AIMS Materials Science, 9(6): 919–938.
DOI: 10.3934/matersci.2022054
Received: 03 September 2022
Revised: 14 October 2022
Accepted: 25 October 2022
Published: 09 November 2022

http://www.aimspress.com/journal/Materials

Research article

Investigation of water desalination/purification with molecular dynamics and machine learning techniques

Christos Stavrogiannis, Filippos Sofos*, Theodoros. E. Karakasidis and Denis Vavougios

Department of Physics, School of Science, University of Thessaly, 35100 Lamia, Greece

* **Correspondence:** Email: fsofos@uth.gr.

Abstract: This paper incorporates a number of parameters, such as nanopore size, wall wettability, and electric field strength, to assess their effect on ion removal from nanochannels filled with water. Molecular dynamics simulations are incorporated to monitor the process and a numerical database is created with the results. We show that the movement of ions in water nanochannels under the effect of an electric field is multifactorial. Potential energy regions of various strength are formed inside the nanochannel, and ions are either drifted to the walls and rejected from the solution or form clusters that are trapped inside low potential energy regions. Further computational investigation is made with the incorporation of machine learning techniques that suggest an alternative path to predict the water/ion solution properties. Our test procedure here involves the calculation of diffusion coefficient values and the incorporation of four ML algorithms, for comparison reasons, which exploit MD calculated results and are trained to predict the diffusion coefficient values in cases where no simulation data exist. This two-fold computational approach constitutes a fast and accurate solution that could be adjusted to similar ion separation models for property extraction.

Keywords: ion separation; desalination; electric field; diffusion coefficient; potential energy; machine learning

1. Introduction

Ion and contaminant removal from aqueous solutions is an ever-growing research trend in the last decades, mainly due to water scarcity all around the globe, which has become prominent nowadays. The idea of removing undesired substances from water through membrane-based

processes, such as reverse osmosis (RO) and nanofiltration (NF), has received a lot of attention in the scientific community [1–4]. Ions are usually extracted from saline solutions with aid of an electric field in electric-assisted desalination [5], while adsorbent-based methods have been also exploited [6]. Moreover, electrodialysis (ED), capacitive deionization (CDI), and flow-electrode capacitive deionization (FCDI) are some of the current methods utilized [7]. Because of their proximity to relevant biological processes, nanometer-sized channels are ideal for such applications.

Novel materials that have been proven suitable for water desalination and purification approaches, both in experimental and theoretical investigations, include graphene, where membrane multilayer structures with slit width greater than 3 molecular diameters can allow unrestricted mixed molecular movement between its sheets [8], while low oxidation and graphene oxide (GO) is a top performer at ion separation efficiency [9]. For improved performance, stacked graphene membranes can be also used as a potential ion separator [10,11]. Moreover, carbon nanotubes (CNTs), metal-organic frameworks (MOFs), and membrane protein channels in block copolymers (MP-BCPs) [12–16], are also choices that have found wide applicability. Of special importance is also the structure of nanochannels, as it could either facilitate or block ion transport [17–19]. Furthermore, the critical role of wall wettability must be emphasized at the nanoscale, as it influences most static and dynamic fluid properties [20–22]. It has been shown that by adjusting the wall wettability, net unidirectional flow can be generated [23,24]. What's more, the water molecule contact angle on the surface, which portrays wall hydrophobicity and hydrophilicity, and ion selectivity [25] is affected in the presence of an external electric field [26].

Apart from experimental methods being incorporated, continuum and atomic-scale simulations have the guiding role in all these applications. The Poisson-Nerst-Planck (PNP) theory is a continuum approach that describes ionic transport effectively [27–29], however, the detail obtained by nanoscale simulations is oftentimes preferable [30,31]. Classical molecular dynamics (MD) simulations have been incorporated in most of these cases, revealing hidden phenomena taking place during ion drift, assisting experimental procedures, and highlighting key critical theoretical issues for the establishment of functional models [32,33], while, hybrid and multiscale techniques have been constantly gaining ground [34]. Other approaches include the application of continuum theory at the nanoscale with satisfying results [35] and the incorporation of a generalized Navier-Stokes equation, which concludes that erroneous force truncation in MD simulations may result in spurious flow effects for static electric fields [36].

The term ML refers to obtaining a computational model of linear/non-linear relationships or complex patterns, with reference only to data. In scientific problems, data are usually gathered from experimental or simulation procedures [30]. ML employs statistical approaches to analyze and build algorithms trained on data and generate predictions based on it. A key question is whether such data-driven methods fit to a problem or not. One could say that ML can be seen as an alternative to costly simulation-based methods and trial-and-error methodologies, providing predictions for only a fraction of the respective computational burden, while keeping the accuracy high. A number of ML algorithms have been incorporated in scientific and technological problems, with tree-based techniques, such as Random Forest, Decision Trees, Gradient Boosting Regressor, to mention a few, providing more than adequate predictions, especially in fluid mechanics research [31], while at the field of membrane fouling, supervised ML algorithms, such as Artificial Neural Networks (ANN), have achieved higher accuracy in less effort and time [32].

MD simulations are exploited in this paper to investigate the ion removal mechanism from

nanochannels in which ion transport is guided by an external electric field perpendicular to the walls [28,31]. The potential energy and electric field maps are created, and diffusion coefficients of the solution are extracted. It is found that the critical parameters that affect ion drift are the channel width, *h*, the external electric field, *E*, and the wall/fluid interaction strength $\varepsilon_{wf}/\varepsilon_{ff}$. These parameters are, then, fed into an ML computational procedure and predictions outside the simulation range are made, with the coefficient of determination approaching values close to unity. MD simulation is a technique that can provide ample numerical data to train an ML algorithm. Towards proposing a framework of combined MD/ML (Figure 1), a new computational approach arises that could be incorporated in property calculation platforms, at least in similar applications, posing as a hybrid technique that could boost calculation times, while making ML an indispensable tool that advances our ability to investigate science and technology problems.

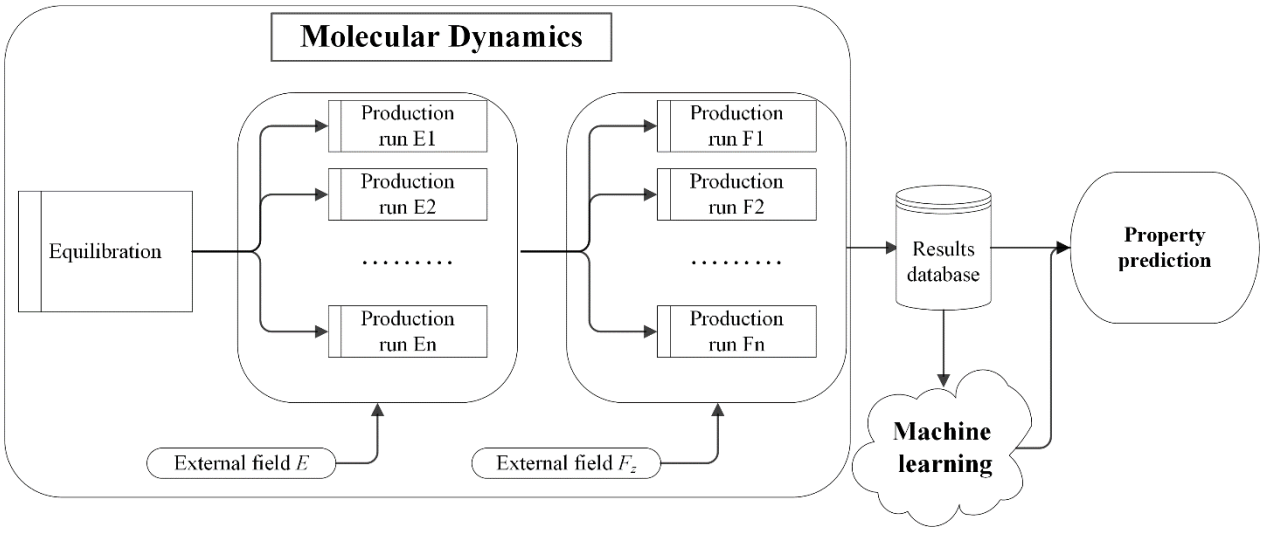


Figure 1. Data flow, including MD simulation and machine learning stages.

