Variance Stabilised Optimisation of Neural Networks: A Case Study in Additive Manufacturing

Scott V. Notley
Automatic Control and Systems Engineering
University of Sheffield
Sheffield, United Kingdom
s.v.notley@sheffield.ac.uk

Yunhui Chen
Dept. of Mechanical Engineering
University College London
London
yunhui.chen@ucl.ac.uk

Peter D. Lee
Dept. of Mechanical Engineering
University College London
London
peter.lee@ucl.ac.uk

George Panoutsos
Automatic Control and Systems Engineering
University of Sheffield
Sheffield, United Kingdom
g.panoutsos@sheffield.ac.uk

Abstract—A new framework is presented for training neural networks that is based on the characterisation and stabilisation of measurement variations. The framework results in a number of useful properties that maximises the use of data as well as aiding in the interpretation of results in a principled manner. This is achieved via variance stabilisation and a subsequent standardisation step. The method is a general approach that may be used in any context where repeatability data is available. Standardisation in this manner allows goodness of fit to be quantified and measurement data to be interpreted from a statistical perspective. We demonstrate the utility of this framework in the analysis of advanced manufacturing data.

Index Terms—Variance stabilisation, neural network, multilayer perceptron, reduced chi-squre, chi-square per degree of freedom, metal additive manufacturing

In this work a neural network is employed, as a generalised regressor, to investigate the relationship between metal additive manufacturing (AM) process parameters and the resultant meltpool geometrical properties for IN718 super alloy. The analysis of additive manufacturing data is here presented as a use case, but the framework itself is general and may be used in any approach where repeatability data is available. Additive manufacturing is a 3D printing process that builds components in a layer-by-layer approach; the meltpool is the volume of molten feedstock and substrate material. Understanding the underlying physics and relationships between material and meltpool is key to process optimisation, however opportunities for in-situ measurements are limited, hence fundamental process understanding is lacking. The use of neural networks in analysing advanced manufacturing process data presents particular difficulties since gathering high quality data is expensive, complex and requires careful planning. This generally results in datasets with a low number of samples [1], [2], that need a systematic methodological approach to aid robust interpretation.

Recent articles have reported attempts at using neural networks, and in particular, the multilayer perceptron (MLP) [1]–[4] in analysing AM data. However, the fundamental issues of small data samples, and crucially, that of characterising and accounting for uncertainties are not addressed. In this work a new methodological framework is proposed, primarily based on accounting for the properties of stochastic variations in measurements, that makes full use of the available data, allows for robust training, and aids in the interpretation of results. From an application-specific perspective this enables high confidence insights into the relationship between process parameters and the fundamentals of transient thermomechanical mechanisms during the AM process, such as meltpool dynamics, and their influence on the final product properties.

Additive Manufacturing is a highly versatile manufacturing approach that generates complex geometries directly from digital design. Directed Energy Deposition Additive Manufacturing (DED-AM), which deposits powder or wire feedstock through a nozzle and melts it with a laser, is one of the most cost-effective AM methods and is specifically used for producing large near-net-shape [5] freeform components and for the repair of high-value components in aerospace, biomedical and automotive industries.

The dataset used in this study was designed to have multiple repeated measurements that allow variations in meltpool morphology to be investigated using Bland-Altman plots [6]. It was found that the stochastic variations in the measurement variations are heteroscedasitc and correlate with the measurement value itself. It is well known that for linear systems, via the Gauss-Markov theorem, that the ordinary least squares (OLS) solution results in the best linear unbiased estimate (BLUE), when the uncertainties are uncorrelated, expectation value of zero and equal variance (homoscedastic). Under a relaxation of these assumptions, as with heteroscedastic variations, the BLUE property can no longer be guaranteed. It may be noted that normality of the residual distribution

is not a requirement to attain the BLUE estimator, however, normality leads to other desirable properties that will be discussed below. For non-linear estimation, as in the estimation of neural network parameters, many of the properties derived from linear estimation theory hold, at least asymptotically [7].

