PROGRAMA DE DOCTORADO EN
INGENIERÍA INDUSTRIAL

TESIS DOCTORAL:

ADVANCED DECISION SUPPORT THROUGH REAL-TIME OPTIMIZATION IN THE PROCESS INDUSTRY

Presentada por **María P. Marcos Núñez**
para optar al grado de
Doctor/a por la Universidad de Valladolid

Dirigida por:
Dr. César de Prada Moraga
Dr. José Luis Pitarch Pérez

This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 723575 (CoPro), from the Spanish MICIU with FEDER funds under projects DPI2015-70975-P and PGC2018-099312-B-C31, and by the regional government of *Castilla y León* plus the EU-FEDER (CLU 2017-09 and UIC 233).

The author thanks the European Social Fund and the *Consejería de Educación de la Junta de Castilla y León*.

A mi yaya

Agradecimientos

A mis directores de tesis, César y José Luis, por todo lo que me han enseñado y por guiarme en este largo camino. Quiero agradecer especialmente a César por confiar en mí aun cuando sólo era una alumna de máster, y a José, por todo el tiempo que me ha dedicado, especialmente todas esas horas evaluando el trabajo realizado con ojo crítico para que todo quede lo mejor posible.

A todos los miembros del grupo de investigación en Control y Supervisión de Procesos, profesores y compañeros de trabajo, tanto los que se han ido como las nuevas incorporaciones, por estar siempre disponibles para aportar ideas y discutir sobre ellas, por agregar nuevos puntos de vista y ayudarme cuando lo he necesitado. Especialmente quiero agradecer a los tres vallisoletanos, Cristian, Suní y Pedro por acogerme en su ciudad y crear un ambiente en el que querer quedarse a pesar de estar lejos de casa.

A mi familia, en especial a mis padres Julio y Paloma, y mi hermana, Rocío, porque nunca habría llegado aquí sin su apoyo y esfuerzo para que yo me dedique a lo que quiera, sin importar si no entienden nada. Por intentar entenderlo.

Al amor de mi vida, Pablo, por estar cada día a mi lado y no dejar que me hunda nunca.

Summary

English

In the process industry, efficiency improvements can come from two main courses of action: replacing older plants, equipment, or processes by more modern and efficient ones; or being more efficient with the current facilities by looking closer at the daily operation, instead of making larger investments with uncertain payback horizons. Focusing on that second way, one can realize that nowadays decision making is conceptually more complex than it was in the past since the rapidly growing technology and communication systems have spawned a large number of alternatives from which a decision maker can select. In addition, a wrong or suboptimal decision with the structural complexity of current problems often results in a magnification of costs along the value chain. Despite that, actual deployment of advanced Decision Support Systems (DSS) remains atypical in process industries since they require big efforts in terms of model development and maintenance, complex mathematical formulations, demanding computational requirements, and integration with the existing control or planning infrastructure.

This thesis contributes to reduce the above barriers by developing efficient formulations for Real-Time Optimization (RTO) in an industrial site. In particular, to improve the operation of three interconnected sections of an

viscose-fiber production facility: an evaporation network, a cooling system one, and a heat-recovery network.

The decisions to take in these systems include the assignment of products to units, the distribution of continuous production among units, and the allocation of shared resources among units. All of these need to consider many constraints due to the actual plant layout, production demands, equipment capacities, environmental conditions and the state of fouling, key aspect in the equipment efficiency. However, it is not enough to formulate the optimization problems, but these must be the core of a tool used by plant operators. Thus, it must be taken into account that these operators are not expert control engineers, so the results must be presented in a way that can be easily and quickly understood.

Another aspect that this thesis deals with is the importance of coordinating the operation of the different sections of the factory, since the systems are coupled, i.e., the operation of one network depends on the operation of the other one and vice versa.

Finally, in any industry, process revamping and plant upgrades over time are very common. However, carrying out such re-designs efficiently is quite complex. The right way to approach the process re-design is to investigate the different possible alternatives together with the future optimal operation, and then to formulate a model which represents the behavior of the plant/network for all the design possibilities, i.e., a superstructure incorporating the different configurations of new equipment. In this case, the cooling-water network could be enhanced incorporating heat pumps.

Objectives

Therefore, the particular objectives of this thesis are: to develop models that represent the operation of each section; to formulate optimization problems using these models to minimize the cost of operation in real time, to develop data driven models which explicitly consider the state of fouling; to develop interfaces for the visualization of the results; to study how to coordinate the operation of the sections; and to obtain the optimal redesign of the cooling-water network.

Methodology

To address the objective of optimal operation of the different sections, the methodology followed has consisted in studying the systems and developing mathematical models that combine discrete decisions with continuous ones, simultaneously considering operation and maintenance tasks. Plant models for optimization are built upon first principles, but they are completed with data-driven sub-models based on available plant data for those relations among variables that are complex to model by physics. Data reconciliation and constrained-regression methods are employed to face the existence of unreliable plant measurements in the development of the data-driven models.

Once the models that represent the systems and their different forms of operation have been obtained, different optimization problems have been formulated. These problems are based on the developed models to which an objective function has been incorporated (in addition to different logical restrictions) that minimizes the operating cost of the corresponding system. In this way, an optimization problem has been obtained capable of establishing the values that the process variables must take so that the operating cost is minimal. It should be noted that the developed models feed on the conditions of the plant, so that the solution obtained depends on these conditions. In addition, the developed formulation allows obtaining results in short computation times, thus allowing their application in real time. Hence, by feeding the optimization with the real-time conditions, an RTO system capable of predicting the optimal operation in real time is obtained.

In addition, different parameters have been included in the developed models to take into account the fouling effects that reduce the equipment efficiency over time. In this way, the RTOs can suggest the cleaning operations that seem more beneficial from an economic point of view, while respecting operation constraints.

Once the RTO problems have been formulated, two decision support systems (DSS) have been developed that help operators in the decision-making process on how to operate the different sections of the factory. Each DSS has as its core one of the developed RTO problems and includes an interface for the presentation of information. These interfaces have been developed tailored to the case study and operators preferences, that lead to a symbiosis of people and computer-based algorithms.

To study the coordination of the operation of the different sections, three different approaches have been analyzed. The first approach is to solve the coupled optimization problems iteratively, the second is to formulate a new optimization problem centralized, and the final fashion is a distributed approach based on Lagrangean decomposition.

In order to obtain the optimal redesign of the cooling network, a superstructure has been developed that not only contemplates all the possible redesigns of the network, but also takes into account the payback time of the purchased equipment. In addition, as predicted benefits depend on the future operation conditions that are somehow uncertain, a two-stage stochastic optimization framework is employed. Thus, a new optimization problem has been formulated based on the superstructure and whose solution indicates what would be the optimal design from an economic point of view. However, as the resulting problem is computationally hard to solve in a monolithic fashion, a decomposition method has been proposed.

Results and conclusions

The results obtained after executing the optimization problems with historical plant data show that if the system variables had been modified according to the solutions provided by the optimizers, significant savings would have been obtained. Of course, as has been proven with historical data, such savings depend on the historical data and how it was acted at the time, but in general, it can be concluded that the developed RTO formulations are very useful to help decision making by indicating the optimal operation. In addition, the results also show that the current policy of cleaning the dirtiest is not the most profitable from an economic point of view, since it depends on many factors. Thus, the developed RTOs also help configuring maintenance task schedule.

The results obtained solving the problems independently versus coordinated verify that the solutions obtained when the problems are solved independently are local, so any of the coordination approaches obtains better results. Furthermore, the comparison of the approaches shows that for this problem size, the centralized formulation is the most suitable.

Analyzing the solution obtained for the network re-design, the conclusion is that the cost of operating the network decreases when the number of heat pumps used increases, up to an extent. Therefore, the number of pumps to buy depends on the payback time of the investment that the company considers acceptable, regardless of the value of such investment. It is worth mentioning that the optimal number of heat pumps to use and their arrangement depend on each scenario.

Finally, it should be noted that this thesis is the result of several years of collaboration between academic and an industrial company, dealing with problems of a real case study. However, although the different technologies developed are tailored for a specific industrial plant, they could easily be adapted and extended to any industry with similar equipment and problems.

Español

En la industria de procesos se puede obtener un aumento de la eficiencia de las plantas de producción, bien mediante la sustitución de procesos o equipos antiguos por otros más modernos y eficientes, o bien operando de forma más eficiente las instalaciones actuales en lugar de realizar grandes inversiones con tiempos de amortización inciertos. Si nos centramos en esta segunda línea de acción, hoy en día la toma de decisiones es conceptualmente más compleja que en el pasado, debido al rápido crecimiento que ha tenido la tecnología últimamente y a que los sistemas de comunicación han generado un gran número de alternativas entre las que se ha de elegir. Además, una decisión incorrecta o subóptima, con la complejidad estructural de los problemas actuales, a menudo resulta en un aumento de los costes a lo largo de la cadena de producción. A pesar de ello, el uso de sistemas de apoyo a la toma de decisiones (DSS) sigue siendo atípico en las industrias de procesos debido a los esfuerzos que se requieren en términos de desarrollo y mantenimiento de modelos matemáticos y al desafío de formulaciones matemáticas complejas, los exigentes requisitos computacionales y/o la difícil integración con la infraestructura de control o planificación existente.

Esta tesis contribuye en la reducción estas barreras desarrollando formulaciones eficientes para la optimización en tiempo real (RTO) en una planta industrial. En particular, esta tesis busca mejorar la operación de tres secciones interconectadas de una fábrica de producción de fibra de viscosa: una red de evaporación, una de sistema de enfriamiento y una red de recuperación de calor.

Las decisiones que se han de tomar en estos sistemas incluyen la asignación de productos a equipos, la distribución de la producción entre equipos paralelos y la distribución de recursos compartidos. Todo ello considerando un alto número de restricciones debido a la configuración de las diferentes redes, las demandas de producción, capacidades de los equipos y condiciones medioambientales, en especial considerando el estado de

ensuciamiento, aspecto clave en la eficiencia de los equipos, Además, no basta con formular los problemas de optimización, sino que estos deben ser la base de una herramienta que pueda ser utilizada por los operarios de la planta, teniendo en cuenta que estos no son expertos ingenieros de control y, por consiguiente, los resultados deben presentarse de manera que se puedan entender fácil y rápidamente.

Otro de los aspectos que trata esta tesis es la importancia de coordinar el funcionamiento de las diferentes secciones de la fábrica, ya que los sistemas están acoplados y, por tanto, la operación de una red depende de la operación de otra y viceversa.

Por último, en cualquier industria, la renovación de procesos y el rediseño de secciones en una planta industrial son muy comunes. Sin embargo, llevar a cabo estos rediseños de manera eficiente es bastante complejo. Para ello, es necesario investigar las diferentes opciones posibles, junto con la futura óptima operación, y formular un modelo matemático que represente el funcionamiento de la red para todas las posibilidades del rediseño, es decir, una superestructura que incorpore diferentes configuraciones de los equipos a incorporar. En este caso, la red de agua de refrigeración podría mejorar su eficiencia incorporando bombas de calor.

Objetivos

Así pues, los objetivos particulares que esta tesis se marca son: desarrollar modelos que representen las diferentes formas de operación de cada sección; formular problemas de optimización usando dichos modelos para minimizar el coste de operación en tiempo real; desarrollar modelos basados en datos para considerar de manera explícita el estado de ensuciamiento de los equipos; desarrollar interfaces para la visualización de los resultados; estudiar la forma de coordinar la operación de las secciones; y obtener el rediseño óptimo de la red de agua de refrigeración.

Metodología

Para hacer frente al objetivo de la operación óptima de las diferentes secciones, la metodología seguida ha consistido en estudiar los sistemas y desarrollar modelos matemáticos que combinan decisiones discretas y continuas sobre la operación y las tareas de mantenimiento. Dichos modelos se basan en primeros principios, pero incluyen submodelos de caja gris basados en datos históricos de operación de la planta para aquellas relaciones entre variables difíciles de obtener dados los datos disponibles. Para afrontar las mediciones de datos poco fiables en el desarrollo de los modelos basados en datos, se han utilizado métodos tales como reconciliación de datos y regresión restringida.

Una vez obtenidos los modelos que representan los sistemas y sus distintas formas de operación, se han formulado diferentes problemas de optimización. Estos problemas se basan en los modelos desarrollados a los que se les ha incorporado (además de diferentes restricciones lógicas) una función objetivo que minimiza el coste de operación del correspondiente sistema. De esta forma, se ha obtenido un problema de optimización capaz de establecer cuáles son los valores que han de tomar las variables del proceso para que el coste de operación sea mínimo. Cabe destacar que los modelos desarrollados se alimentan de las condiciones de la planta, de forma que la solución que se obtiene depende de dichas condiciones. Además, la formulación desarrollada permite obtener resultados para su aplicación en tiempo real, por lo que alimentando la optimización con las condiciones en tiempo real, se obtiene un sistema RTO capaz de predecir la operación óptima en tiempo real.

Además, en los modelos desarrollados se han incluido diferentes parámetros para tener en cuenta la pérdida de eficiencia de los equipos debido a su estado de ensuciamiento. De esta forma, los problemas de RTO también pueden sugerir las tareas de mantenimiento que se deben llevar a cabo desde un punto de vista económico, respetando las restricciones de operación.

Una vez formulados los problemas de RTO, se han desarrollado dos sistemas de apoyo a la decisión (DSS) que ayudan a los operadores en el proceso de toma de decisiones sobre cómo operar las diferentes secciones de la fábrica. Cada DSS tiene como núcleo uno de los problemas de RTO desarrollados e incluyen interfaces para la presentación de la información. Dichas interfaces se han desarrollado según las preferencias de los operarios, de forma que se produce una simbiosis entre los operadores y los algoritmos de control.

Para estudiar la coordinación de la operación de las diferentes secciones se han analizado tres enfoques distintos. El primer enfoque consiste en resolver los problemas de optimización acoplados de forma iterativa, el segundo es desarrollar un nuevo problema de optimización en el que se centralicen ambos sistemas, y un enfoque distribuido basado en la descomposición lagrangeana.

Finalmente, con el objetivo de obtener el rediseño óptimo de la red de refrigeración, se ha desarrollado una superestructura que no sólo contempla todos los posibles rediseños de la red, sino que también tiene en cuenta el tiempo de amortización de los equipos que se van a comprar. Además, los beneficios de dicha inversión dependen de las diferentes condiciones de operación que se pueden dar en el futuro, que son, de algún modo, inciertas, por lo que se ha tenido que emplear una formulación estocástica de dos etapas. Así pues, se ha formulado un nuevo problema de optimización basado en la superestructura y cuya solución indica cuál sería el diseño óptimo desde un punto de vista económico. Sin embargo, el problema resultante era computacionalmente difícil de resolver, por lo que se ha desarrollado un método de descomposición que disminuye el tiempo de resolución.

Resultados y conclusiones

Los resultados obtenidos tras ejecutar los problemas de optimización con datos históricos de la planta muestran que, si se hubieran modificado las variables de los sistemas de acuerdo a las soluciones proporcionadas por los optimizadores, se habrían obtenido ahorros significativos. Por supuesto, como se ha probado con datos históricos, dicho ahorro depende de dichos datos y cómo se actuó en su momento, pero en general, se puede concluir que las formulaciones de RTO desarrolladas son muy útiles para ayudar a la toma de decisiones indicando la operación óptima. Además, los resultados también muestran que la política actual de limpiar los equipos dependiendo de cuál es el último que se limpió no es la más rentable desde el punto de vista económico, puesto que depende de muchos factores. Así pues, los problemas de RTO también son útiles para ayudar a establecer las tareas de mantenimiento.

Los resultados obtenidos al comparar resolver los problemas independientemente y coordinadamente, demuestran que efectivamente al resolverlos de forma independiente se obtienen soluciones locales. Cualquiera de los tres enfoques de coordinación proporciona mejores resultados (se habría obtenido un ahorro mayor). Además, la comparación de los tres enfoques demuestra que, para problemas de este tamaño, la formulación centralizada es la más adecuada.

Analizando la solución obtenida del problema del rediseño de la red, se concluye que el coste de operar la red disminuye conforme aumenta el número de bombas de calor utilizadas, pero solo hasta cierto punto. Por tanto, el número de bombas a comprar depende del tiempo de recuperación de la inversión que la empresa considera aceptable, independientemente de valor de dicha inversión. Cabe mencionar que el número óptimo de bombas de calor a utilizar y su disposición dependen de cada escenario.

Finalmente, cabe destacar que esta tesis es el resultado de varios años de colaboración entre el mundo académico y una industria de procesos, en la que se proponen soluciones a problemas de un caso de estudio real. No obstante, aunque las diferentes tecnologías desarrolladas están personalizadas para una planta industrial en concreto, se podrían adaptar y extender fácilmente a cualquier industria que tenga equipos y problemas similares.

Contents

List of Figures

Chapter 1

Introduction

Nowadays, industrial companies must deal with two main factors: on the one hand, the globalization, which makes the market more competitive and on the other hand, the increasingly restrictive environmental regulation. These factors increase the need to produce as efficiently as possible by reducing the energy and resource consumption (Krämer & Engell, 2017). Furthermore, the European Union and the European Association called SPIRE (Sustainable Process Industry through Resource and Energy Efficiency) have identified energy and resource efficiency as a key step on the path towards a sustainable economy, and the targets for the next few years are to decrease the greenhouse gas emissions and to increase the efficiency of the energy consumption.

To tackle such goals, not only technical innovations, new plants and new technologies are needed, but also the exploitation of the opportunities in existing plants by means of a more optimal operation in terms of energy and resource consumption. Thus, decisions about global production and plant–wide operation are increasingly being considered highly relevant (de Prada et al., 2018). However, not all are constraints in these challenges. The industries can (and must) exploit the recent technological advances in data storage, communications, and computational power, in order to advance in the objective fulfillment. This is linked with the fourth industrial revolution, called Industry 4.0, which is ongoing, and whose goals are to achieve a higher level of operational efficiency and productivity, as well as a higher level of automatization (Lu, 2017). The concept of Industry 4.0 was born in 2011 as an initiative for developing German economy. Such initiative arises as a consequence of the increment of global competition and the need for fast adaptation of production to the ever-changing market requests in the industrial production in the last few years. In order to achieve these requirements, extensive digitalization and organizational changes in current manufacturing technology are needed (Rojko, 2017).

Industry has entered a phase of big change that sees digital technologies as a key factor for the future to design Cyber-Physical Production Systems. These systems are predicted to enable new automation paradigms and improve plant operations in terms of increased facilities effectiveness (Lamnabhi-Lagarrigue et al., 2017). The advances in technology to collect and display data could help in such tasks, but what is really important is *to transform data into sensible information* for really making an impact in the companies. Besides, we could use the information not only to have a better monitoring of the plant, but also to develop different tools which help plant operators and managers to take the best operative decisions in real time. Furthermore, the coordination management of the production at different layers, from real-time actions to medium-term scheduling of production and maintenance, is a key factor in the efficiency of daily operation (Khor & Varvarezos, 2017). For this reason, in the last years, process optimization not only has become an interesting research line from an academic perspective, but also translated into a set of mature technologies with significant impact in engineering practice.

The field of process systems engineering (PSE) has been pushing around such ideas in various ways for over 50 years. This field integrates different areas such as process design, control and operations, and product design. It addressed the development of process models, either steady state or dynamic, strategies of process calculations, and computational methods for optimization that have become key on many industrial automation and control systems (Grossmann & Harjunkoski, 2019). In particular, in the 1980s, the confluence of several developments such as algorithms for numerical optimization, open equation modeling, increment of computer processing capability, etc., allowed for the first time, real-time optimization (Darby et al., 2011).

In industry, the production process usually adopts a hierarchical structure composed of different functional layers. The traditional pyramidal structure described in the international standard ANSI/ISA-95.00.01-2000 (ISA, 2000) and ANSI/ISA-95.00.03-2005 (2005) of process control systems is formed by four different decision levels, each one with a different objective, complexity, and time scale.

At the bottom of the pyramid of Figure 1.1, there is the *Field Level* which refers to the physical production process itself.

Figure 1.1. Pyramid of control

The first layer of the pyramid is the *Basic Control Level*, which defines the basic activities involved in sensing and manipulating the physical processes monitoring the measured variables and maintaining some of them in their range or setpoint, i.e., it controls them, by modifying the manipulated variables. The time scale can range of seconds up to milliseconds such as those required for regulatory control. Some examples of systems at this level are: PIDs, PLCs or DCSs.

The next level in the pyramid is the *Supervision and Advanced Control Level* which undertakes the activities of the workflow to produce the goods with the desired quality, setting suitable setpoints or ranges to achieve production goals. Usually, the time scale of this level is from seconds to hours. In this level we can find different systems like Model Predictive Control (MPC) among others (Camacho & Alba, 2013; P. Santos et al., 2020; Wang et al., 2017).

Above is located the *Management Level* which establishes the basic plant schedule providing the operation point in order to fulfill the demand, avoiding delays or excess of storage. This level takes the decisions of how to operate from an economic point of view. Decision systems at the top of this level are normally named manufacturing execution systems (MES), which support manufacturing operation management (MOM). This level usually operates in a timescale range of hours up to weeks using tools such as scheduling and real-time optimization (RTO).

Finally, the top level is the *Planning Level* or Enterprise Resource Planning (ERP) whose aim is to focus on the economic planning, deciding *what to produce* and how much of it, the resources allocation to do it and the time to start.

The complexity of the decisions to take grows according to the height of the pyramid and, consequently, specialized human intervention increases as the achieved degree of automation decreases. So, in addition to the many technological developments and mathematical background to improve the bottom layers of the pyramid (Åström et al., 2006; Skogestad and Postlethwaite, 2007), it is still needed to develop tools to help in the decision-making process of higher levels, which should be connected to the lower ones in both ways, usual top-bottom but also *bottom up* or feedback.

Consequently, an important issue to consider is how to coordinate the operation of the different layers and systems according to the real-time conditions (Qian et al., 2017). To tackle this issue, an integration of process knowledge with actual data gathered from the plant is needed. The wide-plant optimization is related to process coordination, meanwhile control is related to each process. Hence, integrating process knowledge with plant information is useful not only for control but also for optimization and coordination of the processes.

1.1 Motivation

The motivation of this thesis is aligned with the target mentioned above of the European Union and the SPIRE association to achieve a sustainable economy. The three consolidated targets on this path for 2030 are:

- a 40% cut in greenhouse gas emissions compared to 1990 levels
- a 27% to 30% share of renewable energy consumption
- an increment of 27% to 30% energy savings compared with the business-as-usual scenario

Within this framework, in the process industries, not only technical innovations, new plants and new technologies are needed to achieve these targets but also the exploitation of the opportunities in existing and in future plants by even more energy and resource optimal operation.

Thereby, the challenge is to operate a process/plant efficiently taking decisions in real time, by looking closer at daily operation. However, the industrial processes could be very broad and complex, so their optimal operation is a complex task, often with conflicting objectives (typically the economic benefit versus resource efficiency). Anyway, a clearly wrong or suboptimal decision with the structural complexity of actual process plants often results in a magnification of costs along the products value chain. Currently, many decisions about how to operate the process units are made by the operators and plant managers based on their own expertise. Nevertheless, due to the huge number of alternatives arising from the combinatorial problems that new technologies enable, it is often too complex for a human decision maker to find the best option with tight time requirements. Thus, most of the time, plant managers are happy if just feasible solutions are reached.

In this context, real-time optimization (RTO) appears as a sensible approach to adapt the operation to the mentioned targets and varying production aims. RTO makes use of a rigorous stationary model of the process to compute the operation conditions that optimize a predefined performance index, for example, minimization of the operating costs, maximization of product yields, or minimization of waste (Hernandez et al., 2018).

However, RTO is not very spread in industry due to the efforts it requires in terms of model development and maintenance, and the need of reaching steady-state operation of the plant to successfully update calculations and to see the results, among other reasons. In practice, large-scale plants are rarely in complete steady state and, even if they are, many processes suffer from inherent long-term dynamics (fouling, mechanical wear, catalyst deactivation, etc.) that may invalidate the optimality of the results computed by RTO. Considering how such degradation dynamics affect the behavior of the process is key in order to obtain a realistic model of the system and to determine the best current action that does not worsen the plant state in the long term.

One must bear in mind that RTO should address global aims defined by the planning layer of the company, receiving production aims, prices and constraints imposed by other parts of the process. Hence, its implementation should be done with dedicated software, communicating with the other functional layers through the real-time information system of the site. The standard approach for implementation is to connect the basic structure of RTO in cascade to MPC. However, if perfectly processed plant-model information cannot be assured, the loop plant-MPC-RTO may not be fully closed as the results of the RTO could not be realistic. Thus, an intermediate step is needed.

In this thesis such intermediate step is using a Decision Support System (DSS) that executes the RTO and displays the solution of the optimal operation conditions on a suitable interface. Hence, the human manager is still the one in charge of taking decisions based on expert knowledge, but now helped by the valuable suggestions provided by the DSS. Thereby, personnel are more confident in adopting RTO technologies and the troubles and expenses of closing the loop with the RTO systems are partially avoided.

Note that the benefits of adopting an RTO solution do not constrain to the daily operation, as using the RTO model *offline* may have additional benefits if used for *What if* scenarios or plant improvement studies. There are also benefits from monitoring RTO model parameters such as equipment efficiencies and heat transfer coefficients as a mean of optimally schedule the maintenance activities.

Another important issue to consider is that optimal decisions in a factory require a more global vision than considering in isolation the operation of the individual processes, since what can be good from the point of view of a specific process may not be so good for the whole factory. Thus, seeking for the best operation point in a section of a plant, could be suboptimal from a global point of view as a fully decentralized control architecture does not consider global constraints among processes or sections. Moreover, increasing the complexity of the models to consider the relation between other parts of the plants could cause the computing time to be too large for a practical application. Furthermore, huge models are much more difficult to maintain and update. For that reason, it is very important to study how to coordinate the production of the different sections, which may lead to a coordination of different RTO schemes. Note that larger RTOs have been implemented, for example for oil refineries as the work of (Galan et al., 2019), but these are the exception in the industry.

On the other hand, the operation of a typical process factory is conditioned not only by the decisions made about the way the different processes are conducted and external factors such as markets or raw materials, but also by plant layout and the state of the process units. Plant layout and operation capacities of process units or utilities impose hard constraints on what can be achieved and obtained from a process. Consequently, process revamping and plant upgrades from time to time are quite common, in order to remove bottlenecks or improve capacities.

An approach based on optimization criteria allows to obtain better designs but requires developing rigorous models, often nonlinear, that establish the relationship between variables, as well as the use of the proper software tools. The problem is more complex when alternative plant structures are involved, all of which can be represented in a single model using what is known as a superstructure of the process, where alternative process units or operation modes are included or removed from the active model by means of discrete variables. This, of course, at the price of increasing the complexity of the design-optimization problem (Liong & Atiquzzaman, 2004), often intractable in the past if nonlinear mixed-integer formulations were involved. Fortunately, current advances in hardware, software and optimization algorithms allow to use this approach in a wider class of problems, with clear benefits over designs made mainly from experience and simulation.

Normally, the changes to be implemented in a process are designed so that it performs well in the new conditions in which it is expected to operate. Nevertheless, quite often, the designer forgets to consider that the process has to operate under different circumstances, different loads, materials quality, etc., so that it has to perform well not only for the selected nominal point, but for a whole range of operating conditions that may not be known with precision a priori.

The integration of different operation conditions in the design stage of a plant can be approached in different ways. A well-known method is the so-called multi-point or robust design, in which the model of the plant to be designed has to simultaneously fulfill different operating conditions and constraints, having a common unit sizing or structure and operational variables. This quite often leads to conservative designs linked to the worst-case situation.

Alternatively, the use of two-stage stochastic optimization (TSSO) methods (Birge & Louveaux, 2011) offers a more interesting framework for implementing such integrated design with operation. This approach allows considering the two stages, design and operation under uncertain environments, explicitly in a single framework (Steimel & Engell, 2015). TSSO incorporates the different operating conditions under which a process is expected to operate by means of a set of scenarios, each one corresponding to selected values of the uncertain variables that may appear, with a given likelihood assigned.

In both cases, for the optimal operation and for the optimal re-design, the optimization problem should be model based, so one important task is to develop a good mathematical model which fairly represents the behavior of the system to optimize (de Prada et al., 2017).

This thesis intends to contribute to the advance in the knowledge and implementation of these topics, with a clear focus on industrial application taking as reference an existing Austrian fiber-production plant. Accordingly, this thesis not only proposes models and theoretical RTO formulations, but also addresses the most common issues and limitations above mentioned detected for their implementation in a case study. In particular different RTO schemes for the optimal operation of some sections of an industrial plant are proposed and coordinated. Such RTO schemes take explicit consideration of fouling effects. Their implementation as a part of DSSs is also addressed. Furthermore, an integrated process re-design with RTO is also approached for a section of the plant.

The reference industrial site

In the production of viscose fibers at Lenzing AG (Austria), the most energy intensive process is the recovery of various chemicals by means of an evaporator process and heat recovery units, among other processes. In order to operate these process units optimally, five interdependent tasks have to be performed by the plant staff:

- Assignment of the evaporators to the individual products
- Setting of the individual evaporator loads
- Setting the amount of cooling water that goes through each cooling system attached to the evaporators
- Allocate the hot sources as utilities to heat up the products
- Setting the flow through each device

All this must be done under the presence of a high number of constraints whether due to first principles and logic statements in the model, or to the required operating conditions.

Furthermore, as in many other industries, there are thick layers of unwanted deposits on the surface of the equipment, which strongly reduce their efficiency (fouling effect). Hence, a continuous monitoring of such fouling is needed in order to improve the reliability of the models.

Currently, all decisions about how to operate the process units are made by the operators and plant manager based on their own expertise. Nevertheless, due to the huge number of alternatives arising from such combinatorial problems is often too complex for a human operator/manager. Hence, a DSS based on RTO will help operators and plant managers in the decision-making process on how to operate the process units according to the real-time conditions. The results of the optimization must be presented in an easily and quickly understandable interface.

The different decisions to take can be grouped according to the different networks of the process units (an evaporation network, a cooling system one, and a heat-recovery network). Nevertheless, it is very important to coordinate the operation of the different sections, as they are coupled. Thus, this thesis studies three different ways to coordinate the operation of the different networks, comparing not only the goodness of the results obtained but also the computational time required to achieve such results.

Finally, one of the sections in the plant could be re-designed in order to incorporate new equipment. Thus, a rigorous mathematical model which represents the behavior of the network for all the design possibilities must be formulated. The model must consider not only the operation cost but also the payback time as a constraint to fulfill. It must bear in mind that the new design must be able to operate in different future operation conditions, so the uncertainty of these conditions must be considered.

1.2 State of the technology

In line with the pyramidal structure showed in Figure 1.1, the general hierarchy for control and decision making in a plant according to (Darby et al., 2011) is as Figure 1.2 shows.

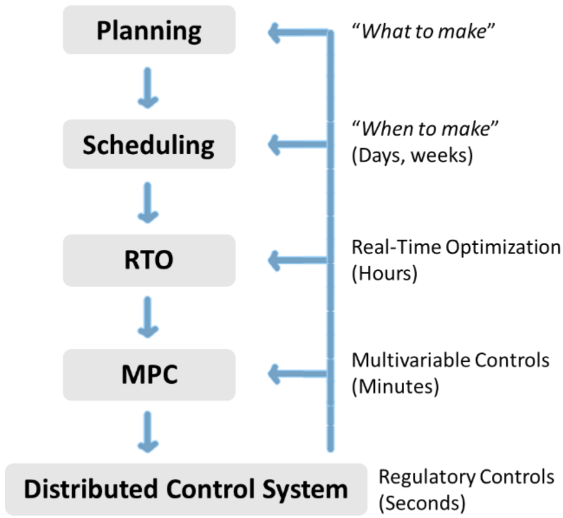


Figure 1.2. Plant decision hierarchy (Darby et al., 2011)

This hierarchy decomposes an otherwise unmanageable problem into a cascade of interconnected solvable problems as each layer determines what is relevant for the specific decisions that must be made and omitting the details that are not relevant. However, it makes necessary sharing relevant plant data, model information, constraints and pricing across all levels of the hierarchy, which is the nowadays main concern.

The DCS layer is responsible for regulatory control, providing stability and safety to the operation of the process systems. The MPC layer not only allows to take care of interactions among process variables, but also considers the constraints about admissible variable values and the presence of complex nonlinear and delayed dynamics of the processes. It is based on a process model which allows to consider future consequences of the interactions and process dynamics, according to the current control actions. The model allows computing the future evolution of the controlled variables over a prediction horizon respecting the constraints and fulfilling the control specifications.

Once the decisions about the technical aspects of how to get to the best operation point from the control point of view have been performed, the next layer is to find such point considering an economical point of view. That is where RTO arises.

Note that, comparing Figure 1.1 and Figure 1.2, the *Management Level* has been split into scheduling and RTO as they are used to solve different problems. In scheduling, the important issue is operational planning *over short-medium time periods* (from hours to weeks), meanwhile RTO computes control setpoints in *real time,* based on current plant conditions.

One must bear in mind that the vertical integration of the automation pyramid is not fully complete, the feedback of information from the lower layers to the upper layers is not well resolved. Nowadays the MES and ERP layers are still very *manual*, and one of the key ideas of Industry 4.0 is to automate them. Hence, the limits of the pyramid layers are blurred in order to achieve an idyllic autonomous and intelligent factory where the well defined and isolated control departments are no longer a model to follow.

1.2.1 Real Time Optimization

Since RTO emerged when online computer control of chemical plants became available, it has been developed closely with industry. Its applications involve not only chemical and petrochemical industries but also food production or biological processes, among others.

According to the hierarchical structure seen in Figure 1.1, RTO provides a bridge between plant scheduling and process control as medium-term decisions are made, by explicitly considering economics in operations decisions (de Prada & Pitarch, 2018). Thus, a real-time optimizer computes the optimal operating point under changing conditions based on an economic criterion through the explicit use of models. Such operating points are the setpoints for a set of controlled variables that are passed on to the lower-level controllers.

Although RTO expands from MPC, the base problem to be solved has two important differences. First, in RTO, the objective function is based on an economic criterion. In addition, meanwhile in the MPC model considers the dynamics of the processes, in RTO the problem is formulated in steady state. Note that in this case real time is referred not in a continuous way, but from time to time, generally according to each time that there are (or could be) changes in the operation conditions, such as changes in raw material quantity or product demand. When RTO is implemented with MPC, the rate at which RTO can be executed depends on the frequency of unmeasured disturbances and the time required for MPC to move the process to a new steady state. Furthermore, as the models only describe the steady-state behavior of the process, the optimizer only should be executed when the process is in steady state.

The benefits of using RTO in addition to MPC comes when we consider the optimization problems from a wider perspective than a process unit, i.e., RTO accounts for tradeoffs that MPC cannot, either by considering economic aspects or a larger envelope. Meanwhile MPC application has a usually local scope, RTO is normally implemented to seek the best operation point that optimizes the global process efficiency and economy.

RTO can consider the amount of interconnected equipment and networks that those industries are made up, where to achieve the optimal management is not a trivial task. Thus, the development of a systematic mechanism, such as RTO, translates into important economic payoffs. So many companies have developed RTO solutions and related software, whose use is increasing due to the highly competitive market. Some examples in the recent literature of the benefits of using an RTO scheme in process-decision making are (Galan et al., 2019; Han et al., 2015; Merino et al., 2018).

The majority of RTOs implemented over the last years use rigorous models based on first principles to describe the *theoretical* steady-state behavior of a process with enough precision. There are many RTO implementations which use commercially available flow sheeting packages based on open-equation modeling techniques that can code such first-principles models. However, care must be taken for a model to represent the actual process accurately, as a wrong RTO solution has large impact on the economy and efficiency of the factory. Thus, it is worth spending time and effort on formulating and customizing the model to the actual plant before implementing RTO. Moreover, the model must be of enough detail and complexity to adequately reflect the key economic tradeoffs and operating constraints, so that the model predictions result in a valid economic optimum, but also considering the computational demands in the optimization as well as the efforts needed by plant engineers to maintain the model.

Therefore, a finer approach is to develop a base model via first-principles laws for the process type under consideration, and then to tailor it to the specific plant with empirical laws derived from process data. Building these *hybrid* or *gray-box* models implies both, knowledge of process behavior based on conservation laws (for mass, energy, and momentum) and constitutive equations that describe phase and chemical equilibrium, transport processes, reaction, etc., and gathering informative enough process data (de Prada et al., 2017). In this way, machine learning can be recalled to approximate some effects that cannot be easily modeled by first principles. Furthermore, the model equations should be coupled with constraints based on process and product specifications, as well as an objective usually driven by an economic criterion (Biegler, 2010). Finally, a set of available decisions, as equipment parameters or operating conditions need to be selected. All these items are translated as a process model with objective and constraint functions, which define the optimization problem, as Figure 1.3.

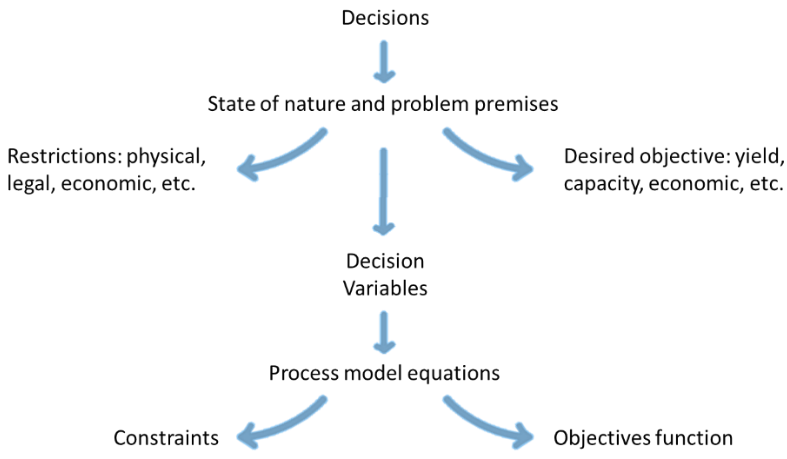


Figure 1.3. Conceptual formulation for process optimization (Biegler, 2010)

Following such scheme, the typical formulation of an optimization problem must contain:

- An objective function that measures the quantitative performance to be minimized or maximized, usually the operational costs, yield or profit.
- Decision variables which are the variables that can be adjusted to satisfy the constraints and find the optimal value for the objective function.
- A model which consists of a set of equations and inequalities named constraints that describe the behavior of the process.
- Operating constraints, i.e., the limitations on the values of some variables due to equipment constraints, safety constraints, quality constraints and/or environmental constraints.

