Machine learning in resource geology – why data quality is critical

F A Pym[1], K E Crook[2], P M Hetherington[3] and M P Murphy[4]

1. MAusIMM, Senior Mine Geologist – Nova Operation, IGO Ltd, South Perth WA 6951. Email: fletcher.pym@igo.com.au
2. MAusIMM, Senior Mine Geologist – Nova Operation, IGO Ltd, South Perth WA 6951. Email: kelsey.crook@igo.com.au
3. Superintendent – Geology – Nova Operation, IGO Ltd, South Perth WA 6951. Email: paul.hetherington@igo.com.au
4. Resource Geology Manager, IGO Ltd, South Perth WA 6951. Email: mark.murphy@igo.com.au

ABSTRACT

IGO Ltd's (IGO's) Nova-Bollinger Deposit (Nova-Bollinger) is a magmatic nickel-copper-cobalt (Ni-Cu-Co) sulfide deposit found 160 km east-north-east of the town of Norseman in Western Australia (WA). The deposit is hosted by the Mesoproterozoic rocks of the Fraser Zone, which is part of the Albany Fraser Orogen (AFO). Since mining commenced in 2016, IGO's Nova Operation (Nova) has mined and processed to 31 December 2020, 5.63 Mt of ore grading 2.04 per cent Ni, 0.86 per cent Cu and 0.07 per cent Co.

Nova's mine geology team (MGT) has used commercially available implicit modelling (IM) and general-purpose resource modelling (RM) industry software systems, to prepare the 31 December 2020 (CY20) update of Nova-Bollinger's JORC Code reportable Mineral Resource estimate (MRE). This CY20 MRE update was predominantly based on ~386 km of diamond core drilling, with the drill holes having a nominal grid spacing of 12.5 by 12.5 m throughout the mineralised zones of the deposit. Using IM tools, Nova's MGT interpreted and modelled 22 separate domains from the drill hole logging, assaying results, and the MGT's underground mapping. These IM closed volumes were then used to control the estimation of MRE variables into conventional industry digital block models using RM and other ancillary software systems.

During 2021, a commercial machine learning (ML) software was trialled by Nova's MGT for preparing block models of Nova-Bollinger's MRE domains. The motivation for the trial was that the ML process could expedite modelling through the direct creation of a domain coded block model from the drill hole logging. Compared to the industry standard approaches of wireframing or IM, where the modeller usually makes many subjective choices to produce the domain model, ML offers a more objective 'hands-off' process for determining the connectivity of domains defined in drill holes. Additionally, the ML process provides a quantitative confidence metric associated with its output that may have utility in mine production and resource classification.

In this paper, the results of the Nova-Bollinger ML modelling trial are compared to the MGT's current workflow for MRE work. The MGT's conclusions as to the advantages and disadvantages of the IM and ML methods for preparation of MRE domains are discussed. The ML software was also tested as a method of rapidly interpreting geology outside the principal zones of MRE interest. These testing results are also presented in this paper.

INTRODUCTION

IGO's Nova is 160 km east-north-east of the town of Norseman in WA, and mines and processes Ni-Cu-Co sulfide ore from an underground mine developed on Nova-Bollinger (Figure 1). The deposit is hosted by the high-grade metamorphic rocks of the Fraser Zone of the AFO. A chonolith-like gabbroic intrusion is interpreted to the be the source of Nova-Bollinger's magmatic sulfide mineralisation (Barnes *et al*, 2020).

During 2021, Nova's MGT trialled a new commercial ML software system that expedites domain block modelling by direct inference from coded drill hole data (Maptek – DomainMCF, 2021).

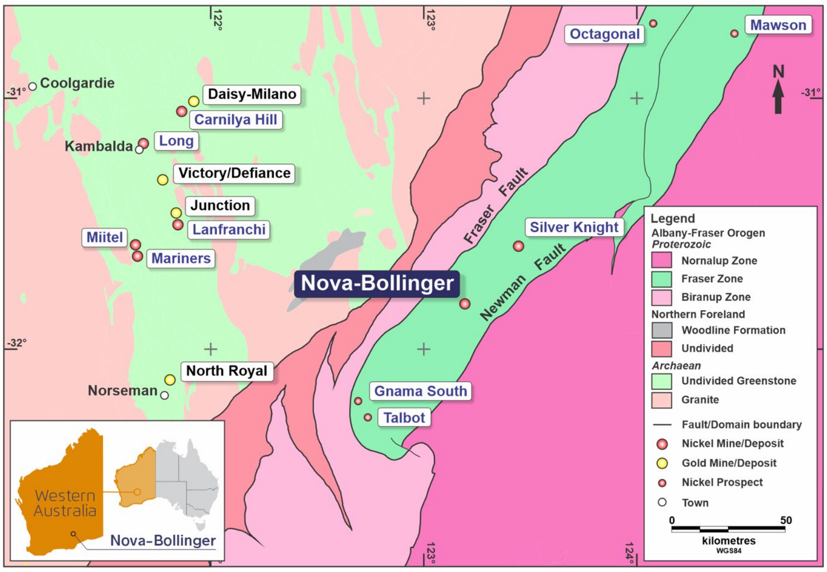


FIG 1 – Nova-Bollinger location and regional geology.

To test whether an ML approach could produce a model like IM for the Nova-Bollinger's CY20 MRE, the MGT provided the ML software with the same zone coded drill hole data that was used for the CY20 estimate. The resulting ML block model was then interrogated, and the advantages and disadvantages of both methods assessed.

Compared to the industry standard approaches of wireframing or IM, where the modeller usually makes many subjective choices to produce the domain model, ML offers a more objective 'hands-off' process (Sullivan *et al*, 2019). This paper details the results of the ML modelling trial at Nova and learnings of the MGT, along with ideas for future work.

GEOLOGY

Nova-Bollinger is in the southern part of the Mesoproterozoic Fraser Zone of the AFO. The Fraser Zone is fault bounded by the Biranup Zone to the north-east and the Nornalup Zone to the south-east (Figure 1). The Arid Basin forms the basement to the Fraser Zone and the Snowys Dam Formation of the Arid Basin, is the basement package to the intrusion complex in the Nova-Bollinger area (Spaggiari *et al*, 2014).

Mafic, ultramafic and granitic intrusions were emplaced in the region during the first phase of the AFO formation at ~1.30 Ga ago. Later, intense tectonic events from 1.14 Ga to 1.12 Ga ago, metamorphosed the Fraser Zone rocks to granulite facies. The Fraser Zone is now characterised by gneissic fabrics, complex refolding and major mylonitic zones.

The Snowys Dam Formation rocks within the Nova-Bollinger region include pelitic to psammitic gneisses, a local carbonate unit, along with metamorphosed mafic-ultramafic and volcanoclastic rocks. Nova-Bollinger's host mafic-ultramafic sill complex is a doubly plunging synform, with the deposit lying on the western side of the structure (Figure 2).

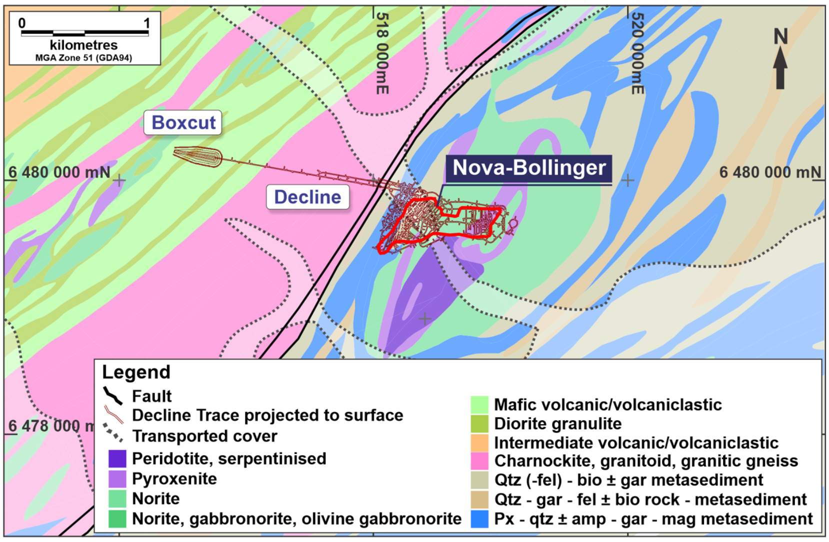


FIG 2 – Nova-Bollinger local regional geology.

RESOURCE MODELLING PRACTICE

Nova's MGT prepared the Nova-Bollinger CY20 MRE using an ensemble of software systems. (Hetherington and Murphy, 2019). Like many companies in the industry, IGO has identified preferred software systems for different stages of the MRE process, rather than adopting a preferred system for all processes.

Briefly, the drill hole data, once verified was exported from IGO's centralised database and the 22 domains of the deposit were modelled in 3D using a commercial IM oriented software system (Leapfrog Geo, 2021). The interpretation process involved preparing non-overlapping and nested 3D closed volumes for each domain by applying radial basis functions to the interpreting geologist's manual interpretation of the presence (or absence) of each domain in the drill hole paths.

Where deemed necessary, manual control strings were digitised to augment the drill hole data domain coding, with some strings interpreted from underground mapping, other strings were prepared to control implicit modelling artefacts, and sometimes strings were prepared to enforce a geologist's assessment of narrow zone connectivity. During this process it was common practice for the modelling geologists to modify the IM digital drill core logs of domains, after assessing the connectivity in the IM software and reassessment of core photographs and/or from litho-geochemical assay results. However, these changes were not back coded into central drill hole database.

When finalised, the IM 3D closed volumes were ported into a general purpose RM software system (Studio RM, 2021), where the drill hole data was coded in the RM software according to the domains interpreted in the IM. The coded drill hole assay intervals from each domain were then exported as text files for 'pre-composite statistical analysis' and missing density multivariate regression imputation using customised statistical software scripts (R Project, 2021). The drill data was then imported back into the RM software and composited to uniform lengths along with calculation of estimation service variables. The composites were then exported to text files again for statistical analysis and parallel continuity modelling in a specialised geostatistical software (Supervisor, 2021).

Over several years, the annual Nova-Bollinger MRE update has been used as an opportunity to train MGT's senior geologists in the MRE process under the supervision of the site-based Competent Person and senior technical management. The full MRE estimation workflow summarised above involved several months of (part time) work for several senior geologists amongst their other daily mine work duties.

