Multi-phase classification by a least-squares support vector machine approach in tomography images of geological samples

Faisal Khan, Frieder Enzmann, and Michael Kersten

Geoscience Institute, Johannes Gutenberg University, Mainz 55099, Germany

Correspondence to: Frieder Enzmann (enzmann@uni-mainz.de)

Abstract. Image processing of X-ray-computed polychromatic cone-beam micro-tomography (μXCT) data of geological samples mainly involves artefact reduction and phase segmentation. For the former, the main beam-hardening (BH) artefact is removed by applying a best-fit quadratic surface algorithm to a given image data set (reconstructed slice), which minimizes the BH offsets of the attenuation data points from that surface. A Matlab code for this approach is provided in the Appendix. The final BH-corrected image is extracted from the residual data or from the difference between the surface elevation values and the original grey-scale values. For the segmentation, we propose a novel least-squares support vector machine (LS-SVM, an algorithm for pixel-based multi-phase classification) approach. A receiver operating characteristic (ROC) analysis was performed on BH-corrected and uncorrected samples to show that BH correction is in fact an important prerequisite for accurate multi-phase classification. The combination of the two approaches was thus used to classify successfully three different more or less complex multi-phase rock core samples.

1 Introduction

Advances in the technological (image resolution) and computational (image size) aspects of X-ray computed micro-tomography (μXCT) technology now enable the acquisition of three-dimensional (3-D) images down to a submicron spatial resolution, which is sufficient to capture the microstructure of geological rock cores (Cnudde and Boone, 2013). Recent research on digital rock physics has successfully combined microscopic imaging with advanced numerical simulations of physical properties for which laboratory measurements are not possible. However, benchmarking tests of commonly used image processing methods have revealed unacceptably large variations in the results, and further development and optimization is therefore clearly warranted (e.g. Andrä et al., 2013). Furthermore, Leu et al. (2014) determined the importance of image analysis of μXCT-generated data to provide accurate parameterization of porosity–permeability relationships. It turned out that this relationship is highly sensitive towards accurate processing of the pore microstructure images used. This accuracy is of great importance in pedohydrology studies of water flux in the critical zone (e.g. Khan et al., 2012; Kumahor et al., 2015), reactive transport modelling (e.g. Schwarz and Enzmann, 2013; Molins et al., 2014), distributions of multi-component fluids (e.g. Berg et al., 2013, 2015), contaminant retardation (e.g. Huber et al., 2012), energy-related activities such as nuclear waste disposal (e.g. Hemes et al., 2015), and the geological sequestration of CO_2 (e.g. Sell et al., 2013; Herring et al., 2015), to name just a few examples.

The research aim of image classification optimization is to obtain representations of structures that can be automatically used for categorization of samples into a finite set of labels (i.e. phases in geological materials). As part of the recent development of computer performance and advanced automated computer algorithms, the classical machine learning technique provides a methodology for (non-)linear function estimation and classification problems (Vapnik, 1995). In general, the supervised machine learning approach involves the construction of a convincing model of the distribution of class labels in terms of predictor features for the whole image on the basis of a reduced example (i.e. training) data set (Alpaydin, 2004; Kotsiantis, 2007). The resulting classifier is then used to optimize a performance criterion and

assign class labels to the testing data. Representative generalization is an important property of a classifier or classification algorithm, because it offers information about as yet unknown data. The support vector machine (SVM) algorithm as one capable example of such classifiers was first developed by Vapnik (1995) as an extension of the generalized portrait algorithm. The SVM algorithm is firmly grounded in the framework of statistical learning theory, which improves the generalization ability of learning machines for unknown data.

The success with any classification technique depends on the quality of the μXCT images used as input; therefore, the utmost care has to be taken to provide less noisy and artefact-free images. In reality, this is seldom the case when using conventional bench-top μXCT. In such polychromatic cone-beam X-ray tomography, the average energy of the X-ray beam increases and the beam spectrum is successively depleted at lower energies as it passes through the geomaterial. The spectrum of the X-ray beam thus hardens and becomes less attenuated by the same phase but further into the material. This effect is called beam hardening (BH) and implies that the grey levels of the projection data are non-linear with respect to the shape of a cross-sectional linear profile. Consequently, the reconstructed μXCT image features some visual distortions such as pronounced edges ("cupping effect") and other artefacts (Schlüter et al., 2014). This cupping artefact shows an apparently higher attenuation coefficient at the outer regions than in the inner part of a round rock core, even if both regions are composed of the same mineral. The presence of BH artefacts therefore causes problems in 3-D image processing and hampers correct image analysis and phase quantification due to a biased multi-phase image segmentation process. Such artefacts apply for polychromatic X-ray sources only; more recently, monochromatic synchrotron-based μXCT has been introduced as a powerful tool for effective BH-free visualization of the microstructural features of geomaterials at a voxel resolution down to the submicron level (Fusseis et al., 2014; Leu et al., 2014). However, access to this advanced technique is limited by experimental sites and available beam time. Therefore, most μXCT studies must still rely on bench-top devices resulting in the artefacts mentioned.

A variety of both hard- and software measures have been developed to eliminate BH artefacts. These include physical pre-filtering to pre-harden the X-ray photon spectrum, the dual-energy approach, and a variety of computational pre- and post-processing image corrections (Schlüter et al., 2014). A major prerequisite for success with the latter software approaches is that the correction method should not rely on any prior knowledge of the material properties; i.e. it should not depend on known attenuation coefficients. Commonly, there is no prior quantitative information available on the number and distribution of phases present in a geological sample. The most common technique for BH correction in pre-processing is linearization, but this is preferable for monophasic mate-

rial cases. Although the most commonly used algorithm for the reconstruction of μXCT data is based on filtered back projection (Feldkamp et al., 1984), an iterative forward projection can be used with modern imaging software like Octopus (https://octopusimaging.eu/), which allows in its latest version 8.9.0 – 64 bit the incorporation of BH modelling algorithms (Brabant et al., 2012). In the latter reconstruction approach, the attenuation coefficients are not simply added but multiplied with factors to simulate BH depending on the accumulated attenuation over the distance the beam has penetrated through the sample. This pre-processing correction thereby minimizes the underestimation of the attenuation coefficient as the beam progresses through it. However, a major drawback of this promising method is that there is currently no way to determine the two necessary iterative parameters α and β automatically (Brabant et al., 2012), resulting in the manually adjusted output being a time-consuming and subjective iteration result.

In principle, the idea of a surface-fitting approach in μXCT image post-processing for BH correction has already been introduced previously (Krumm et al., 2008; Iassonov and Tuller, 2010; Jovanović et al., 2013), albeit without a more detailed outline of how to be realized in practice. Therefore, a quite simple algorithm is suggested that fits a 2-D quadratic polynomial function for accurate removal of BH artefacts upon classically filtered back-projection-reconstructed slices. Our novel BH correction algorithm is followed by a pixel-based phase classification introducing the machine learning algorithm approach. A workflow of our two-stage post-reconstruction image processing approach including BH correction and the novel segmentation approach is presented and validated by three different geological samples.