In the literature in general, two approaches, are commonly used to deal with heteroscedasticity: weighted least squares (WLS) and non-linear data transformations. Weighted least squares appeals to the likelihood function, for normally distributed uncertainties, to derive a weight for each term of the error function to be minimised [8]. The weight is the inverse variance of the measurement uncertainty for each of the dependent variable data points. In contrast, transformation methods attempt to transform the data to a domain where the assumptions of the estimation method are inherently met. Pre-processing and transformation of data is not uncommon in the use of neural networks, for instance, scaling, standardisation, compression and decorrelation of input data [9]–[11]. In this work we specifically consider variance stabilisation, to transform the stochastic variations, such that the data meets the ordinary least squares assumptions. To the authors knowledge variance stabilisation has not been previously applied to training of neural networks.

The main purpose of many transformation methods, where the relationship between mean and variance is known (empirically or theoretically) [12]–[14], is variance stabilisation to obtain a measurement uncertainty that is constant. The method of *Box and Cox* [15] defines a family of transformations and optimises a joint likelihood function, with respect to the transform and model parameters, that maximises the likelihood of the *residuals* of a *linear model* being drawn from a normal distribution. The method requires the estimation of model parameters over a range of transformations. While the underlying theory is not new, it poses a significant limitation with respect to neural networks trained with gradient descent-based optimisation methods, both in terms of computational time and also, in terms of ensuring consistent minima are found over the range of transformations.

To address this challenge, the use of a repeated measures dataset allows a more pragmatic approach to be taken since transformation of the difference of repeats can be made prior to and independent of the estimation of model parameters. Variance stabilising transforms generally have a small number of parameters to be estimated. In this case we use a power law with a single parameter, that is found empirically using a Shapiro-Wilks test for normality.

As observed by *Box and Cox*, variance stabilising transformations not only result in data that more closely meets assumptions but can also be used to find a metric that more succinctly expresses the results. Stabilisation and normalisation (in the sense of transforming to a normal distribution) allows the distribution of the stochastic variations to be robustly summarised with a single statistic; the variance (assuming expectation to be zero). In terms of a neural network the dependent output variable may be standardised by the standard deviation of the *variations* resulting in an output space that has

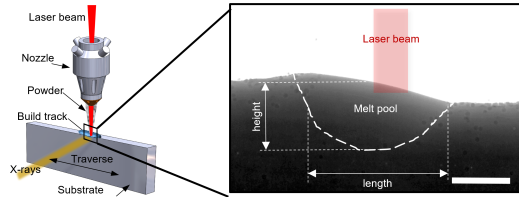| Laser Power (W)	100	150	200
Head Speed (mm/s)	1 & 2.5	1 & 2.5	1 & 2.5
Flow rate (g/min)	1,2, & 3	1, 2, & 3	1, 2, & 3

Fig. 1. Schematic of the meltpool measurement process from the x-ray images

a z-score style of statistical interpretation; the units may now be expressed in standard deviations of 'noise'. Standardisation of the dependent variable in this manner also has implications in interpreting reliability of the regression found by the neural network in that the squared error cost function is now expected to follow a reduced chi-square distribution. This property may be used to objectively assess goodness of fit of the neural network and may further be used, in this sense, as a stopping criterion for the training process without the need for cross validation.

I. METHODS AND DATA

A. Data

The data used in this case study is collected from Synchrotron in situ x-ray imaging of the DED-AM process. A DED-AM process replicator, which faithfully replicates a commercial DED-AM system, was used to capture the key physics during the laser matter interaction of IN718 on a synchrotron beamline. Melt pool morphology (length, height and volume) was mapped from a process map with a full range of process parameters including laser power, traverse speed and powder feedrate. Figure 1 shows a schematic of the experimental set-up and details of the process may be found in [16]. Experiments were performed at a number of combinations of the process parameters shown in Table 1.

For each experiment 3 repeated measurements of the meltpool were made from the x-ray images for length and height. Each measurement was made at fixed points, along each track, spatially separated to ensure no correlation between measurements due to dynamics. Volume estimates were also calculated, as in [16], from the x-ray measurements.

