Scaling Laws From Statistical Data and Dimensional Analysis

Patricio F. Mendez
Department of Metallurgical and Materials Engineering,
Colorado School of Mines,
Golden, CO 80401
e-mail: pmendez@mines.edu

Fernando Ordóñez
Industrial and Systems Engineering,
University of Southern California,
Los Angeles, CA 90089
e-mail: fordon@usc.edu

Scaling laws provide a simple yet meaningful representation of the dominant factors of complex engineering systems, and thus are well suited to guide engineering design. Current methods to obtain useful models of complex engineering systems are typically ad hoc, tedious, and time consuming. Here, we present an algorithm that obtains a scaling law in the form of a power law from experimental data (including simulated experiments). The proposed algorithm integrates dimensional analysis into the backward elimination procedure of multivariate linear regressions. In addition to the scaling laws, the algorithm returns a set of dimensionless groups ranked by relevance. We apply the algorithm to three examples, in each obtaining the scaling law that describes the system with minimal user input. [DOI: 10.1115/1.1943434]

1 Introduction

In engineering design, we are constantly faced with the need to describe the behavior of complex engineered systems for which there is no closed-form solution or exhaustive analysis. This usually leads to a tedious, time consuming, and detailed study of the engineering process in question, delaying the overall process of design and limiting the total number of possibilities that can be investigated. For example, the design of a welding procedure involves so many parameters that it can seldom be predicted reliably; therefore, extensive experimentation must take place in order to determine an ideal process setup. A set of simple and intuitive design laws based only on the most relevant parameters would be of enormous help in this case. Scaling laws in the form of power laws, which we will simply call scaling laws, are particularly well suited for this purpose.

Scaling laws are ubiquitous in engineering. In fact, they have been used to explain the behavior of many physical, biological (e.g., [1,2]), psychophysical [3], geophysical (e.g., [4,5]), Internet traffic [6], and even economic systems [7]. A broad sample of problems that can be described with such scaling laws is presented in [8]. Segel [9] provides a good overview of simplification and scaling. Some reasons for the wide applicability of power law models in engineering are: (i) the combination of units has the form of a power law, (ii) the expressions of many physical phenomena have the form of power laws as noted above, and (iii) many empirical regressions of engineering data in log-log plots tend to give a straight line, which corresponds to a power law.

Scaling laws are of enormous utility during the early stages of design, when the configuration of a system and the choice of materials are still uncertain. In this case, they provide quick estimations of the feasibility of a design, help determine optimal sizes, and contribute to decisions about configuration and materials. These laws are also useful for control systems and for decision-making algorithms, predicting the behavior of a system much faster than computationally intensive models such as finite element analysis or computational fluid mechanics. The design and interpretation of physical models of reduced size, such as reduced-scale aircraft in wind tunnels, are based on scaling laws. When experimental databases or numerical models exist, scaling laws can be used to generalize and extrapolate the results obtained. For existing machines, scaling laws are useful for setup and tuning operations.

Our work aims at facilitating the process of engineering design by providing a computational tool that derives the best power law from experimental data. We propose the algorithm SLAW (Scaling LAWs). This algorithm combines a linear regression model of the experimental data with physical considerations of the process, namely, that the units of the resulting model match the units of the dependent variable. We look for the power law model that minimizes the prediction error only among models that have the correct units. The output of the algorithm is a physically meaningful and simple power law, representing the process and a set of dimensionless groups ordered by their relevance to the problem. The user input in selecting the simple model, and the ability to correct it further using the dimensionless groups, provide the means to construct a model that achieves the desired balance between accuracy and simplicity. An early version of this algorithm was presented in [10]. SLAW grew from that version by incorporating: a rounded model output, where all coefficients are fractions typical in the equations in physics and engineering; user input to select the best scaling law; and ranked dimensionless groups that explain the residual error. In this work, we call "model" any scaling law that captures the main behavioral trends of a process as a function of the parameters of the problem; this should not be confused with other ways of representing the problem, such as finite element models.

The SLAW algorithm differs from classical dimensional analysis in that it selects the scaling law with the smallest predictive error out of all the dimensionally correct models. There are computational implementations of dimensional analysis, such as that described by Kasprzak et al. [11], which are able to construct dimensionally correct models and check the completeness of the set of variables. SLAW differs from these approaches in the systematic search for an expression involving the smallest acceptable number of variables.

SLAW also differs from other statistical simplifications, such as the principal directions of the matrix of correlation, in that these other approaches, besides not necessarily providing the correct units, reduce the mathematical complexity of the problem but still consider all physical parameters, regardless of their importance. Previous works that combines linear regressions and dimensional analysis, such as landmark work by Vignaux and Scott [12], Vignaux [13], Li and Lee [14], and Dovi et al. [15], all use dimensionless groups determined a priori. In contrast, SLAW automatically generates the ranked dimensionless groups.

Contributed by the Applied Mechanics Division of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS for publication in the ASME JOURNAL OF APPLIED MECHANICS. Manuscript received by the Applied Mechanics Division, November 14, 2003; final revision, November 26, 2004. Associate Editor: K. M. Liechti. Discussion on the paper should be addressed to the Editor, Prof. Robert M. McMeeking, Journal of Applied Mechanics, Department of Mechanical and Environmental Engineering, University of California-Santa Barbara, Santa Barbara, CA 93106-5070, and will be accepted until four months after final publication of the paper itself in the ASME JOURNAL OF APPLIED MECHANICS.

The automatic determination of formulas, such as scaling laws, from data has been an active area of research in the artificial intelligence community. Important examples of this research are the algorithm BACON due to pioneering work by Bradshaw et al. [16], algorithms ABACUS [3,17] and COPER [17], and recent work by Washio and Motoda [3,18]. Important differences that the algorithm proposed here has with both BACON and ABACUS are that these latter algorithms produce only scaling laws with integer exponents and require that the data vary one variable at a time. The algorithm COPER and the work by Washio and Motoda differ from the current work in that they do not explicitly construct a scaling law that is the simplest with respect to some criterion, and the dimensionless groups obtained are not necessarily ranked by relevance to the model.

In Sec. 2, we describe the assumptions made on the physical process and the main ideas in the methodology. In Sec. 3, we present the algorithm SLAW. We apply this methodology to three different examples: a pendulum, ceramic-to-metal joining, and the "punch test" (a standard test used to determine mechanical properties of materials). We describe the examples and present the results obtained with our algorithm in Sec. 4. In Sec. 5, we present the conclusions of this work.

2 Methodology

2.1 Scope of Methodology. To illustrate this discussion, consider the problem of joining a ceramic cylinder to a metallic cylinder, as pictured in Fig. 1.

One of the quantities of interest in this problem is the volumetric strain energy in the ceramic, which can be expressed as

$$u(X,Z) = Y(X)u^*(X,Z) \tag{1}$$

Here, $u(X,Z)$ is the volumetric strain energy at each point in the ceramic and depends on the *problem parameters* X and other variables Z. In this example, problem parameters X can include the radius of the cylinders r, the yield strength of the metal σ_Y, and the elastic modulus of the materials E_c and E_m, while other variables Z typically represent the space and time coordinates of a point in the cylinders. The quantity Y is a characteristic value that only depends on problem parameters and has the same units as u. The function u^* is a dimensionless function that shows the volumetric strain energy variations relative to the characteristic value.

Our objective in this paper is to obtain a simple yet meaningful expression for Y from experimental data, as the characteristic value captures important trends of the quantity of interest. For example, in Sec. 4 we show that the characteristic value of the volumetric strain energy in ceramic-to-metal joining as a function of the parameters is $Y=\sigma_Y^2 r^3/E_c$. To use a nomenclature that is consistent with linear regression literature, we will also refer to the characteristic value Y as the *dependent variable* and to the parameters X as the *independent variables*.