As discussed above, a large proportion of the modelling time was spent in the process of the coding of the drill hole data into the 22 separate domains and associated iterative manipulation of the IM model results (Figure 3). The block estimation process also took several weeks (full time) to estimate and fully validate the model. Several more weeks of part time commitments were then required to prepare the detailed MRE modelling documentation.

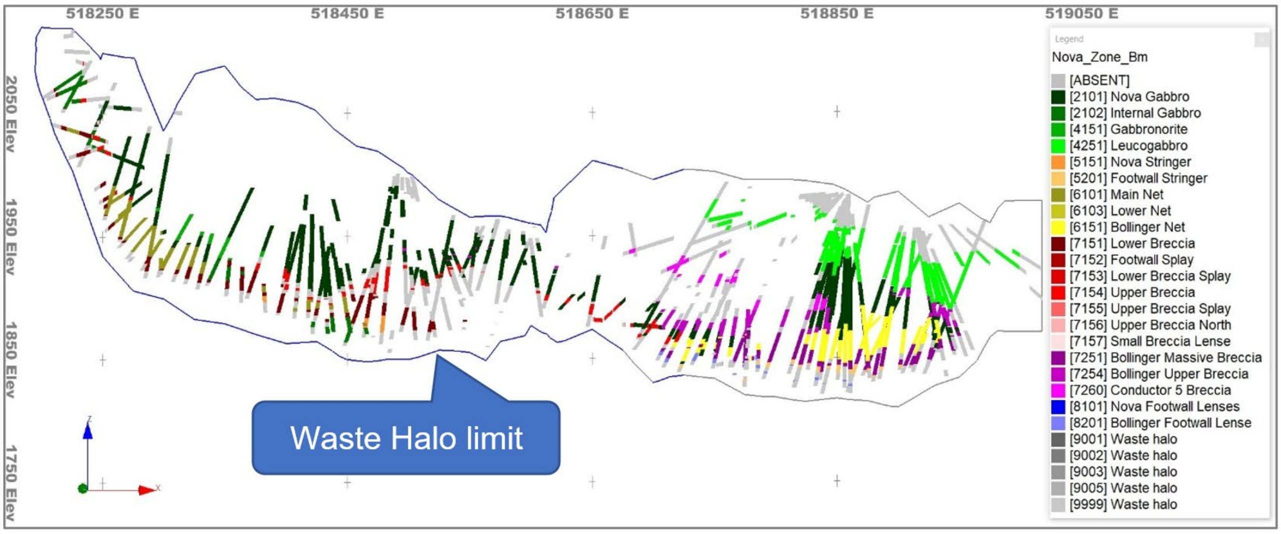


FIG 3 – Nova-Bollinger CY20 MRE long section zone-coded composites looking north on 6 479 700 mN (window ±10 m).

Whilst Nova's MGT has interpreted 22 domains for MRE estimation purposes, the samples outside the Waste Halo limit, were assigned to a single bounding outer background domain. This domain is assigned a default density and grade. As such, the CY20 MRE model is effectively a sulfide only model. No fully integrated attempt has yet been made to prepare a full geological model of deposit and its surrounds using the 99 unique lithological logging codes and the 11 codes recorded for different sulfide mineralisation styles.

MACHINE LEARNING TRIALS

With the increase in computing power in recent years, the application of ML algorithms to the challenges of mining industry resource and ore reserve estimation have progressed past the point of academic experimentation. Several algorithms have now been demonstrated to work in a practical manner on real-world problems (Dutta *et al*, 2010; Nezamolhosseini, Mojtahedzadeh and Gholamnejad, 2017; Sullivan *et al*, 2019; Kaplan and Topal, 2020).

The ML application tested by Nova's MGT in 2021, is a 'black-box' system that runs from a simple local interface, which ports the input data files to cloud servers, where the model is prepared by high performance computers. The commercial in confidence back end of the system is understood to be a deep learning neural network approach that uses the user supervised classification of the drill hole data to prepare a block model of the 3D connectivity of domains (Sullivan *et al*, 2019).

The ML tool tested at Nova only requires three input files as follows:

- A comma delimited text file containing the composite coordinates and domain coding. Optionally, sample assay data or other numeric assay-like fields can also be provided to augment the ML training process. While the domain coding needs to be present for all records, the accessory training data can contain incompletely populated variables.
- A wireframe closed volume (or alternatively upper and lower surfaces), which are used to limit the spatial extents of the block model to be created using ML.
- A block model definition that specifies parameters such as the block model origin, the total 3D spatial extents of the model space, and parent and sub block dimensions.

Nova's MGT ran three ML trials as follows:

- Test 1: The ML software was provided with the drill hole composite files used to estimate Nova-Bollinger's CY20 MRE with the 22 domain codes (as per the legend of Figure 3), provided as the ML's supervised categories. The purpose of this test was to determine the agreement (or not) of the ML model to the CY20 MRE's IM zones. The density metal accumulation service variables of six chemical elements (Ni, Cu, Co, Fe, Mg and S) were also provided to assist with the training phase of the ML algorithm.

- Test 2: The ML software was provided with a drill hole file containing the domain coded composites used for the CY20 MRE, as per Test 1. However, this data was augmented with additional lithology coded data from the drill hole information from outside the estimation limits. Six chemical variables (Ni, Cu, Co, Fe, Mg and S) were also provided to assist with the training of the ML model. The MGT's primary purpose of this test was to assess the utility of the ML-prepared geological model outside the limits of the CY20 MRE. The auxiliary purpose was to assess whether the additional lithology coded samples resulted in any material changes in domain shapes and volumes compared to the Test 1 and IM model results.
- Test 3: The ML software was provided with a drill hole file that contained only mineralisation codes and the grouped lithology used for Test 2. Specifically, this test did not use any of the domain codes used in Test 1. This test examined if a 'hands off' approach would produce a model comparable to a manually domain coded model and whether the resulting model had some utility in the MRE modelling process.

TEST 1 – CY20 MRE SULFIDE MODEL

The ML Test 1 model was prepared using the CY20 MRE's parent block size of 6 m by 6 m in the horizontal plane, with blocks 2 m high. Sub-blocks were specified to be permitted down to one quarter of the parent block size, which is akin to the CY20 MRE model specifications.

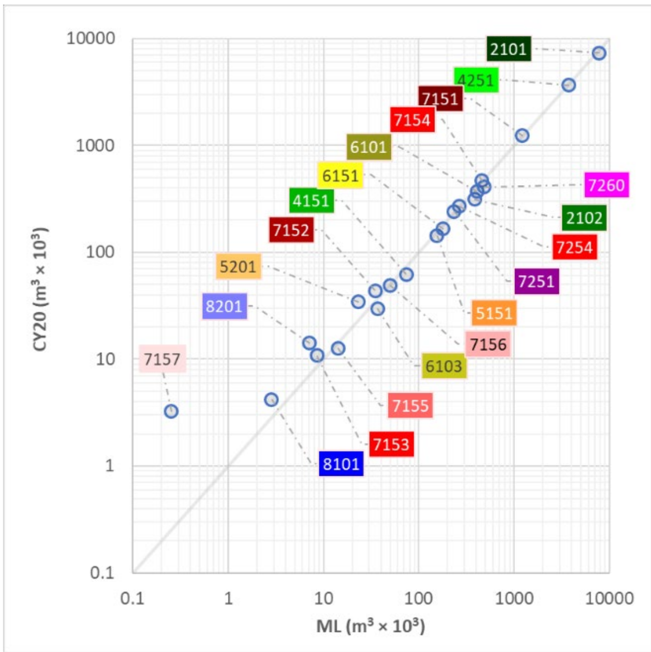
The ML process prepared a domain coded block model in its cloud process platform in 18 minutes and returned a block model of 2.4 million cells including sub blocks. The total workflow time to prepare the model was ~40 minutes, with this time including the time to set-up the input files in the correct format, the time taken to download the model, and the time to port the result into the RM software.

Volume comparisons

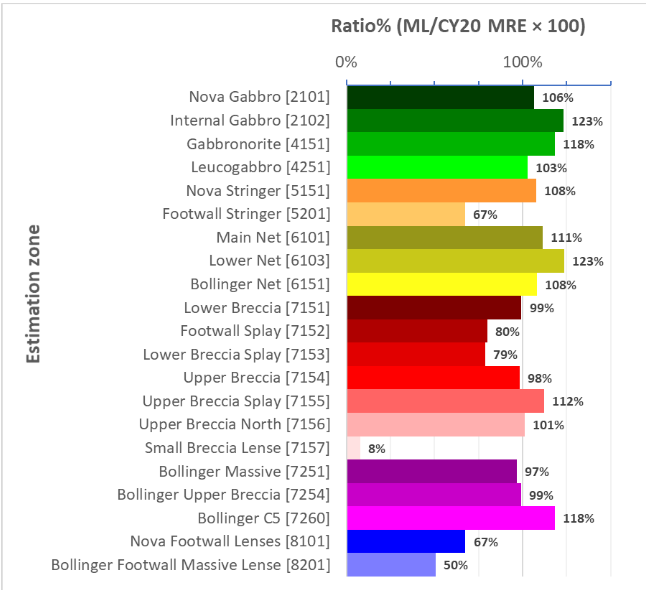
Figure 4a is a cross plot of 21 of the domain volumes from the Test 1 ML compared to the respective domain volumes from the CY20 IM modelling. The values plotted are in thousands of cubic metres. Note that the log10 horizontal axis is with respect to the ML estimated domain volumes and the log10 vertical axis is with respect to the CY20 MRE domains. The large Waste Halo domain has been excluded from this analysis as the focus is on the target mineralised domains. Figure 4b is a bar plot of the percentage ratios returned when dividing the volume of the ML model by the CY20 MRE for the respective domains.

These plots reveal that the comparative volumes of the two alternative MRE model preparation methods are generally similar for all but a few of the small domains. The most noticeable difference between the two modelling methods is the 7157 domain, where the ML approach created only 8 per cent of the CY20 volume. However as discussed further below, the contribution of this domain to the total resource is very minor and the certainty in the CY20 IM interpretation is low. The more surprising result is the larger volumed 5201 domain, where the ML process returned 67 per cent of the CY20 volume. This is also discussed further below.

