Physics-informed deep learning for multi-species membrane separations

Physics-informed deep learning for multi-species membrane separations

Danyal Rehman[a,b], John H. Lienhard[a,*]

[a]*Rohsenow Kendall Heat Transfer Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, U.S.A*
[b]*Center for Computational Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, U.S.A*

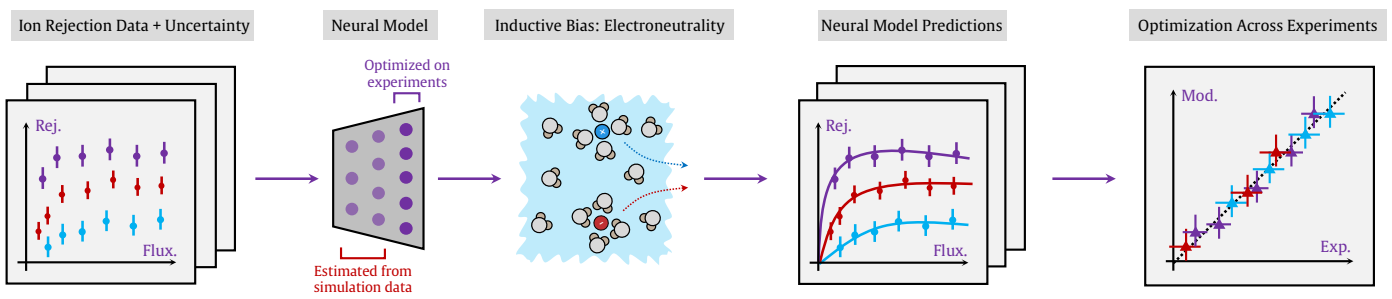
Abstract

Conventional continuum models for ion transport across polyamide membranes require solving partial differential equations (PDEs). These models typically introduce a host of assumptions and simplifications to improve the computational tractability of existing solvers. As a consequence of these constraints, conventional models struggle to generalize predictive performance to new unseen conditions. Deep learning has recently shown promise in alleviating many of these concerns, making it a promising avenue for surrogate models that can replace conventional PDE-based approaches. In this work, we develop a physics-informed deep learning model to predict ion transport across diverse membrane types. The proposed architecture leverages neural differential equations in conjunction with classical closure models as inductive biases directly encoded into the neural framework. The neural methods are pre-trained on simulated data from continuum models and fine-tuned on independent experiments to learn multi-ionic rejection behaviour. We also harness the attention mechanism, commonly observed in language modelling, to learn and infer paired transport relationships. Gaussian noise augmentations from experimental uncertainty estimates are also introduced into the measured data to improve robustness and generalization. We study the neural framework's performance relative to conventional PDE-based methods, and also compare the use of hard/soft inductive bias constraints on prediction accuracy. Lastly, we compare our approach to other competitive deep learning architectures and illustrate strong agreement with experimental measurements across all studied datasets.

Keywords: ion selectivity, membrane separations, physics-informed machine learning, scientific machine learning, deep learning

Number of pages: 19. Number of figures: 6. Number of tables: 5.

Graphical Abstract



*Corresponding author.
Email address: lienhard@mit.edu

1. Introduction and Background

Highly-selective membranes are ubiquitously used across the separations industry, where they play an essential role in the recovery and concentration of valuable metals like lithium and cobalt [1–3]. With the advent of the rapidly growing electric vehicle industry, the demand for these critical metals is expected to increase, necessitating robust solutions that perform well at these immense scales [4, 5]. To meet this burgeoning demand, further optimization of the selectivity and energy efficiency of membrane-based systems across diverse sourcewaters holds substantial industrial interest [6, 7]. However, since building and testing all possible permutations and combinations of these systems is cost-prohibitive, computational models are frequently used to estimate and optimize the performance of larger-scale systems [8–10].

For critical metals recovery, one membrane-based approach of emerging interest is nanofiltration (NF) [5, 11–13]. NF relies on a combination of steric, dielectric, and Donnan exclusion mechanisms to induce separation and metal ion recovery [14]. These mechanisms enable the technology to achieve both size- and charge- based separation of ions across diverse mixtures [15]. NF also typically operates at relatively low pressures, enabling systems to achieve competitive ion selectivities with reduced energy requirements [16]. To optimize NF performance, models that accurately predict and generalize rejection behaviour are essential [17–19].

The first models for NF were derived from irreversible thermodynamics [20, 21]. These frameworks treated membranes like a black-box, which intrinsically neglected the coupled relationship between ion selectivity and membrane properties. Years later, to address this issue, the Donnan–Steric Pore Model (DSPM) was proposed by Bowen and Mukhtar, which leveraged a combination of the Nernst–Planck partial differential equations (PDEs) and hindered transport theory to model transport [22, 23]. The model showed moderate agreement with experiments; a large part of this was because the approach inherently neglected one of the fundamental selectivity mechanisms: dielectric exclusion. To address this, the Donnan–Steric Pore Model with Dielectric Exclusion (DSPM–DE) was proposed in 2002, which introduced the Born model to quantify the effects of ion solvation into nanoporous membranes [24].

Since the development of DSPM–DE, it has become one of the most frequently used models for NF. It has also seen many iterations, most with the objective of addressing its many simplifying assumptions [18]. For example, Bowen and Welfoot considered integrating the effects of pore size distributions into the framework to be more representative of typical membrane morphologies [25]. Yaroshchuk studied the effects of fictitious image forces to improve the predictions associated with dielectric exclusion [26]. Silva et al. investigated the variation of the membrane charge density streamwise of the membrane pores to better approximate the dependence of the charge density on solution composition [27]. However, despite these improvements, DSPM–DE, as well as other continuum-based approaches, have been seen to struggle with generalization performance across different sourcewaters [28]. In other words, membrane parameters regressed from experimental data for a *given* membrane interpolate well, but often extrapolate poorly to new compositions [18]. Given the rapid growth and versatility of deep learning-based methods across the natural sciences [29–31], these approaches may offer a promising avenue to bridge the aforementioned gaps and alleviate many of the issues that plague continuum models today.

Machine learning (ML)-based methods are being increasingly used for many green chemistry and ion separations technologies [32–34]. For NF specifically, interest is rapidly growing in deep learning techniques for the characterization of ion transport [35, 36]. Bowen et al. attempted to model rejection behaviour across NF membranes with feedforward neural networks using a combination of literature data and in-house measurements [37]. Their studies prioritized the design of a neural approach capable of interpolating well. Although successful as used, the method is unable to generalize performance to unseen compositions. Yangali–Quintanilla et al. also considered feedforward neural networks to model transport across polyamide membranes for NF and reverse osmosis (RO) [38]. Their work centered around predicting the rejection of organic contaminants across a wide range of membrane types; the studies also adopted structure-activity relationships to perform dimensionality reduction on experimental data to facilitate the identification of a pertinent set of parameters governing transport [38]. Jeong et al. investigated the use of SHAP values – a game theory-derived explainability technique – to elucidate the role of learned features from neural models on ion separation in polyamide membranes [39]. Their results suggested that neural models are able to learn size- and charge- based exclusion characteristics.

Despite the ongoing interest in deep learning approaches to model transport, the development of a *generalizable* machine learning-based surrogate model for transport across selective membranes remains elusive. The primary impediment lies in gaining access to large amounts of data for training models [18]. Consequently, to develop neural methods in data-constrained settings, alternative strategies are needed. In these scenarios, hybrid deep learning methods that combine mechanistic methods with ML models and/or transfer learning approaches may hold the potential to unlock such approaches. One such instance is presented by Rall et al., who harnessed feedforward neural networks to investigate the interplay between layer-by-layer membrane fabrication conditions and ion separation [40, 41]. Their studies integrate transport models into neural methods by predicting ion rejection from learned charge densities and pore radii. The framework effectively guides neural approaches towards viable solutions, but the method still assumes mechanistic models can be used as ground truths.

In the present work, we mitigate the aforementioned concerns with a generalizable physics-informed deep learning model that captures multi-ionic transport across selective membranes. The neural architecture is more expressive than conventional feedforward neural networks because it leverages neural differential equations [42] in conjunction with the attention mechanism [43] to learn smooth rejection profiles as a function of composition and water flux for diverse mixtures. In addition, we integrate charge conservation laws into the model using orthogonal projection-based inductive biases to improve predictions in data-limited settings. We also leverage a transfer learning strategy by conducting large-scale pre-training on DSPM–DE to improve the quality of intermediate embeddings (these offer better starting rejection estimates in data-constrained settings) [44]. Subsequently, we use experimental data comprising 800 rejection measurements to perform model fine-tuning. We illustrate the model's ability to achieve predictions within $\pm 10\%$ of all data in the test set outperforming conventional continuum models (explicitly shown for DuPont's NF270). We also elucidate the attention mechanism's role in identifying key paired ion transport relationships. Finally, we benchmark our approach against other alternative deep learning methods and mechanistic models: DSPM–DE, extended

DSPM–DE [5], and the solution-friction model [45]. Through this work, we demonstrate the promise of deep learning-based surrogates over conventional models across a diverse range of input conditions.

2. Deep Learning Model

2.1. Neural Differential Equations

The continuous dynamics of the hidden layers, $\mathbf{h}(\cdot)$, capture the change in ion-specific concentrations across the membrane as a function of permeate flux, J_v. As a consequence of the smooth rejection profiles predicted, the hidden layer dynamics are well-suited to being parameterized by a first-order ordinary differential equation (ODE):

$$\frac{d\mathbf{h}(J_v)}{dJ_v} = \text{ODENet}_\theta(\mathbf{h}(J_v), J_v; \theta) \tag{1}$$

for $J_v = \{0 \dots \mathcal{J}_v\}$, and $\mathbf{h} \in \mathbb{R}^d$, where dimension d denotes the maximum number of ionic species present across all experimental datasets[1]. For the studies conducted in this work, $d = 8$. Additionally, ODENet_θ: $[0, \mathcal{J}_v] \times \mathbb{R}^d \to \mathbb{R}^d$. To account for mixtures with different ions in the training, validation, and test datasets, masking is applied prior to being passed into the neural model. Additionally, $\theta \in \Theta$, where Θ represents some finite-dimensional, learnable parameter space [42]. By learning the derivative of the hidden layer output, rejection profiles are uniformly Lipschitz continuous in $\mathbf{h}(J_v)$ and continuous in J_v, enabling facile pre-training on conventional mechanistic models [46, 47].

In addition to masking, polynomial positional encodings are used. The embeddings are concatenated to the masked concentration vector prior to being passed into the neural model[2]. To integrate over the neural ODE, we adopt the Tsitouras 5(4) numerical method with a fixed step size of $\Delta = 0.1$ [48]. Backpropagating through the solver is performed using the continuous adjoint method, originally described by Chen et al. [42].

The network is comprised of five linear layers and $\tanh(\cdot)$ nonlinearities applied to each output (as presented in Fig. 1). Following the last linear layer, no point-wise activations are used (ODENet performance can often be detrimentally impacted through non-linear activations after the final hidden layer). The network is trained using Adam with a batch size of eight and an initial learning rate of 10^{-3} [49]. The learning rate is halved every 100 epochs during both pre-training and fine-tuning processes (additional details are provided in Section 2.4). For all experiments conducted, we evaluate the hidden state dynamics and their derivatives on the GPU using PyTorch [50, 51] for a total of 500 epochs. All studies are performed using NVIDIA M4000 GPUs.

2.2. Attention Mechanism

Transformer networks, typically used for modelling natural language, adopt the concept of attention to provide machine learning models with a mechanism to interpret semantic context in sentences [43]. The attention mechanism has subsequently been translated to other disciplines to enable deep learning methods to better understand context-dependent pairing relationships; these have been demonstrated in molecular models with graph attention networks [52], all the way to attention-guided molecular generative models like the transformer-based variational autoencoder (VAE) [53, 54]. In this work, we introduce an attention head to equip the model with a mechanism to learn paired ion transport relationships (also what we refer to as *ionic context*). We use a slightly modified form of multiplicative (dot-product) attention, in line with the approach presented by Vaswani et al. [43]:

$$\text{Attention}(Q, K, V) \triangleq \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{2}$$

where, Q, K, and V are the query, key, and value matrices, obtained from $W_Q^\top \in \mathbb{R}^{d_k}$, $W_K^\top \in \mathbb{R}^{d_k}$, and $W_V^\top \in \mathbb{R}^{d_k}$, respectively. In the reported work, we set $d_k = 8$. d_k corresponds to the dimension of the key matrix, which serves to normalize the attention computation, providing better empirical performance [43, 55]. This is also done to reduce the possibility of overflow, which can lead to undesirable vanishing gradients.

2.3. Conservation Laws as Hard Inductive Biases

Electroneutrality is a conservation law commonly prescribed in classical mechanistic models for nanofiltration. Electroneutrality requires that a solution's net charge remain zero under equilibrium conditions [2]. Within polyamide membranes, local electroneutrality can

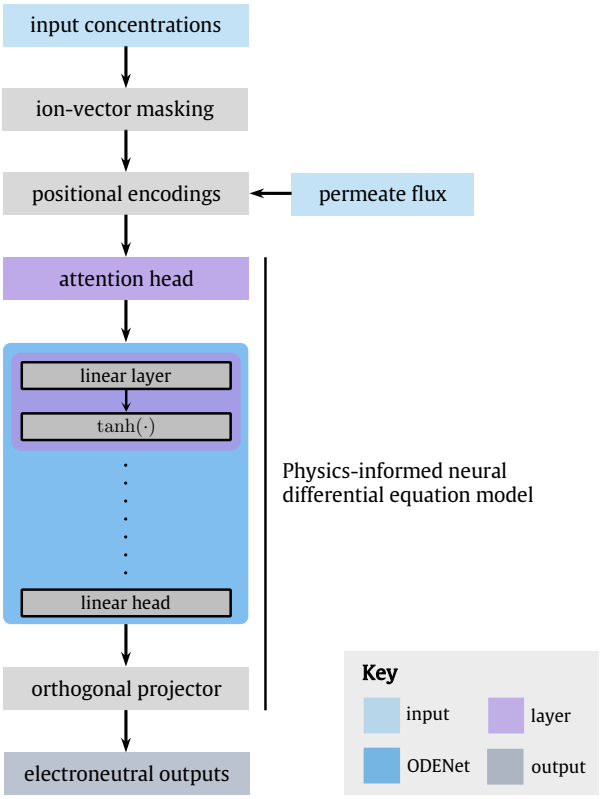


Figure 1: Proposed physics-informed neural differential equation architecture. Initial ion concentrations are masked and combined with positional encodings from the corresponding permeate flux measurements. These values are fed as input into the neural model, which predicts output concentrations. The Tsitouras 5(4) is used to integrate over ODENet_θ. The orthogonality projector subsequently converts permeate concentrations into electroneutral output predictions, which are used to evaluate the ion-specific model rejection, $\mathfrak{R}_j^{\text{mod}}(J_v)$.

