

Description des alignements formés par DTW

François Petitjean

► To cite this version:

| François Petitjean. Description des alignements formés par DTW. 2011. hal-00647522

HAL Id: hal-00647522

<https://hal.archives-ouvertes.fr/hal-00647522>

Submitted on 2 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Description des alignements formés par DTW

François Petitjean

1^{er} décembre 2011

Résumé

Ce document a pour but de décrire des indices permettant de décrire les alignements formés par DTW entre deux séquences. Il se focalisera surtout sur la description des alignements formés par DTW entre une séquence moyenne et une séquence associée, mais le principe resterait identique si les indices étaient calculés entre deux séquences.

I Avant propos : Rappel sur DTW

Dynamic Time Warping (DTW) est inspirée de la distance de Levenshtein, également appelée distance d'édition, et a été introduite par [2, 3] dans le domaine de la reconnaissance vocale. DTW permet de trouver l'alignement global optimal entre deux séquences, c'est-à-dire d'associer chaque élément de chaque séquence à au moins un élément de l'autre séquence en minimisant les coûts d'association. Le coût d'une association correspond à la distance entre les deux éléments ; classiquement une L^p -norm¹. La Figure 1 représente un exemple d'alignement opéré par DTW, illustrant l'alignement de deux sinusoides légèrement déphasées. Le résultat numérique fourni par DTW correspond à la somme des hauteurs des "barreaux" formés par les associations. Les extrémités des alignements de la Figure 1 montrent que DTW est capable de réaligner correctement une séquence par rapport à une autre, et parvient ainsi à saisir des similarités que la distance euclidienne ne peut extraire.

Dynamic Time Warping est définie récursivement par :

$$D(A_i, B_j) = \delta(a_i, b_j) + \min \begin{cases} D(A_{i-1}, B_{j-1}), \\ D(A_i, B_{j-1}), \\ D(A_{i-1}, B_j) \end{cases} \quad (1)$$

où A_i représente la sous-séquence $\langle a_1, \dots, a_i \rangle$. Le coût de l'alignement optimal est alors donné par $D(A_{|A|}, B_{|B|})$. La Figure 2 illustre l'arbre des appels récursifs. On peut remarquer que la résolution du problème principal, symbolisé par la racine, passe par la résolution de sous-problèmes se recouvrant (noeuds de mêmes couleurs apparaissant dans différents sous-arbres). Le principe de programmation dynamique peut alors s'appliquer afin de résoudre cet arbre en partant des feuilles. La fonction DTW peut alors être mémorisée : les différents appels peuvent être retenus afin de ne pas calculer deux fois la fonction appelée avec les mêmes paramètres. Aussi est-il habituel, comme l'arbre contient $|A| \cdot |B|$ noeuds différents, de stocker ces différents résultats intermédiaires dans une matrice $|A| \times |B|$. Le calcul de DTW consiste alors à trouver le chemin de coût minimum dans la matrice, ce qui s'effectue avec une complexité en temps et en mémoire

1. *Nota Bene* : le temps séparant deux éléments n'est pas pris en compte par la distance, ce qui a certaines implications que nous détaillerons par la suite.

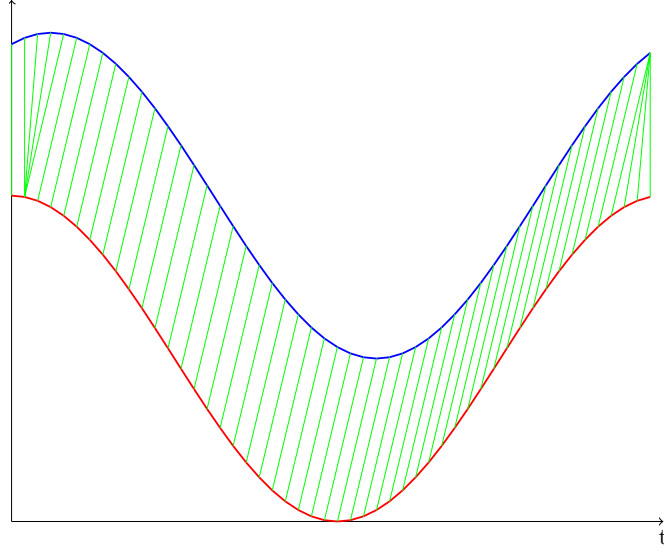


FIGURE 1 – Deux séquences en dimension 1 alignées avec Dynamic Time Warping. Les coordonnées de la séquence du haut et de celle du bas correspondent respectivement à $\cos(t)$ et à $\cos(t + \alpha)$. Pour des questions de visualisation, la séquence du dessus a été décalée vers le haut lors du tracé.

de $\Theta(|A| \times |B|)$ ². Le calcul de l'alignement de deux séquences selon ce principe est illustré en Figure 3.

II Exploiter les alignements de DTW

Quels que soient les travaux utilisant DTW, seul le coût de l'alignement y est utilisé. Cependant, notre hypothèse est que la structure de l'alignement de deux séquences constitue également une information pertinente. Par exemple, deux couples de séquences ayant la même « distance » peuvent avoir été alignées de façon complètement différente. Considérons par exemple trois séquences $A = \langle 0, 1, 0 \rangle$, $B = \langle -\frac{1}{2}, \frac{3}{2}, -\frac{1}{2} \rangle$ et $C = \langle -\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, 1, 0 \rangle$ illustrées en Figure 4. Bien que l'alignement par DTW de A avec B et de A avec C ait le même coût, la structure de ces alignements est complètement différente. Ainsi, le coût de l'alignement ne renseigne-t-il que faiblement sur la structure des alignements obtenus, c'est-à-dire sur le synchronisme des séquences, sur l'avance de l'une par rapport à l'autre, *et caetera*.

Partant de ce constat, la structure des alignements apparaît informative pour l'analyse des données. Ce document se focalise sur la description de ces alignements par la création d'indices de descriptions (également appelés descripteurs par la suite). Bien que d'autres solutions peuvent être étudiées pour l'analyse de l'appariement de deux séquences, la création d'indices de descriptions permet de retranscrire de façon condensée l'information complexe contenue dans l'alignement de deux séquences. De plus, les descripteurs ont l'avantage de représenter une information condensée, sans pour autant aller jusqu'à l'analyse des alignements; cette propriété permet de conserver l'information de distorsion temporelle extraite par DTW, tout en étant ra-

2. La complexité en mémoire peut être réduite à $\Theta(\min(|A|, |B|))$ en ne conservant que deux lignes ou une colonne de la matrice.

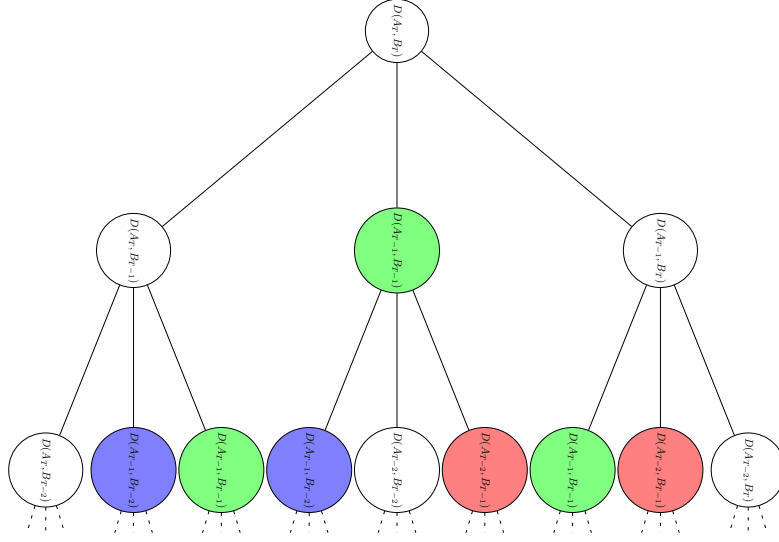


FIGURE 2 – Partie haute de l’arbre des appels récursifs (Équation 1) à DTW. Cette figure met en évidence différents noeuds communs entre trois sous-arbres.

pidement analysable sans avoir à relancer une analyse par DTW. On peut par exemple imaginer que des descripteurs des alignements seront pré-calculés pour chaque zone $(x; y)$ de chaque image acquise par télédétection (coûteux en temps) et que le travail d’analyse des experts se focalisera sur l’étude (possiblement interactive) des cartes de descripteurs, disposant d’une représentation informative et efficace des alignements. Les descripteurs constituent donc un juste équilibre entre représentation informative et efficace des alignements.

S’agissant d’alignements entre deux séquences, plusieurs questions simples permettent d’éclairer le problème (certaines de ces questions sont redondantes) :

1. Quel est de degré de synchronisation des séquences ?
2. Combien de dilatations / contractions sont-elles nécessaires à l’alignement des séquences ?
3. Quel est l’effort³ nécessaire pour aligner les séquences ?
4. Les séquences sont-elles désynchronisées depuis le début ?
5. Quelle est la répartition des alignements entre les deux séquences ?
6. Quelles parties des séquences est-il été nécessaire de réaligner ?
7. Les resynchronisations sont-elles longues ou courtes ?

En réponse à ces questions, plusieurs descripteurs peuvent être envisagés :

1. Le nombre de barreaux nécessaires pour le réalignement, ou la longueur du chemin dans la matrice.
2. La longueur des plateaux et des murs du chemin.
3. Le coût total de l’alignement : le score de DTW.

Les trois dernières questions sont, quant à elles, relatives à la distribution des alignements entre deux séquences. L’étude de la distribution des données fait appel à des outils simples (ex : moyenne, écart-type, histogrammes) ou plus complexes (ex : algorithmes de clustering, algorithmes de réduction de dimensions). Par ailleurs, il est intéressant de noter que le chemin

