|  |  |
| --- | --- |
| cetlogo ***CHEMICAL ENGINEERING TRANSACTIONS***  ***VOL. xxx, 2025*** | A publication of  aidiclogo_grande |
| The Italian Association  of Chemical Engineering  Online at www.cetjournal.it |
| Guest Editors: David Bogle, Flavio Manenti, Piero Salatino  Copyright © 2025, AIDIC Servizi S.r.l. **ISBN** 979-12-81206-21-2; **ISSN** 2283-9216 | |

Experimental Study and Modelling through Artificial Intelligence of the Separation by Azeotropic Distillation of Solvents Used in Industrial Recycling Processes

Carlo Pirolaa\*, Alessio Ferlinb, Cristiano Maesanic, Stefano Alinic, Nicola Mezzettid, Metin Ünlüd, Giulia Tonsia

aUniversità degli Studi di Milano, Dipartimento di Chimica – via Golgi, 19, 20133 Milano (MI), Italy

bUniversità degli Studi della Campania Luigi Vanvitelli, Dipartimento di Architettura e Disegno Industriale - via San Lorenzo, 31, 81031 Aversa (NA), Italy

cRadici InNova S.c. a r.l. – via Fauser, 50, 28100 Novara (NO), Italy

d Téchnéos s.r.l. – via Torre Verde, 21, 38122 Trento (TN), Italy

carlo.pirola@unimi.it

The use of aqueous acidic solvents plays a crucial role in industrial processes, in particular in the recycle operations. Very often, recycling and material recovery operations are essential for the circular economy of the products and materials present in the market. In order to make sustainable the solvent based recovery processes, it is essential to be able to recycle the solvents themselves, bringing them back to the required composition and purity. Aqueous acid solvents are largely applied in industry and their recovery by distillation can be challenging due to the strongly non ideal thermodynamic behaviour of water-acids mixtures. A possible solution is the addition of an entrainer to conduct the azeotropic distillation. As part of the funded project named RE-POLY.AI, conducted in collaboration with Radici InNova s.c. a r.l. (R&D company of RadiciGroup) and Téchnéos s.r.l., experimental tests were performed using process fluids from an industrial recycling. A three-meter continuous distillation column with 15 physical trays was used to evaluate different process conditions and identify the optimal configuration for the separation. This research was performed combining experimental test with a predictive artificial intelligence (AI) framework that integrates Bayesian optimization based on Gaussian Process modelling and ensemble Random Forest regression with uncertainty quantification capabilities. Trained on data from multiple test conditions, this AI-driven approach guided the selection of new experimental setups and provided insights into optimized separation configurations.

* 1. Introduction

Artificial intelligence (AI) is rapidly transforming the field of chemical engineering, particularly in the area of separation processes. By leveraging machine learning, optimization algorithms, and data-driven models, AI enables more efficient design, operation, and control of separation systems. AI's ability to analyze vast amounts of data and predict system behavior has led to breakthroughs in process optimization, energy savings, and the development of novel separation techniques. As a result, AI is becoming a key tool in advancing sustainable and cost-effective chemical engineering solutions. Furthermore, the importance of recycling operations in chemical engineering has grown significantly, as AI helps improve the efficiency and sustainability of recycling processes, contributing to the circular economy. More in detail, the recycle and recovery of products and materials are fundamental operations in the modern industry. Environmental and economic reasons are on the basis of these needs for all the industrial activities in different fields, as plastics (Datta and Kopczyńska, 2016), batteries (Brückner et al., 2020), metals (Wernick and Themelis, 1998) and many other compounds (Hole and Hole, 2019). Recycle processes can be very different depending on the characteristics and amount of the starting products but it is always important to design plant in which the used chemicals are not expensive and can be recycled themselves in the same recovery process, in order to decrease as much as possible the use of virgin materials. In particular, it is imperative to recover and reuse the solvents useful for the process (Triebert et al., 2021). One of the most appealing solvent for recycle operation is water combined with different organic and inorganic acid (Chea et al., 2019), thanks to its low cost, low toxicity and availability. For water acids mixture too, the recovery is fundamental both to avoid the loss of chemical compounds (acids) and to allow simple waste water treatments operations. Distillation based operation are commonly applied for this goal, but in particular with organic acid (for example carboxylic acids), the separation can be challenging due to the strongly non ideal behaviour of the water-acids mixtures (Maurer, 2006). More complex distillation unit, as the azeotropic ones can be used to solve this issue, but again the design and scale up is complex (Lei et al., 2004), in particular considering the need to minimize the cost of the process. Another important point is that in industrial recycle processes, the mixtures of water and acids are not pure (i.e. formed only by the main compounds that need to be separated) but they have often a certain amount of impurities coming from the other unit operations of the whole process. Impurities can alter the distillation operation or concentrate in some fluids that exit from the column and then must be carefully considered. In this scenario, the combination of Artificial Intelligence (AI) with experimental distillation studies enables a more efficient approach to process optimization (Durrani et al., 2018). This methodology not only improves separation efficiency but also enhances solvent recovery, reduces energy consumption, and supports the sustainability of industrial operations (Fang et al., 2024). For AI application is crucial the disposal of a sufficient number of experimental points for its training (Kang and Kang, 2024).

