490. Haldar, A., and Bažant, Z.P. (2009). "Recent Advances in Statistical Analysis of Concrete and Concrete Structures with Implications for Design," *Int. J. of Engrg. under Uncertainty: Hazards, Assessment, and Mitigation* 1, (1-2),81-89.

490.pdf

Recent Advances in Statistical Analysis of Concrete and Concrete Structures with Implications for Design

Achintya Haldar
Professor & da Vinci Fellow, Dept. of Civil Engi. & Engi. Mechanics, University of Arizona, haldar@u.arizona.edu

Zdeněk P. Bažant
McCormick Institute Professor and W.P. Murphy Professor of Civil Engineering and Materials Science Northwestern University, z-bazant@northwestern.edu

Abstract: *Statistical analyses of experimental database play an important role in developing design guidelines. These statistical models satisfy the fundamental theoretical requirements in different levels of sophistication. However, they can often lead to very different conclusions. A model that was rated as superior according to one statistical indicator could be rated as inferior according to another. There are, of course, many debatable points, but they concern only details such as the sampling, weighting, relevance and admissibility of data, rather than the statistical indicator per se. Three procedures are presented in the paper, specifically for the design of concrete structures, to showcase ways the uncertainty analysis procedures can be improved.*

Keywords: *creep, penetration depth in concrete, shear strength of concrete, shrinkage, statistical analysis.*

1. Introduction

The state of the art in uncertainty analysis and risk estimation has improved significantly in the recent years and it is appropriate to evaluate its applicability to the current engineering design. The uncertainty associated with the major design variables has been quantified or established. Various methods with different levels of sophistications or complexities are now available to track the progression of uncertainties from the input element level to the system level. For example, the capacity or resistance of an engineering system is expected to be a function of many uncertain variables with different statistical characteristics (such as the distributions, parameters, correlation characteristics, etc.). The demand or load effect on the system also contains numerous sources of uncertainty. The reliability methods, with various levels of sophistication, are now available to estimate the underlying risk, at least the notional risk. Simulation-based methods are also available for this purpose.

The risk is generally estimated at the element level, as opposed to the system level, and is evaluated with respect to the performance, limit state or critical state function. Sometimes, it is difficult to define the limit states for dynamic problems, including the propagation of cracks. For element level reliability evaluation, all the variables in a limit or critical state function are first transformed to uncorrelated variables, the non-normal variables are transformed to equivalent normal variables at the checking or design points, and then the associated risk is estimated. Some of these steps also introduce additional sources of uncertainty. Whenever possible, all the major design guidelines have been modified or are now in the process of being modified to satisfy an acceptable risk level at the element level, although it is well known that the estimated risk cannot be considered as the absolute risk. Since the loads and the uncertainty in them are beyond the control of engineers, the loads and the load factors are now considered to be the same for almost all designs, and the resistance factors are being modified or adjusted to satisfy the reliability requirements. The next generation design codes are now being developed to satisfy the performance based design (PBD) guidelines. In this approach, for example for the seismic risk estimation, instead of considering seismic load using a return period concept, the risk constraint is enforced on more appropriate performance requirements so that the risk could be different under different performance requirements.

In many cases, it is observed that the estimated risk may not be similar to the risk obtained from the observed data. A super-structure consisting of steel or concrete elements designed for an annual risk of say 10^{-6} or 10^{-7} cannot be integrated with sub-structure (foundation or geotechnical elements) of the same system where the similar risk is expected to be in the order of, say, 10^{-2} to 10^{-4}. The observation clearly indicates that the geotechnical aspects will be the weakest link, which may not always be correct. It is a complex and difficult subject. It needs further attention from the profession.

Some of the ways to improve the uncertainty analysis procedures are presented in this paper. They concern the following: (1) how to treat observational data with specific objectives, (2) how to treat observational data obtained from different sources with different objectives, and (3) how to treat time dependent observational data obtained from many sources with different objectives. These techniques were developed by the authors and are expected to showcase improved uncertainty analysis procedures that can be used in the future.

2. Different Uncertainty Analyses Procedures

In the following sections, the uncertainty analyses of concrete structures with design implications are briefly presented.

2.1 Penetration Depth in Concrete Structures-Data Obtained with Specific Objectives

Nuclear power plants must be protected against flying objects, commonly known as the missiles. Local effects of solid or nondeformable missiles on concrete structures are of interest. The solid missiles may cause perforation, penetration, and scabbing or spallation, or both, depending on many factors such as the strength of the concrete structures, impact energy and the impact area of the missile, time history of impact, etc. The impact mechanism is very complicated and is not completely understood mathematically. An empirical or approximate solution is necessary, preferably based on the available test results. For the lack of space, only estimation of the penetration depth is presented here.

