

EXTENSIONS DE LA DISTANCE LINEAIRE DE FRECHET ET ALGORITHMES D'APPARIEMENT

Sommaire

II.1	Extension 1 : distance moyenne de Fréchet et couples de points homologues..	63
II.2	Extension 2 : distance de Fréchet sur des lignes d'emprises ou de natures différentes.....	67
II.2.1	Distance de Fréchet discrète partielle.....	67
II.2.2	Distance de Fréchet discrète entre deux lignes fermées	68
II.2.3	Distance de Fréchet discrète partielle entre une ligne ouverte et une ligne fermée.....	72
II.2.4	Distance de Fréchet discrète partielle / partielle entre deux lignes ouvertes	73
II.3	Algorithme d'appariement.....	78
II.3.1	Recherche de la nature des lignes comparées.....	79
II.3.2	Relations d'emprises entre les lignes.....	79
II.3.3	Types d'appariement entre lignes.....	81
II.3.3.1	Appariement simple « 1 – 1 » de deux lignes :.....	81
II.3.3.2	Appariement total/partiel « 1 – P » entre deux lignes.....	82
II.3.3.3	Appariement « partiel/partiel » (PP) et « deux partiel/partiel » (2PP) entre deux lignes.....	82
II.3.4	Méthode globale : appariement par fichier.....	86
II.4	Exemple d'applications de la méthode sur la comparaison de traits de côte ...	89
II.4.1	Présentation des données et de la zone d'étude.....	89
II.4.2	Méthode de comparaison des traits de côte	89
II.4.3	Résultats et perspectives.....	90
II.5	Bilan sur les apports des extensions.....	93

Le chapitre Précédent a montré que les éléments caractéristiques importants pour l'intégration de MNT étaient essentiellement linéaires. Il faut donc disposer de distances permettant de comparer, appairer puis intégrer ces éléments. L'algorithme retenu pour réaliser l'intégration de données terre/mer à l'issue de cet inventaire est celui basé sur la distance de Fréchet car elle est la mesure la plus appropriée pour comparer des éléments linéaires complexes (thalwegs crêtes, etc.) Celle-ci, introduite par [Fréchet, 1906], se base sur l'assimilation de toute ligne à une suite de points orientés équivalente à une fonction continue. Le résultat est une distance maximale entre les deux lignes comparées. La difficulté à utiliser ce calcul vient de sa complexité à programmer. [Eiter & Mannila, 1994] proposent une solution en développant une estimation de cette distance : la distance de Fréchet discrète.

Cette distance a pour intérêt de réaliser un appariement basé sur les points homologues des lignes comparées et non les points les plus proches. Cependant, telle qu'elle a été définie, elle est insuffisante pour répondre à toutes les attentes et contraintes d'un appariement de données géographiques. En effet l'algorithme de départ ne s'applique qu'à des lignes ouvertes, sensiblement de même taille et d'orientations identiques. Dans le cadre de l'appariement, il est nécessaire de l'adapter afin de pouvoir comparer des lignes d'emprises, de tailles, de natures et d'orientations différentes.

Dans le cas de données réelles, les lignes numérisées (1) peuvent présenter l'aspect de lignes « fermées » (polygones dont les points sont ordonnés) et (2) ne sont pas obligatoirement saisies dans le même sens. Il convient donc de déterminer avant toute chose la « nature » d'une ligne : si le premier point est identique au dernier, alors la ligne représente une ligne fermée. Ensuite il faut également s'assurer des orientations respectives des lignes entre elles. Il s'agit de vérifier que les suites ordonnées de points de chaque ligne ouverte sont bien dans le même sens et que les points de début et de fin sont homologues et non diamétralement opposés.

Les deux extensions développées dans ce chapitre se proposent donc de répondre aux différents problèmes posés lors de la comparaison de lignes en inventoriant les différents cas de figure rencontrés.

Les algorithmes détaillés par la suite expliquent comment utiliser au mieux ces distances lors des processus d'appariement de lignes. La première partie présente succinctement les lignes et les attributs qui sont utilisés dans les processus de comparaison et d'appariement.

La seconde partie détaille toutes les distances introduites pour la comparaison de lignes :

- la distance moyenne tout d'abord, qui permet d'avoir une estimation de l'écart moyen entre les lignes. Elle donne une information complémentaire à la distance maximale définie par Fréchet ;
- la distance de Fréchet partielle ensuite, définie par [Devogele, 2002] pour la comparaison de lignes de tailles variables ;
- la distance de Fréchet discrète entre deux lignes fermées ;
- la distance de Fréchet discrète partielle destinée à comparer des lignes de tailles variables et de natures différentes (ligne ouverte ou fermée) ;
- et enfin la distance de Fréchet discrète partielle calculée entre deux segments de lignes.

La troisième partie se base sur ces distances pour détailler les algorithmes d'appariement. La première section présente les étapes d'identification de la nature des lignes comparées. La seconde présente les différents types d'inclusion qu'il existe entre les emprises des lignes étudiées. Et la dernière section explique quel appariement est effectué selon les différents cas de figure identifiés et quelle distance est adaptée. Une dernière section réalise une synthèse en intégrant ces appariements simples dans un algorithme de traitement global de deux jeux de polygones.

Finalement, la dernière partie sert de validation des algorithmes appliqués au cas de comparaison de lignes en présentant le résultat d'une étude sur le suivi du trait de côte. Cette étude s'est réalisée au moyen d'indicateurs visuels extraits d'images satellites et de la distance de Fréchet pour la quantification des déplacements entre une ligne de référence et une ligne numérisée à une date donnée.

II.1 Extension 1 : distance moyenne de Fréchet et couples de points homologues

La distance de Fréchet discrète (d_{dF}) représente l'écartement maximal entre deux points homologues des lignes comparées. Afin de compléter l'information fournie par cette mesure, d'autres distances, toutes dérivées de la d_{dF} (voir Chapitre I), ont été introduites. La première distance exposée dans cette section, la **distance de Fréchet moyenne** (d_{aF}), permet l'évaluation de l'écartement moyen global entre les points homologues des lignes. Elle est la moyenne des distances euclidiennes entre couples de points issus du **chemin minimum**.

Le « chemin minimum » est composé d'une suite de couples de points $\{(L_{1,1}, L_{2,1}), \dots (L_{1,n}, L_{2,m})\}$ pour lesquels la distance d'écartement entre les points du couple est inférieure ou égale à d_{dF} . Les points ainsi retenus sont appelés **points homologues**.

Pour reprendre l'exemple du maître et de son chien introduit au chapitre II, plusieurs chemins possibles respectent la distance de Fréchet. Parmi ces chemins candidats, le « chemin minimum » constitue celui pour lequel le promeneur et son chien évoluent le plus proche possible l'un de l'autre à chacun de leur déplacement.

Pour calculer ce chemin, les deux matrices de distance et de Fréchet sont nécessaires ainsi que l'opérateur inférieur ou égal ($\leq$) entre deux couples de réels. Cet opérateur se définit ainsi :

$$\begin{aligned} (a,b) \text{ et } (c,d) &\in \mathbb{R}^2 \\ (a,b) \leq (c,d) &\text{ si } a < c \text{ ou si } a = c \text{ et } b \leq d \end{aligned} \quad (1)$$

où « a » et « c » représentent les valeurs de cellules de la matrice de Fréchet (MF). De même « b » et « d » sont les valeurs des cellules de la matrice de distance (MD). L'opérateur inférieur ou égal ($\leq$) va permettre de choisir le couple de points homologues parmi ceux possibles. Pour les lignes de la figure 29, les matrices de distance et de Fréchet ont été calculées dans le **Erreur ! Source du renvoi introuvable.** A partir de ces matrices sont extraits les couples MF(i, j) et MD(i, j) replacés dans la cellule (i, j) du tableau 6.

Le chemin minimum (C_M) de la figure 30 est construit par « retour arrière » (« backtracking ») à travers les matrices. Il débute donc par le couple de points final : $(L_{1,n}, L_{2,m})$. Le couple suivant, $(L_{1,i}, L_{2,j})$, est recherché parmi trois couples candidats, $(L_{1,i-1}, L_{2,j-1})$, $(L_{1,i-1}, L_{2,j})$ et $(L_{1,i}, L_{2,j-1})$, à l'aide de l'opérateur $\leq$ défini précédemment. La valeur de la MF associée au couple de points considéré signifie que pour aller du couple de départ au couple considéré, la distance de Fréchet vaut MF. Le couple retenu parmi les trois possibles est celui qui permet d'utiliser le sous-chemin le plus petit possible.

Par exemple entre les trois couples $(L_{1,7}, L_{2,7})$, $(L_{1,7}, L_{2,6})$, $(L_{1,8}, L_{2,6})$ de la figure 29, le couple $(L_{1,7}, L_{2,7})$ n'est pas retenu car $(MF_{7,7}) > (MF_{6,7})$ et $(MF_{6,8})$. Entre les deux autres couples, la valeur de la distance de MF est égale ce qui signifie, si l'analogie du maître et de son chien est

reprise, que les deux chemins passant par ces couples utilisent une laisse de même longueur. Dans ce cas, le couple choisi est celui dont les points sont les plus proches (dans la matrice de distance MD) : $(MD_{6,7}) < (MD_{6,8})$.

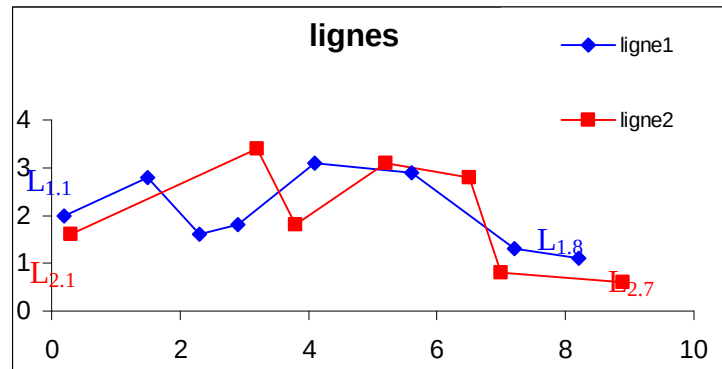


Figure 29 : Couple de lignes homologues.

Tableau 5 : Matrice de distance (euclidienne) et matrice de Fréchet des lignes de la figure 29

x1i	0.2	1.5	2.3	2.9	4.1	5.6	7.2	8.2
y1j	2	2.8	1.6	1.8	3.1	2.9	1.3	1.1

		Matrice de Distance (MD)							
		1	2	3	4	5	6	7	8
1	0.3	0.41	1.70	2.00	2.61	4.09	5.46	6.91	7.92
2	3.2	3.31	1.80	2.01	1.63	0.95	2.45	4.52	5.50
3	3.8	3.61	2.51	1.51	0.90	1.33	2.11	3.44	4.46
4	5.2	5.12	3.71	3.26	2.64	1.10	0.45	2.69	3.61
5	6.5	6.35	5.00	4.37	3.74	2.42	0.91	1.66	2.40
6	7	6.91	5.85	4.77	4.22	3.70	2.52	0.54	1.24
7	8.9	8.81	7.72	6.68	6.12	5.41	4.02	1.84	0.86

		Matrice de Fréchet (MF)							
		1	2	3	4	5	6	7	8
1	0.3	0.41	1.70	2.00	2.61	4.09	5.46	6.91	7.92
2	3.2	3.31	1.80	2.01	2.00	2.00	2.45	4.52	5.50
3	3.8	3.61	2.51	1.80	1.80	1.80	2.11	3.44	4.46
4	5.2	5.12	3.71	3.26	2.64	1.80	1.80	2.69	3.61
5	6.5	6.35	5.00	4.37	3.74	2.42	1.80	1.80	2.40
6	7	6.91	5.85	4.77	4.22	3.70	2.52	1.80	1.80
7	8.9	8.81	7.72	6.68	6.12	5.41	4.02	1.84	1.80

Le processus est itéré selon l'algorithme décrit figure 30 jusqu'au couple de départ $C_{1,1}$. Le chemin minimum obtenu est représenté dans le

tableau 6 sous forme de cellules grisées contenant les couples de réels. C'est une suite ordonnée de neuf couples (voir

tableau 6) : $(C_M) = \{C_{1,1} (L_{1,1}, L_{2,1}), C_{2,2} (L_{1,2}, L_{2,2}), C_{3,3} (L_{1,3}, L_{2,3}), C_{3,4} (L_{1,4}, L_{2,3}), C_{4,5} (L_{1,5}, L_{2,4}), C_{4,6} (L_{1,6}, L_{2,4}), C_{5,6} (L_{1,6}, L_{2,5}), C_{6,7} (L_{1,7}, L_{2,6}), C_{7,8} (L_{1,8}, L_{2,7})\}$.

Cette programmation dynamique avec l'opérateur $\leq$ donne des résultats plus rapidement que dans celle proposée par [Devogele, 2002]. En effet dans [Devogele, 2002], l'ensemble des chemins possibles sont testés avant de choisir celui qui est le plus optimal, ce qui peut être très long pour des lignes comportant énormément de points. Avec la méthode par backtracking présentée ici, seules $(N+M)$ comparaisons au maximum sont effectuées.

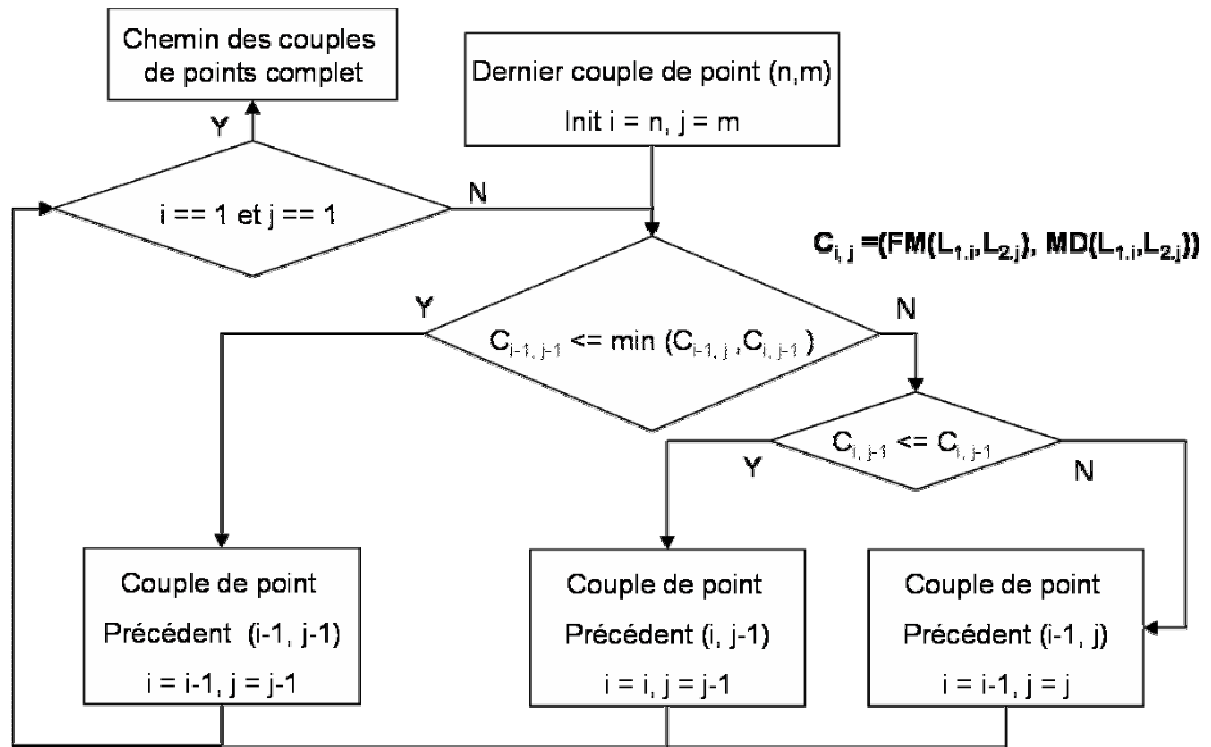
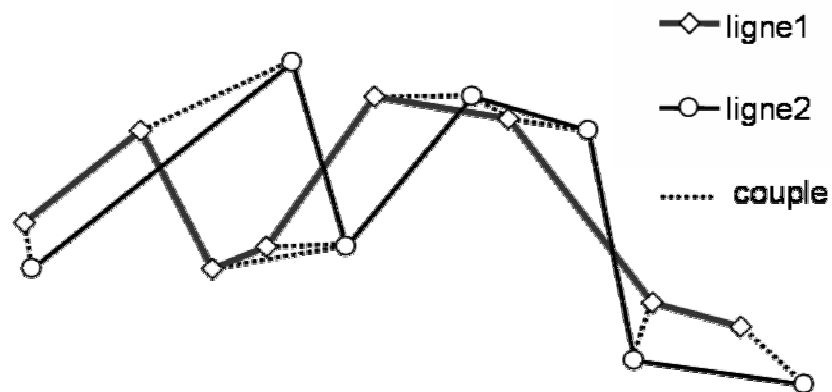


Figure 30 : Algorithme pour le calcul du chemin minimum.

