

Unbiased Statistical Comparison of Creep and Shrinkage Prediction Models

ZDENĚK P. BAŽANT AND GUANG-HUA LI

Structural Engineering Report No. 07-12/A210u

Department of Civil and Environmental Engineering
Northwestern University
Evanston, Illinois 60208, USA

December 21, 2007

Unbiased Statistical Comparison of Creep and Shrinkage Prediction Models

Zdeněk P. Bažant¹, Hon. Mem. ASCE, and Guang-Hua Li²

Abstract: The 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, ACI model, CEB model, and two 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. Several nonstandard statistical methods that have recently been invented to deal with creep and shrinkage data are also examined and their deficiencies are pointed out. The ranking of the models that ensues is quite different from the rankings obtained by the nonstandard methods.

Introduction

Altering the statistical method can often lead to very different conclusions. One such instance, where inventions of various nonstandard statistical indicators have recently sown much confusion, is the use of creep and shrinkage databases to evaluate various prediction models [1, 2, 3, 4, 5, 6, 7, 8, 9]. A model that was rated as superior according to one statistical indicator was rated as inferior according to another.

Are all the statistical methods used in different creep and shrinkage studies correct? 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 which matters for structural safety), it is actually very clear what is the correct statistical approach. It is the method of least squares—the standard method which (as shown by Gauss [10]) maximizes the likelihood function and is consistent with the central limit theorem of the theory of probability (see the Appendix [11, 12]). 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. 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. Five models will be considered:

- 1) Model **B3**, 1995, which was approved as the international RILEM Recommendation [13] and slightly updated in 2000 [7]) (this model is a refinement of the 1978 model BP [2] and of its improvement as model BP-KX [4]).
- 2) **ACI** model [1], based on 1960's research.

¹McCormick Institute Professor and W.P. Murphy Professor of Civil Engineering and Materials Science, Northwestern University, 2145 Sheridan Road, CEE/A135, Evanston, Illinois 60208; z-bazant@northwestern.edu.

²Graduate Research Assistant, Northwestern University.

- 3) Model of Comité européen de béton, labelled **CEB**, which is based on the work of Müller and Hilsdorf [14] (it was adopted in 1990 by CEB [3], updated in 1999 [15], and co-opted in 2002 for Eurocode 2).
- 4) Gardner and Lockman's model, labelled **GL** [9].
- 5) Gardner's earlier model, labelled **GZ** [6].

Sakata's model [5, 8], whose scope is somewhat limited, as well as the crude old models of Dischinger, Illston, Nielsen, Rüsç and Jungwirth, Maslov, Arutyunyan, Aleksandrovskii, Ulickii, Gvozdev, Prokopovich and others [16, 17, 18], will be left out of consideration.

Although there exist certain fundamental theoretical requirements [19], which are essential for choosing the right model, necessitate rejecting some models even before their comparison to test data, and happen to favor model B3, most engineers place emphasis on statistical comparisons with the existing experimental database. Therefore, this paper will deal exclusively with statistics.

The first comprehensive database, comprising about 400 creep tests and about 300 shrinkage tests, was compiled at Northwestern University in 1978 [2], mostly from American and European tests. In collaboration with CEB, begun at the 1980 Rüsç Workshop [20], this database was slightly expanded by an ACI-209 subcommittee chaired by L. Panula. A further slight expansion was undertaken in a subcommittee of RILEM TC-107, chaired by H. Müller. It led to what became known as the RILEM database [21, 14, 22], which contained 518 creep tests and 426 shrinkage tests. Recently, a significantly enlarged database, named NU-ITI database [23] 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 data. A reduced database, consisting of 166 creep tests and 106 shrinkage tests extracted from the RILEM database, has recently been used in Gardner's studies [24, 9, 25].

Among concrete researchers, a popular way to verify and calibrate a 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) $\epsilon_k = y_k - Y_k$ versus time (Fig. 1) [26, 27, 28]. If the model 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. Fig. 1 shows examples of such plots for some of the aforementioned models and the NU-ITI database. One immediately notes that, in this kind of comparison, there is very little difference among the models, even those which are known to give very different long-time predictions. 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 $r_k = 1$ (for problems of such kind of statistics, see comments on Eqs. 15 and 16 which follow). Therefore, such comparisons are ineffectual for our purpose. The causes are four:

- 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 evident from the histograms in Fig. 2.
- 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. Long-duration tests of modern high strength concretes, which creep little, are still rare, as documented by Fig. 2.
- 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 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 found in the databases. The present paper attempts to overcome these deficiencies.

If the time, age and specimen size are transformed to variables that make the trends uniform and the data set almost homoscedastic [29], 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 (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 motivation is the need to compare the existing models using the correct statistical approach and to explain why some previous attempts at such comparisons 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 [30]. 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 span when built in 1977, the long-time deflection reached 5 ft. (1.52m) 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, 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. For example, if one experimenter crowds 1000 readings into the time interval from 1000 to 2000 days and another experimenter takes just 3 readings in that same interval, the unweighted error statistics of the prediction model will be almost independent of the latter experimenter's data and will be totally dominated by the former experimenter's data.

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 [7] 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 (Fig.3a), which means that the subdivisions of $t - t'$ and $t - t_0$ form a geometric progression ($t =$ time, representing the current age of concrete, $t_0 =$ age at the start of drying, and $t - t_0 =$ duration of shrinkage test; all the times are given in days). The reason for this kind of subdivision into intervals is twofold:

- 1) One reason, already invoked, is that the least-square statistical regression must be conducted in variables in which the data appear as approximately homoscedastic, i.e., have

an approximately uniform conditional variance about the regression line [31, 11] (as in Fig. 3c,d). Plotting the creep or shrinkage data in terms of the load duration $t - t'$ or drying duration $t - t_0$, one finds the data to be markedly heteroscedastic. To make them homoscedastic, transformation of the variables is the standard approach [29]. As is generally the case when the relative, rather than absolute, changes of response matter, this transformation happens to be achieved by taking a logarithm.

- 2) Because creep and shrinkage are decaying processes, a time increment of, e.g., 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 $D = 2V/S = 2 \times \text{volume/surface ratio}$ of specimen. Here 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 [19, 32, 33]. As for the environmental humidity H , no transformation seems necessary, although small uniform intervals of H are not possible because of gaps in the distribution of H in the database.

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). However, the use of such boxes has another shortcoming: For the database that exists, it appears that the number of data points in some boxes is 0; no statistics can be taken for such boxes and, therefore, they have to be deleted. But deletion of some boxes implies that the relative weights of the data sets will be unequal and will be proportional to the number of boxes in each set. A data set that contains boxes for longer times or longer ages at loading, or boxes for a wider range of environmental humidity or effective thickness, will thus receive a greater weight.

Since boxes of lesser dimensions have a lesser chance of containing only 1 or 2 data points, two-dimensional boxes of $\log(t - t')$ and H for creep, and $\log(t - t_0)$ and \sqrt{D} for shrinkage (Fig. 3b) appear to be preferable over three- or four-dimensional boxes. One-dimensional boxes, or intervals (Fig. 3a), 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. 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. This makes introduction of such weights unimportant in comparison to the weights w_i for the data boxes, which must differ by more than one order of magnitude in order 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, since accuracy of long-time prediction is of the greatest interest. Again, however, 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.

Reducing Anti-High-Strength Bias

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 (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. 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 = cylindrical compressive strength, we will reduce this bias by the following replacements of all the measured compliance and shrinkage data:

$$y \leftarrow y\sqrt{f_c^0/f'_c} \quad (1)$$

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) = constant factor introduced to retain convenient dimensions.

Standard Regression Statistics of the 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 [34, 31, 11]:

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

where m_i, w_i = number of data points in box number i and the statistical weight assigned to the points in this box; $N = N \sum_{i=1}^n w_i = \sum_{i=1}^n m_i$ = number of all the data points in the database; y_{ij} = measured creep or shrinkage data of which the database is comprised; Y_{ij} = corresponding model predictions, and $y_{ij} - Y_{ij} = \epsilon_{ij}$ = errors of the predictions.

The multiplier $N/(N-p)$, where p = 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 (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 $N \rightarrow \infty$ [34, 31]. Another reason why this multiplier is needed is that a set of only p data points can be fitted exactly, i.e, with $s = 0$.

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

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

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 (C.o.V.) of regression errors, which characterizes ratio of the scatter band width to the mean, and is defined as

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

Here \bar{y} represents the weighted mean of all the measured values y_{ij} in the database (the expression used in [7], namely $w_i = N/nm_i$, might seem to be different but is in fact equivalent to Eq. (3) because $N/n = \text{constant}$).

While the coefficient of variation, ω , characterizes the ratio of the scatter band width to the data mean, the correlation coefficient, ρ , is used in statistics to characterize the ratio of the scatter band width to the overall spread of data. It indicates what percentage of data variation is accounted for by the prediction model. Generalizing the definition of ρ from linear regression [34, 31], we have

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

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

Figs. 4 presents comparisons of the coefficients of variation ω and correlation coefficients ρ of the five aforementioned prediction models, based on using two-dimensional boxes. Table 1a,b,c,d lists the comparisons of the coefficients of variation ω of the five models, based on using different types of data boxes—one-, two- and three-dimensional (four-dimensional boxes given by intervals of $\log(t - t')$, $\log t'$, H and \sqrt{D} , numbering 1400 for compliance and 1120 for shrinkage, have also been tried but found statistically useless because more than half of them were empty). In all these comparisons, model B3 is found to be the best, except for one case where it is one of two equal best. Gardner's newer model GL [9], which modifies his original model GZ [6] by co-opting two key aspects of Bažant and Panula's 1978 model BP [2] (the shrinkage function and dependence on the size or volume-surface ratio), 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, labelled ACI, is the oldest, introduced in 1972 on the basis of 1960's research [1], it is not surprising that it comes out as the worst.

