Comptes rendus du
Troisième atelier de résolution de problèmes industriels de Montréal
Une activité CRM-MITACS

17 au 21 août 2009

Responsable de la rédaction: Odile Marcotte, CRM et UQÀM
Rapport de recherche du CRM no CRM-3315

Proceedings of the
Third Montreal Industrial Problem Solving Workshop
A CRM-MITACS Event

August 17–21, 2009

Editor: Odile Marcotte, CRM and UQÀM
CRM Research Report no. CRM-3315

Contents

Remerciements

J'aimerais exprimer ma reconnaissance au réseau MITACS, qui a été le principal parrain du Troisième atelier de résolution de problèmes industriels de Montréal. La compagnie Pratt & Whitney a aussi apporté une contribution financière à l'atelier. Je suis aussi très reconnaissante aux personnes suivantes, qui nous ont aidés à trouver des problèmes : Catherine Mavriplis, de l'Université d'Ottawa, Fabian Bastin, Anne Bourlioux et Christiane Rousseau (tous trois de l'Université de Montréal), et Louis-Martin Rousseau, de l'École Polytechnique de Montréal. Finalement, je remercie les membres du comité organisateur de leur soutien et de leurs bons conseils. Le comité organisateur était présidé par Jean-Marc Rousseau, Fellow invité CIRANO et PDG du rcm_2, et comprenait Fabian Bastin, Eric Bosco, responsable du développement de MITACS pour le Québec, Michel Gendreau, de l'École Polytechnique de Montréal, Bernard Gendron, directeur du CIRRELT, Roland Malhamé, directeur du GERAD, Odile Marcotte, directrice adjointe du CRM, Catherine Mavriplis et Christiane Rousseau.

Odile Marcotte
Directrice adjointe
Centre de recherches mathématiques

Acknowledgments

I would like to thank the MITACS network, which was the main sponsor of the Third Montreal Industrial Problem Solving Workshop. The CRM also received a grant from Pratt & Whitney Canada in order to help organize the workshop. I am very grateful to the following persons, who helped us find problems for the IPSW: Catherine Mavriplis, from the University of Ottawa, Fabian Bastin, Anne Bourlioux, and Christiane Rousseau (all three from the Université de Montréal), and Louis-Martin Rousseau, from the École Polytechnique de Montréal. Finally, I would like to thank the members of the Organizing Committee for their support and advice. The Organizing Committee was chaired by Dr. Jean-Marc Rousseau, CIRANO Invited Fellow and Chairman of the ncm_2 network, and included Fabian Bastin, Eric Bosco, Director of MITACS Business Development for Québec, Michel Gendreau, from the École Polytechnique de Montréal, Bernard Gendron, Director of CIRRELT, Roland Malhamé, Director of GERAD, Odile Marcotte, Deputy Director of CRM, Catherine Mavriplis, and Christiane Rousseau.

Odile Marcotte
Deputy Director
Centre de recherches mathématiques

Realized Value Optimization in Product Development Post-Certification

Pierre Baptiste and Robert Pellerin
Coordinators, École Polytechnique de Montréal

Yvan Beauregard
Industrial representative, Pratt & Whitney Canada

Mohammed A. Qazi
Professor, Tuskegee University

Adnène Hajji, Mohammad Yousef Maknoon and Benoit Saenz de Ugarte
Students, École Polytechnique de Montréal

Javad Sadr
Report writer, École Polytechnique de Montréal

1. Introduction

With challenging economic conditions, it is more critical than ever to introduce and implement new approaches to help management improve the performance and value delivered by their organizations. It is not easy to develop innovative ways to help organizations understand, measure, manage, and optimize the work of individuals involved in complex engineering activities, such as aerospace product development (PD) [1]. In many regulated sectors, PD efforts usually proceed through different pre-determined phases in order to reach a milestone, represented by the granting of certification from regulatory authorities.

At Pratt & Whitney much effort is spent focusing on timely delivery of quality products within budget in the pre-certification phase, through approaches such as project management and systems engineering [11, 10]. It is not unusual for further engineering resources to be expended in the post-certification phase. These activities are difficult to manage for multiple reasons. First, post-certification activities are often considered to have a lower priority than other activities, and they are not supervised by project managers (who tend to favour the allocation of resources to pre-certification activities). Second, it is difficult to devise an accurate plan for these activities since they are not known in advance. Nevertheless post-certification activities are crucial for Pratt & Whitney for two reasons: they consume a substantial part of the engineering budget and they respond to specific customer needs.

From an operational standpoint, improving the flow of post-certification tasks is crucial for improving the overall PD performance [4, 15, 12, 7]. Arguments have been presented for choosing an appropriately sized work breakdown structure, but no procedure was given for finding optimal size [17]. In this report we use linear programming models to study the influence of diverse factors on PD performance, including the intensity of concurrent PD, the level of multitasking, and the load rate.

From a business standpoint, the strategic and financial value of pre-certification PD activities is generally well understood. Executive attention, decision making, and appropriate processes [5] are available to ensure that resources are available and corporate objectives are met. In the post-certification world, however, the high number of disparate jobs implies that the dimensions of value

are not as well understood or as well coordinated as in the pre-certification context. Also aligning resources is made more difficult by the focus on earnings before interest and taxes.

In the first step, we formulated a simplified version of the actual problem, i.e., a multi-criteria, single-period model with fewer variables and constraints than the actual problem. We obtained an integer linear programming (ILP) problem and solved it by using the Xpress-Mosel solver. The ILP model is presented below. This step helped us gain a better understanding of the problem and we were then able to formulate a multi-criteria, multi-period model incorporating the various dimensions of post-certification decision making. A case study shows that if one considers the multi-faceted dimensions of value in engineering post-certification activities, then one can achieve a greater value (from the resource allocation point of view) than that provided by a throughput maximization approach. At the end of this report, we discuss implementation issues and the results obtained. Note that before reporting on the work carried out during the workshop, we include a section on the background of our study (see the next section).

2. Post-certification, lean engineering performance, and decision making models

The stage-gate process, introduced by Cooper, is a popular decision making approach in the field of portfolio management [21]. According to some widely used project management standards [14], this decision making process can be used (to some extent) even after the product is delivered. For instance, before the PD project is closed, post-mortem reviews are conducted in order to come to conclusions that can be used in future projects. After certification is obtained, one must carry out the remaining activities and several types of jobs compete for limited post-certification engineering resources.

Work on post-certification activities can represent a sizeable portion of the engineering budget. Early detection of problems in PD is less expensive than late detection [20]. Accordingly, much research has been carried out on factors that can improve the PD front end [19]. The nature of post-certification activities and their performance, however, have not received much attention. In the post-certification phase it is difficult to improve continuously the use of the resources. This is so because the tasks are varied, because we lack information about the ways in which value is created, and because the engineers are allocated to new PD projects. Thus "managers have greater difficulty measuring, managing and optimizing work" [1]. Figure 1 illustrates all the steps in the manufacturing of a product (from the step where an order is accepted to the maintenance stage). In Figure 1 post-certification occurs between steps 4 and 5 (which represent engineering tasks in an aerospace manufacturing firm). Figure 2 illustrates the definition of engineering tasks at Pratt & Whitney Canada.

Benchmarking exposes participants to new ideas and encourages them to improve their practices continuously and to be aware of the best ones [3]. The post-certification taxonomy has been developed to help ascertain the source of post-certification work and to ensure the alignment and consistent classification of post-certification engineering PD jobs. The classification scheme of post-certification engineering tasks is influenced by factors such as the origin of need, the clarity and completeness of requirements defined during the fuzzy front-end requirements phase, the effectiveness of the PD process in delivering the expected performance level, and the compliance with engineering PD standards. Jobs are classified into the following 6 categories, according to the above-mentioned factors: pre-certification, product repositioning, product improvement, post-certification, new learning/best practices, and quality.

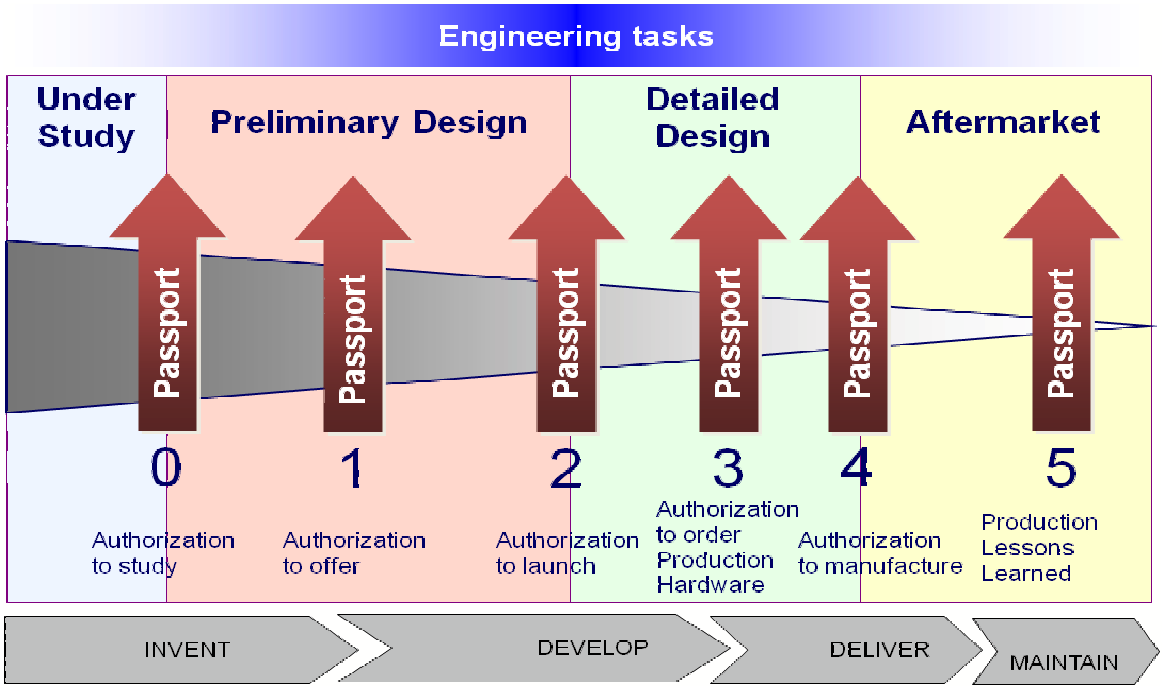


Figure 1. Pre and post-certification sequences for a given project at P&WC.

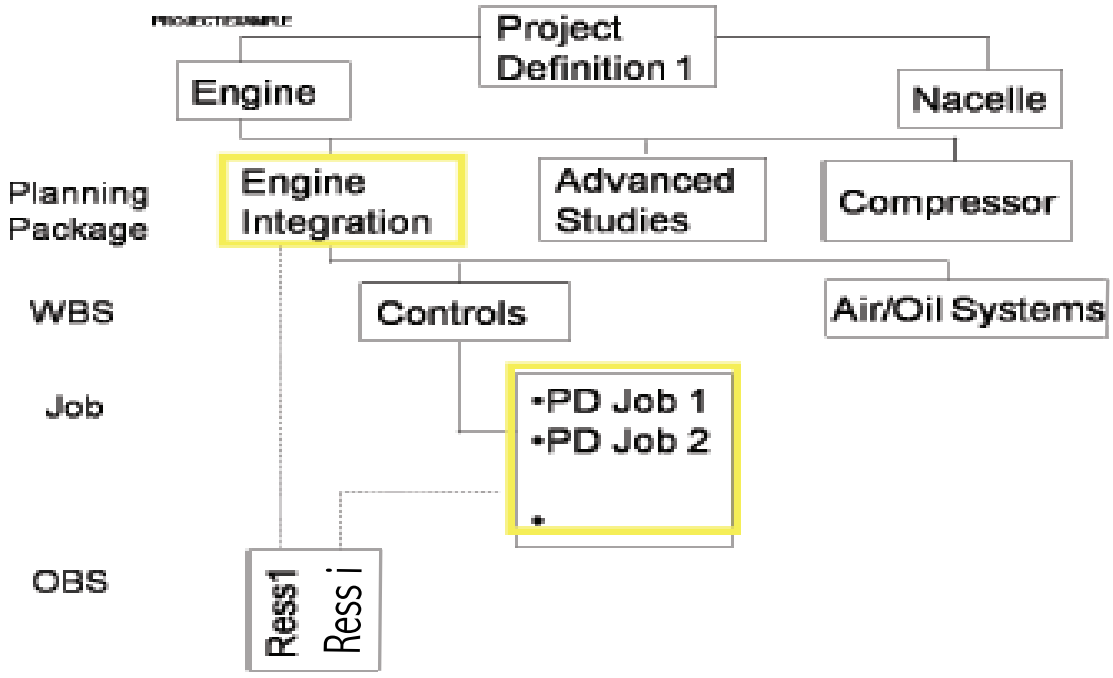


Figure 2. Definition of engineering tasks.

"Lean engineering" is a phrase that was introduced in the 1990's at MIT in order to describe the Japanese production system, where less effort and a lesser use of space and materials are combined with higher output and quality [11]. Womack describes lean thinking as being comprised of the following steps: define value, define value stream, remove barriers to flow, enable pull, and strive for continuous improvement [21]. PD performance can be measured along different dimensions. Enhancing the flow of information in PD is a key goal being actively pursued [3, 18, 12, 6]. McManus referred to the concepts of "better, faster, and cheaper" to describe lean product development [10]. Notions of effectiveness and efficiency have been proposed to measure respectively the ability of PD activities to meet product requirements and the productivity with which these activities are carried out [8]. Browning expanded the notion of PD performance to a value-based PD performance model that takes into account product availability, product affordability, and the right technical performance [4].

An effective PD project is one that "arrives at a new and unique solution that achieves the requirements/specifications of the project" [8]. Lean product development consists of providing the right product through efficient processes and effective integration [10]. While the ultimate success of a PD endeavor may be measured by marketplace response, engineering management requires a more immediate measurement of PD efficiency. Understanding the influence of key PD process parameters on flow performance is part of Womack's lean thinking approach [21]. Given its influence on work in progress, lead time LT_j (where j denotes a specific job) is used to compute an estimator of the flow, denoted by $\hat{F}_s$ (where s is a state variable representing pre-certification or post-certification); LT_j provides a measure of the value stream improvement potential [2], used for comparing post-certification and pre-certification job performances. Here is the formula for $\hat{F}_s$, where E_j (resp. S_j) denotes the end date (resp. start date) of job j.

$$\hat{F}_s = \sum_{j=1}^{m} LT_j = \sum_{j=1}^{m} (E_j - S_j), \qquad j = 1, 2, \ldots, m.$$

Figure 3 illustrates the Pratt & Whitney Canada engineering planning cycle. Effective managerial decisions begin with consideration of the multiple dimensions of value [9]. Explicit decision factors are incorporated into a decision model for improving decision making consistency. An aggregate value index is construced by combining various attributes for each post-certification task into an ordinal scale of value. The benefits of this so-called multi-attribute value index approach include simplicity, consistency of decision making across decision makers, and a way of conveying to engineering personnel the priorities of different jobs.

3. Notations, definitions, and the two mathematical models

In this section we give all the definitions and notations necessary to describe the mathematical models. Then we present the models themselves. Here is a description of the problem to be modelled.

1. Ensure that the project demands correspond to available resources.
2. Optimize project budgets in order to increase shareholder value.
3. Allocate available resources to jobs so as to enhance the value delivered.

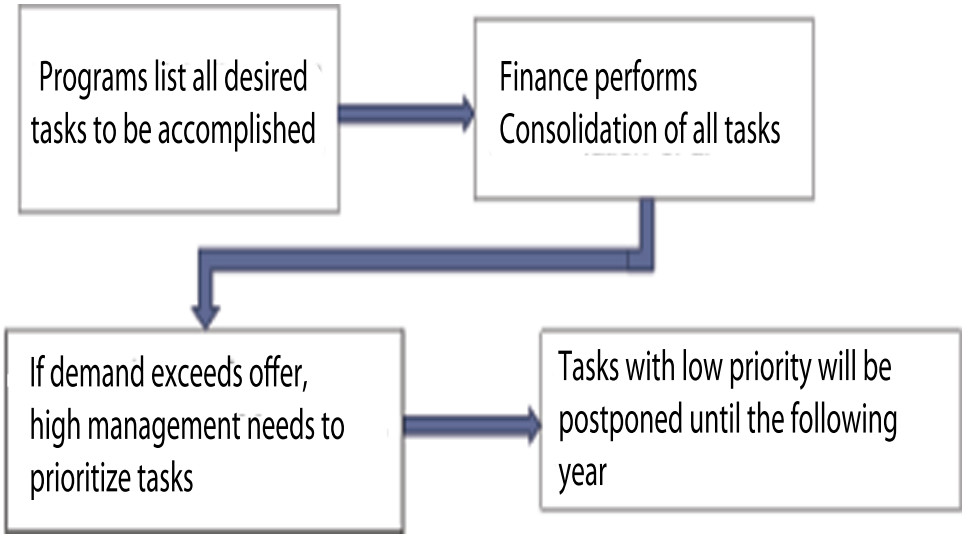


Figure 3. P&WC engineering planning cycle.

4. Increase the frequency of planning exercises while expending less effort.

3.1. Mono-period model

We first consider, within a multi-project problem, a specific period of the planning horizon for PD jobs. This period corresponds to one quarter (i.e., three months) of the activity of the firm. We seek to maximize the realized value. The constraints 1 reflect the limited availability of each resource type. Constraint 2 expresses the fact that the total budget is fixed and cannot be exceeded. The constraints 3 introduce some flexibility within the budget of a specific project. If a project has a high priority, its budget can be increased but that will result in a budget reduction for other projects (see constraint 2). Each of the constraints 4 imposes a lower bound on the consumption of resources by a given activity type.

Definitions of sets

T : set of tasks (or jobs or activities)

K : set of resource types

J : set of value criteria

H : set of activity types

P : set of projects

Parameters

D_{ik} : quantity (in hours) of resource type $k \in K$ required to complete task $i \in T$

C_k : availability (in hours) of resource type $k \in K$

V_i : value of task $i \in T$

B_p : budget (in hours) for project $p \in P$

δ_{ip} : equals 1 if task $i \in T$ is part of project $p \in P$ and 0 otherwise

δ_{ih} : equals 1 if task $i \in T$ is an activity of type $h \in H$ and 0 otherwise

α_h : minimum rate at which a task of type $h \in H$ has to be completed

β_p : maximum proportion of project p budget (for $p \in P$) that must be consumed

γ_p : minimum proportion of project p budget (for $p \in P$) that must be consumed

Decision variables

O_i : equals 1 if we decide to complete task $i \in T$ and 0 otherwise

Objective Function

$$\max \sum_{i \in T} V_i \cdot O_i$$

Constraints

1. Availability constraints
$$\sum_{i \in T} D_{ik} \cdot O_i \leq C_k \qquad \forall k \in K$$

2. Total budget constraint
$$\sum_{i \in T, k \in K} D_{ik} \cdot O_i \leq \sum_{p \in P} B_p \qquad \forall k \in K$$

3. Project budget constraints
$$\gamma_p \cdot B_p \leq \sum_{i \in T, k \in K} \delta_{ip} \cdot D_{ik} \cdot O_i \leq \beta_p \cdot B_p \qquad \forall p \in P$$

4. Activity types constraints
$$\sum_{i \in T, k \in K} \delta_{ih} \cdot O_i \cdot D_{ik} \geq \alpha_h \left(\sum_{i \in T, k \in K} \delta_{ih} \cdot D_{ik} \right) \qquad \forall h \in H$$

5. Nonnegativity constraints
$$O_i \in \{0, 1\} \qquad \forall i \in T$$

Managerial decision making during post-certification involves a number of criteria for prioritizing jobs as well as allocating limited post-certification budgets and engineering resources to activities. Multi-attribute value theory (MAVT) is based on the assumption that decision makers attempt to maximize an implicit value function, denoted $V(.)$. Many definitions of value have been proposed over time (see for instance [6, 13, 16]). Making sure that the decision makers have common value

criteria for prioritizing jobs and allocating resources is a challenge in PD projects [8]. Our objective is to maximize the realized value of all tasks.

This model was a preliminary phase of our studies. Since our data were not very complicated, we were able to use the integer programming algorithm of the Xpress-Mosel solver. The run time was a fraction of a second for an instance with 12 projects, 377 tasks, and 26 resources types. We could not compare the optimal solution returned by the solver with the current solution since the latter was not available. Our experiment, however, showed that the proposed model could be easily implemented within Pratt & Whitney's planning process. A second experiment was conducted with a disaggregated instance, yielding a much larger data set (4000 tasks). Once again, the proposed model was easily solved (within 11 seconds). This led us to consider a more complex model including several periods. This model is described in the next subsection.

3.2. Multi-period model

While the goal of the first model was to choose a group of tasks maximizing the total value under some constraints, the goal of the second model is to choose the numbers of hours dedicated to certain tasks during specific periods in order to maximize the total value. Hence the model includes the variables X_{ikq}, where X_{ikq} denotes the number of worked hours for task $i \in T$, resource $k \in K$, and period $q \in Q$. The connection between X_{ikq} and the value generated, however, is not straightforward, because the latter cannot exceed the number of estimated hours for task $i \in T$, resource $k \in K$, and period $q \in Q$ (denoted E_{ikq}). We thus have to introduce the variable Y_{ikq}, defined as the minimum of X_{ikq} and E_{ikq}. We also have to take into account a penalty for worked hours corresponding to a given combination of task, resource, and period.

Definitions of sets

T : set of tasks (or jobs or activities)

K : set of resource types

J : set of value criteria

H : set of activity types

P : set of projects

Q : set of periods

Parameters

E_{ikq} : estimated quantity (in hours) of resource type $k \in K$ required to complete task $i \in T$ in period $q \in Q$

C_{kq} : availability (in hours) of resource type $k \in K$ in period $q \in Q$

V_{iq} : value of task $i \in T$ in period $q \in Q$

B_{pq} : budget (in hours) for project $p \in P$ in period $q \in Q$

δ_{ip} : equals 1 if task $i \in T$ is part of project $p \in P$ and 0 otherwise

δ_{ih} : equals 1 if task $i \in T$ is an activity of type $h \in H$ and 0 otherwise

α_{hq} : minimum rate at which a task of type $h \in H$ has to be carried out in period $q \in Q$

β_{pq} : maximum proportion of project p budget (for $p \in P$) that must be consumed in period $q \in Q$

γ_{pq} : maximum proportion of project p budget (for $p \in P$) that must be consumed in period $q \in Q$

P_{iq} : penalty for task $i \in T$ in period $q \in Q$

Decision variables

X_{ikq} : number of worked hours for task $i \in T$ and resource type $k \in K$ in period $q \in Q$

Y_{ikq} : earned value for task $i \in T$ and resource type $k \in K$ in period $q \in Q$

Objective Function

The objective function of the multi-period model reflects two goals: the first one is to maximize the total value of all tasks in all periods and the second one to prevent tasks from being completed too late.

$$\max \left(\sum_{i \in T, q \in Q} V_i \cdot \left(\sum_{k \in K} Y_{ikq} \right) \right) - \sum_{i \in T, q \in Q} V_i \cdot P_{iq} \cdot \left(\sum_{k \in K} X_{ikq} \right)$$

Constraints

1. Availability constraints
$$\sum_{i \in T} X_{ikq} \leq C_{kq} \qquad \forall (k, q) \in K \times Q$$

2. Definition of Y_{ikq}
$$Y_{ikq} \leq E_{ikq} \text{ and } Y_{ikq} \leq X_{ikq} \qquad \forall (i, k, q) \in T \times K \times Q$$

3. Constraint on the number of worked hours
$$\sum_{q \in Q} X_{ikq} = \sum_{q \in Q} E_{ikq} \qquad \forall (i, k) \in T \times K$$

4. Overall budget constraint
$$\sum_{i \in T, k \in K} X_{ikq} \leq \sum_{p \in P} B_{pq} \qquad \forall q \in Q$$

5. Project budget constraints
$$\gamma_{pq} \cdot B_{pq} \leq \sum_{i \in T, k \in K} (\delta_{ip} \cdot X_{ikq}) \leq \beta_{pq} \cdot B_{pq} \qquad \forall (p, q) \in P \times Q$$

6. Activity types constraints

$$\sum_{i\in T,k\in K} \delta_{ih} \cdot X_{ikq} \geq \alpha_{hq} \left(\sum_{i\in T,k\in K} \delta_{ih} \cdot E_{ikq} \right) \qquad \forall (h,q) \in H \times Q$$

7. Nonnegativity constraints

$$X_{ikq} \geq 0 \text{ and } Y_{ikq} \geq 0 \qquad \forall (i,k,q) \in T \times K \times Q$$

We used the Xpress-Mosel solver to implement this model and find an optimal solution for an instance of 4000 tasks, 26 resources, and 12 projects. The model included 1664000 variables. For instances of that size, attempting to solve an integer programming model is usually time-consuming and offers no guarantee of optimality. In our case, however, a very good result was obtained within 34 seconds.

