Green Chemistry

PAPER

Check for updates

SolECOs: a data-driven platform for sustainable and comprehensive solvent selection in pharmaceutical manufacturing

Yiming Ma, [ID] [a] Shang Gao, [ID] [a] Neel Mehta,[a] Qinqing Fu, [ID] [b] Wei Li[a] and Brahim Benyahia [ID] *[a]

Solvent selection in pharmaceutical crystallization plays a pivotal role in determining overall manufacturing efficiency while also significantly impacting environmental performance and regulatory compliance. A data-driven solution for sustainable solvent selection, applicable to both single and binary solvent systems, was developed and integrated into *SolECOs* (Solution ECOsystems), a modular and user-friendly platform for Sustainable-by-Design solvent selection in pharmaceutical manufacturing. A comprehensive solubility database containing 1186 active pharmaceutical ingredients (APIs) and 30 solvents was constructed and used in conjunction with thermodynamically informed machine learning models, including the Polynomial Regression Model-based Multi-Task Learning Network (PRMMT), the Point-Adjusted Prediction Network (PAPN), and the Modified Jouyban–Acree-based Neural Network (MJANN), to predict solubility profiles along with associated uncertainties. Sustainability assessment was performed using both midpoint and endpoint life cycle impact indicators (ReCiPe 2016) and industrial benchmarks such as the GSK sustainable solvent framework, enabling a multidimensional ranking of solvent candidates. Experimentally validated case studies involving APIs such as paracetamol, meloxicam, piroxicam, and cytarabine confirmed the approach's robustness, adaptability to various crystallization conditions, and effectiveness in supporting single and binary solvent screening and design.

Green foundation

1. This work advances green chemistry by introducing *SolECOs*, a sustainable-by-design digital platform for solvent and solvent mixture selection, integrating predictive modelling and comprehensive sustainability assessment to support greener pharmaceutical manufacturing.

2. *SolECOs* predicts optimal single or binary solvents for 1186 APIs using a database of 30 000 + solubility points for 30 solvents, ranked *via* 23 Life Cycle Assessment indicators and the GSK Environmental Assessment Framework. Predictions were experimentally validated for four APIs.

3. Greener performance could be achieved by expanding the database to include more bio-based solvents, adding renewable feedstock pathways in LCA, and integrating real-time process data for adaptive, in-process solvent design.

1. Introduction

More than 80% of small-molecule pharmaceuticals are delivered in solid form.[1,2] As a fundamental step in solid–liquid phase transformation, crystallization is pivotal in pharmaceutical manufacturing, where solvent selection serves as a key determinant of process efficiency and product quality.[3,4] An appropriately chosen crystallization solvent affects solubility and supersaturation behavior, which in turn enables control over crystal properties and, more importantly, ensures high product yield.[5,6] With the growing adoption of Green Chemistry[7] and Quality by Design (QbD)[8] in pharmaceutical manufacturing, solvent selection has become central to addressing not only product quality and process efficiency but also sustainability, regulatory compliance, and life cycle assessment (LCA). This shift is reflected in guidelines such as ICH Q8-Q12,[9] the REACH regulation,[10] and initiatives including the Green Pharmacy Initiative and Pharmaceuticals in the Environment (PiE), which emphasize reduced volatile organic compounds (VOCs) emissions, lower carbon footprint, and improved atom economy.[11–14]

On average, it takes approximately 12.5 years and up to £1.15 billion to bring a new drug to market.[15] While many

[a]*Department of Chemical Engineering, Loughborough University, Leicestershire, LE113TU, UK. E-mail: b.benyahia@lboro.ac.uk*
[b]*School of Design and Creative Arts, Loughborough University, Leicestershire, LE113TU, UK*

factors contribute to this painstaking and costly process, inefficiencies in crystallization solvent selection remain a persistent bottleneck, particularly in unit operations such as API synthesis, crystallization, liquid–liquid extraction, wash-filtration, drying, and granulation.[16] Despite decades of accumulated experience, solvent selection in crystallization continues to rely heavily on empirical rules and trial-and-error strategies.[17–19] These approaches are time-consuming, resource-intensive, and heavily reliant on expert judgment, which collectively limits efficiency and scalability in process development.[20,21]

Driven by these challenges, solvent selection is gradually transitioning from traditional empiricism to data-driven intelligent screening and machine learning (ML)-assisted design approaches.[22–26] Technically, the objective of solvent selection aligns closely with solubility prediction, an area that has seen substantial progress in recent years.[27–33] However, based on accurate characterization of solubility behavior, a critical differentiating step lies in effectively linking API dissolution behavior in a given solvent with its environmental footprints under variable real-world production conditions.

Solvent selection approaches are developed to meet various single or multi-objective targets, such as maximizing product yield, controlling crystal polymorphism, and enhancing solvent sustainability.[34,35] From an industrial standpoint, a key and often unavoidable goal is to reduce environmental impact while still achieving the desired product yield. Computer-Aided Molecular Design (CAMD) serves as a systematic approach to identify crystallization solvents.[36–41] Karunanithi et al.[42] developed a framework combining CAMD, database screening, and experiments, with attention to crystal morphology. Wang and Lakerveld[43] presented a systematic approach for the simultaneous optimization of process conditions and solvent selection for continuous crystallization including solvent recycling. Chai et al.[44] introduced the Grand Product Design (GPD) model, incorporating technical, economic, and regulatory factors. Liu et al.[45] proposed an ML-integrated CAMD approach focused on solvent recovery. Watson et al.[46] designed a CAMD-based method for optimal solvent blend selection in pharmaceutical crystallization, capable of simultaneously determining ideal process temperature, solvent and anti-solvent species, and their compositions.

To improve practical applicability, efforts have focused on user-friendly tools that integrate process needs, solvent properties, and environmental constraints.[47–51] Larsen et al.[52] developed a green solvent selection tool for printed electronics, organizing a wide range of solvents based on Hansen solubility parameters and sustainability indicators. Similarly, an interactive tool has been developed to support solvent selection by incorporating chemical functionality, physical properties, regulatory considerations, and Safety, Health, and Environmental (SHE) impacts.[53]

Despite the emergence of various solvent design and selection frameworks in recent years, significant limitations remain. Firstly, the implementation complexity of many methods and models hinders their broader adoption. While computational methods demonstrate strong performance in specific case studies, they typically rely on intricate parameter settings and assumptions, making it difficult to generalize or directly apply the results in real-world scenarios. Without substantial expertise, users may struggle to navigate these tools effectively, thereby diminishing the cost and efficiency advantages of non-experimental approaches. Secondly, the "optimal" solvents identified by computational approaches sometime lack practical feasibility. These designed solvents may face challenges in industrial adoption due to high synthesis costs, limited commercial availability, supply chain constraints, or issues related to transportation and storage. Thirdly, existing methods often lack flexibility and consistency, particularly in sustainability assessment. Current industrial practices rely on diverse and sometimes inconsistent sustainability indicators, each emphasizing different aspects – such as carbon footprint, toxicity, biodegradability, or energy consumption during production. The absence of a unified evaluation framework makes it difficult to comprehensively assess and compare the environmental impacts of solvents or solvent mixtures.

To address these limitations, this study sets out three key objectives. First, to improve usability, we develop a computationally efficient and user-oriented platform that enables solvent selection without requiring advanced modeling expertise or high-performance computing. Second, to enhance practical relevance, the framework focuses on commonly used solvents and their binary combinations, avoiding hypothetical or industrially inaccessible candidates. Third, to accommodate diverse sustainability criteria, the methodology incorporates multiple assessment schemes, allowing engineers and experimentalists to select evaluation criteria aligned with specific environmental, health, or regulatory frameworks.

2. Methodology

A computational framework with a sequential workflow was developed to streamline solvent selection and screening for APIs in both single and binary solvent systems. The process began with the construction of a comprehensive solubility database with over 30k data points for 1183 APIs in organic solvent/water systems, covering both single and binary compositions. Additionally, the environmental impact of solvent usage was systematically assessed by evaluating the sustainability performance of 30 solvents and their mixtures. To enable quantitative solvent selection, the 3D molecular structures of APIs were characterized using 347 molecular descriptors. Key descriptors were identified through a combination of random forest modeling and Monte Carlo sensitivity analysis.

Hybrid modeling approaches integrating ML and theoretical methods were developed. A Polynomial Regression Model-based Multi-Task Learning Network (PRMMT) was designed with multiple shared layers to accommodate different design requirements. The Point-Adjusted Prediction Network (PAPN) was developed for solubility prediction at specific temperatures, while the Modified Jouyban–Acree Model-based Neural Network (MJANN) was tailored to handle
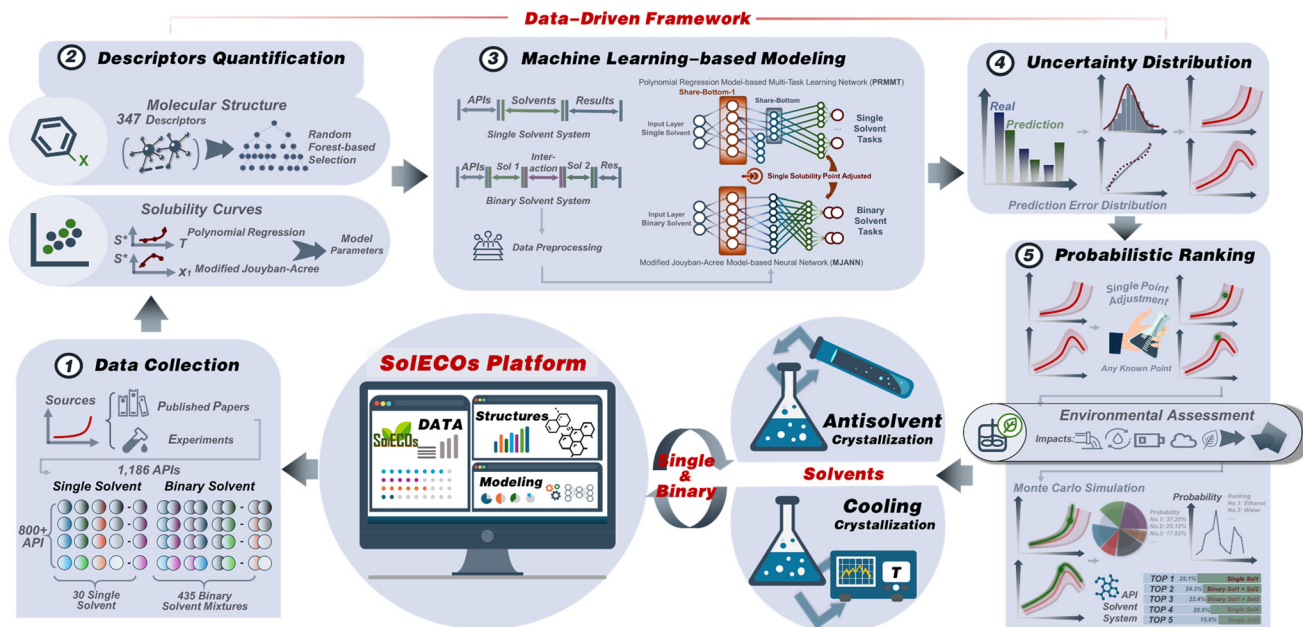
Fig. 1 Data-driven framework for sustainable solvent selection in the *SolECOs* platform.

the complexities inherent to the design of binary solvent systems.

To enhance reliability, discrepancies between predicted and actual solubility values in the validation set were quantified and mapped to optimal probability distributions of prediction residuals. By preserving probability variations across different distribution values, a robust solvent selection framework was established, ensuring reliable solvent recommendations. The entire workflow has been integrated into the user-friendly *SolECOs* platform, providing an efficient tool for solvent screening and selection (Fig. 1).

2.1. Data collection

2.1.1. Database for solubility. To accommodate the diverse range of potential model compounds involved in the crystallization process, the solubility database was curated based on compound value and complexity, selecting a total of 1186 high-value compounds that are essential for the preparation and production of APIs approved by the World Health Organization (WHO). The solubility data for these compounds in 30 commonly used single solvents and binary solvent mixtures were systematically retrieved through comprehensive literature searches and database queries, including published articles and Reaxys.[54] Only data explicitly reporting the use of pure solvents were included to ensure consistency and avoid the influence of mixed-isomer or denatured solvents.

To facilitate model development and validation, the entire dataset was divided into three independent subsets, each serving a specific purpose. Approximately 70% of the data was allocated to the training set for model development, while 30% was used as a validation set to fine-tune model perform-

ance. Additionally, a separate test set, consisting of data from 20 independent APIs, was reserved for final model evaluation.

2.1.2. Database for solvent environmental categories and impact quantification. Thirty solvents widely used were selected for this study (Table S1). The selection process was meticulously designed to balance physicochemical diversity, industrial relevance, and environmental sustainability.[55] Polar protic solvents such as methanol, ethanol, and water were included for their hydrogen-bonding capabilities, while non-polar solvents like hexane and heptane represent low-dielectric environments. Industrial relevance guided the inclusion of widely used solvents such as chloroform, dichloromethane, acetone, and ethyl acetate. Additionally, solvents like acetic acid, pyridine, 1,4-dioxane, and cyclohexanone were incorporated for their roles in tuning polarity and solubility in process-critical applications. Environmental considerations were also incorporated into the selection process. While solvents like chloroform and benzene were retained for benchmarking purposes despite known risks, greener alternatives such as DMSO, oxolane, and selected alcohols were included to promote more sustainable crystallization practices.

The environmental impact of solvents was quantitatively evaluated using SimaPro 9.5 and the ReCiPe 2016 v1.1 method, based on the Ecoinvent 3 database, in accordance with ISO 14040-14043 standards.[56] Both midpoint and endpoint indicators were considered to provide a comprehensive evaluation of environmental impact (Fig. S1). The midpoint approach enabled a detailed examination of each solvent's impact across different environmental categories, while the endpoint approach focused on the overall long-term environmental consequences. In addition to the sustainability indicators provided by the methodology, a weighted sum-

mation of impact factors (eqn (1)) was also considered, where higher a Sustainability Throughput Index (STI) values indicated a greater negative environmental impact.

