

Model Uncertainty, Study-Grouped Validation and Reliability-Based Resistance Factors for the Ultimate Conditions of FRP-Confined Concrete

Aggoun Mahfoudh ^{1*}, Khan Muhammad Danial ²

¹ School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

² School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

*Corresponding Authors: Email : (202422040437@std.uestc.edu.cn.)

Received: 16 December 2025

Accepted: 09 January 2026

Available online: 10 February 2026

Abstract: Design-oriented equations for the ultimate strength and strain of fibre-reinforced polymer (FRP) confined concrete are usually compared through error statistics obtained from a single random split of a test database. Such statistics say little about the safety margin that a designer needs, and they ignore the fact that specimens from the same laboratory are not independent. This study re-analysed a database of 780 circular FRP-confined concrete specimens (530 carbon, 162 glass and 88 aramid FRP, 63 test programmes) compiled for a recent hybrid machine-learning study. Eight strength models and seven strain models were treated as resistance models and their model uncertainty θ = experiment/prediction was characterised. For the strength models the coefficient of variation of θ was 0.177–0.251, while for the strain models it was 0.36–0.82, an order-of-magnitude difference in the information available for design. A random-effects decomposition showed that 53–67% (strength) and 49–62% (strain) of the variance of $\ln \theta$ lies between test programmes rather than within them. Gradient-boosting models were then validated both with random 10-fold cross-validation and with study-grouped cross-validation. The random split reproduced the large accuracy gains reported in the literature (RMSE 0.252 for the strength ratio), but under study-grouped validation the same model gave 0.366, which is no better than the best closed-form equation (0.349); for the strain ratio the machine-learning model was worse than the closed-form equation. Grouped cross-conformal intervals achieved their nominal 90% coverage on average but failed for individual test programmes. Finally, resistance factors were calibrated with EN 1990 Annex D and with the first-order reliability method for an axially loaded confined column. A factor $\phi \approx 0.53$ is needed to reach $\beta_T = 4.0$ with the best strength model, whereas $\phi = 0.65$ gives $\beta \approx 3.18$. Data-driven refinement of the mean prediction does not remove the scatter that controls design, and reporting model uncertainty with study-grouped validation should become standard practice.

Keywords: FRP-confined concrete; model uncertainty; reliability-based calibration; gradient boosting; conformal prediction

1. Introduction

Wrapping a concrete column with a fibre-reinforced polymer (FRP) jacket increases both its compressive strength and its deformation capacity, and it has become a standard retrofit technique for columns that lack confinement (Lam and Teng, 2003; Ozbakkaloglu and Lim, 2013). Design guidance such as ACI 440.2R-17 (ACI Committee 440, 2017) predicts the confined strength f'_{cc} and the ultimate axial strain ϵ_{cu} with closed-form, design-oriented equations calibrated on databases of circular specimens. Over the last two decades many such equations have been proposed, and more recently the same databases have been used to train machine-learning (ML) models: support vector regression coupled with metaheuristic optimisers, neuro-fuzzy systems, genetic programming and tree ensembles (Lim et al., 2016; Mansouri et al., 2017; Ben Seghier et al., 2021).

These studies are almost always reported in the same way. The database is split at random into a training and a test set, a few error statistics (RMSE, MAE, agreement index, Nash-Sutcliffe efficiency) are computed for the new model and for a list of existing equations, and the new model is declared superior. Two questions that

matter for design are left open. First, how large is the scatter of the prediction error, and what resistance factor does it imply? Design codes are calibrated on the *model uncertainty* θ , defined as the ratio between the measured and the predicted resistance, and it is the coefficient of variation (CoV) of θ , not the mean error, that controls the resistance factor (Ravindra and Galambos, 1978; Nowak and Collins, 2013). Second, is the reported accuracy transferable to a new laboratory? Specimens in a confinement database are produced in batches: one test programme contributes 5–40 nominally identical cylinders that share a concrete batch, a jacket material, an instrumentation protocol and an operator. A random split places siblings of almost every test specimen in the training set, so the reported test error measures interpolation within a programme rather than prediction for a new one. In ecology and in the geosciences this problem is well known, and grouped (block) cross-validation is the accepted remedy (Roberts et al., 2017); in structural-engineering model comparisons it is rarely applied.

Reliability-oriented studies of FRP confinement do exist. Keshtegar et al. (2020) fitted distributions to the model error of several confinement models and

computed reliability indices with an improved first-order reliability method (FORM) algorithm; Chen et al. (2022) calibrated stress-strain models for square members probabilistically; and Shayanfar et al. (2026) performed a reliability-based partial-factor calibration for a new strength model. None of these studies addresses the dependence structure of the databases they use, and none compares the model uncertainty of closed-form equations with that of an ML model evaluated honestly.

This paper takes the database and the model set of a recent hybrid-intelligence study (Ben Seghier et al., 2021), in which support vector regression was coupled with genetic, particle-swarm and whale optimisation algorithms, and asks a different question: what do these models mean for design? The contributions are (i) a full characterisation of the model uncertainty of eight strength and seven strain models on 780 specimens, including fibre-specific statistics, distribution fits and residual trends; (ii) a variance decomposition that separates between-programme from within-programme scatter; (iii) a comparison of random and study-grouped cross-validation for gradient-boosting models, including a physics-guided residual formulation; (iv) grouped

cross-conformal prediction intervals with coverage checked per fibre type and per test programme; and (v) calibration of resistance factors with EN 1990 Annex D and with FORM for an axially loaded confined column. All computations are reproducible from the scripts and data released with this paper.

2. Database

The database contains 780 circular FRP-confined concrete specimens tested in axial compression and assembled from the open literature; it is the database used by Ben Seghier et al. (2021), which in turn draws on the compilations of Sadeghian and Fam (2015) and Keshtegar et al. (2017a). Each record gives the specimen diameter D , the jacket thickness t , the elastic modulus E_f and tensile strength f_{tu} of the FRP, the unconfined concrete strength f'_{co} , the confined strength f'_{cc} , the ultimate axial strain ε_{cc} and the hoop rupture strain $\varepsilon_{h,rupt}$ measured at failure, together with the source publication. The specimens come from 63 distinct test programmes: 530 carbon (CFRP), 162 glass (GFRP) and 88 aramid (AFRP) specimens. Table 1 summarises the ranges. The database spans normal- to ultra-high-strength concrete ($f'_{co} = 16.6$ -188.2 MPa) and jacket stiffnesses that differ by two orders of magnitude.

Table 1. Range and statistics of the FRP-confined concrete database by fibre type

Subset	n	Test programmes	D (mm)	f'_{co} (MPa)	E_t (GPa)	$\varepsilon_{h,rupt}$ (%)	Mean f_t/f'_{co}	Mean f'_{cc}/f'_{co}	Mean $\varepsilon_{cc}/\varepsilon_{co}$
CFRP	530	51	51-406	16.6-188.2	25.0-662.5	0.11-2.26	0.307	2.06	7.34
GFRP	162	26	76-406	25.0-188.2	10.5-116.0	0.10-4.98	0.233	1.73	7.68
AFRP	88	11	100-152	29.6-121.2	13.6-128.5	0.60-3.21	0.300	1.95	10.58
All	780	63	51-406	16.6-188.2	10.5-662.5	0.10-4.98	0.291	1.98	7.78

Ranges are minimum-maximum. The strain ratios use ε_{co} from Eq. (5).

The reference strains and the dimensionless confinement ratios follow the definitions used in the source study. The lateral confining pressure at hoop rupture is

$$f_l = \frac{2E_f t \varepsilon_{h,rupt}}{D} \quad (1)$$

and the unconfined strain at peak stress is taken from the expression of Popovics (1973), as adopted by Teng et al. (2009) and in the database itself,

$$\varepsilon_{co} = 9.37 \times 10^{-4} (f'_{co})^{0.25} \quad (2)$$

with f'_{co} in MPa. Four dimensionless ratios are used by the design equations reviewed below: the confinement stiffness ratio ρ_K , the strain ratio ρ_ε , the FRP area ratio ρ_a and the lateral confinement stiffness ratio ρ_E ,

$$\rho_K = \frac{2E_f t}{D (f'_{co}/\varepsilon_{co})}, \quad \rho_\varepsilon = \frac{\varepsilon_{h,rupt}}{\varepsilon_{co}}, \quad \rho_a = \frac{t}{D}, \quad \rho_E = \frac{2E_f}{f'_{co}/\varepsilon_{co}} \quad (3)$$

so that $f_l/f'_{co} = \rho_K \rho_\varepsilon = \rho_E \rho_a$. The jacket stiffness $K_l = 2E_f t/D$ is also used by some models. In 9 specimens the measured strength ratio and in 3 specimens the measured strain ratio are smaller than unity. These records are physically questionable but they were retained, because removing unfavourable observations would bias the model-uncertainty statistics that are the object of this study; their effect is discussed in Section 6.

