



Waste-Heat Recovery from Cement Kilns Using a Cascaded Organic Rankine Cycle with Ultra-Low-GWP Working Fluids: A Plant-Scale Techno-Environmental Case Study

Ismael Badawi Dahash¹

¹ University of Al-Qadisiya /College of Education

Abstract

Cement manufacture is among the most energy- and carbon-intensive industrial activities, yet much of its energy input leaves the kiln line as medium- and low-grade waste heat. This paper reports a plant-scale techno-environmental case study of a water-free, air-cooled cascaded dual-loop organic Rankine cycle (ORC) for a cement plant in Mazandaran, Iran, in which a pure-toluene topping loop recovers the suspension-preheater exhaust and an ultra-low-GWP R1233zd(E)/R1234ze(Z) bottoming loop (0.633/0.367 by mass) recovers the clinker-cooler air together with the topping-loop condensation duty. Fifteen candidate fluids are filtered against explicit numerical criteria and the survivors evaluated on the full cascade model under identical boundary conditions; the design is obtained by surrogate-assisted NSGA-II optimisation with entropy-weighted TOPSIS selection and assessed through five lenses — energy, exergy with avoidable/unavoidable and endogenous/exogenous decomposition, economic, environmental and exergoenvironmental — supported by a digital-twin-oriented operational layer. The cascade absorbs 28.60 MW and delivers 4.851 MW of net shaft power (17.0 % thermal, 33.3 % exergy efficiency) at a levelised cost of 0.0434 \$ kWh⁻¹, a 3.6-year discounted payback at an avoided-outage valuation and 22.0 kt yr⁻¹ of avoided grid CO₂. Monte-Carlo propagation places the levelised cost at 0.031–0.071 \$ kWh⁻¹ (P10–P90) and returns a 99.7 % probability of a positive net present value at that valuation against 0.05 % at the subsidised tariff, so bankability is governed by tariff treatment rather than by engineering. The direct refrigerant impact is 0.0011 % of the avoided emission and the 0.61 K zeotropic glide is worth under 1 %; replacing the optimistic loss assumptions lowers net power by 19.0 % without altering any conclusion.

Keywords: organic Rankine cycle; cement waste-heat recovery; cascaded dual-loop; ultra-low-GWP working fluids; exergoenvironmental analysis; digital twin

Keywords: *organic Rankine cycle; cement waste-heat recovery; cascaded dual-loop; ultra-low-GWP working fluids; digital twin*

1. Introduction

Cement production is one of the most energy-intensive industrial processes and, after iron and steel, the second largest industrial source of carbon dioxide [1,2]. A modern dry-process kiln line rejects much of its fuel energy as sensible heat in two streams: the suspension-preheater (SP) exhaust at 340–360 °C and the air-quenching-cooler (AQC) stream from the clinker grate cooler at 270–335 °C [1,2]. For the plant studied here these carry 45.6 MW relative to a 25 °C ambient, of which 32.8 MW is recoverable above realistic stack and dust-collector floors [3].

The organic Rankine cycle (ORC) is the most mature route for low- and medium-grade sources [4–8], and its performance is governed to first order by the working fluid [9,10]. Fluid selection is now constrained as much by regulation as by thermodynamics: the Kigali Amendment and the European Union F-Gas Regulation are phasing down high-global-warming-



This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited. © 2025 The Authors

potential (GWP) refrigerants, while hydrofluoro-olefins (HFOs) and hydrochlorofluoro-olefins (HCFOs) with GWP_{100} in the single digits are now commercial. Where the source spans a wide range — as in a cement plant, which offers a medium- and a low-grade stream simultaneously — a cascaded dual-loop configuration recovers power a single loop would discard in topping-loop condensation, and zeotropic mixtures add a further degree of freedom through their glide [11–15].

Evaluation practice has matured beyond first-law accounting: advanced exergy analysis splits each component's destruction into avoidable/unavoidable and endogenous/exogenous parts [16,17], and exergoenvironmental analysis couples life-cycle assessment to the exergy balance [12,17]. Design is increasingly posed as a multi-objective problem accelerated by surrogates [15,18], while model-based digital twins have spread across the energy sector [19–21] but remain rare in industrial ORC waste-heat recovery.

To the authors' knowledge no published study combines a cascaded dual-loop ORC matched to both cement streams, ultra-low-GWP fluids selected by an explicit quantitative screen, a complete five-aspect (energy–exergy–economic–environmental–exergoenvironmental) assessment with advanced exergy decomposition, and an operational layer able to assimilate measurements, detect degradation and re-optimize set-points. The objectives are therefore to establish by two-stage screening under identical boundary conditions which fluids are admissible for each loop; to optimize the resulting water-free cascade with an auditable decision step; to quantify performance across the five aspects, including economic uncertainty and the penalty attached to each optimistic assumption; and to verify the reduced-order dynamic model and state estimator that would underpin an operational twin.

The closest study is that of Sanaye et al. [22], whose single-loop toluene ORC on the same Mazandaran plant produced 6.56 MW across two production lines with a 3.20-year payback, but used water cooling, no low-GWP fluids and neither advanced-exergy nor exergoenvironmental analysis. Wang et al. [23] found a dual-loop Kalina arrangement best (6.84–7.17 MW) for the same duty but retained a high-pressure ammonia–water cycle and water cooling; Gioia et al. [24] gained about 25 % more power from a partial-evaporation ORC through turbine-level rather than system-level design; and Fergani and Morosuk [17] applied the same advanced-exergy decomposition without a cement case, a cascade or an operational layer.

2. Materials and Methods

This section defines the plant, the cycle, the property and cost models, the five assessment aspects, the optimisation and decision procedure and the operational layer; the workflow is given as Figure S1.

2.1 Case study and heat-source dataset

The case-study plant in Mazandaran, northern Iran, has the two waste-heat streams typical of modern dry-process cement production. This is a model-based case study rather than an analysis of plant measurements: the primary dataset is the set of source conditions defining the recovery boundary (Table 1), consistent with values reported for the same plant [22]. For a conservative, water-free design suited to water-scarce sites, both sources are cooled only to realistic floors — 120 °C on the SP gas, above the acid dew point, and 100 °C on the AQC air.

The specific heats in Table 1 are not free parameters. A representative dry-process preheater composition of 23.5 % CO_2 , 7.0 % H_2O , 6.0 % O_2 and 63.5 % N_2 by volume [1,2] gives a mass-weighted specific heat from the pure-component correlations of the property library [25], enthalpy-averaged over 120–360 °C, of 1080 $\text{J kg}^{-1} \text{K}^{-1}$, varying only between 1072 and 1088 $\text{J kg}^{-1} \text{K}^{-1}$ for ± 5 percentage points of CO_2 and ± 2 of H_2O ; dry air over 100–300 °C gives 1026 $\text{J kg}^{-1} \text{K}^{-1}$. The adopted values of 1100 and 1050 $\text{J kg}^{-1} \text{K}^{-1}$ are 1.9 % and 2.3 % high, and re-solving with the composition-weighted values changes net power by –2.0 % and the levelised cost by +1.1 %.

Table 1. Table 1. Heat-source operating data defining the recovery boundary (plant case-study dataset). Specific heats are composition-weighted, enthalpy-averaged values; see §2.1.

Source	Medium	$\dot{m}$ (kg s^{-1})	T_{in} (°C)	$T_{\text{out floor}}$ (°C)	c_p ($\text{J kg}^{-1} \text{K}^{-1}$)	Q to floor (MW)	Q to ambient, 25 °C (MW)
SP (preheater exhaust)	flue gas	90.0	360	120	1100	23.76	33.16
AQC (clinker cooler air)	air	42.91	300	100	1050	9.01	12.39
Total						32.77	45.56

2.2 Configuration

The plant comprises two thermally coupled Rankine loops (Figure 1). A high-temperature (HT) topping loop of pure toluene recovers the SP gas and expands the superheated vapour through the HT turbine; a cascade heat exchanger (CHE) passes part of the turbine-exhaust and condensation duty to the bottoming loop while an air-cooled condenser rejects the remainder. The low-temperature (LT) bottoming loop, charged with the R1233zd(E)/R1234ze(Z) blend, is preheated by the CHE, evaporated by the AQC air and expanded through the LT turbine before condensing against ambient air, so the plant is entirely water-free.

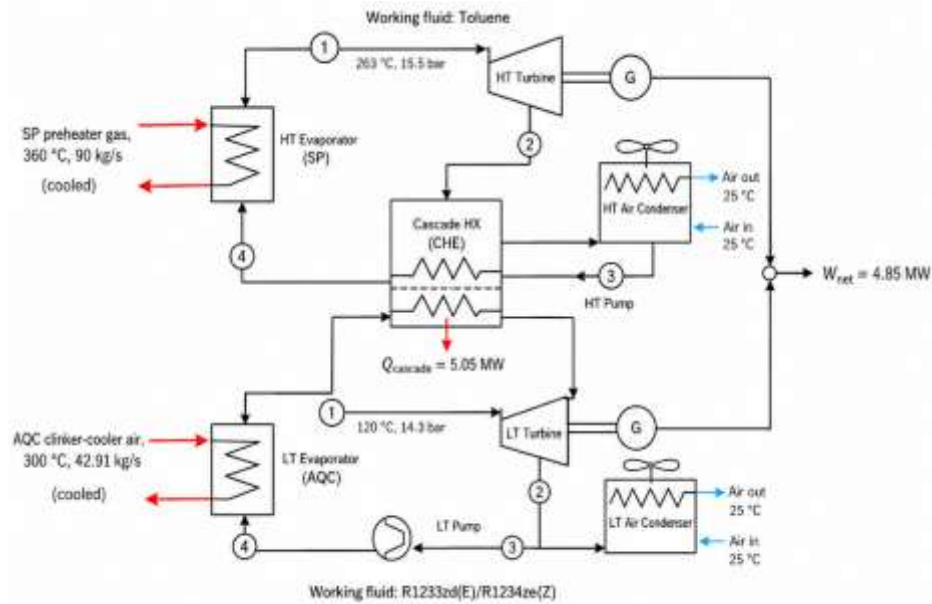


Fig.1: Process flow diagram of the cascaded dual-loop ORC (pure-toluene HT topping loop and R1233zd(E)/R1234ze(Z) LT bottoming loop, linked by a cascade heat exchanger; water-free air-cooled condensation).

