Journal of Membrane Science Letters 6 (2026) 100111

Contents lists available at ScienceDirect

Journal of Membrane Science Letters

journal homepage: www.elsevier.com/locate/memlet

Decoupling solvent features to address spurious correlations in ceramic Organic Solvent Nanofiltration membranes

Wout Linsen [a,b,1], Pieter-Jan Piccard [a,b,*,1], Jef Hooyberghs [a], Anita Buekenhoudt [b]

[a] UHasselt—Hasselt University, Data Science Institute, Theory Lab, Agoralaan, 3590, Diepenbeek Belgium
[b] VITO N.V.—Flemish Institute of Technological Research, Unit MATCH, Boeretang 200, 2400 Mol, Belgium

ARTICLE INFO

Keywords:
Organic solvent nanofiltration
Data science
Ceramic membranes
Spurious correlations
Solvent–solute–membrane interactions

ABSTRACT

Organic Solvent Nanofiltration has emerged as an energy-efficient alternative to traditional thermal methods; yet its widespread implementation is hindered by its poorly understood transport mechanism. Improving the prediction of membrane flux and retention is therefore essential, which has recently been primarily advanced through data-driven modeling. Prediction and interpretation of organic solvent nanofiltration transport, however, are complicated by the collinearity between solvent size and Hansen solubility, as they are dependent on one another for common organic solvent nanofiltration solvents. Considering both solvent size and solubility are known to correlate strongly with flux, the collinearity of these properties obscures the impact of either of them. We break this collinearity by performing flux measurements on outlier solvents in unmodified and methyl-grafted ceramic membranes at room temperature. Breaking collinearity is achieved using propylene carbonate and glycerol-water mixtures, whose molecular size and solubility (at room temperature) significantly deviate from the trend established by common organic solvent nanofiltration solvents. To quantitatively identify the true driver of flux for each membrane separately, linear models using either solvent kinetic diameter, molar volume, or Hansen solubility as predictors are compared using statistical tests. Our analysis indicates that molecular size (specifically, the squared reciprocal kinetic diameter) is the true predictor of flux for unmodified and methyl-grafted titania membranes separately, while molar volume and Hansen solubility add no further predictive power. This method can be extended to, e.g., investigate membrane and temperature dependence. Decoupling solvent size and Hansen solubility can help improve the understanding of organic solvent nanofiltration transport and, via dimensionality reduction, aid the development of data-driven modeling.

1. Introduction

While membrane technology is a recognized energy-efficient alternative to thermal separation (Sholl and Lively, 2016; Rundquist et al., 2012; Adler et al., 2000; Obotey Ezugbe and Rathilal, 2020), its application within organic solvent nanofiltration (OSN) is hindered by the complex interactions between solvent, solute, and membrane (Galizia and Bye, 2018; Ignacz et al., 2023b). Capturing these interactions more effectively is critical, as the industrial viability of any separation process relies heavily on the precise prediction of its key performance metrics.

The performance of membranes in a given separation task is measured by retention and flux. Knowing these measures is essential for industrial implementation, but requires thorough and tedious experimentation. Therefore, predicting retention and flux is crucial and drives extensive research. Traditionally, prediction has been pursued through mechanistic models, attempting to describe the underlying physical processes. However, these models often fall short, either failing to provide accurate predictions or relying on empirical coefficients fitted from experimental data rather than first principles (Marchetti et al., 2014; Vandezande et al., 2008; Piccard et al., 2023). Molecular dynamics simulations have recently demonstrated potential in explaining fundamental transport mechanisms, such as the mediating role of water in solvent sorption (Mahmud et al., 2025) or the interaction between solvent and grafted oligomers (Kyriakou et al., 2023). To bypass the complexity of physical descriptions, and fueled by recent advancements in artificial intelligence and data collection (Van Buggenhout et al., 2025; Ignacz et al., 2022), the field is increasingly turning towards supervised data-driven approaches (Piccard et al., 2023). Data-driven approaches have been used to extract physically interpretable

* Corresponding author at: UHasselt—Hasselt University, Data Science Institute, Theory Lab, Agoralaan, 3590, Diepenbeek Belgium.
E-mail address: pieterjan.piccard@uhasselt.be (P.-J. Piccard).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.memlet.2026.100111
Received 15 December 2025; Received in revised form 9 January 2026; Accepted 12 January 2026
Available online 21 January 2026

trends in solvent–solute–membrane interactions and to identify dominant physicochemical descriptors governing OSN performance (Ignacz et al., 2023c; Piccard et al., 2025; Hu et al., 2021; Ignacz et al., 2023a). Machine-learning models are employed for direct prediction of membrane performance across diverse systems (Hu et al., 2021; Ignacz et al., 2023a, 2025; Lee et al., 2023; Goebel and Skiborowski, 2020; Goebel et al., 2020; Kim et al., 2021; Wang et al., 2023; Gallo-Molina et al., 2023), enabling, e.g., process optimization, exploration of large design spaces, or membrane material development and selection. These approaches offer a powerful alternative for understanding and predicting the complex performance of membranes.

