International Symposium
Non-Destructive Testing in Civil Engineering (NDT-CE)

September 15 - 17, 2015, Berlin, Germany

Comparing investigation approaches and NDT methodologies for concrete strength estimation: an international benchmark

D. BREYSSE[1], J.P. BALAYSSAC[2], S. BIONDI[3], A. BOROSNYOI[4], E. CANDIGLIOTA[5], L. CHIAUZZI[6], V. GARNIER[7], M. GRANTHAM[8], O. GUNES[9], V.A.M. LUPRANO[10], A. MASI[6], V. PFISTER[0], Z.M. SBARTAI[1], K. SZILAGYI[4]
Phone: +33 5 40 00 88 40, Fax: + 33 5 40 00 31 13
[1] University of Bordeaux, CNRS UMR 5295, 33400 Talence, France, denis.breysse@u-bordeaux.fr
[2] INSA-University Paul Sabatier, LMDC, 31077 Toulouse, France, jean-paul.balayssac@insa-toulouse.fr
[3] University Gabriele d'Annunzio, Chieti, Pescara, Italy, s.biondi@pricos.unich.it
[4] Budapest University of Technology, Hungary, borosnyoi.adorjan@epito.bme.hu
[5] ENEA, Bologna, Italy, elena.candigliota@enea.it
[6] University of Basilicata, 85100 Potenza, Italy, leonardochiauzzi@hotmail.it
[7] Université de la Méditerranée, Aix en Provence, France, vincent.garnier@univ-amu.fr
[8] Concrete Solutions, Margate, Kent, U.K., michael.grantham@concrete-solutions.info
[9] Istanbul Technical University, Turkey, ogunes@itu.edu.tr
[10] ENEA, CR Brindisi, Italy, vincenza.luprano@enea.it

Abstract

Strength assessment of concrete in existing structures is a key issue. Many NDT techniques are available which can provide information about the material condition and many models can be used to assess concrete strength from NDT results complemented by core tests, but currently there is no way to establish the level of accuracy/confidence of concrete strength estimates. Since the concept of "knowledge levels" (KL) was introduced in Eurocodes this issue has become more important. A benchmark has been carried out in order to compare (a) how experts define and carry out an NDT investigation programme and (b) how experts derive strength values from the NDT measurements. A third question was to build the relationship between the quality of the assessment and the methodology followed. The benchmark was based on synthetic simulations which reproduced a synthetic data set corresponding to a series of twenty 3m-high columns defining a building floor. On the basis of the contributions of benchmark participants, a model has been established of the investigation strategy, the main parameters of the strategy are identified and the results of each contributor have been analyzed. A companion paper will use Monte-Carlo simulations to analyze and quantify the efficiency of the investigation strategies regarding both average strength and strength variability assessment.

Keywords
Concrete strength, NDT technique, benchmark, engineering practice

1. Introduction

Strength assessment of concrete in existing structures is a key issue, the significance of which has been highlighted with an increasing concern for seismic vulnerability or building ageing [1, 2]. Many non destructive testing (NDT) techniques can provide information about the material condition. Many possibilities also exist for deriving material properties from the measurements carried out [2, 7]. However, three main drawbacks have been identified which prevent common and efficient use of NDT for this purpose:

- the NDT measurements are influenced by many factors other than concrete strength, the most common being carbonation and water content, and there is no universal conversion model for estimating strength from NDT measurement values,
- even if some guidelines or standards describe how an individual measurement with a given technique must be carried out, there is no guideline explaining how to plan the investigation in relation with to what is looked for,
- no rules exist establishing the relationships between (a) the quantity of NDT information made available after tests, (b) the methodology for processing NDT results and (c) the level of accuracy/confidence on concrete strength estimates. Conversely, nobody can be

sure that a pre-defined investigation programme will guarantee the quality of concrete strength estimates.

