



**QUEEN'S  
UNIVERSITY  
BELFAST**

## **Avoiding Overfitting in Inverse Modeling of Chloride Migration in Concrete**

Nanukuttan, S., Ukrainczyk, N., Robinson, D., & Koenders, E. (2016). Avoiding Overfitting in Inverse Modeling of Chloride Migration in Concrete. In *Proceedings of the International RILEM Conference on Materials, Systems and Structures in Civil Engineering 2016* (pp. 599-608)

### **Published in:**

Proceedings of the International RILEM Conference on Materials, Systems and Structures in Civil Engineering 2016

### **Document Version:**

Peer reviewed version

### **Queen's University Belfast - Research Portal:**

[Link to publication record in Queen's University Belfast Research Portal](#)

### **Publisher rights**

Copyright RILEM 2016

### **General rights**

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

### **Open Access**

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

## **AVOIDING OVERFITTING IN INVERSE MODELING OF CHLORIDE MIGRATION IN CONCRETE**

**Sreejith V. Nanukuttan** <sup>(1)</sup>, **Neven Ukrainczyk** <sup>(2)</sup>, **Des J. Robinson** <sup>(1)</sup>, and **Eduardus A.B. Koenders** <sup>(2)</sup>

(1) School of Planning, Architecture and Civil Engineering, Queen's University Belfast, Belfast BT9 5AG, United Kingdom

(2) Institute for Construction and Building Materials, Technical University Darmstadt, Franziska-Braun-Straße 3, 64287 Darmstadt, Germany (ukrainczyk@wib.tu-darmstadt.de)

### **Abstract**

Inverse analysis for reactive transport of chlorides through concrete in the presence of electric field is presented. The model is solved using MATLAB's built-in solvers "pdepe.m" and "ode15s.m". The results from the model are compared with experimental measurements from accelerated migration test and a function representing the lack of fit is formed. This function is optimised with respect to varying amount of key parameters defining the model. Levenberg-Marquardt trust-region optimisation approach is employed. The paper presents a method by which the degree of inter-dependency between parameters and sensitivity (significance) of each parameter towards model predictions can be studied on models with or without clearly defined governing equations. Eigen value analysis of the Hessian matrix was employed to investigate and avoid over-parametrisation in inverse analysis. We investigated simultaneous fitting of parameters for diffusivity, chloride binding as defined by Freundlich isotherm (thermodynamic) and binding rate (kinetic parameter). Fitting of more than 2 parameters, simultaneously, demonstrates a high degree of parameter inter-dependency. This finding is significant as mathematical models for representing chloride transport rely on several parameters for each mode of transport (i.e., diffusivity, binding, etc.), which combined may lead to unreliable simultaneous estimation of parameters.

**Keywords:** rapid chloride migration test, numerical modelling, inverse analysis, parameter estimation, overfitting, Levenberg-Marquardt.

## 1. Introduction

Numerous methods have been developed to determine the chloride ion diffusion coefficient,  $D_{ef}$  [1]. Conventional methods of determining the effective chloride diffusion parameter involves long test duration and hence international standards such as ASTM C1202 and Nordic Test 492, make use of an externally applied voltage to accelerate the test. The parameter determined from these tests is used either for quality control or, in conjunction with Fick's second law, to predict the chloride ingress in concrete. In contrast to the practical development of test and predictive procedures, mathematical models for the reactive transport of ions in concrete have been developed. Recent models have become highly sophisticated taking into account the interaction of the different ionic species within the concrete [2, 3]. This has been achieved by deriving individual transport equations for each species, with the electric field resulting from any applied voltage and from the interaction of the ions themselves being calculated from the solution of a Poisson partial differential equation. In addition, supplementary equations are needed to enforce the electro-neutrality constraint and to relate the activity coefficients to the ionic strength. The mathematical model adopted here derives reactive transport equations for only the chloride ions as proposed by Spiesz et al. [1]. The goal is to bridge the gap between the experimental tests and the mathematical models. The key to achieving this is by the addition of an optimisation procedure to the numerical model, which combined is called numerical inverse modelling. This is in keeping with many mathematical modelling procedures where initially attention focuses on the direct modelling of a problem. When this can be accomplished satisfactorily, attention is then directed to accurate parameter estimation that can be achieved through established optimisation techniques, which should include sensitivity and parameter inter-dependency analysis. Spiesz et al. [1] employed inverse analysis for chloride migration in concrete using a Generalized Reduced Gradient Algorithm optimization method and a forward discretization Finite Difference numerical implementation for solving a reactive-transport model. They simultaneously fitted four parameters simultaneously characterising chloride migration, binding rate and binding capacity, but did not address the parameter estimation accuracy which may be significantly impaired due to over-parametrisation and parameter inter-dependency/correlation problem.

