



OPEN

Nearest neighbor permutation entropy detects phase transitions in complex high-pressure systems

Arthur A. B. Pessa^{1✉}, Leonardo G. J. M. Voltarelli¹, Lúcio Cardozo-Filho^{2,3,4},
Andres G. M. Tamara², Claudio Dariva⁵, Papa M. Ndiaye³, Matjaž Perc^{6,7,8,9,10✉} &
Haroldo V. Ribeiro^{1✉}

Understanding the high-pressure phase behavior of carbon dioxide–hydrocarbon mixtures is of considerable interest owing to their wide range of applications. Under certain conditions, these systems are not amenable to direct visual monitoring, and experimentalists often rely on spectrophotometric data to infer phase behavior. Consequently, developing computationally efficient and robust methods to leverage such data is crucial. Here, we combine nearest neighbor permutation entropy, computed directly from *in situ* near-infrared absorbance spectra acquired during depressurization trials of mixtures of carbon dioxide and a distilled petroleum fraction, with an anomaly detection approach to identify phase transitions. We show that changes in nearest neighbor entropy effectively signal transitions from initially homogeneous mixtures to two-phase equilibria, thereby enabling accurate out-of-sample online predictions of transition pressures. Our approach requires minimum data preprocessing, no specialized detection techniques or visual inspection of the spectra, and is sufficiently general to be adapted for studying phase behavior in other high-pressure systems monitored via spectrophotometry.

Carbon dioxide flooding is widely employed as a miscible displacement technique to enhance oil recovery through mechanisms such as crude vaporization, oil swelling, viscosity reduction, and interfacial tension modification^{1,2}. Furthermore, this technology facilitates the underground storage of substantial quantities of CO₂, thereby contributing to carbon footprint reduction^{3,4}. Conversely, many pre-salt reservoirs along the Brazilian coast contain high concentrations of carbon dioxide, rendering high-pressure phase-equilibrium data essential for petroleum reservoir exploration and planning. Moreover, mixtures of CO₂ and hydrocarbons typically exhibit complex phenomena, including liquid–liquid and vapor–liquid–liquid equilibria, and in some cases, homogeneous mixtures cannot be achieved even at very high pressures^{5–7}.

Traditional approaches to understanding these phase behaviors have typically involved thermodynamic analyses of variations in pressure, volume, and temperature^{8–10}. However, the direct visualization methods employed in these approaches are limited by the opacity of certain crude oils. Studies indicate that most crude oils exhibit a minimum absorbance near 1.6 μm , rendering visualization techniques operating in the visible spectrum inefficient for some assessments^{11,12}. Despite the extensive body of work on the phase behavior of hydrocarbon mixtures, there remains a need to develop new tools and strategies to probe the complex physicochemical phenomena involved.

In this context, the application of spectroscopy to monitor phase equilibrium in complex chemical systems at high pressures is an important subset of analytical methods that minimizes perturbations during sample

¹Departamento de Física, Universidade Estadual de Maringá, 87020-900 Maringá, PR, Brazil. ²School of Engineering, Sao Paulo State University (UNESP), Campus of Sao Joao da Boa Vista, Sao Joao da Boa Vista, SP 13876-750, Brazil. ³Programa de Engenharia Química / COPPE, Universidade Federal do Rio de Janeiro, Cidade Universitária CP: 68502, Rio de Janeiro, RJ 21941-972, Brazil. ⁴Graduate Program in Chemical Engineering, State University of Maringá, Av. Colombo 5790, Maringá, Brazil. ⁵Center for Studies on Colloidal Systems (NUESC), Institute of Technology and Research (ITP), Av. Murilo Dantas, 300, Farolândia, Aracaju, SE, Brazil. ⁶Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia. ⁷Community Healthcare Center Dr. Adolf Drolc Maribor, Ulica talcev 9, 2000 Maribor, Slovenia. ⁸Department of Physics, Kyung Hee University, 26 Kyungheeda-ro, Dongdaemun-gu, 02447 Seoul, Republic of Korea. ⁹Complexity Science Hub, Metternichgasse 8, 1030 Vienna, Austria. ¹⁰University College, Korea University, 145 Anam-ro, Seongbuk-gu, 02841 Seoul, Republic of Korea. ✉email: aabpessa@uem.br; matjaz.perc@gmail.com; hvr@dfi.uem.br

withdrawal^{13–15}. Furthermore, combining spectroscopic techniques with robust, computationally efficient information-theoretic methods – such as permutation entropy¹⁶ and related ordinal approaches^{17–21} – has emerged as a promising strategy for investigating phase behavior under high-pressure conditions, as evidenced by its successful application in diverse spectroscopic and phase behavior studies^{22–30}. For instance, permutation entropy has been employed to segment near-infrared spectra from blood samples, thereby identifying wavelength ranges that improve blood glucose prediction^{22,23}. It has also served as an anomaly detection tool in time series of relative isotope abundances derived from laser absorption spectroscopy of ice cores²⁴. Ordinal methods have been extensively applied to characterize distinct flow regimes – such as slug flow, non-uniform bubble flow, and uniform bubble flow – in two-phase and three-phase systems^{25–28}. Moreover, permutation entropy and ordinal methods have yielded intriguing insights into phase transitions in thermotropic liquid crystals^{29,30}.

Here, we present an application of the recently proposed k -nearest neighbor permutation entropy³¹ to experimental transmittance spectra obtained *in situ* from mixtures of carbon dioxide and a distilled petroleum fraction under controlled high-pressure conditions. Specifically, we develop a methodology to detect phase transitions via abrupt changes in the entropy of the measured spectra. Our approach operates online, detecting phase transitions in real time as new sample spectra are acquired. Furthermore, because the k -nn permutation entropy – as well as other ordinal methods^{16,21} – relies on the relative amplitudes of neighboring data points, our method requires minimum data preprocessing, no specialized detection techniques, and no visual inspection of the spectra, and it can be readily extended to other systems monitored via spectrophotometric measurements.

The remainder of this paper is structured as follows. First, we present the dataset of absorbance spectra collected during depressurization trials of mixtures of carbon dioxide and hydrocarbon fractions. Next, we detail the formalisms underlying both the k -nearest neighbor permutation entropy and permutation entropy before their applications to the spectra. We then describe our anomaly detection approach for inferring phase transitions and present the results of its application. Finally, we discuss our findings and conclude.

