

Evaluation et modélisation des possibilités d'occupation du sol dans un contexte d'incertitude forte.

INTRODUCTION

L'objectif de ce chapitre est double : son but « final » consiste en une évaluation du peuplement possible des zones de forte incertitude, c'est-à-dire les *franges* et les zones de moindre prospection. Afin d'y parvenir, une description et une évaluation des logiques de localisation des sites connus seront nécessaires, dans le but de bâtir les *possibles* sur des critères empiriques, basés sur l'observation des tendances et des régularités dans la répartition des phénomènes, et non uniquement théoriques. Cette compréhension des logiques est donc envisagée à la fois comme un moyen (pour évaluer le peuplement des zones non prospectées) mais également comme une « fin » à part entière, dans le but de comprendre si la distinction de *hot-spots* et d'*outliers*, et donc la régionalisation établie précédemment sur la base des structures spatiales détectées par le calcul des LISA, se justifie d'un point de vue thématique et comment elle se décline dans ses relations avec l'environnement, dans l'espace et dans le temps.

1. METHODE D'APPREHENSION DE L'INCERTITUDE FORTE PAR LA THEORIE DES POSSIBILITES

1.1 Quels niveaux d'information conserver dans la recherche des logiques d'occupation du sol?

Comme cela a déjà été mentionné, notre objectif consiste à appréhender les logiques de localisation des sites archéologiques afin d'estimer les zones de peuplement possibles sur l'ensemble de la zone étudiée. Ces logiques ne seront toutefois pas appréhendées à partir de critères théoriques ou présumés, mais évalués de manière empirique à partir de la localisation des sites connus.

Au vu des résultats obtenus dans le chapitre précédent, nous postulons (et souhaitons vérifier) que les logiques de localisation des sites diffèrent en fonction de critères à la fois thématiques et relatifs aux données:

- Du temps considéré : le contexte socio-environnemental dans lequel évoluent les peuplements passés ayant pu évoluer d'un temps à l'autre, il est également fortement possible que les besoins et les priorités de ces populations aient évolué. Ainsi, toute modélisation du peuplement possible ne peut s'effectuer que du point de vue d'une période donnée, et n'est valable que pour cette période.
- Du type de configuration spatiale à laquelle les sites appartiennent (*hot-spot* ou *outlier*) : si l'on postule que ces deux types de configurations spatiales révèlent des modes d'occupation et éventuellement d'utilisation du sol différents, les peuplements qui y sont associés peuvent également relever de logiques de localisation différentes.
- Du niveau temporel considéré : selon le niveau d'observation temporel, on ajoute de nouveaux sites (attachés au *temps* 1A+1Ba) pouvant, en fonction de leur localisation, faire varier les grandes tendances d'occupation du sol observées.
- Du niveau de fiabilité des données considéré : selon le seuil de fiabilité des données pris en compte, on ajoute ou on retranche des données à l'analyse qui, tout comme pour l'ajout d'un second niveau temporel, peuvent faire varier les tendances observées dans l'occupation du sol.

Chacun de ces niveaux d'information sera donc pris en compte dans l'évaluation des logiques d'occupation du sol.

1.2 Le choix de la logique floue pour la modélisation de connaissances incertaines

D'après L. Gacogne (2003) « la logique floue, ou plus généralement le traitement des incertitudes, a pour objet d'étude la représentation des connaissances imprécises et le raisonnement approché ». Pourtant encore très peu adopté en archéologie (cf chapitre 1) nous postulons que ce cadre théorique et méthodologique est particulièrement adapté à ses problématiques, notamment lorsqu'il s'agit de modéliser le peuplement inconnu sur des zones peu prospectées à partir de connaissances elles-mêmes incertaines (c'est-à-dire ce que l'on appelle communément « l'archéologie prédictive »).

1.2.1 Logique floue et possibilités versus probabilités : quel cadre théorique choisir ?

Le cadre envisagé dans le traitement de l'incertitude en archéologie (mais aussi souvent en géographie) est bien souvent celui des probabilités (cf chapitre 1). L'explication qui suit prend l'exemple de l'estimation de la localisation de sites archéologiques envisagée dans le cadre de la

théorie des probabilités et des possibilités, afin de montrer la différence essentielle qui réside entre les deux approches et de justifier notre choix de la théorie des possibilités.

Dans le cadre de la théorie des probabilités, la somme des probabilités de la localisation d'un site archéologique est obligatoirement égale à 1, et toutes les alternatives doivent être connues et évaluées quantitativement. Ainsi, la probabilité de localisation d'un site n'est pas une valeur intrinsèque de sa propre incertitude, mais une valeur relative à la probabilité de localisation des autres sites. Dans le cadre de la théorie des possibilités, la possibilité de localisation d'un site A n'influe pas sur la possibilité de localisation des autres sites : celle-ci est une valeur intrinsèque à la localisation du site A. De ce fait, l'ensemble des possibilités de l'ensemble des sites n'est généralement pas égal à 1. Zadeh (1978) utilise un exemple simple pour illustrer la différence entre les probabilités et les possibilités, et propose de considérer l'assertion: « Hans a mangé X œufs au petit déjeuner ». L'incertitude de X est évaluée dans le Tableau 23.

Tableau 23 - Logique de possibilités et de probabilités, l'exemple des œufs au petit déjeuner (Dutozia, 2013 d'après Zadeh, 1978).

NOMBRE D'ŒUFS	1	2	3	4	5	6	7	8
POSSIBILITE $\pi_x(u)$	1	1	1	1	0,8	0,6	0,4	0,2
PROBABILITE $P_x(u)$	0,1	0,8	0,1	0	0	0	0	0

On constate ainsi que si la possibilité pour que Hans ait mangé 3 œufs est de 1, la probabilité pour qu'un tel évènement se produise est faible, de 0.1. Ainsi, un faible degré de possibilité n'implique pas forcément un faible degré de probabilité (et inversement). Cependant, si un évènement est impossible, on peut supposer qu'il est improbable.

On voit également que le cadre des probabilités offre un ensemble bien plus restreint de choix que celui des possibilités au nombre d'alternatives plus ouvert, correspondant en fait au cône des localisations probables de sites modélisées classiquement en archéologie « prédictive ». Le tableau ci-dessus nous apprend cependant que certains évènements improbables sont des réalités tout à fait possibles : la théorie des possibilités permet ainsi, au-delà des situations les plus habituelles (également prises en compte par les probabilités) où $X = 1$ à 3, de prendre en compte des situations plus inhabituelles ($X = 4$ à 8), peut être rares, mais possibles exceptionnellement.

Face au haut degré d'incertitude auquel nous sommes confrontés dans les problématiques archéologiques et dans l'évaluation des « passés possibles », l'utilisation de méthodes plus

ouvertes que le cadre des probabilités doit être adoptée. Notamment parce-que ce dernier suppose, en amont une connaissance en amont presque exhaustive du système envisagé, ce dont nous ne pouvons nous targuer. Le cadre des possibilités offre l'opportunité d'intégrer dans l'estimation du peuplement certains facteurs de localisation de sites dont on assume ne pas connaître précisément l'ensemble des composantes et des relations possibles.

P. F. Fisher (2005) distingue les cas où les probabilités et la logique floue (sur laquelle se base la théorie des possibilités) peuvent être mobilisées : « If both the class and the individual are well defined then the uncertainty is caused by errors and is probabilistic in nature. (...) If the uncertainty is attributable to poor definition of class of object or individual object then definition of a class or set within the universe is a matter of vagueness, and this can conveniently be treated with fuzzy set theory ».

1.2.2 Définition d'un sous-ensemble flou

La logique floue a « donné lieu à toute une généralisation des mathématiques classiques fondée sur le concept très simple de sous-ensemble flou » (Gacôgne, 2003).

Un ensemble classique comporte des éléments qui satisfont l'ensemble des propriétés précises. Ainsi, les éléments qui ne satisfont pas ces propriétés ne peuvent appartenir à cet ensemble. De manière plus formelle, un sous-ensemble A d'un ensemble de référence X peut être décrit par sa fonction caractéristique $\chi_A: X \rightarrow \{0, 1\}$ de la manière suivante (Bouchon-Meunier et al., 2003) :

$$\chi_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$$

Par exemple le sous-ensemble A des âges compris entre 20 et 40 ans a pour fonction caractéristique :

$$\chi_A(x) = \begin{cases} 1 & \text{si } 20 \leq x \leq 40 \\ 0 & \text{sinon} \end{cases}$$

Le graphe de cette fonction est représenté dans la Figure 107.

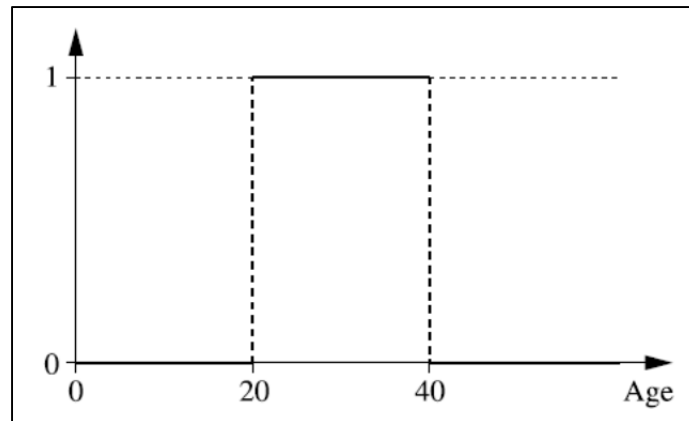


Figure 107 - Fonction caractéristique de l'ensemble A (Bouchon-Meunier et al., 2003)

Si l'on considère désormais un ensemble B des âges avoisinant les 30 ans, la propriété *avoisinant les 30 ans* n'est pas précise : ainsi, contrairement au cas évoqué plus haut, B ne peut être caractérisé par une fonction caractéristique qui séparerait en deux les âges avoisinant les 30 ans de ceux qui ne les avoisinent pas.

Une généralisation de la fonction caractéristique en *fonction d'appartenance* est alors nécessaire dans le but de considérer les âges qui ne sont pas trop éloignés des 30 ans, sans en être vraiment proches non plus. Cette nouvelle fonction permet alors de mettre en évidence les nuances d'appartenance des éléments de l'ensemble de référence X, et de définir un sous-ensemble flou (on dit aussi *ensemble flou* pour simplifier) de X.

La définition d'un sous-ensemble-flou donnée par Bouchon-Meunier et al. (2003) est la suivante :

« Un sous-ensemble flou F de X est défini par une fonction d'appartenance μ_F qui associe à tout élément x de X une valeur réelle $\mu_F(x)$ dans l'intervalle $[0, 1]$.

Ainsi, un sous-ensemble flou est toujours (et seulement) une fonction de X dans $[0, 1]$ ».

La Figure 108 donne un exemple de graphe d'une fonction d'appartenance possible pour l'ensemble B.

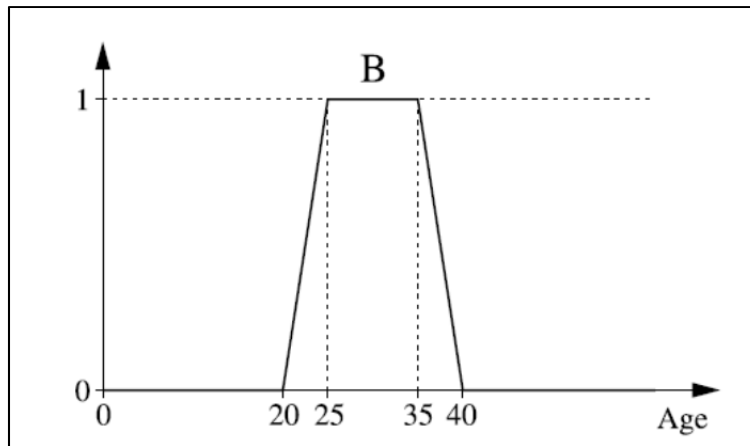


Figure 108 - Fonction d'appartenance de l'ensemble B pour la propriété *avoisinant les 30 ans* (Bouchon-Meunier et al., 2003)

La modélisation de ce que Zadeh (1975) nomme des « variables linguistiques » est ainsi une possibilité offerte par la logique floue.

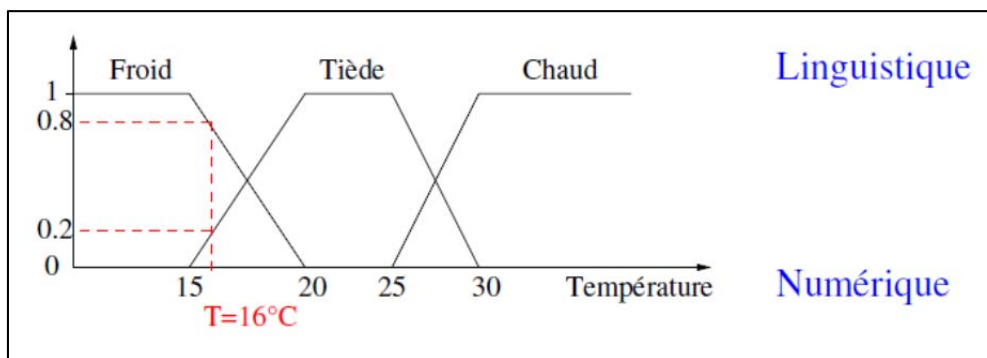


Figure 109 – Exemple d'ensembles flous de la variable *Température* (Guillaume, Charnomordic, 2013)

La Figure 109 montre un exemple de variable linguistique « température » avec trois termes linguistiques, « froid », « tiède », « chaud ». Ainsi, la valeur numérique de température 16°C peut être considérée comme « froide » avec un niveau de 0,8 et comme « tiède » à un niveau de 0,2. Elle n'est pas considérée comme chaude.

1.3 Le choix d'une démarche empirique et exploratoire

La démarche adoptée consiste ainsi à évaluer les paramètres ayant pu entrer en compte dans les préférences de localisation des sites (pour chacun des différents niveaux mentionnés en 1.1), puis à partir de l'observation de ces critères sur l'ensemble de la zone étudiée, transposer ces observations en termes de possibilités d'occupation des zones non prospectées en fonction des caractéristiques de ces zones.

Notre démarche se base donc sur des critères empiriques, et diffère en cela de la démarche adoptée par Bernard Geyer dans l'élaboration de la carte d' « attractivité » pour les Marges Arides (Figure 10), recouvrant une partie de l'espace étudié. L'appréciation des niveaux d' « attractivité » repose en effet sur une évaluation experte des critères théoriques dont on suppose qu'ils aient joué un rôle dans la localisation préférentielle des sites en un lieu donné plutôt qu'un autre. Ce modèle théorique est ensuite confronté à la localisation des sites par période, afin d'en vérifier la validité. Notre démarche est complémentaire à celle-ci en ce que l'évaluation des critères ne se réalise pas par connaissance experte de manière théorique mais en se basant sur les aspects de la localisation des sites connus et en en inférant des possibilités de localisation de sites dans les zones à prospecter.

La démarche est exploratoire en ce que la corrélation entre les critères à tester (c'est-à-dire, les informations dont nous disposons afin d'évaluer les préférences de localisation des sites) et la localisation des sites est évaluée statistiquement en amont du calcul des possibilités. Il s'agit donc en tout premier lieu, de tester l'impact des divers critères dont on suppose qu'ils aient pu jouer un rôle dans la localisation des sites, afin de sélectionner les plus pertinents.

