

Unbiased Statistical Comparison of Creep and Shrinkage Prediction Models

by Zdeněk P. Bažant and Guang-Hua Li

This paper addresses the problem of selecting the most realistic creep and shrinkage prediction model, important for designing durable and safe concrete structures. Statistical methods of standard and several nonstandard types and a very large experimental database have recently been used to compare and rank the existing prediction models, but conflicting results have been obtained by various investigators. This paper attempts to overcome this confusion. It introduces data weighting required to eliminate the bias due to improper data sampling in the database, and then examines Bažant and Baweja's Model B3, the ACI model, the CEB model, and two of Gardner's models. The statistics of prediction errors are based strictly on the method of least squares, which is the standard and the only statistically correct method, dictated by the maximum likelihood criterion and the central limit theorem of the theory of probability, as well as the requirement of noncorrelation of errors. Several nonstandard statistical methods that have recently been invented to evaluate creep and shrinkage models are also examined and their deficiencies are pointed out. The ranking of the models that ensues from the least-square regression statistics is shown to be quite different from the rankings obtained by the nonstandard statistics.

Keywords: creep; design guide; least-square regression; prediction model; shrinkage; statistics.

INTRODUCTION

Altering the statistical method can often lead to very different conclusions. Aside from shear strength statistics,¹ one important statistical problem where inventions of various nonstandard statistical indicators²⁻⁷ have recently sowed much confusion in the statistical comparison of various prediction models for creep and shrinkage⁸⁻²² using large databases. A model that was rated as superior according to one statistical indicator was rated as inferior according to another. The slight differences in the ranges of strength, humidity, and cement type⁷ for which various models were calibrated cannot explain the differences in ranking.

Are all the statistical methods used in different creep and shrinkage studies justified? This paper, whose contents were summarized at a recent conference,²³ shows that most of them are not. In the case of creep and shrinkage, in which one deals with central-range statistics of errors (and not with the far-out distribution tail that matters for structural safety), it is actually very clear what is a rational statistical approach. It is the method of least squares—the standard method that (as shown by Gauss²⁴) maximizes the likelihood function and is consistent with the central limit theorem of the theory of probability (refer to the Appendix).^{25,26} There are, of course, many debatable points, but they concern only details such as the sampling, weighting, omission of outliers, and data relevance or admissibility, rather than the statistical method. This study will attempt to offer correct statistical

comparisons of the main prediction models for creep and shrinkage of concrete and explain why various nonstandard statistical indicators have led to dubious conclusions. The same five models as in Reference 7 will be statistically evaluated:

1. Model B3, 1995, which was approved as the international RILEM Recommendation²⁷ and slightly updated in 2000²¹ (this model is a refinement of the 1978 Model BP⁹ and of its improvement as Model BP-KX¹³);

2. ACI model,⁸ based on 1960s research, reapproved by ACI Committee 209 in 2008;

3. Model of Comité Européen du Béton, labeled CEB, which is based on the work of Müller and Hilsdorf³ (it was adopted in 1990 by CEB,¹² updated in 1999,²⁸ and co-opted in 2002 for Eurocode 2);

4. Gardner and Lockman's model, labeled GL²⁰; and

5. Gardner's earlier model, labeled GZ.¹⁹

Sakata's model,^{18,22} whose scope is somewhat limited, as well as the crude old models of Dischinger, Illston, Nielsen, Rüschi and Jungwirth, Maslov, Arutyunyan, Aleksandrovskii, Ulickii, Gvozdev, Prokopovich and others,²⁹⁻³¹ will not be considered.

There exist certain fundamental theoretical requirements³² that are essential for choosing the right creep and shrinkage model, necessitate rejecting some models even before their comparison to test data, and were taken as the basis of Model B3. Nevertheless, most engineers place emphasis on statistical validation using a large experimental database. Therefore, this paper will deal exclusively with statistics.

The first comprehensive database, comprising approximately 400 creep tests and approximately 300 shrinkage tests, was compiled at Northwestern University in 1978,⁹ mostly from American and European tests. In collaboration with CEB, begun at the 1980 Rüschi Workshop,³³ this database was slightly expanded by an ACI 209 subcommittee. A further slight expansion was undertaken by a RILEM subcommittee. It led to what became known as the RILEM database,^{3,34,35} which contained 518 creep tests and 426 shrinkage tests. Recently, a significantly enlarged database, named the NU-ITI Database³⁶ and consisting of 621 creep tests and 490 shrinkage tests, has been assembled in the Infrastructure Technology Institute of Northwestern University by adding many recent Japanese and Czech experimental results. A reduced database, consisting of 166 creep tests and 106 shrinkage tests extracted from the RILEM database, has recently been used in Gardner's statistical ranking.^{2,20,37}

ACI Materials Journal, V. 105, No. 6, November-December 2008.

MS No. M-2007-422 received December 28, 2007, and reviewed under Institute publication policies. Copyright © 2008, American Concrete Institute. All rights reserved, including the making of copies unless permission is obtained from the copyright proprietors. Pertinent discussion including authors' closure, if any, will be published in the September-October 2009 *ACI Materials Journal* if the discussion is received by June 1, 2009.

Zdeněk P. Bažant *FACI*, is the McCormick Institute Professor and W.P. Murphy Professor of Civil Engineering and Materials Science at Northwestern University, Evanston, IL. He is a Registered Structural Engineer in Illinois. He has received six honorary doctorates. He is a Past Chair and member of ACI Committee 446, Mechanics of Concrete, and a member of ACI Committees 209, Creep and Shrinkage in Concrete; 348, Structural Safety; and Joint ACI-ASCE Committees 334, Concrete Shell Design and Construction; 445, Shear and Torsion; and 447, Finite Element Analysis of Reinforced Concrete Structures. He was the founding Chair of ACI Committee 446, Fracture Mechanics.

Guang-Hua Li is a Graduate Research Assistant and Doctoral Candidate at Northwestern University. His research interests include inelastic and probabilistic mechanics.

Among concrete researchers, a popular way to verify and calibrate a creep and shrinkage model has been to plot the measured values y_k ($k = 1, 2, \dots, n$) 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 (Fig. 1).^{5,6,38} If the models were perfect and the tests scatter-free, the former plot would give a straight line of slope 1, and the latter a horizontal line of ordinate 0. Figure 1, using the NU-ITI Database, shows examples of such plots for some of the aforementioned models. One immediately notes that, in this kind of comparison, there is very little difference among the creep and shrinkage models, even those that are known to give very different long-time predictions. The same is true for another comparison, popular with concrete researchers, where the data-model ratio, $r_k = y_k/Y_k$, is plotted versus time. If the model were perfect and the tests scatter-free, then all the r_k values would lie on a horizontal line, $r_k = 1$ (for deficiencies of this kind of statistics, see comments on Eq. (13) and (14) that follow).

Why are the comparisons in Fig. 1 ineffectual for ranking models? There are four reasons:

1. The statistical trends are not reflected in such plots;
2. Owing to highly nonuniform data distributions (evident from the histograms in Fig. 2), the statistics are dominated by the data for short load durations $t - t'$, low ages t' at loading, and small specimen sizes D , while the main practical interest is in the long-time predictions;
3. Because of their longer test durations and high creep and shrinkage, the statistics are also dominated by the data for old types of concrete having low strength, not used any more. Long-duration test data for modern high-strength concretes, which creep and shrink little, are still quite rare (refer to Fig. 2); and
4. The variability of concrete composition and other parameters in the database causes enormous scatter, masking the much lower scatter in the time evolution of creep and shrinkage.

If the worldwide testing in the past could have been planned centrally so as to follow the proper statistical design of experiments, the chosen sampling of the relevant parameters and reading times of creep and shrinkage tests would have been completely different than those found in the databases. This research attempts to overcome these deficiencies.

If the time, age, and specimen size are transformed to variables that make the trends approximately uniform, 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 approximately the same. However, this is far from true for every existing database (refer to Fig. 2).

Nonetheless, there is no choice but to extract the best information possible from the imperfect database that exists. A quest to do that is what motivates this paper. Another

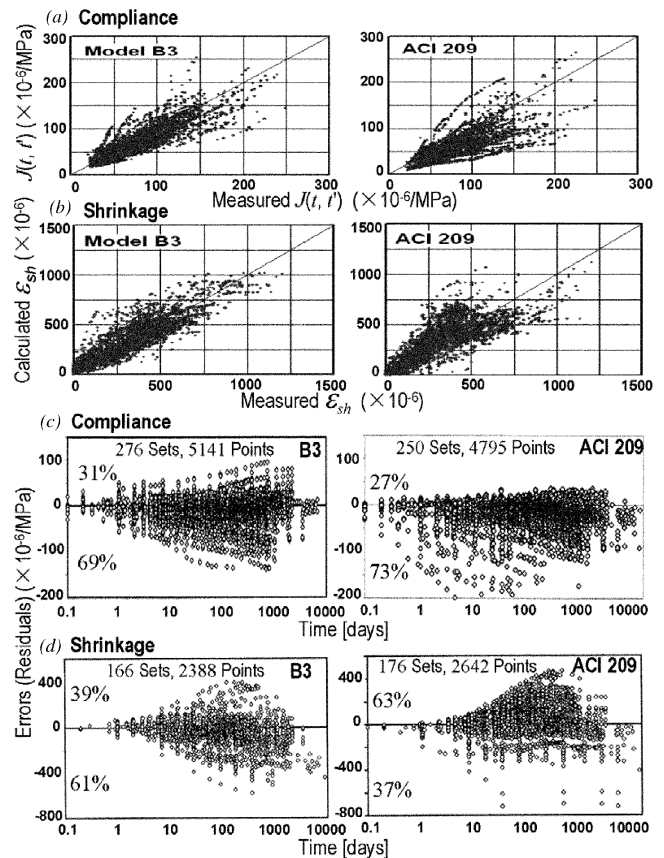


Fig. 1—Examples of ineffectual statistical comparisons in which all the prediction models for: (a) and (c) compliance; and (b) and (d) shrinkage look approximately equally good (or equally bad).

motivation is the need to compare the existing models using the correct statistical approach and to explain why some previous attempts at such comparisons^{2,4-7} were not objective.

