[[1]](#footnote-1)

An automatic detection method of oligoclonal bands on gel electrophoresis of cerebrospinal fluid

S.Boudet, Z. Wang, L.Peyrodie, and G.Forzy\*

*Abstract*— The detection of oligoclonal bands on the cerebrospinal fluid (CSF) electrophorisis is an important diagnostic element for Multiple Sclerosis (MS). The profile images being often highly artifacted and having a very low contrast, the interpretation is often difficult. An automated method of band detection to ease analyze and reduce subjectivity is presented. The method is broken down on multiple steps: convert color image to a grayscale image, realign the profils, remove the artifacts, convert the image 2 D to a signal 1D, detect and thresholding the peaks and remove peak which are seen outside the profile. The results of the method on 36 profiles (21 positive and 15 negative) are compared to the blind analysis of an expert biologist. Considering the expert analyze as a ground truth, the method reaches a sensibility of 0.76 and a specificity of 0.80 which is promising view to the difficulty of the problem.

# INTRODUCTION

In France, Multiple Sclerosis (MS) affects approximately 1 person in 1000. It is the first non-traumatic cause of severe disability in young patients acquired. It is characterized by inflammatory lesions of the white matter of the central nervous system (CNS), spread in space and time. The diagnosis is based on the discovery of 3 of the 4 criteria Barkhof brain and spinal cord on MRI, or on a combination of 2 lesions suggestive MRI associated with a positive cerebrospinal fluid analysis (CSF) [1]. For the cerebrospinal criteria, Freedman and al [2] recommends the determination of immunoglobulin-gamma (IgG) immunoblotting oligoclonal bands in cerebrospinal fluid (CSF) thanks to isoelectric focusing (IEF). The electrophoresis is a method that separate molecules on the basis of their size, electric charge and other physical properties**.**

**Les IgG sont des anticorps produits dans le LCR des patients. Ces anticorps sont multiples et variables pour chaque patient conduisant à des profils electrophorétiques eux aussi très variables. Le profil des IgG est révélé par immunoblot avec un chromogène, produisant un renforcement de coloration là où il y a des protéines en laissant un bruit de fond sur toute la zone de migration des IgG. A l’issue de la migration, ce marquage permet de repérer les immunoglobulines G présentes dans le LCS sous formes de bandes sombres. L’image résultante est issue de la superposition du colorant, des bruits additionnels et des protéines contenues dans le LCR. L’intensité de la coloration est proportionnelle à la concentration locale d’IgG. On appelle ligne de base l’interaction du colorant avec le support et cette ligne de base peut être considérée comme un bruit de fond. Nous pourrons alors observer des bandes de fortes ou faibles intensités noyées dans ce bruit de fond. Pour interpréter, le biologiste compte les bandes d’IgG qu’il identifie par leur aspect fin et homogène**. Profil is then said oligoclonal if there are at least three IgG bands in CSF.

**L’enjeu méthodologique est alors triple : repérer ces bandes de faible intensité, isoler les artéfacts susceptibles de polluer l’isofocalisation et ainsi pouvoir compter sans ambigüité le nombre de bande repérable sur le profil.** To reduce the risk of subjectivity and ease the analysis, it would be interesting to automate the process by computing.

This automated interpretation poses several difficulties:

- low signal noise ratio : bands have often a very low amplitude;

- several image artifacts: black and white stain called salt and pepper noise;

- the blur: which create false lanes which can be seen on the entire membrane

- smiles: horizontal lane deformation

- Important baseline variation on a profile

- several inter-individual differences on baseline level

This paper aims to propose a multiple step method to automatically detect the oligoclonal bands and to evaluate the concordance with an expert biologist.

# State of art

Analyze of DNA is the major field of band detection in electrophoresis. Several methods [3-5] and commercial interfaces exist. They all propose a preprocessing of the gel image captured by camera. Geometry of the whole image is rectified and filtered. Then, each lane is separated individually by image segmentation. Because of the quality of image, automatic filtering and rectification could induce some mistake in peaks detection, some existing bands could disappear and some false detection could appear. The parameters to detect gel bands have to be adapted specifically. Some semi-automatic techniques have been develop but manual settings are still needed such as grayscale threshold selection for a Region of Interest (ROI) or for manually removing specific artifacts.

