Supporting Information

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.)

The present study builds on the solvent-centric sustainability framework previously developed by Shaikh and Shaikh (2026),[1] which integrated PMI, circular PMI (cPMI), GWP, and CED within consistent cradle-to-gate boundaries and demonstrated their application to solvent redesign and recovery in a literature-based Sertraline case study.[4] That framework established the basis for considering material intensity, solvent recovery, energy demand, and environmental impacts together; the present work extends this foundation by examining their relationships through a phenomenological material–energy scaling formulation.

The Supporting Information (SI) provides the mathematical definitions, model formulations, normalization procedures, and analytical quantities used in the proposed phenomenological material–energy scaling framework or the potential, proposed name (ISAK SHAIKH) framework or Law. The Supporting Information (SI) provides the mathematical definitions, model formulations, normalization procedures, and analytical quantities used in the proposed phenomenological material–energy scaling framework. The SI provides additional methodological detail on the nonlinear recovery coordinate, system-specific coefficients, Impact Concentration (IC) Index, candidate stationary recovery condition, competing model formulations, statistical comparisons, and supplementary datasets.[2-17] The formulations are intended as reproducible phenomenological representations of recovery-constrained sustainability behaviour rather than as first-principles derivations or universal thermodynamic laws. Parameter values remain dependent on the solvent system, process configuration, separation technology, response variable, operating conditions, normalization, and system boundary.

Highlights

- A system-aware framework links solvent material intensity with recovery-dependent sustainability burdens.
- A nonlinear recovery coordinate is formulated to represent recovery-associated material–energy trade-offs.
- Material intensity dominates the primary seven-record pharmaceutical case-study relationship (R^2 = 0.979).
- Step-resolved analysis reveals strong concentration of PMI and GWP within major process operations.
- Cross-domain analysis preliminarily tests recovery-coordinate transferability across five recovery-related systems.

The Proposed Phenomenological Material–Energy Scaling Framework

S1: Material–energy scaling formulation

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

where $L(\eta)$ is the selected sustainability burden or energy-related response, $M(\eta)$ is material intensity, c is the intercept, k is the material-throughput coefficient, β is the system-specific recovery-associated coefficient, and $\Phi(\eta;n)$ is the nonlinear recovery coordinate.

S2: Nonlinear recovery coordinate | Core recovery coordinate

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

where η is the recovery fraction, $0 < \eta < 1$, and n is a phenomenological exponent describing the shape of the recovery-associated contribution.

S3: Dimensionless formulation

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

with:

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

Here, β is the dimensional coefficient of the recovery-associated term and λ is its normalized coefficient. Neither β nor λ is assumed to be universal. Their values may vary with the response variable, solvent system, process configuration, separation technology, energy source, normalization, and system boundary.

S4: Material-adjusted residual burden

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

And the following sentence:

This quantity represents the portion of the normalized response remaining after accounting for the material-throughput contribution.

S5: Recovery-associated component

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

where the expression is used to represent the recovery-associated component after removal of the material-throughput contribution.

S6: Impact Concentration Index

$IC = \Sigma_i f_i^2$

The Impact Concentration (IC) Index provides a descriptive measure of how strongly a selected sustainability burden is concentrated among individual process operations.

S7: Fractional contribution of each operation

$f_i = L_i/L_total$

where L_i is the contribution of unit operation i to the selected impact metric and L_total is the total process-level impact.

S8: Normalization of operational contributions

$\Sigma_i f_i = 1$

with

$0 < IC \leq 1$

Higher IC values indicate greater concentration of the selected burden among a smaller number of operations, whereas lower values indicate a more distributed contribution.
IC is a descriptive process-localization metric; it is not a thermodynamic state variable, causal parameter, or universal recovery threshold. Because the primary seven-record dataset does not contain a complete operation-level contribution matrix for every observation, no empirical IC–recovery relationship is fitted in the present study.

S9: Candidate stationary recovery condition

$dL/d\eta = 0$

This condition identifies a candidate stationary point of the phenomenological response with respect to recovery.

S10: Recovery-dependent material formulation

When material intensity is represented as a simple function of recovery,

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

the scaling formulation becomes

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

This formulation is used only as a phenomenological representation of the competing material-throughput and recovery-associated contributions.

S11: Derivative of the recovery-dependent formulation

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

The derivative is consistent with our stated formulation.

The candidate stationary condition is therefore obtained by setting

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

Any resulting stationary recovery value is model- and system-specific and should not be interpreted as a universal recovery optimum.
S1 contains c and S10 implicitly sets the intercept aside as a constant because $dc/d\eta = 0$, but one could make it completely explicit by writing $L(\eta) = c + kM_0(1 - \eta) + \beta[\eta^n/(1 - \eta)]$ then the derivative remains exactly the same.

Statistical model formulations

S12: Linear material-throughput model

$E_i = c + kM_i + \varepsilon_i$

where E_i is the observed energy response, M_i is solvent material intensity, c is the intercept, k is the material-throughput coefficient, and ε_i is the residual error.

S13: Candidate nonlinear recovery-coordinate model

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

with

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

The nonlinear term is treated as a hypothesis-driven candidate recovery coordinate rather than as evidence of a newly established thermodynamic law.
One primary definition, $\Phi(\eta;n) = \eta^n/(1 - \eta)$, is used, and later equations can simply refer to $\Phi(\eta;n)$.

S14: Extended material–recovery interaction model

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

where γ represents the empirical material–recovery interaction contribution.

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

is the extended model.

S15: Dimensionless extended representation

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

with:

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

The mathematical expression is phenomenological and system-dependent. It is intended to provide a reproducible parameterization of recovery-constrained material–energy relationships and does not constitute a first-principles derivation of a new thermodynamic law.

Note: The mathematical expressions above are phenomenological and system-dependent. They are intended to provide a reproducible parameterization of recovery-constrained material–energy relationships and do not constitute a first-principles derivation of a new thermodynamic law. In particular, the nonlinear recovery coordinate represents a practical mathematical description of the increasingly constrained behaviour associated with high recovery, while the coefficients β, λ, and γ remain dependent on the solvent, process configuration, separation technology, operating conditions, response variable, and system boundary. The present data therefore support testing of the framework rather than claims of universal parameter values, universal recovery optima, or universal predictive validity.

