© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1723034

Advanced Thermal Performance Optimization of Refinery Heat Transfer Systems for Saudi Vision 2030

MOHAMMAD ISMAIL ANSAR AHMED ANSARI
0009-0009-7917-935X

Abstract*—The manner in which refinery heat-transfer systems function has a direct bearing on how efficiently high-grade process heat can be recovered before additional fuel, steam, or electricity is consumed; since this affects both operating costs and the achievement of industrial decarbonisation, the performance of these systems is of vital importance. This review gathers the research conducted between 2020 and 2025 on the advanced optimisation of the thermal performance of refinery heat exchangers and heat exchanger networks, with a specific emphasis on the relevance of these findings to Saudi Arabia's Vision 2030 transition. The evidence is organised around five interrelated factors: heat-integration retrofit, fouling-aware operation, equipment-level improvements, data-driven condition monitoring, and low-carbon heat upgrading. Recent studies of refineries have demonstrated that even technically sound measures can be undermined if the network's topology, pressure-drop constraints, changes in crude oil properties, the availability of cleaning operations, and interactions with utility systems are taken into account separately. A more dependable approach is to optimise thermal recovery and hydraulic operability over the entire operating cycle rather than merely at a single point that corresponds to a clean design condition. The review therefore offers a decision-making framework tailored to Saudi Arabia, giving priority to high measurement quality, accurate fouling diagnosis, operational changes that require no capital or only low capital, targeted network retrofit, and finally the integration of electrified or renewable heat. The synthesis indicates that the most beneficial action in the short term is to combine pinch-guided retrofit with fouling prediction and dynamic control, while longer-term decarbonisation will need the use of heat pumps, renewable energy, and flexible utility systems. There are still research gaps relating to plant-scale validation, the management of uncertainty, water-aware optimisation, operation at high ambient temperatures, and explainable machine-learning models. These gaps define a practical research programme for reducing refinery energy intensity without having to compromise reliability, throughput, or product quality.

I. INTRODUCTION

The refining of petroleum remains a thermally intensive process since the steps involved in separation, reaction, hydrogen production, and the preparation of the final products all involve large and continuous transfers of sensible and latent heat. Consequence of this is that a modern refinery should more accurately be seen not as a collection of individual units but as a coupled thermal system in which the loss of duty by one exchanger can have an effect on furnace firing, steam demand, cooling demand, throughput, and emissions. Evaluations recently carried out on a unit-by-unit basis have shown that heat recovery, electrification, the use of renewable hydrogen, digital optimisation, and process integration are some of the most feasible ways of reducing refinery emissions while the existing assets are still used. Moreover, studies at the process level show that the benefit of any single measure is either increased or decreased depending on the way the various parts of the system interact, meaning that an integrated approach is essential if genuine long-term emission reductions are to be achieved.

For Saudi Arabia the issue also has a strategic dimension. The Vision 2030 plan is intended to achieve economic diversification, to improve resource productivity and to establish a more efficient energy system while the country continues to operate a large amount of hydrocarbon-processing infrastructure. The national literature on the energy transition has emphasised the need to reduce dependence on fossil energy for domestic use and to extend the supply of renewable and sustainable energy. The circular-carbon approach also implies that industrial competitiveness and emissions control should be pursued together rather than being seen as opposing objectives. With regard to refineries, heat-transfer systems represent a reasonable point at which to intervene since improved internal heat recovery directly reduces the demand for external hot utility and in most cases attains this result without changing the range of products.

The crude preheat train is the best example of this method; before the crude oil enters the atmospheric furnace, the hot product and pumparound streams transfer their heat to it, and the setup is designed to recover as much process heat as is feasible, considering the temperature driving force, the pressure drop, the metallurgical limitations, the available plot space, and the need for good controllability. Research carried out at a refinery in Kuwait has demonstrated that graphical pinch analysis can spot modifications which are economically attractive for an existing crude-distillation heat exchanger network. Further investigations carried out at various refineries have shown that altering the process and the utilities can result in additional savings beyond those obtained merely by reconnecting the exchangers, and more recent graphical methods have successfully been used in the retrofitting of a residuum hydrogenation network.

Yet simply setting a target for clean-network energy is not enough. The exchangers in a refinery handle crude oil blends, heavy hydrocarbons, salts, corrosion products, and varying velocities, all of which result in deposition and ageing. Fouling increases thermal resistance and may also cause a greater pressure drop, which in turn means a higher furnace duty is required, or places limits on the throughput. Because the network structure causes the response to be non-local, cleaning one exchanger may redistribute the temperature driving forces and the hydraulic load over several units. Consequently, multi-period retrofit methods seek out solutions that will still be viable under changing conditions rather than aiming to optimise a single operating point. It has also been discovered that when low-carbon energy sources are introduced, the design of the utility system and the process recovery should be assessed together.

Economic uncertainty is just as significant; a retrofit that appears to be the better option at one particular fuel or carbon price might see its advantage disappear when long-term energy prices change. It has therefore been proposed that scenario-based evaluations of the different options for a refinery heat exchanger network should be carried out in order to lower the risk of selecting a project that is technically efficient but economically unsound. Similarly, design studies that take fouling into account show that the area margin, the bypass arrangement, the flow velocity, and the cleaning procedure all need to be considered throughout the entire operating cycle since capital and operating decisions are interdependent. The fact that this kind of evidence exists shows that there is a shift taking place from the traditional way of assessing exchangers to lifecycle thermal performance management.

The review focuses on that change. Rather than outlining the various types of exchangers, it examines how refinery heat-transfer systems can be optimised as dynamic, fouling-prone and hydraulically constrained assets within a decarbonising Saudi energy system. It combines recent evidence on network retrofitting, fouling mitigation, exergy, machine learning, heat transfer enhancement, waste-heat upgrading and the integration of renewables. The key conclusion is that fulfilling the requirements of Vision 2030 requires a number of measures: first recover the performance that has been lost, then improve the network configuration, and only thereafter introduce new low-carbon heat sources. This sequence of actions avoids the unnecessary electrification of inefficiencies and provides a technically defensible path from operational excellence to structural decarbonisation.