Starting from these considerations, in this paper the methodology and the results of the experimental activity and AI combined approach are presented for the work conducted in the funded project named RE-POLY.AI. In this project funded by the European Union - NextGenerationEU under the framework of the Italian National Recovery and Resilience Plan (PNRR), a complex separation of water and acids compounds used for recycling process developed in the industrial plants of Radici InNova s.c. a r.l. (R&D company of RadiciGroup, Italy) by azeotropic distillation was investigated. Experimental tests of azeotropic distillation on a mixture water and organic acid with an entrainer were performed in a continuous 15 trays column in bench scale by feeding the industrial mixture. The effect of the industrial impurities on the separation was considered and excluded. The experimental data were used for the training of the AI-model. A specific AI tool was programmed and created by Téchnéos company (Trento, Italy) in order to upload the experimental data, optimize the training of the model, design the most useful new experiments and finally to understand the correlations among the different operative parameters of the column for the final optimization and scale up.

The proposed methodology is, in our opinion, a good example for the emerging applications of AI in the field of the chemical engineering studies. The main goal of the paper is the description of this study and not the specific results obtained for the separation, that are covered by industrial confidentiality.

A diagram of different types of objects

AI-generated content may be incorrect.

*Fig. 1: Schematic representation of the experimental tray column.*

* 1. Experimental and Methodology

**2.1 Continuous distillation experiments**

The mixture water-acid was produced during real industrial processes in the Radici plants and it was used as received as feed for the distillation column. A specific organic compound was selected and used as entrainer and it was received from Merck company. In this paper, we will refer simply to “water”, “acid” and “entrainer” to identify the compounds. It is important to highlight that for the goal of the paper this is not a limitation. Experimental runs were performed using a distillation column equipped with 15 bubble cap trays, total height of 3 meters (Pirola, 2019). The column was designed with an on-off solenoid valve as an actuator, enabling precise control of the reflux ratio. Temperature profiles along the column were monitored using multiple thermocouples strategically placed at various plates. Heat was supplied to the reboiler with two quartz rods inserted directly into the system. A flowmeter was installed to ensure continuous monitoring of the water flow rate through the condenser. The feed points for the water/acid solution and entrainer were located in the middle and near the top of the column (Figure 1).