RESEARCH SIGNIFICANCE

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. The deflections of many large-span prestressed concrete bridges have been far greater than predicted.³⁹ For instance, in the case of the Koror-Babeldaob Bridge in Palau, a prestressed box girder that had the world-record span of 241 m (790.68 ft) when built in 1977, the sag at midspan reached 1.52 m (5 ft) by 1996. An ill-fated attempt to remedy it by additional prestress and jacking led to collapse (with two fatalities). Inadequacy of the creep and shrinkage prediction model available at the time of design is certain to have been one of the causes of excessive deflections of this bridge,⁴⁰ as well as many others. To minimize the chances of repetition, the best among the available prediction models must be identified.

SUPPRESSING DATABASE BIAS DUE TO NONUNIFORM SAMPLING OF PARAMETER RANGES

From Fig. 2, showing the histograms of the available data, it is seen that their distribution in the database is highly nonuniform. This nonuniformity is not an objective property but a result of human choice, and thus leads to biased statistics of data fits.

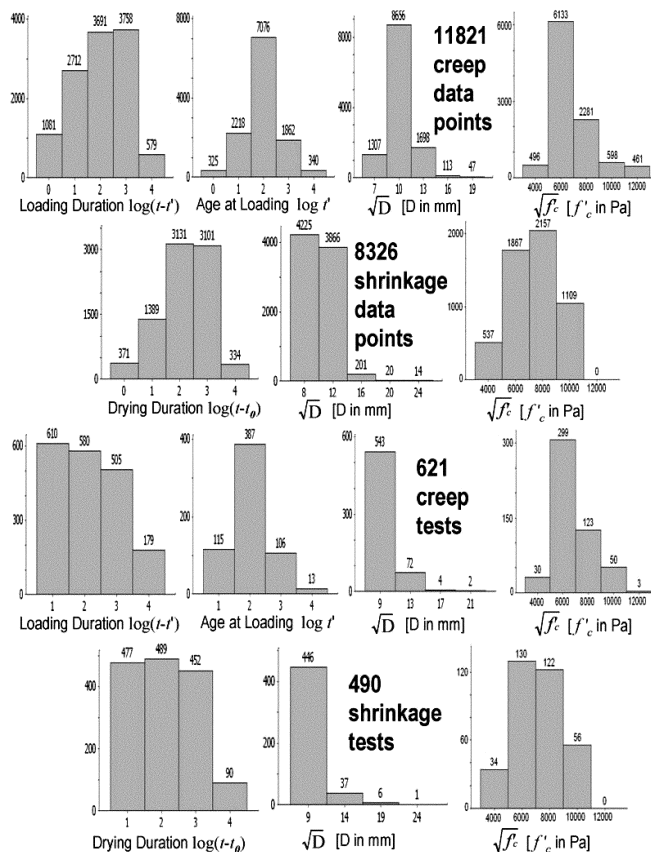


Fig. 2—Histograms of data points and of test curves in the NU-ITI database.

This bias must be counteracted by proper weighting of the data. To this end, one may first subdivide the load duration $t - t'$, age at loading t' , effective specimen thickness D and environmental humidity H into intervals of roughly equal importance that ought to have approximately the same weight in the statistical evaluation.²¹ This is achieved by subdividing $\log(t - t')$ and $\log(t - t_0)$ into equal intervals in the logarithmic scale (refer to Fig. 3(a)), which means that the subdivisions of $t - t'$ and $t - t_0$ form a geometric progression ($t =$ time = current age of concrete, $t' =$ age at loading, $t_0 =$ age at the start of drying, and $t - t_0 =$ shrinkage test duration—all in days).

The main reason for this kind of subdivision into intervals becomes clear upon noting that, because the creep and shrinkage are decaying processes, a time increment of, for example, 10 days, makes much difference when the test duration is 10 days but little difference when it is 1000 days. In other words, intervals forming an arithmetic progression cannot have equal importance. By contrast, extending the duration by, say, 20% is about equally important in both cases, and this corresponds to intervals of equal length, $\log 1.2$, in the logarithmic scale.

A similar argument can be made in regard to the effective thickness (or size) D of the cross section, defined as $D = 2V = S = 2 \cdot \text{volume/surface ratio of the specimen}$. Herein, the proper coordinate transformation, before equal intervals are introduced, is from D to \sqrt{D} . This transformation is indicated by the diffusion theory, which shows that the half-time of drying (or shrinkage) is proportional to D^2 .^{32,41,42} As for the environmental humidity H , no transformation seems necessary.

There are four independent variables that 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 would create four-dimensional boxes (or hypercubes). The use of four-dimensional, as well as three-dimensional, boxes, however, has another shortcoming: For the database that exists, it appears that the number of data points in some boxes is 0 or 1. Such boxes allow no statistics to be taken and, therefore, must be deleted. Even two points in a box is too low for meaningful statistics. Furthermore, deletion of some boxes makes the relative weights of the boxes and of the data sets unequal.

Because boxes of lesser dimensions have a lesser chance of containing only 0, 1, or 2 points, two-dimensional boxes of $\log(t - t')$ and H for creep, and $\log(t - t_0)$ and \sqrt{D} for shrinkage (refer to Fig. 3(b)) appear to be preferable over three- or four-dimensional boxes. One-dimensional boxes, or intervals (refer to Fig. 3(a)), of load or drying durations are even more advantageous in this respect because 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. Maybe they should, but this would be a judgment exposed to criticism. Besides, such differences in weights would certainly be much smaller than an order of magnitude. Introducing such weights would thus be unimportant in comparison to the weights w_i for the data boxes, which must differ by more than one order of magnitude to compensate for the huge differences in the number of data points in different boxes.

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 because accuracy of long-time prediction is of the greatest interest. Again, however, this study does not introduce such additional weights because the appropriate differences in their values would be hard to assess and would 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.

Another bias stems from the difference in numbers N_{ri} and N_{si} of the data readings taken by experimenters r and s within box i . If $N_{ri} \gg N_{si}$, then the statistics would be biased for experimenter r and against experimenter s . For two-dimensional data boxes, however, this bias is not very strong (because, mostly, $0.5 < N_{ri}/N_{si} < 2$), and thus is not considered herein.

CHOICE OF TRANSFORMATIONS OF RANDOM VARIABLES FOR REGRESSION

Data scaling for strength effect

The tests of old types of concretes with high water-cement ratios (w/c), 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 the purpose of this paper because the time curves for low- and high-strength concretes are known to have similar shapes. This is not surprising because, in both, the sole cause of creep is the calcium silicate hydrate (C-S-H). The difference resides merely in the scaling of creep and shrinkage magnitudes. This scaling depends strongly on the w/c and admixtures, in a way that is not yet predictable mathematically (which makes it an important problem for research). Therefore,

the data for old kinds of concrete must be used, but their bias must be counteracted.

Because the overall magnitude of creep and shrinkage strains is roughly proportional to the elastic compliance, and because this compliance is roughly proportional to $1/\sqrt{f'_c}$ where f'_c is the cylindrical compressive strength, one can reduce this bias by replacing the measured data y for the compliance and shrinkage by $y\sqrt{f'_c}/f'_c$, where f'_c is the constant factor that is necessary to get dimensionless ratios and may be chosen arbitrarily because it has no effect on model comparison (the authors chose $f'_c = 5000$ psi [34.5 MPa]). It is by virtue of this simple approximate property that f'_c need not be considered as a fifth independent variable in regression statistics.

The effect of the $1/\sqrt{f'_c}$ scaling of the data is shown, for shrinkage, in Fig. 4(a) and (b). It makes the data band noticeably narrower, but less so for compliance, which is not shown in the figure. Why is the effect of this scaling not more pronounced? Because it is masked by variation of other parameters. For this reason, statistical comparisons will be made both with and without strength scaling.

Relative creep and shrinkage growth with time

Experimental observations show that for concretes of different compositions, the relative increase of deformation with time differs much less than the total increase, provided that other influencing parameters are fixed. For statistical analysis, it thus makes sense to consider the relative compliance or relative shrinkage, defined as the compliance or shrinkage strain divided by its initial value

$$\bar{J}(t, t') = J(t, t')/J_0, \quad \bar{\varepsilon}(t, t_0) = \varepsilon(t, t_0)/\varepsilon_0 \quad (1)$$

where J_0 and ε_0 are the initial values of compliance and shrinkage, which are chosen as the compliance for 3 days of sustained load and the shrinkage for 28 days of drying (the shrinkage at 3 days of drying is too small to be useful, whereas the compliance for 1 day, or even 0.1 day, of sustained loading could serve almost equally well). When the relative compliance and relative shrinkage of the entire database are plotted as a function of load or drying duration, however, the reduction of the scatter band width of the database is disappointingly small. The reason is that taking the relative values suppresses only the effect of composition, not the effects of variation of the age at loading, environmental humidity, specimen size, and specimen shape throughout the database. Unlike a change in composition, these parameters affect the compliance and shrinkage at various times differently. Therefore, the statistics will be calculated for both relative and total values of compliance and shrinkage.

