Towards Data Based Corrosion Analysis of Concrete with Supervised Machine Learning

Conference Paper · July 2018

Christoph Völker
Bundesanstalt für Materialforschung und -prüfung

Sabine Kruschwitz
Bundesanstalt für Materialforschung und -prüfung

Gino Ebell
Bundesanstalt für Materialforschung und -prüfung

Towards Data Based Corrosion Analysis of Concrete with Supervised Machine Learning

Christoph Völker, Corresponding Author
Bundesanstalt für Materialforschung und -prüfung (BAM)
Unter den Eichen 87, 12205 Berlin, Germany
Tel: +49 30 8104-4211, Fax: +49 30 8104-1447; Email: christoph.voelker@bam.de

Sabine Kruschwitz
Technische Universität Berlin
Gustav-Meyer-Allee 25, 13355 Berlin, Germany
Tel: +49 30 314-72109 Fax: +49 30 314-72110; Email: kruschwitz@tu-berlin.de

Jiawei Shen
Southwest Jiaotong University
No. 111, North 1st Section, 2nd Ring Road, 610031, Chengdu, Sichuan, China
Tel: +86-28-66366293 Fax: 86-28-66367260; Email: Jiawei.shen@bam.de

Gino, Ebell,
Bundesanstalt für Materialforschung und -prüfung (BAM)
Unter den Eichen 87, 12205 Berlin, Germany
Tel: +49 30 8104-4353; Email: gino.ebell@bam.de

Word count: 3,024 words

Submission date: July 19th, 2018

Towards Data Based Corrosion Analysis of Concrete with Supervised Machine Learning

Christoph Völker[1], Sabine Kruschwitz[2], Jiawei Shen[3], and Gino Ebell[1]

[1]Bundesanstalt für Materialforschung und -prüfung (BAM)
Unter den Eichen 87, 12205 Berlin, Germany
Tel: +49 30 8104-4211, Fax: +49 30 8104-1447; email: christoph.voelker@bam.de

[2]Technische Universität Berlin
Gustav-Meyer-Allee 25, 13355 Berlin, Germany
Tel: +49 30 314-72109 Fax: +49 30 314-72110; email: kruschwitz@tu-berlin.de

[3]Southwest Jiaotong University
No. 111, North 1st Section, 2nd Ring Road, 610031, Chengdu, Sichuan, China
Tel: +86-28-66366293 Fax: 86-28-66367260; email: Jiawei.shen@bam.de

ABSTRACT

Half-Cell-Potential Mapping (HP) is the most popular non-destructive testing (NDT) method for the detection of active corrosion in reinforced concrete. HP is influenced by parameters such as moisture and chloride gradients in the component. The sensitivity to the spatially small, but dangerous pitting is low. In this study we show how additional measurement information can be used with multi-sensor data fusion to improve the detection performance and to automate data evaluation. The fusion is based on supervised machine learning (SML). SML are methods that recognize relationships in (sensor) data based on given labels. We use SML to distinguish "defective" and "intact" labeled areas in our dataset. It consists of 18 measurement - each contains HP, ground radar, microwave moisture and Wenner resistance data. Exact labels for changing environmental conditions were available in a laboratory study on a reinforced concrete slab, which deteriorated controlled and accelerated. The deterioration progress was monitored continuously and corrosion was generated targeted at a predefined location. The detection results are quantified and statistically evaluated. The SML results shows a significant improvement over the best single method (HP).

Keywords: Data fusion, Half-cell potential mapping, supervised machine learning, non-destuctive corrosion testing of concrete, data based decision making

INTRODUCTION

The corrosion of reinforced concrete is one of the most important damage mechanisms in civil engineering. Infrastructure and buildings in maritime environments are particularly affected, as they are exposed to large quantities of aggressive agents from road salt or seawater. Half-cell potential mapping (HP) [1, 2] is the most commonly used non-destructive testing (NDT) -technique for the detection of active corrosion. A reference electrode at the components surface measures the electrical potential against the rebar. A steep drop of the measurement values indicate corrosion sites. However, small changes in the electrical properties of the concrete cover - e.g. due to moisture or salt ingress – mask the signal, such that the actual measurement information is lost [1, 3]. This makes the detection of locally concentrated pitting corrosion particularly challenging. However, this type of corrosion is particularly dangerous due to rapid cross-section losses. However, this type of corrosion poses a particularly high risk to structural integrity due to the rapid loss of cross-section at the rebar. Although many of the relevant environmental factors can be quantified with additional NDT methods, there is no deeper physical understanding of the interactions between the various sensors that could be used to improve the evaluation.

