Predicting performance of concrete structures by machine learning

Zsuzsanna Szabó

Environmental physicist-geochemist, PhD,
Data Science and R Enthusiast, co-organizer of R-Ladies Budapest

Department of Hydrogeology and Geochemistry, Mining and Geological Survey of Hungary, Budapest
MTA Premium Postdoctorate Research Program, Office of Funded Research Groups, Hungarian Academy of Sciences, Budapest

zsszabo86@gmail.com

1. Introduction

Being able to predict the degradation of concrete is important in economic, environmental and human safety point of view. An academic research project titled "Geochemical interactions of concrete in core samples, experiments and models" started in 2017 October (*Szabó et al. 2017 and 2018*) which aims to understand and/or simulate geochemical interactions (mineral dissolution and precipitation processes) in concrete–rock–water systems. One of the several research tools available for the prediction of these reactions is numerical geochemical modeling (based on chemical equations, equilibrium and rate constants). To support these theoretical models a data based solution is also explored by machine learning algorithms (empirical models), first, applied on a publicly available dataset (*Yeh 1998*, *Kuhn and Johnson 2015*, *UCI ML online*). By reproducing the analysis of these data, lessons to learn are collected for the geochemical perspective project.

2. The dataset and exploratory data analysis

Package **'AppliedPredictiveModeling'**, dataset **'concrete'**, dataframe **'mixtures'**:
Record of 1030 laboratory experiments with proportional concrete compositions and age
Response column: **compressive strength** (*Fig.1.*, *Tab.1.*)

