

RESEARCH ARTICLE

Open Access

# Estimation of diffusion coefficients from voltammetric signals by support vector and gaussian process regression

Martin Bogdan<sup>1,3</sup>, Dominik Brugger<sup>1</sup>, Wolfgang Rosenstiel<sup>1</sup> and Bernd Speiser<sup>2\*</sup>

## Abstract

**Background:** Support vector regression (SVR) and Gaussian process regression (GPR) were used for the analysis of electroanalytical experimental data to estimate diffusion coefficients.

**Results:** For simulated cyclic voltammograms based on the EC, E<sub>qr</sub>, and E<sub>qr</sub>C mechanisms these regression algorithms in combination with nonlinear kernel/covariance functions yielded diffusion coefficients with higher accuracy as compared to the standard approach of calculating diffusion coefficients relying on the Nicholson-Shain equation. The level of accuracy achieved by SVR and GPR is virtually independent of the rate constants governing the respective reaction steps. Further, the reduction of high-dimensional voltammetric signals by manual selection of typical voltammetric peak features decreased the performance of both regression algorithms compared to a reduction by downsampling or principal component analysis. After training on simulated data sets, diffusion coefficients were estimated by the regression algorithms for experimental data comprising voltammetric signals for three organometallic complexes.

**Conclusions:** Estimated diffusion coefficients closely matched the values determined by the parameter fitting method, but reduced the required computational time considerably for one of the reaction mechanisms. The automated processing of voltammograms according to the regression algorithms yields better results than the conventional analysis of peak-related data.

**Keywords:** Support vector regression, Gaussian process regression, Diffusion coefficient, Principal component analysis, Voltammetry, Reaction mechanism

## Background

Voltammetric signals are measurements of the current flowing through an electrode as a function of an externally controlled electrode potential. For example, in a simple case for an initial oxidation, during a single cycle in cyclic voltammetry the electrode potential first increases linearly with time and, upon reaching the switching potential, decreases linearly back to the starting potential [1,2]. It has been argued that voltammetric techniques have found widespread use due to their high sensitivity, adequate selectivity, and ready availability of instrumentation [3]. Measurements of cyclic voltammetric signals provide

detailed information about reactions which include, or are coupled to, electron transfer steps, and thus enable the analysis of the underlying mechanisms [4]. In a special context, these measurements are used, for example, to study the release of neurotransmitters [5], and to characterize the electrochemical properties of recording and stimulation microelectrodes in neuroscience research [6].

Automated acquisition of experimental data [7,8] and computer simulations of electrochemical systems [9,10] play an important role in modern electrochemistry. Due to the wide applicability and high speed of voltammetric experiments [3], data analysis methods are required to aid electrochemists in extracting knowledge about electrochemical systems [11-14]. Recently proposed data analysis methods include, for example, multi-parameter estima-

\*Correspondence: bernd.speiser@uni-tuebingen.de

<sup>2</sup>Institut für Organische Chemie, Universität Tübingen, Auf der Morgenstelle 18, D-72076 Tübingen, Germany

Full list of author information is available at the end of the article

tion from hypersurface models [15,16], artificial neural networks for classifying voltammetric signals by reaction mechanism [17], and bootstrap resampling to extract system parameters and their error distributions [18].

The diffusion coefficient  $D$  is an important physical parameter of the species involved in an electrochemical reaction, that describes diffusional transport. Since Nicholson and Shain's classical treatment [1], diffusion coefficients are directly extracted from voltammetric signals based on theoretical relations (Randles-Sevcik equation), valid for particular electrode reaction mechanisms. Recently analytical solutions for calculating the diffusion coefficient from flux data have also been proposed [19,20], but are restricted to pure diffusive and diffusive-convective conditions. Semiintegral analysis provides a "linearization" method that allows  $D$  to be determined for single electron transfers without kinetic complications [21]. As an alternative, fitting of simulated voltammetric features to experimental data [11,15,16,22], or full current/potential curves [23,24] may provide values for  $D$ . Both approaches have limitations: Theoretical relationships are only valid for certain reaction mechanisms and kinetic schemes, while the fitting of simulated data requires formulation of a reasonable mechanistic hypothesis, substantial computation time and is very sensitive to the initialization of the electrochemical system parameters [15]. Non-electrochemical approaches to determine  $D$  include PGSE-NMR spectroscopy [25,26]. However, these require expensive instrumentation and considerable additional expertise.

To overcome such limitations, we investigate the estimation of diffusion coefficients from experimental cyclic voltammograms by means of two function estimation techniques, support vector regression (SVR) and Gaussian process regression (GPR) [27,28]. Support vector machines, as a tool for both regression and classification, have recently gained popularity across different application fields such as genetics [29], neuroscience [30,31], quantum chemistry [32], spectroscopy [33-35], and electrochemistry [36]. Similar to support vector machines, Gaussian processes have lately seen a revival of interest due to their combination with covariance kernels [28] and were successfully applied to problems in (bio)chemistry and robotics concerning micro-array analysis [37], and decoding of spike trains [38].

## Methods

In the following,  $f$  will denote a scalar function, mapping vectors  $x \in \mathbb{R}^n$  to a scalar  $y \in \mathbb{R}$ . Then, the estimation of diffusion coefficients from voltammetric signals is equivalent to estimating the unknown function  $f(x) \mapsto y$ , where  $x$  is a cyclic voltammogram (CV) and  $y \in \mathbb{R}$  the diffusion coefficient  $D$ . Function  $f$  hence describes the relationship between

experimentally acquired data (CVs) and an unknown physical property ( $D$ ) of the electrochemical species. The following Sections "Support vector regression" and "Gaussian processes" introduce two different techniques for estimating function  $f$ .

### Support vector regression

Support Vector Regression (SVR) [27] is a method to estimate  $f(x) \mapsto y$ , given a set of data points  $(x_i, y_i)$ ,  $i = 1, \dots, m$ . In the application at hand each data point  $(x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}$  consists of a complete CV and the respective diffusion coefficient  $D$ . To introduce the SVR algorithm, we first consider estimation of linear functions  $f(x) = \langle w, x \rangle + b$ , where  $w \in \mathbb{R}^n$  denotes the weight vector and  $b \in \mathbb{R}$  the bias term, or offset. For simple linear regression the parameters  $w$  and  $b$  are determined by minimizing the quadratic loss  $l^2(f(x_i) - y_i) = (f(x_i) - y_i)^2$  (Figure 1A), across all of the data points. In other words, one solves the optimization problem (1).

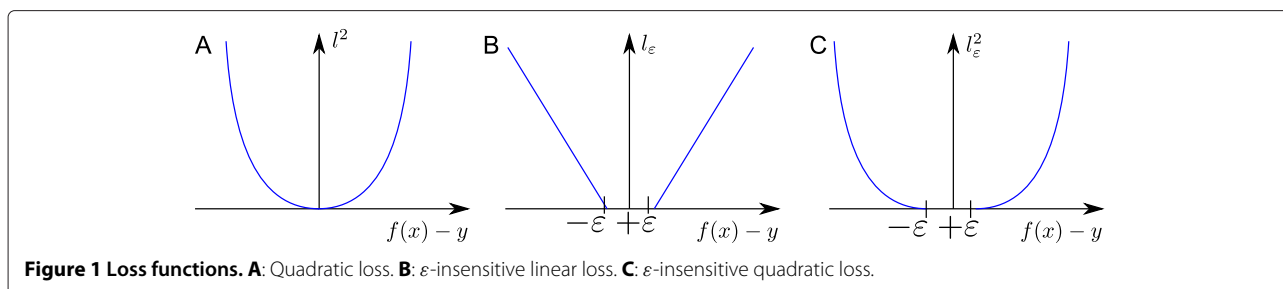
$$\min_{w,b} \sum_{i=1}^m (f(x_i) - y_i)^2 \quad (1)$$

In equation (1), the sum of all  $(f(x_i) - y_i)^2$  is minimized with respect to the weight vector  $w$  and offset  $b$ . After finding  $w$  and  $b$ , diffusion coefficients are estimated for previously unseen cyclic voltammograms by evaluating  $f$ . In general, function  $f$  relating voltammograms and diffusion coefficients will not be linear and we will describe the extension to estimating nonlinear functions later in this paragraph.

