Review of "Machine learning and micromechanics as allies to establish composition-property correlations in cement pastes"

Tulio Honorio, Sofiane Ait Hamadouche, Amélie Fau, Enrico Masoero, Giovanni Di Luzio, Anna Pandolfi

Open Review
Journal of Theoretical, Computational and Applied Mechanics

Identifiers
Open Review
OAI hal-03936241
Reviewed Article
DOI 10.46298/jtcam.9830

History
Review Oct 1, 2022
Rebuttal Oct 13, 2022

Review of "Machine learning and micromechanics as allies to establish composition-property correlations in cement pastes"

Túlio Honório[1], Sofiane Ait Hamadouche[1], Amélie Fau[1], Enrico Masoero[2,R], Giovanni Di Luzio[3,R], and Anna Pandolfi[3,E]

[1] Université Paris-Saclay, CentraleSupélec, ENS Paris-Saclay, CNRS, LMPS – Laboratoire de Mécanique Paris-Saclay, Gif-sur-Yvette, France
[2] School of Engineering, Cardiff University, The Parade, Cardiff, United Kingdom
[3] Department of Civil and Environmental Engineering, Politecnico di Milano, Milan, Italy
[R] Reviewer
[E] Editor

Review of version 1

Permalink: hal-03723418v1

Reviewer 1 (Enrico Masoero)

Reviewer This is an interesting manuscript, exploring how representative and predictive are a number of machine learning techniques and micromechanical models, when applied to existing experimental data on elastic properties of cement pastes. The presented research is original and, as pointed out by the authors, may stimulate various follow-up endeavours. The text is very well structured and the logical progression through it is easy to follow. I only have a few minor comments and requests for clarification: There are various minor typos throughout the paper: please read through and fix them before publication.

Authors We have revised the paper thoroughly and corrected a few typos. We thank the reviewer for the attentive reading.

Reviewer The whole explanation of the machine learning techniques is limited to the overview in Section 2.2. I suspect that each ML method will have associated parameters for the user to decide, e.g. number of layers in a neural network or else. The authors should make sure that all the information required to potentially reproduce their results are included in the manuscript or in the supplemental information.

Authors The implementation of the various methods in Mathematica leaves the user the possibility to impose the associated parameters or use in-built optimized procedures to determine these parameters. The last option seemed more appropriate for us because it reduced the number of cases to be tested. The number of layers in ANN is indeed one critical parameter not cited in the original versions. In the revised version, we have added the following sentences: *The implementation of the various methods in Mathematica leaves the user the possibility to impose the associated parameters or use in-built optimized procedures to determine these parameters. The last option seemed more appropriate for us because it reduced the number of cases to be tested. For reproducibility reasons, we provide the information used by each method as supplementary material. For example, ANN uses 2 layers for E and G predictions and 8 layers for K and ν predictions; DT uses between 23-27 nodes and 12-14 leaves. The specific number of nodes and leaves is provided to GBT*

and RF. Also, we provide a PDF with the detail of the parameterization used.

Reviewer In section 3.3, below equation 1 (as an aside: row numbers would help the review process): the authors write "$C^0 = C_0$". However, C_0 is not defined. Please clarify.

Authors In the case of MT estimate, the matrix is assumed as the phase mediating the interactions. C_0 refers to matrix properties. We have rewritten the sentence for clarity: $C^0 = C_0$ *for the MT scheme, where* C_0 *refers to the matrix stiffness tensor (i.e., the subscript* 0 *stands for matrix properties).*

Reviewer There are issues with Table 3. The caption says that the table contains "fraction of values" but then there are values like 4.1 and 3.5, which would mean 410% and 350% if they were actually fractions. The manuscript text suggests that the values are actually percentages, but then are those for the Poisson modulus really 0.002% (i.e. 2 data point out of 100,000 experimental values?). Please clarify

Authors We have corrected the following issues:

- The values should be given % instead of in "fractions", therefore the caption has been changed accordingly and the "%" added to the values for clarity.
- The values for the Poisson ratio were wrongly reported as fractions. The correct should be 3/365 = 0.82% and 1/365 = 0.27%

Reviewer Still linked to table 3: the authors mention that in some cases a 15% uncertainty has been accounted for. How exactly did the authors account for that?