In contrast, in many of the larger volumed domains, the ML approach returned larger volumes than the CY20 MRE model, especially the internal gabbros and the lower net domain. Overall, summing the volumes for all 21 estimation domains, the ML approach returned 104 per cent of the volume of the CY20 MRE. Figure 5 contains an example long section of the two models. These plots reveal some of the local differences between the modelling approaches, with the overall impression that the ML method is more conservative in terms of connectivity on the smaller volumed domains.



(a)



(b)

FIG 4 – CY20 versus ML volumes – Test 1. (a) Tonnage comparison; (b) Ratio comparison.

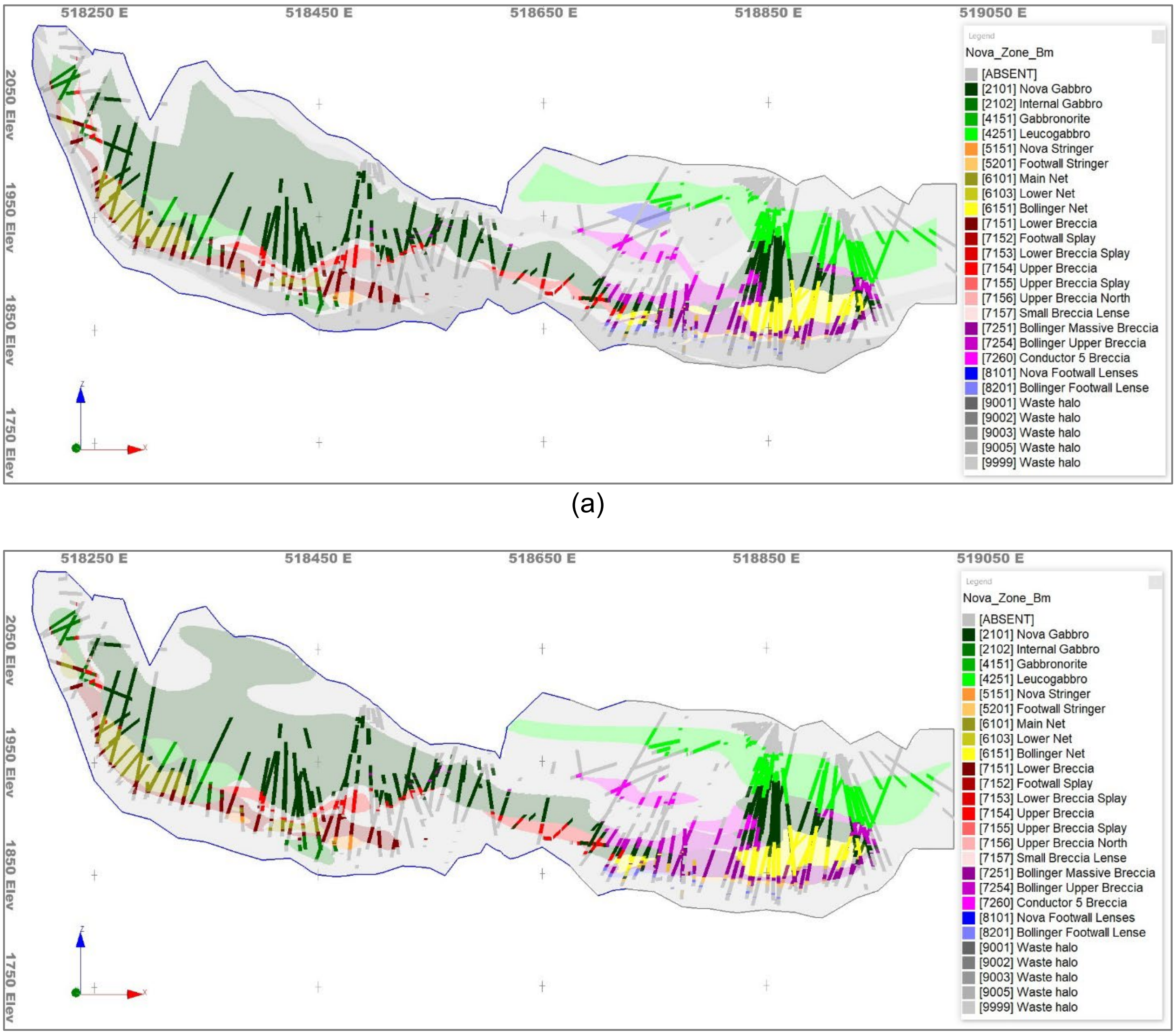
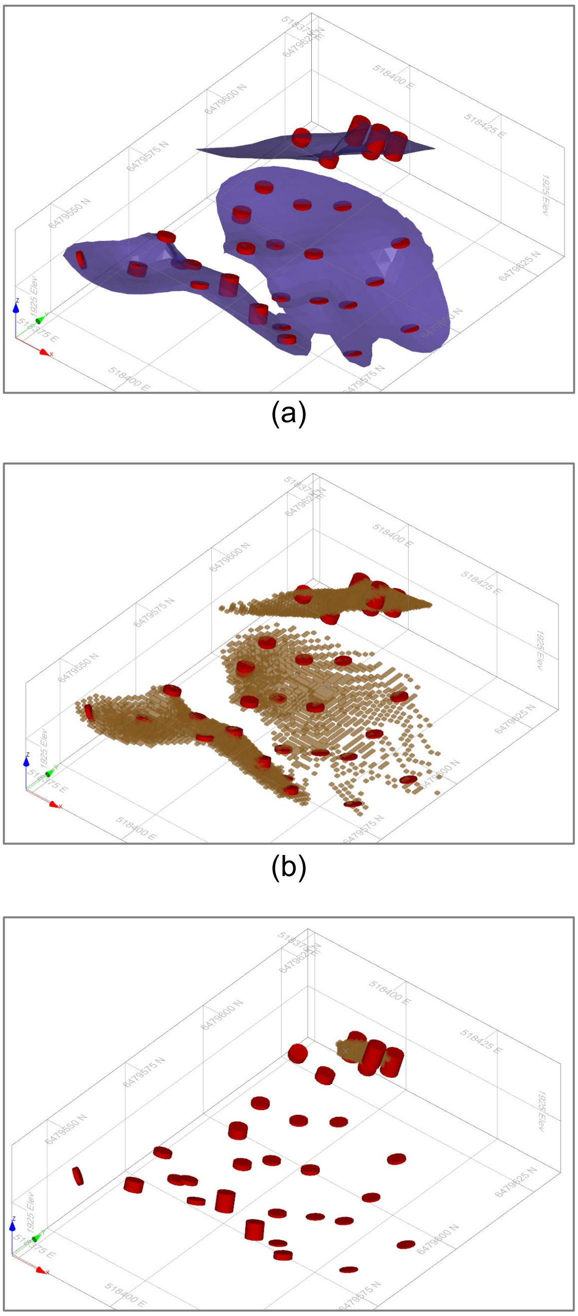


FIG 5 – Comparative IM and ML models on long section 6 479 700 mN (±10 m) looking north. (a) Model section prepared using implicit modelling for CY20 MRE; (b) Model section prepared using machine learning for Test 1.

A closer on-screen inspection of the 7157 domain volume anomaly discussed above, revealed that the differences between the ML and IM methods appear to be a function of domain geometry. Importantly, most of the intercepts defining the 7157 breccia lode structure are very thin (Figure 6a). For the CY20 MRE, where the modeller connected the 7157 domain intercepts using IM tools, the subsequent block fill was largely discontinuous, even when specifying sub blocks down to a resolution of 1 mE × 1 mN × 0.5 mElev (Figure 6b). In contrast, using the ML approach, the only zone where connectivity occurred for the 7157 domain was for a small pod around a cluster of thicker intercepts (Figure 6c).

The finding here is that, as expected, the ML approach is more conservative in domains where the data support is thin and the interpretation more tenuous. Part of the issue here is that at Nova-Bollinger, experience has encouraged modellers to attempt to constrain any narrow high-grade intercepts within broader low-grade domains. This is to help prevent interpolation 'smearing' of higher isolated grades into broader low-grade domains. As such modellers must make subjective decisions between 'lumping' data together or 'clumping' out higher grades. The 7157 domain is a case in point (as is 8101 domain), where the domain was prepared to isolate high-grades, but then it was assigned an Inferred Resource JORC Code confidence due to the understood uncertainty of the interpretation. However, this approach is problematic when Inferred Resource pods end up within domains of Measured and Indicated Resources, which are then used for ore reserve conversion. For the 7151 domain the ML result appears to make more sense in this situation and perhaps grade cutting or distance constraining the interpolation (or both) of thin isolated high-grades makes more sense.



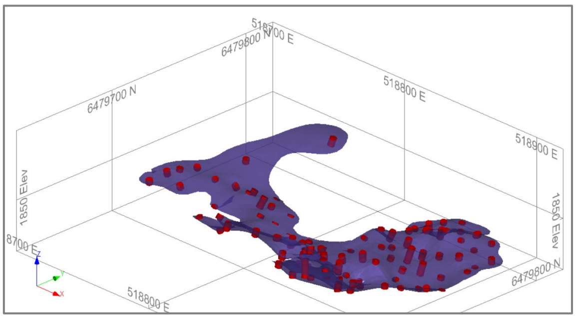
(a)

(b)

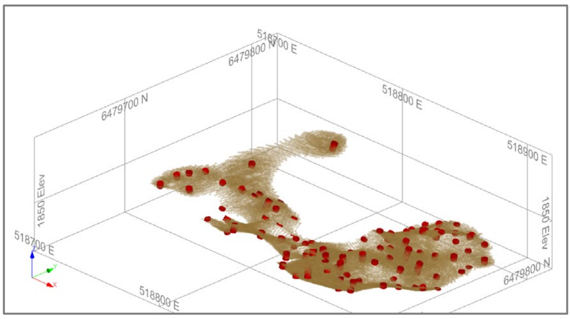
(c)

FIG 6 – 7157 domain IM wireframe and block fills – CY20 and ML Test 1. (a) CY20 MRE IM wireframe; (b) Block fill of IM wireframe; (c) Intercepts and MLA blocks.