Another challenge is the assessment of material variability (either detection of weaker areas or simply assessing the magnitude of variability within the structure, in order to better assess the characteristic and design strength values). It has been advocated that NDT could provide an easy way to answer these questions. However their contribution remains controversial and, in current practice, the optimal combination of destructive testing (DT) (mainly cores) and NDT remains an unresolved issue. Particularly, the European seismic code for existing buildings EC8-3 [8] prescribes that the estimation of in-situ strength has to be mainly based on cores drilled from the structure. However, NDT can effectively supplement coring thus permitting more economical and representative evaluation of the concrete properties throughout the whole structure [9]. For these reasons, RILEM has created the Technical Committee TC-ISC (In situ Strength of Concrete), with the aim of establishing guidelines for the efficient use of NDT for concrete strength assessment. The TC members have decided to carry out a benchmark whose objectives are to identify and compare the expert practices regarding NDT assessment of concrete strength. This covers two points: (a) how experts define and carry out an NDT investigation programme (How many NDT measurements? Where? With what quality? What rules for sampling? How many cores? etc...) and (b) how experts use information made available in order to assess the whole structure and derive strength values. This paper will detail the main elements of this benchmark while a companion paper [10], presented at the same conference, will simulate the strategies proposed by experts in order to evaluate their efficiency regarding concrete assessment.

2. Benchmarking NDT investigation approaches

2.1. Identification of NDT investigation approaches

In the following, “the NDT investigation approach” is considered as all means, methods and models carried out in order to estimate concrete properties. This covers: (a) the definition of the NDT measurement strategy (Where? How? How many?), (b) the measurement analysis and processing, (c) the development and use of a conversion model in order to derive strength estimates. The benchmark was based on synthetic simulations reproducing a series of twenty 3m-high columns defining a building floor. All data was unknown to the contributors who were asked to: (a) define their investigation strategy, (b) derive strength estimates from the data that they received. The process was iterated three times, using an increasing amount of monetary resources, in relation with the concept of knowledge level (KL) introduced in EC8-3. The contributors could choose between taking cores and/or performing NDT measurements, using ultrasonic pulse velocity, rebound hammer or pull-out tests. An important choice was the quality level of the measurement (low / average / high), based on the fact that a lower quality implies a lower cost but larger measurement errors. Each contributor was asked, at each step, to estimate both the average concrete strength and the standard deviation, before addressing his request for data of the next step.

2.2. Synthetic simulations for analyzing the efficiency of NDT

The idea of synthetic simulations has been advocated by one of the authors [10-12]. It consists in developing a “synthetic world” (fig. 1) reproducing as closely as possible the main features of the real world, including: (a) the material physical properties, (b) the NDT properties and measurement process. The main interest of synthetic simulations is twofold. On one hand, simulations can be repeated many times, in order to draw statistical analysis, reducing the influence of chance. On the other hand, the “true values”, while remaining unknown to the expert who assesses strength, can be revealed and the quality of the assessment can be evaluated by comparing “true” and estimated strength.

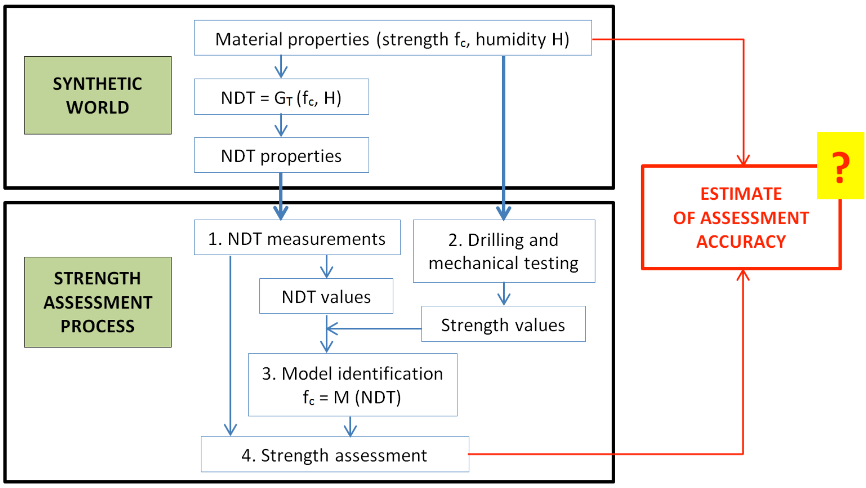

Figure 1. Illustration of the use of synthetic simulation for strength assessment