Selon Hans Kamermans (2006), il semblerait que la démarche adoptée ici corresponde à ce que l'on appelle une démarche "inductive" en archéologie "prédictive": « With the inductive approach to predictive modeling, a model is constructed on the basis of correlations between known archaeological find spots and attributes that are predominantly taken from the current physical landscape. It is only a predictive model when the observed correlations are extrapolated ». Bien que cette définition corresponde globalement à notre démarche au sens méthodologique, les termes employés ne conviennent pas tout à fait à notre philosophie, et nous préférons nous ranger au discours de Thomas G. Whitley (2006), selon lequel "We should not characterize probabilistic modeling with the simplified and incorrectly applied terms of "inductive" and "deductive". Similarly, I have replaced the term "predictive" with "probabilistic" to reflect a greater degree of indeterminacy" bien que la démarche employée ici soit résolument possibiliste et non probabiliste.

2. LES LOGIQUES D'OCCUPATION DU SOL : TEST DES PARAMETRES A INTEGRER

2.1 Préparation des données

2.1.1 Les données relatives aux sites

Si les analyses du chapitre précédent s'effectuaient à l'échelle de la maille, ce découpage n'est désormais plus nécessaire, puisqu'il s'agit de tester la localisation des sites en eux-mêmes en fonction de critères divers eux-mêmes situés dans l'espace. Chaque site s'est donc vu attribuer la configuration spatiale à laquelle il participe, à partir des LISA calculés sur le peuplement des *temps* 1Ba, 1Bb et 1A+1Ba. Pour chacun de ces *temps*, chaque site est ainsi considéré comme relevant d'un *hot-spot*, ou comme un *outlier*.

2.1.2 Les données relatives aux critères à tester

Les données dont nous disposons afin d'évaluer les logiques de localisation des sites (et disponibles dans le SIG) sont relatives aux critères environnementaux de la zone, et correspondent aux critères « classiques » pris en compte dans les modèles d'archéologie « prédictive » (Mehrer, Wescott, 2006). Chacune d'entre elles présente un certain nombre d'avantages et d'inconvénients qui ont été pesés afin de sélectionner le ou les paramètres à tester.

- L'altitude : s'il s'agit d'un critère pouvant s'avérer dans la plupart des cas très efficace, celui-ci ne pouvait, à l'échelle de la zone étudiée, être un paramètre très discriminant et explicatif de la variabilité de la localisation des sites dans la mesure où la majorité (67%) de la zone, et notamment les zones non prospectées, est située entre 300 et 500 mètres d'altitude, et qu'aucune variabilité dans le nombre de sites n'est constatée dans cet intervalle sur ce critère. Evaluer les possibilités de peuplement sur ce critère serait donc en quelque sorte une « solution de facilité » revenant à estimer celui-ci comme fortement possible sur presque toute la zone étudiée.
- La pluviométrie : bien que nous disposions d'un modèle des pluies élaboré par des climatologues (Traboulsi, 2004) pour la zone étudiée, ce modèle se base sur des données actuelles. Nous ne disposons d'aucun modèle paléoclimatique nous permettant d'envisager ce critère dans le passé, et nous avons jugé trop dangereux pour le moment de tester la méthode envisagée sur un critère aussi hasardeux.

- Les sols : une carte des sols précise est disponible dans le SIG pour la zone étudiée, et nous pouvons considérer, érosion mise à part, que ce critère a certainement moins de variabilité entre l'actuel et le passé que la pluviométrie, par exemple. Cependant, la diversité des types de sols et l'évaluation de leur importance dans les localisations de peuplement suppose l'élaboration de typologies de sols (qui pourraient par ailleurs être élaborées à l'aide de la logique floue) que nous n'étions pas, pour le moment, en mesure de fournir, la carte d'« attractivité » disponible (Figure 13) ne couvrant pas la totalité de la zone étudiée.
- Les cours d'eau : ce critère présente l'avantage d'être localisé avec précision et fiabilité dans l'espace. En outre, s'il est établi à partir d'observations actuelles, celui-ci présente certainement moins de variabilité entre l'actuel et le passé que la pluviométrie, par exemple. C'est donc le critère que nous avons choisi de tester afin d'identifier sa corrélation avec les sites existants, et donc d'évaluer son importance dans la localisation des sites archéologiques.

2.2 Test de la corrélation entre la distance au cours d'eau et le nombre de sites par des méthodes statistiques

Le paramètre à tester est la distance des sites aux cours d'eau ; en d'autres termes, on cherche à savoir si les sites vont préférentiellement se localiser à proximité d'un cours d'eau ou si ce critère n'a aucun impact sur la localisation des sites.

A l'aide du SIG, les cours d'eau ont été transformés en buffers de 3, 6, 9, 12, 15, 18 et 21 km, et le nombre de sites contenu dans chaque buffer a été relevé (Figure 110). Le pas choisi pour l'observation des phénomènes est de 3 km : celui-ci permettait de conserver une certaine précision sans avoir de trop nombreux niveaux à manier dans l'analyse. En outre, après test sur des buffers de 1 km, ces derniers introduisaient une précision inutile dans la mesure où les sites situés entre 1 et 3 km des cours d'eau étaient répartis de façon très équilibrée entre ces trois intervalles (0-1 km, 1-2 km, 2-3 km), mais aussi une lourdeur considérable dans l'analyse.

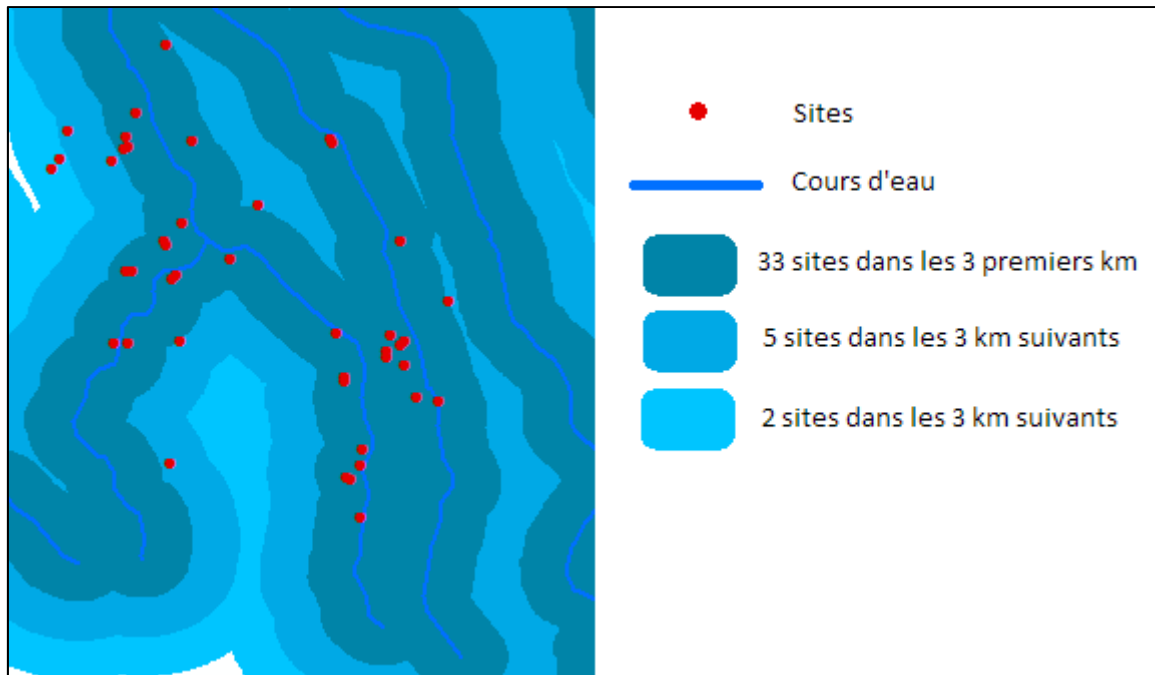


Figure 110 - Comptage du nombre de sites par zone de 3 km autour des cours d'eau.

Les résultats de ce comptage sont présentés sous forme de pourcentages sur les Figure 111, Figure 112 et Figure 113.

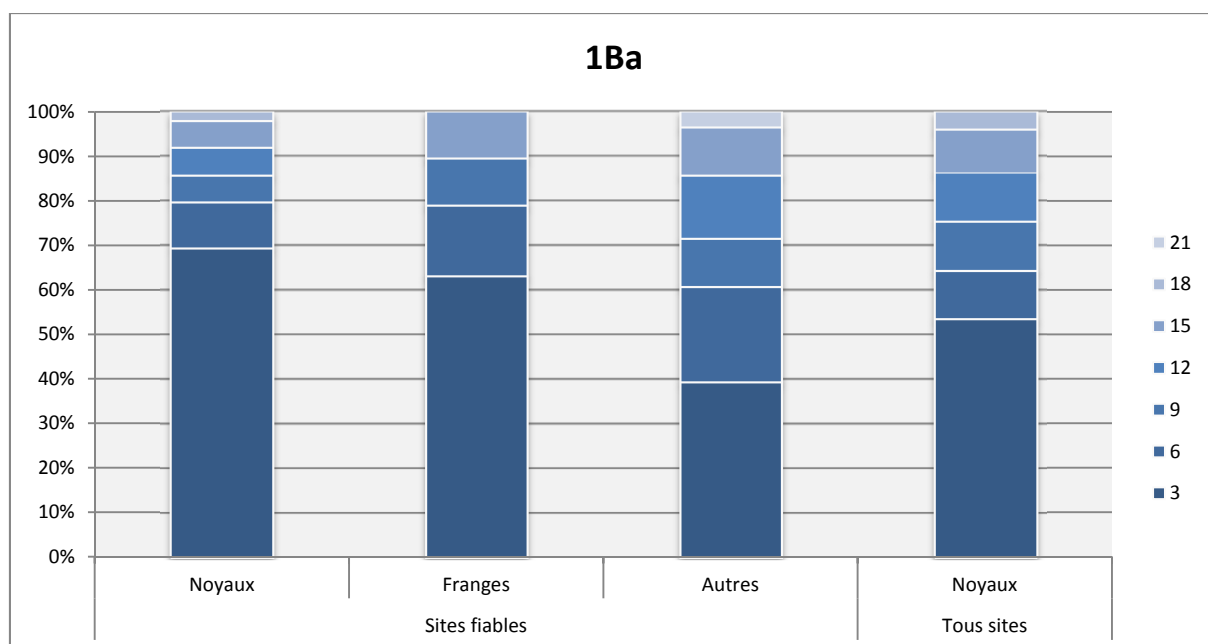


Figure 111 - Pourcentage du nombre de sites du temps 1Ba situés à des distances successives de 3 km des cours d'eau, pour chaque type d'occupation et niveau de fiabilité.

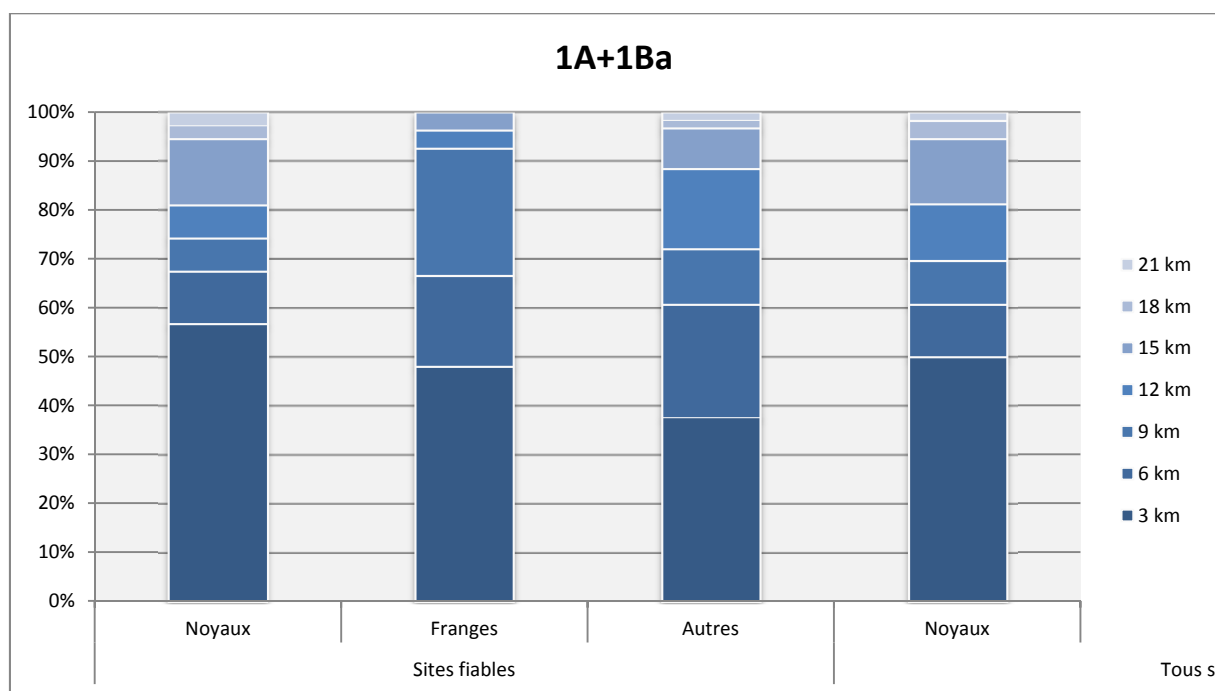


Figure 112 - Pourcentage du nombre de sites du temps 1A+1Ba situés à des distances successives de 3 km des cours d'eau, pour chaque type d'occupation et niveau de fiabilité.

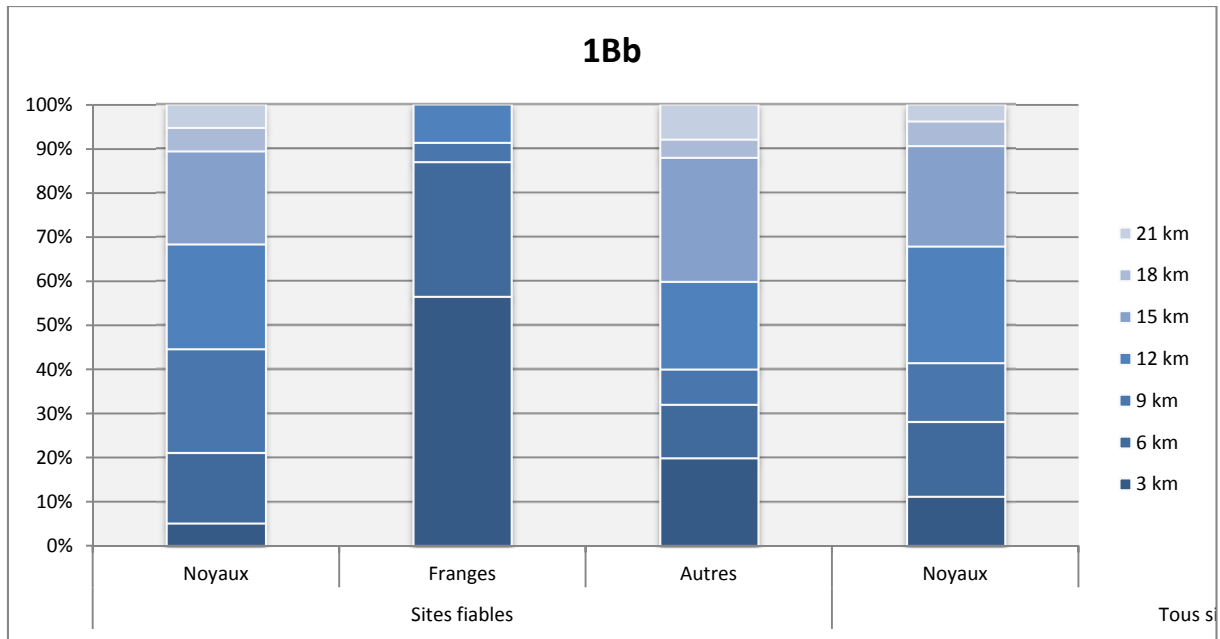


Figure 113 - Pourcentage du nombre de sites du *temps* 1Bb situés à des distances successives de 3 km des cours d'eau, pour chaque type d'occupation et niveau de fiabilité.