Authors We have added the following clarification: *... we also provide bounds estimation accounting for an average* 15% *uncertainty (i.e., the bulk and shear moduli are increased of* 15% *before upper bound calculation).*

Reviewer Again related to Table 3: why did the authors decide to consider only the Powers and TJ hydration models here, and not the KHP?

Authors KHP is based on the Powers model but with a separation from the gel and capillary water. Under the assumption of gel and capillary water having the same behavior, the bounds computed for Powers and KHP are identical. The following clarification has been added:*The bounds are computed for Powers and TJ models (with the assumption of gel and capillary water having the same behavior, the bounds computed for Powers and KHP are identical).*

Reviewer The authors have used MB to create missing data and strengthen ML: this is very interesting. There is also another way in which ML and MB may work together, i.e. starting from a mechanistic template and using ML to fill in for missing parameters or constitutive laws. These kinds of hybrid mechanistic - ML models have been tried in other fields: not sure about cement and concrete research. The authors may want to add this possibility to their outlook in the conclusion (this is not a strong request: happy also if the authors prefer to leave this out).

Authors We thank the reviewer for this suggestion. One of the co-authors has already developed such strategies in other contexts. This idea has been added as a perspective as follows *A reviewer of this work indicated another exciting way in which MB and ML can be allies: starting from a mechanistic template - as the one provided by MB methods - and using ML to provide missing parameters or even the constitutive relation. These kinds of hybrid mechanistic - ML models have been proposed in other fields (e.g., Fuhg et al. 2021; Fuhg et al. 2023), but in the case of cement and concrete research, they are yet to be fully worked out. In this direction, one way of using ML to improve MB estimates would be to use data to determine more appropriate localization relations (i.e., the relations determining the contribution of each phase to the effective behavior) instead of relying on the ones provided in the classic homogenization schemes.*

Reviewer 2 (Giovanni Di Luzio)

Reviewer The paper presents a study that compares machine learning techniques and micromechanics-based methods to predict the material mechanical properties from the material composition. The paper presents a very interesting promising study that perfectly fits the scope of the journal. The manuscript is well written, and the reader can easily follow the presented arguments. The

reviewer has some comments that are reported in the following and need to be resolved before the manuscript publication.

Reviewer Section 2.1. The authors in the construction of the database considered cement composition, water-cement ratio, age, and degree of hydration. The authors should explain the reasons for their choice and, for example, they do not consider admixtures, curing conditions, etc. For example, the effect of the temperature on the type of C-S-H gel that is produced is something that must be taken into account.

Authors We have limited the number of inputs to the ones that could be used in both ML and MB methods (to the best of our knowledge, properly accounting for curing conditions and admixture content is yet to be fully worked out with MB methods). Also, we decided to use inputs for which complete data exists in the studies consulted to build the database. The following sentences have been added to clarify this choice: *Other input of interest for the formulation, such as admixtures, curing conditions (including temperature), etc., are not considered because of the lack of full data and the inadequacy of MB methods to date to take into account these factors properly. Of course, future work might focus on introducing these effects as input in ML-based strategies.*

Reviewer The authors should also comment and discuss better the fact that they assumed the Poisson ratio of the pore equal to 0.5, i.e., perfect in-compressible material (see Table 2). This means that the authors assume the pore fully saturated, and so also the hydration is under saturation condition. Is this the case for all the experimental tests considered in the database?

Authors A Poisson ratio tending to 0.5 is the typical value for fluids with zero shear rigidity (indeed, for an isotropic materials $\nu = (3K - 2G)/(6K + 2G)$, with $G = 0$, $\nu = 0.5$). The same value is generally valid for liquids and gases. Adopting $K = 2.18$ GPa for the porosity indeed presupposes that the pores contribute to the mechanical response and are saturated with liquid. We have also tested also the "inactive porosity assumption" (i.e., $K = 0$ for the porosity), and the results of MB methods deviated slightly more from the experimental data. Since the references consulted for the database do not provide, in all cases details about curing conditions, we decided to adopt $K = 2.18$ GPa. The following sentences have been added to clarify these points: *A Poisson ratio of* 0.5 *is the typical value for fluids with zero shear rigidity. Adopting* $K = 2.18$ *GPa for the porosity presupposes that the pores contribute to the mechanical response and are saturated with liquid water (i.e. the "active porosity assumption"). We have tested also the "inactive porosity assumption" (i.e.,* $K = 0$ *for the porosity), and the results of MB methods exhibited a slightly large deviations from the experimental data. Since the references consulted for the database do not provide in all cases details about curing conditions, we decided to adopt hereon "active porosity assumption".*

Reviewer Section 4.2, it is recommended to estimate the errors also using the coefficient of determination or R-squared (R^2). It is a scale-free quantity and is very precise in telling how well your model explains a phenomenon.