The work reported in this paper makes use of more advanced numerical procedures both for solving the direct model and the inverse optimisation problem. A method of lines (MOL) approach [4] is used for solving the reactive transport system of equations. Optimisation is enriched with a detailed analysis on interdependency and sensitivity of the parameters. In general, as the reactive-transport models increase in complexity, the number of parameters in a model may increase to the point where over-parameterisation can result in overfitting and deterioration in prediction accuracy. This over-fit is due to a high degree of parameter inter-dependency (i.e. mutual correlation) and a lack of sensitivity where the model starts fitting the noise in the experimental data used for calibration. The result of an over-fitted model is a reduced error for the calibration data set, but larger errors in validation and prediction with other data sets, i.e. poor generalisation ability. In fact most of the inverse modelling problems are over-parameterised, particularly when prior information about the many of the parameters involved is not available. In literature there is a lack of discussion on the parameter inter-dependency, particularly in the concrete science community, where the growing interest in application of inverse modelling is generally ignoring this important aspect. This paper demonstrates a relative simple and easy procedure to quantify the level of parameter inter-

dependency and sensitivity in order to decide how many parameters can be estimated simultaneously in a single inverse analysis. Presented procedure is general, and may be applied to parameter estimation problems of a wide variety of different processes described by models with or without clearly defined governing equations.

## 2. Mathematical model

The mathematical model adopted is based on the Nernst-Planck equation [1] given as:

$$\frac{\partial C_t(x,t)}{\partial t} = -D_{ef} \left( \frac{\partial^2 c}{\partial x^2} + \frac{zFE}{RT} \frac{\partial c}{\partial x} \right) - r \quad (1)$$

where  $D_{ef}$  is the effective diffusion coefficient ( $m^2/s$ ),  $C_t$  is the total chloride concentration ( $g/dm^3$  of concrete),  $c$  is the concentration of the diffusing/migrating ion ( $g/dm^3$  in pore solution),  $z$  valence number of the ion,  $F$  Faraday constant ( $9.64846 \times 10^4 C mol^{-1}$ ),  $R$  ideal gas constant ( $8.3143 J mol^{-1} K^{-1}$ ),  $T$  absolute temperature (K),  $x$  concrete depth (m),  $t$  time (s),  $E$  electrical field (Volt  $m^{-1}$ ) and  $r$  reaction term ( $mol m^{-3} s^{-1}$ ). The reaction term accounts for the ion interaction process with the solids, which leads to a binding of chlorides from the liquid phase. Reaction term ( $r$ ) can be considered in two ways depending on a relative rates between chloride binding and transport (migration):

- 1) If a binding rate is faster than transport, an equilibrium, i.e., following Freundlich isotherm, is assured:

$$C_b = K_b c^n \quad (2)$$

from where the equilibrium Cl binding amount is then calculated as:

$$r = (1-p)\rho_s \frac{\partial C_b}{\partial t}; \quad \frac{\partial C_b}{\partial t} = \frac{\partial C_b}{\partial c} \frac{\partial c}{\partial t} = K_b c^{(n-1)} n \frac{\partial c}{\partial t} \quad (3)$$

where  $\rho_s$  is density of solids in concrete and  $C_b$  is in units of g/g of solid, and  $c$  is in  $g/dm^3$  of pore solution.

