A Novel Matheuristic based on Bi-Level Optimization for the Multi-Objective Design of Hydrogen Supply Chains

Victor H. Cantú[a], Catherine Azzaro-Pantel[a,*], Antonin Ponsich[b]

[a] *Laboratoire de Génie Chimique, UMR 5503 CNRS/INP/UPS, Université de Toulouse, Toulouse, France*

[b] *Departamento de Sistemas, Universidad Autónoma Metropolitana Azcapotzalco, Mexico City, Mexico*

Abstract

This work introduces an efficient tool for the design of sustainable hydrogen supply chains (HSCs), considering both economic and environmental concerns, through an appropriate multi-objective strategy. The original problem, being formulated as a bi-objective mixed-integer linear programming (MILP) problem, takes into consideration the availability of different energy sources, the installation and operation of hydrogen facilities of different sizes and technologies, and the transportation of hydrogen from production units to storage facilities. The area of study is divided into grids which have a specific hydrogen demand that evolves over time, thus a multi-period model of the HSC is considered. In order to overcome the computational burden associated to the solution of large size instances of the resulting problem, we proposed a solution strategy consisting of a hybrid algorithm. The original problem is reformulated into a bi-level optimization problem: the upper level (discrete problem) consists of finding the optimal location for production plants and storage facilities, whereas the lower level (continuous problem) minimizes their corresponding costs associated to transportation and facility operation. A multi-objective evolutionary algorithm is employed for the solution of the bi-objective upper level, whereas the bi-objective lower level is decomposed using a scalarizing function, which is then solved using a linear programming solver. The proposed methodology is validated through the comparison of the true Pareto fronts given by CPLEX with ε-constraint method, for six increasing size instances. Numerical results prove that the proposed hybrid approach produces an accurate approximation of the Pareto-optimal fronts, more efficiently than the exact solution approach.

*Corresponding author

Email address: catherine.azzaropantel@toulouse-inp.fr (Catherine Azzaro-Pantel)

1. Introduction

The growing general concern about the depletion of conventional energy sources, such as oil and gas, as well as the degradation of the environment caused by the combustion of these fossil fuels, has motivated the search for a more sustainable energy model based on renewable energy systems. In this context, hydrogen stands as a potential low-carbon alternative since (1) it can be used in multiple sectors such as industry, building and transportation, providing meaningful reductions of CO_2 emissions, (2) it can be produced from a variety of fossil fuels with captured CO_2 in Carbon Capture and Utilization (CUU) processes, but also from renewable sources like biomass, wind or solar photovoltaic, and (3) can thus store surplus power from renewable sources when the grid cannot absorb it (IEA, 2019). According to the Hydrogen Council (McKinsey et al., 2017), hydrogen is expected to play a significant role in low-carbon energy landscape, covering 18% of global energy demand by 2050. These long-term estimations are in part established considering the government policies to limit the global greenhouse gases (GHG) emissions, in line with the 2015 Paris Climate Agreement.

In 2019 the European Green Deal set up a framework of regulation and legislation with targets to reach net zero global warming emissions by 2050 (European Comission, 2019). Hydrogen is considered an important instrument for meeting the Green Deal objectives. Energy storage is often viewed as an electrical power storage through mechanical, electrical and electrochemical storage systems. In the current energy system, grid-scale energy storage is typically short-term and used to maintain stability, in order to address peaks (i.e., on the minute and hour scale) up to daily imbalances. Seasonal storage may be needed in the future for high levels of renewable penetration based mainly on solar and wind generation, and can be achieved at terawatt (TW) level through hydrogen or synthetic methane (Davies et al., 2020). This is why hydrogen is even mentioned as the best option to store electrical energy, even better than pumped hydro, compressed air or batteries. Beyond its potential role in providing chemical storage of electricity, hydrogen can also act as an energy carrier for energy and industrial applications where it is difficult to replace fossil fuels, thus contributing to the decarbonisation of transport, buildings or industry. Hydrogen can thus be viewed as hydrogen as a "coupling sector" technology (Brey, 2020).

In this respect, the transportation sector is of high importance since it is responsible for 26% of GHG emissions in EU, 28% in U.S., and 23% worldwide in recent years (European Environment Agency, 2018; Sims et al., 2014), which is in part explained by the fact that this sector relies almost completely on oil (McKinsey et al., 2017). Consequently, several technologies have been proposed with the aim to decarbonize this sector, and particularly, road transport sector, including hybrid electric vehicles, battery

electric vehicles (BEVs) and fuel cell electric vehicles (FCEVs). Concerning BEVs, on the one hand, they are at present in the market mainly because the electric grid is already available in most areas where cars typically need to be charged, nevertheless, they present primarily two inconveniences, namely, high charging times, and for medium-to-large size vehicles, the need of heavy batteries. On the other hand, it is widely recognized that FCEVs are a necessary complement to BEVs, as FCEVs typically permit users longer ranges and fast fueling times, in comparison to BEVs (Lin et al., 2018), with the drawback that little or none fueling stations are yet available in most areas. Several studies have been conducted showing that BEV and FECV can provide climate benefits, though results depend strongly on several factors including the electrical mix used for battery charging and hydrogen production, the lifetime distance traveled by the vehicle, and the vehicle energy consumption. A Life Cycle Assessment is thus necessary to have a global overview of all the steps involved from well-to-wheel (for instance battery manufacturing and recycling for BEV and Polymer Electrolyte Membrane (PEM) fuel cell for FCEV) (see Cox et al. (2020)).

Hence, even if the potential environmental and technological benefits of hydrogen in the transport sector are encouraging, the change to a hydrogen-based economy is still a challenge; much of the future expansion of hydrogen utilization depends not only on technological developments and energy policies, but also, on the hydrogen supply chain (HSC) deployment. Such a supply chain has to take into consideration the different sources of energy from which hydrogen can be produced, the different production and storage technologies available, the location of hydrogen plants and storage facilities, their respective production and storage capacity sizes, and the mode and rate of transportation between production units and storage facilities in the network. Besides, the design of a cost-efficient hydrogen infrastructure has to consider the varying demand over time, so that a multi-period model of the problem needs to be employed. Due to the aforementioned characteristics, the HSC design problem is often formulated as a difficult optimization problem (NP-hard problem), involving both discrete and continuous variables, accounting for the existence and location of production and storage facilities of a specific capacity, and for the hydrogen flow rates from production to storage units in order to satisfy a given demand. Moreover, the multi-objective nature of the problem needs also to be addressed, so that not only cost minimization but also environmentally-friendly production modes need to be accounted for. Therefore, the optimal design of HSC is not a trivial task.

Most of the multi-objective models proposed in the literature are solved using deterministic or exact techniques, namely ε-constraint or by means of utility functions, e.g., weighted sum. It is noteworthy that these methods can guarantee the optimality of the

solutions found, and even more, some of them can find any Pareto optimal solution by using appropriate parameters, however, they present two main disadvantages, (1) multiple runs must be performed in order to obtain a set of trade-off solutions, which might be computationally prohibitive and (2) the obtained approximation of the Pareto front is not necessarily uniformly distributed, and thus some regions in the Pareto front might be not adequately explored. Therefore, this work aims to present an appropriate solution approach to the multi-objective HSC problem, able to efficiently provide the decision maker with a set of well-distributed non-dominated solutions along the Pareto optimal front. To this end, the original problem is reformulated as a bi-level multi-objective optimization problem, through a master-slave decomposition strategy. In particular, the master sub-problem deals with the facility installation problem, while the slave sub-problem addresses the transport problem corresponding to each structure proposed by the upper-level processes. Even if this decomposition strategy does not reduce the complexity of the original problem, it does enable the efficient solution of each sub-problem by appropriate solution methods, and thus that of the original problem.

The core innovation and contributions of the present work are described in more details as follows:

1. A novel methodology is designed for the solution of the HSC design problem, in which the mathematical structure of the problem can be exploited. In this manner, the multi-objective, combinatorial and linear aspects of the problem can be tackled by appropriate solution methods.
2. Consequently, the original problem is reformulated as a bi-level multi-objective optimization problem, through a master-slave decomposition strategy. The upper-level (master) problem considers the installation of production and storage facilities (multi-objective combinatorial problem), while the lower level examines the problem associated to transportation and production rates (linear programming problem of low complexity).
3. A hybrid solution tool based on both Evolutionary Algorithms and Linear Programming (matheuristic) is developed for the solution of the resulting bi-level optimization problem, able to efficiently achieve a good approximation of the Pareto front in terms of convergence, distribution and number of Pareto solutions. More precisely, the solution of the master problem is performed by a multi-objective evolutionary algorithm (MOEA), taking full advantage of the multi-objective and combinatorial features of the problem. Then, for each individual (partial solution) proposed by the evolutionary algorithm, the slave problem is treated by means of a linear programming (LP) solver. In this manner, the multi-objective bi-level problem is solved in an iteratively manner, in one single run.

4. Finally, the performance of the proposed methodology is evaluated over six increasing size instances of the HSC problem inspired from Almaraz et al. (2014). To this end, the performance of both deterministic and hybrid approaches is presented in terms of the hypervolume indicator, for the first time in the HSC literature. Results according to this indicator show that the proposed hybrid approach outperforms the classical one, i.e., the set of non-dominated solutions obtained by the hybrid approach are better distributed along the Pareto front, offering the decision makers a better picture of all the possible trade-off solutions. Moreover, the proposed methodology can provide approximation sets with a significant number of efficient solutions, in reasonable computational times, thanks to the use of MOEAs.

The remainder of this work is organized as follows. Section 2 provides a literature review of related works highlighting the solution approach employed in each one. In Section 3, we present a description of the problem studied here, along with the main characteristics and assumptions of the model, as well as the mathematical equations governing its formulation (Subsections 3.1 and 3.2, respectively). Then, in Section 4, the proposed methodology for solving the HSC problem is presented, including the bi-level reformulation, the main aspects of the hybrid solution strategy and some specific features of the tools proposed at each level. This methodology is validated in Section 5 through the study of six instances of the HSC of growing size comparing the results with those obtained using a classical approach. Finally, conclusion and perspectives for future work are drawn in Section 6.

2. Literature review

Several approaches have been proposed for the design of hydrogen supply chain in multiple works, each one presenting different assumptions and characteristics of the supply chain.

One of the first studies proposing mathematical modeling tools in this area is found in van den Heever and Grossmann (2003). This work focused on the integration of production planning and reactive scheduling for the optimization of a refinery hydrogen network; it addressed only the operational level of an existing network, so that the hydrogen supply chain design was out of the scope. The problem was formulated as a multi-period mixed-integer nonlinear program (MINLP) and solved through exact techniques (DICOPT++).

