# Using meta-models in simulation-based investment analysis: studying the financing mix of metal mining investments

Savolainen Jyrki, Collan Mikael



**DOI:** 10.25102/fer.2020.01.04

**Using meta-models in simulation-based investment analysis – studying the financing mix of metal mining investments**

**Abstract** — this paper is the first documented research effort on how simple meta-models can be used in simulation-based investment analysis. Modern computers allow the construction and simulation of near real-world emulating models, often referred to as “digital twins”, that offer requisite variety to real world phenomena, such as an industrial investment. These models can be extremely complex and computationally demanding which reduces the scope of their practical applications. This is where meta-models can help.

Meta-models are simple black-box models that are fitted with the input-output - combinations from more complex models to be able to approximate complex model behavior. As the simple meta-models are very fast to solve they may be used to explore much larger solution spaces with considerably higher speed and lower resources needed than the original models.

We demonstrate how the meta-modeling approach can be used in the context of metal mining investment analysis that is originally conducted with a dynamic system model constructed based on a real-world metal mining investment. We show how two simple meta-models, a linear regression model and a regression-tree model, can be used in gaining insight about a suitable financing-mix for the said metal mining investment.

**Keywords**— Meta Models; Simulation; Data Mining; Investment Analysis

**JEL-classification**— C38; C55; D25; L72

## I. INTRODUCTION

There is an on-going transformation towards computer-aided engineering design of processes and products, which means that more and more systems and their components are virtually modeled. These models are often referred to as “digital twins” referring to a digital replica of some physical entity over their lifetime (see discussion in, e.g., Negri, Fumagalli, & Macchi, 2017). When these models are subjected to simulation, the behavior of the underlying designs can be studied under various circumstances, before an actual physical counterpart is built. When the complexity of the digital twin models increases, understanding the key elements that drive analysis-results becomes ever more difficult (Grieves & Vickers, 2017). For this purpose, data mining methods, based on the approximation of computationally-heavy, complex models with simpler more robust models are gaining in popularity. This practice is often called *meta-modeling*. Meta-models are relatively simple “black-box”-models that are used to replace parts of or even whole complex models - meta-models do not include the full mechanics of the models they mimic and are used to replace. The parameters of the simpler meta-models are typically “fitted” (to the problem at hand) by using the simulation results obtained from the to-be-replaced complex model. The goal is to obtain the best possible correspondence between the (part of the) complex original model that is replaced by the meta-model. That is, the results produced with the high-fidelity modeling are used to tune the simpler, much faster, robust models in order to have them approximate unseen data points with a satisficing level of accuracy.

The benefits of meta-modeling include efficiency improvements in terms of reducing the computation-time needed to run complex models and by way of better result generalization (Burrows, Stein, Frochte, Wiesner, & Műller, 2011). Meta-models often offer additional insights about the influence different variables have on results (Kuhlmann, Vetter, Lübbing, & Thole, 2005) and may be helpful in discovering interesting patterns in data (Rupnik, Kukar, & Krisper, 2007). Importantly, meta-modeling can be used to understand why specific solutions (of interest) are reached, which may be as important as the solutions themselves (Geoffrion, 1976).

In this paper we study the use of two different meta-models, a linear regression- and a regression-tree model, in the context of investment analysis and based on a complex techno-economic system dynamic model, or a "digital twin" designed for the purpose of analyzing metal mining investments. The model used as the basis has been presented in detail in (Savolainen, Collan, & Luukka, 2017) and it integrates several aspects of economic analysis of metal mining in a single platform and comprises of *four interlinked subsystems* for production, cash-flow, the balance sheet, and for valuation. Figure 1 illustrates the use of meta-models on the general background of modeling and the approach adopted in this paper.

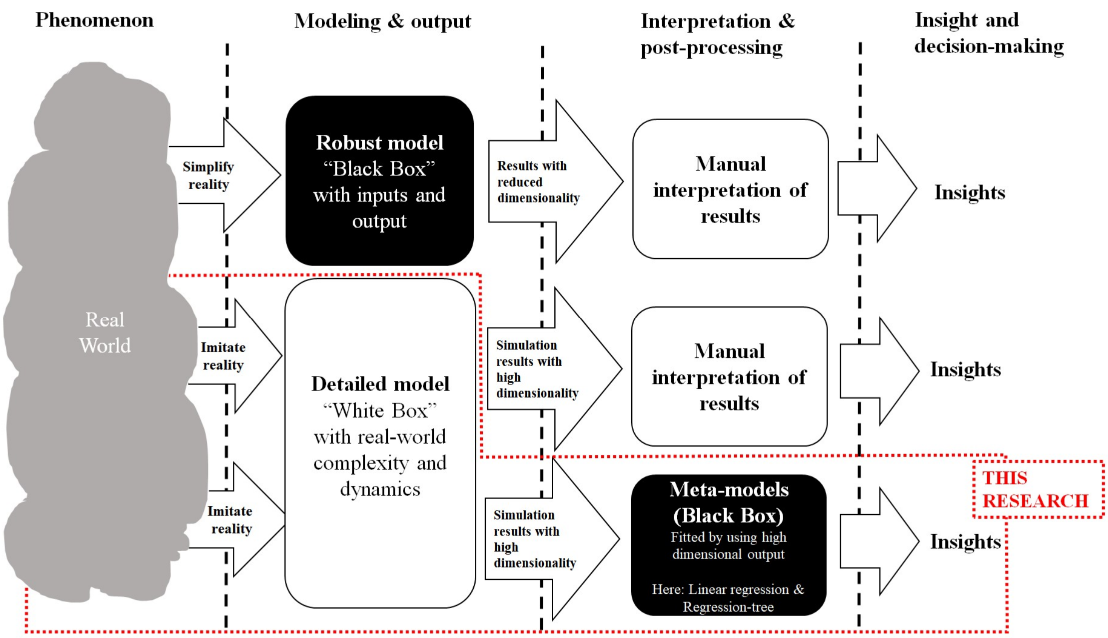


**Fig 1.** Two modeling approaches to real world phenomena and three ways to analyze the results into actionable insights. The approach of this research highlighted with a dotted line.

