action-induced) uncertainty; this can greatly distort the mean. However, there is also risk attached to underlying variation in orebody composition. In Sec. 6.1, we have already witnessed how grade uncertainty enters into grade risk calculations via $\Phi_{\mu_{\mathbf{x}},\sigma_{\mathbf{x}}}$.

Figure 23 illustrated the case of digging at the wrong place and subsequent contamination of a stockpile. This scenario involves single-hop transactions where mistaken action, or inadequate knowledge, has resulted in waste being transferred to a high-grade destination. When material transfer involves multiple hops between the source and final destination, the excavation (or reclamation) might originate from run-of-mine (ROM) stockpiles rather than a mining pit. In such situation, a more advanced model, for instance a discretised state-space model similar to [13], is needed to propagate changes triggered by cascaded and parallel load-dump events. A probabilistic graphical model may be constructed to describe the uncertainties and correlations that exist between [the mass and geochemical] variables which reflect the state of a mine [12]. The ability to identify ore property correlations across different locations can increase situational awareness when there is an issue, and provide multiple avenues for sourcing material to meet production targets.

Fundamentally, material movement induces volumetric and compositional changes at a mine site. These changes are dependent on variation in the underlying geology as well as excavation activities which contribute to the mixing of correlated material throughout a mine. Within a graphical model, a collection of voxels may represent the states (e.g., grade properties) for a small area in a pit, stockpile or even a processing plant. Transportation of excavated material creates linkage between different locations. This is often represented by edges in a graph that permits local updates to be triggered by discrete events which allow the model to evolve. The load and dump actions distribute material and connect a spatiotemporal source with one or more targets. This allows ore properties (statistical moments such as $\mu_{\mathbf{x}}, \sigma_{\mathbf{x}}$) to be projected through multiple links without losing uncertainty information.

A dynamic probabilistic graphical model (DPGM) can track changes and correlations that develop over time and express uncertainties in a mathematical framework. These changes are not always intuitive or easy to understand given the complex dependencies and variety of source-target connections at any given period in time. Hence, a graphical model that provides probabilistic inference and updates its states via learning algorithms can provide an invaluable tool for quantifying risks objectively. Using message passing mechanisms [33], each excavation/reclamation action constitutes a localised update. Belief propagation through the network is typically approached using junction trees [34]; such algorithm allows marginalisation and inferences to be performed as the model evolves based on new information provided [35].

6.3 Synergy and future possibilities

6.3.1 Machine learning using equipment as a sensor

There is also scope for using machine learning techniques [36] in data fusion and inferencing tasks, leveraging their generalisation ability to robustly handle missing data and discover new trends. In [37], Liu et al. apply machine learning to the force, energy and kinematic parameters to infer the material type and physical properties (e.g., hardness, lumpiness) of the excavated material. The digging equipment is used effectively as a sensor to classify material in situ allowing for disturbances due to blast movement [38]. In general, knowing the material type can help mitigate issues associated with storage, transportation, and material handling—including dust-suppression. Conceivably, the inferred material type data could be added as an input to the tracking pipeline to complement the dig position estimates.

6.3.2 Hardware integration

In the equipment automation space, manufacturers have started installing X-ray fluorescence (XRF) sensors on shovels to characterise material in individual buckets [39] and integrating with fleet management systems to minimise incidence of ore-dilution. Multiple sensors capture the spectra of excavated material from different vantage points. Operating at 10Hz, these sensors produce 80-400 spectra per load-dump segment. Measurements are aggregated and converted to in-situ grade estimates. These hardware instrumentation approaches permit model reconciliation [40] and allow the accuracy of assay-based grade predictions to be tested/adjusted to account for blast movement. However, this sensing modality is known to be less effective for trace elements. Aside from Cu, Zn, Pb, Ni and Fe, lighter elements (Al, Si and P are most relevant for iron ore deposits) might also lie outside the detection range. There is currently no published data on their long-term viability in terms of cost (installation and maintenance), scale of operation (ubiquitous vs partial deployment) and robustness (against abrasion and vibration).