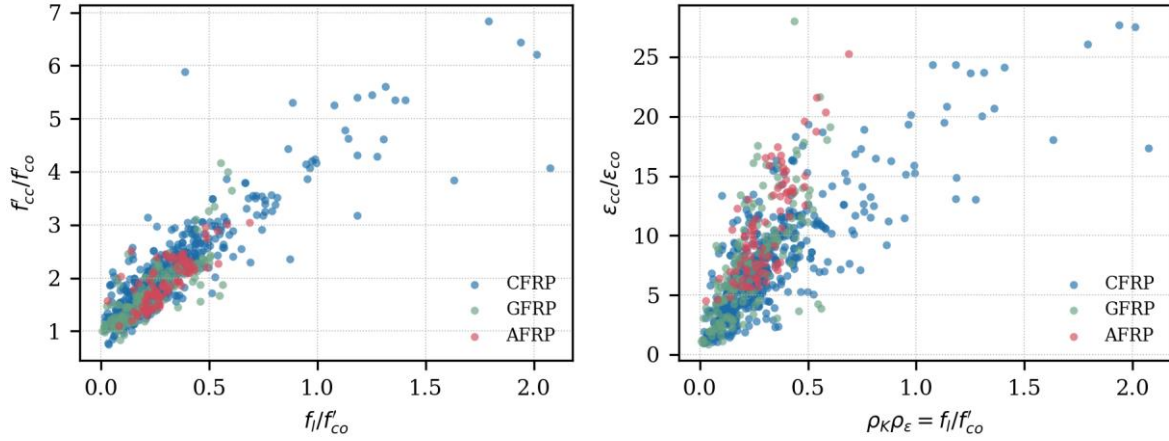


Figure 1. Ultimate conditions of the 780 specimens: (a) strength ratio and (b) strain ratio against the lateral confining ratio, by fibre type.

3. Design-oriented models

Eight design-oriented strength models and seven strain models were implemented exactly in the algebraic form in which they are collected by Ben Seghier et al. (2021); Table 2 lists them. All of them are closed-form expressions in the dimensionless ratios defined above, except the strain models of Mander et al. (1988) and

Toutanji (1999), which need the strength ratio as an input and were evaluated with the measured strength ratio, as in the source study. This gives those two models an advantage that a designer does not have; it is kept here only for comparability and is taken into account in the discussion.

Table 2. Design-oriented models for the ultimate strength and strain of FRP-confined concrete

Model	Expression
Ultimate strength ratio f'_{cc}/f'_{co}	
Newman and Newman (1971)	$1 + 3.7 (f_1/f'_{co})^{0.86}$
Toutanji (1999)	$1 + 3.5 (f_1/f'_{co})^{0.85}$
Guralnick and Gunawan (2006)	$0.616 + f_1/f'_{co} + 1.57 (f_1/f'_{co} + 0.06)^{0.5}$
Teng et al. (2009)	$1 + 3.5 (\rho_K - 0.01) \rho_e$ for $\rho_K \geq 0.01$
Ozbakkaloglu and Lim (2013)	$1 + 0.0058 K_1/f'_{co} + 3.22 (f_1 - f_{i0})/f'_{co}$, $f_{i0} = K_1(0.43 + 0.009 K_1/f'_{co}) \epsilon_{co}$
Sadeghian and Fam (2015)	$1 + (2.77 \rho_K^{0.77} - 0.07) \rho_e^{0.91}$
Keshtegar et al. I (2017a)	$1 + (4.63 + 0.93 \rho_e^{0.53} \rho_e^{1.03}) \rho_a^{0.82} \rho_e^{0.45}$
Keshtegar et al. II (2017b)	$1 + (3.23 - 4.8 \rho_K^{2.5}) (f_1/f'_{co})^{0.95}$
Ultimate strain ratio $\epsilon_{cu}/\epsilon_{co}$	
Mander et al. (1988)	$1 + 5 (f'_{cc}/f'_{co} - 1)$
Toutanji (1999)	$1 + (310.57 \epsilon_{h,rupt} + 1.9)(f'_{cc}/f'_{co} - 1)$
Teng et al. (2009)	$1.75 + 6.5 \rho_K^{0.8} \rho_e^{1.45}$
Ozbakkaloglu and Lim (2013)	$2 - (f'_{co} - 20)/100 + 0.27 (K_1/f'_{co})^{0.9} \epsilon_{h,rupt}^{1.35}/\epsilon_{co}$
Sadeghian and Fam (2015)	$1.5 + 6.78 \rho_K^{0.63} \rho_e^{1.08}$
Keshtegar et al. I (2017a)	$1.5 + (-3.45 + 5.84 \rho_e^{0.6} \rho_e^{0.69}) \rho_a^{0.57} \rho_e^{0.36}$
Keshtegar et al. II (2017b)	$1 + (7.31 + 2.06 \rho_e^{0.8}) (f_1/f'_{co})^{0.6}$

$\epsilon_{h,rupt}$ is used as a decimal fraction; $K_1 = 2E_t/D$ in MPa. The two strain models that contain f'_{cc}/f'_{co} were evaluated with the measured strength ratio.

4. Methods

4.1. Model uncertainty

Each equation is treated as a resistance model. For specimen i the model uncertainty is

$$\theta_i = \frac{r_{e,i}}{r_{t,i}} \quad (4)$$

where r_e is the measured and r_t the predicted ratio (f'_{cc}/f'_{co} or $\epsilon_{cu}/\epsilon_{co}$). The sample mean, CoV, median, and the 1% and 5% quantiles of θ were computed for the whole database and for each fibre type. Lognormality of θ was assessed with the Anderson-Darling statistic on $\ln \theta$ and by comparing empirical with fitted lognormal lower quantiles. Systematic trends were tested with

Spearman rank correlations between θ and ρ_K , ρ_e , f'_{co} , D and f_1/f'_{co} ; a significant correlation means that the model is biased in part of the parameter space, so that a single resistance factor cannot be safe everywhere.

The mean-value correction b and the CoV of the error term V_δ were obtained following EN 1990 Annex D (CEN, 2002): $b = \Sigma r_{e,i} / \Sigma r_{t,i}^2$, $\delta_i = r_{e,i} / (b r_{t,i})$, and V_δ from the variance of $\Delta_i = \ln \delta_i$.

4.2. Between-programme and within-programme variance

Because specimens are grouped by test programme, $\ln \theta$ was modelled as a one-way random-effects model, $\ln \theta_{ij} = \mu + u_i + e_{ij}$, with a programme effect $u_i \sim N(0,$

σ_b^2) and a specimen effect $e_{ij} \sim N(0, \sigma_w^2)$, estimated by restricted maximum likelihood. The intraclass correlation $ICC = \sigma_b^2 / (\sigma_b^2 + \sigma_w^2)$ measures how much of the model uncertainty is a laboratory effect. When ICC is large, the effective number of independent observations is close to the number of programmes, not to the number of specimens, and a model fitted and tested on randomly split specimens is optimistically biased.

4.3. Data-driven models and study-grouped validation

Three data-driven models were compared. (i) *Refit power law*: the closed-form $1 + a(f_i/f'_{co})^b$ for strength, and $a_0 + a \rho_K^b \rho_\varepsilon^c$ for strain, refitted by nonlinear least squares on each training set; this isolates the gain that comes from re-calibrating an existing functional form. (ii) *XGB-M*: a gradient-boosted tree ensemble (Friedman, 2001; Chen and Guestrin, 2016) trained on $\ln$ of the target with the inputs $(\rho_\varepsilon, \rho_a, \rho_E, f'_{co}, D)$ and monotonic constraints imposed on the three confinement ratios, so that the prediction cannot decrease when confinement increases. (iii) *XGB-R*: the same ensemble applied to the logarithm of the model uncertainty of the best closed-form equation (Keshtegar et al., 2017b) with the inputs $(f_i/f'_{co}, \rho_K, \rho_\varepsilon, f'_{co}, D)$; the mechanical equation therefore carries the trend and the ensemble only learns a correction. Hyper-parameters were fixed a priori (learning rate 0.05, depth 3, 300 trees, subsample 0.8, minimum child weight 3); the sensitivity of the conclusions to this choice was checked over depths 2-4 and 150-600 trees.