II. AIM AND OBJECTIVES

The aim of this review is to provide a critical evaluation of the most recent methods aimed at achieving the simultaneous optimisation of thermal recovery, hydraulic resilience, fouling control, operational versatility, and decarbonisation in refinery heat-transfer systems, and also to use this evidence to develop a framework for use in Saudi Arabian refineries by the year 2030. The scope of the review is deliberately broader than that of a standard heat exchanger design survey since the major performance losses are caused by the interactions between the individual pieces of equipment, the network structure, process variability, maintenance decisions, and the supply of utilities.

The study is guided by five objectives. The first of these is to compare the various thermal and fiscal factors through which improvements in performance are obtained via network retrofitting, process changes, cleaning, the redistribution of flow, surface enhancement and the upgrading of waste heat. The second objective is to examine the impact of fouling and the ageing of deposits on both the heat duty and the pressure drop, thereby causing the most favourable condition to vary with time. The third objective consists of assessing the role of data-driven monitoring, machine learning and graph-oriented methods in diagnosing degradation and in identifying possible retrofit options. The fourth objective concerns how electrified and renewable heat can be combined without neglecting the heat-recovery possibilities that are already present in the refinery. The fifth is to determine the research priorities that are specific to Saudi Arabia, having regard to the high ambient temperature, water shortages, the variability of the crude oil, the reliability requirements and the need for staged capital investment.

III. REVIEW METHODOLOGY

So that the evidence base would stay focused and reproducible, a structured critical-review approach was used. The time period covered by the publications was defined as starting on 1 January 2020 and ending on 31 December 2025. Any articles that had been peer-reviewed and were published in English were included provided that they concerned refinery or transferable process-industry heat exchanger networks, crude preheat trains, thermal-hydraulic optimisation, fouling management, heat transfer enhancement, heat recovery, industrial heat pumps, renewable integration, or data-driven condition monitoring. The search terms employed combined a number of terms such as refinery, crude oil, heat exchanger, preheat train, heat exchanger network, pinch analysis, fouling, cleaning scheduling, exergy, retrofit, machine learning, waste heat, heat pump, renewable energy, and Saudi Arabia.

The evidence was looked at to find out whether it was directly relevant to at least one of the following five areas of analysis: network-level heat integration; fouling-aware operation and maintenance; equipment-level thermal-hydraulic enhancement; digital monitoring and optimisation; and low-carbon heat supply. Studies that were outside the date range, lacked a verifiable DOI, were not peer-reviewed, or addressed heat transfer in a manner with no substantial link to industrial energy recovery were excluded. The final synthesised report contains 30 references, all of which date from 2020 to 2025. Although the background papers attached were checked with regard to the structure of the review, the technical approach, and the use of the synthesis tables, only those items that met both the required topic and the publication date criteria were included as cited evidence.

The evaluation of the evidence was carried out qualitatively by means of four questions: What is the physical mechanism which causes the supposed improvement? What constraints were considered, particularly those concerning the temperature approach, the pressure drop, fouling, multi-period operation, and capital cost? Has the method in fact been tested on a case that is relevant to an industrial or refinery environment? And to what extent can the results be applied to the real operating conditions in Saudi Arabia? Although quantitative results from the various studies are used as background information, they are not combined to produce pooled estimates because the network boundaries, utility prices, crude oil properties, baseline conditions, and objective functions differ greatly between the studies. The review thus concentrates on common mechanisms and the decision logic rather than on a single numerical benchmark.

The process begins with an examination of the physics of the individual components, then proceeds to the selection of the network, and finally involves the integration of the energy system. This order is intentional, since even though the overall heat-transfer coefficient may be high, the refinery outcome will not necessarily be more successful if the pressure drop becomes excessive, if fouling is speeded up, if a pinch constraint is moved, or if an uncontrolled network is generated. Conversely, a retrofit of the network might do poorly if the deterioration of the exchangers is ignored. The approach therefore treats thermal duty, exergy loss, pumping demand, maintenance, and the use of external utilities as closely related performance factors. Figure 1 gives a summary of this multi-layer analytical system, and Table 1 compares the various optimisation methods and their relevance to the deployment of the Saudi refinery.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1723034

