

# *Physics-Informed Network Models: a Data Science Approach to Metal Design*

**Amit K. Verma, Roger H. French & Jennifer L. W. Carter**

**Integrating Materials and  
Manufacturing Innovation**

ISSN 2193-9764

Integr Mater Manuf Innov  
DOI 10.1007/s40192-017-0104-5



**Your article is protected by copyright and all rights are held exclusively by The Minerals, Metals & Materials Society. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**



# Physics-Informed Network Models: a Data Science Approach to Metal Design

Amit K. Verma<sup>1</sup> · Roger H. French<sup>1,2</sup> · Jennifer L. W. Carter<sup>1</sup> 

Received: 25 September 2017 / Accepted: 7 December 2017  
© The Minerals, Metals & Materials Society 2017

## Abstract

Functional graded materials (FGM) allow for reconciliation of conflicting design constraints at different locations in the material. This optimization requires a priori knowledge of how different architectural measures are interdependent and combine to control material performance. In this work, an aluminum FGM was used as a model system to present a new network modeling approach that captures the relationship between design parameters and allows an easy interpretation. The approach, in an un-biased manner, successfully captured the expected relationships and was capable of predicting the hardness as a function of composition.

**Keywords** Metal design · Network models · Optimization · Functional gradient materials

## Introduction

Materials design is an iterative, multi-criteria, multi-dimensional optimization process, in which secondary aspects, such as the availability of alloy components, and the achievable microstructures are optimized for the desired alloy performance. This optimization needs a priori knowledge of (Processing | Microstructure | Performance) ((P|M|P)) dependencies to fully control the development of materials. Considering the time and scale needed to develop these dependencies, initiatives such as Materials Genome Initiative [1] and Integrated Computational Materials Engineering [2, 3] came into prominence in the last decade.

This brought along the ideas about how to collect large amount of data effectively [4] in a time-efficient manner, infrastructure and tools needed to analyze this big data [5, 6], and an environment that can integrate experiment, computation, and data [7, 8].

Network models are a broad category of statistical models for studying multi-variate relationships in complex systems [9] and have been developed and applied in many fields such as genomics/proteomics [10], climate modeling [11], and materials [12]. Artificial neural networks (ANN) (i.e., regression analysis) are most common in materials science [13, 14], and they quickly produce models that predict dependent variables based on independent variables. Structural equation modeling (SEM) is a network modeling approach used extensively in the social sciences that captures statistically significant linear relationships among dependent and independent variables and provides easily interpretable insight into latent variables [15]. We have generalized SEM beyond simple linear relationships to allow for more complex functional forms, guided by physics and domain-science knowledge, and refer to this as semi-supervised (by domain knowledge), generalized SEM for (P|M|P) modeling. The generalized SEM network models (netSEM) serve as diagnostic/inference models that illustrate the internal network relationships and pathways among variables [16].

---

✉ Jennifer L. W. Carter  
[jennifer.w.carter@case.edu](mailto:jennifer.w.carter@case.edu)

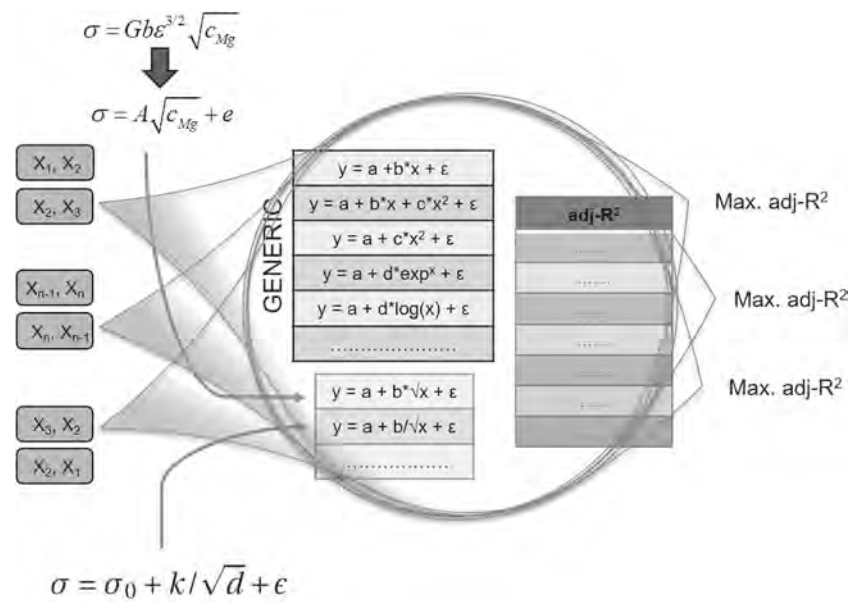
Amit K. Verma  
[amit.a.kumar@case.edu](mailto:amit.a.kumar@case.edu)

Roger H. French  
[roger.french@case.edu](mailto:roger.french@case.edu)

<sup>1</sup> Department of Materials Science and Engineering, Case Western Reserve University, Cleveland, OH 44106, USA

<sup>2</sup> SDLE Research Center, Case Western Research University, Cleveland, OH 44106, USA

**Fig. 1** Schematic of the working of netSEM. First, netSEM form all possible pairs.  $X_1, X_2, \dots, X_n$  are variables. Then, for each pair, netSEM checks all functional forms and ranks them using  $\text{adj-}R^2$ . Functional forms can be generic or may arise from the domain knowledge. Finally, each pair is linked with the functional form with maximum  $\text{adj-}R^2$



## NetSEM Approach

The netSEM approach [16] starts with creating all possible pairs of the variables (i.e.,  $X_i, X_j \forall i, j \in n, i \neq j$ ) and then uses a step-wise regression to move through each pair in a Markovian spirit, whereby the relationship among two variables (i.e.,  $y = X_j = f(X_i) \forall i, j \in n, i \neq j$ ) in the regression is unaffected by prior history, and other variables are considered independent constants [17] (Fig. 1). Then, pathway relationships among the variables are evaluated with the set of permitted functional forms as defined by domain-science predictions (Table 1). The network model assumes that uni-variate correlations exist in parallel and do not have combinatorial or interaction effects. Then, the relationships between variable pairs are rank-ordered using  $\text{adj-}R^2$  values to determine their significance. Finally, the pair-wise system of equations is reported and visualized as a network model for the training dataset, and then models are validated on the testing dataset.