Grouping the decision variables in a vector denoted **x**, and representing the cost function as $J(x)$ and the equality constraints as $h_i(x) = 0$ and the constraints as $g_j(x) \leq 0$, the typical RTO can be associated to the solution of the problem:

$$
\begin{array}{lll}
\min\limits_{x} \; J(\mathbf{x}) & & \\
h_i(\mathbf{x}) = 0 & i = 1,2, \ldots., l \\
g_j(\mathbf{x}) \leq 0 & j = 1,2, \ldots., m
\end{array}
\quad \Rightarrow \quad
\begin{array}{l}
\min\limits_{x} \; J(\mathbf{x}) \\
\mathbf{h}(\mathbf{x}) = 0 \\
\mathbf{g}(\mathbf{x}) \leq 0
\end{array}
$$

Depending on the domain of the variables (binaries or real) and characteristics of the constraints (linear or nonlinear), the optimization problem could be very complex.

One of the major contributions to the use of rigorous models was the arrival of open equation modeling, which allowed engineers to focus on the model equations without worrying about convergence details, or how to structure and solve the model update step vs. the optimization step, as was the case with traditional, flow sheeting packages (Darby et al., 2011). With the increment of computer processing capability, to get an efficient and reliable solution for large problems was possible.

Nevertheless, sometimes, the relation among some variables can be quite complex, difficult to model, or even depending on unknown variables. In such cases, *hybrid models* are needed, where some relationships are extracted from process data. Hence, the development of these models is a critical point in order to have reliable results (de Prada et al., 2017). A more detailed explanation of hybrid modeling is given in the next section (Section 1.2.2).

In addition, field measurements used either as input data to the model or to build the hybrid model, are never perfect. Therefore, it is necessary to correct the raw information before its further use, because if wrong information appears in the model, the solutions of the RTO that must verify these equality equations would not correspond to the real process optimum, even if the structure of the model was perfect. In order to mitigate this problem, RTO is normally accompanied by a module called data reconciliation (Leibman et al., 1992) that computes reliable estimations of the model parameters, i.e., the raw data collected from sensors, using the basic first-principles laws.

One of the biggest challenges concerning RTO is that models never represent the real behavior as processes are always affected by unknown disturbances. When the model has structural differences with reality, it is possible that RTO gives wrong results, obtaining an optimal operation point different from the real one. Hence, different solutions have been studied in literature for dealing with the uncertainty involved in the problem, such as the ISOPE (Integrated System Optimization and Parameter Estimation) (Roberts, 1979) methods or Modifier Adaptation(Marchetti et al., 2009) which use the process measurements to estimate the gradients of the objective function in order to compute some modifiers for the cost function and the constraints. Or also Extremum-Seeking Control (Ariyur & Krstić, 2004) and Neighboring-Extremal Control (Würth et al., 2009) which incorporate the plant information to update the process inputs, replacing the optimization problem by a feedback control problem which tries to satisfy the process optimality conditions.

In addition, although RTO is usually based on steady state models, it may happen that dynamic effects could not be neglected. An example of this case is when the plant is rarely at steady state due to the presence of significant and persistent disturbances. The solution then is to merge the MPC and RTO layers in order to formulate a dynamic RTO, obtaining a dynamic controller with an economic target optimization, called economic MPC. Some examples of such type of controllers can be found in (Engell, 2007; A. I. G. Santos et al., 2001)

Finally, another of the big challenges of the RTO concerns the implementation of the solutions obtained. Usually, such solutions are passed to the MPC as the value of the setpoint, allowing a smooth transition between the current and the optimal operation point. However, it requires consistency between the gains of the RTO and MPC models. But even when such consistency exists, the optimal operation conditions given by the RTO may be in a different operating region, forcing to update the MPC local model. Thereby, alternative approaches of the implementation of the RTO solutions have been studied in the literature. One possibility is described in (Skogestad, 2000), which consists in solving the RTO offline and then to design a control scheme that approximately implements the optimal solutions by maintaining the so-called self-optimizing variables at their setpoints. Another approach is to move the RTO to the MPC layer, using the nonlinear model to update the steady-state gains in the MPC instead of using the model in a separate optimizer, as proposed in (de Gouvêa & Odloak, 1998). Thus, the optimal operating point is obtained based on gains determined by the nonlinear model, without the requirement of achieving plant steady state.

This thesis contributes in the problem of developing hybrid models suitable for RTO of large systems, and in the implementation challenge, i.e., how to put solutions into practice, when the loop is not yet closed and requires human acceptance and participation.

1.2.2 Hybrid modeling

The concepts of *gray-box modeling* and *hybrid semiparametric modeling*, emerged in the 1990s, but in different fields. The first one came up in the systems and control theory field, where the first session on gray-box identification was held at the fourth IFAC Symposium on Adaptive Systems in Control and Signal Processing in 1992. In 1994 a first special issue in International Journal of Adaptive Control and Signal Processing was published on the same topic. The hybrid semiparametric modeling has evolved from the neural network field as a way to introduce structure into the neural network models with the work of Psichogios and Ungar; Johansen and Foss; Kramer, Thompson, and Bhagat; and Su et al. (Glassey & von Stosch, 2018).

Gray-box modeling combines priori knowledge (mainly structural information derived from first principles, i.e., white-box models) and response data from experiments (black-box models). Hybrid semi-parametric models may be classified as a type of grey-box models as its definition, given by (Thompson & Kramer, 1994), is “model structures that combine parametric and nonparametric submodels” considering; *(a)* parametric models those determined a priori based on knowledge about the process and their number of parameters is fixed but, depending on the level of knowledge sophistication, they might have a physical or empirical interpretation; *(b)* nonparametric models those where the number and nature of the parameters are flexible and not fixed in advance by knowledge.

Since its first practical application in the 1990's, hybrid modeling has gained interest and several practical applications was adopted by different industries. A collection of relevant papers of the different applications can be found in (von Stosch et al., 2014).

Current trends encourage the use of pure data-driven approaches due to the rise of *Big Data* at industry. Nevertheless, the process industry is characterized by its knowledge of the physiochemical processes and detailed models for some equipment/plants have been developed successfully in the last decades. Therefore, relying only on data-driven models and throwing out all this deep knowledge would be a waste of information.

The development of first-principles models requires detailed knowledge about the process but, in the process industry, this knowledge is usually partial, as could be unmeasured variables or the effort for computing them is too expensive and the effects of unknown inputs (*disturbances*) are not negligible. Thus, incorporating data-driven models allows to keep an understanding of the system, yet not all parts of the model need to be fundamentally understood. This is the principal benefit of hybrid modeling but not the only one.

The other benefit of using hybrid modeling is that the requirements on the data are lower than for purely data-driven models, as the incorporated fundamental knowledge to build these models may help to describe the relation between variables, reducing the need to investigate these interactions experimentally. Furthermore, the extrapolation properties are better than in purely black-box models, as it has been reported several times (Oliveira, 2004; te Braake et al., 1998; van Can et al., 1998), allowing reliable model predictions beyond experimentally tested process conditions.

Nevertheless, this kind of modeling also raises different challenges, being the major one the identification of the *unknown*. There are many machine-learning methods for such identification from data, being one of the most common the least-squares (LS) regression with regularization in the model coefficients (Kim et al., 2007; Neumaier, 1998). However, it must be considered that the goodness of the obtained model and its guarantees of physical coherence depend strongly not only on the quantity but also on the quality of the data.

Therefore, the design of an identification experiment should be carried out in a way such that the unknown part of the model can almost be completely discovered, though it is rather unrealistic to know that in advance for many cases. If no data at all, nor any knowledge about the system at hand is available, then a systematic exploration of the process design space, through experimental design, can be highly valuable (Chang et al., 2007; S. Gupta et al., 1999; Saraceno et al., 2010; Thibault et al., 2000; Tholudur et al., 2000; Tholudur & Ramirez, 1999).

However, although operation data are available, such raw data obtained from plant measurements might have inconsistencies due to noisy or biased sensors, so the estimation of the real values of the data is a must before any identification in a data-driven model or other further use. Data reconciliation (Leibman et al., 1992) is one of the best ways to ensure the quality of the data. Such methodology consists in estimating the real values of the data by solving an optimization problem that finds the values of the model variables and unknown parameters that better fit the plant measurements. Nevertheless, there are different techniques to approximate the models through the data from the real plant measurements, according to the existing redundancy either in the model (due to additional algebraic constraints) or duplicated sensors.

Recently, (Pitarch et al., 2019a) proposed an interesting approach for building hybrid models in the PSE context. The methodology is split in two-stages, combining robust data reconciliation of the raw data with improved constrained regression for the identification of extra black-box relationships.

Note that black-box models are developed mainly based on data, but if the environment is no longer the same as when these models were identified, it might never be possible for the prediction to correspond to the actual system response.

In this thesis, different hybrid models have been developed for optimization purposes. Such models are essentially based on first principles, expert knowledge on the processes, and their operation constraints. For the relation between variables which are too complex to determine by this way, data-driven models have been developed using the SOS-constrained regression (Pitarch et al., 2019b), both to fit the data and to enforce coherent physical model responses.

Such method consists in assuming that a dataset of N samples over time for some outputs y and some inputs u is available. Then, a candidate model for regression $f(\cdot)$ is sought such that a p-measure of the error (e.g., L_1-regularized or least squares) with respect to the data is minimized over a set of constraints $c(\cdot)$, which specify some desired features on the model (e.g., non-negativity or complex bounds on the model n-degree derivatives). The optimization problem to solve is:

$$\begin{aligned} &\min_x \sum_{t=1}^{N} \left\| y_{[t]} - f\left(x; u_{[t]}\right) \right\|_p \\ &\text{s.t.:} \\ &\qquad c(x; u) \le 0 \qquad\qquad \forall \;\; u \in \mathcal{U} \end{aligned}$$

In this thesis, the data used to develop the models have been obtained under selected operation points of the plant, to ensure the right conditions to develop a proper model. If the required experiments were not possible to perform due to tight production constraints at the factory, data from plant historian was treated by data-reconciliation techniques, when the redundancy in the first-principles base model made it possible.

Sometimes the name hybrid model is also applied for those models which combine discrete and continuous variables, i.e., those models where some of the variables are constrained to take integer values, for instance in scheduling problems (Harjunkoski et al., 2014). The previous concept of gray-box model is used in this thesis to represent equipment behaviors, and this different concept of hybrid models is used to model equipment-network layouts and operation constraints. The optimization problems derived from such models can be solved via Mixed-Integer Programming (MIP). If in addition, the objective function and/or the feasible region of the problem are described by nonlinear function, the problem becomes a Mixed-Integer NonLinear Programming (MINLP) problem.

1.2.3 Mixed-Integer NonLinear Programming (MINLP)

Mixed-Integer NonLinear Programming problems combine the difficulty of optimizing over integer variables with the handling of nonlinear functions, turning out to be NP-hard. Depending on wether all the functions included in the model are convex or not, the MINLP is itself called convex or non-convex. Despite both kinds of MINLP are generally NP-hard, the convex ones are much easier to solve than non-convex ones (Burer & Letchford, 2012). However, process systems which include discrete decisions are usually non-convex, as the physical laws that describe the behavior of the variable so are.

There are quite-effective exact-solution methods for convex MINLP problems which have been devised based on the premise that the continuous relaxation of a convex MINLP is itself convex. Some examples of such solution methods are Generalized Benders' Decomposition (Geoffrion, 1972), Branch-and-Bound (O. K. Gupta & Ravindran, 1985), Outer Approximation (Duran & Grossmann, 1986), LP/NLP-based branch-and-bound (Quesada & Grossmann, 1992), the Extended Cutting Plane (Sawaya & Grossmann, 2005), branch-and-cut (Stubbs & Mehrotra, 1999), and the hybrid methods (Abhishek et al., 2010). These approaches generally rely on the successive solutions of closely related nonlinear programming (NLP) or MIP problems. Nevertheless, all these methods have proved to be capable of solving instances with thousands of variables.

In contrast, the continuous relaxation of a non-convex MINLP is itself a global optimization problem, and therefore likely to be NP-hard. Therefore, using such methods for non-convex MINLP may cause that the NLP subproblems obtained in the relaxation have a local optimum and/or that the MILP master problem may cut-off the global optimum. There are two approaches to handle nonconvexities; (*i*) replace nonconvex terms by underestimates/convex envelopes and solve the global optimization or (*ii*) use *local* methods adding slacks to linearizations and searching until no improvement in NLP problems.

The exact algorithms for non-convex MINLP problems are usually based on the exact-solution methods for convex MINLP problems, specially based on the branch-and-bound method. Some of the most remarkable are: spatial branch-and-bound (Lee & Grossmann, 2001; Smith & Pantelides, 1997), branch-and-reduce (Ryoo & Sahinidis, 1995, 1996), conversion to an MILP (Beale & Tomlin, 1970), branch-and-refine (Leyffer et al., 2008). Using the mentioned exact-solution methods, several software packages have been developed, like BARON (Sahinidis, 1996) or Couenne (Wächter et al., n.d.), among others. Nevertheless, although all these methods are able to solve non-convex MINLP problems to proven optimality, the computational time required is usually too high for real-time demands or the LP relaxation may be infeasible or too expensive.

In contrast, the local methods have been designed to find good and quick solutions, although probably the solution obtained is not globally optimal. Thus, they are recommended when the model is so complex that an exact-solution method cannot get a solution, or, in the author's opinion, in the cases when the computational time to obtain a solution is more important than such solution to be the global optimal, as in RTO. Such local methods can be divided in those based on converting exact-global algorithms for convex MINLP problems into local for non-convex MINLP problems (Leyffer, 2001; Nowak & Vigerske, 2008), and those which adapt classical heuristic (and meta-heuristic) approaches, normally applied to 0–1 LP problems, to the more general case of non-convex MINLP problems (Nannicini & Belotti, 2012; Schlüter et al., 2009; Yiqing et al., 2007). Some software packages that can be used to find local solutions for non-convex MINLP problems are BONMIN (Bonami et al., 2008) or DICOPT (Viswanathan & Grossmann, 1990), which are actually packages for convex MINLP problems.

The different hybrid models developed in this thesis normally mix continuous variables (as the flow that must go through each unit) with discrete decisions (the allocation of products). Furthermore, not only some first principles but also the data-driven models developed are nonlinear functions in the decision variables. Hence, the associated optimization problems are MINLP without guarantee of convexity. As those problems are the core of an RTO for a DSS, the computational time to obtain a solution is key. Therefore, the method used to get a solution is *local*, using the software package BONMIN. However, it should be mentioned that BARON has also been tried in order to get the global optimal solution, but its required computational time to get it made it worthless.

1.2.4 Decomposition methods

When the optimization problem is a non-convex large-scale MINLP problem and neither the global solution methods nor local ones are suitable (due to computational time or limited memory requirements), a last option is to decompose the monolithic problem into several less-complex subproblems.

The key idea of *decomposition* is to solve a large-scale o complex problem by breaking it up into smaller ones and solving each of the smaller ones separately. Nevertheless, usually the smaller problems are coupled (there are *complicating variables* o *complicating constraints*), so they cannot be solved independently. For these cases there are coordination techniques, i.e., decomposition methods, which approach the solution of the centralized original problem by iteratively solving a sequence of smaller subproblems. In Conejo et al. (2006) the principal decomposition methods can be found according to the optimization problem type.

When the subproblems are well defined and the main problem to solve them independently is that there are *shared variables*, not only decomposed methods but also many architectures to coordinate the subproblems in the search for solutions involving the whole process can be applied.

The most common decomposition methods for MINLP problems are based in the Lagrangean relaxation method (Geoffrion, 1974), where the complicating part of the problem is added to the objective function as a penalty and the dual problem is exploited. There is a collection of different updating procedures for the Lagrangean multipliers, being some of the most highlighted the subgradient (Boyd et al., 2003) and the cutting plane method (Kelley James E, 1960).

There are also architectures based on cooperative games applicable to coordinate the solution of the subproblems (Lozano et al., 2013). Most often, the control scheme considers a distributed linear system composed of subsystems coupled with the adjacent subsystems through the variables. The big disadvantage of this method is that it is designed for systems controlled by a small number of variables as the number of options of the cooperative game grows in a combinational way.

Finally, there are also methods based on price-driven coordination algorithm (Cheng et al., 2007). In this approach a multi-layer architecture is needed where the subproblems exchange information with a coordinator (master problem). The harmonization between subproblems is made via price assignment mechanisms, such as price (Lagrangean) coordination methods where the master problem modifies the local cost functions to achieve a global optimal solution. In the case of applied price-driven coordination for shared variables between subproblems, each subproblem must add a constraint to force that the shared variable is equal to a reference. Such constraint is relaxed, i.e., added as a penalty in the objective function with a multiplier (called price). The master problem must compare the results of the shared variables and update the price and the reference for each subproblem. Traditional price-driven coordination algorithms are based on market law to update prices, but nowadays there are different alternatives to enhance the performance as the one proposed in Martí (2015). Nevertheless, these methods are usually not applicable for MINLP.

1.3 Objectives

The objective of this thesis is located in the framework of the work that contributes to operate industrial facilities efficiently by providing plant operators and managers with suitable decision-support systems at the RTO level resource consumption efficiently. Specifically, in this thesis, we face the mathematical modeling, optimal operation, and process re-design problems at several interconnected sections of a fiber-production site, not only by developing separate RTO schemes, but also providing a coordination strategy among them.

To reach this general aim, the following objectives have been considered in particular:

- Studying the operation of different sections of the plant, considering discrete and continuous decisions about both operation and maintenance:
    - An evaporation network
    - A cooling-water distribution system
    - A heat recovery network
- Developing data-based reliable models for optimization purposes, that take explicit consideration of the fouling effects.
- Building hybrid models of the operation alternatives at each considered section.
- Formulating optimization problems using the hybrid models to minimize the operation cost in order to:
    - Assign the products to the evaporation plants
    - Distribute the cooling water in the cooling system
    - Allocate the hot utilities in the heat recovery network
- Modeling a superstructure to represent all the possible configuration for the re-design of one of the networks (the cooling systems attached to the evaporation network).
- Studying different approaches for formulating an optimization problem for the superstructure, taking explicit consideration of the uncertainty in future operation conditions.

- Studying how to coordinate the operation of the different sections in order to get the global optimal point instead of the local optimal point of each one.
- Validating the models and algorithms with real plant data and studying its industrial implementation.
- Developing a DSS which incorporates the developed RTO systems together with suitable interfaces to show the results appropriately to plant personnel.

1.4 Structure of the thesis

The thesis is organized as follows: in the next chapter (Chapter 2), a global description of the fiber-production site is presented as well as the different sections whose operation is the core of the work, highlighting the most relevant details.

Chapter 3 is about the first case study, an evaporation network. After a brief description of the network, previous work to this thesis on this case is summarized. Then, the development of data-based models for predicting the behavior of the important variables is presented. The mathematical formulation of the optimization problem which compound the RTO is presented afterwards, besides the results of using this RTO. Finally, a proposal on how to integrate this RTO scheme with the previous work done is shown.

In Chapter 4, the other case study, which deals with a heat exchanger network, is presented. First, the layout of the network is described, and after that, the fouling effect inside the heat exchangers is modeled based on data. Finally, the hybrid model of the network operation is formulated, and it is incorporated to an RTO scheme, whose validation is demonstrated with the results obtained.

Chapter 5 addresses the formulation of the evaporation network model to consider, not only the operation, but also the optimal re-design for the incorporation of new equipment. As the model includes uncertainty of the future operation conditions, different formulations to solve the problem are presented. Once the optimal re-design is computed, its optimal operation is also guaranteed by a proposed RTO that is coherent with the design phase.

Chapter 6 is about the importance of the implementation of the developed technology, in these cases as decision support tools (DSS). Hence, a brief introduction to DSSs is presented. After that, the implementation of the different RTO schemes presented on previous chapters through the development of suitable DSSs is described. Then, the operator interfaces for each DSS are explained. Finally, how the integration of the DSSs with the control system would be implemented is briefly described.

Finally, Chapter 7 lists again the main contributions of the thesis but giving some general conclusions extracted after the work done. Also, a brief discussion of the future challenges and open issues that could be carried out as future work is included in this chapter.

Chapter 2

Industrial case study

The industrial networks for which this thesis provides decision support solutions correspond to a factory of the Lenzing Group, the biggest European man-made cellulose fiber production company. In particular on its head-quarter plant, located in the small Austrian town that gives name to the company. This industrial site will be used as a reference in the formulation of the problems and as a test site of the tools developed.

In such plant, they mainly produce several types of fibers with cellulose extracted from wood, their raw material. The fiber types are viscose, lyocell and modal fibers, and all of them are used in a variety of textile applications in addition to in hygiene products, cosmetics, or even agricultural applications. Nevertheless, as cellulose only represents the 40% of the wood, the other components are used to obtain different products (for food or pharmaceutical industries) or energy to supply their plants (see Figure 2.1).

Although the process to obtain the different fibers is essentially the same but with specific modifications, in this thesis we are going to focus on the viscose fibers.

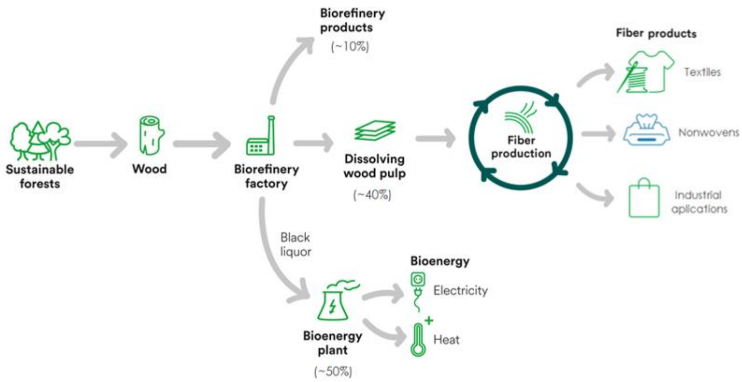


Figure 2.1. Diagram of Lenzing Group production

First, the wood is shredded, and the obtained cellulose pulp is matured and dissolved in a multi-stage physical-chemical process in order to obtain a viscous solution of high purity. After that, such solution is pressed through fine nozzles sunken in an acid bath. This is the key stage, called spinning, where fibers are recovered from the viscose solution by a chemical reaction. Finally, the resulting fibers are subject to different treatments (washed, stretched, cut, dried, and finished with an ended of soap-like substances) in order to obtain a lush and high-quality final product. (See Figure 2.2).

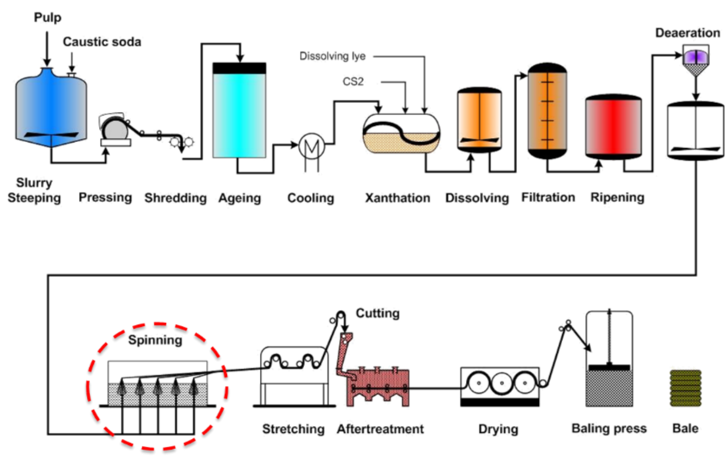


Figure 2.2. Scheme of the different processes to obtain fiber from cellulose pulp.

During the overall process, several chemical products are needed, and, in addition to the fibers, other chemical by-products obtained. In particular, in the spinning, the acid bath used for regeneration of the cellulose in viscose fibers, is composed of sulfuric acid (H_2SO_4), sodium sulfate (Na_2SO_4), zinc sulfate ($ZnSO_4$), alum and water. During regeneration, the alkali present in the viscose reacts with H_2SO_4 to form Na_2SO_4 and water. Hence, there is a continuous depletion of H_2SO_4 and build-up of Na_2SO_4 in the bath. The dilution of acid bath occurs due to large water content in viscose and water generation by the reaction of NaOH and H_2SO_4, causing the degradation of the bath. Therefore, in order to maintain the quality of the fibers and to reduce the wastewater and chemical consumption, the acidity of the baths (called spinbath hereinafter) must be recovered continuously (fiber production is a continuous process).

For that purpose, there is a recovery system which works in parallel to the spinning process. It involves a series of processes and equipment like circulation tanks, spinbath filters, evaporators, a crystallizer and rotary vacuum filters (see Figure 2.3).

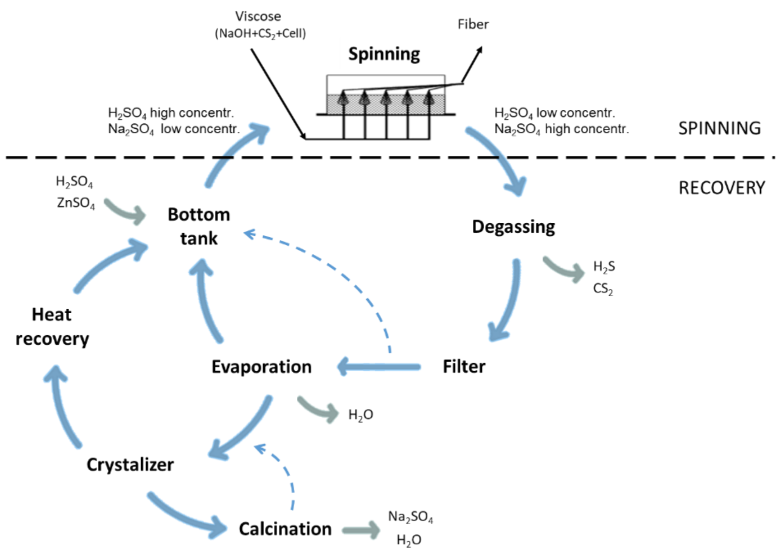


Figure 2.3. Recovery system scheme of the acid bath.

In the process unit *Evaporation*, water is separated from the spinbath, meanwhile the process unit *Crystallization* separates the sodium sulfate. Heat recovery units are used in order to transfer heat from one to the other process unit and in the *Calcination* unit water is removed from the sodium sulfate.

Nevertheless, this thesis is just going to spotlight the evaporation process (and its cooling system) and the heat recovery process, which are going to be briefly explained in the following sections. The complete detail of the networks and all the variables and parameters involved in each one, are in Chapter 3 and Chapter 4. Due to the complexity of each network layout, the amount of decisions to take, and the non-convexity of some model equations (e.g. energy balances) with respect to the decision variables, each network is going to be modeled separately. Then, a solution on how to coordinate the separate optimization problems to approach plant-wide optimization is given.

2.1 Evaporation network

As Figure 2.3 shows, one of the key processes to recover the acidity of the spinbaths is the evaporation of the water, which is carried out by a network of evaporation plants. These plants are multiple-effect evaporation stations (Prada et al., 1987) where each one consists of several separation chambers connected in series, working at low pressure to ease the boiling, among other equipment (see Figure 2.4).

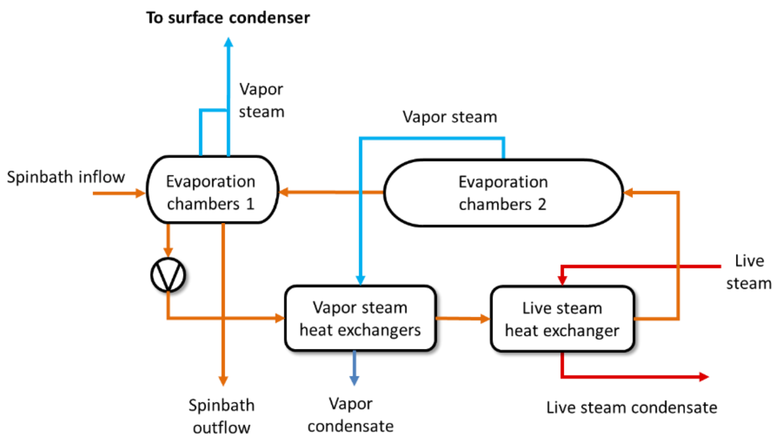


Figure 2.4. Scheme of a single evaporation plant

The spinbath inflow enters the plant by an evaporation chamber (the last in the chain of the evaporation chambers), where it mixes with part of the concentrated spinbath being recirculated. In this last chamber is where the 2nd-

effect vapor is extracted by the vacuum created with a cooling system. The recirculated mixture is then pumped through a series of heat exchangers, where it is preheated with the vapor coming from the rest of evaporation chambers. Nevertheless, to achieve the appropriate temperature to start evaporation, extra live steam coming from boilers is needed (main source of energy consumption). Once the solution is heated to the desired setpoint, it goes to the different evaporation chambers. The concentrated spinbath is finally extracted by overflow in the penultimate chamber, though it is mostly recirculated.

There are twenty-nine evaporation plants working simultaneously in the Lenzing's facility, forming an evaporation network. However, due to the different units that compose each evaporation plant, they are not equivalent, having different evaporation capacities (the amount of evaporation flow that can achieve) and efficiencies. In addition, Lenzing AG produces several types of viscose fibers (five different ones), which means a particular spinbath for each one to obtain the different fiber features. Although the efficiency of an evaporation plant does not directly depend on the spinbath that is treating, the demand of evaporated water does.

Despite the evaporation network is just a part of the recovery system parallel to the main processes of fibers production, it means a high demand of live steam consumption. In fact, it represents about 75% of the whole energy consumption in the factory. Thus, any improvement in such network that implies a reduction of steam consumption will have a big impact on the overall factory operation cost.

Hence, one of the aims of this thesis is to develop a system able to optimize the global steam consumption of the network by considering the coordinate operation of all the evaporation plants, taking into account the operational constraints associated to the process, such as the performance of the cooling system attached to the plants. Furthermore, as it will be exposed in subsequent chapters, there are few process magnitudes monitored and controlled but many different decisions to take (only the evaporation network involves around 150 decision variables, half of them discrete ones).

2.2 Cooling System Network

Attached to each evaporation plant there is a cooling system based on surface condensers. The aim of the surface condensers is, on one hand, to generate a pressure drop by condensation to evaporate the water of the spinbath (the 2nd-effect of the multiple-effect evaporation stations), and on the other hand, to condense such evaporated water to be used later in other parts of the factory. Nevertheless, the efficiency of evaporation plants depends on the behavior of this cooling system: the higher condenser performance, the higher evaporation capacity of a plant and the lower steam consumption. Thus, a good performance of the cooling system could also mean a significant reduction of the steam consumption, and consequently a reduction of the overall expenditure of the site.

The surface condensers are designed as tube and shell heat exchangers (see Figure 2.5) where the cooling medium, which in this case is cooling water from a river nearby, is passed through the tube bundles and the vapor condenses on the pipe wall.

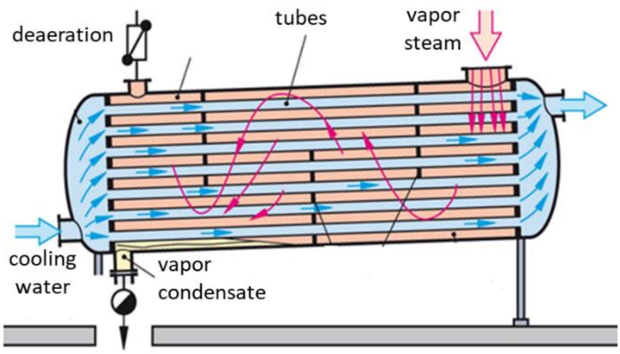


Figure 2.5. Scheme of a surface condenser.

Taking into account that the inlet temperature of the cooling water cannot be controlled, the way to increment the condenser performance is incrementing the cooling-water flow that, at the same time, involves a reduction of the outlet temperature of the cooling water at the output of the surface condenser. Nevertheless, the cooling water available to be taken from the river is limited, so the key is to know how to distribute the cooling water among the most suitable condensers, analyzing how the efficiency of each evaporation plant varies with the cooling power developed by the attached surface condensers.

The surface condensers themselves form a network and, according to the source of the cooling water fed to each condenser, the evaporation plants can be grouped into two sub-networks (see Figure 2.6). The first one can only get cooling water from one source, meanwhile the subnet two can get cooling water from the two available sources. Furthermore, as the outlet cooling water goes back to the river, its temperature must be below a limit imposed by environmental regulation.

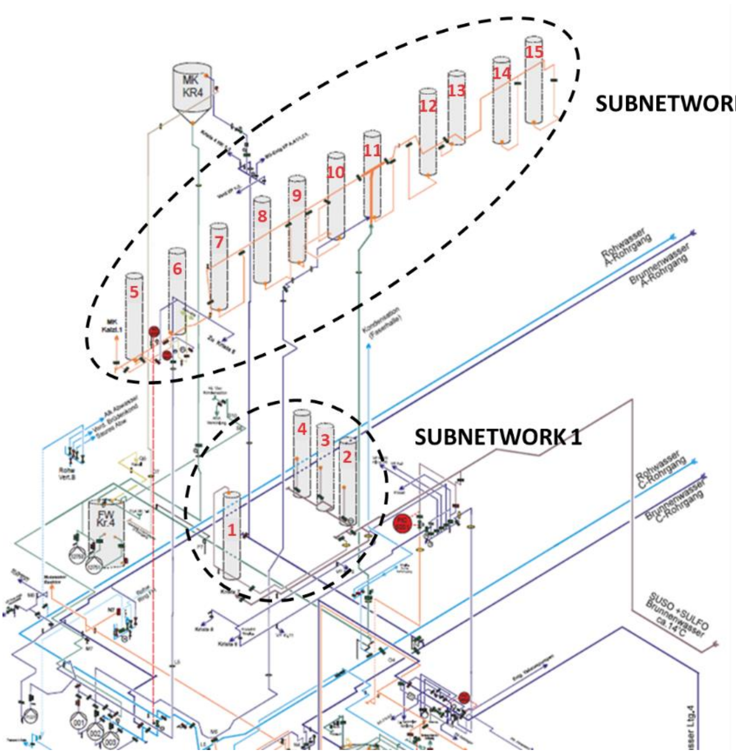


Figure 2.6. Cooling water network

In addition, nowadays the factory is thinking on improving the network by integrating some heat pumps. In brief, a heat pump is a thermal machine that extracts heat from a fluid and transfers it to another. In this case, the heat pumps will cool down the outlet cooling water from surface condensers. Thus, another task is to investigate which is the best way to incorporate heat pumps in the cooling network such that the cooling power is improved (hence, operability margins in the evaporation section too) while fulfilling environmental regulations without incurring in an excessive time lag between investment and payback.

Thus, regarding the cooling-water network, the aims are to develop a system able to provide the cooling-water distribution among the surface condensers not only for the actual network but also the optimal integration of the heat pumps and the cooling-water distribution in the future network layout. All of this, taking into account how it affects the evaporation-plant efficiency and the load allocation of spinbaths to plants. Even though the cooling-water distribution problem is not very large (barely around 15 decision variables), the complexity increases when considering the joint operation including the evaporation network, that not only raises the number of variables but incorporates discrete ones, making the problem non-convex. The heat-pump integration also brings more discrete decisions, as well as the need of incorporating uncertainty in the future operation conditions, making the problem even more complex and non-convex.

2.3 Heat recovery network

The spinbaths concentrated after the evaporation network (and the crystallizer process) must be heated in order to be in the suitable temperature for the spinning. Thus, another important network in the recovery system is the heat recovery network (see Figure 2.3) whose aim is to reduce the energy cost using the remaining heat present in waste streams.

The heat recovery network consists of fifteen heat exchangers. These heat exchangers are plate type, which are widely used industrially due to their extraordinary heat transfer properties. A plate heat exchanger consists of a set of corrugated metal plates in contact with each other where each plate has four holes that serve as inlet / outlet ports and to conduct the fluids along the plate. The plates are grouped together within a frame distributing the hot fluid through one of the plates and the cold fluid through the next one, thus allowing the transfer of heat (as shown in Figure 2.7). Due to its more compact design, they allow greater effectiveness in heat transfer.

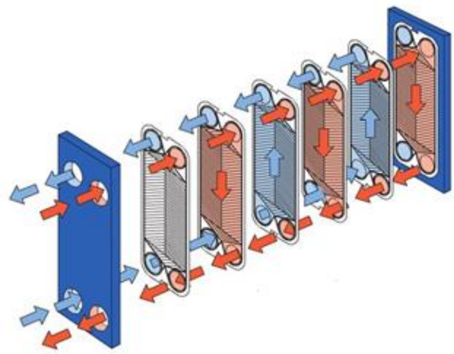


Figure 2.7. Scheme of a plate type heat exchanger

The network should be able to heat the different spinbaths to their suitable temperature by using four diverse heat sources: vapor condensates and different wastewaters that come from other parts of the plant. Nevertheless, not all the heat exchangers have access to the different hot sources. The heat exchangers are mostly disposed in parallel layout, and they can be grouped according to the hot sources that can use. In this regard, in one group the heat exchangers are disposed in a serial layout. In addition, the use of these sources in the heat exchangers involves a cost (pumping and maintenance plus the shadow cost of utilization for other purposes).

Therefore, the key operational decisions are how to allocate and distribute the heat sources (utilities) among the heat exchangers in order to fulfill the different setpoint for the spinbaths. As it will be shown in Chapter 4, there are about 200 decisions to take (some of which are discrete) and roughly twice as many constraints, where some are deeply non-convex.