Two validation schemes were used with ten folds each: *random* cross-validation, in which specimens are assigned to folds at random, and *study-grouped* cross-validation, in which all specimens of a test programme are assigned to the same fold, so that no sibling of a validated specimen is used in training. Both were repeated with five different fold assignments and the mean and standard deviation of the out-of-fold statistics are reported.

4.4. Conformal prediction intervals

A design decision needs an interval, not a point value. Split-conformal prediction converts any point predictor into an interval with finite-sample coverage under exchangeability (Vovk et al., 2005; Lei et al., 2018; Angelopoulos and Bates, 2023). Here the nonconformity score was the absolute logarithmic residual $|\ln(r_e/r_i)|$, and its $(1 - \alpha)(1 + 1/n)$ empirical quantile was computed in a cross-conformal way from the nine folds not containing the specimen, using study-grouped folds so that the calibration set never contains siblings of the predicted specimen (Barber et al., 2021). Two-sided 90% intervals and one-sided 95% lower bounds were produced both for the best closed-form equation and for the XGB-R model. Coverage was evaluated overall, per fibre type, and also with fibre-conditional (Mondrian) calibration sets.

4.5. Reliability-based calibration

Two complementary calibrations were performed. The first follows EN 1990 Annex D and gives the design value of the resistance directly from the test population,

$$\psi_D = \frac{r_d}{r_t(X_m)} = b \exp(-k_{d,\infty} \alpha_{rt} Q_{rt} - k_{d,n} \alpha_\delta Q_\delta - 0.5 Q^2) \quad (5)$$

with $Q_{rt} = [\ln(V_{rt}^2 + 1)]^{0.5}$, $Q_\delta = [\ln(V_\delta^2 + 1)]^{0.5}$, $Q = [\ln(V_r^2 + 1)]^{0.5}$, $V_r^2 = (V_\delta^2 + 1)(V_{rt}^2 + 1) - 1$, $\alpha_{rt} = Q_{rt}/Q$, $\alpha_\delta = Q_\delta/Q$ and $k_{d,\infty} = k_{d,n} = \alpha_R \beta = 0.8 \times 3.8 = 3.04$, which is appropriate because the number of tests is large. V_{rt} is the CoV of the basic variables entering the model, which was varied between 0.05 and 0.20.

The second calibration is a member-level FORM analysis of an axially loaded FRP-confined column. The limit state is

$$g = \theta B_M R_n - D - L \quad (6)$$

in which R_n is the nominal resistance obtained from the confinement model, B_M is a lognormal material and geometry factor with mean 1.0 and CoV V_M (0.10 unless stated otherwise), θ is lognormal with the mean and CoV measured from the database, D is a normal dead load effect with bias 1.05 and CoV 0.10 and L is a Gumbel 50-year maximum live load effect with bias 1.0 and CoV 0.18; these load statistics are those used in the calibration of ACI 318 (Nowak and Szerszen, 2003; Szerszen and Nowak, 2003) and follow the load model of Ellingwood et al. (1982). The value $V_M = 0.10$ is representative of the combined material and geometrical uncertainty of a wrapped column: the Probabilistic Model Code (JCSS, 2001) gives a CoV of 0.07–0.14 for the concrete strength, and comparable values have been used for FRP properties in resistance-factor calibrations for externally bonded composites (Atadero and Karbhari, 2008). The nominal resistance follows from the design equation $\phi R_n = 1.2 D_n + 1.6 L_n$, and the live-to-dead ratio L_n/D_n was varied from 0.25 to 3. The reliability index was computed with the Hasofer-Lind-Rackwitz-Fiessler algorithm and verified by importance sampling; the resistance factor that achieves a target $\beta_T = 4.0$, the value used for columns in the ACI 318 calibration, was found by root-finding.

5. Results

5.1. Model uncertainty of the existing equations

Table 3 gives the model-uncertainty statistics over all 780 specimens. For the strength models the mean of θ ranges from 0.888 (Newman and Newman, 1971) to 1.092 (Teng et al., 2009), so the oldest equations are unconservative on this database by about 10% on average, whereas the four recent equations are almost unbiased. The CoV of θ is remarkably stable, 0.177–0.251, with the single exception of Ozbakkaloglu and Lim (2013) (0.251), whose model produced a few very low predictions for lightly confined high-strength specimens. In other words, thirty years of model development moved the mean of θ towards unity but

barely changed its scatter, and it is the scatter that governs the resistance factor. The share of unconservative predictions ($\theta < 1$) is 82% for Newman and Newman (1971) and close to 50% for the recent models, as expected for mean-value fits.

The strain models behave differently. Their CoV is two to four times larger (0.36–0.82), the two models that require the measured strength ratio are strongly biased (mean $\theta = 1.48$ for Mander et al., 1988), and even the

best equations leave a CoV of about 0.36. The 5% quantile of θ for the strain models is 0.48–0.65: one specimen in twenty reaches less than half of the predicted ultimate strain. Any design rule that relies on the predicted ultimate strain, for example a displacement-based assessment of a confined column, therefore needs a much larger safety margin than a strength check.

Table 3. Model uncertainty $\theta = \text{experiment/prediction}$ over the whole database

Model	<i>n</i>	Mean	CoV	$\theta_{0.05}$ emp.	$\theta_{0.05}$ LN	$\theta < 1$ (%)	<i>b</i>	<i>V_δ</i>
Ultimate strength ratio								
Newman & Newman (1971)	780	0.888	0.183	0.68	0.66	82	0.875	0.171
Toutanji (1999)	780	0.907	0.182	0.69	0.68	78	0.901	0.171
Guralnick & Gunawan (2006)	780	1.082	0.188	0.81	0.79	37	1.121	0.179
Teng et al. (2009)	780	1.092	0.182	0.84	0.81	31	1.035	0.173
Ozbakkaloglu & Lim (2013)	780	1.040	0.251	0.80	0.76	50	1.032	0.185
Sadeghian & Fam (2015)	780	1.005	0.184	0.77	0.75	55	0.989	0.170
Keshtegar et al. I (2017a)	780	1.030	0.177	0.79	0.77	48	1.051	0.168
Keshtegar et al. II (2017b)	780	1.003	0.181	0.77	0.75	56	1.000	0.168
Ultimate strain ratio								
Mander et al. (1988)	778	1.476	0.563	0.65	0.64	25	1.195	0.464
Toutanji (1999)	778	1.361	0.819	0.65	0.64	29	0.980	0.418
Teng et al. (2009)	780	0.935	0.381	0.48	0.48	67	0.824	0.378
Ozbakkaloglu & Lim (2013)	780	1.203	0.393	0.59	0.60	33	1.017	0.390
Sadeghian & Fam (2015)	780	0.999	0.362	0.48	0.51	56	1.025	0.380
Keshtegar et al. I (2017a)	780	0.992	0.365	0.49	0.51	56	1.021	0.378
Keshtegar et al. II (2017b)	780	0.981	0.358	0.49	0.51	59	1.000	0.375

$\theta_{0.05}$ emp. and LN are the empirical and fitted-lognormal 5% quantiles; *b* and *V_δ* follow EN 1990 Annex D.