In comparison with ANN, the differences lie in the architecture of the two network models. In ANN, architecture moves forward with all the independent variables in the first layer and all the dependent variables in the last layer, while all layers in between are hidden [19]. The number of hidden layers and nodes in each layer are user defined and there is no simple answer for how the network progresses from one hidden layer to the next hidden layer [20]. This makes interpretation of the physical meaning of each hidden layer/node difficult. The resulting ANN architecture is frozen and it is not strictly possible to tailor the hidden layers for defining (P|M|P) relationships from one layer to another. In comparison, a netSEM model has no hidden layers, and instead, the approach checks dependability among each variable pair which allows the user to define the layers. For example, processing variables in the first layer give rise to microstructure (tailored hidden layer) for desired properties (last layer) in specific environmental conditions. This increased flexibility of the netSEM network does allow an easy interpretation of

**Table 1** Generic uni-variate functional forms evaluated by netSEM package (v0.4.2.1) [18]

Variable fits	Functional forms
Simple linear ( <b>L</b> )	$y = a + bx + \epsilon$
Quadratic ( <b>Q</b> )	$y = a + bx + c * x^2 + \epsilon$
Simple quadratic ( <b>sQ</b> )	$y = a + c * x^2 + \epsilon$
Exponential ( <b>Exp</b> )	$y = a + d * \exp^x + \epsilon$
Logarithmic ( <b>Log</b> )	$y = a + f * (\log(x)) + \epsilon$
Linear change point	$y = a + b * x + b_1 * (x - c) + \epsilon$
Non-linear exponential (G-up, H-down)	$y = a + d * (1 \pm \exp(g(x - h))) + \epsilon$



the meaning of relationship pathways, but reduces the predictability of the network model as compared to ANN.

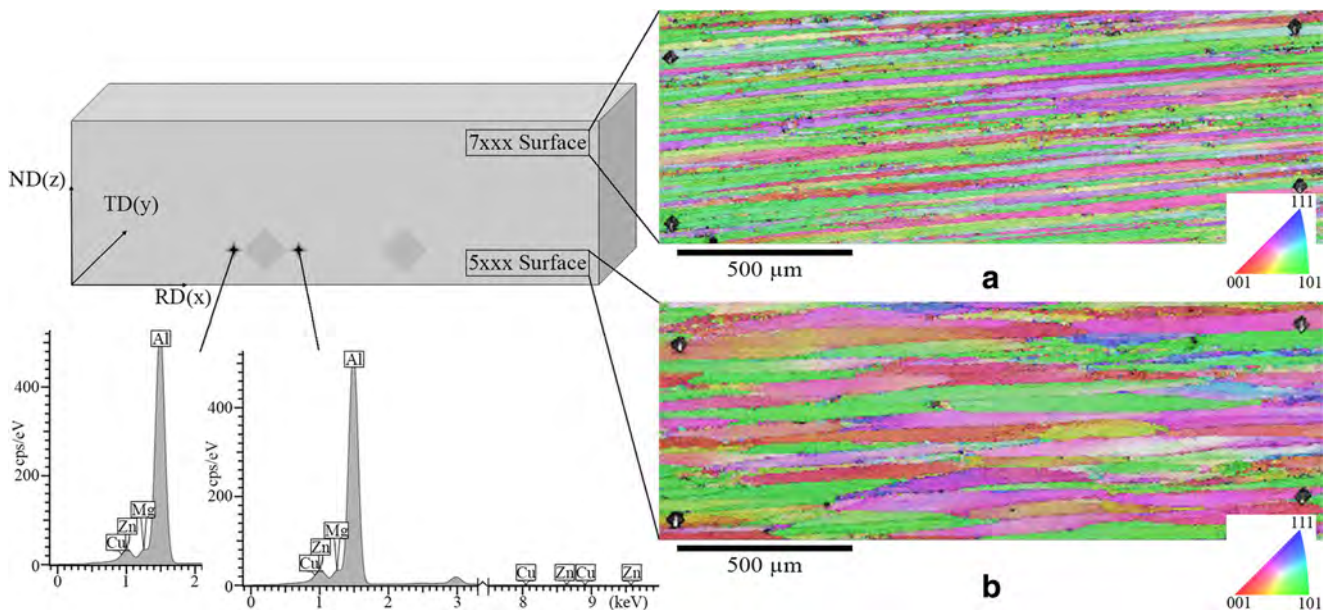
Another key aspect is the use of activation functions. ANN has a family of activation functions for training the model which are a priori decided by the user depending on the data. The usual practice is to use a single activation function to train the model. Once the activation function is selected, the ANN algorithm optimizes the weights between nodes to train the model [21]. Even though the weights are known, the interpretation of the physical meaning of the nodes is difficult [20]. In comparison, netSEM allows the user to explore multiple activation functions (as tabulated in Table 1) for the optimization of the network. The functions in Table 1 are tailorable by the user (Git repository listed at the end of this paper) based on hypothesized functional forms. This further simplifies the interpretation of the network as functional forms are domain guided and arguments are the experimentally derived variables. However, this further reduces the accuracy of prediction of the response as netSEM does not allow the network to self-optimize.