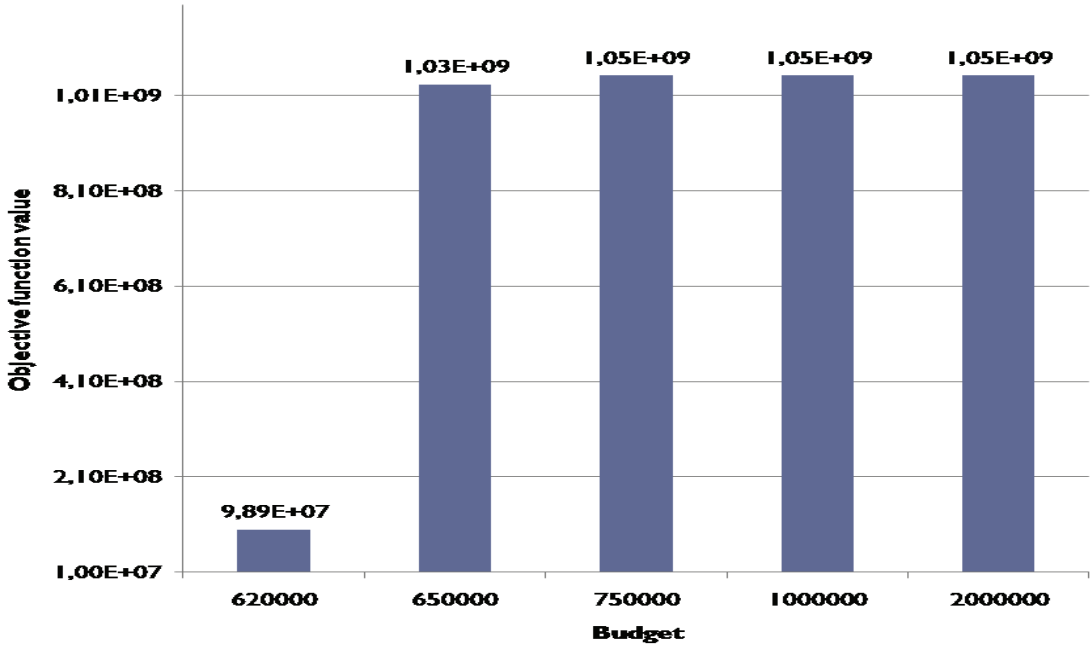


Figure 4. The objective function value vs the total budget.

Figure 4, based on data provided by Pratt & Whitney, shows that the overall budget and project budget constraints are more constraining than the availability and activity types constraints. The objective function attains its maximum value when the overall budget is slightly above 620000.

As explained above, the optimal solution obtained cannot be compared to the current plan, which is not available. This experiment, however, shows that the multi-period model can be implemented within the company planning process (since the solution time is short). This suggests that multiple scenarios (including different priorities and budgets) could be studied within a short period. This is important since results appear to be highly sensitive to the budget allocated to post-certification tasks.

4. Conclusion

In this article, we reviewed the post-certification phase of the PD process, post-certification taxonomy, lean engineering, and decision making models. Then we introduced the concept of lean PD benchmarking for the study of engineering activities (post-certification as well as pre-certification activities). Two mathematical models were presented and applied to two instances, namely a simplified instance of the original problem and the original problem itself. We demonstrated that our approach was worthwhile by implementing and solving the two models through the Xpress-Mosel solver.

Despite the quality of the results described in this article, further work is required before addressing the implementation issues. First, the multi-period model should be tested at the activity level (i.e., the lowest level) in order to identify the maximum number of tasks that can be sequenced within a reasonable time. Obviously, the allowable resolution times must be defined by the Pratt & Whitney managers. Such a model would simplify the data collection procedure since many post-certification jobs are already defined at the activity level. Furthermore, Pratt & Whitney Canada would like to enrich the proposed models by introducing stochastic elements within them. The participants discussed solution methodologies for the stochastic models, in particular: (1) a hybrid solution based on a genetic algorithm and a simulation model, and (2) a stochastic gradient descent approach. Because of time constraints these approaches were not implemented during the workshop.

Future work will involve the development of an enhanced simulation model. In this model the budget could be a stochastic variable. A hybrid model could also be developed by including priorities among jobs and using a sequential scheduling approach to evaluate the objective function. A meta-heuristic approach might be suitable for industrial size problems. Further research should be carried out to determine optimal PD process parameters, in particular the optimal job size.

References

1. Aral, S., Brynjolfsson, E., Van Alstyne, M., 2007, Information, technology and information worker productivity: task level evidence, *National Bureau of Economic Research*, Massachusetts.
2. Beauregard, Y., Thomson, V., Bhuiyan, N., 2008, Lean engineering logistics: load leveling of design jobs with capacity considerations, *Canadian Aeronautics and Space Journal*, Vol. 54, No. 2, pp. 19-33.
3. Beitz, A., Wieczorek, I., 2004, *Applying benchmarking to learn from best practices*, Springer, Berlin.
4. Browning, T., 2000, Value based product development: refocusing lean, *Proceedings of the IEEE Engineering Management Society Conference*, August 13-15, Albuquerque, New Mexico, pp. 168-172.
5. Cooper, R.G., Edgett, S.J., Kleinschmidt, E.J., 2001, *Portfolio management for new products*, 2nd edition, Perseus, Massachusetts.
6. Dowden, T.D., 2005, Exercising a multi-attribute value method for business airplane product assessment, *43rd AIAA Aerospace Sciences Meeting and Exhibit - Meeting Papers*, Jan. 10-13, Reno, Nevada, pp. 8757-8769.
7. Hines, P., Found, P., Griffiths, G., Harrison, R., 2008, *Staying lean: thriving, not just surviving*, Lean Enterprise Research Centre, Cardiff.
8. Kratzer, J., Gemunden, H.G., Lettl, C., 2008, Balancing creativity and time efficiency in multi-team R&D projects: the alignment of formal and informal networks, *R&D management*, Vol. 38, No. 5, pp. 538-549.
9. Mavrotas, G., Panagiotis, T., 2005, Multi-criteria decision analysis with minimum information: combining DEA with MAVT, *Computers & Operations Research*, Vol. 33, No. 2006, pp. 2083-2098.
10. McManus, H., Haggerty, A., and Murman, A., 2005, Lean Engineering: Doing the right thing right, *1st International Conference on Innovation and Integration in Aerospace Sciences*, August 4-5, Queen's University, Belfast, North Ireland, United Kingdom.

11. Murman, E., Allen, T., Bozdogan, K., Cutcher-Gershenfeld, J., McManus, H., Nightingale, D., Rebenstish, E., Shields, T., Stahl, F., Walton, M., Warmkessel, J., Weiss, S., Widnall, S., 2002, *Lean enterprise value: insights from MIT's lean aerospace initiative*, Palgrave, New-York.
12. Oppenheim, B., 2004, Lean product development flow, *Systems engineering*, Vol. 7, No. 4, pp. 352-376.
13. Park, R.J., 1998, *Value engineering: a plan for invention*, St. Lucie Press, London.
14. Project Management Institute, 1996, *A guide to the Project Management Body of Knowledge*, Project Management Institute Standards Committee.
15. Reinertsen, D., 2007, Rethinking lean NPD, *Strategic directions*, Vol. 23, No. 10, pp. 32-34.
16. Slack, R., 1999, The lean value principle in military aerospace product development, *MIT Lean Aerospace Initiative Report Series* RP99-01-16, MIT, Cambridge, Massachusetts.
17. Storch, R.L., 1999, Improving flow to achieve lean manufacturing in shipbuilding, *Production planning and control*, Vol. 10, No. 2, pp. 127-137.
18. Taylor, D.H., 2005, Value chain analysis: an approach to supply chain improvements in the agri-food chain, *International journal of physical distribution and logistics management*, Vol. 35, No.10, pp. 744-761.
19. Walton, M.A., 1999, Strategies for lean product development, *Lean aerospace initiative*, MIT, Massachusetts.
20. Wirthlin, J.R., 2000, *Best practice in user needs/requirements generation*, Master's Thesis, MIT.
21. Womack, J.P., Jones, D., 2003, *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*, revised and updated, Free Press, New York.

Minimum Energy Requirements of the Distillation Process

Étienne Ayotte-Sauvé
Adjunct coordinator, CANMET Energy Technology Centre, Varennes, Natural Resources Canada

Fabian Bastin
Coordinator, Université de Montréal

Angelo Lucia
Professor, University of Rhode Island

Julio Montecinos
Ph.D. Student, École Polytechnique de Montréal-CIRRELT

Shouraz Mirzalizadeh
Ph.D. Student, École Polytechnique de Montréal

1. Introduction

Most of the machinery present in a generic chemical plant (e.g., an oil refinery, a pharmaceutical or polymer factory) aims at purifying materials. As a consequence, a large part of the energy use in many industrial sectors can be attributed to separation processes. These processes involve separating a mixture containing many chemical species into products of specified purity. For example, a petrochemical plant must separate a mixture composed of 50% propane and 50% propylene to sell propane and propylene products at 99% purity. Of all separation technologies available today, distillation dominates the energy demand in many industries. For example, in the U.S., about 18% of the energy use in the manufacturing sector can be attributed to distillation [4].

Since distillation has high energy requirements and many separation needs continue to be addressed with this industrial process, both engineering and industrial communities (including CANMET Energy and the company NOVA Chemicals) have developed an interest in the study of the minimum energy required to accomplish a given distillation task. Methods to determine these energy requirements have gained significant importance in both the engineering and the industrial communities. The recent rise in energy prices and demands further emphasizes the relevance of this problem. Techniques to determine minimum energy requirements for the design of the distillation process usually belong to one of the following two categories: mathematical programming or heuristic approaches based on engineering and thermodynamical insights. For a review of those methods, see the work of Lucia *et al.* [9] and Koehler *et al.* [8].

In this work, we focus on distillation line approaches, that is, mathematical programming and heuristic methods where the properties of liquid and vapour streams along a column are modelled as a discrete dynamical system (see Section 3). Since such approaches may overestimate the minimal energy consumption because of estimation errors on product compositions (see [4]), we investigate the properties of fixed point approaches to minimize these errors.

The next section presents an overview of the distillation process and the minimal energy problem in broad terms. In Section 3, distillation line methods are introduced, and in Section 4, results of our investigations are presented.

2. The distillation process

In a nutshell, distillation is a process by which a mixture is separated into products of specified purity via the addition and removal of heat. Before describing the distillation process in detail, consider the following simple process that attempts to separate a liquid solution composed of two chemical species having different boiling temperatures. In order to separate these two components, one can take advantage of their difference in boiling temperatures by supplying heat to the liquid until it boils. As a result, two phases will coexist: the boiling liquid and an evaporated vapour phase. When compared to the original mixture, the obtained vapour is richer in the component that has a lower boiling temperature (i.e., the most volatile component) while the remaining liquid is richer in the other component (i.e. the less volatile component).

Like simple boiling, the distillation process takes advantage of the difference in volatilities between the various mixture components. Loosely stated, a distillation column creates from the original mixture two streams of matter, one in the vapour phase, and another in the liquid phase, and forces an exchange of matter between these two streams by putting them in counter-current contact (i.e., these streams flow up and down the column respectively, as illustrated in Figure 1). At incremental heights along the column, mass exchange occurs between the vapour and liquid streams until they reach specified purities (designated as products 1 and 2 in Figure 1). The following paragraphs describe the process in greater detail.

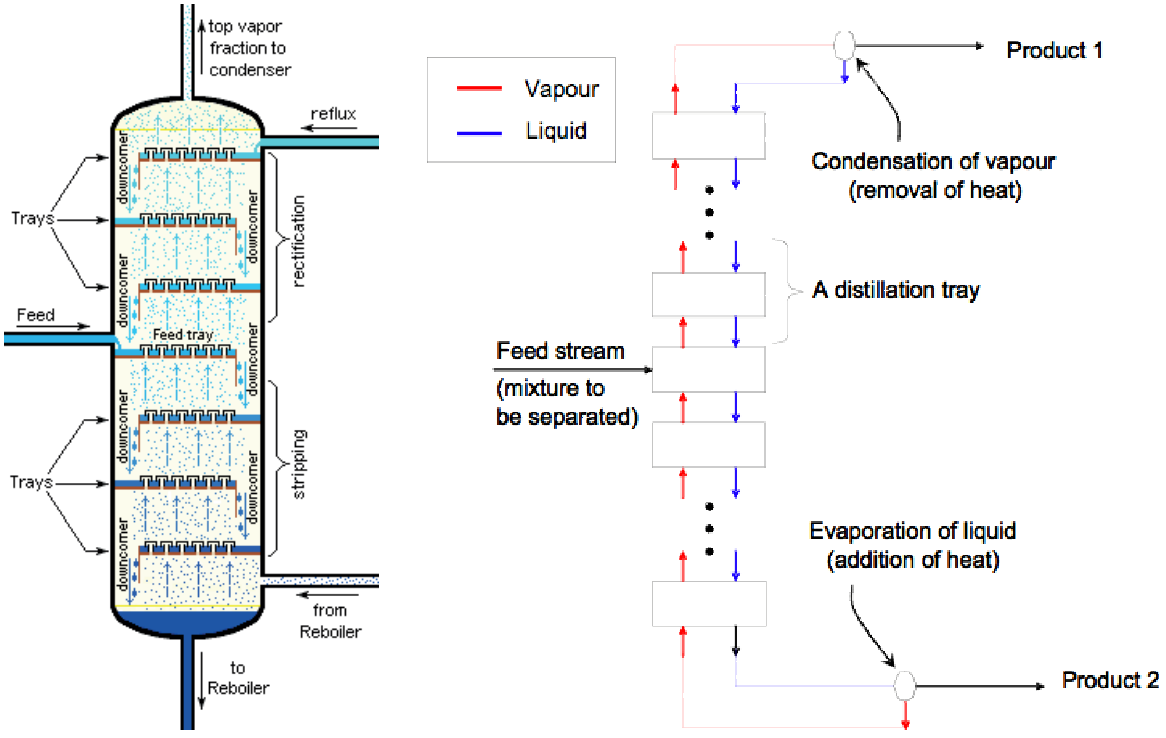


Figure 1. Distillation column

First, a stream containing a mixture of chemical components to be separated is fed to the distillation column. After a short transient phase, the process reaches steady-state operation (i.e., from that point on, the properties of the process don't change with time). The following description and our projected work will focus on the steady-state operation of a distillation column with only

one input stream and two output streams (termed products), obtained at the bottom and at the top of the column. Between these output streams is an arrangement of "steps" (termed trays or stages), which incrementally separate the mixture.

On a distillation tray, a liquid stream, going down the column, and a vapour stream, going up the column, come into contact. As a consequence, an exchange of matter between the two streams entering the tray occurs. The "hot" vapour is partially condensed by the "cold" liquid, while the liquid is partially evaporated by the vapour stream. It follows that the vapour stream exiting the tray is enriched in the more volatile components of the mixture, i.e., the components that evaporate at a lower temperature, while the liquid exiting the tray is enriched in the less volatile components of the mixture, i.e., those having a higher boiling temperature. Going up the column, the vapour stream's composition (i.e., molar fraction) in more volatile components increases at each tray.

Note that the molar fraction of a component in a stream simply describes the proportion of that component present in the stream (on a molar basis). At the top of the column, the vapour stream is condensed (either totally or partially) through a condensing unit, and part of the liquid obtained is returned to the column, while the remainder constitutes the top product of the column (product 1 in Figure 1). That product has a higher purity in more volatile components than the feed stream entering the column. On the other hand, the molar fraction of less volatile components in the liquid stream increases as it comes down the column. At the bottom of the column, that liquid stream is (totally or partially) evaporated through an evaporation unit and part of the vapour obtained is returned to the column, while the remainder constitutes the bottom product of the column (product 2 in Figure 1). That product has a higher purity in less volatile components than the feed stream entering the column.

We will restrict the scope of our work to adiabatic distillation columns, that is, columns where heat is added (respectively removed) only at the bottom (respectively top) of the column. Other columns (diabatic) permit the addition or removal of heat on each column tray. Arguably these (diabatic) columns have received until now little attention by the industry and therefore are not treated here. In the distillation columns we will consider, the energy consumption is thus solely attributed to the addition and removal of heat at the bottom and at the top of the column.

Having reviewed the basic notions underlying the distillation process, we have the tools needed to formulate (in a broad manner) the problem of determining the minimal energy requirements of a distillation column.

Given the specifications of the feed stream and of the two expected products, determine the minimal energy necessary to carry out the given separation with a distillation column.

The specifications of the feed and product streams may include their phase (liquid, vapour, or both, in given proportions), their pressure, their flow rate, their temperature, and their composition (i.e., percentage of each component of the mixture present in the stream). A stream is determined by the aforementioned data. The goal is to determine, before actually building a distillation column, the minimal energy that should be removed at the top of the column and added at its bottom to carry out the specified separation task.

This problem raises various questions. One has to choose the number of trays and the tray on which the feed stream should enter. One also has to establish the proportion of the vapour exiting the evaporation unit (at the bottom of the column) and the proportion of the liquid exiting the condensing unit (at the top) that should be fed back to the column. This directly impacts the energy consumption in the column since the quantity of heat that we need to add or remove depends on the proportions of re-circulated vs. output flow rates (i.e., how much liquid should be evaporated

or how much vapour should be condensed). Note also that the number of trays needed in the column depends on its energy consumption. Intuitively, consuming more energy translates into a smaller number of steps to carry out the separation (and vice versa). The following section presents commonly used solution strategies to tackle the problem; they use mathematical programming or approaches based on engineering.

From a mathematical programming point of view, this problem can be formulated in a natural way as a mixed integer nonlinear optimization problem (MINLP) [2]. The integer part of the problem originates from the choice of the number of distillation trays and the position of the feed stream along the column. The real variables that describe the system are the parameters necessary to define the streams entering and leaving trays in the column (flow rate, pressure, temperature, molar fractions). The distillation process is usually described by a set of equations corresponding to material and energy balances around each tray as well as relations depending on the mixture considered and describing to which extent thermodynamic equilibrium (between the liquid and the vapour) is reached on each tray. This system of equations is complemented by equations describing the energy (enthalpy) of liquid and vapour streams as functions of pressure, temperature, and molar fraction of each component of the mixture. While the material balances are bilinear, the mixture-specific relations (i.e., equilibrium and enthalpy equations) and the energy balances on each tray are nonlinear; their complexity is dictated by the thermodynamic complexity of the mixture considered.

3. Distillation line approaches

For distillation line approaches, the column is modelled as a discrete dynamical system. We will restrict our attention to distillation line approaches where one objective function evaluation (i.e., calculation of the distillation column's energy consumption) involves solving the column's associated dynamical system with initial conditions given by the product characteristics at the bottom of the column, hence simulating the column from bottom to top. Note that the following development can be extended naturally for top-to-bottom approaches. This resolution procedure terminates when the dynamical system passes through a point that is within an acceptable distance of the top product (for bottom-to-top approaches) or the bottom product (for top-to-bottom approaches). We describe the distillation column model and the energy minimization problem below.

3.1. Distillation column model

For the columns considered here, it is assumed that the internal pressure is constant with respect to column height (i.e., no pressure drop along the column), that for each tray the liquid and vapour exiting are at thermodynamic equilibrium and that the constant molar overflow (CMO) hypothesis holds [6]. Note that the CMO hypothesis is equivalent to assuming that liquid and vapour flows are constant with respect to column height for the sections above the column feed (termed rectifying section) and below the column feed (termed stripping section). Most optimization-based methods make these assumptions at the conceptual design level and take into account deviations from these hypotheses afterwards [7]. In the present study, we limit ourselves to the conceptual design of minimal energy single-feed adiabatic distillation columns (with no side-streams).

The ratios of output and recirculated flow rates for the evaporator (also termed reboiler) and the condenser are directly linked to their respective energy consumptions. These ratios are termed reboil and reflux ratio (respectively) and are denoted by s and r (respectively). The distillation

stages are numbered from bottom to top ($j = 1, \ldots, N$), and x_j (respectively y_j) denotes the (c-1)-tuple of molar fractions for the liquid (respectively vapour) exiting stage j, where c is the number of mixture components and N is the total number of stages. The number of stages in the rectifying section (stages from feed to distillate) and the number of stages in the stripping section (stages from bottom to feed, excluding the feed stage) are respectively denoted by N_{rec} and N_{str}. Hence the feed tray is labelled by $N_{str}+1$. The bottoms, distillate, and feed flow rates are denoted respectively by B, D, and F. The bottoms, distillate, and feed composition (c-1)-tuples are denoted respectively by x_B, x_D, and x_F. Finally, the feed's thermal condition (i.e., liquid proportion of the feed) is denoted by q. Note that there are two types of condensers, total or partial, corresponding to whether the output of the condenser is a saturated liquid or saturated vapour (respectively). Accordingly, the distillate composition is written x_D (respectively y_D) for a total (respectively partial) condenser.

Since it is assumed that distillation columns verify the CMO assumption and that equilibrium stages are considered, the following model equations describe the liquid and vapour compositions along the column for a given internal pressure (see [8]).

$$x_{j+1} = \frac{s}{s+1} y_j + \frac{x_B}{s+1}, \quad j = 1, \ldots, N_{str} \quad \text{(stripping section)} \tag{1}$$

$$x_{j+1} = \frac{r+1}{r} y_j - \frac{x_D}{r}, \quad j = N_{str}+1, \ldots, N-1 \quad \text{(rectifying section)} \tag{2}$$

$$r = (s-q+1)\frac{x_{F,i} - x_{D,i}}{x_{B,i} - x_{F,i}} - q \quad \text{for a selected } i \in \{1, \ldots, c\}, \tag{3}$$

$$y_j = f_{eq}(x_j), \quad j = 1, \ldots, N \tag{4}$$

$$x_1 = x_B \tag{5}$$

The symbol f_{eq} denotes the equilibrium function (defined by the given pressure) linking liquid and vapour compositions at equilibrium. Global mass and energy balances around the column yield (3), which is valid for any mixture component (i). In the context of distillation line methods, however, one chooses a particular i, e.g., $i = 1$, which corresponds to the light component when components are indexed in decreasing order of volatility [2].

For a bottom-to-top approach, we assume that the column internal pressure, as well as q, x_B, x_D, and x_F are known. For a given value of N_{str}, if we fix s, we can calculate the stripping profile, i.e., compositions x_j and y_j for $j = 1, \ldots, N_{str}+1$ from the bottom product up to the feed tray ($j = N_{str}+1$), by using equations (1), (4), and (5). The reflux ratio r can then be calculated with equation (3). Using Equations (2) and (4), tray compositions in the rectifying section can be calculated iteratively starting at tray $N_{str}+2$ and ending at a predefined top tray ($j = N$).

To summarize things geometrically, recall that for given column operating conditions, the liquid and vapour streams inside the column can each be described by a tuple of real numbers, part of which corresponds to the molar fraction (or proportion) of each component of the mixture present in the stream. Since these tuples of molar fractions for liquid and vapour streams are linked via equilibrium relations and material balances, one can consider only one of these streams along a column (i.e., either liquid or vapour, but not both) in order to gain insight into a column's performance. The general consensus is to focus on the liquid stream.

For the liquid stream exiting a tray in a column, the tuple of molar fractions corresponds to a point in $\mathcal{R}^c$, where c is the number of components in the mixture. Since molar fractions sum to one, this point is located in the standard (c-1)-simplex contained in $\mathcal{R}^c$ (i.e., in the convex hull of the canonical basis for $\mathcal{R}^c$); the projection of this simplex onto its first (c-1) coordinates is termed

the "composition simplex". The set of tuples corresponding to molar fractions in the liquid exiting each tray of the column then traces a (discrete) path in the composition simplex (see Figure 2).