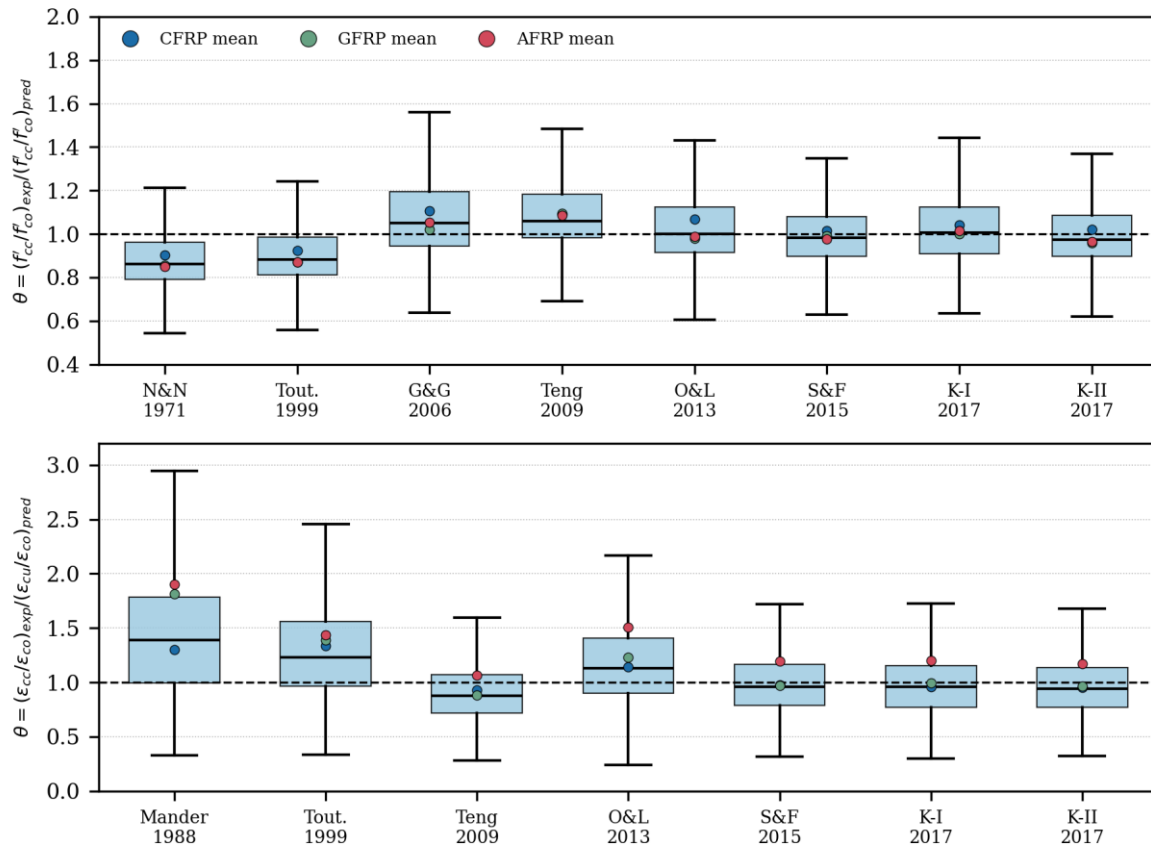


Figure 2. Distribution of the model uncertainty θ for the strength (top) and strain (bottom) models. Boxes show the whole database (median, quartiles, 1.5 IQR whiskers); markers are the fibre-specific means.

Fibre-specific statistics (Table 4) show a systematic difference. All strength models are on average less conservative for GFRP and AFRP specimens than for CFRP specimens; for example the mean θ of Keshtegar et al. (2017b) is 1.022 for CFRP but 0.960 for GFRP and

0.966 for AFRP. The CoV is similar across fibre types for the strength models but differs strongly for the strain models, where the AFRP subset is both the smallest and the most biased (mean $\theta = 1.17$).

Table 4. Fibre-specific model uncertainty for four representative models

Model	CFRP mean	CFRP CoV	GFRP mean	GFRP CoV	AFRP mean	AFRP CoV
Ultimate strength ratio						
Teng et al. (2009)	1.093	0.188	1.096	0.172	1.085	0.170
Ozbakkaloglu and Lim (2013)	1.068	0.272	0.979	0.174	0.990	0.168
Sadeghian and Fam (2015)	1.014	0.187	0.993	0.176	0.976	0.177
Keshtegar et al. II (2017b)	1.022	0.181	0.960	0.176	0.966	0.168
Ultimate strain ratio						
Teng et al. (2009)	0.928	0.360	0.885	0.438	1.067	0.369
Ozbakkaloglu and Lim (2013)	1.144	0.367	1.231	0.448	1.510	0.329
Sadeghian and Fam (2015)	0.975	0.343	0.971	0.423	1.197	0.308
Keshtegar et al. II (2017b)	0.955	0.337	0.963	0.420	1.171	0.310

5.2. Residual trends and distribution

The Anderson-Darling statistic rejects the lognormal hypothesis for most models at the 5% level, but the deviation is concentrated in the centre of the distribution: the fitted lognormal 5% quantiles in Table 3 agree with the empirical ones to within a few per cent, and the 1% quantiles agree to within about 5%. The lognormal model is therefore adequate for the tail that matters in a reliability analysis, which is consistent with the conclusion of Keshtegar et al. (2020).

Spearman correlations reveal residual trends that a single factor cannot capture. For Keshtegar et al. (2017b) the strength model uncertainty increases with ρ_K ($\rho_s = +0.19$, $p < 0.001$) and decreases with the concrete strength ($\rho_s = -0.26$, $p < 0.001$), so the equation becomes unconservative for high-strength concrete under light confinement. For the strain ratio the trends are of the opposite sign in ρ_ε ($\rho_s = +0.16$) and in the diameter ($\rho_s = -0.19$), which reflects the well-known size effect on the hoop rupture strain. Figure 3 shows these patterns.

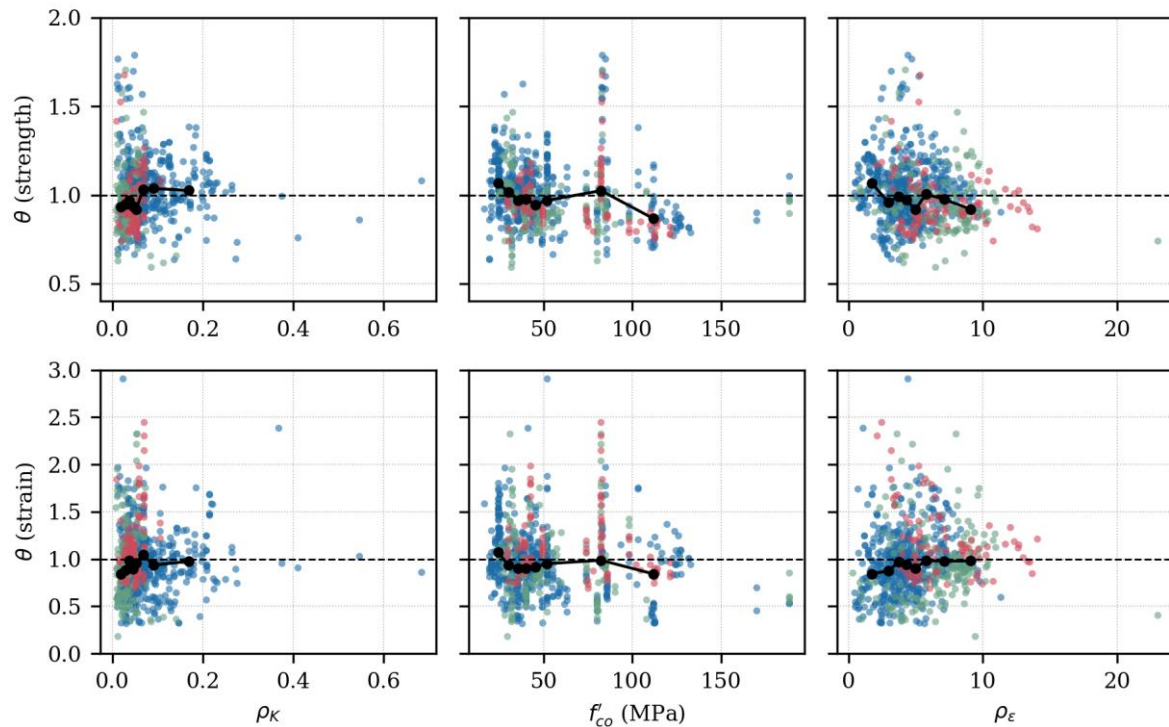


Figure 3. Model uncertainty of Keshtegar et al. (2017b) against ρ_K , f'_{co} and ρ_ϵ for the strength (top) and strain (bottom) ratios. Black lines connect medians of eight equal-count bins.

5.3. How much of the scatter is a laboratory effect?

Table 5 gives the random-effects decomposition of $\ln \theta$. For the strength models the between-programme standard deviation σ_b is 0.143–0.158 and the within-programme value σ_w is 0.108–0.135, giving ICC = 0.53–0.67. For the strain models ICC = 0.49–0.62. Between one half and two thirds of the model uncertainty is thus a property of the test programme rather than of the individual specimen (Figure 4). This has two consequences. For validation, the effective sample size is closer to the 63 programmes than to the 780 specimens. For design, the relevant uncertainty when a model is applied to a new column is the total variance, including the programme effect, which is what the pooled CoV of Table 3 measures; conversely, an ML model that reduces only the within-programme residual can improve error statistics without improving design safety at all.