6.3.3 Uncertainties permeate the mining chain

Looking further afield, error resilience and robustness are important considerations in mine planning. In essence, one can think of these as the ability to create optimal or even just feasible and useful plans in the face of uncertainty [41]. These considerations apply not only to truck dispatch [42] from pit to plant [43] (refer to mining operations described in [21] for context), they also extend to material blending [44], ore car loading / train scheduling [45] and scenario-based network optimisation [46] problems which are generally better served given more complete data and accurate estimates of the uncertainties. The significant difference foreshadowed by probabilistic tracking is the immediacy and precision with which this information is served. Instead of aggregating data over weeks or days for retrospective analysis, emerging trends and fine-grained information can be observed in the evolving model as load-dump estimates or telemetry data stream in. Grade averages and uncertainties are computed with bucket precision rather than averaged over an entire grade-block or stockpile.

6.3.4 Optimisations for mine planning: updating grade uncertainty in a production setting

Various optimisation models for mine planning can be found in the literature [47, 48, 49, 50, 51, 52, 53, 54, 55]. These formulations incorporate resource allocation, production and quality constraints in different ways. In [29],

Lagos et al. consider the evolution of geostatistical uncertainty—using grades of mined blocks to update estimates for unmined blocks—to produce operational policies that minimise risk in the production schedule for short-term planning. In contrast, Del Castillo and Dimitrakopoulos [56] focus on supply uncertainty and investment options, such as equipment purchase decisions to meet service requirements or increase production capacity over the life of a mine.

6.3.5 Particle-based simulation: an alternative for mine-to-mill material tracking

Recently, Servin et al. proposed a particle-based digital-twin model [57] for mine-to-mill material tracking. This is an interesting alternative to creating a graph-based density propagation model in a 3D voxelated world since each pseudo-particle acts as a carrier of observational data generated by the equipment as it interacts with the material. Each simulation presents a view associated with a physical asset—such as a truck, conveyor belt or stockpile—this can be any mobile or fixed entity in the mine. Movements and state changes are dependent on the data coming from control systems, sensors, vehicle telemetry or model synthesis which allow cause and effect to be predicted. This may help answer, for instance, how natural variation in the material at different sources and the actions taken affect the concentration and hardness of the material at the crusher.

6.3.6 Reflection on data analytics and potential for reinforcement learning

Successful integration of load-haul operation spatial analysis with material tracking opens many possibilities. One application is identifying high-risk source locations that warrant further investigation. Concretely, additional assay samples may be obtained in an uncertain region where the geology has been misclassified, or the drill-blast strategies [58] may be revised to improve fragmentation and mitigate risk in the next bench. For instance, Levinson and Dimitrakopoulos [59] have described a reinforcement learning approach to determine additional infill-drilling locations in response to uncertainty. In terms of process efficiency and maintenance, increased production costs and significant delay can occur when viscous materials (e.g., clay-rich minerals) are fed to a screening station or plant due to clogging. If this can be traced back to its origin, it is likely the problem can be prevented in the first place. As mentioned in the introduction, tracking material movement associated with load-haul operations represents a major step for further advances and achieving longer-term objectives in mining automation.

7 Conclusions

This paper demonstrates the benefits of using data analytics more widely in mining automation. For material tracking, estimating dig positions based on loader-truck interactions has led to higher spatial precision and load-haul cycle recall rates. Analysis reveals the existing dig estimates provided by the lh_buckets table are incomplete and at times inconsistent. For comparison, the status quo and proposed method produce bucket position estimates with precision of 84% and 99%, respectively. Furthermore, the load interval recall rate increases from approximately 89% to 97.5% (for wheel-loaders) and from 91.4% to 98.2% (for excavators) when both estimates are used. This represents a conditional improvement of 6% to 8.5%, a margin large enough to make a significant difference at scale. Subsequently, the fixed bucket range assumption is revisited. Experiment results show that neglecting joint movements in the excavator's arm has only a small to negligible effect on performance. Hence, the use of a rigid motion model for the purpose of estimating bucket positions is justified.