The National Defense Research Council (NDRC) (Kennedy, 1976) equations can be used to estimate penetration depth in concrete structures. They are:

$$\frac{x}{d_m} = \left[\frac{4KNW}{d_m}\left(\frac{V}{1{,}000d_m}\right)^{1.8}\right]^{1/2} \; ; \quad \text{when} \quad \frac{x}{d_m} \leq 2.0 \qquad (1)$$

$$\text{and} \; \frac{x}{d_m} = 1 + \left[\frac{KNW}{d_m}\left(\frac{V}{1{,}000d_m}\right)^{1.8}\right] \quad \text{when} \; \frac{x}{d_m} \rangle \, 2.0 \qquad (2)$$

where x is the total penetration depth in inches, d_m is the missile diameter in inches, W is the missile weight in pounds, $K = 180/\sqrt{f_c'}$, f_c' is the ultimate concrete strength in pounds per square inch, V is the striking velocity of the missile in feet per second, N is the missile shape factor which takes the values 0.72 for flat nosed bodies, 0.84 for blunt nosed bodies, 1.00 for spherical end bullets, and 1.14 for a very sharp nose.

Equations 1 and 2 are based on a theory of penetration and can be considered as semi-empirical. A nondeformable cylindrical missile penetrating a massive reinforced concrete target would be the ideal condition for the NDRC equations. Essentially, the formulation neglects the rear boundary effects. The compressive strength of concrete, the size of the aggregate, and the amount of reinforcement are not included in the formulation. The NDRC equations were developed using information on penetration depth for small diameter, light weight, and high impact velocity projectiles.

After an extensive literature survey, Haldar and Hamieh (1984) identified a total of 625 cases of pure penetration where all the parameters in Eqs. 1 and 2 were reported. Out of these 625 cases, 35 are for solid missiles and 590 are for bullets. Also, 161 and 464 cases will satisfy the requirements for Eqs. 1 and 2, respectively. The adequacy of the NDRC equations can be studied with the help of the test data in two ways, as discussed next.

2.1.1 Statistical NDRC Equations

Assuming the functional forms of the different parameters in the NDRC equations are correct, the following regression equation can be used to improve the NDRC equations:

$$E(A|B = b) = C_1 + C_2\, b \qquad (3)$$

where A is represented by x_0/d, x_0 is the observed penetration depth, d is the diameter of the missile, B is the right hand side of Eq. 1 when $x/d_m \leq 2.0$ or the second term in the right hand side of Eq. 2 when $x/d_m \rangle\, 2.0$, C_1 and C_2 are unknown regressions coefficients need to be estimated from the available test data. For Eq. 1, the values of C_1 and C_2 are expected to be 0 and 1.0, respectively. For Eq. 2, they are expected to be 1.0 and

1.0, respectively. When the regression analysis was carried out, they were found to be 0.189 and 0.80, respectively for Eq. 1 and 1.17 and 1.03, respectively for Eq. 2 with the corresponding coefficient of determination (R^2 values) (Haldar and Mahadevan, 2000) of 0.7 and 0.9, respectively. The corresponding residual mean square errors i.e., the residual variation left unexplained by the regression model, are 0.049 and 0.893, respectively. The regression coefficients are different than expected indicating that they may be deficient in predicting the penetration depth.

2.1.2 Proposed Penetration Equations

The functional forms of the NDRC equations might not be ideal. The left-hand sides of these equations are dimensionless but the right-hand side is not; some parameters may receive undue importance and some may not receive enough. A dimensionless impact factor parameter I can be introduced as:

$$I = \frac{W}{g}\frac{N}{d_m^3}\frac{V^2}{f_c'} \tag{4}$$

where g is the acceleration due to gravity and all other parameters were defined earlier. To be consistent with units, Eq. 4 can be rewritten as:

$$I = \frac{12}{32.2}\frac{NWV^2}{d_m^3 f_c'} \tag{5}$$

The primary objective at this stage is to find a functional relationship between the penetration depth and the impact factor. When the ratio x_0/d is plotted against I in the form of a scatter diagram, piecewise linear curves were observed for three ranges of I values. When the linear regression analyses were carried out for the three different ranges, the following relationships are observed:

$$\frac{x}{d_m} = -0.0308 + 0.2251I; \quad 0.3 \le I \le 4.0 \tag{6}$$

$$\frac{x}{d_m} = 0.6740 + 0.0567I; \quad 4.0 \langle I \le 21.0 \text{, and} \tag{7}$$

$$\frac{x}{d_m} = 1.1875 + 0.0299I; \quad 21.0 \langle I \le 455.0 \tag{8}$$

Out of 625 cases, 9, 94, and 522 cases belong to Eqs. 6, 7, and 8, respectively. The R^2 values of these equations are 0.95, 0.70, and 0.90, respectively.