Tableau 6 : Le « chemin minimum » des lignes $(L_{1,1}, L_{2,1})$ et $(L_{1,m}, L_{2,m})$ est défini en sélectionnant les couples de réels en grisés.

	1	2	3	4	5	6	7	8
1	0.41 0.41	1.70 1.70	2.00 2.00	2.61 2.61	4.09 4.09	5.46 5.46	6.91 6.91	7.92 7.92
2	3.31 3.31	1.80 1.80	2.01 2.01	2.00 1.63	2.00 0.95	2.45 2.45	4.52 4.52	5.50 5.50
3	3.61 3.61	2.51 2.51	1.80 1.51	1.80 0.90	1.80 1.33	2.11 2.11	3.44 3.44	4.46 4.46
4	5.12 5.12	3.71 3.71	3.26 3.26	2.64 2.64	1.80 1.10	1.80 0.45	2.69 2.69	3.61 3.61
5	6.35 6.35	5.00 5.00	4.37 4.37	3.74 3.74	2.42 2.42	1.80 0.91	1.80 1.66	2.40 2.40
6	6.91 6.91	5.85 5.85	4.77 4.77	4.22 4.22	3.70 3.70	2.52 2.52	1.80 0.54	1.80 1.24
7	8.81 8.81	7.72 7.72	6.68 6.68	6.12 6.12	5.41 5.41	4.02 4.02	1.84 1.84	1.80 0.86

La figure 31 montre les couples de $(L_{1,i}, L_{2,j})$ associés au chemin minimum. Les points issus des couples sont considérés comme homologues. Cette figure montre qu'un point peut être associé à plusieurs points de l'autre ligne.

Figure 31 : Couples de points $(L_{1,i}, L_{2,j})$ du chemin minimum représentés par les lignes pointillées.

La distance de Fréchet moyenne est la moyenne des distances entre couples de points homologues issus du chemin minimum. Elle donne un indice de l'écartement global moyen des lignes alors que la distance de Fréchet fournit un écart maximal.

Soit k couples de points homologues $(L_{1,i}, L_{2,j}) \in C_M$:

$$d_{aF} = \frac{1}{k} \sum_k d(L_{1,i}, L_{2,j})$$

avec $(L_{1,i}, L_{2,j}) \in C_M$.

Pour l'exemple (d_{aF}) vaut 0,9524. Une valeur forte de la distance de Fréchet (d_{dF}) peut être due :

- soit à un « accident » local important des points d'une des deux lignes (cas **(a)** figure 32). Dans ce cas la valeur de la (d_{aF}) est bien meilleure ;
- soit à un mauvais calage général d'une ligne sur l'autre (cas **(b)** figure 32). La valeur de la (d_{aF}) est alors très proche de celle de la (d_{dF}).

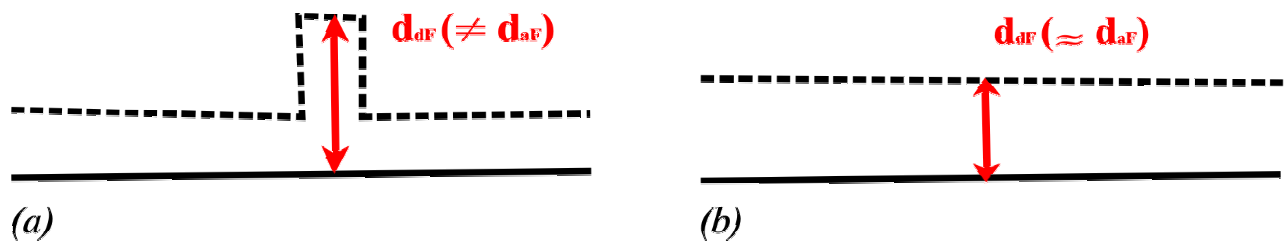


Figure 32 : Cas pour lesquels à distances moyennes égales, les distances de Fréchet maximales correspondantes sont très différentes.

Il est à noter qu'un ré échantillonnage des lignes réduit l'imprécision de la discrétisation sur la valeur de cette distance.

Tableau 7 : Evolution des valeurs de d_{dF} et d_{aF} selon le taux de ré échantillonnage des lignes de la figure 29

	Sans	1	0.1	0.01	0.001
d_{dF}	1.8028	1.2260	1.2015	1.2012	1.2012
d_{aF}	0.9524	0.5843	0.5116	0.5030	0.4997

Le tableau 7 montre la variation des valeurs selon l'échantillonnage effectué. L'architecture des jeux de données de départ ne gère pas le positionnement des points ni leur espacement : l'appariement est alors calculé entre les points constitutifs des lignes mais la longueur des segments comparés est très approximative. Sans ré échantillonnage, la mesure obtenue ne donne pas une valeur « fiable » de la distance maximale réelle entre les segments mais seulement une estimation de l'écartement entre les extrémités des segments.

Pour obtenir une valeur d_{aF} de précision centimétrique, il faut effectuer un échantillonnage avec un pas de l'ordre de quelques centimètres. Ceci s'explique par la corrélation de la distance de Fréchet discrète et de la longueur des segments (« LongMaxSeg », [Eiter & Mannila, 1994]) à partir desquels elle est calculée (voir Chapitre II).

II.2 Extension 2 : distance de Fréchet sur des lignes d'emprises ou de natures différentes

Les distances de Fréchet définies dans cette section ont pour objectif d'apparier des lignes d'emprises et de natures différentes. Elles s'appliquent à différents cas de figure illustrés figure 33 :

- comparaison de lignes ouvertes d'emprises différentes (cas **(a)**) ;
- comparaison entre deux lignes fermées (cas **(b)**) ;
- comparaison entre une ligne ouverte et une ligne fermée (cas **(c)**) ;
- comparaison entre parties de lignes ouvertes (cas **(d)**).

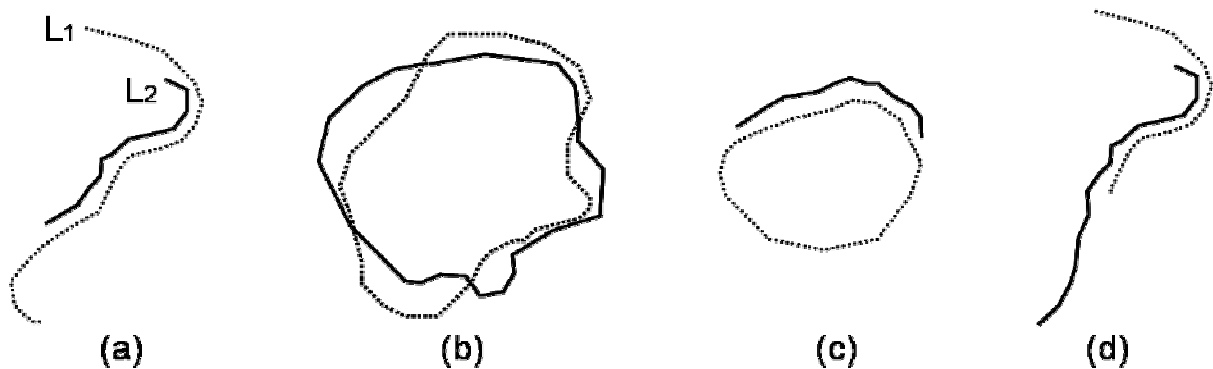


Figure 33 : Exemples de couples de géométries pertinentes pour le calcul des mesures dérivées de la distance de Fréchet.

II.2.1 Distance de Fréchet discrète partielle

Une mesure dérivée de la distance de Fréchet a été introduite par [Devogele, 2002] afin d'apparier partiellement des lignes qui ne sont pas de même emprise (figure 33 **(a)**) : la distance de Fréchet discrète partielle (d_{pdF}).

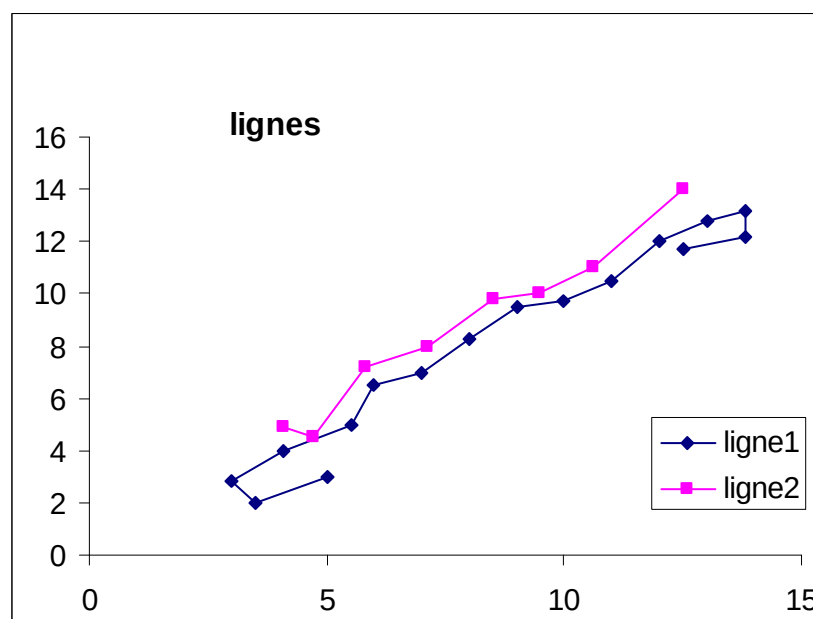


Figure 34 : Lignes partielles homologues représentées avec leur points.

La figure 34 illustre un couple de lignes qui peuvent être partiellement appariées : L_2 avec la partie homologue de L_1 comprise entre les points $< L_{1,4} \dots L_{1,13} >$. Une distance de Fréchet discrète classique ne peut pas être utilisée pour ce cas de figure en terme d'appariement car elle n'a pas de sens. Elle donnerait un résultat exprimé entre l'un des deux couples extrémités des lignes, par conséquent moins bon car non calculé entre points homologues.

Pour calculer la d_{pdf} et déterminer la ligne partiellement homologue, il est nécessaire de détecter automatiquement dans la ligne de plus grande emprise les points de départ ($L_{1,début}$) et de fin ($L_{1,fin}$) de la ligne partiellement homologue. Les points $L_{1,début} \dots L_{1,fin}$ sont testés par ordre de distance aux points de L_2 . Cet ordre est utilisé pour optimiser la recherche de ($< L_{1,début} \dots L_{1,fin} >$) : il évite ainsi de tester systématiquement tous les couples possibles du tableau 8. Le mécanisme de sélection des points homologues parmi les candidats pour $L_{1,début}$ et $L_{1,fin}$ est décrit par [Devogele, 2002]. Ces points sont des points de L_1 tels que la distance de Fréchet discrète entre L_2 et la ligne ($< L_{1,début} \dots L_{1,fin} >$) soit minimale : $d_{pdf} = d_{dF}(L_2, < L_{1,début} \dots L_{1,fin} >)$ où début = 4 et fin = 13 dans l'exemple de la figure 34.

Tableau 8 : Matrice de Distance des lignes homologues partielles représentées figure 34.

Matrice de Distance																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	2.1	2.96	2.37	0.9	1.4	2.48	3.58	5.14	6.72	7.64	8.89	10.6	11.9	12.8	12.1	10.8
2	1.53	2.77	2.4	0.78	0.94	2.39	3.4	5	6.59	7.46	8.7	10.5	11.7	12.6	11.9	10.6
3	4.28	5.69	5.22	3.62	2.22	0.73	1.22	2.44	3.94	4.91	6.16	7.84	9.12	10	9.43	8.07
4	5.42	7	6.62	5	3.4	1.86	1	0.93	2.42	3.39	4.63	6.33	7.61	8.48	7.91	6.55
5	7.65	9.26	8.9	7.28	5.66	4.14	3.18	1.63	0.58	1.5	2.6	4.13	5.41	6.3	5.82	4.43
6	8.32	10	9.7	8.07	6.4	4.95	3.91	2.3	0.71	0.56	1.58	3.2	4.48	5.36	4.83	3.45
7	9.77	11.5	11.2	9.55	7.87	6.44	5.38	3.78	2.19	1.39	0.64	1.72	3	3.88	3.42	2.02
8	13.3	15	14.7	13.1	11.4	9.92	8.9	7.3	5.7	4.93	3.81	2.06	1.3	1.53	2.22	2.3

Le chemin minimal est déterminé de la même manière que pour la distance de Fréchet discrète.

II.2.2 Distance de Fréchet discrète entre deux lignes fermées

La distance de Fréchet discrète entre deux lignes fermées représente la distance de Fréchet discrète entre deux limites de polygones. La figure 33 (b) illustre un exemple pour lequel la distance de Fréchet entre deux lignes fermées doit être utilisée.

La procédure décrite dans [Devogele, 2002] utilise une fonction qui convertit les limites de polygones P_1 et P_2 en suites ordonnées de points, les lignes L_1 et L_2 . La difficulté vient de l'absence de point de début et de fin pour un polygone, contrairement à une ligne ouverte. Chaque ligne issue de la transformation du polygone va donc comporter un point doublé. Ce point marque le caractère « fermé » et symbolise à la fois le point de début et de fin de la ligne ainsi obtenue.

L'objectif est alors d'obtenir des lignes de même sens avec des points de début relativement proches.

L'algorithme présenté figure 35 explique le mécanisme de recherche d'un point de début d'une ligne L_2 en vue de son appariement avec une ligne L_1 . Dans un premier temps, un point de L_1 est choisi de manière à servir de point de début. Ce point dénommé $L_{1,1}$ est arbitrairement sélectionné pour son éloignement maximal vis-à-vis de tous les points de L_2 . Par la suite l'algorithme recherche parmi les points de L_2 le point $L_{2,1}$ le plus proche de $L_{1,1}$.

Ces deux points $L_{1,1}$ et $L_{2,1}$ deviennent alors temporairement les points de début respectifs de L_1 et L_2 localisés à proximité l'un de l'autre : il est alors possible de tester le sens des deux lignes.

Une ligne L_2 est dite **dans le même sens** qu'une ligne L_1 si l'indice du point de L_2 le plus proche du point $L_{1,1}$ est **avant** l'indice du point de L_2 le plus proche de $L_{1,n}$. Si les lignes ne sont pas dans le même sens, la ligne L_2 est inversée. Cette étape permet le calcul de la d_{Fd} entre les deux lignes de façon à ce que la valeur soit la plus petite possible et le chemin minimal continu [Devogele, 2002].

