
Approches statistique et neuronale pour la détermination et la classification de cellules orageuses

N. Ghazzali[†] — M. Parizeau[‡] — S. Dossou-Gbété[⊥] — E. Reny-Nolin[†]

[†]Département de mathématiques et de statistique

[‡]Département de génie électrique et de génie informatique

Université Laval. Québec (PQ). Canada. G1K 7P4.

[⊥]Laboratoire de mathématiques appliquées

Université de Pau et des pays de l'Adour. 64000 Pau. France

ghazzali@mat.ulaval.ca

RÉSUMÉ. Cet article s'intéresse à la détermination et la classification de cellules orageuses à partir des points d'impact de foudre repérés au sol. Le but est d'analyser leurs déplacements en temps réel et prédire leurs mouvements dans le futur afin de prévenir des cibles potentielles (résidents, entreprises,...) du risque d'orage. Nous proposons, dans un premier temps, de classer les points d'impact sur plusieurs périodes successives afin de reconnaître les cellules orageuses. Des méthodes classiques de classification hiérarchique (le saut minimal, le saut maximal, la moyenne et Ward) et non hiérarchique (les centres mobiles) sont utilisées pour analyser les données. Nous classifions ainsi les impacts de foudre sur des tranches de 15 minutes consécutives. Pour répondre à la problématique en temps réel et décrire le mouvement des cellules orageuses, nous proposons une approche originale basée sur les réseaux de neurones artificiels, appelée GNG-RT. Nous présentons les résultats obtenus par la classification et par le GNG-RT et nous discutons de l'apport et des perspectives de ce travail.

MOTS-CLÉS : Classification hiérarchique, algorithme de partitionnement, mélange de densités, réseaux de neurones artificiels, reconnaissance des formes, mouvement de cellules orageuses, GNG-RT.

1. Introduction

La problématique abordée dans cet article nous provient de Météorage, une société qui a été créée en 1986 par Franklin France, constructeur français de matériels de protection contre la foudre. Météorage s'occupe de la détection, de la localisation et du suivi des orages sur le territoire français. Cette société enregistre, grâce à 16 capteurs implantés sur toute la France, les caractéristiques précises de chaque décharge. Elle propose divers services par abonnement dont la cartographie des points d'impact de foudre, la recherche d'impacts dans une zone et une tranche horaire donnée, le suivi en temps réel des mouvements orageux et l'avertissement dès que l'orage franchit une certaine limite¹. Il s'agit donc de déterminer et de reconnaître des cellules orageuses à partir des points d'impact de foudre repérés au sol. Le but est d'analyser leurs déplacements en temps réel et de prédire leurs mouvements dans le futur afin de prévenir les abonnés de Météorage (résidents, entreprises,...) du risque potentiel.

À cet effet, un travail a déjà été effectué en 1989, par Isabelle Desprez [DES 89], sur l'élaboration d'un système de visualisation de cellules orageuses. Une analyse statique basée sur le critère de boules optimisées pour classer les points d'impact de foudre, suivi d'une analyse dynamique permettant de reconnaître une cellule dans son évolution sur plusieurs périodes successives de 15 minutes ont été réalisées. Cependant, ce travail ne répond pas à

1. Le lecteur pourra consulter le site web www.meteorage.com pour plus d'information. Les auteurs tiennent à remercier Météorage, et plus particulièrement Monsieur Stéphane Pédebois, de leur avoir fourni les données et permis de les analyser.

la problématique de reconnaître les cellules orageuses en temps réel et de pouvoir prédire leurs mouvements pour alerter les abonnés d'un risque d'orage potentiel.

Notre étude tente d'apporter quelques éléments de réponse à cette problématique. Nous proposons, en premier lieu, d'élargir le processus de classification des impacts de foudre à plusieurs critères de classification autres que celui des boules optimisées. Le problème propre au domaine de la classification relatif à la détermination du nombre de classes est discuté en considérant plusieurs critères dont celui de la pseudo-F, la pseudo- t^2 , le Cubic Clustering Criterion (CCC) [MIL 85] et celui de Banfield et Raftery [BAN 93] basé sur le mélange de densités [MCL 88]. Les trois premiers critères sont disponibles dans le logiciel SAS alors que le dernier est offert par le logiciel SPLUS. Les résultats obtenus par ces deux logiciels seront discutés pour déterminer des cellules orageuses. Cette extension est réalisée sur des périodes successives de 15 minutes et ne répond pas à la problématique du temps réel et de la prédiction en fonction du mouvement des cellules orageuses.

Pour y remédier nous présentons, en second lieu, une méthode originale basée sur les réseaux de neurones artificiels [HAY 94]. Cette méthode que nous appelons GNG-RT, nous permettra de répondre à la problématique posée et de s'affranchir de la contrainte reliée à la détermination du nombre de classes. Elle s'inspire du fonctionnement du réseau de neurones «Growing Neural Gaz» (GNG) de Fritzke [FRI 95]. Le GNG est un réseau constructif «auto-organisé», c'est-à-dire que sa structure n'est pas définie a priori mais croît plutôt au fil de l'apprentissage. Ses neurones représentent des points dans un espace d'entrée (on peut les assimiler à des centres mobiles), et ses connexions (entre les neurones) définissent des relations d'adjacence dans cet espace. Contrairement, par exemple, au réseau mieux connu de Kohonen [KOH 90], le GNG est capable d'apprendre la topologie d'un échantillon de données sans fixer a priori ni le nombre de neurones ni la structure de cette topologie.