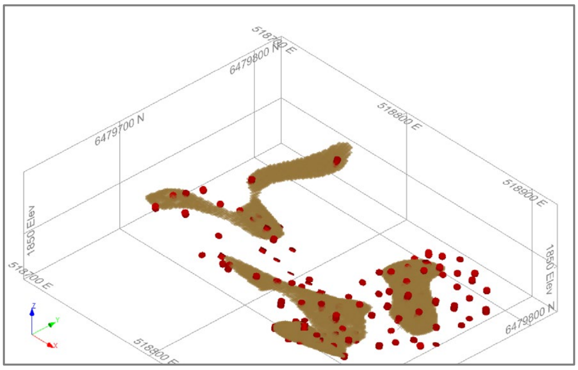
The other estimation domain with a significant volume difference between the CY20 MRE and ML models is the 5201 Footwall Stringer domain (Figure 7a). The 5201 domain has much more drilling and intercepts are generally thicker than in the 7157 domain. However, the ML output appeared to be somewhat erratic with connectivity occurring between wider spaced intercepts in some areas but no connectivity occurring in areas of closed spaced drilling intercepts (Figure 7c). However, again the 5201 domain represents only a small portion of the resource, and the ML modelling of the larger domains was generally consistent with the IM results.



(a)



(b)



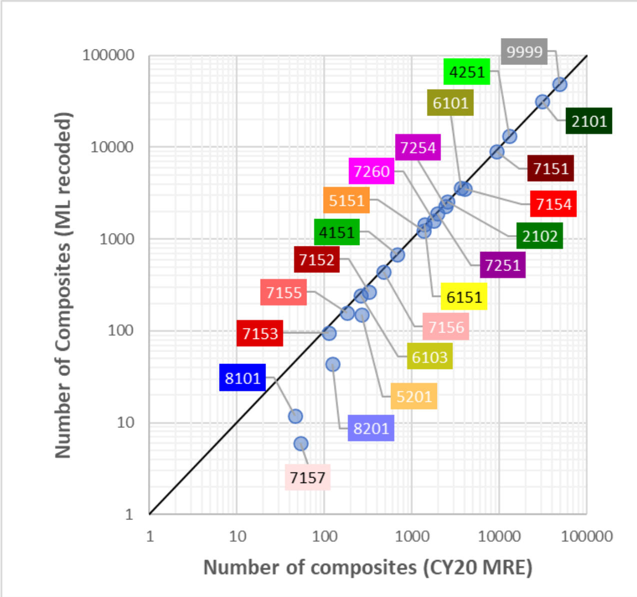
(c)

FIG 7 – 5201 domain IM wireframe and block fills – CY20 and ML Test 1. (a) CY20 MRE IM wireframe; (b) Block fill of IM wireframe; (c) Intercepts and MLA blocks.

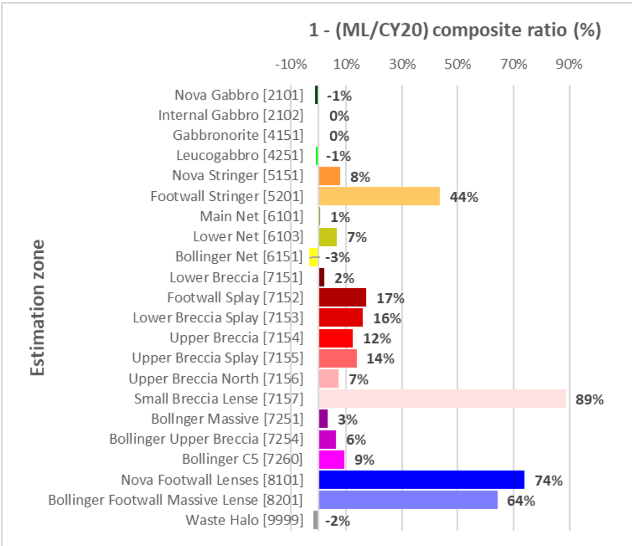
Data classification comparison

Conventional practice when dealing with 'hard-boundary' estimation domains, is to only use composites within an estimation domain to estimate the variable of interest within that domain. This was the approach applied in the Nova-Bollinger CY20 MRE. As the ML approach had defined different spatial volumes for the domains, a new ML-domain class field was added to the composite file to reflect the different ML volumes. A parallel ML domain field was coded into the drill hole file for each composite according to the location of the composite's centre relative to the block defined domain boundaries of the ML block model. The purpose of this exercise was to prepare a proxy confusion matrix for the ML results (Fawcett, 2006).

Figure 8a is a cross plot comparing the number of composites in each domain in the CY20 MRE and the ML recoded field. The horizontal axis is with respect to the CY20 MRE composite file and the vertical axis is with respect to the ML recoded file. Figure 8b is a bar plot of the ratio of the ML to CY20 MRE composites for each domain subtracted from unity, which serves as a 'reclassification index'. For this metric a positive value represents the proportion of composites reclassified to a different domain from what was originally in the CY20 MRE domain coding, and conversely, a negative value indicates the proportion of composites from a CY20 MRE domain moved to a recoded ML domain.



(a)



(b)

FIG 8 – ML/CY20 MRE composite reclassification. (a) Composite count comparison; (b) Reclassification of composites.

As expected from the volume comparison results discussed above, the numbers of composites reclassified is greatest for the small volumed domains such as 7157, 8101, and 8201. However, for the larger volumed estimation domains the reclassification of composites was relatively fewer by proportion, despite the sometimes larger volumes returned in the ML model for these domains. The hypothesised explanation here is that the margins of these larger zones often have wider spaced drilling and are spatially more distant to other domains, so a small change in the boundary position may have only a minor effect in terms of reclassification of surrounding composites.

Of interest is the moderately high reclassification index of the composites from several breccia domains, with values ranging from 7 per cent to 17 per cent in Figure 8b. If these higher grades end up in adjacent zones that are low-grade, these values will invoke problems of grade 'smearing' in the model, which may not be appropriate. As such, further testing is needed to assess the grade interpolation outcome of using the ML domained model and recoded composites. This is part of an ongoing study, and results should be available by the time this paper is published.

Estimation confidence

The ML application prepares a standard confidence metric, which is in the range zero to 100, for each block in the model. Figure 9 is a long section through the Nova-Bollinger Test 1 ML model where the model's blocks have been colour-coded according to the ML confidence inset legend and the drill holes are coded according to the domain legend. The section northing is the same as that used for Figure 5.

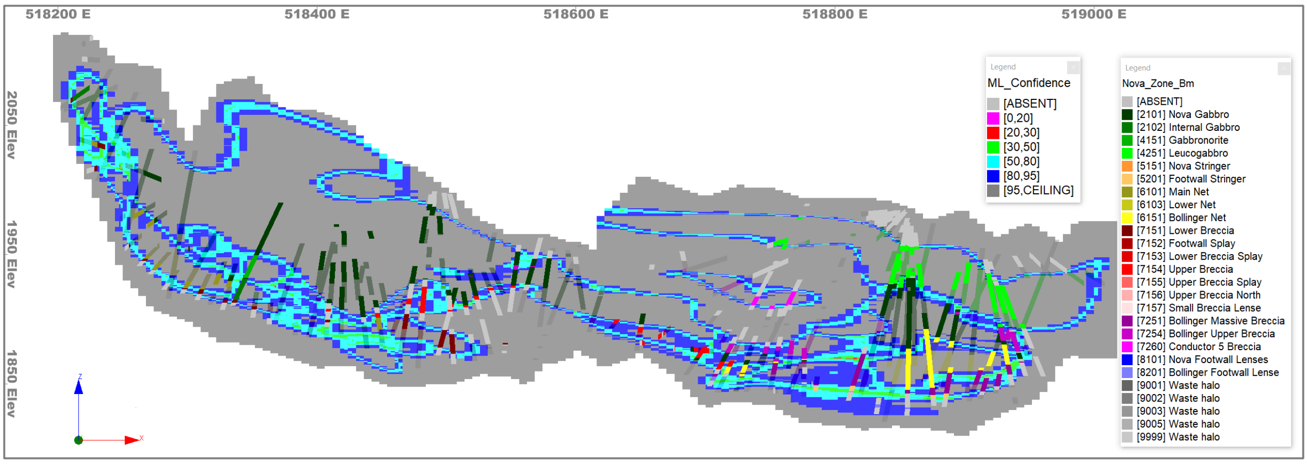


FIG 9 – Machine learning confidence metric on long section 6 479 700 mN (±10 m) looking north – Test 1.

The pattern generated by the ML's confidence metric on the long section included in Figure 5 appears to be similar to the boundary uncertainty quantification, which can be prepared through stochastic simulation or jack-knife estimation of a signed distance function for the boundaries of the modelled estimation zones (Amarante, Rolo and Costa, 2019).

The estimation confidence appears to decrease symmetrically towards a given boundary, and where the local geological complexity is high the overall confidence is lower. For example, the Nova Upper area, and the lower parts of Nova and Bollinger in Figure 9a, where there are multiple relatively thin domains have lower confidence in the boundary locations. Anecdotally, the Nova's MGT have found that these types of mining areas are more likely to be correlated with poor short-term reconciliation. This metric may assist in targeting additional drilling to increase the confidence of these domains.

Interestingly, there is a moderate log-log correlation between the volume of a domain and the ML confidence (Figure 10). This trend is consistent with the observation that domains with higher surface area to volume ratios have lower overall ML confidence as the uncertainty is often higher when approaching the interpreted boundary. The ranges and mean ML confidence values for each domain are depicted in the violin plots of Figure 11.

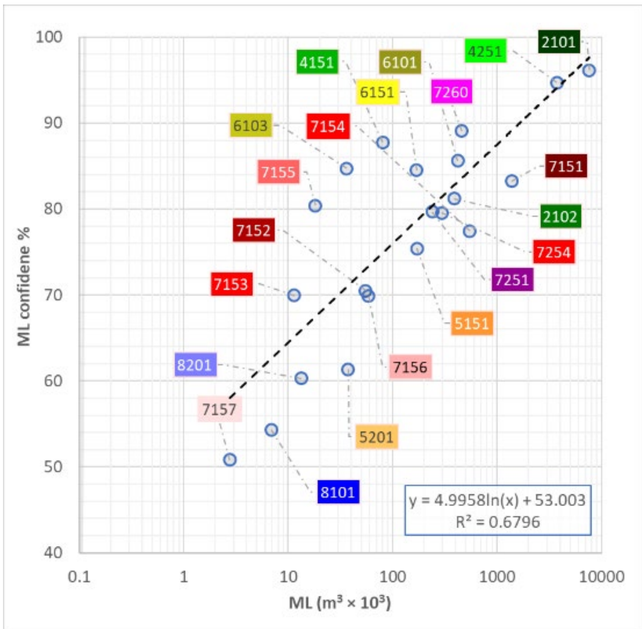


FIG 10 – ML volume versus confidence – Test 1.