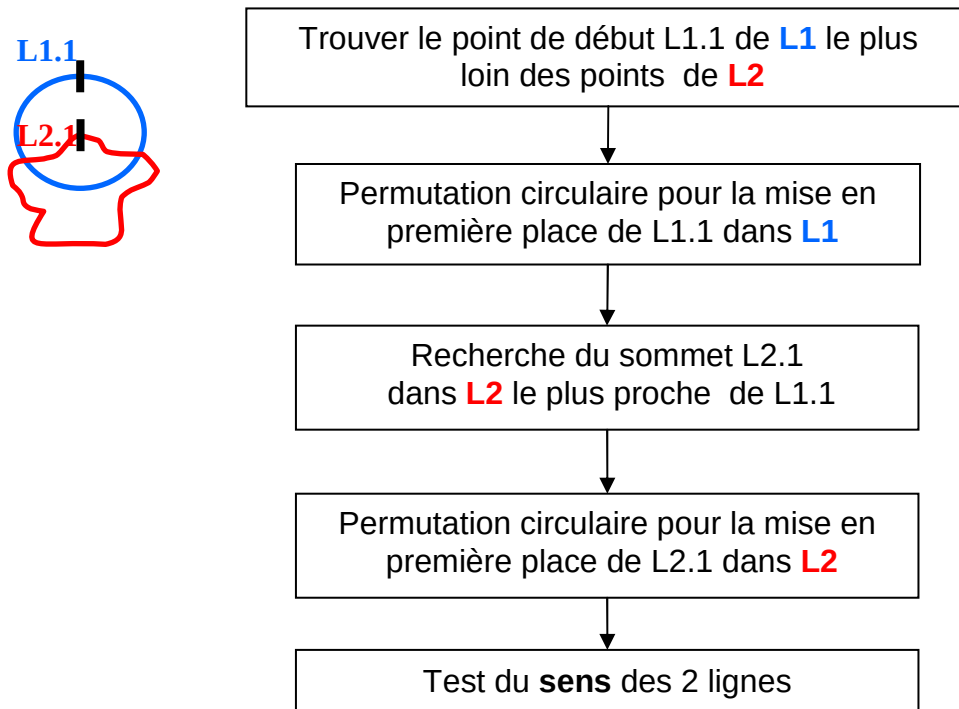


Figure 35 : Algorithme de recherche des points de début de L_1 et L_2 en vue d'un appariement des deux lignes fermées.

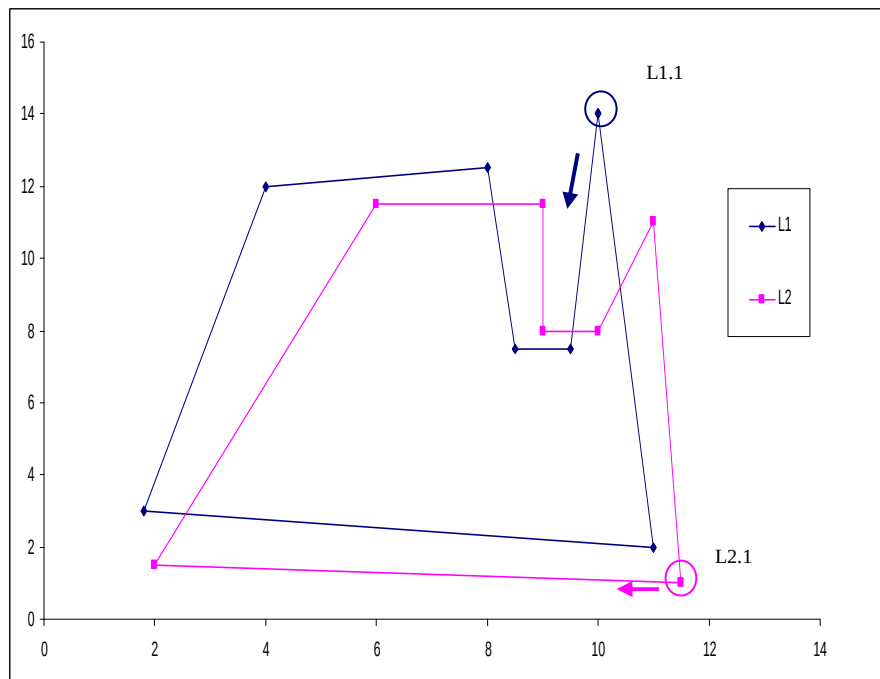


Figure 36 : Ligne fermées représentées avec des sens opposés.

Tableau 9 : Matrices de Distance et de Fréchet des lignes fermées de la figure 36.

Matrice de Distance								
	1	2	3	4	5	6	7	8
1	13.086	14.841	4.717	2.6926	6.0828	6	3.1623	13.086
2	6.8007	9.6047	5.3151	4.0311	0.7071	0.7071	3.8079	6.8007
3	7.1589	8.8459	4.717	4.0311	0.7071	1.5811	4.3012	7.1589
4	12.021	12.53	2.2361	1.4142	4.6098	4.9244	3.3541	12.021
5	13.314	10.689	2.0616	5.0249	6.4031	7.2111	7.0711	13.314
6	9.904	1.5133	9.481	11.14	8.7658	9.6042	12.192	9.904
7	1.118	9.0139	10.735	9.7082	6.3246	6.0828	9	1.118
8	13.086	14.841	4.717	2.6926	6.0828	6	3.1623	13.086

Calcul de Fréchet								
	1	2	3	4	5	6	7	8
1	13.086	14.841	14.841	14.841	14.841	14.841	14.841	14.841
2	13.086	13.086	13.086	13.086	13.086	13.086	13.086	13.086
3	13.086	13.086	13.086	13.086	13.086	13.086	13.086	13.086
4	13.086	13.086	13.086	13.086	13.086	13.086	13.086	13.086
5	13.314	13.086	13.086	13.086	13.086	13.086	13.086	13.314
6	13.314	13.086	13.086	13.086	13.086	13.086	13.086	13.086
7	13.314	13.086	13.086	13.086	13.086	13.086	13.086	13.086
8	13.314	14.841	13.086	13.086	13.086	13.086	13.086	13.086

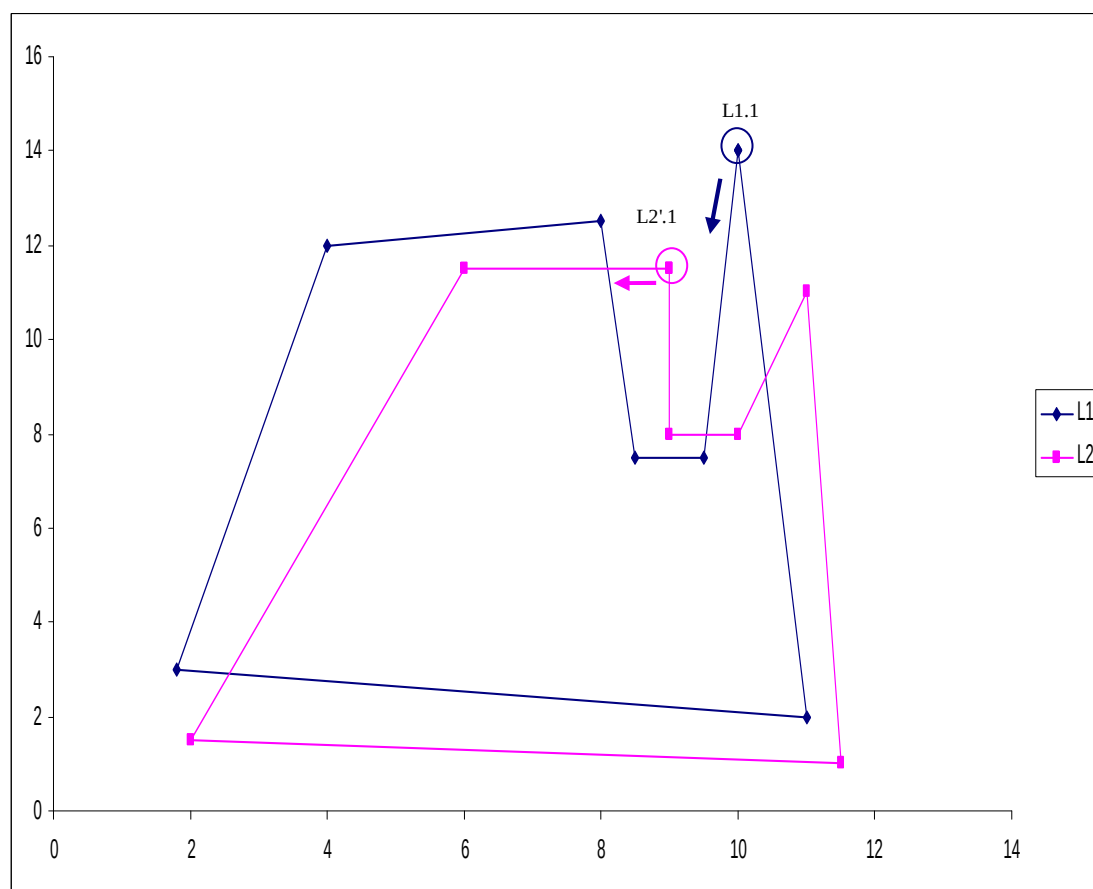
Figure 37 : Etape d'inversion du sens de la ligne fermée L_2 en L_2' .

Tableau 10 : Matrice de distance résultant de l'inversion de la ligne L_2 , présentée figure 37. Le cheminement optimal est représenté en gris. Pour cet exemple $d_{Fd} = 6,08$.

Matrice de Distance								
	1	2	3	4	5	6	7	8
1	2.6926	4.717	14.841	13.086	3.1623	6	6.0828	2.6926
2	4.0311	5.3151	9.6047	6.8007	3.8079	0.7071	0.7071	4.0311
3	4.0311	4.717	8.8459	7.1589	4.3012	1.5811	0.7071	4.0311
4	1.4142	2.2361	12.53	12.021	3.3541	4.9244	4.6098	1.4142
5	5.0249	2.0616	10.689	13.314	7.0711	7.2111	6.4031	5.0249
6	11.14	9.481	1.5133	9.904	12.192	9.6042	8.7658	11.14
7	9.7082	10.735	9.0139	1.118	9	6.0828	6.3246	9.7082
8	2.6926	4.717	14.841	13.086	3.1623	6	6.0828	2.6926

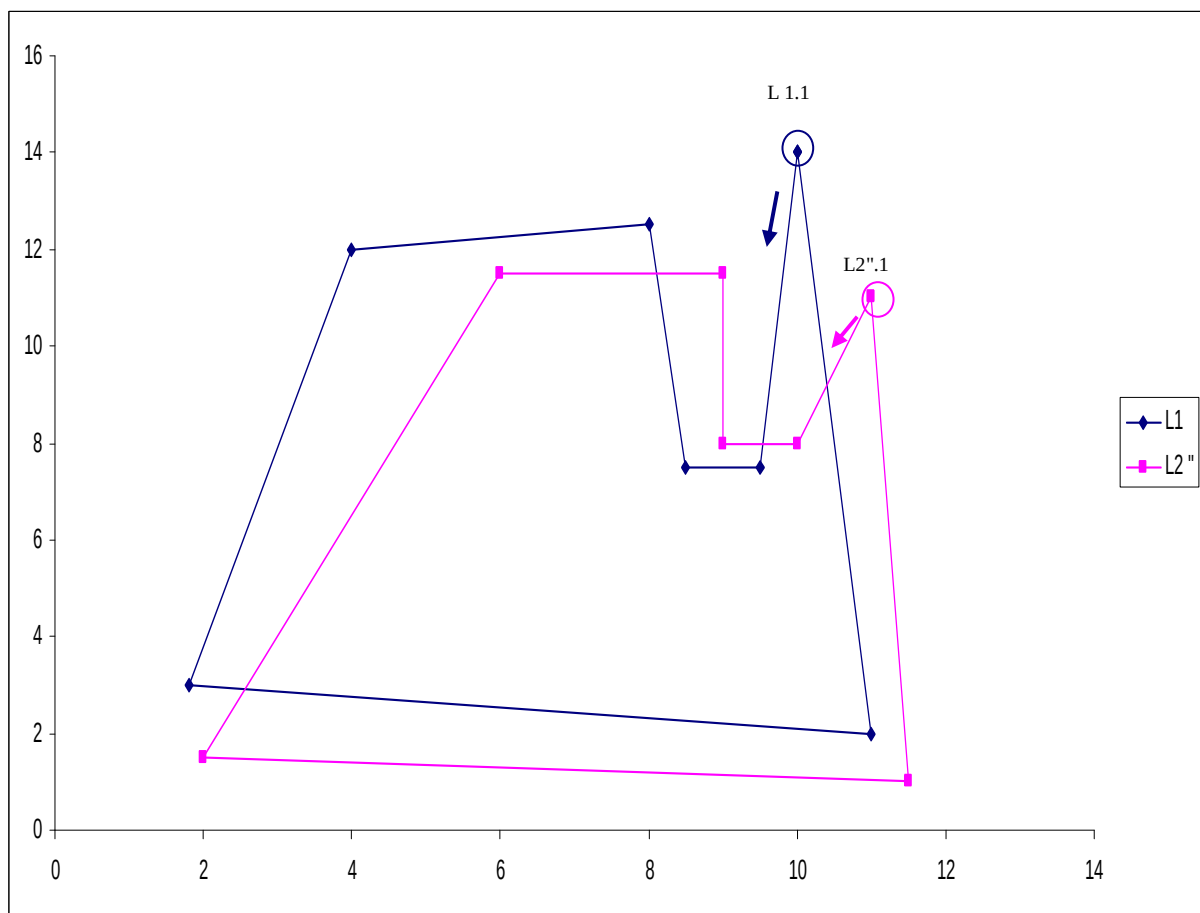


Figure 38 : Recherche du point homologue de $L_{1,1}$ parmi les points de L_2 , et permutation du point trouvé de manière à ce qu'il devienne $L_{2'',1}$.

Les figure 36 et suivantes illustrent le traitement de deux lignes fermées de sens opposés. Le tableau 9 présente les matrices de distance et de Fréchet calculées à partir des points des lignes : aucune recherche ni modification de sens n'a été réalisée. La distance de Fréchet est de : 13,09 ; ce qui est une valeur trop forte pour permettre un appariement. Un nouveau point de début pour L_2 est alors recherché : $L_{2,4}$ devient $L_{2',1}$ après une permutation circulaire. Le test du sens des lignes confirme que celles-ci sont de sens opposés : L_2 est alors inversée.

Tableau 11 : Matrice de Distance résultante de la dernière permutation, distance de Fréchet.

Matrice de Distance								
	1	2	3	4	5	6	7	8
1	3.1623	6	6.0828	2.6926	4.717	14.841	13.086	3.1623
2	3.8079	0.7071	0.7071	4.0311	5.3151	9.6047	6.8007	3.8079
3	4.3012	1.5811	0.7071	4.0311	4.717	8.8459	7.1589	4.3012
4	3.3541	4.9244	4.6098	1.4142	2.2361	12.53	12.021	3.3541
5	7.0711	7.2111	6.4031	5.0249	2.0616	10.689	13.314	7.0711
6	12.192	9.6042	8.7658	11.14	9.481	1.5133	9.904	12.192
7	9	6.0828	6.3246	9.7082	10.735	9.0139	1.118	9
8	3.1623	6	6.0828	2.6926	4.717	14.841	13.086	3.1623

Ordre des points de L2 avant modification

ddF = 13,086

L2.j.x	11.5	2	6	9	9	10	11	11.5
L2.j.y	1	1.5	11.5	11.5	8	8	11	1

Point de L2 inversée au plus proche (L2.4 est le plus proche de L1.1)

ddF = 6,083

L2.j.x	9	6	2	11.5	11	10	9	9
L2.j.y	11.5	11.5	1.5	1	11	8	8	11.5

Point de L2 permutation circulaire (L2.7)

ddF = 3,162

L2.j.x	11	10	9	9	6	2	11.5	11
L2.j.y	11	8	8	11.5	11.5	1.5	1	11

Pas d'autres points à tester car les autres distances (L1.1, L2.j) sont supérieures à 3,162

La figure 37 et ses matrices (tableau 10) montrent l'évolution de la distance de Fréchet après ces changements : la distance obtenue est de 6,08.

Afin d'apparier les points homologues entre eux et non uniquement les plus proches géographiquement, une permutation circulaire est à nouveau réalisée. La figure 38 et le tableau 11 présentent la valeur de la distance obtenue : $d_{Fd} = 3,16$. Des permutations suivantes ne sont pas entreprises car les autres distances dans la matrice de Fréchet sont supérieures à 3,162.