Data

The dataset used in our investigations was originally introduced by Borges et al.⁶ and comprises near-infrared (NIR) absorbance spectra acquired via transmittance measurements during depressurization trials of mixtures of carbon dioxide (CO₂) and distilled petroleum fractions maintained within a high-pressure and variable-volume view cell. The distilled fraction is light yellow in color, with a density of 0.8230 g cm^{−3} at 298 K and a normal boiling point of 569 K; the raw oil composition, determined through standard fractionation analysis, includes 74.7% saturates, 14.7% aromatics, 10.6% resins, and less than 0.5% asphaltenes by weight⁶. The spectra, denoted by $\{x_\lambda\}_{\lambda=\lambda_{\min}, \dots, \lambda_{\max}}$, represent the sample absorbance (x_λ) indexed by 1440 distinct wavelengths (λ) unevenly sampled between $\lambda_{\min} = 1000$ nm and $\lambda_{\max} = 2250$ nm (Fig. 1). Mixtures were prepared using six different CO₂ concentrations, corresponding to relative sample weights ranging from 35% to 90%. Subsequently, several depressurization runs were performed at fixed temperatures ($T = 293$ K, $T = 303$ K, $T = 313$ K, and $T = 333$ K), while the samples were subjected to pressures ranging from 4.1 MPa to 19.5 MPa (see Table 1 for an overview of pressure intervals and the number of runs performed at different CO₂ concentrations and temperatures). Each trial yielded between 21 and 127 spectra, with measurements recorded at successive pressure increments of 0.1 MPa. For each combination of CO₂ concentration and temperature, the transition pressure P_t^{exp} corresponding to the observed phase transitions during depressurization was visually determined by an expert. A more detailed description of the experimental apparatus and techniques used in data collection is provided in Ref.⁶.

The observed phase transitions encompass transitions from homogeneous liquid or gas mixtures (at higher pressures) to different types of two-phase equilibria (at lower pressures), as well as a transition from a two-phase liquid mixture to a multiphase system (at even lower pressures). We categorize these transitions as follows:

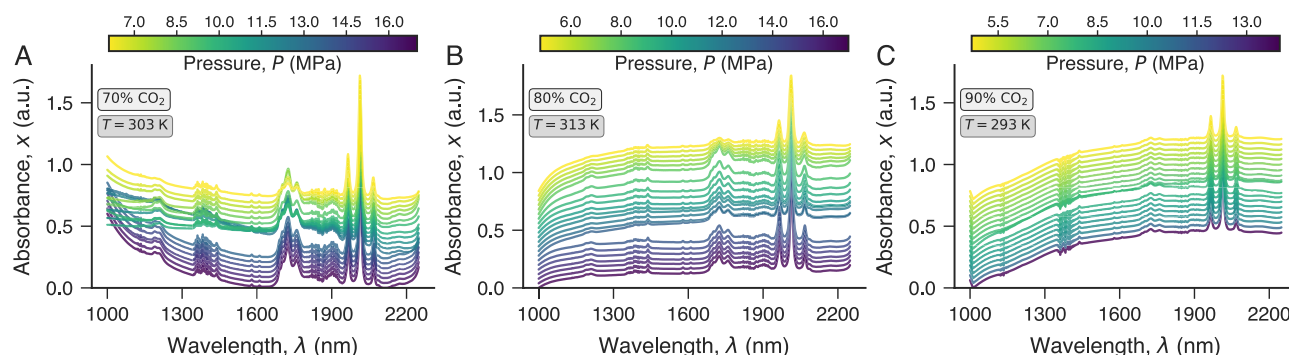


Fig. 1. Near-infrared spectra of mixtures of carbon dioxide and distilled petroleum fractions. Panels (A), (B), and (C) show absorbance spectra (in arbitrary units) for wavelengths between 1000 nm and 2250 nm measured at various pressures for three distinct samples. The spectra are color-coded by pressure (see colorbars) and annotated with the relative CO₂ sample weight and temperature. Absorbance spectra for each depressurization trial were recorded at 0.1 MPa increments and shifted in these visualizations for clarity.

CO ₂	Temperature	Runs	Pressure range
35%	293 K	4	5.3–10.0 MPa
35%	303 K	3	5.4–12.0 MPa
35%	313 K	2	8.0–13.5 MPa
35%	333 K	2	11.0–16.0 MPa
45%	293 K	3	4.9–13.0 MPa
45%	313 K	2	7.2–16.0 MPa
45%	333 K	2	12.6–17.5 MPa
60%	303 K	2	6.1–17.0 MPa
60%	313 K	2	12.8–16.0 MPa
60%	333 K	2	16.0–18.5 MPa
70%	303 K	2	5.8–17.5 MPa
70%	313 K	2	13.5–17.5 MPa
70%	333 K	2	16.4–19.5 MPa
80%	313 K	2	4.8–17.0 MPa
80%	333 K	2	15.5–18.5 MPa
90%	293 K	2	4.1–14.0 MPa
90%	303 K	2	6.2–12.0 MPa
90%	313 K	2	10.5–13.0 MPa
90%	333 K	2	15.0–18.0 MPa

Table 1. Summary of the carbon dioxide and distilled petroleum fraction mixtures used in depressurization trials. Six samples were prepared at six different carbon dioxide (CO₂) concentrations relative to sample weight, with depressurizations conducted at four different temperatures. The table shows the number of depressurization runs for each sample as well as the corresponding pressure intervals.

- liquid-liquid equilibrium (LLE): transitions from a homogeneous liquid mixture to a two-phase liquid mixture;
- vapor-liquid equilibrium via bubble point (VLE-BP): transitions from a homogeneous gas mixture to a two-phase gas-liquid mixture upon crossing the bubble-point curve;
- vapor-liquid equilibrium via dew point (VLE-DP): transitions from a homogeneous liquid mixture to a two-phase gas-liquid mixture upon crossing the dew-point curve;
- vapor-liquid-liquid equilibrium (VLLE): transitions from a two-phase liquid mixture to a multiphase system with three coexisting phases at equilibrium (one gas phase and two immiscible liquids).

Supplementary Table S1 summarizes the key variables extracted from this dataset.

Methods

To investigate whether phase transitions can be detected through changes in the characteristics of sample spectra during depressurization runs, we use the k -nearest neighbor (k -nn) permutation entropy³¹. This information-theoretic measure arises from combining the permutation entropy framework^{16–21} – originally developed for time series analysis – with biased random walks such as those implemented in node2vec³², an algorithm for obtaining high-dimensional vector representations of vertices in complex networks. The k -nn permutation entropy has been primarily proposed to address unstructured, irregularly sampled datasets such as our spectra, and its application to structured data (time series and images) has also yielded excellent results³¹.