The composition of all the flows of the column was determined by titration and HPLC analysis. Moreover, samples were taken along the sampling points of the column (one for each tray) to evaluate the distribution of impurities on the various plates and therefore their possible influence on the distillation. To determine the amount of impurities in the different liquid mixtures taken from the column, these samples were evaporated under vacuum (300 mm Hg) until the volatile components have completely evaporated. In the last phase, the pressure is further reduced to eliminate the last traces of acid and water still retained in the solid residue of impurities. Finally, the sample is left to cool and then weighed, allowing the percentage of heavy impurities present in the analyzed solution to be determined. 17 bench scale experimental tests were conducted changing different operative parameters as feed flowrate and temperature, entrainer flowrate and entrainer, reflux ratio, power to the reboiler. The resulting experimental dataset (distillate flow and composition, residual flow and composition, impurities concentration in the different streams of the column) was used for the training of the AI model.

**2.2 AI Application**

RE-POLY.AI Web Application (RE-POLY.AI) was projected and delivered, with restricted access, for the training and application of AI on the separation process. It is based on advanced Machine Learning techniques to create an intelligent system for process parameters optimization. The system incorporates two distinct modeling approaches and operates in three key phases: 1) Learning Phase: the AI models learn from historical process data, understanding the relationships between different parameters and their impact on process outcomes; 2) Prediction Phase: using the learned patterns, the models make predictions about process performance, including standard deviations to quantify uncertainty; 3) Suggestion Phase: based on different optimization objectives (gains), the system suggests new parameter combinations to improve process performance.

The project implements two complementary modeling approaches, i.e. Ensemble Gaussian Process Model and Random Forest Regressor (Primary Model). The first approach has the aims to implement traditional Bayesian optimization, provides natural uncertainty quantification and built-in exploration-exploitation trade-off, while the second one was selected as the primary model after extensive testing as demonstrated superior prediction accuracy compared to Gaussian Process models, provided uncertainty estimates through ensemble predictions and maintained the ability to quantify prediction confidence while offering better overall performance. After thorough testing across multiple parameter combinations and scenarios, the Random Forest Regressor consistently demonstrated superior predictive performance while still providing the necessary uncertainty quantification for optimization purposes. The models were pre-trained using synthetic data from chemical simulations and optimized by employing the experimental dataset (17 runs); model validation employed data from both the datasets. Despite limited experimental runs due to cost and process complexity constraints, the Random Forest Regressor demonstrated reliable performance on small datasets and provided realistic uncertainty quantification through ensemble predictions, whereas Gaussian Process Models yielded, for given experimental setup, high uncertainty values that led to unrealistic parameter suggestions. Additional experiments will strengthen the model's reliability; however, the current approach balances practical limitations with methodological approach, enabling effective optimization within industrial constraints.

RE-POLY.AI optimizes for multiple objectives, the main target being the minimization of the acid loss in the distillate and the secondary target the minimization of reflux ration and entrainer flowrate for operational cost efficiency. The model monitors and optimizes the following six parameters: feed composition, feed flowrate, feed temperature, entrainer flowrate, reflux ratio and cooling water flowrate (for the cooling of the vapors exiting as distillate from the top of the column). RE-POLY.AI provides three crucial visualization types for parameter analysis, as shown in Fig. 2. The “Mean Prediction” area visualizes the model's expected performance across parameter spaces, where darker regions indicate predicted better performance, and specific markers show in these areas existing and suggested experiments. The “Prediction Uncertainty” area maps the model's uncertainty levels using ensemble predictions, with higher uncertainty in unexplored regions and lower uncertainty near existing data points. Finally, “Expected Improvements” results combined predictions and uncertainty, includes secondary target functions such as reflux ratio and entrainer flow rate influence on the process and balances potential gains with uncertainty. AI-based application is able to provide three types of suggestions for the new experimental test on the basis of the analysis of the available data, called “Exploitation-focused” for fine tuning around well-known results, “Exploration-focused” for the investigation of unexplored regions and “Balance-focused” that combine both exploitation and exploration, recommending points where it is possible to get results while trying new experiment areas. Each suggestion includes comprehensive metrics such as the expected distillate acid composition and its uncertainty and the calculated cost. This last parameter is a normalized, dimensionless number that represents the combined weighted costs of all your selected operative parameters. Success probability for the suggested experiments, i.e. the probability that for these experiments in laboratory it will be possible to reach steady state conditions, are given.