La comparaison de lignes fermées se fait impérativement sur leur totalité car il faudrait sinon rechercher quatre inconnues : deux points de début et deux points de fin. Comme ceci n'est pas réalisable, il n'existe pas de distance de Fréchet discrète partielle pour deux lignes fermées.

II.2.3 Distance de Fréchet discrète partielle entre une ligne ouverte et une ligne fermée

La définition d'une distance de Fréchet discrète partielle (d_{pdF}) entre une ligne L_1 et une partie de frontière de polygone P_2 représente un apport de cette thèse vis-à-vis de ce qui a déjà été présenté. Elle se calcule de façon intermédiaire vis-à-vis des deux mesures décrites précédemment.

Le principe est de découper la ligne fermée pour obtenir la partie homologue de la ligne ouverte et ensuite de calculer une distance de Fréchet discrète. La figure 33 (c) montre un exemple entre une ligne ouverte et une ligne fermée qui impose le calcul d'une d_{pdF} .

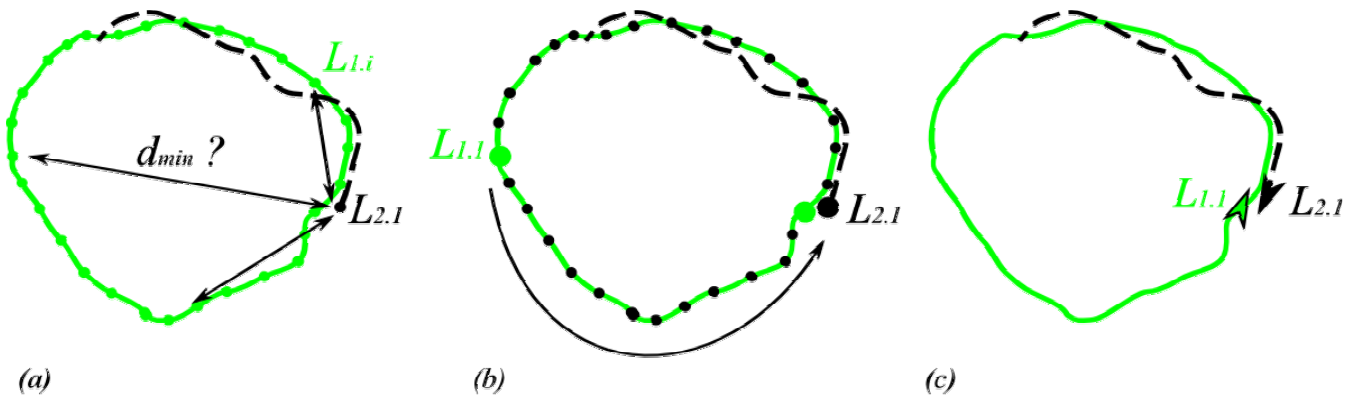


Figure 39 : détail de l'appariement d'une ligne fermée avec une ligne ouverte.

La figure 39 reprend le détail du traitement. Premièrement les distances entre le point de départ de la ligne ouverte $L_{2,1}$ et tous les points de L_1 sont calculées. La distance minimale est conservée (figure 39 (a)). A partir de cette distance, une permutation circulaire de L_1 est effectuée de manière à avoir $L_{1,\text{debut}}$ situé à l'emplacement du point où la distance est minimale (figure 39 (b)). Cette étape de rapprochement géographique, à l'instar de l'exemple précédent entre deux lignes fermées, facilite la détermination du sens de L_1 par rapport à L_2 . S'il s'avère que les lignes ont des sens opposés (figure 39 (c)), une des deux lignes est inversée. La ligne fermée L_1 est ensuite comparée à la ligne ouverte L_2 puis le point homologue à $L_{2,n}$ est recherché. La distance de Fréchet discrète partielle est ensuite calculée de la même manière qu'elle le serait entre deux lignes de même emprise en construisant le chemin minimal et en recherchant les points homologue et non ceux les plus proches.

II.2.4 Distance de Fréchet discrète partielle / partielle entre deux lignes ouvertes

La distance de Fréchet discrète partielle entre deux tronçons de lignes ouvertes diffère de celles présentées précédemment car la relation entre les lignes n'est plus la même. Jusqu'à présent à une ligne entière correspondait la totalité d'une autre ligne (distance de Fréchet discrète) ou au moins un des ses tronçons (distance de Fréchet discrète partielle).

Ici, à une partie d'une ligne ne correspond uniquement qu'une autre partie de la ligne comparée, ce qui amène à définir également une nouvelle relation entre les objets (cf chapitre II) : la relation d'appariement *partiel* « p-p », très liée à la relation n-m définie en partie II.1.1.3. Si à n objets de la BD 1 correspondent m objets de la BD2, cela signifie que pour deux objets comparés, une partie d'un objet géographique du jeu de données 1 équivaut à une partie d'un objet du jeu de données 2 (voir figure 40). D'où la relation partielle/partielle « PP » entre objets. Ce type de relation exclut automatiquement les relations de type 1-1.

Type de relation	
n-m	BD 1 BD 2
p-p	

Figure 40 : Mise en parallèle des relations n-m et p-p entre deux objets.

En pratique cela induit la comparaison de la fin d'une ligne avec le début d'une autre. En effet, contrairement aux mesures précédentes, les **deux** lignes sont comparées partiellement de manière à trouver dans certaines configurations deux points de début et de deux points de fin (voir figure 41 cas **(a)**).

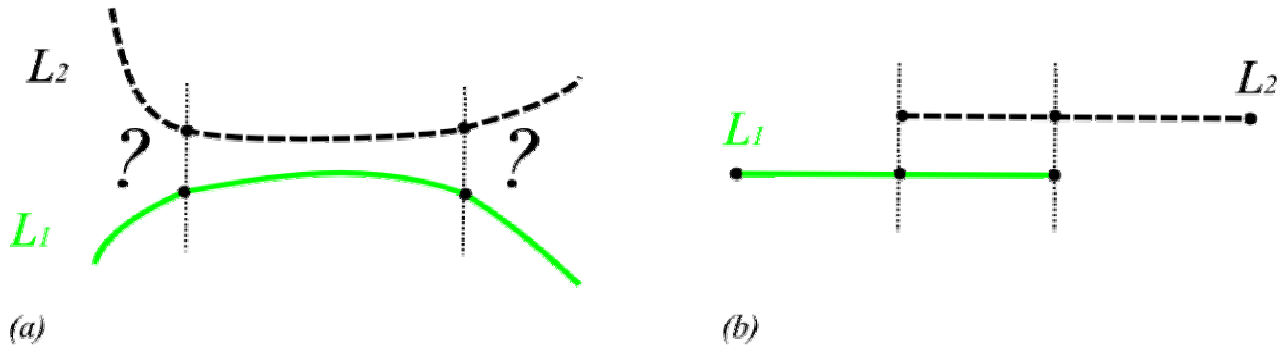


Figure 41 : Cas d'appariement partiel-partiel de lignes ouvertes : a) exemple impossible à résoudre en l'état actuel de nos connaissances ; b) exemple qui possède une solution unique.

Ce type d'appariement vise à identifier un segment de ligne homologue à un autre segment de ligne. Il va alors se poser des **problèmes de contraintes** pour certains cas. L'exemple **(a)** de la figure 41 est insolvable car le nombre d'inconnues est trop grand : la disposition des lignes impose la recherche d'un point de début et de fin pour L_1 et L_2 , soit **quatre points** à déterminer au total. Aucun indice ne permet de déterminer la partie de ligne L_1 à appairer avec la partie de ligne L_2 , il faut donc retenir :

- soit les segments les plus longs possibles et ainsi obtenir une distance de Fréchet discrète très grande ;
- soit rechercher les segments les plus courts afin d'avoir une distance de Fréchet bien meilleure, au risque de réduire les segments aux deux points les plus proches.

Entre ces deux extrêmes, il existe une infinité de solutions possibles. Pour apporter une solution à ce problème, il est donc nécessaire d'introduire deux autres contraintes :

- un des points de début est le point d'origine de la ligne ;
- un des points de fin est le point de fin de l'autre ligne considérée.

Ces deux contraintes sont illustrées par l'exemple **(b)** figure 41 qui ne demande la recherche que d'un point de début sur une ligne L_1 et un point de fin sur la ligne L_2 .

D'un point de vu calcul matriciel, ces contraintes équivalent à chercher les points candidats dans une sous-matrice située dans l'angle haut droit de la matrice globale. La résolution mathématique comporte en effet de fortes similitudes avec le calcul d'une distance entre lignes partielle/totale :

- dans un cas partiel/totale, les points de début et de fin sont cherchés dans une sous matrice incluant la première et dernière ligne de la matrice ;
- pour un cas PP les points de débuts et de fin candidats sont recherchés uniquement parmi les points de la **première ligne** et de la **dernière colonne** de la matrice globale (tableau 13).

Sous certaines conditions, ces règles de recherches de points peuvent être étendues à la recherche de quatre points si les lignes possèdent entre elles non plus **une** extrémité homologue (cas « PP ») mais **deux** (exemple **(b)** figure 42) : la relation définie est appelée « **deux partielle-partielle** » (2PP) et réalise deux appariements PP.

A partir de ces considérations, les cas de figure qui peuvent être traités sont illustrés figure 42 :

- (a) : relation PP ;
- (b) : relation 2PP ;
- (c) : relation ambiguë d'un cas PP.

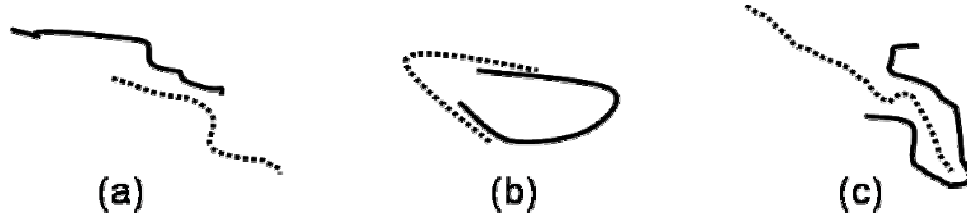


Figure 42 : Illustration des différents positionnements des lignes lors de leur appariement.

Les exemples suivants montrent des cas de lignes pour lesquelles il y a impossibilité à déterminer des points de début et de fin : 1) pour des lignes ouvertes/fermées (figure 43), et 2) pour des lignes fermées/fermées (figure 44).

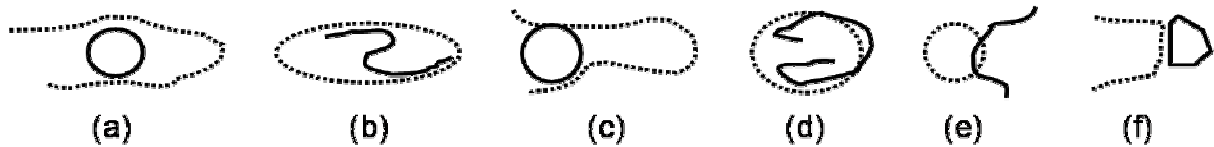


Figure 43 : Inventaire des cas d'appariement impossibles à traiter entre une ligne ouverte et une ligne fermée. En (a), (b) et (c) dans un contexte d'appariement deux partiel/partiel ; et en (d), (e), (f) pour des cas partiel/partiel.

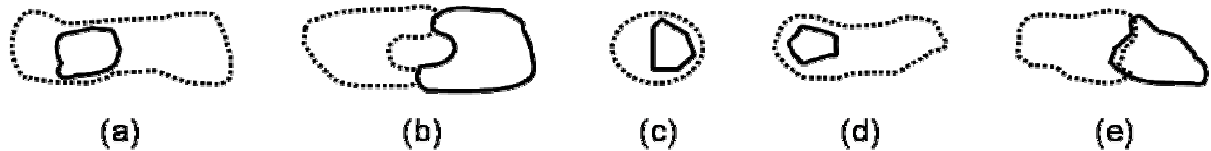


Figure 44 : Inventaire des cas impossibles à traiter entre deux lignes fermées. En (a), (b) et (c) pour un appariement deux partiel/partiel ; et en (d), (e) pour un appariement partiel/partiel.

L'exemple (b) « 2PP » de la figure 42 est présenté ci-après de manière à étayer plus précisément la méthode de traitement.

Visuellement les lignes ne sont homologues que partiellement, mais simultanément sur leur deux extrémités.

L'outil implémenté ne permet pas lors du processus de comparaison de déterminer quelle extrémité de la ligne est homologues à l'autre ni si elle est homologues sur une partie ou sur deux. Quatre distances partielles sont donc calculées :

- L_1 et L_2 ;
- L_{1inv} et L_2 ;
- L_1 et L_{2inv} ;
- L_{1inv} et L_{2inv} .

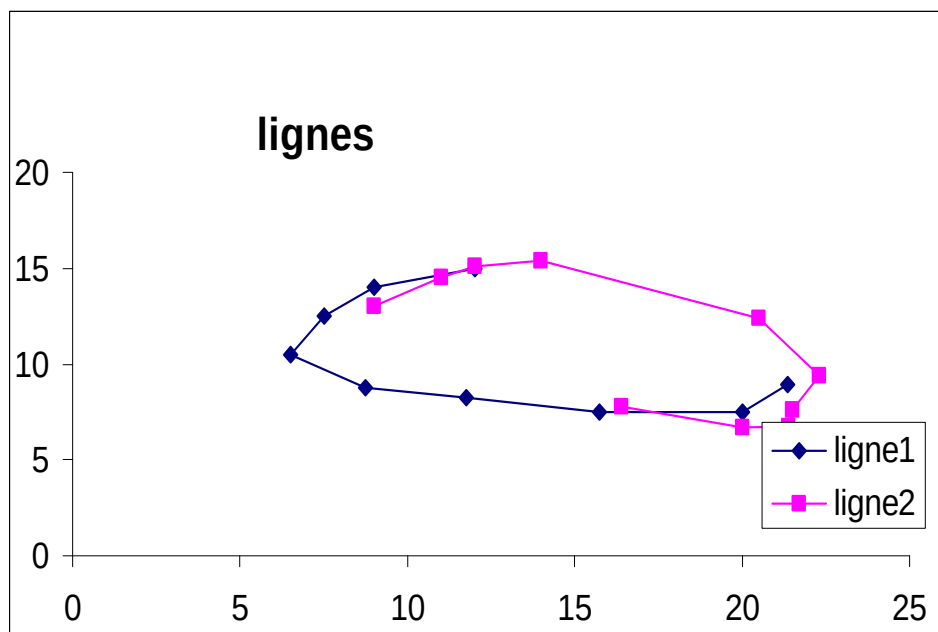


Figure 45 : Exemple de lignes pour lesquelles les deux extrémités sont homologues (cas 2PP).

Tableau 12 : Matrice de Distance et Matrice de Fréchet des lignes L_1 et L_2 présentées figure 45.

Matrice de Distance (MD)									
	1	2	3	4	5	6	7	8	9
1	3.606	1	1.581	3.536	4.257	5.489	8.701	12.3	13.02
2	1.118	2.062	4.031	6.021	6.175	6.295	8.451	11.4	11.77
3	0.1	3.195	5.197	7.17	7.133	6.855	8.466	11.03	11.21
4	2.04	5.192	7.118	8.959	8.473	7.496	8.082	9.92	9.8
5	8.889	11.61	13	14.13	12.3	9.684	6.817	4.925	3.568
6	11.72	14.07	15.12	15.84	13.57	10.61	6.817	2.983	1.038
7	12.04	14.04	14.83	15.28	12.8	9.772	5.751	1.503	1.346
8	12.52	14.37	15.05	15.37	12.81	9.769	5.704	1.598	2.21
9	11.52	13.2	13.78	14.02	11.43	8.393	4.325	0.79	2.617
10	8.438	9.654	10.06	10.26	7.709	4.672	0.712	3.612	5.099

Matrice de Fréchet (MF)									
	1	2	3	4	5	6	7	8	9
1		1	1.581	3.536	4.257	5.489	8.701	12.3	13.02
2		1.118	4.031	6.021	6.175	6.295	8.451	11.4	11.77
3		0.1	5.197	7.17	7.133	6.855	8.466	11.03	11.21
4		2.04	7.118	8.959	8.473	7.496	8.082	9.92	9.92
5		8.889	13	14.13	12.3	9.684	7.496	7.496	7.496
6		11.72	15.12	15.84	13.57	10.61	7.496	7.496	7.496
7		12.04	14.83	15.28	13.57	10.61	7.496	7.496	7.496

Les matrices de distance et de Fréchet des lignes de la figure 45 sont présentées dans le tableau 12.