The initial step in calculating the k -nn entropy involves fixing the number of neighbors k to construct a nearest neighbor graph on which random walks are performed (Fig. 2A and B). In our dataset, each absorbance value x_λ of a spectrum $\{x_\lambda\}_{\lambda=\lambda_{\min}, \dots, \lambda_{\max}}$ is mapped to a node in the graph. Both the absorbance and its corresponding wavelength (λ, x_λ) are considered when defining the nearest neighbors via the Euclidean distances between all datapoints (λ, x_λ) in the spectrum. However, the range of variation and the typical scale of differences between wavelengths and absorbances are not equivalent. In fact, the locally smooth behavior of the spectrum renders neighboring data points considerably closer in terms of absorbance than in terms of wavelength, thereby disproportionately weighting the absorbance values during graph construction. To mitigate this imbalance, we rescale the wavelength and absorbance values to the interval $[0, 1]$ via min-max normalization prior to constructing the nearest-neighbor graph. Subsequently, n random walks of length w are initiated from each x_λ (Fig. 2B and C). These random walks are second-order Markov processes, wherein the probability ρ_{bc} of transitioning from node b to node c depends on the previous node a visited before b according to the unnormalized transition probabilities

$$\rho_{bc} = \begin{cases} 1/\alpha, & \text{if } s_{ac} = 0 \\ 1, & \text{if } s_{ac} = 1 \\ 1/\beta, & \text{if } s_{ac} = 2 \end{cases}, \quad (1)$$

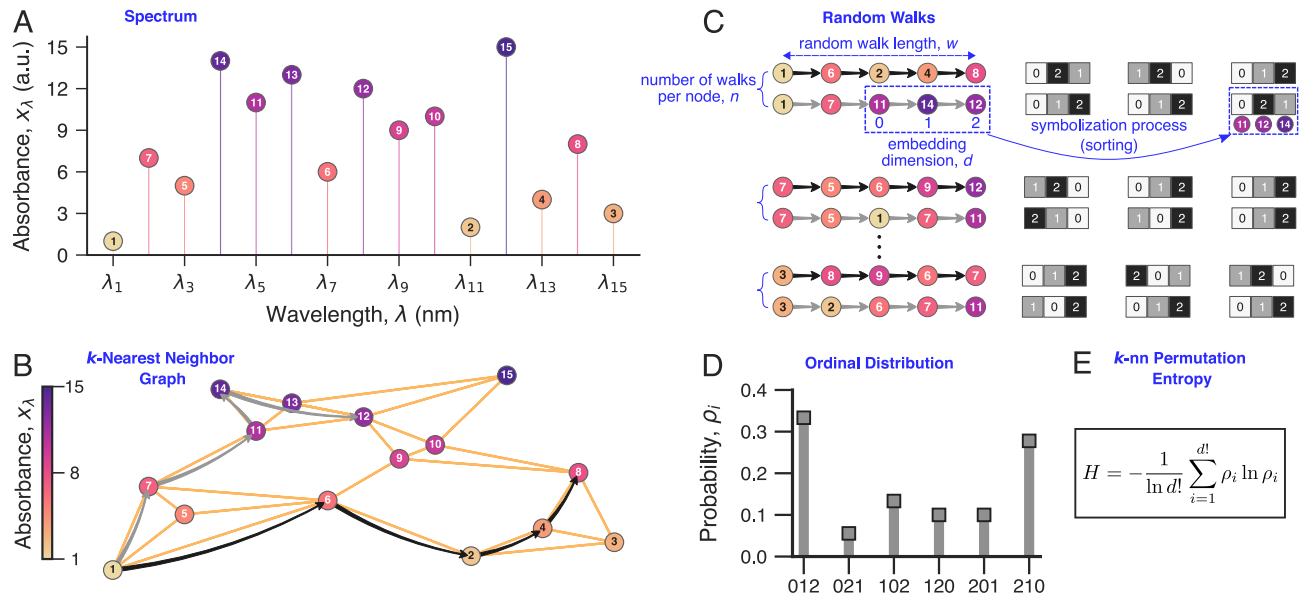


Fig. 2. Calculation of the k -nearest neighbor permutation entropy for sequential data. **(A)** Illustration of a short, unevenly sampled, hypothetical spectrum $\{x_\lambda\}_{\lambda=\lambda_1, \dots, \lambda_{15}}$. **(B)** From this hypothetical spectrum, we construct a k -nearest neighbor graph ($k = 3$ in this example) using the data coordinates (λ, x_λ) to define neighborhood relationships. In this graph, each observed absorbance value x_λ is represented by a node, with undirected edges connecting pairs of observations $x_{\lambda_i} \leftrightarrow x_{\lambda_j}$ when $(\lambda_j, x_{\lambda_j})$ is among the k -nearest neighbors of $(\lambda_i, x_{\lambda_i})$. For ease of interpretation, the min-max normalization step has been omitted from this illustration. **(C)** Subsequently, we execute n biased random walks of length w starting from each node, sampling the absorbance values (x_λ) to generate time series ($n = 2$ and $w = 5$ in this example). We then apply the Bandt-Pompe symbolization approach to each of these time series. This symbolization entails creating overlapping partitions of length d ($d = 3$ in this example) and arranging partition indices in ascending order to determine the sorting permutation for each partition. **(D)** Finally, we evaluate the probability of occurrence of each of the $d!$ permutation types (ordinal distribution) and calculate its **(E)** Shannon entropy, thereby defining the k -nearest neighbor permutation entropy.

where s_{ac} denotes the shortest path (minimum number of edges) connecting nodes a (previously occupied node) and c (subsequent node), while $\alpha > 0$ and $\beta > 0$ are parameters controlling the walker's bias. Smaller values of α favor more exploitative walks that sample local neighborhoods, while smaller values of β favor more exploratory ones that sample deeper neighborhoods.

After the walks are performed, we obtain $N \times n$ time series of length w , which we denote by $\{\tilde{x}_m\}_{m=1, \dots, N \times n}$. We then apply a symbolization process derived from the permutation entropy framework¹⁶ to each time series as follows. First, we select a sliding window of size d (the embedding dimension) that traverses each $\{\tilde{x}_m\}$ (Fig. 2C), dividing each trajectory into overlapping partitions p_m represented by

$$p_m = (\tilde{x}_m, \tilde{x}_{m+1}, \dots, \tilde{x}_{m+d-1}), \quad (2)$$

with $m = 1, \dots, w - d + 1$. For each partition p_m , we determine the permutation $\pi_m = (r_0, r_1, \dots, r_{d-1})$ of the index numbers $(0, 1, \dots, d-1)$ that sorts the elements of p_m , such that $\tilde{x}_{m+r_0} \leq \tilde{x}_{m+r_1} \leq \dots \leq \tilde{x}_{m+r_{d-1}}$. In the event of equal values within a partition, that is, $\tilde{x}_{m+r_{k-1}} = \tilde{x}_{m+r_k}$, we impose $r_{k-1} < r_k$ for $k = 1, \dots, d-1$ ³³, thereby effectively treating earlier observations as smaller than later ones. As an illustration of the symbolization, consider a short walk $\{\tilde{x}_m\} = \{6, 1, 6, 4, 8\}$ with sliding window size $d = 3$. The first partition is $p_1 = (6, 1, 6)$ and sorting its elements yields $1 < 6 \leq 6$, that is, $x_{1+1} < x_{1+0} \leq x_{1+2}$. Thus, the permutation associated with p_1 is $\pi_1 = (1, 0, 2)$. It is important to note that the number of different permutation types each π_m can assume is equal to $d!$. Accordingly, the complete set of all permutation types is represented by $\{\Pi_j\}_{j=1, \dots, d!}$. For instance, if $d = 3$, there are $3! = 6$ permutations types: $\Pi_1 = (0, 1, 2)$, $\Pi_2 = (0, 2, 1)$, $\Pi_3 = (1, 0, 2)$, $\Pi_4 = (1, 2, 0)$, $\Pi_5 = (2, 0, 1)$, and $\Pi_6 = (2, 1, 0)$.

