Green Chemical Technology

SciEPublish

Article

A Proposed Phenomenological Material–Energy Scaling Framework for Recovery-Constrained Solvent Sustainability

Isak Rajjak Shaikh [1,2,3,*]

[1] School of Chemical Sciences, Swami Ramanand Teerth Marathwada University, Nanded 431606, India
[2] Razaaq University Consortium, Paulsgatan 8 D, 21446 Malmö, Sweden
[3] Razak Institution of Skills, Education & Research (RISER; Reg. HQs at Nanded), [Aḥmad-un-Nisāʾ (ash-Shahīdah) wa ʿAbd ar-Razzāq Shaikh Majmaʿ al-Jāmiʿāt], Pune 411008, India

* Corresponding author. E-mail: isak.rajjak@gmail.com or isak@razaaq.se (I.R.S.)

Received: 7 July 2026; Revised: 8 October 2026; Accepted: 2 November 2026; Available online: 20 November 2026

ABSTRACT: Solvents constitute a major material and environmental burden in pharmaceutical manufacturing, yet sustainability assessment is still largely based on proportional metrics such as Process Mass Intensity (PMI) and aggregated life-cycle indicators. Building on our previously published solvent-centric sustainability framework, which integrated PMI, circular PMI (cPMI), Global Warming Potential (GWP), and Cumulative Energy Demand (CED) for solvent-focused process assessment, the present study develops a phenomenological material–energy scaling formulation for recovery-constrained solvent sustainability. The proposed relationship is formulated as a testable phenomenological scaling framework, with its present scope limited to a system-specific relationship rather than a universal physical law. The framework represents the total sustainability burden, $L(\eta)$, as a combination of a material-driven contribution and a recovery-associated nonlinear contribution: $L(\eta) = c + kM(\eta) + \beta[\eta^n/(1 - \eta)]$, where $M(\eta)$ is solvent material intensity, η is the recovery fraction, c is an intercept, k is a material-impact coefficient, β represents system-specific recovery-associated sensitivity, and n is a phenomenological recovery exponent. A dimensionless form is: $L^* = c^* + \alpha M^* + \lambda[\eta^n/(1 - \eta)]$, where $L^* = L/L_0$, $M^* = M/M_0$, $c^* = c/L_0$, $\alpha = kM_0/L_0$, and $\lambda = \beta/L_0$. The nonlinear coordinate, $\Phi(\eta;n) = \eta^n/(1 - \eta)$, is used as a practical representation of recovery-associated asymptotic behaviour rather than as a first-principles thermodynamic equation. The parameters β, λ, and n are therefore expected to depend on the solvent, process configuration, recovery technology, operating conditions, response being measured, and system boundary. A seven-record pharmaceutical case-study dataset was used to examine the linear material relationship, the proposed nonlinear recovery coordinate, and an extended material–recovery interaction. Across the records, solvent mass decreases from 150 to 45 kg·kg^{-1} API as recovery increases from 30% to 92%, while energy use decreases from 520 to 380 MJ·kg^{-1} API. The linear model explains most of the observed variation (R^2 = 0.979), whereas adding the nonlinear recovery term provides only a small improvement (R^2 = 0.981), is penalized by information criteria, and gives an exponent at its imposed upper limit. An extended interaction model gives a closer fit to the seven observations, but its five-parameter structure relative to only seven observations does not support claims of independent predictive performance. These results therefore provide a basis for treating the proposed formulation as a testable phenomenological scaling hypothesis, rather than as evidence for a universal material–energy law. A complementary step-resolved analysis

© 2026 The authors. This is an open access article under the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

examines how solvent-related material and environmental burdens are distributed across individual process operations and how recovery may shift the balance between material use and recovery-associated energy requirements. This analysis identifies high-contributing operations and defines an Impact Concentration (IC) Index: $IC = \Sigma_i f_i^2$, where $f_i = L_i/L_{total}$ represents the fractional contribution of operation i to a specified sustainability burden. IC is used as a descriptive measure of burden concentration and process localization, rather than as a thermodynamic variable, causal parameter, or universal recovery criterion. Thus, the step-resolved analysis identifies where burdens are concentrated, while the scaling formulation examines how material intensity and recovery-associated burden may be related. Solvent-level descriptor and impact analyses show substantial variation in GWP and CED among the supplied solvents, while exploratory machine-learning analysis indicates the potential for relating molecular descriptors to solvent sustainability. These analyses remain exploratory because of their small sample sizes and differences in data sources and system boundaries. A separate desalination comparison examines whether similar recovery–energy patterns can be observed across other recovery-constrained separation systems. Because membrane, thermal, and solvent-recovery processes differ in their mechanisms, feed characteristics, product requirements, and energy use, this comparison is treated as an empirical comparison of recovery–energy relationships rather than evidence of a common physical mechanism or universal master curve. Finally, a compact cross-domain analysis using 24 observations from independent recovery-related datasets examines whether the proposed recovery coordinate can also provide a useful comparative description beyond pharmaceutical solvent systems. The exponent was not refitted to create a new universal relationship. Furthermore, the broader, forward-looking exergy-based Computational Sustainability Physics (CSP) framework, representing an original conceptual development by the present author, has not included in this study. Overall, the results indicate a system-specific and testable material–energy scaling framework in which increasing solvent recovery can reduce virgin-material demand while changing the relative importance of recovery-associated energy and process-level burden concentration. Further studies using independently documented process data, wider recovery ranges, consistent system boundaries, and independent validation across different solvents and recovery technologies are needed to determine the broader predictive value and transferability of the proposed formulation.

Keywords: Nonlinear scaling framework; Solvent recovery; Material–energy coupling; Circularity; Sustainability assessment; Process intensification; Recovery-constrained sustainability

1. Introduction

1.1. Solvent Dominance and Limitations of Proportional Sustainability Metrics

Solvents constitute the dominant material component in pharmaceutical manufacturing, often exceeding the mass of Active Pharmaceutical Ingredients (APIs) by one to two orders of magnitude [1]. As a result, Process Mass Intensity (PMI) has become a widely adopted metric for evaluating material efficiency and sustainability [2]. PMI provides an important measure of the amount of material required to manufacture a unit mass of product; however, material intensity alone does not capture the environmental consequences of solvent production, recovery, purification, reuse, or the energy required to operate these processes. In this sense, a simple proportional interpretation between material usage and environmental impact can become incomplete when solvent recovery and circularity are introduced.

Industrial solvent management commonly involves recovery and recycling operations, including distillation, membrane separation, extraction, evaporation, and other purification processes [3]. These operations can reduce the requirement for virgin solvent, but they also introduce additional energy demand and process-specific environmental burdens. The relationship between solvent use, recovery efficiency, and

overall environmental performance therefore depends not only on the quantity of solvent circulating through the process but also on solvent properties, separation technology, target purity, process configuration, heat integration, and the energy supply.

1.2. Recovery Constraints and Energy–Material Trade-Offs

From a thermodynamic and separation-engineering perspective, near-complete solvent recovery represents an increasingly constrained separation problem. As recovery efficiency approaches unity, the residual unrecovered fraction becomes progressively smaller, and separation may become increasingly demanding depending on mixture composition, purity requirements, separation mechanism, and operating conditions. Consequently, the energy required for high recovery need not decrease proportionally with the amount of virgin solvent displaced. This behaviour is not proposed here as a newly discovered thermodynamic law; rather, increasing separation difficulty and asymptotic energy requirements near limiting recovery or purity conditions are established features of separation processes [3]. The question addressed in the present study is whether this known physical tendency can be represented within a practical sustainability framework that accounts simultaneously for solvent material intensity and recovery-associated energy demand. The precise form and magnitude of this relationship remain system-specific, depending on solvent properties, mixture characteristics, separation technology, operating conditions, and system boundaries.

1.3. Advances in Solvent-Centric Sustainability and Recovery Technologies

Recent advances in solvent-centric sustainability assessment have sought to integrate mass-based metrics with life-cycle and energy indicators within harmonized system boundaries [4]. Such integrated approaches recognize that solvent use influences not only material intensity but also Cumulative Energy Demand (CED) and Global Warming Potential (GWP), particularly when solvent recovery, reuse, and circularity are explicitly considered [4]. This integration is important because a reduction in virgin solvent demand does not necessarily translate into a proportional reduction in total environmental burden when recovery energy and upstream solvent-specific impacts are included. The foundational solvent-centric sustainability framework developed by Shaikh and Shaikh [4] integrated process mass intensity (PMI), circular PMI (cPMI), GWP, and CED within consistent cradle-to-gate boundaries and demonstrated their application to solvent redesign and recovery through a literature-based Sertraline case study. The present study builds on that foundation by extending integrated sustainability assessment toward a phenomenological description of the relationship between solvent material intensity, recovery, and associated energy and environmental burdens.

Data-driven approaches and cheminformatics tools have further improved predictive solvent selection, solvent-property estimation, and process optimization [5,6]. Machine-learning approaches have, for example, been applied to solvent discovery and solubility prediction, while hybrid modelling has been used to compare chemical separation technologies in terms of energy consumption and carbon dioxide emissions [7]. Recent developments further demonstrate that solvent recovery is strongly dependent on separation technology and process configuration. Organic-solvent Nanofiltration (OSN) has been investigated for catalyst recovery and pharmaceutical purification, while membrane-enabled and continuous-flow reaction–separation strategies provide alternative routes for integrating recovery with processing [8,9]. Recent work on recyclable polymer membranes further illustrates the role of membrane design, molecular structure, and material properties in separation performance and sustainability [10]. These developments demonstrate that solvent recovery is not solely a material-efficiency intervention but involves coupled material, separation, energy, and environmental considerations. However, the underlying relationship between material throughput, recovery efficiency, and environmental or energy impact remains largely empirical rather than

described through a common phenomenological formulation. The present study addresses this gap by extending the earlier solvent-centric sustainability framework toward a phenomenological, system- and technology-dependent description of the relationship between material intensity, recovery, and associated energy and environmental burdens. Rather than assuming a mechanism-independent or universal recovery relationship, the study treats recovery as a process variable whose quantitative behaviour depends on solvent properties, separation technology, process configuration, operating conditions, and the response variable considered. The proposed formulation is therefore intended as a testable, reproducible framework for describing material–energy coupling under recovery constraints, rather than a newly established universal physical law.

1.4. Remaining Modelling Gap

Despite progress in solvent selection, recovery technologies, and process-level sustainability assessment, the relationship between solvent material throughput, recovery efficiency, and system-level environmental burden remains largely addressed through separate metrics or empirical, process-specific relationships. PMI describes material efficiency, life-cycle indicators quantify environmental impacts, and circularity metrics incorporate recovery effects. However, a compact formulation that explicitly examines the potential coupling between material throughput and recovery-associated energy demand remains useful for systematic comparison.

The key question addressed in this study is not whether high recovery can become increasingly demanding—a behaviour already established in separation science—but whether material intensity and recovery efficiency can be represented together through a transparent and testable phenomenological formulation within defined process and sustainability boundaries. To address this methodological gap, the study evaluates linear material scaling, a nonlinear recovery coordinate, and a material–recovery interaction formulation using a primary pharmaceutical case-study dataset, complemented by step-resolved burden localization and exploratory comparisons with other recovery-related systems. The formulations are assessed in terms of explanatory performance, parameter identifiability, model complexity, and transferability. Accordingly, the contribution lies in evaluating alternative phenomenological representations for the analyzed data, not in introducing a new thermodynamic principle. Any broader application remains dependent on data quality, system boundaries, and independent validation.

1.5. Present Study and Model Formulation

At low or moderate recovery, material throughput may provide a dominant contribution to the observed burden, approximated conceptually by:

$$L \propto M \tag{1}$$

where L denotes a specified system-level burden, and M denotes solvent material throughput.

As recovery efficiency increases, recovery-associated separation requirements may introduce an additional dependence on recovery efficiency, such that the burden is no longer adequately represented by a single constant material-impact proportionality. In this regime:

$$L = f(M, \eta) \tag{2}$$

where η is the solvent recovery fraction.

The resulting transition can therefore be interpreted as a shift from a predominantly material-controlled regime toward a regime in which recovery-associated energy and process configuration exert increasing influence. Importantly, the extent of this transition is expected to differ among solvent systems and separation technologies. The purpose of the present framework is therefore not to assume identical

behaviour across all systems, but to provide a mathematically explicit and testable representation of material–energy coupling.

For continuity with this framework, "the candidate phenomenological material–energy scaling relation" is written as:

$$L(\eta) = c + kM(\eta) + \beta[\eta^n/(1 - \eta)] \quad (3)$$

where $L(\eta)$ represents the specified total environmental or sustainability burden associated with solvent recovery at recovery efficiency η, $M(\eta)$ denotes the material-throughput term, c is an intercept, k is the material-impact proportionality coefficient, β is a system-specific coefficient representing the magnitude of the recovery-associated contribution, and $n > 0$ is a fitted nonlinearity exponent controlling the shape of the recovery coordinate.

The nonlinear recovery coordinate is defined as:

$$\Phi(\eta;n) = \eta^n/(1 - \eta) \quad (4)$$

Here, $\Phi(\eta;n)$ is used as a phenomenological coordinate for representing recovery dependence. Its divergence as $\eta \rightarrow 1$ is a mathematical property of the selected functional form and is intended to parameterize the increasing difficulty that may accompany near-complete recovery; it is not a first-principles derivation of a universal energy divergence. The fitted exponent n controls the curvature of the empirical coordinate: $n > 1$ produces stronger curvature toward high recovery, whereas $0 < n < 1$ produces weaker curvature. The physical interpretation of n therefore remains conditional on the response variable, system boundary, solvent–process combination, and quality of the available data.

For normalized analysis, the corresponding "dimensionless formulation" is written as:

$$L^* = c^* + \alpha M^* + \lambda[\eta^n/(1 - \eta)] \quad (5)$$

with

$$L^* = L/L_0$$
$$M^* = M/M_0$$
$$c^* = c/L_0$$
$$\alpha = kM_0/L_0$$
$$\lambda = \beta/L_0$$

Here, the asterisk belongs to the normalized variables and coefficients. L_0 and M_0 are defined reference values, α represents the normalized material contribution, and λ represents the normalized coefficient of the recovery-associated term. Unlike a universal constant, λ is expected to vary with the response variable, solvent system, separation technology, process configuration, energy source, and system boundary. Consequently, β and λ are not assumed to have the same values for different solvents or recovery processes.

