Short-Term Planning of Industrial Waste Heat Recovery Using a McCormick-Based MILP-Conic Reformulation

MELIN Baptiste[a], BAUDET Philippe[b], THERY HETREUX Raphaele[c] and HETREUX Gilles[d]

[a] *Laboratoire de Génie Chimique, Toulouse, France, baptiste.melin@proesis.net*
[b] *PROESIS SAS, Toulouse, France, philippe.baudet@proesis.net*
[c] *Laboratoire de Génie Chimique, Toulouse, France, raphaele.thery@inp-toulouse.fr*
[d] *Laboratoire de Génie Chimique, Toulouse, France, gilles.hetreux@ensiacet.fr*

Abstract:

In industry, short-term energy planning aims to align energy consumption with production forecasts to reduce costs, improve reliability, and support decarbonization. However, detailed modelling of enthalpy balances, heat transfer, and thermal storage naturally leads to MINLP formulations that are difficult to solve within operational time constraints. The objective of this work is to extend the MILP model of the *MERLIN* platform [15] by replacing bilinear terms with McCormick envelopes [1,2] and approximating the heat transfer equation $Q = U \cdot A \cdot \Delta TML$ using the Paterson method [4], expressed as a conic constraint for heat exchangers. The resulting mixed-integer convex formulation remains computationally tractable for short-term planning. Its performance is illustrated on a waste heat recovery chain coupling an industrial steam plant with a district heating network, showing computation times compatible with operational use while preserving sufficient accuracy.

Keywords:
District Heating; Energy; Operation; McCormick Envelopes.

1. Introduction

District heating networks (DHNs) play a central role in the decarbonization of thermal energy uses. On the one hand, they enable the pooling of energy demand at the scale of an urban or industrial area, and on the other hand, they allow the recovery of waste heat from industry or waste-to-energy plants, as well as the integration of renewable energy sources into the energy mix. In this context, the operation of these district heating networks can be complex and may require the implementation of high-performance Energy Management Systems (EMS) capable of coordinating, over a short-term horizon, both supply (boilers, cogeneration units, heat pumps, heat recovery systems) and demand, while managing the hydraulic and thermal constraints of the network, as well as potential interactions with energy markets.

Consequently, the operation of urban district heating networks must simultaneously meet several objectives: ensuring that thermal demand from end users is satisfied, minimizing operating costs and CO_2 emissions, and providing the flexibility required for the integration of renewable energy sources. From a modeling perspective, these objectives lead to a complex optimization problem, notably involving energy balances that couple flow rate, temperature, and thermal power. These enthalpy balances (1) in particular, introduce bilinear terms (products of two variables), such as:

$$Q_p = \dot{m} C_p \left(T - T_{ref}\right) \tag{1}$$

where the thermal power Q_p simultaneously depends on the mass flow rate $\dot{m}$ and the temperatures T and T_{ref}. The presence of these nonlinear terms, combined with the introduction of discrete variables related to operating mode selection or equipment activation, naturally leads to a *Mixed-Integer Nonlinear Programming* (MINLP) formulation. However, MINLP problems are known for their high computational complexity, resulting both from the combinatorial explosion associated with integer variables and from the presence of non-convexities. For large-scale systems, these problems become particularly computationally expensive, sometimes intractable, and may also exhibit strong sensitivity to initialization [4]. To overcome these difficulties, our approach consists in reverting to a linear formulation, whose solvers are now well established, by approximating the nonlinearities of the system. This strategy avoids the need for MINLP solvers but requires reformulating the problem by replacing the nonlinear equations with combinations of linear constraints.

2. Approximation of Bilinear Terms

2.1. Approximation of bilinear terms $w_{ij} = y_i.x_j$ using McCormick envelopes

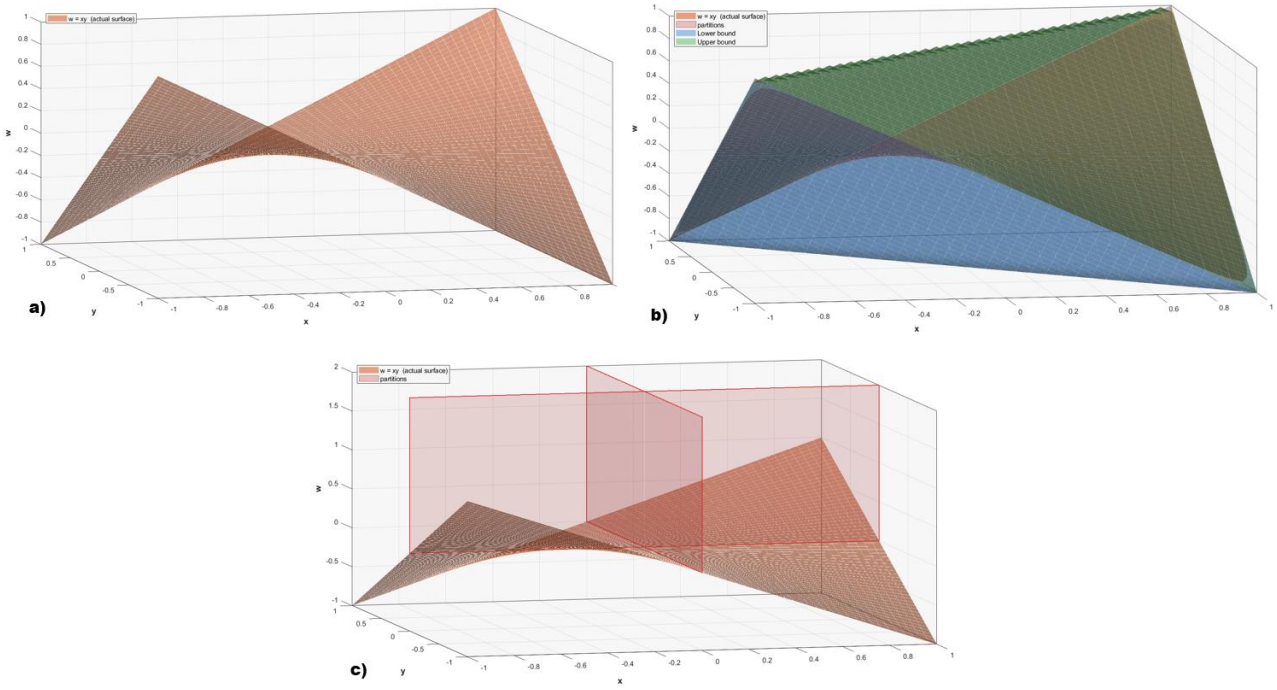

Figure. 1. *a) Surface of the product* $w = xy$ *over* $[-1\,;1] * [-1,1]$ *b) Representation of the convex envelope of the product surface c) Illustration of the partitioning boundaries*

McCormick envelopes are a common approach for relaxing constraints involving bilinear terms. In the standard approach, each bilinear term is replaced by a new auxiliary variable $w_{ij} = y_i.x_j$ and bounded by four linear inequalities (2)–(5), which define the convex envelope of the term over a bounded domain such that $x_j \in [x_j^L; x_j^U]$ and $y_i \in [y_i^L; y_i^U]$ (see Figures 1a and 1b):

$$\forall i \in 1..I, \forall\, j \in 1..J, \qquad \begin{cases} w_{ij} \geq y_i.x_j^L + y_i^L.x_j - y_i^L.x_j^L \\ w_{ij} \geq y_i.x_j^U + y_i^U.x_j - y_i^U.x_j^U \\ w_{ij} \leq y_i.x_j^L + y_i^U.x_j - y_i^U.x_j^L \\ w_{ij} \leq y_i.x_j^U + y_i^L.x_j - y_i^L.x_j^U \end{cases} \qquad (2)-(5)$$