The examples that follow illustrate the steps involved in tracking volumetric and compositional changes associated with load-dump events. The objective is to emphasize what can be achieved by linking bucket positions with orebody chemistry using various forms of visualisation. Sankey diagram is used to illustrate the complexity of material flow in a real mine. To distil information, stacked histograms are embedded in composition and material transfer plots to make sense of the transactions between a source and multiple destinations and vice-versa. Region-based grade models are constructed using kriging and blasthole assay samples to provide mean and variance estimates relevant to an excavated area. Specifically, the case study demonstrates how digging at the wrong location—and perhaps over-reliance on grade estimates with inadequate spatial resolution—can lead to contamination of a stockpile when waste material is mistaken for ore. The ensuing discussion outlines applications that may benefit from high fidelity accounts of grade risk resulting from material transfer and describes the role dynamic probabilistic graphical models can play in quantifying uncertainties (risks) and correlations as the connections become more complicated. Monte Carlo simulation and level-sets have been used to visualise grade risk associated with excavation.

8 Appendices

Appendix A presents a Monte Carlo simulation procedure for modelling positional and grade uncertainties. Appendix B describes how the theoretical grade standard deviation is computed and the associated hyperparameters.

A Monte Carlo simulation for positional and grade uncertainties

Monte Carlo (MC) simulation provides an effective way to synthesize the effects of positional and grade uncertainties. Positional uncertainty may arise from GPS localisation errors or material movement due to blasting. More generally, given a region of influence $\Omega_{\mathbf{x}}$ centered at $\mathbf{x}$, the points $\mathbf{s} \in \Omega_{\mathbf{x}}$ may contribute $f(\mathbf{s}) \in \mathbb{R}^+$ to the composition of a dig bucket at location $\mathbf{x}$. In the absence of a mixture or dispersion model, Procedure 2 describes one possible way to generate samples assuming: (a) non-stationary geostatistics, (b) $\Omega_{\mathbf{x}}$ is isotropic and each point carries equal weight.

Procedure 2 Monte Carlo simulation of positional and grade uncertainties

Require: Access to tables *gradeblock_properties* and *hole_properties* in database

Input: $A \in \mathbb{R}^{m \times n}$: blasthole geochemical assays within modelled region (m samples, n chemical components).
G: mean grade prediction model. Default: ordinary kriging.
$\Omega_{\mathbf{x}}$: region of influence. Default: a circle with radius r centered at $\mathbf{x} \in \mathbb{R}^2$.
$f_{\mathbf{x}}(\mathbf{s})$: weight of contribution from position $\mathbf{s}$ to $\mathbf{x}$. Default: $f_{\mathbf{x}}(\mathbf{s}) = 1 \, \forall \mathbf{x}, \forall \mathbf{s}$.
n_1: number of positions to draw from each region of influence.
n_2: number of particles to sample for each queried location.

Definitions: $\mathbf{c} = A(:, c) \in \mathbb{R}^m$ contains the assay grades for the chemical of interest (e.g. Fe).
$n_i^{\text{empirical}}$: effective number of blasthole samples enclosed in a local neighbourhood with radius ρ. These samples are used to compute the robust s.d. using the MCD estimator [60].