Usually, one is interested in a high prediction accuracy on data not available during the optimization process, that is, one wants a function that generalizes well beyond the given set of training data points. To improve the generalization performance of the estimated function the space of solutions for  $w$  is restricted by minimizing  $\|w\|^2$  in addition to the squared loss (equation 2)

$$\min_{w,b} \|w\|^2 + C \sum_{i=1}^m (f(x_i) - y_i)^2, \quad (2)$$

where the parameter  $C$  controls the complexity of the solution. Large values of  $C$  lead to a smaller error on the training data points at the expense of a complex function, while small values of  $C$  result in simple (flat) linear functions at the expense of larger training errors. The ridge regression [39] problem in Equation 2 can be transformed into the SVR optimization problem by replacing the quadratic loss with the  $\varepsilon$ -insensitive linear loss,



$l_\epsilon(f(x_i) - y_i) = \max\{0, f(x_i) - y_i\}$  which is shown in Figure 1B:

$$\min_{w,b} \|w\|^2 + C \sum_{i=1}^m \xi + \xi^* \quad (3)$$

subject to:

$$\begin{aligned} f(x_i) - y_i &\leq \epsilon + \xi \\ y_i - f(x_i) &\leq \epsilon + \xi^* \\ \xi, \xi^* &\geq 0 \end{aligned}$$

From Equation 3 it is clear that only data points with  $|f(x_i) - y_i| > \epsilon$  contribute to the solution, since otherwise the slack variables  $\xi, \xi^*$  are zero. The choice of the  $\epsilon$ -insensitive loss function hence induces a sparse solution that only depends on data points with non-zero loss, which are called 'support vectors' [27]. In practice the  $\epsilon$ -zone of the loss makes the function estimation more robust against measurement noise in the target values  $y_i$ , and the  $\epsilon$  parameter is set to match the level of noise in the target values, if known. The automatic choice of parameters  $C$  and  $\epsilon$  will be explained later. Robustness of  $f$  with respect to outliers in the target values is achieved by the linear part of the loss function (Figure 1B). Since outliers are not an issue for the envisaged estimation of diffusion coefficients, where the training data set consists of simulated cyclic voltammograms, the loss function is replaced by the  $\epsilon$ -insensitive quadratic loss,  $l_\epsilon^2(f(x_i) - y_i) = \max\{0, (f(x_i) - y_i)^2 - \epsilon\}$ , shown in Figure 1C. This exchange of the loss function allows to solve the SVR optimization problem by the Newton algorithm for linear [40] and nonlinear function estimation [41,42]. For the  $\epsilon$ -insensitive quadratic loss the optimization problem in Equation 3 transforms into the unconstrained optimization problem (4):

$$\min_{w,b} \|w\|^2 + C \sum_{i=1}^m l_\epsilon^2(\langle w, x_i \rangle - y_i). \quad (4)$$

Linear functions might not provide the necessary flexibility for the estimation of diffusion coefficients from experimental data. To extend SVR to nonlinear function estimation one assumes that the function  $f(x)$  resides in a Hilbert space  $\mathcal{H}$ . Under this assumption the minimization of  $\|w\|^2$  is replaced by the minimization of the squared

function norm  $\|f\|_{\mathcal{H}}^2$  in Hilbert space  $\mathcal{H}$ , and Equation 4 can be reformulated as:

$$\min_f \|f\|_{\mathcal{H}}^2 + C \sum_{i=1}^m l_\epsilon^2(f(x_i) - y_i). \quad (5)$$

In this form the optimization problem (5) is not solvable, since  $f$  is unknown. Yet, according to the representer theorem [43] the evaluation of  $f$  at point  $x_i$  is given by a linear combination of kernel functions:

$$f(x_i) = \sum_{i=1}^m \beta_i k(x_i, x_j). \quad (6)$$

This permits the minimization of  $l_\epsilon^2(f(x_i) - y_i)$  in terms of the coefficients  $\beta_i$  instead of  $f$ . Further, Equation 6 allows one to rewrite the squared norm of the function:

$$\|f\|_{\mathcal{H}}^2 = \langle f, f \rangle_{\mathcal{H}} = \sum_{i,j} \beta_i \beta_j \langle k(x_i, \cdot), k(\cdot, x_j) \rangle_{\mathcal{H}}.$$

In the final step the dot product between kernel functions can be expressed as  $\langle k(x_i, \cdot), k(\cdot, x_j) \rangle_{\mathcal{H}} = k(x_i, x_j)$ , where we exploited the reproducing property [44] of the Hilbert space given by  $f(x_i) = \langle f, k(\cdot, x_i) \rangle_{\mathcal{H}}$ . By combining these reformulations, the nonlinear SVR optimization problem is:

$$\min_{\beta,b} \beta^T K \beta + C \sum_{i=1}^m l_\epsilon^2(K_i \beta + b - y_i), \quad (7)$$

where  $K_{ij} = k(x_i, x_j)$  is the kernel matrix and  $K_i$  denotes its  $i$ -th row. Similar to the linear case, the objective function in (7) contains a regularization term,  $\|f\|_{\mathcal{H}}^2 = \beta^T K \beta$ , and a loss function term,  $l_\epsilon^2(K_i \beta + b - y_i)$ . As discussed above for the linear case, parameter  $C$  controls the complexity of the estimated function.

Table 1 lists the two kernel functions which are subsequently used to estimate diffusion coefficients from

**Table 1 Kernel functions**

Type	Function
Linear	$k(x_i, x_j) = \langle x_i, x_j \rangle$
RBF	$k(x_i, x_j) = \exp(-\gamma \ x_i - x_j\ ^2)$

cyclic voltammograms. The parameter  $\gamma$  for the radial basis function (RBF) kernel, together with the regularization parameter  $C$ , and the loss function parameter  $\varepsilon$  were automatically chosen during SVR function estimation by minimizing a bound on the leave-one-out error. The leave-one-out is the average of errors across single data points that were removed from the set before the function estimation. It is an almost unbiased estimate of the expected error on unseen data, but requires the function to be estimated  $m$  times. To avoid this, we minimized a bound on the leave-one-out error with a Quasi-Newton algorithm [45,46]. The described algorithms were implemented within MATLAB®.

### Gaussian processes

A Gaussian process is defined as a collection of random variables, any finite number of which have consistent joint Gaussian distributions [28]. A Gaussian process generalizes the concept of the Gaussian distribution over vectors to a distribution over functions and is fully defined by its mean function  $\bar{m}(x)$  and covariance function  $k(x, x')$ . In order to draw samples from a Gaussian process one first evaluates the mean and covariance function at a finite set of data points to obtain a mean vector  $\mu_i = \bar{m}(x_i) \in \mathbb{R}^m$  and covariance matrix  $\Sigma_{ij} = k(x_i, x_j) \in \mathbb{R}^{m \times m}$ , and subsequently draws a vector of function values  $f \sim \mathcal{N}(\mu, \Sigma)$  where  $\mathcal{N}(\mu, \Sigma)$  denotes a multi-dimensional Gaussian distribution with mean vector  $\mu$  and covariance matrix  $\Sigma$ . Specifying the mean and covariance function thus reflects prior knowledge about the properties, for example, the smoothness of the estimated function.

Finding the function values  $f_*$  for previously unseen test data points is possible by considering the joint distribution:

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma_* \\ \Sigma_*^T & \Sigma_{**} \end{bmatrix} \right) \quad (8)$$

where  $\mu_*$  is the vector of test means,  $\Sigma_*$  the covariance for training-test data points and  $\Sigma_{**}$  the covariance for test data points. Since the joint distribution is Gaussian, the posterior distribution of  $f_*$ , given the known function values at the training data points, is again Gaussian:

$$f_* | f \sim \mathcal{N} \left( \mu_* + \Sigma_*^T \Sigma^{-1} (f - \mu), \Sigma_{**} - \Sigma_*^T \Sigma^{-1} \Sigma_* \right) \quad (9)$$

Thus calculating the distribution of  $f_*$  just requires evaluation of the mean vectors and covariance matrices, and the inversion of the training set covariance matrix by a Cholesky decomposition [47].

The choice of a particular mean and covariance function corresponds to the training of a Gaussian process. In the absence of precise prior information about the functional

relationship underlying the data it is best to parameterize the mean and covariance function and estimate the parameters from the available data. Usually the training is restricted to identifying a suitable covariance function, after subtracting the empirical mean from the regression targets  $y_i$ . Table 2 lists the covariance functions considered for the estimation of diffusion coefficients. An additional term  $\sigma_n \delta_{ij}$  is added to each covariance function, with  $\delta_{ij}$  being Kronecker's delta, in order to model Gaussian noise in the regression targets.

The parameters  $\theta$  of the covariance function, e.g.  $\theta = (\sigma^2, l)$  for the squared exponential covariance function, are determined by maximizing the probability of the data given the parameters. Since the data distribution is assumed to be Gaussian the logarithm of this probability is [28]:

$$\begin{aligned} L &= \log p(y|x, \theta) \\ &= -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (y - \mu)^T \Sigma^{-1} (y - \mu) - \frac{m}{2} \log(2\pi) . \end{aligned} \quad (10)$$

After calculating the partial derivative of Equation 10 with respect to  $\theta$  one can use a conjugate gradients algorithm to optimize the parameters. It should be noted that the first term in the objective function (10) regularizes the solution, while the second term measures the quality of the data fit, and the third term is a constant independent of the data. In contrast to the SVR algorithm (Section "Support vector regression") there is no regularization parameter  $C$  that needs to be set, since there is an implicit trade-off between function complexity and data fit. For the Gaussian process regression we used the freely available GPML toolbox for MATLAB® [28].

