



HAL
open science

Atelier SITraDA (Systèmes d'Information et Traitements de Données Archéologiques) : compte rendu 2016 (année universitaire 2015-2016)

Bruno Desachy, Julie Gravier, Pablo Ciezar, Lydie Dussol, Laurine Guyot,
Léa Hermenault, Luc Sanson, Christophe Tuffery

► To cite this version:

Bruno Desachy, Julie Gravier, Pablo Ciezar, Lydie Dussol, Laurine Guyot, et al.. Atelier SITraDA (Systèmes d'Information et Traitements de Données Archéologiques) : compte rendu 2016 (année universitaire 2015-2016). [Rapport de recherche] Université Paris 1 - Panthéon Sorbonne; École Doctorale d'Archéologie (ED 112); UMR 7041 Arscan. 2016. hal-02456921

HAL Id: hal-02456921

<https://hal.science/hal-02456921v1>

Submitted on 27 Jan 2020

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

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

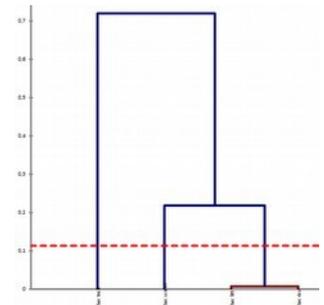
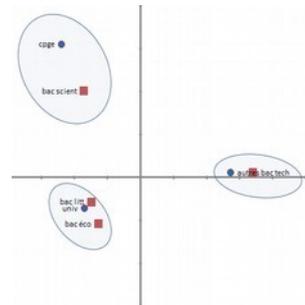
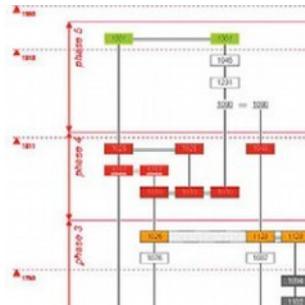
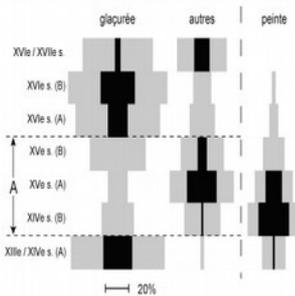


atelier SITraDA

(Systèmes d'Information et Traitement
des Données Archéologiques)

compte rendu 2016

(année universitaire 2015/2016)



coordonné et mis en forme par Bruno Desachy et Julie Gravier

Paris 1 – ED 112 – École doctorale d'archéologie

atelier SITraDA – compte rendu 2016

L'atelier SITraDA est un lieu d'échanges méthodologiques autour de travaux et projets de systèmes d'information et / ou de traitements des données appliqués à l'archéologie. L'atelier est ouvert aux doctorants et étudiants, ainsi qu'aux chercheurs porteurs de projets dans un cadre professionnel. Il comprend quatre thématiques principales :

- 1 - méthodes et outils d'enregistrement et de traitement des données de terrain ;
- 2 - synthèses archéologiques urbaines et territoriales ;
- 3 - sériations et typologies des éléments matériels ;
- 4 - exploration multidimensionnelle de sources croisées pour l'archéologie

L'atelier est animé par Bruno Desachy (MCC, Paris 1, UMR 7041 ArcScAn - bruno.desachy@univ-paris1.fr) avec l'aide de Julie Gravier (doctorante, Paris 1, UMR 8504 Géographie-Cités - julie.gravier@parisgeo.cnrs.fr).

<https://cours.univ-paris1.fr/fixe/03-40-doctorat-archeologie-atelier-sitrada>

sommaire

| | |
|---|-----------|
| participants 2015/2016..... | 5 |
| <i>contributeurs (présentations de projets).....</i> | <i>5</i> |
| <i>auditeurs.....</i> | <i>5</i> |
| réunion préliminaire samedi 5/12/2015..... | 6 |
| <i>actualité et informations diverses.....</i> | <i>6</i> |
| <i>organisation de l'atelier en 2016 :.....</i> | <i>6</i> |
| <i>colloques auxquels des membres de l'atelier ont assisté ou participé :.....</i> | <i>6</i> |
| <i>colloque ou formations à venir :.....</i> | <i>6</i> |
| <i>liens utiles :.....</i> | <i>7</i> |
| <i>présentation de projet : transactions foncières à Paris au XVIe s. (L. Hermenault).....</i> | <i>7</i> |
| séance 1 samedi 23/01/2016..... | 10 |
| <i>actualité et informations diverses.....</i> | <i>10</i> |
| <i>colloque ou formations à venir :.....</i> | <i>10</i> |
| <i>publications et travaux :.....</i> | <i>10</i> |
| <i>boîte à outils : le package Explor sur R (J. Gravier).....</i> | <i>10</i> |
| <i>présentation de projet : réflexions autour de la dynamique de l'habitat et de sa représentation dans la plaine de Troyes (L. Sanson).....</i> | <i>13</i> |
| <i>présentation de projet : le rôle des flux de circulation pour l'évolution du tissu urbain parisien entre le XVIe et le XIXe siècle (L. Hermenault).....</i> | <i>16</i> |
| séance 2 samedi 20/02/2016..... | 19 |
| <i>actualité et informations diverses.....</i> | <i>19</i> |
| <i>organisation de l'atelier :.....</i> | <i>19</i> |
| <i>colloques, rencontres et formations :.....</i> | <i>19</i> |
| <i>boîte à outils : sans décoder !? Quelques points à propos de programmation (B. Desachy).....</i> | <i>20</i> |
| <i>présentation de projet : outil Le Stratifiant (future version) – choix de développement (B. Desachy).....</i> | <i>23</i> |
| séance 3 samedi 19/03/2016..... | 28 |
| <i>actualité et informations diverses.....</i> | <i>28</i> |
| <i>recherches en cours :.....</i> | <i>28</i> |
| <i>colloques :.....</i> | <i>28</i> |
| <i>publications :.....</i> | <i>28</i> |
| <i>présentation de projet : traitement des indicateurs de datation et inscription dans le temps quantifié des unités stratigraphiques et des entités fonctionnelles (B. Desachy).....</i> | <i>28</i> |

| | |
|--|-----------|
| séance 4 samedi 23/04/2016 | 39 |
| actualité et informations diverses | 39 |
| recherches en cours :..... | 39 |
| colloques :..... | 39 |
| formations et outils :..... | 39 |
| présentation de projet : Interroger la taille des villes sur le temps long : premiers éléments de discussion (J. Gravier) | 39 |
| présentation de projet : reprise des données stratigraphiques d'une fouille ancienne (Orléans – Saint-Pierre-Lentin) (L. Guyot) | 48 |
| séance 5 samedi 21/05/2016 | 52 |
| actualité et informations diverses | 52 |
| colloques :..... | 52 |
| Boite à outils 1 : AnalyseSHS du PIREH (Pôle Informatique de Recherche et d'Enseignement en Histoire de Paris 1) (J. Gravier) | 52 |
| Boite à outils 2 : Un peu d'analyse spatiale avec R (J. Gravier) | 54 |
| Présentation de projet : traitement statistique de données anthracologiques (L. Dussol) | 58 |
| séance 6 samedi 25/06/2016 | 61 |
| actualité et informations diverses | 61 |
| colloques :..... | 61 |
| Présentation : Stratigraphie et SIG : quelques rappels, réflexions et suggestions sur les conceptions et représentations du temps et de l'espace et sur les liens entre chronologie et spatialisation dans les SIG (C. Tuffery) | 61 |
| Présentation de projet : traiter statistiquement les phénomènes de cycles et répétitions de pas variable : le cas des dérivées de sigillée d'Argonne décorées à la molette (P. Ciezar) | 74 |
| Index des projets :..... | 86 |
| Travaux universitaires :..... | 86 |
| Autres projets de recherche discutés dans l'atelier :..... | 86 |
| plan-type de notice de présentation de projet | 87 |

participants 2015/2016

contributeurs (présentations de projets)

Pablo Ciezar (INRAP) ; pablo.ciezar@inrap.fr

Bruno Desachy (MCC, Paris 1, UMR 7041 ArcScAn, équipe Archéologies Environnementales) ;
bruno.desachy@univ-paris1.fr

Lydie Dussol (doctorante, Paris 1, ED 112, UMR 8096 Archéologie des Amériques) ;
lydie.dussol@mae.u-paris10.fr

Julie Gravier (doctorante, Paris 1, ED 434, UMR 8504 Géographie-Cités) ;
julie.gravier@parisgeo.cnrs.fr

Laurine Guyot (étudiante en master, université de Nantes) ; laurineguyot@hotmail.fr

Léa Hermenault (doctorante, Paris 1, ED 112, UMR 7041 ArcScAn, équipe Archéologies Environnementales) ; lea.hermenault@univ-paris1.fr

Luc Sanson (INRAP Grand-Est-Nord) ; luc.sanson@inrap.fr

Christophe Tufféry (ingénieur, INRAP) ; christophe.tuffery@inrap.fr ;

auditeurs

Marie Anton (doctorante, Paris 1, ED 112, UMR 7041 ArcScAn) ; anton.marie@wanadoo.fr

Sophie Benhaddou (INRAP) ; sophie.benhaddou@inrap.fr

François Capron (INRAP) ; francois.capron@inrap.fr

Anne Chaillou (MCC) ; achail@9online.fr

Federico Defendenti (doctorant, EPHE, Università di Roma III, UMR 8167 Orient & Méditerranée) ;
fede.defe@gmail.com

Robert Demaille (chef de projet, SOGETI) ; robert.demaille@sogeti.com

Fabienne Dugast (ingénieure, CNRS, UMR 8167 Orient & Méditerranée) ; fabienne.dugast@cnrs.fr

Jérôme Haquet (ingénieur, CNRS, UMR 8167 Orient & Méditerranée) ; jfhaquet@cnrs.fr

Anaïs Pinhède (archéologue géomaticienne, ville de Chartres) ; anais.pinhede@hotmail.fr ;

Ingrid Renault (UMR 8167 Orient & Méditerranée) ; ingrid.renault@orange.fr

Emilie Roux-Capron (service archéologique municipal, ville d'Orléans) ; eroux@ville-orleans.fr

réunion préliminaire samedi 5/12/2015

9h30-12h30 IAA 3, rue Michelet 75006 Paris salle 319

présents : Robert Demaille, Bruno Desachy, Julie Gravier, Léa Hermenault, Anaïs Pinhède, Christophe Tuffery

actualité et informations diverses

organisation de l'atelier en 2016 :

- Mise en place d'un document annuel de compte rendu des séances, complété au fur et à mesure et finalisé lors de la réunion de bilan et de préparation de l'année suivante ;
- Programme prévisionnel : voir [page de l'atelier](#) sur le site de Paris 1 (espaces pédagogiques interactifs).

colloques auxquels des membres de l'atelier ont assisté ou participé :

Metrology for Archaeology

[Metrology for Achaeology](#), Benevent, Italie, 20-22 octobre 2015 ([contribution](#) de B. Desachy dans la session « *Time in the Archaeological Information Systems* ») ; actes du colloque en format .pdf bientôt disponibles.

Ontologies

Les journées «[Ontologies en Sciences Humaines et Sociales](#)» se sont tenues à Tours les 9-10 novembre 2015.

Les archives de fouilles

[Les archives de fouilles - Modes d'emplois](#), Paris, Fondation Hugot du Collège de France, 27-28 novembre 2015 ; une publication électronique des contributions est prévue.

colloque ou formations à venir :

Prochain CAA

Le [CAA 2016](#) se tiendra à Oslo du 29 avril au 2 mars 2016.

Formation de statistiques

[Un projet de stage](#) de formation CNRS « Statistiques multidimensionnelles appliquées à l'archéologie : traitement des tableaux de données qualitatives et quantitatives », prévu à Nanterre (MAE) du 9 au 13 mai 2016, animé par B. Desachy est en cours de finalisation.

Les JIAP

Les [Journées archéologique et informatiques de Paris](#) se tiendront du 7 au 10 juin 2016.

Dernière minute

L'atelier SIG de l'EHESS, animé par Laurent Costa, Sandrine Robert et Éric Mermet (atelier validé comme séminaire de parcours doctoral de l'ED 112) tiendra ses prochaines séances les mardi 16/1, 2/12, 8/3, 12/4, 3/5 et 7/6 2016.

liens utiles :

- Consortium [MASA](#) (mémoire des archéologues et des sites archéologiques) ;
- Programme européen [ARIADNE](#) (*Advanced Research Infrastructure for Archaeological Dataset Networking in Europe*) ;
- TGIR (Très Grande Infrastructure de Recherche) [HUMA-NUM](#) ;
- GDR [MAGIS](#) (groupement de recherche CNRS « Méthodes et applications pour la géomatique et l'information spatiale »).

présentation de projet : transactions foncières à Paris au XVe s. (L. Hermenault)

intitulé du projet dans lequel s'inscrit la présentation :

thèse en cours : « Ville et mouvements, le rôle des flux de circulation pour l'évolution du tissu urbain parisien aux périodes médiévale et moderne »

type de projet :

mémoire de thèse (Paris 1 – ED 112)

porteur(s) / auteur(s) du projet :

Léa Hermenault ; doctorante, Paris 1, ED 112, UMR 7041 ArcScAn, équipe Archéologies Environnementales ; lea.hermenault@univ-paris1.fr

thématiques de l'atelier concernées par la présentation

2 (synthèses archéologiques urbaines et territoriales) ; 4 (exploration multidimensionnelle de sources croisées pour l'archéologie)

précédentes présentations dans l'atelier du projet :

année 2014/2015 : séances du 15/11/2014 et du 27/06/2015

compte-rendu de présentation – éléments discutés et résultats de la discussion

(voir [diaporama de la présentation](#) du 5/12/2015).

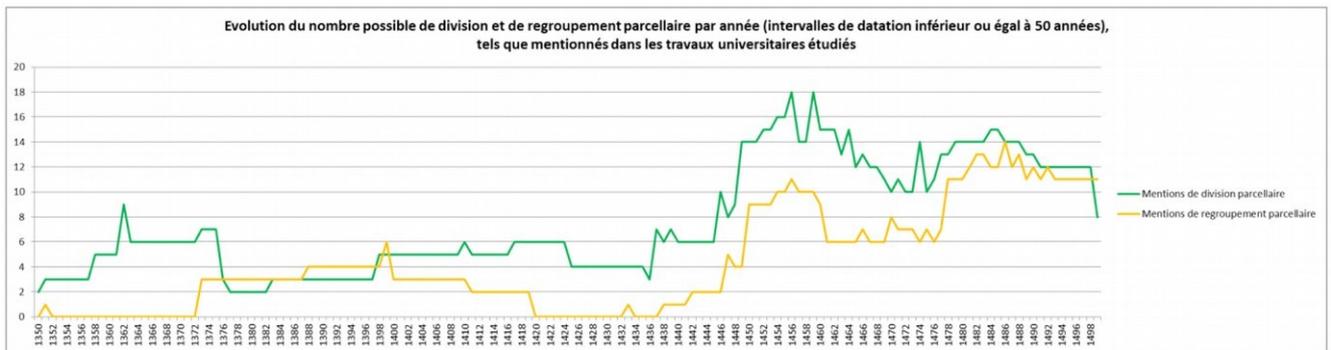
Dans le cadre de mon travail de recherche de thèse qui porte sur l'analyse des interactions, qui s'établissent à plusieurs échelles, entre la matérialité urbaine et les potentialités d'échanges induites par la proximité d'un flux à Paris entre le 15e et le premier tiers du 19e siècle, la question objet de la présentation est le traitement d'un corpus de données écrites.

Je cherche à analyser dans le temps 107 transactions foncières qui se déroulent dans plusieurs quartiers de Paris au 15e siècle. Ces transactions foncières (divisions ou regroupements parcellaires) ne sont pas datées précisément la plupart du temps. Je dois donc travailler avec des incertitudes de datation pouvant varier entre 1 et 50 années. Je cherche à identifier les périodes et les lieux où ces transactions ont préférentiellement lieu.

Je rencontre plusieurs problèmes dans le traitement de ces données :

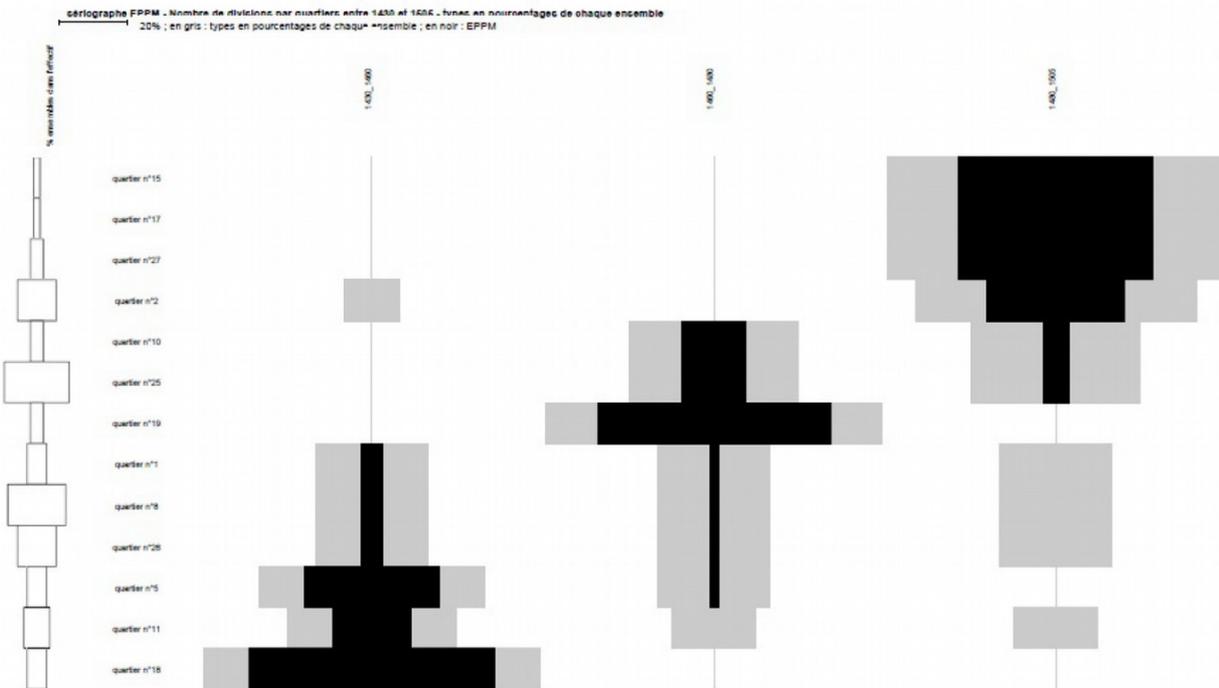
- 1) Comment gérer l'absence de continuité dans la série de données que j'exploite ?
- 2) Comment gérer ces incertitudes de datations ?
- 3) Quelle méthode utiliser pour analyser l'éventuelle corrélation entre deux des variables que j'étudie (évolution des transactions et revenus de l'affermage des chaussées, dans l'hypothèse de liens économiques et historiques entre l'évolution de la circulation et des échanges dont témoignent ces

revenus et celle du (ré) aménagement urbain dont témoignent les transactions foncières – p. 7 du diaporama) ?



©Léa Hermenault

4) L'utilisation de l'outil *Sériographe* est-elle la bonne solution pour mettre en valeur des similarités d'évolution dans le temps d'unités spatiales (quartiers et rues) à partir des données ici étudiées (p. 8, 10 et 11 du diaporama) ?



©Léa Hermenault

La discussion collective à partir de l'exposé et du diaporama de présentation a fait émerger les points suivants, concernant les 4 questions posées ci-dessus :

1) : ajouté à la simple mise en graphique (courbes) des séries chronologiques étudiées (nombre de transactions, comparé aux revenus de l'affermage des chaussées par unité de temps), le calcul de courbes de régression permettrait d'apprécier les tendances (à la hausse ou à la baisse) pour une période donnée, et de pallier ainsi aux lacunes limitées de données. Ces régressions (linéaires ou polynomiales) peuvent être obtenues simplement avec un tableur (Excel ou LibreOffice Calc : options de graphes XY). Toutefois il n'y a pas de miracle : en cas de lacune trop importante, une régression à partir de trop peu de points trop inégalement répartis ne permet plus de fonder valablement un commentaire.

2) : pour tenir compte de l'imprécision qui affecte la datation de certaines transactions sur l'échelle de temps en années (cas où l'on connaît l'intervalle d'années dans laquelle la transaction a eu lieu, mais pas l'année précise), il est de toutes façons préférable d'éviter de recompter les transactions pour chaque année dans leur intervalle d'imprécision (la courbe est faussée par le poids ainsi artificiellement multiplié des transactions imprécisément situées). Une solution simple mais peu satisfaisante consisterait à résumer chaque intervalle d'imprécision par un point moyen (année médiane de l'intervalle), et de ne représenter que ces points sur mon graphique. Une solution sensiblement meilleure et qui reste simple serait de réaliser 3 courbes :

- une courbe avec répartition de chaque transactions imprécisément située, suivant une valeur (1/nombre d'années de l'intervalle d'imprécision) pour chaque année de cet intervalle d'imprécision (ce qui correspond à une hypothèse de probabilité égale d'existence de la transaction pour chaque année de cet intervalle d'imprécision). Par exemple un intervalle d'incertitude de 4 années serait par exemple représenté par une valeur de 0,25 transaction par an (probabilité d'occurrence de 25% par an pour cette transaction) ;

- deux courbe figurant les deux hypothèses chronologique extrêmes, au plus ancien, et au plus récent (transactions comptées pour la première ou la dernière année de leur intervalle d'imprécision).

Ces 3 graphiques superposés sur la même échelle de temps permettraient d'apprécier l'impact de l'imprécision des données et de dégager les phénomènes signifiants perceptibles malgré cette imprécision.

3) : au vu des courbes, une association est effectivement perceptible entre l'évolution des transactions et celle des revenus de l'affermage des chaussées, pour une partie de la période étudiée. Toutefois il vaut mieux ne pas parler de « corrélation » avant d'avoir vraiment mesuré celle-ci ; or, vu les importantes lacunes dans l'une des deux séries, le calcul du coefficient de corrélation et le test de cette corrélation risquent fort de ne pas donner de résultat satisfaisant. L'hypothèse d'un lien signifiant entre les deux phénomènes peut néanmoins être proposée, comme simple induction (ce qui ne veut pas dire « faux » ou « non fondé », mais seulement « non évaluable statistiquement en l'état actuel des données »), et argumentée notamment à l'aide d'informations extrinsèques de contexte historique.

4) : le principe graphique du *Sériographe*, amène à considérer les lignes du tableau (les unités chronologiques) comme la référence par rapport à laquelle analyser les colonnes (les variables caractérisant les unités chronologiques). Sachant cela, il serait plus clair de transposer les tableaux de comptage (nombre de transactions) considérés (quartiers X unité de temps et rues X unités de temps), en plaçant les unités de temps en lignes, pour le traitement à l'aide de *Sériographe*. Comme outil graphique complémentaire, il serait possible d'utiliser *Matrigraphe* (qui permet une visualisation des écarts à l'indépendance sans privilégier les lignes par rapport aux colonnes, ni les écarts positifs par rapport aux écarts négatifs). Au delà de cette exploration graphique des données, pour établir un regroupement en classes des quartiers ou rues, il est possible d'opérer une analyse factorielle (AFC), puis une classification ascendante hiérarchique (CAH) sur les coordonnées factorielles de l'AFC ; les classes sont obtenues en choisissant de découper l'arbre de classification à une valeur donnée de similarité des individus (quartiers ou rues).

suite et développements postérieurs

Il est prévu que nous fassions le point sur ce traitement de données lors de la prochaine séance du séminaire qui aura lieu le 23 janvier 2016. Je présenterai également ce jour-là un autre corpus pour le traitement duquel j'aurais également besoin de conseils.

références

Poster présenté aux Journées informatiques et archéologiques de Paris 2016

séance 1 samedi 23/01/2016

9h30-12h30 IAA 3, rue Michelet 75006 Paris salle 319

présents : Marie Anton, Sophie Benhaddou, Pablo Ciezar, Robert Demaille, Bruno Desachy, Lydie Dussol, Julie Gravier, Jérôme Haquet, Léa Hermenault, Luc Sanson, Christophe Tuffery

actualité et informations diverses

colloque ou formations à venir :

Projet de réseau national sur les statistiques au CNRS

C. Tuffery signale un projet de constitution d'un réseau au sein du CNRS, destiné aux statisticiens et aux utilisateurs des statistiques, et visant dans un premier temps à « recenser les besoins (formations, conseils, bonnes pratiques...) de la communauté, les attentes de chacun par rapport à son métier et identifier, au niveau du CNRS, les personnes impliquées aussi bien dans l'analyse statistique proprement dite, en passant par l'utilisation de logiciels (SAS, R...), que dans le développement méthodologique. ». Il s'agit d'un projet interdisciplinaire, à champ plus large que le réseau [CAI-RN](#) existant spécialisé en archéométrie et archéologie. Le projet est porté par Amandine Blin (amandine.blin@mnhn.fr), Bérengère Saliba-Serre (berengere.saliba-serre@univ-amu.fr), Sandrine Cabut (cabut@mmsch.univ-aix.fr), Thibault Laurent (thibault.laurent@tse-fr.eu) et Cécile Lecoœur (cecile.lecoeur@cnrs.fr).

(dernière minute) : analyse de réseaux en histoire et archéologie

Signalée par C Tuffery, une session sera consacrée à l'analyse de réseaux en histoire et archéologie dans le cadre de la prochaine conférence européenne sur les réseaux sociaux qui aura lieu à Paris du 14 au 17 juin : <http://eusn2016.sciencesconf.org/89651> ; clôture de l'appel à contributions le 16 février.

(dernière minute) : colloque : Creuser au Mésolithique

Signalé (et coorganisé) par L. Sanson, un colloque sur les fosses mésolithiques, dont l'étude renvoie au choix des méthodes d'enregistrement et des unités d'analyse du terrain et concerne donc la thématique 1 de l'atelier ; fiche d'inscription [sur le site de l'INRAP](#).

(dernière minute) : MOOC « Introduction à la statistique avec R »

Le MOOC (cours en ligne) gratuit « Introduction à la statistique avec R », de l'Université de Paris Sud se déroulera du 29 février au 1^{er} avril :

<https://www.france-universite-numerique-mooc.fr/courses/UPSUD/42001S05/session05/about>

publications et travaux :

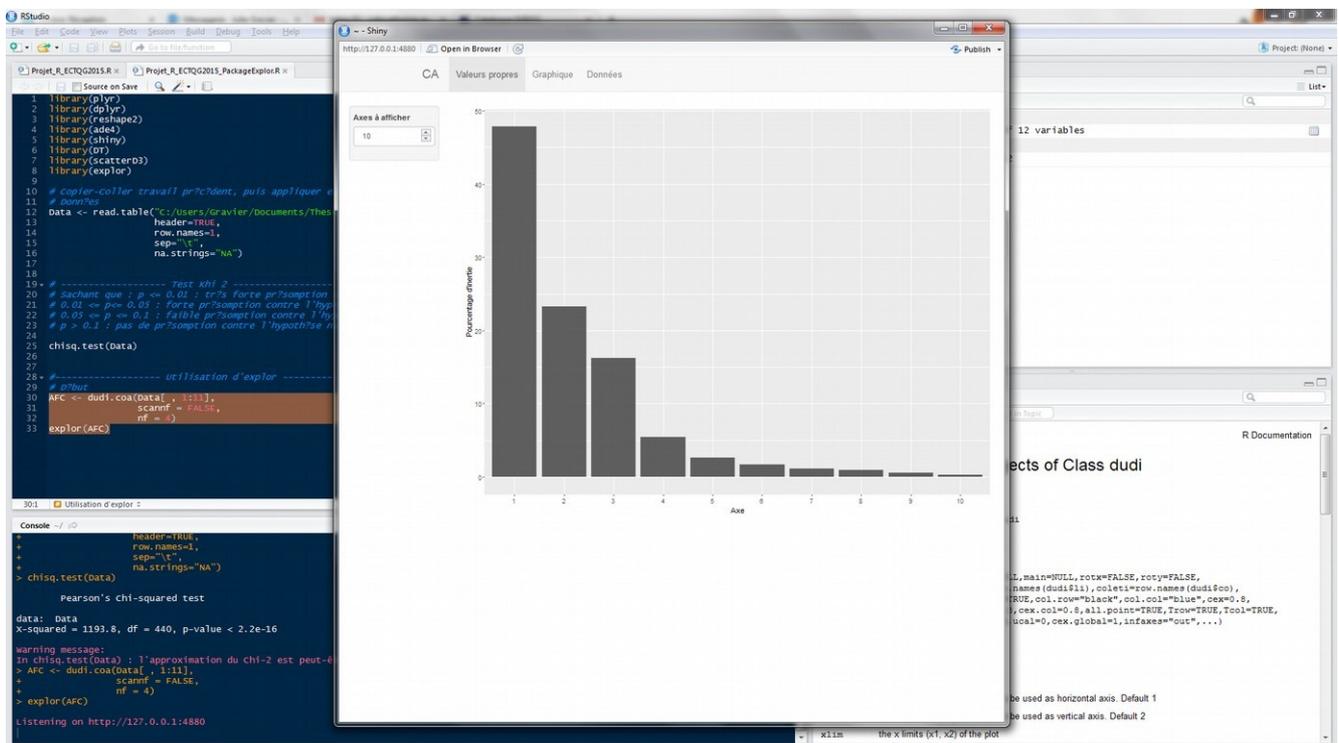
(dernière minute) : HDR de Xavier Rodier

Le dossier d'HDR de Xavier Rodier, soutenue le 11 janvier à Besançon, avec sa synthèse de travaux de recherches intitulée « Espace-Temps des sociétés du passé », est [disponible en téléchargement sur HAL](#).