- 2) If a binding rate is slower than transport (due to acceleration by imposed voltage), equilibrium is not achieved and binding reaction kinetic process should be considered. In a simplest form the reaction kinetics can be described as [1]:

$$r = k(c - c_{eq}) = k \left( c - \left( \frac{C_b}{K_b} \right)^{1/n} \right) \quad (4)$$

where  $k$  is a kinetic constant of the binding reaction,  $c$  is the pore solution concentration and  $c_{eq}$  is equilibrium concentration following Freundlich binding isotherm eq. 2. The total chloride concentration ( $C_t$ , g/g of concrete) can be defined as a sum of free and bound chlorides, where

the units are transformed by considering concrete porosity ( $p$ ), density of solids in concrete ( $\rho_s$ ) and saturated concrete density  $\rho_c$ :

$$C_t = \frac{p c + (1-p) \rho_s C_b}{\rho_c} \quad (5)$$

The following system of coupled partial differential is to be solved for the free chlorides (eq. 6, where  $c$  is in g/dm<sup>3</sup> of pore solution) and bound chlorides (eq. 7, where  $C_b$  is in g/g of solid as calculated by Freundlich isotherm):

$$p \frac{\partial c}{\partial t} = -D_{ef} \frac{\partial^2 c}{\partial x^2} - u \frac{\partial c}{\partial x} - k \left( c - \left( \frac{C_b}{K_b} \right)^{1/n} \right) \quad (6)$$

$$(1-p) \rho_s \frac{\partial C_b}{\partial t} = k \left( c - \left( \frac{C_b}{K_b} \right)^{1/n} \right) \quad (7)$$

where  $u = zFE/RT$  is the migration term (in units of m/s) induced by potential gradient  $\Delta E$ , but is analogous to convective (advection) term.

This set of coupled differential equations (6 and 7) is to be used to model the accelerated diffusion test where the chloride ions are driven through a concrete specimen of thickness  $L$  under the presence of an electric field caused by applying a potential difference  $V$  across the specimen. One side ( $x = 0$ ) of the concrete is kept at a fixed concentration of chloride ions  $c_0$  (64.95 g/dm<sup>3</sup>) while the other ( $x = L = 0.05$ m) is assumed to have a zero concentration of chloride ions. Zero initial concentration of free and bound chlorides is set to  $c_i$  and  $C_{bi}$ . Experimental data on accelerated migration (RCM) test on mortar are taken from [1] with following experimental details: RCM applied voltage  $\Delta U = 35$ V for 24h, mortar age 27 days, cement type Cem I 52.5 N (785.6 kg/m<sup>3</sup>), w/c = 0.263, sand (0-2mm) 1382 kg/m<sup>3</sup>, mortar saturated density 2450 g/dm<sup>3</sup>, mortar solids density 2710 g/dm<sup>3</sup>,  $c_0 = 70.9$  g/dm<sup>3</sup>, concrete thickness  $L = 0.05$ m, porosity  $p = 0.1511$ . The parameters to the problem are kinetic  $D_{ef}$ ,  $k$ , and thermodynamic (binding isotherm)  $n$  and  $K_b$ . The values of these parameters are to be estimated (best fitted) to match with experimental results taken from the test. To assess the accuracy of the estimates one needs to quantify the sensitivity of each of the parameters in terms of the proximity and relative importance of each, as well as their interdependency (mutual correlation) in determining the model solution concentrations of chloride ions.