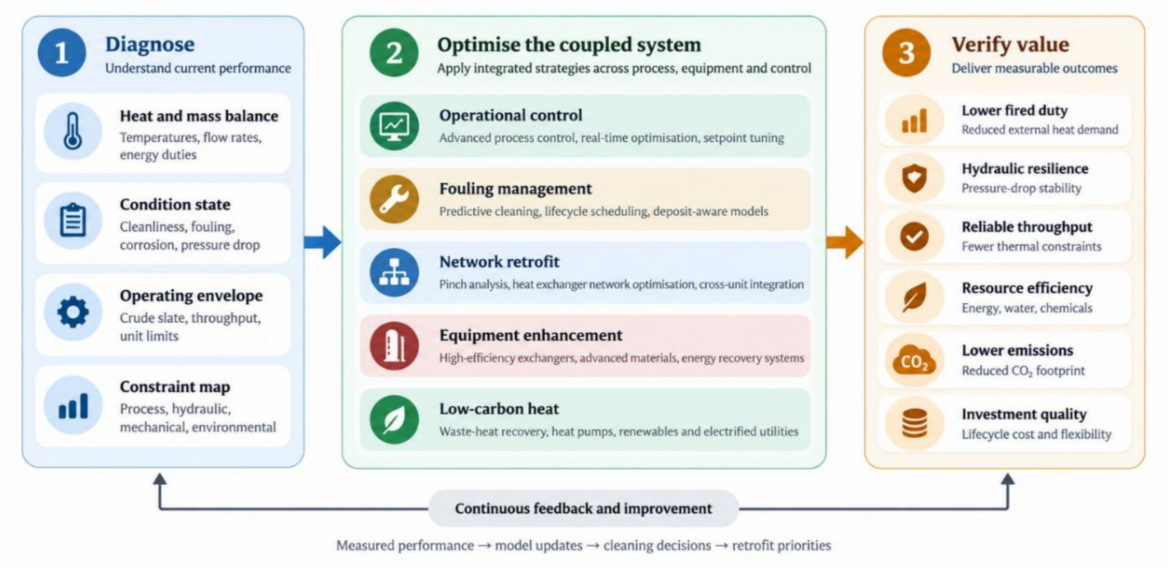


Fig. 1. Integrated architecture for refinery *thermal*-performance optimisation
From trusted plant data to lifecycle energy, reliablity and decarbonisation decisions

Table 1. Critical synthesis of advanced thermal-performance optimisation pathways.

Domain	Primary mechanism	Critical evidence from 2020-2025	Saudi refinery implication
Network heat integration	Increase internal heat recovery through rematching, process changes and multi-period retrofit.	Pinch and graph methods expose thermodynamic bottlenecks; robust solutions must respect existing topology and varying operation .	Prioritise brownfield modifications that remain feasible across crude slates, throughput and seasonal utility constraints.
Fouling-aware operation	Restore effective heat-transfer area and manage thermal-hydraulic degradation over the operating cycle.	Cleaning, flow redistribution, ageing models and lifecycle design should be co-optimised rather than scheduled independently.	Include water, cleaning resources, furnace penalties, pressure-drop margin and turnaround windows in the objective.
Equipment enhancement	Improve local heat transfer or reduce irreversibility using geometry, area, velocity and exchanger-type changes.	Local gains are constrained by pressure drop, cleanability and network temperature driving forces.	Use selective enhancement after diagnosing whether the true bottleneck is fouling, area, driving force or hydraulics.
Digital monitoring	Estimate clean baseline, fouling state, anomalies and promising retrofit paths from plant data and network structure.	Machine learning and graph methods can accelerate diagnosis and candidate screening, but industrial generalisation remains limited .	Require uncertainty, explainability and validation across crude changes, sensor drift and seasonal modes.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1723034

Domain	Primary mechanism	Critical evidence from 2020-2025	Saudi refinery implication
Waste-heat upgrading	Lift low-grade residual heat to a useful sink temperature after internal recovery has been improved.	Industrial heat-pump performance depends strongly on source/sink temperatures and electricity conditions .	Map stable heat sources and sinks; incorporate summer utility constraints and grid-carbon sensitivity.
Low-carbon utilities	Replace residual external energy with renewable or electrified supply and storage.	Refinery energy-system optimisation shows that variable renewable supply and demand flexibility must be solved together.	Size new low-carbon utilities against an efficiency-improved baseline to avoid electrifying avoidable demand.

IV. THERMAL PERFORMANCE AS A NETWORK PROBLEM

The performance of an exchanger is determined by its heat duty, the temperature driving force, the overall heat-transfer coefficient, the area, and the hydraulic resistance. In practice, a refinery operator is concerned not only with the exchanger carrying out its designed heat duty but also with it doing so without requiring excessive pumping power, without violating pressure limits, or without causing instability in the downstream temperatures. Studies using exergoeconomic methods on shell-and-tube exchangers show that thermal and hydraulic variables should be optimised together since designs which improve heat transfer may also lead to greater pressure losses and higher operating costs. Advanced exergy analysis offers a complementary means of diagnosis by separating avoidable irreversibility from that which is structurally unavoidable, thus aiding in the distinction between potential areas for improvement and losses that are inherent to the configuration chosen .

At the network level, the various trade-offs become more complicated since each match influences the temperature available to the following exchangers. Pinch analysis is still useful since it shows where heat recovery is limited by the thermodynamic temperature approach rather than by preferences regarding the local equipment. The example of the Kuwaiti refinery demonstrated that graphical analysis can identify practical retrofit modifications and opportunities for improving the temperature driving force in an existing network. Furthermore, a stream-specific temperature-enthalpy analysis is able to detect process or utility changes which conventional composite curves might conceal. In the case of residuum hydrogenation, graphical retrofitting increased heat recovery and resulted in a short reported payback period, thus showing that transparent and easily interpretable methods remain relevant in an industrial context.

Yet if only one target condition is considered, the resulting design can be fragile. There are variations in refinery throughput, in the composition of the crude oil, in the product yields, in the ambient conditions, and in the availability of equipment. Optimisation based on graph theory has been applied in order to generate a number of alternative options for retrofitting a refinery while taking into account the existing structure, and large-scale optimisation techniques aim at finding computationally efficient solutions for complex network synthesis. Tools that deal with multiple time periods specifically examine how a network reacts to changes in the inlet conditions and flow rates, whereas hybrid evolutionary methods include practical constraints and allow for flexible configurations when operating over several periods. Together, these methods show that thermal performance should be assessed according to the feasible recovery over the entire operating envelope, not on the basis of the maximum recovery at a single nominal point.

This point is especially pertinent to the Saudi refineries, since high summer ambient temperatures can lead to a decrease in the performance of the air coolers and cause changes in the bottlenecks within the utility systems. A retrofit which increases heat recovery but transfers more of the duty to a cooling system that is already constrained might not result in a lower site energy intensity. Likewise, a reaction that proves attractive with one kind of crude oil could end up having a higher viscosity, a greater pressure drop, or more deposition when using a different crude. The desired objective is therefore not merely to achieve the maximum amount of recovered megawatts; it must be to minimise the total lifecycle utility and maintenance requirements while satisfying the constraints relating to production, hydraulics, metallurgy, controllability, and reliability.