To close this gap we present a data-driven multi-sensor fusion approach that enables the joint evaluation of multi-sensory data. Data-driven means that inter-senor-relationships are derived from measured data without the requirement of a physical model. Supervised machine learning (SML) algorithms can be used to recognize (or learn) these relations by optimizing mathematical functions between pairs of sensor data and labels (that contain the sought information, e.g. defect/intact) in so-called training.

To obtain training data we conducted a large-scale experiment that acceleratedly simulates the life cycle of a salt-exposed concrete component under various controlled environmental conditions. The experiment was monitored

with NDT methods that are commonly used for corrosion detection, namely ground-penetrating radar (GPR) [4], HP, Wenner-resistivity (WR) [5]and microwave-moisture (MW) [6].
We compare the differences in the decision making process between a basic SML-method named Logistic Regression (LR) to a more advanced method named Decision Tree (DT). Their performance is quantified in terms of true-positive-rate (TPR) and false-positive rate (FPR) and compared to the evaluation according to the American standard ASTM C876 [2] and the German standard DGZfP Merkblatt B3 [1].

EXPERIMENT

The key requirements for suitable training data are that they contain the variability that represents all expected scenarios in the later application (to ensure the genrelizability of the inference) and a clear reference (to be able to assign a label to each data point). To address the first, data was collected on a large scale concrete slab that has progressively been exposed to moisture and salts until eventual corrosion (see figure 1, left). The experiment covers the deterioration stages in a components life cycle, namely an undamaged reference condition (specimen after concreting), a used condition (specimen contains chlorides, no corrosion) and a defective condition (specimen contains chlorides and corrodes). Additionaly, the influence of different degrees of compaction was taken into account by conducting two measurement campaigns: the first was carried out on the less dense top side - and the second was repeated on the denser bottom side of the slab.
Because penetration of salts into considerable depths is a process that may take years without further measures [7] the process was accelerated by the polarization method [8, 9]. To assure unambigious data labeling and prevent randomly occurring corrosion the salt ingress was continuously monitored. Eventually corrosion was initiated at a dedicated rebar rod (see figure 1, right) by anodic polarization beyond the critical corrosion-inducing potential. After initiation, the corrosion activity was monitored by means of the current flowing between the corrosion-rod and rebar cage via a shunt resistor. The experimental procedure is further detailed in [10, 11].
GPR measurements were carried out using a GSSI, Inc. SIR20 device with a 2 GHz palm antenna [12]. The data consist of two perpendicular polarizations per measurement that were collected subsequently with an automated scanner system with a lateral measurement spacing of 5 mm and a line spacing of 2 cm. The HP, WR and MW measurements were collected manually along a predefined measuring grid with a spacing of 10 cm. The PM data were collected using a Canin+ corrosion analysis system from Proceq [13]. The reference electrode is a copper sulfate rod electrode. The WR measurements were collected using the Resipod probe from Proceq [13] and the MW measurements were performed using an ID10 probe from HF-Sensors [14].

Figure1: Corrosion specimen. Left: specimen with saline water solution during potentiostatic chloride ingress. Right: Corrosion rods for top and bottom of the specimen before concreting.