2. Materials and methods

2.1. Molecular Dynamics model

Water separation from Na^+ and Cl^- ions is performed with an MD model consisted of two impermeable carbon walls (a nanochannel) and an electrical field acting as the driving force towards the walls. The system is considered periodic in the *x*-, *y*-, and *z*- directions (Figure 2). Table 1 lists all the parameters and variables of the dimensions of the materials used. Wall and fluid particles are given Lennard-Jones (LJ) parameters to make the simulation less computationally demanding. Equation 1 describes the total potential as sum of the LJ potential between two particles *i* and *j* and the coulombic potential due to electrostatic forces (where present), as

$$u(r_{ij}) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] + \frac{Cq_iq_j}{\varepsilon_0 r_{ij}} & r_{ij} < r_c \\ 0 & r_{ij} \geq r_c \end{cases} \quad (1)$$

where ε represents the interconnection's strength, σ the length range, r_c is the cut-off radius, C is the energy conversion constant, q_i and q_j are the charges of interacting atoms, and ε_0 is the dielectric constant. Arithmetic Lorentz-Berthelot combining rules are applied to all interatomic interconnections based on the values in Table 2. Wall wettability is determined by the $\varepsilon_{wf}/\varepsilon_{ff}$ ratio of wall-to-fluid interaction. The wall is considered strongly hydrophilic when $\varepsilon_{wf}/\varepsilon_{ff}$ = 1.0, hydrophilic when $\varepsilon_{wf}/\varepsilon_{ff}$ = 0.5, hydrophobic when $\varepsilon_{wf}/\varepsilon_{ff}$ = 0.2, strong hydrophobic when $\varepsilon_{wf}/\varepsilon_{ff}$ [33–35]. Water molecules conform to the SPC/E (extended simple point charge) pair potential, which has been found to adequately reproduce water's structural and dynamic properties [36].

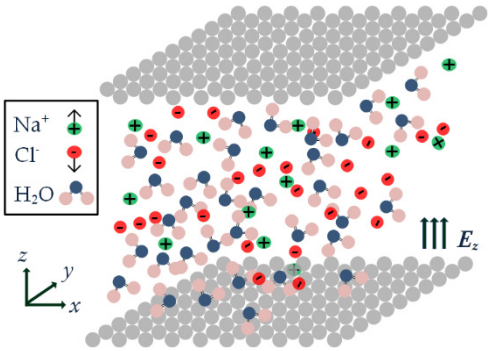


Figure 2. Ion/water flow between two carbon walls, channel model (periodic in x- and y-dimensions, while, h, is the channel height, in the range $3\ nm \leq h \leq 62\ nm$). Green and red circles are Na^+ and Cl^- ions, respectively. The external electric field E_z is applied perpendicular to the flow (driven by the external force F_x).

Table 1. Simulation parameters.

Model property	Value
Channel height, h (nm)	3–21 nm
Computational dimension in x, L_x (nm)	3.1
Computational dimension in y, L_y (nm)	1.5
Computational dimension in z, L_z (nm)	3.72–23.72
External electric field, E_z (V/Å)	0.0–1.0
External force, F_x (kcal/molÅ)	0.25
Wall-to-fluid ε ratio, $\varepsilon_{wf}/\varepsilon_{ff}$	0.1–1.0

Table 2. Parameters (ε, σ, and m) for each atom in the solution.

Atom	ε (Kcal/mol)	σ (Å)	mass (a.u.)
H	0	0	1.008
O	0.155	3.166	15.99
C	0.056	3.400	12.01
Cl	0.107	4.446	35.45
Na	1.607	1.897	22.99

The SPC/E water model is incorporated here, which is composed of LJ terms and Coulombic potential for the ion/water solution. Water bonds are constrained using the SHAKE algorithm [37] and no dissociation phenomena are observed. A particle-particle particle-mesh (PPPM) method has been used to calculate the long-range electrostatic force. An external electric field (similar to a solution between two charged surfaces) should result in anion drift towards the "positive" surface and cation drift towards the "negative" surface. The electric field is caused by the homogeneous distribution of opposite sign charges on the two walls, which produces an electric force $F_e = qE_z$ acting on the z-direction. In this work, the applied E_z is investigated over the range 0.0–1.0 V/Å.

The MD procedure involves the following steps, as can be also seen from Figure 1, (a) an equilibrium MD run of t_{eq} = 10 ns is incorporated, (b) in the first production runs (parallel independent runs), the desired desalination procedure takes place when most ions approach the walls due to the effect of E_z, (c) then, on each fluid particle, an external driving force F_x is applied along the x-direction, and flow is induced (in the same manner as pressure difference acts on Poiseuille flow), and (d) as ions are forced to diverge from the solution, clean water flows through the channel interior and must be collected appropriately at the outlet (not shown here).

The results are saved, averaged (we have considered n = 5), and secured in the simulation database for later use. Wall atoms absorb the increase in fluid kinetic energy caused by the application of external fields, and Nosé-Hoover thermostats are used at the walls to maintain a constant system temperature (T = 300 K).

2.2. Computational details

The impact of an external electric field on ion/water flows in conduits is investigated. A multivariate interaction environment is considered, taking into account the hydrophobic/hydrophilic nature of the wall as well as the repulsive/attractive forces between positive and negative ions. As a result, potential energy maps are extracted to reveal probable ion positions. The computational space is divided into *mxnxk* bins along (*x*,*y*,*z*)-dimensions and local calculations are conducted.

Potential energy as a local quantity, U_{mnk}, is calculated in each one of the *mxnxk* bins as

$$U_{mnk} = U_{mnk}^{LJ} + V_{mnk}^{c} + U_{mnk}^{E} \tag{2}$$

U_{mnk}^{LJ} refers to the local Lennard-Jones interaction potential within each computational bin, which includes wall/fluid interactions in terms of hydro/ion-phobicity and hydro/ion-philicity. The potential energy associated with long-range electrical interactions between all charged atoms, H^+ and O^- from water molecules, and Na^+, Cl^- as free-running ions in the solution, is denoted by the second term, V_{mnk}^{c}. The result of applying the external field perpendicular to the walls, E_z, is associated with U_{mnk}^{E}.

The potential energy map inside the channels has been extracted using Eq 2. Because the goal of the proposed model is to present a method of ion drift towards the walls, this would be useful information, because low (negative) potential regions can reveal possible ion caging positions inside the channel, which would obstruct their transport.

The induced electric field caused by the presence of several N_c charged atoms within a bin is calculated.

$$\vec{E}_{ion}(\vec{r}_i) = -\frac{\partial V^c}{\partial x}\Big|_{\vec{r}_i} \vec{\mathrm{x}} - \frac{\partial V^c}{\partial y}\Big|_{\vec{r}_i} \vec{\mathrm{y}} - \frac{\partial V^c}{\partial z}\Big|_{\vec{r}_i} \vec{\mathrm{z}} \tag{3}$$

or equivalently

$$\vec{E}_{ion}(\vec{r}_i) = \frac{1}{4\pi\varepsilon_0\varepsilon_r} \sum_{j=1, j\neq i}^{N_c} \frac{q_j}{\left|\vec{r}_i - \vec{r}_j\right|^3}\left(\vec{r}_i - \vec{r}_j\right) \tag{4}$$

where ε_r denotes the dielectric constant of water. For each *mxnxk* bin, the electric field is calculated locally. We get the local value $\vec{E}_{mnk}$ by adding the external field $\vec{E}_{ext}$, which is distributed evenly throughout the channel.

$$\vec{E}_{mnk} = \vec{E}_{ion} + \vec{E}_{ext} \tag{5}$$

Using Einstein's relation, the average diffusion coefficient throughout the channel is calculated by time-averaging the mean square displacement of *N* fluid particles.

$$MSD = \frac{1}{N}\left\langle \sum_{j=1}^{N} \left[\boldsymbol{r}_j(t) - \boldsymbol{r}_j(0)\right]^2 \right\rangle \tag{6}$$

and

$$D = \lim_{t\to\infty} \frac{1}{6Nt}\left\langle \sum_{j=1}^{N} \left[\boldsymbol{r}_j(t) - \boldsymbol{r}_j(0)\right]^2 \right\rangle \tag{7}$$

where $\mathbf{r}_j$ is the j^{th} atom's position vector. Equation 7 is usually used for extracting the diffusion coefficients for systems at equilibrium. Here, we have a non-equilibrium system, where forces are applied on the *x*-dimension to drive the flow. However, it can be used in such cases, provided one excludes the drift contribution from the flow [38].

