

RESEARCH

Open Access



Relating plasticity to dislocation properties by data analysis: scaling vs. machine learning approaches

Stefan Hiemer¹, Haidong Fan² and Michael Zaiser^{1*}

*Correspondence:
michael.zaiser@fau.de

¹ Department of Materials Science, WW8-Materials Simulation, Friedrich-Alexander Universität Erlangen-Nürnberg, Dr.-Mack-Str. 77, 90762 Fürth, Germany

² Department of Mechanics, Sichuan University, Chengdu, People's Republic of China

Abstract

Plasticity modelling has long relied on phenomenological models based on ad-hoc assumption of constitutive relations, which are then fitted to limited data. Other work is based on the consideration of physical mechanisms which seek to establish a physical foundation of the observed plastic deformation behavior through identification of isolated defect processes ('mechanisms') which are observed either experimentally or in simulations and then serve to formulate so-called physically based models. Neither of these approaches is adequate to capture the complexity of plastic deformation which belongs into the realm of emergent collective phenomena, and to understand the complex interplay of multiple deformation pathways which is at the core of modern high performance structural materials. Data based approaches offer alternative pathways towards plasticity modelling whose strengths and limitations we explore here for a simple example, namely the interplay between rate and dislocation density dependent strengthening mechanisms in fcc metals.

Keywords: Dislocation plasticity, Scaling invariance, Machine learning

Introduction

Plasticity modelling has long proceeded along two independent lines. On the one hand, engineers are seeking tools to predict the behavior of engineering components during processing and under in-service loads. In this case, the material is assumed as given and the task is to predict, based on available experimental data, as accurately as possible the behavior of this material within complex-shaped parts with equally complex boundary loadings. This has led to phenomenological models which seek, often with an abundance of parameters, to reproduce a set of experimental data as accurately as possible. The task of these models is to reproduce the behavior of a given material under a typically rather limited range of deformation conditions, and in meeting this task they often achieve an impressive degree of accuracy. Their predictive power beyond the material and range of deformation conditions for which they have been parametrized, on the other hand, is extremely limited. Therefore, materials scientists tasked with the development of new and improvement of existing materials tend to use a different approach. By analysing deformation on the level of the defect microstructure, they seek to identify the physical

mechanisms that control macroscopic features of the plastic deformation process such as the flow stress. A mathematical description of the corresponding microstructure-property relationships, so it is hoped, may provide generic insights that can be used as a basis for predictive modelling. But this approach, which is supported by microstructure characterization tools of increasing sophistication, is itself beset by pitfalls.

On a most elementary level, the flow stress of a dislocated crystal can be related to the density and arrangement of crystal lattice dislocations. This is well established since the seminal paper of Taylor (1934) who analyzed crystal plasticity in terms of the motion of dislocations and established a fundamental relationship between the flow stress τ_f – here understood as critical resolved shear stress on the active slip system(s) – and the density of dislocations ρ , $\tau_f = \alpha \mu b \sqrt{\rho}$ where μ is the shear modulus, b the length of the Burgers vector of the active slip systems, and α a numerical factor which was put by Taylor into the range of $\alpha \approx 0.2 \dots 0.3$ where it has remained ever since. However, while the structure of the Taylor relationship has never been in question, the precise nature of the dislocation arrangements that give rise to this dependency has been a subject of controversy for decades. Taylor considered a checkerboard pattern of positive and negative dislocations ('Taylor lattice') – an arrangement which, while analytically tractable, has the disadvantage that it has never been observed in experiment. Based on surface observations (blocking of slip lines), Mott (1953) and Seeger et al. (1957) proposed that the Taylor stress is produced by large pile ups of dislocations at indestructible Lomer-Cottrell Barriers, an idea which, while consistent with surface observations, cannot be reconciled with TEM where such pile ups are very hard to find. Bailey and Hirsch (1960) instead suggested that Taylor's law can be explained by the stress needed to cut forest dislocations, whose spacing scales like the square root of dislocation density. Finally, Hirsch and Warrington (1961) pointed out that Taylor-type behavior can also be explained by the dragging of jogs whose spacing in turn reflects the forest spacing.

The controversies surrounding the mechanism of hardening illustrate an inherent weakness in the quest for 'mechanism-based' interpretations of complex collective phenomena. Virtually all of the above mentioned dislocation configurations and mechanisms (with possible exception of the original Taylor lattice) can be observed in TEM imaging, but possible selection bias makes it difficult to quantify their relevance based on published data. Another example of the problematics of mechanistic thinking concerns the nature and role of dislocation sources. Every dislocations textbook contains images of Frank-Read or spiral sources, and numerous attempts to explain the flow stress of materials from the macro to the micro scale are based on the concept of a 'weakest source'. Yet, while dislocation sources can be observed in TEM, they are surprisingly rare and the actual process of dislocation multiplication does *not* proceed by sequential emission of discrete loops from dislocation sources but in a much more diffuse and mechanistically less tangible manner which, despite recent works such as the excellent simulation study of Weygand and co-workers (Stricker et al. 2018), is still not fully understood. Thus, even simple and elegant mechanisms on the single-dislocation level do not necessarily help to obtain an adequate understanding of the inherently collective and complex dynamics of dislocation networks.

Here we illustrate two alternative approaches towards quantifying the relationship between plastic deformation behavior and properties of the dislocation microstructure.

High-throughput simulations and experimentation offer the perspective to establish a sound data base which allows data analytic methods to be used for identification and classification of recurrent features and structures. Mathematical analysis allows to formulate symmetries and invariance principles that reduce data complexity and assist in analysis. In the present study, we illustrate these approaches on a simple example, namely the superposition of rate and dislocation density effects in controlling the flow stress of fcc single crystals.