To further strengthen the sustainability assessment, the platform also incorporated the solvent evaluation framework proposed by the regularly updated GSK Solvent Sustainability Guide.[57,58] This method categorizes solvents into ten distinct subcategories, which are subsequently aggregated into four major sustainability category scores and ultimately synthesized into a composite sustainability score (G), as described in eqn (2)–(5). All scores range from 1 to 10, where a low score indicates poor sustainability, while a high score reflects favorable environmental performance. Overall, the platform offers 23 different sustainability indicators for users to choose from.

$$STI = \sum_{i=1}^{N} w_i \cdot I_i \quad (1)$$

$$H = \sqrt{HH \times EP} \quad (2)$$

$$S = \sqrt{FE \times RS} \quad (3)$$

$$E = \sqrt{air \times aqua} \quad (4)$$

$$W = \sqrt{I \times R \times BT \times VOC} \quad (5)$$

$$G = \sqrt[4]{H \times S \times E \times W} \quad (6)$$

2.2. Descriptors determination

The Molecular Operating Environment (MOE) software[59] was employed to calculate molecular descriptors, encompassing both 2D and 3D properties which include topological, geometric, and electronic properties. After computation, the descriptors were reviewed and exported for further analysis. To identify and select the most independent descriptors, a random forest model and Monte Carlo simulations based on random forest were utilized. More information can be found in the SI.

2.3. Modeling

2.3.1. Thermodynamic and empirical modeling. Classical thermodynamic and empirical models provide an effective means to describe solubility variations with temperature in single solvents, and with both temperature and composition in binary mixtures. The proposed digital platform employs an empirical Polynomial Regression (PR) model to describe solubility variations with temperature in single-solvent systems. The general form of this model is provided in eqn (7):

$$S = \alpha_0 + \alpha_1 \cdot T + \alpha_2 \cdot T^2 + \cdots + \alpha_n \cdot T^n \quad (7)$$

where S is the solubility value, T is the temperature. α_n are the model's coefficients. Given that solubility often exhibits nonlinear behavior with temperature, employing a quadratic function ($n = 2$) can effectively capture this trend.

The Jouyban–Acree (JA) model is widely utilized to correlate the solubility of solutes with both temperature and the initial composition of binary solvent mixtures. This model effectively captures the dependence of solution behavior on solvent composition and temperature in multi-solvent systems. One of the key advantages of the JA model is its simplified three-parameter structure, which significantly enhances computational efficiency and makes it well-suited for integration with ML frameworks.[60–62] The general form of the JA model is presented in eqn (8).

$$\ln(x_1) = x_2{}^0 \cdot \ln(x_1)_2 + x_3{}^0 \cdot \ln(x_1)_3$$
$$+ x_2{}^0 \cdot x_3{}^0 \cdot \left(\frac{\beta_0}{T} + \frac{\beta_1}{T}(x_2{}^0 - x_3{}^0) + \frac{\beta_2}{T}(x_2{}^0 - x_3{}^0)^2 \right) \quad (8)$$

where x_1 is the mole fraction solubility of the solute; $x_2{}^0$ and $x_3{}^0$ are the initial mole fractions of two solvents in the solute-free, binary solvent mixtures, respectively; β_0, β_1 and β_2 are model parameters; and $(x_1)_2$ and $(x_1)_3$ refer to the corresponding mole percentage solubility of compound in two solvents. Thermodynamic model parameters were considered as target variables for ML modeling. Additional information is available in the SI.

2.3.2. ML modeling. The Polynomial Regression Model-based Multi-Task Learning Network (PRMMT) developed in this study adopts a shared-bottom architecture to predict the solubility-temperature profiles of compounds across multiple solvents. This approach improves computational efficiency by enabling simultaneous predictions across 30 different tasks while leveraging shared representations. The model consists of a fully connected shared-bottom layer followed by task-specific branches. Each branch comprises two dense layers optimized through hyperparameter tuning, with dropout layers applied to mitigate overfitting. The model predicts three solubility-related outputs per task, leading to a total of 90 outputs. To ensure physically meaningful predictions, custom loss constraints enforce non-negativity and monotonicity of solubility with respect to temperature. Hyperparameter tuning, including the number of units, dropout rates, and learning rates, is performed using Keras Tuner with a random search strategy across 400 trials. The final model, trained for up to 1000 epochs, adaptive optimizer is applied to minimize the Mean Absolute Error (MAE). Post-training, inverse transformation is applied to restore the standardized solubility predictions to their original scale. Model performance is evaluated based on the MAE across tasks. By comparing the predicted and actual solubility values across each task in the validation set, the uncertainty in each prediction is evaluated. This uncertainty reflects the variability inherent in the model's predictions, and with the current scale of data, it represents the predictive uncertainty of the model. This uncertainty can be used to assess the reliability of the predictions, especially when extending the model to predict solubility for new compounds or solvents.

A Point-Adjusted Prediction Network (PAPN) and the Modified Jouyban–Acree-based Neural Network (MJANN) were also developed to predict the solubility of APIs in solvents at a single temperature point, as well as their solubility in binary

mixed solvents. These models follow the same framework as previous studies.[60,63] The inputs and outputs of the models are presented in Tables 1 and S2, and the procedure for residual distribution fitting and probability estimation is detailed in the SI.

2.3.3. User interface. The user interface is developed with PySide6 and provides an interactive and user-friendly platform that integrates data input, model execution and result visualization. It adopts a multi-tab layout that organizes solubility prediction, uncertainty analysis, single-point adjustment, binary solvent evaluation and sustainability assessment in a structured manner. Users can load and save files, configure model parameters and initiate computations through an intuitive graphical environment with real-time feedback using clickable buttons, progress indicators and status tracking. Matplotlib-based visualization supports scatter plots, uncertainty distributions and 3D representations of solubility trends and sustainability indicators. The sustainability module categorizes solvents based on selected indicators and provides graded recommendations using classification and radar charts. To ensure efficiency and responsiveness, computational tasks run in the background using QThread and QRunnable for smooth multitasking.

2.4. Comparison: simulation and experimental solubility determination

To assess the accuracy and reliability of the computational framework, prediction results were systematically compared with experimental data and widely used existing methodologies. Experimental solubility measurements were conducted using the *Crystalline* instrument (Technobis, Netherlands) to provide a direct comparison with predicted results. Additionally, the solubility prediction module in PSE gPROMS, a widely used commercial process simulation software for pharmaceutical process modeling, was also employed for comparison.

The predictive performance of the model was evaluated using multiple statistical metrics, including MAE, Root Mean Squared Error (RMSE), Root Mean Squared Log Error (RMSLE), and the coefficient of determination (R^2). Given the variability in scale across thermodynamic and empirical parameters and solubility values, there is a risk that solvents with lower solubility might be underestimated by the model, potentially leading to biased exclusion in decision-making. To address this, RMLSE was adopted as a key performance metric, as it penalizes underprediction more strongly than conventional

metrics. Further details on the experimental procedures, computational methodologies, and evaluation metrics can be found in the SI.

3. Results and discussion

3.1. Data construction and determination of input descriptors

3.1.1. Data curation. The types of solvents and their occurrence frequency in the solubility database, along with the distribution of solubility data, are illustrated in Fig. 2. While most small-molecule pharmaceuticals operate within moderate temperature ranges, we have comprehensively compiled and visualized all available data capturing temperatures up to 250 °C to serve as a foundational database for potential future studies. However, during the model construction phase, we restricted our dataset to solubility data at temperatures below 70 °C to align with practical pharmaceutical conditions.

The density and size of the circles in Fig. 2(a) represent the frequency of solubility data points across different solvent-temperature combinations. A noticeable clustering of data is observed in the solubility range of 10^{-6} to 10^1 mole percent and within the temperature range of 0–50 °C, indicating that most data falls within these conditions. Although some data points exist at temperatures above 100 °C, predominantly in the 1–10 mole percent solubility range, statistical analysis reveals that these cases account for less than 2% of the total dataset.

Compared to aqueous solubility data, solubility data measured in organic solvents are relatively limited (Fig. 2b). The 30 solvent-specific tasks defined in the PRMMT model align with the most frequently occurring solvents in Fig. 2b (see Table S1 for details). Analyzing the logarithmic solubility values ($\log S$) of the collected data indicates that water and ethanol are the most extensively represented solvents, collectively accounting for over 30% of the total dataset. In contrast, solvents such as propyl acetate and 1,2-xylene appear far less frequent, contributing to less than 5% of the dataset.

The solubility data in water exhibit a relatively narrow distribution, primarily falling within the $\log S$ range of −4 to 4 (in mole percent). Conversely, solvents like ethanol and Propan-2-one show a broader solubility distribution, suggesting that solubility variations are more pronounced across different solutes. Importantly, even for solvents with lower data avail-

Table 1 Summary of the key features of the proposed ML models

Model name	Polynomial regression model-based multi-task learning network (PRMMT)	Point-adjusted prediction network (PAPN)	Modified Jouyban–Acree-based neural network (MJANN)
Input	Representative API molecular descriptors	Representative API and solvent molecular descriptors	Representative API and solvent molecular descriptors, interaction between solvents, pure solvent solubility values
Output	PR model parameters	Solubility of API in temperature T	JA model parameters
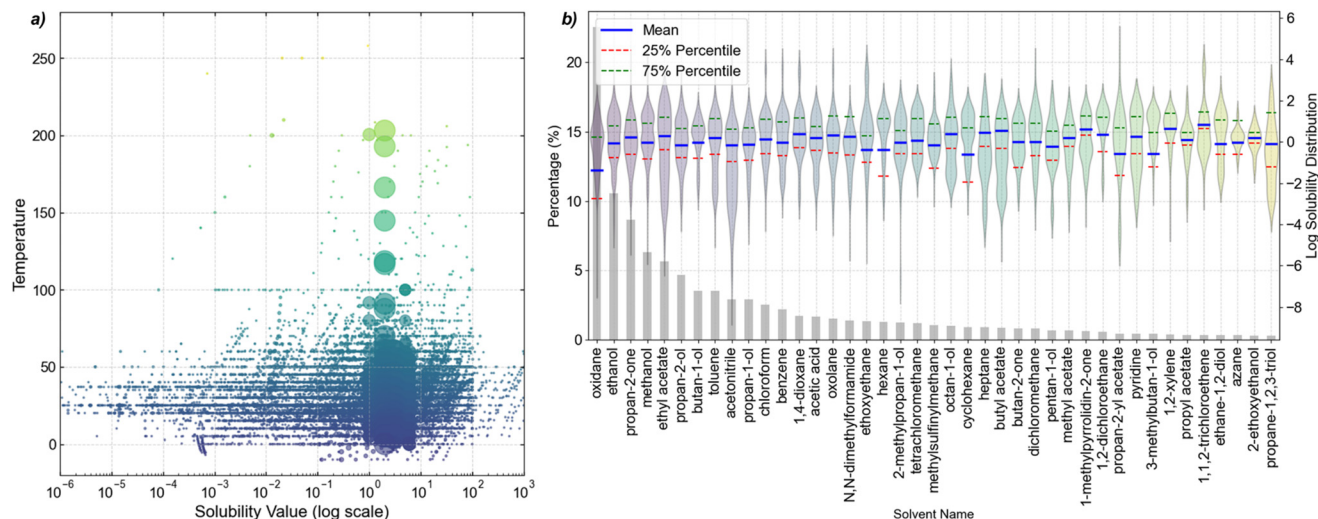
Fig. 2 Solubility and solvent frequency analysis. (a) Solubility distribution across different temperatures. The *x*-axis represents solubility values (log scale), and the *y*-axis shows temperature (°C). Each bubble corresponds to a solubility data point, with bubble size representing the occurrence frequency of solubility values and color intensity mapped to temperature, increasing with higher thermal conditions. (b) Combined violin and bar plots for solvent frequency and solubility distribution. The violin plot shows solubility distributions for solvents with data frequency > 50, where width represents solubility range. The blue line marks the mean, while red and green dashed lines indicate the 25th and 75th percentiles. Below, the gray bar chart displays solvent occurrence frequency. The right *y*-axis shows solubility on a log scale. Colors of the violin plots are assigned using the *viridis* palette solely for distinguishing different solvents.

ability, the dataset does not exhibit an overly concentrated distribution, maintaining a relatively diverse range of solubility values. This diverse solubility distribution underscores the effectiveness and representativeness of the dataset.

Ideally, for model training, a uniform distribution of log*S* values across the entire dataset would be preferred. However, due to practical limitations in data availability, no log*S*-based pruning was applied to the organic solvent dataset. This ensures that the dataset retains its inherent diversity, which is crucial for robust model performance and generalizability.

3.1.2. Descriptor development. The descriptors consist of two categories: quantitative characterization of solute/solvent 3D structures and temperature-dependent solubility curve. The temperature-dependent solubility profiles of compounds in single and binary solvent systems were parameterized using the PR model and JA model, each defined by three fitted parameters.

For single-solvent solubility prediction, since each solvent prediction task was assigned to independent parallel tasks, explicit solvent descriptors were not required. Instead, the selected descriptors needed to comprehensively capture API molecular characteristics. To determine the most relevant descriptors, random forest modeling combined with Monte Carlo sensitivity analysis and an independent random forest approach were applied to assess descriptor importance. Tables S3 and S4 list the top 25 molecular descriptors, along with their definitions. Their importance rankings, after Unit Vector Normalization, are visualized in Fig. 3. While some variations in ranking exist, most high-ranking descriptors exhibit consistent trends.

Key descriptors include GCUT_SLOGP, which incorporates both structural features (*via* graph cut) and hydrophobicity (*via*

log*P*), descriptors related to the heat of formation of the compound, distance and adjacency matrices of heavy atoms, descriptors describing mass distribution relative to the molecular center of mass, and those characterizing molecular flexibility. To minimize redundancy, a representative heat of formation descriptor was selected, and a Pearson correlation analysis (Fig. S2) was performed to ensure descriptor independence.

For binary solvent systems, additional considerations were made for solvent importance and solvent–solvent interactions. Key solvent descriptors included molecular weight, AM1_dipole, and ASA (accessible surface area). The AM1_dipole represents the dipole moment calculated using the AM1 Hamiltonian, while ASA quantifies the solvent-accessible surface area.