V. ADVANCED OPTIMISATION PATHWAYS

5.1 Fouling-aware operation and cleaning

Fouling is the main dynamic limitation in the recovery of heat in a refinery; as deposits build up the thermal resistance increases and, as the deposits age, their conductivity and mechanical properties can change. Research into the visualisation of crude oil preheat trains has demonstrated that ageing of the deposits can alter the balance between thermal and hydraulic effects, and that it is necessary to monitor the pressure drop if the network's behaviour is to be interpreted correctly. This is significant since a drop in outlet temperature may result from a number of factors, such as fouling, changes in flow, sensor bias, or disturbances upstream. It therefore follows that treating a fall in temperature alone as an indication of fouling may result in cleaning being carried out too early or intervention being delayed.

Operational optimisation provides a means of obtaining value before equipment is replaced. Methods that are used in real time and combine cleaning schedules with control of flow distribution have shown lower operating costs than if the two actions were dealt with separately, since the optimiser can temporarily redistribute the workload while deciding whether or not cleaning creates sufficient economic value. Similarly, measurement-aided scheduling treats cleaning as an avoided-cost decision rather than as a fixed task on a calendar. The economic and environmental implications are so significant that this amount of attention is warranted: updated evaluations of fouling in the crude-preheat process indicate large sector-wide costs together with the emissions resulting from the extra fuel required.

The design and maintenance decisions must therefore follow the same lifecycle approach. An optimisation of heat exchanger networks which takes fouling into account has included area margin, bypass arrangements, velocity, and operating-cycle cost, thus demonstrating that a network designed solely for clean conditions might be a long way from being optimal throughout its service life. It has also been studied to replace some of the existing shell-and-tube heat exchangers with welded plate heat exchangers as a means of reducing fouling, the velocity distribution, heat transfer, pressure drop, and exchanger location all being optimised together. The main takeaway is not that one type of exchanger is generally better; rather it is that surface shear, wall temperature, cleanability, pressure capacity, and the local tendency to foul should decide where intensification is worthwhile.

With regard to the Saudi facilities, fouling management also has a aspect relating to water usage and availability. The frequent need for mechanical or chemical cleaning uses up labour, water, chemicals, and upsets the scheduled outages. A cleaning strategy which is attractive from a thermal point of view and involves aggressive cleaning may be at variance with water-efficiency or reliability aims. The most beneficial optimisation objective should therefore take into account furnace-fuel penalties, pumping penalties, cleaning cost, the demand for cleaning resources, production risk, and emissions. In this way, fouling would cease to be an afterthought in the maintenance programme and instead become a quantified part of energy management.

5.2 Heat-transfer enhancement and equipment-level optimisation

Equipment enhancement may either reduce the area required or increase the amount that can be recovered, but it should be used selectively. An exergoeconomic optimisation of the shell-and-tube geometry demonstrates that the baffle arrangement, flow rate, area, pressure drop, and cost all have a strong interaction [20]. Similarly, advanced exergy analysis shows that a large portion of the exchanger's irreversibility can be avoided in some cases, which in turn supports making targeted changes to the configuration and operating variables [21]. Heat-transfer enhancement has also been incorporated directly into multi-period HEN synthesis and retrofit, the aim being to reduce utility demand and environmental impact without disregarding the existing equipment [22].

When providing refinery service, it is necessary to balance improvement against the problem of fouling.

Although higher velocities can enhance convection and raise the wall shear stress and thus possibly reduce deposition, they also lead to a greater pressure drop and might worsen erosion or vibration. While compact or plate technologies can provide better heat transfer in a limited area, the actual suitability for use depends on the metallurgy, the sealing arrangements, the inspection capabilities, the pressure rating, and the nature of the feed contaminants [19]. As a result, the most effective retrofit usually involves a moderate amount of equipment improvement together with rematching at the network level, rather than aiming for the highest possible local heat-transfer coefficient.

A good way to carry out a screening is to find out whether the bottleneck lies with the area, the driving force, or cleanliness. In the case of a unit which has sufficient area but suffers from poor duty due to fouling, cleaning or velocity management will then be the main issue. When there is a thermodynamic mismatch in the network, increasing the area of the wrong exchanger brings about only a small advantage. When both the driving force and the area are favourable but the pressure drop is a limitation, a configuration with lower resistance might be better. Using this diagnostic approach stops capital being invested in treating symptoms rather than in addressing the actual constraints.

5.3 Computational and graph-based network optimisation

The larger the refinery networks become, the unfeasible it becomes to carry out a complete manual evaluation of retrofits. Nowadays, various methods employ graph theory, mathematical programming, evolutionary algorithms, and reduced-order formulations in order to search out complex topology spaces. When it comes to refinery retrofitting from a graph-theoretical point of view, it has been shown that it is possible to offer positional swaps as well as several near-optimal solutions rather than insisting on a single mathematically optimal network. This is useful in the case of operating plants, since factors such as constructability, tie-in windows, plot congestion, and the reuse of equipment can cause the second-best theoretical solution to be the actually best project.

Large-scale synthesis frameworks achieve a reduction in computational demand without compromising the main structural decisions. By introducing another level, multi-period algorithms insist on feasibility in various operating states. Moreover, when carrying out an economic assessment under historical variations in energy prices it becomes clear that the ranking of solutions should be subject to stress testing rather than being based on a single price forecast. All of these developments indicate a move towards robust optimisation: the candidate networks should be evaluated over a number of operating periods, fouling states, utility price scenarios, and maintenance constraints.

