Chapter 12
Application of Large Neighborhood Search to Strategic Supply Chain Management in the Chemical Industry

Pedro J. Copado-Méndez, Christian Blum,
Gonzalo Guillén-Gosálbez, and Laureano Jiménez

Abstract. Large neighborhood search is a popular hybrid metaheuristic which results from the use of a complete technique—such as dynamic programming, constraint programming or MIP solvers—for finding the best neighbor within a large neighborhood of the incumbent solution. In this work we present an application of large neighborhood search to a strategic supply chain management problem from the Chemical industry, namely the configuration of a three-echelon hydrogen network for vehicle use with the goal of minimizing the total cost. Traditionally, these large-scale combinatorial optimization problems have been solved by means of mathematical programming techniques. Our experimental results show that large neighborhood search has the potential to be a viable alternative, especially when the complexity of the problem grows.

12.1 Introduction

Supply chain management (SCM) problems [18, 15] can be classified into strategic, tactical and operational according to the temporal and spatial scales considered in the analysis [7]. In this work we will focus on the strategic level, which deals with decisions that have a long lasting effect on the firm, such as those related with the establishment of new facilities and transportation links between the supply

---

Pedro J. Copado-Méndez · Gonzalo Guillén-Gosálbez · Laureano Jiménez
Departament d'Enginyeria Quimica, Universitat Rovira Virgili, Tarragona, Spain
e-mail: {pedrojesus.copado,gonzalo.guillen}@urv.cat
laureano.jimenez@urv.cat

Christian Blum
ALBCOM Research Group, Universitat Politècnica de Catalunya, Barcelona, Spain
e-mail: cblum@lsi.upc.edu

E.-G. Talbi (Ed.): Hybrid Metaheuristics, SCI 434, pp. 335–352.
springerlink.com

chain entities. Spatially explicit models have recently gained wider interest in SCM. These formulations are particularly suited for strategic SCM problems in which the supply chain performance shows a strong geographical dependence. They give rise to large-scale MILP models with three types of variables: (1) integers representing the number of facilities opened in a given location, (2) binary variables denoting the existence of transportation links between two sub-regions, and (3) continuous variables that quantify the materials flows and inventory levels. In this work we will deal with a spatially explicit model that concerns the strategic planning of hydrogen supply chains for vehicle use [1, 9, 10, 12, 19].

In spatially explicit SCM models a trade-off exists between modelling accuracy and computational burden. Realistic models require the definition of a large number of discrete variables. Mathematical programming is probably the prevalent approach for solving SCM problems. Hereby, decomposition strategies that exploit the mathematical structure of the problem are sometimes used to make the problem tractable. A general review on the application of mathematical programming techniques in SCM can be found in [14], whereas more specific reviews devoted to process industries have been presented in [8, 16]. Apart from mathematical programming, metheuristics have also been applied so strategic SCM problems. In [21], for example, a method to solve the vehicle routing problem (VRP) is proposed that combines genetic algorithms with mathematical programming. The authors of [5] examine the open vehicle routing problem with time windows (OVRPTW) using tabu search. Several evolutionary algorithms for the application fo SCM models have been proposed in [2], while in [6] the authors employed genetic algorithms for solving the coordinated scheduling of production and air transportation. Other applications can be found in [24, 3].

The goal of this work is the application of a popular algorithm from the field of hybrid metaheuristics to the above mentioned SCM problem. Hybrid metaheuristics [4] are algorithms for optimization that combine metaheuristics with components of other techniques for optimization. Examples are combinations of metaheuristics with dynamic programming, contraint programming, and branch & bound. The specific algorithm that is applied in this work is known as *large neighborhood search* (LNS) [17]. The characteristic feature of LNS algorithms is the use of complete techniques for searching large neighborhoods within a metaheuristic framework. Our method, as shown by means of numerical examples, produces near optimal solutions in a fraction of the computational time required by stand-alone deterministic branch and cut techniques applied to the original full-space MILP. The same approach can be easily extended to tackle similar engineering problems with large numbers of discrete decisions, expediting current solution approaches for a certain class of process systems engineering models.

The remainder of this chapter is organized as follows. In Section 12.2, we provide a generic formulation of spatially explicit supply chain models. The full description of the mathematical model of the hydrogen supply chains for vehicle use is given in Appendix A. In Section 12.3 we describe the proposed LNS approach, whereas in

Section 13.5 the experimental results are outlined in detail. Finnally, the conclusions of the work are presented in Section 12.5.

12.2 Spatially Explicit Supply Chain Models

As mentioned before, we address the solution of MILPs resulting from the formulation of spatially explicit models used in SCM. The problem under study can be formally stated as follows (see also Figure 12.1). Given are a set of available production, storage and transportation technologies that can be adopted in different locations of a region in order to fulfill the demand of a product of interest. We are also given economic and environmental data associated with the establishment and operation of these facilities. The goal of the analysis is to determine the optimal supply chain configuration, including the type of technologies selected, the capacity expansions over time, and their optimal location, along with the associated planning decisions that optimize a predefined objective function.

The strategic planning problem presented above can be described in mathematical terms as an MILP of the following form:

$$\min_{x,Y,N} f(x,Y,N)$$

such that

$$h(x,Y,N) = 0$$
$$g(x,Y,N) \leq 0$$
$$x \subset \mathbb{R}, \quad Y \subset \{0,1\}, \quad N \subset \mathbb{Z}^+$$

This generic formulation includes three types of variables: continuous variables x, denoting capacity expansions, production rates, inventory levels and materials flows; discrete variables N, representing the number of transportation units and production and storage facilities opened in a given region; and binary variables Y employed for modelling the establishment of transportation links between two potential locations within the overall region of interest. The inequality and equality constraints, denoted by $g(x,Y,N)$ and $h(x,Y,N)$ respectively, represent mass balances, capacity limitations and objective function calculations. In this work, without loss of generality, we address the solution of a spatially explicit SCM model that was introduced in [13, 11, 9, 20]. The solution of this multi-period model provides the optimal supply chain structure along with the capacity expansions over time required to follow a given demand pattern.

