Surveys in Geophysics (2020) 41:531–548
https://doi.org/10.1007/s10712-019-09558-4

A Machine Learning-Based Data Fusion Approach for Improved Corrosion Testing

Christoph Völker[1] · **Sabine Kruschwitz**[1,2] · **Gino Ebell**[1]

Received: 26 September 2018 / Accepted: 15 July 2019 / Published online: 24 August 2019

Abstract
This work presents machine learning-inspired data fusion approaches to improve the nondestructive testing of reinforced concrete. The principal effects that are used for data fusion are shown theoretically. Their effectiveness is tested in case studies carried out on large-scale concrete specimens with built-in chloride-induced rebar corrosion. The dataset consists of half-cell potential mapping, Wenner resistivity, microwave moisture and ground penetrating radar measurements. Data fusion is based on the logistic regression algorithm. It learns an optimal linear decision boundary from multivariate labeled training data, to separate intact and defect areas. The training data are generated in an experiment that simulates the entire life cycle of chloride-exposed concrete building parts. The unique possibility to monitor the deterioration, and targeted corrosion initiation, allows data labeling. The results exhibit an improved sensitivity of the data fusion with logistic regression compared to the best individual method half-cell potential.

Keywords Data fusion · Concrete · Pitting corrosion · Half-cell potential mapping · Machine learning

1 Introduction

Geophysical methods are typically used to study the geological situation of the near subsurface or the physical structure of the Earth. However, they have also become very popular for the nondestructive investigation of concrete infrastructure buildings (McCann and Forde 2001). Ground penetrating radar (GPR) (Jol 2009) is, e.g., commonly used for locating rebars and other steel built-in parts in reinforced concrete. Ultrasonic techniques—which are the high-frequency equivalent to seismic methods in geophysical applications—are applied for structural inspections such as thickness measurements, the localization of voids, delaminations or other types of flaws. However, the inspection tasks in nondestructive testing (NDT) of concrete objects are often quite complex and the measured data

✉ Christoph Völker
christoph.voelker@bam.de

[1] Bundesanstalt für Materialforschung und -prüfung, Unter den Eichen 87, 12205 Berlin, Germany

[2] Technische Universität Berlin, Institut für Bauingenieurwesen, Gustav-Meyer-Allee 25, 13355 Berlin, Germany

depend on different boundary and environmental conditions. Therefore, it is often difficult to reliably locate flaws or damages by only one individual sensor. For example, half-cell-potential (HP) (ASTM 2015) measurements are established for the detection of actively corroding areas of reinforcing steel in concrete. The measured potentials strongly depend on the respective moisture and chloride contents as well as the concrete cover thickness at each reading point. Without a priori information, the data interpretation is highly ambiguous (Kotan and Müller 2015; Gucunski et al. 2013). Consequently, and because automated and fast scanner systems are becoming more and more available, multisensor data are collected. Data evaluation and interpretation though are usually carried out for each sensor individually. And this is despite the fact that much more powerful algorithms for the evaluation of multisensor data were developed in other areas of science and are current state of knowledge.

In 1936, Fisher's famous work on the classification of iris flowers (Fisher 1936) described the basic principles of information extraction from multi-source data. Over the decades, his work has evolved into comprehensive theories that take into account aspects of real-world data such as imperfection and uncertainty (Vapnik 2000). The key concept is the representation of data in vectors, where each column corresponds to a different source. This is useful beyond data organization, as a comprehensive set of tools from stochastic, linear algebra and geometry is available to understand the data (Blum et al. 2018). The practical application of the theories, e.g., in machine learning, is very successful across disciplines and reveals a surprisingly high universal validity (Witten et al. 2017).

Our paper now provides a literature review that shows that the quality and amount of multi-source data fusion for NDT of concrete are still very low. We compile and compare several general theoretical considerations concerning the nature of different sensor data and therefore make an important contribution to the conceptual understanding of data fusion. We present an example of data-driven multisensor fusion based on laboratory corrosion data to demonstrate how the information content of a multisensor dataset can be increased in practice. Data-driven means that inter-sensor relationships are derived from measured data without the requirement of a physical model. Instead, the parameters of mathematical functions are optimized between pairs of sensor data and labels (that contain the sought information, e.g., defect/intact) in so-called training.

We use a supervised machine learning (SML) algorithm, more specifically logistic regression to recognize (or learn) corrosion and distinguish these data from those of non-corroding areas. The training data were obtained in a large-scale experiment that simulates the accelerated 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 GPR, HP, Wenner resistivity (WR) (Morris et al. 2002) and microwave moisture (MW) (Leschnik and Schlemm 1999). The results of the best individual sensor (HCP) and data fusion are compared in terms of true positive rate (TPR) and false positive rate (FPR).

1.1 Literature Review

Data fusion is a relatively new field in NDT of concrete, a fact that is also reflected in the relatively small number of references. The first comprehensive book on data fusion in NDT was written by Gros (1997). It describes particularities and challenges for the implementation of data fusion in NDT and contains a few examples of applications, but not for concrete testing.

The first comprehensive literature review on data fusion in NDT was published in 2001—also by Gros (2001). It contains an article on the NDT of prestressed steel in reinforced concrete (Fiedler 2001). In a laboratory study, complementary information from magnetic flux leakage and residual magnetism measurements was fused with a fuzzy-logic-based combination rule. This contribution contains neither a quantification of the result nor a measurable improvement compared to the individual measurements. Nevertheless, contributions from other areas of NDT showed the potential of data fusion, for example by improving the signal-to-noise ratio, obtaining additional information and increasing detection accuracy.

The second comprehensive literature review on data fusion in NDT was published in 2008 (Liu et al. 2008). One article describes a common three-dimensional visualization of different NDT methods for concrete analysis (Kohl et al. 2003)—a method from sensor integration (not data fusion).

Garnier et al. (2011) published data fusion approaches for material characterization of concrete in a field study and laboratory studies (Ploix et al. 2011) in 2011. The results showed a good agreement with the reference values in some cases, but a comparison with results from the individual measurements as evidence for an actual improvement through the fusion is missing.

Cotič et al. (2014) use cluster algorithms to detect near-surface built-in parts in a concrete specimen in a laboratory study using various NDT methods in 2013. Data fusion has only partially improved the result and in some cases was worse than the best single sensor. In the same publication, a field study on the detection of corrosion with cluster-based data fusion is presented. However, there is no adequate "ground-truth" to verify the results.

In the same year, Cui et al. (2013) report on a laboratory study in which the degradation of concrete by chloride ingress was simulated. The deterioration was monitored using the two NDT methods radar and half-cell potential mapping. The fusion is based on a simple combination of extracted signal features. An exact reference for the validation of the measurements and the fusion result is missing as well.