Authors In the revised version, we present the coefficient of determination (R2) for ML (new Table 6) and MB (new Table 9) estimates. Accompanying the new Table 6, we have added: *Table 6 shows the (mean) coefficient of determination* R^2 *of the elastic constants obtained from K-fold cross-validation technique based on 5-fold or 10-fold. The* R^2 *is a scale-free quantity, quantifying how a model explains a phenomenon. All models yield high* R^2 *(closer or higher than* 0.9*) for the elastic moduli, while the* R^2 *for* ν *are overall lower. ANN predictions exhibit higher* R^2 *in most cases.* Accompanying the new Table 9, we have added: *Table 9 the coefficients of determination (*R^2*) of elastic constants. The estimates with ellipsoidal inclusions using Powers model,* MT_{ellip} *Powers and* SC_{ellip} *Powers, showed a higher* R^2 *for the elastic moduli. SC Powers (with spherical inclusions) also exhibit one of the higher* R^2*. The* R^2 *is overall low, showing the difficulty of the model to properly capture this elastic constant. Since two elastic constants are sufficient to fully determine an isotropic behavior, this last observation suggests that dealing with the elastic moduli (E, K or G) is a strategy less prone to errors.*

Reviewer Section 4.2, in the homogenization predictions the elastic constants are, in most cases, better only for a DOH $\leqslant 0.7$ (young ages), Fig. 4. The authors should better comment and discuss this point because this is due to the limitation of the MB models that don't take into account aging

processes, such as C-S-H gel densification, C-S-H polymerization, etc.

Authors We have corrected the sentence below which should refer to later ages *better when only the observations in the training dataset with DOH* > 0.7 *(i.e. associated with late ages).* And the following sentence has been added pointing out the limitation in MB early-age estimates *The late ages estimates exhibiting lower errors overall means that a better description of hydration kinetics, phase assemblage, and particular effects associated, such as C-S-H gel densification and C-S-H structural and compositional variability, could improve the estimates.*

Reviewer The authors showed well the performance in the prediction of both ML and MB methods. However, it is not clear how the two approaches can be "allies". The only "alliance" proposed by the authors is in the "extension of the database" using the MB. However, the MB approaches present some limitation (see the previous point).

Authors In the original version, we have proposed the following points in ML and Mb could work as allies:

1. MB as a filter for curating databases for ML (Sections 3.4 and 4.2).
2. MB to provide missing data for ML (section 4.3).

We agree with the reviewer that MB approaches present some limitations. As pointed out by Reviewer 1, an interesting way in which ML and MB can be used as allies is by using ML to provide missing parameters or even the constitutive relations (obtained, for example, with MB methods). Some sentences have been added to highlight this idea as a perspective of this work.

Review of version 2

Permalink: hal-03723418v2

Reviewer The reviewers are fully satisfied with the revised version of the manuscript, which is now accepted for publication.

Editor's assessment (Anna Pandolfi)

The paper presents an interesting study perfectly in line with the scope of the journal. The manuscript is well written, and the arguments are easily followed by the readers. Both reviewers made some comments that have been clearly and exhaustively addressed in the revised version of the paper. It is overall a very nice contribution.

References

Fuhg, J., C. Böhm, N. Bouklas, A. Fau, P. Wriggers, and M. Marino (2021). Model-data-driven constitutive responses: Application to a multiscale computational framework. *International Journal of Engineering Science*. [DOI], 2104.02650.

Fuhg, J. N., A. Fau, N. Bouklas, and M. Marino (2023). Enhancing phenomenological yield functions with data: challenges and opportunities. *European Journal of Mechanics - A/Solids*. [DOI], hal-03619186.