[1] Since the magnitude of typical permeate fluxes is on the order of 10^{-5}-10^{-6} m/s, this can lead to numerical stability issues. Consequently, we normalize the flux by the largest observed water flux from the training, validation, and test set, ensuring the model only sees fluxes bounded between 0 and 1.

[2] Although not essential for the proposed method to succeed, it was empirically observed that the positional encodings improved predictive performance.

break down [56]; however, in the bulk solution, $\forall J_v$, the constraint is expected to hold [14]. The conservation law, also shown in dot-product form, can be expressed as follows:

$$\sum_{j=1}^{d} z_j \mathbf{h}_j(J_v) = z^\top \mathbf{h} = 0 \tag{3}$$

where $z \in \mathbb{Z}^d$ is a vector comprised of ion valences. To encode electroneutrality into the neural model as a *hard* constraint, we seek the orthogonal projection/component of the hidden layer output as follows:

$$z^\top \mathbf{h}_\perp = z^\top \mathbf{h} - z^\top \mathbf{h}_\parallel, \qquad \mathbf{h}_\perp = \mathbf{h} - \frac{\langle z, \mathbf{h} \rangle}{\langle z, z \rangle} z \tag{4}$$

By using $\mathbf{h}_\perp$ instead of appending the inductive bias to the loss as a soft constraint, the model guarantees inter-ionic coupling between ions from electroneutrality irrespective of the inputs, substantially improving generalization performance. Here, $\mathbf{h}_\parallel$ is evaluated using the projection operator: $\mathbf{h}_\parallel = \text{proj}_z(\mathbf{h})$. During ablation studies, we test both hard and soft constraints to illustrate the improvement in generalization performance. The soft loss formulation is detailed in Appendix A.

2.4. Transfer Learning Approach and Training Regimes

To improve predictive performance in the data-constrained regime (in which we are operating), we propose the use of transfer learning [57]. Here, we decompose the training process into two distinct stages: (**1**) pre-training on conventional PDE-based models; and (**2**) fine-tuning on experimental data from independent measurements comprising over 800 rejection data points. The reason behind doing so is twofold. Firstly, although 800 data points appears to be a significant amount of data, the distribution of ions across datasets is quite heterogenous, and the likelihood that the model sees out-of-distribution compositions at test time is high [58]; as a result, despite what appears to be a substantial amount of data, on its own is insufficient (we demonstrate this clearly in Section 3.2). Secondly, by using synthetic data from these simulations, we can substantially improve the quality of the model's learned embeddings. In other words, by using conventional PDE-based models to expose the neural approach to compositions likely to be seen during test, we can substantially improve predictive performance by narrowing the scope of feasible solutions. Consequently, we propose combining numerical and experimental efforts by freezing the weights of the first three layers and only updating the last two during fine-tuning. In doing so, we can leverage the moderate-quality embeddings learned from pre-training to yield strong starting estimates for learning in the low-data regime [59].

2.4.1. Pre-training on Synthetic Data from Continuum Models

To pre-train the neural method, we use simulated data generated from the well-established Donnan–Steric Pore Model with Dielectric Exclusion (DSPM–DE). In this model, multi-ionic transport through the polyamide membrane is expressed using the extended Nernst–Planck partial differential equations [60]:

$$J_j = -D_j K_{j,d} \partial_x C_j + K_{j,c} C_j J_v - \frac{K_{j,d} D_j C_j z_j F}{RT} \partial_x \psi, \quad x \in [0, \Delta x_e] \tag{5}$$

Here, J is the flux, with indices j and v used to denote ionic species j and water, respectively. D, F, R, and T are the diffusion coefficient, Faraday's constant, the universal gas constant, and absolute temperature, respectively. The dependent variables are C, the ion concentration, and ψ, the electric potential. The hindered form of the PDE is valid between $x = 0$ and $x = \Delta x_e$, where the latter term corresponds to the effective membrane thickness (the ratio of the membrane's thickness to its porosity). Lastly, $K_{j,c}$ and $K_{j,d}$ are the convective and diffusive hindrance factors for species j, respectively, both of which are evaluated using the correlations originally derived by Dechadilok and Deen [61]. Complete model details are provided in Appendix B.

In DSPM–DE, four latent variables are typically used to parameterize the membrane: $\mathcal{Z}_\ell \in \{r_p, \Delta x_e, \zeta_p, \chi_d\}$ [62]. These variables correspond to the membrane pore radius, effective thickness, dielectric constant in the pores, and volumetric charge density, respectively. Multiple regressions to novel sets of experimental data are most often performed to ascertain values of these parameters. To solve the extended Nernst–Planck equations, an iterative, under-relaxed numerical scheme is used (implementation details of the numerical approach are rigorously outlined in previous work by Geraldes and Brites Alves [60]).

The objective function used to ascertain the values of the membrane parameters is provided below:

$$\mathcal{Z}_\ell^* = \underset{r_p, \Delta x_e, \zeta_p, \chi_d}{\text{argmin}} \frac{1}{kd} \sum_{i=1}^{k} \sum_{j=1}^{d} \left(\frac{\mathfrak{R}_{ij}^{\text{mod}}(r_p, \Delta x_e, \zeta_p, \chi_d) - \mathfrak{R}_{ij}^{\text{exp}}}{\sigma_{ij}} \right)^2 \tag{6}$$

where σ_{ij} corresponds to the experimental uncertainty of each rejection measurement, and k is the total number of flux measurements taken per species. Using this formulation, less weight is attributed to experimental measurements that have a larger variance; conversely, the optimization focuses on minimizing the loss associated with points that we are more confident in. Additionally, i cycles through all experimental flux measurements, while j cycles through all ions in solution. Lastly, the asterisks, '*', denotes globally optimal parameter estimates, and the mean values used for pre-training are provided in Table 1 (as well as their 95% confidence intervals):

Table 1: DSPM–DE parameters and their 95% confidence intervals. The mean values were used for pre-training the physics-informed neural solver.

r_p^* [nm]	Δx_e^* [µm]	ζ_p^* [-]	χ_d^* [mol/m^3]
0.51 ± 0.04	1.27 ± 0.21	43.56 ± 4.15	-51.23 ± 11.54

To generate data for pre-training, we sample a d-dimensional initial concentration vector from the pre-training bounds reported in Fig. C.6 using low-discrepancy Sobol sequences and project them to log-space [63]. This was done to improve model predictions at lower concentrations [5]. A total of 42,183 points were generated. The process, although parallelized over GPUs to expedite data generation, still took just under four days to produce the full set of pre-training data used. The mean squared error (MSE) minimized during pre-training is expressed below:

$$\mathcal{L}^{\text{cm}}(\mathbf{h}, \mathbf{h}^{\text{cm}}) = \frac{1}{kd} \sum_{i=1}^{k} \sum_{j=1}^{d} \left[\mathbf{h}_j(J_{v,i}) - \mathbf{h}_j^{\text{cm}}(J_{v,i}) \right]^2 \tag{7}$$

where $k \in \mathcal{S}_\ell$ is the total number of flux measurements prescribed per species. Additionally, $\mathcal{S}_\ell = \{4, 5, 6, 7\}$, and $k \sim \mathcal{U}(\mathcal{S}_\ell)$, given that most training data typically had between 4-7 flux measurements. Subscripts i and j cycle through all flux measurements and ionic species, respectively. Selected pre-training ranges are detailed in Appendix C.

2.4.2. *Fine-tuning on Experimental Measurements*

Across all solutions in the studied datasets, ions included were: $\mathcal{S}^{\text{exp}} = \{Na^+, K^+, Li^+, Ca^{2+}, Mg^{2+}, Cl^-, NO_3^-, SO_4^{2-}\}$. Following pre-training, the network was fine-tuned on experimental data comprising 800 ion concentration measurements. The collected data spans studies conducted by numerous authors as well as in-house experiments [3, 19, 64–67]. All datasets analyzed the same polyamide membrane (DuPont's FilmTec™ NF270) for ion separation. The salinities considered were all at or below that of seawater, given that beyond these concentrations, DSPM–DE may not offer reasonable rejection predictions (DSPM–DE does not consider ion complexation, which has been seen to occur at elevated salinities; details of DSPM–DE's assumptions and simplifications are rigorously discussed in prior work by Wang and Lin [18]). Lastly, a training, validation, and test split of 600, 100, and 100 rejection measurements was prescribed.

To improve generalization performance, we also use measured uncertainties across the training data to fit Gaussian statistics to individual data points (μ, σ^2). To evaluate the loss function, values of output concentration were sampled from this distribution[3]. This type of data augmentation technique is commonly used in image settings, where small amounts of noise are added to each image to improve model robustness at inference time [70]. By adopting this approach, we can simulate the presence of 'more' data, improving generalization performance in data-constrained settings. The loss on the measurement data can be expressed as:

$$\mathcal{L}^{\text{exp}}(\mathbf{h}, \mathbf{h}^{\text{exp}}) = \frac{1}{n_f d} \sum_{i=1}^{n_f} \sum_{j=1}^{d} \left[\mathbf{h}_j(J_{v,i}) - \mathbf{h}_j^{\text{exp}}(J_{v,i}) \right]^2,$$
$$\mathbf{h}_j^{\text{exp}}(J_{v,i}) \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2) \; \forall i, j \quad (8)$$

Here, n_f is the total number of flux measurements taken, and serves the same purposes as k. In this case, n_f depends on the individual dataset used, as opposed to being a sampled quantity.

2.5. *Benchmarking against Alternative Deep Learning Models*

To benchmark the performance of our physics-informed ODENet relative to alternative deep learning architectures, we study Convolutional Neural Networks (CNNs) [71], Residual Networks (ResNets) [72], U-Nets [73], and conventional feedforward neural networks (FFNN). Hyperparameter tuning details are provided in Appendix D. The number of parameters used in each architecture was maintained relatively constant to ensure a fair comparison between methods, with methodological details summarized in Appendix E.

3. Results and Discussion

3.1. *Deep Learning Predictions and Data Requirements*

To evaluate the predictive performance of the proposed neural framework, we compare rejection predictions against conventional DSPM–DE on two samples from the test set (illustrated in Fig. 2A and Fig. 2B). During training, since Gaussian error estimates are fitted to experimental points to reflect measurement uncertainty, we are able to learn different models due to the intrinsic stochasticity of the sampling process [74]. Different learned models, as well as the one that minimizes validation error, are exemplified by the translucent shaded regions, and solid lines, respectively. In both Fig. 2A and Fig. 2B, we observe that the neural model outperforms DSPM–DE when a fixed set of membrane parameters is used[4]. This test illustrates generalization performance when new unseen data is presented to both the PDE-based model and neural approach (without having to perform new computationally-expensive, and experimentally-demanding regressions). In this case, the neural model is able to provide strong performance benefits relative to conventional continuum modelling approaches, which are known to struggle with generalizing performance to unseen conditions [14].

Although we use a fixed set of membrane parameters for the previous comparison, more commonly, a new set of DSPM–DE parameters is regressed from new batches of experimental data, with the model then used for interpolation purposes; however, even under these conditions, the assumptions and simplifications built into the model often prevent it from: (**1**) learning physically-representative membrane parameters (regressed pore radii can often be unreasonably small/large, or the effective membrane thicknesses can substantially underestimate/overestimate the true membrane thickness); or (**2**) regressing membrane parameters that are able to yield reasonable fits to the measured experiments [62]. As a result, we elucidate the neural model's superior predictive capabilities across solutions in the test set, even when conventional DSPM–DE is regressed to individual sets of data, as shown in the parity plot in Fig. 2C. Here, we observe that even when the continuum model is fitted to each set of data, it frequently fails to provide reasonable predictions for rejection, whereas the neural model has accuracy within $\pm 10\%$ of all measurements in the test set (despite never having seen the data *a priori*; the confidence bound is illustrated by the shaded envelope). In recent work by Jeong et al. on deep learning methods for ion transport across polyamide membranes, they note that despite operating in data-constrained settings, machine learning models can learn the importance of size- and charge-based exclusion [39]. Although their study demonstrates these findings through SHAP values, we show that we can draw the same conclusions by accurately predicting ion rejection behaviour across diverse mixtures.

In the inset of Fig. 2C, we present the training and validation losses across hyperparameter sweeps (we apply Bayesian optimization for hyperparameter tuning; details are provided in Appendix D). The descent curves presented only correspond to those obtained from the experimental fine-tuning step. In addition, we experimented with different amounts of pre-training and noted that typically more pre-training led to higher quality embeddings and improved predictive performance, albeit at the cost of increased computation runtime[5]. Conversely, with the experimental data, we noted that the model began to overfit to the rejection measurements after ~220 epochs (labelled by the dotted vertical line), beyond which the validation loss curves are seen to steadily increase in agreement with expectation.