3.1.3. Solvent environmental assessment. The ReCiPe method was employed to calculate midpoint and endpoint indicators, serving as the sustainability assessment framework in this study to quantitatively evaluate the environmental impact of 30 solvents under the same usage conditions. The results are presented in Fig. 4 and Table S5.

Under midpoint indicators, solvents such as pyridine, propan-1-ol, and oxolane exhibit significant environmental burdens across multiple impact categories, including global warming potential, marine and freshwater ecotoxicity, and ozone layer depletion. Furthermore, propan-1-ol and oxolane demonstrate notable effects in human health-related categories, particularly in carcinogenic and non-carcinogenic toxicity. These solvents not only pose potential risks to workers and end-users throughout their life cycle but also contribute to long-term environmental degradation due to waste emissions that impact ecosystems. In contrast, water and solvents such as
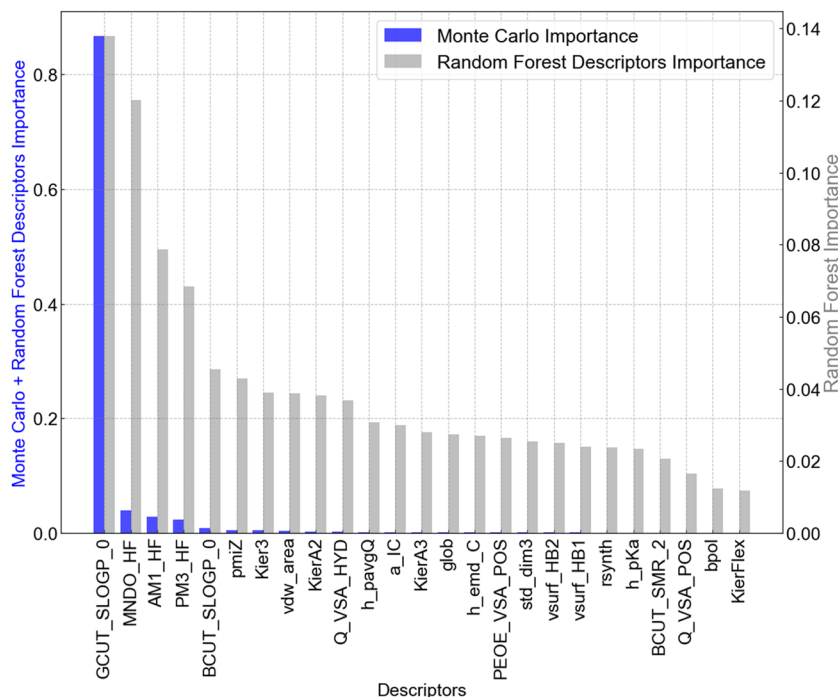
Fig. 3 Statistical analysis of descriptor importance values determined by combined random forest model and monte carlo *vs.* random forest model.

toluene, and 1,2-xylene which exhibit relatively lower impact values across most categories, may be more environmentally sustainable options for API purification and production (Fig. 4a).

The endpoint indicators integrate the midpoint assessment results, providing a more comprehensive evaluation of the overall environmental impact (Fig. 4b). The endpoint analysis reveals that propan-1-ol, acetonitrile, pyridine, and *N*-methyl-2-pyrrolidone exhibit the most pronounced environmental impacts across multiple categories, particularly concerning human health and ecosystem damage. In contrast, solvents such as water, heptane, hexane, and ethanol demonstrate relatively lower overall environmental impact values, especially in resource depletion categories, indicating potential advantages in environmental sustainability.

A comparison between the ReCiPe method and the GSK method (Table S1 and Fig. 4c) reveals both similarities and discrepancies in solvent rankings. These differences primarily arise from variations in evaluation frameworks, numerical processing methodologies, and data sources. Although an attempt was made to establish a correspondence between the GSK classification and the Midpoint indicators in Fig. S1, complete alignment remains challenging due to fundamental differences in category definitions. Moreover, numerical processing methodologies differ between the two approaches. In Fig. 4a and b, the ReCiPe method assigns equal weighting to all Midpoint and Endpoint indicators, followed by a direct summation of impact scores, whereas the GSK method applies a square-root transformation (eqn (6)) to normalize variations across subcategories. Differences in data sources also contribute to the observed ranking discrepancies. The GSK Solvent

Sustainability Guide is based on industry-specific data accumulated within GSK, using a simplified scoring system tailored to manufacturing operations, whereas the ReCiPe method provides a broader environmental perspective but remains susceptible to regional policy influences and assumptions embedded in its methodological framework.

It is important to recognize that no single green assessment method can fully address the inherent challenges of quantifying qualitative sustainability attributes, and the prioritization of solvent selection criteria may vary depending on the specific application context. This study aims to establish a multifaceted evaluation platform as a complementary approach to well-established sustainability guidelines that are widely recognized and trusted by users.

3.2. Model integration

The performance of the PRMMT, PAPN, and MJANN models on the whole dataset and testing set is summarized in Tables 2, S6 and Fig. S3. The close agreement between the results on the whole dataset and the independent test set demonstrates that all three models achieve consistent predictive accuracy across unseen data, effectively capturing the solubility behavior of APIs in diverse solvents and showing no evidence of overfitting. Since the predicted values correspond to model parameters, achieving absolute numerical accuracy does not necessarily indicate an improvement in predictive performance. The accuracy of single-temperature-point predictions is generally higher than that of the overall solubility curve fitting, as evidenced by the superior accuracy demonstrated by the PAPN model compared to the other two models.
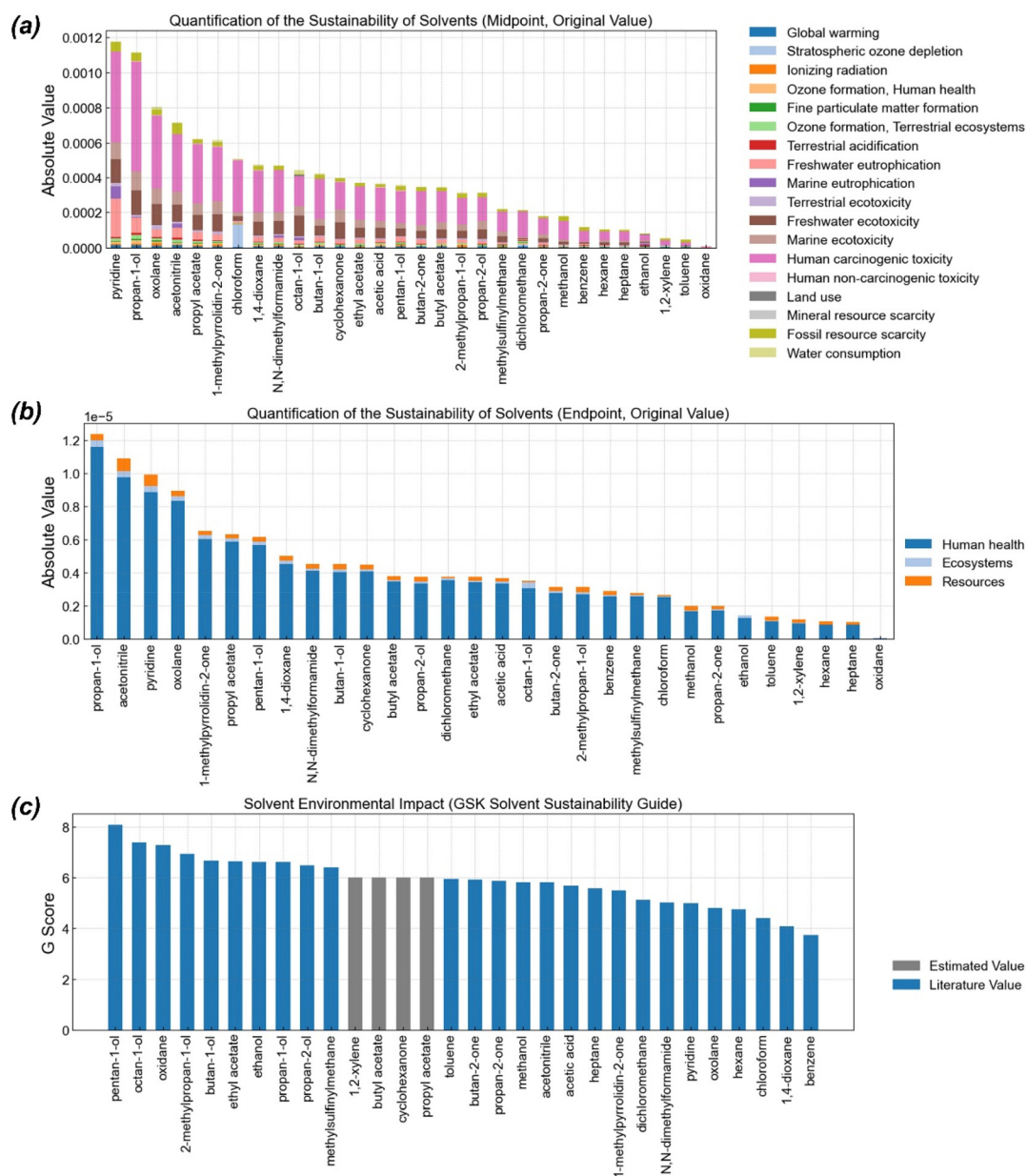
Fig. 4 The environmental impact of solvents analyzed using: (a) the ReCiPe midpoint method, (b) the ReCiPe endpoint method, and (c) the GSK solvent sustainability guide.

**Table 2** Prediction performance of PRMMT, PAPN and MJANN Models on the testing set

Average evaluation metrics	PRMMT model	PAPN model	MJANN model
MAE	0.584	0.472	0.994
RMSE	0.963	0.821	1.391
RMSLE	0.268	0.380	0.351

To further evaluate prediction reliability, the differences between predicted and real values were analyzed to determine the error distribution shown in Fig. 5. The probability value (*p*-value) was used as an indicator of the confidence in the accuracy of the model's description. The error distribution fit for all tasks within the PRMMT model resulted in *p*-values predominantly concentrated between 0.8 and 1, with an average exceeding 0.6, indicating a high degree of accuracy in describing prediction errors. The t-distribution was observed most frequently, suggesting that the statistical treatment of errors places greater emphasis on the tail regions, allowing for a more flexible and conservative estimation by accommodating variations in the degrees of freedom across different tasks. By mapping the error distribution to specific tasks, it is possible to determine the probability distribution of the predicted values within ±*x* intervals. Theoretically, restricting the range
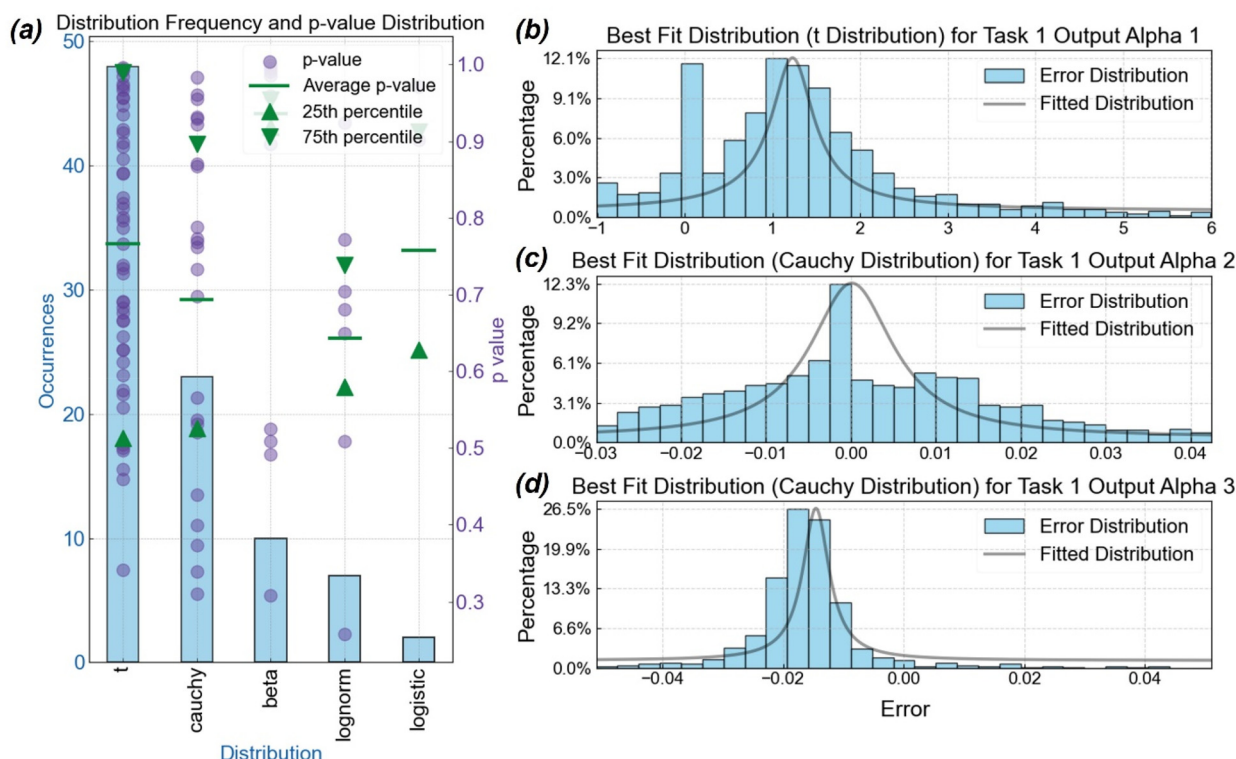
Fig. 5 Optimal fit of error distributions and frequency-p-value analysis of various distributions: (a) frequency and p-value distribution of 90 outputs; (b)–(d) best fit of the error distribution for task 1 output alpha 1 to output alpha 3.

of output parameters could reduce the prediction uncertainty. However, the objective of this study is to establish a predictive framework that provides a broader range of possibilities rather than aiming for extreme precision. This aligns with the principle in pharmaceutical solvent selection, where R&D departments aim to avoid overlooking potential solvents or solvent combinations. Consequently, the PRMMT model outputs three parameters, accompanied by error distributions incorporating t-distribution, Cauchy distribution, Beta distribution, log–normal distribution, and logistic distribution. Similar error distributions are established for the PAPN and MJANN models. However, to optimize computational efficiency, only the uncertainty range from the PRMMT model is considered in subsequent applications.