DATA-BASED DECISION MAKING

The data-based multi-sensor fusion is conducted on feature level. Features are signal parameters that are significantly influenced by the respective defect. Table 1 lists the features that were extracted from the NDT-methods. The decision making between intact and defect areas (so called classififcation) is conducted in feature space (FS) - a mathematical space in which the coordinates of each pixel correspond to the normalized feature values at one measurement point. The course of the function that distinguishes between intact and defect pixels in

FS is called decision boundary (DB). SML-algorithms differ in how the DB is obtainend, but most impotantly in the possible complexity of the DBs shape.
LR is a binary classifier first published by Strother et al. [15] in 1967. Its DB is described by a simple linear regression function. During training the regression parameters theta $\boldsymbol{\theta}$ are optimized to minimize incorrect classification. The classification error is estimated by a cost function that scores a wrongly classified pixel according to its minimum distance to the DB with a sigmoid function. This limits the maximum error per pixel to less than one (as opposed to the unlimited error in regular regression analysis). The influence of outliers is reduced, which makes LR particularly robust.
The DT algorithm [16] compares feature values to thresholds based on a hirachical tree structure. Starting from the "stomp node", threshold queries determine the path along the outgoing" branch nodes" until a final query at the "leaf node" reveals the classification result. One reason for the popularity of DTs is that complex decision making is broken into many smaller "if-then" based problems. The ability to follow the querry path of the process makes the interpretation of DT intuitively easy. However, depending on the number of leafs, the DB geometry in FS may be very complex.

Benchmarking

The performance of a classification method is typically characterized by error rates, such as TPR and FPR. For data based approaches the data set is divided into a training partition (to train the classifier) and an **independent** testing partition, which is used to blind test the classifier to obtain the performance. In the case of variable model complexity, e.g. through the number of nodes in DT, the optimal model can be determinet with a so-called bias-variance-decomposition. The training error, so-called bias, measures the general ability of a method to solve the classification problem. The testing error, so-called variance, measures the generalizability of the approach. Typically, the bias decreases with the complexity of the model, whereas the variance increses. The preferable model is the one with the lowest sum of bias and variance error. In small, heterogenous data sets with high variability optimal partitioning can be challenging; depending on wether rare scenarios are in the training- or testing-partition the performance is over- or underestimated, respectively. To obtain a fair assesment cross validation methods, such as k-fold, use various different combinations of training and testing data and output mean performances. In k-fold the data is partitionened into *k* numbers of independent subsets named folds. Each fold functions as the testing partition and the rest as the training partition for *k* number of times.
Another remarkable characteristic of a classifier (besides the performance) is how it makes use of the available feature set. Although it is likely that features have different informative value, it can be shown that algorithms that exploit the full dimensionality of FS potentially perform better [11]. For LR and DT the use of each feature can be determinend differently. For LR (as with regular regression analysis) the regression coefficients theta $\boldsymbol{\theta}$ may be taken as a measure of the respective importance of a feature. If the theta value for a feature is close to zero, it contributes little to the separability of the classes. Conversely, non-zero thetas indicate relevant features. For DT a benchmark named “predictor importance” (PI) can be estimated. This measure values the risks of changes depending on the variability of a feature in the joint probabilities of a class (such as defect or intact) at a node and the probability that a node is reached. If the PI is zero, the feature plays no role in the decision making process. Conversley, the feature with the highest PI value has the highest impact.

RESULTS

The data set contains 18 independently collected measurements with four different NDT methods on a specimen with varying moisture, salt content, concrete quality and corrosion activity. Seven features were extracted. Table 1 lists the features (for further details on the feature extraction see [10, 11]).

Table 1: List of the features that were extracted from the NDT signals

	NDT-Method	Parameter
Feature 1	Ground-penetrating radar (GPR)	energy of direct wave
Feature 2	GPR	dominant frequency of direct wave
Feature 3	Half-cell potential (HP)	rebar potential
Feature 4	HP	change of rebar potential
Feature 5	Wenner-resitivity (WR)	electrical resistivity
Feature 6	Microwave-moisture (MW)	relative moisture of concrete
Feature 7	GPR	depth-corrected top rebar reflection