2.3 Working-fluid screening and selection

Working-fluid selection is a two-stage, criterion-driven procedure rather than an assumption: stage 1 filters all fifteen candidates against explicit numerical criteria, and stage 2 evaluates every survivor on the complete cascade model under identical boundary conditions.

The stage-1 criteria, applied to Table S13, are: (C1) a Helmholtz-energy equation of state in the open property library [25], without which no consistent mixture flash is possible; (C2) $GWP_{100} \leq 150$, the European Union F-Gas threshold; (C3) ozone-depletion potential $\leq 1 \times 10^{-3}$; (C4) for the HT loop a critical temperature of 230–360 °C, so that subcritical evaporation is possible against a 360 °C source while the condensing pressure at the cascade interface stays above a deep vacuum; (C5) for the LT loop a critical temperature of 120–200 °C; (C6) a condensing pressure of at least 0.5 bar at the LT condensing temperature; and (C7) for the HT loop a thermal-stability limit of at least 260 °C [32,33]. Safety class qualifies rather than eliminates.

Stage 1 removes four candidates and explains the exclusions raised in the literature. R1336mzz(Z) and R1224yd(Z) are attractive on GWP and safety grounds but have no Helmholtz-energy equation of state in the open library, so they cannot be evaluated on the same footing and are the highest-priority additions to any future study with a reference-fluid database. R245fa fails C2 with $GWP_{100} = 858$ and is kept only as a legacy benchmark, while R1234ze(E), R1234yf and R290 fail C5 and cyclopentane and R601a survive C4 only marginally, so none can evaporate at the design temperature of its loop. Cyclohexane and n-decane pass every criterion and are carried into stage 2 rather than excluded a priori.

Stage 2 evaluates the survivors on the full cascade model with every other boundary condition held at the converged design — the same sources, sink, evaporation and condensing temperatures, superheat, pinch and approach differences and isentropic efficiencies (Table 2, §3.2).

The bottoming-loop charge is R1233zd(E)/R1234ze(Z) at 0.633/0.367 by mass, the value returned by the optimisation of §2.8 with the mass fraction treated as a continuous variable over [0.25, 0.75]. No fitted binary interaction parameter exists for this pair, so an estimated (linear) mixing rule is used and the resulting glide is cross-validated against an ideal-solution calculation in §3.9.

Table 2. Stage-2 like-for-like evaluation of the shortlisted fluids at identical boundary conditions.

Fluid	Class	GWP_{100}	ODP	Safety	T_{crit} (°C)	NBP (°C)	Loop	Status
Toluene	HC	3	0	flammable	318.6	110.6	HT	selected (HT)
R1233zd(E)	HCFO	1	4×10^{-4}	A1	166.5	18.3	HT/LT	selected (LT)
R1234ze(Z)	HFO	0.3	0	A2L	150.1	9.7	LT	selected (LT)
Cyclohexane	HC	1	0	flammable	280.5	80.7	HT	screened
Cyclopentane	HC	11	0	flammable	238.6	49.2	HT	screened
Benzene	HC	3	0	flammable, toxic	288.9	80.1	HT	screened (safety flag)
n-Decane	HC	3	0	flammable	344.5	174.1	HT	screened
R601a (isopentane)	HC	4	0	A3	187.2	27.8	HT	screened
R245fa	HFC	858	0	B1	153.9	15.0	HT	legacy baseline
R1234ze(E)	HFO	0.3	0	A2L	109.4	-19.0	LT	screened
R1234yf	HFO	0.5	0	A2L	94.7	-29.5	LT	screened
R600a (isobutane)	HC	3	0	A3	134.7	-11.7	LT	screened

R290 (propane)	HC	3	0	A3	96.7	-42.1	LT	screened
R1336mzz(Z)	HFO	2	0	A1	171.3	33.5	HT	reference-fluid only
R1224yd(Z)	HCFO	0.88	1.2×10^{-4}	A1	155.5	14.6	LT	reference-fluid only

All cases solved on the full cascade model with the same sources, sink, condensing temperatures, superheat, pinch and approach differences and isentropic efficiencies as the reported design; only the fluid of the loop under test changes. “Infeasible” means the required saturation state does not exist, not that the result is poor.

2.4 Thermodynamic model and assumptions

Each component is modelled as a steady-state open control volume with negligible kinetic and potential energy changes, adiabatic boundaries except at the specified heat exchangers, and negligible line pressure losses. Mass and energy conservation applied to component k gives

$$\sum_i \dot{m}_{i,in} = \sum_i \dot{m}_{i,out}, \dot{Q}_k - \dot{W}_k = \sum_{out} \dot{m} h - \sum_{in} \dot{m} h. \quad (1)$$

The net power and the two energy-efficiency metrics follow as $\dot{W}_{net} = \sum_t \dot{W}_t - \sum_p \dot{W}_p - \dot{W}_{fan}$, with the thermal efficiency $\eta_{th} = \dot{W}_{net} / \dot{Q}_{in}$ referred to the absorbed heat and the waste-heat utilization $\eta_{WHR} = \dot{W}_{net} / \dot{E}_{avail}$ referred to the heat available to ambient.

Heat exchangers are discretised into 40 counter-current cells; the minimum local temperature difference is constrained to the specified pinch and the conductance accumulated as $UA = \sum \dot{Q}_i / \Delta T_{lm,i}$, from which the costing area follows as $A = UA/U$ with the coefficients of Table 3. The cascade is closed by enforcing the CHE energy coupling with the pinch and approach constraints, solved by a damped fixed-point iteration on the CHE cold-side outlet enthalpy. Properties come from Helmholtz-energy equations of state for both the pure fluids and the mixture [25].

Four assumptions in Table 3 are optimistic and are quantified rather than asserted in §3.7: the condenser-fan parasitic taken as 1.5 % of gross turbine power, the absence of a generator and mechanical efficiency, frictionless piping and adiabatic casings. Each is replaced by a conservative alternative — a fan power $\dot{m}_{air} \Delta p / (\rho_{air} \eta_{fan})$ with $\Delta p = 150$ Pa and $\eta_{fan} = 0.65$; a generator-plus-mechanical efficiency of 0.96×0.98 ; pressure drops of 3 %, 5 % and 3 % across the superheater, condenser and preheater lines; and a 2 % casing heat loss — giving the penalty ladder of Table 8.

Table 3. Principal modelling assumptions and techno-economic/environmental parameters.

Category	Parameter	Value
Reference (dead) state	T_0, P_0	25 °C, 1.013 bar
Turbomachinery	Turbine isentropic efficiency	0.85
	Pump isentropic efficiency	0.75
Heat exchangers	Evaporator pinch ΔT	design variable (optimized)
	Air-condenser approach	10 K
Parasitics	Condenser-fan power	1.5 % of gross turbine power
	Generator/mechanical efficiency	not applied (net-shaft basis)
Economic	Costing method	Turton module factors (CEPCI-updated)
	Plant life, n	20 yr
	Discount/interest rate, i	10 %
	O&M (fraction of CapEx)	≈ 2 % yr ⁻¹
	Capacity factor	0.85 (≈ 7450 full-load h yr ⁻¹)
Environmental	Electricity tariff scenarios	0.015 / 0.060 / 0.120 \$ kWh ⁻¹
	Grid emission factor	610 kg CO ₂ MWh ⁻¹
	LCA basis	cradle-to-gate material CO ₂ -eq factors

Turbomachinery efficiencies, interest rate and O&M fraction are the values adopted in the analysis; confirm against the as-implemented configuration if they differ.

2.5 Exergy and advanced exergy analysis

The second-law analysis follows the specific-exergy-costing (SPECO) fuel–product formalism. The physical exergy of a stream, the component-level destruction and the system exergy efficiency are

$$e_{ph} = (h - h_0) - T_0(s - s_0), \dot{E}_{D,k} = \dot{E}_{F,k} - \dot{E}_{P,k}, \varepsilon_{sys} = \frac{\dot{W}_{net}}{\dot{E}_{F,sys}}, \quad (2)$$

where $\dot{E}_{F,k}$ and $\dot{E}_{P,k}$ are the exergy of fuel and product defined consistently for each component (e.g., the enthalpy-exergy drop of the hot stream as fuel and the gain of the cold stream as product for a heat exchanger; the shaft work as product for a turbine). To move beyond a simple ranking of irreversibilities, the destruction in each component is decomposed into avoidable and unavoidable parts, and independently into endogenous and exogenous parts,

$$\dot{E}_{D,k} = \dot{E}_{D,k}^{AV} + \dot{E}_{D,k}^{UN} = \dot{E}_{D,k}^{EN} + \dot{E}_{D,k}^{EX} \quad (3)$$

The unavoidable part is obtained by re-evaluating each component under near-ideal limiting conditions (minimum approach temperatures and best-practice isentropic efficiencies), so that the avoidable part isolates the realistically recoverable irreversibility; the endogenous part is the destruction generated by the component operating with all other components ideal, the remainder being exogenous and attributable to component interactions. This decomposition identifies not merely where exergy is destroyed but where redesign can usefully reduce it.

For a pump the shaft work is the fuel and the exergy rise of the liquid the product; $\dot{E}_{F,\text{sys}}$ is the exergy of the two source streams relative to the dead state at $T_0 = 298.15 \text{ K}$ and $P_0 = 1.013 \text{ bar}$. The near-ideal conditions used to isolate the unavoidable destruction are approach differences of 2 K and isentropic efficiencies of 0.92 (turbines) and 0.88 (pumps).