By repeatedly applying the previously described symbolization method to all sampled walks and aggregating the corresponding permutations, we obtain a single symbolic sequence $\{\pi_m\}_{m=1, \dots, M}$ comprising $M = N \times n \times (w - d + 1)$ permutation symbols. The probability $\rho_i(\Pi_i)$ for each permutation type is then estimated as

$$\rho_i(\Pi_i) = \frac{\text{number of permutations of type } \Pi_i \text{ in } \{\pi_m\}_{m=1, \dots, M}}{N \times n \times (w - d + 1)}. \quad (3)$$

Finally, given the probability distribution $\mathcal{P} = \{\rho_i(\Pi_i)\}_{i=1,\dots,d!}$, the nearest neighbor permutation entropy³¹ is computed via the normalized Shannon entropy (Fig. 2D and E):

$$H[\mathcal{P}] = -\frac{1}{\ln d!} \sum_{i=1}^{d!} \rho_i(\Pi_i) \ln \rho_i(\Pi_i). \quad (4)$$

According to this equation, the k -nearest neighbor permutation entropy is restricted to $H \in [0, 1]$, with H quantifying the degree of irregularity in the spectrum $\{x_\lambda\}_{\lambda=\lambda_{\min},\dots,\lambda_{\max}}$. In general, we expect $H \approx 1$ when a single permutation dominates the probability distribution, whereas $H \approx 0$ indicates a uniform distribution across all permutation types. The procedure for computing the k -nn permutation entropy is illustrated in Fig. 2, and for the numerical implementation, we rely on the Python module `knnpe`³⁴.

In applying the k -nearest neighbor permutation entropy to our absorbance spectra, we adopt the default values for the four parameters associated with the random walks³¹: $n = 10$, $w = 10$, $\alpha = 10$, and $\beta = 0.001$. In contrast, the parameters k and d are extensively varied to identify optimal combinations. Additionally, to provide a basis for comparison, we also calculate the conventional permutation entropy of the spectra¹⁶. Briefly, the permutation entropy is calculated by symbolizing a spectrum $\{x_\lambda\}_{\lambda=\lambda_{\min},\dots,\lambda_{\max}}$ using a sliding window of size d to partition the complete sequence of absorbance values. Next, we identify the permutations that sort these partitions (forming the permutation sequence $\{\pi_t\}$), determine the probabilities associated with each permutation type $[\rho_i(\Pi_i), i = 1, \dots, d!]$, and compute the corresponding Shannon entropy, implemented using the Python module `ordpy`³⁵. Supplementary Table S2 summarizes the parameters and key variables related to the k -nearest neighbor permutation entropy.

Results

Leveraging the absorbance spectra collected during our depressurization trials, we investigate phase transitions in mixtures of carbon dioxide and a distilled petroleum fraction via an anomaly detection approach. Specifically, we calculate the k -nearest neighbor entropy for each spectrum recorded at a given pressure and monitor the dispersion of these entropy values as an indicator of phase transitions.

Formally, we interpret a sequence of entropy values $\{H_P\}_{P=P_{\max},\dots,P_{\min}}$ obtained from a depressurization trial as a time series indexed by decreasing pressure values, from P_{\max} to P_{\min} . Thus, the trend and fluctuations in H_P are quantified iteratively by calculating the average entropy

$$\mu_P = \sum_{P'=P_{\max}}^P \frac{H_{P'}}{N} \quad (5)$$

and its standard deviation

$$\sigma_P = \sqrt{\sum_{P'=P_{\max}}^P \frac{(H_{P'} - \mu_{P'})^2}{N}} \quad (6)$$

evaluated over an expanding window indexed by pressures $P \in \{P_{\max}, \dots, P_{\min}\}$ as the depressurization proceeds. We assume that the transition pressure P_t^{knn} associated with a phase transition is signaled by an abrupt change in H_P . Specifically, P_t^{knn} is identified as the pressure at which H_P first falls outside the confidence band defined by the thresholds $H_P^- = \mu_P - \gamma\sigma_P$ and $H_P^+ = \mu_P + \gamma\sigma_P$, where γ controls the band width. Supplementary Table S3 summarizes the parameters and key variables used in our anomaly detection framework.

Figure 3 illustrates this procedure for three different samples. For instance, Fig. 3A presents a single depressurization trial involving a mixture containing 45% CO₂ at $T = 313$ K. The initially homogeneous liquid mixture transitions to a vapor-liquid equilibrium upon crossing the bubble-point curve (VLE-BP) at $P_t^{\text{exp}} = 10.2$ MPa. Notably, the spectra entropy values first exceed the upper threshold H_P^+ at $P_t^{knn} = 10.1$ MPa, signaling the onset of the VLE-BP phase in close agreement with the visually inferred transition pressure. Figure 3B shows a transition from an initially homogeneous gas mixture to vapor-liquid equilibrium by crossing the dew-point curve (VLE-DP), while Fig. 3B depicts a transition from an initially homogeneous liquid mixture to a two-phase liquid-liquid equilibrium (LLE). For the VLE-DP transition in Fig. 3B, our method identifies $P_t^{knn} = 11.7$ MPa, marginally smaller than the visually determined $P_t^{\text{exp}} = 11.8$ MPa, and for the LLE transition in Figure 3C it predicts $P_t^{knn} = 8.1$ MPa, slightly lower than the visually determined $P_t^{\text{exp}} = 8.4$ MPa.

We next apply our approach to characterize phase transitions across all 42 depressurization trials. As the approach depends on the threshold parameter γ and the entropy parameters k and d , we systematically explore all possible combinations of $\gamma \in \{1.00, 1.05, \dots, 3.50\}$, $k \in \{10, 15, \dots, 300\}$, and $d \in \{3, 4, 5\}$ to optimize the predictive performance and maximize the number of accurately detected transitions. To identify the best-performing parameter set, we adopt a leave-one-out cross-validation strategy³⁶ in which each depressurization trial is iteratively used as a validation set, while the remaining trials constitute the training set. This procedure ensures that the parameter optimization process remains strictly independent of the evaluation data, thereby enabling out-of-sample predictions that reflect our method's generalizability and predictive power. Through this