B. Bland-Altman Plots

Bland-Altman plots were originally proposed as a method of comparing the agreement between two measurement methods by studying the mean difference [6], [17]–[20]. The x-axis of the plots is the average of two measurements and the y-axis show the difference between two measurements. Thus,

the spread in the y-axis direction shows the behaviour of measurement variations as a function of the average of the measurements. The method was further proposed as a method of assessing measurement error [21] in terms of bias and heteroscedasticity [22], [23] via repeated measurements.

It is assumed that errors due to the actual measurement process are negligible in comparison to the natural variations occurring in the meltpool. It is also assumed that the variations in the meltpool are stochastic and not dynamically related to previous measurements made along the same track, for example due to material deposition variation.

C. Variance Stabilisation

Optimisation methods often attempt to minimise a least squares loss function. Inherent in the assumptions of a least square approach, via the link to the maximum likelihood approach, is the assumption that the residuals are normally distributed and have constant variance across the measurement domain. Deviations from these assumptions leads to loss of statistical efficiency in estimation of model parameters and possible misleading results e.g. measurements with large uncertainty can unduly weight regressions in comparison to more accurate measurements with less stochastic perturbations.

An approach to handling data with non-constant variability and non-Gaussian distribution, is to use variance stabilisation techniques [15] to non-linearly transforms the variables to a space where the noise has constant measurement accuracy; known as the equal variance domain. After transformation the data is in a form more closely matched to the assumptions in an ordinary least squares setting and is here referred to as transformed OLS (TOLS).

We define a transformation function, similar in nature to that of *Box and Cox* [15], as $f(x; \theta) = x^{\theta}$. The value of θ for each measurement was found empirically such that a Shapiro-Wilk test [24] indicated the *difference of repeats* has a distribution that is Gaussian. Since this method is reliant on the difference of repeats, the transformation function is independent of the model and may be found prior to training. It may be noted that this method requires the estimation of only two parameters, the transform and a variance, from the entire dataset in contrast to a method such as WLS that requires multiple variance estimates from *subsets* of the full dataset.

D. Standardisation

In general measurements have an associated accuracy that is dependent on the measurement process itself and other factors. In this work we consider the variations in each measured variable, as determined by the repeatability differences, to be a summation of variances from all non-deterministic sources and/or processes not directly associated with the process parameter.

It is common to scale variables by their standard deviation as a method of standardisation, however, this does not distinguish between signal power and noise power. Due to *stabilised* repeatability measurements, with constant variance, signals may be standardised by the standard deviation of the *repeatability* noise, as in equation 1, such that

$$\hat{x}_n = \frac{f(x_n; \theta)}{\sigma_{f(\eta;\theta)}} \tag{1}$$

where x_n is a single variable, η is the repeatability differences and $\sigma_{f(\eta;\theta)}$ is estimated as

$$\sigma_{f(\eta;\theta)} = \sqrt{\frac{1}{N} \sum_{n}^{N} (f(x'_n; \theta) - f(v_n^*; \theta))^2} \tag{2}$$

where x'_n is the n-th measurement and x_n^* is the corresponding n-th repeated measurement. Since this is a repeated measurement the mean difference is assumed to be zero.

In this form measurements may now be expressed in units of standard deviations of 'noise' [1]. This allows disparate, and non-commensurate, measurements with differing units and accuracies to be compared and considered in a principled manner. For instance, differences in this standardised space now have an interpretation that is similar to a z-score or a Welch's t-test.

E. Neural based Regression Analysis

To understand the relationship between the process parameters and the resultant meltpool morphology measures, neural networks are employed as a generalised regressors. The structures used are multilayer perceptrons (MLPs) with a single hidden layers, consisting of Tanh activation functions, and an output layer consisting of a linear activation function. Three separate networks were trained, with the process parameters of laser power, head speed and powder flow rate as inputs; outputs of each network were length, height and volume respectively.