3. Effort est ici utilisé au sens large de coût.

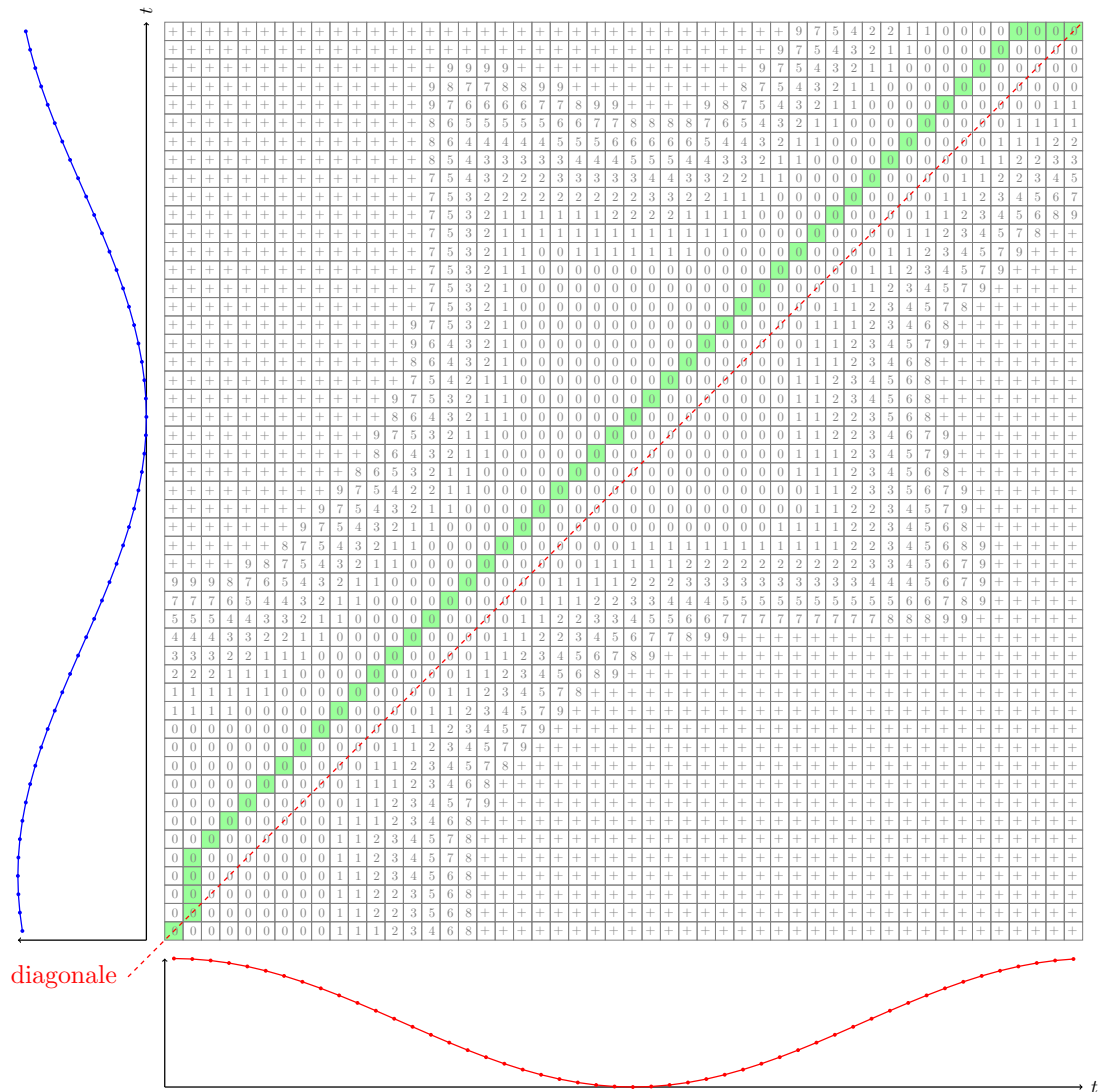


FIGURE 3 – Exemple de chemin d’alignement calculé par DTW utilisant le principe de programmation dynamique entre les deux séquences de la Figure 1. *N.b.* : Les chiffres sont arrondis à l’entier le plus proche, et “+” représente les chiffres supérieurs à 9.

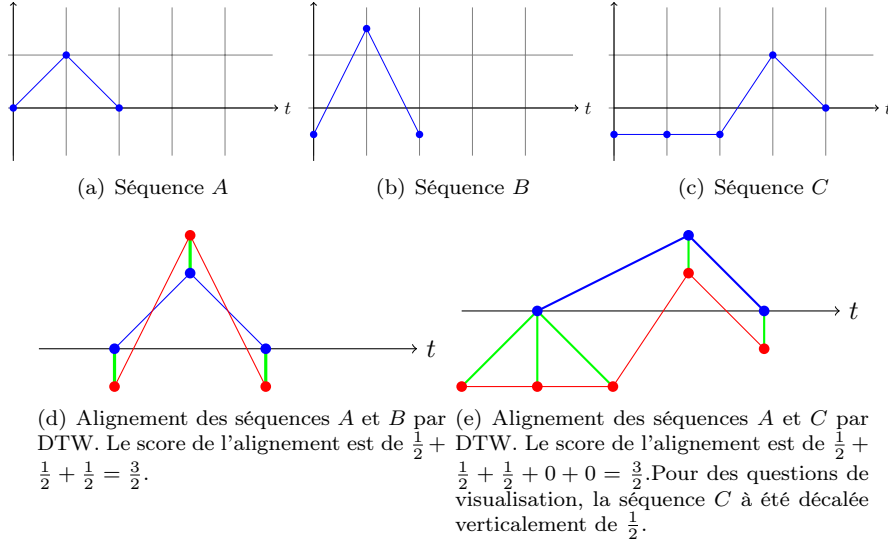


FIGURE 4 – Trois séquences sont illustrées en (a) (b) et (c). Les alignements respectifs de A avec B et de A avec C sont illustrés en (d) et (e). Cette figure met en exergue deux alignements complètement différents dont les scores sont identiques : $DTW(A, B) = DTW(A, C)$.

d'alignement (*warping path*) dans la matrice constitue une représentation des alignements. Il peut donc être utile de décrire les alignements en décrivant ce chemin directement. En effet, ce chemin peut être vu comme le tracé d'une fonction⁴; la description des alignements peut donc se faire par l'étude de cette fonction (dérivée, intégrale, etc.).

III Descripteurs séquentiels d'alignements

Cette section présente plusieurs indices proposés pour la description de l'alignement de deux séquences. Les deux premiers indices utilisent la représentation de l'alignement sous forme de graphe d'association des différents éléments des deux séquences. Les indices suivants utilisent quant à eux la représentation des alignement induite par le chemin de coût optimal d'alignement. La première partie présente pour chaque indice son principe, sa définition, l'information qu'il représente. La seconde partie présente quant à elle la caractérisation d'une série d'images satellites grâce à ces indices.

1 Théorie

1 a) Nombre d'alignements

Le nombre d'alignements correspond au nombre de “barreaux” permettant d'aligner les deux séquences; il correspond également à une information très proche de la longueur du chemin d'alignement (*warping path* en anglais) dans la matrice de calcul de l'alignement (voir le chemin en vert en Figure 3).

4. Il s'agit en effet d'une fonction croissante, définie sur l'intervalle $[1, |A|]$ (ou $[1, |B|]$) et passant par les deux points $(1, 1)$ et $(|A|, |B|)$.

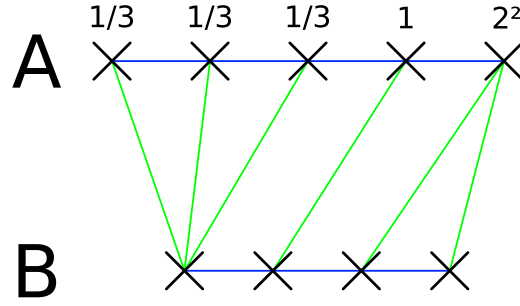


FIGURE 5 – Exemple de calcul du Page Rank. Ce descripteur aurait pour résultat $PageRank(A, B) = \sqrt[6]{\left(\frac{1}{3}\right)^3 \times 1 \times 2^2} = \sqrt[6]{\frac{4}{27}} \approx 1.37$.

Ce descripteur des alignements donne une information de déformation globale entre les deux séquences. Ainsi, un chemin très court comme celui formé par la diagonale dans la matrice indique-t-il une synchronisation complète des deux séquences. À l'inverse, un chemin très long comme celui suivant les bords de la matrice indique que les deux séquences ne sont pas du tout synchrones et que beaucoup de déformations sont nécessaires pour aligner les deux séquences.

1 b) Page Rank

Cet indice reprend en partie le principe de l'indice de classement des pages web développé par Google. Son objectif est de fournir un descripteur prenant en compte la longueur des paliers, c'est-à-dire tenant compte du fait que les prises de directions "haut" ou "bas" dans la matrice (correspondant aux accélérations / décélérations de la section précédent) font partie d'un palier ou d'une verticale d'une certaine longueur. Cette information renseigne sur les décalages successifs opérés entre les deux séquences pour l'alignement ; l'objectif étant de différencier les alignements ayant subi des déformations contiguës des autres. Aussi, le descripteur proposé consiste-t-il à associer à chaque élément du chemin, sa fraction de participation au palier courant. La Figure 5 illustre le principe. L'ensemble de ces fractions est ensuite résumé en utilisant la moyenne harmonique, s'agissant de ratios. Il est important de noter que ce descripteur n'est pas symétrique. Cette asymétrie peut cependant ne pas être contraignante si cet indice est utilisé pour comparer l'alignement de deux séquences à une troisième séquence de référence. Ce principe sera utilisé pour obtenir des résultats comparables en décrivant deux séquences par rapport à une séquence d'évolution moyenne.

1 c) Intégrale du chemin

Bien que le premier indice utilisait déjà implicitement les notions de dérivées et d'intégrale pour calculer la longueur du chemin d'alignement, ce groupe d'indice repose sur l'utilisation directe de l'information fournie par l'intégrale. L'idée sous-jacente est d'étudier l'aire sous la courbe du chemin (Figure 6), par rapport à la diagonale (représentant un alignement sans déformations et donc un synchronisme des deux séquences).

Cette information d'éloignement à la diagonale permet de différencier certains types d'alignements qui ne le sont pas en utilisant simplement la longueur du chemin. La Figure 7 illustre deux alignements différents dont les longueurs de chemins sont identiques. On comprend ici en quoi la longueur du chemin est insuffisante pour la caractérisation des alignements. L'intégrale apporte ici une information relative à la durée de désynchronisation. Plusieurs indices peuvent

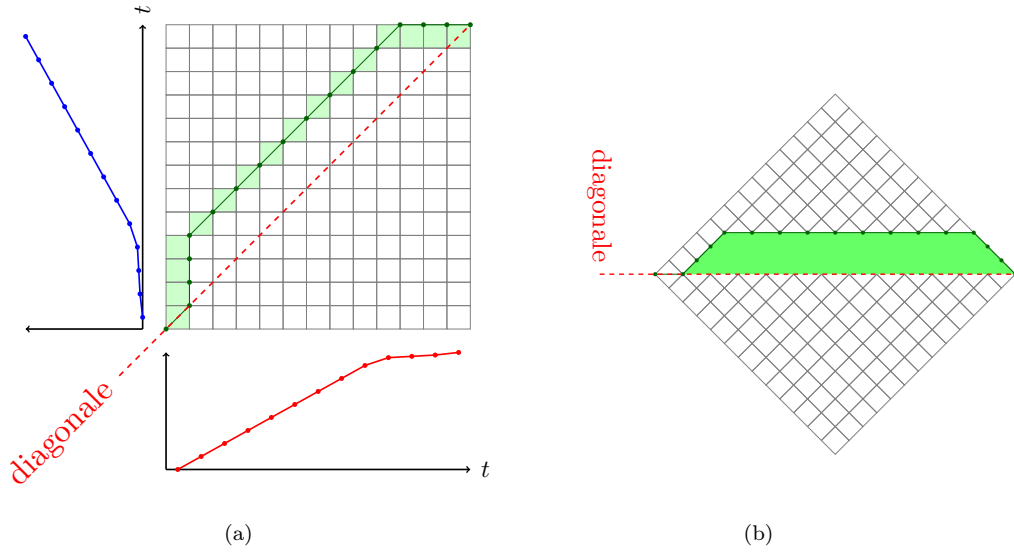


FIGURE 6 – La Figure (a) montre la fonction définie par le chemin d'alignement (vert foncé). La Figure (b) illustre quant à elle l'aire sous la courbe de cette fonction par rapport à la diagonale. Par abus de notation, l'aire dessinée est l'aire sous la courbe de la fonction continue interpolée par rapport à la fonction discrète décrite par le chemin d'alignement.