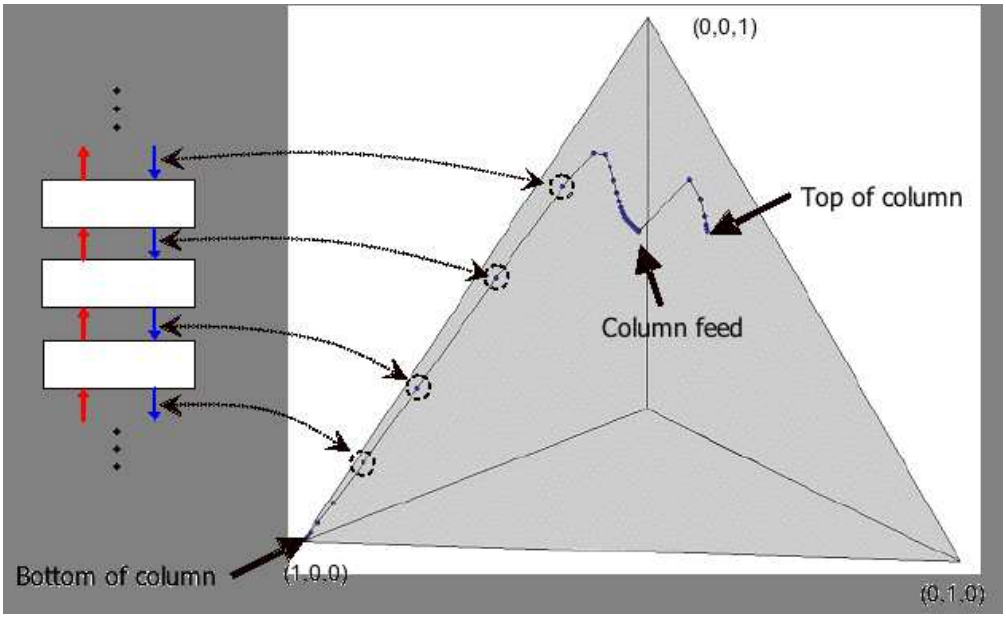


Figure 2. Liquid composition trajectory in the composition simplex for a 4-component mixture ($c = 4$) corresponding to given values of s, N_{str}, N, q, x_B, x_D, and x_F.

The bottom-to-top procedure ends when either the liquid composition profile goes out of the composition simplex, when it converges, or when a composition (liquid or vapour) reaches a neighborhood of the specified distillate. The number of trays N and the calculated distillate composition, which is x_N in the case of a total condenser and y_N in the case of a partial condenser, are thus determined by the value of the reboil ratio (s).

3.2. Optimization problem

Recall that for a given column pressure and a given separation task, i.e., the thermal state of the feed (q), its composition (x_F), the composition of the bottoms (x_B), and a target value for the distillate composition (x_D or y_D, depending on the type of condenser), our goal is to find the minimal energy requirement for distillation. This can be achieved by minimizing various objective functions with respect to the model equations; for example, [9] and [1] propose objective functions based on geometric and thermodynamic heuristics, respectively. In the present study, we will use the reboil ratio (s) as the objective function. It is common practice to do so since the column's energy consumption is an increasing function of the reboil ratio [3]. Therefore, we consider the following MINLP optimization problem. For a given pressure and given values of q, x_B, x_F, and x_D or y_D (depending on the type of condenser), we wish to solve

$$\begin{aligned} \min \quad & s \\ \text{s.t.} \quad & \text{Equations (1) to (5)}, \\ & \|x_N - x_D\|^2 \leq \zeta \text{ or } \|y_N - y_D\|^2 \leq \zeta \text{ (depending on the type of condenser)}, \end{aligned}$$

where ζ is a fixed positive real number and $\|\cdot\|$ is the Euclidean norm. Given values of N_{str}, q, x_B, x_F, and x_D, Figure (3) shows the behaviour of the calculated distillate composition (x_N) as a function of the reboil ratio (s). Assuming a distillate composition of $(0.99, 10^{-10})$, we see that fixing a value of ζ determines a set of values of s for which the constraint $\|x_N - x_D\|^2 \leq \zeta$ is verified. The next section investigates a simple fixed point method in order to satisfy the aforementioned constraint. Mathematical properties of this approach are also presented.

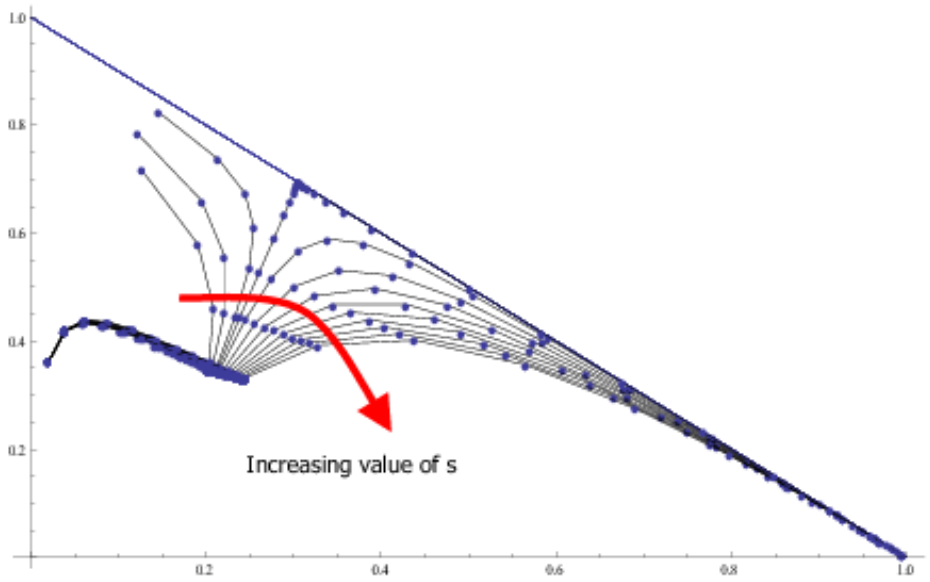


Figure 3. Liquid composition trajectory in the composition simplex for a 3-component mixture ($c = 3$) corresponding to given values of N_{str}, q, x_B, x_D, x_F, and increasing values of s.

4. A fixed point method to satisfy product composition constraints

4.1. Algorithm description

This section presents a specific version of the fixed point method introduced in [9] and known as the iteratively refined approach. For simplicity in writing the procedure, we will assume a partial condenser throughout this section. The following parameters are assumed to be given: q, x_B, x_F, N_{str}, and s. The goal is to find a value of the distillate composition y_D such that equations (1) to (5) are satisfied up to a predefined precision and for which $\|y_N - y_D\|^2 \leq \zeta$ holds. The successive refinement method [5] tries to solve this problem by direct substitution of the distillate composition and by starting from an initial guess.

Set q, x_B, x_F, s, N_{str}, ζ_1, ζ_2, ϵ, where q belongs to $\mathcal{R}$, x_B and x_F belong to the composition simplex $\Delta := \left\{x \in \mathcal{R}^{c-1} \mid \sum_{i=1}^{c-1} x_i \leq 1 \text{ and } x_i \geq 0,\ i = 1, \ldots, c-1\right\}$, and ζ_1, ζ_2, and ϵ are predefined tolerances, with $\zeta_1 \geq \zeta_2$.

Step 0 Preliminary step: calculation of the stripping composition profile. For the given bottoms specification (5), calculate the liquid and vapour compositions in the stripping section, i.e., calculate x_j and y_j, for $j = 1, \ldots, N_{str} + 1$, by alternating between the equilibrium relation (4)

and the stripping operating line equation (1). We assume here that for the given specifications, the stripping composition profile is contained in the composition simplex.

Step 1 Initialisation: input an initial guess of the distillate composition y_D^0 and set $k = 0$. Since mass balances link input and output compositions of the column, the distillate composition should be on the line segment $\{x \in \Delta \,|\, x = (1-t)x_B + tx_F,\ t \geq 1\}$ given by a global mass balance along the column [3].

Step 2 Calculation of the rectifying profile and distillate composition correction. Note that we assume that the column has at least one rectifying stage. It is also implicitly assumed that the image of the composition simplex by the equilibrium function (f_{eq}) is contained within the composition simplex.

Step 3 Calculate $r^{(k)} = r^{(k)}\left(y_D^{(k)}\right)$ using (3) for $i = 1$ (light component). Set $j = N_{str} + 1$.

Step 4 Calculate $x_{j+1}^{(k)}$ with (2).

Step 5 If $x_{j+1}^{(k)}$ lies outside the composition simplex or if the liquid composition profile has converged (i.e., $\|x_{j+1}^{(k)} - x_j^{(k)}\| \leq \epsilon$ holds), then set $N_{rec}^{(k)} := j - N_{str}$, $y_D^{(k+1)} := y_j^{(k)}$ and go to Step 8. Else go to Step 6.

Step 6 Calculate $y_{j+1}^{(k)}$ with (4) and set $y_D^{(k+1)} := y_j^{(k)}$.

Step 7 If $\|y_D^{(k+1)} - y_D^{(k)}\| \leq \zeta_1$ holds, then set $N_{rec}^{(k)} := j + 1 - N_{str}$ and go to Step 8; else set $j := j + 1$ and go to Step 4.

Step 8 If $\|y_D^{(k+1)} - y_D^{(k)}\| \leq \zeta_2$ holds, then go to Step 9; otherwise set $k := k+1$ and return to Step 3.

Step 9 End of procedure: set $y_D = y_D^{(k)}$ and $N_{rec} = N_{rec}^{(k)}$.

4.2. Mathematical Investigations

Having precisely defined the context in the previous subsection, we aim to understand better the convergence properties of the fixed point approach. As before, we assume a partial condenser and given values for q, x_B, x_F, s, N_{str}, ζ_1, and ζ_2. The building block of the iteratively refined method is the function of the composition simplex into itself that defines the relation between input and output vapour compositions on a tray in the rectifying section (i.e., above the column feed). More precisely, consider the function $G : \Lambda \subseteq \Delta \times \Delta \to \Delta$ defined by $G(y, y_D) = f_{eq}(f_{rec}(y, y_D))$, where, for (y, y_D) in $\Delta \times \Delta$,

$$f_{rec}(y, y_D) = \frac{(r(y_D) + 1)y - y_D}{r(y_D)},$$

and $r(y_D)$ is computed as in (3) (where we use the light component of the mixture ($i = 1$)). Note that G is not necessarily defined on the whole set $\Delta \times \Delta$ since the image of the function f_{rec} may contain points in $\mathcal{R}^{c-1}\backslash\Delta$, on which f_{eq} is not necessarily defined. Accordingly, the domain of the function G, denoted by Λ, is defined as

$$\Lambda = f_{rec}^{-1}(\Delta) = \{(y, y_D) \in \Delta \times \Delta \,|\, f_{rec}(y, y_D) \in \Delta\}.$$

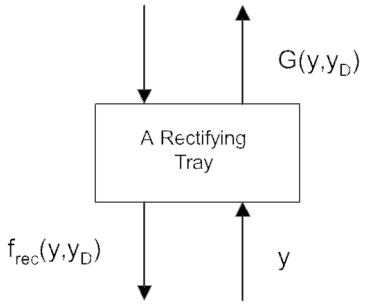


Figure 4. Schematic representation of the function G.

Figure 4 gives a schematic representation of the function G.

Since x_B, s, and N_{str} are known, liquid and vapour compositions for trays $j = 1$ to $j = N_{str} + 1$ are determined by using equations (1) and (4). Studying the sensitivity of the vapour composition k trays above the feed tray as a function of the distillate composition y_D amounts to analyzing the functions $H_k : \Lambda_k \subseteq \Delta \to \Delta$, defined recursively by

$$H_1(y_D) = G(y_{N_{str}+1}, y_D),$$

and

$$H_k(y_D) = G(H_{k-1}(y_D), y_D),$$

for $k \geq 2$. The domains of the functions H_k are defined recursively by

$$\Lambda_1 = \{y_D \in \Delta \,|\, (y_{N_{str}+1}, y_D) \in \Lambda\},$$

and

$$\Lambda_k = \{y_D \in \Lambda_{k-1} \,|\, (H_{k-1}(y_D), y_D) \in \Lambda\},$$

for $k \geq 2$. Sparing the reader the technical details, one obtains the following sensitivity result.

PROPOSITION 1. *Assume that q, x_B, x_F, s, N_{str}, ζ_1, and ζ_2 are fixed as in Section 4.1 and that the functions G and H_k are defined as above. The Jacobian of the iterative map H_k is recursively given by:*

$$\begin{aligned} J(H_1; y_D) &= J(f_{eq}; x_{N_{str}+2}) M(y_{N_{str}+1}; y_D), \\ J(H_k; y_D) &= J(f_{eq}; x_{N_{str}+k+1}) \left[\frac{r(y_D)+1}{r(y_D)} J(H_{k-1}; y_D) + M(y_{N_{str}+k}; y_D) \right], \text{ for } k \geq 2, \end{aligned}$$

where $J(f; z)$ denotes the Jacobian of the function f at point z and $M(y; y_D)$ is the square matrix of order $(c-1)$ whose (α, β)-entry is

$$M(y; y_D)_{\alpha,\beta} = \begin{cases} 0 & \text{if } \beta \neq 1 \text{ and } \alpha \neq \beta, \\ \frac{1}{r(y_D)^2} \frac{s-q+1}{x_{F,1}-x_{B,1}} (y_{D,\alpha} - y_\alpha) & \text{if } \beta = 1 \text{ and } \alpha \neq \beta, \\ -\frac{1}{r(y_D)} & \text{if } \alpha = \beta \neq 1, \\ \frac{1}{r(y_D)^2} \frac{s-q+1}{x_{F,1}-x_{B,1}} (y_{D,\alpha} - y_\alpha) - \frac{1}{r(y_D)} & \text{if } \alpha = \beta = 1. \end{cases}$$

The proof of the preceding proposition follows from the definitions of the functions G, H_k, f_{rec}, and r as well as the chain rule. Suggested future work includes deriving information on the eigenvalues of the Jacobian of the iterative map (e.g., upper and lower bounds, spectral radius), by studying the equilibrium function, which encapsulates the thermodynamic model of the mixture. Such sensitivity results may help in understanding the convergence properties of the iteratively refined approach and may lead to improvements of this method.

5. Conclusions and perspectives

In this text, we presented the context in which distillation line methods attempt to determine minimum energy requirements of single-feed adiabatic distillation (with no side-streams). Since the quality of the results provided by such methods often depends on global mass balance errors, sensitivity information on these errors may provide a way to control the quality of the results obtained. In this work, we considered the bottom-to-top iteratively refined method of Hassan and Lucia [5]. Casting this method in a more general context, we derived a recursive formula for the Jacobian of the iterative map used to calculate rectifying vapour composition profiles.

6. Acknowledgement

The left image in Figure 1 is due to H. Padleckas and is used under the Creative Commons Attribution ShareAlike 2.5 License (`http://en.wikipedia.org/wiki/File:Tray_Distillation_Tower.PNG`).

References

1. Etienne Ayotte-Sauvé and Mikhail Sorin. Energy requirements of distillation: exergy, pinch points and the reversible column. Industrial Engineering Chemistry Fundamentals, 49(11):5439-5449, 2010.
2. Mariana Barttfeld, Pio A. Aguirre and Ignacio E. Grossmann. Alternative Representations and Formulations for the Economic Optimization of Multicomponent Distillation Columns. Computers and Chemical Engineering, 27(3):363-383, 2003.
3. Michael F. Doherty and Michael F. Malone. Conceptual design of distillation systems. McGraw-Hill, Boston, MA, USA, 2001.
4. R. Bruce Eldridge, A. Frank Seibert, and Sharon Robinson. Hybrid separations/distillation technology: Research opportunities for energy and emissions reduction. Industrial Technologies Program, U.S. Department of Energy, Energy Efficiency and Renewable Energy, 2005.
5. Chris Hassan and Angelo Lucia. An iteratively refined distillation line method. AIChE Journal, to appear.
6. C. Judson King. Separation Processes, Second edition. McGraw-Hill, New York, NY, USA, 1980.
7. Jennifer R. Knight and Michael F. Doherty. Design and synthesis of homogeneous azeotropic distillations. 5. Columns with nonnegligible heat effects. Industrial & Engineering Chemistry Fundamentals, 25(2):279-289, 1986.
8. Juergen Koehler, Peter Poellmann, and Eckhart Blass. A review on minimum energy calculations for ideal and nonideal distillations. Industrial & Engineering Chemistry Research, 34(4):1003-1120, 1995.
9. Angelo Lucia, Amit Amale, and Ross Taylor. Distillation pinch points and more. Computers and Chemical Engineering, 32(6):1350-1372, 2008.

Reacting Flows and Vortices

Anne Bourlioux
Coordinator, Université de Montréal

Pierre Gauthier
Industrial representative, Rolls-Royce Canada

Sanaz Arabzade, Mario Morfin Ramirez and Nasim Shahbazian
Students, University of Toronto

Alexandre Desfossés Foucault and Louis-Xavier Proulx
Students, Université de Montréal

Apala Majumdar, Stephen Peppin and Tim Reis
Mathematical Institute, University of Oxford

Catherine Mavriplis
Professor, University of Ottawa

Mary Pugh
Professor, University of Toronto

1. Statement

Rolls Royce manufactures plane engines that can also be used as turbines at some point in the life cycle of the engines. Thus the company takes an interest in the chemical reactions that occur in a gas turbine combustor. Specifically, Rolls Royce engineers wish to evaluate the quantities of some of the chemicals, such as carbon monoxide (CO) and nitric oxide (NO), that come out of the "tail pipe". Ideally, in order to achieve this, one would perform a direct numerical simulation of the compressible reactive Navier-Stokes equations (including all the chemical reactions). In practice, the chemical reactions happen on a time scale that is much shorter than the time scales of the fluid flow, resulting in a very stiff, large system. Rolls Royce has carried out large eddy simulations of a reactive Navier-Stokes model in which only a few chemical reactions are considered. When one compares these simulations to experiments, one observes that the simulations are in qualitative agreement with them but are not accurate enough (from a quantitative point of view) to be of industrial use.

For this reason, Rolls Royce is considering some sort of hybrid computation. Specifically, the engineers' approach can be described as follows.

1. Use a CFD software package to carry out a large eddy simulation of the reactive Navier-Stokes model with simplified chemistry.

2. Given the simulation results, carry out temporal averaging over a time window in order to create spatially-dependent but time-independent values for the velocity (u, v, w), temperature T, pressure p, carbon monoxide CO, nitric oxide NO, carbon dioxide CO2, water H20, and oxygen O2.

3. Given these spatially-dependent quantities, do some sort of post-processing to identify spatial regions that might be called "vortices" (regions of space in which a fluid particle would spend a good chunk of time before leaving the region).

4. For each of these regions, a "residence time" and the spatial averages of CO, NO, CO2, H20, O2, temperature, and pressure are computed.

5. These spatial averages, along with the geometry of the decomposition (describing the adjacency relationship between regions) are then passed on to a chemistry software package, Chemkin, which computes the chemical reactions in detail.

In this way, a stiff PDE would be handled in three stages: the resolution of a less stiff PDE via a CFD software (in Step 1), the post-processing of the CFD solution (in Steps 2-4), and the computation of a detailed chemical network model via Chemkin (in Step 5).

The workshop team has been asked to help with Step 3: the decomposition of the fluid domain into subdomains that, in some way, correspond to largely independent reaction chambers. Figure 1 illustrates such a decomposition.

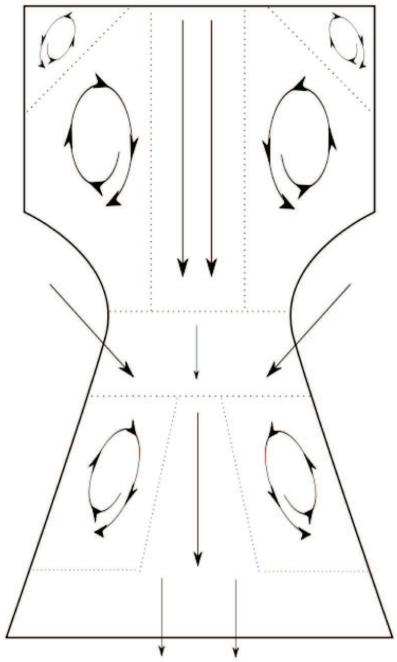


Figure 1. Combustion chamber containing some vortices

For a given steady CFD solution computed for a gas turbine combustor, information such as the velocity field (u, v, w), the temperature T, and the pressure P are known everywhere in the domain. From these data, regions in the flow where large and small vortical structures are formed or "recirculating zones" must be determined.

There has already been extensive research on identifying vortices in fluid flows; a variety of approaches have been proposed. The workshop team considered a number of identification methods

that are in the literature. In a first step, these methods were tested on sample two-dimensional flows.

2. Approaches

A quick literature survey reveals a variety of existing methods for various applications. The team's focus was on understanding and testing some of those (classical or not so classical) approaches. Here is a list of the various methods that were tested during the week.

1. Greek methods: Q-criterion, Δ-criterion, λ_2-criterion, and Eigen Helicity.
2. Critical points methods.
3. Symbolic dynamic method.
4. Geometric methods.
5. Haller's method (from a reference suggested by the industrial partner).

In this section, the basic ideas behind those methods will be explained. The results obtained by these methods will then be presented. Note that the Eigen Helicity method belongs to the family of so-called Greek methods but is described in a separate subsection.

2.1. Greek methods: general framework

The Greek methods are all based on the analysis of the deformation matrix of a vector field. Given the vector field $\vec{u} := (u(x,y), v(x,y))$, the deformation matrix is the following:

$$\nabla \vec{u} := \nabla \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} \tag{1}$$

If we define the symmetric component S and antisymmetric component Ω as

$$S = \frac{1}{2}\left[\nabla \vec{u} + (\nabla \vec{u})^T\right] \tag{2}$$

and

$$\Omega = \frac{1}{2}\left[\nabla \vec{u} - (\nabla \vec{u})^T\right] \tag{3}$$

respectively, $\nabla \vec{u}$ can written as follows.

$$\nabla \vec{u} = S + \Omega = \frac{1}{2}\left(\nabla \vec{u} + (\nabla \vec{u})^T\right) + \frac{1}{2}\left(\nabla \vec{u} - (\nabla \vec{u})^T\right) \tag{4}$$

More specifically, S and Ω are the following matrices.

$$S = \begin{pmatrix} u_x & \frac{1}{2}(u_y + v_x) \\ \frac{1}{2}(u_y + v_x) & v_y \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0 & \frac{1}{2}(u_y - v_x) \\ -\frac{1}{2}(u_y - v_x) & 0 \end{pmatrix}.$$

2.1.1. *The Q-criterion*
This method segments the domain by applying the following rule: the points at which Ω is "larger" than S are within a vortex and the points at which Ω is "smaller" than S are outside a vortex. How do we measure the size of a matrix? Hunt *et al.* [6] use the norm

$$\|\Omega\|^2 := \mathrm{Tr}\left(\Omega\ \Omega^T\right) = \Omega_{11}^2 + \Omega_{12}^2 + \Omega_{21}^2 + \Omega_{22}^2.$$

In this fashion $\nabla\vec{u}$ is viewed as a vector in R^4 and the dot product has the usual meaning. In this context S and Ω are orthogonal.

$$\nabla\vec{u} = \begin{pmatrix} u_x \\ u_y \\ v_x \\ v_y \end{pmatrix}\ ,\quad S \in \mathrm{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}\ ,\quad \Omega \in \mathrm{span}\left\{ \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \right\}$$

The Q-criterion is defined as follows:

$$\|S(x,y)\|^2 < \|\Omega(x,y)\|^2 \quad \Longrightarrow \quad (x,y) \text{ is in a vortex,}$$

that is,

$$u_x^2 + 2u_y v_x + v_y^2 < 0 \quad \Longrightarrow \quad (x,y) \text{ is in a vortex.}$$

The above can be written in terms of the matrix invariants, $\mathrm{Tr}(\nabla\vec{u})$ and $\mathrm{Det}(\nabla\vec{u})$:

$$Q = (\mathrm{Tr}(\nabla\vec{u}))^2 - 2\mathrm{Det}(\nabla\vec{u}) < 0 \quad \Longrightarrow \quad (x,y) \text{ is in a vortex.} \tag{5}$$

Observe that if the flow is incompressible, we have $\mathrm{Tr}(\nabla\vec{u}) = 0$. Therefore, if the flow is incompressible **and** $\mathrm{Det}(\nabla\vec{u})$ is greater than 0, we may conclude that (x,y) is in a vortex.