### Nicholson-Shain equation approach

The analysis of voltammetric measurements relates a system parameter [11], diffusion coefficient  $D$ , and the experimental variables, such as the initial concentration  $c_0$ , the electrode area  $A$ , scan rate  $\nu$ , and temperature  $T$ , as well as other parameters (here: number of transferred electrons  $n$ ), of the electrochemical system to the electric current  $i$  flowing through the electrode. For the dimensionless current function  $\chi$  the relationship (11) holds [1].

$$\sqrt{\pi} \chi = \frac{i}{nFAc_0 \sqrt{D \frac{nF}{RT} \nu}} , \quad (11)$$

**Table 2 Covariance functions**

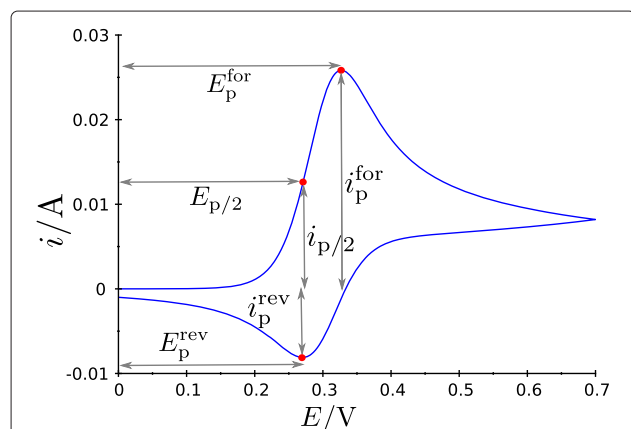
Type	Function
Linear	$k(x, x') = \sigma^2 (1 + \langle x, x' \rangle)$
Squared exp.	$k(x, x') = \sigma^2 \exp(-\ x - x'\ ^2 / 2l^2)$

with Faraday constant  $F = 96485.339 \text{ C mol}^{-1}$ , and gas constant  $R = 8.314472 \text{ J mol}^{-1} \text{ K}^{-1}$ . If the reaction under investigation is a simple reversible electron transfer, the dimensionless current at the peak approaches a value [1], i.e.  $\sqrt{\pi}\chi_p = 0.4463$ , independent of any parameter describing the electrochemical system. During voltammetric experiments the current is measured, while  $\nu$ ,  $T$ ,  $A$ , and  $c_0$  are known or under control of the experimenter. Therefore, the diffusion coefficient of the electrochemical species can be determined by solving Equation 11, in particular at the voltammetric peak:

$$D = \frac{(i_p^{\text{for}})^2 RT}{(0.4463 n c_0 A F)^2 n F \nu} \quad (12)$$

where the current of the forward peak  $i_p^{\text{for}}$  (Figure 2) is extracted from the experimental cyclic voltammogram.

Although diffusion coefficients can be calculated from Equation 12 given an experimental cyclic voltammogram, the assumption of a known dimensionless current  $\sqrt{\pi}\chi_p$  is violated for electrode reactions deviating from the simple diffusion-controlled one-electron transfer. For more complex cases,  $\sqrt{\pi}\chi$  depends on various variables [1], including rate constants that are often unknown, and examples are the  $E_{\text{qr}}$  (quasi-reversible electron transfer), the EC (reversible electron transfer with irreversible chemical follow-up reaction), and the  $E_{\text{qr}}\text{C}$  (quasi-reversible electron transfer with irreversible chemical follow-up reaction) mechanisms, described in Section "Results and discussion". Then, the peak current  $\chi_p$  changes in a nonlinear fashion depending on the kinetic rate constants of the electron transfer or the follow-up reaction. For the case of the EC mechanism, the dependence on the dimensionless follow-up rate constant  $\kappa_1 = k_1/a$  (with  $k_1$  being the first order rate constant,



**Figure 2 Example cyclic voltammogram.** The forward peak, half peak, and reverse peak potentials ( $E_p^{\text{for}}$ ,  $E_{p/2}$ ,  $E_p^{\text{rev}}$ ), and currents ( $i_p^{\text{for}}$ ,  $i_{p/2}$ ,  $i_p^{\text{rev}}$ ), which are used to calculate the manually extracted features are indicated.

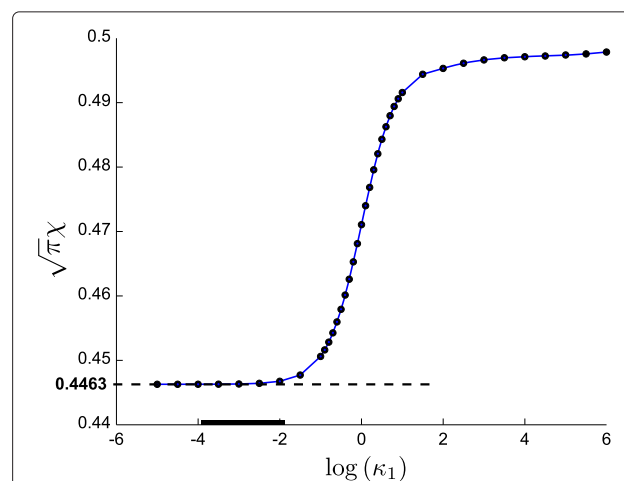
and  $a = nF\nu/RT$ ) is shown in Figure 3. In this case calculation of the diffusion coefficient by the Nicholson-Shain equation is only possible if the rate constant of the EC mechanism has a very small value of  $\log(\kappa_1) < -3$ . If the exact value of the rate constant is unknown, it might still be possible to estimate the diffusion coefficient by regression algorithms such as SVR (Section "Support vector regression"), or GPR (Section "Gaussian processes").

### Simulations

Voltammetric measurements were simulated by the CVSIM program included in the EASIEEST software package [48]. Common parameters used in all simulations are listed in Table 3 while the remaining parameter values of the electrochemical system are given separately in Section "Estimation from simulated data" for each analyzed mechanism. In all simulation runs the CVSIM program was configured to use the METAN1 integrator and the technique of spline collocation [49] with 10 collocation points.

### Fitting of simulation parameters

Fitting simulation parameters by globally minimizing the sum of squared errors between experimental and simulated cyclic voltammograms was used to identify the formal potential  $E^0$ , the heterogeneous electron transfer rate constant  $k_s$ , and  $D$  for the  $E_{\text{qr}}$  and  $E_{\text{qr}}\text{C}$  mechanisms, as well as the homogeneous chemical rate constant  $k_1$  for the  $E_{\text{qr}}\text{C}$  mechanism from the experimental cyclic voltammograms. The resulting  $D$  were used as approximations



**Figure 3 Variation of the dimensionless peak current  $\sqrt{\pi}\chi_p$  with the dimensionless rate constant  $\kappa_1$  for the EC reaction mechanism.**

The dimensionless peak current  $\sqrt{\pi}\chi_p$  is constant only for very small ( $\log(\kappa_1) < -3$ ) and very large ( $\log(\kappa_1) > 4$ ) values of the rate constant. In the former case, the limiting value of 0.4463 is approached; for an explanation of the black bar on the abscissa, see text, Section "EC mechanism — dependence on  $k_1$ ".

**Table 3 Common simulation parameters for all mechanisms**

Parameter and unit	Value
Scan rate $v$ ( $V s^{-1}$ )	0.2
Potential step size $\Delta E$ (mV)	1
Initial concentration $c_0$ (mmol/L)	0.4
Temperature $T$ ( $^{\circ}K$ )	293.15
Electrode area $A$ ( $cm^2$ )	0.064
Symmetry factor $\alpha$	0.5

to the real value. To achieve a homogeneous fit across all experimental voltammograms and avoid large deviations for small-amplitude voltammograms, the currents of simulated and experimental voltammograms were scaled to the interval  $[-1, 1]$ , prior to computing the objective function. The minimization of the sum of squared errors measure was carried out by an interior point algorithm [50] as implemented in the KNITRO software library [51]. Values for the diffusion coefficients obtained by this approach served as a reference for judging the accuracy of coefficients estimated by SVR and GPR for the experimental cyclic voltammograms of the organometallic complexes (Section “Estimations from experimental data”).

## Results and discussion

In a first step (Section “Estimation from simulated data”) the approach based on the Nicholson-Shain equation and the regression algorithms SVR and GPR were used to estimate diffusion coefficients for simulated cyclic voltammograms with known diffusion coefficients. This allowed us to compare the performance of the different methods in terms of accuracy of the estimated diffusion coefficients. Furthermore, the simulated data helped to analyze the dependence of accuracy on the rate constants of the underlying reaction mechanism. In a second step (Section “Estimations from experimental data”) the regression algorithms, trained on the simulated data, were used to estimate  $D$  for experimental cyclic voltammograms with unknown diffusion coefficients.