Dimensional analysis states that the characteristic value Y can be decomposed in one power law expression with the same units of u and a function f of m dimensionless groups $\Pi_1,\ldots,\Pi_m$, which also have a power law expression as a function of the parameters. If we assume that there are n parameters for the problem $X_1,\ldots,X_n$, this means that

$$Y = a_0 \prod_{j=1}^{n} X_j^{a_{0j}} f(\Pi_1,\ldots,\Pi_m) \tag{2}$$

with $\Pi_i = a_i \Pi_{j=1}^{n} X_j^{a_{ij}}$. For simplicity, we will assume that the function f can be approximated by a power law. This is reasonable, given that most of the behavior of the dependent variable is typically captured by the power law expression and f shows small, smooth, and monotonic variations within a regime. Therefore, we can write Eq. (2) as

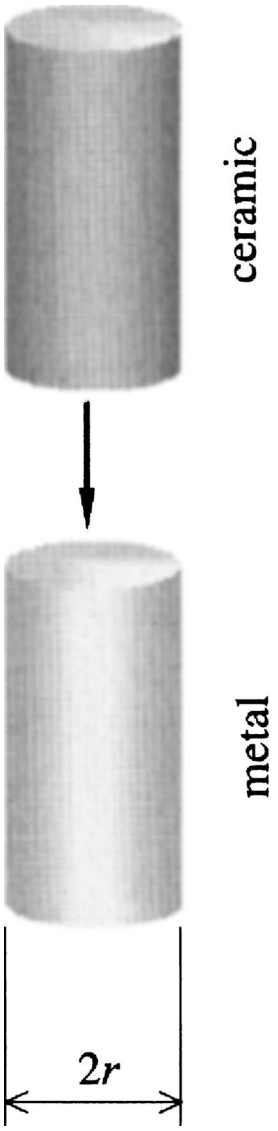

Fig. 1 Geometry of the ceramic and metal parts to be joined

$$Y = a_0 \prod_{j=1}^{n} X_j^{a_{0j}} \prod_{i=1}^{m} \left(a_j \prod_{j=1}^{n} X_j^{a_{ij}} \right) \tag{3}$$

$$= a \prod_{j=1}^{n} X_j^{\sum_{i=0}^{m} a_{ij}} \tag{4}$$

with $a = \Pi_{i=0}^{m} a_i$. Equation (4) is a valid model for the problem under two additional assumptions on the system, which are standard assumptions in dimensional analysis: (i) We assume that at least all parameters that determine the problem are considered. This assumption is also necessary in the analysis of regressions. Omitting a relevant parameter can result in ignoring a dominant effect and, thus, missing the correct model. Considering more parameters than are strictly necessary is not a problem, since SLAW can efficiently discard the less relevant parameters. (ii) We assume that the physical system is studied under a single regime. This means that the same physical factors (there is no need to know exactly which) are dominant for all of the observations used to build the input data set. This implies that f is of the order of magnitude of 1.

An important observation, which is enforced explicitly in the algorithm, is that the model given by Eq. (4) has the same units as the characteristic value Y. This additional constraint, denoted the *units constraint*, is expressed by

$$\text{units of } Y = \prod_{j=1}^{n} (\text{units of } X_j)^{\Sigma_{i=0}^{m} a_{ji}} \tag{5}$$

For this approach to be well defined, it is necessary that some combination of the parameters considered can yield the units of the characteristic value, and thus it is possible to satisfy the units constraint.

2.2 Constrained Linear Regression. For any application, the model postulated is faced with uncertainties that arise, for example, from working with experimental data or considering only n independent variables and disregarding the possibly tiny effect of other variables. By taking the logarithm of Eq. (4), and considering the existing uncertainties in the model, we can express the model as

$$\log Y = \beta_0 + \sum_{j=1}^{n} \beta_j \log X_j + \varepsilon \tag{6}$$

where the coefficients are $\beta_0 = \log a$ and $\beta_j = \Sigma_{i=0}^{m} a_{ij}$, and ε is an error term that captures the model uncertainties. Additive errors in logarithms of measurements is a common assumption in fitting scaling laws, see [15,19], and it is justified by Benford's law [20,21], which states that variations of physical quantities are evenly distributed in a logarithmic scale.

Considering p experimental observations of the physical process, we obtain estimators of the coefficients of Eq. (6) using standard linear regression machinery. We denote the p observations of the dependent variable Y by $y_1, \ldots, y_p$, and the observations for the jth independent variable X_j by $x_{1j}, \ldots, x_{pj}$. We assume independent experimental observations, which implies that observed errors $\varepsilon_1, \ldots, \varepsilon_p$ are independent identically distributed (IID) random variables, where $\varepsilon_i = \log y_i - \beta_0 - \Sigma_{j=1}^{n} \beta_j \log x_{ij}$. Using matrix notation, we have

$$\tilde{y} = \begin{pmatrix} \log y_1 \\ \vdots \\ \log y_p \end{pmatrix}, \quad \text{and } \tilde{X} = \begin{bmatrix} 1 & \log x_{11} & \cdots & \log x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \log x_{p1} & \cdots & \log x_{pn} \end{bmatrix}$$

The estimate for the coefficients in model (6) that minimizes the residual sum of squares is the solution to the system of normal equations $\tilde{X}^T \tilde{X} \beta = \tilde{X} \tilde{y}$, where the superscript T denotes the transpose of a matrix. We denote this estimate by the $n+1$ dimensional vector $\hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_n)$, and the estimate of the independent variable becomes $\hat{Y} = e^{\hat{\beta}_0} \Pi_{j=1}^{n} X_j^{\hat{\beta}_j}$.

The estimate $\hat{\beta}$, however, will generally not satisfy the units constraint. Therefore, we have to select the coefficients that minimize the residual sum of squares only among those that satisfy Eq. (5), which as we explain below is equivalent to a linear constraint of the form $R\beta = b$. With this additional constraint, the estimate of the coefficients in model (6) that minimizes the residual sum of squares and satisfies the units constraint is the solution to the problem

$$\begin{aligned} &\min_{\beta} (\tilde{y} - \tilde{X}\beta)^T (\tilde{y} - \tilde{X}\beta) \\ &\text{s.t. } R\beta = b \end{aligned} \tag{7}$$

To represent the units constraint in linear form, assume that q reference units (m, kg, s,…) are the building blocks for the units of the dependent and all independent variables in the problem. The units constraint can be expressed by $R\beta = b$, where b is a q-dimensional vector such that b_i is the exponent of reference unit i in the units of the dependent variable Y, and matrix R is q by $n+1$ such that R_{ij} is the exponent of reference unit i in the units of variable X_j, for $j = 0, 1, \ldots, n$. Note that we include a variable X_0, which accounts for the constant dimensionless term, thus $R_{i0} = 0$ for all i. For example, this notation implies that $R\bar{a} = b$, for $\bar{a}^T = (a_0, a_{01}, \ldots, a_{0n})$.

2.3 Generation of Dimensionless Groups. The solution to problem (7), denoted with the vector $\beta^0 = (\beta_0^0, \beta_1^0, \ldots, \beta_n^0)$, estimates the coefficients that construct the model in Eq. (4). This solution, however, can have all coefficients different from zero, which leads to a model that, although satisfies the units constraint and has minimal residual sum of squares, is not simple as it includes all independent variables and lacks physical interpretation. We aim to separate this complicated model into a simpler power law and dimensionless groups, as in Eq. (3). For this, consider $\tilde{\beta}$ another $n+1$ dimensional vector that satisfies the units constraint and is simpler, i.e., has few nonzero coefficients, and let $\delta = \beta^0 - \tilde{\beta}$. Using this decomposition, we express the model with β^0 in the following form, in line with Eq. (3):

$$\hat{Y} = e^{\beta_0^0} \prod_{j=1}^{n} X_j^{\beta_j^0} = e^{\tilde{\beta}_0} \prod_{j=1}^{n} X_j^{\tilde{\beta}_j} \left(e^{\delta_0} \prod_{j=1}^{n} X_j^{\delta_j} \right)$$

The first factor in the right side of this equation corresponds to a simple power law as $\tilde{\beta}$ has few nonzero coefficients and, since $R\delta = 0$ by construction, the second corresponds to a dimensionless group.