The comparison highlights that while network models such as ANN primarily focus on the prediction of a response, netSEM provides a robust way for interpreting and attributing physics-based meaning to the relationship pathways. Each method is useful for different goals of materials engineering and design. ANN provides better prediction in very complex problems while netSEM can be used to explore potential mechanisms by the incorporation of expected functional forms.

## Material

Functional graded materials (FGM) are a class of materials with engineered continuous compositional gradients through the plate thickness [22]. This work applies the netSEM approach [16] on an aluminum FGM, produced via sequential alloy casting using planar solidification [23, 24], to quantify the  $\langle P|M|P \rangle$  relationships. The material has a continuous gradient in zinc (Zn) and magnesium (Mg) concentrations through the plate thickness (ND(z) in Fig. 2). This subsequently produces a gradient in strengthening mechanisms from a dominant precipitate-strengthened aluminum alloy (AA) (Zn-based AA-7055 [25]) to a dominant strain-hardenable aluminum alloy (Mg-based AA-5456 [25]). Therefore, the material is simultaneously strengthened via solid solution strengthening and precipitation strengthening.

After casting, the aluminum FGM was artificially aged to produce the desired precipitate distribution in the AA-7055, which causes the AA-5456 side of the FGM to undergo an annealing treatment. Figure 2 shows that in addition to the gradient in composition, the processing route also produces a gradient in grain size, with the Mg-rich region having much larger grain size than the Zn-rich side of the FGM. The room temperature ternary phase diagram [26] indicates that within the compositional range of the aluminum FGM (4.6–3.5 at%Mg and 2.9–4.2 at%Zn), the equilibrium precipitate should be  $Mg_{32}Zn_{31.9}Al_{17.1}$ , with remainder of Zn and Mg in solid solution. Precipitation aging experimental results presented by Balderach et al.



**Fig. 2** Sample schematic showing RD(x), TD(y), and ND(z) directions. Small tilted squares represent microhardness indents, and small crosses (x) next to indents represent XEDS measurements. Gradient

in concentration is along the Z-axis. Inverse pole figure maps show a small grain size in Zn-rich regions (a) and a large grain size in Mg-rich regions (b)

**Table 2** Uni-variate functional forms evaluated by netSEM. Asterisk-marked equations were added to the netSEM package (v0.4.2.01) [18] based on domain knowledge of strengthening mechanisms

Variable fits	Functional forms
Simple linear ( <b>L</b> )	$y = a + bx + \epsilon$
Quadratic ( <b>Q</b> )	$y = a + bx + c * x^2 + \epsilon$
Simple quadratic ( <b>sQ</b> )	$y = a + c * x^2 + \epsilon$
Exponential ( <b>Exp</b> )	$y = a + d * \exp^x + \epsilon$
Logarithmic ( <b>Log</b> )	$y = a + f * (\log(x)) + \epsilon$
Linear change point	$y = a + b * x + b_1 * (x - c) + \epsilon$
Non-linear exponential (G-up, H-down)	$y = a + d * (1 \pm \exp(g(x - h))) + \epsilon$
Solid solution* ( <b>SS</b> )	$y = a + b * x^{\frac{2}{3}} + \epsilon$
Inverse square root* ( <b>ISQR</b> )	$y = a + b/\sqrt{x} + \epsilon$

[27] show the formation of metastable  $MgZn_2(\eta)$  over equilibrium  $Mg_{32}Zn_{31.9}Al_{17.1}$ .

## Strengthening Mechanism

Existing strengthening models may provide guidance for the potential functional forms for netSEM. Solid solution strengthening is caused by the interaction between stress fields caused by solute atoms and moving dislocations and has been described by Fleischer [28] and Labusch [29]. A general equation for the increase in yield strength due to a particular solute type can be simplified to

$$\Delta\sigma = AG(T)\epsilon^{\frac{4}{3}}x^{\frac{2}{3}} \quad (1)$$

where  $G$  is the matrix shear modulus,  $\epsilon$  is the mismatch parameter,  $x$  is the solute mole fraction, and  $A$  is a fitting parameter. Assuming  $G$  and  $\epsilon$  are constant in the aluminum FGM (because solute types are constant and in low concentration), the governing equation simplifies to **SS** in Table 2.

Similarly, precipitation hardening due to bypassing of particles by dislocation loops have been refined based on Orowan's 1948 model [30]

$$\sigma_{Or} = M \frac{0.4Gb}{\pi(1-\nu)^{1/2}} \frac{\ln(\bar{d}/b)}{\bar{\lambda}} \quad (2)$$

where  $M$  is the Taylor factor,  $G$  is the matrix shear modulus,  $b$  is the Burgers vector magnitude,  $\nu$  is the Poisson's ratio that is assumed constant with precipitate type, and  $\bar{d} = d\sqrt{2/3}$  and  $\bar{\lambda} = \bar{d}(\sqrt{\pi/4f} - 1)$  are the mean precipitate size and interprecipitate spacing, respectively, and  $d$  is the average precipitate diameter and  $f$  is the precipitate volume fraction. The model indicates that the strength is not directly related to solute concentration but is dependent on several interrelated variables, particularly  $d$  and  $f$ , that are indirectly related to solute concentration [31]. Without insight into the detailed processing parameters, it is difficult to simplify this model to the compositional variables,

since composition and processing parameters interact to determine volume fraction, distribution, morphology, and precipitate size.