Later, in a pioneering work Hugo et al. (2005), authors proposed a generic mathematical model for the HSC, which includes different primary energy feedstocks, different production and storage technologies, distribution types, potential sites for location of

production units, and the dynamic change of hydrogen demand over time. The resulting mixed integer linear programming (MILP) model considered two conflicting objectives (economic and environmental). The authors do not specify which solution approach they employed, but it is presumably an exact one.

In Almansoori and Shah (2006), authors provided a formal model encompassing every echelon of the supply chain with an illustrative case study in UK. Initially, the model only considered a constant deterministic demand with the minimization of the total cost of the hydrogen infrastructure as the objective function. Later, in Almansoori and Shah (2009), the original model was modified to involve multiple periods, each one with a different hydrogen demand, and then, uncertainty on demand was taken into consideration in Almansoori and Shah (2012). Besides, this model has been extended in several works to consider multiple objectives simultaneously, namely, the greenhouse gas emissions (GHG) and safety, with cases of study in Germany (Almansoori and Betancourt-Torcat, 2016), France (Almaraz et al., 2015), Korea (Kim and Moon, 2008) and Portugal (Câmara et al., 2019). It must be highlighted that, regarding the handling of multiple objectives, the solution approach adopted in all these works is always the same, i.e., ε-constraint method, and so, because of the combinatorial aspect of the problem, large-scale instances or accurate Pareto front approximations are neglected due to prohibitive computational times.

With the intention of alleviating the numerical difficulties associated with the solution of multi-objective large-scale instances, two remarkable strategies based on model decomposition have been proposed in Guillén-Gosálbez et al. (2010) and Sabio et al. (2010). In Guillén-Gosálbez et al. (2010), the authors minimize the total cost and an environmental criterion according to the principles of life cycle assessment, related to the supply chain. The original problem is reformulated into a bi-level (master-slave) optimization problem. In the master problem, the integer variables representing the existence of production plants and storage facilities are removed, those associated to the number of transportation units are relaxed (considered as continuous), and auxiliary binary variables are added to represent the selection of each type of production unit, storage facility and transportation mode. The master problem provides as an output the type of production, storage and transport technologies that should be used in the supply chain (it does not indicate their optimal values, but only if they will be used or not). Concerning the lower level, it is represented by the original MILP model, but this latter is solved only for the subset of technologies predicted at the upper level. In this manner, the master and slave problems are solved iteratively until the bounds of each sub-problem converge within a specified tolerance. It is important to note that the multi-objective aspect of the original problem remains in both master and slave sub-problems, tackled

via ε-constraint method in both levels. So, in order to speed up the solution process, authors propose the use of integer and logic cuts in the upper level. In another notable work (Sabio et al., 2010), the authors considered the total cost and financial risks of the HSC as the two objectives to be minimized. They also proposed a master-slave decomposition approach, based on the assumption that, in practical problems, the continuous relaxation of the integer variables of the full-space model provides tight lower bounds. Consequently, the master sub-problem solves the relaxed original model by reformulating all the integer variables as continuous, while the slave problem solves the original problem but considers the relaxed variables in the master problem as parameters once they have been rounded to the next integer. This decomposition strategy allows obtaining lower and upper bounds of the original problem efficiently. It must be emphasized that both master and slave problems are still modeled as bi-objective MILP problems that exhibit a reduced complexity in comparison to the original model. The multi-objective aspect in both sub-problems is addressed again using the ε-constraint method.

More recently, in Woo and Kim (2019) the authors employed genetic algorithms coupled with exact techniques to solve the optimal design of the HSC with replenishment cycles, modeled as a mixed-integer nonlinear programming problem (MINLP). A two-level approach is proposed as the solution strategy, the upper-level is solved by a binary-coded genetic algorithm that handles some variables, in such a way that the lower level solves a linearized model as a result of considering the upper level variables as parameters. For solving the MILP problem in the lower level, the CPLEX solver is employed. However, authors only addressed a single-objective problem (economic criterion) and several characteristics of the supply chain are neglected, e.g., availability of different energy feed-stock and facility sizes.

Approaches similar to that introduced in Woo and Kim (2019) have been carried out in other areas of engineering. On the one hand, the integration of biomass technologies is investigated in a systematic way in Fazlollahi and Maréchal (2013), taking into account multi-period and multi-criteria features of the problem. The resulting model is of MINLP type and considers three objectives to be minimized: the annual investment cost, the operating cost and the overall CO_2 emissions of the system. The solution methodology proposes decomposing the original problem according to a master-slave structure. The upper level is solved using a dominance-based MOEA which considers the type of district conversion technologies as well as their corresponding maximum sizes (continuous and discrete variables). The lower level, formulated as a MILP problem, optimizes the utilization rate and the operation strategy of district conversion technologies; it is solved using a branch and bound technique. In addition, in Setak et al. (2019), a three-echelon generic supply chain is modeled considering manufacturing plants, distribution centers

and multiple retailers with an application dedicated to a pharmaceutical company. The problem, modeled as a Stackelberg game, is formulated as a mixed-integer nonlinear bi-level optimization problem with an objective targeted on distribution centers in the upper level, and one devoted to retailers in the lower level. The authors proposed the use of a hybrid algorithm as the solution method as follows: the upper level is solved by a genetic algorithm, which handles all integer and some continuous variables, whereas the lower level solves a quadratic programming problem for each upper-level candidate solution using a deterministic local optimizer.

Summarizing, several works have been proposed in the literature with the aim of mitigating the computational difficulties (prohibitive CPU times) highlighted when solving medium to large-size instances of the multi-objective HSC design problem by a classical approach. These works propose to transform the original problem using a master-slave decomposition, nevertheless, the resulting sub-problems are still of MILP type and both contain the same number of objectives as the original problem. Then, some works employed ε-constraint method for solving the multi-objective sub-problems, however, if more objectives or larger size instances are to be considered, the computational times might still be prohibitive to obtain an accurate approximation of the Pareto front. Other works, in other areas of application, have proposed the use of metaheuristic techniques for constructing the Pareto front approximation in the master problem, and exact methods for solving the MILP lower problem. Therefore, we propose a hybrid methodology similar to that of Fazlollahi and Maréchal (2013) as a solution method for the HSC problem. However, our work differentiates from all others because (1) the lower-level sub-problems is of low computational complexity, since only continuous variables are present (linear programming problems), and can therefore be efficiently solved by exact solvers; (2) the multi-objective aspect in the lower level is tackled using a utility function that is rigorous to non-convex problems and, (3) the upper-level is solved by a performance-based MOEA, particularly suited for solving bi-criterion problems.

3. Problem statement

The HSC design model considered in this work is taken from Almansoori and Shah (2009) and its extension from Almaraz et al. (2014). This pioneering model permits a good representation of the HSC based on a time-growing demand over a geographical area of study.

3.1. Description of the problem

The HSC design problem considered consists of a four-echelon supply chain for transport sector (energy source–production–transportation–storage). It minimizes the total

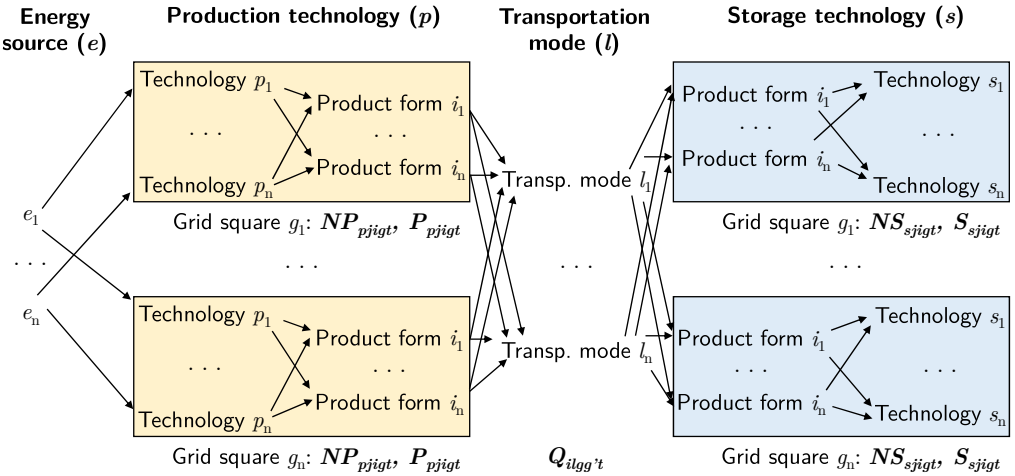

Figure 1: Superstructure of the hydrogen supply chain.

daily cost and the global warming potential simultaneously, that is, the objective is to provide the decision-makers with a number of different trade-off configurations between the cost efficiency of the supply chain and the CO_2 emissions related to it. The superstructure describing all options in the model is represented in Figure 1. The network considers several energy sources as a feedstock to different production technologies, each production technology having different possible sizes with corresponding capacity bounds, capital costs and product unit costs. A set of options is available for transportation from production sites to storage facilities. Before and after transportation, it can be necessary to condition hydrogen to a suitable physical form for transportation and then, for storage. Depending on this final physical form, several storage technologies are available with different capacity sizes. Besides, the whole system is designed to provide hydrogen supply over a given territory, which is divided into grids (nodes). Each grid has a hydrogen demand for a given period of time (input data), and serve as a node for potential production and storage locations.

The model is established under the following assumptions:

1. The model is demand-driven and operates at steady-state conditions, contrary to a more realistic dynamic supply-demand model. This clearly implies that the optimal configuration of the network is the one that minimizes the objectives, subject to demand satisfaction in each grid for each time period.
2. Only some sizes of fixed capacities for production and storage facilities are available.

The HSC problem can therefore be stated as: given a geographical area divided into grids, each grid with an explicit hydrogen demand at each period that must be satis-

fied, investment and operational cost for production and storage facilities as well as their capacity constraints, costs and availability of energy sources, transportation costs and environmental information related to the network operation and installation; the objective is to determine the structure and transportation flows within the hydrogen supply chain that minimizes its total costs and environmental impact. Such a supply chain is determined by the location, type, capacity and number of production and storage facilities in each grid (nodes); the transportation links (directed arcs) between grids, as well as the type and number of transportation units; production rates for each production plant as well as inventory amounts for storage units.

3.2. Mathematical model

As mentioned previously, the mathematical model used as a basis in this work is taken from Almansoori and Shah (2009) and Almaraz et al. (2014). In Almansoori and Shah (2009), the total daily cost (TDC) of the supply chain is considered as the objective to be minimized, including appropriate equations accounting for investment costs related to plant installation and transportation routes, operational costs for production, storage and transportation, and also constraints for plant capacity, mass balance between grids and demand satisfaction. In Almaraz et al. (2014), the model is adapted to consider the global warming potential (GWP) as a second objective. To this end, authors proposed mathematical relations for computing gas emissions due to production, storage and transportation.