2.3. Machine Learning Algorithms

2.1.1 Decision trees

The decision tree (DT) ML algorithm has recently gained popularity as a supervised machine learning (ML) method for predicting processes and events. This algorithm is useful in problems involving data sets that must be classified. A decision tree is made up of several transitional (decision-makers) and leaf nodes. Each decision node separates the data into two parts based on one or more input variables, and this process is repeated hierarchically from middle to leaf nodes. Each leaf node specifies the final clause value [39]. The tree can be described by the following function:

$$F_K(x) = \sum_{I \in leaves(T)} k_I \, I_I(x) \tag{8}$$

The three critical variables for the DT classifier are the maximum depth of the tree, the minimum number of samples required for splitting, and the number of maximum features. Decision Tree (DT) algorithm is often used as a comparative candidate among other ML methods, without always having the best performance [40].

2.1.2 Random Forest

The Random Forest (RF) algorithm is a grouping of DTs that predicts the final value using the averaging method

$$Y = \frac{1}{b} \sum_{j=1}^{b} Y_b(X') \tag{9}$$

where Y_b is each DT and X' is the number of unknown scenarios. The total number of DTs is presented by the variable b. A segment of training dataset known as the "bootstrapped dataset" creates each tree in the forest. To build the trees, a random set of characteristics is also chosen. There is no need to perform input normalization before training the RF model. The RF method can automatically determine the most influential parameters. In comparison to the DT model, the forest developed with randomness has a better prediction performance and has one additional critical value which is the number of trees. The RF algorithm was utilized for predicting salt absorption capacity with great success [41], and it has also been the best performing ML method to predict salt passage [42].

2.1.3 Gradient Boosting

The Gradient Boosting Regressor (GBR) is a supervised ML method where an initial basic learner is formed by a function $f_0(x)$, using a sample of known x, y values as trainer. The goal is to estimate the function $F(x)$, which maps variables x to their output values y, by decreasing the expected value of a given loss function, $F_w(x) = L(y, F(x))$, forming a node tree-like schematic. Gradient boosting produces a weighted sum of functions as an additive estimate of $F(x)$, as

$$F_w(x) = F_{w-1}(x) + p_w h_w(x) \tag{10}$$

where p_w is the weight of the w_{th} function, h_w(x). This method, due to its weakness at overfitting, is best suited with data that are regularized or combined with other ML methods for optimum performance. Gradient Boosting tree algorithm was successfully used in Thin Film Nanocomposite RO membranes to predict and form models of water permeability and salt rejection rate [41]. The GB method was the best performer among other ML algorithms in predicting properties of a 2D membrane during MD simulations, while a hybrid GBR implementation has performed well in predicting the internal concentration polarization in Forward Osmosis membranes [43].

2.1.4 Multi-Layer Perceptron

One of the most fundamental and significant ANN models is the Multi-Layer Perceptron (MLP), a supervised ML method [44], which mimics the transcriptional patterns of human brain. These networks usually have an input layer, one or more middle-hidden layers, and an output layer. The number of hidden layers can be altered and calibrated based on the task's sophistication to achieve the best results for process simulation. The depth (or shallowness) of a machine learning model is determined by the number of hidden layers. The input signal is routed through the network layer by layer, so that each sensory unit analyzes the signal after receiving it from a neural or non-neural neuron and conveys the result to another sensory unit. This behavior will continue until a valid conclusion is reached, a selection will be made, and further processing will begin. The majority of the network behavior of the human brain and signal transmission are scrutinized in this model.

$$Y = f(w\, x_i + \, b) \tag{11}$$

where Y, w, x_i, and b represent output data, weighted vector data, input data, and the input bias term, respectively. A supervised learning method is employed to calculate weights. MLP can understand nonlinear function approximation for classification and regression given a set of parameters. The MLP model has shown high capability and accuracy in predicting the amount of water produced by the CDI method [45], and CO_2 solubility in salt solutions [46].

3. Results and discussion

3.1. Potential energy maps

To study the effects of electric field strength on ion localization we employ a hydrophobic ($\varepsilon_{wf}/\varepsilon_{ff} = 0.1$) nanochannel of size h = 3 nm. At first, pure water solution for E_z = 0 V/Å is studied, and Figure 3a shows the potential energy map. We remind here that the lowest negative potential energy regions are preferred by fluid particles [37]. The negative potential energy regions are randomly distributed throughout the channel, while regions near the walls may attract water molecules due to the hydrophobic walls [34]. When an external field E_z = 0.01 V/Å is applied to the z-direction, this tendency becomes slightly stronger (Figure 3b). The solution is then enriched with Na^+ and Cl^- ions for E_z = 0.1 V/Å (Figure 3c). A new behavior occurs, as low/high potential energy regions have been formed near the upper and lower walls, along with scarce negative potential energy regions near the channel center. It becomes clear that the presence of ions has altered the potential energy distribution, and, at least for the given electric field strength value, all ions may not drift towards the wall, as negative potential regions near the channel center may prevent their movement. For E_z = 0.5 V/Å (Figure 3d), "shady" potential regions are formed primarily near the walls, hinting ion separation.

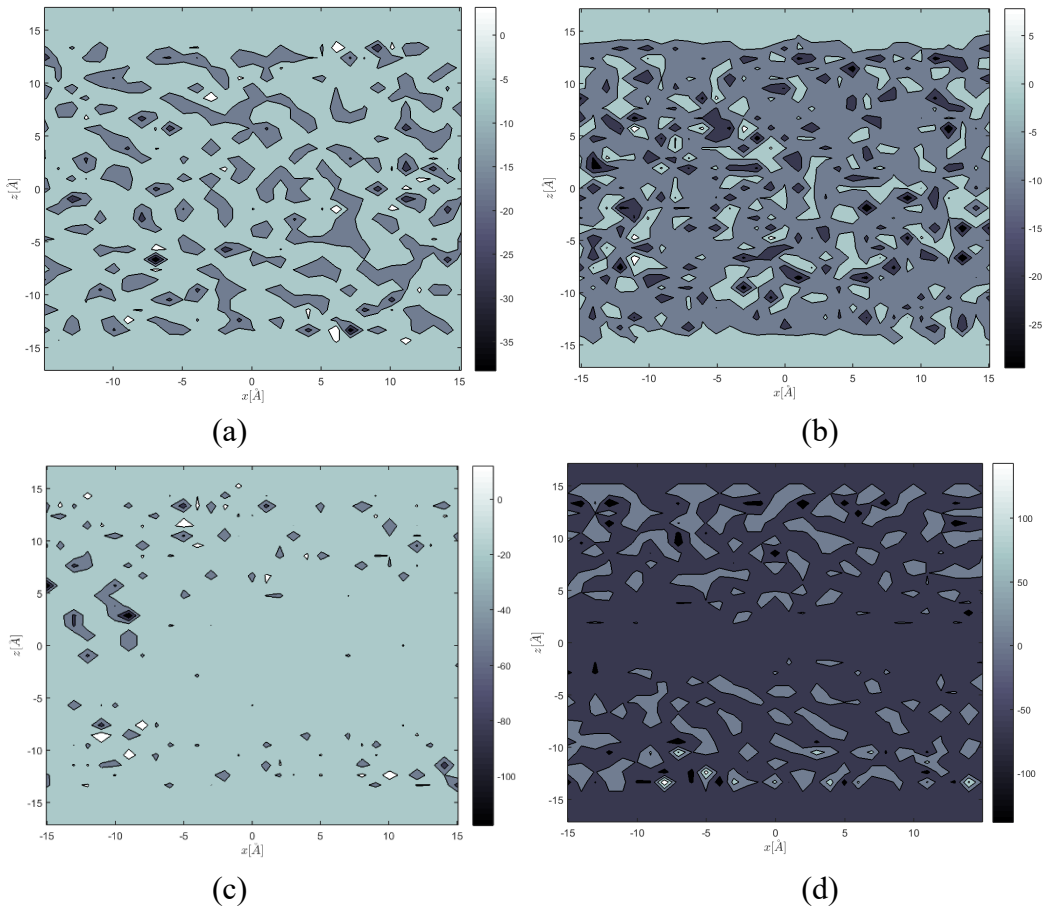
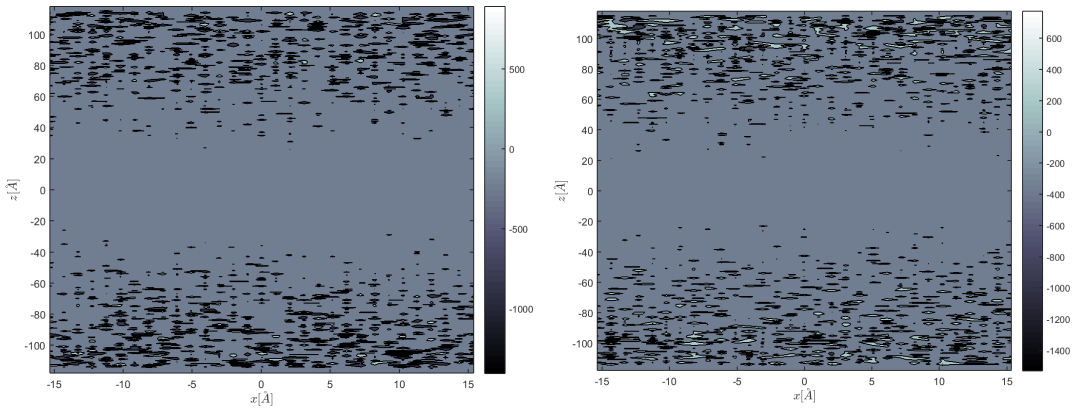


