

**Comptes rendus du
Premier atelier de résolution de problèmes industriels
de Montréal**

20 au 24 août 2007

Parrainé par MITACS et le réseau rcm_2
Responsable de la rédaction: Odile Marcotte, CRM et UQÀM
Rapport de recherche du CRM no CRM-3276



**Proceedings of the
First Montreal Industrial Problem Solving Workshop**

August 20–24, 2007

Sponsored by MITACS and the rcm_2 network
Editor: Odile Marcotte, CRM and UQÀM
CRM Research Report no. CRM-3276

Contents

| | |
|--|----|
| Remerciements | 3 |
| Acknowledgments | 3 |
| Extracting Autofluorescence from Diffuse Optical Images | 5 |
| Nonlinear Continuous Deformation of an Image Based on a Set of Intersecting Straight Lines | 13 |
| Strategic Planning at Kruger Products | 15 |
| Routes de collecte du lait de la Fédération des producteurs de lait du Québec | 23 |
| Recognized Maritime Picture: Geofeasibility Scores | 31 |

Remarque. Les rapports sont présentés dans la langue dans laquelle ils ont été soumis. Le lecteur trouvera les descriptions de tous les problèmes en français sur le site

<http://crm.math.ca/probindustriels2007/>.

Remark. The reports are presented in the language in which they were submitted. The problem descriptions in English can be found on the site

http://crm.math.ca/probindustriels2007/index_e.shtml.

Remerciements

J'aimerais exprimer ma reconnaissance au réseau MITACS, qui a été le principal parrain du Premier atelier de résolution de problèmes industriels de Montréal. Le rcm_2 a également fourni une subvention qui a permis au CRM d'organiser cet atelier. Je suis aussi très reconnaissante aux personnes suivantes, qui nous ont aidés à trouver des problèmes : Frédéric Lesage, de l'École Polytechnique de Montréal, Frédéric Leblond, de Dartmouth College, Jean-Marc Lina, de l'École de Technologie Supérieure, Arnaud Lina, de Matrox Electronic Systems, Vincent Béchard, de Différence S.E.N.C., Louis-Martin Rousseau, de l'École Polytechnique de Montréal, Elisa Shahbazian, de Lockheed Martin Canada, et Anne Bourlioux, de l'Université 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 Alan Bernardi, directeur général des Laboratoires universitaires Bell, Anne Bourlioux, de l'Université de Montréal, Myriam Bouroche, directrice des Laboratoires universitaires Bell, Michel Gendreau, directeur du CIRRELT, Alexandra Haedrich, représentante de MITACS, Pierre Hansen, de HÉC Montréal, François Lalonde, directeur du CRM, et Roland Malhamé, directeur du GERAD.

Odile Marcotte
Directrice adjointe
Centre de recherches mathématiques

Acknowledgments

I would like to thank the MITACS network, who was the main sponsor of the First Montreal Industrial Problem Solving Workshop. The CRM also received a grant from the ncm_2 network in order to organize the workshop. I am very grateful to the following persons, who helped us find problems for the IPSW: Frédéric Lesage, from the École Polytechnique de Montréal, Frédéric Leblond, from Dartmouth College, Jean-Marc Lina, from the École de Technologie Supérieure, Arnaud Lina, from Matrox Electronic Systems, Vincent Béchard, from Différence S.E.N.C., Louis-Martin Rousseau, from the École Polytechnique de Montréal, Elisa Shahbazian, from Lockheed Martin Canada, and Anne Bourlioux, from the Université 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 Alan Bernardi, director general of Bell University Laboratories, Anne Bourlioux, from the Université de Montréal, Myriam Bouroche, director of Bell University Laboratories, Michel Gendreau, director of CIRRELT, Alexandra Haedrich, MITACS representative, Pierre Hansen, from HÉC Montréal, François Lalonde, director of CRM, and Roland Malhamé, director of GERAD.

Odile Marcotte
Deputy Director
Centre de recherches mathématiques

Extracting Autofluorescence from Diffuse Optical Images

Frédéric Lesage

Coordinator, École Polytechnique de Montréal

Frédéric Leblond

Industrial representative, ART Inc.

Olesya Peshko

Report writer, McMaster University

Shidong Shan

Student, University of British Columbia

1. Introduction

The problem studied by the team was proposed by the company ART Advanced Research Technologies Inc., which “designs, develops, and markets optical imaging products for the medical and pharmaceutical industries” (see the web site <http://www.art.ca/en/art/profile.php>). Also, “ART has developed several products that relate to medical imaging, diagnostics, medical research and drug discovery”. Among these products one finds Optix, “an *in vivo* optical molecular imaging device designed to monitor physiological changes in small animals during preclinical drug studies”. Optix is used worldwide by industry and academic leaders. The students in the team, Olesya Peshko and Shidong Shan, were supposed to investigate the algorithms used in this product and their work led to the improvement of one aspect of the optical data processing carried out by Optix. In the rest of this section we discuss biomedical imaging in general. Section 2 describes the approaches investigated by the team and the results obtained.

The advent of new biomedical imaging techniques has brought significant progresses in how we image humans and animals. In particular, the emergence in the mainstream of new modalities, such as diffuse optical imaging (see [13], [8], [12], [4], [5]), has enabled researchers to image hemodynamics *in vivo* in a non-invasive fashion. The relatively low absorption and low scattering in the 600-1000 nm spectral range allow detection of photons travelling through several centimetres of biological tissue. Coupled with accurate models of light propagation, the NIR (Near Infra-Red) techniques enable imaging of deep tissue with boundary measurements using non-ionizing, low dose radiation. Applications of clinical significance have emerged for this modality; among those, breast cancer detection and neuronal activation measurement for cognitive studies are applications currently being actively pursued. In particular, optical imaging is currently arousing interest in the neurosciences. Absorption of near infra-red light by blood (oxy- and deoxy-haemoglobin), see [13], [8], and [12], can be used to measure functional information during brain activation. For example, using different wavelengths, it is possible to map changes in hemodynamics and oxygenation levels to changes in light intensity during some neuronal task (see [4] and [5]). The advantages of this modality are that one can distinguish between oxygenated and deoxygenated blood, and that NIR techniques are more convenient and lightweight. Its main weaknesses concern spatial resolution and quantification.

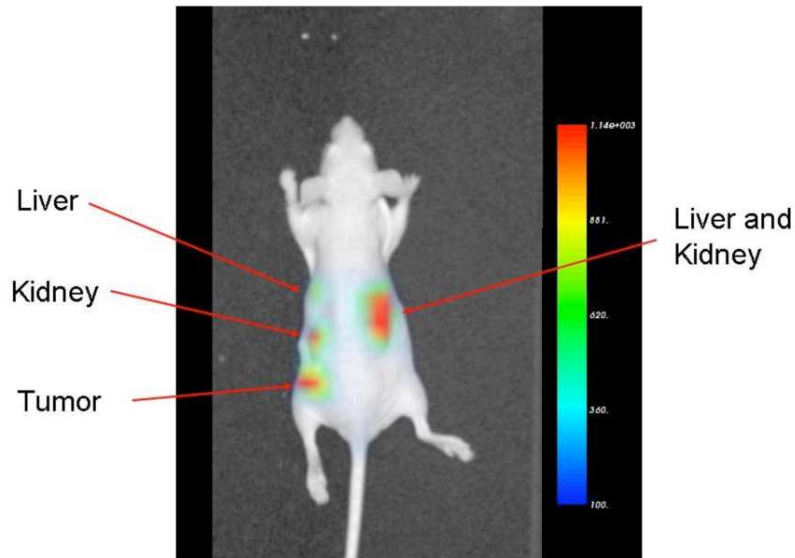


Figure 1. Small animal imaging with fluorescent probes targeting cancer

In parallel, associated with the emergence of endogenous diffuse optical imaging in humans, a new approach is emerging that enables to follow molecular products injected in animals, in vivo, through fluorescence. This new imaging technique (see [6], [9], [14], and [3]), referred to as molecular imaging, holds great promises: whereas one would detect diseases through anatomical changes and anomalies, we can now use new methods to provide a minimally invasive characterization and measurement of biological processes at the cellular and molecular levels of living organisms. Applications abound: early detection of diseases, real-time assessment of therapy effectiveness, and specific and selective delivery of drugs to targeted tissues. The recent emergence of genomic information clears the way for more individualized and specific drugs. Angiogenesis, hypoxia and pH, hyper-metabolism, micro-calcifications, and protein expression are all potential targets for these new techniques. Optical biomedical imaging has, over the last few years, enabled researchers to take significant steps towards this new paradigm. Fluorescent optical imaging, however, is subject to inherent difficulties. Tissues also generate fluorescence, and one needs to determine whether the measurements originate from the molecular compound that was injected or from autofluorescence.

Whether it is necessary to localize structures originating from targeted inclusions in the images or to filter out autofluorescence to improve image quality, the problem consists of segmenting out components of the image whose properties allows us to characterize them as either inclusions or autofluorescence. Figure 1 displays a small animal with fluorescent probes targeting cancer, and Figure 2 a model exhibiting a target and interference generated by simulated autofluorescence. A defining characteristic is that inclusions have a relatively high-contrast, well-defined, almost round shape, while autofluorescent objects are characterized by irregular, lower-contrast shapes. In this workshop, several filtering techniques have been implemented to identify the structures, such as wavelet transform, principal component analysis, and normalized cross-correlation and shape fitting.

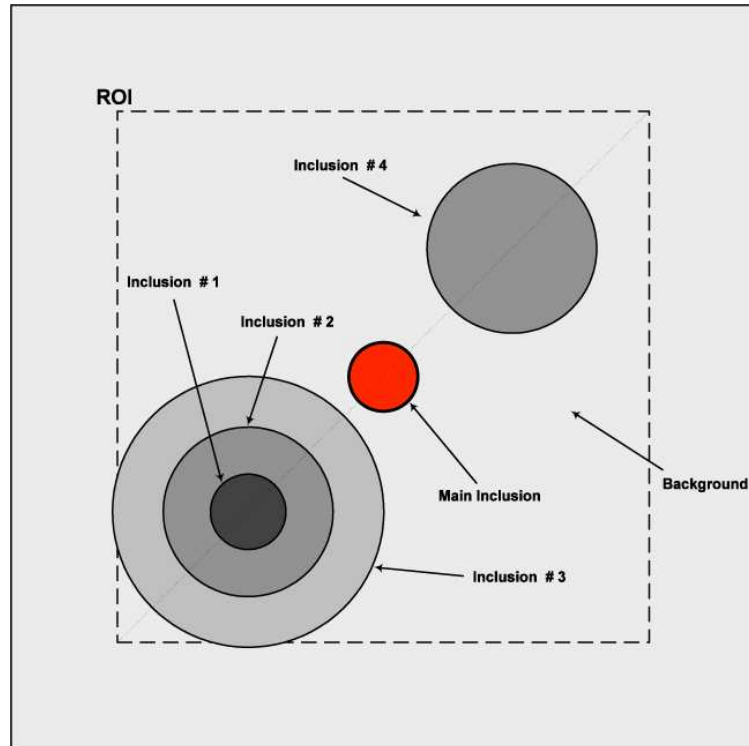


Figure 2. Numerical model exhibiting a target and interference generated by simulated autofluorescence

2. Problem Approaches

2.1. WAVELET TRANSFORM

Method

The general denoising procedure using a wavelet transform involves three steps: decompose, threshold, and reconstruct (see [7] and [10]). Firstly, a wavelet transform and a decomposition level N are chosen, and the wavelet decomposition of the signal at the level N is computed. Secondly, for each level from 1 to N , a threshold is selected and soft or hard thresholding is applied to the detail coefficients. Finally, a wavelet reconstruction is computed using the original approximation coefficients of the level N and the modified detail coefficients of levels from 1 to N . When this technique is applied to an image of fluorescence, the reconstructed image can be considered as the main inclusion from targeted probes after extracting the autofluorescence from tissues.

Results

In this project, the wavelet transform approach was carried out by using the Matlab wavelet toolbox [1]. Specifically, we have utilized the 2D discrete stationary wavelet transform (swt2) with “db1” wavelets to carry out the wavelet decomposition. The reconstructed images were produced by the inverse stationary wavelet transform (iswt2) after soft thresholding.

Figure 3 illustrates a 3-level wavelet decomposition on the synthetic image obtained in Case 1 with Detector 1. Figure 4 shows a comparison of one original image and the reconstructed main component of the image using the inverse wavelet transform. Note that the structure of the main inclusion can be identified more easily in the reconstructed image than in the original image.

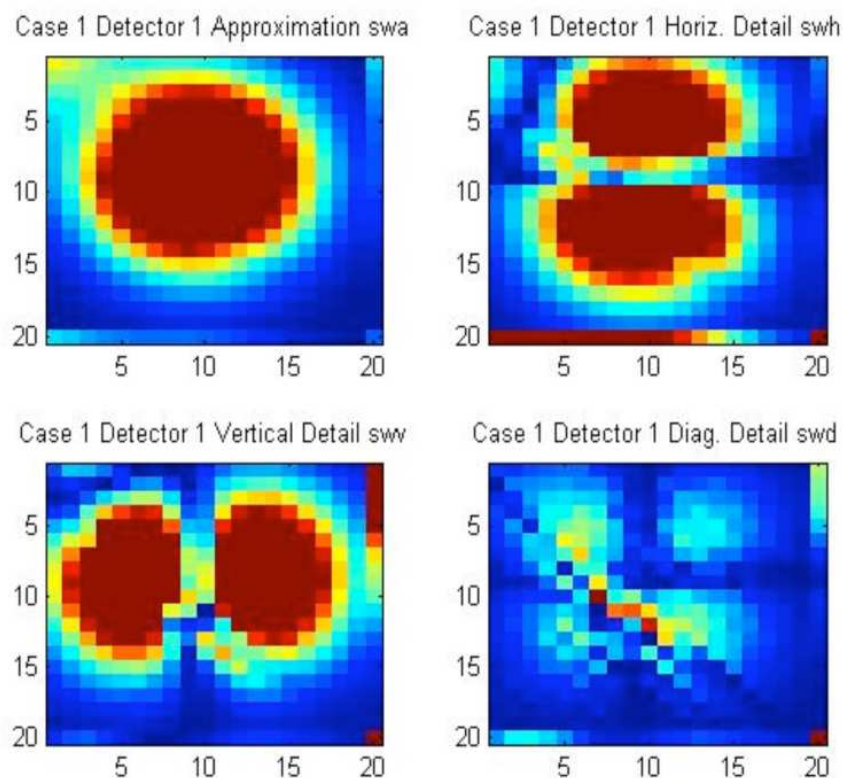


Figure 3. 3-level wavelet decomposition for the synthetic image (case 1 detector 1)

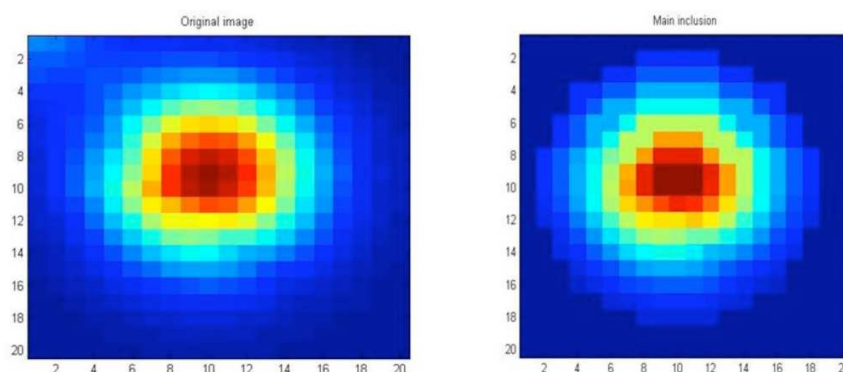


Figure 4. Original and reconstructed image of the main component

Discussion

The well-known wavelet package Wavelab is available at

<http://www-stat.stanford.edu/~wavelab/>

(see [2]). Because of the time limitation, we could use the Matlab wavelet toolbox only. It would be interesting to explore the use of Wavelab for extracting autofluorescence from the diffuse optical images. Wavelet thresholding is an efficient technique to filter additive noise and restore original signals. The quality of the reconstructed signal depends on the choice of threshold and the method of thresholding. The specific thresholding strategies should be selected according to the basic

autofluorescence model. Moreover, different thresholding techniques need to be tested in order to determine which method produces the best results for the diffuse optical images of fluorescence.