### Estimation from simulated data

Cyclic voltammograms were simulated as described in Section “Simulations” for the following three reaction mechanisms with the respective model parameters (Table 4):

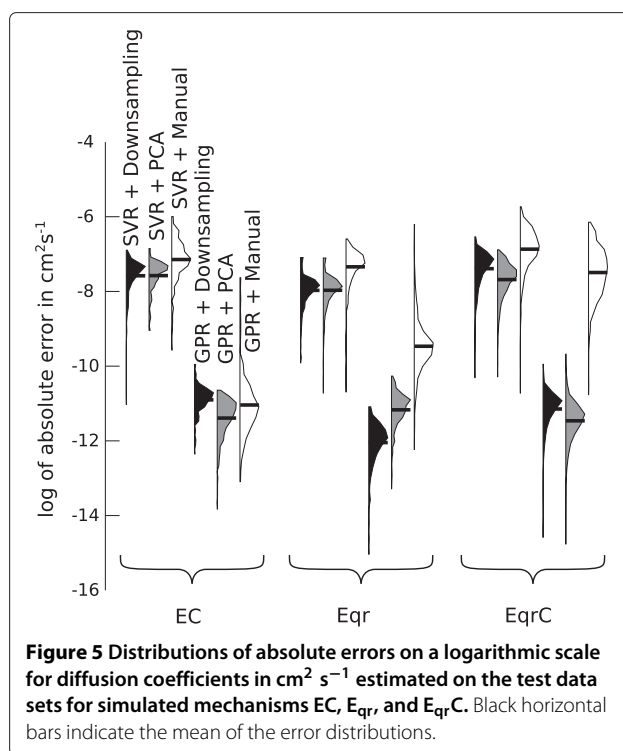
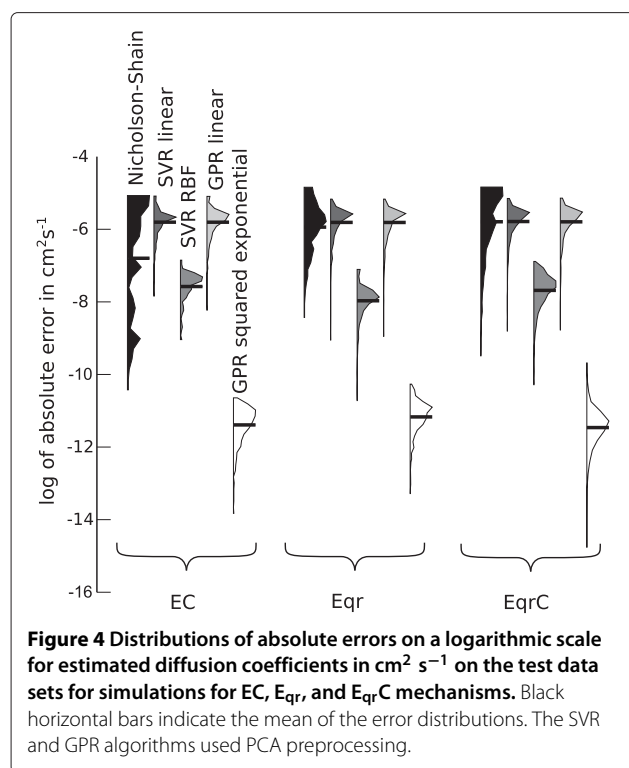


**Table 4 Simulation parameters for the EC,  $E_{qr}$ , and  $E_{qr}C$  mechanism**

EC: $A \xrightleftharpoons{\pm e} B \xrightarrow{k_1} C$	
$k_1$ ( $s^{-1}$ )	0.001, 0.01, 0.1, 1, 10, 100, 1000
$D$ ( $cm^2 s^{-1}$ )	$1 \cdot 10^{-6}$ , $1.5 \cdot 10^{-6}$ , ..., $5 \cdot 10^{-5}$ , $5.05 \cdot 10^{-5}$
$E^0$ (V)	0.3
$E_{start}$ (V)	0
$E_{rev}$ (V)	0.7
$E_{qr}: A \xrightleftharpoons{\pm e} B$	
$k_s$ ( $cm s^{-1}$ )	0.001, 0.005, 0.01, 0.02, ..., 0.1, 0.5, 1
$D$ ( $cm^2 s^{-1}$ )	as EC
$E^0$ (V)	0.2108
$E_{start}$ (V)	0
$E_{rev}$ (V)	0.5
$E_{qr}C: A \xrightleftharpoons{\pm e} B \xrightarrow{k_1} C$	
$k_1$ ( $s^{-1}$ )	as EC
$k_s$ ( $cm s^{-1}$ )	as $E_{qr}$
$D$ ( $cm^2 s^{-1}$ )	as EC
$E^0$ (V)	0.2775
$E_{start}$ (V)	0
$E_{rev}$ (V)	0.6

For each mechanism one combination of diffusion coefficient and rate constant(s) was used per simulation run (Table 4). The resulting simulated data set comprised a total of 700 simulated voltammograms for the EC mechanism, 1400 for the  $E_{qr}$  mechanism, and 2800 for the  $E_{qr}C$  mechanism. This full data set was randomly partitioned into training and test data sets, each containing 50% of the simulated cyclic voltammograms. Only the training data set was used for the function estimation by SVR and GPR, while the performance of each algorithm was assessed on the test data set.

First we compared the accuracy of the diffusion coefficients calculated by the approach based on the Nicholson-Shain equation, SVR with linear kernel, SVR with RBF kernel (Table 1), GPR with linear covariance function, and GPR with squared exponential covariance function (Table 2) for each of the three reaction mechanisms (Figure 4). For the simulated data the true value of the diffusion coefficients is known and can be used as a reference. Prior to applying the SVR and GPR algorithm we reduced the dimensionality of the simulated CVs from 1401 (each dimension corresponds to one current value of the CV) to 5, by projecting the data to the subspace spanned by the 5 dominant principal components. This preprocessing by principal component analysis (PCA) explained 99% of the variance in the EC mechanism data,

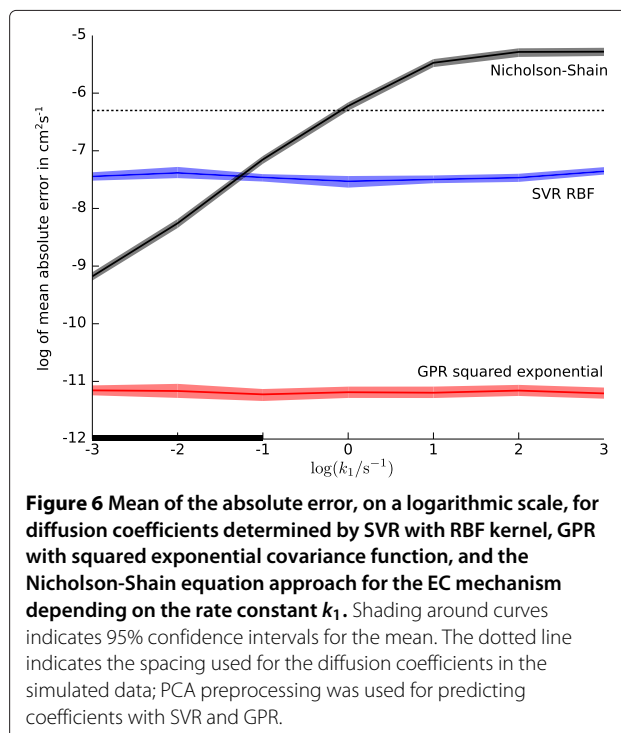


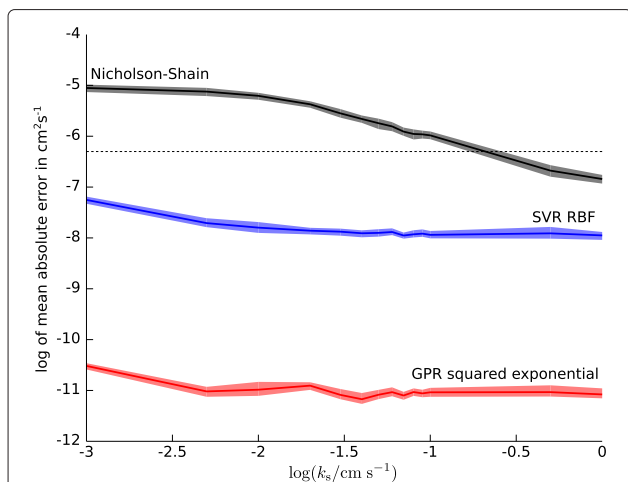
and 99%/98% of the variance in the  $E_{qr}/E_{qr}C$  mechanism data respectively.