2.6 Economic analysis and uncertainty propagation

Capital cost is estimated component-by-component using the Turton module-costing method, with purchased-equipment costs sized on heat-transfer area (heat exchangers) or power (turbines and pumps), corrected to the analysis year through the Chemical Engineering Plant Cost Index and aggregated to the total module cost Z_{tot} . The levelized cost of electricity is then

$$\text{LCOE} = \frac{\text{CRF} \cdot Z_{\text{tot}} + \dot{C}_{\text{om}}}{\dot{W}_{\text{net}} t_{\text{op}}}, \text{CRF} = \frac{i(1+i)^n}{(1+i)^n - 1} \quad (4)$$

where $\dot{C}_{\text{om}}$ is the annual operation-and-maintenance cost, t_{op} the annual operating hours, CRF the capital-recovery factor, i the discount rate, and n the plant life. Because LCOE is independent of the sale price of electricity, project-level viability (net present value, internal rate of return, and discounted payback) is evaluated separately under three explicit electricity-tariff scenarios — a deeply subsidized industrial tariff, an avoided-grid/export valuation, and an avoided-outage valuation reflecting the reliability value of self-generation to a continuous-process plant — to address the wide spread of credible tariffs in the Iranian context.

Numerically, $C_{\text{om}} = 3\%$ of Z_{tot} per year, $t_{\text{op}} = 7446 \text{ h yr}^{-1}$, $i = 8\%$ and $n = 20 \text{ yr}$; purchased-equipment costs follow the Turton correlations [26] with multi-shell splitting above 1000 m^2 per shell, a Chemical Engineering Plant Cost Index of 800 and a 1.18 contingency-and-fee factor.

Project-level viability follows from NPV = $-Z_{\text{tot}} + \sum_{t=1}^n (c_e E - C_{\text{om}})(1+i)^{-t}$, (6) with c_e the value of the recovered electricity and E the annual generation.

A single deterministic result is not sufficient for a plant-scale study, so the economics are also propagated probabilistically. Six inputs are sampled 20 000 times by Monte Carlo: the error of the cost correlation itself (lognormal, median 1, $\sigma_{\ln} = 0.30$, the $\pm 30\%$ accuracy quoted for a factored study estimate [26]); the CEPCI (triangular 760/800/860); the discount rate (6/8/14 %); the capacity factor (0.72/0.85/0.93); the O&M fraction (2.0/3.0/4.5 %); and the plant life (15/20/25 yr). The electricity value carries a further $\pm 25\%$ triangular factor.

2.7 Environmental and exergoenvironmental analysis

The environmental assessment combines an operational (use-phase) credit with a life-cycle (construction-phase) burden. The operational benefit is the avoided grid emission, $\dot{m}_{\text{CO}_2,\text{av}} = \varepsilon_{\text{grid}} \dot{W}_{\text{net}}$, evaluated at the grid emission factor of Table 3; the direct refrigerant contribution is quantified through the total-equivalent-warming-impact (TEWI), combining the direct charge-leakage term with the indirect operational emissions. The embodied burden of each component is obtained from a cradle-to-gate life-cycle inventory of its material mass, expressed as a component environmental rate $\dot{Y}_k$ (kg CO₂-eq per unit time). These are integrated in the exergoenvironmental balance, in which an environmental impact is assigned to every exergy stream and the product impact of each component equals the fuel impact plus the component's own embodied rate,

$$\dot{B}_{P,k} = \dot{B}_{F,k} + \dot{Y}_k f_{b,k} = \frac{\dot{Y}_k}{\dot{Y}_k + \dot{B}_{D,k}} \quad (5)$$

The exergoenvironmental factor $f_{b,k} \in [0,1]$ partitions each component's environmental impact between its construction ($\dot{Y}_k$) and its exergy destruction ($\dot{B}_{D,k}$), thereby indicating whether improvement effort should target component manufacture or thermodynamic irreversibility. Because the waste-heat fuel carries an essentially zero specific CO₂ burden, this analysis yields the embodied environmental intensity of the delivered electricity, which is contrasted with the grid intensity it displaces. TEWI uses an annual leakage fraction of 5 %, an end-of-life recovery fraction of 85 % and the global warming potentials of Table S13; the life-cycle inventory uses cradle-to-gate CO₂-equivalent factors on the computed material mass of every component..

2.8 Surrogate-assisted optimisation and decision-making

The design was obtained by surrogate-assisted multi-objective optimisation with an explicit decision step, in five stages (Algorithm S1).

Design space and sampling. Ten continuous variables are optimised: HT evaporation temperature (180–250 °C), HT superheat (0–20 K), HT condensing temperature (95–120 °C), LT evaporation temperature (60–110 °C), LT superheat (0–15 K), LT condensing temperature (41–50 °C), the R1233zd(E) mass fraction (0.25–0.75) and the evaporator, condenser and cascade-exchanger pinches (5–25, 5–12 and 5–25 K). A 72-point Latin-hypercube design spans this box, each point evaluated on the full cascade solver.

Surrogate construction. A Gaussian-process model with a squared-exponential kernel with automatic relevance determination is fitted to each target, and a binary classifier predicts feasibility. Five-fold cross-validation gives $R^2 = 0.998$ for η_{th} , 0.992 for $\dot{W}_{net}$, ε_{sys} and avoided CO_2 , 0.925 for the LCOE and 0.864 for the specific cost (Table S2). This quantifies only fidelity to the underlying physical model; that model is validated separately in §2.10.

Multi-objective search. NSGA-II [27] is run on the surrogate with a population of 200 over 500 generations over three objective pairings: maximise ε_{sys} against minimise LCOE (the reported front), ε_{sys} against specific cost, and $\dot{W}_{net}$ against LCOE. Environmental performance is not a further objective, not through neglect but through redundancy: the avoided emission is exactly proportional to net power, so the net life-cycle CO_2 rate is an affine, strictly decreasing function of $\dot{W}_{net}$. This is verified in §3.4 with a genuine three-objective search.

Decision step. The non-dominated set is ranked by TOPSIS [28] with Shannon-entropy weights. TOPSIS needs no preference information beyond the criteria themselves, ranks alternatives by distance to both the ideal and the anti-ideal point, and is the method most used for ORC design selection, which makes the result comparable with the literature; entropy weighting derives the weights from criterion dispersion. Because any such rule embeds a preference, the selection is tested in §3.4 against equal and CRITIC weights [29], min–max normalisation, simple additive weighting, LINMAP and VIKOR [30], and a continuous weight sweep.

Verification loop. Because the cost surrogate is the least accurate, the ranked front is traversed in order and every candidate re-solved on the full physical model; the first that passes the pinch, pressure and vapour-quality limits is adopted. This prevents an optimistic surrogate artefact from being reported as an optimum.

2.9 Operational (digital-twin-oriented) layer

The operational layer gives the optimised design an online, model-based capability. It is described as a digital-twin-oriented framework rather than a deployed digital twin: it implements the full assimilate–detect–re-optimize loop on a physical model, but has not been calibrated against measurements from the operating kiln line.

Reduced-order model. The foundation is a two-zone moving-boundary model of the source evaporator, in which the sub-cooled and two-phase zones are lumped energy balances and the phase-boundary position is tracked by an ordinary differential equation for the two-phase length fraction, with a lumped wall thermal mass [31]. The gas- and refrigerant-side conductances follow from the steady design solve; the wall thermal mass does not, and governs the slow time scale.

State estimation. A discrete state-space realisation of this model is coupled to a Kalman filter that assimilates a noisy condensing-temperature measurement, propagating the state estimate by the standard predict–update recursion $\hat{x}_t^- = A \hat{x}_{t-1} + B u_{t-1}$ followed by the measurement correction $\hat{x}_t = \hat{x}_t^- + K_t (z_t - C \hat{x}_t^-)$, (7) in which $K_t = P_t^- C^T (C P_t^- C^T + R)^{-1}$ is the Kalman gain formed from the propagated error covariance P_t^- and the measurement-noise covariance R .

Fault detection and re-optimisation. A residual-based threshold test flags degradation when the estimated condensing temperature exceeds its nominal value by a margin, and the set-point is then re-optimised on the physical model. The margin trades detection delay against false-alarm rate, both quantified in §3.10 (Algorithm S2).

The assumed measurement model is a four-wire Pt100 class-A sensor on the LT condensing line with $\sigma = 0.8$ K sampled every 60 s, with class-1 thermocouples on the source streams, a 0.1 %-FS pressure transmitter and a 0.2 %-of-rate Coriolis meter (Table S6); the process-noise variance is $0.02 \text{ K}^2 \text{ s}^{-1}$.

Three things are established quantitatively: that the two-zone representation reproduces the duty of the 40-cell distributed exchanger at the same conductance; that the Kalman filter is statistically consistent and reduces the measurement error by a known factor; and that detection delay and false-alarm rate follow a known trade-off with the threshold. What is not established is agreement with plant transients, because no such dataset exists for this site; the commissioning protocol is set out in §4.

2.10 Computational framework and validation of the thermodynamic model

The framework is implemented in MATLAB®, with properties from CoolProp through Helmholtz-energy equations of state [25]. The cascade solver, the five-aspect analyses, the optimiser and the operational loop form a single pipeline in which plant and property data feed the thermodynamic model, which drives the five parallel analyses, while the optimisation and operational layers close the loop between design and operation.

Cross-validating the surrogate demonstrates only that it reproduces the underlying cascade model, so the physical model is validated separately in four layers. (1) Property layer: critical points and normal boiling points are compared with the published reference values of the source equations of state. (2) Conservation layer: the cascade must close the first law to numerical precision, the SPECO balance must satisfy $\Sigma(\dot{E}_F - \dot{E}_P - \dot{E}_D) = 0$ component by component, all stream impacts must be non-negative with $f_b \in [0,1]$, and the thermal efficiency must lie below both the Carnot and the tighter finite-source Lorenz bound. (3) Component layer: published single-loop operating points are reproduced and the turbine, pump and heat-duty terms independently recomputed from raw property calls. (4) Plant layer: the cascade is re-solved over a range of source throughputs and compared with the published result for the same plant [22]. Outcomes are summarised in §3.11 and tabulated as Table S1; no experimental data exist for this plant and none is claimed.

Table 5. Four-layer verification and validation of the thermodynamic model.