The predictability of the NDRC, statistical NDRC, and impact factor-based new equations can be checked by comparing the predicted penetration depths and the corresponding observed penetration depths. Three factors F_1, F_2, and F_3 are introduced such that

$$F_1 = \frac{x_{NDRC}(Eqs. 1, 2)}{x_{0\,observed}} \tag{9}$$

$$F_2 = \frac{x_{Statistical\,NDRC}(Eq. 3)}{x_{0\,observed}} \tag{10}$$

$$F_3 = \frac{x_{Proposed}(Eqs. 6, 7, 8)}{x_{0\,observed}} \tag{11}$$

For bullets, the mean values of F_1, F_2, and F_3 are found to be 1.00, 1.016, and 1.047, respectively, and the corresponding coefficients of variation (CoVs) are 0.436, 0.441, and 0.462. For large missiles, the mean values of the three factors are 1.733, 1.99, and 1.039, respectively, and the corresponding CoVs are 1.140, 1.350, and 0.677. The mean value of F_3 is closer to 1.0 for both bullets and large missiles, particular of interest to the nuclear industry. The corresponding CoV values are also smaller.

This exercise clearly indicates that by treating the parameters differently but essentially using the same statistical techniques, the predictability of the equations can be improved significantly.

2.2 Size Effect in Concrete Shear Strength Evaluation - From the Experimental Database

Bažant and Yu (2005, 2007) observed that the fracture of concrete, an archetypical quasi-brittle material, typically exhibits a rather large fracture process zone (FPZ), typically 0.5 m long. This causes that small structures (having cross sections less than a fully developed FPZ) fail in a quasi-ductile manner with a plastic yield plateau and exhibit almost no size effect, while very large structures (having cross sections much larger than the FPZ length) fail in concrete rather than steel behave in an almost perfectly brittle manner. The scatter band in the ACI-445F database with 398 data points has a downward trend as d, the effective depth of a beam, increases. Similar observations were made by Bažant and Kim (1984) and Bažant and Sun (1987). The database shows a decrease of scatter band width as the size increases, but is obtained from specimens with different geometries in which parameters other than size are varied significantly. Thus, the entire database cannot be used for statistical evaluation.

Bažant and Yu (2007) commented that sound arguments for a realistic design formula capturing the size effect on shear strength of beams must be based on fracture mechanics, verified by properly designed experiments, and statistically calibrated by a broad database. They added that, because of different objectives of various original experiments, it is not surprising that the existing ACI database has major gaps and a strong subjective statistical bias caused by crowding of the test data in the small-size range, scant data in the large size range, and no data at all for the largest sizes of practical interest (depths >2 m). Consequently, simple bivariate statistical regression of all the points of the ACI-445F database yields a misleading trend (Bažant and Yu, 2005). Eliminating the bias is important for a realistic update of the code provisions for the design of concrete beams.

The size effect can be defined as the size dependence of the nominal strength of structure when geometrical similarity is maintained and all the parameters other than the size are kept constant. In the case of beam shear, the size may be measured by the beam depth d, the nominal strength of structure may be taken as the average concrete shear strength in the cross section, v_c, and the parameters that must be kept constant comprise all the concrete properties including the maximum aggregate size d_a, the longitudinal reinforcement ratio ρ_w, and the shear span ratio a/d (here a is the distance of the load from the support). So the question is how to minimize the statistical bias in regard to the size effect. From the size effect viewpoint, this database has a bias of two kinds:

Kind 1 - Crowding of the data in the small size range —86% of the 398 data points pertain to three-point-loaded beams of depths less than 0.5 m, and 99% to depths less than 1.1 m, while only 1% of data pertain to depths from 1.2 to 2 m.

Kind 2 - Strongly dissimilar means and distributions among different size intervals of the subsidiary influencing parameters, particularly ρ_w, a/d, and d_a.

To reach any meaningful statistical conclusion on the size effect, both kinds of bias must be filtered out. Instead of the standard multivariate least-square nonlinear regression analysis in which all the parameters are optimized simultaneously, another approach can be tried which makes the statistical trend conspicuous without any mathematics.