Logarithmic transformation of random data

The creep or shrinkage data plotted in terms of the load duration $t - t'$ or drying duration $t - t_0$ are generally found to be markedly heteroscedastic (that is, the conditional variance is not constant but varies with time). The regression statistics, however, works best when the data are homoscedastic⁴³ (that is, the conditional variance is almost uniform). To make the data homoscedastic, transformation of the variables is the standard approach. As is generally the case when the relative, rather than actual, changes of response matter, approximate homoscedasticity of compliance data happens to be achieved by taking the logarithm of the data $y = J$ ($\ln y$ is preferred over

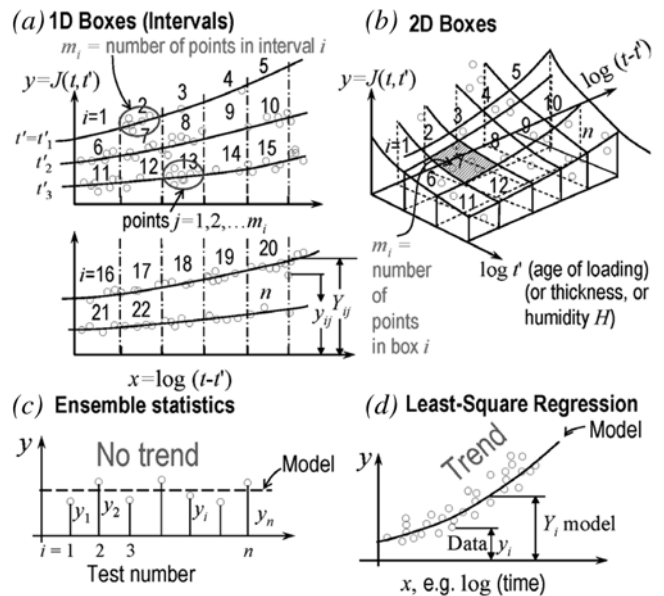


Fig. 3—Sketches explaining: (a) and (b) subdivision of database variables into one-dimensional intervals and two-dimensional boxes of equal importance; and (c) and (d) difference between ensemble (or population) statistics and regression statistics.

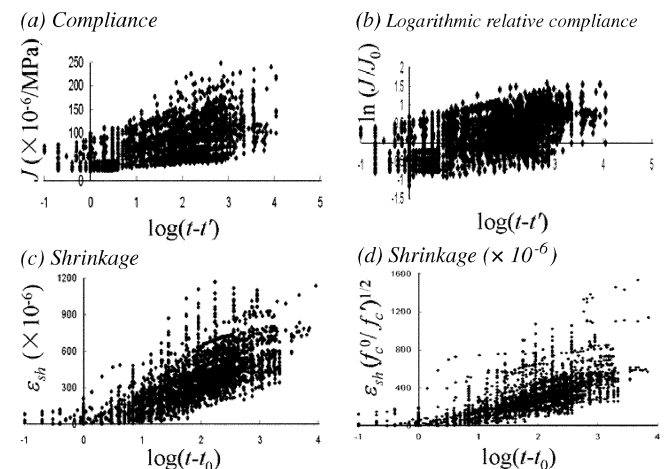


Fig. 4—(a) and (b) Effect of logarithmic ordinate transformation on compliance database; and (c) and (d) effect of scaling ordinate by $(\text{strength})^{-1/2}$ on shrinkage database.

$\log_{10} y$ because, for small errors, the standard deviation of $\ln y$ is equal to the coefficient of variation of y , as $d \ln Y_k = dY_k/Y_k$).

A comparison of Fig. 4(c) and (d) shows that the compliance data indeed become almost homoscedastic upon logarithmic transformation, which is better applied to the relative, rather than actual, compliance. The scatter band becomes very wide (compared with the scatter band rise over time). Therefore, statistical comparisons will be made for both J and $\ln(J/J_0)$.

STANDARD REGRESSION STATISTICS OF DATABASE

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^{25,44,45}

$$s = \sqrt{\frac{N}{N-p} \sum_{i=1}^n w_i \sum_{j=1}^m (Y_{ij} - y_{ij})^2} \quad (2)$$

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, respectively; $N = N \sum_{i=1}^n w_i = \sum_{i=1}^n m_i$ is the number of all the data points in the database; y_{ij} are the measured creep or shrinkage data of which the database is comprised; Y_{ij} are the corresponding model predictions; and $y_{ij} - Y_{ij} = \varepsilon_{ij}$ are the errors of the predictions.

The multiplier $N/(N-p)$ where p is the number of input parameters of the model ($p = 12$ for Model B3), is very close to 1 because $N \gg p$ (and could thus be dropped). This multiplier is used in Eq. (2) to eliminate a different (and much milder) kind of bias, namely, to prevent the variance of regression errors of the database with a finite number N of data points from being systematically smaller than the variance of a theoretical database with $N \rightarrow \infty$.^{44,45} Another reason why this multiplier is necessary is that a set of only p data points can be fitted exactly (that is, with no errors).

Let the intervals or boxes of data be labeled by one index, $i = 1, 2, \dots, n$, running consecutively through all the data sets in the database, as illustrated in Fig. 3(a) and (b). To counteract the human bias, every box of every data set must be assigned 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 $\sum_{i=1}^n w_i = 1$,

$$w_i = \frac{1}{m_i \bar{w}}, \quad \bar{w} = \sum_{i=1}^n \frac{1}{m_i} \quad (3)$$

To compare various models, one must use dimensionless statistical indicators of scatter. In regression statistics, two kinds of such dimensionless indicators are recognized. One is the coefficient of variation of regression errors, which characterizes ratio of the scatter band width to the mean, and is defined as

$$\omega = \frac{s}{\bar{y}}, \quad \bar{y} = \frac{\bar{w}}{n} \sum_{i=1}^n w_i \sum_{j=1}^m y_{ij} \quad (4)$$

where \bar{y} represents the weighted mean of all the measured values y_{ij} in the database (the expression used in Reference 21, namely $w_i = N/nm_i$, might seem to be different but is, in fact, equivalent to Eq. (3) because N/n is constant).

While the coefficient of variation ω characterizes the ratio of the scatter band width to the data mean (and should be minimized), the correlation coefficient ρ (which should be maximized) is used in statistics to characterize the ratio of the scatter band width to the overall spread of data, including the spread caused by systematic statistical trend. This coefficient (or the coefficient of determination) indicates what percentage of data variation is accounted for by the prediction model. Generalizing the definition of ρ from linear regression^{44,45}

$$\rho = \sqrt{1 - \frac{s^2}{\bar{s}^2}}, \quad s^2 = \sum_{i=1}^n w_i \sum_{j=1}^m (y_{ij} - Y_{ij})^2, \quad \bar{s}^2 = \sum_{i=1}^n w_i \sum_{j=1}^m (y_{ij} - \bar{y})^2 \quad (5)$$

where s is the overall weighted standard error of predictions and \bar{s} is the overall weighted standard deviation of all the data.

Figure 5 presents comparisons of the coefficient of variation (ω , Eq. (2) and (4)) and of the correlation coefficients (ρ , Eq. (5)) for the five aforementioned prediction models, based on many types of two-dimensional and one-dimensional boxes. Compared are statistics of the total values, relative values, and strength-scaled values. Furthermore, Tables 1(a), (b), (c), and (d) list the comparisons of the coefficient of variation ω of the five models based on other types of data boxes—one- and three-dimensional (checked were also four-dimensional boxes given by intervals of $\log(t-t')$, $\log t'$, H , and \sqrt{D} , numbering 1400 for compliance and 1120 for shrinkage, but they appeared statistically useless because more than half of them were empty).

In each of these comparisons, Model B3 is found to be the best, except for one value in Table 1 where Model B3 is one of two equal best. Gardner's newer Model GL,²⁰ which modifies his original Model GZ¹⁹ by introducing two key aspects of Bažant and Panula's 1978 Model BP⁹ (the shrinkage function, and the dependence on the size or volume-surface ratio) comes out as second best. Considerably worse but the third best overall is seen to be the CEB model. Because the current ACI 209 model, labeled ACI, is the oldest (introduced in 1972⁸ on the basis of 1960s research), it is not surprising that it comes out as the worst.

SCATTER BAND PLOTS AND OBSTACLES TO REDUCING THEIR WIDTH

The coefficient of variation of compliance as well as shrinkage is quite small ($\leq 8\%$) when the concrete type and other parameters are fixed.^{46,47} The high coefficient of variation values, evident in Fig. 5, are caused by the variability of concrete composition, curing, and other parameters throughout the database, as schematically portrayed in Fig. 3(c) and (d). The consequence is a very broad scatter band in plotting the trend with time, as seen in Fig. 4 and 6(a) and (b). In Fig. 6, the logarithmic time scale is subdivided into five decades and the centroid of data located in each decade is shown by the diamond point. The solid curves connect the points of the decade centroid \pm standard deviation of the data in that decade, and the dashed curves represent the interval centroid \pm standard deviation of the predictions corresponding to the database points in the same decade (if Gaussian distribution is assumed, 14% of the data or predictions would lie above the upper curve, and 14% below the lower curve). As can be seen, the scatter bands of both the data and the predictions are so wide that it is impossible to distinguish among even very different shapes of creep or shrinkage curves of various models.

Nevertheless, for the relative compliance and relative shrinkage, the comparison in Fig. 6(c) and (d) of the bands of interval centroids \pm standard deviation of errors is somewhat more indicative than the analogous comparison for the actual values in Fig. 6(a) and (b). For Model B3, the band of predictions (dashed curves) lies mostly within the band of data (solid curves) and exceeds this band only slightly in a few cases. For Model GL, the band of data is exceeded slightly more. For the ACI 209 model, the band of predictions spreads grossly outside the band of data.