To reduce the number of parameters to be determined simultaneously with the inverse analysis we propose to use experimental data on the binding isotherm for the case of material's boundary condition ( $L=0$ ). Parameter  $K_b$  can be expressed as a function of  $n$  by combining Freundlich's eq. (2) and eq. (5) for the case of concrete surface ( $x=0$ ):

$$K_b = \left( \frac{\rho_c C_{t,(x=0)} - p c}{(1-p) \rho_s} \right) / c^n \quad (8)$$

where  $C_{t,(t=0)}$  is a total chloride concentration at the concrete surface, and the fraction in a bracket represents the  $C_b$  from the Freundlich isotherm (eq 2).  $C_{t,(t=0)}$  can be estimated from experimental profile data, where an equilibrium is established as observed by reaching a maximum in the concentration profile at the near surface of the concrete. Note that the unit for free chloride is  $\text{g/dm}^3$  of pore solution, and bound chloride is expressed in  $\text{g/100g}$  (or  $\text{mas } \%$ ) of concrete. For chloride binding (Freundlich) isotherm the parameters are obtained by using following units: free chloride is also in  $\text{g/dm}^3$  of pore solution, but bound chloride is expressed in  $\text{g/g}$  of solids.

## 2.1 Numerical implementation

The Matlab's built-in non-linear solver "*pdepe.m*" based on Petrov-Galerkin Method of Lines (MOL) [4] is used. The partial differential equation are first converted to ordinary differential equations using a second-order spatial discretisation based on a fixed set of user defined nodes. The success of the MOL arise from its simplicity of implementation and the availability of high-quality time integrators for solving a wide range of ordinary differential equations problems. The time integration is done with the Matlab's built-in solver "*ode15s.m*" for solving stiff system of differential algebraic equations. This routine changes both the time step and the computing formula dynamically in order to minimise computational time for a desired accuracy. The setup of the analysis saw the domain divided into 500 (logarithmically spaced) cells and the time increment was automatically varied in a way to assure a numerical relative and absolute tolerance set to  $10^{-4}$ . The second-order algorithm "*pdepe.m*" and the variable time step integration "*ode15s*" provide a good convergence rate and are robust enough to handle non-linearities and the couplings of the equations.

## 3. Optimisation approach (inverse model)

The optimisation package interfaces with the analysis through a user defined objective function which will return the value of squared 2-norm (i.e. a sum) of the residuals (where, residual is the difference between measured and predicted) for a specified set of design parameters. The function uses the method of lines to provide the chloride concentration at a grid of positions of depth and time. To calculate the residuals the analysis needs to find the concentrations at the same time and position as the test results. Conveniently MATLAB provides a function that will interpolate from a regular grid, as supplied from the analysis, and will estimate the concentration at the depths and time values used in the test measurements and hence allow the residuals to be determined. Cubic interpolation was used to establish the values. Minimisation of an objective function, OF written in matrix notation looks like:

$$\text{OF}(\mathbf{p}) = [\mathbf{C}_{t,\text{measured}} - \mathbf{C}_t(\mathbf{p})]^T [\mathbf{C}_{t,\text{measured}} - \mathbf{C}_t(\mathbf{p})] \quad (9)$$

where  $\mathbf{C}_{t,\text{measured}}$  and  $\mathbf{C}_t$  are the vectors containing the measured and simulated concentrations, respectively, and the superscript T indicates the transpose of a vector. To minimise the least squares norm given by eq. 9 the derivative of  $\text{OF}(\mathbf{p})$  with respect to the unknown parameter is equated to zero.

$$\mathbf{J}(\mathbf{p}) = \left[ \frac{\partial \mathbf{C}_i(\mathbf{p})}{\partial \mathbf{p}} \right] = \begin{bmatrix} \frac{\partial \mathbf{C}_{i,1}}{\partial p_1} & \frac{\partial \mathbf{C}_{i,1}}{\partial p_2} & \dots & \frac{\partial \mathbf{C}_{i,1}}{\partial p_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \mathbf{C}_{i,m}}{\partial p_1} & \frac{\partial \mathbf{C}_{i,m}}{\partial p_2} & \dots & \frac{\partial \mathbf{C}_{i,m}}{\partial p_n} \end{bmatrix} \quad (10)$$