2.4 Fouling effect

One of the most important issues that process industries that have fluids flowing through pipes usually have to deal with is the fouling effect.

Fouling is the accumulation of unwanted deposits on the surfaces of an equipment (see for example a fouled plate heat exchanger in Figure 2.8). In the case of heat exchangers, these deposits are related mainly to the fluid velocity, the compounds of the dissolution and the temperature. It increases the resistance to heat transfer, thus reducing the efficiency of such equipment over time. Consequently, more flow from the utilities is needed to reach the product temperature setpoint. Therefore, taking into account that the effect of fouling is key in order to obtain a realistic model of the process behavior and to determine the best way of action.

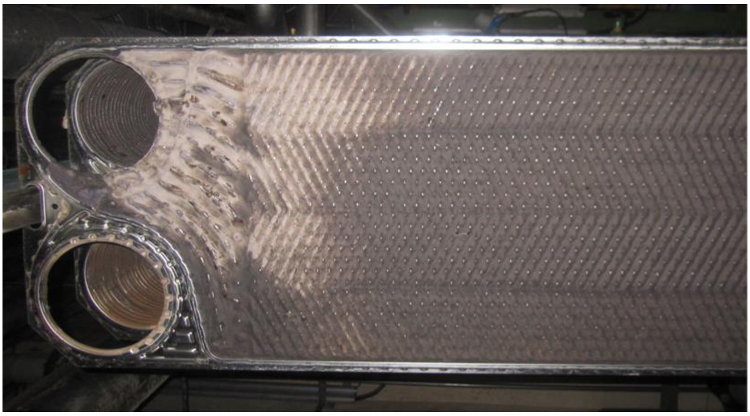


Figure 2.8. Fouling of a heat exchanger plate

Moreover, eventually the exchangers need to be cleaned in order to recover its nominal efficiency. But knowing when it is the best moment to schedule maintenance tasks is not a trivial task. Indeed, production-maintenance scheduling has been widely studied in the literature. See for instance the works of (Ghaleb et al., 2020; Palacín et al., 2018; Wang et al., 2022). Nevertheless, in those works the scheduling problems are not solved in real time, or the formulation is approximated to get a linear optimization problem. This second way is not feasible in this case study, at least without being extremely conservative or limiting too much the prediction range. Thus, instead of modeling the fouling dynamics to state a full production-maintenance scheduling over a future time window, this thesis focuses on developing proper models that estimate the heat-transfer efficiency in real time, and comparing it against the nominal values without fouling. By this way, we can monitor the fouling state and even suggest which heat exchanger should be cleaned in order to increase the efficiency of the network considering the cost of the cleaning task. Such suggestions increase the complexity of the problems, but it is possible to solve them in real time.

Chapter 3

Optimal operation of the evaporation network

This chapter is focused on the evaporation network and its cooling systems, briefly described in Chapter 2. In particular, this chapter addresses the optimal operation of such two continuous industrial systems interconnected.

First, previous work of the optimal operation of the evaporation network is presented in Section 3.1, along with the main operation problems of the cooling system attached. The main objectives for this case study will be listed in Section 3.2, and, after that, Section 3.3 will show the data-driven models developed for the unknown relations between decision variables, and the considerations taken into account. In Section 3.4 the mathematical formulation of the optimization problem associated to the cooling system and some results obtained are presented. Section 3.5 studies how to connect the two optimization problems in order to get a coordinated solution for the whole network. Finally, a summary of the work done in this case study and the conclusions are presented in Section 3.6.

3.1 Previous work and remained problems

As it was introduced in Chapter 2, the Lenzing site has an evaporation network used to concentrate the spinbaths, which is the main energy consumer section of the site. Thus, one important concern was to minimize the amount of steam required by each evaporation plant taking into consideration that the evaporation demand of each spinbath must be fulfilled.

The optimal operation of an individual evaporation plant was addressed in (Pitarch et al., 2017). In such work, the authors determined that, for a given demand of evaporated water (the so-called evaporation load, *EC*), the optimal operation point can be accomplished by setting the temperature of the recirculated flow (denoted T in Figure 3.1) at the highest admissible value, and then adjusting the recirculated flow (F in Figure 3.1) to reach the *EC* demand. A self-optimizing controller was implemented to ensure that operation always works this way. Note that the evaporation ratio for a given spinbath is known and fixed, so the flow of diluted spinbath (F_{in}) is not a control variable according to the control system at Lenzing. Instead, the evaporation flow to be reached (*EC*) determines the plant setpoint, given a spinbath inlet flow (F_{in}) to concentrate.

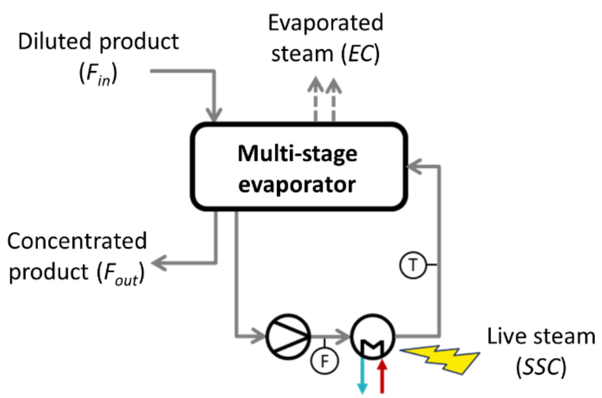


Figure 3.1. Simplified scheme of an evaporation plant

Once known the optimal operation of a single plant, the problem of how to allocate the different spinbaths to plants and the amount of water to be evaporated (evaporation load) that each plant must reach, has also been studied

(Kalliski et al., 2019; Palacín et al., 2018). The models proposed in such works consider that the evaporation plants have different nominal efficiencies. Moreover, depending on operating point, the specific steam consumption at each plant changes.

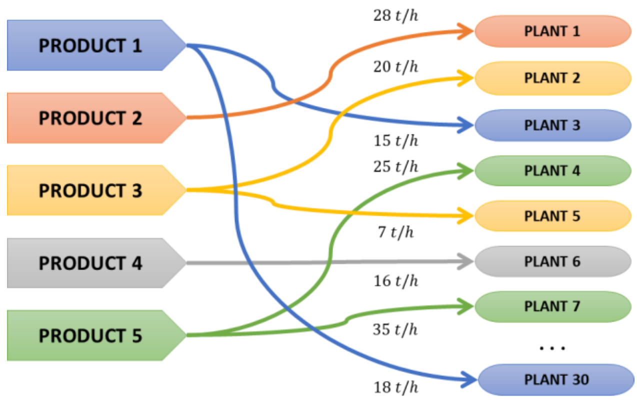


Figure 3.2. Example of distribution of production load among evaporation plants.

Do not forget that the efficiency of the evaporation plants also depends strongly on the performance of the cooling system. In those works, the studied network is composed of plants that have attached a cooling tower as cooling system. Thus, the performance of such cooling system is normally an active constraint, as they always work at maximum cooling power. Nevertheless, in the network studied in this thesis, the evaporation plants use surface condensers as cooling system, where the cooling power depends both on the cooling water flow through the condensers (control variable) and its inlet temperature (disturbance). Hence, the more and colder cooling water is provided to the condensers, the more efficient the evaporation plant becomes as less live steam is needed to achieve the evaporation setpoint. Nevertheless, the total available cooling water is limited, making impossible to run all surface condensers even at their maximum achievable cooling power given an inlet water temperature. Therefore, a proper cooling-water distribution through the surface condensers in real time could lead to increase the efficiency of the evaporation plants and, consequently, to decrease their operating cost.

The amount of cooling water available depends on two sources. According to the evaporation network layout, the surface condensers are grouped into two subnetworks where the first one is composed by the ones that only have access to Source 1, meanwhile the surface condensers of the second subnetwork have access to both sources (see Figure 3.3).

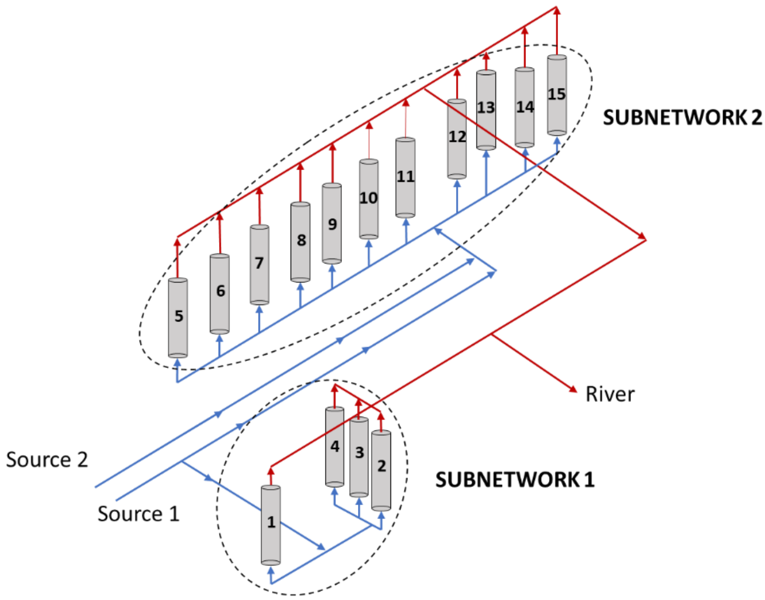


Figure 3.3. Simplification of the cooling-water network

Such sources are not only limited (the amount of cooling-water available per hour is limited) but also shared with other departments of the factory, consequently its use implies a cost. Thus, the operating cost of the network is a trade-off between the cost of live steam and cooling water usage.

In addition, due to environmental regulations, the temperature of the cooling water at the output of the cooling-system network must be below a limit value, as it goes back to the river (see Figure 3.3). Thus, a model to predict such temperature depending on the cooling-water flow through each surface condenser is needed for the optimal cooling-water distribution problem formulation.

Furthermore, one must bear in mind that industries which work with fluids usually have to deal with fouling problems that can lead to a loss of efficiency. In the case of surface condensers, the cooling water comes from the river, and although it is mainly water, it could include slight amounts of other organic particles that form unwanted deposits. Such deposit formation is related mainly to different factors as fluid velocity and temperature. The fouling layer decreases the heat transmission, so knowing how it affects the performance of the surface condensers is a must in order to have a realistic prediction of the system behavior.

Thus, an optimization problem that optimally distributes the cooling water through the surface condensers taking into account the network layout, the cooling-water availability and the outlet temperature should be formulated. The objective will be to minimize the operation cost given not only by the cooling-water flow, but also considering how such distribution affects the live steam consumption of the evaporation plants. In this regard, given an evaporation load allocation, using for example the proposed RTO tool by (Kalliski et al., 2019), the optimal cooling-water distribution can be obtained by the new problem formulation proposed in this chapter.

Nevertheless, with the new distribution of cooling water, the initially computed evaporation load allocation could be sub-optimal. Thus, it is necessary to integrate the cooling-water distribution problem jointly with the evaporation load allocation problem. In this chapter, different approaches have been studied in order to get the optimal operation for both systems.

The first approach is to solve both problems in an iterative way as the solution of each optimization changes the value of the parameters needed for the other optimization. This procedure is easy to implement, and the termination condition would be the equality of the computed cooling power in the solution of both problems. Nevertheless, such procedure has no global optimality guarantees.

The second approach is to merge both formulations into a centralized problem. However, as the cooling-water distribution is an NLP problem and the evaporation load allocation involves discrete decisions, the centralized problem becomes an MINLP one. Therefore, the complexity of this approach is expected to be higher than the one of the separate problems.

In order to overcome the above issues, the third approach proposed is to address the optimization in a distributed fashion reusing the existent formulations for the separate problems. This is done via Lagrangean decomposition and price-coordination schemes (Cheng et al., 2007), adding the shared variables as a penalty in the objective function for each individual problem. Furthermore, a smart rule for updating the *resource prices* in each iteration is employed to speed up the resolution, providing a decision-support tool with real-time capabilities.

3.2 Objectives

The global aim is to operate the whole evaporation network in the most efficient way, i.e., with the lower operating cost, now taking into account not only the performance of the evaporation system but also its cooling system.

This implies targets at the plant level and targets at the network level. Initially, the optimal operation of the cooling system according to the cooling water distribution must be studied. As one constraint of the network operation is the outlet temperature of the cooling water, which goes to the river, a data-driven model to know how it depends on the cooling-water flow through the surface condenser must be obtained. Moreover, such models must consider the decrease of heat transfer due to fouling in real time.

Furthermore, the efficiency of the evaporation network can be measured by its specific steam consumption. Thus, it is necessary to develop a model to know how the live steam consumption depends on the cooling water distribution (cooling water flow through each surface condenser).

Once such data-driven models for the outlet temperature and the live-steam consumption have been obtained, they can be used to develop a reliable hybrid model which represents all possible cooling-system responses according to the plant conditions measured in real time. Such model will be the base for the formulation of an optimization problem which predicts the optimal operation of the cooling system in real-time (an RTO). The objective of the

RTO is to distribute the available cooling water minimizing a trade-off between the live steam consumption and the cooling water.

After that, a study of how to integrate such distribution with the spinbath allocation problem to the plants will be included, so that both problems can be solved in a coordinated way in order to get the optimal operating point of the whole network in real time.

In summary, this chapter aims to:

- Obtain reliable data-driven models of the dependency of the steam consumption and the outlet temperature with the cooling-water flow.
- Formulate an optimization problem to distribute the cooling water according with real-time conditions.
- Implement and test the proposed RTO scheme with plant data.
- Study the best way to solve jointly the allocation of spinbath and the cooling water distribution.

3.3 Data-driven models

The cooling system associated to the evaporation plants uses cooling water to condense the vapor extracted. After passing through the surface condensers, the cooling water is returned to the river as shown in Figure 3.4.

Nevertheless, such temperature must be below certain value in order to avoid environmental problems (it is regulated by environmental laws). Thus, it is necessary to know how the outlet temperature (To) depends on the cooling water flow (F) in order to incorporate such constraint into an optimization problem of the specific steam consumption at the evaporation plants.

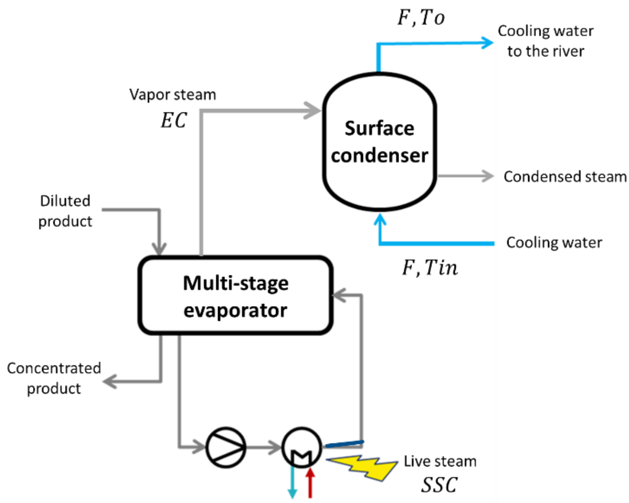


Figure 3.4. Scheme of an evaporation plant with its surface condenser

Thus, the relation between the outlet temperature (To) and the live steam consumption (SSC) with respect to the flow of the cooling water (F) must be studied. Modeling such relation based on first principles could be difficult due to the reduced flexibility of such kind of models (few parameters to be adjusted by the user to fit the actual plant), the many assumptions stated in the related literature, as well as the lack of information on some key variables that are not measured in the plant. Therefore, such relations have been modeled as black-box equations upon different experiments carried out onsite.

3.3.1 Modeling methodology

The experiments have consisted in running the surface condensers in different conditions, i.e., covering their usual range of flow operation. All the experiments have been done under three specific conditions:

- The elapsed time between experiments must be short to ensure that the inlet temperature of the cooling water is kept constant.
- The evaporation load must also keep constant.
- The experiments must take place after a cleaning to ensure that there is no fouling in the plant heat exchangers.

Applying these requisites allows to perform additional experiments in the future and to combine them with the already collected data with coherence, in order to improve and/or update the models over time. In addition, if the surface condensers are fully clean, it can be assured that the live steam consumption of each plant is the minimum achievable. Note that the specific live steam consumption of each plant also depends on the recirculated flow and its temperature, but such variables are already controlled in order to reach the optimal operation at each moment.

Furthermore, the experiments have been done with the plant in steady state as dynamics of pressure drop and heat transfer in the surface condensers are neglected (they are quite fast in comparison to the time constants of the plant). Thus, the relations between variables can be simply identified fitting a polynomial model to the experimental data by least squares. However, as such models have been obtained when the surface condensers are fully clean and for a given inlet water temperature, they must be adapted to be able to represent the system under other conditions.

It must be borne in mind that these experimental conditions do not prevent possible outlier points in the experimental data due to disturbances or sensor noise. Hence, these fits might not be the best to represent the real behavior of the system if such points are not avoided in the identification. Additionally, as there is not redundancy in the initial model due to additional algebraic constraints or duplicated sensors, data reconciliation methods (mentioned in Section 1.2.2) cannot be applied.

Thus, a more sophisticated modeling method to obtain these curves (instead of fitting the data to polynomials by least squares) is proposed based on the ideas collected in (Cozad et al., 2014). Such modeling routine can be divided into two steps:

i) *Adjust:*

First, given a series of N data (outlet temperature and specific live steam consumption) order them with respect to the values of the variable we want to relate them (cooling-water flow).

After that, find the best polynomial fit by solving the following nonlinear mixed-integer optimization problem, where the objective function follows the Akaike corrected criterion (Hurvich & Tsai, 1993) which is a trade-off between model complexity (determined by the variable D) and the fitting error got by the candidate polynomial functions $f: \mathbb{R} \to \mathbb{R}$ and $g: \mathbb{R} \to \mathbb{R}$. Such functions relate the outlet temperature and the specific steam consumption with the cooling water flow and the cooling power respectively, for each surface condenser.

$$\min_{\substack{\alpha,\beta,D \in \mathbb{R}^9 \\ y \in \{0,1\}^4}} N \log(N^{-1} J) + 2D + \frac{2D(D+1)}{N-D-1} \tag{3.1a}$$

$$s.t.$$

$$y_1 + y_2 + y_3 + y_4 \leq D \tag{3.1b}$$

$$\alpha^l y_1 \leq \alpha_2 \leq \alpha^u y_1, \qquad \alpha^l y_2 \leq \alpha_3 \leq \alpha^u y_2 \tag{3.1c}$$

$$\beta^l y_3 \leq \beta_2 \leq \beta^u y_3, \qquad \beta^l y_4 \leq \beta_3 \leq \beta^u y_4 \tag{3.1d}$$

$$f(F_{i+1}) - f(F_i) < 0 \qquad i \in \mathcal{N} \tag{3.1e}$$

$$g\left({Q^{cp}}_{i+1}\right) - g\left({Q^{cp}}_{i}\right) < 0 \qquad i \in \mathcal{N} \tag{3.1f}$$

$$g \circ f(F_{i+1}) - g \circ f(F_i) < 0 \qquad i \in \mathcal{N} \tag{3.1g}$$

Where $\mathcal{N} \coloneqq \{1, \dots, N-1\}$ and parameters α and β refer to the coefficients of the base functions that compose the candidate models f and g respectively. Such coefficients are limited by their lower and upper bounds, (3.1c)-(3.1d). In addition, they can be overridden to be zero, as the bounds are multiplied by the binary variables y allowing to limit the complexity of the model given by the decision variable D in (3.1b), being D the number of different base functions that will finally compose the models f and g.

Symbol $g \circ f$ is the composition of both functions, i.e., function g is applied to the result of applying the function f to F_i. In this case, the specific steam consumption depends on the cooling power, which at the same time depends on the cooling water flow and its outlet temperature. The next sections (sections 3.3.2 and 3.3.3) explain the development of the models, and $g \circ f$

refers to the complete model, being used to compute the fitting error in the objective function, (3.13a).

Besides, J computes the *fair* function (Huber, 1981):

$$J \coloneqq C^2 \sum_{i=1}^{N} \left(\frac{|\epsilon_i|}{C} - \log\left(1 + \frac{|\epsilon_i|}{C}\right)\right) \tag{3.2}$$

Being $C \in \mathbb{R}^+$the parameter that defines the sensitivity of the *fair* function to incoherent measures, and ϵ_i is obtained in each case as the differences between the collected data and the predicted one by the model. In this formulation the least squares function could be also used, but we prefer the fair function as it is robust to gross measurement errors (Nicholson et al., 2014).

The fundamental contribution of this methodology is the addition of constraints (3.1e)-(3.1g), where it is imposed that the numerical derivative of the three functions forming the model (i.e., f, g and $g \circ f$) must be negative, since it is known from physics that at higher water flows, lower outlet temperature and lower specific steam consumption. Note that it is also possible to obtain good polynomial models with restrictions in their derivatives using the SOS-constrained regression with regularization, but it has been decided to present this methodology in order to use the fair function instead of least squares and to have the possibility of using base functions not polynomials in the candidate model (although it has not been necessary).

ii) *Validation*

Obtained the model coefficients α and β, analytically check the fulfillment of the model constraints in the possible range of cooling-water flow (region of operation) by solving respectively:

$$\min_F f(F) \qquad s.t.: \underline{F} \le F \le \overline{F} \tag{3.3a}$$

$$\min_{Q^{cp}} g(Q^{cp}) \qquad s.t.: \underline{Q^{cp}} \le Q^{cp} \le \overline{Q^{cp}} \tag{3.3b}$$

$$\min_F g \circ f(F) \qquad s.t.: \underline{F} \le F \le \overline{F} \tag{3.3c}$$

If the solution of any of the previous problems is equal to the upper bound of the variable, the obtained model is monotonically decreasing for any flow value, and consistent with the physics of the process, so the algorithm ends. Otherwise, add each value found into the ordered dataset (so $\mathcal{N}$ grows) and repeat the adjust step.

3.3.2 Outlet cooling water temperature

Models providing the outlet temperature of cooling water *(To)* with respect to the cooling water flow (*F)* have been obtained fitting the data to a polynomial curve (see an example in Figure 3.5). It has to be emphasized that there is a different model for each evaporation plant, so, a set of 15 different models was obtained after performing the corresponding experiments.

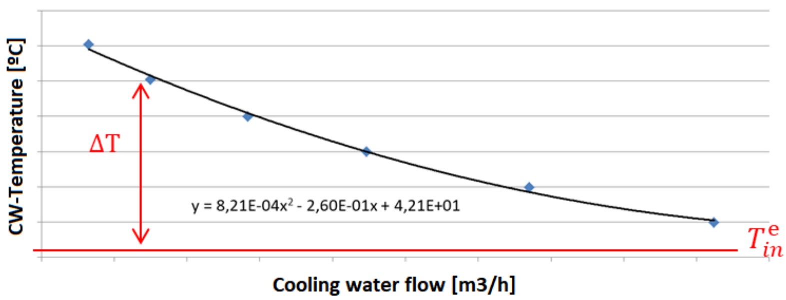


Figure 3.5. Experimental model for outlet temperature vs. flow

Nevertheless, these models output predict with respect to a reference temperature. That reference is the inlet water temperature when the experiments were carried out, T_{in}^{e} . Therefore, removing such reference temperature allows getting an incremental model (3.4). Hence, given a real-time measurement of the inlet temperature, denoted by $\widehat{\mathrm{T}}_{\mathrm{in}}$, the outlet temperature can be predicted at any weather condition by (3.5).

$$\Delta T \; = f(F) - \mathrm{T}_{in}^{e} \tag{3.4}$$

$$To \; = \Delta T + \widehat{\mathrm{T}}_{\mathrm{in}} \tag{3.5}$$

Where $f\colon \mathbb{R} \to \mathbb{R}$ is a nonlinear function that represents the experimental model, e.g., the polynomial curve depicted in Figure 3.5.

In addition, due to the fouling that the surface condensers suffer, the heat transfer decreases with time and, consequently, the outlet temperature of the cooling water will also decrease. Thus, assuming the tests were carried out with the surface condensers fully clean, the actual outlet temperature measured at any time instant will be always lower or equal than the predicted by (3.5), as shown in Figure 3.6.

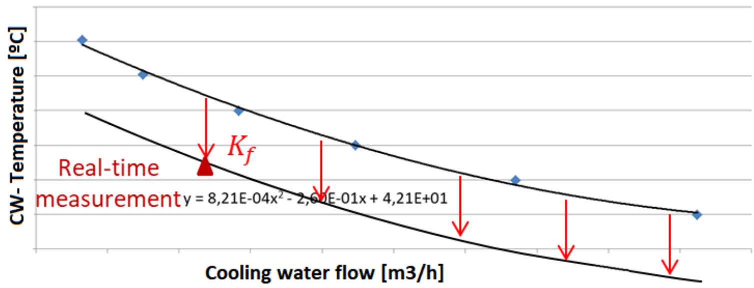


Figure 3.6. Model adaptation to current fouling state

Then, the idea is to add a bias parameter K_f to the above base model (3.5), which adjusts the curve to the real-time measurement:

$$To = \left(f(F) - T_{in}^{exp}\right) + \hat{T}_{in} - K_f \tag{3.6}$$

In this way, the current state of fouling will be considered in the network optimization. The bias K_f can be easily updated from run to run with real-time measurements of the outlet temperature $\hat{T}_o$ and the flow $\hat{F}$ by:

$$K_f = f(\hat{F}) - T_{in}^{e} + \hat{T}_{in} - \hat{T}_o \tag{3.7}$$

Note that this approach allows to isolate the fouling effects in the cooling system from the ones in the spinbath heating line, which also affect the overall specific steam consumption. Fouling monitoring in the heating line was treated in a work before this thesis; see (Kalliski et al., 2019) for details.

3.3.3 Specific Steam Consumption

From the collected data of inlet/outlet temperature and volumetric flow of the cooling water, the actual cooling power (Q^{cp}) in the surface condensers can be computed by (3.8), where the water density (ρ) and the specific heat capacity (Cp) are assumed to be constant.

$$Q^{cp} = F\,\rho\,\mathrm{Cp}\left(To - \widehat{\mathrm{T}}_{\mathrm{in}}\right) \tag{3.8}$$

Moreover, by recording the live steam consumption of the evaporation plant in during the tests with the SCs, the specific steam consumption (SSC) can be depicted versus the cooling power Q^{cp} and, hence, to fit a model for prediction, as Figure 3.7 shows.

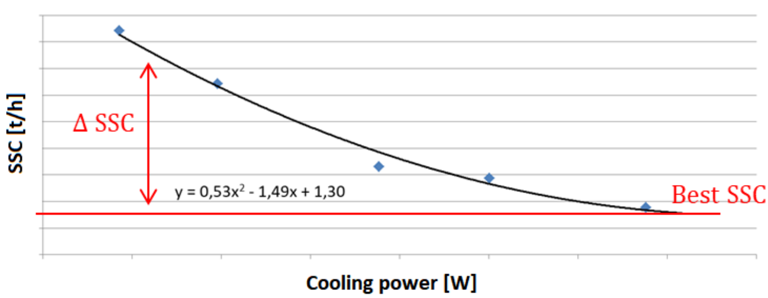


Figure 3.7. Specific steam consumption vs. cooling power

Note that, as well as for the outlet temperature models, there is a different model for each surface condenser.

Analogous to the previous case, we also need to remove the dependency on the operating point (evaporation load) during the experiments from the fitted model. To do so, the simplest idea is to compute the best specific steam consumption (BSSC), i.e., the minimum SSC obtained in the experiments, and to build an incremental model:

$$\Delta SSC = g(Q^{cp}) - BSSC \tag{3.9}$$

Where $g: \mathbb{R} \to \mathbb{R}$ is a nonlinear function that represents the experimental model, e.g., the polynomial curve depicted in Figure 3.7. For that to be true, it is needed to assume that the tests were carried out with the surface condensers fully clean, and that the model for ΔSSC (i.e., the shape of the curve in Figure 3.7 for instance) does not vary significantly from one operation point to another (plant evaporation loads).

Based on the data, polynomials with degree no greater than three are enough to get a good fit. These models represent satisfactorily the system in the operation range and are suitable for real-time optimization.

Once both data-driven models $f(F)$ and $g(Q^{cp})$ have been found by identification, they can be used for prediction, receiving the water flows F and the inlet temperatures $\widehat{\mathrm{T}}_{\mathrm{in}}$ as inputs and providing the water outlet temperatures To and the increase of specific steam consumption ΔSSC as outputs. Note that, the current measured outlet temperature $\widehat{\mathrm{T}}_{\mathrm{o}}$ and flow $\widehat{\mathrm{F}}$ can be used to update the parameter $\mathrm{K_f}$ online, but they are different from the predictions, i.e., the decision variables for optimization.

3.4 Cooling system optimization

Once the different data-driven models for the outlet temperature and specific steam consumption are obtained and adapted to be able to predict the behavior under different conditions (inlet temperature and state of fouling), the next step is to formulate a mathematical model which represents the behavior of the network. Such model is based on first principles, i.e., mass balances, but including the black-box models developed. Thus, the resulting model is a gray-box model which predicts the performance of the cooling-water system.

As mentioned in section 3.1, this evaporation network can be grouped into two sub-networks according to the source of the cooling water (resource) that feeds each surface condenser. The first one, formed by four evaporation plants, can only get cooling water from Source 1, meanwhile the subnet two, formed by the remaining eleven plants, can get cooling water from the two sources (see Figure 3.3). The surface condensers are connected in parallel, i.e., the cooling water used by one surface condenser is not used in another surface condenser. Thus, as the cooling water to take from each source is limited, a problem of shared resources arises.

The next step is to incorporate an economic target as objective function that minimizes the operation costs, thus obtaining an optimization problem which seeks for the optimal plant operation. Such problem is formulated in terms of mathematical programming.

The optimal distribution strongly depends on the efficiency of each plant, i.e., the specific-steam consumption of each plant and the cooling-water consumption. Thus, the problem objective deals with the trade-off between the cost of live steam and water usage, distributing the cooling water in an optimal way so that the more efficient plants are prioritized as long as the evaporation demand for each spinbath is met. The solution will give the optimal flow of cooling water for each surface condenser given the real-time conditions (fouling state, demand and inlet cooling-water temperature). Automatic differentiation and current gradient-based NLP algorithms will allow to get solutions in real-time.

3.4.1 Mathematical formulation

The first step is to define the set $\mathcal{E}$ of all the evaporation plants which use a surface condenser as cooling system. Hence, the set $\mathcal{E}$ is also the set of the surface condensers. Moreover, this set is divided into two subsets $\mathcal{E} = \{\mathcal{E}_1 \cup \mathcal{E}_2\}$ according to those surface condensers that have access to cooling-water Source 2 or not (see Figure 3.8).

The set of decision variables is $F_{ev} \in \mathbb{R}^+$ which denotes the cooling-water flow that should go through each surface condenser attached to the evaporation plant *ev*. Moreover, the exceeding water which can go from subnet 1 to subnet 2, but not backwards, is a decision variable denoted by $F_L \in \mathbb{R}^+$.

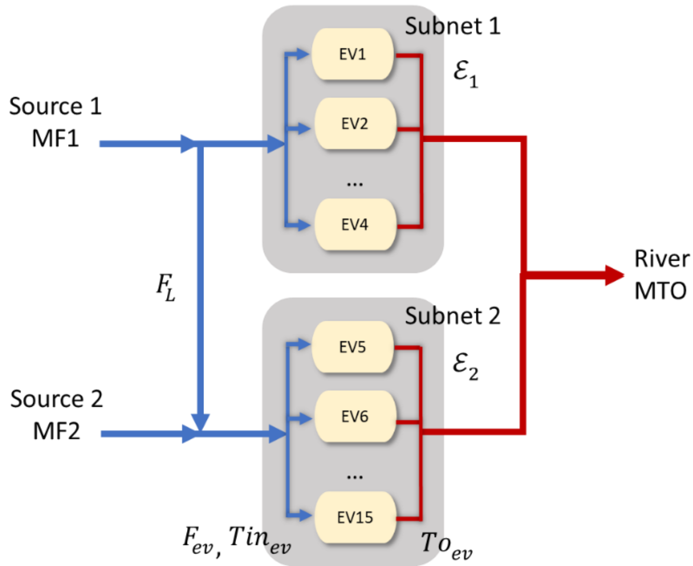


Figure 3.8. Scheme of cooling-water network

The problem constraints are:

1. Mass balance of cooling water in each subnet. For subnet 1, the total flow that feeds the surface condensers in this subnet must be as maximum the available flow at source 1 (MF1) minus the flow that goes to subnet 2 (F_L); meanwhile in subnet 2, the maximum water that the surface condensers can use is the available at source 2 (MF2) plus the remaining water from source 1 (F_L). (See equations (3.13b) and (3.13c), respectively).
2. Lower and upper flow bounds defined for each surface condenser ($\underline{\mathrm{F}_{ev}}, \overline{\mathrm{F}_{ev}}$), i.e., suitable operation range in order to avoid the presence of the spinbath in the surface condenser pipes (3.13d). This effect is particularly harmful for the equipment, so it needs to be avoided.
3. The outlet water temperature resulting from the surface condensers has to be lower than the maximum allowed (MTO), as such outlet water goes to the river and it has to fulfill the constraints imposed by environmental regulations, (3.13e).

Note that, in this formulation, italic symbols represent variables and sets whilst plain ones are known values (i.e., inputs or parameters) for the optimization.

Finally, the objective function is to minimize the operational cost, i.e., a trade-off between the cost of live steam and cooling-water consumption. Thus, it is needed to compute the absolute steam consumption of each plant (ASC_{ev}) according to the cooling-water distribution. It can be obtained from the predicted specific steam consumption and the assigned evaporation load at each plant (EC_{ev}), as shown in:

$$ASC_{ev} = SSC_{ev} \cdot \mathrm{EC_{ev}} \tag{3.10}$$

However, the black-box model (3.9) provides the *increment* of the specific steam consumption, so the absolute steam consumption must be computed using the minimum specific steam consumption (BSSC) that each plant can reach, plus its increment due to the performance of the cooling system (ΔSSC), (3.11). Nevertheless, as the BSSC is a fixed term, for the resource tradeoff optimization problem we must only account for the increment of the absolute steam consumption (ΔASC), defined as shown in (3.12).

$$ASC_{ev} = (\mathrm{BSSC} + \Delta SSC_{ev})\mathrm{EC_{ev}} \tag{3.11}$$

$$\Delta ASC_{ev} = \Delta SSC_{ev}\ \mathrm{EC_{ev}} \tag{3.12}$$

In other words, maximum water flows to SCs correspond to best SSC (i.e., *zero steam cost* in the tradeoff optimization problem) and minimum water consumption leads to the highest steam cost.

Hence, the operational cost tradeoff is given by: *(i)* the absolute steam consumption increment of all the plant times the price of generating a unit mass of live steam in boilers ($\mathrm{P_{steam}}$) plus *(ii)* the net amount of water consumed from the sources times the cost of pumping and pipes operation ($\mathrm{P_{water}}$). Both price values are calculated internally in the company and their values are omitted here due to confidentiality agreements.

The formulated optimization problem is:

$$\min_{F_{ev} \in \mathbb{R}+} J = \sum_{ev}^{\varepsilon} (\Delta ASC_{ev}\ \mathrm{P_{steam}} + F_{ev}\ \mathrm{P_{water}}) \tag{3.13a}$$

s.t.:

$$\sum_{ev}^{\mathcal{E}_1} F_{ev} \leq \mathrm{MF1} - F_L \tag{3. 13b}$$

$$\sum_{ev}^{\mathcal{E}_2} F_{ev} \leq \mathrm{MF2} + F_L \tag{3. 13c}$$

$$\underline{\mathrm{F_{ev}}} \leq F_{ev} \leq \overline{\mathrm{F_{ev}}} \qquad \forall ev \in \mathcal{E} \tag{3. 13d}$$

$$\frac{\sum_{ev}^{\mathcal{E}} To_{ev} \, F_{ev}}{\sum_{ev}^{\mathcal{E}} F_{ev}} \leq \mathrm{MTO} \tag{3. 13e}$$

Where *To* could be calculated by the experimental model of each plant as stated in section 3.3.2, formula (3.6) expanded to all the evaporators (3.13f) , and ΔSSC can be obtained combining (3.8) and (3.9), as (3.13g) shows.

$$To_{ev} = f(F_{ev}) - \mathrm{T}_{in}^{exp} + \widehat{\mathrm{T}}_{\mathrm{in}} - \mathrm{K}_{\mathrm{f}_{ev}} \tag{3. 13f}$$

$$\Delta SSC_{ev} = g\left(\rho \, \mathrm{Cp} \, F_{ev} \cdot \left(To_{ev} - \widehat{\mathrm{T}}_{\mathrm{in,ev}}\right)\right) - \mathrm{BSSC_{ev}} \tag{3. 13g}$$

Finally, as the experimental models are C^1 nonlinear functions, usually quadratic polynomials, and the evaporation load ($\mathrm{EC_{ev}}$) is a known value taken from the plant control system, the optimization problem (3.13) is easily handled via nonlinear programming (NLP) using an interior point optimizer, e.g. IPOPT (Wächter & Biegler, 2006).

3.4.2 Implementation considerations

Once the problem is formulated, some concerns have to be taken into account for its implementation on the real plant as an RTO scheme that supports operators and plant managers in the optimal distribution of the cooling water.

First, as the experimental models have been built for a specific flow range, if the future real-time cooling-water flows $\left(\hat{\mathrm{F}}_{ev}\right)$ are out of such range, the optimization should not be executed, as the estimation of the fouling parameter (K_f) via (3.7) may be wrong due to plant-model mismatch. Therefore, a warning message should appear to inform of such situation.