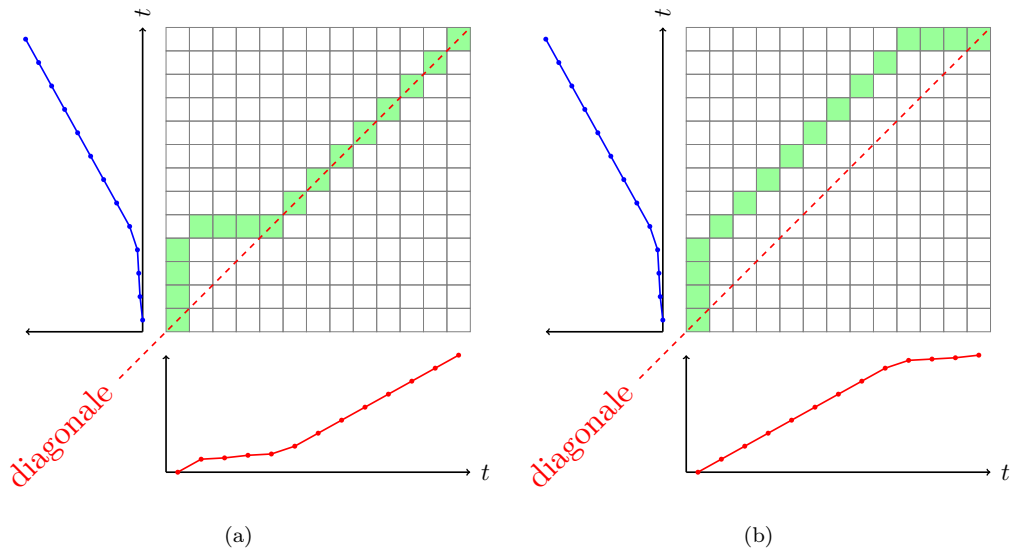


FIGURE 7 – Deux exemples d'alignements sont illustrés en (a) et (b). Ces deux alignements ont le même nombre de barreaux (*i.e.*, la longueur du chemin est identique) mais le comportement d'alignement est complètement différent. Les séquences alignées en (a) sont désynchronisées uniquement au début alors que les séquences en (b) sont désynchronisées du début à la fin.

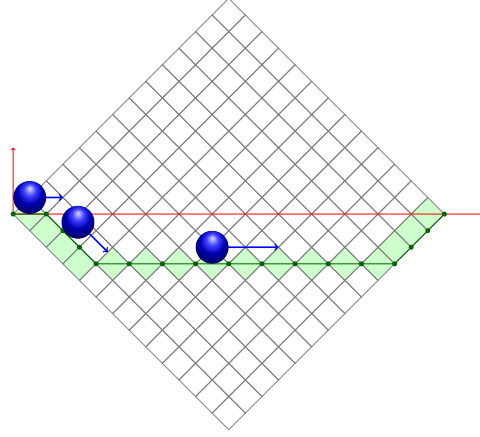


FIGURE 8 – Illustration du comportement d'une balle évoluant sur le chemin d'alignement.

être dérivés en utilisant l'intégrale comme outil de base :

1. l'intégrale directe, *i.e.*, la somme des distances de chaque élément du chemin à la diagonale (perpendiculairement), permettant d'évaluer la déformation globale nécessaire pour aligner les deux séquences ;
2. la moyenne des intégrales partielles (pour chaque élément), correspondant à la distance moyenne du chemin à la diagonale, et permettant d'évaluer la déformation moyenne nécessaire à l'alignement des deux séquences ;
3. l'écart-type des intégrales partielles, permettant de mesurer la dispersion des distances autour de la distance moyenne.

Par ailleurs, il est intéressant de noter l'intégrale aura des valeurs positives lorsque la fonction sera au dessus de la diagonale et inversement. Cette information permet de différencier l'avance ou le retard d'une séquence par rapport à une autre.

1 d) Dérivée / Parcours du chemin

Dans la continuité des indices précédents étudiant le comportement de la fonction décrite par le chemin d'alignement, cette section s'intéresse à la dérivée de cette fonction, projetée dans le repère ayant la diagonale pour axe des abscisses et sa perpendiculaire pour axe des ordonnées (voir Figure 8). Notons ici qu'à l'inverse de l'intégrale de la fonction, la fonction discrète n'est pas dérivable ; nous considérerons donc la dérivée de l'interpolation linéaire de la fonction. L'idée sous-jacente est d'étudier le temps de parcours d'une balle déposée sur le chemin d'alignement, comme illustré en Figure 8. En pratique, la balle subit autant de variations d'accélération (y compris nulles) qu'il y a de segments le long du chemin d'alignement. Débutant avec une vitesse v , celle-ci augmente à chaque fois que la dérivée de la fonction est négative et diminue lorsque la dérivée est positive. Par ailleurs, la vitesse initiale v doit être fixée de façon à ce que la vitesse ne puisse pas être négative. On peut alors étudier deux variables sur ce principe : le temps de parcours et la vitesse moyenne du parcours. Cependant, chacune de ces variable couplée à la

longueur du chemin représentent la même information puisqu’elles sont reliées par :

$$v_{moy} = \frac{d}{t} \quad (2)$$

Dans la même lignée, il serait également possible d’utiliser le nombre d’accélérations pour caractériser les alignements ; le nombre de décélérations est inutile, étant fortement corrélé au nombre d’accélérations et à la longueur des séquences.

2 Résultats

Cette section a pour objectif de présenter une évaluation qualitative des descripteurs décrits dans la section précédente. Nous disposons pour cela d’un ensemble de séquences, et aimerions caractériser chacune d’entre elles par ces indices. Aussi, l’idée est-elle de construire une séquence moyenne représentant toutes les classes de cultures confondues. Pour cela, la carte de référence est utilisée afin de construire un masque culture/non-culture. L’idée est ainsi de pouvoir mettre en évidence les distorsions temporelles présentes entre les différentes cultures (déphasage des pics de cultures, vitesses de maturation différentes, etc.). Une séquence moyenne d’évolution des cultures est donc construite à partir de toutes les séquences non-masquées. Puis, chaque séquence est alignée par DTW à cette séquence moyenne et tous les indices sont alors calculés pour chaque séquence.

Dans notre étude, s’agissant de séquences construites à partir de séries d’images satellites, chaque séquence correspond à l’évolution radiométrique d’une zone géographique atomique, identifiée dans les images par ses coordonnées (x, y) . Comme il est possible de décrire chaque séquence par rapport à la séquence moyenne d’évolution des cultures, il est possible de re-spatialiser les résultats ; une image à niveaux de gris peut donc être construite pour chaque indice. Il est ainsi possible d’évaluer qualitativement la cohérence spatiale et thématique des résultats. Les cartes associées aux différents indices sont présentées en Annexe A. Il est intéressant d’en étudier la capacité à différencier les différents comportements temporels, que ce soit pour séparer différents types de cultures, ou pour retranscrire des phénomènes temporels (ex : avance/retard de maturation) à l’intérieur d’une même classe thématique. Par ailleurs, la cohérence spatiale renseignera sur la dispersion des valeurs.

De façon quantitative, les résultats seront évalués dans le texte en étudiant la distribution des valeurs des indices en fonction des classes considérées ; nous utiliserons pour cela l’histogramme de distribution.

Les résultats sont principalement présentés sur la série d’images la plus longue (2006) ; la Figure 9 illustre la vérité terrain de la série d’images satellites utilisée ainsi que le masque de cultures, sur lequel une unique séquence moyenne est calculée.

Pour référence et comparaison avec les autres indices développés, l’Annexe B présente l’utilisation de la valeur de DTW pour séparer les différentes classes.

2 a) Nombre d’alignements

La Figure 10 présente les résultats du nombre d’alignements permettant d’aligner chacune des séquences (identifiée par leurs coordonnées (x, y)) à la séquence moyenne des cultures, *i.e.*, au comportement moyen d’évolution des cultures. Cet indice présente des résultats cohérents, avec de légères différences entre les différentes classes, mais aucune des cultures n’est nettement séparée des autres et même les parcelles de maïs sont confondues avec les parcelles de blé. Cet indice a donc un comportement cohérent mais pas assez discriminant pour être utilisé pour séparer les différentes classes en fonction de leurs valeurs uniquement. De plus, il ne permet pas

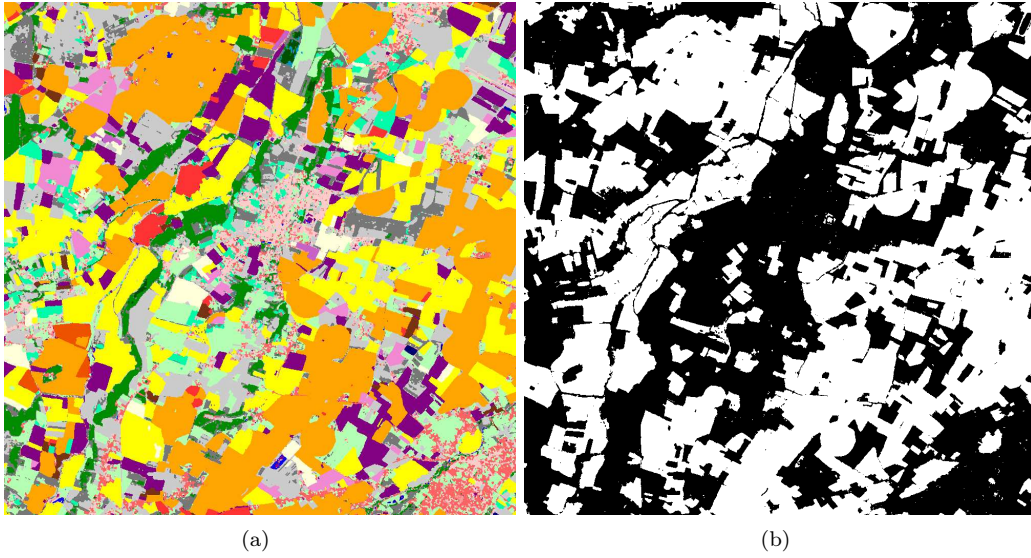


FIGURE 9 – (a) Carte de classification de référence de la SITS de 2006 et (b) masque de cultures sur lequel la séquence moyenne est calculée.

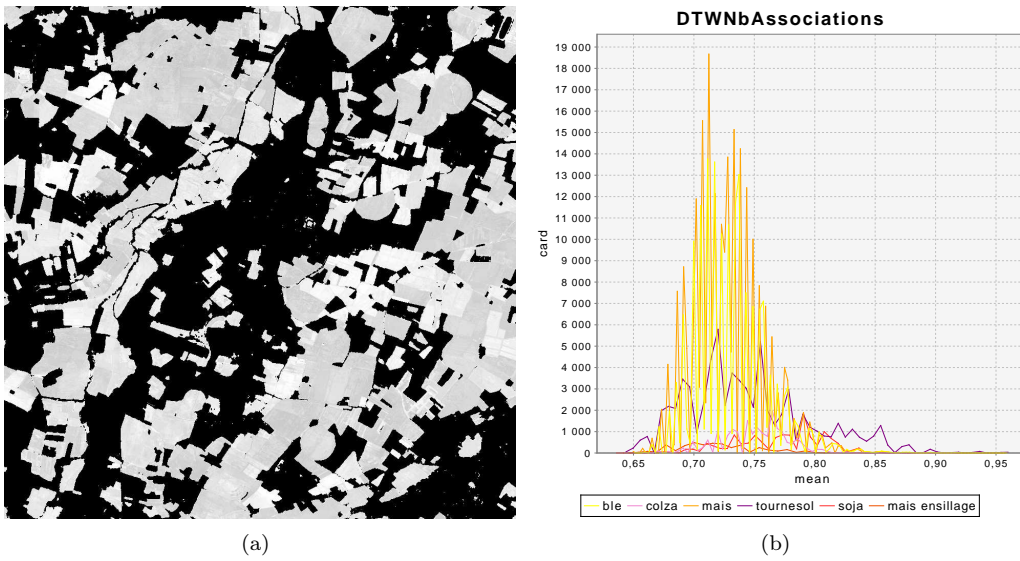


FIGURE 10 – (a) Carte résultat du descripteur correspondant au nombre d’alignements, normalisé par le nombre maximum d’associations ($|A|+|B|-1$). (b) Histogramme de distribution des valeurs par rapport aux différentes classes.