Moreover, RE-POLY.AI generates an in-depth analysis of the numerical results of the uploaded experiments, providing quantitative insights into several aspects. In the “Failure Analysis” section, the system employs a secondary AI model to understand what separates successful experiments from failures. In this context, “successful experiment” means an experimental test able to reach steady state conditions. The application learns from both successful and failed experiments and ranks features by their impact on success probability. In the “Importance Analysis” section, RE-POLY.AI reveals how changes in each operative parameters of the column affect your target variable (Distillate acid composition) and helps on identifying key process parameters by integrating Random Forest model’s feature importance capability. Complementary information is available in the “Main Effect Analysis” windows, where on the basis of a statistical approach, the reported results isolate each operative parameter individual contribution to the whole process outcomes by systematically varying parameters while holding others constant.

In the web AI-based application is possible to visualize in graphical maps, called “Pair visualizations”, the most important results of the numerical elaborations, showing how the operative parameters interact to influence the process results, as discussed in the following “Results” paragraph.

* 1. Results

An example of the numerical results, obtained using the laboratory scale rectification column shown in Fig.1, is reported in the following Table 1.

Table 1: Experimental results of an example of continuous distillation run. These data were collected during the same experimental test at different working times.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Measure time data and time | Distillate composition (entrainer %wt) | Distillate composition (Acid %wt) | Bottom composition (entrainer %wt) | Bottom composition (Acid %wt) | Distillate flowrate (ml/min) | Bottom flowrate (ml/min) |
| 12:28 pm | 83.1 | 0.37 | 0.19 | 57.08 | 16 | 5 |
| 13:00 pm | 83.5 | 0.16 | 0.19 | 59.67 | 16 | 5 |
| 13:38 pm | 79.4 | 0.09 | 0.00 | 57.95 | 16 | 5 |
| 14:15 pm | 81.5 | 0.03 | 0.00 | 59.47 | 16 | 5 |
| 14:50 pm | 84.6 | 0.03 | 0.00 | 58.64 | 16 | 5 |

The data reported in Table 1 is only an extract, due to space reason, of all available data and is given as an example. Actually, for each experimental performed run all the following data are available, collected for all the duration of the experiment (typically between 2 and 6 hours): room temperature, feed composition, feed flowrate, feed temperature, feed tray, impurity presence in the feed, impurities concentration, kind of entrainer, azeotrope/entrainer boiling point, azeotropic composition, entrainer mutual solubility in water, entrainer flowrate, entrainer temperature, entrainer feed tray, reflux ratio, heat provided in the reboiler, cooling water flowrate for the distillate vapor, condenser temperature, reboiler temperature, distillate composition, reboiler composition, distillate flowrate, bottom flowrate, all the temperatures of all the trays of the column. Some experiments weren’t able to reach steady state conditions, for example for a too high or not well controlled flowrate of the feed and the entrainer that caused flooding condition in the column that caused the stop of the experiment before its predicted conclusion. In the best test (for example the run whose data are reported in Table 1), the specifications required by the industrial process were reached and kept stable for the entire duration of the test. In the first two samples, the column had probably not yet reached complete stationarity, resulting in slightly elevated values of high-boiling solvent in the distillate and entrainer in the bottom stream. However, in the three subsequent samples, these values have been aligned with the expected parameters. The analysis of the distillate and bottom flow rates also confirms the material balance of the distillation process, further validating the reliability of the test. RE-POLY.AI AI model was trained and optimized employing different datasets produced along the R&D process. Some examples of the results and indication of the AI-model are reported in Table 2 for the “Feature Importance” are and in Table 3 for the “Next Experiment Suggestions” windows of the RE-POLY.AI (described in the methodology section).

Looking at the results in Table 2, the most critical operating parameters for the correct execution of the experiment identified by the model are the temperatures of the column incoming flows (feed and entrainer). Too low temperatures can unbalance the internal column flows and therefore more easily lead to flooding conditions. Instead, the most important parameters for achieving the separation specifications are the entrainer and feed flows, or rather the ratio between these two flows. This result is easily understandable since the liquid-vapor equilibria strongly depends on the quantity of entrainer compared to that of the main binary mixture to be separated.