Data base

High throughput discrete dislocation dynamics (DDD) simulations covering 9 orders of magnitude in dislocation density ρ ($10^7 \text{m}^{-2} < \rho < 10^{16} \text{m}^{-2}$) and 7 orders of magnitude in strain rate ($10^{-1} \text{s}^{-1} \leq \dot{\epsilon} \leq 10^6 \text{s}^{-1}$) were conducted by Fan and co-workers (Fan et al. 2021) with the aim of establishing the joint influence of dislocation density and strain rate on the flow stress of fcc metals. These simulations were complemented by MD simulations of highly dislocated crystals which further extend the range of dislocation densities and strain rates to densities of $2.2 \times 10^{16} \text{m}^{-2}$ and strain rates of $2.5 \times 10^8 \text{s}^{-1}$. For each set of parameters multiple simulations with different, but statistically equivalent, initial dislocation configurations were conducted, amounting to a total of about 200 simulations. In all these simulations the flow stress, defined as the stress at a fixed plastic strain ϵ_y^p , was recorded alongside the imposed strain rate and dislocation density at the same strain. A default value $\epsilon_y^p = 0.5\%$ was used for the offset plastic strain, though lower offsets were considered at the lowest strain rates. In addition to flow stress values, other characteristics such as the probability distribution $p(v)$ of dislocation segment velocities and the plastic strain pattern at the global strain ϵ_y^p were determined for all simulations.

The simulations were complemented by an extensive literature search to retrieve records where simultaneous measurements of strain, strain rate, flow stress and dislocation density were reported for monocrystalline specimens. This search yielded about 120 datasets, mostly from the literature of the 1960s to 1980s. Few examples from the recent literature could be found, since unfortunately the number of publications which report quantitative measurements of dislocation densities alongside mechanical data has over the past decades decreased in inverse proportion with the increasing number of plasticity models that use dislocation densities as internal variables. A downloadable compilation of all data can be found in the supplementary material of Fan et al. (2021).

In the following we investigate the performance of different prediction strategies in relating the flow stress to parameters such as strain, dislocation density and strain rate. As a measure of prediction performance we use the coefficient of determination

$$R^2 = 1 - \frac{\sum_i (x_i - x_i^{\text{pred}})^2}{\sum_i (x_i - \langle x_i \rangle)^2} \quad (1)$$

which we apply to the logarithm of the flow stress, which we try to predict based on the remaining variables. (Note that in view of the range of variation of all variables, which encompasses many orders of magnitude in dislocation density, strain rate and flow stress, a logarithmic measure is required).

Predicting flow stresses: superposition of forest hardening and rate effects

We first consider a simple question: How do dislocation interactions and rate-dependent flow stress contributions, which can ultimately be traced back to the stress needed to move dislocations with a given imposed velocity, superimpose in controlling the flow stress of fcc metals? In other words, what is the function that relates the flow stress σ_f to dislocation density ρ and strain rate $\dot{\epsilon}$? In the literature, there exists an abundance of phenomenological relationships introduced by different authors in a more or less ad-hoc manner. In particular we mention the form popularized by Mecking and Kocks (1981),

$$\sigma_f = \left(\frac{\dot{\epsilon}}{\dot{\epsilon}_0} \right)^{1/m} \alpha_0 \mu b \sqrt{\rho} \tag{2}$$

where $\dot{\epsilon}_0$ is an arbitrary reference strain rate, α_0 an accordingly determined nondimensional factor, μ the shear modulus, and b the Burgers vector length. The exponent m is called the strain rate sensitivity. We shall probe the usefulness of this and similar equations in reproducing data based on a combination of theoretical and data based analysis. (For a discussion of other aspects, such as thermodynamic consistency, see Wu and Zaiser (2022)).

Scaling analysis

In the following we show how to exploit generic scaling invariance properties of dislocation systems in order to establish constraints on the possible form of constitutive laws that connect statistically averaged properties of dislocation systems such as stress, strain rate, and dislocation density. The main arguments were for the first time formulated by Zaiser and Sandfeld (2014) and are here briefly repeated.

We consider a system of N dislocations $i \in \{1 \dots N\}$. The dislocation with Burgers vectors \mathbf{b}^i moves by glide on the slip plane \mathcal{P}^i with slip plane normal \mathbf{n}^i . The unit slip vector is $\mathbf{s}^i = \mathbf{b}^i / b$ where b is the Burgers vector length. The dislocations form closed loops \mathcal{C}^i , each contained within a single slip plane. These loops are parameterized by $\mathbf{r}(s^i)$ with local tangent vector $\mathbf{t}(s^i) = d\mathbf{r}/ds^i$. Junctions are described in terms of local alignment of segments of different loops. We consider bulk behaviour, i.e., we assume that the dislocation loops are contained within a quasi-infinite crystal where the boundaries are remote such that image stresses can be neglected, or that the system is replicated periodically.

Dislocation motion is assumed to occur by glide and to be controlled by phonon drag with drag coefficient B . Thus, the local velocity is given by

$$\frac{\partial \mathbf{r}(s^i)}{\partial t} = [\mathbf{n}^i \times \mathbf{t}(s^i)] v(s^i) \quad , \quad v(s^i) = \frac{b}{B} \tau^i(s^i) \quad , \quad \tau^i(s^i) = \mathbf{M}^i : \boldsymbol{\sigma}(\mathbf{r}(s^i)) \tag{3}$$

where $\mathbf{M}^i = [\mathbf{n}^i \otimes \mathbf{s}^i]^{\text{sym}}$ is the slip system projection tensor. The stress is composed of an 'external' stress imposed by remote boundary tractions and considered constant over the volume of interest, and a dislocation-related internal stress which can be computed in terms of line integrals over the dislocation lines:

$$\begin{aligned} \boldsymbol{\sigma} &= \boldsymbol{\sigma}^{\text{ext}} + \boldsymbol{\sigma}^{\text{int}} \\ \sigma_{kl}^{\text{int}}(\mathbf{r}) &= -\frac{\mu}{8\pi} \sum_k \int_{C^k} \left\{ \frac{2}{1-\nu} \left(\frac{\partial^3 R}{\partial r_n \partial r_k \partial r_l} - \delta_{kl} \frac{\partial}{\partial r_n} \nabla^2 R \right) b_o \epsilon_{nom} t_m \right. \\ &\quad \left. + \left(\frac{\partial}{\partial r_n} \nabla^2 R \right) b_o [\epsilon_{nok} t_l + \epsilon_{nol} t_k] \right\} ds^k, \end{aligned} \tag{4}$$

where $R = |\mathbf{r} - \mathbf{r}(s^k)|$. It is now easy to see that the above equations are invariant upon re-scaling by an arbitrary factor λ according to

$$\mathbf{r} \rightarrow \lambda \mathbf{r} \quad , \quad \boldsymbol{\sigma} \rightarrow \lambda^{-1} \boldsymbol{\sigma} \quad , \quad t \rightarrow \lambda^2 t \quad . \tag{5}$$

which implies the auxiliary transformations

$$\nu \rightarrow \lambda^{-1} \nu \quad , \quad \rho \rightarrow \lambda^{-2} \rho \quad , \quad \dot{\boldsymbol{\epsilon}}^P \rightarrow \lambda^{-3} \dot{\boldsymbol{\epsilon}}^P \quad . \tag{6}$$

where the transformation rule for dislocation density follows directly from its definition as line length per volume. The transformation rule for the plastic strain rate follows from its definition in terms of slip rates on the different slip systems, $\dot{\boldsymbol{\epsilon}}^P = \sum_{\beta} \mathbf{M}^{\beta} \dot{\gamma}^{\beta}$, where the slip rates $\dot{\gamma}^{\beta} = b \dot{A}^{\beta} / V$ are products of Burgers vector length and rate of change in slipped area per dislocation loop, divided by the system volume.

Crystal plasticity constitutive equations

We now outline some consequences of the above formulated invariance principle. It is clear that invariance under the transformation (5) does not depend on any mechanisms or specific processes, nor does it depend on the scale on which the dislocation system is considered. Scaling invariance must not only hold on the macroscopic scale, but must also apply to any emergent statistical signatures of the evolving dislocation system. This is in well known for the characteristic wavelength of dislocation patters, which obeys the so-called law of similitude (Rudolph et al. 2005; Sauzay and Kubin 2011) as well as for the mesh length distribution in fractal dislocation networks (Zaiser and Hähner 1999; Hähner and Zaiser 1999) and the distribution of dislocation velocities (Fan et al. 2021). In particular, any constitutive equations that derive from the micro-dynamics of interacting dislocations via an averaging procedure are bound to possess the same invariance properties. This provides us with a useful rule-of-thumb for assessing the validity of phenomenological dislocation-based constitutive equations proposed in the literature. For instance, simple power counting demonstrates that Eq. (2) is invariant under (5) only in the rate independent limit $m \rightarrow \infty$, and can therefore not meaningfully describe rate effects in dislocation plasticity.

Conversely, a useful strategy for formulating constitutive equations is to cast these in the form of relationships between *invariant parameters* that by construction show invariance under the transformation (5). We demonstrate this strategy for the superposition of rate and dislocation density effects in dislocation plasticity. Thus, we define invariant slip rate and dislocation density variables on the different slip systems β via

$$\mathbf{P}^{\beta} = \left(\frac{\mu b^3}{B} \right)^{2/3} \frac{\rho^{\beta}}{(\dot{\gamma}^{\beta})^{2/3}} \quad , \quad \Gamma^{\beta} = \frac{B}{\mu b^3} \frac{\dot{\gamma}}{(\rho^{\beta})^{3/2}} = (\mathbf{P}^{\beta})^{-3/2} \tag{7}$$

Similarly, we define invariant stress and strain measures via

$$\Sigma_\rho = \frac{\sigma}{\mu b \sqrt{\rho}} \quad , \quad \Sigma_{\dot{\gamma}} = \frac{\sigma}{\mu^{1/3} (B \dot{\gamma})^{2/3}} \quad , \quad \Gamma^\beta = \frac{\gamma^\beta}{b \rho^{1/2}} \tag{8}$$

from which invariant shear stresses derive via

$$T_\rho^\beta = \mathbf{M}^\beta : \Sigma_\rho \quad , \quad T_{\dot{\gamma}}^\beta = \mathbf{M}^\beta : \Sigma_{\dot{\gamma}}.$$

Note that we have non-dimensionalized all quantities using the material constants which govern dislocation motion and interactions, in order to allow for a material independent formulation of constitutive behavior associated with collective dynamics of dislocations.

Next we study asymptotic cases. We note that the slip rates scale on the active slip systems β are given by $\dot{\gamma}^\beta = \rho^\beta b v^\beta$ where ρ^β is the dislocation density on a given slip system and $v^\beta = \dot{\gamma}^\beta / (\rho^\beta b)$ the average velocity of these dislocations.

First we envisage the quasistationary limit $v^\beta \rightarrow 0$ of near-zero strain rates or of very high dislocation densities. In this limit the stresses on all dislocation lines asymptotically vanish, $\sigma^{\text{ext}} + \sigma^{\text{int}} \rightarrow 0$, from which scaling invariance dictates that $\tau^\beta \propto \sqrt{\rho}$ for all active slip systems β . In terms of the invariant variable P^β this behavior is expressed as

$$T_{\dot{\gamma}}^\beta = \alpha^\beta (P^\beta)^{1/2} \quad , \quad T_\rho^\beta = \alpha^\beta \quad , \quad P^\beta \rightarrow \infty, \dot{\Gamma}^\beta \rightarrow 0. \tag{9}$$

where the parameters α^β may depend on the distribution of dislocations over the different slip systems as expressed by the ratios $f^\beta = \rho^\beta / \rho$.