For the sake of brevity, a detailed description of the mathematical model for the hydrogen supply chains problem for vehicle use can be found in Appendix A. Moreover, further details on the complete MILP formulation can be found in the original works. From now on, we will refer to this model as HYDROGEN.

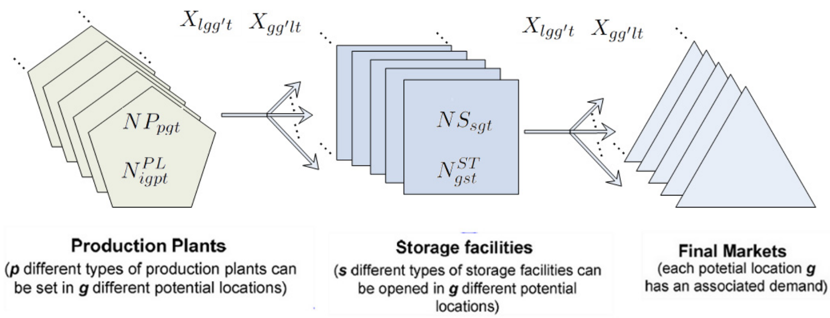

Fig. 12.1 Main sets of decision variables involved in spacially explicit supply chain management

12.3 The Proposed LNS Algorithm

The difficulty in solving model exposed in the previous section is highly dependent on the number of integer and binary variables since they are responsible for the combinatorial complexity of the problem. The number of discrete variables required increases with the number of time periods and sub-regions considered in the model. The MILP can be solved via standard branch-and-cut techniques implemented in software packages such as CPLEX. Models accounting for a large number of time periods and/or sub-regions may lead to branch-and-bound trees with a prohibitive number of nodes, thus making the MILP computationally intractable. We next present a hybrid method, LNS, that combines local search with standard branch and cut for the efficient solution of the tackled problem. LNS was first introduced by [22]. In LNS, an initial solution is gradually improved by alternately destroying and repairing it. This approach combines components from different search techniques, and has many potential applications in the fields of operations research and artificial intelligence. Classifications, taxonomies and overviews on the subject can be found in the work by [4, 23].

All LSN algorithms are based on the observation that searching a large neighbourhood results in finding local optima of high quality. Specifically, LNS decomposes the original problem into a number of smaller sub-problems that are solved in a sequential way. Each sub-problem emerges from a partial solution, in which some decision variables are fixed and others released. A partial solution defines a neighbourhood of solutions that can be explored rather fast by either tailored (e.g., another heuristic or meta-heuristic) or general purpose algorithms (e.g., branch and cut MIP solvers). LNS is a general framework that must be adapted to the particularities of the problem under study. Hence, the definition of the large neighbourhood is highly dependent on the problem of interest. In the simplest case, an appropriate portion of the decision variables is fixed to the values that they have in the current solution, and only the remaining "free" variables are considered by the optimization algorithm (typically, a MIP-solver). If the MIP-solver finds an improved solution, it becomes the new current solution, a new large neighbourhood is defined around it, and the process is repeated in subsequent iterations. Obviously, the selection of the variables that remain fixed and the ones that are subject to optimization, respectively, plays a crucial role in the performance of the algorithm. Particularly, the number of free variables directly defines the size of the neighbourhood. Too restricted neighbourhoods—that is, sub-problems—are unlikely to yield improved solutions, while too large neighbourhoods might result in excessive running times for solving the sub-problems by the MIP-solver. Therefore, a strategy for dynamically adapting the number of free variables is sometimes used. Furthermore, the variables to be optimized might be selected either purely at random or in a more sophisticated guided way by considering the variables with largest potential impact on the objective function and their relatedness. The section that follows describes the main features of our algorithm.

12.3.1 Algorithm

In this section we describe the LNS implementation for our particular problem. The algorithm requires the following input data:

- t_{max}: a maximum execution time of the algorithm;
- n_{max}: a maximum number of variables to be released;
- m_{max}: a maximum number of attempts (the meaning of this parameter is described below).

The algorithm works as follows (see Algorithm 11). First, the initial solution is generated in function *generate_initial_solution*(). The HYDOGENE model includes three main types of discrete variables that are relevant for our algorithm:

- **Integer variables** N_{igpt}^{PL}**:** Number of facilities producing hydrogen in form i using technology p established in location g at period t.
- **Integer variables** N_{gst}^{ST}**:** Number of storage facilities of type s opened in location g at period t.
- **Binary variables** $X_{gg'lt}$**:** Equals 1 if there is a link between g and g' using transportation mode l in period t and 0 otherwise.

The initial solution is generated by solving the HYDROGENE model with the variables NP_{igpt}^{PL}, N_{gst}^{ST} and $X_{gg'lt}$ fixed to the values obtained from a reduced-space model that considers a single time period with a demand equal to the average demands over all the time periods. We have used CPLEX for this purpose. The pseudo-code of this procedure is given in Algorithm 12).

