

N-745

ANALYSE AUTOMATIQUE DE GELS
D'ELECTROPHORESE BIDIMENSIONNELLE
ET MORPHOLOGIE MATHEMATIQUE

S. BEUCHER

FONTAINEBLEAU

FEVRIER 1982

*ANALYSE AUTOMATIQUE DE GELS D'ELECTROPHORESE
BIDIMENSIONNELLE ET MORPHOLOGIE MATHÉMATIQUE*

RAPPORT DE FAISABILITE

INTRODUCTION

L'électrophorèse bidimensionnelle est une technique de séparation de protéines en fonction de leur point isoélectrique d'une part et de leur poids moléculaire d'autre part. Cette technique très puissante permet de disposer pour un organisme d'un grand nombre de marqueurs. Plusieurs systèmes de révélation des gels peuvent être utilisés. L'information fournie par un gel a un double aspect. C'est d'abord une information de position, puisque l'identification des protéines est obtenue par calcul des coordonnées des taches sur le gel. Ensuite, l'étendue des taches et leur intensité sont des paramètres importants car directement liés à la quantité et l'activité de la protéine.

Le Centre de Géostatistique et de Morphologie Mathématique a entrepris une étude de faisabilité de la lecture automatique des gels d'électrophorèse à partir d'autoradiographies fournies par le laboratoire d'amélioration des plantes, INRA à Orsay. Le présent rapport a pour objet de commenter les résultats obtenus et d'explicitier la contribution de la Morphologie Mathématique à la résolution des problèmes posés. Ce rapport comprend deux parties : la première illustre quelques procédés utilisés pour résoudre les différents problèmes posés. La seconde partie traite des questions qui n'ont pas encore été abordées ou résolues et des perspectives offertes par la morphologie mathématique dans ces domaines. Dans un souci de simplicité et de lisibilité, ce rapport ne s'embarrasse pas de considérations techniques complexes. En particulier, on ne donnera pas de description détaillée des algorithmes utilisés. Le lecteur est prié de se reporter à la documentation existante.

ANALYSE AUTOMATIQUE DES GELS D'ELECTROPHORESE

Avertissement préliminaire

Les méthodes et algorithmes qui seront présentés dans la suite de cette étude ont été testés sur un analyseur de textures Leitz-TAS. Cet

appareil ne disposant pas, dans la configuration utilisée, de mémoires numériques d'images, les traitements ont été effectués sans stockage préalable des images analysées. Cette précision est d'importance si l'on désire apprécier correctement la qualité des résultats. En effet, cette façon de faire introduit un bruit non négligeable tout au long du traitement.

Détection des taches

La détection des taches et le calcul de leur coordonnées constituent la première étape du traitement.

La figure 1 représente une vue grossie d'une autoradiographie de gel. Les taches y apparaissent comme des objets convexes de taille et d'intensité variable. Certaines taches ont tendance à se recouvrir partiellement.