Use of mixtures models like Gaussian Mixtures [6] or Wavelet transforms [7] are commonly used to fine tune the quantification of bands on DNA. Artifact removing is still a challenge in DNA band detection in addition to the high computation time.

Compared to DNA electrophoresis, the problem of IgG isofocalisation does not require a fine band location detection, but only to count bands. However, the contrast is much lower and the band amplitude can reach the limit of eye perception abilities. At this threshold, any small artifact perturbation can cause a false detection. For thoses reason, the method to process DNA electrophoresis cannot be applied directly on IgG isofocalisation without adapting methods and parameters. We did not find actual reference on literature [8] which deals about this problem. The following section present an automated method inspired from DNA processes but adapted to isofocalisation.

# Band Detection process

This section present the various step for automated band dection on IgG isofocalisation of CSF. The Figs. 1 and 2 illustrate those different steps on a high contrasted positive profile with some artifacts and smile distortion.

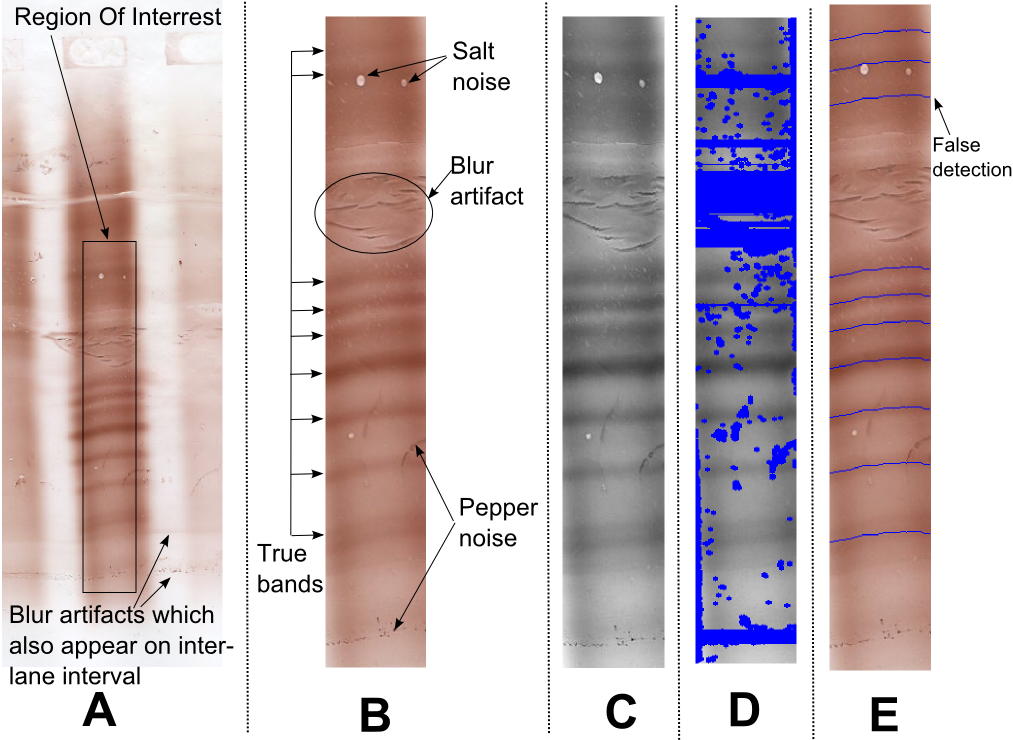


Figure 1.Illustration of image processing steps of band detection. – A. profile image on membrane – B. Annotated profile image of ROI – C. results of grayscale conversion and re-alignment steps – D. noise removing E. final detected bands reported on original image.