Le point de départ retenu est $L_{1,2}$ et le point d'arrivée $L_{2,7}$, le résultat de la distance de Fréchet discrète partielle obtenue est de : **7,496**.

La ligne L_1 est ensuite inversée (L_{1inv}) et le même calcul est réitéré avec L_2 (voir tableau 13).

Tableau 13 : Matrice de Distance et de Fréchet des lignes de la figure 45 avec les points de la ligne L_1 inversée.

Matrice de Distance									
	1	2	3	4	5	6	7	8	9
1	13.02	12.3	8.701	5.489	4.257	3.536	1.581	1	3.606
2	11.77	11.4	8.451	6.295	6.175	6.021	4.031	2.062	1.118
3	11.21	11.03	8.466	6.855	7.133	7.17	5.197	3.195	0.1
4	9.8	9.92	8.082	7.496	8.473	8.959	7.118	5.192	2.04
5	3.568	4.925	6.817	9.684	12.3	14.13	13	11.61	8.889
6	1.038	2.983	6.817	10.61	13.57	15.84	15.12	14.07	11.72
7	1.346	1.503	5.751	9.772	12.8	15.28	14.83	14.04	12.04
8	2.21	1.598	5.704	9.769	12.81	15.37	15.05	14.37	12.52
9	2.617	0.79	4.325	8.393	11.43	14.02	13.78	13.2	11.52
10	5.099	3.612	0.712	4.672	7.709	10.26	10.06	9.654	8.438

Matrice de Fréchet									
	1	2	3	4	5	6	7	8	9
1								1	3.606
2								2.062	1.118
3								3.195	1.118

Le point de départ retenu est $L_{1,8}$ et le point d'arrivée $L_{2,3}$, le résultat de la distance de Fréchet discrète partielle obtenue est de : **1,118**.

Les deux autres calculs sont effectués exactement sur le même modèle et les distances de Fréchet obtenues sont :

- pour L_1 et L_{2inv} = **1,598** ;
- pour L_{1inv} et L_{2inv} = **7,495**.

Dans cet exemple, six appariements sont donc possibles : deux appariements 2PP et quatre appariements PP. Dans le cas d'un appariement 2PP les résultats des couples retenus sont les suivants :

- couple L_1 et L_2 et L_{1inv} et L_2 = {**7,496** ; **7,495**} ;
- couple L_{1inv} et L_2 et L_1 et L_{2inv} = {**1,118** ; **1,598**}.

Pour un appariement PP seules les distances entre les deux lignes sont retenues :

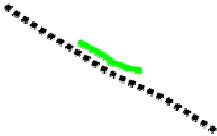
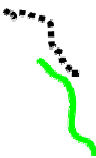
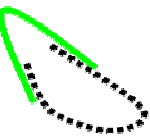
- L_1 et L_2 = 7, 496 ;
- L_1 et L_{2inv} = 1, 598 ;
- L_{1inv} et L_2 = 1, 118 ;
- L_{1inv} et L_{2inv} = 7, 495 ;

Se pose alors la question du choix entre un appariement 2PP ou PP. Il a été décidé de préférer un appariement 2PP pour les cas où il existe **un couple** possédant deux valeurs de distances de Fréchet inférieures à ε . Dans le cas où une seule distance est inférieure, un appariement PP est choisi.

Pour cet exemple les couples L_{1inv}/L_2 , et L_1/L_{2inv} , présentent des distances de Fréchet très proches en valeurs et assez faibles. L'appariement est donc effectué entre les points homologues des lignes retenues dans les chemins minimaux et de type 2PP. Cependant il ne faut pas prendre en compte les couples de points qui possèdent deux fois la même extrémité (exemple figure 42 (c)). Pour ces types de cas, un appariement PP doit être retenu en fonction des d_{pF} et du contexte (défini par l'utilisateur).

Le tableau 14 synthétise les différents types d'appariement résolus grâce à soit l'utilisation de distances de Fréchet déjà existantes, soit la définition de nouvelles distances.

Tableau 14 : Récapitulatif des distances d'appariements déjà traitées et celles introduites dans cette section.

	Total/total 1 - 1	Partiel 1 - P	Partiel - Partiel	
			PP	2PP
Distances déjà existantes				
Apports de la thèse				

Les différentes catégories d'appariements, basés sur les multiples distances déjà existantes ou créées selon les nouveaux cas de comparaisons effectuées, vont maintenant être présentées.

II.3 Algorithme d'appariement

L'objectif est de proposer le meilleur appariement lors de la comparaison de deux lignes quelconques. Pour ce faire, l'algorithme procède à une séquence de tests réalisés sur les lignes de manière à identifier quelle distance de Fréchet doit être calculée en vue de leur appariement final.

Cette section présente les trois étapes essentielles intervenant lors du processus d'appariement de lignes (voir figure 46). La première est la détermination du type de lignes comparées. Cette étape est effectuée soit manuellement par l'utilisateur (comparaison et appariement simple) soit automatiquement (appariement global par fichier). Dans la seconde étape il s'agit d'identifier le type de relations entre les emprises des lignes, symbolisées par les rectangles englobants. Enfin, la troisième et dernière étape consiste à calculer la distance de Fréchet choisie en fonction du type de lignes identifié en première étape et de la position relative des lignes déterminée lors de l'étape deux. En fonction du résultat de cette distance, un appariement approprié est effectué ou non.

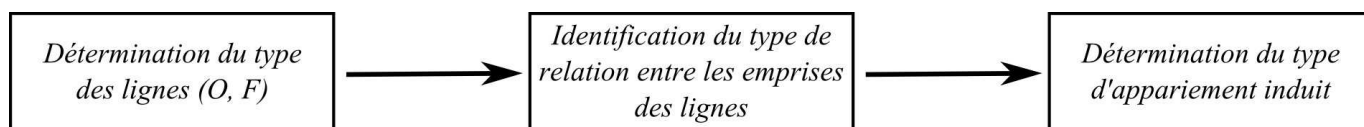


Figure 46 : Description des trois principales étapes définies pour l'appariement.

Cette chaîne de traitement vise à résoudre automatiquement le plus grand nombre de cas possibles lors d'un processus de comparaison. Ceux impossibles à résoudre par cette méthode sont également décrits en fin de section.

II.3.1 Recherche de la nature des lignes comparées

Les algorithmes d'appariement testés ne sont pas les mêmes selon qu'il s'agisse de comparer deux lignes ouvertes de même emprise, deux lignes ouvertes d'emprise différentes, deux lignes fermées ou une ligne ouverte avec une ligne fermée. Par souci d'optimisation du nombre de calculs effectués, l'algorithme va effectuer un premier tri : l'identification de la nature des lignes à comparer. En effet certaines catégories de comparaison de lignes vont obligatoirement exclure certains cas d'appariement qu'il n'est donc pas nécessaire de tester. La première étape consiste donc à vérifier pour une ligne donnée si les points de début et de fin sont les mêmes. Une fois que le type de ligne est connu, plusieurs types de comparaison s'offrent alors :

- ouvert/ouvert total ;
- ouvert/ouvert partiel ;
- ouvert/fermé (toujours partiel) ;
- fermé/fermé (toujours total).

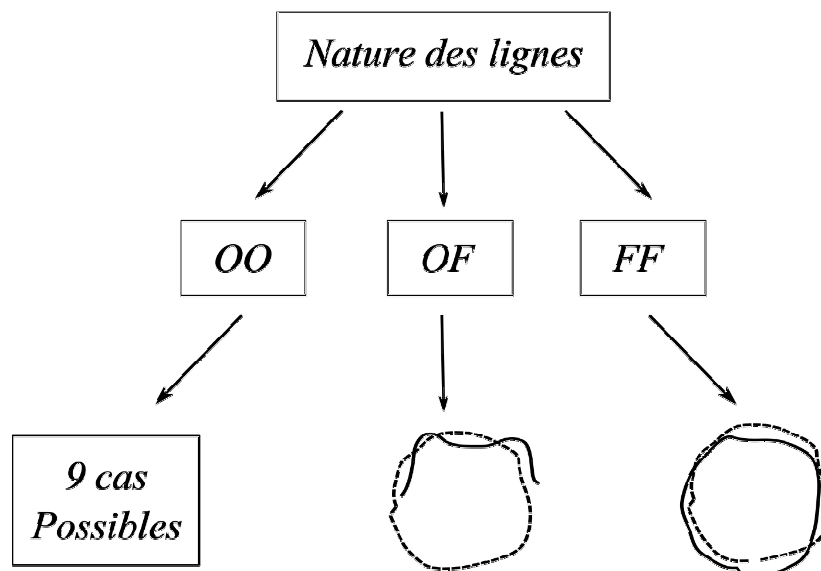


Figure 47 : Inventaire des appariements possibles selon la nature des lignes comparées.

La figure 47 récapitule les différents appariements possibles selon la nature des lignes comparées. Dans le cas de deux lignes ouvertes, tous les types d'appariement sont possibles. Pour déterminer quelle distance de Fréchet calculer, un autre tri doit être réalisé grâce aux relations des emprises des lignes comparées.

La comparaison d'une ligne ouverte avec une ligne fermée se rapporte à calculer une distance de Fréchet partielle et donc induire un appariement partiel : soit 1-P, soit PP soit 2PP.

La comparaison de deux lignes fermées est la plus simple car elle n'entraîne qu'un appariement total.

II.3.2 Relations d'emprises entre les lignes

Les différents appariements possibles sont déduits à partir des relations d'inclusion des emprises des lignes. Ces emprises sont représentées par le rectangle simple d'une ligne (BB : Bounding Box) et le rectangle englobant élargi d'un epsilon (BBE : Bounding Box Extended) de l'autre ligne comparée. Cet élargissement a pour but de créer une zone tampon

d'incertitude liée à la qualité des données. Cette valeur epsilon (ϵ) est choisie par l'utilisateur selon la précision de positionnement. Si les données sont très précises, ϵ peut être plus petit. Si au contraire les données sont peu précises, il va être nécessaire d'utiliser un ϵ plus grand de manière à s'affranchir des problèmes de calage qui peuvent aller jusqu'à plusieurs dizaines de mètres. Dans ce cas, l'appariement est effectué si la valeur de la distance de Fréchet obtenue est inférieure à une dizaine de mètres.

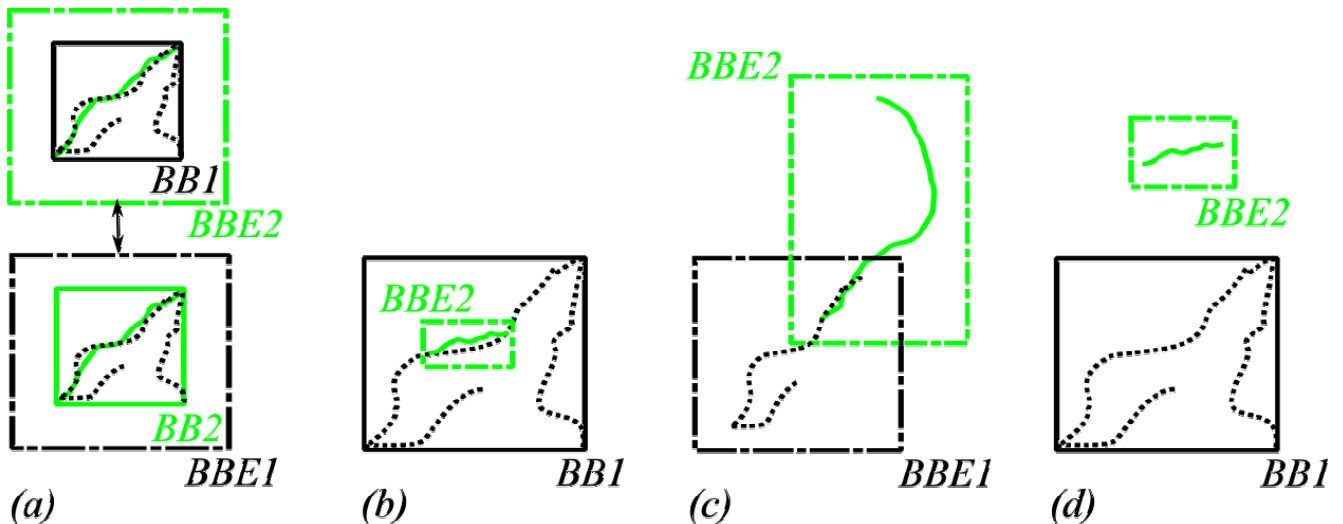


Figure 48 : Inventaire des relations d'inclusion existant entre objets géographiques : (a) emprises équivalentes, (b) emprise incluse dans une autre, (c) emprises possédant une zone de recouvrement, (d) emprises sans intersection.

Quatre types de relations ont été définis :

- l'inclusion réciproque de deux lignes entre elles dans le cas où elles sont d'emprises relativement **équivalentes** (cas (a), figure 48) : $BB1 \subset BBE2$ **et** $BB2 \subset BBE1$. L'appariement est alors soit total (1-1), soit partiel (1-P, PP ou 2PP) ;
- l'**inclusion** simple d'une ligne dans le rectangle englobant d'une autre (cas (b), figure 48) : $BB1 \subset BBE2$. Sa relation complémentaire se nomme **est inclus** : $BB2 \subset BBE1$. L'appariement est toujours partiel (1-P, PP ou 2PP) ;
- le **recouvrement** partiel (cas (c), figure 48) : $BBE1 \cap BBE2 \neq 0$. L'appariement effectué est partiel de type PP ou 2PP ;
- **aucune inclusion** (cas (d), Figure 48) : $BBE1 \cap BBE2 = 0$. Dans ce cas il n'y a pas d'appariement possible.

Ces relations sont appelées relations d'inclusion entre les emprises (RIE). Pour chaque type d'inclusion et selon la nature des lignes comparées, les appariements engagés et les distances de Fréchet discrètes utilisées ne sont pas les mêmes. Par exemple pour la cas (a), le principe est que si les emprises des lignes sont équivalentes, il y a une très forte probabilité pour que les lignes soient au même endroit et à peu près de la même taille. La distance calculée est une distance de Fréchet totale et l'appariement testé est de type 1-1. Si la d_F obtenue est inférieure à ϵ , l'appariement est validé, sinon il est éliminé des appariements possibles.

De la même manière, si la relation entre les emprises est l'inclusion, cela exclut automatiquement un appariement de type total.

S'il n'y a pas d'intersection entre les emprises, aucune distance de Fréchet n'est calculée et aucun appariement n'est effectué.

La section suivante rappelle les types d'appariement existants et quels types de lignes et de relations d'emprise sont concernés.

II.3.3 Types d'appariement entre lignes

II.3.3.1 Appariement simple « 1 – 1 » de deux lignes

Ce type d'appariement concerne les lignes d'emprises équivalentes qui ont une grande probabilité d'avoir une relation 1-1 dans leur comparaison : à une ligne L_1 correspond une et une seule ligne L_2 .

Deux lignes ouvertes

L'équivalence des emprises des lignes est vérifiée par test croisé : $BB1$ doit être inclus dans $BBE2$ et inversement. Ensuite la distance de Fréchet est calculée entre L_1/L_2 . Si la distance minimale trouvée est supérieure à ϵ , alors aucun appariement n'est effectué.