The network was trained by minimisation of the least squares error

$$E = \sum_{n=1}^{N} (g(\mathbf{p}_n; \boldsymbol{\phi}) - \hat{x}_n)^2 \tag{3}$$

using the backpropagation algorithm, in stochastic mode, where $g(\cdot, \boldsymbol{\phi})$ is the network function, $\boldsymbol{\phi}$ is the network parameters, $\mathbf{p}$ is the input vector of process parameters and $\hat{x}_n$ is the stabilised and standardised meltpool measurements given by (1). Due to the process of stabilisation and standardisation the squared error is expected, at the optimal parameters, to follow a reduced chi-squared distribution (chi-square per-degree-of-freedom) given by

$$\chi_{red}^2 = \chi_{N-k}^2 = \frac{E}{N-k} \tag{4}$$

where N is the number of training data points and k is the number of degrees of freedom of the network. The work of *Bartlett* [25] showed that for large networks the generalisation

---

[1] Noise in this sense refers to stochastic variations that are primarily natural meltpool variations.

performance depends on the size of weights and not the size of the network and gives some justification to using networks with more parameters than the number of training data points. Building on this work Ingrassia and Morlini [26] further showed that the *equivalent* number of degrees of freedom, k, in a multilayer perceptron is $k \leq p + 1$ where p is the number of hidden units. For estimation of error variance, with with early stopping as a regularisation technique, they suggest setting $k = p + 1$.

In this work we make use of these properties and use χ^2_{red} as a measure of goodness of fit. For networks whose parameter sets are providing a good fit to the data the value of χ^2_{red} is close to 1. The variance of a chi-squared distribution with v degrees of freedom, χ^2_v, is given by $\sigma^2 = 2v$. Thus we consider values in the two sigma range of

$$(1 - 2\sqrt{2/(N-k)}) < \chi^2_{red} < (1 + 2\sqrt{2/(N-k)}) \quad (5)$$

to be a good fit, given the limits of the sample size, with a low risk of over- or under-fitting[2]. It should be noted that the analysis in this work is primarily interested in regression to gain insight into the relationship between the process parameters and meltpool morphology. As such the χ^2_{red} criterion, (5), is used purely as a goodness of fit measure and does not necessarily relate to generalisation.

II. RESULTS

The following sections show the results of applying the methodology described above to the AM data. In this application we wish to robustly identify any relationships that may exist between the process parameters of laser power, heed speed and powder flow rate and meltpool morphology summarised by length, height and volume measurements. In total 18 experiments were carried out with the DED-AM replicator, for all combinations of settings given in table I, with 3 repeat measurements giving a total of 54 training samples for each of 3 networks mapping process parameters to length, height and volume respectively.

A. Bland-Altman plots and Variance Stabilisation

Figure 2 shows the Bland-Altman plots for length, Height and Volume. The results of Shapiro-Wilk tests for normality are shown in table II; the results pre- and post- stabilising transformations are shown for comparison.

Figures 2(b) and (c) show clear evidence of variability that is related to the mean value (heteroscedastic), for height and volume, with smaller repeatability differences for small values and larger repeatability differences for large values. This type of variance characteristic tends to result in distributions that are leptokurtic (kurtosis > 3); sharp peak with long tails in

[2]The chi-square distribution is actually skew. For degrees of freedom above 50 it is acceptable to approximate the distribution as Gaussian. For degrees of freedom below this the true two sigma upper limit will be greater than $1 + 2\sqrt{2/(N-k)}$. Thus, the upper limit is safe for goodness of fit, however, some care must be taken with this approximation, especially with the lower limit, when the number of hidden units is comparable to the size of the training data set.
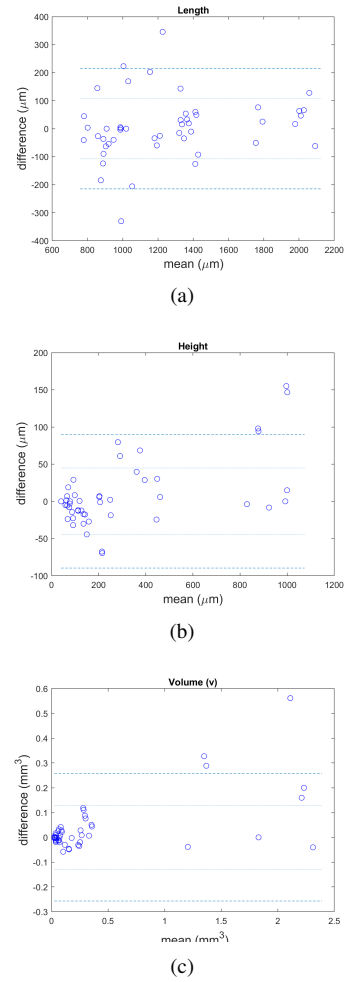