The mathematical coordinate therefore provides a common way of representing recovery dependence, whereas its fitted parameters remain system-specific. A shared functional form should consequently not be interpreted as evidence of a universal master curve or identical physical behaviour across chemically and technologically different systems. The present formulation is therefore referred to as the proposed phenomenological material–energy scaling framework. This terminology replaces the earlier "Isak Shaikh Law" designation used in preprints and reflects the intended role of the formulation as a testable, system-specific model.

1.6. Scope and Hypotheses

The central proposition tested in this study is that material throughput and recovery efficiency may be jointly represented through alternative empirical formulations within defined process systems. The analysis therefore tests whether a linear material term is sufficient, whether the nonlinear recovery coordinate

provides additional explanatory value, and whether an explicit material–recovery interaction improves representation of the observed sustainability response. Any coefficients used to parameterize a material–recovery relationship are expected to be system- and response-dependent rather than universal constants. In particular, β and its dimensionless form λ may vary with solvent properties, recovery technology, process configuration, operating conditions, response metric, energy source, and system boundary. The nonlinear exponent n is likewise treated as an estimated model parameter rather than a universal physical constant. The accompanying step-resolved analysis provides a complementary basis for identifying where material and environmental burdens become concentrated within individual process operations, while solvent-level descriptor and life-cycle analyses provide additional context for solvent-specific differences. These analyses are intended to complement, rather than replace, established sustainability metrics such as PMI, life-cycle assessment, GWP, and CED. The broader comparisons included in the study are treated as exploratory transferability and external-check analyses, rather than as proof of a universal relationship. In particular, differences in process mechanisms, response units, system boundaries, and data provenance limit direct quantitative comparison across domains. The analysis therefore does not assume that pharmaceutical solvent recovery, desalination, carbon capture, battery recycling, or metal recovery obey identical physical mechanisms or share identical parameter values [11–25]. Independent rare-earth-element (REE) recovery benchmark dataset [26], which provides an independent assessment of the applicability of the proposed framework beyond the solvent-recovery S1–S3 dataset from Sertraline case-study and cross-domain different S1–S5 benchmark with different chemistries, is used as an external qualitative validation dataset for the constrained-circularity hypothesis. The independent REE recovery benchmark records are provided in Table S6 of the Supplementary Materials. The framework has not yet established a universal master curve, a universal recovery optimum, or a universal physical law; instead, its applicability must be evaluated against the quality and provenance of the underlying data, parameter identifiability, defined system boundaries, and independent validation. For now, the proposed framework is positioned as a testable phenomenological approach for examining material–energy trade-offs under recovery-constrained process conditions, building on established sustainability metrics and separation principles rather than replacing them.

2. Methods

2.1. Framework and System Boundaries

The solvent-centric sustainability framework was developed using step-resolved analysis of individual unit operations, including reaction, work-up, separation, crystallization, and solvent recovery. The framework was designed to examine the relationship between solvent material throughput, recovery efficiency, and associated environmental and energy burdens. Where life-cycle indicators were available, sustainability metrics were considered within cradle-to-gate system boundaries, including upstream solvent production and process-level energy use. The specific system boundary and response variable were retained for each dataset because the solvent-level and process-level datasets do not provide identical inventories or life-cycle metadata. Accordingly, comparisons between datasets were treated as system-specific analyses rather than as evidence of a common universal response function.

The primary process-level analysis used the supplied seven-record case-study/scenario series containing solvent mass intensity, solvent recovery, energy use, VOC emissions, and product yield for annual-labelled records from 2018–2024. The source file does not document a facility, measurement protocol, simulation procedure, or original publication; therefore, these records were treated as supplied case-study/scenario records rather than independently verified industrial measurements or a validated temporal series. A separate nine-solvent dataset containing physicochemical descriptors and solvent-level GWP and CED was analyzed independently. Because a traceable process-specific solvent inventory and

fully compatible life-cycle boundaries were not available for these records, solvent-level impacts were not converted to process-level impacts per kilogram of API.

Data Provenance and Evidential Classification

The datasets used in this study differ in origin, scale, and evidential status and are therefore not treated as interchangeable observations. The seven-record pharmaceutical dataset constitutes the primary case-study dataset used to examine the candidate material–recovery formulations. The solvent-level descriptor and environmental-impact dataset is used for exploratory comparisons between solvent properties and sustainability-related indicators. The desalination dataset comprises computational process scenarios rather than independently measured experimental observations. The 24-observation cross-domain dataset is used as a preliminary transferability check across heterogeneous recovery-related systems. Finally, the step-resolved analysis is used to identify the distribution and concentration of process burdens across individual operations. Each dataset is interpreted within its own system boundary and limitations; consequently, the combined analyses are not presented as a single homogeneous experimental dataset or as independent validation of a universal relationship.

2.2. Decomposition of Environmental Impact

The total environmental impact is expressed conceptually as the sum of material-driven and recovery-associated contributions:

$$L(\eta) = L_m(\eta) + L_e(\eta) \tag{6}$$

where $L_m(\eta)$ denotes the material-driven contribution and $L_e(\eta)$ denotes the recovery-associated contribution at recovery efficiency η.

The material contribution is represented as:

$$L_m(\eta) = kM(\eta) \tag{7}$$

where $M(\eta)$ denotes the solvent material-throughput term and k is the material-impact proportionality coefficient.

The formulation recognizes that increasing solvent recovery can reduce virgin solvent demand and therefore reduce the material-throughput contribution. In contrast, the recovery-associated contribution may increase with increasing recovery requirements because separation and purification can become progressively more demanding at high recovery or purity. The precise magnitude of this contribution depends on solvent properties, mixture composition, separation technology, target purity, process configuration, heat integration, and energy supply.

For phenomenological representation, the recovery-associated term is written as:

$$L_e(\eta) = \beta\Phi(\eta;n) \tag{8}$$

where β is a system-specific coefficient and $\Phi(\eta;n)$ is the candidate nonlinear recovery coordinate defined as:

$$\Phi(\eta;n) = \eta^n/(1 - \eta) \tag{9}$$

Combining the material and recovery-associated contributions yields the candidate scaling relation:

$$L(\eta) = c + kM(\eta) + \beta[\eta^n/(1 - \eta)] \tag{10}$$

This formulation is intended as a phenomenological scaling representation of recovery-associated energetic amplification rather than a first-principles derivation of separation thermodynamics. The asymptotic behaviour of the coordinate as η approaches unity provides a mathematical representation of increasing separation difficulty near limiting recovery; it is not, by itself, evidence that the fitted environmental burden must physically diverge at $\eta = 1$.

2.3. Metrics and Normalization

Material intensity is represented primarily by solvent mass intensity, expressed as:

$$M = \text{kg solvent} \cdot \text{kg}^{-1}\ \text{API}$$

where appropriate, Process Mass Intensity (PMI) provides the broader material-efficiency context. Where circularity is explicitly evaluated, cPMI may be operationalized; this is our operationalized framework metric, not a universally standardized sustainability metric, which is defined as: cPMI = PMI (1 − R), where R is the solvent recovery fraction. In the present model comparison, however, the primary response variable is the supplied solvent mass intensity M and the corresponding energy use E; cPMI is not substituted for M unless explicitly stated for a separate circularity analysis.

Environmental impact is represented, where available, using life-cycle indicators including Global Warming Potential (GWP) and Cumulative Energy Demand (CED). For the primary seven-record regression analysis, energy use E ($MJ \cdot kg^{-1}$ API) is used as the response variable because it is the directly available process-level energy measure.

Normalized variables are defined as:

$$M^* = M/M_0 \tag{11}$$

$$L^* = L/L_0 \tag{12}$$

where M_0 and L_0 are defined reference values.

Normalization is used to facilitate comparison of material and environmental terms across a defined system. It does not imply that different solvents, process configurations, or environmental indicators share identical scaling coefficients.

2.4. Scaling Behaviour and Residual Analysis

At low or moderate recovery, material throughput may provide a dominant contribution to the observed burden and can be represented conceptually by:

$$L \propto M \tag{13}$$

At higher recovery, recovery-associated process requirements may introduce an additional dependence on η such that the material-only proportionality is no longer sufficient:

$$L = f(M,\eta) \tag{14}$$

Accordingly, the derivative

$$dL/dM$$

need not remain constant when recovery efficiency, process configuration, or other coupled variables change.

To examine deviation from a simple material baseline, a normalized residual can be defined as:

$$\Delta L^* = L^* - c^* - \alpha M^* \tag{15}$$

$$\Delta L^* = \lambda \Phi(\eta;n)$$

where α is the normalized material contribution defined below.

This residual should be interpreted as the portion of the normalized response not represented by the material-throughput term under the specified model, rather than being assumed a priori to represent only energy. Other unmeasured process effects may also contribute to the residual.

2.5. Phenomenological Formulation and Dimensionless Representation

The observed material–recovery scaling behaviour is represented by the proposed nonlinear material–energy scaling formulation as a phenomenological candidate model:

$$L(\eta) = c + kM(\eta) + \beta\Phi(\eta;n) \quad (16)$$

with

$$\Phi(\eta;n) = \eta^n/(1 - \eta) \quad (17)$$

For normalized analysis, the corresponding dimensionless formulation is:

$$L^* = c^* + \alpha M^* + \lambda\Phi(\eta;n) \quad (18)$$

where

$$L^* = L/L_0$$
$$M^* = M/M_0$$
$$c^* = c/L_0$$
$$\alpha = kM_0/L_0$$
$$\lambda = \beta/L_0$$

Here, β is the dimensional coefficient associated with the recovery-dependent contribution, whereas λ is its dimensionless counterpart. Both parameters are specific to the selected response variable, normalization, solvent system, process configuration, separation technology, and system boundary. The use of a common functional coordinate does not imply universal parameter values or a universal physical response. A consolidated summary of the equations, variables, model parameters, and framework-specific quantities is provided in Table S7 of the Supplementary Materials.

2.6. Coupling Parameter (λ)

The coupling parameter λ characterizes the normalized contribution of the recovery-associated term within the specified model:

$$\lambda = \beta/L_0 \quad (19)$$

where β is the coefficient of the recovery-dependent term, and L_0 is the reference response value.

Under the normalized formulation:

$$\Delta L^* = \lambda\Phi(\eta;n) \quad (20)$$

only when ΔL^* is defined specifically as the residual after subtracting the normalized material contribution αM^*. Thus, λ represents the coefficient linking the defined residual response to the candidate recovery coordinate $\Phi(\eta;n)$.

The parameter should therefore be interpreted as a system-specific empirical coupling coefficient rather than a universal physical constant. Its estimated value can differ between solvents, environmental indicators, separation technologies, process configurations, and system boundaries.

The present analysis does not use simple ratios such as:

$$\lambda \approx \Delta GWP/\Delta PMI$$

or

$$\lambda \approx \Delta CED/\Delta PMI$$

as definitions of λ, because such ratios do not in general provide the normalized coefficient of Equation (16) and may not be dimensionless without explicit normalization.

2.7. Coupling Function and Scaling Representation

To represent the candidate recovery dependence in normalized form, the coupling function is defined as:

$$\Phi(\eta;n) = \eta^n/(1 - \eta) \quad (21)$$

The residual scaling relation is then:

$$\Delta L^* = \lambda\Phi(\eta;n) \quad (22)$$

when the residual is defined according to Equation (15) $\Delta L^* = L^* - c^* - \alpha M^*$, and the response follows the candidate model.

This formulation provides a convenient coordinate for testing whether recovery-associated responses exhibit systematic curvature beyond a material-throughput baseline. It does not by itself establish that solvent-system data collapse onto a universal master curve. Any apparent collapse must be evaluated against solvent-specific parameters, process-family differences, data provenance, system boundaries, and independent validation.

2.8. Impact Concentration (IC) Index and Step-Resolved Burden Localization

To quantify the concentration of sustainability burden among individual process operations, an Impact Concentration Index (IC) was defined as:

$$IC = \Sigma_i f_i^2$$

where $f_i = L_i/L_{total}$, L_i is the contribution of unit operation i to the selected impact metric, and L_{total} is the total process-level impact. Thus, $\Sigma_i f_i = 1$ and $0 < IC \leq 1$. Higher IC values indicate that the selected burden is concentrated in a smaller number of unit operations, whereas lower values indicate a more distributed contribution.

Here, IC is used as a descriptive, process-localization metric rather than a thermodynamic state variable, causal parameter, or universal recovery criterion. Its value is necessarily dependent on the selected impact metric, system boundary, and level of process disaggregation. The step-resolved analysis was therefore used to identify the major contributing operations, while IC provides a complementary quantitative representation of the degree of burden concentration when a complete operation-level contribution matrix is available. Because the primary seven-record dataset does not contain a reproducible operation-level contribution matrix for every observation, no empirical IC–recovery relationship was fitted, and no universal IC threshold or recovery optimum was inferred. IC is instead retained as a defined methodological quantity for process-level burden localization and for future applications in which operation-resolved data are available.

2.9. Candidate Optimum Recovery Condition

A stationary point of the fitted response with respect to recovery efficiency can be defined by:

$$dL/d\eta = 0 \quad (23)$$

This condition represents a mathematical balance between the marginal material contribution and the marginal recovery-associated contribution. It identifies a candidate stationary point rather than establishing an observed or universal optimum.

For the purpose of deriving an analytical stationary condition, solvent material intensity is assumed to decrease proportionally with recovery efficiency:

$$M(\eta) = M_0(1 - \eta) \quad (24)$$

The candidate model becomes the proposed phenomenological formulation:

$$L(\eta) = c + kM_0(1 - \eta) + \beta[\eta^n/(1 - \eta)] \quad (25)$$

where c is the intercept, kM_0 represents the material-related contribution, and $\beta[\eta^n/(1 - \eta)]$ represents the recovery-associated contribution.