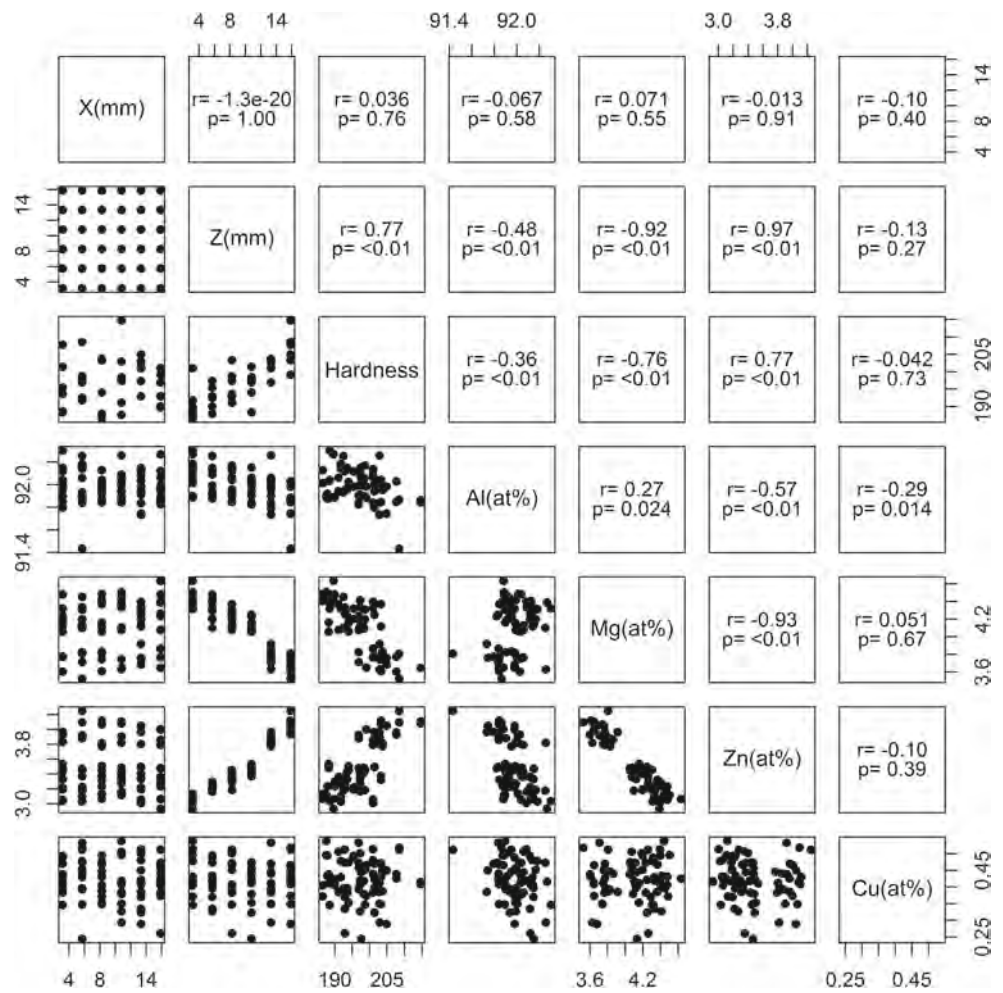
The Hall-Petch relationship [32, 33] describes the changes in strength due to changes in grain size, which simplifies to the **ISQR** functional form in Table 2. In conclusion, strengthening mechanisms provided two functional forms, which led to the modification of netSEM package v0.4.2.01 to v0.4.7 [18].

## Data Collection

Samples ( $19 \times 19 \times 4$  mm) were machined from the FGM plate such that the compositional gradient was parallel to the 19-mm edge (ND(z) in Fig. 2). Samples were mechanically polished through 1  $\mu$ m diamond paste using an Allied Multiprep System, culminating with a final vibratory polish with 0.05  $\mu$ m colloidal silica. A Buehler 1600-4963 indenter with 300 gf load was used to measure Vickers hardness (HV) on polished samples as a function of position along the compositional gradient (z-direction). The direct measurement of localized yield strength was not done due to the constraints imposed by sample dimensions and the compositional gradient of the FGM. The relation  $HV \cong 3\sigma_y$  [34], where HV and  $\sigma_y$  have the units of MPa, was used for the transformation. The compositional gradient was characterized using X-ray Energy Dispersive Spectroscopy (XEDS) in an FEI Nova NanoLab 200 at 10 KV beam energy. All parameters like working distance, accelerating voltage, current, dwell time, and dead time were held constant so that XEDS measurements from different samples/locations could be compared without bias.

Data for conducting exploratory data analysis, and model training, was collected at six different RD(x)-locations (same composition along RD(x)) for each of the six different ND(z)-locations (different composition along ND(z)), which reflects the composition and hardness data plotted in the pair-wise correlation plot in Fig. 3. For each

**Fig. 3** Pairs-wise linear correlation plot between composition and performance variables for exploratory data analysis (training dataset). X and Z represent the sample orthogonal coordinate system where the compositional gradient is along the ND(z)-axis as indicated in Fig. 2.  $r$  and  $p$  values in the upper triangle are linear correlation coefficients and probability value from hypothesis testing, calculated using R-package “stats” [35] respectively



scatter plot in the lower triangle of a pair-wise correlation plot (Fig. 3), the y-axis variable is defined by the variable listed horizontally on the matrix main diagonal, and the x-axis variable is defined by the variable listed vertically on the matrix main diagonal. Similarly, the  $r$  and  $p$  values for a linear fit between variables are presented in the upper triangle in a mirrored fashion. Once developed on training data, the network models were validated against an additional test dataset. The test dataset was collected on a different sample, at different sampling density (three ND(z)-levels with eleven different locations at each ND(z)-level).