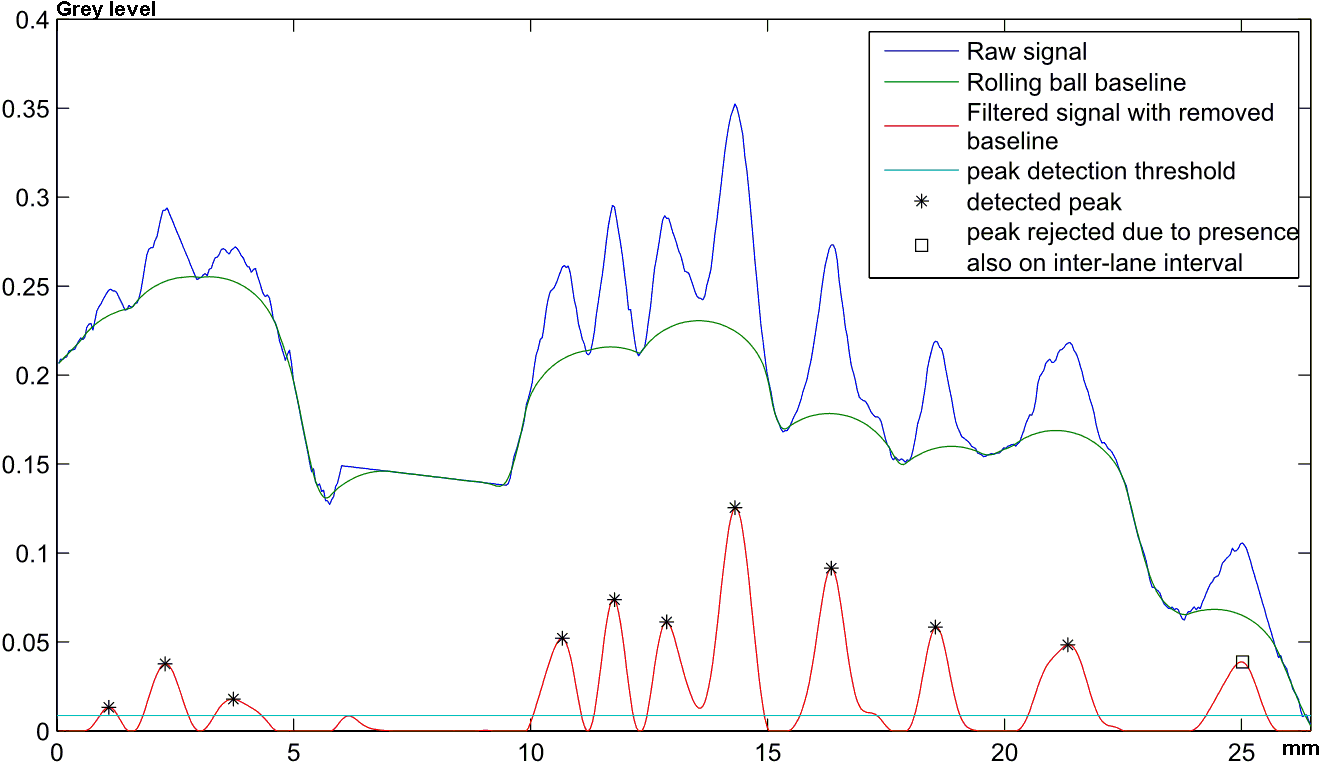


Figure 2. Result of signal processing steps of band detection.

## Electrophorese

**Pour séparer les immunoglobulines G, nous utilisons la technique Helena d’isofocalisation sur gel d’agarose et de transfert sur une membrane de dimensions 10cm\*8cm (Helena Biosciences, United Kingdom)**.

The membrane image contains ten profiles of various patients recorded simultaneously.

## B. Profile image acquisition

The physical membrane is scanned [Epson V750 PRO] in 48 bit color (the 3 color values are considered from 0 to 1) at a resolution of 600 dpi. A specialist has to choose a profile and select manually a Rectangular Region of Interest (ROI) thanks to a specific Matlab interface (Fig. 1.A-B).

## C. Grayscale conversion

The color image is converted to grayscale thanks to a linear conversion (Gray=R\*Red+G\*Green+B\*Blue). In order to find the coefficients which maximize the contrast, a Principal Component Analysis (PCA) has been realized on some positive profiles (considering all pixels as samples and the three colors as data). The various coefficient obtained on the various profiles are normalized and averaged. The final coefficients thus obtained for all images are R=0.16, G=0.52, B=0.32 (Fig. 1.C). This is coherent with the fact that bands are most visible on the green image with our acquisition system.

## D. Band re-alignement

Some images are subject to distortion called smiles which causes band not being straight and not being horizontal. A re-alignment process is then applied which consist to determine an offset for each column of pixel. This set of offset make a curve) and this curve is determined to maximize correlation between each column of pixels while being smooth and with minimum distortion.