Data-driven modeling introduces its own set of challenges. In high-dimensional, data-scarce regimes, models suffer from the curse of dimensionality. They are then prone to overfitting, resulting in poor generalization on unseen conditions. Hybrid modeling can reduce degrees of freedom and thus overfitting. Similarly, selecting a compact set of informative features can improve generalization (Chollet and Chollet, 2021). A related challenge is spurious correlations: when features are highly correlated, separate interpretation becomes impossible. Models may value features appearing informative only due to correlations with underlying drivers. Retaining spurious features undermines hybrid modeling and inflates dimensionality, hurting generalization. Relying on correlation introduces bias, causing failure when extrapolating to unexplored regions where this correlation may not be present. Consequently, removing spurious correlations is essential for robust data-driven models (Wang and Jordan, 2024; Altman and Krzywinski, 2018).

A potential pitfall that can hinder OSN prediction and interpretability was exposed in a careful investigation of these features by a previous study (Piccard et al., 2025). They identified three solvent properties that strongly correlate with flux, but are also highly mutually *collinear* for solvents commonly used in OSN, here referred to as "*traditional solvents*". In other words, because these properties tend to vary simultaneously across the dataset, it becomes mathematically difficult to disentangle which specific property is actually driving the performance in OSN. This raises the possibility of *spurious correlations* (apparent relationships lacking genuine physical causation) among these solvents. Consequently, it is conceivable that one or more of these features possess little to no genuine predictive capability, and that their apparent significance arises merely as an artifact of the training data, a risk that extends to many studies utilizing similar solvents.

The primary objective of this study is therefore to decouple the effects of multiple solvent properties, specifically size and solubility descriptors, on flux to quantitatively identify and eliminate potential spurious correlations in unmodified and methyl-grafted titania membranes. This work breaks the collinearity between the solvent properties by introducing a targeted dataset of "*untraditional solvents*", specifically selected because their properties deviate from the previously observed correlations. By analyzing this decoupled dataset, we will evaluate the differences in predictive power when modeling flux, using different feature sets. If a feature fails to provide additional predictive value for these outlier solvents, it indicates that the previously observed linear correlations were likely spurious. Consequently, excluding this feature can improve the generalization capability of future models.

2. Materials and methods

2.1. Experiments

Experiments were performed on native and methyl-grafted 0.9 nm TiO_2 membranes at room temperature. The native membranes are commercially available (manufactured by Inopor), while the methyl-grafted versions were modified from their native counterparts in-house using our proprietary method of Grignard chemistry (Buekenhoudt et al., 2015; Van Heetvelde et al., 2013; Mustafa et al., 2016). A rationale for why these are selected and a detailed explanation of the setup are provided in the supplementary information (SI). The volumetric flux F was measured and reported in L/(h· m^2).

2.2. Solvent size and Hansen solubility in OSN

Both solvent size descriptors (in this study, *Molar Volume* v and *Kinetic Diameter* d_{kin}), as well as *Hansen solubility* δ_{tot} (Hansen, 2007), have been valuable for both predicting and interpreting OSN transport. Therefore, it would be valuable to disentangle the effect of both.

Hansen solubility parameters are commonly used to interpret and model OSN, through both mechanistic (Marchetti et al., 2014; Buekenhoudt et al., 2013; Karan et al., 2015), and data-driven methods (Ignacz et al., 2023c; Goebel and Skiborowski, 2020; Goebel et al., 2020; Hu et al., 2021; Lee et al., 2023; Gallo-Molina et al., 2023). Hansen solubility can also provide valuable interpretation: when combined with the membrane and solute solubility, information on the affinity among them is revealed, which can, in turn, be used to deduce flux behavior.

Both molar volume and kinetic diameter describe the size of solvents. While molar volume measures the bulk solvent size, kinetic diameter is a more direct measure of molecular size. Molar volume commonly and naturally emerges in models of membrane transport when thermodynamic relations are employed (Mulder, 1996; Mason and Lonsdale, 1990; Piccard et al., 2023). Kinetic diameter is instead a more direct measure of size (similar to molecular weight for mass), and can be used to describe steric effects (Piccard et al., 2025). Its role (or that of related direct size measures) in mechanistic modeling is more limited, although it has been used by Darvishmanesh et al. in their resistance-in-series model (Darvishmanesh et al., 2009).

2.3. Normalization of flux

To reliably compare experimental data and isolate the specific behavior of flux in relation to the features under study, it is helpful to eliminate the influence of other variables. For instance, flux is known to be strongly influenced by transmembrane pressure and solvent viscosity with known dependencies. To partly account for this, we can fall back on theoretical models for flux based on hydrodynamics, such as the Hagen–Poiseuille equation that predicts flux inversely proportional to solvent viscosity η, and proportional to trans-membrane pressure Δp, according to Mulder (1996):

$$F = \frac{A}{8\pi\tau z}\frac{\Delta p}{\eta} \quad \text{(Hagen–Poiseuille)}, \tag{1}$$

with pore cross-sectional area A, tortuosity τ, and pore length z. Thus, we can account for these known dependencies by dividing flux by $\Delta p/\eta$:

$$F \cdot \frac{\eta}{\Delta p} \quad \text{(To account for known hydrodynamics)}. \tag{2}$$

This adjusted flux yields a significantly higher correlation with the solvent features of interest (Piccard et al., 2025).