In the opposite limit $v^\beta \rightarrow \infty$ of low dislocation densities or high strain rates, the resolved shear stresses acting on the dislocations must become very high. This is only possible when the externally applied stresses are high and in the asymptotic limit, the internal stresses are asymptotically irrelevant. Thus, in the asymptotic limit $v^\beta = \dot{\gamma}^\beta / (\rho^\beta b) = \tau^\beta B / b$. From this relation we obtain

$$T_{\dot{\gamma}}^\beta = (P^\beta)^{-1} \quad , \quad T_\rho^\beta = (P^\beta)^{-3/2} \quad , \quad P^\beta \rightarrow 0, \dot{\Gamma}^\beta \rightarrow \infty. \tag{10}$$

An equivalent formulation is of course possible by substituting $P^\beta = (\dot{\Gamma}^\beta)^{-3/2}$. A generic constitutive law must interpolate between the asymptotic limits given above. The simplest way to do so is to simply add up the asymptotic expressions, as proposed by Fan et al. (2021):

$$T_{\dot{\gamma}}^\beta = (P^\beta)^{-1} + \alpha^\beta (P^\beta)^{1/2} \quad , \quad T_\rho^\beta = \alpha^\beta + (P^\beta)^{-3/2} = \alpha^\beta + \dot{\Gamma}^\beta. \tag{11}$$

Equation (11) is compared in Fig. 1 with the data of Fan et al. (2021), who in their simulations consider uniaxial tensile tests of Cu and Al deformed in [100] orientation, and with experimental data referring to uni-axial tests compiled from the literature (Fan et al. 2021) which have been corrected for orientation factors. It can be seen that the simple relationship (11) provides an acceptable description of the data. We measure prediction performance in terms of the coefficient of determination

$$R_\tau^2 = 1 - \frac{\sum_i (\ln \tau_i - \ln \tau_i^{\text{pred}})^2}{\sum_i (\ln \tau_i - \langle \ln \tau_i \rangle)^2} \tag{12}$$

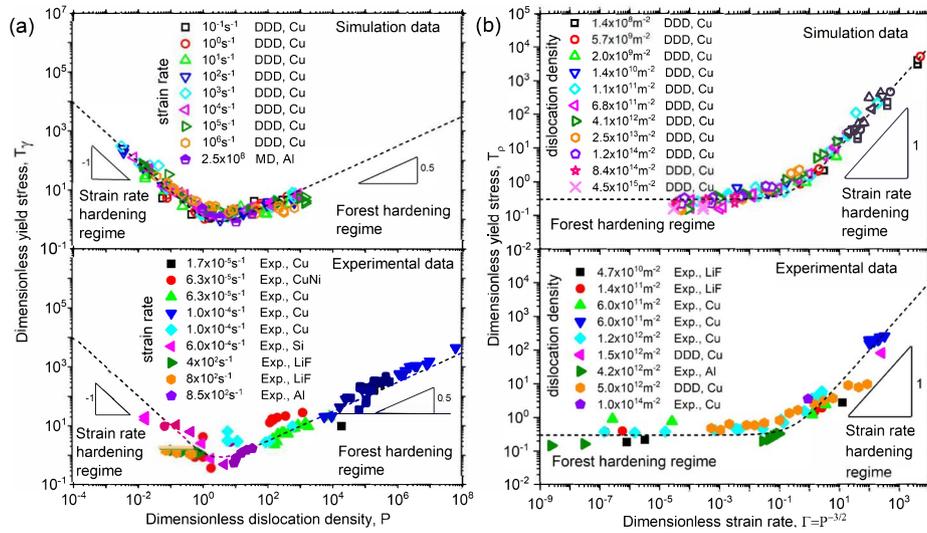


Fig. 1 Comparison between the scaling function, Eq. (11) (dashed lines) and data describing the dependency of flow stress on strain rate and dislocation density; materials and type of data (discrete dislocation dynamics simulation - DDD, molecular dynamics simulation (MD), or experiment (Exp.), are indicated in the legends; figure reproduced after Fan et al. (2021)

where τ_i are the individual flow stress values within the set to be predicted, $\langle \ln \tau_i \rangle$ is their logarithmic average, and τ_i^{pred} are the predicted values for τ_i . Note that we use logarithmic quantities (i.e., we consider *relative* deviations) to ensure that all stress values in datasets which typically span 4-5 orders of magnitude are weighted equally (not doing so would imply that our performance measure is dominated by the largest stresses).

For our entire set of data (simulation and experiment), a fit of Eq. (11) results in a high coefficient of determination ($R^2 = 0.958$). However, it is clear that the simple expression (11) is not the only possible form of a constitutive law that is consistent with scaling. For example, the transition regime between the asymptotic regimes might be represented by a modified scaling law which leaves the asymptotics unchanged, such as

$$T_\rho^\beta = \alpha^\beta + \dot{\Gamma}^\beta + \delta(\dot{\Gamma}^\beta)^\eta. \tag{13}$$

Fitting this law to all data produces a slightly improved fit ($R^2 = 0.967$) as shown in Fig. 2.

Importantly, Eq. (13) implies a different leading-order strain rate dependency of flow stress in the regime of low strain rates. Equation (11), which has been used often in the literature, predicts that the strain rate increases linearly with the 'effective stress' $\tau_{\text{eff}}^\beta = \tau^\beta - \alpha^\beta \mu b \sqrt{\rho}^\beta$, provided that the resolved shear stress exceeds the friction-like 'Taylor stress': $\dot{\gamma} \propto \tau_{\text{eff}}^\beta$ if $|\tau^\beta| \geq \alpha^\beta \mu b \sqrt{\rho}^\beta$. Equation (13), on the other hand, predicts a nonlinear increase, $\dot{\gamma} \propto (\tau_{\text{eff}}^\beta)^{1/\eta}$, as has been reported in simulations, see e.g. Miguel et al. (2002).