2.2. PRINCIPAL COMPONENT ANALYSIS

Method

Another way of segmenting the inclusions and other objects in the image is by using a Principal Component Analysis (PCA). Here, it is carried out by performing a Singular Value Decomposition (SVD) and choosing only the components corresponding to the largest singular values (to reveal the most important structures in the image). The image is represented by a matrix A of size $m \times n$. The SVD of A is a matrix factorization of the form $A = USV^t$. Let $s_1 \geq \dots \geq s_i \geq \dots \geq s_r \geq 0$ be the singular values of A , where $r = \min(m, n)$. S is diagonal with singular values in non-decreasing order and U and V have orthonormal columns; each column of U (resp. V) is a left-singular (resp. right-singular) vector of A . It is well known that the first p ($1 \leq p \leq r$) singular values capture the most important features of A . Thus we define the i th component A_i as US_iV^t , where S_i is a matrix obtained from S by zeroing all the diagonal elements except for s_i .

Results

In our experiments, we have empirically chosen the first three components, because they contain almost all the meaningful information. For some images, considering as few as two components is enough. Figure 5 displays the results of the PCA for a synthetic image. In the original image (see Figure 5 top left), one can see an inclusion with some background fluorescence in the upper left corner. Figures 5 top right, 5 bottom left, and 5 bottom right represent the A_1 , A_2 , and A_3 components, respectively. In this case, the inclusion is almost completely contained in A_1 , while the background fluorescence is in A_2 . Component A_3 contains some negligible details. For this image, the background fluorescence can be removed easily.

Figure 6 top left demonstrates an image of the mouse with an inclusion (black arrow). You can observe a lot of fluorescence from the organs in addition to the target inclusion. Figures 6 top right, 6 bottom left, and 6 bottom right represent the A_1 , A_2 , and A_3 components, respectively. In this case, the number of fluorescent objects in the image is relatively large and separating them out is not as straightforward as in the previous example.

Discussion

If any preliminary knowledge about the objects in the image is available, the information from the three components can be combined in a certain way, i.e. $\alpha A_1 + \beta A_2 + \gamma A_3$, where A_1 , A_2 , and A_3 are the image components, and α , β , and γ the coefficients that reflect the weight of each component. While the success of the PCA segmentation depends on image sophistication, the target inclusion is almost always present in the A_1 component because it has a relatively well-defined shape and shows high-contrast fluorescence. Separating out the components, however, cannot be considered an acceptable solution in many cases, since it often degrades the overall quality of the image. Below we develop a more robust approach, which is capable of separating out objects from the image. Moreover, it does not depend upon the number of the objects and can even handle partially overlapping objects.

2.3. CROSS-CORRELATION AND SHAPE FITTING

Method

In this approach, we try to separate out the elliptical shapes in the image. This is justified by the following assumptions: 1) target inclusions are almost always round, and 2) autofluorescent

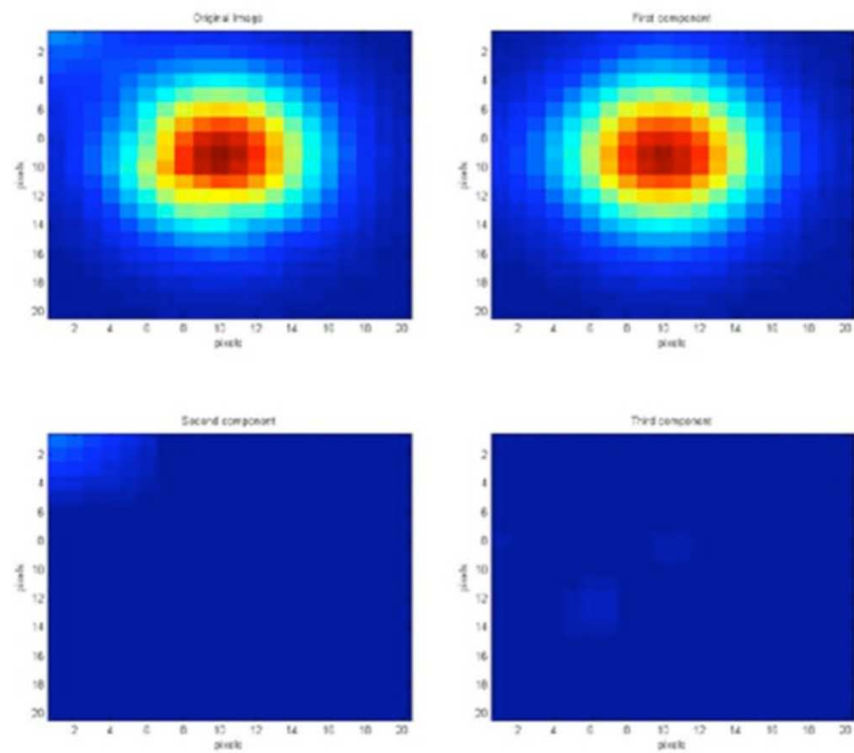


Figure 5. Components of the PCA for the simulated image

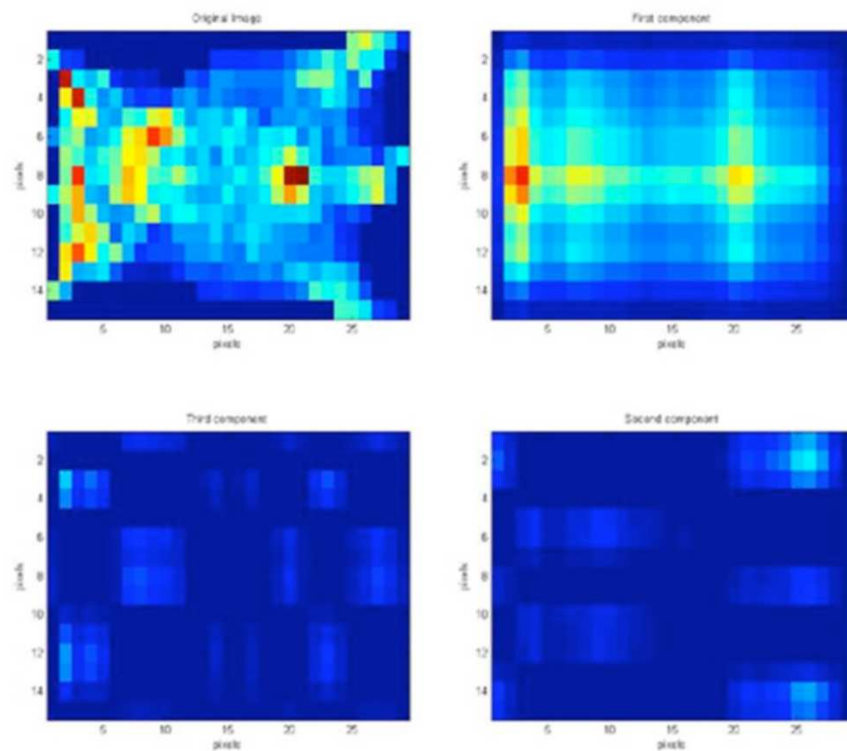


Figure 6. PCA analysis of a real image

objects, though generally of irregular shapes, are low-contrast, and in many cases can be treated as consisting of one or several partially overlapping elliptical shapes. The general idea of the algorithm is to separate out objects from the image one by one, and estimate their shapes to determine what kind of object is extracted at each iteration; the nature of the object will help one decide whether it should be removed from the image or not. As there can be many fluorescent objects, a smaller region of the image is chosen at each iteration. We use a Gaussian model f_p to define an elliptical shape: $f_p(x, y) = v_0 + v_1 e^{-s(x, y)}$, where $s(x, y)$ is defined as $a_1(x - b_x)^2 + 2a_2(x - b_x)(y - b_y) + a_3(y - b_y)^2$, x and y are the coordinates of the pixels in the image, (b_x, b_y) is the centre of the bump, and a_1 , a_2 , and a_3 define the elliptical shape. Note that $a_1 a_3 - a_2^2 > 0$. We choose several highly probable bump sizes and compute the location of the bumps in the image by maximizing normalized cross-correlation of the model and the image data. Then we choose the parameters of the bump that conform to the largest number of template sizes or to the one containing the brightest pixels. These estimated parameters serve as a first guess for refining the optimization procedure.

In order to remove the bump from the image, we fit the image data $\Phi_{x, y}$ to the model $f_p(x, y)$ by solving an optimization problem with the model parameters b_x , b_y , a_1 , a_2 , a_3 , v_0 , and v_1 :

$$\min \|f_p(x, y) - \Phi_{x, y}\|^2.$$

A similar idea, for a different model, is explored in [11], where it is described in more detail.

Results

Figure 7 top left demonstrates a light intensity function from a synthetic image that contains one well defined bump and a background fluorescence caused by another (close) bump. Figure 7 top center shows the fitted model and Figure 7 top right the result of the removal of the detected bump. The remaining bump can be separated from the image in the same fashion. Figure 7 bottom shows the results of a similar procedure carried out on an *in vivo* image of a mouse. At the first iteration a target inclusion was detected (see Figure 7, bottom left and center). Figure 7 bottom right displays the bumps that can be detected at the subsequent iterations of the algorithm.

Discussion

Generally speaking, this algorithm requires a fine tuning of the parameters, based on what is known about imaging data; this is so because the success of the optimization procedure depends on the quality of the first guess. For example, it is useful to have a notion of the possible bump sizes in order to obtain a good fit. The choice of the stopping criteria is another important issue. For instance, the number of iterations may be fixed beforehand in the case where the number of extracted objects is known or can be guessed reasonably well. On the other hand, the algorithm may be stopped if it can be inferred from the image or bump parameters that there are no inclusions or autofluorescent objects left in the image.

References

1. *Matlab wavelet toolbox documentation*.
2. <http://www-stat.stanford.edu/~wavelab>.
3. S. Achilefu, R. Dorshow, J. Bugaj, and R. Rajagopalan, *Novel receptor-targeted fluorescent contrast agents for in-vivo tumor imaging*, *Invest. Radiol.* **35** (2000), 479–485.
4. D. A. Boas, G. Strangman, J. P. Culver, R. D. Hoge, G. Jaszewski, R. A. Poldrack, B. R. Rosen, and J. B. Mandeville, *Can the cerebral metabolic rate of oxygen be estimated with near-infrared spectroscopy?*, *Physics in Medicine and Biology* **48** (2003), 2405–2418.
5. J. P. Culver, A. M. Siegel, J. J. Stott, and D. A. Boas, *Volumetric diffuse optical tomography of brain activity*, *Optics Letters* **28** (2003), no. 21, 2061–2063.