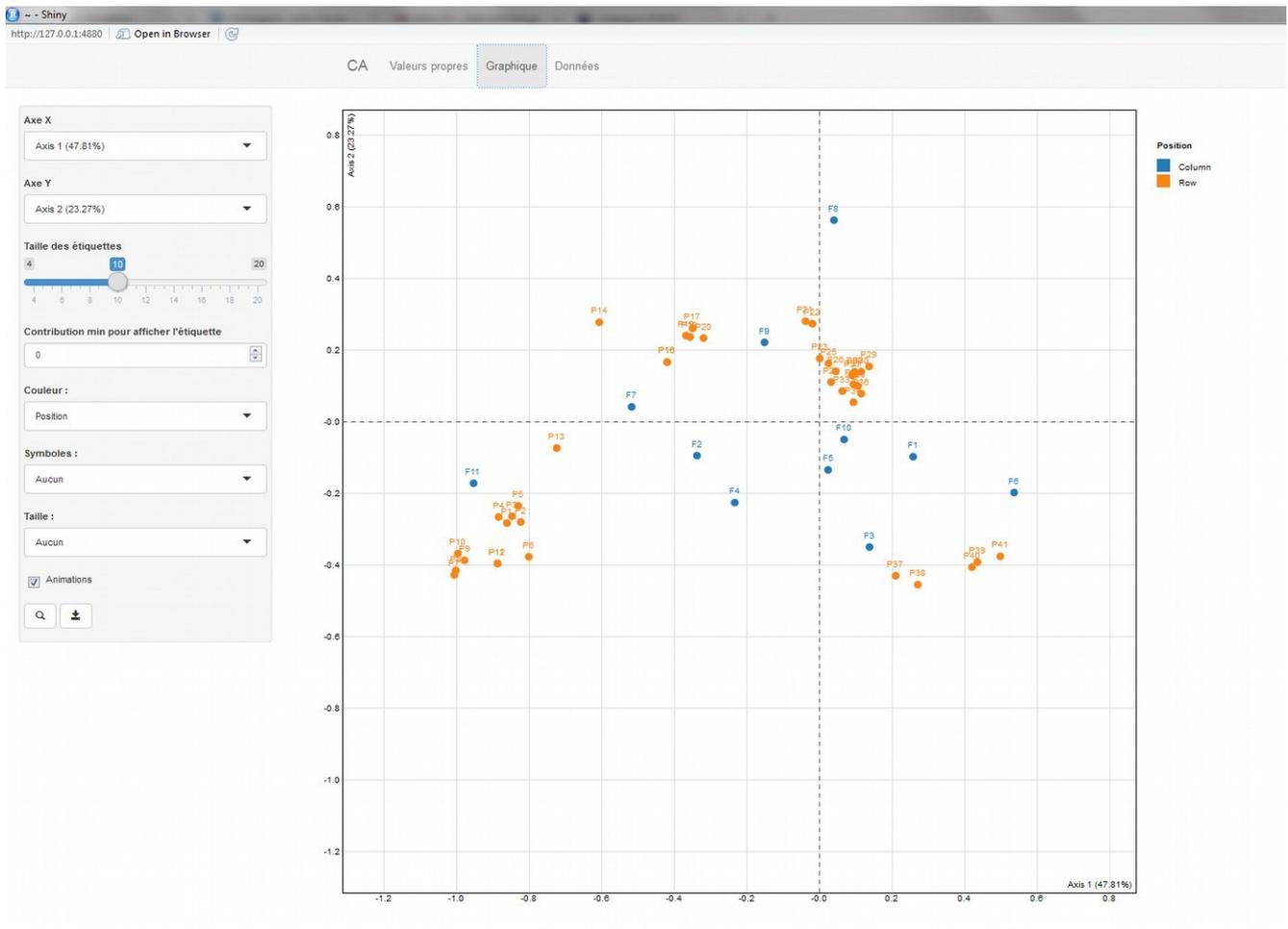
boite à outils : le package Explor sur R (J. Gravier)

qu'est-ce que R ?

Il s'agit d'un logiciel libre, gratuit et multiplateforme. Sans entrer dans de longs détails parfaitement développés dans de nombreuses publications, il faut au moins avoir en tête que R est fondé sur : 1) une base commune (dénommée r-base) et 2) des *packages*, l'équivalent d'extensions, que l'on peut définir comme « une bibliothèque de fonctions implémentées par les utilisateurs et mises à disposition de tous par l'intermédiaire de dépôts regroupés dans le cadre du *Comprehensive R Archive Network* (CRAN) » (Commenges et al. 2014 : 1-2). Or, ces packages sont très polyvalents puisqu'ils couvrent des domaines variés comme l'analyse de graphe, la cartographie... ce qui tend à devoir comprendre R non pas comme un logiciel-langage uniquement dédié aux analyses statistiques, mais bien plus largement aux traitements de données. Notons également que la structure de R est bâtie sur une modularité importante, où les fonctions d'un package vont dépendre d'un autre package – et ainsi de suite – formant une logique en arbre. Cette structure est bien entendu la force du langage (créativité, adaptabilité...). Cependant, elle implique pour les novices une prise en main qui n'est pas évidente car un même traitement pourra être fait de plusieurs manières. De plus, cette organisation modulaire implique une forte diversité dans les formats d'objets, chaque package pouvant être doté de ses propres classes de données.



visualisation avec *Explor* : valeurs propres



visualisation avec Explor : graphique de plan factoriel (AFC)

présentation d'explor

Le package est développé par Julien Barnier, informaticien à l'ENS Lyon, rattaché à l'équipe *Dispositions, pouvoirs, cultures et socialisations* de l'UMR 5283 – Centre Max Weber. La version 0.1 d'explor a été déposée sur CRAN le 15 décembre 2015.

les fonctionnalités

Explor propose d'effectuer des visualisations et explorations de traitements statistiques multidimensionnelles : analyse en composantes principales (ACP), analyse factorielle des correspondances (AFC) et analyse des correspondances multiples (ACM). L'intérêt majeur d'explor est que l'exploration interactive et dynamique de données est possible au moyen de l'utilisation des packages *shiny*, *DT* et *scatter3D*. Un autre des atouts d'explor est que le package est utilisé avec *FactoMineR* ou *ade4*, qui sont les deux packages de référence pour effectuer des traitements statistiques multidimensionnelles avec R.

test sur le traitement des analyses factorielles des correspondances

Nous avons actuellement uniquement testé et présenté le traitement des AFC avec explor, tout simplement car nous n'avons pas encore eu l'intérêt d'utiliser des ACP et des ACM pour traiter les données de nos travaux.

Le package permet de visualiser à la fois le diagramme des valeurs propres, le tableau des coordonnées, contributions et cosinus carré des différents axes factoriels, ainsi que le graphique

factoriel lui-même. L'aspect dynamique de l'exploration s'effectue au moyen de nombreuses fonctions visuelles : zoom/dézoom ; jeux sur les transparences ; les étiquettes des lignes (observations) et colonnes (variables) du tableau de contingence initial peuvent être déplacées sur le graphique, leurs tailles ajustées et même être affichées selon un seuil minimal de contribution ; la taille des points relatifs aux lignes et aux colonnes peut également être ajustée sous forme de cercles proportionnels selon la somme de leurs contributions sur les deux axes représentés, et une légende est directement associée au graphique.

De plus, les graphiques des AFC peuvent être directement sortis en format .svg, utile pour des potentielles retouches sur Inkscape par exemple. D'après ce que nous avons pu observer, ce n'est pas le cas pour le diagramme des valeurs propres qui sert cependant à l'argumentation dans le cadre du choix des axes étudiés. Cela nous semble particulièrement important au regard des données archéologiques – généralement lacunaires à la fois dans le temps et dans l'espace – et pour lesquelles l'argumentation de l'analyse statistique multidimensionnelle doit bien souvent se fonder sur les quatre premiers facteurs d'une AFC. En cela, nous conseillons d'associer dans l'écriture de sa propre chaîne de traitement dans R les éléments relatifs à l'exploration de données grâce à *explor*, ainsi que l'automatisation des traitements et sorties les plus utiles (diagramme des valeurs propres et tableaux des contributions des lignes et colonnes).

quel intérêt pour l'archéologue ?

Il nous semble que ce nouveau package est intéressant et utile pour les archéologues à deux principaux titres :

- En complément par exemple de l'interface d'exploration Web *Analyse SHS* développée par le PIREH – et plutôt bien connue des archéologues de l'Université Paris 1 – *explor* permet de travailler en local. Cela nous apparaît intéressant au regard des contextes des terrains archéologiques où l'accessibilité à internet peut encore faire largement défaut ;
- Aussi, ce package permet d'obtenir une interface dynamique d'exploration de données en très peu de lignes de code. Cet aspect nous semble important à mettre en avant afin que son utilisation soit facilement élargie dans le cadre de la communauté archéologique.

références

Commenges et al. 2014

COMMENGES H. dir., BEAUGUITTE L., BUARD É., CURA R., LE NÉCHET F., LE TEXIER M., MATHIAN H., REY S., *R et espace. Traitement de l'information géographique* [en ligne], Lyon : Framasoft, URL : <http://framabook.org/r-et-espace/>.

À propos du package :

<http://cran.r-project.org/web/packages/explor/explor.pdf>

<http://github.com/juba/explor>

À propos d'Analyse :

Leur Wiki : http://pireh.univ-paris1.fr/wiki/doku.php?id=ressources_pour_analyse

<http://analyse.univ-paris1.fr/>

présentation de projet : réflexions autour de la dynamique de l'habitat et de sa représentation dans la plaine de Troyes (L. Sanson)

intitulé du projet dans lequel s'inscrit la présentation :

La Plaine de Troyes : évolution d'un territoire rural des premiers agriculteurs au premier réseau villageois

type de projet :

projet collectif de recherches (PCR)

porteur(s) / auteur(s) de la présentation et du projet :

Luc Sanson (INRAP Grand-Est-Nord) ; luc.sanson@inrap.fr

projet coordonné par Vincent Riquier avec participation INRAP Grand-Est-Nord, SRA, UMR 8215, UMR 6289, UMR 7209

thématiques de l'atelier concernées par la présentation

1 (méthodes et outils d'enregistrement et de traitement des données de terrain) ; 2 (synthèses archéologiques urbaines et territoriales)

précédentes présentations dans l'atelier du projet :

année 2014/2015

compte-rendu de présentation – éléments discutés et résultats de la discussion

La présentation portait principalement sur la question de la visualisation de l'évolution des occupations archéologiques au sein de la plaine de Troyes, dans le cadre du PCR. La représentation phase par phase permet d'observer des phases temporelles figées et délimitées par leurs bornes temporelles.

Plusieurs essais ont été pratiqués, guidés par l'objectif de mieux apprécier les dynamiques, la représentativité des données et de trouver le ou les modes de représentation visuelle adaptés. Dans l'immédiat, il apparaît qu'une analyse par maille semble une base robuste pour représenter la plupart des données quantitatives et permet de dépasser la simple projection de points ou de polygones.

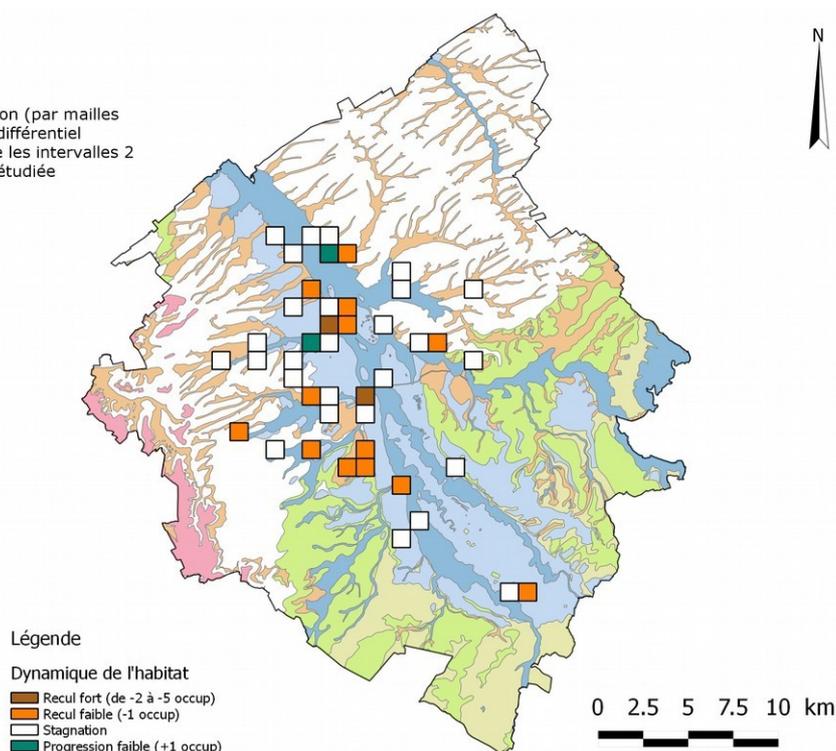
Toutefois, dès lors que l'on souhaite travailler sur une base temporelle, il est nécessaire de coupler plusieurs modes de représentation. Des logiciels ou extensions de QGIS (comme *TimeManager*) sont des pistes de réflexion pour permettre une visualisation des différentes dynamiques temporelles ayant lieu dans la plaine de Troyes.

La discussion, pendant et après la présentation, a porté sur plusieurs points.

La question de l'organisation du travail collaboratif a été abordée ; l'information spatiale fait l'objet d'une structuration globale commune en trois couches majeures (emprises totales des opérations, aires réellement ouvertes, et unités d'observation) permettant de traiter de façon comparable les données d'opérations récentes et le dépouillement des données anciennes. Ce niveau commun d'informations est partagé au moyen d'un serveur de données spatiales (postGis). L'enregistrement analytique détaillé au niveau de chaque opération reste lui propre à cette opération et n'est pas partagé.

La question de la quantification de l'occupation archéologique, à l'intérieur du maillage kilométrique retenu comme grille d'analyse spatiale, a aussi été abordée. Les unités d'occupation archéologique décomptées sont définies par une fonction (suivant une grille fonctionnelle à 5 catégories majeures) ; par une continuité spatiale (marquée par un polygone, pouvant résulter de la fusion de plusieurs polygones si le chercheur estime que plusieurs observations issues de plusieurs opérations proches forment une même occupation fonctionnelle) ; et par une continuité chronologique dans un intervalle de temps quantifié. On note que ce type d'unité se rattache conceptuellement à la notion d'entité archéologique (dont la première formalisation en France fut celle des « repères urbains » des documents d'évaluation du centre national d'archéologie urbaine), et semble bien constituer l'unité conceptuelle fondamentale de toute carte archéologique, déclinée suivant des échelles et des catégories fonctionnelles différentes.

carte de visualisation (par mailles kilométriques) du différentiel d'occupation entre les intervalles 2 et 3 de la période étudiée



Concernant l'essai présenté de représentation de l'évolution, la période entre -100 et + 200 a été découpée en 3 intervalles de temps successifs ; les cartes montrent le différentiel d'occupation entre les intervalles, en termes de recul ou de progression en nombre d'occupations d'habitat, par maille kilométrique.

En réponse au questionnement sur d'autres représentations possibles de la dynamique d'évolution (outre l'utilisation envisagée de *TimeManager*), plusieurs avis sont émis :

- le recours à des graphes sur échelle de temps quantifiés, en regard des cartes ; en effet, par définition ces dernières sont limitées à la représentation à un état ou une étape donnés ;
- pour chaque intervalle de temps, la mise en classes et en échelle de couleur de la « profondeur de temps d'occupation » affectant chaque maille kilométrique occupée, en fonction des bornes anciennes des occupations existantes dans l'intervalle considéré ; la première et plus récente classe étant celle des nouvelles créations, c'est-à-dire les occupations dont les bornes anciennes sont égales ou plus récentes que celle de l'intervalle considéré, la classe suivante étant celle des occupations remontant à l'intervalle précédent, la 3^e classe étant celle des occupations remontant à l'intervalle encore précédent, etc. ;
- ces représentations de profondeur de temps ou de différentiel chronologique ont pour inconvénient de ne pas tenir compte de la densité d'occupation lors de l'intervalle considéré ; cette information peut être figurée par ajout d'une symbologie appropriée, ou de façon vraisemblablement plus pratique, au moyen d'une carte de densité (suivant le même maillage chronologique), en regard de chaque carte chronologique.

Un point technique a été abordé, celui du découpage en classes des valeurs affectées aux mailles kilométriques et traduites par une échelle de couleurs. La méthode de Jenks sur les données de chaque carte (découpage de classes correspondant à la variance interclasses maximale et aux variances intraclasses minimales, c'est-à-dire de sorte que la distance moyenne entre individus à l'intérieur de chaque classe soit la plus faible possible et la distance moyenne entre individus de classes différentes soit la plus élevée possible) est la plus adaptée aux données d'une carte en particulier, mais ne favorise pas la comparaison car le découpage en classes change d'un intervalle chronologique et d'une carte à l'autre. Des solutions alternatives sont évoquées :

- mise en classes à intervalles fixes, basés sur la médiane et les quantiles (plutôt que la moyenne et la variance) des séries de valeur à mettre en classes (les indicateurs non paramétriques que sont la médiane et les quantiles peuvent être plus parlants et plus adaptés aux lacunes et imprécisions grévant inévitablement les données archéologiques) ;

- découpage en classe portant sur l'ensemble des valeurs pour tous les intervalles et les cartes traités : la mise en classes serait ainsi commune à toute l'évolution étudiée.

En addition à la discussion menée en séance, les aspects de sémiologie graphique ont fait l'objet de remarques communiquées par C. Tuffery, notant sur les cartes jointes au rapport de PCR (Riquier 2015 : 118 à 121, 124 à 128) des améliorations possibles pour remédier à des problèmes de cohérence sémiologique (harmonisation de la nature et de la précision des fonds de carte, meilleure adaptation des couleurs et symbologies aux données quantitatives à représenter, utilisation de flèches pour représenter des flux dynamiques), et de légendes à compléter.

références

Riquier 2015

RIQUIER V. coord., *Projet Collectif de Recherche La Plaine de Troyes : « évolution d'un territoire rural des premiers agriculteurs au premier réseau villageois »*, Troyes.

présentation de projet : le rôle des flux de circulation pour l'évolution du tissu urbain parisien entre le XVe et le XIXe siècle (L. Hermenault)

intitulé du projet dans lequel s'inscrit la présentation :

thèse en cours : « Ville et mouvements, le rôle des flux de circulation pour l'évolution du tissu urbain parisien aux périodes médiévale et moderne »

type de projet :

mémoire de thèse (Paris 1 – ED 112)

porteur(s) / auteur(s) du projet :

Léa Hermenault ; doctorante, Paris 1, ED 112, UMR 7041 ArcScAn, équipe Archéologies Environnementales ; lea.hermenault@univ-paris1.fr

thématiques de l'atelier concernées par la présentation

2 (synthèses archéologiques urbaines et territoriales) ; 4 (exploration multidimensionnelle de sources croisées pour l'archéologie)

précédentes présentations dans l'atelier du projet :

année 2014/2015 : séances du 15/11/2014 et du 27/06/2015 ;

année en cours : réunion du 5/12/2015

compte-rendu de présentation – éléments discutés et résultats de la discussion

(voir [diaporama de la présentation](#) du 23/01/2016).

Mon travail de recherche de thèse porte sur l'analyse des interactions, qui s'établissent à plusieurs échelles, entre la matérialité urbaine et les potentialités d'échanges induites par la proximité d'un flux à Paris entre le 15e et le premier tiers du 19e siècle.

Lors de la séance du 5/12/2015, j'ai présenté un corpus de données écrites que je cherche à exploiter dans le cadre de cette recherche (cf. [compte-rendu de présentation](#)). Plusieurs solutions m'avaient été

proposées concernant les questions posées lors de cette présentation. L'objectif de la présente intervention est de faire le point sur ce traitement de données.

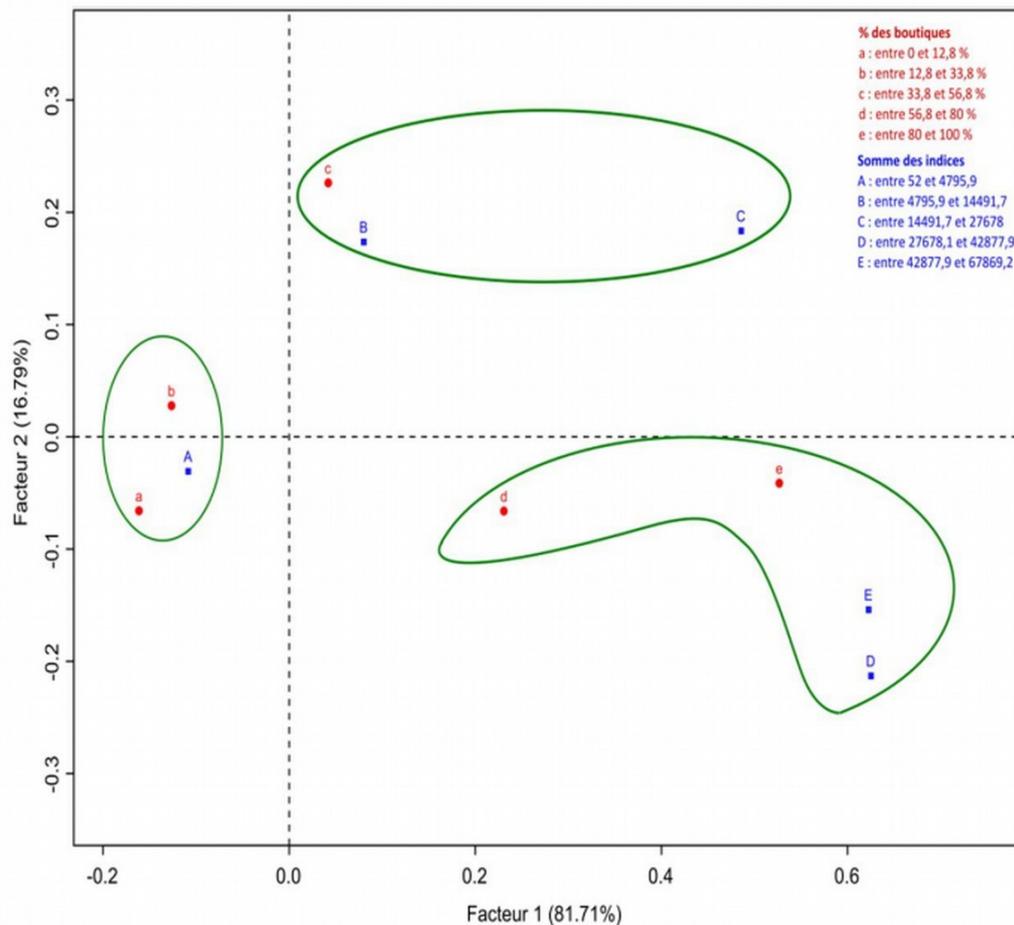
1 – Concernant une première question posée – comment gérer les discontinuités dans la série de données que j'exploite – il m'avait été conseillé de calculer des courbes de régression pour permettre d'apprécier les tendances à la hausse ou à la baisse. Le calcul d'une courbe de régression linéaire sur la série de données non continue des montants tirés de l'affermage des chaussées donne un résultat difficilement interprétable, ainsi que le montre le coefficient de régression qui est très bas. Je divise la série en trois parties pour mettre de côté les intervalles de temps pour lesquels nous n'avons pas de données, et je recalcule une courbe de régression linéaire sur chacune de ces parties. Le résultat est bien meilleur et permet de bien apprécier la tendance à la hausse ou à la baisse de la variable. Le résultat serait sans doute plus intéressant si je divisais la série de donnée seulement en deux parties et non trois. Le calcul d'une courbe de régression polynomiale est intéressant puisque cette courbe rend bien compte de la dynamique générale de la variable sur le temps long, mais avec de forts étalements et atténuations principalement dues à l'importante lacune de données entre 1428 et 1440. Il serait intéressant que je calcule là aussi des courbes de régression polynomiale sur la série de données séparée en deux parties (de part et d'autres de cette lacune). Dans le cadre de mon travail, il faut que je présente ces deux types de courbes de régression.

2 – Concernant la deuxième question posée lors de la précédente présentation – comment gérer les incertitudes, ou plus exactement les imprécisions de datations – plusieurs solutions m'avaient été conseillées pour gérer différemment les intervalles de datation auxquels je suis confrontée (*cf. [compte-rendu de présentation](#)*). Les trois méthodes testées (point médian, vision superposées des chronologies au plus tôt calée sur les bornes inférieures et au plus tard calée sur les bornes supérieures, répartition « probabiliste » des transactions imprécises par fractions annuelles à l'intérieur de l'intervalle d'imprécision) donnent des résultats assez différents et montrent bien la difficulté qu'il y a à exploiter ces intervalles. Elles montrent toutes cependant en premier lieu une forte tendance à la hausse des occurrences de transactions dans la seconde moitié du 15^e siècle. La méthode des points médians et la répartition par fractions annuelles permettent d'identifier en second lieu un très léger décalage dans le temps de l'occurrence des types de transaction. Il faut prendre garde à un biais de cette méthode de répartition par fractions annuelles, due aux différences d'amplitude des intervalles d'imprécision : l'état des sources fait que dans la seconde moitié du 15^e siècle, les dates d'occurrence des transactions sont plus précisément connues qu'elles ne le sont auparavant, ce qui renforce artificiellement les occurrences de transaction par années, par rapport à la période précédente.

3 – Concernant la question de l'utilisation du sériographe elle aussi posée lors de la précédente présentation, le traitement réalisé sur la base de comptages annuels (incluant les transactions imprécisément situées dans le temps et réparties par fractions annuelles – ou « probabilités d'occurrences annuelles » – dans leurs intervalles d'imprécision) montre une évolution par quartier ; qui serait toutefois plus claire si l'on mettait de côté un de ces quartiers dont le comportement semble cyclique et qui vient perturber la sériation. Le même traitement en prenant les rues et non plus les quartiers comme unités spatiales donne lui un très bon résultat (sériation claire) à partir duquel il est possible de sortir des analyses intéressantes.

4 – À côté de ces transactions foncières, je travaille sur un autre corpus de données, basé lui sur le comptage des boutiques et des portes cochères dans les rues de la rive droite de Paris au début du 18^e siècle. Je cherche à analyser la localisation des fortes densités de chacun de ces équipements dans la ville. L'unité d'analyse spatiale est le segment de rue (défini entre deux croisements). Chaque segment de rue est caractérisé par son appartenance à l'une des 5 classes de densité de boutique, à l'une des 5 classes de densités de portes cochères, et à l'une des 5 classes d'indice de centralité. Deux AFC ont été réalisées sur des tableaux de comptages (de segments de rue), la première croisant classes de densité de boutiques et classes d'indices de centralité, la seconde croisant classes de densité de portes cochères et classes d'indices de centralité.

AFC boutiques/indice de centralité



La première AFC montre une claire partition en trois ensembles de rues. On me conseille de réfléchir à la composition d'un seul tableau de données intégrant les données boutiques/portes cochères/indices de centralité (auxquelles il pourrait également être intéressant d'ajouter des données concernant la largeur des rues) pour mieux faire ressortir des profils de rues en fonction de leurs caractéristiques. Il pourrait s'agir d'une analyse des correspondances multiples (ACM).

références

[Poster](#) présenté aux Journées informatiques et archéologiques de Paris 2016

séance 2 samedi 20/02/2016

9h30-12h30 IAA 3, rue Michelet 75006 Paris salle 319

présents : Anne Chaillou, Federico Defendenti, Robert Demaille, Bruno Desachy, Julie Gravier, Léa Hermenault, Christophe Tuffery

actualité et informations diverses

organisation de l'atelier :

Pour laisser le plus possible la discussion se dérouler autour d'une présentation de projet sans avoir à couper court en raison du temps, on essaiera désormais de ne pas prévoir plus d'une présentation de projet importante par séance ; une même séance pourra en revanche contenir aussi une présentation plus courte, de type « boîte à outils » (exposé ou retour d'expérience pratique sur l'utilisation de logiciels ou procédures informatiques), ou une discussion initiale sur un projet à son début (opportunité et types de traitements de données à envisager et/ou structuration des données).

colloques, rencontres et formations :

réseau RIGMA (et projet ISHTAR)

Anne Chaillou signale la tenue d'une table ronde du réseau interprofessionnel de gestionnaires de mobilier (RIGMA) à Besançon en janvier 2015, où les questions de structuration des données relatives à la gestion des mobiliers ont été abordées. Les actes de la table ronde sont en cours de publication et une prochaine table ronde aura lieu en 2017.