After the generation of the initial solution, the main loop of the algorithm starts. While the maximum computation time limit is not reached, in each trial m the following is done. First, a set V of n variables that are to be released is chosen in

Algorithm 11. LNS for solving the HYDROGENE model

Require: The model HYDROGENE to be solved
  AND $t_{max} > 0$ **AND** $m_{max} > 0$ **AND** $n_{max} > 0$
Ensure: s
$s :=$ generate_initial_solution()
while computation time limit t_{max} not reached **do**
  $n := 1$
  improved := **FALSE**
  while $n \leq n_{max}$ **AND NOT** *improved* **do**
    $m := 1;$
    while $m \leq m_{max}$ **AND NOT** *improved* **do**
      $V := choose\_variables\_to\_be\_released(n)$
      $s' := release\_variables(s, V)$
      $s'' := MIP\_solve(s')$
      if $f(s'' < f(s)$ **then**
        $s := s''$
        improved := **TRUE**
      end if
      $n := n + 1;$
    end while
    $m := m + 1;$
  end while
end while

Algorithm 12. Generating the initial solution

for all g **do**
  $D_g := \frac{\sum_t^T D_{gt}}{T}$
end for
Solve HYDROGENE considering one period $(t = 1)$ with demand D_g
Solve HYDROGENE for all the time periods fixing
$\langle N^{PL}_{igpt}, N^{ST}_{gst}, X_{gg'lt} \rangle := \langle N^{PL}_{igp1}, N^{ST}_{gs1}, X_{gg'l1} \rangle$

function $choose\_variables\_to\_be\_released(n)$. Second, solution s is copied, resulting in solution s'. Next, the n variables from V are released in s'. Third, the CPLEX solver is invoked. The solver determines the best solution that can be obtained on the basis of the partial solution s'. In case $f(s' < f(s)$—where $f(\cdot)$ refers to the value of the objective function—variable *improved* is assigned the value true.

12.4 Experimental Evaluation

In the following subsection we present numerical results that illustrate the performance of LNS as compared to the commercial full-space branch and cut code implemented in CPLEX. We have selected different instances of the HYDROGENE model concerning the number of time periods. More specifically, we tested $t \in \{2, 4, 6, 8, 10, 12, 14, 16\}$. As computation time limits for the resulting eight models of HYDROGENE we chose $\{1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000\}$

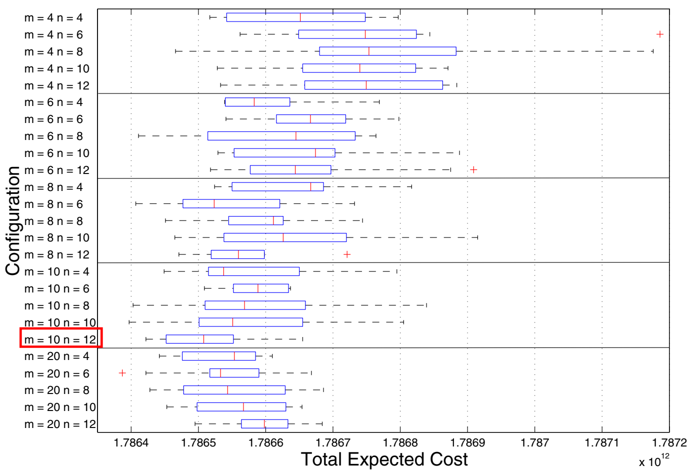

Fig. 12.2 Tuning of algorithm LNS. Note that n and m refer to parameters n_{max} and m_{max}.

seconds. All experiments were performed on a PC Intel (R) Core (TM) Quad CPU Q9550@2.83 GHz and 3 GB of RAM.

12.4.1 Algorithm Tuning

In order to obtain reasonable values for parameters n_{max} and m_{max}, we applied LNS for each combination of n_{max} and m_{max} 10 times to each of the eight different models (resulting from eight different time periods). The values considered for n_{max} are taken from $\{4,6,8,10,20\}$, while the values considered for m_{max} are taken from $\{4,6,8,10,12\}$. The results are shown for each combination of n_{max} and m_{max} in the form of boxplots in Figure 12.2. This is a standard and convenient way of graphically depicting sets of numerical data through their five-number summaries: the smallest observation (sample minimum), lower quartile (Q1), median (Q2), upper quartile (Q3), and largest observation (sample maximum). A boxplot also indicates which observations, if any, are to be considered as outliers. When observing these results, the general impression is that the results become better when m_{max} grows. Concerning n_{max}, no conclusions can be drawn. The final setting that we chose based on these results is marked by a box. In particular, we chose the setting of $n_{max} = 10$ and $m_{max} = 12$.