The quality of the relaxation directly depends on the choice of bounds for the variables y_i and x_j, since the maximum error over an envelope $[y_i^L, y_i^U] * [x_j^L, x_j^U]$ is given by relation (6):

$$\forall i \in 1..I, \forall\, j \in 1..J, \qquad \varepsilon_{i,j} = \frac{(y_i^U - y_i^L)(x_j^U - x_j^L)}{4} \qquad (6)$$

2.2. Partitioned McCormick Envelopes

To strengthen the relaxation, one can refine the domain of y_i and/or x_j by dividing it into several subintervals (see Figure 1c). This discretization process introduces domain partitions and naturally leads to formulations in disjunctive programming, in which the global domain is described as the union of several tighter linearized regions.

Three alternatives exist for formulating a linear disjunctive programming model: the *Big-M formulation*, the *convex hull relaxation technique* [5,6] *and indicator constraints*. In this work, we adopted the *Convex Hull formulation* as in [1], in which the variable domain is divided into P partitions, thus allowing the definition of a union of smaller local polyhedra. This formulation relies on the aggregation/disaggregation of variables. Disaggregation consists of decomposing each variable $v \in [v^L; v^U]$ defined over the entire domain, into a sum of local variables $\widehat{v_n} \in [\widehat{v^L}\,; \widehat{v^U}]$, associated with each partition n, as shown in equation (7).

$$\forall i \in 1..I, \forall j \in 1..J, \forall n \in 1..P, \quad w_{ij} = \sum_{n=1}^{P} \hat{w}_{i,j,n} \quad x = \sum_{n=1}^{P} \hat{x}_{j,n} \quad y = \sum_{n=1}^{P} \hat{y}_{i,n} \quad z = \sum_{n=1}^{P} \hat{z}_{i,j,n} \qquad (7)$$

Each partition $n \in 1..P$ has its own local variables $(\hat{x}_{j,n}, \hat{y}_{i,n}, \hat{w}_{i,j,n})$, which are active only if the corresponding binary selection variable $\hat{z}_{i,j,n}$ equals 1, thanks to constraints (8) and (9):

$$\forall i \in 1..I, \forall j \in 1..J, \forall n \in 1..P, \quad \begin{cases} x_{j,n}^{L} \hat{z}_{i,j,n} \leq \hat{x}_{j,n} \leq x_{j,n}^{U} \hat{z}_{i,j,n} \\ y_{i,n}^{L} \hat{z}_{i,j,n} \leq \hat{y}_{i,n} \leq y_{i,n}^{U} \hat{z}_{i,j,n} \end{cases} \qquad (8) - (9)$$

For each partition n, the disaggregated bilinear term $\hat{w}_{i,j,n}$ is bounded by the local McCormick envelopes, defined over the specific bounds $[x_{j,n}^{L}, x_{j,n}^{U}]$ and $[y_{i,n}^{L}, y_{i,n}^{U}]$, as given by equations (10) to (13).

$$\forall i \in 1..I, \forall j \in 1..J, \forall n \in 1..P, \quad \begin{cases} \hat{w}_{i,j,n} \geq \hat{y}_{i,n}.x_{j,n}^{L} + y_{i,n}^{L}.\hat{x}_{j,n} - x_{j,n}^{L} y_{i,n}^{L} \hat{z}_{i,j,n} \\ \hat{w}_{i,j,n} \geq \hat{y}_{i,n}.x_{j,n}^{U} + y_{i,n}^{U}.\hat{x}_{j,n} - x_{j,n}^{U} y_{i,n}^{U} \hat{z}_{i,j,n} \\ \hat{w}_{i,j,n} \leq \hat{y}_{i,n}.x_{j,n}^{U} + y_{i,n}^{L}.\hat{x}_{j,n} - x_{j,n}^{U} y_{i,n}^{L} \hat{z}_{i,j,n} \\ \hat{w}_{i,j,n} \leq \hat{y}_{i,n}.x_{j,n}^{L} + y_{i,n}^{U}.\hat{x}_{j,n} - x_{j,n}^{L} y_{i,n}^{U} \hat{z}_{i,j,n} \end{cases} \qquad (10) - (13)$$

2.3. Bilinear Domain Partitioning Strategies

To minimize the McCormick relaxation error, three domain partitioning techniques have been developed depending on the context of use (see Appendix A).

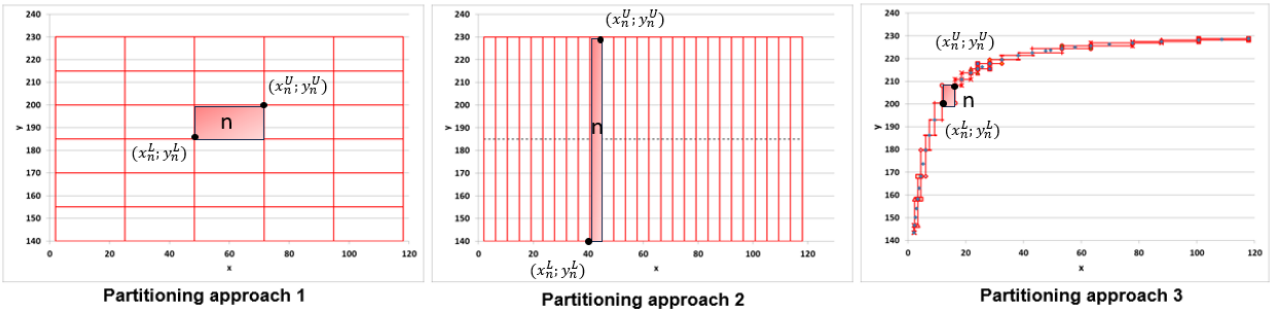

Figure. 2. *Illustration of the partitioning on the xy-plane projection of the convex polyhedron formed by the bilinear term over* $[0\,;120] * [140\,;230]$ *for Algorithms 1, 2, and 3 (from left to right)*

- **Partitioning Approach 1:** The most general case for a bilinear product is the one shown in Figure 1a, where flow rate and temperature can vary independently over a continuous domain, thus defining a surface in space. In this case, a uniform partitioning is applied, as illustrated in Figure 2a, and implemented using Algorithm 1 (Appendix A). This algorithm performs a **bivariate partitioning**, dividing the set along two dimensions (F and T in our case). At each iteration, the dimension (F or T) whose refinement yields the largest error reduction is subdivided first.

- **Partitioning Approach 2:** Implemented via Algorithm 2 (Appendix A), this approach performs a **univariate partitioning**, dividing the set along only one variable (F or T) while keeping the other variable at its global bounds (see Figure 2b). This approach is particularly useful when the goal is to improve accuracy for a specific variable while maintaining a coarser discretization for the other. It is also relevant when one variable has a dominant effect on the error, or when one aims to reduce model complexity while retaining sufficient accuracy.