Nous discuterons des résultats obtenus et de la pertinence de l'approche proposée et nous indiquerons comment l'approche par la classification pourrait être élaborée pour prendre en considération la notion du temps et le mouvement des cellules orageuses.

2. Description des données

Les données proviennent de Météorage Franklin. Elles représentent une situation orageuse de 6 jours consécutifs du 2 au 7 août 1999 sur l'Aquitaine, où ont été enregistrés 19012 points d'impact sur le sol. On dispose donc d'un tableau de données où chaque point d'impact est décrit par 4 variables, qui sont: 1) la date de l'impact exprimée en jour, mois et année 2) le moment de l'impact dans la journée exprimée en heure, minutes et secondes, 3) la longitude exprimée en degrés décimaux et 4) la latitude exprimée en degrés décimaux. Avant de présenter les analyses statistiques, nous aimerions définir la notion de cellules orageuses, au sens de Krider, telle qu'elle a été présentée dans Desprez ([DES 89], page 9): «Une cellule orageuse est un regroupement d'impacts de foudre proches aussi bien dans l'espace que dans le temps».

3. Analyses statistiques

Les analyses présentées dans cette section sont celles issues des méthodes de classification et de la méthode neuronale GNG-RT. Nous analyserons, plus spécifiquement, la journée du 2 août 1999 pour la période de 8h00 à 12h00. Bien entendu, ce choix ne représente aucunement une limitation des approches proposées et l'analyse aurait pu être faite sur n'importe quels jour, zone ou période horaire.

3.1. Méthodes classiques de classification

Reconnaître les cellules orageuses consiste, dans cette première approche, à appliquer une méthode de classification permettant de regrouper les points d'impact selon un critère de proximité. Pour prendre en considération la proximité temporelle, le temps est partitionné en périodes de 15 minutes consécutives. Ces intervalles sont soit disjoints (par exemple]8h00,8h15],]8h15,8h30],...), soit décalés de 5 minutes (par exemple,]8h00,8h15],]8h05,8h20],...). La classification des points d'impact se fait alors par tranche de 15 minutes en fixant le temps et en ne considérant que les coordonnées spatiales à savoir la longitude et la latitude. La mesure de proximité entre ces points d'impact est, par conséquent, la distance euclidienne. Les critères d'agrégation des classes sont le saut minimal, le saut maximal, la moyenne et Ward. Mais également, nous avons considéré la méthode de partitionne-