Au cours de cette table ronde, a notamment été présenté le [projet ISHTAR](#) de système de gestion du mobilier et de la documentation scientifique, développé en région des Pays de la Loire sous la conduite de Yann Lejeune.

L'adresse Internet du réseau RIGMA (<http://www.rigma.fr/>) ne semble malheureusement pas être opérationnelle actuellement.

CAA Oslo (29/3/2016 – 2/4/2016)

Le prochain [colloque CAA](#) aura lieu à Oslo, sur le thème « *Exploring oceans of data* ». Christophe Tuffery y présentera une communication dans le cadre de la session 11 (*Supporting researchers in the use and re-use of archaeological data: continuing the ARIADNE thread*) consacrée au programme européen [ARIADNE](#). Cette communication évoquera les résultats d'une formation faite en juillet 2015 à Prato (Pôle de l'Université de Florence) dans le cadre du programme ARIADNE à propos de l'utilisation du CIDOC-CRM, en particulier de son extension [CIDOC-CRMarchaeo](#) adaptée pour l'archéologie.

ontologie CIDOC-CRM appliquée à l'archéologie

Outre la communication ci-dessus, Christophe Tuffery signale [l'application en ligne 3M](#) qui permet de tester un modèle de BD archéo avec l'ontologie du CIDOC-CRM.

conférence Spatial Humanities 2016 (Lancaster)

Signalée par Julie Gravier, une conférence qui porte sur l'utilisation des SIG dans les SHS, avec un appel particulier pour l'histoire et l'archéologie : http://www.lancaster.ac.uk/fass/projects/spatialhum.wordpress/?page_id=1347

Notamment : "*We welcome submissions on all aspects of using GIS in humanities research from database development to applied research in which spatial technologies have made a contribution to understanding of the past.*"

(dernière minute) : à propos de probabilités

À noter un débat en cours, signalé par Julie Gravier, sur le seuil de probabilités à partir duquel on peut décider d'écarter le risque que le phénomène observé soit aléatoire (et donc non significatif ; on prend traditionnellement le seuil 0.01 – soit 1 % de risques d'avoir affaire à des variations aléatoires ; ou, en étant un peu plus permissif, le seuil 0.05 – soit 5 % de risques ; voire, si l'on est très joueur, le seuil 0.1 en acceptant 10 % de risques). Voici la position de l'American Statistical Association dans ce débat, [ici](#) et [là](#) ; et un [exemple rigolo](#) trouvé par Julie. Ce qui est bien sûr en jeu derrière, ce sont les conditions d'utilisation des tests qui donnent ces probabilités...

boite à outils : sans décoder !? Quelques points à propos de programmation (B. Desachy)

Pour que la présentation de l'avancement de la version future du *Stratifiant* (outil de traitement des données stratigraphiques et chronologiques – voir ci-dessous) soit mieux partagée, il a paru utile d'aborder au préalable quelques éléments de programmation informatique.

À la base de la programmation, il y a l'algorithmique, c'est-à-dire la formalisation logique de processus de traitement – en ce qui nous concerne, de traitements archéologiques. Coder un algorithme, c'est l'écrire en langage informatique (ou même en « pseudo code » qui pourra ensuite être traduit dans différents langages réels). Cette écriture participe à la conception de cet algorithme ; car elle oblige à préciser la formalisation en éliminant toutes les lacunes et ambiguïtés. Le code informatique est fondamentalement une forme de langage mathématique et logique. En tant que tels, algorithmes mais aussi code ont une dimension méthodologique qui intéresse les archéologues.

Plus généralement, cet intérêt méthodologique de l'application de notions mathématiques et logiques aux processus de traitement que l'on trouve derrière le code informatique, fait l'objet aujourd'hui d'une certaine (re)prise de conscience, comme le montre l'entrée (ou plutôt le retour) du code dans l'enseignement général secondaire et supérieur : le code est à la mode !

Bien sûr l'objectif ne saurait être de transformer tout le monde en développeur professionnel ; mais pour l'archéologue du XXI^e siècle (voire du III^e millénaire) comme pour d'autres métiers, quelques aperçus au niveau du code et des langages informatiques ne sont pas inutiles, afin de mieux comprendre ou mieux suivre l'informatisation des processus de travail, voire pour participer à cette informatisation. En effet, comme chacun le sait plus ou moins, il existe maintenant des langages de développement et de macrocommandes accessibles aux utilisateurs avertis, contenus dans (ou utilisables depuis) la plupart des logiciels courants, qui permettent d'adapter plus efficacement ces logiciels à son travail (à condition bien sûr d'éviter les bricolages brouillons). Le but de cette présentation était d'évoquer ces possibilités.

aperçu historique

on peut distinguer schématiquement quatre grandes étapes du code informatique :

- le code binaire (puis hexadécimal), avec lequel, dans les années 1940, il fallait directement programmer les premières machines électroniques au moyen d'interrupteurs (pour des usages militaires, de cryptographie ou de balistique par exemple) ;
- les premiers vrais langages de programmation, formés d'instructions prédéfinies plus aisément manipulables par les opérateurs humains pour implémenter un algorithme, et traduites en code binaire par la machine elle-même. Ces premiers langages (FORTRAN, COBOL, etc.) sont

apparus dans les années 1950 pour des applications scientifiques et financières. Au niveau du matériel, ils sont liés au développement des moyens de stockage de l'information (cartes perforées, mémoires magnétiques) et des interfaces utilisateur/ordinateur (clavier, imprimantes, puis écrans cathodiques). Au niveau du logiciel – ces deux niveaux, matériel et logiciel, étant dès lors distincts – ils sont liés à la programmation dite procédurale et à ses composantes de base pour gérer les procédures (i.e. les suites d'opérations) :

- des variables de différents types (texte, nombre, etc.) pour stocker et transformer les données,
- des instructions pour maîtriser le déroulement d'un programme en interrompant, orientant ou relançant des procédures : les célèbres instructions de condition (« if...then... »), et de répétition (boucles).

Variables, conditions et boucles sont aujourd'hui encore les trois piliers de la partie procédurale des langages de programmation et de macrocommandes ;

- le développement de l'informatique dans les années 1960 et 1970 voit l'avènement de la programmation dite structurée : c'est à dire l'enrichissement de la programmation procédurale par la conception en systèmes de programmes et sous programmes liés par des appels. En pratique cela permet par exemple le recours par un nouveau programme à des bibliothèques de sous-programme déjà écrits. Mais surtout, conceptuellement, cela permet une structure plus claire (donc plus fiable) des programmes complexes ;
- l'étape suivante est celle dite de la « programmation orientée objet ». Au-delà de la programmation procédurale avec ses variables « passives » (auxquelles il faut appliquer une procédure écrite dans le programme pour le moindre traitement), et dans le prolongement des bibliothèques de sous-programmes rendues possible par la programmation structurée, l'idée est de pouvoir manipuler directement des « briques » élémentaires de stockage et de traitement des données, contenant à la fois une structure descriptive de la catégorie (ou « classe ») de données concernée (les « propriétés »), et des traitements ou éléments de procédure applicable à cette classe (les « méthodes »). Par exemple une liste de valeurs peut être une variable (de type « tableau ») en programmation procédurale ; si on veut la trier, il faut écrire des instructions spécifiques de tri (et, en programmation structurée, stocker celles-ci dans un sous-programme et y faire appel en tant que besoin). En programmation objet, si l'on dispose d'une classe (un type d'objet) permettant de stocker une liste de valeurs, et munie d'une méthode permettant de trier ces valeurs, on peut créer une occurrence de cette classe (par exemple « MaListe ») contenant les valeurs voulues, aussi simplement que l'on aurait créé une variable tableau ; un simple recours à la méthode voulue de cet objet permet alors d'obtenir la liste triée (c'est à dire une seule instruction dans le programme, de type « MaListe.Sort »). Au delà d'un avantage pratique et économique évident (gain de temps de programmeur !), objets, propriétés, méthodes, classes et relations entre classes sont conceptuellement de puissants outils de modélisation et d'intégration des différents domaines de la programmation informatique.

des outils accessibles à l'utilisateur averti

Dès les années 1960-1970 sont apparus des langages simplifiés conçus dans un objectif pédagogique d'initiation à la programmation, mais permettant aussi de créer des programmes opérationnels : d'abord le BASIC offrant les bases de la programmation procédurales (variables, conditions et boucles) ; puis le PASCAL pour l'apprentissage de la programmation structurée. Dans les années 1980, ce furent les vecteurs d'une première incursion de la programmation informatique dans les collèges et les lycées, et d'un boom de la programmation amateur avec les premiers micro-ordinateurs personnels (les éléments les plus vieillissants de l'atelier se souviennent peut-être avec émotion du rudimentaire GWBasic Microsoft ou du plus classieux TurboPASCAL Borland, sous le système d'exploitation MSDOS et son écran noir en mode texte...)

Aujourd'hui, grâce à la programmation orientée objet, ces langages et leurs descendants peuvent manipuler des logiciels d'application (ces logiciels et leurs commandes étant eux-même des objets), remplaçant les langages spécifiques de macrocommandes. Ils constituent ainsi un niveau de développement et de programmation « infra-professionnel » (c'est à dire accessible aux utilisateurs avertis, au prix d'un minimum de formation) constitué des « API » (application programming interface) offertes par un grand nombre d'applications, et des « EDI » (environnements de développement intégré) permettant une utilisation conviviale de ces langages .

deux produits-phares : VisualBasic et Python

L'offre actuelle est largement libre et/ou gratuite mais paraît à première vue complexe et pléthorique. Concernant les outils généralistes, en mettant de côté le cas spécifique du SQL (langage spécialisé dans la gestion de bases de données, qui mérite par ailleurs que l'utilisateur averti s'y intéresse – on y reviendra probablement lors de futures séances), ainsi que le cas de certains logiciels qui continuent à avoir leur propre langage de macrocommandes spécifique (au hasard : Filemaker...), on retiendra ici deux outils dominants : du coté du logiciel libre : le langage Python ; et du coté du logiciel commercial dominé par Windows : le langage VisualBasic de Microsoft.

- Dans le monde Windows, le langage VisualBasic (un lointain descendant du Basic historique) s'est imposé, sous deux déclinaisons :
 - le VBA (VisualBasic for Application) est le langage de développement interne des applications de la suite Microsoft Office. Puissant, mais très convivial, il a supplanté les langages de macrocommandes plus limités qui existaient précédemment (dans Excel notamment). C'est devenu un standard : les autres éditeurs de logiciel sous Windows l'ont aussi adopté (par exemple ESRI pour le SIG ArcGis). Ainsi le *Sériographe*, le *Stratifiant*, mais aussi par exemple l'additif d'analyse statistique multidimensionnelle XLstat, sont des développements réalisés à l'aide du VBA contenu dans Excel. Dans tous les logiciels qui le contiennent, la commande Alt+F11 permet de passer en coulisses, dans l'environnement de développement intégré VBA ;
 - le VB.net est un langage destiné à la création de logiciels autonomes exécutables, non lié à une application en particulier (même si il peut aussi piloter les applications compatibles VBA). Il est très semblable au VBA, mais avec de plus grandes possibilités liées à une orientation objets plus poussée et surtout à l'environnement de développement – en quelque sorte l'atelier logiciel – VisualStudio avec lequel il est distribué. Dans sa version professionnelle payante, VisualStudio inclut, au choix, plusieurs langages outre VB.net (y compris du Python, mais aussi par exemple le très pointu et professionnel langage C++) pouvant utiliser un même ensemble d'objets et classes (le « framework »). Bonne nouvelle, il en existe une version gratuite (VisualStudio Community) limitée, mais déjà très complète.

VBA et VB.net ont pour avantages la simplicité d'origine du langage Basic, un grand confort d'utilisation (aide en ligne, auto-complétion du code...), une documentation abondante, et le fait d'être des langages compilés, c'est-à-dire traduits en langage machine après avoir été écrits, permettant une exécution plus rapide (et dans le cas du VB.net, permettant la création d'applications autonomes de type .exe que l'on peut déployer et utiliser sous Windows sans installer le langage ni aucun autre logiciel). L'inconvénient est la limite à l'univers Windows.

- Le Python, lui, né dans les années 1990, a d'abord été utilisé sur le système d'exploitation Linux et ses dérivés. Il est aujourd'hui multi-plateformes (Linux, Windows et Mac). Outre ce premier atout, on peut résumer ses avantages à son caractère libre et gratuit ; à sa grande diffusion (en particulier dans le monde universitaire et scientifique) garante d'une certaine pérennité ; à sa reconnaissance par d'importants logiciels libres qui l'acceptent comme langage de développement interne et de macrocommandes : par exemple Inkscape (dessin), QGis (SIG), Libre/OpenOffice (qui dispose par ailleurs de son propre langage interne, clone – moins puissant hélas – du VBA Microsoft) ; et à ses qualités formelles : c'est un langage très concis (à

la différence du Basic), permettant de créer du code considéré comme particulièrement clair. Enfin, il dispose de nombreuses bibliothèques de fonctions et sous-programmes (interfaces graphiques, calcul scientifique, etc.). Dans ce qui peut apparaître comme des inconvénients – bien qu'ils s'agisse de choix volontaires des créateurs et non de défauts – le python est un langage interprété et non compilé, ce qui signifie notamment qu'on ne peut pas fabriquer d'exécutables indépendant : pour exécuter un programme Python, il faut que le langage et toutes les bibliothèques de fonction nécessaire aient été chargés dans l'ordinateur concerné. D'autre part, certains aspects d'utilisation sont encore peu ou mal documentés (par exemple l'utilisation comme outil de développement d'applications internes à LibreOffice...), ce qui les réserve de fait à des développeurs très expérimentés. Enfin, l'environnement de développement quoique complet, est d'apparence et d'ergonomie nettement plus spartiates que ce que propose Microsoft... R. Demaille précise cependant que le Python est en soi un langage très accessible et bien documenté, avec une communauté active. On peut donc se mettre sans problème à son apprentissage ; c'est d'ailleurs le langage qui a été retenu pour l'enseignement du code au lycée.

Liens:

site et téléchargement python : www.python.org ;

site et téléchargement [VisualStudio Community](http://VisualStudioCommunity.com) ;

parmi la documentation très abondante accessible en ligne, on peut signaler entre autres le site OpenClassRoom (<https://openclassrooms.com/>) qui propose des MOOC d'initiation gratuits (sur inscription).

présentation de projet : outil *Le Stratifiant* (future version) – choix de développement (B. Desachy)

intitulé du projet dans lequel s'inscrit la présentation

développement de l'outil de traitement de données stratigraphiques et chronologiques *Le Stratifiant* (cadre actuel du projet : UMR 7041 ArScAn, programme Archéologie du Bassin Parisien)

type de projet :

projet de recherche professionnel

porteur(s) / auteur(s) du projet :

Bruno Desachy ; MCC, Paris 1, UMR 7041 ArcScAn, équipe Archéologies Environnementales et programme Archéologie du Bassin Parisien ; bruno.desachy@univ-paris1

thématiques de l'atelier concernées par la présentation

1 (méthodes et outils d'enregistrement et de traitement des données de terrain)

contenu de la présentation et de la discussion

La présentation (voir [diaporama](#)) comprenait :

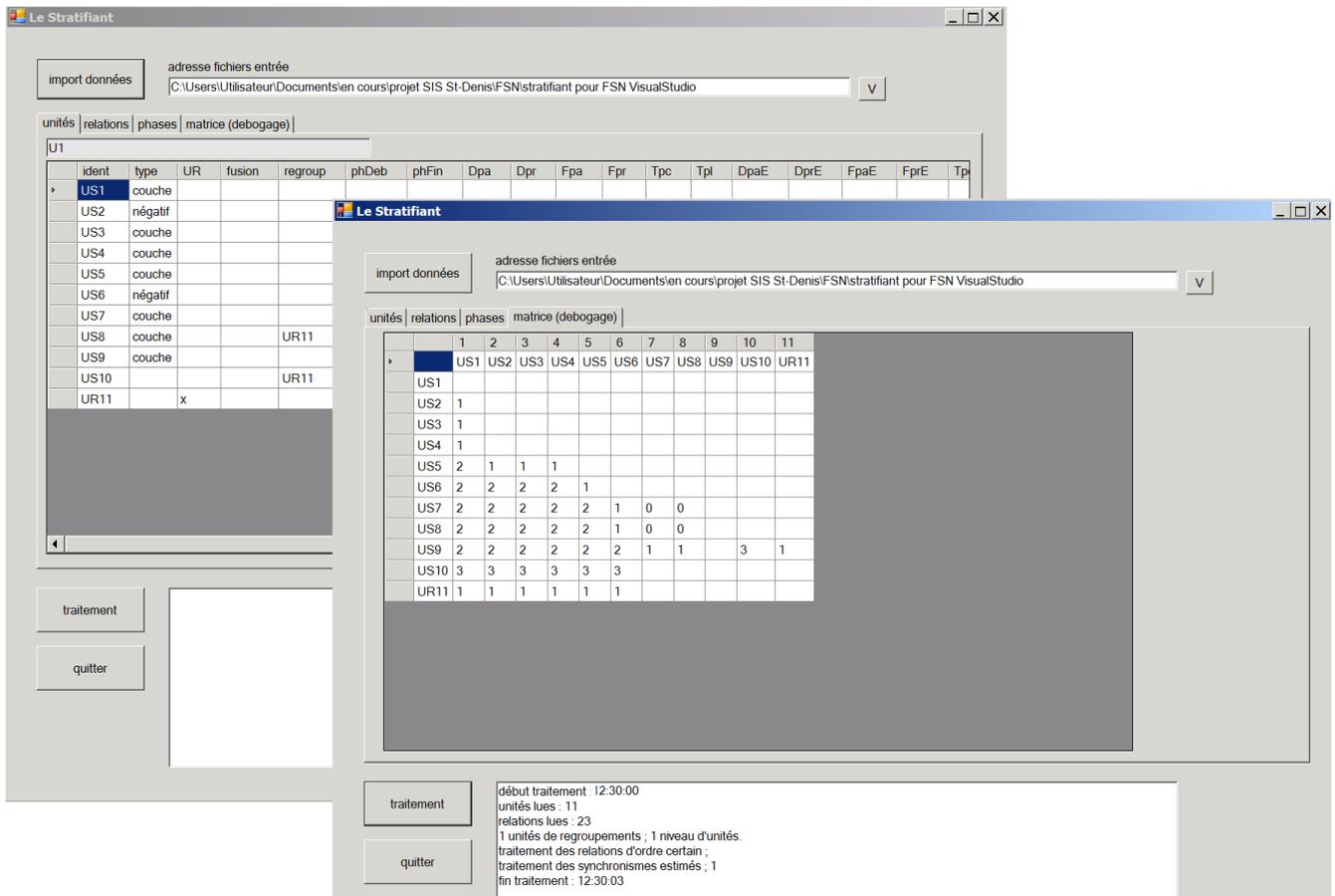
- un historique rapide du développement du *Stratifiant* ; l'algorithme est issu à l'origine d'un travail réalisé en 1989 dans le cadre du séminaire de Fr. Djindjian exploitant l'idée apportée par ce dernier de traiter une matrice d'adjacence codée avec les relations stratigraphiques enregistrées ;
- un rappel des principales caractéristiques techniques du *Stratifiant* actuel, créé en 2008 dans le cadre d'une thèse. Cette version qui se présente sous la forme d'un additif au logiciel Microsoft Excel (programmé en VBA) est utilisée par différents archéologues au sein d'institutions

diverses ; elle a pu bénéficier de retours d'expérience variés. Après avoir reçu de nombreuses corrections, elle est aujourd'hui globalement stabilisée.

- Un aperçu des travaux en cours. En effet, au delà de l'actuelle version, un triple objectif d'améliorations pratiques répondant à des demandes d'utilisateurs (par exemple gestion plus souple des regroupements stratigraphiques actuellement limité à une mise en phase générale), de recherche méthodologique sur la formalisation des concepts mobilisés par les archéologues dans le raisonnement chronologique de terrain, et d'évolution technique vers une solution informatique la plus libre possible, ont conduit à entamer le développement d'une nouvelle version aux fonctionnalités étendues.

Cet aperçu a concerné :

- les choix pratiques de développement d'une part ; les objectifs indiqués ci-dessus impliquaient de ne pas rester sous Excel, et de ne pas se contenter de « rustines » rajoutées à l'actuelle application. Une reprise en profondeur – impliquant le développement d'une nouvelle application – est nécessaire. Pour des raisons de continuité de principe (intégrer le traitement des données stratigraphique dans un environnement familier, et profiter de toutes les fonctions d'un logiciel bureautique déjà existant) l'interface principale reste un tableur, à savoir Calc LibreOffice (concurrent libre d'Excel, de plus en plus répandu). Le langage de programmation interne propre à LibreOffice (OpenBasic) est utilisé en fin de traitement pour le tracé automatique des diagrammes stratigraphiques et des graphes de chronologie quantifiée (dans le module de dessin Draw de LibreOffice), mais est malheureusement trop peu performant pour assurer le traitement principal (calculs sur matrices d'adjacence). Le recours à Python aurait été la solution la plus universelle, la plus cohérente et la plus élégante ; mais, sans renoncer à terme à cette solution, il a été choisi, pour de basses raisons d'utilisation des acquis et de confort de programmation, de recourir à un module exécutable développé en VB.net qui sera appelé depuis LibreOffice ;
- les fonctionnalités d'autre part ; les évolutions les plus importantes sont :
 - le traitement des regroupements stratigraphiques avec la possibilité de créer de UR (unités de regroupement) incluant d'autres unités, sans limitation de niveaux de regroupements (il s'agit en d'autres termes de l'introduction de relations d'inclusion dans l'ensemble des unités ordonnées par les relations d'ordre et d'équivalence stratigraphiques) ; et la possibilité d'obtenir des graphes de synthèse en fusionnant les UR choisies ;
 - la prise en compte étendue de la gestion des intervalles d'inscription des unités dans le temps quantifié (actuellement testée avec l'application expérimentale *Chronographe* sur laquelle on reviendra dans une séance ultérieure)
- par ailleurs le principe robuste d'une communication avec différentes bases de données d'enregistrement via des fichiers d'échange (pour l'instant en format CSV) a été conservé.



Module de traitement (état actuel) : importation des unités et traitement de la matrice des relations

La discussion (prolongée par quelques courriels) a porté notamment sur les choix de traitement des unités de regroupement (UR) ; en particulier :

- les modalités d'héritage de relations d'antéro-postériorité entre unités de regroupement et unités incluses dans ces UR. Celles sont liées à un double choix :
 - la possibilité de saisir des relations à tous les niveaux de regroupements. Il s'agit de s'adapter aux cas (en grand décapage notamment) où l'on est amené à enregistrer d'abord un ensemble stratigraphique (par ex. un fossé et son remplissage) et à enregistrer des relations à ce niveau (par ex. cet ensemble recoupe visiblement une fosse – notons dans ce cas que si l'on identifie un recouvrement en surface, cela signifie logiquement que l'on est déjà sûr qu'il s'agit bien d'un fossé creusé, et dès lors il faut distinguer le négatif du creusement du fossé de son remplissage, ce dernier restant jusqu'à plus d'information une unité globale) – et ensuite seulement à procéder à une observation stratigraphique plus fine (une coupe dans le fossé par exemple) permettant de détailler sa stratification ;
 - la possibilité d'afficher le graphe à différents niveaux de détail : soit au plus petit niveau d'US enregistré (auquel cas les unités élémentaires contenues dans les UR héritent des relations éventuellement enregistrées au niveau de ces UR), soit en fusionnant les unités contenues dans une UR pour n'afficher que le niveau de cette UR dans le graphe (auquel cas, à l'inverse, l'UR hérite des relations enregistrées au niveau de ses unités incluses) ;
- les modalités du contrôle de cohérence entre indications de regroupement et relations stratigraphiques ont été aussi été discutées. Deux situations peuvent être détectées :

- des indications de regroupement contradictoires avec la stratigraphie enregistrée ; exemple : pour les relations x sous y sous z , on a x et z inclus dans une UR A , et y (en position stratigraphique intermédiaire) inclus dans une autre UR B (elle-même non incluse dans l'UR A). Ces données entraînent une intrication des UR A et B , donc une faute de logique stratigraphique, A et B devenant à la fois postérieure et antérieure l'une à l'autre ;
- des indications de regroupement logiquement incomplètes par rapport à la stratigraphie enregistrée ; exemple : pour les relations x sous y sous z , on a x et z inclus dans une UR A , mais y non indiqué comme inclus dans cette UR A ;

Le contrôle de cohérence des données saisies est actuellement prévu pour laisser le maximum de responsabilités à l'utilisateur, dans l'idée que c'est l'utilisateur (et non pas la machine) qui doit prendre, consciemment, les décisions d'enregistrement. Dans le 1er cas (contradiction logique) il donc est prévu un arrêt du programme avec marquage de la faute ; et dans le second cas (incomplétude logique), il est prévu un arrêt du programme avec proposition (à confirmer par l'utilisateur avant relance du traitement) d'attribution de l'unité z à l'UR A .

La possibilité d'une application moins « psycho-rigide » de ce parti-pris de responsabilisation de l'utilisateur a été évoquée en discussion ; ce qui pourrait se traduire par exemple par une réattribution automatique de y à l'UR A dans le second cas (avec signalisation de cette réattribution), voire de fusion automatique des UR A et B en intrication dans le premier cas (là encore avec signalisation).

suite et développements postérieurs

Le développement se poursuit depuis janvier 2016 dans un cadre plus stable que précédemment : celui du programme Archéologie du Bassin Parisien (UMR 7041) ; avec en perspective une première version bêta permettant des retours d'expérience. Parallèlement, un développement dans un but d'intégration au projet FSN, expérimenté à l'unité d'archéologie de Saint-Denis, va être effectué (aussi en VB.net) par l'équipe de la SOGETI chargée du projet FSN et conduite par R. Demaille, avec des échanges d'information entre les deux développements.

références

Desachy 1989

DESACHY B., « Aide au traitement des données stratigraphiques des sites archéologiques très stratifiés : automatisation des diagrammes de Harris », in DJINDJIAN F. (éd.), *séminaire « informatique et mathématiques appliquées en archéologie » - années 1987-1988, 1988-1989 - rapport d'activités (Paris I - Paris X - Ecole Normale Supérieure)*, s.l. : reprographié, pp. 250-281.

Desachy 2005

DESACHY B., « Du temps ordonné au temps quantifié: application d'outils mathématiques au modèle d'analyse stratigraphique d'Edward Harris », *Bull. la Société préhistorique française* [en ligne], 102, 4, pp. 729-740, URL : http://www.persee.fr/web/revues/home/prescript/article/bspf_0249-7638_2005_num_102_4_13176.

Desachy 2008

DESACHY B., « Le Stratifiant, un outil de traitement des données stratigraphiques », *Archeol. e Calc.* [en ligne], 19, pp. 187-194, URL : http://soi.cnr.it/archcalc/indice/PDF19/15_Desachy.pdf.

Desachy 2010

DESACHY B., « Le Stratifiant : a simple tool for processing stratigraphic data », in BÖRNER W., UHLIRZ S. (éd.), *Proceedings of the 14th International Congress « Cultural Heritage and New Technologies » Held in Vienna, Austria November 2009* [en ligne], Vienne : Museen der Stadt Wien, pp. 687-699, URL : http://www.chnt.at/wp-content/uploads/eBook_WS14_Part5_Poster.pdf.

Desachy 2012a

DESACHY B., « Formaliser le raisonnement chronologique et son incertitude en archéologie de terrain », *CyberGeo Eur. J. Geogr.* [en ligne], 597, URL : <http://cybergegeo.revues.org/25233> [lien valide au 28 février 2013].

Desachy, Djindjian 1990

DESACHY B., DJINDJIAN F., « Sur l'aide au traitement des données stratigraphiques des sites archéologiques », *Hist. Mes.* [en ligne], 5, 1, pp. 51-88, URL : http://www.persee.fr/doc/hism_0982-1783_1990_num_5_1_1365.