3. Case-study: estimation of concrete strength of a one-storey RC building

3.1. Rules of the game

3.1.1. Description of the structure and target of the assessment

The purpose of the benchmark was to analyze a simple structure, made of a one-level concrete frame, in which only columns were considered. Three levels of investigation were defined, respectively corresponding to three amounts of resource that could be spent for getting data (and to the three KL defined by EC8-3). The data (i.e. synthetic measurement results) were all affected by some measurement uncertainties, just as happens in practice on site. The synthetic structure was a one-level structure, made of a column-beam frame in which only columns were considered. The building shape was rectangular, with 20 columns arranged in five plane frames (from A to E) of four columns (from 1 to 4), with a regular spacing of six meters and a height of 3 meters. The cross-section of the columns was 35x35 cm^2. The building was un-heated in a temperate climate. From visual inspection, as expected considering casting modalities, a vertical gradient of properties is suspected in the columns, while there can be a humidity gradient between the extreme areas of the structure. The carbonation effect was neglected. For each data set (thus for each KL), two results were expected: (a) the estimate of the average strength on the whole structure, (b) the estimate of the variability in concrete strength, including, if possible, an estimate of lowest values.

3.1.2. Investigation techniques and costs

All investigation costs were defined in Cost Units (CU). The total amount of resource available varied with the KL value. It was respectively 20 CU for KL1, 40 CU for KL2 and 60 for KL3. The unit cost of a test depends on the test type and on the quality of the test (higher quality increases the cost but reduces the uncertainty of measurements – see Table 1). The definition of an investigation plan thus comes to combine the number, type and quality of techniques in order to respect the resource amount available. Once a contributor received a dataset of measurement results, he was free to use any method for deriving strength estimates (average, variability, extreme values). The available techniques were (the device is fixed):

- ultrasound measurements (US) direct from one side of the column to the opposite side, which provided a velocity in m/s,
- rebound hammer measurement (which could be performed on any of the four sides of the column), and provided a rebound number,
- pull-out test (Capo), which provided a load in kN,
- coring and mechanical tests (the cost covers both drilling, conservation and mechanical compression tests), provided a compressive strength in MPa. The location of cores was either predefined or conditional, based on the analysis of available NDT test results.

Table 1 . Unit costs and magnitude of uncertainty of all tests

Quality level		High	Average	Low
Magnitude of measurement uncertainty		X 0.5	1	X 2
Drilling and compression test	One core	14	10	7
Pull-out test	One test	5.6	4	2.8
US velocity test	Test result = average of 2 measurements	2.8	2	1.4
Rebound test	Test result = average of 10 measurements	1.4	1	0.7

3.2. Illustration of synthetic data

3.2.1 Generation process of synthetic data

This section briefly describes how the synthetic data was generated. It is assumed that only one concrete mix was used for all columns, but that due to the casting process three types of variability could exist: (a) a batch-to-batch variability corresponding to a difference between various series of columns. In a given batch, concrete strength f_c was generated by assuming a Gaussian distribution $N(f_{cm}, s(f_c))$, (b) a within-batch variability, corresponding to differences between columns of the same batch, (c) a within-column variability, with a degree of compaction that could vary according to the elevation. Moisture content was also assumed to vary between columns while remaining uniform in each column. Synthetic values for the ultrasonic velocity V, rebound number R and pull-put force F were produced using non linear relationships established after an in-depth literature review on the physics involved. In these relationships, velocity and rebound depend on both concrete strength and humidity [11, 12]. Just as in the real world, measurement errors were added to the synthetic values of V, R and F whose magnitudes depended on the quality of measurements. These errors were assumed to have Gaussian distribution $N(0, s(R) \text{ or } s(V) \text{ or } s(F))$. The standard deviation was respectively 100 m/s, 2 units and 1 kN for US, R and F for average quality measurements and changed according to Table 1 for low quality or high quality measurements. The same was true for compressive strength measured in the laboratory with $s(f_c)$ = 1.5 MPa for average quality measurements. The range of measurement errors was in close agreement with data from the literature. Finally, the “true” synthetic dataset was a series of NT sets of 4-values (V, R, F, f_c), with NT = 620 corresponding to 31, 10cm-spaced test locations on each of the 20 columns. The “measured” synthetic dataset was simply derived from this initial dataset by adding random errors. For deriving compression strength results, it was assumed that coring induced no bias or noise, and that changes in strength were only due to conservation conditions, with three possibilities: saturated, as on site and finally, air-dried.