The performance of linear LR and the optimal DT from a k-fold (k=8) evaluation is summarized in table 2 and compared against the ASTM evaluation. We used fitglm [17] for LR and the fitctree [18] for DT from the MATLAB statistical toolbox to obtain the models. For DT the optimal number of splits and the optimal split criterion was determinend in a bias variance decomposition. The number of splits were varied between 1 and 60. The choice of the optimal split criterion among Gini's diversity index, twoing rule and cross entropy reduction fell on the latter. The best training results for LR were achieved with a data set containing unbalanced classes (~200 corrosion pixels VS. ~50k intact pixels). For DT the best training results were achieved with a more balanced data set (~2k corrosion pixels VS. ~50k intact pixels). Balancing was achieved through higher spatial sampling in the defect areas. However, the testing was conducted with the unbalanced data to assure comparability.
The fusion results are compared in table 2 in means of TPR and FPR with the conventional corrosion assessment based on the absolute potential measurements (according to ASTM C876 [2]) or based on the potential gradients (according to DGZFP Merkblatt B3 [1]).
The best result in terms of lowest number of misclassified pixels is obtainend with the LR-algorithm in a k-fold cross validation. Although the sensitivity of the DT algorithm is higher than the LR it is prone to make fals alarms. According to the ASTM standard, there are two possible ways of interpreting the results, depending on whether potentials in a transition range (of measured potential values between -200 and -350 mV) are considered corrosion or not. If not, the TPR for our data set is 5 % and the TNR 100 percent. If they are, the TPR and TNR are 100 % and 19 % respectively. According to Merkblatt B3, corrosion sites are identified by a negative potential change of several hundreds of millivolts. However, the maximum negative potential change in the current data set is approximately 40 millivolts. The evaluation according to DGZFP Merkblatt B3 is therefore completely inadequate for the present testing problem.

Table 2: Performance of different corrosion detection methods in terms of TPR and FPR with absolute number of misclassified pixels in brackets.

Method	TPR (num. misclas. pixels)	FPR (num. misclas. pixels)
LR (testing performance)	0.49 (100)	0 (24)
DT (testing performance)	0.51 (96)	0.006 (270)
ASTM (corrosion riks for umbigious potentials rounded up/down)	1/0.05 (0/186)	0.2/0 (9000/0)
DGZfP Merkblatt B3	0 (196)	0 (0)

To asses how LR and DT make use of the differen featueres table 3 lists the average regression coefficients from LR and the average PI values from DT. The table suggests that LR uses all features, since all theta are clearly different from zero. The importance factor PI for DT on the other hand is zero for two features (feature 2 and 6) and close to zero for a third (feature 7). The most important features for both methods are feature 3 and 4, which are both extracted from the HP-data.

Table 3: Importance of features for classification in terms of regression coefficient theta θ for LR-algorithm and PI-value for DT-algorithm.

Feature number/test method	Average regression coefficient theta θ	Average PI-value
Null	-11.23	-
1/GPR	0.134	0.0027
2/GPR	0.17	0
3/HP	-1.23	0.0052
4/HP	-2.86	0.0127
5/WR	-0.78	0.0028
6/MW	1.52	0
7/GPR	-1.69	0.0001

CONCLUSION

With this article we demonstrated that data-based approaches can improve the corrosion analysis of concrete. The reason is the possibility to consider different environmental conditions through the FS-representation. The data basis was created in an experiment that simulates the variability of field scenarios and provides a label with the corrosion state for each measuring point. For the simulated pitting corrosion scenarios the standard evaluation is either not sensitive enough (as in the case of Merkblatt B3 and ASTM-C876 with rounded off probabilities) or shows too many false-positive pixels (as in the case of ASTM-C876 with rounded up probabilities). Data-based corrosion assesment with the LR algorithm detects the damaged area with a TPR of 49% while maintaining a low FPR. The potentially more complex DT algorithm does not succeed in maintaining a small false alarm rate. The reason for this is assumed to be the poorer use of the available features.

This work shows the great potential for data driven methods. It exemplarily confirms that algorithms that find relatively complex DBs in relatively small but heterogeneous data sets are inferior to simple but robust approaches. In future work, the result may be improved by strategies such as ensemble-learning. Due to the higher demands on the validation and generalizability of data fusion, their development requires great experimental effort. However, if the results are transferred into practice, data fusion contributes to the development of fully automated NDT systems.