- **Partitioning Approach 3:** This special case arises when additional physical equations allow the initial domain to be reduced. Instead of a surface of possible solutions, one obtains a unique relationship between F and T, i.e., a curve composed of admissible (F, T) pairs. Algorithm 3 (Appendix A) is applied in this case. It uses a **sequential and adaptive approach** to determine a minimal number of partitions based on simulation or experimental data. The current partition is extended as long as the McCormick relaxation error remains below a specified threshold; a new partition is created once this threshold is exceeded. Unlike classical uniform partitioning (Algorithms 1 and 2), as in [1,7], this approach adapts to the nonlinear shape of the curve. The mesh is refined around relevant areas of the domain, avoiding the introduction of binary variables in physically impossible regions.

Regardless of the approach, a trade-off must be found between accuracy and computation time: increasing the number of envelopes improves the quality of the approximation but also leads to a higher number of binary variables and, consequently, longer solving times. In the case of a uniform mesh, the relaxation error decreases proportionally to $\frac{1}{N_x N_y}$. However, this reduction is significant only for a small number of partitions; beyond a certain threshold, further refinement provides only a marginal gain in accuracy while increasing computation time.

3. Approximation of the Fundamental Heat Equation

3.1 Paterson Approximation

In tools dedicated to the modeling or design of heat exchanger networks, the fundamental heat equation (14) is formulated using the logarithmic mean (15) of the temperature differences.

$$Q = UA\Delta TML \quad \text{with} \quad \Delta TML = \frac{\theta_1 - \theta_2}{\ln\left(\frac{\theta_1}{\theta_2}\right)} \qquad (14) - (15)$$

where:

- U is the overall heat transfer coefficient,
- A is the heat exchange area,
- $\theta_1 = T_{hot,in} - T_{cold,out}$
- $\theta_2 = T_{hot,out} - T_{cold,in}$

This mean introduces a strong nonlinearity into the system of equations, significantly complicating the numerical solution of optimization problems, particularly when multiple temperatures or flow rates are decision variables. Moreover, it has an indeterminate form when $\theta_1 = \theta_2$, as the denominator becomes $\ln(1) = 0$, making the equation undefined (although the limit exists and equals θ_1).

[8] proposes an alternative to simplify the nonlinearity of the equation without significantly sacrificing accuracy. It was observed that the logarithmic mean always lies between the arithmetic mean $AMTD$ (equation 16) and the geometric mean $GMTD$ (equation 17).

$$AMTD = \frac{\theta_1 + \theta_2}{2} \qquad (16)$$

$$GMTD = \sqrt{\theta_1 \theta_2} \qquad (17)$$

Using a Taylor series expansion, Paterson demonstrated that a good approximation of ΔTML can be obtained using the expression (18):

$$\Delta TML_{approx} = \frac{1}{3} AMTD + \frac{2}{3} GMTD \qquad (18)$$

Paterson mathematically compared this approximation with the logarithmic mean by varying the ratio $R = \frac{\theta_2}{\theta_1}$, and showed that the error is negligible in an industrial context. For $R = 2$, the error is practically zero, and for $R = 10$, the error is less than 1%. Moreover, the approximation remains perfectly stable when θ_1 approach θ_2.

3.2 Quadratic Constraint

The main advantage of Paterson's approximation lies in its ability to be reformulated as a system of linear and convex quadratic constraints by separating the arithmetic mean (linear component) from the geometric mean (nonlinear component). Since the geometric mean cannot be directly included as an equality in a convex solver because it is not linear, an auxiliary variable $GMTD$ is introduced, and a rotated quadratic cone constraint (19) is used to bound it:

$$GMTD^2 \leq \theta_1 \theta_2 \qquad (19)$$

because the geometric mean satisfies $GMTD \leq \sqrt{\theta_1 \theta_2}$ his is equivalent to constraint (19) provided that $GMTD \geq 0, \theta_1 \geq 0 \; et \; \theta_2 \geq 0$.

However, to ensure physically correct heat calculation, constraint (19) must be active at the optimum. As noted in [9], if the auxiliary variable $GMTD$ is included in a maximization objective function, the solver will naturally push $GMTD$ toward its upper bound.

4. Case study: District Heating Networks (DHN)

4.1 Evolution of District Heating Networks

Figure 3 [10] highlights the evolution of district heating networks toward lower transport temperatures and more decarbonized energy mixes, from the early coal-fired steam networks (1st generation) to low-temperature networks that heavily use renewables and heat pumps (4th generation). However, the 3rd generation represents a significant turning point: these were the first networks to operate with hot water at temperatures below 100 °C, while significantly integrating the recovery of industrial waste heat and so-called “clean” energies such as solar thermal and biomass. This reduction in temperature also decreases thermal losses across the network and facilitates the integration of lower-temperature heat sources, thereby preparing the transition toward 4th-generation networks.

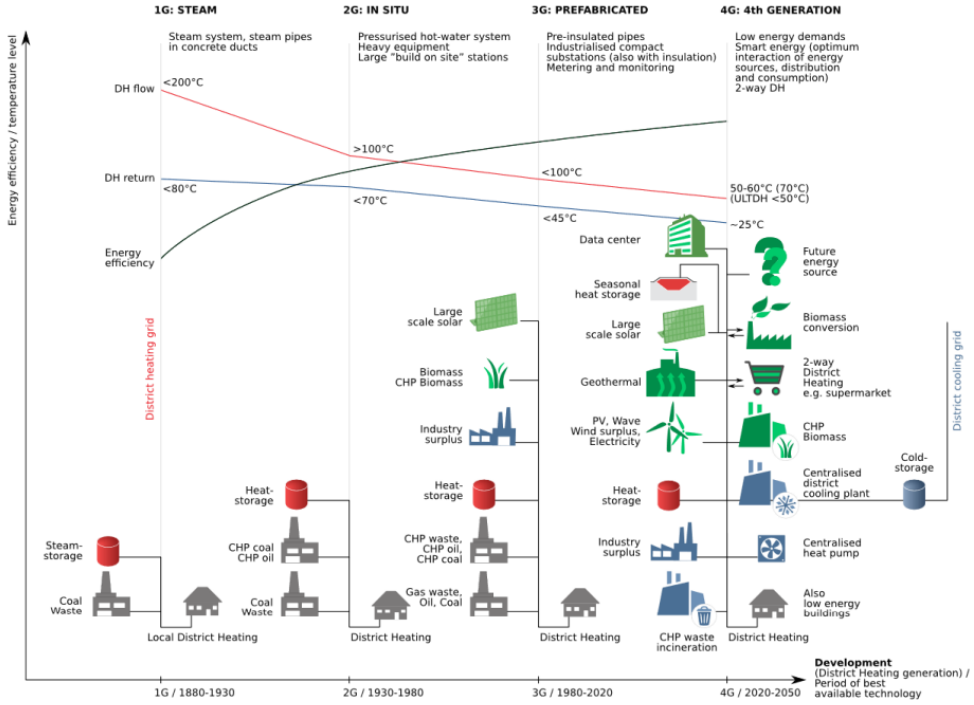

Figure. 3. *Evolution of district heating networks, energy mixes, and temperature levels*

4.2 Control Mode

The operation of a thermal production plant must adapt to fluctuating demand, dictated by consumption profiles and weather conditions. The key operational challenge lies in optimizing the supply temperature: it must remain as low as possible to minimize thermal losses while still being sufficient to meet the needs of the most unfavorable (or critical) point in the network. To address these operational constraints, two control modes for district heating networks [11] are possible:

1. **Fixed Temperatures and Variable Flow:** T_s and T_r are fixed and assumed constant. The total flow rate $\dot{m}$ is then adjusted to ensure that the total power delivered to consumers Q_p meets the demand, according to the relation (20):

$$Q_p = \dot{m} C_p (T_s - T_r) \tag{20}$$

2. **Fixed Flow and Variable Temperatures:** This mode is used particularly when the flow rate reaches its saturation limit in the pipes. Power is then modulated by adjusting the temperature of the heat transfer fluid, according to the relation (21). Unlike the first mode, the storage system can no longer be considered isothermal:

$$T_s = T_{s,min} + \left(T_{s,max} - T_{s,min}\right).\frac{Q_p}{Q_{p,max}} \tag{21}$$

Finally, a mixing operation is introduced at the storage outlet, allowing simultaneous adjustment of both flow rate and temperature to precisely meet thermal demand.