Problems in Suppressing Scatter Due to Variability of Concrete Type and Test Parameters

The high values of the coefficient of variation, evident in Figs. 4, are caused by the variability of concrete composition, curing and other parameters throughout the database, as schematically portrayed in Fig. 3c,d. The consequence is a very broad scatter band in plotting the trend with time, shown for the database in Figs. 5a,b. In these plots, the logarithmic time scale is subdivided into five decades and, for each decade, the centroid of data is shown by the diamond point; the solid curves connect the points of the interval centroid \pm standard deviation of data for each decade, and the dashed curves represent the interval centroid \pm standard deviation of predictions corresponding to the database points in the same decade (for Gaussian distribution, 14% of the data or predictions would lie above the upper curve, and 14% below the lower curve).

As we see, 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, as illustrated in Fig. 3c.

To reduce the scatter band width, one may take advantage of the experimental observation that a change of concrete composition causes nearly the same relative changes of deformation due to creep or shrinkage for both short and long times. Thus the scatter band width can be reduced by considering the relative compliance or relative shrinkage, defined as the compliance or shrinkage strain divided by its initial value;

$$\bar{J}(t, t') = J(t, t')/J_0, \quad \bar{\epsilon}(t, t_0) = \epsilon(t, t_0)/\epsilon_0 \quad (6)$$

where J_0 and ϵ_0 are the initial values of compliance and shrinkage. These initial values 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, while the compliance for 1 day, or even 0.1 day, of sustained loading could serve almost equally well for creep data).

However, when the relative compliance and relative shrinkage of the entire database are plotted as a function of load or drying duration, the reduction of scatter band width is disappointingly small. This is clear from Fig. 5c,d, giving plots analogous to Fig. 5a,b. The reason is that the division by initial values suppresses only the effect of composition but 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, extracting from the database the statistics of the relative values helps only little.

Nevertheless, the comparison in Figs. 5c,d of the bands of interval centroids \pm standard deviation of errors is, for the relative compliance and relative shrinkage, somewhat more indicative than the comparison in Figs. 5a,b. Note that the only prediction model for which the band of predictions (dashed curves) lies fully within the band of data (solid curves) is model B3. 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, $\log t'$, $\log t_0$ and H or \sqrt{D} . Then one could do the same for the predictions of each 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 we would see a much greater difference among different models.

However, the problem is that, 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. For each of these cubes, the rating of the five prediction models would be different. One could obtain the root-mean-square of the coefficients of variation from all these 280 cubes, but that would be equivalent to the statistics for four-dimensional which, as already mentioned, are not very indicative. 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 by human choice, and thus would be poisoned by human bias. Therefore, it is preferable not to engage in such comparisons.

Nonstandard Statistical Indicators Used in Recent Studies

1) Gardner's Linear Coefficient of Variation (C.o.V.)

In [25], the logarithmic scales of load duration $t - t'$ and drying duration $t - t_0$ are divided into intervals equal to decades, labelled as $j = 1, 2, \dots$. For each decade, the overall mean of data for all the intervals is obtained as $\bar{y} = (1/N) \sum_{i=1}^n \bar{y}_i$ where $\bar{y}_i = (1/m_i) \sum_{j=1}^{m_i} y_{ij}$. Here \bar{y}_j = mean of the data in interval j ; \bar{y} represents the standard expression for a weighted mean, giving equal weight to each decade of time. However the calculation of the overall coefficient of variation of prediction errors, ω_G , is nonstandard:

$$\omega_G = \frac{s_G}{\bar{y}}, \quad s_G = \frac{1}{n} \sum_{i=1}^n s_i, \quad (7)$$

$$\text{where} \quad s_i = \sqrt{\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (y_{ij} - Y_{ij})^2} \quad (8)$$

The bias due to having different numbers m_i of points in different intervals is here compensated by using the coefficient of variation for each interval, which correctly gives to each time interval the same weight. However, the expression in (7) for the overall standard deviation \bar{s} of the data from the model predictions is not statistically correct because, instead of the averaging of squared errors s_i^2 , the averaging is linear in s_i . Correctly, the averaging must be applied to the squared errors. Linear averaging of s_i is tantamount to denying the validity of the central limit theorem of the theory of probability, underpinning the Gaussian distribution (see 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 [12], but that was in problems of extreme value statistics, to which the central limit theorem of the theory of probability does not apply).

For the standard error definition in comparisons of prediction models to be correct, the minimization of the squared standard error must yield the optimum data fit. In the case of Eq. (7), one would have to minimize the expression:

$$s_G^2 = \frac{1}{n^2} \left(\sum_{i=1}^n \sqrt{\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (y_{ij} - Y_{ij})^2} \right)^2 \neq \text{quadratic} \quad (9)$$

Since we deal 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 statistical approach is not population statistics but the least-square statistical regression only, as generally accepted [35, 36, 37, 38, 39, 40, 41, 42]. 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. (7) would give the following expression to be minimized:

$$s_G^2 = \frac{1}{n^2} \left(\sum_{i=1}^n \sqrt{\frac{1}{m_i - 1} \sum_{j=1}^{m_i} [y_{ij} - (a + bX_{ij})]^2} \right)^2 \quad (10)$$

where X_{ij} are the coordinates [e.g., the values of $\log(t - t')$] 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 so must, in general, be expected to have a non-unique solution, despite linearity of the regression problem. The nonlinearity of these equations confirms again that Eq. (7) 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

$$s^2 = \frac{N}{N-p} \sum_{i=1}^n w_i \sum_{j=1}^{m_i} [y_{ij} - (a + bX_{ij})]^2 = \min \quad (11)$$

Here the minimizing conditions $\partial s^2/\partial a = 0$ and $\partial s^2/\partial b = 0$ yield linear equations, whose 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 (7) be significant? Indeed it can. To document it, consider again the special limit case of a linear model $Y = a + bX$, for which we know that the correct optimum data fit is obtained if and only if the linear regression is used. Let us consider 2 sets of 3 pairs of data points shown in 2 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 $X = 2$; and 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. (7) are indicated. For set 1 (left diagram), the correct coefficient of variation (based on linear regression) is 14% while Eq. (7) gives 16%. This is not a great discrepancy. However, for set 2 (right diagram), the correct coefficient of variation is 57%, which is 21% larger than the value given by Eq. (7), which is 47%. This is not an insignificant discrepancy.

2) CEB Coefficient of Variation

In [14] (cf. [27, 28]), the coefficient of variation of prediction model errors was defined as:

$$\omega_{CEB} = \sqrt{\frac{1}{n} \sum_{i=1}^n \omega_i^2}, \quad \omega_i = \frac{1}{\bar{y}_i} \sqrt{\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (Y_{ij} - y_{ij})^2}, \quad \bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij} \quad (12)$$

(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 since the statistics of creep and shrinkage data are here calculated as the population statistics. In principle, converting a statistical problem having a data trend with respect to some variable to a problem of population statistics (Fig. 3e,f) is not a correct 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 incorrect if the data exhibit a statistical trend with respect to some parameter (in our 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 to 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. (12) before all ω_i

are combined into one coefficient of variation. An interval with nearly vanishing \bar{y}_i gives a very large ω_i and thus, incorrectly, dominates the entire statistics.

Is the difference from the correct statistical indicator in Eq. (2) significant? Very much so. To demonstrate it, we consider again the limiting special case of a linear model and the example of 2 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%.

3) CEB Mean-Square Relative Error

In [14] (cf. [27, 28]), another comparison is made on the basis of the relative error defined as

$$S_{CEB} = \sqrt{\frac{1}{n} \sum_{i=1}^n S_i^2}, \quad S_i^2 = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} \left(\frac{Y_{ij}}{y_{ij}} - 1 \right)^2 = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} w_{ij} (Y_{ij} - y_{ij})^2 \quad (13)$$

where $w_{ij} = 1/y_{ij}^2$. Unlike the previous case, this definition of error is consistent with the least-square regression. However, it 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 get greatly overemphasized, and the errors in the large values to get greatly underemphasized. Yet, the long-time predictions are the most important, while the short-time ones are the least.

4) CEB Mean Relative Deviation

Still another comparison in [14] (cf. [27, 28]) employed what was called the mean deviation:

$$M_{CEB} = \frac{1}{n} \sum_{i=1}^n M_i, \quad M_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{Y_{ij}}{y_{ij}} \quad (14)$$

This does not correspond to the method of least squares and, for the special case of a linear model, the minimization of $(M_{CEB} - 1)^2$ does not reduce to linear regression. So this approach is afflicted by all the previously described problems that arise for such nonstandard statistics.

5) Coefficient of Variation of the Data/Prediction Ratios

Noting that, in a perfect model, the 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 is an incorrect approach to statistics, which has been endemic in concrete research. To show the problem, let us replace, for the sake of brevity, the double indices ij by a single index $k = 1, 2, \dots, K$ where $K = \sum_{i=1}^n m_i$. The variance s_R^2 of the population of all $r_k = y_k/Y_k$ is

$$s_R^2 = \sum_{k=1}^K w_k \left(\frac{y_k}{Y_k} - \bar{r} \right)^2, \quad \bar{r} = \sum_{k=1}^K w_k \frac{y_k}{Y_k} \quad (15)$$

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 , i.e., consider the replacement $Y_k \leftarrow cY_k$. Then the variance changes from s_R to \tilde{s}_R as follows:

$$\tilde{s}_R^2 = \sum_{k=1}^K w_k \left(\frac{y_k}{cY_k} - \sum_{m=1}^K w_m \frac{y_m}{cY_m} \right)^2 = \frac{1}{c^2} s_R^2 \quad (16)$$

So, as we note, the variance of the prediction-data ratios can be made arbitrarily small by multiplying the prediction formula by a sufficiently large number. Since 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 [43].

Therefore, 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 prediction models, was incorrect and misleading. Further it follows that the plots of data/prediction ratios r_k versus time (or versus k) should not be used for visual comparison of the goodness of data fits by various creep prediction models.

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 differences $1 - r_{ij}$ tend to be the greatest for short times, which thus dominate the statistics although the long times are of main interest.