Ramos et al. (2015) determined the Young's modulus of a granite church wall in a field study, considering a priori knowledge. The approach based on Bayesian statistics determines a "trust factor" as a confidence measure in addition to the modulus of elasticity. However, no reference is available and the change in the fusion result has little significance after three iterations.

Ahmed et al. (2018) compare a wavelet transform-based image fusion approach with Bayesian networks using concrete bridge deck condition data. The data are taken from an Iowa highway research project report and contain information on two defects, namely corrosion and delamination (Gucunski et al. 2011). The article introduces different qualitative benchmarks for each fusion method. A comparison of the individual measurement methods is not possible as they yield different statements at different scales. The referencing by only a few qualitatively described drill cores is not suitable for estimating the degree of damage for the bridge—as claimed in the paper—because delamination and corrosion defects occur nonsystematically (randomly) and are localized.

When examining the literature, it is noticeable that the improvement in the information quality by data fusion for the NDT of reinforced concrete (or other mineral building materials) has not been demonstrated so far. However, potential improvements can be demonstrated in terms of statistical parameters, as shown in the next section.

2 Theoretical Background

The following example is based on Fisher's famous discriminant analysis equations (Fisher 1936) which we use to graphically illustrate the concept of data vectors and their usefulness for the fusion of NDT data. Despite strong assumptions, this simple model facilitates the meaningful selection of methods and signal features for data fusion in NDT.

2.1 Data Fusion Example

Suppose a NDT sensor observes an event c (e.g., the occurrence of a defect) through a characteristic feature X (e.g., radar amplitudes, resistivity measures, etc.). Its information quality is high when the feature values generate fixed probability distributions C that are significantly different for each event. The set of all feature values associated with one event is called class. In NDT examples, this means:

- For the HP method, the electrochemical potential of the rebar is measured against a reference electrode at the concrete surface. The measured feature X is the electrical potential. Measurements at corrosion sites tend to scatter around more negative potentials, and measurements at intact sites tend to scatter around more positive potentials. There are two distributions C for the classes "corrosion" and "no corrosion."
- In ferromagnetic concrete cover measurements, the depth of a rebar is measured using the energy loss of an induced magnetic field. The observed event is the depth of a rebar. The observed feature X is the energy loss of the magnetic field. The energy loss diminishes with increasing concrete cover thickness. There is one distribution C per observable depth.

The task of NDT is to determine the underlying (unknown) distribution C of each class, e.g., through laboratory investigations with representative samples. This allows to estimate the probability of an event c based on features X using the event distributions C according to Eq. (1):

$$p(X|C) = c. \tag{1}$$

The sensor performance depends on the overlap between the class distributions; the overlapping areas correspond to the probability of a misclassification (Johnason and Wichern 2014). Figure 1 shows a histogram of the feature distributions of two classes (red and blue) for two sensors. Both sensors are correlated with the classes (so-called marginal correlation)—this means that there is a direct relationship between the value of the sensor and the class affiliation (small values for blue; large values for red). The significance of sensor 1 is better than sensor 2 because the classes overlap less. For normal distributions, the significance increases with increasing distance of the mean values and decreasing variance.

Fisher recognized that the separability of different classes improves with the representation of the data in a higher dimensional feature space (FS). The FS is a real vector space in which the coordinates of each point correspond to the measured feature values at a position. The feature matrix F represents j features x at n measuring points together (see Eq. 2):

$$F = \left[x_1, x_2, \ldots, x_j\right] \quad F \in \mathbb{R}^{nxj}. \tag{2}$$

This representation can be regarded as a point cloud, where each point describes a measurement location and its position in space depends on the respective feature values.

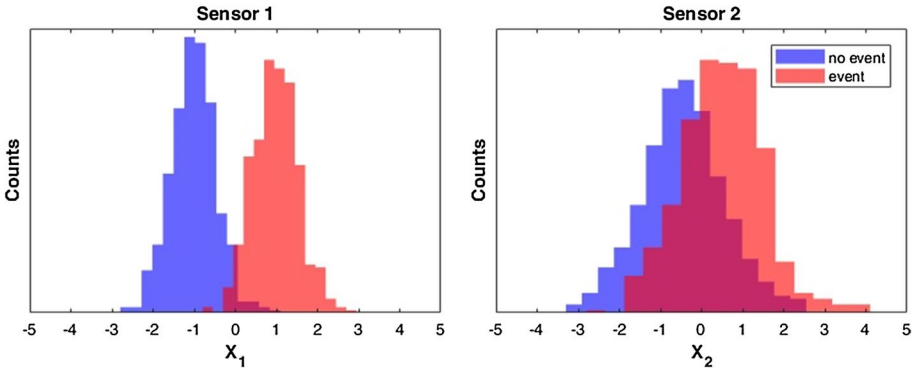

Fig. 1 Feature distributions from two sensors for classes "no event" (blue) and "event" (red); left: sensor with relatively high significance; right: sensor with relatively low significance

Points from different classes C are scattered at different locations in the FS. Fisher derives the separability of the classes by analogy to the univariate case by means of the distance between the class centers $\bar{X}$ and the covariance matrix Σ as the dispersion measure. Fisher recognizes that the potential significance increases in FS. The underlying assumption is that the intrinsic dimension of F is in fact much smaller and that the characteristic features of the observed event can be optimally represented by a lower-dimensional subspace $E \subseteq F$. E can be interpreted as a feature derived from the set of original features, which for two classes is expressed according to Eq. (3) (Johnason and Wichern 2014):

$$E = w_0 + w^{\mathrm{T}} x \tag{3}$$

with

$$w = \left(n_1 \Sigma_1 + n_2 \Sigma_2\right)^{-1} \left(\bar{X}_1 - \bar{X}_2\right)$$
$$w_0 = \frac{1}{n} \sum_{i=1}^{n} x_i - w^{\mathrm{T}} x_i$$

where $E \epsilon \mathbb{R}^{n \times 1}$ is the vector of projected feature values, $x_i \in \mathbb{R}^{j \times 1}$ is the ith line of the feature matrix with j features, $w_0 \in \mathbb{R}^{1 \times 1}$ is the intersection of projected features with $x_i = 0$, $w \in \mathbb{R}^{j \times 1}$ is the projection vector, $\Sigma \in \mathbb{R}^{j \times j}$ is the covariance matrix and $\bar{X} \in \mathbb{R}^{j \times 1}$ is the vector of class centroids. The eventual course of E is governed by the three statistical parameters:

$\left(\bar{X}_1 - \bar{X}_2\right)$	Distance between the mean values of the classes; a measure of how far away the classes are from each other
(σ^2)	Variance; diagonal entries of covariance matrix and a measure for marginal (individual feature) scattering
cov	Covariance coefficient; nondiagonal entries of covariance matrix and a measure for joint scattering of two random variables