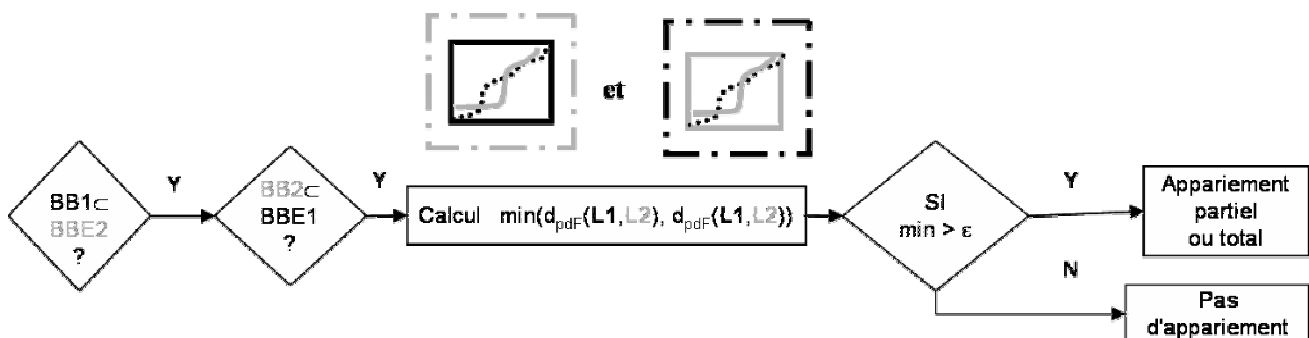


Figure 49 : Algorithme d'appariement total « 1-1 » entre deux lignes ouvertes. Les illustrations indiquent quel cas est alors identifié lors du calcul de la d_{pdf} .

Deux lignes de natures différentes : ouverte/fermée

La comparaison d'une ligne ouverte avec une ligne fermée se réalise en découpant la ligne fermée selon l'emprise de la ligne ouverte. Le calcul de la distance de Fréchet partielle est ensuite lancé. Si la valeur est satisfaisante ($< \epsilon$), un appariement partiel entre les deux tronçons est alors effectué.

Deux lignes fermées

Le cas des lignes fermées est probablement le plus simple car seules les lignes de même emprise sont appariables (figure 50).

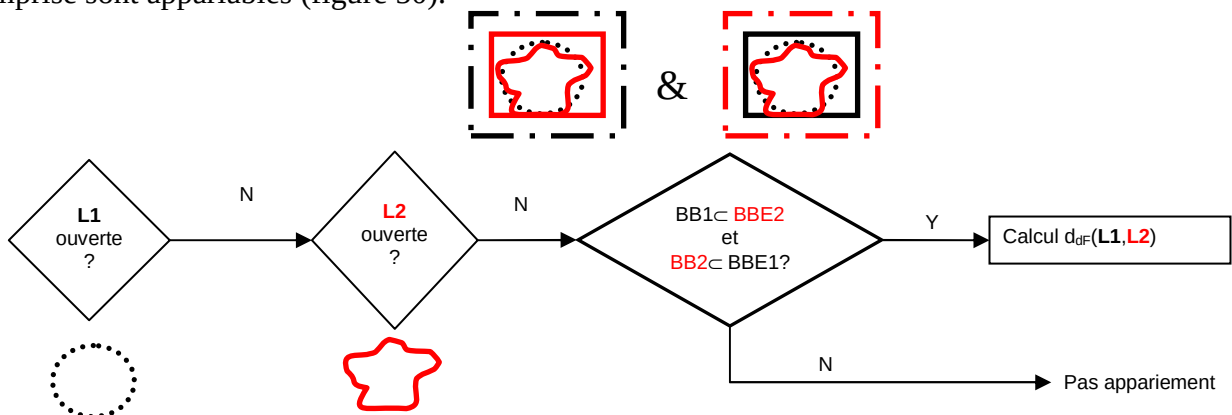


Figure 50 : Algorithme d'appariement de deux lignes fermées.

II.3.3.2 Appariement total/partiel « 1 – P » entre deux lignes

Cette catégorie d'appariement concerne les lignes qui ont une relation 1-n : à une ligne d'une BD1 correspond une, plusieurs lignes ou une partie d'une ligne de la BD2. Un découpage va donc être nécessaire, ce qui implique que cet appariement ne se réalise que sur des lignes ouvertes.

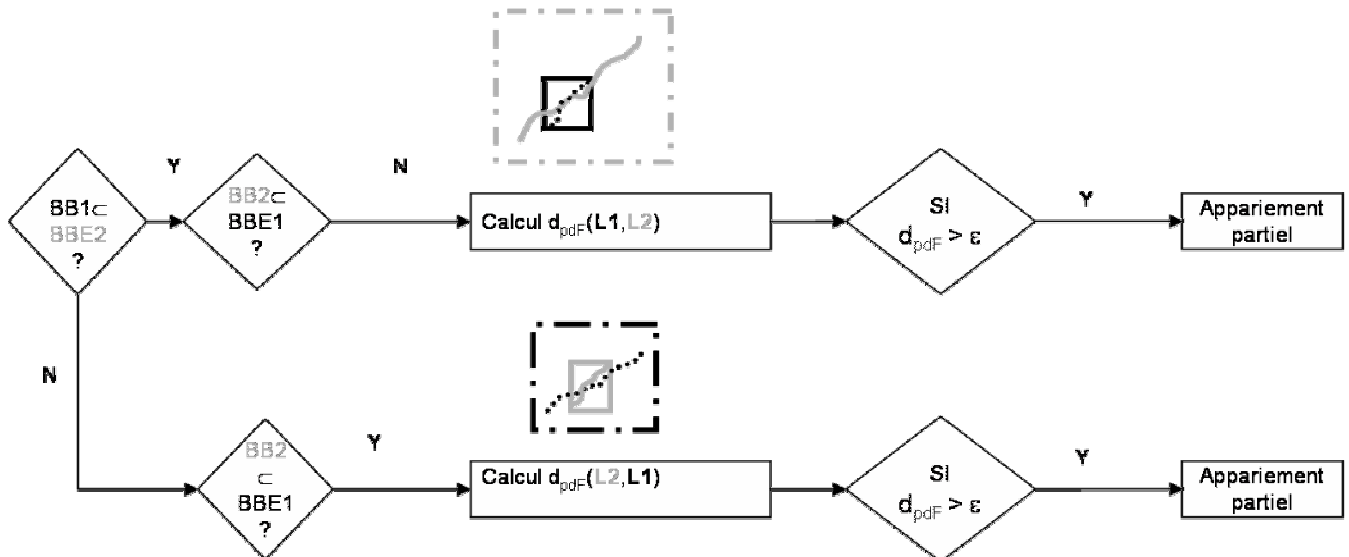


Figure 51 : Algorithme d'appariement total/partiel entre deux lignes ouvertes.

La méthodologie présentée figure 51 détaille les différentes actions effectuées pour le calcul de la distance de Fréchet partielle entre des lignes incluses. La première étape consiste à vérifier la relation d'inclusion entre les rectangles englobants simple de L_1 et élargi de L_2 . Si la relation est vraie, alors la droite la plus grande (L_1 dans notre exemple) est découpée selon le rectangle englobant élargi de la plus petite (L_2) à l'aide de la fonction « sélection rectangle » (voir figure 54, étape 2).

Celle-ci renvoie le segment de L_1 candidat à l'appariement avec L_2 . L'orientation des lignes est par la suite vérifiée. A partir de ces lignes, les deux matrices, de distance euclidienne et de Fréchet, sont alors construites comme dans l'exemple présenté figure 29 et **Erreur ! Source du renvoi introuvable.** Les couples de points homologues de début et de fin entre L_1 et L_2 sont recherchés, et les distances de Fréchet discrète partielle et moyenne sont calculées.

II.3.3.3 Appariement « partiel/partiel » (PP) et « deux partiel/partiel » (2PP) entre deux lignes

L'appariement de type partiel/partiel se calcule à l'aide de la distance de Fréchet partielle et peut s'appliquer à tout type de lignes quelles que soient leur emprise respective (voir figure 55 pour le récapitulatif de tous les cas possibles). Cependant, dans notre problématique d'intégration, il n'a été implémenté que pour les lignes ouvertes possédant des emprises avec un recouvrement partiel.

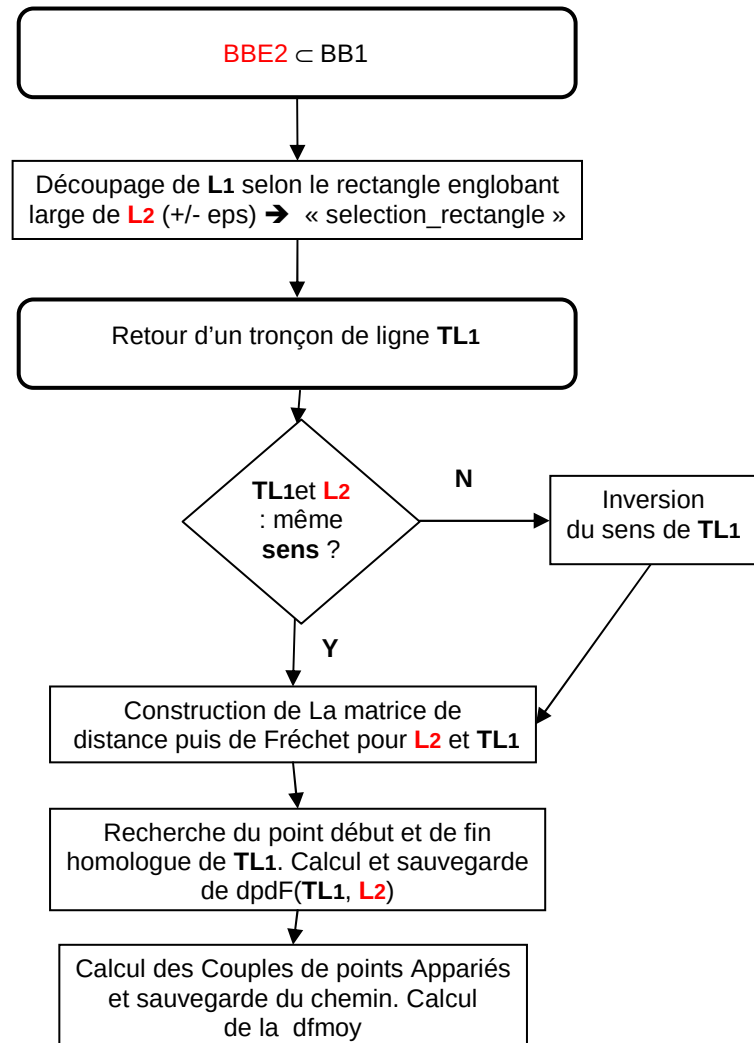
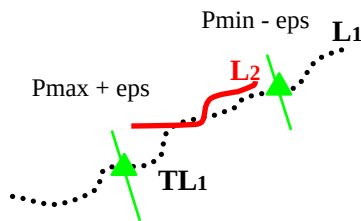
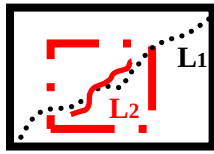


Figure 52 : Méthodologie de découpage d'une ligne L_1 selon une ligne L_2 d'emprise plus petite en vue du calcul de la distance de Fréchet partielle d_{pdF} .

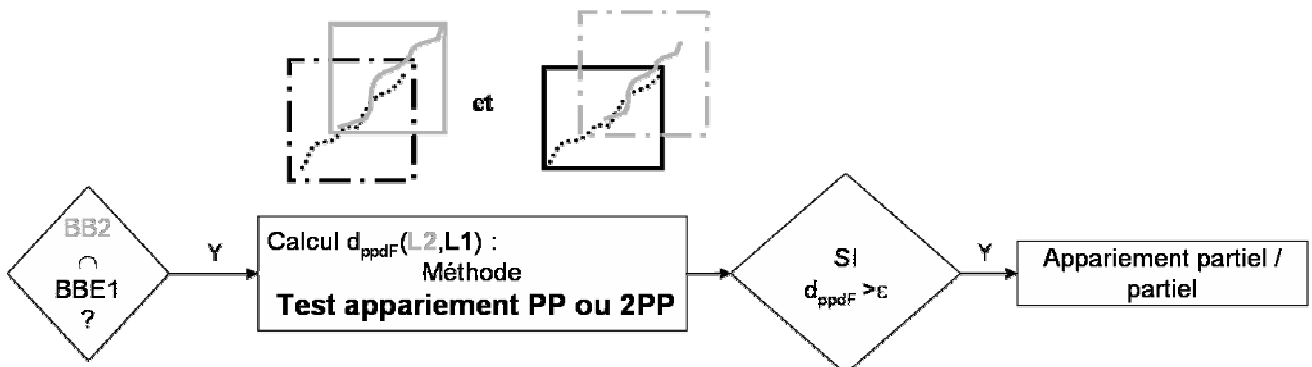


Figure 53 : Algorithme d'appariement partiel/partiel ou deux partiel/partiel de lignes possédant un domaine de recouvrement entre leurs emprises.

Lors du test du type d'appariement entre PP ou 2PP (« **Test appariement PP ou 2PP** » figure 53), neuf catégories sont identifiées (voir détail figure 54) :

- quatre concernent uniquement l'inclusion d'une seule extrémité d'une ligne dans le rectangle élargi de la seconde (illustration figure 42 **(a)**) : appariement partiel/partiel ;
- une concerne les deux extrémités des lignes (illustration figure 42 **(b)**) : appariement « deux partiel/partiel » ;
- quatre concernent une extrémité d'une ligne L_1 et deux extrémités d'une ligne L_2 (illustration figure 42 **(c)**) : appariement partiel/partiel.

Il est à noter que l'algorithme effectue le test des neuf cas en parallèle et que la (ou les) meilleure(s) valeur(s) de d_{pdF} est/sont retenue(s) à la fin après tri.

Le premier test de l'algorithme présenté figure 54 vise tout d'abord à identifier pour deux lignes (L_1 et L_2), dont le recouvrement n'est ni total ni inclus, si au moins une des extrémités (le début par exemple) de L_1 est comprise dans le rectangle englobant élargi de L_2 (fin non incluse). Si oui, le test suivant regarde si l'inverse est vrai, à savoir qu'une des extrémités de L_2 (par exemple le début) est comprise dans le rectangle englobant élargi de L_1 et pas l'autre (fin de L_2). Si c'est le cas, la d_{pdF} est calculée et conservée pour comparaison avec les autres valeurs.

Les trois tests suivants vérifient uniquement l'orientation des lignes entre elles en calculant les d_{pdF} pour chaque sens de ligne.

Le quatrième test concerne le cas où les deux extrémités (début et fin) de chaque ligne (L_1 et L_2) sont proches. Elles sont alors comprises dans les rectangles englobants élargis respectifs. Comme il n'est pas possible ici de déterminer à priori le sens des lignes, les quatre d_{pdF} sont calculées et la meilleure est gardée (se référer pour cela à l'exemple exposé figure 45).

Le test cinq traite le cas illustré figure 42 **(c)** où il est difficile de déterminer si c'est le début ou la fin de la ligne L_1 qui est homologue à l'extrémité de L_2 .

Les trois autres tests réitèrent le calcul en changeant les extrémités concernées de façon à vérifier l'orientation des lignes entre elles.

Le tableau 15 résume l'algorithme de la figure 54 :

- si un seul des cas A, B, C ou D est vrai, alors seul un appariement PP est effectué ;
- si A et D ou B et C sont vrais, alors un appariement 2PP est testé ;
- si un autre couple est vrai, la décision de l'appariement à réaliser est laissée au choix de l'utilisateur.

Tableau 15 : Récapitulatif des cas possibles d'appariement PP et 2PP.