1: **for** each location $\mathbf{x}$ in the modelled region **do**
2: Initialise set of particles $\mathcal{P}$ to $\{\,\}$.
3: Draw n_1 random positions $\mathbf{s}_i$ uniformly distributed in $\Omega_{\mathbf{x}}$. Let $\mathcal{S}_{\mathbf{x}} = \{\mathbf{s}_i \mid \mathbf{s}_i \in \Omega_{\mathbf{x}}\}_{1 \le i \le n_1}$
4: **for** each position $\mathbf{s}_i$ in set $\mathcal{S}_{\mathbf{x}}$ **do**
5: Look up grade mean: $\hat{\mu}_i \leftarrow G(\mathbf{s}_i)$
6: Look up theoretical grade standard deviation (s.d.): $\hat{\sigma}_i^{\text{theoretical}} \leftarrow \texttt{theoretical\_stdev}(\hat{\mu}_i, c)$ $\star$
7: Estimate grade s.d. Apply the MCD estimator to blasthole samples in the neighbourhood of $\mathbf{s}_i$: $\hat{\sigma}_i^{\text{empirical}}, n_i^{\text{empirical}} \leftarrow \texttt{robust\_sample\_stdev}(\mathbf{s}_i \,|\, \mathbf{c})$
8: Compute the effective grade s.d. as a convex combination of $\hat{\sigma}_i^{\text{theoretical}}$ and $\hat{\sigma}_i^{\text{empirical}}$: Set $\hat{\sigma}_i = \lambda \hat{\sigma}_i^{\text{empirical}} + (1-\lambda)\hat{\sigma}_i^{\text{theoretical}}$ where $\lambda = \frac{n_i^{\text{empirical}}}{n_i^{\text{empirical}} + n^{\text{theoretical}}}$ and $n^{\text{theoretical}} = 6$.
9: Draw $n_2 \cdot f_{\mathbf{x}}(\mathbf{s}_i)$ samples from $\mathcal{N}(\cdot \mid \hat{\mu}_i, \hat{\sigma}_i)$ and append these values v_k to $\mathcal{P}$.
10: **end for**
11: Estimate grade mean and variance at $\mathbf{x}$ from $\mathcal{P}$ $\hat{\mu}_{\mathbf{x}} = \frac{1}{N}\sum_k v_k$ and $\hat{\sigma}^2_{\mathbf{x}} = \frac{1}{N}\sum_k (v_k - \hat{\mu}_{\mathbf{x}})^2$ where $N = \sum_i f_{\mathbf{x}}(\mathbf{s}_i) n_2 = n_1 n_2$.
12: **end for**

Output: Collection of $[\hat{\mu}_{\mathbf{x}}, \hat{\sigma}_{\mathbf{x}}]$

$\star$ Refer to formulas and parameters given in Appendix B

A crucial part of this procedure is local grade variance estimation. This needs to reflect natural grade variability (the aleatoric uncertainty) of the deposit rather than blasthole spacing (the epistemic uncertainty). In line 8, $\hat{\sigma}_i$—the standard deviation (s.d.) at each influential location $\mathbf{s}_i$ around the point of query $\mathbf{x}$—is a convex combination of $\hat{\sigma}_i^{\text{empirical}}$ and $\hat{\sigma}_i^{\text{theoretical}}$. The term $\hat{\sigma}_i^{\text{empirical}}$ represents a robust estimate of the local s.d. which dominates when the number of blasthole samples is plentiful ($n^{\text{empirical}} \gg n^{\text{theoretical}}$). Conversely, when a region is devoid of blasthole assays, $\hat{\sigma}_i^{\text{theoretical}}$ provides a sensible default value. To compute $\hat{\sigma}_i^{\text{empirical}}$, the MCD robust estimator [60] is used, as it suppresses outliers (incongruous measurements) within a local neighbourhood. This avoids over-estimation of the variance which in turn prevents smearing (inflation of $\hat{\sigma}_{\mathbf{x}}$) in the final result.

B Estimating grade variability: control parameters over σ

When a location is devoid of blasthole assays, $\texttt{theoretical\_stdev}((\hat{\mu}_i, c)$ computes a baseline value $\hat{\sigma}_i^{\text{theoretical}}$ in accordance with Procedure 3. A set of hyperparameters is associated with each chemical, c. The dependence on $\hat{\mu}_i$ ensures the variance is adjusted in a manner that is geologically reasonable and consistent with empirical observation. For instance, Fe variance is generally much higher in low grade material than in high-grade (mineralised) material.