To enhance robustness, a predictive framework was developed by integrating the three models. Given the PAPN model's superior accuracy in single-point temperature predictions, the framework prioritizes its predictions. The PRMMT model serves as the foundation, providing initial predictions along with corresponding uncertainty estimates. The PAPN model is then used to refine the predictions at specific temperature points, acting as correction anchors. A tolerance value (T_v) is introduced to ensure that the solubility curve predicted by the PRMMT model falls within the confidence interval of the PAPN model's single-temperature predictions. T_v represents a user-defined error margin, which can be adjusted based on the prediction confidence of the PAPN models. For example, if

PAPN predictions are considered highly reliable, a lower T value can be set to enforce stricter constraints. Alternatively, an approximate predictive error of 10% ($T_v = 0.1$) can be used as a default tolerance for correction in the platform. The influence of different T values on predictive performance is further explored in case studies. The computational precision (step size) is defined by the number of Monte Carlo simulation samples, with 10^6 samples chosen to balance accuracy and computational efficiency.

For binary solvent mixtures, the PAPN-corrected single-point predictions serve as curve endpoints in the MJANN model. In real-world applications, solubility values may vary depending on measurement methodologies. This study accounts for this variability by offering users the flexibility to manually define correction points, T_v, and error distributions. In this scenario, correction points can be derived from experimental data, and the predictive error distribution is replaced by actual experimental error.

Based on the selected 30 pure solvents, a theoretical total of 435 binary solvent combinations is possible. However, due to partial or complete immiscibility of certain solvent pairs at specific temperatures, some binary mixtures were excluded from this study, and the final selection of binary solvent systems is provided in Fig. S4. The computational step size for binary solvent mixtures is another critical parameter. Given that APIs may exhibit limited solubility in mixed solvents, a step size of 0.1 (*i.e.*, solvent fraction increments of 0.05) was

chosen to accurately capture potential extreme solubility points. More precise calculations, such as a step size of 0.01, are feasible but would require significantly greater computational resources. In most cases, the endpoint solubility values for binary mixtures are obtained from model predictions. However, since some users may prefer to input their own solubility data, the framework also allows for the manual definition of binary solvent system endpoints, providing greater flexibility in practical applications.

3.3. Case study

Four APIs were selected to validate the robustness of the methodology, considering structural diversity and relevance to crystallization processes. Paracetamol (*N*-acetyl-*para*-aminophenol, APAP) was included due to the extensive literature data and research experience available for comparison. Meloxicam (MLX) and Piroxicam (PXC) can also get historical results from published papers for consideration and these two API exhibit a certain degree of similarity in chemical structure. In contrast, Cytarabine (AraC), with its relatively complex structure and limited literature reports, was selected to facilitate comparative analysis and provide insights into alternative solvent choices and experimentally verified. The study examined variations in cooling temperature ranges and different sustainability considerations. These variables are detailed in Table 3.

3.3.1. Accuracy assessment with established data. The crystallization process considered in APAP case **a1** involved cooling from 40 °C to 15 °C. The first step was to predict the thermodynamic solubility of the API in target solvents over a broad temperature range. Fig. 6 and Table S7 present solubility predictions using different models, where subfigures (a)–(h) illustrate the performance of the PRMMT and PAPN models in pure solvents, and subfigures (e)–(l) demonstrate the MJANN model's performance in binary solvent mixtures. A key insight derived from the RMSLE result is that the PRMMT model, after PAPN correction, retained competitive predictive accuracy. However, in specific cases, the corrected model exhibited enhanced capability in capturing absolute solubility variations at discrete temperature intervals. Noteworthy discrepancies were observed between the PRMMT model and single-point PAPN corrections, particularly in predicting APAP solubility in toluene, where deviations arose due to intrinsic model differences and sensitivity to tolerance thresholds. Despite these variations, the general solubility-temperature trend was effectively captured, indicating that the corrected solubility deviations at individual temperature points remained within an acceptable range. In contrast, the SAFT-γ Mie GC method exhibited substantial deviations from experimental data in over half of the evaluated cases, failing to reliably reproduce the monotonic increase in solubility with temperature in single-solvent systems. For binary solvent systems, solubility predictions were inherently dependent on the accuracy of endpoint solubility estimations, with the models in this study effectively capturing both monotonic solubility trends and potential co-solvent effects.

Subsequent analyses were conducted to evaluate green solvent selection for API crystallization. In case **a1**, the ReCiPe method Midpoint indicators were employed, where individual indicators were aggregated using an equal-weight summation approach to compute the STI, assigning equal weights to all parameters. The objective was to identify solvent systems with the lowest possible environmental impact. To systematically assess the sustainability of both single and mixed solvents, the sustainability rankings were categorized into ten distinct grades, with higher grades indicating superior environmental performance.

The APAP screening results, presented in Fig. 7, illustrate the classifications through an interactive computational interface. The left panel presents a 2D visualization, where probability and STI values are plotted against the count across different grades for each combination. Concurrently, the middle and right panels exhibit potential single-solvent and binary-solvent selections. In Grades 1 to 3, oxolane, acetonitrile, propan-1-ol, 1,4-dioxane, and *N,N*-dimethylformamide were identified as predominant solvents. These solvents are well-documented for their superior solubility performance and extensive industrial applicability; however, they are frequently associated with suboptimal green chemistry attributes. As sustainability rankings increased, solvents such as pentan-1-ol, butyl acetate, acetic acid, and methylsulfinylmethane were more frequently observed. At the highest sustainability levels (Grades 8 to 10), solvents including dichloromethane, methanol, benzene, water, heptane, toluene, hexane, and ethanol became dominant. Notably, water and ethanol emerged as particularly competitive due to their low environmental impact and high biodegradability. A holistic approach to solvent selection necessitates a multifaceted evaluation beyond solubility and environmental attributes alone. Rather than evaluating

Table 3 Summary of API cooling temperature ranges and sustainability considerations

API name	Molecular formula	Temperature range	Sustainability considerations
Paracetamol (APAP)	$C_8H_9NO_2$	**a1**: 40 °C–15 °C	**a1**: midpoint, weighted sum STI
Meloxicam (MLX)	$C_{14}H_{13}N_3O_4S_2$	**b1**: 50 °C–10 °C	**b1**: midpoint, human carcinogenic toxicity
		b2: 30 °C–5 °C	**b2**: midpoint, human carcinogenic toxicity
Piroxicam (PXC)	$C_{15}H_{13}N_3O_4S$	**c1**: 30 °C–10 °C	**c1**: endpoint, resources
		c2: 30 °C–10 °C	**c2**: GSK methodology
Cytarabine (AraC)	$C_9H_{13}N_3O_5$	**d1**: 50 °C–5 °C	**d1**: endpoint, human health
		d2: 40 °C–15 °C	**d2**: endpoint, human health
		d3: 40 °C–15 °C	**d3**: GSK methodology
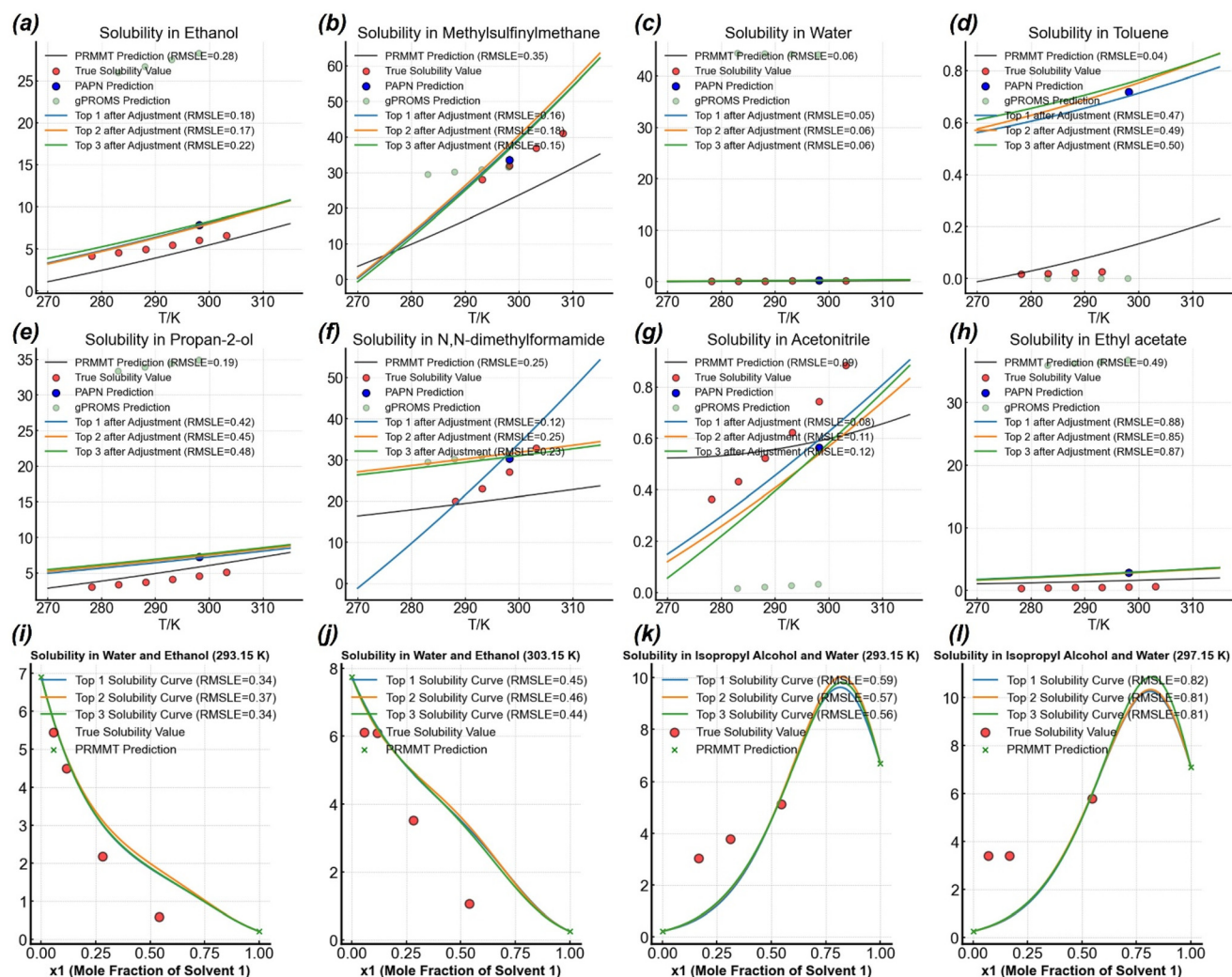
Fig. 6 Comparison of solubility predictions for various single (a)–(h) and binary solvents (i)–(l) using PRMMT, PAPN, and MJANN models across temperature (T/K) and mole fraction (x_1) conditions. Subfigures (a)–(h) show PRMMT model prediction curves along with the top three PAPN-adjusted predictions ranked by probability of accuracy compared to actual values, with RMSLE as the measure of prediction error in parentheses. Each subfigure includes scatter points representing actual solubility data, PAPN predictions, and values calculated using the SAFT-γ Mie group-contribution (GC) method. Subfigures (i)–(l) display three solubility prediction curves generated by the MJANN model along with scatter points indicating actual solubility data.

solubility or sustainability in isolation, an optimal solvent or solvent mixture should be selected based on minimizing environmental burden while maintaining adequate solubility within a target temperature range. For instance, while N,N-dimethylformamide exhibited the highest solubility potential, its substantial environmental impact relegated it to lower sustainability grades.

In consideration of binary solvent system, in Grade 1 to Grade 3, solvents such as oxolane, propan-1-ol, and pentan-1-ol remained dominant components. However, at intermediate sustainability levels, the nonlinear thermodynamic behavior of binary solvent mixtures resulted in the emergence of pentan-1-ol, hexane, benzene, and acetonitrile across multiple grades, each exhibiting relatively high probability values. At higher sustainability grades, binary solvent mixtures predominantly

incorporated solvents previously identified in top-ranked single-solvent selections, such as water, ethanol, and dichloromethane. The probability distributions across different sustainability grades also exhibited some fluctuations, as these values were influenced by the accuracy of the ML model predictions. In this case study, lower sustainability-grade mixtures generally displayed higher probability values, indicating potential uncertainties in model predictions at different sustainability levels.

