Harnessing data complexity – how machine learning applies all project data for accurate resource modelling

S Sullivan[1]

1. FAusIMM, Senior Technical Sales Specialist, Maptek Pty Ltd, Adelaide SA 5065.
   Email: steve.sullivan@maptek.com.au

ABSTRACT

A recent review of resource reports lodged to statutory bodies found geological models with oversmoothed or unrealistic lithological boundaries, many of which do not represent geological appearance in the field, pit or underground drive. The underlying data is often too complex or messy to be incorporated into the resource model and is thus a simplification of reality, to the detriment of the mining operation and its shareholders. Volumes can be over- or underestimated and subsequent grade predictions can be misplaced, combining to cause over- or understatement in resource reporting to stakeholders and statutory authorities.

Domain models can be improved with new modelling techniques such as machine learning. However, with poor data or improper use of machine learning techniques, the conclusions can be just as misleading as with current techniques. Machine learning rarely offers a single 'right' outcome, rather a range of possible outcomes from the data provided. But importantly it can take advantage of the full richness of data for more informed models. Machine learning also provides confidence in its predictions so that uncertainty in the resultant models can be quantified and passed through for risk assessment during mine planning and operation.

This paper will discuss the factors accentuating complexity in deposit modelling: data diversity, structural controls, chemistry, data volumes, process workflows and external non-geological constraints. A case study illustrates the risk of reducing or ignoring complexity, which can result in an overly simplified geological model.

INTRODUCTION

All ASX listed companies are compelled to produce geological models and ore resource statements using the guidelines established in The JORC Code (2012). The code is not prescriptive, allowing the Competent Person to use their knowledge and experience to generate relevant content for resource statements. Table 1 of the JORC Code is used as a template for ease of presentation of the key observations and assumptions used for the resource report. This covers a wide range of subjects including: database integrity; geological interpretation; estimation and modelling techniques; mining, metallurgical and environmental factors; classification; audits and reviews; and a discussion on the relative accuracy or confidence in the report.

In this paper, we concern ourselves with several aspects of the modelling and interpretation aspects of the process which include: confidence in the interpretation; the use of geology to guide and control grade estimation; factors affecting the continuity of grade and geology; generation of domains; interpolation parameters; extrapolation away from sample data; and the validation of the model against the input data.

CONSIDERING DATA

Firstly, we will briefly consider six aspects of data which need to be considered prior to interpretation and modelling – data diversity, structural controls, chemistry, data volumes, process workflows and external constraints.

Data diversity

Models are built to represent the underlying geology for use in resource estimation, reporting, geotechnical studies, geometallurgical work and mine planning. Relevant data can be sourced from a range of technologies such as field mapping, geological logging, geochemistry, geophysics, geotechnical, hyperspectral, pXRF, photography and lidar. Each source uses different data formats and provides differing levels of accuracy and relevance. Data interpreted from remotely sensed sources should be checked and calibrated against factual and observable features in the rocks themselves. As well as merging data from different backgrounds, an environment for analysis, interpretation and modelling must be able to process this complex array of disparate data.

Structural controls

Many mineralised systems have been formed by controlling structures while others have been modified since emplacement by post-mineralising events such as folding, faulting and/or shearing, sometimes of multiple generations. The sequence or hierarchy of complex events such as post-deposit faulting or cross-cutting dyke emplacement will impact geotechnical competency and mine planning studies. Understanding the structural framework is important in deposit modelling.

Chemistry

Mineral deposits are natural enrichments of elements or compounds of economic interest. Deposits can be simple in geometry but complex in mineralogy and chemical composition. This complexity can prevail in the economically important minerals where they are locked in refractory mineral species. Additionally, where deleterious elements are intimately entwined within the mineralisation their distribution also needs to be understood.

Data volumes

The increased availability and variety of sensors collecting data for geological modelling has led to a significantly greater volume of data. A decade ago, input data sizes may have been measured in megabytes or gigabytes; the current generation of core imagery from hyperspectral sensors can generate terabytes in minutes. Managing the validation, integration and usefulness of this data is a significant challenge for operations and adds to the complexity matrix.

Process workflows

Many mining companies use a lot of different software to manage and process their geological data. In the author's experience, across hundreds of mine sites, the most prolific usage observed at a single mine site was the use of eight different geological software vendors for managing data from capture to final resource model. There is increased potential for error as data transfers between software programs, which rarely share a common data model or platform. Additionally, learning multiple new software interfaces with their own method of operating can be confusing, so it requires more effort to induct new personnel into the geological team.

Complex software interactions also add an overhead for IT application management and data flow, requiring validation that the process is not broken by individual software components and/or operating system upgrades. Data security can also be compromised through the use of a diverse array of applications. Each data transfer or interaction can be a point of weakness, open to unauthorised hacking.

External constraints

In addition to technical constraints, external complexities can impact the production of geological models. Continuity of geological staff is a principal concern for many operations. Multi-generational mines require stewardship through the hiring/firing cycles that follow the highs and lows of commodity prices. Shorter lived or remote operations using fly-in fly-out staff can experience an inconsistent and distracted workforce. The use of company geologists in comparison to short-term contractors or consultants also needs to be managed to maintain consistent standards.

It is most important to maintain geological data integrity, regardless of who is collecting, recording and managing the data.

GEOLOGICAL INTERPRETATION

Geological interpretation is subject to personal decisions by the practising geologist on how particular pieces of observational data are linked in three dimensions. We can never see the entire rock mass at any time during the mining cycle. Initial interpretation may come from field mapping, where historically it was easy to draw lines around outcrop extents in solid lines and then use dashed or dotted lines between observations to indicate the uncertainty in interpretation between the known data. In three dimensions, as drilling or mining progresses, the uncertainty is extended below the surface as interpretations go from physical data points into areas of incomplete and often sparsely spaced data.

Sometimes the sample data is combined with proxies such as geophysical or chemical information to assist in the interpretation of highly weathered or altered material. Anecdotally it could be said that if you provide your data to three different geologists, you will get four different results. This theory was discussed by Sullivan, Oldham and Buchanan (2020) and then further analysed by Sullivan (2021) using a simple 2D section of five drill holes with two different rock types as shown in Figure 1.

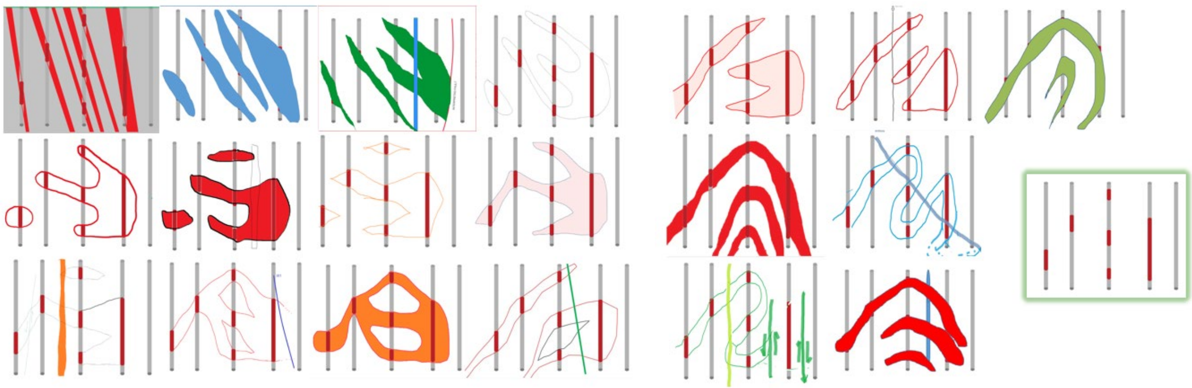

FIG 1 – Cross-sections of geological interpretations from a single hypothetical section of five drill holes (lower right), (Sullivan, 2021).