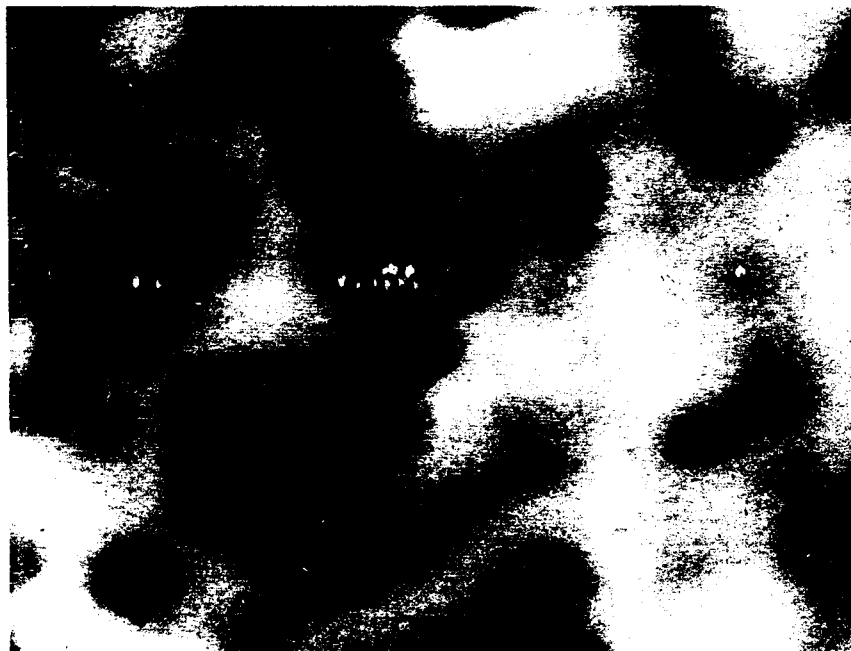


Figure 1 : Gel d'électrophorèse. Vue grossie.

Le graphe de la fonction $f(x)$ des différentes valeurs de gris en chaque point x peut être représenté dans l'espace comme une surface topographique. Plus la valeur de gris au point x est sombre, plus élevé sera le point correspondant à cette valeur sur le graphe. La Figure 2 représente l'aspect de cette surface topographique.

L'image (b) visualise en perspective l'ensemble des niveaux de gris présents dans le champ d'analyse visible sur l'image (a).

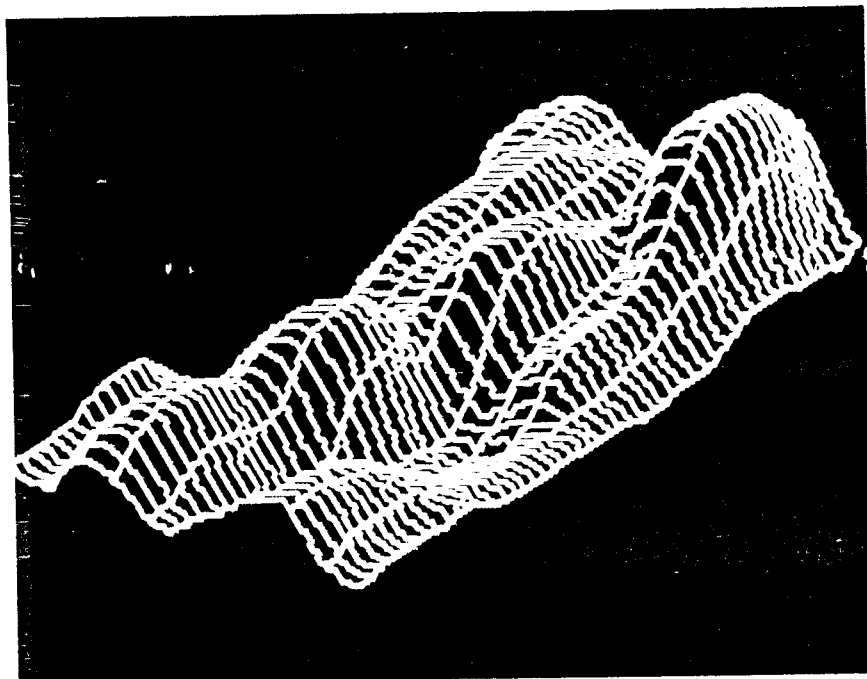
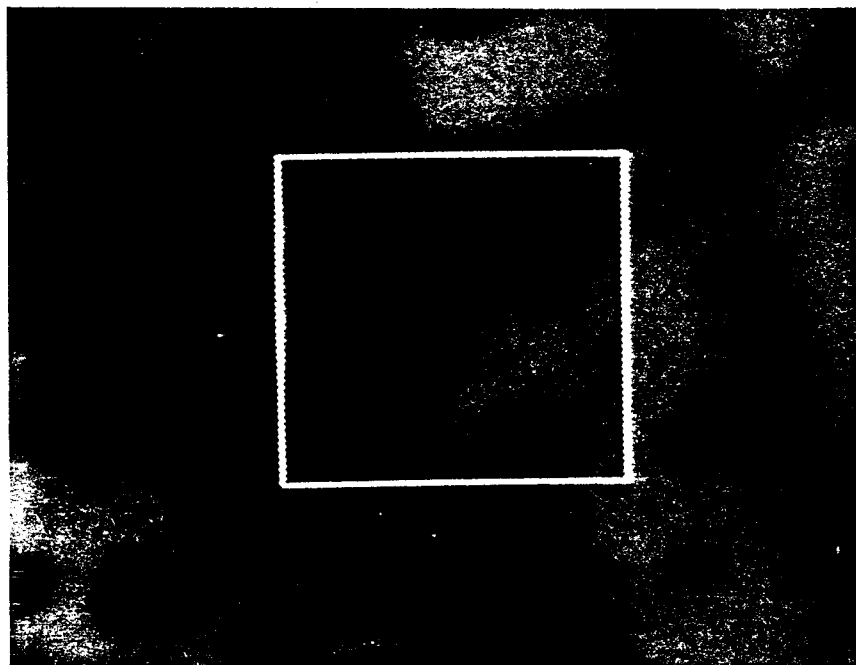