The proposed algorithm simplifies the model, and in the process identifies the dimensionless groups by removing independent variables from the model one at a time. A variable X_j is removed from the model by forcing the solution to satisfy $\beta_j = 0$, which is enforced by the linear constraint $e_j^T \beta = 0$, where e_j is the jth canonical vector in $n+1$ dimensions. After k iterations of the algorithm, exactly k independent variables, let us say $X_{i_1}, \ldots, X_{i_k}$, have been removed from the model. Therefore, feasible models now must also satisfy $e_{i_1}^T \beta = 0, \ldots, e_{i_k}^T \beta = 0$, which can be written in matrix form as $M_k \beta = 0$ for the k by $n+1$ matrix $M_k = [e_{i_1}, \ldots, e_{i_k}]^T$. For each iteration k, let $\beta^k = (\beta_0^k, \ldots, \beta_n^k)$ be the solution to Eq. (7) that satisfies, in addition, $M_k \beta = 0$. To remove an additional variable X_j with $\beta_j^k \neq 0$ from the model, we simply add the constraint $\beta_j = 0$ to the problem. Therefore, the reduced model is given as the solution to the optimization problem

$$\begin{aligned} z_k^*(j) = &\min_{\beta} (\tilde{y} - \tilde{X}\beta)^T (\tilde{y} - \tilde{X}\beta) \\ &\text{s.t. } R\beta = b \\ &M_k \beta = 0 \\ &\beta_j = 0 \end{aligned} \tag{8}$$

Equation (8) minimizes a strictly convex function over linear constraints and therefore has a unique solution that can be computed solving the first-order optimality conditions, which for this problem are a linear system of equation [22].

From all variables with $\beta_j^k \neq 0$, we eliminate at iteration $k+1$ the variable i_{k+1}, which makes $z_k^*(j)$ smallest. The resulting model, with coefficients β^{k+1}, best fits the given data in a least-squares sense, satisfies the units constraint, and has $k+1$ coordinates equal to zero. This last constraint is encoded in the matrix $M_{k+1}^T = [M_k^T e_{i_{k+1}}]$. After eliminating k parameters from the model, we obtain a simplified model β^k and the dimensionless groups $\delta^i = \beta^{i-1} - \beta^i$ for $i = 1, \ldots, k$. These vectors satisfy $\beta^0 = \Sigma_{i=1}^{k} \delta^i + \beta^k$, which implies the following expression:

$$\hat{Y} = e^{\beta_0^k} \prod_{j=1}^{n} X_j^{\beta_j^k} \left[\prod_{i=1}^{n} \left(e^{\delta_0^i} \prod_{j=1}^{n} X_j^{\delta_j^i} \right) \right]$$

The process of sequentially eliminating variables is possible while the linear system of constraints has a solution. We show below that the dimensionless groups generated δ^k are linearly independent; therefore, the number of iterations can be at most $n+1-\text{rank}(R)$.

To see that $\delta^1, \ldots, \delta^K$ are linearly independent, assume without loss of generality that coordinate k is eliminated to construct β^k, then by construction we have that $\delta_j^k = 0$ for all $j < k$ and $\delta_k^k \neq 0$, thus the matrix $[\delta^1, \ldots, \delta^K]$ is lower triangular with nonzero diagonal elements, which implies the linear independence. In addition we know there are $\text{rank}(R)$ linearly independent rows of R, and thus the vector space orthogonal to the rows of R is of dimension $n+1-\text{rank}(R)$. Since all δ^k are orthogonal to the rows of R and linearly independent, there can be at most $n+1-\text{rank}(R)$ of them.

The process of generating solutions β^k, each with k variables removed, can be interpreted as backward elimination for linear regression, see [23]. Here, the backward elimination is applied to a linearly constrained linear regression problem.

3 Algorithm SLAW

The algorithm SLAW (Scaling LAWs) uses experimental data, information regarding the units of variables, and information about the accuracy needed. It outputs a physically meaningful simple model and a set of dimensionless groups that explain the dependent variable in order of importance.

This algorithm can be broken down into four steps:

1. Find the sequence of models $\{\beta^k\}$ through the backward elimination process that solves Eq. (8).
2. Determine, with user input, which model of the sequence $\{\beta^k\}$ to select, say its $\hat{\beta}$.
3. Round the coefficients in $\hat{\beta}$, obtaining a physically meaningful simple model β^*.
4. Perform backward elimination again, to identify dimensionless groups in what is not explained by β^*.

Step 4 is needed to identify the correct dimensionless groups because the rounding procedure in Step 3 creates a model β^* that is slightly different from the model derived in the original regressions. This second application of backward elimination has a significant difference from the first. While the goal of the first application of backward elimination is to find the smallest meaningful scaling law, we start the second set of iterations with a scaling law already in place, β^*. Although the change in scaling law is typically small, the dimensionless groups corresponding to the new law are not necessarily the same, and we perform another backward elimination to find them. The goal of this second backward elimination is simply to find a scaling law for the error between the input data and the rounded scaling law, characterized by $\delta = \beta^0 - \beta^*$. We now describe each step of the SLAW algorithm.

3.1 Model Reduction Algorithm. The following algorithm identifies the dimensionless groups of the linear regression model in order of significance to the dependent variable. The inputs are the experimental data and units constraint data; the output are the sequence of estimators $\{\beta^k\}$ and dimensionless groups $\{\delta^k\}$.

—Algorithm MODEL REDUCTION $(\tilde{X}, \tilde{y}, R, b)$:

Step 1: Solve (7), let β^0 be the solution. $k=0$.

Step 2: Find the coordinate j_k that minimizes $z_k^*(j)$ [problem (8)]. Let β^{k+1} be the solution to $z_k^*(j_k)$. Let $\delta^{k+1} = \beta^k - \beta^{k+1}$.

Step 3: Let $k=k+1$. Repeat Step 2 while some variable can be eliminated from β^k.

—Output: $\{\beta^k\}_{k=0}^{n-\text{rank}(R)}$ and $\{\delta^k\}_{k=1}^{n-\text{rank}(R)}$.

3.2 Selecting Model. By definition, β^0 defines the model that best explains the dependent variable out of the sequence generated by the DIMENSION REDUCTION algorithm. This solution has the smallest residual sum of squares (RSS) but is not a simple solution, as it uses all n variables and constant term. In contrast, $\beta^{n-\text{rank}(R)}$ is the simplest solution that satisfies the units constraints and at the same time has the largest RSS. User input is used to select a model that balances simplicity with accuracy.

The inputs are the experimental data, the sequence of estimators $\{\beta^k\}$, and a user-supplied tolerance TOL. The output is the estimator $\hat{\beta}$, the simplest estimator in the sequence with average relative error $<$TOL.

—Algorithm Selection $(\tilde{X}, \tilde{y}, \{\beta^k\}_k, \text{TOL})$

Step 1: Let $\hat{\beta}$ be the model β^k with less coefficients $\neq 0$ that satisfies

$$\alpha_k = \sqrt{\frac{1}{p}(\tilde{y} - \tilde{X}\beta^k)^T(\tilde{y} - \tilde{X}\beta^k)} \leqslant \text{TOL}$$

—Output: $\hat{\beta}$.

The quantity α_k used to determine the cutoff is the square root of an average RSS of the linear models considered. We opted for this criteria due to the engineering interpretation of this quantity, outlined below. There are a number of classic statistical tests, such as the F test, that are used in linear regressions. However, such tests do not have the direct interpretation of α_k, and do not apply to the linearly constrained regressions we are considering; additionally, they require extra assumptions on the distribution of the errors.