2　Material and methods

2.1　Samples and tomographic set-up

The custom-built μXCT scanner used at our laboratory (ProCon CT-Alpha, Germany) is equipped with a micro-focus X-ray tube (Feinfocus, Germany) and contains a diamond-coated anode target with a focal spot size of a few micrometres. X-ray data acquisition is performed with a 2048×2048 pixel ("2 k") flat-panel CCD detector of size $105\,\text{mm} \times 105\,\text{mm}$ (Hamamatsu, Japan). The first test sample (sample A) is comprised of sand grains filled into a tube made of Plexi-glass material with 50 mm height and 10 mm inner diameter. The X-ray source voltage was set to 100 kV, and the beam was slightly pre-hardened with 0.15 mm aluminum foil. A rotation step of $0.45°$ with 2 s exposure time corresponded to 800 projections for full $360°$ data acquisition at a spatial resolution of $5.89\,\mu\text{m}$. The other two geological samples B and C were cylindrical rock cores composed of (B) a clay mineral matrix with a network of pores/cracks and

anhydrite veins, and (C) a clay mineral matrix with halite-sealed veins. Detailed mineralogical analysis of the samples was published previously (Enzmann et al., 2009). These two rock cores were measured at a slightly higher 130 kV source voltage with a 0.15 mm silver foil. Exposure time was at a lower 1 s each, but with the same rotation step of 0.45° as for the sample A, which yields 800 projections at a lower resolution of 42 μm per pixel. Precise centro-symmetrical alignment of the rock cores along the vertical axis is an important prerequisite for success with the BH correction procedure.

The reconstruction of the 3-D data set was performed on the fly by the Feldkamp filtered back-projection algorithm (Feldkamp et al., 1984). This classical 3-D cone-beam reconstruction algorithm follows three main steps: (i) pre-weighting the projection rays according to their position within the beam cone; (ii) filtering the projections along horizontal detector lines using a discrete filtering kernel; and (iii) performing a weighted back projection of the filtered projections along the cone with a weighting factor. Raw projections were corrected for dark current and flat-field variations, followed by non-local mean filtering, ring removal, and filtered back-projection reconstruction using the imaging software package Octopus (https://octopusimaging.eu/; Vlassenbroeck et al., 2007).

The μXCT images are rarely perfect representations of the attenuation coefficients, because they are also biased by scatter and noise. Therefore, image denoising was additionally performed as an initial correction operation. A suitable smoothing filter should reduce the noise level with minimal alteration of edged features in the image. We applied a 3-D median filter technique with window size mask ($3 \times 3 \times 3$ in 3-D), which replaces a singular pixel value with the median value considering the nearest-neighbourhood pixels. The median filter acted to smooth noisy regions and to improve the preservation of their boundary structures (Gallagher et al., 1981), and it is routinely implemented on μXCT images (Culligan et al., 2004; Kaestner et al., 2008; Khan et al., 2012; Sell et al., 2013; Landry et al., 2014; Herring et al., 2015). After reconstruction of the raw data from all three samples, 2-D digital images of dimensions $x, y = 1327 \times 1212$ pixels of sample A, 1600×1600 pixels of sample B, and 1417×1417 pixels of sample C were chosen to evaluate our image correction and novel classification scheme.

2.2 Mathematical basis of the surface-fitting algorithm

Our post-reconstruction method corrects the BH artefact by fitting a 2-D polynomial, i.e. a quadratic surface to the reconstructed μXCT image data (2-D slice). The surface-fitting (i.e. second-order polynomial) approach has a mathematical expression of the form

$$P(x_k, y_k) = a_1 + a_2 x + a_3 y + a_4 x^2 + a_5 xy + a_6 y^2,$$
$$k = 1, 2, \ldots, N \tag{1}$$

for some choice of unknown coefficients $a_1, a_2, \ldots, a_6$. The solution for all a_n coefficients determines the best fit of the polynomial of Eq. (1) to a given set of data points (reconstructed grey-scale values). The final BH-corrected image is the residual of the data points, i.e. the difference between the surface elevation values and the original image values. Consider $f_k \in (x_k, y_k), k = 1, 2, \ldots N$ as arbitrary data points on the 2-D slice (μXCT image); then the normal equations for fitting a polynomial (Eq. 1) can be expressed in a matrix–vector form:

$$\mathbf{M} = \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & x_1 y_1 & y_1^2 \\ 1 & x_2 & y_2 & x_2^2 & x_2 y_2 & y_2^2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_N & y_N & x_N^2 & x_N y_N & y_N^2 \end{bmatrix},$$

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix},$$

$$\boldsymbol{f} = \begin{bmatrix} f_1 \\ f_2 \\ . \\ . \\ . \\ f_N \end{bmatrix}. \tag{2}$$

Equation (2) can be solved to yield the solution vector $\boldsymbol{a}$ by

$$\mathbf{M}^{\mathrm{T}}\mathbf{M}\boldsymbol{a} = \mathbf{M}^{\mathrm{T}}\boldsymbol{f}. \tag{3}$$

The solution of Eq. (3) for the vector $\boldsymbol{a}$ determines the best fit of the polynomial of Eq. (1) to a given set of data points. The Matlab code of this surface-fitting approach for BH correction is listed in the Appendix.

2.3 Mathematical basis of the LS-SVM classification algorithm

Once BH correction of a μXCT image by surface fitting has been accomplished, a pixel-based multi-classification can efficiently be performed by utilizing supervised machine learning of the SVM type. The basic mathematical formulation of the SVM classifier and its non-linear or least-squares versions is presented here only in brief. For further details, please refer to the classical literature and textbooks (Vapnik, 1995; Chapelle et al., 1999; Suykens et al., 2002). In general, the non-linear support vector machine (NL-SVM) method

maps the input vector into the high-dimensional feature space by non-linear mapping associated with a kernel function (often called "kernel trick"; refer to Eqs. 14 and 15). The aim is to construct an optimal separating hyperplane, also known as maximum-margin hyperplane in the higher-dimensional feature spaces. For classification problems, let $\{y_i, x_i\}_{i=1}^{N}$ be a given training set of N data points (here: each points are accounted as image pixel values), where $x_i \in \Re^n$ is the ith input in n-dimensional vector space, and $y_i \in \Re$ is the associated output class of labels such that $y_i \in \{-1, +1\}$. A graphical scheme is depicted in Fig. 1 for a better understanding of the mathematical background of the NL-SVM implementation. The input data points of two classes $\{+1, -1\}$ are non-linearly separable in a feature x space ($x1$ and $x2$ in Fig. 1a). In this scheme, every data point (blue or green colour dot point in Fig. 1a) has a location in the space and represents a vector in that space with the assigned class. In other words, if the features $x1$ and $x2$ represent two dimensions, then each data point represents a two-dimensional vector space. In Fig. 1b, the data points (vectors) are transformed from the original input x space into a higher-dimensional z space ($z1$, $z2$). This means that by the induction of non-linear kernel function the data points, which were not linearly separable in original x space, are now separated by a linear hyperplane defined by a normal vector $\boldsymbol{w}$, with the help of a kernel function using the kernel trick. In Fig. 1b, the safety space showed by dotted lines between two classes is a margin, and the points on the margin are called the support vectors.