Figure 2 : a) Image initiale et zone analysée
b) Représentation tridimensionnelle du graphe de la fonction de gris.

Dans ce relief ainsi dessiné, les taches des protéines apparaissent donc comme des dômes ou monts d'altitude variable. La détection des taches

consiste donc à détecter les maxima régionaux du relief, puisque à chaque mont correspond une tache et une seule.

Les maxima de la fonction de gris sont fournis à la Figure 3.

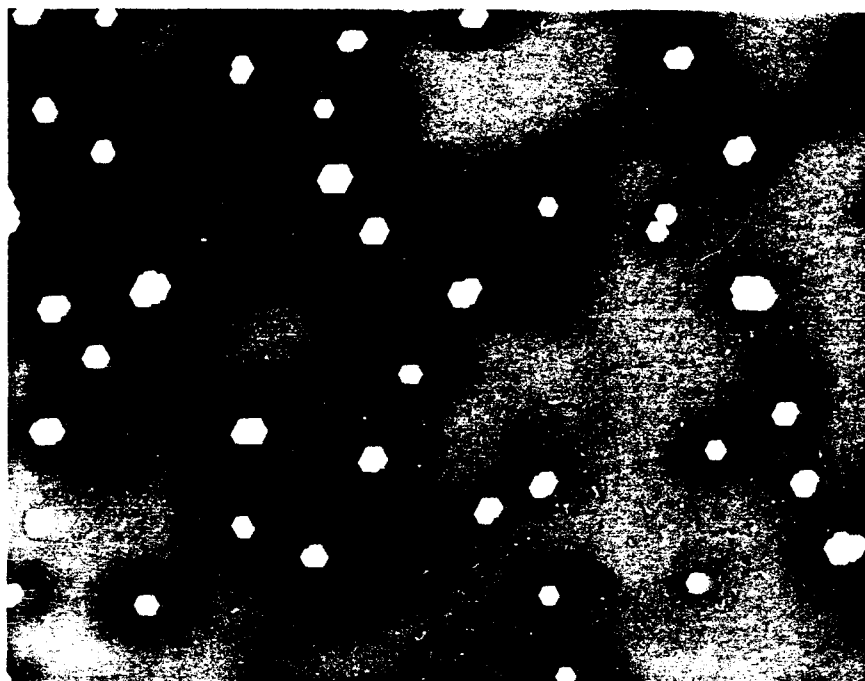


Figure 3 : Maxima de la fonction de gris et détection des taches.

Il est alors facile de générer les centroïdes correspondant à ces maxima (Figure 4), centroïdes dont les coordonnées donnent la position des taches de protéines.