On peut constater que la majorité des sites fiables sont, pour le *temps* 1Ba, situés entre 0 et 6 km des cours d'eau, quelle que soit la configuration spatiale dont ils relèvent : 80% pour les *hot-spots* et les *outliers*. L'intégration des sites non fiables abaisse quelque peu cette proportion pour les *hot-spots* (65%) mais ne change pas pour les *outliers*. En outre, le plus grand pourcentage de sites est invariablement observé entre 0 et 3 km des cours d'eau.

L'observation des sites au niveau temporel 1A+1Ba diminue quelque peu les pourcentages, mais conserve les rapports : le plus grand nombre de sites est toujours localisé entre 0 et 3 km des cours d'eau.

Le *temps* 1Bb, quant à lui, nous réservait une surprise : seuls 5% des sites fiables (10% en introduisant les non fiables) situés dans un *hot-spot* sont localisés entre 0 et 3 km des cours d'eau, au profit de l'intervalle situé entre 3 et 15 km. Quant aux *outliers*, ceux-ci atteignent en revanche des scores plus élevés que pour les temps précédents : 88% des sites fiables sont situés entre 0 et 6 km. Afin d'observer ce phénomène de plus près, nous avons réalisé un principe de comptage similaire à celui-ci mais à partir des sites qui apparaissent et qui se maintiennent entre les *temps* 1Ba et 1Bb, et entre 1A+1Ba et 1Bb (Figure 114 à Figure 117).

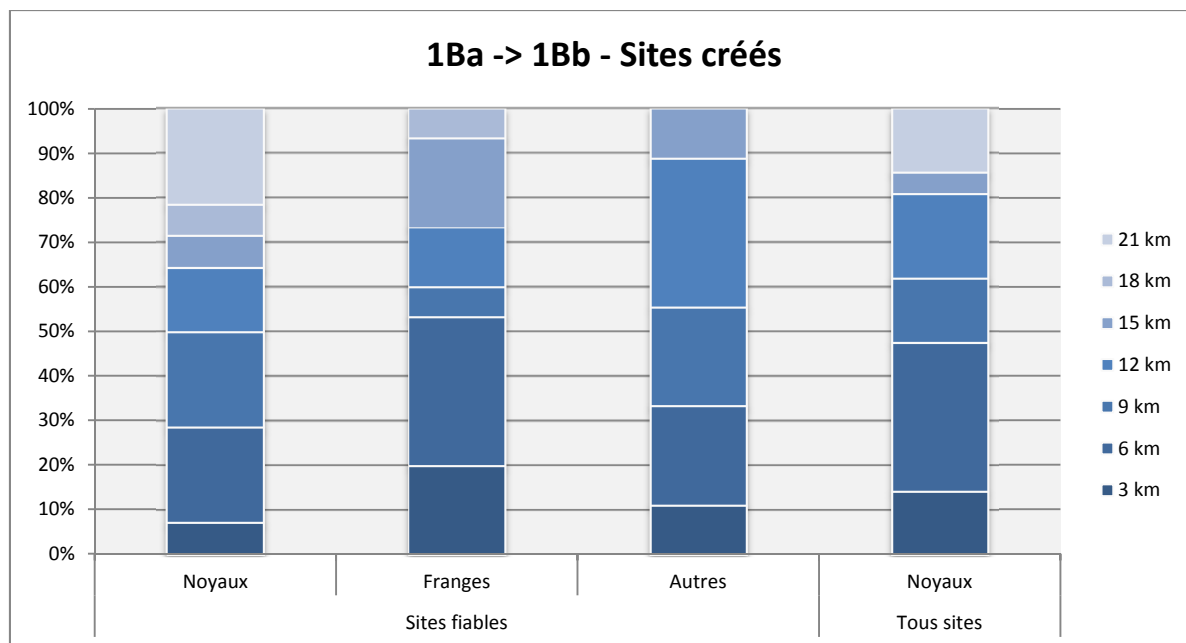


Figure 114 - Pourcentage du nombre de sites apparaissant entre les *temps* 1Ba et 1Bb situés à des distances successives de 3 km des cours d'eau, pour chaque type d'occupation du sol et niveau de fiabilité.

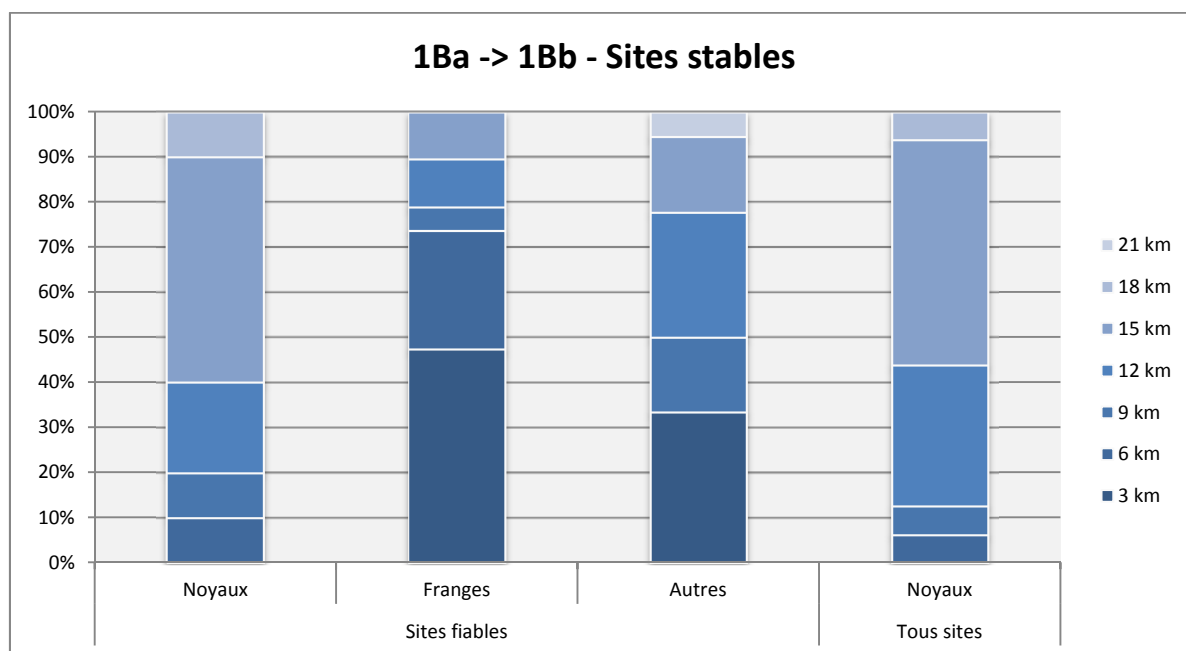


Figure 115 - Pourcentage du nombre de sites demeurant stables entre les *temps* 1Ba et 1Bb situés à des distances successives de 3 km des cours d'eau, pour chaque type d'occupation du sol et niveau de fiabilité.

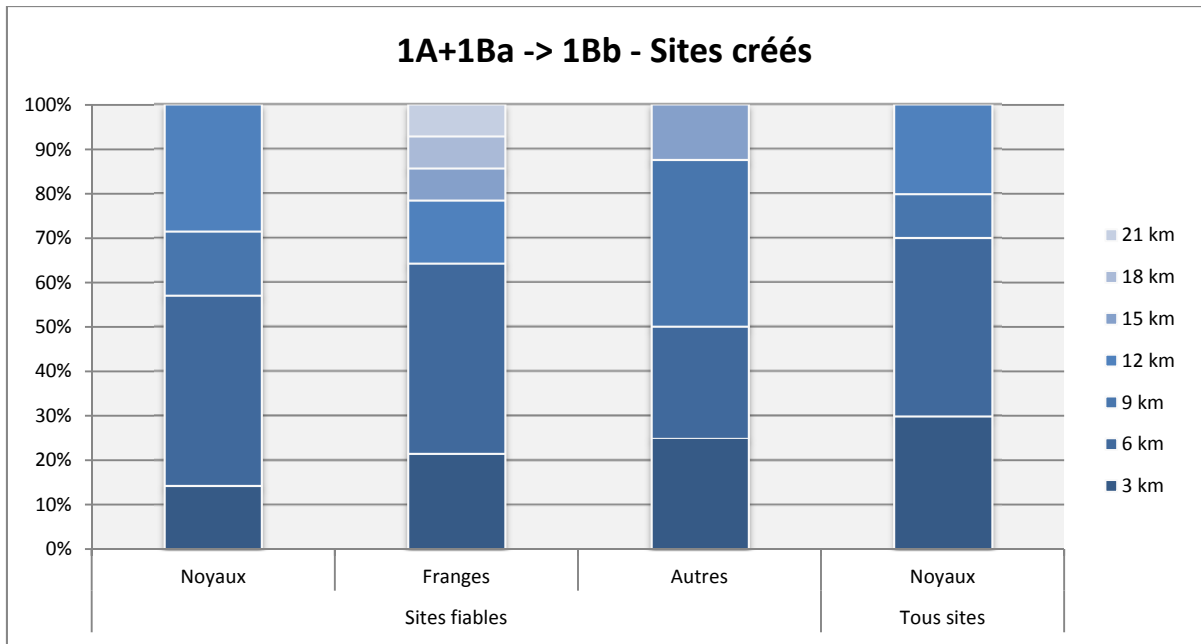


Figure 116 - Pourcentage du nombre de sites apparaissant entre les *temps* 1A+1Ba et 1Bb situés à des distances successives de 3 km des cours d'eau, pour chaque type d'occupation du sol et niveau de fiabilité.

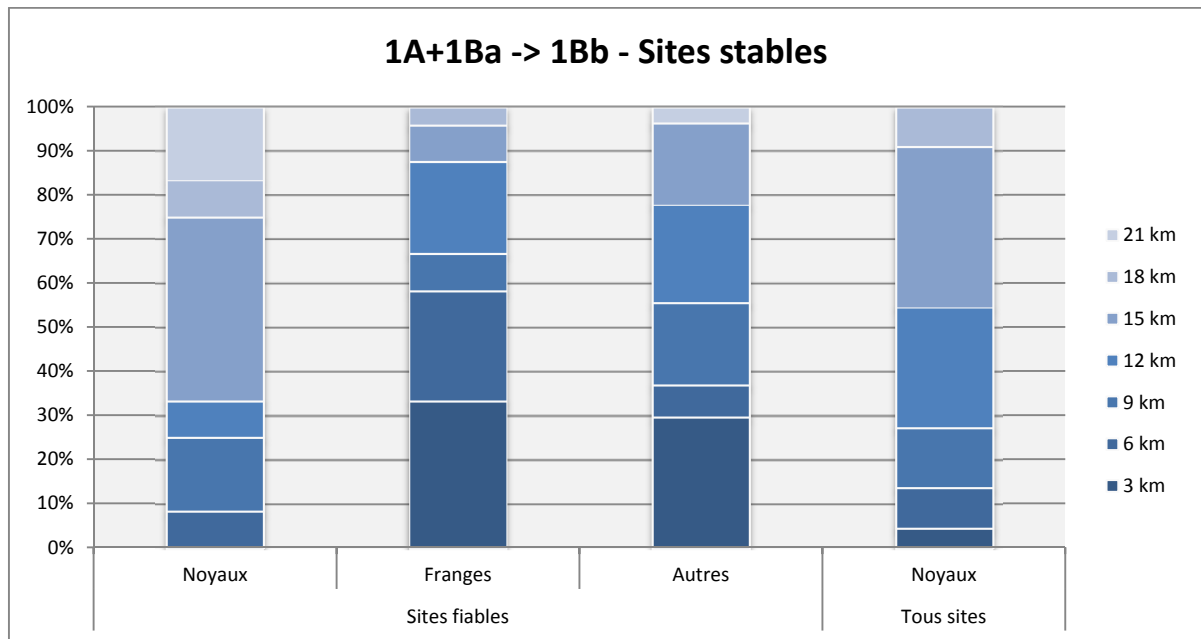


Figure 117 - Pourcentage du nombre de sites demeurant stables entre les *temps* 1A+1Ba et 1Bb situés à des distances successives de 3 km des cours d'eau, pour chaque type d'occupation du sol et niveau de fiabilité.

On constate que les sites qui apparaissent entre les *temps* 1Ba et 1Bb ne se localisent pas préférentiellement à proximité d'un cours d'eau. Il semble même que les sites qui perdurent entre les *temps* 1Ba et 1Bb ne soient pas les sites localisés au plus près d'un cours d'eau.

Le changement radical de tendance de localisation entre les *temps* 1Ba et 1Bb montre bien tout l'intérêt d'une distinction temporelle de ces logiques : il se pourrait en effet qu'un changement dans les besoins ou les priorités des populations ait également changé leur mode d'occupation du sol.

Les informations ci-dessus ont été à nouveau testées par le calcul du coefficient de corrélation entre la distance aux cours d'eau (3, 6, 9, 12...) et le nombre de sites associé à chaque intervalle (Tableau 24).

Tableau 24 - Coefficients de corrélation entre le nombre de sites correspondant à chaque type d'occupation et la distance aux cours d'eau.

1Ba - Coefficients de corrélation			
Sites fiables		Tous sites	
<i>Hot-spots</i>	<i>Outliers</i>	<i>Hot-spots</i>	<i>Outliers</i>
-0,72	-0,77	-0,74	-0,94
1Bb - Coefficients de corrélation			
Sites fiables		Tous sites	
<i>Hot-spots</i>	<i>Outliers</i>	<i>Hot-spots</i>	<i>Outliers</i>
-0,21	-0,92	-0,33	-0,98
1A+1Ba - Coefficients de corrélation			
Sites fiables		Tous sites	
<i>Hot-spots</i>	<i>Outliers</i>	<i>Hot-spots</i>	<i>Outliers</i>
-0,69	-0,89	-0,73	-0,98

Tous les coefficients de corrélation calculés sont négatifs : en effet, dans tous les cas, la distance aux cours d'eau et le nombre de sites associés évoluent en sens inverse. En d'autres termes, plus on s'éloigne d'un cours d'eau, moins on compte de sites. En outre, si toutes les configurations spatiales, présentent des coefficients de corrélation supérieurs à 70% quel que soit le niveau de fiabilité ou le niveau temporel considéré, les *hot-spots* associés au *temps* 1Bb présentent une corrélation très faible aux cours d'eau, alors que les *outliers* présentent des taux de corrélation très élevés.

Dans l'attente de réaliser des analyses destinées à comprendre ce changement d'occupation du sol et les différenciations internes au *temps* 1Bb entre *hot-spots* et *outliers*, nous avons été contrainte

d'écarter ce temps de l'analyse, et de n'envisager les possibilités de localisation des sites en fonction du paramètre « distance au cours d'eau » que pour les *temps* 1Ba et 1A+1Ba.

2.3 Test par des méthodes morphologiques

La distance au cours d'eau est ici envisagée sous une forme linéaire, par l'intermédiaire des buffers représentant le dilaté du réseau hydrographique à des pas successifs de 3 km. Nous souhaitons à présent tester la corrélation de ces paramètres du point de vue de leur forme.

En effet, la localisation préférentielle des sites à un certain niveau des cours d'eau (en amont ou en aval par exemple) pourrait être observée, et gommée par les buffers qui ne font que suivre leur linéaire.

Nous avons donc à nouveau mobilisé la morphologie mathématique et l'algorithme du graphe perceptuel (cf. Figure 48) afin de comparer le squelette des sites (fiabes, puis en introduisant les sites non fiabes) des *temps* 1Ba et 1A+1Ba au linéaire du réseau hydrographique (Figure 118 à Figure 121). Deux pas de squelettisation ont été testés et sont présentés : 3 km et 4 km (en d'autres termes, les sites ont été successivement dilatés par un élément structurant circulaire de 3 km et de 4 km de rayon).