The idea of the maximum-margin hyperplane is obtained from statistical learning theory and provides a probabilistic test error bound that is minimized when the margin is maximized (see graphical representation of NL-SVM, Fig. 1). The parameters of the maximum-margin hyperplane are derived by solving a quadratic programming (QP) optimization problem. Suykens and Vandewalle (1999) proposed the idea of modifying Vapnik's SVM formulation by adding a least-squares term to the cost function, which transformed the problem from solving a QP problem to the practically more convenient solving of a set of linear equations. This modification significantly reduces the effort in complexity and thus the computational cost, which may otherwise become excessive. Consider the feature (kernel) function $\phi(x_i): \Re^n \rightarrow \Re^{nb}$ representing a non-linear mapping to a high-dimensional feature space which is formulated as

$$\boldsymbol{w}^{\mathrm{T}}\varphi(x_i) + b \geq 1, \quad \text{if } y_i = +1, \tag{4}$$

and

$$\boldsymbol{w}^{\mathrm{T}}\varphi(x_i) + b \geq -1, \quad \text{if } y_i = -1, \tag{5}$$

which is equivalent to

$$y_i\left[\boldsymbol{w}^{\mathrm{T}}\varphi(x_i) + b\right] \geq 1, \quad i = 1, \dots, N, \tag{6}$$

where $\boldsymbol{w} \in \Re^n$ is an adjustable weight vector parameter, and $b \in \Re$ is a bias term. The slack variable $\xi_i \geq 0$ is introduced in the case of the violation of Eq. (6).

$$y_i\left[\boldsymbol{w}^{\mathrm{T}}\varphi(x_i) + b\right] \geq 1 - \xi_i, \quad i = 1, \dots, N. \tag{7}$$

In real data classification problems, a perfect linear separation is impossible due to overlapping classes. Therefore, a limited number of misclassifications should be tolerated around the margin. The simplified least-squares support vector machine (LS-SVM) algorithm is thus derived from the general non-linear support vector machine approach (Suykens and Vandewalle, 1999). In LS-SVM for function estimation, the following optimization problem is formulated to minimize the empirical risk of misclassification in Eq. (7):

$$\min_{\boldsymbol{w}, b, e} \mathrm{J}_p(\boldsymbol{w}, e) = \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + \gamma\frac{1}{2}\sum_{i=1}^{N}e_i^2, \tag{8}$$

subject to the equality constraints

$$y_i\left[\boldsymbol{w}^{\mathrm{T}}\varphi(x_i) + b\right] = -1 + e_i, \quad i = 1, \dots, N, \tag{9}$$

where $e_i = ([e_1, e_2, \dots e_N]^{\mathrm{T}})$ represents the estimation error for some misclassification tolerance in the case of overlapping distributions and γ is a positive regularization constant in the cost function defining the trade-off between a large margin and misclassification error. In the case of the primal problem expressed in terms of the feature map, the parameter $\boldsymbol{w}$ may have a range over an "infinite-dimensional" parameter set. Therefore, the dual problem for the LS-SVM represents a solution in terms of the kernel function by means of Lagrange multipliers $\alpha_i = \gamma e_i$, which can be positive or negative due to the equality constraints. This means that no sparseness property remains in the LS-SVM formulation, and every training data value is treated as a support vector. The Lagrangian

$$\ell(\boldsymbol{w}, b, e; \alpha) = J_p(\boldsymbol{w}, b, e)$$
$$- \sum_{i=1}^{N}\alpha_i\{y_i\left[\boldsymbol{w}^{\mathrm{T}}\varphi(x_i) + b\right] - 1 + e_i\} \tag{10}$$

is given by the following conditions for optimality:

$$\begin{cases} \frac{\partial\ell}{\partial\boldsymbol{w}} = 0 & \rightarrow \quad \boldsymbol{w} = \sum_{i=1}^{N}\alpha_i y_i\varphi(x_i), \\ \frac{\partial\ell}{\partial b} = 0 & \rightarrow \quad \sum_{i=1}^{N}\alpha_i y_i = 0, \\ \frac{\partial\ell}{\partial e_i} = 0 & \rightarrow \quad \alpha_i = \gamma e_i, \\ \frac{\partial\ell}{\partial\alpha_i} = 0 & \rightarrow \quad y_i\left[\boldsymbol{w}^{\mathrm{T}}\varphi(x_i) + b\right] - 1 + e_i = 0, \\ i = 1, \dots, N. \end{cases} \tag{11}$$

These can be written as a linear system:

$$\begin{bmatrix} \boldsymbol{I} & 0 & 0 & -\boldsymbol{Z}^{\mathrm{T}} \\ 0 & 0 & 0 & -\boldsymbol{Y}^{\mathrm{T}} \\ 0 & 0 & \gamma\boldsymbol{I} & -\boldsymbol{I} \\ \boldsymbol{Z} & \boldsymbol{Y} & \boldsymbol{I} & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{w} \\ b \\ e \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vec{\boldsymbol{1}} \end{bmatrix}, \tag{12}$$
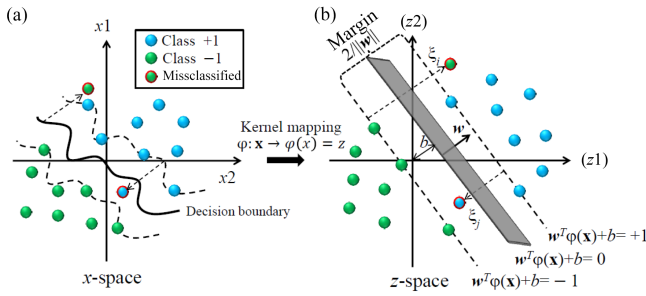
Figure 1. Graphical presentation of the support vector machine classifier with a non-linear kernel, **(a)** complex binary pattern classification problem in input space, and **(b)** non-linear mapping into high-dimensional feature space where a linearly separable data classification takes place.

where $\mathbf{Z} = [\varphi(x_1)^{\mathrm{T}} y_1 \ldots \varphi(x_N)^{\mathrm{T}} y_N]^{\mathrm{T}}$, $\mathbf{Y} = [y_1 \ldots y_N]^{\mathrm{T}}$, $\vec{\mathbf{1}} = [1 \ldots 1]^{\mathrm{T}}$, $\mathbf{e} = [e_1 \ldots e_N]^{\mathrm{T}}$, and $\boldsymbol{\alpha} = [\alpha_1 \ldots \alpha_N]^{\mathrm{T}}$. Elimination of $\mathbf{w}$ and $\mathbf{e}$ gives

$$\begin{bmatrix} 0 - Y^{\mathrm{T}} \\ Y\Psi + \gamma^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix}. \tag{13}$$