[3] Hinge loss terms based on the Hofmeister series were originally included in the loss function but provided mixed results [68]. Given that ion rejection has been seen to diverge from the Hofmeister series under certain conditions, it was removed entirely from the loss [69].

[4] The chosen set of values was quantified by regressing DSPM–DE to measurement data from over 15 experimental studies on DuPont's NF270 membrane, all at pH 7 (experimental tests had a wide range of ions present, typically ranging from 2-8, all with salinities at or below that of seawater). Our previous approach, which harnesses hybrid global-local optimization with maximum likelihood estimation (MLE), was employed to quantify the latent membrane parameters across measurement data [62].

[5] There is a point beyond which model pre-training should overfit to the synthetic training data, however, we did not observe this limit during our training procedure given the high-dimensional solution space of DSPM–DE [75].

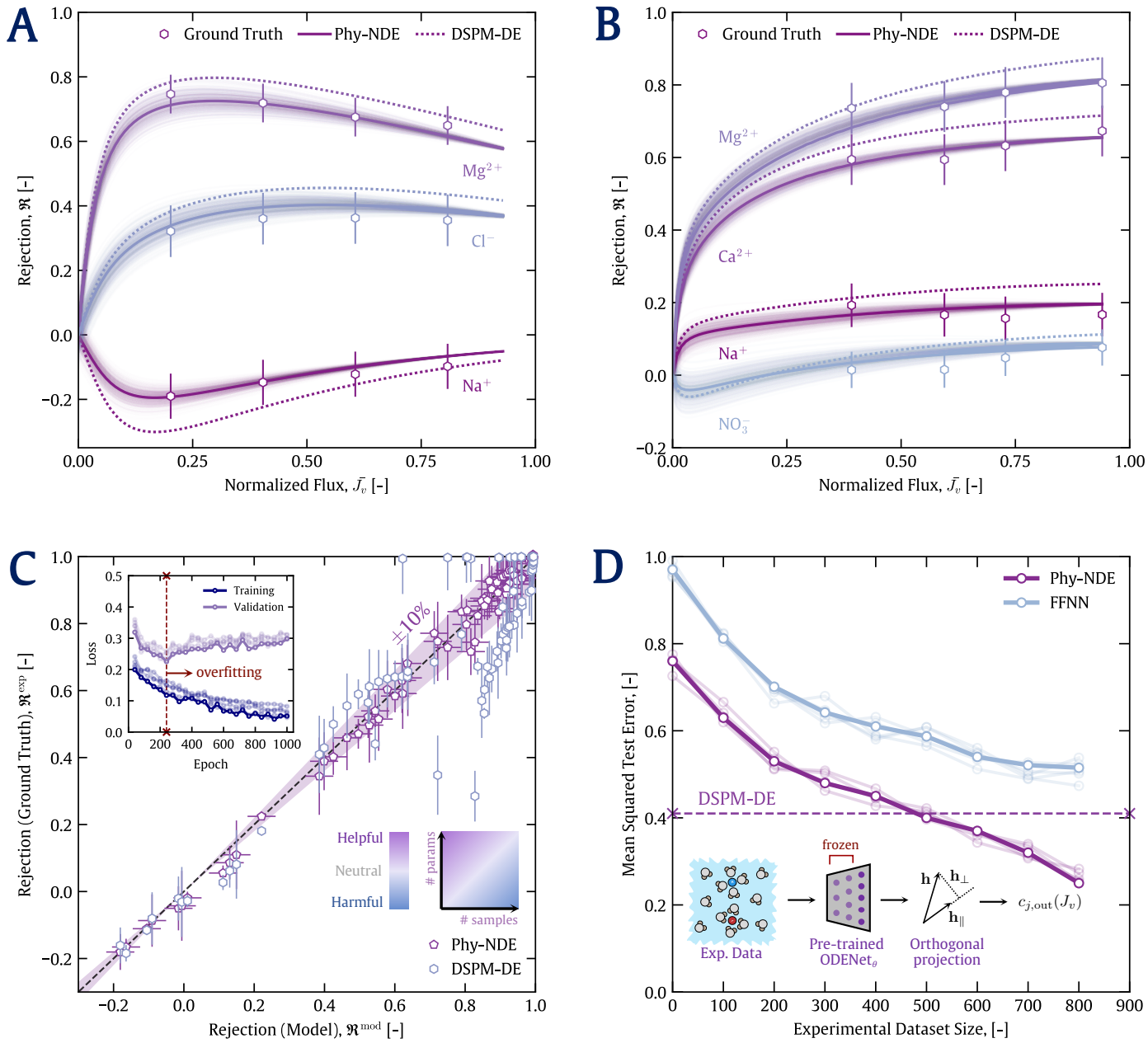


Figure 2: **A**) and **B**) Rejection against normalized flux (normalized by the largest flux measurement present in the test set) for a ternary-component and quaternary-component mixture with the physics-informed ODENet contrasted against conventional DSPM–DE, respectively. **C**) Performance of the proposed model benchmarked against data-tailored DSPM–DE across all mixtures in the test set illustrating the neural model's predictive capabilities. The inset includes the descent curves of the loss obtained on both training and validation sets. Overfitting to the training data occurs beyond ∼220 epochs, as observed by the increasing validation error. At this point model training is stopped. **D**) The change in the MSE on the test set as a function of the amount of experimental data used; here, we illustrate our method's ability to outperform conventional modelling approaches after ∼500 data points, in contrast to the feedforward neural network, which never achieves superior performance to DSPM–DE irrespective of the amount of data used.

Lastly, in Fig. 2D, we illustrate the amount of data needed to outperform tailored DSPM–DE using our proposed physics-informed neural solver and an optimized feedforward neural network. Here, we note that the error uncertainties stem from the sampling process, similarly to earlier simulations. We also observe that the proposed neural framework eventually outperforms DSPM–DE on the test data after ∼500 experimental rejection measurements. The descent slopes also appear to align with those arising from probably-approximately correct (PAC) learnability estimates [76–78] (these methods expect test error to decrease with $1/n$, where n is the amount of data used). Similar behaviour is noted with the feedforward network; however, we see that the conventional feedforward network is never able to outperform DSPM–DE irrespective of the number of experimental points used during fine-tuning. These simulations speak to the promise of our proposed neural framework in serving as an alternative to conventional continuum modelling approaches for ion transport across selective membranes.

3.2. Implications of Attention, Pre-training, and Inductive bias

The attention mechanism, most frequently used for modelling natural language in transformer networks [43], serves as a means for understanding the semantic context of words in a sentence. Here, we leverage the attention mechanism to learn the relative importance of ions in complex mixtures prior to being passed into the neural model, similarly to the way graph attention networks (GATs) leverage attention to model inter-molecular interactions [52]. In Fig. 3A, we benchmark the MSE achieved across a series of competitive deep learning models with and without its use. Our approach outperforms other ML-based alternatives with U-Nets achieving the closest MSE. This is largely attributed to the smooth profiles generated by ODENet (since the network learns rejection gradients rather than standalone rejection) for unseen fluxes that closely mirror experimental observation; other methods are unable to capture this continuity leading to inferior performance on the test data [42]. For all conducted tests, we maintained a similar number of model parameters across benchmarks to ensure a fair comparison (details of these architectures are in Appendix E). In all cases, we note that the inclusion of the attention layer improves predictive performance, with the reduction in predictive performance ranging between 5-20%, depending on the architecture tested.

Next, in Fig. 3B, we illustrate an example of the learned attention matrix for our set of studied ions. Here, we note that the attention mechanism clearly identifies the importance of valence and ionic coupling on transport. For example, we see that the attention given to both Cl^- and NO_3^- by Na^+ is high; this makes physical sense as SO_4^{2-} is often too large and immobile to be transported through the membrane meaning that Cl^- and NO_3^- are the primary anions carried across the membrane to conserve electroneutrality [79]. This is also clearly validated by the fact that the attention given to SO_4^{2-} appears low relative to the other monovalent anions[6]. This is similarly observed in Fig. 2B, where the model correctly predicts negative rejection values for NO_3^- at low fluxes, given that as the only anion in solution, its transport must be expedited through the membrane to ensure electroneutral permeate concentrations. These findings clearly illustrate the value of the attention mechanism in learning and characterizing ion transport across selective membranes.

Another observation in Fig. 3B is that Li^+ and Mg^{2+} are given substantial amounts of attention by ODENet. In other words, when both lithium and magnesium are present in the sourcewater, the model prescribes substantial weight to their relative quantities prior to predicting rejection behaviour. This also makes physical sense given that Li^+ and Mg^{2+} are challenging to separate given their similar solubility products and ionic radii [80, 81]. As a result, additional emphasis on their relative compositions is propagated through the model to ensure accurate predictions of Li^+ and Mg^{2+} rejection. This also aligns with uncertainty quantification studies using DSPM–DE that clearly highlight the elevated sensitivity of Li^+/Mg^{2+} selectivity towards changing feedwater composition and membrane parameters shown in our prior work [14].

Next, in Fig. 3C, we investigate the importance of pre-training on PDE-based models (PT) and including electroneutrality as a hard (HIB) vs. soft (SIB) constraint across competitive deep learning architectures. We note that by foregoing pre-training (NPT), the resultant MSE is at least 70% higher than when it is included. This trend is in agreement with expectation given that these models are being trained in highly data-limited regimes, where there is an insufficient amount of data to generalize well; as a result, transferring knowledge from continuum models to the neural architecture substantially boosts performance when the models experience new unseen compositions. Furthermore, we note that by neglecting pre-training on PDE-based models, it appears not to be possible to outperform conventional DSPM–DE irrespective of the architecture used. Consequently, it is essential to recognize that despite the shortcomings that PDE-based solution methods impose [18], using them to improve the quality of embeddings for deep learning-based alternatives is a critical step in achieving superior generalization performance.

In the case where pre-training is not performed (i.e., the data-limited setting), we found that the model has difficulty learning the importance of electroneutrality. In this case, although soft inductive biases encourage electroneutral outputs, they do not guarantee them [82, 83]. As a result, we see a trade-off emerge, where additional training on the experimental data can lead to the model becoming more likely to predict electroneutral permeate concentrations, yet at the cost of becoming increasingly likely to overfit to the training data, adversely impacting generalization performance. Consequently, when we enforce electroneutral outputs through the orthogonal projection (hard constraint), the improvement in performance is notable (a drop in MSE of at least 20% is seen across deep learning methods). Overall, however, despite the hard inductive bias, the result still fails to outperform conventional DSPM–DE in predicting rejection, necessitating an alternative solution.

Lastly, in Fig. 3C, we note that although there is a difference in predictive performance when using hard projection constraints relative to soft regularization terms (with pre-training), the differential error may not be as high as we may expect (as was the case when pre-training was foregone). The primary reason behind this is that the predictions from the continuum model are *already* electroneutral; as a result, by conducting judicious pre-training on DSPM–DE, we are already exposing the model to large quantities of training data that is already conditioned on electroneutral outputs. Consequently, even through the use of the soft regularization loss, we do a reasonable job at encouraging the model to prioritize electroneutral predictions. Similar to the previous case, however, using soft regularization terms does not guarantee electroneutral predictions for new unseen compositions; as a result, the optimal performance is observed when hard inductive biases are imposed. Overall, we observe the best performance using our proposed physics-informed ODENet with pre-training and hard inductive biases integrated into the architecture.

3.3. Benchmarking Performance against PDE-based Models

In this section, we conduct a more detailed assessment between the neural model and conventional PDE-based methods. Specifically, we study the neural method relative to three different continuum modelling approaches: (**1**) conventional DSPM–DE with the optimal set of regressed parameters for the given solution composition (termed DSPM–DE); (**2**) a modified version of DSPM–DE, where in addition to the four typical membrane parameters, we also regress ion-specific convective and diffusive hindrance factors (amounting to a total of $2n+4$ regression parameters for a solution composition with n ions; termed extended DSPM–DE); and lastly, (**3**) the solution-friction model, in which the Nernst–Planck PDEs are also used to model ion transport, except that the hindrance factors are mapped into a

[6] SO_4^{2-} typically has a limiting rejection above 90% when using DuPont's FilmTec™ NF270 at pH 7 [16].