3. Results and Discussion

3.1 Cascade architecture and heat-source matching

The optimised plant consists of a pure-toluene topping loop and an R1233zd(E)/R1234ze(Z) bottoming loop (0.633/0.367 by mass) connected by a direct working-fluid cascade heat exchanger. The topping loop recovers the SP gas (90 kg s⁻¹, 360 → 148.6 °C) and the bottoming loop is driven jointly by the topping-condenser heat delivered through the CHE and by the AQC air (42.91 kg s⁻¹, 300 → 137.9 °C). The design absorbs 28.60 MW and generates 4.851 MW of net power with toluene and blend mass flows of 33.72 and 52.63 kg s⁻¹.

The four exchanger profiles are shown in Figure 2 and discussed in turn below.

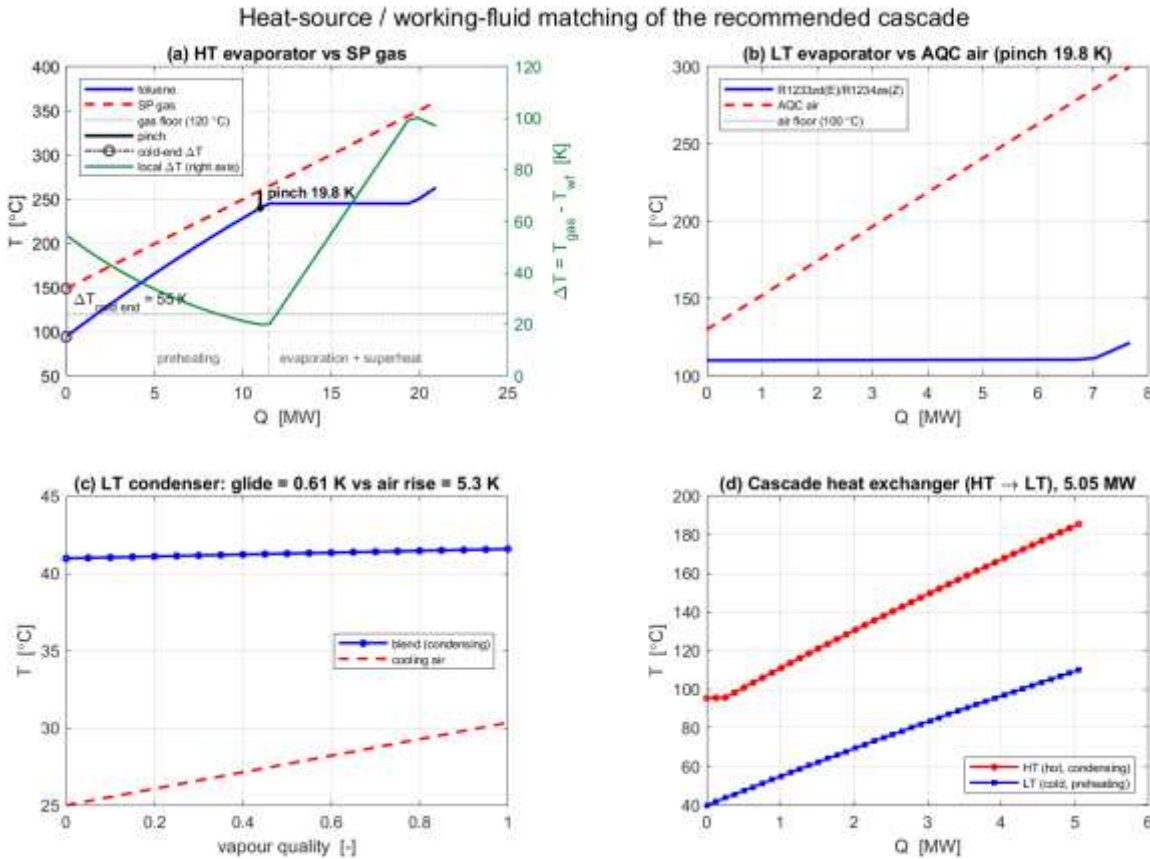


Fig. 2: Heat-source/working-fluid matching: (a) HT evaporator against the SP gas, with the preheating and evaporation zones marked, the pinch annotated, the 120 °C acid dew-point floor drawn and the local driving temperature difference on the right-hand axis; (b) LT evaporator against the AQC air; (c) LT condenser, comparing the zeotropic glide with the cooling-air rise; (d) cascade heat exchanger.

Figure 2(a) shows the HT evaporator against the SP gas. The toluene line rises from 94.0 °C to the 244.9 °C bubble point over the first 10.99 MW of duty, runs almost isothermally through evaporation and rises again through 18.1 K of superheat to the 263.0 °C turbine inlet, while the gas falls from 360 °C to 148.6 °C. The local driving temperature difference, on the right-hand axis, is a minimum of 19.8 K exactly at the onset of evaporation, 97 K across the two-phase plateau, and — the feature the reader notices first — widens monotonically to 54.6 K at the cold end.

That cold-end gap is neither an artefact nor a design error but the geometric consequence of a subcritical cycle recovering heat from a sensible source. The pinch is internal, at the bubble point, because the heat-capacity rate of the liquid toluene (74 kW K⁻¹) is smaller than that of the gas (99 kW K⁻¹), so the working-fluid line is steeper and the two diverge as the duty decreases. The same geometry fixes how far the gas can be cooled: the evaporator duty is set by the toluene mass flow, which is set by the pinch, so the gas leaves at 148.6 °C rather than at the 120 °C floor. The unrecovered tail is 2.83 MW of enthalpy — 11.9 % of the SP heat available above the floor — but only 0.76 MW of exergy; converting it at the present exergy efficiency would add about 0.25 MW, or 5.2 % of net output. This is the largest identified improvement opportunity (§4) and the physical origin of the 1.17 MW destroyed in the HT evaporator, which the advanced decomposition classifies as entirely unavoidable.

In the LT evaporator (Figure 2b) the mismatch runs the other way: the blend evaporates almost isothermally at 110 °C while the air falls from 300 °C, so the exchanger is compact (1134 m² in two shells) but thermodynamically costly — the origin of the 1.22 MW destroyed there. Figure 2(c) compares the condensation glide with the cooling-air rise in the LT condenser, 0.61 K against 5.3 K, the quantitative reason the mixture cannot materially improve the match (§3.9), and Figure 2(d) shows the CHE, in which the HT stream is condensed from 185.6 to 95.2 °C while the LT liquid is preheated from 41.0 to its 109.9 °C bubble point over a 5.05 MW duty.

The closed-cycle state points are listed in Table 4: the toluene loop expands from 263.0 °C/15.52 bar to 185.6 °C/0.641 bar, the residual 92 K of superheat confirming dry expansion, and the blend loop from 120.3 °C/14.30 bar to 68.2 °C/2.535 bar.

Table 4. State-point summary (HT: toluene, 33.72 kg s⁻¹; LT: R1233zd(E)/R1234ze(Z), 52.63 kg s⁻¹). State numbering follows Figure 3.

State	<i>T</i> (°C)	<i>P</i> (bar)	<i>h</i> (kJ kg ⁻¹)	<i>s</i> (J kg ⁻¹ K ⁻¹)	Loop
1 condenser-out	93.2	0.641	-34.25	-91.2	HT
2 pump-out	94.0	15.52	-31.76	-89.5	HT
3 turbine-in	263.0	15.52	588.88	1219.8	HT
4 turbine-out	185.6	0.641	483.14	1271.1	HT
1 condenser-out	39.0	2.535	248.32	1210.4	LT
2 pump-out	39.9	14.30	249.67	1211.6	LT
X CHE-out (→ AQC)	109.9	14.30	345.69	1487.6	LT
3 turbine-in	120.3	14.30	491.40	1867.2	LT
4 turbine-out	68.2	2.535	462.59	1888.6	LT

The temperature–entropy diagrams (Figure 3) place the saturation domes and process paths in context. In both loops the pump process 1 → 2 is so nearly isentropic that states 1 and 2 are indistinguishable at the scale of the main axes, so each panel carries an inset that resolves them: the toluene pump raises the pressure from 0.641 to 15.52 bar with an entropy rise of only 1.7 J kg⁻¹ K⁻¹, the blend pump from 2.535 to 14.30 bar with 1.2 J kg⁻¹ K⁻¹. The toluene cycle expands entirely in the superheated region, so there is no droplet-erosion risk in the HT turbine.

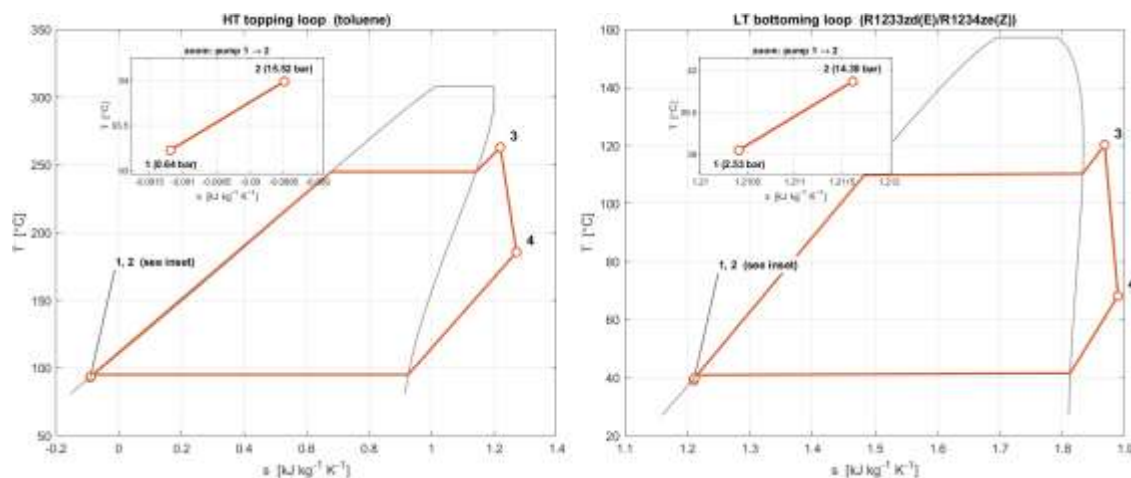


Fig. 3: Temperature–entropy diagrams of (a) the HT topping loop (toluene) and (b) the LT bottoming loop (R1233zd(E)/R1234ze(Z) blend), with state points corresponding to Table 4. The insets resolve states 1 and 2, which coincide at the scale of the main axes because the pump compression is almost isentropic.