System dynamic models are suitable for representing industrial systems with feedback-loops and time delays (Größler, Thun, & Milling, 2008) and they address some of the difficulties of economic (real options) modeling, such as intuitive understandability and the ability to include multiple different types of uncertainties in a single model (Lander & Pinches, 1998). System dynamic modeling has been used in the economic evaluation of mining investment also previously (Sontamino & Drebenstedt, 2011, 2014; Inthavongsa, Drebenstedt, Bongaerts, & Sontamino, 2016), and research about the financing mix used in mining investments has also been previously conducted (Savolainen, Collan, Kyläheiko, & Luukka, 2017).

More specifically, with the help of meta-modeling, we are interested in increasing our understanding of two issues. First, how the financing mix of a metal mining investment affects the value (net present value) of the investments for equity-holders and, second, the probability of default of the debt taken, in situations, where the mine-management cannot optimize the value of the mine by temporarily shutting it down, due to obligatory debt repayments (and interest payments on the debt). The real option of the debtors to "force the mine to stay open" is in conflict with the real option to "temporarily shut-down the mine" that the mine management (equity holder) has.

The debtor´s real option, in the used model, to force the mine to stay open is triggered by the low cash-balance of the mining investment, which means that the mine will remain open and generate cash-flow to service the debt even, when it would not be the optimal policy from the equity-holder (value) point-of-view. Our goal is to study, whether a debt-level (financing mix) can be found that simultaneously maximizes the value (net present value) of the investment for the equity-holder, while minimizing the debt default-risk, calculated as the number of defaults over simulated scenarios, and would thus indicate an "optimal" capital structure for the project. In the valuation with the complex full model we adopt the approach used in Savolainen, Collan, Kyläheiko, & Luukka (2017), where the discount rate used in calculating the net present value changes as a function of leverage. In vein with what is discussed in Smith (2002) and as a benchmark, a valuation with 100% financing is performed before going into the leverage optimization.

This paper introduces, to the best of our knowledge, for the first time the use of meta-modeling in the context of investment analysis. Using meta-modeling to study complex systems with real options is also a first. The results presented open new avenues for academic research and provide new insight also for the practitioner. The rest of the this paper is structured as follows: section two provides a literature review on metamodeling; in section three we provide a short introduction to the underlying case, the method and the specific meta-models used, and present meta-model fitting and simulation results. The paper closes with section four, where findings are summarized and discussed.

## II. Literature Review on Meta-modeling and Simulation Data-mining

In the broad context of scientific enquiry, numerical simulation of complex phenomena has become the “third mode of science” to complement theory and experiments (Hu, Yin, Peng, & Li, 2006; Mei & Thole, 2008). According to Rupnik et al. (2007) data mining can be shortly defined as “the process of analyzing data in order to discover implicit, but potentially useful information and uncover patterns and relationships hidden in the data”, or “as a collection of analytics-driven techniques and technologies” supporting knowledge discovery (Painter, Erraguntla, Hogg Jr., & Beachkofski, 2006). Although some recently published work tends to emphasize a combination of *simulation* and *data mining* as a novel invention, the idea of is in no way new in the context of operations research (OR). As early as 1970s Kleijnen (1979) paid attention to the fact that simulations are often of an ad-hoc character, and there is a broader need to generalize the results. Other early efforts of meta-modeling in OR include the work of Lawless and others (1971) who generalized the insights of a complex disaster-planning simulation with alternative patient-treatment policies by using linear regression.

A meritorious and widely cited review of meta-modeling and metamodel-based design optimization is provided by Wang and Shan (2006). They list the roles of meta-modeling as: model approximation, design space approximation, problem formulation, and optimization support. Geoffrion (1976) describes a general methodological approach to auxiliary model building and Kleijnen (1979) provides an analytical introduction to regression meta-models and their implementation. The use of meta-models in simulation optimization is covered in a review of Xu and others (2015).

A lot of the published research under *simulation data mining* concentrates on engineering applications of the automotive industry and in particular on car-crash simulation data analysis. Carcrashes are a non-repeatable phenomenon (indeterministic system), where the quantification of "crashworthiness" cannot be meaningfully based on a single simulation (Mei & Thole, 2008). Kuhlmann and others (2005) discuss the process of data re-use from crash simulations and Mei and Thole (2008) identify critical points in car design that cause bifurcation of simulation results in indeterministic systems from good to bad scenarios at certain points of time in the dynamic simulation. They show that weak points of the car model may be re-designed with subtle changes in parameters to reduce the scattering of results. Zhao and others (2010) use the decision tree algorithm to detect feasible occupant restraint-system configurations (including parameters of seatbelts, airbags, etc.) from simulation data, while Bohn and others (2013) use a three-step process (clustering, dimensionality reduction, and analysis) to draw quick insights from crash simulation data. They underline that the traditionally used principal component analysis (PCA) does not work on models with higher order parameter interactions.

In economic analysis metamodeling techniques have already been applied to solve complex manufacturing system design problems (see, e.g., Negahban & Smith, 2014) and in the optimization of flexible manufacturing systems (see, e.g. Dengiz, İç, & Belgin, (2016); Kuo, Yang, Peters, & Chang, (2007). These topics, similar to investment analysis, deal with multidimensional models with a high degree of freedom. In a recent paper of Liu and others (2019), discussed the need to integrate design optimization (static) and operational optimization (dynamic) aspects into a single "digital twin" -model, which highlight the increasing role of meta-modeling in their interpretation. Weinberger & Moshfegh (2018) present an investment case, where a polynomial regression meta-model is applied to the profitability analysis of a combined heat and power production investment. In this work we concentrate on the system design of an investment and exclude the dynamic optimization aspects.