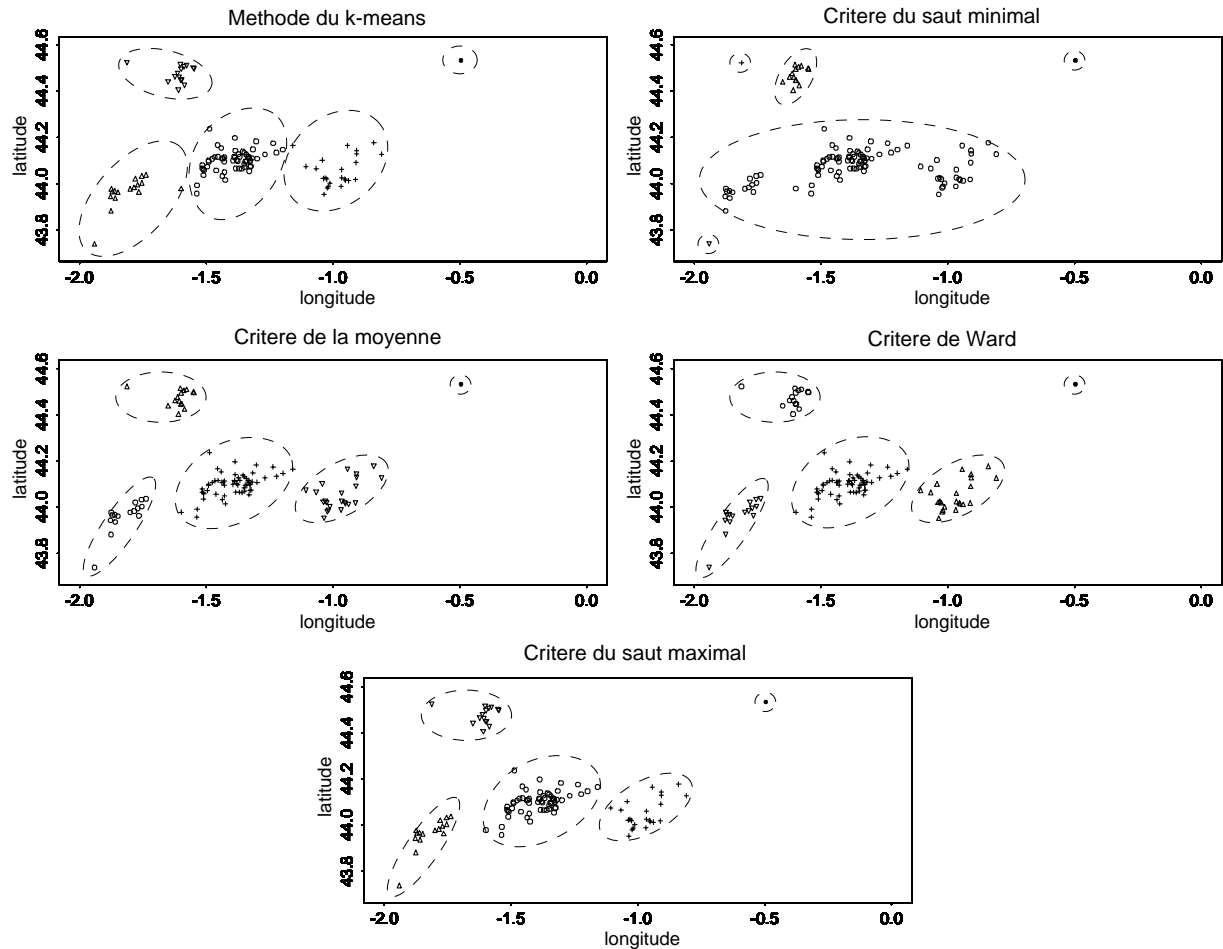


Figure 1. Classification des impacts de foudre en 5 classes de 9h15 à 9h30 par les critères du k-means, saut minimal, moyenne, Ward et saut maximal.

ment autour des centres mobiles et plus particulièrement la méthode du k-means. Nous avons utilisé les logiciels SAS et SPLUS pour faire ces analyses.

Le problème de la détermination du nombre de classes s'est posé aussi bien pour les méthodes de classification hiérarchique que pour le k-means. Pour éviter toute procédure manuelle ou subjective, nous avons analysé les résultats obtenus par différentes statistiques: la pseudo-F, la pseudo- t^2 , le CCC et Banfield & Raftery (le critère est décrit dans [BAN 93], page 807) pour choisir le bon nombre de classes.

Pour se fixer les idées, concentrons-nous sur la période de 9h15 à 10h15 que nous subdivisons en 4 intervalles disjoints de 15 minutes chacun. Au total, nous disposons de 586 impacts à analyser dont 116 sont entre [9h15,9h30], 184 entre [9h30,9h45], 156 entre [9h45,10h00] et 130 entre [10h00,10h15]. L'analyse des différents critères de détermination du nombre de classes nous conduit à retenir 5 classes.

La figure 1 présente la partition en 5 classes pour les 5 critères de classification considérés, et ce de 9h15 à 9h30. Nous remarquons que les critères de Ward, du saut maximal et de la moyenne donnent exactement les mêmes résultats. Le k-means donne des résultats similaires à deux impacts près. On obtient alors des classes plutôt sphériques, de même taille ou de même variance. Alors que le saut minimal donne, comme on s'y attendait, une grosse classe allongée et des classes isolées. Ces deux structures de classes existent effectivement dans les données. Ce qui sera illustré par la méthode neuronale (cf. §3.2) où toutes les données seront représentées en temps réel.