Therefore, the bi-objective HSC optimization problem can be represented mathemat-

ically as:

$$\min_{\mathbf{x}} \quad [TDC(\mathbf{x}), GWP(\mathbf{x})]^T \tag{1}$$

$$\text{s.t.} \quad \sum_{p\in P}\sum_{j\in J}\sum_{g\in G} PCap^{\min}_{pji} NP_{pjigt} - \sum_{g\in G} D_{igt} \leq 0 \qquad \forall i \in I,\, \forall t \in T \tag{2}$$

$$\sum_{g\in G} D_{igt} - \sum_{p\in P}\sum_{j\in J}\sum_{g\in G} PCap^{\max}_{pji} NP_{pjigt} \leq 0 \qquad \forall i \in I,\, \forall t \in T \tag{3}$$

$$\sum_{s\in S}\sum_{j\in J} SCap^{\min}_{sji} NS_{sjigt} - S^T_{igt} \leq 0 \qquad \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{4}$$

$$S^T_{igt} - \sum_{s\in S}\sum_{j\in J} SCap^{\max}_{sji} NS_{sjigt} \leq 0 \qquad \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{5}$$

$$PCap^{\min}_{pji} NP_{pjigt} - P_{pjigt} \leq 0 \qquad \forall p \in P,\, \forall j \in J,\, \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{6}$$

$$P_{pjigt} - PCap^{\max}_{pji} NP_{pjigt} \leq 0 \qquad \forall p \in P,\, \forall j \in J,\, \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{7}$$

$$\sum_{p\in P}\sum_{j\in J} P_{pjigt} - \sum_{l\in L}\sum_{g'\in G, g'\neq g} (Q_{ilgg't} - Q_{ilg'gt}) - D_{igt} = 0 \qquad \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{8}$$

$$NP_{pjigt}, NS_{sjigt} \in \mathbb{N} \qquad \forall p \in P,\, \forall s \in S,\, \forall j \in J,\, \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{9}$$

$$P_{pjigt}, Q_{ilgg'} \in \mathbb{R}_{\geq 0} \qquad \forall p \in P,\, \forall j \in J,\, \forall i \in I,\, \forall l \in L,\, \forall g \in G,\, \forall t \in T \tag{10}$$

$$\mathbf{x} = [NP_{pjigt}, NS_{sjigt}, P_{pjigt}, Q_{ilgg't}]^T \tag{11}$$

where $\mathbf{x}$ in equation (1) represents the vector of decision variables, i.e., $\mathbf{x} = [NP_{pjigt}, NS_{sjigt}, P_{pjigt}, Q_{ilgg't}]^T$, as stated in equation (11). The exhaustive model nomenclature is presented in Appendix A.

The installed production plants in the network (NP_{pjigt}) must allow to exactly satisfy the total hydrogen demand, for each type of hydrogen physical form i and for each period t. This is enforced by constraints (2) and (3). Similarly, constraints (4) and (5) ensure that the installed storage facilities (NS_{sjigt}) will guarantee the total inventory within certain limits, for each storing product form i, each grid g and each period t.

Constraints (6) and (7) force the production rate of a given plant (P_{pjigt}) to be within the allowed (minimum and maximum) production capacities.

Equation (8) states that hydrogen demand must be satisfied exactly (that is, not overproduction is permitted), through local production and/or importation/exportation from/to other grid ($Q_{ilgg't}$). This mass balance needs to be fulfilled for each product physical form i, each grid g and each period t.

Constraints (9) and (10) sets the variable nature (integer for the production and storage facilities, and non-negative for production and flow rates).

In the following, the $TDC(\mathbf{x})$ is written as TDC for short. The same for $GWP(\mathbf{x})$.

Accounting for $|T|$ time periods, the TDC objective function is calculated as the sum of the total daily cost over all periods as:

$$TDC = \sum_{t \in T} TDC_t \tag{12}$$

The total daily cost for each period t takes into account investment costs related to plant installation (FCC) and transportation (TCC), as well as operational costs related to production and storage (FOC), transportation (TOC) and use of energy source (ESC). It is computed as:

$$TDC_t = \frac{FCC_t + TCC_t}{\alpha CCF} + FOC_t + TOC_t + ESC_t \qquad \forall t \in T \tag{13}$$

where FCC_t represents the facility capital cost at each period t and thus considers only the cost associated to the installation of new production plants (IP_{pjigt}) and new storing facilities (IS_{sjigt}) at a given period as:

$$FCC_t = \frac{1}{LR_t} \sum_{j \in J} \sum_{i \in I} \sum_{g \in G} \left(\sum_{p \in P} PCC_{pji} IP_{pjigt} + \sum_{s \in S} SCC_{sji} IS_{sjigt} \right) \qquad \forall t \in T \tag{14}$$

where LR_t represents the learning rate, that is, a cost reduction associated to acquired experience over time by technology manufacturers. The decision variables NP_{pjigt} and NS_{sjigt} are related to IP_{pjigt} and IS_{sjigt}, by the following two equations, respectively, as:

$$NP_{pjigt} = \sum_{t=1}^{t} IP_{pjigt} \qquad \forall p \in P, \forall j \in J, \forall i \in I, \forall g \in G, \forall t \in T \tag{15}$$

$$NS_{sjigt} = \sum_{t=1}^{t} IS_{sjigt} \qquad \forall s \in S, \forall j \in J, \forall i \in I, \forall g \in G, \forall t \in T \tag{16}$$

The transportation capital cost (TCC_t) considers the flow rate of product between grids ($Q_{ilgg't}$), the transportation mode availability (TMA_l), the capacity of the transport container ($TCap_{il}$), the average distance connecting two grids ($L_{lgg'}$), the average speed (SP_l) as well as the loading/unloading time (LUT_l). Also, a factor accounting for establishing a given transportation mode is used (TMC_{il}). It is calculated as:

$$TCC_t = TMC_{il} \sum_{i \in I} \sum_{l \in L} \sum_{g \in G, g' \neq g} \frac{Q_{ilgg't}}{TMA_l TCap_{il}} \left(\frac{2L_{lgg'}}{SP_l} + LUT_l \right) \qquad \forall t \in T \tag{17}$$

The facility operating cost (FOC_t) considers the efficient operation of each production

plant and storage facility. It is directly related to the amount of production and storage as:

$$FOC_t = \sum_{j\in J}\sum_{i\in I}\sum_{g\in G}\left(\sum_{p\in P} UPC_{pji}P_{pjigt} + \sum_{s\in S} USC_{sji}S^T_{igt}\right) \qquad \forall t \in T \tag{18}$$

where UPC_{pji} and USC_{sji} are the unit cost of production and storage, respectively.

The transportation operating cost (TOC_t) consists of fuel, maintenance, labor and general costs as:

$$\begin{aligned} TOC_t = \sum_{i\in I}\sum_{l\in L}\sum_{g\in G, g'\neq g} \frac{2L_{lgg'}Q_{ilgg't}}{TCap_{il}}\left(\frac{FP_l}{FE_l} + ME_l \right. \\ \left. + \left(DW_l + \frac{GE_l}{TMA_l}\right)\left(\frac{1}{SP_l} + \frac{LUT_l}{2L_{lgg'}}\right)\right) \qquad \forall t \in T \end{aligned} \tag{19}$$

where the terms inside the parenthesis accounts for the fuel costs (FP_l, FE_l), maintenance expenses (ME_t), driver wages (DW_l) and general maintenance costs (GE_l, TMA_t), which depends on the working time (related to the speed SP_l, the distance $L_{lgg'}$, and the loading/unloading times LUT_l).

The cost associated to the transportation of primary energy sources is computed as:

$$ESC_t = \sum_{e\in E}\sum_{g\in G} UIC_e IPES_{egt} \qquad \forall t \in T \tag{20}$$

where UIC_e is the unit import cost of energy source and $IPES_{egt}$ is the corresponding amount of imported energy source, which depends on the production rate and the availability of energy sources (A_{egt}) for a given grid g, at each period t, according to:

$$\begin{aligned} IPES_{egt} = \max\left(0, SSF\sum_{p\in P}\sum_{j\in J}\sum_{i\in I}\gamma_{epj}P_{pjigt} - A_{egt}\right) \qquad \forall e \in E, \\ \forall g \in G, \forall t \in T \end{aligned} \tag{21}$$

where SSF is a safety stock factor for storing a small inventory of energy source while γ_{epj} represents the rate of utilization of primary energy source.

Concerning the second objective, the overall effects of GHG of the network are considered by the production, storage and transportation of hydrogen according to the following

equation:

$$GWP = \sum_{t \in T} (PGWP_t + SGWP_t + TGWP_t) \tag{22}$$

where $PGWP_t$ is related to production, and it is computed as the production rate of each production plant in the network times a global warming potential factor associated as:

$$PGWP_t = \sum_{p \in P} \sum_{j \in J} \sum_{i \in I} \sum_{g \in G} GW_p^{prod} P_{pjigt} \qquad \forall t \in T \tag{23}$$

Similarly, the global warming potential due to product storage, is expressed as the production rate of each production plant in the network times a global warming potential factor as:

$$SGWP_t = \sum_{p \in P} \sum_{j \in J} \sum_{i \in I} \sum_{g \in G} GW_i^{stock} P_{pjigt} \qquad \forall t \in T \tag{24}$$

The third term in equation (22) refers to the global warming potential due to transportation, it considers the average distance and the product flow rate between grids, the capacity of the transportation mode employed, its weight (W_l), and a transportation global warming potential factor. It is computed as:

$$TGWP_t = \sum_{i \in I} \sum_{l \in L} \sum_{g \in G, g' \neq g} \left(\frac{2L_{lgg'} Q_{ilgg't}}{TCap_{il}} \right) GW_i^{transp} W_l \qquad \forall t \in T \tag{25}$$

4. An efficient solution strategy

The complexity of the HSC problem, due to its combinatorial and multi-objective nature, deserves the development of an adapted solution strategy. To this end, the original problem (A)[1] is reformulated into a bi-level optimization problem. The idea behind this reformulation is to decompose the original problem into two sub-problems of lower complexity, in such a manner that each aspect of the original problem can be tackled by appropriate solution approaches. The two main parts of problem (A) are (1) the combinatorial problem related to facility installation and (2) the continuous problem associated to production and transportation rates. Then, in such a decomposition, the combinatorial problem can be tackled apart by a population-based metaheuristic, so that good-quality

[1]For the sake of readability, in this section the problem represented by equations (1-25) is named problem (A), it constitutes the original or classical formulation of the HSC problem, according to the models of Almansoori and Shah (2009); Almaraz et al. (2014).

solutions can be obtained even for large-size instances in reasonable computational times. Additionally, the multi-objective nature of this sub-problem can be considered by that metaheuristic, in particular, by using a multi-objective evolutionary algorithm (MOEA). Now, regarding the continuous problem, it can be easily transformed into a canonical transport problem applying a simple heuristic (detailed in Subsection 4.4), and so it can be solved efficiently by exact methods.