To further account for slight structural heterogeneity between individual membranes of the same type, the data is scaled against a reference standard: the pure water flux (Buekenhoudt et al., 2013). By creating a ratio of the solvent permeability to the pure water permeability, we obtain a dimensionless, normalized flux F_{norm}, expressed as (Buekenhoudt et al., 2013; Piccard et al., 2025):

$$F_{norm} = \left[F \cdot \frac{\eta}{\Delta p}\right] / \left[F_{H_2O} \cdot \frac{\eta_{H_2O}}{\Delta p_{H_2O}}\right] \tag{3}$$

Through this non-dimensionalization, experimental results obtained under varying transmembrane pressures, distinct membrane samples, and solvents of differing viscosity can now be effectively compared.

2.4. Properties of binary mixtures

Binary solvent mixtures are treated as effective single-solvents, whose properties are calculated as weighted averages of the individual properties. Molar volume is calculated via mole-fraction-averaging, kinetic diameter via volume-fraction-averaging, and Hansen solubility from the individually volume-fraction-weighted HSP's.

This averaging approach is based on empirical evidence of flux measurements in the 0.9 nm TiO_2 membranes. First, no change in mixture composition was observed during filtration for the titania membranes, indicating that the components permeate as bulk fluid. Second, analysis of traditional experimental flux measurements (detailed in the SI) demonstrates that binary mixtures characterized by weighted-averaged d_{kin}, ν, and δ_{tot} follow the same flux trends as pure solvents.

Not considering membrane transport, the averaging approach is experimentally verified as a good approximation for Hansen solubility (Hansen, 2007), and molar volume in the considered mixtures (Negadi et al., 2017; Bai et al., 1998; Guevara-Carrion et al., 2021; Scharlin et al., 2002). However, applying a volume-fraction-weighted average to the kinetic diameter, being a molecular-scale property rather than a macroscopic one, may not be physically meaningful. Caution is therefore required when treating binary mixtures as effective single solvents, and this limitation is recognized.

For mixture viscosities (used for the calculation of F_{norm} in Eq. (3)), experimental values from literature are used, as this can strongly deviate from a simple weighted average (Thompson et al., 2006; del Carmen Grande et al., 2007; Bernal-García et al., 2008).

2.5. Statistical method

To evaluate the predictive power of ν, d_{kin}, and δ_{tot}, linear regression models were created using each feature separately as the predictor, and normalized flux F_{norm} as target, making no distinction between "traditional" and "untraditional" data. To ensure that the performance of these linear models is statistically robust (and not a result of chance), each model was recreated 50 times, with each iteration using a different set of test and training data (80%/20% train/test). The models' performance was evaluated using the *Mean Absolute Error* (MAE) of the test set, because of its high interpretability. Moreover, in machine learning, MAE can be used to model experimental data that might contain measurement uncertainty (Hastie et al., 2009). The resulting MAE distributions (across the 50 different iterations) were then analyzed using a one-way *ANOVA* followed by a *Tukey's honest significant difference (HSD)*. The ANOVA simply determines whether at least one of the means of those MAE distributions is significantly different from the others. The Tukey's HSD test then assesses which of the three linear models are significantly different. In essence, this procedure tells us which of ν, d_{kin}, or δ_{tot} is better at predicting flux F_{norm}. However, as the random-split test sets are not completely mutually independent, p-values might be overly optimistic. Such tests are nevertheless often used to evaluate models when data is sparse (Rainio et al., 2024). Therefore, to strengthen the analysis, it is supplemented with Cliff's delta as an effect-size measure, which quantifies the lack of overlap between the MAE distributions (Meissel and Yao, 2024).

To investigate the combined effect of two features, two-dimensional linear regression models were trained using all possible pairs of ν, d_{kin}, or δ_{tot}. The same statistical procedure was followed; however, here, the ANOVA was followed by a *Dunnett's test* to compare the models to the 1D model with the lowest MAE (Hirotsu, 2017). This additional procedure is followed because even if one single feature may be best at predicting flux on its own, it does not exclude possible contributions from the others.

Table 1
Overview of the organic solvents used in this study, as well as water and organic-water mixtures, with their total Hansen solubility δ_{tot}, kinetic diameter d_{kin}, molar volume ν, and viscosity η at room temperature. The mixtures are expressed in volume percentages, and their properties are averaged as described in Section 2.4. (*) A list of binary traditional solvents part of this dataset is provided in the SI. (**) Values for kinetic diameter are gathered from literature (Bowen et al., 2003; Iliyas et al., 2007; Shao and Huang, 2007).

Solvent	δ_{tot} [$MPa^{0.5}$]	d_{kin} [Å]**	ν [cm^3/mol^{-1}]	η [mPa s]
Traditional solvents				
Acetone	20.0	4.6	74.0	0.32
Acetonitrile	24.3	3.4	52.5	0.34
Dichloromethane (DCM)	20.3	4.9	64.1	0.43
Dimethylformamide (DMF)	24.8	5.0	77.0	0.82
Dimethyl sulfoxide (DMSO)	26.7	4.4	71.2	2.00
Ethyl acetate	18.1	5.2	97.8	0.43
Ethanol	26.5	4.4	52.8	1.08
Heptane	15.3	4.3	147.4	0.40
Isopropanol (IPA)	23.5	4.7	76.6	2.15
Methanol	29.6	3.8	40.7	0.54
Tetrahydrofuran (THF)	19.4	4.8	81.1	0.46
Toluene	18.2	5.5	106.3	0.57
Water	47.8	2.6	18.0	0.92
Various binary mixtures (*)				
Solvents and solvent mixtures used in this study				
Water/Glycerol 92.8/7.2	46.81	2.90	19.05	1.11
Water/Glycerol 85/15	45.70	3.17	20.31	1.44
Water/Glycerol 65/35	42.89	3.86	24.47	3.20
Water/Glycerol 50/50	40.82	4.38	28.90	6.86
Propylene Carbonate	27.20	5.10	84.72	2.50