The sensitivity matrix (eq. 10) has a size  $m$  by  $n$ , where  $m$  is the number of data points and  $n$  is the number of parameters. The elements of the sensitivity matrix, called the sensitivity coefficients are thus defined as the first derivative of the estimated total concentration at depth  $d_m$  with respect to the unknown parameter  $p_n$ . The estimation of the parameter for a small value of the magnitude of  $J_{i,j}$  is difficult, because basically the same value for concentration would be obtained for a wide range of parameter values. The estimation problem is ill-conditioned near the initial guess used for the unknown parameter, creating difficulties in the applications of Gauss method. The Levenberg-Marquardt method alleviates such difficulties by damping oscillations and instabilities due to the ill-conditioned character of the problem, by making it large as compared to  $\mathbf{J}^T \mathbf{J}$  if necessary.

The addition of constraints (Tikhonov regularisation) allows the inclusion of prior information about the parameters to be estimated, including equality constraints and limits on the range of parameters. Constraints can alleviate over-parameterisation by constraining the feasible region of parameters. The Levenberg-Marquardt algorithm does not handle bound constraints. Therefore, a Matlab's Trust-region-reflective algorithm was also employed, which implements the Levenberg-Marquardt algorithm using a trust-region approach to solve bound-constrained optimisation problems.

### 3.1 Parameter sensitivity and inter-dependency obtained by eigen value analysis

The relative roles played by parameters in achieving the fit is of extreme importance. The Hessian matrix,  $\mathbf{H}$  can be used to assess the sensitivity of the solution to the value of parameters, however this can be approximated by ignoring the second-order terms in Hessian as:

$$\text{Hessian } \mathbf{H} \approx 2\mathbf{J}^T \mathbf{J} \quad (11)$$

The Hessian contains the second derivatives or curvature of the objective function with respect to the parameters. The greater the curvature the more sensitive the function will be to change in the parameters. An eigen-value analysis of the Hessian will additionally establish the principal directions where maximum sensitivity will occur. A Matlab function,  $[\mathbf{v}, \mathbf{d}] = \text{eig}(\mathbf{H})$ , was used to obtain a diagonal matrix  $\mathbf{d}$  of eigenvalues and matrix  $\mathbf{v}$  whose columns are the corresponding right eigenvectors, so that:

$$\mathbf{H} \mathbf{v} = \mathbf{v} \mathbf{d} \quad (12)$$

Generally, if directions of the eigen vectors correspond to unit vectors, this is taken to suggest that the parameters are independent of each other. This is usually not the case as the formulation of the differential equation would indicate that there is a connection between the parameters. With the eigen vectors being near unity the eigen value can be taken as the curvature of the objective function due to a unit change in a corresponding parameter. Care should however be taken when comparing the eigen values, as the scaling of the parameters may make it non-comparable. A scaling of a Jacobian vectors was employed here by multiplying it by a value of the estimated parameter [5: p.57].