Table 5. Random-effects decomposition of $\ln \theta$ into between-programme and within-programme components

Model	μ	σ_b	σ_w	ICC
Ultimate strength ratio				
Newman & Newman (1971)	-0.114	0.156	0.109	0.67
Toutanji (1999)	-0.093	0.155	0.109	0.67
Guralnick & Gunawan (2006)	+0.078	0.149	0.121	0.60
Teng et al. (2009)	+0.096	0.158	0.110	0.67
Ozbakkaloglu & Lim (2013)	+0.040	0.143	0.135	0.53
Sadeghian & Fam (2015)	+0.012	0.154	0.108	0.67
Keshtegar et al. I (2017a)	+0.032	0.146	0.114	0.62
Keshtegar et al. II (2017b)	+0.008	0.148	0.110	0.64
Ultimate strain ratio				
Mander et al. (1988)	+0.253	0.351	0.339	0.52
Toutanji (1999)	+0.188	0.313	0.319	0.49
Teng et al. (2009)	-0.136	0.294	0.256	0.57
Ozbakkaloglu & Lim (2013)	+0.087	0.319	0.248	0.62
Sadeghian & Fam (2015)	-0.071	0.286	0.260	0.55
Keshtegar et al. I (2017a)	-0.081	0.284	0.261	0.54
Keshtegar et al. II (2017b)	-0.086	0.280	0.258	0.54

Restricted maximum likelihood; ICC = $\sigma_b^2 / (\sigma_b^2 + \sigma_w^2)$.

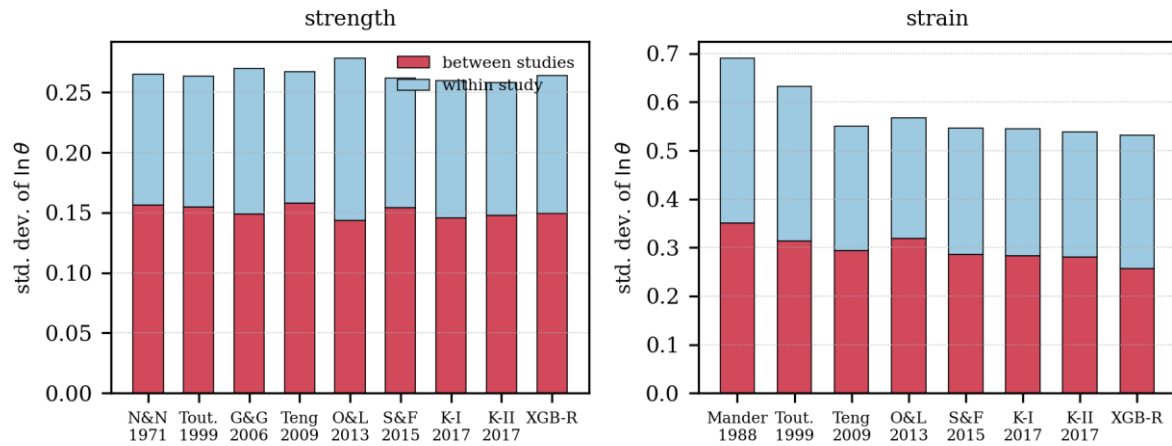


Figure 4. Between-programme and within-programme standard deviation of $\ln \theta$ for the strength (left) and strain (right) models, including the study-grouped gradient-boosting model.

5.4. Random versus study-grouped validation of the data-driven models

Table 6 and Figure 5 contrast the two validation schemes. Under random 10-fold cross-validation the monotonic gradient-boosting model reaches an RMSE of 0.255 for the strength ratio and 1.94 for the strain ratio, that is, 27% and 25% better than the best closed-form equation. These gains are of the same order as those reported for hybrid SVR models on this database (Ben Seghier et al., 2021). Under study-grouped cross-validation the same model gives 0.524 and 2.86, that is, 2.1 times and 1.5 times worse than the random-split estimate, and clearly worse than the closed-form equation. The physics-guided residual model XGB-R is

much more robust: its grouped RMSE is 0.366 for strength, only 5% above the equation it corrects, because the mechanical form carries the extrapolation while the ensemble only adds a bounded correction. Simply re-fitting the two coefficients of a power law generalises better than the free-form ensemble as well (0.398).

The result is not an artefact of hyper-parameters. Over depths 2-4 and 150-600 trees the grouped RMSE of XGB-M stays between 0.484 and 0.563 for the strength ratio and that of XGB-R between 0.349 and 0.368; neither approaches the random-split value. The optimistic bias comes from the grouping structure of the database, not from over-fitting in the usual sense.

Table 6. Out-of-fold performance under random and study-grouped cross-validation

Model	Validation	RMSE	R^2	Mean θ	CoV of θ
Strength ratio f_{co}/f_{co}					
Refit power law	random CV	0.376 ± 0.002	0.778	0.993	0.181
XGB-M	random CV	0.255 ± 0.002	0.898	1.007	0.121
XGB-R	random CV	0.252 ± 0.003	0.900	1.006	0.118
Refit power law	study-grouped CV	0.398 ± 0.002	0.751	0.995	0.185
XGB-M	study-grouped CV	0.524 ± 0.011	0.568	1.044	0.220
XGB-R	study-grouped CV	0.366 ± 0.008	0.790	1.033	0.189
Sadeghian and Fam (2015)	fixed equation	0.362	0.794	1.005	0.184
Keshtegar et al. II (2017b)	fixed equation	0.349	0.809	1.003	0.181
Strain ratio $\varepsilon_{co}/\varepsilon_{co}$					
Refit power law	random CV	2.576 ± 0.003	0.697	1.036	0.764
XGB-M	random CV	1.941 ± 0.014	0.828	1.031	0.262
XGB-R	random CV	2.006 ± 0.018	0.816	1.029	0.255
Refit power law	study-grouped CV	2.654 ± 0.017	0.678	1.017	0.880
XGB-M	study-grouped CV	2.855 ± 0.062	0.627	1.085	0.406
XGB-R	study-grouped CV	2.851 ± 0.031	0.628	1.097	0.426
Sadeghian and Fam (2015)	fixed equation	2.594	0.692	0.999	0.362
Keshtegar et al. II (2017b)	fixed equation	2.579	0.696	0.981	0.358

Mean $\pm$ standard deviation over five fold assignments. The fixed equations are evaluated on all 780 specimens.

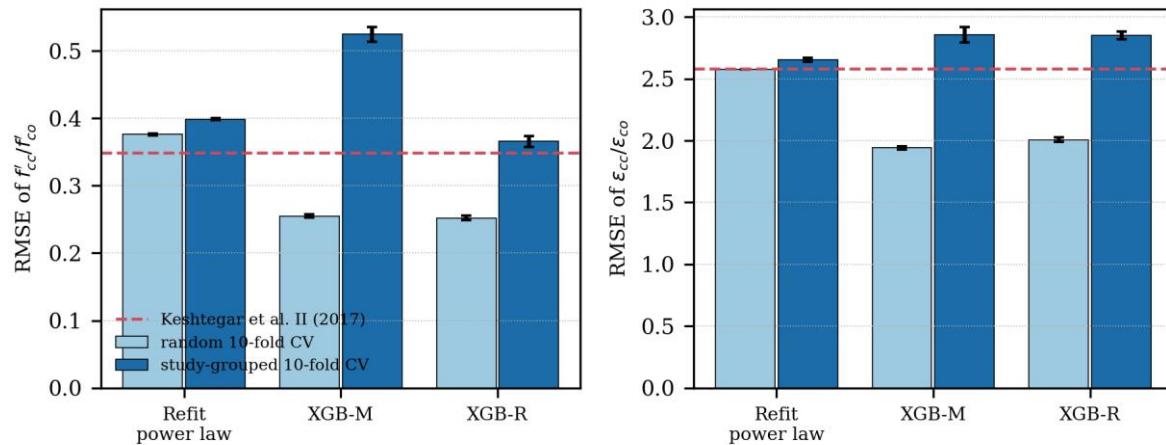


Figure 5. RMSE of the data-driven models under random and study-grouped cross-validation for (a) the strength ratio and (b) the strain ratio. The dashed line is the closed-form equation of Keshtegar et al. (2017b).