Compared to existing literature, the solvent systems identified by our framework follow consistent trends. For instance, solvents such as ethanol, methanol, and Propan-2-one, defined in Grades 7 to 10, have been widely reported as effective crystallization media, particularly for obtaining the stable and metastable polymorph.[64–66] Ethanol, in particular,
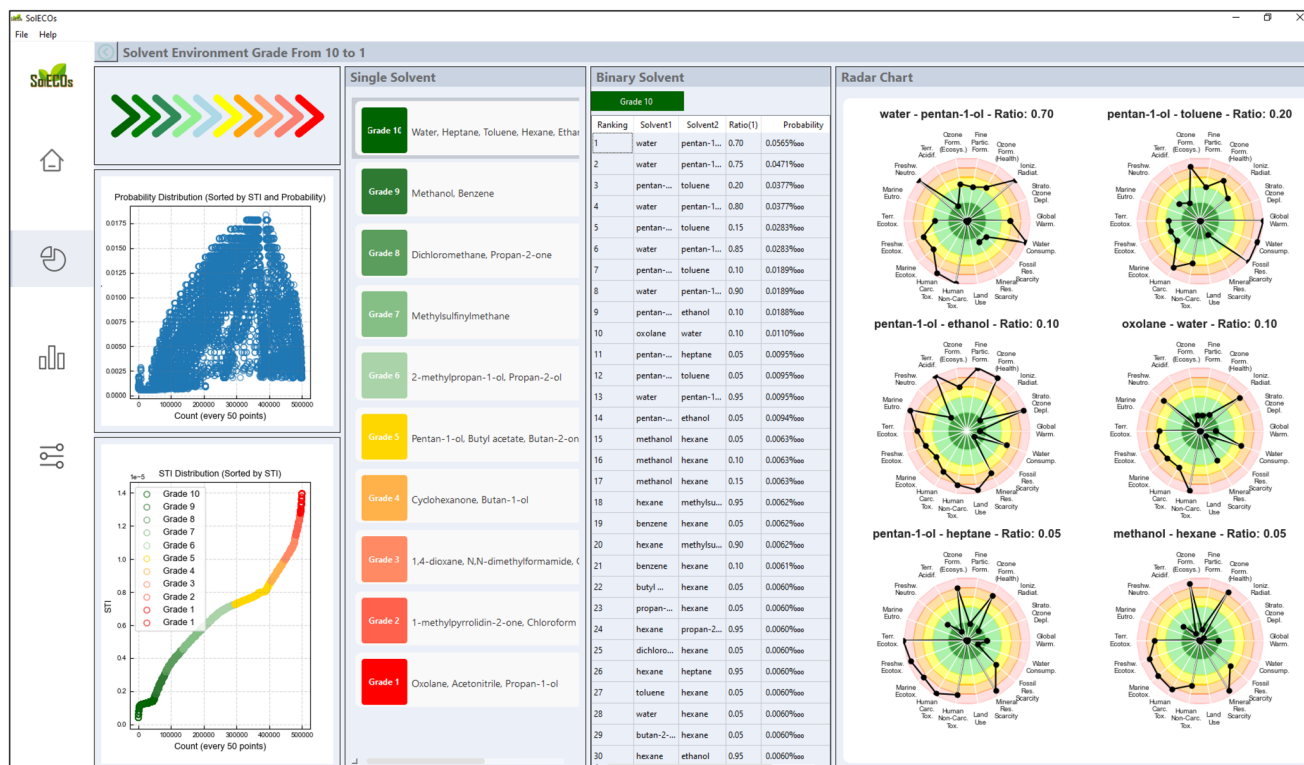
Fig. 7 *SolECOs* interface – APAP case study screenshot. The interface displays the classification of single and binary solvents by sustainability grade, including solvent identity, composition, and probability. The right panel shows radar plots of the top six binary solvent combinations with the highest probabilities, where each axis represents one of 18 midpoint indicators normalized to a 0–1 scale.

is widely used in APAP crystallization system for its strong solvating power and industrial applicability, and it also ranks as the top-performing single solvent in our framework.[64,67] Green solvents such as water and isopropanol, which were highlighted at intermediate to high sustainability grades, are also commonly employed in the literature for polymorphic control and crystallization kinetics optimization.[68] Furthermore, binary solvent systems, including water-alcohol combinations identified in our results (Table S8), have shown favorable performance in modulating solubility and directing polymorphic outcomes, in agreement with previous experimental studies.[69,70]

3.3.2. Temperature sensitivity in solvent selection. Although the predicted thermodynamic solubility-temperature profile of the API remains unchanged across different temperature gradient settings, variations in temperature conditions directly influence the theoretical solubility differentials, which in turn affect the required solvent volume for API production. Consequently, this may lead to variations in solvent grading. The single-solvent grade classification and binary solvent selection results for MLX cases **b1** and **b2** are presented in Table 4. The single-solvent classification for MLX demonstrated a high degree of consistency across different conditions. Specifically, when the temperature gradient was changed from a cooling range of 50 °C to 10 °C to a narrower range of 30 °C to 5 °C, propyl acetate was no longer present in

Grade 2, while heptane disappeared from Grade 10. This exclusion serves as a direct reflection of the model's stability mechanisms.

To ensure the reliability of the predictive outcomes, a critical threshold was established in this study, where a solvent was classified as "uncertain" if its grade ranking fell outside the top P% of solvents ranked by probability (set at 10% in this case). This approach accounts for potential classification uncertainty, acknowledging that while a solvent can be numerically assigned to a specific grade, its classification may still be subject to variability due to probabilistic ranking constraints.

The comparative analysis between MLX **b1** (50–10 °C) and MLX **b2** (30–5 °C), both evaluated using the same sustainability indicator, reveals a high degree of consistency in the classification of single solvents across Grades 1 to 10. Only minor differences were observed, such as the inclusion of propyl acetate in Grade 2 under MLX **b1**. For binary solvent selection, both cases identified highly similar solvent pairs within corresponding grade levels, with slight variations in the optimal composition ratios. Notably, the predicted probabilities for the top binary combinations were marginally higher under the lower temperature gradient in MLX **b2**, suggesting improved predictive confidence under a narrower cooling range. These results confirm that, within the proposed framework, when the sustainability evaluation method is held constant, the

Table 4 Single-solvent grading and probable green binary mixtures under case study scenarios

Case	Single solvent		Top-ranked binary solvent combinations with predicted probabilities (in ‰)	
	Grade num	Solvent name	Rank	Solvent combination
APAP a1	Grade 1	Oxolane, acetonitrile, propan-1-ol	1	Water (Solvent 1: 0.7) pentan-1-ol (Solvent 2: 0.3), probability: 0.00565
	Grade 2	1-Methylpyrrolidin-2-one, chloroform	2	Water (Solvent 1: 0.75) pentan-1-ol (Solvent 2: 0.25), probability: 0.00471
	Grade 3	1,4-Dioxane, N,N-dimethylformamide	3	Pentan-1-ol (Solvent 1: 0.2) toluene (Solvent 2: 0.8), probability: 0.00377
	Grade 4	Cyclohexanone, butan-1-ol	4	Water (Solvent 1: 0.8) pentan-1-ol (Solvent 2: 0.2), probability: 0.00377
	Grade 5	Pentan-1-ol, butyl acetate, butan-2-one, acetic acid, ethyl acetate	5	Pentan-1-ol (Solvent 1: 0.15) toluene (Solvent 2: 0.85), probability: 0.00283
	Grade 6	2-Methylpropan-1-ol propan-2-ol	6	Water (Solvent 1: 0.85) pentan-1-ol (Solvent 2: 0.15), probability: 0.00283
	Grade 7	Methylsulfinylmethane	7	Pentan-1-ol (Solvent 1: 0.1) toluene (Solvent 2: 0.9), probability: 0.00189
	Grade 8	Dichloromethane, propan-2-one	8	Water (Solvent 1: 0.9) pentan-1-ol (Solvent 2: 0.1), probability: 0.00189
	Grade 9	Methanol, benzene	9	Pentan-1-ol (Solvent 1: 0.1) ethanol (Solvent 2: 0.9), probability: 0.00188
	Grade 10	Water, heptane, toluene, hexane, ethanol	10	Oxolane (Solvent 1: 0.1) water (Solvent 2: 0.9), probability: 0.00110
MLX b1	Grade 1	Pyridine, oxolane, propan-1-ol	1	Water (Solvent 1: 0.55) dichloromethane (Solvent 2: 0.45), probability: 0.00064
	Grade 2	1-Methylpyrrolidin-2-one, chloroform, propyl acetate, acetonitrile	2	Water (Solvent 1: 0.6) dichloromethane (Solvent 2: 0.4), probability: 0.00057
	Grade 3	1,4-Dioxane, N,N-dimethylformamide	3	Dichloromethane (Solvent 1: 0.35) toluene (Solvent 2: 0.65), probability: 0.00050
	Grade 4	Butan-1-ol	4	Water (Solvent 1: 0.65) dichloromethane (Solvent 2: 0.35), probability: 0.00050
	Grade 5	Pentan-1-ol, butyl acetate, butan-2-one, acetic acid, ethyl acetate	5	Dichloromethane (Solvent 1: 0.3) ethanol (Solvent 2: 0.7), probability: 0.00048
	Grade 6	Octan-1-ol	6	Dichloromethane (Solvent 1: 0.3) 1,2-xylene (Solvent 2: 0.7), probability: 0.00044
	Grade 7	Dichloromethane, 2-methylpropan-1-ol, cyclohexanone	7	Dichloromethane (Solvent 1: 0.3) toluene (Solvent 2: 0.7), probability: 0.00043
	Grade 8	Propan-2-ol, methylsulfinylmethane	8	Water (Solvent 1: 0.7) dichloromethane (Solvent 2: 0.3), probability: 0.00043
	Grade 9	Propan-2-one, methanol	9	Dichloromethane (Solvent 1: 0.25) ethanol (Solvent 2: 0.75), probability: 0.00041
	Grade 10	Water, heptane, 1,2-xylene, benzene, toluene, hexane, ethanol	10	Dichloromethane (Solvent 1: 0.25) 1,2-xylene (Solvent 2: 0.75), probability: 0.00037
MLX b2	Grade 1	Pyridine, oxolane, propan-1-ol	1	Water (Solvent 1: 0.55) dichloromethane (Solvent 2: 0.45), probability: 0.00065
	Grade 2	1-Methylpyrrolidin-2-one, chloroform, acetonitrile	2	Dichloromethane (Solvent 1: 0.4) toluene (Solvent 2: 0.6), probability: 0.00058
	Grade 3	1,4-Dioxane, N,N-dimethylformamide	3	Water (Solvent 1: 0.6) dichloromethane (Solvent 2: 0.4), probability: 0.00058
	Grade 4	Butan-1-ol	4	Dichloromethane (Solvent 1: 0.35) 1,2-xylene (Solvent 2: 0.65), probability: 0.00052
	Grade 5	Pentan-1-ol, butyl acetate, butan-2-one, acetic acid, ethyl acetate	5	Dichloromethane (Solvent 1: 0.35) toluene (Solvent 2: 0.65), probability: 0.00051
	Grade 6	Octan-1-ol	6	Water (Solvent 1: 0.65) dichloromethane (Solvent 2: 0.35), probability: 0.00051
	Grade 7	Dichloromethane, 2-methylpropan-1-ol, cyclohexanone	7	Dichloromethane (Solvent 1: 0.3) ethanol (Solvent 2: 0.7), probability: 0.00048
	Grade 8	Propan-2-ol, methylsulfinylmethane	8	Dichloromethane (Solvent 1: 0.3) 1,2-xylene (Solvent 2: 0.7), probability: 0.00045
	Grade 9	Propan-2-one, methanol	9	Dichloromethane (Solvent 1: 0.3) toluene (Solvent 2: 0.7), probability: 0.00043
	Grade 10	Water, 1,2-xylene, benzene, toluene, hexane, ethanol	10	Water (Solvent 1: 0.7) dichloromethane (Solvent 2: 0.3), probability: 0.00043
PLX c1	Grade 1	Pyridine, acetonitrile, propan-1-ol	1	Dichloromethane (Solvent 1: 0.95) methylsulfinylmethane (Solvent 2: 0.05), probability: 0.00663
	Grade 2	2-Methylpropan-1-ol, oxolane, propan-2-ol, butan-1-ol	2	Dichloromethane (Solvent 1: 0.95) ethanol (Solvent 2: 0.05), probability: 0.00661
	Grade 3	1,4-Dioxane, pentan-1-ol, methanol	3	Chloroform (Solvent 1: 0.05) dichloromethane (Solvent 2: 0.95), probability: 0.00660
	Grade 4	N,N-Dimethylformamide, cyclohexanone	4	Water (Solvent 1: 0.05) dichloromethane (Solvent 2: 0.95), probability: 0.00660
	Grade 5	Benzene, 1-Methylpyrrolidin-2-one, butan-2-one	5	Octan-1-ol (Solvent 1: 0.05) dichloromethane (Solvent 2: 0.95), probability: 0.00660
	Grade 6	Acetic acid, ethyl acetate	6	Dichloromethane (Solvent 1: 0.9) ethanol (Solvent 2: 0.1), probability: 0.00626
	Grade 7	Butyl acetate, toluene, 1,2-xylene	7	Chloroform (Solvent 1: 0.1) dichloromethane (Solvent 2: 0.9), probability: 0.00626
	Grade 8	Hexane, heptane, propan-2-one	8	Water (Solvent 1: 0.1) dichloromethane (Solvent 2: 0.9), probability: 0.00625
	Grade 9	Methylsulfinylmethane	9	Octan-1-ol (Solvent 1: 0.1) dichloromethane (Solvent 2: 0.9), probability: 0.00625
	Grade 10	Water, chloroform, dichloromethane, octan-1-ol, ethanol	10	Dichloromethane (Solvent 1: 0.1) Ethanol (Solvent 2: 0.15), probability: 0.00592
PLX c2	Grade 1	Chloroform, 1,4-dioxane, benzene, hexane, oxolane	1	Pentan-1-ol (Solvent 1: 0.6) dichloromethane (Solvent 2: 0.4), probability: 0.00278
	Grade 2	Dichloromethane, pyridine, N,N-dimethylformamide	2	Pentan-1-ol (Solvent 1: 0.65) dichloromethane (Solvent 2: 0.35), probability: 0.00243
	Grade 3	1-methylpyrrolidin-2-one	3	Pentan-1-ol (Solvent 1: 0.7) dichloromethane (Solvent 2: 0.3), probability: 0.00209
	Grade 4	Acetic acid, heptane	4	Pentan-1-ol (Solvent 1: 0.75) dichloromethane (Solvent 2: 0.25), probability: 0.00174
	Grade 5	Propan-2-one, acetonitrile, methanol	5	Water (Solvent 1: 0.8) dichloromethane (Solvent 2: 0.2), probability: 0.00139
	Grade 6	1,2-Xylene, butyl acetate, butan-2-one, toluene, cyclohexanone	6	Pentan-1-ol (Solvent 1: 0.8) dichloromethane (Solvent 2: 0.2), probability: 0.00139
	Grade 7	N/A	7	Octan-1-ol (Solvent 1: 0.8) dichloromethane (Solvent 2: 0.2), probability: 0.00139
	Grade 8	Propan-2-ol, methylsulfinylmethane	8	Water (Solvent 1: 0.85) dichloromethane (Solvent 2: 0.15), probability: 0.00104
	Grade 9	Ethyl acetate, butan-1-ol, ethanol, propan-1-ol	9	Pentan-1-ol (Solvent 1: 0.85) dichloromethane (Solvent 2: 0.15), probability: 0.00104
	Grade 10	2-Methylpropan-1-ol, water, pentan-1-ol, octan-1-ol	10	Octan-1-ol (Solvent 1: 0.85) dichloromethane (Solvent 2: 0.15), probability: 0.00104
AraC d1	Grade 1	Acetonitrile, propan-1-ol	1	Water (Solvent 1: 0.60) 1,2-xylene (Solvent 2: 0.40), probability: 0.00637
	Grade 2	2-Methylpropan-1-ol, oxolane, propan-2-ol, butan-1-ol	2	1,2-Xylene (Solvent 1: 0.350) ethanol (Solvent 2: 0.650), probability: 0.00581
	Grade 3	1,4-Dioxane, pentan-1-ol, methanol	3	Water (Solvent 1: 0.650) 1,2-xylene (Solvent 2: 0.350), probability: 0.00558
	Grade 4	N,N-Dimethylformamide	4	1,2-Xylene (Solvent 1: 0.300) ethanol (Solvent 2: 0.700), probability: 0.00504
	Grade 5	Benzene, 1-methylpyrrolidin-2-one, butan-2-one, cyclohexanone	5	Water (Solvent 1: 0.700) 1,2-xylene (Solvent 2: 0.300), probability: 0.00479
	Grade 6	Acetic acid, ethyl acetate	6	1,2-Xylene (Solvent 1: 0.250) ethanol (Solvent 2: 0.750), probability: 0.00427
	Grade 7	Butyl acetate, propan-2-one, toluene, 1,2-xylene	7	Water (Solvent 1: 0.750) 1,2-xylene (Solvent 2: 0.250), probability: 0.00401
	Grade 8	Hexane, methylsulfinylmethane	8	1,2-Xylene (Solvent 1: 0.200) ethanol (Solvent 2: 0.800), probability: 0.00350
	Grade 9	Dichloromethane	9	Water (Solvent 1: 0.800) 1,2-xylene (Solvent 2: 0.200), probability: 0.00322
	Grade 10	Water, chloroform, octan-1-ol, ethanol	10	Octan-1-ol (Solvent 1: 0.800) 1,2-xylene (Solvent 2: 0.200), probability: 0.00321

Table 4 (Contd.)