Figure 3. 2D potential energy map (along the zx-plane) inside the h = 3 nm, hydrophobic ($\varepsilon_{wf}/\varepsilon_{ff} = 0.1$) channel, and (a) pure water and E_z = 0.0 V/Å, (b) pure water and E_z = 0.01 V/Å, (c) water/on solution (Na^+ and Cl^-) and E_z = 0.1 V/Å, and (d) water/on solution (Na^+ and Cl^-) and E_z = 0.5 V/Å.

Another important factor affecting fluid particle localization is wall/fluid interaction, which is stronger near the walls, inside the interaction radius of the wall particles [47–49]. Two cases are presented next, a highly hydrophobic case ($\varepsilon_{wf}/\varepsilon_{ff} = 0.1$) in Figure 4a, and a highly hydrophilic case ($\varepsilon_{wf}/\varepsilon_{ff} = 1.0$) in Figure 4b. The channel size is h = 21 nm and the electrical field is E_z = 0.5 V/Å. The $\varepsilon_{wf}/\varepsilon_{ff} = 0.1$ case indicates that regions of lower, negative potential appear near the walls, along with few positive potential regions. When we consider highly hydrophilic walls, $\varepsilon_{wf}/\varepsilon_{ff} = 1.0$, in Figure 4b, both low and high potential regions appear near the upper and lower walls, forming an interaction region where ions could be trapped. This could also hinder the presence of Na^+ and Cl^- ions in this region. In any case, there appears to be no ionic interactions near the channel center, as potential energy values are close to zero. As a result, ion removal from the channel center region has been accomplished.



(a) (b)

Figure 4. 2D potential energy map (along the zx-plane) inside the h = 21 nm channel, for E_z = 0.5 V/Å and (a) highly hydrophobic walls ($\varepsilon_{wf}/\varepsilon_{ff} = 0.1$, (b) highly hydrophilic walls ($\varepsilon_{wf}/\varepsilon_{ff} = 1.0$).

We now turn our attention on the effect of channel size on the potential energy map. Ion/ water flow simulations in this work include the investigation of channels spanning from $3\,\text{nm} \leq h \leq 21\,nm$. We consider $\varepsilon_{wf}/\varepsilon_{ff} = 0.1$ and E_z = 0.1 V/Å. Low and close to zero potential regions are revealed near the upper wall of the h=6nm channel, along with a strong negative potential region at the channel center (Figure 5a). As the channel size increases to h = 9 nm (Figure 5b), negative potential regions appear near the upper wall, and some negative regions still remain near the channel center. For the h = 15 nm channel, a multiple interaction environment is also observed near the walls (Figure 5c).

Finally, in Figure 5d, where the channel dimension has been increased to h = 21 nm, lower potential regions are found throughout the channel, making it more difficult for ions to reach the walls. To summarize the effect of channel size, the presence of negative potential energy regions near the walls, which indicate probable ion position during the simulation, is not solely affected by the strength of the external electric field. Although wider channels receive more electric energy when the same E_z is applied, ions must travel a greater distance and may become trapped inside a low potential region away from the walls.

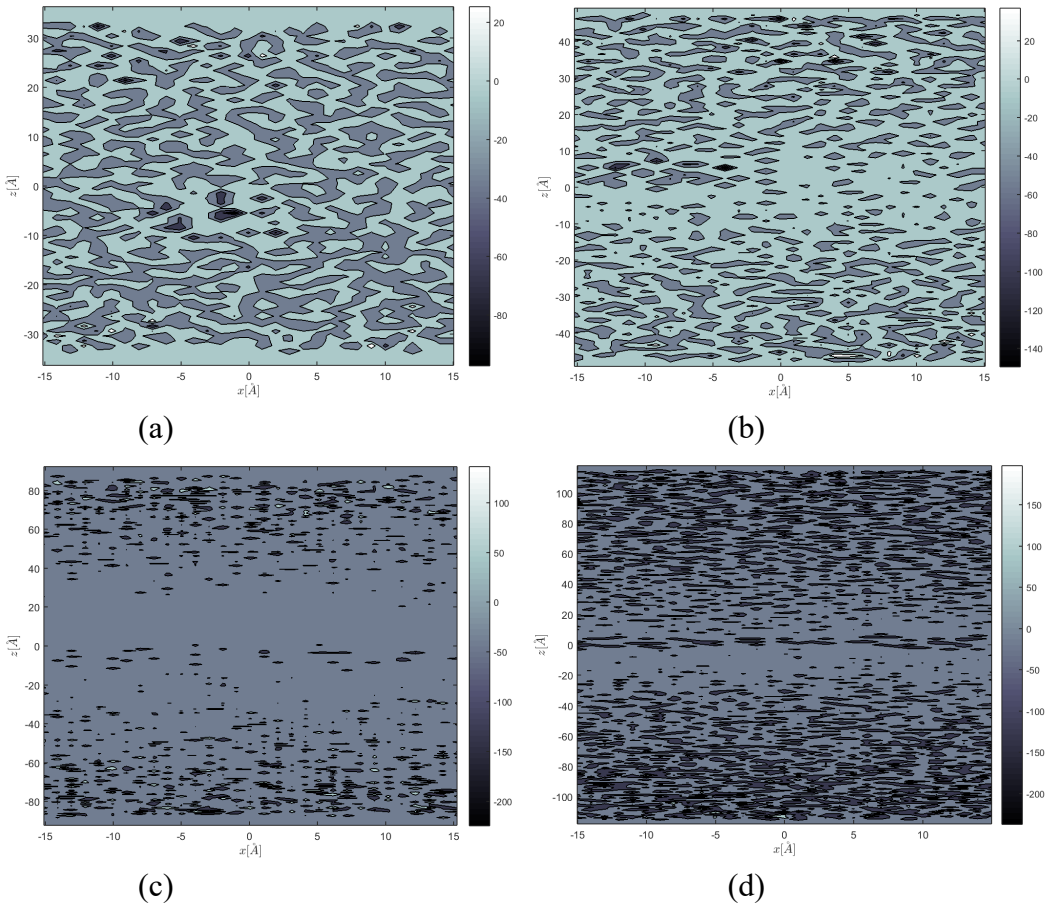


Figure 5. Energy potential regions, with E_z = 0.1 V/Å, $\varepsilon_{wf}/\varepsilon_{ff} = 0.1$, and for channels equal to (a) h = 6 nm, (b) h = 9 nm, (c) h = 15 nm, and (d) h = 21 nm.