3.2.2. Influence of measurement error

In the benchmark, the dataset generated by the random process was fixed (corresponding to a unique simulation), in order to provide the same database to all contributors. However, the test results available for each contributor differed for two main reasons: (a) the choice of different techniques, locations, quality and number of measurements, (b) chance, which may generate very different random measurement errors at two close test locations. A very important statement is that, because of this weight of chance, it is formally impossible to derive definitive conclusions about the absolute or relative efficiency of any investigation approach. This type of conclusion would require a large number of repetitions of the same process, which will be developed in the companion paper. The following series of figures illustrates some features of the synthetic data that have to be considered for assessment. Figure 2 provides the cumulative distribution of true and measured velocities, while figures 3 and 4 respectively illustrate the existing correlations between strength and rebound values for true values (fig. 3) and measured values (fig. 4). The scatter of properties (resulting from the variability at three scales) is clearly visible on figure 2 which also highlights the contribution of measurement errors. For instance, while the average velocity is 3985 m/s, the standard deviation amounts to 63 m/s on true values and increases to 116 m/s on measured values (average quality). The measurement uncertainty results in a doubling of apparent variability.

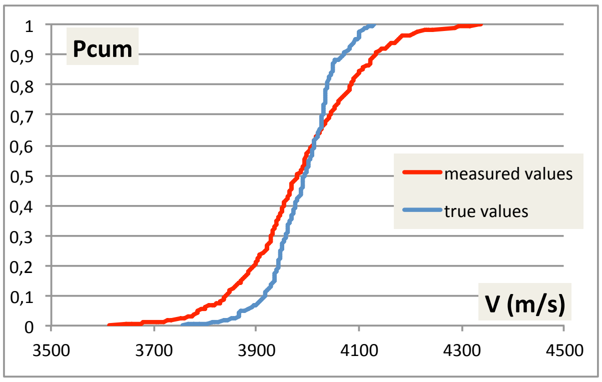

Figure 2. Cumulative distribution of true and measured velocities

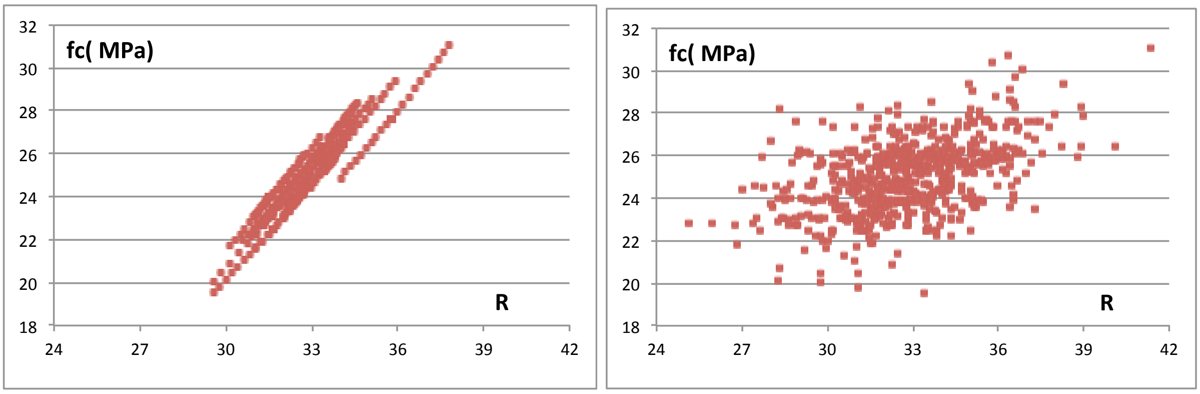

Figure 3. Correlation between strength and rebound (true values)

Figure 4. Correlation between strength and rebound (measured values)

Figures 3 and 4 show how the relationship visible in true synthetic data (here between strength and rebound) becomes much less visible because of measurement uncertainties. While on figure 3 each curve corresponds to a given column (different curves are explained by small variations in moisture content), all data form an extended cloud on figure 4. The correlation that will be identified from sampling this dataset may be signiticantly different than the true one. Since the conversion models will also be identified and calibrated from these data, these uncertainties will propagate. It must be noted that this fact, clearly highlighted here on synthetic data, is also relevant for real on-site investigation (the only difference being there that the « true » values can never be identified.