Figure 2 illustrates the impact of each of these parameters on the separability of the different classes in three exemplary cases. Although genuine feature sets are often not

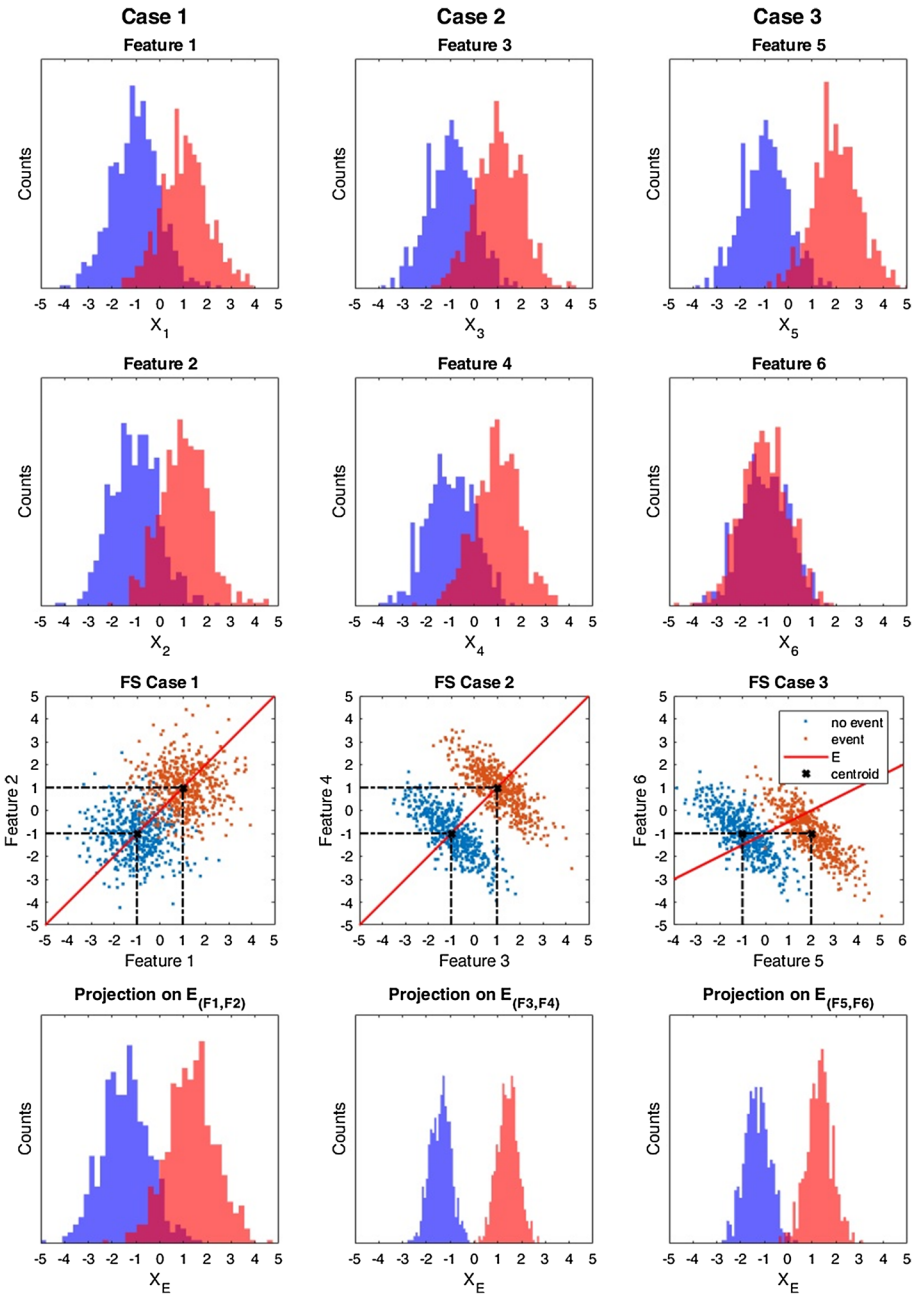

Fig. 2 Three exemplary cases of marginal feature distributions (row one and two), their joint distribution in FS (row 3) and the distribution of the projected feature values on the subspace E (red line) according to Eq. (3) (row 4) for two classes (blue, red). Left: marginally correlated features in FS, centre: marginally and conditionally correlated feature in FS. Right: conditionally correlated feature with only one marginally correlated feature

normally distributed, this simple model facilitates the understanding of otherwise counter-intuitive mathematical spaces.

In case 1 (Fig. 2, left), the features are (marginally) correlated with the classes but not among each other; all nondiagonal members of the covariance matrix are zero. The class distributions scatter in circular, standing or lying elliptical shapes in FS depending on the variance of the marginal features. The optimal course of E is through the class centers. This means that an increased significance results solely from $\bar{X}_1 - \bar{X}_2$, which is greater in higher dimensions due to its inclined course through space.

The features in case 2 (Fig. 2, center) are conditionally correlated with the classes; the covariance matrix is fully populated. Class distributions scatter in shapes of diagonally inclined ellipses. The particularity of the presented case is that the correlation between the feature values within one class and the course of the class centers is exactly orthogonal. The marginal distributions are identical to case 1. While the course of E is also identical to the first case, the separability along E increased noticeably. The reason is that the separability is dependent on the scattering between the features (described by the covariance coefficient) which by definition is smaller than the variance.

In case 3 (Fig. 2, right), a combination of all three factors has an impact on the course of E according to Eq. (3). Interestingly, in this example, feature 5 in itself is completely insignificant (marginal class distributions overlap entirely), but contributes to separability through its conditional correlation with feature 6.

The value of the FS representation for data fusion is on the one hand the increased significance through redundant information as a function of the distance between the mean values of the classes (as in case 1). On the other hand, complementary information in the form of characteristic correlation patterns as a function of variance and covariance may be identified (as in cases 2 and 3).

The implications for NDT vary casewise: Case 1 occurs if all features are independent. The independence is often reached through measurements with different NDT methods. While the improvement of case 1 is the lowest of the three, its predictability depends on only one variable $\left(\bar{X}_1 - \bar{X}_2\right)$ and is therefore particularly attractive in practical applications. Case 3 occurs if one feature (such as feature 6) measures an influencing factor that solely affects the other feature (such as feature 5). This may include the measurement of environmental conditions such as temperature, humidity and surface roughness. The features in case 2 are also correlated, but both show significance individually. This is the case if both the feature and the event itself are influenced by the same factor [e.g., temperature influences chemical reactions (event) and sensors (feature)]. The information content is best disclosed if the entire set of dependencies is represented. The depicted examples are two-dimensional, but can be arbitrarily high-dimensional, e.g., if more features are mutually dependent in so-called correlations of higher order.