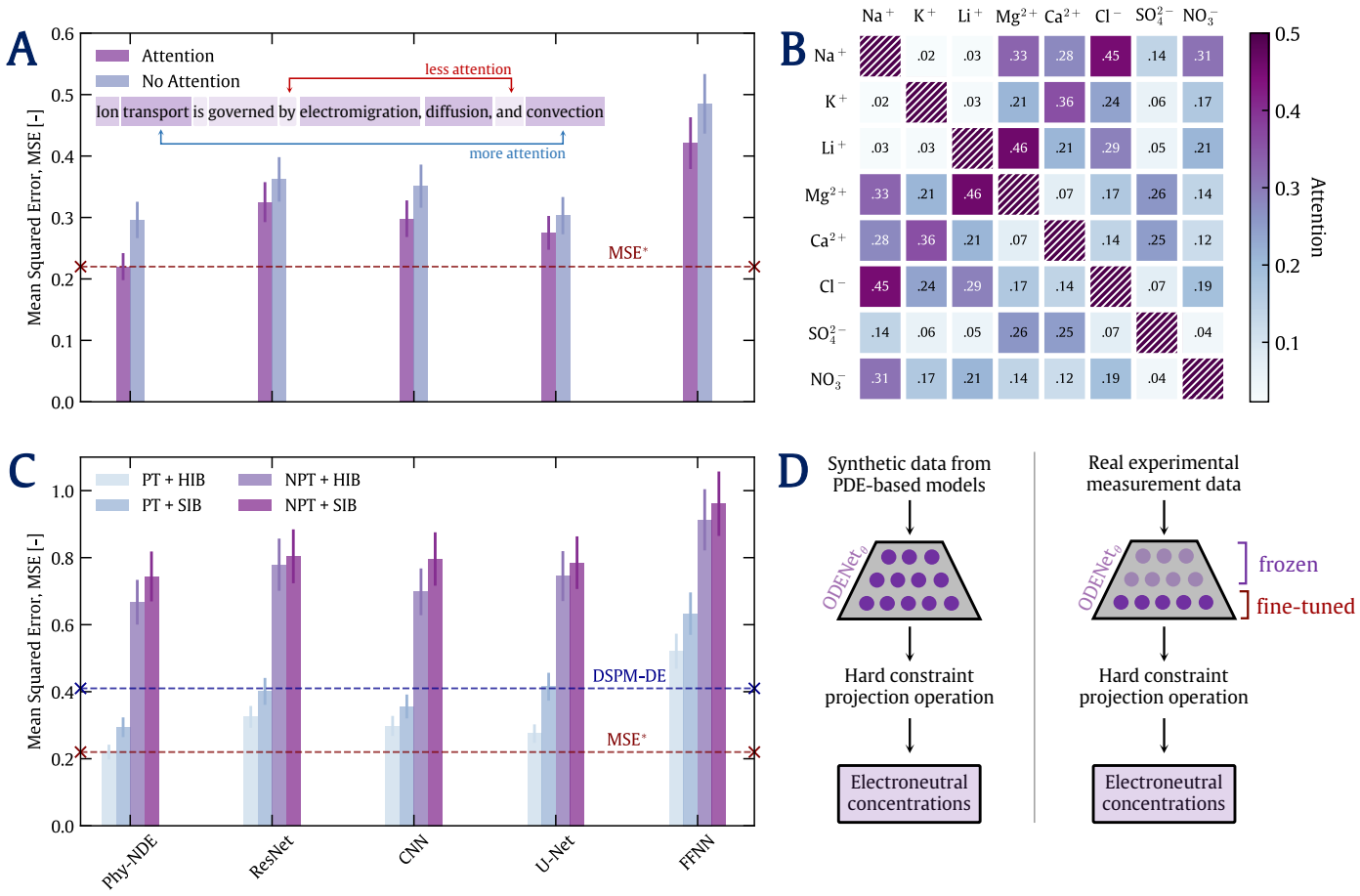


Figure 3: **A**) illustrates the performance benefits that arise from the attention mechanism across a series of competitive deep learning-based methods. Results suggest that including attention layers is beneficial across all studied ML approaches. **B**) provides an example of the attention matrix for a given solution composition illustrating the model's ability to learn the importance of both valence and solute size in predicting ion rejection. Heightened values of attention are attributed to ion pairs that can significantly contribute to one anothers' rejection predictions, as seen with charge conservation in Na^+ and Cl^-, for example. **C**) illustrates the performance benefits from pre-training on synthetic data and treating electroneutrality as a hard vs. soft constraint across the studied deep learning alternatives. **D**) A schematic drawing illustrating the transfer learning approach taken to improve the quality of learned embeddings using neural methods in the data-limited regime.

series of known friction factors. In solution-friction, these coefficients are frequently known for given ion species [84]. In addition to this modification, the Born and steric exclusion terms are combined into a size-dependent partition coefficient that is regressed for each ionic species j (this means that the dielectric constant in the membrane pores is no longer regressed, yielding a total of $n+3$ learnable parameters for solutions with n ions) [45][7]. Solution-friction model details are in Appendix G.

In Fig. 4A and Fig. 4B, we illustrate the residual error between conventional DSPM–DE, extended DSPM–DE, and the neural model for the two worst performing ions in a given composition from the test set[8]; for this example, these errors correspond to Na^+ (left) and Cl^- (right). The errors are defined as the difference in predictions between the method of interest and an exponential-fit curve. In this case, we note that conventional DSPM–DE provides the least optimal performance for both ions. More specifically, with Na^+, the initial errors are substantially larger than those from both extended DSPM–DE and the neural approach. In addition, as we transition to higher fluxes, we note that the residuals appear to plateau steadily at 0.04, indicating that the model clearly overpredicts sodium rejection across higher fluxes. The neural approach also overpredicts the data at lower fluxes, but then provides competitive predictive performance near the experimental data points. These findings elucidate the promise of neural methods for predicting ion rejection performance, despite having never seen the experimental data before. Lastly, we see the best-fit performance with extended DSPM–DE; however, it is important to note that by using $2n+4$ parameters to fit the rejection curves, the model results are likely an overfit, substantially impeding generalization performance. In other words, although the error can be minimized quite significantly on the regressed measurements, the model very likely overfits to the experimental data suggesting that the same set of learned parameters for other solutions may yield misleading findings (we prove these findings clearly in Fig. 4D).

In Fig. 4B, we note similar, but even more exaggerated findings with Cl^-. In this case, conventional DSPM–DE exhibits substantial deviations from the experimental data with prediction errors worsening with increasing flux. Similar to the case with Na^+, the

[7] In the conventional DSPM–DE model, hindrance factor expressions derived by Higdon and Muldowney [85], as well as, Mavrovouniotis and Brenner [86] are used. Additional details are provided in Appendix F.

[8] Note, to avoid overcrowding the figure, we exclude residual error predictions from the solution-friction model given that the results were nearly identical to those obtained from extended DSPM–DE.

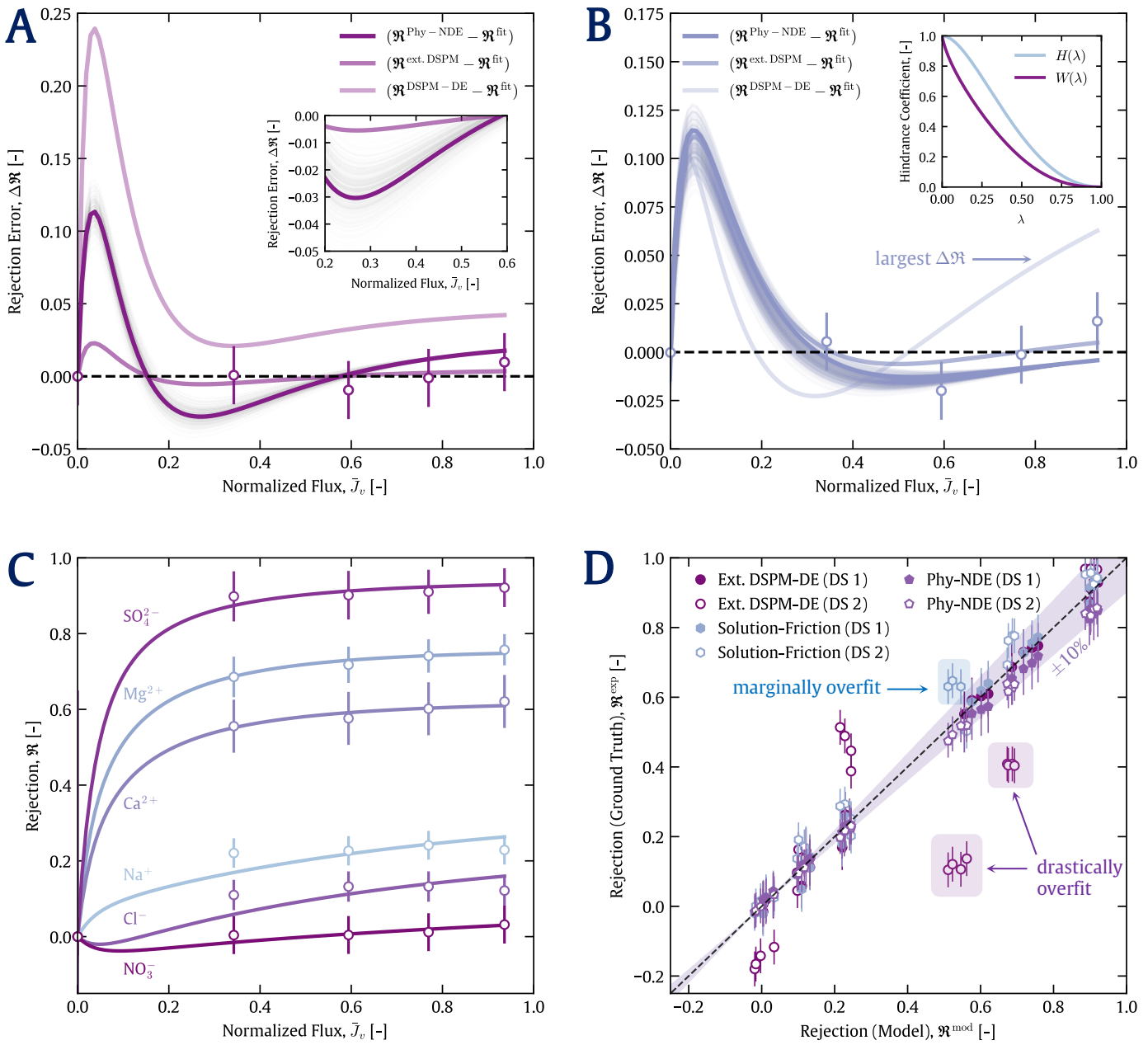


Figure 4: **A**) and **B**) illustrate the differences between the predicted rejection (using both continuum modelling approaches and the proposed neural method) and an exponential fit curve for the two worst performing ions (Na^+ and Cl^-). The predictions indicate that the neural model outperforms conventional DSPM–DE, while not meeting the prediction accuracies attained by extended DSPM–DE. The inset includes the convective and diffusive hindrance factor expressions used in DSPM–DE as a function of λ, the ratio of solutes' Stokes radius to the membrane pore radius. **C**) High-accuracy fits obtained from extended DSPM–DE, which are largely attributed to overfitting the model to experimental measurement data using the $2n+4$ fitting parameters present in the regression formulations. **D**) Generalization performance comparisons between extended DSPM–DE, the solution-friction model, and the neural solver using two datasets from the test set.

neural approach and the extended DSPM–DE model achieve strong performance. The over-parameterized argument previously applied to Na^+ also translates to Cl^-, given that with extended DSPM–DE, the solute flux is an entirely constrained function. More specifically, by learning the convective and diffusive hindrance factors, the model has complete control of the solute flux expression in the extended Nernst-Planck PDE. Furthermore, given that the solute flux and permeate flux are decoupled in the mathematical formulation, the model has substantial numerical freedom to predict ion rejection, with the only required constraint being charge conservation from electroneutrality. As a result, the likelihood that extended DSPM–DE overfits to the measurement data is relatively high. To illustrate the accuracy of the regression using extended DSPM–DE, we present rejection predictions for a six ion mixture in Fig. 4C. The regressed membrane parameters and hindrance factors are summarized for the analyzed sample composition in Appendix H.

Finally, in Fig. 4D, we elucidate the effects of over-parameterization and overfitting that can occur in conventional continuum models. Specifically, we can see that the agreement between extended DSPM–

DE and experimental data from dataset 1 to be near perfect (nearly all points lie on the parity line, in agreement with the observations from Fig. 4C); however, if we hold the regressed parameters fixed and run inference on dataset 2, we observe substantial deviation between predictions from the model and experiments. These are annotated in the figure with the label 'drastically overfit'. As stated previously, these findings are largely attributed to the fact that mathematically regressing hindrance factors gives the model significant dexterity, which can work effectively on a given dataset, but fail to generalize well to new concentrations and compositions.

When we consider the solution-friction model, we find similar performance on dataset 1, whereby the model does an excellent job at predicting ion rejections (points lie extremely close to the parity line, signifying excellent prediction accuracy). In this case, when we constrain the regressed parameters and run inference on dataset 2, we find that the model does not overfit to the same degree as extended DSPM–DE. We annotate points that lie outside our $\pm 10\%$ bounds, with a label stating 'marginally overfit' (since the deviations are nowhere near as significant as those obtained from extended DSPM–DE). As a result, we see that the solution-friction model generalizes more effectively than extended DSPM–DE. We also observe that the performance of the neural solver consistently achieves rejection predictions that are within $\pm 10\%$ for both dataset 1 and dataset 2. These studies elucidate the neural approach's efficacy in generalizing rejection predictions to new unseen concentrations and compositions, demonstrating the value of neural methods as alternatives to conventional PDE-based modelling approaches.

4. Implications and Summary

In this work, we present a physics-informed, attention-enhanced neural differential equation model pre-trained on synthetic data from conventional PDE-based models and fine-tuned on measurement data from across the literature and in-house experiments. We find that:

1. The neural approach outperforms conventional DSPM–DE across solutions in the test set, achieving accuracies within $\pm 10\%$ across the set of studied ions.
2. Attention layers, similar to those used in modelling natural language, play an important role in improving rejection predictions. Our studies suggest that including the attention mechanism consistently improves the model's ability to generalize: on average, we see that excluding the attention layer increases the MSE by nearly 20% across deep learning methods studied.
3. The learned attention matrices are also capable of identifying key paired transport relationships that govern ion transport phenomena across polyamide membranes. This is illustrated for a given solution composition in the test set.
4. In cases where pre-training is foregone, none of the deep learning-based methods are able to outperform conventional DSPM–DE. We attribute this to the data-limited regime in which we operate, where improving the scope of feasible solutions through transfer learning approaches ends up substantially improving predictive performance on the test set.
5. Including charge conservation-based inductive biases into the neural model consistently improves rejection predictions, whereby hard orthogonal projector constraints outperform soft regularization terms. In data-constrained regimes, the hard constraints offer substantial performance benefits to guarantee electroneutral predictions, in line with expectations.
6. When we contrast the neural approach to three different continuum models: DSPM–DE, extended DSPM–DE, and the solution-friction model, the neural approach achieves lower residual errors compared to conventional PDE-based models when hindrance factors are not regressed. In the event that convective and diffusive hindrance factors are fitted as well, the deep learning approach fails to yield a lower MSE. We observe that this is due to the over-parameterization of the extended DSPM–DE and solution-friction models; in other words, using the same set of regressed parameters for different mixtures can be seen to impede generalization performance.