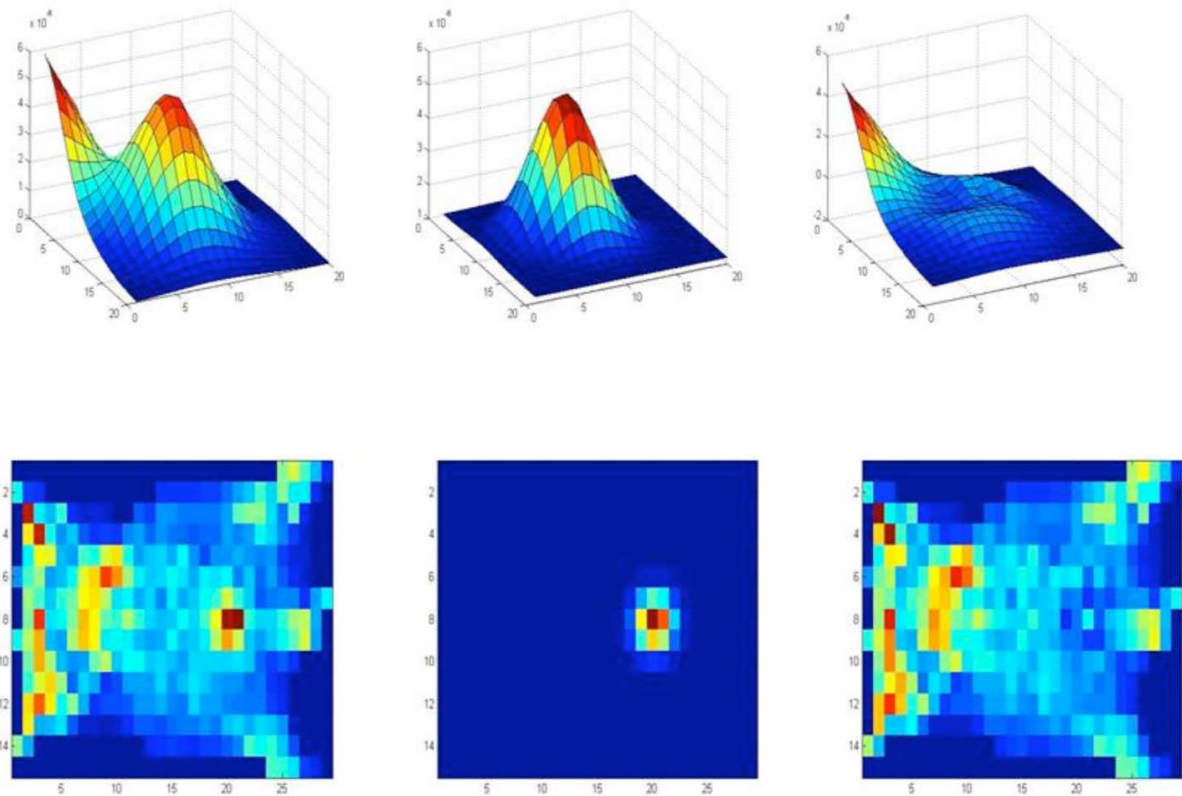


Figure 7. Results of the cross-correlation technique

6. J. V. Frangioni, *In vivo near-infrared fluorescence imaging*, *Curr. Opin. Chem. Biol.* **7** (2003), 626–634.
7. A. Graps, *An introduction to wavelets*, *IEEE* **2** (1995), no. 2.
8. X. Intes and B. Chance, *Non-PET functional imaging techniques: optical*, *Radiologic Clinics of North America* **43** (2005), 221–234.
9. K. Licha, *Contrast agents for optical imaging*, *Topics in Current Chemistry* **222** (2002), 1–29.
10. Stéphane Mallat, *A wavelet tour of signal processing*, 2nd edition, Academic Press, 1999.
11. O. Peshko, C. Anand, and T. Terlaky, *Surface reconstruction from structured-light images for radiation therapy*, *Photonic Applications in Biosensing and Imaging*, Proceedings of SPIE, no. 5969, 2005, pp. 427–436.
12. R. Weissleder and U. Mahmood, *Molecular imaging*, *Radiology* **219** (2001), 316–333.
13. A. Yodh and B. Chance, *Spectroscopy and imaging with diffusing light*, *Phys. Today* **48** (1995), 34–40.
14. G. Zheng, Y. Chen, X. Intes, B. Chance, and J. Glickson, *Contrast-enhanced NIR optical imaging for subsurface cancer detection*, *J. Porphyrin and Phthalocyanines* **8** (2004), 1106–1118.

Nonlinear Continuous Deformation of an Image Based on a Set of Intersecting Straight Lines

Dominique Orban

Coordinator, École Polytechnique de Montréal

Vincent Zalzal

Industrial representative, Matrox Electronic Systems

Paul Armand

Professor, Université de Limoges

Dominique Brunet

Student, Université Laval

Philippe Chaput

Student, McGill University

Arthemy Kiselev

Professor, Université de Montréal and University of Utrecht

Odile Marcotte

CRM and UQAM

Alexi Morin-Duchesne

Student, Université de Montréal

Nikolaj van Omme

Student, École Polytechnique de Montréal

Summary

The problem addressed by this team was proposed by Matrox Electronic Systems. Let I be a closed area within the plane, i.e., an image, which can be assumed to be a rectangle. Let D_i be a set of equidistant points belonging to a straight line and included in I . Let $A = \{D_i \mid i = 1, 2, \dots, M\}$ be a collection of M straight lines belonging to the same plane (i.e., all points in $\bigcup_{i=1}^M D_i$ belong to the same plane). Some of the intersection points of lines in A may belong to I . Finally, let $g(x, y)$ be a function from the plane into itself representing a nonlinear, continuous, and smooth deformation of I . Given the image under g of the union of all the D_i , the problem is to find a method for estimating g . The function g to be estimated represents the nonlinear image deformation produced by the lens of a camera.

The problem just described arises in camera calibration but has features that set it apart from the problems considered so far in the literature. Indeed, the original points and the relative positions of the lines are unknown, and in most cases, the set $\bigcup_{i=1}^M D_i$ is not dense in the plane containing the lines. The members of the team proposed several approaches for solving the problem, and we refer the reader to the presentation on the web site of the workshop. Some team members pursued their work and submitted an article that has now been accepted. The reference is the following: *Nonlinear Continuous Deformation of an Image Based on a Set of Coplanar Straight Lines - A Two-Stage Camera Self-Calibration Procedure*, by P. Armand, A. Kiselev, O. Marcotte, D. Orban, V. Zalzal, to appear in the electronic journal **Mathematics-in-Industry Case Studies** (cf. the site <http://www.micsjournal.ca/index.php/mics>).

Acknowledgements

The team members are very grateful to Arnaud Lina, from Matrox Electronic Systems, for proposing this problem.

Strategic Planning at Kruger Products

Bernard Gendron

Coordinator, Université de Montréal

Vincent Béchar

Industrial representative, Différence S.E.N.C.

John O'Hara

Industrial representative, Kruger Products

Mervat Chouman, Xiaorui Fu and Paul-Virak Khuong

Students, CIRRELT and Université de Montréal

Jean-François Cordeau

Professor, HEC Montréal

Mehdi El Ouali

Student, McGill University

Hamid Ghaffari and Yang Li

Students, McMaster University

Xi Liu

Student, University of Alberta

Stefan Ropke

Postdoctoral fellow, CIRRELT and HEC Montréal

Robert Warren

Student, University of Waterloo

1. Introduction

Kruger Products is a major provider of tissue consumer products to the North American market with leading consumer brands such as Cashmere, Sponge Towels, and Scotties. A major challenge for the company is to make decisions on where and when major investments in capital should be made. Currently, these decisions are based on estimates of costs and revenues that do not take into account the whole logistics network and its evolution over time. Therefore, the company would like to acquire a strategic planning tool that would help it decide whether to expand existing plants, acquire new ones, or close existing ones.

It is in this context that John O'Hara from Kruger Products and Vincent Béchar from Différence S.E.N.C. (a consultant for Kruger Products) submitted the strategic planning problem at Kruger Products to the First Montreal Industrial Problem Solving Workshop (August 20-24, 2007), organized by the CRM and financed by MITACS and the ncm₂. Along with Bernard Gendron and Jean-François Cordeau (two researchers specialized in operations research applied to logistics and transportation), nine students worked on various aspects of the problem during the workshop: Stefan Ropke, Mervat Chouman, Mehdi El Ouali, Xiaorui Fu, Hamid Ghaffari, Paul-Virak Khuong, Yang Li, Xi Liu, and Robert Warren.

During the workshop, the following activities were performed by the team.

- *Modeling the problem.* Following an initial discussion with the industrial partner on the first day of the workshop, a formal definition of the problem was proposed by the participants and validated by the industrial partner. A first mathematical model, called the *Basic Model*, was then formulated. During that time, some students performed a literature review on related strategic planning problems. Although the Basic Model is an accurate representation of the problem, it has potential deficiencies, as demonstrated by existing studies on similar problems. Hence, two alternative models, the *Disaggregated Model* and the *Path Model*, were proposed.
- *Developing computer codes.* The industrial partner did not provide any real-world instances. We were able, however, to propose a data file format based on the description of the problem, in order to allow the team to test solution methods on toy problems. One student also developed a data generator that randomly constructs instances according to this data file format. Two students programmed a C++ code implementing the Basic Model; their code uses the ILOG Concert Library and calls the state-of-the-art mixed integer programming (MIP) solver ILOG CPLEX. Another student implemented the Basic Model using the modeling tool AMPL, which generates input files for CPLEX. Due to a lack of time, these codes could not be tested.
- *Designing solution methods.* For realistic instances, the Disaggregated and Path models yield very large-scale formulations, which cannot be solved directly by a MIP solver. Hence the researchers and postdoctoral students started to investigate decomposition methods for these models, based on cutting-plane, Lagrangian relaxation, and column generation approaches.

The rest of this report is organized as follows. In Section 2, we give a description of the strategic planning problem at Kruger Products. Section 3 first introduces the notation used in the different formulations and then presents the Basic Model. Section 4 and Section 5 describe the Disaggregated and Path Models, respectively. We conclude this report by outlining avenues for future work.

2. Problem Description

The strategic planning problem at Kruger Products can be represented using a two-echelon logistics network comprising three nodes sets: the suppliers, the plants, and the customers. The only possible connections link suppliers to plants and plants to customers. The suppliers represent the sites that provide wood (or raw material) to the plants, which are either existing ones or potential facilities. The wood is processed at the plants to generate the finished products. There is a preliminary processing stage at the plants that creates intermediate products; those can be processed at the same plant or delivered to another plant in order to be transformed into finished products. At each stage, we know how many units of raw material (resp. intermediate product) are used to produce one unit of intermediate product (resp. final product).

The company seeks to optimize its investments over a period of ten years, with new decisions about the plants being evaluated every year. The company wishes to enforce the constraint that a plant that is closed (resp. open) at the beginning of the planning horizon and is then opened (resp. closed) at some period remains so for the rest of the planning horizon.

For each year of the planning horizon, we assume that we have estimates of the demand of each customer for each finished product, as well as capacities and lower bounds at the suppliers and at the plants, both global and for each type of product (raw material or intermediate product). There are fixed costs for opening and operating the potential plants, but also for supplying some amount of raw material. The variable costs consist of three components: manufacturing costs,

transportation costs, and benefits from sales. The latter being larger than the sum of the first two, we can capture the variable cost for each link-product combination at each period by using a generic, negative, variable cost.

Given these fixed and variable costs, the problem consists of minimizing the resulting objective function, such that demands and production bounds are satisfied over a planning horizon of ten years, with the additional constraint that any plant that is opened (or closed) at some point during the planning horizon remains so for the rest of the horizon.

3. Notation and Basic Model

The notation used to model the problem and the three groups of variables are given in the following tables. To obtain a compact representation of the constraints, we consider all types of products (raw material, intermediate or finished products) as members of a generic commodities set. Each commodity has several origins and several destinations, where an origin-destination pair corresponds to a link in the network. To model the situation where intermediate and finished products are generated at the same plant, we create two instances of each plant with a link between the two instances.

| Symbol | Description |
|------------------------|--|
| Sets | |
| S | Set of suppliers. |
| P | Set of plant locations. |
| C | Set of customers. |
| R | Set of raw materials. |
| Q | Set of intermediate products. |
| F | Set of finished products. |
| T | Set of time periods. |
| $O = S \cup P$ | Set of origin nodes. |
| $D = P \cup C$ | Set of destination nodes. |
| $K = R \cup Q \cup F$ | Set of all commodities. |
| $O^k \subseteq O$ | Set of origins that provide commodity $k \in K$. |
| $D^k \subseteq D$ | Set of destinations that require commodity $k \in K$. |
| $P^k \subseteq P$ | Set of plants that potentially require commodity $k \in K$ to make their products. |
| $K^o \subseteq K$ | Set of commodities that are produced at origin $o \in O$. |
| $P_1 \subseteq P$ | Set of plants that are open at the start of the planning horizon. |
| $P_0 \subseteq P$ | Set of plants that are closed at the start of the planning horizon. |
| Data | |
| a_{cf}^t | Max. demand of customer $c \in C$ for product $f \in F$ in period $t \in T$. |
| \bar{q}_{ok}^t | Maximum production of commodity $k \in K$ at origin $o \in O$ in period $t \in T$. |
| \underline{q}_{ok}^t | Minimum production of commodity $k \in K$ at origin $o \in O$ in period $t \in T$ if origin o is selected. |
| \bar{u}_o^t | Overall capacity at origin $o \in O$ in time period $t \in T$. |
| \underline{u}_o^t | Minimum production at origin $o \in O$ in time period $t \in T$ if origin o is selected. |

| Symbol | Description |
|------------------------------|---|
| Data | |
| b_{pkj} | Number of units of $k \in K$ needed to have one unit of $j \in K$ at plant $p \in P$. |
| c_o^t | Fixed cost of selecting origin $o \in O$ in time period $t \in T$. |
| c_{ok}^t | Fixed cost of producing commodity $k \in K$ at origin $o \in O$ in time period $t \in T$. |
| c_{odk}^t | Cost of transporting one unit of commodity $k \in K$ from $o \in O$ to $d \in D$ in time period $t \in T$. If $d \in C$ then the cost also ‘‘includes’’ the revenue of selling one unit of commodity k to customer d . Thus c_{odk}^t can be negative when $d \in D$. The cost also includes the cost of producing one unit of commodity k if $o \in P$ and the cost of buying one unit of commodity k if $o \in S$. |
| L_p^k | Set of products that use commodity k at plant p . |
| Decision variables | |
| $y_o^t \in \{0, 1\}$ | equals 1 iff origin $o \in O$ is selected in period $t \in T$. |
| $v_{ok}^t \in \{0, 1\}$ | equals 1 iff origin $o \in O$ provides commodity $k \in K$ in period $t \in T$. |
| $x_{odk}^t \in \mathbb{R}^+$ | Quantity of commodity $k \in K$ transported from origin $o \in O$ to destination $d \in D$ in period $t \in T$. |

The variables representing the commodities correspond to links (or arcs) of the network. The resulting model is a multiperiod multicommodity generalized arc-based network flow formulation.

$$\min \sum_{t \in T} \sum_{o \in O} c_o^t y_o^t + \sum_{t \in T} \sum_{k \in K} \sum_{o \in O^k} c_{ok}^t v_{ok}^t + \sum_{t \in T} \sum_{k \in K} \sum_{o \in O^k} \sum_{d \in D^k} c_{odk}^t x_{odk}^t \quad (1)$$

subject to

$$\sum_{p \in P} x_{pcf}^t \leq a_{cf}^t \quad \forall c \in C, \forall t \in T, \forall f \in F \quad (2)$$

$$\sum_{o \in O^k} x_{opk}^t - \sum_{j \in L_p^k} \sum_{d \in D^j} b_{pkj} x_{pdj}^t = 0 \quad \forall k \in R \cup Q, \forall p \in P^k, \forall t \in T \quad (3)$$

$$\sum_{k \in K^o} \sum_{d \in D^k} x_{odk}^t \leq \bar{u}_o^t y_o^t \quad \forall o \in O, \forall t \in T \quad (4)$$

$$\sum_{k \in K^o} \sum_{d \in D^k} x_{odk}^t \geq \underline{u}_o^t y_o^t \quad \forall o \in O, \forall t \in T \quad (5)$$

$$\sum_{d \in D^k} x_{odk}^t \leq \bar{q}_{ok}^t v_{ok}^t \quad \forall o \in O, \forall k \in K^o, \forall t \in T \quad (6)$$

$$\sum_{d \in D^k} x_{odk}^t \geq \underline{q}_{ok}^t v_{ok}^t \quad \forall o \in O, \forall k \in K^o, \forall t \in T \quad (7)$$

$$y_o^{t+1} \geq y_o^t \quad \forall o \in P_0, \forall t \in T \quad (8)$$

$$y_o^{t+1} \leq y_o^t \quad \forall o \in P_1, \forall t \in T \quad (9)$$

$$y_o^t \in \{0, 1\} \quad \forall o \in O, \forall t \in T \quad (10)$$

$$v_{ok}^t \in \{0, 1\} \quad \forall o \in O, \forall t \in T, \forall k \in K \quad (11)$$

$$x_{odk}^t \in \mathbb{R}^+ \quad \forall o \in O, \forall d \in D, \forall t \in T, \forall k \in K \quad (12)$$

Inequalities (2) enforce the customer demand constraint. Equalities (3) are the flow conservation constraints. Inequalities (4) and (5) enforce upper and lower bounds on overall production at origin $o \in O$. Inequalities (6) and (7) enforce upper and lower bounds on production of each commodity $k \in K$ at origin $o \in O$. Inequalities (8) ensure that once a plant has been opened, it is kept open for the rest of the planning horizon. Inequalities (9) ensure that once a plant has been closed, it is kept closed for the rest of the planning horizon.