(a)

(b)

(c)

Fig. 2. Bland-Altman plots showing the noise properties as a function of the mean value for (a) length , (b) Base-Surface Height and (c) Volume. The horizontal lines show 1 and 2 standard deviations of the repeatability noise.

comparison to Gaussian. This is reflected in the Shapiro-Wilk tests showing clear evidence (p-values<<0.001) to reject the null-hypothesis of a normal distribution.

For the length measurements, figure 2(a), the measurement errors do not show any clear relationship to the mean values. The estimated kurtosis value is greater than 3 indicating leptokurtic behaviour but with the amount of samples available this estimate is an unreliable stand alone test. A transformation was neverless found that improved the results of the Shapiro-Wilks test.

Figure 3 shows the Bland-Altman plots after stabilisation by (1). These results show that after stabilisation there is no evidence of the meltpool variations being related to the mean values. This is reflected in table II with all Shapiro-Wilk scores closer to one after stabilisation and no evidence to reject the null-hypothesis at the 3σ level. It may also be noted that the kurtosis scores for height and volume are close to 3; however for the length measurments the kurtosis, although slightly improved, remains leptokurtic.

TABLE II
SHAPIRO-WILK SCORES, P-VALUES AND KURTOSIS FOR RAW AND STABILISED DATASETS. STABILISED DATASETS ARE INDICATED WITH A T IN PARENTHESIS. THE ESTIMATED STABILISATION PARAMETER FOR THE FUNCTION $f(x) = x^{\theta}$ IS ALSO SHOWN.
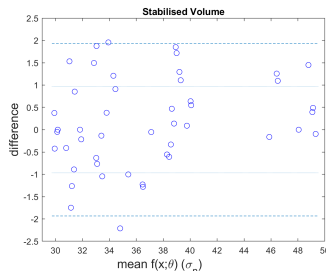
	S-W	p-value	kurtosis	θ	S-W (T)	p (T)	kurtosis (T)
height	0.84	$2e^{-5}$	5.8	0.2	0.97	0.24	2.7
length	0.92	$5e^{-3}$	5.3	1.6	0.94	0.02	4.9
volume	0.60	$3e^{-9}$	10.9	0.1	0.98	0.46	2.4
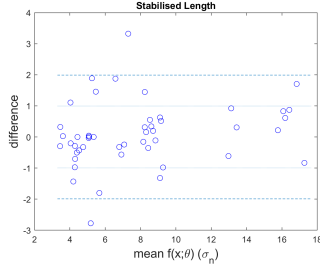
(a)
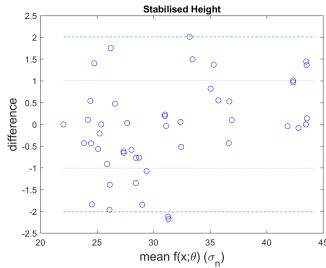
(b)

(c)

Fig. 3. Bland-Altman plots showing the noise properties as a function of the mean value for stabilised versions (a) length , (b) Base-Surface Height and (c) Volume.

B. Regression Analysis

The properties given by the reduced chi-square were also utilised in model order selection. For each network, multiple training runs were performed from random initial conditions, with increasing network size until a reduced chi-square, within the range give by (5), was found. Training was performed with a large number of epochs (1,000,000) to ensure that a representative minimum of the squared error was found. A learning rate of 0.000005 was used. This was chosen empirically to give the highest rate that gave smooth error curves.