Differentiating with respect to η gives:

$$dL/d\eta = -kM_0 + \beta\{n\eta^{n-1}(1 - \eta) + \eta^n\}/(1 - \eta)^2 \quad (26)$$

Equivalently:

$$dL/d\eta = -kM_0 + \beta\eta^{n-1}[n + (1 - n)\eta]/(1 - \eta)^2$$

The intercept does not appear in the derivative because $dc/d\eta = 0$. A candidate stationary recovery condition is therefore obtained by setting:

$$-kM_0 + \beta\eta^{n-1}[n + (1 - n)\eta]/(1 - \eta)^2 = 0 \quad (27)$$

A solution within the feasible operating range identifies a candidate stationary point for the specified parameter set. It should be interpreted as a minimum only if the second-derivative condition and relevant process constraints are also satisfied.

This derivation is conditional on the assumed relationship $M(\eta) = M_0(1 - \eta)$, which is not established as a universal solvent-recovery relationship. Furthermore, the primary seven-record dataset does not independently identify recovery-associated energy penalties from material throughput. Therefore, the present study does not claim an observed, intrinsic, or universal optimal recovery efficiency. A candidate optimum is meaningful only when supported by system-specific data, uncertainty analysis, and feasible process constraints.

2.10. Statistical Model Comparison and Parameter Estimation

To evaluate the candidate scaling relation, model comparison was performed using the same seven primary case-study/scenario records for each competing regression specification. The response variable was process energy use, E ($MJ \cdot kg^{-1}$ API), with solvent mass intensity, M, as the principal material predictor and recovery fraction, η, as the recovery variable.

Three model forms were evaluated.

(i) Linear material-throughput model

The linear material-throughput model was expressed as:

$$E_i = c + kM_i + \varepsilon_i \quad (28)$$

where c is the intercept, k is the material-throughput coefficient, and ε_i is the residual error.

(ii) Candidate nonlinear recovery model

The candidate nonlinear recovery model was expressed as:

$$E_i = c + kM_i + \beta\Phi(\eta_i;n) + \varepsilon_i \quad (29)$$

$$\text{where } \Phi(\eta_i;n) = \eta_i^n/(1 - \eta_i) \quad (30)$$

and n is the fitted nonlinearity exponent.

(iii) Extended material–recovery interaction model

$$E_i = c + kM_i + \beta\Phi(\eta_i;n) + \gamma M_i\eta_i + \varepsilon_i \quad (31)$$

where γ represents an empirical material–recovery interaction coefficient.

The interaction term γMη is introduced phenomenologically to capture higher-order coupling between solvent throughput and recovery intensity. It is not interpreted as a first-principles thermodynamic term.

Nonlinear least-squares fitting was performed with n bounded between 0.01 and 10. The parameter bounds were predefined for the regression analysis. A fitted solution reaching a boundary was interpreted as evidence of weak parameter identification rather than confirmation of the candidate functional form.

Model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC), using a common Gaussian least-squares likelihood for model comparison. AIC and BIC were used to account for differences in model complexity. Because the primary analysis contains only seven observations, none of the models is treated as independently predictive. In particular, the extended interaction model contains a high number of regression parameters relative to the number of observations and is therefore considered susceptible to overfitting. Its close in-sample fit is consequently interpreted as an empirical approximation rather than evidence of a validated physical relationship.

2.11. Solvent-Level Descriptor and Machine-Learning Analysis

A separate exploratory solvent-level analysis was conducted using nine solvents with matched physicochemical descriptors and GWP values. Four solvent descriptors—dipolarity index, hydrogen-bond capacity index, viscosity, and biodegradability score—were used as predictors in a Random Forest Regressor model. The model used 500 trees and random_state = 42. Leave-one-out cross-validation (LOOCV) was performed across the nine solvent observations, and predictive performance was evaluated using held-out R^2, RMSE, and MAE. Because only nine observations were available, this analysis was treated as exploratory and was not used as validation of the primary recovery–energy scaling model. Descriptor–impact correlations were likewise treated as descriptive associations rather than causal relationships or general solvent-selection rules.

The desalination comparison was based on a supplied scenario dataset comprising 1316 process-level records generated within the accompanying desalination analysis package. The records represent modelled desalination process scenarios rather than 1316 independently measured experimental observations [25]. Each record contains process-recovery and specific-energy information used to examine how recovery–energy relationships vary across desalination configurations. The dataset includes multiple process configurations and therefore contains substantial within-family and between-family structure. A subset of 1244 records corresponds to the zero-liquid-discharge (ZLD) and minimal-liquid-discharge (MLD) scenarios used for the high-recovery analysis. For process-family comparison, the supplied dataset was also reduced to one maximum-recovery record per process-train label, giving 100 representative process-train records. These reductions were performed to examine process-family structure and were not treated as independent external validation. Because the supplied scenario files do not provide a complete experimental provenance for every individual record, including original facility measurements, measurement protocols, or independently reported experimental uncertainty, the desalination dataset is treated here as a computational scenario dataset. It is therefore used strictly as an empirical transferability and process-heterogeneity assessment. The analysis does not assume that desalination and pharmaceutical solvent recovery share identical separation mechanisms, thermodynamic boundaries or energy definitions.

2.12. Desalination Comparison as an Empirical Transferability Analysis

A separate desalination dataset was analyzed to examine whether recovery-associated energy relationships observed in another constrained-recovery process family could be represented using comparable empirical coordinates.

Recovery was converted to a fractional value:

$$\eta = \text{recovery (\%)}/100$$

Specific energy consumption (SEC) remained in $kWh \cdot m^{-3}$ of produced water.

Pooled linear, quadratic, cubic, and candidate nonlinear-coordinate models were compared using the same desalination records. Family-aware models were also evaluated to account for differences among process configurations. Five-fold grouped cross-validation was performed by holding out complete process-train labels. This analysis was designed as an empirical comparison of recovery–energy relationships rather than as validation of a common thermodynamic mechanism. Membrane desalination, thermal desalination, and organic-solvent recovery differ in feed composition, separation driving forces, product specifications, process configuration, and energy integration. Accordingly, any similarity in empirical scaling behaviour is interpreted cautiously and does not establish a universal master curve.

2.13. Scope of the Framework and Reproducibility of the Analysis

The present framework distinguishes between a mathematical scaling coordinate and the empirical evidence required to support its parameters. The candidate coordinate $\Phi(\eta;n)$ provides a testable representation of recovery dependence, while k, β, n, α, and λ remain system- and response-dependent quantities. The primary seven-record dataset is therefore used to compare competing model forms rather than to establish universal parameter values. The nine-solvent analysis provides solvent-specific environmental context, while the desalination analysis provides a separate process-family comparison. These analyses are complementary and are not treated as interchangeable observations within a single universal regression. All equations, variables, parameter definitions, model assumptions, and system boundaries are specified so that subsequent studies can reproduce the analysis using independently documented datasets. Stronger claims concerning transferability, predictive validity, universal parameter values, or an optimum recovery efficiency require independent process datasets with documented provenance, comparable system boundaries, independently measured recovery-associated energy requirements, and held-out validation.

3. Results

3.1. Predictive Solvent-Sustainability Modelling and Molecular Descriptor Integration

The overall process-level trends across the seven case-study records are shown in Figure 1, highlighting the concurrent changes in solvent mass intensity, recovery, energy use, VOC emissions, and product yield. The relationship between solvent mass intensity and energy use is further examined in Figure 2, which illustrates the strong material–energy association while also indicating that changes in material intensity alone do not fully capture the observed sustainability response. Figure 3 extends the analysis to the solvent level by comparing the GWP and CED profiles of the nine investigated solvents, providing a molecular- and solvent-specific context for the subsequent cheminformatics and machine-learning analysis (see the supplementary solvent-descriptor and environmental-impact diagnostics in Figure S3 and the underlying nine-solvent dataset in Table S3). Figure 4 summarizes the exploratory cheminformatics–machine-learning analysis used to examine whether solvent physicochemical descriptors can represent solvent-related environmental impacts.

Figure 4 summarizes the exploratory cheminformatics–machine-learning analysis used to examine whether solvent physicochemical descriptors can represent solvent-related environmental impacts. In contrast to the apparent agreement reported in the initial analysis, the revised evaluation uses leave-one-out cross-validation (LOOCV) and gives an R^2 of 0.483, with RMSE = 1.295 and MAE = 1.112. These values indicate moderate explanatory capability but do not support strong predictive generalization from the small solvent-level dataset. The descriptor analysis nevertheless identifies systematic relationships between solvent properties and sustainability-related outcomes. Polarity, hydrogen-bonding characteristics, viscosity-related behaviour, and biodegradability-related descriptors contribute to the differentiation of solvent systems. These results are interpreted as exploratory evidence that molecular-scale solvent

properties can provide useful information for solvent selection and sustainability assessment, rather than as evidence of a universal molecular-to-process scaling relationship. The machine-learning component is therefore treated as a supporting decision and interpretation layer rather than as an independent validation of the proposed material–energy scaling formulation. The corresponding leave-one-out cross-validation diagnostics are provided in Figure S4 of the Supplementary Materials.

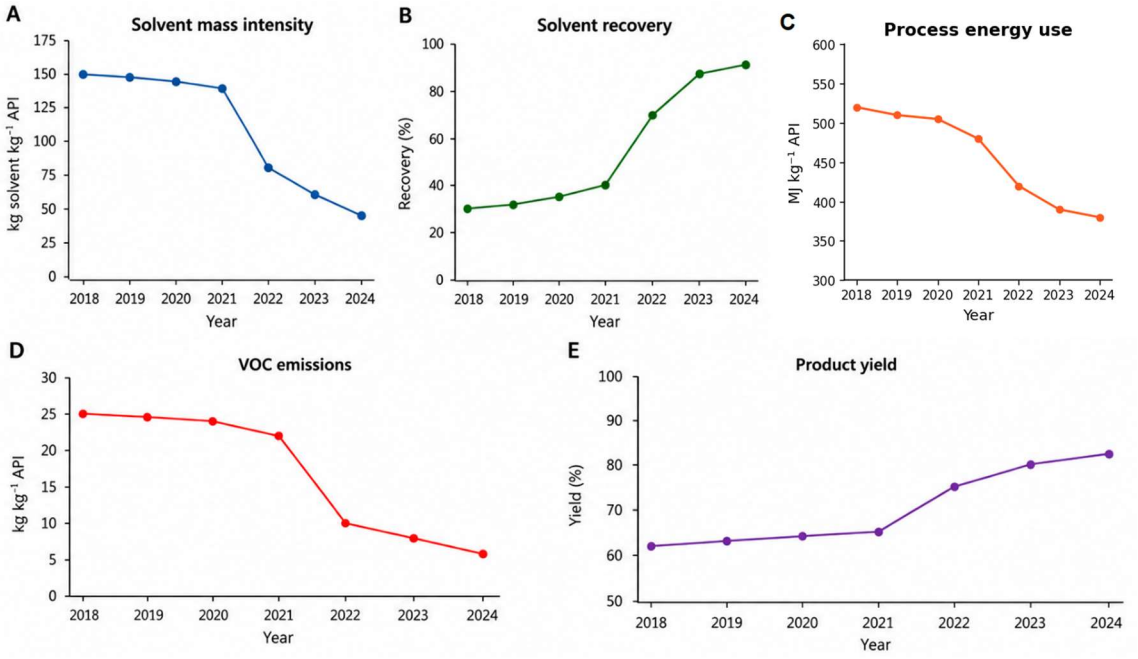


Figure 1. Indicators in the seven supplied annual-labelled case-study records. (**A**) Solvent mass intensity, (**B**) solvent recovery, (**C**) process energy use, (**D**) VOC emissions, and (**E**) product yield for the seven annual-labelled records supplied in Dataset_S1.csv (2018–2024). The records are normalized, literature-derived case-study/scenario data used for methodological demonstration. The observed changes are descriptive and are not interpreted as independently verified temporal industrial measurements or proprietary manufacturing data.

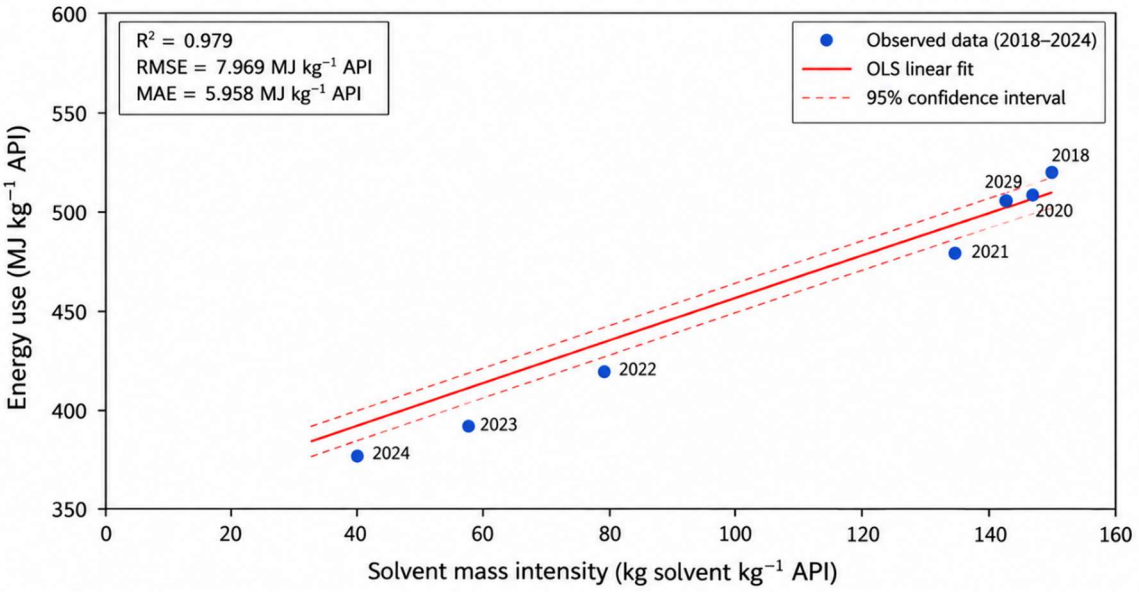


Figure 2. Energy use versus solvent mass for the seven supplied case-study records. Energy use (MJ·kg^{-1} API) is plotted against solvent mass intensity (kg·solvent·kg^{-1} API), with an ordinary least-squares linear fit. The in-sample association is $R^2 = 0.979$, with RMSE = 7.969 MJ·kg^{-1} API and MAE = 5.958 MJ·kg^{-1} API. Because the seven observations are few and solvent mass, recovery, and energy change simultaneously, the association is descriptive and does not establish causality or a universal material–energy relationship.