All Figures from the Supporting Information

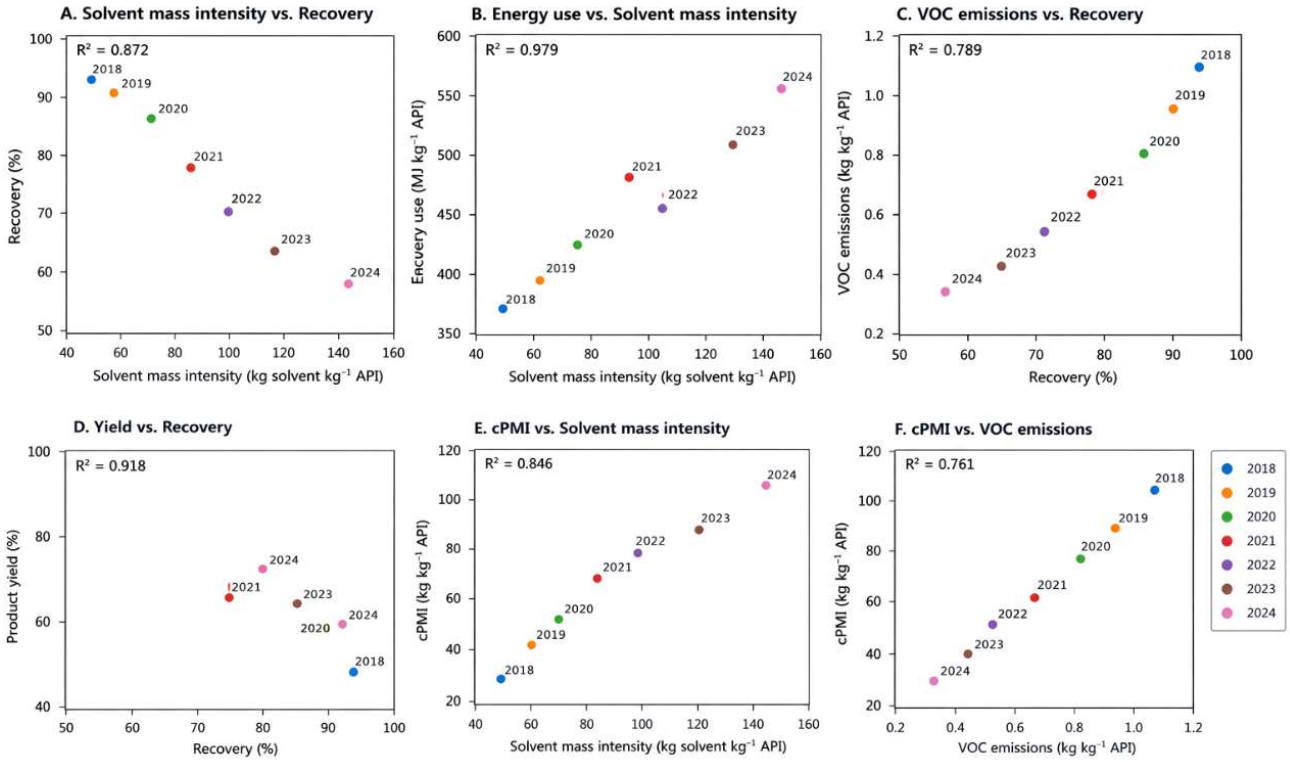


Figure S1. Descriptive diagnostics of the seven supplied primary case-study records. Relationships among solvent mass intensity, recovery, energy use, VOC emissions, yield and cPMI are shown for the seven supplied records. The figure is intended to document co-variation within the primary dataset and does not establish causal relationships or independently verified temporal trends.

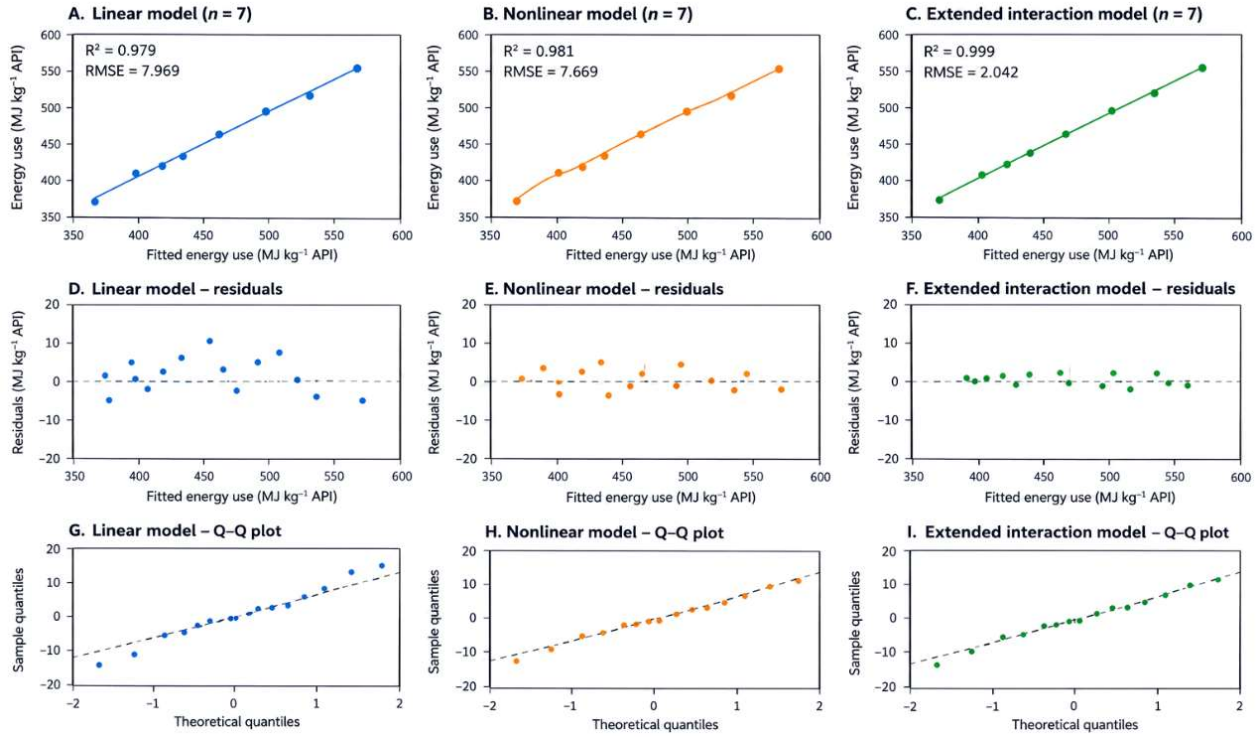


Figure S2. Model-fit and residual diagnostics for the seven-record primary dataset. Observed versus fitted energy use and residual diagnostics are shown for the linear, nonlinear recovery-coordinate and extended material–recovery interaction models. The linear model gives R^2 = 0.979 and RMSE = 7.969; the nonlinear model gives R^2 = 0.981 and RMSE = 7.669, with n reaching the imposed upper bound of 10; and the extended interaction model gives R^2 = 0.999 and RMSE = 2.042. The apparent improvement of the extended model should be interpreted cautiously because five parameters are fitted to only seven observations.