Comparison of the scatter plots of Mg(at%) - Z(mm), and Zn(at%) - Z(mm), variable pairs in Fig. 3 shows that the compositional gradients are approximately linear along the Z-axis ( $|r| > 0.92$ ). The scatter plot of Mg(at%) - Zn(at%) show that the gradient in Mg concentration and Zn concentration is not independent, but inversely related ( $|r| > 0.93$ ). The Cu(at%) and to a lesser extent the Al(at%) concentration are invariant with Z(mm) position ( $|r| < 0.48$ ), and therefore not significant components for univariate network modeling analysis. The Hardness - Z(mm)

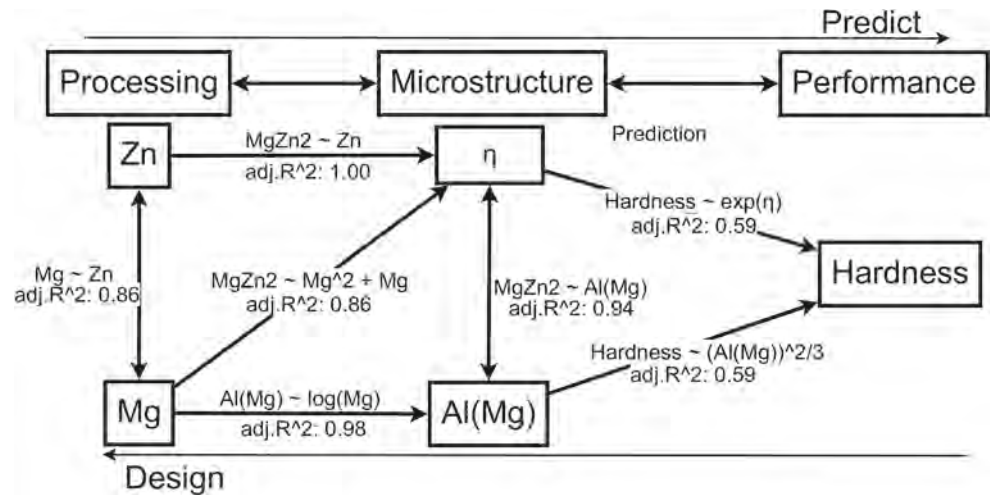
variable pair plot shows a continuous increase in strength as the compositional gradient changes from Mg-rich to Zn-rich. Visual interpretation of the scatter plots indicates that correlations between Hardness - Mg(at%) and Hardness - Zn(at%) show higher scatter as compared to Hardness - Z(mm) variable pair plot. This could reflect the aggregate sum of uncertainties of two different techniques used to collect datasets. The grain size data was not collected under this study.

## Physics-Informed Network Model

The modified netSEM package (v0.4.7) uses the general functional forms and the functional forms from strengthening mechanism domain knowledge in Table 2 to assess the most significant relationship (highest adj- $R^2$  value) for each variable pair. The potential precipitate for the FGM have a specific stoichiometry, which can be used for evaluating the distribution of atoms. The aging experiments [27] show  $MgZn_2$  ( $\eta$ ) as a potential precipitate. The volume



**Fig. 4**  $\langle P|M|P \rangle$  network model of the training dataset. Units are in at% and VHN for compositional and hardness variables, respectively. The best functional form, with adj- $R^2$  values for the different variable pairs, is presented



fraction of  $\eta$  precipitates require a 2:1|Zn:Mg stoichiometry and therefore the volume fraction of precipitates are limited by the minimum of either the Zn/2 or Mg concentrations. For the compositional range of the FGM, this minimum is always Zn/2; therefore,  $\eta$  concentration is proportional to the Zn(at%) concentration. This means that no Zn was retained in solid solution, and  $\eta$  and Al(Mg) (i.e., Mg retained in solid solution) were used as variables for network model.

Un-biased analysis of the uni-variate correlations with netSEM package indicates that Hardness exhibits the highest adj- $R^2$  between (1) an exponential relationship with Zn and (2) a logarithmic and (3) solid solution relationships (both have the same adj- $R^2$ ) with Al(Mg) concentration. Visualization of the data regime indicate that functional forms (**Log** and **SS**) have similar shapes within ranges of the variables collected. The domain knowledge biases the selection of **SS** functional form for the network model between Al(Mg) and Hardness over the selection of the **Log** function, as it is consistent with the Mg concentration primarily contributing through solid-solution hardening. The **Exp** function with respect to the  $\eta$  concentration could be consistent with diffusional mechanisms controlling the precipitate distribution and thus the strengthening behavior. The simplified network model derived from the netSEM quantification is shown in Fig. 4

and rank ordered relationships for each variable pair can be found in the supplemental materials in the data repository <https://cwru-msl.github.io/MRL17/>. The nodes in the Fig. 4 indicate that a uni-variable model of  $\eta$  or Al(Mg) can account for 0.59 of the variability in the training data as indicated by their adj- $R^2$  values.

The predictive power of the  $\eta$  and Al(Mg) models (Table 3) were validated against a test data set. This was done by comparing measured hardness values from a test dataset with predicted hardness values calculated using models and measured compositional values of test dataset. Mg and Zn concentrations are inversely dependent by alloy design, therefore adding complexity to the model (i.e., compare the  $\eta$  and Al(Mg) +  $\eta$  models), does not increase model predictive power. The combination model Table 3: Al(Mg) +  $\eta$ , imposes simultaneous mechanisms (Mg: solid solution, and Zn: precipitation) as dictated by domain knowledge but provides no statistical benefit to the predictive power of the models as indicated by no change in predictive- $R^2$  or the Akaike Information Criterion (AIC) values (which are two measures of the relative quality of statistical models). Predictive- $R^2$  is a variation of  $R^2$  which adjusts  $R^2$  to determine how well the model predicts for new observations [36]. The first two models in Table 3 can be used for prediction, as long as Al(Mg) and  $\eta$  concentration in the FGM remains inversely dependent to each other.