3.2 Working-fluid selection: like-for-like comparison

Table 2 reports the stage-2 evaluation of §2.3: every fluid surviving stage 1 is run through the complete cascade model with all other boundary conditions held at the converged design point.

For the topping loop, four of the eight candidates are infeasible at the 244.9 °C design evaporation temperature because their critical temperatures lie below it. Of the four feasible fluids benzene gives the highest net power (5.057 MW, +4.3 % on toluene) and cyclohexane is marginally above toluene (+0.8 %), while n-decane is 5.8 % below; on the objective actually minimised the ranking reverses, toluene giving the lowest levelised cost at 0.0434 \$ kWh⁻¹ against 0.0441, 0.0501 and 0.0532 for benzene, cyclohexane and n-decane, because the higher pressures and larger conductances the alternatives require inflate cost faster than they raise output. Toluene also carries the highest thermal-stability limit of the group (about 400 °C), and benzene — the only fluid that beats it on power — is an IARC Group 1 human carcinogen. Toluene is therefore selected on cost, stability and safety simultaneously, not by assumption.

For the bottoming loop the picture is flatter: three candidates are infeasible at the 109.9 °C design evaporation temperature, and among the five feasible fluids net power ranges only from 4.842 MW to 4.972 MW, a spread of 2.7 %, with levelised cost varying by 0.35 %. Selection is therefore properly governed by the non-thermodynamic criteria: R245fa is excluded by its GWP₁₀₀ of 858, and R600a is ASHRAE class A3, so a hydrocarbon charge of the required size carries a fire risk its 2.5 % power advantage does not justify. The blend was retained over pure R1234ze(Z), which would deliver 0.75 % more power, because its majority component is class A1, which lowers the flammability classification of the charge, because R1233zd(E) has the higher critical temperature and so preserves evaporation-temperature headroom, and because the finite glide was of interest as a design lever — a motivation tested and found worth less than 1 % in §3.9.

3.3 Energy and exergy performance

The headline indicators are collected in Table 5. The cascade delivers 4.851 MW of net shaft power: 5.082 MW of gross turbine output (3.566 MW HT, 1.516 MW LT) less 0.155 MW of pumping and 0.076 MW of fan parasitics. Referred to the

28.60 MW absorbed this is a first-law efficiency of 17.0 %, and referred to the 45.6 MW available to ambient a waste-heat utilisation of 10.6 %. Power is reported on a net-shaft basis; the busbar figure would be about 0.30 MW lower at a combined generator and mechanical efficiency of 0.94, and this and the other optimistic assumptions are quantified in §3.7.

Table 5. Key performance indicators of the recommended cascade.

Indicator	Value	Indicator	Value
Net power, $\dot{W}_{\text{net}}$	4.851 MW	LCOE	0.0434 \$ kWh ⁻¹
Thermal efficiency, η_{th}	17.0 %	NPV (avoided outage tariff)	+27.17 M\$
Wasteheat utilization, η_{WHR}	10.6 %	IRR	33.4 %
System exergy efficiency, ε_{sys}	33.4 %	Discounted payback	3.56 yr
Total exergy destruction	5.771 MW	CO ₂ avoided	22.0 kt yr ⁻¹
Cascade duty, $\dot{Q}_{\text{CHE}}$	5.054 MW	TEWI (direct, refrigerant)	4.84 t
Capital cost (Turton)	11.88 M\$ (2450 \$ kW ⁻¹)		

Second-law accounting on a fuel–product (SPECO) basis gives a system exergy efficiency of 33.3 % with 5.771 MW destroyed; the component breakdown is given in Table 6 and Figure S11 and the full fuel–product balance in Table S3.

The largest destroyers are the HT air condenser (1289 kW, 22.3 %), the LT AQC evaporator (1225 kW, 21.2 %), the HT SP evaporator (1171 kW, 20.3 %) and the CHE (664 kW, 11.5 %), with the two turbines accounting for a further 14.8 %. The advanced decomposition is far more informative than the ranking: the two source evaporators are 100 % unavoidable and the two air condensers 94 % unavoidable, because transfer across the finite gas-to-fluid difference and the air-cooling approach are imposed by the source and the sink, whereas the HT and LT turbines are 88.7 % and 89.9 % avoidable — worth about 0.76 MW, or 16 % of net power. Irreversibility is 96.7 % endogenous, supporting a component-by-component improvement strategy, although that split comes from a reduced isolation model (§4) and should be read as corroborative. The fuel–product efficiencies (Table S3) agree: the turbines and the HT evaporator (0.819–0.874) are high-quality converters, while the air condensers (0.454 and 0.157) are low by construction, their product being heat rejection.

Table 6. Component exergy destruction with the avoidable/unavoidable and endogenous/exogenous decomposition (kW).

Component	$\dot{E}_D$	Share (%)	Avoidable	Unavoidable	Endogenous	Exogenous
HT air condenser	1289.1	22.3	74.7	1214.5	1214.5	74.7
LT AQC evaporator	1224.7	21.2	0	1224.7	1224.7	0
HT evaporator (SP)	1170.5	20.3	0	1170.5	1170.5	0
CHE	664.1	11.5	102.3	561.8	561.8	102.3
LT air condenser	535.7	9.3	34.3	501.4	501.4	34.3
HT turbine	516.4	9.0	458.2	58.1	516.3	0.1
LT turbine	335.0	5.8	301.1	33.9	335.0	0
LT pump	18.9	0.3	18.0	1.0	18.9	0
HT pump	17.0	0.3	16.0	1.0	17.0	0

3.4 Optimisation, environmental-objective redundancy and decision robustness

Surrogate fidelity is high for the thermodynamic targets but lower for the cost targets (§2.8), with a concrete consequence: the ε_{sys} -versus-specific-cost front produced optimistic, physically implausible points that the verification loop rejected when re-solved on the full model (2160 \$ kW⁻¹ predicted at $\varepsilon_{\text{sys}} \approx 0.367$ where the physics gives 2758–2822 \$ kW⁻¹). The recommendation was therefore taken by walking the well-fitted ε_{sys} –LCOE front to the first physics-feasible solution.

The three Pareto frontiers (Figure S12) track the expected efficiency–cost and power–cost trade-offs, the LCOE rising monotonically with both ε_{sys} and $\dot{W}_{\text{net}}$ as larger exchangers are bought. The chosen compromise ($\varepsilon_{\text{sys}} \approx 33.3$ %, LCOE ≈ 0.0434 \$ kWh⁻¹) lies on the low-curvature shoulder of the front, where further efficiency gains require disproportionate capital; the front spans ε_{sys} from 31.7 % to 36.5 % and LCOE from 0.0414 to 0.0478 \$ kWh⁻¹.

Because the study is presented as techno-environmental, it is fair to ask why environmental performance is not itself an objective. The answer is redundancy, and it can be demonstrated. The avoided emission is exactly proportional to delivered energy, so the net life-cycle CO₂ rate is affine and strictly decreasing in net power, with a slope of –4.53 kg CO₂-eq yr⁻¹ per watt; the embodied intensity of 1.37 kg CO₂-eq MWh⁻¹ is 0.22 % of the 610 kg CO₂-eq MWh⁻¹ grid factor and cannot reverse it. Across the whole design of experiments the rank correlation between the net CO₂ rate and both net power and exergy efficiency is exactly –1.000.

A genuine three-objective search (max ε_{sys} , min LCOE, min net life-cycle CO₂) confirms this: its Pareto set projects onto the (ε_{sys} , LCOE) plane essentially on top of the two-objective front, with a mean levelised-cost offset at equal efficiency of 1.4×10^{-4} \$ kWh⁻¹ and a maximum of 7.9×10^{-4} \$ kWh⁻¹, within the surrogate's own RMSE (Figure S2). Adding the environmental criterion therefore does not change the set of efficient designs, though carried as a third TOPSIS criterion it effectively doubles the weight on the power-related direction and moves the selection to $\varepsilon_{\text{sys}} = 36.3$ %. The two-criterion selection is retained because the third criterion carries no independent information.

The decision rule was tested in the same spirit (Table S4, Figure S3). Re-ranking the 70-point front with equal weights, CRITIC weights, min–max normalisation, simple additive weighting, LINMAP and VIKOR moves the selection over $\varepsilon_{\text{sys}} = 33.3$ –36.5 %; six of the seven rules select within 0.8 percentage points of the reported design, and only VIKOR, which minimises a regret measure, departs materially. Sweeping the TOPSIS weight on exergy efficiency from 0.05 to 0.95 moves the selection only across the full width of a front that is itself narrow, so the reported design reflects the shape of the trade-off rather than the choice of decision method.

3.5 Parametric sensitivity of the recommended design

Local sensitivities around the chosen point (Figure S4) confirm the controlling variables. Decreasing the evaporator pinch from 25 K to 5 K raises ϵ_{sys} from about 31.8 % to 37.5 % but the LCOE from about 0.042 to 0.063 \$ kWh⁻¹ — the steepest trade-off in the design space and the reason the optimiser settled on 19.8 K. Raising either evaporation temperature improves ϵ_{sys} while lowering the LCOE, win-win directions exploited to the upper bound of the searched space. The blend composition has little influence: sweeping the R1233zd(E) mass fraction from 0.3 to 0.7 changes ϵ_{sys} by only 0.13 percentage points.

3.6 Economic analysis and uncertainty

Capital cost was estimated with Turton module factors as 11.88 M\$, or 2450 \$ kW⁻¹, comfortably within the 1800–3200 \$ kW⁻¹ band reported for 1–10 MWe industrial ORC retrofits. It is concentrated in three components — the HT SP evaporator (3.17 M\$ for 6190 m² across seven shells), the HT turbine (2.67 M\$) and the LT turbine (2.17 M\$) — which together carry 67 % of the total.