In the next subsection, the reformulation of problem (A) into a bi-level problem is formally provided. Then, the proposed methodology for its solution is discussed in Subsection 4.2. The details of this solution strategy for each sub-problem are presented in the subsequent subsections.

4.1. Bi-level formulation

In order to propose an efficient strategy to solve the optimization problem (A), a decomposition of the problem is proposed in this subsection. The resulting bi-level optimization problem is represented as follows:

$$\min_{\mathbf{x}} \quad [TDC(\mathbf{x}), GWP(\mathbf{x})]^T \tag{26}$$

$$\text{s.t.} \quad \sum_{p\in P}\sum_{j\in J}\sum_{g\in G} PCap^{\min}_{pji} NP_{pjigt} - \sum_{g\in G} D_{igt} \leq 0 \qquad \forall i \in I,\, \forall t \in T \tag{27}$$

$$\sum_{g\in G} D_{igt} - \sum_{p\in P}\sum_{j\in J}\sum_{g\in G} PCap^{\max}_{pji} NP_{pjigt} \leq 0 \qquad \forall i \in I,\, \forall t \in T \tag{28}$$

$$\sum_{s\in S}\sum_{j\in J} SCap^{\min}_{sji} NS_{sjigt} - S^T_{igt} \leq 0 \qquad \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{29}$$

$$S^T_{igt} - \sum_{s\in S}\sum_{j\in J} SCap^{\max}_{sji} NS_{sjigt} \leq 0 \qquad \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{30}$$

$$NP_{pjigt}, NS_{sjigt} \in \mathbb{N} \qquad \forall p \in P,\, \forall s \in S,\, \forall j \in J,\, \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{31}$$

$$\min_{\mathbf{y}} \quad [TDC(\mathbf{y}), GWP(\mathbf{y})]^T \tag{32}$$

$$\text{s.t.} \quad PCap^{\min}_{pji} NP_{pjigt} - P_{pjigt} \leq 0 \qquad \forall p \in P,\, \forall j \in J,\, \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{33}$$

$$P_{pjigt} - PCap^{\max}_{pji} NP_{pjigt} \leq 0 \qquad \forall p \in P,\, \forall j \in J,\, \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{34}$$

$$\sum_{p\in P}\sum_{j\in J} P_{pjigt} - \sum_{l\in L}\sum_{g'\in G, g'\neq g} (Q_{ilgg't} - Q_{ilg'gt}) - D_{igt} = 0 \qquad \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{35}$$

$$P_{pjigt}, Q_{ilgg'} \in \mathbb{R}_{\geq 0} \qquad \forall p \in P,\, \forall j \in J,\, \forall i \in I,\, \forall l \in L,\, \forall g \in G,\, \forall t \in T \tag{36}$$

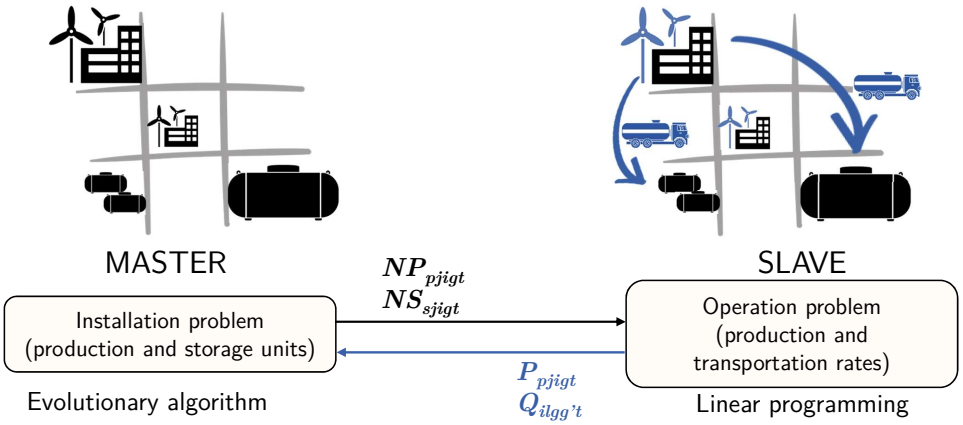

Figure 2: Hybrid (master - slave) approach diagram.

where $\mathbf{x}$ in equation (26) represents the vector of decision variables in the upper level, i.e., $\mathbf{x} = [NP_{pjigt}, NS_{sjigt}, P_{pjigt}, Q_{ilgg'}]^T$ as in (1), and $\mathbf{y}$ is the vector of decision variables in the lower level, i.e., $\mathbf{y} = [P_{pjigt}, Q_{ilgg'}]^T$. In addition to equations (26-36), the complete formulation of the bi-level problem includes equations (12-25). This constitutes problem (B). It is worth noting that the optimal Pareto sets of original problem (A) and reformulated problem (B) are the same.

The upper-level (master) problem addresses the combinatorial problem associated to the structure of the supply chain, that is, it finds the optimal location, size and technology type of production and storage facilities. Concerning the lower-level (slave) problem, it consists in finding the optimal production rates and transportation flows between grids for the network configuration predicted by the upper-level.

A general description of the proposed methodology for the solution of problem (B) is provided in the next subsection.

4.2. Global description of the strategy

As stated previously, the proposed solution strategy to the optimization problem (B), consists in a hybrid approach, more precisely, on a MOEA coupled with a linear programming solver. In this way, the upper-level integer variables NP_{pjigt} and NS_{sjigt} are handled by the evolutionary algorithm. Since these variables are generated through stochastic operators, they might require a repair mechanism in order to fulfill constraints (27-30), which state that production and storage facilities must allow for meeting the total demand in the network. Then, for each feasible (partial) solution, that is, for each feasible structure of the HSC provided by the evolutionary algorithm, the optimal continuous variables P_{pjigt} and $Q_{ilgg't}$ are computed by solving the multi-period operation problem,

```
1  initialize MOEA
2  assign integer variables NP_pjigt, NS_sjigt
3  for each period t ∈ T
4      if individual violates equations (27-30)
5          repair infeasible solution
6      end
7      construct transportation problem (identify sink and source grids)
8      solve transportation problem (LP solver)
9  end
10 compute master problem's objective functions
11 assign fitness value according to MOEA's working paradigm
12 evolve population according to MOEA's working paradigm
13 if termination criteria not reached
14     go to step 2
15 else
16     output current Pareto set approximation
17 end
```

Figure 3: Algorithmic scheme of the hybrid solution strategy proposed.

at the lower-level (see Figure 2). At this point, it is important to note that, even if the lower-level sub-problem constitutes a bi-objective problem, only one solution is required to evaluate a given upper-level candidate solution. To this end, a utility function with a random weight vector is used (this is discussed in more details in Subsection 4.4). Moreover, in order to speed up the solution of the slave problem, only grids for which the installed production units can satisfy local demand are considered as potential exporting grids (sources), otherwise they are considered as demand (sinks) grids. This constitutes a heuristic that reduces the complexity of the linear programming problem. Then, once the slave problem has been solved for every master problem's candidate solution, the MOEA recovers the continuous variables (P_{pjigt}, $Q_{ilgg't}$) from the LP solver, in order to compute the upper-level objective functions. Once evaluated, each individual in the population has its fitness function value assigned by the MOEA, which evolves the population to the next generation according to its working mode, e.g., according to Pareto dominance principles, or through a decomposition function or a performance indicator. Finally, the process is repeated iteratively until a stopping criterion is reached. The proposed methodology is presented in an algorithmic scheme in Figure 3.

In Figure 3, all lines, excepting lines 7 and 8 (slave problem), correspond to the solution of the master sub-problem by means of a MOEA. This is detailed in Subsection 4.3. Then, in Subsection 4.4, the solution of the slave problem is discussed.

4.3. Upper-level problem solution approach

In this work, the SMS-EMOA algorithm (Emmerich et al., 2005) is used for solving the bi-objective master problem. This quality-indicator based MOEA seeks to maximize the

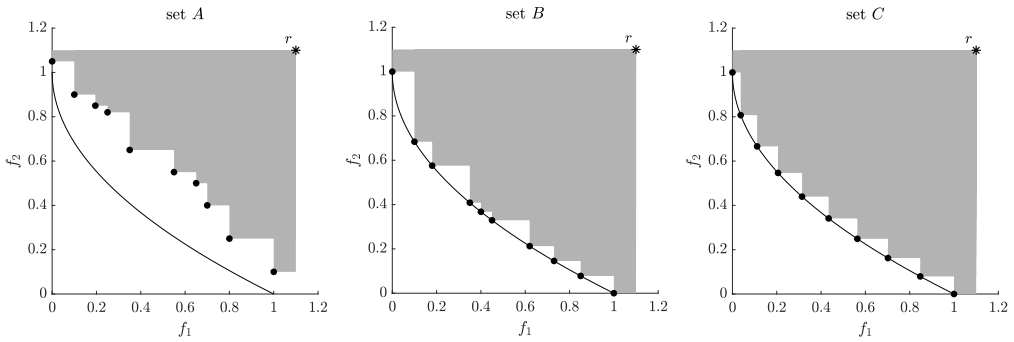

Figure 4: Hypervolume representation for the ZDT1 minimization problem (shaded area). HV(A) = 0.5808; HV(B) = 0.8101; HV(C) = 0.8237.

hypervolume measure (HV), also known as $\mathcal{S}$ metric, which is an indicator that reflects the volume enclosed by a Pareto front approximation and a reference point (Zitzler and Thiele, 1998). This performance indicator has the remarkable property that its maximum value is yielded only by the optimal Pareto set. A graphical representation of the hypervolume, for a classical benchmark problem, is provided in Figure 4. It can be observed that both convergence to the Pareto front and diversity among solutions are addressed by this indicator: solutions in set C show a good convergence and are well distributed along the Pareto front so that the highest hypervolume value is obtained among the three sets.