This research is a new addition to the literature of using meta-models and takes their use to a new direction – namely to the economic analysis of investments with multiple stakeholders and to the context of mining industry feasibility modeling.

## III. Case description, method, and meta-model fitting and simulation results

### A. Case

The metal mining industry and the underlying dynamic techno-economic mining investment model provide an excellent context for a meta-model testing. The life of mining investments usually spans tens or even hundreds of years and the profitability of these investments depends on metal prices and on how technical uncertainties play out. The models used can be highly complex and therefore there is a place for using meta-models in making analyses faster, especially when multiple different metal-mine investments are studied simultaneously, e.g., in a portfolio of mines analysis.

The case used is that of a nickel mine, presented previously by Savolainen and others (2017). Here we assume that the mining operation is ongoing (already started), because this reduces the simulation rounds needed to arrive at meaningful results. The goal of using the model is to maximize value for the equity holders, while ensuring the servicing of debt with interest. The model includes three real options with interactions: i) flexibility to temporarily shut-down mining, if metal prices go below a set limit (mine management, equity holders); ii) option to forbid dividend pay-out (debt-holders); iii) option to force the mine to keep operating, when the cash-balance reaches a minimum limit (debt-holders). The real options ii and iii held by the debt-holders are in reality loan covenants and in the simulation they are removed when the loan is fully paid back. Three key variable values, Ni-price reversion level, the interest rate on the debt, and the leverage (debt-level) are varied and the used input (and the resulting output) values are stored for each simulation run. The loan payment schedule is assumed to be fixed and to start from month thirteen (beginning of the second year), where the debt is paid back in installments of 10 MEUR, while the final debt re-payment is the remaining debt-balance. The used values for the key variables in the model are listed in Appendix 1. For full details on the model and model-mechanics we refer to Savolainen and others (2017).

## B. Meta-models used and fitting the meta-models

As discussed above, the underlying model used in this research is the dynamic system-model of a metal mining investment presented in Savolainen, Collan, & Luukka (2017). The model is used to run a 100-round Monte Carlo simulation for each design-parameter combination, this means in the case of running 792 parameter combinations, as is done here, a total of 79 200 simulation rounds is run. We use one hundred randomly selected price-scenarios in the simulations that is, the same one hundred scenarios for all tested parameter-combinations in order to reach true comparability of results. The number of simulations has not been optimized and it is a normatively set value – finding the optimal number of simulations is left outside the scope of this research. The dynamic system-model is a non-linear “digital twin”-model and the contents of the model are presented and visible as a function-block diagram. The model is able to consider complex real option interactions.

In this work, the selected meta-models are a linear regression model and a regression-tree model that are fitted using the input parameters of the original model and its respective reached simulation outcomes. Our focal points of interest are the *value for equity-holders* in terms of net present value (NPV) and the *probability of default on the debt* as a percentage of simulated outcomes. The two selected meta-models are fitted separately based on the same six inputs out of which three are critical investment variables (reversion level of the Ni-price on the long term, debt interest rate, and initial debt level) and the three aforementioned real options (in use or not).

This means that four separate meta-models in total are built and analyzed as both the value for equity-holders and the probability of default on the debt both have a separate model and they are tested with two sampling-methods each (two by two). Typically, both meta-models are run with the same input variable values at the same time to be able to find value-combinations that give good outputs simultaneously.

Goodness of the meta-model fitting can be tested either by comparing generated observations from the original simulations, with the meta-model generated “forecasts”, or by testing with mean residual sum of squares (Kleijnen, 1979) - in this research we apply both these techniques. The two types of meta-models used in this study are *linear regression* and the *regression-tree* models. Linear regression is a simple statistical approach to modeling the relationship between a dependent variable and a set of explanatory variables. The regression-tree is a supervised machine learning method capable of predicting the outcome of a dependent variable, based on a number of independent variables. The regression-tree implementation used here is an "off the shelf" package available in the Matlab software, for details we refer to the documentation available on their website (Mathworks, 2019). For a complete technical description of the regression-tree method we refer the interested reader to Loh (2011). There are a number of other possible methods, such as, e.g., Radial Basis Functions (Bagheri, Konen, Emmerich, & Bäck, 2017) that can also serve as meta-models, but they are left outside the scope of this research.

The meta-model fitting is "slave" to the decisions made about the design of the underlying more complex model, because the fitting is done based on a selected sample from the original model results. In this case, the simulated results from the underlying complex model acts as the universe from which a sample of outcomes is drawn and used in tuning the meta-models. The way the sampling is done affects the end result and puts limits on the design space and is a part of the design of meta-model analyses (Kleijnen, 1979; Simpson, Lin, & Chen, 2001). In general, one can say that the number of samples used should depend on the complexity of the approximated system (function) (Wang & Shan, 2006).