**Table 3** Comparison of predictive models of hardness ( $y$ ) with measures such as predictive- $R^2$  and AIC values.  $\epsilon$  is the mean-zero random error term

Variables	Functional form ( $y=$ )	Predictive- $R^2$	AIC
$\eta$	$169.64 + 4.72 * \exp(\eta) + \epsilon$	0.57	416.35
Al(Mg)	$235.90 - 22.10 * Al(Mg)^{2/3} + \epsilon$	0.58	415.20
Al(Mg) + $\eta$	$208.95 - 13.24 * Al(Mg)^{2/3} + 1.96 * \exp(\eta) + \epsilon$	0.58	416.14



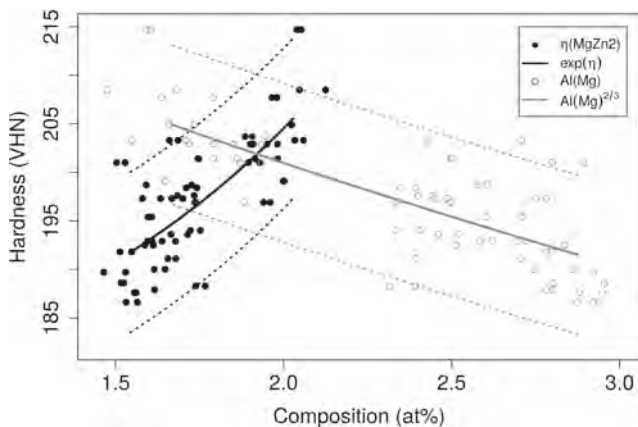
## Discussion

Data-driven netSEM modeling provides insight into the statistically relevant pathway relationships and uni-variate functional forms between composition and performance, but it is essential to consider the limitations and applicability of the approach to design. Hardness - Mg(at%) variable pair plot in Fig. 3 is counter to the simple expectation that increasing solute concentration (Mg(at%)) increases hardness since the Zn-Mg concentrations are not independent. Therefore, extrapolating the network model to different systems where the Zn/Mg concentrations are independently controlled would not be appropriate. For example, AA-5059 [25] which contains 6–6.5 at% Mg and 0.2–0.3 at% Zn (nominally the rest Al) has an average measured hardness value of  $115.5 \pm 1.4$  VHN at 300 gf while the  $\eta$  model and Al(Mg) model predict a hardness of 175 VHN and 164 VHN, respectively. This lack of correlation indicates that there are additional latent variables that need to be considered for the model to be generalizable to other aluminum alloy products.

## Scatter

High variability in the hardness values as a function of composition for the training dataset (Fig. 5) explains the low adj- $R^2$  value of the models (Table 3). The variability not captured by the models could arise from two different uncertainties: aleatoric and epistemic.

Aleatoric uncertainty is caused by visual assessment of indent size at a particular magnification, or uncertainty in the compositional measurements. Epistemic uncertainty is due to hidden or latent variables (e.g., grain size). To check aleatoric uncertainty, first, hardness measurement uncertainty was assessed by measuring the same indent ten



**Fig. 5** Models (lines) do not fit the high scatter in the Hardness measurements as a function of compositional variables (points). Top and bottom dotted lines represent 95% confidence interval of each model

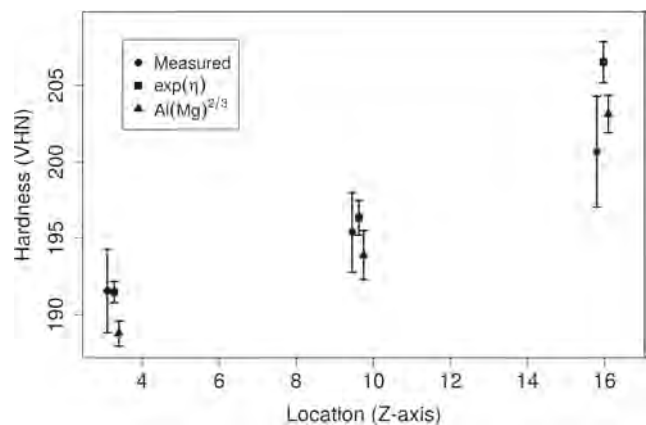
times, which amounted to  $\pm 1.4$  VHN, whereas scatter in the hardness measurements (Fig. 5) is 2.4 order of magnitude greater than this aleatoric uncertainty. Next, aleatoric uncertainty propagated through the models due to variation in compositional measurements was explored. For each compositional measurement, the hardness was predicted based on the models (Table 3). Presented in Fig. 6 is the average and standard deviation, for both measured and predicted hardness values for each model.

The comparison (Fig. 6) provides a measure of the spread in the model predictions based on the observed compositional variation. Figure 6 confirms that the variability in the compositional measurements is insufficient to describe the variability in the measured hardness values. Therefore, the additional variability in the measured hardness values must be attributed to epistemic uncertainty (hidden or latent variables).

## Latent Variables

Grain size could be the additional variable needed to capture the variability in hardness. As indicated in Fig. 6, the scatter in the hardness values appears random at a particular composition. This shows that a missing variable should account for this variation, while composition remains constant. From processing history (also Fig. 2), we can infer that the grain size is dependent on the composition; therefore, adding grain size as a variable using conventional Hall-Petch relationship would not account for the scatter in the data seen in Fig. 6.

Inverse pole figure in Fig. 2 shows the grain size in the material is of the same order of magnitude as of the size of the hardness indents, i.e., the black diamonds in the corners of the images. Therefore, each indent effectively samples a few crystallites instead of enough grains to get a homogeneous measure of the properties needed to utilize



**Fig. 6** Predicted hardness values and uncertainties for the  $\eta$  and Al(Mg) models based on measured compositional gradient in comparison to measured hardness values/variability for the test dataset

Hall-Petch or Taylor factors [37]. This could introduce a scatter to the hardness, localized strength.