Figure 4 : Centres des taches des protéines

Cette méthode met en évidence le centre optique de chaque tache, c'est-à-dire l'endroit où elle est la plus sombre, et non son centre géométrique. Deux taches qui se recouvrent partiellement seront séparées par ce procédé, à condition qu'elles soient discernables à l'oeil (Figure 5)

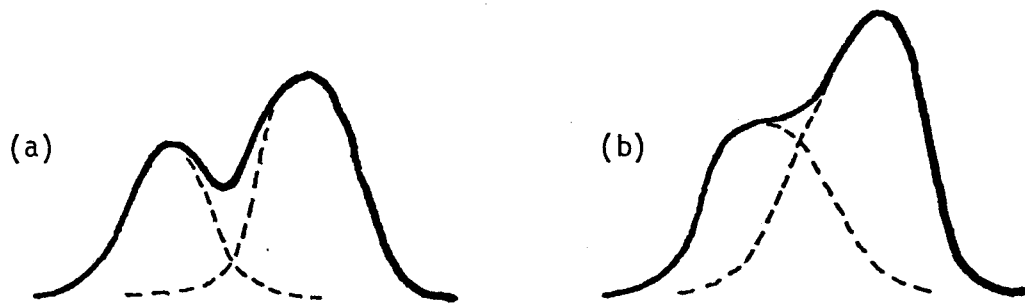


Figure 5 : Séparabilité des taches
a) cas favorable
b) la séparation est impossible

Dans les cas défavorables, il est possible d'effectuer cette séparation en faisant intervenir certains critères géométriques. Ces procédés n'ont pas été testés, car on court le risque de faire apparaître des éléments de l'image tels les traînées présentes sur le gel.

Détection du contour des taches

Cette détection du contour des taches constitue la deuxième étape du traitement. En effet, muni de ce contour, il va être possible de définir l'extension de la tache et son intensité. La détection du contour des taches utilise un algorithme de segmentation assez complexe, appelé "ligne de partage des eaux". En effet, revenant à la description du graphe de la fonction de gris en termes de relief, cet algorithme met en évidence les lignes de partage des eaux de ce relief (ou les lignes de col, si on utilise une représentation inversée).

La Figure 6 en particulier représente le résultat d'un tel algorithme appliqué à la fonction initiale.

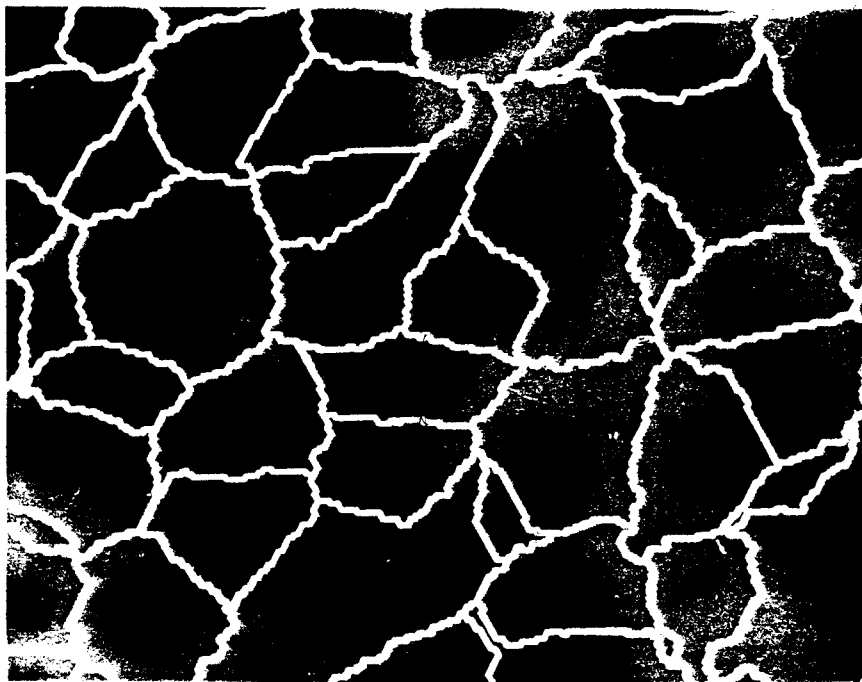


Figure 6 : Ligne de partage des eaux de la fonction de gris.