A new approach has been introduced through the use of graph machine learning. This involves employing a programmed retrofit method which models the HEN as a graph and makes use of thermally guided searches to find promising routes for modification. The advantage goes beyond just faster computation; graph representations reflect the actual physical connections in a network, allowing algorithms to learn structural characteristics that are linked to internal heat recovery. In the context of refinery engineering, the most reasonable application is decision support that reduces the number of retrofit options while still maintaining thermodynamic verification. It would be difficult to place any trust in a completely automated generation of topology without engineering constraints in the case of safety-critical brownfield plants.

5.4 Data-driven fouling monitoring and predictive control

Digital optimisation is most useful when it enhances the state estimate that is used by operations and maintenance. Reports on machine-learning-based fouling monitoring show that there is increasing interest in support-vector methods, nearest-neighbour techniques, recurrent neural networks, and similar predictors, but they also point out the small number of actual industrial applications. The existence of this gap is important since refinery data include sensor drift, changes in operating mode, missing measurements, shifts in crude properties, and maintenance activities which are seldom found in laboratory datasets.

Hybrid modelling provides a practical solution. By combining data generated by CFD with machine learning it is possible to speed up the prediction of fouling resistance throughout the exchanger's lifecycle. Studies involving real-time monitoring of plate exchangers have found that recurrent neural networks are able to detect temporal degradation patterns, though the reliability of the models is still affected by variations in operating conditions and noise. With regard to the refining industry, a 2025 study on the preheat train compared long short-term memory networks, gradient boosting, and a semi-empirical effectiveness model for the purpose of virtual sensing of outlet temperatures; fouling was then inferred from the deviations between the predictions of the clean model and the actual performance.

When assessing digital fouling tools for Saudi refineries, judgement should be based on the quality of the decisions they enable, not just on prediction accuracy. A model featuring a low statistical error is of only limited value if it does not lead to a change in operational practice—for example, by adjusting the flow, checking a sensor, scheduling a cleaning, modifying a bypass, or looking into the effect of a crude-blend. The models in question therefore need to provide uncertainty limits and diagnostic characteristics rather than just a single unclear fouling score. The training data must include examples from both summer and winter conditions, different levels of throughput, changes in the crude oil type, the period after cleaning, and instrument recalibrations. Since decisions regarding production or maintenance could be triggered by predicted degradation, explainability is particularly important in such cases.

5.5 Waste-heat upgrading and low-carbon heat integration

Once the internal heat recovery has been optimised, the remaining low-grade heat may be looked at for upgrading. Research into the use of heat pumps in industrial settings shows that waste heat at moderate temperatures can be raised to useful process temperatures, the level of performance and the climate benefits depending on the source temperature, the requirements of the heat sink, the carbon intensity of the electricity used, and the economic conditions. In the case of refineries, heat pumps are most attractive when there is a steady low-grade heat source and a suitable heat sink that is available for long periods of operation. It is for this reason that their use should be based on source-sink mapping rather than on a technology-driven approach.

The need caused by renewable integration is just as great for system thinking. Studies involving hybrid renewable energy to meet refinery power requirements indicate that solar, wind, biomass, power from the grid, and conventional generation can all be optimised together when economic uncertainty is taken into account. Moreover, multi-period models of refineries show that renewable generation and storage must be sized with variable demand and the need for flexibility in mind [30]. In the case of thermal systems, this means that electrification, heat pumps, electric boilers, renewable electricity, and possibly renewable thermal supply should all be jointly optimised together with the heat exchanger network; otherwise, the new clean energy will merely meet demand which could have been eliminated through better heat recovery.

A hierarchical approach is therefore the preferred sequence of actions. The first step should be to restore both the cleanliness of the exchangers and the functionality of the instrumentation. The second step involves optimising the flow distribution and the operating targets. The third step is to retrofit the HEN in order to improve internal recovery. The fourth step consists of upgrading the waste heat which cannot be avoided. The fifth step is to replace the remaining external energy demand with a lower-carbon energy source. Although it does not mean that all the projects have to wait until the previous stage has been completed, since parallel implementation may be more economical, the sequence does ensure that the baseline against which the size of new low-carbon utilities is determined is thermally credible.

VI. SAUDI VISION 2030 DEPLOYMENT CONTEXT

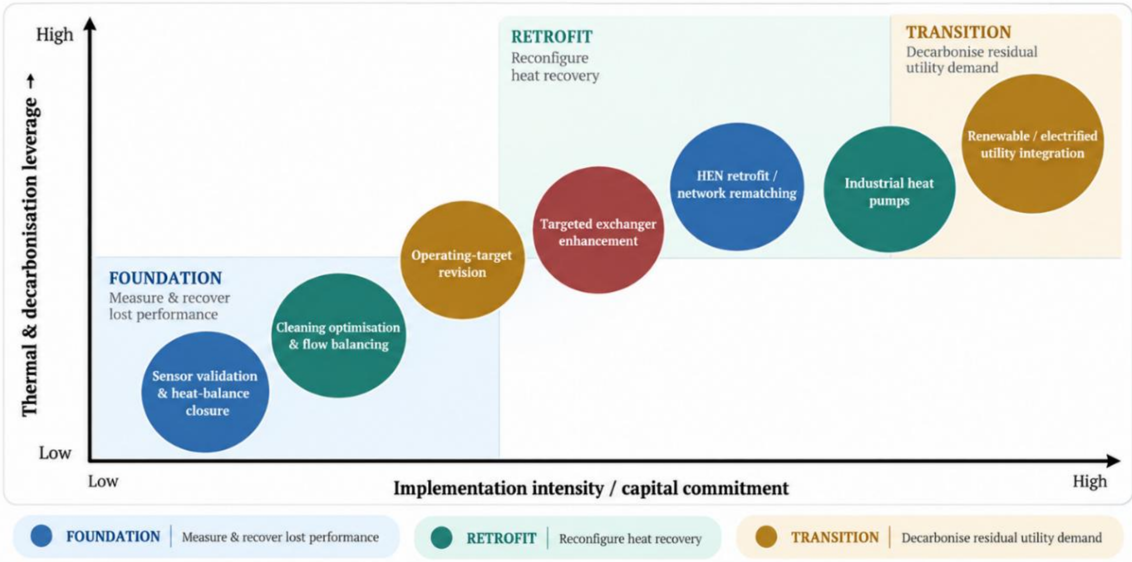
A strategy that is specific to Saudi conditions should take into account the fact that optimising a refinery is limited by climate, the availability of resources, the maturity of the assets, and the need to maintain reliable production. High ambient temperatures can reduce the efficiency of air-cooled heat rejection and lead to a higher demand for auxiliary electricity. Because of the scarcity of water, the need for additional cooling water or for cleaning becomes a matter of strategic importance. Moreover, dust and the severe outdoor conditions can have an impact on instrumentation and on renewable energy assets. At the same time, the flexibility of the crude oil can affect the tendency to foul, the viscosity, and the flow rates of heat capacity. These factors make it more appropriate to use a multi-period optimisation approach that takes uncertainty into account rather than relying on imported designs based on a single set of conditions.