The yield strength of an arbitrarily oriented single crystal (in fcc metals) should range between 2 and 3.57 times the critical resolved shear stress (CRSS) [38], resulting in a 56% difference in potential strengths. The variation in measured values can be calculated with the help of 95% confidence interval shown in Fig. 5. For example, at 1.5 at% of  $\eta$ , hardness value ranges between 183.5 and 200 VHN, with mean at 191.75 VHN. On calculating, this amounts to 8% variation. This “sudo-single crystal sampling” of a textured material could describe the inherent random scatter in the hardness values, as the 95% confidence limits on the measured values show a variation of 8% difference in hardness values.

## Conclusions

The netSEM modeling approach was applied to assess the strength of coupling coefficients in functional forms to develop a diagnostic  $\langle P|M|P \rangle$  model with predictive capabilities. This domain-guided diagnostic model (in Table 3) was trained and validated over a limited compositional range, showing statistical significance with an predictive- $R^2$  of 0.57 ( $\eta$ ) and 0.58 ( $Al(Mg)$ ). The methodology was able to capture the SS function for excess Mg in solid solution and an Exp function for  $\eta$  concentration with the hardness of the material. These functional forms capture the internal mechanisms and reflect why the material behaves the way it does, consistent with domain knowledge of alloy strengthening mechanisms. A comparison between ANN and netSEM was presented to allude to the values of conducting one or both network modeling approaches depending on user requirements. Though not explicitly shown in this paper, the netSEM methodology is generalizable to multi-variable problems where interpretation of the meaning of relationship pathways is important, while also allowing the user to tailor the tested functional forms to the system under investigation [39].

**Acknowledgments** The material for this study was provided by the Alcoa/Arconic Alloy Technology Division. This work was funded by the Army Research Office Short Term Innovative Research program: W911NF-14-0549. The data and R-analytics code utilized for the paper can be found at <https://cwru-msl.github.io/MRL17/>.

## References

- (2011) Materials genome initiative for global competitiveness. Tech. rep. <https://www.mgi.gov/about>
- Arnold SM, Holland F, Gabb TP, Nathal M, Wong TT (2013) American institute of aeronautics and astronautics. <https://doi.org/10.2514/6.2013-1850>
- Arnold SM, Holland F, Bednarczyk BA (2014) American institute of aeronautics and astronautics. <https://doi.org/10.2514/6.2014-0460>
- Miracle D, Majumdar B, Wertz K, Gorsse S (2017) Scr Mater 127(Supplement C):195. <https://doi.org/10.1016/j.scriptamat.2016.08.001>. <http://www.sciencedirect.com/science/article/pii/S1359646216303657>
- Lingerfelt EJ, Belianinov A, Endeve E, Ovchinnikov O, Somnath S, Borreguero JM, Grodowitz N, Park B, Archibald RK, Symons CT, Kalinin SV, Messer OEB, Shankar M, Jesse S (2016) Procedia computer science 80(Supplement C):2276. <https://doi.org/10.1016/j.procs.2016.05.410>. <http://www.sciencedirect.com/science/article/pii/S1877050916308869>
- Kalidindi SR, De Graef M (2015) Materials Data Science: Current Status and Future Outlook. Annu Rev Mater Res 45(1):171–193. <https://doi.org/10.1146/annurev-matsci-070214-020844>
- Jacobsen MD, Benedict MD, Foster BJ, Ward CH, Foster BJ, Jacobsen MD, Benedict MD (2015) In: Proceedings of the 3rd world congress on integrated computational materials engineering (ICME 2015). Springer, Cham, pp 285–292. [https://doi.org/10.1007/978-3-319-48170-8\\_34](https://doi.org/10.1007/978-3-319-48170-8_34). [https://link.springer.com/chapter/10.1007/978-3-319-48170-8\\_34](https://link.springer.com/chapter/10.1007/978-3-319-48170-8_34)
- Jacobsen MD, Fourman JR, Porter KM, Wirrig EA, Benedict MD, Foster BJ, Ward CH, Jacobsen MD, Foster BJ, Fourman JR, Benedict MD, Porter KM, Wirrig EA (2016) Creating an integrated collaborative environment for materials research. Integr Mater Manuf Innov 5(1):12. <https://doi.org/10.1186/s40192-016-0055-2>. <https://link.springer.com/article/10.1186/s40192-016-0055-2>
- Havlin S, Kenett DY, Ben-Jacob E, Bunde A, Cohen R, Hermann H, Kantelhardt JW, Kertész J, Kirkpatrick S, Kurths J, Portugali J, Solomon S (2012) The European Physical Journal Special Topics 214(1):273. <https://doi.org/10.1140/epjst/e2012-01695-x>. <http://link.springer.com/article/10.1140/epjst/e2012-01695-x>
- Nibbe RK, Koyutürk M, Chance MR (2010) PLOs Comput Biol 6(1):e1000639. <https://doi.org/10.1371/journal.pcbi.1000639>
- Steinhaeuser K, Chawla NV, Ganguly AR (2011) Statistical Analysis and Data Mining 4(5):497. <https://doi.org/10.1002/sam.10100>. <http://onlinelibrary.wiley.com/doi/10.1002/sam.10100/abstract>
- Bhadesia HKDH (1999) ISIJ Int 39(10):966. <https://doi.org/10.2355/isijinternational.39.966>
- Bhadesia HKDH (2009) Neural Networks and Information in Materials Science. Statistical Analy Data Mining 1(5):296–305. <https://doi.org/10.1002/sam.10018>. <http://onlinelibrary.wiley.com/doi/10.1002/sam.10018/abstract>
- Sha W, Edwards KL (2007) Mater Des 28(6):1747. <https://doi.org/10.1016/j.matdes.2007.02.009>. <http://www.sciencedirect.com/science/article/pii/S0261306907000520>
- Hoyle RH (2012) Handbook of structural equation modeling, 1st edn. The Guilford Press, New York
- French RH, Podgornik R, Peshek TJ, Bruckman LS, Xu Y, Wheeler NR, Gok A, Hu Y, Hossain MA, Gordon DA, Zhao P, Sun J, Zhang G-Q (2015) Curr Opinion Solid State Mater Sci 19(4):212. <https://doi.org/10.1016/j.cossms.2014.12.008>. <http://www.sciencedirect.com/science/article/pii/S1359028614000989>
- Faraway JJ (2004) Linear models with R, 1st edn. Chapman and Hall/CRC, Boca Raton
- Wheeler N, Xu Y, Du W, Gok A, Ma J, Bruckman L, Elsaeti M, Sun J, French R (2013) Semi-supervised generalized structural equation modeling. <https://github.com/vuvlab/sgsem>