However, if this happens because the plants are in maintenance, i.e., they are not processing any spinbath, the flow of this evaporator must not be optimized, but the rest of the network can be optimized. To do that, if the evaporation load of an evaporator (EC) is less than a small value, e.g., 1 m^3/h, the flow of that surface condenser is set to its measured value, i.e., constraint (3.13d) is automatically replaced by the following expressions:

$$\begin{cases} \underline{\mathrm{F}_{\mathrm{ev}}} \leq F_{ev} \leq \overline{\mathrm{F}_{\mathrm{ev}}} & \forall ev \in \{ev | \mathrm{EC}_{\mathrm{ev}} > 1\} \\ F_{ev} = \hat{\mathrm{F}}_{ev} & \forall ev \notin \{ev | \mathrm{EC}_{\mathrm{ev}} > 1\} \end{cases} \tag{3.14}$$

Finally, it may be the case that, due to an aggressive state of fouling and/or water inlet temperature to the surface condensers, the optimum cooling power is out of the range where the experimental models were built. In this case, the optimization should not run, and a warning should be displayed in order to inform the operators of this situation.

The optimization problem (3.13) and the above extra considerations have been modeled and coded in MATLAB using the open-source tool for nonlinear optimization CasADi (Andersson et al., 2019). The MATLAB environment was chosen by Lenzing preference, as it allows their experienced engineers to collect the real-time data of the needed parameters from the data acquisition system, for example the PI system (PI System, OSIsoft, n.d.), as well as to maintain the code.

3.4.3 First-test results

The optimization formulated in the previous section has been tested offline with real data taken from the Lenzing database. The values of the parameters used to solve the problem are omitted due to confidentiality reasons. The obtained results are shown along with the real-time measured data for the time that the parameters were recorded.

Comparing the optimized flows with the measured ones (Figure 3.9), it is observed that, although the total amount of cooling water did not decrease, it has been redistributed. This can look inconsistent a priori (especially for those plants that increase their use) as the cooling water has a cost. However, as shown in Figure 3.10, the consumption of steam has decreased. Note that the steam price is 10 times higher than the water one, so when adding both costs in the tradeoff objective function, the result is that benefits have been obtained. Indeed, the cooling-water cost was almost negligible by that time, and the key is to distribute the total available cooling water from the two sources in an optimal way in order to reduce the steam consumption as much as possible.

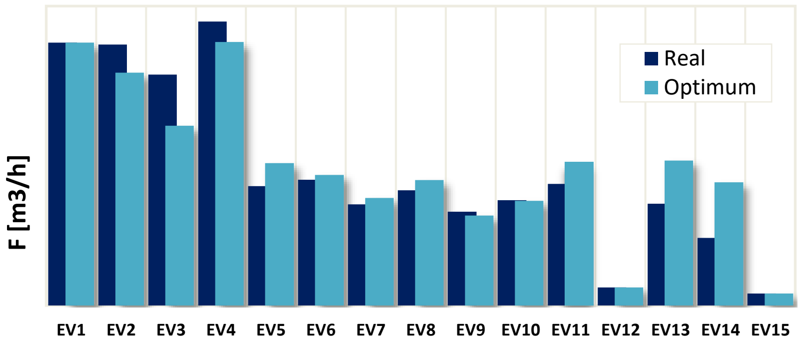


Figure 3.9. Cooling-water distribution

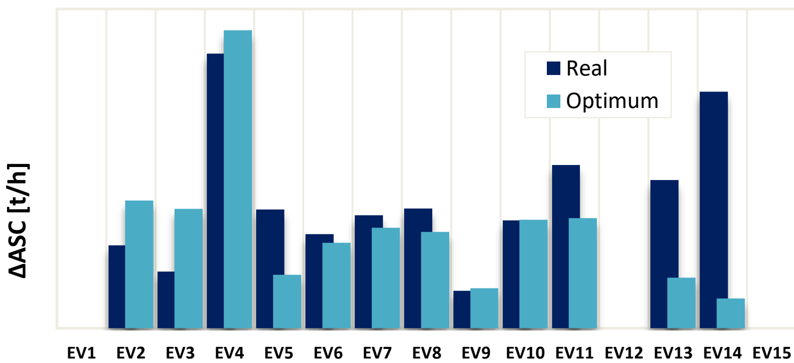


Figure 3.10. Increment of specific steam consumption

It should be noted that the ΔASC of plants 1, 12 and 15 is zero. Analyzing their corresponding cooling-water flows, it can be determined that evaporation plants 12 and 15 are in maintenance, so such flow has not been optimized and there is no steam consumption. Furthermore, a warning message appears in the designed decision-support system indicating it. With respect to plant 1, there is indeed a real steam consumption recorded in the database, but the SC was running at its maximum flow, so the predicted SSC coincides with the BSSC, and the increment of steam consumption is zero.

Comparing the operational current cost and the predicted one applying the optimization results, the savings would be around 16%. Nevertheless, note that such value only represents a snapshot in a particular time instant and only refers to the cost of the cooling water consumption without considering the live steam consumption of the evaporators whose cost is ~30 times greater. Nonetheless, the annual potential benefit that would be obtained by applying this optimization tool in daily operation looks promising.

3.5 Full network optimization

Once the optimization problem for the cooling system has been developed, the goal now becomes to recompute the optimal load allocation for the evaporation network where the power values developed by the cooling systems are also decision variables.

In the previous work (Kalliski et al., 2019) surrogate models were developed to estimate the steam consumption by an evaporation plant. After straightforward manipulations with such models, the steam consumption is found to depend on the evaporation load (EC), on the cooling power of the cooling system (Q^{cp}) and on the state of fouling of the evaporation plant (K_v), i.e., in the state of fouling of the heat exchangers used to heat the spinbath (different fouling that in the surface condensers, K_f). Note that, the specific steam consumption also depends on the recirculated flow and its temperature, but such variables are already controlled in an optimal way (see section 3.1).

$$SSC = \mathrm{a}_1\, K_v \;+\; \mathrm{a}_2\, Q^{cp} +\; \mathrm{a}_3\, EC \tag{3.15}$$

In models (3.15), $\mathrm{a}_1, \mathrm{a}_2$ and a_3 are regression parameters obtained from those in (Kalliski et al., 2019).

The evaporation-load allocation problem consists in assigning the amount of evaporation load that each plant must process, for the five different spinbaths, in order to minimize the overall live steam consumption. Two sets are defined to build the problem formulation; one for all the evaporation plants $\mathcal{E}$ (same set that in the cooling-water distribution problem), and a set for all the different spinbaths that must be processed (called products hereinafter), $\mathcal{P}$.

The decision variables that relate the above introduced sets are:

- $EC_{ev,p} \in \mathbb{R}^+$: evaporation load of product p that the evaporation plant *ev* must achieve.
- $X_{ev,p} \in \{0,1\}$: links the plant *ev* to process product *p*.

The constraints being considered are the following:

Each plant can only process one product at a time (3.16), and each product requires a different amount of water to be removed, i.e., evaporation load demand that must be fulfilled (3.17).

$$\sum_{p}^{\mathcal{P}} X_{ev,p} \leq 1 \qquad \forall ev \in \mathcal{E} \tag{3.16}$$

$$\sum_{ev}^{\mathcal{E}} EC_{ev,p} \geq \mathrm{SP_p} \qquad \forall p \in \mathcal{P} \tag{3.17}$$

Furthermore, not all plants can process all the different products, mainly because not all connections from plants to products are physically feasible (3.18).

$$X_{ev,p} = 0 \qquad (ev,p) \in \mathcal{M} \tag{3.18}$$

Table 1 shows the available links between products and plants, where the ticks indicate that a link is feasible. Thus, the set $\mathcal{M}$ is composed of the combination of products and plants which are not physically connected and therefore cannot be linked.

Table 1. Connections between products and plants. The ticks indicate feasible connection.

	EV1	EV2	EV3	EV4	EV5	EV6	EV7	EV8	EV9	EV10	EV11	EV12	EV13	EV14	EV15
$\mathcal{P}_1$	✓	✓	✓	✓		✓						✓			
$\mathcal{P}_2$					✓		✓	✓			✓	✓	✓		
$\mathcal{P}_3$	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\mathcal{P}_4$		✓		✓											
$\mathcal{P}_5$							✓		✓		✓	✓	✓	✓	

Finally, the achievable evaporation load in each plant is bounded due to actual equipment features:

$$\underline{\mathrm{EC}_{\mathrm{ev}}} X_{ev,p} \le EC_{ev,p} \le \overline{\mathrm{EC}_{\mathrm{ev}}} X_{ev,p} \qquad \forall ev \in \mathcal{E}, \forall p \in \mathcal{P} \tag{3.19}$$

Where $\underline{\mathrm{EC}_{\mathrm{ev}}}$ and $\overline{\mathrm{EC}_{\mathrm{ev}}}$ state the minimum and maximum evaporation load for a plant to operate correctly. Moreover, to ensure that $EC_{ev,p}$ take zero value for the products not linked to plant *ev*, the limits are multiplied by $X_{ev,p}$.

The objective function is to minimize the live steam consumption given by the specific steam consumption times the evaporation load (see (3.10)). Thus, the evaporation-load allocation problem is summarized in (3.20).

$$\begin{aligned} \min_{\substack{EC_{ev,p} \in \mathbb{R}+ \\ X_{ev,p} \in \{0,1\}}} \quad & J = \sum_{ev}^{\mathcal{E}} \sum_{p}^{\mathcal{P}} SSC_{ev,p}\, EC_{ev,p} \\ s.t.: \quad & (3.16) - (3.19) \end{aligned} \tag{3.20}$$

To solve the optimization problem, one needs to know the values of the state of fouling of each evaporation plant (K_v) and of the cooling power developed by the cooling system (Q^{cp}). The first one can be obtained online following the procedure in (Kalliski et al., 2019), comparing the model predictions with the measured values. Taking the cooling power as a parameter, the overall problem becomes a mixed-integer quadratic programming (MIQP) problem as it involves binary and continuous variables, and quadratic terms in the objective function. However, the cooling power by the surface condensers is now variable through the cooling water distribution (see equation (3.8)).

At the same time, in the cooling-water distribution problem (3.13), the evaporation load is taken as a fixed parameter. However, both problems are coupled (see Figure 3.11) and should be solved together. This can be done following different alternatives.

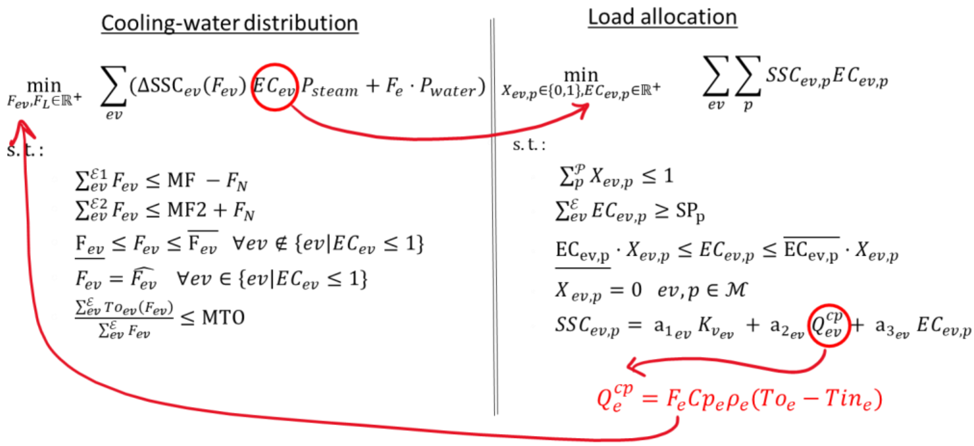


Figure 3.11. Relation between cooling-water distribution and load allocation problems

In this thesis, we propose three different approaches:

1. Solve both problems consecutively in an iterative way, using the results of one problem as the input data to the other one until convergence (see Figure 3.12).

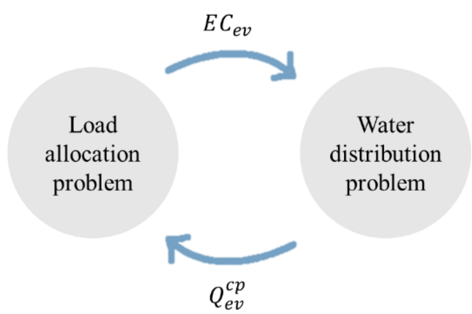


Figure 3.12. Sequential procedure scheme

2. Combine both formulations into a single centralized optimization problem.

3. Address the optimization in a distributed fashion, i.e., adding the shared variables as a penalty in the objective function for each individual problem and including a coordination layer that compares the shared variables in each iteration and updates the resource prices (see Figure 3.13).

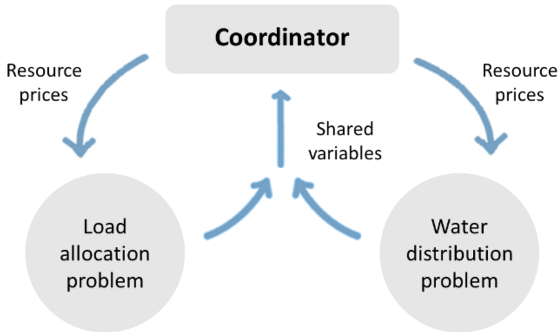


Figure 3.13. Distributed approach scheme

How these different approaches are applied to this case is going to be explained hereunder.

3.5.1 Sequential approach

The first alternative is to solve both problems sequentially and iteratively. Hence, a first local problem can handle the plants evaporation load allocation, whereas a second one optimizes the water distribution, where only the magnitudes EC and Q^{cp} are shared between both problems.

The key idea is to solve the evaporation-load allocation problem (3.20) taking the Q^{cp} as known (computed by (3.8) using the real-time measures of cooling water flow and temperature). Then, from the obtained solution, to take the values of the EC and solve the cooling-water distribution problem (3.13), getting a new set of values to compute the Q^{cp} and so on (see Figure 3.12). It will last until the solution of the problems does not change. The procedure is formalized in Algorithm 1.

Algorithm 1. Sequential optimization

1: Compute Q_{ev}^{cp} from current $\widehat{\mathrm{F}}_{ev}$, $\widehat{\mathrm{T}}\mathrm{o}_{\mathrm{ev}}$, $\widehat{\mathrm{T}}_{\mathrm{in}}$ and set ξ = 1
2: While ξ > 0.001 do
3: Solve (3.20) and save the computed allocation EC_{ev}
4: Solve (3.13) using the EC_{ev} from step 3
5: Compute $\boldsymbol{Q}_{ev}^{cp}$ from the F_{ev} and To_{ev} obtained in step 4
6: Compute $\xi = \sum_{ev}^{\mathcal{E}} \left|\left|Q_{ev}^{cp} - \boldsymbol{Q}_{ev}^{cp}\right|\right|_2^2$
7: Update $Q_{ev}^{cp} = \boldsymbol{Q}_{ev}^{cp}$

The reader may note that, in the evaporation-load allocation problem, $EC_{ev,p}$ depends on the evaporation plant *ev* and the product p, meanwhile in the cooling-water distribution problem EC_{ev} only depends on the evaporation plants. Nevertheless, such connection is just given by (3.21).

$$EC_{ev} = \sum_{p}^{\mathcal{P}} EC_{ev,p} \tag{3.21}$$

Finally, as the sub-problems are NLP and MIQP, the computational time to solve each of them is few seconds, so the time to reach a solution following the Algorithm 1 only depends on the number of iterations, that are also a few. Of course, following this strategy there is no global optimality guarantee.

3.5.2 Centralized approach

Another alternative consists in mixing both formulations in order to get a monolithic problem. Thus, the decision variables are the ones of each problem, i.e.:

- $EC_{ev,p} \in \mathbb{R}^+$: evaporation load of product *p* that the evaporation plant *ev* must achieve.
- $X_{ev,p} \in \{0,1\}$: links the plant *ev* to process product *p*.

- $F_{ev} \in \mathbb{R}^{+}$: cooling-water flow that should go through the surface condenser attached to the evaporation plant *ev*.
- $F_L \in \mathbb{R}^{+}$: exceeding cooling water from source 1 that goes to subnet 2.

The objective function to minimize comes from the operational cost of the overall network (live-steam and cooling-water consumption):

$$J = \sum_{ev}^{\mathcal{E}} (ASC_{ev}\,\mathrm{P_{steam}} + F_{ev}\,\mathrm{P_{water}}) \tag{3.22}$$

However, in contrast to the cooling-water distribution problem, equation (3.15) must be used here to compute the specific steam consumption instead of the increment of it. Hence, the absolute steam consumption is obtained using (3.10).

The problem constraints are a collection of both problems:

- Each plant can only process one spinbath at a time (3.16).
- Total evaporation demand per spinbath must be fulfilled (3.17).
- There are impossible connections between products and plants (3.18).
- Evaporation load of each plant is bounded (3.19).
- Total water consumption in each subnet must be lower than the available at the sources, considering that exceeding water can go from subnet 1 to subnet 2 but not backwards (3.13b)-(3.11c).
- Cooling-water flows are also bounded, considering that if a plant has not got any product assigned, such flow should be zero. Thus, equation (3.11e) is modified as follows:

$$\underline{\mathrm{F}_{\mathrm{ev}}} \textstyle\sum_{p}^{\mathcal{P}} X_{ev,p} \leq F_{ev} \leq \overline{\mathrm{F}_{\mathrm{ev}}} \textstyle\sum_{p}^{\mathcal{P}} X_{ev,p} \qquad ev \in \mathcal{E} \tag{3.23}$$

- Outlet-water average temperature must be lower than a limit stated by the environmental regulation (3.11e), where the outlet temperature of each surface condenser is obtained using (3.6).

The presence of discrete and continuous variables as well as the nonlinear dependency to compute the outlet-water average temperature and the absolute steam consumption (the specific steam consumption is also nonlinear) makes the centralized problem an MINLP one. However, it is computationally affordable for real-time implementations with current computer power and branch & bound NLP solvers.

3.5.3 Distributed approach

The final alternative is to solve both problems independently but in a coordinated fashion, i.e., to address the optimization in a distributed way using Lagrangean decomposition and price-coordination schemes (see section 1.2.4).

Thus, as in the sequential approach, there is a local problem which handles the plants evaporation load allocation and another which optimizes the water distribution. Nevertheless, instead of treating the shared magnitudes ($R_{i,j} \coloneqq \{EC_{ev}, Q_{ev}^{cp}\}$) as fixed parameters whose value depends on the solution of the corresponding problem, such magnitudes are handled as shared variables (R) between both problems. Then, a coordination layer must be added to progressively force both sets of decision variables to be equal in both local optimizations (see Figure 3.14).

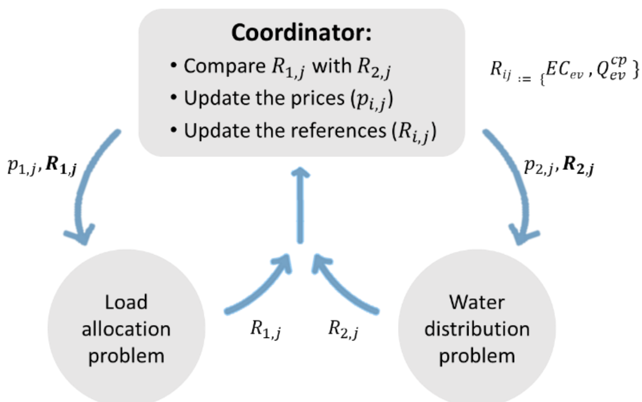


Figure 3.14. Distributed approach expanded scheme

First, the optimization problems (3.20) and (3.13) should add a constraint imposing that the shared variables ($R_{i,j}$)[1] must be equal to a reference ($\boldsymbol{R_{i,j}}$). As usual in Lagrangean decomposition, the constraints will be added as a penalty in the respective objective functions multiplied by a shadow price for *resource* utilization ($p_{i,j}$). Thus, the new objective function (J') of each sub-problem will be as (3.24) shows.

$$J'_i \coloneqq J_i + \frac{1}{2}\sum_{j}^{J} p_{i,j}\left(R_{i,j} - \boldsymbol{R_{i,j}}\right)^2 \tag{3.24}$$

Furthermore, as the shared variables are not exactly the decision variables of the sub-problems, it is needed to add the corresponding constraints that compute such shared variables. In the evaporation-load allocation problem, $EC_{ev,p}$ depends on the evaporation plant *ev* and the product *p*, meanwhile in the cooling-water distribution problem EC_{ev} only depends on the evaporation plants, so in the evaporation-load allocation problem we need to add the constraint (3.21).

Hence, the sub-problems to solve are:

1. Evaporation-load allocation problem:

$$\min_{\substack{EC_{ev,p}\in\mathbb{R}+ \\ X_{ev,p}\in\{0,1\} \\ R_{1,j}\in\mathbb{R}+}} \quad J'_1 \coloneqq \sum_{ev}^{\mathcal{E}} ASC_{ev}\,\mathrm{P}_{\mathrm{steam}} + \frac{1}{2}\sum_{j}^{J} p_{1,j}\left(R_{1,j} - \boldsymbol{R_{1,j}}\right)^2 \tag{3.25}$$

$$s.t.: \qquad (3.16) - (3.19), (3.21)$$

[1] The sub-index *i* refers to the sub-problem, meanwhile the sub-index *j* refers to the shared variables, i.e., $R_{1,1} = EC_{v1}$ in the load allocation subproblem.

2. Cooling-water distribution problem:

$$\min_{\substack{F_{ev}\in\mathbb{R}+\\R_{2,j}\in\mathbb{R}+}} \quad J'_2 \coloneqq \sum_{ev}^{\varepsilon}(ASC_{ev}\,\mathrm{P}_{\mathrm{steam}} + F_{ev}\,\mathrm{P}_{\mathrm{water}}) + \frac{1}{2}\sum_{j}^{J} p_{2,j}\left(R_{2,j} - \boldsymbol{R_{2,j}}\right)^2 \tag{3.26}$$

$$s.t.: \qquad (3.8), (3.11\mathrm{b}) - (3.11\mathrm{e})$$

In each iteration, the coordinator layer receives the values for the shared variables got by each sub-problem, compares them, and updates the reference values as well as the shadow prices for the next iteration according to the following rules:

$$p_{i,j}^{[k+1]} = p_{i,j}^{[k]} + \mu^{[k]}\left(R_{i,j}^{[k]} - \boldsymbol{R_{i,j}^{[k]}}\right)^2 \tag{3.27}$$

$$\mu^{[k+1]} = \mu^{[k]}\,\lambda \tag{3.28}$$

$$\boldsymbol{EC_{i,ev}^{[k+1]}} = EC_{1,ev}^{[k+1]} \tag{3.29}$$

$$\boldsymbol{Q_{i,ev}^{cp\,[k+1]}} = Q_{2,ev}^{cp\,[k+1]} \tag{3.30}$$

Where notation $R_{i,j}^{[k]}$ denotes the values of shared variables R_j, solution of sub-problem i at iteration k. Progressive hedging (Rockafellar & Wets, 1991) is used in (3.28) to update the factor μ in each iteration, via the user-defined parameter λ. Nevertheless, there are many ways to update the prices, being *parametric sensitivities* based on Newton's method (Ganesh & Biegler, 1987) one with the best results in terms of convergence speed.

In this particular case, the references $\boldsymbol{R_{i,j}}$ are updated according to the sub-problem which optimizes such variables: the load allocation problem provides the evaporation load, and the cooling-water distribution problem provides the cooling-water flows used to obtain the cooling powers. Nonetheless, any other methodology from the literature can be applied, as the one proposed by (Martí, 2015) that minimizes the worst constraint violations that a modification of the local control inputs would cause.

The overall procedure is summarized in Algorithm 2.

Algorithm 2. Distributed optimization

1:	Set $\boldsymbol{R_{i,j}}$ to current values from the plants and $p_{i,j} = \xi = 0.1$
2:	While ξ > 0.001 do
3:	Solve problems (3.25) and (3.26), and get new values of $R_{i,j}$
6:	Compute $\xi = \sum_{ev}^{\mathcal{E}} \left\|\left\|R_{i,j} - \boldsymbol{R}_{i,j}\right\|\right\|_2^2$
7:	Update prices $p_{i,j}$ and references $\boldsymbol{R}_{i,j}$

In this way, the sub-problems continue being MIQP and NLP but, due to the methodology used to update the prices and references, it is expected that improved optimality can be reached with this procedure in comparison to the sequential methodology. Note also that this strategy has computational benefits as well, as the sub-problems can be solved in parallel.

3.5.4 Results and discussion

For the sake of comparison/analysis, the same problem has been solved from a particular network situation following the three presented methods above, all coded in Pyomo/Python (Hart et al., 2011, 2017). For the sequential and distributed approaches, the solver used for the evaporation-load allocation problem is GUROBI (Gurobi Optimization, 2018), meanwhile the cooling water subproblem has been solved using IPOPT (Wächter & Biegler, 2006). The centralized problem has been solved using two different solvers: based on a global-deterministic method, BARON (Sahinidis, 1996), and based on local-deterministic method, BONMIN (Bonami et al., 2008). For this problem, both solvers get practically the same solutions, which means that using the faster BONMIN for real-time purposes provides solutions near the global optimum.

Table 2 and Figure 3.15 show the results of the evaporation load allocation and the cooling-water distribution obtained by the different approaches. Note that the label *Real* refers to the historical data used to run the optimizations and, as both results obtained with the centralized approach (the ones obtained using BARON and the ones obtained with BONMIN) are practically the same, they are only once reflected to simplify the comparison between approaches.

Table 2. Evaporation load allocation obtained with the different approaches

	EV1	EV2	EV3	EV4	EV5	EV6	EV7	EV8	EV9	EV10	EV11	EV12	EV13	EV14	EV15
Real	19.4	16.8	16.1	27.0	7.7	7.5	7.3	9.2	6.2	7.3	6.9	0.0	6.9	6.0	0.0
Sequential	12.1	16.2	24.0	27.0	0.0	5.5	4.1	7.7	8.8	10.4	8.0	0.0	9.0	11.0	0.0
Centralized	0.0	17.2	24.0	26.0	7.0	5.5	4.9	8.7	11.5	11.0	8.0	0.0	9.0	11.0	0.0
Distributed	16.9	16.9	24.0	26.4	0.0	5.5	6.1	7.7	6.7	5.6	8.0	0.0	9.0	11.0	0.0

P1 P2 P3 P4 P5

Comparing the results of the evaporation load allocation, one can see that the product demand can be reached using only twelve plants, i.e., in addition to plant 12 and 15 that are not being used due to be in maintenance, the optimal operation is to have another plant in standby. This means that another maintenance task could be carried out or that the production could be increased. Furthermore, the sequential and the distributed approach give the same allocation with respect to the products, but the evaporation load of some plants varies.

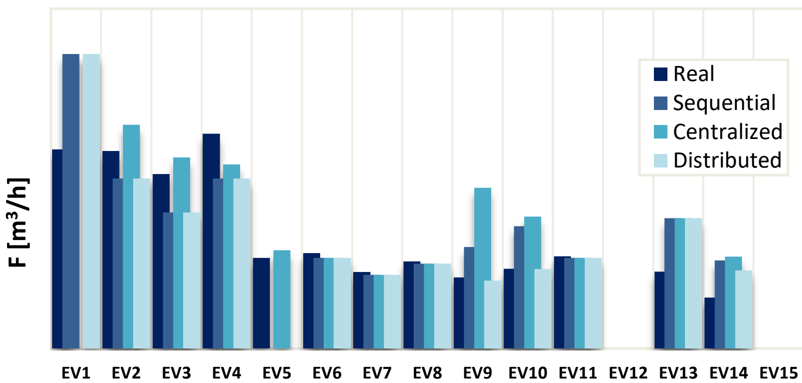


Figure 3.15. Cooling-water distribution obtained with the different approaches

With respect to the cooling water distribution showed in Figure 3.15, one can see that it changes with respect to Figure 3.9, as now the evaporation load allocation is optimal for the new water distribution. Nevertheless, the total amount of cooling water used is once again the maximum available, as the price of the cooling water is lower than the price of the live steam, so its high usage compensates the reduction of live-steam consumption. But this situation may change in the future.

Finally, Table 3 shows a brief summary of the computational time and the value of the objective function for each alternative in order to compare the obtained results. Comparing the values of the objective function for the different cases, the conclusion is that, as expected, the three approaches give significant savings with respect to current practices in Lenzing. Nevertheless, the centralized problem reaches the best solution (savings around 25%), meanwhile the sequential and distributed approaches get a solution more than a 10% worse (savings around 17% and 10%, respectively). The difference between the solutions obtained with the global-deterministic based solver and the local based one is negligible, but the computational time is not.

Table 3. Operation cost and computational effort for the three approaches.

	Total CPU time[2] (s)	**Operation cost (€/h)**
Real	-	501.96
Sequential	33.4	416.11
Centralized (BARON)	7908	379.08
Centralized (BONMIN)	7.78	379.34
Distributed	293	452.95

A suitable time window between real-time optimizations in this network is about 30 minutes, so the centralized approach solved with a global method is not suitable. Hence, any of the other approaches could be admissible. As the centralized formulation solved with BONMIN gives the best results, it is the selected to be the best option. However, consider that, if the problem grows by considering fouling predictions over time (fouling models) or by including more parts of the factory (the water network is larger and also serves

[2] Over an Intel® i7-7700CPU machine with 32Gb of DDR4 RAM memory.

to other processes, and the heat-recovery network also interacts with the evaporation one), the centralized solution may be far for getting solutions within the required time period.

3.6 Summary and conclusions

In this chapter, a problem on resource efficiency has been addressed. First, different black-box models have been obtained which allow not only to know the relation between key process variables, but also to execute an automatic monitoring of the surface condensers fouling state based on real-time measurements. Moreover, as typical regression procedures are quite sensitive to disturbances in the data and sensor noises, a more sophisticated modeling routine to improve the computation of these curves was described.

Such models have been incorporated to the mathematical formulation that describes the cooling system of the Lenzing's evaporation network. The resulting network model is then used in an RTO scheme that solves an NLP problem according to the current production constraints and the plants fouling states. The aim of the RTO is supporting the network operators to take better decisions in real time by an optimal cooling-water distribution, which minimizes the resource utilization cost.

Nevertheless, the instantaneous RTO does not take into account any prediction of the fouling effect, so the proposed control actions may be suboptimal in the long term if they affect the fouling process. Furthermore, there are out-of-range situations due to the limited experiments done to get the training data in the black-box model development.

Finally, as the surface condensers network is attached to the evaporation one, the cooling-water distribution problem and the evaporation-load allocation problem are coupled. Thus, it is mandatory to solve them jointly in order to get the optimal operation of the overall system.

Both models have been incorporated in three different optimization schemes, analyzing the convenience of the decomposition approaches versus the centralized one. In this case study, the three tested approaches solve the problem in acceptable CPU time for real-time purposes, although the centralized approach has arisen as the most efficient one, getting the best solution. Apparently, the sequential and distributed approaches do not have any advantage here: the predicted cost is higher, and they elapse more time than the centralized formulation solved with a local-search MINLP algorithm. This is because the centralized problem is not so large scale in this case, many constraints are affine in the decision variables, and the advantages of parallel computation cannot be exploited with just two local optimization subproblems. Nevertheless, formulating two effortless independent problems (MIQP and NLP) will clearly beat the centralized formulation if the problem is extended to consider uncertainty in an explicitly way via, for instance, two-stage stochastic optimization. Comparing the two decompositions approaches one can conclude that the formulation based on price-coordination gives worse results and elapses more time. Thus, although it is less sophisticated, for this case study the sequential approach results a better option. This may be due to the characteristics of the problems, as usually price-coordination is used only for problems without discrete variables.

It is worth mentioning that using any of both RTO developed schemes (the one for the cooling water distribution and the one that also considers the load allocation) in a real plant will be very beneficial. Comparison the operation between the historical data and the provided solution for the same conditions shows that if the solutions were implemented significant saving would be obtained. Thus, implementing the scheme in the plant will reach the aim of supporting the operators to take better decisions on how to operate the networks. Nonetheless, as the cooling water distribution RTO only considers the cooling water cost, it would be more worthy the implementation of RTO scheme formulated with the centralized approach for the whole system.

The work presented in this chapter has been presented in a national and two international congresses: the development of the data driven models and the cooling system optimization are addressed in (M. P. Marcos, Pitarch, Jasch, et al., 2018; M. P. Marcos, Pitarch, Prada, et al., 2018) , meanwhile the study of the different formulations for the joint optimization of the cooling system and the evaporation network is broached in (M. P. Marcos, Pitarch, Jasch, et al., 2019).

Chapter 4

Optimal operation of the heat-recovery network

This chapter is focused on the heat-recovery network of the Lenzing site described in Chapter 2. In here, the modeling of the network and the development of a real-time optimization are addressed, taking explicit consideration of the state of fouling in the heat exchangers. In order to do that, models for the heat-transfer coefficients of each equipment have been developed as well.

First, a more detailed description of the network is presented in Section 4.1 and the main objectives for this case study will be listed in Section 4.2. Specific domain knowledge on heat transfer is used together with machine learning in Section 4.3 to build gray-box models for heat exchangers. After that, in Section 4.4 the mathematical formulation of the optimization problem for the optimal operation of the network is described. Section 4.5 analyzes the optimal operation of the heat-recovery network jointly with the evaporation network. Finally, in Section 4.6, the most important aspects found, and goals achieved in this chapter are summarized.

4.1 Problem identification

The heat-recovery network must heat up some of the spinbath concentrated streams after the evaporation network (called product streams hereinafter) as it was described in Chapter 2. The fifteen heat exchangers that form the network, should be able to heat up the products to the desired temperatures in order to be reintroduced in the spinning machines, Four different hot streams are used as utilities to this aim. However, as each one has traces of different chemical components, they cannot be mixed, so, each heat exchanger can only be working with one heat source at a time, despite most of them can be connected to more than one (see Figure 4.).

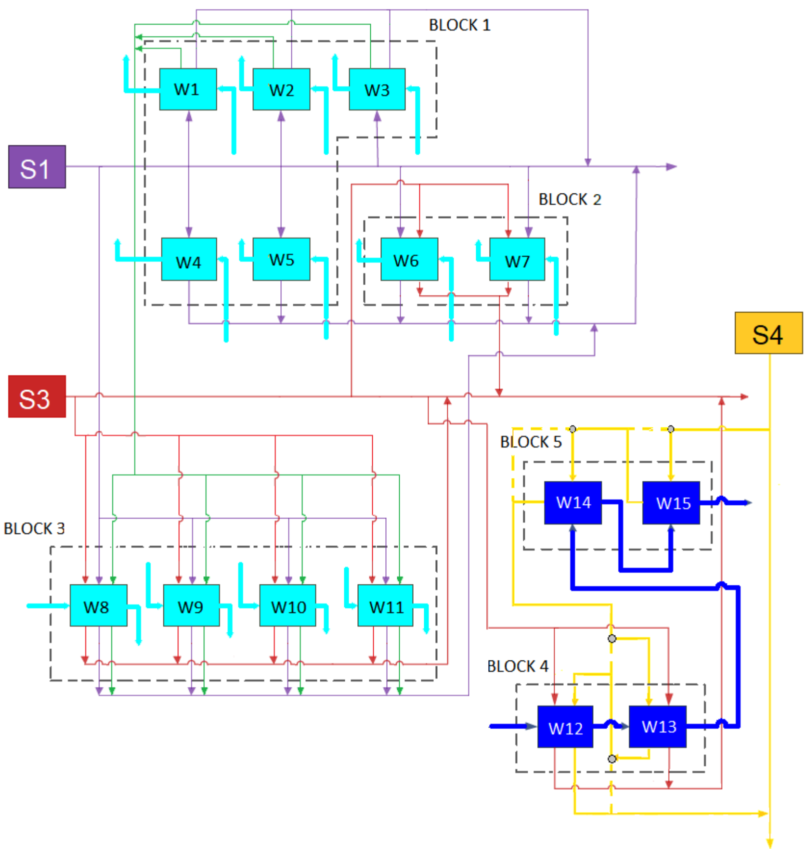


Figure 4.1. Scheme of the network layout. W represents heat exchangers while S heat sources. Thick lines are the product streams meanwhile thin lines are the heat sources streams. Green outlets from exchangers in Block 1 can be reused to feed exchangers in Block 3. The products of cyan heat exchangers are connected in parallel, meanwhile the product of the dark blue heat exchangers goes in series.

The available heat sources are:

- Alkaline wastewater, called S1 hereinafter, represented in purple in Figure 4.1.
- Alkaline wastewater from source S1 that has been *already used* in an exchanger belonging to Block 1. It will be treated as a *virtual heat source* denoted by S2 hereinafter and represented by the green pipes in Figure 4.1.
- Vapor condensate, called S3 hereinafter, represented in red in Figure 4.1.
- Acid wastewater, called S4 hereinafter, represented in yellow in Figure 4.1.

Note that, as these streams come from other processes and are shared among other parts of the plant, the use of these sources involves a cost (pumping and maintenance plus the shadow cost of utilization for other purposes) and its availability is limited too.

The heat exchangers can be grouped, depending on their connectivity to sources (see Figure 4.1). The possible connections between heat sources and heat exchangers are given by Table 4.

Table 4. Allowed source connections to heat exchangers

Block	**Heat exchanger**	**Heat source**
$\mathcal{B}_1$	w1, w2 w3, w4, w5	S1
$\mathcal{B}_2$	w6, w7	S1, S3
$\mathcal{B}_3$	w8, w9, w10, w11	S1, S2, S3
$\mathcal{B}_4$	w12, w13	S3, S4
$\mathcal{B}_5$	w14, w15	S4

From the point of view of the heat-exchanger connections, the network can be divided into two groups. The first one involves the heat exchangers being connected in parallel, i.e., each one heats a different product and the heat sources feed these exchangers in parallel. These are the ones depicted in cyan in Figure , i.e., the exchangers w1 to w11.

The other group is composed of the heat exchangers connected in series from the point of view of product, i.e., the same product goes through all of them (represented in dark blue in Figure 4.1). Hence, the temperature setpoint of the product only has to be fulfilled at the outlet of w15. Note that the S4 source goes in series through these heat exchangers too, but backwards: first it goes to w15 and subsequently passes through w14, w13, and finally through w12. An important feature to take into account in this subset is that, as the product is connected in series, there is no need to use all the heat exchangers as long as the setpoint can be reached with just some of them. If one heat exchanger of the chain is not used, it means that there is not any source passing through it, so there will not be heat transfer and, consequently, the outlet temperature of the product stream will be assumed the same that the one at its inlet. This operation mode is possible because in the source inlet of each heat exchanger there is a bypass valve that can only be in two positions, open and closed.