The levelised cost of electricity is 0.0434 \$ kWh⁻¹ and is, by construction, independent of the tariff, so viability depends entirely on how the recovered electricity is valued. Under the deeply subsidised industrial tariff (0.015 \$ kWh⁻¹) the project is unviable (NPV -10.1 M\$, IRR -9.2 %, payback never reached); under an avoided-grid valuation (0.060 \$ kWh⁻¹) it is marginally attractive (+5.9 M\$, 14.2 %, 9.7 yr); and under the avoided-outage valuation that reflects the value of self-generated reliability to a continuous-process plant (0.120 \$ kWh⁻¹) it is strongly attractive (+27.2 M\$, 33.4 %, 3.6 yr), as tabulated in Table S10. Bankability in Iran is therefore governed by tariff treatment rather than by engineering performance.

A local tornado analysis (Figure S5) ranks capacity factor first for the levelised cost and capacity factor and electricity price first for the net present value. One-at-a-time perturbation is not sufficient for a plant-scale investment case, so the economics were also propagated with the 20 000-sample Monte-Carlo scheme of §2.6 (Table 7, Figure 4).

The levelised-cost distribution has a median of 0.0467 \$ kWh⁻¹ and a P10–P90 interval of 0.0308–0.0712 \$ kWh⁻¹, a coefficient of variation of 33 %; the deterministic result sits at the 34th percentile, central-to-slightly-optimistic rather than a best case. Viability is far more sensitive to the valuation than to technical uncertainty: the probability of a positive net present value is 0.05 % at the subsidised tariff, 75.9 % at the avoided-grid valuation and 99.7 % at the avoided-outage valuation, where the median discounted payback is 4 years. The deterministic conclusion therefore survives the uncertainty analysis intact.

The global indices tell a different story from the local tornado. Ranked by standardised rank-regression coefficient, the levelised cost is driven overwhelmingly by the accuracy of the cost correlation itself (0.92), then the discount rate (0.27), plant life and capacity factor (-0.16 each). The largest single reduction in economic uncertainty would therefore come not from refining the thermodynamics but from replacing the factored estimate with vendor-quoted costs for the two turbines and the SP evaporator, which carry 67 % of the capital.

Table 7. Monte-Carlo economic uncertainty (20 000 samples) and global sensitivity indices.

Quantity	P10	P50	P90	Deterministic	Note
LCOE (\$ kWh ⁻¹)	0.0308	0.0467	0.0712	0.0434	CoV = 33 %; deterministic at the 34th percentile
NPV, subsidised tariff 0.015 \$ kWh ⁻¹ (M\$)	-16.85	-9.93	-5.12	-10.06	P(NPV > 0) = 0.05 %
NPV, avoided-grid 0.060 \$ kWh ⁻¹ (M\$)	-3.64	+4.02	+10.44	+5.89	P(NPV > 0) = 75.9 %
NPV, avoided-outage 0.120 \$ kWh ⁻¹ (M\$)	+12.61	+22.45	+32.84	+27.17	P(NPV > 0) = 99.7 %
Discounted payback, avoided-outage (yr)	—	4	7	3.56	P90 value
Global sensitivity (SRRC)	on LCOE		on NPV (avoided-outage)		
Cost-correlation error ($\sigma_{\ln} = 0.30$)	0.918		-0.555		dominant driver
Discount rate (6/8/14 %)	0.270		-0.495		
Plant life (15/20/25 yr)	-0.161		+0.299		
Capacity factor (0.72/0.85/0.93)	-0.158		+0.247		

O&M fraction (2.0/3.0/4.5 %)		0.108		-0.068		
CEPCI (760/800/860)		0.074		-0.047		

Economic uncertainty and global sensitivity analysis

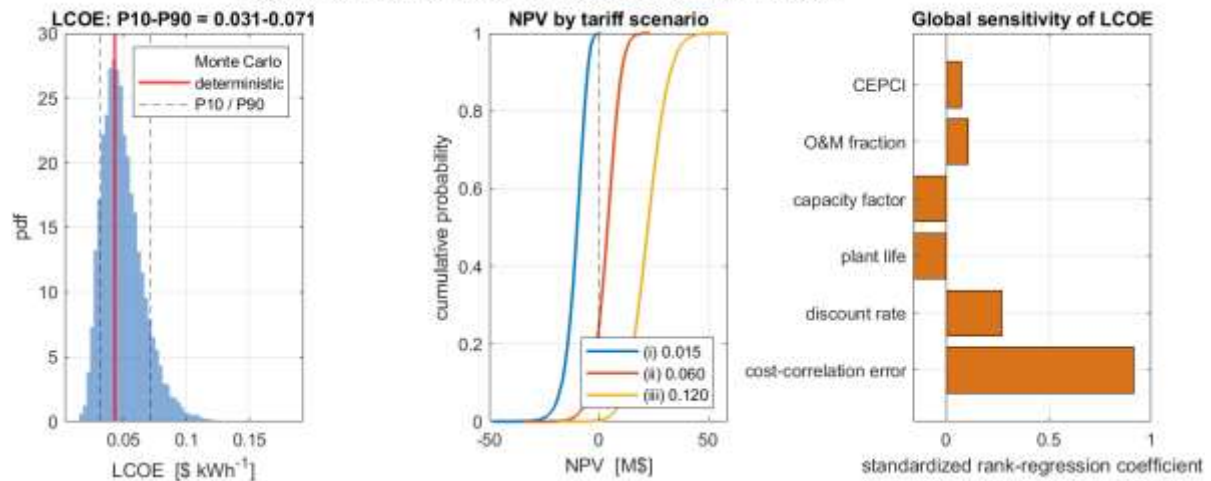


Fig.4: Economic uncertainty and global sensitivity: (a) LCOE distribution with the deterministic value and the P10/P90 bounds; (b) cumulative distribution of NPV in the three tariff scenarios; (c) standardised rank-regression coefficients of the LCOE.

3.7 Sensitivity to the modelling assumptions

Several assumptions in Table 3 are optimistic, and their combined effect deserves to be quantified rather than acknowledged. Table 8 and Figure S9 report a penalty ladder in which each is replaced in turn by a conservative alternative, with the converged geometry held fixed.

The largest single penalty comes from the condenser fans. Sizing them on the air-side pressure drop (150 Pa at 2313 kg s^{-1} of air, fan-motor-drive efficiency 0.65) gives 464 kW rather than the 76 kW implied by the placeholder, reducing net power by 8.0% to 4.463 MW — the price of the design's principal advantage. A generator and mechanical efficiency of 0.96×0.98 takes the cumulative reduction to 14.4% , pressure losses cost a further 2.7 points and a 2% casing heat loss a further 1.9 , giving 3.930 MW — 19.0% below the headline — at $0.0535 \text{ \$ kWh}^{-1}$; degrading the turbomachinery as well gives 3.691 MW at $0.0567 \text{ \$ kWh}^{-1}$.

The important observation is what does not change. Even fully penalised the project retains a net present value of $+17.1 \text{ M\$}$ and a 5.0 -year discounted payback at the avoided-outage valuation, the levelised cost stays below the $0.060 \text{ \$ kWh}^{-1}$ avoided-grid valuation, and the exergy efficiency remains 25.4% . The component ranking, the identification of the turbines as the avoidable-loss components, the negligible role of the glide and the dominance of tariff treatment are all unaffected; the headline numbers should be read as an idealised upper bound whose realistic counterpart is roughly 19% lower.

Table 8. Penalty ladder: effect of replacing each optimistic modelling assumption with a conservative alternative.

Case	$\dot{W}_{\text{gross}}$ (MW)	$\dot{W}_{\text{pump}}$ (MW)	$\dot{W}_{\text{fan}}$ (MW)	$\dot{W}_{\text{net}}$ (MW)	η_{th} (%)	ε_{sys} (%)	LCOE ($\text{\$ kWh}^{-1}$)	NPV (M\\$)	PBP (yr)	$\Delta \dot{W}_{\text{net}}$ (%)
L0 re-reported base case	5.082	0.155	0.076	4.851	16.96	33.35	0.0434	+27.17	3.56	0.0
L1 + physically sized ACC fans	5.082	0.155	0.464	4.463	15.61	30.68	0.0471	+23.77	3.95	-8.0
L2 + generator/mechanical loss (0.96×0.98)	5.082	0.155	0.464	4.153	14.52	28.55	0.0507	+21.05	4.34	-14.4
L3 + working-fluid pressure losses ($3/5/3 \%$)	4.951	0.160	0.464	4.024	14.07	27.66	0.0523	+19.91	4.52	-17.1
L4 + 2% casing heat loss	4.951	0.160	0.464	3.930	13.74	27.02	0.0535	+19.10	4.67	-19.0
L5 + conservative	4.708	0.172	0.467	3.691	12.91	25.37	0.0567	+17.07	5.04	-23.9

turbomachinery (η_s -4 pts, η_p -5 pts)										
--	--	--	--	--	--	--	--	--	--	--

3.8 Environmental and exergoenvironmental analysis

Because the cascade displaces grid electricity, the avoided combustion emission dominates every other environmental term by orders of magnitude (Figure 5): 22.0 kt CO₂ yr⁻¹, or 441 kt over a 20-year life. The direct refrigerant contribution is negligible — 4.84 t CO₂-eq over the plant life, 0.0011 % of the avoided term — a direct consequence of selecting fluids with GWP₁₀₀ ≈ 1; the same cascade charged with R245fa (GWP₁₀₀ = 858) would carry a direct term roughly 850 times larger and would no longer comply with the F-Gas phase-down.