Machine learning

In the following, we study the performance of different machine learning methods in predicting flow stresses (resolved shear stresses) based on dislocation density, strain rate, and strain as well as essential materials parameters (shear modulus, dislocation

drag coefficient, Burgers vector length). The scaling analysis presented in the previous paragraph serves as a benchmark for prediction performance. All features and targets were logarithmically transformed and standardized before training, testing and prediction. Specifically, to ensure comparability between simulation and experimental data, experimental resolved shear stresses determined for some material X deformed in $[hkl]$ orientation with Schmid factor $M_{[hkl]}$ were referred to Cu deformed in $[100]$ orientation by multiplication with the ratio of Schmid factors, $M_{[100]}/M_{[hkl]}$ and with the ratio of shear moduli, μ_{Cu}/μ_X . Similarly, strain rates were referred to $[100]$ deformed Cu by multiplication with the inverse ratio of Schmid factors, $M_{[hkl]}/M_{[100]}$, and the ratio of drag coefficient over Burgers vector, $b_{Cu}B_X/(b_XB_{Cu})$. Drag coefficients were taken from the experimental studies compiled by Fan et al. (2021); where possible, data obtained from pulse loading of single dislocations were used. No other pre-processing of the data was performed, in particular, we did *not* use transformation to scaling invariant parameters as in Eqs. (7, 8).

ML methods

To analyze the data, we use three different methods, namely kernel ridge regression (KRR), a decision tree, and a simple neural network. This choice covers different classes of machine learning approaches: i) KRR is representative for kernel approaches (support vector regression, kernel principal component analysis coupled to other regressors, etc.) and memory/instance based regressors (k-nearest neighbors); ii) decision trees cover tree based approaches (gradient boosted trees, random forests, etc.); iii) the multilayer perceptron covers for neural networks, we use this type because our data are not images (which rules out convolution based neural networks) and have no graph structure (which rules out graph neural networks). The purpose is to analyse generic issues of the application of ML approaches to dislocation problems that are not specific to the class of ML model used.

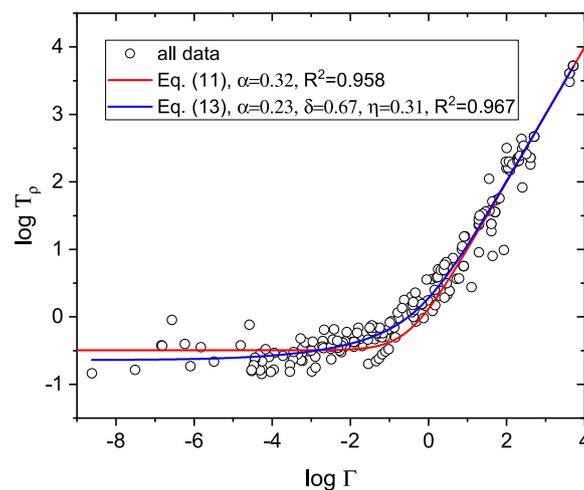


Fig. 2 Fitting the data of Fig. 1 with a modified law corresponding to a different asymptotic strain rate exponent; data points: all data in Fig. 1, right; red curve: best fit according to Eq. (11), blue curve: best fit according to Eq. (13), fit parameters see legend

Kernel ridge regression is a memory based method that makes predictions for a new data point x_i through its similarity to samples in the training set x_j . Similarity is quantified by a kernel function k and a distance $d(x_i, x_j)$, such that new predictions are made by a linear combination of weighted kernel functions

$$y_i = \sum_j^N w_j k(d(x_i, x_j), \gamma). \quad (14)$$

γ is here a generic kernel parameter. The weights w are inferred by minimizing an L_2 regularized least squares problem which yields a closed form solution (Bishop and Nasrabadi 2006). We use the standard Euclidean distance and the radial basis function kernel. The regularization parameter is varied between 10^{-5} and 10^5 with one hundred logarithmically even spaced increments as well as the kernel parameter γ . The combination with the best performance in the test set is chosen as final parameter set. Decision trees partition the feature space in greedy fashion and make predictions through the average value of training points within the partition to which a new data point x_i is assigned. This partitioning is done in a sequential per-feature manner. The maximum tree depth, minimum number of samples per partition and minimum number of samples to induce a split are tested with values 2^n , $1 \leq n \leq 5$ by exhaustive combination. The final model is a multilayer perceptron. Here the model is a set of stacked layers consisting of individual units/neurons with the nonlinear activation function $f(x)$. Each neuron (i, j) is connected to all units $(i - 1, k)$ of the previous layer via a weighted connection of weight w and is further modified with a bias b . The intermediate value in the i -th layer on the j -th neuron then is given by

$$z_{i,j} = f\left(\sum_k [b_{i,j} + w_{k,j} z_{i-1,k}]\right) \quad (15)$$

Weights and biases are trained in a stochastic gradient approach via backpropagation. In this work the architecture is kept very small due to the limited number of samples: The structure is varied between two to four layer depth with a width of ten neurons. We test both the relu and sigmoid activation functions. Training is done with the Adam stochastic optimizer (Kingma and Ba 2014). For more details on KRR, the interested reader is referred to the book of Bishop and Nasrabadi (2006), whereas for the other methods we refer to the standard textbook of Hastie et al. (2009). For KRR and the decision tree, the scikit-learn package was used (Pedregosa et al. 2011) while for the multilayer perceptron the Keras library was used with Tensorflow as backend (Chollet et al. 2015; Abadi et al. 2015).

Training schemes

Our data base collates data from different sources (experiment, simulation) and containing different features. In our ML study, we used different training schemes which differ in the manner how the data base is divided into training and test data, and in the features which are actually being used by the ML algorithms for flow stress prediction. These training schemes are compiled in Table 1. The first training scheme follows the classical idea of building a model based on simulation data and then validating it based

on experimental data. The second scheme uses all data and splits them randomly into subsets for training and testing. Schemes 3 and 4 work similarly, however, Scheme 3 considers only experimental data for training and testing, and Scheme 4 considers only simulation data. All schemes are considered in two versions, one including plastic strain among the features used for training and prediction and the other discarding the strain feature.