The efficiency of the heat exchangers depends on different factors, such as constructive materials, dimensionality factors or their operation conditions (Boccardi et al., 2010). The most important ones are the transmission area (i.e., exchanger size) and the fouling state. This network uses plate-type heat exchangers, i.e., they are composed of many thin, slightly separated plates that have very large surface areas so that alternate fluids (hot and cold) flow between them to maximize the heat transfer. The total heat-transfer area in these heat exchangers depends on the number of plates. In addition, the heat exchangers suffer from fouling that increases their thermal resistance, so that their efficiency progressively reduces over time. Consequently, more flow from the heat source is needed to reach the product setpoint. Eventually, the exchanger needs to be cleaned in order to recover its nominal efficiency. The usual cleaning policy in the factories is exclusively based on operators experience, prioritizing first the heat exchangers being more days in operation since last cleaning. In this way it is hard to infer whether the taken decisions really lead to an optimal operation regarding economics.

It is noteworthy that, although the decisions to take may be repetitive and are within a well-defined and stable context, achieving an efficient network operation is a *semi-structured* decision problem, because there is a large number of decision alternatives regarding the network configuration and the cleaning policy (that are difficult to simultaneously manage by a human operator) and the implications of these are not well known, as prediction models for the heat transfer with fouling in these exchangers are not readily available.

4.2 Objectives

The main objective is improving the heat-recovery network operation by optimizing the resource utilization in real time while satisfying a set of production constraints and taking the fouling state of the equipment into explicit consideration.

This requires knowing the dependence of the heat exchanger heat transfer on flows and temperatures of the streams that pass through them. With that purpose in mind, data-driven models of the heat transfer coefficient have been obtained, that must also consider the state of fouling and how it affects the heat transmission.

Once the heat transfer models are obtained, they can be used on a bigger hybrid model which represents the heat-recovery network layout and all possible alternatives of operation. The problem can then be formulated as an optimization one, operating in real-time, with an objective function that reflects the operation costs to be minimized under the constraints imposed by the model and other process restrictions. The resulting RTO problem gives the optimal distribution of the heat sources among the heat exchangers.

Furthermore, as the model takes explicit consideration of the state of fouling of the heat exchangers, the optimization problem also includes a suggestion of which heat exchanger should be cleaned.

In summary, we aim:

- To obtain reliable data-driven models of the heat transfer coefficients of the heat exchangers.
- To formulate an optimization problem to distribute the heat sources according to measured real-time conditions.
- To include in the formulation of the problem a variable which suggests the next heat exchanger that should be cleaned.
- To implement and test the proposed RTO solution.

4.3 Modeling the heat transfer

The aim of the heat-recovery network is to heat up some products with several heat sources. In order to obtain a mathematical model that represents the possible alternatives of network operation, the heat transmission in the heat exchangers must be studied.

It is commonly admitted that the heat transfer through a heat exchanger (Q') can be computed by:

$$Q' = U \cdot \mathrm{A} \cdot \ LMTD \tag{4.1}$$

Where A is the total heat-transfer surface, *LMTD* is the logarithmic mean temperature difference between both inlet and outlet of the two streams through the heat exchanger, and U is the overall heat-transfer coefficient.

Once in operation, the overall heat-transfer coefficient is the key variable from which the exchanger efficiency depends on. However, it is not constant as it depends on the streams densities, viscosities, flow velocities, etc., i.e., it depends on the operating conditions. Moreover, the fouling in the exchanger surfaces also modifies the conduction coefficient. Since the aim of the proposed RTO is to compute the right flows through each heat exchanger for economic operation, a prediction model for U with respect to these variables is required.

Well established empirical laws based on dimensionless groups could be recalled for such a task. However, this approach requires a precise knowledge on the heat-exchanger dimensions and constructive materials, as well as sensible data on the stream properties (dynamic viscosity, density, etc.), in order to find the more suitable correlation among the Nusselt number with the Prandtl, Reynolds or Peclet ones, plus with other of kinematic and geometric nature (Boccardi et al., 2010; Dorao & Fernandino, 2017). The drawbacks with this approach are twofold:

Formulas for the convection coefficient are well known in the case of usual fluids, like water, steam, or most common refrigerants, but no significant studies are available for other scarcer fluids, mixture of several components at varying concentrations, like the acid spinbaths in this case study. Hence, laboratory experiments are required to collect the relevant data for regression.

In *dirty* industrial environments, streams and heat-exchanger features vary with the time and operation conditions (concentrations, fouling, etc.), so the regressions gotten from the lab data may be quickly outdated if such data is not representative enough for the whole operating region, which barely can be.

In addition, these highly non-convex models rely on empirical data in the end, but the only data available online in the plant are flows and temperatures. Therefore, it was decided to avoid this way choosing a more flexible data-driven approach.

The idea is to build a data-based model for the *clean exchanger* gathering data from the plant historian in different operating conditions (flow ranges) but always after a cleaning task. Nevertheless, a data reconciliation must be done before building the black models in order to correct the raw data according to basic energy balances.

The developed models will establish the relation between the overall heat-transfer coefficient and the flows and temperatures of the streams, without taking the fouling state into account.

4.3.1 Data reconciliation

Usually, raw plant data may present inconsistencies, for instance due to noisy or biased sensors, making it necessary to correct the raw data before its further use to take decisions. Thus, the first step is performing a reconciliation of the raw data collected from sensors using the basic first-principles laws (e.g., mass balances). The procedure is solving an optimization problem in which the objective function to minimize is the weighted sum of squares of the deviations between measured data with respect to their estimated (i.e., corrected) values, subject to the physical laws (process model) and other additional imposed relations (Sarabia et al., 2012).

In this case, the measured variables that are necessary to correct are the inlet and outlet temperatures of the product streams (Tin_c, To_c), those of the heat sources (Tin_s, To_s) and their respective flows (F_c, F_s). Thus, in the data reconciliation problem, for a given dataset of N different operating instants i after a cleaning task of the previous measured variables $\widehat{y_i} \coloneqq \{\widehat{Tin}_{c,i}, \widehat{To}_{c,i}, \widehat{Tin}_{s,i}, \widehat{To}_{s,i}, \widehat{F_{c,i}}, \widehat{F_{s,i}}\}$, there is an equivalent reconciliated variable $y_i \coloneqq \{Tin_{c,i}, To_{c,i}, Tin_{s,i}, To_{s,i} F_{c,i}, F_{s,i}\}$. It must be mentioned that the variables and parameters have been normalized in order to avoid that the variables that take higher values influence the optimization algorithms with a greater weight, as well as to facilitate their convergence. This normalization has been done through (4.2) where μ_i and σ_i are the mean value and the standard deviation of the variable y_i respectively and Y_i is the value of the corresponding normalized variable.

$$Y_i = \frac{y_i - \mu_i}{\sigma_i} \tag{4.2}$$

The objective function to minimize is the error between the variables that fulfill the constraints and the corresponding measured values (4.3). Note that in this case we are using least squares, but the fair function (3.2) could be also used for a more robust estimation.

Finally, data reconciliation is constrained to the energy balance in the heat exchanger (4.4) where, assuming that there is no heat loss to the ambient, the heat that the heat source gives (Q_s) has to be equal to the heat that the product gains (Q_c). Both heat magnitudes can be computed using (4.5) where ρ and Cp are the density and the specific heat of the stream respectively, and ΔT is the difference between the inlet and the outlet flow temperatures[3].

The optimization problem for the data reconciliation of the collected data of each heat exchanger is:

$$\min_{Y_i} \quad J = \sum_{i}^{N} \left(\frac{\hat{Y}_i - Y_i}{\hat{Y}_i} \right)^2 \tag{4.3}$$

$$\text{s.t.:} \quad Q_{s,i} = -Q_{c,i} \tag{4.4}$$

$$Q_{k,i} = F_{k,i}\, \rho_k\, Cp_k\, \Delta T_{k,i} \qquad \forall k \in \{s, c\} \tag{4.5}$$

Note that in this formulation the sub-index s and c correspond to the heat source and product stream respectively, and the sub-index i for the operating point.

A sample of the results of the data reconciliation is shown in Figure 4.2, where the product flow data obtained from measurements and the values corrected by reconciliation are compared. From the analysis of the reconciliation results, one can realize the existence of inconsistencies in the data. In this case there are two facts that may cause inconsistencies in the data: on one hand the presence of a bypass valve after the location of the inlet product flowmeter that makes that the real flow could be lower than the measured, on the other hand the heat-source flowmeter becomes saturated beyond a value, so the real temperature could be higher than the measured.

[3] Note that the minus sign in (4.4) is to compensate for the negative value of ΔT, as $Tout_c$ is greater than Tin_c. Consequently, Q_c will also be negative.

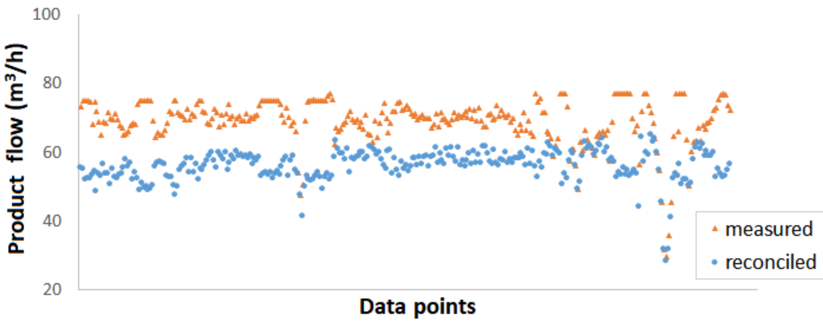


Figure 4.2. Comparison between the measured and corrected values for F_c

4.3.2 Regression models

Once reliable data have been obtained, the overall heat-transfer coefficient in each operating point (U_i) can be estimated through (4.1) which is rewritten as (4.6), where the heat transferred (Q') is the heat that the heat source gives (Q_s) computed with (4.5).

$$U_i = \frac{\mathrm{F_{c,i}}\, \rho_\mathrm{c}\, Cp_c\, (\mathrm{To_{c,i}} - \mathrm{Tin_{c,i}})}{A \cdot \mathrm{LMTD}} \tag{4.6}$$

$$\mathrm{LMTD} = \frac{(\mathrm{Tin_{s,i}} - \mathrm{Tin_{c,i}}) - (\mathrm{To_{s,i}} - \mathrm{To_{c,i}})}{\ln\left(\frac{\mathrm{Tin_{s,i}} - \mathrm{Tin_{c,i}}}{\mathrm{To_{s,i}} - \mathrm{To_{c,i}}}\right)} \tag{4.7}$$

Note that the temperatures and flows used to compute the heat-transfer coefficient are the obtained with the data reconciliation instead of the measured ones.

Now, with the *computed* values of U, the aim is to find an experimental relationship (black-box model) between the overall heat-transfer coefficient (U) and the flows and temperatures of both streams, $U = f(F, T)$. For this goal, $f: \mathbb{R} \to \mathbb{R}$ is constrained to be a polynomial in its arguments, such that SOS-constrained regression (see section 1.2.2) can be used both to fit the data and to enforce coherent physical model responses.

The best model obtained[4] is (4.8), a third-degree polynomial where the overall heat-transfer coefficient depends just on the flows, as the temperature influence has resulted to be negligible in our case study.

$$U = \gamma_0 + \gamma_1 F_s + \gamma_2 F_c + \gamma_3 F_s^2 + \gamma_4 F_s F_c + \gamma_5 F_s^3 \qquad (4.8)$$

The model parameters $\theta = \{\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5\}$ are independent of the operation conditions, but they depend on the heat-exchanger features. Hence, the exchangers have been grouped in three sets according to their sizes (small, medium and large heat-transfer areas) so that the procedure had to be repeated to get, at least, three different sets of values for the model parameters.

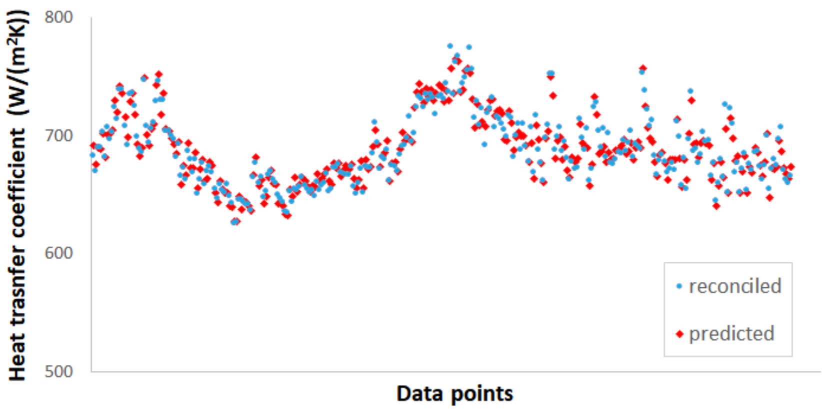


Figure 4.3. Goodness of fit for U: estimated values by data reconciliation (blue) with their corresponding model predictions (red).

In order to give an insight of the goodness of fit, Figure 4.3 shows some values of U estimated by data reconciliation and their corresponding ones predicted by (4.8), for a medium-size heat exchanger.

[4] Best according to the selected regularization coefficients to limit an excessive model complexity.

4.3.3 Fouling contribution

As mentioned before, the heat-transfer capacity of a heat exchanger is limited by the fouling. Therefore, as the regression models (4.8) have been fitted using data from clean heat exchangers, the actual U over time will always be lower than the predicted one. To tackle this issue, (4.8) has been corrected by adding an additional term, K_h, that accounts for the *state of fouling* of the heat exchangers. Indeed, the state of fouling K_h can be monitored online by just comparing the actual heat transfer with the predicted by the *clean model* (Lenzing, 2019). In this way, and assuming that the fouling state does not evolve significantly during a few hours, measurements in real time can be used to update the bias parameter K_h in (4.5) from day to day.

$$U = \gamma_0 - K_h + \gamma_1 F_s + \gamma_2 F_c + \gamma_3 F_s^2 + \gamma_4 F_s F_c + \gamma_5 F_s^3 \tag{4.9}$$

4.4 Optimal operation

Once the data-driven models for the heat-transfer coefficient have been obtained, they can be added to a hybrid model that represents the plant operation in mathematical programming terms. The core of the model is based not only on first principles (mass and energy balances), but also considering logic statements and the gray-box models for the heat transfer. As many production changes within a day are scarce, and the system dynamics is fast enough to be neglected, the model focuses on representing the steady state reliably.

As mentioned before, the model also includes other operational constraints that have to be taken into account. The objective function is to minimize the operation cost of the network given by the resource utilization (heat sources) and the decision variables are the distribution of the heat sources, that is, the amount of each heat source that goes through each heat exchanger able to fulfill the product temperature demand. All these components of the optimization problem will be described next.

4.4.1 Network model

The first aim of the optimization is to distribute such sources among exchangers accounting all the allocation alternatives, in order to minimize the cost of the heat sources consumption.

The following sets of entities and variables are previously defined to build a mathematical model of the network operation possibilities.

The set $\mathcal{W}$ is defined to contain all heat exchangers. From the point of view of heat-sources connectivity in each heat exchanger, $\mathcal{W}$ has already been divided into five subsets $\mathcal{W} = \{\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5\}$ as Table 4 shows. However, $\mathcal{W}$ can also be split in other two different subsets, $\mathcal{W} = \{\mathcal{V}_1 \cup \mathcal{V}_2\}$ where $\mathcal{V}_1$ includes the heat exchangers with the products connected in parallel and $\mathcal{V}_2$ lists those where the product flows in serial connection through them.

The set $\mathcal{S}$ gathers all heat sources in the network. Note that there is not a set defined for product streams, as they are fixed for each heat exchanger, and they are just input data for the optimization. In the model constraints, parameters and input data are written in plain style whereas decision variables are written in italics.

The sets of decision variables employed in the model are:

- $X_{s,w} \in \{0,1\}$: Binary variables that link the heat source s to the heat exchanger w, when active $(X_{s,w} = 1)$.
- $F_{s,w} \in \mathbb{R}^{+}$: Continuous variables that set the flow of heat-source s to heat exchanger w.

Once sets and variables are defined, we can write the constraints that describe the network operation. They are based on first principles and logic statements, in addition to the production goals.

1. *Demand:* the outlet product temperatures ($To_{c,w}$) have to reach certain setpoints ($\mathrm{SP_w}$). Note that, these constraints will affect only the heat

exchangers connected in parallel and the last heat exchanger in the chain of those connected in series:

$$To_{c,w} \geq \mathrm{SP}_w \qquad \forall\, w \in \mathcal{V}_1 \cup \mathrm{w15} \tag{4.10}$$

2. *Exclusivity:* since the heat sources cannot be mixed, the heat exchangers can only use a single source (if any) at a time:

$$\sum_{s}^{\mathcal{S}} X_{s,w} \leq 1 \qquad \forall\, w \in \mathcal{W} \tag{4.11}$$

3. *Flow bounds:* if a heat source is not connected to a heat exchanger, its flow is zero, otherwise it is bounded by an upper limit $\overline{\mathrm{F_w}}$ (different upper bound may be set up for each heat exchanger):

$$F_{s,w} \leq \overline{\mathrm{F_w}}\, X_s \qquad \forall\, w \in \mathcal{W} \tag{4.12}$$

4. *Non feasible connections:* not all heat exchangers can be physically connected to all heat sources (just the ones described in Table 4), so there must be a constraint in the model that blocks the non-existent links from sources to heat exchangers, denoted by the set $\mathcal{I}$ in (4.13).

$$X_{s,w} = 0 \qquad \forall s, w \in \mathcal{I} \tag{4.13}$$

5. *Energy balances:* analogous to the data reconciliation problem, the energy balances in each heat exchanger must be fulfilled, equations (4.1)-(4.5). Nevertheless, as now there may be more than one heat source able to feed an exchanger, these equations are expanded as follows:

$$\sum_{s}^{\mathcal{S}} Q_{s,w} = -Q_{c,w} \qquad \forall w \in \mathcal{W} \tag{4.14}$$

$$Q_{k,w} = F_{k,w}\, \rho_\mathrm{k}\, Cp_k\, \Delta T_{k,w} \qquad \forall k \in \{s, c\} \tag{4.15}$$

$$Q_{s,w} = U_{s,w}\, \mathrm{A_w}\, LMTD_{s,w} \quad \forall s \in \mathcal{S}, \forall w \in \mathcal{W} \tag{4.16}$$

Where the experimental expression for U obtained in Section 4.3.3, U_w, is recalled to compute the heat transferred in each heat exchanger. Moreover, in order to fulfill (4.14) when the source s is not used in the heat exchanger w (i.e., $Q_{s,w} = 0$), the term $\left(\gamma_{0,w} - K_{h_w}\right)$ is multiplied by $X_{s,w}$. Thus, in combination with (4.12) which forces the flow to be zero when $X_{s,w} = 0$, U_w and the heat transfer will be forced to be zero too.

$$U_{s,w} = (\gamma_{0_w} - K_{h_w})X_{s,w} + \gamma_{1_w} F_{s,w} + \gamma_{2_w} F_{c,w} + \gamma_{3_w} F_{s,w}^2 + \gamma_{4_w} F_{s,w} F_{c,w} + \gamma_{5_w} F_{s,w}^3 \quad \forall s \in \mathcal{S}, \forall w \in \mathcal{W} \tag{4.17}$$

6. *Mass balances:* the maximum flow available at the heat sources ($\mathrm{MF_s}$) cannot be exceeded. For the source S4, as it is connected in series, the flow in every exchanger has to be lower than the maximum available:

$$\begin{cases} \sum_{w}^{\mathcal{W}} F_{s,w} \leq \mathrm{MF_s} & \forall s \in \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}\} \\ F_{s,w} \leq \mathrm{MF_s} & \forall\, w \in \mathcal{W} \;\; \forall s = \mathrm{S4} \end{cases} \tag{4.18}$$

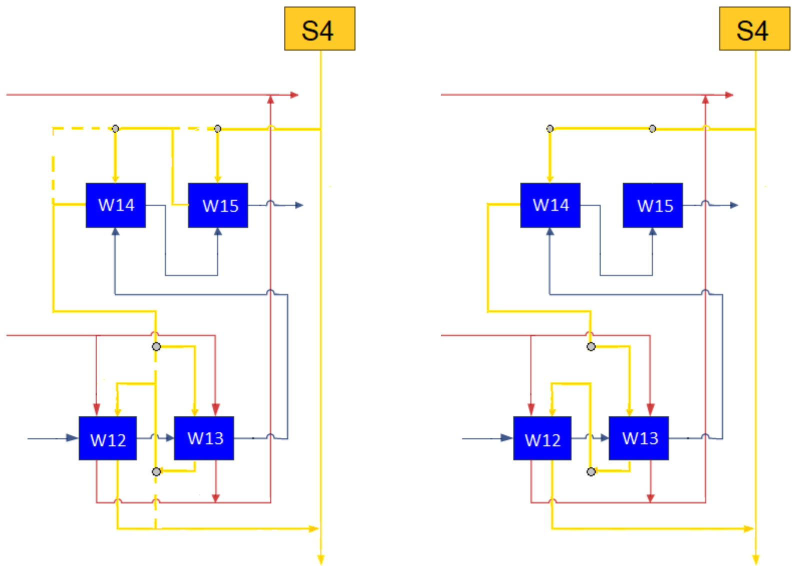
7. *S2-source considerations:* the total amount of heat source S2 (MF_{S2}) depends on the amount of *used* S1 in the heat exchangers $\mathcal{W}_1$ (see Figure 4.):

$$MF_{\mathrm{S2}} = \sum_{w}^{\mathcal{W}1} F_{\mathrm{S1},w} \tag{4.19}$$

Consequently, the inlet temperature of source S2 depends on the S1 temperatures at the outlet of exchangers $\mathcal{W}_1$:

$$T\mathrm{in}_{\mathrm{S2}} = \frac{\sum_{w}^{\mathcal{W}1} To_{\mathrm{S1},w}\, F_{\mathrm{S1},w}}{\sum_{w}^{\mathcal{W}1} F_{\mathrm{S1},w}} \tag{4.20}$$

In addition to this constraint, other two must be added, that allude to the series connection. The different connection possibilities of the heat source S4 in $\mathcal{V}_2$ are modeled by the following constraints, taking into account that the bypass valves can switch between feeding the exchanger or not, see Figure 4.4.a).



(a) **(b)**

*Figure 4.4. Detail of the heat exchangers connected in series. Grey dots represent the location of the bypass valves. (**a**) Shows the possible connections. (**b**) Presents the layout when w12, w13 and w14 are connected but w15 is not.*

Firstly, the S4 flow must be the same in all connected heat exchangers. To state this constraint, we formulate the system of equations shown in (4.21), where M is a big enough value (for instance M can be set to three times MF_{S4}).

$$\sum_{w \neq \beta}^{\mathcal{V}_2} F_{S4,w} \leq F_{S4,\beta} \sum_{w \neq \beta}^{\mathcal{V}_2} X_{S4,w} + M\left(1 - X_{S4,\beta}\right) \quad \forall \beta \in \mathcal{V}_2 \tag{4.21}$$

With (4.21) the model has the possibility of switching off any of the exchangers from w12 to w15, ensuring that the ones in operation get the right heat flow in series through them.

To clarify this last constraint, consider for instance that heat exchangers w12, w13 and w14 are connected but w15 is not (see Figure 4.4.b). Then, by developing (12):

- for β = w12: $$F_{S4,w13} + F_{S4,w14} \leq F_{S4,w12}\, 2 + M \tag{4.22}$$
- for β = w13: $$F_{S4,w12} + F_{S4,w14} \leq F_{S4,w13}\, 2 + M \tag{4.23}$$

- for β = w14: $$F_{S4,w12} + F_{S4,w13} \leq F_{S4,w14}\, 2 + M \tag{4.24}$$

- for β = w15: $$F_{S4,w12} + F_{S4,w13} + F_{S4,w14} \leq M \tag{4.25}$$

Thus, leading to the only possible solution of:

$$F_{S4,w13} = F_{S4,w14} = F_{S4,w12} \tag{4.26}$$

Secondly, the inlet temperatures for both product and heat source depend on the outlet of the previous heat exchanger in the stream direction. For the product stream, as it goes through all the exchangers, the inlet temperatures can be computed directly by (4.27). If a heat exchanger of the chain is not working, the heat gained by the product stream will be zero, and so is $\Delta T_{p,w}$.

$$Tin_{c,w+1} = To_{c,w} \qquad \forall w \in \mathcal{V}_2 \tag{4.27}$$

For the hot-stream side, if a heat exchanger is not connected, the heat source through it will be zero, so $\Delta T_{s,w}$ can take any value in the formulation. However. the inlet temperature will depend on the outlet of the *previous but connected* heat exchanger:

$$\begin{aligned} Tin_{S4,w} = \sum_{j=w15}^{w-1} \left(To_{S4,j} X_{S4,j} \prod_{j=w15}^{w-1} \left(1 - X_{S4,j+1}\right) \right) \\ + Tin_{S4,w15} \prod_{j=w15}^{w-1} \left(1 - X_{S4,j}\right) \forall w \in \mathcal{V}_2 \end{aligned} \tag{4.28}$$

An illustrative example of this constraint is presented below, where it is developed for the w13 inlet temperature, which depends on w14 using S4 as heat source. If not, it depends on w15 working with S4. Otherwise, the temperature will be directly the one at the source S4.

$$\begin{aligned} Tin_{S4,w13} = To_{S4,w14} X_{S4.w14} + To_{S4.w15} X_{S4.w15}\left(1 - X_{S4,w14}\right) \\ + Tin_{S4,w15}\left(1 - X_{S4,w15}\right)\left(1 - X_{S4,w14}\right) \end{aligned} \tag{4.29}$$

In this mathematical representation of the heat recovery network, the scalars A, ρ, Cp, and $\overline{\mathrm{F_w}}$, as well as the coefficients a_i are known fixed values.

The values of the parameters $\mathrm{F}_{\mathrm{p,w}}, \mathrm{SP}_{\mathrm{w}}, \mathrm{Tin}_{\mathrm{s}}$ and MF_{s} change over time, but they can be read from the information system of the factory in real time.

4.4.2 Optimization problem

The optimization aim is to minimize the normalized cost of operation per time unit. This cost comes from the consumption of each source (i.e., the total used flow) times its price[5] (P_s). Note that, for the sources connected in parallel, the total flow is the sum of the flows used in each heat exchanger, meanwhile for S4 the consumption is just the flow that goes through one of the connected exchangers.

$$Cost_{operation} = \sum_{s}^{\mathcal{S}\backslash S4} \sum_{w}^{\mathcal{W}} F_{s,w}\, \mathrm{P}_s + \max\left(F_{S4,w}\right) \mathrm{P}_{S4} \tag{4.30}$$

Hence, the optimization problem is to minimize (4.30) subject to (4.10)-(4.21), (4.27) and (4.28). Note that this is a MINLP problem, as there are discrete decisions ($X_{s,w}$) and $LMTD(\cdot), U(\cdot)$, are nonlinear functions. Once the problem is formulated, it has been coded in Pyomo-Python and solved using the NLP-based branch and-bound algorithm BONMIN (Bonami et al., 2008).

The problem size is 195 decision variables (40 binaries) with 421 constraints, and a solution with zero relative gap is proven in about 30 seconds over an Intel® 7-7700 CPU machine with 32Gb of DDR4 RAM memory. The optimization has been tested offline with plant historical data and some results are shown in Table 5.

[5] Note that S2 gets zero cost, as it is reused from S1.

Table 5. Example of the result for the optimal heat-sources distribution

	Heat exchanger	F_{S1} [m³/h]	F_{S2} [m³/h]	F_{S3} [m³/h]	F_{S4} [m³/h]
$\mathcal{B}_1$	w1	19.94	0.00	0.00	0.00
	w2	33.85	0.00	0.00	0.00
	w3	49.62	0.00	0.00	0.00
	w4	54.43	0.00	0.00	0.00
	w5	61.97	0.00	0.00	0.00
$\mathcal{B}_2$	w6	0.00	0.00	32.27	0.00
	w7	0.00	0.00	32.69	0.00
$\mathcal{B}_3$	w8	0.00	165.02	0.00	0.00
	w9	0.00	0.00	50.95	0.00
	w10	0.00	0.00	25.11	0.00
	w11	0.00	38.76	0.00	0.00
$\mathcal{B}_4$	w12	0.00	0.00	0.00	17.76
	w13	0.00	0.00	0.00	0.00
$\mathcal{B}_5$	w14	0.00	0.00	0.00	0.00
	w15	0.00	0.00	0.00	0.00

In this case, as there is enough availability in the hot sources, the optimization decides to connect each heat exchanger to the cheaper source within its allowed links. We can observe that, for $\mathcal{B}_3$ (w8 to w11), the RTO tries to connect as many exchangers as possible to S2, as this source is cost free. For the heat exchangers connected in series (w12 to w15), the RTO shows that the product temperature setpoint can be reached with just one heat exchanger connected.

Finally, note that, as model (4.9) can predict the efficiency of a heat exchanger according to its actual fouling state, it is possible to compare the current heat transfer with the one given by the heat exchanger if fully clean, in order to suggest if it should be cleaned. To add such suggestions into the optimization problem, the objective function must also take into account the cost of the cleaning tasks, that must be normalized to be comparable with the operation cost.

Therefore, the resulting optimization problem not only gives the optimal allocation of the heat sources according to the real-time conditions but also suggests the heat exchangers that should be cleaned, providing the best global economic trade-off.

4.4.3 Suggestions on cleaning tasks

The problem described above takes into account the performance degradation due to fouling, as (4.17) includes the parameter Kh_w which accounts for the state of fouling. In addition, by monitoring the state of fouling, we can not only decide over the hot-sources distribution, but also suggest which heat exchanger should be cleaned according to an economic trade-off criterion.

To do that, the previous formulation is extended with a new set of binary variables:

- $Y_w \in \{0,1\}$: Binary variables that activate ($Y_w = 1$) or deactivate ($Y_w = 0$) the cleaning in heat exchanger w.

Thus, the fouling factor Kh_w can be multiplied by $(1 - Y_w)$ in order to provide the model with the ability to compare between operating exchanger w with the actual efficiency or with the nominal one, as if it would be freshly clean. Therefore, (4.17) is extended to:

$$U_{s,w} = \left(\gamma_{0_w} - K_{h_w}(1 - Y_w)\right) X_{s,w} + \gamma_{1_w} F_{s,w} + \gamma_{2_w} F_{c,w} + \gamma_{3_w} F_{s,w}^2 + \gamma_{4_w} F_{s,w} F_{c,w} + \gamma_{5_w} F_{s,w}^3 \quad \forall s \in \mathcal{S}, \forall w \in \mathcal{W} \tag{4.31}$$

Moreover, the objective function must also consider the cost of the potential cleaning tasks to perform. However, the cleaning tasks have a fixed cost, P_C, which must be normalized to be comparable with the instantaneous operation cost of the network. Hence, P_C must be amortized over the operation time of the heat exchanger since its last cleaning, t_w. Thus, the normalized cost of the cleaning tasks will be:

$$Cost_{clean} := \sum_{w}^{W} \frac{Y_w \ P_c}{t_w} \tag{4.32}$$

In this way, the economic objective function defines an instantaneous trade-off between operation and cleaning, where the cleaning cost is depreciated over time whilst the cost of operation progressively increases due to fouling (more heat flow is needed to reach the product temperature setpoints).

Note that, if the optimization sets $Y_w = 1$ for some heat exchanger w, it is considered that such exchanger performs as it was fully clean by (4.31) and its corresponding costs of cleaning are included in the objective function by (4.32). This way, the optimization problem provides suggestions for cleaning in real time, which aim to provide the best global economic trade-off.

In summary, the optimization problem to be solved in real time is:

$$\begin{aligned} \min_{F_{s,w}, X_{s,w}, Y_w, F_{S4}} \quad & Cost_{operation} + Cost_{clean} + \tau \sum_{s}^{S} \sum_{w}^{W} X_{s,w} \\ \text{s.t.:} \quad & \text{model constraints (4.10)-(4.16), (4.18)-(4.21),} \\ & \text{(4.27), (4.28). (4.31)} \\ & F_{s,w}, F_{S4} \in \mathbb{R}^+ \\ & X_{s,w}, Y_w \in \{0,1\} \end{aligned} \tag{4.33}$$

Where an additional term in the objective function ($\tau > 0$ user-defined weight) is included to penalize the connection of serial heat exchangers if it is not strictly necessary to reach the product setpoints. In this way, α serves as a tuning parameter to control the *nervousness* (Dalle Ave et al., 2019) of the RTO solutions.

This optimization involves 211 decision variables (75 binaries) and 425 constraints. It is solved with the NLP-based branch-and-bound algorithm BONMIN (Bonami et al., 2008) elapsing about 120 seconds average in an Intel® 7-7700 CPU workstation with 32Gb of DDR4 RAM memory to get a solution with zero relative gap. Therefore, as this RTO runs in an hourly basis (sensible frequency to account for realistic changes in setpoints or available source flows), the computational time does not represent a major issue.

The problem was tested with the same data as the used in the previous section, in order to be able to compare the solution with the suggestions of cleanings. Unfortunately, since the fouling state was not monitored in the past, we do not have values for K_{h_w} recorded in the plant historian, but, based on preliminary data monitoring of one exchanger, we found that a good enough initial approximation is to set K_{h_w} as the double of the operation time since last cleaning, $K_{h_w} = 2t_w$. Of course, this rough approximation is just for evaluation purposes, and it will be replaced by actual data when the monitoring system will be fully implemented.

Comparing the results of Table 6 with the ones showed in Table 5, we can observe that the heat-source distribution is the same. However, by cleaning heat exchangers w2 and w7, the needed flow of the heat sources to reach the product temperature setpoint is lower (4.3% and 6.2% lower respectively). Although the cleaning tasks involve a cost, the reduction of the utility consumption is beneficial from an economic point of view. Indeed, the total cost will decrease about 0.5% if the cleaning tasks are performed.

Such results also show that the best policy not always is to clean the dirtiest (w8, w13). Instead, the suggestion is to clean w2 and w7. This can be explained taking into account that the source used in w8 is cost free and w13 is not being used, so no operation cost is incurred for these two.

Table 6. Results considering the suggestions of cleanings.

	W	F_{S1} [m³/h]	F_{S2} [m³/h]	F_{S3} [m³/h]	F_{S4} [m³/h]	Cleaning suggestion	K_h [W/(m²K)]
$\mathcal{B}_1$	*w1*	19.94	0.00	0.00	0.00	0	20
	w2	32.40	0.00	0.00	0.00	1	30
	w3	49.62	0.00	0.00	0.00	0	10
	w4	54.43	0.00	0.00	0.00	0	0
	w5	61.97	0.00	0.00	0.00	0	6
$\mathcal{B}_2$	*w6*	0.00	0.00	32.27	0.00	0	6
	w7	0.00	0.00	30.68	0.00	1	34
$\mathcal{B}_3$	*w8*	0.00	167.82	0.00	0.00	0	44
	w9	0.00	0.00	50.95	0.00	0	12
	w10	0.00	0.00	25.11	0.00	0	24
	w11	0.00	38.94	0.00	0.00	0	2
$\mathcal{B}_4$	*w12*	0.00	0.00	0.00	17.76	0	18
	w13	0.00	0.00	0.00	0.00	0	36
$\mathcal{B}_5$	*w14*	0.00	0.00	0.00	0.00	0	22
	w15	0.00	0.00	0.00	0.00	0	24

To test the consistency of the solutions, i.e., to evaluate the so called nervousness level, the optimization was rolled out offline for several consecutive days, updating the input data with the actual plant state in each run: each day the fouling state is updated according to the operation and the inlet stream temperatures vary within a maximum of 3°C with respect to the actual measurements (i.e., to make the test reasonably demanding, it is assumed that the inlet temperature will be modified by a random number between ±3°C). In addition, we assume that the cleaning tasks suggested are performed on the day, thus resetting the operation time and the fouling state for the next execution.

Remark that this is not a maintenance-prediction scheduling, which is out of the scope of this work, as there is not a single optimization for a set of future days (time horizon), but several RTO runs done sequentially, one per day. The evolution for fifteen consecutive days is shown in Figure 4.5.

	Day/ W	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
B1	w1	41,26	41,22	40,88	40,91	39,36	39,30	19,13	19,13	19,20	19,28	41,07	19,44	31,11	19,60	40,47
	w2	37,30	36,88	35,57	35,31	38,96	39,00	32,49	32,59	32,68	32,77	39,63	32,96	33,06	33,16	37,20
	w3	49,62	49,76	49,91	50,06	50,21	50,35	50,50	48,91	48,91	49,05	49,19	49,33	49,48	49,62	49,76
	w4	54,43	54,47	54,52	54,56	54,61	54,65	54,70	54,74	54,79	54,84	54,88	54,93	54,98	55,03	55,08
	w5	61,93	62,03	62,10	62,17	62,24	62,31	62,38	62,45	62,52	62,59	62,66	61,77	61,77	61,84	61,90
B2	w6	32,27	32,33	32,39	32,45	32,51	32,58	32,64	32,70	32,76	32,83	32,10	32,10	32,15	32,21	32,27
	w7	30,68	30,68	30,79	30,90	31,02	31,13	31,25	31,36	31,48	31,60	31,72	30,68	30,68	30,79	30,90
B3	w8	130,6	130,4	131,7	131,8	133,3	133,4	162,8	164,4	34,62	34,71	134,5	34,90	144,4	159,4	131,9
	w9	50,95	51,10	51,24	50,13	50,13	50,26	50,40	50,54	50,67	50,81	50,95	51,10	51,24	50,13	50,13
	w10	79,06	79,08	76,33	76,16	76,87	76,93	24,52	24,59	102,8	102,5	77,41	102,6	86,03	24,98	76,47
	w11	34,84	34,86	35,01	35,05	35,23	35,27	39,87	39,96	44,01	44,24	35,52	44,31	16,94	40,43	36,00
B4	w12	21,86	17,82				17,84	12,99			21,94		21,97	22,00	22,02	22,05
	w13				21,85			16,32	17,85	21,93		17,89				
B5	w14															
	w15			17,76		17,78										

S1 S2 S3 S4

Figure 4.5 Overview of the suggestions provided by the RTO during fifteen consecutive days. Cell color indicates the heat source linked to the exchanger whilst color intensity points the fouling state: the higher the intensity, the dirtier the exchanger. A sudden reduction in the intensity indicates a cleaning operation was performed.