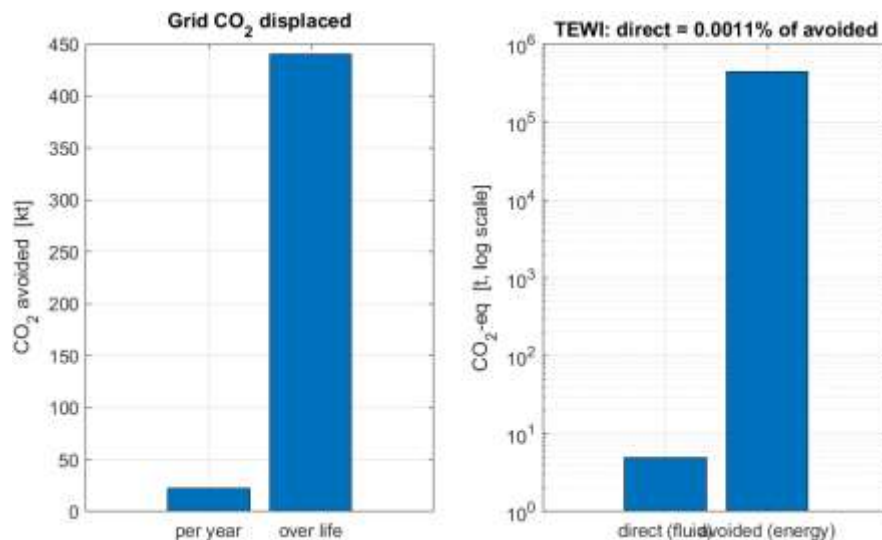


Fig.5: (a) Grid CO₂ displaced per year and over the plant life; (b) TEWI decomposition (logarithmic scale) showing the direct refrigerant term relative to the avoided-energy term.

Combining the second-law and life-cycle results yields the exergoenvironmental balance of Table S5 and Figure S6. Because the waste-heat fuel carries an essentially zero specific CO₂ burden, the embodied impact of the delivered electricity is only 1.37 kg CO₂-eq MWh⁻¹ — 0.22 % of the grid intensity displaced — giving an emission payback of well under two months. The exergoenvironmental factor f_b then points improvement effort in a specific direction: the HT SP evaporator dominates the impact rate with $f_b = 1$, its burden entirely constructive (a 279 t steel mass), so it calls for a lighter exchanger rather than thermodynamic measures, whereas the turbines show the irreversibility-driven signature ($f_b = 0.42$ and 0.20) where the avoidable destruction of §3.3 is the actionable lever.

3.9 Assessment of the zeotropic working-fluid choice

The role of the zeotropic blend is real but marginal for this source and sink. The 0.61 K condensation glide of Figure 2(c) was cross-validated against an independent ideal-solution (Raoult) model, which reproduces a known high-glide pair (R600a/R601a, 12.8 K) correctly, so the small value is a genuine physical result and not a property-routine artefact: R1233zd(E) and R1234ze(Z) have near-parallel vapour-pressure curves and hence low relative volatility. A like-for-like comparison at equal pinch (Table S7) shows that substituting the blend for pure R1233zd(E) changes the LT condenser exergy destruction by less than 0.3 %, raises ϵ_{sys} by about 1.0 % and lowers the LCOE by about 0.4 %: a 0.61 K glide against a 5.3 K air-side rise is simply too small to improve the temperature match materially.

3.10 Operational layer: verification, state estimation and fault response

The operational layer is assessed in three steps — verification of the reduced-order model, characterisation of the estimator, and demonstration of the closed loop — with supporting detail in Table S6 and Figures S7 and S8.

Verification. The two-zone moving-boundary representation was tested against the 40-cell distributed exchanger by giving both models only the total conductance and the inlet conditions and asking them to predict the duty. The reduced-order model returns 21.06 MW against 20.93 MW (0.62 %) and a gas outlet temperature of 147.3 °C against 148.6 °C (0.89 %), whereas a single-zone lumped model errs by 18.5 % and 26.3 % because the preheating zone carries 54.7 % of the duty. The two-zone discretisation is therefore both necessary and sufficient as the real-time model.

Dynamics. Under a +25 K step in SP gas temperature the wall temperature moves from 297 to 309 °C and the two-phase length fraction from 93.95 % to 94.49 %. The two states operate on very different time scales: the boundary migrates in a few seconds and is effectively quasi-algebraic, whereas the wall responds with a first-order time constant of 56 s and a 95 % settling time of 167 s, moving to 83–250 s when the wall thermal mass is varied by ± 50 %. At a 60 s assimilation interval the model is fast enough for real-time emulation, but the wall thermal mass is the first quantity a plant step-response test would have to identify.

Estimator performance. The scalar Kalman filter was characterised by Monte Carlo over 500 noise realisations at each of three sensor noise levels. At the assumed $\sigma = 0.8$ K it reduces the root-mean-square temperature error from 0.799 K to 0.401 K, with a mean normalised innovation squared of 1.09 — close to the theoretical unity, confirming statistical consistency — and it degrades gracefully, giving 0.225 K at $\sigma = 0.4$ K and 0.696 K at $\sigma = 1.5$ K. Fault detection with a 3 K margin has a mean delay of 35.3 ± 2.4 samples (35 min) with no false alarms in 500 runs; that delay is a tuning choice, since a 2 K margin gives 25.3 samples at a 0.014 % false-alarm rate and a 1 K margin 14.9 samples at 0.27 %.

Closed-loop demonstration. The full loop was demonstrated on a synthetic progressive condenser-fouling fault (Figure 6). The filter tracks the rising LT condensing temperature, the detector raises the flag after the expected 33-sample delay, and the optimiser re-solves the physical model and raises the LT evaporation set-point from 110 to 128 °C. The closed loop recovers 150 % of the power lost to fouling, lifting output above the pre-fault value (4.85 $\rightarrow$ 4.96 MW, against 4.64 MW if untreated) because the retuning also exploits headroom the nominal set-point had left unused. This establishes the operational character of the layer, though not agreement with plant behaviour.

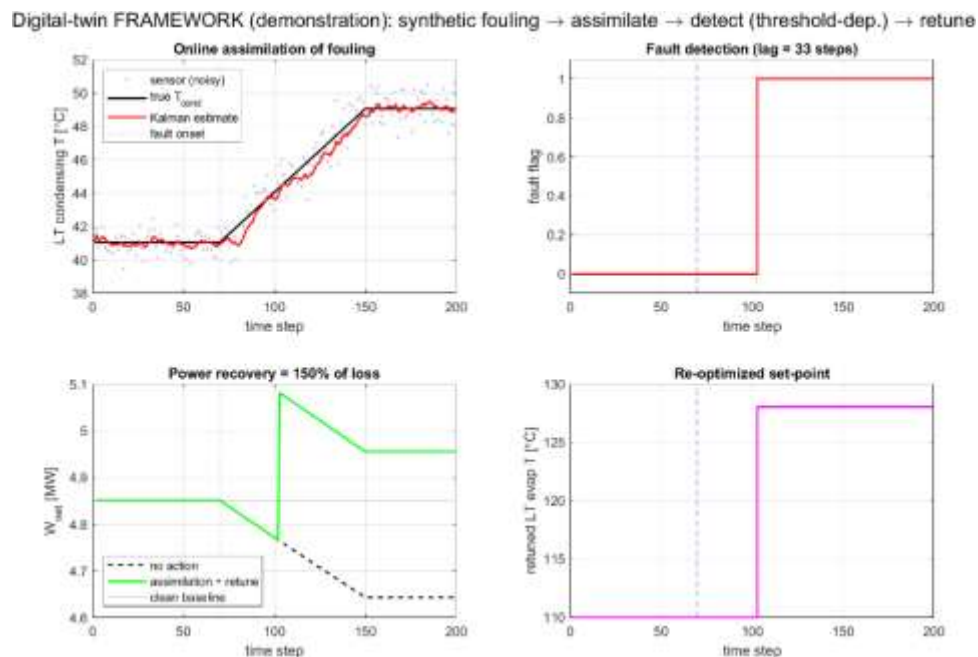


Fig. 6: Operational-layer demonstration on a synthetic fouling fault: (a) Kalman assimilation of the LT condensing temperature from noisy measurements; (b) fault-detection flag (33-sample lag); (c) net-power trajectory with and without assimilation-driven retuning (recovery = 150 % of the loss); (d) re-optimised LT evaporation set-point.

3.11 Model validation and comparison with related work

Table S1 collects the four-layer validation of §2.10. Critical points and normal boiling points agree with published reference values to within 0.03 % for toluene, cyclohexane, R1234ze(Z) and R245fa, the largest deviation being R1233zd(E) at 0.44 % and 1.13 % because the library implements an earlier equation of state. The cascade closes the first law to 1.0×10^{-8} W against a 28.6 MW duty, the SPECO balance closes component by component, and the thermal efficiency of 17.0 % lies below both the Carnot bound of 52.9 % and the tighter Lorenz bound of 41.7 %, a second-law ratio of 0.41 typical of a well-designed subcritical ORC. Independently recomputed turbine, pump and heat-duty terms match the solver to within 10^{-6} relative, and two published single-loop operating points are reproduced within 0.6 % and 3.6 %.

The plant layer is the most demanding test available without measurements, and it resolves the apparent power deficit against the same-plant literature. Sanaye et al. [22] reported 6.56 MW across two production lines, against which the present 4.851 MW is 26 % lower at face value; the difference is one of source basis and cooling technology, not cycle quality, since their figure covers the full two-line plant and assumes water cooling. Re-solving over a range of source throughputs with every

intensive variable fixed (Table S8) gives a net power per unit of available thermal input that is exactly flow-invariant at $0.1065 \text{ MW MW}^{-1}$ and reproduces 6.56 MW at an SP throughput of 121.7 kg s^{-1} — a 35 % larger inventory, entirely consistent with a two-line basis, so the two systems are at parity per unit of recovered heat.

Table S9 places this work alongside the most closely related studies with the basis of each comparison stated explicitly, since levelised costs in particular are reported on incompatible bases in the literature (full LCOE here, EPC-only elsewhere).

4. Limitations and Future Work

This is a model-based, steady-state design study with an operational layer demonstrated in simulation, and many limitations qualify it. First, properties, source conditions and techno-economic parameters come from published libraries rather than site measurements, and neither the reduced-order model nor the estimator has been confronted with plant transients — which is why the layer is called digital-twin-oriented. Second, the SP mass flow of 90 kg s^{-1} is a plant-scale estimate and propagates linearly into absolute power, although all intensive results are invariant to it. Third, as quantified in §3.7, the reported performance rests on a fan-power placeholder, a net-shaft basis, frictionless piping and adiabatic casings, which together overstate net power by about 19 %.