3. Results and discussion

3.1. Beyond traditional solvent property combinations

Size and solubility properties, here represented by ν^{-1} & d_{kin}^{-2} and δ_{tot} respectively, are highly mutually collinear for "traditional solvents" commonly used in OSN (see Table 1). This is clearly illustrated by the gray dots in Fig. 1, where these three features appear to be linearly dependent on one another. These gray dots represent the pure solvents and mixtures among them. This collinearity is no coincidence: chemicals outside these linear trends tend to be either highly viscous liquids or gases at room temperature (Piccard et al., 2025). However, this collinearity limits the extraction of useful information (see Section 2.2), as results may be spurious. Indeed, in the SI, it is quantitatively demonstrated that, *using only traditional solvents, it is not possible to identify whether solvent size (ν^{-1}, d_{kin}^{-2}) or Hansen solubility (δ_{tot}) is a better predictor of flux*.

This study, therefore, identified solvents that lie outside these established trends. Glycerol and propylene carbonate are found to be most suitable outliers (Fig. 1, green cross and blue dot respectively), as they deviate strongly from the trend of the "traditional solvents" in all feature combinations plotted in Fig. 1. Glycerol at room temperature, however, is impractical for experimental use due to its *high viscosity* (which is, at the same time, the reason that it deviates from the linear trend; Glycerol is large and polar, causing high viscosity). For this reason, water–glycerol mixtures are selected instead. These are included in Fig. 1 (red dots), and their properties are listed in Table 1. In the SI, the deviation of these solvents from the trend is shown quantitatively.

3.2. Decoupling the effect of solvent properties on flux

Fig. 2 shows scatterplots of the collinearity-breaking flux measurements in function of solubility and size properties, revealing distinct variations among these features. Previous measurements on "traditional solvents" (gray dots) are also displayed, all of which contain a solute of varying species (explaining the high-variance gray dots).

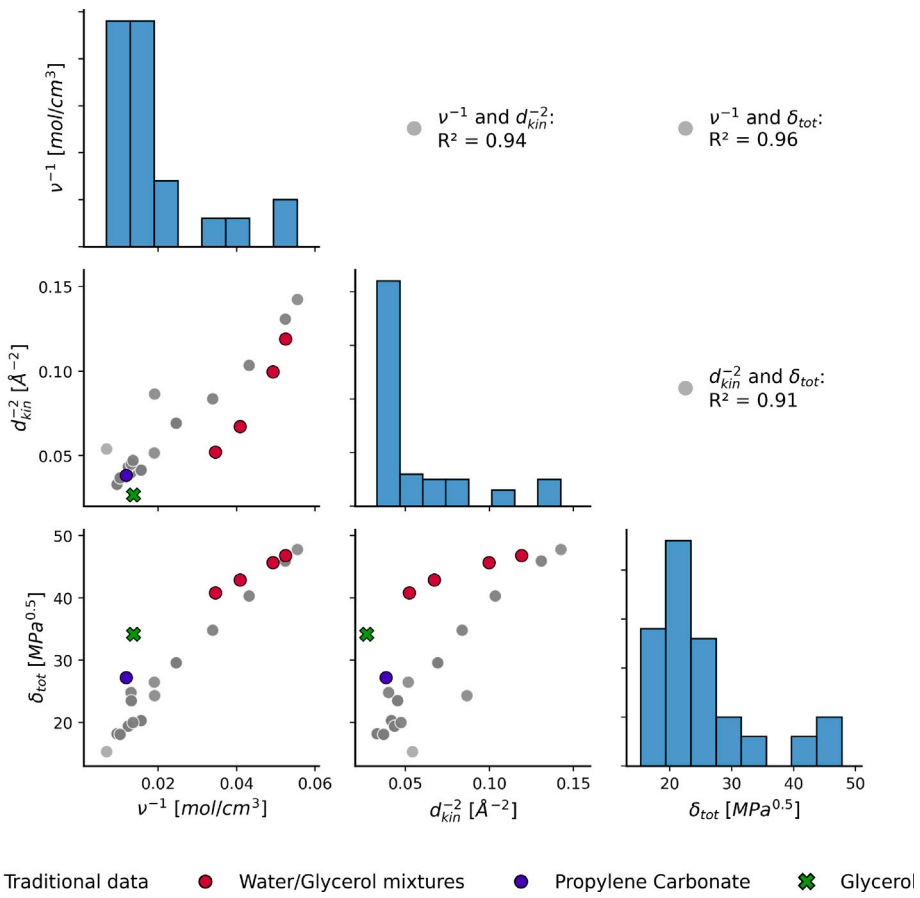

Fig. 1. Scatter-correlation matrix of the used features. The upper-right elements of the matrix present the Pearson correlation coefficient of the "traditional data". The lower triangle presents the scatter plots of the feature combinations. New solvents (mixtures) used for this study are indicated. The main diagonal of the matrix contains histograms of the separate features, also using the "traditional data".