ML results

In our first training scheme we use the simulation data as training data and use the experimental data for testing (or validation), following the classical scheme of using experimental data to validate a model based on theory or simulation. In this case, a fit of the simulation data using Eq. (12) allows to reproduce the actual values with a coefficient of determination of $R^2 = 0.972$, i.e., the fit is slightly better than for the case considered in the previous section where *all* data were used for fitting. This value provides our benchmark for the ML algorithms. The same parameters then reproduce the experimental data with a value of $R^2 = 0.782$, indicating that the scaling analysis captures the experimental data well but also showing that the agreement is not perfect. A comparison of the scaling predictions and actual values for this training strategy is shown in Fig. 3, top left, using different symbols for predictions related to experimental and simulation data.

The reduced reliability in predicting simulation data is to be expected, since the simulations represent highly idealized situations where deformation is controlled by dislocation interactions and dislocation drag alone, whereas the experiments are necessarily influenced by presence of other defects such as impurities or point defects, even though only experiments using single crystal specimen were considered. In line with this argument, deviations are strongest in the regime of low flow stresses, where the relative influence of such confounding factors is highest and accordingly the scaling prediction tends to underestimate the actual flow stress. However, the crucial question is whether the simulations correctly represent the essential aspects of the reality which underlying the experimental data, which is mathematically reflected by the scaling relations which we have formulated, such that the additional influences can be considered as added noise. Alternatively, it might be possible that the additional factors entering the experiments are essential in the sense that the experiments reflect a fundamentally more complex reality which cannot be adequately represented in the simplified simulation setting. Machine learning approaches offer a new perspective on this fundamental question, which we explore in the following.

Table 1 Training schemes used in our ML study; scheme variants labeled with an asterisk exclude the plastic strain from the considered features

No.	training data	test data	features
1/1*	all simulation data	all experimental data	$\rho, \dot{\gamma}, (\epsilon_p)$
2/2*	subset of all data (88%)	subset of all data (12%)	$\rho, \dot{\gamma}, (\epsilon_p)$
3/3*	subset of experimental data (84%)	subset of experimental data (16%)	$\rho, \dot{\gamma}, (\epsilon_p)$
4/4*	subset of simulation data (90%)	subset of simulation data (10%)	$\rho, \dot{\gamma}, (\epsilon_p)$

If we apply the same training scheme 1 but replace the scaling fit by a machine learning algorithm, the results are at first glance quite disastrous (Fig. 3, top right). The R^2 values compiled in Table 2 indicate that, while notably the perceptron algorithm is able to well reproduce the simulation data it has been fed, the performance of the trained algorithms in reproducing the test (experimental) data is either low ($R^2 < 0.5$) or non existing ($R^2 < 0$).

In order to establish the reasons for this poor performance, we first look at the feature set provided to the algorithms. In this set, one variable (the plastic strain) is not actually used in the scaling analysis. This is in line with the materials physics idea that plastic strain is not a meaningful variable characterizing the internal state of a material. Of course, plastic strain can nevertheless be related to flow stress IF the initial state of the material (the initial dislocation density) is known, and evidently flow stress tends to increase with plastic strain in a strain hardening material. However, here the problem is

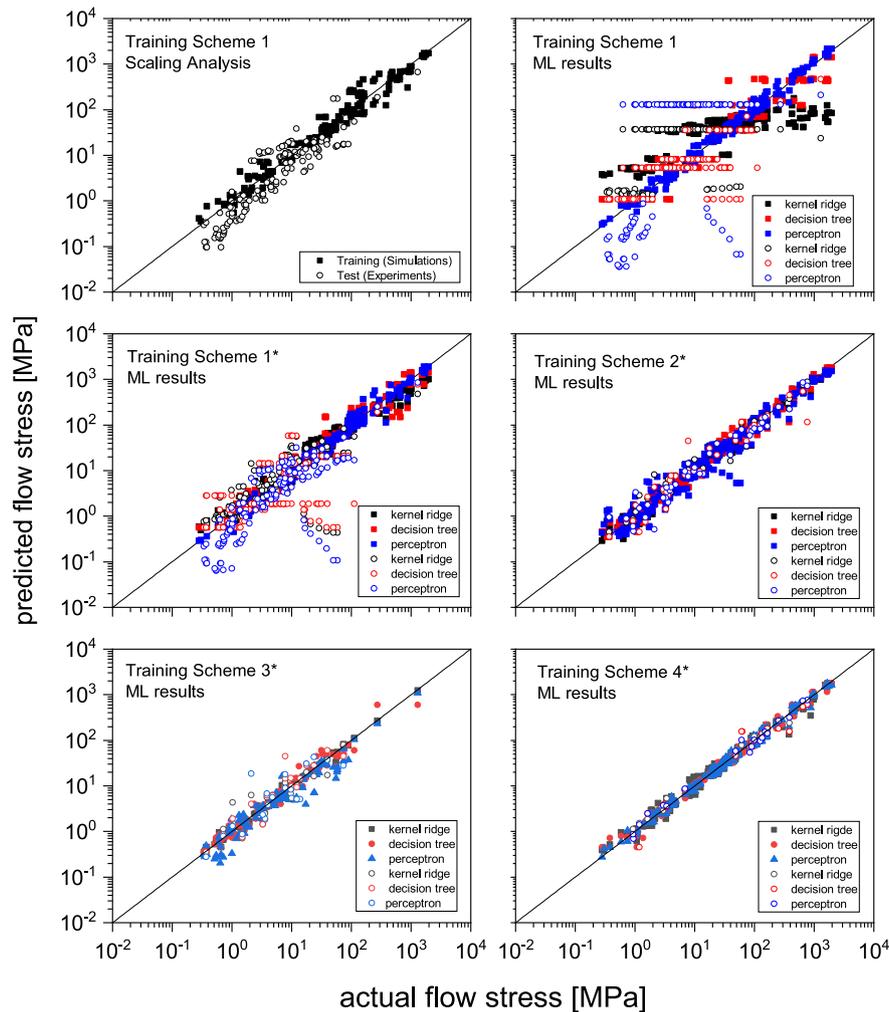


Fig. 3 Comparison of the performance of different prediction strategies, for description of the different training schemes, see Table 1, for the respective prediction scores, see Table 2; full symbols refer to the ability of a scheme to reproduce training data, open symbols illustrate the ability of a scheme to reproduce test data, each data point corresponds to the actual value and predicted value for one data record