5. Conclusions

A water-free, air-cooled cascaded dual-loop ORC with ultra-low-GWP working fluids was designed, optimised and assessed for waste-heat recovery from a cement kiln line in Mazandaran, Iran, through an integrated five-aspect framework with a digital-twin-oriented operational layer. The recommended cascade — a pure-toluene topping loop on the preheater exhaust and an R1233zd(E)/R1234ze(Z) bottoming loop (0.633/0.367 by mass) on the clinker-cooler air, linked by a 5.05 MW cascade heat exchanger — absorbs 28.60 MW and delivers 4.851 MW of net shaft power, for a thermal efficiency of 17.0 % (41 % of the finite-source Lorenz bound), a waste-heat utilisation of 10.6 % and an exergy efficiency of 33.3 %.

The contribution is therefore the integrated, auditable package — quantitative fluid screening, a water-free ultra-low-GWP cascade, five-aspect assessment with advanced exergy and exergoenvironmental decomposition, uncertainty-propagated economics and a verified operational layer — rather than the absolute power output, which is set by the assumed source inventory. The clearest next steps are to recover the 2.83 MW unrecovered gas tail, to relax the active evaporation-temperature bounds, and to commission the operational layer against plant data so that it can properly be called a digital twin.

References

- [1] Kasaeian, A., Afshari, F., Mahmoudkhani, M., Masoumi, A., & Esmaeili Bidhendi, M. (2025). Waste heat recovery by thermodynamic cycles in cement plants: A review. *Energy*, 314, 134176. doi: 10.1016/j.energy.2024.134176
- [2] Karellas, S., Leontaritis, A.-D., Panousis, G., Bellos, E., & Kakaras, E. (2013). Energetic and exergetic analysis of waste heat recovery systems in the cement industry. *Energy*, 58, 147–156. doi: 10.1016/j.energy.2013.03.097
- [3] Forman, C., Muritala, I. K., Pardemann, R., & Meyer, B. (2016). Estimating the global waste heat potential. *Renewable and Sustainable Energy Reviews*, 57, 1568–1579. doi: 10.1016/j.rser.2015.12.192
- [4] Quoilin, S., Van Den Broek, M., Declaye, S., Dewallef, P., & Lemort, V. (2013). Techno-economic survey of organic Rankine cycle (ORC) systems. *Renewable and Sustainable Energy Reviews*, 22, 168–186. doi: 10.1016/j.rser.2013.01.028
- [5] Tchanche, B. F., Lambrinos, G., Frangoudakis, A., & Papadakis, G. (2011). Low-grade heat conversion into power using organic Rankine cycles – A review of various applications. *Renewable and Sustainable Energy Reviews*, 15(8), 3963–3979. doi: 10.1016/j.rser.2011.07.024
- [6] Lecompte, S., Huisseune, H., van den Broek, M., Vanslambrouck, B., & De Paepe, M. (2015). Review of organic Rankine cycle (ORC) architectures for waste heat recovery. *Renewable and Sustainable Energy Reviews*, 47, 448–461. doi: 10.1016/j.rser.2015.03.089
- [7] Xu, B., Rathod, D., Yebi, A., Filipi, Z., Onori, S., & Hoffman, M. (2019). A comprehensive review of organic Rankine cycle waste heat recovery systems in heavy-duty diesel engine applications. *Renewable and Sustainable Energy Reviews*, 107, 145–170. doi: 10.1016/j.rser.2019.03.012
- [8] Colonna, P., Casati, E., Trapp, C., Mathijssen, T., Larjola, J., Turunen-Saaresti, T., & Uusitalo, A. (2015). Organic Rankine cycle power systems: From the concept to current technology, applications, and an outlook to the future. *Journal of Engineering for Gas Turbines and Power*, 137(10), 100801. doi: 10.1115/1.4029884
- [9] Bao, J., & Zhao, L. (2013). A review of working fluid and expander selections for organic Rankine cycle. *Renewable and Sustainable Energy Reviews*, 24, 325–342. doi: 10.1016/j.rser.2013.03.040
- [10] Drescher, U., & Brüggemann, D. (2007). Fluid selection for the organic Rankine cycle (ORC) in biomass power and heat plants. *Applied Thermal Engineering*, 27(1), 223–228. doi: 10.1016/j.applthermaleng.2006.04.024
- [11] Zhang, Y., Ren, X., Duan, X., Gong, L., & Hung, T.-C. (2024). Strategy for the zeotropic organic Rankine cycle operation to match the heat sink variation. *Energy*, 286, 129541. doi: 10.1016/j.energy.2023.129541
- [12] Fergani, Z., Morosuk, T., & Touil, D. (2021). Exergy-based multi-objective optimization of an organic Rankine cycle with a zeotropic mixture. *Entropy*, 23(8), 954. doi: 10.3390/e23080954

- [13] Wang, E. H., Zhang, H. G., Fan, B. Y., Ouyang, M. G., Zhao, Y., & Mu, Q. H. (2011). Study of working fluid selection of organic Rankine cycle (ORC) for engine waste heat recovery. *Energy*, 36(5), 3406–3418. doi: 10.1016/j.energy.2011.03.041
- [14] Shu, G., Liu, L., Tian, H., Wei, H., & Yu, G. (2014). Parametric and working fluid analysis of a dual-loop organic Rankine cycle (DORC) used in engine waste heat recovery. *Applied Energy*, 113, 1188–1198. doi: 10.1016/j.apenergy.2013.08.027
- [15] Li, Y., Li, Z., Zhang, H., Zhang, J., He, X., Qiao, Y., & Yu, Z. (2024). Comparative exergy analysis of series and parallel dual-pressure auto-cascade organic Rankine cycles. *Processes*, 12(12), 2872. doi: 10.3390/pr12122872
- [16] Kelly, S., Tsatsaronis, G., & Morosuk, T. (2009). Advanced exergetic analysis: Approaches for splitting the exergy destruction into endogenous and exogenous parts. *Energy*, 34(3), 384–391. doi: 10.1016/j.energy.2008.12.007
- [17] Fergani, Z., & Morosuk, T. (2023). Advanced exergy-based analysis of an organic Rankine cycle (ORC) for waste heat recovery. *Entropy*, 25(10), 1475. doi: 10.3390/e25101475
- [18] Imran, M., Haglind, F., Asim, M., & Zeb Alvi, J. (2018). Recent research trends in organic Rankine cycle technology: A bibliometric approach. *Renewable and Sustainable Energy Reviews*, 81, 552–562. doi: 10.1016/j.rser.2017.08.028
- [19] Yu, W., Patros, P., Young, B., Klinac, E., & Walmsley, T. G. (2022). Energy digital twin technology for industrial energy management: Classification, challenges and future. *Renewable and Sustainable Energy Reviews*, 161, 112407. doi: 10.1016/j.rser.2022.112407
- [20] Al-Shetwi, A. Q., Atawi, I. E., El-Hameed, M. A., & Abuelrub, A. (2025). Digital twin technology for renewable energy, smart grids, energy storage and vehicle-to-grid integration: Advancements, applications, key players, challenges and future perspectives in modernising sustainable grids. *IET Smart Grid*, 8, e70026. doi: 10.1049/stg2.70026
- [21] Abdessadak, A., Ghennioui, H., Thirion-Moreau, N., Elbhiri, B., Abraim, M., & Merzouk, S. (2025). Digital twin technology and artificial intelligence in energy transition: A comprehensive systematic review of applications. *Energy Reports*, 13, 3245–3268. doi: 10.1016/j.egy.2025.02.026
- [22] Sanaye, S., Khakpaay, N., Chitsaz, A., Hassan Yahyanejad, M., & Zolfaghari, M. (2020). A comprehensive approach for designing, modeling and optimizing of waste heat recovery cycle and power generation system in a cement plant: A thermo-economic and environmental assessment. *Energy Conversion and Management*, 205, 112353. doi: 10.1016/j.enconman.2019.112353
- [23] Wang, Y., Xu, Y., Qiu, Y., & Ning, S. (2025). Thermodynamics, economy and environment analyses and optimization of series, parallel, dual-loop Kalina cycles for double-source heat recovery in cement industry. *PLOS ONE*, 20(2), e0315972. doi: 10.1371/journal.pone.0315972
- [24] Gioia, R., Ottaviano, S., Romei, A., Peretto, A., Branchini, L., & Spinelli, A. (2026). Optimal cycle and turbine design for MW-scale waste heat recovery organic Rankine cycle with partial evaporation. *Journal of Engineering for Gas Turbines and Power*, 148(3), 031014. doi: 10.1115/1.4069578
- [25] Bell, I. H., Wronski, J., Quoilin, S., & Lemort, V. (2014). Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp. *Industrial & Engineering Chemistry Research*, 53(6), 2498–2508. doi: 10.1021/ie4033999
- [26] Turton, R., Shaeiwitz, J. A., Bhattacharyya, D., & Whiting, W. B. (2018). *Analysis, synthesis and design of chemical processes* (5th ed.). Prentice Hall.
- [27] Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197. doi: 10.1109/4235.996017
- [28] Hwang, C.-L., & Yoon, K. (1981). *Multiple attribute decision making: Methods and applications*. Springer. doi: 10.1007/978-3-642-48318-9
- [29] Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The CRITIC method. *Computers & Operations Research*, 22(7), 763–770. doi: 10.1016/0305-0548(94)00059-H
- [30] Opricovic, S., & Tzeng, G.-H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *European Journal of Operational Research*, 156(2), 445–455. doi: 10.1016/S0377-2217(03)00020-1
- [31] Rasmussen, B. P. (2012). Dynamic modeling for vapor compression systems—Part I: Literature review. *HVAC&R Research*, 18(5), 934–955. doi: 10.1080/10789669.2011.582916
- [32] Calderazzi, L., & Colonna di Paliano, P. (1997). Thermal stability of R-134a, R-141b, R-131I, R-7146, R-125 associated with stainless steel as a containing material. *International Journal of Refrigeration*, 20(6), 381–389. doi: 10.1016/S0140-7007(97)00043-1
- [33] Invernizzi, C., Iora, P., & Silva, P. (2007). Bottoming micro-Rankine cycles for micro-gas turbines. *Applied Thermal Engineering*, 27(1), 100–110. doi: 10.1016/j.applthermaleng.2006.05.003