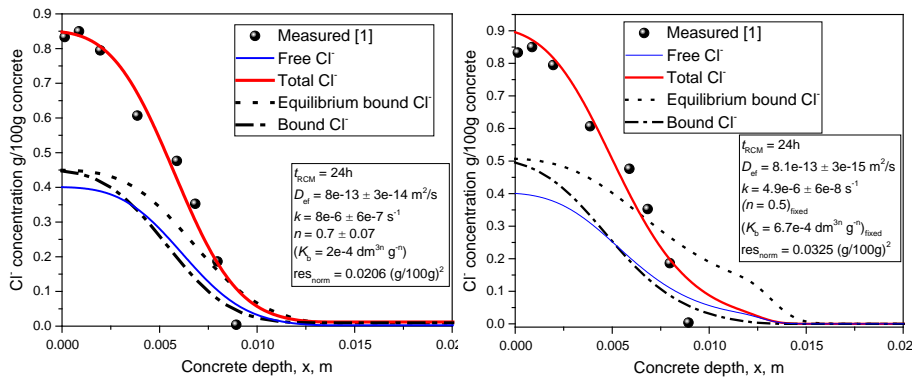


Figure 1. Chloride concentration profiles obtained after 24h RCM test: **(Left)** simultaneous estimation of three parameters : effective diffusivity ( $D_{ef}$ ), binding rate constant ( $k$ ) and binding parameter ( $n$ ); **(Right)** simultaneous estimation of two parameters  $D_{ef}$  and  $k$ , while binding parameters were fixed to  $n=0.5$  and  $K_b=6.7 \cdot 10^{-4} \text{ dm}^{3n} \text{ g}^{-n}$  [1].

Table 1: Results of inverse analysis: estimated parameter values and eigen vectors ( $v_i$ ) with corresponding eigen values.

Number of simultaneously estimated parameters	Parameters	Estimated value	$v_1$	$v_2$	$v_3$	$R_{cond}$
2	$D_{ef}, 10^{-12} \text{ m}^2/\text{s}$	8.1	-0.998	0.059	-	1.3 $10^{-6}$
	$k, 10^{-6} \text{ s}^{-1}$	4.9	-0.059	-0.998	-	
	(fixed: $n=0.5$ ; $K_b=6.7 \cdot 10^{-4} \text{ dm}^{3n} \text{ g}^{-n}$ )	eigen value:	3.359	0.208	-	
3	$D_{ef}, 10^{-12} \text{ m}^2/\text{s}$	8.0	-0.955	-0.234	-0.184	2.0 $10^{-19}$
	$k, 10^{-6} \text{ s}^{-1}$	8.0	0.077	0.564	0.774	
	$n$	0.70	-0.288	-0.792	0.606	
	( $K_b = f(n)$ via eq. 8)	eigen value:	5.403	0.156	0.001	



### 3.1 Results and discussion

Figure 1 shows examples of the simulated total-, actually bound- and free-chloride concentration profiles, all expressed in g/100g of concrete in order to enable their direct relative comparison. In this way, the total chloride content is a direct sum of actually bound and free chloride content. Furthermore, an equilibrium chloride content is plotted as well, which refers to the virtual case when the binding reaction would be fast, i.e. instantaneous (rate constant  $k \rightarrow \infty$ ), that an equilibrium can be attained. However, because the binding reaction has a limited rate, i.e. in comparison to accelerated migration transport rate, the amount of bound chlorides is less than what is possibly in equilibrium. In other words, the high velocity of ions during high-rate (non-steady state) migration transport accelerated by an electrical field, depending on the experimental conditions, hinders significantly binding compared to the equilibrium state that takes place during corresponding diffusion tests without migration. The fit (Figure 1) is very encouraging with the model being able to predict the features of the measured values, in both fitting scenarios. The implementation of the advanced (stiff) time integration and the central differencing was selected to achieve this stability and smoothness of the results. However, due to the dominance of the migration term there may be merit in considering up-winding (adaptive or solution-sensitive) approximation for the derivatives to reflect the directional nature of the ion movement.

The relative roles played by the parameters in achieving the fit is of extreme importance and to help assess this the run of the optimisation uses a set of a different number of parameters. The values of thermodynamic (binding isotherm) parameters  $n$  and  $K_b$  were fixed as obtained from [1] in agreement with independent measurement of chloride binding isotherm. However, the question arises whether, the binding parameters could be simultaneously estimated together with rate parameters  $D_{ef}$  and  $k$ , or whether the parameters are only being too sensitive to the experimental noise in the data, or too correlative with each other.