TABLE III
THE ESTIMATED MODEL ORDER FOR MAPPING THE PROCESS PARAMETERS TO MELTPOOL HEIGHT, LENGTH AND VOLUME AND THE AN ASSOCIATED REPRESENTATIVE χ^2_{red} VALUE FOUND AFTER 1 MILLION TRAINING EPOCHS. FOR HEIGHT NO MAPPING WAS FOUND THAT COULD PROVIDE A SATISFACTORY FIT TO THE DATA.

|	N$^{\circ}$ hidden nodes	χ^2_{red}
height	N/A	>30
volume	7	1.13
length	5	1.36

Table III shows the results of systematically increasing the number of hidden nodes of the neural network until a chi-squared per degree of freedom could be found within the two sigma range given by (5). For the mappings to the meltpool height no reliable mapping was found.

Figure 4 shows the regression surfaces relating the stabilised and standardised length measurements to the flow rate and head speed at three power settings. For all power settings there is no significant relationship between the powder flow rate and the meltpool length and at low power (100W) there is no head speed that is able to significantly change the length of the meltpool. For a laser power setting of 144W there is an increase in the meltpool length, that is approximately 2.5 times the standard deviation of the noise, in reducing the head speed from 2.5mm/s to 1mm/s and is thus on the limits of being a significant change. At a laser power of 188W this change in the meltpool length increases to approximately a 10 standard deviation effect.

Figure 5 shows similar regression surfaces for the meltpool volume. The 'base' level for each regression is at approximately 33 standard deviations of noise. Thus, the meltpool volume has significantly larger ratio of volume to variations than the length. At laser power of 100W there is no combination of head speed or powder flow rate that significantly changes the meltpool volume. At 144W both the powder flow rate and the head speed can been seen to significantly change the meltpool volume with decreased head speed and increased powder flow increasing the meltpool volume. For the higher laser power of 188w only the head speed has a significant impact of the meltpool volume.
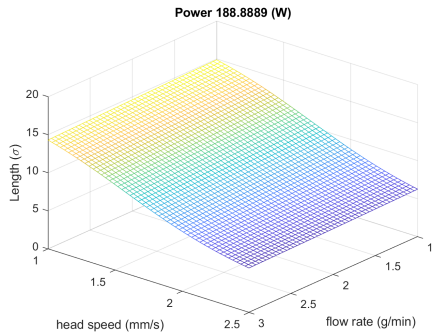
DISCUSSION AND CONCLUSIONS

In this work we have presented a new systematic methodological framework for training a MLP using repeatability data. The framework is general in its approach and may be applied to any datasets where repeatability information is
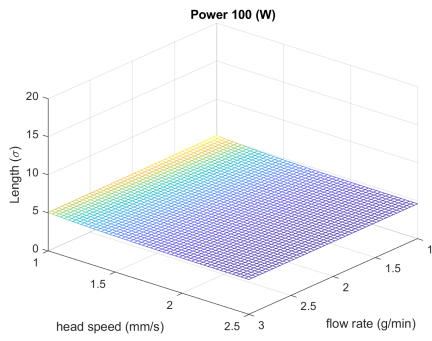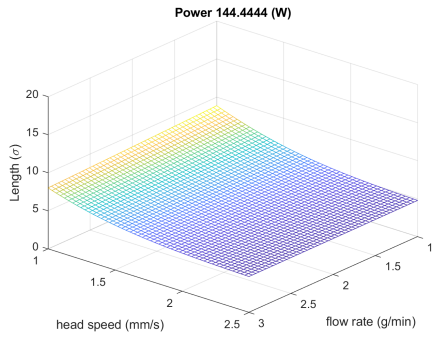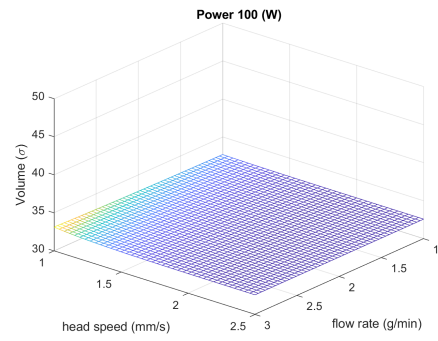
Fig. 4. Regression plots of head speed/flow vs standardised length at three different power settings.
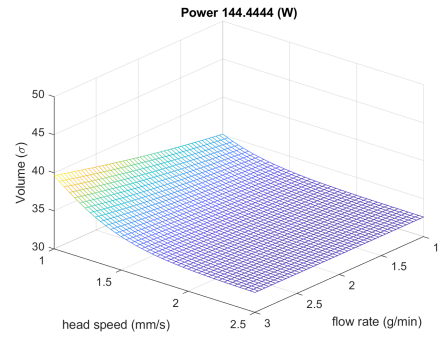