4. Forcing Constraints and Disaggregated Model

The Basic Model is an accurate representation of the problem, but its linear programming (LP) relaxation is weak, especially when fixed charges are high and capacities are tight. Consequently, solving realistic instances by traditional LP-based branch-and-bound methods seems to be a difficult task. To improve the model, one might add valid inequalities, which for fixed-charge models often take the form of so-called *forcing constraints*, linking fixed-charge binary variables to other variables: thus the latter are forced to take the value 0 whenever a corresponding fixed-charge variable is equal to 0.

The first obvious forcing constraints relate the two families of binary variables:

$$v_{ok}^t \leq y_o^t \quad \forall o \in O, \forall k \in K^o, \forall t \in T. \quad (13)$$

We can also define forcing constraints relating the flow of finished products at each customer and the binary variables:

$$x_{pcf}^t \leq a_{cf}^t v_{pf}^t \quad \forall c \in C, \forall f \in F, \forall p \in O^f, \forall t \in T. \quad (14)$$

Similar forcing constraints can be defined for arcs between plants or between a supplier and a plant, but these valid inequalities are weak, since their right hand-side is the sum of the demands of finished products for several customers. We can obtain stronger forcing constraints by introducing disaggregated flow variables binding together the original flow variables and the destination of the flow (i.e., a “customer-finished product” combination). More precisely, for arcs between two plants, we define the variables x_{lpqcf}^t , where x_{lpqcf}^t denotes the “amount of intermediate product $q \in Q$ transported from plant $l \in P$ to plant $p \in P$ and entering into the composition of final product $f \in F$ delivered to customer $c \in C$ at time $t \in T$ ”.

These variables are related to the original flow variables as follows:

$$\sum_{l \in P} \sum_{q \in Q} (b_{pqf})^{-1} x_{lpqcf}^t = x_{pcf}^t \quad \forall c \in C, \forall f \in F, \forall p \in O^f, \forall t \in T. \quad (15)$$

We can then add the following forcing constraints:

$$x_{lpqcf}^t \leq (b_{pqf} a_{cf}^t) v_{lq}^t \quad \forall c \in C, \forall f \in F, \forall p \in O^f, \forall q \in Q, \forall l \in O^q, \forall t \in T. \quad (16)$$

For arcs between suppliers and plants, we define the variables $x_{slrpqcf}^t$, where $x_{slrpqcf}^t$ denotes the “amount of raw material $r \in R$ transported from supplier $s \in S$ to plant $l \in P$, transformed into intermediate product $q \in Q$, then transported to plant $p \in P$ and entering into the composition of final product $f \in F$ delivered to customer $c \in C$ at time $t \in T$ ”.

These variables are related to the disaggregated flow variables between plants by the following equations:

$$\sum_{s \in S} \sum_{r \in R} (b_{lrq})^{-1} x_{slrpqcf}^t = x_{lpqcf}^t \quad \forall c \in C, \forall f \in F, \forall p \in O^f, \forall q \in Q, \forall l \in O^q, \forall t \in T. \quad (17)$$

We can then derive the forcing constraints below.

$$x_{slrpqcf}^t \leq (b_{lrq}b_{pqf}a_{cf}^t)v_{sr}^t \quad \forall c \in C, \forall f \in F, \forall p \in O^f, \forall q \in Q, \quad (18)$$

$$\forall l \in O^q, \forall r \in R, \forall s \in S^r, \forall t \in T$$

Note that S^r is the set of suppliers that can supply the commodity r . The disaggregated model is obtained by adding constraints (13)-(18) to the basic model. Using equations (17) and (15), it is possible to remove from the model all flow variables except the variables $x_{slrpqcf}^t$; indeed, the following equations express each type of flow variable as a function of the $x_{slrpqcf}^t$.

$$x_{pcf}^t = \sum_{s \in S} \sum_{r \in R} \sum_{l \in P} \sum_{q \in Q} (b_{lrq}b_{pqf})^{-1} x_{slrpqcf}^t \quad \forall c \in C, \forall f \in F, \forall p \in O^f, \forall t \in T \quad (19)$$

$$x_{lpq}^t = \sum_{s \in S} \sum_{r \in R} \sum_{c \in C} \sum_{f \in F} (b_{lrq})^{-1} x_{slrpqcf}^t \quad \forall p \in O^f, \forall q \in Q, \forall l \in O^q, \forall t \in T \quad (20)$$

$$x_{slr}^t = \sum_{p \in P} \sum_{q \in Q} \sum_{c \in C} \sum_{f \in F} x_{slrpqcf}^t \quad \forall l \in P, \forall r \in R, \forall s \in S^r, \forall t \in T \quad (21)$$

Since this reformulation results in a very large-scale model, decomposition methods are required to solve it. We have considered three decomposition approaches: a cutting-plane method and two Lagrangian relaxations.

The cutting-plane method would start by solving the LP relaxation of the Basic Model. At each iteration, it would generate disaggregated flow variables, along with their corresponding flow conservation equations; this procedure would be carried out only when these variables have a chance to lead to violated forcing constraints. It would then solve the resulting LP relaxation, add violated forcing constraints and iterate until there are no more violated forcing constraints. This method would solve the LP relaxation of the Disaggregated Model; it could be complemented by heuristic techniques and embedded in a branch-and-cut algorithm in order to produce (nearly) optimal solutions to very large-scale problems.

Two Lagrangian relaxations were considered. The first one consists of relaxing the constraints linking consecutive time periods. The resulting Lagrangian subproblem decomposes by time period, and can be solved using the same cutting-plane technique as above. A second Lagrangian relaxation appears more promising, since the resulting Lagrangian subproblem is easier to solve. It consists of relaxing the demand and flow conservation constraints; the resulting subproblem decomposes by node and by time period. It can be shown that both Lagrangian relaxations provide better lower bounds than the LP relaxation.

5. Path Model

Assume we define an *expanded network* consisting of the following types of nodes: couples of the form “supplier-raw material”, couples “plant-intermediate product”, couples “plant-finished product”, and customers. For the first three types of nodes, we will use the notation (o, k) to represent some origin o (supplier or plant) and some commodity k (raw material, intermediate product, or finished product). In this expanded network, there is an arc between two nodes (o, k) and (p, l) if and only if there is an arc between o and p in the original network and $b_{pkl} > 0$ holds. Finally, there is an arc between (p, f) and c if and only if there is an arc between p and c in the original network and $a_{cf}^t > 0$ holds for at least one period t .

Let W denote the set of all paths in this expanded network. Also, we will define the following sets: W_{cf} is the set of all paths delivering finished product f to customer c , and W_{ok} is the set of all paths having (o, k) as starting or intermediate node. We can characterize a path $w \in W$ by its nodes; hence we will use the notation $w = (slrpqcf)$ to denote the path $(s, r) \rightarrow (l, q) \rightarrow (p, f) \rightarrow c$. For such a path, we define *conversion factors* as follows: $b_{wr} = 1$, $b_{wq} = b_{lrq}$, and $b_{wf} = b_{lrf}b_{pqf}$. The transportation cost c_w^t for the path w at time t can then be computed as follows:

$$c_w^t = c_{slr}^t b_{wr} + c_{lpq}^t (b_{wq})^{-1} + c_{pcf}^t (b_{wf})^{-1}.$$

To formulate the path model, we introduce the following path variables: x_w^t is the amount of flow (of raw material) transported on path $w \in W$ at time $t \in T$. The path model can then be written as follows. (Note that the flow conservation constraints are not needed any more.)

$$\min \sum_{t \in T} \sum_{o \in O} c_o^t y_o^t + \sum_{t \in T} \sum_{k \in K} \sum_{o \in O^k} c_{ok}^t v_{ok}^t + \sum_{t \in T} \sum_{w \in W} c_w^t x_w^t \quad (22)$$

subject to

$$\sum_{w \in W_{cf}} (b_{wf})^{-1} x_w^t \leq a_{cf}^t \quad \forall c \in C, \forall t \in T, \forall f \in F \quad (23)$$

$$\sum_{k \in K^o} \sum_{w \in W_{ok}} (b_{wk})^{-1} x_w^t \leq \bar{u}_o^t y_o^t \quad \forall o \in O, \forall t \in T \quad (24)$$

$$\sum_{k \in K^o} \sum_{w \in W_{ok}} (b_{wk})^{-1} x_w^t \geq \underline{u}_o^t y_o^t \quad \forall o \in O, \forall t \in T \quad (25)$$

$$\sum_{w \in W_{ok}} (b_{wk})^{-1} x_w^t \leq \bar{q}_{ok}^t v_{ok}^t \quad \forall o \in O, \forall k \in K^o, \forall t \in T \quad (26)$$

$$\sum_{w \in W_{ok}} (b_{wk})^{-1} x_w^t \geq \underline{q}_{ok}^t v_{ok}^t \quad \forall o \in O, \forall k \in K^o, \forall t \in T \quad (27)$$

$$y_o^{t+1} \geq y_o^t \quad \forall o \in P_0, \forall t \in T \quad (28)$$

$$y_o^{t+1} \leq y_o^t \quad \forall o \in P_1, \forall t \in T \quad (29)$$

$$y_o^t \in \{0, 1\} \quad \forall o \in O, \forall t \in T \quad (30)$$

$$v_{ok}^t \in \{0, 1\} \quad \forall o \in O, \forall t \in T, \forall k \in K \quad (31)$$

$$x_w^t \in \mathbb{R}^+ \quad \forall w \in W, \forall t \in T \quad (32)$$

The forcing constraints (13), (14), (16), and (18) can then be added to this formulation to strengthen its LP relaxation. In terms of the path variables, the last three types of forcing constraints can be written as follows:

$$\sum_{w \in W_{cf} \cap W_{pf}} (b_{wf})^{-1} x_w^t \leq a_{cf}^t v_{pf}^t \quad \forall c \in C, \forall f \in F, \forall p \in O^f, \forall t \in T \quad (33)$$

$$\sum_{w \in W_{cf} \cap W_{pf} \cap W_{lq}} (b_{wq})^{-1} x_w^t \leq (b_{pqf} a_{cf}^t) v_{lq}^t \quad \forall c \in C, \forall f \in F, \forall p \in O^f, \forall q \in Q, \forall l \in O^q, \forall t \in T \quad (34)$$

$$x_w^t \leq (b_{wf} a_{cf}^t) v_{sr}^t \quad \forall c \in C, \forall f \in F, \forall p \in O^f, \forall q \in Q, \quad (35)$$

$$\forall l \in O^q, \forall r \in R, \forall s \in S^r, \forall t \in T, w = (slrpqcf)$$

It can be shown that the LP relaxation bounds provided by the Path Model and by the Disaggregated Model are equal. To solve the LP relaxation of the Path Model, a column generation approach can be used. This technique starts by generating a small subset of all paths, and at each iteration, solves the resulting restricted LP relaxation. New (non-basic) path variables with negative reduced costs are then generated by solving the pricing subproblem. We still need to investigate the structure of this pricing subproblem to determine whether or not it can be solved efficiently.

6. Conclusion

In this report, we have presented the work carried out by the team of researchers and students who studied the strategic planning problem of Kruger Products during the First Montreal IPSW. The team proposed three formulations for this problem: the Basic, Disaggregated, and Path Models. The first one can be solved by a state-of-the-art MIP solver, but its LP relaxation is notoriously weak when fixed charges are high and capacities are tight. The two other models provide better LP relaxations, but require decomposition methods if one is to solve large-scale instances. The team has outlined three approaches for the Disaggregated Model: a cutting-plane method and two Lagrangian relaxations; it has also outlined a column generation approach for the Path Model.

Before implementing these decomposition approaches, it is first necessary to test the Basic Model, both on randomly generated instances and on real data. Then the two other models could be implemented and tested on small instances, without using decomposition, in order to determine to what extent their LP relaxation value improves the bound obtained by solving the LP relaxation of the Basic Model. Finally, decomposition approaches could be implemented and tested on large-scale instances.

Acknowledgments

We gratefully acknowledge the collaboration of John O'Hara, from Kruger Products, and Vincent B  chard, from Diff  rence S.E.N.C.. We also gratefully acknowledge the support of MITACS, the main sponsor of this workshop.

Routes de collecte du lait de la Fédération des producteurs de lait du Québec

Louis-Martin Rousseau

Coordonnateur, CIRRELT et École Polytechnique de Montréal

Réjean Robitaille

Représentant industriel, FPLQ

Michel Gendreau

Professeur, CIRRELT et École Polytechnique de Montréal

Guillaume Provencher

Étudiant, Université de Montréal

Alessandro Zanarini

Étudiant, École Polytechnique de Montréal

Marie-Ève Rancourt

Rédactrice, HEC Montréal

1. Introduction

L'industrie laitière est la première industrie agricole en importance au Québec. À la fin de 2006, la production de lait des fermes laitières québécoises s'élevait à près de trois milliards de litres et correspondait à près de 2 milliards de dollars. La gamme des produits laitiers fabriqués au Québec ne cesse de s'élargir : la grande variété des fromages québécois est sûrement le meilleur exemple illustrant la diversité des produits dérivés du lait au Québec. Les exportations de produits laitiers sont devenues plus importantes que les importations, ce qui est un autre signe de la vitalité de l'industrie laitière.