Desachy, Djindjian 1991

DESACHY B., DJINDJIAN F., « Matrix Processing of Stratigraphic Graphs: a New Method », in RAHTZ S., LOCKYEAR K. (éd.), *CAA90. Computer Applications and Quantitative Methods in Archaeology 1990* [en ligne], Oxford : Tempus Reparatum, pp. 29-37, URL : http://proceedings.caaconference.org/files/1990/06_Desachy_Djindjian_CAA_1990.pdf.

séance 3 samedi 19/03/2016

9h30-12h30 IAA 3, rue Michelet 75006 Paris salle 319

présents : François Capron, Pablo Ciezar, Robert Demaille, Bruno Desachy, Julie Gravier, Jérôme Haquet, Ingrid Renault, Emilie Roux, Luc Sanson, Christophe Tuffery

actualité et informations diverses

recherches en cours :

Emilie Roux signale un projet en cours d'étude des caves et sous-sols d'Orléans (SICAVOR), basé sur une base de données *Access* couplée à un système d'information géographique, et mené en collaboration avec le laboratoire Archéologie et Territoires (UMR Citeres) ;

colloques :

JIAP 2016

Les journées informatiques et archéologiques de Paris (JIAP) se dérouleront du 8 au 10 juin. Ces journées ont lieu tous les deux ans ; l'édition 2014 avait laissé la place au CAA tenu à Paris. La date de clôture d'appels à contributions (originellement 21 mars) a été prolongée jusqu'au 10 avril. Plusieurs communications vont être proposées concernant des projets discutés dans le cadre de l'atelier SITraDA. Les formulaires d'inscription et renseignements sont accessibles sur le [site des JIAP 2016](#).

publications :

À signaler une publication récente :

Barcelo J. A., Bogdanovic I. (éd.), *Mathematics and Archaeology*, CRC Press, 2015, 514 p. ; ce recueil préfacé par J. Doran comprend 27 contributions (dont une de F. Fjindjian sur l'histoire des mathématiques en archéologie) concernant les différents aspects actuels de l'application de méthodes mathématiques à notre discipline

présentation de projet : traitement des indicateurs de datation et inscription dans le temps quantifié des unités stratigraphiques et des entités fonctionnelles (B. Desachy)

intitulé du projet dans lequel s'inscrit la présentation

développement de l'outil de traitement de données stratigraphiques et chronologiques Le Stratifiant (cadre actuel du projet : UMR 7041 ArScAn, programme Archéologie du Bassin Parisien)

type de projet :

projet de recherche professionnel

porteur(s) / auteur(s) du projet :

thématiques de l'atelier concernées par la présentation

1 (méthodes et outils d'enregistrement et de traitement des données de terrain)

précédente présentation du projet dans l'atelier :

- séance précédente (20/02/2016) pour le développement du *Stratifiant* ;
- 18 octobre 2014 (présentation de l'outil *Chronophage*) ;

contenu de la présentation et de la discussion

L'exposé présentait un état de la recherche de formalisation de l'inscription dans le temps quantifié des unités stratigraphiques. Cette recherche, liée au développement de l'outil *Le Stratifiant*, vise à rendre totalement explicite (donc informatisable), le plus simplement et le plus complètement possible, le passage du temps relatif de succession au temps quantifié. En effet, dans la pratique, cette étape basique du raisonnement chronologique de terrain reste souvent en partie ambiguë et implicite. Une fois explicité et formalisé, ce processus pourrait s'appliquer à toute succession d'étapes chronologiques à inscrire dans la chronologie « absolue » en ne disposant que d'une information imparfaite (lacunaire et/ou incertaine).

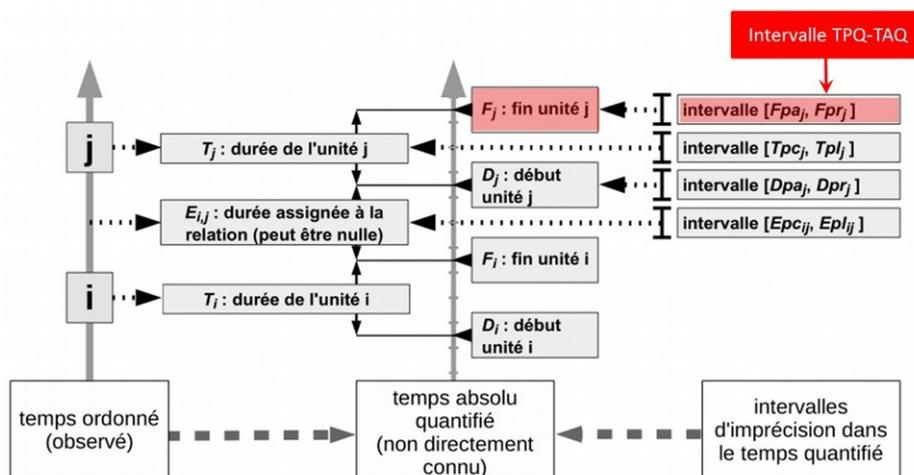


Figure 1

Les acquis actuels, présentés dans des publications récente (Desachy 2015) ou à paraître, peuvent être résumés par les points suivants :

- la localisation dans le temps quantifié d'une unité stratigraphique (et d'une façon générale de toute unité archéologique observée en chronologie relative) se fait traditionnellement au moyen d'un intervalle d'imprécision [date au plus ancien (ou *terminus post quem* ou date plancher), date au plus récent (ou *terminus ante quem* ou date plafond)]. Les bornes de cet intervalle d'imprécision sont fournies par les indications de datation, quelle que soit leur nature (physico-chimique, typologique, historique, etc.) ; à noter que dans certains cas favorables mais rares, la datation de l'événement cherché est directement connue de sorte que le *terminus post quem* se confond avec le *terminus ante quem* et que l'intervalle d'imprécision est nul (par exemple la destruction de Pompéi) ;
- cependant cet unique intervalle [TPQ, TAQ] peut être trop simplificateur. En effet il ne tient pas compte des durées, et réduit donc l'unité de chronologie relative considérée à un seul point ou instant dans le temps. Cela peut être gênant dans le raisonnement chronologique si la durée (d'existence ou de formation) de l'unité est importante ;

- la Recherche Opérationnelle (mathématiques – en particulier théorie des graphes – appliquées à la production industrielle) a défini dans les années 1950 le cadre de temps quantifié applicable aux tâches successives d'un processus de production (et pas conséquent à toute chronologie relative discrétisée en unités finies liées par des contraintes de succession – comme la stratigraphie). Quatre variables élémentaires suffisent : le début D_i d'une tâche i , sa fin F_i , sa durée T_i (temps d'exécution), et la durée T_{ij} affectée à la contrainte de succession entre deux tâches successives i et j (temps de séchage intermédiaire si i est la peinture de la carrosserie et j la pose du pare brise dans une chaîne de montage de voitures par exemple). Ce cadre simple est bien sûr transposable aux unités et relations stratigraphiques (la durée T_i correspondant au temps de formation de l'unité i , et la durée T_{ij} au temps représenté par la relation stratigraphique lorsque les unités i et j ne se succèdent pas immédiatement) ;
- on doit cependant prendre en compte l'imperfection inhérente à l'information archéologique, c'est à dire le caractère lacunaire et incomplet (dans la plupart des cas) de l'ensemble des indications de datation quantifiée disponibles. Autour de ces 4 variables chronologiques, ce n'est donc pas un seul, mais 4 intervalles d'imprécision qui sont nécessaires (figure 1) : [début au plus ancien Dpa_i , début au plus récent Dpr_i], [fin au plus ancien Fpa_i , fin au plus récent Fpr_i], [durée au plus court Tpc_i , durée au plus long Tpl_i] pour chaque US i ; et [durée au plus court Epc_{ij} , durée au plus long Epl_{ij}] pour « l'entretemps » séparant une US i de l'US postérieure j . Rappelons que l'intervalle classique $[TPQ, TAQ]$ correspond à l'intervalle $[Fpa, Fpr]$, le *terminus post quem* donné par le mobilier contenu dans une US donnant une date au plus ancien pour la fin de la formation de cette US (et non son début) ; ce système d'intervalles ainsi que les contraintes de succession (ici les relations stratigraphiques) peuvent être exprimés et traités comme un système d'inéquations, permettant de déduire et préciser les bornes de tous les intervalles de la succession étudiée à partir des bornes renseignées. On ne revient pas ici sur ce traitement algébrique (simple, mais assez fastidieux...) présenté par ailleurs (*cf.* références) et en partie implémenté dans l'outil Chronophage (disponible avec son mode d'emploi en téléchargement sur la page de l'atelier SITraDA) ;
- il en résulte, pour chaque unité de chronologie relative, des plages de temps de statut différent, « certain » ou « possible », l'imprécision des intervalles engendrant de l'incertitude :
 - la formation de chaque US est contenu dans un « temps possible », entre son début au plus ancien et sa fin au plus récent (fig 2 – a). Si l'imprécision de datation est telle que le début au plus récent n'est pas antérieur à la fin au plus ancien, ces deux bornes (Dpr et Fpa) n'apportent rien à la connaissance de la position chronologique de l'US, qui reste incertaine à l'intérieur du temps possible : à chaque instant de ce temps possible, la formation de cette unité peut ne pas avoir commencé ou a déjà pu s'achever. La seule certitude est négative : avant le début au plus ancien et après la fin au plus récent s'étend le « temps impossible » dans lequel l'unité ne peut s'être formée ;

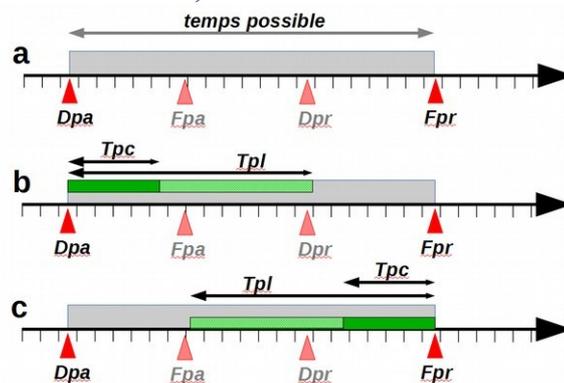


Figure 2

- la connaissance de durées au plus court (T_{pc}) et au plus long (T_{pl}) permet dans ce cas de bâtir des hypothèses chronologiques au plus ancien (fig 2 – b) et au plus récent (fig 2 – c) ;
- dans le cas favorable d'imprécision moindre, où le début au plus récent D_{pr} est antérieur à la fin au plus ancien F_{pa} , alors ces deux bornes définissent une plage de temps « certain », car il est certain que l'unité est en formation durant ce laps de temps (fig 3). Toutefois, le début de la certitude n'est pas la certitude du début : entre le début au plus ancien et le début au plus récent peut subsister une plage de temps possible incertain où l'US peut avoir déjà, ou non, débuté sa formation ; et la fin de la certitude n'est pas la certitude de la fin : entre la fin au plus ancien et la fin au plus récent peut de même subsister une plage incertaine (temps possible) où la formation de l'US peut être achevée, ou non. Le cas optimal de précision totale, où les intervalles d'imprécision sont nuls et où le temps possible est entièrement absorbé par le temps certain, est on l'a dit rarissime en archéologie ;

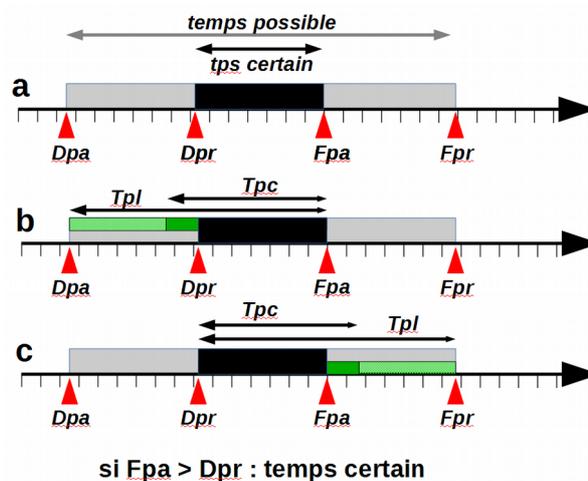


Figure 3

- dans ce cas favorable d'imprécision moindre, les hypothèses au plus ancien (fig 3 – b) et au plus récent (fig 3 – c) mobilisant les intervalles de durées sont calées en fonction du temps certain (celui-ci pouvant d'ailleurs, inversement, être déduit du chevauchement des positions au plus ancien et au plus récent d'une durée au plus court connue et suffisamment longue : c'est l'un des résultats possibles du système d'inéquations évoqué plus haut) ;

l'exposé est revenu plus particulièrement sur deux points actuellement en cours de réflexion et d'approfondissement :

- premier point : la dualité de la temporalité des vestiges matériels *in situ*. Le raisonnement chronologique de terrain considère le temps stratigraphique, exclusivement lié à la formation des traces et vestiges observés ; mais il envisage aussi une temporalité plus large, celle des vestiges considérés comme entités fonctionnelles historiques, incluant leur usage postérieur à leur formation. Par exemple, un mur, en tant qu'unité stratigraphique, ne revoie qu'au temps de construction de ce mur ; l'unité stratigraphique sera ainsi antérieure à la construction d'un sol voire de plusieurs sols successifs s'appuyant sur ce mur (cf. figure 4). En revanche, ce mur (et le bâtiment auquel il appartient) connaît un temps d'usage, donc d'existence historique, qui s'étend jusqu'à sa destruction ou son abandon. Ce temps « historique » (ou temps total d'usage) des vestiges *in situ* est assimilable à la position « en contexte systémique » d'un objet (mobilier ou immobilier) définie par M. Schiffer (objet en interaction avec une société par laquelle il est d'une manière ou d'une autre utilisé) par opposition à la position en contexte archéologique (objet culturellement mort, n'ayant plus aucune interaction avec une société vivante)

(Schiffer 1972). Il a déjà été noté que le processus d'inscription dans la chronologie quantifiée par intervalles d'imprécision, rappelé ci-dessus, peut s'appliquer à ces deux temporalités (stratigraphique et historique), à condition de lier la durée T_i d'une unité stratigraphique i à son strict temps de formation, et de lier la durée T_j d'un vestige (ou ensemble de vestiges) j considéré comme entité historique à son temps d'usage fonctionnel (Desachy 2015). Ce cadre quantifié peut aussi être appliqué aux entités historiques formellement définies par une fonction, un lieu et un temps (et non nécessairement liées à des vestiges matériels observés archéologiquement), utilisées comme unités d'analyse de l'occupation de l'espace à l'échelle d'une carte archéologique ou d'une synthèse urbaine, ainsi qu'aux éventuelles contraintes de succession reliant ces entités (Gravier 2015) Cette distinction entre temps stratigraphique et temps historique amène à soumettre plusieurs propositions à la discussion :

1. cette différence de temporalité entre temps stratigraphique (de formation) et temps historique (d'usage) existe implicitement dans la pratique des fouilleurs : en effet elle est sous-jacente à certaines distinctions notionnelles que l'on trouve dans différents systèmes d'enregistrement :
 - entre US dites « d'occupation » dont le temps de formation se confond avec un temps d'usage (par exemple le dépôt de surface formé par la circulation sur un sol), et unités de construction ou plus généralement d'aménagement ; dans le premier cas le temps stratigraphique coïncide avec le temps historique (d'usage) ; dans le second, le temps d'usage de la structure aménagée excède le temps de formation de l'US correspondante (cf exemple du mur ci-dessus) ;
 - entre les US et leurs regroupements chronologiques (« séquences », « phases », « périodes ») d'une part, et des regroupements structurels et fonctionnels (« faits », « structures »...) d'autre part ; ces deux ordres de regroupements synthétiques, illustrant la double vision diachronique de l'évolution du site (la « fabrique ») et synchronique de l'usage de l'espace à un moment donné (le « fonctionnement »), apparaissent dans la première (et seule) tentative de synthèse des systèmes d'enregistrement stratigraphique en France (CNAU 1987), et sont utilisés depuis de façon plus ou moins explicite. Ces deux ordres de regroupement correspondent en fait à cette double temporalité : la vision diachronique (mise en séquences, phases, périodes) est celle du temps stratigraphique et de la formation du site, la vision synchronique (faits, structures...) est celle du temps historique d'usage ;
2. il existe une différence fondamentale entre les temporalités stratigraphique et historique : la première exclut les cycles (la même unité ne se forme jamais deux fois : un cycle dans une succession stratigraphique est nécessairement une faute logique), la seconde les admet. En effet M. Schiffer souligne le caractère possiblement cyclique du séjour des objets en contexte systémique (un objet peut passer en contexte archéologique, puis réintégrer à nouveau le contexte systémique : ainsi un objet archéologique trouvé en fouille et intégrant les collections d'un musée) ; des phénomènes cycliques apparaissent de même dans la temporalité d'usage des vestiges *in situ*, après leur formation, car le même aménagement matériel peut resservir dans plusieurs états structurels successifs, entraînant (du point de vue de la durée d'usage) une position à la fois antérieure et postérieure à celle d'unités à durée d'usage plus courte (cf. de nouveau le mur 3 par rapport à la succession de sols construits sur la coupe de la figure 4). Cela a pour conséquence que si l'on peut formaliser les regroupements strictement stratigraphiques et chronologiques par un simple relation d'inclusion 1-n (par exemple : une US ne peut appartenir qu'à une seule séquence, qui ne peut appartenir qu'à une seule phase, qui ne peut appartenir qu'à une seule période), la relation entre cette temporalité stratigraphique et les regroupements structurels est nécessairement complexe (n-n), la même unité stratigraphique (plus exactement sa matérialité) pouvant être « recrutée » par plusieurs états structurels successifs. C'est bien

cette différence de temporalité entre « fabrique » et « fonctionnement », entre formation du terrain et usage de l'espace, bref entre temps stratigraphique et temps historique, qui explique le besoin pratique de cette relation n-n entre US et « faits », déjà gérée par quelques systèmes.

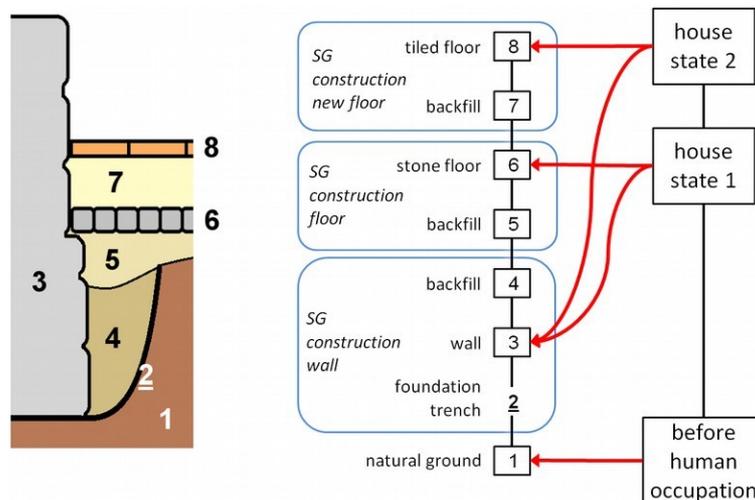


Figure 4 : relation n-n entre US et regroupements stratigraphiques (temps de la formation du terrain) (à gauche) et regroupements structurels (temps de l'usage de l'espace) (à droite)

3. Temporalité stratigraphique (de la formation du terrain) et temporalité historique (de l'usage de l'espace) correspondent à deux temps successifs du raisonnement chronologique de terrain. En effet, ce qui est accessible à l'archéologue, c'est d'abord le temps stratigraphique, observé puis quantifié. Le temps d'usage historique est lui induit, c'est une reconstruction intellectuelle qui intervient dans une deuxième étape, à partir (mais pas seulement) de la stratigraphie ; de ce point de vue, il est à noter que la dialectique chronologie relative / chronologie « absolue » quantifiée, à laquelle on ramène souvent le raisonnement chronologique de terrain, masque cette distinction temporalité stratigraphique / temporalité historique, pourtant au fond plus importante ; de même l'assimilation du temps stratigraphique à la seule chronologie relative est inadéquate, le temps stratigraphique pouvant s'inscrire dans le temps quantifié distinctement du temps historique (durées de formation et non d'usage) ;
4. il serait utile, du point de vue de cette double temporalité et de l'articulation majeure qu'elle représente dans le raisonnement chronologique de terrain, de clarifier et d'expliciter les pratiques d'enregistrement et le discours de terrain :
 - en précisant si les unités manipulées dans l'enregistrement et le discours relèvent de la temporalité stratigraphique (US et regroupements d'US), ou de la temporalité historique d'usage de l'espace ; comme on l'a vu, à la première catégorie se rattache clairement la notion de « séquence » (dans les enregistrements qui l'utilisent) ; à la deuxième, formant ce que l'on appellera ici des « entités matérielles historiques », semblent se rattacher plus fréquemment les termes (déjà débattus dans l'atelier) de « faits » ou « structures » (les fouilleurs reprenant ainsi la distinction reconnue dans le document du Cnau de 1987) ;
 - il est à noter qu'une ambiguïté spécifique s'attache à ces termes « faits » ou « structures » lorsqu'ils sont utilisés, en contexte de grand décapage et de site peu densément stratifié, pour enregistrer des phénomènes stratigraphiques d'abord perçus

globalement en plan (bassins sédimentaires artificiels ou non, détecté par contraste avec le terrain antérieur qui les contient), et ensuite (éventuellement) analysés plus finement par unités stratigraphiques constitutives ; ces « structures » directement observées sont bien des regroupements stratigraphiques, inscrites dans le temps de formation du terrain ; les confondre avec les entités matérielles historiques (inscrites dans le temps d'usage de l'espace) auxquelles elles sont reliées, peut créer des problèmes chronologiques car cette relation est on l'a dit complexe, et doit être explicitée (ainsi un « trou de poteau » observé renvoie, via le temps de formation de ses US constitutives, à l'abandon, ou à la construction d'une structure dont l'utilisation marque un usage de l'espace, mais pas directement à ce temps d'usage lui-même ; celui-ci doit être inféré).

5. Dernière proposition plus générale (et quelque peu provocatrice) : cette distinction de ces deux temporalités, et la nécessité pour l'archéologue d'élucider d'abord le temps stratigraphique pour ensuite inférer le temps d'usage de l'espace, m'amène à considérer que le traitement du temps stratigraphique est indépendant de, et préalable à, celui des données spatiales. L'espace (plus exactement, le volume sédimentaire du site) n'intervient, au départ, que comme gisement d'où le temps stratigraphique est extrait ; puis il n'intervient ensuite qu'au passage du temps stratigraphique au temps historique, quand il faut effectivement inférer les usages de l'espace après en avoir mis en ordre les composantes chronologiques ; c'est à dire passer de l'élucidation de la diachronie stratigraphique, à la vision de fonctionnement synchrone, à un instant donné (et à une collection d'instant successifs), qui relève du temps historique. Conséquence pratique : c'est à cette vision synchrone de l'usage de l'espace que me semble adapté un SIG, bien plus qu'au temps stratigraphique. Le SIG n'est donc à mes yeux pas un outil central de l'enregistrement archéologique (mais c'en est un pour la synthèse historique issue de cet enregistrement) ;
- le deuxième point de réflexion concerne l'imperfection de l'information chronologique en archéologie, incluant l'incertitude ci-dessus évoquée provenant de l'imprécision des indications de temps quantifiée, mais aussi l'incertitude affectant parfois les données de chronologie relative ou quantifiée ; cette dernière due aux difficultés d'observation de terrain ainsi qu'à l'hétérogénéité qualitative de la documentation disponible (historique ou archéologique). L'originalité du *Stratifiant* – et la justification de son développement par rapport à l'existence préalable d'outils de même type – est de gérer ce statut d'incertitude (dont l'attribution est de la responsabilité du chercheur), au moyen d'une logique à deux modalités : « certain » et « estimé » (= incertain). Cette forme simple de logique modale s'applique au cadre de chronologie quantifiée présenté ci-dessus, les bornes d'intervalles d'imprécision pouvant alors être « certaines » ou « estimées » (ainsi dans l'outil *Chronophage*). Au delà, il existe une perspective de passer de cette logique modale réduite à trois modalités (vrai, faux, incertain), à l'univers continu de la logique floue ; nous tentons ici un premier aperçu de l'application possible de cette voie à notre système d'intervalles d'imprécision :
 - la logique floue (plus exactement théorie des sous-ensembles flous) a été proposée en 1965 par le mathématicien Lofti Zadeh, comme une évolution de la théorie des ensembles admettant un degré d'appartenance variable à un ensemble. Dans cette logique, la valeur de vérité n'est donc pas binaire (vrai/faux) comme en logique classique, ni discrétisée en modalités finies comme en logique modale, mais graduée dans un intervalle réel de 0 à 1 (ou de 0 à 100%) (Cohen 2004). Elle est à la base des « systèmes experts » informatiques développés à partir des années 1980. Les applications au raisonnement archéologique sont peu nombreuses ; au premier rang se trouvent les travaux pionniers de l'équipe de Jean-Claude Gardin dès les années 1980 (Gardin et al. 1987) et plus récemment, concernant la localisation spatiale et chronologique, les travaux de Cyril de Runz (Runz 2008) ;

- dans cette approche appliquée à notre système d'intervalles, la position chronologique d'une unité n'est plus représentée de façon linéaire sur l'axe gradué du temps (cf. figure 2 et 3 ci-dessus), mais sur un plan défini par l'axe du temps en abscisse et la valeur de vérité en ordonnée. Dans le cas d'une unité munie d'un temps certain et d'un temps possible (figure 5), cette valeur est logiquement 0 avant le début au plus ancien (Dpa), puis 1 à partir du début au plus récent (Dpr) jusqu'à la fin au plus ancien (Fpa), puis de nouveau 0 après la fin au plus récent (Fpr). On peut ainsi positionner les 4 points ABCD correspondant à ces bornes d'intervalle sur le plan défini par le temps et la valeur de vérité, et tracer la ligne joignant successivement ces points. Cette ligne limite un polygone de vérité chronologique (le trapèze gris sur la figure 5) pour l'unité considérée. En d'autres termes, on passe de valeurs discrètes « vrai » (0) ou « faux » (1) à une fonction continue de vérité $v(t)$ qui prend ses valeurs entre 0 et 1 (ou 0 et 100%), qui donne la valeur de vérité v correspondant à chaque instant t pour l'unité considérée, et dont l'intégrale sur l'intervalle [A,D] (début au plus ancien – fin au plus récent) donne la zone de vérité chronologique de cette unité ;

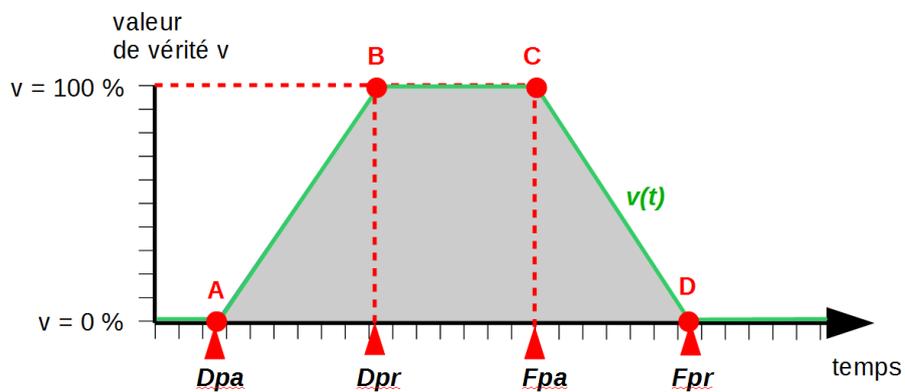


Figure 5

- dans le cas défavorable d'une forte imprécision telle que le début au plus récent d'une unité est postérieur à sa fin au plus ancien (cf. fig. 3 ci-dessus), les points sont cotés de la même façon sur l'axe de vérité, mais l'interversion du début au plus récent et de la fin au plus ancien sur l'axe du temps réduit géométriquement le polygone de vérité de l'unité considérée à un triangle limité par une fonction de vérité qui n'atteint jamais 100 % (fig.6) : on est bien dans le temps possible mais jamais certain, décrit plus haut, qui caractérise ce cas de grande imprécision. Plus le début au plus récent est postérieur à la fin au plus ancien, plus le triangle s'aplatit et la valeur maximale de vérité pour cette unité baisse : traduction géométrique du fait énoncé ci dessus que plus d'imprécision engendre plus d'incertitude ;