**Table 1.** Variable-combinations simulated: (I) scenario-sampling 792 combinations, (II) Latin hypercube-sampling 704 combinations (20% * 3520), and (III) test-points with 150 combinations.

| **PARAMETER** | | | | | | |
|---|---|---|---|---|---|---|
| Sampling-method | | I: Scenarios | | II: Hypercube | | III: Test |
| **Parameter** | **Unit** | **Range** | **N** | **Range** | **n** | **Range** |
| *Real Option(s)* | | | | | | |
| Mothballing | - | 0/1 | 2 | 0/1 | 2 | 0/1 |
| *Covenant(s)* | | | | | | |
| Dividends | - | 0/1 | 2 | 0/1 | 2 | 0/1 |
| Force open | - | 0/1 | 2 | 0/1 | 2 | 0/1 |
| Dividends + Production | - | 0/1 | 2 | 0/1 | 2 | 0/1 |
| *Financial* | | | | | | |
| Initial Leverage | % | 0:10:100 | 10 | 0:10:100 | 10 | 0, 15:10:95 |
| Nickel price reversion level | k$/tn | 11, 15, 18 | 3 | 11:1:18 | 3 | 13, 16.5 |
| Loan interest rate | % | 4, 8, 12 | 3 | 4:1:12 | 3 | 5 |
| TOTAL COMBINATIONS | | | 792 | | 3520 | 150 |
| Sampled, % of TOTAL COMB. | | | 100% | | 20% | |
| **TOTAL to be simulated** | | | **792** (100%*792) | | **704** (20%*3520) | |

Here we use two sampling methods, “scenario-sampling” and “Latin hypercube-sampling”. Scenario-sampling is a commonly applied method in investment analysis, where the sampling concentrates on the limits of the given design space based on typically three scenarios (min, max, middle). The Latin hypercube-sampling method is a semi-random sampling method that strives to ensure that the drawn samples are a good representation of the original population, for details we refer to the original publication of McKay, Beckman, & Conover (1979).

As a result of fitting the meta-models with the sampled data from the original model, the meta-model is able to yield approximately “same” solutions as the original model with the chosen inputs, but within the limits of the sampling used. Properly fitted and well-functioning meta-models enclose the key insights of a complex system, while they typically require only a fraction of the computation time. Naturally, any changes in the original model typically require the meta-models to be re-fitted.

The simulations were run on a modern laptop computer and each set of one hundred simulation rounds took approximately one and a half minutes. This means, in the case of the Latin hypercube-sampling, a total simulation time of seventeen and a half hours that comes from 704 variable-configurations times 1.5 minutes each. Specifically, the sampling of design space was performed with Matlab by using in built functions ("*combvec*" and "*lhsdesign*"). The "*lhsdesign*" function returns a continuous variable value from an interval 0 to 1 and the outputs were rounded either to 0, or to 1. There is a possibility that the same simulation-point is simulated more than once, but in our case (with 20% sampling) no duplicates were produced. Table 1 presents the simulated variable-combinations.

### C. Simulation results with a linear regression meta-model

Linear regression can be thought of as a starting point for using meta-models, because it has been widely applied and well-documented in the earlier literature related to the use of meta-models (Kleijnen, 1979; Lawless et al., 1971; Painter et al., 2006). We fitted a linear regression model to the simulated data with the data sampled with both sampling methods, to our surprise the $R^2$ was at a very high level for the value for equity-holders meta-model with both scenario-sampling (.954) and with Latin hypercube-sampling (.951). In the case of the probability of default on the debt meta-model, the $R^2$ was significantly lower in both scenario-sampling (.757) and with Latin hypercube-sampling (.814), suggesting a comparatively higher level of non-linearity in the original model. Statistics for the fitting of the meta-models are available in Table 2.

**TABLE 2** Resulting meta-models from linear regression fitting using scenarios and hypercube-sampling approaches both for the equity holder and the creditor. *Legend:* tStat = Statistical significance of the variable; RMSE = Root Mean Squared Error.

| **EQUITY HOLDER** | ***tStat*** | | | ***Estimate*** | | |
|---|---|---|---|---|---|---|
| | **Scen** | **HC** | **Diff** | **Scen** | **HC** | **Diff** |
| RMSE | | | | 16.7 | 11.6 | -5.1 |
| R-squared | | | | 0.954 | 0.961 | 0.007 |
| *Variables (continuous)* | | | | | | |
| Ni-price (reversion) | 125.900 | 128.010 | 2.110 | 0.026 | 0.028 | 0.002 |
| Loan Interest Rate | -4.851 | -3.889 | 0.962 | -0.882 | -0.725 | 0.157 |
| Initial Debt | 0.095 | 0.958 | 0.863 | 0.113 | 0.836 | 0.723 |
| *Real Options (binomial)* | | | | | | |
| Mothballing | 0.722 | 0.192 | -0.530 | 0.857 | 0.167 | -0.690 |
| No equity dividends | -0.612 | -2.047 | -1.435 | -0.727 | -1.791 | -1.064 |
| Force open | -17.789 | -25.307 | -7.518 | -0.334 | -0.384 | -0.050 |

| **CREDITOR** | ***tStat*** | | | ***Estimate*** | | |
|---|---|---|---|---|---|---|
| | **Scen** | **HC** | **Diff** | **Scen** | **HC** | **Diff** |
| RMSE | | | | 13.1 | 8.7 | -4.4 |
| R-squared | | | | 0.757 | 0.814 | 0.057 |
| *Variables (continuous)* | | | | | | |
| Ni-price (reversion) | 39.564 | 40.707 | 1.143 | 0.006 | 0.007 | 0.000 |
| Loan Interest Rate | -3.797 | -5.126 | -1.329 | -0.540 | -0.715 | -0.175 |
| Initial Debt | 0.900 | 0.292 | -0.608 | 0.836 | 0.191 | -0.645 |
| *Real Options (binomial)* | | | | | | |
| Mothballing | 1.178 | 2.246 | 1.069 | 1.093 | 1.468 | 0.375 |
| No equity dividends | -0.781 | 0.909 | 1.689 | -0.725 | 0.595 | 1.320 |
| Force open | -29.481 | -35.251 | -5.770 | -0.433 | -0.400 | 0.033 |

The linear regression meta-models were further studied and we found that the nickel-price is a dominating variable in both the value for equity-holders and the probability of default on the debt meta-model cases, but the effect nickel-price has is little less pronounced in the probability of default on the debt model case. As expected, a higher interest rate and a higher initial leverage-level have a negative effect on the equity-holder value and on the ability of the investment to service the debt. The option to force the mine to be kept open (loan covenant) does not increase the value for the equity-holders, nor does it enhance the debt servicing ability, which rules it out from the possible future simulation efforts. The covenant that allows the debt-holders to veto dividend-payments has different implications depending on the sampling method used: in the case of hypercube-sampling a small positive effect on the debt-servicing ability is observed.