Qualitatively, the scatterplots indicate that d_{kin}^{-2} is a robust linear predictor, as the experimental data for the outlier solvents adhere closely to the linear trend established by the traditional data. In contrast, deviations are clearly pronounced for ν^{-1} and δ_{tot}, where the new solvents diverge from the traditional trend lines. To objectively verify these observations and identify the true driver of flux, we apply the statistical method described in Section 2.5.

3.2.1. 1D statistical analysis

To quantitatively assess which of the three solvent properties (d_{kin}^{-2}, ν^{-1} or δ_{tot}) is the 'true' predictor of flux, we applied our statistical method, described in detail in Section 2.5, on the data presented in Fig. 2. The results of this statistical analysis are displayed in Fig. 3 using boxplots.

Fig. 3A shows the MAE of linear models, trained using the indicated property as predictor. Kinetic diameter (d_{kin}^{-2}) is the best predictor, showing the lowest MAE, followed by molar volume for both native and methyl-grafted membranes. This is quantitatively supported by their p-values determined by Tukey's HSD. This indicates that the MAE of a linear model in function of d_{kin}^{-2} is significantly different compared to either ν^{-1} (with $p = 2.51 \cdot 10^{-8}$ (native) and $p = 2.59 \cdot 10^{-10}$ (methyl-grafted)) and δ_{tot} (with $p = 2.75 \cdot 10^{-14}$ (native) and $p = 2.15 \cdot 10^{-14}$ (methyl-grafted). These conclusions are supported by the values of the Cliff's delta measure, provided between boxplots in Fig. 3A. All values are large, except for the medium value for the native membrane between δ_{tot} and ν (Meissel and Yao, 2024).

Noteworthy about this result is the difference between the inverse molar volume and the inverse square of the kinetic diameter. Both are parameters assumed to describe the size of the solvent. However, the scatter data for both molar volume and HSP deviate significantly from the linear trend in a similar way, although the deviation is stronger for HSP. This can be explained by realizing that molar volume is a measure of bulk solvent size, and the kinetic diameter is a measure of molecular size. A measure of bulk solvent is also dependent on the intermolecular interactions present, which also affect Hansen solubility.

3.2.2. 2D statistical analysis

While this analysis suggests kinetic diameter (specifically, d_{kin}^{-2}) to be the true linear predictor, it does not exclude possible contributions from ν^{-1} and δ_{tot}. Therefore, two-dimensional linear models using these properties in combination with d_{kin}^{-2} are made and compared to the one-dimensional model to see if additional predictive power can be obtained.

The two-dimensional analysis, shown in Fig. 3B, reveals that no combination of features lowers the MAE of the one-dimensional model based on kinetic diameter. No significant difference, compared to the reference 1D model of kinetic diameter, is observed when adding δ_{tot} or ν^{-1}. Also, their Cliff's deltas are negligibly small. This indicates that the other features exert little to no additional influence and d_{kin}^{-2} is the true predictor.

The SI validates the feature exponents used in our linear regression model. Specifically, we show that varying the exponents for Hansen solubility yields consistent results, confirming that our main conclusions are not dependent on this specific parameterization.

4. Conclusion

Solvent features in OSN are decoupled by employing outlier solvents to break previously observed collinearity. Properties of solvents

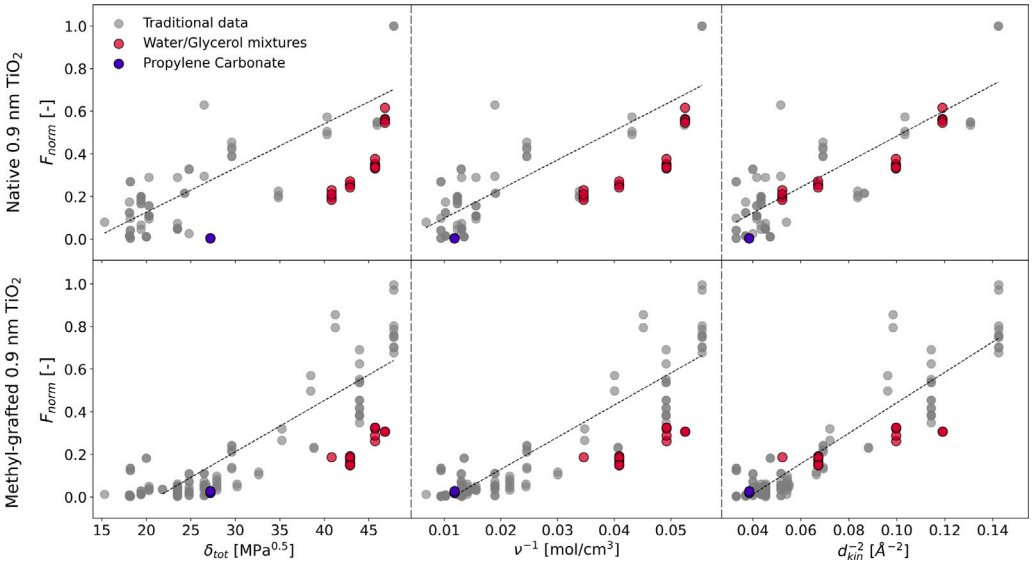

Fig. 2. Normalized flux as a function of the Hansen solubility parameter (left column), inverse molar volume (middle column), and the inverse squared solvent kinetic diameter (right column) for a native 0.9 nm TiO_2 membrane (upper row), and a methyl-grafted version of the same membrane (bottom row). The dashed trend lines are linear fits constructed from only the "traditional data".