4. Analysis of investigation approaches

4.1. Main degrees of freedom in an investigation approach

A total of thirteen RILEM ISC-TC members have been involved in this benchmark. Some of them have proposed variants in their approaches. The resulting number of contributions is respectively 11, 11 and 10 with 14, 13, 12 variants for KL1, KL2 and KL3. Table 2 synthesizes how contributors combine destructive and non-destructive tests at the three KL. The two last lines indicate when NDT has been carried out on cores and when core location has been specified from a previous NDT series of measurements.

Table 2. Number of variants for each type of approach

	KL1	KL2	KL3
Without cores	4	1	0
Cores + V	3	1	2
Cores + R	3	5	3
Cores + combined	4	6	7
Total	14	13	12
NDT on cores	2	5	4
Conditional cores	2	2	2

Because of the limited amount of resources, assessment without cores was common at KL1 but became marginal for KL2. The overall average expense for cores was respectively 42 %, 50% and 53% of the total expense for the three KL. With increasing KL, the combination of two NDT techniques was more frequent and became dominant. Figure 5 synthesizes for all contributions the relative cost of NDT, while figure 6 illustrates the balance between cores and two most common NDT techniques (rebound and US velocity), at KL3 level.

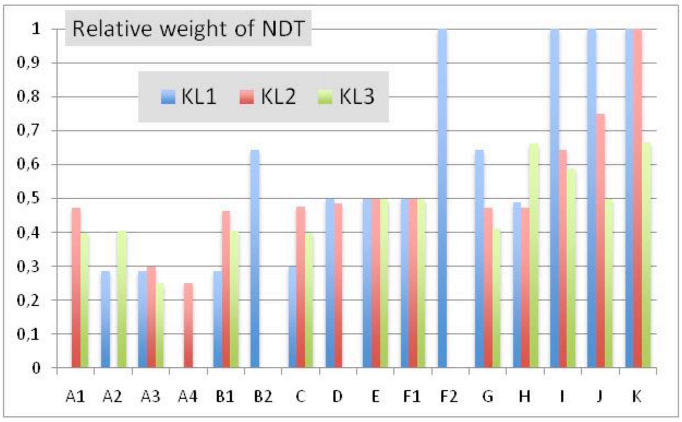

Figure 5. Relative weight (in cost) of NDT for all contributors and all Knowledge Levels (KL)

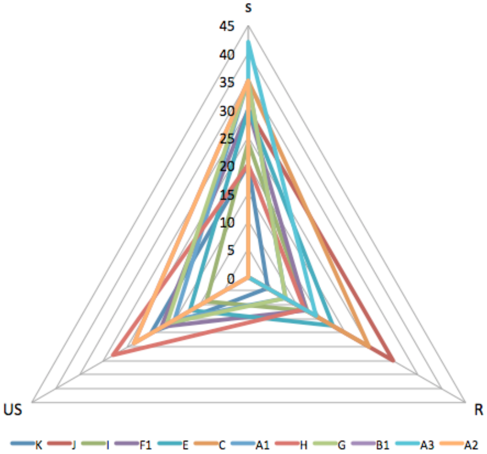

Figure 6. Cost balance at KL3 between cores, rebound and US velocity tests

For keeping anonymity, the contributions have been coded with letters, from A to K, according to the increasing percentage of NDT in cost. Numbers are used after the letter in the code for describing the variants. Since the total cost, for any technique, results from the multiplication of unit cost and number of tests, there is a compromise, causing a decrease in quality if one wants to increase the number. Some strategies concentrated on the quantity of data while others preferred to limit the data to a smaller number of high quality measurements.