The fibre-specific results (Table 7) add a practical caveat: under grouped validation the ensemble is competitive for CFRP, for which half of the programmes are available, but it loses to the closed-form equation for

the GFRP and AFRP strain ratios, where only a few programmes exist. A data-driven model inherits the coverage of its database, and the AFRP subset is represented by 11 programmes only.

Table 7. Fibre-specific out-of-fold RMSE and CoV of θ (study-grouped cross-validation)

Model	RMSE CFRP	RMSE GFRP	RMSE AFRP	CoV CFRP	CoV GFRP	CoV AFRP
Strength ratio						
Keshtegar et al. II (2017b)	0.361	0.337	0.289	0.181	0.176	0.168
Refit power law (grouped CV)	0.426	0.344	0.295	0.184	0.177	0.167
XGB-M (grouped CV)	0.581	0.420	0.328	0.222	0.227	0.181
XGB-R (grouped CV)	0.368	0.342	0.274	0.187	0.187	0.146
Strain ratio						
Keshtegar et al. II (2017b)	2.26	3.20	3.06	0.337	0.420	0.310
Refit power law (grouped CV)	2.32	3.27	3.18	0.375	0.682	0.346
XGB-M (grouped CV)	2.55	3.49	3.92	0.388	0.499	0.385
XGB-R (grouped CV)	2.33	3.61	3.76	0.387	0.495	0.397

5.5. Conformal prediction intervals

Grouped cross-conformal intervals were generated for the closed-form equation and for XGB-R. Table 8 reports the empirical coverage of the nominal 90% intervals. Pooled over the database the coverage is 0.904 for the strength ratio and 0.897 for the strain ratio, as the theory predicts. The intervals are wide: the median ratio between the upper and the lower limit is 1.66 for the strength ratio and 3.30 for the strain ratio, i.e. the ultimate strain of a confined column is only known to within a factor of about three. Coverage is not uniform

across fibre types: for the strain ratio the GFRP subset is covered 82% of the time with pooled calibration, and fibre-conditional (Mondrian) calibration restores it to 91% at the price of wider intervals. More seriously, coverage conditional on the test programme is poor: 22% of the programmes have less than 70% of their specimens inside the 90% interval for the strength ratio. Exchangeability, which conformal prediction requires, does not hold across laboratories, and this is the same laboratory effect that the variance decomposition quantified.

Table 8. Empirical coverage of grouped cross-conformal intervals (nominal 90%)

Base model	Subset	Coverage	Width ratio	Coverage (Mondrian)	One-sided 95% bound
Strength ratio					
Keshtegar et al. II (2017b)	CFRP	0.896	1.69	0.902	0.951
Keshtegar et al. II (2017b)	GFRP	0.895	1.69	0.907	0.932
Keshtegar et al. II (2017b)	AFRP	0.943	1.71	0.898	0.977
Keshtegar et al. II (2017b)	All	0.901	1.69	0.903	0.950
XGB-R (grouped CV)	CFRP	0.909	1.66	0.900	0.951
XGB-R (grouped CV)	GFRP	0.864	1.66	0.914	0.907
XGB-R (grouped CV)	AFRP	0.943	1.67	0.943	0.966
XGB-R (grouped CV)	All	0.904	1.66	0.908	0.944
Strain ratio					
Keshtegar et al. II (2017b)	CFRP	0.908	3.56	0.902	0.951

Base model	Subset	Coverage	Width ratio	Coverage (Mondrian)	One-sided 95% bound
Keshtegar et al. II (2017b)	GFRP	0.821	3.50	0.907	0.914
Keshtegar et al. II (2017b)	AFRP	0.955	3.56	0.795	1.000
Keshtegar et al. II (2017b)	All	0.895	3.56	0.891	0.949
XGB-R (grouped CV)	CFRP	0.919	3.30	0.906	0.947
XGB-R (grouped CV)	GFRP	0.846	3.30	0.914	0.901
XGB-R (grouped CV)	AFRP	0.864	3.41	0.761	1.000
XGB-R (grouped CV)	All	0.897	3.30	0.891	0.944

Width ratio is the median of the upper-to-lower interval limit. The last column is the fraction of tests above the one-sided 95% lower bound (nominal 0.95).

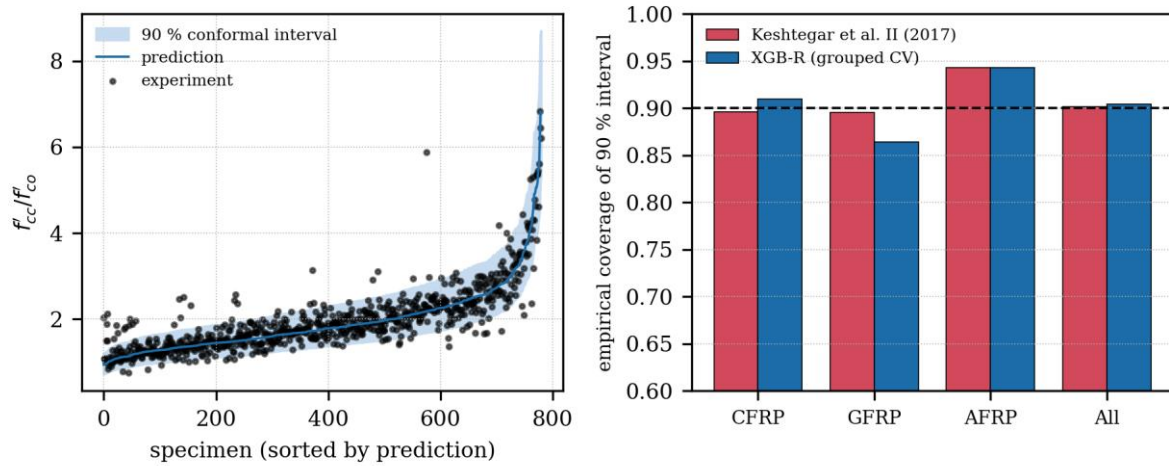


Figure 6. Grouped cross-conformal prediction for the strength ratio: (a) 90% intervals for the XGB-R model, specimens sorted by prediction; (b) empirical coverage by fibre type.

5.6. Design factors and resistance factors

Table 9 gives the Annex D design factor $\psi_D = r_d/r_t(X_m)$, the fraction of the predicted mean resistance that may be used as a design value. With a basic-variable CoV $V_{rt} = 0.10$ the strength models give $\psi_D = 0.47$ – 0.59 and the strain models only 0.24 – 0.30 . The ranking is not the ranking of the error statistics: Teng et al. (2009), which has a larger RMSE than Keshtegar et al. (2017b), yields a *higher* design value (0.55 against 0.54) because it is conservative on average. The grouped-CV gradient-boosting correction gives 0.55, essentially the same as the equation it corrects.

The member-level FORM results are summarised in Table 10 and Figure 7. For an axially loaded column with $L_n/D_n = 1$ and $V_M = 0.10$, the resistance factor needed to reach $\beta_T = 4.0$ is $\phi = 0.47$ for Newman and Newman (1971), $\phi = 0.53$ for Keshtegar et al. (2017b) and $\phi = 0.58$ for Teng et al. (2009); the study-grouped ensemble gives $\phi = 0.55$. Conversely, applying the ACI 318 factor for tied columns, $\phi = 0.65$, to the mean prediction of the best equation yields $\beta = 3.18$, and $\phi = 0.75$ yields $\beta = 2.58$. An importance-sampling check with 4×10^5 samples reproduced the FORM index to within 0.05 ($\beta_{FORM} = 3.18$, $\beta_{IS} = 3.13$). The model uncertainty dominates the sensitivity: its FORM importance factor α^2 is 0.55, against 0.17 for the material and geometry factor and 0.24 for the live load. Increasing V_M from 0.05 to 0.20 lowers the required ϕ from 0.57 to 0.42, and the dependence on the load ratio is weak (0.50–0.52).

Fibre-specific factors (Figure 7b) differ by up to 0.09 between fibre types for the same equation, because of the fibre-dependent bias reported above: with Keshtegar et al. (2017b), $\phi = 0.54$ for CFRP but 0.52 for GFRP. A single factor calibrated on the pooled database is therefore unconservative for glass-FRP applications and conservative for carbon-FRP applications.