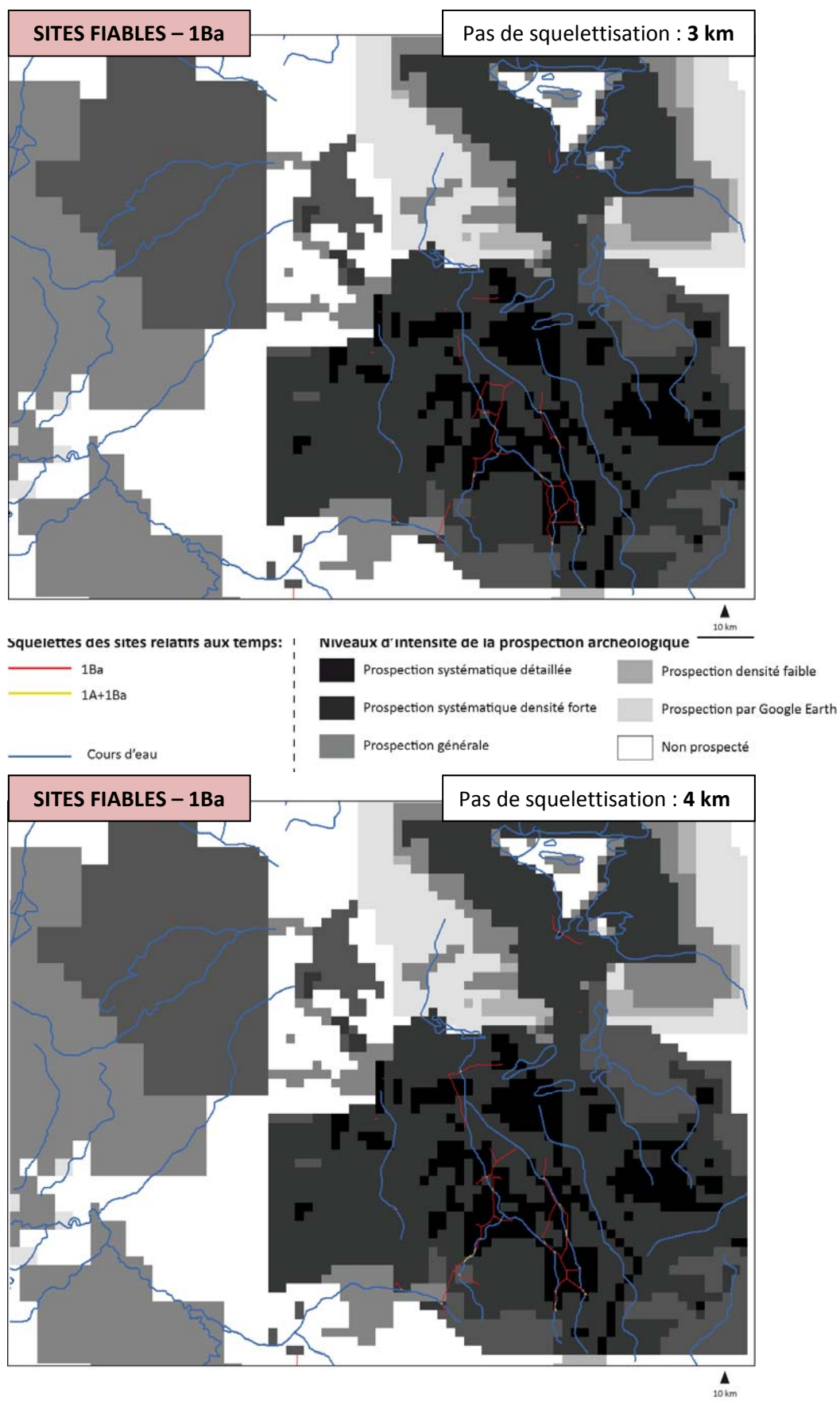


Figure 118 - Squelettes des sites fiables aux pas 3 km et 4 km, pour le *temps* 1Ba

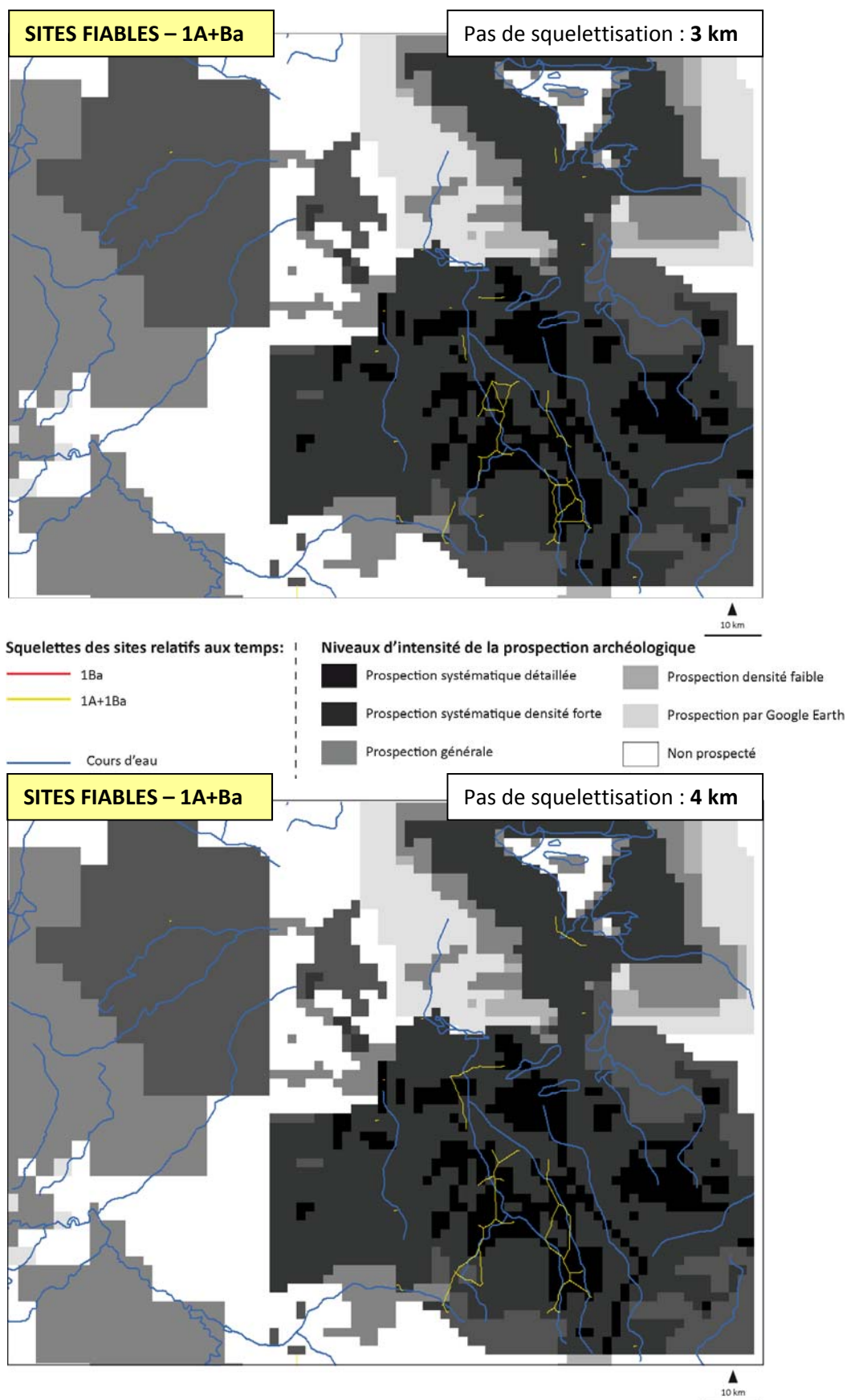


Figure 119 - Squelettes des sites fiables aux pas 3 km et 4 km, pour le temps 1A+1Ba

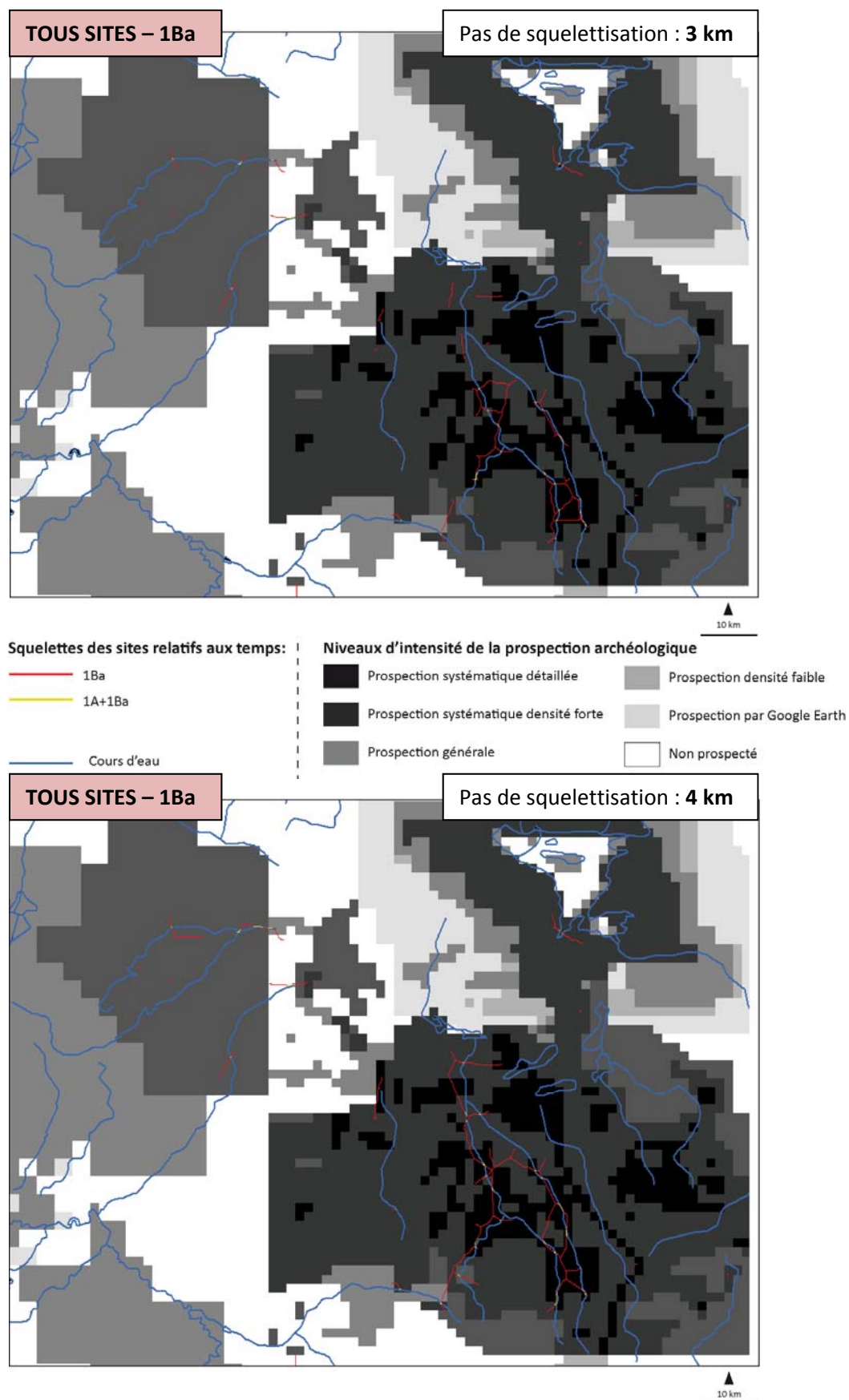


Figure 120 - Squelettes de la totalité des sites aux pas 3 km et 4 km, pour le *temps* 1Ba

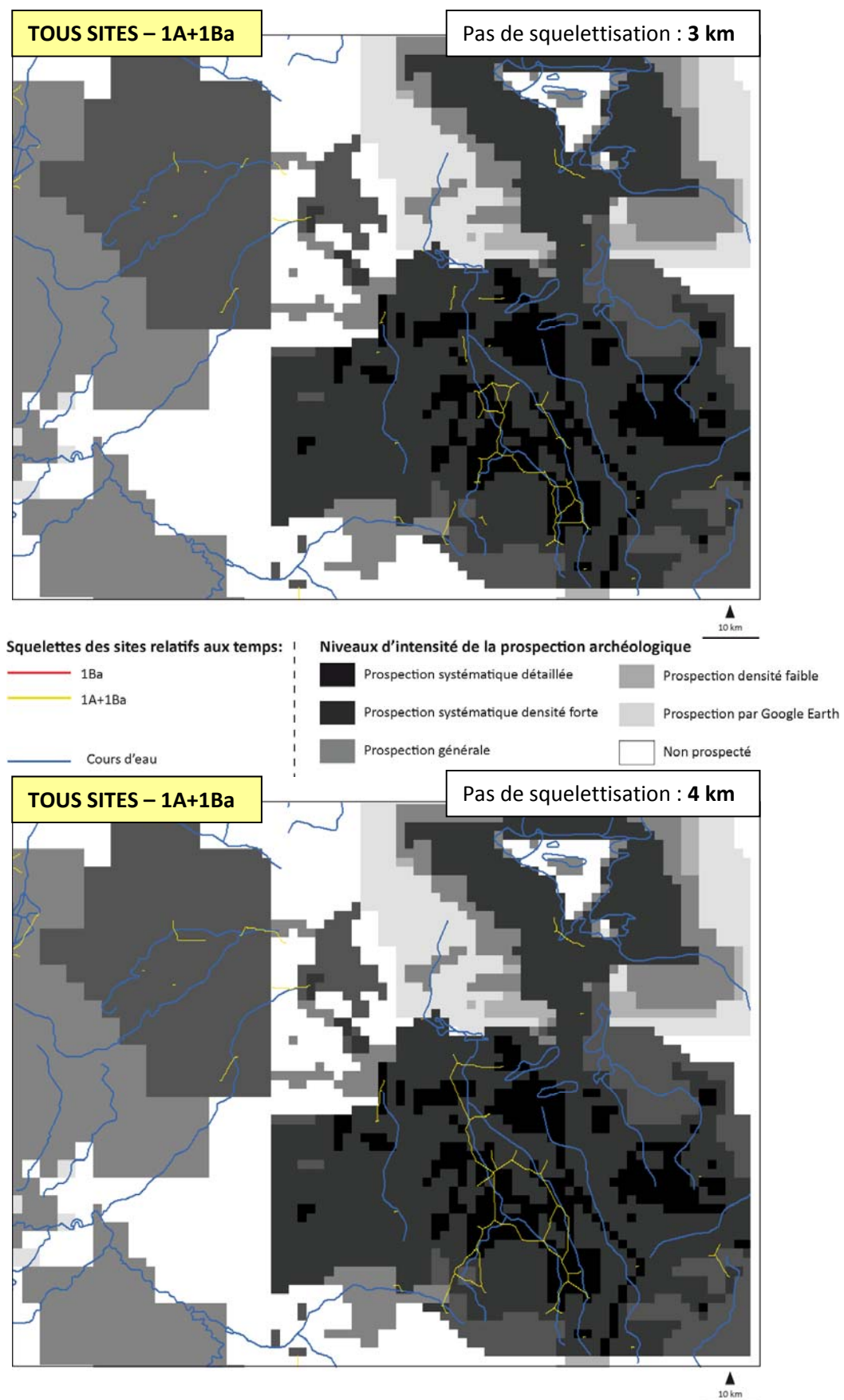


Figure 121 - Squelettes de la totalité des sites aux pas 3 km et 4 km, pour le *temps* 1A+Ba

On constate que quel que soit le pas, le niveau de fiabilité ou le temps considéré, le squelette des sites dans la zone la plus peuplée suit parfaitement le linéaire du cours d'eau. De manière générale, les sites sont répartis indifféremment sur toute sa longueur (dans les zones les mieux prospectées).