Segment inclus de L_1	Segment inclus de L_2	distance	cas
$L1.n \subset BBE2$	$L2.1 \subset BBE1$	$d_{pdF}(L_1, L_2) < \epsilon$	A
$L1.1 \subset BBE2$	$L2.1 \subset BBE1$	$d_{pdF}(L_{1inv}, L_2) < \epsilon$	B
$L1.n \subset BBE2$	$L2.m \subset BBE1$	$d_{pdF}(L_1, L_{2inv}) < \epsilon$	C
$L1.1 \subset BBE2$	$L2.m \subset BBE1$	$d_{pdF}(L_{1inv}, L_{2inv}) < \epsilon$	D

Méthode test appariement PP ou 2PP

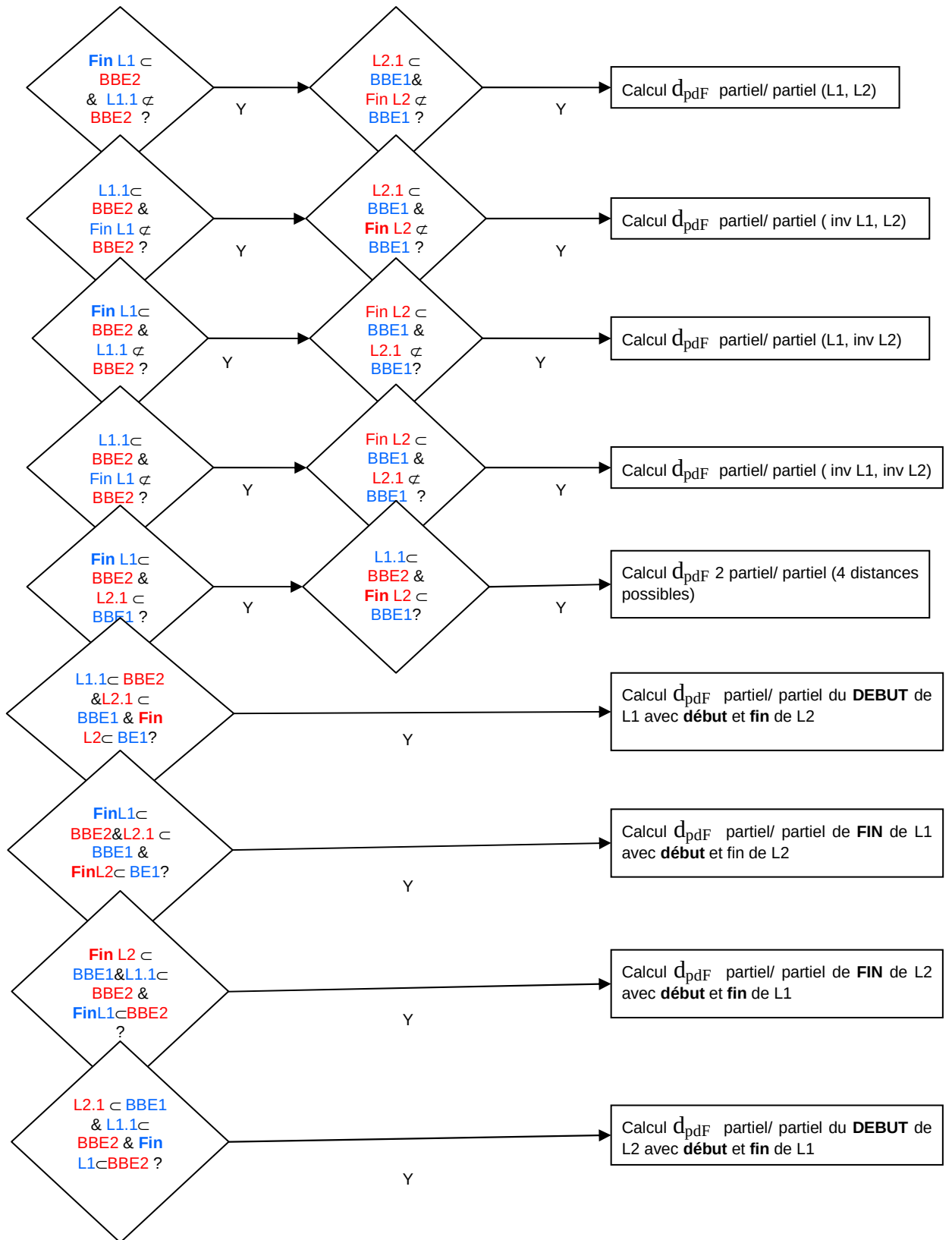


Figure 54 : Algorithme de test d'appariement partiel/partiel (PP) et deux partiel/partiel (2PP).

Pour proposer une synthèse des différents appariements possibles et des distances de Fréchet calculées dans chacun des cas, voici un tableau récapitulatif exprimé en fonction des relations d'inclusion :

Tableau 16 : Résumé des différents appariements présentés et de leurs distances de Fréchet associées.

Type appariement	Type L_1	Type L_2	RIE (Relation d'inclusion entre les emprises) (L_1, L_2)	Distances
1-1	O	O	équivalence	$d_F(L_1, L_2) < \epsilon$
	F	F		
1-P (non 1-1)	O	O	équivalence	$d_{pdF}(L_1, L_2) < \epsilon$ ou $d_{pdF}(L_2, L_1) < \epsilon$. Si les deux sont vrai $\rightarrow$ choix utilisateur
			inclut	$d_{pdF}(L_1, L_2) < \epsilon$
			est inclus	$d_{pdF}(L_2, L_1) < \epsilon$.
	O	F	est inclus	$d_{pdF}(L_2, L_1) < \epsilon$.
PP et 2PP (non 1-P)	O	O	Non disjoint	7 cas possibles
2PP	O	O	recouvrement	$d_{ppdF}(L_1, L_2) < \epsilon$ ou $d_{ppdF}(L_1, L_{2inv}) < \epsilon$ ou $d_{ppdF}(L_{1inv}, L_2) < \epsilon$ ou $d_{ppdF}(L_{1inv}, L_{2inv}) < \epsilon$.

Les différents cas de figure rencontrés lors de la comparaison de lignes et les appariements induits sont repris dans l'algorithme qui décrit la méthode globale d'appariement par fichier. Les lignes ne sont plus comparées individuellement selon le choix de l'utilisateur mais par l'intermédiaire de fichiers regroupant l'ensemble des lignes.

II.3.4 Méthode globale : appariement par fichier

La méthode globale reprend tous les modules de comparaison et appariement décrits précédemment et les enchaîne selon un ordre précis. Les bases de données de lignes à comparer sont regroupées dans deux fichiers. Chaque fichier est importé dans un tableau (ou vecteur) de lignes.

L'algorithme de la figure 55 présente la méthode de traitement des fichiers de lignes dans son ensemble. Une comparaison deux à deux (L_i, L_j) avec L_i dans le fichier 1 et L_j dans le fichier 2 est effectuée. Après importation des lignes, leur nature (fermée ou ouverte) est ensuite déterminée : en effet les comparaisons successives impliquées ne sont pas les mêmes selon la situation considérée. Les trois cas possibles évoqués précédemment (ouvert/ouvert, ouvert/fermé ou fermé/fermé) déterminent les types d'appariement testés : total (ou partiel), total/partiel seul, partiel/partiel ou deux partiel/partiel.

Les distances de Fréchet partielles inférieures à ϵ sont sauvegardées dans un fichier dans un format qui permet une visualisation rapide des résultats via un SIG.

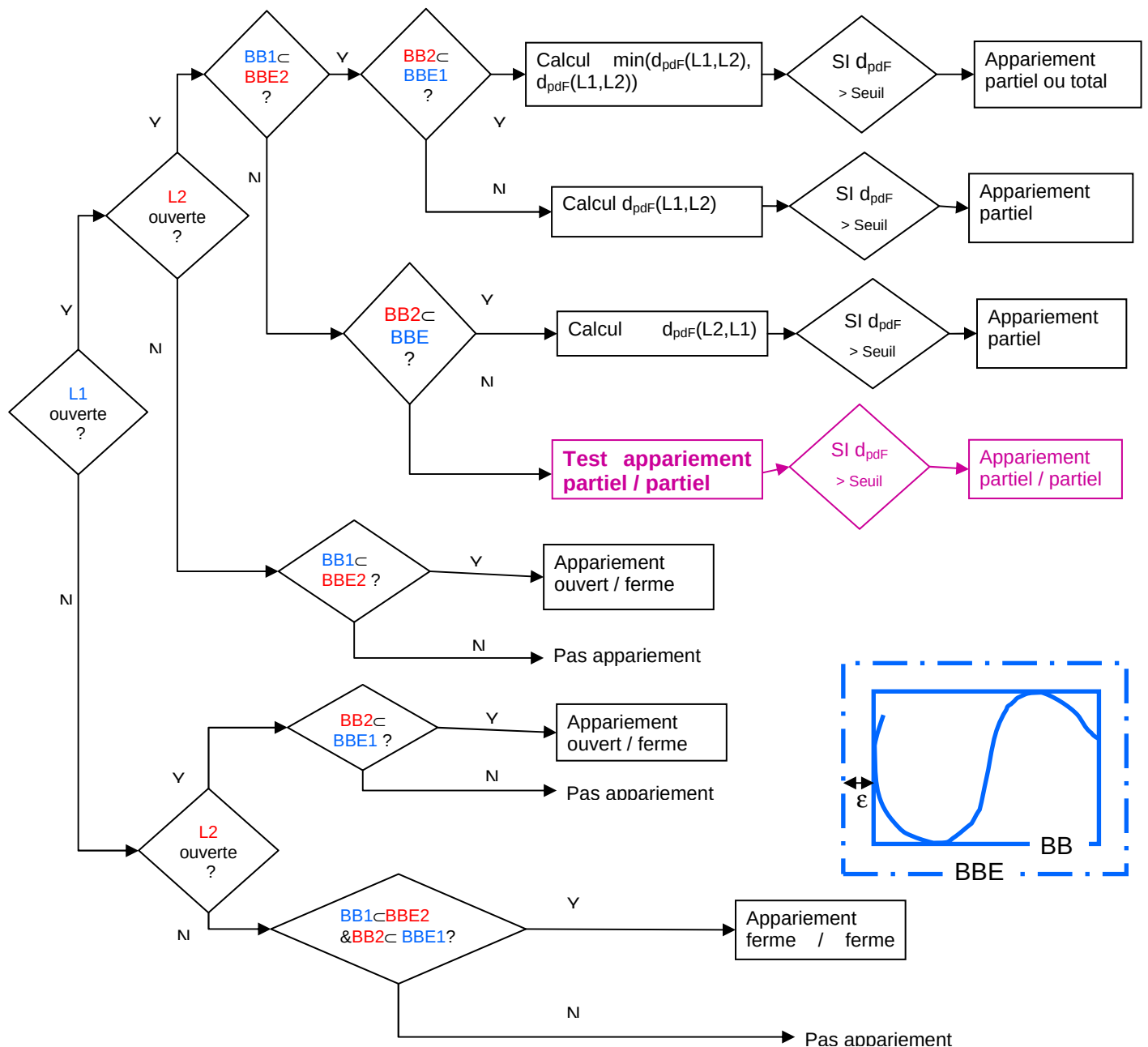


Figure 55 : Algorithme d'appariement global.

Le tableau 17 dresse un inventaire de tous les cas existants : ceux identifiés et traités, et ceux impossibles à résoudre. Les cas sont classés en fonction de la nature des deux lignes comparées (ouverte/fermée), de leur emprise (totale, incluse ou partielle) et du type d'appariement. Les rectangles englobants simples sont en trait plein et ceux étendus en trait pointillé.

Les cases contenant seulement une croix sont celles pour lesquelles il n'existe pas de cas possible. Les dessins grisés barrés d'une croix sont les situations pour lesquelles il n'y a pas de solution en l'état actuel de la recherche. Les dessins en gris sont les cas traités algorithmiquement mais non encore codés dans les outils d'intégration.

Cette méthode globale permet de traiter la majorité des cas parmi ceux identifiés : sont exclus ceux qui ne possèdent pas de solution simple (cas en gris clair dans le tableau).

L'implémentation de toutes ces méthodes de détection de cas, de tests des différents types de lignes, de leurs sens respectifs, et de calculs des appariements possibles, représente environ 5.300 lignes de codes réparties en 11 classes et 209 méthodes.

Afin de vérifier la validité de ce code au travers de la méthodologie globale, il est impératif de la tester sur des données simulées, représentatives de données réelles. De ces données sont extraites les lignes caractéristiques nécessaires à l'intégration, mais le problème de recouvrement est exclu.

Tableau 17 : Récapitulatif des différents cas de figure rencontrés lors de la comparaison de lignes (1) pour chaque type de ligne (ouverte ou fermée) ; (2) pour chaque type de recouvrement : total, inclus ou partiel (rectangles englobants simple et étendu associés).

	Total/total	Partiel/total	Partiel/ partiel	2 partiel /partiel
OO				
OO	X			
OO	X	X		
OF	X			X
OF	X		X	
OF	X	X		
FF			X	X
FF	X	X		
FF	X	X		

Le chapitre suivant développe plus en détail ces deux aspects : l'extraction d'éléments caractéristiques (présentés au chapitre III) de MNT possédant un recouvrement total, puis une intégration établie sur des modèles Terre/Mer avec un recouvrement partiel.

L'étude suivante détaille dans la section suivante une application de cet algorithme à la comparaison de traits de côte. L'objectif est de quantifier les différences entre une ligne de référence et une ligne obtenue avec des données plus récentes mais de moins bonne résolution.

II.4 Exemple d'applications de la méthode sur la comparaison de traits de côte

Cette étude, détaillée dans les travaux de [Le Berre et al., 2005], a été réalisée dans le cadre de l'appel d'offre CNES-IFEN d'août 2002 relatif au suivi du littoral à grande échelle par imagerie SPOT 5. L'étude présente l'expérimentation réalisée pour la délimitation et la caractérisation du « trait de côte géomorphologique » tel que défini par le Conseil National de l'Information Géographique (CNIG). Ce trait est défini comme étant « *la ligne de discontinuité morphologique caractéristique de la limite probable atteinte par la mer sur la côte* » [CNIG., 2003]. Une comparaison des traits de côte issus de la photo-interprétation de la BD ORTHO® et d'une image SPOT 5 avec des données de référence produites lors de levés topographiques sur le terrain est réalisée à l'aide de méthodes basées sur la distance linéaire de Fréchet. Cette comparaison produit un résultat numérique quantitatif qu'il est possible de comparer à une évaluation qualitative du numéricien.

II.4.1 Présentation des données et de la zone d'étude

Les données de référence pour la validation des résultats sont extraites des bases d'information géographique développées au laboratoire GEOMER sur le littoral finistérien ou ont été fournies par le Ministère de l'Équipement¹³ via le CETMEF et la DDE du Finistère, partenaires de ce projet. La comparaison quantitative des différents traits de côte produits et acquis est réalisée par l'institut de recherche de l'école navale (IRENav) à l'aide de méthodes basées sur la distance de Fréchet. Cette comparaison complète ainsi l'évaluation qualitative de la numérisation déjà réalisée.

Le projet a bénéficié de la programmation du satellite SPOT 5 pour l'acquisition d'une image au cours d'une basse mer de vive-eau sur le littoral nord-ouest finistérien. Une image en mode multispectral à 2,5 m de résolution a été acquise le 17 avril 2003, lors de la basse mer, ce qui permet une très bonne visualisation de l'estran.

La zone d'application retenue pour la comparaison des traits de côte se situe dans deux secteurs du nord Finistère. Ils ont été choisis du fait (1) de leur représentativité des différents types de côtes de la région, (2) de la disponibilité des données de référence et (3) de la présence sur le terrain de l'équipe dans la période encadrant l'acquisition de l'image.

II.4.2 Méthode de comparaison des traits de côte

La méthode de comparaison des traits de côte employée s'appuie sur des levés topographiques, sur l'interprétation de l'image SPOT 5 et de la BD ORTHO® IGN, puis sur l'analyse comparative des différents traits de côte ainsi produits.