Note that if we let $\hat{y}^k$ be such that $\log \hat{y}^k = \tilde{X}\beta^k$, and define $\Delta y^k = \hat{y}^k - y$, then we have

$$\alpha_k^2 = \frac{1}{p}(\tilde{y} - \tilde{X}\beta^k)^T(\tilde{y} - \tilde{X}\beta^k) = \frac{1}{p}\sum_{i=1}^{p} (\log y_i - \log \hat{y}_i^k)^2$$

$$= \frac{1}{p}\sum_{i=1}^{p} \log^2\left(1 + \frac{\Delta y_i^k}{y_i}\right)$$

Since for small values of $\Delta y_i^k / y_i$ we have that $\log(1 + (\Delta y_i^k / y_i)) \sim (\Delta y_i^k / y_i)$, an interpretation for the cutoff criteria of algorithm selection is that the quantity α_k^2 corresponds to the average squared relative error

$$\alpha_k^2 = \frac{1}{p}\sum_{i=1}^{p} \log^2\left(1 + \frac{\Delta y_i^k}{y_i}\right) \sim \frac{1}{p}\sum_{i=1}^{p} \left(\frac{\Delta y_i^k}{y_i}\right)^2$$

We refer to α_k as the average relative error (avg. RE).

3.3 Rounding the Model. The input to function Round is an estimator $\hat{\beta}$, and its output is a related estimator β^* that satisfies the units constraints, has all coefficients rounded to a number with decimals, either 0, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, or $\frac{3}{4}$, and minimizes the increase in error. Note that since 0 is a round number, β^* has all the zeros of $\hat{\beta}$. The choice of using quarters and thirds as the finest division strikes a balance between independence from experimental error and physical meaning. A large number of known laws for engineering problems, probably the vast majority, involve exponents consistent with this choice. For example, scaling laws for a boundary layer involve whole numbers and halves in the exponents, and if the boundary layer involves heat transfer, we obtain exponents with thirds [24]. Additionally, rounding the exponents reduces their variation with experimental error.

To obtain β^*, the algorithm sequentially fixes the closest exponent to its rounded version, and solves a linearly constrained linear regression similar to problem (8). The problems solved at each iteration include the units constraint and linear constraints that fix

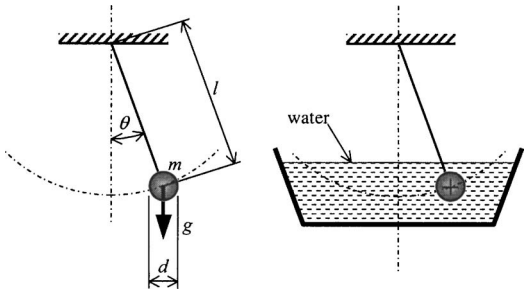

Fig. 2 Representation of a simple pendulum and its elements

coordinates to their rounded values. This is iterated until all exponents are fixed; the resulting β^* satisfies the units constraint, as this is enforced at very iteration.

3.4 Overall SLAW Algorithm. The overall algorithm performs the four steps outlined at the start of this section. For an input of statistical data $\tilde{y}$, $\tilde{X}$, and units constraint R and b, the algorithm SLAW does

Step 1: Run Dimension Reduction $(\tilde{X},\tilde{y},R,b)$, and obtain the sequence $\{\beta^k\}_k$, $\{\delta^k\}_k$.

Step 2: Run Selection $(\tilde{X},\tilde{y},\{\beta^k\},\mathrm{TOL})$, and obtain $\hat{\beta}$.

Step 3: Run Round $(\hat{\beta})$, obtaining β^*.

Step 4: Run Dimension Reduction $(\tilde{X},\tilde{y}-\tilde{X}\beta^*,R,0)$, to obtain $\{\tilde{\beta}^k\}_k$ and $\{\tilde{\delta}^k\}_k$.

—Output: β^* and $\{\tilde{\delta}^k\}_k$.

4 Examples

In this section, we present the results of applying the algorithm SLAW to three different physical experiments: a pendulum, the bonding of ceramics to metals, and a "punch test" recently developed at Exponent, Inc. As we show below, in each case SLAW "rediscovered," with a minimum of human input, the scaling laws that match previous analysis of the problem.

The input files for these three examples can be downloaded from [25]. For each problem, there are two text input files, one containing the experimental data and the other describing the units matrix.

4.1 Period of a Pendulum. Figure 2 shows a schematic of a simple pendulum and some of its elements. When the only relevant force acting on the pendulum is the force of gravity, and for small oscillations, the period of the pendulum is given by the following formula:

$$T = 2\pi\sqrt{l/g} \tag{9}$$

where T is the period (the dependent variable in this problem), l is the length of the string, and g is the acceleration of gravity. In what follows, we denote by T the observed period, and by T_{theory} the quantity in Eq. (9). We now use the SLAW algorithm to discover this relationship, assuming that the period of the pendulum is determined by the parameters in Table 1.

Table 1 Parameters involved in the pendulum example

Symbol	Units	Description
T	s	period (dependent variable)
m	kg	mass of the bob
l	m	length of the pendulum
θ	rad	initial angle
d	m	characteristic dimension of the bob
ρ	$\mathrm{kg/m^3}$	density of fluid surrounding the bob
g	$\mathrm{m/s^2}$	acceleration of gravity

These parameters account not only for inertial forces and gravity, but also for drag, rotational inertia of the bob, and initial angle. If these last three effects are neglected, dimensional analysis alone can solve Eq. (9), with the exception of the constant, which can be estimated by a single experiment. When these three effects are considered, dimensional analysis alone does not provide a unique formula. In this work, we considered these effects, which make the problem more complex, but more representative of real situations.

4.1.1 Input for SLAW. The input for SLAW are the set of experimental data and a matrix describing the units of the dependent and independent variables. The set of experimental data consists of a table listing the measured value of the period for different values of the parameters. Table 2 below displays the matrix R, which contains the units of the dependent variable and the parameters. This matrix was constructed using the units listed in Table 1; each element corresponds to the exponent of a unit (listed in the left column) for a given parameter (listed in the top row).

4.1.2 Output From SLAW. To illustrate the workings of SLAW, we here illustrate the output of the different steps of the SLAW algorithm. Table 3 displays the result of the iterations to obtain the initial simplest, meaningful, and dimensionally correct scaling law.

In this table, the first iteration corresponds to a power law without the units constraint. This model corresponds to what is typically used in engineering to fit experimental data. It is the most mathematically accurate of all possible power laws; however, it is physically incorrect, as the estimates of the model do not have the right units. The second iteration is a modification of the first power law, chosen to provide the correct units with the minimum increase in fitting error. These first two iterations use all parameters of the problem, regardless of their relevance. The third to sixth iterations remove the least significant independent variables one at a time; for example, the exponent for θ in the third iteration

Table 2 Matrix of reference units *R* for the pendulum example

Units	T	m	l	θ	d	ρ	g
m	0	0	1	0	1	−3	1
kg	0	1	0	0	0	1	0
s	1	0	0	0	0	0	−2

Table 3 Results of Step 1 of SLAW, for the pendulum example

Iter.	Param.	Constant	m	l	θ	d	ρ	g	avg. RE
1	6	0.117	0.000	0.502	0.023	−0.004	0.021	0.266	0.0197
2	6	2.050	−0.021	0.506	0.025	0.057	0.021	−0.500	0.0197
3	5	1.998	−0.016	0.500	0.000	0.049	0.016	−0.500	0.0197
4	4	1.861	−0.006	0.517	0.000	0.000	0.006	−0.500	0.0200
5	3	1.872	0.000	0.500	0.000	0.000	0.000	−0.500	0.0260
6	2	0.000	0.000	0.500	0.000	0.000	0.000	−0.500	1.8582

Table 4 Result of Step 3 of SLAW, for the on pendulum example

Iter.	Param.	Constant	m	l	θ	d	ρ	g	avg. RE
1	6	−0.001	0.000	0.002	0.023	−0.004	0.021	−0.002	0.020
2	5	0.178	−0.021	0.006	0.025	0.057	0.021	0.000	0.021
3	4	0.195	−0.021	0.000	0.022	0.064	0.021	0.000	0.022
4	3	0.125	−0.016	0.000	0.000	0.048	0.016	0.000	0.026
5	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.044

is set to zero after determining that the corresponding average relative error in this case was smaller than that of setting to zero the exponent of any other parameter. The sixth iteration is the minimum expression that provides the correct units after removing parameters in this order. Since the average relative error (avg. RE) jumps from 2.6% to 185.8% in the last iteration, we decide to use Iteration 5 as the basis of the final scaling law. This corresponds to the simplest model that provides an acceptable error. The selection of this model constitutes Step 2 of SLAW and it is automatically performed using a criteria that the average relative error (avg. RE) has a tolerance of TOL=0.2.