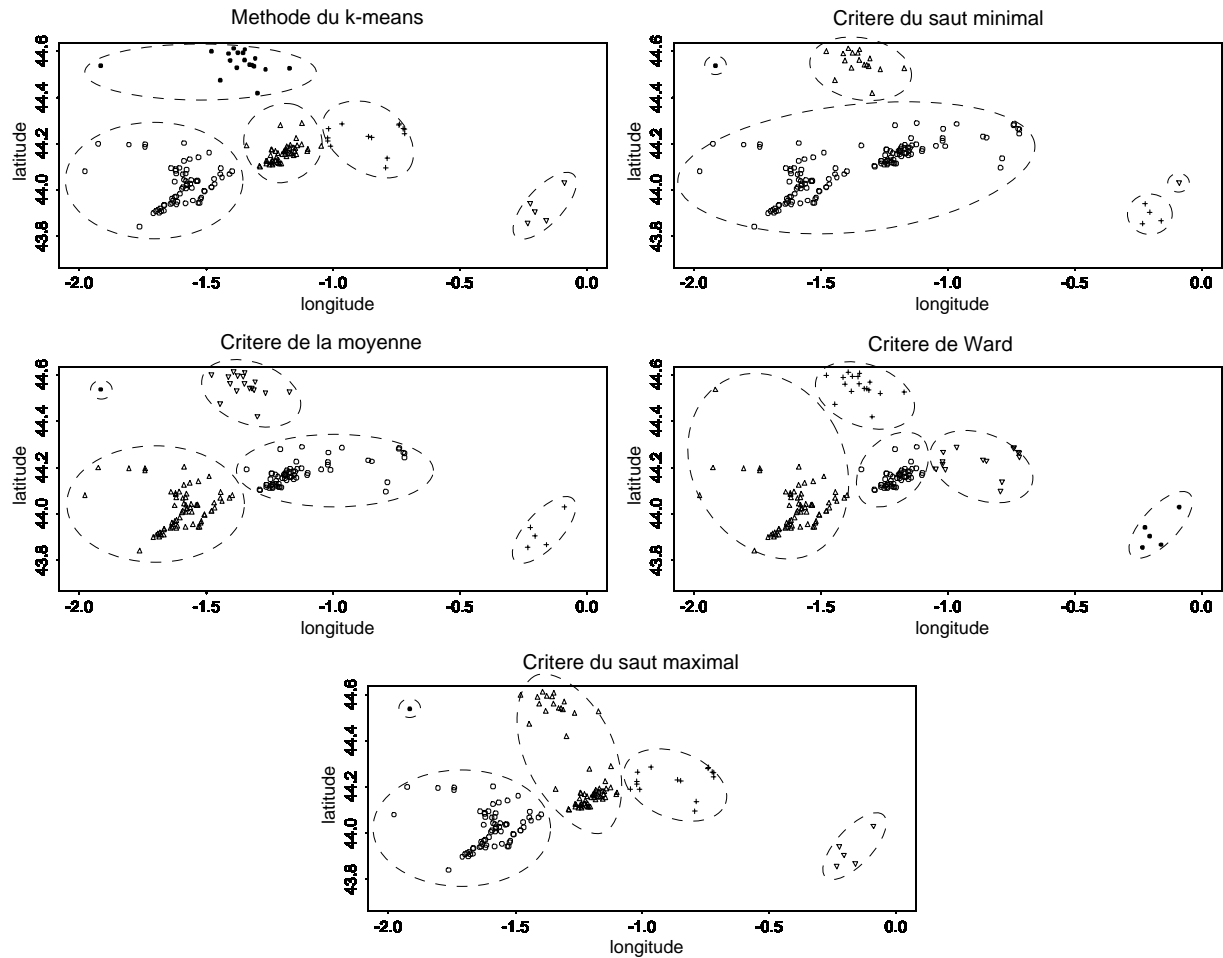


Figure 2. Classification des impacts de foudre en 5 classes de 9h45 à 10h00 par les critères du k-means, saut minimal, moyenne, Ward et saut maximal.

Les résultats obtenus entre 9h15 et 9h30 ne sont pas nécessairement les mêmes pour les autres intervalles. En ce sens, les partitions peuvent différer de par leur structure d'un critère à l'autre. La figure 2 donne la partition en 5 classes pour les 5 critères de classification, et ce entre 9h45 et 10h00. Le critère du saut minimal étant à part, nous analysons les quatre autres critères. Ces derniers donnent deux types de résultats. Le premier, obtenu par le critère du saut maximal et celui de la moyenne, révèle une classe avec un seul point. Alors que le deuxième, formé par le k-means et Ward, agrège ce même point à une autre classe. De plus, pour un type donné, les classes des partitions ne sont pas nécessairement les mêmes.

Les classes évoluent et changent d'une période de temps à une autre. Certaines classes meurent alors que d'autres naissent ou enregistrent moins d'impacts de foudre. En particulier, comment interpréter une classe formée d'un seul impact? Subsisterait-elle si on réduisait le nombre de classes à 4 ou à 3? Comment évolue-t-elle à travers le temps?

Le mouvement des cellules orageuses doit être pris en compte pour mieux les reconnaître. La figure 3 donne la partition en 5 classes pour le critère de la moyenne à travers les 4 intervalles à l'étude, soit de 9h15 à 10h15. Notons P_1, P_2, P_3, P_4 les partitions pour chaque intervalle de temps. Soit $P_i = \{C_{i1}, C_{i2}, C_{i3}, C_{i4}, C_{i5}\}$, $i = 1, \dots, 4$, les classes composant chacune des partitions, numérotées de gauche à droite. On remarque que P_1 diffère de P_2 puisqu'on assiste à la mort de la classe C_{15} formée d'un seul impact et à la naissance de la classe C_{25} formée de 3 impacts. En effet, la classe C_{15} ne donne pas suite à une activité orageuse alors que la classe C_{25} montre