exacerbated by the fact that the plastic strains in the dislocation dynamics simulations are quite small (typically 0.2%) and no systematic study of strain hardening has been conducted, whereas in the experiments strains may be much larger but data stem from a wide range of sources where initial conditions may differ. Thus, the training (simulation) and test (experiment) data show poor overlap and non-systematic coverage in feature space as far as this variable is concerned.

If we remove plastic strain from the feature set in a modified training scheme 1*, the predictive power of the ML algorithms increases (Fig. 3, center left). First of all, the 'training' performance is improved as most algorithms obtain better scores in reproducing the simulation data. Second, now all algorithms achieve positive prediction scores ($R^2 > 0$) for the experimental (test) data, though their performance still falls below the performance of the scaling analysis. It is therefore fair to call the plastic strain in the present context a *confounding variable*.

Even if plastic strain is removed from the feature set, the limited overlap of training and test data in feature space remains a problem that can cause poor prediction performance. Looking at Fig. 1, we see that the parameter range covered by the simulations corresponds to an interval of lower P parameters than that of the experiments. The reason is that simulations are typically conducted at much higher strain rates than experiments. This is a simple consequence of the required effort: The computational effort to conduct a DDD simulation increases tremendously with decreasing strain rate, because the numerical stiffness of the simulations increases. The reason is that the simulation time step is controlled by the motion of fast nodes on close, strongly interacting dislocations, and therefore only weakly strain rate dependent, whereas the overall simulated time to reach a given strain is inversely proportional to the imposed strain rate. In experiment, the opposite is true: While a low strain rate of, say, 10^{-5} s^{-1} is completely standard in a tensile test, achieving a high strain rate $> 10^4 \text{ s}^{-1}$ in controlled test requires non standard equipment and significant effort.

To resolve the problem of poor overlap of training and test data in feature space, we devise a second training scheme ('training scheme 2') where training and test data are chosen randomly from the pool of *all* datasets, ensuring an approximately equal coverage of feature space by the training and test data. As seen in Fig. 3, center right, this leads to an improved performance which now matches the results of the scaling analysis. This is also manifest from the R^2 values compiled in Table 1 for this training scheme: First, the improved overlap

Table 2 Values of coefficient of determination for the different prediction algorithms and training schemes; schemes with asterisk do not consider the plastic strain for prediction, note that the scaling fits which serve as reference do not consider plastic strain

training	scaling fit		kernel ridge		decision tree		perceptron	
	training	test	training	test	training	test	training	test
1	-	-	0.388	- 0.45	0.879	0.430	0.989	-2,32
1*	0.972	0.782	0.961	0.514	0.963	0.237	0.988	0.202
2	-	-	0.988	0.958	0.990	0.927	0.958	0.946
2*	0.967	0.944	0.937	0.955	0.994	0.958	0.948	0.957
3	-	-	0.989	0.902	0.975	0.907	0.954	0.847
3*	-	-	0.994	0.890	0.977	0.881	0.909	0.822
4	-	-	0.994	0.996	0.997	0.988	0.995	0.996
4*	-	-	0.996	0.996	0.997	0.987	0.998	0.996

of training and test data ensures that also the scaling fit works better in reproducing the test set. Second, now all machine learning algorithms achieve comparable prediction performance for training and test data, and this performance matches the results of the scaling analysis. This provides evidence that the data derived from simulation and experiment, even though they tend to be located in different areas of parameter space, share a common underlying structure and that this structure is well captured by the scaling analysis.

This conjecture is further corroborated when we consider training schemes that consider either only experimental data for training and testing (scheme 3) or only simulation data (scheme 4). The performance of scheme 3, which considers only experimental data (Fig. 3, bottom left) is worse than that of scheme 2 (all data). This can be understood that, while the data share a fundamentally similar structure, the experimental data are more noisy. Accordingly, the best performance is achieved by scheme 4 (only simulation data) as shown in Fig. 3, bottom right, where the data show near-perfect predictability irrespective of the algorithm used – a result of the highly controlled and highly idealized nature of the simulations.

Discussion and Conclusions

The present example is simple and surely not a critical test of the potential of machine learning approaches – after all, we are dealing with the representation of a comparatively simple functional relationship in a low-dimensional parameter space. Nevertheless, it illustrates some of the chances and pitfalls that arise in the interplay between high-throughput simulation, experiment, and machine learning.

The first and obvious conclusion points to the necessity of ensuring adequate overlap between simulation and experiment, reflecting the observation that most machine learning approaches are better at interpolation than at extrapolation (Webb et al. 2020). This can be facilitated by carefully analysing the mathematical structure underlying the simulations: our scaling analysis actually demonstrates that a discrete dislocation dynamics simulation at high strain rate and high dislocation density may be equivalent to one at lower strain rate and lower dislocation density. Recognizing this fact can evidently accelerate the exploration of parameter space and allows to cover, at the same cost, a wider range of parameters. More generally speaking, it is helpful to exploit any symmetries in the mathematical formulation of the simulation problem, of which the scaling relations studied here are a nontrivial example. In our work on machine learning approaches to materials mechanics, we generally observe the importance of accounting for symmetries as well as symmetry breaking phenomena. Examples include the breaking of the translational symmetry of space due to strain localization in the run-up to creep failure (Biswas et al. 2020), and the use of feature functions of reduced symmetry for predicting bond breaking in statistically isotropic glasses under axial load (Font-Clos 2022).