Moreover, SMS-EMOA is a steady-state greedy algorithm, meaning that only one individual is generated at each iteration and no decrease in the hypervolume covered by the current population is permitted. This implies that new offspring solutions can only integrate the current population if replacing a member increases the hypervolume covered by the population, or, more precisely, if the new offspring (1) dominates at least one individual in the current population, or (2) does not contribute the least to the hypervolume computation provided that it belongs to the last Pareto front. Condition (1) is taken from the well-known NSGA-II algorithm (Deb et al., 2002), and is easy to grasp: among two solutions, the one that dominates the other is preferred at any point of the evolutionary process. Condition (2) introduces contribution $\Delta_{\mathcal{S}}$ of each individual to the hypervolume as the selection criterion for those individuals in the last front (according to the non-dominated sorting procedure in NSGA-II). That is, the individual that contributes the least to the HV is discarded from the worst ranked front. In a bi-objective case, this contribution can be computed efficiently as the product of the difference of the objective values between two subsequent solutions, once the set has been sorted in ascending order according to the values of the first objective function f_1. Note that, since extreme points are to be maintained in the population, the contribution to the hy-

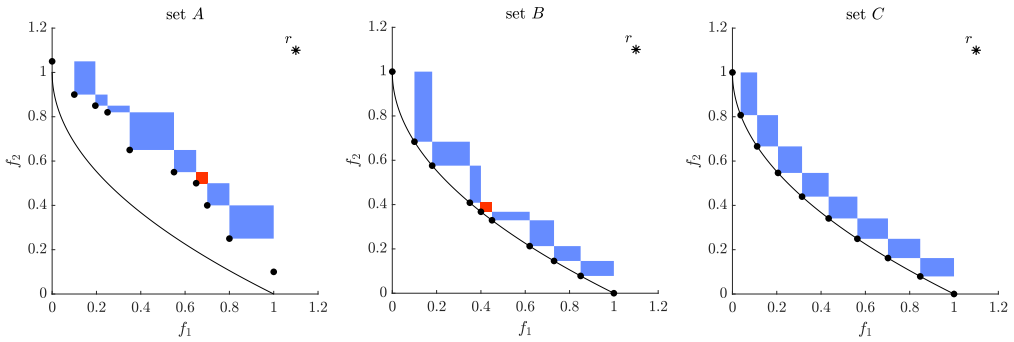

Figure 5: Representation of the contribution to the hypervolume for the ZDT1 problem (shaded rectangles).

pervolume is only computed for internal solutions of the last non-dominated rank. Then, considering the last front $\mathcal{R}_I = \{s_1, \ldots, s_{|\mathcal{R}_I|}\}$ with $|\mathcal{R}_I| > 2$, the $\Delta_{\mathcal{S}}$ of each internal solution s_n is computed as:

$$\Delta_{\mathcal{S}}(s_n, \mathcal{R}_I) = (f_1(s_{n+1}) - f_1(s_n)) \cdot (f_2(s_{n-1}) - f_2(s_n)) \tag{37}$$

for all $n \in \{2, \ldots, |\mathcal{R}_I| - 1\}$. In Figure 5 the contribution of the hypervolume of three approximation sets for the ZTD1 problem is displayed graphically. The area of each rectangle corresponds to the $\Delta_{\mathcal{S}}$ of each candidate solution. The red rectangle in sets A and B represents the $\Delta_{\mathcal{S}}$ of the solution that contributes the least to the HV. Thus, the best possible approximation to the Pareto front for a given number of efficient solutions is a set in which each solution contributes in an equal extent to the hypervolume total value, such as set C in Figure 5.

The main drawback of SMS-EMOA is the high computational cost of $\Delta_{\mathcal{S}}$ for more than two objectives, which undermines its efficiency for many objectives or large sets (Knowles and Corne, 2002). However, this does not constitute an issue for the bi-objective HSC design problem tackled here.

In what follows, some relevant aspects of the implementation of SMS-EMOA in our framework are discussed, considering all lines excepting 7 and 8 of Figure 3.

Line 1: Initialization. The initialization step of SMS-EMOA consists in declaring the initial parameters: population size μ, maximum number of generations as the stopping criterion (or maximum number of candidate solutions) and the evaluation of the initial random population. This initial population is created in such a manner that production and storage facilities are added randomly into the grids in the network, one by one, until a feasible solution is obtained (which ensures the respect of constraints 27-30).

Line 2: Offspring. One offspring solution is generated by means of stochastic operators, namely, SBX crossover with a polynomial mutation (Deb and Agrawal, 1995; Deb, 2001). The integer decision variables NP_{pjigt}, NS_{sjigt} are encoded as continuous variables and are rounded to the next integer only in the evaluation steps. This contributes to diversity in the population.

Lines 4-6: Reparation. If the offspring violates constraints (27-30), it is repaired by randomly adding or removing production and storage units accordingly. Note that (27-30) are inequality constraints which define large feasible regions, and thus are not difficult to fulfill following this procedure.

Line 10: Evaluation. The optimal variables $P_{pjigt}, Q_{ilgg't}$ are recovered from the LP solver in order to compute the master problem's objective functions. Note that these variables are not encoded as decision variables in any individual, at any step of the evolutionary algorithm.

Line 11: Fitness assignment. The μ+1 individuals, which include the current population and the candidate offspring, are sorted following the non-dominated sorting procedure proposed in Deb et al. (2002). Then, $\Delta_{\mathcal{S}}$ is computed for all internal individuals in the last front according to equation (37).

Line 12: Selection. If the worst ranked front contains more than two individuals, that with least $\Delta_{\mathcal{S}}$ is discarded from the population (no external archive is used). Otherwise, if the last front contains exactly two solutions, one is randomly chosen to be discarded.

Lines 13-17: Evolution/Termination. If the termination criteria is not met, the μ best fitted individuals survive to the next generation and the process is repeated, otherwise the SMS-EMOA algorithm terminates and the current population is output.

4.4. Lower-level problem solution approach

A first key point for the solution of the slave problem is the construction of a feasible transportation problem using the integer variables provided by the evolutionary module. To this end, we propose the following heuristic. All grids in the network are considered as potential importing grids, that is, as potential "sinks" or "demand" nodes. Then, for each period t, a set of potential exporting grids G_t^E is distinguished depending on the number of production plants installed as follows: if, for a given grid g, the sum of the maximum production capacities of all plants installed is greater than its hydrogen

demand, the grid g is considered as a potential "source" node and then $g \in G_t^E$. This can be mathematically expressed as:

$$D_{igt} \leq \sum_{p \in P} \sum_{j \in J} PCap_{pji}^{\max} NP_{pjigt} \implies g \in G_t^E \qquad \forall i \in I,\, \forall g \in G,\, \forall t \in T \tag{38}$$

In this way, the complexity of the resulting LP problem is considerably reduced, since only transportation links can occur between sources and sinks nodes. As it can be noted, this heuristic relies on the fact that grids that cannot produce enough hydrogen to satisfy its own demands (even if the installed facilities operate at their maximum capacity), are unlikely to export hydrogen to other grids, otherwise more transportation units would be necessary, resulting in a more expensive solution.

Now, regarding the multi-objective aspect of the LP sub-problems, this is tackled employing a utility function. Such a function decomposes the original problem into a single-objective problem by means of a weight vector $\mathbf{w}$ associated to the importance of each objective, and thus assigns a target direction in the objective space. In this work, the augmented achievement scalarizing function (AASF) is chosen, due to its ability to deal with non-convex front shapes and weakly dominated solutions (Miettinen, 2012; Pescador-Rojas et al., 2017). Therefore, the bi-objective function in equation (32) is converted into the following scalar function u:

$$u(\mathbf{f}; \mathbf{w}) = \max_i \left\{ \frac{f_i}{w_i} \right\} + \alpha \sum_{i=1}^{k} \left\{ \frac{f_i}{w_i} \right\} \tag{39}$$

where $\mathbf{f} = [TDC, GWP]^T$, thus $k = 2$ and α takes small values (Pescador-Rojas et al., 2017)(in this work $\alpha = 1\text{e}{-3}$). Since only one solution (not a set of non-inferior solutions) is needed to evaluate the upper-level problem, it is solved only once with a weight vector randomly generated such that $\mathbf{w} \in \mathbb{R}_+^k$ and $\sum_{j=1}^{k} w_j = 1$. It is worth mentioning that the random generation of weight vectors might seem inappropriate to deal consistently with the lower-level sub-problem. However, the experimental results (see Subsection 5.3) validate this strategy for the studied problem. Anyway, a smart weight vector tuning mechanism is under the scope of future work.

5. Computational experiments

The efficiency of the proposed methodology for solving the bi-objective HSC problem is validated through a performance comparison with a classical approach, namely, a branch-and-bound algorithm with ε-constraint method. For this purpose, six growing size instances of the problem are studied.

5.1. Case study

The instances for the HSC problem considered in this work correspond to the data for the case study of Midi-Pyrénées territory in France (Almaraz et al., 2014), see Figure 6. Accordingly, the characteristics of the HSC considered are:

- Five primary energy sources are available, namely, solar photovoltaic, wind, hydro, nuclear (electricity from the grid) and natural gas.
- Three different production technologies: steam methane reforming (SMR), central and distributed electrolysis. Renewable energies can only serve as feedstock to electrolysis plants, while natural gas is exclusively destined to SMR technology.
- Three different production scales: small, medium and large. The large-scale size plant is only available for central electrolysis and SMR technologies.
- Once hydrogen is produced, it will be delivered to storage facilities via tanker trucks in liquid form.
- Like for production plants, different sizes for storage facilities are accessible: mini, small, medium and large.
- The learning rate cost reductions due the accumulated experience is considered as 2% per period.

The detailed input data is presented in the supplementary material. Also, in order to validate the solution strategy, different size instances of the problem are generated, varying the number of grids and time periods. The network is divided into 8 or 22 grids. The entire planning horizon is set to year 2050, with three different periods divisions: 1 (2050), 4 (2020-2050, ten-year periods) and 7 periods (2020-2050, five-year periods). The combination of these two aspects yields the six different instances studied here. The nomenclature used for instance names indicates the number of grids and periods, as an example, the instance *HSC22g04p* considers 22 demand grids and 4 time periods.