To reduce the number of parameters to be determined simultaneously with the inverse analysis we propose to use experimental data on the binding isotherm for the case of material's boundary condition (catholyte exposure solution) according to eq 8. From the surface free chloride concentration, equal to amount in catholyte exposure solution, one can obtain the corresponding equilibrated bound concentration from the measured total concentration. Fitting of 2 ( $D_{ef}$  and  $k$ ) or 3 ( $D_{ef}$ ,  $n$ ,  $k$ ) parameters took about 6 or 16 iterations, respectively, with 3 or 5 function evaluations per iteration. Optimality criteria was below  $3e-4$  (2 parameter fit) or below  $6e-5$  (for 3 parameters fit), indicating that an optimum was always possible. Comparison of values on estimated parameters (Table 1) between the different scenarios (with different amount of simultaneously fitted parameters) shows no significant variance in  $D_{ef}$ , but high variance in  $k$  and  $n$ . This indicates a confidence and robustness of the inverse analysis for the estimation of  $D_{ef}$ , which can be explained by the fact that  $D_{ef}$  exhibits the highest sensitivity and lowest interdependency with other parameters. However, this results should still be taken with care when fitting more than 2 parameters ( $D_{ef}$  and  $k$ ) to other experimental data with different measurement errors. To assure accuracy simultaneous parameter estimation should always be accompanied by the eigen value analysis in order to demonstrate the degree of interdependency and sensitivity of  $D_{ef}$ . The accuracy of  $k$ ,  $n$  and  $K_b$  should be taken with care, and their estimates are only a first approximation, and fitting of 4 simultaneous parameters is definitely not recommended. Eigen values matrix is a sparse matrix with a non-zero (eigen) values positioned on its diagonal. These eigen values rate the relative contribution of eigen vectors on the

objective function. The eigen vectors comprise the effect of the principal parameter together with interactions with other parameters. Therefore, the eigen values actually quantify the relative impact of the change of a principal parameter together with a combination of other inter-dependent parameters within each eigen vector. When parameters are independent (orthogonal), eigen vectors have unity values, where the position of the unity element (a  $v_i$  row in Table 1) within the vector indicates to which parameter is this vector attributed.

For example case with 2 fitting parameters in Table 1, eigen vector (column)  $v_1$  has a maximal (-0.998) value of first element, whose vector position (first row) corresponds to the principal parameter  $D_{ef}$ . Deviation from unity absolute value of the principal parameter, and deviation from zero absolute value of the other row elements (corresponding to the other parameters, here -0.059) quantifies their correlation, i.e. the parameter inter-dependency. The dependency in  $v_1$  is very low, 0.02, which means that only about 2% of the variance in  $D_{ef}$  is due to its interaction with the parameter  $n$ . This vector exhibits the highest eigen value (higher than for  $v_2$ ) indicating the highest relative sensitivity. By the values within the eigen vector  $v_1$  one can determine that this highest sensitivity corresponds predominately to  $D_{ef}$  (highest value of first row, -0.998) with very low interdependency with other parameters (close to zero values of other rows). If parameters have been interdependent, eigen vectors will not be orthogonal, i.e. more elements have significant deviations from non-zero values. Eigen values show the sensitivity of the eigen vector, which is a combination of the interdependent parameters. For example in simultaneous estimation of 3 parameters (Table 1), vector  $v_2$  has a 2<sup>nd</sup> place in vector sensitivity ranking which is moreover attributed mainly to parameter  $n$  (79.2%) but also to its significant interdependence with  $k$  and  $D_{ef}$ . In this dependency, we see that  $n$  is more correlated to  $k$  than to  $D_{ef}$  as indicated by a difference between absolute values of 2<sup>nd</sup> (0.564) and 1<sup>st</sup> (0.234) element of  $v_2$  vector. This is also confirmed in  $v_3$ , where the main parameter  $k$  (0.774) is more correlated to  $n$  (0.606) than to  $D_{ef}$  (0.184). The eigen vector  $v_3$ , has the least relative sensitivity to minimisation of objective function, indicated by an eigen value of 0.001, compared to eigen values of  $v_1$  (5.403) and  $v_2$  (0.156). This means that the combined effect of (primarily) parameter  $k$  (in that eigen vector), together with interdependency between  $k$  and  $D_{ef}$ , has the least relative sensitivity. Likewise, the eigen vector (column)  $v_1$  has a maximal (-0.955) absolute value of the first eigen vector element, whose vector position (first row) corresponds to the principal parameter  $D_{ef}$ . Deviation level from unity of the principal parameter, while deviation from zero of the other row elements corresponding to the other parameters, quantifies their correlation, i.e. the parameter inter-dependency. By reducing the number of fitting parameters, correlation/dependency and  $R_{cond}$  (inverse of a condition number for  $\mathbf{J}^T\mathbf{J}$ ) values improve rapidly (Table 1): dependency rapidly reduces and  $R_{cond}$  rapidly increases. Simultaneous fitting of four parameters results in a highly ill-conditioned system (high  $R_{cond}$ ) due to overfitting and overparametrisation, where parameters show extreme degree of interdependency (results not shown in Table here). Generally,  $D_{ef}$  was found to exhibit the highest sensitivity and lowest interdependency, i.e. a highest eigen value for an eigen vector that indicates  $D_{ef}$  as a primary parameter. This indicates a confidence and robustness of the inverse analysis for the estimation of  $D_{ef}$ . However, care should be taken when fitting more than 2 parameters ( $D_{ef}$  and  $k$ ) to simultaneously also estimate the binding parameters  $n$  and  $K_b$  defining the Freundlich isotherm.