Des levés topographiques détaillés de sections du trait de côte ont été réalisés sur le terrain dans la perspective de comparer les tracés obtenus à ceux issus de l'interprétation de l'image SPOT 5 et de l'orthophotographie. Ces levés ont été effectués le jour du passage du satellite dans la presqu'île de Sainte Marguerite en Landeda dans des zones de falaises et de plages. Dans les secteurs de falaises, les deux lignes de référence définissant généralement le trait de

¹³ Nouvellement renommé le Ministère de l'écologie, de l'énergie, du développement durable et de la mer.

côte, à savoir le pied de falaise [Bonnot-Courtois & Levasseur, 2003] et le rebord sommital de la partie érodée du versant [Robin, 2002], ont été mesurées. Dans le cas du secteur sableux, seul le contact entre la limite de la végétation dunaire a été pris en considération pour la cartographie du trait de côte [Robin, 2002]. Sur le secteur artificialisé c'est la base des ouvrages qui a été relevée.

II.4.3 Résultats et perspectives

L'image SPOT 5 tout comme la BD ORTHO[®] IGN sont traitées par interprétation assistée par ordinateur afin d'étudier, de manière fine, les possibilités de délimitation et de caractérisation du trait de côte.

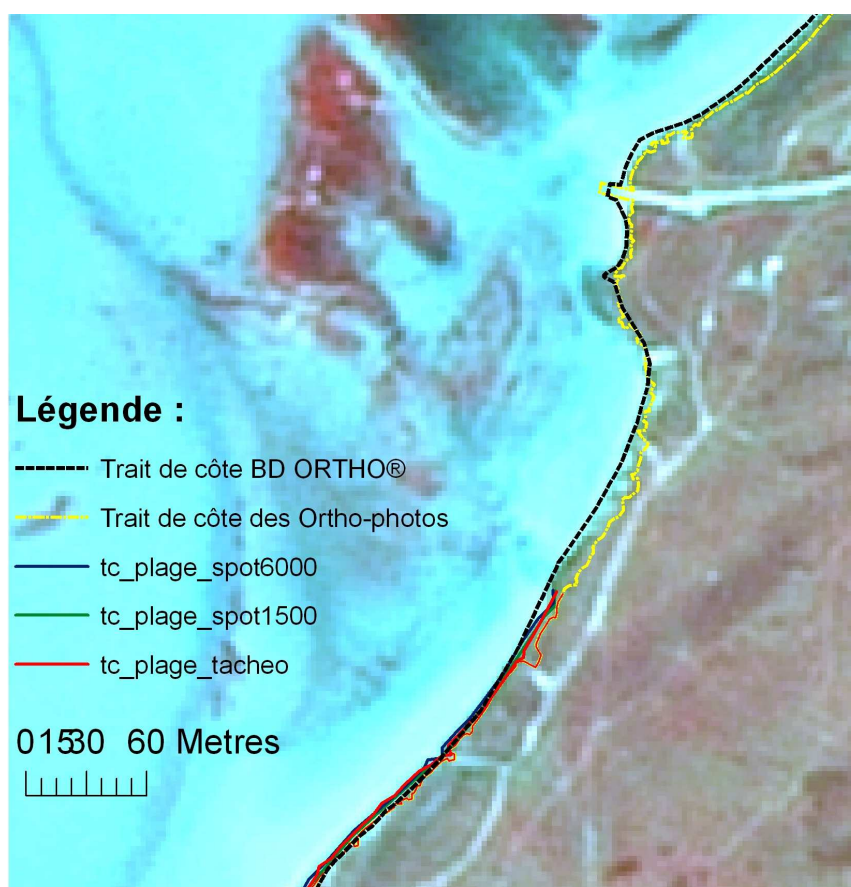


Figure 56 : Représentation des différents traits de côtes issus de la BD ORTHO[®], des levés de terrain et photo interprétés à partir des ortho-photographies, et de l'image SPOT 5.

L'image SPOT 5 a fait l'objet de deux numérisations séparées, la première à 1/6000 qui constitue une échelle confortable pour l'interprétation de l'image, la seconde à 1/1500 qui constitue l'échelle à partir de laquelle les pixels deviennent nettement visibles et commencent à constituer une gêne pour l'interprétation de l'image. La BD ORTHO[®] pour sa part a été numérisée à 1/1500, échelle de confort pour son interprétation.

Chaque numérisation aboutit à la production de couches d'information géographique vectorielles (voir figure 56). Deux traits de côte sont photo-interprétés d'après l'image SPOT 5 (aux échelles de 1/6000 et de 1/1500 : « tc_plage_spot6000 » et « tc_plage_spot1500 », légende Figure 56) et un troisième d'après la BD ORTHO[®] (à l'échelle de 1/1500 : « Trait de côte des Ortho-photos »). Les traits de côte de référence sont acquis par tachéométrie ou par GPS différentiel (nommés « tc_plage_tacheo »).

Les points des extrémités des segments constituant ces traits de côte sont exportés dans un fichier texte à l'aide d'Arc/Info. Ces lignes sont traitées par un exécutable développé en C++. Elles sont suréchantillonnées tous les demi-mètres pour être comparées à l'aide de mesures dérivées de la distance de Fréchet [Devogele, 2000]. La précision de numérisation entraîne une estimation décimétrique de l'écart entre une ligne de référence et une ligne saisie. Elles qualifient la précision de numérisation en comparant les traits de côte numérisés avec les traits de côte de référence.

Tableau 18 : Distance en mètres entre les lignes saisies et les lignes de référence : distance de Fréchet discrète partielle moyenne suivie de la distance de Fréchet discrète partielle entre parenthèses (précision décimétrique).

	Trait de côte de référence (levé tachéomètre ou GPS)					
Photo- interprétation	Rebord sommital de falaise (Brouennou)	Falaise végétalisée (baie des Anges)	Dune (Ste Marguerite)	Grève (baie des Anges)	Cordon littoral (baie des Anges)	Ouvrages côtiers (baie des Anges)
SPOT 5, 1/6000	3,34 (6,07)	4,54 (11,98)	2,25 (6,08)	0,96 (3,90)	2,01 (4,18)	2,37 (6,26)
SPOT 5, 1/1500	1,62 (3,78)	2,77 (10,84)	1,17 (3,79)	1,22 (5,20)	1,45 (3,39)	3,58 (14,01)
BD ORTHO®, 1/1500	1,03 (3,10)	2,8 (7,25)	2,1 (8,64)	1,41 (5,38)	1,82 (5,08)	1,77 (7,37)

Sur la zone de falaise vive et verticale, les résultats du tableau 18 montrent qu'en moyenne la précision de photo-interprétation réalisée à partir de la BD ORTHO® est meilleure que celle issue de SPOT 5. Pour la zone de falaise végétalisée, les résultats sont relativement proches et montrent que la végétation constitue manifestement une contrainte pour une numérisation précise.

En zone dunaire, les traits de côte issus de SPOT 5 sont légèrement meilleurs que ceux issus de la BD ORTHO®. La simple évolution qu'a connue la ligne de rivage en haut de plage entre le moment de la prise orthophotographique (juin 2000) et des acquisitions SPOT 5 et tachéométrique (avril 2003), explique la meilleure précision obtenue ici à partir de l'image SPOT 5. Elle démontre ainsi l'intérêt de l'imagerie SPOT 5 pour le suivi à court terme de l'évolution des littoraux meubles.

De même, les meilleurs résultats obtenus à l'aide de SPOT 5 pour l'interprétation du cordon de galet et de la zone de grève (figure 57) doivent être interprétés avec prudence et à la lumière d'une possible évolution du trait de côte ou de la végétation qui la recouvre entre les deux prises de vues.

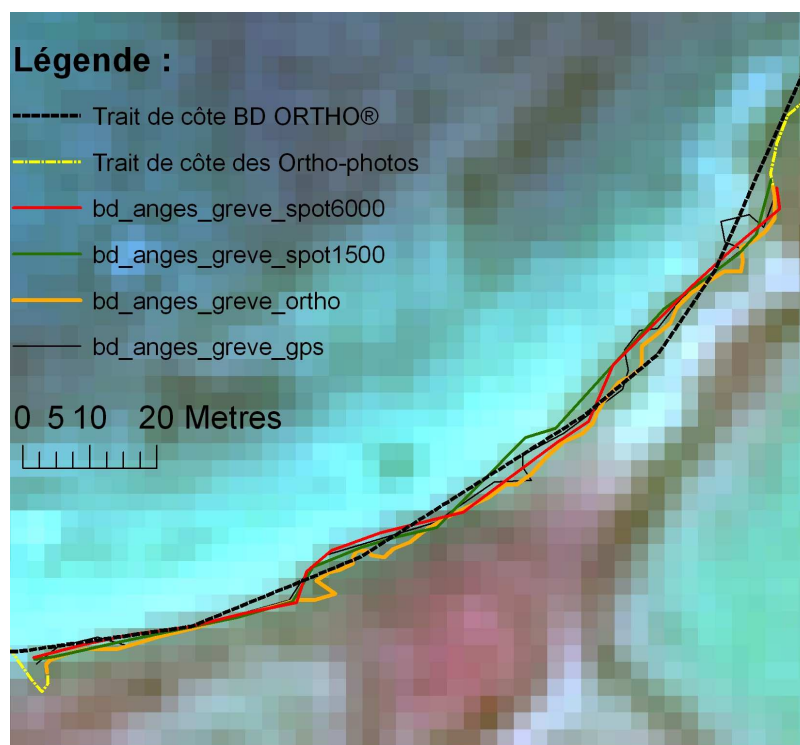


Figure 57 : Illustration des différents traits de côte numérisés au 1/1500 et 1/6000 comparés à la BD ORTHO® et à des levés GPS établis sur une grève.

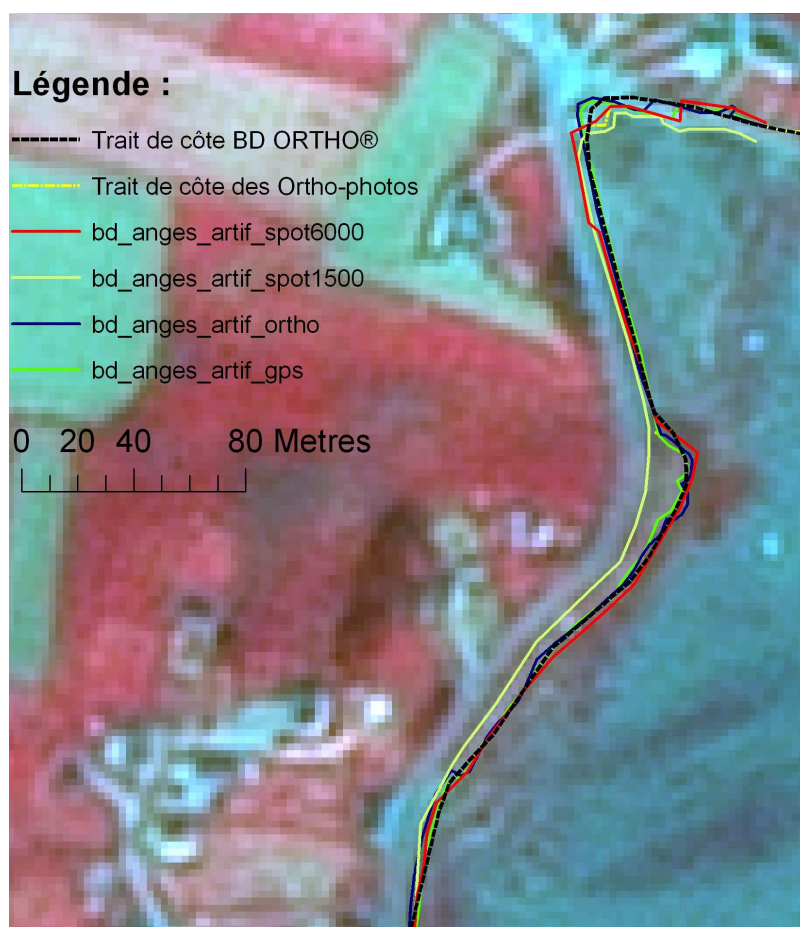


Figure 58 : Comparaison des traits de côte issus de photo interprétation à partir de l'image SPOT 5 avec le trait de côte de la BD ORTHO® sur une côte artificielle.

En zone artificielle l'ortho photographie permet l'obtention de résultats sensiblement meilleurs (25% d'amélioration).

Le faible résultat obtenu pour le trait de côte numérisé à 1/1500 (trait bleu figure 58) d'après SPOT 5 est essentiellement lié à une différence dans l'interprétation et dans le choix de l'objet à numériser, qui est également à l'origine de la forte distance de Fréchet discrète (écart maximal) dans les trois cas. Cette valeur témoigne de la difficulté à identifier précisément les ouvrages côtiers.

Ainsi, globalement, sur cette zone d'étude, si l'on excepte la numérisation effectuée au 6000ème d'après l'image SPOT 5, les photo-interprétations de SPOT 5 ont une précision du même ordre de grandeur que celles de la BD ORTHO®.

Au vu des résultats obtenus, tant en comparaison visuelle qu'à l'aide des valeurs obtenues avec la distance de Fréchet, il apparaît que l'utilisation de SPOT 5 semble appropriée pour la remise à jour du trait de côte. En effet, malgré des incertitudes de délimitation liées notamment aux effets d'ombre, certains secteurs donnent de bons résultats. Il s'agit principalement des zones dunaires sur lesquelles le contraste entre la partie terrestre et le domaine marin est souligné par la végétation, qui permet une bonne détermination de la limite côtière. Compte tenu de la mobilité de ce type de côte, une voie potentiellement intéressante peut être explorée.

D'après les tests de numérisation et la comparaison des mesures effectuées in situ, il apparaît que ces secteurs de côte peuvent être cartographiés avec une précision satisfaisante au regard de celle obtenue à partir de l'ortho-photographie. L'image SPOT 5 offre donc un certain nombre de perspectives en matière de suivi à court terme de l'évolution du trait de côte qui s'additionnent aux potentialités offertes par la répétitivité d'acquisitions de données.

II.5 Bilan sur les apports des extensions

Ce chapitre introduit tout d'abord les différents types de lignes présentes lors d'une comparaison : ligne ouverte, fermée ou d'orientations différentes. Ces notions sont essentielles pour comprendre la nécessité des extensions apportées à la distance de Fréchet. En effet la méthode définie par [Eiter & Mannila, 1994] ne s'applique qu'à des lignes de tailles similaires et de même orientation. Les données géographiques remplissent rarement ces deux impératifs.

Par ailleurs, la comparaison de lignes de forme complexe amène à définir d'autres distances complémentaires nécessaires aux besoins de cette thèse : distance moyenne, distance partielle entre lignes, distance entre lignes fermées, ou encore entre des extrémités de ligne uniquement (distance « deux partiel/partiel »).

La section trois de ce chapitre reprend donc ces distances et les intègre dans les algorithmes d'appariement de lignes en décrivant les trois étapes du processus d'appariement. La première étape détaille tout d'abord le processus de recherche de la nature des lignes comparées et vérifie leur orientation. La seconde étape inventorie les types d'appariement mis en œuvre dans la troisième étape au moyen des relations d'inclusion existant entre les lignes :

- total,
- total/partiel,
- partiel/partiel ou deux partiel/partiel.

La dernière partie synthétise les cas possibles au travers de la description d'une méthode de comparaison et d'appariement globale entre des fichiers de lignes.

L'exemple d'application détaillé en dernière section permet d'illustrer une application possible de la distance de Fréchet tout en validant les divers algorithmes de comparaison de lignes présentés en section 4.1. En revanche cette comparaison s'est faite sur des lignes 2D numérisées choisies par l'utilisateur.

Il s'agit donc maintenant de tester l'algorithme d'appariement global sur des lignes extraites des MNT (thalwegs, crêtes, ruptures de pente) en l'intégrant à une méthodologie d'intégration générique. Cette méthodologie doit permettre d'intégrer tous types de données quel que soit le cas de figure rencontré. Pour chaque élément apparié, des mesures de distances moyenne, maximale ainsi que les couples de points homologues sont obtenues. Cette étape originale est absente des autres méthodes d'intégration présentées en état de l'art. Elle permet d'apparier des éléments morphologiques caractéristiques extraits des modèles numériques de terrain. La méthodologie globale d'intégration des MNT terrestre et maritime va maintenant être détaillée ainsi que son application à des données simulées.