Case	Single solvent		Top-ranked binary solvent combinations with predicted probabilities (in ‰)	
	Grade num	Solvent name	Rank	Solvent combination
AraC d2	Grade 1	Acetonitrile, propan-1-ol	1	Water (Solvent 1: 0.60) 1,2-xylene (Solvent 2: 0.40), probability: 0.00637
	Grade 2	2-Methylpropan-1-ol, oxolane, propan-2-ol, butan-1-ol	2	1,2-Xylene (Solvent 1: 0.350) ethanol (Solvent 2: 0.650), probability: 0.00581
	Grade 3	1,4-Dioxane, pentan-1-ol, methanol	3	Water (Solvent 1: 0.650) 1,2-xylene (Solvent 2: 0.350), probability: 0.00558
	Grade 4	N,N-Dimethylformamide	4	1,2-Xylene (Solvent 1: 0.300) ethanol (Solvent 2: 0.700), probability: 0.00504
	Grade 5	Benzene, 1-methylpyrrolidin-2-one, butan-2-one, cyclohexanone	5	Water (Solvent 1: 0.700) 1,2-xylene (Solvent 2: 0.300), probability: 0.00479
	Grade 6	Acetic acid, ethyl acetate	6	1,2-Xylene (Solvent 1: 0.250) ethanol (Solvent 2: 0.750), probability: 0.00427
	Grade 7	Butyl acetate, propan-2-one, toluene, 1,2-xylene	7	Water (Solvent 1: 0.750) 1,2-xylene (Solvent 2: 0.250), probability: 0.00401
	Grade 8	Hexane, methylsulfinylmethane	8	1,2-Xylene (Solvent 1: 0.200) ethanol (Solvent 2: 0.800), probability: 0.00350
	Grade 9	Dichloromethane	9	Water (Solvent 1: 0.800) 1,2-xylene (Solvent 2: 0.200), probability: 0.00322
	Grade 10	Water, chloroform, octan-1-ol, ethanol	10	Octan-1-ol (Solvent 1: 0.800) 1,2-xylene (Solvent 2: 0.200), probability: 0.00321
AraC d3	Grade 1	Benzene, 1,4-dioxane, chloroform	1	Pentan-1-ol (Solvent 1: 0.500) 1,2-xylene (Solvent 2: 0.500), probability: 0.00803
	Grade 2	Dichloromethane, hexane, N,N-dimethylformamide	2	Pentan-1-ol (Solvent 1: 0.550) 1,2-xylene (Solvent 2: 0.450), probability: 0.00725
	Grade 3	1-Methylpyrrolidin-2-one	3	Pentan-1-ol (Solvent 1: 0.600) 1,2-xylene (Solvent 2: 0.400), probability: 0.00647
	Grade 4	Acetic acid	4	Pentan-1-ol (Solvent 1: 0.650) 1,2-xylene (Solvent 2: 0.350), probability: 0.00569
	Grade 5	Propan-2-one, acetonitrile, methanol	5	Pentan-1-ol (Solvent 1: 0.700) 1,2-xylene (Solvent 2: 0.300), probability: 0.00491
	Grade 6	1,2-Xylene, butyl acetate, butan-2-one	6	Pentan-1-ol (Solvent 1: 0.750) 1,2-xylene (Solvent 2: 0.250), probability: 0.00414
	Grade 7	N/A	7	Octan-1-ol (Solvent 1: 0.750) 1,2-xylene (Solvent 2: 0.250), probability: 0.00399
	Grade 8	Propan-2-ol, methylsulfinylmethane	8	Pentan-1-ol (Solvent 1: 0.800) 1,2-xylene (Solvent 2: 0.200), probability: 0.00336
	Grade 9	2-Methylpropan-1-ol, butan-1-ol, propan-1-ol	9	Water (Solvent 1: 0.800) 1,2-xylene (Solvent 2: 0.200), probability: 0.00322
	Grade 10	Water, pentan-1-ol, octan-1-ol	10	Octan-1-ol (Solvent 1: 0.800) 1,2-xylene (Solvent 2: 0.200), probability: 0.00321

influence of temperature on solvent classification is limited. However, temperature can still affect the fine-tuning of binary solvent compositions and the associated selection probabilities.

The final solvent screening results for the MLX cases underscore the practical relevance and industrial compatibility of the proposed framework (Tables S9 and S10). Ethanol and methanol, ranked in Grades 10 and 9 respectively, have been experimentally validated for meloxicam dissolution and crystallization.[71,72] Notably, ethanol-water mixtures (Grade 10) are highlighted in patent EP-1462451A1 as preferred media for Form I crystallization, offering controlled polarity and enhanced purity.[73] Such alcohol-water co-solvent systems are widely used in industrial crystallization processes, where temperature control enables high yield, polymorphic stability, and improved solid properties.[74] These examples confirm that high-grade solvents identified by the framework are not only environmentally favorable but also well-aligned with established industrial practices.

3.3.3. Sustainability metrics and method-driven variability.
A comparison of the solvent grading results obtained using the ReCiPe method Endpoint indicator (Resources, PLX **c1**) and the GSK method (PLX **c2**) reveals significant discrepancies in solvent classification and prioritization. One of the most significant differences is observed in the classification of lower-grade solvents. In PLX c1, solvents such as pyridine, acetonitrile, and propan-1-ol are categorized within Grade 1, whereas in PLX **c2**, chloroform, 1,4-dioxane, benzene, hexane, and oxolane are assigned to the same category. This distinction suggests that the Endpoint indicator primarily evaluates solvents based on resource consumption and environmental toxicity, quantifying their sustainability through environmental impact scores. In contrast, the GSK method can give a more industry-oriented vision, incorporating additional considerations such as process compatibility, regulatory compliance, and environmental, health, and safety factors.

Significant differences also emerge in the classification of mid-tier solvents. In PLX **c1**, Grade 5 includes benzene, 1-methylpyrrolidin-2-one, and butan-2-one, whereas in PLX **c2**, Grade 5 consists of propan-2-one, acetonitrile, and methanol. Notably, benzene is assigned a relatively high grade in the Endpoint indicator but is ranked significantly lower in the GSK method, suggesting differences in risk perception between the two methodologies. Both methods, however, classify DMSO at a relatively high grade, reflecting its recognition as an environmentally preferable solvent due to its low toxicity and high biodegradability.

Water, widely regarded as a green solvent, is consistently assigned Grade 10 in both methods, reinforcing its high priority for sustainability. However, notable discrepancies exist in the classification of chlorinated solvents. In PLX **c1**, chloroform and dichloromethane are also categorized as Grade 10, whereas in PLX **c2**, dichloromethane is assigned a significantly lower ranking at Grade 2, likely due to the stricter regulatory constraints imposed on chlorinated solvents within the GSK framework.

The binary solvent selection results further emphasize the methodological divergence between the two approaches. In PLX **c1**, the top-ranked binary solvent combinations are predominantly characterized by a high proportion of dichloromethane mixed with small amounts of other solvents. In contrast, PLX **c2** follows a different ranking trend, where pentan-1-ol and dichloromethane mixtures dominate, and solvent ratios vary more significantly.

From a probability distribution perspective, the binary solvent combination probabilities calculated in PLX **c1** are notably higher than those in PLX c2 (PLX **c1** maximum:

0.00663 *vs.* PLX **c2** maximum: 0.00278). This suggests that the Endpoint indicator is more likely to identify high-probability solvent combinations, whereas the GSK method, due to its broader consideration of multiple influencing factors and smaller numerical differentials across criteria, results in lower overall probability variations among binary solvent combinations. Fig. 8 illustrates the distribution of binary solvent combinations within Grade 10 for PLX **c1** (a) and PLX **c2** (b). In PLX **c1**, a limited number of combinations, particularly those involving dichloromethane, show markedly higher probabilities. This indicates a strong preference for dichloro-
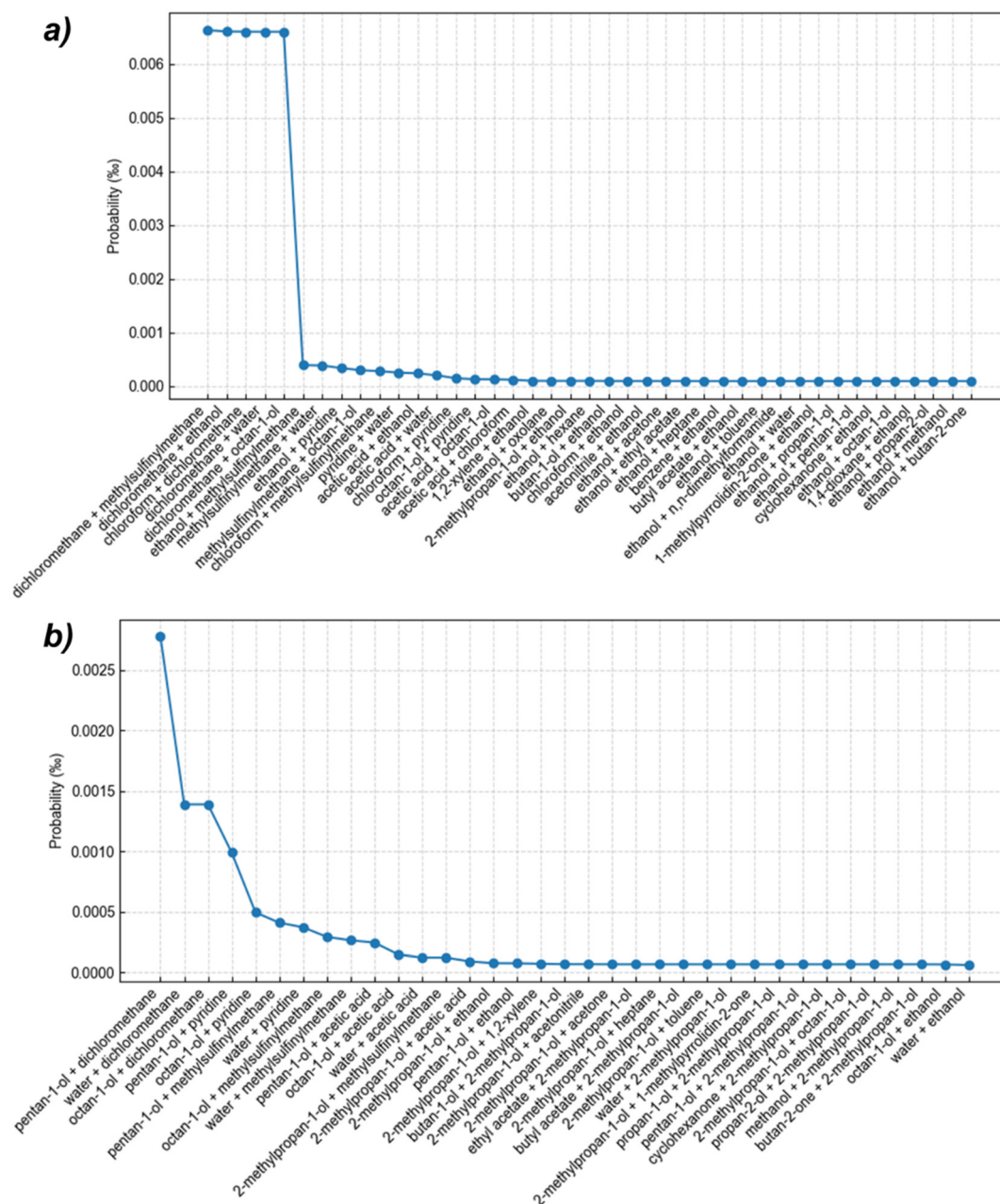
Fig. 8 Statistical analysis of occurrence probabilities for solvent combinations and corresponding dominant constituents in Grade 10 in Case **c1** (a) and Case **c2** (b).

methane-based mixtures under the ReCiPe Endpoint indicator. In contrast, PLX **c2** displays a more balanced probability distribution across several solvent systems. Although dichloromethane remains among the top candidates, the wider spread suggests that the GSK method allows greater flexibility and supports more diverse solvent selection strategies. When focusing on traditional green solvents such as water or ethanol as one component in binary mixtures, the Endpoint indicator (PLX **c1**) yields not only more concentrated high-probability combinations but also a greater number of qualifying binary systems within Grade 10. In contrast, the GSK-based method (PLX **c2**) identifies fewer combinations but distributes probability more evenly. See Tables S11 and S12 for detailed listings.