If these caveats are addressed, our little study demonstrates that machine learning approaches can correctly infer relationships between simulation and/or experimental data and can thus be used to represent constitutive relationships governing material behavior. Because this representation is not governed by inherent biases such as physically unmotivated traditions regarding the structure of constitutive laws (see Eq. (2)), they may in fact do a better job than many human researchers (Hiemer and Zapperi

2021). The presented study also illustrates that it is even possible to use both kinds of data (simulation + experiment) in conjunction, such as to achieve comprehensive coverage of parameter space, provided that there is reason to believe that the simulations capture essential aspects of the experimental reality. Moreover, by comparison of different training schemes it is possible to check, even without recourse to theoretical arguments (here provided by our scaling analysis) whether or not this conjecture is valid.

Acknowledgements

Not applicable.

Authors' contributions

S.H. designed and performed the machine learning analysis and wrote the section on machine learning, H.F. provided the data base from high-throughput DDD simulations and experimental literature searches. M.Z. and H.F. performed the scaling analysis, M.Z. drafted the manuscript. All authors corrected and edited the manuscript final version. The author(s) read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. M.Z. and S.H. acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG) under Grants no. Za 171/13-1 and Za 171/15-1. S.H. also acknowledges participation in the training activities of the DFG graduate school FRASCAL (GRK 2423/1).

Availability of data and materials

Primary data and machine learning code can be downloaded from <https://simlab.wm.uni-erlangen.de/publications/relating-plasticity-to-dislocation-properties-by-data-analysis-scaling-vs-machine-learning-approaches>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 14 July 2022 Accepted: 13 December 2022

Published online: 03 January 2023

References

- M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous systems. arXiv:1603.04467. (2016)
- J. Bailey, P. Hirsch, The dislocation distribution, flow stress, and stored energy in cold-worked polycrystalline silver. *Philos. Mag.* **5**(53), 485–497 (1960)
- C.M. Bishop, N.M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4 (Springer, 2006)
- S. Biswas, D. Fernandez Castellanos, M. Zaiser, Prediction of creep failure time using machine learning. *Sci. Rep.* **10**(1), 1–11 (2010)
- F. Chollet, et al. keras. (2015), <https://keras.io>. Accessed 20 Dec 2022
- H. Fan, Q. Wang, J.A. El-Awady, D. Raabe, M. Zaiser, Strain rate dependency of dislocation plasticity. *Nat. Commun.* **12**(1845), 1–11 (2021)
- F. Font-Clos, M. Zanchi, S. Hiemer, S. Bonfanti, R. Guerra, M. Zaiser, S. Zapperi, Predicting the failure of two-dimensional silica glasses. *Nat. Commun.* **13**(2820), 1–11 (2022)
- P. Hähner, M. Zaiser, Dislocation dynamics and work hardening of fractal dislocation cell structures. *Mater. Sci. Eng. A* **272**, 443–454 (1999)
- T. Hastie, R. Tibshirani, J.H. Friedman, J.H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2 (Springer, New York, 2009), pp. 241–249
- S. Hiemer, S. Zapperi, From mechanism-based to data-driven approaches in materials science. *Mater. Theory* **5**(1), 1–9 (2021)
- P. Hirsch, D. Warrington, The flow stress of aluminium and copper at high temperatures. *Philos. Mag.* **6**(66), 735–768 (1961)
- D.P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
- H. Mecking, U. Kocks, Kinetics of flow and strain-hardening. *Acta Metall.* **29**(11), 1865–1875 (1981)
- M.C. Miguel, A. Vespignani, M. Zaiser, S. Zapperi, Dislocation jamming and andrade creep. *Phys. Rev. Lett.* **89**(16), 165501 (2002)
- N.F. Mott, Bakerian lecture: dislocations, plastic flow and creep. *Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci.* **220**(1140), 1–14 (1953)
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
- P. Rudolph, C. Frank-Rotsch, U. Juda, F. Kiessling, Scaling of dislocation cells in gas crystals by global numeric simulation and their restraints by in situ control of stoichiometry. *Mater. Sci. Engng. A* **400–401**, 170–174 (2005)
- M. Sauzay, L.P. Kubin, Scaling laws for dislocation microstructures in monotonic and cyclic deformation of fcc metals. *Prog. Mater. Sci.* **56**(6, SI), 725–784 (2011). <https://doi.org/10.1016/j.pmatsci.2011.01.006>

- A. Seeger, J. Diehl, S. Mader, H. Rebstock, Work-hardening and work-softening of face-centred cubic metal crystals. *Phil. Mag.* **2**(15), 323–350 (1957)
- M. Stricker, M. Sudmanns, K. Schulz, T. Hochrainer, D. Weygand, Dislocation multiplication in stage ii deformation of fcc multi-slip single crystals. *J. Mech. Phys. Solids* **119**, 319–333 (2018)
- G.I. Taylor, The mechanism of plastic deformation of crystals. part i.—theoretical. *Proc. R. Soc. Lond. Ser. A Containing Pap. Math. Phys. Character* **145**(855), 362–387 (1934)
- T. Webb, Z. Dulberg, S. Frankland, A. Petrov, R. O'Reilly, J. Cohen, Learning representations that support extrapolation. in *International conference on machine learning* (PMLR, 2020), pp. 10136–10146
- R. Wu, M. Zaiser, Thermodynamic considerations on a class of dislocation-based constitutive models. *J. Mech. Phys. Solids* **159**, 104735 (2022)
- M. Zaiser, P. Hähner, The flow stress of fractal dislocation arrangements. *Mater. Sci. Eng. A* **270**, 299–307 (1999)
- M. Zaiser, S. Sandfeld, Scaling properties of dislocation simulations in the similitude regime. *Model Simul. Mater. Sc.* **22**(6), 065012 (2014)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