2.1.2. Δ*-criterion*
The Δ-criterion identifies a vortex as a region where the eigenvalues of $\nabla\vec{u}$ are complex. For a 2-dimensional flow, the eigenvalues would be complex conjugates of one another and for a 3-dimensional flow, there would be one real eigenvalue and two complex conjugates. In the case of a 2-dimensional flow, the eigenvalues of $\nabla\vec{u}$ are

$$\frac{1}{2}\mathrm{Tr}(\nabla\vec{u}) \pm \frac{1}{2}\sqrt{\mathrm{Tr}(\nabla\vec{u})^2 - 2\mathrm{Det}(\nabla\vec{u})}.$$

If we compare the projection approach (now called the Q-criterion) to the eigenvalue approach (based on the Δ-criterion), we see that the following implication still holds.

$$\Delta = (\mathrm{Tr}(\nabla\vec{u}))^2 - 2\mathrm{Det}(\nabla\vec{u}) < 0 \quad \Longrightarrow \quad (x,y) \text{ is in a vortex} \tag{6}$$

As before, if the flow is incompressible **and** $\mathrm{Det}(\nabla\vec{u})$ is greater than 0, we may conclude that (x,y) is in a vortex.

2.1.3. λ_2*-criterion*
Jeong and Hussain [7] proposed seeking minima of the pressure as a method for identifying vortices. Expressing the gradient of the Navier-Stokes equations in terms of the matrices S and Ω yields

$$\frac{DS_{ij}}{Dt} - \nu S_{ij,kk} + \Omega_{ik}\,\Omega_{kj} + S_{ik}\,S_{kj} = -\frac{1}{\rho}p_{ij}, \tag{7}$$

where $\frac{D}{Dt}$ denotes the Lagrangian derivative. In this way, the Hessian of the pressure is given in terms of a material derivative of S, the Laplacian of S, and the matrix $\Omega^2 + S^2$. A local pressure minimum requires that the Hessian of p have two positive eigenvalues.

If the flow is steady and the viscosity is zero, then the Hessian of the pressure is simply $\Omega^2 + S^2$ and we have the following implication.

$$\Omega^2 + S^2 \text{ has two negative eigenvalues} \quad \Longrightarrow \quad (x, y) \text{ is in a vortex} \tag{8}$$

We note that the λ_2-criterion (8) has also been applied in the case of unsteady flows and viscous flows. If the flow is incompressible, the eigenvalues of $\Omega^2 + S^2$ are both equal to $-\text{Det}(\nabla\vec{u})$. We may conclude again that if the flow is incompressible and $\text{Det}(\nabla\vec{u}) > 0$ holds, (x, y) is in a vortex. Therefore the Q, Δ, and λ_2 methods are identical in the case of 2-dimensional incompressible flows but could yield different results in the case of compressible flows.

2.2. Eigen Helicity method

2.2.1. *Introduction*

In commonly used vortex identification methods, attention is usually focused on the vorticity tensor $\Omega = \frac{1}{2}\left((\nabla\vec{u} - (\nabla\vec{u})^T\right)$ and the strain rate tensor $S = \frac{1}{2}\left((\nabla\vec{u})^T + \nabla\vec{u}\right)$, where $\vec{u}$ is the velocity vector. Eigenvectors of the strain rate tensor S, however, do not necessarily align with the vortex tubes. As pointed out by Haller [3], this can lead to spurious predictions of the location of vortex boundaries. Zhang *et al.* [17] note that "vorticity production also contributes to strain generation, so an alignment scheme based on the eigenvectors of S is not appropriate." Therefore they propose a new scheme that presents a different alignment of the vorticity vector with the full eigen system of the velocity gradient tensor. They refer to this scheme as the "Eigen Helicity Density" or H_e scheme.

In order to introduce the Eigen Helicity density scheme, we first review the Helicity method, introduced by Levy *et al.* [11]. They use the normalized helicity H_n to extract vortex core lines. H_n is simply defined to be the cosine of the angle between the velocity $\vec{v}$ and the vorticity $\vec{\omega} = \nabla \times \vec{u}$, that is,

$$H_n = \frac{\vec{v} \cdot \vec{\omega}}{|\vec{v}||\vec{\omega}|}. \tag{9}$$

The underlying assumption is that near vortex core regions, the angle between $\vec{v}$ and $\vec{\omega}$ is very small, and in the limiting case where the two vectors are parallel or antiparallel, H_n is ± 1. The authors extract vortex core lines by locating the maximum points of H_n on cross-sectional planes. The sign of H_n indicates the direction of swirl, clockwise or anticlockwise, and the sign changes whenever there is a transition between primary and secondary vortices. One criticism of the Helicity method is that it depends on the velocity, ensuring that the method is not Galilean invariant.

The Eigen Helicity method is a refinement of the Helicity method that makes use of the full eigensystem of the velocity gradient tensor. The quantity H_e is defined as follows.

$$H_e = \frac{\vec{n}_{swirl} \cdot \vec{\omega}}{|\vec{n}_{swirl}||\vec{\omega}|}, \tag{10}$$

where $\vec{n}_{swirl} = -(\vec{e}_1 \times \vec{e}_2)i/2$, and $\vec{e}_1$ and $\vec{e}_2$ are the two eigenvectors corresponding to the complex conjugate eigenvalues of $\nabla\vec{u}$. By analogy with the Helicity method, the vortex core corresponds to regions where the magnitude of H_e is maximum.

We have tested six different methods: the vorticity maximum method, the Q-method, the Δ-method, the λ_2-method, the Helicity method, and the Eigen Helicity method on 3-dimensional Richtmyer-Meshkov (RM) simulations, where vortices and their interactions dominate the flow [17]. This flow is compressible, and has variable density and a baroclinic effect. In the next section we report spurious results with the first five methods: some methods like the pressure minimum method fail to find the vortex tubes, whereas others like the Q-method detect parts of the vortex and also identify non-vortex regions as being vortices.

The Eigen Helicity method has been found to be the most accurate within the scope of our study. The desired vortex features are obtained by plotting the contours of H_e. The peaks of H_e correspond to the vortex centres, and as in the Helicity method, they also give the direction of rotation within the vortex. Furthermore, the Eigen Helicity method is not based on any assumption on the flow field, and has the potential to converge (i.e., may be a method for automatic vortex identification). It is proposed to use the contours of threshold values such as $H_e = 0.9$ (or $Q = 0$ in the Q-method, $\Delta = 0$ in the Δ-method, $\lambda_2 = 0$ in the λ_2-method, where λ_2 is the second largest eigenvalue of $\nabla\vec{u}$) to provide good approximations to the boundaries of the vortex regions.

2.2.2. *Lattice Boltzmann*

In the next section we will present tests carried out on different kinds of data, including a flow around a cylinder that is a Lattice-Boltzmann solution of the 2-dimensional Navier-Stokes equation. For this reason we present a brief discussion of this concept. The Lattice Boltzmann equation can be classified as an explicit, Lagrangian, finite-hyperbolic approximation to the Navier-Stokes equation that has been derived within the framework of kinetic theory [4]. The LBE is characterized by a lattice and some rule describing the manner in which particles move along lattice directions from one node to another.

Let $f_i(\vec{x}, t) = f(\vec{x}, \vec{c_i}, t)$ be the single particle distribution function at node $\vec{x}$ and time t with the discrete microscopic velocity $\vec{c_i}$. The Lattice Boltzmann evolution equation in rectilinear coordinates is

$$f_i(\vec{x} + \vec{c_i}, t + 1) = f_i(\vec{x}, t) - \omega \left(f_i - f_i^{(e)} \right), \tag{11}$$

where $f_i^{(e)}$ is a suitably defined equilibrium function and ω is the relaxation parameter of the fluid. Mass and momentum are defined to be the first two moments of the distribution function respectively, and conservation of these quantities require that the following relations hold.

$$\rho = \sum_i f_i = \sum_i f_i^{(e)}, \tag{12}$$

$$\rho\vec{u} = \sum_i f_i \vec{c_i} = \sum_i f_i^{(e)} \vec{c_i}. \tag{13}$$

By specifying a sufficiently symmetric lattice (i.e., the so-called D2Q9 lattice) and taking into account the physical conservation laws, we find the equilibrium function to be

$$f_i^{(e)} = \rho w_i \left[1 + 3\vec{c_i} \cdot \vec{u} - \frac{3}{2}\|\vec{u}\|^2 + \frac{9}{2} \left(\vec{c_i} \cdot \vec{u} \right)^2 \right], \tag{14}$$

where $w_0 = 4/9$, $w_i = 1/9$ $(i = 1, \ldots, 4)$, and $w_i = 1/36$ $(i = 5, \ldots, 8)$.

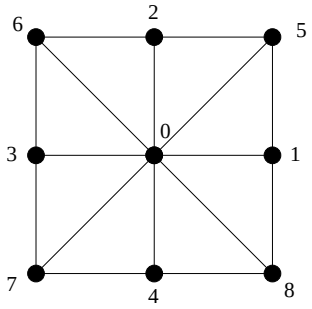


Figure 2. The D2Q9 lattice

In the hydrodynamic and incompressible (small Mach number) limit, Equation (11) can be shown to be a second order approximation to

$$\partial_t \vec{u} + \vec{u} \cdot \nabla \vec{u} \;=\; -\frac{1}{\rho}\nabla p_0 + \nu \nabla^2 \vec{u}, \tag{15}$$

$$\nabla \cdot \vec{u} \;=\; 0, \tag{16}$$

where

$$\nu = \frac{1}{3}\left(\frac{1}{\omega} - \frac{1}{2}\right) \tag{17}$$

is the kinematic viscosity (see Higuera *et al.* [4]).

2.2.3. *The equivalence of the Greek methods in 2D and their relation to the Eigen Helicity method* The Q-criterion for a vortex is given by

$$Q = \frac{1}{2}\left(u_{i,i}^2 - u_{i,j}u_{j,i}\right) > 0, \tag{18}$$

which for incompressible, 2-dimensional flows reduces to

$$Q = -\frac{1}{2}u_{i,j}u_{j,i} = -u_x^2 - u_y v_x, \tag{19}$$

since $u_x + v_y = 0$. We also note that for a 2-dimensional incompressible flow, the invariant $R = \mathrm{Det}\,(\nabla \vec{u})$ is given by

$$R = u_x v_y - u_y v_x = -u_x^2 - u_y v_x, \tag{20}$$

so that $Q = R$ in this case. This in turn implies that the Δ-criterion

$$\Delta = \left(\frac{Q}{3}\right)^3 + \left(\frac{R}{2}\right)^2 > 0 \tag{21}$$

can be expressed in terms of Q alone and $Q > 0$ necessarily implies that $\Delta > 0$. Similarly, the λ_2-method is based on the eigenvalues of the matrix $S^2 + \Omega^2$; this matrix has three ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$, and the vortex identification criterion is equivalent to the relation $\lambda_2 \leq 0$. A direct computation shows that for a 2-dimensional incompressible flow, the characteristic equation for the eigenvalues is simply

$$(Q + \lambda)^2 = 0, \tag{22}$$

so that if $Q > 0$ holds, then we necessarily have two negative eigenvalues. We deduce that it suffices to consider the Q-method for this class of vortex identification criteria.

For the Helicity method (see Levy *et al.* [11]), we note that for a 2-dimensional velocity field

$$\vec{u} = (u(x,y), v(x,y), 0)\,,$$

the vorticity is in the z-direction, so that the quantity H_n defined in (9) is necessarily zero and this method does not give us any information. For the Eigen Helicity method (see Zhang *et al.* [17]), we note that the quantity H_e in (10) is only defined whenever $\nabla\vec{u}$ has a pair of complex conjugate eigenvalues. We denote the corresponding eigenvectors by $\vec{e}_1 = (e_{1,1}, e_{1,2})$ and $\vec{e}_2 = (e_{2,1}, e_{2,2})$ respectively. A direct calculation shows that

$$H_e = -\frac{i}{2}\left[e_{1,1}e_{2,2} - e_{2,1}e_{1,2}\right](v_x - u_y)\,, \tag{23}$$

so that

$$|H_e|^2 = \frac{1}{4}\left(e_{1,1}e_{2,2} - e_{2,1}e_{1,2}\right)\overline{\left(e_{1,1}e_{2,2} - e_{2,1}e_{1,2}\right)}\left(v_x - u_y\right)^2,$$

where $\overline{(e_{1,1}e_{2,2} - e_{2,1}e_{1,2})}$ denotes the complex conjugate. The vortices correspond to the maxima of the function $|H_e|^2$.

2.3. Critical points methods

This approach is also based on the deformation matrix of a given flow. It considers critical points in the flow and tries to determine whether they are saddle points or vortex centres. The Jacobian of a critical point (x, y) is actually the deformation matrix at that point (see Equation (1)). The classification of points is based on the trace $\mathrm{Tr}(J)$ and determinant $\det(J)$ of this matrix. If $\det(J) < 0$ holds, the point is a saddle point, and if $\det(J) > \frac{(-\mathrm{Tr}(J))^2}{2}$, it is a vortex centre.

2.4. Symbolic dynamic method

This approach uses graph theory to try to follow particles around the combustion chamber as they move under the influence of a given velocity field, denoted by $(u(x,y), v(x,y))$. It assumes that if a particle stays inside a region for a long enough time, it is inside a vortex. The first step is to partition our domain. Then, we must determine if the particles in a region move to another region in a given time step. We construct the connectivity matrix in the following way.

$$M(i,j) = \begin{cases} 1 & \text{if some of the gas in partition member } i \text{ went to} \\ & \text{partition member } j \text{ in one time step} \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

The entries equal to 1 in the jth column indicate which members of the partition send material to the j^{th} member of the partition. The matrix M is the adjacency matrix of a directed graph. We can then calculate the powers of M. The entry $M^p(i,j)$ equals the number of paths of length p from member i to member j. Columns with larger values of $M^p(i,j)$ indicate members receiving more mass; cycles of mass will be detected as recurrences.

For example consider the following matrix.

$$M = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} \tag{25}$$

Then M^2 and M^3 can be computed easily.

$$M^2 = \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 \end{pmatrix} \tag{26}$$

$$M^3 = \begin{pmatrix} 0 & 1 & 4 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 4 & 0 \end{pmatrix} \tag{27}$$

The entries $M(4,1)$ and $M(1,4)$ form a cycle in the underlying directed graph. This indicates the presence of a vortex. In general, after computing the powers of M up to the pth power, we must localize the members of the partition that are receiving the largest amount of mass by adding up the elements in every column of the matrix for every power. This will give us a function F defined for every member of the domain partition. $F(j)$ gives us the number of ways in which one can send mass to j from every member in p iterations or fewer than p iterations. The members maximizing F are the points receiving the largest amount of mass. The next step is to decide, for every other point in the domain, to which local maximum of F this point sends the largest amount of mass. The following procedure can be used.

1. Take the domain and color the local maxima of F with a dark shade.
2. Using the first power of the matrix, look at the points around each local maximum that gave mass to this point and color them with a lighter shade.
3. Repeat the previous operation for the other powers of the matrix (using a lighter shade at each iteration).
4. Repeat this process for the points that already have a colour.
5. If we attempt to assign to a point a colour that is darker than the one it already has, this means that the point in question gives more mass to another local maximum. So we have found part of the border between two vortices.

This method, alhthough conceptually interesting, entails expensive computations. The original matrix M is very large, and calculating the powers of M consumes a lot of time.

2.5. Geometric methods

The main idea behind geometric methods is that geometric properties of the streamlines are used (instead of physical quantities related to the flow). For example, Sadarjoen and Post [13] have developped the curvature centre method.

2.5.1. *Curvature centre method*
The idea underlying this method is simple: it is based on the computation of the curvature centre for different points on the streamlines. The curvature centre points are accumulated on a new grid, and vortex regions can be identified on this grid by making use of the fact that curvature centre points accumulate in the center of vortices. In non-vortical regions, the curvature centre points are scattered. This method would detect circular vortices easily, but it has some limitations since centre points might be spread out (for instance in the case of elliptic vortices).

2.5.2. *Streamline method*
Petz *et al.* [12] have developped a new algorithm that detects and extracts 2-dimensional vortex regions. The extracted regions correspond to what we could call swirling motion regions. This type of approach has many advantages, listed below.

- The algoritm is parameter-free.
- Arbitrarily shaped vortex regions can be extracted.
- The approach can handle vector fields with and without divergence.
- The approach is grid-independent.

The only required inputs are streamlines and critical points, especially saddle points. A vortex region is defined in the case of a divergence-free vector field as the union of closed streamlines winding around a common center. For vector fields with divergence, the streamlines are not closed, but the same idea can be generalized. It uses lines whose tangents have a constant incident angle with respect to the vector field. Those lines correspond to streamlines in the divergence-free vector field.

The method applies a rotation matrix to the vector field. It then uses a loop, i.e., a positively oriented closed streamline in the rotated vector field (closed streamlines detection is discussed in Theisel's article [15]). Such a loop intersects the original vector field at a constant angle since rotating the vectors around that angle yields a vector field tangential to the loop.

Applying varied rotations to the vector field yields a set of loops. This set can be split into equivalence classes of loops (denoted by r_i) that are homotopic with respect to $\mathbb{R}^2 \setminus \{x \in \mathbb{R}^2 \,|\, V(x) = 0\}$. A vortex region candidate R_i is defined as the area covered by the lines of the equivalence class r_i. The union of lines that wind around a common center defines a vortex region candidate. A vortical behaviour of the flow in such a region is not enforced by the definition. For a restriction to vortex regions, an additional filtering step must be applied. Filter criterias can be based, for instance, on the presence of complex eigenvalues at the enclosed critical points, on vorticity, or on a predominance of curl within a region.

Extracted vortex region candidates are bounded by loops that start and end at saddle points. Due to continuity, at least one saddle point is included in the closure of a bounded region R_i. The vortex region candidates R_i have disjoint boundaries but may be nested. Note that this method can be applied to 2-dimensional slices cut through a 3-dimensional field. Details of the algorithm can be found in the article of Petz *et al.* [12].

2.6. Haller's method

We now present a method defined precisely for 3-dimensional flows and that uses a definition of vortex not depending upon the frame of reference. By using a decomposition similar to the one we used at the beginning of the subsection on Greek methods, the gradient of a smooth velocity field $u(x,t)$ can be written as

$$\nabla\vec{v} = S + \Omega, \tag{28}$$

where S and Ω are defined as in Subsection 2.1. We can now define the strain acceleration vector as follows.

$$M = S_t + (\nabla S)\vec{v} + S(\nabla\vec{v}) + (\nabla\vec{v})^T S \tag{29}$$

In this expression S_t denotes the partial derivative of S with respect to t. The cone field is defined as follows.

$$Z(x,t) = \{\xi \mid \langle\xi, S(x,t)\xi\rangle = 0\} \tag{30}$$

We assume that $detS(x,t) \neq 0$ holds. By the assumption of incompressibility, the symmetric tensor S has trace zero and hence one positive eigenvalue and one negative eigenvalue. Thus we can assume, without loss of generality, that $sign(s_1) = sign(s_2)$ and $sign(s_2) \neq sign(s_3)$. In the following coordinates

$$\xi = \eta_1 e_1 + \eta_2 e_2 + \eta_2 e_3, \tag{31}$$

the cone field Z satisfies

$$\eta_3^2 = a\eta_1^2 + (1-a)\eta_2^2, \tag{32}$$

where $a(x,t) = -\frac{s_1(x,t)}{s_3(x,t)} \in (0,1)$.

Depending on the value of $sign(s_3)$, the flow goes to the origin ($s_3 < 0$) or leaves the origin ($s_3 > 0$). This defines an elliptic cone in the strain basis. Observe that this cone field is not in the spatial coordinates, but rather in its tangent space. Pulling it down into spatial coordinates (by looking at its image under $exp : TX \to X$) is not trivial, but the computation can be carried out. If we can compute that image locally for every point, we can restrict our function M to this image at every point. We call it M_Z.

The hyperbolic domain $H(t)$ is the set of points x at which M_Z is positive definite. This domain is subdivided into two separate domains, $H(t)^-$ and $H(t)^+$, where the superscript of each domain depends on the sign of the eigenvector s_3. The elliptic domain $E(t)$ is the set where M_Z is indefinite. We now have everything we need to define our criterion for the existence of vortices (called the M_Z-criterion): a vortex is a set of fluid trajectories along which M_Z is indefinite.

3. Tests and Results

Some methods presented in the previous section have been implemented and tested on four different 2-dimensional (planar or axisymmetric) flows. They correspond to the following test cases:

1. ABC flow: 2-dimensional planar analytical solution of 2-dimensional Euler equations;
2. Immersed droplet flow (test case no. 1 in Haller's article): analytical axisymmetric solution for Stokes equations;
3. Driven cavity flow: direct numerical solution of 2-dimensional Navier-Stokes equations for different values of the Reynolds number Re (downloaded from www.cavityflow.com);

4. Flow around a cylinder: Lattice-Boltzmann solution of the 2-dimensional Navier-Stokes equations, computed by Tim Reis.

Note that we present results for the Q-method only since it yields results identical to those of the other Greek approaches.

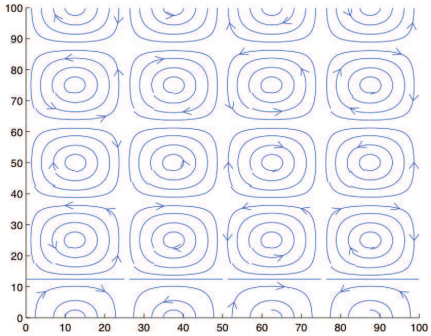


Figure 3. Streamlines of the ABC flow.

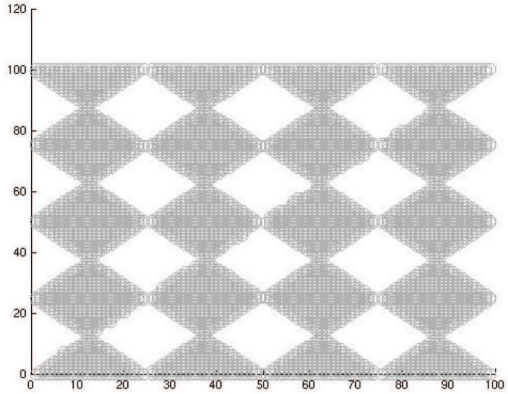


Figure 4. Q-criterion applied on the ABC flow.

Figure 3 shows the streamlines for the ABC flow to which the Q-method was applied in order to detect the vortices. Figure 4 shows how the Q-criterion detects the vortices in the ABC flow. Once again, the Q-method is used to identify the vortices and results are presented in Figure 6. The streamlines of the immersed droplet flow, the second test case, are shown in Figure 5. The third test case corresponds to the driven cavity flow. Experiments were carried out for different values of Reynolds number. Streamlines are pictured in Figures 7, 11, and 13. The Q-method was tested in each case as presented in Figures 8, 12, and 14. The critical point method was also applied to a 2-dimensional driven cavity flow. The zero-contour lines of u and v intersect within the potential cell, showing that the critical point is located inside the cell (see Figure 9). Applying this method, the location of the vortex centre within the domain is obtained, as shown in Figure 10.

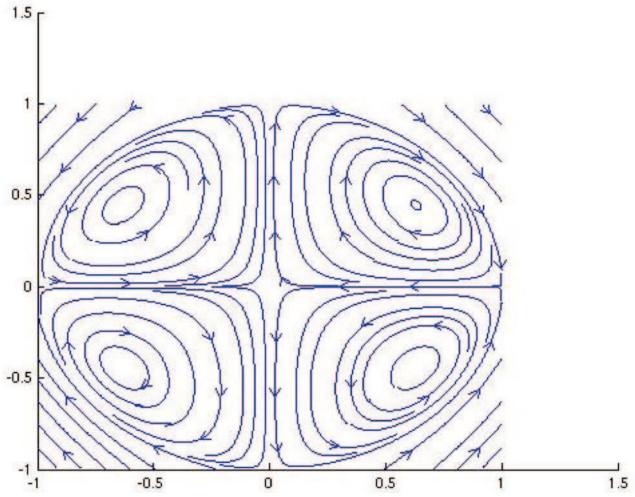


Figure 5. Streamlines for the immersed droplet flow.