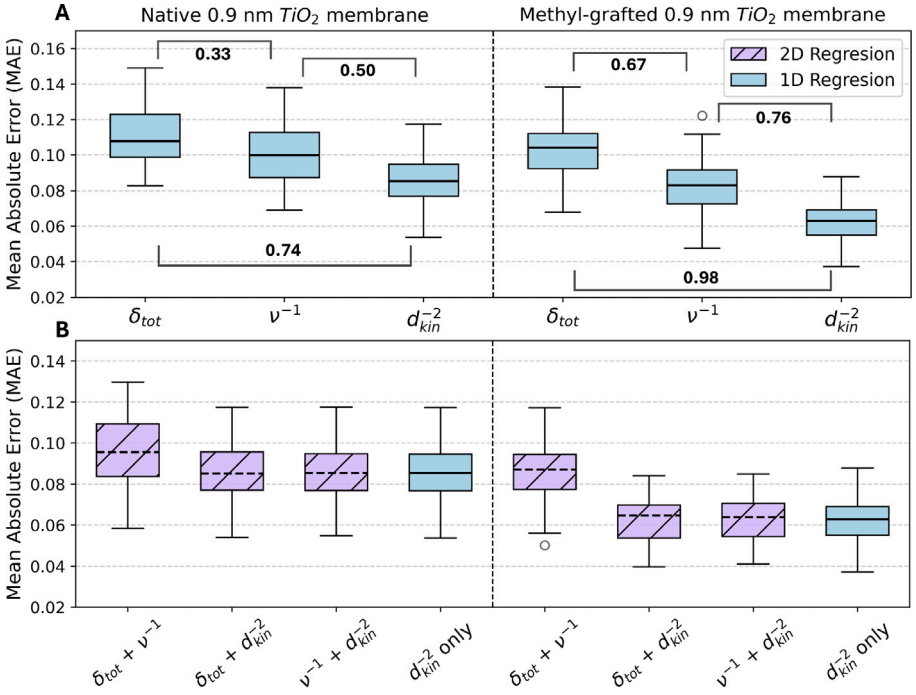

Fig. 3. Boxplots displaying the distribution of the MAE of linear regression models trained using the indicated properties as flux predictors for a native 0.9 nm TiO_2 membrane (left) and its methyl-grafted counterpart (right). Top **(A)**: 1D linear regression models using a single input, supplemented with the Cliff's delta values. Bottom **(B)**: 2D linear regression models using two predictors (for comparison, the 1D boxplot using only d_{kin}^{-2} is repeated).

describing size (e.g., molar volume, kinetic diameter) or solubility (e.g., Hansen solubility) are known to correlate highly with flux, but because they are also mutually dependent on one another, it was not possible to distinguish their effects. Here, we designed experiments that decouple this collinearity and demonstrate that the kinetic diameter (specifically, d_{kin}^{-2}) is the primary predictor of flux for a fixed unmodified and methyl-grafted 0.9 nm TiO_2 membrane at room temperature, significantly outperforming Hansen solubility (δ_{tot}) and molar volume (specifically, ν^{-1}).

Decoupling solvent size and Hansen solubility can help improve the understanding of OSN transport and, via dimensionality reduction, aid the development of data-driven modeling. Importantly, when OSN interpretation is implemented within data-driven modeling, i.e., in hybrid models, this understanding is crucial to consider.

An interesting future endeavor is to assess whether this conclusion holds in general, or is specific to the unmodified and methyl-grafted 0.9 nm TiO_2 membranes separately. Particularly, Hansen solubility may be powerful in modeling membrane dependence, which was not investigated in this study.

CRediT authorship contribution statement

Wout Linsen: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Pieter-Jan Piccard:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Conceptualization. **Jef Hooyberghs:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Anita Buekenhoudt:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.memlet.2026.100111.

Data availability

The data of flux measurements are available in the Supplementary Information.

References

Adler, S., Beaver, E., Bryan, P., Robinson, S., Watson, J., 2000. Vision 2020: 2000 Separations Roadmap. Tech. Rep., EERE Publication and Product Library, Washington, DC (United States).

Altman, N., Krzywinski, M., 2018. The curse(s) of dimensionality. Nature Methods 15, 399–400. http://dx.doi.org/10.1038/s41592-018-0019-x.

Bai, T.-C., Yao, J., Han, S.-J., 1998. Excess molar volumes for binary and ternary mixtures of (n, n-dimethylformamide+ ethanol+ water) at the temperature 298.15 k. J. Chem. Thermodyn. 30 (11), 1347–1361.

Bernal-García, J.M., Guzmán-López, A., Cabrales-Torres, A., Estrada-Baltazar, A., Iglesias-Silva, G.A., 2008. Densities and viscosities of (n, n-dimethylformamide+ water) at atmospheric pressure from (283.15 to 353.15) k. J. Chem. Eng. Data 53 (4), 1024–1027. http://dx.doi.org/10.1021/je700671t.