In the Nicholson-Shain Equation 12 the diffusion coefficient is a quadratic function of the forward peak current  $i_p^{for}$ . It is therefore not surprising that the nonlinear functions estimated by SVR with RBF kernel and GPR with the squared exponential covariance function are better suited to describe the relationship between cyclic voltammogram and diffusion coefficient for all investigated mechanisms. There is a significant difference between the means of the error distributions of SVR with linear/RBF kernel, and GPR with linear/squared exponential covariance function, as shown in Figure 4. In addition, the nonlinear functions estimated by SVR and GPR consistently yield lower errors on average than the Nicholson-Shain equation approach for all the reaction mechanisms. Please note that the broad range of errors induced by the Nicholson-Shain equation based approach is not surprising, due to the non-constant dimensionless peak current  $\chi_p$  in the test voltammograms, although this method assumes a constant value (Figure 3).

After finding an appropriate kernel (RBF) and covariance function (squared exponential) for the regression algorithms, we analyzed the influence of different preprocessing methods on the estimated diffusion coefficients (Figure 5). For the downsampling method the number of dimensions in each simulated cyclic voltammogram was reduced by a factor of 20, i.e. retaining only every

20th sample, while preprocessing by PCA worked as described above. The manual preprocessing method used the seven features derived from the potentials and currents of the cyclic voltammogram shown in Figure 2, which were chosen as those being most prominent and





**Figure 7** Mean of the absolute error, on a logarithmic scale, for diffusion coefficients determined by SVR with RBF kernel, GPR with squared exponential covariance function and the Nicholson-Shain equation for the  $E_{qr}$  mechanism depending on the rate constant  $k_s$ . Shading around curves represent 95% confidence intervals for the mean. The dotted line indicates the spacing used for the diffusion coefficients in the simulated data; PCA preprocessing was used for predicting coefficients with SVR and GPR.

commonly used for analysis. These manually extracted features include the forward peak, half peak, and reverse peak potentials ( $E_p^{for}$ ,  $E_{p/2}$ ,  $E_p^{rev}$ ), the difference between forward and reverse peak potential  $E_p^{for} - E_p^{rev}$ , the forward peak current  $i_p^{for}$ , and the ratio between forward and reverse peak current  $i_p^{for}/i_p^{rev}$ . Note, that this is *not* the peak current ratio as defined by Nicholson [52].

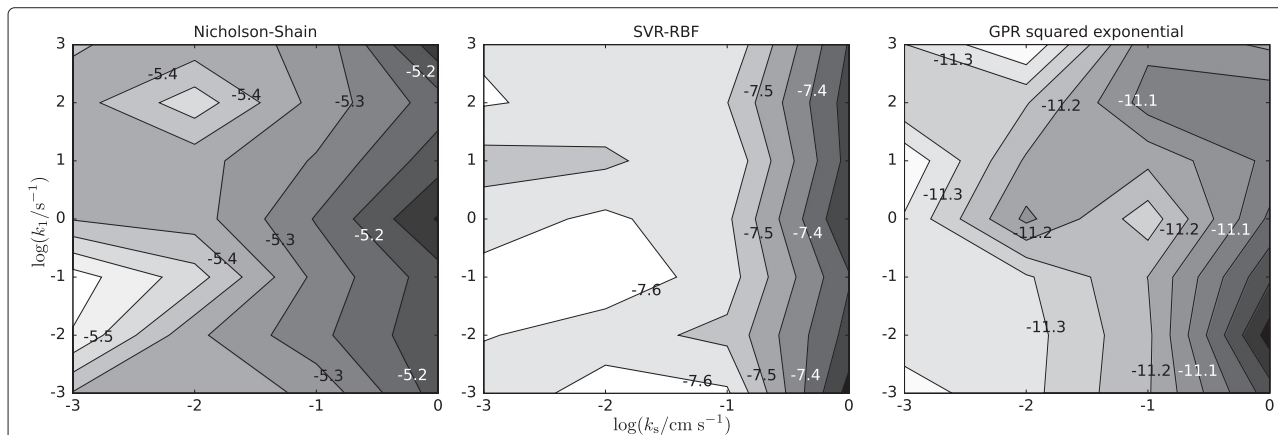
As shown in Figure 5 the manual preprocessing method yields the lowest accuracy of the estimated diffusion coefficients for both regression algorithms and all reaction mechanisms. This indicates that, albeit being helpful for a human observer, the manually extracted features discard

too much of the information contained in the full cyclic voltammogram. The performance differences between the PCA and downsampling method are small, yet PCA works best for the  $E_{qr}$ C mechanism, while there is no difference between the preprocessing methods on the EC and  $E_{qr}$  mechanism in conjunction with the SVR algorithm. For the GPR algorithm PCA is slightly better for the EC mechanism, while downsampling is better for the  $E_{qr}$  mechanism. We used PCA preprocessing for both regression algorithms when estimating diffusion coefficients from real data, as it allows to judge the quality of the data reduction depending on the amount of explained variance.

### EC mechanism — dependence on $k_1$

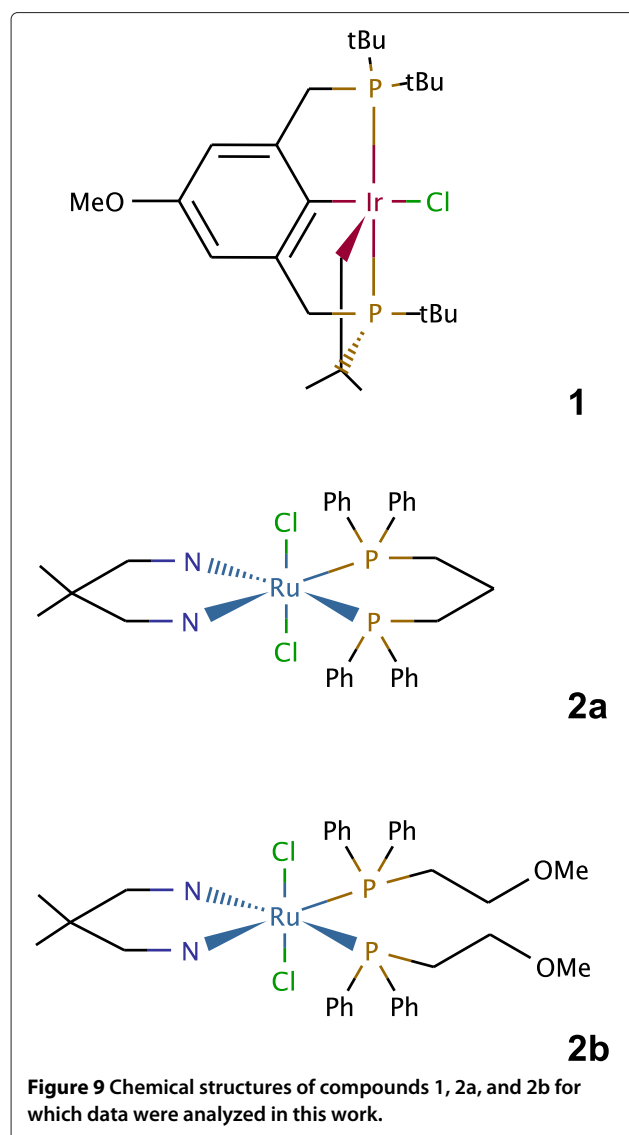
Figure 6 shows the average absolute error between estimated and true diffusion coefficient values depending on the rate constant  $k_1$  for the EC mechanism. The dotted line in Figure 6 marks the spacing used for  $D$  in the simulations and can be considered as the baseline error of a simple table lookup, e.g. if the diffusion coefficient is determined from a table listing values of  $D$  for different rate constants  $k_1$ . Confidence intervals for the average absolute error at the 95% level were computed by a bootstrap method with 1000 bootstrap samples [53]. While the accuracy of the diffusion coefficients estimated by the regression algorithms is virtually independent of the rate constant value, as indicated by the flat error curves, the accuracy of diffusion coefficients calculated with the Nicholson-Shain equation degrades with increasing  $k_1$  and the error increases above the baseline error for  $k_1 > 1 \text{ s}^{-1}$ .

This behaviour of the results from the Nicholson-Shain equation based approach is expected due to the dependence of the dimensionless peak current  $\sqrt{\pi} \chi_p$  on the dimensionless rate constant  $\kappa_1$  described in



**Figure 8** Contour plots showing the dependence the average absolute error on the rate constants  $k_1$  and  $k_s$  ( $E_{qr}$ C mechanism) on a logarithmic scale in  $\text{cm}^2 \text{ s}^{-1}$ . The average absolute error is calculated between estimated and true diffusion coefficients.