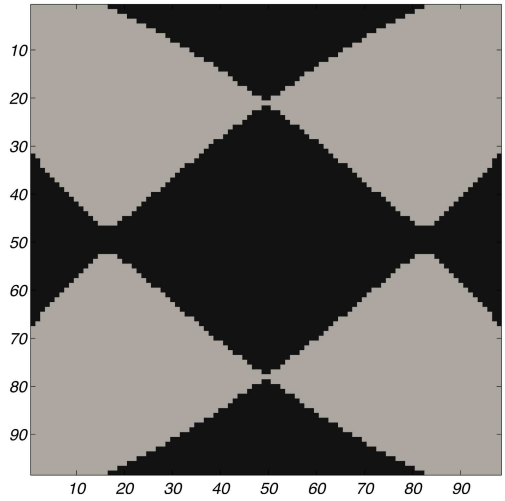


Figure 6. Q-criterion applied on the immersed droplet flow.

We now present test results for data constructed by Lattice Boltzmann methods. We first tested the Q-method on our data. The velocity plots and the vorticity plots are shown in Figure 15. In what follows, we compute the contours $Q = 0$ for the Q-method and the contours $H_e = 0$ for the Eigen Helicity method. We use these contours as a crude approximation to the boundaries of the vortex regions. The rationale for using $H_e = 0$ as an approximation is that the quantity H_e is only defined within these regions. The quantity H_e is not defined outside these contours, which implies that $\nabla\vec{u}$ has real eigenvalues outside these closed regions. This, in turn, rules out the presence of vortices. The contours are illustrated in Figure 16.

From Figure 16, we can see that the Q-method and the H_e-method identify the vortex regions in a qualitative fashion. They also omit parts of the vortex regions and identify certain vortex-free

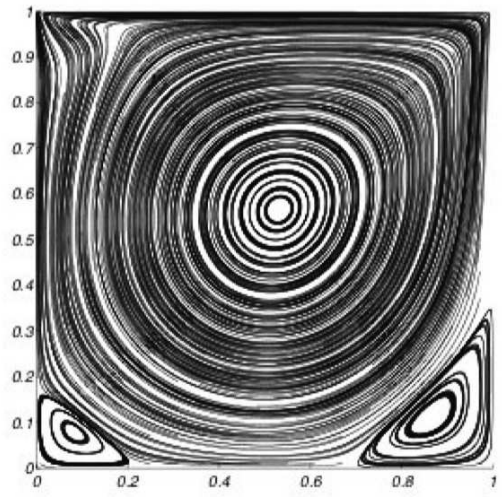


Figure 7. Streamlines of the driven cavity flow with $Re = 1000$.

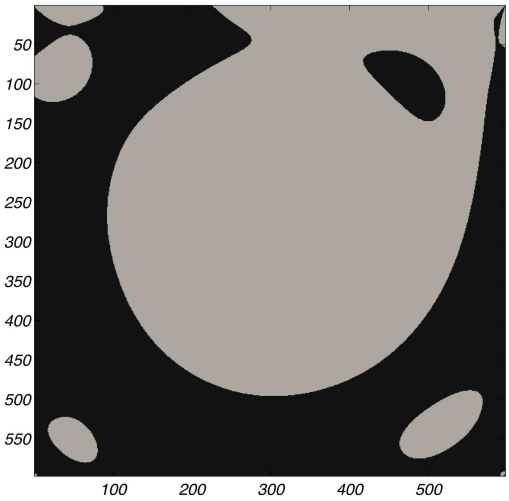


Figure 8. Q-criterion applied on the driven cavity flow with $Re = 1000$.

regions as being vortices. This is consistent with the findings of Zhang *et al.* [17], who report similar problems with the Q-method. It is disappointing to find that we do not get better results with the H_e-method. To improve it, we could use a normalized version of the Eigen Helicity, i.e., the expression

$$H_e^* = \frac{|H_e|^2}{\max\left\{|H_e|^2\right\}}, \tag{33}$$

which takes values in the range $[0, 1]$. The vortices correspond to the points such that $H_e^* = 1$ and one could use the threshold value $H_e^* = 0.95$ in order to trace the boundaries of the vortex regions (see Zhang *et al.* [17]). The next two figures (17 and 18) were obtained using the Boundary Loop

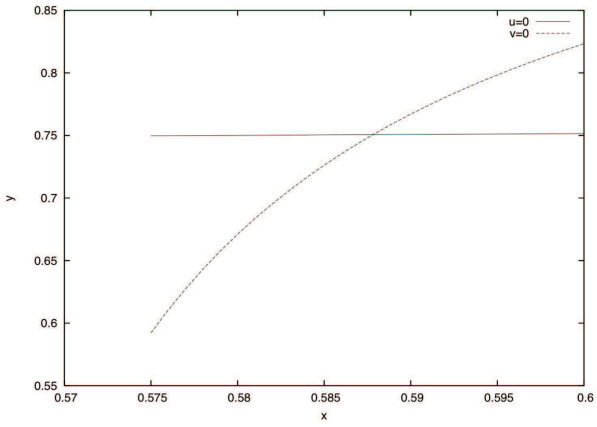


Figure 9. Critical point's location in the potential cell.

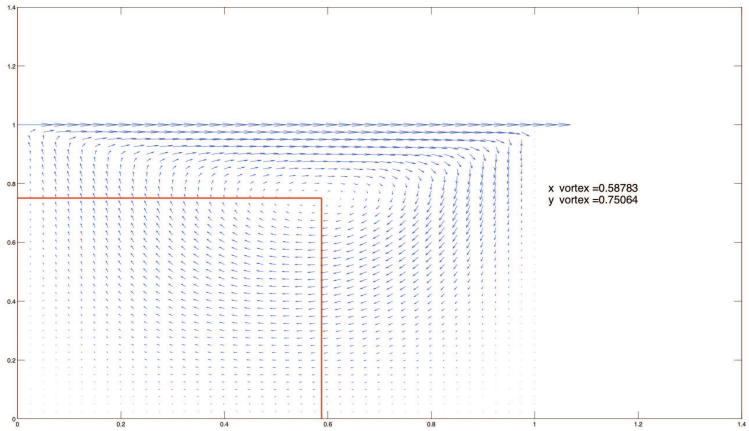


Figure 10. Centre of the vortex in the driven cavity flow vector field with $Re = 900$.

and the λ_2-methods to detect vortices for time-dependent simulation of the Kármán vortex behind a cylinder. They are based on the results of Petz *et al.*, and were not implemented by the workshop team.

Interestingly, Figures 17 and 18 show how the vortex regions extracted with the Boundary Loop method correspond to the intuitive idea of a vortex region more closely than the regions extracted with the λ_2-method. This might be due to the fact that the Boundary Loop method tries to extract the maximum region of swirling flow, while the λ_2-method defines more precisely the core of the vortex region. Note that this method can be applied to 2-dimensional slices cut through a 3-dimensional field.

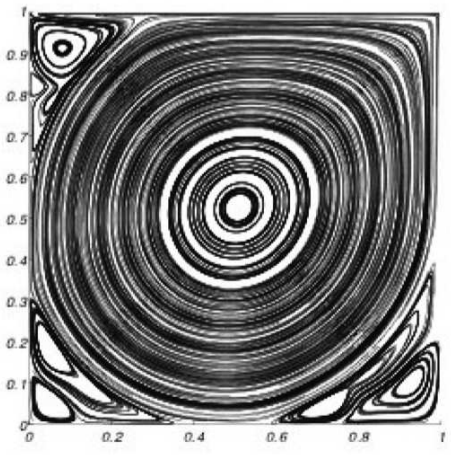


Figure 11. Streamlines of the driven cavity flow with $Re = 10000$.

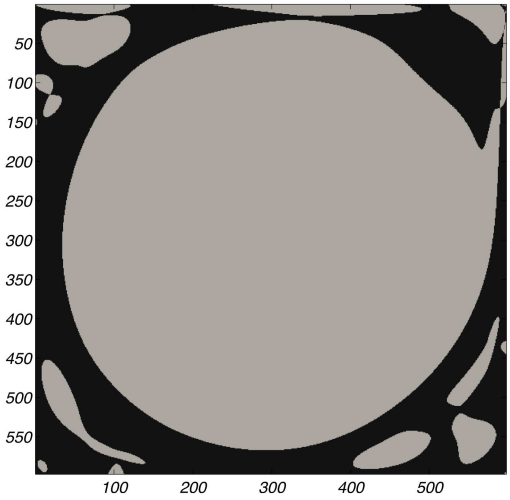


Figure 12. Q-criterion applied on the driven cavity flow with $Re = 10000$.

4. Conclusions and Further Work

After spending a week surveying the literature and implementing some existing classical methods, our conclusions are that it is fairly easy and cheap to determine a coarse approximation to regions of recirculating flow. These methods have been tested on relatively simple 2-dimensional synthetic and classical flows, for which they provide adequate solutions. We could, however, foresee cases where distinguishing regions would be tricky. The methods did not succeed in delineating the recirculating zones with accuracy. Adjustments to the method with a threshold value did not improve the results.

As pointed out in Haller's article [3], it may be necessary to go beyond the classical methods and implement his method or to combine the classical methods with some other detection method or postprocessing. Given the further approximations that must be made in the coupling with

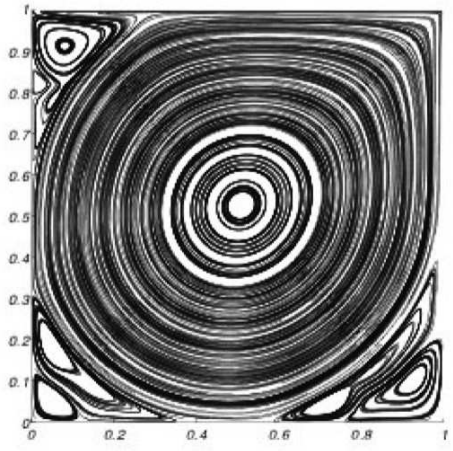


Figure 13. Streamlines of the driven cavity flow with $Re = 20000$.

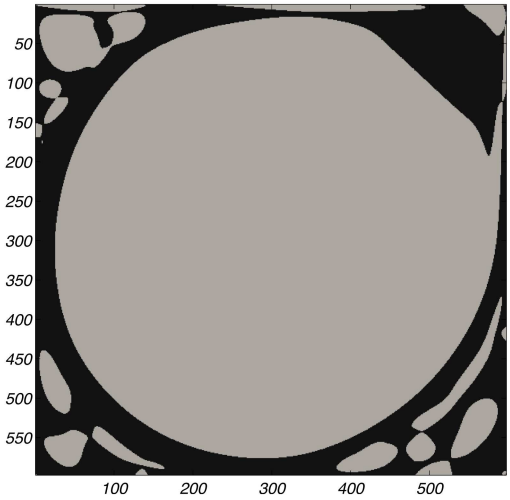


Figure 14. Q-criterion applied on the driven cavity flow with $Re = 20000$.

the detailed chemistry (Chemkin) model, we believe that the cheap approximate methods are a viable option, perhaps in conjunction with the temperature and mixture fraction indicators used previously. Of course, the validation of the subsequent results using Chemkin will determine the viability of these proposals. The extension to 3-dimensional flows appears to be feasible but the team did not have the time to study the concepts involved during the week of the workshop. The critical point method and the geometric methods are not viable options for the 3-dimensional case. Compressibility is not expected to be an issue except in the case of Haller's method.

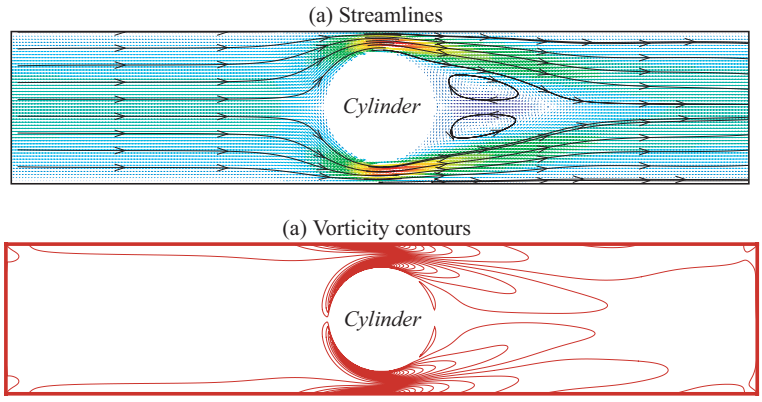


Figure 15. Results for the Q-method

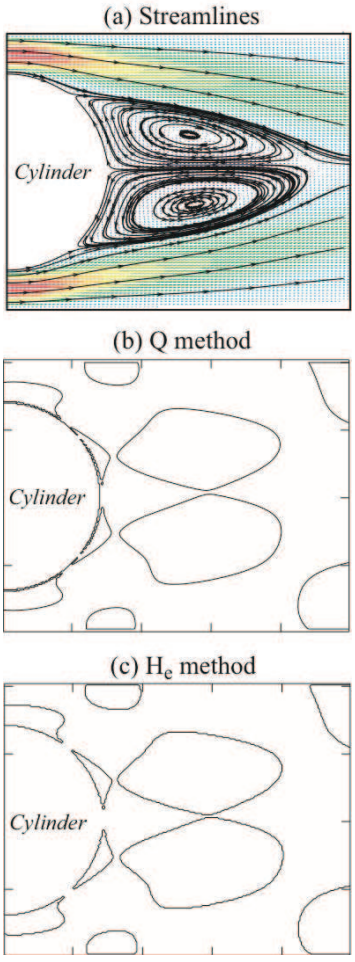


Figure 16. Streamlines of the flow behind a cylinder and contours for $Q = 0$ and $H_e = 0$.

Here are some recommendations and possibilities for further work on this topic. As mentioned before, Haller's method should be implemented, if easily tractable. The detection methods could

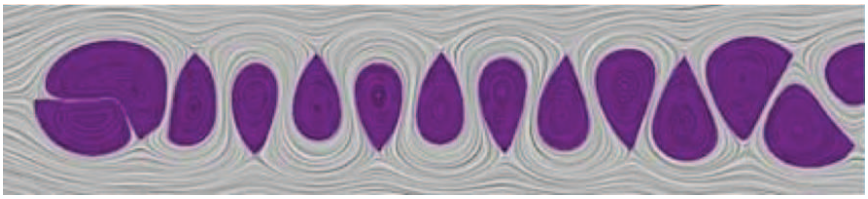


Figure 17. Boundary Loop method applied on Kármán vortex behind a cylinder.

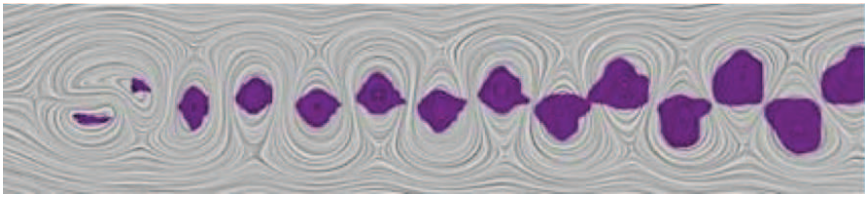


Figure 18. λ_2-method applied on Kármán vortex behind a cylinder.

be combined with a technique to separate the recirculation regions. This technique could be based on a crude expansion of the identified regions, perhaps using some detection method based on local velocities. Another idea would be to design a procedure to find the streamline "separating" two regions; one could identify either a normal or low normal flux line (streamline) or a line along which the curl of neighbouring velocities is zero (parallel) and the scalar product is negative (opposite directions). This approach would only cover cases of abutting recirculation zones of opposite direction, which happens often but not always.

Also, the team thinks there might exists some method to find a trough on a topological map. Basically this would involve the computation of the path of lowest flux (or equivalently the locus of local minima). We did also see references to dividing streamlines in meteorology but could not pinpoint methods to compute them. Once these regions are defined, finding the fluxes between the regions will be easy: one can use the CFD software if it is indeed flux-based. The faces of the cells along the delineating curves could be used as locations across which fluxes could contribute to the total flux out or into the region. Residence times are also assumed to be computed as in the previous Rolls Royce work. We did not have the time to look into this aspect although we were aware of it. Not all members of the team agreed on the level of difficulty involved. It would be interesting to compare two sequences: the zone-extraction post-processing followed by the time-averaging of the CFD data, and the time-averaging of the CFD data followed by the zone-extraction post-processing. Because of reaction nonlinearities, the output from Chemkin might be very sensitive to the sequence.

References

1. M.S. Chong, A.E. Perry, B.J. Cantwell, A General Classification of Three-Dimensional Flow Fields, *Phys. Fluids A*, **2**, 765-777, 1990.
2. U. Dallmann, Topological Structures of Three-Dimensional Flow Separation, *DFVLR-IB Report*, No. 221-82 A07.
3. G. Haller, An objective definition of vortex, *J. Fluid Mech.*, **525**, 1-26, 2005.
4. F. Higuera, S. Succi and R. Benzi, *Europhys. Lett.*, **9**, 345 (1989); http://dx.doi.org/10.1209/0295-5075/9/4/008
5. H.W. Hirsch, C.C. Pugh, M. Shub, Invariant manifolds, *Lect. Notes Math.*, **583**, 1977.

6. J.C.R. Hunt, A.A. Wray, P. Moin, Eddies, Stream and Convergence Zones in Turbulent Flows, *Center for Turbulence Research Report*, CTR-S88, 193-208.
7. J. Jeong, F. Hussain, On the identification of a vortex, *J. Fluid Mech.*, **285**, 69-94, 1995.
8. R. D. Knowles, M. V. Finnis, A. J. Saddington, K. Knowles, Novel Post-Processing Techniques for Vortical Flows, Second international symposium on integrating CFD and experiments, Cranfield University, 5-6 September, 2005.
9. V. Kolar, Vortex identification: New requirements and limitations, *International Journal of Heat and Fluid Flow*, **28**, p. 638, 2007.
10. V. Kolar, Compressibility effect in vortex identification, *AIAA Journal*, **47**, p. 473, 2009.
11. Y. Levy, D. Degani, A. Seginer, Graphical visualization of vortical flows by means of helicity, *AIAA J.*, **28**, 1347, 1990.
12. C. Petz *et al.*, Hierarchical Vortex Regions in Swirling Flow, *Eurographics-IEEE-VGTC Symposium on Visualization*, 863-870, 2009.
13. I.A. Sadarjoen, F.H. Post, Geometric methods for vortex extraction, *Eurographics-IEEE TVCG Symposium on Visualization*, 53-62, 1999.
14. A. Surana, O. Grunberg, G. Haller, Exact theory of three-dimensional flow separation. Part I. Steady separation, *J. Fluid. Mech.*, **564**, 57-103, 2006.
15. H. Theisel *et al.*, Grid-Independent Detection of Closed Stream Lines in 2D Vector Fields, *Proc. Vision, Modeling and visualization 2004*, 421-428, 2004.
16. H. Vollmers, H.-P. Kreplin, H.U. Meier, Separation and vortical-type flow around a prolate spheroid - Evaluation of relevant parameters, *Proc. of the AGARD Symposium on Aerodynamics of Vortical Type Flows in Three Dimensions*, Rotterdam, AGARD-CP-342, 1983.
17. Shuang Zhang, Dipankar Choudhury, Eigen helicity density: A new vortex identification scheme and its application in accelerated inhomogeneous flows, *Physics of Fluids*, **18**, 2006.
18. Jialin Zhong, Thomas S. Huang, Extracting 3D Vortices in Turbulent Fluid Flow, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, Number 2, February 1998.

Optimisation des limites d'un réseau de transport d'énergie

Michel Gendreau
Coordonnateur, École Polytechnique de Montréal

Jean-Claude Rizzi et Guy Vanier
Représentants industriels, IREQ, Varennes

Saeid Ghafouri et Vahid Famildardashti
Étudiants, CIRRELT et École Polytechnique de Montréal

Sivan Altinakar et Martin Turcotte
Étudiants, GERAD et École Polytechnique de Montréal

Peter Miasnikof
Étudiant, University of Toronto

Maaike Schakenbos
Étudiante, GERAD et École Polytechnique de Montréal

Pierre-Luc Carpentier
Rédacteur, CIRRELT et École Polytechnique de Montréal

1. Introduction

Avec plus de 33000 km de lignes électriques à haute tension et plus de 500 postes de distribution, le réseau québécois de transport d'électricité compte parmi les plus importants en Amérique du Nord. Comme l'indique la figure no 1, le réseau québécois de transport d'électricité a une topologie fortement radiale avec deux axes de transmission alimentant en électricité le sud-ouest de la province. Un des axes de transmission est orienté vers les complexes hydroélectriques de la Baie-James et un autre axe est orienté vers la Côte-Nord, où sont situés les complexes Manicouagan, Outardes, Sainte-Marguerite et Churchill Falls. L'énergie est transportée environ 1000 km plus au sud dans la vallée du Saint-Laurent, où se trouvent les zones densément peuplées. Plusieurs postes d'interconnexion avec les États-Unis et le reste du Canada sont également situés au sud du Québec et permettent l'exportation d'énergie sur les marchés nord-américains de l'électricité.

Pour exploiter le réseau de transport d'électricité de manière fiable et sécuritaire, il est essentiel de connaître la capacité (ou limite) de transport du réseau à tout moment afin d'éviter de surcharger celui-ci. Dans les faits, plusieurs facteurs peuvent réduire substantiellement la capacité de transport du réseau. Dans un premier temps, les opérations d'entretien qui doivent être effectuées sur les équipements du réseau obligent la division TransÉnergie d'Hydro-Québec à mettre hors service des portions du réseau électrique pendant certaines périodes. De plus, les conditions météorologiques ont une influence directe sur la capacité des lignes électriques. La production de chaleur associée au transport de l'énergie électrique a pour effet d'allonger les lignes à haute tension. La puissance transmise doit être limitée de manière à empêcher que l'allongement des fils électriques ne soit trop grand.

À l'exception de la période de pointe hivernale, le réseau est donc généralement incomplet (notamment parce qu'il faut entretenir les équipements électriques). Des bris d'équipement peuvent aussi survenir inopinément sur le réseau, ce qui limite la capacité de transport du réseau. Tous ces

effets doivent être pris en compte pour évaluer à tout moment, et de façon réaliste, la capacité du réseau. La combinaison formée par l'état physique du réseau et les facteurs limitant sa capacité est appelée «configuration». Certaines configurations, correspondant au bris d'un petit nombre d'équipements, sont appelées «configurations élémentaires». Le problème d'optimisation des limites de transport est décrit de façon détaillée dans la deuxième section de ce rapport. Un modèle mathématique, formulé par les chercheurs de l'IREQ, est présenté dans la troisième section. En soumettant leur problème aux participants de l'atelier, les chercheurs de l'IREQ désiraient obtenir des propositions d'algorithmes pour le résoudre. Par conséquent, dans la quatrième section, on trouvera les méthodes d'optimisation proposées pendant l'atelier. La cinquième section contient les conclusions de cette étude.

Figure 1. Carte du réseau de transmission à haute tension d'Hydro-Québec

2. Description du problème d'optimisation

La division TransÉnergie d'Hydro-Québec a la responsabilité d'exploiter et entretenir le réseau de transmission québécois le plus efficacement possible, à un coût raisonnable, de façon sécuritaire et de manière à assurer la pérennité de ses installations. Pour ce faire, les exploitants de TransÉnergie établissent la limite de transport sur chacune des cinq sections du réseau québécois (qui sont relativement indépendantes). Les tables utilisées pour déterminer ces limites contiennent des données sur environ 100 configurations élémentaires, dont chacune a un impact substantiel et connu sur la capacité du réseau. À chacune de ces configurations élémentaires correspond une restriction, obtenue grâce à un ensemble de simulations qui représentent le comportement du réseau dans divers scénarios plausibles. Cependant, dans la pratique, on observe des configurations qui ne sont

pas élémentaires mais qui peuvent être considéréres comme des combinaisons de configurations élémentaires. Les tables mentionnées ci-dessus ne contiennent donc pas toutes les configurations parce qu'il y en a un nombre exponentiel.

Comme l'illustre la figure no 2, lorsque le réseau est complet, la limite de transport de base L^B peut être affectée par la mise hors service des équipements A, B et C. Le retrait de l'équipement A a pour conséquence de faire augmenter la limite de transport (en d'autres termes $L^S_{\overline{A}}$ est supérieure à L^B). Notez que la limite $L^S_{\overline{A}}$ est obtenue par simulation. Si on retire l'équipement B (respectivement C) du réseau, la limite de transport diminue à $L^S_{\overline{B}}$ (respectivement $L^S_{\overline{C}}$). Comme le révèlent aussi les résultats de simulation, le retrait simultané de plusieurs équipements n'a pas un effet additif sur la limite de transport. Notez que certaines configurations élémentaires correspondent à un état du réseau dans lequel plusieurs équipements (formant un groupe) ont été retirés.