Comparison of Models 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 limited 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. 6. As seen, the comparisons of prediction models by nonstandard indicators give very different conclusions than before. According to the standard indicator (Eq. 4), mode B3 appears to be the best. The classical ACI-209 model and the GZ model are by far the worst, although not according to the nonstandard indicators in Eqs. (7), (12) and (13).

The five creep and shrinkage prediction models considered here were statistically compared in one recent committee report and in [25, 28]; the results are listed in Table 2a,b. Unfortunately, the nonstandard statistical indicators were considered as equally relevant, and thus it is no surprise that each different statistical indicator placed a different prediction model on top or bottom (using Gardner's database, which he kindly made available, his indicators were recalculated and yielded somewhat different results, which are also listed in the table).

Other Aspects to Consider in Model Evaluation and Comparison

Since 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 [13, 7]. Using a statistically correct extrapolation method [13, 7] is an 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 time curves of dependence on the duration of load or drying, on the age at the start of loading or drying, on the environmental humidity and on the 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 the scatter due to 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, it is suitable for extrapolation of short-time data; see Fig. 8 for a few examples of such fits with model B3 [7, 4].

Conclusions

1. The highly nonuniform data distribution in the database is a result of human choice. It introduces bias, which must be suppressed. This can be accomplished by weighting.
2. Although the precise weighting to use is debatable, 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 the present statistics of the least bias.
3. The nonstandard statistical indicators considered here are not valid approaches for comparing the accuracy of prediction models. They do not yield the estimates of maximum likelihood, conflict with the principles of least-square regression and are equivalent to denying the central limit theorem of the theory of probability.
4. Therefore the previous rankings of various prediction models obtained by these nonstandard indicators cannot be taken seriously.
5. In all the comparisons based on standard statistical indicators, model B3 comes out as the best, except for one case where it is one of two equal best.

Acknowledgment. *Financial support from the U.S. Department of Transportation through the Infrastructure Technology Institute of Northwestern University, under grant 0740-357-A210, is gratefully appreciated.*

Appendix: Why Is the Method of Least Squares the Only Correct Approach?

The method of least squares was first published by Legendre in 1806 [44] but its rigorous derivation is due to Gauss [10] (who is known to have used it already before 1806). For brevity, let us again replace the double indices ij by a single index $k = 1, 2, \dots, K$ where $K = \sum_{i=1}^N m_i$. The errors are defined as $\epsilon_k = y_k - F(X_k)$ where $X_k =$ coordinates of data points, $F(X_k) = Y_k =$ predicted values, and function F defines the prediction model. The joint probability density distribution of all the measured data, also called the likelihood function \mathcal{L} [11], is:

$$\mathcal{L} = f(y_1, y_2, \dots, y_K) = [\phi(y_1)]^{w_1} [\phi(y_2)]^{w_2} \cdots [\phi(y_K)]^{w_K} \quad (17)$$

where $\phi(y_m) =$ probability density distribution of measurement y_m alone, and exponent w_m means that we imagine a w_m -fold repetition of the m -th measurement, which is equivalent to applying weight w_m to data point m . Let us first assume the errors are approximately normally distributed; then

$$\phi(y_k) = \frac{1}{s\sqrt{2\pi}} e^{-[y_k - F(X_k)]^2 / 2s^2} \quad (k = 1, 2, \dots, K) \quad (18)$$

where s = standard deviation. The objective of optimal data fitting is to maximize the likelihood function \mathcal{L} [11]. This is equivalent to minimizing $-\ln \mathcal{L}$, i.e.,

$$-\ln \mathcal{L} = -\ln \left[(s\sqrt{2\pi})^{-\sum_k w_k} e^{-\sum_k w_k (y_k - Y_k)^2 / 2s^2} \right] = C_1 \sum_{k=1}^M w_k (y_k - Y_k)^2 + C_2 = \min \quad (19)$$

where $C_1 = 1/2s^2$ and $C_2 = \ln(s\sqrt{2\pi})$ are constant, and normalized weights, i.e. $\sum_k w_k = 1$, are assumed. Eq. (19) represents minimization of the sum of squared errors, and thus proves validity of the method of least squares.

The histograms of data plotted on the normal probability paper demonstrate that the distributions or errors ϵ_{ij} in creep and shrinkage are approximately normal. But small deviations from normality exist. So what if the distributions $\phi(y_k)$ of data y_k are not normal?

In that case we may subdivide the data set into groups labelled as $r = 1, 2, \dots, N_g$, each group r containing a sufficient but not excessive number n_r of closely located adjacent points ($n_r = 6$ is reasonable). The variance s_R^2 of the mean of each data group is the variance of a sum of random variables. According to the *central limit theorem* of the theory of probability [11, 12], the distribution of this sum, and thus the group mean, converges to the normal distribution (albeit one with a scaled standard deviation). The convergence is so rapid that, for $n_r = 6$, the distribution is virtually indistinguishable from normal (indeed, a sum of 6 rectangularly distributed random variables is nowadays generally used for efficient computation of the normal distribution in Monte Carlo simulations).

So we may now reinterpret Eq. (17) as a weighted product of distributions of all the group means, and the rest of calculation up to Eq. (19) remains formally the same. The conclusion is that method of least squares is what must be used to fit the means of the data groups. Finally, since each mean minimizes the sum of squares of its differences from the individual data points, it follows that the method of least squares maximizes the likelihood function \mathcal{L} even for a (sufficiently large) set of data whose distributions are not normal.

Hence, regardless of the type of distribution of data, the method of least squares is generally the only correct approach to statistics. The reason, to recapitulate, is that the maximum likelihood fit of the all the data in the database is virtually the same as the fit of the means of all small groups of closely located adjacent points, and that, according to the central limit theorem, these means have a normal distribution regardless of the individual data distribution.

References

- [1] ACI Committee 209 (1972). "Prediction of creep, shrinkage, and temperature effects in concrete structure." ACI-SP-194; reapproved in 1982 (ACI-SP-76) and 1992 (ACI-209 T92).
- [2] Bažant, Z.P., and Panula, L. (1978–79). "Practical prediction of time-dependent deformations of concrete." *Materials and Structures* Vol.11, pp. 307–328, 415–434, Vol. 12, 169–183.
- [3] *CEB-FIP Model Code 1990. Model Code for Concrete Structures*. Thomas Telford Services Ltd., London, Great Britain; also published by Comité euro-international du béton (CEB), Bulletins d'Information No. 213 and 214, Lausanne, Switzerland.
- [4] Bažant, Z.P., Kim, Joong-Koo, and Panula, L. "Improved prediction model for time-dependent deformations of concrete." *Materials and Structures* (RILEM, Paris) Vol. 24 (1991), pp. 327–345, 409–421, Vol. 25 (1992), pp. 21–28, 84–94, 163–169.

- [5] Sakata, K. (1993). “Prediction of concrete creep and shrinkage.” *Creep and Shrinkage of Concrete* (Proc., 5th Intern. RILEM Symp., Barcelona 1993), Z.P. Bažant and I. Carol, eds., E&F Spon, London, pp. 649–654.
- [6] Gardner, N.J. and Zhao, J.W. (1993). “Creep and shrinkage revisited.” *ACI Materials Journal* 90, 236–246. Discussion by Bazant and Baweja, (1994) *ACI Materials Journal* 91, 204–216.
- [7] Bažant, Z.P., and Baweja, S. (2000). “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, A. Al-Manaseer, ed., pp. 1–83 (update of RILEM Recommendation published in *Materials and Structures* Vol. 28, 1995, pp. 357–365, 415–430, and 488–495).
- [8] Sakata, K., Tsubaki, T., Inoue, S., and Ayano, T. (2001). “Prediction equations of creep and drying shrinkage for wide-ranged strength concrete”. *Creep, Shrinkage and Durability Mechanics of Concrete and Other Quasi-Brittle Materials* (Proc., 6th Intern. Conf., CONCREEP-6, Cambridge, MA), F.-J. Ulm, Z.P. Bažant and F.H. Wittmann, eds., Elsevier, Amsterdam, 753–758.
- [9] Gardner N.J. and Lockman M.J. (2001). “Design provisions for drying shrinkage and creep of normal strength.” *ACI Materials Journal* 98 (2), Mar.-Apr., 159–167.
- [10] Gauss, K.F. (1809). *Theoria Motus Corporum Caelestium*. Hamburg (reprinted by Dover Publ., New York 1963).
- [11] Bulmer, M.G. (1979). *Principles of Statistics*. Dover Publ., New York (chapter 12).
- [12] Bouchaud, J.-P., and Potters, M. (2000). *Theory of financial risks: From statistical physics to risk management*. Cambridge University Press, Cambridge, U.K.
- [13] Bažant, Z.P. and Baweja, S. (1995). “Creep and shrinkage prediction model for. analysis and design of concrete structures: ModelB3.” *Materials and Structures* 28, 357–367.
- [14] Müller, H.S., and Hilsdorf, H.K. (1990). “Evaluation of the time-dependent behaviour of concrete: summary report ont the work of the General Task Force Group No. 199.” CEB (Copmité euro-internationale du béton), Lausanne (201 pp.).
- [15] FIB (1999). *Structural Concrete: Textbook on Behaviour, Design and Performance, Updated Knowledge of the of the CEB/FIP Model Code 1990*. Bulletin No. 2, Fédération internationale de béton (FIB), Lausanne, Vol. 1, pp. 35–52.
- [16] Bažant, Z.P. (1975). “Theory of creep and shrinkage in concrete structures: A precis of recent developments”, *Mechanics Today*, ed. by S. Nemat-Nasser (Am. Acad. Mech.), Pergamon Press 1975, Vol. 2, pp. 1–93.
- [17] Bažant, Z.P. (1982). “Mathematical models of nonlinear behavior and fracture of concrete,” in *Nonlinear Numerical Analysis of Reinforced Concrete*, ed. by L. E. Schwer, Am. Soc. of Mech. Engrs., New York, 1–25.
- [18] RILEM Committee TC-69 (1988). (Z.P. Bažant, Chairman and princ. author), “State of the art in mathematical modeling of creep and shrinkage of concrete,” in *Mathematical Modeling of Creep and Shrinkage of Concrete*, ed. by Z.P. Bažant, J. Wiley, Chichester and New York, 1988, 57–215.
- [19] Bažant, Z.P. (2000) “Criteria for rational prediction of creep and shrinkage of concrete.” *Adam Neville Symposium: Creep and Shrinkage—Structural Design Effects*, ACI SP-194, A. Al-Manaseer, ed., pp. 237-260 (update of an article from *Revue Française de Génie Civil* 3 (3–4), 1999, pp. 61–89).
- [20] “Conclusions of the Hubert Rüsç Workshop on Creep of Concrete”. *ACI Concrete International* 2 (11), p. 77.