To further reduce the scatter in the time evolution, one might filter from the database all the data belonging to a certain small cube (or three-dimensional box) defined by chosen intervals of three parameters, of $\log t'$, $\log t_0$, and H or \sqrt{D} . Then one could do the same for the predictions of each

Table 1—Standard coefficients of variation of errors of various prediction models in: (a) compliance; (b) shrinkage; (c) relative compliance; and (d) relative shrinkage*

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

*Cubes are in $\log(t - t')$, $\log t'$, and H for compliance or \sqrt{D} for shrinkage.

model corresponding to each extracted point. This would lead to rather narrow scatter bands of the data points and of the corresponding predictions, and in this way, one would see a much greater difference among different models.

There is a problem, though. For the presently chosen parameter intervals, there are as many as 280 such cubes, each of them giving one scatter band of data and one scatter band of the corresponding predictions. The rating of the five prediction models would not be the same for each cube. One could obtain the root-mean-square of the coefficients of variation from all these 280 cubes, but that would be analogous to the regression statistics for four-dimensional boxes, whose significance is debatable for reasons already mentioned. The ranking of the models would get clearer by selecting from the 280 cubes a few typical ones, but such a selection would have to be made intuitively, and thus would be effected by human bias. Therefore, it is preferable not to engage in such comparisons.

NONSTANDARD STATISTICAL INDICATORS USED IN RECENT STUDIES

Gardner's linear coefficient of variation

In Reference 2, the logarithmic scales of load duration $t - t'$ and drying duration $t - t_0$ are divided into intervals equal to decades, labeled as $i = 1, 2, \dots$. The overall mean of data for all the intervals is obtained as $\bar{y} = (1/n)\sum_{i=1}^n \bar{y}_i$ where $\bar{y}_i = (1/m_i)\sum_{j=1}^{m_i} \bar{y}_{ij}$. Herein, \bar{y}_i is the mean of the data in interval i and \bar{y} represents the standard expression for a weighted mean, giving equal weight to each decade of time. The calculation of the overall coefficient of variation of prediction errors, ω_G , however, is nonstandard

$$\omega_G = \frac{s_G}{\bar{y}}, s_G = \frac{1}{n} \sum_{i=1}^n s_i \quad (6)$$

where

$$s_i = \sqrt{\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (y_{ij} - Y_{ij})^2} \quad (7)$$

The bias due to having different numbers m_i of points in different intervals is herein compensated by using the coefficient of variation for each interval, which correctly

gives to each time interval the same weight. The expression in Eq. (6) for the overall standard deviation \bar{s} of the data from the model predictions, however, is not statistically justified because, instead of averaging the squared errors s_i^2 , the averaging is linear in s_i . Properly, the averaging must be applied to the squared errors. The linear averaging of s_i is tantamount to denying the validity of the maximum likelihood criterion and of the central limit theorem of the theory of probability, underpinning the Gaussian distribution (see the Appendix). This implicit denial is untenable (it is true that linear averaging of errors has recently been used for some special purposes in financial statistics,²⁶ but that was in problems of extreme value statistics, to which the least-square regression and the central limit theorem of the theory of probability do not apply).

The definition of error used for ranking of various prediction models is correct only if the minimization of error yields the optimum data fit. Otherwise, a smaller error would not mean a better model. In the case of Eq. (6), one would have to minimize the expression

$$s_G^2 = \frac{1}{n^2} \left(\sum_{i=1}^n \sqrt{\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (y_{ij} - Y_{ij})^2} \right)^2 \neq \text{quadratic} \quad (8)$$

Because this paper deals with a statistical problem in which the data represent not merely a population (or ensemble) of realizations of one stochastic variable but the realizations of a variable with a statistical trend, the correct, generally accepted, statistical approach is not population statistics but the least-square statistical regression (Fig. 3(c) and (d)).⁴⁸⁻⁵⁷ Therefore, in the special limit case of a linear model, the statistical method must reduce to linear regression statistics. This is a simple but fundamental check on the soundness of the statistical approach to the comparison of prediction models.

In the special case of a two-dimensional linear model, $Y_{ij} = a + bX_{ij}$, Eq. (6) would give the following expression to be minimized

$$s_G^2 = \frac{1}{n^2} \left(\sum_{i=1}^n \sqrt{\frac{1}{m_i - 1} \sum_{j=1}^{m_i} [y_{ij} - (a + bX_{ij})]^2} \right)^2 \quad (9)$$

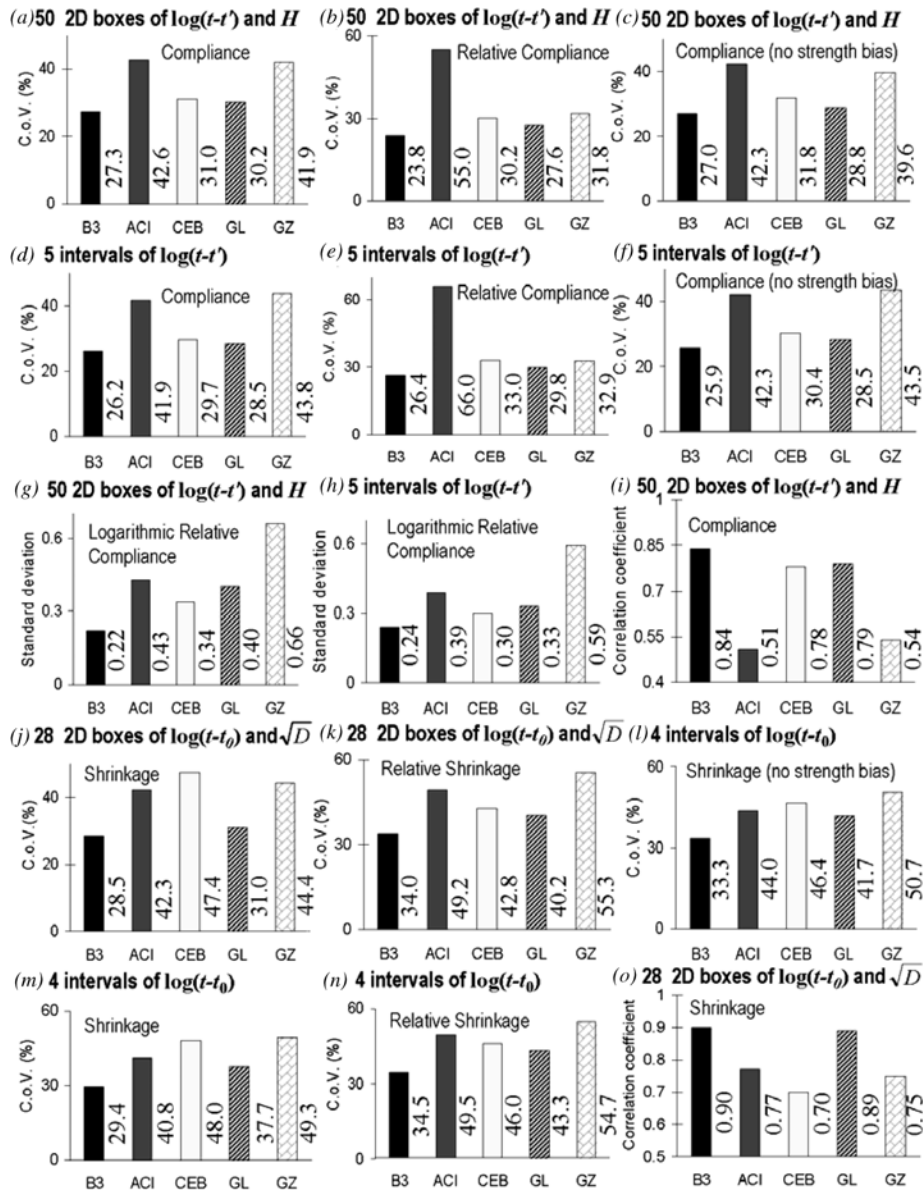


Fig. 5—Coefficients of variation of errors (a) through (h) and (j) through (n), which should be minimized, and correlation coefficients (i) and (o), which should be maximized, for five prediction models and NU-ITI database; (a), (d), (i), (j), (m), and (o) for actual data; (b), (e), (k), and (n) for relative data; (g) and (h) for logarithmic relative data; and (c), (f), and (l) for data scaled by $(\text{strength})^{-1/2}$.

where X_{ij} are the coordinates (for example, the values of $\log(t - t^i)$) of data points Y_{ij} . The minimizing conditions $\partial s_G^2 / \partial a = 0$ and $\partial s_G^2 / \partial b = 0$ would then yield two equations for a and b . It is easy to see that these equations will be nonlinear, and thus will not guarantee a unique solution, despite linearity of the regression problem. The nonlinearity of these equations confirms again that Eq. (6) is invalid.

On the other hand, in the case of the standard error expression in Eq. (2), substitution of $Y_{ij} = a + bX_{ij}$ yields

$$s^2 = \frac{N}{N-p} \sum_{i=1}^n w_i \sum_{j=1}^{m_i} [y_{ij} - (a + bX_{ij})]^2 = \text{minimum} \quad (10)$$

Herein, the minimizing conditions $\partial s^2 / \partial a = 0$ and $\partial s^2 / \partial b = 0$ yield linear equations, and their solution gives the well known expressions for slope b and intercept a of the regression line.