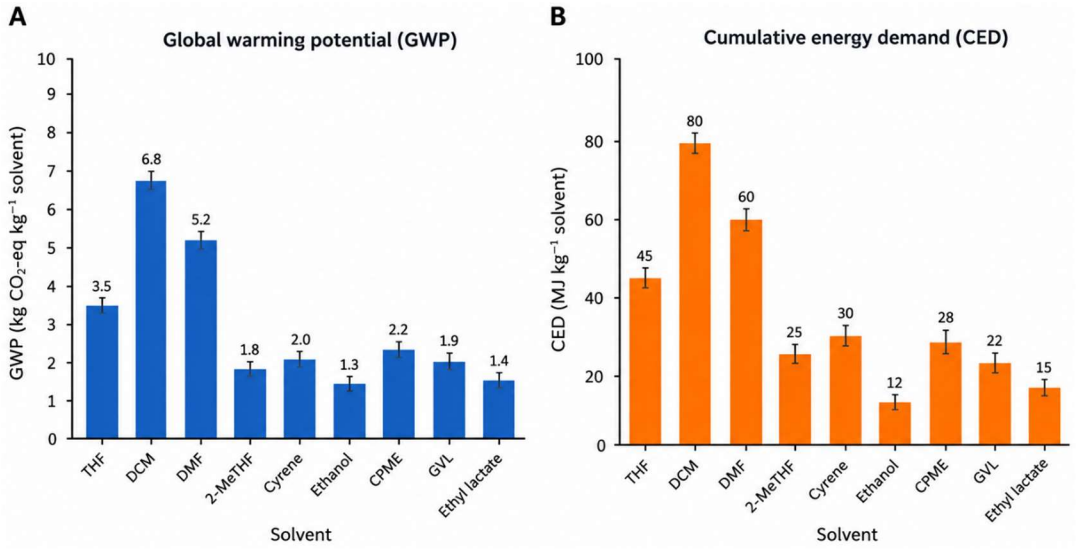


Figure 3. Solvent-level global warming potential and cumulative energy demand. (**A**) Global warming potential (GWP); (**B**) cumulative energy demand (CED) for the nine solvents in the supplied solvent-impact dataset, expressed on a per-kilogram-solvent basis. The figure provides environmental-impact context for the solvent-selection analysis. Values are not converted to per-API impacts because a traceable, process-specific inventory and common system boundary are not available for all solvents.

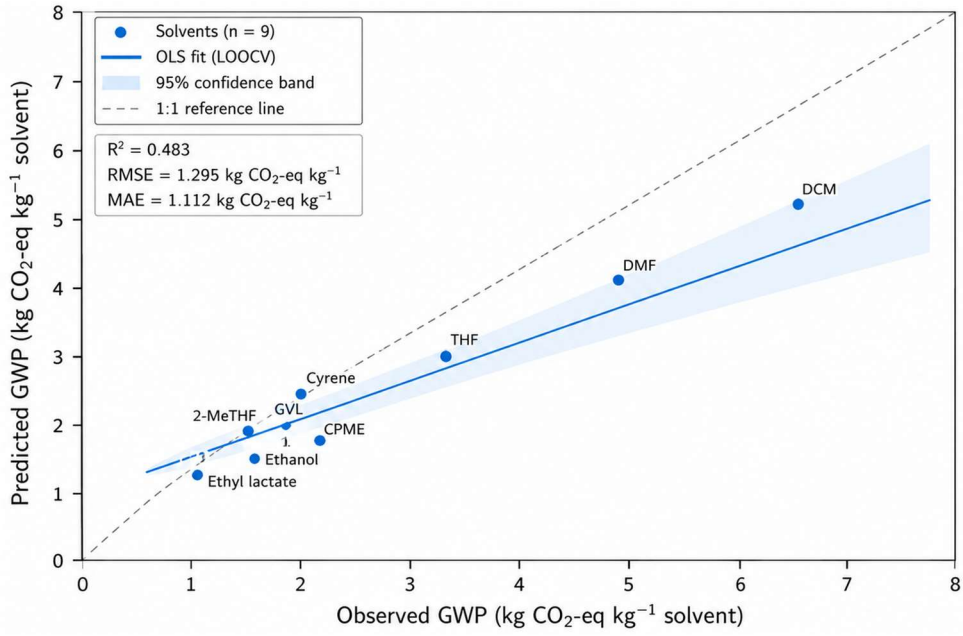


Figure 4. Observed versus leave-one-out predicted GWP for nine solvents. Observed GWP values are compared with predictions from the exploratory Random Forest model using four predefined solvent descriptors: dipolarity index, hydrogen-bond capacity, viscosity, and biodegradability. Leave-one-out cross-validation produced R^2 = 0.483, RMSE = 1.295 kg CO_2-eq·kg^{-1} solvent and MAE = 1.112 kg CO_2-eq·kg^{-1} solvent. The nine-observation analysis is a proof-of-concept demonstration rather than a validated or universal predictive model.

3.2. Process-Level Origin of Non-Proportional Sustainability Behaviour

The Step-Resolved analysis provides a process-level view of how solvent-related burdens are distributed across individual operations. Within the analyzed process configuration, a limited number of unit operations account for a substantial fraction of the total solvent-related burden. In particular, the two highest-contributing operations account for approximately 65–75% of PMI and approximately 60% of GWP within the analyzed case-study configuration. This is presented as burden localization, not as proof of a universal recovery law. This concentration is important because aggregate PMI, GWP, and CED values

alone do not identify where the burden originates. The Step-Resolved analysis indicates that solvent-intensive reaction, work-up, separation, and recovery operations do not contribute equally. Instead, a relatively small subset of operations can dominate the material and environmental burden. As recovery efficiency increases, the contribution of solvent separation and recovery operations becomes increasingly important to the overall sustainability balance. This observation provides the process-level basis for examining material–energy coupling, but it does not by itself establish a universal recovery threshold or a universal energetic penalty. The supplementary process-level illustration of burden localization and the methodological basis of the Impact Concentration (IC) Index are provided in Figure S6.

3.3. Primary Pharmaceutical Case-Study Dataset and Breakdown of Proportional Scaling

The primary dataset contains seven normalized case-study/scenario records spanning 2018–2024. The records describe solvent mass intensity, solvent recovery efficiency, energy demand, VOC emissions, and product yield. The dataset is normalized case-study/scenario records compiled from the supplied dataset and used for methodological comparison/demonstration and should not be interpreted as proprietary manufacturing measurements from an individual industrial facility. Across the seven records, solvent mass intensity decreases from 150 to 45 kg·kg^{-1} API, while solvent recovery increases from 30% to 92%. Energy demand decreases from 520 to 380 MJ·kg^{-1} API over the same range, while VOC emissions decrease from 25 to 6 kg·kg^{-1} API. Thus, the material reduction is substantially larger than the corresponding reduction in energy demand over the observed range. The underlying seven-record dataset, together with the corresponding derived cPMI values, is summarized in Table 1 (see also Figure S1 and Table S1 in the Supplementary Materials), while Table 2 presents the parameter estimates and comparative performance metrics of the three candidate formulations.

Table 1. Primary Seven-Record Solvent-Recovery Dataset * and Derived Circularity Metric.

Year	Solvent Mass Intensity, M (kg·Solvent·kg^{-1} API)	Solvent Recovery, η (%)	Energy Use (MJ·kg^{-1} API)	VOC Emissions (kg·kg^{-1} API)	Yield (%)	cPMI (kg·kg^{-1} API)
2018	150	30	520	25.0	62	105.00
2019	148	32	510	24.5	63	100.64
2020	145	35	505	24.0	64	94.25
2021	140	40	480	22.0	65	84.00
2022	80	70	420	10.0	75	24.00
2023	60	88	390	8.0	80	7.20
2024	45	92	380	6.0	82	3.60

* Seven primary pharmaceutical case-study/scenario records.

The table above reports the seven annual-labelled records supplied in Dataset_S1.csv, comprising solvent mass intensity, solvent recovery, process energy use, VOC emissions, and product yield. cPMI was calculated as cPMI = PMI × (1 − R), where R is the recovery fraction. The records constitute a normalized, literature-derived case-study/scenario dataset used for methodological demonstration; they are not presented as independently verified annual industrial measurements or proprietary Pfizer Inc., New York (USA) manufacturing data.

cPMI was calculated as cPMI = PMI × (1 − R), where R is the recovery fraction. In the primary regression analysis, however, the response variable is energy use and the principal material predictor is solvent mass intensity; cPMI is therefore provided as a complementary circularity indicator rather than substituted for M.

The corresponding normalized values show that PMI decreases to 0.30 of its 2018 reference value, whereas normalized energy demand remains at approximately 0.73. This difference demonstrates a decoupling between material-intensity improvement and energy-demand improvement. Importantly,

however, the seven-record dataset does not show an increase in absolute energy demand as recovery increases. Therefore, the observed data support material–energy decoupling but do not independently establish an energetic divergence near $\eta = 1$.

The relationship between environmental burden (L) and material intensity (M) was therefore evaluated first using a proportional baseline:

$$L \propto M$$

The relationship between energy use and solvent mass intensity for the seven case-study records is shown in Figure 2. The strong linear association confirms that material intensity remains an important determinant of the observed sustainability burden. However, the difference between the reduction in PMI and the smaller reduction in energy demand indicates that material metrics alone do not capture the complete process-level sustainability response. The corresponding changes in solvent mass intensity, recovery, energy use, VOC emissions, and product yield are shown in Figure 1.

3.4. Residual Analysis and Candidate Nonlinear Recovery Coordinate

To examine deviations from proportional material scaling, the normalized residual contribution was evaluated as:

$$\Delta L^* = L^* - c^* - \alpha M^*$$

where $c^* + \alpha M^*$ represents the normalized baseline contribution defined in the Methods.

The residual analysis was used diagnostically to determine whether recovery efficiency contains information beyond material intensity. The results indicate that the residual component is not fully represented by a simple proportional PMI relationship. However, the seven-record dataset does not provide sufficient evidence for interpreting this residual as a universal asymptotic energy penalty.

The proposed nonlinear recovery coordinate,

$$\Phi(\eta;n) = \eta^n/(1 - \eta)$$

was therefore retained as a candidate phenomenological coordinate. Its functional form is motivated by the known difficulty of achieving increasingly complete separation and recovery as the unrecovered fraction becomes small. The coordinate should not be interpreted as a newly discovered thermodynamic divergence. Rather, it provides a compact parameterization that can be tested against recovery-dependent data. This distinction is important because asymptotic energetic behaviour near complete recovery is already recognized in separation engineering and thermodynamic analyses. The contribution of the present framework is to test whether such recovery dependence can be incorporated into material-intensity accounting through a common phenomenological coordinate.

3.5. Evaluation of the Proposed Phenomenological Material–Energy Scaling Framework and Its Formulation

The proposed phenomenological material–energy scaling formulation is expressed in the present study as the phenomenological relationship:

$$L(\eta) = c + kM(\eta) + \beta[\eta^n/(1 - \eta)]$$

The first term represents the material-associated contribution and the second term represents a candidate nonlinear recovery-associated contribution.

The seven-record data provide a strong baseline fit for the linear material model ($R^2 = 0.979$). The addition of the nonlinear recovery coordinate produces only a marginal increase in R^2 to 0.981. Information criteria do not favour the nonlinear model over the simpler linear formulation, and the fitted exponent n reaches the imposed upper bound of the parameter range. The boundary solution indicates weak

identification of n within the available seven observations rather than empirical confirmation of a particular asymptotic exponent. The mathematical sensitivity of the recovery coordinate to alternative values of the exponent n is illustrated in Figure S8 of the Supplementary Materials. These curves represent mathematical sensitivity rather than fitted universal relationships or experimental energy trajectories. Accordingly, the present results do not establish that the nonlinear term is required to explain the primary pharmaceutical dataset. Instead, they demonstrate that the proposed coordinate is mathematically testable and can be compared directly with a material-only baseline. The coefficient β should likewise be interpreted as a system-specific parameter. Its magnitude depends on the response variable, process configuration, system boundary, and normalization used. It is therefore not expected to have a universal value across solvents or application domains. Additional observed-versus-fitted and residual diagnostics are provided in Figure S2 and the corresponding fitted coefficients and model-comparison statistics in Table S2 of the Supplementary Materials.

3.6. Dimensionless Representation and System-Specific Coupling

The proposed relationship can be expressed in dimensionless form as:

$$L^* = c^* + \alpha M^* + \lambda \Phi(\eta;n)$$

where:

$$L^* = L/L_0$$

$$M^* = M/M_0$$

$$\alpha = kM_0/L_0$$

$$\lambda = \beta/L_0$$

and

$$\Phi(\eta;n) = \eta^n/(1 - \eta)$$

This formulation provides a common mathematical coordinate for comparing recovery-constrained systems while retaining system-specific coefficients. In particular, λ is not treated as a universal constant. Its value depends on the response variable and system under investigation. The normalized representation therefore permits comparison of the relative contribution of material intensity and recovery dependence without assuming that different solvents, process configurations, or environmental indicators have identical scaling parameters. The revised interpretation differs from the earlier master-curve formulation. Normalization provides a common coordinate system, but the available data do not demonstrate that all solvent systems collapse onto a single universal trajectory.

3.7. Material-Dominated and Recovery-Influenced Regimes

The results support a qualitative distinction between material-dominated and increasingly recovery-influenced operating conditions. At relatively low recovery, the material-associated term provides a substantial fraction of the total response, and the relationship between material intensity and environmental burden is approximately proportional:

$$L \approx kM$$

As recovery increases, material intensity decreases substantially, whereas the corresponding improvement in the selected energy/environmental response is smaller. The relative importance of recovery-associated operations therefore increases. This transition is interpreted as a change in the relative contribution of material and recovery terms, rather than as evidence for universal numerical recovery thresholds. The previously stated

thresholds of $\eta < 0.6$, $0.6 \leq \eta \leq 0.8$, and $\eta > 0.8$ are therefore not retained as general regime boundaries because the present dataset is insufficient to establish them as transferable constants.