Let being the image, being the signal of the k column of pixel. Eq.1 designs a cost function which has to be minimized by an optimal.



corresponds to the norm 1 of the signal (sum of absolute value), is the second derivative of t, is the squared correlation between the two signals, is the i-th column of the image shifted by pixels, and are balancing terms set empirically.

This cost is composed by three balanced terms:

- (sum of absolute value of the curve) aims to minimize the change if there is no significant distortion on the image;

- aims to minimize the second derivation of the curve in order to avoid discontinuities while allowing straight but not horizontal shifting.

- is the negative of the sum of the correlation of each pair of the shifted pixels column.

The optimization process consists on a greedy algorithm: each column of pixels is taken individually in a random order, the optimal shifting of this column is determine by testing all values on a range taking into account the current shifting of other columns. This process is repeated several times so that each column is optimized 4 times. This process is programmed in C and it takes a neglecting computation time (<0.5s by profile). A result is illustrated on Fig. 1-C.

## E. Cleaning salt and pepper noise

This part aims to reject those artifacted parts on the image in order to get a clean signal when averaging horizontally.

For each pixel of grey value , we compute the average of an ellipse of 10 pixel wide and 5 pixel height centered on If the absolute difference between and is over a threshold, the pixel is considered as artifact as well as the other pixels around on a radius of 3 pixels.

The ellipse has a larger width than height to avoid rejecting contrasted horizontal bands. This rejection process is set to be highly sensible and poorly specific since the information of a band can be retrieve on all the width of the profile.

At the end, the rows of pixels which have more than 50% of artifacted pixels are completely rejected (Fig. 1-D).

## F. Convert 2D signal to mono-dimensional signal

Once image (2D signal) has been re-aligned and cleaned for artifacts, the mono-dime nsional signals can be simply obtained by averaging each row of the profile image. There can be missing values on the signal due to rejected rows of the previous step. If there are less than 15 pixels (0.65mm on membrane) between two missing values, the signal between those values is also considered as unreliable and rejected. At the end, the blocks of missing values are filled with a linear interpolation on the borders of the blocks (Fig. 2 blue curve).

Once the signal obtained, the band detection consists in detecting peaks on a signal.

## G. Filtering and baseline removal

To remove the small noise and to have smooth curve with reliable local extremum, a low pass filter is applied on the signal. The filter used is a 4-th order Butterworth at a cutoff period of 0.67mm (16 pixels) with a forward-backward process to avoid phase-shifting.

Then, the signal baseline is removed thanks to a rolling ball algorithm (Fig. 2 red curve). The radius of the ball is set to 1.48mm on distance and 0.041 on grey level (=10.5 at 256 level). This rolling ball has three interests: ease the amplitude computation of a peak, amplifies peak which are on slopped baseline, and remove the peaks which are two wide to correspond to a lane.

## H. Detect peak and threshold setting

After filtering and baseline removal, the peak signals are clean and the set of peaks can be obtained by taking the local maxima of signal which are over a threshold. This threshold is set to 0.0087 (=2.21 at 256 level), which is very close to the human eye perception limit (Fig. 2 cyan curve).

It is possible that two close local maxima appear above the threshold without returning to the baseline. It can either correspond to two bands which are very close or just a shape effect. On those cases, if the minimum between those two peaks is over 2/3 of the lower of the two peaks, this lowest peak is then removed.

## I. Removing peaks which are also detected on inter-lane interval

There is another type of artifact called blur which can have exactly the same shape than a band and produce false detection. Most of time, those artifacts appears on the white inter-lane interval of the membrane each side of the profile. It is then possible to apply the same detection process (step F, G H) on those inter-lane intervals to detect the possible artifacts. The images taken correspond to an 11-pixels wide rectangle in the middle of the inter-lane interval each side of the profile. If a band is detected either on the left or on the right, all bands in the profile within 15-pixels range are then considered as possible blur artifact and removed (Fig. 2 square point).

# Evaluation

## A Method

To evaluate the method we have realized a blind comparison with the analysis of an expert biologist. Dans un premier temps, le caractère oligoclonal ou non des profils est défini de façon consensuel par un groupe d’experts du laboratoire du GHICL, directement sur la membrane (before scanning). Notre base de données comprend 36 profils, 21 positifs et 15 négatifs, analysés sur 5 membranes différentes.

In second step, the biologist selects the ROI on the scanned profile and locates precisely the bands by click thanks to a specific Matlab interface.