The kernel trick is applied to the matrix $\Psi = \mathbf{Z}^{\mathrm{T}}\mathbf{Z}$.
  Hence,

$$\Psi_{il} = y_i y_l \varphi(x_i)^{\mathrm{T}} \varphi(x_l) = H(x_i, x_l), \tag{14}$$

$$\Psi_{il} = y_i y_l H(x_i, x_l), \quad i, l = 1, \ldots, N. \tag{15}$$

According to Mercer's condition (Aronszajn, 1950), any positive-definite kernel function can be expressed as the inner product of two vectors in some feature space and can be used in LS-SVM. This relation is also often termed kernel trick since no explicit construction of the mapping $\phi(x_i)$ is needed. It enables the LS-SVM to work in a high-dimensional feature space, without actually performing calculation in this space. Computations are done in another space after applying this kernel trick. In short, one starts from a formulation in the primal weight space with a high-dimensional feature space by applying transformations $\phi(\cdot)$. The solution is calculated not in this primal weight space, but in the dual space of Lagrange multipliers after applying the kernel trick. In this way, classification is done implicitly in a higher-dimensional feature space rather than in the original input space. Hence, the non-linear LS-SVM classifier in dual space ultimately takes the form

$$y(x) = \mathrm{sign}\left[\sum_{i=1}^{N} \alpha_i y_i H(x, x_i) + b\right], \tag{16}$$

where the sum is taken over the non-zero α_i values which correspond to N numbers of the training data set.
  In our model approach, only the Gaussian radial basis function (RBF) kernel is implemented in the LS-SVM classifier due to its high accuracy in function estimation and

data set classification (Van Gestel et al., 2002; Selvaraj et al., 2007; Caicedo and Van Huffel, 2010; Ghorbanzade and Fatemi, 2012):

$$H(x, x_i) = \exp\left(-||x - x_i||^2 / \sigma^2\right), \tag{17}$$

where $||x - x_i||^2$ is the squared Euclidean distance between the two feature vectors x and x_i, and σ^2 is the squared variance of the kernel function. If the kernel results in a higher value near the point x following the Gaussian with some spread σ, this indicates similarities between the data points x and x_i in the feature space. The value of the RBF kernel decreases with distance between 0 and 1 in the range of σ threshold values. In other words, for a large enough σ value, all data points are similar to the kernel ($x = x_i$), while, for a small number of σ, only a few points will be in similarity to any particular data points x. In the LS-SVM model, the Mercer condition holds for all σ values of the RBF kernel. Also, the data point x contains only one data point, whereas x_i can have N data points of the same dimension as x. For the LS-SVM approach to be realized in practice, the public-domain toolbox LS-SVMlab v1.8 is used (http://www.esat.kuleuven.be/sista/lssvmlab/), which contains Matlab–C implementations for a number of relevant algorithms. The overall workflow of our two-stage post-reconstruction image processing approach including BH correction and the LS-SVM segmentation approach is illustrated in Fig. 2.

3 Results

3.1 Correction for beam-hardening effects

In the presence of a BH artefact, the reconstructed grey-scale values vary across the rock core from higher values at the periphery to lower values in the central region for the same mineral phase. Visual inspection of the images A, B, and C of our three samples showed that the grey-scale values of the minerals in the central region may overlap with the grey-scale values of other minerals at the periphery (Fig. 3), which would significantly hamper the correct differentiation between both minerals. In order to adjust unequivocally a unique grey-scale level for each mineral phase, we applied the quadratic 2-D polynomial function (Eq. 1) to our images. This polynomial approximation constructs the surface that best fits the cloud of data points subject to the coefficients determined by Eq. (3). The residual data values were extracted as the difference between the values of the original data and those of the fitted surface (Fig. 4c). Due to the BH effect, the attenuation cross-section function across a sample is a parabolic curve rather than a linear line (Fig. 5, light-grey colour lines). For a better visualization of the fitted surface to the frequency (distribution) of grey-scale values (cloud data) of each phase in an image, we illustrated here grey-
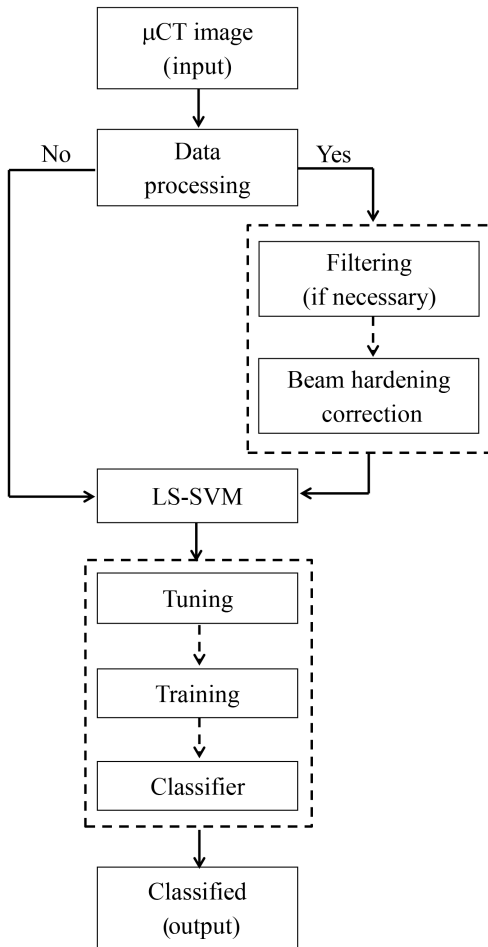
Figure 2. Workflow chart of our proposed μXCT image post-processing method that combines BH correction with an LS-SVM segmentation algorithm.

scale values row-wise at the centre of each image in y direction (Fig. 5). For image A, the fitted function passed through the centre of grey-scale values because of equally distributed phases (sand grains and pores, Fig. 5a). On the other hand, in the case of images B and C, the function was fitted to the abundance of the phase grey-scale values (clay minerals in Fig. 5b, c) and missed out on the fitting to other phases. However, due to the high-contrast grey values of each phases, the grey-level differences in residual grey values were possible. For instance, in the distribution of residual grey-scale values in a 2-D image (Fig. 4c) and the corresponding plot (Fig. 5c, single line black colour), the peaks representing a higher grey-scale level of anhydrite mineral are clearly differentiated from the base level data values representing clay minerals. The images were again reconstructed from the residual data values, which ultimately leads to the efficient removal of the BH artefact in comparison with the original images (compare Fig. 4b and d).