4.3 Case Study Description

In this case study, we focus on a third-generation district heating network (temperature < 100 °C) that includes multiple heat production units as well as short-term thermal storage. The primary production unit is a low-CO_2 biomass boiler, one of the most widely used technologies for DHNs in France [12]. This biomass boiler is sized to cover peak demand. A backup wood boiler is included in the system to provide heat in case of a failure or biomass supply issue. Finally, the network is located near an industrial site that emits waste heat. The heat from the flue gases of the site's natural gas boiler is recovered either via water directly supplied to the DHN or stored in a tank upstream. Since the industrial site's steam demand cannot be controlled, the production of this recovered hot water may not be correlated with the DHN's heat demand, making storage necessary.

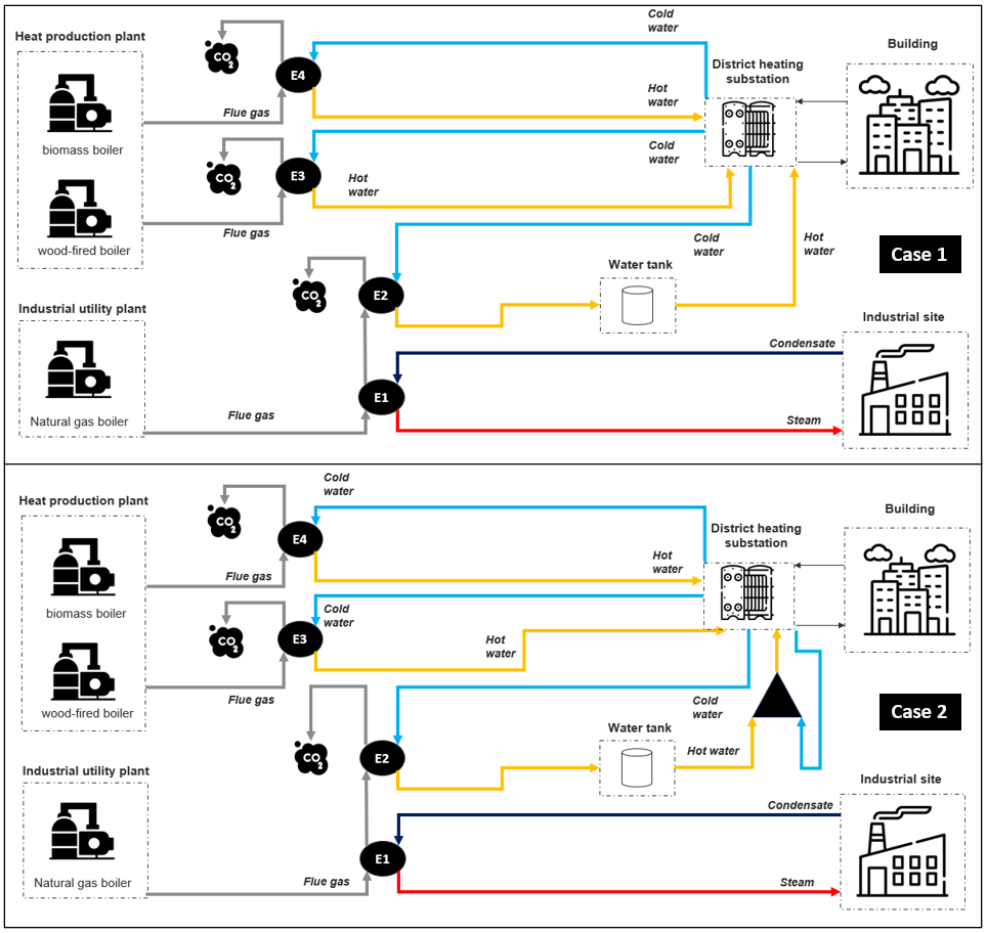

Figure. 4. *District heating plant with constant temperature regulation (Case 1) and variable temperature regulation with mixer (Case 2)*

Although the production units can be coupled in series or in parallel, the parallel configuration is generally preferred in practical applications; it is therefore applied here to the three units under study.

5. Modeling and Parameterization of Equipment

This section presents the models used to represent the various components of the energy system. The main characteristics of the equipment are described, along with the associated mathematical formulations. Special attention is given to areas requiring approximation of bilinear terms and how to optimally partition them.

5.1 Boilers

Three boilers are included in the case study, but the modeling details are illustrated only for the *natural gas boiler*; the other two are modeled in the same way. The natural gas boiler model used in this study is based on the model developed in [13] within *ProSimPlus*. This model has been adapted to ensure a minimum temperature difference in the exchanger ($\Delta Tmin$) greater than 10 °C.

The boiler consists of two main subsystems: the *combustion chamber* and the *flue gas–steam heat exchanger*. The schematic of the boiler is shown in Figure 5, with a focus on the stream managed using McCormick envelopes, where a partitioning strategy according to Algorithm 3 is applied.

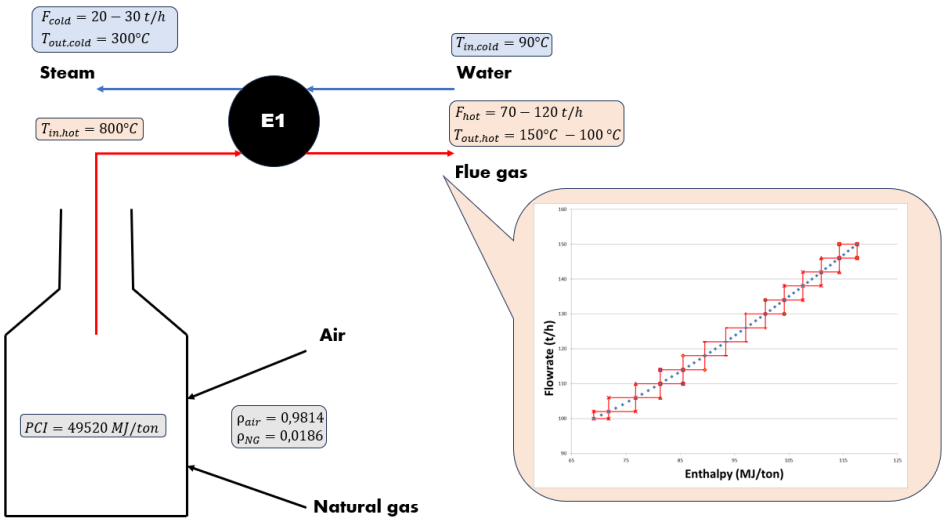

Figure. 5. *Schematic representation of the combustion chamber and the flue gas–steam heat exchanger*

5.1.1 Combustion

The combustion of natural gas produces high-temperature flue gases that feed the heat exchanger. An enthalpy balance on the streams entering and leaving the combustion system leads to the equation (22):

$$h_{flue\ gas} = (h_{NG} + PCI_{NG}).\rho_{NG} + h_{air}.(1 - \rho_{NG}) \tag{22}$$

The mass fraction of natural gas at the combustion inlet ρ_{NG}, is calculated to ensure a fixed outlet temperature from the boiler. The specific enthalpies of the different components $h_{flue\ gas}$, h_{NG}, h_{air} (corresponding respectively to the flue gas, natural gas, and air) are computed using a process simulator.