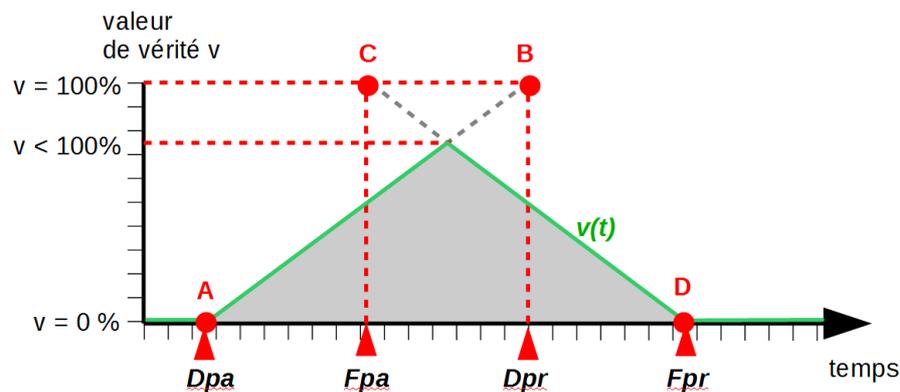


Figure 6

- il faut insister sur le fait qu'accepter ainsi l'hypothèse du continu liée à la logique floue, c'est à dire accepter la valeur de vérité déduite de cette fonction $v(t)$ en n'importe quel point du temps, y compris en dehors des bornes d'intervalles renseignées, n'est pas un mince choix intellectuel de la part de l'archéologue. Par exemple, dans le cas de la figure 5, cela revient à accepter une existence vraisemblable à 50 % de l'entité considérée à l'instant médian entre le début connu au plus ancien et le début connu au plus récent. On voit bien qu'une fonction de vérité chronologique aussi simpliste et linéaire ne saurait être acceptée par les archéologues fins et subtils que nous sommes ! C'est là qu'intervient la possibilité de modifier cette fonction en introduisant des bornes d'intervalles dotées d'une certaine valeur de vérité entre 0 et 100 %. Ainsi le mode logique « estimé » (incertain), intermédiaire entre le vrai et le faux, que nous avons admis pour le traitement des relations stratigraphiques et pour les intervalles d'imprécision dans le temps quantifié, peut être simplement pris en compte avec une valeur de vérité de 50 %, avec pour résultat un polygone et une fonction de vérité déjà moins simplistes (fig. 7) ;

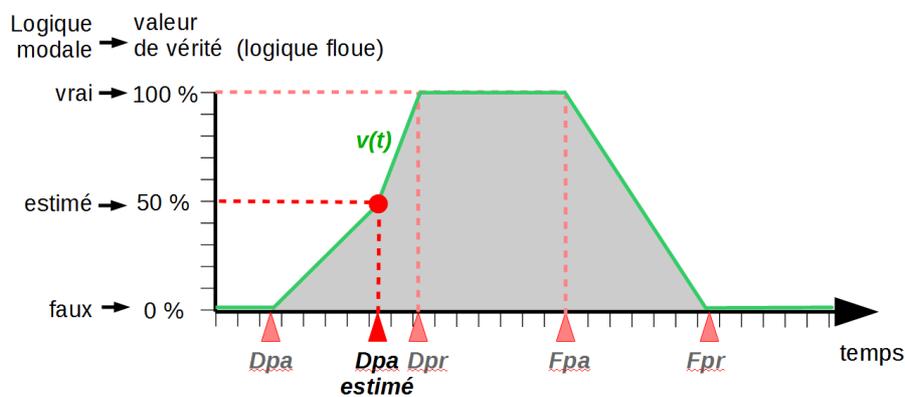


Figure 7

- Au delà, le passage de la logique modale à la logique floue permet de saisir des valeurs de vérité fixées par expertise, c'est à dire par une argumentation proprement archéologique convertie en estimation chiffrée de vraisemblance (c'est la voie ouverte pour l'archéologie par Jean-Claude Gardin il y a une trentaine d'années). Dans notre système d'intervalles chronologiques, cela correspond à la possibilité de saisir autant de points que l'on veut pour chaque borne, chacun correspondant à une valeur de vérité et un instant différents (le début de la formation de l'US considérée – ou de l'existence fonctionnelle de l'entité historique considérée – est vraisemblable à tant pour cent à telle date, à tant + n pour cent à telle autre date, etc.) ;

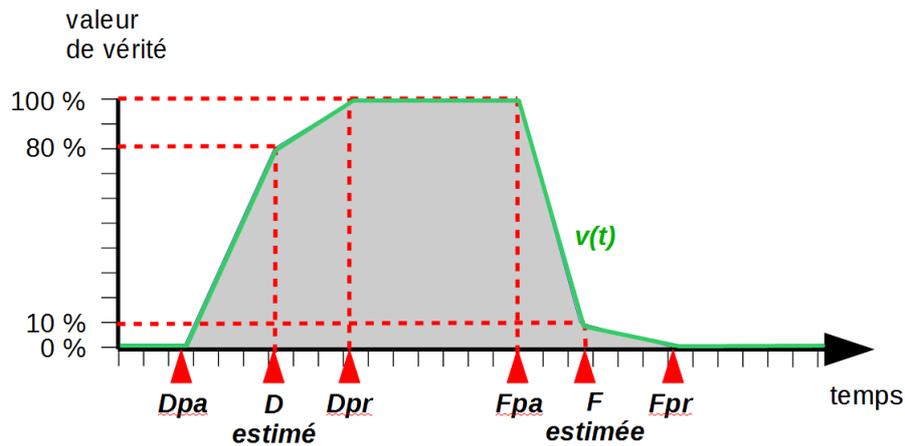


Figure 8

- en amont de la représentation graphique et géométrique ici abordée, d'autres aspects sont à développer ; en particulier l'intégration de ces valeurs de vérité saisies comme coefficients dans le traitement du système d'inéquations, permettant de calculer les valeurs de vérités des bornes déduites. L'idée générale dans laquelle s'inscrit cet aperçu est celle de la recherche d'une meilleure prise en compte des nuances de l'expertise et de l'estimation archéologiques dans l'utilisation des systèmes d'information, nécessairement formalisés ; permettant, par exemple, des requêtes chrono-spatiales sur SIG en fonction de ces nuances, de type « à telle date, quel est l'usage de l'espace connu au dessus du seuil de vraisemblance de n %? »

Les éléments de discussion pendant et après la présentation ont porté notamment sur les rapports du temps et de l'espace archéologiques, débat qui bénéficiera de l'apport de la présentation de C. Tuffery le 25 juin, et pour lequel l'importance des concepts de Fernand Braudel sur le temps long et le temps cyclique (Braudel 1958) a été souligné par L. Sanson ; ainsi que sur les questions pratiques de compatibilité avec les modèles de données d'enregistrement existants : ainsi la relations n-n entre les « faits » et les US qui existe déjà dans le système FSN est parfaitement compatible avec l'approche proposée (qui distingue les « faits » structurels des US et de leurs regroupements en assimilant les premiers à la temporalité historique d'usage) ; cette approche suppose en outre de ne pas autoriser la saisie de relations stratigraphiques sur des « faits » ainsi entendus.

suite et développements postérieurs

les débouchés pratiques sont envisagés en deux temps :

- l'intégration du traitement complet des intervalles d'imprécision (avec modalités certain/estimé) – actuellement en partie implémenté dans l'outil *Chronophage* - dans la version en cours de développement du Stratifiant ;
- à plus long terme, intégration de fonctions de logique floue (saisie et traitement de valeurs de vérité) ;

références

outre les références indiquées pour la présentation de projet de la séance précédente du 19/03/2016 :

Braudel 1958

BRAUDEL F., « Histoire et Sciences sociales: La longue durée », *Ann. ESC* [en ligne], 13, 4, pp. 725-753, URL : http://www.persee.fr/doc/ahess_0395-2649_1958_num_13_4_2781.

CNAU 1987

CNAU, *Enregistrements des données de fouilles urbaines* [en ligne], Tours : Centre national d'archéologie urbaine, URL : http://www.culturecommunication.gouv.fr/var/culture/storage/pub/enregistrements_des_donnees_de_fouilles_urbaines_1e_partie/files/docs/all.pdf.

Cohen 2004

COHEN G. (éd.), *La logique : le vrai, le faux et l'incertain*, Paris : Tangente, Hors-série n°15.

Desachy 2015

DESACHY B., « A Simple Way to Formalize the Dating of Stratigraphic Units », in GILIGNY F., DJINDJIAN F., COSTA L., MOSCATI P., ROBERT S. (éd.), *CAA 2014, 21st Century Archaeology, Concepts, Methods and Tools. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology* [en ligne], Oxford : Archaeopress, pp. 365-370, URL : <http://www.archaeopress.com/ArchaeopressShop/Public/displayProductDetail.asp?id=%7BE35F9954-5653-493D-884B-4A7D2DE66610%7D>.

Gardin et al. 1987

GARDIN J.-C., GUILLAUME O., HERMAN P. Q., HESNARD A., LAGRANGE M.-S., RENAUD M., ZADORA-RIO É., *Systèmes experts et sciences humaines : le cas de l'archéologie*, Paris : Eyrolles.

Gravier 2015

GRAVIER J., « Recognizing Temporalities in Urban Units from a Functional Approach: Three Case Studies », in GILIGNY F., DJINDJIAN F., COSTA L., MOSCATI P., ROBERT S. (éd.), *CAA 2014, 21st Century Archaeology, Concepts, Methods and Tools. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology* [en ligne], Oxford : Archaeopress, pp. 371-380, URL : <http://www.archaeopress.com/ArchaeopressShop/Public/displayProductDetail.asp?id=%7BE35F9954-5653-493D-884B-4A7D2DE66610%7D>.

Runz 2008

RUNZ C. de, *Imperfection, temps et espace : modélisation, analyse et visualisation dans un SIG archéologique* [en ligne], (s.l., Université de Reims Champagne-Ardenne, Thèse de doctorat sous la direction de M. Herbin et F. Piantoni), URL : <https://tel.archives-ouvertes.fr/tel-00560668/document>.

Schiffer 1972

SCHIFFER M., « Archaeological Context and Systemic Context », *Am. Antiq.*, 37, 2, pp. 156-165.

séance 4 samedi 23/04/2016

9h30-12h30 INHA salle Jullian

présents : Pablo Ciezar, Robert Demaille, Bruno Desachy, Fabienne Dugast, Julie Gravier, Jérôme Haquet, Léa Hermenault, Ingrid Renault, Luc Sanson, Christophe Tuffery

actualité et informations diverses

recherches en cours :

Fabienne Dugast et Ingrid Renault signalent le projet en cours sur la Vallée de l'Eure : [Cité des Carnutes – prospections dans la Vallée de l'Eure](#) dans le cadre de l'UMR 8167.

colloques :

Creuser au Mésolithique / Digging in the Mesolithic

le [colloque « Creuser au Mésolithique »](#) s'est tenu à Châlons-en-Champagne du 29 au 30 mars 2016 avec une communication inaugurale sur l'analyse morphologique, stratigraphique et la distribution spatiale des fosses mésolithiques : *Pits by hundreds, a renewed picture of the Mesolithic in Champagne - Analysis and mapping of an unexpected phenomenon* (Des fosses par centaines, une nouvelle vision du Mésolithique en Champagne - Analyse et cartographie d'un phénomène insoupçonné) par Nathalie Achard-Corompt, Emmanuel Ghesquière, Christophe Laurelut, Charlotte Leduc, Arnaud Rémy, Isabelle Richard, Vincent Riquier, Luc Sanson (participant à l'atelier), Julia Wattez.

JIAP 2016

Rappel : les journées informatiques et archéologiques de Paris (JIAP) se dérouleront du 8 au 10 juin. Ces journées ont lieu tous les deux ans ; l'édition 2014 avait laissé la place au CAA tenu à Paris. Les formulaires d'inscription et renseignements sont accessibles sur le [site des JIAP 2016](#).

formations et outils :

À signaler le démarrage du MOOC : « Échanges et proximité : la première loi de la géographie » présenté par l'Université Paris I : <https://www.fun-mooc.fr/courses/Paris1/16004/session01/about>

présentation de projet : Interroger la taille des villes sur le temps long : premiers éléments de discussion (J. Gravier)

intitulé du projet dans lequel s'inscrit la présentation

Caractérisation spatio-temporelle de la ville sur le temps long. Approche diachronique de l'articulation du système intra-urbain de Noyon et de son système de villes (1er – 21e s.)

type de projet :

thèse en cours sous la direction de Lena Sanders et Nicolas Verdier

porteur(s) / auteur(s) du projet :

Julie Gravier UMR 8504 Géographie-Cités

thématiques de l'atelier concernées par la présentation

2 (synthèses archéologiques urbaines et territoriales) ; 4 (exploration multidimensionnelle de sources croisées pour l'archéologie)

contenu de la présentation et de la discussion (voir [diaporama](#))

L'objectif de la présentation était de questionner les indicateurs des systèmes de villes étudiés dans la diachronie, depuis le 1^{er} s. apr. J.-C. jusqu'au début du 19^e s. On pose la question de la pertinence des critères surfaciques de l'espace urbain pour aborder la question de la taille des villes grâce à une étude statistique des relations entre surfaces et populations.

1) Contexte de questionnement

La présentation s'inscrit au sein d'un travail de doctorat dont l'objet est la caractérisation spatio-temporelle de la ville sur le temps long. Plus précisément, la question porte sur l'articulation d'une ville avec son système de villes sur deux millénaires. Pour y répondre, l'approche est ainsi diachronique et multi-scalaire et concerne le cas de Noyon dans le nord de la France.

La trajectoire de la ville de Noyon apparaît plutôt particulière au regard des profils de villes du même contexte socio-historique, tel qu'en France (Desachy 1999). Noyon était tout d'abord une agglomération secondaire de la cité des Viromanduoens (Ben Redjeb et al. 1992). Elle a ensuite été le siège de l'Évêché, ce statut étant attesté dans un texte daté du début du 7^e s.¹ (Collart 1984, Collart, Gaillard 2004, Gravier 2012). À la Révolution Noyon a perdu ce statut – au profit de la ville de Beauvais – et n'avait plus qu'un statut local puisqu'elle n'était que chef-lieu de canton². Elle conserve aujourd'hui cette condition cantonale. Au regard des statuts de Noyon au cours du temps de son existence, il semble donc que sa trajectoire n'est pas linéaire. Ainsi il nous apparaît nécessaire d'étudier sa place dans son système de villes pour décrire et comprendre cette trajectoire.

Discussion : La proposition de représentation de la trajectoire de la ville sur une courbe graduée au cours du temps a été discutée (*cf.* support de présentation : <https://cours.univ-paris1.fr/mod/folder/view.php?id=94036>). En particulier, le fait que l'axe des ordonnées n'ait pas de définition précise a posé problème. Il s'agissait initialement de présenter une forme d'ordonnement qualitative. Des pistes ont été proposées, notamment auprès des travaux de J.-Y. Blaise portant sur la visualisation de données spatio-temporelles (*cf.* en particulier : http://www.map.cnrs.fr/jyb/puca/seminaire/jyBlaisePUCA2014_pdf.pdf).

Des éléments à caractère événementiel permettent de questionner la place de la ville dans ses réseaux politiques et économiques. À titre d'exemple, Charlemagne a été couronné à Noyon ; les évêques étaient comtes et pairs de France à partir du 12^e s. Ces éléments, par exemple, permettent d'identifier des liens politiques privilégiés entre le roi et la ville au Moyen Âge. Il semble dès lors que Noyon était relativement importante au regard de son intégration politique. Quant au caractère économique de la place de la ville dans son système, d'autres événements peuvent être mentionnés. On remarque par exemple qu'elle ne figurait pas dans la Hanse des XVII^e au 13^e et 14^e s. – qui associait des drapiers des villes flamandes et du nord de la France – tandis que l'on y retrouvait dès le début Beauvais et Saint-Quentin, puis d'autres villes du nord telles qu'Abbeville, Amiens, Arras, Douai, Lille, Montreuil, Péronne etc. (Carolus-Barré 1965). On note par ailleurs que de nombreux maires de Noyon étaient des

1 Le caractère particulier de la trajectoire de Noyon est ici clairement identifiable puisque ce sont les chefs-lieux de cités qui deviennent généralement sièges d'Évêchés, notamment dans les Trois Gaules.

2 La ville n'a même pas été retenue en 1800 pour être chef-lieu d'arrondissement parmi les neuf districts qui existaient dans l'Oise. En effet, les quatre chefs-lieux ont été Beauvais, Clermont, Compiègne et Senlis (pour les districts et les arrondissements, voir notamment http://cassini.ehess.fr/cassini/fr/html/8d_glossaire.htm).

tanneurs, et ce, jusqu'au début du 20^e s. Les notables de la ville détenaient donc des ateliers artisanaux n'ayant qu'une zone de chalandise micro-régionale, voire locale.

Le processus de réflexion permettant d'interroger la place d'une ville dans son système de villes sur le temps long est ainsi décomposé en deux phases : 1) rechercher des événements historiques caractérisant des sous-systèmes sociaux d'ordre différents (ici d'ordre politico-administratif et éco-démographique) ; 2) poser l'hypothèse de l'intégration de la ville dans ses réseaux et sa portée selon les sous-systèmes sociaux observés. On passe donc de connaissances ponctuelles à une hypothèse continue dans le temps. Ainsi, au moyen de nombreux événements historiques connus, nous émettons l'hypothèse que la ville de Noyon a été intégrée au sein d'un réseau politico-administratif de portée macro-régionale au Moyen Age et à l'époque moderne, mais que la portée de la ville d'un point de vue économique n'a jamais dépassé celle d'une étendue micro-régionale au cours du temps.

Une fois l'hypothèse posée, on est dès lors confronté à la question de savoir quels indicateurs permettraient de définir la place éco-démographique d'une ville par rapport à d'autres. Pour y répondre, il est possible de se tourner vers les travaux des géographes, qui ont notamment développé la notion de taille de ville depuis la seconde moitié du 20^e s. On peut rapprocher le concept de taille de celui d'une mesure de l'importance des villes à l'intérieur d'un système. Cette taille est généralement évaluée par les géographes grâce à la population recensée, entendue comme un résumé des propriétés fonctionnelles des villes (Pumain 1997).

Si l'on considère que la taille des villes peut être un concept pertinent pour étudier la place éco-démographique de Noyon, on se retrouve face au problème de l'absence de données démographiques pour les périodes anciennes. Ainsi, quels seraient les indicateurs archéologiques des tailles des villes sur le temps long ? Il est en outre nécessaire que ces indicateurs soient pérennes et aient une continuité de sens entre l'Antiquité et l'époque moderne, tandis que ces contextes sociaux sont très différents. Nous proposons dès lors de prendre en considération les surfaces de l'espace urbain qui peuvent être approximées³ depuis des connaissances archéologiques ponctuelles.

Discussion : Lors de la discussion d'autres formes d'indicateurs ont été proposés, en particulier l'idée d'établir un rapport de surface entre l'espace dédié aux activités publiques et l'espace urbain. Ainsi, il serait envisageable de faire des comparaisons de ville à ville par période chronologique.

³ L'approximation est ici à entendre au sens mathématiques d'une représentation imprécise ayant toutefois un lien étroit avec la quantité ou l'objet qu'elle reflète.

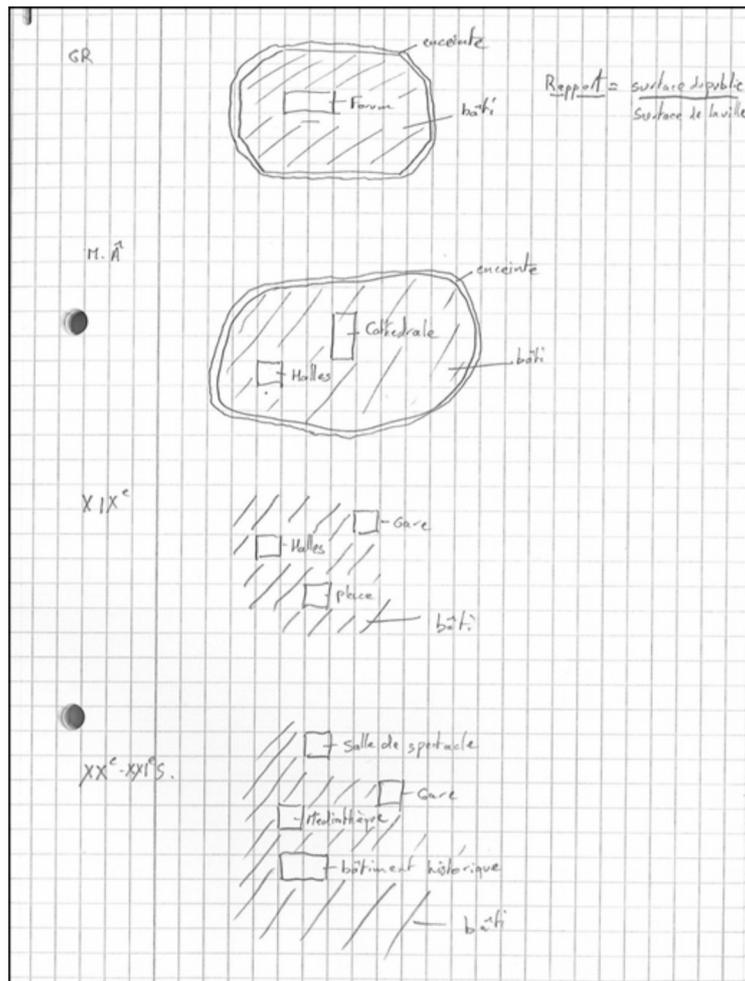


Figure 1 : schéma : rapports surfaciques entre espaces dédiés aux activités publiques et espace urbanisé (Luc Sanson)

2) Les populations des villes déterminent-elles les surfaces densément occupées de l'espace urbain ? Les formes statistiques des relations entre surface et population dans le nord de la France entre la fin du 18e et le début du 19e s.

Il apparaît que la considération selon laquelle les surfaces de l'espace urbain seraient *de facto* relatives aux populations des villes est un postulat très fort. Dès lors, il nous semble nécessaire de discuter les relations entretenues entre les populations et les surfaces. Il est notamment possible de se demander si les populations des villes déterminent les surfaces densément occupées de l'espace urbain au moyen d'explorations statistiques. Plus précisément, il s'agit d'étudier la corrélation entre les deux variables, la variable explicative étant ici la population et la variable à expliquer la surface.

2.1. Définition du corpus étudié

Le corpus de données date de la fin du 18e et du début du 19e s., à partir du moment où les recensements sont exprimés à échelle individuelle et non par foyer (feux), et à un moment où l'on détient des données cartographiques précises et harmonisées à échelle interurbaine. Par ailleurs, nous avons privilégié des données préindustrielles à des données contemporaines qui auraient pourtant été plus simples à recueillir. On se situe en effet avant l'arrivée du chemin de fer et donc précédemment à la profonde réduction des distances-temps liée à l'accroissement des vitesses – ayant notamment entraîné un étalement urbain marqué (Bretagnolle 1999, Rodier et al. 2010). Ainsi, le contexte des relations espace-temps à la fin du 18e et au début du 19e s. est semblable aux périodes anciennes. Les données population sont issues du recensement de 1793 effectué à échelle de la municipalité. Plus

précisément, les données sont extraites de la base de données mise en place par A. Bretagnolle. Elles ont été complétées par les données du site Cassini, développé par le Laboratoire de démographie historique – CNRS/EHESS (*cf.* http://cassini.ehess.fr/cassini/fr/html/10_maj.htm et Motte, Vouloir 2007), pour les villes ayant une population inférieure à 2 000 habitants puisqu'elles n'avaient pas été prises en compte dans la BDD d'A. Bretagnolle.

Les données surface sont fondées sur la carte d'état-major dont les levées topographiques ont été effectuées entre 1818 et 1836 pour l'espace considéré. Les surfaces ont été enregistrées manuellement via le géoportail, avec un repérage des villes au 1/136 000 et un enregistrement au 1/17 000. Le corpus des villes est issu de la définition des lieux donnée par la carte d'état-major elle-même, et comporte au total un semi de 148 villes. *Discussion* : La pertinence du corpus a été discutée quant à la synchronie des données. L'écart possible entre la réalité démographique à un instant T et les surfaces à un instant T+1 pose en effet problème. Il a été proposé de mettre à jour les données démographiques selon des recensements postérieurs. Il serait sans doute plus pertinent de prendre en compte les recensements de 1821 et de 1836. Pour ce faire, il sera nécessaire de récupérer les données manuellement via le site Cassini.

2.2. *Quelle définition morphologique de la ville est-elle la plus pertinente ?*

Dans une étude précédente, réalisée sur les 77 villes françaises évoquées par Charles de Fourcroy dans la table poléométrique (Dainville 1958), nous avons testé plusieurs protocoles d'enregistrement manuel morphologique. Trois manières d'enregistrer les surfaces des villes ont été mises en place : 1) la surface de l'espace intra-urbain dense qui correspond aux îlots bâtis sur la carte d'état-major ; 2) la surface intra-muros correspondant à l'espace à l'intérieur de l'enceinte urbaine ; 3) la surface de l'espace urbain que nous considérons concorder à l'espace intra-muros adjoint des faubourgs.

Les tests de corrélation entre les populations et ces différents types de surface ont permis de considérer que le troisième protocole est le plus pertinent. En effet, les ajustements linéaire et puissance associés à l'espace urbain présentent les meilleurs R^2 , respectivement de 0,83 et 0,74 (c et c' sur la figure 2). Les délimitations morphologiques des 148 villes du corpus relatif au nord de la France ont donc été faites selon le troisième protocole.

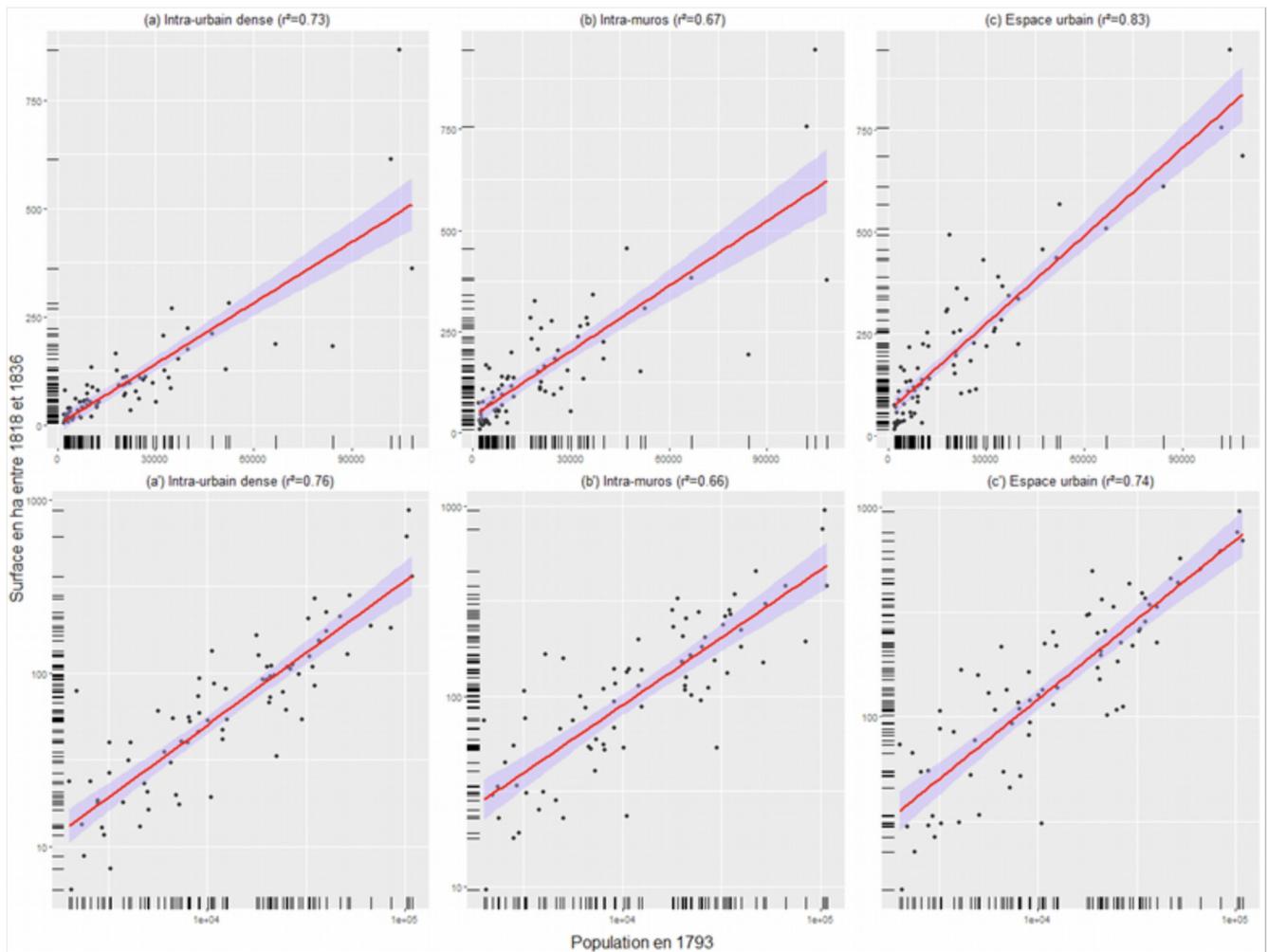


Figure 2 : Ajustements linéaire (haut) et puissance (bas) de la population et la surface des 77 villes de la table poléométrique de Charles de Fourcroy selon trois critères de définition morphologique

2.3. Étude statistique

Nous avons dans un premier temps procédé à un test de corrélation selon une relation linéaire entre population et surface. Les distributions des variables sont particulièrement dissymétriques, ainsi le coefficient de corrélation de Bravais-Pearson (supérieur à 0,98) est très sensible à quelques valeurs extrêmes. Il est notable au regard du nuage de points que la droite de régression sous la forme $y = ax + b$ n'est pas un bon ajustement de la relation entre les variables.