The behavior of exhibiting dramatic increases in the fitting error of models as variables are removed is the norm, not the exception, for physical processes that can be described with scaling laws. In this case, there is little error as the less relevant terms are discarded, and the error jumps suddenly when one of the essential terms of the power law is eliminated. This sudden jump in error can often be used by SLAW to automatically identify a cutoff point in the simplifications. The scaling law automatically obtained by SLAW for this problem is

$$T_{\mathrm{SLAW1}} = e^{1.872}\sqrt{l/g} \tag{10}$$

where T_{SLAW1} is the estimation of the period based in this scaling law. The constant factor obtained is within a 3% error from the exact coefficient 2π, and the functional dependence is the same as predicted by theory in Eq. (9).

Table 4 displays the iterations that determine the dimensionless groups ranked by relevance. Each line of this table corresponds to a dimensionless group that can be used to improve the predictions of the scaling law.

The first iteration corresponds to fitting the error of the scaling law (10) with a dimensionless group that considers all parameters. The fourth iteration corresponds to the simplest dimensionless group obtained after eliminating the less relevant parameters: $\Pi_1=(\rho d^3/m)^{0.016}$. The last iteration (the fifth in this case) always corresponds to a constant numerical factor. In this case, this factor is one (this row contains only zeros) because the scaling law obtained in the first iteration already has a constant numerical factor. Incorporating the most relevant dimensionless group yields the following scaling law:

$$T_{\mathrm{SLAW2}} = e^{1.997}\sqrt{\frac{l}{g}}\left(\frac{\rho d^3}{m}\right)^{0.016} \tag{11}$$

which has a better predictive value for these experiments and incorporates the effect of fluid drag. Fluid drag is relevant in this case because we measured pendulums surrounded by air and water, which differ by a factor of three orders of magnitude in their density. In the set of relevant parameters, we did not include the fluid viscosity. This choice is based on engineering insight that viscous drag is negligible. A metric for the relative relevance of viscosity is the Reynolds number, which in our experiments varied between 10^2 and 10^4. These values correspond to a flow in which viscous drag is unimportant.

Figure 3 illustrates the predictive capabilities of the scaling law obtained with the SLAW algorithm. Figure 3 plots the observed period on the vertical axis versus the known law T_{theory} and the scaling laws, T_{SLAW1} and T_{SLAW2}, on the horizontal axis. We also plot the identity for comparison purposes. We note that all three models predict the observed period of the pendulum very accurately when the pendulum is surrounded by air. However, both T_{theory} and T_{SLAW1} are slightly off the observed values when the bob was submerged in water. This effect is corrected by including the most important dimensionless group that accounts for the effect of drag, and can be seen in the plot of T_{SLAW2}.

Table 5 presents the correlation with the observed period and the average residual sum of squares for T_{theory}, T_{SLAW1}, and T_{SLAW2}. The correlation provides an indication of how well the different models capture the functional dependence; in engineering terms, it captures the *precision* of the models. The average relative error captures not only the trends, but also how close the predictions and the measured data are. In engineering terms, this would be the *accuracy* of the models.

4.2 Strain Energy in Ceramic-to-Metal Joining. Figure 1 shows the geometry of the problem, which consists of two long cylinders; one made of ceramic and the other of metal. These two cylinders are joined at their circular bases at high temperature. The temperature variation between the hot joining temperature and the cooler room temperature causes the ceramic and the metallic cylinder to decrease slightly in size. Typically, the metallic cylinder will shrink more than the ceramic cylinder, causing very large stresses on and around the interface of the joint. These stresses weaken the joint; therefore, the calculation of these

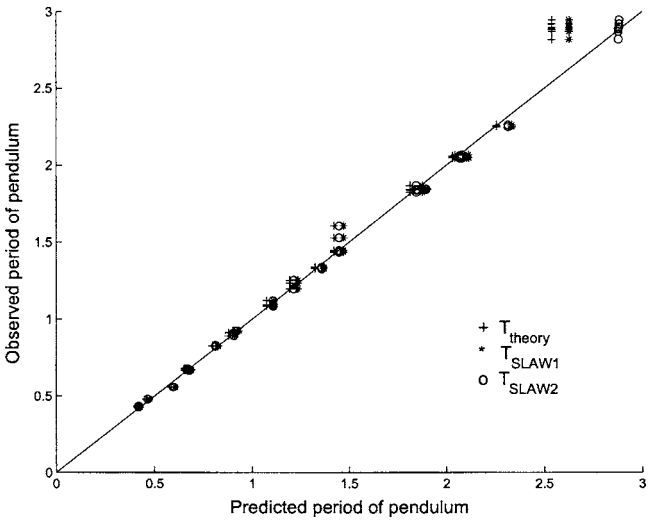

Fig. 3 Theoretical and SLAW-generated scaling laws for the pendulum example

Table 5 Correlation and error of the scaling laws for the pendulum example

Model	Correlation with T	avg. RE
T_{theory}	0.992	0.056
T_{SLAW1}	0.992	0.044
T_{SLAW2}	0.999	0.025

Table 6 Parameters involved in the ceramic-to-metal bonding example

Symbol	Units	Description
U	m^3 Pa	total elastic strain energy in the ceramic (dependent variable)
E_c	Pa	elastic modulus of the ceramic
E_m	Pa	elastic modulus of the metal
σ_Y	Pa	yield stress of metal
r	m	radius of cylinders
ε_T		differential thermal shrinkage

stresses is essential. The metric for these stresses is the "total elastic strain energy" U (units=Pa·m^3) accumulated in the ceramic. The total strain energy is the integral of the volumetric strain energy over the total volume of the ceramic cylinder.

Scaling factors exist for cases in which the metallic cylinder behaves elastically [26]. Similar scaling factors for when the metallic cylinder experiences nonlinear plasticity have been obtained only recently, by manual analysis of computational experiments [27]. In this paper, we will show how SLAW automatically obtains the same scaling factor of [27]. Similar to what is done in [27], we consider that the parameters of Table 6 represent the total strain energy in the ceramic. These parameters correspond to an elastic-plastic metal and a linear elastic ceramic.

4.2.1 Input for SLAW. Table 7 lists the elastic strain energy, which is the dependent variable that we wish to analyze, and the problem parameters for nine numerical simulations of ceramic-to-metal joints. In this problem, an engineering criterion was used to discard the length of the cylinders as a relevant parameter. The reason is that the cylinders considered are long enough, such that the far end does not influence the joined faces.

Table 8 displays the matrix of units R for this example.

4.2.2 Output From SLAW. We now present the output obtained from the SLAW algorithm for the above input data. The main outputs of SLAW are a simple scaling law, with rounded coefficients, and dimensionless groups that identify the most relevant parameters to describe the fitting error of this scaling law. The following scaling law was obtained after Step 1 of SLAW, using an avg. RE criteria with a tolerance of TOL=0.2, and the rounding procedure of Step 3.

$$U_{\text{SLAW1}} = \frac{\sigma_Y^2 r^3}{E_c} \tag{12}$$

This is the same scaling law that is obtained in [27] through ad hoc analysis and physical considerations. Table 9 displays the second set of iterations that determine the dimensionless groups and

Table 8 Matrix of reference units R for the ceramic-to-metal bonding example

Units	U	E_c	E_m	σ_Y	r	ε_T
Pa	1	1	1	1	0	0
m	3	0	0	0	1	0

numerical constant ranked by relevance.