Bowen, T.C., Kalipcilar, H., Falconer, J.L., Noble, R.D., 2003. Pervaporation of organic/water mixtures through b-zsm-5 zeolite membranes on monolith supports. J. Membr. Sci. 215 (1–2), 235–247.

Buekenhoudt, A., Bisignano, F., De Luca, G., Vandezande, P., Wouters, M., Verhulst, K., 2013. Unravelling the solvent flux behaviour of ceramic nanofiltration and ultrafiltration membranes. J. Membr. Sci. 439, 36–47.

Buekenhoudt, A., Wyns, K., Meynen, V., Maes, B., Cool, P., 2015. Surface modified inorganic matrix and method for preparation thereof. URL https://patents.google.com/patent/US8980096B2/en.

Chollet, F., Chollet, F., 2021. Deep Learning with Python. Simon and Schuster.

Darvishmanesh, S., Buekenhoudt, A., Degrève, J., Van der Bruggen, B., 2009. Coupled series–parallel resistance model for transport of solvent through inorganic nanofiltration membranes. Sep. Purif. Technol. 70 (1), 46–52.

del Carmen Grande, M., Juliá, J.A., García, M., Marschoff, C.M., 2007. On the density and viscosity of (water+ dimethylsulphoxide) binary mixtures. J. Chem. Thermodyn. 39 (7), 1049–1056. http://dx.doi.org/10.1016/j.jct.2006.12.012.

Galizia, M., Bye, K.P., 2018. Advances in organic solvent nanofiltration rely on physical chemistry and polymer chemistry. Front. Chem. 6, 511.

Gallo-Molina, J.P., Claessens, B., Buekenhoudt, A., Verliefde, A., Nopens, I., 2023. Capturing unmodelled phenomena: A hybrid approach for the prediction of the transport through ceramic membranes in organic solvent nanofiltration. J. Membr. Sci. 686, 122024.

Goebel, R., Glaser, T., Skiborowski, M., 2020. Machine-based learning of predictive models in organic solvent nanofiltration: Solute rejection in pure and mixed solvents. Sep. Purif. Technol. 248, 117046.

Goebel, R., Skiborowski, M., 2020. Machine-based learning of predictive models in organic solvent nanofiltration: Pure and mixed solvent flux. Sep. Purif. Technol. 237, 116363.

Guevara-Carrion, G., Fingerhut, R., Vrabec, J., 2021. Density and partial molar volumes of the liquid mixture water+ methanol+ ethanol+ 2-propanol at 298.15 k and 0.1 mpa. J. Chem. Eng. Data 66 (6), 2425–2435.

Hansen, C.M., 2007. Hansen Solubility Parameters: A User's Handbook. CRC Press.

Hastie, T., Tibshirani, R., Friedman, J., et al., 2009. The elements of statistical learning.

Hirotsu, C., 2017. Advanced Analysis of Variance, first ed. In: Wiley Series in Probability and Statistics, John Wiley & Sons.

Hu, J., Kim, C., Halasz, P., Kim, J.F., Kim, J., Szekely, G., 2021. Artificial intelligence for performance prediction of organic solvent nanofiltration membranes. J. Membr. Sci. 619, 118513. http://dx.doi.org/10.1016/j.memsci.2020.118513.

Ignacz, G., Alqadhi, N., Szekely, G., 2023a. Explainable machine learning for unraveling solvent effects in polyimide organic solvent nanofiltration membranes. Adv. Membr. 3, 100061.

Ignacz, G., Beke, A.K., Szekely, G., 2023b. Data-driven future for nanofiltration: Escaping linearity. J. Membr. Sci. Lett. 3 (1), 100040.

Ignacz, G., Beke, A.K., Szekely, G., 2023c. Data-driven investigation of process solvent and membrane material on organic solvent nanofiltration. J. Membr. Sci. 674, 121519.

Ignacz, G., Beke, A.K., Toth, V., Szekely, G., 2025. A hybrid modelling approach to compare chemical separation technologies in terms of energy consumption and carbon dioxide emissions. Nat. Energy 10 (3), 308–317.

Ignacz, G., Yang, C., Szekely, G., 2022. Diversity matters: Widening the chemical space in organic solvent nanofiltration. J. Membr. Sci. 641, 119929.

Iliyas, A., Eić, M., Zahedi-Niaki, M.H., Vasenkov, S., 2007. Towards observation of single-file diffusion using tzlc. Diffus. Fundam. 6.

Karan, S., Jiang, Z., Livingston, A.G., 2015. Sub–10 nm polyamide nanofilms with ultrafast solvent transport for molecular separation. Science 348 (6241), 1347–1351.

Kim, C., You, C., Park, M., Jang, D., Lee, S., Kim, J., et al., 2021. Machine learning-based approach to identify the optimal design and operation condition of organic solvent nanofiltration (osn). In: Computer Aided Chemical Engineering, vol. 50. Elsevier, pp. 933–938.

Kyriakou, N., Boorsma, E., Aardema, G.-J., van Eck, G.R., Drobek, M., de Beer, S., Nijmeijer, A., Winnubst, L., Pizzoccaro-Zilamy, M.-A., 2023. Hybrid ceramic nanofiltration membranes prepared by impregnation and solid-state grafting of organo-phosphonic acids. J. Membr. Sci. 687, 122041.