3.3.4. Experimental validation in under-explored systems. The AraC case study evaluated single- and binary-solvent grading under two temperature gradients and two distinct sustainability criteria. The results suggest that, particularly in the present case, temperature is not the primary determinant of solvent classification. Under the Endpoint Human Health evaluation, despite cases **d1** and **d2** employing different temperature ranges, the overall solvent grading trends remained consistent. Solvents such as acetonitrile, 1-methylpyrrolidin-2-one, and 1,4-dioxane were consistently ranked in lower grades,

whereas water, ethanol, and methanol, widely recognized as green solvents, were consistently assigned higher grades. While temperature settings can help refine solvent classification, the key consideration remains the sustainability focus of the evaluation methodology. By contrast, the differences between solvent evaluation methods were more pronounced. In AraC **d1** and 2, benzene, 1,4-dioxane, and N,N-dimethylformamide were categorized within Grade 3–5, whereas in AraC **d3** these solvents were assigned lower rankings, falling into Grade 1 or 2.

Fig. 9 compares the Grade 10 binary solvent systems with water as a fixed component under AraC **d1** and AraC **d3**. While both methods identify common co-solvents such as 1,2-xylene and acetone, AraC **d1** includes less sustainable options like dichloromethane, whereas AraC **d3** favors greener solvents such as 2-methylpropan-1-ol. This reflects the broader tolerance of Endpoint-based screening *versus* the stricter sustainability constraints of the GSK metric. Compositionally, AraC d1 allows wider water ratio ranges, indicating greater flexibility, while AraC **d3** yields narrow, sharply defined optima, suggesting higher selectivity. Both methods consistently rank water + 1,2-xylene highest, though optimal ratios differ. Systems like water + ethanol show lower probabilities and narrower ranges, reflecting limited suitability. These observations are consistent
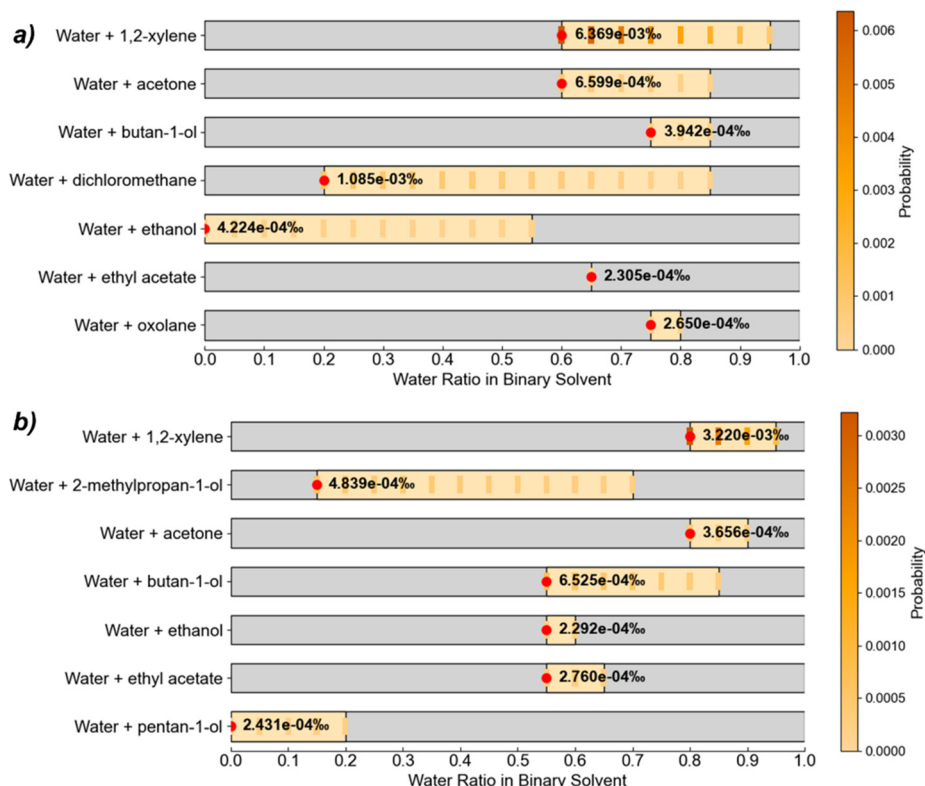
Fig. 9 Compositional distributions of water–containing binary solvent systems identified in grade 10 for case **d1** (a) and case **d3** (b). Each subplot presents binary solvent systems comprising water and a co-solvent, selected under Grade 10 criteria for two cases. The *x*-axis represents the water fraction in the binary mixture, and the *y*-axis lists the corresponding solvent combinations. Orange bars indicate compositions included in Grade 10, with darker shades representing higher occurrence probabilities. Gray bars denote excluded compositions. Red circles mark the most probable composition for each combination.

with the trend discussed in Section 3.3.3, where the GSK indicator led to a more selective and compositionally constrained solvent space compared to the more inclusive Endpoint approach. A detailed statistical summary of Grade 10 binary solvent systems containing water or ethanol as one component is provided in the SI (Tables S13 and S14).

To further validate the predictive accuracy, experimental verification was conducted, with results shown in Fig. 10. The model demonstrated a high level of predictive performance. In the binary solvent combination design, the Grade 10 combination of dichloromethane (Solvent 1: 0.05) and 1,2-xylene (Solvent 2: 0.95) (D–X combination) exhibited the highest probability, but this does not imply that it is the most sustainable choice. In subfigures (e–h) the D–X combination exhibited higher environmental impacts in categories such as ozone formation, global warming, and fossil resource scarcity compared to water–ethanol combinations at any ratio. This underscores the fact that although Endpoint 1 Human Health was selected

as the evaluation criterion, it does not mean that all binary solvent combinations within the same grade exhibit identical environmental impacts. The solvent selection process should be tailored to the user's specific sustainability priorities, ensuring a balance between high predictive robustness (probability) and optimal sustainability impact within the selected evaluation framework.

3.4. Future developments and current limitations

Prediction accuracy is a decisive factor influencing solvent selection and design. In this study, the developed platform allows users to customize key parameters, such as correction tolerance, and provides the capability to manually incorporate prior thermodynamic knowledge of specific systems, enabling tailored adjustments to prediction outcomes. In addition to these user-defined parameters, factors such as the reliability of thermodynamic databases, the availability of experimental data, and the structure of machine learning models also
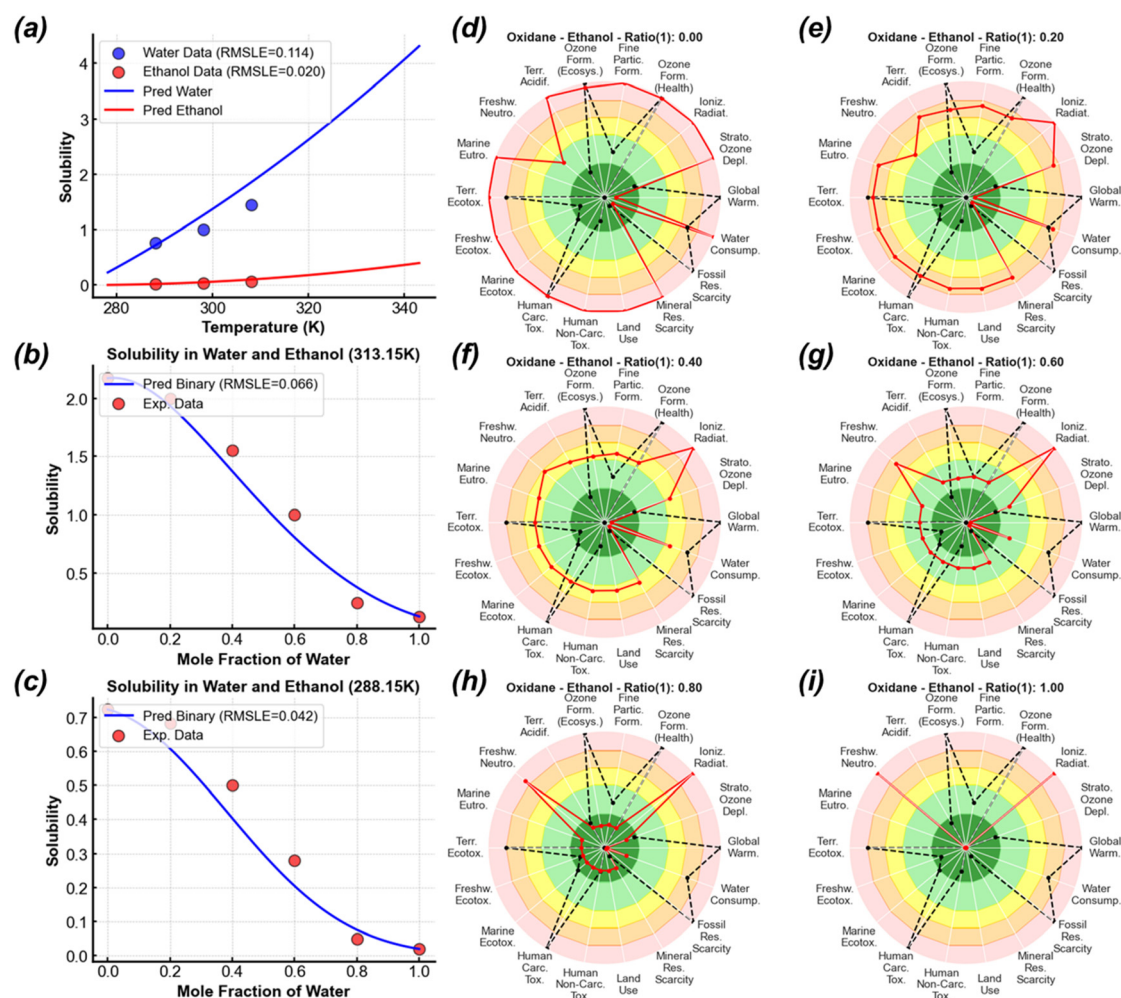
Fig. 10 Solubility behavior and environmental impact assessment of water–ethanol binary solvent systems. (a–c) illustrate the solubility of AraC in water and ethanol, comparing experimental data with model predictions. (d–i) present the environmental impact assessment for producing an equivalent amount of AraC using binary water–ethanol mixtures at solvent ratios of 0.00–1.00, shown as radar plots of 18 midpoint environmental impact indicators normalized to a 0–1 scale. The black dashed line represents the baseline (highest-probability solvent system), while the red solid lines correspond to the sustainability assessment results for the current study's ethanol-water compositions.

impact model performance and the final solvent classification results. However, in designing an efficient and user-friendly ecosystem, the platform aims to provide an intuitive and accessible interface, ensuring that users can achieve solvent screening and optimization without requiring in-depth knowledge of complex model hyperparameters. By allowing adjustments to critical variables while maintaining an optimized structural framework, the platform balances flexibility and usability, enabling users to focus on the practical application of single and binary solvent selection for a given API, rather than engaging in intricate model tuning.

Despite the robustness demonstrated by the developed methodology, the most reliable solvent selection and optimization strategy still necessitates experimental calibration to refine predictive accuracy. Pre-calibrated experimental data help control error margins, mitigating the risk of cumulative inaccuracies arising from model approximations and prediction errors. Furthermore, while machine learning-based solvent screening has demonstrated strong predictive capabilities, its accuracy remains inherently constrained by the quality and diversity of training data. Expanding high-quality experimental datasets will be critical for further enhancing the predictive reliability of the model.

Additionally, the choice of sustainability assessment methodologies significantly influences the final solvent rankings and recommendations. Different evaluation frameworks and solvent-specific sustainability priorities, such as toxicity concerns, resource consumption, or process safety considerations, may yield varying rankings for the same solvents. Future research will focus on integrating commonly used sustainability assessment frameworks into the platform and exploring multi-objective optimization approaches. By incorporating a broader set of sustainability indicators, the solvent evaluation framework can be expanded to ensure that solvent selection accounts for both industrial applicability and environmental impact.

Building upon the platform developed in this study, future research will further focus on the digitalization of solid–liquid separation processes, integrating prediction, design, and optimization into the comprehensive, intelligent solvent selection and process optimization *SolECOs* platform. This development will not only enhance solvent selection efficiency but also improve the overall effectiveness of crystallization and separation processes. By incorporating real-time process monitoring and adaptive optimization, the platform will evolve into a data-driven intelligent tool, offering more precise, efficient, and sustainable solutions for pharmaceutical and chemical process design.

4. Conclusion

This study presents the solvent design and selection module of *SolECOs*, a comprehensive data-driven platform for sustainable pharmaceutical manufacturing. *SolECOs* integrates a curated solubility database of over 30 000 data points covering 1186 APIs and 30 solvent systems with thermodynamically informed machine learning models to support solvent-related decision-making in crystallization processes.

The modeling framework includes a polynomial regression-based multi-task learning network (PRMMT) for temperature-dependent solubility profiling, a point-adjusted prediction network (PAPN) for single-temperature correction, and a modified Jouyban–Acree neural network (MJANN) for binary solvent prediction. These models enable interpretable and uncertainty-aware predictions across a wide range of crystallization conditions. To further support environmentally informed decision-making, *SolECOs* incorporates comprehensive sustainability evaluations based on both the ReCiPe 2016 life cycle impact framework and the GSK Solvent Sustainability Guide, allowing users to balance solubility performance with environmental priorities.

The entire workflow is implemented in an interactive graphical interface, facilitating user-friendly data input, model execution, and visualization of solubility curves, confidence intervals, and sustainability indicators. Case studies involving representative APIs, including paracetamol, meloxicam, piroxicam, and cytarabine, validate the robustness and applicability of this module across varying crystallization scenarios. As a foundational part of the broader *SolECOs* platform, this module demonstrates how data-driven modeling and sustainability metrics can be integrated to guide solvent selection in early-stage pharmaceutical process development.

Author contributions

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Abbreviations

APAP	Paracetamol
AraC	Cytarabine
CAMD	Computer-aided molecular design
JA Model	Jouyban–Acree model
MJANN	Modified Jouyban–Acree-based neural network
MAE	Mean absolute error
ML	Machine learning
MLX	Meloxicam

MSE	Mean squared error
PR Model	Polynomial regression (PR) model
PRMMT	Polynomial regression model-based multi-task learning network
PXC	Piroxicam
R^2	Coefficient of determination
RMSE	Root mean square error
RMSLE	Root mean squared logarithmic error
STI	Sustainability throughput index
p-Value	Probability value (confidence level of predictive error distribution fitting)
T_v	Tolerance value

Data availability

Trial use of the *SolECOs* platform is available upon request from the corresponding author. All primary data underlying this study are included in the article and its supplementary information (SI).