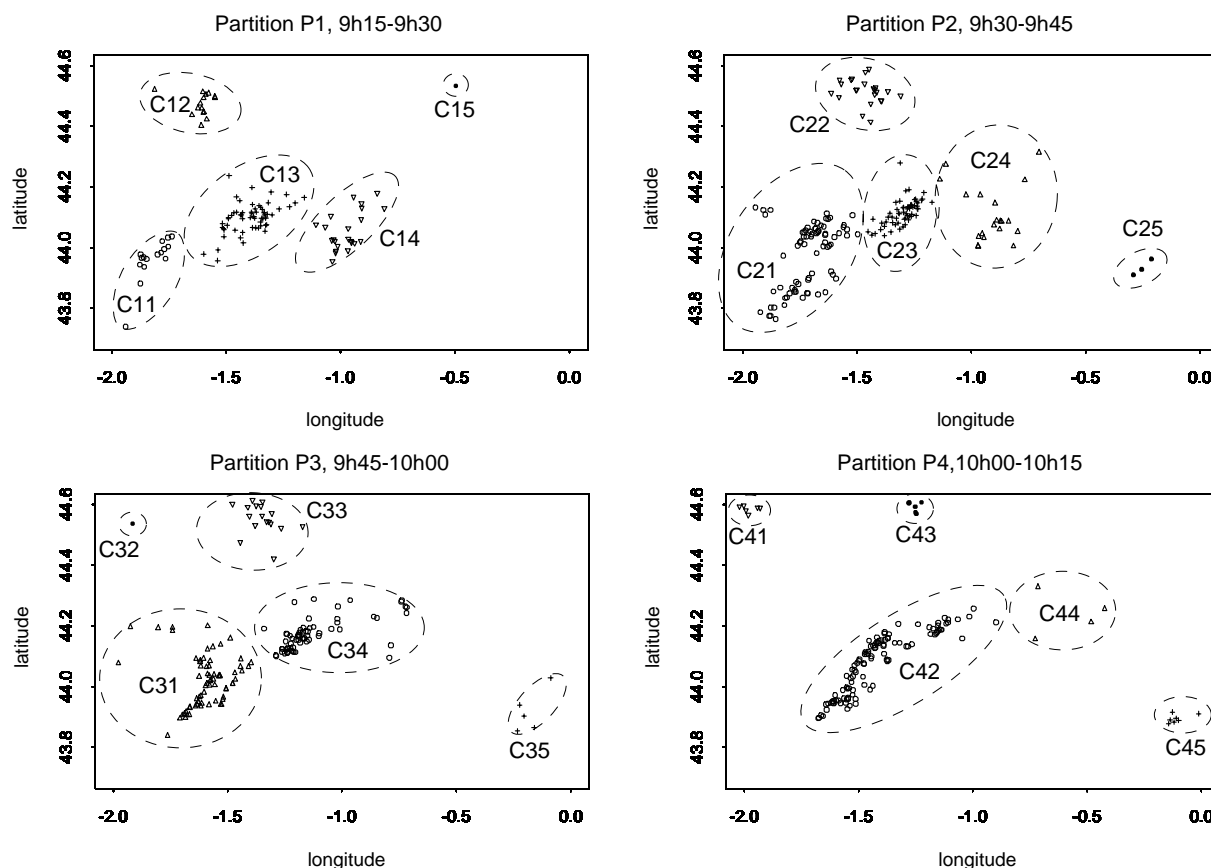


Figure 3. Classification des impacts de foudre en 5 classes de 9h15 à 10h15 par le critère de la moyenne.

la naissance d'une cellule orageuse qui s'intensifie avec le temps en *C35* et *C45*. Les classes *C11*, *C12*, *C13* et *C14* définissent des cellules orageuses qui s'amplifient dans le deuxième intervalle de temps. Cependant, de 9h45 à 10h00, on observe un ralentissement de l'activité orageuse de *C23* et *C24* qui s'agrègent en *C34*, et la naissance de la cellule *C32*. Dans le dernier quart d'heure, cette dernière s'intensifie alors que *C31* et *C34* ralentissent leur activité et se transforment en *C42*. Finalement, on assiste à la naissance d'une nouvelle cellule *C44*.

3.2. Méthode neuronale

En ce qui concerne cette deuxième approche, nous ne fixons pas le temps par périodes de 15 minutes mais en temps réel, c'est-à-dire que chaque point d'impact est analysé en tenant compte du moment où il se produit. Le GNG dont nous nous inspirons peut servir à traiter des données temporelles, mais à condition que celles-ci soient échantillonnées à intervalle régulier. Or puisque la foudre peut frapper à n'importe quel moment et que le temps qui sépare deux impacts possède une signification importante pour notre analyse, nous avons modifié quelque peu l'algorithme d'apprentissage du GNG pour en faire un GNG-RT («real-time»).

Les neurones du GNG-RT représentent les régions de l'espace où dans un passé «récent», il y a eu des impacts. À chaque neurone est associé un rayon spécifiant l'étendue de la région correspondante (i.e. les régions sont circulaires); il existe un paramètre ρ qui spécifie l'étendue maximale d'une région. La position du neurone peut changer dynamiquement au fil du temps, de même que l'étendue de la région qui lui est associée.

Une connexion reliant deux neurones indique qu'il y a eu des impacts récents quelque part à proximité des deux neurones en question. Chaque connexion possède un âge. Les connexions vieillissent donc avec le temps et peuvent finir par mourir si elle ne sont jamais rajeunies. Lorsqu'un impact se produit à proximité d'une connexion,

l'âge de celle-ci est réinitialisé à zéro. L'âge maximal d'une connexion est fixé par le paramètre τ . Passé cet âge, la connexion meurt et disparaît.