On peut approximer l'effet du retrait simultané de certains équipements ou groupes d'équipements en additionnant les effets des retraits indépendants de chacun des équipements ou groupes d'équipements sur le réseau de transport. Cette approximation est facile à calculer et sécuritaire. Son inconvénient est de borner de manière trop rigoureuse la capacité de transport du réseau, ce qui fait perdre des revenus à la division TransÉnergie. Par exemple, retirer les équipements B et C du réseau a pour conséquence de faire diminuer la limite de transport à $L^S_{\overline{BC}}$ dans les simulations. On peut constater que la limite approximée $L^R_{\overline{BC}}$ est inférieure à la limite simulée $L^S_{\overline{BC}}$; l'écart positif entre ces deux quantités, $L^S_{\overline{BC}} - L^R_{\overline{BC}}$, est dénoté $E_{\overline{BC}}$.

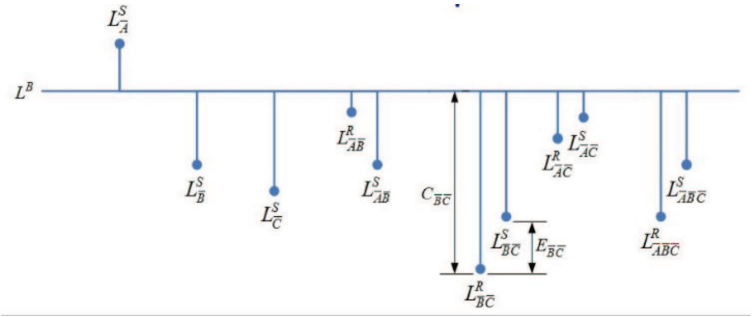


Figure 2. Représentation graphique des limites de transfert sur le réseau de transmission électrique

Un très grand nombre de combinaisons (c'est-à-dire de configurations non élémentaires) sont possibles, et le problème d'optimisation étudié pendant l'atelier consiste à choisir un ensemble de configurations élémentaires permettant de représenter le mieux possible toutes les configurations, c'est-à-dire tous les états possibles du réseau. Ceci doit être fait de façon sécuritaire en évitant les cas où on obtiendrait des écarts négatifs entre la limite simulée et la limite approximée.

3. Formulation mathématique du problème d'optimisation

L'objectif du problème d'optimisation est de sélectionner, parmi les configurations élémentaires i disponibles, celles qui permettent d'estimer de manière sécuritaire les limites de transfert pour toutes les configurations possibles du réseau. De plus, cet objectif doit être atteint en minimisant les pertes induites par l'approximation des limites de transfert, et ce, en tenant compte de la probabilité d'occurrence de chaque scénario. Pour formuler ce problème, nous introduisons les notations suivantes : L^R_k dénote la limite approximée pour la configuration k, L^S_k la limite simulée

pour la configuration k, E_k l'écart entre ces deux limites et C_k l'effet combiné des restrictions élémentaires pour la configuration k. Les variables de décision du modèle sont les y_i et les n_i. La variable binaire y_i prend la valeur 1 si et seulement si la configuration élémentaire i est choisie. La variable n_i représente un facteur entier d'ajustement de restriction pour la configuration élémentaire i.

Les chercheurs de l'IREQ ont formulé le problème comme ci-dessous. Notez que toutes les quantités non définies dans le modèle lui-même, à part les variables y_i et n_i, sont des constantes dont les définitions seront données ou rappelées plus loin. Notez aussi que Z dénote l'ensemble des nombres entiers et que par convention, n_i est égal à 0 lorsque y_i est lui-même égal à 0.

$$\min_{y_i,n_i} \sum_{k\in K} \omega_k P\left(E_k\right) \tag{1}$$

sous les contraintes

$$E_k = L_k^S - L_k^R \ \forall k \in K$$

$$L_k^R = L_k^B - C_k \ \forall k \in K$$

$$C_k = \sum_i y_i u_{ki} \left(R_i + g \cdot n_i\right) \ \forall k \in K$$

$$\sum_i y_i \leq N_{max}^R$$

$$n_{min} \leq n_i \leq n_{max}, n_i \in Z, \forall i \in I$$

$$y_i \in \{0,1\}, \forall i \in I$$

La fonction-objectif à minimiser est obtenue en additionnant, pour chacune des configurations k possibles, le produit d'un écart E_k par la pénalité $P(E_k)$ auquel il est associé. Chaque terme de la série est pondéré par un poids ω_k qui reflète la probabilité d'occurrence de la configuration k. Les configurations qui surviennent rarement en pratique auront un poids proche de 0 alors que les configurations qui surviennent souvent auront un poids proche de 1. L'écart E_k correspond à la différence entre la limite simulée et la limite approximée (obtenue en combinant différentes configurations élémentaires). La fonction de pénalité $P(E_k)$ a une forme convexe et linéaire par morceaux. Le minimum de cette fonction survient lorsque E_k est égal à 0. La fonction $P(E_k)$ pénalise fortement les configurations correspondant à un écart négatif de manière à éviter les situations non sécuritaires. Dans une moindre mesure, la fonction $P(E_k)$ pénalise aussi les écarts positifs ; elle a une forme linéaire pour ces écarts-là. Le tableau 1 regroupe les définitions des quantités utilisées dans le modèle.

Durant la semaine, l'équipe de travail a proposé d'utiliser la fonction de pénalité convexe et linéaire par morceaux suivante.

$$P(E_k) = \begin{cases} -4.95 \times 10^5 - 10^4 E_k, & \text{si } E_k < -50; \\ -100 E_k, & \text{si } -50 \leq E_k < 0; \\ E_k, & \text{si } 0 \leq E_k < 150; \\ -1350 + 10 E_k, & \text{si } 150 \leq E_k. \end{cases} \tag{2}$$

Cette fonction de pénalité vise dans un premier temps à empêcher les configurations non sécuritaires (correspondant à un écart négatif) en pénalisant fortement ces configurations. Un coefficient de 10000 s'applique à E_k lorsque cette quantité est "trop" négative. La fonction de pénalité est conçue

Tableau I. Définitions des quantités utilisées dans le modèle

Variables	Définition
k	Indice représentant une configuration
i	Indice représentant une configuration élémentaire
K	Ensemble des configurations
I	Ensemble des configurations élémentaires
y_i	Variable booléenne indiquant si la configuration élmentaire i est choisie
C_k	Cumul des restrictions associées aux configurations élémentaires pour la configuration k
E_k	Écart pour la configuration k
g	Granularité d'ajustement de restrictions
u_{ki}	Facteur de respect des règles d'application des restrictions
L_k^B	Limite simulée de la configuration de base utilisée pour dériver la configuration k
L_k^S	Limite simulée pour la configuration k
L_k^R	Limite obtenue par cumul de restrictions pour la configuration k
n_i	Facteur d'ajustement de restrictions pour la configuration élémentaire i
n_{max}	Borne supérieure sur n_i
n_{min}	Borne inférieure sur n_i
N_{max}^R	Nombre maximal de configurations élémentaires choisies
$P(E_k)$	Pénalité associée à l'écart E_k
R_i	Valeur minimale de la restriction associée à la configuration élémentaire i
ω_k	Poids reflétant la probabilité d'occurrence de la configuration k

aussi de manière à pénaliser les solutions réalisables en fonction de leur écart. Il pourrait être intéressant de considérer d'autres modèles mathématiques, c'est-à-dire des modèles où il n'est pas nécessaire de déterminer les coefficients u_{ki} a priori.

4. Méthode de solution proposée

Étant donné la structure hautement combinatoire et la taille du problème d'optimisation considéré, l'équipe a pensé qu'il ne serait pas judicieux de tenter de résoudre le problème par une méthode exacte. Elle a plutôt choisi de concentrer ses efforts sur le développement d'algorithmes heuristiques. Dans un premier temps, une heuristique basée sur les méthodes de descente et développée par les chercheurs de l'IREQ a été présentée par ceux-ci. Une approche métaheuristique a aussi été étudiée afin de résoudre le problème d'optimisation de manière efficace. Parmi les métaheuristiques basées sur les trajectoires d'une seule solution («métaheuristiques de type S»), on compte les différentes versions de la recherche locale, les méthodes de type tabou, la recherche à voisinage variable, et les méthodes de type «Ruin and Recreate». Parmi les métaheuristiques utilisant une population de solutions («métaheuristiques de type P»), on compte les algorithmes génétiques, les algorithmes de colonies de fourmis et les algorithmes mémétiques.

4.1. Algorithme de recherche avec tabous

La méthode de résolution qui a été retenue est une recherche avec tabous, dont le pseudocode est donné ci-dessous. Cet algorithme part de la solution initiale $s^0 = (y^0, n^0)$. Soit s la solution courante. L'algorithme explore le voisinage $N(s)$ de la solution courante (voir ci-dessous pour la description de ce voisinage). La meilleure solution du voisinage, dénotée s', est comparée à la meilleure solution s^* trouvée jusqu'ici (le coût de s^* est dénoté f^*). Si la solution s' a un coût inférieur au meilleur

coût trouvé jusqu'ici, alors la meilleure solution trouvée s^* est mise à jour, c'est-à-dire qu'elle est remplacée par s'. Ces étapes sont répétées tant que le critère d'arrêt n'est pas satisfait. L'algorithme retourne la meilleure solution trouvée $s^* = (y^*, n^*)$ ainsi que le coût de celle-ci.

Tableau I.

Algorithme de recherche avec tabous

1. Entrées
$s^0 = (y^0, n^0)$: solution initiale

2. Initialisation
$s \leftarrow s^0$: solution courante
$s^* \leftarrow s^0$: meilleure solution trouvée jusqu'ici
$f^* = f(s^*)$
$T \leftarrow \emptyset$: liste

3. Calcul itératif d'une solution heuristique
TANTQUE Critère_arrêt = FAUX
    Choisir un voisinage $N(s)$ pour la solution courante s
    $s' \leftarrow \arg\min_{s'' \in N(s)} f(s'')$: trouver la meilleure solution permise dans le voisinage $N(s)$
    SI $f(s') \leq f^*$
    ALORS
        $s^* \leftarrow s'$
        $f^* \leftarrow f(s')$
    FINSI
    Ajouter le déplacement à la liste T pour éviter d'examiner la solution une autre fois
    $s \leftarrow s'$
    SI le nombre maximal d'itérations est atteint
    ALORS
        Critère_arrêt ← VRAI
    FINSI
FINTANTQUE

4. Sorties
$s^* = (y^*, n^*)$: solution heuristique
f^* : valeur finale de la solution heuristique

4.2. Formulation du voisinage

À chaque itération, l'algorithme de recherche avec tabous explore un voisinage $N(s)$ de la solution courante s. L'algorithme proposé par l'équipe choisit au hasard un des deux voisinages suivants : le voisinage 2-échanges $N_e(s)$ et le voisinage d'ajustement $N_a(s)$. Le voisinage $N_e(s)$ consiste de toutes les solutions réalisables obtenues en transposant les valeurs de deux variables de décision dans la solution courante. Voici la définition en termes mathématiques de $N_e((y^t, n^t))$, où (y^t, n^t) représente la solution courante à l'étape t.

$$\{(y,n)\,|\,\exists(i_1,i_2) \text{ tel que } i_1 \neq i_2, y^t_{i_1} = 1, y^t_{i_2} = 0, y_{i_1} = 0, y_{i_2} = 1, n_{i_1} = 0, n^t_{i_2} = 0, \text{ et } y^t_i = y_i, n^t_i = n_i, \text{ pour } i \neq i_1, i_2\}$$

Le voisinage $N_a((y^t, n^t))$, quant à lui, consiste des solutions obtenues à partir de la solution courante en modifiant la valeur de n_{i_1} pour un indice i_1 donné.

5. Conclusion

Pendant cet atelier de résolution de problèmes industriels, notre équipe a étudié le problème d'optimisation des limites de transfert d'énergie. Dans un premier temps nous avons essayé de comprendre le problème pratique auquel est confrontée la division TransÉnergie d'Hydro-Québec. Nous avons ensuite évalué les caractéristiques du programme mathématique proposé par les chercheurs de l'IREQ et proposé un algorithme de recherche avec tabous. En effet, la complexité du problème ne nous permet pas d'espérer trouver la solution optimale à l'aide d'un algorithme exact.

Références

1. Gendreau, M. et J.-Y. Potvin. "Tabu Search", Search Methodologies : Introductory Tutorials in Optimization and Decision Support Techniques, E. Burke et G. Kendall (éd.), Springer, New York, NY (USA), pp. 165-186, 2005.
2. Glover, F. et M. Laguna. Tabu Search, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.
3. Talbi, E. G. Metaheuristics : From Design to Implementation. Wiley, 2009.
4. Valette, A., J. A. Huang, S. Guillon, L. Loud, G. Vanier, F. Lévesque, L. Riverin, J.-C. Rizzi et F. Guillemette. An Integrated Approach for Optimizing Dynamic Transfer Limits at Hydro-Quebec. Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, IEEE, 20-24 juillet 2008.

Real-time Placement of Labels on a Geographical Map

Odile Marcotte
Coordinator, CRM and UQÀM

Patrick St-Louis
Industrial representative, GIRO Inc.

Mo'tassem Al-arydah
Student, University of Ottawa

Asma Mdimagh and Katayoon Moazzami
Students, École Polytechnique de Montréal

Duc Minh Vu
Student, Université de Montréal

Kamel Bentahar
Mathematical Institute, University of Oxford

Alain Hertz
Professor, École Polytechnique de Montréal

1. Context of the problem

GIRO Inc. is a company specializing in the production of routing software for public transit and postal distribution. It is present in 23 countries on 4 continents, has 220 employees and is growing steadily. For instance, in the case of postal distribution, it may provide a client with a set of routes that cover all the deliveries, and such that each route corresponds to a specific driver. The client must be allowed to modify the routes provided by GIRO. Thus GIRO has designed a module for displaying a city map (or part thereof) and the routes constructed by its software. This map contains many entities (street names, delivery codes, routes, etc.). As noted before, the client also has the possibility of modifying some entities displayed on the map (routes, for instance). When some entities are modified, their labels must be redisplayed very quickly. GIRO thus faces a combinatorial problem similar to the classical label placement problem. In the following sections of this report, we define formally the classical problem, describe the features of the problem faced by GIRO, and outline a proposed solution.

2. Formal definitions of the classical placement problem

We are given a city map (or part thereof), containing several types of entities: streets, rivers, houses, etc. For each entity there is a finite set of positions where the entity label may be displayed. The classical label placement problem consists of placing the labels on the map such that it be legible and contain as much information as possible. Among other things, the "legibility" requirement means that

- a label must be placed fairly close to the corresponding entity (i.e., the label position must be of "high quality"), and

- there is not much overlap (or none at all) between the labels.

We now attempt to formalize this problem. Let G be an undirected graph (the *conflict graph*) defined as follows:

- the vertices of G are of the form (i, j) (or ij for short), where i denotes a label and $j \in P_i$ one of the positions for label i,
- there is an edge between ij and $i\ell$ for any pair $\{j, \ell\}$ of distinct positions, and
- there is an edge between ij and $k\ell$ if the labels i and k are different and the positions j and ℓ overlap.

Each label i has a priority p_i. Assume first that we don't allow overlaps. The problem is then to choose a *stable set* S of vertices of G such that the quantity $\sum_{ij \in S} p_i$ is maximum. Note that the requirement that S be a stable set implies that S contains at most one vertex of the form ij for each i.

For an illustration, consider a problem where there are three entities and there are 4 (resp. 3, 4) alternative positions for the first (resp. second, third) entities. Thus P_1 (resp. P_2, P_3) equals $\{1, 2, 3, 4\}$ (resp. $\{1, 2, 3\}$, $\{1, 2, 3, 4\}$). Here is a picture of the conflict graph for this example.

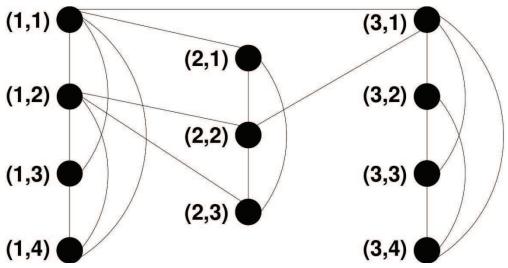


Figure 1. Conflict graph.

Assume now that we allow overlaps between labels. We first define the notions of *vertex cost* and *edge cost*.

- The cost of vertex ij, denoted by c_{ij}, is defined as $p_i \cdot weight(j)$, where $weight(j)$ is a measure of the quality of position j.
- The cost of edge $\{ij, k\ell\}$, denoted by $c_{ij,k\ell}$, is defined as

$$\max\{c_{ij}, c_{k\ell}\} \cdot (c_{ij} + c_{k\ell}) \cdot \text{ percentage of overlap.}$$

- The cost of a subgraph is the sum of the costs of all its edges.

The problem consists of choosing a subset S such that

- S contains exactly one vertex of the form ij for each i, and
- the cost of the subgraph induced by S is minimum.

Finally we propose a mathematical programming formulation for this problem. The (binary) variables of this formulation are the x_{ij} and the $y_{ij,k\ell}$.

- x_{ij} equals 1 if and only if the vertex (i, j) belongs to S.
- $y_{ij,k\ell}$ equals 1 if and only if (i, j) and (k, ℓ) belong to S and there is a conflict between those two vertices.

The label placement problem (in the version where one minimizes the cost of the overlap) can thus be formulated as follows.

$$\text{minimize } \sum c_{ij,k\ell}\, y_{ij,k\ell}$$

$$\text{such that } y_{ij,k\ell} \geq x_{ij} + x_{k\ell} - 1 \text{ for each edge}$$

$$\sum_j x_{ij} = 1 \text{ for each } i$$

$$x_{ij} \in \{0, 1\}, y_{ij,k\ell} \in \{0, 1\}$$

Note that there is an extensive literature on the map labeling problem, and one may find an online bibliography on this topic on the site [1].

3. Features of the GIRO problem

We now discuss the features specific to the problem faced by GIRO. As noted above, the user may modify interactively the solution produced by the routing software and select the labels he wants displayed. A crucial point is that the *display module takes some time to compute an alternative (i.e., non preferred) position for a given label*; for instance, in order to do so it has to take into account the conflicts between labels and entities. This means that producing the set P_i (for a given entity i) is costly, and P_i may not be known beforehand. This feature is the most important difference between the classical label placement problem and the problem faced by GIRO. The challenge is then to design a label placement algorithm that finds a good placement while sending as few requests as possible to the display module.

To make the context more precise, here is an example of the interaction between the display module and the placement algorithm.

- The user selects the labels he wants displayed (he may have modified some entities beforehand).
- The display module gathers the requested data.
- The algorithm obtains label priorities and tries to compute a good placement.
- If necessary, the algorithm requests more information from the display module, i.e., information on overlaps in the current solution or on alternative positions.
- The algorithm produces a placement for the display module.

4. Outline of the proposed solution

The mathematical programming formulation given above can be viewed as a generalization of the maximum weight stable set problem, which is of course NP-complete. In practice, this means that a problem with a few hundred labels cannot be solved in a reasonable amount of time. Actually the problem must be solved in real time, which forces us to consider small conflict graphs only. The team members have thus proposed to decompose the problem along "geographical lines", i.e., to decompose the map into relatively small areas, so that few requests are made to the display module and the conflict graph is of moderate size.

The first step is thus to identify the basic *blocks* of the map. Indeed, the map contains streets, rivers and other features that can represent block boundaries. We may consider the map as a fine grid, in the sense that it consists of small blocks (city blocks, for instance). Some blocks may contain too few entities to be of interest, and in this case we may have to create larger blocks by merging two or more blocks. This can be achieved by a decomposition algorithm or a clustering algorithm. In this fashion we obtain a coarse grid in which every block contains enough entities to represent an interesting subproblem (but not too many because the subproblem must be solved quickly). After the construction of the coarse grid, each subproblem is solved independently by assuming that t_1 positions are known for each entity (where t_1 is a parameter). Then the global placement obtained by juxtaposing the placements for the individual blocks is improved; intuitively the improvements will be needed only for entities close to block boundaries.

We now present a high-level view of the proposed method for placing the labels.

1. Construct a coarse grid whose subproblems can be solved quickly.
2. For each block, ask the display module for t_1 positions for each label whose entity lies within the block.
3. For each block, compute a placement of the labels for that block (using an exact algorithm or a heuristic followed by a post-optimization procedure).
4. Identify the labels that are poorly placed.
5. For each poorly placed label, ask the display module for the next t_2 positions and compute a better placement of the poorly placed labels.

The sub-algorithm consisting of the last two steps can be iterated.

Before giving more details on the steps of this method, we illustrate the construction of a coarse grid. Figure 2 displays a portion of the Montreal city map, with black dots representing features that may be labelled. Note that an ordinary city block does not contain many entities, and that an element of the coarse grid will in general consist of several city blocks.

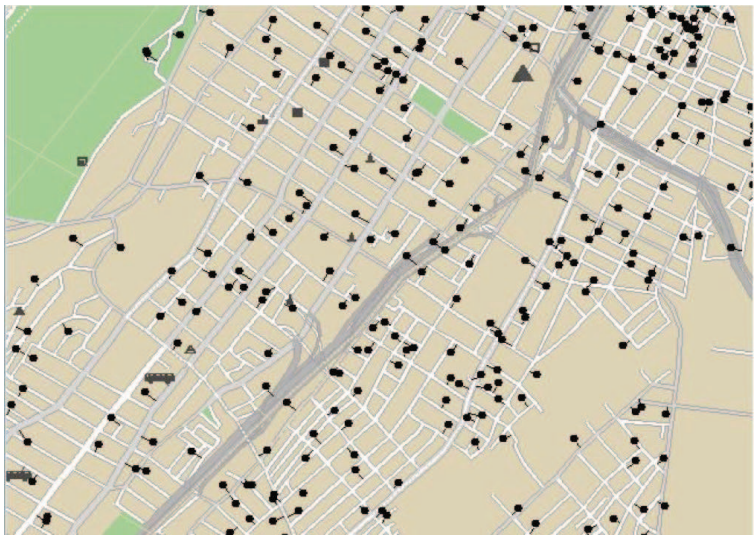


Figure 2.

To decompose the map into smaller regions, we first use the natural boundaries contained in it. In Figure 3 the natural boundaries are displayed in red.

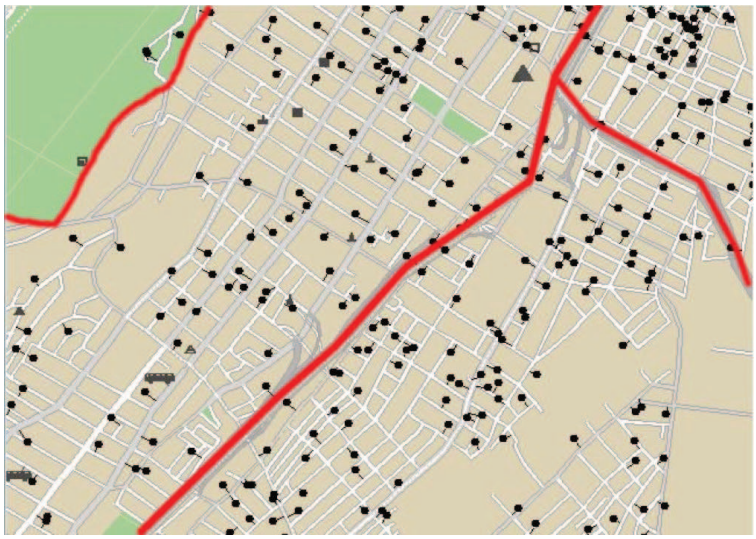


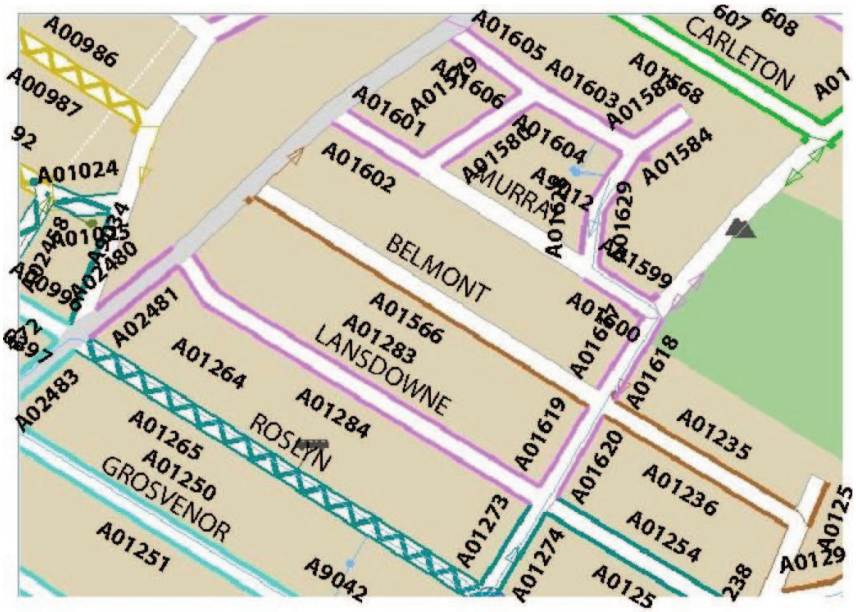
Figure 3.