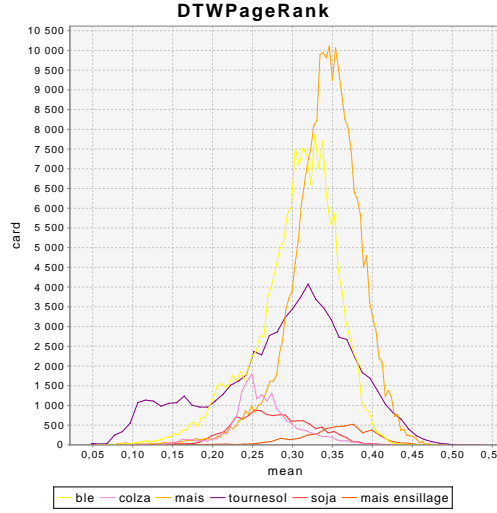


FIGURE 11 – Histogramme associé au descripteur inspiré du PageRank.

non-plus l’identification de comportements intra-classes caractéristiques, avec des teintes globalement homogènes à l’intérieur d’une même classe. Néanmoins, l’étude de la carte en Figure 10(a) donne deux indications. Premièrement, couplée à l’information spatiale, la valeur de l’indice permet de distinguer certaines parcelles très claires correspondant à des parcelles de colza (rose). Deuxièmement, les parcelles se distinguent les unes des autres, *i.e.*, la teinte des différentes parcelles est spatialement cohérente, ce qui conforte la pertinence de l’étude des alignements pour la description des évolutions.

2 b) Page Rank

Le descripteur inspiré du PageRank est illustré en Figure 11. Cet indice ne permet pas beaucoup mieux de discriminer les différentes cultures que le simple nombre d’alignements, à l’exception de quelques valeurs associées au tournesol. Puisque cet indice est sensible à la longueur morceaux de chemin d’alignement dont la direction reste constante, ce résultat tend à démontrer que cette information n’est pas discriminante pour les cultures, ce qui est conforté par l’étude de la carte présentée en Figure 21. Cependant, ce résultat n’invalide pas l’utilisation de cet indice pour d’autres données moins bruitées. En effet, avec les données radiométriques, les morceaux de chemin sans changement de direction sont plutôt rares car les réalignements sont plus nuancés/progressifs et rarement opérés sur des données consécutives.

2 c) Intégrale du chemin

Les trois graphiques présentés en Figure 12 correspondent respectivement à la distance moyenne à la diagonale, à l’écart-type de la distance à la diagonale, et à l’intégrale signée sous la diagonale.

La Figure 12(a) montre que l’indice basé sur la distance moyenne à la diagonale “démélange” plus efficacement les différentes gaussiennes mais ne les sépare pas complètement.

Concernant l’écart-type de la distance à la diagonale (Figure 12(b)), on peut remarquer de façon évidente que cet indice est très corrélée à l’indice précédent, ce qui peut s’expliquer assez facilement : une distance moyenne faible correspond à un chemin d’alignement le long

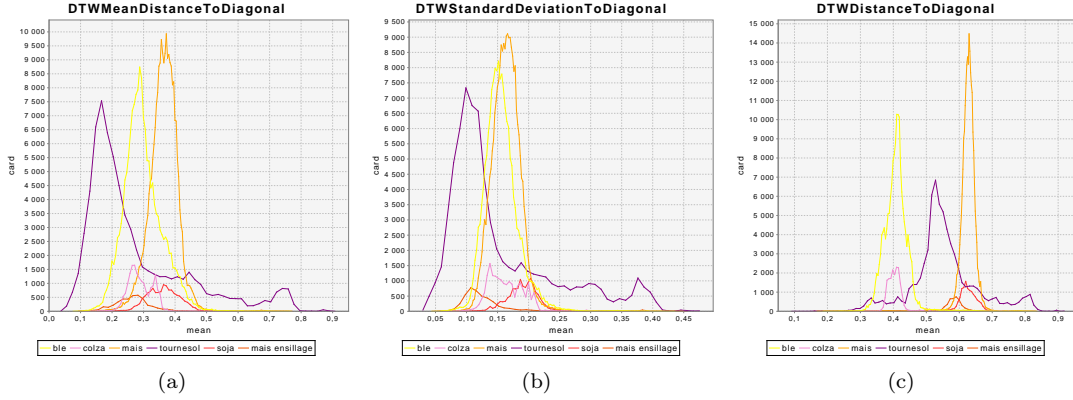


FIGURE 12 – Histogrammes de répartition des classes correspondant aux indices dérivés de la distance à la diagonale : (a) distance moyenne à la diagonale, (b) écart-type de la distance à la diagonale et (c) intégrale signée de la courbe par rapport à la diagonale.

de la diagonale, et donc à un chemin ne s'en éloignant pas ; or si celui-ci ne s'en éloigne pas, la distance de chaque case à la diagonale reste également faible, d'où un écart-type également faible. À l'inverse, dans le cas d'un chemin suivant les bords de la matrice et correspondant à une distance moyenne à la diagonale élevée, le chemin doit obligatoirement débuter aux coordonnées $(1, 1)$ et finir aux coordonnées $(|A|, |B|)$: deux points à distance nulle à la diagonale. Ainsi, pour des distances à la diagonale élevées, les distances de chaque élément du chemin à la diagonale sont très variables, d'où un écart-type élevé.

L'intégrale⁵ signée sous la diagonale est quant à elle illustrée en Figure 12(c). Les intégrales négatives (sous la diagonale) sont représentées par des valeurs de l'indice dans l'intervalle $[0, \frac{1}{2}]$, tandis que les intégrales positives sont représentées par des valeurs de l'indice dans l'intervalle $[\frac{1}{2}, 1]$ pour des raisons de normalisation. Aussi, même si la classe tournesol apparaît ici dans les valeurs proches de 0.5, ce résultat est cohérent avec les résultats obtenus précédemment puisqu'il correspond à une intégrale nulle, c'est-à-dire dont le chemin d'alignement est proche de la diagonale. On peut remarquer ici que cet indice parvient à démêler de façon plus efficace les trois principales classes de cultures de maïs, de blé et de tournesol. Par ailleurs, le positionnement de ces trois gaussiennes est assez cohérent d'un point de vue thématique puisque les cultures de blé et de colza sont plus précoces (avril-mai) que les cultures de maïs et de soja (juillet-août), tandis que les dates de cultures de tournesol sont plus disséminées dans le temps. L'ordre d'apparition des gaussiennes correspond donc à une réalité temporelle d'apparition des cultures. La principale différence avec les autres résultats dérivés de la distance à la diagonale, est que ce résultat discerne une part importante des cultures de tournesol : la moitié pixels correspondant à des cultures de tournesol présente une teinte qui n'est partagée avec aucune autre classe (dans les tons médians et très clairs). Cet indice est donc le plus discriminant.

5. L'intégrale d'une fonction discrète peut être calculée de différentes façons : comme la distance de chaque point à l'axe des abscisses ou encore comme l'aire sous la courbe interpolée de cette fonction discrète (selon différents degrés). Dans cette expérience, nous avons choisi d'utiliser la définition classique de l'intégrale, correspondant à la somme des distances de chaque élément du chemin à sa projection orthogonale sur la diagonale.

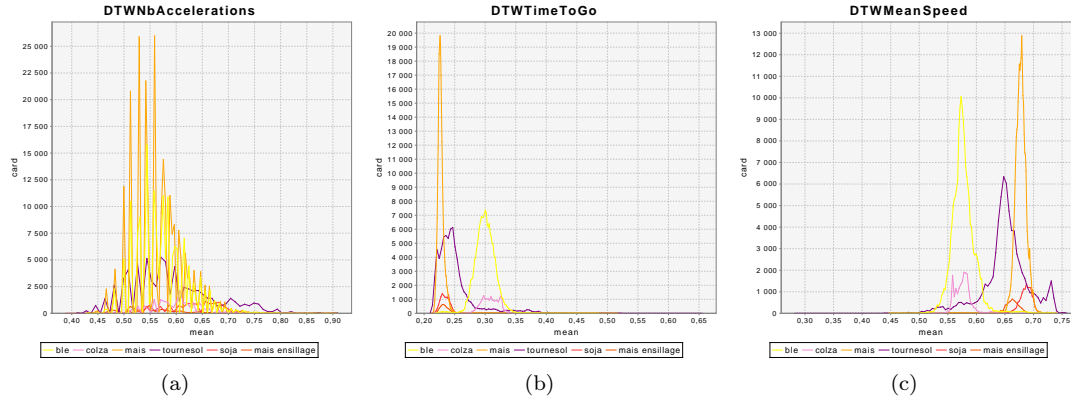


FIGURE 13 – Histogrammes de répartition des classes correspondant aux indices dérivés de la simulation de parcours du chemin par un solide : (a) nombre d’accélération sur le chemin, (b) temps de parcours du chemin et (c) vitesse moyenne le long du chemin.

2 d) Dérivée / Parcours du chemin

La Figure 13 illustre les indices dérivés de la simulation de parcours du chemin par un solide.

Le nombre d’accélération illustré en Figure 13(a) est très proche du nombre de d’associations illustré en Figure 10, ce qui paraît cohérent car :

$$\text{Nb Associations} = \text{Nb Accelerations} + \text{Nb Decelerations} + \text{Nb Diagonales} \quad (3)$$

De plus, le nombre d’accélération, de décélérations et de diagonales⁶ sont également liés car la longueur du chemin est bornée. Aussi, connaître le nombre d’accélération détermine-t-il directement le nombre de décélérations ainsi que de diagonales, en fonction de la taille de la matrice.

Le temps de parcours du chemin est illustré en Figure 13(b). Cet indice est assez intéressant car, même s’il confond la culture de maïs avec la culture de tournesol, les cultures de blé et de colza sont plus nettement séparées des autres cultures qu’avec les précédents indices.

La vitesse moyenne du parcours est illustrée en Figure 13(c). Ce descripteur porte une information assez proche de l’intégrale signée sous la diagonale, ce qui est cohérent puisqu’une intégrale négative (ou positive en fonction du signe choisi) correspond à des vitesses de parcours faibles et inversement. En revanche, le tournesol est moins séparé du maïs que dans ce précédent.

Conclusion

Nous avons vu différents indices permettant de mettre en évidence différentes classes de comportements. Nous avons également vu que plusieurs indices contiennent des informations similaires. Les indices apparaissant les plus intéressants dans cette étude sont l’intégrale par rapport à la diagonale et le temps de parcours. Les autres indices se sont révélés être soit fortement corrélés à ces deux indices, soit non-discriminant pour cette étude.