Incorporating this last dimensionless group, the constant factor, in the scaling law, we obtain

$$U_{\text{SLAW2}} = e^{-0.333}\frac{\sigma_Y^2 r^3}{E_c} \tag{13}$$

In Table 10, we see that the average residual sum-of-square error is greatly reduced by considering this constant factor. This table also shows the correlation between the observed strain energy and what is predicted by U_{SLAW1} and U_{SLAW2}. Here, we additionally consider the scaling law $U_{\text{SLAW-lin}}$, given by the simple model selected by the avg. RE criteria, but prior to the rounding procedure.

$$U_{\text{SLAW-lin}} = \frac{\sigma_Y^{2.045} r^3}{E_c^{1.045}} \tag{14}$$

The model $U_{\text{SLAW-lin}}$ is the best simple linear regression model, but because the coefficients are not rounded, the model lacks physical interpretation. We notice that adding this physical interpretation by rounding the coefficients somewhat deteriorates the predictive value of our model; however, we still keep very high correlation with the observed strain energy and practically all the increased error can be recovered by incorporating the constant factor.

The predictive significance of these scaling laws is also observed in Fig. 4, which plots the observed strain energy versus what is predicted for the different models: $U_{\text{SLAW-lin}}$, U_{SLAW1}, and U_{SLAW2}. The figure shows that the experimental points fall closely around a straight line of slope one. We notice that the best prediction is given by model $U_{\text{SLAW-lin}}$, and that rounding the coefficients consistently overestimates the strain energy. This effect is compensated by incorporating the constant term in U_{SLAW2}.

4.3 Maximum Stress in the Punch Test. The punch test is an ASTM standard test, developed to determine mechanical properties of materials, such as ultrahigh molecular weight polyethylene used in surgical implants [28]. The test consists of using a spherical tip to push the center of a disk constrained at the edge. Figure 5 shows the geometry of the problem, where a sphere of radius r is pushing up on the center of a disk of radius L and width t.

The goal in this example is to obtain a scaling law that reproduces the maximum stress in the disk (stress at point A in Fig. 5)

Table 7 Input database containing the results of nine numerical experiments [27] for the ceramic-to-metal bonding example

Ceramic	Metal	U 10^{-2} Pa·m^3	E_c 10^{11} Pa	E_m 10^{11} Pa	σ_Y 10^8 Pa	r 10^{-3} m	ε_T 10^{-3}
Si_3N_4	Cu	0.423	3.04	1.28	7.58	6.25	6.85
Si_3N_4	Ni	1.52	3.04	2.08	1.48	6.25	5.15
Si_3N_4	Nb	2.80	3.04	1.03	2.40	6.25	2.10
Si_3N_4	Inco600	3.78	3.04	2.06	2.50	6.25	5.15
Si_3N_4	AISI 304	3.88	3.04	2.06	2.56	6.25	7.10
Si_3N_4	AISI 316	4.91	3.04	1.94	2.90	6.25	7.00
Al_2O_3	Ti	1.04	3.58	1.20	1.72	6.25	0.505
Al_2O_3	Inco600	3.00	3.58	2.06	2.50	6.25	2.95
Al_2O_3	AISI 304	3.16	3.58	2.00	2.56	6.25	4.90

Table 9 Result of Step 3 SLAW for the ceramic-to-metal bonding example

Iter.	Param.	Constant	E_c	E_m	σ_Y	r	ε_T	avg. RE
1	5	0.138	−0.151	0.192	−0.224	−0.701	0.128	0.026
2	4	−1.051	0.050	0.171	−0.221	−0.000	0.143	0.028
3	3	−1.087	−0.000	0.221	−0.221	0.000	0.131	0.030
4	2	0.642	−0.000	0.000	−0.000	−0.000	0.174	0.082
5	0	−0.333	−0.000	0.000	−0.000	0.000	−0.000	0.163

as a function of the displacement induced by the spherical tip. This problem has been addressed in [29], by an ad hoc analysis of computational experiments. Here, we show how SLAW automatically recreates the results obtaining the same scaling law. Similar to the work in [29], we consider a number of problem parameters that describe the elastic regime of the problem; these problem parameters and the dependent variable are listed in Table 11.

Table 10 Correlation and error of the scaling laws for the ceramic-to-metal bonding example

Model	Correlation with U	avg. RE
$U_{\text{SLAW-lin}}$	0.9840	0.1694
U_{SLAW1}	0.9835	0.3704
U_{SLAW2}	0.9835	0.1626

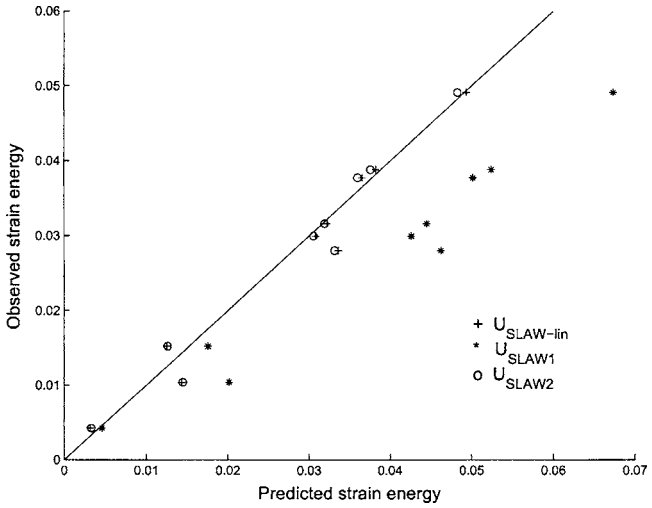

Fig. 4 SLAW-generated scaling laws for the ceramic-to-metal bonding example

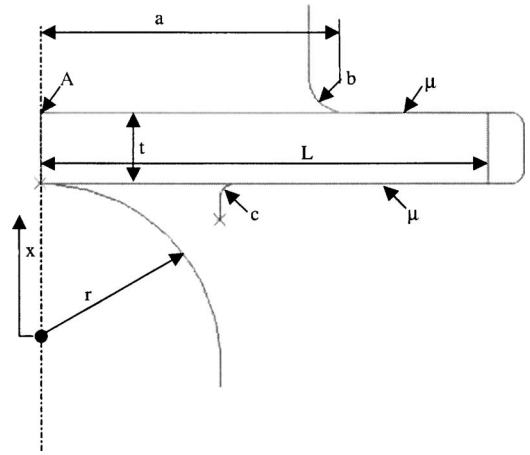

Fig. 5 Geometry of the punch test

4.3.1 Input for SLAW. This example considers the effect of Poisson's modulus υ. In solid mechanics, this parameter typically appears as a combination of υ, $1-\upsilon$, and $1+\upsilon$; therefore, we assigned three columns corresponding to each possibility. Inputs for SLAW are the table of experimental data and the units information of the variables, which is displayed in Table 12 below.