The regions determined by the natural boundaries are subdivided until the number of entities within each subregion is small enough for the subproblem to be solved quickly. For instance, in Figure 4, at least two regions contains too many entities and must be subdivided; the result is displayed in Figure 5.

Figure 4.

Figure 5.

Now we illustrate the label placement algorithm and compare its results with those produced by the current GIRO software. First, in Figure 6, we present the solution produced by this software for a relatively small area in Montreal.



A solution provided by the software at GIRO

Figure 6.

Figure 7 illustrates the partitioning of this map into relatively large blocks (i.e., the construction of the coarse grid).

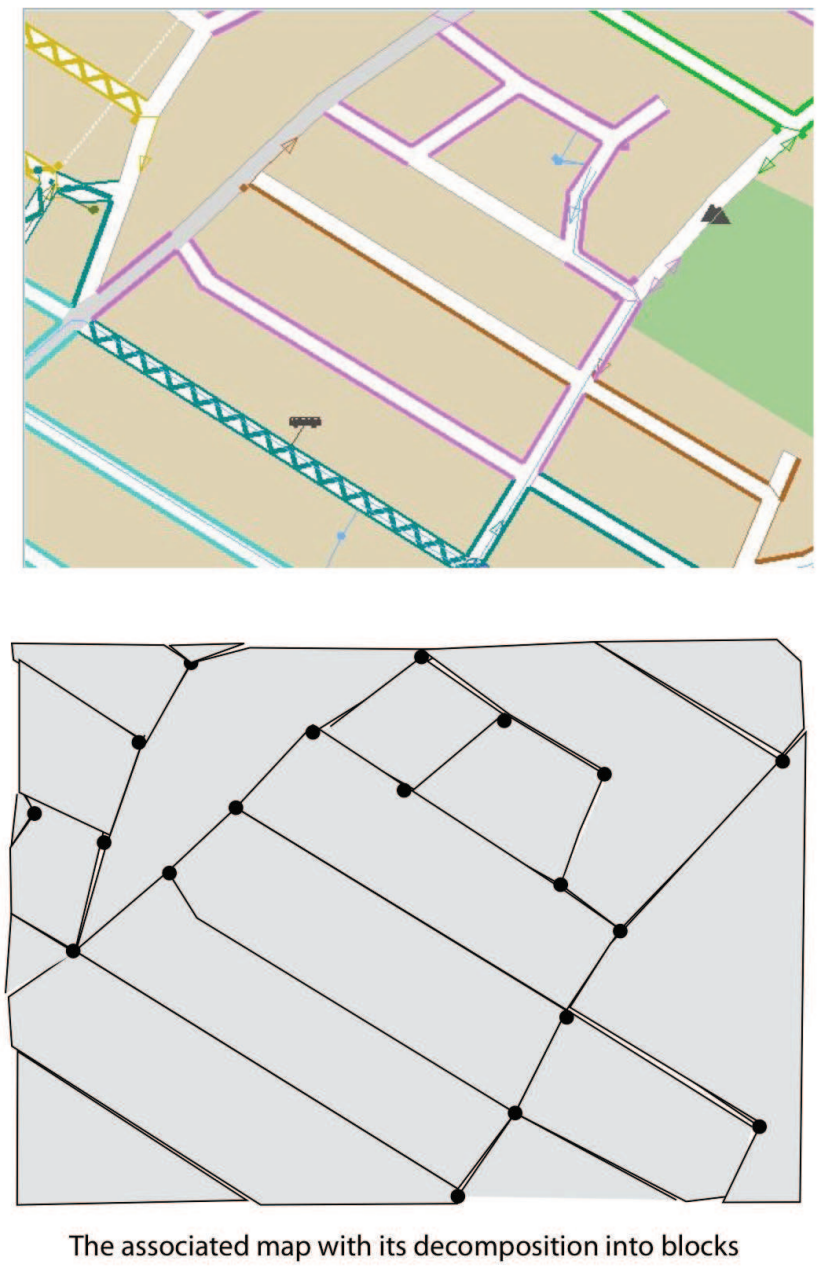


Figure 7.

Figure 8 illustrates the coarse grid with the preferred positions for the labels.

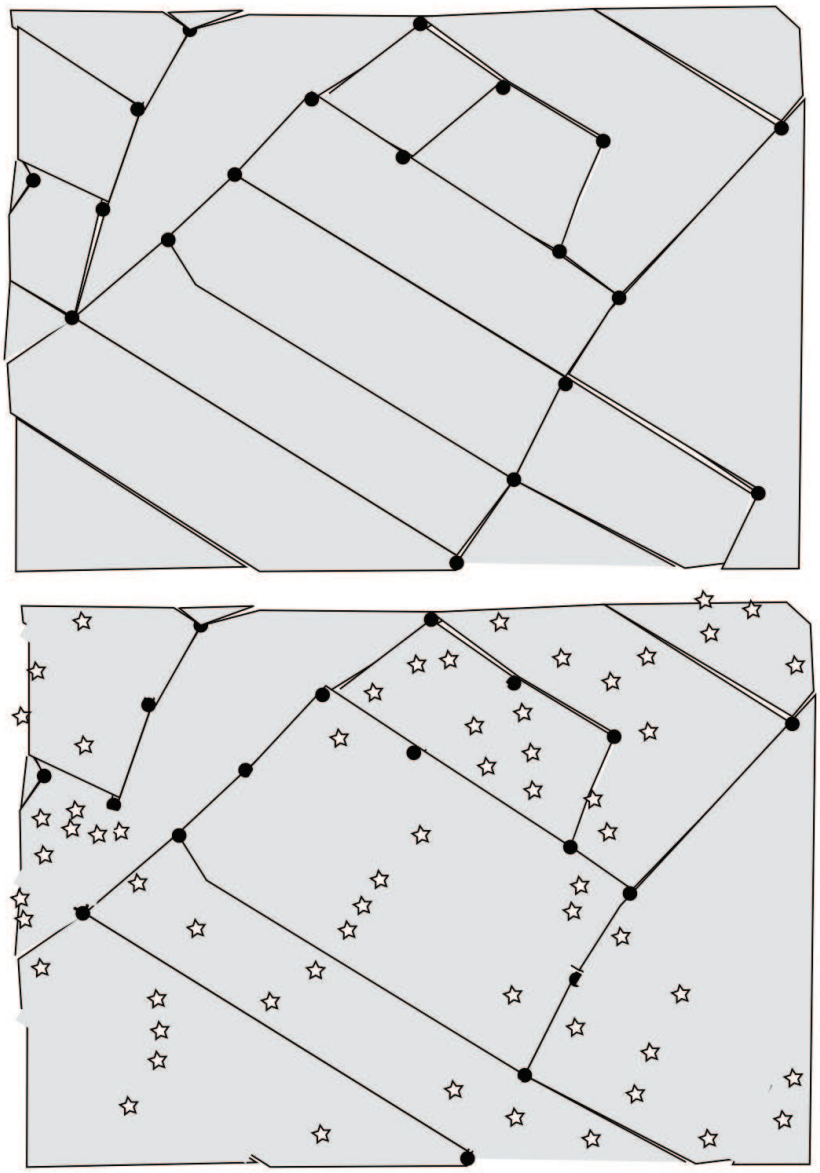


Figure 8.

Figure 9 displays the solution obtained after optimizing each block independently.

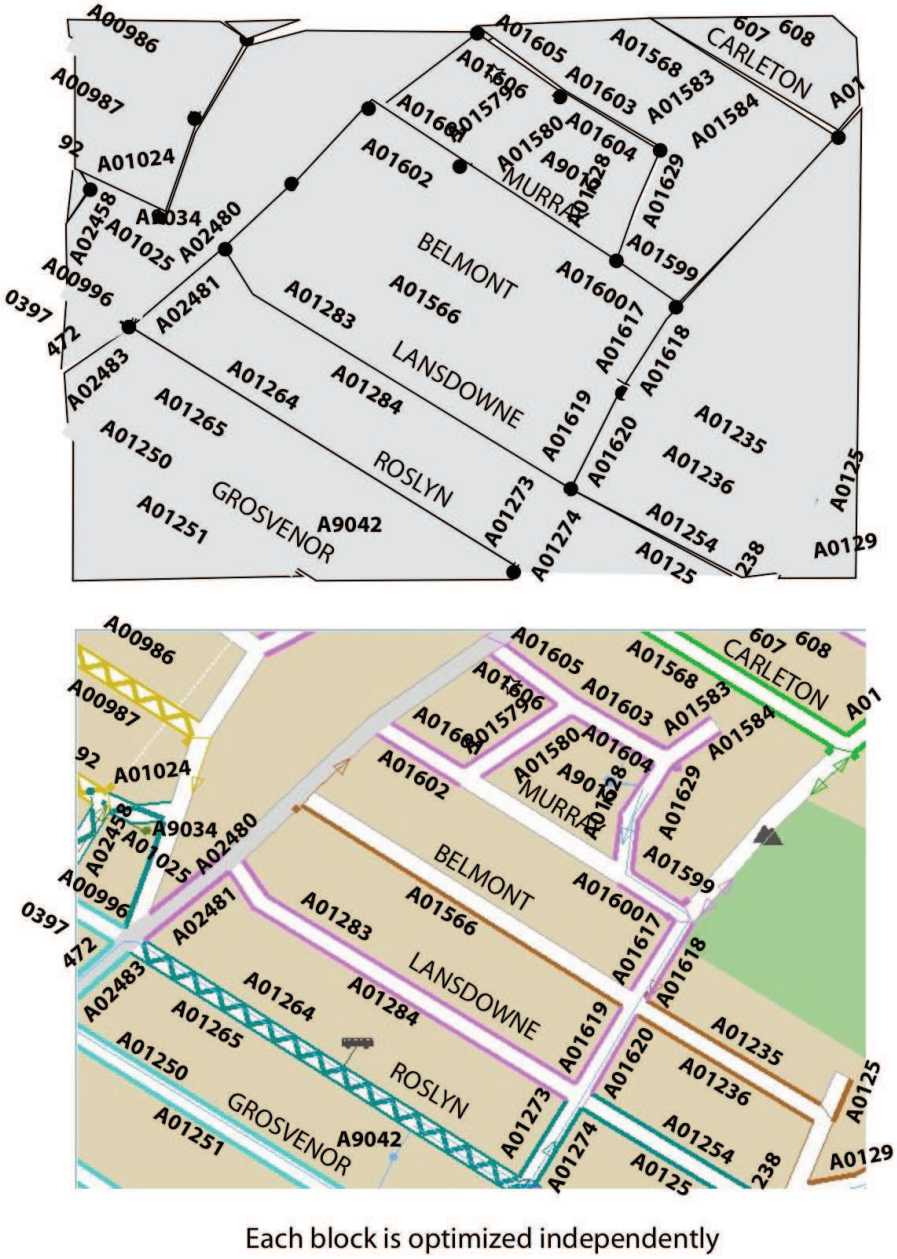


Figure 9.

Figure 10 displays the result of the post-optimization procedure, which attempts to improve the solution by modifying the placement "along the boundary" of adjacent blocks.

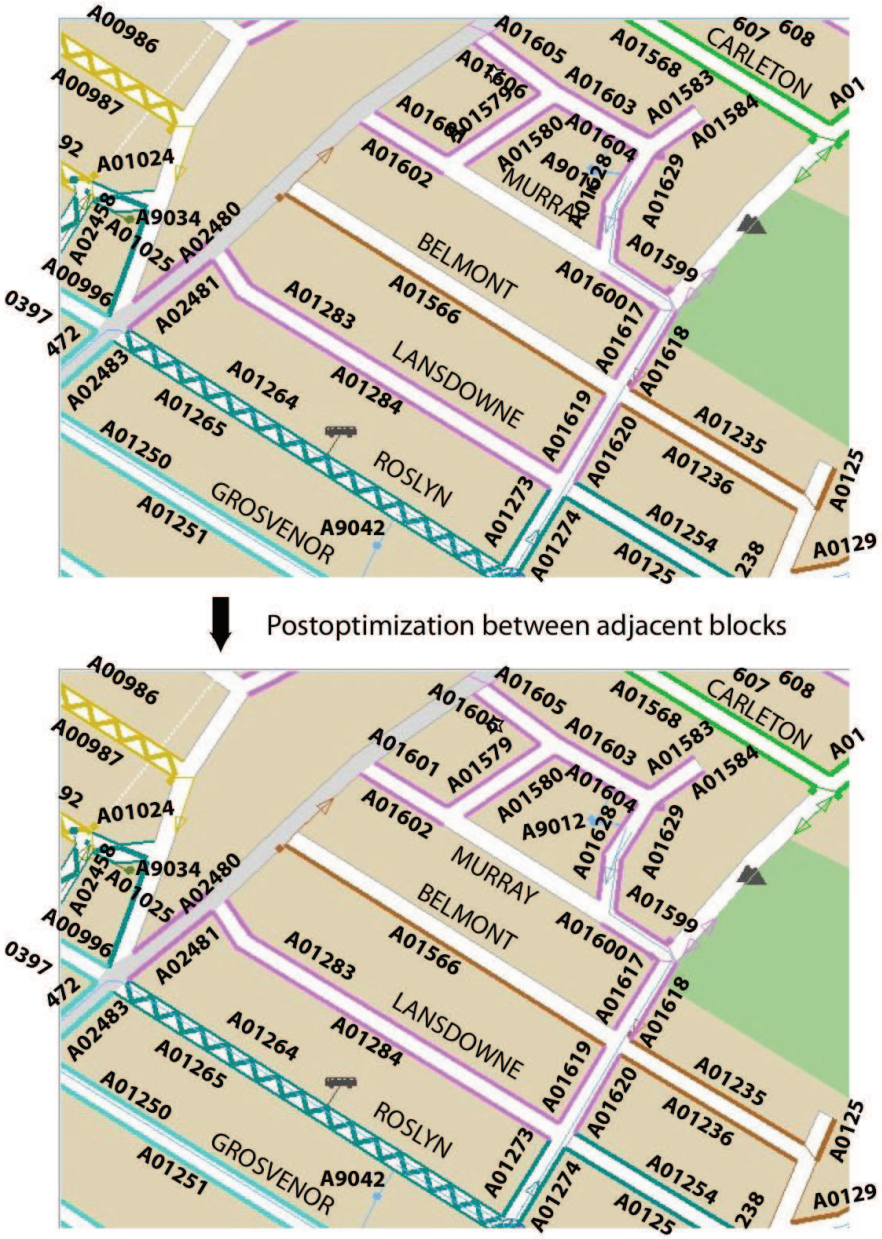


Figure 10.

Finally Figure 11 displays the solution produced by the current GIRO software and that produced by our algorithm.

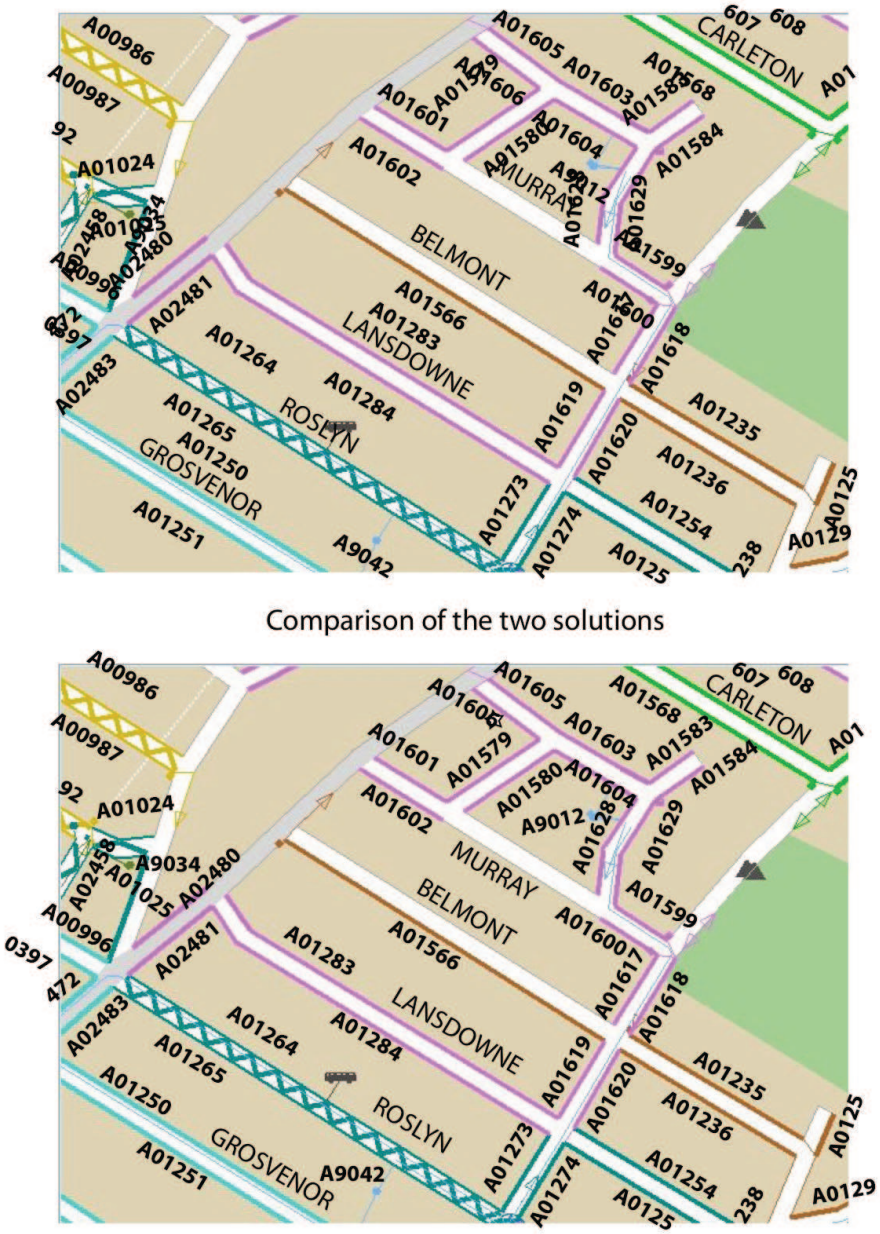


Figure 11.

During the workshop, participants suggested several (exact or heuristic) algorithms to implement the third step of the above method. Here is a list of these suggested algorithms.

- The Tabu Search Heuristic proposed by Yamamoto et al. [4]
- The GRASP Algorithm ("Greedy Randomized Adaptive Search Procedure") proposed by Cravo et al. [2]
- An integer programming algorithm (that of CPLEX, for instance)
- A dual ascent procedure adapted from Erlenkotter's procedure for the plant location problem (see [3])

There were also several suggestions for the fourth step of the method (i.e., identifying poorly placed labels). Within a given block, let S denote the assignment of positions to labels within the block, i.e., the set

$$\{(i, j) \mid i \text{ is a label and } j \text{ its position}\}.$$

Recall that $c_{ij,k\ell}$ denotes the cost of edge $\{ij, k\ell\}$. Here are some criteria for "selecting" poorly placed labels.

- Choose a vertex (i, j) maximizing $\sum_{(k,\ell)\in S} c_{ij,k\ell}$ and request more positions from the display module.
- Choose a subset S' of S such that for each (i, j) in S, $\sum_{(k,\ell)\in S} c_{ij,k\ell}$ is at least a certain threshold; then request more positions for each (i, j) in S'.
- Find a maximum cardinality stable set in S and request more positions for the vertices in S that do not belong to this stable set.

5. Implementation of the proposed solution at GIRO

The GIRO graphics team could not devote enough time (at least this year) to the implementation of all the ideas put forward by the workshop participants. The workshop, however, was very useful, in the sense that the proposed solution addressed the problem faced by GIRO; to our knowledge this problem has not been addressed in the literature. The graphics team implemented a grid based on spatial coordinates (instead of city blocks). In each of the cells the label placement problem is solved in a greedy fashion by an algorithm similar to MIN (a greedy algorithm for finding a large stable set in a given graph). A post-optimization procedure is then applied within each cell, but not between cells. Finally, conflicts are "solved" by removing some labels among those that overlap; the label priorities are taken into account to decide which labels are removed.

References

1. http://i11www.iti.uni-karlsruhe.de/map-labeling/bibliography/
2. Gildásio Lecchi Cravo, Glaydston Mattos Ribeiro, and Luiz Antonio Nogueira Lorena. A greedy randomized adaptive search procedure for the point-feature cartographic label placement. Comput. Geosci., 34(4):373-386, 2007.

3. Donald Erlenkotter. A Dual-Based Procedure for Uncapacitated Facility Location. Operations Research, 26(6):992-1009, 1978.
4. Missae Yamamoto, Gilberto Camâra, and Luiz Antonio Nogueira Lorena. Tabu search heuristic for point-feature cartographic label placement. GeoInformatica, 6(1):77-90, 2002.

Solving a Combined Routing and Scheduling Problem in Forestry

Louis-Martin Rousseau
Coordinator, École Polytechnique de Montréal

Joseph Nader
Industrial representative, FPInnovations, Feric

Jean-François Audy
Main report writer, Université Laval

Nizar El Hachemi
Report writer, École Polytechnique de Montréal

Bernard Gendron
Professor, Université de Montréal

Paul Khuong
Student, Université de Montréal

Laurent Michel
Professor, University of Connecticut

Greg Rix
Student, École Polytechnique de Montréal

David Titley-Peloquin
Student, McGill University

1. Introduction

Transportation planning in forestry involves many decisions that are commonly managed according to four temporal horizons: strategic (up to 5 years), tactical (half a year to 5 years), operational (1 to 180 days), and real-time (less than one day). We refer to [15], [7], and [4] for an exhaustive survey of these issues. In the industrial problem proposed by Feric, three main decisions occurring at the operational level must be addressed: i) allocation, ii) routing, and iii) scheduling. We describe each of them, along with its explicit (i.e., obligatory) objective and its implicit (i.e., desirable) objective.

- Allocation decisions are answers to the following questions: "Which forests will supply which mills?" and "How much of each product type will be supplied?". Each mill has a demand profile (expressed as the number of full truck loads by product type) that must be fulfilled by the end of the week. Meeting the entire demand of each mill is the explicit objective of the industrial problem. A set of forests is available at the beginning of the week in order to meet the demand (with no additional volume from harvesting operations during the week). Each forest has a supply profile (also expressed as the number of full truck loads by product type). A mill (resp. forest) may have a demand (resp. supply) consisting of more than one product type and no product substitution is allowed.

- Routing involves choosing a set of circuits in order to satisfy the entire demand of the mills. A circuit is a sequence of routes corresponding to one truck, each route starting after the

previous one ends (note that a truck can be on duty 24 hours per day). The first route starts at the beginning of the week, while the last one must be completed before the end of the week. A route is defined as a sequence of trips, where each trip visits two sites (an origin and a destination). One specifies the product type (in a volume corresponding to a full truck load) to be picked up (at an origin) or delivered (at a destination). Each route represents the working shift of a driver, while a circuit represents the utilization of a truck over the week.

The explicit objective of the industrial problem is to minimize the total distance corresponding to dead trips (i.e., trips with no product loaded on the truck). When only trips with a full truck load are considered (as in the proposed industrial problem), such an objective is equivalent to the minimization of the total distance covered in the routes of all circuits. The duration of any route cannot exceed 13 hours because of trucking regulations. Also it is not appropriate, in the "real world", to assign a duty consisting of few consecutive hours; hence a route must last at least 8 hours.

Typically truck drivers use their own cars to travel from their home to a mill before their first route and from the mill to their home after their last circuit. Therefore each route must begin and end at the same mill and this mill becomes the home base of the truck for the week (i.e., all routes in a circuit start and end at the same mill). The implicit objective of the industrial problem is to maximize the use of any given truck over the whole week in order to transport the products with the minimum number of trucks, but also to provide the drivers, inasmuch as possible, with circuits extending on the whole week (i.e., providing them with full-time jobs). Hence the minimum (resp. maximum) duration of each circuit is set to 96 hours or 4 days (resp. 120 hours or 5 days). In the proposed industrial problem the number of trucks at a given mill is unknown, but as we have just mentioned, the implicit objective is to use as few trucks as possible.