Section “Nicholson-Shain equation approach”. The black bars on the abscissa of Figures 3 and 6 mark the region where the dimensionless peak current does not deviate significantly from the constant asymptotic value of 0.4463. It should be noted that the scales on the abscissa in both, Figures 3 and 6, are equivalent apart from a constant offset since, for  $n = 1$ ,  $\log(\kappa_1) = \log(k_1/s^{-1}) - \log(a/s^{-1})$  and  $\log(a/s^{-1}) \approx 0.9$ . The quality of the diffusion coefficients calculated by the Nicholson-Shain equation for rate constants in this range ( $\log(k_1/s^{-1}) \in (-\infty, -1]$ ) is even better than the coefficient values estimated by the SVR algorithm with RBF kernel (Figure 4). Since the exact value of the rate constant is often not known in practice, however, it seems to be better to resort to one of the regression algorithms for finding the diffusion coefficient in general.

#### $E_{qr}$ mechanism — dependence on $k_s$

For the  $E_{qr}$  mechanism the error incurred by the SVR and GPR algorithms is constant for electron transfer rate constant values  $\log(k_s/cm\ s^{-1}) > -2.5$  (Figure 7). Below this value one can observe a slight increase in the average absolute error from  $10^{-8}$  to  $10^{-7.3}$  for SVR and from  $10^{-11}$  to  $10^{-10.5}$  for GPR.

The error of the Nicholson-Shain equation approach, on the other hand, increases from  $10^{-7}$  to  $10^{-5}$  for electron transfer rates  $\log(k_s/cm\ s^{-1})$  in the range  $[-3, -2]$  and thus shows a stronger dependence of diffusion coefficient accuracy on the rate constant. The absolute error approaches the order of magnitude of the values of  $D$ . Overall, the regression algorithms SVR and GPR yield a more accurate estimate of the diffusion coefficient for simulated  $E_{qr}$  voltammograms in comparison to the Nicholson-Shain equation and to table look-up.

#### $E_{qr}C$ mechanism — dependence on $k_1$ and $k_s$

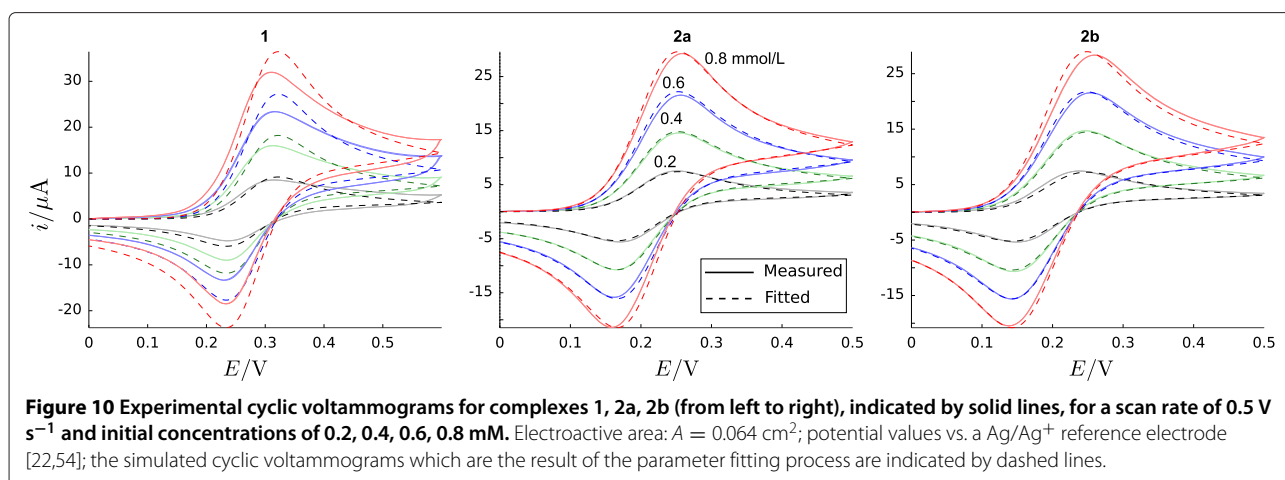
In contrast to the EC and  $E_{qr}$  reaction mechanisms, the  $E_{qr}C$  mechanism is governed by two rate constants  $k_1$  and  $k_s$  (Table 4). For the three tested methods the error surfaces are rather flat and only slightly increase for  $\log(k_s/cm\ s^{-1})$  between -1.5 and 0 (Figure 8). The largest difference between two points on the logarithmic error surface is 0.48 for the Nicholson-Shain equation approach, 0.36 for SVR, and 0.53 for GPR. Notably, the global error level for the  $E_{qr}C$  mechanism is on the same scale as the error level for the  $E_{qr}$  and EC mechanism (Nicholson-Shain:  $[-5.6, -5.1]$ , SVR:  $[-7.6, -7.3]$ , GPR:  $[-11.4, -10.9]$ ), which indicates that the proposed estimation of diffusion coefficients is extensible to more complex reaction mechanisms.

#### Estimations from experimental data

The estimation of diffusion coefficients was applied to three experimental data sets, each containing 80 experimental cyclic voltammograms. The first data set consisted of measurements for iridium complex **1** [22], the second and third of those for ruthenium complexes **2a** and **2b** [54,55] (see Figure 9 and Section “Experimental”). The reaction mechanisms ( $E_{qr}C$  for complex **1**, and  $E_{qr}$  for complexes **2a** and **2b**) were established earlier [22,54].

**Table 5** Parameter values yielding the best fit between simulated and experimental cyclic voltammograms for the three metal complexes

Parameter	<b>1</b>	<b>2a</b>	<b>2b</b>
$E^0$ (V)	0.2767	0.2084	0.1938
$k_s$ (cm s <sup>-1</sup> )	0.0232	0.0199	0.0118
$k_1$ (s <sup>-1</sup> )	0.1473		
$D$ (cm <sup>2</sup> s <sup>-1</sup> )	1.5846e-5	1.0535e-5	1.0824e-5



Since the true value of the diffusion coefficient is unknown for each of the experimental data sets, we fitted simulated cyclic voltammograms to the experimental signals by optimizing the formal potential  $E^0$ , the rate constants  $k_1$ ,  $k_s$ , and the diffusion coefficient  $D$  as described in Section “Fitting of simulation parameters”. The fitted diffusion coefficients serve as a reference point for comparing the values calculated by the regression algorithms and the Nicholson-Shain equation approach. Table 5 lists the parameter values that yield the best fit between simulated and experimental cyclic voltammograms and Figure 10 gives an impression of the fit quality. The best fit was obtained for the  $E_{qr}$  reaction of complex **2a** with an average absolute error between simulated and experimental signals of  $0.75 \mu\text{A}$ , followed by the  $E_{qr}$  reaction of **2b** ( $1.09 \mu\text{A}$ ), and the  $E_{qr}C$  reaction of **1** ( $3.23 \mu\text{A}$ ).

Based on the results with simulated data (Section “Estimation from simulated data”) we used SVR with RBF kernel and GPR with squared exponential covariance function in conjunction with the PCA preprocessing method to estimate diffusion coefficients for the experimental data sets. For complex **1**, the training data consisted of all 2800 simulated cyclic voltammograms created for the  $E_{qr}C$  mechanism (Section “ $E_{qr}C$  mechanism — dependence on  $k_1$  and  $k_s$ ”), while 1200 simulated cyclic voltammograms for the  $E_{qr}$  mechanism served as training

**Table 6 Diffusion coefficients in  $10^{-5} \text{ cm}^2 \text{ s}^{-1}$  determined by different methods for the experimental cyclic voltammograms; bold values: best matches with respect to parameter fitting results**

	<b>1</b>	<b>2a</b>	<b>2b</b>
Parameter-Fit	1.58	1.05	1.08
Nicholson-Shain	1.07	0.84	0.82
SVR	2.32	<b>1.07</b>	<b>1.09</b>
GPR	<b>1.54</b>	1.10	1.13

data for **2a/2b**. In order to have the voltammograms on a comparable scale the current was normalized by multiplying the signal with the factor  $(c_0\sqrt{v})^{-1}$ .

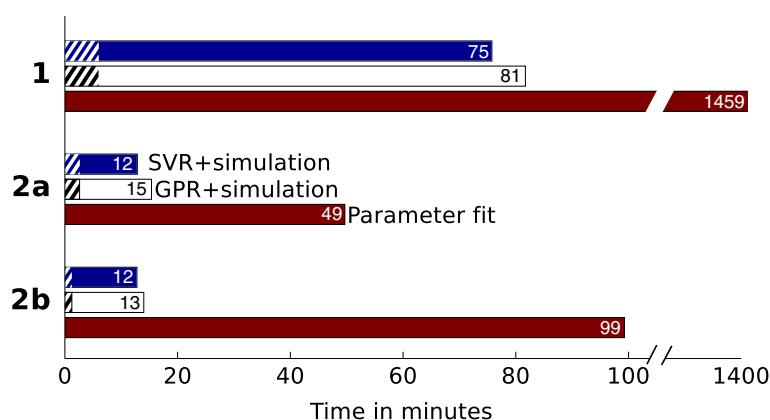
The trained regression algorithms and the approach based on the Nicholson-Shain equation were then used to calculate the diffusion coefficient for each of the 80 experimental voltammetric curves. Since the diffusion coefficient of the electrochemically active species should be constant across measurements with different scan rates and initial concentrations, we averaged the 80 calculated coefficients to arrive at the final estimate. Table 6 lists the diffusion coefficients determined by parameter fitting, the Nicholson-Shain equation approach, and the regression algorithms.