3.2. Ion localization

When an electric field is applied perpendicular to the walls in similar models, water molecules are aligned, with hydrogen atoms pointing in the negative direction [38]. Without ions in the solution, the electric field lines would be straight lines parallel along the z-direction. Transport along the z-axis (towards the upper or lower wall) is expected in the presence of ions, if E_z is strong enough to overcome interatomic interactions. Ion separation is not always clear; Na^+/Cl^- ion pairs form at low E_z strength and may move together towards the walls or, in the worst-case scenario, wander around the channel centerline. Furthermore, Na^+ and Cl^- ions have distinct transport mechanisms and diffuse at different rates [50]. Larger ions (Cl^-) are less bonded to water molecules and thus more susceptible to being driven by an electric field than smaller Na^+ ions [15]. As a result, ion movement is somewhat complicated. As ions move towards the upper or lower wall, a new electric field, E_{ion}, opposite to E_z, emerges and competes with E_z. Further research is required to elucidate this mechanism.

We evaluate the electric field effect inside the ion/water solution. At first, we consider E_z = 0.1 V/Å and h = 3 nm, in Figure 6a. It is observed that E_z = 0.1 V/Å is not strong enough to anticipate high ion drift movement towards the walls as ions are mostly clustered just above the channel's center. Another observation is that the presence of ions has redirected the external field

(z-direction) flow lines in nearby regions, which appears to be impeding their movement towards the walls. The potential energy map shown previously in Figure 3c has captured this effect by revealing local minima in the same region where ions are now set. As a result, ions must absorb more energy to overcome this barrier. In Figure 6b, we increase the electric field strength to E_z = 1.0 V/Å. In this case the ion movement is clear, electric field lines are straight, and this makes ions be localized near the upper and lower walls.

Consequently, the effect of wall/fluid-particle interaction is investigated. We also note the distance d which refers to the minimum ion distance from the lower wall, inside the h = 21 nm channel for two cases, a hydrophobic wall ($\varepsilon_{wf}/\varepsilon_{ff} = 0.2$), in Figure 6c, and a hydrophilic wall ($\varepsilon_{wf}/\varepsilon_{ff} = 0.5$), Figure 6d. We have incorporated a strong electric field, E_z = 1.5 V/Å. Parameter d shows no significant difference between the two cases. Therefore, we come into conclusion that effect of wall wettability on ionic separation from water solutions is not critical for the proposed desalination model at this scale. Nevertheless, it may be significant in other situations, such as those resembling biological processes that take place at smaller scales, around 1–2 nm, as it has been found elsewhere [28,51] .

From another point of view, it has been shown that in confined nanochannels, where electric fields are applied perpendicular to the walls, one wall acts as hydrophobic and the other as hydrophilic [52]. In such cases, small external electric fields do not affect hydrogen bonding, while a large E can break these bonds. Water molecules are oriented, with hydrogen atoms pointing on the negative direction and oxygen on the positive direction of the field. We have to note here that ion concentration considered in this work is rather small to affect the electrical behavior of internally induced electric fields, so water behavior near the walls is mainly due to molecules orientation. Breaking of electric lines continuity observed in Figure 6a,c,e,f is a local phenomenon and is attributed to increased local ion concentration, which takes place inside a small region of the channel.

The impact of channel size (h = 6 nm and h = 9 nm) on ionic separation in water solutions is shown in Figure 6e–f, respectively, for E_z = 1.0 V/Å. We note that the electric field E_z forms parallel lines across the z-direction, indicating that all ions are drifted near the walls for h = 6 nm. This finding validates the proposed desalination procedure, at least for this system size. However, field orientation is not parallel to the z-direction for the h = 9 nm channel, and ions appear to be blocked near the channel center.

To summarize our findings, the electric field distribution studied in the previous cases shows that ion removal in this scale is independent of wall wettability and channel size for cases $h < 6$ nm, but it is strongly affected for $h > 6$ nm. It seems that the most important parameter for controlling ion separation for this solution is the strength of the electric field, which is the driving force of the ion drift movement. Channel size is important, as well, since smaller channels achieve better ion separation, and this is consistent to observations presented in similar works [10,25].

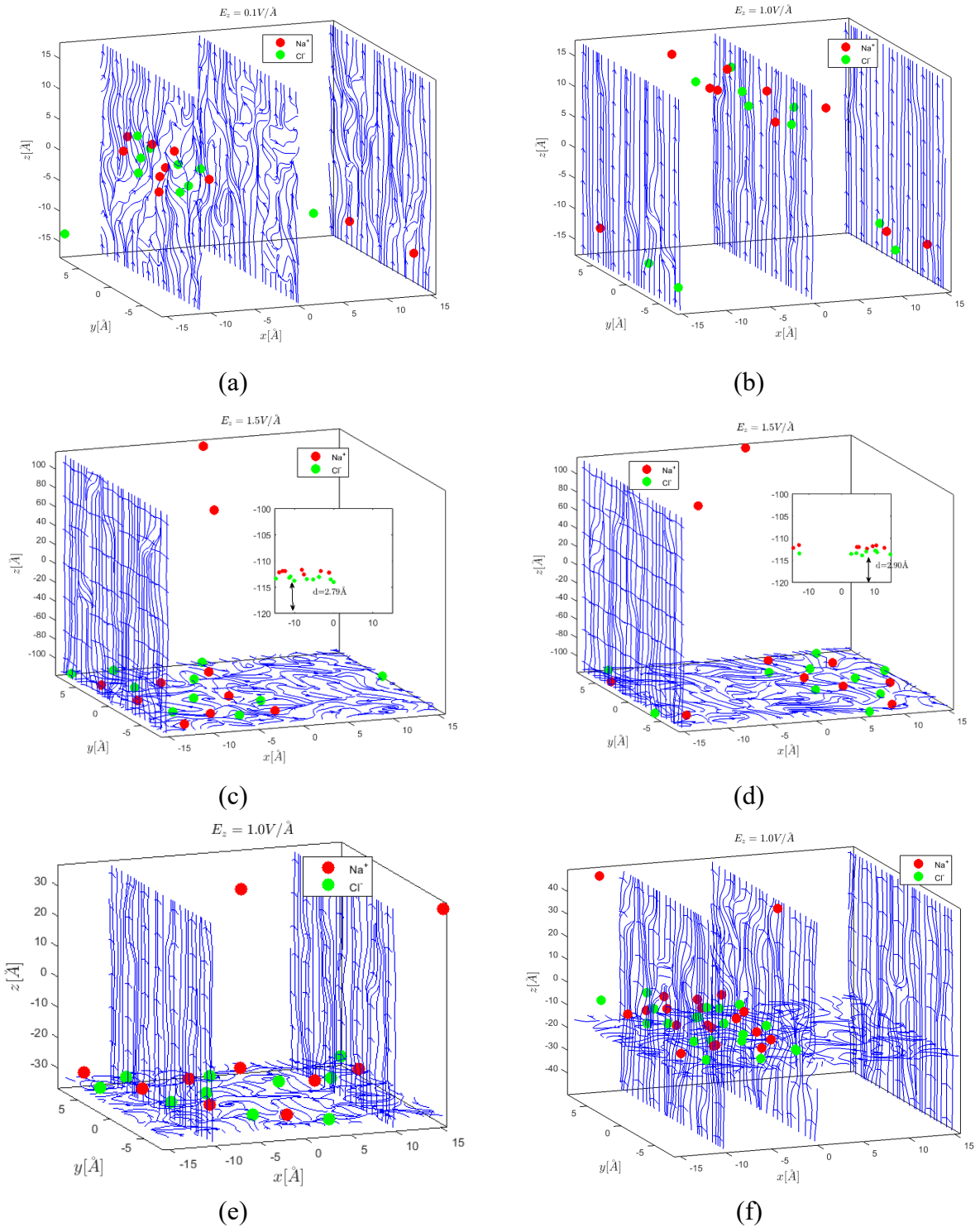


Figure 6. Ion drift in water solution, channel size for (a) h = 3 nm, $\varepsilon_{wf}/\varepsilon_{ff} = 0.1$, E_z = 0.1 V/Å, (b) h = 3 nm, $\varepsilon_{wf}/\varepsilon_{ff} = 0.1$, E_z = 1.0 V/Å, (c) h = 21 nm, E_z = 1.5 V/Å, $\varepsilon_{wf}/\varepsilon_{ff} = 0.1$, (d) h = 21 nm, E_z = 1.5 V/Å, $\varepsilon_{wf}/\varepsilon_{ff} = 0.5$. Inset figures show how close ions approach the lower wall (d) parameter), (e) h = 6 nm, E_z = 1.0 V/Å, $\varepsilon_{wf}/\varepsilon_{ff} = 1.0$, and (f) h = 9 nm, E_z = 1.0 V/Å, $\varepsilon_{wf}/\varepsilon_{ff} = 1.0$.