The profile is then analyzed by the described method and the two analyses are then compared. There is a matching between bands positioned by the expert and by the computer if there are less than 15 pixels (0.65mm on membrane) between them. In case of several bands can match, the matching is established on the nearest band.

## B Result

The figures 3 illustrate the results on 5 profiles.

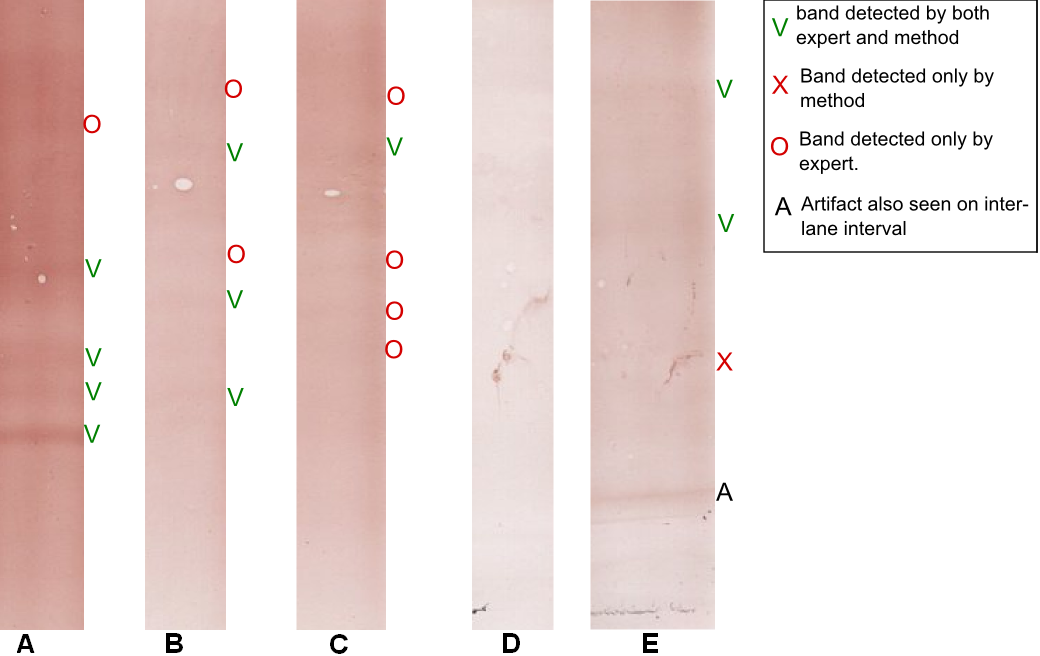


Figure 3. Result of band detection on 5 profiles: A, B & C. are positive;   
D & E are negative.

The profile A is judged positive by expert due to the presence of 5 bands. The contrast is relatively important so only one band is missed by the method. Profiles B and C are positive profile with very low contrast. They show the difficulty of the problem for both human and computer analysis. B is well detected positive with 3 bands detected on 5 but C is detected negative with 1 band detected on 5. Profile D and E are negative profiles. On D, the method does not detect any band as well as the expert. On E, a false detection due to an artifact residue makes the profile wrongly detected as positive.

Considering the expert analyze as a ground truth, we mean by “well classified (vs misclassified)” an agreement (vs disagreement) between expert and automatic analysis. The global results are summarized in table one.

1. Results of band detection

| Profile | Well classified | Misclassified | Matching bands | Forgotten bands | Added bands |
| --- | --- | --- | --- | --- | --- |
| Positive  (21) | 16 | 5 | 103 | 37 | 19 |
| Negative (15) | 12 | 3 | 7 | 9 | 17 |
| Total (36) | 28 | 8 | 110 | 46 | 36 |

The method reaches a sensibility of 0.76 and a specificity of 0.80 for the entire profile classification. For the band detection, the method reaches a sensibility of 0.67 and a Positive Predictive Value (PPV) of 0.75. The pixel specificity (eq. 2) reaches 0.91 and the pixel Negative Predictive Value (eq. 3) 0.89.

(2)

(3)

## C Discussion

At this step of our works (proof of concept), the results are encouraging in view to the difficulties of the problem. By looking for the a posteriori results of each step, it seems that each step is efficient enough but there is still a small loss of information which cumulated reduce the classification rate:

- step B: even if scanning enables the user to zoom and to maximize the contrast, the bands are often more visible on membrane. Moreover, the step of ROI definition could be automated but it did not appear as a priority for biologists.