Cette transformation partitionne l'espace en polygones, appelés bassins versants. A chaque bassin versant correspond une tache de protéine, du moins pour celles qui correspondent à un maximum (Figure 7).

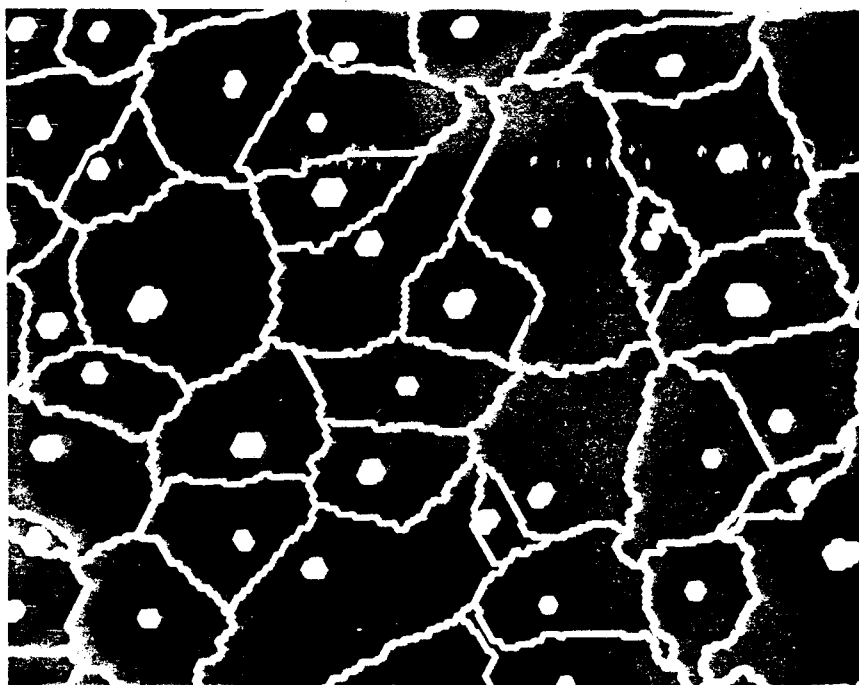


Figure 7 : Correspondance biunivoque entre les taches et les bassins versants.

Cette partition est intéressante à plus d'un titre. Elle est en effet utilisée pour détecter les contours des taches, mais elle fournit également des relations d'adjacence entre les taches. Cette particularité peut être très utile pour résoudre les problèmes de déformations liés à l'irrégularité du gel.

Les contours des taches sont obtenus en appliquant l'algorithme de lignes de partage des eaux à la fonction gradient de l'image initiale. La procédure utilisée prend en compte l'information fournie par la segmentation de la figure 7. Les contours des taches détectées apparaissent alors. Ces contours sont toujours séparés, même lorsque les taches ont tendance à se recouvrir (Figure 8). Cela est normal car, comme l'explique la Figure 9, les contours correspondent toujours à des maxima d'inflexion de la fonction de gris initiale.