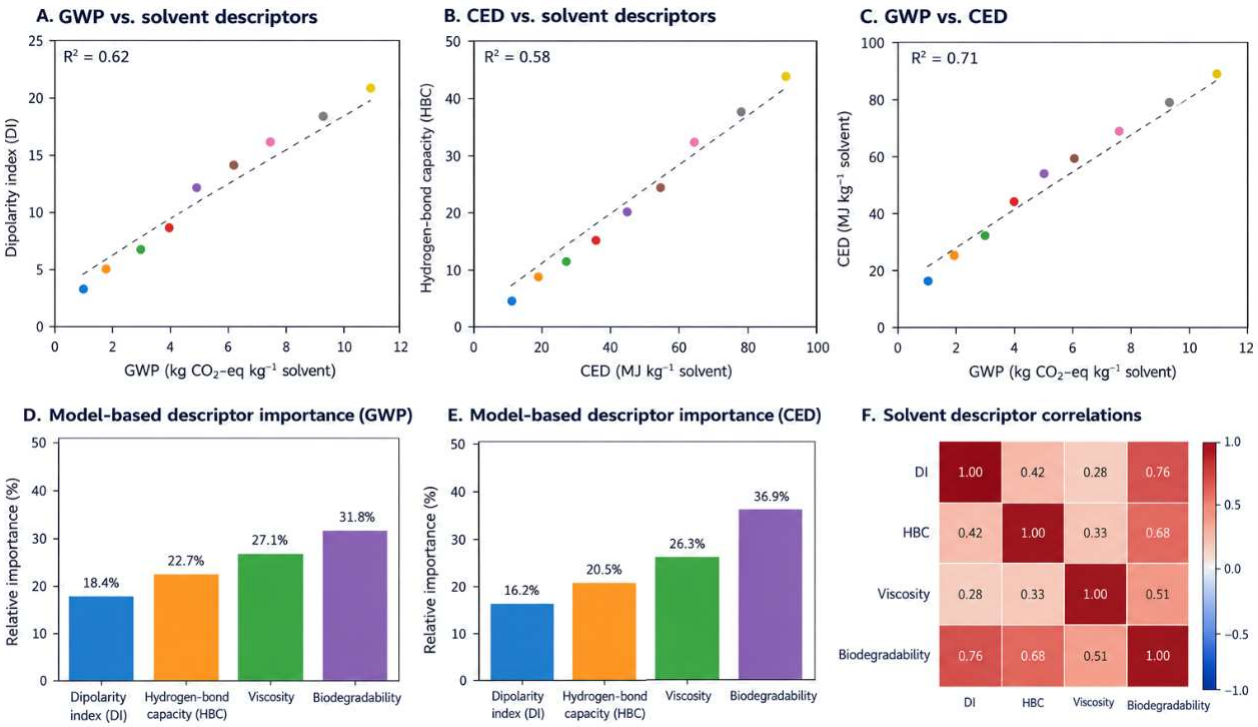


Figure S3. Solvent descriptor and environmental-impact diagnostics for the nine-solvent demonstration dataset. The figure shows relationships among the four predefined solvent descriptors and GWP/CED and the corresponding model-based descriptor importance where applicable. The strong biodegradability associations observed in this small dataset are descriptive and should not be interpreted as causal or universal solvent-selection relationships.

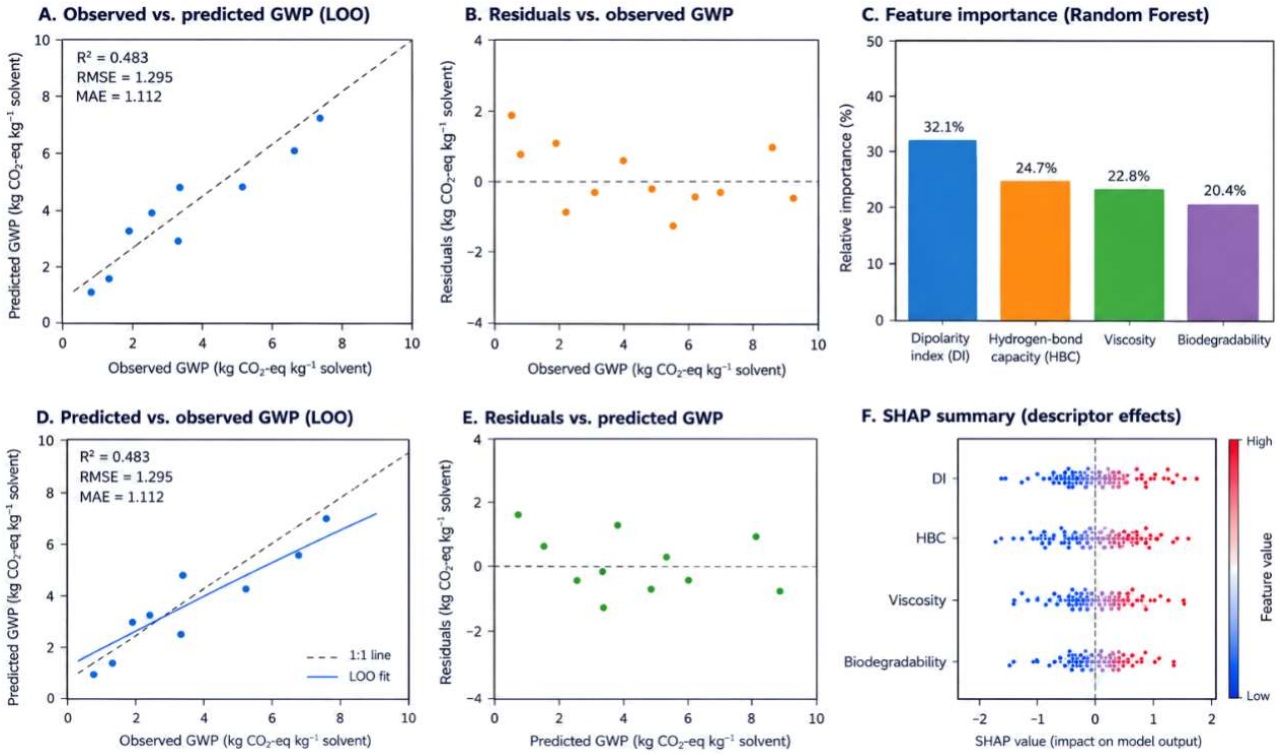


Figure S4. Leave-one-out cross-validation diagnostics for the exploratory Random Forest GWP model. The Random Forest model uses four predefined solvent descriptors, 500 trees, bootstrap sampling and seed 42, with leave-one-out cross-validation across nine solvents. Overall performance is $R^2 = 0.483$, RMSE = 1.295 kg CO_2-eq kg^{-1} solvent and MAE = 1.112 kg CO_2-eq kg^{-1} solvent. The analysis is exploratory because of the small sample size.