Commented [A1]: I am removing this sentence as it is confusing and I couldn't correct it.

#### 4. Conclusion

Inverse modeling of the transport of chloride ions through cement based materials in the presence of an electrical field has been analysed using the MOL discretisation/integration modeling approach within the Levenberg-Marquard optimisation loop, focusing on a parameter sensitivity and interdependency. The technique is very flexible and performed well producing stable, smooth and convergent results. The parameters can readily be specified as being fixed or included as variables allowing the approach to model and optimise a range of situations. A confidence and robustness of the inverse analysis was demonstrated for the estimation of diffusivity,  $D_{ef}$ , by the fact that  $D_{ef}$  exhibits the highest sensitivity and lowest interdependency with the other simultaneously parameters. However, care should be taken when fitting more than 2 parameters ( $D_{ef}$  and  $k$ ) to simultaneously estimate the binding parameters ( $n$  and  $K_b$  defining the Freundlich isotherm). The estimates of  $n$  and  $K_b$  are only a first approximation, as they are firstly highly correlated to each other, and secondly also with parameter  $k$ . In this case, the proposed eigen value and parameter interdependency analysis should be part of the inverse analysis in order to demonstrate the degree of interdependency, sensitivity and accuracy of the estimates.

An approach was proposed to reduce the number of simultaneously fitted parameters from four to three, where  $n=f(K_b)$  following eq. 8 considers an establishment of a binding equilibrium at the concrete surface.

#### References

- [1] Truc O., Ollivier J.P., Nilsson L.O., Numerical simulation of multi-species transport through saturated concrete during a migration test - MsDiff code, *Cem Concr Res*, 30 (2000), 1581-1592
- [2] Samson E., Marchand J., Modeling the transport of ions in unsaturated cement-based materials, *Comp Struct*, 85(2007), 1740–1756
- [3] Schiesser W.E., Griffiths G.W., A Compendium of Partial Differential Equation Models, Method of Lines Analysis with MATLAB, Cambridge University Press 2009.
- [4] Spiesz P., Brouwers H.J.H., Influence of the applied voltage on the Rapid Chloride Migration (RCM) test, *Cem Concr Res* 42 (2012), 1072–1082
- [5] Steffes-Lai D., Approximation Methods for High Dimensional Simulation Results Parameter Sensitivity Analysis and Propagation of Variations for Process Chains, Logos Verlag Berlin GmbH, 2014. ISBN 978-3-8325-3696-1, pp.233