From a sample population of over a hundred participants in the interpretation challenge conducted by Sullivan there were at least 20 different outcomes, all of which were geologically possible. Several respondents declined to make any interpretation at all, indicating that there was insufficient information to make a unique geological interpretation. That is a sound judgment and justifies further expenditure to acquire more information to reduce the uncertainty to a level at which a geological interpretation would be possible.

Analysing this outcome, which was expected but not to the degree in which it manifested itself in the study, allows us to conclude that the interpretations being made are potentially a reflection on one's experience. If the interpreter has spent a lot of time working in narrow vein deposits, the mineralised intercepts will be easily joined to reflect that prior bias. A geologist highly experienced in systems with folds, may intuitively choose that route to interpret the cross-section. Recognition of our personal experience and biases is important and the use of peer and third-party review is essential, especially when working on new projects, where open cut or underground exposures do not yet exist to substantiate the geological model.

GEOLOGICAL MODELLING

Reflecting on the generation and use of models, in 1987, statisticians Box and Draper wrote:

> *'Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind...'*

And they followed this up in 1997 (Box and Luceño):

> *'It has been said that 'all models are wrong but some models are useful'. In other words, any model is at best a useful fiction—there never was, or ever will be, an exactly normal distribution or an exact linear relationship. Nevertheless, enormous progress has been made by entertaining such fictions and using them as approximations.'*

And further refined it in 2009 (Box, Luceño and del Carmen Paniagua-Quiñones):

> *'All models are approximations. Assumptions, whether implied or clearly stated, are never exactly true. All models are wrong, but some models are useful. So the question*

you need to ask is not "Is the model true?" (it never is) but "Is the model good enough for this particular application?"

So why are we quoting statisticians in a geological context? Many, if not all, of the 3D geological models we now produce are derived using underlying mathematical algorithms driven by parameters and assumptions. Kentwell (2019) summarised the progress of geological volume modelling from the days of paper-based interpretations from hand drawn sections through unconstrained block models, digitised strings and wireframes to algorithmic implicit modelling. His accompanying presentation noted that to control implicit algorithms in an attempt to approximate geological scenarios requires the use of up to 50 different parameters. Even with this over-parameterisation, accompanied by simplification of the 'warts and all' data, generating results which are credible to an external geological auditor can sometimes be challenging

An example is shown in Figure 2 whereby an implicit model is presented by Devlin *et al* (2021) in their resource report. This deposit is hosted in an ultramafic structural setting where the stress regimes generally create elongate mineralisation envelopes with steeply plunging shoots. This geological environment is highly unlikely to host spherical to lobate ore geometries as shown in Figure 2. Field mapping and core logging provide evidence to use when synthesizing the geological setting. In this type of setting, you expect to see slickensides in the field or in drill core, which are highly anisotropic in dimensions with strong shear fabric. In the author's experience, this type of microstructure provides clues (as well as red herrings) as to the macro geological setting and these should be used as input when building any geological model to produce credible results, as shown in Figure 2.

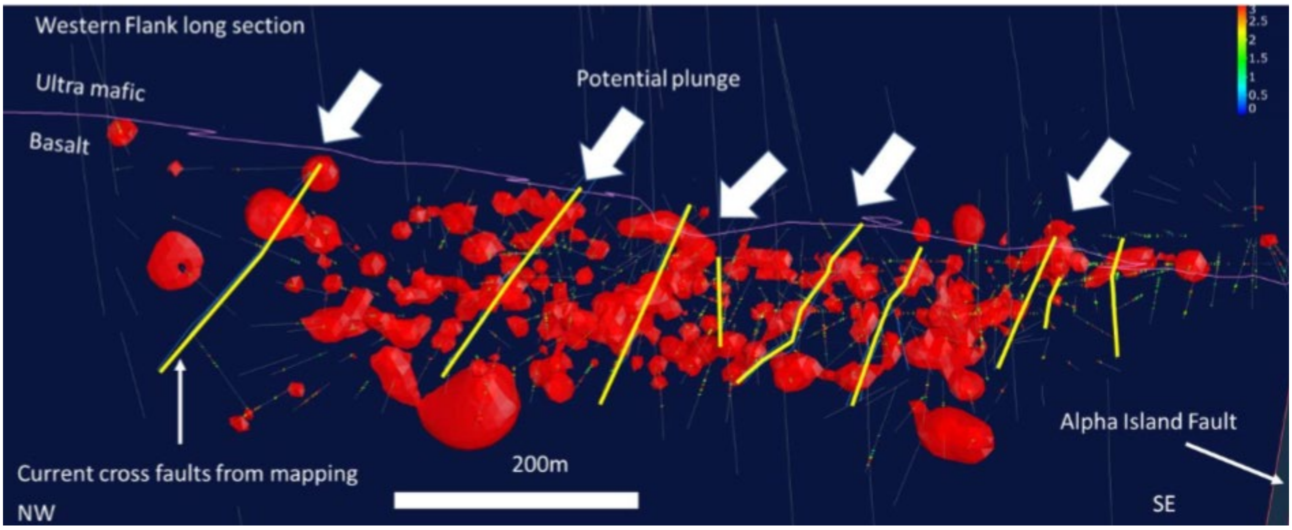

FIG 2 – Potential ore shoot geometry of the Western Flank gold mineralisation at Beta Hunt in Western Australia (Devlin *et al*, 2021).

In summary it is important to ensure that your computer-generated model(s) firstly, honour(s) your input data and secondly, represents your best understanding of the mineralising system and its host setting.

CASE HISTORY

Introduction

The Lisheen mine in County Tipperary was mined for 17 years, from 1999 until mine closure in December 2015. It produced 22.4 Mt at 11.63 wt% Zn and 1.96 wt% Pb (Torremans *et al*, 2018). Mineralisation at Lisheen consists of several, largely stratiform, massive sulfide orebodies at or within thirty metres of Lower Carboniferous Waulsortian transgressive marine carbonates and breccias. The mineralisation is manifest by semi-massive, disseminated, and vein-hosted sulfides in a complex of laterally discontinuous normal faults (Hitzman, 1999). The deposit comprised six distinct orebodies as shown in Figure 3 after Kyne *et al* (2019). These orebodies are referred to as Main Zone (MZ, MZW, MWE), Derryville, Island Zone, and Bog Zone West, Bog Zone Central, and Bog Zone East. E-W fault zones shown in red and labelled using abbreviations, eg DF refers to the Derryville Fault.