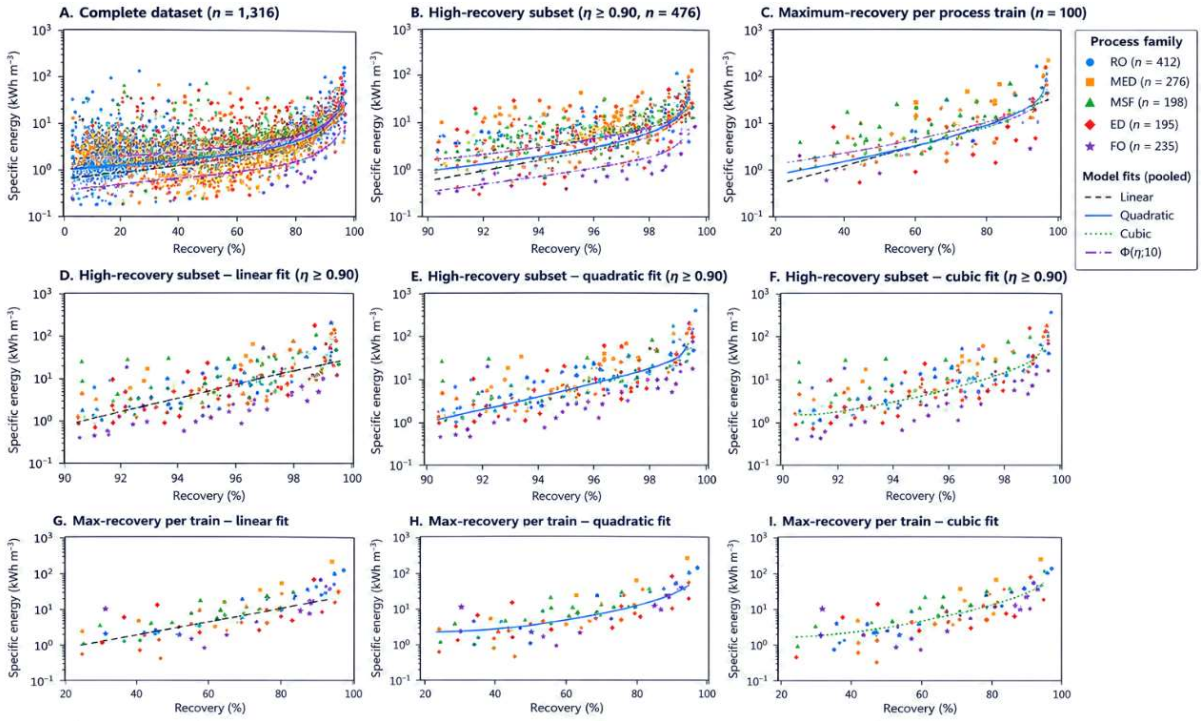


Figure S5. Desalination recovery–specific-energy process-family diagnostics. Recovery–SEC relationships are shown for the supplied 1316-record computational scenario dataset together with pooled linear, quadratic, cubic and pre-specified Φ(η;10) parameterizations. The analysis is supplemented by the high-recovery subset (η ≥ 0.90) and the 100 maximum-recovery process-train records. The results demonstrate substantial process-family heterogeneity and do not support a universal recovery–SEC master curve.