Although in more than three dimensions the information is no longer visually accessible, an improvement can be observed through increased fusion performance (e.g., with respect to lower error rates).

2.2 Logistic Regression for Supervised Machine Learning

Real-world data—especially sensor data—are often more complex than Fisher assumed in his example. Outliers, multimodal, nonsymmetric feature distributions, nonlinear correlations, etc., require more robust algorithms, e.g., from SML. SML recognizes (or learns) the unknown inter-feature relations by optimizing mathematical functions between pairs

of representative data and labels (that contain the sought information, e.g., defect/intact) in so-called training. It is often more effective to estimate a function that distinguishes different classes, than to estimate a hyperplane for data projection as in the example above. The course of the distinguishing function is called decision boundary (DB). (For the scenarios in Fig. 2, the DB and E are orthogonal.) SML algorithms differ in how the DB is obtained and how complex the shape of the DB gets.

We use logistic regression (LR) which is a binary classifier algorithm first published by Walker and Duncan (1967). Its DB is described by a linear regression function (LRF), such as $\mathrm{LRF} = \theta_0 + \theta_1 F_1, \ldots, \theta_j F_j$. In training, the regression parameters theta $\boldsymbol{\theta}$ are optimized to minimize misclassification. The classification error is estimated by a cost function that weights misclassified pixels according to their 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 without the sigmoid function). This reduces the influence of outliers, which makes LR particularly robust. An advantageous property of LR is the possible linearly inclined course of the DB through FS (as opposed to an axially parallel course). As demonstrated in Fig. 2, this is advantageous for an optimal separation of classes in FS. We used fitglm (https://de.mathworks.com/help/stats/fitglm.html) for LR from the MATLAB statistical toolbox to obtain the model.

The key requirements for training data are that they represent the expected variability of applications (to ensure the generalizability of the inference) and a clear reference (to be able to assign a label to each data point). An experiment that addresses both aspects is described in the next section.

3 Corrosion Experiment for Data Fusion

The experiment has two main objectives: (1) variation in the environmental factors moisture and salt as expected in the life cycle of an affected structure (which is described in Sect. 3.2) and (2) provision of a reference value for the corrosion condition at each measurement location to allow for latter data labeling (which is described in Sect. 3.3). While the focus of the measures in (1) is on creating conditions conducive to corrosion, (2) aims at preventing corrosion and only initiating it at predefined locations. In this context, conventional corrosion experiments [as in Stoppel (2011), Gucunski et al. (2015), Nygaard and Geiker (2012), Hong et al. (2015), Hubbard et al. (2003) and Yeih and Huang (1998)] are not effective because either the exact location of the corrosion or the environmental conditions cannot be controlled. To close this gap, we have developed a procedure which is described in the following.

3.1 Description of the Specimen

The specimen was laid out as a large-scale concrete slab with the dimensions $1 \times 1.4 \times 0.3$ m^3 (LWH) (see Fig. 3, left). The design parameters are based on a traffic structure as these structures are exposed to large amounts of road salt and are therefore particularly at risk (CC Technologies Laboratories 2002). The concrete cover is 4.5 cm. The specimen consists of well-compacted concrete with a target strength of C30/37. The fresh concrete porosity is 1.4%. The water–cement ratio is 0.49. A Portland pozzolan CEM II/A-LL 32.5 *R* was used. The slab reinforcement consists of 10-mm-thick construction steel with a spacing of about 15 cm crosswise. Data were collected in two campaigns.

Fig. 3 Corrosion specimen. Left: specimen with saline water solution during potentiostatic chloride ingress. Right: corrosion anodes at top and bottom of the specimen before concreting

3.2 Variation in Environmental Factors Salt and Moisture

In practice, salts migrate into the concrete by diffusion and adsorption. It takes several years until an increased chloride concentration is reached at a considerable depth (Oh and Jang 2007). The chloride ingress is hence accelerated. We use the polarization method—a method well established in the quality assurance of concrete (Worldwide 2012; Stanish et al. 2001). It is based on the endeavor of ions to move toward the oppositely charged pole in an electric field. The electric field is typically induced potentiostatic between a working electrode (WE) and a counter electrode (CE). A potentiostat is a DC power source that keeps the potential of the WE constant with respect to a reference electrode (RE). The required current is dynamically generated (Stern and Geary 1957). All three components are connected to a galvanic element during operation. The key feature of this design is that the amount of chlorides passing through the field correlates with the induced current. This helps to control the migration, as a sudden increase in the polarization current could, for example, warn of a premature penetration of chloride through cracks and flaws. In the present configuration, the reinforcement is the WE and functions as the positive pole. It is polarized against a manganese mixed-oxide mesh in a sodium chloride solution on top of the test specimen as CE, which functions as the negative pole. An ERE 20 (Force Technology 2015) manganese dioxide (MnO_2) reference electrode is installed at the top and bottom of the specimen. A wiring diagram of the setup is shown in Fig. 5 (top).

3.3 Referencing the Corrosion Condition

The migration of chlorides is monitored with depassivation sensors. Similar to the anode ladder (Sensortec 2015), carbon steel rods are installed at different depths. The variation in the rods potential against the reference electrode is logged with a multi-channel data logger (multiplexer). A steeply decreasing potential indicates the onset of corrosion—caused by the presence of a critical corrosion-initiating chloride concentration in the respective depth. The sensor consists of two rows of stainless-steel welding wires with an attached, 1-cm-long carbon steel tip. In the centre of each of the four sides of the test specimen, we installed one sensor reaching five centimeters into the concrete (see Fig. 4). They cover a depth of 4 cm with a vertical resolution of 0.5 cm. This setup contains several systematic uncertainties:

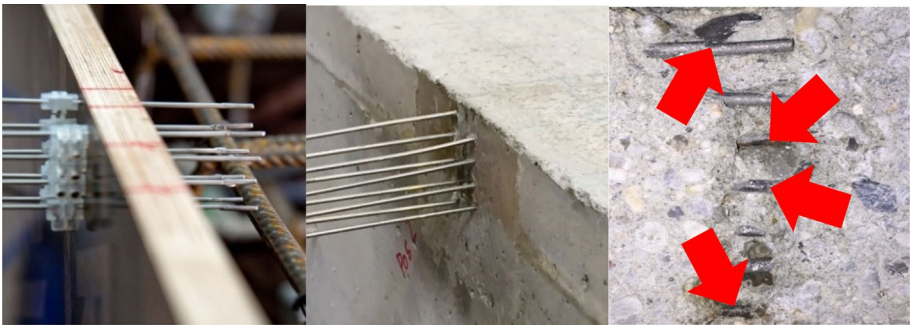

Fig. 4 Depassivation sensors in two rows; left: installation before concreting; middle: sensors in use on the specimen; right: sensor tips after completion of chloride penetration with marked corrosion spots