6. Par “diagonales”, nous entendons ici le nombre de transitions dans la matrice empruntant la diagonale de la case.

IV Descripteurs temporels d'alignements

1 Théorie

Nous avons vu dans la section précédente que les résultats sont intéressants et permettent de caractériser la “forme” des alignements entre deux séquences. Cependant, nous avons vu que les “retard”, “avance”, “décalages” et “distorsions”, ne peuvent être qualifiés qu’en terme de *nombre de cases/barreaux*. Cette dé-temporalisation des données est inhérente à DTW et en fait à la fois une force, lui permettant d’admettre des séries à échantillonnage irrégulier ainsi que des séries de longueurs différentes, et une faiblesse en termes de caractérisation temporelle des phénomènes. Aussi, même s’il est difficile de modifier l’équation de DTW afin de prendre en compte les dates dans le calcul sans perdre ses propriétés (paramétrisation, flexibilité)⁷, est-il possible d’imaginer plonger le chemin d’alignement dans un espace temporel daté. En effet, dans les sections précédentes, le chemin d’alignement a été étudié directement dans la matrice; ses coordonnées correspondaient aux indices dans les matrice. Aussi la fonction n’évoluait-elle que par pas de un sur ses deux axes (simultanément ou non). Par ailleurs, chaque point de la fonction correspond à la mise en correspondance *i.e.*, à l’alignement de deux éléments provenant de deux séquences. Or, ces deux éléments correspondent à l’état d’une zone géographique, acquis à une certaine date. Aussi est-il possible de décrire le chemin d’alignement, non-plus par les indices dans la matrice, mais par les différentes dates.

La Figure 14 montre deux séquences alignées par DTW et la forme du chemin d’alignement plongé ou non dans l’espace *temporalisé* des séquences; espace induit par les dates d’acquisitions des différents éléments des deux séquences. De façon évidente, il apparaît que le chemin d’alignement plongé, non-plus dans un espace séquentiel (Figure 14(a)) mais dans un espace temporel (Figure 14(b)), n’a pas la même forme.

1 a) Utiliser les dates d’acquisition des images

Dans le cadre de notre étude, étudier ce chemin temporalisé permettrait de réduire les sur- et sous-évaluations des indices, suite à des différences de dates. Considérons par exemple deux parcelles de blé Ble^1 et Ble^2 ayant une date de maturation identique, par exemple le 15 mai, et dont nous étudions l’évolution radiométrique. Nous disposons pour Ble^1 d’une image en avril, d’une image en mai et de deux images en juin, et pour Ble^2 d’une image en avril, de deux images en mai et d’une image en juin. En comparant ces deux séquences avec DTW, l’alignement entre les différents stades de maturations va s’opérer correctement. En revanche, le chemin d’alignement va montrer un décalage constant de deux séquences à partir du mois de mai. Aussi, l’étude de ce chemin d’alignement va-t-il nous mener à la conclusion que la maturation des deux parcelles est différente, puisque le chemin ne suit pas la diagonale de la matrice. Or, la phénologie de ces parcelles étant identique, cette conclusion est erronée. Cependant, une fois le chemin d’alignement plongé dans l’espace temporel des séquences, le décalage sera effacé et alors beaucoup plus proche de la diagonale. Le but de cette temporalisation du chemin d’alignement est donc de corriger les erreurs commises par le calcul des indices lorsque l’échantillonnage des séquences n’est pas régulier. Il nous est ainsi possible de profiter de la flexibilité de DTW induite par le calcul de l’alignement dans un espace séquentiel non daté, tout en exploitant la forme des alignements dans un espace temporel, par l’étude du chemin d’alignement plongé dans un espace tenant compte des dates d’acquisitions des images.

Dans la suite de ce document, nous parlerons de descripteurs atemporels ou temporels pour désigner les descripteurs tenant compte ou non des dates d’acquisitions des séquences.

7. Nos précédents travaux ont néanmoins montré qu’il est possible de contraindre l’alignement de DTW dans un sous-ensemble de la matrice.

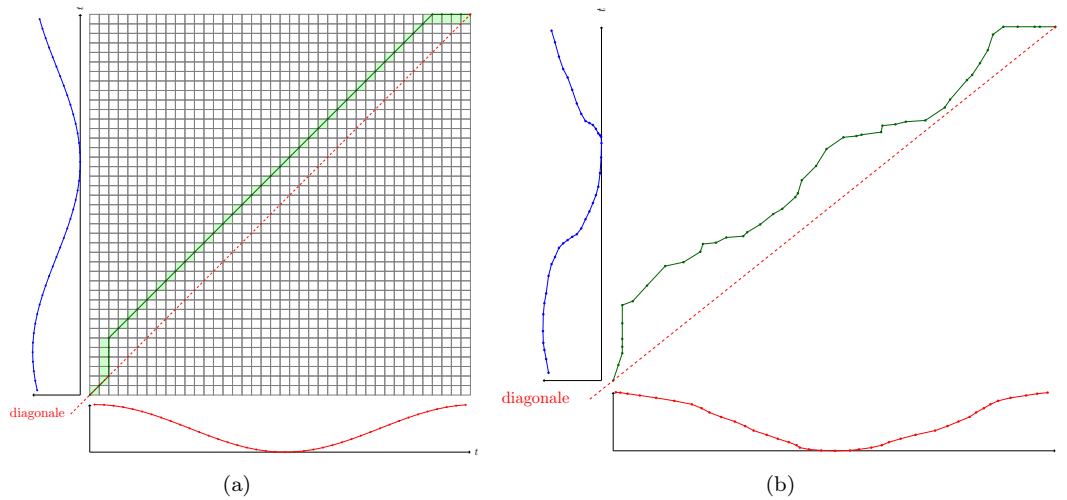


FIGURE 14 – Temporalisation d’un chemin d’alignement : (a) illustre un chemin induit par les indices de la matrice et (b) son équivalent plongé dans l’espace temporel des séquences, induit par les dates d’acquisitions des différents éléments des séquences. *Nb* : Les échelles de temps des deux séquences sont différentes, ce qui peut être remarqué par la forme des sinusoïdes en Figure (b).

1 b) Le problème du centroïde atemporel

En pratique, les indices précédemment décrits peuvent être appliqués à des séquences dont l’échantillonnage temporel est irrégulier. Cependant, dans notre cas, les indices ne sont pas calculés entre différentes séquences de la zone observée, mais entre chaque séquence et le profil moyen d’évolution des cultures (toutes cultures confondues). De plus, alors que chaque élément de chaque séquence du jeu de données correspond à un pixel acquis à une date donnée, les éléments de la séquence moyenne n’ont pas de dates associées, puisqu’ils correspondent à des barycentres moyens de différentes valeurs acquises à différentes dates. Aussi est-il nécessaire de temporaliser les éléments constituant la séquence moyenne afin de pouvoir exploiter plus facilement l’information temporelle des séquences du jeu de données. Bien qu’imparfaite, la solution retenue consiste à calculer une date moyenne pour chaque élément de la séquence moyenne, en fonction des éléments des séquences y ayant été associés par DTW⁸. Par ailleurs, l’étude actuelle se focalisant sur une série d’images annuelle, l’échelle de temps retenue correspond au nombre de jours écoulés depuis le début de l’année.

8. \triangle Cette solution permet, de par sa simplicité, une utilisation directe du principe décrit précédemment. Un autre choix eut été de conserver l’ensemble des dates associées à chaque élément de la moyenne. Un tel choix aurait cependant compliqué la description des chemins d’alignements, puisqu’un chaque élément du chemin d’alignement aurait été décrit de façon floue sur une des deux dimensions temporelles. La date moyenne constitue donc une solution plus simple, et reste de plus cohérente avec le principe de fonctionnement voulu. Il est cependant important de noter que cette solution peut aboutir à un phénomène dégénératif : un élément de la moyenne peut précéder un autre élément mais avoir pourtant une date postérieure à ce-dernier. En pratique, cela peut conduire à un chemin d’alignement dont la monotonie n’est plus garantie. Ce type de phénomène ne représente cependant pas le cas général, et ne modifie pas la définition ni la pertinence des descripteurs.

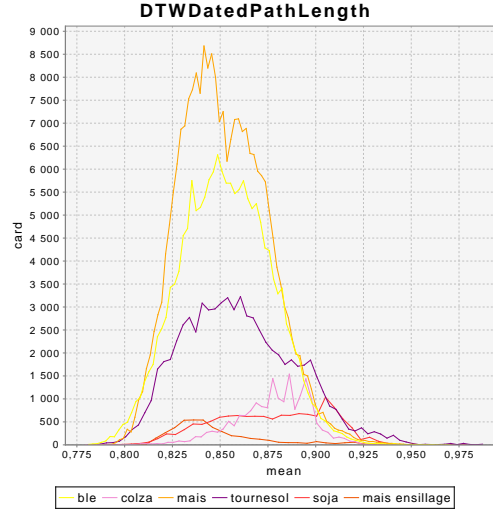


FIGURE 15 – Carte résultat du descripteur correspondant à la longueur du chemin temporalisé, normalisé par la longueur maximale.

2 Descripteur temporels d'alignements

Puisque les indices ne subissent que des modifications mineures de principe par rapport à leurs versions atemporelles, la présentation des indices et de leurs résultats se fera dans cette même partie.

2 a) Longueur du chemin / Nombre d'alignements

La version atemporelle du nombre d'alignements est presque identique à la longueur du chemin en terme de pouvoir discriminant, c'est pourquoi nous n'avons présenté que le nombre d'alignements. Or, le nombre d'alignements étant identique que ce soit dans le cas temporel ou dans le cas atemporel, nous présentons ici le descripteur basé sur la longueur du chemin, différant entre les deux versions.

La Figure 15 présente l'histogramme de répartition de l'indice par rapport aux classes de cultures. Comme dans le cas atemporel, cet indice ne permet pas de discriminer les différentes classes de culture.

2 b) Page Rank

Le descripteur inspiré du PageRank est identique à sa version atemporelle.

2 c) Intégrale du chemin

Les indices temporels liés à la position du chemin par rapport à la diagonales diffèrent en deux points :

- les coordonnées des éléments constitutifs du chemin d'alignement sont plongés dans un espace séquentiel ;

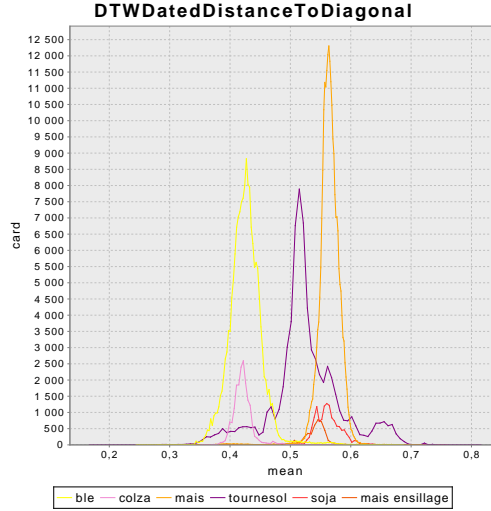


FIGURE 16 – Histogrammes de répartition des classes correspondant à l’intégrale du chemin d’alignement temporelisé par rapport à la diagonale.

- la diagonale n’a pas pour coordonnées les extrémités de la matrice, mais les points (1, 1) et (365, 365) ; la diagonale représentant l’axe de synchronisation.⁹

Par ailleurs, l’information de distance moyenne étant un condensé de l’information d’intégrale de la fonction, nous ne présentons pas la distance moyenne à la diagonale. De plus, l’indice d’écart-type étant fortement corrélés, par construction, à la distance moyenne, sa version temporelle n’est pas détaillée.