4.2. Identification of a NDT investigation approach model

The analysis of all contributions makes it possible to derive a simplified model describing what could be called an "investigation approach". The total amount of resource S being defined, the model follows a 5-step approach (Figure 7):

- defining [1] what part of available resources is devoted to destructive tests (cores - DT) and to non destructive tests (NDT);
- defining [2] the quality of NDT and DT tests (low / average / high). Table 3 summarizes advantages and drawbacks of number *vs* quality;
- choosing [3] the specific location (column, Z) for each test. This can follow a more or less systematic spatial distribution or some kind of random distribution;
- choosing the type of NDT tests (rebound R , US velocities V or pull-out F), if some cores have to be taken, define (a *vs* b) if the location of cores is pre-defined or conditioned to NDT test results [4];
- if cores are to be taken, the last decision regards the conservation conditions [5], and especially the target moisture condition at the time of compression tests.

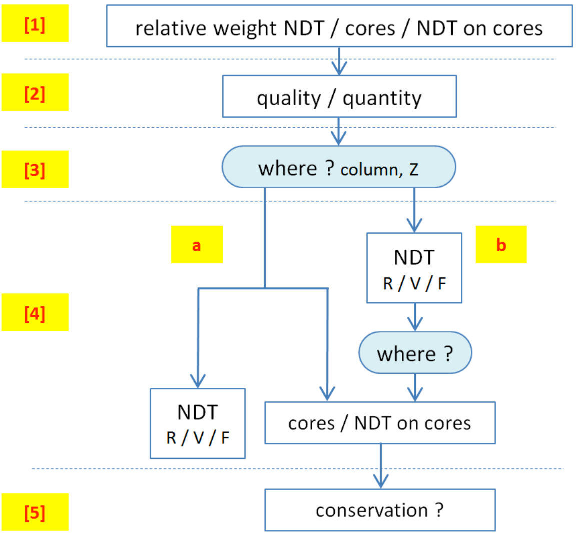

Figure 7. Model describing an investigation strategy

Table 3. Respective advantages and drawbacks of privileging number against quality

	Advantages	Drawbacks
Large number / low quality	Better coverage of the structure (spatial variation) Location of extrema	Standard deviation over estimated Problems for model fitting (uncertainty, influence of errors)
Small number / high quality	More accurate values Possibility to estimate the standard deviation	Problems for model fitting (low number of calibration data)

This series of decision steps quickly leads to a combined set of solutions, and the number of possible alternatives may be very large, especially when the amount of resources increases. For instance, regarding the two first steps, with the simplifying assumption that the two NDT methods have the same quality level, the number of possibilities at KL1 is very large. With low quality cores, as much as 68 alternatives are possible. Strategy F2 which combines 1 core (10 CU) to 3 US V measurements (6 CU) and 4 R measurements (4 CU), is only one of them.

4.3. Methods for deriving strength estimates

Generally speaking, three types of general approaches can be used in order to assess strength values from NDT measurements: (a) use a pre-defined model (prior model), without any calibration, (b) use a pre-defined model after calibration, (c) build a specific model. The calibration option is only possible when some reference values are available, usually core strength values. These options were respectively chosen by 5, 4 and 4 contributors at KL1, 3 and 5, by 4 contributors at KL2, and 1, 4 and 6 contributors at KL3. This shows a clear tendency towards the development of specific models when more information is available. Reversely, prior models were more used at KL1 and KL2. A prior model may be a law (or a chart) drawn from the literature or the expert's own experience. In the following sections the two other approaches will be detailed.

4.3.1. Approaches using a calibrated model

These approaches require a prior model M(NDT), from which the in situ (i) estimated (est) compressive strength $f_{c\ est\ i}$ is derived at each test location where an NDT test result $f_{c\ est\ m}$ is available. The prior model can be a chart or a law taken from literature or derived from the expert knowledge. The mean estimated value is calculated from the N estimated local values:

$$f_{c\ est\ m} = (\Sigma f_{c\ est\ i}) / N$$

A calibration factor k is then calculated, by comparing the estimated strength values $f_{c\ est\ m}$ to strength values $f_{c\ m}$ directly measured on cores (at NC test locations where a core has been taken): $k = (\Sigma f_{c\ m}) / (\Sigma f_{c\ est\ m})$

where $f_{c\ m}$ is the average of strength values obtained on cores. Finally, the model is updated:

$$M'(NDT) = k \text{ x } M(NDT)$$

A variant consisted of limiting the calculation of the estimated mean to points where NDT values correspond to the drilling of cores.