5.2. Parameter settings

The performance of the proposed solution approach is compared to that of the classical approach usually considered in the literature. To this end, the original MILP model, defined by equations (1-25), was solved using CPLEX solver with ε-constraint in GAMS environment. Because of the multi-objective nature of the problem, the performance comparisons of solution sets of both approaches are carried out using the hypervolume indicator. The reference point for the HV computation is equal to $[1.1, 1.1]^T$ in the normalized objective space $[0, 1]^2$, for all instances. As it has been previously mentioned,

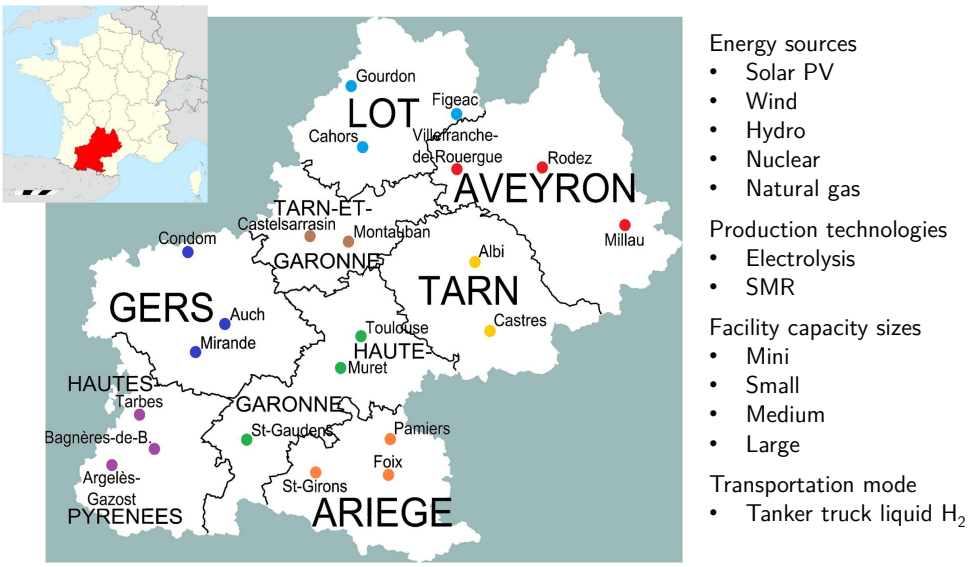

Figure 6: Midi-Pyrénées region case study. Names in capitals indicate departments (grids) for 8-grid instances. Color points correspond to grid cities (in minuscules) in 22-grid instances.

a maximum value of the hypervolume is preferred, since it reflects both convergence to the Pareto front and spread of solutions in the objective space. Solution sets of equal size are necessary to perform a fair comparison. For the exact approach, an approximation set containing 21 solutions is obtained for each instance using the ε-constraint method. We considered this arbitrary size (21) for the approximation set as adequate to display the shape of the true Pareto front, and at the same time, in order to avoid an untractable computational burden to the exact method. The stopping criteria for each CPLEX execution (21 executions for each instance) are (1) the solution found so far has an optimality gap lower than 0.01%, or (2) the computational time exceeds a given limit. Regarding the hybrid algorithm, the population size of the MOEA is set to 200 for 8-grid instances and 800 for 22-grid instances, according to a previous sensitivity analysis; the stopping criterion (number of function evaluations) is set so that similar computational times to those of the exact method are employed. The variation operators use the following (standard) parameters: SBX operator's probability and distribution $p_c = 1$ and $\eta_c = 20$, respectively; polynomial mutation's probability and distribution $p_m = 1/n$ and $\eta_m = 20$, respectively, where n is the number of decision variables. Also, since the Pareto set approximation provided by the MOEA typically contains more than 21 solutions, the final population is pruned so that only 21 well-distributed solutions are considered to calculate the HV. Finally, in order to perform appropriate (statistical) comparison of results, the hybrid approach is run 11 times for each instance.

Table 1: Numerical results of both approaches for all instances.

Instance	Time limit p/point (s)	CPU time(s)	Optimality gap (%)	HV(ε-constr.)	HV(hybrid)
HSC08g01p	100	142.50	0.01	0.9552	0.9834 (0.0003)
HSC08g04p	100	1 711.65	0.06	0.8008	0.8099 (0.0002)
HSC08g04p	1 000	7 852.27	0.02	0.8008	0.8100 (0.0002)
HSC08g07p	100	2 056.78	0.23	0.7639	0.7680 (0.0004)
HSC08g07p	1 000	14 629.53	0.03	0.7639	0.7690 (0.0008)
HSC22g01p	100	1 674.72	0.41	0.9616	0.9794 (0.0022)
HSC22g01p	1 000	12 526.37	0.09	0.9618	0.9822 (0.0013)
HSC22g04p	100	2 432.25	2.31	0.7736	0.7776 (0.0035)
HSC22g04p	1 000	22 416.78	0.71	0.7908	0.7934 (0.0023)
HSC22g07p	100	2 425.38	–	Infeasible	0.7394 (0.0027)
HSC22g07p	1 000	22 950.42	0.95	0.7555	0.7563 (0.0014)

The exact method was implemented in GAMS environment (v23.9.5) using CPLEX solver v12.4.0.1. For the hybrid approach, SMS-EMOA was implemented in MATLAB language (vR2019a), and the solution of the LP sub-problems is performed using CPLEX solver v12.8.0 called from MATLAB. Both deterministic and hybrid approaches were carried out within the same computer hardware, with a processor Intel Xeon E3-1505M v6 at 3.00 GHz and 32 Go RAM.

5.3. Results and discussion

The obtained numerical results are displayed in Table 1. The column indicating the optimality gap refers to the average optimality gap over the 21 CPLEX executions required to produce an approximation set. Besides, for the exact approach, the time limit per point is set to 1 000 seconds, but also, in order to track the any-time performance of each solution approach, Table 1 presents intermediate results setting the time limit per point at 100 seconds. For the hybrid approach (last column), the results represent the mean HV value and, in parentheses, the corresponding standard deviation over 11 runs. As mentioned previously, for the HV computation only 21 solutions from the last population are considered.

First, small size instances were studied for validating the hybrid solution method. The Midi-Pyrénées region is thus divided into its 8 main districts. From Table 1, it can be observed that the first instance, which considers a constant demand, required only 142 seconds from the exact method to obtain the Pareto front approximation, each solution being proved as optimal since the average optimality gap is equal to 0.01%.

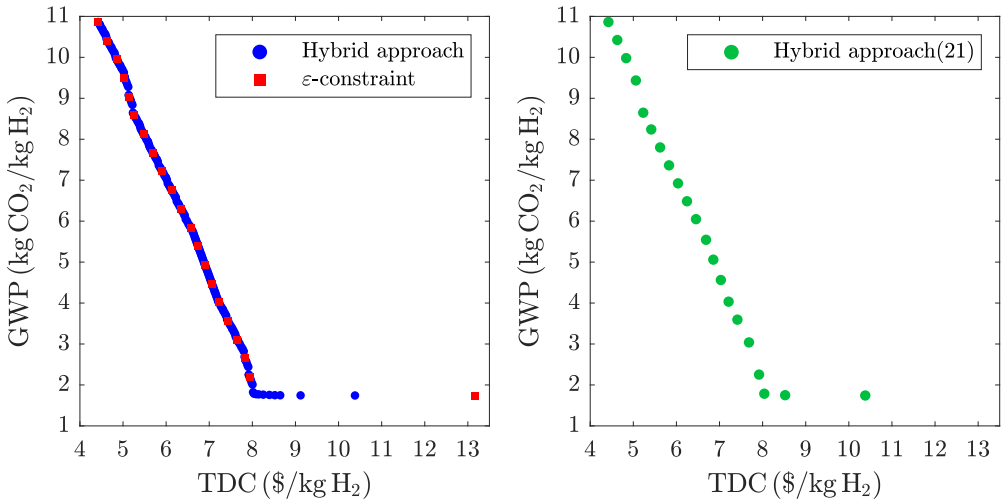

Figure 7: Pareto front approximations for HSC08g01p instance. For the hybrid algorithm the results of the median run with respect to the HV value are presented.

Consequently, the hybrid approach was stopped after a similar runtime. The Pareto front approximation of both approaches is displayed in Figure 7. It can be observed that the hybrid approach reproduces the non-dominated solutions found by CPLEX (red squares), and besides, it provides a much more detailed Pareto front approximation (blue points), finding 200 efficient solutions. From these 200 points, 21 are selected for the hypervolume computation in order to carry out a fair comparison with the output of the exact method. However, it should be clear to the reader that, without pruning the final population for the above-mentioned purpose, the hybrid strategy provides a much more detailed approximation of the complete front. Although not computed here, the hypervolume corresponding to the complete approximation produced by the hybrid algorithm would drastically outperform those shown in the table for 21 points. From the numerical results in the table, it can be concluded that the 21-point approximation set proposed by the hybrid approach (HV 0.9834) is a better representation of the true Pareto front than that of the exact method (HV 0.9552), because solutions are better distributed along the Pareto front. Also, since the standard deviation is not significant (0.0003), it is concluded that the algorithm is consistent in each run.

Then, a 4-period instance of the 8-grid network is considered. This time, some points solved by ε-constraint exceeded the time limit of 1 000 seconds and the final approximation set has an average optimality gap of 0.02%, i.e., CPLEX already experiences convergence troubles for this small size case. Also, from the intermediate results (time limit per point equals to 100 seconds) shown in Table 1, it is appreciated that both approximation sets obtained by the exact method present the same value of the hyper-

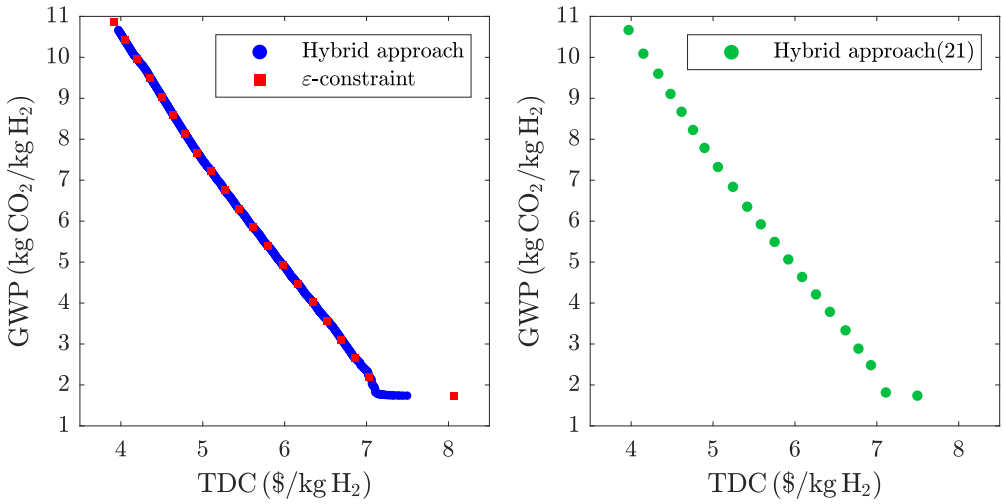

Figure 8: Final Pareto front approximations for HSC08g04p instance. For the hybrid algorithm the results of the median run with respect to the HV value are presented.

volume (0.8008), which demonstrates that both sets contain optimal solutions even if convergence is not achieved (see their corresponding values of optimality gap). With respect to the hybrid approach, the mean hypervolume values at both instants of the optimization (0.8099 and 0.8100, respectively) indicate a good performance for finding trade-off solutions of the HSC problem in all runs (standard deviation 0.0002). Note that at both instants, that is, at approximately 1 711 and 7 852 seconds, the hybrid approach outperforms the classical one because the solution set has a better distribution in the objective space. Moreover, it is important to emphasize that the evolutionary algorithm provides an approximation of the Pareto front that, if judged visually, could be considered as continuous (it contains 200 efficient solutions). This can be appreciated in Figure 8. In comparison to the single-period instance, the TDC objective for the HSC08g04p instance is significantly lower, which is explained by the learning rate factor that reduces capital costs in the subsequent time periods, accounting for gained experience by technology manufacturers over time.