The range of beam depths d of the existing test data can be subdivided into 5 size intervals [vertical strips in Fig. 1(a-c)]. They range from 0.075 to 0.15 m (3 in to 6 in), 0.15 to 0.3 m (6 in to 12 in), 0.3 to 0.6 m (12 in to 24 in), 0.6 to 1.2 m (24 in to 48 in), and 1.2 to 2.4 m (48 to 96 in). In the ACI database, these intervals contain 26, 251, 80, 38, and 3 data points, respectively. To filter out the effect of influencing parameters other than d, each interval of d must include only the data within a certain restricted range of ρ_w-values such that the average would be almost the same for each interval of d. Similarly, the range of a/d and d_a in each interval must be restricted so that their averages would also be about the same for each interval of d. The filtering of data must be done in an objective manner without any human preference. With the help of a computer optimization algorithm, the data points in each interval are deleted until uniformity of each subsidiary influencing parameter is optimally approached. Fig. 1(a-c) shows the restricted (filtered) data points by bigger circles, and those filtered out by the tiny circles.

As seen in Fig. 1, there are only three test data in the size interval spanning 1.2 to 2.4 m. They have the longitudinal steel ratio of ρ_w of 0.14%, 0.28% and 0.74%. The extremely low ρ_w of the first two makes it impossible to find similar data in other intervals of d. For example, the minimum ρ_w is 0.91% within the first interval of d, and 0.46% within the third interval. Therefore, one must consider the size range from 0.075 to 1.2 m. Formulating a statistical optimization algorithm for database filtering, one finds 7, 68, 17, and 36 data points within the admissible ranges for each interval of d (ideally, of course, the number of data in each interval should be the same, and the fact that it is not shows that complete elimination of statistical bias is impossible; nevertheless, for obtaining reliable means, 7 data certainly suffice).

After filtering, the mean values of ρ_w for the restricted ranges are 1.51%, 1.5%, 1.5%, and 1.5%, the mean values of a/d are 3.45, 3.33, 3.33 and 3.23, respectively, and the mean values of d_a are 16.8, 17.0, 16.8 and 16.5 mm. This provides data samples with minimum bias in terms of ρ_w, a/d and d_a. An important point to note is that, for different averages ρ_w, a/d and d_a, the trend of the interval centroids is the same, and closely matches the size effect law. This demonstrates objectivity of the data filtering approach.

Kinds 1 and 2 of bias afflict not only the mean trend of the full database, but also its scatter. The scatter may be measured by an unbiased CoV of the errors of the optimum fit curve compared to the individual data points.

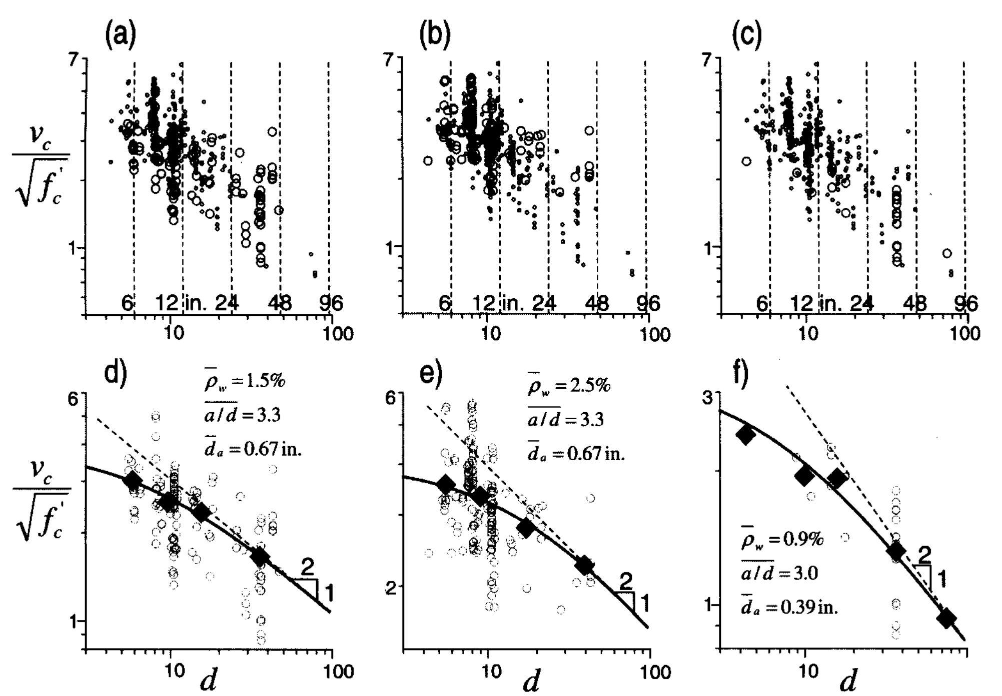

Figure 1: ACI-445F Database Subdivided into Intervals of Equal Size Ratio; (a-c) Full Database (the data retained are shown by larger circles and those filtered out in various cases by tiny circles); (d-f) Filtered Restricted Data Giving the Indicated Combinations of Uniform Mean Values of Subsidiary Parameters, Their Centroids and Regression Curves

This error must be considered for safe design and the following method can be used.