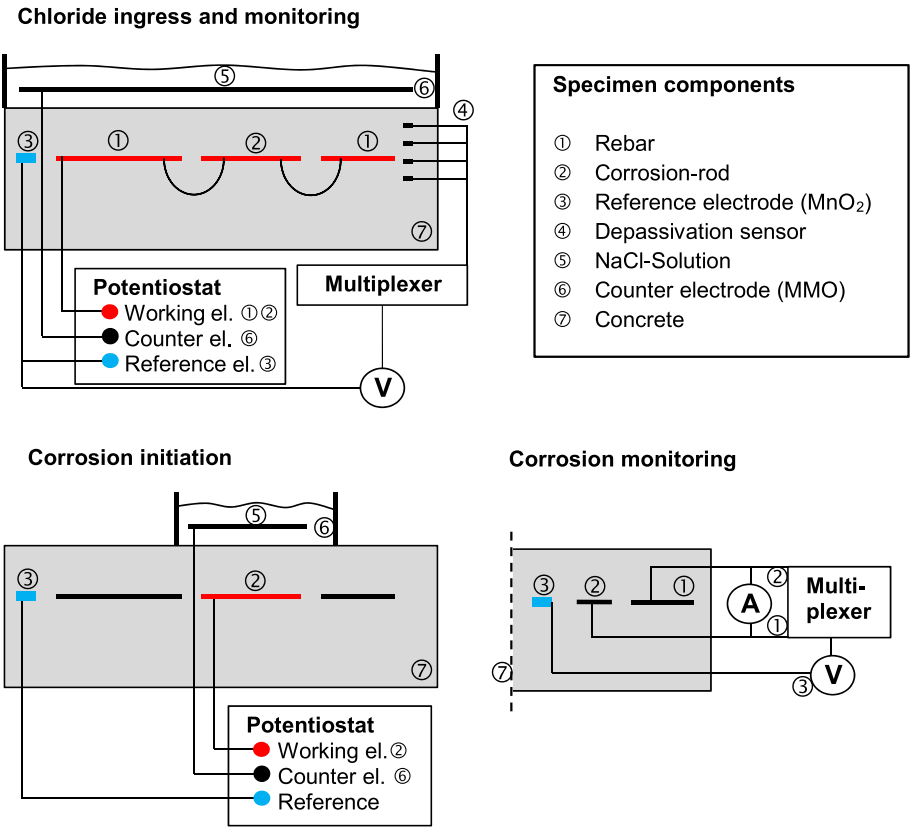

Fig. 5 Wiring diagrams for the steps to control the deterioration conditions; top left: chloride ingress and monitoring; top right: specimen components; bottom left: corrosion initiation; bottom right: corrosion monitoring

- The critical chloride concentration for thin wires is higher than for rebars (Dauberschmidt 2006).
- Although the sensors at the edge of the specimen are in the region of the strongest polarization field strength, point-like measurements are subject to the typical uncertainties of a random sample and might therefore not be representative for the entire surface area.
- Other causes that are not necessarily related to the penetrating chloride may also cause corrosion at the sensor tip (Ann and Song 2007).
- After polarization, chloride penetration progresses further, e.g., through the equalization of concentration with deeper pores during drying (Johannesson 2003).

Accelerated chloride ingress has been terminated after the first indication from a depth of 2 cm. A wiring diagram of the measurement setup is shown in Fig. 5 (top).

The penetration of the chlorides was destructively verified at the end of the experiment. Eight drill cores have been taken from the specimen, and their cross sections were analyzed with laser-induced breakdown spectroscopy (LIBS) (Miziolek 2006). LIBS is a spectral analysis method that is used to map chemical element distributions on surfaces. The critical corrosion-initiating chloride content of 4 wt% relative to the cement mass[1] reaches a depth of approximately 2 cm in the first campaign and about 5 mm in the second campaign. Thus, increased chloride concentrations at reinforcement level could be ruled out for data labeling.

After areal chloride migration, the corrosion is initiated by anodic polarization of the reinforcement of an electrically insulated corrosion rod (see Fig. 3, right) beyond the critical corrosion-inducing potential. A schematic diagram of the structure is shown in Fig. 5 (bottom, left). A titanium mixed-metal oxide grid was used as the counter electrode and a 3 molar sodium chloride solution as the electrolyte. The polarization current has been logged. A sudden increase indicates the beginning of active corrosion (Revie and Uhlig 2008). After initiation, the corrosion activity was monitored by measuring the corrosion current between the anode and the noncorroding rebar cage (as the cathode) via a 10 Ω shunt resistor (see Fig. 5, bottom, right). The experimental procedure is further detailed in Völker (2017).

3.4 Description of NDT Measurements and Feature Extraction

The experimental procedure has been monitored by regular NDT measurements. The final dataset contains 18 independently collected measurements with four different NDT methods on the specimen described above. The GPR measurements were carried out using a GSSI, Inc. SIR20 device with a 2 GHz palm antenna (GSSI 2017). 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 20 mm. The HP, WR and MW measurements were collected manually along a predefined measuring grid with a spacing of 100 mm. The HP data were collected using a Canin + corrosion analysis system from Proceq (2018). The reference electrode is a copper/copper sulfate rod electrode. The WR data were collected using a

[1] There is a wide range of specifications for the critical corrosion-causing chloride content. The value 4 wt% is based on the DIN-EN 206 standard.

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

	NDT-method	Feature	Range of values		
			Min.	Mean	Max.
Feature 1	GPR	Energy of surface wave	1292	2699	3750
Feature 2	GPR	Dominant frequency of surface wave	0.69	1.51	2.16 (GHz)
Feature 3	HP	Rebar potential	− 352	− 253	− 108 (mV)
Feature 4	HP	Change of rebar potential	− 42	0	52 (mV)
Feature 5	WR	Electrical resistivity	47	212	977 (Ω)
Feature 6	MW	Relative moisture of concrete	0	2.6	6.9 (wt%)
Feature 7	GPR	Depth-corrected top rebar reflection	0.4	1	1.3

Resipod probe from Proceq (2018), and the MW measurements were performed using an ID10 probe from HF sensor (HF-Sensors 2018).

A list of the features, which were extracted from the NDT methods, their range of values and their respective units are given in Table 1.