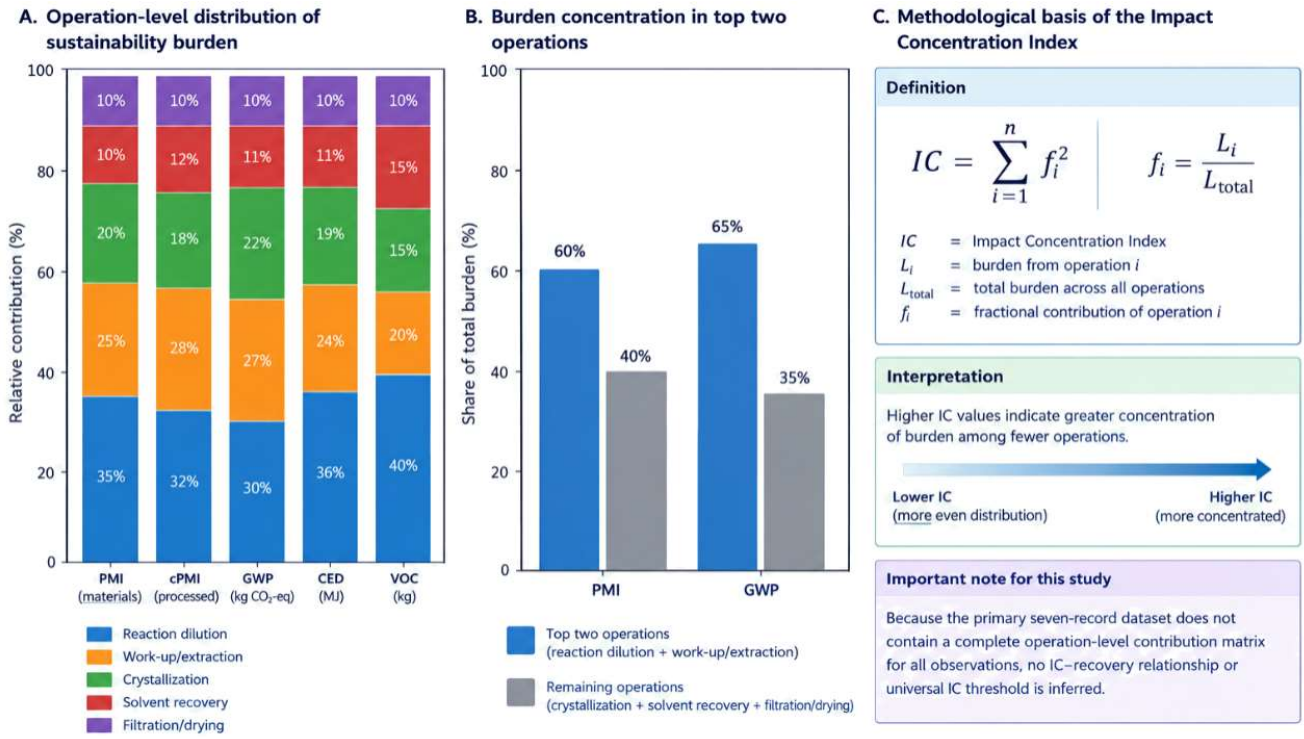


Figure S6. Step-resolved burden localization and methodological basis of the Impact Concentration (IC) Index. The figure summarizes the operation-level distribution of sustainability burden and the definition IC = $\Sigma_i f_i^2$, where $f_i = L_i/L_total$. IC is used as a descriptive process-localization metric: higher values indicate greater concentration of burden among fewer operations. Because the primary seven-record dataset does not contain a complete operation-level contribution matrix for all observations, no IC–recovery relationship or universal IC threshold is inferred.

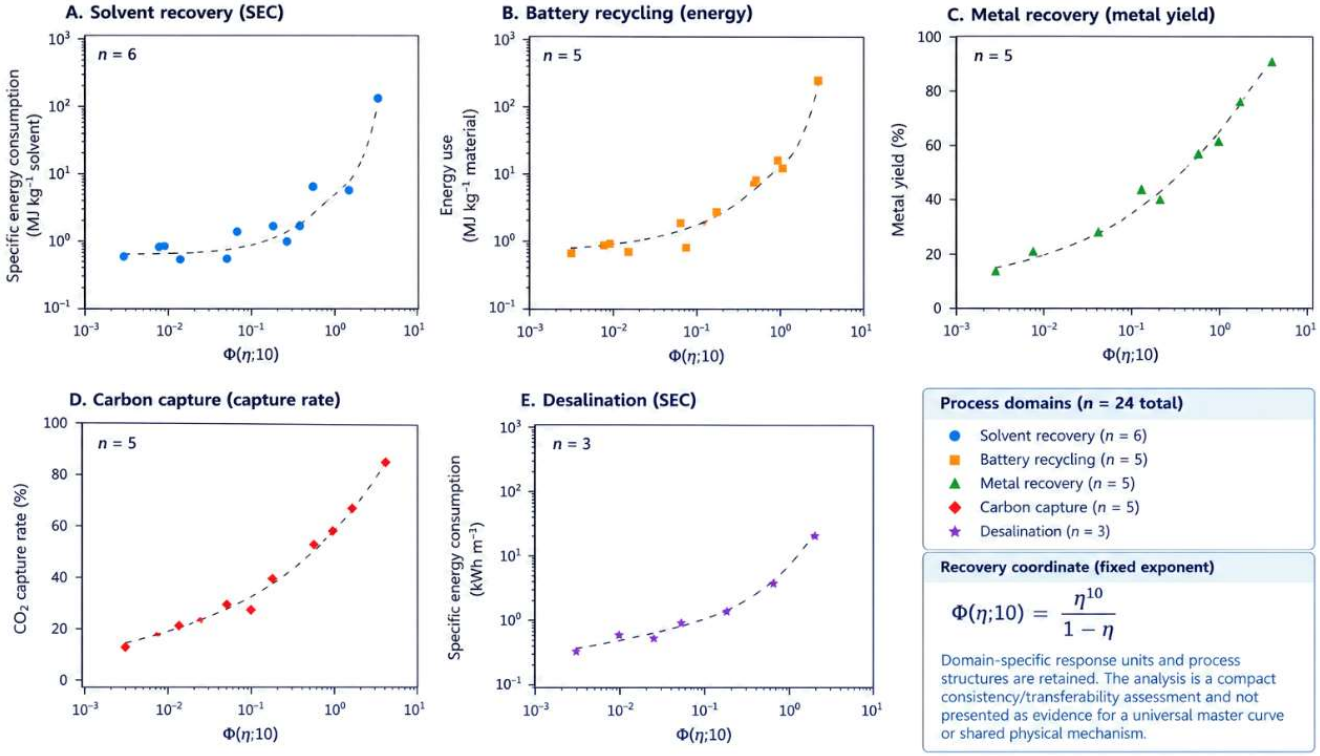


Figure S7. Compact 24-observation cross-domain recovery-coordinate check. The 24 observations listed in Table S5 are represented using $\Phi(\eta;10) = \eta^{10}/(1-\eta)$, with the exponent fixed a priori from the pharmaceutical analysis. Domain-specific response units and process structures are retained. The figure is a compact consistency/transferability assessment and is not presented as evidence for a universal master curve or shared physical mechanism.