- step D: the band re-alignement worked on all cases on our study. However, we know that there are some uncommon cases where the distortion of each band is not constant on the profile and the method is not designed for that.

- step E: the noise removing method seems efficient, and it did never remove a true peak on our study. However, the residue on border of important artifacts often causes false detection.

- step F: the converted mono-dimensional signal seems most often conform to the image but there is some cases where double-peaks are merged and some cases where the contrast seems less important on signal than on image. The reason is not indentified yet.

- step G&H: low-pass filtering, baseline removal and thresholding should be the steps with the greatest improving capacities. We have observed that most of forgotten band can be seen on curve but have the amplitude under the threshold. Unfortunately, lowering the threshold will add false detections by every small artifact. Future research is envisaged on using mixed methods analysis like GMM.

- step I: removing peaks which are also detected on inter lane interval remove some false detection. This process could be generalized for bands which appear on every profiles of the membrane.

# Conclusion

Confrontation between ground truth and automated analysis gives promising results, considering the difficulties of the problem. We hope that this work could be used as basis for future research to optimize independently each step. Particularly, our future work will focus on the reliability of the detection of peaks by methods based on mixture models.

We hope also testing the method on a more important variety of profile and to compare the results with the inter-expert and intra-expert variability in order to have a better idea of true discordances and the questionable ones.

Ce type de travail permettra peut-être de définir une nouvelle sémiologie de lecture de l’électrophorèse.

Acknowledgment

We thanks Christophe Herlin for is technical assistance Doctor David Perry for English review

References

1. Polman and al. “Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria”, *Ann Neurol*, 69(2), Feb 2011; pp 292–302.
2. Freedman and al. “Recommended standard of cerebrospinal fluid analysis in the diagnosis of multiple sclerosis: a consensus statement,” in *Arch Neurol*, vol 6, june*,*2005, pp865-70.
3. N. Kaabouch and R Schultz,"A 2-D gel electrophoresis DNA image analysis algorithm with automatic thresholding," *The international Society for Optical Engineering Proceedings (SPIE)*, Vol. 6508,2007, pp. 65081H-I to 65081H-12,
4. A PSU, A Bhornra, "Automatic analysis of agarose gel images," *Bioinformatics* ,vol 17(11), 2001,pp. 1084-1090,.
5. I Bajla, L Hollander, S. Fluch, K Burg, Kollar, "An alternative method for electrophoretic gel image analysis in the Ge!Master software," *Computer Methods and Programs in Biomedicine* , vol 77, 2005, pp. 209 -231,.
6. Polanski A, Marczyk M, Pietrowska M, Widlak P, Polanska J (2015) Signal Partitioning Algorithm for Highly Efficient Gaussian Mixture Modeling in Mass Spectrometry. PLoS ONE 10(7): e0134256. doi:10.1371/journal.pone.0134256.
7. B. Moreira, A Sousa, A Mendonça, and A Campilho, “Automatic Lane Segmentation in TLC Images Using the Continuous Wavelet Transform,” Computational and Mathematical Methods in Medicine, vol. 2013, 19 pages. doi:10.1155/2013/218415.
8. A. Gillain, A. Fumal, C.Nève, J.-M. Minon , A. Maertens de Noordhout,’Oligoclonal bands and IgG index interpreted according to Reiber in the inflammatory diseases of the central nervous system’ Immunoanalyse &amp; Biologie Spécialisée, vol 24(3), 2009,pp 135-

1. \*Research supported by PHRC POLAR.

   S.Boudet. is with FMM-UTSB, 56 rue du Port, 59000 Lille (France) (corresponding author to provide phone: 00330328384858 fax: 00330328384838; e-mail: samuel.boudet@gmail.fr).

   Z. Wang., is with FMM-UTSB, (e-mail: zefeng.wang@hotmail.fr).

   L.Peyrodie is with UTSB-HEI, 13 rue de Toul, 59000, Lille (e-mail: [Laurent.peyrodie@hei.fr](mailto:Laurent.peyrodie@hei.fr)).

   G. Forzy is with FMM and GHICL, rue du Grand But, BP 249, 59642 Lomme, (France). [↑](#footnote-ref-1)