Les quelque 7000 fermes laitières québécoises mettent collectivement en marché leur lait par l'intermédiaire du programme à frais partagés des producteurs de lait du Québec. La Fédération des producteurs de lait du Québec (FPLQ) est une fédération de 14 syndicats fondée en 1983. Les producteurs de lait ont délégué à la FPLQ la responsabilité de négocier l'ensemble des conditions de vente avec les transformateurs. La FPLQ a adopté un système de gestion de l'offre afin d'établir un équilibre entre l'offre et la demande de lait et de produits laitiers. Par ailleurs, les conditions de transport du lait sont négociées avec les transporteurs par la FPLQ et font partie d'une convention nationale. Plus précisément, la FPLQ a la responsabilité d'organiser et de réglementer les étapes suivantes du transport : le ramassage du lait chez les producteurs, l'acheminement du lait aux usines de transformation et la distribution des produits laitiers aux commerçants. C'est dans le but d'optimiser le processus de transport du lait avant sa transformation que la Fédération a voulu travailler en collaboration avec les participants du Premier atelier de résolution de problèmes industriels de Montréal, parrainé par MITACS et le rcm₂.

2. Description du problème

Dans le cadre de cet atelier, nous avons donc cherché à développer des stratégies afin d'aider la FPLQ à résoudre le problème du ramassage de lait chez les producteurs et de l'acheminement du lait aux

usines de transformation. En 2006, le Québec comptait près de 7390 fermes laitières dispersées sur un vaste territoire. Pour assurer le ramassage de lait chez les producteurs et son acheminement aux usines de transformation, la FPLQ compte sur des transporteurs privés avec lesquels elle a signé des contrats. Ainsi, une flotte totale de 274 véhicules de différents types, effectuant 592 circuits chaque jour, assure le transport du lait. Ces précisions donnent une idée de l'envergure du problème auquel la Fédération fait face. Actuellement, les tournées de véhicules sont partiellement déterminées par une procédure automatisée ; elles sont ensuite peaufinées par des employés qui sont des experts de la répartition des véhicules.

Le problème proposé par la FPLQ est un cas particulier de problème de tournées de véhicules (PTV), selon la terminologie de la recherche opérationnelle dans le domaine des transports. Dans ce qui suit, les caractéristiques du problème en question seront examinées de plus près. On constatera que ce problème complexe n'a pas le format d'un problème classique de tournées de véhicules, tel qu'étudié dans les articles scientifiques. On dira donc que ce problème concret est un problème *riche* de tournées de véhicules.

2.1. CARACTÉRISTIQUES DU PROBLÈME DE LA FPLQ

Pour effectuer le ramassage du lait, les transporteurs disposent d'une flotte hétérogène comportant cinq types de véhicules. Ceux-ci peuvent contenir de 14000 litres à 35000 litres de lait, et leurs frais d'exploitation varient d'un modèle à l'autre. Des contraintes restreignent l'accès des camions volumineux à certains sites de ramassage (fermes) ou de distribution (usines). De plus, la journée de travail d'un camionneur ne peut pas dépasser 15 heures et lorsque celle-ci dépasse 12 heures, le salaire du camionneur augmente d'environ 40%. Dans les articles sur le sujet, ces caractéristiques sont résumées de la manière suivante : le problème de la FPLQ est un PTV avec une flotte de véhicules hétérogène, des contraintes de capacité et des contraintes de durée des tournées de véhicules.

On appelle *circuit* la tournée effectuée par un véhicule et qui consiste à ramasser le lait chez différents producteurs et à l'acheminer à une usine de transformation. On considère qu'un circuit est scindé en quatre segments, décrits ci-dessous. Notez qu'un véhicule part d'un dépôt et retourne au même dépôt une fois sa tournée terminée.

- segment 1 : du dépôt jusqu'au premier producteur
- segment 2 : du premier producteur jusqu'au dernier producteur
- segment 3 : de la fin du ramassage jusqu'à l'usine de transformation
- segment 4 : de l'usine de transformation jusqu'au dépôt

La détermination du circuit d'un véhicule donné implique qu'il faut choisir non seulement les producteurs desservis par ce véhicule, mais aussi l'usine à laquelle le lait sera acheminé avant que le véhicule ne retourne à son dépôt. La figure 1 illustre une tournée de véhicule. Dans un problème classique de tournées de véhicule, il n'est pas nécessaire de choisir une usine par tournée ; la confection de tournées est donc plus complexe dans le cas de la FPLQ.

Par ailleurs, la formule donnant les coûts de transport, adoptée à la fois par la FPLQ et les transporteurs, est particulièrement complexe. La formule en question a pour but d'estimer correctement les frais d'exploitation des transporteurs. Par exemple, elle tient compte des coûts d'amortissement de l'équipement et des coûts du carburant et de la main-d'oeuvre. La description complète de ces coûts se trouve dans un document officiel d'une vingtaine de pages (cf. [2]). Toutefois, dans le cadre

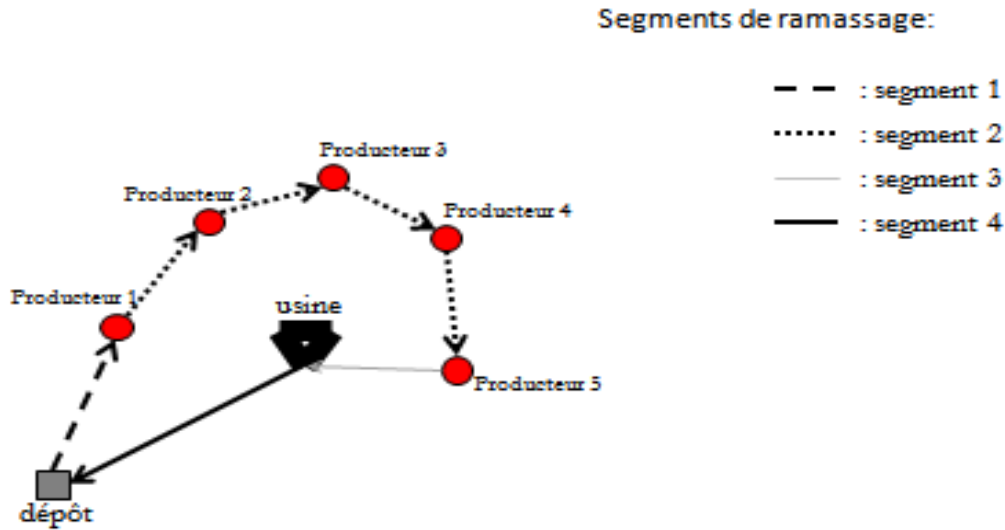


Figure 1. Illustration des quatre segments dans un circuit de ramassage de lait

de ce projet, nous ne nous intéresserons qu'aux coûts les plus importants, c'est-à-dire les coûts de la main-d'oeuvre et du carburant. Notons que pour estimer le coût d'un circuit, il suffit d'estimer sa longueur et la durée de son parcours, étant donné que le prix du carburant et le taux horaire d'un camionneur sont connus a priori. Or, pour une suite donnée de producteurs à visiter suivie d'une usine à desservir, il est facile d'évaluer la longueur du circuit associé à cette suite, mais il est plus difficile d'estimer le temps nécessaire pour parcourir le circuit.

Afin de tenter de simuler les délais encourus par un camion lors du ramassage en milieu rural, la convention de transport évalue la vitesse des camions en fonction des déplacements qu'ils effectuent. Cette évaluation tient compte des opérations requises ainsi que du fait que les véhicules vont moins vite lorsqu'ils sont remplis. Rappelons que les trajets des véhicules sont subdivisés en quatre segments. Les formules ci-dessous permettent d'évaluer le temps de conduite des camionneurs pour chaque segment.

- segment 1 : de la base jusqu'au premier ramassage

$$t_1 = 2,8 + 1,106505x_1 - 0,10766x_1^{1,3},$$

où x_1 dénote le nombre de kilomètres de la base au premier ramassage. La vitesse de ce segment, $v_1 = x_1/t_1$, ne peut être supérieure à 85 km/h ni inférieure à 10 km/h. Si c'est le cas, le nombre de minutes retenues pour ce segment est déterminé en utilisant la borne la plus proche.

- segment 2 : durant le ramassage

$$t_2 = 17,5 + 13,58662x_2 - 9,20408x_2^{1,05} - 18,6158\frac{x_2}{y} + 1,401566\frac{x_2^{1,7}}{y},$$

où x_2 dénote le nombre de kilomètres entre le premier et le dernier producteur, moins le nombre de kilomètres au-delà de 34 km entre deux producteurs consécutifs, et y dénote le nombre de producteurs sur ce segment. La vitesse ne peut être supérieure à 45 km/h ni inférieure à 10 km/h. Si c'est le cas, le nombre de minutes retenues pour ce segment est déterminé en utilisant la borne la plus proche.

- segment 3 : de la fin du ramassage jusqu'à l'usine de transformation

$$t_3 = 3,8 + 1,901233x_3 - 0,96033x_3^{1,05},$$

où x_3 dénote le nombre de kilomètres entre le dernier producteur et l'usine, plus le nombre de kilomètres au-delà de 34 km entre deux producteurs consécutifs. Cette vitesse ne peut être supérieure à 85 km/h ni inférieure à 10 km/h. Si c'est le cas, le nombre de minutes retenues pour ce segment est déterminé en utilisant la borne la plus proche.

- segment 4 : de l'usine à la base

$$t_4 = 0,4 + 1,378538x_4 - 0,17215x_4^{1,03},$$

où x_4 dénote le nombre de kilomètres entre l'usine et la base. La vitesse ne peut être supérieure à 85 km/h ni inférieure à 10 km/h sur ce segment. Si c'est le cas, le nombre de minutes retenues pour ce segment est déterminé en utilisant la borne la plus proche.

L'objectif visé par la FPLQ est évident : concevoir un ensemble de tournées de véhicules qui utilise les véhicules disponibles afin de transporter à moindre coût le lait des producteurs vers les usines, et ce, tout en respectant les contraintes. Le but de ce projet est donc de mettre au point des stratégies susceptibles de produire de meilleures solutions que celles présentement obtenues par la FPLQ.

3. Méthodes de résolution du problème et résultats préliminaires

Pour résoudre le problème de la FPLQ, deux approches de résolution ont été développées. Dans un premier temps, nous avons travaillé sur une approche basée sur l'application d'une métaheuristique ; dans un deuxième temps, nous avons conçu une approche multiphases.

3.1. APPROCHE DIRECTE

La première approche de résolution du problème est une méthode de recherche avec tabous, c'est-à-dire une métaheuristique de recherche locale qui explore l'espace des solutions en se déplaçant d'une solution courante s à la meilleure solution dans son voisinage $N(s)$. Contrairement aux méthodes d'amélioration traditionnelles, la méthode de recherche avec tabous permet d'examiner temporairement des solutions moins bonnes que la meilleure solution trouvée jusque là. Des mécanismes imposant des tabous pour certains déplacements sont alors mis en oeuvre pour éviter que la recherche ne cycle.

L'algorithme que nous avons utilisé pour résoudre le problème étudié est celui développé par Cordeau, Laporte et Mercier [1]. Dans cette méthode, le voisinage d'une solution est obtenu en déplaçant un client de sa tournée courante vers une autre tournée et en choisissant la position qui minimise les coûts dans la nouvelle tournée. Ainsi, pendant chaque itération, tous les déplacements d'un seul producteur d'une tournée à l'autre sont examinés. Si un déplacement non tabou améliore la solution, on l'applique, et si aucun déplacement n'améliore la solution, on applique le meilleur déplacement et on décrète qu'il sera tabou par la suite. De plus, des mécanismes de diversification et d'aspiration sont utilisés afin d'aider l'algorithme à mieux explorer l'espace des solutions. Pour plus de détails sur l'algorithme, nous suggérons au lecteur de se reporter à l'article de Cordeau, Laporte et Mercier [1].

Dans les formules proposées par la FPLQ pour évaluer les temps de transport sur les segments d'une tournée de véhicule, on remarque que les temps ne varient pas linéairement en fonction des distances. Cette complexité dans l'évaluation des coûts fait en sorte qu'il est difficile de déterminer rapidement l'impact d'une modification dans une tournée. Si on considère que le coût d'une tournée est fonction de sa longueur totale seulement, on peut évaluer plus facilement le coût d'une modification. Par exemple, le coût d'insérer un client k dans une tournée entre deux clients i et j peut être évalué en calculant le coût $c(i, k, j)$ de faire le détour par le client k , ce qui donne

$$c(i, k, j) = C(d(i, k) + d(k, j) - d(i, j)),$$

où $d(i, j)$ dénote la distance entre le client i et le client j et C dénote une constante quelconque, par exemple, une constante proportionnelle au prix de l'essence. Afin de simplifier les calculs, nous avons donc cherché à déterminer si les fonctions de temps de parcours données ci-dessus ne pouvaient pas être estimées à l'aide de fonctions linéaires. Nous nous sommes surtout penchés sur le segment de ramassage du lait chez les producteurs puisque celui-ci est le plus important dans une tournée de véhicule. Le graphique de la figure 2 représente le temps de parcours d'un véhicule en fonction de la distance parcourue pour un nombre donné de fermes (c'est-à-dire de producteurs de lait).

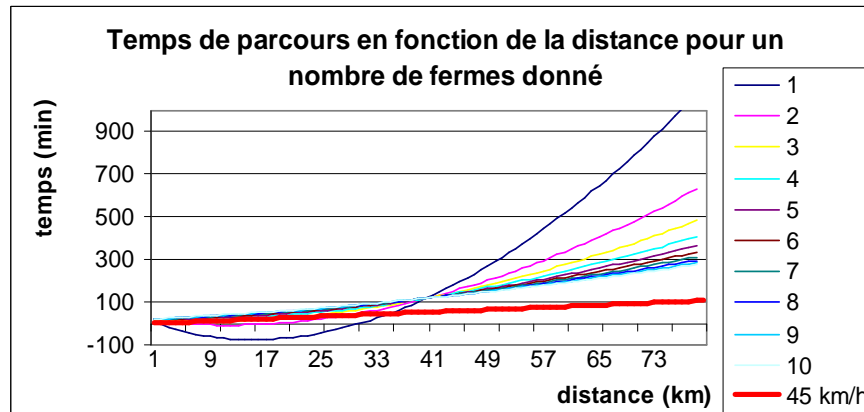


Figure 2. Graphique illustrant le temps de parcours en fonction de la distance pour un nombre donné de producteurs de lait

On constate que plus le nombre de producteurs augmente, plus la fonction ressemble à une fonction linéaire. Comme un camion dessert en général plus de six producteurs, nous avons jugé raisonnable de supposer que le temps de parcours est proportionnel à la distance parcourue. Notons qu'à un camion donné est affectée une usine, c'est-à-dire qu'un camion visite un certain nombre de producteurs et livre leur lait à une usine spécifiée à l'avance.