Procedure 3 Computing a theoretical value for the local grade standard deviation

Require: Hyperparameters $(\alpha, \beta, L, R, \lambda)$ in Table 9
Definition: The function $\texttt{theoretical\_stdev}(\hat{\mu}_i, c)$ returns a default s.d. value $\hat{\sigma}_i^{\text{theoretical}}$.
Input: c: chemical component, such as Fe, SiO_2, Al_2O_3, LOI or P.
$\hat{\mu}_i$: expected grade of c at location $\mathbf{s}_i \in \mathbb{R}^2$.
1: Compute $g_{\text{logistic}} = [1 + \exp(-\alpha_c(\hat{\mu}_i - \beta_c))]^{-1}$ ⋆
2: Compute $\nu = L_c + (R_c - L_c)(1 - g_{\text{logistic}})$ ⋆
3: Draw a random sample from the Laplacian distribution: $x \sim \texttt{Laplacian}(\nu, \lambda_c)$ ⋆
4: Let $\hat{\sigma}_i^{\text{theoretical}} = \max\{x, 0\}$
Output: $\hat{\sigma}_i^{\text{theoretical}}$
⋆ These functions are plotted in Fig. 27

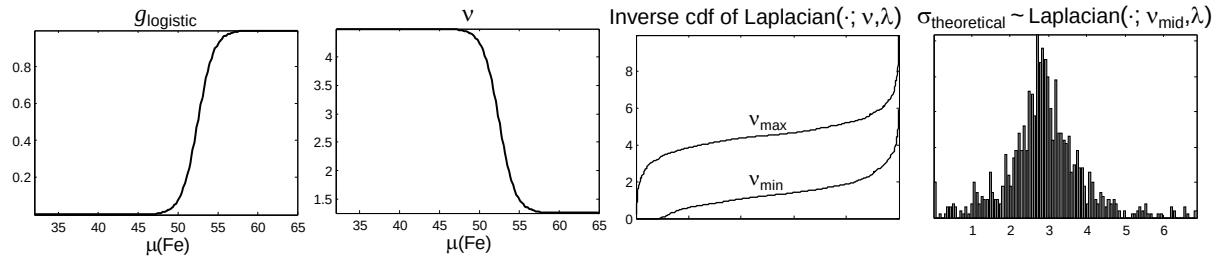

Figure 27: Plot of functions used in Procedure 3

Table 9: Hyperparameters used in Procedure 3

Chemical c	typical range	α_c	β_c	L_c	R_c	λ_c
Fe	$[32, 65]$	1	52.5	1.25	4.5	1
SiO_2	$[5, 40]$	1	22	4.5	1.25	0.9
Al_2O_3	$[0.8, 7.2]$	5	3	1.1	0.35	0.1
LOI	$[1.5, 12]$	3	6	3	1.2	0.25
P	$[0.06, 0.22]$	35	0.14	0.05	0.02	0.0075

Acknowledgements

This work was supported by the Australian Centre for Field Robotics and the Rio Tinto Centre for Mine Automation.

References

[1] Danish Ali. Advanced analytics for surface mining. In *Advanced Analytics in Mining Engineering*, pages 169–179. Springer, 2022.

[2] Tracey L Kerr, Anthony P O'Sullivan, Peter Waters, Richard Turner, and Darryl C Podmore. Geophysics and iron ore exploration: examples from the Jimblebar and Shay Gap-Yarrie regions, Western Australia. *ASEG Extended Abstracts*, 1994(1):355–368, 1994.

[3] Katherine Silversides and Arman Melkumyan. Machine learning for classification of stratified geology from MWD data. *Ore Geology Reviews*, page 104737, 2022.

[4] Konstantin M Seiler, Andrew W Palmer, and Andrew J Hill. Flow-achieving online planning and dispatching for continuous transportation with autonomous vehicles. *IEEE Transactions on Automation Science and Engineering*, 19(1):457–472, January 2022.

[5] Paweł Stefaniak, Piotr Kruczek, P Śliwiński, Norbert Gomolla, A Wyłomańska, and Radosław Zimroz. Bulk material volume evaluation and tracking in belt conveyor network based on data from SCADA. In *Proceedings of the 27th International Symposium on Mine Planning and Equipment Selection-MPES 2018*, pages 335–344. Springer, 2019.

[6] Guillaume Caumon, PLCD Collon-Drouaillet, C Le Carlier de Veslud, Sophie Viseur, and J Sausse. Surface-based 3d modeling of geological structures. *Mathematical Geosciences*, 41(8):927–945, 2009.