4.3.2. Approaches developing a specific model

A variety of models can be used, according to: (a) the number and type of NDT parameters used, (b) the mathematical expression of the model. The specificities of each model are not analyzed here, while we will limit to more general statements:

(a) since model parameters are identified by minimizing the error between estimated and measured strength on a limited set of cores, at least two cores are required for a monovariate conversion model (i.e. $f_c = a R^b$ or $f_c = a' \exp (b'V)$) while a bivariate model needs at least three parameters (i.e. $f_c = a R^b V^c$).

(b) the quality of fit tells nothing about the predictive ability of the model for estimating strength where only NDT results are available. Here, the role played by the sample size (N cores) and the measurement error is of prime importance.

(c) conditional coring, used for instance to take cores where NDT results revealed extreme values may be beneficial, since it can increase the range of variation of strength covered by the regression and have stabilizing effects on the model.

Some experts have developed interesting variants, illustrated on the few following examples.

Contributor K used data fusion to analyze strength estimates derived from different types of NDT measurements, by considering their overall consistency. The principle of data fusion is to, at the same time, calibrate all the different conversion models (for the different NDT used), on all measurement points, by maximizing the consistency between the assessment obtained by all techniques independently.

Contributor B used conditional coring, on the basis of an extensive series of NDT measurements (US velocities), the choice of core location being defined in order to cover the

range of variation of material properties as revealed by NDT results. Another interesting option was to compute the strength estimation with a weighted average between direct strength values (on cores) and strength estimates (through NDT and conversion models) while considering that these two sources of information have a different level of confidence.
Contributor I developed a specific approach combining the three possible NDT (R, V and pull-out) and cores (1 core at KL2 and 2 cores at KL3). The small number of cores prevents the identification of any relevant regression model, but it however provides some estimate of the expected average which are then used for calibration at KL2 and KL3, while pull-out is considered as a possible reference test (1 pull out at KL1, 3 pull out at KL3) which can provide an alternative « reference strength », by using a prior model, derived from literature. Where R and V measurements are available, a bi-variate model was calibrated against reference strength value (on cores or, if no core is available on pull out values). At points where only R is available a specified linear model $f_c(R)$ is fitted and used for estimating strength. At the end of the process, at KL2 and KL3 levels, one has four types of strength estimates: direct measurements on cores (respectively 0, 1 and 2 at the three KL), from pull out (resp. 1, 1, 3), from R and V combined (resp. 4, 4, 5), from R only (resp. 4, 5, 6). This approach takes profit from available information.

5. Analysis of benchmark results and perspectives

We focus here on few results obtained during the benchmark. Figures 8 and 9 respectively plot the estimates of average strength and standard deviation obtained at KL1 and KL3 (one contribution has been excluded, which predicted very large values).

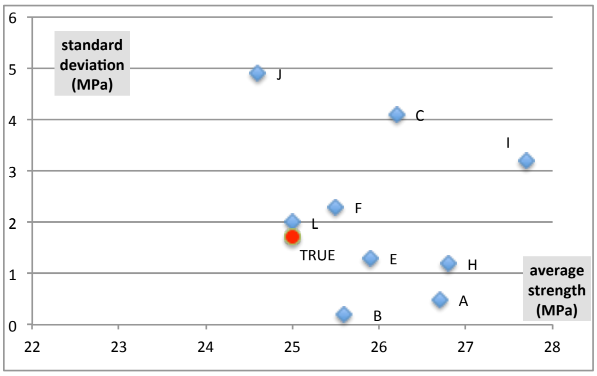

Figure 8. Comparison of estimations (average strength and standard deviation) to real values (KL1)

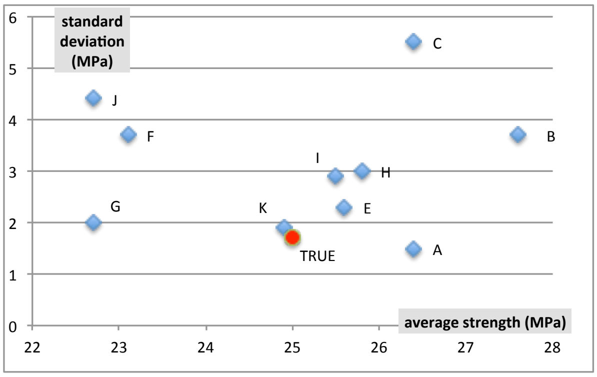

Figure 9. Comparison of estimations (average strength and standard deviation) to real values (KL3)