Through this work, we illustrate the ability of neural methods to accurately predict ion separation across NF membranes at new unseen compositions and salinities. These findings open up the potential for further exploration of deep learning methods for ion transport across membranes. In particular, it would be interesting to see whether the proposed method pre-trained on alternative mechanistic models, could be useful in capturing ion transport across other selective separation systems, like selective electrodialysis (SED), or membrane capacitive deionization (SCDI). Alternatively, we would be curious to study whether the model trained on one membrane could translate its performance to a new membrane (with appropriate modifications introduced to account for membrane parameterization). Another potential avenue could be investigating the model's ability to predict ion transport at different temperatures, or at high salinity. By continuing to explore the emerging capabilities of machine learning methods for ion transport, there may be hope that deep learning-based solutions have the potential to entirely replace their PDE-based counterparts in modelling transport phenomena for diverse separations.

5. Acknowledgements

The authors thank the Centers for Mechanical Engineering Research and Education at MIT and SUSTech (MechERE Centers at MIT and SUSTech) for partially funding the research reported in this paper. Danyal Rehman also acknowledges financial support provided by a fellowship from the Abdul Latif Jameel World Water and Food Systems (J-WAFS) Lab and fellowship support from the Martin Family Society of Fellows.

Nomenclature

Greek Symbols

χ_d	Volumetric Membrane Charge Density
$\Delta\psi_{\mathrm{D}}$	Donnan Potential
Δx_e	Effective Membrane Thickness
Δ	Step Size in Tsitouras Method
γ_j	Activity Coefficient of Solute j
λ_j	Ratio of Solute j's Stokes Radius to Pore Radius
$\lambda_{\mathcal{E}}$	Lagrange Multiplier for Soft Loss Constraint
μ_{ij}	Experimental Rejection Mean for Flux i and Solute j
$\phi_{j,\mathrm{Di}}$	Dielectric Exclusion Partition Coefficient of Solute j
$\phi_{j,\mathrm{Do}}$	Donnan Exclusion Partition Coefficient of Solute j
$\phi_{j,\mathrm{F}}$	Solution-Friction Partition Coefficient of Solute j
$\phi_{j,\mathrm{S}}$	Steric Exclusion Partition Coefficient of Solute j
ψ	Electric Potential
σ_{ij}^2	Experimental Rejection Variance for Flux i and Solute j
Θ	Finite Dimensional Learnable Parameter Space
θ	Learnable Parameters in ODENet
ξ	Linearized Electric Potential Gradient
ζ_b	Dielectric Constant of Water in the Bulk Solution
ζ_p	Dielectric Constant in the Membrane Pores

Roman Symbols

0^+	Membrane Side at Solution-Membrane Interface (Feed)
0^-	Solution Side at Solution-Membrane Interface (Feed)
a	Activity
C_j	Molar Concentration of Solute j
d	Dimension of Hidden Layer Input/Output
D_j	Diffusion Coefficient of Solute j
d_k	Dimension of Key Matrix in Attention
FFNN	Feed-forward Neural Network
F	Faraday's Constant
$\mathbf{h}$	Hidden Layer Output
HIB	Hard Inductive Biases Used
$H_{j,c}$	Integrated Convective Hindrance Coefficient of Solute j
$H_{j,d}$	Integrated Diffusive Hindrance Coefficient of Solute j
$\bar{J}_v$	Normalized Permeate Flux Across Dataset
J_j	Molar Flux of Solute j
J_v	Permeate Water Flux
K	Key Matrix in Attention Calculation
k	Number of Flux Measurements during Pre-training
$\bar{k}_{c,j}$	Modified Mass Transfer Coefficient of Solute j
k_B	Boltzmann's Constant
$K_{j,c}$	Convective Hindrance Coefficient of Solute j
$K_{j,d}$	Diffusive Hindrance Coefficient of Solute j
$K_{j,f}$	Friction Coefficient of Solute j
$\mathcal{L}$	Loss Function
MSE	Mean Squared Error Loss
n	Number of Solutes in Solution
$\mathcal{N}$	Normal Distribution
NPT	No Pre-training Conducted
n_f	Number of Flux Measurements during Fine-tuning
N_A	Avogadro's Constant
Pe_j	Péclet Number of Solute j
$\mathrm{proj}_v(u)$	Projection Operation of u onto v
PT	Pre-training Conducted
Q	Query Matrix in Attention Calculation
q	Fundamental Electronic Charge
R	Universal Gas Constant
r_j	Stokes Radius of Solute j
r_p	Pore Radius of Membrane
$\mathcal{S}_\ell$	Set of Total Number of Possible Measurements Taken
SIB	Soft Inductive Biases Used
T	Absolute Temperature
V	Value Matrix in Attention Calculation
x	Spatial Coordinate Orthogonal to Membrane
$\mathcal{Z}_\ell$	Set of All Latent Membrane Parameters
z_j	Valence of Solute j

Superscripts

cm	Continuum Model
exp	Experimental Measurement
mod	Model Prediction
SF	Solution-Friction Model

Subscripts

a, X	Anion
b	Bulk Solution
c, M	Cation
f	Feed Stream
j	Species Index
m	Solution-Membrane Interface
p	Permeate Stream
$\parallel$	Parallel Component of Concentration Vector
$\perp$	Orthogonal Component of Concentration Vector

Fraktur Symbols

$\mathfrak{R}_j^{\mathrm{exp}}$	Experimental Rejection of Solute j
$\mathfrak{R}_j^{\mathrm{lim}}$	Limiting Rejection of Solute j
$\mathfrak{R}_j^{\mathrm{mod}}$	Model Rejection Prediction of Solute j

Appendix A. Soft Loss with Lagrange Multipliers

When evaluating the loss during pre-training or fine-tuning, hard inductive biases were used given their superior generalization performance (as reported and discussed in Section 3.2). Since soft loss constraints have also shown promise in many physics-informed neural networks [29], albeit typically in data-rich settings [87], we also benchmark against them. The loss function used during fine-tuning, which reframes the problem using the method of Lagrange multipliers [88], is shown below:

$$\mathcal{L}^{\text{cm}}(\mathbf{h},\mathbf{h}^{\text{cm}}) = \frac{1}{kd}\sum_{i=1}^{k}\sum_{j=1}^{d}\left[\mathbf{h}_j(J_{v,i}) - \mathbf{h}_j^{\text{cm}}(J_{v,i})\right]^2 + \lambda_{\mathcal{E}}\sum_{i=1}^{k}||z^\top \mathbf{h}(J_{v,i})||_2^2 \quad \text{(A.1)}$$

where $\lambda_{\mathcal{E}}$ is the Lagrange mulitplier used to weight the importance of electroneutrality relative to minimizing the ℓ_2 empirical risk term. We experimented with $\lambda_{\mathcal{E}} \in \{0.01, 0.1, 1, 2.5, 5\}$, where optimal generalization performance was observed for $\lambda_{\mathcal{E}} = 1$.

Appendix B. Donnan–Steric Pore Model with Dielectric Exclusion (DSPM–DE)

The Donnan–Steric Pore Model with Dielectric Exclusion (DSPM–DE) was originally proposed by Bowen and Welfoot, and has become one of the most frequently used transport models for NF [24]. The model adopts the extended Nernst–Planck equations to quantify ion transport across NF membranes. The hindered formulation of the equation, which accounts for restricted transport through nanoporous membranes is expressed in one-dimension in Eq. (B.1) below:

$$J_j = -D_j K_{j,d}\frac{dC_j}{dx} + K_{j,c}C_j J_v - \frac{K_{j,d}D_j C_j z_j \mathfrak{F}}{RT}\frac{d\psi}{dx} \quad \text{(B.1)}$$

In the above equation, J is the flux, where subscripts j and v denote species j and water, respectively. D is the ion's bulk diffusion coefficient in water, and $K_{j,d}$ and $K_{j,c}$ account for the reduced ion mobilities inside the membrane pores [23]. These two terms correspond to diffusive and convective hindrance factors for species j, respectively. In addition, the two dependent variables used for modelling purposes are C and ψ, which correspond to the molar concentration and electric potential, respectively. Here, $x \in [0, \Delta x_e]$, is the direction orthogonal to the solution-membrane interface and spans the effective membrane thickness, Δx_e. Lastly, the remaining variables are $\mathfrak{F}$, z, R, and T, which are Faraday's constant, ion valence, the universal gas constant, and absolute temperature, respectively.

The unhindered version of the extended Nernst–Planck PDE is linearized and used to account for ion transport in the feed-side boundary layer [60]. A mass transfer coefficient, $\bar{k}_{c,j}$ is also introduced to account for boundary layer effects that arise from the spacers present in conventional NF systems [16]. The linearized equation is:

$$J_j = -\bar{k}_{c,j}\left[C_{j,f,m} - C_{j,f,b}\right] + J_v C_{j,f,m} - z_j C_{j,f,m} D_j \frac{\mathfrak{F}\xi}{RT} \quad \text{(B.2)}$$

Here, subscripts f and p denote the feed and permeate streams, respectively. Additionally, as a result of the linearization, ξ serves as the linearized electric potential gradient. Depending on the system configuration, various mass transfer correlations are imposed to quantify the mass transfer coefficient [60]. For all the reported work, we use the following mass transfer correction:

$$\bar{k}_{c,j} = k_{c,j}\left[\omega_w + \left(1 + 0.26\omega_w^{1.4}\right)\right]^{-1.7} \quad \text{(B.3)}$$

where $\omega_w \triangleq J_v/k_{c,j}$. The modifications presented in the above equation account for the membrane suction effect [89].

In the permeate stream, the common assumption is that the solution is dilute enough to not have sufficient concentration polarization to necessitate a separate PDE. As a result, only convective transport is assumed in the permeate stream, whereby the solute flux is a product of the permeate concentration, $C_{j,p}$ and the water flux, J_v.

Since the differential equations governing transport across the membrane have been completely defined, we now require boundary conditions and electroneutrality constraints to fully close the system of equations [62]. The boundary conditions in the feed bulk correspond to those from the feed composition. At the solution-membrane interface on the feed side, steric, dielectric, and Donnan exclusion partition coefficients are introduced [90]. The product of these partition coefficients quantifies the discontinuous concentrations at the interface:

$$\frac{\gamma_j(0^-)C_{j,f,m}(0^-)}{\gamma_j(0^+)C_{j,f,m}(0^+)} = \phi_{j,\text{S}}\phi_{j,\text{Di}}\phi_{j,\text{Do}} \quad \text{(B.4)}$$

Here, 0^- and 0^+ correspond to the solution side and membrane side, respectively. A reciprocating expression is present on the permeate side. ϕ is a partition coefficient, with subscripts S, Di, and Do, relating to steric, dielectric, and Donnan exclusion, respectively. The complete formulation of the partition coefficients is described in previous work [14]. γ are the activity coefficients used to quantify the non-ideality of the mixture. Although the Davies model is typically used in DSPM–DE, we adopt the Pitzer–Kim model given its increased accuracy at elevated salinities [91, 92].

Finally, the electroneutrality terms are prescribed in the feed boundary layer, membrane, and permeate [93]. Inside the membrane, the equation takes the following form:

$$\chi_d + \sum_{j=1}^{d} z_j C_j = 0 \quad \text{(B.5)}$$

where χ_d is the volumetric membrane charge density. The summation iterates from $j = 1, 2, \ldots, d$, where the d corresponds to the total number of solutes in solution.

At pH 7, zeta potential measurements of NF270 indicate that the membrane is negatively charged, meaning that χ_d takes on values below zero [94]. As the pH is reduced, making the surrounding solution more acidic, the membrane goes past its iso-electric point, transitioning from a negatively-charged state to neutral to positively-charged. This is often performed as a pretreatment step to improve cation selectivities for metals recovery applications [14]. The electroneutrality relations used in the feed stream and permeate stream are equivalent (barring the values of the concentrations used) and presented below:

$$\sum_{j=1}^{d} z_j C_j = 0 \quad \text{(B.6)}$$

We illustrate the discretized form of the membrane in Fig. B.5.

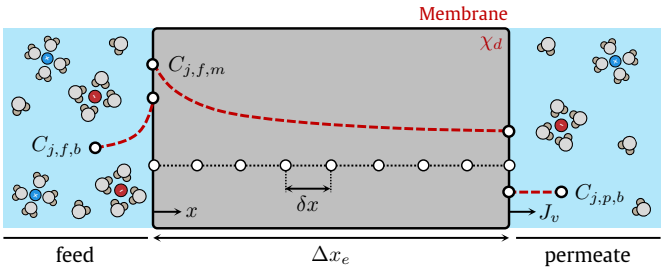


Figure B.5: Discretization of the membrane by δx in DSPM–DE. The boundary layer forms on the feed side and is absent in the permeate. The concentration discontinuity arising from a continuous chemical potential gradient is evident in both feed and permeate streams. The x direction is also shown.

Appendix C. Salinity Distribution and Pre-training Ranges

In Fig. C.6, we illustrate the distribution of ion concentrations present in the training, validation, and test datasets. The salinity ranges studied spanned 2–35 g/L with varying compositions in between (studying concentrations above those of seawater could lead to low quality rejection predictions from the continuum model, and were hence not performed). We also report the concentration ranges over which pre-training was performed (denoted by solid lines in Fig. C.6).