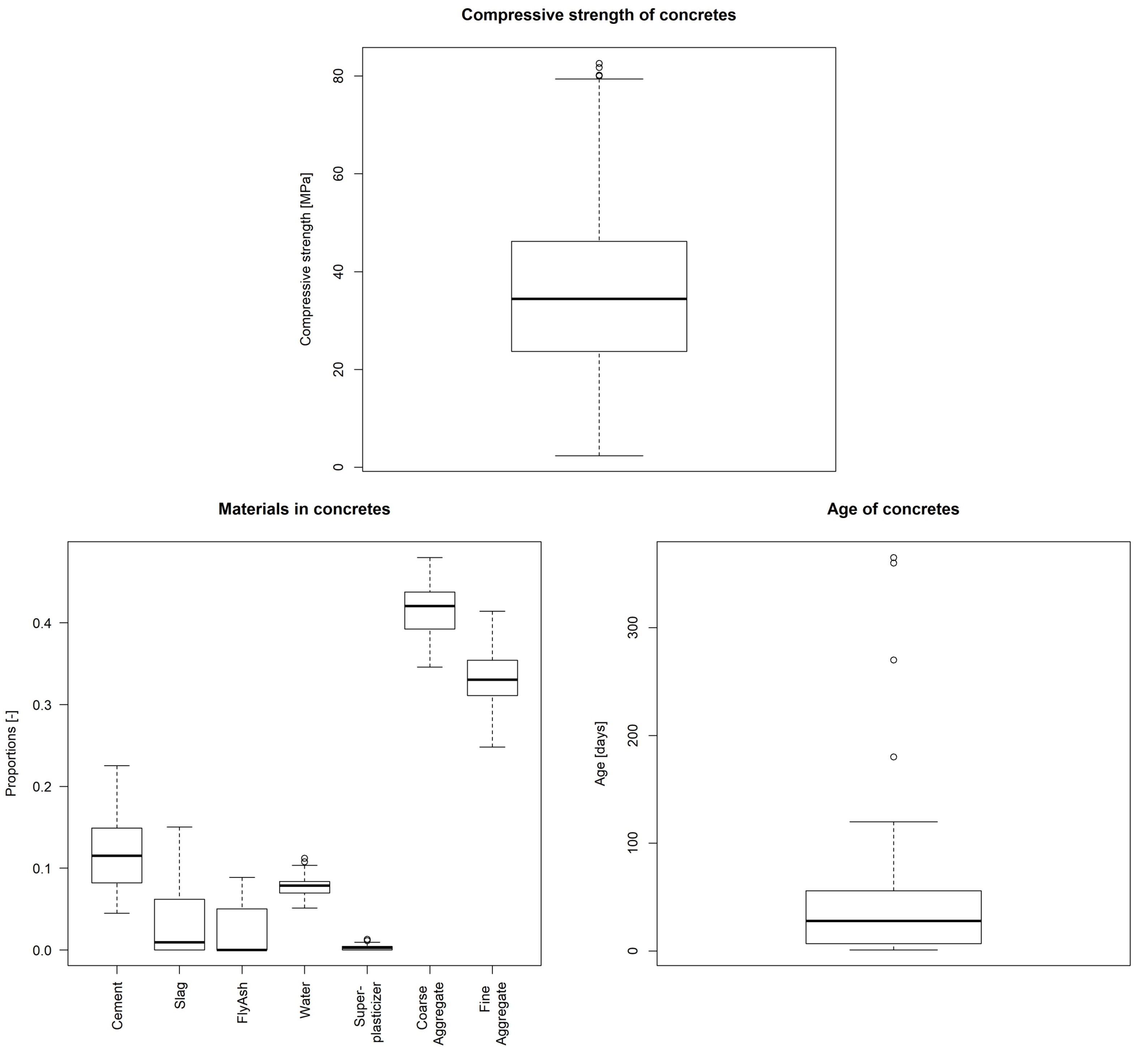

Fig.1.: Boxplots of concrete compressive strength, proportional composition and age in the 'mixtures' dataframe

Cement	BlastFurnaceSlag	FlyAsh	Water	Superplasticizer	CoarseAggregate	FineAggregate	Age
0.48	0.12	-0.11	-0.31	0.35	-0.32	-0.29	0.33

Tab.1.: Pearson correlation coefficients for compressive strength data

3. Preparations and tested machine learning algorithms

train set : test set
80% 20%

GLM:
Generalized Linear Models
Package 'caret'

GBM:
Gradient Boosting Machine
Package 'caret'

RF:
Random Forest
Package 'caret'

NN:
Artificial Neural Network
Package 'neuralnet'
min-max scaling
predictor selection
parameterization (e.g. hidden = 100)
etc.

Ahmad et al. 2017
NN performs better
RF for parameterization support

4. Results

4.1. Overall performance of tested algorithms: Tab.2.

	GLM		GBM		RF		NN		NN in *Yeh 1998*	
	train set	test set	train set	test set	train set	test set	train set	test set	train set**	test set**
R^2	0.615	0.606	0.931	0.927	0.981	0.938	0.97	0.879	0.917-0.945	0.814-0.922
RMSE	10.233	10.975	4.357	4.789	2.375	4.487	0.036*	0.083*	-	-

* *due to min-max scaling not comparable to other RMSE*
** *different train and test sets than in this study*

Tab.2.: Coefficients of determination (R^2) and root mean square errors (RMSE) of tested machine learning algorithms in this study and NN results of Yeh (1998)