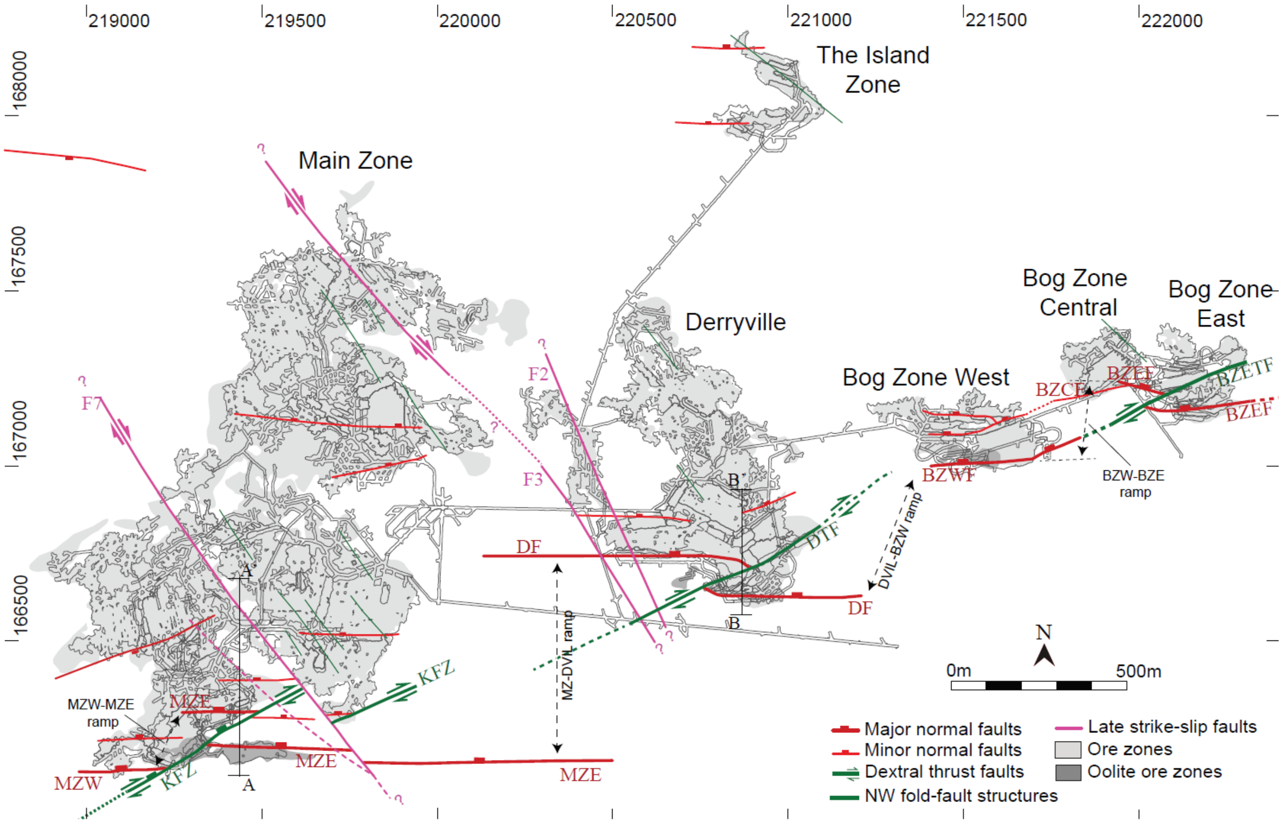

FIG 3 – Location of the Derryville South orebody in context with neighbouring orebodies at Lisheen (after Kyne *et al*, 2019).

A comprehensive 3D synthesis of the Lisheen orebodies was put forward by Kyne *et al* (2019) using a combination of traditional sectional interpretation, wireframing and fault interpolation to define geometric and kinematic links between faulting, structure and mineralisation. This paper documents building an analogous 3D model using the latest commercial machine learning engine, Maptek DomainMCF. For the initial project, one of the orebodies, Derryville South was selected for modelling, analysis and comparison with previous work.

Data analysis

The Derryville South deposit contains 1149 published holes, mostly drilled from surface during exploration and project evaluation and subsequently added to during production with underground delineation drilling. Holes were geologically logged with 34 different codes being used to describe the host rocks and mineralised intervals. The drill data was imported from csv format into Maptek Vulcan GeologyCore software for investigation and analysis. The frequency of the most common lithologies is shown in Figure 4.

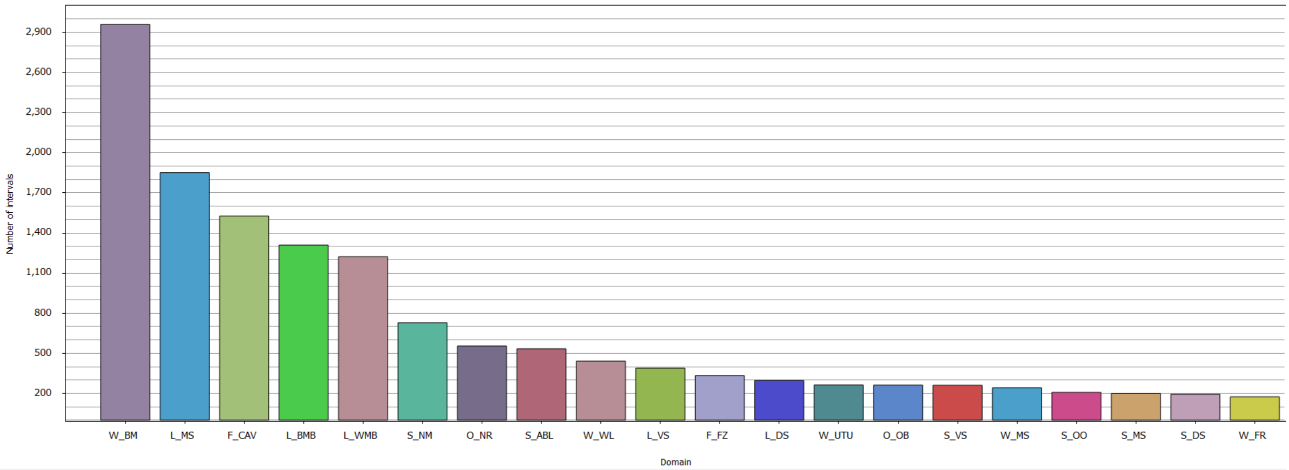

FIG 4 – Derryville South domain frequency from combined exploration and production drilling.

Chemical analysis of mineralised intervals and surrounding host rocks includes determinations for Zn, Pb, Fe, As, Mg, Mn, Ag, Cu, Cd, Hg, Ba, Co, Ni and Tl. Specific gravity measurements were also performed on selected samples. This analytical data identified zinc-lead mineralisation in almost half the logged geological codes, with the majority of economic value lying in the MSA (LM_S) massive sulfide unit as shown in Figure 5.