3.2   LS-SVM multi-classification for phase analysis

Upon successful removal of the BH artefacts, the LS-SVM algorithm was tested for the multi-classification task. The performance of the LS-SVM algorithm was ultimately evaluated also in the same image but with uncorrected BH artefacts. The two μXCT images B and C were chosen for the LS-SVM model validation. In our LS-SVM approach, the direct (voxelized) input of an original μXCT image was mapped into a feature vector for training and for testing data points. The rock core μXCT image B was classified into binary phases: clay minerals and cracks/pores; image C was classified into the three major phases: halite, anhydride, and clay minerals. To perform a pixel-based classification, certain regions at different locations (four for image B and six for image C) were manually selected, as marked by letters in Fig. 6a, d. The selection of all pixel values for each phase was performed carefully to avoid boundaries overlapping with each other phase and to limit the misclassification rate. For image B, the total number of data points thus trained for binary phases was 1190 (training data), which is only 0.06 % of the entire pixel data set of the rock core in a 2-D slice. The remaining 1 905 290 pixels were treated as an unknown data set (test data). In the case of image C, a total of 1755 pixels were trained, which is 0.01 % of all object pixels in a 2-D slice, and the remaining 1 570 149 pixels were used as a test data set. It is important to include a possible range of grey-scale level in a training data set, in order to provide maximum information with true class labels; otherwise the classifier considers the output to be undecided. In our LS-SVM approach, the generalization performance of the algorithm requires tuning of a set of hyperparameters (e.g. the regularization constant γ and the RBF kernel parameter σ^2). These tuning parameters were obtained by combining a coupled simulated annealing (CSA) and a standard simplex method. First, CSA was used to determine the appropriate starting points to be transferred to the simplex optimization routine to tune the result. Finally, for both images B and C, an optimal value set of $\gamma = 3.94$, $\sigma^2 = 1.62$, and $\gamma = 4.6$, $\sigma^2 = 1.7$, respectively, was determined in the training data set B and C by applying a leave-one-out routine with a 10-fold cross-validation score function and encoding scheme of one versus one. The remaining data sets of both images B and C were tested based on the predictor feature vector (data points) of the training class labels thus obtained. The output of the data values classified in this way was again reconstructed to give an image in which each distinguished attenuation level was labelled by a single integer value. In Fig. 6b and c, the binary values of "0" and "1" represent the phases of clay minerals and cracks/pores, respectively, of a rock core object inside a 2-D slice. In addition, the labelled values of "1", "2", and "3" for the three phases halite vein, anhydrite, and clay minerals, respectively, are illustrated in Fig. 6e and f. From visual inspection, the LS-SVM performs quite well on the BH-corrected image, in which the label class of each phase distribution is
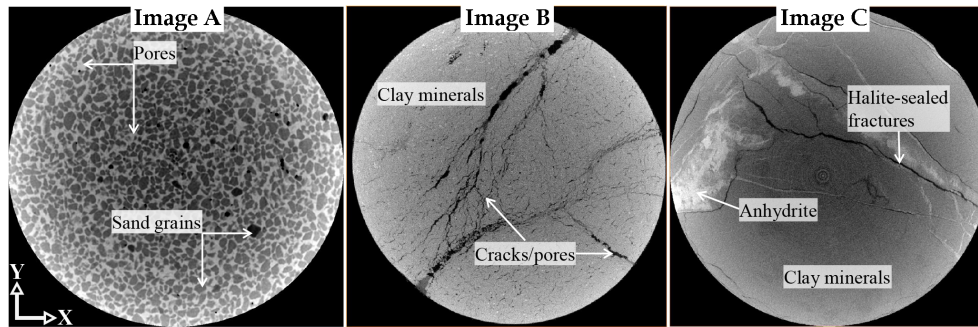
Figure 3. Reconstruction of a μXCT image of computational domain sizes of (A) a 16 bit image of 1327 × 1212 dimensions with each pixel length 5.9 µm of a sand-filled Plexiglass tube, and of (B, C) 8 bit images of dimensions 1600 × 1600 and 1417 × 1417 pixels of two different rock cores (42 µm per pixel).
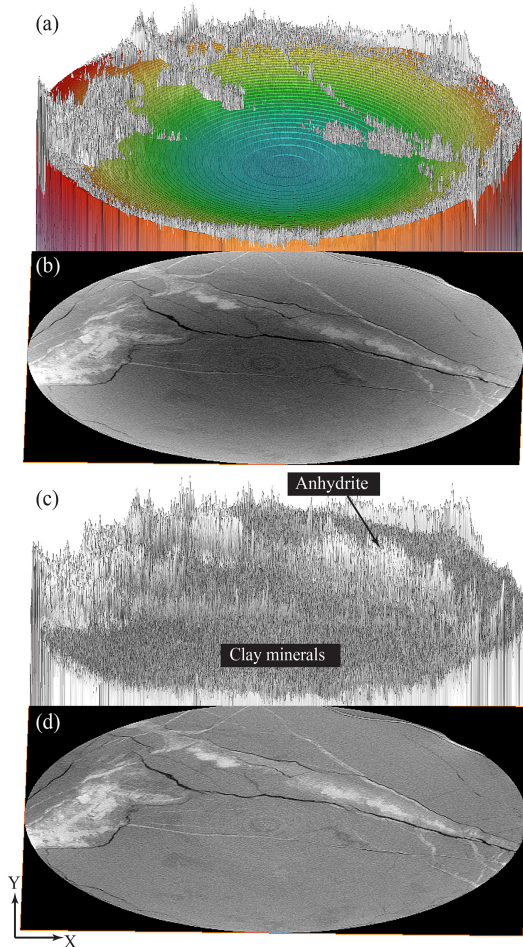

Figure 4. BH correction after noise filtering, where **(a)** depicts the 2-D polynomial surface, fitted to the original image grey-scale values **(b)**. The red–blue colour range in **(a)** represents the elevation of the fitted surface from higher to lower grey-scale values. **(c)** depicts the plot representing the residual grey-scale range of values as a result of the surface fitting, and **(d)** the reconstruction of the BH-corrected image.

well matched with the mineral distribution in the original image but fails to perform in this way on the image with BH artefacts.