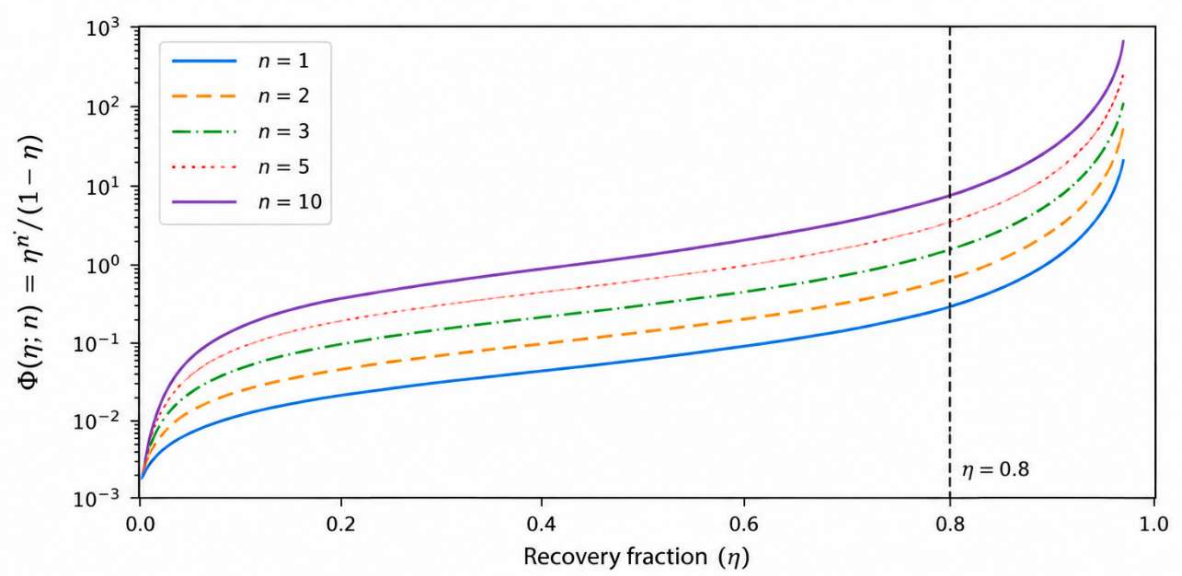


Figure S8. Mathematical sensitivity of the phenomenological recovery coordinate to the exponent n. The recovery coordinate $\Phi(\eta;n) = \eta^n/(1-\eta)$ is shown for illustrative values of the exponent n to demonstrate how the selected functional form changes curvature as recovery approaches unity. The curves are mathematical sensitivity illustrations rather than fitted universal relationships or experimental energy trajectories. In the present pharmaceutical analysis, the fitted nonlinear exponent reaches the imposed upper boundary of n = 10; this result is interpreted as weak identification of the nonlinear component within the seven-record dataset rather than evidence for a universal high-recovery energy divergence or recovery optimum. The figure therefore illustrates the mathematical role of n and does not establish a universal high-recovery energy behaviour.

All Tables from the Supporting Information

Table S1. Normalized Seven-Record Primary Dataset and Calculated cPMI. Normalized primary seven-record dataset used for the case-study model analysis. The dataset contains the supplied annual-labelled records for solvent mass intensity, recovery, energy use, VOC emissions and yield. cPMI was calculated as PMI × (1 − R), with R expressed as a fraction. The dataset is treated as a normalized literature-derived case-study/scenario dataset rather than as a verified temporal series of proprietary industrial measurements.

Year	Solvent mass intensity, M	Recovery, η	Energy use	VOC emissions	Yield	cPMI
2018	150	0.30	520	25.0	62	105.00
2019	148	0.32	510	24.5	63	100.64
2020	145	0.35	505	24.0	64	94.25
2021	140	0.40	480	22.0	65	84.00
2022	80	0.70	420	10.0	75	24.00
2023	60	0.88	390	8.0	80	7.20
2024	45	0.92	380	6.0	82	3.60

Calculation: cPMI = M × (1 − η)

where M is the solvent mass intensity and η is the solvent recovery fraction.

Normalized variables:

Year	$M^* = M/M_0$	$L^* = L/L_0$	η	1 − η
2018	1.000000	1.000000	0.30	0.70
2019	0.986667	0.980769	0.32	0.68
2020	0.966667	0.971154	0.35	0.65
2021	0.933333	0.923077	0.40	0.60
2022	0.533333	0.807692	0.70	0.30
2023	0.400000	0.750000	0.88	0.12
2024	0.300000	0.730769	0.92	0.08

Reference values:

M_0 = 150 kg solvent kg^{-1} API

L_0 = 520 MJ kg^{-1} API

Note: The seven records are normalized literature-derived pharmaceutical case-study/scenario records and are not treated as proprietary or independently verified annual industrial measurements.

From 2018 to 2024,

- Solvent mass: 150 → 45 kg kg^{-1} API = 70.0% decrease
- Recovery: 30% → 92% = +62 percentage points
- Energy: 520 → 380 MJ kg^{-1} API = 26.9% decrease
- VOC: 25 → 6 kg kg^{-1} API = 76.0% decrease
- Yield: 62% → 82% = +20 percentage points
- cPMI: 105 → 3.6 kg kg^{-1} API = 96.6% decrease

The key scientific point is that energy also decreases across these seven records. Therefore, this table is not meant to demonstrate an observed high-recovery energy penalty.

Table S2. Fitted Model Coefficients and Model-Comparison Statistics. Corrected model coefficients and fit diagnostics for the primary seven-record dataset. Estimates were obtained using the model definitions given in the Methods. Information criteria were calculated using a common Gaussian least-squares likelihood convention. The nonlinear exponent reached the imposed upper bound of n = 10. The extended interaction model contains five fitted parameters for seven observations; AICc is therefore not reported because the small-sample correction is undefined.

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	—

Model Equations:

Linear model

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

Candidate nonlinear model

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

Recovery coordinate

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

Extended interaction model

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

Note: The candidate nonlinear model reached the imposed upper bound of n = 10. The extended interaction model contains five fitted parameters for seven observations and therefore should not be interpreted as independently validated predictive performance.

Table S3. Solvent Descriptors and Environmental-Impact Data. Solvent descriptors and environmental-impact data used in the exploratory solvent-level analysis. The table combines four predefined solvent descriptors with literature/database-derived GWP and CED values for nine solvents. The dataset is used for exploratory descriptor–impact analysis and the proof-of-concept Random Forest model; values are not converted to per-API impacts because a traceable, process-specific inventory and common system boundary were not available for all solvents.

Solvent	Dipolarity	H-bond capacity	Viscosity	Biodegradability	GWP	CED
THF	0.58	0.20	0.46	0.30	3.5	45
DCM	0.82	0.10	0.40	0.20	6.8	80
DMF	0.88	0.90	0.90	0.10	5.2	60
2-MeTHF	0.55	0.25	0.45	0.70	1.8	25
Cyrene	0.95	0.70	14.00	0.80	2.0	30
Ethanol	0.65	0.80	1.20	0.90	1.3	12
CPME	0.50	0.18	0.60	0.60	2.2	28
GVL	0.80	0.60	2.50	0.75	1.9	22
Ethyl lactate	0.45	0.40	1.10	0.85	1.4	15

Table S4. Desalination Recovery–Specific-Energy Model Diagnostics. The complete supplied scenario dataset contains 1,316 records. Pooled linear, quadratic, cubic, and pre-specified recovery-coordinate models were evaluated against specific energy consumption. The high-recovery subset contains 476 records with $\eta \geq 0.90$. Reduction to the maximum-recovery record for each process-train label produced 100 representative records spanning 20.789–98.631% recovery. The records represent computational desalination scenarios rather than independently measured experimental observations.