Nous avons dans un second temps procédé à une transformation bi-logarithmique (en \log_{10}) des valeurs des variables afin d'explorer d'autres formes de relations de ces dernières. L'ajustement puissance, écrit sous la forme $y = ax^\beta$ est meilleur au regard du nuage de points, bien que le coefficient de corrélation soit de l'ordre de 0,79 et la qualité de l'ajustement de 0,63 (R^2). Ainsi, la relation entre population et surface n'est pas proportionnelle.

Ces résultats sont en outre comparables avec les travaux de géographes portant sur les lois d'échelle (Pumain 2007, Pumain et al. 2010). L'étude de la valeur de β – inférieure, supérieure ou proche de 1 – permet d'identifier trois types de régimes de relations entre la taille des villes (exprimés par les logarithmes des populations) et les quantités observées (ici, les surfaces). Dans le cas où β est inférieur à 1, comme ici où $\beta = 0,6$, le régime est qualifié de sublinéaire, signifiant qu'il existe une concentration des quantités observées dans les plus petites villes (Finance en cours). Dans le cas présenté, il y a donc une concentration des surfaces dans les plus petites villes. Ainsi, il apparaît que les grandes villes ont tendance à se densifier plutôt qu'à s'étaler, peut-être à cause d'un effet de la distance à partir d'un

certain seuil de surface occupée. Il est par ailleurs intéressant de noter que ce type de relation entre population et surface est également observé par les géographes pour les villes européennes contemporaines (Bretagnolle et al. 2016, Cottineau et al. 2016).

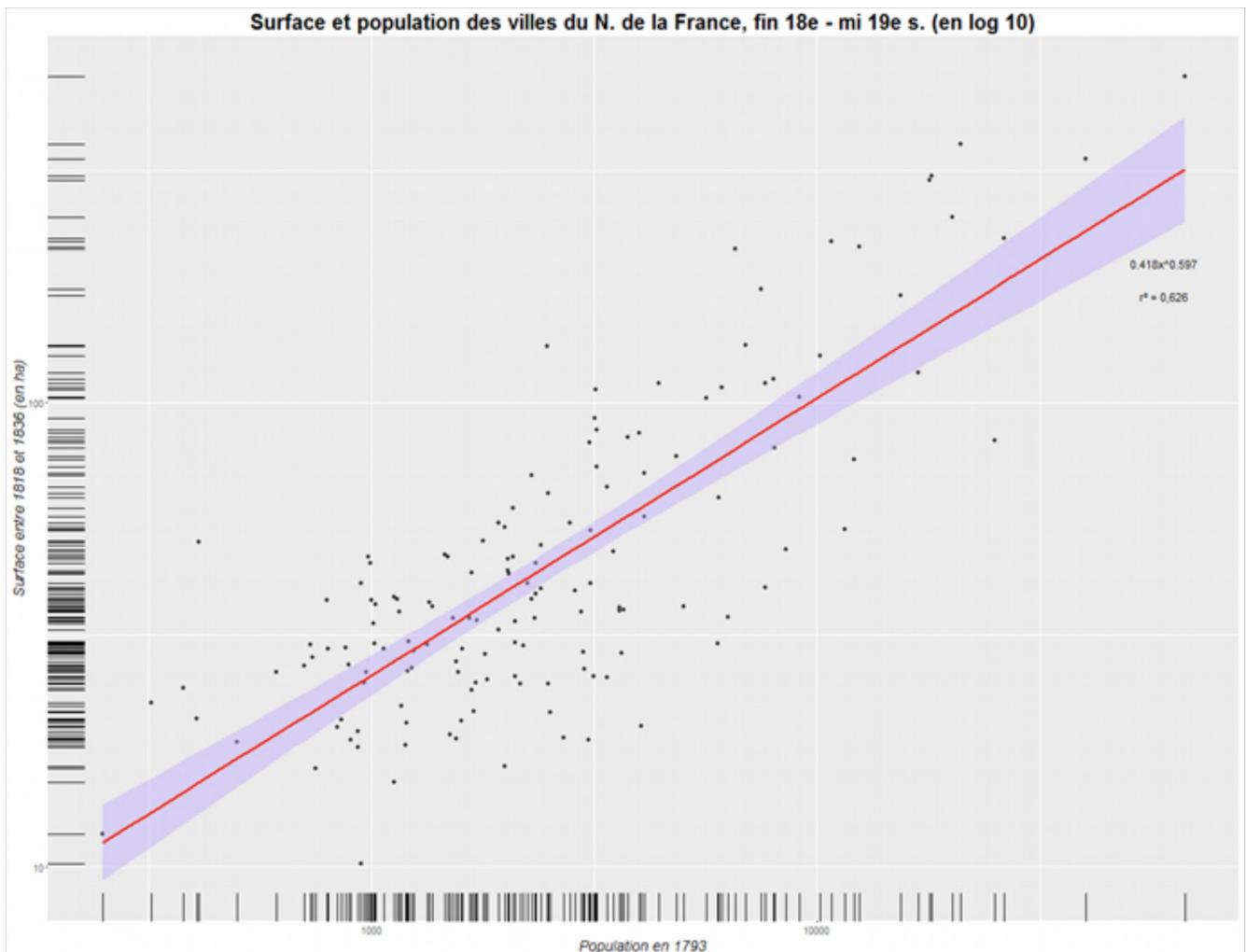


Figure 3 : Transformation bi-logarithmique et ajustement puissance entre population et surface des 148 villes étudiées

Il est enfin possible d'étudier les écarts au modèle, soit les écarts entre chaque ville (point) et la droite d'ajustement puissance – généralement appelés « résidus ». Il est ainsi possible d'étudier la distribution des résidus et leur répartition spatiale. De ce dernier point de vue, nous pouvons observer différentes concentrations des résidus positifs et négatifs.

Notons au préalable que si une ville a un résidu négatif cela signifie que, par rapport au modèle, la surface de la ville est particulièrement restreinte par rapport à sa population – et inversement si elle a un résidu positif. Nous observons tout d'abord qu'un nombre important de villes du nord ont des écarts négatifs le long de l'actuelle frontière avec la Belgique ; ensuite, des villes dans le nord-est, dans l'est – l'Aisne en particulier – et dans le sud ont des résidus positifs ; aussi, nous remarquons que des villes aux résidus négatifs sont présentes dans la région de Dieppe. Il est enfin intéressant de noter que les villes considérées comme importantes pour l'espace considéré ont de forts résidus positifs, telles qu'Abbeville, Amiens, Arras, Beauvais, Douai, Lille et Saint-Quentin. Les causes de ces écarts au modèle sont multiples. Dès lors, il est très compliqué de proposer précisément des facteurs selon l'observation de leur répartition spatiale. Il nous semble que plusieurs pistes pourraient être explorées : des effets de frontière spécifiques au nord (?) ; des causes géomorphologiques, comme dans la région

de Dieppe où les espaces sont très marécageux et où les villes auraient tendance à se densifier plutôt qu'à s'étaler afin d'éviter les coûts importants – en hommes et en temps – de travaux d'assainissement et d'aménagement (?) ; des causes relatives à l'ancienneté et l'importance de certaines villes qui auraient ou aurait eut tendance à s'étaler plus que les autres (?).

suite et développements postérieurs

Discussion : La discussion a permis de mieux définir les perspectives envisageables. Il est ainsi apparu que deux possibilités pouvaient être retenues afin d'étudier les tailles de villes grâce aux surfaces sur le temps long. La première consisterait à transformer les surfaces observées en populations selon la fonction puissance définie dans le cadre du test réalisé. Ainsi, nous obtiendrions des hypothèses de populations des villes, à partir desquelles nous pourrions travailler. La seconde possibilité, plus liée aux données de terrain et que nous souhaitons de ce fait privilégier, consisterait simplement à pouvoir comparer les villes les unes aux autres grâce aux surfaces à différents instants dans le temps puisque nous avons maintenant établi, du moins selon le test réalisé dans un contexte préindustriel, qu'il existe un lien entre les surfaces des villes et leur réalité démographique.

Bien entendu, une fois le travail de collecte des surfaces des villes effectué, il est possible de concrétiser ces deux perspectives et de les mettre en regard l'une de l'autre car une simple opération de transformation est un travail rapide à entreprendre.

Par ailleurs, la discussion a porté sur le statut et la « validité » de la démarche partant de la constatation de non-disponibilité de chiffres fiables de population urbaine pour les périodes les plus anciennes, alors que le critère de surface occupée par l'espace urbain (suivant les trois types de périmètres distingués ci-dessus) est lui accessible au chercheur (ou au moins plus accessible) pour toutes les périodes ; la recherche d'une corrélation menant à une fonction de régression (qui en l'occurrence n'est pas une fonction linéaire simple), permettant d'estimer une population à partir d'une surface est certes un pari, mais l'intégration de cette recherche de régression statistique dans un argumentaire archéologique et historique explicite donne à cette proposition le statut d'une « construction guidée » gardinienne (cf. Gardin 1979).

Suite concrète : un chapitre de la thèse, [contribution à la Journée Informatique et Archéologie de Paris en juin 2016](#).

références

Ben Redjeb et al. 1992

BEN REDJEB T., AMANDRY M., ANGOT J.-P., DESACHY B., TALON M., BAYARD D., COLLART J.-L., FAGNARD J.-P., LAUBENHEIMER F., WOIMANT G. P., « Une agglomération secondaire des Viromanduois : Noyon (Oise) », *Rev. archéologique Picardie* [en ligne], 1, 1-2, pp. 37-74, URL : http://www.persee.fr/web/revues/home/prescript/article/pica_0752-5656_1992_num_1_1_1643.

Bretagnolle 1999

BRETAGNOLLE A., *Les systèmes de villes dans l'espace-temps : effets de l'accroissement des vitesses de déplacement sur la taille et l'espacement des villes*, (s.l., Université Paris I - Panthéon-Sorbonne, Thèse de doctorat sous la direction de D. Pumain).

Bretagnolle et al. 2016

BRETAGNOLLE A., GUÉROIS M., LE NÉCHET F., MATHIAN H., PAVARD A., « La ville à l'échelle de l'Europe. Apports du couplage et de l'expertise de bases de données issues de l'imagerie satellitale », *Rev. Int. géomatique*, 1, pp. 55-77.

Carolus-Barré 1965

CAROLUS-BARRÉ L., « Les XVII villes, une hanse vouée au grand commerce de la draperie », *Comptes-rendus des séances l'Académie des Inscriptions B.-lett.* [en ligne], 109, 1, pp. 20-30, URL : http://www.persee.fr/web/revues/home/prescript/article/crai_0065-0536_1965_num_109_1_11795.

Collart 1984

COLLART J.-L., « Le déplacement du chef lieu des Viromandui au Bas-Empire, de Saint-Quentin à Vermand », *Rev. archéologique Picardie* [en ligne], 3-4, pp. 245-258, URL : http://www.persee.fr/web/revues/home/prescript/article/pica_0752-5656_1984_num_3_1_1446.

Collart, Gaillard 2004

COLLART J.-L., GAILLARD M. collab., « Vermand, Saint-Quentin et Noyon : le chef lieu d'une cité à l'épreuve de la christianisation », in FERDIÈRE A. (éd.), *Capitales éphémères: des capitales de cités perdent leur statut dans l'Antiquité tardive*, Tours : FERACF, pp. 83-102.

Cottineau et al. 2016

COTTINEAU C., HATNA E., ARCAUTE E., BATTY M., « Diverse cities or the systematic paradox of Urban Scaling Laws », *Comput. Environ. Urban Syst.* [en ligne], 59, URL : <http://www.sciencedirect.com/science/article/pii/S0198971516300448>.

Dainville 1958

DAINVILLE F. de, « Grandeur et population des villes au XVIIIe siècle », *Population (Paris)*. [en ligne], 13, 3, pp. 459-480, URL : http://www.persee.fr/doc/pop_0032-4663_1958_num_13_3_5673.

Desachy 1999

DESACHY B., « Noyon », *Rev. archéologique Picardie* [en ligne], spécial 16, 1, pp. 171-177, URL : http://www.persee.fr/issue/pica_1272-6117_1999_hos_16_1.

Finance [sans date]

FINANCE O., *Les villes françaises investies par les capitaux étrangers : des entreprises en réseau aux établissements localisés*, (s.l., Université Paris I - Panthéon-Sorbonne, Thèse de doctorat sous la direction de D. Pumain).

Gardin 1979

GARDIN J.-C., *Une archéologie théorique*, Evreux : Hachette.

Gravier 2012

GRAVIER J., *Synthèse Archéologique Urbaine. Topographie historique de la ville de Noyon du Ier s. apr. J.-C. au début du XXIe s. (Oise)* [en ligne], (s.l., Université Paris I - Panthéon-Sorbonne, mémoire de master recherche sous la direction de J. Burnouf), URL : <http://le-nid-du-stratifiant.ouvaton.org/spip.php?article23>.

Gravier 2015

GRAVIER J., « Recognizing Temporalities in Urban Units from a Functional Approach: Three Case Studies », in GILIGNY F., DJINDJIAN F., COSTA L., MOSCATI P., ROBERT S. (éd.), *CAA 2014, 21st Century Archaeology, Concepts, Methods and Tools. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology* [en ligne], Oxford : Archaeopress, pp. 371-380, URL : <http://www.archaeopress.com/ArchaeopressShop/Public/displayProductDetail.asp?id=%7BE35F9954-5653-493D-884B-4A7D2DE66610%7D>.

Motte, Vouloir 2007

MOTTE C., VOULOIR M.-C., « Le site Cassini.ehess.fr. Un instrument d'observation pour une analyse du peuplement », *Bull. du Com. français Cartogr.* [en ligne], 191, pp. 68-84, URL : <http://www.lecfc.fr/new/articles/191-article-7.pdf>.

Pumain 1997

PUMAIN D., « Pour une théorie évolutive des villes », *Espac. géographique* [en ligne], 26, 2, pp. 119-134, URL : http://www.persee.fr/doc/spgeo_0046-2497_1997_num_26_2_1063.

Pumain 2007

PUMAIN D., « Lois d'échelle et mesure des inégalités en géographie », *Rev. Eur. des Sci. Soc.* [en ligne], XLV, 138, pp. 55-65, URL : <http://ress.revues.org/193?lang=en>.

Pumain et al. 2010

PUMAIN D., PAULUS F., VACCHIANI-MARCUZZO C., LOBO J., « An evolutionary theory for interpreting urban scaling laws », *CyberGeo Eur. J. Geogr.* [en ligne], 343, URL : <http://cybergeog.revues.org/2519>.

Rodier, Saligny 2010

RODIER X., SALIGNY L., « Modélisation des objets historiques selon la fonction, l'espace et le temps pour l'étude des dynamiques urbaines dans la longue durée », *CyberGeo Eur. J. Geogr.* [en ligne], 502, URL : <http://cybergeog.revues.org/pdf/23175> [lien valide au 2 mai 2013].

présentation de projet : reprise des données stratigraphiques d'une fouille ancienne (Orléans – Saint-Pierre-Lentin) (L. Guyot)

intitulé du projet dans lequel s'inscrit la présentation :

Le site de Saint-Pierre-Lentin (Orléans, Loiret) : reprise des données et nouvelle synthèse

type de projet :

mémoire de master 2 (université de Nantes)

porteur(s) / auteur(s) du projet :

Laurine Guyot sous la direction d'Yves Henigfeld (maître de conférences, Université de Nantes) et le tutorat de Sébastien Jesset (SAMO)

thématiques de l'atelier concernées par la présentation

1 (méthodes et outils d'enregistrement et de traitement des données de terrain) ; 2 (synthèses archéologiques urbaines et territoriales)

précédente présentation dans l'atelier du projet :

Année 2014-2015 : samedi 2 mai 2015

contenu de la présentation et de la discussion

Voir [diaporama](#).

Les données faisant ici l'objet d'une reprise sont issues des opérations de sauvetage programmé menées de 1977 à 1979 sur le site orléanais de Saint-Pierre-Lentin. Cette fouille s'inscrit dans le vaste programme des « abords de la cathédrale » qui comprend quatre zones de fouilles toutes situées au nord de la cathédrale à l'exception de Saint-Pierre-Lentin, seule opération au sud de l'édifice. La découverte majeure résultant de l'étude de ce dernier est celle des vestiges d'une église attestée dans les textes à partir du XIII^e siècle sous l'appellation « Sanctus Petrus lactentium ». Sa mise au jour n'avait rien d'inattendu, son emprise demeurant toujours visible dans les plans de l'Ecole Supérieure de Jeunes filles détruite en 1960. Néanmoins la restitution de son plan en tau et son attribution au IX^e siècle, soit bien plus tôt que sa première mention, sont apparus comme des éléments bien plus surprenants. Toutefois les méthodes et le contexte de l'époque, l'état des connaissances en matière de céramiques alto-médiévales notamment, n'avaient pas permis d'en proposer une datation précise fondée sur des arguments suffisamment probants. Un des intérêts de revenir à la source, soit l'enregistrement stratigraphique du site, réside ici dans l'identification des couches de construction de l'église et de ses premières couches d'occupation afin de pouvoir en ré-étudier le mobilier. Cette entreprise gage donc également sur l'accessibilité et la conservation du matériel, céramique pour l'essentiel, pour en proposer une nouvelle interprétation à la lumière des avancées faites en la matière ces dernières décennies.

Ma démarche est celle d'une opération de sauvetage de l'information archéologique quarante ans après son archivage. Ce travail est donc également l'occasion d'un constat sur la pérennité des données

enregistrées sur le terrain. Certaines lacunes sont liées à des défaillances intervenant dès la phase de terrain (erreur d'enregistrement...) mais des facteurs extérieurs ont sans doute plus fortement altéré l'intégrité de ces données. Il s'agit pour l'essentiel des pertes survenues après leur stockage. On pense en premier lieu à celle des diagrammes stratigraphiques qui selon des témoignages oraux se présentaient pourtant sous la forme de documents particulièrement imposants et aux fiches d'enregistrement stratigraphiques de trois des huit zones du site absentes des archives (ont-elles par ailleurs jamais existé ?). Il manque une part importante de la documentation graphique, les minutes de terrain ne se trouvant pas dans les archives ou en proportion extrêmement réduite, seule celles publiées dans le rapport et comportant quelques erreurs ont pu être exploitées pour ce travail. Si on peut supposer que des déménagements, comme les emprunts de diapositives pour des expositions ou des publications, ont dû engendrer une partie de ces disparitions il reste difficile de faire un bilan de la documentation en distinguant ce qui a disparu de ce qui n'a jamais existé...

Les cahiers d'enregistrement ont été les premiers supports exploités. Ces derniers sont composés de l'assemblage des fiches US mises au point pour la fouille en 1977. Ces documents, figurant parmi les tout premiers exemplaires du genre, constituent par ailleurs des pièces historiographiques précieuses. Le traitement de ce contenu a débuté par sa saisie dans la base de données relationnelle CADoc. Cet outil a été conçu par Thomas Guillemard (INRAP) en 2009 et est actuellement utilisé par l'INRAP et le SAMO pour l'enregistrement stratigraphique des opérations préventives. La table des US se présente comme le noyau central autour duquel gravitent toutes les autres tables couvrant tous les aspects du chantier (mobilier, minutes, photographies, faits, ...). La question de l'adaptabilité de la base à ce genre de travail a été soulevée au cours de l'atelier. Pour l'heure l'outil s'est avéré répondre efficacement à des besoins qui sont finalement les mêmes que lors d'un enregistrement stratigraphique classique, la corrélation entre les champs des fiches US de 1977 et ceux de la base de données n'ont notamment pas posé de problème. Néanmoins l'évolution de la base vers une structure plus adaptée à un travail de reprise de données, dont le support n'est plus celui de l'enregistrement sur le terrain pour lequel elle est conçue, pourrait être envisagée avec l'accord de son concepteur. Il a été proposé de rajouter un champ commentaire à la table des relations permettant d'enregistrer la fiabilité de ces dernières et donc la critique de l'enregistrement stratigraphique ainsi qu'une table destinée à l'enregistrement de la documentation archéologique incluant notamment un champs renseignant l'état de conservation.

Un autre intérêt de CADoc est sa liaison avec l'outil de traitement des données stratigraphiques *le stratifiant*, permettant l'export direct des relations et des US depuis la base. L'utilisation du stratifiant a été la deuxième étape faisant directement suite à la saisie des données et soulevant de nouveaux questionnements méthodologiques. Face à l'ampleur des circuits – ensembles d'unités dont les relations sont en contradiction logique – (une trentaine dont deux comprenant chacun mille relations environ) il a fallu passer par une reconstruction manuelle d'une partie de la stratigraphie. Dès lors que des erreurs apparaissaient (fautes logiques, ...) il a fallu déterminer quelles sources retenir, quand cela s'avérait possible, pour les solutionner entre les cahiers d'enregistrement éventuellement accompagnés de croquis, les relevés présents dans le rapport et les photographies. Une fois ces corrections faites dans CADoc un nouvel export a été fait vers le stratifiant en isolant chaque zone. Les diagrammes ainsi obtenus restent toutefois à retravailler, des erreurs qui ne représentent pas de fautes logiques et qui n'ont donc pas été détectées demeurant présentes. Ces derniers manquent également de lisibilité, beaucoup de relations « parasites » et, pour l'essentielle d'entre elles, physiques viennent perturber la cohérence de l'ensemble sans pour autant être erronées. Ces enregistrements de relations physiques sont fréquents voire systématiques dans les cahiers.

Il reste une étape importante avant de pouvoir proposer et exposer de nouveaux résultats : l'étude du mobilier. En raison des difficultés actuelles liées à l'accès au CCE, ce dernier n'a pas encore pu être déplacé. L'état de conservation demeurant la grande inconnue, il est sans doute prématuré de définir un protocole d'étude. Néanmoins il paraît souhaitable de continuer à utiliser CADoc pour cette étape,

des tables dévolues au mobilier existant dans la base et permettant d'enregistrer les lots en les reliant à un numéro d'US.

La méthodologie qui est ici mise en place vient répondre à une problématique initiale relative à l'ancienneté supposée de l'église Saint-Pierre-Lentin. Dans le cadre d'un mémoire et du temps imparti à cet exercice cette finalité ne doit pas être perdue de vue. Toutefois le problème soulevé par la reprise des enregistrements de 1977 et 1978 apparaît finalement comme une question méthodologique dont l'intérêt scientifique égale la problématique historique. Au-delà de la nécessité d'effectuer ce genre de travaux sur des documents souvent mis en péril par un oubli prolongé dans des conditions plus ou moins favorables à leur conservation, le contexte actuel semble également justifier le bien-fondé de ce type d'étude. Dans le cas présent des rapports ont été publiés, ils sont sujet à réinterprétation mais ils existent et sont accompagnés de quelques articles. Mais pour certaines opérations menées au cours de cette décennie et des suivantes aucun rapport ne vient synthétiser les résultats qui ne sont donc accessibles qu'à partir de la donnée brute. Cette documentation abondante, souvent collectée et enregistrée dans des conditions précaires ne permettant par leur étude complète, constitue un gisement offrant sans doute à l'archéologie urbaine d'intéressantes perspectives d'avenir. Il peut être envisager que l'acquisition des connaissances dans des centres urbains de moins en moins investis par l'archéologie puisse en grande partie reposer sur la ré-étude de ces données dormantes. Il y a donc un intérêt certain à poser les questions méthodologiques aboutissant à la définition d'un protocole opérant et efficace à ce genre d'exercice. Les variations sont nombreuses voire infinies d'une fouille à l'autre et doivent être prises en compte. La nature du site, la méthode d'enregistrement ou les conditions de conservation sont autant de facteurs qui rendent chaque cas différent. Même si il s'avère possible de trouver des traits universels inhérents à ces archives, la démarche pour pouvoir être applicable, doit impérativement être modulable.



Fouilles de St-Pierre-Lentin (photo P. Gleize)

Au cours de la discussion qui a suivi l'exposé de ce travail de reprise de données primaires, plusieurs perspectives ont été évoquées :

- d'une façon générale, la reprise de la documentation et du mobilier de fouilles relativement anciennes et en particulier des premières grandes opérations urbaines des années 1980 et 1990, pas toujours totalement étudiées, publiées ni rapportées faute de moyens à l'époque, apparaît

comme une problématique majeure : c'est une nécessité préalable à des travaux de synthèse urbaine, et comme l'a fait remarquer L Guyot, c'est un gisement de données à ne pas négliger dans la mesure où les interventions importantes dans les centres-ville historiques sont désormais plus rares ; un certain nombre de travaux en cours ou envisagés à brève échéance (sur la fouille de la place Clemenceau de Beauvais évoquée par B. Desachy, dans le cadre d'opérations extra métropolitaines évoquées par J. Haquet, et à Orléans donc) s'inscrivent de fait dans cette problématique. Elle s'inscrit aussi en complément des travaux menés par le consortium MASA, mais sans se confondre totalement avec, l'objectif ici n'étant pas en soi la conservation et la gestion des archives de fouilles, mais l'extraction depuis ces archives d'une information archéologique restée non (ou non totalement) exploitée jusqu'à présent, en vue d'un usage actuel. Les pistes d'un PCR, et/ou d'un axe de travail par exemple dans le cadre du programme « Archéologie du Bassin Parisien » subventionné par le MCC, sont à explorer ;

- dans les aspects pratiques de ce type de travail, le traitement des lacunes ou incertitudes de la documentation a été évoqué : les deux niveaux de statut de l'information stratigraphique (« certain » ou « incertain ») utilisés dans l'outil *le stratifiant* pourraient être généralisés pour qualifier les sources documentaires ; on peut ainsi prendre le parti de considérer a priori la documentation comme fiable ; mais si une source documentaire (par exemple des fiches d'US) présente des contradictions manifestes, alors le statut des informations contradictoires pourrait être dégradé en « incertain » ; de sorte que l'incertitude contradictoire ainsi distinguée n'affecte pas les autres indications, et puisse être éventuellement tranchée par un autre document original (par exemple un relevé de terrain) qui lui n'est pas affecté de contradiction et a gardé son statut de source fiable (certaine). D'une façon générale, il paraît utile d'aller vers des procédures d'examen critique du document archéologique, à l'instar de la diplomatie des historiens des textes ;
- dans cet ordre d'idée, dans le prolongement de ce qui est fait sur les données de la fouille de Saint-Pierre-Lentin, il convient d'adapter les outils d'enregistrement actuels (bases de données de terrain) à cette tâche de reprise de données anciennes en distinguant clairement deux strates : les données dans l'état où elles ont été recueillies à l'époque, et les commentaires actuels ;
- la question de l'enquête orale – le complément d'information auprès des participants à la fouille, du moins ceux qui fort heureusement sont toujours parmi nous (ce qui n'est hélas pas le cas de toute l'équipe des fouilles d'Orléans) – se pose dans la même perspective : sans du tout mettre en cause la sincérité ou la mémoire des personnes concernées, un souvenir recueilli aujourd'hui auprès d'un témoin est assimilable à un commentaire actuel, pas à une donnée d'origine...
- l'intérêt épistémologique et historiographique de ce type d'enquêtes (surtout concernant les fouilles d'Orléans, point clé parmi quelques autres – Tours, Saint-Denis... – dans l'arrivée en France des méthodes de l'archéologie urbaine britannique des années 1960-1970), rejoignant là des préoccupations du consortium MASA, a aussi été souligné.

références

Collectif 1983

COLLECTIF, « Archéologie de la ville : Orléans », *Rev. archéologique du Loiret*, 9.

Kisch 1977

KISCH Y. de, *Orléans 3 rue Saint-Pierre-Lentin, sauvetage programmé, : abords de la cathédrale, rapport de fouilles campagne 1977*, Orléans.

Kisch 1978

KISCH Y. de, *Abords de la cathédrale, rapport de fouilles campagne 1978*, Orléans.

séance 5 samedi 21/05/2016

9h30-12h30 IAA 3, rue Michelet 75006 Paris salle 319

présents : Anne Chaillou, Robert Demaille, Bruno Desachy, Lydie Dussol, Julie Gravier, Léa Hermenault, Luc Sanson, Christophe Tuffery

actualité et informations diverses

colloques :

Le programme des journées informatiques et archéologiques de Paris (JIAP) du 7 au 10 juin est maintenant disponible. À signaler plusieurs interventions des participants à l'atelier : une communication de J. Gravier, un workshop de B. Desachy, des posters de Léa Hermenault et de l'équipe du projet FSN.