Évidemment, avant que nous puissions résoudre le problème avec l'algorithme de recherche avec tabous, les données fournies par la FPLQ ont dû être importées, converties et mises en forme. Par la suite, le problème a été résolu grâce à l'algorithme de recherche avec tabous en utilisant une approximation linéaire du temps de parcours. Après avoir laissé l'algorithme tourner pendant une dizaine de secondes, nous avons obtenu un ensemble de tournées de véhicules que nous avons comparées avec les tournées effectuées par les transporteurs de la FPLQ. Pour certains exemplaires du problème, nous avons pu obtenir des tournées dont la longueur était de 10% inférieure à celle des tournées de la FPLQ. Ces résultats sont très prometteurs et nous pensons que l'application de cette méthode de résolution pourrait être bénéfique pour la FPLQ.

3.2. APPROCHE MULTI-PHASES

Nous avons aussi exploré la possibilité d'une autre approche pour résoudre le problème. Cette méthode comporte deux phases : une phase de regroupement des producteurs et d'affectation des regroupements aux usines, et une phase d'ordonnancement.

La phase de regroupement consiste à former des sous-groupes de producteurs tels que la quantité totale de lait à ramasser dans un sous-groupe ne dépasse pas la capacité Q_k d'un véhicule k . Notons que N dénote l'ensemble des producteurs, K l'ensemble des véhicules, q_i la quantité de lait à ramasser chez le producteur no i , et L_k l'ensemble des usines qui peuvent être affectées au véhicule k . Chaque sous-groupe a un "centre", c'est-à-dire un producteur donné appartenant à ce sous-groupe. La solution d'un modèle de programmation mathématique permettra de déterminer les sous-groupes de producteurs.

Avant de présenter ce modèle, notons que le coût $c_1(i, j)$ d'affecter un producteur j à un sous-groupe centré en i est estimé par $c_1(i, j) = d(i, j)$. D'autre part, le coût $c_2(i, l, k)$ d'affecter une usine l à un sous-groupe centré en i et desservi par le véhicule k est estimé par $c_2(i, l, k) = d(i, l) + d(l, b_k)$, où b_k est le dépôt du véhicule k . Les variables du modèle sont les x_{ij} (pour $i \in N$ et $j \in N$), les y_i (pour $i \in N$), et les w_{ikl} (pour $i \in N$, $k \in K$ et $l \in L_k$). Ces variables sont définies ci-dessous.

$$x_{ij} = \begin{cases} 1 & \text{si le producteur } j \text{ appartient au sous-groupe centré en } i \\ 0 & \text{sinon} \end{cases}$$

$$y_i = \begin{cases} 1 & \text{si le producteur } i \text{ est le centre d'un sous-groupe} \\ 0 & \text{sinon} \end{cases}$$

$$w_{ikl} = \begin{cases} 1 & \text{si le sous-groupe centré en } i \text{ est desservi par le véhicule } k \text{ apportant le lait à l'usine } l \\ 0 & \text{sinon} \end{cases}$$

Nous pouvons maintenant présenter le modèle de programmation mathématique mentionné ci-dessus, où m dénote le nombre de sous-groupes (fixé a priori).

$$\text{minimiser } \sum_{i \in N} \sum_{j \in N} c_1(i, j)x_{ij} + \sum_{i \in N} \sum_{k \in K} \sum_{l \in L} c_2(i, l, k)w_{ikl} \quad (36)$$

$$\sum_{i \in N} y_i = m \quad (37)$$

$$x_{ij} \leq y_i \quad \forall i \in N \quad \forall j \in N \quad (38)$$

$$\sum_{i \in N} x_{ij} = 1 \quad \forall j \in N \quad (39)$$

$$\sum_{k \in K} \sum_{l \in L_k} w_{ikl} = y_i \quad \forall i \in N \quad (40)$$

$$\sum_{i \in N} \sum_{l \in L_k} w_{ikl} \leq 1 \quad \forall k \in K \quad (41)$$

$$\sum_{j \in N} q_j x_{ij} \leq Q_k + M \left(1 - \sum_{l \in L_k} w_{ikl} \right) \quad \forall i \in N \quad \forall k \in K \quad (42)$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in N \quad \forall j \in N \quad (43)$$

$$y_i \in \{0, 1\} \quad \forall i \in N \quad (44)$$

$$w_{ikl} \in \{0, 1\} \quad \forall i \in N \quad \forall k \in K \quad \forall l \in L_k \quad (45)$$

La contrainte (37) exprime le fait qu'il y a m sous-groupes. Les contraintes (38) expriment le fait qu'un producteur ne peut être affecté au sous-groupe centré en i que si ce sous-groupe existe. Les contraintes (39) expriment le fait que tout producteur doit être affecté à un sous-groupe. Les contraintes (40) expriment le fait que s'il existe un sous-groupe centré en i , alors un camionneur et une usine devront être affectés à ce sous-groupe. Les contraintes (41) expriment le fait qu'un camionneur ne peut pas effectuer plus d'une tournée. Enfin, les contraintes (42) expriment le fait que si le camionneur k est affecté au sous-groupe centré en i , la quantité de lait qu'il ramassera ne dépassera pas sa capacité. Notons que M est une très grande constante, c'est-à-dire une constante plus grande que la somme de tous les q_j .

Naturellement, la première phase permet de déterminer des sous-groupes seulement ; le modèle ne spécifie pas l'ordre dans lequel les producteurs d'un sous-groupe sont desservis par le camionneur. La deuxième phase consiste donc à déterminer l'ordre dans lequel les producteurs d'un sous-groupe donné seront desservis par le camionneur, qui devra ensuite apporter le lait à l'usine choisie et retourner à son dépôt. Ce sous-problème est en réalité un problème de commis-voyageur, avec la contrainte que les deux derniers lieux de la tournée (l'usine et le dépôt) sont fixés. Plusieurs algorithmes sophistiqués ont été proposés pour résoudre le problème du commis-voyageur (cf. Laporte [3]). Malheureusement nous n'avons pas pu mettre en oeuvre la deuxième phase, mais nous pensons qu'elle pourrait donner des résultats intéressants.

4. Conclusion

Dans le cadre de cet atelier, nous avons proposé deux stratégies pour résoudre le problème du ramassage de lait de la FPLQ. Le problème de la FPLQ est un problème riche de tournées de véhicules caractérisé par des fonctions complexes estimant le temps de parcours des tournées et par le fait qu'à la fin d'une tournée de ramassage de lait, le camion doit apporter le lait à une usine. La première méthode proposée pour résoudre le problème est l'algorithme de recherche avec tabous de Cordeau, Laporte et Mercier. Celui-ci a été implanté en approximant le temps de parcours d'une tournée par une fonction linéaire de la distance. Les résultats obtenus sont encourageants puisque dans certains cas, nous avons pu trouver des tournées dont la longueur est de 10% inférieure à la longueur des tournées de la FPLQ.

La deuxième méthode comporte deux phases et est basée sur la programmation mathématique. Malheureusement, elle n'a pas été testée, mais nous pensons qu'elle devrait l'être dans un avenir proche. D'autre part, il serait sûrement profitable d'adapter la procédure de recherche avec tabous afin qu'elle prenne en compte les caractéristiques du problème de la FPLQ. Par exemple, on pourrait utiliser les "vraies" fonctions et non pas seulement des approximations linéaires. Le problème auquel fait face la FPLQ est un problème riche qui mérite d'être étudié plus en profondeur. Nous espérons que le présent rapport sera la point des départ d'autres recherches.

Références

1. J.-F. Cordeau, G. Laporte et A. Mercier, *A unified tabu search heuristic for vehicle routing problems with time windows*, Journal of the Operational Research Society **52** (2001), 928-936.
2. FPLQ, *Annexe F : Formule d'établissement des tarifs de transport pour les circuits de ramassage et de livraison du lait*, Convention de transport du lait, novembre 2006.
3. G. Laporte, *The traveling salesman problem : An overview of exact and approximate algorithms*, European Journal of Operational Research **40** (1992), 1086-1094.

Recognized Maritime Picture: Geofeasibility Scores

Mark Coates

Coordinator, McGill University

Marie-Odette St-Hilaire

Industrial representative, Lockheed Martin Canada

Nando de Freitas

Professor, University of British Columbia

Francis Moreau

Student, Université de Montréal

Boris Oreshkin

Student, McGill University

Mary Pugh

Professor, University of Toronto

1. Introduction

Maritime security is a considerable challenge for Canada due to its extensive coastline (150,000 miles) and its vast area of responsibility (over 6.7 million square miles). There are 250 ports in Canada and between 1500 and 2000 reported ships in its area of responsibility [3]. It may be the case that many ships do not self-report because they are far from port and not covered by the vessel traffic management system.

A Recognized Maritime Picture (RMP) is a plot compiled to depict maritime activity. The term “recognized” means that the picture has been evaluated prior to its distribution. A central authority compiles and fuses the data from many stations and evaluates their validity. The construction of an RMP consists of detecting objects of interest (ships), classifying them, assessing what they are doing, and deciding whether any type of follow-on action is required [9].

The Canadian Department of National Defense receives data from numerous sensor and information systems, operated by both civilian and military authorities. These systems include automatic self-reporting positional systems, air patrol surveillance systems, high frequency surface radars, electronic intelligence systems, radar space systems, and high frequency direction finding sensors [9]. One of the primary challenges of the Department is fusing the data from all of these sources, whilst trying to make full, accurate use of the information in a timely fashion.

Human operators are a key component of the current fusion system, but they are overloaded. There is insufficient automation in the system, resulting in inadequate fusion of the data, and the integration is relatively poor. The human operators are presented with too much information and must make too many decisions, a fact that significantly reduces the efficacy of the system. The result is that the initial version of the RMP is overloaded with multiple tracks, derived from different sources and sensor systems, that represent the same ship. The operators must decide whether two tracks are generated by the same ship, basing their decision on ship attributes and geographic considerations. It is highly desirable that this process be partially automated, so that the human operators have to make assessments about borderline cases only.

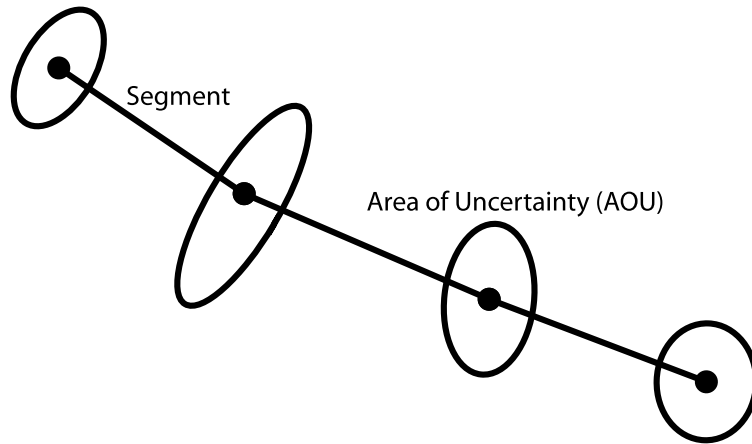


Figure 1. An example track showing the 2σ areas of uncertainty derived at each time instant

In each report, a source associates an area of uncertainty (AOU) of elliptic shape to a ship, delimiting a 2σ probability region (corresponding to a Gaussian distribution). Figure 1 depicts an example track. During the workshop, we considered the problem of reaching decisions based solely on geographic considerations. Our goal was to derive metrics that reflect the probability that two tracks are generated by the same ship (based solely on geographic considerations). A metric, or geofeasibility score g , must have a value comprised between zero and one, with a higher value representing a greater probability that the tracks are generated by the same ship. There are three basic constraints on the function g :

$$g(x) \in [0, 1], \quad (46)$$

$$g(x) = 0 \quad \text{if ellipses do not touch, and} \quad (47)$$

$$g(x) = 1 \quad \text{if ellipses overlap totally.} \quad (48)$$

In this report, we focus on the generation of a geofeasibility score based on a single, common time instant. This score is based solely on the ellipses and the underlying Gaussian distributions. In practice, multiple such geofeasibility scores would be combined (probably by using a harmonic or geometric mean) in order to derive a total score for the tracks.

We now discuss briefly the interest of Lockheed Martin Canada in the present work. The Research and Development Department of Lockheed Martin Canada was part of the DRDC Multi-Sensor Integration within a Common Operating Environment (MUSIC) Technical Demonstration Project [8]. As part of this project, a team of government and contract researchers developed a track fusion application that is expected to be fielded in the Canadian Navy's operation centers within the current year. Geofeasibility scoring plays an important part in the track-to-track fusion system developed within the MUSIC project for two main reasons: the score is evaluated a huge number of times (benchmark: 100,000 score evaluations per hour) and is visualized by operators on the RMP. Because of the second reason, the score must be intuitive enough for an operator to understand it, and therefore trust it. The descriptions of the track modeling approach and the track-to-track fusion process can be found in [10] and [5], respectively.

2. Proposed Solutions

2.1. SELECTED METRIC

We select as a metric the normalized overlap area between two ellipses, which can be defined as follows:

$$g = \frac{\int \mathbf{1}_{\{\mathbf{x} \in \mathcal{S}_1 \cap \mathcal{S}_2\}} d\mathbf{x}}{\int \mathbf{1}_{\{\mathbf{x} \in \mathcal{S}_{\min}\}} d\mathbf{x}}, \quad (49)$$

where the numerator is the area of the intersection between the ellipses defined by regions \mathcal{S}_1 , \mathcal{S}_2 , and \mathcal{S}_{\min} is the ellipse having the smallest area:

$$\mathcal{S}_{\min} \triangleq \arg \min_{\mathcal{S}_1, \mathcal{S}_2} \left[\int \mathbf{1}_{\{\mathbf{x} \in \mathcal{S}_1\}} d\mathbf{x}, \int \mathbf{1}_{\{\mathbf{x} \in \mathcal{S}_2\}} d\mathbf{x} \right]. \quad (50)$$

Note that (49) satisfies the requirements for a metric.