L’histogramme de la version temporelle de l’intégrale de la fonction par rapport à la diagonale est donné en Figure 16. L’étude de cet histogramme ne montre pas de différences significative avec la version atemporelle de l’indice. Cette similitude entre les deux versions s’explique par la fréquence d’échantillonnage relativement élevée en 2006, comme le montre la Figure 17(a). La différence est en effet plus sensible lorsque la fréquence d’échantillonnage temporel diminue, comme par exemple en 2007 (voir Figure 17(b)). Les résultats des deux versions de l’indice correspondant à l’intégrale sous le chemin pour l’année 2007 sont présentés en Figure 18. On peut remarquer dans ce cas que la version temporalisée de l’indice a tendance à affiner les histogrammes des classes, *i.e.*, à réduire leurs variances ; les classes sont ainsi mieux séparées dans l’espace.

2 d) Dérivée / Parcours du chemin

L’extension des indices atemporels à leurs versions temporelles n’est pas immédiate. Aussi, le calcul du temps de parcours dans sa version temporalisée est-il détaillé en Annexe C ; la vitesse moyenne correspondant quant à elle à diviser le temps de parcours par la longueur du chemin temporalisé. Ces résultats sont présentés en Figure 19. On peut remarquer sur ces résultats qu’ils sont différents de ceux obtenus avec leurs versions atemporelles, mais qu’ils ne permettent pas seuls de séparer les différentes cultures. Cependant, la distribution du tournesol pour l’indice de vitesse moyenne est assez intéressante car elle diffère des distributions obtenues jusqu’alors.

9. Par abus de représentation, la Figure 14(b) représentait la diagonale dans l’axe de la matrice, afin de faciliter la lecture de ce premier graphique de temporalisation.

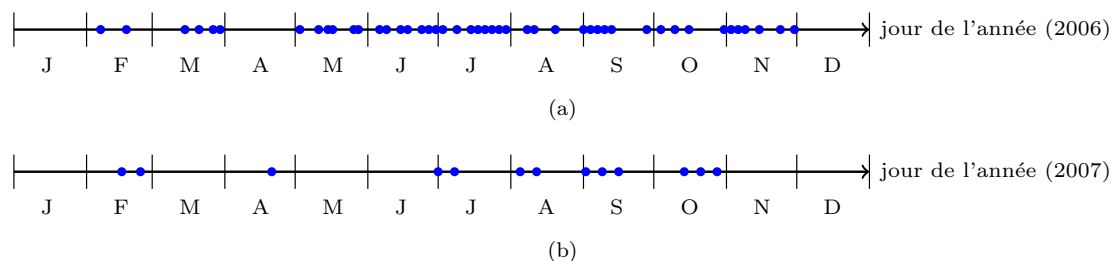


FIGURE 17 – Échantillonnage des séries d’images satellites des années 2006 et 2007 ; chaque point représente une image acquise dans l’année.

Cet indice pourrait donc, en complément, être utile pour discriminer les différentes cultures. Par ailleurs, les mêmes observations ont pu être faites sur des séries à faible fréquence d’échantillonnage, que pour l’indice basé sur l’intégrale du chemin d’alignement, à savoir que la pertinence que l’amélioration de la version temporelle augmente lorsque les séries sont moins fournies.

3 Conclusion sur les descripteurs temporels d’alignements

Nous avons vu que les versions temporelles des différents indices modifient légèrement les résultats obtenus avec les versions séquentielles (ou atemporelles) des indices. De plus, nous avons vu que l’amélioration liée à la temporalisation des indices est d’autant plus notable que l’échantillonnage des séries est faible et irrégulier.

Par ailleurs, même si les résultats temporels ne diffèrent pas beaucoup des résultats atemporels pour la série de l’année 2006, dont la fréquence d’échantillonnage est élevée, la justesse de ces indices est un argument important en leur faveur. En effet, un point du chemin d’alignement correspondant à l’alignement de deux valeurs, les indices atemporels ne mesurent pas la proximité temporelle d’acquisition de ces deux valeurs. Aussi, ces indices retranscrivent-ils la même information, que ces deux valeurs aient été acquises à un mois d’intervalle ou à un jour. De ce fait, les indices temporels devraient être utilisés de préférence pour l’analyse de profils radiométriques à échantillonnage irréguliers. Ce choix est de plus motivé par les résultats obtenus sur les séries à échantillonnage plus faible et plus irrégulier de 2007.

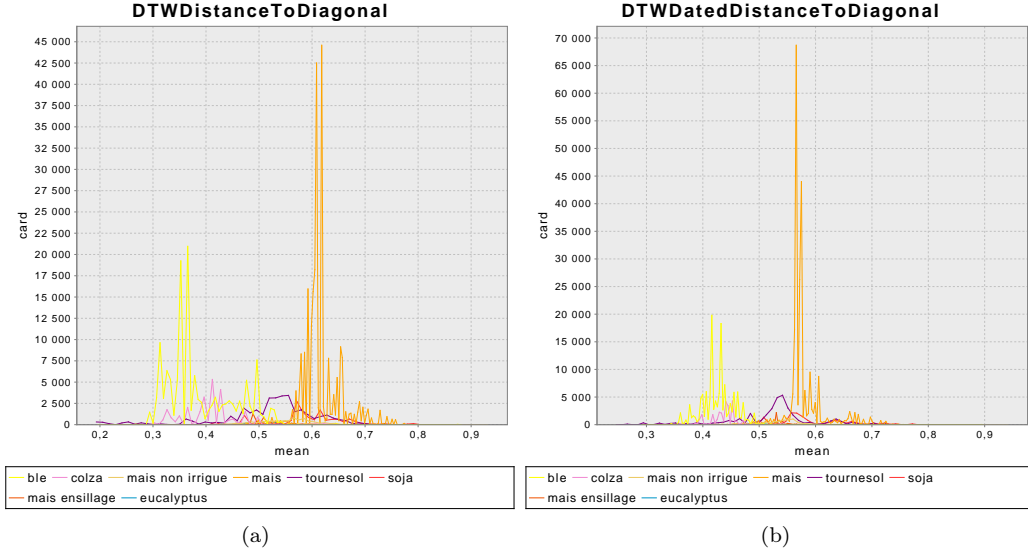


FIGURE 18 – Histogrammes de répartitions des classes correspondant aux deux versions de l'intégrale sous le chemin d'alignement. En (a) et (b) sont respectivement présentés les résultats de la version atemporelle et de la version temporelle. Ces indices sont calculés sur la série d'images acquise en 2007.

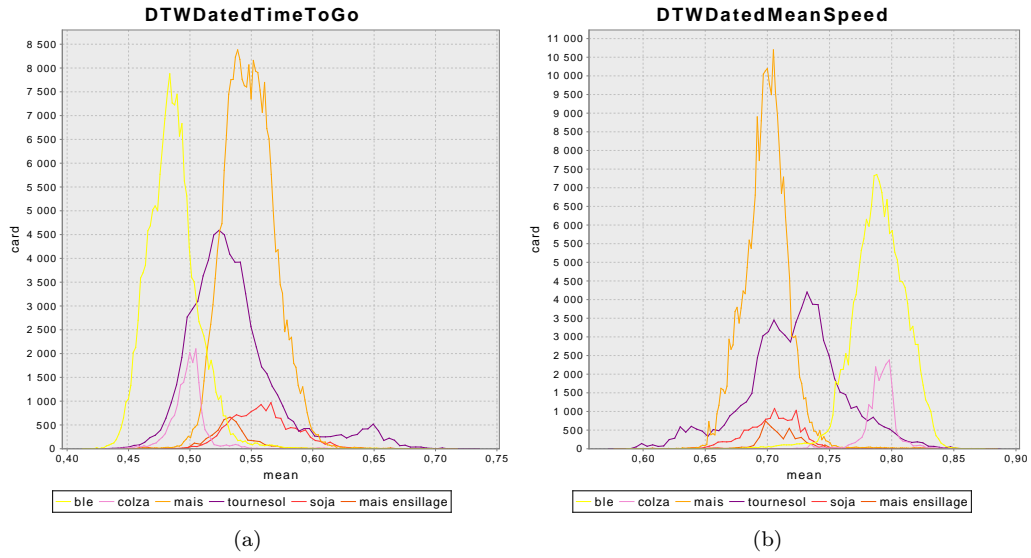


FIGURE 19 – Histogrammes de répartition des classes liés au parcours d'un solide sur le chemin d'alignement, plongé dans son espace temporel. (a) Temps de parcours et (b) Vitesse moyenne pendant le parcours.

V Conclusion générale sur la description des alignements

La problématique traitée dans ce document pourrait être formulée comme suit :

« Comment utiliser l’information contenue dans l’alignement construit par DTW de deux séquences d’évolution, afin de différencier des comportements identiques au sens du coût de l’alignement ? ».

Une objection intervient alors directement, puisque si l’objectif est de discriminer des comportements en fonction de leurs distorsions temporelles, DTW n’est pas la bonne mesure de similarité et la distance euclidienne devrait être préférée. Cependant, deux argument important motivent l’utilisation de DTW.

1. Les données que nous traitons ne peuvent pas être comparées avec la distance Euclidienne car l’échantillonnage (impliquant la longueur) de deux séquences à comparer peut être différent.
2. Même si deux séquences ont exactement le même échantillonnage, l’espace induit par la distance euclidienne n’est pas structuré par le temps et n’exploite donc pas l’information temporelle. Il n’y a donc une garantie que de différencier la fréquence des occurrences, et non des distorsions temporelles.

Aussi, notre étude a-t-elle montré que l’alignement de deux séquences peut être décrit à partir de deux représentations : sous forme de graphe, ou sous la forme de chemin d’alignement dans la matrice.

La représentation sous la forme du chemin d’alignement nous a permis de développer les indices les plus intéressants, même si la représentation sous forme de graphe a séparé les valeurs de façon particulière.

La seconde partie de cette étude s’est intéressé au ré-ajustement des indices lorsque l’échantillonnage n’est pas régulier dans le temps. La représentation du chemin d’alignement n’a donc plus été faite dans l’espace séquentiel induit par la succession des valeurs, mais dans l’espace temporel induit par les dates d’acquisition des valeurs. La topologie induite par le séquençement des valeurs a donc été plongée dans un espace “géométrique” induit par les dates des acquisitions des différentes valeurs. Les résultats se sont avérés cohérents avec les résultats précédents et la correction est apparue d’autant plus utile que l’échantillonnage des séquences est irrégulier.

La Figure 20 en page 21 illustre une composition colorée de trois indices. Cette figure montre que les indices sont porteurs d’informations discriminantes pour la séparation des différentes cultures. Et, même si le découpage n’est pas parfait, ce type de carte constitue une description synthétique des 46 images en terme de comportement temporel et permet une analyse prospective rapide des données.

Enfin, il est intéressant de noter que les indices utilisant la fonction discrète sans interpolation (ex : intégrale du chemin) se sont avérés plus pertinents que ceux utilisant une interpolation linéaire de celle-ci (ex : dérivée). De fait, interpoler le chemin revient à interpoler les alignements et donc à considérer un alignement continu des différentes portions des séquences. Or, ceci peut s’avérer erroné lorsque les séquences subissent des changements abrupts, comme par exemple dans le cas de la récolte d’un champs. Ainsi, dans le cas où l’on disposerait d’une image avant une récolte d’une parcelle et d’une image après cette récolte, considérer un chemin interpolé reviendrait-il à considérer que le passage de l’état “culture levée” à l’état “sol nu” s’est fait progressivement, ce qui n’est pas le cas. L’argumentation en faveur de la non-interpolation des informations manquantes, souvent soutenue en géosciences, apparaît encore une fois pertinente dans cette étude.