Table 9. EN 1990 Annex D design factor $\psi_D = r_d/r_t(X_m)$ for four levels of basic-variable uncertainty

Model	$V_{rt} = 0.05$	$V_{rt} = 0.10$	$V_{rt} = 0.15$	$V_{rt} = 0.20$
Ultimate strength ratio				
Newman & Newman (1971)	0.503	0.471	0.429	0.382
Toutanji (1999)	0.519	0.486	0.442	0.394
Guralnick & Gunawan (2006)	0.629	0.591	0.539	0.482
Teng et al. (2009)	0.591	0.555	0.505	0.451
Ozbakkaloglu & Lim (2013)	0.569	0.535	0.489	0.438
Sadeghian & Fam (2015)	0.571	0.535	0.487	0.434
Keshtegar et al. I (2017a)	0.610	0.571	0.519	0.463
Keshtegar et al. II (2017b)	0.581	0.544	0.494	0.441
XGB-R (grouped CV)	0.588	0.551	0.500	0.446
Ultimate strain ratio				
Mander et al. (1988)	0.280	0.272	0.260	0.244
Toutanji (1999)	0.264	0.256	0.243	0.227
Teng et al. (2009)	0.251	0.243	0.230	0.214
Ozbakkaloglu & Lim (2013)	0.299	0.289	0.274	0.255
Sadeghian & Fam (2015)	0.310	0.300	0.284	0.264
Keshtegar et al. I (2017a)	0.311	0.301	0.285	0.265
Keshtegar et al. II (2017b)	0.307	0.297	0.281	0.261
XGB-R (grouped CV)	0.314	0.303	0.287	0.267

Table 10. Reliability of an axially loaded FRP-confined column ($L_n/D_n = 1$, $V_M = 0.10$) and required resistance factors

Strength model	Mean θ	CoV of θ	β at $\phi = 0.65$	β at $\phi = 0.75$	ϕ for $\beta_T = 3.5$	ϕ for $\beta_T = 4.0$
Newman & Newman (1971)	0.888	0.183	2.65	2.05	0.53	0.47
Toutanji (1999)	0.907	0.182	2.75	2.15	0.54	0.48
Guralnick & Gunawan (2006)	1.082	0.188	3.42	2.84	0.64	0.56
Teng et al. (2009)	1.092	0.182	3.52	2.93	0.65	0.58
Ozbakkaloglu & Lim (2013)	1.040	0.251	2.67	2.17	0.51	0.44
Sadeghian & Fam (2015)	1.005	0.184	3.16	2.57	0.60	0.53
Keshtegar et al. I (2017a)	1.030	0.177	3.33	2.73	0.62	0.55
Keshtegar et al. II (2017b)	1.003	0.181	3.18	2.58	0.60	0.53
Refit power law (grouped CV)	0.994	0.183	3.12	2.53	0.59	0.52
XGB-M (grouped CV)	1.044	0.219	2.96	2.42	0.56	0.49
XGB-R (grouped CV)	1.034	0.183	3.29	2.70	0.62	0.55

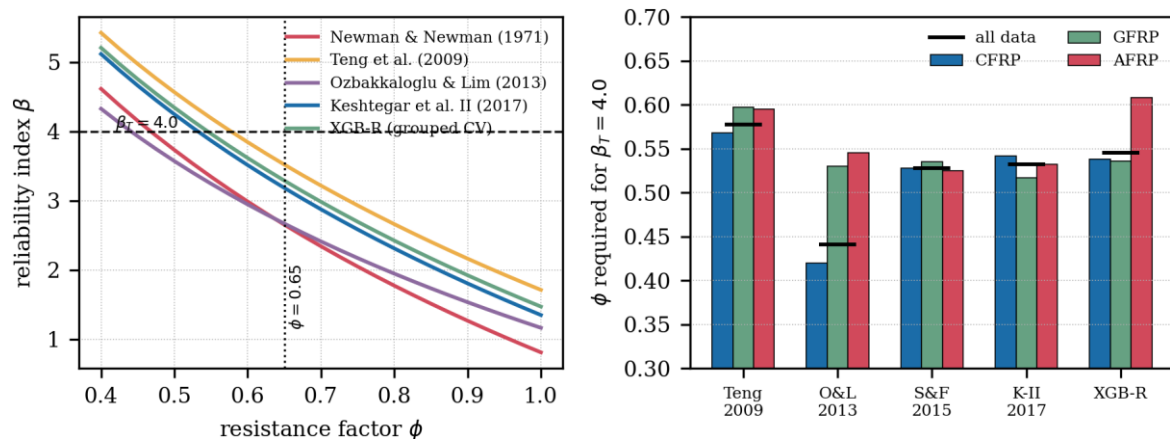


Figure 7. Reliability-based calibration: (a) reliability index against the resistance factor for five strength models; (b) resistance factor required for $\beta_T = 4.0$ by fibre type (dashes: all data).

6. Discussion

The three analyses point in the same direction. The scatter of the model uncertainty of FRP confinement models has barely improved since the first design-oriented equations, the largest part of that scatter is a laboratory effect, and a machine-learning model that is validated without respecting the laboratory structure appears to remove scatter that it cannot actually remove. The practical consequence is that the resistance factor implied by any of these models is low: with the load model used in the calibration of ACI 318 and a target reliability appropriate for columns, the mean prediction of the best available equation must be multiplied by about 0.53. ACI 440.2R-17 reaches a comparable level by combining the strength-reduction factor for the member with an additional factor $\psi_f = 0.95$ on the confinement contribution and an efficiency factor $\kappa_s = 0.55$ on the rupture strain; the present results suggest that the reduction is dominated by the model uncertainty of the confinement equation and by the variability of the rupture strain, and that an additional factor of the order of 0.82 on the predicted confined strength would be needed if the hoop rupture strain were known exactly.

It is important to state what the study does *not* show. It does not show that machine learning is useless for confinement problems. The physics-guided residual

model was almost as good as the closed-form equation under the strict grouped protocol while being clearly better under the random protocol, and it removed the residual trends in ρ_K and f'_{co} ; with a database that covers more laboratories, in particular for glass and aramid jackets, a data-driven correction of a mechanical equation is a reasonable way forward. What the study shows is that the *reported* superiority of flexible ML models over design equations on this database is largely an artefact of the validation protocol, and that even a perfect mean predictor would not change the resistance factor much, because the between-laboratory scatter would remain.

Several limitations should be kept in mind. The database contains no specimen height, so the size-dependent form of ε_{co} used by some authors could not be reproduced and the database convention (Eq. 5) was applied to all models; models calibrated with a different ε_{co} convention are therefore slightly penalised. The hoop rupture strain is treated as a measured input, whereas in design it must be predicted from the coupon strength with an efficiency factor, so the model uncertainty reported here is a lower bound on the uncertainty of a real design calculation. The FORM calculation uses a gravity load combination and a lumped material and geometry factor; a seismic retrofit application would

need a different load model. The nine specimens with a measured strength ratio below unity and the three with a strain ratio below unity were retained; removing them raises the mean of θ by less than 0.01 and lowers its CoV by about 0.005 for the strength models, which does not affect any conclusion. Finally, the equations were implemented in the algebraic form in which they are collected by the source study; two of them could not be checked against the original publications because those equations are reproduced there as images, and small differences in exponents would change the individual statistics but not the pattern across models.

7. Conclusions

A database of 780 circular FRP-confined concrete specimens was re-analysed from the point of view of design rather than of prediction accuracy. The main findings are:

(1) The coefficient of variation of the model uncertainty of eight design-oriented strength models is 0.177–0.251 and of seven strain models 0.36–0.82. Recent equations are nearly unbiased but are not less scattered than equations proposed forty years earlier.

(2) Between 53% and 67% of the variance of $\ln \theta$ (strength) lies between test programmes. The effective sample size of the database is therefore closer to its 63 test programmes than to its 780 specimens.

(3) Gradient boosting reduced the RMSE of the strength ratio to 0.255 under random cross-validation but only to 0.524 under study-grouped cross-validation, against 0.349 for the best closed-form equation. For the strain ratio the grouped-CV ensemble was worse than the closed-form equation. A physics-guided residual formulation was the most robust data-driven variant.

(4) Grouped cross-conformal intervals attain their nominal coverage on average but not for individual test programmes; the 90% interval for the ultimate strain spans a factor of about 3.3.

(5) Resistance factors of $\phi \approx 0.53$ ($\beta_T = 4.0$) or $\phi \approx 0.60$ ($\beta_T = 3.5$) are required with the best strength model, and the EN 1990 Annex D design factor for the strain models is below 0.31. Fibre-specific factors differ by up to 0.09.