2.2. MONTE-CARLO INTEGRATION

The direct calculation of (49) is not viable analytically because it is not possible to define the limits of integration in the numerator. The integration limits are defined by the intersection points of two ellipses and an analytical formula for the general case is not available. Some of the team members used Newton's method to try to compute the points of intersection between the two ellipses (see the presentation on the web site of the workshop). Here we discuss Monte-Carlo integration, a powerful tool that can be used to resolve problems of this kind. It is suitable for the computation of integrals having the following form:

$$I = \int \int_{\mathcal{S}} \int f(x_1, x_2, \dots, x_K) dx_1 dx_2 \dots dx_K, \quad (51)$$

where \mathcal{S} is the set of integration limits. Monte-Carlo integration is useful when the integral in (51) cannot be evaluated analytically for some reason, but $f(x_1, x_2, \dots, x_K)$ can be evaluated pointwise. Moreover, the set of integration limits can be defined by a set of rules or otherwise evaluated in an efficient way during the execution of a numerical calculation routine. The simplest Monte-Carlo algorithm samples N random vectors $\mathbf{x}^{(i)} = (x_1^i, x_2^i, \dots, x_K^i)^T$ (for $i = 1, \dots, N$) uniformly from the volume defined by \mathcal{S} and approximates the integral (51) by the sum:

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}^{(i)}). \quad (52)$$

The estimator (52) is unbiased [2]. Moreover, the variance of this estimator is shown to decay with the rate N [2], which means that (52) is a consistent estimator of (51) [7].

The direct application of the naive Monte-Carlo algorithm is limited by the fact that the integration volume \mathcal{S} itself is often hard to evaluate analytically. Thus a more general algorithm is commonly used. In this algorithm one defines a probability density $p(\mathbf{x})$ that supports the volume of interest \mathcal{S} and is easy to sample from. The random vectors $\mathbf{x}^{(i)}$ are now sampled from this density and the integral (51) can be easily represented as

$$I = \int \int \int_{-\infty}^{+\infty} \mathbf{1}_{\{\mathbf{x} \in \mathcal{S}\}} \frac{f(\mathbf{x})}{p(\mathbf{x})} p(\mathbf{x}) d\mathbf{x}, \quad (53)$$

where $\mathbf{1}_{\{\cdot\}}$ is the Riemann kernel. The application of the importance sampling [2] paradigm results in an estimator of the following form:

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{\mathbf{x}^{(i)} \in \mathcal{S}\}} \frac{f(\mathbf{x}^{(i)})}{p(\mathbf{x}^{(i)})}. \quad (54)$$

Thus we propose to estimate the value of the normalized overlap area metric (introduced above) by the following Monte–Carlo algorithm:

$$g_M = \frac{\sum_{i=1}^N \mathbf{1}_{\{\mathbf{x}^{(i)} \in \mathcal{S}_1 \cap \mathcal{S}_2\}} \frac{1}{p_{\mathcal{S}_{\min}}(\mathbf{x}^{(i)})}}{\sum_{i=1}^N \mathbf{1}_{\{\mathbf{x}^{(i)} \in \mathcal{S}_{\min}\}} \frac{1}{p_{\mathcal{S}_{\min}}(\mathbf{x}^{(i)})}}, \quad (55)$$

where the $\mathbf{x}^{(i)}$ are samples drawn from the Gaussian distribution $p_{\mathcal{S}_{\min}}(\mathbf{x})$ with parameters defined by the ellipse \mathcal{S}_{\min} having the smallest area. In general the 2σ ellipse and corresponding \mathcal{S} can be parameterized by its semi-axes r_x, r_y , bearing angle φ , and translation of the center x_0, y_0 . The mapping between the ellipse parameters and the parameters of the Gaussian distribution $p_{\mathcal{S}}(\mathbf{x}) \triangleq \mathcal{N}(\mu, \mathbf{R})$ consists of the following two parts. Firstly, the covariance \mathbf{R} is connected to r_x, r_y , and φ by the following:

$$\mathbf{U} = \begin{bmatrix} \sin(\varphi) & \cos(\varphi) \\ \cos(\varphi) & -\sin(\varphi) \end{bmatrix} \quad (56)$$

$$\mathbf{\Lambda} = \begin{bmatrix} \frac{r_x^2}{4} & 0 \\ 0 & \frac{r_y^2}{4} \end{bmatrix} \quad (57)$$

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T. \quad (58)$$

Secondly, μ is connected to x_0, y_0 by the following:

$$\mu_1 = x_0, \quad \mu_2 = y_0. \quad (59)$$

2.3. GENERALIZED LIKELIHOOD RATIO

In this section we describe an alternative metric. We assume that the source provides us with noisy measurements $\mathbf{y}_1, \mathbf{y}_2$ of the true target coordinates μ_1, μ_2 , and with Maximum Likelihood (ML) estimators of measurement covariance matrices $\hat{\mathbf{R}}_1, \hat{\mathbf{R}}_2$. We choose to construct a test statistic to discriminate between the two hypotheses:

$$H_1 : \mu_1 = \mu_2, \quad (60)$$

$$H_0 : \mu_1 \neq \mu_2. \quad (61)$$

It is known [7] that in this case a Uniformly Most Powerful test does not exist. However, we can resort to a suboptimum statistic that is called the Generalized Likelihood Ratio (GLR):

$$\Lambda(\mathbf{y}_1, \mathbf{y}_2) = \frac{\max_{\Theta} p(\mathbf{y}_1, \mathbf{y}_2 | \Theta, H_1)}{\max_{\Theta} p(\mathbf{y}_1, \mathbf{y}_2 | \Theta, H_0)}. \quad (62)$$

Here Θ stands for the unknown parameters $\mu_1, \mu_2, \mathbf{R}_1, \mathbf{R}_2$. Given the Gaussian and independence assumptions on the underlying distribution $p(\cdot)$, we have

$$p(\mathbf{y}_1, \mathbf{y}_2 | \Theta, H) = \frac{1}{(2\pi)^2 \sqrt{\det(\mathbf{R}_1) \det(\mathbf{R}_2)}} \exp \left[-\frac{1}{2} \left((\mathbf{y}_1 - \mu_1)^T \mathbf{R}_1^{-1} (\mathbf{y}_1 - \mu_1) + (\mathbf{y}_2 - \mu_2)^T \mathbf{R}_2^{-1} (\mathbf{y}_2 - \mu_2) \right) \right]. \quad (63)$$

We deduce immediately that the following equation holds under H_0 .

$$\max_{\Theta} p(\mathbf{y}_1, \mathbf{y}_2 | \Theta, H_0) = \frac{1}{(2\pi)^2 \sqrt{\det(\hat{\mathbf{R}}_1) \det(\hat{\mathbf{R}}_2)}} \quad (64)$$

On the other hand, to find the ML estimator of the mean under H_1 , we use $\mu \triangleq \mu_1 = \mu_2$ and equate the partial derivative to 0:

$$\frac{\partial}{\partial \mu} p(\mathbf{y}_1, \mathbf{y}_2 | \Theta, H_1) = \mathbf{R}_1^{-1} (\mathbf{y}_1 - \mu) + \mathbf{R}_2^{-1} (\mathbf{y}_2 - \mu) = 0. \quad (65)$$

This results in the following expression for the ML estimator of the mean.

$$\hat{\mu} = \left[\mathbf{R}_1^{-1} + \mathbf{R}_2^{-1} \right]^{-1} \left[\mathbf{R}_1^{-1} \mathbf{y}_1 + \mathbf{R}_2^{-1} \mathbf{y}_2 \right] \quad (66)$$

Substituting (64) and (66) into the likelihood ratio (62), we get the statistic of the form

$$\Lambda(\mathbf{y}_1, \mathbf{y}_2) = \exp \left(-\frac{1}{2} \Delta^T \hat{\mathbf{R}}_{\Delta}^{-1} \Delta \right), \quad (67)$$

where Δ is the difference between the measurements:

$$\Delta = \mathbf{y}_1 - \mathbf{y}_2, \quad (68)$$

and the estimate of the covariance inverse is given by the formula

$$\begin{aligned} \hat{\mathbf{R}}_{\Delta}^{-1} &= \hat{\mathbf{R}}_2^{-1} \left[\hat{\mathbf{R}}_1^{-1} + \hat{\mathbf{R}}_2^{-1} \right]^{-1} \hat{\mathbf{R}}_1^{-1} \left[\hat{\mathbf{R}}_1^{-1} + \hat{\mathbf{R}}_2^{-1} \right]^{-1} \hat{\mathbf{R}}_2^{-1} \\ &\quad + \hat{\mathbf{R}}_1^{-1} \left[\hat{\mathbf{R}}_1^{-1} + \hat{\mathbf{R}}_2^{-1} \right]^{-1} \hat{\mathbf{R}}_2^{-1} \left[\hat{\mathbf{R}}_1^{-1} + \hat{\mathbf{R}}_2^{-1} \right]^{-1} \hat{\mathbf{R}}_1^{-1}. \end{aligned} \quad (69)$$

It is straightforward to show, by using the rule that the inverse of the product is equal to the product of the inverses (in reverse order), that the following holds:

$$\hat{\mathbf{R}}_{\Delta}^{-1} = \left[\hat{\mathbf{R}}_1 + \hat{\mathbf{R}}_2 \right]^{-1}. \quad (70)$$

Hence the geofeasibility score based on GLR admits the following simple and intuitive form:

$$g_G = \exp \left(-\frac{1}{2} \Delta^T \left[\hat{\mathbf{R}}_1 + \hat{\mathbf{R}}_2 \right]^{-1} \Delta \right). \quad (71)$$

Note that (71) does not satisfy the requirements for a metric. In Section 3, however, it will be argued that in most cases this issue can be resolved by using simple thresholding.

It is interesting to observe that (71) is closely related to the Mahalanobis distance, which is defined as follows [6]:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{Q}^{-1} (\mathbf{x} - \mathbf{y})}, \quad (72)$$

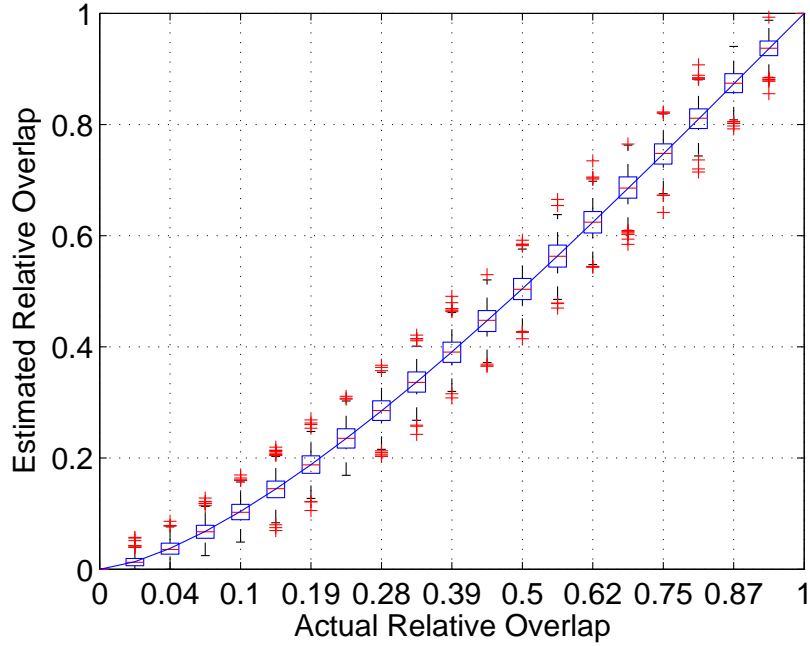


Figure 2. A comparison between the actual relative overlap and the estimated relative overlap using Monte-Carlo integration with number of samples $N = 500$. The line depicts the true value; the error bars are derived from 1000 trials of the Monte-Carlo integration technique. The scenario used to generate these results is the special case of overlapping circles, which is the only case where an analytical expression for the overlap can be derived.

and where it is assumed that \mathbf{x} and \mathbf{y} are jointly Gaussian random vectors drawn from the same distribution with covariance \mathbf{Q} . In our case \mathbf{Q} is unknown and we outline a statistical procedure (see (70)) that is used to estimate $\mathbf{Q} \triangleq \widehat{\mathbf{R}}_{\Delta}$ from the available data $\widehat{\mathbf{R}}_1$ and $\widehat{\mathbf{R}}_2$.

3. Simulation Results

In this section we present the simulation results showing the performance of the proposed solutions. First, the error bars for the Monte-Carlo integration method are shown in Fig. 2. Figure 2 indicates that reasonable error values can be obtained with a number of Monte-Carlo samples equal to 500. Fig. 3 shows the dependence of the computing time on the number of Monte-Carlo samples used; the proposed algorithm, with 500 samples, requires slightly more than 1 ms per normalized overlap area evaluation. The simpler GLR statistic requires 0.04 ms per evaluation. Thus the latter looks more attractive from a computational point of view.

The next set of figures explains the behaviour of the Monte-Carlo estimator in the situations when the approximated normalized overlap area and sample size vary. In particular, Fig. 4 shows how Relative Mean Squared Error (RMSE) scales with the normalized area overlap for a fixed Monte-Carlo sample size. Because in this case we consider the ratio of absolute MSE to the square of the true value, the relative error increases as the normalized area of overlap decreases. This implies that the absolute error is approximately constant (see also Fig. 2) and it is stable in the range of interest. To generate this figure we compared the value estimated via Monte-Carlo integration to the normalized overlap area that can be analytically derived for the case when ellipses simplify to circles. Figure 5 shows how MSE scales with Monte-Carlo sample size N for a

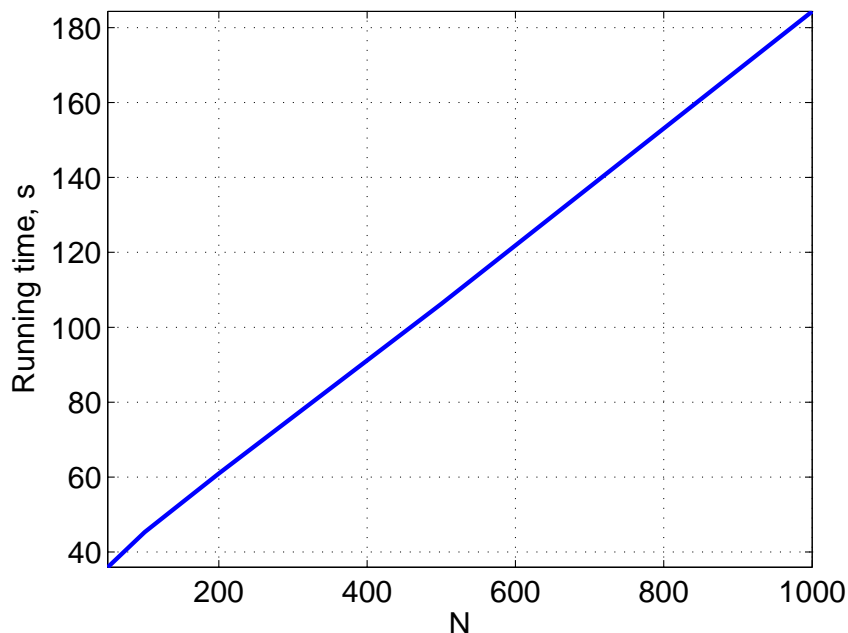


Figure 3. Computing time, as a function of the number of samples, N , for 100,000 evaluations of the normalized overlap area using the Monte–Carlo algorithm.

fixed overlap area. As expected, the approximation error can be easily controlled by changing the sample size. The doubling of the sample size leads to a 3 dB decrease in the MSE error. Thus the performance of the Monte–Carlo method is a trade-off between the approximation error introduced (see Fig. 5) and the computational effort involved (see Fig. 3).