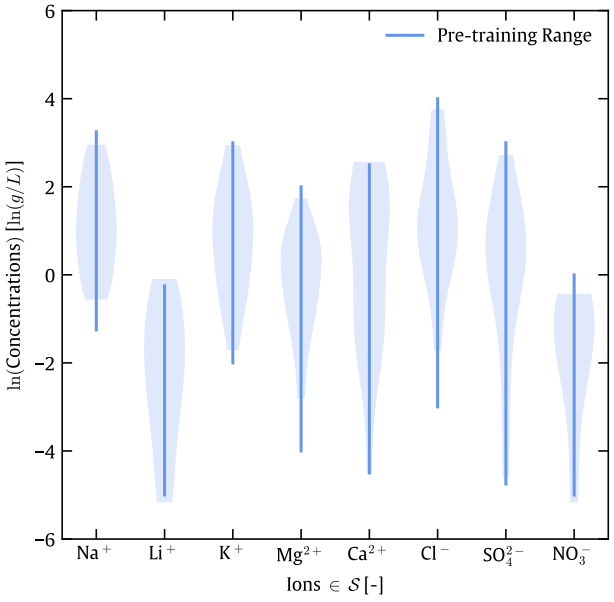


Figure C.6: The distribution of ion concentrations present in the training, validation, and test data (shaded regions), as well as the pre-training ranges used (solid lines).

We explicitly selected our pre-training ranges to coincide with the measured data to prevent the model from having to significantly extrapolate at inference time. This approach can also be applied in other transfer learning settings, where the model can be pre-trained on moderate-accuracy continuum models based on the expected operating ranges [95]. Subsequently, the neural model can then be fine-tuned on limited data after moderate-quality embeddings are obtained from the pre-training stage [96].

In the studies performed, larger pre-training ranges could have been used; however, we observed that majority of the computation time was consumed during the pre-training stage (the pre-training step took just under a week to conduct using our computational setup) i.e., to improve the quality of the learned embeddings, minor increases in the pre-training range substantially exacerbated training runtime.

Appendix D. Bayesian Optimization for Hyperparameter Tuning and Selection

To determine an optimal set of hyperparameters, ensuring efficient learning and a set of feasible solutions, we use Bayesian optimization with BoTorch [97]. Hyperparameters considered were number of layers, hidden layer dimension, batch size, non-linear activation functions, optimizer choice, and learning rate [98].

For the number of hidden layers, we experimented with ranges between 3–7. Despite the additional expressive power of deeper networks [99], rejection predictions worsened with larger models. With the hidden layer dimension, we noticed that a layer dimension of 12 consistently provided stable performance. As a result, it was fixed during the fine-tuning process.

When considering the batch size and learning rate, we observed that the model appeared to be quite sensitive to these parameters. Stable performance was noted for a batch size of eight and a learning rate scheduler was adopted to improve stability during fine-tuning [100]. Without the scheduler, obtaining stable convergence was a challenge. The scheduler halved the learning rate every 100 epochs starting with a learning rate of 1×10^{-3}.

For the non-linear activations, we chose $\tanh(\cdot)$ non-linearities to ensure that concentrations were appropriately normalized prior to passage into the subsequent layer. This also prevented estimates from growing unboundedly and engendering vanishing gradients. The Adam optimizer was used across performed tests, with instabilities arising when stochastic gradient descent (SGD) was applied. The final set of converged hyperparameters is provided in Table D.2 below:

Table D.2: Finalized hyperparameters for training our proposed ODENet.

num. layers	layer dim.	batch.	activation	optim.	lr (init.)
5	12	8	$\tanh(\cdot)$	Adam	1×10^{-3}

Appendix E. Deep Learning Architecture Details

We adopt CNNs, ResNets, and U-Nets, for model benchmarking. Architecture details are summarized in Table E.3. Hyperparameter optimization following Appendix D was also performed. Batch normalization was used in all models, with no padding and a stride of 1 for all convolutional layers. The tuple in the layer dimensions corresponds to the convolutional filters and fully-connected layers, respectively.

Table E.3: Deep learning architecture details. 'fc' is a fully-connected layer.

model	num. layers	layer dim.	activation	optim.	lr (init.)
CNN	3 conv. 2 fc.	3, 12	$\tanh(\cdot)$	Adam	1×10^{-3}
ResNet	5 fc.	12	$\tanh(\cdot)$	Adam	1×10^{-3}
U-Net	2 conv. 3 fc.	3, 12	$\tanh(\cdot)$	Adam	1×10^{-3}

Appendix F. Restricted Transport with Hindrance Factors

In hindered transport theory, solute molecules are most frequently treated as particles, while the solvent is modelled like a continuum [61]. In addition, given the small length and velocity scales present, the Reynolds number is sufficiently small to render Stokes' equation applicable [61]. Under these conditions, solute–solute interactions are assumed negligible and long-range interactions can also be ignored.

Under the assumption that these conditions hold, restricted transport inside the membrane pores can be characterized using hindrance terms, often referred to as *enhanced drag coefficients* (drag coefficients relative to those in an unbounded fluid). There exist hindrance coefficients for both convection and diffusion, $K_{j,c}$ and $K_{j,d}$, respectively, which capture the reduced transport of solute j inside the membrane. The key quantity used to parameterize these coefficients is λ_j, which corresponds to the ratio of species j's hydrodynamic radius to the membrane pore radius, r_p. In cylindrical pores, as the solute size decreases relative to the pore size, diffusion can be treated as if it were in the bulk solution i.e. $K_{j,d} \to 1$ for $\lambda_j \to 0$. Conversely, as the solute becomes of comparable size to the membrane pore, diffusive effects are seen to vanish i.e. $K_{j,d} \to 0$ for $\lambda_j \to 1$. In the convective case, although $K_{j,c} \to 1$ for $\lambda_j \to 0$, $K_{j,c}$ only tends to 0 for $\lambda_j \to 1$ when the pores are cylindrical [61]. The convective hindrance coefficient is also not necessarily a monotonic function of λ_j due to the dominant effects of steric exclusion near the membrane wall.

From studies conducted by Higdon and Muldowney (which characterize solute transport through porous membranes for $0 \leq \lambda_j \leq 0.95$), the following expression can be used to evaluate the diffusive hindrance coefficient [85]:

$$K_{j,d} \triangleq K_d(\lambda_j) = \frac{1 + (9/8)\lambda_j \ln \lambda_j - 1.56034\lambda_j + 0.528155\lambda_j^2 + 1.91521\lambda_j^3 - 2.81903\lambda_j^4 + 0.270788\lambda_j^5 + 1.10115\lambda_j^6 - 0.435933\lambda_j^7}{(1-\lambda_j)^2} \tag{F.1}$$

For larger solutes, where $\lambda_j > 0.95$, Mavrovouniotis and Brenner applied asymptotic matching to evaluate the diffusive hindrance coefficient in the limit of $\lambda_j \to 1$ [86]. In this case, the diffusive hindrance coefficient takes the following form:

$$K_d(\lambda_j) = 0.984 \left(\frac{1-\lambda_j}{\lambda_j} \right)^{5/2} \tag{F.2}$$

In the convective case, Ennis et al. used a Padé approximation in conjunction with lubrication results obtained from Bungay and Brenner to obtain the following expression for the convective hindrance coefficient for all $\lambda_j \in [0, 1]$ [101, 102]:

$$K_{j,c} \triangleq K_c(\lambda_j) = \frac{1 + 3.867\lambda_j - 1.907\lambda_j^2 - 0.834\lambda_j^3}{1 + 1.867\lambda_j - 0.741\lambda_j^2} \tag{F.3}$$

Appendix G. Solution-Friction Model

The solution-friction model was originally posed by Wang et al. as a means for capturing salt and water transport across RO and NF membranes [45]. Similarly to DSPM–DE, the model also adopts the extended Nernst–Planck equations to model ion transport across polyamide membranes; however, instead of using the convective and diffusive hindrance factors derived from perturbation theory, the model adopts one friction factor or hindrance function, $K_{j,f}$ per ionic species [84]. The modified version of the extended Nernst–Planck equations in its one-dimensional form is:

$$J_j = -D_j K_{j,f} \frac{dC_j}{dx} + K_{j,f} C_j J_v - \frac{K_{j,f} D_j C_j z_j \mathfrak{F}}{RT} \frac{d\psi}{dx} \tag{G.1}$$

The friction function, $K_{j,f}$, is a function of the interactions between the ion and fluid, as well as the ion and membrane [45]. Since these interactions have been quantified in prior work, $K_{j,f}$ is known quantity in the solution-friction model [103].

In addition to the ion-solution and ion-membrane friction terms, there is also a regressed friction term between the fluid and membrane, $f_{f,m}$; however, this term only shows up in the pressure gradient evaluation, which we do not use in the proposed model. As a result, it does not contribute to the number of latent variables to be regressed.

In addition to the modifications made to the governing differential equation, the solution-friction model also alters the partitioning relationships at the solution-membrane interface [45]. Instead of treating the steric and Born exclusion mechanisms separately, the model prescribes a new partition coefficient, $\phi_{j,\mathrm{F}}$, which characterizes their combined effect [84]. The new coefficient is motivated by the fact that steric and dielectric exclusion are coupled (yet without an understanding of what the parametric form between them looks like) [104, 105]. As a result, $\phi_{j,\mathrm{F}}$ replaces the product of the steric and Born partition coefficients used in DSPM–DE as shown below:

$$\phi_{j,\mathrm{S}} \phi_{j,\mathrm{Di}} \phi_{j,\mathrm{Do}} \overset{\mathrm{SF}}{=} \phi_{j,\mathrm{F}} \phi_{j,\mathrm{Do}} \tag{G.2}$$

Furthermore, since it is challenging to measure information about the size (or state) of the ions inside the membrane after partial dehydration, the authors treat $\phi_{j,\mathrm{F}}$ as a latent variable to be regressed [45]. Consequently, for new membranes, Δx_e, χ_d, and r_p, as well as $\phi_{j,\mathrm{F}}$ for each ion species j, are regressed. This yields a total of $n+3$ learned parameters, for a solution comprising n ions (when the fluid-membrane term is neglected; if it is included, this would constitute $n+4$ parameters in total) [84]. We also note that in certain cases when Δx_e, χ_d, and r_p have been previously regressed for a given membrane, they can be re-used, leaving n parameters to be regressed for new compositions on the same membrane. This was consistent with the approach taken by Wang et al. [45].

The original solution-friction model also neglects the activity coefficients present in the concentration discontinuity at the solution-membrane interface [45, 84]. In the reported work, we include activity coefficients to be more accurate, while ensuring a fair comparison between DSPM–DE, extended DSPM–DE, and the solution-friction approach. As a result, on the feed side, the boundary condition takes the following form (with a similar expression present on the permeate

side):

$$\frac{\gamma_j(0^-)C_{j,f,m}(0^-)}{\gamma_j(0^+)C_{j,f,m}(0^+)} = \phi_{j,\mathrm{F}}\phi_{j,\mathrm{Do}} \tag{G.3}$$

The remaining details of the model are mostly in agreement with those in DSPM–DE (in some implementations, the handling of the feed side boundary layer, and/or the mass transfer correlations used, may vary). In the reported work, to maintain consistency between the two models, we adopt the same mass transfer correlations, and how the feed stream boundary layer is modelled [106].

Appendix H. Learned Parameters and Hindrance Factors from Extended DSPM–DE

In this section, we summarize the set of regressed parameters obtained when calibrating the extended DSPM–DE model on the salinity and composition studied in Section 3.3. The regression methodology used was in alignment with the approach detailed in our previous work [62]. In this framework, we apply simulated annealing for global optimization in conjunction with the Nelder–Mead local search option to determine our set of regressed parameters. The membrane parameter values obtained are provided in Table H.4 below:

Table H.4: The set of regressed membrane parameters when using the extended DSPM–DE model on the composition studied in Section 3.3.

r_p [nm]	Δx_e [µm]	ζ_p [-]	χ_d [mol·m^{-3}]
0.37	2.38	80.82	−100.29

When using the extended DSPM–DE model, we also regress ion-specific convective and diffusive hindrance factors, as stated *a priori*. The values regressed are summarized in Table H.5 below:

Table H.5: The set of regressed convective and diffusive hindrance parameters when calibrating the extended DSPM–DE model on the composition studied in Section 3.3.

	Na^+ [-]	Ca^{2+} [-]	Mg^{2+} [-]	Cl^- [-]	NO_3^- [-]	SO_4^{2-} [-]
K_c	0.26	0.61	0.77	0.66	1.10	0.28
K_d	0.21	0.23	0.02	0.45	0.30	1.10

References

[1] Ryan M. DuChanois, Nathanial J. Cooper, Boreum Lee, Sohum K. Patel, Lauren Mazurowski, Thomas E. Graedel, and Menachem Elimelech. Prospects of metal recovery from wastewater and brine. *Nature Water*, 1(1):37–46, 2023. URL https://doi.org/10.1038/s44221-022-00006-z.

[2] Mohammad A. Alkhadra, Xiao Su, Matthew E. Suss, Huanhuan Tian, Eric N. Guyes, Amit N. Shocron, Kameron M. Conforti, J. Pedro de Souza, Nayeong Kim, Michele Tedesco, Khoiruddin Khoiruddin, I Gede Wenten, Juan G. Santiago, T. Alan Hatton, and Martin Z. Bazant. Electrochemical methods for water purification, ion separations, and energy conversion. *Chemical Reviews*, 122(16):13547–13635, 2022. URL https://doi.org/10.1021/acs.chemrev.1c00396.

[3] Ruoyu Wang, Rongrong He, Tao He, Menachem Elimelech, and Shihong Lin. Performance metrics for nanofiltration-based selective separation for resource extraction and recovery. *Nature Water*, 1(3):291–300, 2023. URL https://doi.org/10.1038/s44221-023-00037-0.

[4] Yong-Uk Shin, Weiyi Pan, Sohum K. Patel, Jihun Lim, Lea R. Winter, Wen Ma, Seungkwan Hong, and Menachem Elimelech. Electrochemical chromium removal using nanodiamond enhanced flow-through electrosorption. *Chemical Engineering Journal*, 474:145756, 2023. URL https://doi.org/10.1016/j.cej.2023.145756.