Step-Resolved Burden Localization

The step-resolved analysis provides an important process-level complement to the aggregate material–energy comparison. In the analyzed solvent-intensive configuration, the top two unit operations account for approximately 65–75% of total PMI and approximately 60% of GWP, indicating that sustainability burdens are structurally concentrated rather than uniformly distributed across the synthetic sequence. The relative importance of individual operations is also metric-dependent: reaction dilution and work-up contribute strongly to material intensity, whereas downstream recovery, crystallization, and associated energy-intensive operations can contribute disproportionately to energy and climate indicators. These observations motivate the use of an Impact Concentration (IC) Index as a descriptive localization metric, but they do not establish a universal threshold or causal relationship between impact concentration and recovery efficiency.

The IC provides a complementary way of quantifying this concentration when operation-level contribution data are available. In the present study, however, the aggregate seven-record dataset does not contain a complete and reproducible operation-level contribution matrix for all observations. Accordingly, IC was not fitted against recovery, and no empirical IC–recovery relationship or universal IC threshold was claimed. The directly supported result is the concentration of PMI and GWP within the major contributing operations identified by the step-resolved analysis.

3.8. Impact Concentration and Localization of Environmental Burden

The Impact Concentration (IC) Index was used to quantify the structural concentration of environmental burden across process operations:

$$IC = \Sigma_i \, f_i^2$$

where

$$f_i = L_i / L_{total}$$

The index provides a compact measure of whether environmental burden is distributed relatively evenly across process operations or concentrated within a smaller number of operations. Higher IC values indicate greater concentration of the total burden within fewer operations. In the present study, IC is used primarily as a localization metric for the Step-Resolved analysis. The available operation-level data do not justify claiming a monotonic increase in IC with recovery efficiency across the full primary dataset. Accordingly, no universal IC–η relationship is inferred. The Step-Resolved results nevertheless show that solvent-intensive processes can become increasingly sensitive to a limited number of high-contribution operations. This provides a process-level complement to the aggregate material–energy analysis and identifies potential high-leverage operations for process improvement. The corresponding operation-level burden distribution and IC methodology are illustrated in Figure S6 of the Supplementary Materials.

The desalination analysis comprised 1316 supplied scenario records rather than 1316 independent experimental observations. The dataset contains multiple process configurations, including ZLD/MLD scenarios, and therefore preserves substantial process-family structure. A 1244-record ZLD/MLD subset and a 100-record maximum-recovery-per-process-train representation were used for complementary high-recovery and family-aware analyses. The results are consequently interpreted as an empirical assessment of process-family-dependent recovery–energy relationships rather than as independent validation of the proposed phenomenological material–energy scaling framework. This is presented as an empirical comparison of recovery–energy relationships rather than validation of a common thermodynamic mechanism.

The corresponding desalination recovery–specific-energy comparisons across the full, high-recovery, and process-train datasets are shown in Figure 5 (see the supplementary process-family diagnostics in Figure S5). The step-resolved distribution of material and environmental burden across the investigated solvent-recovery process is shown in Figure 6. The Pearson and Spearman correlations between solvent descriptors and environmental-impact indicators are summarized in Table 3. The quantitative comparison of the alternative desalination recovery–specific-energy formulations and dataset subsets is summarized in Table 4 (see the supplementary desalination model diagnostics in Table S4).

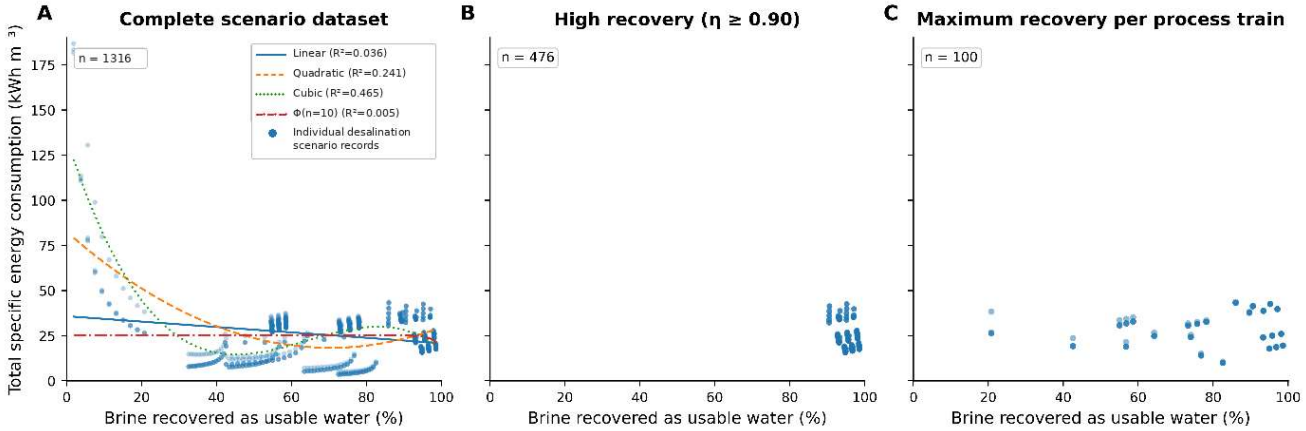


Figure 5. Desalination recovery–specific-energy records stratified by process configuration. (**A**) Complete supplied desalination scenario dataset (n = 1316), (**B**) high-recovery subset (n = 476; η ≥ 0.90), and (**C**) one maximum-recovery record per process-train label (n = 100; recovery range 20.789–98.631%). Pooled linear, quadratic, cubic, and pre-specified recovery-coordinate models demonstrate substantial process heterogeneity and do not support a universal recovery–specific-energy master curve. The records represent computational desalination scenarios rather than 1316 independently measured experimental observations. The comparison is therefore empirical and transferability-oriented rather than a claim of identical separation mechanisms or thermodynamic boundaries across desalination and pharmaceutical solvent recovery. [Blue dots represent individual desalination scenario records; the three panels show the complete scenario dataset, the high-recovery subset, and the maximum-recovery record for each process train, respectively].

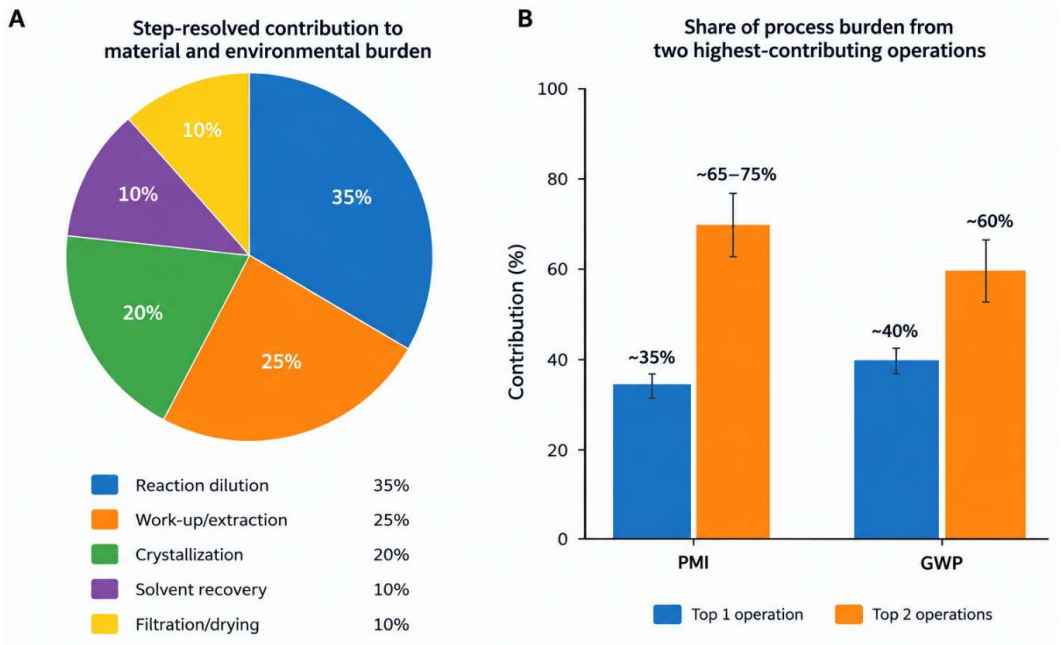


Figure 6. Step-resolved localization of material and environmental burden in the solvent-recovery process. (**A**) Relative contribution of reaction dilution (35%), work-up/extraction (25%), crystallization (20%), solvent recovery (10%), and filtration/drying (10%) within the documented step-resolved process framework. (**B**) The two highest-contributing operations

together account for approximately 65–75% of PMI and approximately 60% of GWP, illustrating concentration of process burden among major operations. This step-resolved analysis is distinct from the seven-record aggregate Dataset_S1 analysis and is used to localize process burdens rather than to establish a universal recovery relationship. The Impact Concentration Index is treated separately as a descriptive process-localization metric.

3.9. Candidate Recovery Optimum and Circularity Limits

The mathematical formulation permits a candidate stationary recovery condition to be defined by:

$$dL/d\eta = 0$$

Using $M(\eta) = M_0(1 - \eta)$, the phenomenological model becomes:

$$L(\eta) = kM_0(1 - \eta) + \beta[\eta^n/(1 - \eta)]$$

The derivative condition can therefore be evaluated for a specified parameter set. However, a mathematically defined stationary point should not automatically be interpreted as an experimentally observed optimum.

For the primary seven-record dataset, solvent mass intensity and energy demand both decrease over the observed recovery range from 30% to 92%. Consequently, the data do not demonstrate a measured minimum in total environmental or energy burden at an intermediate recovery. The present analysis therefore refers to an analytical candidate optimum condition, not an intrinsic or universal optimum recovery efficiency. Any practical optimum must be determined from system-specific measurements that include the relevant material, energy, environmental, economic, and quality constraints. This distinction prevents the mathematical stationary condition from being interpreted as evidence that maximizing or minimizing recovery has a universal sustainability optimum.

3.10. Solvent-Level Descriptors and Broader Solvent-System Behaviour

The solvent-level analysis provides complementary evidence that solvent identity and physicochemical properties influence sustainability-related outcomes. The nine-solvent dataset includes THF, DCM, DMF, 2-MeTHF, Cyrene, ethanol, CPME, GVL, and ethyl lactate. The descriptor analysis indicates that solvent properties contribute to differences in environmental impact and process suitability. The machine-learning model provides exploratory evidence for descriptor–impact relationships, but the LOOCV performance (R^2 = 0.483) indicates that the available dataset is insufficient to establish a highly predictive solvent-level model. The observed versus leave-one-out predicted GWP values from the exploratory Random Forest analysis are shown in Figure 4. These findings reinforce an important feature of the phenomenological material–energy scaling framework: the recovery coordinate does not replace solvent-specific chemistry or process configuration. Instead, the proposed scaling formulation is intended to provide a higher-level representation of material–recovery coupling while solvent-specific descriptors remain relevant to the magnitude and direction of the sustainability response. Thus, the observed behaviour should be interpreted as a system-level phenomenological relationship whose parameters may differ among solvents and processes. The solvent-level GWP and CED values for the nine-solvent dataset are summarized in Figure 3, providing the environmental-impact context for the accompanying descriptor analysis.

3.11. Comparative Evaluation of Linear, Nonlinear, and Extended Coupling Models

The primary seven-record dataset was used to compare three formulations: a linear material model, the proposed nonlinear recovery model, and an extended model incorporating an interaction between material intensity and recovery efficiency. Model performance was assessed using R^2, RMSE, MAE, AIC, BIC, and, where defined, AICc.

The linear model,

$$E_i = c + kM_i + \varepsilon_i,$$

provides a strong and parsimonious baseline description, with $R^2 = 0.979$. This result confirms that material throughput is the dominant explanatory variable within the primary dataset.

The nonlinear formulation,

$$E_i = c + kM_i + \beta\Phi(\eta_i;n) + \varepsilon_i,$$

increases R^2 only marginally, from 0.979 to 0.981. However, AIC and BIC increase relative to the simpler linear model, indicating that the additional nonlinear term does not improve the statistical performance sufficiently to offset its added complexity. Furthermore, the fitted exponent reaches the imposed upper bound (n = 10), indicating weak identification of the nonlinear component within this small dataset. The nonlinear recovery coordinate is therefore retained as a phenomenological hypothesis rather than as a statistically preferred replacement for the linear baseline.

The extended model,

$$E_i = c + kM_i + \beta\Phi(\eta_i;n) + \gamma M_i\eta_i + \varepsilon_i,$$

provides the highest in-sample explanatory performance, approximately $R^2 = 0.999$ in the revised fit. This formulation allows material intensity and recovery efficiency to interact and provides a substantially closer in-sample representation of the observations. However, the model contains five fitted parameters for only seven observations. Its coefficients are consequently sensitive to the limited sample size, and the model has not been independently validated. AICc is not defined for this parameterization. The higher in-sample R^2 should therefore not be interpreted as evidence of superior predictive performance. Table 2 presents the parameter estimates and comparative performance metrics of the three candidate formulations.

The model comparison provides an important result for the interpretation of the phenomenological material–energy scaling framework. The available pharmaceutical dataset does not demonstrate that the nonlinear asymptotic term is the principal determinant of observed variation. Instead, most of the explanatory power arises from the material term, with additional improvement obtained when material intensity and recovery efficiency are allowed to interact. The linear model provides the strongest parsimonious descriptive baseline for the seven-record dataset, while the nonlinear $\Phi(\eta;n)$ term represents a scientifically motivated hypothesis for recovery-dependent curvature. Its weak identification and boundary-constrained exponent indicate that the present dataset does not provide strong statistical support for treating the nonlinear term as the dominant contribution. Although the extended interaction model provides the closest in-sample fit, its high parameter-to-observation ratio limits its inferential value. Taken together, these results indicate that the present dataset supports comparison and further testing of the proposed formulations but does not establish a universal nonlinear recovery law. The results therefore support comparison of the formulations rather than selection of a universal governing model. The principal purpose, interpretation, and limitations of the three candidate formulations are summarized comparatively in Table S8 of the Supplementary Materials.