The last plot, Fig. 6, shows the comparison between the statistics we studied. In particular, in this plot, the thick blue (dark grey) line corresponds to the Monte–Carlo approximation of the normalized overlap area g_M defined in (55), the black line shows the values of the GLR statistic g_G defined in (71), the red (light grey) line corresponds to the Monte–Carlo evaluation of the product of the two Gaussian distributions, and the green (lightest grey, horizontal) line shows the threshold setting at 0.5. In this figure we use the Monte–Carlo estimator with $N = 1000$ as a reference and examine how well this reference can be approximated by the simpler GLR statistic throughout 10,000 random trials. In each trial a random configuration of two ellipses is generated. We can see that the GLR never reaches 0 or 1, even if there is complete overlap or no overlap at all between the ellipses. On the other hand, it is clear that by applying proper thresholding we can force g_G to zero or one in these regions. Values of the normalized overlap area between 0 and 1 are approximated by the GLR statistic reasonably well, given the modest computational effort involved in the computation of g_G . Moreover, in general, g_G tends to overestimate the overlap area in the region $[0, 0.7]$.

We know from [9] that acceptance of false track pairs in this particular problem is not penalized as much as the miss of a true, but distant track pair. Thus as an approximation of operator intuition g_G exhibits desirable properties in the region $[0, 0.7]$. On the other hand, in the region $[0.7, 1]$, g_G underestimates the true normalized overlap area, but it can be seen clearly that it is always greater than the threshold 0.5. From a psychological perspective [9], this threshold is the best borderline between a track pair that is certainly a false match and a pair that should be

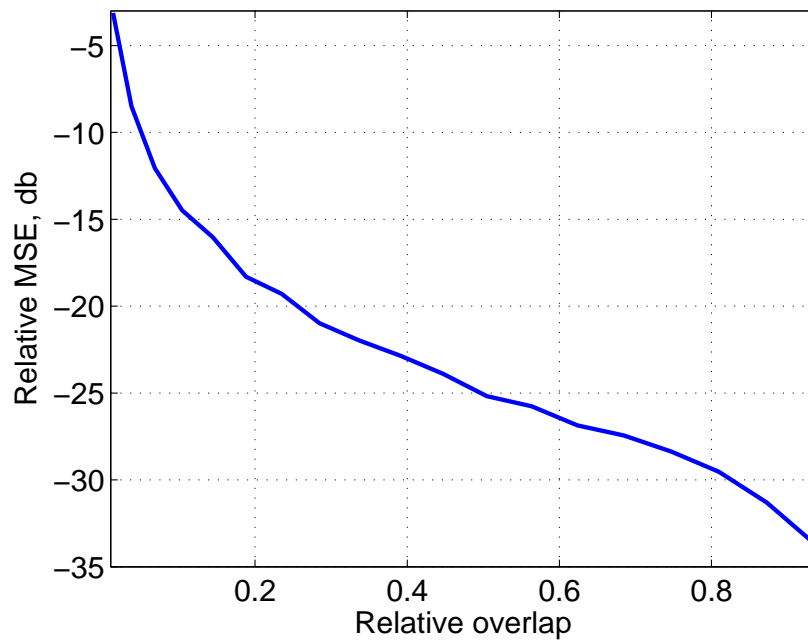


Figure 4. The relative mean-squared error (RMSE) of the Monte–Carlo integration algorithm for the case $N = 500$. The absolute mean squared error is the average squared difference between the estimate and the true value. The RMSE is the ratio between this value and the square of the true overlap area. The figure indicates that the RMSE increases as the overlap area decreases, which reflects the desirable property that the absolute mean-squared error remains approximately constant.

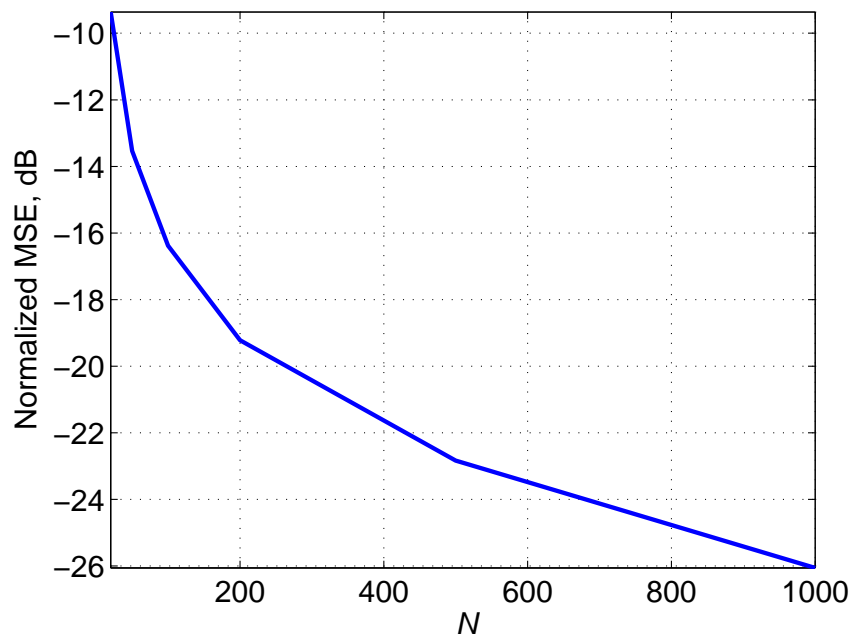


Figure 5. The Monte–Carlo error as a function of the number of samples, for a fixed value of the relative overlap area equal to 0.391.

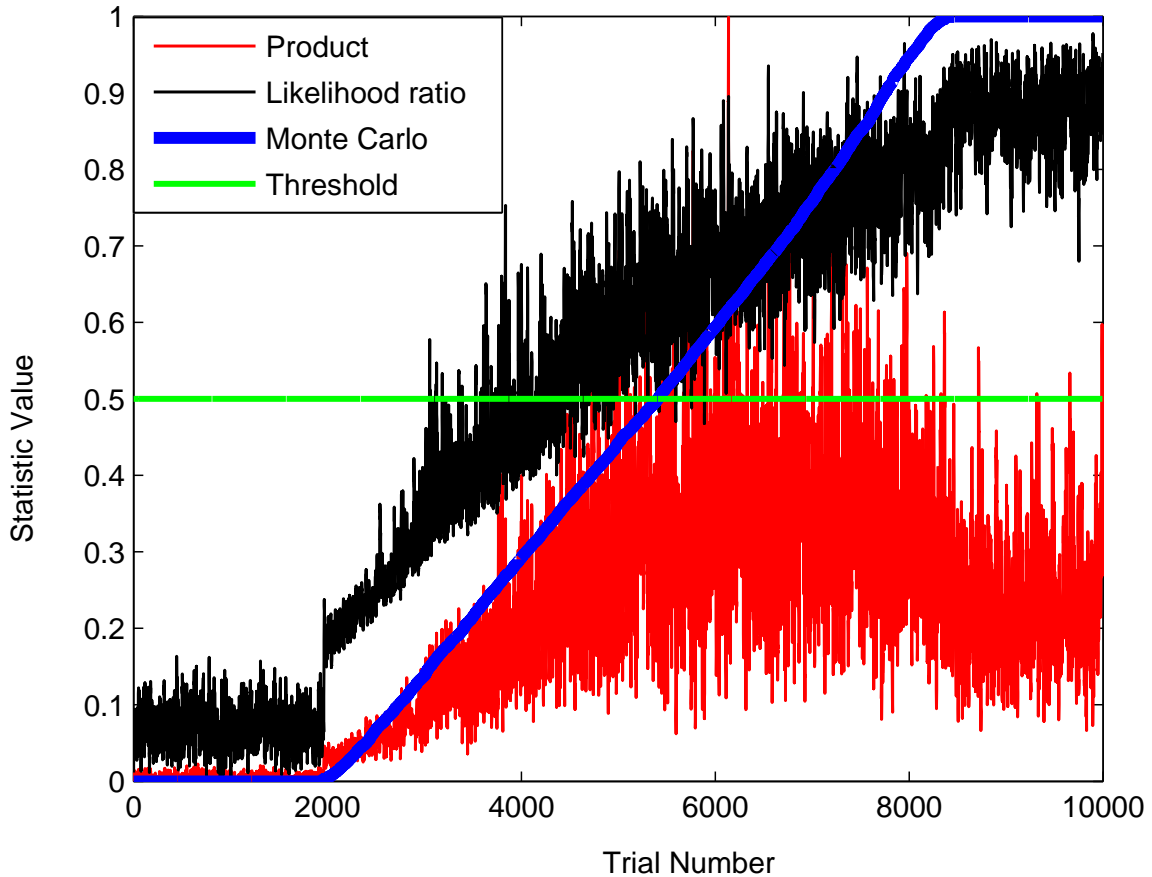


Figure 6. A comparison of statistics over ten thousand trials, each consisting of two random ellipses with varying size, orientation, and overlap. In the figure, trials are sorted in order of the Monte–Carlo estimation of the overlap. The primary comparison is between the Monte–Carlo estimate of the area of overlap and the likelihood ratio statistic. The likelihood ratio is generally slightly larger in magnitude than the Monte–Carlo estimate for relative overlaps between 0 and 0.7, and drops below the estimate for relative overlaps in the range 0.7 to 1.

considered as a potential match. This is also a desirable property. The obvious drawback of the g_G statistic is that a larger number of track pairs are considered as potential matches as compared with the Monte–Carlo statistic g_M . In turn this increases the load on the operator who needs to make a final decision regarding the track matching.

4. Conclusions

In this report we focused on the generation of two alternative metrics that can be used to evaluate a geofeasibility score of two tracks. We considered the situation where the only information available for feasibility assessment is geographical information consisting of two 2σ ellipses representing the corresponding target position measurements and their uncertainty. The first metric is based on direct evaluation of the normalized intersection area between the two ellipses, based on Monte–Carlo integration. The second metric is based on the GLR statistic under the Gaussian assumption

on measurement uncertainty. The experiments conducted on synthetic data showed the viability of both metrics in terms of the computational burden required to perform numerical calculations and obtain accurate results.

The accuracy and computational resources required for the implementation of the Monte–Carlo based geofeasibility score are easily controlled by varying the Monte–Carlo sample size. Reasonably accurate and fast results can be achieved for sample sizes of 500, requiring slightly more than 100 seconds for 100,000 geofeasibility score evaluations. The GLR based geofeasibility score approximates the normalized intersection area metric reasonably well on average. The computational effort involved in the calculation of the GLR based score is significantly less than that of the Monte–Carlo based score. It amounts to approximately 4 seconds per 100,000 evaluations. The choice between the two proposed metrics depends on the available computational resources, the accuracy required by the application, and the closeness of the metrics to the desired track match indicator.

Although the approach described in this report is, from a mathematical point of view, more satisfying than the geometrical approach, it might not be ideal from a psychological point of view. Recall that the geometrical approach consists of computing the area of intersection of two ellipses. One of the major factors underlying the utilization of a complex and high–criticality system, such as the MUSIC track-to-track fusion system, is the level of trust the operator has in it [4]. Operators may be unwilling to use a reliable automated system leading to high-precision results if they consider it as untrustworthy. An automated system will be trusted to a greater extent if the operators understand the algorithms underlying the system [1]. Maritime surveillance operators perceive AOU as simple two-dimensional ellipses, not as Gaussian distributions, which most of them do not understand. Since operators have to interpret the results, the geofeasibility scoring chosen for the track-to-track fusion system developed within the MUSIC project is a linear approximation of the normalized area of intersection of two ellipses (see the details in [5]). Although this approach may seem too simple from a statistical point of view, it is based on the operator’s perception, arising from many years of field experience.

References

1. H. Atoyan, J.-R. Duquet, and J.-M. Robert, *Trust in new decision aid systems*, 2006, pp. 115 – 122.
2. A. Doucet, N. de Freitas, and N. Gordon (eds.), *Sequential Monte Carlo Methods in Practice*, Springer–Verlag, Berlin, 2001.
3. Capt. Laurence Hickey, *Background brief — the recognized maritime picture*, Presentation to Senate Committee for National Security and Defence, Canada, Sept./Nov. 2003, available at <http://www.parl.gc.ca/37/2/parlbus/commbus/senate/Com-e/defe-e/witn-e/hickey2-e.htm>.
4. J.D. Lee and K.A. See, *Trust in automation: Designing for appropriate reliance*, **46** (2004), 50–80.
5. E. Lefebvre, C. Helleur, and N. Kashyap, *The use of a multidimensional space for fusion candidate representation in a maritime domain awareness application*, vol. 6974, March 2008.
6. P. C. Mahalanobis, *On the generalised distance in statistics*, Proceedings of the National Institute of Science of India **12** (1936), 49–55.
7. V. H. Poor, *An introduction to signal detection and estimation*, 2 ed., Springer-Verlag, Berlin, 1994.
8. Defense Research and Development Canada, *Multi-Sensor Integration within a Common Operating Environment (MUSIC), Technology Demonstration (TD)*, May 2006.
9. M.-O. St-Hilaire, *Recognized maritime picture — Montreal Industrial Problem Solving Workshop (IPSW) problem statement*, personal communication, August 2007.
10. M.-O. St-Hilaire, E. Lefebvre, and C. Helleur, *Track modeling for maritime surveillance*, Proceedings of 11th International Conference on Information Fusion, July 2008.