Features 1 and 2 were extracted from the GPR surface wave (SW). The SW is significantly influenced by changes in the dielectric properties of the concrete at the surface [as demonstrated in Klysz et al. (2005) and Chen (2012)]. With increasing pore water content in the concrete, its permittivity increases and the propagation velocity of the SW and thus its dominant frequency decrease (Laurens et al. 2005). Dissolved ions increase the conductivity of the concrete, which reduces the SW amplitude—especially of the higher frequencies in the emitted spectral band. Chlorides further increase attenuation. Changes in the permittivity and conductivity affect the SW energy (feature 1) and SW frequency (feature 2). The feature extraction is based on unfiltered raw signals without depth-dependent amplification. The power spectrum is calculated using the discrete Fourier transformation. The energy of the SW (feature 1 in Table 1) was estimated as areal integral over the frequencies from 0 to 4 GHz:

$$E = \int_{0\,\mathrm{GHz}}^{4\,\mathrm{GHz}} E\left(\bar{f}\right), \mathrm{d}\bar{f} \tag{4}$$

where E is the energy in the frequency spectrum and $\bar{f}$ is the discrete frequency. The center frequency of the SW (feature 2 in Table 1) is determined from the discrete power spectrum by the following equation:

$$\bar{f}_M = \frac{1}{\sum E_{f(0,\ldots,4)}} \int_{0\,\mathrm{GHz}}^{4\,\mathrm{GHz}} \bar{f} E\left(\bar{f}\right), \mathrm{d}\bar{f} \tag{5}$$

where $\bar{f}_M$ is the discrete center frequency.

Features 3–6 are derived from parametric measurement methods. Because their measurement signals are scalars, feature extraction is not required. Feature 7 is based on the attenuation of the rebar reflections—a common method for the detection of deteriorated concrete with GPR [e.g., in Gucunski et al. (2011), Barnes and Trottier (2004)].

Table 2 Subtrahends and factors $\left(\mathrm{SD}(f\text{-median}(f))^{-1}\right)$ for feature scaling (the subtrahend is the number to be subtracted)

	Feat. 1	Feat. 2	Feat. 3	Feat. 4	Feat. 5	Feat. 6	Feat. 7
Subtrahend $S = \mathrm{median}(f)$	5875	1.372	−263.790	−0.137	192.260	2.450	0.9910
Scaling factor $F = \frac{1}{\mathrm{SD}(f-S)}$	0.0011	0.855	0.0196	0.1306	0.0085	0.9720	6.9307

The GPR amplitudes were depth-corrected with the automated "Distance Amplitude Correction" (DAC).

The feature set has been normalized to ensure comparability among each other. The normalization parameters are summarized in Table 2. A spatially down-sampled version of the features set is available for download (Völker 2018).

4 Results

For features 3, 4 and 7, the distance between the mean values $\bar{X}_{\mathrm{corrosion}} - \bar{X}_{\mathrm{intact}}$ of corrosion and intact pixels is with 1.2, 2.6 and 1.5 significantly different. This means they are marginally correlated with the classes "corrosion" and "intact," which corresponds to the case 1 in Fig. 2. Features 1, 2, 5 and 6 aim at characterizing the environmental conditions. This corresponds to case 3 in Fig. 2. The distances of the means between the classes are smaller (0.7, 0.3, 0.5 and 0.5, respectively). The covariance matrix of the normalized features in Eq. (6) reveals that there are no strong correlations between any of the corrosion and environmental features:

$$\mathrm{cov}(F) = 10^{-2} \times \begin{bmatrix} 99 & 59 & 7 & 5 & 26 & -54 & -2 \\ & 98 & -22 & 3 & 12 & -26 & -2 \\ & & 101 & 16 & -10 & -25 & -1 \\ & & & 108 & 9 & -5 & 7 \\ & & & & 99 & -48 & 3 \\ & \mathrm{sym} & & & & 102 & -2 \\ & & & & & & 98 \end{bmatrix} \begin{matrix} F1 \\ F2 \\ F3 \\ F4 \\ F5 \\ F6 \\ F7 \end{matrix} \quad (6)$$

$$F1 \quad F2 \quad F3 \quad F4 \quad F5 \quad F6 \quad F7$$

However, the covariance coefficient indicates only linear correlations between two variables. As mentioned above, higher order correlations, i.e., between more than two features, are still possible as are nonlinear correlations. Figure 6 shows an example of the distributions of features 4 and 7 and their common feature space. In comparison with Fig. 2, it is evident that the distributions are more irregular. Nevertheless, their distribution in FS is similar to that of case 1 in Fig. 2 as expected. The separability can be improved by the use of more features (in a higher dimensional space, which cannot be visualized).

The performance of a classification method is typically characterized by error rates, such as true positive rate (TPR) and false positive rate (FPR). For data-based approaches, the dataset 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.

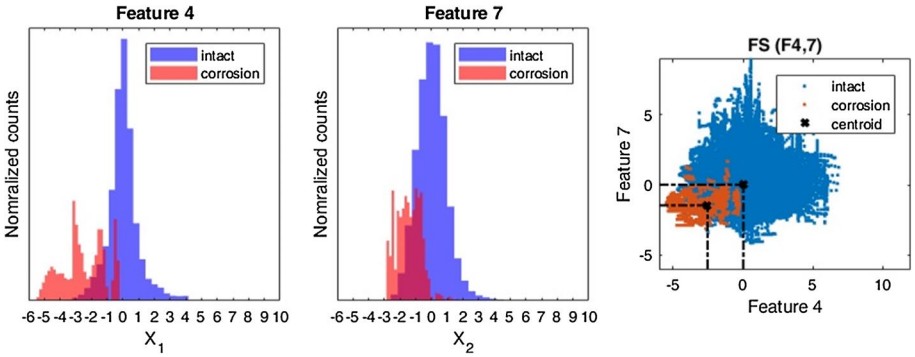

Fig. 6 Feature distributions for classes "intact" (blue) and "corrosion" (red); left and center: class-normalized histogram of features 4 (change of rebar potential) and 7 (depth-corrected top rebar reflection); right: FS from features 4 and 7

Table 3 Model parameters from logistic regression

	Theta 0	Feat. 1	Feat. 2	Feat. 3	Feat. 4	Feat. 5	Feat. 6	Feat. 7
Θ	− 11.23	0.134	0.17	− 1.23	− 2.86	− 0.78	1.52	− 1.69

Table 4 Sensitivity and false alarm rate of fusion and individual methods

Method for corrosion detection	TPR	FPR
Data fusion: LR with seven features	0.49	0
Data fusion: LR with five features (features 1–4, feature 7)	0.58	0
Best LR result for single feature (feature 4)	0.2	0
ASTM C-876 (50% corrosion risk rounded up/rounded down)	1/0.05	0.2/0
DGZfP B 03 E	0	0

The training error is a measure for the general ability of a method to solve the classification problem. The testing error measures the generalizability of the approach. In small, heterogeneous datasets with high-variability optimal partitioning can be challenging, depending on whether rare scenarios are in the training—or testing—partition the performance is over- or underestimated, respectively. To obtain a fair assessment, cross-validation methods, such as *k*-fold, use different combinations of training and testing data and give mean performances (e.g., in terms of mean TPR and FPR) (Witten et al. 2017). In *k*-fold, the data are partitioned 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. LRs parameters θ can be obtained as means from the *k* trainings or by a single training with the entire dataset. We have chosen the latter for the parameters in Table 3.