3.3. Diffusion coefficients

Diffusion coefficient values for various values of h, E_z, and $\varepsilon_{wf}/\varepsilon_{ff}$ have been calculated according to Eq 7 with MD simulations and all values are given in Table 3. The MSD values are extracted from Eq 6 and the respective diagrams can be found in the Supplementary Information. To eliminate statistical errors, we have run five independent simulations (see Figure 1) and have taken D as the average value, with $\sigma(D)$ being the error margin.

Table 3. Diffusion coefficients (values are given in $10^9 \times m^2 s^{-1}$) for channels in the range $3\,nm < h < 21\,nm$, wall/fluid interactions $0.1 < \varepsilon_{wf}/\varepsilon_{ff} < 1.5$ and external electric field values $0.0 < E_z < 1.5\,V/A$. $\sigma(D)$ is the error margin calculated from MD simulations.

h	E	$\varepsilon_{wf}/\varepsilon_{ff}$	D	$\sigma(D)$	h	E	$\varepsilon_{wf}/\varepsilon_{ff}$	D	$\sigma(D)$
3.00	0.00	0.10	2.789	0.09	6.00	0.10	0.50	3.106	0.25
3.00	0.01	0.10	2.826	0.07	6.00	0.10	1.00	3.011	0.42
3.00	0.01	0.20	2.587	0.11	6.00	0.50	0.10	3.365	0.80
3.00	0.01	1.00	2.535	0.17	6.00	0.50	0.20	3.052	0.73
3.00	0.10	0.10	2.138	0.15	6.00	0.50	0.50	3.128	0.34
3.00	0.10	0.20	2.252	0.15	6.00	0.50	1.00	3.106	0.69
3.00	0.10	0.50	2.174	0.20	6.00	1.00	0.10	3.529	0.28
3.00	0.10	0.68	1.763	0.08	6.00	1.00	0.20	3.002	0.25
3.00	0.10	1.00	1.976	0.15	6.00	1.00	0.50	2.846	0.43
3.00	0.20	0.10	1.985	0.04	6.00	1.00	1.00	2.986	0.17
3.00	0.20	0.20	1.942	0.09	6.00	1.50	0.10	3.041	0.22
3.00	0.20	0.50	1.659	0.31	9.00	0.00	0.10	1.860	1.66
3.00	0.20	1.00	1.662	0.01	9.00	0.10	0.10	2.022	0.11
3.00	0.50	0.10	2.565	0.26	9.00	1.00	0.10	2.055	0.01
3.00	0.50	0.20	2.314	0.07	9.00	1.00	0.20	2.173	0.15
3.00	0.50	0.50	2.169	0.20	9.00	1.00	1.50	2.080	0.07
3.00	0.50	1.00	2.189	0.31	9.00	0.50	0.50	2.313	0.29
3.00	1.00	0.10	2.365	0.31	9.00	0.50	0.50	2.256	0.14
3.00	1.00	0.20	2.279	0.18	9.00	0.50	0.50	2.251	0.22
3.00	1.00	0.50	2.133	0.28	9.00	1.50	1.50	2.275	0.19
3.00	1.00	1.00	2.088	0.18	9.00	1.50	1.50	2.137	0.18
3.00	0.75	0.10	2.361	0.29	9.00	1.50	1.50	2.283	0.15
3.00	0.75	0.20	2.110	0.38	15.00	0.00	0.10	2.809	0.11
3.00	0.75	0.50	2.010	0.26	15.00	1.00	0.10	2.742	0.94
3.00	1.50	0.10	2.149	0.31	15.00	0.10	0.10	2.752	1.04
3.00	1.50	0.20	2.042	0.10	15.00	0.10	0.20	2.438	0.72
3.00	1.50	0.50	2.094	0.16	15.00	0.10	0.50	2.316	0.12
6.00	0.00	0.10	2.155	0.01	15.00	0.10	1.00	2.408	0.09
6.00	0.01	0.10	2.220	0.56	21.00	0.00	0.10	2.019	0.11
6.00	0.01	0.20	2.121	0.06	21.00	0.10	0.10	2.352	0.30
6.00	0.01	0.50	2.205	0.05	21.00	0.10	0.20	2.330	0.15
6.00	0.01	1.00	2.118	0.01	21.00	0.10	0.50	2.862	0.09
6.00	0.10	0.20	3.433	0.38	21.00	0.10	1.00	2.068	0.08

The bulk diffusion coefficient value for pure water at T = 300 K ranges from $D = 2.19 - 2.30 \times 10^{-9}\ m^2 s^{-1}$ [53], while in a confined channel h = 3 nm it spans as $D = 1.659 - 2.826 \times 10^{-9} m^2 s^{-1}$, for most E_z and $\varepsilon_{wf}/\varepsilon_{ff}$ values studied, which is, within statistical accuracy, close to the water bulk values, taking into the multifactorial environment. In general, an increasing E_z decreases the average D value. Under the same E_z value, D decreases monotonically as $\varepsilon_{wf}/\varepsilon_{ff}$ turns from strongly hydrophobic to strongly hydrophilic. For h = 6 nm, we have obtained greater statistical errors, and D spans as $D = 2.121 - 3.129 \times 10^{-9} m^2 s^{-1}$, while, for h > 6 nm, where wall effects are minimized, D is set around $2.2 - 2.7 \times 10^{-9} m^2 s^{-1}$. This means that calculated D values range around water bulk values as channel dimensions increase.

3.4. Machine learning predictions for diffusion coefficients

The four ML algorithms implemented in the context of this paper. i.e., DT, RF, GBR, and MLP, have been trained on data taken from MD simulations, as shown in Table 3, and suggest predictions with accuracy metrics presented in Figure 7a–d. The coefficient of determination. R^2, is shown for every algorithm. We observe that R^2 > 0.700 for all cases, and this shows more than adequate algorithm performance. The three algorithms that are based on decision trees (DT, RF, GBR) in Figure 7a–c have achieved higher R^2 compared to MLP, with DT achieving R^2 = 0.853.

The main outcome from the application of ML techniques to this investigation is that ML can pose as an alternative approach to obtaining properties in similar flow systems, complementary to simulations and/or experiments. The diffusion coefficient has been chosen as the candidate property. It is a fact that only scarce experimental data are available in the literature for diffusion coefficients, and, as far as we know, it is particularly difficult to perform experiments. Moreover, simulation data provided in the literature are most of the times calculated over specific conditions and lack in details that could be fed as input parameters to train an ML model, at least in the problem we are dealing with in this paper. In this direction, MD simulations seem as the proper to obtain diffusion data, and they can be boosted by employing ML prediction techniques.

The amount of data incorporated for training an ML algorithm is one of the main considerations for obtaining an accurate result. However, as our dataset is uniformly distributed between a short range of channel widths, and most indicative cases of wall wettability, and electric field values (at least, in cases where no chemical phenomena occur), we can achieve nice-fitting results. We also believe that the problem of overfitting in ML methods is anticipated here by employing multiple ML algorithms, that are based on different internal mechanisms. Furthermore, we did not acquire "perfectly" fit for any algorithm, which might have been an indication of overfitting.