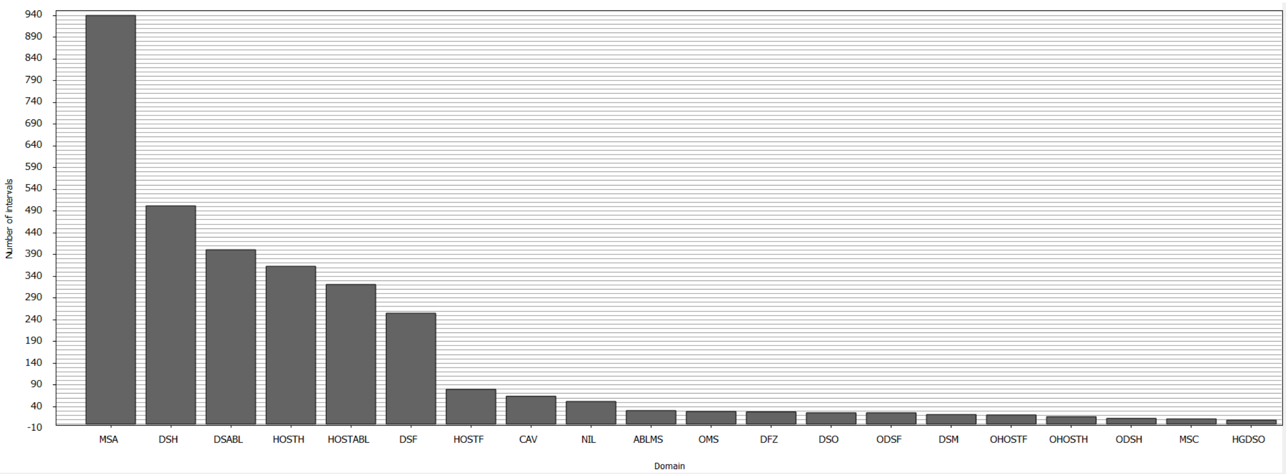

FIG 5 – Derryville South mineralised interval frequency per domain from exploration and production drilling.

Geological modelling

Previous modelling

Modelling during the operation of the Derryville mine used sectional interpretation of exploration and production drilling and underground channel sampling/mapping observations. Explicit wireframes were created for the footwall and hanging wall of each mineralised horizon. These two surfaces were then combined to make a solid model of each horizon, which was then used to flag an ore code into a 3D block model as shown in Figure 6.

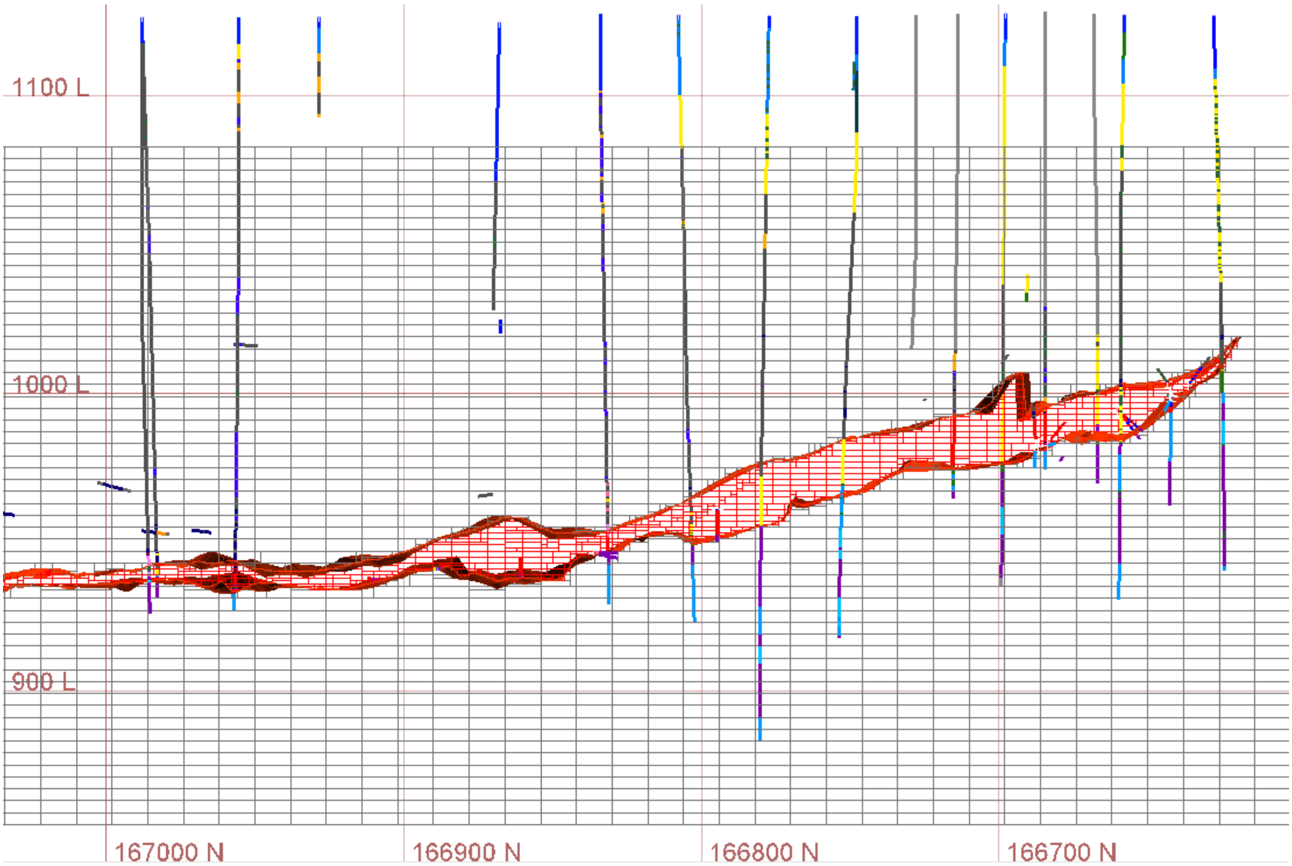

FIG 6 – Section on 220810E displaying Derryville South zinc-lead mineralisation wireframe and consequent domain coding into the mine production block model.

The Derryville mine geologists would update their resource model on a 6 or 12 month cycle. Interpretation and correlation between drill holes on a sectional basis and update of wireframes was a manual process which would take at least a week (Colin Badenhorst, personal communication). The comprehensive 3D modelling and structural synthesis by Kyne *et al* (2019), did not document the time involved in building their models. A plan view from their work is shown in Figure 7.