4   Discussion

Due to the nature of the BH artefact present in μXCT images, the reconstructed grey-scale data values for the same mineral phase show a non-linear (parabolic) attenuation curve from the periphery to the centre of the rock cores. A 2-D polynomial surface function was fitted to a slice image in order to extract residual data values in terms of the difference between the original data values and the fitted surface points. This BH correction approach is quite flexible for any geomaterial of any shape. In principle, this method could also be applied to non-cylindrical samples and is computationally fast. Any further classification results can be biased by the BH effect if not corrected for, and this bias can be evaluated by a performance measure based on the receiver operating characteristic (ROC) method. The ROC is a statistical measure of the performance of a binary classification test and provides tools to select optimal models in the analysis of decision making (Brown and Davis, 2006; Fawcett, 2006). An ROC curve can be constructed by plotting the specificity ("false-positive rate") against the sensitivity ("true-positive rate") by varying the decision threshold over its entire range. In our LS-SVM model scheme, only binary classification ROC function is integrated. Therefore, the multi-phase classification problem was first decomposed into binary classification tasks; i.e. for an image C, phases of anhydrite and halite were arranged as one class group and compared with the clay minerals as another phase to measure the ROC relationship for LS-SVM both with and without BH-artefact-corrected images. Similarly, the ROC method was implemented on image B with binary phases classified as clay minerals and cracks/pores (see Fig. 6b, c). Note that ROC was implemented only on the training set data to minimize computational costs. In addition to the ROC parameters of
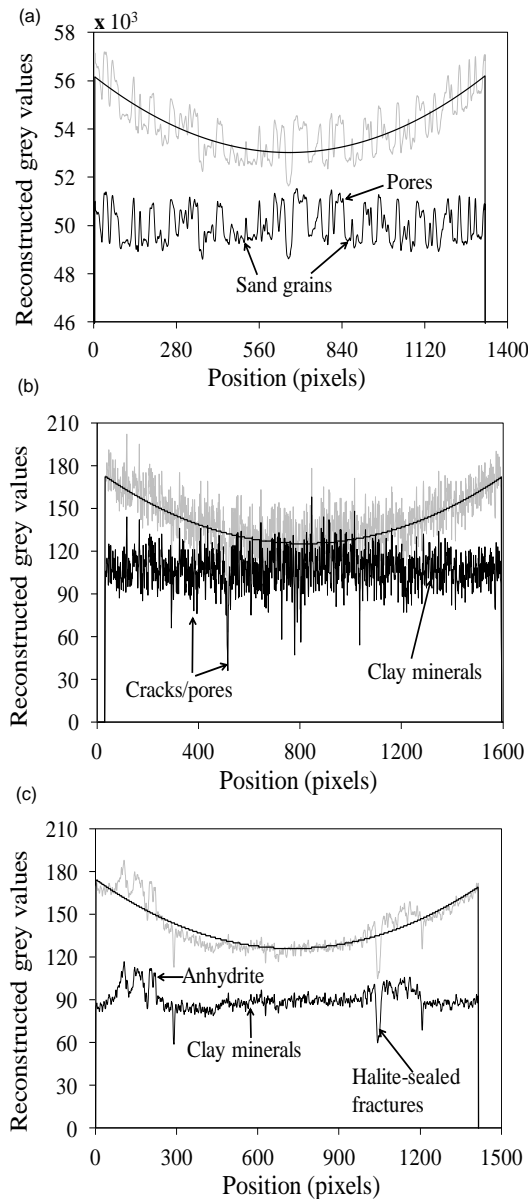
Figure 5. Plot of grey values of BH correction by our proposed method. The curves in plot **(a)**, **(b)**, and **(c)** represent the BH correction corresponding to the images in Fig. 3a, b, and c, respectively. The light-grey colours show the grey values of the BH-uncorrected images, which are extracted row-wise (horizontally) at the centre of each image. The black solid curves represent the residual grey-scale range of values as a result of the fitting.

Figure 6. Pixel-based image classification using LS-SVM, where **(a)** and **(d)** depict locations of pixels selected for training in the original μXCT image, **(b)** and **(e)** the output of the multi-classification on the BH-corrected image, **(c)** and **(f)** the output of multi-classification in the presence of BH artefacts. In image **(b)** and **(c)**, the dark colour represents cracks/pores, while the light colour shows the clay mineral regions of sample A; in images **(e)** and **(f)**, the dark colour represents the halite veins, grey colour the anhydride phase, and light colour the clay mineral region of the evaporite rock cores B and C.

sensitivity and specificity, other important performance measures calculated were the area under the ROC curve (AUC; Hanley and McNeil, 1982; Selvaraj et al., 2007; Luts et al., 2010) and accuracy, which represented the ratio of correctly assigned classes (Brown and Davis, 2006). A typical plot of the ROC curve is shown in Fig. 7.

From the performance classification plot in Fig. 7a and b, the calculated parameters of AUC and accuracy were 0.989
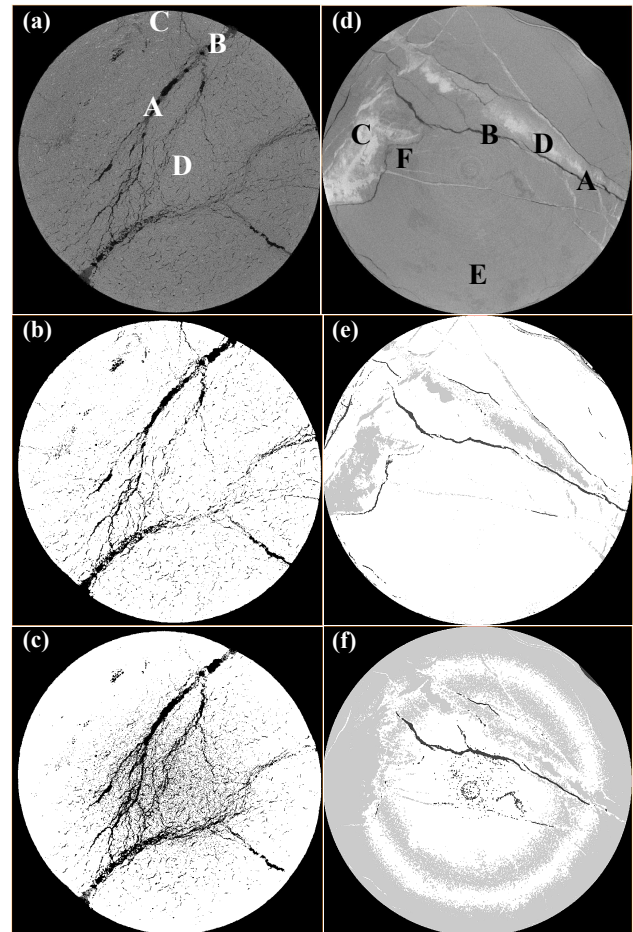
and 99.41 % for the BH-corrected image (Fig. 6b), but as low as 0.901 and 81.10 % in the presence of a BH artefact (Fig. 6c). Similarly, the parameters of AUC and accuracy were calculated from the performance classification plot shown in Fig. 7c and d. The AUC and accuracy were 0.999 and 99.82 % for the BH-corrected image (Fig. 6e) and 0.963 and 88.71% in the presence of a BH artefact, respectively (Fig. 6f). Therefore, the performance measure results based on the pixel-based grey value training data set demonstrate that the probabilistic bias rate was higher in the BH-affected images, and this consequently caused misclassification of the test data. This finding provides evidence that BH correction is an important prerequisite in obtaining a good classifier per-
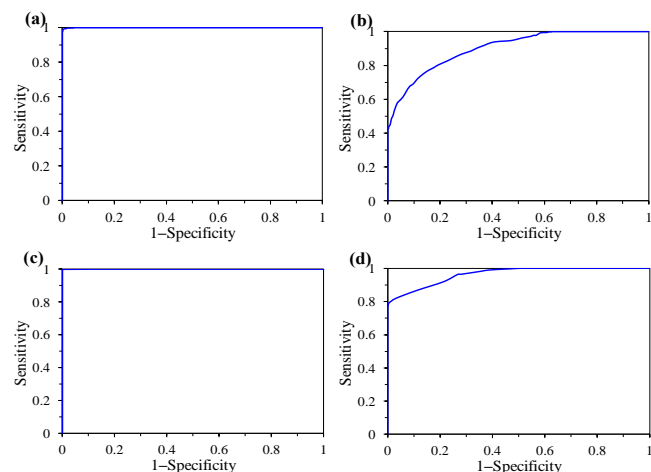
Figure 7. ROC curve analysis of LS-SVM classifier performance on the basis of training set. **(a)** and **(c)** are the result of BH-corrected image, and **(b)** and **(d)** are the outcome of BH-uncorrected image.