But can the difference between the statistical indicators \bar{s} in Eq. (2) and (6) be significant? Indeed it can. To document it, consider again the special limit case of a linear model $Y = a + bX$, for which it is known that the correct optimum data fit is obtained if and only if the linear regression is used. Consider two sets of three pairs of data points shown in two diagrams in Fig. 7. (For Set 1, the data are $Y = 0.1$ and 0.3 for $X = 0$, $Y = 1.0$ and 1.3 for $X = 1$, and $Y = 2.1$ and 2.4 for $X = 2$. For Set 2, the data are $Y = 0.1$ and 0.3 for $X = 0$, $Y = 0.2$ and 1.8 for $X = 1$, and $Y = 1.7$ and 1.9 for $X = 2$). In each diagram, the regression line is drawn and the values of the coefficient of variation obtained according to the least-square linear regression and according to Eq. (6) are indicated. For Set 1 (left diagram), the correct coefficient of variation (based on linear regression) is 14%, whereas Eq. (6) gives 16%. This is not a great discrepancy. For Set 2 (right diagram), however, the correct coefficient of variation is

57%, which is 21% larger than the value given by Eq. (6), which is 47%. This discrepancy is not insignificant.

CEB coefficient of variation

In Reference 3 (compare with References 5 and 38), the coefficient of variation of prediction model errors was defined as

$$\omega_{CEB} = \sqrt{\frac{1}{n} \sum_{i=1}^n \omega_i^2}, \quad \omega_i = \frac{1}{\bar{y}_i} \sqrt{\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (Y_{ij} - y_{ij})^2}, \quad \bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij} \quad (11)$$

Note that because $m_i - 1$ appears in the denominator and is 0 for a box with only one point, $m_i = 1$, not only the empty boxes but also those with a single point have to be deleted in calculating this statistic.

This statistic has a different shortcoming: The statistical trend is ignored because the statistics of creep and shrinkage data are herein calculated as the population (or ensemble) statistics. In principle, converting a statistical problem having a data trend with respect to some variable to a problem of population statistics (refer to Fig. 3(e) and (f)) is not a legitimate statistical approach and leads to misleading comparisons.

The conversion from least-square regression to population statistics was effected by treating the ω_i in the individual intervals as the coefficients of variation of different groups of realizations of one and the same statistical variable with no trend. But this is unreasonable if the data exhibit a statistical trend with respect to some parameter (in this case, the time) because the errors must be measured with respect to the trend and not the data mean.

Another objectionable aspect is that, compared with the least-square statistical regression, the short-time data get overemphasized and the long-time data get underemphasized. This is caused by the appearance of \bar{y}_i (rather than \bar{y}) in the denominator of Eq. (11) before all ω_i are combined into one coefficient of variation. An interval with a nearly vanishing \bar{y}_i (which occurs, for example, for short-time shrinkage) gives a very large ω_i and thus, incorrectly, dominates the entire statistics.

Can the difference from the correct statistical indicator in Eq. (2) be significant? Very much so. To demonstrate it, the authors again consider the limiting special case of a linear model and the example of two sets of data in Fig. 7. The coefficient of variation for Set 1 (the left diagram) is found to be 44%, which is 214% larger than the correct value of 14% from linear regression. The coefficient of variation for Set 2 (the right diagram) is found to be 77%, which is 35% larger than the correct value of 57%.

CEB mean-square relative error

In Reference 3 (compare with References 5 and 38), another comparison is made on the basis of the relative error defined as

$$S_{CEB} = \sqrt{\frac{1}{n} \sum_{i=1}^n S_i^2}, \quad S_i^2 = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} \left(\frac{Y_{ij}}{y_{ij}} - 1 \right)^2 = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} w_{ij} (Y_{ij} - y_{ij})^2 \quad (12)$$

where $w_{ij} = 1/y_{ij}^2$. Unlike the previous case, this definition of error is based on the method of least squares. But it is applied to the model-data ratio, which implies unrealistic weighting of the data. As shown by the last expression, it means that the weights w_{ij} are inversely proportional to y_{ij}^2 . This causes the errors in the small compliance or shrinkage values to be greatly overemphasized, and the errors in the large values to

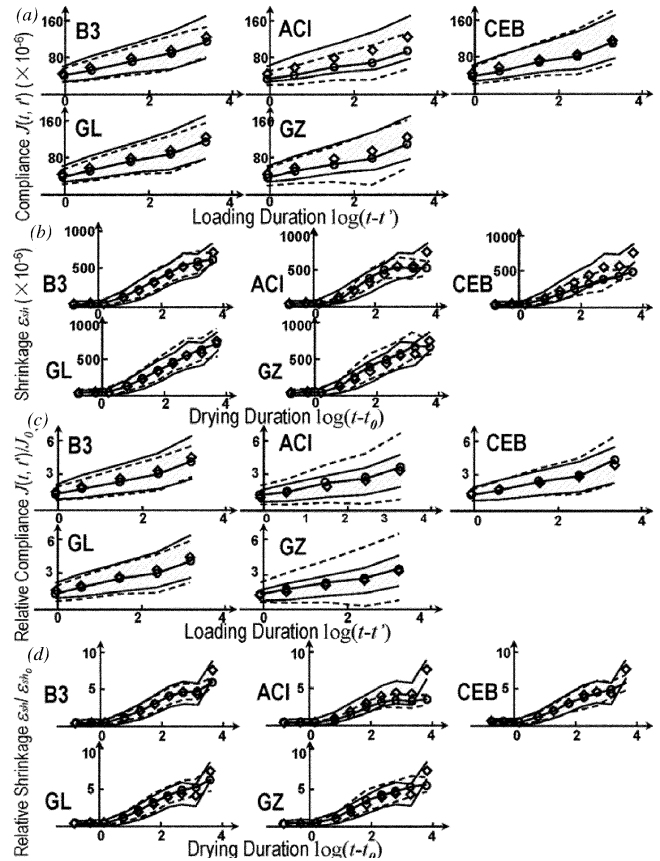


Fig. 6—Bands of interval centroids \pm standard deviation for actual data (solid lines) and predicted values (dashed lines) for: (a) compliance; (b) shrinkage; (c) relative compliance; and (d) relative shrinkage for various prediction models.

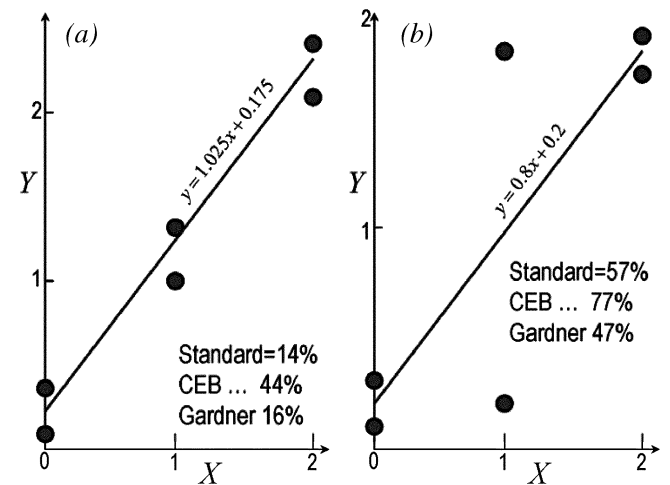


Fig. 7—Differences in coefficients of variation of errors between standard and nonstandard statistical methods for examples of linear regression.

be greatly underemphasized. For example, if the compliance increases four times, its weight will be 16 times smaller. Yet, the long-time predictions are the most important, whereas the short-time ones are the least important. A short-time value 10 times smaller than a long-time value will have a weight 100 times larger, and will thus totally dominate the statistics, making the long-term data irrelevant.

Table 2—Comparison of standard and nonstandard statistical indicators of errors used by various authors to compare and rank four predictions models, for (a) compliance and (b) shrinkage

	(a) Compliance, %						(b) Shrinkage, %				
	Indicator	ACI	B3	CEB	GL		Indicator	ACI	B3	CEB	GL
Bažant ²¹ basic creep	ω	58	24	35	—	Bažant ²¹	ω	55	34	46	—
Bažant ²¹ drying creep	ω	45	23	32	—	Al-Manaseer ⁵	ω_{CEB}	46	41	52	37
Al-Manaseer ⁵	ω_{CEB}	48	36	37	35		S_{CEB}	83	84	60	84
	S_{CEB}	32	35	31	34		M_{CEB}	122	107	75	126
	M_{CEB}	86	93	92	92	Gardner ²	ω_G	41	20	—	19
Gardner ²	ω_G	30	27	—	22	Gardner ² recalculated	ω_G	41	20	44	22

Coefficient of variation of data/prediction ratios

Noting that, in a perfect model, the data-model ratios $r_{ij} = y_{ij}/Y_{ij}$ should be as close to 1 as possible, some studies calculate the coefficient of variation of r_{ij} and use it to compare the prediction models. But this approach to statistics, endemic in concrete research, is incorrect. To show the problem, replace, for the sake of brevity, the double indices ij by a single index $k = 1, 2, \dots, K$ where $K = \sum_{i=1}^K m_i$. The variance s_R^2 of the population of all $r_k = y_k/Y_k$ is

$$s_R^2 = \sum_{k=1}^K w_k \left(\frac{y_k}{Y_k} - \bar{r} \right)^2, \quad \bar{r} = \sum_{k=1}^K w_k \frac{y_k}{Y_k} \quad (13)$$

where w_k are the weights such that $\sum_{k=1}^K w_k = 1$ and \bar{r} is the weighted mean of all r_k . Consider now that the prediction formula giving Y_k is multiplied by any constant factor c , that is, $Y_k \leftarrow cY_k$. Then the variance changes from s_R^2 to \tilde{s}_R^2 as follows

$$\tilde{s}_R^2 = \sum_{k=1}^K w_k \left(\frac{y_k}{cY_k} - \sum_{m=1}^K w_m \frac{y_m}{cY_m} \right)^2 = \frac{1}{c^2} s_R^2 \quad (14)$$

So, as can be seen, the variance of the model-data ratios can be made arbitrarily small by multiplying the prediction formula by a sufficiently large number. Because the mean \bar{r} is replaced by \bar{r}/c , the coefficient of variation $\omega_r = s_R/\bar{r}$ is found to be independent of c .⁵⁸

Therefore, the minimization of s_R^2 cannot be used for the purpose of data fitting. It follows that the use of the coefficient of variation ω_r in some studies, intended for statistical comparison of different models, is unreasonable and misleading. Further, it follows that the plots of data-model ratios r_k versus time (or versus k) should not be used for visual comparisons of the goodness of data fits by various creep prediction models.