- [21] Müller, H.S., Bažant, Z.P., and Kuttner, C.H. (1999). “Data base on creep and shrinkage tests.” *RILEM Subcommittee 5 report RILEM TC 107-CSP*, RILEM, Paris (81 pp.)
- [22] Müller, H.S. (1993). “Considerations on the development of a database on creep and shrinkage tests.” *Creep and Shrinkage of Concrete* (Proc., 5th Intern. RILEM Symp., Barcelona 1993), Z.P. Bažant and I. Carol, eds., E&F Spon, London, 859–872.
- [23] Bažant, Z.P., and Li, G.-H. (2008). “Database on Concrete Creep and Shrinkage.” Infrastructure Technology Institute (ITI), Northwestern University, Report (in preparation, to be posted on ITI website).
- [24] Gardner, N.J. (2000). “Design provisions fo shrinkage and creep of concrete.” *Adam Neville Symposium: Creep and Shrinkage—Structural Design Effects*, ACI SP-194, A. Al-Manaseer, ed., pp.101-104.
- [25] Gardner, N.J. (2004). “Comparison of prediction provisions for drying shrinkage and creep of normal strength concretes.” *Canadian Jour. of Civil Engrg.* 31 (5), Sep.-Oct., 767-775.
- [26] McDonald, D.B., and Roper, H. (1993). “Accuracy of prediction models for shrinkage of concrete.” *ACI Materials Journal* 90 (3), May-June, 265–271.
- [27] Al-Manaseer, A., and Lakshmikanthan, S. (1999). “Comparison between current ant future design code models for creep and shrinkage”. *Revue française de génie civil* 3 (3–4), 39–40.
- [28] Al-Manaseer, A., and Lam, J.-P. (2005). “Statistical evaluation of creep and shrinkage models.” *ACI Materials Journal* 102 (May-June), 170–176.
- [29] Ang, A.H.-S., and Tang, W.H. (1984). *Probability concepts in engineering planning and design. Vol II. Decision, Risk and Reliability*. J. Wiley, New York.
- [30] Křístek, V., Bažant, Z.P., Zich, M., and Kohoutková (2006). “Box girder deflections: Why is the initial trend deceptive?” *ACI Concrete International* 28 (1), 55–63. ACI SP-194, 237–260.
- [31] Ang, A. H.-S., and Tang, W.H. (1976). *Probability concepts in engineering planning and design. Vol. 1, Sec.7*, J. Wiley, New York.
- [32] Bažant, Z.P., and Kim, Joong-Koo (1991). “Consequences of diffusion theory for shrinkage of concrete.” *Materials and Structures* (RILEM, Paris) 24 (143), 323–326.
- [33] Bažant, Z.P., and Raftshol, W. J. (1982). “Effect of cracking in drying and shrinkage specimens.” *Cement and Concrete Research*, 12, 209–226.
- [34] A. Haldar and S. Mahadevan (2000). *Probability, reliability and statistical methods in engineering design*. J. Wiley, New York
- [35] Beck, J.V., and Arnold, K.J. (1977). *Parameter estimation in engineering science*. J. Wiley, New York.
- [36] Draper, N., and Smith, F. (1981). *Applied regression analysis*. 2nd ed. J. Wiley, New York.
- [37] Fox, J. (1997). *Applied regression analysis, linear models and related methods*. Sage Publications. (also: <http://socser.socsci.mcmaster.ca/jfox/Books/Applied-Regression/in...>).
- [38] Lehmann, E.L. (1959). *Testing statistical hypotheses*. J. Wiley, New York.
- [39] Mandel, J. (1984). *The statistical analysis of experimental data*. Dover Publications.
- [40] Plackett, R.L. (1960). *Principles of regression analysis*. Clarendon Press.
- [41] Crow, E.L., Davis, F.A., and Maxfield, M.W. (1960). *Statistics Manual*. Dover Publ., Neew York.

[42] Benjamin, J.R., and Cornell, C.A. (1970). *Probability, Statistics and Decision for Civil Engineers*. McGraw-Hill, New York.

[43] Bažant, Z.P. (2004). Discussion of “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* 101 (Feb.), 139-140.

[44] Legendre, A.M. (1806). *Novelle méthodes pour la détermination des orbites des comètes*, Paris.

List of Tables

1	Standard coefficients of variation of errors of various prediction models in a) compliance, b) shrinkage, c) relative compliance, and d) relative shrinkage (cubes are in $\log(t - t')$, $\log t'$ and H for compliance or \sqrt{D} for shrinkage).	17
2	Comparison of standard and nonstandard nonstandard statistical indicators of errors used by various authors for various prediction models, for a) compliance, and b) shrinkage.	17

Table 1: Standard coefficients of variation of errors of various prediction models in a) compliance, b) shrinkage, c) relative compliance, and d) relative shrinkage (cubes are in $\log(t - t')$, $\log t'$ and H for compliance or \sqrt{D} for 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
5 intervals, $\log(t - t')$	26.2	41.9	29.7	28.5	43.8	26.4	66.0	33.0	29.8	32.9
4 intervals, $\log t'$	27.4	37.1	29.9	28.8	48.2	26.9	74.3	33.3	30.5	33.0
7 intervals, \sqrt{D}	23.3	36.9	27.3	23.3	33.2	20.1	55.9	24.4	21.9	22.6
10 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
4 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
4 intervals, $\log t_0$	42.8	48.6	56.0	53.9	64.2	44.9	52.8	57.6	54.0	64.7
7 intervals, \sqrt{D}	27.2	37.3	49.2	29.1	38.9	33.7	46.4	45.0	39.9	52.9
10 intervals, H	38.4	52.0	46.9	54.4	46.6	41.6	55.6	43.0	41.9	45.6

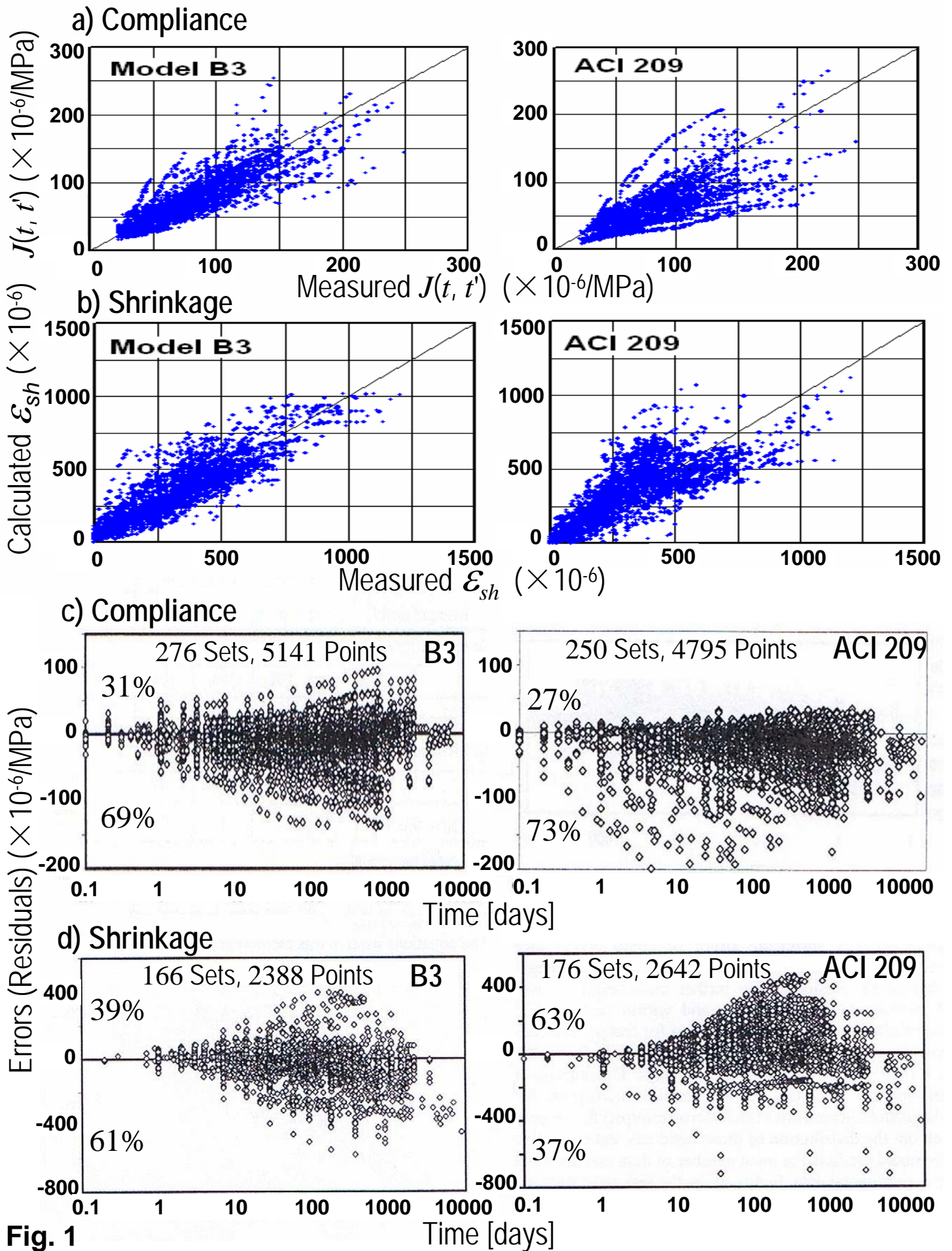
Table 2: Comparison of standard and nonstandard nonstandard statistical indicators of errors used by various authors for various prediction models, for a) compliance, and b) shrinkage.