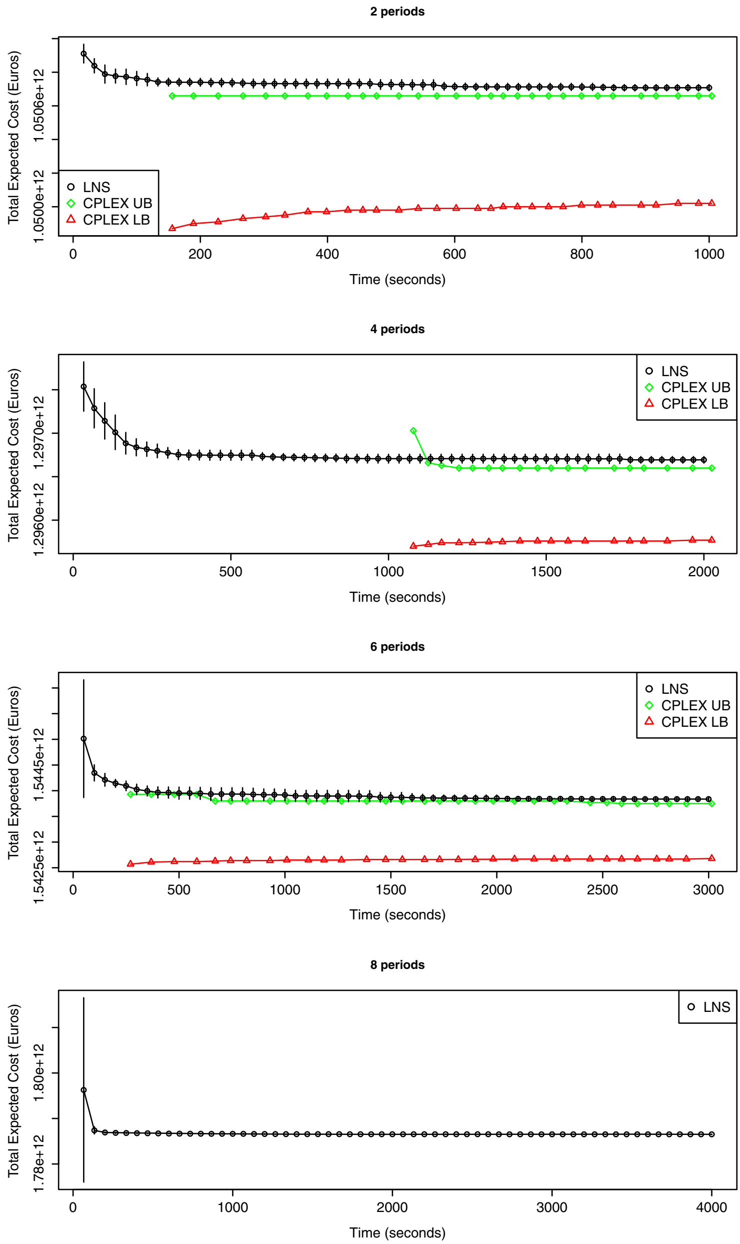

Fig. 12.3 Commparison of LNS with CPLEX (lower and upper bound) over time. The four graphics show the results for a different number of periods (value of t). From top to down, t takes values $\{2,4,6,8\}$). The vertical bars show the standard deviation of LNS over 10 runs. In the cases in which CPLEX results are missing, CPLEX was not able to find any solution within the given time.

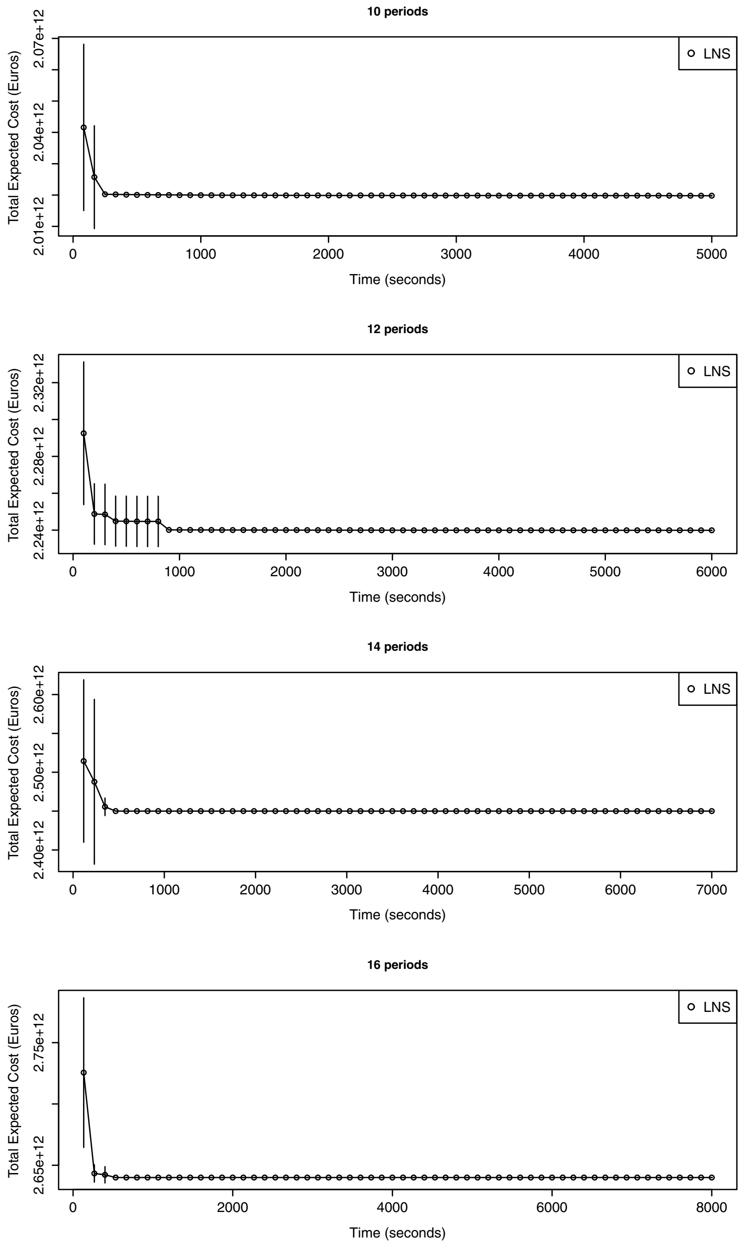

Fig. 12.4 Commparison of LNS with CPLEX (lower and upper bound) over time. The four graphics show the results for a different number of periods (value of t). From top to down, t takes values $\{10, 12, 14, 16\}$). The vertical bars show the standard deviation of LNS over 10 runs. In the cases in which CPLEX results are missing, CPLEX was not able to find any solution within the given time.