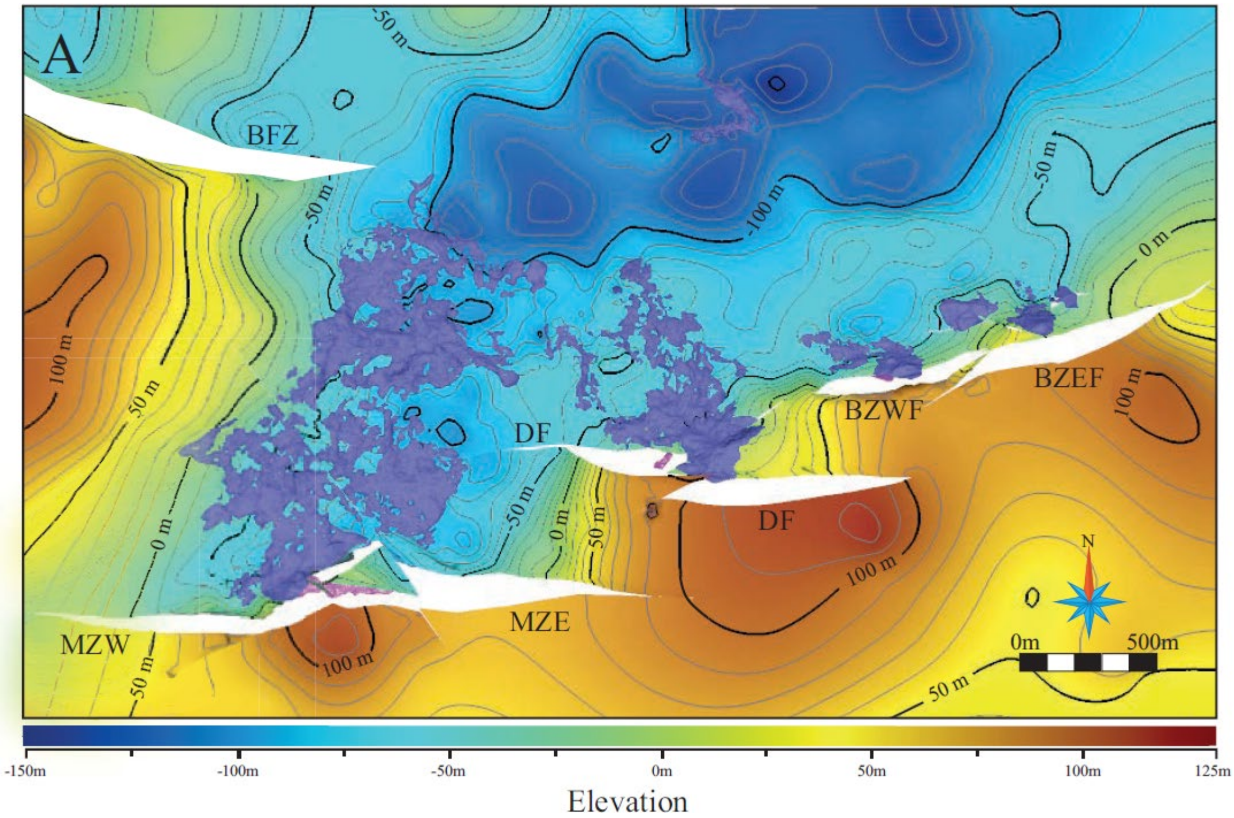

FIG 7 – Plan view of the 3D model generated by Kyne *et al* (2019) with the Derryville deposit (DF) in the centre of the screen. The white E-W trending shapes are interpreted faults and are equivalent to the faults marked and annotated in red in Figure 3. The contours represent the elevation above sea level (which is equivalent to the 1000 m Z value on the mine grid displayed on the cross-sections in Figure 6 and Figures 8 to 13).

Modelling using machine learning

The origin of the concept of data driven modelling is discussed in detail by Solomatine, See and Abrahart (2009) and summarised in Montánsa *et al* (2019). The rapid advance of computing power facilitates computer-based discovery and analysis. In the geological context this allows us to generate a documented workflow which updates the model when the input data changes. This could be a manual process but the most benefit is derived from an automated system which is triggered by a change in state in the underlying data.

The Derryville South mine data was processed using the commercially available DomainMCF machine learning engine that was launched from collaborative research and development with the mining industry as described in Sullivan *et al* (2019). The DomainMCF system processes the input geological data using multi-threaded cloud computing with cyber secure data transfer for data upload and model download.

Before data is uploaded, it needs to be validated. There are hosts of simple rules which can be applied in this process: checking the validity of collar x, y, z coordinates; checking *from and to* sample intervals are not missing or overlapping; and checking that analytical data uses the same units of measurement. Data validation rules are inbuilt into DomainMCF so that data going through to the modelling process is as clean as required. This process is critical as the output models will be directly impacted by erroneous input data.

In the Derryville South data, several drill holes were pre-collared through the overburden and were not geologically logged. When composited, these intervals are given the default code of NONE. If unlogged intervals remain in the data driven process, a domain model will be generated for NONE as per Figure 8.

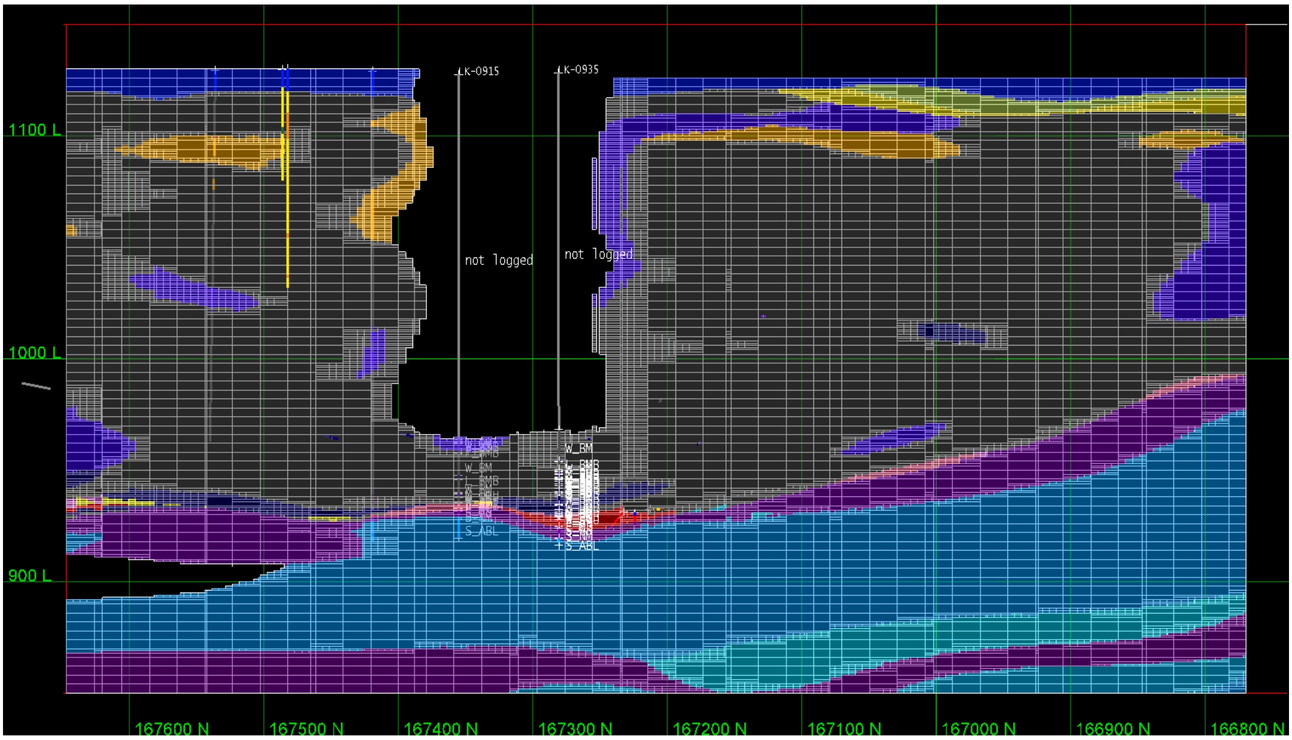

FIG 8 – A cross-section through Derryville South overburden showing the impact of leaving unlogged intervals in the input data to machine learning.