	a) Compliance (%)						b) Shrinkage (%)				
	Indicator	ACI	B3	CEB	GL		Indicator	ACI	B3	CEB	GL
Bažant [7] basic creep	ω	58	24	35	-	Bažant [7]	ω	55	34	46	-
Bažant [7] drying creep	ω	45	23	32	-	Al-Manaseer [28]	ω_{CEB}	46	41	52	37
Al-Manaseer [28]	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		ω_{BP}	102	55	90	46
	ω	87	61	75	47	Gardner [25]	ω_G	41	20	-	19
Gardner [25]	ω_G	30	27	-	22	Gardner [25] recalculated	ω_G	41.6	20.5	43.8	22.2

List of Figures

- 1 Examples of ineffectual statistical comparisons of prediction models for compliance (a,c) and shrinkage (b,d). 19
- 2 Histograms of data points and of test curves in the NU-ITI database. 19
- 3 Sketches explaining: a,b) subdivision of database variables into one-dimensional intervals and two-dimensional boxes of equal importance; c,d) differences in time evolution of scatter between actual and relative data; e,f) Difference between ensemble and regression statistics. 19
- 4 Coefficients of variation of errors (a,b,c,d) and correlation coefficients (e,f) of prediction models for actual data (a,b,e,f) and relative data (c,d), for NU-ITI database. 19

5	Bands of interval centroids \pm standard deviation for actual data (solid lines) and predicted values (dashed lines) for compliance (a), shrinkage (b), relative compliance (c) and relative shrinkage (d), for various prediction models.	19
6	Comparison (in %) of standard and nonstandard statistical indicators for various prediction models; 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.	19
7	a) Differences in coefficients of variation of errors between standard and nonstandard statistical methods for examples of linear regression.	19
8	Fits of characteristic long-time compliance and shrinkage data by the formulae of model B3.	19
9	Cumulative histograms of errors of B3 model compared to NU-ITI database, plotted on normal probability paper; left: compliance, right: shrinkage, top: unweighted, bottom: weighted.	19



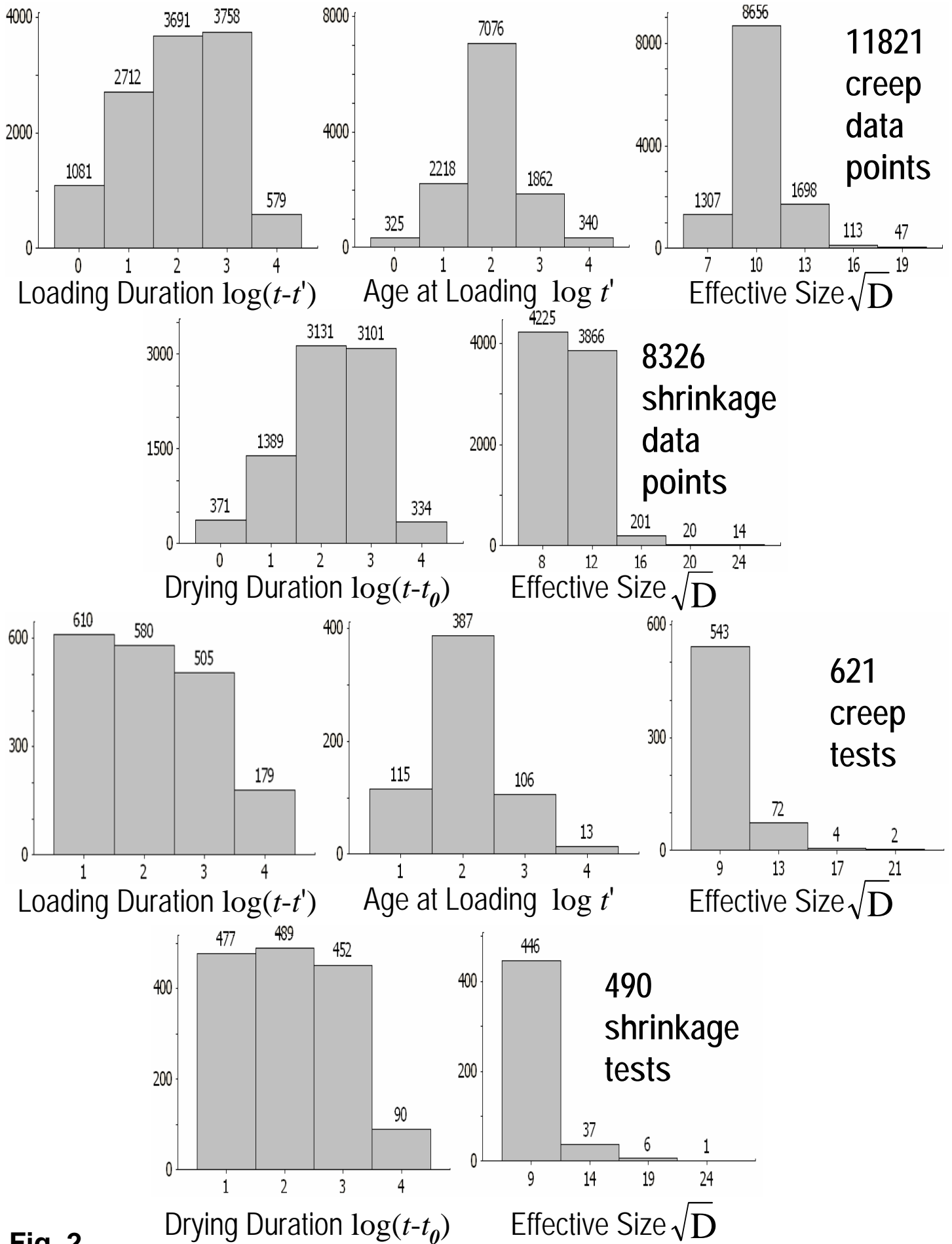
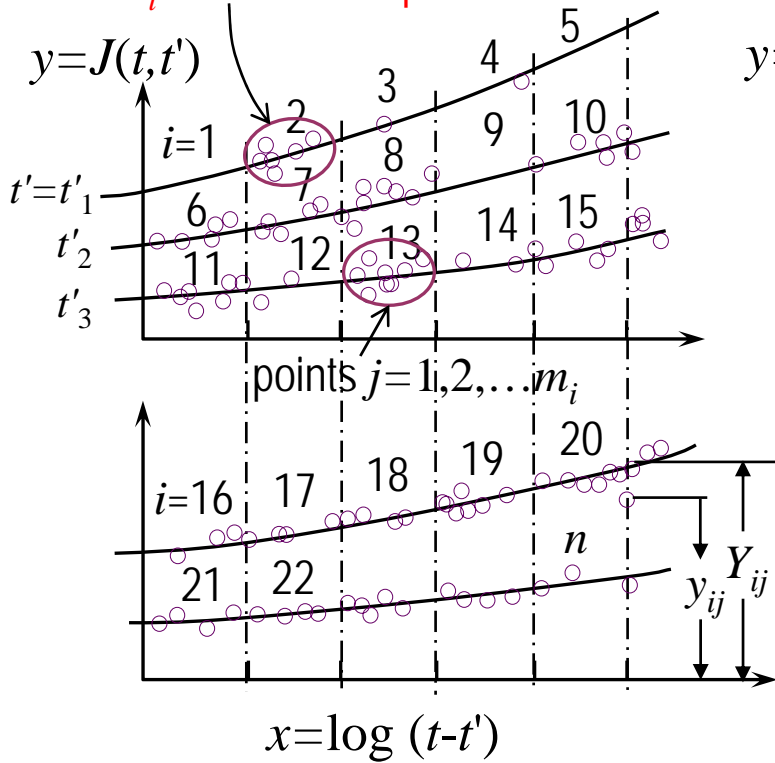


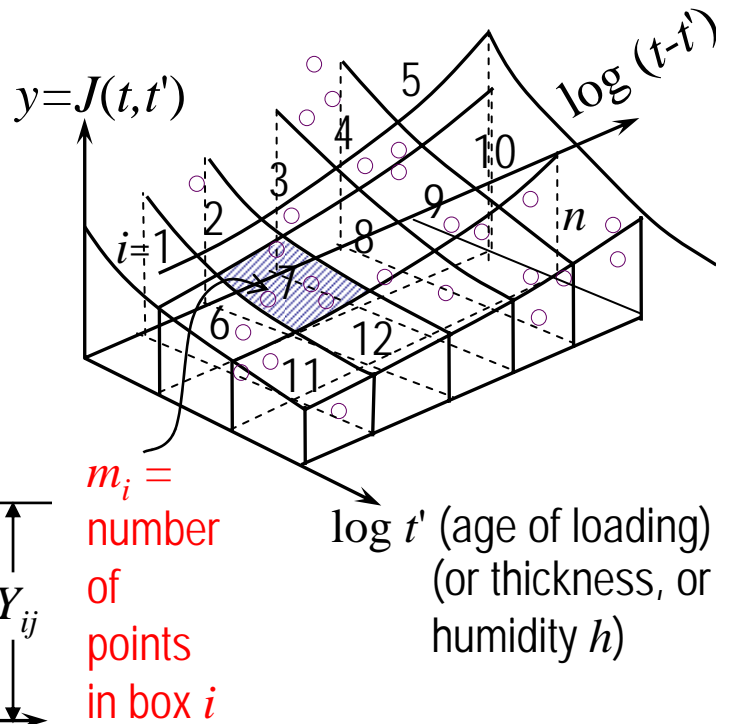
Fig. 2

a) 1D Boxes (Intervals)