Table 12.1 Results of LNS for the eight different HYDROGEN models

# time periods	time limit	avg obj function value	std (obj)	avg comp time	std (time)
2	1000	1.050.707.800.000	18.860.894,29	477,97	248,27
4	2000	1.296.693.500.000	37.997.806,95	748,38	581,41
6	3000	1.543.836.900.000	42.019.704,37	1156,19	784,40
8	4000	1.786.516.900.000	58.333.238,10	1651,62	1196,73
10	5000	2.019.793.600.000	111.899.955,32	2751,97	1657,26
12	6000	2.239.906.300.000	128.670.164,03	2973,20	1508,65
14	7000	2.450.000.000.000	0,0	195,28	125,80
16	8000	2.640.000.000.000	0,0	227,95	116,70

Table 12.2 Optimality gaps of CPLEX and LNS. GAP's are calculated with respect to the best lower bound found by CPLEX when the 12h time limit was applied. **No result** indicates that in the given time CPLEX was not able to obtain any feasible solution.

# time periods	CPLEX (12h)	CPLEX	LNS (avg)	LNS (best)
2	0.05	0.05	0.05	0.05
4	0.05	0.06	0.07	0.06
6	0.06	0.07	0.07	0.07
8	0.08	No result	0.08	0.08
10	0.10	No result	0.09	0.09
12	No result	No result	No result	No result
14	No result	No result	No result	No result
16	No result	No result	No result	No result

12.4.2 Final Comparison

After the above-mentioned tuning procedure we applied CPLEX with the same computation time limits (and additionally with the computation time limit of 12 hours) to all eight HYDROGENE models. Figures 12.3 and 12.4 show—for all eight different time periods—the evolution of the lower and upper bounds found by CPLEX as a function of time, along with the performance of the proposed LNS algorithm. As can be seen, for a rather low number of time periods (up to 6), CPLEX performs slightly better than the proposed algorithm, finding better solutions in shorter CPU times. For more than 6 time periods, CPLEX cannot find any solution within the given time, whereas LNS is always able to provide at least one feasible solution. Note that the variability of the results obtained with our algorithm is rather low.

Numerical results of LNS are shown in Table 12.1. The first column of this table indicates the number of periods, while the second table column states the computation time limit. The four remaining columns contain, respectively, the average of the objective function values of the best solutions found in ten runs, the corresponding standard deviation, the average computation times necessary to obtain these solutions, and, again, the corresponding standard deviation.

Finally, Table 12.2 displays the optimality gaps obtained by the following algorithms: the best solution calculated by CPLEX after 12 hours of CPU time (column labelled **CPLEX (12h)**) and after the same CPU time limit applied to LNS (column labelled **CPLEX**), the average solution quality obtained by LNS (column labelled **LNS (avg)**), and the value of the best solution found by LNS (column labelled **LNS (best)**). The optimality gap is determined with respect to the best lower bound calculated by CPLEX within 12 hours. Note that in some cases (12, 14, and 16 periods), CPLEX is unable to provide any bound even after 12 hours.

Summarizing, we can say that LNS appears to be a useful alternative to pure mathematical programming for what concerns the application to large-scale models from the Chemical industries.

12.5 Conclusions

In this work we have introduced an efficient hybrid algorithm for a spatially explicit supply chain management model. Our algorithm combines mathematical programming techniques with local search, and is known as large neighborhood search in the literature. The capabilities of the proposed method were illustrated through its application to the strategic planning of infrastructures for hydrogen production. Our algorithm was shown to outperform the stand-alone branch and cut method implemented in CPLEX especially for large-scale problems. Numerical examples have demonstrated that our method is particularly suited for tackling large scale problems with a high number of time periods and potential locations (and, therefore, a high number of integer and binary variables).

Future work will particularly focus on investigating how to incorporate the information obtained after solving sub-problems of the mathematical program into the original model in order to expedite the solution of the full space formulation.

Acknowledgements. Pedro J. Copado wishes to acknowledge support of this research work from the Spanish Ministry of Education and Science (DPI2008-04099/DPI). The authors also wish to acknowledge support from the CON- 844 ICET (Argentina), the Spanish Ministry of Education and Science (DPI2008-04099/DPI, CTQ2009-14420 and ENE2008-06687-C02-01), and the Spanish Ministry of External Affairs (projects A/8502/07, A/023551/09 and HS2007-0006). Finally, Christian Blum acknowledges support from grant TIN2007-66523 (FORMALISM) and from the *Ramón y Cajal* program of the Spanish Government.

References

1. Almansoori, A., Shah, N.: Design and Operation of a Future Hydrogen Supply Chain: Snapshot Model. Chemical Engineering Research and Design 84(6), 423–438 (2006)
2. Amodeo, L., Prins, C., Sánchez, D.R.: Comparison of Metaheuristic Approaches for Multi-objective Simulation-Based Optimization in Supply Chain Inventory Management. In: Giacobini, M., Brabazon, A., Cagnoni, S., Di Caro, G.A., Ekárt, A., Esparcia-Alcázar, A.I., Farooq, M., Fink, A., Machado, P. (eds.) EvoWorkshops 2009. LNCS, vol. 5484, pp. 798–807. Springer, Heidelberg (2009)
3. Baykasoglu, A., Gocken, T.: Multi-objective aggregate production planning with fuzzy parameters. Advances in Engineering Software 41(9), 1124–1131 (2010)
4. Blum, C., Puchinger, J., Raidl, G.R., Roli, A.: Hybrid metaheuristics in combinatorial optimization: A survey. Applied Soft Computing Journal 11(6), 4135–4151 (2011)