Model comparisons on grouped experimental databases should report the model uncertainty, use grouped validation, and state the resistance factor that the model implies. Without these, better error statistics do not translate into safer or more economical design.

Data and code availability

The cleaned database, the implementation of all models, the cross-validation, conformal and reliability scripts and the scripts that generate every figure and table are provided as supplementary material.

References

ACI Committee 440 (2017). Guide for the Design and Construction of Externally Bonded FRP Systems for Strengthening Concrete Structures (ACI 440.2R-17). American Concrete Institute, Farmington Hills, MI.

- Angelopoulos, A.N., & Bates, S. (2023). Conformal prediction: a gentle introduction. *Foundations and Trends in Machine Learning*, 16(4), 494–591. <https://doi.org/10.1561/22000000101>
- Atadero, R.A., & Karbhari, V.M. (2008). Calibration of resistance factors for reliability based design of externally-bonded FRP composites. *Composites Part B: Engineering*, 39(4), 665–679. <https://doi.org/10.1016/j.compositesb.2007.06.004>
- Barber, R.F., Candès, E.J., Ramdas, A., & Tibshirani, R.J. (2021). Predictive inference with the jackknife+. *The Annals of Statistics*, 49(1), 486–507. <https://doi.org/10.1214/20-AOS1965>
- Ben Seghier, M.E.A., Keshtegar, B., Nait Amar, M., Correia, J.A.F.O., & Trung, N.-T. (2021). Simulation of the ultimate conditions of fibre-reinforced polymer confined concrete using hybrid intelligence models. *Engineering Failure Analysis*, 128, 105605. <https://doi.org/10.1016/j.engfailanal.2021.105605>
- CEN (2002). EN 1990:2002 Eurocode – Basis of Structural Design, Annex D: Design assisted by testing. European Committee for Standardization, Brussels.
- Chen, Q.-S., Yu, B., & Li, B. (2022). Probabilistic calibration of compressive stress–strain models for FRP-confined concrete in square cross section members. *Journal of Composites for Construction*, 26(3), 04022015. [https://doi.org/10.1061/\(ASCE\)CC.1943-5614.0001202](https://doi.org/10.1061/(ASCE)CC.1943-5614.0001202)
- Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
- Ellingwood, B., MacGregor, J.G., Galambos, T.V., & Cornell, C.A. (1982). Probability based load criteria: load factors and load combinations. *Journal of the Structural Division*, 108(5), 978–997. <https://doi.org/10.1061/JSDAEG.0005959>
- Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29(5). <https://doi.org/10.1214/aos/1013203451>
- Guralnick, S.A., & Gunawan, L. (2006). Strengthening of reinforced concrete bridge columns with FRP wrap. *Practice Periodical on Structural Design and Construction*, 11(4), 218–228. [https://doi.org/10.1061/\(ASCE\)1084-0680\(2006\)11:4\(218\)](https://doi.org/10.1061/(ASCE)1084-0680(2006)11:4(218))
- JCSS (2001). Probabilistic Model Code. Joint Committee on Structural Safety.
- Keshtegar, B., Gholampour, A., Ozbakkaloglu, T., Zhu, S.-P., & Trung, N.-T. (2020). Reliability analysis of FRP-confined concrete at ultimate using conjugate search direction method. *Polymers*, 12(3), 707. <https://doi.org/10.3390/polym12030707>
- Keshtegar, B., Ozbakkaloglu, T., & Gholampour, A. (2017a). Modeling the behavior of FRP-confined concrete using dynamic harmony search algorithm.

- Engineering with Computers, 33(3), 415–430.
<https://doi.org/10.1007/s00366-016-0481-y>
- Keshtegar, B., Sadeghian, P., Gholampour, A., & Ozbakkaloglu, T. (2017b). Nonlinear modeling of ultimate strength and strain of FRP-confined concrete using chaos control method. *Composite Structures*, 163, 423–431.
<https://doi.org/10.1016/j.compstruct.2016.12.023>
- Lam, L., & Teng, J.G. (2003). Design-oriented stress–strain model for FRP-confined concrete. *Construction and Building Materials*, 17(6–7), 471–489.
[https://doi.org/10.1016/S0950-0618\(03\)00045-X](https://doi.org/10.1016/S0950-0618(03)00045-X)
- Lei, J., G’Sell, M., Rinaldo, A., Tibshirani, R.J., & Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523), 1094–1111.
<https://doi.org/10.1080/01621459.2017.1307116>
- Lim, J.C., Karakus, M., & Ozbakkaloglu, T. (2016). Evaluation of ultimate conditions of FRP-confined concrete columns using genetic programming. *Computers & Structures*, 162, 28–37.
<https://doi.org/10.1016/j.compstruc.2015.09.005>
- Mander, J.B., Priestley, M.J.N., & Park, R. (1988). Theoretical stress-strain model for confined concrete. *Journal of Structural Engineering*, 114(8), 1804–1826.
[https://doi.org/10.1061/\(ASCE\)0733-9445\(1988\)114:8\(1804\)](https://doi.org/10.1061/(ASCE)0733-9445(1988)114:8(1804))
- Mansouri, I., Kisi, O., Sadeghian, P., Lee, C.-H., & Hu, J.W. (2017). Prediction of ultimate strain and strength of FRP-confined concrete cylinders using soft computing methods. *Applied Sciences*, 7(8), 751.
<https://doi.org/10.3390/app7080751>
- Newman, K., & Newman, J.B. (1971). Failure theories and design criteria for plain concrete. In M. Te’eni (Ed.), *Structure, Solid Mechanics and Engineering Design* (pp. 963–995). Wiley-Interscience, London.
- Nowak, A.S., & Collins, K.R. (2013). *Reliability of Structures* (2nd ed.). CRC Press, Boca Raton, FL.
- Nowak, A.S., & Szerszen, M.M. (2003). Calibration of design code for buildings (ACI 318): Part 1 – statistical models for resistance. *ACI Structural Journal*, 100(3), 377–382.
<https://doi.org/10.14359/12613>
- Ozbakkaloglu, T., & Lim, J.C. (2013). Axial compressive behavior of FRP-confined concrete: experimental test database and a new design-oriented model. *Composites Part B: Engineering*, 55, 607–634.
<https://doi.org/10.1016/j.compositesb.2013.07.025>
- Popovics, S. (1973). A numerical approach to the complete stress-strain curve of concrete. *Cement and Concrete Research*, 3(5), 583–599.
[https://doi.org/10.1016/0008-8846\(73\)90096-3](https://doi.org/10.1016/0008-8846(73)90096-3)
- Ravindra, M.K., & Galambos, T.V. (1978). Load and resistance factor design for steel. *Journal of the Structural Division*, 104(9), 1337–1353.
<https://doi.org/10.1061/JSDEAG.0004981>
- Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillerá-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., & Dormann, C.F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929. <https://doi.org/10.1111/ecog.02881>
- Sadeghian, P., & Fam, A. (2015). Improved design-oriented confinement models for FRP-wrapped concrete cylinders based on statistical analyses. *Engineering Structures*, 87, 162–182.
<https://doi.org/10.1016/j.engstruct.2015.01.024>
- Shayanfar, J., Baharloo, S., Barros, J.A.O., & Mohammadi, A. (2026). Strength model for FRP-confined concrete: comprehensive database and reliability-based partial factor calibration. *Journal of Composites for Construction*, 30(2), 04026006.
<https://doi.org/10.1061/JCCOF2.CCENG-5448>
- Szerszen, M.M., & Nowak, A.S. (2003). Calibration of design code for buildings (ACI 318): Part 2 – reliability analysis and resistance factors. *ACI Structural Journal*, 100(3), 383–391.
<https://doi.org/10.14359/12614>
- Teng, J.G., Jiang, T., Lam, L., & Luo, Y.Z. (2009). Refinement of a design-oriented stress–strain model for FRP-confined concrete. *Journal of Composites for Construction*, 13(4), 269–278.
[https://doi.org/10.1061/\(ASCE\)CC.1943-5614.0000012](https://doi.org/10.1061/(ASCE)CC.1943-5614.0000012)
- Toutanji, H. (1999). Stress-strain characteristics of concrete columns externally confined with advanced fiber composite sheets. *ACI Materials Journal*, 96(3), 397–404. <https://doi.org/10.14359/639>
- Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York. <https://doi.org/10.1007/b106715>