Table 2. Comparison of Candidate Material–Energy Scaling Models.

Model	Parameter	Estimate	R^2	RMSE	MAE	AIC	BIC	AICc
Linear	c	317.465	0.979	7.969	5.958	33.059	32.950	36.059
Linear	k	1.280	0.979	7.969	5.958	33.059	32.950	36.059
Nonlinear	c	307.544	0.981	7.669	5.609	36.521	36.304	56.521
Nonlinear	k	1.350	0.981	7.669	5.609	36.521	36.304	56.521
Nonlinear	β	1.927	0.981	7.669	5.609	36.521	36.304	56.521
Nonlinear	n	10.000	0.981	7.669	5.609	36.521	36.304	56.521
Extended	c	483.598	0.999	2.042	1.293	19.996	19.726	–

Extended	k	1.017	0.999	2.042	1.293	19.996	19.726	–
Extended	β	−7.549	0.999	2.042	1.293	19.996	19.726	–
Extended	n	9.478	0.999	2.042	1.293	19.996	19.726	–
Extended	γ	−2.574	0.999	2.042	1.293	19.996	19.726	–

The linear model was specified as $E = c + kM + \varepsilon$; the nonlinear model as $E = c + kM + \beta\Phi(\eta;n) + \varepsilon$, where $\Phi(\eta;n) = \eta^n/(1 - \eta)$; and the extended model additionally included the material–recovery interaction term $\gamma M\eta$. Model performance is reported using R^2, RMSE, MAE, AIC, BIC, and, where defined, AICc. Information criteria were calculated using a common Gaussian least-squares likelihood convention. The nonlinear exponent reached the upper fitting bound (n = 10), indicating that the asymptotic coordinate was weakly identified by the seven-record dataset. The extended model has five fitted parameters for seven observations; therefore, AICc is not reported because the small-sample correction is undefined. The substantially lower information criteria for the extended model should be interpreted cautiously because of the very small sample size and high parameter-to-observation ratio.

Pearson and Spearman correlations were calculated for the nine-solvent demonstration dataset using four predefined solvent descriptors and the corresponding literature/database-derived GWP and CED values. Table 3 shows the correlations between solvent descriptors and environmental-impact indicators. The strong negative association between biodegradability and both GWP and CED is descriptive of this small dataset and should not be interpreted as a causal or universal solvent-selection relationship. The analysis is exploratory and is intended to demonstrate how solvent-level descriptors can be integrated with environmental-impact indicators.

Table 3. Correlations between Solvent Descriptors and Environmental-Impact Indicators.

Descriptor	GWP Pearson r	GWP *p*-Value	GWP Spearman ρ	Spearman *p*-Value	CED Pearson r	CED *p*-Value	CED Spearman ρ	Spearman *p*-Value
Dipolarity index	0.452	0.222	0.450	0.224	0.471	0.200	0.517	0.154
H-bond capacity	−0.205	0.597	−0.367	0.332	−0.250	0.516	−0.317	0.406
Viscosity	−0.233	0.546	−0.483	0.187	−0.153	0.695	−0.500	0.170
Biodegradability	−0.913	0.001	−0.917	0.001	−0.916	0.001	−0.883	0.002

The supplied desalination scenario dataset contained 1316 process-level records. Pooled linear, quadratic, cubic, and pre-specified nonlinear recovery-coordinate models were evaluated against specific energy consumption (SEC). The high-recovery subset comprised 476 records with $\eta \geq 0.90$. A separate reduction to one maximum-recovery record per process-train label produced 100 representative records spanning 20.789–98.631% recovery. The desalination records are computational scenario records rather than 1316 independent experimental measurements. Desalination recovery–energy model comparison and dataset structure are shown in the Table 4. The results are used to examine empirical process heterogeneity and transferability of recovery–energy parameterizations, not to imply that desalination and pharmaceutical solvent recovery share identical physical mechanisms or thermodynamic boundaries.

Table 4. Desalination Recovery–Energy Model Comparison and Dataset Structure.

Model/Analysis	Number of Records	R^2	RMSE (kWh·m^{-3})	Recovery Subset/Range
Pooled linear	1316	0.036	17.024	Full dataset
Pooled quadratic	1316	0.241	15.107	Full dataset
Pooled cubic	1316	0.465	12.686	Full dataset
Pooled Φ(η;10)	1316	0.005	17.297	Full dataset
High-recovery subset	476	–	–	η ≥ 90%
Maximum-recovery record per process train	100	–	–	20.789–98.631%

3.12. Compact Cross-Domain Scaling Check

A compact cross-domain analysis was undertaken to examine whether the recovery coordinate has qualitative transferability beyond pharmaceutical solvent recovery. The check comprised 24 quantitative observations from five recovery-related domains: solvent recovery, battery recycling, metal recovery, carbon capture, and desalination. These observations were treated as an independent, pre-specified transferability check rather than as an extension of the primary pharmaceutical regression.

The purpose of this analysis was deliberately limited. The same recovery coordinate,

$$\Phi(\eta;n) = \eta^n/(1-\eta)$$

was used as a common comparison coordinate, with the pharmaceutical parameterization treated as prespecified rather than refitting a new universal exponent to the combined dataset. The analysis therefore asks whether a common recovery coordinate provides a useful descriptive ordering of normalized energetic or burden-related observations across domains.

The cross-domain observations show that recovery efficiency can be associated with substantially different energetic or environmental responses across process families. However, the observations remain separated by domain, reflecting differences in process architecture, physical mechanisms, system boundaries, and response definitions. The data therefore do not support a universal master curve. The resulting 24-observation cross-domain comparison using the fixed recovery coordinate is shown in Figure 7 (see the supplementary cross-domain scaling check in Figure S7). The result is consequently interpreted as limited evidence of coordinate-level transferability, rather than proof of universal scaling. The 24 observations, their recovery fractions, response values, and corresponding recovery-coordinate values are provided in Table 5, with the corresponding supplementary dataset reported in Table S5.

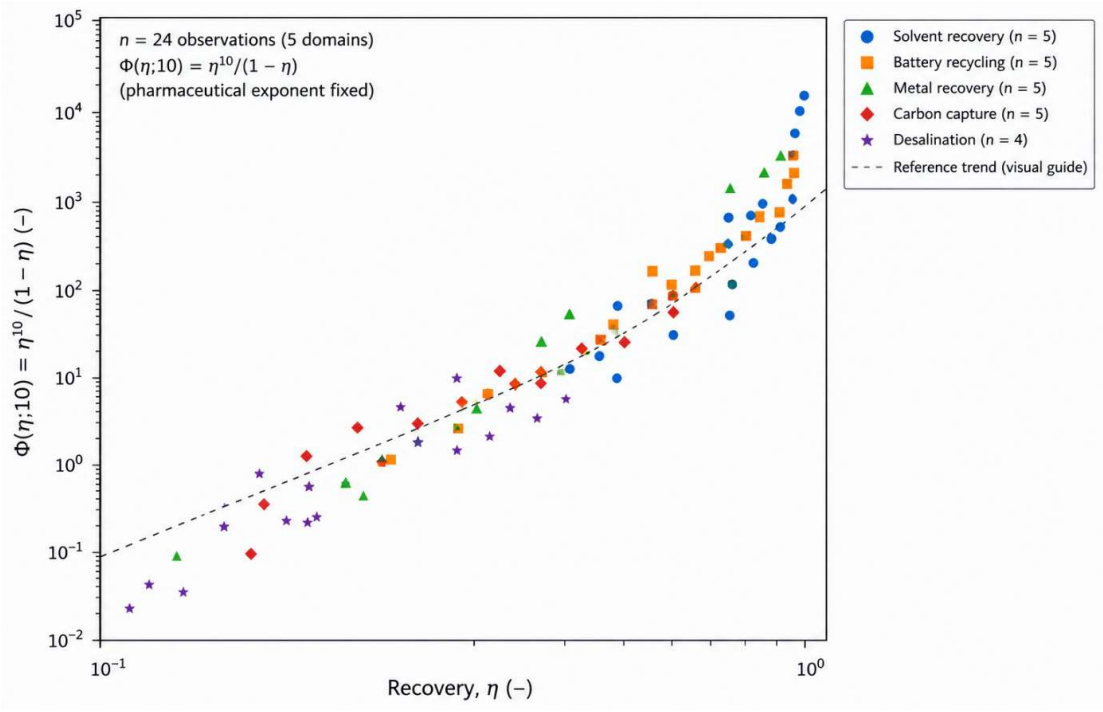


Figure 7. Compact cross-domain check of the candidate recovery coordinate. Twenty-four observations spanning solvent recovery, battery recycling, metal recovery, carbon capture and desalination are represented using the pre-specified recovery coordinate $\Phi(\eta;10) = \eta^{10}/(1-\eta)$, with the pharmaceutical exponent fixed rather than refitted to the pooled observations. Domain-specific response scales and process configurations are retained rather than collapsed into a single universal regression. The analysis is therefore a compact transferability/scaling check and does not establish a universal master curve, universal parameter values, or a common physical mechanism.

Table 5. Compact 24-Observation Cross-Domain Dataset Used for the Recovery-Coordinate Scaling Check.

No.	Domain	System	η	Response	Unit	Φ(η;10)
1	Solvent recovery	Ethanol	0.70	1.80	$MJ·kg^{-1}$	0.094158
2	Solvent recovery	Ethanol	0.85	3.00	$MJ·kg^{-1}$	1.312496
3	Solvent recovery	IPA	0.90	5.20	$MJ·kg^{-1}$	3.486784
4	Solvent recovery	Acetonitrile	0.95	8.50	$MJ·kg^{-1}$	11.974739
5	Battery recycling	Lithium hydrometallurgy	0.70	1.20	$MJ·kg^{-1}$	0.094158
6	Battery recycling	Lithium hydrometallurgy	0.90	2.50	$MJ·kg^{-1}$	3.486784
7	Battery recycling	Cobalt hydrometallurgy	0.95	3.80	$MJ·kg^{-1}$	11.974739
8	Battery recycling	Nickel DES-assisted	0.98	5.50	$MJ·kg^{-1}$	40.853640
9	Metal recovery	Cu hydrometallurgy	0.75	12.00	$MJ·kg^{-1}$	0.225254
10	Metal recovery	Ni recovery	0.82	18.00	$MJ·kg^{-1}$	0.763600
11	Metal recovery	Co recovery	0.88	25.00	$MJ·kg^{-1}$	2.320841
12	Metal recovery	REE extraction	0.91	35.00	$MJ·kg^{-1}$	4.326846
13	Metal recovery	Zn recovery	0.67	10.00	$MJ·kg^{-1}$	0.055238
14	Metal recovery	Al recycling	0.95	8.00	$MJ·kg^{-1}$	11.974739
15	Carbon capture	Amine_low	0.60	2.50	$GJ·tCO_2^{-1}$	0.015117
16	Carbon capture	Amine_mid	0.75	3.20	$GJ·tCO_2^{-1}$	0.225254
17	Carbon capture	Amine_high	0.90	5.50	$GJ·tCO_2^{-1}$	3.486784
18	Carbon capture	DAC_low	0.50	4.00	$GJ·tCO_2^{-1}$	0.001953
19	Carbon capture	DAC_mid	0.70	6.50	$GJ·tCO_2^{-1}$	0.094158
20	Carbon capture	DAC_high	0.90	9.00	$GJ·tCO_2^{-1}$	3.486784
21	Desalination	RO	0.30	0.75	$kWh·m^{-3}$	0.000008
22	Desalination	RO	0.50	3.00	$kWh·m^{-3}$	0.001953
23	Desalination	High-recovery RO	0.70	5.50	$kWh·m^{-3}$	0.094158
24	Desalination	Thermal	0.85	9.50	$kWh·m^{-3}$	1.312496

The dataset comprises 24 observations spanning solvent recovery, battery recycling, metal recovery, carbon capture, and desalination. The recovery coordinate was calculated as $\Phi(\eta;10) = \eta^{10}/(1 - \eta)$, using the pharmaceutical exponent as a fixed, pre-specified coordinate rather than refitting a universal exponent to the pooled observations. Response units differ among domains and are therefore retained explicitly; the table is intended as a compact transferability/scaling check rather than as evidence of a common absolute response scale, universal master curve, or shared physical mechanism.

4. Discussion

The present study builds on the previously published solvent-centric sustainability framework developed by us (Shaikh & Shaikh, 2026) [4], which integrated material-efficiency metrics with environmental and energy indicators to evaluate solvent selection, process design, and recovery. While that framework established an integrated basis for solvent-focused sustainability assessment, the present study examines the relationship between material intensity and recovery-associated contributions through an explicitly formulated phenomenological scaling approach. The distinction is methodological: the earlier framework provides the integrated assessment context, whereas the present work evaluates alternative mathematical representations of material–recovery coupling, their parameter identifiability, and their limitations when applied to a small primary pharmaceutical dataset.

Process-level interpretation: The step-resolved analysis reinforces the importance of evaluating solvent recovery within the broader process configuration rather than as an isolated circularity variable. Solvent quantity, recovery efficiency, separation technology, purity requirements, and energy demand interact at the unit-operation level. Accordingly, a recovery coordinate can provide a useful analytical representation of material–energy trade-offs only when its parameters are estimated for a defined solvent–process system

and system boundary. This interpretation is consistent with the broader development of hybrid modelling [7] approaches that combine process data with mechanistic separation information rather than assuming a single response function across technologies.

4.1. Mechanistic Interpretation of Non-Proportional Sustainability Behaviour

The present results indicate that solvent-system sustainability cannot be fully interpreted through proportional material-efficiency metrics alone. Increasing solvent recovery reduces solvent consumption and PMI, but the associated recovery operations introduce additional energetic and process constraints. The resulting relationship between material intensity and environmental burden is therefore not necessarily proportional across all operating conditions. The thermodynamic basis for increasing difficulty at high recovery is well established in separation engineering. As the unrecovered fraction becomes progressively smaller, further separation can require increasingly demanding process conditions, additional separation stages, greater selectivity, or higher energy input. This behaviour is therefore not proposed here as a newly discovered thermodynamic law. Rather, it provides the physical motivation for examining whether a compact recovery-dependent coordinate can improve material–energy accounting in solvent-recovery systems.