formance, e.g. by using the LS-SVM approach. For an optimal classification result, it is always desirable to include the full grey-scale range ("pixel value") of each individual phase to be trained in order to avoid misclassification, i.e. an undecided data classification as undesired output.

5 Conclusions

In this study, polychromatic cone-beam X-ray-source-generated μXCT images of cylindrically shaped samples (sand-filled tube and rock cores) were evaluated for the efficient removal of the beam-hardening artefact and subsequent optimized multi-phase classification. A drawback is that in cases of multi-component geological material of extremely low contrast between different phases, the fitting of the surface function to the cloud of data points may fit one phase well but may miss out another one. Consequently, this leads to over- or underestimation of the range of grey-scale values of each individual phase in the corrected image, which will subsequently affect the correct phase classification. Therefore, it is necessary to apply 2-D-polynomial fitting on the grey range of each individual phase separately.

The advanced least-squares support vector machine (kernel-based learning) method is proposed as an efficient routine to segment the μXCT images on the basis of a direct pixel-based classification task. Without any reduction in dimensionality, i.e. dealing with the large dimensional classification problem, the radial basis function kernel yields overall good classification results for BH-corrected images with a high accuracy rate (less misclassification), but it fails to classify phases in the presence of beam-hardening artefacts as quantified by a ROC analysis. Our method is sensitive to the selection of data points (pixels) at different locations, and to the number of data values of each individual mineral selected for training. Therefore, the presence of artefacts and inadequate data value selection for a specific mineral may affect correct image classification and may become computationally costly as the result of the higher dimensionality of the feature vector. In a companion paper, a comparison is presented of our LS-SVM method with other supervised and unsupervised machine learning techniques to demonstrate that it is best suited for μXCT image segmentation (Chauhan et al., 2016).

Appendix A: Matlab code of beam-hardening (BH) correction by the quadratic surface-fitting approach

```
function [M_corr Surfacefit] = BHC_function(A)
   % Quadratic surface equation of second-order polynomial
   % P(x_k, y_k) = a_1 + a_2 x + a_3 y + a_4 x^2 + a_5 xy + a_6 y^2, k =
1, 2, ..., N
   % To find coefficients "a" of best fit to a function ex-
pressed by M^T Ma = M^T f
   % M_corr       = BH-corrected image
   % Surfacefit   = Surface fit values
   % First convert µXCT image grey-scale values into a ma-
trix "A"
   % Image input parameters
   nX=1417;                    % X dimension of the input
image
   nY=1417;                    % Y dimension
   limitval=1;
   zshift=15 000;    % This can be changed according to im-
age grey-scale range (here 16 bit: 0–65 535)
   % Main function
   [r,c,v]=find(A > limitval);
   M=zeros(size(c,1),6);
   M(:,1)=1;
   M(:,2)=c;                   % x indices
   M(:,3)=r;                   % y indices
   M(:,4)=c.^2;
   M(:,5)=c.*r;
   M(:,6)=r.^2;
   cyl=A > limitval;    % To extract the grey-scale value of
only the object material of the 2-D slice.
   R=cyl.*A;
   [m,n,f]=find(R);
   a=(M'*M)^(−1)*(M'*f);
   p=a(1).*M(:,1)+      a(2).*M(:,2)+      a(3).*M(:,3)+
a(4).*M(:,4)+ a(5).*M(:,5)+ a(6).*M(:,6);
   corr=f-p + zshift;
   S= sparse(r, c,corr, nX,nY);
   M_corr=full(S);
   p1=sparse(r, c,p, nX,nY);
   Surfacefit=full(p1);
   M_corr=uint16(M_corr);
   end
```

Edited by: H. Steeb

References

Alpaydin, E.: Introduction to machine learning, MIT Press, Cambridge, 2004.

Andrä, H., Combaret, N., Dvorkin, J., Glatt, E., Han, J., Kabel, M., Keehm, Y., Krzikalla, F., Lee, M., Madonna, C., Marsh, M., Mukerji, T., Saenger, E. H., Sain, R., Saxena, N., Ricker, S., Wiegmann, A., and Zhan, X.: Digital rock physics benchmarks – Part I: Imaging and segmentation, Comput. Geosci., 50, 25–32, 2013.

Aronszajn, S.: Theory of reproducing kernels, T. Am. Math. Soc., 686, 337–404, 1950.

Berg, S., Ott, H., Klapp, S. A., Schwing, A., Neiteler, R., Brussee, N., Makurat, A., Leu, L., Enzmann, F., Schwarz, J. O., Kersten, M., Irvine, S., and Stampanoni, M.: Real-time 3d imaging of Haines jumps in porous media flow, P. Natl. Acad. Sci. USA., 110, 3755–3759, 2013.

Berg, S., Armstrong, R. T., Georgiadis A., Ott, H., Schwing, A., Neiteler, R., Brussee, N., Makurat, A., Rücker M., Leu, L., Wolf M., Khan F., Enzmann, F., and Kersten, M.: Onset of oil mobilization and nonwetting-phase cluster-size distribution, Petrophysics, 56, 15–22, 2015.

Brabant, L., Pauwels, E., Dierick, M., Van Loo, D., Boone, M. A., and Van Hoorebeke, L.: A novel beam hardening correction method requiring no prior knowledge, incorporated in an iterative reconstruction algorithm, NDT&E Int., 51, 68–73, 2012.

Brown, C. D., and Davis, H. T.: Receiver operating characteristics curves and related decision measures: a tutorial, Chemometr. Intell. Lab., 80, 24–38, 2006.

Caicedo, A. and Van Huffel, S.: Weighted LS-SVM for function estimation applied to artifact removal in bio-signal processing, Eng. Med. Biol. Soc. Ann., 2010, 988–91, doi:10.1109/IEMBS.2010.5627628, 2010.

Chapelle, O., Haffner, P., and Vapnik, V. N.: Support vector machines for histogram-based image classification, IEEE T. Neural Networ., 10, 1055–1064, 1999.

Chauhan, S., Rühaak, W., Khan, F., Enzmann, F., Mielke, P., Kersten, M., and Sass, I.: Processing of rock core microtomography images: using seven different machine learning algorithms, Comput. Geosci., 86, 120–128, 2016.