Model / analysis	Number of records	R^2	RMSE (kWh m^{-3})
Pooled linear	1,316	0.036	17.024
Pooled quadratic	1,316	0.241	15.107
Pooled cubic	1,316	0.465	12.686
Pooled $\Phi(\eta;10)$	1,316	0.005	17.297
High-recovery subset, $\eta \geq 0.90$	476	—	—
Maximum-recovery record per process train	100	—	—

Table S5. 24-Observation Cross-Domain Dataset (grouped as S1-S7).

No.	Domain	Recovery, η	Φ(η;n)
1	Solvent recovery	0.70	0.094158
2	Solvent recovery	0.85	1.312496
3	Solvent recovery	0.90	3.486784
4	Solvent recovery	0.95	11.974739
5	Battery recycling	0.70	0.094158
6	Battery recycling	0.90	3.486784
7	Battery recycling	0.95	11.974739
8	Battery recycling	0.98	40.853640
9	Metal recovery	0.75	0.225254
10	Metal recovery	0.82	0.763600
11	Metal recovery	0.88	2.320841

12	Metal recovery	0.91	4.326846
13	Metal recovery	0.67	0.055238
14	Metal recovery	0.95	11.974739
15	Carbon capture	0.60	0.015117
16	Carbon capture	0.75	0.225254
17	Carbon capture	0.90	3.486784
18	Carbon capture	0.50	0.001953
19	Carbon capture	0.70	0.094158
20	Carbon capture	0.90	3.486784
21	Desalination	0.30	0.000008
22	Desalination	0.50	0.001953
23	Desalination	0.70	0.094158
24	Desalination	0.85	1.312496

Coordinate:

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

The pharmaceutical nonlinear fit reaches the imposed upper boundary n = 10. Therefore, this cross-domain table is described as a compact coordinate-level transferability check, not as evidence for a universal exponent or universal physical law.

Table S6. Independent External REE Recovery Dataset. Independent external rare-earth-element recovery dataset used for qualitative external check. Recovery values were compiled from Peelman et al. (2018) and related recovery/separation examples reported in the supplied external-validation dataset. Because comparable energy-response variables and harmonized system boundaries are not available for these observations, the dataset is not included in the 24-observation quantitative cross-domain scaling analysis. It is used only as a qualitative external check stream to assess whether recovery ranges encountered in an independent resource-recovery domain are broadly compatible with the recovery regimes considered in the proposed framework.

Material/process	Recovery (%)	Domain
Y	70–100	Heavy rare earth
Dy	70–100	Heavy rare earth
Ho	70–100	Heavy rare earth
Er	70–100	Heavy rare earth
Tm	70–100	Heavy rare earth
Yb	70–100	Heavy rare earth
Lu	70–100	Heavy rare earth
Phosphorus	>99	Recovery/separation
Calcium	65	Recovery/separation
REE solvent extraction	50	Solvent extraction
Nd hydrometallurgical recovery	91	Hydrometallurgy
Nd leaching	99.2	Leaching
REE precipitation	92	Precipitation
Pyro-hydrometallurgical Nd	90	Pyro-hydrometallurgy
REE transfer to slag	>95	Pyrometallurgy
Fe recovery	>90	Metal recovery
Fe–Nd separation	>99	Separation

Table S7. Equation and Model-Definition Summary. Definitions of variables, model parameters and framework-specific quantities. The table consolidates the notation used throughout the manuscript. cPMI and IC are operationalized quantities within the present framework and are not presented as universally standardized thermodynamic state variables. β and λ are system- and response-dependent coefficients and should not be interpreted as universal solvent-independent constants.

Symbol / expression	Definition	Interpretation
M	Material/solvent mass intensity	Process material burden
η	Recovery fraction, $0 < \eta < 1$	Fraction recovered
R	Recovery fraction	Used in cPMI
cPMI	PMI × (1 − R)	Operationalized circular PMI metric
$\Phi(\eta;n)$	$\eta^n/(1-\eta)$	Nonlinear recovery coordinate
L*	Normalized response – the model: $L^* = c^* + \alpha M^* + \lambda\Phi(\eta;n)$	Energy or selected burden metric
c*	Intercept c/L_0	Baseline response term
k	Material coefficient	Material-associated response contribution
β	Recovery-coordinate coefficient	System-specific nonlinear contribution
γ	Interaction coefficient	Material–recovery interaction
n	Recovery exponent	Shape parameter of the candidate coordinate
IC	$\Sigma_i f_i^2$	Burden-concentration index
f_i	L_i/L_total	Fractional contribution of operation i
L	L/L_0^*	Dimensionless response
M*	M/M_0^*	Dimensionless material intensity
α	kM_0/L_0	Dimensionless material coefficient
λ	β/L_0	Dimensionless recovery-coordinate coefficient

The source material explicitly describes this dataset as an independent rare-earth-element recovery benchmark dataset compiled from Peelman et al. (2018) and states that it was not used in model development, making it appropriate as an external qualitative assessment.

The individual source entries and values are also explicitly reported: heavy REEs 70–100%, phosphorus >99%, calcium 65%, REE solvent extraction 50%, Nd recovery 91%, Nd leaching 99.2%, REE precipitation 92%, pyro-hydrometallurgical Nd recovery 90% with ~700 kWh t^{-1}, REE transfer to slag >95%, Fe recovery >90%, and Fe–Nd separation >99%.

Table S8. Comparison of the Three Candidate Formulations. Interpretation and intended use of the candidate material–energy formulations.

Formulation	Principal purpose	Interpretation in the present study	Limitation
Linear material-throughput model, $E_i = c + kM_i + \varepsilon_i$	Parsimonious baseline	Describes the dominant material-associated contribution in the seven-record pharmaceutical dataset	Does not explicitly represent recovery-dependent curvature
Nonlinear recovery-coordinate model, $E_i = c + kM_i + \beta\Phi(\eta_i;n) + \varepsilon_i$	Test recovery-associated curvature	Hypothesis-driven phenomenological extension of the linear baseline	Small R^2 improvement; AIC/BIC increase; n reaches imposed upper bound; weak identification
Extended material–recovery interaction model, $E_i = c + kM_i + \beta\Phi(\eta_i;n) + \gamma M_i\eta_i + \varepsilon_i$	Examine material–recovery coupling	Provides the closest in-sample representation of the seven observations	Five fitted parameters for seven observations; limited inferential and predictive value

Note: The three formulations are compared to determine how different representations account for the observed data. The results do not establish a universally preferred model, universal parameter values, or universal recovery behaviour.