4.3.2 Output From SLAW. We now present the output obtained from the SLAW algorithm for the input data presented above. The principal outputs from SLAW are the simple, physically meaningful, scaling law and the most relevant dimensionless groups in describing the fitting error of the rounded scaling law. By running Step 1, using the avg. RE criteria with a tolerance of TOL=0.2, and the rounding procedure of Step 3, we obtain the following scaling law:

$$\sigma_{\text{SLAW1}} = \frac{E x_{\max}}{a} \tag{15}$$

We note that Step 1 of the algorithm realized 13 total iterations, and the last model with an avg. RE less than the tolerance of 0.2 is iteration 12. This scaling law is consistent with [29]. Again, the SLAW algorithm found this law automatically, while in [29] it took several hours of identifying trends manually. There is, however, a significant difference between Eq. (15) and the result in [29], which is the absence of Poisson's coefficient in the former, while the latter indicates that $1-\upsilon$ should appear in an analytical expression. SLAW indicates that the error of neglecting υ (or any of its variations) in the error of the approximation is similar to the error of neglecting other parameters that [29] also ignores. This does not mean that υ would not appear in an algebraic deduction; it means that for the data set analyzed, neglecting the effect of υ does not introduce a significant error. For comparison purposes, we will present the correlation and predictive value of the model found in [29], which we refer to as σ_{BK}.

We can reduce the error of σ_{SLAW1} by using the dimensionless groups and numerical constants obtained in Step 4 of SLAW. Table 13 displays a subset of the second group of iterations to determine the dimensionless groups and numerical constant, ranked by

Table 11 Parameters involved in the punch test example

Symbol	Units	Description
$\sigma_{\max}$	Pa	maximum stress at A (dependent variable)
L	m	radius of circular sample
t	m	thickness of circular sample
r	m	radius of spherical punch tip
a	m	radius of contact of top constraint
b	m	corner radius of top constraint
c	m	corner radius of bottom constraint
E	Pa	Elastic modulus of sample material
μ		Friction coefficient
$x_{\max}$	m	Maximum displacement of punch tip
υ		Poisson's modulus
$1-\upsilon$		1−Poisson's modulus
$1+\upsilon$		1+Poisson's modulus

Table 12 Matrix of reference units *R* for the punch test example

Units	σ_{max}	L	t	r	a	b	c	E	μ	x_{max}	v	$1-v$	$1+v$
Pa	1	0	0	0	0	0	0	1	0	0	0	0	0
m	0	1	1	1	1	1	1	0	0	1	0	0	0

Table 13 Summary of results of Step 4 of SLAW for the punch test example

Iter.	Param.	Constant	L	t	r	μ	x_{max}	v	$1+v$	avg. RE
1	12	1.19	0.40	−0.10	−0.14	−0.10	−0.17	0.72	−0.72	0.030
2	11	0.69	0.37	−0.10	−0.12	−0.08	−0.16	0.96	−2.57	0.031
3	10	0.72	0.37	−0.10	−0.11	−0.08	−0.16	0.97	−2.67	0.031
4	9	0.69	0.38	−0.10	−0.11	−0.08	−0.16	0.96	−2.62	0.032
5	8	0.72	0.38	−0.11	−0.11	−0.07	−0.15	0.96	−2.63	0.032
6	7	0.78	0.38	−0.11	−0.11	−0.07	−0.16	0.98	−2.73	0.032
7	6	−0.66	0.37	−0.12	−0.10	−0.08	−0.15	0.37	0.00	0.033
8	5	−0.49	0.35	−0.11	−0.10	0.00	−0.14	0.36	0.00	0.038
9	4	0.01	0.22	−0.12	−0.10	−0.00	−0.00	0.39	−0.00	0.044
10	3	0.10	0.15	−0.15	−0.00	−0.00	−0.00	0.43	−0.00	0.049
11	1	0.35	−0.00	0.00	−0.00	−0.00	−0.00	0.39	−0.00	0.061
12	0	−0.14	0.00	0.00	0.00	−0.00	0.00	0.00	−0.00	0.092

relevance.

The constant numerical factor compensates for the error introduced in the rounding of the exponents. Including this constant factor, we obtain the following scaling law:

$$\sigma_{SLAW2} = e^{-0.138}\frac{Ex_{max}}{a} \tag{16}$$

In Table 14, we show the correlation to maximum stress and average residual sum-of-square error of σ_{BK}, σ_{SLAW1}, and σ_{SLAW2}. We note that the average residual sum-of-square error is greatly reduced by adding the constant factor to σ_{SLAW1}. Although this table also shows that the scaling laws obtained from SLAW are slightly less correlated than σ_{BK} to the observed maximum stress, they are comparable and were obtained through an automatic methodology. In fact, σ_{SLAW2} is slightly less correlated to σ_{max} than σ_{BK}, but has smaller average residual sum of squares.

The predictive significance of these scaling laws is also observed in Fig. 6, which plots the observed maximum stress versus what is predicted for the different models: σ_{BK}, σ_{SLAW1}, and σ_{SLAW2}. The figure shows that the experimental points fall closely around a straight line of slope one; with model σ_{BK} overestimating the maximum stress, and model σ_{SLAW1} underestimating the maximum stress. Here, again, the model that incorporates the most significant dimensionless group (in this case, a constant factor) is seen as more representative of the observed σ_{max}, which is quantified with a smaller average residual sum of squares.

5 Conclusions

In this paper, we propose a new algorithm, named SLAW, to obtain scaling laws for complex systems from experimental data. The algorithm SLAW combines the ideas of dimensional analysis with statistical linear regression to obtain a representative model of the complex system. The algorithm automatically generates models of the system that not only achieve a high correlation with the observed experimental data, but are at the same time simple and physically meaningful. From linear regression, SLAW uses the concept of backward elimination to simplify the models considered, and to obtain a rank of the importance of problem parameters. Dimensional analysis is used to obtain a model that satisfies the units constraints and that considers only simple exponents. The proposed SLAW algorithm is able to:

1. find the simple scaling law that rules many real-life engineering problems.
2. provides a ranking of the significance of the parameters in the problem.

These two features of SLAW make it a useful tool for engineering design, where simple and approximate laws can be used by engineers to narrow down on a configuration during the conceptual stage of design. Even though the SLAW algorithm is based on the standard tools of dimensional analysis and linear regressions, it outperforms both of these techniques in identifying useful mod-

Table 14 Correlation and error of the scaling laws for the punch test example

Model	Correlation with σ_{max}	avg. RE
σ_{BK}	0.9592	0.1527
σ_{SLAW1}	0.9573	0.2507
σ_{SLAW2}	0.9573	0.1326

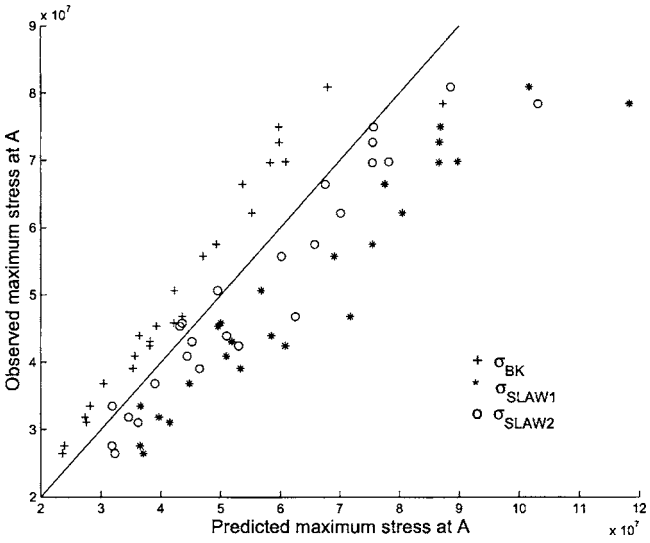

Fig. 6 SLAW-generated scaling laws for the punch test example

els for complex systems. The reason for this is that, more often than not, complex engineering problems can be explained using only a few of the physical parameters, and SLAW is able to identify these important parameters.

The SLAW algorithm makes two central assumptions on the physical system: (i) that a single physical regime is being modeled and (ii) that at least all the relevant parameters of the system have been included. The first assumption is necessary to obtain the simplest possible model. Since different regimes are characterized by different scaling laws, a data set that includes more than one regime would necessarily consider more parameters; this increases the complexity of the expression without adding to the physical understanding of the problem. The second assumption is standard for both dimensional analysis and linear regression models for physical systems. Clearly, it is not possible to explain a system if a key parameter is omitted.