The literature on national transition advocates for combining efficiency with cleaner energy rather than treating these as separate initiatives. For refineries, improvements in energy efficiency have a multiplicative effect in that each unit of heat recovered internally will lead to a reduction in the need for fuel combustion, steam generation, cooling requirements, and in certain configurations also for electrical auxiliary systems. The 2025 review of refinery decarbonisation lists residual heat recovery and digital process optimisation as measures which can be put into place without having to carry out a complete structural replacement. Hence, heat-transfer optimisation is especially compatible with the staged approach to Vision 2030 investments since a number of the actions can be carried out during planned turnarounds or by means of operational changes before the large utility-system projects are started.

Figure 2 presents the evidence in the form of a deployment map; prioritisation should be given to low-capital actions such as sensor validation, cleaning optimisation, flow balancing and operating-target revision in those cases where performance has moved away from design. Medium-capital actions are targeted exchanger replacement, the addition of area, bypass modification and network rematching. Higher-capital measures consist of extensive HEN reconstruction, the use of industrial heat pumps and the integration of renewable or electrified utilities.

Fig. 2. Saudi refinery deployment map: thermal leverage versus implementation intensity

Qualitative synthesis for sequencing decisions; positions show relative deployment logic, not audited project returns



Performance governance is just as important; the heat-management dashboard of a refinery oriented towards Vision 2030 should monitor fired-heater duty on a per unit of throughput basis, the amount of process heat recovered, indicators of exchanger cleanliness, the margin of network pressure drop, cooling demand, the use of cleaning resources, and the amount of carbon emissions avoided. The economic indicators should comprise the discounted lifecycle cost and sensitivity to changes in fuel, electricity, and carbon values rather than simply the payback period.

VII. INTEGRATED DECISION FRAMEWORK

The evidence points to the use of a five-stage decision framework. In the first stage, data credibility is established by means of sensor reconciliation, achieving heat and mass balance closure, and identifying any missing pressure-drop or temperature measurements. The second stage distinguishes between reversible degradation and structural caps by calculating the level of cleanliness, the fouling resistance, the hydraulic margin, and the exchanger effectiveness. The third stage entails assessing a variety of operational interventions, such as flow redistribution, bypass settings, cleaning schedules, and modifications to utility targets. For the fourth stage, the structural retrofit is evaluated using pinch analysis, graph-based methods, or mathematical optimisation over a number of operating periods. Finally, in the fifth stage, heat pumps, renewable energy, and electrified utilities are considered in relation to the reduced residual demand.

Gates should seek to maintain options rather than compel early commitment. For instance, a heat exchanger that has been identified as thermally weak could be cleaned, had its tubes replaced, be substituted with a more compact unit, have additional area added, or be bypassed as a result of a network rematch. The right option to choose will depend on whether the limiting factor is fouling, area, temperature approach, pressure drop, metallurgy, or network topology. In a similar way, a heat pump should not be selected on the basis of its coefficient of performance alone; instead, source stability, sink temperature, electricity price, grid emissions, and the way it integrates with steam levels all have to be taken into account.

The framework also puts digitalisation on a par with engineering decisions. Machine learning should be used to improve state assessment, anomaly detection, and scenario screening, while first-principles balances and thermodynamic limits continue to serve as acceptance tests. Although graph-based methods can speed up the generation of candidate solutions and virtual fouling sensors can be used to prioritise inspections, any changes to the plant must still be checked against hydraulic, mechanical, control, and safety requirements. This approach delivers a realistic route from conventional performance monitoring to semi-autonomous thermal optimisation without the need for an opaque digital system to control a refinery's energy backbone.

8. RESEARCH GAPS AND FUTURE AGENDA

The initial problem concerns industrial validation over long time periods. Although a great number of optimisation studies show promising results in the case of a specific study, even fewer examine performance over multiple fouling and cleaning cycles, throughout various crude campaigns, and under seasonal conditions. It is therefore necessary for upcoming studies in Saudi Arabia to publish anonymised multi-period datasets or validated digital replicas which retain the records of temperature, flow, pressure drop, cleaning, and crude properties. Otherwise algorithms risk overfitting the narrow range of operating conditions.

Second, uncertainty should be incorporated directly into the optimisation model rather than being left to sensitivity analysis. The prices of fuel and electricity, the value assigned to carbon, the rates of fouling, the timing of turnarounds, and the throughput are all uncertain. Robust or random formulations could then be used to identify retrofit solutions that give up only a small portion of the nominal recovery in order to achieve much greater resilience.