C'est donc le paramètre τ qui distingue le passé récent de celui qui ne l'est plus. Un neurone qui ne possède aucune connexion meurt et disparaît lui aussi. Ainsi, une région orageuse anciennement active devient inactive dès que durant une période τ , aucun impact ne survient autour de son centre. Et nous définissons les cellules orageuses comme les composantes connexes du graphe non orienté engendré par le GNG-RT.

Nous présentons maintenant plus en détail l'algorithme d'apprentissage du GNG-RT. À chaque intervalle de temps élémentaire δ , par exemple à chaque seconde, les étapes suivantes sont exécutées dans l'ordre:

1. Incrémenter de δ l'âge de toutes les connexions;
2. Éliminer les connexions dont l'âge est supérieur à τ ;
3. Retirer les neurones isolés (sans connexion);
4. Réduire le rayon r de tous les neurones: $r = \alpha r$, où $0 < \alpha < 1$ est le facteur de réduction.

Les étapes précédentes sont indépendantes de la foudre. Elles font en sorte que le réseau s'adapte en temps réel. Lorsque la foudre frappe à la position \mathbf{x} , les étapes additionnelles suivantes sont exécutées:

5. Si le réseau est vide, créer une nouvelle composante connexe de deux neurones centrés sur \mathbf{x} ;
6. Soit N_1 le neurone le plus près de \mathbf{x} , et N_2 le deuxième plus près;
7. Additionner au carré du rayon de N_1 , le carré de la distance entre \mathbf{x} et N_1 ;
8. Rapprocher N_1 de \mathbf{x} :

$$N_1 = N_1 + \epsilon_g (\mathbf{x} - N_1) \quad [1]$$

où $0 < \epsilon_g < 1$ est le facteur de rapprochement pour le neurone «gagnant»;

9. Rapprocher aussi de \mathbf{x} tous les neurones N , voisins immédiats de N_1 :

$$N = N + \epsilon_v (\mathbf{x} - N) \quad [2]$$

où $0 < \epsilon_v < 1$ est le facteur de rapprochement pour les voisins du neurone gagnant;

10. Si N_1 et N_2 sont déjà connectés, alors remettre à zéro l'âge de la connexion; Autrement, créer une nouvelle connexion entre N_1 et N_2 ;
11. Soit N_a le neurone dont le rayon est maximal parmi tous les neurones, et N_b le neurone parmi ses voisins immédiats et dont le rayon est également maximal;
12. Si le rayon de N_a est supérieur à ρ :
 - (a) Insérer un nouveau neurone N_c à mi-chemin entre N_a et N_b ;
 - (b) Insérer une connexion entre N_c et N_a , et une autre entre N_c et N_b ; retirer la connexion entre N_a et N_b ;
 - (c) Réduire de moitié les rayons de N_a et N_b , et affecter le nouveau rayon de N_a au rayon de N ;

13. S'il existe encore un neurone dont le rayon est supérieur à ρ , retourner à l'étape 6;

L'étape 5 permet d'initialiser le réseau avec le premier impact disponible, soit au temps $t = 0$, soit après une période d'accalmie ayant provoqué la mort de toutes les connexions et, par conséquent, de tous les neurones. L'étape 6 permet de déterminer le neurone «gagnant», c'est-à-dire le neurone associé à la région de l'espace qui est la plus rapprochée de l'impact. L'étendue de cette région est ensuite augmentée à l'étape 7 en fonction du carré de la distance à l'impact. Ainsi, la région couverte par le neurone s'ajuste dynamiquement selon l'erreur. Plus loin, à l'étape 12, les neurones qui couvrent une trop grande étendue seront subdivisés. Les étapes 8 et 9, quant à elles, consistent à déplacer le neurone gagnant ainsi que ses voisins immédiats (les neurones reliés par des connexions) dans la direction de l'impact. Les paramètres ϵ_g et ϵ_v déterminent l'ampleur de ce déplacement. L'étape 10 vient remettre à zéro l'âge de la connexion entre les deux neurones les plus près de l'impact, ou encore elle crée cette connexion si celle-ci n'existe pas déjà. Une connexion entre deux neurones représente un lien entre les deux régions sous-jacentes. Un nouvel impact situé quelque part entre ces deux régions vient donc confirmer l'existence de ce lien et donc l'appartenance des deux régions à une même cellule orageuse. Finalement, les étapes