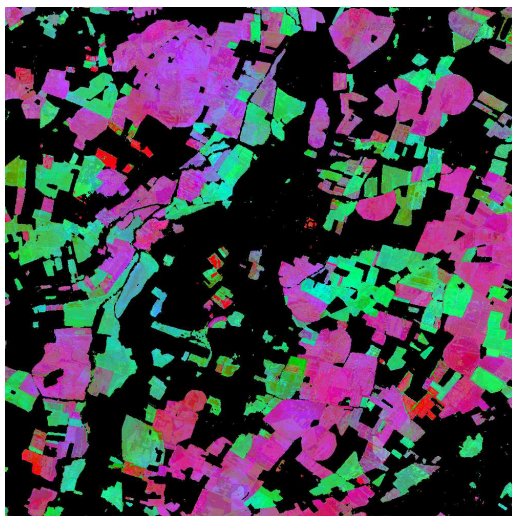


FIGURE 20 – Composition colorée des trois indices temporels : le rouge correspond à l'intégrale sous la diagonale, le vert correspond à la vitesse moyenne du solide sur le chemin d'alignement et le bleu correspond au PageRank.

A Cartes correspondant aux indices

1 Indices atemporels

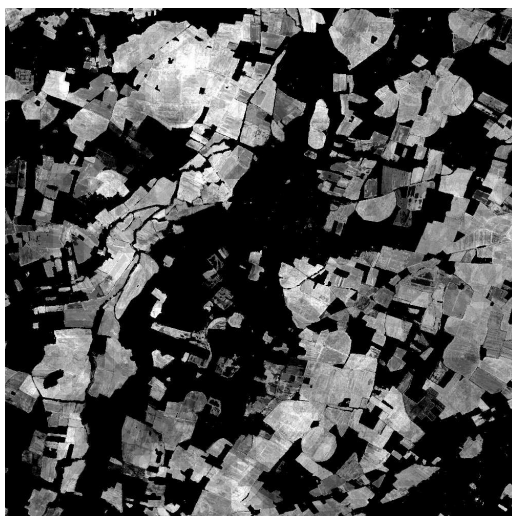


FIGURE 21 – Carte résultat du descripteur correspondant au PageRank.

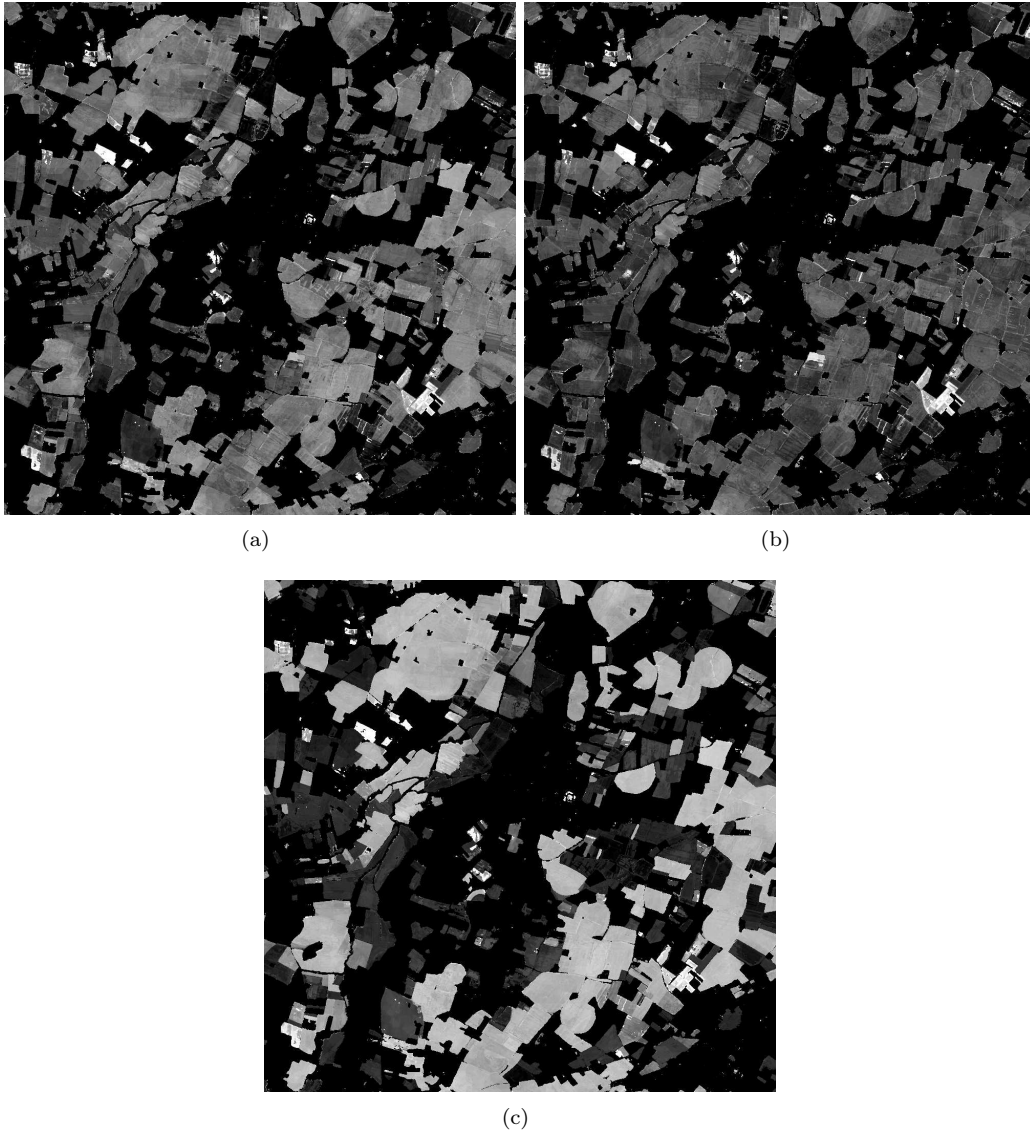


FIGURE 22 – Cartes résultat correspondant aux indices dérivés de la distance à la diagonale : (a) distance moyenne à la diagonale, (b) écart-type de la distance à la diagonale et (c) intégrale signée de la courbe par rapport à la diagonale.

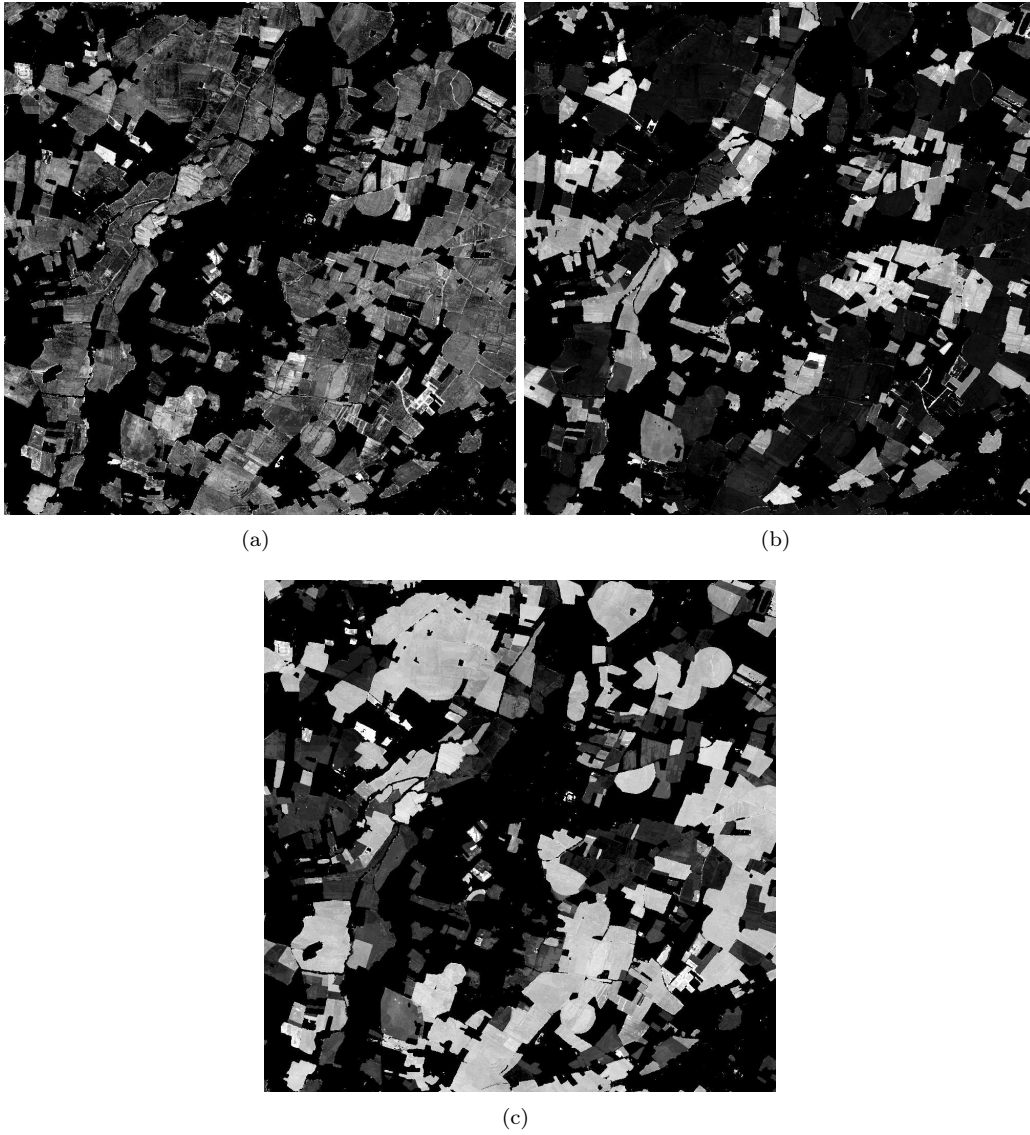


FIGURE 23 – Cartes résultat correspondant aux indices dérivés de la vitesse de parcours du chemin : (a) nombre d'accélération sur le chemin, (b) temps de parcours du chemin et (c) vitesse moyenne le long du chemin.

2 Indices temporels

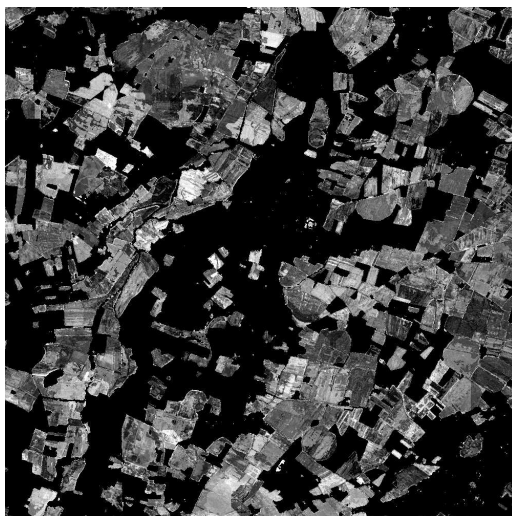


FIGURE 24 – Carte résultat du descripteur correspondant à la longueur du chemin temporalisé, normalisé par la longueur maximale.

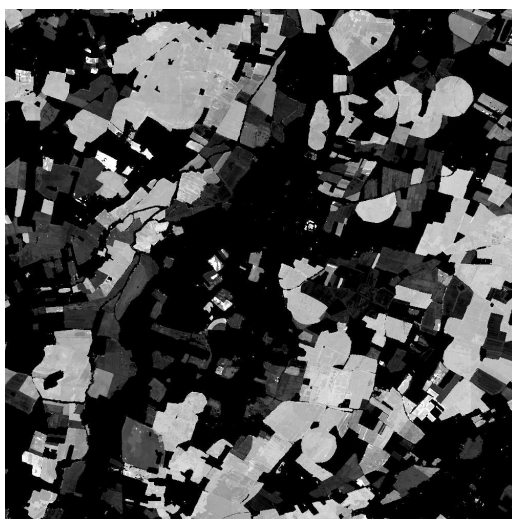


FIGURE 25 – Carte résultat du descripteur correspondant à l'intégrale du chemin d'alignement temporalisé par rapport à la diagonale.