5. Chiang, W., Russell, R., Xu, X., Zepeda, D.: A simulation/metaheuristic approach to newspaper production and distribution supply chain problems. International Journal of Production Economics 121(2), 752–767 (2009)
6. Delavar, M.R., Hajiaghaei-Keshteli, M., Molla-Alizadeh-Zavardehi, S.: Genetic algorithms for coordinated scheduling of production and air transportation. Expert Systems with Applications 37(12) (2010)
7. Fox, M.S., Barbuceanu, M., Teigen, R.: Agent-oriented supply-chain management. International Journal of Flexible Manufacturing Systems 12(2), 165–188 (2000)
8. Grossmann, I.: Enterprise-wide optimization: A new frontier in process systems engineering. AICHE Journal 51(7), 1846–1857 (2005)
9. Guillén-Gosálbez, G., Mele, F.D., Grossmann, I.E.: A bi-criterion optimization approach for the design and planning of hydrogen supply chains for vehicle use. AICHE Journal 56(3), 650–667 (2010)
10. Kim, J., Lee, Y., Moon, I.: Optimization of a hydrogen supply chain under demand uncertainty. International Journal of Hydrogen Energy 33(18), 4715–4729 (2008)
11. Kostin, A.M., Guillén-Gosálbez, G., Mele, F.D., Bagajewicz, M.J., Jiménez, L.: A novel rolling horizon strategy for the strategic planning of supply chains. Application to the sugar cane industry of argentina. Computers & Chemical Engineering (2010)
12. Li, Z., Gao, D., Chang, L., Liu, P., Pistikopoulos, E.N.: Hydrogen infrastructure design and optimization: A case study of china. International Journal of Hydrogen Energy 33(20), 5275–5286 (2008)
13. Mele, F.D., Kostin, A.M., Guillen-Gosalbez, G., Jimenez, L.: Multiobjective model for more sustainable fuel supply chains. A case study of the sugar cane industry in argentina. Industrial & Engineering Chemistry Research (2011)
14. Mula, J., Peidro, D., Diaz-Madronero, M., Vicens, E.: Mathematical programming models for supply chain production and transport planning. European Journal of Operational Research 204(3), 377–390 (2010)
15. Naraharisetti, P.K., Adhitya, A., Karimi, I.A., Srinivasan, R.: From pse to pse2-decision support for resilient enterprises. Computers and Chemical Engineering 33(12), 1939–1949 (2009)
16. Papageorgiou, L.G.: Supply chain optimisation for the process industries: Advances and opportunities. Computers and Chemical Engineering 33(12), 1931–1938 (2009)
17. Pisinger, D., Ropke, S.: Large Neighborhood Search. International Series in Operations Research & Management Science, vol. 146. Springer, US (2010)
18. Puigjaner, L., Guillén-Gosálbez, G.: Towards an integrated framework for supply chain management in the batch chemical process industry. Computers & Chemical Engineering 32(4-5), 650–670 (2008)
19. Sabio, N., Gadalla, M., Jimnez, L., Guillén-Gosálbez, G.: Risk management on the design and planning of a hydrogen supply chain for vehicle use under uncertainty in production prices: A case study of spain (2009)
20. Sabio, N., Gadalla, M., Guillén-Gosalbéz, G., Jimenez, L.: 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(13), 6836–6852 (2010)
21. Sadjadi, S.J., Jafari, M., Amini, T.: A new mathematical modeling and a genetic algorithm search for milk run problem (an auto industry supply chain case study). International Journal of Advanced Manufacturing Technology 44(1-2), 194–200 (2009)
22. Shaw, P.: Using Constraint Programming and Local Search Methods to Solve Vehicle Routing Problems. In: Maher, M.J., Puget, J.-F. (eds.) CP 1998. LNCS, vol. 1520, pp. 417–431. Springer, Heidelberg (1998)
23. Talbi, E.-G.: A Taxonomy of Hybrid Metaheuristics. Journal of Heuristics 8(5), 541–564 (2002)
24. Warren Liao, T., Chang, P.C.: Impacts of forecast, inventory policy, and lead time on supply chain inventorya numerical study. International Journal of Production Economics 128(2), 527–537 (2010)

Appendix A

In this appendix we provide the complete mathematical model of the three-echelon hydrogen network problem for vehicle use with the goal of minimizing the total cost. Moreover, we give a brief description of its components. Further details may be found in [20, 9, 11].

Notation

Indices

e	scenarios
i	hydrogen form
g	potential locations
l	transportation mode
p	manufacturing technologies
s	storage technologies
t	time period

Sets

$IL(l)$	set of hydrogen forms that can be transported via transportation mode l
$IS(s)$	set of hydrogen forms that can be stored via technology s
$LI(i)$	set of transportation modes that can transport hydrogen form i
$SI(i)$	set of storage technologies that can store hydrogen form i

Parameters

av_l	availability of transportation mode l
cc_{lt}	capital cost of transport mode l in period t
cud_{lt}	maintenance cost of transportation mode l in period t per unit of distance traveled
$\overline{D_{gt}}$	total demand of hydrogen in location g in period t
$distance_{gg'}$	average distance traveled between locations g and g'
$dsat$	demand satisfaction level to be fulfilled
$fuelc_l$	fuel consumption of transportation mode l
$fuelp_{lt}$	price of the fuel consumed by transportation mode l in period t
ge_{lt}	general expenses of transportation mode l in period t
ir	interest rate
$lutime_l$	loading/unloading time of transportation mode l
$\overline{PC_p^{PL}}$	upper bound on the capacity expansion of manufacturing technology p

$\underline{PC_p^{PL}}$ lower bound on the capacity expansion of manufacturing technology p

$\overline{QC_{gg'l}}$ upper bound on the flow of materials between locations g and g' via transportation model l