We apply the SLAW algorithm to three very different examples, and note that SLAW automatically obtains the correct scaling law in each one. The resulting scaling laws are simple, correlate well with the experimental data, and are obtained using minimal user expertise. The only alternative method to obtain scaling laws in the last two examples involve time-consuming expertise-intensive analysis of the data.

Future work in this algorithm will involve, for example, investigating efficient alternative methods of selecting simple models. Currently, the algorithm eliminates one parameter at a time through backward elimination, and therefore does not consider all possible combinations of parameters. Therefore, the algorithm can, in theory, miss the most adequate model; to evaluate the impact of this on engineering practice, it is important to investigate other methods of generating simple models such as forward selection or even an exhaustive analysis of all combinations with correct units for small problems. An implementation of forward selection (albeit for predetermined dimensionless groups) was developed by Li and Lee [14].

Another potential improvement is a refined analysis of the dimensionless groups obtained. Currently, the less relevant dimensionless groups consider a larger number of parameters, which are also considered by the most relevant dimensionless groups. The determination of simpler dimensionless groups, minimizing the overlap of the parameters considered, would enhance their intuitive meaning.

Finally, the capability of SLAW of grouping a large amount of data under a single, simple law could be potentially used for clustering algorithms, separating experimental data into fewer and well-defined groups characterized by their different simple scaling laws. Further work will be necessary to explore the application of SLAW to data sets that cover more than one regime, perhaps drawing inspiration from previous multiple-regime work by Li and Lee [14] and the artificial intelligence algorithm ABACUS [17].

Acknowledgment

The authors thank Dr. Jorgen Bergström from Exponent, Inc., for the concepts and calculations used in the punch test example, and to Dr. Jin-Woo Park of Central R&D Institute in Samsung Electro-Mechanics for providing the data related to the ceramic-to-metal bonding.

References

[1] Kokshenev, V. B., 2003, "Observation of Mammalian Similarity Through Allometric Scaling Laws," Physica A, **322**(1-4), pp. 491–505.
[2] Azad, R. K., Bernaola-Galvan, P., Ramaswamy, R., and Rao, J. S., 2002, "Segmentation of Genomic DNA Through Entropic Divergence: Power Laws and Scaling," Phys. Rev. E, **65**(5), p. 051909.
[3] Washio, T., and Motoda, H., 1999, "Extension of Dimensional Analysis for Scale-Types and Its Application to Discovery of Admissible Models of Complex Processes," in *2nd Int. Workshop on Similarity Method*, University of Stuttgart, Stuttgart, pp. 129–147.
[4] Housen, K. R., Schmidt, R. M., and Holsapple, K. A., 1983, "Crater Ejecta Scaling Laws—Fundamental Forms Based on Dimensional Analysis," J. Geophys. Res., **88**(B3), pp. 2485–2499.
[5] Cho, J. Y. N., Newell, R. E., and Sachse, G. W., 1999, "Anomalous Scaling of Mesoscale Tropospheric Humidity Fluctuations," Geophys. Res. Lett., **27**(3), pp. 377–380.
[6] Carlson, J. M., and Doyle, J., 2000, "Power Laws, Highly Optimized Tolerance and Generalized Source Coding," Phys. Rev. Lett., **84**(24), pp. 5656–5659.
[7] de Jong, F. J., and Quade, W., 1967, *Dimensional analysis for economists*, Contributions to Economic Analysis 50, North Holland, Amsterdam.
[8] Szirtes, T., and Rózsa, P., 1997, *Applied Dimensional Analysis and Modeling*, McGraw Hill, New York.
[9] Segel, L. A., 1972, "Simplification and Scaling," SIAM Rev., **14**(4), pp. 547–571.
[10] Mendez, P. F., and Ordóñez, F., 2002, "Determination of Scaling Laws From Statistical Data," in *Fifth Int. Workshop on Similarity Methods*, University of Stuttgart, Stuttgart, pp. 21–31.
[11] Kasprzak, W., Lysik, B., and Rybachuk, A. M., 1999, "Dimensional Analysis in the Identification of Mathematical Models," http://www.immt.pwr.wroc.pl/kniga
[12] Vignaux, G. A., and Scott, J. L., 1999, "Simplifying Regression Models Using Dimensional Analysis," Aust. N. Z. J. Stat., **41**(1), pp. 31–41.
[13] Vignaux, G. A., 2001, "Some Examples of Dimensional Analysis in Operations Research and Statistics," in *4th Int. Workshop on Similarity Methods*, University of Stuttgart, Stuttgart, pp. 247–265.
[14] Li, C. C., and Lee, Y. C., 1990, "A Statistical Procedure for Model-Building in Dimensional Analysis," Int. J. Heat Mass Transfer, **33**(7), pp. 1566–1567.
[15] Dovi, V. G., Reverberi, A. P., Maga, L., and De Marchi, G., 1991, "Improving the Statistical Accuracy of Dimensional Analysis Correlations for Precise Coefficient Estimation and Optimal Design of Experiments," Int. Commun. Heat Mass Transfer, **18**(4), pp. 581–590.
[16] Bradshaw, G., Langley, P., and Simon, H. A., 1980, "Bacon 4: The Discovery of Intrinsic Properties," in *Proc. of the Third Nat. Conf. of the Canadian Society for Computational Studies of Intelligence*, Victoria, BC, pp. 19–25.
[17] Kokar, M. M., 1986, "Determining Arguments of Invariant Functional Descriptions," Mach. Learn., **1**(4), pp. 403–422.
[18] Washio, T., and Motoda, H., 1996, "Discovery of Possible Law Formulae Based on Measurement Scale," in *Proc. of Fourth Int. Workshop on Rough Sets, Fuzzy Sets and Machine Discovery*, Tokyo, pp. 209–216.
[19] Li, C. C., and Lee, Y. C., 1989, "Computational Aspects of Dimensional Analysis," Int. Commun. Heat Mass Transfer, **16**(2), pp. 315–321.
[20] Newcomb, S., 1881, "Note on the Frequency of Use of the Different Digits in Natural Numbers," Am. J. Math., **4**, pp. 39–40.
[21] Benford, F., 1938, "The law of anomalous numbers," Proc. Am. Philos. Soc., **78**(4), pp. 551–572.
[22] Bazaraa, M. S., Sherali, H. D., and Shetty, C. M., 1993, *Nonlinear Programming, Theory and Algorithms*, 2nd Edition, Wiley, New York.
[23] Freedman, D., Pisani, R., and Purves, R., 1998, *Statistics*, 3rd Edition, W.W. Norton, New York.
[24] Schlichting, H., 1987, *Boundary-Layer Theory*, 7th Edition, McGraw-Hill Classic Textbook Reissue Series, McGraw-Hill, New York.
[25] Mendez, P. F., and Ordóñez, F., 2004, "SLAW, a Package for Scaling LAWs From Statistical Data," http://illposed.usc.edu/~pat/SLAW
[26] Blackwell, B. E., 1996, *A Framework for Determining the Mechanical Properties of Dissimilar Material Joints*, Doctor of Philosophy, Massachusetts Institute of Technology, Cambridge, MA.
[27] Park, J.-W., Mendez, P. F., and Eagar, T. W., 2002, "Strain Energy Distribution in Ceramic to Metal Joints," Acta Mater., **50**, pp. 883–899.
[28] American Society for Testing and Materials, 2002, *ASTM F 2183-02. Standard Test Method for Small Punch Testing of Ultra-High Molecular Weight Polyethylene Used in Surgical Implants*, W. Conshohocken, PA.
[29] Bergström, J., and Kurtz, S., 2003, "Extraction of Strength Properties of Bone Cement From Uniaxial Tension, Uniaxial Compression, and Small Punch Data," Technical Report MT00046.0MD/C0F2/0103/0001, Exponent, January 31.