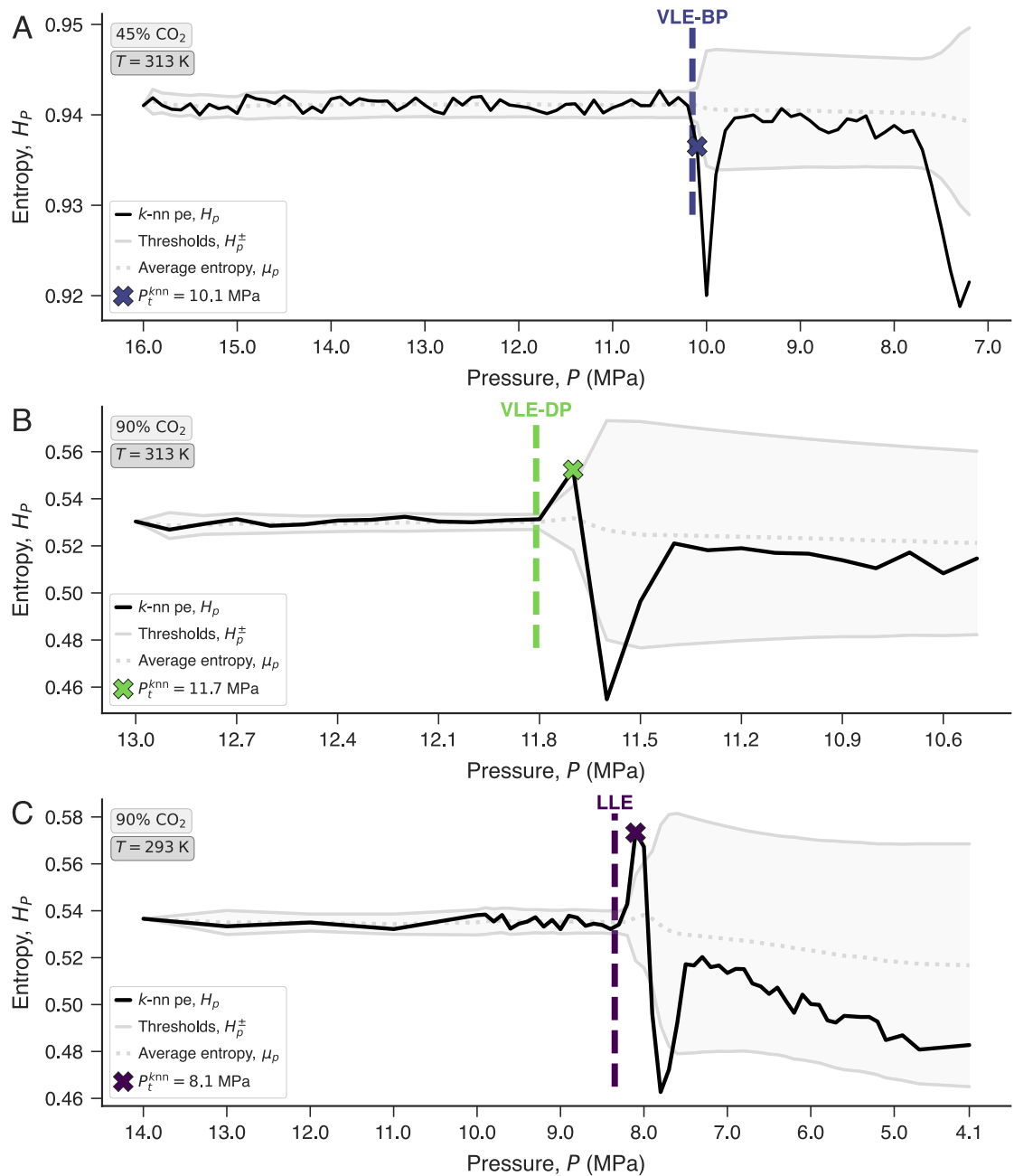


Fig. 3. Phase transitions manifest as anomalies in entropy values. Panels (A), (B), and (C) show the values of k -nearest neighbor permutation entropy (H_P , solid black line) calculated from spectra of carbon dioxide and distilled petroleum fraction mixtures collected during depressurization trials. Light gray lines denote the confidence bands defined by threshold values H_P^+ and H_P^- , whose exceedance indicates a phase transition. Gray dotted lines display the average entropy values (μ_P) calculated within expanding windows. Transition pressures visually determined by experts are indicated by colored vertical lines alongside the type of initiating phase, while colored cross markers show transition pressures identified by entropy values outside the confidence bands. CO_2 concentration levels and temperatures are annotated within the panels. In all panels, the entropy parameters are set to $k = 265$ and $d = 5$, with the threshold parameter $\gamma = 2.25$.

extensive search, we find that the best results are consistently obtained for $\gamma = 2.25$, $k = 265$, and $d = 5$ for 41 out of the 42 validation trials. The only exception is a trial prepared with $\text{CO}_2 = 45\%$ and $T = 333$ K, where no phase transition was detected and a slightly different parameter set emerged.

Figure 4A compares the predicted transition pressures (P_t^{knn}) with the visually determined reference values (P_t^{exp}) for the 41 trials in which a phase transition was detected (see also Supplementary Table S4). The predictions closely match the experimental values, with a coefficient of determination $R^2 = 0.96$ and a mean absolute percentage error (MAPE) of 4.51%. Figure 4B further shows that this high predictive performance

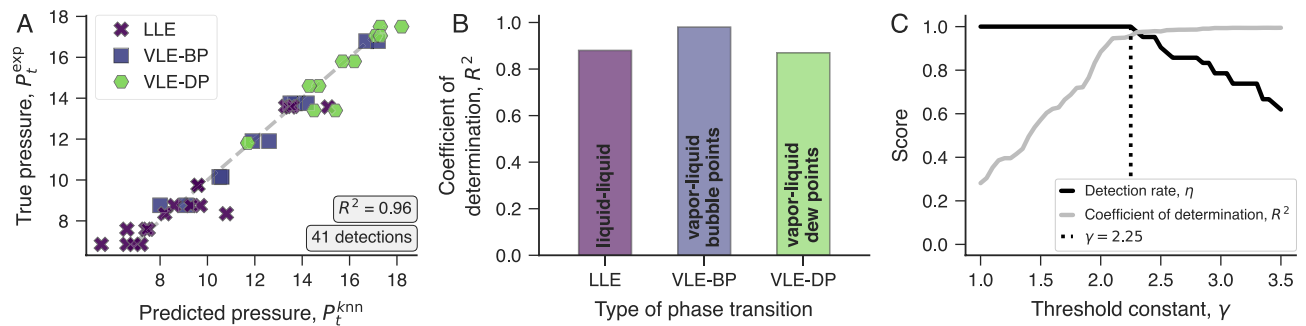


Fig. 4. Accuracy of out-of-sample predictions for transition pressures in mixtures of carbon dioxide with a distilled petroleum fraction. **(A)** Relationship between the true P_t^{exp} (experimentally determined) and predicted P_t^{knn} transition pressures for 41 of 42 depressurization trials in which a transition was successfully identified. Colors indicate the transition types observed during the depressurization of homogeneous mixtures. The coefficient of determination (R^2) quantifying the agreement between true and predicted pressures is shown within the panel. **(B)** Coefficient of determination (R^2) stratified by the transition types. Predictions were obtained using a leave-one-out cross-validation approach, with the threshold parameter γ and entropy parameters k and d optimized independently of the evaluation data. **(C)** Fraction of detected transitions η (black line) and coefficient of determination R^2 (gray line) as functions of the threshold parameter γ , with entropy parameters fixed at their optimal values ($k = 265$ and $d = 5$).

holds across different types of initiating phase equilibria. For example, transitions to LLE yield $R^2 = 0.88$ and MAPE = 6.25%, transitions to VLE-BP lead to $R^2 = 0.98$ and MAPE = 3.16%, and transitions to VLE-DP produce $R^2 = 0.87$ and MAPE = 3.15%. In 22 out of 41 detected transitions, the predicted transition pressures are equal to or smaller than the true values, indicating an approximately symmetric chance of anticipating or lagging the experimentally determined transitions. On average, the absolute difference between the predicted pressures and the pressures sampled nearest to the experimental estimates is 0.5 MPa. These discrepancies may partly arise from the finite temporal resolution of the experimental measurements and the windowed nature of the k -nn permutation entropy calculation, both of which can hinder the detection of abrupt changes; gradual variations in system dynamics near the transition may also contribute. Additionally, given the inherent difficulty in exhaustively monitoring these rapid transitions, visually determined transition pressures are themselves susceptible to experimental errors.