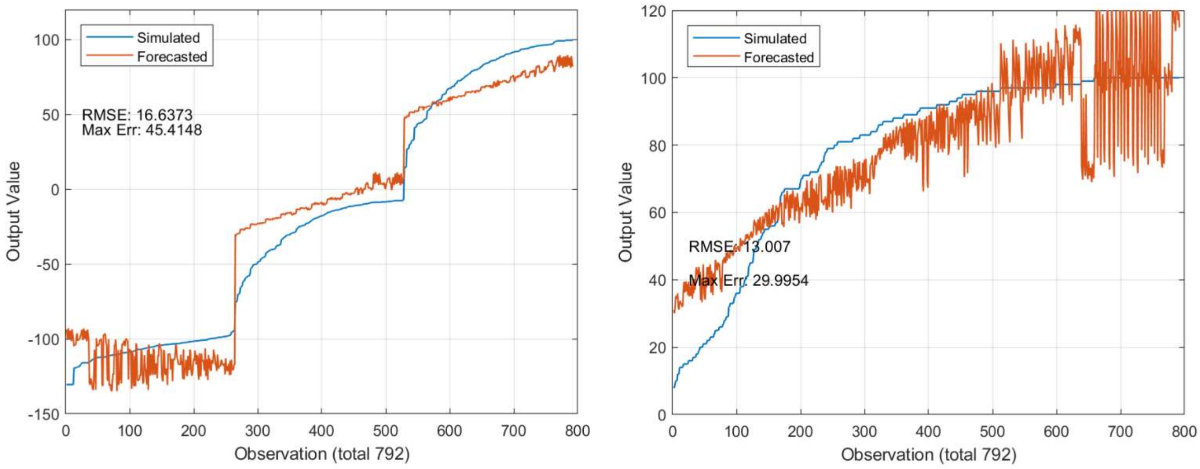


**Fig 2.** Original (in sample) simulated and meta-model forecasted results from the linear regression meta-model, with scenario-sampling. *Left:* Value for equity-holders; *Right:* Probability of default on the debt.

Goodness of fit testing for the linear regression meta-model with scenario-sampling by using the original data also used for teaching (fitting) the meta-model is visualized in Figure 2 – from the figure one can see that the linear regression model based on scenario-sampling does not generalize the original model behavior very accurately. The results from the Latin hypercube-sampling are better, but the difference is not remarkable. When new input variable combinations are tested (combinations not used in the fitting of the meta-model) the Latin hypercube-sampling based linear regression meta-model gives better results than the scenario-sampling based meta-model.

#### D. Simulation results with a regression-tree meta-model

As a side-product of meta-model fitting, regression-tree models provide a graphical presentation of the tree-structure that is useful for interpreting results. Figure 3 shows a tree-structure that resulted from the value for equity-holders model fitting. For illustration, the tree-structure has been forced to ten splits and is at a rough level, when compared with the original model tree with more than a hundred splits. The simplified presentation can be used to make some rough conclusions about the mean reversion price-level and the leverage level - the tree implies (first split) that equity-holder value deteriorates at prices below ~15 000$/ton.

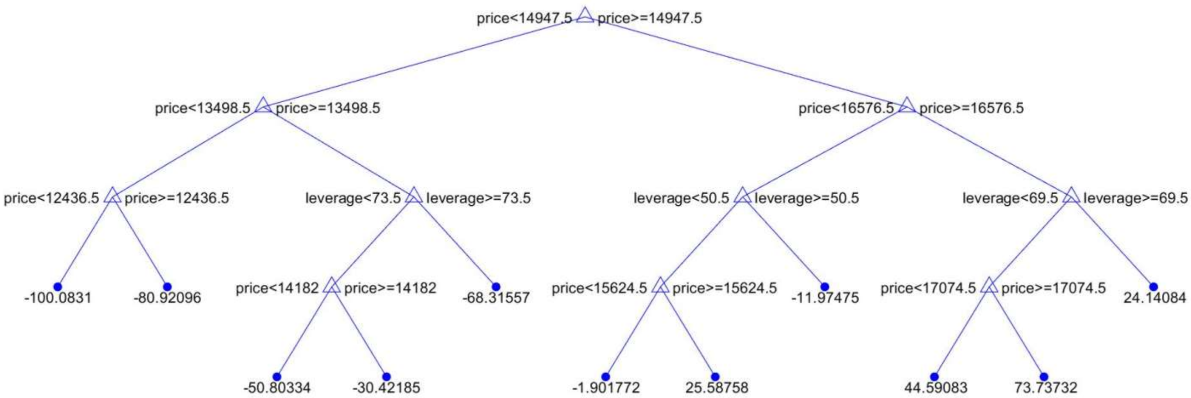


**Fig 3.** Ten-split tree-structure from the value to equity-holders model fitting.

There is a significant difference between the results received with the scenario-sampling and with Latin hypercube-sampling – scenario-sampling produces clearly inferior and poor results for predicting outcomes from in-sample variable combinations, while the results obtained with the Latin hypercube-sampling for both models are fairly accurate.