To justify Eq. (12) or (13), it is reasoned that the relative errors $\Delta Y_k/Y_k$ are more important than actual errors ΔY_k . But if so, and if the data in $\ln Y_k$ are closer to being homoscedastic than those in Y_k , then the correct approach is the least-square regression of $\ln Y_k$ rather than Y_k (because $d \ln Y_k = dY_k/Y_k$).

Minimization of s_R^2 also fails the test that, in the case of a linear model, minimization of the coefficient of variation of all r_{ij} must reduce to linear regression. Another problem is that the difference $1 - r_{ij}$ tends to be the greatest for short times, which thus dominate the statistics, although the long times are of main interest.

MODEL COMPARISONS BY STANDARD AND NONSTANDARD STATISTICAL INDICATORS

The aforementioned standard and nonstandard statistical indicators have been calculated for all the presently considered

prediction models using the present database, as well as the RILEM database and Gardner’s drastically reduced database. From the last two databases, it was necessary to delete a few data sets for which the parameters required for evaluating some of the prediction models were not known.

The results are shown by the column diagrams in Fig. 8. As seen, the nonstandard indicators give a very different ranking of prediction models than the standard indicators. According to the standard indicator (Eq. (4)), Model B3 appears as the best, and both the classical ACI 209 model and the GZ model as by far the worst, although not according to the nonstandard indicators in Eq. (6), (11), and (12). Model GL is the second best according to the standard indicator when the present database is used, and the best⁷ according to Gardner’s indicator (Eq. (6)) when his reduced database is used.

The five creep and shrinkage prediction models considered herein were statistically also compared by Gardner and Al-Manaseer,^{2,5} and these comparisons were featured in a recent report⁷ (see the results listed in Table 2(a) and (b) reproduced from Reference 7). Nonstandard indicators were used for models other than Model B3. The RILEM database, which is a slightly reduced subset of the present one, was used, except that a drastically reduced subset of the RILEM database was used by Gardner² (recalculations using Gardner’s database, kindly made available to the writers, showed that one value needs to be corrected, as shown by shrinkage prediction of Model GL by ω_G in the last row of Table 2), that is, the last value should be 22 instead of 19. Unfortunately, the nonstandard statistical indicators were considered as equally relevant, and so it is no surprise that each different statistical indicator placed a different prediction model on top or bottom.

OTHER ASPECTS OF MODEL EVALUATION AND COMPARISON, AND CROSS-VALIDATION

Because the variations in concrete strength, composition, and curing cause by far the greatest random scatter of creep and shrinkage predictions, good long-time predictions can be achieved only by extrapolating short-time tests, or updating of the prediction model according to such tests, or a combination of both. Realistic extrapolation of shrinkage and drying creep data requires measuring weight loss of test specimens.^{21,27} Using a statistically correct extrapolation method^{21,27} is one essential requirement for reliable long-time predictions. The second requirement is to use a model of a form that allows easy fitting of short-time data by adjustments of its parameters according to linear regression. The third requirement is to use a model having correct shapes of the curves of creep and shrinkage versus time, the age at the start of loading or drying, the environmental humidity, and the

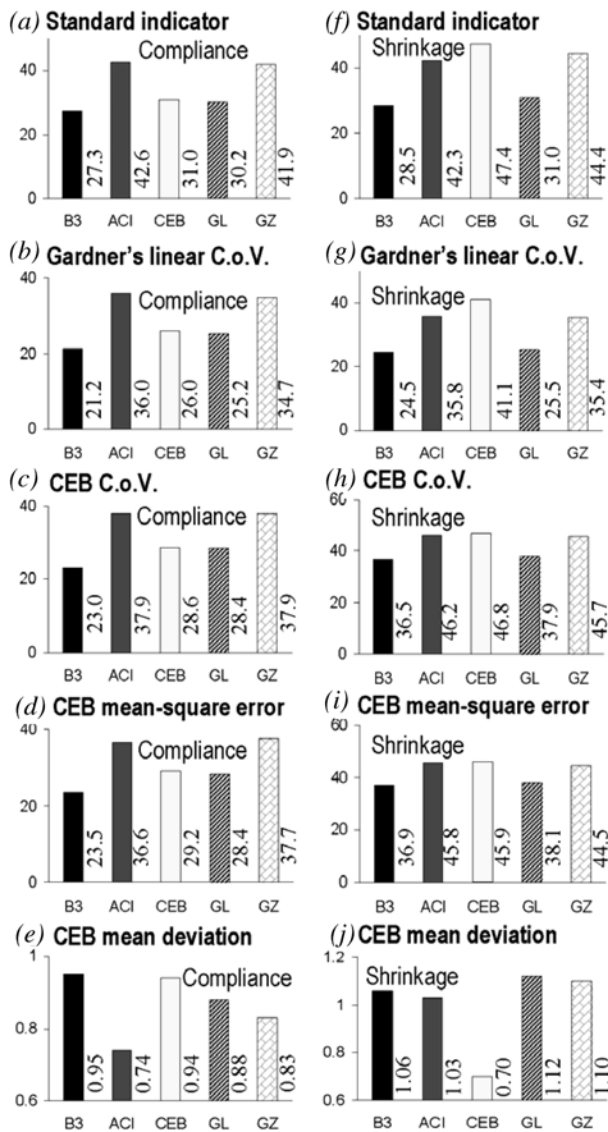


Fig. 8—Comparisons of coefficient of variation (in percent) of various models by means of standard and nonstandard statistical indicators; based on NU-ITI database, with 50 boxes of $\log(t - t')$ and H for creep, and 28 boxes of $\log(t - t_0)$ and \sqrt{D} for shrinkage.

effective thickness of cross section. Unlike others, Model B3 satisfies all these requirements.

Correctness of the shape of time curves cannot be judged by comparisons with the entire database because it is masked by huge scatter resulting from variations of strength, composition, curing, and other parameters. It can be appraised only by comparisons with the creep and shrinkage curves for one and the same concrete, conducted in one and the same laboratory, for one and the same precisely controlled curing. Only if the model can fit such curves closely, is it suitable for extrapolation of short-time data. A few examples of such fits with Model B3, extracted from the examples in References 13 and 21, are presented in Fig. 9.

Model B3 was calibrated by the RILEM database, which was smaller than the present one. The fact that it is evaluated herein by an enlarged database represents a part of the procedure of cross-validation.⁵⁹ The fact that the fits of enlarged database remain equally close supports Model B3.

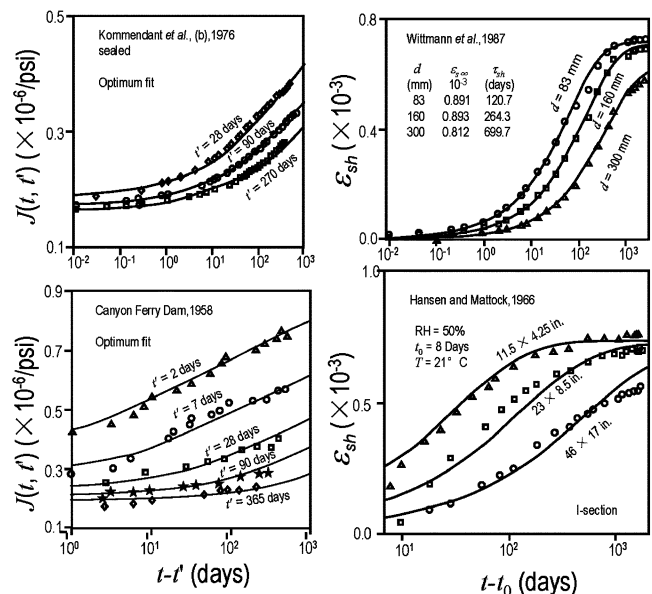


Fig. 9—Fits of characteristic long-time compliance and shrinkage data by formulas of Model B3.

CONCLUSIONS

Based on the research, the following conclusions can be made:

1. The highly nonuniform data distribution in the database is a result of human choice. It introduces an unintended bias, which must be suppressed. This can be accomplished by data weighting;

2. Although the precise weighting to use is debatable and weight differences less than an order of magnitude are not very important, it is reasonable to assign the same weight to the total of all test data within each interval of time, size, humidity, and age at loading or start of drying. This is the basic premise of minimizing the statistical bias;

3. The nonstandard statistical indicators examined herein are not rational approaches for comparing the accuracy of prediction models. They do not yield estimates of maximum likelihood, conflict with the principles of least-square regression, are tantamount to denying the central limit theorem of the theory of probability, and do not lead to a model for which the errors would be uncorrelated;

4. Therefore, the previous rankings of various prediction models obtained by these nonstandard indicators cannot be taken seriously⁷; and

5. In the present comparisons of five prediction models based on the standard statistical indicators and the complete database, Model B3 comes out as the best and Model GL as the second. The old 1972 ACI Model, reappraised by ACI in 2008, comes out as the worst.

ACKNOWLEDGMENTS

Financial support from the U.S. Department of Transportation through the Infrastructure Technology Institute of Northwestern University, provided under Grant 0740-357-A222, is gratefully appreciated.