3. PASSAGE DE L'OBSERVATION A L'ESTIMATION DE L'OCCUPATION DU SOL DANS UNE LOGIQUE POSSIBILISTE

3.1 Définition du cadre des sous-ensembles flous de type 1

L'objectif final étant d'obtenir des cartes représentant les possibilités de localisation des sites en fonction de la présence ou non d'un cours d'eau, la méthode adoptée consiste à présent à calibrer les sous-ensembles flous correspondant à un nombre de sites « fort », « moyen » ou « faible » (c'est-à-dire, à terme, des possibilités « forte », « moyenne » ou « faible » de localisation de sites) à partir du nombre de sites relevés précédemment pour chaque distance au cours d'eau.

Cependant, dans un but de reproductibilité de la méthode, et de comparabilité des résultats, la définition des sous-ensembles flous « fort », « moyen » et « faible » doit s'effectuer de manière identique pour chaque temps, niveau de fiabilité, et configuration spatiale envisagée.

La méthode élaborée est la suivante : les bornes de chaque sous-ensemble flou, seront définies par le pourcentage que représente le nombre de sites à chaque distance du cours d'eau par rapport à la moyenne des sites pour chaque temps, niveau de fiabilité et type de configuration spatiale.

Le cadre général auquel sera appliqué chaque cas est présenté sur la Figure 122 : Les couleurs rouge, bleue et verte correspondent aux valeurs faibles, moyennes et fortes. Les deux zones colorées en orange et en jaune correspondent aux zones d'incertitude où on ne peut avec certitude définir si le nombre est « faible » ou « moyen », ou « élevé » ou « moyen ».

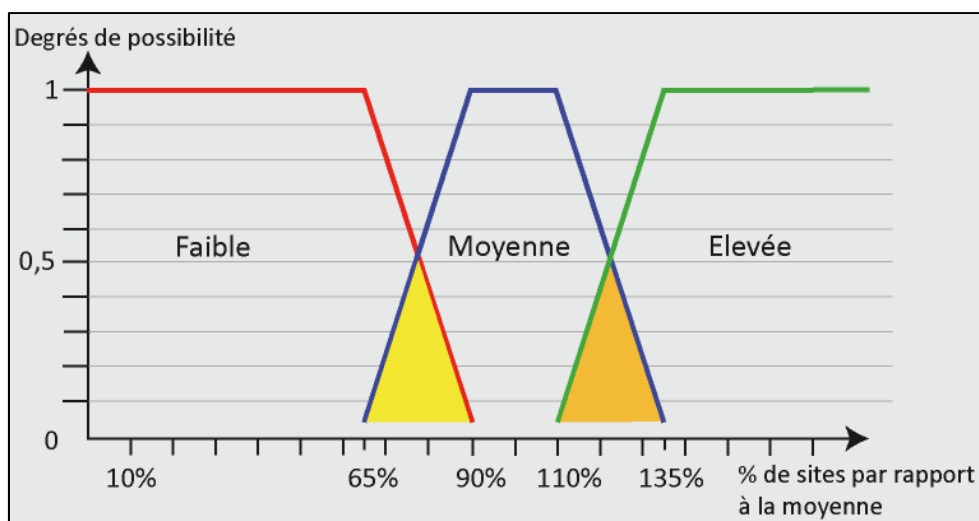


Figure 122 - Définition du cadre des sous-ensembles flous permettant d'évaluer si la quantité de sites est "faible", "moyenne" ou "élevée" par rapport à la moyenne.

Ainsi, afin d'estimer si le nombre de sites à chaque pas de dilatation du cours d'eau est plutôt « fort », « moyen » ou « faible », on calcule le pourcentage que représente chaque nombre de sites présents dans chaque dilaté par rapport à la moyenne.

Si le nombre de sites est noté N , on calcule donc $N * 100 / moyenne$.

Le Tableau 25 montre en exemple cette étape des calculs réalisés pour les sites fiables du temps 1Ba.

Tableau 25 - Calcul du pourcentage de sites par rapport à la moyenne : l'exemple des sites fiables du temps 1Ba

1Ba – sites fiables		
Distance aux cours d'eau	Nombre de sites <i>hot-spots</i>	Nombre de sites <i>outliers</i>
3 km	34	12
6 km	5	3
9 km	3	2
12 km	3	0
15 km	3	2
18 km	1	0
Moyenne	8	4



1Ba – sites fiables		
Distance aux cours d'eau	Pourcentage de sites <i>hot-spots</i> par rapport à la moyenne	Pourcentage de sites <i>outliers</i> par rapport à la moyenne
3 km	416%	316%
6 km	61%	79%
9 km	37%	53%
12 km	37%	0
15 km	37%	53%
18 km	12%	0

A partir des nombres de sites à chaque distance des cours d'eau pour chaque temps et niveau de fiabilité, on définit les sous-ensembles flous correspondant au « fort », « moyen » ou « faible » nombre de sites par rapport à la moyenne.

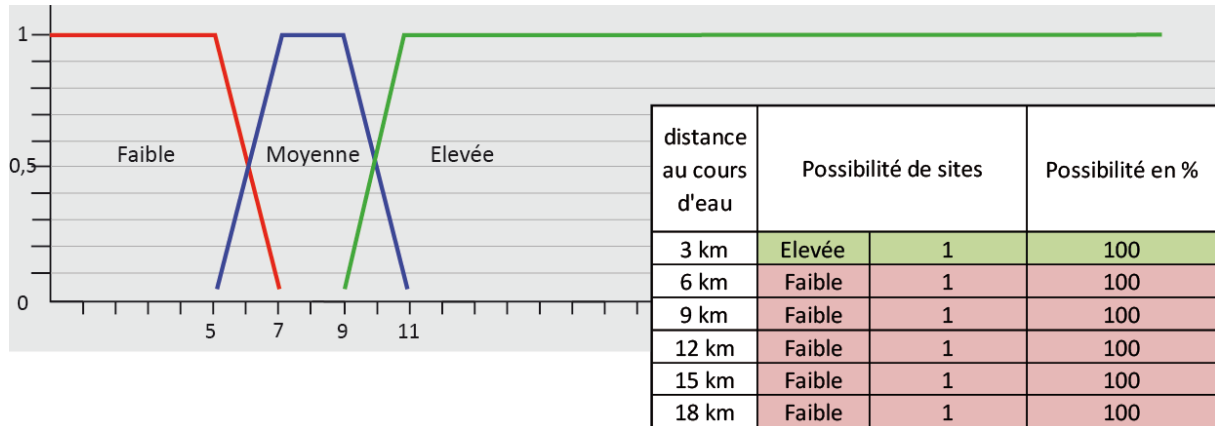
Cette procédure est décrite à titre d'exemple dans le Tableau 26 pour les sites fiables du temps 1Ba.

Tableau 26 - Calcul des sous-ensembles flous à partir des sites fiables du temps 1Ba pour chaque type d'occupation.

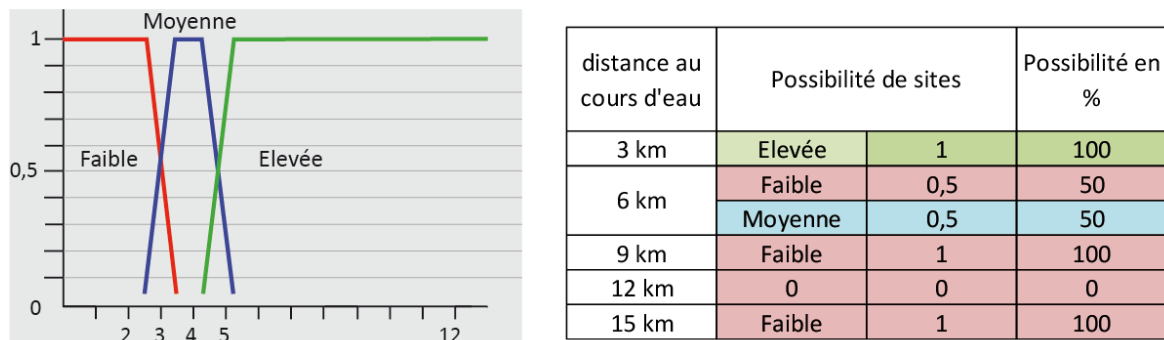
Limites des sous-ensembles flous	<i>Hot-spots</i>	<i>Outliers</i>
65%	5 (moyenne*65/100, soit 8*65/100)	2,6
90%	7	3,6
100% (moyenne)	8	4
110%	9	4,4
135%	11	5,4

On peut ainsi, à partir de ce tableau, tracer les sous-ensembles flous qui permettront d'évaluer les possibilités pour que le nombre de sites présent à chaque distance des cours d'eau et inhérent à chaque temps, niveau de fiabilité et type de configuration spatiale soit « fort », « moyen », ou « faible » (Figure 123).

1Ba , sites fiables - Hot-spots - Sous-ensembles flous de type 1



1Ba , sites fiables - Outliers - Sous-ensembles flous de type 1



1Ba – sites fiables		
Distance aux cours d'eau	Nombre de sites hot-spots	Nombre de sites outliers
3 km	34	12
6 km	5	3
9 km	3	2
12 km	3	0
15 km	3	2
18 km	1	0
21 km	0	0
Distance aux cours d'eau	Possibilité d'occupation du sol de type hot-spot	Possibilité d'occupation du sol de type outliers
3 km	Elevée	Elevée
6 km	Faible	Faible (0,4) Moyenne (0,6)
9 km	Faible	Faible
12 km	Faible	Faible
15 km	Faible	Faible
18 km	Faible	Faible
21 km	Faible	Faible

Figure 123 - Sous-ensembles flous calculés à partir des sites fiables du temps 1Ba pour chaque type d'occupation.

3.2 Définition du cadre des sous-ensembles flous de type 2

Le *flou* mobilisé jusqu'à présent est défini comme du « flou de type 1 » : les fonctions d'appartenance des sous-ensembles flous sont en effet des valeurs précises. On peut noter sur la Figure 123, pour les sites *hot-spot* par exemple, qu'une coupe à 6 sites correspond à un niveau de possibilité faible à 0,5 et moyenne à 0,5 également : ces deux valeurs sont bien des valeurs précises.

On pourrait objecter à ce constat qu'il est paradoxal qu'une méthode dite *floue* déduise une valeur d'appartenance parfaitement précise, comme c'est le cas dans le *flou de type 1* (Mendel, 2004 ; Mendel, John, 2002 ; Fisher, Arnot, 2006). L'introduction du *flou de type 2* (Zadeh, 1978) permet de s'affranchir de cet écueil. En effet, cette logique permet de considérer à la fois le *flou* sur les valeurs de références et le flou des valeurs d'appartenance, qui ne sont désormais plus définies par des valeurs précises mais par des intervalles (Figure 124). Dans le *flou de type 2*, on parle alors d'empreinte de l'incertitude (« *footprint of uncertainty* »).

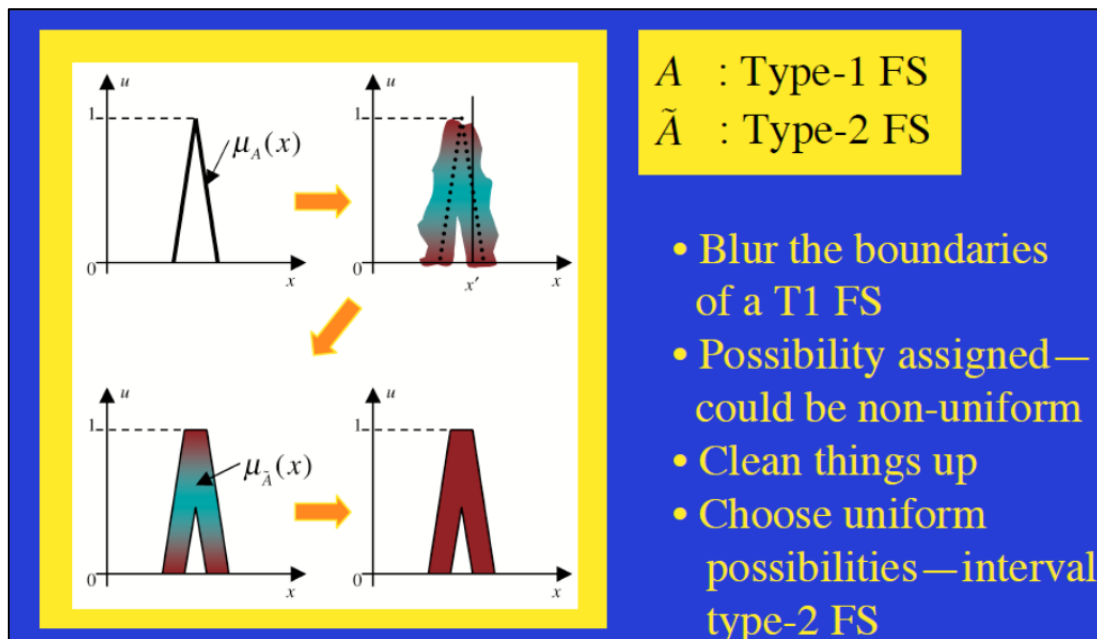


Figure 124 - Le passage du *flou de type 1* au *flou de type 2* (Mendel, 2004).

Ainsi, « Type-2 fuzzy sets can be seen as a response to the philosophical issue of higher order vagueness. It is shown that the type 2 analysis yields a variety of answers to any query about the amount of change, reflecting the higher order vagueness in the uncertainty if the change in a vague phenomenon » (Fisher, Arnot, 2006). Nous postulons que le cadre du *flou de type 2* est particulièrement approprié aux questionnements archéologiques.

Dans notre cas, cela suppose de transformer les valeurs uniques des bornes des sous-ensembles définis précédemment en intervalles. Plutôt que de définir ces intervalles de façon arbitraire, ce qui serait contraire à la démarche empirique adoptée jusqu'à présent et ne serait porteur que d'une complexification inutile et alourdissant la démarche sans véritable apport dans la compréhension et l'analyse, nous avons choisi de définir les intervalles en nous basant sur la différenciation établie dans la fiabilité des sites. La Figure 125 explique la procédure suivie pour l'élaboration des *sous-ensembles flous de type 2*.