Furthermore, given the stability of the optimal entropy parameters across validation trials, we investigate the effect of the threshold parameter γ on predictive performance while fixing $k = 265$ and $d = 5$ – their optimal values. As shown in Figure 4C, values of γ in the range $1 \leq \gamma \leq 2.25$ yield the highest detection rate η , albeit at the cost of reduced accuracy in the predicted transition pressures. Conversely, for $2.00 \leq \gamma \leq 3.50$, the predictions of transition pressures improve, but the fraction of detected transitions decreases with increasing γ . Since γ controls the width of the confidence band, these findings indicate that transition pressures are more accurately identified when entropy changes are more abrupt. Thus, enhancing the temporal resolution of the experimental measurements would likely improve both detection accuracy and performance, as the confidence band would become narrower and smoother.

For further comparison, we apply our anomaly detection approach using the standard permutation entropy in place of the k -nearest neighbor entropy. As before, we employ a leave-one-out cross-validation strategy to identify the optimal parameter combination, considering $\gamma \in \{1.00, 1.05, \dots, 3.50\}$ and $d \in \{3, 4, 5\}$. This modification, however, results in markedly poorer predictive performance, with $R^2 = 0.67$ and MAPE = 15.0%, despite correctly detecting all 42 transitions (Supplementary Figure S1A). Stratifying results by transition type further reveals that standard permutation entropy performs particularly poorly for LLE and VLE-BP transitions (Supplementary Figure S1B). These findings indicate that the superior performance of our approach stems from the k -nn permutation entropy's ability to exploit local neighborhood relationships via the graph-based sampling strategy that captures amplitude variations and wavelength gaps ignored by the standard permutation entropy, thereby enhancing noise resilience and increasing sensitivity in detecting phase transitions.

Continuing our analysis, we observe that nine depressurization trials exhibit a second transition, each corresponding to a change from liquid-liquid equilibrium (at higher pressures) to vapor-liquid-liquid equilibrium (at lower pressures). Figure 5A illustrates the evolution of the k -nearest neighbor permutation entropy in one such trial, along with the confidence band and the detected transition obtained using our approach with previously optimized parameters ($\gamma = 2.25$, $k = 265$ and $d = 5$). However, unlike this particular example, and despite detecting all transitions, our method generally performs poorly in identifying this class of transitions, yielding a negative coefficient of determination ($R^2 = -2.42$) and a substantial prediction error (mean absolute percentage error, MAPE = 16.1%). In an attempt to address this limitation, we re-optimize the parameter set using a leave-one-out cross-validation strategy tailored to this class of transitions. Yet, as shown in Fig. 5B (see also Supplementary Table S5), even re-optimized parameters fail to significantly improve performance, with the coefficient of determination remaining negative ($R^2 = -1.68$) and a high prediction error persisting (MAPE = 13.0%), based on 7 detected transitions out of the 9 observed. Furthermore, replacing k -nearest

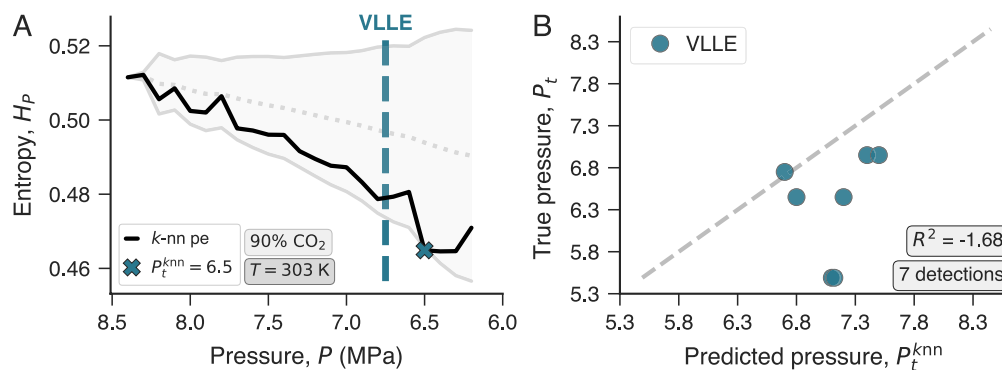


Fig. 5. Challenges in predicting transition pressures between liquid–liquid and vapor–liquid–liquid equilibria. **(A)** Values of the k -nearest neighbor permutation entropy (H_P , solid black line) calculated from spectra of carbon dioxide and distilled petroleum fraction mixtures during depressurization trials involving a transition from a liquid–liquid equilibrium to vapor–liquid–liquid equilibrium. Light gray lines represent threshold values whose exceedance indicates a phase transition (H_P^- and H_P^+), while gray dotted lines show average entropy values (μ_P). The experimentally identified transition pressure (P_t^{exp}) is marked by a vertical colored line, and the predicted transition pressure (P_t^{knn}) is shown by a colored cross. Mixture composition and temperature are indicated within the panel. **(B)** Relationship between the true P_t^{exp} (experimentally determined) and predicted P_t^{knn} transition pressures for 7 of 9 depressurization trials exhibiting a second phase transition. The coefficient of determination (R^2) quantifying the agreement between true and predicted pressures is shown within the panel.

neighbor entropy with standard permutation entropy results in even less accurate predictions. These results indicate that transitions between liquid–liquid equilibrium (LLE) and vapor–liquid–liquid equilibrium (VLLE) are considerably more difficult to detect via abrupt changes in entropy than transitions between homogeneous mixtures and other types of high-pressure phase equilibria. This increased difficulty likely stems from the greater complexity of the LLE–VLLE transition – characterized by vapor nucleation and mass transfer between phases – which renders phase homogeneity more difficult to achieve experimentally⁶. Additionally, due to the spatial arrangement of phases within the optical cell, the NIR beam may predominantly sample only one of the coexisting phases – typically the denser or more spectrally active liquid – while underrepresenting the emerging vapor phase or the less optically absorbing liquid. The relatively weak absorbance of the vapor phase, combined with its limited residence time or distribution along the optical path, may further diminish its contribution to the spectra. As the k -nn entropy reflects spectral complexity rather than intensity alone, the absence of abrupt or structured spectral changes across scans makes such transitions inherently more difficult to resolve. Finally, because these transitions occur later in the depressurization process, the number of spectra preceding the transition is limited, reducing the temporal resolution and the robustness of confidence band estimation – further constraining the effectiveness of our detection approach.