[7] Flávia Cristina Silveira Braga, Carlos Alberto Rosiere, João Orestes Schneider Santos, Steffen G Hagemann, and Pedro Valle Salles. Depicting the 3D geometry of ore bodies using implicit lithological modeling: An example from the Horto-Baratinha iron deposit, Guanhães block, MG. *REM-International Engineering Journal*, 72:435–443, 2019.

[8] SG Hagemann, T Angerer, Paul Duuring, CA Rosière, RC Figueiredo e Silva, L Lobato, AS Hensler, and DHG Walde. BIF-hosted iron mineral system: a review. *Ore Geology Reviews*, 76:317–359, 2016.

[9] Nadia Mery, Xavier Emery, Alejandro Cáceres, Diniz Ribeiro, and Evandro Cunha. Geostatistical modeling of the geological uncertainty in an iron ore deposit. *Ore Geology Reviews*, 88:336–351, 2017.

[10] Arja Jewbali, Fabio T. Ramos, and Arman Melkumyan. A non-parametric Bayesian framework for automatic block estimation. In *Proceedings, Applications of Computers and Operations Research in the Minerals Industries (APCOM)*, number 056, pages 1–20, 2011.

[11] Anna Chlingaryan, Arman Melkumyan, and Raymond Leung. IntegralGP: Volumetric estimation of subterranean geochemical properties in mineral deposits by fusing assay data with different spatial supports. *Expert Systems with Applications*, 298A:Article 129429, 2026.

[12] Tim Bailey. Probabilistic graphical model for tracking ore properties from pit to plant (unpublished). Technical report, The University of Sydney, 2021.

[13] Christopher Innes, Eric Nettleton, and Arman Melkumyan. Estimation and tracking of excavated material in mining. In *14th International Conference on Information Fusion*, pages 1–8. IEEE, 2011.

[14] Christopher Innes. *A stochastic method for representation, modelling and fusion of excavated material in mining*. PhD thesis, The University of Sydney, 2012.

[15] Mehala Balamurali. A Bayesian method for estimating uncertainty in excavated material. *International Journal of Mining, Reclamation and Environment*, 36(2):125–141, 2022.

[16] Shi Zhao, Tien-Fu Lu, Ben Koch, and Alan Hurdsman. Stockpile modelling using mobile laser scanner for quality grade control in stockpile management. In *2012 12th International Conference on Control Automation Robotics & Vision (ICARCV)*, pages 811–816. IEEE, 2012.

[17] Mehala Balamurali and Konstantin M Seiler. Improving geometry dynamic prediction of in-pit stockpiles using geospatial data and polygon models. In *Proceedings, Applications of Computers and Operations Research in the Minerals Industries (APCOM)*, pages 105–116. The Southern African Institute of Mining and Metallurgy, 2021.

[18] Georgia R Markham. *Automated load-haul cycle segmentation with unsupervised machine learning*. B.E.(Hons) thesis, The University of Sydney, 2021.

[19] Jinwoo Kim and Seokho Chi. Multi-camera vision-based productivity monitoring of earthmoving operations. *Automation in Construction*, 112:103121, 2020.

[20] Danish Ali and Samuel Frimpong. Artificial intelligence, machine learning and process automation: existing knowledge frontier and way forward for mining sector. *Artificial Intelligence Review*, 53(8):6025–6042, 2020.

[21] Raymond Leung, Andrew J. Hill, and Arman Melkumyan. Automation and artificial intelligence technology in surface mining: a brief introduction to open-pit operations in the Pilbara. DOI: 10.1109/MRA.2023.3328457. *IEEE Robotics & Automation Magazine*, 32(3):1–19, 2025.

[22] Marta Polak, Paweł Stefaniak, Radosław Zimroz, Agnieszka Wyłomańska, Paweł Śliwiński, and Marek Andrzejewski. Identification of loading process based on hydraulic pressure signal. *International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management*, 2:459–466, 2016.

[23] Daniel Wedge, Andrew Lewan, Mark Paine, Eun-Jung Holden, and Thomas Green. A data mining approach to validating drill hole logging data in pilbara iron ore exploration. *Economic Geology*, 113(4):961–972, 2018.