REFERENCES

- Reineck, K.-H.; Kuchma, D. H.; Kim, K.-S.; and Marx, S., "Shear Database for Reinforced Concrete Members without Shear Reinforcement," *ACI Structural Journal*, V. 100, No. 2, Mar.-Apr. 2003, pp. 240-249.
- Gardner, N. J., "Comparison of Prediction Provisions for Drying Shrinkage and Creep of Normal Strength Concretes," *Canadian Journal of Civil Engineering*, V. 31, No. 5, Sept.-Oct. 2004, pp. 767-775.
- Müller, 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," Comité Euro-Internationale du Béton, Lausanne, Switzerland, 1990, 201 pp.

4. Al-Manaseer, A., and Lakshmi Kantan, S., "Comparison between Current and Future Design Code Models for Shrinkage and Creep," *Revue Française de Génie Civil* 3, V. 3-4, 1999, pp. 39-59.

5. Al-Manaseer, A., and Lam, J.-P., "Statistical Evaluation of Creep and Shrinkage Models," *ACI Materials Journal*, V. 102, No. 3, May-June 2005, pp. 170-176.

6. McDonald, D. B., and Roper, H., "Accuracy of Prediction Models for Shrinkage of Concrete," *ACI Materials Journal*, V. 90, No. 3, May-June 1993, pp. 265-271.

7. ACI Committee 209, "Guide for Modeling and Calculating Shrinkage and Creep in Hardened Concrete (209.2R-08)," American Concrete Institute, Farmington Hills, MI, 2008, 45 pp.

8. ACI Committee 209, "Prediction of Creep, Shrinkage, and Temperature Effects in Concrete Structures (ACI 209R-92)," American Concrete Institute, Farmington Hills, MI, 1992 (Reapproved 2008), 47 pp.

9. Bažant, Z. P., and Panula, L., "Practical Prediction of Time-Dependent Deformations of Concrete. Part I: Shrinkage. Part II: Creep," *Materials and Structures*, V. 11, No. 65, 1978, pp. 307-328.

10. Bažant, Z. P., and Panula, L., "Practical Prediction of Time-Dependent Deformations of Concrete. Part III: Drying Creep. Part IV: Temperature Effects on Basic Creep," *Materials and Structures*, V. 11, No. 66, 1978, pp. 415-434.

11. Bažant, Z. P., and Panula, L., "Practical Prediction of Time-Dependent Deformations of Concrete. Part V: Temperature Effects on Drying Creep. Part VI: Cyclic Creep," *Materials and Structures*, V. 12, No. 69, 1979, pp. 169-183.

12. Comité Euro-International du Béton, "CEB-FIP Model Code 1990: Design Code," Thomas Telford Ltd., London, UK, 1993, 464 pp.

13. Bažant, Z. P.; Kim, J.-K.; and Panula, L., "Improved Prediction Model for Time-Dependent Deformations of Concrete: Part 1—Shrinkage," *Materials and Structures*, V. 24, No. 143, 1991, pp. 327-345.

14. Bažant, Z. P., and Kim, J.-K., "Improved Prediction Model for Time-Dependent Deformations of Concrete: Part 2—Basic Creep," *Materials and Structures*, V. 24, No. 144, 1991, pp. 409-421.

15. Bažant, Z. P., and Kim, J.-K., "Improved Prediction Model for Time-Dependent Deformations of Concrete: Part 3—Creep and Drying," *Materials and Structures*, V. 25, No. 145, 1992, pp. 21-28.

16. Bažant, Z. P., and Kim, J.-K., "Improved Prediction Model for Time-Dependent Deformations of Concrete: Part 4—Temperature Effects," *Materials and Structures*, V. 25, No. 146, 1992, pp. 84-94.

17. Bažant, Z. P., and Kim, J.-K., "Improved Prediction Model for Time-Dependent Deformations of Concrete: Part 5—Cyclic Load and Cyclic Humidity," *Materials and Structures*, V. 25, No. 147, 1992, pp. 163-169.

18. Sakata, K., "Prediction of Concrete Creep and Shrinkage," *Creep and Shrinkage of Concrete*, Proceedings of the 5th International RILEM Symposium, Barcelona, Spain, Z. P. Bažant and I. Carol, eds., E&F Spon, London, UK, 1993, pp. 649-654.

19. Gardner, N. J., and Zhao, J. W., "Creep and Shrinkage Revisited," *ACI Materials Journal*, V. 90, No. 3, May-June 1993, pp. 236-246, with discussion by Z. P. Bažant and S. Baweja, *ACI Materials Journal*, V. 91, No. 2, Mar.-Apr. 1994, pp. 204-216.

20. Gardner, N. J., and Lockman, M. J., "Design Provisions for Drying Shrinkage and Creep of Normal Strength Concrete," *ACI Materials Journal*, V. 98, No. 2, Mar.-Apr. 2001, pp. 159-167.

21. Bažant, Z. P., and Baweja, S., "Creep and Shrinkage Prediction Model for Analysis and Design of Concrete Structures: Model B3," *The Adam Neville Symposium: Creep and Shrinkage-Structural Design Effects*, ACI SP-194, A. Al-Manaseer, ed., American Concrete Institute, Farmington Hills, MI, 2000, pp. 1-83.

22. Sakata, K.; Tsubaki, T.; Inoue, S.; and Ayano, T., "Prediction Equations of Creep and Drying Shrinkage for Wide-Range Strength Concrete," *Creep, Shrinkage and Durability Mechanics of Concrete and Other Quasi-Brittle Materials*, Proceedings of the 6th International Conference, CONCREEP-6, F.-J. Ulm, Z. P. Bažant, and F. H. Wittmann, eds., Elsevier, Amsterdam, the Netherlands, 2001, pp. 753-758.

23. Bažant, Z. P.; Li, G.-H.; and Yu, Q., "Prediction of Creep and Shrinkage and Their Effects in Concrete Structures: Critical Appraisal," *Proceedings of the 8th International Conference on Concrete Creep and Shrinkage*, CONCREEP-8, T. Tanabe, ed., Ise-Shima, Japan, Oct. 2008, pp. 1275-1289.

24. Gauss, K. F., *Theoria Motus Corporum Caelestium*, Hamburg, Germany, 1809.

25. Bulmer, M. G., *Principles of Statistics*, Chapter 12, Dover Publications, New York, 1979, 252 pp.

26. Bouchaud, J.-P., and Potters, M., *Theory of Financial Risks: From Statistical Physics to Risk Management*, Cambridge University Press, Cambridge, UK, 2000, 400 pp.

27. Bažant, Z. P., and Baweja, S., "Creep and Shrinkage Prediction Model for Analysis and Design of Concrete Structures: Model B3," *Materials and Structures*, V. 28, 1995, pp. 357-367.

28. FIB, "Structural Concrete: Textbook on Behaviour, Design and Performance, Updated Knowledge of the of the CEB/FIP Model Code 1990," *Bulletin* No. 2, V. 1, Fédération internationale du béton (FIB), Lausanne, Switzerland, 1999, pp. 35-52.

29. Bažant, Z. P., "Theory of Creep and Shrinkage in Concrete Structures: A Précis of Recent Developments," *Mechanics Today*, V. 2, S. Nemat-Nasser, ed., Pergamon Press, 1975, pp. 1-93.

30. Bažant, Z. P., "Mathematical Models of Nonlinear Behavior and Fracture of Concrete," *Nonlinear Numerical Analysis of Reinforced Concrete*, L. E. Schwer, ed., ASCE, New York, 1982, pp. 1-25.

31. RILEM Committee TC-69, "State of the Art in Mathematical Modeling of Creep and Shrinkage of Concrete," *Mathematical Modeling of Creep and Shrinkage of Concrete*, Z. P. Bažant and J. Wiley, eds., Chichester, UK, 1988, pp. 57-215.

32. Bažant, Z. P., "Criteria for Rational Prediction of Creep and Shrinkage of Concrete," *The Adam Neville Symposium: Creep and Shrinkage—Structural Design Effects*, SP-194, A. Al-Manaseer, ed., American Concrete Institute, Farmington Hills, MI, 2000, pp. 237-260

33. Hillsdorf, H. K., and Carreira, D. J., "ACI-CEB Conclusions of the Hubert Rüscher Workshop on Creep of Concrete," *Concrete International*, V. 2, No. 11, Nov. 1981, p. 77.

34. Müller, H. S.; Bažant, Z. P.; and Kuttner, C. H., "Database on Creep and Shrinkage Tests," *RILEM Subcommittee 5 Report*, RILEM TC 107-CSP, 1999, 81 pp.

35. Müller, H. S., "Considerations on the Development of a Database on Creep and Shrinkage Tests," *Creep and Shrinkage of Concrete*, Proceedings of the 5th International RILEM Symposium, Barcelona, Spain, Z. P. Bažant and I. Carol, eds., E&F Spon, London, UK, 1993, pp. 859-872.

36. Bažant, Z. P., and Li, G.-H. "Database on Concrete Creep and Shrinkage," Infrastructure Technology Institute (ITI), Northwestern University, Evanston, IL, 2008, http://www.iti.northwestern.edu/research/completed/tea-21/bazant/bazant_a423.html.

37. Gardner, N. J., "Design Provisions for Shrinkage and Creep of Concrete," *The Adam Neville Symposium: Creep and Shrinkage—Structural Design Effects*, SP-194, A. Al-Manaseer, ed., American Concrete Institute, Farmington Hills, MI, 2000, pp. 101-104.

38. Al-Manaseer, A., and Lakshmi Kantan, S., "Comparison between Current and Future Design Code Models for Creep and Shrinkage," *Revue française de génie civil*, V. 3, No. 3-4, 1999, pp. 39-40.