As a simulation model is not available for biomass and wood boilers, the flue gas exiting the combustion chamber is assumed to be at 150 °C and of the same composition as flue gases from industrial boilers. The specific enthalpies of the flue gas and air are identical to those mentioned above, while the lower heating value (LHV) and the mass fractions differ.

$$h_{flue\ gas} = (h_{wood} + PCI_{wood}).\rho_{wood} + h_{air}.(1 - \rho_{wood}) \tag{23}$$

$$h_{flue\ gas} = (h_{biomass} + PCI_{biomass}).\rho_{biomass} + h_{air}.(1 - \rho_{biomass}) \tag{24}$$

Table 1. Features of boilers.

Material	$h\ \left(\frac{MJ}{ton}\right)$	$\rho(-)$	$PCI(\frac{MJ}{ton})$
Natural gas	-1.066	0.0186	49520
Wood	0	0.0302	7700
Biomass	0	0.0163	14300
Flue gas	232.47		
Air	0.23381		

5.1.2 Flue Gas–Steam Heat Exchanger

The flue gases from combustion transfer their heat to a water stream entering at 90 °C to produce steam in the heat exchanger *E1* (Figure 5). The mass flow rates of the hot fluid F_{hot} and the cold fluid F_{cold} are treated as decision variables in the optimization system. In contrast, $T_{in,hot}$, $T_{in,cold}$ and $T_{out,cold}$ are fixed input data, while $T_{out,hot}$ is a variable determined by the energy balances (Figure 6).

The energy balance on the hot side introduces a bilinear relationship between the variables F_{hot} and $T_{out,hot}$ in the model. This nonlinear constraint therefore requires the implementation of a McCormick approximation. Since it is possible to obtain all admissible (flow, temperature) pairs for this boiler using a *ProSimPlus* simulation, the bilinear relationship is partitioned using Algorithm 3. Figure 7 shows the results for different values of ε_{max}.

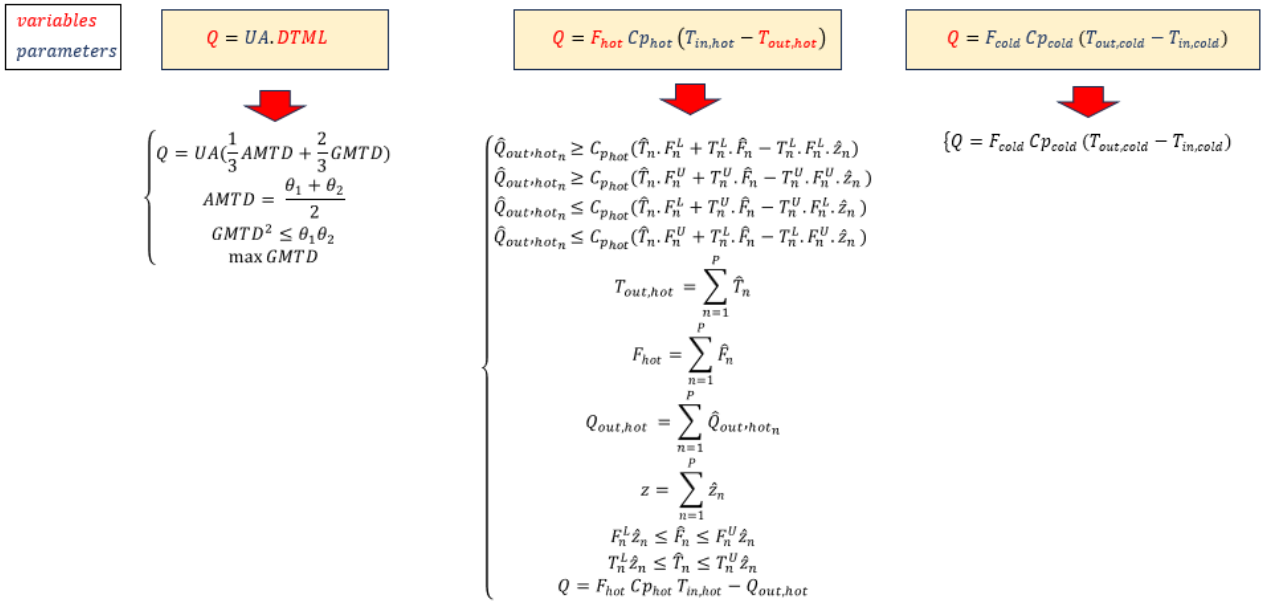

Figure. 6. *Reformulation of nonlinear constraints as a set of linear constraints for the natural gas boiler*

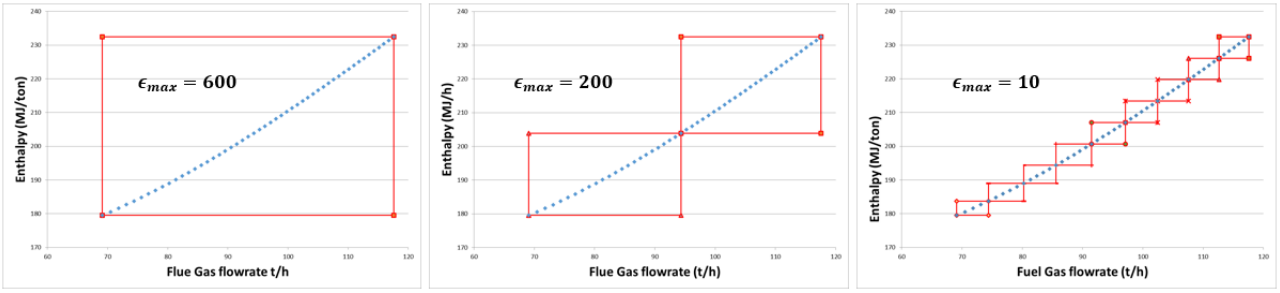

Figure. 7. *Example of partitioning according to the threshold set by the modeler*

5.1.3 Water–Flue Gas Heat Exchangers

Exchangers *E2*, *E3*, and *E4* are water–flue gas heat exchangers. *E3* is connected to the wood boiler, *E4* to the biomass boiler, and *E2* is the heat recovery exchanger. All three have the same UA value of 1755.3 W·K.