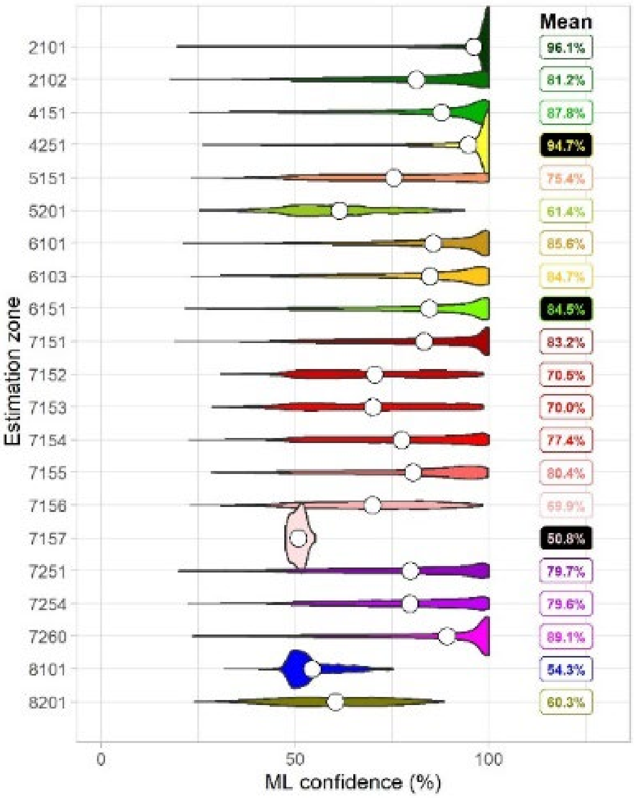


FIG 11 – ML confidence by zone – Test 1.

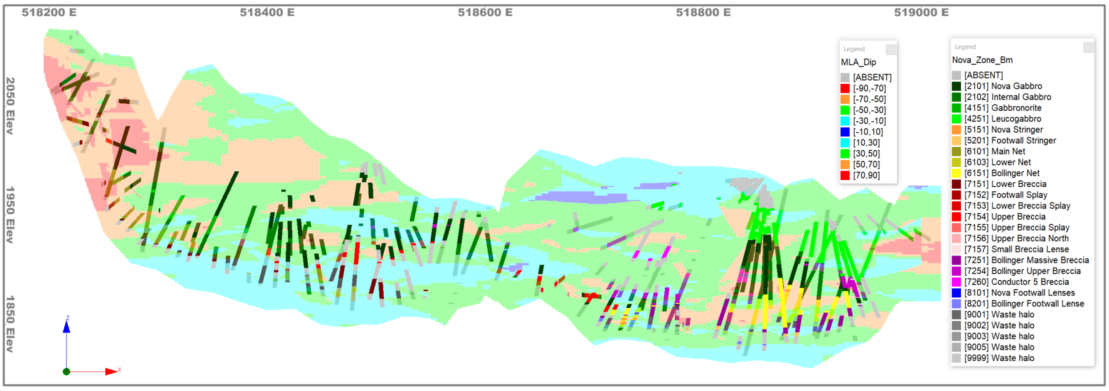
The shapes of the distributions in Figure 11, highlight the risker domains such as 7157, but also reveal the more interesting result that many of the thin breccia domains such as domains 7152 to 7156, have a large proportion of blocks with low ML confidence values. Nova's MGT plan to investigate this metric in relation to mine reconciliation results to determine its utility in forecasting reconciliation issues that are currently thought to reflect local domain complexity.

Locally varying angle field

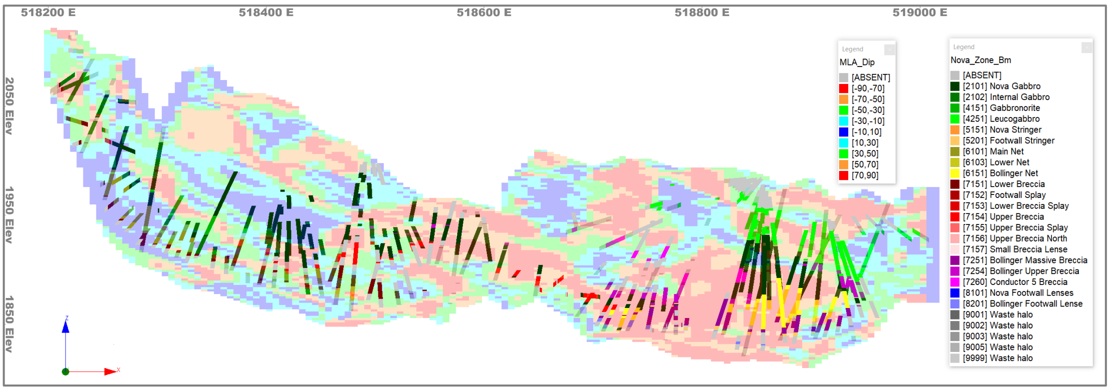
Nova's MGT prepared a locally varying angle (LVA) vector field to facilitate the CY20 MRE estimation of block grades and density using a dynamic anisotropy (DA) sample search methodology. DA requires that every block in the model have a defined plunge vector (dip and dip direction) of the major axis of the sample search ellipsoid (Machuca-Mory, Rees and Leuangthong, 2015).

The ML process also prepares a LVA field, which potentially could be used to replace the time consuming and somewhat subjective methods used to prepare the LVA field in kriging estimation. Figure 12a contains images of the LVA dip variable as interpolated from estimation wireframe boundary surfaces for the CY20 MRE, and Figure 12b is the LVA dip variable from the ML model.

The dip fields for the ML and DA methods are quite different. The DA dip field is generally less steep in the range 10° to 30°, compared to the ML equivalent blocks. Figure 12b demonstrates that the ML dip field tends to have broad areas of very steep dip (>70°), which seems counter to the MGT's understanding of the geological and grade continuity in many areas of the deposit. Nova's MGT plan to run an estimate using the ML LVA field, but initial impressions are that the field is unlikely to be a substitute for the standard DA approach. Certainly, a better understanding of how and why the ML LVA field is created would be helpful prior to additional testing work.



(a)



(b)

FIG 12 – LVA field dip on long section 6 479 700 mN (±10 m) looking north – Test 1. (a) CY20 MRE true dip; (b) ML plunge angle.

TEST 2 – SULFIDE AND FULL GEOLOGY MODEL

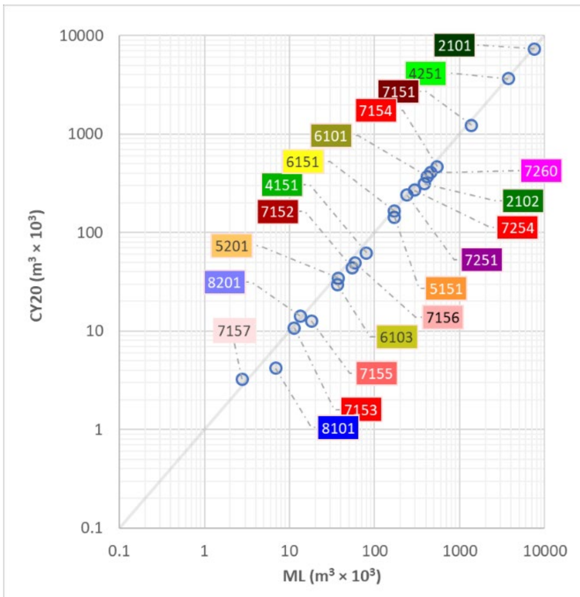
For the CY20 MRE, Nova's MGT restricted its modelling to sulfide domains within the bounds of the all-encompassing but spatially constrained 'Waste Halo' domain. The purpose of the Waste Halo was to prepare a waste dilution model with robust grade and density estimates, which would be required in the ore reserve estimation process. The model space outside of the Waste Halo was assigned to the background model.

Nova's MGT prepared the 'Test 2' model to assess how well the ML application might model the geology of the model space outside the 21 sulfide domains. Additionally, Test 2 facilitated the assessment of whether having additional information outside the sulfide domains might have any material difference to the Test 1 ML model results.

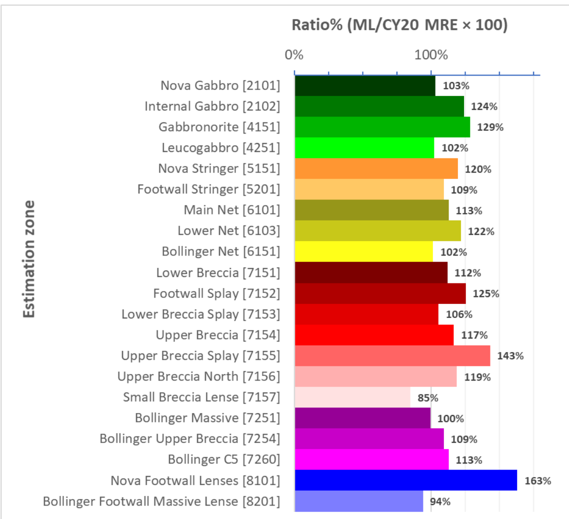
To prepare the Test 2 ML model, Nova's MGT first grouped the 120 alphanumeric codes into a more tractable 25 categories using the team's current understanding of the geology outside mineralisation. An example of this was to combine several of the internal hanging wall sedimentary logging codes that represent different mineral proportions within the same unit. For example, the unit SQQ (quartz dominated) and SQG (quartz and garnet dominated) were combined into a single 'HW Sediments' domain. Another example included combining the 15 different logging codes for felsic intrusive into a single 'Granite' domain code. These 25 categories were then combined with the sulfide model's codes and the ML model was prepared from 51 separate geology and domain codes. The ML model was also provided with the grades of six chemical variables (Ni, Cu, Co, Mg, Fe and S) to augment the ML training. The Test 2 ML model specifications in terms of model extents, dimensions of parent and subblocks, were set to be the same as used in Test 1. The total workflow time to prepare the model was ~1.5 hours.

Volume comparisons

Figure 13a is a cross plot of the volumes of the Nova-Bollinger CY20 MRE domains compared to the Test 2 ML model – akin to the analysis above prepared for Test 1 in Figure 8. Figure 13b is a bar plot of the percentage ratios returned when dividing the volume of the ML model and grouped geology model by the CY20 MRE for the respective domains.