For **1** the diffusion coefficient estimated by GPR is the best match with respect to the fitted coefficient value. Although there is only a small difference in the estimates of SVR and GPR, the best diffusion coefficient estimates for **2a/2b** are provided by SVR. In contrast to the regression algorithms, the Nicholson-Shain equation consistently underestimates the diffusion coefficient value on all data sets.

To further assess the quality of the estimated values we repeated the simulation of cyclic voltammograms with the estimated diffusion coefficients and calculated the discrepancy between simulated and experimental voltammetric signals (Table 7). In comparison to the parameter fitting method the average absolute error increases only

**Table 7 Average absolute error of currents in  $\mu\text{A}$  between simulated and experimental cyclic voltammograms**

	<b>1</b>	<b>2a</b>	<b>2b</b>
Parameter-Fit	3.23	<b>0.75</b>	<b>1.09</b>
Nicholson-Shain	3.74	1.41	1.74
SVR	4.66	0.76	1.10
GPR	<b>3.19</b>	0.803	1.14



**Figure 11** CPU time in minutes required by the parameter fitting method, and the regression algorithms for the three organometallic complexes **1**, **2a**, and **2b**. Hatched bars indicate the portion of time required by the SVR and GPR algorithm without the simulations. All measurements were made on an INTEL® XEON® 5150 processor with 2.66 GHz and 8 GB of main memory.

slightly for the coefficients estimated by SVR for **2a/2b**, and GPR on all organometallic complexes. The diffusion coefficients obtained by the Nicholson-Shain equation for **1**, **2a**, and **2b**, and by the SVR algorithm for **1** are of inferior quality.

The parameter fitting approach usually yields reliable estimates of the diffusion coefficients in practice, but at the expense of long computational times (Figure 11). In contrast, the creation of simulated data followed by regression algorithm training and estimation of diffusion coefficients only takes a small percentage of the parameter fitting time (3–20%). If simulated data is already available, this percentage is further reduced to 0.01–0.06%, which is beneficial if large amounts of experimental data need to be analyzed.

## Experimental

Voltammetric signals in each data set in Section “Estimations from experimental data” were acquired twice for ten scan rates of 0.02, 0.05, 0.1, 0.2, 0.5, 1.003, 2.007, 5.120, 10.240, and 20.480 V s<sup>-1</sup>, and four different initial concentrations  $c_0$  of 0.2, 0.4, 0.6, 0.8 mmol L<sup>-1</sup> in a dichloromethane electrolyte with 0.1 M tetra-*n*-butylammonium hexafluorophosphate as supporting electrolyte at a Pt electrode (for further experimental details, see [22,54]). The scanning potential varied between 0 and 0.6 V for **1**, and between 0 and 0.5 V for **2a/2b** with an increment of 1 mV in each case.

## Conclusion

The results presented in this work show the feasibility of estimating diffusion coefficients from experimental cyclic voltammograms by regression algorithms trained on simulated data. This approach is generic in the sense that it is not restricted to a particular reaction mechanism and

range of rate constants, as demonstrated by the results obtained on simulated data for the EC, E<sub>qr</sub>, and E<sub>qr</sub>C mechanisms. On simulated data the accuracy of diffusion coefficients estimated by SVR with RBF kernel and GPR with squared exponential covariance function is higher as compared to the Nicholson-Shain equation approach over a wide range of rate constants. The best preprocessing method for estimating  $D$  with the regression algorithms turned out to be the principal component projection of the cyclic voltammograms. Projecting the data to the subspace spanned by the first five principal components apparently retains important shape information that is discarded by the manual extraction of prominent peak features. This indicates that the commonly used evaluation of the limited set of human recognizable features related to voltammetric peaks might not be optimal for data evaluation in all cases. For the three experimental data sets, estimation with GPR yielded diffusion coefficients that closely matched the values determined by the classical parameter fitting approach, whereas SVR showed comparable performance only for **2a/2b**. These results indicate that GPR with a squared exponential covariance function is better suited than SVR to reliably determine diffusion coefficients from experimental data. Furthermore the GPR based determination of the diffusion coefficient requires less computational time in contrast to the parameter fitting approach.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

The concept of this research was conceived by MB and BS. The derivation of equations and computational work was performed by DB in the context of his doctoral thesis. MB, WR, and BS supervised the project from the informatics (MB, WR) and electrochemical (BS) point of view. The sequence of authors is determined alphabetically by last name. All authors read and approved the final manuscript.

## Acknowledgements

We thank the Deutsche Forschungsgemeinschaft, Bonn-Bad Godesberg, Germany, for financial support of this work within the Graduiertenkolleg 441 "Chemie in Interphasen". We are grateful to Filip Novak, Institut für Organische Chemie, Universität Tübingen, for preparing and providing the experimental data sets. We acknowledge support by the Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of the Universität Tübingen to cover the processing charges of the article. This paper is part 3 of the series "Chemical Information from Electrochemical Data"; for part 2, see [15].

## Author details

<sup>1</sup>Technische Informatik, Universität Tübingen, Sand 13, D-72076 Tübingen, Germany. <sup>2</sup>Institut für Organische Chemie, Universität Tübingen, Auf der Morgenstelle 18, D-72076 Tübingen, Germany. <sup>3</sup>Present address: Technische Informatik, Universität Leipzig, Augustusplatz 10, D-04109 Leipzig, Germany.