It is a common practice in supervised machine learning (SML) to systematically remove features from the training data. This is due to the phenomenon called "Curse of Dimensionality." The curse lies in the fact that spaces in higher dimensions become increasingly sparse. Consequently, a given number of data points become increasingly insignificant and

SML converges increasingly worse. Removing features from the dataset is an effective way to avoid this. The performance of linear LR for our dataset from a k-fold ($k=8$) evaluation is summarized in Table 4 and compared against the best single method, the best combination of any features and conventional corrosion assessment based on absolute potential measurements by HP [according to ASTM C876 (ASTM 2015)] or based on the potential gradients (according to the German standard DGZFP B03 E (Deutsche 2014)].

Data fusion with LR detects about half of the corrosion pixels while maintaining a low false alarm rate. LR achieves the best result for any combination of features using features 1–4 and 7. The true positive rate of 0.58 marks a performance increase of nearly 20% over the full dataset.

The best result achieved with a single feature is by using the relative potential change from the HP method (feature 4). However, with a TPR of 21%, the sensitivity is low. The corrosion detection for the first campaign fails due to a relatively low corrosion activity. According to the ASTM standard, there are two possible ways of interpreting the measured values, depending on whether potentials in a transition range (of potential values between -200 and -350 mV) are considered as displaying corrosion or not. If not, the TPR is 5% and the TNR 100%. If they are, the TPR is 100% and the TNR is 19%, respectively. According to DGZfP B 03 E (Deutsche 2014), corrosion sites are identified by a negative potential change of several hundred mV. The maximum negative potential change in the current dataset is approximately 40 mV. The evaluation according to DGZfP B 03 E is therefore inadequate for the present testing problem.

5 Conclusions

The present work has shown how the mathematical relationship developed by Fisher in 1936 for the separation of two classes of data in the feature space can be interpreted for the fusion of NDT results. Simple statistical relationships are explained and visualized using three descriptive cases: (1) the features correlate marginally with the classes but not with each other, (2) the features correlate marginally with the classes and conditionally with each other and (3) one out of two features correlates marginally with the classes, both correlate conditionally with each other. In all three cases, the results of the data evaluation improve in the feature space. Practical examples of NDT sensor combinations are given for all three cases.

On the example of a laboratory study, the requirements for real data evaluation in the feature space are described. The combined accelerated deterioration and targeted corrosion initiation on a large concrete specimen represents a novel corrosion experiment, which is controlled at all times. The particular advantage is the possibility to tag the collected NDT data in a variety of environmental conditions with the condition labels “intact” or “corrosion” depending on the position and time where they were collected. This allowed for practical use of algorithms from the field of supervised machine learning in the NDT of concrete.

We subsequently divided the dataset into training and testing data subsets with a wide variability. For the evaluation of the classification performance of the classifier developed on the basis of the training data, the relatively small number of 18 feature sets was accounted for by using k-fold validation. It could be shown that the fusion outperformed each individual NDT method and standard evaluation approaches. Whereas the best individual sensor (here half-cell potential mapping for the detection of corrosion events)

showed a true positive rate of 21%, the combined use of seven features (here generated by the four NDT methods ground penetrating radar, microwave, resistivity and half-cell potential mapping) increased the true positive rate to 49%. The false positive rate was 0% in both approaches. However, the best result of any combination of features was obtained with features 1–4 and feature 7 with a TPR of 0.58 and an FPR of 0. This combination of features reduces the number of methods needed for a useful data fusion result down to only two: HP and GPR.

Due to the higher demands on the validation and generalizability of data fusion, their development requires great experimental effort. However, if these procedures are used in practice, data fusion contributes to the development of fully automated NDT systems and increases the reliability, robustness and completeness of the data evaluation result significantly.

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

References

Ahmed M, Moselhi O, Bhowmick A (2018) Two-tier data fusion method for bridge condition assessment. Can J Civ Eng 45:197–214

Ann KY, Song H-W (2007) Chloride threshold level for corrosion of steel. Corros Sci 49:4113–4133

ASTM International—Standards Worldwide (2015) ASTM-C876: standard test method for corrosion potentials of uncoated reinforcing steel in concrete, USA

Barnes CL, Trottier J-F (2004) Effectiveness of ground penetrating radar in predicting deck repair quantities. J Infrastruct Syst 10:69–76

Blum A, Hopcroft J, Kannan R (2018) Foundations of data science. Cornell University, New York

CC Technologies Laboratories (2002) Incorporated; NACE International; Federal Highway Administration, "Corrosion Costs and preventive Strategies in the United States (FHWA-RD-01-156, R315-01)". National Technical Information Service, Alexandria

Chen C-C (2012) Lateral waves in ground penetrating radar applications. In: Proceedings of the 14th international conference on ground penetrating radar, Shanghai

Cotič P, Jagličić Z, Niederleithinger E, Stoppel M, Bosiljkov V (2014) Image fusion for improved detection of near-surface defects in NDT-CE using unsupervised clustering methods. J Nondestruct Eval 33(3):384–397

Cui J, Huston DR, Arndt RW (2013) Data fusion for multiple-sensor nondestructive evaluation on concrete bridge deck. In: Proceedings of TRB 2013 annual meeting, Washington, DC

Dauberschmidt C (2006) Untersuchungen zu den Korrosionsmechanismen von Stahlfasern in chloridhaltigem Beton. Dissertation, RWTH Aachen, Aachen

Deutsche Gesellschaft für Zerstörungsfreie Prüfung (2014) DGZfP B 03 E—electrochemical half-cell potential measurements for the detection of reinforcement corrosion, Berlin

Fiedler U (2001) NDT data fusion in civil engineering. In: Gros XE (ed) Applications of NDT data fusion. Kluwer, Norwell, pp 193–204

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugen 7:179–188

Force Technology (2015) ERE 20 reference electrode (Online). http://www.forcetechnology.com/en/Menu/Products/Concrete-monitoring/Concrete-monitoring-probes/ere20referenceelektrode.htm. Accessed 12 Mar 2015

Garnier V, Ploix MA, Breysse D (2011) Data fusion to improve the concrete diagnosis. Nondestruct Test Mater Struct 6:1241–1246