Supplementary Note: Datasets

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 1,316 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 analysis scripts, source documentation and dataset-construction information should be archived with the final submission to permit independent reproduction. The cross-domain files supplied with this study – grouped separately as Dataset S1-S7 – including the desalination and rare-earth-element recovery datasets used for transferability analysis and qualitative external checking.

Supplementary Note: Scope, applicability and limitations of the proposed framework

The proposed phenomenological material–energy scaling framework provides a system-specific mathematical representation of recovery-constrained sustainability behaviour. It is not intended as a first-principles thermodynamic derivation of solvent-recovery energetics or as a universal physical law. The quantitative contribution of the recovery-associated term depends on solvent properties, solvent–mixture characteristics, separation technology, process configuration, operating conditions, energy source, response variable, normalization procedure, and system boundary. Consequently, the coefficients β, λ, and γ should be estimated for the specific system and response being analysed and should not be interpreted as universal solvent-independent constants.

The nonlinear recovery coordinate, $\Phi(\eta;n) = \eta^n/(1 - \eta)$, provides a convenient phenomenological coordinate for representing recovery-dependent curvature. Its divergence as η approaches unity is a mathematical property of the selected functional form and should not be interpreted as an experimentally established universal divergence of energy demand. Established separation-science considerations indicate that increasingly high recovery or purity can become progressively more demanding, but the magnitude and functional form of this behaviour remain process-specific.

The seven-record pharmaceutical dataset provides a case-study demonstration rather than sufficient evidence for universal parameter estimation. In particular, the nonlinear formulation provides only a marginal improvement in R^2 relative to the linear model, while AIC and BIC increase and the fitted exponent reaches the imposed upper bound of $n = 10$. The extended interaction model provides a closer in-sample fit but contains five fitted parameters for seven observations. These characteristics limit inference and do not establish independent predictive validity. The step-resolved analysis provides complementary information on where material and environmental burdens are concentrated within the investigated process configuration. The Impact Concentration Index is therefore used as a descriptive process-localization metric and is not interpreted as a causal recovery parameter or universal threshold. The solvent-level descriptor analysis and Random Forest model are exploratory because the available solvent dataset is small and the associated life-cycle boundaries are not fully harmonized. Similarly, the desalination dataset is treated as a computational scenario dataset rather than as a collection of independently measured experimental observations. Its purpose is to examine process heterogeneity and coordinate-level transferability rather than to provide direct physical validation of pharmaceutical solvent recovery.

The 24-observation cross-domain analysis is likewise a compact transferability check. The pharmaceutical recovery exponent is retained as a fixed coordinate for comparison rather than being refitted to establish a new universal exponent. Differences in response units, process mechanisms, data provenance, and system boundaries prevent interpretation of the cross-domain observations as a universal master curve. The independent rare-earth-element dataset provides a qualitative external check because

comparable energy-response variables and harmonized system boundaries are not available. It is therefore not incorporated into the quantitative model fitting. Overall, the framework should be regarded as a testable and reproducible phenomenological formulation for examining material–energy trade-offs under defined recovery constraints. Broader predictive applicability requires independently documented datasets with larger sample sizes, wider recovery ranges, consistent system boundaries, traceable process inventories, and independent validation across solvents, separation technologies, and process configurations.

References

1. Shaikh, I. R.; Shaikh, M. M. 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.
2. Sheldon, R. A. The E factor 25 years on: the rise of green chemistry and sustainability. Green Chem. 19, 18–43 (2017).
3. Byrne, F. P., Jin, S., Paggiola, G., Petchey, T. H. M., Clark, J. H., Farmer, T. J., Hunt, A. J., McElroy, C. R. & Sherwood, J. Tools and techniques for solvent selection: green solvent selection guides. Sustain. Chem. Process. 4, 7 (2016).
4. Constable, D. J. C. et al. Key green chemistry research areas – a perspective from pharmaceutical manufacturers. Green Chem. 9, 411–420 (2007).
5. Marchetti, P., Solomon, M. F. J., Szekely, G. & Livingston, A. G. Molecular separation with organic solvent nanofiltration: A critical review. *J. Membr. Sci.* 454, 17–53 (2014).
6. Harper, G., Sommerville, R., Kendrick, E., Driscoll, L., Slater, P., Stolkin, R., Walton, A., Christensen, P., Heidrich, O., Lambert, S., Abbott, A., Ryder, K., Gaines, L. & Anderson, P. Recycling lithium-ion batteries from electric vehicles. *Nature* 575, 75–86 (2019).
7. Rinne, M., Mäkinen, J., Dahl, O. & Watkins, G. Sustainability assessment of hydrometallurgical recycling technologies for lithium-ion batteries. *Int. J. Life Cycle Assess.* 29, 1157–1174 (2024).
8. 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.* 146, 101338 (2024). https://doi.org/10.1016/j.pmatsci.2024.101338
9. Reuter, M. A., van Schaik, A., Gutzmer, J., Bartie, N. & Abadías-Llamas, A. Challenges of resource efficiency and recycling in complex metal systems. *Miner. Eng.* 52, 3–19 (2013).
10. Reuter, M. A. et al. Challenges of resource efficiency and recycling in complex metal systems. *Miner. Eng.* 160, 106656 (2021).
11. Rochelle, G. T. Amine scrubbing for CO_2 capture. *Science* 325, 1652–1654 (2009).
12. Keith, D. W., Holmes, G., St. Angelo, D. & Heidel, K. A process for capturing CO_2 from the atmosphere. *Joule* 2, 1573–1594 (2018).
13. Elimelech, M. & Phillip, W. A. The future of seawater desalination: Energy, technology, and the environment. *Science* 333, 712–717 (2011).
14. Stillwell, A. S. & Webber, M. E. Technical analysis of a river basin-based model of advanced power plant cooling technologies for mitigating water management challenges. *Water* 8, 601 (2016).
15. Mistry, K. H., McGovern, R. K., Thiel, G. P., Summers, E. K., Zubair, S. M. & Lienhard, J. H. Entropy generation analysis of desalination technologies. *Entropy* 13, 1829–1864 (2011).
16. O'Connell, M. G., Rajendran, N., Elimelech, M., Gilron, J. & Dunn, J. B. Analysis of energy, water, land and cost implications of zero and minimal liquid discharge desalination technologies. *Nat. Water* **2**, 1116–1127 (2024). https://doi.org/10.1038/s44221-024-00327-1
17. Peelman, S., Sun, Z. H. I., Sietsma, J. & Yang, Y. Recycling of rare earths: A critical review. *J. Clean. Prod.* 186, 8–27 (2018).