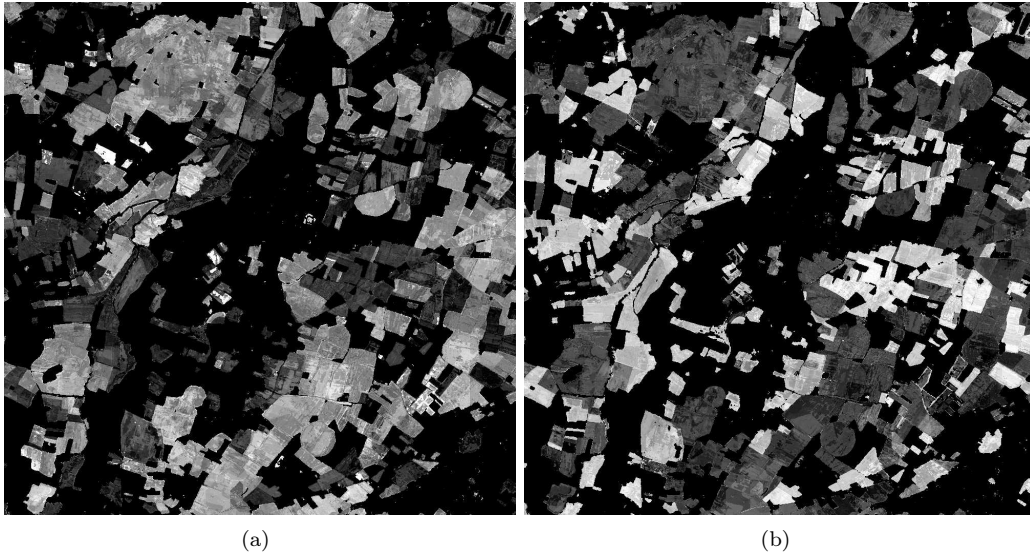


FIGURE 26 – Cartes résultat correspondant aux indices liés au parcours d’un solide sur le chemin d’alignement, plongé dans son espace temporel. (a) Temps de parcours et (b) Vitesse moyenne pendant le parcours.

B Séparation de référence

La Figure 27 illustre la distribution du score de DTW dans la comparaison des séquences à la séquence moyenne. On peut remarquer deux points : premièrement, la valeur moyenne de la distance est assez faible, ce qui indique la pertinence de la méthode de moyenne utilisée DBA [1]. Deuxièmement, la répartition des séquences autour de la séquence moyenne est assez uniforme ; ne permettant pas de discriminer un type particulier de culture qui se situerait également réparti en “cercle” autour de la séquence moyenne.

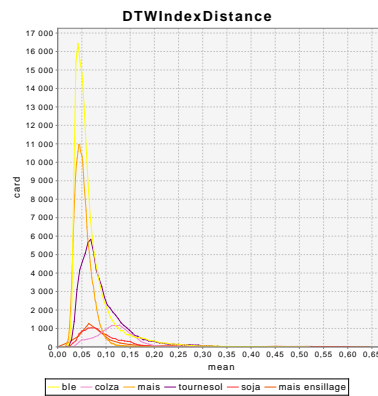


FIGURE 27 – Histogramme de répartition des valeurs de la distance DTW par rapport aux différentes classe.

C Temps de parcours temporel : détail du calcul

1 Principe

Sur chaque segment du chemin d'alignement, le solide virtuel suit un mouvement rectiligne uniformément accéléré. Et, contrairement à sa version atemporelle, il existe plus de trois types de plans inclinés¹⁰ et la longueur de ces plans varie. Aussi est-il nécessaire d'utiliser les théories issues de la cinématique pour construire cet indice. Voici les différentes étapes du calcul de cet indice :

1. calculer le chemin d'alignement ;
2. plonger le chemin dans l'espace temporel (et non-plus séquentiel) ;
3. faire une rotation du chemin pour que l'axe des abscisses soit confondu avec la diagonale¹¹ ;
4. calculer la vitesse initiale du solide, lui permettant de franchir tous les chemins possibles ;
5. calculer le temps de parcours du chemin (calcul des accélérations pour chaque segment, puis du temps de parcours segment après segment) ;
6. normaliser par rapport au temps maximum.

2 Chemin d'alignement : calcul, plongement et rotation

Le chemin d'alignement est calculé classiquement en utilisant DTW. Une fois calculé, chaque point de jonction (x, y) sur le chemin d'alignement est plongé dans l'espace temporel associé : x est associé à la $x^{\text{ième}}$ date de la première séquence et y est associé à la $y^{\text{ième}}$ date de la deuxième séquence. On effectue ensuite une rotation de 45° dans le sens horaire de tous les points du chemins.

3 Rappels de cinématique

Rappelons que l'accélération sur un plan incliné de pente α est donné par :

$$a = g \cdot \sin(\alpha) \quad (4)$$

avec g l'accélération de la pesanteur. Rappelons également que, dans le cas d'un mouvement rectiligne uniformément accéléré, le solide est régi par :

$$v(t) = a \cdot t + v_0 \quad (5)$$

$$x(t) = \frac{1}{2} \cdot a \cdot t^2 + v_0 \cdot t \quad (6)$$

4 Déterminer v_0

Comme nous l'avons précédemment évoqué, la vitesse initiale du solide doit être suffisante pour pouvoir franchir le plan incliné correspondant au bord supérieur de la matrice. Ce bord, une fois plongé dans l'espace temporel et reprojété par rapport à la diagonale, est incliné d'un angle α avec l'axe des abscisses (cet axe est confondu avec la diagonale dans notre référentiel). La

10. La version atemporelle était une simplification du cas général puisque le solide pouvait soit descendre sur un plan à 45° , soit monter sur un plan à 45° , soit être sur un plan. De plus, nous avons considéré que la longueur du plan était constante, ce qui permettait de simplifier le calcul de l'indice, en incrémentant ou décrémentant simplement la vitesse à chaque segment.

11. En pratique, la diagonale forme toujours un angle de 45° avec l'axe des abscisses puisqu'elle passe par les points de coordonnées $(1, 1)$ et $(365, 365)$.

valeur de cet angle peut être calculé en utilisant sa tangente. La longueur du palier est également calculée¹². Nous cherchons donc v_0 telle que la vitesse au sommet du plan incliné soit nulle. Soit d la longueur du palier, et a_α la valeur de l'accélération sur ce palier dont la pente est α . Trouver v_0 consiste à résoudre le système de deux équations à deux inconnues (t et v_0) suivant :

$$\begin{cases} \frac{1}{2} \cdot a_\alpha \cdot t^2 + v_0 \cdot t = d & (7) \\ a_\alpha \cdot t + v_0 = 0 & (8) \end{cases} \Leftrightarrow \begin{cases} \frac{1}{2} \cdot a_\alpha \cdot t^2 + v_0 \cdot t = d & (9) \\ t = -\frac{v_0}{a_\alpha} & (10) \end{cases}$$

$$\Leftrightarrow \begin{cases} \frac{1}{2} \cdot a_\alpha \cdot \left(-\frac{v_0}{a_\alpha}\right)^2 - v_0 \cdot \frac{v_0}{a_\alpha} = d & (11) \\ t = -\frac{v_0}{a_\alpha} & (12) \end{cases} \Leftrightarrow \begin{cases} v_0 = \sqrt{-2 \cdot d \cdot a_\alpha} & (13) \\ t = -\frac{v_0}{a_\alpha} & (14) \end{cases}$$

$$\Leftrightarrow \begin{cases} v_0 = \sqrt{-2 \cdot d \cdot a_\alpha} & (15) \\ t = -\frac{\sqrt{-2 \cdot d \cdot a_\alpha}}{a_\alpha} & (16) \end{cases}$$

5 Calcul des caractéristiques de chaque segment (pente et longueur)

La pente de chaque segment doit être connue afin de pouvoir en calculer l'accélération de la pesanteur associée. Étant les deux points A et B aux extrémités d'un segment, la pente est donnée par :

$$\alpha = \tan^{-1} \left(\frac{y_B - y_A}{x_B - x_A} \right) \quad (17)$$

d'où la valeur de l'accélération a_{AB} sur ce segment :

$$a_{AB} = g \cdot \sin(\alpha) = g \cdot \sin \left(\tan^{-1} \left(\frac{y_B - y_A}{x_B - x_A} \right) \right) \quad (18)$$

Par ailleurs, la longueur d_{AB} du segment $[AB]$ est donnée par :

$$d_{AB} = \sqrt{(y_B - y_A)^2 + (x_B - x_A)^2} \quad (19)$$

6 Calcul du temps de parcours de chaque segment

Pour calculer le temps de parcours d'un segment $[AB]$, il faut trouver à quel temps t_{arr} le solide sera arrivé à la fin du segment, par rapport au moment où il sera parti. Il s'agit donc de résoudre :

$$\begin{aligned} x(t_{arr}) &= d_{AB} \\ \Leftrightarrow \frac{1}{2} \cdot a_{AB} \cdot t_{arr}^2 + v_{prec} \cdot t_{arr} &= d_{AB} \\ \Leftrightarrow \frac{1}{2} \cdot a_{AB} \cdot t_{arr}^2 + v_{prec} \cdot t_{arr} - d_{AB} &= 0 \end{aligned} \quad (20)$$

Cette équation du second degré a pour discriminant :

$$\Delta = v_{prec}^2 + 2 \cdot a_{AB} \cdot d_{AB} \quad (21)$$

12. Par abus de notation, et ce pour conserver la notation de physique, nous considérons que les longueurs des segments sont exprimés en mètres, même si ces distances correspondent en fait à des durées. Il apparaîtrait en effet perturbant d'exprimer les distances en nombre de jours.

et a donc pour solutions :

$$t_{arr} = \frac{-v_{prec} - \sqrt{\Delta}}{a_{AB}} \text{ ou } t_{arr} = \frac{-v_{prec} + \sqrt{\Delta}}{a_{AB}} \quad (22)$$

Nous retiendrons la solution positive la plus petite des deux¹³. Par ailleurs, dans le cas où l'accélération serait nulle sur le segment (palier), on peut directement calculer t_{arr} tel que :

$$t_{arr} = \frac{d_{AB}}{v_{prec}} \quad (23)$$

Il faudra également calculer la vitesse à du solide à la fin du segment donné directement par :

$$v_{arr} = v(t_{arr}) = a_{AB} \cdot t_{arr} + v_{prec} \quad (24)$$

7 Normalisation

Finalement, le temps de parcours est normalisé par rapport à un temps de parcours maximum correspondant au trajet sur le bord supérieur de la matrice.

Références

- [1] François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for Dynamic Time Warping, with applications to clustering. *Pattern Recognition*, 44(3) :678–693, March 2011.
- [2] H. Sakoe and S. Chiba. A dynamic programming approach to continuous speech recognition. In *Proceedings of the Seventh International Congress on Acoustics*, volume 3, pages 65–69, 1971.
- [3] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1) :43–49, 1978.

13. Deux solutions positives peuvent en effet exister dans le cas d'une accélération négative avec une vitesse élevée.