Note that the heat-sources allocation changes over time, clearly seen in blocks $\mathcal{B}_3$ and $\mathcal{B}_4$ where the heat source used in each heat exchanger needs to adapt to the operation conditions of the day. The most significant changes related to the solution nervousness are the ones observed in the heat exchangers connected in series, where the solution shows that not only the heat source but also the heat exchanger used changes. This is because the product heat demand in these exchangers is not high (usually due to a low input flow) in comparison with the total heating capacity of these four exchangers, so the optimization tries to switch between the less fouled ones in order to save a bit of the flow of source usage. Nonetheless, this behavior, if unacceptable, can be easily attenuated by increasing the value of the tuning parameter α in the objective function of problem (4.33).

4.5 Plant-wide optimization

The heat-recovery network is related to the evaporation network whose optimal operation has been discussed in Chapter 3, as the products that must be heated up in the heat-exchanger network are the concentrated products obtained after the evaporation process (see Figure 4.6). Hence, the input temperatures of the products at the heat-recovery network depend on the performance of the evaporation network (including the cooling system).

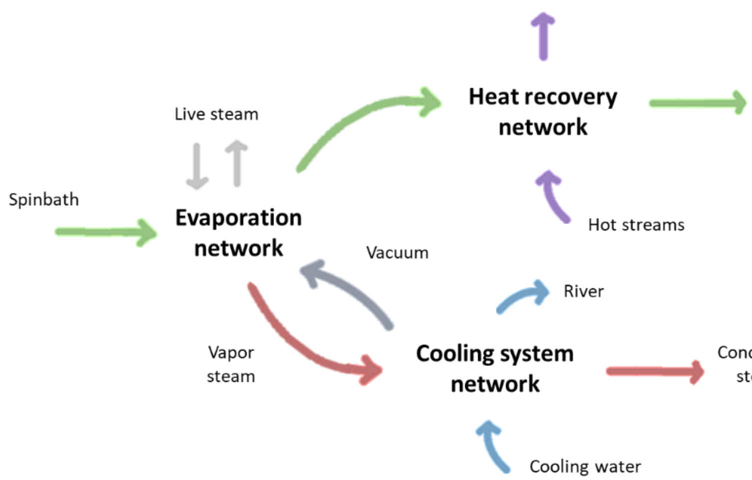


Figure 4.6. Relation between the different networks

Thus, one might think that a plant-wide optimization of the three networks together (the evaporation, the cooling system, and the heat-recovery one) would lead to the optimal operation of the whole process. Nevertheless, it is not necessary as one may conclude by doing an engineering analysis of the behavior of the different systems.

In order to increase the temperature of the concentrated product (to reduce the use of heat sources in the heat-recovery system), it is needed to increase the live steam consumption. Note that the live steam is used in the evaporation plant to heat up the product (see Figure 4.7). However, the price of the live steam is much bigger than the price of the heat sources (about 20 times bigger), so it would be self-defeating.

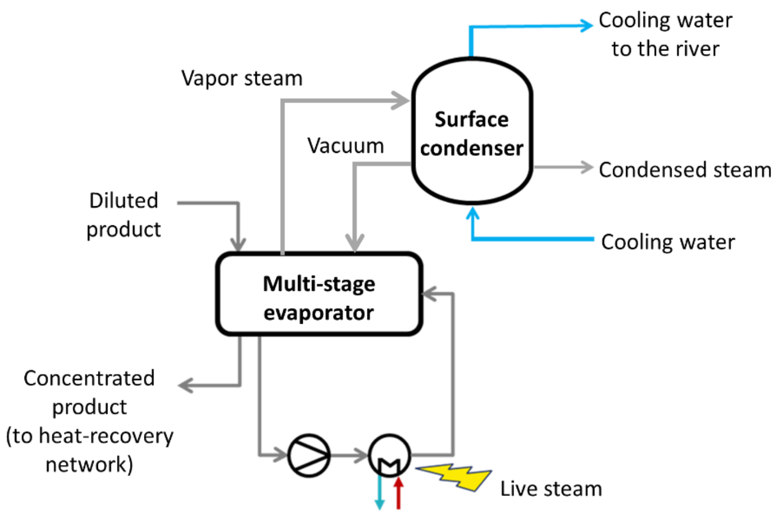


Figure 4.7. Scheme of an evaporation plant

The live-steam consumption is the key utility from an economic point of view and it does not depend on the heat-recovery network. Hence, the global optimal operation can be achieved by performing the local optimizations presented in this chapter: running the evaporation network (including the cooling system) optimization, and, with the resulting conditions, running the heat-recovery optimization subsequently.

4.6 Summary and conclusions

In this chapter the problem of optimal operation of a heat-recovery network has been addressed considering the efficiency of the heat exchangers and the use of shared resources simultaneously. The key operational decisions are how to allocate and distribute the heat sources (utilities) among the heat exchangers that compose the network.

First, the heat transfer through the heat exchangers has been studied and a gray-box model that relates the overall heat transfer coefficient with the operation conditions has been developed. Such model has been obtained based on experimental data collected under suitable plant conditions. Nevertheless, to guarantee the reliability of the data points, a data reconciliation has been done.

Such gray-box models have been incorporated into a rigorous mathematical model of the network, which is the core of an MINLP optimization accounting for the current production constraints. Besides, as the model also takes into account the equipment fouling states, the optimization problem has been extended to provide suggestions on cleaning tasks. In this way, the proposed RTO will support the plant operators in such a complex decision-making process in real time, saving resources and reducing the personnel workload.

The preliminary evaluation showed that the decisions proposed by the RTO outperformed those taken by the human operator. A relevant outcome from this evaluation is that the current cleaning policy of *dirtiest first* has been proven not to be the best in all cases. Indeed, knowing which heat exchanger should be cleaned for optimal economic performance is not an easy task, as it depends a lot on different interconnected factors. Hence, the proposed RTO not only improves the heat-source distribution, but also helps to configure the cleaning schedule.

Nevertheless, this is an instantaneous RTO which does not predict the future evolution of the fouling, so the proposed maintenance actions may be suboptimal in the long term. In return for this, the proposed RTO solution does not suffer from the always undesirable turnpike effect (Carlson & Haurie, 2013) typically arising in dynamic optimization and scheduling solutions over a finite future time horizon.

Finally, the operation of the heat-recovery section jointly with the evaporation network has been discussed. The inference obtained from the analysis is that the evaporation process is the resource-intensive process, as the live-steam consumption is the most expensive utility. Thus, as the heat-recovery section does not directly affect the evaporation process, the procedure for an optimal operation is first, performing the optimization of the evaporation network based on the actual conditions of the plant and, after that, performing the heat-recovery optimization with the resulting conditions.

The work presented in this chapter led to three scientific papers, one in a national congress (M. P. Marcos, Pitarch, & Prada, 2019), one in an international conference (M. P. Marcos et al., 2020a), and one published in a Q1 JCR-indexed journal (M. P. Marcos et al., 2020b).

Chapter 5

Integrated process re-design

This chapter goes beyond joint operation of the interconnected evaporation and cooling-system networks by incorporating process design decisions as well. In this regard, the previous mathematical formulation to solve the real-time optimization problem described in Chapter 3 is twofold extended: a) to include the possibility of attaching heat pumps to the cooling system in order to improve the overall resource efficiency; and b) incorporating uncertainty in future operation conditions.

First, the problem identification for the re-design of the cooling system in the evaporation network is presented in Section 5.1, where all the configuration possibilities of heat-pump incorporation are described. It is followed by the main objectives of the chapter in Section 5.2. After that, in Section 5.3 the mathematical formulation of the optimization problem for the re-design of the network is described. Then, the integration of uncertainty in the formulation and the analysis of the results are presented in Section 5.4. The daily optimal operation of such network through an RTO tool is shown in Section 5.5. Finally, in Section 5.6, the most important aspects achieved in this chapter are summarized.

5.1 Problem identification

As we have concluded after the results obtained in Chapter 3, the surface condenser must be working at maximum cooling power (i.e., using the maximum cooling water available) to reduce the steam consumption in the evaporation process. In addition, as the cooling water after the surface condensers goes back to the river, the distribution of it among the surface condensers must be done in a way that its temperature does not exceed the maximum allowed by environmental legislation. A solution to increase the efficiency of the network under such constraint is placing a group of heat pumps that take part of the outlet cooling water from the surface condensers to cool it down. Hence, it could be reused again at the surface condensers inlet as a way of recovering energy and limiting the temperature of the water flowing to the river.

Recall that a heat pump is a thermal machine that extracts heat from a fluid and transfers it to another (see Figure 5.1). In this case, they would remove heat from the desired streams, and such heat might be used for other purposes. Nevertheless, in this thesis, possible uses for such heat are left out of the scope, even though they would bring extra savings.

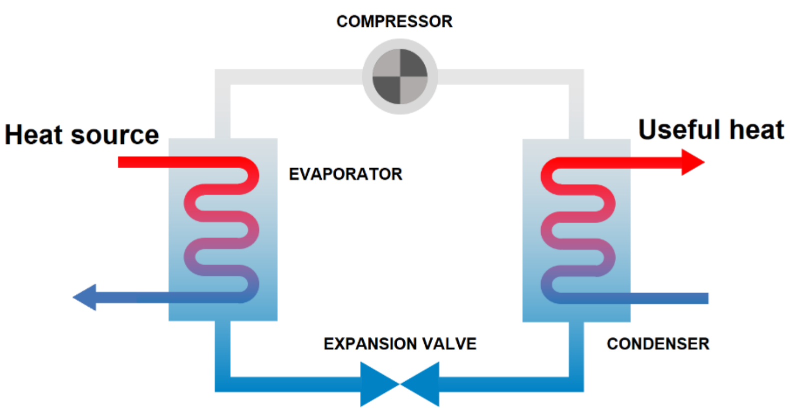


Figure 5.1. Scheme of a heat pump

Thereupon, the task here is to investigate which is the best way to incorporate the heat pumps in the cooling network such that the cooling power is improved (hence, operability margins in the evaporation section will improve too) while fulfilling environmental regulations, but without incurring in an excessive time lag between investment and payback[6].

Specifically, there are three main questions to answer:

- How many heat pumps are needed to be installed?
- Which is the best installation layout? I.e., connections among them.
- How should the heat pumps be incorporated into the actual network layout? In this regard, connection among heat pumps can be:
    - In series: the input flow of a heat pump is the output of the previous one (see Figure 5.2.a).
    - In parallel: all heat pumps receive an inlet from the same stream (see Figure 5.2.b).
    - Hybrid: a combination of the above as a matrix of heat pumps, where some heat pumps are arranged in series forming blocks that are connected in parallel (see Figure 5.2.c).

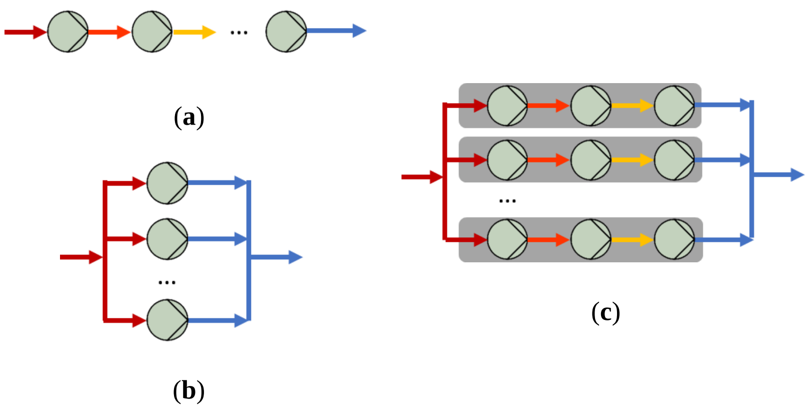


*Figure 5.2. Schematic sketch of the three ways to connect the heat pumps. (**a**) shows a series configuration. (**b**) presents parallel structure. (**c**) reflects the hybrid configuration.*

[6] Note that the payback in this context is the period of time elapsed before the investment is recouped without considering the useful life of the heat pumps (supposed to be long enough).

With respect to the outlet flow from the heat-pump system, there are two main configurations:

- To be mixed with the remaining outlet flow from the SCs (amount that is not directed to the heat pumps), in order to decrease the temperature of the water flow that goes to the river (see Figure 5.3.a).
- To be mixed with the inlet flow of the surface condensers, i.e., as if it was an additional water source (see Figure 5.3.b). However, note that this option is only possible for the evaporation plants in Subnet 2, because circulation of water from Source 2 to Subnet 1 is forbidden.

Nevertheless, both options can be happening simultaneously, if a flow-control valve is placed at the output of the heat-pump superstructure to divide the flow. Thus, part of the flow can be used to cool down the stream that goes to the river, and the rest is used as an additional source (see Figure 5.3.c).

In summary, there are quite a few decisions to take from the heat-pump integration design side, added to those related to the operation of the existing evaporation network (Chapter 3). Furthermore, these decisions must not be taken separately, but integrated considering the future conditions at which the re-designed network will operate.

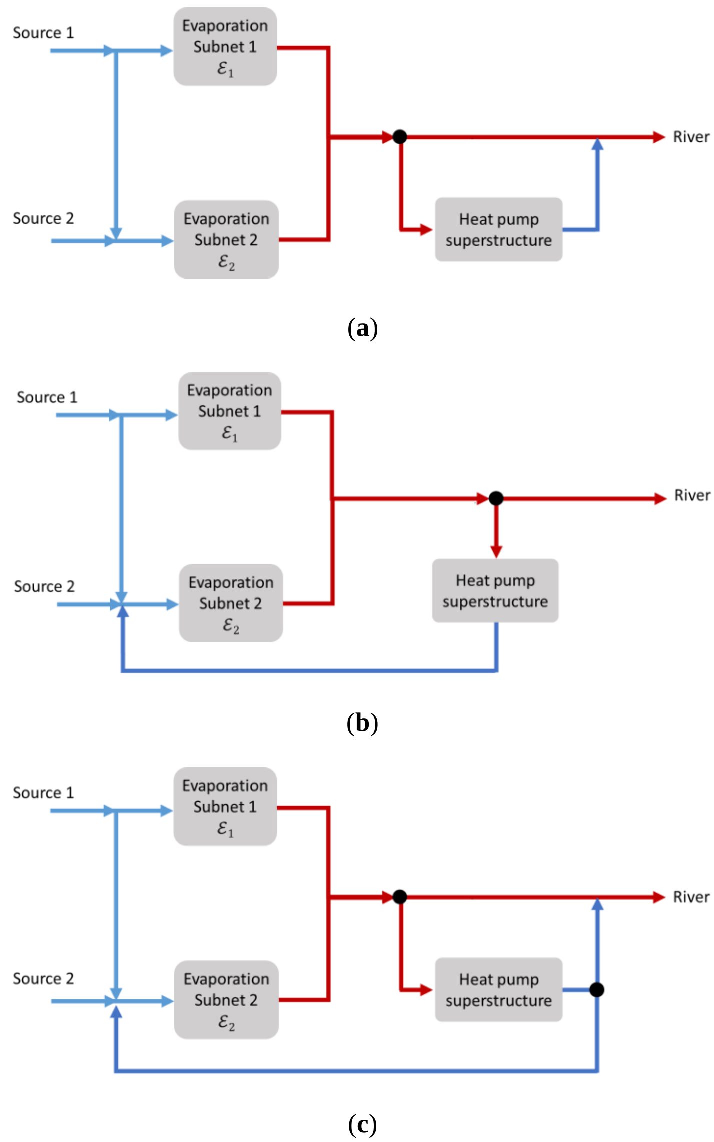


*Figure 5.3. Possible configuration of the cooling system with heat-pump incorporation. (**a**) Shows the layout when the heat pumps are used to cool down the temperature flow that goes to the river. (**b**) Shows the option of recirculating the output flow of the heat pumps as a new source to the cooling system. (**c**) Combines (a) and (b) options if the output flow can be split.*

5.2 Objectives

The general aim is incorporating a superstructure of heat pumps to model all alternatives in order to improve the evaporation process efficiency. For such aim, the different possibilities described in the previous section are going to be mathematically modeled into what is known as a *superstructure*. Such hybrid model must also consider the physical laws of the process and the operation constraints. It is going to be the core of an optimization problem whose objective function considers the operation cost of the network, computing in addition the optimal number of heat pumps to use and their connection layout.

Such optimization problem must include different scenarios according to the expected future operation conditions. Therefore, the problem must also consider the uncertainty of some initial conditions. However, the complexity of the resulting optimization problem could turn out to be excessive to have a solution in acceptable time, especially if the designer would like to evaluate different payback horizons. Hence, a balance between complexity and easiness of solution must be taken into account.

Finally, once the optimal number of heat pumps to purchase and their configuration have been computed, the RTO which seeks for the best operation conditions must be updated to integrate the heat pumps.

In summary, we aim:

- To model the network including the superstructure of the heat-pump integration.
- To formulate an optimization problem which seeks for the best operation conditions and the optimal heat-pump integration.
- To incorporate the uncertainty of some initial conditions.
- To update the RTO with the heat-pump integration and test it.

5.3 Optimization problem for re-design

The heat pump add-on is to be designed in such a way that optimizes the operation of the complete system in the more usual situations along the year. In this regard, although there are detailed models and performance studies on heat pumps available in the research literature (Chua et al., 2010), only nominal features are often considered at the industrial design level. Then, it is assumed that the heat pumps are arranged in such a way that they operate at (or close to) design conditions. Therefore, with the aim of not adding unnecessary complexity to the mathematical formulation, in this chapter its dynamics are neglected, and a heat pump is simply characterized by:

- Its *capacity*, i.e., the heat flow that is able to transfer from one fluid to the other (Q) in nominal conditions.
- Its *energy efficiency ratio* (EER), that is the ratio between the capacity and the electrical power consumption required to achieve it.
- The *nominal flow* FB, for which the heat pump achieves its capacity in steady state.

Hence, in this case study, the heat-pump model is based on its energy balance (5.1), where Cp and ρ are the specific heat and density of the cooling water through the heat pump respectively, and ΔT is the temperature difference between the inlet and outlet on the heat pump. Note that Cp and ρ are the same as the ones used in Chapter 3 referring to the cooling water.

$$\mathrm{Q} = \mathrm{FB}\ \rho_{\mathrm{c}}\ \mathrm{Cp}_{\mathrm{c}}\ \Delta T_{hp} \tag{5.1}$$

5.3.1 Network hybrid model

The formulation of the mathematical model which represents the superstructure of the network operation is based on the centralized formulation described in Chapter 3 (Section 3.5.2). Here, it is adapted in order to incorporate the decisions about heat-pump integration, according to the layout showed in Figure 5.4.

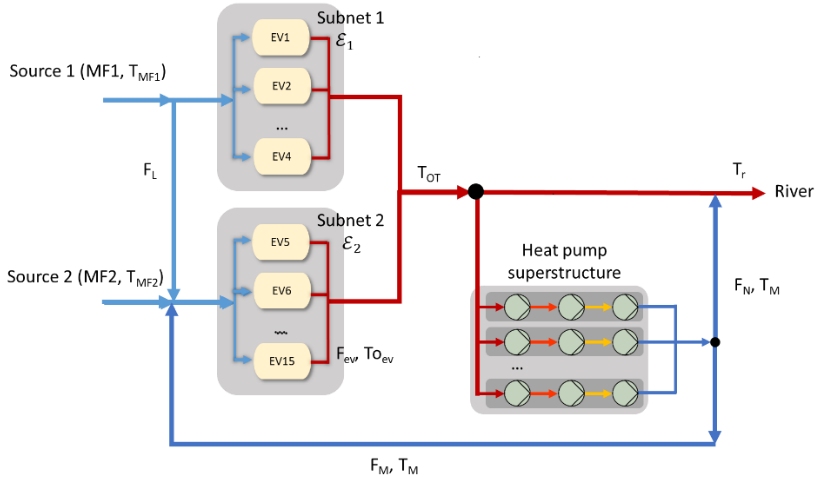


Figure 5.4. General layout of the cooling system with heat pump integration

Therefore, the different sets now are:

- $ev \in \mathcal{E}$: set of all the evaporation plants. This set is divided into two subsets $\mathcal{E} = \{\mathcal{E}_1 \cup \mathcal{E}_2\}$ according to the existing surface condenser subnet arrangement, i.e., those that have access to cooling-water Source 2 or not.
- $p \in \mathcal{P}$: set of the different products.
- $b \in \mathcal{B}$: set of the blocks of heat pumps, to be connected in parallel.
- $hp \in \mathcal{H}$: set of the heat pumps connected in series within each block, considering that all blocks have the same length.

The variables that relate the above introduced sets are now defined:

- $EC_{ev,p} \in \mathbb{R}^+$: evaporation load of product p that plant ev must achieve.
- $X_{ev,p} \in \{0,1\}$: indicates if product p is treated in plant ev.
- $F_{ev} \in \mathbb{R}^+$: cooling-water flow that goes to plant ev.
- $F_L \in \mathbb{R}^+$: remaining cooling water from source 1 that goes to subnet 2.
- $Y_{b,hp} \in \{0,1\}$: indicates if a heat pump hp in block b is in operation ($Y_{b,hp} = 1$) or not ($Y_{b,hp} = 0$)

- $F_M \in \mathbb{R}^+$: recirculated flow to the surface condensers after the heat pumps.
- $F_N \in \mathbb{R}^+$: flow that goes to the river after the heat pumps.

The constraints of the model that represent the operation of the evaporation network in steady state are based on first principles and some logic statements, detailed below. As well as in the previous chapters, in this formulation, italic symbols represent variables and sets whilst plain ones are known values (i.e., inputs or parameters) for the optimization.

1. An evaporation plant only can process one product at a time:

$$\sum_{p}^{\mathcal{P}} X_{ev,p} \leq 1 \qquad \forall ev \in \mathcal{E} \tag{5.2}$$

2. Some connections between plant and products are not allowed (see Table 1):

$$X_{ev,p} = 0 \qquad (ev, p) \in \mathcal{M} \tag{5.3}$$

3. The connection in series of the heat pumps imply that, in order to use the next heat pump of the chain, the previous one must be connected[7]:

$$Y_{b,hp+1} \leq Y_{b,hp} \qquad \forall b \in \mathcal{B}, \forall hp \in \mathcal{H} \backslash \{hp_{|\mathcal{H}|}\} \tag{5.4}$$

4. In the same way, to avoid different solutions with the same number of heat pumps it is forced to connect a heat pump of a block only if the heat pump of the previous block in the same position in the chain is connected:

$$Y_{b+1,hp} \leq Y_{b,hp} \qquad \forall b \in \mathcal{B} \backslash \{b_{|\mathcal{B}|}\}, \forall hp \in \mathcal{H} \tag{5.5}$$

5. The outlet flow from the heat-pump blocks can be recirculated to the SCs inlet or delivered to the river stream:

[7] Having an ordered set $x \in \mathcal{X} \coloneqq \{1, 2, \dots, n\}$, notation $x_{|x|}$ refers to the last element in such set, i.e. $x_{|x|} = n$ in this example.

$$F_M + F_N = \mathrm{FB} \sum_{b}^{\mathcal{B}} Y_{b,1} \qquad \forall b \in \mathcal{B} \tag{5.6}$$

In (5.6), FB is the flow that goes through each block of heat pumps, whose value is the same for all the heat pumps (all the heat pumps to be purchased will be of same characteristics) and it is set by the nominal specifications of the heat pumps. As it is a known parameter, to ensure that there is flow passing through the block, it is multiplied by the sum of $Y_{b,1}$. Note that, because (5.5), if any heat pump of a block is connected the first heat pump such block must be connected too.

6. The evaporation demand for each product must be fulfilled:

$$\sum_{ev}^{\mathcal{E}} EC_{ev,p} \geq \mathrm{SP}_p \qquad \forall p \in \mathcal{P} \tag{5.7}$$

Where SP_p denotes the evaporation setpoint for product p required by the factory to fulfill the demand.

7. The evaporation load in each plant is bounded due to actual equipment features:

$$\underline{\mathrm{EC}_{ev}} X_{ev,p} \leq EC_{ev,p} \leq \overline{\mathrm{EC}_{ev}} X_{ev,p} \qquad \forall ev \in \mathcal{E}, \forall p \in \mathcal{P} \tag{5.8}$$

Where $\underline{\mathrm{EC}_{ev}}$ and $\overline{\mathrm{EC}_{ev}}$ state the minimum and maximum evaporation load for a plant to operate correctly. Multiplying by $X_{ev,p}$, it can be assured that $EC_{ev,p}$ takes zero value for the products not linked to plant *ev*.

8. The cooling water flow through the surface condenser is also bounded:

$$\underline{\mathrm{F}_{ev}} \sum_{p}^{\mathcal{P}} X_{ev,p} \leq F_{ev} \leq \overline{\mathrm{F}_{ev}} \sum_{p}^{\mathcal{P}} X_{ev,p} \qquad \forall ev \in \mathcal{E} \tag{5.9}$$

Where $\underline{\mathrm{F}_{ev}}$ and $\overline{\mathrm{F}_{ev}}$ states the minimum and maximum cooling-water flow that could pass through the surface condenser attached to the evaporation plant *ev*. Like the previous constraint, the cooling water flow is forced to be zero if there is not any product linked to the evaporation plant, in this case through $X_{ev,p}$ added for all products.

9. The total flow that feeds the surface condensers in Subnet 1 must be, as maximum, the available flow of Source 1 (MF1) minus the flow that goes to Subnet 2 (F_L). In Subnet 2, now there are three inputs of cooling water: Source 2 (MF2), remaining water from Source 1 (F_L), and the flow recirculated from heat pumps (F_M).

$$\sum_{ev}^{\mathcal{E}_1} F_{ev} \leq \mathrm{MF1} - F_L \tag{5.10}$$

$$\sum_{ev}^{\mathcal{E}_2} F_{ev} \leq \mathrm{MF2} + F_L + F_M \tag{5.11}$$

10. Energy balance for each heat pump:

$$\mathrm{Q}\, Y_{b,hp} = \mathrm{FB\ Cp}\ \rho\ \Delta T_{b,hp} \qquad \forall b \in \mathcal{B}, \forall hp \in \mathcal{H} \tag{5.12}$$

Where Q, FB, Cp and ρ are as in (1), and $\Delta T_{b,hp}$ is the temperature difference between the flow inlet and outlet in each heat pump. Note that such temperature difference is set by (5.12), so it can be rewritten as in (5.13) where the outlet water temperature at each heat pump ($T_{b,hp}$) can be computed from the inlet one. That is directly the outlet of the previous heat pump connected in series ($T_{b,hp-1}$).

$$\begin{cases} T_{b,hp} = T_{b,hp-1} - \dfrac{\mathrm{Q}\, Y_{b,hp}}{\mathrm{FB\ Cp\ \rho}} & \forall b \in \mathcal{B}, \forall hp \in \mathcal{H}\{1\} \\ T_{b,hp} = T_{OT} \coloneqq \dfrac{\sum_{ev}^{\mathcal{E}} To_{ev}\ F_{ev}}{\sum_{ev}^{\mathcal{E}} F_{ev}} & \forall b \in \mathcal{B}, hp = 1 \end{cases} \tag{5.13}$$

By this way, if a heat pump is not connected ($Y_{b,hp} = 0$), the outlet temperature is the same as the inlet one. Note that for the first heat pump in each block, $T_{b,1}$ is the temperature resulting from the mixture of flows leaving the surface condensers (T_{OT}).

The outlet water temperature from each surface condenser (To_{ev}) is computed with the data-based model (3.6) developed in Chapter 3, section 3.3.2, according to the inlet water temperature at the time the experiments were carried out ($\mathrm{T}_{in_{ev}}^{exp}$), the inlet water temperature to the surface condenser ($T_{in_{ev}}$), the flow (F_{ev}) and the state of fouling ($\mathrm{K}_{\mathrm{f}_{\mathrm{ev}}}$) of the surface condensers, this last being monitored online.

$$To_{ev} = \left(f(F_{ev}) - \mathrm{T}_{in_{ev}}^{exp}\right) + T_{in_{ev}} - \mathrm{K}_{\mathrm{f}_{\mathrm{ev}}} \tag{5.14}$$

Note that, now, $T_{in_{ev}}$ depends not only on the known water source temperatures ($\mathrm{T}_{\mathrm{MF1}}$ and $\mathrm{T}_{\mathrm{MF2}}$), but also on the temperature of the water potentially recirculated from the heat pump superstructure (mixture of the outlets from each heat pump block that is chosen to recirculate water via F_M):

$$T_{in_{ev}} = \mathrm{T}_{\mathrm{MF1}} \qquad \forall ev \in \mathcal{E}_1 \tag{5.15}$$

$$T_{in_{ev}} = \frac{F_L\,\mathrm{T}_{\mathrm{MF1}} + \left(\sum_{ev}^{\mathcal{E}_2} F_{ev} - F_L - F_M\right)\mathrm{T}_{\mathrm{MF2}} + T_M\,F_M}{\sum_{ev}^{\mathcal{E}_2} F_{ev}} \qquad \forall ev \in \mathcal{E}_2 \tag{5.16}$$

$$T_M = \frac{\sum_b^{\mathcal{B}} T_{b,hp_{|\mathcal{H}|}} Y_{b,1}}{\sum_b^{\mathcal{B}} Y_{b,1}} \qquad \forall ev \in \mathcal{E}_2 \tag{5.17}$$

11. The stream temperature that goes to the river (T_r) should be lower that the maximum temperature allowed by environmental regulation (MTO):

$$T_r \leq \text{MTO} \tag{5.18}$$

Such stream can be computed from the mixture between the surface-condenser outlet flows that do not feed the heat pumps, and the outlet of the heat-pump blocks that is not recirculated to the inlet, i.e., the stream F_N, whose temperature is the same as the recirculated one T_M, (see Figure 5.4):

$$T_r = \frac{(\sum_{ev}^{\mathcal{E}} F_{ev} \; - F_M - F_N \;) T_{OT} + F_N \; T_M}{\sum_{ev}^{\mathcal{E}} F_{ev} \; - F_M} \tag{5.19}$$

Once the mathematical model which represents the system has been described, objective function must be defined representing the operational cost, but accounting for the desired payback time for the purchased equipment as well.

5.3.2 Objective function

The operation cost of the network could be computed as the sum of three different terms: the cost due to the steam consumption in the evaporation plants; the one associated to the cooling-water consumption; and the cost of the electricity consumed by the heat pumps.

Once again, the steam consumption can be computed using the adapted model developed by (Kalliski et al., 2019) already described in Chapter 3, Section 3.5.

$$SSC_{ev} = \mathrm{a}_{1_{\mathrm{ev}}} \, K_{v_{ev}} \; + \; \mathrm{a}_{2_{\mathrm{ev}}} \, Q_{ev}^{cp} + \; \mathrm{a}_{3_{\mathrm{ev}}} \, EC_{ev} \tag{5.20}$$

$$ASC_{ev} = SSC_{ev} \sum_{p}^{\mathcal{P}} EC_{ev,p} \tag{5.21}$$

The steam-consumption contribution to the total cost ($Cost_{steam}$) is computed then as the overall ASC times the price of generating a unit mass of live steam in boilers ($\mathrm{P_{steam}}$). In the same way, the cooling-water cost ($Cost_{water}$) is the net amount of water consumed from the sources times the cost of pumping and pipes operation ($\mathrm{P_{water}}$). The electric power consumed by the connected heat pumps is known from their capacity Q and their Energy Efficiency Ratio (EER).

$$Cost_{steam} \coloneqq \mathrm{P_{steam}} \sum_{\mathrm{ev}}^{\mathcal{E}} ASC_{ev} \tag{5.22}$$

$$Cost_{water} \coloneqq \mathrm{P_{water}} \left(\sum_{ev}^{\mathcal{E}} F_{ev} - F_M \right) \tag{5.23}$$

$$Cost_{elect} \coloneqq \mathrm{P_{elect}} \frac{\mathrm{Q}}{\mathrm{EER}} \sum_{b}^{\mathcal{B}} \sum_{hp}^{\mathcal{H}} Y_{b,hp} \tag{5.24}$$

Thus, the objective function to minimize is the operation cost, which is given by the sum of the three terms:

$$OF \coloneqq Cost_{steam} + Cost_{water} + Cost_{elect} \tag{5.25}$$

Furthermore, the amortization of the heat pumps must also be considered. Hence, a new constraint is needed to consider the payback time (RP), that must be, at most, the one required by the company. Note that the company does not limit the initial investment, but the time needed to recoup it. The payback time is computed as the price of a heat pump[8] ($\mathrm{P_{HP}}$) divided by the savings over a year, where such savings can be computed as the difference between the operation cost of the current network ($\widehat{OF}$) and the new cost with heat pumps predicted by the model (OF).

$$\frac{\mathrm{P_{HP}} \; \sum_b^{\mathcal{B}} \sum_{hp}^{\mathcal{H}} Y_{b,hp}}{\widehat{OF} - OF} \leq \mathrm{RP} \tag{5.26}$$

Note that the potential savings obtained by using heat pumps could be higher than the values shown in this thesis, as the heat removed from the outlet stream might be used in other section of the plant.

---

[8] According to (de Kleijn Energy Consultants & Engineers, 2014) a reasonable price for this kind of heat pumps is 40 000€ each one ($\mathrm{P_{HP}}$).

5.3.3 Model parameters

In many of the above model constraints there are parameters whose value is fixed and known, and measurable inputs which depend on the operating conditions.

The prices of steam, cooling water and electricity are calculated internally in the company, so their values are omitted here due to confidentiality agreements.

The heat-pumps selected capacity ($\mathrm{Q} = 500\ \mathrm{kW}$) and the energy efficiency ratio ($\mathrm{EER} = 4$) were taken from the spec sheets provided by the manufacturers of this equipment (Danfoos Engineering Tomorrow, n.d.; de Kleijn Energy Consultants & Engineers, 2014; *Ochsner Energie Technik*, n.d.). Note that the number of heat pumps to use and their layout are decision variables in the model, but the maximum number of heat pumps to install depends on the elements stated for sets $\mathcal{B}$ and $\mathcal{H}$. For this case study, the first estimation is with $|\mathcal{B}| = 15$ and $|\mathcal{H}| = 3$ (thus, there are up to 45 heat-pump slots), which provides enough design flexibility to find optimal solutions. If the results of the optimization would indicate that either all heat pumps in a block or fifteen blocks in parallel are required, such sets would be expanded.

Finally, water availability at the two sources and their temperatures, the state of fouling of each evaporation plant, and the demand of each product, can change over time. Thus, different scenarios have been proposed accordingly, in order to consider different expected operation conditions. Note that the cost of operating the network without heat pumps also depends on the value of these parameters and conditions so, to be fair, it would be computed also via the proposed model forcing the no inclusion of heat pumps.

5.4 Two-stage stochastic formulation

The aim of stochastic programming is to find optimally robust decisions in problems that involve uncertainty.

In this case, the main objective is to know how many heat pumps to buy, considering that there are some input variables in the model whose future evolution is uncertain. Hence, different scenarios arise depending on the values assigned to such time-varying parameters or model inputs. Table 7 summarizes the scenarios that have been considered for this case study, according to the model inputs that may take different values in future uncertainty realizations.

Table 7. Scenarios defined according to expected operating conditions. Most probable values are in plain text whilst the fewer probable ones are colored in red.

$\mathcal{J}$	MF1 [m³/h]	T_{MF1} [ºC]	MF2 [m³/h]	T_{MF2} [ºC]	MTO [ºC]	SP_A [t/h]	SP_B [t/h]	SP_C [t/h]	SP_D [t/h]	SP_E [t/h]	φ [%]	$\widehat{OF}$ [€/h]
1	680	10.0	690	9.00	30.0	42.2	13.9	66.0	15.6	14.2	20.0	386
2	680	16.0	690	18.0	30.0	42.2	13.9	66.0	15.6	14.2	20.0	516
3	680	13.0	690	13.5	30.0	20.5	24.7	33.0	43.2	12.9	20.0	358
4	680	13.0	690	13.5	30.0	38.2	29.3	62.3	18.2	15.4	20.0	515
5	580	13.0	620	13.5	30.0	42.2	13.9	66.0	15.6	14.2	11.0	457
6	680	13.0	690	13.5	27.0	42.2	13.9	66.0	15.6	14.2	6.00	448
7	580	16.0	620	18.0	30.0	42.2	13.9	66.0	15.6	14.2	2.20	-
8	680	16.0	690	18.0	27.0	42.2	13.9	66.0	15.6	14.2	0.700	-

The eight scenarios have been selected according to the most common conditions observed over a year of operation, adding the worst possible cases. In the table, φ indicates the percentage realization probability of each scenario respect to the rest. $\widehat{OF}$ is the computed operating cost without heat pumps.

Note that there are two scenarios (nº 7 and nº 8) where it is not possible to fulfill all the constraints *without heat pumps* for the given operation conditions (infeasible problems). Thus, the payback computation via (5.26) cannot be performed for these scenarios ($\widehat{OF}$ cannot be computed). Nevertheless, note that they will be included in the design optimization anyway (in the objective function and rest of model constraints). Moreover, as the sum of their realization probability is lower than 3%, excluding them from the payback constraint will not vary the economic estimations significantly.

The two-stage stochastic approach distinguishes two types of decision variables: the first-stage ones correspond to the sizing of units or structure selection of the process, which has to be selected so that the process can cope with the constraints imposed by the different scenarios. And then, the set of second-stage variables is tailored to each scenario, so that they can be used to adapt the effect of the first-stage ones to the particular needs of each scenario. These variables correspond normally to operational or control actions and represent the adaptation that these systems might have against a particular scenario. Hence, they provide extra flexibility to the optimization, which is able to compute fewer conservative solutions.