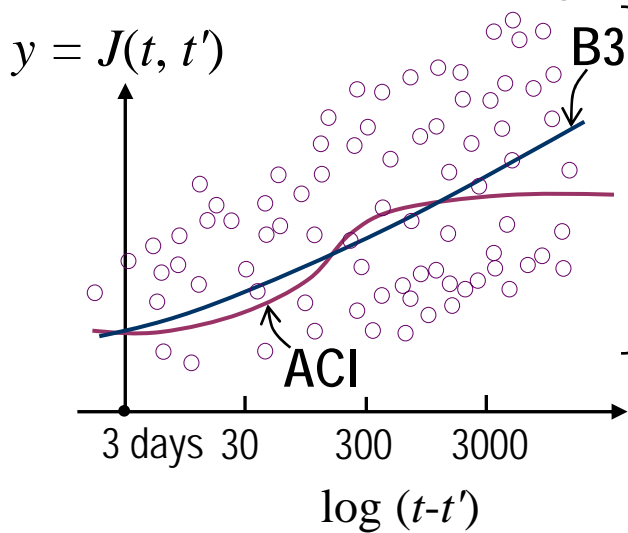
$m_i =$ number of points in interval i



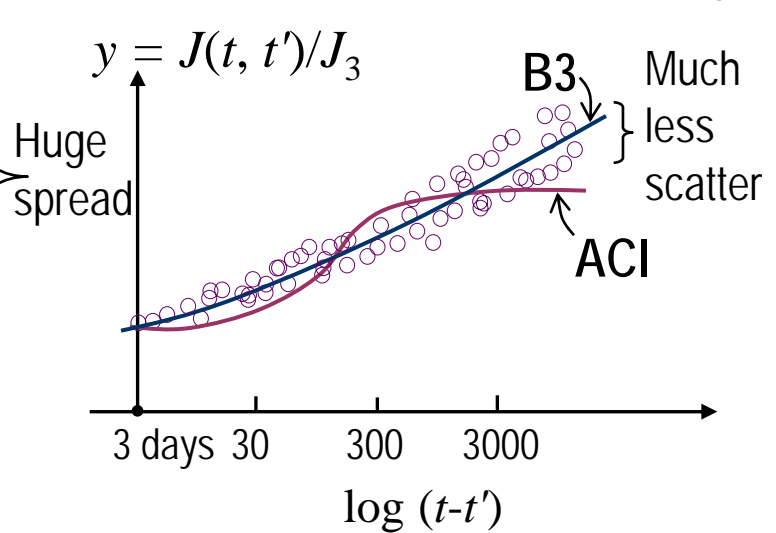
b) 2D Boxes



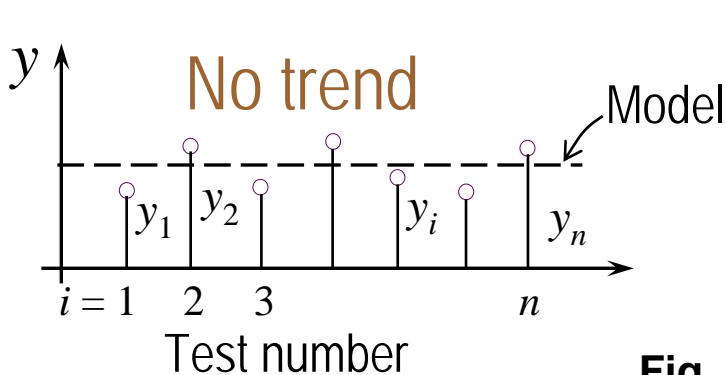
c) Compliance or Shrinkage



d) Relative Compliance or Shrinkage



e) Ensemble statistics



f) Least-Square Regression