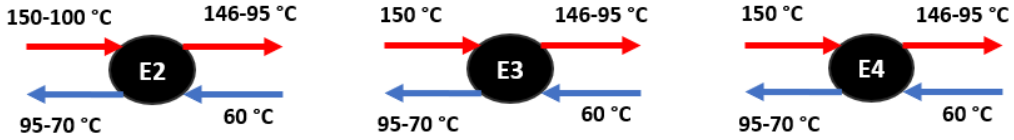

Figure. 8. *Characteristics of the Heat Exchangers*

5.2 Thermal Storage (Hot Water Tank)

The storage system (hot water tank) is integrated into the model using two distinct approaches, depending on the control mode chosen for the district heating network.

5.2.1 Assumptions and Storage Dynamics

The energy dynamics of the storage system are governed by Equation (25) (neglecting thermal losses for this study):

$$M^t T_{storage}^t = M^{t-1} T_{storage}^{t-1} + \dot{F_{in}} T_{in}^t - \dot{F_{out}} T_{storage}^t \tag{25}$$

5.2.2 Behavior According to Regulation Modes

The modeling approach varies depending on the network control strategy:

- ***Mode 1* (Flow Regulation):** The storage is considered at a constant temperature.
- ***Mode 2* (Variable Temperature and Mixing):** The storage operates at a variable temperature. A mixing operation is introduced at the outlet to simultaneously adjust both flow rate and temperature in order to precisely meet the DHN's demand.

5.2.3 Linearization of Bilinear Terms

The discretization equation introduces critical bilinear products that require McCormick relaxation to be integrated into the MIQCP model (Mixed-Integer Quadratically Constrained Programming):

- **Term** $M^t T^t_{storage}$**:** In the absence of a predefined operating curve for the storage temperature, Algorithm 2 (univariate partitioning) is applied to linearize this product, as it involves fewer binary variables than Algorithm 3 (see Figure 9).
- **Term** $\dot{F_{out}} T^t_{storage}$ **:** This product is treated using Algorithm 2 for the same reason as above.

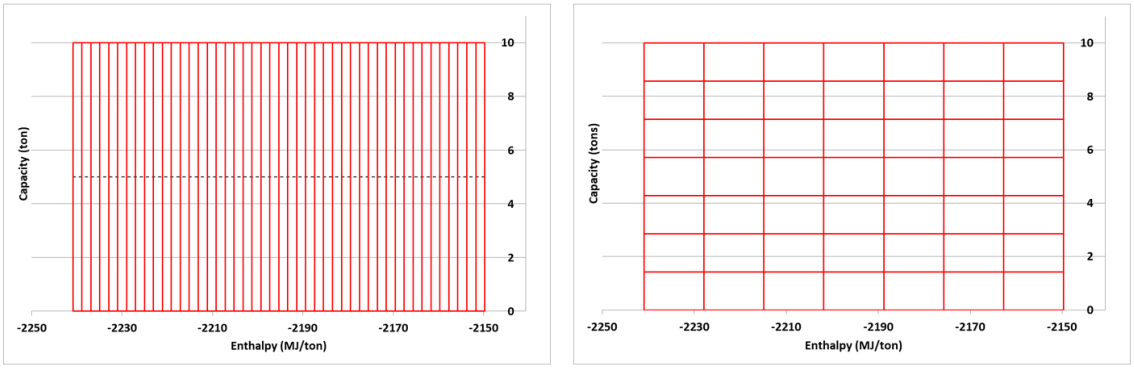

Figure. 9. *Comparison of univariate (left) and bivariate (right) partitioning for* ε_{max} *= 5 (46 envelopes versus 49 envelopes)*

6. Problem Resolution

6.1 Scenarios

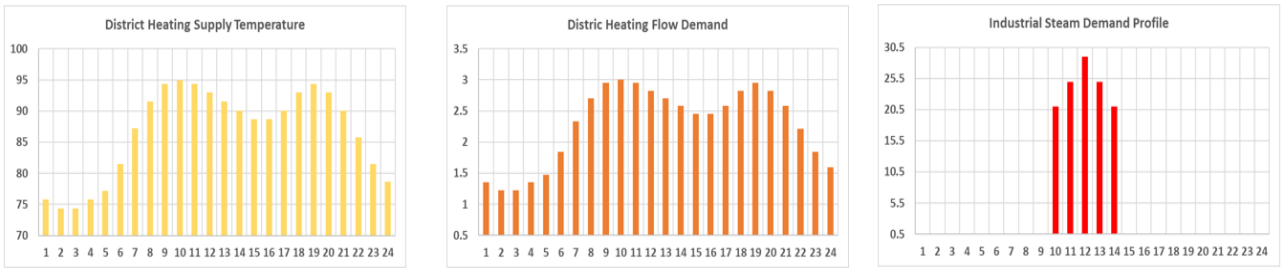

Figure. 10. *a) Supply temperature of the district heating network b) Flow demand of the district heating network c) Profile of industrial steam demand*

To illustrate the model's behavior, a power demand profile representative of real conditions is considered. This profile exhibits a morning peak around 8 a.m. and a second evening peak, corresponding to periods of higher occupancy in residential buildings. In *Mode 1*, the requested power is translated into flow rate (Figure 10b), with supply and return temperatures assumed constant. In *Mode 2*, the flow is maintained constant, and regulation is achieved by adjusting the source temperature (Figure 10a). The return temperature is assumed constant and fixed in both configurations.

Additionally, a constraint is imposed on the supply temperature, with a minimum value set at 70 °C. Finally, the industrial steam demand is considered variable but limited to the time window between 10 a.m. and 2 p.m. (Figure 10c). All bilinear products have been partitioned with a maximum threshold value of $\boldsymbol{\varepsilon_{max}} = \mathbf{5}$. A sensitivity analysis was performed but is not the focus of this study.

6.2 Results for Mode 1

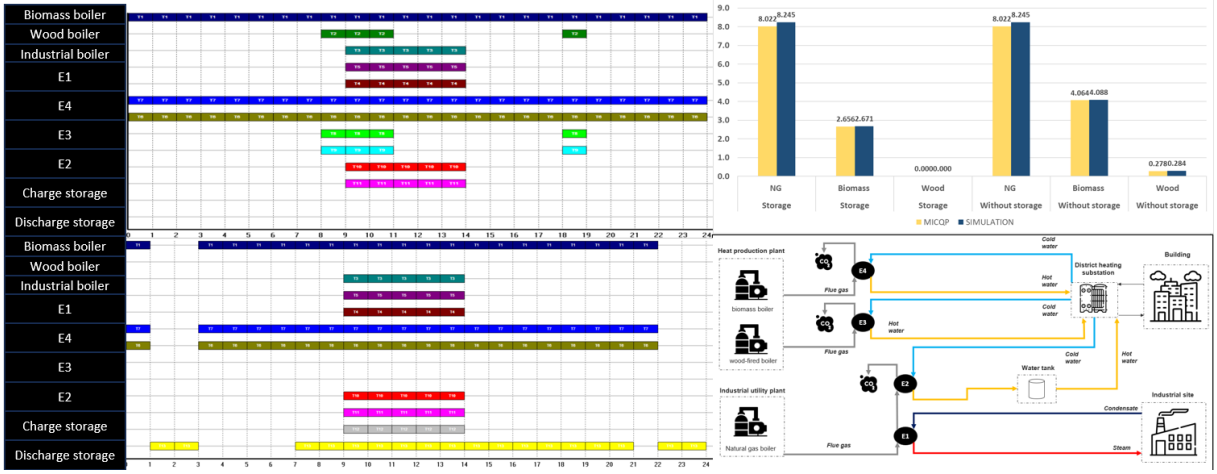

Figure. 11. *Gantt diagram of the boiler plant and fuel consumption*

The comparative analysis of the operational scenarios presented in Figure 7 highlights the crucial role of thermal storage. In a configuration without storage, the wood boiler must intervene as a supplement to the biomass boiler to cover heat demand peaks, particularly during the morning peak at 8 a.m. and in the early evening. Integrating a storage unit allows effective peak shaving: it completely eliminates the load on the wood boiler and achieves a substantial saving of approximately 1.5 tons of biomass. At the same time, natural gas consumption remains stable, as its use is strictly dedicated to meeting industrial steam demand.