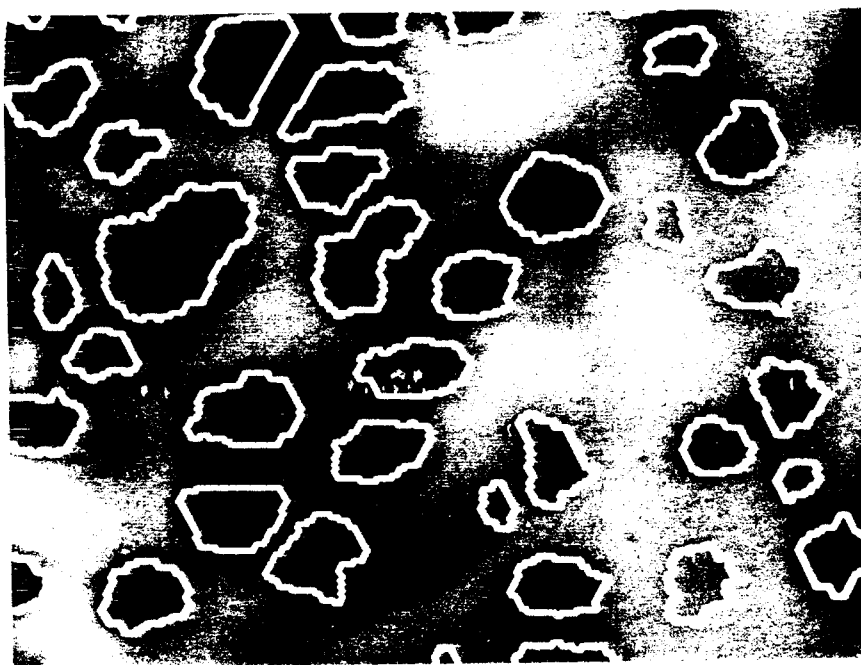


Figure 8 : Contours des taches

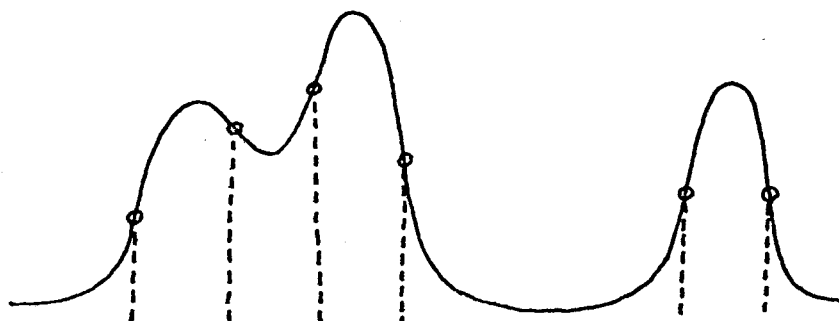


Figure 9 : Détection des contours sur les points d'inflexion

Ce procédé a peut-être tendance à biaiser quelque peu les mesures de l'intensité globale de la tache, mais il n'est pas difficile d'envisager de la corriger par quelques transformations morphologiques peu coûteuses en temps. (L'une d'entre elles, par exemple, pouvant consister à étendre le contour par dilatations de taille inversement proportionnelle au gradient au point d'inflexion).

PROBLEMES NON RESOLUS ET PERSPECTIVES

De nombreux problèmes liés à la lecture automatique des gels d'électrophorèse n'ont pas été abordés. Ce n'était d'ailleurs pas le but de cette étude dans laquelle on s'est surtout préoccupé de savoir si la reconnaissance automatique des taches était possible. Néanmoins, les résultats déjà obtenus et les outils utilisés apportent des éléments de réponse à ces problèmes. Deux problèmes méritent encore une attention particulière : les problèmes liés au recouvrement des taches d'une part et ceux liés à la comparaison des gels d'autre part.

Le recouvrement des taches

Comme on l'a indiqué précédemment, le recouvrement des taches induit deux phénomènes : d'abord une possible mauvaise détection des taches, ensuite un biais dans la quantification de ces taches. Nous avons déjà donné quelques ébauches de solution concernant ce dernier problème. Quant au premier, des outils morphologiques adaptés existent et ont fait leurs preuves dans des problèmes de segmentations similaires. Ce sont notamment tous les procédés dits d'érosion ultime (Figure 10).

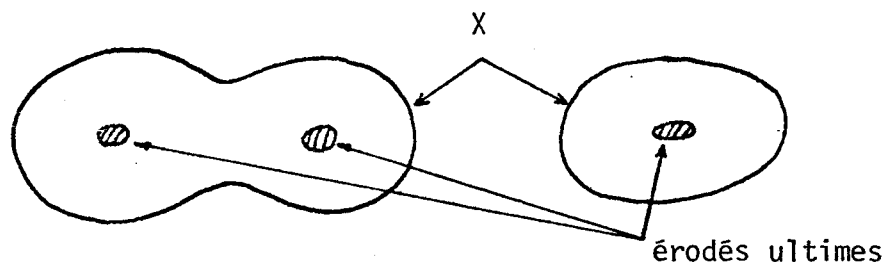


Figure 10 : Erodés ultimes d'un ensemble

L'érosion ultime d'une composante connexe formée de deux ou plusieurs taches se recouvrant aura tendance à la segmenter, car cette composante connexe est composée de l'union d'objets convexes (les taches).