The three main results are that:

- the estimates appear to be scattered both on average strength and standard deviation,
- all approaches estimate the average strength with less than a 3MPa error (mean estimate = 24.0 MPa, sd on estimate = 1.6 MPa), which is clearly acceptable for practical purposes (however, the conclusion could be different for a more heterogeneous contrast in concrete properties),
- results on standard deviation are very scattered and when the KL increases show a general tendency to overestimation (which can be larger than 100%). This is clearly a handicap if the standard deviation estimation is used for estimating lower percentiles (or characteristic values).

The role of error measurements is highlighted by these conclusions: it is because of error measurements that the concrete variability tends to be overestimated (a refined analysis of

results confirms that the contributor who had privileged quantity of tests against quality of measurement tended to have the larger overestimations of variability).
It is however too soon to derive any type of conclusions on the basis of a single simulation. How and why each type of assessment strategy (defined as on Figure 7 and including how the strength estimates are derived) requires an in-depth analysis of all contributing factors. To go further, we will keep the synthetic simulations tool and manage to simulate the investigation strategies, which will make it possible to “play” many times with the same approach. Doing so, statistical results will be derived (average strength estimates and variability estimates), and the stability/robustness of the approach will be analyzed. In a second step, the focus will be given to a more systematic analysis of key factors: range of material variability, type of conversion model, magnitude of the measurement error. The first part of this work is presented in our companion paper [10].

Acknowledgements

This work was carried out by the RILEM Technical Committee TC-249 ISC which aims at preparing guidelines for more efficient and reliable NDT strength assessment of concrete.

References

[1] Masi A. and Vona M., Vulnerability assessment of gravity-load designed RC buildings: evaluation of seismic capacity trough non linear dynamic analyses, Engineering Structures, 45, 257–269, 2012.

[2] Biondi S, Candigliota E., In situ tests for seismic assessment of RC structures, in. 14th World Conf. on Earthquake Engineering, Beijing, 12–17/10/2008.

[3] Facaoaru I., Draft recommendation for in situ concrete strength determination by combined non-destructive methods, RILEM Draft Recommendation, Mat. Str., 26:43–9, 1993.

[4] IAEA International Atomic Energy Agency, Guidebook on non-destructive testing of concrete structures, testing training course series n° 17, 2002.

[5] Luprano V., Caretto F., Labia N., Ciniglio G., Tatì A., Tundo A., Pfister V., Non destructive Test Methods for evaluation of concrete: the case study of Punta Perotti, Int. Conf. on Durability of Building Materials and Components ISTANBUL, Turkey 11-14 May 2008

[6] Szilagyi K., Borosnyoi A., 50 years of experience with the Schmidt rebound hammer, Concrete Structures, vol. 10, 2009, p. 46-56.

[7] EN 13791, Assessment of in situ compressive strength in structures and precast concrete, CEN, Brussels, 2007.

[8] CEN, 2005. Eurocode 8: design of structures for earthquake resistance—Part 3: assessment and retrofitting of buildings, EN 1998-3, June 2005, Brussels.

[9] Masi A., Chiauzzi L., An experimental study on the within-member variability of in-situ concrete strength in RC building structures, Const. Build. Mat., 47, 951–961, Oct. 2013.

[10] Al Wash M., Breysse D., Sbartai D., Analysis of key influent factors of investigation for concrete strength estimation in relation with an international benchmark, NDT-CE conf. Berlin, Sept. 2015.

[11] Breysse D., Nondestructive evaluation of concrete strength: an historical review and new perspective by combining NDT methods Const. Build. Mat., vol. 33, n° 21, 2012, p. 139-163.

[12] Breysse D., Martinez-Fernandez J., Assessing concrete strength with rebound hammer: review of key issues and ideas for more reliable conclusions, Mat. Str., 47, 9, 1589-1604, 2014.

Author Index