4.2. The best performing model: RF, Fig.2.

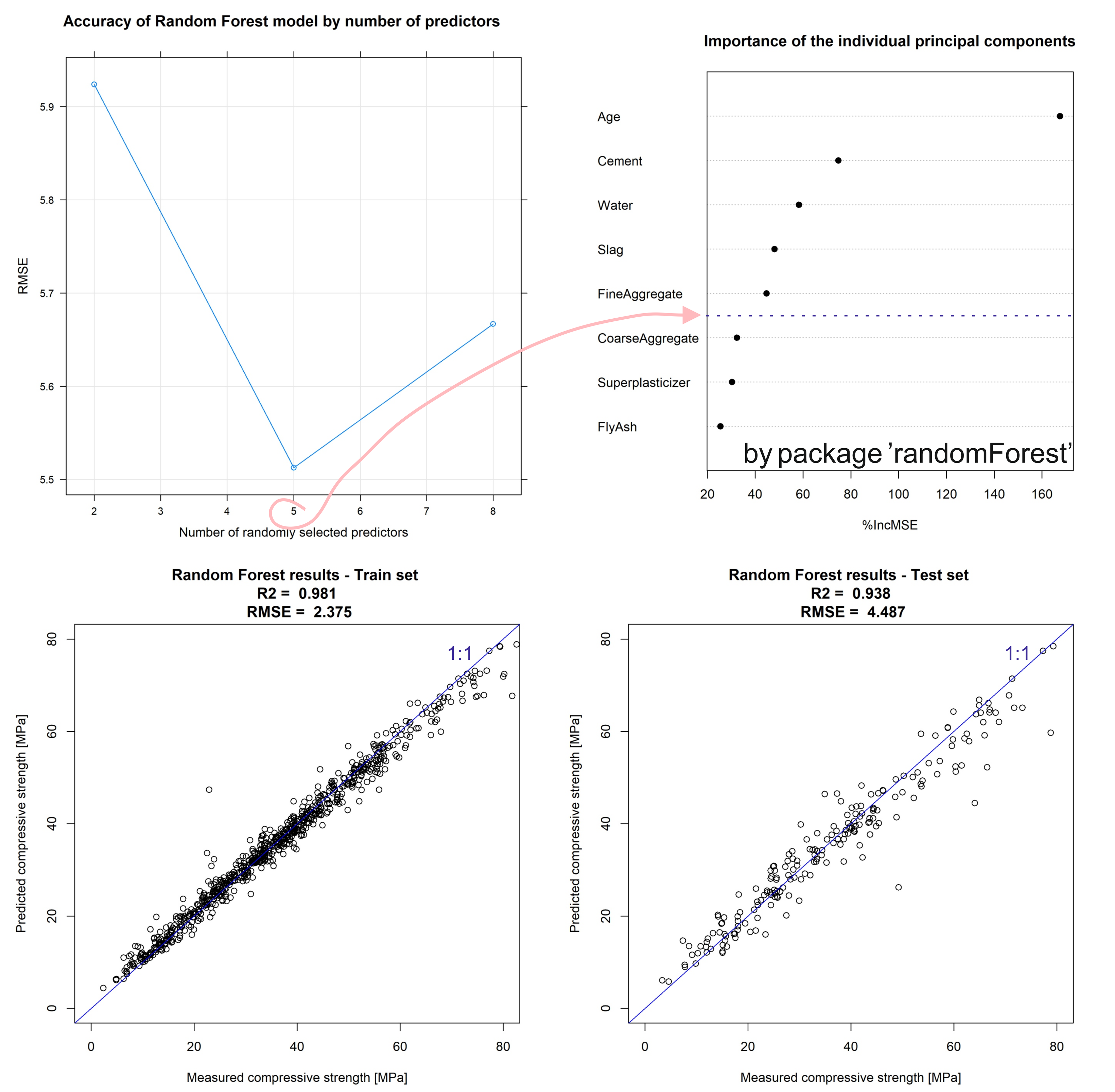

Fig.2.: Results of the RF model which takes into account the 5 most important components (age, proportions of cement, water, slag and fine aggregate; %IncMSE - % increase in mean square error) and produces the best fits for both the train and test sets

5. Conclusions

The analyzed dataset contains composition and compressive strength data for **concretes younger than a year** which limits the general applicability of any fitted models. The **Random Forest** algorithm ('caret' package) provides the best fit among the tested machine learning methods. It well predicts (**R^2=0.938**) the **compressive strength** of similar concretes (composition, age) **based on** their **age and proportions of cement, water, slag and fine aggregate** used in their preparation. To conclude about parameter importance, Pearson correlation coefficients are misleading. Based on the experiences of this analysis, efforts are made to access a local dataset which is suitable to use for prediction of concrete integrity for longer ages and in different reactive conditions.

Acknowledgment

The author thanks the guidance of Eszter Windhager-Pokol in the preparation of the analysis. Further thanks goes to Coursera and Johns Hopkins University who published the Data Science Specialization and made possible to easily learn about R and machine learning. The study was done within the framework of project titled "Geochemical interactions of concrete in core samples, experiments and models" financed by the MTA Premium Postdoctorate Research Program, Hungarian Academy of Sciences.

References

Ahmad, M.W., Mourshed, M., Rezgui, Y. (2017) *Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings, 147, 77-89.*

Kuhn, M., Johnson, K. (2015) *R Package 'AppliedPredictiveModeling', CRAN, ftp://cran.r-project.org/pub/R/web/packages/AppliedPredictiveModeling/AppliedPredictiveModeling.pdf*

Szabó, Zs., Udvardi, B., Kónya, P., Gál, N., Király, E., Török, P., Szabó, Cs., Falus, Gy. (2017) *Geokémiai folyamatok a Bátaapáti Nemzeti Radioaktívhulladék-tároló gránit-beton határfelületén. 154. In: Dégi et al., 8. Kőzettani és Geokémiai Vándorgyűlés, MFGI, ISBN: 978-963-671-311-9*

Szabó, Zs., Király, Cs., Szabó, Cs., Falus, Gy. (2018) *Optical and electron microscopic observations at concrete–granite interface. (in Hungarian with English abstract).153. In: Török et al., Engineering Geology Rock Mechanics 2018, BME, ISBN: 978-615-5086-11-3*

UCI ML *(UCI Machine Learning Repository), http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength*

Yeh, I.C. (1998). *Modeling of strength of high-performance concrete using artificial neural networks. Cement and Concrete Research, 28(12), 1797-1808.*