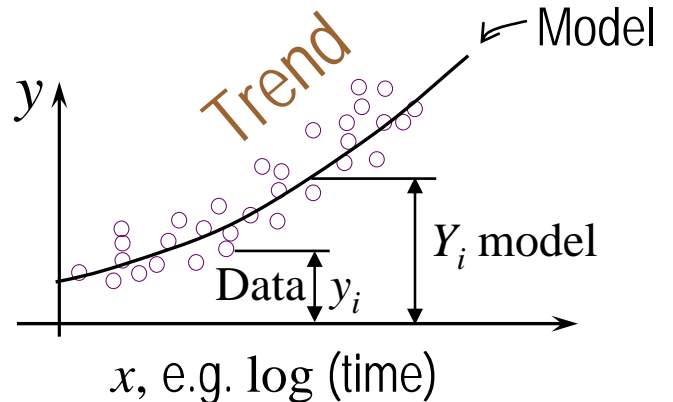


Fig. 3

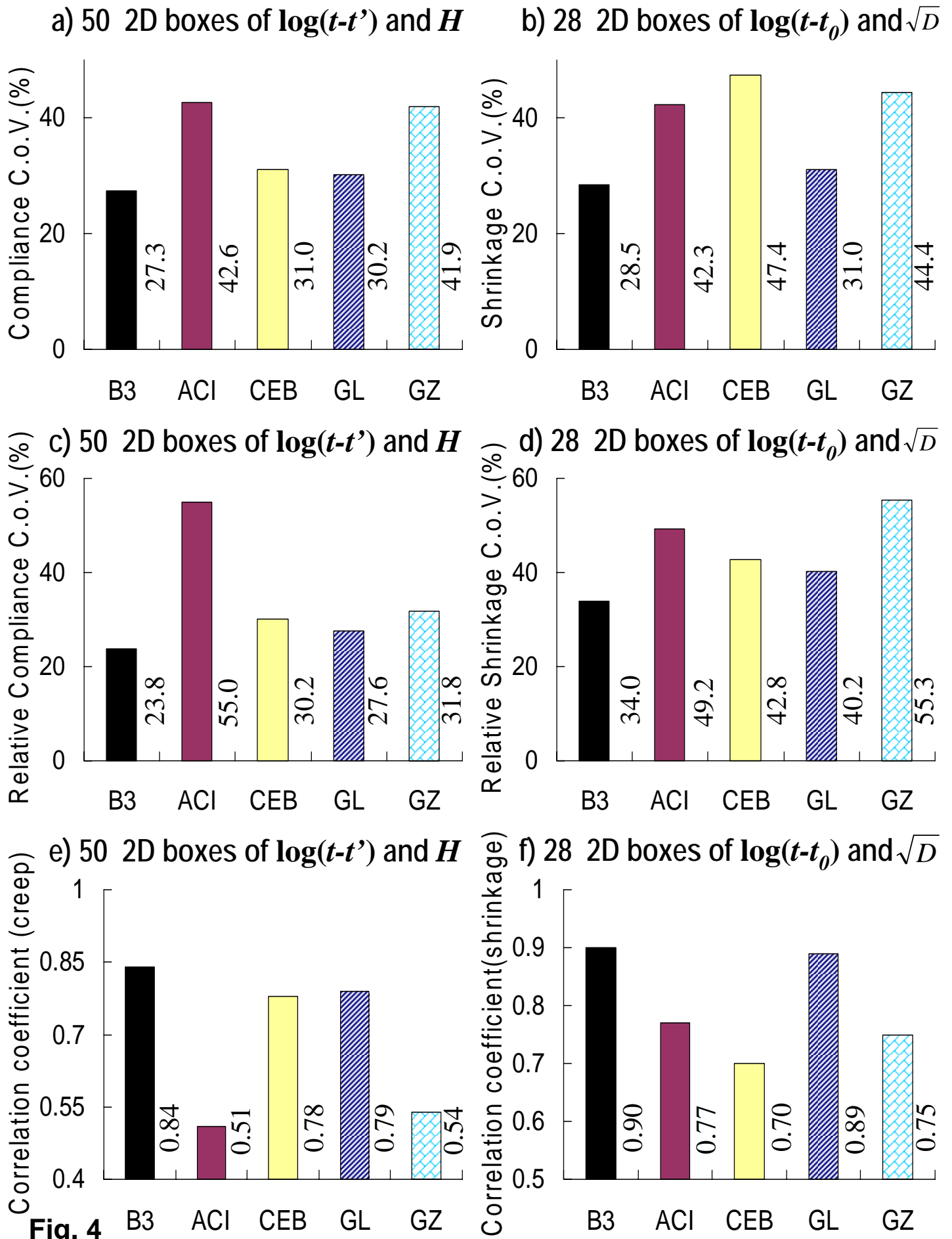


Fig. 4

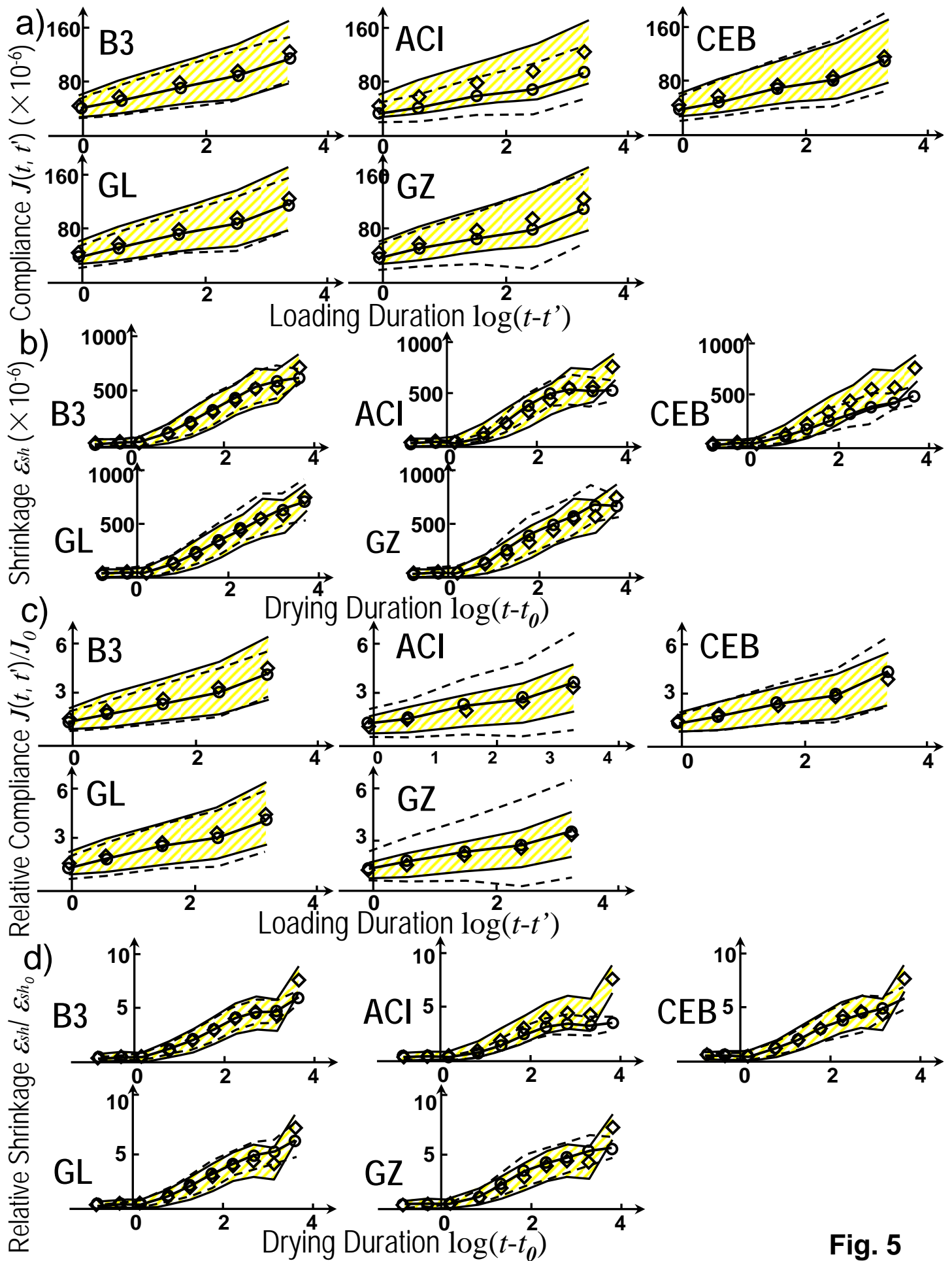


Fig. 5

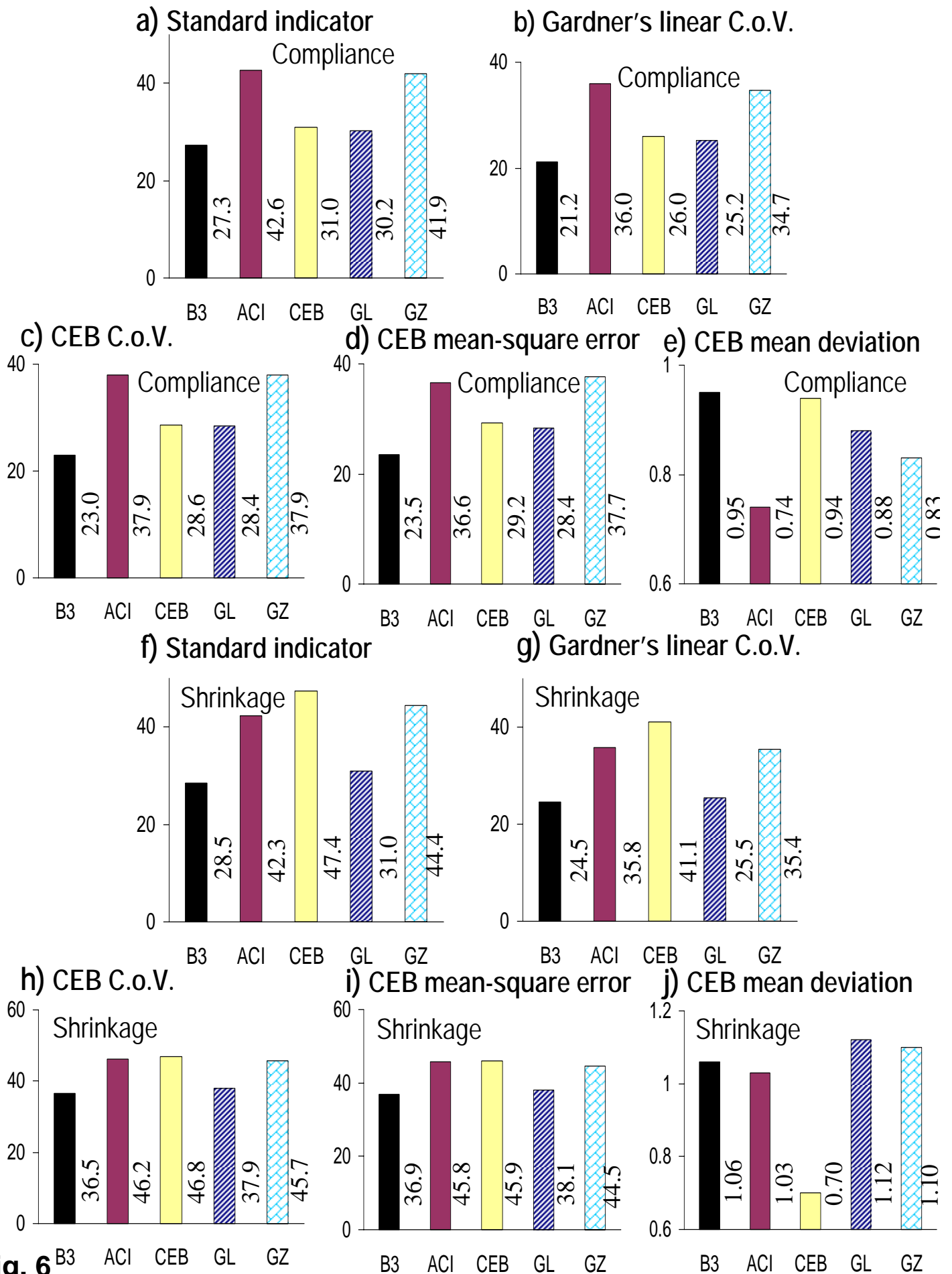


Fig. 6

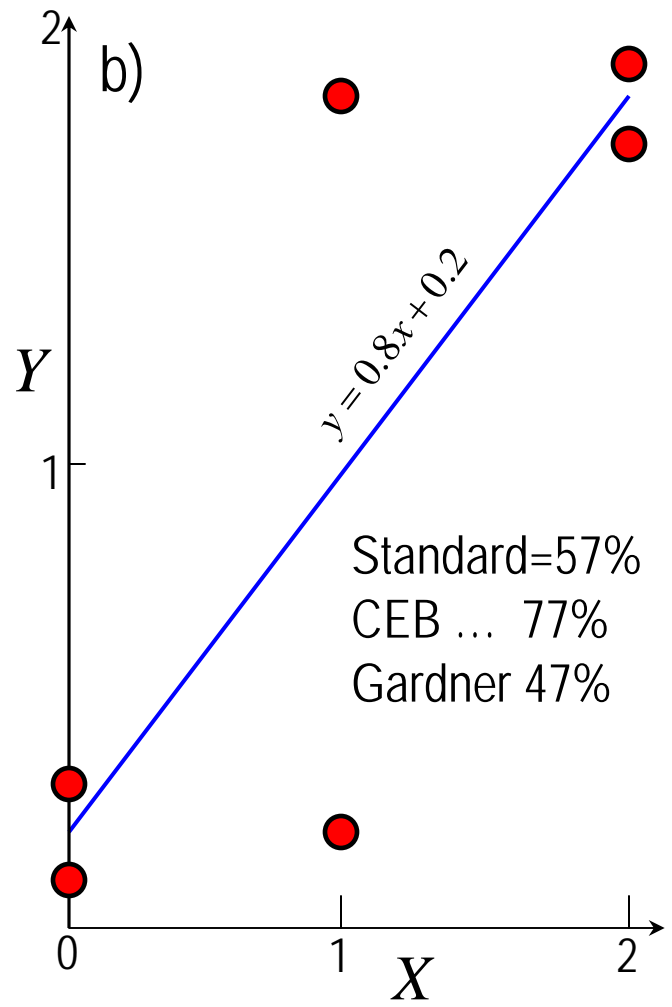
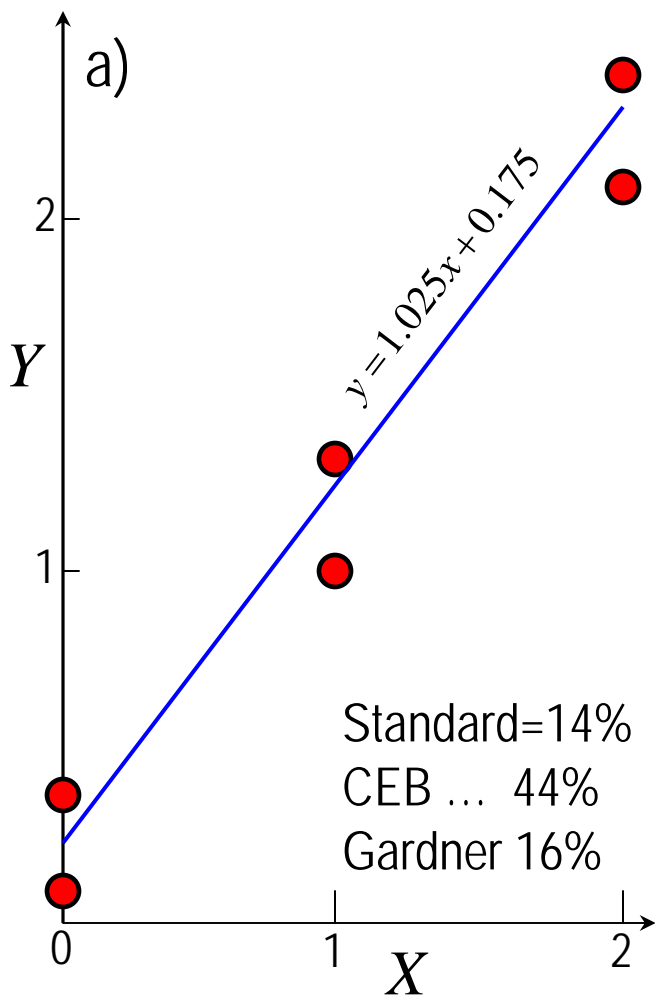