Finally, to anticipate the fact that diffusion data presents high statistical errors in some cases, even after statistical averaging of their simulation values after 5 independent runs (see Figure 1 the need of exploiting such a two-fold approach arises

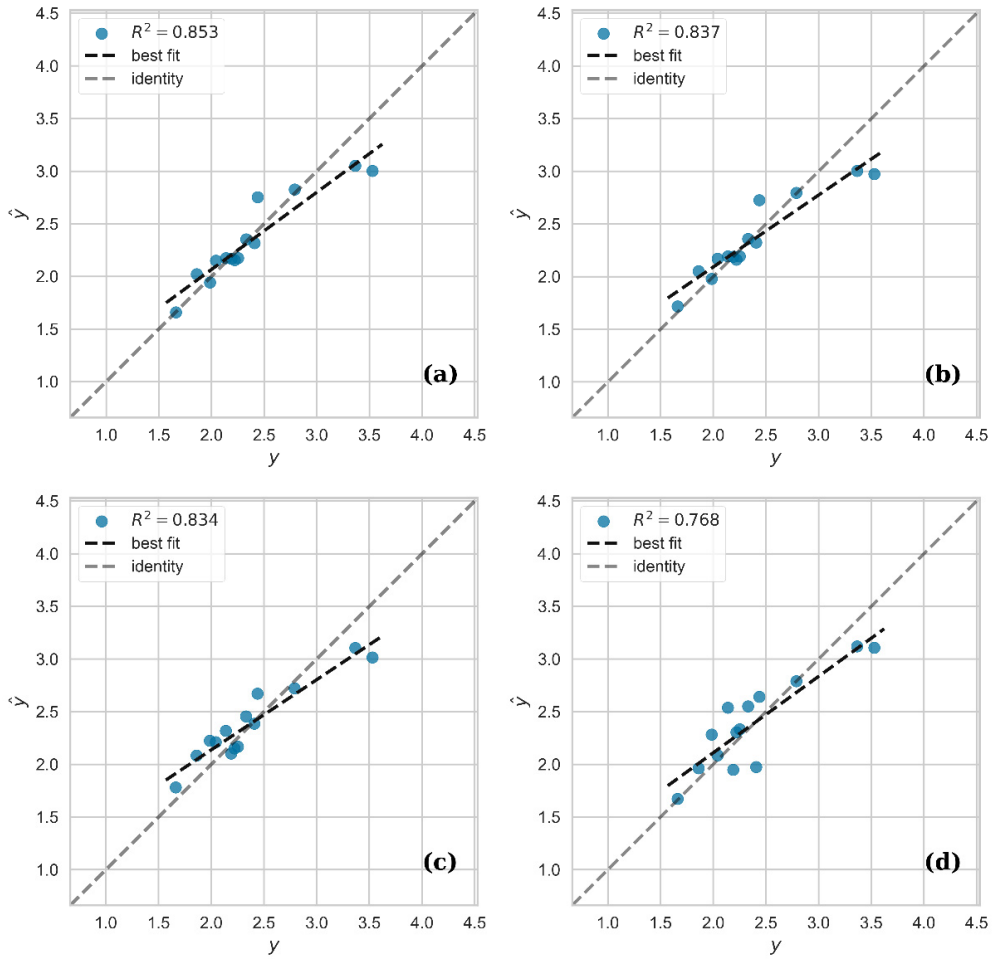


Figure 7. Simulation vs. predicted values for diffusion coefficient, after the incorporation of four ML algorithms. (a) decision-tree, (b) random forest, (c) gradient-boosting, and (d) multi-layer perceptron. Grey-dotted lines (45º) denote the perfect match, while black lines are the model regression lines.

4. Conclusions

This paper has tried to reveal the ion behavior mechanism within water solutions, when a number of factors are taken into account, such as the electric field strength, the wall wettability strength, and the channel width. Molecular Dynamics simulations have been incorporated to provide detailed potential maps inside the nanochannels, spot possible ion position in the solution, and calculate the diffusion coefficients of the solution.

There have been cases where Na^+ and Cl^- ions along with water molecules have been found to form clusters and/or hydrated ions, preventing ion rejection towards the walls. The increase of the electric field value is the main force that ensures ion drift to the walls, as it overcomes internal energy barriers that keep ions away from the walls. It has also been observed that in smaller nanochannels, the ratio of wall/fluid interaction may affect ion/water flow characteristics. In terms of diffusion coefficients, calculated values with MD simulations for a wide range of system conditions are shown here. These values are also exploited to train various ML models in order to construct a

computational prediction scheme that can act in parallel to simulations and use their output to maximize the span of the obtained results. Here, we have achieved good accuracy in the prediction of diffusion coefficient values with tree-based ML algorithms.

We envision that the proposed two-fold simulation scheme could be improved and incorporated in various materials science computational methods, especially for properties that are hard to obtain, such as shear viscosity, thermal conductivity, and electrical conductivity in nanofluidic devices.

Acknowledgments

F.S. and C.S. acknowledge support from the project CAMINOS, which is implemented in the context of a grant by the Center of Research Innovation and Excellence of U.Th., funded by the Special Account for Research Grants of U.Th. We also acknowledge computational time granted from the National Infrastructures for Research and Technology S.A. (GRNET S.A.) in the National HPC facility—ARIS.

Conflict of interest

All authors declare no competing financial interest.

References

1. Abraham J, Vasu KS, Williams CD, et al. (2017) Tunable sieving of ions using graphene oxide membranes. *Nat Nanotechnol* 12: 546–550. https://doi.org/10.1038/nnano.2017.21
2. Padmavathy N, Behera SS, Pathan S, et al. (2019) Interlocked graphene oxide provides narrow channels for effective water desalination through forward osmosis. *ACS Appl Mater Interfaces* 11: 7566–7575. https://doi.org/10.1021/acsami.8b20598
3. Yang T, Lin H, Loh KP, et al. (2019) Fundamental transport mechanisms and advancements of graphene oxide membranes for molecular separation. *Chem Mater* 31: 1829–1846. https://doi.org/10.1021/acs.chemmater.8b03820
4. Landon J, Gao X, Omosebi A, et al. (2019) Progress and outlook for capacitive deionization technology. *Curr Opin Chem Eng* 25: 1–8. https://doi.org/10.1016/j.coche.2019.06.006
5. Barbosa GD, Liu X, Bara JE, et al. (2021) High-salinity brine desalination with amine-based temperature swing solvent extraction: A molecular dynamics study. *J Mol Liq* 341: 117359. https://doi.org/10.1016/j.molliq.2021.117359
6. Mahmoud A, Nassef E, Salah H, et al. (2020) Use of hydrazide derivative of poly methylacrylate for the removal of cupric ions from solutions. *AIMS Mater Sci* 7: 420–430. https://doi.org/10.3934/matersci.2020.4.420
7. Yang F, He Y, Rosentsvit L, et al. (2021) Flow-electrode capacitive deionization: A review and new perspectives. *Water Res* 200: 117222. https://doi.org/10.1016/j.watres.2021.117222
8. Muscatello J, Jaeger F, Matar OK, et al. (2016) Optimizing water transport through graphene-based membranes: Insights from nonequilibrium molecular dynamics. *ACS Appl Mater Interfaces* 8: 12330–12336. https://doi.org/10.1021/acsami.5b12112
9. Cohen-Tanugi D, Lin L-C, Grossman JC (2016) Multilayer nanoporous graphene membranes for water desalination. *Nano Lett* 16: 1027–1033. https://doi.org/10.1021/acs.nanolett.5b04089