Thirdly, water should be regarded as a resource that is optimised together with other factors. Although most thermal studies concentrate on fuel and electricity, the decisions regarding cooling and cleaning can affect water demand. Saudi research should combine the optimisation of heat-exchanger networks with constraints related to the cooling system and the water network, in order that an energy-saving measure may not be accepted if it would result in a greater burden on scarce water resources. The high ambient temperature should be included explicitly since it affects the performance of the cooling system and may change the optimum target for heat recovery.

Fourthly, research in the area of machine learning should focus on explainability and transferability. It has already been noted in recent reviews that there is a lack of real-world industrial applications [24]. A model developed for one exchanger or one crude slate might not generalise to another. Physics-informed features, uncertainty assessment, drift detection, and transfer-learning techniques should therefore be assessed in terms of their impact on maintenance decisions, not just on the prediction metrics. The most effective model could in fact be a simpler one that the operations staff are able to examine and place their trust in.

In conclusion, upcoming research should involve linking the optimisation of the refinery's HEN with the overall decarbonisation strategy of the site. The various initiatives—such as heat recovery, modifications to the steam system, the use of heat pumps, the incorporation of renewable electricity, meeting hydrogen demand, and carrying out carbon-management projects—compete for capital and have thermodynamic interactions with one another.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1723034

Although multi-period planning for refinery operations provides a basic foundation [30], additional research is still required in order to incorporate the degradation of exchangers and the flexibility of heat recovery into those energy-system models. Table 2 gives a summary of the main gaps and the decision outputs needed for a research programme focused on Saudi Arabia.

Table 2. Priority research gaps and decision outputs for Saudi refinery heat-transfer optimisation.

Research gap	Why it matters	Recommended research approach	Required decision output
Long-horizon plant validation	Algorithms may appear effective on short case studies but fail across fouling-cleaning cycles or crude campaigns.	Multi-year, anonymised operating datasets and validated digital replicas that retain temperature, flow, pressure-drop, cleaning and crude-property history.	Evidence of persistent energy savings, reliability and model validity.
Uncertainty-aware investment	Nominally optimal retrofits can be fragile to energy prices, fouling rates, throughput and turnaround timing.	Robust or stochastic multi-period optimisation with lifecycle economics and uncertainty bounds.	A portfolio of retrofit options ranked by resilience, not only nominal payback.
Water-energy coupling	Cleaning or cooling changes can reduce fuel use while increasing scarce-water demand.	Coupled HEN, cooling-system and water-network optimisation under Saudi resource constraints.	Energy savings reported together with water and cleaning-resource impacts.
High-ambient-temperature performance	Summer conditions can shift cooling bottlenecks and auxiliary loads.	Seasonal utility models with realistic air-cooler, cooling-water and electrical constraints.	Retrofit targets that remain feasible during peak summer operation.
Explainable digital fouling models	High predictive accuracy does not guarantee trustworthy maintenance decisions.	Physics-informed features, drift detection, transfer learning, uncertainty quantification and operator-facing diagnostics.	Actionable cleaning, inspection and flow-control recommendations with confidence levels.
Site-wide decarbonisation integration	HEN retrofit, steam, heat pumps, renewable power, hydrogen and carbon-management projects compete for capital.	Integrated multi-period refinery energy-system models that include exchanger degradation and heat-recovery flexibility.	Coordinated 2030 investment pathway with explicit thermal dependencies.

VIII. CONCLUSIONS

The idea of optimising heat transfer in refineries cannot any longer be completely captured by referring merely to exchanger design or pinch analysis. As shown in the literature from 2020 to 2025, there is a clear trend towards making decisions based on the entire lifecycle, on networks, and using data. The pinch and graph methods improve structural heat recovery; multi-period optimisation takes account of variations in operation; fouling-aware scheduling links up maintenance with energy performance; exergy and thermal-hydraulic analysis show the potential for improvements at the component level; and machine learning offers ever more useful capabilities for virtual sensing and screening. Waste-heat upgrading and the combination of renewables then extend the scope of the optimisation from the HEN to the whole site energy system.

The approach outlined in Saudi Vision 2030 is one that is staged yet integrated. The refineries should commence by setting up reliable methods of measurement and recovering performance that has deteriorated, followed by optimising their operations and cleaning processes, then retreating the bottlenecks in the network, and finally sizing the low-carbon heat and power systems in relation to the reduced residual demand. This order maintains reliability is maintained while preventing capital spending on wastefulness that are unnecessary. The key research requirement is validation at plant scale under Saudi conditions, with a particular focus on high ambient temperatures, water limitations, variations in the crude oil feedstock, and long fouling cycles. Only when these factors are incorporated into robust optimisation and supported by clear digital monitoring can the refinery's heat-transfer systems become a concrete platform for reducing energy intensity, lowering emissions, and maintaining industrial competitiveness.

REFERENCES

[1] Seijo-Bestilleiro E, Arias-Fernández I, Carro-López D, Naveiro M. Opportunities for emission reduction in the transformation of petroleum refining. *Fuels.* 2025;6:66. MDPI

[2] Talaei A, Oni AO, Ahiduzzaman M, Roychaudhuri PS, Rutherford J, Kumar A. Assessment of the impacts of process-level energy efficiency improvement on greenhouse gas mitigation potential in the petroleum refining sector. *Energy.* 2020;191:116243. Crossref

[3] Amran YHA, Amran YHM, Alyousef R, Alabduljabbar H. Renewable and sustainable energy production in Saudi Arabia according to Saudi Vision 2030: Current status and future prospects. *Journal of Cleaner Production.* 2020;247:119602. Crossref

[4] Alshammari YM. Achieving climate targets via the circular carbon economy: The case of Saudi Arabia. *C.* 2020;6(3):54. MDPI

[5] Alhajri IH, Gadalla MA, Abdelaziz OY, Ashour FH. Retrofit of heat exchanger networks by graphical Pinch Analysis - A case study of a crude oil refinery in Kuwait. *Case Studies in Thermal Engineering.* 2021;26:101030. Crossref