Gros XE (1997) NDT data fusion. S. E. Press, Bodmin

GSSI Geophysical Survey Systems, Inc. (2017) GSSI 2.0 GHz antenna (Online). https://www.geophysical.com/antennas#lightbox-3

Gros XE (2001) Applications of NDT data fusion. Kluwer, Norwell

Gucunski N, Romero F, Kruschwitz S (2011) Comprehensive bridge deck deterioration mapping of nine bridges by nondestructive evaluation technologies (project SPR-NDEB(90)–8H-00). Iowa Highway Research Board, Iowa Department of Transportation, Federal Highway Administration, Iowa

Gucunski N, Imani A, Romero F, Nazarian S, Yuan D, Wiggenhauser H, Shokuohi P, Taffe A, Kutrubes D (2013) Nondestructive testing to identify concrete bridge deck deterioration. TRB, Washington, DC
Gucunski N, Kee SH, La H, Basily B, Maher A (2015) Delamination and concrete quality assessment of concrete bridge decks using a fully autonomous RABIT platform. Struct Monit Maint 2(1):19–34
HF-Sensors GmbH, Leipzig (2018) Microwave moisture products (Online). http://www.hfsensor.de/english/index.html
Hong S, Lai WL, Helmerich R (2015) Experimental monitoring of chloride-induced reinforcement corrosion and chloride contamination in concrete with ground-penetrating radar. Struct Infrastruct Eng Maint Manag Life Cycle Des Perform 11(1):15–26
Hubbard SS, Zhang J, Monteiro JM, Peterson JE, Rubin Y (2003) Experimental detection of reinforcing bar corrosion using nondestructive geophysical techniques. Mater J 100(6):501–510
Johannesson BF (2003) A theoretical model describing diffusion of a mixture of different types of ions in pore solution of concrete coupled to moisture transport. Cem Concr Res 33(4):481–488
Johnason RA, Wichern DW (2014) Applied multivariate statistical analysis, 6th edn. Pearson Education, Upper Saddle River
Jol HM (2009) Ground penetrating radar theory and applications. Elsevier, Amsterdam
Klysz G, Ferries X, Balayssac J, Laurens S (2005) Simulation of direct wave propagation by numerical FDTD for a GPR coupled antenna. NDT&E Int 39:338–347
Kohl C, Krause M, Maierhofer C, Wiggenhauser H (2003) 3D-visualisation of NDT data using a data fusion technique. OR Insight 45(12):800–804
Kotan VME, Müller H (2015) Application of a modified half-cell potential mapping procedure for the condition assessment of a partially coated pre-stressed concrete pedestrian bridge. In: International symposium on non-destructive testing in civil engineering (NDT-CE), Berlin
Laurens S, Balayssac J, Rhazi J, Klysz G, Arliguie G (2005) Non-destructive evaluation of concrete moisture by GPR: experimental study and direct modeling. Mater Struct 3:827–832
Leschnik W, Schlemm U (1999) Dielektrische Untersuchung mineralischer Baustoffe in Abhängigkeit von Feuchte- und Salzgehalt bei 2.45 GHz. Umwelt · Meßverfahren · Anwendungen, Berlin
Liu Z, Forsyth DS, Komorowski JP, Hanasaki K, Kirubarajan T (2008) Survey: state of the art in NDE data fusion techniques. IEEE Trans Instrum Meas 56:2435–2451
Mathworks Inc., Fitglm matlab function (Online). https://de.mathworks.com/help/stats/fitglm.html. Accessed 2018
McCann DM, Forde MC (2001) Review of NDT methods in the assessment of concrete. NDT&E Int 34:71–84
Miziolek AE (2006) Laser induced breakdown spectroscopy (LIBS)—fundamentals and applications. Cambridge University Press, Cambridge
Morris W, Vico A, Vazques M, Sanchez S (2002) Corrosion of reinforcing steel evaluated by means of concrete resistivity measurements. Corros Sci 44(1):81–99
Nygaard PV, Geiker MR (2012) Measuring the corrosion rate of steel in concrete—effect of measurement technique, polarisation time and current. Mater Corros 63(3):200–214
Oh BH, Jang SY (2007) Effects of material and environmental parameters on chloride penetration profiles in concrete structures. Cem Concr Res 37(1):47–53
Ploix MA, Garnier V, Breysse D, Moysan J (2011) NDE data fusion to improve the evaluation of concrete structures. NDT&E Int 44:442–448
Proceq (2018) Portable non-destructive concrete testing instruments (Online). https://www.proceq.com/uploads/tx_proceqproductcms/import_data/files/Concrete%20Testing%20Products_Sales%20Flyer_English_high.pdf
Ramos LF, Miranda T, Mishra M, Fernandes FM, Manning E (2015) A Bayesian approach for NDT data fusion: the Saint Torcato church case study. Eng Struct 84:120–129
Revie RW, Uhlig HH (2008) Corrosion and corrosion control—an introduction to corrosion science and engineering. Wiley, New York
Sensortec (2015) Sensortec anode ladder (Online). http://www.sensortec.de/sensoren-sensors/anodenleiter-anode-ladder. Accessed 13 Mar 2015
Stanish KD, Hooton RD, Thomas MDA (2001) Testing the chloride penetration resistance of concrete: a literature review. Department of Civil Engineering, University of Toronto, Canada: United States. Federal Highway Administration
Stern M, Geary AL (1957) Electrochemical polarization I. A theoretical analysis of the shape of polarization curves. J Electrochem Soc 104(1):56–63
Stoppel M (2011) Differenzpotentialfeldmessung in der automatisierten Prüfung von Stahlbetonbauteilen, vol 75. BAM-Dissertationsreihe, Berlin
Vapnik VN (2000) The nature of statistical learning theory. Springer, Red Bank

Völker C (2017) Datenfusion zur verbesserten Fehlstellendetektion bei der zerstörungsfreien Prüfung von Betonbauwerken. Dissertation. Universität des Saarlandes, Deutschland

Völker C (2018) Labeled non-destructive testing data set for corrosion detection (Online). https://doi.org/10.13140/rg.2.2.24881.89448. Accessed 03 Mar 2018

Walker SH, Duncan DB (1967) Estimation of the probability of an event as a function of several independent variables. Biometrika 54(1–2):167–179

Witten IH, Frank E, Hall MA, Pal CJ (2017) Data mining—practical machine learning tools and techniques, 4th edn. Morgan Kaufmann, Cambridge

Worldwide AI-S (2012) ASTM C1202: standard test method for electrical indication of concrete's ability to resist chloride ion penetration. American Society for Testing and Materials, New York

Yeih W, Huang R (1998) Detection of the corrosion damage in reinforced concrete members by ultrasonic testing. Cem Concr Res 28(7):1071–1083

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com