Boite à outils 1 : AnalyseSHS du PIREH (Pôle Informatique de Recherche et d'Enseignement en Histoire de Paris 1) (J. Gravier)

L'équipe du PIREH, et en particulier Stéphane Lamassé, a développé une nouvelle interface web de l'outil gratuit d'analyse de données AnalyseSHS (URL : <http://analyse.univ-paris1.fr/>). L'objectif de cet outil est triple :

- Permettre l'exploration statistique de données à travers différentes méthodes, telles que les AFC, les ACP, les classifications hiérarchiques...
- ... ceci, grâce à l'utilisation du logiciel-langage R...
- ... en le rendant facilement accessible et utilisable par la communauté des SHS – en particulier des historiens, des archéologues et des épistémologues faisant partie du groupe de chercheurs de Paris 1 qui travaillent avec des statisticiens et des informaticiens dans le projet « Analyse de données » (voir notamment les chroniques de ce projet sur <http://www.openedition.org/9483>). Ainsi l'accessibilité est mise en œuvre au moyen d'une interface web, actuellement très conviviale, et la possibilité d'exporter simplement les données et les visualisations statistiques réalisées⁴.

Nous avons actuellement uniquement testé et présenté le traitement des AFC et des CAH avec AnalyseSHS, car nous n'avons pas eu l'intérêt d'utiliser des ACP et des ACM pour traiter les données de nos travaux.

Notons tout d'abord que l'importation de données a l'avantage de comprendre les formats de tableaux les plus usités (.csv, .ods, .txt et .xls). Concernant le traitement des tableaux de contingence par la méthode de l'AFC, AnalyseSHS permet de visualiser le graphique factoriel, mais également le diagramme des valeurs propres et le tableau de décroissance des facteurs qui lui est associé. Les tableaux des coordonnées, des contributions et des cosinus carré des différents axes factoriels sont également affichables dans des onglets complémentaires (fig. 1). En outre, différents paramétrages de l'AFC sont réalisables et proposés directement à droite du graphique factoriel. Les nouveaux paramétrages proposés par l'interface – d'après ce que nous avons pu observer – regroupent l'ajout d'un titre au graphique factoriel et la réalisation de classifications selon la méthode de Kohonen. Il

⁴ L'export de ces éléments a été facilitée dans la nouvelle interface en donnant accès aux fichiers produits et présents dans un seul répertoire via l'onglet « télécharger les résultats ».

semble que tous les autres paramétrages de l'ancienne interface aient été repris, hormis la possibilité de visualiser la taille des points relatifs aux lignes et aux colonnes sous la forme de cercles proportionnels selon leurs contributions.

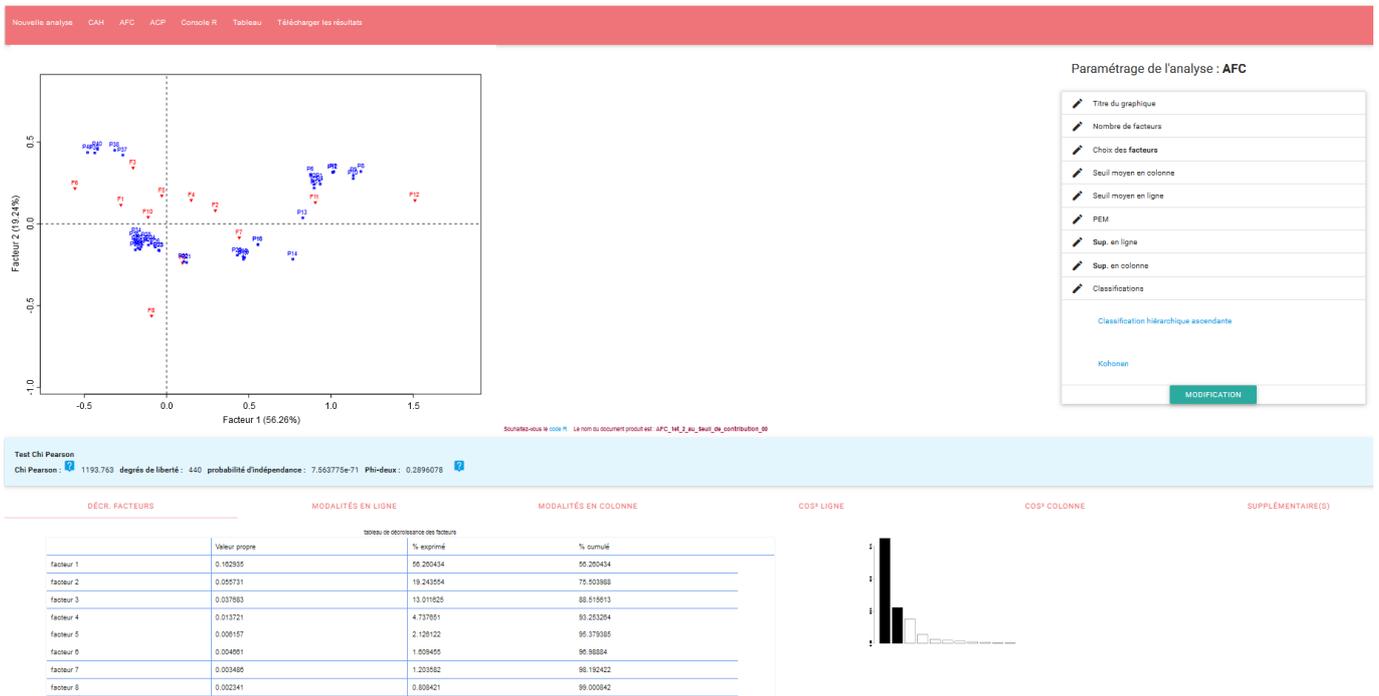


Figure 1 : Interface principale d'AnalyseSHS lors de la réalisation d'une AFC

Classement Hierarchique (CH)

Distance : euclidian

Méthode : ward (inertie)

Indiquez la hauteur à laquelle vous souhaitez couper :

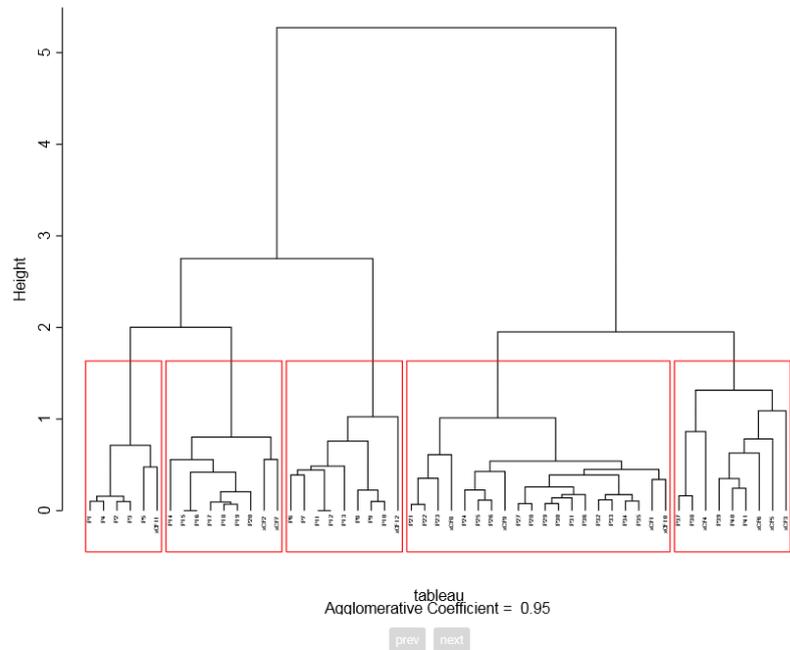
1.5

Nombres de facteurs pour la méthode

10

RÉSULTATS

Classement hiérarchiques des coordonnées



Souhaitez-vous le code R Le nom du document produit est : CHAgnesManhattanWard Télécharger l'ensemble des résultats

Figure 2 : CAH sur les coordonnées des facteurs de l'AFC. Visualisations des classes choisies à une hauteur de 1.5 sous forme d'arbre

Il est intéressant de noter que la visualisation de la CAH sur les coordonnées des facteurs a été améliorée. En particulier, il est dorénavant possible d'observer sur le graphique factoriel les classes obtenues grâce à la CAH (fig. 2 et 3).

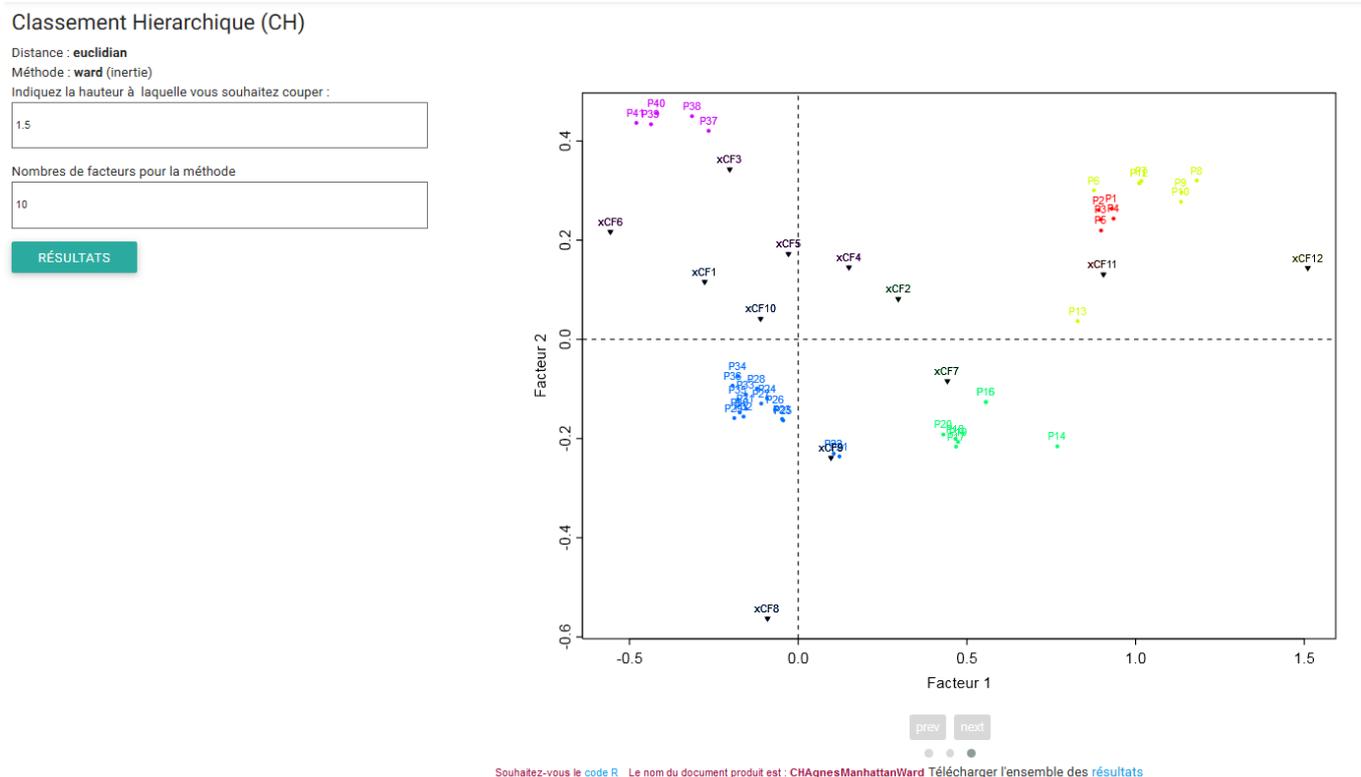


Figure 3 : CAH sur les coordonnées des facteurs de l'AFC. Visualisations des classes choisies à une hauteur de 1.5 sur le graphique factoriel des deux premiers axes (chaque couleur représente une classe)

Par ailleurs, le traitement des tableaux directement en CAH n'était pas faisable au moment de la présentation à SITRADA pour cause de dysfonctionnement interne de l'interface. Grâce à Léa Hermenault qui a informé Stéphane Lamassé de ce bug, les CAH sur les jeux de données initiaux sont maintenant réalisables avec AnalyseSHS.

Boite à outils 2 : Un peu d'analyse spatiale avec R (J. Gravier)

Les éléments d'analyse spatiale évoqués lors de la présentation sont relatifs au chapitre 11 de l'ouvrage *R et espace. Traitement de l'information géographique* écrit par les membres du groupe ElementR (Commenges et al. 2014). Nous avons utilisé ici les packages `sp`, `rgdal`, `maptools`, `RANN` et `spdep`.

R permet de gérer directement des fichiers spatiaux (.shp), aussi simplement que des tableaux. Les appels et les modes de traitements ne diffèrent en effet pas de l'ensemble de la logique sur laquelle est fondée le langage. Ainsi par exemple, la fonction élémentaire `plot()` s'applique à des .shp et permet de les visualiser, telle la figure suivante qui représente les surfaces des villes du nord de la France selon la carte de l'état-major⁵ (figure 1).

Il est en outre possible de récupérer les coordonnées des semis de points et des polygones. À partir des coordonnées, il est ainsi envisageable de faire de l'analyse spatiale en s'intéressant aux positions

⁵ Pour plus d'informations sur le traitement des surfaces de la carte d'état-major, voir la présentation SITRADA « Interroger la taille des villes sur le temps long : premiers éléments de discussion » d'avril 2016.

relatives des lieux les uns par rapport aux autres, et notamment aux distances entre ces derniers. Le calcul des distances aux plus proches voisins selon une distance euclidienne a par exemple été effectué pour en sortir ensuite des histogrammes (fig. 2, en bas).

Dans le cas présenté ici, nous avons comparé les distances aux plus proches voisins de deux semis de villes issus de sources et d'époques différentes, mais associés globalement au même espace. D'une part, nous avons dépouillé *la guide des chemins de France* de Charles Estienne datée de 1552 (Estienne 1552), et d'autre part, nous avons saisi le semis de la carte d'état-major dont les levés pour l'espace considéré datent des années 1818-1836.

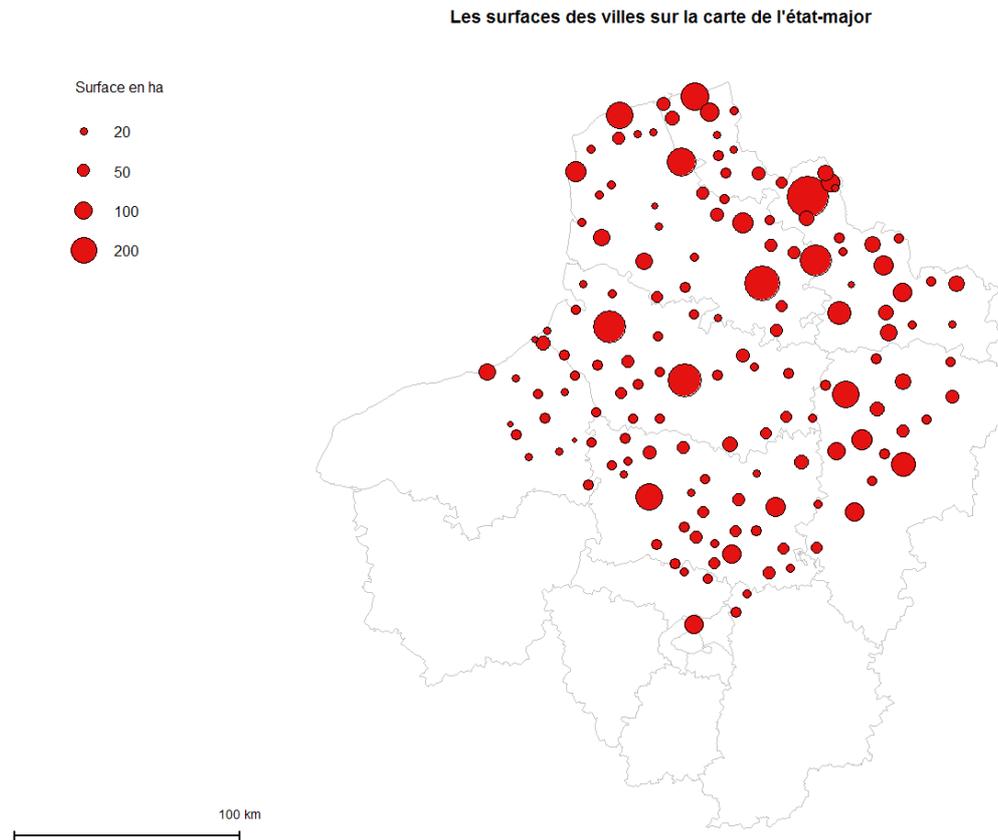


Figure 1 : Les surfaces des villes selon la carte de l'état-major (1818-1836)

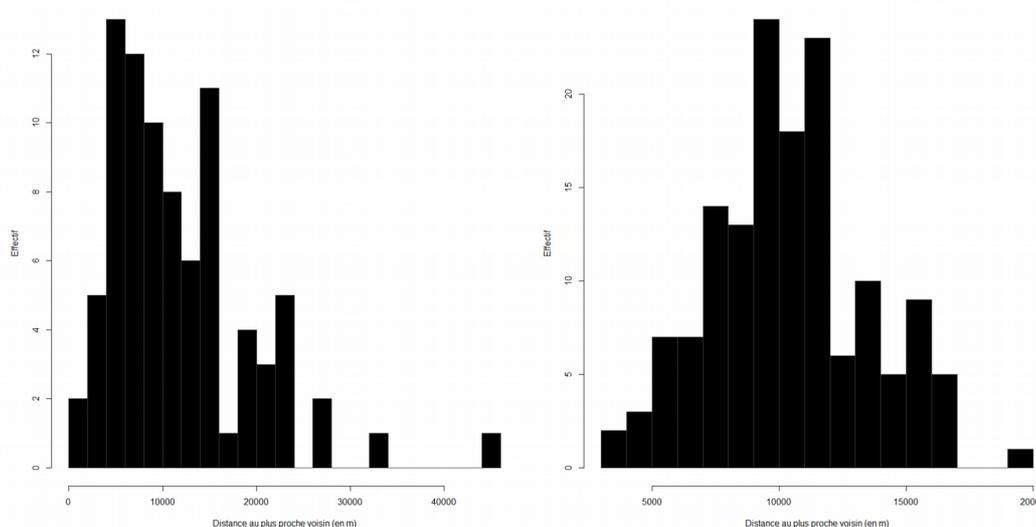
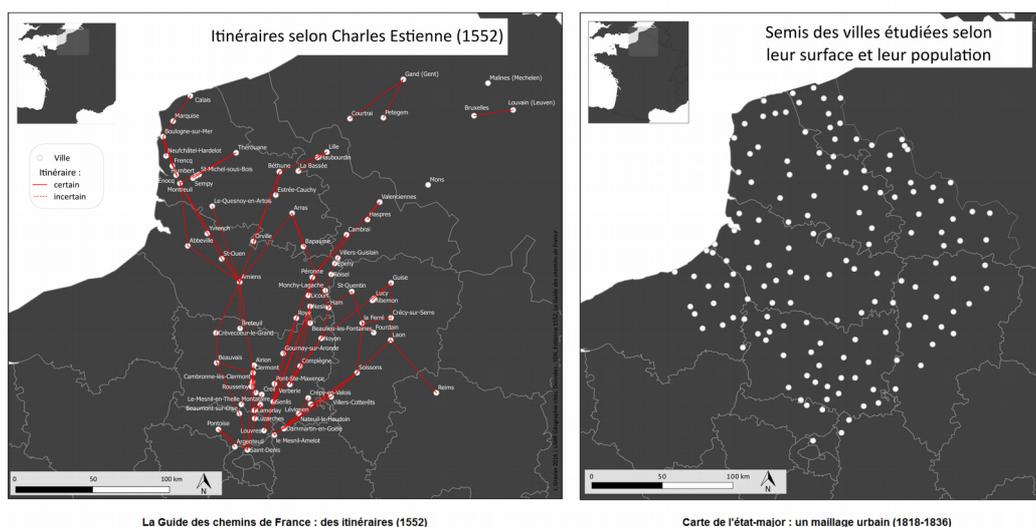


Figure 2 : Comparaison des semis de villes et des distances aux plus proches voisins de la Guide des chemin de France (à gauche) et de la carte de l'état-major (à droite) pour le nord de la France

La première source est un livre évoquant les grands itinéraires de France menant à Paris – sous forme de liste – selon une logique rayonnante depuis la capitale. En effet, l’auteur précise dès son introduction relative à Paris que c’est « sous ceste partie de la France que prenons pour point milieu de la description des chemins d’icelle » (Estienne 1552 : 4). La seconde source est composée d’une série de cartes harmonisées et relativement précises puisque les minutes ont été faites au 1/40 000 pour l’ensemble de la France (Costa, Robert 2008, Plumejeaud, Ruas 2013). L’observation des histogrammes des plus proches voisins permet de mettre en évidence deux manières de (se) représenter l’espace, chacune d’elle étant liée aux objectifs des auteurs. Nous pouvons remarquer pour la guide de Charles Estienne que la distribution est clairement dissymétrique à gauche. En cela, les distances sont majoritairement faibles entre les villes et il existe quelques valeurs extrêmes, telles que Reims ou Mons qui sont très éloignées d’autres villes. L’histogramme met ainsi en exergue une représentation de l’espace sous forme d’itinéraire où l’objectif est avant tout d’évoquer les points stratégiques d’un chemin, que ce soit les points d’étapes qui peuvent être relativement proches les uns des autres et les « hauts-lieux » qui peuvent être relativement éloignés. Quant à l’histogramme associé au semis de la carte d’état-major, nous pouvons observer que la distribution est plus symétrique, avec des distances aux plus proches voisins majoritairement comprises entre 9 et 12 km. Cette forme de la distribution représente ici plus une vision maillée de l’espace.

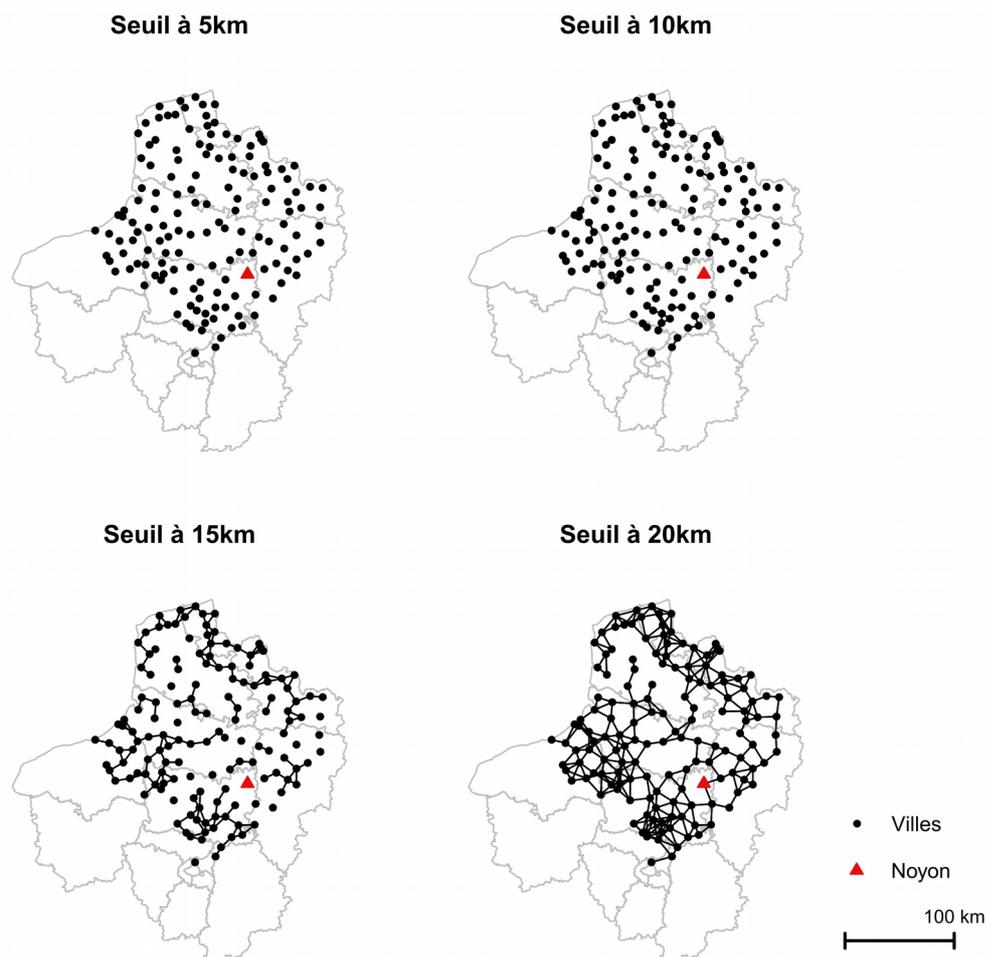


Figure 3 : Graphes de voisinage avec des seuils de distance euclidienne selon le semis de villes de la carte d'état-major (1818-1836)

Il est par ailleurs possible de traiter les données de voisinage sous la forme de graphes avec différents seuils de distance (fig. 3). Dans le cas ci-dessous, si une ville est située à moins de 5/10/15 ou 20 km d'une autre – selon une distance euclidienne – alors il existe un lien entre les deux villes. Ainsi se dessinent des maillages et des réseaux potentiels entre les lieux selon leurs situations dans l'espace géographique. Dans le cas d'un seuil à 15 km nous observons la présence de plusieurs composantes connexes dans le graphe. Trois majeures semblent se dessiner : 1) au nord, le long de la frontière avec la Belgique actuelle ; 2) une à l'ouest depuis Dieppe jusqu'à Amiens ; 3) une dernière au sud aux alentours de Creil et de Senlis. Dans le cas d'un seuil à 20 km, considéré dans la littérature comme l'équivalent d'une journée de voyage aller-retour entre une ville et une autre selon les moyens de transports hippomobiles, nous remarquons que le graphe est connexe. Il est en outre possible de s'intéresser à des lieux en particulier, comme c'est le cas dans cette figure pour la ville de Noyon. Ainsi, il est envisageable de se demander quelle est la place d'une ville dans ses réseaux potentiels, en observant par exemple le nombre de liens que possède un lieu en comparaison avec les autres.

références

Commenges et al. 2014

COMMENGES H. dir., BEAUGUITTE L., BUARD É., CURA R., LE NÉCHET F., LE TEXIER M., MATHIAN H., REY S., *R et espace. Traitement de l'information géographique* [en ligne], Lyon : Framasoft, URL : <http://framabook.org/r-et-espace/>.

Costa, Robert 2008

COSTA L., ROBERT S., *Guide de lecture des cartes anciennes*, Paris : Errance.

Estienne 1552

ESTIENNE C., *La guide des chemins de France* [en ligne], Paris : Charles Estienne, Imprimeur du Roy, URL : <http://gallica.bnf.fr/ark:/12148/bpt6k102662d>.

Plumejeaud, Ruas 2013

PLUMEJEAUD C., RUAS A., *Étude comparative du contenu des cartes de Cassini, d'Etat-Major et de la BDTopo en vue de l'intégration sémantique de l'information*, Paris.

Présentation de projet : traitement statistique de données anthracologiques (L. Dussol)

titre de la présentation

Traitement statistique des assemblages anthracologiques en milieu forestier tropical : réflexions préliminaires sur les problèmes liés à préparation des données.

intitulé du projet dans lequel s'inscrit la présentation

Économie du bois et gestion des forêts à Naachtun (Guatemala), site maya de la période Classique (150-950 apr. J.-C.)

type de projet :

thèse de doctorat en cours sous la direction de Dominique Michelet et de Michelle Elliott

porteur(s) / auteur(s) du projet :

L. Dussol ; doctorante, Paris 1, ED 112, UMR 8096 Archéologie des Amériques ; lydie.dussol@mae.u-paris10.fr

thématiques de l'atelier concernées par la présentation

1 (méthode de traitement des données de terrain) ; en vue de 3 (sériation et typologie des éléments)

contenu de la présentation et de la discussion

Les diagrammes anthracologiques issues de fouilles en milieu forestier tropical présentent souvent un très grand nombre de taxons, conséquence directe de l'extraordinaire biodiversité végétale inhérente à ce type de milieu. Les méthodes de traitement statistique des données anthracologiques couramment employées pour les sites des milieux tempérés et bien développées en Europe (notamment sériographies et analyses multivariées sur des tableaux de contingence) ne peuvent donc s'appliquer directement aux assemblages taxonomiques issues de milieux tropicaux, puisqu'un trop grand nombre de variables (ici les taxons) présentant des effectifs trop faibles (nombre de charbons) rend ce type d'analyse inefficace. Il n'est pas rare, en effet, de devoir traiter plus d'une centaine de taxons, alors que la richesse taxonomique maximale sur un site européen ne dépasse généralement pas trente taxons. Le traitement statistique des assemblages anthracologiques en milieu tropical implique donc un important travail de préparation des données passant par le regroupement des taxons.