[6] Lai YQ, Wan Alwi SR, Manan ZA. Graphical customisation of process and utility changes for heat exchanger network retrofit using individual stream temperature versus enthalpy plot. *Energy.* 2020;203:117766. Crossref

[7] Zhi K, Wang B, Guo L, Chen Y, Li W, Ocłoń P, Wang J, Chen Y, Tao H, Li X, Varbanov PS. Graphical pinch analysis-based method for heat exchanger networks retrofit of a residuum hydrogenation process. *Energy.* 2024;299:131538. Crossref

[8] Teng SY, Orosz A, How BS, Jansen JJ, Friedler F. Retrofit heat exchanger network optimization via graph-theoretical approach: Pinch-bounded N-best solutions allows positional swapping. *Energy.* 2023;283:129029. Crossref

[9] Zhao K, Zhao L, Tang QQ, Chen QL, He C, Zhang BJ. A novel optimization framework integrating multiple initialization, automatic topologization and MINLP reduction to accelerate large-scale heat exchanger network synthesis. *Energy.* 2024;307:132508. Crossref

[10] Langner C, Svensson E, Harvey S. A computational tool for guiding retrofit projects of industrial heat recovery systems subject to variation in operating conditions. *Applied Thermal Engineering.* 2021;182:115648. Crossref

[11] Stampfli JA, Ong BHY, Olsen DG, Wellig B, Hofmann R. Applied heat exchanger network retrofit for multi-period processes in industry: A hybrid evolutionary algorithm. *Computers & Chemical Engineering.* 2022;161:107771. Crossref

[12] Halmschlager D, Beck A, Knöttner S, Koller M, Hofmann R. Combined optimization for retrofitting of heat recovery and thermal energy supply in industrial systems. *Applied Energy.* 2022;305:117820. Crossref

[13] Boldyryev S, Gil T, Ilchenko M. Environmental and economic assessment of the efficiency of heat exchanger network retrofit options based on the experience of society and energy price records. *Energy.* 2022;260:125155. Crossref

[14] Hang P, Zhao L, Liu G. Optimal design of heat exchanger network considering the fouling throughout the operating cycle. *Energy.* 2022;241:122913. Crossref

[15] Ishiyama EM, Pugh SJ, Wilson DI. Incorporating deposit ageing into visualisation

of crude oil preheat train fouling. *Process Integration and Optimization for Sustainability.* 2020;4:187-200. Springer

[16] Lozano Santamaria F, Macchietto S. Online integration of optimal cleaning scheduling and control of heat exchanger networks under fouling. *Industrial & Engineering Chemistry Research.* 2020;59(6):2471-2490. ACS

[17] Trafczynski M, Markowski M, Urbaniec K. Energy saving and pollution reduction through optimal scheduling of cleaning actions in a heat exchanger network. *Renewable and Sustainable Energy Reviews.* 2023;173:113072. Crossref

[18] Ishiyama EM, Pugh SJ, Zettler HU. Economic and environmental implications of fouling in crude preheat trains. *Heat Transfer Engineering.* 2024;45(15):1277-1285. Taylor & Francis

[19] Wu Y, Wang Y, Liu R, Feng X. Applying plate heat exchangers in crude preheat train for fouling mitigation. *Chemical Engineering Research and Design.* 2021;165:150-161. Crossref

[20] Jamil MA, Goraya TS, Shahzad MW, Zubair SM. Exergoeconomic optimization of a shell-and-tube heat exchanger. *Energy Conversion and Management.* 2020;226:113462. Crossref

[21] Prajapati P, Raja BD, Savaliya H, Patel V, Jouhara H. Thermodynamic evaluation of shell and tube heat exchanger through advanced exergy analysis. *Energy.* 2024;292:130421. Crossref

[22] Isafiade AJ. Synthesis and retrofit of multiperiod heat exchanger networks considering heat transfer enhancement and environmental impact. *Process Integration and Optimization for Sustainability.* 2023;7:1031-1053. Springer

[23] Chen X, Huang J, Zhang L. A new programmed method for retrofitting heat exchanger networks using graph machine learning. *Applied Thermal Engineering.* 2024;242:122427. Crossref

[24] Villa L, Brusamarello CZ. Application of machine learning in monitoring fouling in heat exchangers in chemical engineering: A systematic review. *Canadian Journal of Chemical Engineering.* 2025;103(4):1786-1801. Wiley

[25] He S, Ye Y, Wang M, Zhang J, Tian W, Qiu S, Su GH. A machine learning and CFD based approach for fouling rapid prediction in shell-and-tube heat exchanger. *Nuclear Engineering and Design.* 2025;432:113759. Crossref

[26] Hou G, Zhang D, Yan Q, Wang S, Ma L, Jiang M. Application of machine learning algorithms in real-time fouling monitoring of plate heat exchangers. *International Communications in Heat and Mass Transfer.* 2025;164:108809. Crossref

[27] Ujević Andrijić Ž, Rimac N. Data-driven fouling detection in refinery preheat train heat exchangers using neural networks and gradient boosting. *Sensors.* 2025;25(16):4936. MDPI

[28] Jovet Y, Lefèvre F, Laurent A, Clausse M. Combined energetic, economic and climate change assessment of heat pumps for industrial waste heat recovery. *Applied Energy.* 2022;313:118854. Crossref

[29] Toghyani M, Saadat A. From challenge to opportunity: Enhancing oil refinery plants with sustainable hybrid renewable energy integration. *Energy Conversion and Management.* 2024;305:118254. Crossref

[30] Lau CYF, Chew YE, How BS, Andiappan V. Planning and optimisation of renewable energy systems for decarbonising operations of oil refineries. *Process Integration and Optimization for Sustainability.* 2025;9:93-116. Springer