Fig. 7

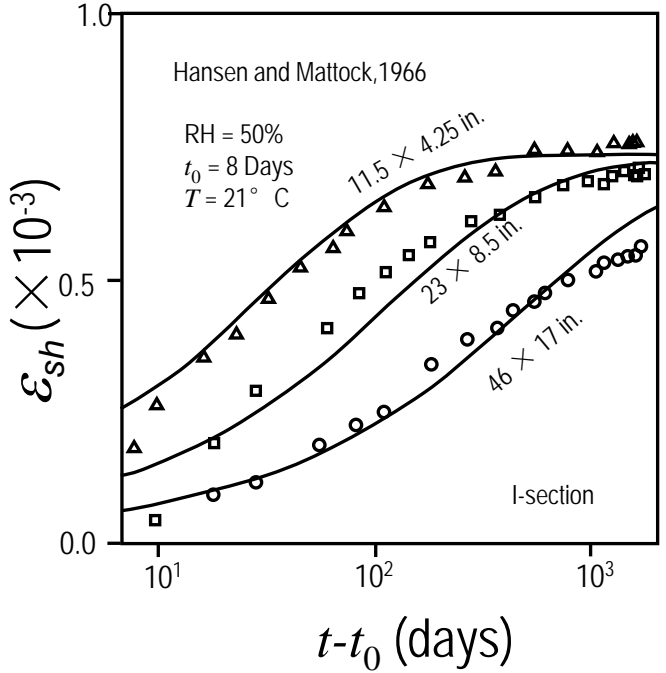
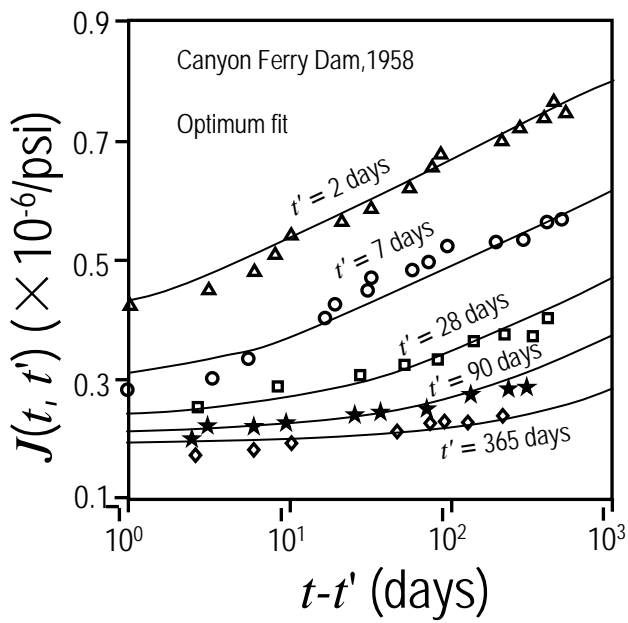
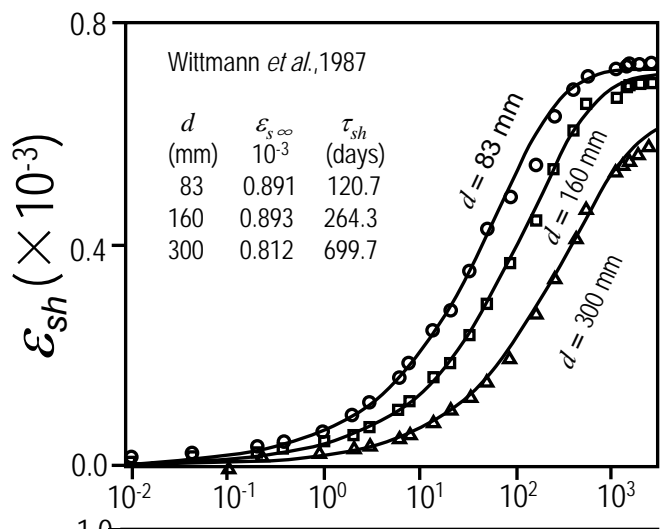
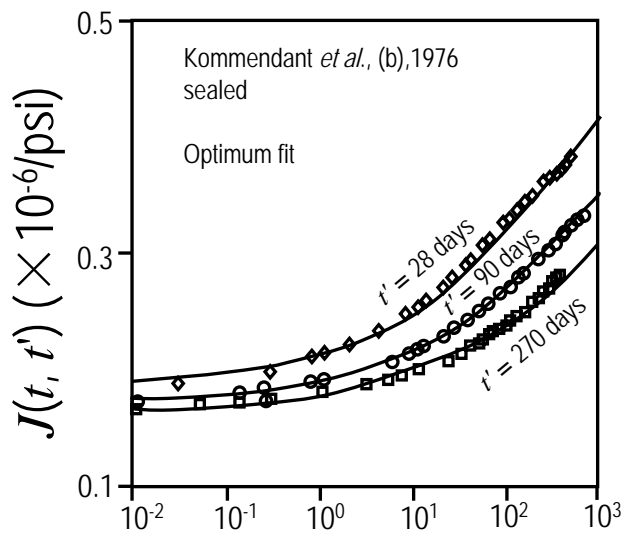
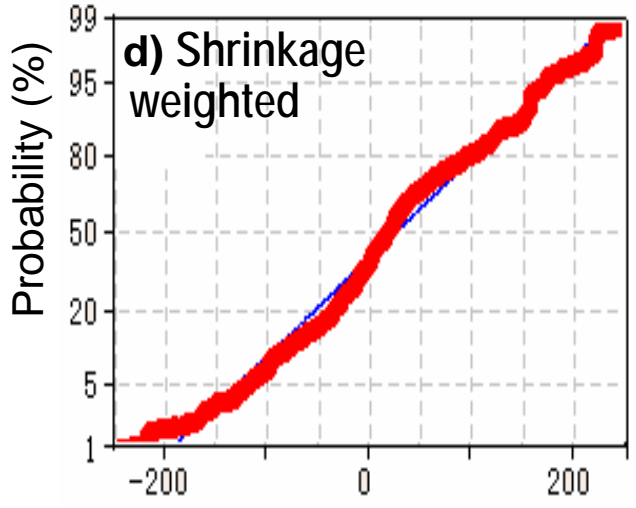
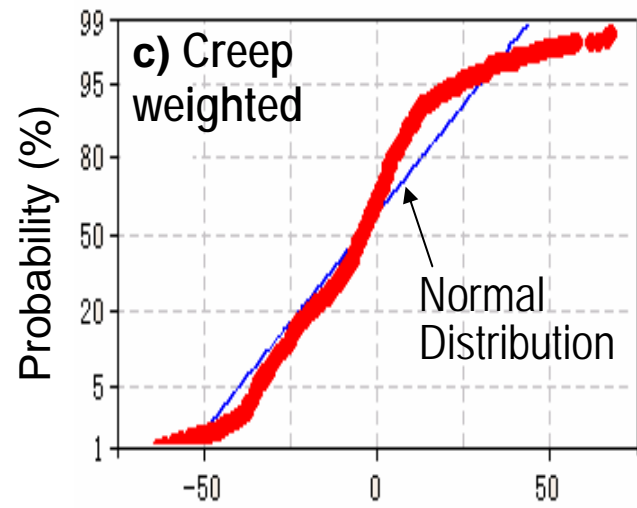
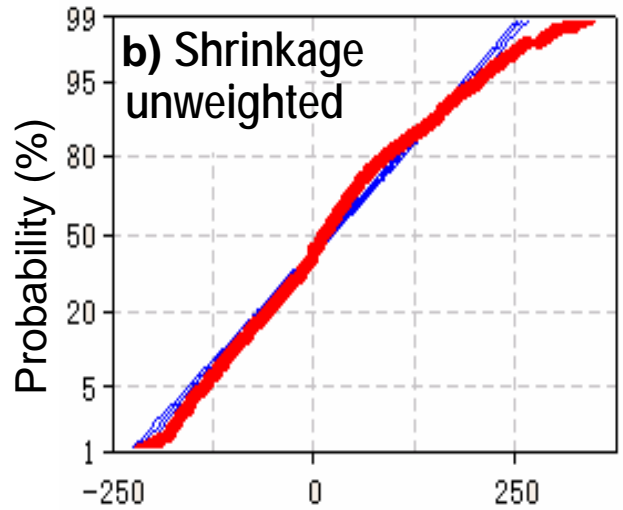
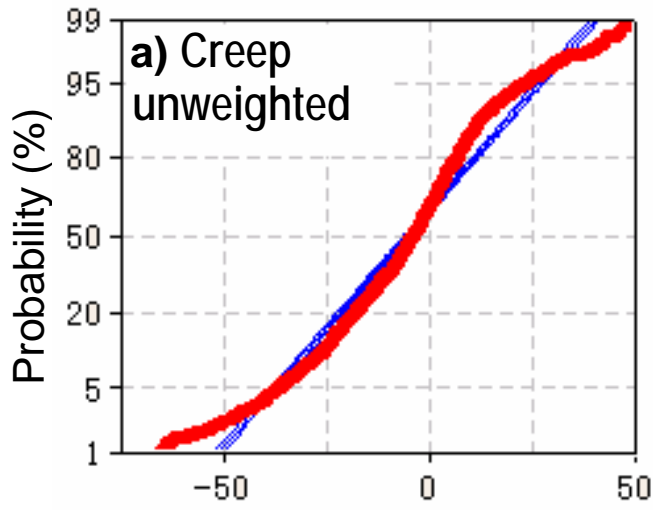


Fig. 8



Error (normal probability paper)

Fig. 9