ACKNOWLEDGEMENT

We greatly acknowledge the generous financial support provided by the Indio German Science and Technology Centre (IGSTC) through DLR (German Aerospace Center).

REFERENCES

1. Deutsche Gesellschaft für Zerstörungsfreie Prüfung, DGZfP Merkblatt B3: Merkblatt für elektrochemische Potentialmessungen zur Detektion von Bewehrungsstahlkorrosion, Berlin, Germany, 2014.
2. ASTM International - Standards Worldwide, ASTM-C876: Standard Test Method for Corrosion Potentials of Uncoated Reinforcing Steel in Concrete, USA: American Society for Testing and Materials, 2015.
3. Kessler, S., Zur Verwertbarkeit von Potentialfeldmessungen für die Zustandserfassung und -prognose von Stahlbetonbauteilen – Validierung und Einsatz im Lebensdauermanagement (Dissertation, Munich, Germany: Technische Universität München, Munich, 2015.
4. Jol, H. M., Ground Penetrating Radar Theory and Applications, Amsterdam, Netherlands: Elsevier Science BV, 2009.
5. Morris, W., and A. Vico, M. Vazques und S. Sanchez, "Corrosion of reinforcing steel evaluated by means of concrete resistivity measurements," *Corrosion Science,* vol. 44, no. 1, pp. 81-99, 2002.
6. Leschnik, W., and U. Schlemm, Dielektrische Untersuchung mineralischer Baustoffe in Abhängigkeit von Feuchte- und Salzgehalt bei 2,45 GHz, Berlin, Germany: Umwelt · Meßverfahren · Anwendungen, 1999.
7. Oh, B. H., and S. Y. Jang, "Effects of material and environmental parameters on chloride penetration profiles in concrete structures," *Cement and Concrete Research,* vol. 37, no. 1, pp. 47-53, 2007.
8. ASTM International - Standards Worldwide, ASTM C1202: Standard Test Method for Electrical Indication of Concrete's Ability to Resist Chloride Ion Penetration, USA: American Society for Testing and Materials, 2012.
9. Stanish, K. D., and R. D. Hooton, M. D. A. Thomas, Testing the Chloride Penetration Resistance of Concrete: A Literature Review, Department of Civil Engineering, University of Toronto, Canada: United States. Federal Highway Administration, 2001.
10. Völker, C., and S. Kruschwitz, G. Ebell, "Data-driven multi-sensor-fusion for improved corrosion assessment of reinforced (submitted, in review)," *Journal of Nondestructive Evaluation,* 2018.
11. Völker, C., Datenfusion zur verbesserten Fehlstellendetektion bei der zerstörungsfreien Prüfung von Betonbauwerken (Dissertation), Deutschland : Universität des Saarlandes, 2017.
12. GSSI Geophysical Survey Systems, Inc., "GSSI 2.0 GHz Antenna," 2017. [Online]. Available: https://www.geophysical.com/antennas#lightbox-3.
13. Proceq, "Portable Non-Destructive Concrete Testing Instruments," 2018. [Online]. Available: https://www.proceq.com/uploads/tx_proceqproductcms/import_data/files/Concrete%20Testing%20Products_Sales%20Flyer_English_high.pdf.
14. HF-Sensors GmbH, Leipzig, "Microwave Moisture Products," 2018. [Online]. Available: http://www.hfsensor.de/englisch/index.html.
15. Walker, S. H., and D. B. Duncan, "Estimation of the Probability of an Event as a Function of Several

Independent Variables," *Biometrika,* vol. 54, no. 1-2, pp. 167-179, 1967.

16. Quinlan, J., "Induction of Decision Trees," *Machine Learning,* vol. 1, pp. 81-106, 1986.
17. Mathworks Inc., "fitglm Matlab function," [Online]. Available: https://de.mathworks.com/help/stats/fitglm.html. [Accessed 2018].
18. Mathworks Inc., "fitctree Matlab function," [Online]. Available: https://de.mathworks.com/help/stats/fitctree.html. [Accessed 2018].