A simple bivariate nonlinear regression of our filtered restricted database, in which the kind 2 bias is already suppressed, can be used for this purpose. To suppress the kind 1 bias, one needs to give the same weight to the data in each size interval i, regardless of the number m_i of the points that fall into that interval. This may be achieved by assigning to the data in each interval i the normalized weight $w_i = (1/m_i)/\sum_k (1/m_k)$. Nonlinear regression of the weighted data yields the CoV of 22.3% and 23.6% for the average ρ_w of 1.5%, and 2.5%, respectively, as shown in Fig. 2.

The effect of data weighting can further be clarified by Fig. 2(a, b). As one can see, almost undistinguishable curves (dashed ones) are obtained by the weighted nonlinear bivariate statistical regression of the filtered database. An unweighted regression of the same data points is shown by the dash-dot curves. The dash-dot curve is again hardly distinguishable from the regression curve of the centroids in Fig. 2(a), but is very different in Fig. 2(b). One reason for this difference is that the vertical ranges of the restricted data in the individual size intervals, marked by vertical bars, are in Fig. 2(a) nearly symmetric with respect to the centroid curve, but not in Fig. 2(b). Another reason is that the restricted database in Fig. 2(a) is roughly homoscedastic, while that in Fig. 2(b) is not.

2.3 Time Dependent Observational Data Obtained from Many Sources–Creep and Shrinkage

Creep and shrinkage have been a pervasive cause of damage and excessive deflections in structures, and long-time creep buckling has caused a few collapses. Although there exist certain fundamental theoretical requirements

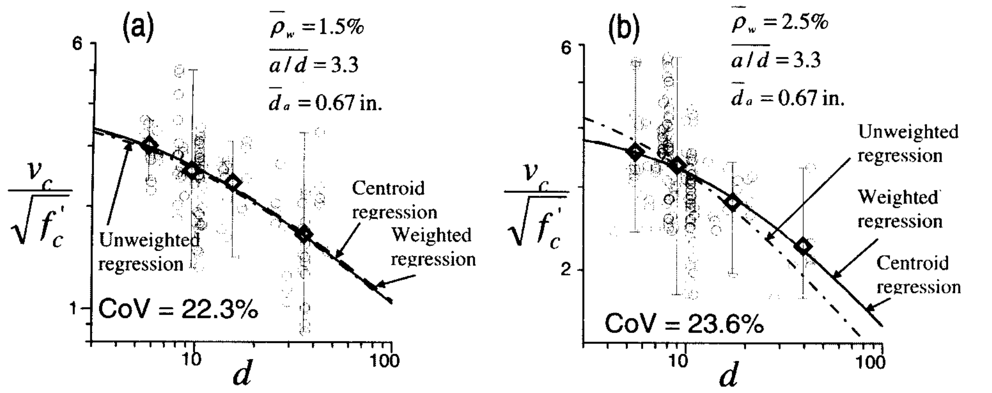

Figure 2: Regression Curves Corresponding to Weighted Fitting (Dashed Curves), Unweighted Fitting (Dash-dot Curves) and Fitting on Centroids (Solid Curves) for Filtered Database of (a) Average Steel Ratio = 1.5%; and (b) Average Steel Ratio = 2.5%.

(Ba•ant, 2000) which are essential for choosing the right model, necessitate rejecting some models even before their comparison to test data, most engineers place emphasis on statistical comparisons with the existing experimental database. Recently, a significantly enlarged database, named NU-ITI database (Ba•ant and Li, 2008) consisting of 621 creep tests and 490 shrinkage tests, has been assembled in the Infrastructure Technology Institute of Northwestern University by adding recent Japanese and Czech data.

Altering the statistical methods can often lead to very different conclusions. One such instance, where introduction of various nonstandard statistical indicators have recently sown much confusion, is the use of creep and shrinkage databases to evaluate various prediction models (Ba•ant and Li, 2007). A model that was rated as superior according to one statistical indicator was rated as inferior according to another. There are, of course, many debatable points, but they concern only details such as the sampling, weighting, relevance and admissibility of data, rather than the statistical indicator per se.