10. Giri AK, Cordeiro MNDS (2021) Heavy metal ion separation from industrial wastewater using stacked graphene Membranes: A molecular dynamics simulation study. *J Mol Liq* 338: 116688. https://doi.org/10.1016/j.molliq.2021.116688
11. Yu Y, Tan R, Ding H (2020) Controlling ion transport in a C_2N-based nanochannel with tunable interlayer spacing. *Phys Chem Chem Phys* 22: 16855–16861. https://doi.org/10.1039/D0CP02993A
12. Shao C, Zhao Y, Qu L (2020) Tunable graphene systems for water desalination. *ChemNanoMat* 6: 1028–1048. https://doi.org/10.1002/cnma.202000041
13. Abdullah N, Yusof N, Ismail AF, et al. (2021) Insights into metal-organic frameworks-integrated membranes for desalination process: A review. *Desalination* 500: 114867. https://doi.org/10.1016/j.desal.2020.114867
14. Presumido PH, Primo A, Vilar VJP, et al. (2021) Large area continuous multilayer graphene membrane for water desalination. *Chem Eng J* 413: 127510. https://doi.org/10.1016/j.cej.2020.127510
15. Hinds BJ, Chopra N, Rantell T, et al. (2004) Aligned multiwalled carbon nanotube membranes. *Science* 303: 62-65. https://doi.org/10.1126/science.1092048
16. Agrawal KV, Shimizu S, Drahushuk LW, et al. (2016) Observation of extreme phase transition temperatures of water confined inside isolated carbon nanotubes. *Nat Nanotechnol* 12: 267–273. https://doi.org/10.1038/nnano.2016.254
17. Hou D, Qiao G, Wang P (2021) Molecular dynamics study on water and ions transport mechanism in nanometer channel of 13X zeolite. *Chem Eng J* 420: 129975. https://doi.org/10.1016/j.cej.2021.129975
18. Liu Y, Cheng Z, Song M, et al. (2021) Molecular dynamics simulation-directed rational design of nanoporous graphitic carbon nitride membranes for water desalination. *J Membrane Sci* 620: 118869. https://doi.org/10.1016/j.memsci.2020.118869
19. Zhao Y, Huang D, Su J, et al. (2020) Coupled transport of water and ions through graphene nanochannels. *J Phys Chem C* 124: 17320–17330. https://doi.org/10.1021/acs.jpcc.0c04158
20. Chen L, Wang SY, Xiang X, et al. (2020) Mechanism of surface nanostructure changing wettability: A molecular dynamics simulation. *Comput Mater Sci* 171: 109223. https://doi.org/10.1016/j.commatsci.2019.109223
21. Mahmood A, Chen S, Chen L, et al. (2020) Spontaneous propulsion of a water nanodroplet induced by a wettability gradient: A molecular dynamics simulation study. *Phys Chem Chem Phys* 22: 4805-4814. https://doi.org/10.1039/C9CP06718C
22. Ranathunga DTS, Shamir A, Dai X, et al. (2020) Molecular dynamics simulations of water condensation on surfaces with tunable wettability. *Langmuir* 36: 7383-7391. https://doi.org/10.1021/acs.langmuir.0c00915
23. De Luca S, Todd BD, Hansen JS, et al. (2013) Electropumping of water with rotating electric fields. *J Chem Phys* 138: 154712. https://doi.org/10.1063/1.4801033
24. Kazemi AS, Nataj ZE, Abdi Y, et al. (2021) Tuning wettability and surface order of MWCNTs by functionalization for water desalination. *Desalination* 508: 115049. https://doi.org/10.1016/j.desal.2021.115049
25. Giri AK, Teixeira F, Cordeiro MNDS (2019) Salt separation from water using graphene oxide nanochannels: A molecular dynamics simulation study. *Desalination* 460: 1–14. https://doi.org/10.1016/j.desal.2019.02.014

26. Zong D, Yang Z, Duan Y (2017) Wettability of a nano-droplet in an electric field: A molecular dynamics study. *Appl Therm Eng* 122: 71–79. https://doi.org/10.1016/j.applthermaleng.2017.04.064
27. Bruus H (2008) *Theoretical Microfluidics*, Oxford, New York: Oxford University Press.
28. Bartzis V, Sarris IE (2020) A theoretical model for salt ion drift due to electric field suitable to seawater desalination. *Desalination* 473: 114163. https://doi.org/10.1016/j.desal.2019.114163
29. Bartzis V, Ninos G, Sarris IE (2022) Water purification from heavy metals due to electric field ion drift. *Water* 14: 2372. https://doi.org/10.3390/w14152372
30. Sofos F (2021) A water/ion separation device: Theoretical and numerical investigation. *Appl Sci* 11: 8548. https://doi.org/10.3390/app11188548
31. Sofos F, Karakasidis T, Sarris IE (2020) Molecular dynamics simulations of ion drift in nanochannel water flow. *Nanomaterials* 10: 2373. https://doi.org/10.3390/nano10122373
32. Kandezi MK, Lakmehsari MS, Matta CF (2020) Electric field assisted desalination of water using B- and N-doped-graphene sheets: A non-equilibrium molecular dynamics study. *J Mol Liq* 302: 112574. https://doi.org/10.1016/j.molliq.2020.112574
33. Lynch CI, Rao S, Sansom MSP (2020) Water in nanopores and biological channels: A molecular simulation perspective. *Chem Rev* 120: 10298–10335. https://doi.org/10.1021/acs.chemrev.9b00830
34. Steinhauser MO (2017) Multiscale modeling, coarse-graining and shock wave computer simulations in materials science. *AIMS Mater Sci* 4: 1319–1357. https://doi.org/10.3934/matersci.2017.6.1319
35. Huang DM, Cottin-Bizonne C, Ybert C, et al. (2008) Aqueous electrolytes near hydrophobic surfaces: Dynamic effects of ion specificity and hydrodynamic slip. *Langmuir* 24: 1442–1450. https://doi.org/10.1021/la7021787
36. Bonthuis DJ, Horinek D, Bocquet L, et al. (2009) Electrohydraulic power conversion in planar nanochannels. *Phys Rev Lett* 103: 144503. https://doi.org/10.1103/PhysRevLett.103.144503
37. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys* 117: 1–19. https://doi.org/10.1006/jcph.1995.1039
38. Karniadakis G, Beşkök A, Aluru NR (2005) *Microflows and Nanoflows: Fundamentals and Simulation*, Springer.
39. Dimiduk DM, Holm EA, Niezgoda SR (2018) Perspectives on the impact of machine learning. deep learning, and artificial intelligence on materials, processes, and structures engineering. *Integr Mater Manuf I* 7: 157–172. https://doi.org/10.1007/s40192-018-0117-8
40. Sofos F, Stavrogiannis C, Exarchou-Kouveli KK, et al. (2022) Current trends in fluid research in the era of artificial intelligence: A review. *Fluids* 7: 116. https://doi.org/10.3390/fluids7030116
41. Abbaspour M, Akbarzadeh H, Jorabchi MN, et al. (2022) Investigation of doped carbon nanotubes on desalination process using molecular dynamics simulations. *J Mol Liq* 348: 118040. https://doi.org/10.1016/j.molliq.2021.118040
42. Voronov RS, Papavassiliou DV, Lee LL (2006) Boundary slip and wetting properties of interfaces: Correlation of the contact angle with the slip length. *J Chem Phys* 124: 204701. https://doi.org/10.1063/1.2194019
43. Ibrar I, Yadav S, Braytee A, et al. (2022) Evaluation of machine learning algorithms to predict internal concentration polarization in forward osmosis. *J Membrane Sci* 646: 120257. https://doi.org/10.1016/j.memsci.2022.120257

44. Odabaşı Ç, Dologlu P, Gülmez F, et al. (2022) Investigation of the factors affecting reverse osmosis membrane performance using machine-learning techniques. *Comput Chem Eng* 159: 107669. https://doi.org/10.1016/j.compchemeng.2022.107669
45. Salari K, Zarafshan P, Khashehchi M, et al. (2022) Modeling and predicting of water production by capacitive deionization method using artificial neural networks. *Desalination* 540: 115992. https://doi.org/10.1016/j.desal.2022.115992
46. Yin G, Alazzawi FJI, Bokov D, et al. (2022) Multiple machine learning models for prediction of CO_2 solubility in potassium and sodium based amino acid salt solutions. *Arab J Chem* 15: 103608. https://doi.org/10.1016/j.arabjc.2021.103608
47. Sofos F, Karakasidis TE, Liakopoulos A (2012) Surface wettability effects on flow in rough wall nanochannels. *Microfluid Nanofluid* 12: 25–31. https://doi.org/10.1007/s10404-011-0845-y
48. Jiang H, Müller-Plathe F, Panagiotopoulos AZ (2017) Contact angles from Young's equation in molecular dynamics simulations. *J Chem Phys* 147: 84708. https://doi.org/10.1063/1.4994088
49. Jorgensen WL, Chandrasekhar J, Madura JD, et al. (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935. https://doi.org/10.1063/1.445869
50. Bagheri M, Akbari A, Mirbagheri SA (2019) Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: A critical review. *Process Saf Environ* 123: 229–252. https://doi.org/10.1016/j.psep.2019.01.013
51. Behnam P, Faegh M, Khiadani M (2022) A review on state-of-the-art applications of data-driven methods in desalination systems. *Desalination* 532: 115744. https://doi.org/10.1016/j.desal.2022.115744
52. Bratko D, Daub CD, Leung K, et al. (2007) Effect of field direction on electrowetting in a nanopore. *J Am Chem Soc* 129: 2504–2510. https://doi.org/10.1021/ja0659370
53. Yeo CSH, Xie Q, Wang X, et al. (2020) Understanding and optimization of thin film nanocomposite membranes for reverse osmosis with machine learning. *J Membrane Sci* 606: 118135. https://doi.org/10.1016/j.memsci.2020.118135

AIMS Press