The supplementary Information includes detailed descriptions of the model development methodology, case study configurations, simulation workflows, and additional supporting tables and figures. See DOI: **https://doi.org/10.1039/d5gc04176g**.

Additional materials and datasets can be made available by the corresponding author upon reasonable request.

Acknowledgements

References

1 B. M. Couillaud, P. Espeau, N. Mignet and Y. Corvis, *ChemMedChem*, 2019, **14**, 8–23.
2 M. Gibson, *Drugs Pharm. Sci.*, 2001, **199**, 199.
3 B. Benyahia, R. Lakerveld and P. I. Barton, *Ind. Eng. Chem. Res.*, 2012, **51**, 15393–15412.
4 J. Liu and B. Benyahia, *Ind. Eng. Chem. Res.*, 2024, **63**, 7300–7314.
5 Y. Liu, Y. Ma, C. Yu, Y. Gao, K. Li, L. Tong, M. Chen and J. Gong, *Green Chem.*, 2022, **24**, 5779–5791.
6 Y. Ma, W. Li, J. Liu, G. Shang, H. Yang, J. Gong, Z. K. Nagy and B. Benyahia, *AIChE J.*, 2025, e18931.
7 P. T. Anastas and J. C. Warner, *Green chemistry: theory and practice*, Oxford university press, 2000.
8 U. S. Food and Drug Administration (FDA), Pharmaceutical Quality for the 21st Century: A Risk-Based Approach, 2004. Available from: **https://www.fda.gov/about-fda/reports/pharmaceutical-quality-21st-century-risk-based-approach-progress-report**.
9 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), ICH quality guidelines Q8-Q12: Pharmaceutical development, quality risk management, pharmaceutical quality system, development and manufacture of drug substances, and lifecycle management, ICH, 2009–2017. Available from: **https://www.ich.org/page/quality-guidelines**.
10 European Commission, Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Regulation, 2006. Available from: **https://environment.ec.europa.eu/topics/chemicals/reach-regulation_en**.
11 International Pharmaceutical Federation (FIP), FIP Statement of Policy on Environmental Sustainability within Pharmacy Practice: Green Pharmacy Practice, 2016. Available from: **https://www.fip.org/file/5618**.
12 European Federation of Pharmaceutical Industries and Associations (EFPIA), The Management of Pharmaceuticals in the Environment (PIE) - Questions and Answers, 2019. Available from: **https://www.efpia.eu/media/25281/pie-qa.pdf**.
13 D. Kralisch, D. Ott and D. Gericke, *Green Chem.*, 2015, **17**, 123–145.
14 D. Ott, D. Kralisch, I. Denčić, V. Hessel, Y. Laribi, P. D. Perrichon, C. Berguerand, L. Kiwi-Minsker and P. Loeb, *ChemSusChem*, 2014, **7**, 3521–3533.
15 C. Burns, *Pharm. J.*, 2019, **https://pharmaceutical-journal.com/article/feature/drug-development-the-journey-of-a-medicine-from-lab-to-shelf** (accessed March 21, 2025).
16 B. Benyahia, in *Computer Aided Chemical Engineering*, Elsevier, 2018, vol. 41, pp. 141–157.
17 M. Escriba-Gelonch, V. Hessel, M. C. Maier, T. Noel, M. F. Neira d'Angelo and H. Gruber-Woelfler, *Org. Process Res. Dev.*, 2018, **22**, 178–189.
18 L. Zhang, H. Mao, Q. Liu and R. Gani, *Curr. Opin. Chem. Eng.*, 2020, **27**, 22–34.
19 T. J. Campbell, C. D. Rielly and B. Benyahia, in *Computer Aided Chemical Engineering*, Elsevier, 2024, vol. 53, pp. 343–348.
20 S. Jia, P. Yang, Z. Gao, Z. Li, C. Fang and J. Gong, *CrystEngComm*, 2022, **24**, 3122–3135.
21 Y. Ma, S. Wu, E. G. J. Macaringue, T. Zhang, J. Gong and J. Wang, *Org. Process Res. Dev.*, 2020, **24**, 1785–1801.
22 T. Zhou, Z. Lyu, Z. Qi and K. Sundmacher, *Chem. Eng. Sci.*, 2015, **137**, 613–625.
23 P. Cysewski, T. Jelinski, M. Przybylek, W. Nowak and M. Olczak, *Pharmaceutics*, 2022, **14**, 2828.
24 S. J. Urwin, M. W. S. Chong, W. Li, J. McGinty, B. Mehta, S. Ottoboni, M. Pathan, E. Prasad, M. Robertson, M. McGowan, M. al-Attili, E. Gramadnikova, M. Siddique, I. Houson, H. Feilden, B. Benyahia, C. J. Brown, G. W. Halbert, B. Johnston, A. Nordon, C. J. Price, C. D. Reilly, J. Sefcik and A. J. Florence, *Chem. Eng. Res. Des.*, 2023, **196**, 726–749.

25 L. König-Mattern, E. I. Sanchez Medina, A. O. Komarova, S. Linke, L. Rihko-Struckmann, J. S. Luterbacher and K. Sundmacher, *Chem. Eng. J.*, 2024, **495**, 153524.

26 T. Yamaki, T. T. H. Nguyen, N. Hara, S. Taniguchi and S. Kataoka, *Green Chem.*, 2024, **26**, 3758–3766.

27 S. Chinta and R. Rengaswamy, *Ind. Eng. Chem. Res.*, 2019, **58**, 3082–3092.

28 R. Wang, J. Chen, Z. Song and Z. Qi, *Ind. Eng. Chem. Res.*, 2023, **62**, 5382–5393.

29 E. Rahimpour and A. Jouyban, *Liquids*, 2023, **3**, 512–521.

30 F. H. Vermeire, Y. Chung and W. H. Green, *J. Am. Chem. Soc.*, 2022, **144**, 10785–10797.

31 K. Ge and Y. Ji, *Ind. Eng. Chem. Res.*, 2021, **60**, 9259–9268.

32 A. L. Perryman, D. Inoyama, J. S. Patel, S. Ekins and J. S. Freundlich, *ACS Omega*, 2020, **5**, 16562–16567.

33 C. De Stefano, G. Lando, C. Malegori, P. Oliveri and S. Sammartano, *J. Mol. Liq.*, 2019, **282**, 401–406.

34 S. Mascia, P. L. Heider, H. Zhang, R. Lakerveld, B. Benyahia, P. I. Barton, R. D. Braatz, C. L. Cooney, J. M. Evans, T. F. Jamison, K. F. Jensen, A. S. Myerson and B. L. Trout, *Angew. Chem., Int. Ed.*, 2013, **52**, 12359–12363.

35 C. L. Mustoe, A. J. Turner, S. J. Urwin, I. Houson, H. Feilden, D. Markl, M. M. Al Qaraghuli, M. W. S. Chong, M. Robertson, A. Nordon, B. F. Johnston, C. J. Brown, J. Robertson, C. Adjiman, H. Batchelor, B. Benyahia, M. Bresciani, C. L. Burcham, J. Cardona, C. Cottini, A. S. Dunn, D. Fradet, G. W. Halbert, M. Henson, P. Hidber, M. Langston, Y. S. Lee, W. Li, J. Mantanus, J. McGinty, B. Mehta, T. Naz, S. Ottoboni, E. Prasad, P. O. Quist, G. K. Reynolds, C. Rielly, M. Rowland, W. Schlindwein, S. L. M. Schroeder, J. Sefcik, E. Settanni, H. Siddique, K. Smith, R. Smith, J. S. Srai, A. A. Thorat, A. Vassileiou and A. J. Florence, *Int. J. Pharm.*, 2025, 125625, DOI: **10.1016/j.ijpharm.2025.125625**.

36 L. Achenie, V. Venkatasubramanian and R. Gani, *Computer aided molecular design: theory and practice*, Elsevier, 2002.

37 R. Gani, *Comput. Chem. Eng.*, 2004, **28**, 2441–2457.

38 A. T. Karunanithi, C. Acquah, L. E. K. Achenie, S. Sithambaram and S. L. Suib, *Comput. Chem. Eng.*, 2009, **33**, 1014–1021.

39 N. G. Chemmangattuvalappil, *Curr. Opin. Chem. Eng.*, 2020, **27**, 51–59.

40 C. S. Adjiman and A. Galindo, *Curr. Opin. Chem. Eng.*, 2025, **47**, 101073.

41 S. Chai, E. Li, L. Zhang, J. Du and Q. Meng, *AIChE J.*, 2022, **68**, e17499.

42 A. T. Karunanithi, L. E. Achenie and R. Gani, *Chem. Eng. Sci.*, 2006, **61**, 1247–1260.

43 J. Wang and R. Lakerveld, *AIChE J.*, 2018, **64**, 1205–1216.

44 S. Chai, Q. Liu, X. Liang, Y. Guo, S. Zhang, C. Xu, J. Du, Z. Yuan, L. Zhang and R. Gani, *Comput. Chem. Eng.*, 2020, **135**, 106764.

45 Q. Liu, L. Zhang, K. Tang, L. Liu, J. Du, Q. Meng and R. Gani, *AIChE J.*, 2021, **67**, e17110.

46 O. L. Watson, A. Galindo, G. Jackson and C. S. Adjiman, in *Computer Aided Chemical Engineering*, Elsevier, 2019, vol. 46, pp. 949–954.

47 S. Kalakul, L. Zhang, Z. Fang, H. A. Choudhury, S. Intikhab, N. Elbashir, M. R. Eden and R. Gani, *Comput. Chem. Eng.*, 2018, **116**, 37–55.

48 P. Harten, T. Martin, M. Gonzalez and D. Young, *Environ. Prog. Sustainable Energy*, 2020, **39**, 1–13331.

49 Q. Liu, L. Zhang, L. Liu, J. Du, A. K. Tula, M. Eden and R. Gani, *Comput. Chem. Eng.*, 2019, **124**, 285–301.

50 J. S. Tan, L. R. Hilden and J. M. Merritt, *J. Pharm. Sci.*, 2019, **108**, 2621–2634.

51 V. Odegova, A. Lavrinenko, T. Rakhmanov, G. Sysuev, A. Dmitrenko and V. Vinogradov, *Green Chem.*, 2024, **26**, 3958–3967.

52 C. Larsen, P. Lundberg, S. Tang, J. Rafols-Ribe, A. Sandstrom, E. Mattias Lindh, J. Wang and L. Edman, *Nat. Commun.*, 2021, **12**, 4510.

53 L. J. Diorazio, D. R. J. Hose and N. K. Adlington, *Org. Process Res. Dev.*, 2016, **20**, 760–773.

54 Reaxys, (n.d.). Retrieved from **https://www.reaxys.com**.

55 S. Rohani, S. Horne and K. Murthy, *Org. Process Res. Dev.*, 2005, **9**, 858–872.

56 PRé Sustainability, *Simapro 9.5*, 2023, Available from **https://www.pre-sustainability.com**.

57 C. M. Alder, J. D. Hayler, R. K. Henderson, A. M. Redman, L. Shukla, L. E. Shuster and H. F. Sneddon, *Green Chem.*, 2016, **18**, 3879–3890.

58 R. A. Sheldon, *Green Chem.*, 2017, **19**, 18–43.

59 Molecular Operating Environment (MOE), *Version 2019.0102*, Chemical Computing Group ULC, Montreal, 2019.

60 Y. Zhao, W. Li, C. Zhao, L. Fang, H. Yang and Y. Ma, *Ind. Eng. Chem. Res.*, 2024, **63**, 16529–16544.

61 A. Jouyban and W. E. Acree, *J. Drug Delivery Sci. Technol.*, 2007, **17**, 159–160.

62 A. Jouyban and W. E. Acree, *J. Mol. Liq.*, 2018, **256**, 541–547.

63 Y. Ma, Z. Gao, P. Shi, M. Chen, S. Wu, C. Yang, J. Wang, J. Cheng and J. Gong, *Front. Chem. Sci. Eng.*, 2021, **16**, 523–535.

64 N. A. Mitchell, C. T. Ó'Ciardhá and P. J. Frawley, *J. Cryst. Growth*, 2011, **328**, 39–49.

65 Y. N. Thi, K. Rademann and F. Emmerling, *CrystEngComm*, 2015, **17**, 9029–9036.

66 P. c. C. Cruz, F. A. Rocha and A. M. Ferreira, *Org. Process Res. Dev.*, 2019, **23**, 2592–2607.

67 Y. Kim, Y. Kawajiri, R. W. Rousseau and M. A. Grover, *Ind. Eng. Chem. Res.*, 2023, **62**, 2866–2881.

68 A. Saleemi, I. I. Onyemelukwe and Z. Nagy, *Front. Chem. Sci. Eng.*, 2013, **7**, 79–87.

69 L. Keshavarz, R. R. E. Steendam, M. A. R. Blijlevens, M. Pishnamazi and P. J. Frawley, *Cryst. Growth Des.*, 2019, **19**, 4193–4201.

70 S. Chewle, F. Emmerling and M. Weber, *Crystals*, 2020, **10**, 1107.

71 N. Bolourchian, M. Nili, S. Shahhosseini, A. Nokhodchi and S. M. Foroutan, *J. Drug Delivery Sci. Technol.*, 2021, **66**, 102926.

72 D. A. Tinjacá, F. Martínez, O. A. Almanza, A. Jouyban and W. E. Acree, *J. Solution Chem.*, 2021, **50**, 667–689.

73 L. Coppi, M. Bartra Sanmarti and M. Closa Calvo, Novel crystalline forms of meloxicam and the preparation and interconversion methods thereof (EP1462451A1), *European Patent Office*, EP1462451A1, 2004.

74 L. McElroy, L. Cromie and M. Garrett, Process for the purification of meloxicam (EP1923392B1), *European Patent Office*, EP1923392B1, 2009.