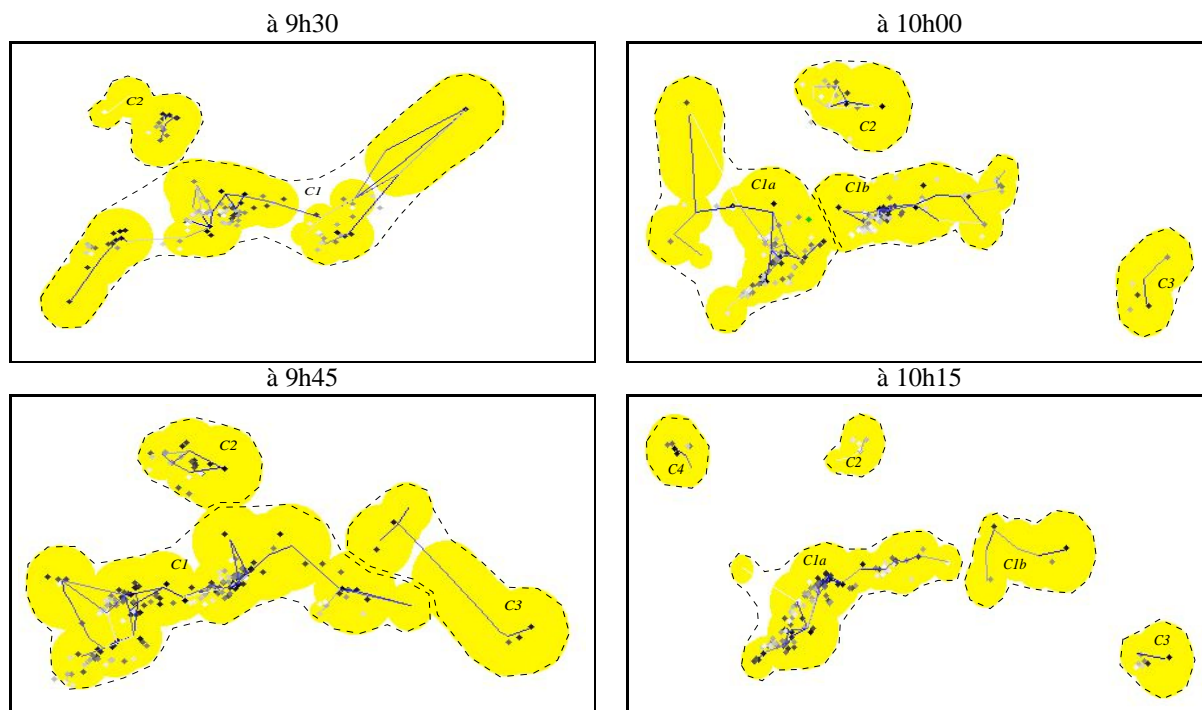


Figure 4. Classification des impacts de foudre par la méthode GNG-RT.

11 et 12 permettent d'identifier le neurone dont la région associée est la plus étendue. Si cette étendue dépasse un certain seuil préétabli (ρ), un nouveau neurone est inséré à mi-chemin entre le neurone en question et son voisin possédant lui aussi, parmi l'ensemble des voisins, la région maximale. Cette étape vient donc placer un nouveau neurone dans une cellule orageuse où l'erreur est maximale.

La figure 4 présente l'état du réseau à quatre instants dans la journée du 2 août 1999: à 9h30, à 9h45, à 10h00 et à 10h15. Chacun des graphiques montre la position des impacts durant les 15 minutes précédentes, ainsi que l'étendue des régions associées aux neurones et les connexions entre ces régions. La position des neurones n'est pas illustrée pour ne pas encombrer la figure, mais peut être déduite approximativement des brisures dans les lignes des connexions. La teinte de gris des points d'impact représente le temps écoulé depuis l'impact en question. Un point noir indique un impact récent, alors qu'un point blanc (invisible) indique un impact qui remonte à plus de 15 minutes. De même, la teinte de gris des connexions représente l'âge de celles-ci (les connexions sur le point de mourir sont presque invisibles). Pour ces résultats, l'âge maximal des connexions a été fixé à 15 minutes également. Pour les autres paramètres, nous avons choisi: $\rho = 0.4$, $\alpha = 0.999$ (typiquement nous fixons $\alpha = \frac{\tau-1}{\tau}$ où τ est exprimé en secondes), $\epsilon_g = 0.8$ et $\epsilon_v = 0.1$.

À 9h30, le réseau GNG-RT indique qu'il existe en fait deux cellules orageuses représentées par deux composantes connexes dans le graphe (ces dernières sont encerclées par des lignes pointillées). La première cellule $C1$ est plutôt horizontale et semble être composée de quatre sous-cellules reliées les unes aux autres par des connexions plus ou moins âgées. Mais les liens entre ces quatre sous-cellules sont bien réels et se confirment dans la minute suivante par plusieurs impacts qui viennent les souder solidement ensemble. La deuxième cellule, de plus petite taille, même si elle est à proximité de la première se déplace quant à elle de manière plutôt indépendante comme le montrent les graphiques à 9h45, 10h00 et 10h15.