Fig. 5. Regression plots of head speed/flow vs standardised volume at three different power settings.

available. The approach uses variance stabilisation techniques to transform the *repeatability differences* to a domain where the variance is constant and has a distribution that is close to Gaussian. The well established method of *Box and Cox* [15] uses a similar transform, however, their approach attempts to optimise a transform such that the *residuals* of a linear model fit are stabilised. This is problematic in application to neural networks, with non-linear optimisation by gradient descent, both in terms of the computational load and in the finding consistent parameter optimisations across transforms. The transformation of repeatability differences allows the stabilisation to be performed independent of the model and prior to parameter optimisation.
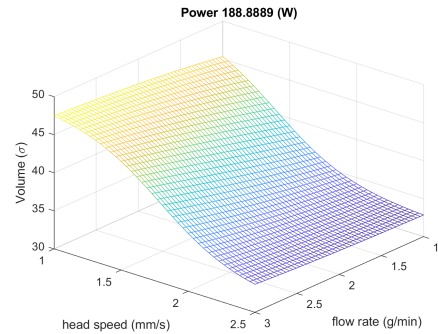
Attempting to maintain the properties of constant variance

and Gaussian distribution leads to a number of advantages and properties. Firstly, the data more closely matches the assumptions of least squares optimisation leading to more statistically efficient estimation of the network parameters. Statistical efficiency is of general importance but gains greater significance in applications with small datasets, such is in AM, where significant relationships between variables may require the full use of the information contained in the data to be revealed.

Secondly, the data may be standardised by the standard deviation of the stochastic variations. This standardisation means that the squared error cost function is now expected to follow a reduced chi-square distribution at the optimal solution. This may be utilised to assess the goodness of fit of

the regressions allowing poor fits to be rejected and not falsely interpreted as showing a significant input-output relationship. In this work, we further use this property as a criterion for early stopping of training and thus makes full use of the small dataset. We emphasise that this method is not necessarily an indication of generalisation and is here purely used as a goodness of fit measure for regression analysis.

Finally, standardisation, as above, allows output variables to be interpreted in terms of standard deviations of noise. This gives the output space a statistical interpretation that is similar to a z-score and allows only significant changes to be considered.

The framework has been employed on a repeated measures dataset derived from experiments investigating the laser-matter interface in advanced additive manufacturing processes. The neural approach in this work uses the MLP as a generalised regressor to gain insight into the dependencies between meltpool morphology and the process parameters. The work of *Caiazzo and Caggiano* [3] attempted to capture these dependancies, for 2024 Al Alloy, using a MLP but did not attempt to characterise or account for the nature of meltpool variations.

Variations in the meltpool morphology were investigated, independently of the model and prior to any training, using Bland-Altman plots. The analysis shows clear indication of heteroscedastic variations that correlate with the values itself and does not meet the assumptions of a least squares minimisation approach.

The variance of the measurements was stabilised by empirically finding a non-linear transformation that maximised the Shapiro-Wilk's test score for normality. The properties described above were then used to train neural networks to find good fits to the data if possible. The results showed that only the length and volume of the meltpool could be meaningfully related to the process parameters using this data. The framework also showed that the powder flow rate is not a significant factor in determining the meltpool length but does have significance in determining the volume of the meltpool.

ACKNOWLEDGMENT

REFERENCES

[1] C.Wang, X. Tan, S. B. Tor, and C. S. Lim, "Machine learning in additive manufacturing: State-of-the-art and perspectives," *Additive Manufacturing*, vol. 36, p. 101538, 2020.

[2] X. Qi, G. Chen, Y. Li, X. Chen, and C. Li, "Applying neural-network-based machine learning to additivemanufacturing: Current applications, challenges, and futureperspectives," *Engineering*, vol. 5, pp. 721–729, 2019.