$\underline{QC_{gg'l}}$ lower bound on the flow of materials between locations g and g' via transportation model l

$\overline{SC_s^{ST}}$ upper bound on the capacity expansion of storage technology s

$\underline{SC_s^{ST}}$ lower bound on the capacity expansion of storage technology s

$speed_l$ average speed of transportat mode l

$tcap_l$ capacity of transport mode l

upc_{igpte} mean value of unit production cost of hydrogen form i produced via technology p in location g in period t in scenario e

$Vupc_{igpte}$ Variance associated to the probability distribution of upc_{igpte}

usc_{igst} unit storage cost of hydrogen form i stored via technology s in location g in period t

$wage_{lt}$ driver wage of transportation mode l in period t

α_{gpt}^{PL} fixed investment term associated with manufacturing technology p installed in location g in period t

α_{gst}^{ST} fixed investment term associated with storage technology s installed in location g in period t

β_{gpt}^{PL} variable investment term associated with manufacturing technology p installed in location g in period t

β_{gst}^{ST} variable investment term associated with storage technology s installed in location g in period t

θ average storage period

τ minimum desired percentage of the capacity that must be used

$prob_e$ occurrence probability of scenario e

Variables

C_{gpt}^{PL} capacity of manufacturing technology p in location g in period t

C_{gst}^{ST} capacity of storage technology s in location g in period t

CE_{gpt}^{PL} capacity expansion of manufacturing technology p in location g in period t

CE_{gst}^{ST} capacity expansion of storage technology s in location g in period t

D_{igt} amount of hydrogen form i distributed in location g in period t

FC_t fuel cost in period t

FCC_t facility capital cost in period t

FOC_{te} facility operating cost in period t in scenario e

GC_t general cost in period t

LC_t labor cost in period t

MC_t maintenance cost in period t

$TPIC$ capital cost of pipelines establishment (euros/km)

UTP unit transportation cost of pipelines (euros/kg day)

$UTCB$ unit transportation cost of ship rental (euros/h kg)

$PICC_t$	pipeline capital cost (euros/yr)
PIC_t	pipeline operating cost (euros/yr)
$TOCB_t$	ship operating cost
N^{PL}_{gpt}	number of plants of type p installed in location g in period t (integer variable)
N^{ST}_{gst}	number of storage facilities of type s installed in location g in period t (integer variable)
N^{TR}_{lt}	number of transportation units of type l purchased in period t (integer variable)
PR_{igpt}	production of hydrogen mode i via technology p in period t in location g
$Q_{igg'lt}$	flow of hydrogen mode i via transportation mode l between locations g and g' in period t
S_{igst}	amount of hydrogen in physical form i stored via technology s in location g in period t
TC_{te}	total amount of money spent in period t for scenario e
TCC_t	total transportation capital cost in period t
TDC_e	total discounted cost for scenario e
TOC_t	transportation operating cost in period t
$X_{gg'lt}$	binary variable (1 if a link between locations g and g' using transportation technology l is established, 0 otherwise)

Equation 12.1 defines the mass balance for the grids considered in the analysis, whereas Equation 12.2 forces the model to fulfill a minimum demand satisfaction level. Equation 12.3 limits the production capacity between lower and upper bounds. Equation 12.4 determines the production capacity in a time period from the previous one plus the expansion in capacity executed in the same period. Equation 12.5 limits the capacity expansions within lower and upper bounds given by the number of facilities opened.

$$\begin{aligned}&\sum_{s\in SI(i)} S_{igst-1} + \sum_{p} PR_{igpt} + \sum_{g'\neq g}\sum_{l} Q_{ilg'glt}\\ &= \sum_{s\in SI(i)} S_{igst} + D_{igt} + \sum_{g'\neq g}\sum_{l} Q_{ilgg'lt} \forall i,g,t\end{aligned} \tag{12.1}$$

$$\overline{D_{gt}} dsat \leq \sum_{i} D_{igt} \leq \overline{D_{gt}} \forall g,t \tag{12.2}$$

$$\tau C^{PL}_{gpt} \leq \sum_{i} PR_{igpt} \leq C^{PL}_{gpt} \forall g,p,t \tag{12.3}$$

$$C^{PL}_{gpt} = C^{PL}_{gpt-1} + CE^{PL}_{gpt} \quad \forall g,p,t \tag{12.4}$$

$$\underline{PC^{PL}_{p}} N^{PL}_{gpt} \leq CE^{PL}_{gpt} \leq \overline{PC^{PL}_{p}} N^{PL}_{gpt} \forall g,p,t \tag{12.5}$$

Equations 12.6 to 12.9 are equivalent to equations 12.3 to 12.5, but apply to warehouses. Particularly, equation 12.6 limits the amount of materials stored to be lower than the existing capacity. Equation 12.7 forces the average inventory level, which is determined from the demand and turnover ratio, to be lower than the existing capacity. Equation 12.8 provides the storage capacity in a time period from the previous one and the expansion in capacity in the previous period, whereas equation 12.9 limits the expansion in capacity between lower and upper limits given by the number of storage facilities installed.

$$\sum_{i \in IS(s)} S_{igst} \leq C^{ST}_{gst} \forall g, s, t \tag{12.6}$$

$$2(\theta D_{igt}) \leq \sum_{s \in SI(i)} C^{ST}_{gst} \forall i, g, t \tag{12.7}$$

$$C^{ST}_{gst} = C^{ST}_{gst-1} + CE^{ST}_{gst} \forall g, s, t \tag{12.8}$$

$$\underline{SC^{ST}_{s}} N^{ST}_{gst} \leq CE^{ST}_{gst} \leq \overline{SC^{ST}_{s}} N^{ST}_{gst} \forall g, s, t \tag{12.9}$$