Received: 5 December 2013 Accepted: 24 April 2014  
Published: 28 May 2014

## References

- Nicholson RS, Shain I: **Theory of stationary electrode polarography. Single scan and cyclic methods applied to reversible, irreversible, and kinetic systems.** 1964, **36**:706–723.
- Heinze J: **Cyclovoltammetrie — die "Spektroskopie" des Elektrochemikers.** *Angew Chem* 1984, **96**:823–840. **Cyclic Voltammetry — the "Spectroscopy" of the Electrochemist.** *Angew Chem Int Ed Engl* 1984, **23**:831–847.
- Economou A, Bolis SD, Efsthathiou CE, Volikakis GJ: **A "virtual" electroanalytical instrument for square wave voltammetry.** *Anal Chim Acta* 2002, **467**:179–188.
- Speiser B: **Methods to investigate mechanisms of electroorganic reactions.** In *Encyclopedia of Electrochemistry, Vol. 8 Organic Electrochemistry*, Chap. 1. Edited by Bard AJ, Stratmann M, Schäfer HJ. Weinheim: Wiley-VCH; 2004:1–23.
- Venton BJ, Wightman RM: **Psychoanalytical electrochemistry: Dopamine and behavior.** *Anal Chem* 2003, **75**:414A–421A.
- Cogan SF: **Neural stimulation and recording electrodes.** *Annu Rev Biomed Eng* 2008, **10**:275–309.
- Bessant C, Saini S: **An object oriented approach to electroanalytical technique implementation.** *Electroanalysis* 1997, **9**:926–931.
- Bieniasz LK: **Towards computational electrochemistry — a kineticist's perspective.** In *Mod. Asp. Electrochem., Vol. 35*. Edited by Conway BE, White RE. New York: Kluwer Academic/Plenum Publishers; 2002:135–195.
- Britz D: **Digital simulation in electroanalytical chemistry.** In *Encyclopedia of Electrochemistry, Vol. 3 Instrumentation and Electroanalytical Chemistry*, Chap. 1.3. Edited by Bard AJ, Stratmann M, Unwin P. Weinheim: Wiley-VCH; 2003:51–71.
- Speiser B: **Numerical simulations in electrochemistry.** In *Encyclopedia of Applied Electrochemistry*. Edited by Savinell RF, Ota K, Kreysa G. New York: Springer; 2014. doi:10.1007/SpringerReference\_303485.
- Speiser B: **Multiparameter estimation: Extraction of information from cyclic voltammograms.** *Anal Chem* 1985, **57**:1390–1397.
- Bieniasz LK, Speiser B: **Use of sensitivity analysis methods in the modelling of electrochemical transients. Part 3. Statistical error/uncertainty propagation in simulation and in nonlinear least-squares parameter estimation.** *J Electroanal Chem* 1998, **458**:209–229.
- Richards E, Bessant C, Saini S: **Multivariate data analysis in electroanalytical chemistry.** *Electroanalysis* 2002, **14**:1533–1542.
- Lu X, Mo J, Yang C, Kang J, Gao J: **Orthogonal wavelets analysis of electroanalytical signals.** *Anal Lett* 2000, **33**:1167–1186.
- Speiser B: **Chemical information from electroanalytical data. Part 2. Determination of the rate and equilibrium constants of a chemical reaction preceding a reversible electron transfer from cyclic voltammetric data.** *J Electroanal Chem* 1991, **301**:15–35.
- Scharbert B, Speiser B: **Chemical information from electroanalytical data. Part 1 — Determination of system parameters for quasi-reversible electron transfer reactions from cyclic voltammetric test data and data for the reduction of cerium(IV) bis(octaethylporphyrinate).** *J Chemometrics* 1989, **3**:61–80.
- Sapozhnikova EP, Bogdan M, Speiser B, Rosenstiel W: **EChem++ — An object-oriented problem solving environment for electrochemistry. Part 3. Classification of Voltammetric signals by the fuzzy ARTMAP neural network with respect to reaction mechanism.** *J Electroanal Chem* 2006, **588**:15–26.
- Bieniasz LK, Rabitz H: **Extraction of parameters and their error distributions from cyclic voltammograms using bootstrap resampling enhanced by solution maps: computational study.** *Anal Chem* 2006, **78**:8430–8437.
- Hasanov A: **Identification of unknown diffusion coefficient in pure diffusive linear model of chronoamperometry. I. The theory.** *J Math Chem* 2010, **48**:491–507.
- Hasanov A: **Identification of unknown diffusion and convection coefficients in ion transport problems from flux data: An analytical approach.** *J Math Chem* 2010, **48**:413–423.
- Oldham KB, Myland JC: **Extracting parameter values from quasireversible cyclic voltammograms.** *J Solid State Electrochem* 2012, **16**:3691–3693.
- Novak F, Speiser B, Mohammad HAY, Mayer HA: **Electrochemistry of transition metal complex catalysts. Part 10. Intra- and intermolecular electrochemically activated C–H addition to the central metal atom of a P–C–P-pincer iridium complex.** *Electrochim Acta* 2004, **49**:3841–3853.
- Rudolph M, Reddy DP, Feldberg SW: **A simulator for cyclic voltammetric responses.** *Anal Chem* 1994, **66**:589A–600A.
- Bott AW, Feldberg SW, Rudolph M: **Fitting experimental cyclic voltammetry data with theoretical simulations using digisim 2.1.** *Curr Sep* 1996, **15**:67–71.
- Moon K, Kaifer AE: **Dimeric molecular capsules under redox control.** *J Am Chem Soc* 2004, **126**:15016–15017.
- Janisch J, Ruff A, Speiser B, Wolff C, Ziggelli J, Benthin S, Feldmann V, Mayer HA: **Consistent diffusion coefficients of ferrocene in some non-aqueous solvents: Electrochemical simultaneous determination together with electrode sizes and comparison to pulse-gradient spin-echo NMR results.** *J Solid State Electrochem* 2011, **15**:2083–2094.
- Schölkopf B, Smola AJ: *Learning with Kernels.* Cambridge: MIT Press; 2002.
- Rasmussen CE, Williams CKI: *Gaussian Processes for Machine Learning.* Cambridge: The MIT Press; 2006.
- Sonnenburg S, Zien A, Philips P, Rätsch G: **POIMs: positional oligomer importance matrices—understanding support vector machine-based signal detectors.** *Bioinformatics* 2008, **24**(13):i6–i14.
- Brugger D, Butovas S, Bogdan M, Schwarz C: **Real-time adaptive microstimulation increases reliability of electrically evoked cortical potentials.** *IEEE Trans Bio-med Eng* 2011, **58**:1483–1491.
- Brugger D, Butovas S, Bogdan M, Schwarz C, Rosenstiel W: **Direct and inverse solution for a stimulus adaptation problem using SVR.** In *ESANN Proceedings*. Bruges; 2008:397–402.
- Balabin RM, Lomakina EI: **Support vector machine regression (LS-SVM)—an alternative to artificial neural networks (ANNs) for the analysis of quantum chemistry data?** *Phys Chem Chem Phys* 2011, **13**:11710–11718.
- Balabin RM, Lomakina EI: **Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data.** *Analyst* 2011, **136**:1703–1712.
- Brereton RG, Lloyd GR: **Support vector machines for classification and regression.** *Analyst* 2010, **135**:230–267.
- Devos O, Ruckebusch C, Durand A, Duponchel L, Huvenne J-P: **Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation.** *Chem Int Lab Syst* 2009, **96**:27–33.
- Niazi A, Sharifi S, Amjadi E: **Least-squares support vector machines for simultaneous voltammetric determination of lead and tin: a comparison between LS-SVM and PLS in voltammetric data.** *J Electroanal Chem* 2008, **623**:86–92.
- Chu W, Ghahramani Z, Falciani F, Wild DL: **Biomarker discovery in microarray gene expression data with gaussian processes.** *Bioinformatics* 2005, **21**:3385–3393.
- Shpigelman L, Singer Y, Paz R, Vaadia E: **Spikernels: predicting arm movements by embedding population spike rate patterns in inner-product spaces.** *Neural Comput* 2005, **17**:671–690.

39. Saunders C, Gammerman A, Vovk V: **Ridge regression learning algorithm in dual variables**. In *Proceedings of the 15th International Conference on Machine Learning*. San Francisco; 1998:515–521.
40. Keerthi SS, DeCoste D: **A modified finite newton method for fast solution of large scale linear SVMs**. *JMLR* 2005, **6**:341–361.
41. Chapelle O: **Training a support vector machine in the primal**. *Neural Comput* 2007, **19**:1155–1178.
42. Bo L, Wang L, Jiao L: **Recursive finite newton algorithm for support vector regression in the primal**. *Neural Comput* 2007, **19**:1082–1096.
43. Kimeldorf GS, Wahba G: **A correspondence between bayesian estimation on stochastic processes and smoothing by splines**. *Ann Math Stat* 1970, **41**:495–502.
44. Aronszajn N: **Theory of reproducing kernels**. *Trans Amer Math Soc* 1950, **68**:337–404.
45. Chang M-W, Lin C-J: **Leave-one-out bounds for support vector regression model selection**. *Neural Comput* 2005, **17**:1188–1222.
46. Bertsekas DP: *Nonlinear Programming, 2nd edn*. Belmont: Athena Scientific; 2003.
47. Golub GH, van Loan CF: *Matrix Computations, 3rd edn*. Baltimore: The John Hopkins University Press; 1996.
48. Speiser B: **EASIEST — A program system for electroanalytical simulation and parameter estimation — I. Simulation of cyclic voltammetric and chronoamperometric experiments**. *Comput Chem* 1990, **14**:127–140.
49. Hertl P, Speiser B: **Electroanalytical investigations. Part VI. The simulation of fast chemical equilibrium reactions in cyclic voltammetric reaction-diffusion models with spline collocation**. *J Electroanal Chem* 1987, **217**:225–238.
50. Waltz RA, Morales JL, Nocedal J, Orban D: **An interior algorithm for nonlinear optimization that combines line search and trust region steps**. *Math Program A* 2006, **107**:391–408.
51. Waltz RA, Plantenga TD: *Knitro User's Manual, Version 6.0*. Ziena Optimization, Inc.; (2004–2009). [http://www.ziena.com/docs/Knitro60\_UserManual.pdf]
52. Nicholson RS: **Semiempirical procedure for measuring with stationary electrode polarography rates of chemical reactions involving the product of electron transfer**. *Anal Chem* 1966, **38**:1406.
53. DiCiccio TJ, Efron B: **Bootstrap confidence intervals**. *Stat Sci* 1996, **11**:189–228.
54. Novak F, Speiser B, Lindner E, Lu Z-L, Mayer HA: **Elektrochemie von Übergangsmetallkatalysatoren. Teil 11. Eine chemisch modifizierte Platinelektrode als zweizähliger Diaminligand zur Bildung wohldefinierter immobilisierter Bis( $\eta^1$ -P-ether-phosphan)(diamin)ruthenium(II)-Komplexe**. *Angew Chem* 2004, **116**:2059–2062.  
**Electrochemistry of Transition Metal Complex Catalysts. Part 11. A Chemically Modified Platinum Electrode as a Bidentate Diamine Ligand for Forming Well-Defined, Immobilized Bis( $\eta^1$ -P-ether-phosphane)(diamine)ruthenium(II) Complexes**. *Angew Chem Int Ed* 2004, **43**:2025–2028.
55. Lindner E, Al-Gharabli S, Mayer HA: **Supported organometallic complexes. Part 31: Diaminediphosphineruthenium(II) precursor complexes for parallel synthesis in interphases**. *Inorg Chim Acta* 2002, **334**:113–121.

doi:10.1186/1758-2946-6-30

**Cite this article as:** Bogdan et al.: Estimation of diffusion coefficients from voltammetric signals by support vector and gaussian process regression. *Journal of Cheminformatics* 2014 **6**:30.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

“Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge.”

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
http://www.chemistrycentral.com/manuscript/



**ChemistryCentral**