The proposed recovery coordinate,

$$\Phi(\eta;n) = \eta^n/(1 - \eta)$$

should consequently be interpreted as a phenomenological parameterization of recovery dependence rather than as a first-principles derivation of separation energetics. The present seven-record pharmaceutical dataset does not independently demonstrate energetic divergence as η approaches unity. Indeed, energy demand decreases from 520 to 380 MJ·kg^{-1} API while recovery increases from 30% to 92%. The strongest empirical finding is therefore the decoupling between the much larger reduction in solvent mass intensity and the more modest reduction in energy demand, rather than an observed divergence of energy at high recovery.

This distinction also explains the model-comparison results. The linear material model already explains most of the observed variance (R^2 = 0.979), whereas incorporation of the nonlinear recovery coordinate increases R^2 only marginally to 0.981. The fitted nonlinear exponent reaches the imposed upper parameter boundary, indicating weak identification rather than confirmation of a particular universal exponent. By contrast, the extended material–recovery interaction model gives a substantially higher in-sample fit, but its five fitted parameters relative to seven observations make it susceptible to overfitting and prevent interpretation as independently validated predictive evidence. Accordingly, the present results support further investigation of material–recovery interactions within the investigated pharmaceutical regime, rather than already establishing an interaction-dominated interpretation. Recovery efficiency is relevant, but its practical contribution is expressed through its interaction with material throughput and process configuration rather than through a demonstrably dominant universal asymptotic term.

4.2. System-Specific Parameters and the Role of β and λ

An important implication of the revised formulation is that β and λ should not be interpreted as universal constants.

In

$$L(\eta) = c + kM(\eta) + \beta[\eta^n/(1 - \eta)]$$

β represents the fitted magnitude of the recovery-associated contribution for a specified response variable, system boundary, process configuration, and normalization. In the dimensionless representation,

$$L^* = c^* + \alpha M^* + \lambda[\eta^n/(1 - \eta)]$$

$$\lambda = \beta/L_0$$

Thus, β and λ can differ among solvents, separation technologies, environmental indicators, and process configurations. Differences may arise from solvent volatility, boiling behaviour, phase equilibria, solvent–solute interactions, separation selectivity, heat integration, membrane performance, process scale, recovery technology, and the definition of the response variable itself.

The use of a common recovery coordinate therefore does not imply identical coefficients across systems. Instead, the framework separates the mathematical coordinate describing recovery dependence from the system-specific parameters determining its magnitude. This distinction is particularly important when comparing solvent systems with different physicochemical properties. Recent advances in organic-solvent nanofiltration illustrate how membrane design and solvent–membrane interactions can substantially alter recovery and purification performance. For example, high-flux OSN membranes have been developed for catalyst recovery and pharmaceutical purification, demonstrating that the practical energy and separation response depends strongly on membrane architecture and operating characteristics. Similarly, recent continuous-flow photoredox work demonstrates that catalyst recovery can be integrated with reaction processing through recyclable materials and nanofiltration, reinforcing the importance of process architecture when assessing recovery sustainability. Accordingly, the proposed framework should be understood as a system-aware scaling representation rather than a claim that all solvent systems possess a common numerical β, λ, or n.

4.3. Molecular Descriptors, Process Integration, and Predictive Sustainability

The solvent-level descriptor analysis provides complementary evidence that molecular properties influence sustainability-related outcomes. However, the revised LOOCV analysis gives $R^2 = 0.483$, RMSE = 1.295, and MAE = 1.112. These values support an exploratory descriptor–impact relationship but do not justify the stronger claim that solvent sustainability can be accurately predicted from the available descriptor set. Physicochemical properties remain relevant because they influence solvent selection, process compatibility, phase behaviour, separation selectivity, and the choice and performance of recovery technologies. The relationship between molecular characteristics and process-level energy demand is therefore mediated by process design rather than being direct or universal. This interpretation is consistent with recent hybrid modelling approaches to chemical separation. Ignacz et al. (2025) [7] demonstrated that combining data-driven and mechanistic models can support technology selection among nanofiltration, evaporation, extraction, and hybrid configurations, with energy and carbon-dioxide performance varying substantially across separation conditions. Such work supports the broader principle that molecular and process information should be considered together when evaluating separation sustainability. The present machine-learning analysis therefore functions as a supporting layer within the overall framework rather than as validation of the phenomenological material–energy scaling framework itself. The cheminformatics, machine-learning, and digital-twin components can support scenario analysis and process optimization, but they should remain conceptually distinct from the empirical testing of the material–energy scaling formulation.

4.4. Implications for Solvent-Centric Sustainability Frameworks

Existing sustainability frameworks commonly evaluate material efficiency, life-cycle impacts, solvent selection, energy use, and circularity through related but partially independent indicators. The present analysis indicates that these dimensions can become coupled when solvent recovery is considered explicitly. The material–energy scaling framework provides a phenomenological structure for representing this coupling through three complementary components: material intensity, a recovery-dependent coordinate, and process-level burden localization.

As f_i represents the fraction of total burden attributable to operation I, the equation of the Impact Concentration (IC) Index becomes as follows:

$$IC = \Sigma_i f_i^2 \text{ with } \Sigma_i f_i = 1$$

provides an additional structural descriptor of how environmental burden is distributed among process operations. Importantly, the framework does not replace established metrics such as PMI, E-factor, LCA, GWP, or CED. Rather, it provides a higher-level modelling structure through which these metrics can be examined in relation to solvent recovery and process configuration. The present results also indicate why a single sustainability metric can be insufficient for process decisions. A substantial reduction in solvent mass intensity does not necessarily translate into a proportional reduction in energy demand or environmental impact. Conversely, a recovery technology that increases solvent circularity may introduce additional energy, membrane, heat-integration, or process-control requirements.

Recent developments in membrane-based recovery and integrated reaction–separation systems further illustrate this systems perspective. OSN has been investigated for pharmaceutical purification and catalyst recovery, while recyclable catalyst systems have increasingly integrated reaction and separation within continuous processing. Emerging biobased membrane systems also demonstrate that membrane material design can affect selectivity and separation performance, further emphasizing the system-specific nature of recovery technologies. The practical implication is therefore not that solvent recovery should be maximized unconditionally. Instead, recovery should be optimized jointly with material use, energy demand, environmental burden, product quality, and process constraints.

Impact Concentration as a Process-Level Complement

The step-resolved findings further indicate that material and environmental burdens are not uniformly distributed across the process. The concentration of approximately 65–75% of PMI and approximately 60% of GWP within the two highest-contributing operations illustrates why aggregate recovery metrics should be interpreted together with process-level burden localization. The Impact Concentration (IC) index provides a simple quantitative and descriptive means of characterizing this distribution by quantifying the concentration of a specified burden across process steps, provided that a complete operation-level contribution matrix is available. It is not intended as a thermodynamic state variable, a universal recovery threshold, or an independent recovery criterion, nor does it establish causal relationships. Because the present seven-record dataset does not provide the required operation-level matrix across all observations, an empirical IC–recovery relationship cannot be established here. Its contribution to the present framework is therefore methodological and diagnostic: it offers a reproducible descriptor for identifying operations where recovery integration, process redesign, solvent substitution, or energy-efficiency interventions may have the greatest potential to influence the overall process burden.

4.5. Recovery Constraints, Circularity Limits, and Process Design

The mathematical formulation allows a stationary condition to be evaluated for specified model parameters: $dL/d\eta = 0$. However, the existence of a mathematical stationary point does not, in itself, establish an experimentally observed or intrinsic optimum. The seven-record pharmaceutical dataset does not provide sufficient evidence to identify a universal optimum recovery efficiency. Across the observed range, both solvent mass intensity and energy demand decrease as recovery increases. Accordingly, the results support the use of a material–recovery trade-off framework but do not demonstrate that the overall process burden reaches a minimum at a particular recovery level. The practical implication is that recovery should be evaluated alongside its associated material and energy requirements, rather than assuming that either maximum recovery or minimum solvent consumption is universally optimal. A candidate optimum may arise when the marginal benefit of reducing material consumption becomes comparable to the

additional burden associated with recovery operations. Its location, however, is expected to depend on the process configuration, solvent properties, separation technology, and system boundaries.

The model-comparison results further qualify this interpretation. Although the extended interaction formulation provides a closer in-sample fit, the limited number of observations does not establish that the interaction term identifies a reproducible optimum or a validated recovery-dependent energy penalty. The extended formulation should therefore be regarded as a candidate representation of material–recovery coupling, rather than evidence of an empirically established optimum. Approaching near-complete recovery may also be constrained by separation thermodynamics, transport limitations, membrane selectivity, solvent properties, heat requirements, and the degree of process integration. These factors provide the physical and operational context for interpreting the phenomenological recovery coordinate developed in this study. The framework does not establish a universal circularity threshold; rather, it provides a basis for examining the competing material and recovery contributions within a defined process. Consequently, recovery targets should be assessed in relation to process-specific performance, resource requirements, and system boundaries.

4.6. Cross-Domain Transferability: What the Present Analysis Does and Does Not Demonstrate

The compact cross-domain analysis involving 24 observations provides a limited test of whether the recovery coordinate can be used beyond pharmaceutical solvent recovery. The five represented domains—solvent recovery, battery recycling, metal recovery, carbon capture, and desalination—differ substantially in chemistry, transport mechanisms, process architecture, and system boundaries. The analysis therefore should not be interpreted as demonstrating a universal master curve. A common coordinate,

$$\Phi(\eta;n) = \eta^n/(1 - \eta)$$

provides a means of placing observations on a common mathematical axis, but it does not imply that the underlying mechanisms are identical. This distinction is particularly important for desalination. Aqueous membrane desalination and organic-solvent recovery involve different physical mechanisms and cannot be treated as mechanistically equivalent systems. Any comparison between them is therefore restricted to an empirical curve-fitting or coordinate-transferability exercise. It is not evidence that they obey an identical separation mechanism.

The cross-domain analysis consequently supports a narrower but more defensible conclusion: recovery efficiency provides a potentially useful common organizing variable for comparing constrained recovery systems, while the response coefficients and physical mechanisms remain domain-specific. The full exergy-harmonized cross-domain analysis, including broader statistical validation and generalized scaling, should remain outside the scope of the present solvent-focused manuscript [27,28]. That analysis constitutes a separate theoretical development rather than necessary evidence for the present phenomenological formulation.

4.7. Why the Nonlinear Term Is Retained Despite Its Limited Statistical Improvement

The relatively small improvement produced by the nonlinear term raises significant concerns. The present results do not justify retaining the term on the basis of statistical superiority alone. The nonlinear term is retained for three reasons:

First, its functional form encodes a physically motivated recovery coordinate that increases strongly as the unrecovered fraction becomes small. This behaviour is consistent with the established conceptual difficulty of approaching complete separation.

Second, the term allows recovery efficiency to be represented explicitly rather than treating solvent mass intensity as the sole explanatory variable. This is important because the primary data show that improvements in PMI and energy demand are not proportional.

Third, retaining the coordinate makes the framework testable against larger and independently measured datasets. The present seven observations are not sufficient to identify the nonlinear exponent robustly, but they establish a clear hypothesis that can be evaluated in future process-specific datasets.

The nonlinear term is therefore retained as a hypothesis-driven component of the framework**,** not because the present statistical analysis proves its necessity. This distinction also resolves the apparent tension between the model comparison and the theoretical formulation. The linear model is currently the better-supported empirical baseline for the seven-record dataset, whereas the nonlinear coordinate provides a structured hypothesis for future testing under recovery regimes and process configurations where nonlinear effects may become more pronounced.

The extended interaction model provides an exploratory indication that material throughput and recovery efficiency may jointly describe the observed response. However, its improved in-sample fit, based on five regression coefficients estimated from seven observations, is insufficient to establish interaction-driven dominance or independent predictive performance. The interaction term is therefore retained as a candidate extension for evaluation using larger and independently documented datasets.

4.8. Limitations and Future Validation

Several limitations should be recognized as follows:

First, the primary pharmaceutical dataset contains only seven normalized records. This restricts statistical power and makes parameter-rich models particularly vulnerable to overfitting.

Second, the nonlinear exponent is weakly identified and reaches the imposed upper parameter boundary. This prevents a strong empirical interpretation of n.

Third, the primary dataset does not directly demonstrate energetic divergence near complete recovery. The proposed asymptotic behaviour is therefore a hypothesis motivated by separation constraints rather than an observed divergence established by the present data.

Fourth, the molecular descriptor analysis is exploratory and has only moderate cross-validated predictive performance.

Fifth, the cross-domain dataset contains heterogeneous process types and response definitions. Its purpose in the present manuscript is limited to a compact transferability check rather than full universal model validation.

Sixth, the Impact Concentration Index requires operation-level burden data and should not be interpreted as an independently validated recovery law.

Future studies should therefore use larger, independently measured process datasets with explicit energy balances, recovery efficiencies, solvent identities, separation technologies, system boundaries, and uncertainty estimates. Independent validation should include out-of-sample prediction, process-family-specific parameter estimation, sensitivity analysis, and where appropriate, thermodynamic or exergy-based harmonization. Such work would provide a stronger basis for determining whether the proposed recovery coordinate remains useful across different separation classes and whether more general scaling relationships can be established.