[24] Raymond Leung. Measuring sampling fairness and geochemical consensus for blasthole assays within grade-block mining units. *Computers & Geosciences*, 171:Article 105286, 2023.

[25] Roussos Dimitrakopoulos. Conditional simulation algorithms for modelling orebody uncertainty in open pit optimisation. *International Journal of Surface Mining, Reclamation and Environment*, 12(4):173–179, 1998.

[26] Raymond Leung. Modelling orebody structures: Block merging algorithms and block model spatial restructuring strategies given mesh surfaces of geological boundaries. *Journal of Spatial Information Science*, 21:137–174, 2020.

[27] Raymond Leung, Alexander Lowe, Anna Chlingaryan, Arman Melkumyan, and John Zigman. Bayesian surface warping approach for rectifying geological boundaries using displacement likelihood and evidence from geochemical assays. *ACM Transactions on Spatial Algorithms and Systems*, 8(1):1–23, March 2022.

[28] Hans Wackernagel. Ordinary kriging. In *Multivariate geostatistics*, pages 79–88. Springer, 2003.

[29] Tomás Lagos, Margaret Armstrong, Tito Homem-de Mello, Guido Lagos, and Denis Sauré. A framework for adaptive open-pit mining planning under geological uncertainty. *Optimization and Engineering*, 23:111–146, 2020.

[30] Georges Mathéron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, 1963.

[31] Mirko Mälicke. SciKit-GStat 1.0: a SciPy flavoured geostatistical variogram estimation toolbox written in Python. *Geoscientific Model Development*, 15(6):2505–2532, 2022.

[32] Abbas Aghajani Bazzazi, Morteza Osanloo, and Behrooz Karimi. Deriving preference order of open pit mines equipment through MADM methods: Application of modified VIKOR method. *Expert Systems with Applications*, 38(3):2550–2556, 2011.

[33] Martin J Wainwright. Graphical models and message-passing algorithms: Some introductory lectures. In *Mathematical Foundations of Complex Networked Information Systems*, pages 51–108. Springer, 2015.

[34] David Barber. Probabilistic modelling and reasoning: The junction tree algorithm. *Course Notes*, 2014.

[35] Tim Bailey. Stabilised moment-form junction tree for Gaussian Bayesian networks with semidefinite models. Available at: `https://ssrn.com/abstract=4176478`. *SSRN Preprint*, 2022.

[36] Yosoon Choi, Hoang Nguyen, Xuan-Nam Bui, and Trung Nguyen-Thoi. Optimization of haulage-truck system performance for ore production in open-pit mines using big data and machine learning-based methods. *Resources Policy*, 75:102522, 2022.

[37] Liyang Liu, Mehala Balamurali, Katherine L Silversides, and Rami Khushaba. Inference of geological material groups using structural monitoring sensors on excavators. In *Australasian Joint Conference on Artificial Intelligence*, pages 787–797. Springer's Lecture Notes in Computer Science Series, 2022.

[38] Zongshan Zou and Yang Jun. Modelling blast movement and muckpile formation with the position-based dynamics method. *International Journal of Mining, Reclamation and Environment*, 35(4):306–317, 2021.

[39] M Haest, K Hume, M Bradshaw, H Lang, F Faraj, and M Pal. ShovelSense measuring grade at bucket resolution – the new tool in the mine geologist toolbox. In *2022 International Mining Geology Conference*, pages 449–452. AusIMM, 2022.

[40] H Lang, R Botha, K Hume, S Sandler, and M Haest. MineSense technology empowering bucket resolution mine to mill reconciliation. In *2022 International Mining Geology Conference*, pages 375–379. AusIMM, 2022.

[41] Andrew W Palmer, Andrew J Hill, and Steven J Scheding. Modelling resource contention in multi-robot task allocation problems with uncertain timing. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3693–3700. IEEE, 2018.

[42] Chung H Ta, James V Kresta, J Fraser Forbes, and Horacio J Marquez. A stochastic optimization approach to mine truck allocation. *International journal of surface mining, reclamation and environment*, 19(3):162–175, 2005.