(a)



(b)

FIG 13 – CY20 versus ML volumes – Test 2. (a) Volume comparison; (b) Ratio comparison.

The bar plot that is Figure 15, where each bar represents the percentage ratio of the respective domain volumes in Test 2 compared to Test 1, reveals that including the geology coding in the ML training, increases the volume assigned to the smaller volumed destination zones, such as the 7157 zone (Figure 14). While the change in approach did not fill the volume as much as the IM approach as depicted previously in Figure 6c, there were significantly more blocks allocated to the 7157 zone in Test 2. Additionally, the Test 2 model had smaller volumes created in Bollinger Net domain (6151) and the Bollinger C5 domain (7260) when compared to Test 1.

The addition of the grouped geology in Test 2 had a direct influence upon the ML software's ability to further define the 7157 and 5201 domains compared to the Test 1 model. Figure 14 shows the increased volume of 7157 produced with addition of the grouped geology and Figure 16 shows the increased volume of the 5201 domain compared to depicted previously in Figure 7c.

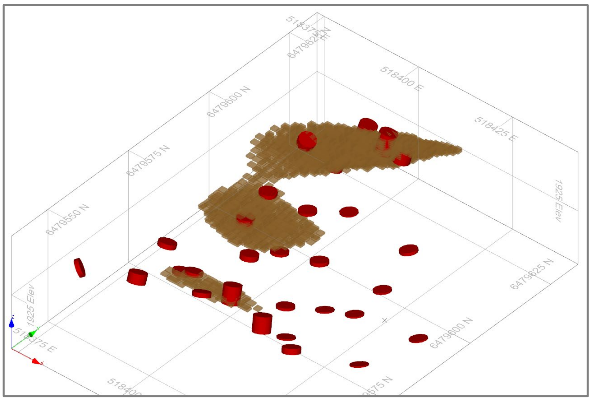


FIG 14 – 7157 domain block fills – Test 2.

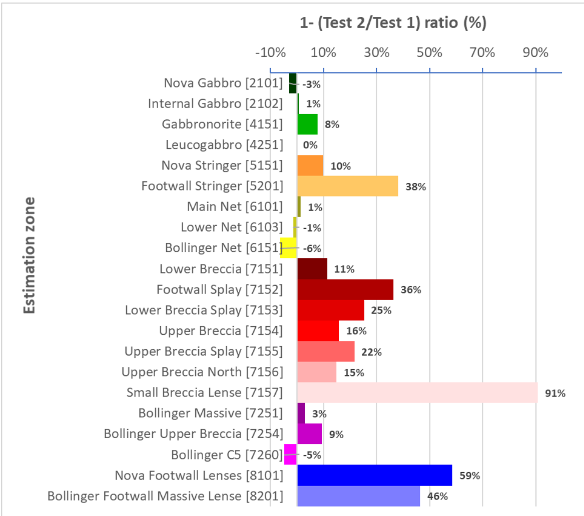


FIG 15 – ML Test 2/Test 1 percentage ratio.

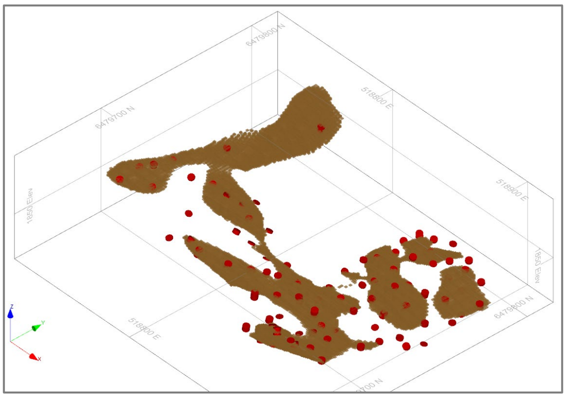


FIG 16 – 5201 domain block fills – Test 2.

Due to the black box nature of the ML software, understanding the drivers that increase the narrow lode volumes in Test 2, compared to the Test 1 results, is difficult. The reason may be that the additional samples surrounding these composited domains increase confidence in larger volumetric estimations, whilst constraining the thicker volumed domains. Nova's MGT plan to investigate this effect with further experimentation.

Test 2 – model cross-section

Figure 17, which is a cross-section through the Test 2 ML model, has a strong similarity in geometry and connectivity of the sulfide domains when compared to both the CY20 MRE and Test 1 models (as previously depicted in Figure 5). There is, however, significantly higher geological complexity modelled in the waste domains, both inside the Waste Halo limit (Figure 17), and outside this limit of the CY20 MRE model (Figure 18).

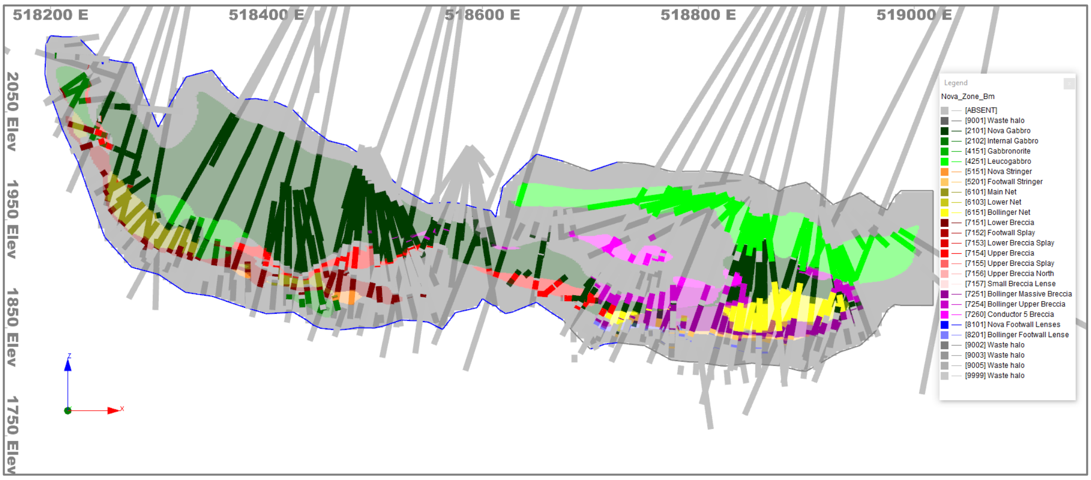


FIG 17 – Test 2 long section 6 479 700 mN (±10 m) looking north – domains only.

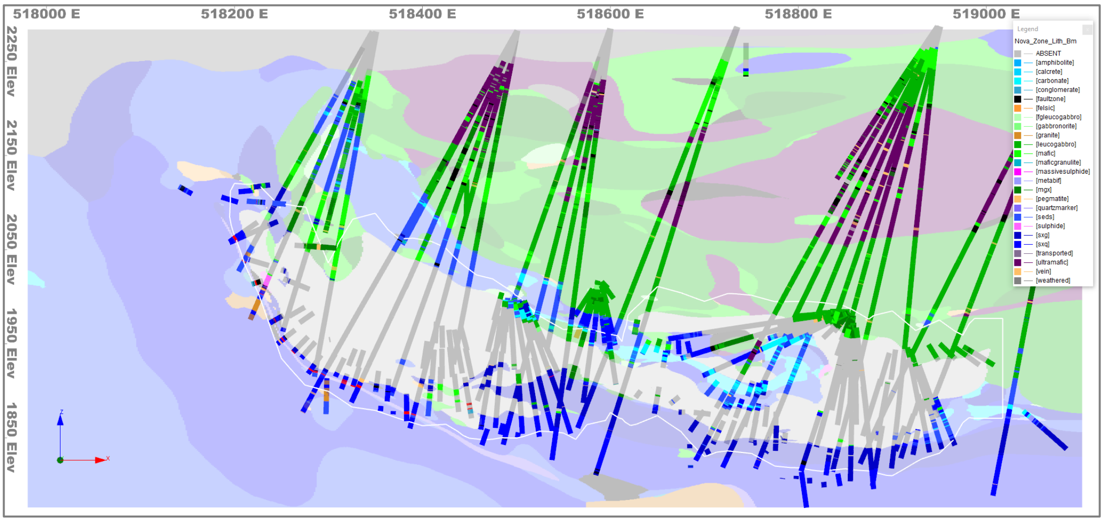


FIG 18 – Test 2 long section 6 479 700 mN (±10 m) looking north – lithology only.

When compared to a Nova-Bollinger geological interpretation prepared by one of IGO's geochemical experts as depicted in Figure 19, the Test 2 model has modelled sedimentary domains in the area to the east of C5 as opposed to gabbro domain being interpreted in the manually interpreted geological model. Test 2 modelled sediments in this area display complex folding structures, and this is supported by underground mapping observations in the area.

The Test 2 model has increased lithological complexity in the intrusive complex in the hanging wall of the deposit when compared to the geological model. This is due to the intended scale of definition the geological model was produced at (50–120 m intervals) and having used drilling with full geochemical data only.

The fault zone domain modelled in Test 2 is often discontinuous and dipping in the direction of the surrounding geology. This result disagrees with the orientations of known faults, where the faults are observed to be cross-cutting structures in face mapping. Spatially, Test 2 did however, model fault zones in similar areas to the known structures. Further investigation is required to determine how structural data could be correctly modelled using ML or if a different approach is necessary.