Discussion and conclusions

We have presented an investigation of phase transitions in synthesized mixtures of carbon dioxide with a distilled petroleum fraction using an anomaly detection methodology. During depressurization trials, our approach repeatedly calculates the k -nearest neighbor permutation entropy – an information-theoretic measure directly computed from near-infrared absorbance spectra – and signals the occurrence of a phase transition by identifying large variations in entropy values. In line with previous studies on high-pressure systems^{13,15}, our work integrates synthetic and analytical methods to elucidate phase behavior in high-pressure conditions, employing a modified experimental setup that enables continuous system monitoring without requiring sample extraction⁶.

Our methodology yields robust results, accurately identifying transitions from initially homogeneous mixtures to different types of two-phase equilibria. Moreover, the approach can be tailored to specific mixtures of carbon dioxide and hydrocarbon fractions (for instance, 60% CO_2) or particular types of phase transitions, allowing for further improvements in performance. Indeed, detecting phase transitions in high-pressure hydrocarbon mixtures with elevated CO_2 content has proven challenging, typically requiring cumbersome data preprocessing, specialized detection techniques based on field expertise, and qualitative visual assessments⁶, difficulties that our method overcomes through an online, streamlined, and computationally efficient strategy.

Although we have used the same dataset as Borges et al.⁶, the results are not directly comparable. In their study, transition pressures were estimated via visual identification of inflection points in cumulative absorbance curves, whereas our fully automated approach employs a leave-one-out cross-validation scheme to enable out-of-sample predictions. Moreover, their analysis reflects average behaviors across isothermal depressurization trials at fixed CO_2 concentrations, while our method produces individualized predictions for each trial. We have further verified that the predictive power of our approach arises from the ability of k -nearest neighbor permutation

entropy to exploit local neighborhood structure through a graph-based sampling strategy that captures both amplitude variations and wavelength gaps – features not adequately resolved by standard permutation entropy, which yielded markedly inferior results. While future studies may explore the performance of other generalized variants of permutation entropy, we emphasize that the capacity to simultaneously account for gaps, amplitude values, and non-sequential sampling remains a distinctive feature of the k -nearest neighbor permutation entropy³¹.

Transitions between liquid-liquid and vapor-liquid-liquid equilibria, however, were not effectively captured by changes in the k -nearest neighbor permutation entropy. A closer examination reveals that, in these cases, entropy values tend to shift gradually, limiting our method's ability to detect abrupt changes. This observation represents an intriguing avenue for future studies aimed at defining or applying more sensitive information-theoretic quantifiers capable of detecting gradual shifts. Beyond the intrinsic complexity of the liquid-liquid to vapor-liquid-liquid equilibria transition, part of the challenge may stem from the fact that the mixtures were not allowed to stabilize after reaching liquid-liquid equilibrium; consequently, the experimental conditions preceding this phase transition differ from those preceding the transition of the initially homogeneous mixture into liquid-liquid equilibrium. These transitions also occur later in the depressurization process, resulting in fewer spectra preceding their onset and limiting the reliability of confidence band estimation – factors that further hinder the application of our approach in these cases. We believe that experiments specifically designed to probe such transitions, incorporating slower depressurization rates and using higher sampling rates for collecting the spectra, may yield more conclusive results. We also note that our anomaly-detection framework is somewhat computationally intensive due to the exhaustive search over a three-dimensional parameter space, which may become cumbersome for large spectral datasets.

Despite these limitations, our approach is sufficiently general to be adapted for studying phase behavior in other chemical systems subjected to high-pressure conditions and monitored by spectrophotometry. Thus, our work underscores the noninvasive nature of employing permutation entropy and ordinal methods in the investigation of complex physicochemical systems^{29,30}.

Data availability

All code and data necessary to reproduce the results and figures in this work are available at the GitLab repository <https://gitlab.com/complexlab/nir-co2oil-knnpe>.

Received: 16 February 2025; Accepted: 5 June 2025

Published online: 01 July 2025

References

- Almobarak, M. et al. A review of chemical-assisted minimum miscibility pressure reduction in CO₂ injection for enhanced oil recovery. *Petroleum*. **7**, 245–253. <https://doi.org/10.1016/j.petlm.2021.01.001> (2021).
- Li, B. et al. Molecular dynamics simulation of CO₂ dissolution in heavy oil resin-asphaltene. *Journal of CO₂ Utilization*. **33**, 303–310. <https://doi.org/10.1016/j.jcou.2019.06.011> (2019).
- Rodrigues, H. W., Mackay, E. J. & Arnold, D. P. Multi-objective optimization of CO₂ recycling operations for CCUS in pre-salt carbonate reservoirs. *International Journal of Greenhouse Gas Control*. **119**, 103719. <https://doi.org/10.1016/j.ijggc.2022.103719> (2022).
- Zhao, X. et al. A multi-medium and multi-mechanism model for CO₂ injection and storage in fractured shale gas reservoirs. *Fuel*. **345**, 128167. <https://doi.org/10.1016/j.fuel.2023.128167> (2023).
- Romero Yanes, J. F. et al. Study of liquid-liquid and liquid-liquid-vapor equilibria for crude oil mixtures with carbon dioxide and methane using short-wave infrared imaging. Experimental and thermodynamic modeling. *Energy & Fuels*. **34**, 14109–14123. <https://doi.org/10.1021/acs.energyfuels.0c03064> (2020).
- Borges, G. R. et al. Near infrared spectroscopy applied for high-pressure phase behavior measurements. *The Journal of Supercritical Fluids*. **104**, 221–226. <https://doi.org/10.1016/j.supflu.2015.06.015> (2015).
- Cruz, A. A. et al. CO₂ influence on asphaltene precipitation. *The Journal of Supercritical Fluids*. **143**, 24–31. <https://doi.org/10.1016/j.supflu.2018.08.005> (2019).
- Alher, A. A., Aljawad, M. S. & Alyasery, A. A. Estimating the PVT properties for crude oil from a Southern Iraqi oil field. *IOP Conference Series: Materials Science and Engineering*. **433**, 012076. <https://doi.org/10.1088/1757-899X/433/1/012076> (2018).
- Regueira, T., Glykioti, M.-L., Kottaki, N., Stenby, E. H. & Yan, W. Density, compressibility and phase equilibrium of high pressure-high temperature reservoir fluids up to 473 K and 140 MPa. *The Journal of Supercritical Fluids*. **159**, 104781. <https://doi.org/10.1016/j.supflu.2020.104781> (2020).
- Lucas, M. A. et al. Use of real crude oil fractions to describe the high pressure phase behavior of crude oil in carbon dioxide. *The Journal of Supercritical Fluids*. **118**, 140–147. <https://doi.org/10.1016/j.supflu.2016.08.004> (2016).
- Daridon, J.-L., Lin, C.-W., Carrier, H., Pauly, J. & Fleming, F. P. Combined investigations of fluid phase equilibria and fluid–solid phase equilibria in complex CO₂–crude oil systems under high pressure. *Journal of Chemical & Engineering Data*. **65**, 3357–3372. <https://doi.org/10.1021/acs.jced.0c00144> (2020).
- Romero Yanes, J. F. et al. Phase behavior investigation of a live presalt crude oil from short-wave infrared observation, acoustic wave sensing, and equation of state modeling. *Energy & Fuels*. **35**, 18504–18517. <https://doi.org/10.1021/acs.energyfuels.1c02980> (2021).
- Dohrn, R., Fonseca, J. M. & Peper, S. Experimental methods for phase equilibria at high pressures. *Annual Review of Chemical and Biomolecular Engineering*. **3**, 343–367. <https://doi.org/10.1146/annurev-chembioeng-062011-081008> (2012).
- Braeuer, A. *In situ Spectroscopic Techniques at High Pressure*, vol. 7 of *Supercritical Fluid Science and Technology* (Elsevier, 2015).
- Dohrn, R., Peper, S., Secuianu, C. & Fonseca, J. M. High-pressure fluid-phase equilibria: Experimental methods, developments and systems investigated (2013–2016). *Fluid Phase Equilibria*. **579**, 113978. <https://doi.org/10.1016/j.fluid.2023.113978> (2024).
- Bandt, C. & Pompe, B. Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*. **88**, 174102. <https://doi.org/10.1103/PhysRevLett.88.174102> (2002).
- Zanin, M., Zunino, L., Rosso, O. A. & Papo, D. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy*. **14**, 1553–1577. <https://doi.org/10.3390/e14081553> (2012).
- Riedl, M., Müller, A. & Wessel, N. Practical considerations of permutation entropy. *The European Physical Journal Special Topics*. **222**, 249–262. <https://doi.org/10.1140/epjst/e2013-01862-7> (2013).