Lee, Y.J., Chen, L., Nistane, J., Jang, H.Y., Weber, D.J., Scott, J.K., Rangnekar, N.D., Marshall, B.D., Li, W., Johnson, J., et al., 2023. Data-driven predictions of complex organic mixture permeation in polymer membranes. Nat. Commun. 14 (1), 4931.

Mahmud, M., Van der Bruggen, B., Artzner, F., Ghoufi, A., Szymczyk, A., 2025. Deciphering the role of water in ethanol uptake within pim-1 membranes. J. Membr. Sci. 124314.

Marchetti, P., Jimenez Solomon, M.F., Szekely, G., Livingston, A.G., 2014. Molecular separation with organic solvent nanofiltration: a critical review. Chem. Rev. 114 (21), 10735–10806.

Mason, E., Lonsdale, H., 1990. Statistical–mechanical theory of membrane transport. J. Membr. Sci. 51 (1–2), 1–81.

Meissel, K., Yao, E.S., 2024. Using cliff's delta as a non-parametric effect size measure: an accessible web app and r tutorial. Pr. Assess. Res. Eval. 29 (1).

Mulder, M., 1996. Basic Principles of Membrane Technology, second ed. Kluwer Academic Publishers, Dordrecht, The Netherlands, http://dx.doi.org/10.1007/978-94-009-1766-8.

Mustafa, G., Wyns, K., Buekenhoudt, A., Meynen, V., 2016. New insights into the fouling mechanism of dissolved organic matter applying nanofiltration membranes with a variety of surface chemistries. Water Res. 93, 195–204.

Negadi, L., Feddal-Benabed, B., Bahadur, I., Saab, J., Zaoui-Djelloul-Daouadji, M., Ramjugernath, D., Negadi, A., 2017. Effect of temperature on density, sound velocity, and their derived properties for the binary systems glycerol with water or alcohols. J. Chem. Thermodyn. 109, 124–136.

Obotey Ezugbe, E., Rathilal, S., 2020. Membrane technologies in wastewater treatment: A review. Membranes 10, http://dx.doi.org/10.3390/membranes10050089.

Piccard, P.-J., Borges, P., Cleuren, B., Buekenhoudt, A., Hooyberghs, J., 2025. Exploratory data analysis reveals the impact of solvent and solute on the performance of native and grafted ceramic membranes in organic solvent nanofiltration. J. Membr. Sci. 735, 124509. http://dx.doi.org/10.1016/j.memsci.2025.124509.

Piccard, P.-J., Borges, P., Cleuren, B., Hooyberghs, J., Buekenhoudt, A., 2023. Organic solvent nanofiltration and data-driven approaches. Separations 10 (9), http://dx.doi.org/10.3390/separations10090516.

Rainio, O., Teuho, J., Klén, R., 2024. Evaluation metrics and statistical tests for machine learning. Sci. Rep. 14 (1), 6086.

Rundquist, E.M., Pink, C.J., Livingston, A.G., 2012. Organic solvent nanofiltration: a potential alternative to distillation for solvent recovery from crystallisation mother liquors. Green Chem. 14 (8), 2197–2205.

Scharlin, P., Steinby, K., Domańska, U., 2002. Volumetric properties of binary mixtures of n, n-dimethylformamide with water or water-d2 at temperatures from 277.13 k to 318.15 k. J. Chem. Thermodyn. 34 (6), 927–957.

Shao, P., Huang, R., 2007. Polymeric membrane pervaporation. J. Membr. Sci. 287 (2), 162–179.

Sholl, D., Lively, R., 2016. Seven chemical separations to change the world. Nature 532, 435–437. http://dx.doi.org/10.1038/532435a.

Thompson, J.W., Kaiser, T.J., Jorgenson, J.W., 2006. Viscosity measurements of methanol–water and acetonitrile–water mixtures at pressures up to 3500 bar using a novel capillary time-of flight viscometer. J. Chromatogr. A 1134 (1–2), 201–209. http://dx.doi.org/10.1016/j.chroma.2006.09.006.

Van Buggenhout, S., Ignacz, G., Caspers, S., Dhondt, R., Lenaerts, M., Lenaerts, N., Hosseinabadi, S.R., Nulens, I., Koeckelberghs, G., Ren, Y., et al., 2025. Open and fair data for nanofiltration in organic media: A unified approach. J. Membr. Sci. 713, 123356.

Van Heetvelde, P., Beyers, E., Wyns, K., Adriaensens, P., Maes, B., Mullens, S., Buekenhoudt, A., Meynen, V., 2013. A new method to graft titania using grignard reagents. Chem. Commun. 49 (62), 6998–7000.

Vandezande, P., Gevers, L.E., Vankelecom, I.F., 2008. Solvent resistant nanofiltration: separating on a molecular level. Chem. Soc. Rev. 37 (2), 365–405.

Wang, Y., Jordan, M.I., 2024. Desiderata for representation learning: A causal perspective. J. Mach. Learn. Res. 25 (275), 1–65, URL http://jmlr.org/papers/v25/21-107.html.

Wang, C., Wang, L., Soo, A., Pathak, N.B., Shon, H.K., 2023. Machine learning based prediction and optimization of thin film nanocomposite membranes for organic solvent nanofiltration. Sep. Purif. Technol. 304, 122328.