Equation 12.10 limits the transportation links between lower and upper bounds provided the link is finally established. Equations 12.11 and 12.12 are defined for the construction of pipelines. Equation 12.13 is a logic constraint that makes the formulation tighter. Equations 12.14 and 12.15 avoid the transportation between certain maritime grids, whereas equation 12.16 is a symetric cut. Finally, equations 12.17 to 12.31 allow to determine the cost of the network.

$$\begin{aligned} &\underline{QC_{lgg'}} X_{gg'lt} \leq \sum_{i} Q_{ilgg't} \leq \overline{QC_{lgg'}} X_{gg'lt} \\ &\forall g, g' (g \neq g'), l \in LI(i) \cup NPL, t \end{aligned} \tag{12.10}$$

$$\begin{aligned} &\sum_{t' \leq t+1} \underline{QC_{lgg'}} X_{gg'lt'} \leq \sum_{i} Q_{ilgg't} \leq \sum_{t' \leq t+1} \overline{QC_{lgg'}} X_{gg'lt} \\ &\forall g, g' (g \neq g'), l = pipeline, t \end{aligned} \tag{12.11}$$

$$\sum_{t' \leq t+1} X_{gg'lt'} \leq 1 \qquad \forall g, g' (g \neq g'), l = pipeline, t \tag{12.12}$$

$$X_{gg'lt} + X_{g'glt} \leq 1 \qquad \forall g, g' (g \neq g'), l \in LI(i, t \tag{12.13}$$

$$\begin{aligned} &X_{lgg't} = 0 \qquad \forall l, g, g' \in LG' \\ &LG' = \{l, g, g' : (l = ship) \wedge ((g, g') \notin SGG(gg'))\} \end{aligned} \tag{12.14}$$

$$\begin{aligned} &X_{lgg't} = 0 \qquad \forall l, g, g' \in LG \\ &LG = \{l, g, g' : (l \neq ship) \wedge ((g, g') \in SGG'(gg'))\} \end{aligned} \tag{12.15}$$

$$X_{lgg't} = 0 \qquad \forall l, g = g' \tag{12.16}$$

$$TDC = \sum_t \frac{TC}{(1+ir)^{t-1}} \tag{12.17}$$

$$TC_t = FCC_t + TCC_t + FOC_t + TOC_t \qquad \forall t \tag{12.18}$$

$$\begin{aligned} FOC_t &= \sum_i \sum_g \sum_p upc_{igpt} PR_{igpt} \\ &+ \sum_i \sum_g \sum_s \in SI(i) usc_{igst} \left(\theta D_{igt}\right) \qquad \forall t \end{aligned} \tag{12.19}$$

$$\begin{aligned} FCC_t &= \sum_g \sum_p \left(\alpha^{PL}_{gpt} N^{PL}_{gpt} + \beta^{PL}_{gpt} CE^{PL}_{gpt}\right) \\ &+ \sum_g \sum_s \left(\alpha^{ST}_{gst} N^{ST}_{gst} + \beta^{ST}_{gst} CE^{ST}_{gst}\right) \qquad \forall t \end{aligned} \tag{12.20}$$

$$TCC_t = \sum_{l \neq ship, pipeline} N^{TR}_{lt} \cdot cc_{lt} + PCC_t \tag{12.21}$$

$$PCC(t) = \sum_g \sum_{g' \neq g} \sum_{l \in LI(i)} upcc_t X_{lgg't} distance_{gg'} \qquad \forall t \tag{12.22}$$

$$\sum_{t' \leq t+1} N^{TR}_{lt'} \geq \sum_{i \in IL(l)} \sum_g \sum_{g' \neq g} \sum_t \frac{Q_{igg'lt}}{av_l tcap_l} \left(\frac{2distance_{gg'}}{speed_l} + lutime_l\right) \\ \forall l \neq ship, pipeline \tag{12.23}$$

$$TOC_t = ROC_t + POC_t + SOC_t \qquad \forall t \tag{12.24}$$

$$ROC_t = FC_t + LC_t + MC_t + GC_t \qquad \forall t \tag{12.25}$$

$$FC_t = \sum_i \sum_g \sum_{g' \neq g} \sum_{l \in LI(i)} fuelp_{lt} \frac{2distance_{gg'} Q_{ilgg't}}{fuelc_l tcap_l} \qquad \forall t \tag{12.26}$$

$$\begin{aligned} LC_t &= \sum_i \sum_g \sum_{g' \neq g} \sum_{l \in LI(i)} wage_{lt} \\ &\times \left[\frac{Q_{ilgg't}}{tcap_l} \left(\frac{2distance_{gg'}}{speed_l} + lutime_l\right)\right] \qquad \forall t \end{aligned} \tag{12.27}$$

$$MC_t = \sum_i \sum_g \sum_{g' \neq g} \sum_{l \in LI(i)} cud_l \frac{2distance_{gg'} Q_{ilgg't}}{tcap_l} \qquad \forall t \tag{12.28}$$

$$GC_t = \sum_l \sum_{t' \leq t} g_{lt} N^{TR}_{lt'} \qquad \forall t \tag{12.29}$$

$$POC(t) = \sum_i \sum_g \sum_{g' \neq g} \sum_{l \in LI(i)} upoc_t Q_{ilgg't} \qquad \forall t \tag{12.30}$$

$$SOC_t = \sum_i \sum_g \sum_{g' \neq g} \sum_{l \in LI(i)} usoc_t \left(\frac{distance_{gg'}}{speed_l} \right) Q_{ilgg't} \qquad \forall t \tag{12.31}$$