[5] Li Wang, Danyal Rehman, Peng-Fei Sun, Akshay Deshmukh, Liyuan Zhang, Qi Han, Zhe Yang, Zhongying Wang, Hee-Deung Park, John H. Lienhard, and Chuyang Y. Tang. Novel positively charged metal-coordinated nanofiltration membrane for lithium recovery. *ACS Applied Materials & Interfaces*, 13(14):16906–16915, 2021. URL https://doi.org/10.1021/acsami.1c02252.

[6] Baoying Wang, Fei Liu, Feng Zhang, Ming Tan, Heqing Jiang, Yang Liu, and Yang Zhang. Efficient separation and recovery of cobalt(ii) and lithium(i) from spent lithium ion batteries (LIBs) by polymer inclusion membrane electrodialysis (PIMED). *Chemical Engineering Journal*, 430:132924, 2022. URL https://doi.org/10.1016/j.cej.2021.132924.

[7] Ramesh Kumar, Chengjia Liu, Geon-Soo Ha, Young-Kwon Park, Moonis Ali Khan, Min Jang, Sang-Hyoun Kim, Mohammed A. Amin, Amel Gacem, and Byong-Hun Jeon. Downstream recovery of Li and value-added metals (Ni, Co, and Mn) from leach liquor of spent lithium-ion batteries using a membrane-integrated hybrid system. *Chemical Engineering Journal*, 447:137507, 2022. URL https://doi.org/10.1016/j.cej.2022.137507.

[8] Jasir Jawad, Alaa H. Hawari, and Syed Javaid Zaidi. Artificial neural network modeling of wastewater treatment and desalination using membrane processes: A review. *Chemical Engineering Journal*, 419:129540, 2021. URL https://doi.org/10.1016/j.cej.2021.129540.

[9] Te Tu, Shuo Liu, Qiufang Cui, Liqiang Xu, Long Ji, and Shuiping Yan. Techno-economic assessment of waste heat recovery enhancement using multi-channel ceramic membrane in carbon capture process. *Chemical Engineering Journal*, 400:125677, 2020. URL https://doi.org/10.1016/j.cej.2020.125677.

[10] Varun H. Hegde, Michael F. Doherty, and Todd M. Squires. A two-phase model that unifies and extends the classical models of membrane transport. *Science*, 377(6602):186–191, 2022. URL https://doi.org/10.1126/science.abm7192.

[11] Hongwei Li, Ying Wang, Tingyu Li, Xiang-Kui Ren, Jixiao Wang, Zhi Wang, and Song Zhao. Nanofiltration membrane with crown ether as exclusive Li^+ transport channels achieving efficient extraction of lithium from salt lake brine. *Chemical Engineering Journal*, 438:135658, 2022. URL https://doi.org/10.1016/j.cej.2022.135658.

[12] Rongrong He, Shanshan Xu, Ruoyu Wang, Bingyang Bai, Shihong Lin, and Tao He. Polyelectrolyte-based nanofiltration membranes with exceptional performance in Mg^{2+}/Li^+ separation in a wide range of solution conditions. *Journal of Membrane Science*, 663:121027, 2022. URL https://doi.org/10.1016/j.memsci.2022.121027.

[13] Ruoyu Wang, Rayan Alghanayem, and Shihong Lin. Multipass nanofiltration for lithium separation with high selectivity and recovery. *Environmental Science & Technology*, 57(38):14464–14471, 2023. URL https://doi.org/10.1021/acs.est.3c04220.

[14] Danyal Rehman, Fareed Sheriff, and John H. Lienhard. Quantifying uncertainty in nanofiltration transport models for enhanced metals recovery. *Water Research*, 243:120325, 2023. URL https://doi.org/10.1016/j.watres.2023.120325.

[15] Prashant S. Kulkarni. Recovery of Uranium(VI) from acidic wastes using tri-n-octylphosphine oxide and sodium carbonate based liquid membranes. *Chemical Engineering Journal*, 92(1):209–214, 2003. URL https://doi.org/10.1016/S1385-8947(02)00255-3.

[16] Omar Labban, Chang Liu, Tzyy Haur Chong, and John H. Lienhard. Fundamentals of low-pressure nanofiltration: Membrane characterization, modeling, and understanding the multi-ionic interactions in water softening. *Journal of Membrane Science*, 521:18–32, 2017. URL https://doi.org/10.1016/j.memsci.2016.08.062.

[17] Deniz Rall, Artur M. Schweidtmann, Benedikt M. Aumeier, Johannes Kamp, Jannik Karwe, Katrin Ostendorf, Alexander Mitsos, and Matthias Wessling. Simultaneous rational design of ion separation membranes and processes. *Journal of Membrane Science*, 600:117860, 2020. URL https://doi.org/10.1016/j.memsci.2020.117860.

[18] Ruoyu Wang and Shihong Lin. Pore model for nanofiltration: History, theoretical framework, key predictions, limitations, and prospects. *Journal of Membrane Science*, 620:118809, 2021. URL https://doi.org/10.1016/j.memsci.2020.118809.

[19] Zi Hao Foo, Danyal Rehman, Andrew T. Bouma, Sebastian Monsalvo, and John H. Lienhard. Lithium concentration from salt-lake brine by Donnan-enhanced nanofiltration. *Environmental Science & Technology*, 57(15):6320–6330, 2023. URL https://doi.org/10.1021/acs.est.2c08584.

[20] Ora Kedem and Aharon Katchalsky. Permeability of composite membranes. part 1.—electric current, volume flow and flow of solute through membranes. *Transactions of the Faraday Society*, 59:1918–1930, 1963. URL http://dx.doi.org/10.1039/TF9635901918.

[21] Kurt S. Spiegler and Ora Kedem. Thermodynamics of hyperfiltration (reverse osmosis): criteria for efficient membranes. *Desalination*, 1(4):311–326, 1966. URL https://doi.org/10.1016/S0011-9164(00)80018-1.

[22] Richard W. Bowen and Hilmi Mukhtar. Characterisation and prediction of separation performance of nanofiltration membranes. *Journal of Membrane Science*, 112(2):263–274, 1996. URL https://doi.org/10.1016/0376-7388(95)00302-9.

[23] William M. Deen. Hindered transport of large molecules in liquid-filled pores. *AIChE Journal*, 33(9):1409–1425, 1987. URL https://doi.org/10.1002/aic.690330902.

[24] Richard W. Bowen and Julian S. Welfoot. Modelling the performance of membrane nanofiltration—critical assessment and model development. *Chemical Engineering Science*, 57(7):1121–1137, 2002. URL https://doi.org/10.1016/S0009-2509(01)00413-4.

[25] Richard W. Bowen and Julian S. Welfoot. Modelling of membrane nanofiltration—pore size distribution effects. *Chemical Engineering Science*, 57(8):1393–1407, 2002. URL https://doi.org/10.1016/S0009-2509(01)00412-2.

[26] Andrij E. Yaroshchuk. Dielectric exclusion of ions from membranes. *Advances in Colloid and Interface Science*, 85(2):193–230, 2000. URL https://doi.org/10.1016/S0001-8686(99)00021-4.

[27] Verónica Silva, Álvaro Martín, Fernando Martínez, Jorge Malfeito, Pedro Prádanos, Laura Palacio, and Antonio Hernández. Electrical characterization of nf membranes. a modified model with charge variation along the pores. *Chemical Engineering Science*, 66(13):2898–2911, 2011. ISSN 0009-2509. URL https://doi.org/10.1016/j.ces.2011.03.025.

[28] Jan Straatsma, Gerrald Bargeman, Hilje C. van der Horst, and Johannes A. Wesselingh. Can nanofiltration be fully predicted by a model? *Journal of Membrane Science*, 198(2):273–284, 2002. URL https://doi.org/10.1016/S0376-7388(01)00669-X.

[29] Mazier Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. URL https://doi.org/10.1016/j.jcp.2018.10.045.

[30] Maziar Raissi, Alireza Yazdani, and George E. Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367(6481):1026–1030, 2020. URL https://doi.org/10.1126/science.aaw4741.

[31] Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3):88, 2022. URL https://doi.org/10.1007/s10915-022-01939-z.

[32] Saif Al Aani, Talal Bonny, Shadi W. Hasan, and Nidal Hilal. Can machine language and artificial intelligence revolutionize process automation for water treatment and desalination? *Desalination*, 458:84–96, 2019. URL https://doi.org/10.1016/j.desal.2019.02.005.

[33] Mingyi Fan, Jiwei Hu, Rensheng Cao, Wenqian Ruan, and Xionghui Wei. A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence. *Chemosphere*, 200:330–343, 2018. URL https://doi.org/10.1016/j.chemosphere.2018.02.111.

[34] Cody L. Ritt, Mingjie Liu, Tuan Anh Pham, Razi Epsztein, Heather J Kulik, and Menachem Elimelech. Machine learning reveals key ion selectivity mechanisms in polymeric membranes with subnanometer pores. *Science Advances*, 8(2):eabl5771, 2022. URL https://doi.org/10.1126/sciadv.abl5771.

[35] Naif A. Darwish, Nidal Hilal, Habis Al-Zoubi, and Abdul W. Mohammad. Neural networks simulation of the filtration of sodium chloride and magnesium chloride solutions using nanofiltration membranes. *Chemical Engineering Research and Design*, 85(4):417–430, 2007. URL https://doi.org/10.1205/cherd06037.

[36] Grishma R. Shetty and Shankararaman Chellam. Predicting membrane fouling during municipal drinking water nanofiltration using artificial neural networks. *Journal of Membrane Science*, 217(1):69–86, 2003. URL https://doi.org/10.1016/S0376-7388(03)00075-9.

[37] Richard W. Bowen, Meirion G. Jones, Julian S. Welfoot, and Haitham N.S. Yousef. Predicting salt rejections at nanofiltration membranes using artificial neural networks. *Desalination*, 129(2):147–162, 2000. URL https://doi.org/10.1016/S0011-9164(00)00057-6.

[38] Victor A. Yangali-Quintanilla, Arne R.D. Verliefde, Tae uk Kim, Anawr H.M. Sadmani, Maria D. Kennedy, and Gary L. Amy. Artificial neural network models based on qsar for predicting rejection of neutral organic compounds by polyamide nanofiltration and reverse osmosis membranes. *Journal of Membrane Science*, 342(1):251–262, 2009. URL https://doi.org/10.1016/j.memsci.2009.06.048.

[39] Nohyeong Jeong, Razi Epsztein, Ruoyu Wang, Shinyun Park, Shihong Lin, and Tiezheng Tong. Exploring the knowledge attained by machine learning on ion transport across polyamide membranes using explainable artificial intelligence. *Environmental Science & Technology*, 0(0):null, 2023. URL https://doi.org/10.1021/acs.est.2c08384.

[40] Deniz Rall, Daniel Menne, Artur M. Schweidtmann, Johannes Kamp, Lars von Kolzenberg, Alexander Mitsos, and Matthias Wessling. Rational design of ion separation membranes. *Journal of Membrane Science*, 569:209–219, 2019. URL https://doi.org/10.1016/j.memsci.2018.10.013.

[41] Deniz Rall, Artur M. Schweidtmann, Maximilian Kruse, Elizaveta Evdochenko, Alexander Mitsos, and Matthias Wessling. Multi-scale membrane process optimization with high-fidelity ion transport models through machine learning. *Journal of Membrane Science*, 608:118208, 2020. URL https://doi.org/10.1016/j.memsci.2020.118208.

[42] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018. URL https://proceedings.neurips.cc/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[44] Danyal Rehman and John H. Lienhard. Physics-constrained neural differential equations for learning multi-ionic transport. *International Conference on Learning Representations (ICLR) – Physics for Machine Learning Workshop*, 2023. URL https://arxiv.org/abs/2303.04594.

[45] Li Wang, Tianchi Cao, Jouke E. Dykstra, Slawomir Porada, P. M Biesheuvel, and Menachem Elimelech. Salt and water transport in reverse osmosis membranes: Beyond the solution-diffusion model. *Environmental Science & Technology*, 55(24):16665–16675, 2021. URL https://doi.org/10.1021/acs.est.1c05649.

[46] Patrick Kidger. On neural differential equations, 2022. URL https://arxiv.org/abs/2202.02435.

[47] Anudhyan Boral, Zhong Y. Wan, Leonardo Zepeda-Núñez, James Lottes, Qing Wang, Yi fan Chen, John R. Anderson, and Fei Sha. Neural ideal large eddy simulation: Modeling turbulence with neural stochastic differential equations, 2023. URL https://arxiv.org/abs/2306.01174.

[48] Charalampos Tsitouras. Runge–Kutta pairs of order 5(4) satisfying only the first column simplifying assumption. *Computers & Mathematics with Applications*, 62(2):770–775, 2011. ISSN 0898-1221. URL https://doi.org/10.1016/j.camwa.2011.06.002.

[49] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. URL https://arxiv.org/abs/1412.6980.

[50] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272, 2020. URL https://doi.org/10.1038/s41592-019-0686-2.

[51] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf.

[52] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lió, and Yoshua Bengio. Graph attention networks, 2018. URL https://arxiv.org/abs/1710.10903.

[53] Orion Dollar, Nisarg Joshi, David A. C. Beck, and Jim Pfaendtner. Attention-based generative models for de novo molecular design. *Chemical Science*, 12(24):8362–8372, 2021. URL https://doi.org/10.1039/D1SC01050F.

[54] Yuki Matsukiyo, Chikashige Yamanaka, and Yoshihiro Yamanishi. De novo generation of chemical structures of inhibitor and activator candidates for therapeutic target proteins by a transformer-based variational autoencoder and bayesian optimization. *Journal of Chemical Information and Modeling*, 0(0):null, 2023. URL https://doi.org/10.1021/acs.jcim.3c00824.

[55] Danyal Rehman and John H. Lienhard. Attention-enhanced neural differential equations for physics-informed deep learning of ion transport. *Advances in Neural Information Processing Systems – Machine Learning and the Physical Sciences Workshop*, 2023. URL https://arxiv.org/abs/2312.02871.