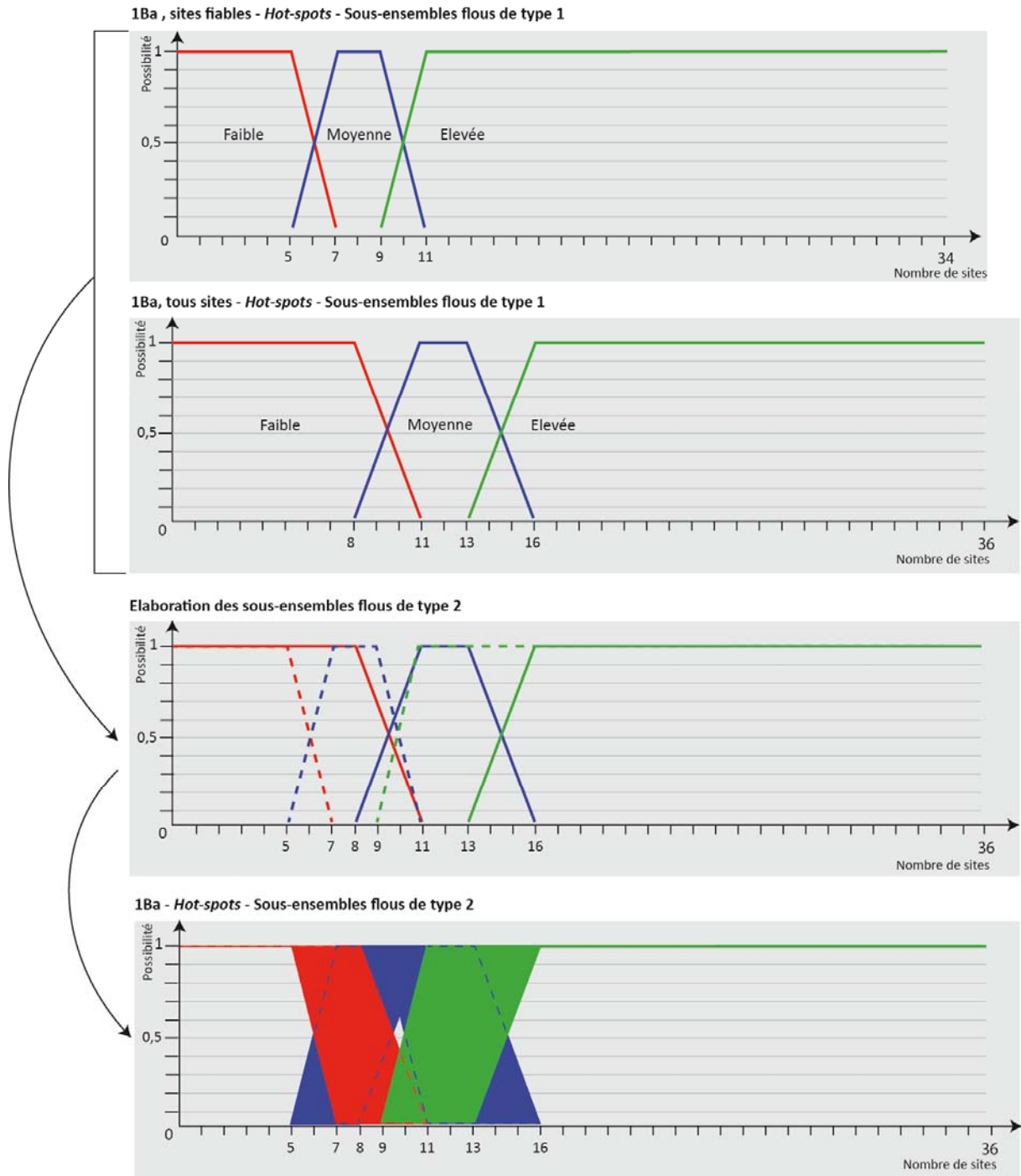


Figure 125 - Passage des sous-ensembles flous de type 1 aux sous-ensembles flous de type 2.

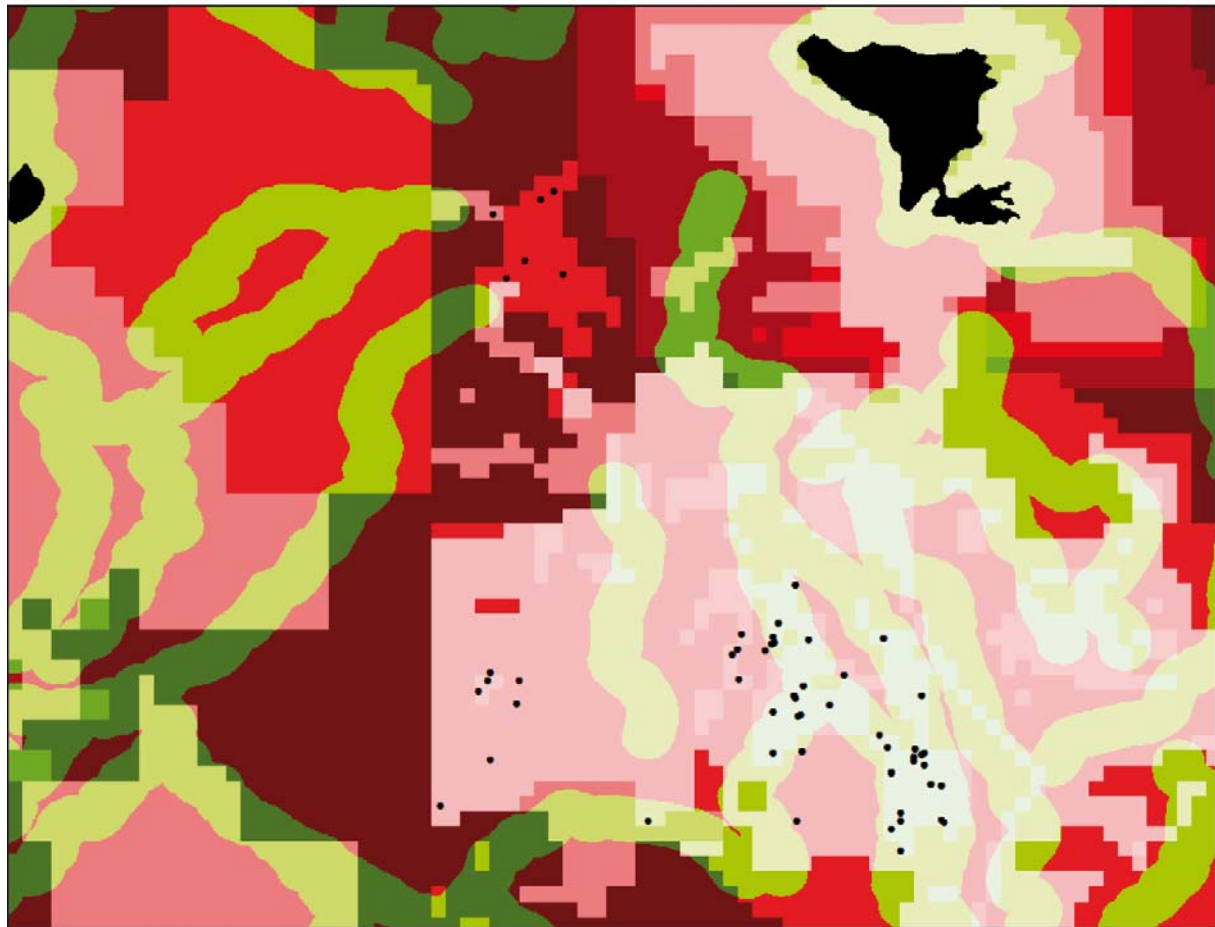
Le *sous-ensemble flou de type 2* n'est donc plus une ligne mais une surface : l'incertitude présente une empreinte dont les limites sont définies par la prise en compte des limites des *sous-ensembles flous de type 1* pour les sites fiables (limite inférieure du *sous-ensemble flou de type 2*) des limites des *sous-ensembles flous de type 1* pour la totalité des sites (limite supérieure du *sous-ensemble flou de type 2*).

Les *sous-ensembles flous de type 2* ainsi définis présentent l'avantage de synthétiser dans une même évaluation des possibilités de localisation des sites l'information relative à la fiabilité des sites.

3.1 Cartographie des résultats

Les cartes produites à partir des sous-ensembles flous définis sur la Figure 125 sont présentées sur les Figure 126 à Figure 128 pour les sites *hot-spots* du temps 1Ba, et sur les Figure 129 à Figure 131 pour les sites *outliers* du temps 1Ba. Les cartes relatives au temps 1A+1Ba sont présentées dans l'Annexe 2.

1Ba , sites fiables - Hot-spots - Sous-ensembles flous de type 1



● Sites fiables

▲ 10 km

distance au cours d'eau	Possibilité de sites		Possibilité en %
3 km	Elevée	1	100
6 km	Faible	1	100
9 km	Faible	1	100
12 km	Faible	1	100
15 km	Faible	1	100
18 km	Faible	1	100

Intensité de prospection de la zone

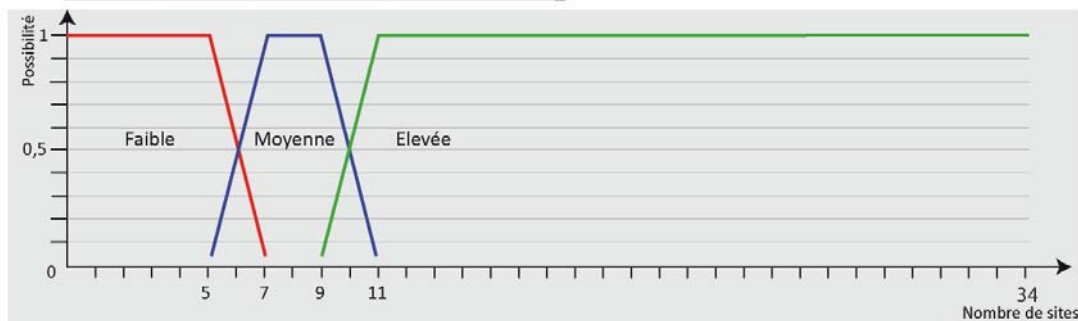
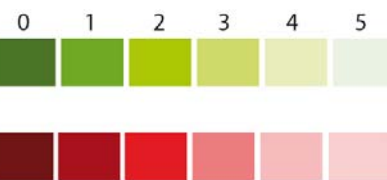


Figure 126 - Possibilités de localisation des sites évaluées à partir des *sous-ensembles flous de type 1* des sites *hot-spot* fiables du temps 1Ba.

1Ba, tous sites - Hotspots - Sous-ensembles flous de type 1

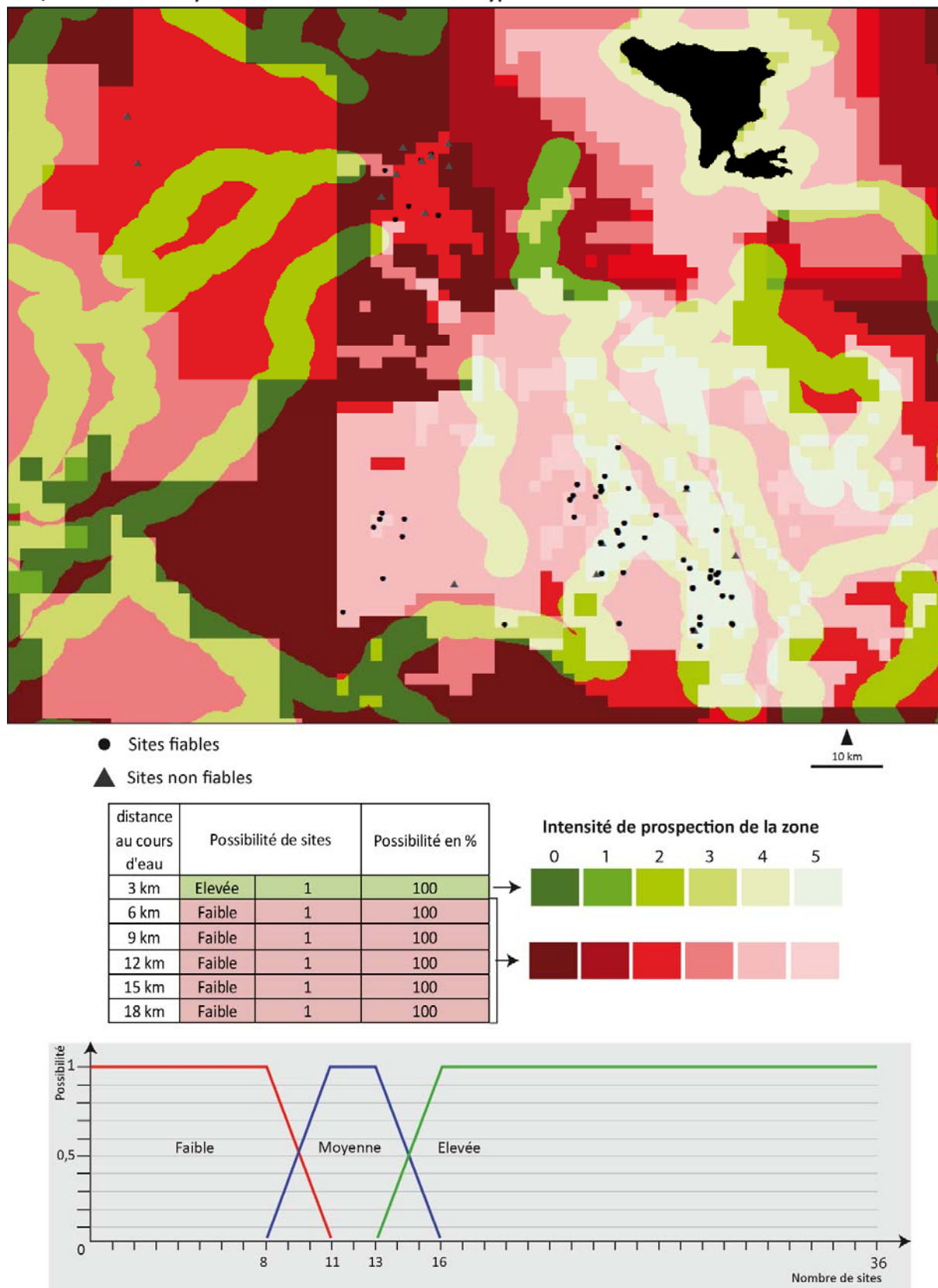
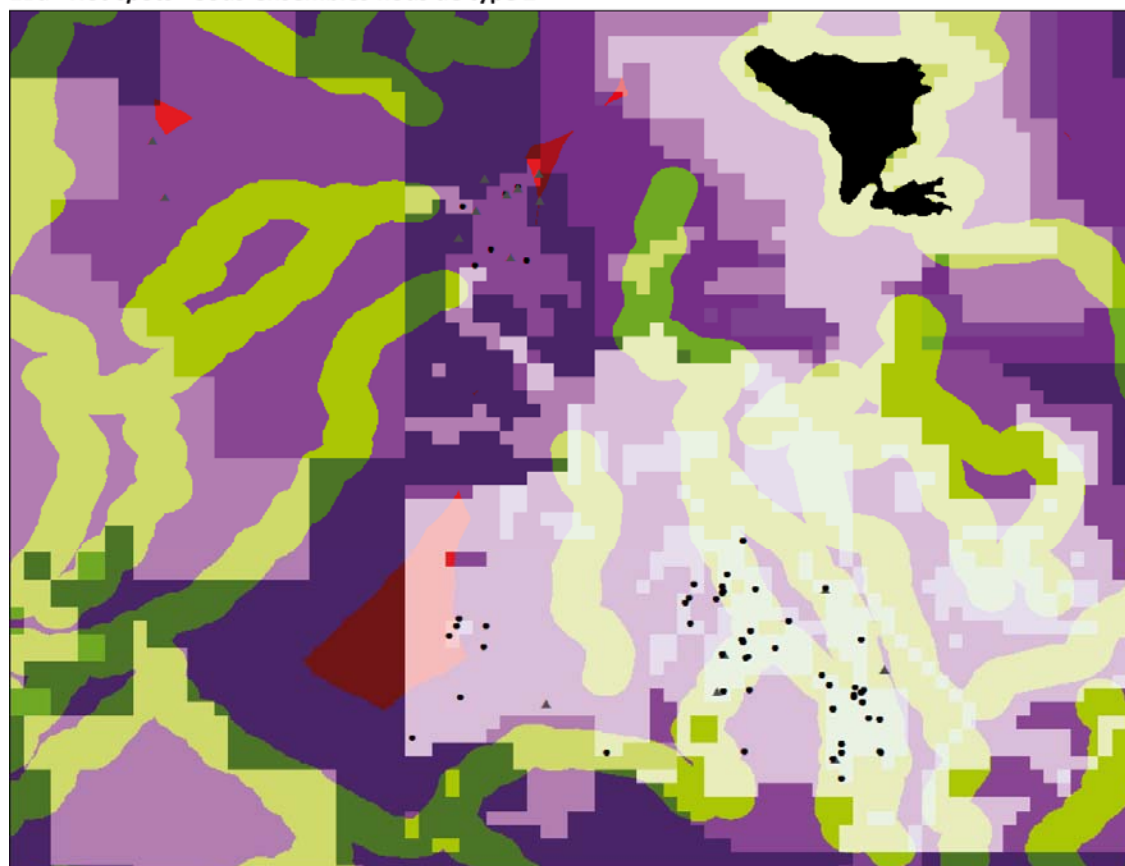


Figure 127 - Possibilités de localisation des sites évaluées à partir des *sous-ensembles flous de type 1* de la totalité des sites *hot-spot* du temps 1Ba.