Table 2: Feature Importance results with the optimized AI model.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Analysis | Entrainer Temperature | Reflux Ratio | Feed Temperature | Entrainer Flowrate | Feed to Entrainer Ratio | Feed Flowrate | Cooling Water Flowrate |
| Failure Analysis (%) | 46.2 | 14.3 | 12.5 | 9.2 | 8.5 | 6.1 | 3.2 |
| Importance Analysis (%) | 0.0 | 6.2 | 5.6 | 20.0 | 12.6 | 48.1 | 7.5 |

Table 3: Next experiments suggestions (example for some parameters) with the optimized AI model.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Suggestions | Entrainer Flowrate (ml/min) | Entrainer Temperature (°C) | Feed flowrate (ml/min) | Feed Temperature (°C) | Reflux Ratio | Expected Distillate Acid composition  (wt%) | Expected Distillate Acid uncertainty  (wt%) |
| 1 (Exploitation) | 7.13 | 43.52 | 2.30 | 57.65 | 0.41 | 0.0829 | 0.0708 |
| 2 (Balanced) | 7.00 | 47.40 | 1.91 | 80.20 | 0.47 | 0.0829 | 0.0708 |
| 3 (Exploration) | 13.50 | 35.39 | 2.39 | 106.46 | 1.46 | 0.0940 | 0.0719 |

The test conditions suggested by the AI ​​model reported in Table 3 allow the design and execution of new experimental tests aimed at exploring the less data-rich areas in the multidimensional space of the many operating parameters. The method has already been successfully followed for the progressive data collection for a constant refinement and optimization of the model itself.

A screenshot of a computer screen

AI-generated content may be incorrect.

*Fig. 2: Pair Visualization maps for the: a) “Mean Prediction of Acid composition”, b) “Uncertainty Mapping”; c) “Expected Improvements”. Example obtained by visualizing the combination entrainer flowrate vs. Reflux ratio.*

Finally, the AI application produces Pair Visualization maps, that are the most intuitive way to understand the process, showing how parameters interact to influence the results. These visualizations come in three complementary views, reported as example in Figure 2. More in detail, Fig. 2a shows the “Mean Prediction of Acid composition”, with predicted target value across parameters ranges, the parameter interaction effects, the optimal operating regions and clear pathways to improvement. Fig. 2b shows the “Uncertainty Mapping” that Reveals prediction confidence across your parameter space with areas of high and low certainty, reliability of predictions, risk assessment tool and guide exploration decisions. Fig. 2c reports the “Expected Improvement” and combines all the optimization objectives to show potential improvement regions, balance of multiple objective and strategic guidance for next steps. The example reported in Fig. 2 has been obtained by visualizing the “entrainer flowrate vs. Reflux ratio” combination, but many other visualizations are available by combining the different operative parameters of the column. The validation of the model is currently underway thanks to the continuous collection of new experimental data useful both for this purpose and for the progressive training of the model itself.

* 1. Conclusions

A separation process for azeotropic multistage continuous distillation has been investigated and optimized by combining experimental tests and analyses with artificial intelligence techniques. The experimental tests have demonstrated the feasibility of the azeotropic distillation process, obtaining the distillate and residue purities required by the industrial recycling process under study. Furthermore, the experimental data have been successfully used in training the artificial intelligence model that has allowed to program new experimental tests with different objectives and to identify the regions still to be explored in the space of the process operating parameters. The main correlations between the different operating parameters and the weight of each of them on the column performance and optimization have been defined and will be increasingly better defined thanks to the progressive refinement of the artificial intelligence model with the new experimental data.