To remedy the unlogged intervals, the geologist can either manually enter appropriate logging codes based on surrounding drill holes, or preferably, delete all unlogged intervals in the input data and allow the machine learning to predict and interpolate relevant domains into the unlogged space as shown in Figure 9.

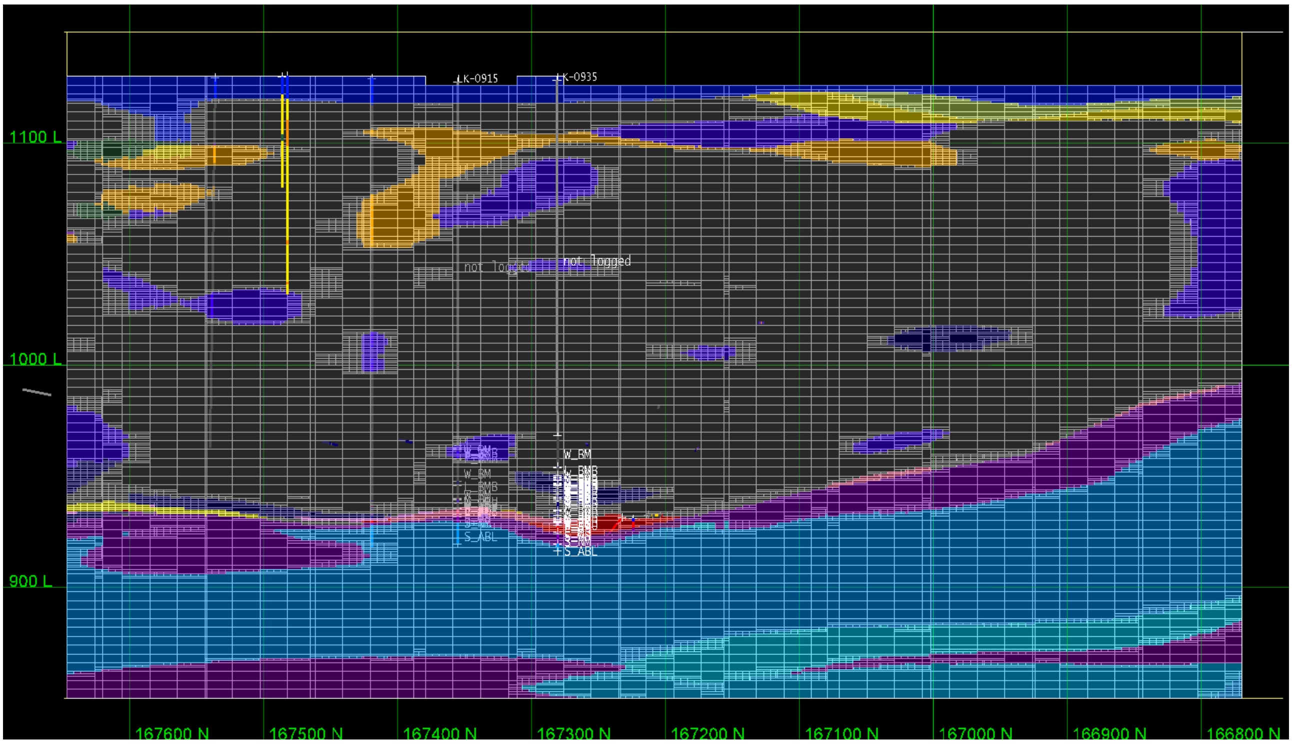

FIG 9 – A cross-section of Derryville South overburden showing the machine learning prediction for domain codes where pre-collars of surface drill holes were not logged.

A second observation from modelling straight from the geological codes is the use of the code W_MS to denote the mineralised horizon in some of the exploration holes and the code L_MS in others.

These are shown in yellow and red respectively in Figure 10. The underground production data exclusively uses the code L_MS.

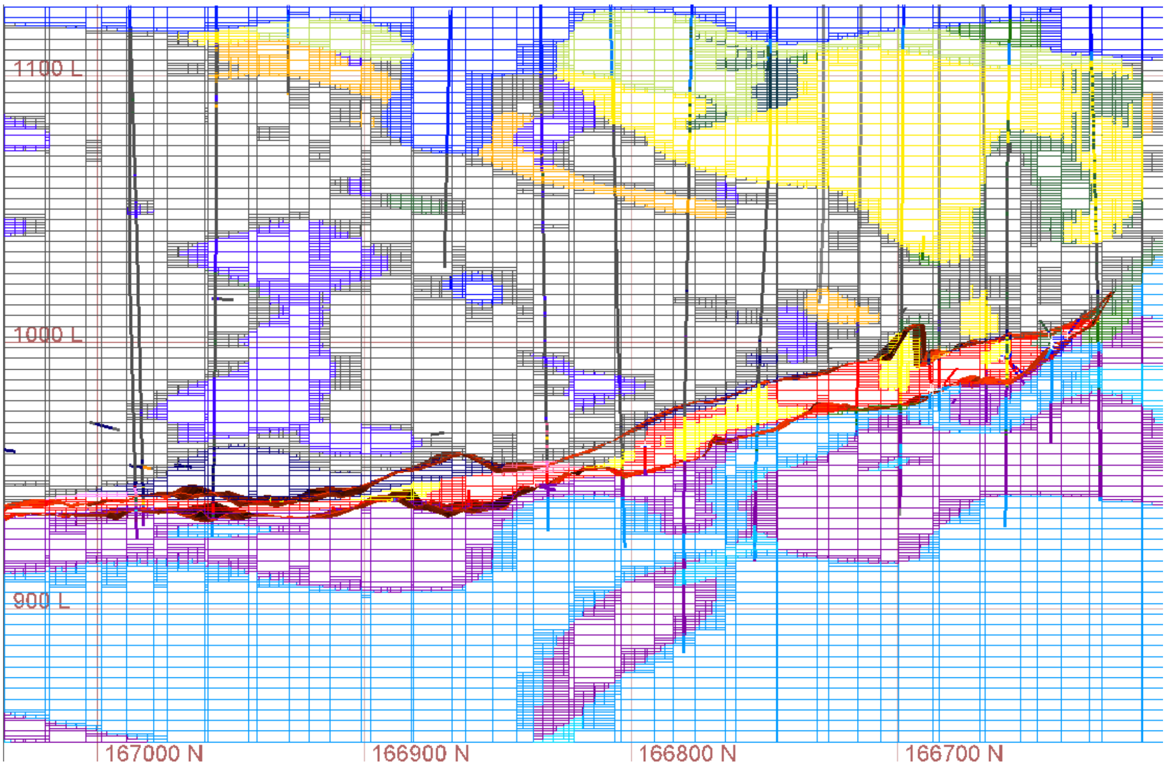

FIG 10 – Section on 220810E displaying Derryville zinc-lead mineralisation modelled using DomainMCF and the explicit wireframe from mine production used for context. Note the alternating codes of domain W_MS (yellow blocks) and domain L_MS (red blocks).