Comparaison des gels

Ce problème n'a pas été abordé car il nous manquait le matériel nécessaire. La comparaison des gels peut s'effectuer à plusieurs niveaux, niveaux de difficulté croissante. On peut vouloir comparer plusieurs vues du même gel. C'est le niveau le plus élémentaire. Il recouvre d'ailleurs deux aspects, un aspect comparaison proprement dit, mais aussi un aspect amélioration d'image, puisque plusieurs vues du même gel peuvent être combinées pour réduire sensiblement le bruit sur l'image. On peut également comparer des gels provenant du même sujet. C'est à ce niveau qu'apparaissent les problèmes liés à l'hétérogénéité des gels. Cette hétérogénéité produit des déformations géométriques conduisant à des erreurs de positionnement. Enfin, le troisième niveau est celui de la comparaison de gels provenant de sujets différents. Il convient là de repérer les taches communes et les différences. Les deux derniers niveaux de comparaison sont, de loin, les plus difficiles. Des éléments de réponse à ces problèmes apparaissent avec la transformation ligne de partage des eaux introduite plus haut. La partition de la fonction de gris en ses différents bassins versants peut servir de point de départ pour apprécier les décalages des taches d'un gel à l'autre. En effet, il est raisonnable d'envisager qu'on acceptera un décalage d'autant plus important que la tache est isolée, autrement dit que son bassin versant est important. Un décalage identique de deux taches n'a pas les mêmes conséquences et n'apporte pas les mêmes difficultés selon que la tache est dans une zone très dense ou au contraire relativement isolée.

CONCLUSIONS

De nombreux autres problèmes n'ont pas été envisagés. Il n'était certes pas possible de tout traiter dans le cadre relativement restreint de cette pré-étude.

Nous n'avons pas effectué, entre autres, d'analyse comparative des résultats obtenus par les méthodes que nous utilisons, et ceux produits par les logiciels existants (en provenance des Etats-Unis) et ceci pour plusieurs raisons : en premier lieu, on ne dispose pas

à l'heure actuelle de standard en matière de gels. En second lieu, cette comparaison s'avère illusoire. Il est en effet difficile de comparer des procédés de traitement faisant appel à des outils théoriques très différents et surtout à des technologies aussi dissemblables que sont un ordinateur classique et un analyseur de texture.

Plusieurs verrous technologiques bloquent la lecture automatique des gels d'électrophorèse. Les plus importants sont les difficultés de standardisation de l'obtention du gel, d'une part, et les problèmes liés à la lecture du gel par un analyseur d'images, d'autre part. Cette étude démontre que les outils morphologiques peuvent apporter une contribution non négligeable au crochetage du deuxième verrou. Il est même raisonnable de penser, sans faire preuve d'un optimisme outrancier, que l'ensemble des problèmes posés par l'analyse des gels peuvent être résolus par l'ensemble des outils de la Morphologie Mathématique. Les résultats d'ores et déjà disponibles pourront être considérablement améliorés par l'utilisation de technologies plus adaptées telles que capteurs de haute qualité, mémoires numériques d'images, etc...

Cette amélioration promet d'être particulièrement sensible au niveau des temps de traitement. Ici nous avons utilisé un T.A.S. commercial dont la finalité est loin d'être celle de la lecture d'électrophorèses. La discrimination des tâches y prend environ 5 minutes. Il est facile d'imaginer des processeurs de morphologie, de facture plus moderne, qui pour un coût de 200.000 à 300.000 francs, effectueraient toutes les opérations (discrimination + quantification) en quelques dizaines de secondes. Cela est d'autant plus facile, qu'en fait ils sont déjà en chantier aux U.S.A.

Toutes ces informations doivent être confrontés aux besoins réels des utilisateurs biologistes, de façon à concevoir, puis réaliser, un système optimal.