Among concrete researchers, a popular way to verify and calibrate a model has been to plot the measured values y_k $(k = 1, 2, \ldots n)$ from an experimental database against the corresponding model predictions Y_k, or to plot the errors (or residuals) $\varepsilon_k = y_k - Y_k$ versus time. If the models were perfect and the tests were scatter-free, the former plot would give a straight line of slope 1, and the latter a horizontal line of ordinate 0. If one makes such plots for some of the available models using the NU-ITI database, one finds very little difference among the models, even those which are known to give very different long-time predictions (Ba•ant and Li, 2007). The same is true for another popular comparison where the ratio $r_k = y_k/Y_k$ is plotted versus time, for which, if the model were perfect and the tests scatter-free, one would get a horizontal plot $r_k = 1$. Such comparisons are ineffectual for the following four reasons:

(1) The statistical trends are not reflected in such plots.

(2) The statistics are dominated by the data for short times $t - t'$, low ages t' at loading and small specimen sizes D, while predictions for long times are of main interest for practice. This is due to highly nonuniform data distributions.

(3) Because of their longer test durations and high creep and shrinkage, the statistics are also dominated by the data for old low-strength concretes not in use any more. But long-duration tests of modern high strength concretes, which creep little, are still rare.

(4) The variability of concrete composition and other parameters in the database causes enormous scatter masking the scatter of creep and shrinkage evolution.

If the time, age and specimen size are transformed to variables that make the trends uniform and the data set almost homoscedastic, i.e., have an approximately uniform conditional variance about the regression line (Ang and Tang, 1984), and if these variables are subdivided into intervals of equal importance, the number of tests and the number of data points within each interval should ideally be about the same. However, this is far from true for every existing database. From the histograms of the available data, Bažant and Li (2007) observed that the distribution in the database is highly nonuniform. This bias must be counteracted by proper weighing of the data. One may first subdivide the load duration $t - t'$, age at loading t', effective specimen thickness D (Bažant and Baweja, 2000) and environmental humidity H into intervals of roughly equal importance, which ought to have approximately the same weight in the statistical evaluation. This is achieved by subdividing $\log(t - t')$ and $\log(t - t_0)$ into equal intervals in the logarithmic scale, which means that the intervals of $t - t'$ and $t - t_0$ form a geometric progression (t is the time, representing the current age of concrete, t_0 is the age at the start of drying, and $t - t_0$ is the duration of shrinkage test; all the times are given in days).

There are four independent variables which need to be subdivided into intervals of equal statistical weight: $t - t'$, t', D, and H for creep, and $t - t_0$, t_0, D, and H for shrinkage. Ideally, all these subdivisions should be introduced simultaneously, which creates four-dimensional boxes (or hypercubes). Since boxes of lesser dimensions have a lesser chance of containing insufficient number of data points (0, 1 or 2), two-dimensional boxes of $\log(t - t')$ and H for creep, and $\log(t - t_0)$ and $\sqrt{D}$ for shrinkage appear to be preferable over three- or four-dimensional boxes. One-dimensional boxes, or intervals, of load or drying durations are even more advantageous in this respect, since the existing database has many points in every such interval. Differences in weights might also be considered for data sets obtained on different concretes and in different laboratories. Another debatable point is whether the boxes for long creep or shrinkage durations should not actually receive a greater weight than those for short durations. Maybe they should, since accuracy of long-time prediction is of the greatest interest. Again, we choose not to introduce such additional weights because the appropriate differences in their values would be hard to assess and would anyway be much less than an order of magnitude, being dwarfed by differences in weights w_i compensating for differences in the number of data points in different boxes.

The anti-high-strength bias also needs to be reduced at this stage. The tests of old types of concretes with high water-cement ratios, lacking modern admixtures, dominate the database. Of little relevance though such concretes are today, these tests cannot be ignored because they supply most of the information on very long creep and shrinkage durations.

Besides, these tests are not completely irrelevant for our purpose because the time curves for low and high strength concretes are known to have similar shapes. This is not surprising since, in both, the sole cause of creep is the calcium silicate hydrate, or C-S-H. The difference resides merely in the scaling of creep and shrinkage magnitudes. This scaling depends strongly on the water-cement ratio and admixtures, in a way that is not yet predictable mathematically. Therefore, the data for old kinds of concrete must be used, but their bias must be counteracted. Since the overall magnitude of creep and shrinkage strains is roughly proportional to the elastic compliance, and since this compliance is roughly proportional to $1/\sqrt{f_c'}$, where f_c' is the cylindrical compressive strength, the bias can be reduced by replacing all y data by $y\sqrt{f_c^0 / f_c'}$ i.e., by scaling all the measured compliances and shrinkage strains y in inverse proportion to $\sqrt{f_c'}$; here $f_c^0 = 5000$ psi (34.5 MPa), a constant factor introduced to retain convenient dimensions.