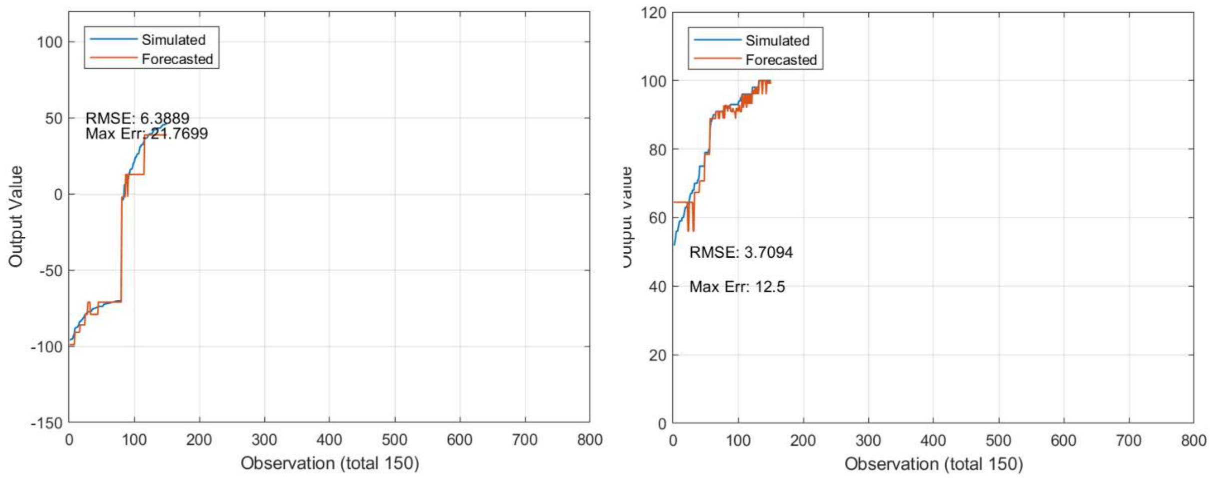


**Fig 4.** Out of sample (new) simulated and meta-model forecasted results from the regression-tree meta-model, with Latin hypercube-sampling. *Left:* Value for equity-holders; *Right:* Probability of default on the debt.

With Latin hypercube-sampling the predictions for out of sample (new) variable combinations are quite accurate (see Figure 4), while the predicting performance of both meta-models with scenario-sampling remains of low quality. All in all, the regression-tree meta-model with Latin hypercube-sampling seems to be the most accurate of the four tested meta-models.

### E. Meta-model use for obtaining further results

Now, that we have tested two meta-models, both with two different sampling-methods, and found the best meta-model, we can use it to derive further insights, because the time constraints that are present with running the original dynamic system model are no longer binding us. What we are interested in is finding out what is the input variable combination-space that returns situations where the value to the equity-holders is positive (NPV > 0) and the probability that there is a default on the debt is very low ($p < 0.01$). To do this we input the whole design space into the meta-model to cover the whole space and to find any “sweet spots” and then limit the results according to the positive NPV and very low debt default requirement.

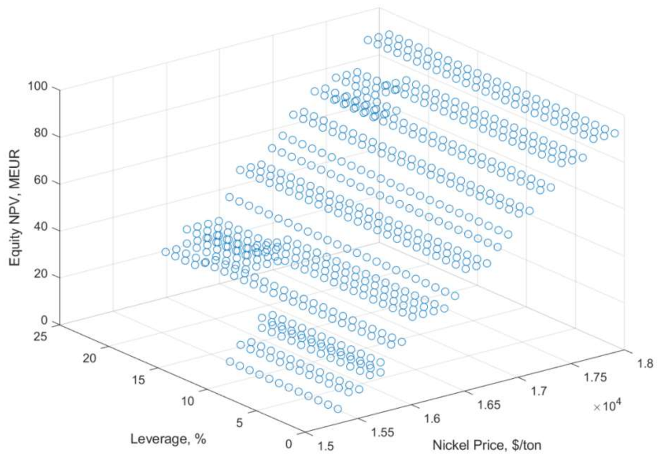


**Fig. 5** A 3D-illustration of the solutions fulfilling the positive value for equity-holders and very low debt-default requirements created with the regression-tree meta-model with Latin hypercube-sampling. Overlapping points are due to the effect of other variables not included in the two variable axes.

The solution space set is graphically illustrated in Figure 5, where one can clearly see that leverage between zero and approximately twenty-five percent returns a very low probability for default with a positive value for equity-holders. The value for equity-holders depends heavily on the long-term reversion level of the metal-price (Ni). This kind of information is very important from the point of view of setting up a metal mine with the characteristics like the one modeled here. As the financing mix is most often ex-ante negotiable, making the right decisions with adequate balance of risk and return makes a lot of sense for all involved stakeholders.

### F. Discussion

Based on the case-example, we make the following implications. First, it is shown that the existing knowledge on meta-models can be applied to enhance the application of investment analysis models that may currently be regarded as too complex to be applied. Second, the ability to effectively use complex investment models in economics, via the meta-modeling, enables the use of otherwise too-heavy-to-run, multi-disciplinary, integrated, and generic investment models that can accommodate the points-of-view of several stakeholders at simultaneously (here equity holders and lenders). From the practitioner and policy-maker point of view, the improved ability to model reality in a better way can reduce the problems associated with the typical parallel use of owner-specific, simplified, and stand-alone models in the analysis of complex investments.

## IV. SUMMARY AND CONCLUSIONS

In this paper, we have presented the use of meta-models in the context of investment analysis. To the best of our knowledge, this is a novel contribution. We have used two simple meta-models, a linear regression meta-model and a regression-tree meta-model to mimic the input-output relationship produced by a complex dynamic techno-economic system model designed for the analysis of metal mining investments. We have shown that the simple meta-models used are able to produce a usable level of accuracy in mimicking the results from the more complex model, even with the limited number of simulation rounds used in generating results for the fitting phase. The limitations in the approach make all results preliminary, however, one can state that the results seem to demonstrate that meta-modeling is a viable approach for complementing the use of complex investment analysis models with several interacting design parameters. Despite the promising results of this study, we observe the previous conclusion by Wang and Shan (2006) and Zang and others (2013) that for now there are no mathematically rigorous methods to actually quantify the uncertainty in meta-modeling, and we suggest the reader not to draw any definitive conclusions based on these results.