- Scheduling consists of assigning a start time and an end time to each activity in every circuit, while taking into account the key resources, namely, the loaders. There are four activities that define a trip: the travel from the mill (resp. forest) to the forest (resp. mill) and the loading (resp. unloading) at the forest (resp. mill). Thus a route is a loop in which the following activities are repeated: travel from a mill to a forest, loading at the forest, travel from the forest to the mill, and unloading at the mill. In the industrial problem the explicit objective is to minimize the total waiting time of the trucks, which depends upon the availability of resources. Since there is no crane on a truck, the latter requires a loader when picking up products or delivering them. Each site has at least one loader and the loading or unloading activity for one truck takes 0.6 hour. Trucks may have to wait if there are more trucks than loaders on a given site at a given time.

  Figure 1 shows an example with two circuits in which a truck is waiting for a loader at a forest (F3). The waiting time of the truck is added to the schedule of this circuit. Loaders are assigned to a site for the whole week, and between the loading of two trucks, they are not allowed to move to a busy site. In the industrial problem, the number of loaders per site is unknown. The implicit objective is to use as few loaders as possible, while assigning the highest priority to the explicit objective, i.e., the objective of minimizing the total waiting time of the trucks.

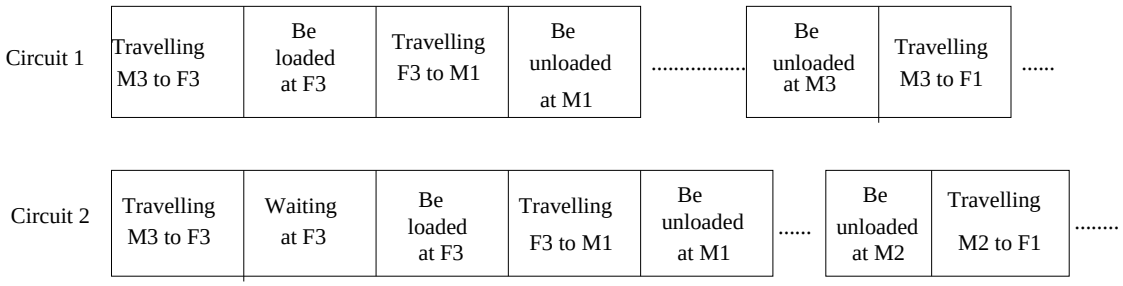


Figure 1. Waiting time at forest 3 in circuit 2

2. Literature Review

No planning method addressing the industrial problem was found in the literature. Several articles, however, address at least two of the subproblems (decisions) and several of the constraints, e.g., [11], [16], [13], [14], [10], [12], [5], [6], [8], and [1]. The planning method used by the company proposing the industrial problem is called "MaxTour"([9]); it is a tool that supports routing decisions. Unfortunately the solution provided by MaxTour does not address all the constraints and decisions mentioned above (see Section 4). MaxTour provides, however, so-called backhaul routes. The use of backhaul routes enables one to create a route combining two or more trips, thus reducing the total distance corresponding to dead trips (if one compares this route to the solution consisting of consecutive round trips). We refer the reader to [2] for a review of backhauling in forestry.

Based on an adaptation of the well-known savings heuristic of [3], MaxTour computes a set of potential backhaul routes by combining trips of full truck load of one product type. All the trips (and the number of times they must be repeated, denoted by NbOfTime) must be specified before MaxTour is used, since the method does not address allocation decisions. This set of potential backhaul routes is then sorted in order of decreasing savings. One then obtains the number of times each backhaul route must be carried out by replicating the first backhaul route as many times as possible (i.e., a number of times equal to the minimum value of NbOfTime for all trips included in the backhaul route), then replicating the second backhaul route, and so on (note that the value of NbOfTime must be decreased after each step). After exhausting the backhaul route, one makes each remaining trip into a round trip. About ten case studies with MaxTour have been conducted within Canadian companies and potential savings of 2% to 7% have been reported ([12]).

3. Solution Methodology

In this section we develop a solution based on a three-phase approach that addresses, in turn, the generation of routes for the trucks, the selection of a subset of routes to meet the needs of the mill, and finally the scheduling of the activities.

3.1. Generation of the routes

The first phase consists of an enumerative method using constraint programming, which allows the generation of a large set of feasible routes visiting a site more than once. This method contrasts

with the MaxTour method, in which a site can be visited only once except for the last site of the route, which must be identical to the first one. The method we used has the following properties:

- it generates any route lasting between 8 and 12 hours;
- it considers forest → mill and mill → forest arcs only;
- it repeats an arc no more times than required by the total availability (resp. demand) of the commodity at the forest (resp. mill);
- it breaks symmetries by forcing the lowest-ordered mill at the beginning of the route and using other rules; and
- it generates any route lasting between 8 and 11 hours.

The last property overrides the first one. It introduces more flexibility during the scheduling phase in order to respect the constraint on the 13-hour maximum duration for a driver's shift. Specifically, the reduction of one hour allows at least two hours (instead of one) of waiting time on a generated route. Accordingly, moving from simply requiring routes of a certain length to enforcing all the requirements reduces the set of generated routes from over 3 million to a few thousands.

3.2. Selection of the routes in each circuit

In the following we describe a mixed integer model that selects, by truck and by home base, a set of routes satisfying the demand and supply constraints. $Mills$, $Routes$ and $Trucks$ denote, respectively, the sets of mills, routes, and trucks. Also $time_r$ denotes the duration of route r, D_{mp} the weekly demand of product p associated with wood mill m, S_{fp} the weekly stock of product p at forest site f, and Q_{rlp} the number of shipments of product p from forest f (resp. to mill m) in route r; note that $l = f$ if the location is a forest and $l = m$ if it is a mill. Finally X_{brt} denotes the number of routes r assigned to truck t based at mill b and Y_{brt} is a binary variable equal to 1 if and only if truck t based at mill b is used in the plan.

$$\text{Min} \sum_{b\in Mills} \sum_{r\in Routes} \sum_{t\in Trucks} time_r X_{brt}$$

subject to

$$\sum_{b\in Mills} \sum_{r\in Routes} \sum_{t\in Trucks} X_{brt} Q_{rmp} \geq D_{mp}, \quad \forall m \in Mills,\ \forall p \in Products \tag{1}$$

$$\sum_{b\in Mills} \sum_{r\in Routes} \sum_{t\in Trucks} X_{brt} Q_{rfp} \leq S_{fp}, \quad \forall f \in Forests,\ \forall p \in Products \tag{2}$$

$$\sum_{r\in Routes} time_r X_{brt} \geq 96\, Y_{bt}, \quad \forall b \in Mills,\ \forall t \in Trucks \tag{3}$$

$$\sum_{r\in Routes} time_r X_{brt} \leq 120\, Y_{bt}, \quad \forall b \in Mills,\ \forall t \in Trucks \tag{4}$$

Constraint (1) states that the supply of each product at each mill must satisfy the demand in each period, whereas constraint (2) states that the supply of each product at each forest site must be respected. The last two constraints, (3) and (4), state that the only allowable circuits (i.e. sets

of routes for specific trucks and home bases) are those in which a "truck duty" ends after the fourth day of the week and before the fifth day.

Note that one could consider a simpler model that does not explicitly account for the individual trucks but aggregates them into fleets. This means that in the last two constraints of the previous model, instead of verifying that the total duration of the routes for each truck lies between the lower and upper bounds, we check whether the average total duration of the selected set of routes for a given combination of truck and home base lies between the lower and upper bounds. The simpler model is a relaxation whose solution may include fractional numbers of trucks and therefore be infeasible in the scheduling phase. Unfortunately this model does not yield any solution within a reasonable amount of time and we discarded it in favour of a two-step approach. The idea behind the decomposition is to reduce the set of routes (shifts) included in the model and obtain a reduced set covering problem. The original model is solved with the reduced set of routes.

$$\text{Min} \sum_{r \in Routes} time_r X_r$$

subject to

$$\sum_{r \in Routes} X_r Q_{rmp} \geq D_{mp}, \quad \forall m \in Mills, \ \forall p \in Products \tag{5}$$

$$\sum_{r \in Routes} X_r Q_{rfp} \leq S_{fp}, \quad \forall f \in Forests, \ \forall p \in Products \tag{6}$$

In the above model X_r represents the number of routes r selected to satisfy the demand and supply constraints while minimizing the total working time. In the second step, we solve the mixed integer model but only with the routes selected by the set covering problem.

3.3. Scheduling of the Circuits

This section presents a brief description of the scheduling model used in this last phase. The scheduling part produces the starting and ending times of each activity. It takes into account the availability of resources (loaders), which was not considered in the previous models. Three different components determine the scheduling model: jobs, activities, and resources. Here are the constraints that must be respected.

- Each route in a circuit is a job.
- Each trip in a route has four activities with strict precedence (i.e., "travelling from a mill to a forest", "being loaded at a forest", "travelling from a forest to a mill", "being unloaded at a mill").
- Each location (mill or forest site) has loaders, considered as discrete resources. Trucks, however, are modelled as alternative resources. Each activity (loading or unloading) can be processed by a single resource (a loader), and there is no preemption, meaning that no activity can be interrupted during processing.

The model is solved by a constraint programming approach, which consists of assigning alternatively trucks to routes, ranking all unary resources, and using the procedure *setTimes* to schedule activities as soon as possible. The solution space is explored through a depth-first search (DFS) strategy.

Figure 2. Relative geographic distribution of forests and mills

Products		MAPW	MHPW	SFGP	SFKP	SFPW	SFSL	SFSW	WPSL
Mills	DDM							115	
	GLT					10	600		
	R&S		5	5		80			27
	SXCP	2	5	5	25	25	5		
Forests	2071026						21		
	2071179						1		
	2072584								
	2072586								
	2073627					25			
	2073661								
	3070606								
	3070607						126		
	3070633								9
	3070787						5		
	3070954	2	10				18		9
	3071001						5		
	3071006						6		
	3072315						5		
	FH00577						19	115	
	NACKY			41	59	95	384		
	SH00379								9
	SXCPY						22		

Figure 3. Supply and demand in the instance

4. Results

The data for the industrial problem to be solved were supplied by FPInnovations-FERIC, and were related to an instance provided by an eastern Canadian forestry company. This instance includes 4 mills, 14 forests, and 8 product types. Figure 2 displays the relative geographic distribution of the forests and mills in the Canadian province of New Brunswick.

Mills may receive supplies from more than one forest only in the case of a few product types. Therefore few allocation decisions have to be made (see the columns of product types SFPW and SFSL in Figure 3). The other allocation decisions are straightforward. For each product type the supply is at least equal to the demand. The first phase model takes less than 5 seconds to generate about 8000 routes. The second phase takes less than one minute to select a subset of 385 routes in 34 circuits. The third phase takes less than 4 minutes to provide a first feasible solution. The solution, however, does not satisfy 10 of the demands. Further work will be required to fix this problem.

Only the potential backhaul routes designed by MaxTour for the instance in question were provided. Even if this solution is a relaxation (i.e., the backhaul routes provided by MaxTour do not respect all the constraints mentioned previously and do not address all the decisions), comparisons can be made by using some indicators: for instance the total time (TT) and the total

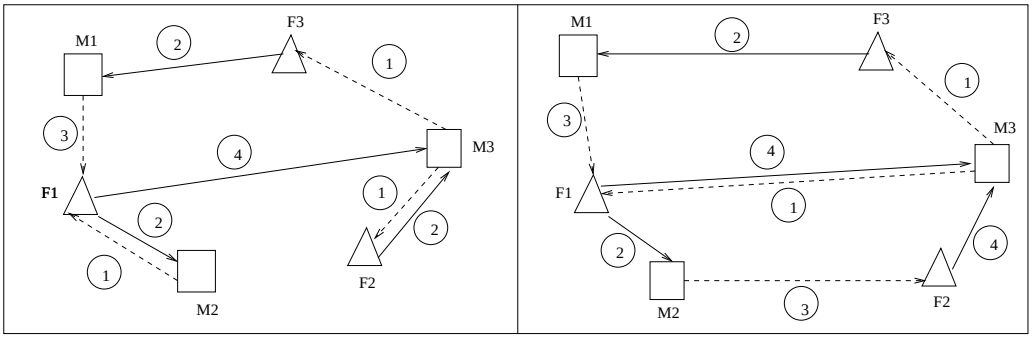


Figure 4. Constraints relaxation in the MaxTour solution

distance of dead trips (TETD). Table I shows that the solution of our second phase (scenario called "Second phase") has 0.1% more time (resp. 2.42% more distance) in TT (resp. TETD) than the MaxTour solution relaxation (scenario called "MaxTour").

Table I. **Routing part solution**

Scenario	Time (h)	Reduction (%)	Empty Time (h)	Reduction (%)
Round trips	3685.8	0	1297.5	0
MaxTour	3636.7	1.33	1230.4	5.17
Second phase	3640.3	1.23	1261.8	2.75

The quality of the MaxTour solution is mainly explained by the fact that it does not take into account two constraints of the second phase: first, the constraint on the design of circuits and routes so that they satisfy the specified durations, and second, the constraint on a truck (which must have the same home base for the whole week). Figure 4 shows the impact of reintroducing these two constraints (i.e., route duration and truck's home base) in a small instance of four trips (i.e. F1-M2; F1-M3, F2-M3 and F3-M1). On the left, the actual empty (broken line) and loaded (complete line) distances computed in the solution of MaxTour are illustrated. On the right, the loaded and empty solution distances satisfying the two constraints are illustrated. The two combined trips (i.e. F1-M3 and F3-M1) are not modified since they satisfy both constraints, while the two round trips previously planned (F1-M2 and F2-M3) are combined to obtain a route whose duration is within the specified range and in which the trip F1-M2 starts and ends at the home base M3.

For comparison purposes, another incomplete solution for the instance can also be computed (see the scenario entitled "Round trips" in Table I). First, all the trips to be carried out are defined by allocating, in a sequential fashion, the minimum available (resp. required) volume for the supply (resp. demand) to the closest forest-mill pair. Second, all the trips are planned in individual round trips. Table I shows that the solution of our second phase (the scenario entitled "Second phase") has 1.23% (resp. 2.75%) less time (resp. distance) in TT (resp. TETD) than the solution of the "Round trips" scenario. The solution of our third phase results in a total waiting time of 111.3 hours for the 34 trucks used. This represents 3.05% of TT, which is quite reasonable. Table II displays the distribution of the 34 trucks used by home base as well as the number of routes for each home base. The distribution of the 34 trucks per home base (mill) follows roughly the distribution

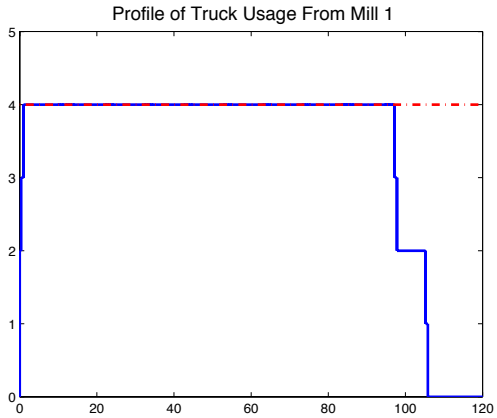


Figure 5. Weekly profile of the working and waiting times of the trucks assigned to the mill DDM

of the 909 demands per mill, e.g., the mill GLT has 67.1% of the total demand and 61.8% (21) of the 34 trucks used.

Table II. **Trucking capacities used by home base**

Home base	Number of Trucks	Number of Routes
1	4	38
2	21	262
3	5	48
4	4	37

Figures 5 and 6 show the weekly profile of the working and waiting times of the trucks assigned to, respectively, the home bases DDM and GLT. Note that the working time is the time during which a truck is travelling or is being loaded or unloaded. We can see that at the busy mill GLT, the routes of all 21 trucks start at the beginning of the week and no truck finishes its circuit before Friday morning, while at the mill DDM, half of the trucks start at the beginning of the week and all the trucks have completed their circuit by Friday morning.

A total of 30 loaders have been used. One loader has been assigned to every site except the mill GLT and the forests 3070607 and NACKY, each of which has 5 loaders. Figures 7 and 8 show the weekly profile of the working and waiting times of, respectively, a site with one loader (DDM) and a site with 5 loaders (GLT). The presence of small waiting times, spread out over the whole week as much as possible, should not be considered as a negative result (as in the weekly profile of the truck). In practice, the volume available at a forest site is typically spread out over several wood piles on each side of one or several roads within a harvesting area. The presence of small waiting times allows the loader (a machine that travels at a slow rate) to move from one pile to another, as required.

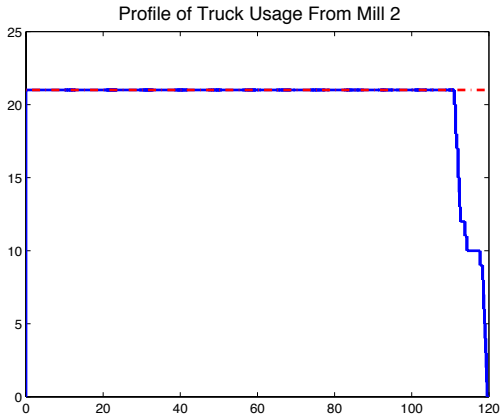


Figure 6. Weekly profile of the working and waiting times of the trucks assigned to the mill GLT

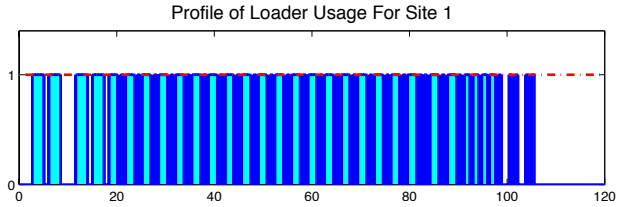


Figure 7. Weekly profile of the working and waiting times of the loader assigned to the mill DDM

5. Discussion and Conclusion

There are several practical aspects that must be evaluated before providing a complete allocation, routing, and scheduling log truck planning method. We discuss a number of them.

- For forests with a low supply (e.g., 2071179, 3070787), it makes sense to allocate a loader in one or few consecutive routes. Thus, the same loader can be used in several "low supply" forests and the total loader cost will be reduced. This will require a detailed schedule for the loader, including transportation time between forests.

- One should take into account the potential capacity constraint(s) in the number of trucks (resp. loaders) per home base (resp. site), for the whole problem and/or by sub-area (i.e., subset of sites).

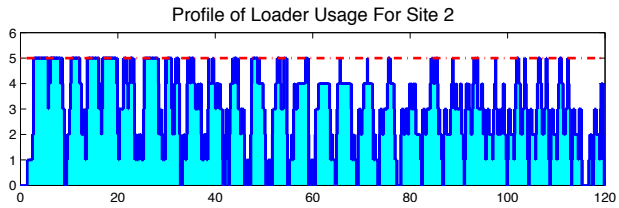


Figure 8. Weekly profile of the working and waiting times of the five loaders assign to the mill GLT

- The method used here does not take into account the fact that, in practice, a loader (especially a loader assigned to a mill) can (un)load a truck outside the schedule, leading to potential additional waiting times in the scheduling. One solution is to add random additional time to the static 0.6 (un)loading time. The same observation can be made for the static travelling distance (e.g., it could vary because of poor weather conditions).

- To add flexibility in the design of routes with a duration within the range of a driver's shift, one might allow a driver to end his route with a loaded truck, the next driver then completing the delivery.

- One could consider the possibility to schedule different types of trucks (with different trailer capacities), including trucks with a crane (not requiring a loader) or trucks with multi-product trailers. By allowing the transportation of different types of products within the same truck, multi-product trailers could reduce empty travelling distances (see e.g. [9]).

- In the method used here, all sites are in operation 24 hours a day during the whole week. In practice, some sites could have different operating hours (e.g., 7 AM to 5 PM from Monday to Thursday and 7 AM to 1 PM on Friday), so that pickup and/or delivery would be restricted to these operating hours.

In this article, we presented the weekly log-truck scheduling problem, in which one seeks to minimize the makespan. To address this problem, we decomposed it into three phases. The first enumerates a set of potential routes, and the second selects the routes in circuits to ensure that supply and demand constraints are satisfied. The final phase consists of scheduling all circuits in order to minimize the makespan. Future research directions include solving the last two phases in one step with additional constraints.

6. Acknowledgements

The report writers wish to thank the other members of the team for their comments on the first version of the article. They wish to thank as well the industrial partner, FPInnovations-FERIC. The team would like to acknowledge the organizing team and sponsors of the Third Montreal Industrial Problem Solving Workshop.

References

1. Audy, J.-F., D'Amours, S., Rousseau, L.-M., 2010. Virtual Transportation Manager - a collaborative decision support system for pickup and delivery problem in forestry. CIRRELT Working Paper.
2. Carlsson, D., Rönnqvist, M., 2007. Backhauling in forest transportation: models, methods, and practical usage. Canadian Journal of Forest Research, 37(12): 2612-2623.
3. Clarke, G., Wright, J.R., 1964. Scheduling of vehicles from a central depot to a number of delivery points. Operations Research 12(4): 568-581.
4. D'Amours, S., Rönnqvist, M., Weintraub, A., 2008. Using operational research for supply chain planning in the forest products industry. Information Systems and Operational Research, 46(4): 265-281.
5. El Hachemi, N., Gendreau, M., Rousseau, L.-M., 2008. Solving a log-truck scheduling problem with constraint programming. In: Perron, L., Trick, M.A.., (Eds.). 5th International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems, Lecture Notes in Computer Science, Vol. 5015. Berlin-Heidelberg: Springer. 293-297.

6. El Hachemi, N., Gendreau, M., Rousseau, L.-M., 2009. A heuristic to solve the weekly log-truck scheduling problem. In: International Conference on Industrial Engineering and Systems Management, 13-15 May 2009, Montreal, Canada.
7. Epstein, R., Karlsson, J., Rönnqvist, M., Weintraub, A., 2007. Forest transportation. In: Weintraub, A., Romero, C., Bjorndal, T., Epstein, R., (Eds.). Handbook on Operations Research in Natural Resources, International Series in Operations Research and Management Science, Vol. 99, Chapter 20. New York: Kluwer Academic Publishers. 391-403.
8. Flisberg, P., Lidén, B., Rönnqvist, M., 2009. A hybrid method based on linear programming and tabu search for routing of logging trucks. Computers and Operations Research, 36(4): 1122-1144.
9. Gingras, C., Cordeau, J.-F., Laporte, G., 2007. Un algorithme de minimisation du transport à vide appliqué à l'industrie forestière. Information Systems and Operational Research, 45(1): 41-47.
10. Gronalt, M., Hirsch, P., 2007. Log-truck scheduling with tabu search strategy. In: Doerner, K.F., Gendreau, M., Greistorfer, P., Gutjahr, W.J., Hartl, R.F., Reimann, M. (Eds.). Metaheuristics - Progress in Complex Systems Optimization. New York: Springer. 65-88.
11. Linnainmaa, S., Savola, J., Jokinen, O., 1995. EPO: A knowledge based system for wood procurement management. In: Aikins, J., Shrobe, H., (Eds.). 7th Conference on Innovative Applications of Artificial Intelligence, 20-23 August 1995, Montreal, Canada. Menlo Park: AAAI Press. 107-113.
12. Marier, P., Audy, J.-F., Gingras, C., D'Amours, S., 2007. Collaborative wood transportation with the Virtual Transportation Manager. In: Blanchet, P., (Eds.). International Scientific Conference on Hardwood Processing, 24-26 September 2007, Quebec, Canada. Quebec: FPInnovations-Forintek. 191-198.
13. Palmgren, M., Rönnqvist, M., Varbrand, P., 2003. A solution approach for log truck scheduling based on composite pricing and branch and bound. International Transactions in Operations Research. 10(5): 433-447.
14. Palmgren, M., Rönnqvist, M., Varbrand, P., 2004. A near-exact method to solve the log-truck scheduling problem. International Transactions in Operations Research. 11(4): 447-464.
15. Rönnqvist, M., 2003. Optimization in forestry. Mathematical Programming, 97(1-2) : 267-284.
16. Weintraub, A., Epstein, R., Morales, R., Seron, J., Traverso, P., 1996. A truck scheduling system improves efficiency in the forest industries. Interfaces, 26(4): 1-12.