The reliability of this approach is confirmed by the convergence between the model results and the simulation. The computation time is around 30 seconds in both cases, which is incomparable to the solution times for MINLP models. The low relative error observed validates the relevance of the adopted methodology.

However, some limitations have been identified. The approximation may underestimate or overestimate the amount of water charged into the storage. Consequently, any error may accumulate over successive periods, potentially leading to an unreliable production plan.

6.3 Results for Mode 2

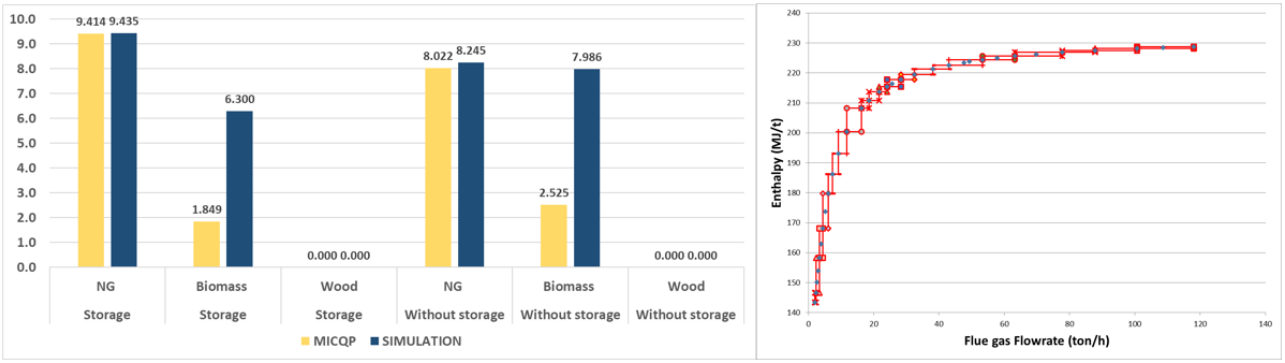

Figure. 12. *Amount of energy input and the relationship between enthalpy and cooled flue gas flow rate*

A significant underestimation of biomass consumption is observed. Indeed, the presence of a plateau in the previous figure shows that, for a nearly identical demand temperature, the flow rate can vary substantially. As a result, the problem relaxation and associated errors allow the optimizer to converge to a solution with a flue gas flow — and therefore biomass consumption — lower than in reality. An improvement could be achieved by increasing the number of envelopes, at the expense of higher computational cost. Another approach would be to rerun the calculation, concentrating a greater number of envelopes around the previously identified point. Furthermore, natural gas consumption increases, as the optimizer favors steam production to raise the temperature of the stored water and optimize subsequent use of the storage.

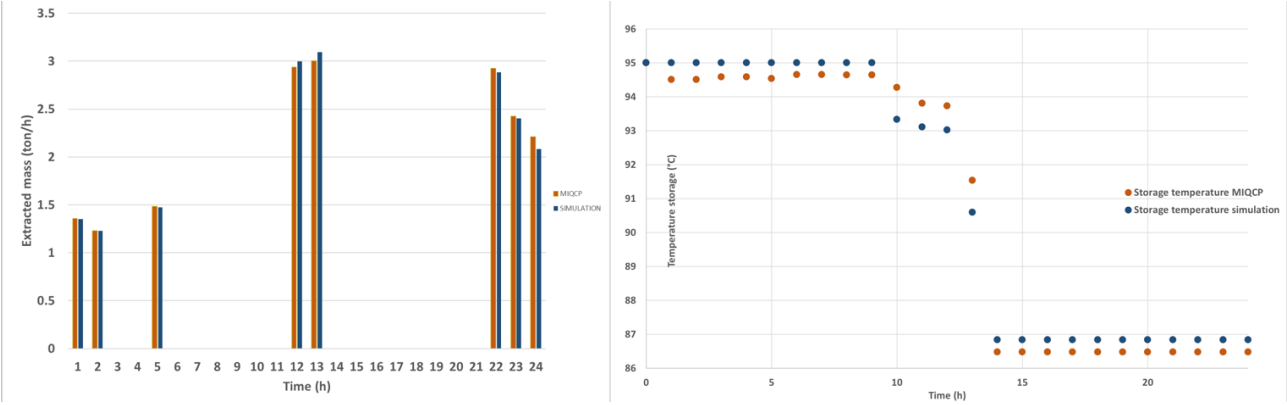

Figure. 13. *Amount extracted from storage and storage temperature over the time horizon*

The relative error on the storage temperature remains below 1 % in this scenario. To meet the temperature demand, the optimizer adjusts the outlet flow from the storage (figure on the left) at a variable temperature, and then performs a mixing operation with water at 60 °C, which also results in a very small error.

The model with variable-temperature storage was stopped after 2 hours of computation, with a gap of 2.65 % (the first feasible solution was obtained after 20 minutes). It was decided not to further refine the partitioning, since the biomass boiler is located on a circuit independent of the storage. The optimizer can therefore activate this boiler, and a recalculation via simulation remains possible.

7. Conclusion and Perspectives

The main contributions of this work lie in the modeling of variable-temperature storage using McCormick approximations, as well as in the introduction of a quadratic constraint to represent a geometric mean, applied to the modeling of heat exchangers.

The results of this study show that the error associated with McCormick envelopes strongly depends on the variable bounds and the level of partitioning. This approach is particularly effective when intervals are narrow but becomes more challenging to control as they widen. Furthermore, precise quantification of the error remains difficult, as it is generally bounded without any guarantee on its exact value. In the case of storage, the error on the flow from the heat recovery exchanger varies depending on steam production, reinforcing the dependence on the operating point. Nevertheless, judicious partitioning of critical zones, such as storage, allows obtaining feasible solutions with error levels compatible with industrial applications.

Computation times differ significantly between the two studied modes: *Mode 2*, which includes more free temperature variables, induces greater complexity due to the resolution of the quadratic part, making the problem more challenging for the solver. The increase in the number of binary variables also constitutes a major limitation for large-scale problems or problems with multiple bilinear nonlinearities. To overcome this difficulty, one potential approach is to use bound contraction methods, as in [14], which can reduce the number of binary variables while progressively tightening the search space around relevant solutions.

Despite these challenges, this approach remains competitive and promising compared to classical MINLP solvers. Indeed, our MINLP model was validated over a 5-period horizon, but for 24 periods, no solution is found in either mode for a computation time below 2 hours.

Acknowledgments

We would like to thank PROESIS SAS and ANRT for the financial support

References

[1] Castro PM. Tightening piecewise McCormick relaxations for bilinear problems. Computers & Chemical Engineering. 2015;72:300–311.