Cnudde, V. and Boone, M. A.: High-resolution X-ray computed tomography in geosciences: A review of the current technology and applications, Earth-Sci. Rev., 123, 1–17, 2013.

Culligan, K. A., Wildenschild, D., Christensen, B. S., Gray, W. G., Rivers, M. L., and Tompson, A. B.: Interfacial area measurements for unsaturated flow through porous media, Water Resour. Res., 40, W12413, doi:10.1029/2004WR003278, 2004.

Enzmann, F., Meier, T., Janz, M., Jovanovic, Z., Rheingans, K., Schwarz, J., Göbbels, J., and Kersten, M.: Bestimmung der durchflusswirksamen Porosität an Bohrkernproben mittels Computer-Tomographie, Deutsche Gesellschaft für Geowissenschaften, Heft 242, 90–96, 2009.

Fawcett, T.: An introduction to ROC analysis, Pattern Recogn. Lett., 27, 861–874, 2006.

Feldkamp, L. A., Davis, L. C., and Kress, J. W.: Practical conebeam algorithm, J. Opt. Soc. Am. A., 1, 612–619, 1984.

Fusseis, F., Xiao, X., Schrank, C., and De Carlo, F.: A brief guide to synchrotron radiation-based microtomography in (structural) geology and rock mechanics, J. Struct. Geol., 65, 1–16, 2014.

Gallagher, N. C. and Wise, G. L.: A theoretical analysis of the properties of median filters, IEEE T. Acoust. Speech, 29, 1136–1141, 1981.

Ghorbanzade, M. and Fatemi, M. H.: Classification of central nervous system agents by least squares support vector machine based on their structural descriptors: A comparative study, Chemometr. Intell. Lab., 110, 102–107, 2012.

Hanley, J. A. and McNeil, B. J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology, 143, 29–36, 1982.

Hemes, S., Desbois, G., Urai, J. L., Schröppel, B., and Schwarz, J.-O.: Multi-scale characterization of porosity in Boom Clay (HADES-level, Mol, Belgium) using a combination of X-ray μ-CT, 2-D BIB-SEM and FIB-SEM tomography, Mat. Res. S. C., 208, 1–20, 2015.

Herring, A. L., Andersson, L., Schlüter, S., Sheppard, A., and Wildenschild, D.: Efficiently engineering pore-scale processes: The role of force dominance and topology during nonwetting phase trapping in porous media, Adv. Water Resour., 79, 91–102, 2015.

Huber, F., Enzmann, F., Wenka, A., Bouby, M., Bentz, M., and Schäfer, T.: Natural micro-scale heterogeneity induced solute nanoparticle retardation in fractured crystalline rock, J. Contam. Hydrol., 133, 40–52, 2012.

Iassonov, P. and Tuller, M.: Application of segmentation for correction of intensity bias in X-ray computed tomography images, Vadose Zone J., 9, 187–191, 2010.

Jovanović, Z., Khan, F., Enzmann, F., and Kersten, M.: Simultaneous segmentation and beam-hardening correction in computed microtomography of rock cores, Comput. Geosci., 56, 142–150, 2013.

Kaestner, A., Lehmann, E., and Stampanoni, M.: Imaging and image processing in porous media research, Adv. Water Resour., 31, 1174–1187, 2008.

Khan, F., Enzmann, F., and Kersten, M.: 3-D simulation of the permeability tensor in a soil aggregate on basis of nanotomographic imaging and LBE solver, J. Soil. Sediment., 12, 86–96, 2012.

Kotsiantis, S. B.: Supervised machine learning: A review of classification techniques, Informatica, 31, 249–268, 2007.

Krumm, M., Kasperl, S., and Franz, M.: Reducing non-linear artifacts of multi-material objects in industrial 3-D computed tomography, NDT&E Int., 41, 242–251, 2008.

Kumahor, S. K., de Rooij, G.H., Schlüter, S., and Vogel, H.-J.: Water flow and solute transport in unsaturated sand – A comprehensive experimental approach, Vadose Zone J., 14, doi:10.2136/vzj2014.08.0105, 2015.

Landry, C. J., Karpyn, Z. T., and Ayala, O.: Pore-scale lattice Boltzmann modeling and 4-D X-ray computed microtomography imaging of fracture-matrix fluid transfer, Transport Porous Med., 103, 449–468, 2014.

Leu, L., Berg, S., Enzmann, F., Armstrong, R. T., and Kersten, M.: Fast X-ray micro-tomography of multi-phase flow in Berea sandstone: a sensitivity study on image processing, Transport Porous Med., 105, 451–469, 2014.

Luts, J., Ojeda, F., Van de Plas, R., De Moor, B., Huffel, S. V., and Suykens, J. A. K.: A tutorial on support vector machine-based methods for classification problems in chemometrics, Anal. Chim. Acta, 665, 129–145, 2010.

Molins, S., Trebotich, D., Yang, L., Ajo-Franklin, J. B., Ligocki, T. J., Shen, C., and Steefel, C. I.: Pore-scale controls on calcite dissolution rates from flow-through laboratory and numerical experiments, Environ. Sci. Technol., 48, 7453–7460, 2014.

Schlüter, S., Sheppard, A., Brown, K., and Wildenschild, D.: Image processing of multiphase images obtained via X-ray microtomography: A review, Water Resour. Res., 50, 3615–3639, 2014.

Schwarz, J.-O. and Enzmann, F.: Simulation of fluid flow on fractures and implications for reactive transport simulations, Transport Porous Med., 96, 501–525, 2013.

Sell, K., Enzmann, F., Kersten, M., and Spangenberg, E.: Microtomographic quantification of hydraulic clay mineral displacement effects during a CO_2 sequestration experiment with saline aquifer sandstone, Environ. Sci. Technol., 47, 198–204, 2013.

Selvaraj, H., Selvi, S. T., Selvathi, D., and Gewali, L.: Brain MRI slices classification using least squares support vector machine, IC–MED, 1, 21–33, 2007.

Suykens, J. A. K. and Vandewalle, J.: Least squares support vector machine classifiers, Neural Process. Lett., 9, 293–300, 1999.

Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., and Vandewalle, J.: Least Squares Support Vector Machines, World Scientific, Singapore, ISBN 981-238-151-1, 2002.

Van Gestel, T., Suykens, J. A. K., Lanckriet, G., Lambrechts, A., De Moor, B., and Vandewalle, J.: Multiclass LS-SVMs: moderated outputs and coding-decoding schemes, Neural Process. Lett., 15, 45–58, 2002.

Vapnik, V.: The nature of statistical learning theory, Springer, New York, 1995.

Vlassenbroeck, J., Dierick, M., Masschaele, B., Cnudde, V., Van Hoorebeke, L., and Jacobs, P.: Software tools for quantification of X-ray microtomography, Nucl. Instrum. Meth A., 580, 442–445, 2007.