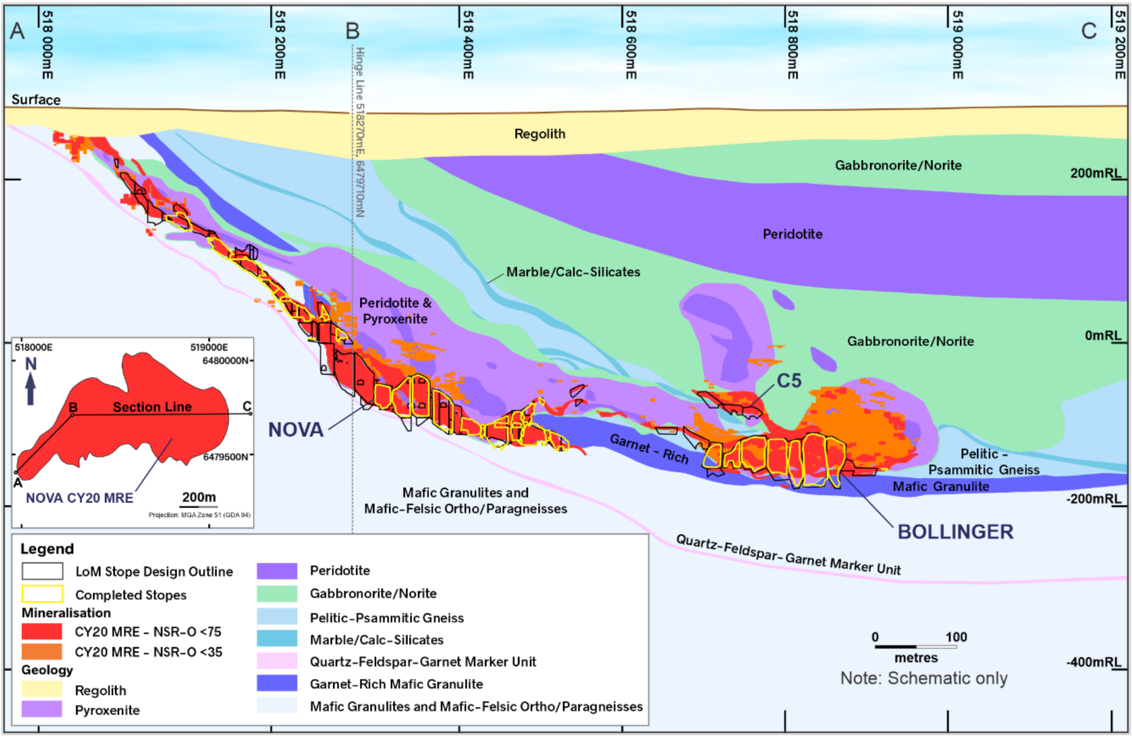


FIG 19 – Nova Bollinger CY20 MRE hinged cross-section 6 479 710 mN looking north (manual geological interpretation Ben Cave).

TEST 3 – MINERALISATION ONLY CODING

For ML Test 3, Nova's MGT provided the program with a (pre-domaining) major lithology component and assay result drill hole file. The major component lithology was designated the supervised domain variable. The goal of this test was to determine if a 'hands off' approach would produce a model that could be used for early-stage development or estimation prior to undertaking a mineralisation domain rationalisation similar to that applied for the CY20 MRE.

The mineralisation codes logged in the Nova-Bollinger drill hole database are shorthand alphanumeric abbreviations. For example, '$X' is used for net textured mineralisation and '$S' for stringer. In these cases, the codes were recast as 'net texture' and 'stringer' respectively for the purposes of the test. Many similar adjustments were made for other shorthand mineralisation codes. All other non-mineralisation lithology codes were grouped into the same 34 grouped geological codes using the same assumptions as applied in Test 2. The workflow time to prepare this model was ~1.5 hours.

As expected, removing the CY20 MRE domain coding to train the ML model, resulted in more complex relationships between the ML interpreted domains. For example, Figure 20 is a bar plot of the proportions of Test 3's ML domains inside the wireframe of the CY20 MRE 7151 Lower Breccia domain. This plot confirms that within this CY20 MRE domain, there are near equal proportions of breccia, massive and stringer mineralisation that have been grouped into breccia style, reflecting the variability of logged lithologies with corresponding grades that have been 'lumped' into the unit.

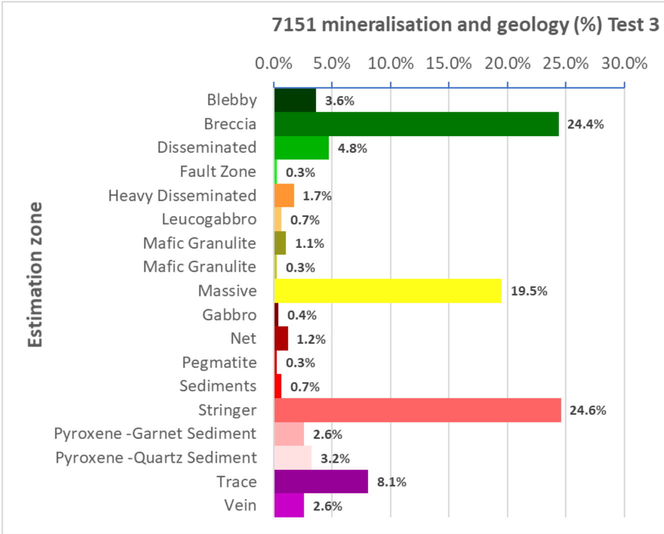


FIG 20 – Test 3 proportions of the Lower Breccia.

Test 3 – model cross-section

Figure 21 is a cross-section through the Test 3 ML model showing mineralisation and geology. Like Test 2, there is again, an increase in the complexity in the waste domains outside of the CY20 MRE Waste Halo limit compared to the Test 2 model shown in Figure 17. The mineralisation domains are also significantly more complex compared to Test 2, with some of the Test 3's mineralisation occurring outside the CY20 MRE Waste Halo limit.

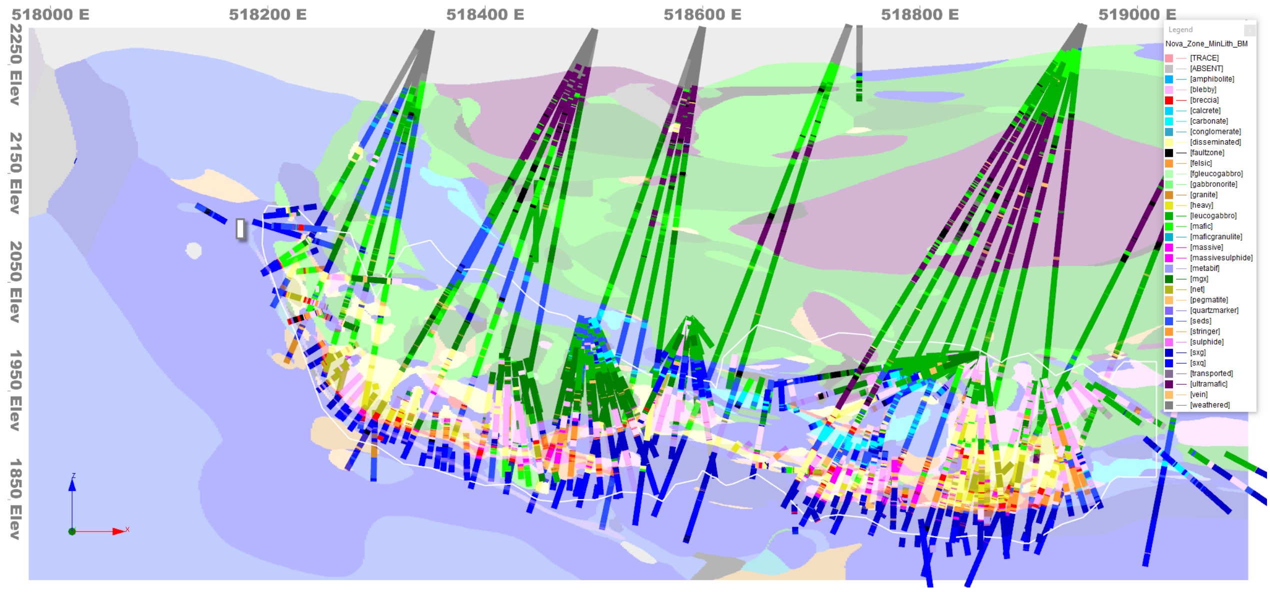


FIG 21 – Test 3 long section 6 479 700 mN (±10 m) looking north.

Compared to the Test 2 model, the Test 3 model typically has larger volumes of stringer mineralisation interpreted to occur in place of breccia domains. For example, this effect occurs for the Lower Breccia and Upper Breccia domain, which contain most of the nickel metal in the Nova Bollinger. This effect likely indicates a tendency for geologists to model stringer material into the primary breccia domains when highly mineralised or when stringer xenoliths occur near, or internal to, the breccia boundary.

The Test 3 domains also contain less breccia mineralisation compared to Test 2 in the hanging wall of Nova. Where breccia is modelled in Test 2 in the hanging wall, this is often interpreted to be gabbroic and net textured mineralisation in the Test 3 model. The net textured mineralisation in Test 3 is also more steeply dipping compared to the Test 2 model. This may be a function of the high dip angles present in the ML model's LVA field but could equally be a valid connectivity effect. More investigation is needed.

The Test 3 model also contains significant amounts of disseminated and blebby mineralisation that is not present in the Test 2 model, particularly in the hanging wall of Bollinger. The gabbroic mineralisation frequently contains disseminated very low-grade sulfides in the location where this extra mineralisation is defined in the Test 3 model. This effect is likely explained by subeconomic mineralisation not being modelled separately to the Leucogabbro in the CY20 MRE.

LEARNINGS FROM NOVA

The ML software trial has provided Nova's MGT with an opportunity to test several different hypotheses on how to model the geology and mineralisation of Nova-Bollinger. The trial has also provided an opportunity to understand if ML techniques can be applied to enhance the current MRE workflow.

Drill hole file

Moving to ML software from using manual or IM software for domaining, highlights the criticality of having a robust method of drill hole logging. Nova is a relatively young operation, with only one change of ownership since its discovery in 2012 (Bennett *et al*, 2014). However, the logging methodology and lithology coding has changed several times, which has resulted in 99 different lithologies being interpreted to occur within the CY20 MRE drill hole file. Unfortunately, multiple codes are often used for the same lithology and inaccuracies also often occur in the logged lithology. These issues highlight the problems of inadequate training, and mentoring junior staff tasked with logging drill core. As such, a robust quality control stage is needed to review the logging before MRE work can commence.

The MGT's use of IM modelling tools has allowed logging errors within the input drill hole files to be noticed and corrected during MRE preparatory work. However, this approach is time-consuming and rarely are the changes updated in the primary drill hole database due to the time-pressure to complete the estimate and subsequent MRE documentation.

Increases in the speed of modelling in ML compared to the current MRE workflow, could provide additional time for the senior and resource geologists to mentor junior staff and better review drill core logging activities. ML results, however, provide a different opportunity to pick up on logging anomalies by using cross-section checks compared to 3D checks in IM modelling.