[2] McCormick GP. Computability of global solutions to factorable nonconvex programs: Part I — Convex underestimating problems. Mathematical Programming. 1976;10(1):147–175.

[3] Pettersson F. Heat exchanger network design using geometric mean temperature difference. Computers & Chemical Engineering. 2008;32(8):1726–1734.

[4] Lee J, Leyffer S. Mixed Integer Nonlinear Programming. Springer Science & Business Media; 2011.

[5] Balas E. Disjunctive Programming and a Hierarchy of Relaxations for Discrete Optimization Problems. SIAM J on Algebraic and Discrete Methods. 1985;6(3):466–486.

[6] Grossmann IE, Trespalacios F. Systematic modeling of discrete-continuous optimization models through generalized disjunctive programming. AIChE Journal. 2013;59(9):3276–3295.

[7] Karuppiah R, Grossmann IE. Global optimization for the synthesis of integrated water systems in chemical processes. Computers & Chemical Engineering. 2006;30(4):650–673.

[8] Paterson WR. A replacement for the logarithmic mean. Chemical Engineering Science. 1984;39(11):1635–1636.

[9] Alizadeh F, Goldfarb D. Second-order cone programming. Math Program, Ser B. 2003;95(1):3–51.

[10] Lund PD, Lindgren J, Mikkola J, et al. Review of energy system flexibility measures to enable high levels of variable renewable electricity. Renewable and Sustainable Energy Reviews. 2015;45:785–807.

[11] Haurant P. Exploitation de la donnée énergétique. Nantes: Nantes Université; 2025.

[12] Elhafaia M. Réseau concept : contribution à l'optimisation dynamique des réseaux de chaleur incluant le stockage thermique [Internet] [Theses]. Université de Pau et des Pays de l'Adour; 2024 [cited 2026 Jan 7]. Available from: https://theses.hal.science/tel-04760751.

[13] Théry Hétreux R, Hétreux G, Floquet P, et al. The energy Extended Resource Task Network, a general formalism for the modeling of production systems:Application to waste heat valorization. Energy. 2021;214:118970.

[14] Optimal Operation of Integrated Heat and Electricity Systems: A Tightening McCormick Approach. Engineering. 2021;7(8):1076–1086.

[15] Mallier L, Hétreux G, Thery-Hétreux R, et al. A modelling framework for energy system planning: Application to CHP plants participating in the electricity market. Energy, 2021;214:118976.

Appendix A

Algorithm 1. Uniform Grid Partitioning of a Bilinear Domain

1: Inputs:

$x^{min}, x^{max}, y^{min}, y^{max}$: variable bounds
ε_{max}: maximum allowable relaxation error (absolute)

2: Initialization:

$N_x \leftarrow 1, N_y \leftarrow 1$
$\Delta x = x_{max} - x_{min}, \Delta y = y_{max} - y_{min}$

3: repeat

4: Compute McCormick relaxation error over the uniform grid:

$$\varepsilon = \frac{(x_{max} - x_{min})(y_{max} - y_{min})}{4 \cdot Nx \cdot Ny}$$

If $\varepsilon \leq \varepsilon_{max}$ Then

Exit loop

5: else — increment the dimension yielding the greatest error reduction:

If $\frac{\Delta x \Delta y}{4(Nx+1)\cdot Ny} < \frac{\Delta x \Delta y}{4 Nx (Ny+1)}$ $\quad then\ N_x \leftarrow N_x + 1\ else\ N_y \leftarrow N_y + 1$

6: end if

7: until criterion met

8: Output: Set of $N_x \times N_y$ sub-rectangles $[x^k_{min}, x^k_{max}] \times [y^k_{min}, y^k_{max}]$ defined by:

$$x^k_{min} = x_{min} + (k-1)\frac{\Delta x}{N_x}, x^k_{max} = x^{min} + k\frac{\Delta x}{N_x}, \quad k = 1,..,N_x$$

$$y^l_{min} = y_{min} + (l-1)\frac{\Delta y}{N_y}, y^k_{max} = y^{min} + l\frac{\Delta y}{N_y}, \quad l = 1,..,N_y$$

Algorithm 2. Uniform Grid Partitioning — Single Variable Partitioning

1: Inputs:

$x^{min}, x^{max}, y^{min}, y^{max}$: variable bounds
ε_{max}: maximum allowable relaxation error (absolute)

2: Initialization:

$N_x \leftarrow 1, N_y \leftarrow 1$ (fixed)
$\Delta x = x_{max} - x_{min}, \Delta y = y_{max} - y_{min}$

3: repeat

4: Compute McCormick relaxation error over the uniform grid:

$$\varepsilon = \frac{(x_{max} - x_{min})(y_{max} - y_{min})}{4 \cdot Nx}$$

5: if $\varepsilon \leq \varepsilon_{max}$ Then

Exit loop

6: else

$N_x \leftarrow N_x + 1$

7: end if

8: until criterion met

9: Output: Set of $N_x \times N_y$ sub-rectangles $[x^k_{min}, x^k_{max}] \times [y^k_{min}, y^k_{max}]$ defined by:

$$x^k_{min} = x_{min} + (k-1)\frac{\Delta x}{N_x}, x^k_{max} = x^{min} + k\frac{\Delta x}{N_x}, \quad k = 1,..,N_x$$

y_{min} *and* y_{max} remain constant across all partitions

Algorithm 3. Adaptive Sequential Partitioning Algorithm

1: Inputs: $P = \{(x_i, y_i)\}\ \forall i \in [1\,;N]$ *(ordered data points)*,
ε_{max} *(maximum allowable relaxation error)*
2: Initialization: $k \leftarrow 1$ (partition index);

$$x^{min}_k \leftarrow min(x_1, x_2), x^{max}_k \leftarrow max(x_1, x_2)$$

$$y^{min}_k \leftarrow min(y_1, y_2), y^{max}_k \leftarrow max(y_1, y_2)$$

3: for $i = 3$ ***to*** N **do**
4: Save current bounds: $\widehat{x^{min}}, \widehat{x^{max}}, \widehat{y^{min}}, \widehat{y^{max}} \leftarrow x^{min}_k, x^{max}_k, y^{min}_k, y^{max}_k$
5: Update temporary bounds with point i :

$$x^{min}_k \leftarrow min(x^{min}_k, x_i), x^{max}_k \leftarrow max(x^{max}_k, x_i)$$

$$y^{min}_k \leftarrow min(y^{min}_k, y_i), y^{max}_k \leftarrow max(y^{max}_k, y_i)$$

6: Calculate theoretical McCormick error:

$$\epsilon = \frac{(x^{max}_k - x^{min}_k)(y^{max}_k - y^{min}_k)}{4}$$

7: If $\epsilon > \epsilon_{max}$
8: Store partition k using saved bounds $(\hat{x}, \hat{y})$;
9: Start new partition: $k \leftarrow k + 1$;
10: Initialize partition k with points $i-1$ and i :

$$x^{min}_k \leftarrow min(x_{i-1}, x_i), x^{max}_k \leftarrow max(x_{i-1}, x_i)$$

$$y^{min}_k \leftarrow min(y_{i-1}, y_i), y^{max}_k \leftarrow max(y_{i-1}, y)$$

11: end if
12: end for
13: Output: Set of k partitions defined by $[x^{min}_k, x^{max}_k]$ and $[y^{min}_k, y^{max}_k]$