A new field was added to the drill hole database and the W_MS and L_MS domain codes were unified to a single L_MS domain code for remodelling as shown in Figure 11.

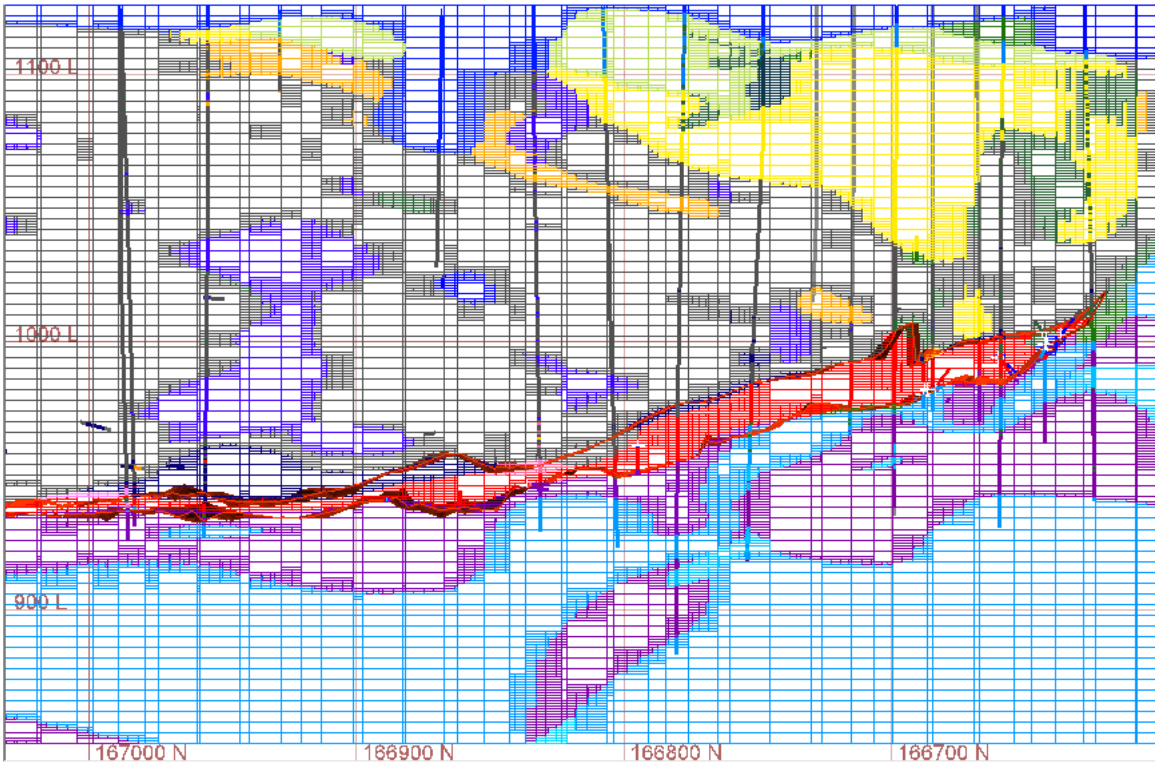

FIG 11 – Section on 220810E displaying unified Derryville zinc-lead mineralisation modelled using DomainMCF and the explicit wireframe from mine production used for context.

Once a model has been created it is important to compare it with the input data. For domain codes this is most easily accomplished by viewing cross-sections in two perpendicular orientations and also in plan view. Data driven modelling can be brutal as it will honour intervals of inliers and outliers in the data, which otherwise would be glossed or smoothed over during manual interpretation. Examples of this are seen in Figures 12 and 13, whereby internal waste intervals in the mineralised zone are encompassed by the manual wireframes resulting in incorrect flagging of domain codes for subsequent grade estimation.

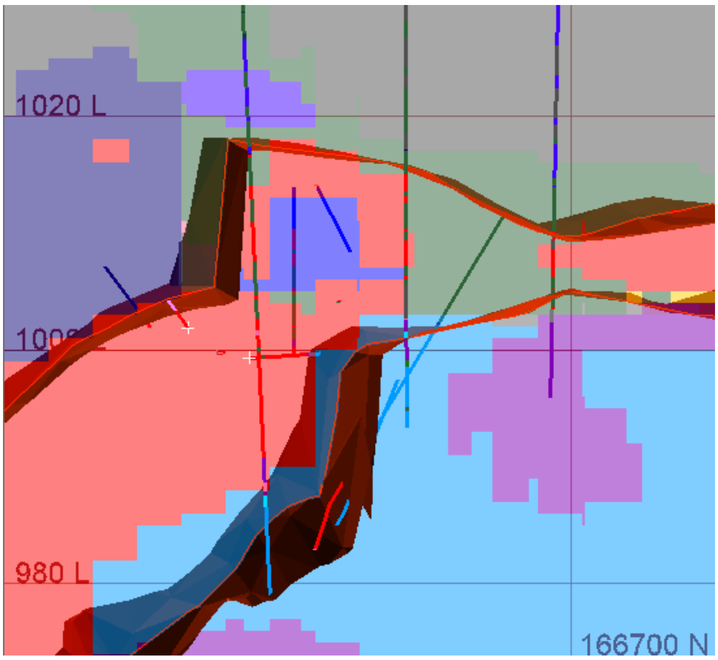

FIG 12 – Cross-section on 220892E showing DomainMCF block domain codes in the background with drill data and the manual interpretation in red wireframes. The mineralisation is coloured red in the drill holes and red in the block model slice. The manual wireframes have ignored the internal waste bands (shown in green and medium blue between 1000 and 1200RL).