19. Mitchell TM (1997) Machine learning, 1st edn. McGraw-Hill Education, New York
20. White H, Gallant AR, Hornik K, Stinchcombe M, Wooldridge J (1992) Artificial neural networks: approximation and learning theory, illustrated edition edn. Blackwell Pub, Oxford
21. Skinner AJ, Broughton JQ (1995) Model Simul Mater Sci Eng 3(3):371. <https://doi.org/10.1088/0965-0393/3/3/006>. <http://stacks.iop.org/0965-0393/3/i=3/a=006>
22. Rioja RJ, Sawtell RR, Chu MG, Karabin M, Cassada WA, Karabin M (2012) Functional Gradient Products Enabled by Planar Solidification Technologies. Springer, Pittsburgh, pp 1383–1388. [https://link.springer.com/chapter/10.1007/978-3-319-48761-8\\_211](https://link.springer.com/chapter/10.1007/978-3-319-48761-8_211). [https://doi.org/10.1007/978-3-319-48761-8\\_211](https://doi.org/10.1007/978-3-319-48761-8_211)
23. Chu M, Giron A, Cassada W (2012) The Minerals, Metals and Materials Society, Pittsburgh, pp 1367–1375
24. Chu MG, Yu H, Giron A, Kallaher K (2007) Method of unidirectional solidification of castings and associated apparatus. <http://www.google.com/patents/US7264038>. U.S. Classification 164/122.1, 164/133, 164/337; International Classification B22D35/04, B22D27/04, B22D37/00, B22D35/00; Cooperative Classification B22D27/045, B22D7/00; European Classification B22D27/04A
25. (2015) International alloy designations and chemical composition limits for wrought aluminum and wrought aluminum alloys. The Aluminum Association, Arlington, pp 38. [http://www.aluminum.org/sites/default/files/TEAL\\_1\\_OL\\_2015.pdf](http://www.aluminum.org/sites/default/files/TEAL_1_OL_2015.pdf)
26. Deshpande NU, Ray KK, Mallik AK (1986) 2:108. [http://mio.asminternational.org/apd/viewPicture.aspx?dbKey=grantami\\_apd&id=10757508&revision=421381](http://mio.asminternational.org/apd/viewPicture.aspx?dbKey=grantami_apd&id=10757508&revision=421381)
27. Balderach DC, Hamilton JA, Leung E, Cristina Tejeda M, Qiao J, Taleff EM (2003) Mater Sci Eng A 339(1–2):194. [https://doi.org/10.1016/S0921-5093\(02\)00158-2](https://doi.org/10.1016/S0921-5093(02)00158-2). <http://www.sciencedirect.com/science/article/pii/S0921509302001582>
28. Fleischer RL (1963) Acta Metallurgica 11(3):203. [https://doi.org/10.1016/0001-6160\(63\)90213-X](https://doi.org/10.1016/0001-6160(63)90213-X). <http://www.sciencedirect.com/science/article/pii/000161606390213X>
29. Labusch R (1970) Phys Stat Sol (b) 41(2):659. <https://doi.org/10.1002/pssb.19700410221>. <http://onlinelibrary.wiley.com/doi/10.1002/pssb.19700410221/abstract>
30. Orowan E (1948) Institute of metals, London, pp 451–453
31. Guo Z, Sha W (2002) Mater Trans 43(No. 6):1273
32. Hall EO (1951) Proc Phys Soc B 64(9):747. <https://doi.org/10.1088/0370-1301/64/9/303>. <http://stacks.iop.org/0370-1301/64/i=9/a=303>
33. Petch NJ (1953) J Iron Steel Inst Jpn 174:25
34. Tiryakioğlu M (2015) Materials science and engineering: A 633:17. <https://doi.org/10.1016/j.msea.2015.02.073>. <http://www.sciencedirect.com/science/article/pii/S0921509315002014>
35. R Core Team (2015) R: a language and environment for statistical computing, R Foundation for Statistical Computin, Vienna. <https://www.R-project.org/>
36. RPubs - Predictive R-squared according to Tom Hopper. <https://rpubs.com/RatherBit/102428>
37. Bunge H (1970) Krist Techn 5:145. <https://doi.org/10.1002/crat.19700050112>
38. Schmid E, Boas W (1968) Plasticity of crystals with special reference to metals. Chapman & Hall, London
39. Bruckman LS, Wheeler NR, Ma J, Wang E, Wang CK, Chou I, Sun J, French RH (2013) IEEE Access 1:384. <https://doi.org/10.1109/ACCESS.2013.2267611>. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6527980>