[43] M Samavati, AW Palmer, AJ Hill, and KM Seiler. Improvements in plan-driven truck dispatching systems for surface mining. In *Mining Goes Digital*, pages 357–366. CRC Press, 2019.

[44] Haonan Zhou, Mehran Samavati, and Andrew J Hill. Heuristics for integrated blending optimisation in a mining supply chain. *Omega*, 102:102373, 2021.

[45] William Jones and Philip Gun. Train timetabling and destination selection in mining freight rail networks: A hybrid simulation methodology incorporating heuristics. *Journal of Simulation*, pages 1–14, 2022.

[46] E Bakhtavar and H Mahmoudi. Development of a scenario-based robust model for the optimal truck-shovel allocation in open-pit mining. *Computers & Operations Research*, 115:104539, 2020.

[47] PJ Ravenscroft. Risk analysis for mine scheduling by conditional simulation. *Transactions of the Institution of Mining and Metallurgy. Section A. Mining Industry*, 101, 1992.

[48] Marcelo Moretti Fioroni, Luiz Augusto G Franzese, Tales Jefferson Bianchi, Luiz Ezawa, Luiz Ricardo Pinto, and Gilberto de Miranda. Concurrent simulation and optimization models for mining planning. In *2008 Winter simulation conference*, pages 759–767. IEEE, 2008.

[49] Natashia Boland, Irina Dumitrescu, and Gary Froyland. A multistage stochastic programming approach to open pit mine production scheduling with uncertain geology. *Optimization online*, pages 1–33, 2008.

[50] R Dimitrakopoulos and S Ramazan. Uncertainty based production scheduling in open pit mining. *SME Transactions*, 316, 2004.

[51] SP Upadhyay and H Askari-Nasab. Truck-shovel allocation optimisation: a goal programming approach. *Mining Technology*, 125(2):82–92, 2016.

[52] Adrien Rimélé, Roussos Dimitrakopoulos, and Michel Gamache. A stochastic optimization method with in-pit waste and tailings disposal for open pit life-of-mine production planning. *Resources Policy*, 57:112–121, 2018.

[53] Adrien Rimélé, Roussos Dimitrakopoulos, and Michel Gamache. A dynamic stochastic programming approach for open-pit mine planning with geological and commodity price uncertainty. *Resources Policy*, 65:101570, 2020.

[54] Mehrnaz Mohtashama, Hossein Mirzaei-Nasirabad, Hooman Askari-Nasab, and Behrooz Alizadeh. A multi-objective model for fleet allocation schedule in open-pit mines considering the impact of prioritising objectives on transportation system performance. *International Journal of Mining, Reclamation and Environment*, 35(10):709–727, 2021.

[55] Karo Fathollahzadeh, Elham Mardaneh, Mehmet Cigla, and Mohammad Waqar Ali Asad. A mathematical model for open pit mine production scheduling with grade engineering and stockpiling. *International Journal of Mining Science and Technology*, 31(4):717–728, 2021.

[56] M Fernanda Del Castillo and Roussos Dimitrakopoulos. Dynamically optimizing the strategic plan of mining complexes under supply uncertainty. *Resources Policy*, 60:83–93, 2019.

[57] Martin Servin, Folke Vesterlund, and Erik Wallin. Digital twins with distributed particle simulation for mine-to-mill material tracking. *Minerals*, 11(5):524, 2021.

[58] S Klerkx, J Sattarvand, and M S Shishvan. Development of an explosive energy distribution optimization system to accommodate drilling errors by adjusting blasthole charges. In *Proceedings, Applications of Computers and Operations Research in the Minerals Industries (APCOM)*, pages 243–255, 2021.

[59] Zachary Levinson and R Dimitrakopoulos. Self-learning infill drilling for strategic mine planning: Simultaneously optimising the value of additional information in a mining complex under grade uncertainty. In *Proceedings, Applications of Computers and Operations Research in the Minerals Industries (APCOM)*, pages 347–354, 2021.

[60] Peter J Rousseeuw and Katrien Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999.