39. Krístek, V.; Bažant, Z. P.; Zich, M.; and Kohoutková, "Box Girder Deflections: Why is the Initial Trend Deceptive?" *Concrete International*, V. 28, No. 1, Jan. 2006, pp. 55-63.

40. Bažant, Z. P.; Li, G.-H.; and Yu, Q., "Explanation of Excessive Long-Time Deflections of Collapsed Record-Span Box Girder Bridge in Palau," *Structural Engineering Report No. 08-09/A222e*, Northwestern University, Evanston, IL, 2008, http://www.iti.northwestern.edu/activities/research/completed/tea-21/bazant/bazant_a423.html.

41. Bažant, Z. P., and Kim, J.-K., "Consequences of Diffusion Theory for Shrinkage of Concrete," *Materials and Structures*, V. 24, No. 143, 1991, pp. 323-326.

42. Bažant, Z. P., and Raftshol, W. J., "Effect of Cracking in Drying and Shrinkage Specimens," *Cement and Concrete Research*, V. 12, 1982, pp. 209-226.

43. Ang, A. H.-S., and Tang, W. H., *Probability Concepts in Engineering Planning and Design: Decision, Risk and Reliability*, V. II, John Wiley and Sons, New York, 1984, 562 pp.

44. Haldar, A., and Mahadevan, S., *Probability, Reliability and Statistical Methods in Engineering Design*, Wiley, New York, 1999, 320 pp.

45. Ang, A. H.-S., and Tang, W. H., *Probability Concepts in Engineering Planning and Design*, V. I, Wiley, New York, 1976, 406 pp.

46. Bažant, Z. P.; Wittmann, F. H.; Kim, J.-K.; and Alou, F., "Statistical Extrapolation of Shrinkage Data—Part I: Regression," *ACI Materials Journal*, V. 84, No. 1, Jan.-Feb. 1987, pp. 20-34.

47. Wittmann, F. H.; Bažant, Z. P.; Alou, F.; and Kim, J.-K., "Statistics of Shrinkage Test Data," *Cement, Concrete and Aggregates*, V. 9, No. 2, 1987, pp. 129-153.

48. Beck, J. V., and Arnold, K. J., *Parameter Estimation in Engineering Science*, John Wiley and Sons, New York, 1977, 501 pp.

49. Draper, N., and Smith, F., *Applied Regression Analysis*, second edition, Wiley-Interscience, New York, 1981, 736 pp.

50. Fox, J., *Applied Regression Analysis, Linear Models and Related Methods*, Sage Publications, Inc., 1997, 624 pp.

51. Feller, W., *Introduction to Probability Theory and Its Applications*, second edition, Wiley, New York, 1957, 704 pp.

52. Soong, T. T., *Fundamentals of Probability and Statistics for Engineers*,

Wiley-Interscience, New York, 2004, 400 pp.

53. Lehmann, E. L., *Testing Statistical Hypotheses*, John Wiley, New York, 1959, 400 pp.

54. Mandel, J., *The Statistical Analysis of Experimental Data*, Dover Publications, 1984, 448 pp.

55. Plackett, R. L., *Principles of Regression Analysis*, Oxford UP, 1960.

56. Crow, E. L.; Davis, F. A.; and Maxfield, M. W., *Statistics Manual*, Dover Publications, New York, 1960, 288 pp.

57. Benjamin, J. R., and Cornell, C. A., *Probability, Statistics and Decision for Civil Engineers*, McGraw-Hill, New York, 1970, 640 pp.

58. Bažant, Z. P., discussion on “Shear Database for Reinforced Concrete Members without Shear Reinforcement,” by K.-H. Reineck, D. A. Kuchma, K. S. Kim, and S. Marx, *ACI Structural Journal*, V. 101, No. 2, Mar.-Apr. 2004, pp. 139-140.

59. Geisser, S., *Predictive Inference: An Introduction*, CRC Press, Boca Raton, FL, 1993, 240 pp.

60. Legendre, A. M., *Nouvelle méthode pour la détermination des orbites des comètes*, Paris, France, 1806, pp. 576-579.

61. Rousseeuw, R. J., and Leroy, A. M., *Robust Regression and Outlier Detection*, John Wiley and Sons, New York, 1987, 329 pp.

APPENDIX—WHY IS THE LEAST-SQUARE REGRESSION THE ONLY RATIONAL METHOD FOR COMPARING CONCRETE DESIGN MODELS?

The method of least squares was first published by Legendre in 1806⁶⁰ but its rigorous derivation, proving its necessity, is due to Gauss²⁴ (who is known to have used it before 1806). For brevity, let us again replace the double indexes ij by a single data index $k = 1, 2, \dots, K$ where $K = \sum_{i=1}^N m_i$. Let $\varepsilon_k = y_k - Y_k =$ model errors, $k = 1, 2, \dots, K$ where $Y_k = F(X_k, Y_k, \dots) =$ predicted values, $X_k, Y_k, \dots =$ independent variables (for example, influencing parameters such as the load duration, age at loading, thickness, and humidity) associated with data point k , and $F(\dots) =$ function defining the prediction model. The joint probability density distribution of all the data, called the likelihood function L ,²⁵ is

$$L = f(y_1, y_2, \dots, y_k) = \prod_{k=1}^K [\phi(y_k)]^{W_k} \quad (15)$$

where $\phi(y_k)$ is probability density distribution (pdf) of measurement y_k , and W_k is assigned weight (equivalent to W_k -fold repetition of the k -th measurement).

First assume the errors to be approximately normally distributed, that is,

$$\phi(y_k) = e^{-\frac{(y_k - Y_k)^2}{2s_k^2}} / \sqrt{2\pi s_k^2} \quad (k = 1, 2, \dots, K) \quad (16)$$

where $s_k^2 =$ (conditional) variance of y_k , which is a constant known (or knowable) a priori. The optimal data fit must maximize L ,²⁵ that is, minimize $-\ln L$;

$$\begin{aligned} -\ln L &= -\ln \left[e^{-\sum_k W_k (y_k - Y_k)^2 / 2s_k^2} (2\pi s_k^2)^{-W_k/2} \right] \quad (17) \\ &= \sum_k W_k (y_k - Y_k)^2 + C = \min \end{aligned}$$

where $w_k = W_k / 2s_k^2 =$ redefined weights (not normalized), and

$C = \sum_k W_k \ln \sqrt{2\pi s_k^2} =$ constant. Equation (17) shows that minimization of the weighted sum of squared errors is the only

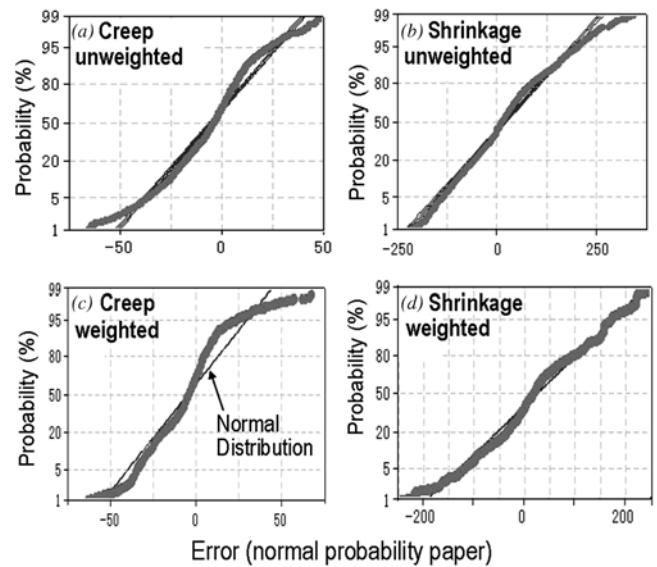


Fig. A—Cumulative histograms of errors of B3 model compared to NU-ITI database, plotted on normal probability paper: left—creep; right—shrinkage; top—unweighted; bottom—weighted.

correct approach if one adheres to the standard requirement of maximum likelihood.

The histograms of data plotted on the normal probability paper (Fig. A) indicate that the distributions or errors ε_{ij} in creep and shrinkage are approximately normal. But the histograms are not broad enough to prove it. So what if the distributions $\phi(y_k)$ of data y_k are not normal? In that case, the database may be subdivided into data groups labeled as $r = 1, 2, \dots, N_g$, such that each group r contains a sufficient but not excessive number n_r of adjacent data points located so closely that the statistical trends within each group are negligible ($n_r \approx 6$ appears suitable). The mean of each data group is a scaled sum of random variables, and according to the *central limit theorem*,^{25,26} the distribution of this sum, and thus the group mean, converges to the normal distribution, albeit one with a scaled standard deviation. We may now logically expect that the best fit of y_k can be obtained as the best fit of all the group means, each of which has a Gaussian distribution. The remaining derivation up to Eq. (17) is the same and leads to the same conclusion. Q.E.D.

Another requirement that inevitably leads to the least-square regression is that, for optimum fit, the random errors $\varepsilon_k = y_k - Y_k$ must be uncorrelated, or else the fit would not capture the statistical trend. So, the correlation matrix of random variables ε_k must be diagonal. If the statistics used are not based on the method of least squares, this requirement will be violated.

To be rigorous, it must be admitted that there exist special problems where the least-square regression is insufficient or even inappropriate. One example is the extreme value statistics, leading to Weibull distribution of strength of brittle structures.²⁶ Another is the extension of the least-square approach to Bayesian optimization, in which the posterior data are supplemented by some sort of prior information.⁴⁴ A third example is the robust regression,⁶¹ used to emphasize the role of numerous outliers of heavily tailed non-Gaussian distributions. But these special problems do not arise for the typical regression problems of concrete design equations discussed herein.