Following research will be on developing a comprehensive digital twin model using Kolmogorov-Arnold Networks and other neural network architectures. The current AI application optimizes controlled parameters like reflux ratio, entrainer flowrate, while the next generation model will expand to predict process outcomes based on both controlled and uncontrolled process variables. This digital twin model will incorporate additional information such as process sensors and account for naturally occurring variations such as impurities, which affects the distillation performance but not are directly manipulated. The Kolmogorov-Arnold Networks architectures are promising for our process due to its capability to model complex non-linear relationships with maintained interpretability, which is a crucial feature in industrial processes. This advancement will represent transition from parameter optimization to a process prediction system that can improve process understanding of the fundamental relationships governing distillation process.

Acknowledgments

The authors are grateful to the project RE-POLY.AI funded by the European Union - NextGenerationEU under the framework of the Italian National Recovery and Resilience Plan (PNRR) - CUP D43C22003120001. Bando a cascata relativo al progetto “MICS - Made in Italy Circolare e Sostenibile”, PE00000004, CUP B43C22000740006 (D.D. Rep. n. 641/2024, Prot. n.10715/2024 del 18/01/2024 e successive rettifiche ed integrazioni).

References

Brückner, L., Frank, J., Elwert, T., 2020. Industrial Recycling of Lithium-Ion Batteries—A Critical Review of Metallurgical Process Routes. Metals (Basel). 10, 1107. https://doi.org/10.3390/met10081107

Chea, J.D., Christon, A., Pierce, V., Reilly, J.H., Russ, M., Savelski, M., Slater, C.S., Yenkie, K.M., 2019. Framework for Solvent Recovery, Reuse, and Recycling In Industries. pp. 199–204. https://doi.org/10.1016/B978-0-12-818597-1.50032-1

Datta, J., Kopczyńska, P., 2016. From polymer waste to potential main industrial products: Actual state of recycling and recovering. Crit. Rev. Environ. Sci. Technol. 46, 905–946. https://doi.org/10.1080/10643389.2016.1180227

Durrani, M.A., Ahmad, I., Kano, M., Hasebe, S., 2018. An Artificial Intelligence Method for Energy Efficient Operation of Crude Distillation Units under Uncertain Feed Composition. Energies 11, 2993. https://doi.org/10.3390/en11112993

Fang, S., Tu, W., Lu, W., 2024. Artificial intelligence vision technology application in sustainability evaluation of solar-driven distillation device. Environ. Technol. Innov. 36, 103731. https://doi.org/10.1016/j.eti.2024.103731

Hole, G., Hole, A.S., 2019. Recycling as the way to greener production: A mini review. J. Clean. Prod. 212, 910–915. https://doi.org/10.1016/j.jclepro.2018.12.080

Kang, M., Kang, S., 2024. Knowledge distillation with insufficient training data for regression. Eng. Appl. Artif. Intell. 132, 108001. https://doi.org/10.1016/j.engappai.2024.108001

Lei, Z., Li, C., Li, Y., Chen, B., 2004. Separation of acetic acid and water by complex azeotropic distillation. Sep. Purif. Technol. 36, 131–138. https://doi.org/10.1016/S1383-5866(03)00208-9

Maurer, G., 2006. Modeling the liquid–liquid equilibrium for the recovery of carboxylic acids from aqueous solutions. Fluid Phase Equilib. 241, 86–95. https://doi.org/10.1016/j.fluid.2005.11.005

Pirola, C., 2019. Learning distillation by a combined experimental and simulation approach in a three steps laboratory: Vapor pressure, vapor-liquid equilibria and distillation column. Educ. Chem. Eng. 28, 54–65. https://doi.org/10.1016/j.ece.2019.05.003

Triebert, D., Hanel, H., Bundt, M., Wohnig, K., 2021. Solvent-Based Recycling. pp. 33–59. https://doi.org/10.1021/bk-2021-1391.ch003

Wernick, I.K., Themelis, N.J., 1998. RECYCLING METALS FOR THE ENVIRONMENT. Annu. Rev. Energy Environ. 23, 465–497. https://doi.org/10.1146/annurev.energy.23.1.465