We highlight that the use of meta-models is a generalized technical shortcut to speed up and simplify simulations, once a reliable mass of results has been generated with a complex white-box model that is not limited to specific scientific discipline. In the context of investment analysis, interesting future research directions include creation of several meta-models that mimic a portfolio of (multiple) investments with underlying complex investment models and running them simultaneously with the goal of being able to understand super-system dynamics and outcomes through observing the aggregate results from the meta-models. Also, complex market models exhibiting regular patterns could be modeled. These types of endeavors might prove to be extremely slow with the complex original models due to high computing power demand and computing time restrictions. Another already emerging area of meta-modeling in the context of product/process development using digital twin simulation is the teaching of metamodels for control system applications instead of real data. On a conceptual level, the control logic of these systems could be enhanced with the metamodels of economic value presented in this paper to improve their productivity and autonomous operation in non-stationary environments/conditions.

## V. References

Bagheri, S., Konen, W., Emmerich, M., & Bäck, T. (2017). Self-adjusting parameter control for surrogate-assisted constrained optimization under limited budgets. *Applied Soft Computing*, *61*, 377–393. https://doi.org/https://doi.org/10.1016/j.asoc.2017.07.060

Bohn, B., Garcke, J., Iza-teran, R., Paprotny, A., Peherstorfer, B., Schepsmeier, U., & Thole, C.-A. (2013). Analysis of car crash simulation data with nonlinear machine learning methods. *Procedia Computer Science*, *18*, 621–630. https://doi.org/10.1016/j.procs.2013.05.226

Burrows, S., Stein, B., Frochte, J., Wiesner, D., & Műller, K. (2011). Simulation Data Mining for Supporting Bridge Design. In P. Vamplew, A. Stranieri, K.-L. Ong, C. Christen, & P. J.

Kennedy (Eds.), *AusDM ’11 Proceedings of the Ninth Australasian Data Mining Conference - Volume 121* (pp. 163–170). Ballarat, Australia: Australian Computer Society, Inc., Darlinghurst, Australia, Australia.

Dengiz, B., İç, Y. T., & Belgin, O. (2016). A meta-model based simulation optimization using hybrid simulation-analytical modeling to increase the productivity in automotive industry. *Mathematics and Computers in Simulation*, *120*, 120–128. https://doi.org/https://doi.org/10.1016/j.matcom.2015.07.005

Geoffrion, A. M. (1976). The Purpose of Mathematical Programming is Insight, Not Numbers. *Interfaces*, *7*(1), 81–92. https://doi.org/10.1287/inte.7.1.81

Grieves, M., & Vickers, J. (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems BT - Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. In F.-J. Kahlen, S. Flumerfelt, & A. Alves (Eds.) (pp. 85–113). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-38756-7_4

Größler, A., Thun, J.-H., & Milling, P. M. (2008). System dynamics as a structural theory in operations management. *Production and Operations Management*, *17*(3), 373–384. https://doi.org/10.3401 / poms.1080.0023

Hu, J., Yin, J., Peng, Y., & Li, D. (2006). Knowledge Discovery from Multidisciplinary Simulation to Support Concurrent and Collaborative Design. In *2006 10th International Conference on Computer Supported Cooperative Work in Design* (pp. 1–6). https://doi.org/10.1109/CSCWD.2006.253177

Inthavongsa, I., Drebenstedt, C., Bongaerts, J., & Sontamino, P. (2016). Real options decision framework: Strategic operating policies for open pit mine planning. *Resources Policy*, *47*, 142–153. https://doi.org/10.1016/j.resourpol.2016.01.009

Kleijnen, J. P. C. (1979). Regresssion Metamodels for Generalizing Simulation Results. *IEEE*

*Transactions on Systems, Man and Cybernetics*, *9*(2), 93–96.

Kuhlmann, A., Vetter, R.-M., Lübbing, C., & Thole, C.-A. (2005). Data Mining on Crash Simulation Data. In P. Perner & A. Imiya (Eds.), *Machine Learning and Data Mining in Pattern Recognition* (pp. 558–569). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kuo, Y., Yang, T., Peters, B. A., & Chang, I. (2007). Simulation metamodel development using uniform design and neural networks for automated material handling systems in semiconductor wafer fabrication. *Simulation Modelling Practice and Theory*, *15*(8), 1002–1015. https://doi.org/https://doi.org/10.1016/j.simpat.2007.05.006

Lander, D. M., & Pinches, G. E. (1998). Challenges to the practical implementation of modeling and valuing real options. *The Quarterly Review of Economics and Finance*, *38*(3), 537–567. https://doi.org/10.1016/S1062-9769(99)80089-1

Lawless, R. W., Williams, L. H., & Richie, C. G. (1971). A sensitivity analysis tool for simulation with application to disaster planning. *Simulation*, *17*(6), 217–223. https://doi.org/10.1177/003754977101700602

Liu, Q., Zhang, H., Leng, J., & Chen, X. (2019). *Digital twin-driven rapid individualised designing of automated flow-shop manufacturing system*. *International Journal of Production Research* (Vol. 57). Taylor & Francis. https://doi.org/10.1080/00207543.2018.1471243

Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 14–23. https://doi.org/10.1002/widm.8

Mathworks. (2019). Fit binary decision tree for regression. Retrieved April 18, 2019, from https://se.mathworks.com/help/stats/fitrtree.html

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, *21*(2), 239–245. https://doi.org/10.1080/00401706.1979.10489755

Mei, L., & Thole, C. A. (2008). Data analysis for parallel car-crash simulation results and model optimization. *Simulation Modelling Practice and Theory*, *16*(3), 329–337. https://doi.org/https://doi.org/10.1016/j.simpat.2007.11.018