Modelling

ML challenges the traditional methodology of modelling mineral deposits. As demonstrated in the example of Test 1, continuous domains modelled using IM sometime do not maintain the same connectivity in an ML model. However, the addition of the geology 'samples' to the domain composites in Test 2 showed that providing additional information to the ML produced greater connectivity in thinner domains. While the reasons for this improvement are not yet fully understood, the important observation is that a better model was returned by not truncating the information for the ML approach.

The Test 3 example using mineralisation codes showed that without the influence of geologist-modelled domains, the ML software could create a comprehensive spatial model of all mineralisation styles and geologies. The utility of this type of model needs more assessment and testing, particularly when considering subsequent grade and density estimation.

Test 3 also highlighted that the CY20 MRE breccia domains often contain many different types of mineralisation logged within a MRE grouped domain. Intuitively, this variability is due to the mixed nature of brecciated orebodies where a drill hole can pass through several xenoliths exhibiting different lithologies located within the same mineralised envelope.

Estimation

In an iron ore deposit, the ML software has been tuned to estimate block grades for multiple variables, in conjunction with the geological domaining, with grade estimation and geological domains both determined by an ML algorithm (Sullivan *et al*, 2019). Given the early-stage of development of the ML software at the time of the trial, calibration to estimate metal grade for a Ni-Cu-Co deposit such as Nova had not been completed. However, the MGT tested an internal process to use the domain coded Test 2 ML model blocks and apply the standard ordinary kriging CY20 MRE estimation process. The results of this first pass estimation were highly comparable to the CY20 MRE.

Hopefully, future development of the ML software will involve the calibration of the grade estimation algorithms for Ni-Cu-Co deposits and thereby remove the need for the current estimation stage. However, the authors of this paper consider that developing a clearer understanding of how these ML estimates are generated will be required before these estimations can be routinely accepted and adopted by the industry.

Future work

One of the key areas of investigation for future work is to understand how the ML software can be improved to model thin domains. The results of Test 1 raised the question whether Nova's MGT were overly optimistic with the connectivity of thin domains. The results from Test 2 showed that connectivity improved with additional drill hole information. Further work to determine how to optimise how the ML software models thin domains is required so that meaningful comparisons can be made between manual and ML interpretations.

Given the large number of lithology codes within the CY20 MRE drill hole file, grouping of lithology codes in Tests 2 and 3 was completed out of necessity to help produce a reasonable result. Future work will investigate if data conditioning can be improved and expanded to include the use of spatial or geochemical data contained within the drill hole file. This may lead to defining different groupings or domains for testing using pre-process ML analysis or rules based domaining, without the need for 3D modelling work.

The LVA angles contained within the ML models were significantly different to the DA LVA field prepared for the CY20 MRE. Future work will investigate how this effects the resultant grade models.

Advantages and disadvantages

What follows is a list of what Nova's MGT interpreted to be the current advantages of ML over IM modelling:

- The ML pay-by-use business model may be more cost-effective than maintaining IM software systems.
- The simple ML inputs can be readily prepared in most commercial general RM software systems.
- The ML modelling times are relatively short, albeit the update times are longer than updating a model for a few new holes in IM systems because the ML entire model needs to be re-run when new data is appended.
- The ML model returns an objective measure of uncertainty in the geological model, which is likely to have some utility in mineral resource classification and mining reconciliation work.
- Multiple different geological models can be prepared in parallel to assess the JORC Code Table 1 item (Section 3) as 'The effect, if any, of alternative interpretations on Mineral Resource estimation'(Joint Ore Reserves Committee, 2012).

The MGT's assessment as to the current IM advantages over ML modelling are as follows:

- The ML black-box technology can only be judged on the utility of its results. Currently there appears to be limited user scope to tune the ML's control parameters to a geologist's expert understanding. In contrast the theory and practice of IM modelling is well documented in the scientific literature and the mature technology is now appearing in most general purpose RM software systems.

- The IM method is currently better able to model thin domains with complex geometry and permits a geologist to enforce connectivity even if the confidence is low, which may be acceptable for Inferred Mineral Resource modelling.
- IM provides capability to model structural offsets, which appears to require estimations into different fault blocks in the ML process based on some preliminary tests (Maptek – Machine learning for fault identification, 2021).
- IM accommodates string data more simply, such as face mapping boundaries and permits local adjustments based on a geologist expert assessment of continuity.

CONCLUSION

Advancements in the application of ML algorithms to resource modelling, has returned focus to the fundamental importance of the accuracy of the foundational drill hole file. IM methods of modelling have provided a set of tools to help overcome errors in logging but have also resulted in them being continually retained within drill hole databases. Errors contained within the drill hole file inevitably propagate within current models and will also affect ML models. This highlights the criticality of developing robust logging processes and training junior staff in the importance of accurate logging.

Moving from an interpretive 3D domain-based process to a drill hole file data process, capable of building multiple interpretations within the same model, represents a fundamental shift in the way geologists can model a mineral deposit. However, having the ability to build multiple conditioned interpretations within ML models requires robust and rapid methods of model reconciliation to assist in testing different ML scenarios.

Nova's MGT considers that the maturity of the ML software trialled at Nova is not yet currently at a stage that allows for complete substitution of current MRE processes. Notwithstanding this conclusion, as further developments in the ML modelling capability are made, the authors consider that it is likely that ML will become the preferred modelling method for mineral deposits. ML will become particularly attractive if the process can not only model geological domains, but also return reliable grade estimates for mine planning across for the full range of mineralisation styles, while also providing a well understood confidence measure that can assist in risk quantification of both geology and grade.

ACKNOWLEDGEMENTS

The authors of the paper thank IGO for permission to publish results relating to the Nova-Bollinger deposit and Steve Sullivan at Maptek for assistance and feedback over the trials that are the basis of this paper. We also thank Ben Cave for the contribution of the manual geological interpretation of Nova-Bollinger.

REFERENCES

Amarante, F A N, Rolo, R M and Costa, J F C L, 2019. 'Assessing geologic model uncertainty – a case study comparing methods', *REM – International Engineering Journal*, 72(4), pp. 643–653. doi: 10.1590/0370–44672019720037.

Barnes, S J, Taranovic, V, Miller, J M, Boyce, G and Beresford, S, 2020. 'Sulfide emplacement and migration in the nova-bollinger Ni-Cu-Co deposit, Albany-Fraser Orogen, Western Australia', *Economic Geology*, 115(8), pp. 1749–1776. doi: 10.5382/ECONGEO.4758.

Bennett, M, Gollan, M, Staubmann, M and Bartlett, J, 2014. 'Motive, Means, and Opportunity: Key Factors in the Discovery of the Nova-Bollinger Magmatic Nickel-Copper Sulfide Deposits in Western Australia', in Building Exploration Capability for the 21st Century, pp. 301–320. doi: 10.5382/sp.18.15.

Dutta, S, Bandopadhyay, S, Ganguli, R and Misra, D, 2010. 'Machine Learning Algorithms and Their Application to Ore Reserve Estimation of Sparse and Imprecise Data', *Journal of Intelligent Learning Systems and Applications*, 02(02), pp. 86–96. doi: 10.4236/jilsa.2010.22012.

Fawcett, T, 2006. 'An introduction to ROC analysis', *Pattern Recognition Letters* 27, pp. 861–874. doi:10.1016/j.patrec.2005.10.010

Hetherington, P and Murphy, M, 2019. 'Nova-Bollinger deposit mineral resource estimation process; Mark P Murphy Resource Geology Manager-IGO Perth', in Kambalda Symposium. IGO, pp. 1–27. Available at: https://www.igo.com.au/site/PDF/8b65f5f3–163c-4e7d-81d73e89ae39d91a/KambaldaSymposiumNovaUndergroundMappingbyGlennBoyceProjectMineGeologist (Accessed: 10 August 2021).

Joint Ore Reserves Committee, 2012. The JORC Code 2012 Edition Effective 20 December 2012 and mandatory from 1 December 2013 Australasian Code for Reporting of Exploration Results, Mineral Resources and Ore Reserves Prepared by the Joint Ore Reserves Committee of The Australasian Institute. Available at: http://www.jorc.org (Accessed: 13 September 2021).

Kaplan, U E and Topal, E, 2020. 'A New Ore Grade Estimation Using Combine Machine Learning Algorithms', Minerals 2020, Vol 10, Page 847, 10(10), p. 847. doi: 10.3390/MIN10100847.

Leapfrog Geo, 2021. Available at: https://www.seequent.com/products-solutions/leapfrog-geo/ (Accessed: 3 September 2021).

Machuca-Mory, D F, Rees, H and Leuangthong, O, 2015. 'Grade Modelling with Local Anisotropy Angles: A Practical Point of View', in 37th Application of Computers and Operations Research in the Mineral Industry (APCOM 2015), Vol 1, pp. 1118–1128. doi: 10.13140/RG.2.1.2425.3281.

Maptek – DomainMCF, 2021. Available at: https://www.maptek.com/products/domainmcf/ (Accessed: 3 September 2021).

Maptek – Machine learning for fault identification, 2021. Available at: https://www.maptek.com/forge/march_2020/machine_learning_for_fault_identification.html (Accessed: 13 September 2021).

Nezamolhosseini, S A, Mojtahedzadeh, S H and Gholamnejad, J, 2017. The Application of Artificial Neural Networks to Ore Reserve Estimation at Choghart Iron Ore Deposit, *Analytical and Numerical Methods in Mining Engineering*, vol 6, pp 73–83.

R Project, 2021. Available at: https://www.r-project.org/ (Accessed: 3 September 2021).

Spaggiari, C V, Kirkland, C L, Smithies, R H and Wingate, M T D, 2014. Tectonic links between Proterozoic sedimentary cycles, basin formation and magmatism in the Albany-Fraser Orogen, Western Australia, Geological Survey of Western Australia Report 133. Available at: https://www.researchgate.net/publication/265907986_Tectonic_links_between_Proterozoic_sedimentary_cycles_basin_formation_and_magmatism_in_the_Albany-Fraser_Orogen_Western_Australia.

Studio RM, 2021. Available at: https://www.dataminesoftware.com/solutions/studio-rm-resource-modelling/ (Accessed: 3 September 2021).

Sullivan, S, Green, C, Carter, D, Sanderson, H and Batchelor, J, 2019. *Deep Learning – A New Paradigm for Orebody Modelling*. Available at: www.maptek.com (Accessed: 22 August 2021).

Supervisor, 2021. Available at: https://snowdengroup.com/supervisor/ (Accessed: 3 September 2021).