4.9. Overall Interpretation and Significance

The principal contribution of this study is not the discovery of a new thermodynamic law. The contribution is the formulation of a phenomenological material–energy scaling framework that explicitly incorporates solvent recovery into sustainability assessment. The primary seven-record analysis indicates that material intensity provides a strong descriptive baseline for the observed energy response, while recovery efficiency provides an additional dimension for examining possible material–energy coupling. The comparative models show that the nonlinear recovery contribution is weakly identified in the available

data, whereas an explicit material–recovery interaction improves the in-sample fit but remains susceptible to overfitting. The Step-Resolved analysis further shows that a limited number of process operations can dominate environmental burden, offering a practical pathway to identify high-leverage interventions. The proposed material–energy scaling formulation should therefore be interpreted as a testable phenomenological framework for recovery-constrained solvent sustainability. Its mathematical form is general, but its coefficients and predictive performance are system-specific. This positioning also clarifies the relationship between the present work and broader developments in sustainable separation. Hybrid mechanistic–data-driven approaches, advanced organic-solvent nanofiltration, integrated reaction–separation systems, and emerging recyclable membrane technologies all point toward the importance of jointly considering material, energy, molecular, and process variables when optimizing chemical separations. The framework developed here provides one such mathematical representation, while recognizing that its broader generalization requires independent experimental and process-level validation.

5. Conclusions

This study develops a proposed phenomenological material–energy scaling framework for recovery-constrained solvent sustainability. Its central formulation is:

$$L(\eta) = c + kM(\eta) + \beta[\eta^n/(1 - \eta)]$$

with the corresponding dimensionless representation:

$$L^* = c^* + \alpha M^* + \lambda[\eta^n/(1 - \eta)]$$

where β and λ are system-specific parameters rather than universal constants. The framework provides a quantitative structure for examining the relationship between material intensity, recovery efficiency, and the associated sustainability response without implying a universal physical or thermodynamic law.

The primary seven-record pharmaceutical dataset demonstrates a strong association between material intensity and the selected sustainability response, with the linear model explaining most of the observed variation ($R^2 = 0.979$). Solvent mass intensity decreases from 150 to 45 kg·kg^{-1} API as recovery increases from 30% to 92%, while energy demand decreases more modestly, from 520 to 380 MJ·kg^{-1} API. These results indicate that material intensity and energy demand change at different rates across the observed recovery range; they do not demonstrate energetic divergence near complete recovery. The nonlinear recovery coordinate provides only a marginal improvement in fit ($R^2 = 0.981$), and its exponent is weakly identified, reaching the imposed upper fitting bound. The extended material–recovery interaction model achieves a closer in-sample fit, but its five regression coefficients relative to seven observations create a substantial risk of overfitting. Accordingly, the linear model provides the most parsimonious descriptive baseline, while the nonlinear and interaction formulations remain candidates for further investigation rather than independently validated predictive models.

The step-resolved analysis provides an additional process-level contribution by identifying the concentration of material and environmental burdens within a limited number of high-contribution operations. The Impact Concentration Index offers a compact descriptive measure of this structural localization, complementing the material–recovery framework without implying a universal relationship between impact concentration and recovery efficiency. The stationary condition, $dL/d\eta = 0$, is retained as an analytical means of identifying a candidate system-specific stationary point; however, the available data do not establish an intrinsic or universal optimum recovery efficiency.

The complementary solvent-level analysis provides additional exploratory insight into solvent descriptors and their potential relevance to sustainability-related behaviour, without constituting independent validation of the proposed material–energy scaling relationship. The 24-observation cross-domain analysis further explores the use of recovery efficiency as a common descriptive coordinate across

solvent recovery, battery recycling, metal recovery, carbon capture, and desalination. Given the heterogeneity of these observations, the results represent a preliminary transferability check rather than evidence of a universal master curve or a common physical mechanism. Together, the primary, step-resolved, solvent-level, and cross-domain analyses support examining material intensity, recovery efficiency, energy demand, and process-level burden localization within an integrated, system-aware framework. They do not establish a universal recovery threshold, a universally applicable scaling relationship, or a new fundamental thermodynamic law.

Further work should focus on independent process-level validation using larger, experimentally characterized datasets; explicit energy and, where appropriate, exergy accounting; system-specific parameter estimation; and out-of-sample testing across different solvents and separation technologies. Such studies will be necessary to assess the predictive scope of the proposed formulations and determine whether the recovery coordinate can support a broader generalized scaling framework.

Supplementary Materials

The following supporting information can be found at: https://www.sciepublish.com/article/pii/1340, Figure S1: Primary dataset diagnostics; Figure S2: Model-fit/residual diagnostics; Figure S3: Solvent descriptor/environmental-impact diagnostics; Figure S4: RF/LOOCV diagnostics; Figure S5: Desalination process-family diagnostics; Figure S6: Step-resolved burden localization/IC; Figure S7: 24-observation cross-domain scaling check; Figure S8: Mathematical sensitivity of the phenomenological recovery coordinate to the exponent n; Table S1: Normalized seven-record primary dataset and calculated cPMI; Table S2: Fitted model coefficients and model-comparison statistics; Table S3: Nine-solvent descriptor, GWP, and CED dataset; Table S4: Desalination model diagnostics; Table S5: 24-observation cross-domain dataset; Table S6: Independent External REE Recovery Dataset; Table S7: Equation and Model-Definition Summary; Table S8: Candidate-model interpretation.

Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript, the author used OpenAI solely for limited language refinement, specifically to appropriately position adverbial and other grammatical elements to improve transitions between sections and overall readability, without changing the authors' wording, scientific content, interpretation, or intended meaning. Nothing was copied and pasted into the submitted article. No generative AI or AI-assisted technology was used to generate, fabricate, alter, or interpret research data, figures, experiments, experimental results, or scientific conclusions. The author takes full responsibility for the content of the published article.

Acknowledgments

The author gratefully acknowledges the scientific literature and research community whose contributions have informed the development of the sustainability assessment and separation concepts discussed in this study. This paper is lovingly dedicated to the cherished memory of my angelic Mother, whose boundless love, strength, compassion, and unwavering faith in me have been a guiding light throughout my academic journey. Though she is no longer by my side, her spirit and values live on in my heart, transcending physical existence and time, forever inspiring my path. This work is a humble tribute to her eternal presence in my life.

Author Contributions

The author conceived the study, developed the phenomenological material–energy scaling framework, performed the analyses, prepared the figures and tables, and drafted and revised the manuscript.

Ethics Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The process-level and solvent-level files supplied with this study are Dataset_S1.csv, Dataset_S2.csv, and Dataset_S3.csv. The desalination scenario dataset comprises 1316 process-level scenario records and is provided in the accompanying desalination analysis package, including the ZLD/MLD subset and process-train information used for the analyses reported here. The desalination records are treated as computational scenario data rather than independently verified experimental measurements. Because provenance and system-boundary metadata are incomplete for some supplied datasets, the corresponding analyses are presented as empirical and exploratory rather than as independently validated measurements. The datasets and analysis scripts can be provided separately in the accompanying data package upon request.

Funding

The author's affiliated university provided no financial support for this study. The research received no external funding.

Declaration of Competing Interest

The author declares no competing interests.

References

1. Roschangar F, Sheldon RA, Senanayake CH. Overcoming barriers to green chemistry in the pharmaceutical industry—The Green Aspiration Level™ concept. *Green Chem.* **2015**, *17*, 752–768. DOI:10.1039/C4GC01563K
2. Trost BM. The atom economy—A search for synthetic efficiency. *Science* **1991**, *254*, 1471–1477. DOI:10.1126/science.1962206
3. Udugama I, Mansouri S, Mitic A, Flores-Alsina X, Gernaey K. Perspectives on resource recovery from bio-based production processes: From concept to implementation. *Processes* **2017**, *5*, 48. DOI:10.3390/pr5030048
4. Shaikh IR, Shaikh MM. Solvent-centric sustainability framework for pharmaceutical process chemistry: Integrated metrics, circularity, and digital tools demonstrated through a Sertraline case study. *Green Chem. Technol.* **2026**, *3*, 10007. DOI:10.70322/gct.2026.10007
5. Cenci F, Diab S, Ferrini P, Harabajiu C, Barolo M, Bezzo F, Facco, P. Predicting drug solubility in organic solvents mixtures: A machine-learning approach supported by high-throughput experimentation. *Int. J. Pharm.* **2024**, *660*, 124233. DOI:10.1016/j.ijpharm.2024.124233
6. Zhou F, Shi D, Mu W, Wang S, Wang Z, Wei C, Li R, Mu, T. Machine learning models accelerate deep eutectic solvent discovery for the recycling of lithium-ion battery cathodes. *Green Chem.* **2024**, *26*, 7857–7868. DOI:10.1039/D4GC01418A
7. Ignacz G, Beke AK, Toth V, Szekely G. A hybrid modelling approach to compare chemical separation technologies in terms of energy consumption and carbon dioxide emissions. *Nat. Energy* **2025**, *10*, 308–317. DOI:10.1038/s41560-024-01668-7
8. Che R, Song W, Peng Y, Wang B, Li C, Su B. Tris(hydroxymethyl)aminomethane-modulated high-flux organic solvent nanofiltration membrane for catalyst recovery and pharmaceutical purification. *J. Membr. Sci.* **2026**, *747*, 125328. DOI:10.1016/j.memsci.2026.125328
9. Gede M, Ignacz G, De Castro CSP, Dereli B, Ferree M, Laquai F, Cavallo, L, Szekely, G. Tailoring polymers of intrinsic microporosity as photoredox catalysts for continuous-flow reaction–separation processes. *Nat. Commun.* **2026**, *17*, 7062. DOI:10.1038/s41467-026-73833-3

10. Oldal DG, Hardian R, Vovusha H, Peskov MV, Schwingenschlögl U, Shi C, Chen, EYX, Szekely, G. Stereoregular self-assembly of biobased recyclable polymer membranes for enantioselective nanofiltration. *Matter* **2026**, *9*, 102797. DOI:10.1016/j.matt.2026.102797
11. Sheldon RA. The E factor 25 years on: The rise of green chemistry and sustainability. *Green Chem.* **2017**, *19*, 18–43. DOI:10.1039/C6GC02157C
12. Byrne FP, Jin S, Paggiola G, Petchey THM, Clark JH, Farmer TJ, Hunt, AJ, McElroy, CR, Sherwood, J. Tools and techniques for solvent selection: Green solvent selection guides. *Sustain. Chem. Process.* **2016**, *4*, 7. DOI:10.1186/s40508-016-0051-z
13. Constable DJC, Dunn PJ, Hayler JD, Humphrey GR, Leazer JL, Linderman RJ, Lorenz, K, Manley, J, Pearlman, BA, Wells, A, et al. Key green chemistry research areas—A perspective from pharmaceutical manufacturers. *Green Chem.* **2007**, *9*, 411–420. DOI:10.1039/B703488C
14. Marchetti P, Jimenez Solomon MF, Szekely G, Livingston AG. Molecular separation with organic solvent nanofiltration: A critical review. *Chem. Rev.* **2014**, *114*, 10735–10806. DOI:10.1021/cr500006j
15. Harper G, Sommerville R, Kendrick E, Driscoll L, Slater P, Stolkin R, Walton, A, Christensen, P, Heidrich, O, Lambert, S, et al. Recycling lithium-ion batteries from electric vehicles. *Nature* **2019**, *575*, 75–86. DOI:10.1038/s41586-019-1682-5
16. Rinne M, Aromaa-Stubb R, Elomaa H, Porvali A, Lundström M. Evaluation of hydrometallurgical black mass recycling with simulation-based life cycle assessment. *Int J Life Cycle Assess* **2024**, *29*, 1582–1597. DOI:10.1007/s11367-024-02304-y
17. Zhou K, Dai X, Li P, Zhang L, Zhang X, Wang C, Wen, J, Huang, G, Xu, S. Recent advances in deep eutectic solvents for next-generation lithium batteries: Safer and greener. *Prog. Mater. Sci.* **2024**, *146*, 101338. DOI:10.1016/j.pmatsci.2024.101338
18. Reuter MA, van Schaik A, Gutzmer J, Bartie N, Abadías-Llamas A. Challenges of the circular economy: a material, metallurgical, and product design perspective. *Annu. Rev. Mater. Res.* **2019**, 49, 253–274. DOI:10.1146/annurev-matsci-070218-010057.
19. Reuter MA, Heiskanen K, Boin U, Van Schaik A, Verhoef E, Yang Y, et al. *The Metrics of Material and Metal Ecology: Harmonizing the Resource, Technology and Environmental Cycles*; Elsevier: Amsterdam, The Netherlands, 2005; pp. 3–706.
20. Rochelle GT. Amine scrubbing for CO_2 capture. *Science* **2009**, *325*, 1652–1654. DOI:10.1126/science.1176731
21. Keith DW, Holmes G, St. Angelo D, Heidel K. A process for capturing CO_2 from the atmosphere. *Joule* **2018**, *2*, 1573–1594. DOI:10.1016/j.joule.2018.05.006
22. Elimelech M, Phillip WA. The future of seawater desalination: Energy, technology, and the environment. *Science* **2011**, *333*, 712–717. DOI:10.1126/science.1200488
23. Stillwell AS, Clayton ME, Webber ME. Technical analysis of a river basin-based model of advanced power plant cooling technologies for mitigating water management challenges. *Environ. Res. Lett.* **2011**, *6*, 034015. DOI:10.1088/1748-9326/6/3/034015
24. Mistry KH, McGovern RK, Thiel GP, Summers EK, Zubair SM, Lienhard JH. Entropy generation analysis of desalination technologies. *Entropy* **2011**, *13*, 1829–1864. DOI:10.3390/e13101829
25. O'Connell MG, Rajendran N, Elimelech M, Gilron J, Dunn JB. Analysis of energy, water, land and cost implications of zero and minimal liquid discharge desalination technologies. *Nat. Water* **2024**, *2*, 1116–1127. DOI:10.1038/s44221-024-00327-1
26. Peelman S, Sun ZHI, Sietsma J, Yang Y. *Leaching of Rare Earth Elements: Review of Past and Present Technologies*; Elsevier: Amsterdam, The Netherlands, 2016; pp. 319–334. ISBN 9780128023280.
27. Shaikh IR. Toward Computational Sustainability Physics: A Thermodynamic Scaling Framework for Constrained Circularity Across Recovery Systems. Preprint 2026. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=7051826 (accessed on 7 July 2026).
28. Shaikh I. A Scaling Law for Solvent-System Sustainability in Pharmaceutical Manufacturing. Preprint 2026. Available online: https://www.researchsquare.com/article/rs-9611365/v1 (accessed on 5 May 2026).