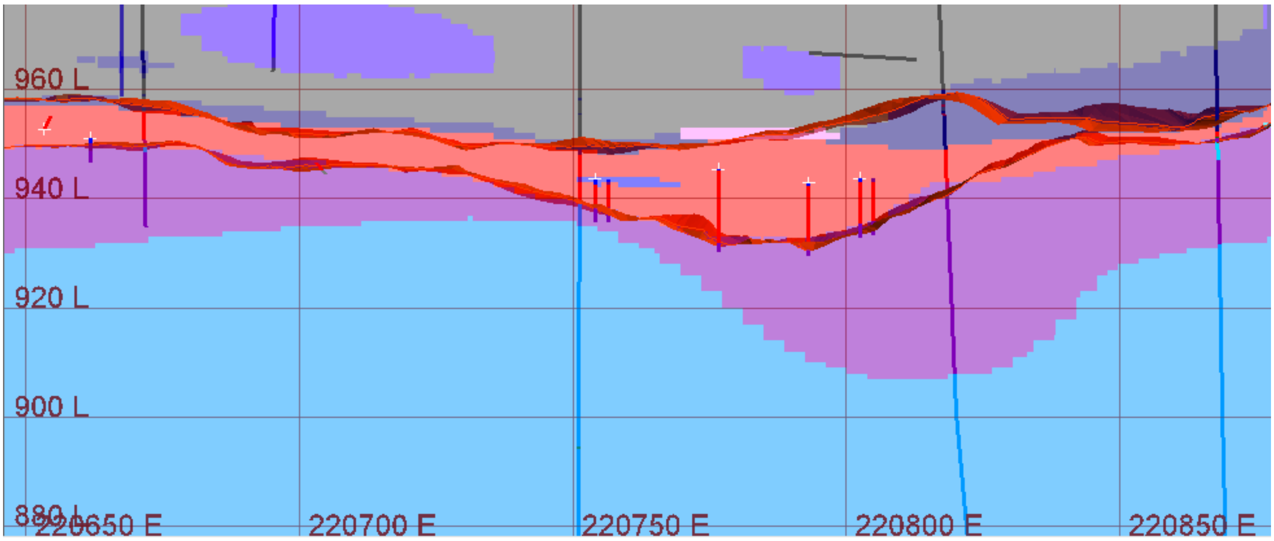

FIG 13 – Cross-section on 166870N showing DomainMCF block domain codes in the background with drill data and the manual interpretation in red wireframes, which ignores the internal waste bands (shown in light and dark blue). Mineralisation is shown in red in both the drill hole traces and the block model slices.

A comprehensive review across all sections and plans shows that the DomainMCF data driven model honours the mineralised data and its surrounding host rocks better than the manual interpretation and consequent wireframe models. The data driven model is providing finer discrimination of ore and waste domains, resulting in more relevant volumetric reports for resource statements.

DISCUSSION

The Derryville mine data has been subjected to a data driven modelling process and the resultant models appear superior in mapping the input data compared with a manually controlled process. The correlation between drill data has been derived purely from the input data without bias from the operator. Another advantage of adopting the data driven modelling approach at Derryville is the speed at which results are generated. The modelling of all mineralised and host rock domains from the raw input data occurred in less than 30 minutes. As this is now an automated process, as new data is collected, a model update can be triggered by a change in the data state and the model can always be up to date, no matter the speed or frequency of data collection.

CONCLUSIONS

Data driven geological modelling has not yet been adopted by mainstream mine and resource geologists but as the benefits of its use become widespread, we will look back and wonder why it took so long to embed into our daily processes. Machine learning is not just for traditional mines, as shown in Kapageridis *et al* (2021), which discussed the application of machine learning for domain classification in the quarrying environment. The new approach will become ubiquitous in our world, just as mobile phones and the internet have changed the way we work.

ACKNOWLEDGEMENTS

The author acknowledges the Exploration and Mining Division of Minerals Ireland for making the Derryville South mine data available for use in this paper. The author is also appreciative of the input from internal peer review within Maptek.

REFERENCES

Box, G E P and Draper, N R, 1987. *Empirical Model-Building and Response Surfaces*, John Wiley and Sons.

Box, G E P and Luceño, A, 1997. *Statistical Control: By Monitoring and Feedback Adjustment*, John Wiley and Sons.

Box, G E P, Luceño, A and del Carmen Paniagua-Quiñones, M, 2009. *Statistical Control By Monitoring and Adjustment*, John Wiley and Sons.

Devlin, S, McLeay, S, Wielligh, A, Glacken, I and Cheyne, R, 2021. Karora Resources NI 43–101 Technical Report, Higginsville-Beta Hunt Operation, (Western Australia) 2021 to the Canadian Securities Administrators, 523 p.

Hitzman, M W, 1999. Extensional faults that localize Irish syndiagenetic Zn-Pb deposits and their reactivation during Variscan compression. Geological Society, London, Special Publications, v 155, p 233–245.

JORC, 2012. Australasian Code for Reporting of Exploration Results, Mineral Resources and Ore Reserves (The JORC Code) [online]. Available from: <http://www.jorc.org> (The Joint Ore Reserves Committee of The Australasian Institute of Mining and Metallurgy, Australian Institute of Geoscientists and Minerals Council of Australia).

Kapageridis, I, Albanopoulos, C, Sullivan, S, Buchanan, G and Gialamas, E, 2021. Application of Machine Learning to Resource Modelling of a Marble Quarry with Domain, *MCF International Conference on Raw Materials and Circular Economy (RawMat2021)*, Athens, Greece, 5 to 9 September, 2021.

Kentwell, D J, 2019. Destroying the distinction between explicit and implicit geological modelling. *Proceedings International Mining Geology Conference 2019*, p 154–160 (The Australasian Institute of Mining and Metallurgy: Melbourne).

Kyne, R, Torremans, K, Guven, J, Doyle, R and Walsh, J, 2019. 3-D Modeling of the Lisheen and Silvermines Deposits, County Tipperary, Ireland. *Insights into Structural Controls on the Formation of Irish Zn-Pb Deposits Economic Geology*, v 114, no 1, p 93–116.

Montánsa, F, Chinesta, F, Gómez-Bombarelli, R and Kutz, J, 2019. Data-driven modeling and learning in science and engineering. *Comptes Rendus Mecanique*, vol 347, p 845–855.

Solomatine, D, See, L and Abrahart, R, 2009. Data-Driven Modelling: Concepts, Approaches and Experiences. In: R J Abrahart, L M See and D P Solomatine (eds), *Practical Hydroinformatics Water Science and Technology Library*, vol 68. Springer, Berlin, Heidelberg https://doi.org/10.1007/978–3-540–79881–1_2

Sullivan, S, 2021. Recognising the impact of uncertainty in resource models, in *Proceedings of AEGC's 3rd Australasian Exploration Geoscience Virtual Conference*, 15–17th September 2021.

Sullivan, S, Green, C, Carter, D, Sanderson, H and Batchelor, J, 2019. Deep Learning – A New Paradigm for Orebody Modelling, *Proceedings International Mining Geology Conference 2019* (The Australasian Institute of Mining and Metallurgy: Melbourne).

Sullivan, S, Oldham, J and Buchanan, G, 2020. Towards Greater Certainty in Resource Modelling [online], https://www.maptek.com/video/towards-greater-certainty-in-resource-modelling/ (33:44 min).

Torremans, K, Kyne, R, Doyle, R, Güven, J F and Walsh, J J, 2018. Controls on metal distributions at the Lisheen and Silvermines deposits: Insights into fluid-flow pathways in Irish-type Zn-Pb deposits. *Economic Geology*, vol 113, no 7, pp 1455–1477.