1Ba - Hot-spots - Sous-ensembles flous de type 2



- Sites fiables
- ▲ Sites non fiables

10 km

distance au cours d'eau	Possibilité de sites		Intensité de prospection de la zone					
			0	1	2	3	4	5
3 km	Elevée	1						
	Faible	1						
6 km	Moyenne	1						
	Faible	1						
9 km	Moyenne	1						
	Faible	1						
12 km	Faible	1						
	Moyenne	1						
15 km	Faible	1						
	Moyenne	1						
18 km	Faible	1						
	Faible	1						

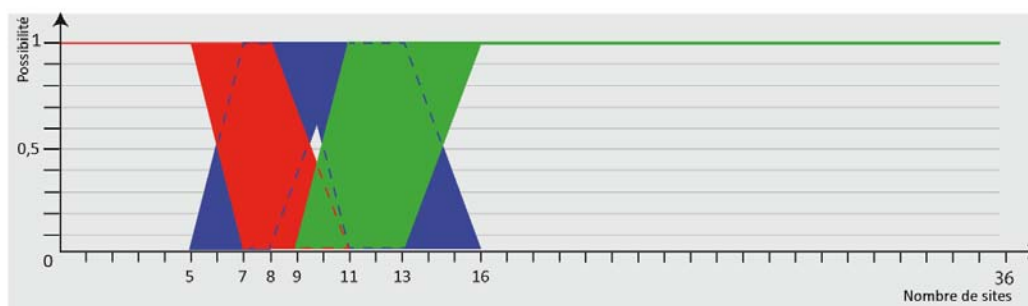
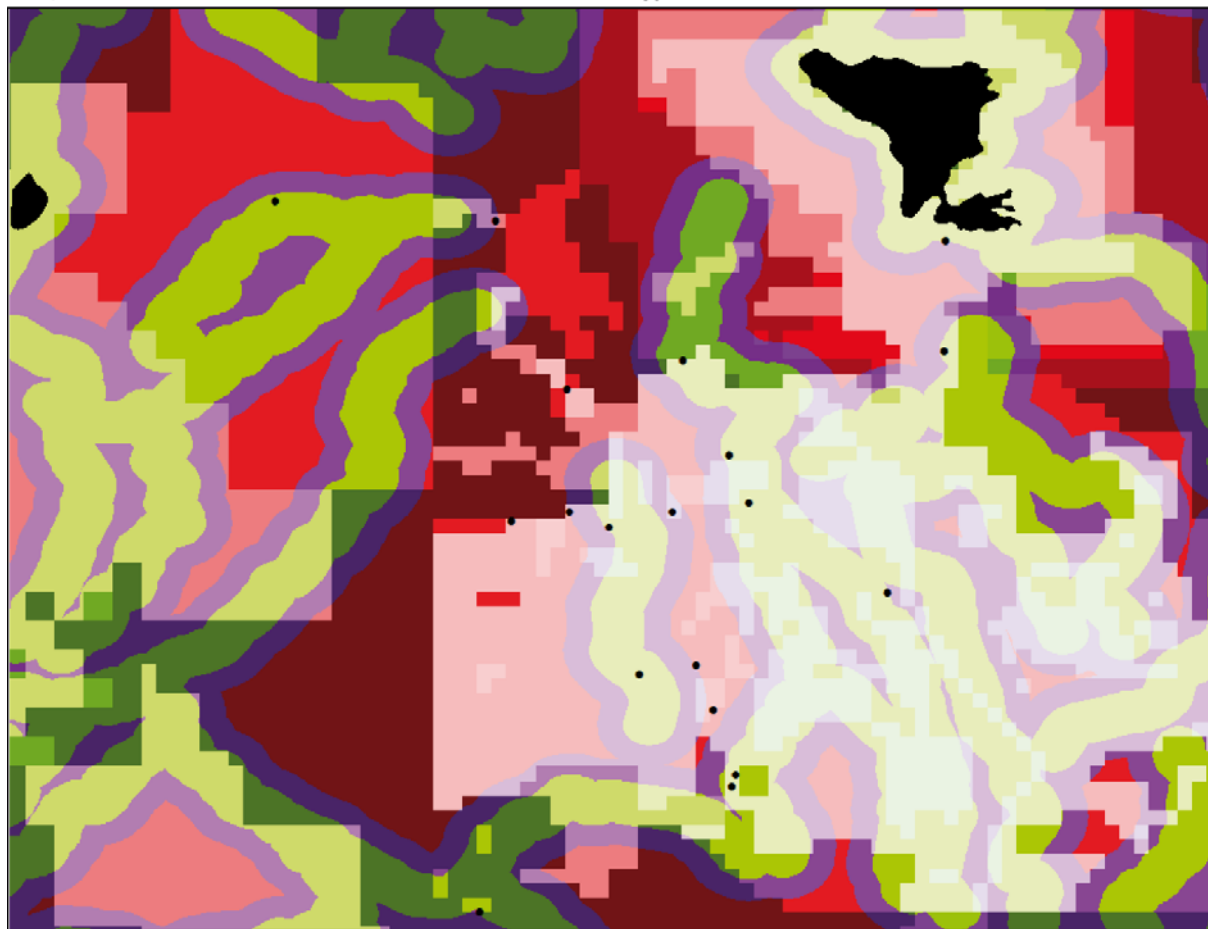


Figure 128 - Possibilités de localisation des sites évaluées à partir des *sous-ensembles flous de type 2* des sites *hot-spot* du temps 1Ba.

1Ba, sites fiables - *Outliers* - Sous-ensembles flous de type 1



● Sites fiables

▲
10 km

distance au cours d'eau	Possibilité de sites		Possibilité en %	Intensité de prospection de la zone					
				0	1	2	3	4	5
3 km	Elevée	1	100						
	Faible	0,5	50						
6 km	Moyenne	0,5	50						
	Faible	1	100						
9 km	Faible	0	0						
12 km	Faible	1	100						

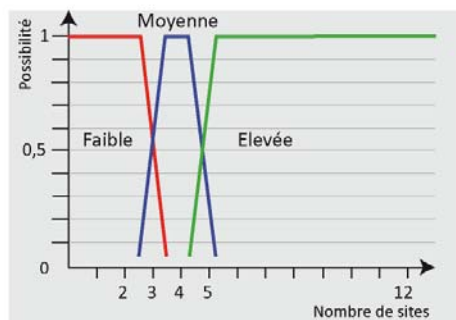
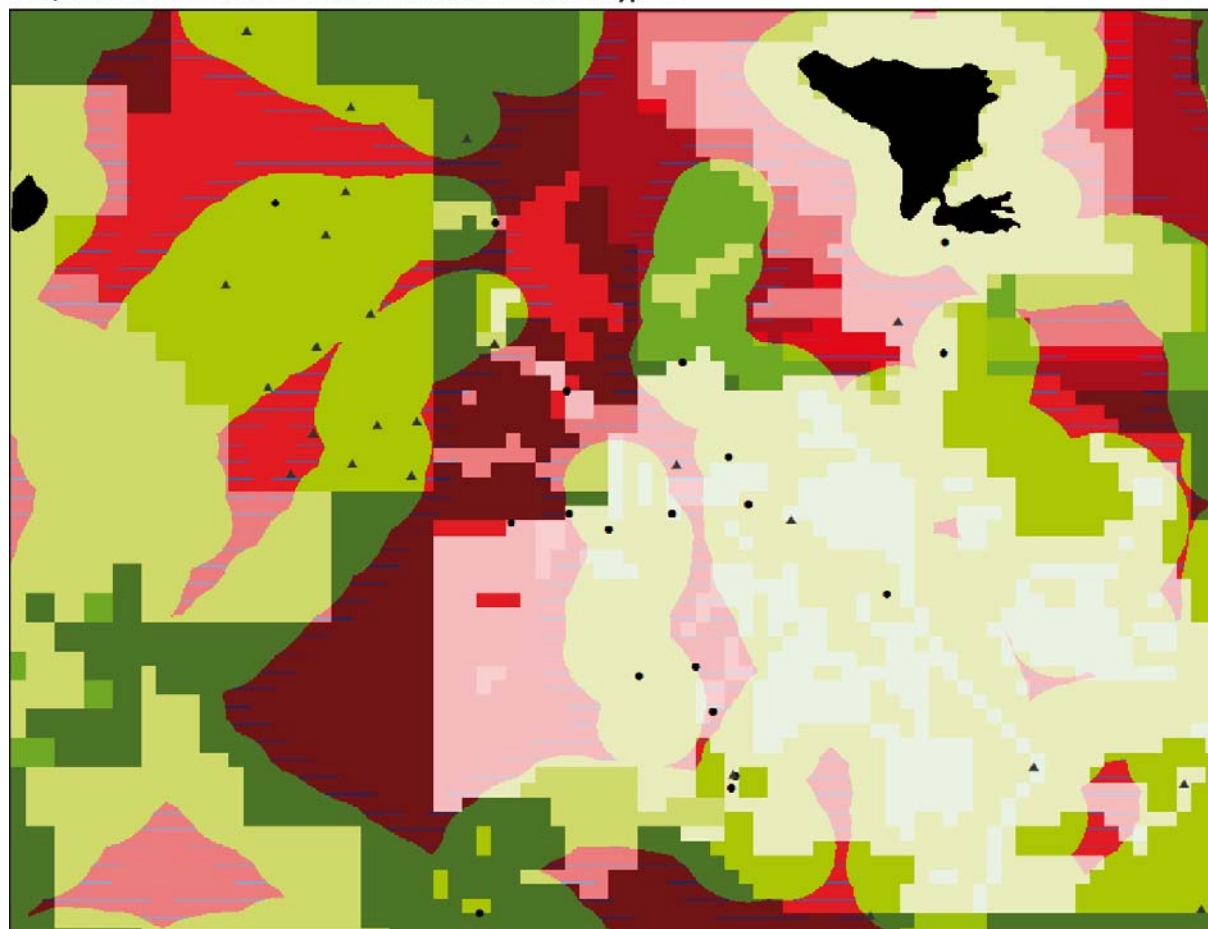


Figure 129 - Possibilités de localisation des sites évaluées à partir des *sous-ensembles flous de type 1* des sites *outliers* fiables du temps 1Ba.

1Ba, tous sites - *Outliers* - Sous-ensembles flous de type 1



- Sites fiables
- ▲ Sites non fiables

▲
10 km

distance au cours d'eau	Possibilité de sites		Possibilité en %	Intensité de prospection de la zone						
				0	1	2	3	4	5	
3 km	Elevée	1	100	Dark Green	Green	Light Green	Yellow-Green	Yellow	Light Yellow	White
6 km	Elevée	1	100	Dark Green	Green	Light Green	Yellow-Green	Yellow	Light Yellow	White
9 km	Faible	0,8	80	Dark Red	Red	Light Red	Pink	Light Pink	White	White
	Moyenne	0,2	20	Dark Red	Red	Light Red	Pink	Light Pink	White	White
12 km		0	0	Dark Red	Red	Light Red	Pink	Light Pink	White	White
15 km	Faible	1	100	Dark Red	Red	Light Red	Pink	Light Pink	White	White

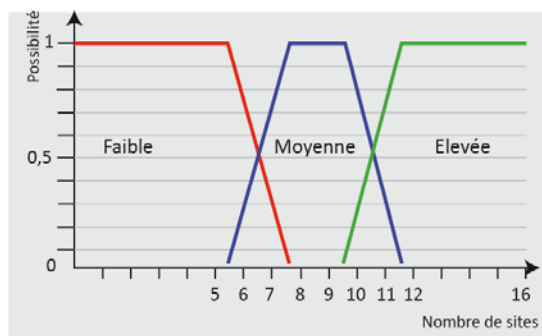
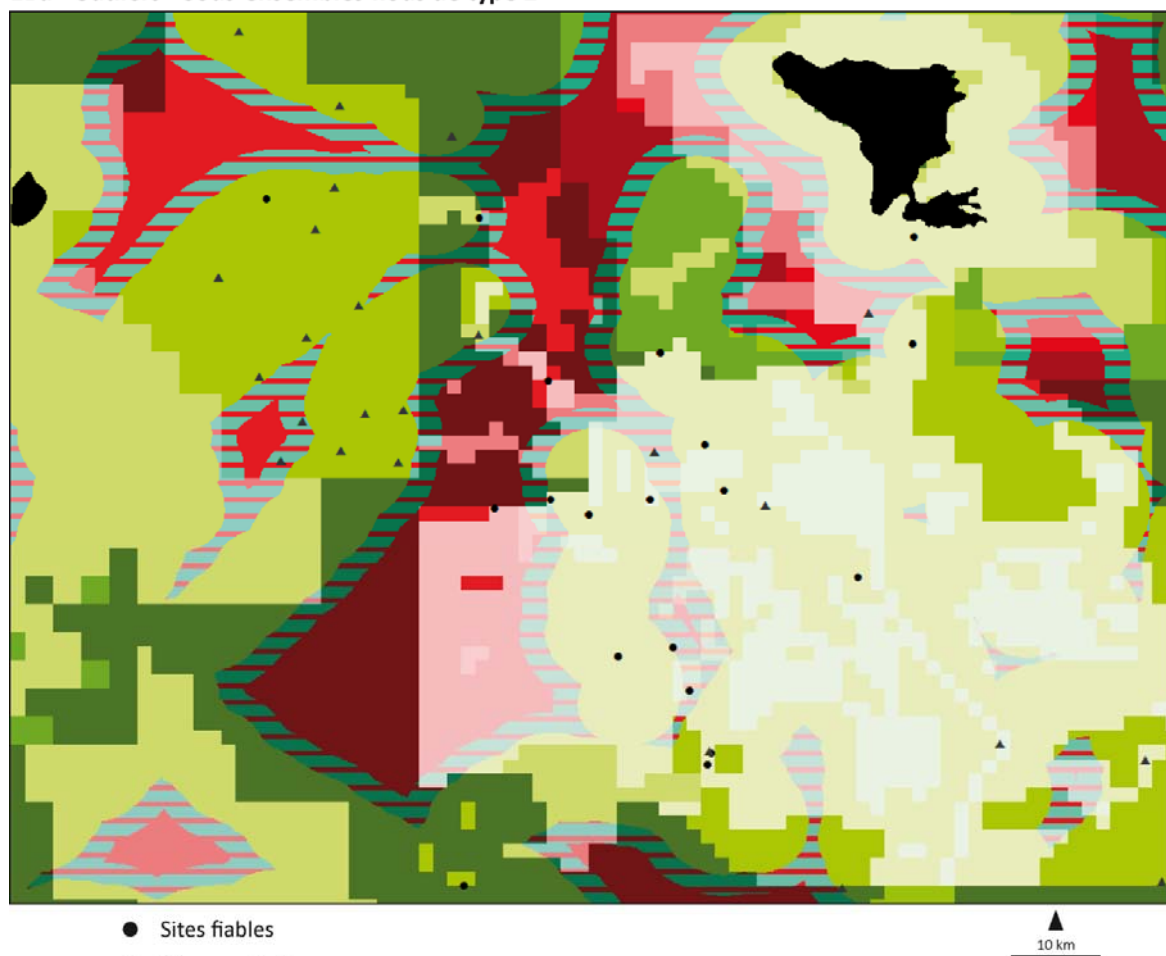


Figure 130 - Possibilités de localisation des sites évaluées à partir des *sous-ensembles flous de type 1* de la totalité des sites *outliers* du temps 1Ba.