Based on the subdivision into boxes of equal weight, the standard error s of the prediction model (representing the standard error of regression) is defined as follows (Ang and Tang 1984, Haldar and Mahadevan, 2000):

$$s = \sqrt{\frac{N}{N-p}\sum_{i=1}^{n} w_i \sum_{j=1}^{m_i}\left(Y_{ij} - y_{ij}\right)^2} \qquad (12)$$

where m_i and w_i are the number of data points in box number i and the statistical weight assigned to the points in this box; $N = N\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} m_i$ is the number of all the data points in the database; y_{ij} is the measured creep or shrinkage data; Y_{ij} is the corresponding model predictions, and $y_{ij} - Y_{ij} = \varepsilon_{ij}$ is the error of the prediction, and p = number of free parameters in the model. Let the intervals or boxes of data be labeled by one index, $i = 1, 2, \ldots n$, running consecutively through all the data sets in the database. To counteract the human bias, every box of every data set must have the same weight. This is achieved by considering the statistical weights w_i of the individual data points in each box to be inversely proportional to the number m_i of data points in that box. Normalizing the weights so that $\sum_{i=1}^{n} w_i = 1$, one obtains:

$$w_i = \frac{1}{m_i \overline{w}}, \qquad \overline{w} = \sum_{i=1}^{n} \frac{1}{m_i} \tag{13}$$

To compare various models, one must use dimensionless statistical indicators of scatter. In regression statistics, CoV of regression errors δ, which characterizes the ratio of the scatter band width to the mean, is commonly used, and it can be defined as:

$$\delta = \frac{s}{\overline{y}}, \qquad \overline{y} = \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} y_{ij} \tag{14}$$

where $\overline{y}$ represents the weighted mean of all the measured values y_{ij} in the database.

While δ characterizes the ratio of the scatter band width to the data mean, the coefficient of determination, R^2, a ratio of the variability accounted for by the regression model and the total variability in the response variable indicates the adequacy of the regression relationship. It can be defined as:

$$R^2 = 1 - \frac{s^2}{\overline{s}^2}, \quad s^2 = \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} \left(y_{ij} - Y_{ij}\right)^2, \quad \overline{s}^2 = \sum_{i}^{n} w_i \sum_{j=1}^{m_i} \left(y_{ij} - \overline{y}\right)^2 \tag{15}$$

where s^2 is the residual variation left unexplained by the regression model and is the total variability in the response variable or the overall weighted standard deviation of all the data. Bažant and Li (2007) analyzed five predictive models. They are: (1) B3 (Bažant and Baweja, 2000), (2) ACI (1972), (3) CEB (Muller and Hilsdrof, 1990), (4) GL (Gardner and Lockman, 2001), and (5) GZ (Gardner and Zhao, 1993). Table 1 presents comparisons of the CoV and correlation coefficients of these prediction models, based on using different types of data boxes -one-, two- and three-dimensional. In all these comparisons, model B3 is found to be the best, except for one case where it is one of two equal best. GL model comes out as the second best. Considerably worse but the third best overall is seen to be the CEB model. Since the current ACI-209 model, labeled ACI, is the oldest, introduced in 1972 on the basis of 1960's research, it is not surprising that it comes out as the worst.

Table 1
Standard CoV of Errors of Various Prediction Models

	B3	*ACI*	*CEB*	*GL*	*GZ*
			(a) Compliance (%)		
200 cubes	28.3	38.8	30.6	28.5	39.5
5 intervals, log (t-t')	26.2	41.9	29.7	28.5	43.8
4 intervals, log t'	27.4	37.1	29.9	28.8	48.2
7 intervals, $\sqrt{D}$	23.3	36.9	27.3	23.3	33.2
10 intervals, H	24.4	44.2	29.0	30.7	44.6
			(b) Relative compliance (%)		
200 cubes	24.4	59.0	29.3	27.3	35.7
5 intervals, log (t-t')	26.4	66.0	33.0	29.8	32.9
4 intervals, log t'	26.9	74.3	33.3	30.5	33.0
7 intervals, $\sqrt{D}$	20.1	55.9	24.4	21.9	22.6
10 intervals, H	21.0	52.6	28.0	25.4	28.6
			(c) Shrinkage (%)		
112 cubes	37.4	44.4	48.1	43.3	50.0
4 intervals, log (t-t_0)	29.4	40.8	48.0	37.7	49.3
4 intervals, log t_0	42.8	48.6	56.0	53.9	64.2
7 intervals, $\sqrt{D}$	27.2	37.3	49.2	29.1	38.9
10 intervals, H	38.4	52.0	46.9	54.4	46.6
			(d) Relative shrinkage (%)		
112 cubes	41.8	51.8	47.9	48.3	58.1
4 intervals, log (t-t_0)	34.5	49.5	46.0	43.3	54.7
4 intervals, log t_0	44.9	52.8	57.6	54.0	64.7
7 intervals, $\sqrt{D}$	33.7	46.4	45.0	39.9	52.9
10 intervals, H	41.6	55.6	43.0	41.9	45.6