Following such approach, in this case only the number of heat pumps to purchase (N) is the first-stage variable (here and now decision) and the rest of variables in the model are of second stage, as they can be adapted to each uncertainty realization (recourse variables). Thus, all these variables and the involved constraints are extended with a new sub-index $i \in \mathcal{I}$ that refers to the scenario numbers listed in Table 7 (e.g., the evaporation load now will depend on the evaporation plant *ev*, the product *p*, and the scenario i: $EC_{ev,p,i}$).

5.4.1 Monolithic formulation

It may happen that the optimal number of heat pumps to switch on in a particular scenario could be lower than the purchased heat pumps, i.e., using all the installed heat pumps could not be economically optimal if water availability and temperature are favorable, for instance. Hence, a new variable N must be defined which can be computed as the maximum number of heat pumps used in all scenarios:

$$N := \max_{i} \sum_{b}^{\mathcal{B}} \sum_{hp}^{\mathcal{H}} Y_{b,hp,i} \tag{5.27}$$

To define the payback constraint, different approaches can be followed depending on the desired robustness level. It is clear that the payback time horizon has to be computed with the number of purchased heat pumps, but this constraint could be applied for each scenario (hence it would consider the worst case) or just once, computing the weighted average for the different cases according to the probability φ assigned to each one. In this case, the proposed formulation follows this second approach, less conservative. Thus, (5.26) is rewritten as (5.28).

$$\frac{\mathrm{P_{HP}}\, N}{\sum_{i}^{\mathcal{I}\setminus\{7,8\}} \varphi_i\left(\widehat{\mathrm{OF}}_i - OF_i\right)} \leq \mathrm{RP} \tag{5.28}$$

Finally, the objective function to minimize is the weighted sum of the operation cost for each scenario according to its probability. Therefore, the resulting two-stage stochastic optimization problem is summarized by:

$$\min_{X_{ev,p,i},EC_{ev,p,i},F_{L_i},F_{ev,i},Y_{b,hp,i},F_{M_i},F_{N_i}} J \coloneqq \sum_{i}^{\mathcal{I}} OF_i\, \varphi_i \tag{5.29}$$

$$s.t.: (5.2) - (5.18), (5.28)$$

The results of the optimization will give the optimal number of heat pumps to purchase, as well as, for each scenario: the heat pumps to be used and their connection layout, the cooling-water distribution, and the product allocation to plants. Everything ensuring the average payback time RP in (5.28) is fulfilled.

Nevertheless, the presence of discrete and continuous variables as well as the nonlinear dependency among them in many constraints, make (5.29) be a non-convex mixed-integer nonlinear programming (MINLP) problem, with a high number of decision variables (defining set $\mathcal{B}$ with 10 elements ($|\mathcal{B}| = 10$) and set $\mathcal{H}$ with 5 elements ($|\mathcal{H}| = 5$), there are 218 variables for scenario, 125 binaries, so for the 6 feasible scenarios there will be 1 308 variables). In order to overcome the typical issues of slow solution convergence and high computational demands in a centralized formulation, a suitable problem decomposition is proposed.

5.4.2 Decomposed formulation

The reader may have noted that the only shared variable among all scenarios in the optimization problem (5.29) is the number of heat pumps to buy (*N*), arising in the payback complicating constraint. Thus, removing (5.28), the optimization of each scenario minimizing (5.29) subject to (5.2)-(5.18) can be solved independently for a given N, hence computed in parallel. Therefore, the key is to develop an upper-layer master problem to select the proper value of N that ensures the payback constraint.

The payback time can be computed once the results of each scenario are obtained, and then be compared with the upper RP value. However, this simple calculus does not ensure that the constraint is fulfilled. To solve this, an iterative procedure could be applied. First, each scenario is solved independently for the same initial value N and, with the locally obtained results, (5.27) and (5.28) can be computed. If (5.28) is fulfilled, that is it, an optimal[9] solution is reached. If not, it means that the number of heat pumps to purchase needs to be lower than the computed in the current iteration k. To impose this, it is necessary to incorporate a new constraint set (5.30) onto the model that forces the maximum number of heat pumps to use being lower than a predefined parameter $\mathrm{N}^{[k]} \in \mathbb{N}$, whose value has been obtained by (5.27) in the previous iteration $(k-1)$ and updated to $\mathrm{N}^{[k]} = \mathrm{N}^{[k-1]} - 1$.

$$\sum_{b}^{\mathcal{B}} \sum_{hp}^{\mathcal{H}} Y_{b,hp,i} \leq \mathrm{N}^{[k]} \qquad \forall \mathrm{i} \in \mathcal{I} \tag{5.30}$$

In the next iteration, we will have the results of each scenario according to this constraint, so the master problem can check again if (5.28) is fulfilled. This iterative procedure is formalized in Algorithm 3 below.

[9] Of course, the provided guarantees are just for local optimality if gradient-based MINLP solvers are used, as the problem is non-convex.

Algorithm 3. Decomposed iterative optimization

1: Set $k = 0$ and minimize subject to (5.2)-(5.18) for all scenarios
2: Compute $\mathrm{N}^{[k]}$ through N defined in (5.27)
3: While (5.28) is not fulfilled do:
4: $\mathrm{N}^{[k]} = \mathrm{N}^{[k-1]} - 1$
5: Minimize (5.25) subject to (5.2)-(5.18) and (5.30) for all scenarios
6: Compute the overall weighted operation cost of (5.30) from the scenario costs obtained in Step 5
7: Go back to Step 2

Note that this decomposition approach provides (local) optimality guarantees because: parameter $N^{[k]}$ only can take non-negative integer values; and RP is monotonic with respect to $N^{[k]}$, i.e., the payback time increases with the number of purchased heat pumps. This is because (5.28) is affine in N and the savings increment provided per year attenuate with the number of purchased heat pumps, see results in next section. In fact, thanks to these features, Algorithm 3 converges to a solution in a finite number of iterations, at most $\mathrm{N}^{[0]}$ iterations.

Note that this procedure is equivalent to the stochastic centralized formulation (5.30) but the computational demand and time to obtain a solution is significantly reduced, as the scenario optimizations in each iteration are solved independently, in a simultaneous fashion if parallelized. Thus, the computation time in each iteration is the one of the most demanding scenario, but now the size of each independent problem is of 218 variables (for a set $\mathcal{B}$ with 10 elements and set $\mathcal{H}$ with 5 elements). The maximum time to get a solution with zero relative gap of an iteration is ~10 minutes, and the computational time to solve the whole algorithm over an Intel® i7-7700 CPU machine with 32Gb of DDR4 RAM memory is ~15 minutes.

5.4.3 Results and discussion

Following the methodology described in the previous section, results showed in Figure 5.5 and Figure 5.6 have been obtained, where the percentage of savings is the difference between the operation cost of the current network

($\widehat{\text{OF}}$) and the new cost with heat pumps predicted by the model (OF), normalized by the operation cost $\widehat{\text{OF}}$.

The behavior of the system is that, as more cooling water goes through the surface condensers the more efficient the evaporation plants are and, consequently, the less steam consumption is needed by the evaporation plants. Therefore, the more heat pumps, the more cooling water may be recirculated being available to use in the cooling system. Nevertheless, at some point, the savings for the reduction of steam consumption are not high enough to compensate the electricity costs to switch on many heat pumps. Hence, each scenario has a different amount of heat pumps on to reach optimal operation, i.e., the lowest operation cost.

In this regard, the numbers in Figure 5.5 show how the percentage of savings obtained in each scenario increases with the number of purchased heat pumps, but only until such optimal number of heat pumps switched on is reached. Thus, when the number of purchased heat pumps exceeds the number of those that are optimal for a scenario, the savings remain constant.

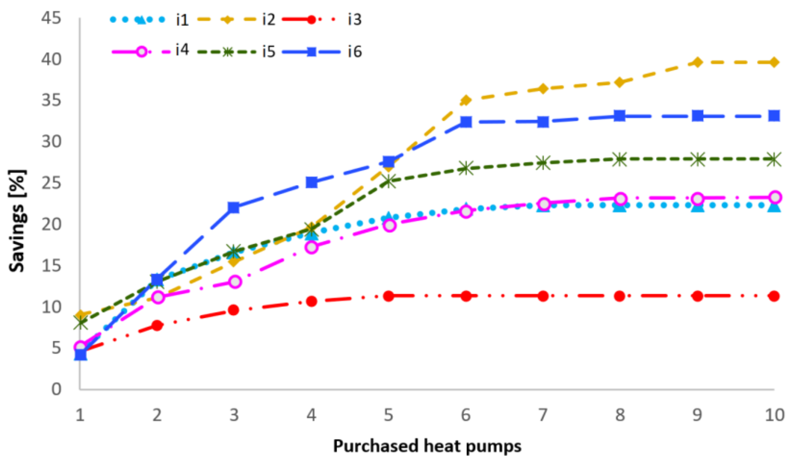


Figure 5.5. Savings percentage according to the purchased heat pumps for each scenario.

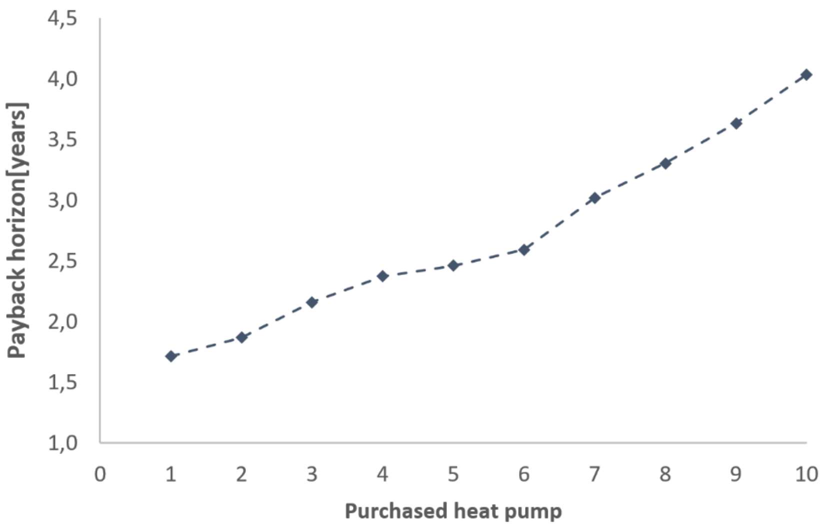


Figure 5.6. Payback time according to the number of purchased heat pumps.

If we focus in the payback time, showed in Figure 5.6, it decreases with the number of purchased heat pumps, as expected due to the monotonic behavior of (5.28) with respect to N. Although buying more heat pumps increases the savings, they progressively loose influence to compensate the increased initial investment, so that the payback time also increases. In addition, this relation is not linear, as when a scenario reaches the highest possible savings, increasing the purchased heat pumps will only increment the payback time.

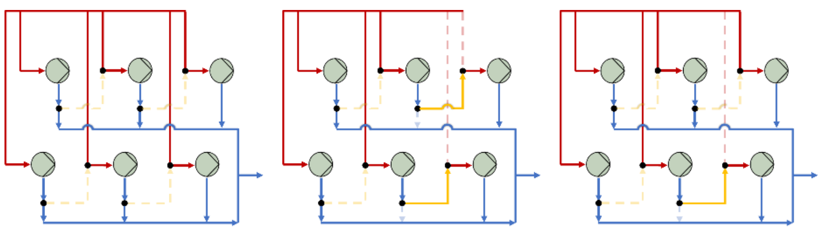
Note that Figure 5.5 and Figure 5.6 depict values in a range $N \in [1,10]$. This is because the first iteration applying Algorithm 3 already computed $N^{[0]} = 10$, i.e., the highest number of heat pumps suggested to be used was 10, obtained for scenario i4. Therefore, purchasing more heat pumps does not make sense because it does not translate in more steam savings, as the trends in Figure 5.5 clearly show. From that upper value, the proposed procedure has been followed, decreasing N in each iteration until $N = 1$ just to show a complete view of the system behavior. For the actual case, where RP must be 3 years as maximum, the number of heat pumps to buy is six, stopping the algorithm after five iterations. This choice predicts obtaining 22.7% of steam savings. Note that some scenarios would obtain more savings with more heat pumps, but their realization probability is not enough to compensate the required investment to be amortized in three and a half years. Nonetheless, the door to buy additional heat pumps in the future is open if conditions vary.

Note importantly that scenarios i7 and i8, which are not included in the payback computation because they are infeasible with the current system, would be indeed feasible with the heat pump integration. In this way, investing in the heat pump integration can ensure not only mere resource savings but also that the site would keep production capacity in extreme conditions, that is not possible currently.

Remark that, although the optimal number of heat pumps to purchase is six, the optimal number of them to switch on will vary with time, according to the operation conditions.

With respect to the installation layout, for scenarios i1 to i6 the optimizer chooses to connect all suggested heat pumps in parallel (Figure 5.7.a). Nevertheless, for scenarios i7 and i8, the layout would be:

- Scenario i7: four heat pump blocks in parallel where two of them have two heat pumps connected in series (Figure 5.7.b).
- Scenario i8: five heat pump blocks in parallel, with only one of them having two heat pumps connected in series (Figure 5.7.c).



(a) (b) (c)

Figure 5.7. Installation layout according to the different scenarios: (a) represents the superstructure with all the heat pumps connected in series; (b) shows the layout for four heat pumps connected in parallel and two in series; (c) reflects the configuration for five heat pumps connected in parallel, one in series. Dashed lines indicate existing connection but not used in the scenario. Yellow lines indicate the outlet flow from a heat pump which is used as inlet for another (serial connection).

Therefore, the proposed installation layout is a matrix of six heat pumps, initially connected in parallel, but with two of them having the possibility to drive its outlet into the inlets of other two, in order to connect these heat pumps in series. This piping configuration is easily implementable with on/off valves in the inlets and outlets of the involved heat pumps (represented with black circles in Figure 5.7).

Finally, the only question that remains open is how the heat pumps should be incorporated into the actual network layout, regarding what to do with the outlet flow from the heat pump section. For most scenarios, suggestion for such flow is complete recirculation to the inlet of Subnet 2. Nonetheless, when the inlet temperature from the sources is quite high, or in the hypothetical cases where the maximum temperature allowed by the environmental regulation (MTO) decreases, the computed results suggest recirculating only part of such flow as the most beneficial course of action. The remaining part will join the water going to the river. Therefore, the optimal installation design is to incorporate a flow control valve at the outlet of the heat pump section, able to modify the flow that goes each way, i.e., the values of F_M and F_N in Figure 5.4.

5.5 Network optimal operation

After the incorporation of the new equipment into the network, the only remaining issue is to operate the network in the best efficient way. In addition to knowing the evaporation load that each evaporation plant must process regarding the demand of the different products, and the cooling water distribution according to the availability of the sources, now it is also needed to know how many heat pumps to switch on, as this depends on the actual conditions arising in real time.

Therefore, the solution provided in Chapter 3 must be widen in order to take into account the incorporation of heat pumps. The increment of available cooling water due to the heat pumps, not only would change the cooling-water distribution, but also might change the evaporation load allocation. Furthermore, one needs to incorporate the decision on how many heat pumps to connect in the RTO formulation.

To do that, the mathematical formulation of the previous section is reused to develop an RTO tool that determines the optimal operation setpoints according to the real-time conditions. The optimization problem to solve is the same as the one in the decomposed approach for a particular scenario, i.e., to minimize the total cost of operating the network subject to (5.2)-(5.18) and (5.30), taking into account that in (5.30), the total number of heat pumps (N) is now fixed to six. Furthermore, to speed up the resolution, the number of useless decision variables is decreased by setting that the set $\mathcal{B}$ only includes 6 elements ($|\mathcal{B}| = 6$) and set $\mathcal{H}$ will include just 3 elements ($|\mathcal{H}| = 3$), according to the layout configuration of Figure 5.7.

The parameters and input values which depend on the operation conditions (the availability of the cooling water sources and their inlet temperatures, the state of fouling of each plant, and the demand of the different products) will be supplied to the RTO by the site information technology (IT) system. Currently, as the purchase of heat pumps is not materialized yet, the RTO has been tested offline, accessing to historical plant data. Nevertheless, the idea is that each time the end-user wants to execute the optimization, it automatically gets the real-time values of the needed parameters and inputs.

Using six heat pumps, the number of decision variables in the RTO is 186, from which 93 are binary, and there are 320 constraints. As an RTO requires obtaining solutions in short time periods, the algorithm used to solve the problem is NLP-based branch-and-bound algorithm BONMIN (Bonami et al., 2008), at the price of possibly reaching a local optimum from time to time.

With this setup, the time to get a solution with zero relative gap is ~40 seconds (depends on the scenario input data) over an Intel® i7-7700 CPU machine with 32Gb of DDR4 RAM memory. This computational time is considered suitable, as the frequency to account for sensibly observable changes in the evaporation demand or sources availability is more than twenty minutes.

Finally, Figure 5.8 and Figure 5.9 show a comparison between the historical data (used to perform the optimization), the results obtained with the centralized formulation described in Chapter 3, and the results obtained in the case that six heat pumps were integrated according to the layout above mentioned.

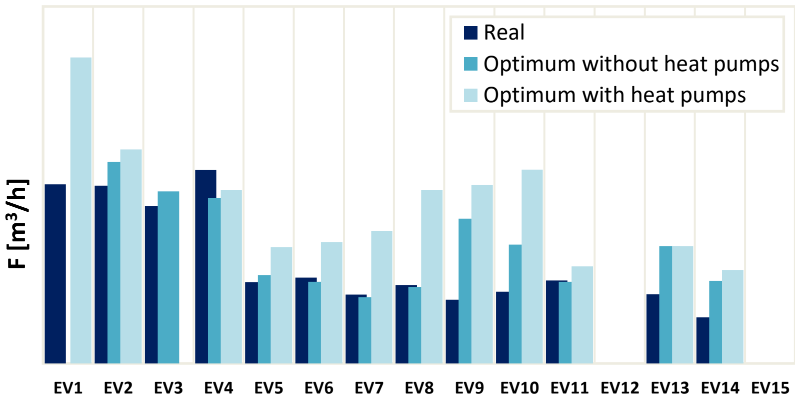


Figure 5.8. Comparison of the cooling water distribution

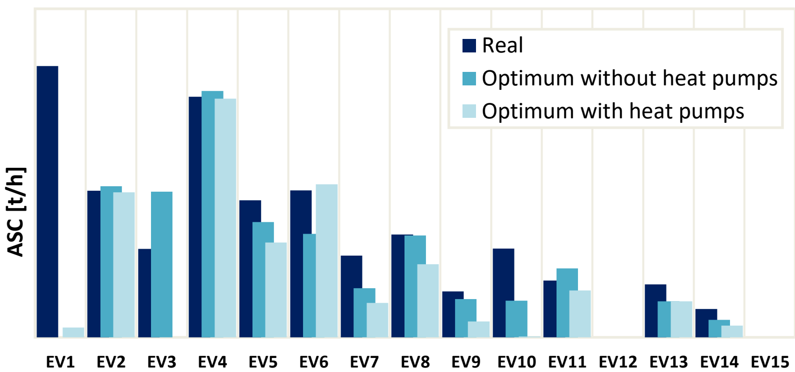


Figure 5.9. Comparison of the steam consumption

As it was expected, increasing the amount of cooling-water availability, the steam consumption decreases. Such decrease compensates the electricity cost of the heat pumps, being the total operational cost of the evaporation process 9.8% lower than without heat pumps (32% lower than the historical one). It confirms that the integration of heat pumps will be very beneficial once they are amortized.

5.6 Summary and conclusions

This chapter revisited two important problems in process systems engineering from a holistic perspective: the optimal design for the incorporation of new equipment into an existing facility, and the optimal operation of the modified system.

Considering the evaporation system, whose optimal operation has been discussed in previous chapters, a rigorous mathematical model of the network has been developed including all the possible layouts of the new equipment to be incorporated, as well as their potential connections. Based on this hybrid model, an optimization problem that indicates the optimal layout of the equipment such that the operation cost of the whole network is minimum, has been formulated. However, as the results depend on the operation conditions, whose future realizations are somehow uncertain, a two-stage stochastic formulation has been built for different scenarios that intend to cover reasonably the plausible operation conditions. The first approach was to formulate the optimization problem in a monolithic way, but due to the characteristics of the resulting optimization problem (MINLP problem), an equivalent decomposed approach was proposed. Such approach was selected because the individual formulations for each scenario are only linked by a single constraint (payback time horizon) involving a shared variable (number of heat pumps to purchase). Moreover, as the shared variable is an integer positive number and the payback-time constraint is monotonic with such variable, the solution is proven to be obtained in few iterations with the proposed decomposition algorithm.

Analyzing the solutions, the best configuration is gotten considering the payback time that the company considers acceptable. In general, the benefits increase with the number of used heat pumps, up to an extent, as more recirculated water involves lower steam consumption. However, at some point the reduction cost of steam does not compensate the electrical cost of installing more heat pumps. It is worth mentioning that the optimal number of heat pumps to use and its layout depend on each scenario.

Consequently, once the optimal design is found assuming optimal operating conditions, the daily operation should be addressed accordingly. For such a task, an RTO was built based on the same model of the network, already developed. From current plant situation (real-time collected data), the RTO suggests the optimal connection of heat pumps, the evaporation-load allocation and the cooling-water distribution according to an economic criterion. Consequently, the proposed system architecture is able to quickly react to disturbances or evaporation load changes.

The core of the work presented in this chapter has been published in an indexed peer reviewed journal (M. P. Marcos et al., 2021)

Chapter 6

Decision support systems

In the previous chapters, the reader may have realized the difficulties associated to the operation of different sections of a plant due to the complexity given by the large number of feasible alternatives. Different RTO schemes have been developed to indicate the best setpoints for the manipulated variables, in order to give the lower operational cost. Such RTOs have been designed to be used as the base of two Decision Support Systems (DSS), which help operators in the decision-making process on how to operate the heat-recovery and the evaporation network respectively.

In this chapter, a brief introduction of DSS is given, followed by the implementation of the different components that compose the DSS. Then, the interfaces developed for the corresponding DSS are explained. After that, how the DSS are integrated with the plant control system is briefly described. Finally, some conclusions are presented.

6.1 Introduction

In 1965, with the emergence of digital computers, building management information systems in large corporations became more practical and cost-effective. Thus, model-oriented decision support systems (DSS) became practical in order to help decision makers to simulate different possibilities and select the best course of action (Rashidi et al., 2018).

Nowadays, due to the increased level of structural complexity of industrial problems, decision making in a plant is more complicated for human operators and managers than it was in the past. Furthermore, if an error or bad decision is taken, a chain reaction of magnification of costs can occur. Hence, DSSs have experienced a noticeable attention growth, not only in industry but also in academia over the past two decades (Rashidi et al., 2018).

A DSS is defined by (Sprague, 1980) as "*interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems*". Another simplified and less restrictive definition of a DSS can refer to any computer-based system that provides information that enables to make decisions. According to (Aronson et al., 2005) there are four phases in the decision-making process: (*i*) the intelligence phase where the problem identification is carried out; (*ii*) the design phase, where the alternative options are generated; (*iii*) the choice phase, where, after an analysis of the alternatives, an option is chosen and; (*iv*) the implementation of the chosen alternative, i.e., the solution and its monitorization in order to guarantee the successful response. It is noteworthy that, usually, the final system cannot be fully automated, as a perfectly processed information and an optimum model is needed. However, in industry, models are simplified representations of reality and, consequently, subject to uncertainty. Therefore, the end user is still the one in charge of taking decisions, but combining its expert knowledge with the valuable suggestions given by the DSS.

The two main DSS types that we can find in industry are data-based DSS and model-based ones. The first type just analyzes high amount of data (current and past) from different sources, usually stored in a database, i.e., what nowadays is called *Big Data* (Camargo-Vega et al., 2015). The extracted information from the analysis of these large data pools helps managers in making better decisions. On the other hand, the model-based DSS helps managers to perform simulations and what-if analysis based on a reliable model which represents all the feasible alternatives of a system.

This last kind of DSS consists of three main components (see Figure 6.1): (*i*) a database or data management, which includes not only the data from the process, but also the knowledge in form of a model or expert system that synthesize knowledge of the manager and operators; (*ii*) a software system based on mathematical or analytical models, used to simulate the alternatives and to give an output; and (*iii*) a user interface where such output is displayed in order to give the end user an appropriate view of the recommended actions (Haag & Cummings, 2009).

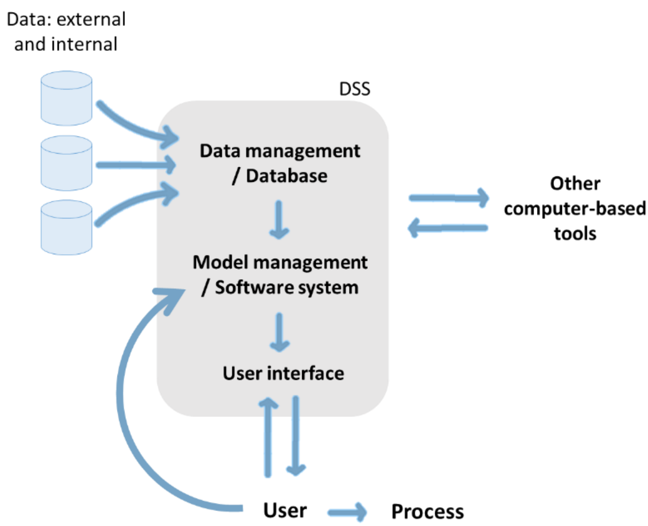


Figure 6.1. Scheme of a DSS

According to the type of model used in the software system component, they can be classified as: statistical models, sensitivity analysis models, optimization analysis, forecasting models, and backward analysis sensitivity models (Power, 2002). The optimization-analysis ones are used to find optimum values for target variables under given circumstances. Its widest use is to help in the decision-making process of optimal use of resources. Hence, an RTO could be the software system component of a DSS where the optimization analyzes the best values for the decision variables, and the results are displayed in the user interface.

Indeed, in the literature there are many examples of specific DSS concepts based on mathematical programming, not only in industrial applications (Galan et al., 2021) but also in other areas as medical routines (Kandakoglu et al., 2019), disaster response (Cavdur & Sebatli, 2019) or transportation (Erdoğan et al., 2019).

6.2 Implementation

In the particular case of DSS developed in this thesis, the first component is the data base which contains information of two types: the one from expert knowledge, more static with time and directly reflected by a model (e.g., physical laws, possible connections of heat sources to exchangers, etc.); and the operation data that changes over time (inlet stream temperatures, temperature setpoints, the state of fouling, prices, etc.).

The second component is the core of the DSS, the DSS software system. In this thesis, two DSS have been developed and their respective cores are the optimization problems described in Chapter 4 and Chapter 5. The first one determines the optimal operation of the heat-recovery and the second one the optimal operation of the evaporation network including the cooling-water system. As described in the previous chapters, such optimization problems have been coded in Pyomo-Python (Hart et al., 2017), a toolbox for efficient numerical optimization and control.

Nevertheless, to be able to solve the optimization problems according to the real time conditions, they need to be supplied continuously with plant data. Thus, the DSS software must be integrated with the information technology (IT) infrastructure of the plant as well as with the operation one (OT). For this purpose, we made use of PIconnect (*PIconnect · PyPI*, n.d.), a Python interface to the OSIsoft Plant Information system of the plant (OSIsoft, 2020), to access (and periodically update) the relevant data in order to feed the DSS software system in real time, i.e., before each optimization run.

Despite proposing an RTO software, the loop was not closed, as perfectly processed plant-model information cannot be assured. Therefore, the human manager is still the one in charge of taking decisions based on expert knowledge, but now helped by the valuable suggestions provided by the DSS.

Finally, the DSS was completed with a user interface which gives not only the results of the optimization but also a suitable overview with relevant information of the process during production. Thus, a friendly interface where the results of the optimization translate also into clear directives to be performed by the human is needed. By this way, the person in charge can take a look at the presented suggestions and combined with his/her expert knowledge, decide how to proceed.

According to the plant personnel preferences, the interface was designed in MS Excel. It displays the important information and suggested actions, previously computed by the RTO in the backend. The connection between Python and MS Excel is done via OpenPyXL (E. Gazoni and C. Clark, 2020). It is noteworthy that this is a bi-level interface, where plant engineers have further access to modify some optimization parameters, model coefficients and equations in case that something in the network layout changes (for example the incorporation of another heat pump).

Figure 6.2 depicts a scheme of the proposed workflow, showing the interactions of information between the DSS components and end users. Only the plant engineers will have access to the backend (i.e., to the DSS software system) in case model parameters need to be updated, or some constraints need to be modified or added. For example, to include a constraint to set up a minimum operation time before cleaning or limiting the number of heat exchangers that can be cleaned at once for the DSS of the heat-recovery process.

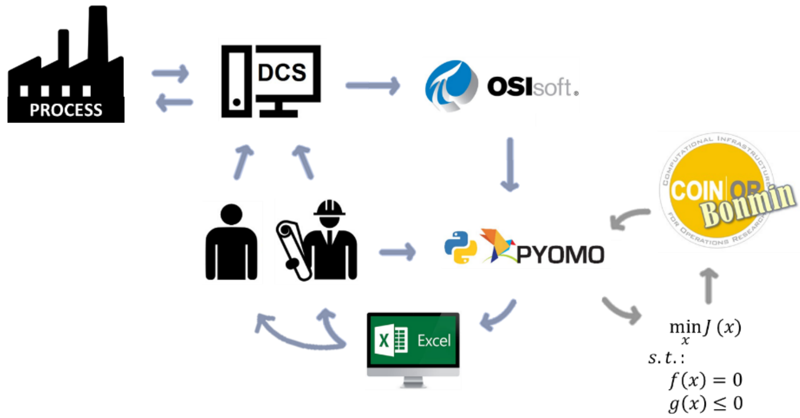


Figure 6.2. Proposed workflow within the DSS components and the end users

6.3 Interfaces

The results obtained with the RTO need to be presented in a simple and understandable way to the human operators and plant managers. Consequently, friendly DSS interfaces tailored to the end-users' background and responsibility level (e.g., plant operators versus maintenance personnel or plant managers), need to be provided (Rashidi et al., 2018).

Remark that the developed interfaces allow to interpret the given solution of the RTOs, but the end user is who decides whether to follow the recommendations or not.

6.3.1 Heat recovery network interface

For helping in the decision-making process of the heat recovery network operation, the operators and managers are supplied with the dashboard concept shown in Figure 6.3, which displays both the computed optimal solution for the current time (the optimization problem described in Chapter 4) and the process data used to obtain it.

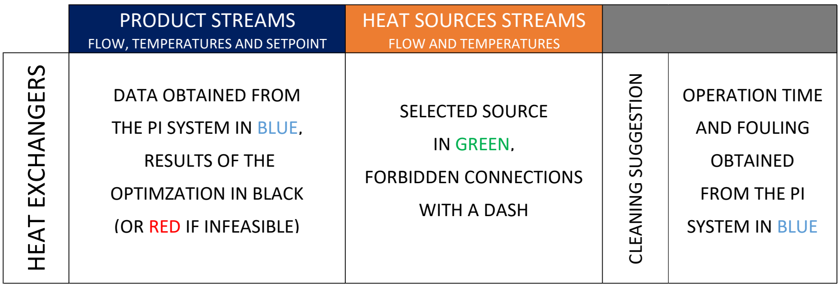


Figure 6.3. Designed concept for the DSS dashboard, tailored to plant operators.

The heat exchangers present in the network will be ordered in rows, where each column will then represent the features of the heat-exchanger streams, which can be separated in four sections. The first column lists the heat exchangers while the second set of columns provides information about the cold streams to be heated: the flow, inlet temperature, setpoint to achieve and outlet temperature. The third section gives information about the heat sources with the suggested connections to exchangers (they will be highlighted in green, and the no feasible connection between source and exchanger will be in dark gray, see Figure 6.4). The last columns devote to the cleaning suggestions: the operation time since last cleaning, the estimated state of fouling (in a red scale as it increases), and the exchangers recommended to clean (a green tick if the heat exchanger should be cleaned or a red cross otherwise, see Figure 6.4).

Moreover, to clearly differentiate the data read from the PI system from the results computed by the RTO, the first are displayed in blue italics and the second one in black plain text. Furthermore, in the case that the optimization problem resulted infeasible for any reason (i.e., some heat exchangers would not fulfill the product temperature setpoints), the computed values below the demanded temperature will be highlighted in red. With this visual design, the end user can quickly identify the recommendations given by the RTO and analyze them at a glance.

Figure 6.4 shows an example of how the interface looks with the results obtained by optimizing for some conditions read from historical data.

		Product stream				Hot source streams						Cleaning?	Op Time (days)	Fouling [W/(m²K)]
		flow (m³/h)	inflow (ºC)	SP (ºC)	outflow (ºC)	S1 (m³/h)	S2 (m³/h)	S3 (m³/h)	S4 (m³/h)	inflow (ºC)	outflow (ºC)			
HEAT EXCHANGERS	1	15	35,0	60,0	67,0	40,92				70,0	58,2	✗	10	20,0
	2	60	25,0	40,0	40,7	36,29				70,0	43,9	✗	15	30,0
	3	50	30,0	50,0	50,0	49,62				70,0	49,7	✗	5	10,0
	4	140	47,0	55,0	55,0	54,43				70,0	49,3	✗	0	0,0
	5	130	45,0	55,0	55,0	61,93				70,0	48,9	✗	3	5,0
	6	50	45,0	60,0	60,0	0,00		32,27		75,0	51,6	✗	3	6,0
	7	25	40,0	60,0	60,0	0,00		30,68		75,0	58,6	✓	17	35,0
	8	60	20,0	40,0	40,0	0,00	131,67	0,00		50,0	40,9	✓	22	45,0
	9	60	20,0	45,0	45,0	0,00	0,00	50,95		75,0	45,4	✗	6	12,0
	10	60	25,0	40,0	40,0	0,00	76,58	0,00		50,0	38,2	✓	12	25,0
	11	45	17,0	35,0	35,0	0,00	34,95	0,00		50,0	26,7	✗	1	3,0
	12	50	25,0	-	45,0			21,86	0,00	75,0	29,0	✗	12	24,0
	13	50	45,0	-	45,0			0,00	0,00	0,0	0,0	✗	11	22,0
	14	50	45,0	-	45,0			0,00	0,00	0,0	0,0	✗	18	36,0
	15	50	45,0	45,0	45,0			0,00	0,00	0,0	0,0	✗	9	18,0

Figure 6.4. Interface for the DSS of the heat-recovery section

6.3.2 Evaporation-process interface

The interface developed for the evaporation process extends the one that Lenzing already has to manage the plants evaporation load. It shows the solution obtained with the RTO on a panel but, in addition to the evaporation load allocation, it also incorporates information about the cooling-water distribution and the heat-pump connections.

Figure 6.5 shows how the interface will look when the integration of the six heat pumps will be realized. Note that the reflected data are the results obtained for some conditions read from the historian, as the updated tool has not been implemented yet.

		EVAPORATION PLANTS															TOTAL [t/h]	DEMAND [t/h]
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
EC [t/h]		21,00	15,60	21,18	0,00	13,92	0,00	10,00	0,00	12,00	11,00	8,00	3,21	9,00	11,00	15,99	**151,90**	**151,90**
PRODUCTS	A																42,18	42,18
	B																13,92	13,92
	C																65,99	65,99
	D																15,60	15,60
	E																14,21	14,21
Cool. Water [m3/h]		300,0	163,9	139,7	0,0	98,5	0,0	65,0	0,0	155,7	110,2	80,0	50,0	115,0	82,1	125,0		

	HEAT PUMPS					
	HP1	HP2	HP3	HP4	HP5	HP6
Connected	P	P	S	P	S	S

FM [m3/h]	FN [m3/h]	Open valve [%]
255,0	0,0	100,0

Actual cost [€/h]	Optimal cost [€/h]	Benefit [€/h]	Benefit [%]
543,2	431,3	111,85	**20,59**

Figure 6.5. Designed interface for the evaporation process.

As Figure 6.5 shows, the interface is divided in different sections.

The first panel reflects the evaporation-load allocation and the cooling-water distribution, where columns represent the evaporation plants. The first row indicates the evaporation load in t/h. The next five rows correspond to the products assignment to the evaporation plants. The optimal evaporation load assignment arises in the green boxes, meanwhile dark gray boxes indicate forbidden connections. Finally, the last row displays the cooling water to each surface condenser in m^3/h.

There is a two-column panel at the right side. The first column indicates the total evaporation load processed by all the plants, and the second one is the evaporation demand for each product. These two need to match for the allocation to be right.

At the bottom, three more display panels are included. The first one on the left indicates if each heat pump should be switched on (also in green) and how it should be connected, displaying a P if the connection is in parallel or an S if it is in series. If the heat pump should be off, the cell will be in dark gray. The next panel shows the flow that should go to each side after the heat pumps (to be recirculated or to be mixed with the outflow that goes to the river).

Finally, the right-side panel shows the total cost of operating the network in the current conditions versus the cost predicted with the proposed distribution. Moreover, the savings that will be presumably achieved by applying the optimized distribution are explicitly displayed in order to encourage the operator to follow the recommendations proposed by the tool (Kujanpää et al., 2017).

6.4 Integration with the control system

The developed DSSs suggest the values of the manipulated variables (cooling-water flow and evaporation load for each plant, and heat-sources flow to heat exchangers in the heat-recovery process) so that the controlled variables (total evaporation demand and product temperatures respectively) would *ideally* reach the desired setpoints according to the models.

However, note that none of the models in this thesis include dynamics, so the actual implementation must be driven by the existing distributed control system (DCS). Moreover, this control system is necessary to correct the probable plant-model mismatch in the formulated optimization problems. Thus, the integration of the DSS recommendations with the existent DCS is as follows: the RTO computes feasible decisions (according to the model) regarding to the allocation of sources to heat exchangers for the heat-recovery process and the assignment of the evaporation load and the cooling-water flow to surface condensers at minimum cost, such that this provides the plant PID controllers with *reachable* setpoints.