Now, regarding the 8-grid 7-period instance, the numerical results in Table 1 are once again in favor of the hybrid strategy. For the ε-constraint method, the optimality gap decreases between 100 and 1 000 seconds while the hypervolume remains steady, leading to the same conclusions as previously: the feasible solutions found might be optimal, but convergence cannot be guaranteed in a reasonable time. Moreover, due to the fact that ε-constraint method performs an even partition of the Pareto front with respect to a given objective, the obtained efficient solutions, despite the possibility of being proven as optimal, might not be well distributed along the Pareto front. This explains the

superiority of the hybrid approach.

The influence of grid number representing the spatial granularity has also been studied, thus, instances considering each main city in the Midi-Pyrénées region are now discussed. Concerning the HSC22g01p instance, the single time-period makes its combinatorial complexity similar to that of HSC08g07p. This can be confirmed by comparing the CPU times and optimality gaps for both instances in Table 1. The hypervolume indicator suggests a better performance of the hybrid approach for this instance, at both instants, that is, after 1 674 and 12 526 seconds. The standard deviation values of the 11 runs performed by the MOEA are higher for the 22-grid instances than for those of 8 grids. This is somewhat related to the optimality gap for the exact method, in the sense that it gives some insights about the convergence of the SMS-EMOA algorithm. However, the HV' standard deviation are not significant considering the mean values, meaning that even the worst run outperforms the exact method.

For the HSC22g04p instance, the required computational times increase considerably as well as the optimality gap for CPLEX. There is a significant difference between the obtained HV values at each instant, for both approaches: approx. 0.77 and 0.79, at 2 432 and 22 416 seconds, respectively. The mean HV values of the hybrid approach are greater than that of the ε-constraint method, however its standard deviation values do not allow to prove a significant superiority, when comparing equal number of solutions. Nevertheless, it is worth highlighting that the SMS-EMOA algorithm is able to provide 800 non-dominated solutions to decision makers. The median solution according to the hypervolume obtained by the stochastic approach is plotted in Figure 9 together with the obtained solution by the deterministic method.

Finally, the largest instance, HSC22g07p describes the more realistic case study, with a forecast of hydrogen demand over 30 years split into 7 five-year periods considering the most important cities (in terms of population density) in the Midi-Pyrénées territory. Obtaining the proven optimal solution for every point in the Pareto approximation set using an exact approach would take prohibitive computational times, because of the large size of the search space. In these cases, a near-optimal solution is accepted. Table 1 highlights that for short CPU times the exact method does not find any feasible solution, contrary to the hybrid approach. For greater computational times (1 000 s per point), the average optimality gap of the exact method's solution set is lower than 1%. The mean HV value of the 11 runs performed by the hybrid approach is slightly superior to that of ε-constraint method, additionally, the final approximation set (P) yielded by SMS-EMOA contains 800 efficient solutions, which yield a mean hypervolume far superior to that of CPLEX (0.7555), HV(P) = 0.7771 (not presented in Table 1). The obtained Pareto fronts are shown in Figure 10. It can be observed that the "bad" distribution

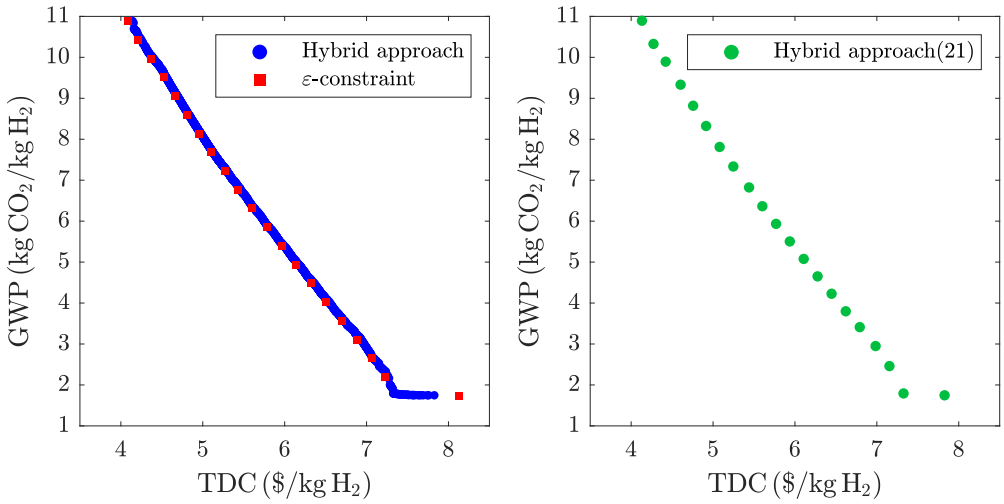

Figure 9: Final Pareto front approximations for HSC22g04p instance. For the hybrid algorithm the results of the median run with respect to the HV value are presented.

of the solution set obtained by CPLEX is located in the lower part of the Pareto front, where this method lacks to identify accurately the knee-points of the front (as appreciated also for the other instances). Note that missing the knee part of the Pareto front might be critical for the subsequent decision-making process, which in many cases chooses this knee solution.

6. Conclusions

The optimal design of the hydrogen supply chain constitutes a current challenge to society, since it gives the basis for the evaluation of a cost-efficient hydrogen-based economy. Its design is far from being a trivial task, in particular when both economical and environmental aspects are considered. The mathematical model of the HSC involves combinatorial aspects that could make classical optimization methods inefficient for solving large-sized instances of the problem, that is, for providing an accurate Pareto front approximation to the decision-makers. Therefore, in this work a novel methodology for solving the HSC problem has been presented.

The hybrid solution methodology proposed is a matheuristic combining the benefits of multi-objective evolutionary algorithms and linear programming. This methodology efficiently solves the multi-objective HSC problem, providing the decision-makers with a set of efficient solutions well distributed along the Pareto front. Besides, this hybrid approach has proven to explore regions of the Pareto front that might be ignored when using ε-constraint method, because this classical method can be constrained by

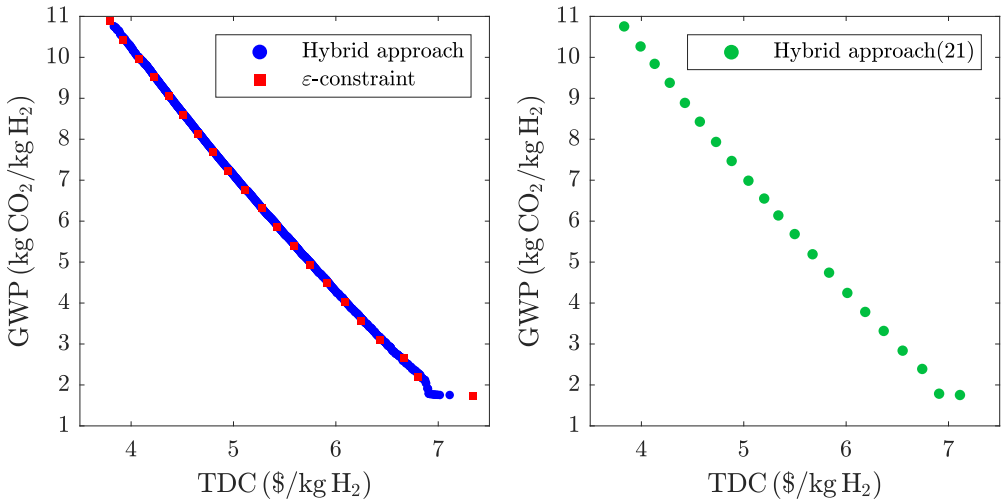

Figure 10: Final Pareto front approximations for HSC22g07p instance. For the hybrid algorithm the results of the median run with respect to the HV value are presented.

the computational burden associated to the need for multiple executions to produce an approximation of the Pareto front. Moreover, according to the obtained results, the methodological framework is able to find good-quality solutions in short computational times even for large-size instances, contrary to ε-constraint method, which rapidly can become computationally expensive.

The computational efficiency exhibited by the proposed approach is based on the following aspects:

1. The decomposition of the original model into a bi-level optimization problem, such that combinatorial and linear aspects of the problem are treated separately. This allows the efficient solution of each sub-problem (level) by appropriate solution methods.
2. The use of a MOEA for solving the multi-objective installation problem. Contrary to ε-constraint, the population-based algorithm solves the problem in a collaborative way, i.e., information among the population is shared to evolve it towards the true Pareto front.
3. The heuristic employed to construct the linear transport problem. This reduces the complexity of the problem as it reduces the number of potential exporting grids.

The solution strategy proposed here for the HSC design problem yet presents some limitations, thus providing potential improvement features. The first limitation regards the possibility of considering additional objectives: although the general structure of the hybrid algorithm can be maintained unchanged, SMS-EMOA might not be suited

anymore, mainly because of the computational cost corresponding to the hypervolume calculation. As a consequence, other paradigms for handling multiple objectives should be contemplated, such as, for instance, decomposition-based MOEAs (see MOEA/D, Zhang and Li (2007)). Besides, the plant installation costs involve, in many practical contexts, nonlinear terms. Therefore, an extension of the present work to deal with nonlinear formulations constitutes a significant room for improvement, in order to address more realistic representations of the HSC. These two points (changing the MOEA paradigm and considering nonlinear formulations) are actually under the scope of future work.

Appendix A: Model notation

Sets/Indices

E/e	Primary energy sources
G/g	Grid squares
I/i	Product physical form
J/j	Size of production plants and storage facilities
L/l	Type of transportation modes
P/p	Plant type with different production technologies
S/s	Storage facility type with different storage facilities
T/t	Time period of the planning horizon

Parameters

A^0_{eg}	Initial average availability of primary energy sources e in grid g, unit d^{-1}.
CCF	Capital charge factor – payback period of capital investment, y.
D^T_{igt}	Total demand for product form i in grid g during time period t, kg d^{-1}.
DW_l	Driver wage of transportation mode l, \$ h^{-1}.
FE_l	Fuel economy of transportation mode l, km L^{-1}.
FP_l	Fuel price of transportation mode l, \$ L^{-1}.
GE_l	General expenses of transportation mode l, \$ L^{-1}.

$L_{lgg'}$ Average delivery distance between grids g and g' by transportation mode l, km.

LR Learning rate – cost reduction as technology manufacturers accumulate experience, %.