1Ba - Outliers - Sous-ensembles flous de type 2



- Sites fiables
- ▲ Sites non fiables

10 km

distance au cours d'eau	Possibilité de sites		Intensité de prospection de la zone					
			0	1	2	3	4	5
3 km	Elevée	1						
	Elevée	1						
6 km	Elevée	1						
	Moyenne	1						
9 km	Faible	0 - 0,7						
	0	0						
12 km	Faible	1						
15 km	Faible	1						

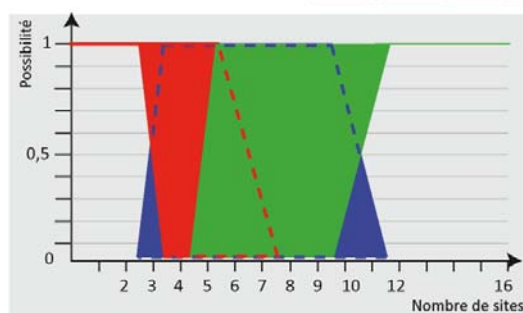


Figure 131 - Possibilités de localisation des sites évaluées à partir des *sous-ensembles flous de type 2* des sites *outliers* du temps 1Ba.

Nous avons choisi de distinguer, dans la cartographie des résultats, les possibilités définies par les *sous-ensembles flous de type 1* et 2, et la manière dont ceux-ci se déclinent spatialement en fonction des zones d'intensité de la prospection.

En effet, plus une zone a été prospectée, plus une zone « vide de sites » a des chances de traduire l'absence de peuplement plus qu'une absence de connaissance sur la zone. Moins la zone a été prospectée, plus une zone vide pourrait être susceptible de receler des sites... ce qui ne signifie pas pour autant que la présence de sites soit possible aussi intensément sur l'ensemble des zones non prospectées (c'est pour définir la variabilité de cette possibilité sur l'ensemble de la zone d'étude que les sous-ensembles flous ont été élaborés).

Afin de traduire ces deux informations (possibilités de localisation des sites évaluées sur le paramètre « distance aux cours d'eau », et intensité de la prospection) sur une même cartographie des résultats, nous avons choisi de décliner les possibilités « forte » (couleur verte), « moyenne » (couleur bleue) et « faible » (couleur rouge) dans des gammes d'intensité décroissant en fonction du niveau d'intensité de prospection de la zone concernée.

Par exemple, une zone non prospectée (intensité 0 de prospection) où la possibilité de localisation de sites serait évaluée comme « forte » se verrait attribuer la couleur vert foncé.

Une zone où la possibilité de sites serait toujours évaluée comme forte, mais qui a été très bien prospectée (intensité de la prospection 5) se verrait quant à elle attribuée la couleur vert clair. Ce dernier cas nous donne ainsi un autre type d'information, permettant de discuter à la fois la pertinence des cartes de possibilité, et les caractéristiques du peuplement : conserver les zones prospectées dans la cartographie des résultats permet de constater directement dans quelle mesure les sites correspondent spatialement au modèle de possibilités réalisé, et d'observer où se situent les écarts au modèle.

La densité de l'information représentée sur ces cartes peut néanmoins rendre la lecture difficile si l'on tente de tout regarder en même temps. Afin de faciliter la compréhension des résultats présentés, nous pourrions globalement les synthétiser ainsi, avec en entrée l'information que l'on souhaite en extraire :

- Si nous souhaitons porter attention à la dimension « possibiliste » des cartes, il convient de se focaliser sur les couleurs foncées (zones non prospectées, et dont le peuplement possible est recherché).

- Si notre intérêt est plutôt « descriptif », il convient plutôt de porter attention aux couleurs plus claires (zones bien prospectées, et sur lesquelles repose l'élaboration du modèle).

CONCLUSION : DISCUSSION SUR LA METHODE ET LES RESULTATS

Cartographier l'ensemble de la zone étudiée (et pas uniquement les zones non prospectées, c'est-à-dire les zones à investiguer) présente ainsi l'avantage de donner à l'observateur deux informations simultanées - pourvu qu'on replace les sites connus sur la carte :

- Les possibilités de localisation des sites dans les zones non prospectées ;
- La localisation des zones et des sites qui correspondent au modèle réalisé, et la localisation de ceux qui s'en écartent : ces informations nous sont extrêmement précieuses, car ce sont elles qui vont nous indiquer dans quelle mesure il serait nécessaire d'affiner ou de rectifier le modèle élaboré, tout en s'intéressant de plus près aux sites qui s'en écartent, et de se demander si ceux-ci relèvent d'un type de configuration spatiale ou d'un mode d'utilisation du sol différent, par exemple. C'est d'ailleurs la démarche suggérée par Hans Kamermans (2006): "These extrapolation models are most commonly used in archaeological heritage management archaeology, but may also have their use in scientific research, for example to analyze anomalies in an observed pattern".

En outre, ces cartes montrent bien, selon nous, l'intérêt et l'importance de l'intégration de multiples niveaux d'analyse dans l'évaluation des « passés possibles » :

- Niveaux temporels : que l'on se base sur des critères théoriques ou empiriques, les informations modélisées sont toujours censées représenter les modes d'occupation du sol des populations concernées, traduits en termes de besoins, de priorités, d'accessibilité, etc... Construire des modèles généraux et statiques va donc totalement à l'encontre de la variabilité des populations et de leurs dynamiques de changement. L'évolution temporelle des résultats de la corrélation entre la distance aux cours d'eau et le nombre de sites illustre clairement ce propos. Ainsi, un modèle possibiliste n'est inhérent qu'à un temps donné, mais également à un niveau temporel donné : les modèles réalisés au *temps* 1Ba (c'est-à-dire, concernant les observations réalisées à partir des sites de l'âge du Bronze Ancien IV) ou au *temps* 1A+1Ba (concernant la totalité des sites de l'âge du Bronze Ancien) ne nous donnent pas les mêmes résultats : la prise en compte de la totalité des sites attribuables à un niveau de sous-période

donné transforme de toute évidence les statistiques et les structures spatiales observées, qui rétroagissent sur les résultats du modèle empirique. Faire varier les niveaux temporels d'analyse permet d'enrichir les questionnements sur les biais que peut introduire le découpage en périodes par l'intermédiaire notamment des typologies céramiques, en visualisant concrètement et spatialement les impacts de ce découpage.

- Niveaux de fiabilité : une fois encore, la distinction des divers niveaux de fiabilité présente l'avantage de nous indiquer les cas où cette distinction n'introduit pas de différence considérable dans le modèle (comme c'est le cas pour les *hot-spots* du *temps* 1Ba par exemple : les modèles élaborés ne changent pas avec l'intégration des sites non fiables dans le calcul) et les cas où cette distinction est fondamentale. Dans notre cas, bien que les résultats du *temps* 1Ba ne changent pas quel que soit le niveau de fiabilité introduit dans l'analyse, nous considérons que celle-ci reste tout de même fondamentale étant donnée la configuration spatiale des données : l'intégration des sites non fiables modifie en effet considérablement les structures spatiales dans la mesure où ceux-ci occupent presque exclusivement une vaste zone au Nord-Ouest de la carte. Se priver de leur intégration, c'est donc aussi se priver d'informations précieuses sur l'étendue possible du peuplement à l'échelle de la zone étudiée.
- Types de configurations spatiales : les différenciations entre les sites, dans les modèles d'occupation du sol passée, sont généralement établies sur des critères inhérents au site en lui-même, c'est-à-dire à l'échelle même du site. Les modèles seront donc élaborés pour les sites d'habitat, funéraires ou à vocation militaire par exemple, si c'est la fonction qui est mise en avant. On pourra également distinguer les sites par leurs matériaux de construction ou bien encore leur taille. Mais plus rares sont les tentatives de différenciation des sites à partir de leurs structures spatiales, c'est-à-dire à une échelle plus petite (au sens géographique), par exemple, celle de la zone d'étude, et plus rare encore est l'intégration de ces informations dans la modélisation des possibilités de peuplement. Considérer dans la modélisation les sites archéologiques comme un simple semis de points dont les attributs ne sont qu'intrinsèques revient à ne considérer que deux échelles d'étude très extrêmes dans l'analyse : celle du site, et celle de la zone étudiée. Selon l'étendue de la zone étudiée, cela peut s'avérer dangereux, cela revenant à ignorer que tous les sites ne relèvent pas forcément (c'est le cas dans notre étude) de modes et de critères d'occupation du sol similaires. Une analyse des types de configurations spatiales (par le calcul des LISA par exemple) peut ainsi permettre, outre l'intérêt en lui-même de l'information produite pour la compréhension des systèmes de peuplement passés, et par l'intermédiaire des régionalisations établies sur la base des structures spatiales, une échelle intermédiaire dans l'observation (et donc, la modélisation) des

phénomènes et des possibles. L'intérêt de la démarche est illustré notamment en ce qui concerne le *temps* 1Bb : ne pas distinguer ce que nous avons appelé les *hot-spots* des *outliers* dans le calcul de la corrélation entre la distance aux cours d'eau et le nombre de sites et donc, prendre en compte la totalité des sites dans le calcul nous aurait conduit à un résultat du coefficient de corrélation très élevé (- 0,95 pour les sites fiables et -0,89 pour la totalité des sites) influencé par les *outliers* et donc, à considérer le paramètre « distance aux cours d'eau » comme un paramètre pertinent de modélisation pour le *temps* 1Bb.

Quoi qu'il en soit, les modèles présentés ici sont de toute évidence très incomplets, et doivent s'inscrire dans une dimension plus large, dépassant le cadre de cette étude qui, en tant qu'exercice académique, n'avait vocation qu'à en élaborer le cadre méthodologique et amorcer le mouvement, et non à en saisir la multitude des dimensions thématiques qui doivent être élaborées dans un cadre profondément interdisciplinaire. La complexité visée ici consistait plus dans la prise en compte des multiples niveaux spatio-temporels et d'incertitude que dans l'intégration de multiples renseignements thématiques au sein du modèle. Les nombreuses tentatives de cet ordre n'ont pu porter leurs fruits car les renseignements dont nous disposions ne pouvaient être traités directement, de manière brute (carte des sols ou des pluies par exemple) mais nécessitaient d'autres traitements en amont pour pouvoir être intégrés dans toute tentative de modélisation, traitements que nous n'étions, pour des raisons de temps principalement, pas en mesure de pouvoir fournir. Ces tentatives concernaient principalement le paramétrage d'un système d'inférence *flou* permettant, à la manière d'un système expert, de gérer une multitude de règles définies à partir de la quantification *floue* de plusieurs paramètres dont les combinaisons diverses permettent d'évaluer les divers sous-espaces de zone étudiée en fonction de la plus ou moins grande représentation de ces critères. Ces tentatives ont été réalisées à l'aide du freeware FisPro élaboré par Serge Guillaume (Guillaume, Charnomordic, 2011 ; 2013) -ne permettant toutefois pour le moment que d'envisager le *flou de type 1*.

La démarche adoptée dans ce chapitre s'inscrit dans une chaîne de raisonnement exploratoire plus large où l'on tente tout d'abord de dégager les niveaux spatiaux, temporels et d'incertitude auxquels s'expriment les données en allant du général au particulier, avant d'en modéliser certaines inconnues sous la forme de *possibles* (Figure 132).

Il en résulte qu'aucune des informations produites, quelle que soit l'étape considérée, n'a la vocation –ni la prétention- d'être considérée comme un « résultat » arrêté et immuable, mais est plutôt considérée comme une sorte de « moteur » propulsant l'étude et les analyses dans diverses directions. A chacune de ces étapes, nous avons tenté de conserver le plus possible de niveaux

d'analyse afin d'aborder les questionnements qui émergeaient sous des angles variés, et de telle sorte que chacune de ces dimensions nous fournisse un maximum de recul et d'esprit critique sur les informations obtenues.

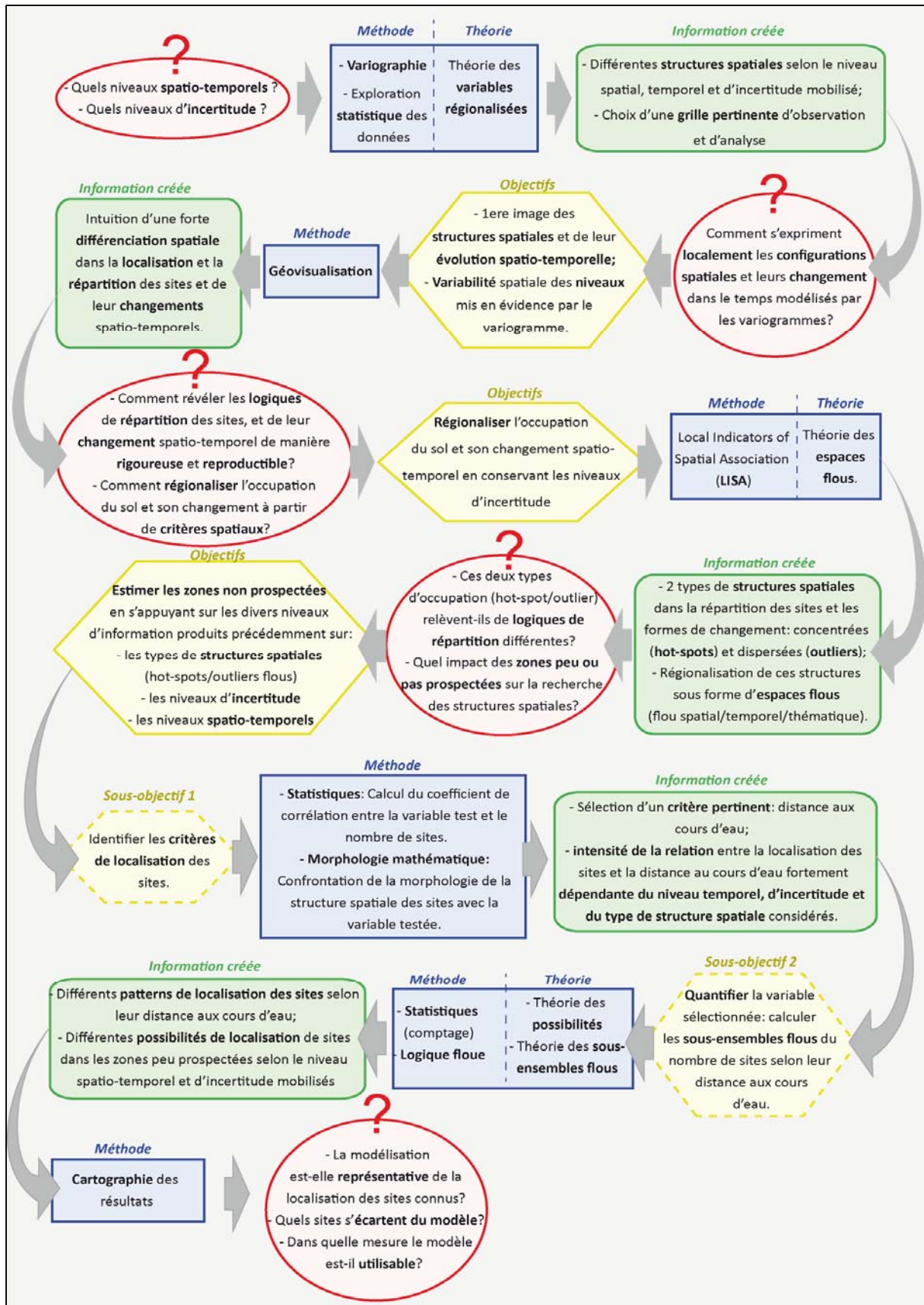


Figure 132 - La chaîne de raisonnement développée dans la partie III.