Regarding the heat-recovery network, as temperature dynamics is normally slower than the fluid-mechanics one, a minor modification of the existent DCS is proposed to include a cascade-control structure. In there, the internal PID loop sets the flow of heat source (good initial guess given by the DSS) and the external PID loop just modifies such flow setpoint if necessary to reach the temperature setpoint, compensating thus any small plant-model mismatch. In this way, the proposed implementation allows a faster response against operation changes in real time.

6.5 Conclusions

Once the RTO schemes for the optimal operation of the different networks are developed, the proposed solution is to incorporate them in specific DSSs that keep the human operator in the loop. Therefore, such DSSs are based on rigorous mathematical models of the networks, which are the core of an MINLP optimization accounting for the current production constraints and the equipment fouling states in each case.

The DSSs should be fed in real time by the plant PI system and present the results of the optimization through a simple interface, designed according to the operator visualization preferences (Excel in this case). In this way, the proposed DSSs will support the plant operators in such a complex decision-making process in real time, saving resources and reducing the personnel workload. This work is a proof that DSSs based on mathematical models and mixed-integer nonlinear optimization can unlock the potential benefits associated to complex combinatorial problems arising in the daily management of industrial sites.

A drawback of the proposed DSSs is relative to the ease of adaptability and flexibility. Ideal DSSs should be formed of simpler pieces such that end users could be able to build and modify them easily. This does not mean that the proposed DSSs in this thesis are rigid black boxes, but the complexity of the underlying models requires to be adapted by process experts, with assistance from qualified engineers on mathematical optimization. Consequently, future development can focus on developing a library of components with models for the individual equipment and a method to provide plug and play features to ease the inclusion and modification of the model constraints.

The interfaces presented in this chapter can be also found in (M. P. Marcos et al., 2020b) for the case of the heat-recovery network DSS and in (M. P. Marcos et al., 2021) for the evaporation process one.

Chapter 7

Contributions, final conclusions & outlook

In the process industry, optimal energy and resource operation is key. However, the optimal operation in real time of systems having a wide variety of alternatives is very complex, as it is widely discussed in the related literature. Despite this, after a deep study of the literature published on the topic, one can find out that current research often provides academic case studies to develop and prove their formulations and algorithms. Hence, on the one hand, when one tries to apply such methods to a real industrial case, they often do not fit properly or do not get the expected results. This is because the formulations are too generic or do not scale well enough. On the other hand, the industrial-size cases reported in the literature are very specific for a particular environment, so it is not feasible to apply them directly in other contexts.

There is a wide variety of industrial environments and the complexity of each one makes nearly impossible to find a single approach. Consequently, every situation has to be studied to choose the best-suited methodology and formulation for every case. In addition, some of the most common problems regarding optimal operation in the process industry are related to the development of the models that represent the real behavior of the systems, and the coordination and implementation of the obtained solutions.

This thesis is the last result of years of collaboration between academia and an industrial company, dealing with these problems in a real case study, providing innovative solutions and efficient formulations for the system considered. Hence, this thesis contributes to the development of the field in Process Systems Engineering and in the implementation of the principles and aims of the so-called Industry4.0.

Nevertheless, even though the proposed solutions and technologies are customized to some existing problems in a viscose fiber-production plant, other process and food industries have similar equipment and issues (heat-exchanger networks, allocation of products to plants, equipment degradation, etc.). Therefore, the ideas and developments here described can be adapted and extended to similar scenarios with reduced effort.

7.1 Thesis contributions

The core contribution of this thesis is the formulation of RTO-DSS solutions related to the optimal operation of different sections (an evaporation network, a cooling-water network and a heat-recovery one) on a viscose fiber production site, where there are several products to process, shared resources, varying equipment efficiencies (also over time) and multiple arrangement alternatives. The end goal is improving the resource efficiency by better coordination of the operation, current and future.

With regard to the proposed objectives for this thesis, mentioned in Chapter 1, the main contributions have been:

- To study the problems of optimally operating different sections of a plant in real time incorporating discrete and continuous decision variables, as well as the extra issues derived from equipment degradation, fouling in heat exchangers in this case. Such topics have been covered in Sections 3.1 and 4.1 for the different case studies.

- To develop data-driven models for some difficult to fit by first principles relations among process variables (the cooling-water temperature leaving the surface condensers with respect to the flow, the specific-steam consumption with respect to the cooling power, and the overall heat transfer coefficient in the heat exchangers with respect to the stream flows). All the details can be found in Sections 3.3 and 4.3.
- To expand the developed data-driven models to explicitly consider current fouling states. This is also addressed in Sections 3.3 and 4.3.
- To develop efficient hybrid models for the optimal operation of the different networks in real time, dealing with limited shared resources. The formulations of the different optimization problems can be found in Sections 3.4 and 4.4.
- To incorporate suggestions on maintenance tasks according to the efficiency loss due to fouling. It was included in the model for the optimal operation in Section 4.4.
- To study and develop different formulations for the coordination of the operation of two plant sections. The different approaches are explained in Sections 3.5 and 4.5.
- To analyze the incorporation of new equipment in order to increase the efficiency of a section. Such topic has been covered in Section 5.1. For this aim, the primary RTO formulation is expanded with a superstructure that models the operation of the network in every possible configuration of the new equipment. This is presented in Section 5.3.
- To further elaborate the formulation to tackle the uncertainty in future operation conditions when doing re-design of an existing system.
- To provide efficient ways to solve the complex and/or large-scale optimizations derived from the addressed problems. In particular providing alternatives to the usual monolithic approach that are based

on decomposition strategies. Coordination of plant sections is addressed in Section 5.3, and decomposition of stochastic optimization for process re-design is in Section 5.4.

- To develop DSS interfaces to properly show the results of RTOs to plant operators and managers. Such interfaces have been described in Section 6.3.
- To validate the proposed methods and algorithms with industrial data, and to guide the implementation of the developed RTO schemes via DSSs and the existing DCS in the plant (with minor modifications in some cases). The results of the validation have been presented at the end of Sections 3.4, 4.4 and 5.5, and the implementation as part of the DSS in Section 6.2.

The research work performed within the framework of this thesis has led to several publications in peer-reviewed national and international conferences as well as in indexed JCR journals, listed hereunder:

- Marcos, M. P., Pitarch, J. L., & de Prada, C. (2021). Integrated Process Re-Design with Operation in the Digital Era: Illustration through an Industrial Case Study. *Processes*, *9*(7), 1203.
- Marcos, M. P., Pitarch, J. L., & de Prada, C. (2020). Decision support system for a heat-recovery section with equipment degradation. *Decision Support Systems*, 137, 113380.

o Marcos, M. P., Pitarch, J. L., & de Prada, C. (2020). Modelling and real-time optimisation of a heat-exchanger network. *21st IFAC World Congress 53*(2), (pp. 11780-11785). Berlin, Germany.

o Marcos, M. P., Pitarch, J. L., & de Prada, C. (2019). Real-time optimisation for a heat recovery section with equipment degradation. *XL Jornadas de Automática*, (pp. 513-519). Ferrol, Spain.

o Marcos, M. P., Pitarch, J. L., Jasch, C., & de Prada, C. (2019). Optimal distributed load allocation and resource utilisation in evaporation plants. *Computer Aided Chemical Engineering, 46*, (pp.979-984). Eindhoven, The Netherlands.

- Marcos, M. P., Pitarch, J. L., de Prada, C., & Jasch, C. (2018). Modelling and real-time optimisation of an industrial cooling-water network. *International Conference on System Theory, Control and Computing (ICSTCC*), (pp. 591-596). Sinaia, Romania.
- Marcos, M. P., Pitarch, J. L., de Prada, C., & Jasch, C. (2018). Modelado para operación óptima de un sistema de refrigeración industrial. *XXXIX Jornadas de Automática*, (pp. 702-709). Badajoz, Spain

The developments and solutions in this thesis have been motivated and applied mainly thanks to the European H2020 project *Improved energy and resource efficiency by better coordination of production in the process industries* (CoPro), which intends for a better resource usage in the European process industry, contributing to reduce the climate footprint. This is in line with the two national research projects in which the author also collaborated: *Integración de optimización y control en plantas de procesos* [Optimization and control integration in process plants] *(INOPTCON)* and *Integrated plant wide control and optimization for Industry4.0* (InCO4In). Both aim at providing solutions to the theoretical and practical challenges that the implementation of the Industry 4.0 concepts in the process industry brings.

7.2 Author's conclusions

Here below my main conclusions from the different technologies developed and presented in this thesis are exposed, as well as some limitations that I personally think that need to be further addressed:

- The gray-box models developed allow to avoid the typical problems of first-principles models when applied in industry (excessive complexity and lack of flexibility when not all data is available), but still imposing some physical coherence to the relations between process variables, the main weakness of black-box models. However, such models are still sensitive to disturbances and noise error present in the data used to obtain them, so two compatible solutions have been proposed: a modeling routine to improve the computation of the models and performing a data reconciliation when there is redundancy on data or due to additional algebraic constraints.

- Although the formulated problems are MINLP due to the intrinsic systems characteristics, the proposed approaches are able to provide results in relative short periods of time. This is positive not only from an implementation point of view, where results must be obtained faster than the actual process changes, but also from a usability point of view since end users will be more willing to use tools if they do not have to wait for long time periods to obtain results. Nevertheless, it must be borne in mind that this quick solution only can be obtained using a local-deterministic method, which may lead to not globally optimal solutions that depend on the initial guess. In the author's opinion, the best option is to use the real-time conditions of the plant as initial guess. How to systematically set a good initial guess for unmeasured variables, rather than based on expertise, is still a topic which has room for improvement.
- The proposed RTO schemes have been successfully tested with historical plant data. It has been predicted that operating costs can be reduced if the solutions proposed by the RTO are executed. These results reflect the complexity of the decision-making process and how an adequate tool can help humans in such task. However, even though their potential benefit has been shown, the on-site implementation of such tools must address other challenges like software integration and maintenance.
- The suggestions of the cleaning tasks provided by the heat exchanger network DSS have proven that the current policy of *dirtiest fir*st is not the best in all cases. The decisions about maintenance to keep performance from an optimal economic point of view are not an easy task, as they depend a lot on different interconnected factors. Note that, the developed models allow to monitor the fouling effect, but they do not predict the future evolution of the fouling, so the proposed actions may be suboptimal in the long term.

- The study of the different formulations to coordinate the two RTO schemes referred to the evaporation network has proven, on the one hand that solving each scheme independently gets sub-optimal solutions, and on the other hand that the best formulation for the coordination depends on the particular problem. Although the three presented approaches solve the problem in acceptable CPU time for real-time purposes, in this case the centralized formulation has arisen as the most efficient one, getting the best solution (lower cost) and being the quickest. However, this may be not the case for instance considering time dynamics and/or different scenarios of the uncertain conditions, so each case needs to be studied in depth to be addressed in the more convenient way. In the author's opinion, standardized solutions are not able to unlock the true savings potential in complex process plants.
- The hybrid model developed for the re-design of the cooling-system network shows the actual complexity of such problem as there are so many operation and layout alternatives. In addition, not only the payback time must be considered with current conditions, but the probably different future conditions in which the network must be able to work.
- The uncertainty on such future operation conditions for the re-design of the network has been explicitly considered with the two-stage stochastic optimization. This was a case where a monolithic formulation was not practical to solve due to the high amount of CPU time required. This is a usual drawback with stochastic simulation and optimization problems. Thus, the problem was decomposed developing a formulation, tailored to the particular characteristics of the problem, that were exploited after a good previous analysis of the case study.
- The decomposed formulation for process re-design allows to solve each scenario independently. The solution is proven to be obtained in a finite number of iterations, where the resolution time of each iteration only depends on the worst scenario. These are nice features for a decomposition approach, especially when involving non-convex optimization.
- The proposed software architecture presented in Chapter 6 is an example to illustrate a possible industrial implementation of the RTO schemes

as part of DSS. It allows to directly connect the optimization problem to the PI system in order to get the necessary data to compute a solution based on real-time conditions, and to show the results in a suitable interface. Thereby, the common problems of implementing RTO as part of the automatic control system are partially avoided as the loop with the plant is not fully closed. The end user is the one that must take the decisions based on the information provided by the DSS. This also allows to use the proposed tool for other purposes, like operator training or *What-if* analysis.

- Finally, it should be noted that, although the models represent the actual behavior of the different networks in the best case, they must be updated in the future for sure. The technology proposed in this thesis has been proven to be able to get significant benefits if it is correctly used, but if the models do not represent well the systems behavior, the optimal operation point provided by the optimizations will be far from the real one, leading to the progressive abandonment of the DSSs by the operators and plant managers. Hence, an important open issue is model upkeep: monitoring the goodness of model predictions and automatically update models when deviations exceed acceptable thresholds.

7.3 Open research & development lines

This thesis focused on the optimal operation of different networks in real time and considers one of the most common problems of the operation in process industry: the fouling. But as it was concluded before, the developed models do not predict the future evolution of the fouling. Thus, future research can focus on extending the problem formulation to include a prediction of the fouling dynamics over time, hence allowing the computation of a full production-maintenance schedule of the networks over a suitable time horizon. However, it requires a dynamic model to predict the fouling state with time and, possibly, with respect to the flows fed to the surface condensers or the heat exchangers, depending on the network, but in both cases they are decision variables. Furthermore, an estimation of the future operation conditions will be required too, but these estimations are uncertain over a large time horizon, so robust scheduling formulations are foreseen somehow mandatory. In any case, the computational complexity of these MINLP optimization problems will increase exponentially, so decomposition and convexification techniques will be required to cast the problems in a more tractable form.

On the other hand, importance of maintaining the models over time has been remarked. In order to ease such update, they should be formed of simpler pieces where each piece could be monitored and modified easily. Consequently, future work should be developing a components library with models for the individual equipment and a method to provide plug-and-play features to ease the inclusion and modification of the model constraints. Moreover, automatic routines of model-quality monitoring, parameter estimation, as well as constrained regression to update the black-box submodels when consistent plant-model deviations are detected, will be a fundamental pillar to support and generate acceptance of the RTO technologies proposed in this thesis.

References

Abhishek, K., Leyffer, S., & Linderoth, J. (2010). FilMINT: An outer approximation-based solver for convex mixed-integer nonlinear programs. *INFORMS Journal on Computing*, *22*(4), 555–567.

Andersson, J. A. E., Gillis, J., Horn, G., Rawlings, J. B., & Diehl, M. (2019). CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, *11*(1), 1–36.

Ariyur, K. B., & Krstić, M. (2004). Real-Time Optimization by Extremum-Seeking Control. *Real-Time Optimization by Extremum-Seeking Control*.

Aronson, J. E., Liang, T.-P., & MacCarthy, R. V. (2005). *Decision support systems and intelligent systems* (Vol. 4). Pearson Prentice-Hall Upper Saddle River, NJ, USA:

Åström, K. J., Hägglund, T., & Astrom, K. J. (2006). *Advanced PID control* (Vol. 461). Research Triangle Park: ISA-The Instrumentation, Systems, and Automation Society.

Beale, E. M. L., & Tomlin, J. A. (1970). Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables. *OR*, *69*(447–454), 99.

Biegler, L. T. (2010). *Non-Linear Programing: Concepts, Algorithms, and Applications to Chemical Processes* (S. for I. and A. Mathematics., Ed.).

Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

Boccardi, G., Celata, G. P., Lazzarini, R., Saraceno, L., & Trinchieri, R. (2010). Development of a heat transfer correlation for a Scraped-Surface Heat Exchanger. *Applied Thermal Engineering*, *30*(10), 1101–1106.

Bonami, P., Biegler, L. T., Conn, A. R., Cornuéjols, G., Grossmann, I. E., Laird, C. D., Lee, J., Lodi, A., Margot, F., Sawaya, N., & others. (2008). An algorithmic framework for convex mixed integer nonlinear programs. *Discrete Optimization*, *5*(2), 186–204.

Boyd, S., Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Boyd, S., Xiao, L., & Mutapcic, A. (2003). Subgradient methods. *Lecture Notes of EE392o, Stanford University, Autumn Quarter*, *2004*, 2004–2005.

Burer, S., & Letchford, A. N. (2012). Non-convex mixed-integer nonlinear programming: A survey. *Surveys in Operations Research and Management Science*, *17*(2), 97–106.

Camacho, E. F., & Alba, C. B. (2013). *Model predictive control*. Springer Science & Business Media.

Camargo-Vega, J. J., Camargo-Ortega, J. F., & Joyanes-Aguilar, L. (2015). Conociendo big data. *Facultad de Ingeniería*, *24*(38), 63–77.

Carlson, D. A., & Haurie, A. (2013). *Infinite horizon optimal control: theory and applications* (Vol. 290). Springer Science & Business Media.

Cavdur, F., & Sebatli, A. (2019). A decision support tool for allocating temporary-disaster-response facilities. *Decision Support Systems*, *127*, 113145.

Chang, J.-S., Lu, S.-C., & Chiu, Y.-L. (2007). Dynamic modeling of batch polymerization reactors via the hybrid neural-network rate-function approach. *Chemical Engineering Journal*, *130*(1), 19–28.

Cheng, R., Forbes, J. F., & Yip, W. S. (2007). Price-driven coordination method for solving plant-wide MPC problems. *Journal of Process Control*, *17*(5), 429–438.

Chua, K. J., Chou, S. K., & Yang, W. M. (2010). Advances in heat pump systems: A review. *Applied Energy*, *87*(12), 3611–3624.

Conejo, A. J., Castillo, E., Mínguez, R., & García-Bertrand, R. (2006). Decomposition techniques in mathematical programming: Engineering and science applications.

In *Decomposition Techniques in Mathematical Programming: Engineering and Science Applications*.

Cozad, A., Sahinidis, N. V, & Miller, D. C. (2014). Learning surrogate models for simulation-based optimization. *AIChE Journal*, *60*(6), 2211–2227.

Dalle Ave, G., Alici, M., Harjunkoski, I., & Engell, S. (2019). An Explicit Online Resource-Task Network Scheduling Formulation to Avoid Scheduling Nervousness. In A. A. Kiss, E. Zondervan, R. Lakerveld, & L. Özkan (Eds.), *29th European Symposium on Computer Aided Process Engineering* (Vol. 46, pp. 61–66). Elsevier.

Danfoos Engineering Tomorrow. (n.d.). *Coolselector 2*. https://www.danfoss.com/es-es/service-and-support/downloads/dcs/coolselector-2/

Darby, M. L., Nikolaou, M., Jones, J., & Nicholson, D. (2011). RTO: An overview and assessment of current practice. *Journal of Process Control*, *21*, 874–884.

de Gouvêa, M. T., & Odloak, D. (1998). One-layer real time optimization of LPG production in the FCC unit: procedure, advantages and disadvantages. *Computers & Chemical Engineering*, *22*(SUPPL.1), S191–S198.

de Kleijn Energy Consultants & Engineers. (2014). *Industrial Heat Pumps*. http://tools.industrialheatpumps.nl/warmtepompwijzer/EN_index.html.

de Prada, C., Mazaeda, R., & Podar, S. (2018). Optimal operation of a combined continuous–batch process. In *Computer Aided Chemical Engineering* (Vol. 44, pp. 673–678). Elsevier.

de Prada, C., & Pitarch, J. L. (2018). Real-Time Optimization (RTO) Systems. In *Resource Efficiency of Processing Plants* (pp. 265–292). John Wiley & Sons, Ltd.

de Prada, C., Sarabia, D., Gutierrez, G., Gomez, E., Marmol, S., Sola, M., Pascual, C., & Gonzalez, R. (2017). Integration of {RTO} and {MPC} in the Hydrogen Network of a Petrol Refinery. *Processes*, *5*(1).

Dorao, C. A., & Fernandino, M. (2017). Dominant dimensionless groups controlling heat transfer coefficient during flow condensation inside pipes. *International Journal of Heat and Mass Transfer*, *112*, 465–479.

Duran, M. A., & Grossmann, I. E. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming*, *36*(3), 307–339.

Gazoni, E. & Clark, C. (2020). *OpenPyXL 3.0.5*.

https://pypi.org/project/openpyxl/

Engell, S. (2007). Feedback control for optimal process operation. *Journal of Process Control*, *17*(3), 203–219.

Erdoğan, G., Stylianou, N., & Vasilakis, C. (2019). An open source decision support system for facility location analysis. *Decision Support Systems*, *125*, 113116.

Galan, A., de Prada, C., Gutierrez, G., Sarabia, D., & Gonzalez, R. (2021). Real-time reconciled simulation as decision support tool for process operation. *Journal of Process Control*, *100*, 41–64.

Galan, A., Prada, C. de, Gutierrez, G., Sarabia, D., Grossmann, I. E., & Gonzalez, R. (2019). Implementation of {RTO} in a large hydrogen network considering uncertainty. *Optimization and Engineering*, 1–30.

Ganesh, N., & Biegler, L. T. (1987). A reduced hessian strategy for sensitivity analysis of optimal flowsheets. *AIChE Journal*, *33*(2), 282–296.

Geoffrion, A. M. (1972). Generalized benders decomposition. *Journal of Optimization Theory and Applications*, *10*(4), 237–260.

Geoffrion, A. M. (1974). Lagrangean relaxation for integer programming. In *Approaches to integer programming* (pp. 82–114). Springer.

Ghaleb, M., Taghipour, S., Sharifi, M., & Zolfagharinia, H. (2020). Integrated production and maintenance scheduling for a single degrading machine with deterioration-based failures. *Computers & Industrial Engineering*, *143*, 106432.

Glassey, J., & Von Stosch, M. (2018). *Hybrid modeling in process industries*. CRC Press.

Grossmann, I. E., & Harjunkoski, I. (2019). Process Systems Engineering: Academic and Industrial Perspectives. *Computers & Chemical Engineering*, *129*, 474–484.

Gupta, O. K., & Ravindran, A. (1985). Branch and bound experiments in convex nonlinear integer programming. *Management Science*, *31*(12), 1533–1546.

Gupta, S., Liu, P.-H., Svoronos, S. A., Sharma, R., Abdel-Khalek, N. A., Cheng, Y., & El-Shall, H. (1999). Hybrid first-principles/neural networks model for column flotation. *AIChE Journal*, *45*(3), 557–566.

Gurobi Optimization, L. L. C. (2018). *Gurobi Optimizer Reference Manual*.

Haag, S., & Cummings, M. (2009). *Management information systems for the information age*. McGraw-Hill, Inc.

Han, J., Forman, G. S., Elgowainy, A., Cai, H., Wang, M., & DiVita, V. B. (2015). A comparative assessment of resource efficiency in petroleum refining. *Fuel*, *157*, 292–298.

Harjunkoski, I., Maravelias, C. T., Bongers, P., Castro, P. M., Engell, S., Grossmann, I. E., Hooker, J., Méndez, C., Sand, G., & Wassick, J. (2014). Scope for industrial applications of production scheduling models and solution methods. *Computers & Chemical Engineering*, *62*, 161–193.

Hart, W. E., Laird, C. D., Watson, J.-P., Woodruff, D. L., Hackebeil, G. A., Nicholson, B. L., & Siirola, J. D. (2017). *Pyomo--optimization modeling in python* (Second, Vol. 67). Springer Science & Business Media.

Hart, W. E., Watson, J.-P., & Woodruff, D. L. (2011). Pyomo: modeling and solving mathematical programs in Python. *Mathematical Programming Computation*, *3*(3), 219–260.

Hernandez, R., Dreimann, J., Vorholt, A., Behr, A., & Engell, S. (2018). Iterative real-time optimization scheme for optimal operation of chemical processes under uncertainty: Proof of concept in a miniplant. *Industrial & Engineering Chemistry Research*, *57*(26), 8750–8770.

Huber, P. J. (1981). *Robust statistics* (Vol. 523). John Wiley & Sons.

Hurvich, C. M., & Tsai, C. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis*, *14*(3), 271–279.

ISA, A. (2000). ISA-95.00. 01-2000: Enterprise-control system integration part 1: Models and terminology. *Technical Report, ISA, The Instrumentation, Systems, and Automation Society*.

Kalliski, M., Pitarch, J. L., Jasch, C., & Prada, C. de. (2019). Support to decision-making in a network of industrial evaporators. *Revista Iberoamericana de Automática e Informática Industrial*, *16*(1), 26–35.

Kandakoglu, A., Sauré, A., Michalowski, W., Aquino, M., Graham, J., & McCormick, B. (2019). A decision support system for home dialysis visit scheduling and nurse routing. *Decision Support Systems*, 113224.

Kelley James E, J. (1960). The cutting-plane method for solving convex programs. *Journal of the Society for Industrial and Applied Mathematics*, *8*(4), 703–712.

Khor, C. S., & Varvarezos, D. (2017). Petroleum refinery optimization. *Optimization and Engineering*, *18*(4), 943–989.

Kim, S.-J., Koh, K., Lustig, M., Boyd, S., & Gorinevsky, D. (2007). An interior-point method for large-scale $\ell_1 $-regularized least squares. *IEEE Journal of Selected Topics in Signal Processing*, *1*(4), 606–617.

Krämer, S., & Engell, S. (2017). *Resource Efficiency of Processing Plants: Monitoring and Improvement*. John Wiley & Sons.

Kujanpää, M., Hakala, J., Pajula, T., Beisheim, B., Krämer, S., Ackerschott, D., Kalliski, M., Engell, S., Enste, U., & Perez, J. L. P. (2017). Successful Resource Efficiency Indicators for process industries: Step-by-step guidebook. *Clinical Epigenetics*.

Lamnabhi-Lagarrigue, F., Annaswamy, A., Engell, S., Isaksson, A., Khargonekar, P., Murray, R. M., & Hof, P. (2017). Van den Systems & Control for the future of humanity, research agenda: Current and future roles, impact and grand challenges. *Annual Reviews in Control*, *43*, 1–64.

Lee, S., & Grossmann, I. E. (2001). A global optimization algorithm for nonconvex generalized disjunctive programming and applications to process systems. *Computers & Chemical Engineering*, *25*(11–12), 1675–1697.

Leibman, M. J., Edgar, T. F., & Lasdon, L. S. (1992). Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques. *Computers & Chemical Engineering*, *16*(10), 963–986.

Lenzing. (2019). *D2.5 Final report on equipment degradation modelling*. www.spire2030.eu/copro

Leyffer, S. (2001). Integrating SQP and branch-and-bound for mixed integer nonlinear programming. *Computational Optimization and Applications*, *18*(3), 295–309.

Leyffer, S., Sartenaer, A., & Wanufelle, E. (2008). Branch-and-refine for mixed-integer nonconvex global optimization. *Preprint ANL/MCS-P1547-0908, Mathematics and Computer Science Division, Argonne National Laboratory*, *39*, 40–78.

Liong, S.-Y., & Atiquzzaman, M. (2004). Optimal design of water distribution network using shuffled complex evolution. *Journal of the Institution of Engineers, Singapore*, *44*(1), 93–107.

Lozano, S., Moreno, P., Adenso-Díaz, B., & Algaba, E. (2013). Cooperative game theory approach to allocating benefits of horizontal cooperation. *European Journal of Operational Research*, *229*(2), 444–452.

Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, *6*, 1–10.

Marchetti, A., Chachuat, B., & Bonvin, D. (2009). Modifier-Adaptation Methodology for Real-Time Optimization. *Industrial and Engineering Chemistry Research*, *48*(13), 6022–6033.

Marcos, M. P., Pitarch, J. L., Jasch, C., & Prada, C. de. (2018). Modeling for optimal operation of an industrial cooling system. *XXXIX Jornadas de Automática*, 702–709.

Marcos, M. P., Pitarch, J. L., Jasch, C., & Prada, C. de. (2019). Optimal distribution load allocation and resource utilisation in evaporation plants. *Computer Aided Chemical Engineering*.

Marcos, M. P., Pitarch, J. L., & Prada, C. de. (2019). *Real-time optimisation for a heat-recovery section with equipment degradation*. 513–519.

Marcos, M. P., Pitarch, J. L., & Prada, C. de. (2020a). Modelling and real-time optimisation of a heat-exchanger network. *IFAC-PapersOnLine*, *53*(2), 11780–11785.

Marcos, M. P., Pitarch, J. L., & Prada, C. de. (2020b). Decision support system for a heat-recovery section with equipment degradation. *Decision Support Systems*, *137*.

Marcos, M. P., Pitarch, J. L., Prada, C. de, & Jasch, C. (2018). Modelling and real-time optimisation of an industrial cooling-water network. *2018 22nd International Conference on System Theory, Control and Computing (ICSTCC)*, 591–596.

Marcos, M. P., Pitarch, J. L., & Prada, C. de. (2021). Integrated Process Re-Design with Operation in the Digital Era: Illustration through an Industrial Case Study. *Processes*, *9*(7), 1203.

Martí, R. (2015). *Price Coordination Strategies in Large-Scale Process Control*. Escuela de Ingenierías Industriales, Universidad de Valladolid.

Merino, A., Acebes, L. F., Alves, R., & de Prada, C. (2018). Real Time Optimization for steam management in an evaporation section. *Control Engineering Practice*, *79*, 91–104.

Nannicini, G., & Belotti, P. (2012). Rounding-based heuristics for nonconvex MINLPs. *Mathematical Programming Computation*, *4*(1), 1–31.

Neumaier, A. (1998). Solving ill-conditioned and singular linear systems: A tutorial on regularization. *SIAM Review*, *40*(3), 636–666.

Nicholson, B., López-Negrete, R., & Biegler, L. T. (2014). On-line state estimation of nonlinear dynamic systems with gross errors. *Computers & Chemical Engineering*, *70*, 149–159.

Nowak, I., & Vigerske, S. (2008). LaGO: a (heuristic) branch and cut algorithm for nonconvex MINLPs. *Central European Journal of Operations Research*, *16*(2), 127–138.

Ochsner Energie Technik. (n.d) Retrieved July 27, 2020 https://ochsner-energietechnik.com/hoechsttemperatur-waermepumpen/

Oliveira, R. (2004). Combining first principles modelling and artificial neural networks: a general framework. *Computers & Chemical Engineering*, *28*(5), 755–766.

OSIsoft. (2020). *{OSIsoft PI} System*. www.osisoft.com/pi-system/

Palacín, C. G., Pitarch, J. L., Jasch, C., Méndez, C. A., & Prada, C. de. (2018). Robust integrated production-maintenance scheduling for an evaporation network. *Computers & Chemical Eng.*, *110*, 140–151.

PI System - Connecting data, operations & people | OSIsoft. (n.d.). Retrieved December 5, 2021, from https://www.osisoft.com/pi-system/

PIconnect · PyPI. (n.d.). Retrieved August 20, 2021, from https://pypi.org/project/PIconnect/

Pitarch, J. L., Palacín, C. G., de Prada, C., Voglauer, B., & Seyfriedsberger, G. (2017). Optimisation of the Resource Efficiency in an Industrial Evaporation System. *Journal of Process Control*, *56*, 1–12.

Pitarch, J. L., Sala, A., & de Prada, C. (2019a). A Systematic Grey-Box Modeling Methodology via Data Reconciliation and {SOS} Constrained Regression. *Processes*, *7*(3).

Pitarch, J. L., Sala, A., & de Prada, C. (2019b). A Sum-Of-Squares Constrained Regression Approach for Process Modeling. *IFAC-PapersOnLine*, *52*(1), 754–759.

Power, D. J. (2002). *Decision support systems: concepts and resources for managers*. Greenwood Publishing Group.

Prada, C., Villar, E., & Aleixandre, V. (1987). Dynamic Simulation of a Multiple-effect Evaporation Station. *IFAC Proceedings Volumes*, *20*(12), 129–132.

Qian, F., Zhong, W., & Du, W. (2017). Fundamental Theories and Key Technologies for Smart and Optimal Manufacturing in the Process Industry. *Engineering*, *3*(2), 154–160.

Quesada, I., & Grossmann, I. E. (1992). An LP/NLP based branch and bound algorithm for convex MINLP optimization problems. *Computers & Chemical Engineering*, *16*(10–11), 937–947.

Rashidi, M., Ghodrat, M., Samali, B., & Mohammadi, M. (2018). Decision Support Systems. In M. Pomffyova (Ed.), *Management of Information Systems*. IntechOpen.

Roberts, P. D. (1979). An algorithm for steady-state system optimization and parameter estimation. *International Journal of Systems Science*, *10*(7), 719–734.

Rockafellar, R. T., & Wets, R. J.-B. (1991). Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, *16*(1), 119–147.

Rojko, A. (2017). Industry 4. *0 Concept Background and Overview International Journal of Interactive Mobile Technologies*, *11*(5), 77–90.

Ryoo, H. S., & Sahinidis, N. V. (1995). Global optimization of nonconvex NLPs and MINLPs with applications in process design. *Computers & Chemical Engineering*, *19*(5), 551–566.

Ryoo, H. S., & Sahinidis, N. V. (1996). A branch-and-reduce approach to global optimization. *Journal of Global Optimization*, *8*(2), 107–138.

Sahinidis, N. v. (1996). BARON: A general purpose global optimization software package. *Journal of Global Optimization*, *8*(2), 201–205.

Santos, A. I. G., Cosme, J. M. Z., & de Prada Moraga, C. (2001). Nonlinear model predictive control in a batch fermentator with state estimation. *2001 European Control Conference, ECC 2001*, 426–431.

Santos, P., Pitarch, J. L., Vicente, A., de Prada, C., & Garcia, A. (2020). Improving operation in an industrial MDF flash dryer through physics-based NMPC. *Control Engineering Practice 104213, 94.*

Sarabia, D., de Prada, C., Gómez, E., Gutierrez, G., Cristea, S., Sola, M., & Gonzalez, R. (2012). Data reconciliation and optimal management of hydrogen networks in a petrol refinery. *Control Engineering Practice*, *20*(4), 343–354.

Saraceno, A., Curcio, S., Calabrò, V., & Iorio, G. (2010). A hybrid neural approach to model batch fermentation of “ricotta cheese whey” to ethanol. *Computers & Chemical Engineering*, *34*(10), 1590–1596.

Sawaya, N. W., & Grossmann, I. E. (2005). A cutting plane method for solving linear generalized disjunctive programming problems. *Computers & Chemical Engineering*, *29*(9), 1891–1913.

Schlüter, M., Egea, J. A., & Banga, J. R. (2009). Extended ant colony optimization for non-convex mixed integer nonlinear programming. *Computers & Operations Research*, *36*(7), 2217–2229.

Skogestad, S. (2000). Plantwide control: the search for the self-optimizing control structure. *Journal of Process Control*, *10*(5), 487–507.

Skogestad, S., & Postlethwaite, I. (2007). *Multivariable feedback control: analysis and design* (Vol. 2). Wiley New York.

Smith, E. M. B., & Pantelides, C. C. (1997). Global optimisation of nonconvex MINLPs. *Computers & Chemical Engineering*, *21*, S791–S796.

Steimel, J., & Engell, S. (2015). Conceptual design and optimization of chemical processes under uncertainty by two-stage programming. *Computers & Chemical Engineering*, *81*, 200–217.

Stubbs, R. A., & Mehrotra, S. (1999). A branch-and-cut method for 0-1 mixed convex programming. *Mathematical Programming*, *86*(3), 515–532.

Te Braake, H. A. B., Van Can, H. J. L., & Verbruggen, H. B. (1998). Semi-mechanistic modeling of chemical processes with neural networks. *Engineering Applications of Artificial Intelligence*, *11*(4), 507–515.

Thibault, J., Acuña, G., Pérez-Correa, R., Jorquera, H., Molin, P., & Agosin, E. (2000). A hybrid representation approach for modelling complex dynamic bioprocesses. *Bioprocess Engineering*, *22*(6), 547–556.

Tholudur, A., & Ramirez, W. F. (1999). Neural-network modeling and optimization of induced foreign protein production. *AIChE Journal*, *45*(8), 1660–1670.

Tholudur, A., Ramirez, W. F., & McMillan, J. D. (2000). Interpolated parameter functions for neural network models. *Computers & Chemical Engineering*, *24*(11), 2545–2553.

Thompson, M. L., & Kramer, M. A. (1994). Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal*, *40*(8), 1328–1340.

Van Can, H. J. L., Te Braake, H. A. B., Dubbclman, S., Hellinga, C., Luyben, K. Ch. . A. M., & Heijnen, J. J. (1998). Understanding and applying the extrapolation properties of serial gray-box models. *AIChE Journal*, *44*(5), 1071–1089.

Viswanathan, J., & Grossmann, I. E. (1990). A combined penalty function and outer-approximation method for MINLP optimization. *Computers & Chemical Engineering*, *14*(7), 769–782.

Von Stosch, M., Oliveira, R., Peres, J., & de Azevedo, S. F. (2014). Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers & Chemical Engineering*, *60*, 86–101.

Wächter, A., & Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, *106*(1), 25–57.

Wächter, A., Margot, F., Bonami, P., Vigerske, S., Berthold, T., Biegler, Larry T., Cornuéjols, G., & Grossmann, I. E. (n.d.). *Couenne*.

Wang, X., Mahalec, V., Li, Z., & Qian, F. (2017). Real-time optimization and control of an industrial ethylbenzene dehydrogenation process. *Chemical Engineering Transactions*, *61*, 331–336.

Wang, X., Wang, S., & Xu, Q. (2022). Simultaneous Production and Maintenance Scheduling for Refinery Front-End Process with Considerations of Risk Management and Resource Availability. *Industrial & Engineering Chemistry Research*, *61*(5), 2152–2166.

Würth, L., Hannemann, R., & Marquardt, W. (2009). Neighboring-extremal updates for nonlinear model-predictive control and dynamic real-time optimization. *Journal of Process Control*, *19*(8), 1277–1288.

Yiqing, L., Xigang, Y., & Yongjian, L. (2007). An improved PSO algorithm for solving non-convex NLP/MINLP problems with equality constraints. *Computers & Chemical Engineering*, *31*(3), 153–162.