Par ailleurs, on remarque en haut à droite (dans le graphique de 9h30) un point d'impact isolé mais récent (celui qui correspond à la classe $C15$ de la figure 3) qui a été rattaché à la cellule $C1$ parce que celle-ci était la plus proche. L'algorithme du GNG-RT rattache toujours un nouvel impact à une cellule existante. C'est le temps et les impacts subséquents qui détermineront le sort de ce nouvel appendice. Lorsqu'il s'agit d'un impact réellement

isolé, la ou les connexions qui relient les neurones sous-jacents à ceux de la cellule principale finiront par mourir et l'appendice formé disparaîtra. Ce sera une fausse alerte qui durera τ secondes. De même, s'il s'agit de la formation d'une nouvelle cellule orageuse, celle-ci se séparera de la cellule principale τ secondes plus tard. Lorsqu'il s'agit d'un mouvement subit de la cellule orageuse, alors celui-ci sera anticipé, ce qui est fortement souhaitable.

Quinze minutes plus tard, à 9h45, une troisième cellule s'est formée en bas à droite, et celle-ci s'est fusionnée avec l'appendice de la cellule $C1$ faussement formée à 9h30. La cellule $C2$, quant à elle, s'est légèrement déplacée vers la droite. On remarque aussi qu'en bas à gauche la cellule $C1$ a tendance à se déformer vers le haut, ce qui se confirmera à 10h00, puis à 10h15 avec la formation d'une nouvelle cellule $C4$.

À 10h00, la cellule $C1$ s'est scindée en deux cellules notées $C1a$ et $C1b$. Cette scission résulte simplement d'une absence d'impact dans la zone qui les sépare. Elle est cohérente avec les résultats obtenus par l'approche classique. Un impact isolé en haut à gauche a aussi créé un appendice tout comme à 9h30. Mais cette fois-ci, cet appendice se confirmera à 10h15 comme étant une cellule orageuse distincte.

Finalement, à 10h15, le GNG-RT détecte 5 cellules orageuses. Il s'agit des cellules $C1a$ et $C2$ correspondant respectivement aux cellules $C1$ et $C2$ présentes à 9h30 et qui se sont déplacées vers la droite. On remarque sur le dos de $C1a$ une bosse qui correspond à un vieux neurone dont la connexion est sur le point de mourir. En effet, cette connexion n'a été associée à aucun impact depuis presque 15 minutes. Les cellules $C3$ et $C4$ sont vraisemblablement de nouvelles cellules orageuses apparues depuis ce temps. La cellule $C1b$, quant à elle, correspond probablement à la portion la plus droite de la cellule $C1$ originelle qui se déplace plus rapidement que le reste vers la droite.

4. Conclusion

Cet article présente deux approches, statistique et neuronale, pour déterminer et reconnaître des cellules orageuses à partir de points d'impacts de foudre repérés sur le sol. Le but est d'analyser le déplacement de ces cellules en temps réel et de prédire leurs mouvements dans le futur afin de prévenir ces cibles potentiels du risque d'orage. L'approche statistique, basée essentiellement sur les méthodes de classification, s'appuie sur une proximité géographique. L'aspect temporel est pris en considération en analysant les cellules orageuses sur des périodes successives de 15 minutes. Alors que l'approche neuronale, basée sur le GNG-RT, s'affranchit de cette limitation en analysant les données en temps réel et en construisant un graphe dont les arêtes naissent, vieillissent et, éventuellement, meurent pour engendrer des composantes connexes dynamiques. Ce sont ces composantes connexes qui modélisent les cellules orageuses. Le problème de prédiction est alors envisageable puisque, à l'intérieur d'une même zone géographique, il suffit de prévenir les cibles potentielles d'un risque si l'activité orageuse se rapproche d'un certain seuil τ , plus ou moins important. L'une des perspectives de ce travail s'inscrit dans ce cadre pour définir plus précisément cette notion de zone géographique et de seuil temporel. Nous aimerions également développer l'approche statistique en incorporant le temps dans la mesure de proximité et en considérant les méthodes de mélange de densités à des fins de classification.

Références

- [BAN 93] BANFIELD J.D. & RAFTERY A.E., «Model-Based Gaussian and Non-Gaussian Clustering», *Biometrics*, vol. 49, 1993, p. 803–821.
- [DES 89] DESPREZ I., «Élaboration d'un système de visualisation des cellules orageuses», *Mémoire d'ingénieur*, Conservatoire National des Arts et Métiers, Institut d'informatique d'entreprise, 1989.
- [FRI 95] FRITZKE B., «A Growing Neural Gas Network Learns Topologies», D.S. Touretzky et T.K. Leen (éditeurs), *Advances in Neural Information Processing Systems 7*, MIT Press, Cambridge MA, 1995.
- [HAY 94] HAYKIN S., *Neural Networks: A comprehensive Foundation*, IEEE Press, 1994.
- [KOH 90] KOHONEN T., «The Self-Organizing Map», *Proceedings of the IEEE*, Vol. 78, no. 9, 1990, pp. 1464–1480.
- [MCL 88] Mc LACHLAN G. & BASFORD K., *Mixture Models: Inference and Applications to Clustering*, Marcel Dekker, 1988.
- [MIL 85] MILLIGAN, W.G. & COOPER, M.C., «An Examination of Procedures for Determining the Number of Clusters in a Data Set», *Psychometrika*, vol. 50, no.2, 1985, p.59–179.