19. Amigó, J. M., Keller, K. & Unakafova, V. A. Ordinal symbolic analysis and its application to biomedical recordings. *Philosophical Transactions of the Royal Society A*. **373**, 20140091. <https://doi.org/10.1098/rsta.2014.0091> (2015).
20. Keller, K., Mangold, T., Stolz, I. & Werner, J. Permutation entropy: New ideas and challenges. *Entropy*. **19**, 134. <https://doi.org/10.3390/e19030134> (2017).
21. Pessa, A. A. B. & Ribeiro, H. V. Ordpy: A Python package for data analysis with permutation entropy and ordinal network methods. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. **31**, 063110. <https://doi.org/10.1063/5.0049901> (2021).
22. Li, X. & Li, C. Application of permutation entropy in feature extraction for near-infrared spectroscopy noninvasive blood glucose detection. *Journal of Spectroscopy*. **2017**, 9165247. <https://doi.org/10.1155/2017/9165247> (2017).
23. Li, X. & Li, C. Pretreatment and wavelength selection method for near-infrared spectra signal based on improved ceemdan energy entropy and permutation entropy. *Entropy*. **19**, 380. <https://doi.org/10.3390/e19070380> (2017).
24. Garland, J. et al. Anomaly detection in paleoclimate records using permutation entropy. *Entropy*. **20**, 931. <https://doi.org/10.3390/e20120931> (2018).
25. Ren, W., Zhang, J. & Jin, N. Rescaled range permutation entropy: a method for quantifying the dynamical complexity of gas-liquid two-phase slug flow. *Nonlinear Dynamics*. **104**, 4035–4043. <https://doi.org/10.1007/s11071-021-06468-2> (2021).
26. Zhang, J., Jin, N. & Ren, W. Detecting the fluid flow transitions in oil-gas-water three-phase flows using multivariate rescaled range permutation entropy analysis. *The European Physical Journal Plus*. **137**, 808. <https://doi.org/10.1140/epjp/s13360-022-03026-6> (2022).
27. Du, M., Wei, J., Li, M.-Y., Gao, Z.-K. & Kurths, J. Interconnected ordinal pattern complex network for characterizing the spatial coupling behavior of gas-liquid two-phase flow. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. **33**, 063108. <https://doi.org/10.1063/5.0146259> (2023).
28. Du, M., Zhang, L., Niu, X. & Grebogi, C. Detecting gas-liquid two-phase flow pattern determinism from experimental signals with missing ordinal patterns. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. **30**, 093102. <https://doi.org/10.1063/5.0016401> (2020).
29. Sigaki, H. Y. D., de Souza, R. F., de Souza, R. T., Zola, R. S. & Ribeiro, H. V. Estimating physical properties from liquid crystal textures via machine learning and complexity-entropy methods. *Physical Review E*. **99**, 013311. <https://doi.org/10.1103/PhysRevE.99.013311> (2019).
30. Pessa, A. A., Zola, R. S., Perc, M. & Ribeiro, H. V. Determining liquid crystal properties with ordinal networks and machine learning. *Chaos, Solitons & Fractals*. **154**, 111607. <https://doi.org/10.1016/j.chaos.2021.111607> (2022).
31. Voltarelli, L. G. J. M. et al. Characterizing unstructured data with the nearest neighbor permutation entropy. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. **34**, 053130. <https://doi.org/10.1063/5.0209206> (2024).
32. Grover, A. & Leskovec, J. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, 855–864. <https://doi.org/10.1145/2939672.2939754> (Association for Computing Machinery, New York, NY, USA, 2016).
33. Cao, Y., Wen, Tung, W., Gao, J. B., Protopopescu, V. A. & Hively, L. M. Detecting dynamical changes in time series using the permutation entropy. *Physical Review E*. **70**, 046217. <https://doi.org/10.1103/PhysRevE.70.046217> (2004).
34. Voltarelli, L. G. J. M. et al. knnpe: A Python package implementing the k-nearest neighbor permutation entropy. Available: <https://github.com/hvrbeiro/knnpe> (2024). Accessed: 16 Feb 2025.
35. Pessa, A. A. B. & Ribeiro, H. V. A Python package for data analysis with permutation entropy and ordinal network methods (ordpy). Available: <https://github.com/arthurpessa/ordpy> (2021). Accessed: 16 Feb 2025.
36. James, G., Witten, D., Hastie, T., Tibshirani, R. & Taylor, J. *An Introduction to Statistical Learning: with Applications in Python*. Springer Texts in Statistics (Springer International Publishing, 2023).

Acknowledgements

The authors acknowledge the support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – Grant 303533/2021-8), and Slovenian Research and Innovation Agency (Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije) (Grants P1-0403 and N1-0232).

Author contributions

A.A.B.P., L.G.J.M.V., L.C.F., A.G.M.T., C.D., P.M.N., M.P., and H.V.R. designed research, performed research, analyzed data, and wrote the paper.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-06016-7>.

Correspondence and requests for materials should be addressed to A.A.B.P., M.P. or H.V.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025