3. Conclusions

Statistical techniques are routinely used to develop design guidelines. In many cases, competing models are suggested. When the predictability of these models is compared with the available experimental database, they may all appear to be the same in the overall sense. However, if they are not considered appropriately, they may produce unacceptable results with severe design implications. Three specific cases related to the concrete structures are presented here to document how the predictability of design equations can be improved significantly.

References

ACI Committee 209. "Prediction of Creep, Shrinkage, and Temperature Effects in Concrete Structure," ACI-SP-194, 1972, Reapproved in 1982 (ACI-SP-76) and 1992 (ACI-209 T92).

Ang, A. H. S., and Tang, W. H., *Probability Concepts in Engineering Planning and Design, Vol II. Decision, Risk and Reliability*, (1984), John Wiley, New York.

Bažant, Z. P., "Criteria for Rational Prediction of Creep and Shrinkage of Concrete," *Adam Neville Symposium: Creep and Shrinkage-Structural Design Effects*, ACI SP-194, (2000), A. Al-Manaseer, ed., 237-260.

Bažant, Z. P., and Baweja, S., "Creep and Shrinkage Prediction Model for Analysis and Design of Concrete Structures: Model B3." *Adam Neville Symposium: Creep and Shrinkage-Structural Design Effects*, ACI SP-194, (2000), A. Al-Manaseer, ed., 1-83.

Bažant, Z. P., and Kim, J. K., "Size Effect in Shear Failure of Longitudinally Reinforced Beams," *Am. Concrete Institute Journal*, (1984). 81, 456-468, Disc. and Closure (1985), 82, 579-583.

Bažant, Z. P., and Li, G. H., *Unbiased Statistical Comparison of Creep and Shrinkage Prediction Models*, Structural Engineering Report No. 07-12/A210u, Department of Civil and Environmental Engineering, Northwestern University, (2007).

Bažant, Z. P., and Li, G. H., *Database on Concrete Creep and Shrinkage*, Infrastructure Technology Institute (ITI), Northwestern University Report, (in preparation, to be posted on ITI website), (2008).

Bažant, Z. P., and Sun, H. H., "Size Effect in Diagonal Shear Failure: Influence of Aggregate Size and Stirrups", *ACI Materials Journal*, (1987), 84 (4), 259-272.

Bažant, Z. P., and Yu, Q., "Consequences of Ignoring or Mis-Judging the Size Effect in Concrete Design Codes and Practice," Proc., *TCI 2007 Concrete Technology Conference*, 1st Annual Meeting of Taiwan Concrete Institute (plenary opening lecture), (2007), Taipei, 1-23.

Bažant, Z. P., and Yu, Q., "Designing against Size Effect on Shear Strength of Reinforced Concrete Beams without Stirrups. I. Formulation", *J. of Structural Engineering ASCE*, (2005), 131 (12), 1877-1885.

Gardner N. J. and Lockman M. J., "Design Provisions for Drying Shrinkage and Creep of Normal Strength." *ACI Materials Journal*, (2001), 98 (2), 159-167.

Gardner, N. J. and Zhao, J. W. "Creep and Shrinkage Revisited," *ACI Materials Journal*, (1993), 90, 236-246. Discussion by Bažant and Baweja, *ACI Materials Journal*, (1994), 91, 204-216.

Haldar, A., and Hamieh, H. A., "Local Effect of Solid Missiles on Concrete Structures," *J. of Structural Engineering, ASCE*, (1984), 110(5), 948-960.

Haldar, A., and Mahadevan, S, *Probability, Reliability and Statistical Methods in Engineering Design*. J. Wiley, New York, (2000).

Kennedy, R. P., "A Review of Procedures for Analysis and Design of Concrete Structures to Resist Missile-Impact Effects," Nuclear Engineering and Design, (1976), 37, 183-203.

Muller, H. S., and Hilsdorf, H. K., Evaluation of the Time-dependent Behaviour of Concrete: Summary Report on the Work of the General Task Force Group No. 199, CEB, (1990).