Negahban, A., & Smith, J. S. (2014). Simulation for manufacturing system design and operation: Literature review and analysis. *Journal of Manufacturing Systems*, *33*(2), 241–261. https://doi.org/https://doi.org/10.1016/j.jmsy.2013.12.007

Negri, E., Fumagalli, L., & Macchi, M. (2017). A Review of the Roles of Digital Twin in CPS-based Production Systems. *Procedia Manufacturing*, *11*, 939–948. https://doi.org/https://doi.org/10.1016/j.promfg.2017.07.198

Painter, M. K., Erraguntla, M., Hogg Jr., G. L., & Beachkofski, B. (2006). Using Simulation, Data Mining, and Knowledge Discovery Techniques for Optimized Aircraft Engine Fleet Management. In *Proceedings of the 38th Conference on Winter Simulation* (pp. 1253–1260). Winter Simulation Conference. Retrieved from http://dl.acm.org/citation.cfm?id=1218112.1218341

Qureshi, A. M. (2007). System dynamics modelling of firm value. *Journal of Modelling in Management*, *2*(1), 24–39. https://doi.org/10.1108/17465660710733031

Rupnik, R., Kukar, M., & Krisper, M. (2007). Integrating data mining and decision support through data mining based decision support system. *Journal of Computer Information Systems*, *47*(3), 89–104.

Savolainen, J., Collan, M., Kyläheiko, K., & Luukka, P. (2017). On the trade-off between the leverage effect and real options thinking: A simulation-based model on metal mining investment. *International Journal of Production Economics*, *194*, 43–51. https://doi.org/10.1016/j.ijpe.2017.06.002

Savolainen, J., Collan, M., & Luukka, P. (2017). Analyzing operational real options in metal mining investments with a system dynamic model. *The Engineering Economist*, *62*(1), 54–

72. https://doi.org/10.1080/0013791X.2016.1167988

Simpson, T. W., Lin, D. K. J., & Chen, W. (2001). Sampling Strategies for Computer Experiments: Design and Analysis. *International Journal of Reliability and Applications*, *2*(3), 209–240.

Smith, L. D. (2002). Discounted Cash Flow Analysis Methodology and Discount Rates. *CIM Bulletin*, *95*(1062), 101–108.

Sontamino, P., & Drebenstedt, C. (2011). Decision Support System of Coal Mine Planning Using System Dynamics Model: Introduction and Reviews. In *6th Freiberg-St. Petersburg Kolloquium junger Wisseshaftler 15-17 Juni 2011, Freiberg, Germany* (pp. 1–6). Retrieved from http://www.researchgate.net/publication/233959780_Decision_Support_System_of_Coal_Mine_Planning_Using_System_Dynamics_Model_Introduction_and_Reviews/file/9fcfd50d56cd10a622.pdf

Sontamino, P., & Drebenstedt, C. (2014). A Prototype Decision Making Tool of a Coal Mine Planning Using System Dynamics Model. In C. Drebenstedt & R. Singhal (Eds.), *Mine Planning and Equipment Selection*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-02678-7

Wang, G. G., & Shan, S. (2006). Review of Metamodeling Techniques in Support of Engineering Design Optimization. *Journal of Mechanical Design*, *129*(4), 370–380. https://doi.org/10.1115/1.2429697

Weinberger, G., & Moshfegh, B. (2018). Investigating influential techno-economic factors for combined heat and power production using optimization and metamodeling. *Applied Energy*, *232*, 555–571. https://doi.org/https://doi.org/10.1016/j.apenergy.2018.09.206

Xu, J., Huang, E., Chen, C.-H., & Lee, L. H. (2015). Simulation Optimization: A Review and Exploration in the New Era of Cloud Computing and Big Data. *Asia-Pacific Journal of*

*Operational Research*, *32*(3), 34.

Zhang, S., Zhu, P., Chen, W., & Arendt, P. (2013). Concurrent treatment of parametric uncertainty and metamodeling uncertainty in robust design. *Structural and Multidisciplinary Optimization*, *47*(1), 63–76. https://doi.org/10.1007/s00158-012-0805-5

Zhao, Z., Jin, X., Cao, Y., & Wang, J. (2010). Data mining application on crash simulation data of occupant restraint system. *Expert Systems With Applications*, *37*(8), 5788–5794. https://doi.org/10.1016/j.eswa.2010.02.029

## APPENDIX 1

**Table A1.** Key variable values used in the original (underlying) dynamic system model (modified from (Savolainen, Collan, Kyläheiko, et al., 2017)).

| VARIABLE | Pessimistic | Best guess | Optimistic | Vol., % | Unit |
|---|---|---|---|---|---|
| *Technical* | | | | | |
| Reserve size | - | 140 000 | - | - | Tons |
| Metal yield | 1 000 | 1 200 | 1 400 | 10 | Tons/month |
| *Operating costs* | | | | | |
| Unit cost | 4 000 | 3 500 | 3 000 | - | EUR/ton (of Ni) |
| Fixed cost (production) | | 2 500 000 | | | EUR/month |
| *Mothballing costs* | | | | | |
| Shutdown cost | - | 1 200 000 | - | | EUR (per shutdown) |
| Fixed cost (mothballed) | - | 500 000 | - | | EUR/month |
| Re-start production cost | - | 500 000 | - | | EUR (per shutdown) |
| *Investment costs* | | | | | |
| Initial Investment | - | 60 000 000 | - | - | EUR |
| Initial cash | - | 20 000 000 | - | | EUR |
| *Financials* | | | | | |
| Payment ratio of metal | - | 60 | - | | % |
| Exchange rate | - | 1.1 | - | - | USD/EUR |
| Cost discount rate | - | 5.0 | - | | % |
| Abandon cost | - | 5 000 000 | - | | % |