[3] F. Caiazzo and A. Caggiano, "Laser direct metal deposition of 2024 al alloy: Trace geometry prediction via machine learning," *Materials*, vol. 11, no. 3, 2018.

[4] A. Singh, D. Cooper, N. Blundell, G. Gibbons, and D. Pratihar, "Modelling of direct metal laser sintering of eos dm20 bronze using neural networks and genetic algorithms," in *Proceedings of the 37th International MATADOR Conference*, 2012.

[5] T. Sercombe and G. Schaffer, "Rapid manufacturing of aluminum components," *Science*, vol. 301, pp. 1225–1227, 2003.

[6] D. G. Altman and J. M. Bland, "Measurement in medicine: The analysis of method comparison studies," *Journal of the Royal Statistical Society. Series D*, vol. 32, no. 3, pp. 307–317, 1983.

[7] H. Bunke, "Parameter estimation in nonlinear regression," in *Handbook of Statistics* (P. R. Krishnaiah, ed.), vol. 1, pp. 593–615, North-Holland Publishing Company, 1980.

[8] T. Strutz, *Data Fitting and Uncertainty*. Springer Viewweg, 2nd ed., 2016.

[9] M. Shanker, M. Y. Hu, and M. S. Hung, "Effect of data standardization on neural network training," *Omega, Int. J. Mgmt Sci.*, vol. 24, no. 4, pp. 385–397, 1996.

[10] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural Networks: Tricks of the Trade* (G. Montavon, G. B. Orr, and K.-R. Müller, eds.), pp. 9–48, Springer, 2012.

[11] K. Kuźniar and M. Zajac, "Some methods of pre-processing input datafor neural networks," *Computer Assisted Methods in Engineering and Science*, vol. 22, p. 141–151, 2015.

[12] M. S. Bartlett, "The use of transformations," *Biometrics*, vol. 3, no. 1, 1947.

[13] F. J. Anscombe, "The transformation of poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3–4, p. 246–54, 1948.

[14] F. J. Anscombe and J. W. Tukey, "The examination and analysis of residuals," *Technometrics*, vol. 5, no. 2, 1963.

[15] G. E. P. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society*, vol. 26, no. 2, pp. 211–252, 1964.

[16] C. Yunhui, S. J. Clark, Y. Huang, L. Sinclair, C. L. A. Leung, S. Marussi, T. Connolley, O. V. Magdysyuk, R. C. Atwood, G. J. Baxter, M. A. Jones, I. Todd, and P. D. Lee, "In situ x-ray quantification of melt pool behaviour during directed energy deposition additive manufacturing of stainless steel," *Materials Letters*, vol. 286, p. 129205, 2021.

[17] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Stat. Methods. Med. Res.*, vol. 8, p. 135–160, 1999.

[18] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Int. J. Nurs. Stud.*, vol. 47, pp. 931–936, 2010.

[19] J. M. Bland and D. G. Altman, "Statistical method for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 327, p. 307–310, 1986.

[20] D. Giavarina, "Understanding bland altman analysis," *Biochemia medica*, vol. 25, no. 2, pp. 141–151, 2015.

[21] J. Bland and D. G. Altman, "Transforming data," *BMJ*, vol. 312, p. 770, 1996.

[22] J. Bland and D. G. Altman, "Measurement error proportional to the mean," *BMJ*, vol. 312, p. 1654, 1996.

[23] M. Brehm, V. A. Scholtes, A. J. Dallmeijer, J. Twisk, and J. Harlaar, "The importance of addressing heteroscedasticity in the reliability analysis of ratio-scaled variables: an example based on walking energy-cost measurements," *Dev Med Child Neurol.*, vol. 54, no. 3, pp. 267–273, 2012.

[24] S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3–4, p. 591–611, 1965.

[25] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network," *IEEE trans. Inf. Theory*, vol. 44, no. 2, pp. 525–535, 1998.

[26] S. Ingrassia and I. Morlini, "Neural network modelling for small datasets," *Technometrics*, vol. 47, no. 3, pp. 1537–2723, 2005.