LUT_l Load/unload time of product for transportation mode l, h.

ME_l Maintenance expenses of transportation mode l, \$ km^{-1}.

$PCap_{pji}^{\min}$ Minimum production capacity of plant type p and size j for product form i, kg d^{-1}.

$PCap_{pji}^{\max}$ Maximum production capacity of plant type p and size j for product form i, kg d^{-1}.

PCC_{pji} Capital cost of establishing plant type p and size j producing product form i, \$.

$Q_{il}^{\min}$ Minimum flow rate of product form i by transportation mode l, kg d^{-1}.

$Q_{il}^{\max}$ Maximum flow rate of product form i by transportation mode l, kg d^{-1}.

$SCap_{sji}^{\min}$ Minimum storage capacity of storage type s and size j for product form i, kg.

$SCap_{pji}^{\max}$ Maximum storage capacity of storage type s and size j for product form i, kg.

SCC_{sji} Capital cost of establishing storage type s and size j storing product form i, \$.

SP_l Average speed of transportation mode l, km h^{-1}.

SSF Safety stock factor of primary energy sources within a grid, %.

$TCap_{il}$ Capacity of transportation mode l transporting product form i, kg mode^{-1}.

TMA_l Availability of transportation mode l, h d^{-1}.

TMC_{il} Cost of establishing transportation mode l transporting product form i, \$ mode^{-1}.

UIC_e Unit cost of importing primary energy source e from overseas, \$ unit^{-1}.

UPC_{pji} Unit production cost for product form i by plant type p and size j, \$ kg^{-1}.

USC_{sji} Unit storage cost for product form i at storage type s and size j, \$ kg^{-1} d^{-1}.

α Network operating period, d y^{-1}.

β Storage holding period – average number of days' worth of stock, d.

γ Rate of utilization of primary energy source e by plant type p and size j, unit resource per unit product.

Decision variables

IP_{pjigt} Investment of plants of type p, size j, producing product form i in grid g, during period t.

IS_{sjigt} Investment of storage facilities of type s, size j, for product form i in grid g, during period t.

P_{pjigt} Production rate of product form i produced by plant type p of size j in grid g, during period t, $\mathrm{kg\,d}^{-1}$.

$Q_{ilgg't}$ Flow rate of product form i by transportation mode l between grids g and g', during period t, $\mathrm{kg\,d}^{-1}$.

Dependent variables

A_{egt} Average availability of primary energy source e in grid g during time period t, unit d^{-1}.

D^L_{igt} Demand for product form i in grid g satisfied by local production during time period t, kg d^{-1}.

D^I_{igt} Imported demand of product form i to grid g during time period t, kg d^{-1}.

ESC Primary energy source cost, \$ d^{-1}.

FC Fuel cost, \$ d^{-1}.

FCC Facility capital cost, \$.

FOC Facility operating cost, \$ d^{-1}.

GC General cost, \$ d^{-1}.

I_{egt} Import of primary energy source e to grid g from overseas during time period t, unit d^{-1}.

L_{gt}^{ave} Average delivery distance within grid g during time period t, km.

LC Labor cost, \$ d^{-1}.

LTC_{gt} Local transportation cost within grid g during time period t, \$.

MC Maintenance cost, \$ d^{-1}.

P_{igt}^{T} Total production rate of product form i in grid g during time period t, kg d^{-1}.

NP_{pjigt} Number of plants of type p, size j, producing product form i in grid g, during period t.

NS_{sjigt} Number of storage facilities of type s, size j, for product form i in grid g, during period t.

NTU Number of transport units.

$QE_{egg't}$ Flow rate of primary energy source e between grids g and g' during time period t, unit d^{-1}.

S_{sjigt} Average inventory of product form i stored in storage type s and size j in grid g during time period t, kg.

S_{igt}^{T} Total average inventory of product form i in grid g during time period t, kg.

TCC Transportation capital cost, \$.

TDC Total daily cost of the network, \$ d^{-1}.

TOC Transportation operating cost, \$ d^{-1}.

References

Almansoori, A., Betancourt-Torcat, A., 2016. Design of optimization model for a hydrogen supply chain under emission constraints - A case study of Germany. Energy 111, 414–29.

Almansoori, A., Shah, N., 2006. Design and operation of a future hydrogen supply chain: Snapshot model. Chemical Engineering Research and Design 84, 423 –38. doi:https://doi.org/10.1205/cherd.05193.

Almansoori, A., Shah, N., 2009. Design and operation of a future hydrogen supply chain: multi-period model. International Journal of Hydrogen Energy 34, 7883–97.

Almansoori, A., Shah, N., 2012. Design and operation of a stochastic hydrogen supply chain network under demand uncertainty. International Journal of Hydrogen Energy 37, 3965–77.

Almaraz, S.D.L., Azzaro-Pantel, C., Montastruc, L., Boix, M., 2015. Deployment of a hydrogen supply chain by multi-objective/multi-period optimisation at regional and national scales. Chemical Engineering Research and Design 104, 11–31.

Almaraz, S.D.L., Azzaro-Pantel, C., Montastruc, L., Domenech, S., 2014. Hydrogen supply chain optimization for deployment scenarios in the Midi-Pyrénées region, France. International Journal of Hydrogen Energy 39, 11831–45.

Brey, J., 2020. Use of hydrogen as a seasonal energy storage system to manage renewable power deployment in Spain by 2030. International Journal of Hydrogen Energy .

Câmara, D., Pinto-Varela, T., Barbósa-Povoa, A.P., 2019. Multi-objective optimization approach to design and planning hydrogen supply chain under uncertainty: A Portugal study case, in: Computer Aided Chemical Engineering. Elsevier. volume 46, pp. 1309–14.

Cox, B., Bauer, C., Beltran, A.M., van Vuuren, D.P., Mutel, C.L., 2020. Life cycle environmental and cost comparison of current and future passenger cars under different energy scenarios. Applied Energy 269, 115021.

Davies, J., Dolci, F., Klassek-Bajorek, D., OrtizCebolla, R., Weidner, E., 2020. Current status of Chemical Energy Storage Technologies. Technical Report. European Commission. doi:doi:10.2760/280873.

Deb, K., 2001. Multi-objective optimization using evolutionary algorithms. volume 16. John Wiley & Sons.

Deb, K., Agrawal, R.B., 1995. Simulated binary crossover for continuous search space. Complex Systems 9, 115–48.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6, 182–97. doi:10.1109/4235.996017.

Emmerich, M., Beume, N., Naujoks, B., 2005. An EMO algorithm using the hypervolume measure as selection criterion, in: International Conference on Evolutionary Multi-Criterion Optimization, Springer. pp. 62–76.

European Comission, 2019. The European Green Deal. techreport. European Comission. URL: https://ec.europa.eu/info/sites/default/files/european-green-deal-communication_en.pdf.

European Environment Agency, 2018. Greenhouse Gas Emissions from Transport. Technical Report. European Environment Agency. URL: {https://www.eea.europa.eu/data-and-maps/indicators/transport-emissions-of-greenhouse-gases/transport-emissions-of-greenhouse-gases-12}.

Fazlollahi, S., Maréchal, F., 2013. Multi-objective, multi-period optimization of biomass conversion technologies using evolutionary algorithms and mixed integer linear programming (MILP). Applied Thermal Engineering 50, 1504 –13. doi:https://doi.org/10.1016/j.applthermaleng.2011.11.035. combined Special Issues: ECP 2011 and IMPRES 2010.

Guillén-Gosálbez, G., Mele, F.D., Grossmann, I.E., 2010. A bi-criterion optimization approach for the design and planning of hydrogen supply chains for vehicle use. AIChE Journal 56, 650–67. doi:10.1002/aic.12024.

van den Heever, S.A., Grossmann, I.E., 2003. A strategy for the integration of production planning and reactive scheduling in the optimization of a hydrogen supply network. Computers & Chemical Engineering 27, 1813–39.

Hugo, A., Rutter, P., Pistikopoulos, S., Amorelli, A., Zoia, G., 2005. Hydrogen infrastructure strategic planning using multi-objective optimization. International Journal of Hydrogen Energy 30, 1523–34.

IEA, 2019. The Future of Hydrogen. Technical Report. International Energy Agency.

Kim, J., Moon, I., 2008. Strategic design of hydrogen infrastructure considering cost and safety using

multiobjective optimization. International Journal of Hydrogen Energy 33, 5887–96.

Knowles, J., Corne, D., 2002. On metrics for comparing nondominated sets, in: Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600), IEEE. pp. 711–6.

Lin, Z., Ou, S., Elgowainy, A., Reddi, K., Veenstra, M., Verduzco, L., 2018. A method for determining the optimal delivered hydrogen pressure for fuel cell electric vehicles. Applied Energy 216, 183–94.

McKinsey, et al., 2017. Hydrogen scaling up, A sustainable pathway for the global energy transition. Technical Report. Hydrogen Council.

Miettinen, K., 2012. Nonlinear multiobjective optimization. volume 12. Springer Science & Business Media.

Pescador-Rojas, M., Hernández Gómez, R., Montero, E., Rojas-Morales, N., Riff, M.C., Coello Coello, C.A., 2017. An overview of weighted and unconstrained scalarizing functions, in: International Conference on Evolutionary Multi-Criterion Optimization, Springer. pp. 499–513.

Sabio, N., Gadalla, M., Guillén-Gosálbez, G., Jiménez, L., 2010. Strategic planning with risk control of hydrogen supply chains for vehicle use under uncertainty in operating costs: a case study of spain. International Journal of Hydrogen Energy 35, 6836–52.

Setak, M., Feizizadeh, F., Tikani, H., Ardakani, E.S., 2019. A bi-level stochastic optimization model for reliable supply chain in competitive environments: Hybridizing exact method and genetic algorithm. Applied Mathematical Modelling 75, 310 –32. doi:https://doi.org/10.1016/j.apm.2019.05.037.

Sims, R., Schaeffer, R., Creutzig, F., Cruz-Núñez, X., D'agosto, M., Dimitriu, D., Figueroa Meza, M., Fulton, L., Kobayashi, S., Lah, O., et al., 2014. Transport Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical Report. IPCC. URL: https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_chapter8.pdf.

Woo, Y.B., Kim, B.S., 2019. A genetic algorithm-based matheuristic for hydrogen supply chain network problem with two transportation modes and replenishment cycles. Computers & Industrial Engineering 127, 981–97.

Zhang, Q., Li, H., 2007. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. IEEE Transactions on Evolutionary Computation 11, 712–31. doi:10.1109/TEVC.2007.892759.

Zitzler, E., Thiele, L., 1998. Multiobjective optimization using evolutionary algorithms—a comparative case study, in: International Conference on Parallel Problem Solving from Nature, Springer-Verlag. pp. 292–301.