[56] Pedro J. de Souza, Amir Levy, and Martin Z. Bazant. Electroneutrality breakdown in nanopore arrays. *Physical Review E*, 104:044803, Oct 2021. URL https://doi.org/10.1103/PhysRevE.104.044803.

[57] Behnam Neyshabur, Hanie Sedghi, and Chiyuan Zhang. What is being transferred in transfer learning? In *Advances in Neural Information Processing Systems*, volume 33, pages 512–523. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/0607f4c705595b911a4f3e7a127b44e0-Paper.pdf.

[58] Lei Li, Shuming Rong, Rui Wang, and Shuili Yu. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. *Chemical Engineering Journal*, 405:126673, 2021. URL https://doi.org/10.1016/j.cej.2020.126673.

[59] Dan Lu, Xuanchao Ma, Jiancong Lu, Yukun Qian, Yifang Geng, Jing Wang, Zhikan Yao, Lijun Liang, Zhilin Sun, Songmiao Liang, and Lin

Zhang. Ensemble machine learning reveals key structural and operational features governing ion selectivity of polyamide nanofiltration membranes. *Desalination*, 564:116748, 2023. URL https://doi.org/10.1016/j.desal.2023.116748.

[60] Vítor Geraldes and Ana M. Brites Alves. Computer program for simulation of mass transport in nanofiltration membranes. *Journal of Membrane Science*, 321(2):172–182, 2008. URL https://doi.org/10.1016/j.memsci.2008.04.054.

[61] Panadda Dechadilok and William M. Deen. Hindrance factors for diffusion and convection in pores. *Industrial & Engineering Chemistry Research*, 45(21):6953–6959, 2006. URL https://doi.org/10.1021/ie051387n.

[62] Danyal Rehman and John H. Lienhard. Global optimization for accurate and efficient parameter estimation in nanofiltration. *Journal of Membrane Science Letters*, 2(2):100034, 2022. URL https://doi.org/10.1016/j.memlet.2022.100034.

[63] Stephen Joe and Frances Y. Kuo. Constructing Sobol sequences with better two-dimensional projections. *SIAM Journal on Scientific Computing*, 30(5):2635–2654, 2008. URL https://doi.org/10.1137/070709359.

[64] Habis Al-Zoubi, Nidal Hilal, Naif A. Darwish, and Abdul W. Mohammad. Rejection and modelling of sulphate and potassium salts by nanofiltration membranes: Neural Network and Spiegler-Kedem model. *Desalination*, 206(1):42–60, 2007. URL https://doi.org/10.1016/j.desal.2006.02.060.

[65] Marina Micari, Dionysia Diamantidou, Sebastiaan G.J. Heijman, Massimo Moser, Amir Haidari, Henri Spanjers, and Valentin Bertsch. Experimental and theoretical characterization of commercial nanofiltration membranes for the treatment of ion exchange spent regenerant. *Journal of Membrane Science*, 606:118117, 2020. URL https://doi.org/10.1016/j.memsci.2020.118117.

[66] Habis Al-Zoubi and Waid Omar. Rejection of salt mixtures from high saline by nanofiltration membranes. *Korean Journal of Chemical Engineering*, 26(3):799–805, 2009. URL https://doi.org/10.1007/s11814-009-0133-7.

[67] Razi Epsztein, Evyatar Shaulsky, Nadir Dizge, David M. Warsinger, and Menachem Elimelech. Role of ionic charge density in donnan exclusion of monovalent anions by nanofiltration. *Environmental Science & Technology*, 52(7):4108–4116, 2018. URL https://doi.org/10.1021/acs.est.7b06400.

[68] Anissa Somrani, Ahmed H. Hamzaoui, and Maxime Pontie. Study on lithium separation from salt lake brines by nanofiltration (NF) and low pressure reverse osmosis (LPRO). *Desalination*, 317:184–192, 2013. URL https://doi.org/10.1016/j.desal.2013.03.009.

[69] Jianquan Luo and Yinhua Wan. Effect of highly concentrated salt on retention of organic solutes by nanofiltration polymeric membranes. *Journal of Membrane Science*, 372(1):145–153, 2011. URL https://doi.org/10.1016/j.memsci.2011.01.066.

[70] Luke Taylor and Geoff Nitschke. Improving deep learning with generic data augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1542–1547, 2018. URL https://doi.org/10.1109/SSCI.2018.8628742.

[71] Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995. URL https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=e26cc4a1c717653f323715d751c8dea7461aa105.

[72] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. URL https://arxiv.org/abs/1512.03385.

[73] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation, 2015. URL https://arxiv.org/abs/1505.04597.

[74] Gabriel Sigmund, Mehdi Gharasoo, Thorsten Hüffer, and Thilo Hofmann. Deep learning neural network approach for predicting the sorption of ionizable and polar organic pollutants to a wide range of carbonaceous materials. *Environmental Science & Technology*, 54(7):4583–4591, 2020. URL https://doi.org/10.1021/acs.est.9b06287.

[75] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[76] Luiz Chamon and Alejandro Ribeiro. Probably approximately correct constrained learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 16722–16735. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/c291b01517f3e6797c774c306591cc32-Paper.pdf.

[77] Long Tian, Peng Zhou, Zhaoyang Su, Ting Liu, Nigel Graham, Tom Bond, and Wenzheng Yu. Insights into the properties of surface waters and their associated nanofiltration membrane fouling: The importance of biopolymers and high molecular weight humics. *Chemical Engineering Journal*, 451:138682, 2023. URL https://doi.org/10.1016/j.cej.2022.138682.

[78] Agnese Marcato, Gianluca Boccardo, and Daniele Marchisio. A computational workflow to study particle transport and filtration in porous media: Coupling cfd and deep learning. *Chemical Engineering Journal*, 417:128936, 2021. URL https://doi.org/10.1016/j.cej.2021.128936.

[79] Yvana D. Ahdab, Georg Schücking, Danyal Rehman, and John H. Lienhard. Treatment of greenhouse wastewater for reuse or disposal using monovalent selective electrodialysis. *Desalination*, 507:115037, 2021. URL https://doi.org/10.1016/j.desal.2021.115037.

[80] María L. Vera, Walter R. Torres, Claudia I. Galli, Alexandre Chagnes, and Victoria Flexer. Environmental impact of direct lithium extraction from brines. *Nature Reviews Earth & Environment*, 4(3):149–165, 2023. URL https://doi.org/10.1038/s43017-022-00387-5.

[81] Lei Wu, Changyong Zhang, Seoni Kim, T. Alan Hatton, Hengliang Mo, and T. David Waite. Lithium recovery using electrochemical technologies: Advances and challenges. *Water Research*, 221:118822, 2022. URL https://doi.org/10.1016/j.watres.2022.118822.

[82] Stéphane d'Ascoli, Hugo Touvron, Matthew L Leavitt, Ari S Morcos, Giulio Biroli, and Levent Sagun. Convit: Improving vision transformers with soft convolutional inductive biases. In *International Conference on Machine Learning*, pages 2286–2296. PMLR, 2021. URL https://arxiv.org/abs/2103.10697.

[83] Yongyue Zhou, Yangmin Ren, Mingcan Cui, Fengshi Guo, Shiyu Sun, Junjun Ma, Zhengchang Han, and Jeehyeong Khim. Sonoelectrochemical system mechanisms, design, and machine learning for predicting degradation kinetic constants of pharmaceutical pollutants. *Chemical Engineering Journal*, 478:147266, 2023. URL https://doi.org/10.1016/j.cej.2023.147266.

[84] Li Wang, Jinlong He, Mohammad Heiranian, Hanqing Fan, Lianfa Song, Ying Li, and Menachem Elimelech. Water transport in reverse osmosis membranes is governed by pore flow, not a solution-diffusion mechanism. *Science Advances*, 9(15):eadf8488, 2023. URL https://doi.org/10.1126/sciadv.adf8488.

[85] James L. Higdon and George P. Muldowney. Resistance functions for spherical particles, droplets and bubbles in cylindrical tubes. *Journal of Fluid Mechanics*, 298:193–210, 1995. URL https://doi.org/10.1017/S0022112095003272.

[86] Gretchen M. Mavrovouniotis and Howard Brenner. Hindered sedimentation, diffusion, and dispersion coefficients for brownian spheres in circular cylindrical pores. *Journal of Colloid and Interface Science*, 124(1):269–283, 1988. URL https://doi.org/10.1016/0021-9797(88)90348-7.

[87] Pablo Márquez-Neila, Mathieu Salzmann, and Pascal Fua. Imposing hard constraints on deep networks: Promises and limitations, 2017. URL https://arxiv.org/abs/1706.02025.

[88] Dimitri P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods (Optimization and Neural Computation Series)*. Athena Scientific, 1 edition, 1996. ISBN 1886529043. URL https://www.mit.edu/~dimitrib/Constrained-Opt.pdf.

[89] Byron B. Bird, Warren E. Stewart, Edwin N. Lightfoot, and Robert E. Meredith. Transport phenomena. *Journal of The Electrochemical Society*, 108(3):78C, mar 1961. URL https://dx.doi.org/10.1149/1.2428074.

[90] Endre Nagy, Imre Hegedüs, Danyal Rehman, Quantum J. Wei, Yvana D. Ahdab, and John H. Lienhard. The need for accurate osmotic pressure and mass transfer resistances in modeling osmotically driven membrane processes. *Membranes*, 11(2), 2021. ISSN 2077-0375. URL https://doi.org/10.3390/membranes11020128.

[91] Karan H. Mistry, Harrison A. Hunter, and John H. Lienhard. Effect of composition and nonideal solution behavior on desalination calculations

for mixed electrolyte solutions with comparison to seawater. *Desalination*, 318:34–47, 2013. URL https://doi.org/10.1016/j.desal.2013.03.015.

[92] Kenneth S. Pitzer. Thermodynamics of electrolytes. I. Theoretical basis and general equations. *The Journal of Physical Chemistry*, 77(2):268–277, 1973. URL https://doi.org/10.1021/j100621a026.

[93] Danyal Rehman, Yvana D. Ahdab, and John H. Lienhard. Monovalent selective electrodialysis: Modelling multi-ionic transport across selective membranes. *Water Research*, 199:117171, 2021. URL https://doi.org/10.1016/j.watres.2021.117171.

[94] Manas Ranjan Puhan, Bhaumik Sutariya, and Santanu Karan. Revisiting the alkali hydrolysis of polyamide nanofiltration membranes. *Journal of Membrane Science*, 661:120887, 2022. URL https://doi.org/10.1016/j.memsci.2022.120887.

[95] Chen Xu, Ba Trung Cao, Yong Yuan, and Günther Meschke. Transfer learning based physics-informed neural networks for solving inverse problems in engineering structures under different loading scenarios. *Computer Methods in Applied Mechanics and Engineering*, 405:115852, 2023. URL https://doi.org/10.1016/j.cma.2022.115852.

[96] Shaan Desai, Marios Mattheakis, Hayden Joy, Pavlos Protopapas, and Stephen Roberts. One-shot transfer learning of physics-informed neural networks, 2022. URL https://arxiv.org/abs/2110.11286.

[97] Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. BoTorch: A framework for efficient Monte-Carlo Bayesian optimization, 2020. URL https://arxiv.org/abs/1910.06403.

[98] Chengyu Dong, Liyuan Liu, Zichao Li, and Jingbo Shang. Towards adaptive residual network training: A neural-ODE perspective. volume 119, pages 2616–2626. PMLR, 2020. URL https://proceedings.mlr.press/v119/dong20c.html.

[99] David Rolnick and Max Tegmark. The power of deeper networks for expressing natural functions, 2018. URL https://arxiv.org/abs/1705.05502.

[100] Rahul Yedida, Snehanshu Saha, and Tejas Prashanth. LipschitzLR: Using theoretically computed adaptive learning rates for fast convergence, 2020. URL https://arxiv.org/abs/1902.07399.

[101] Jonathan Ennis, Heyou Zhang, Geoff W. Stevens, Jilska M. Perera, Peter J. Scales, and Steven L. Carnie. Mobility of protein through a porous membrane. *Journal of Membrane Science*, 119(1):47–58, 1996. URL https://doi.org/10.1016/0376-7388(96)00112-3.

[102] Peter M. Bungay and Howard Brenner. Pressure drop due to the motion of a sphere near the wall bounding a poiseuille flow. *Journal of Fluid Mechanics*, 60(1):81–96, 1973. URL https://doi.org/10.1017/S0022112073000054.

[103] Arindam K. Das, Murli Manohar, and Vinod K. Shahi. Cation-exchange membrane with low frictional coefficient and high limiting current density for energy-efficient water desalination. *ACS Omega*, 3(8):10331–10340, 2018. URL https://doi.org/10.1021/acsomega.8b01403.

[104] Yvana D. Ahdab, Danyal Rehman, and John H. Lienhard. Brackish water desalination for greenhouses: Improving groundwater quality for irrigation using monovalent selective electrodialysis reversal. *Journal of Membrane Science*, 610:118072, 2020. URL https://doi.org/10.1016/j.memsci.2020.118072.

[105] Maarten P. Biesheuvel, Li Zhang, Pauline Gasquet, Bastiaan Blankert, Menachem Elimelech, and Walter G.J. van der Meer. Ion selectivity in brackish water desalination by reverse osmosis: Theory, measurements, and implications. *Environmental Science & Technology Letters*, 7(1):42–47, 2020. URL https://doi.org/10.1021/acs.estlett.9b00686.

[106] Yvana D. Ahdab, Georg Schücking, Danyal Rehman, and John H. Lienhard. Cost effectiveness of conventionally and solar powered monovalent selective electrodialysis for seawater desalination in greenhouses. *Applied Energy*, 301:117425, 2021. URL https://doi.org/10.1016/j.apenergy.2021.117425.