Les données présentées lors de la séance du 21/05/2016 proviennent du site maya de Naachtun, cité majeure de la période Classique (AD 150-950) située au nord de l'État du Petén au Guatemala, dans les

Basses Terres centrales. Ces données s'intègrent dans une thèse de doctorat en cours dont l'objectif est de comprendre l'économie globale des bois et les stratégies de gestion des espaces forestiers sur ce site urbain au cours de huit siècles d'occupation. Ceci passe par une reconstitution des usages des bois selon différentes approches : fonctionnelle, sociale, spatiale et diachronique. Jusqu'à présent, l'analyse de plus de 9200 charbons, provenant de 88 contextes archéologiques (représentant les individus) a produit un spectre anthracologique très riche et d'autant plus problématique, puisque 8 taxons dominants représentent 61% du spectre, alors que les 39% restants sont constitués de 97 taxons mineurs, parmi lesquels 78 taxons sont représentés chacun par moins de 10 charbons (taxons rares). Le principal problème de ce spectre de données est que si ces taxons rares, considérés séparément, n'ont aucun poids statistique, certains d'entre eux n'en sont pas moins des marqueurs écologiques ou socio-culturels de premier ordre. De plus, c'est souvent l'association des taxons entre eux qui est représentative d'une zone écologique ou d'une pratique culturelle particulière. Ne pas considérer ces taxons rares dans les analyses reviendrait donc, de façon regrettable, à limiter le potentiel informatif de l'assemblage initial.

Cette première présentation visait donc à réfléchir sur l'étape de préparation des données, en vue de réaliser une sériation significative de l'usage global des bois au cours de l'occupation du site, ainsi que des analyses multivariées des différents combustibles utilisés en fonction de paramètres extrinsèques (type de feu, type de contexte, phase chronologique, zone du site, matériel associé, etc...). La discussion a d'abord porté sur l'uniformisation préalable des données. La nécessité d'uniformiser les effectifs (nombre de charbons analysés) s'est imposée du fait que les volumes de sédiments prélevés pour la flottation sont différents entre les contextes archéologiques, et que certains échantillons de charbons particulièrement riches ont dû être sous-échantillonnés en laboratoire avant l'analyse. Par conséquent, les assemblages d'origine (comptages bruts) ne sont pas comparables entre les contextes archéologiques. La méthode proposée consiste à travailler soit à partir de la présence-absence des taxons, soit à partir des taux de charbons, c'est-à-dire le nombre de charbons calculé pour 10 L de sédiment par contexte. À partir du tableau de contingence uniformisé ou du tableau de présence-absence, il s'agit ensuite d'extraire des associations significatives de taxons ou de contextes via une ou plusieurs AFC suivies de CAH. Les groupes mis en évidence, s'ils existent, peuvent ensuite être réinjectés en tant que variable qualitative dans la matrice globale. Les taxons non significatifs ou les taxons ubiquistes doivent être supprimés des analyses ou bien maintenus comme variables supplémentaires.

Deux approches, probablement complémentaires, doivent être testées en parallèle. Une approche globale intégrerait l'ensemble des contextes et des taxons et viserait à mettre en évidence la structure statistique globale de la matrice de données, notamment via la méthode de la matrice graphique appliquée à un tableau de présence-absence. Une approche plus contextuelle analyserait séparément les variables entre elles (par exemple taxons/type de contexte ou taxon/phase chronologique) à partir de tableaux de contingences synthétiques, de façon à faire émerger des groupes de taxons potentiellement différents en fonction du facteur extrinsèque considéré. Il a en effet été suggéré au cours de la discussion que la recherche d'une structure globale incluant l'ensemble des variables extrinsèques n'est peut-être pas adaptée, et que, au contraire, la mise en évidence de groupes de taxons différents en fonction de chaque variable extrinsèque pourrait enrichir avantageusement l'interprétation archéologique. Cette remarque fait totalement sens puisqu'il est en effet peu probable que la sélection des bois de feu ait été conditionnée par tous les facteurs socio-culturels et environnementaux à la fois. Quoi qu'il en soit, les premières tentatives montrent qu'il est difficile d'extraire des groupes de taxons statistiquement significatifs sans un premier regroupement des taxons fondé sur des paramètres intrinsèques, tels que les familles de taxons ou les affinités écologiques.

suite et développements postérieurs

Traitement statistique efficace et réaliste du corpus et soutenance de thèse probablement en décembre 2016 ou janvier 2017.

références

Bertin 1977

BERTIN J., *La graphique et le traitement graphique de l'information*, Paris : Flammarion.

séance 6 samedi 25/06/2016

9h30-12h IAA 3, rue Michelet 75006 Paris salle 319

présents : Anne Chaillou, Pablo Ciezar, Robert Demaille, Bruno Desachy, Léa Hermenault, Jérôme Haquet, Christophe Tuffery

actualité et informations diverses

colloques :

Le poster de Léa Hermenault présenté aux journées informatiques et archéologiques de Paris (JIAP) du 8 au 10 juin ([programme](#)) « percevoir les rythmes d'évolution du bâti parisien dans la seconde moitié du XVIe s. » (travail qui a fait l'objet de présentations dans l'atelier le 5/12/2015 et le 23/1/2016) est [téléchargeable](#) sur la page de l'atelier.

Présentation : Stratigraphie et SIG : quelques rappels, réflexions et suggestions sur les conceptions et représentations du temps et de l'espace et sur les liens entre chronologie et spatialisation dans les SIG (C. Tuffery)

type de projet :

réflexions méthodologiques – appel à débats

porteur(s) / auteur(s) du projet :

Christophe Tuffery (ingénieur, INRAP) ; christophe.tuffery@inrap.fr

thématiques de l'atelier concernées par la présentation

1 (méthodes et outils d'enregistrement et de traitement des données de terrain)

contenu de la présentation et de la discussion

Cette présentation a commencé par rappeler que les liens entre chronologie et spatialisation de données dans les SIG est un sujet déjà ancien. Il peut être abordé sous divers angles, parmi lesquels celui de la représentation et de l'analyse d'entités spatio-temporelles ainsi que celui de la gestion du temps dans les SIG. C'est sous ces deux angles, qu'ont été évoquées les interrogations suivantes qui sont proposées comme autant de points de discussion dans un débat qui pourrait être prolongé dans le cadre du séminaire :

- comment font les SIG avec le temps ?
- comment peut-être prise en compte la chrono-stratigraphie dans des modélisations spatiales planaires ?
- quand et comment faut-il spatialiser la stratigraphie ?

Il est utile de rappeler que la modélisation du temps et de l'espace a fait l'objet de nombreux travaux depuis les années 90. La bibliographie très partielle proposée ci-dessous mentionne certains de ces travaux en France ou à l'étranger.

En archéologie, l'étude des relations entre temps et espace est centrale à tel point qu'on peut se demander s'il est possible d'étudier l'un sans l'autre. Cette interrogation, de nature épistémologique, explique l'importance des modèles de description et de représentation des données dans leurs dimensions spatiales ET temporelles, que ce soit à l'échelle du site ou à des échelles plus petites.

A l'échelle du site, la figure 1 indique clairement pour un auteur comme Ph. Boissinot (2015) les deux grandes questions posées à un site (ou agrégat) archéologique au cours de son « démontage ».

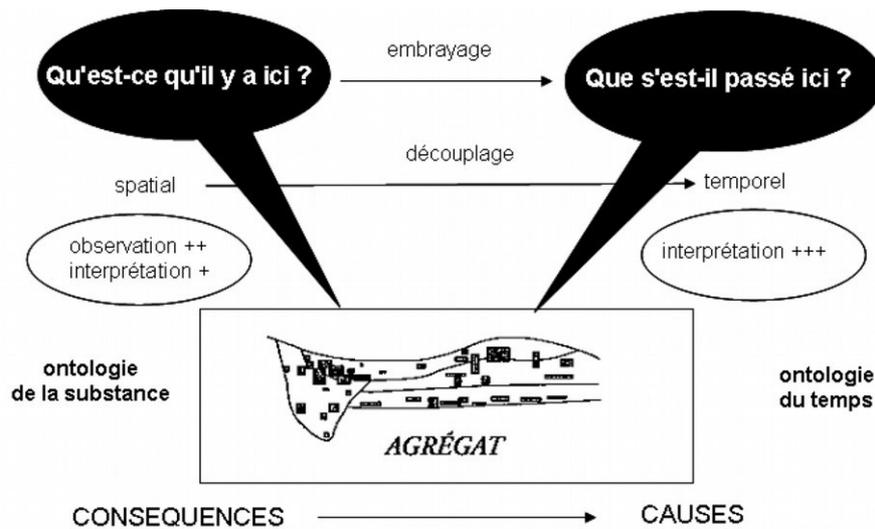


Figure 1 : Les questions posées à un site (agrégat) archéologique lors de son démontage (in Boissinot 2015)

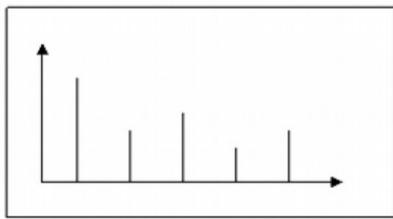
Les deux questions posées dans l'ordre sont :

- « qu'est-ce qu'il y a ici ? » à laquelle doit répondre la description des observations, en laissant peu de place à l'interprétation,
- « que s'est-il passé ici ? » qui est le domaine de l'interprétation

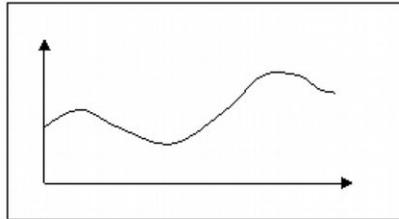
Ce double questionnement s'effectue selon un processus d'« embrayage » ou de « découplage » qui fait passer l'archéologue de l'ontologie de la substance qui propose un modèle décrivant l'espace, à l'ontologie du temps qui décrit la dimension temporelle du site archéologique lors de son « démontage » (ex. : plans phasés).

A d'autres échelles, plus petites que celle du site archéologique, la représentation de l'espace et de ses dynamiques, utilise des catégories de regroupement (ensembles de sites) et/ou de subdivision (discontinuité du temps) pour pouvoir rendre compte d'évolutions (ex. : faciès chrono-culturels), pour lesquels il existe de nombreux modes de représentation (dynamiques d'occupation de territoires, chrono-chorèmes).

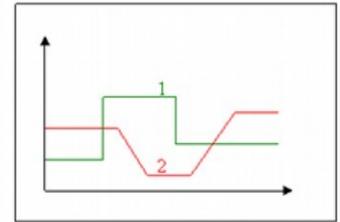
Lorsqu'on évoque la façon de rendre compte du temps, de sa mesure et de processus évolutifs, il est utile aussi de rappeler la différence entre intervalle temporel et résolution temporelle. L'intervalle temporel I est la durée d'observation du phénomène (ex. 100 ans), alors que la résolution temporelle R est la granularité de cette observation (ex. année ou décennie). C'est ainsi que les trois grands types d'évolution peuvent être décrits (figure 2) :



Evolution discrète : $I < R$



Evolution continue : $I > R$



Evolution par paliers

Figure 2 : Les trois grands types d'évolution (in Paque 2004)

Toute modélisation, de l'espace ou du temps, implique de passer d'un espace ou d'un temps continu à un espace ou un temps discontinu. Des choix doivent donc être faits pour le découpage de ces deux dimensions (discrétisation) et introduire de la discontinuité. Cette nécessité méthodologique invite à s'interroger pour savoir si les intervalles temporels et les résolutions temporelles nécessaires pour décrire des chrono-stratigraphies peuvent être compatibles avec celles mobilisées pour décrire des espaces, quelle qu'en soit l'échelle.

Une autre question qui apparaît lorsqu'on souhaite représenter des données temporelles est la suivante : comment faire avec l'incertitude des données temporelles dans la datation de la stratigraphie ? On pense ici bien évidemment aux notions de TPQ/TAQ, qu'il s'agisse de données certaines ou estimées. Les travaux de B. Desachy sur le sujet proposent une méthode et des outils pour traiter ces qualités diverses des données temporelles. A ces travaux, il convient d'ajouter ceux sur l'usage de la logique floue pour la gestion de l'incertitude (Runz 2008) ou encore ceux sur le continuum, notion issue du web sémantique pour la modélisation spatio-temporelle d'un monde ouvert mais incomplet (Harbelot 2015).

Concernant les données temporelles en archéologie, et en particulier pour constituer des diagrammes chronostratigraphiques, il faut souligner ici que les termes d'événements et de faits concernent des notions aux temporalités variables et aux relations pas toujours identifiables.

Pour aller plus avant dans le questionnement sur le temps archéologique, on propose de repartir du modèle synthétique minimal proposé par B. Desachy (2012b) pour toute opération archéologique (figure 3).

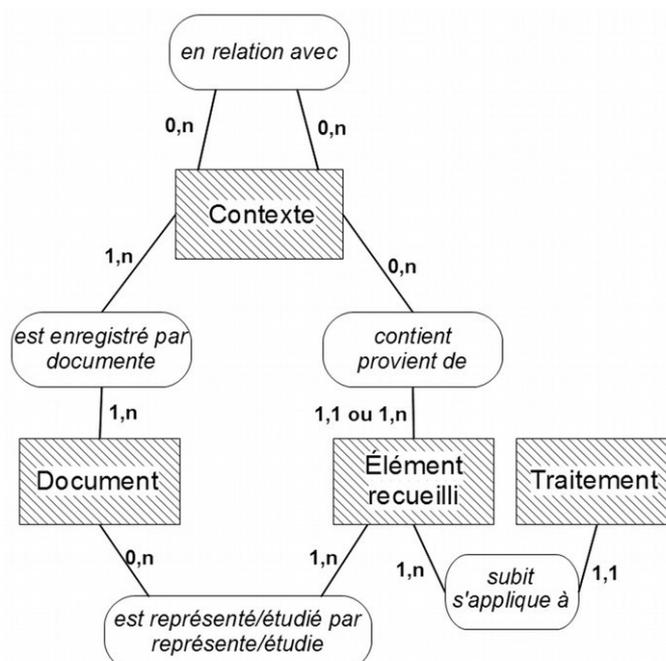


Figure 3 : Modèle minimal des éléments constitutifs d'un site archéologique (Desachy 2012b)

Ce graphique propose "[une] structure générique sous forme d'un modèle entité-association simplifié, tel que l'on peut dire qu'en l'absence de ces entités et relations minimales, un système d'enregistrement de terrain est incomplet" (Desachy 2012b).

Si on s'interroge sur les éléments qui, dans ce modèle minimal, relèvent de la dimension temporelle et de la dimension spatiale, la réponse proposée par B. Desachy (2012b) est la suivante : "Les relations entre documents d'une part, contextes et éléments recueillis d'autre part, incluent la spatialisation (topographique, voire topologique) de ces contextes et éléments, lorsque les documents concernés sont des relevés et représentations spatiales, ou qu'ils précisent une localisation ; laquelle spatialisation peut recourir à des outils de SIG".

Les éléments qui relèvent de la dimension temporelle sont, au cours de la phase terrain, le contexte et l'élément recueilli (en post-fouille s'y ajoute le traitement) entre lesquels la relation établie relève de la dimension spatiale. Chacun de ces éléments est décrit dans un document qui enregistre la dimension temporelle et spatiale de ces éléments, la dimension spatiale étant décrite par des relevés et des représentations spatiales et une localisation. Ce processus de spatialisation des éléments s'effectue largement aujourd'hui à l'aide des outils de type SIG (figure 4).

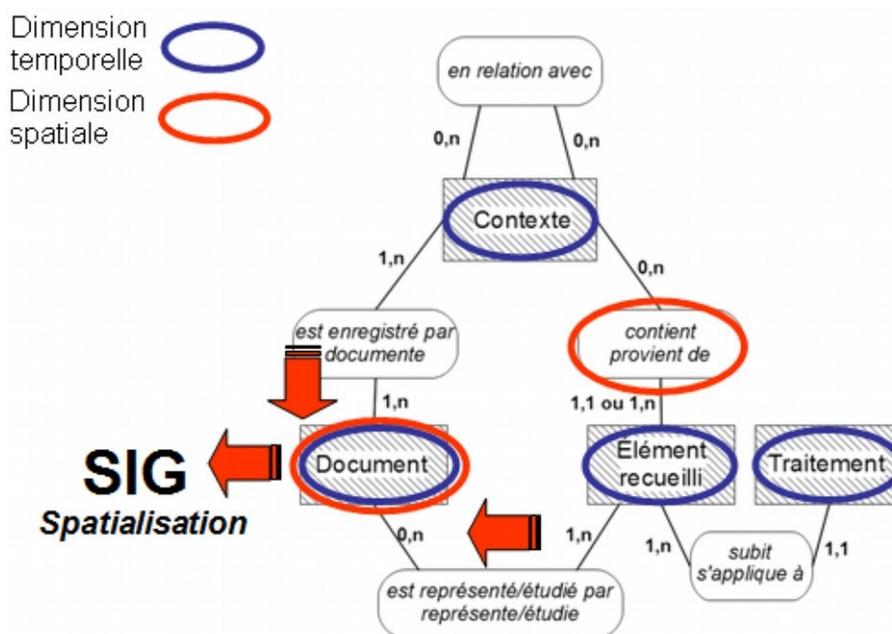


Figure 4 : Éléments temporels et éléments spatiaux dans le modèle minimal des éléments constitutifs d'un site archéologique (d'après Desachy 2012b)

Il existe différentes solutions pour exploiter et afficher le temps dans les SIG. On peut citer ici entre autres :

- la représentation des valeurs d'une ou de plusieurs variables temporelles associées à des entités spatiales (ex. : changement de localisation, de forme, de symbole,...),
- la représentation d'une ou de plusieurs variables d'évolution d'entités spatiales (ex. : croissance, décroissance, stagnation, disparition, apparition) entre plusieurs dates ou sur une période,
- le regroupement des entités spatiales appartenant à une même unité temporelle (ex. phasage de la stratigraphie, plan phasé),
- l'utilisation des "objets géo-historiques", définis par les trois dimensions F (fonction), E (espace), T (temps) (Rodier, Saligny 2010),

- l'implémentation d'une modélisation de type CIDOC-CRM dans une BDD spatiale (Le Goff et al. 2015)

On peut aussi mentionner ici le courant de la géographie temporelle ou "espace-temps" et ses cubes spatio-temporels (Hägerstrand 1970) qui ont eu, semble-t-il, peu de retombées en archéologie.

Lorsqu'on veut traiter plus spécifiquement du temps archéologique dans les SIG, nous proposons de repartir des travaux de B. Desachy qui, s'inspirant de J. Bertin (1967), propose une démarche en trois temps :

- traiter les unités d'enregistrement stratigraphique indépendamment de leurs géométries et de leur localisation
- établir leurs relations d'ordre chronologique (équivalence, antéro-postériorité)
- enfin spatialiser les unités d'enregistrement avec leurs relations chronologiques

Comme l'indique B. Desachy (2012b), à propos des sites densément stratifiés « *la stratification tridimensionnelle qui les caractérise impose, pour être pleinement représentée, de traiter des volumes (des polyèdres et non plus des polygones) et leur topologie (les faces communes traduisant des relations stratigraphiques) : ce que ne sait pas faire aujourd'hui un simple SIG cartographique* ». Là se trouve une limite incontestable de l'utilisation des SIG pour rendre compte correctement de la dimension temporelle des éléments constituant la stratigraphie, qui devraient donc être disponibles dans les SIG comme des volumes avec une topologie volumétrique et non comme des entités géométriques à deux dimensions, décrites uniquement dans une topologie planaire.

Il est techniquement possible d'afficher depuis ou à l'intérieur des logiciels de SIG comme ArcGIS ou QGIS des diagrammes stratigraphiques :

- importer dans le logiciel de SIG le fichier .xls du Stratifiant de B. Desachy (2008) et faire une jointure avec la couche géoréférencée des unités stratigraphiques par un champ commun,
- réorganiser l'ordre de superposition des géométries représentant les US en fonction du rang stratigraphique donné par le Stratifiant (Th. Guillemard),
- co-visualiser le fichier .xls du Stratifiant depuis le logiciel de SIG,
- insérer dans l'outil de mise en page du logiciel de SIG (mode mise en page dans ArcGIS, composeur d'impression dans QGIS) un objet en liaison dynamique avec le graphe du Stratifiant,
- développer dans un logiciel des scripts pour produire un chronographe des entités fonctionnelles (cf. application OH_FET développée par O. Marlet (L.A.T. UMR CITERES) pour QGIS).

Mais ces diverses solutions, qui tiennent pour certaines d'un « bricolage », ne suffisent pas pour disposer d'un modèle volumétrique. Mais là où les limites de la topologie planaire sont encore plus évidents c'est dans les opérateurs spatiaux disponibles dans les logiciels de SIG comme ceux mentionnés ci-dessus ou dans la plupart des autres disponibles à ce jour, qu'ils soient libres ou propriétaires.

A l'image du modèle de topologie planaire dans lequel sont décrites les entités géométriques, les opérateurs spatiaux sont adaptés à une topologie planaire. Ils ne prennent pas en compte les relations de chronologie relative qu'un diagramme stratigraphique permet de restituer.

C'est pourquoi nous avons entrepris de passer en revue les principaux opérateurs spatiaux les plus courants et d'évaluer les possibilités de les utiliser pour des unités stratigraphiques avec diverses relations de chronologie relative.

Les résultats que nous avons obtenus et que nous souhaitons partager et mettre en débat, montrent que toutes les combinaisons des opérateurs spatiaux avec les divers cas où les unités stratigraphiques ont des relations chronologie relative variées, ne sont pas possibles. Seule la combinaison de règles de chronologie relative et de distance relative permet d'obtenir des opérateurs spatio-temporels, adaptés aux besoins des archéologues.

Cette limitation des opérateurs spatiaux de topologie planaire prenant en compte des éléments de chronologie relative, est due au fait que les opérateurs spatiaux comportent implicitement des relations spatiales pour des entités se trouvant dans un même espace plan temporel, c'est-à-dire auquel il manque la profondeur temporelle que procure la chronologie relative. Ces opérateurs « embarquent » avec eux des relations de chronologie relative qui empêchent de les utiliser pour certaines relations chronologiques établies entre US.

Les neuf figures 5 à 14 proposent pour les différents cas de figure des relations entre plusieurs US, d'identifier ceux où l'usage des principaux opérateurs spatiaux du logiciel de SIG QGIS (mais un exercice similaire est envisageable pour ArcGIS).

Les figures correspondent aux cas de figure suivants (le cas de figure est en tête de la figure) :

- les deux unités d'enregistrement (UE 1 et UE 2) sont synchrones ou antérieures ou postérieures l'une à l'autre
- idem avec les deux unités d'enregistrement (UE 1 et UE 3)
- idem avec les deux unités d'enregistrement (UE 2 et UE 3)

Chacune des figures se lie de gauche à droite avec :

- à gauche : l'opérateur spatial (encadré)
- au centre : la représentation des UE sous forme de volumes schématisés en stratigraphie. En cas d'impossibilité d'utiliser l'opérateur spatial pour la configuration stratigraphique des UE, nous faisons apparaître une croix rouge traduisant l'impossibilité de la combinaison, de l'opérateur spatial et de la position en stratigraphie des UE,
- à droite : la représentation des UE sous forme de polygones schématisés en planimétrie sous réserve que l'opérateur spatial soit compatible avec la configuration stratigraphique.

Ces figures correspondent donc à toutes les combinaisons possibles et identifient celles qui nous paraissent impossibles. Pour quelques unes d'elles, nous convenons que l'impossibilité se discute.

A l'issue de ces travaux exploratoires, nous pouvons affirmer la nécessité pour les SIG utilisés en archéologie pour restituer l'évolution spatio-temporelle d'US, de faits, de structures à l'échelle d'un site, d'un ensemble de sites à l'échelle locale ou régionale ou à des échelles plus petites, qu'ils intègrent des opérateurs spatio-temporels, c'est-à-dire des opérateurs spatiaux sous contrainte de cohérence temporelle. Ces opérateurs devraient vérifier que les relations spatiales sont possibles "dans la durée de vie" ("durée de formation", "durée d'usage", etc.) des unités spatiales concernées.

UE1 est synchrone avec UE2

□ UE 1 ■ UE 2

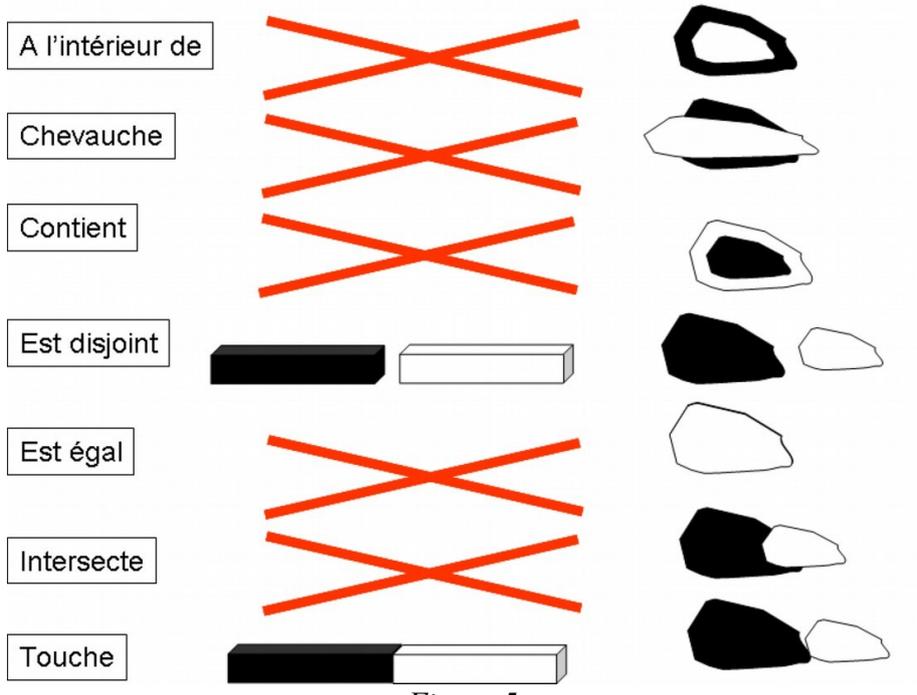


Figure 5

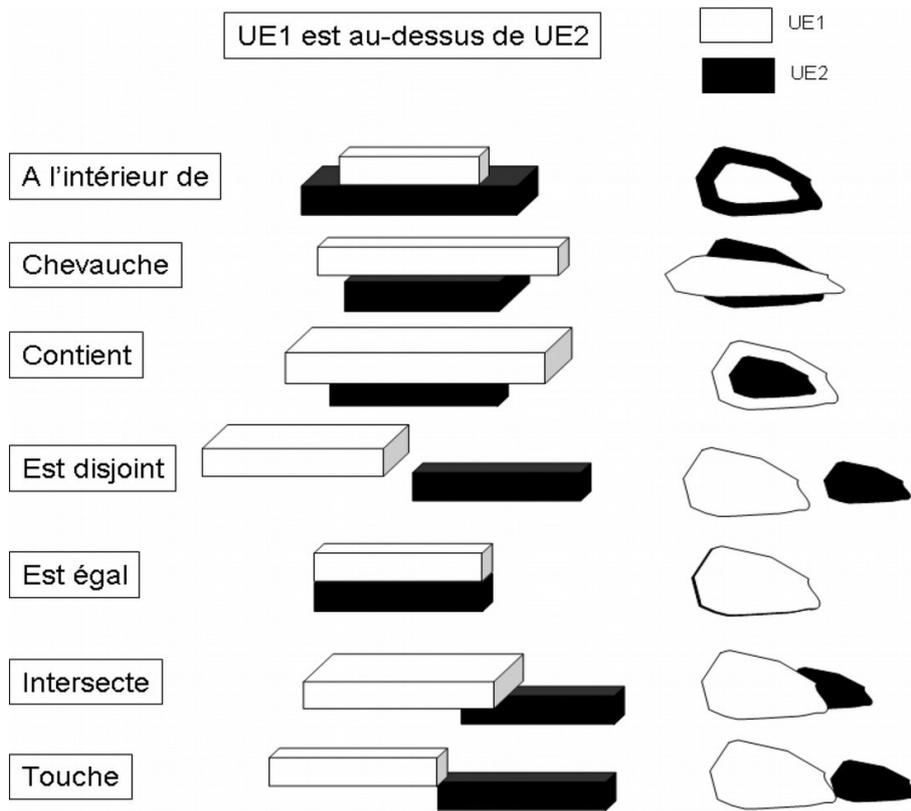


Figure 6

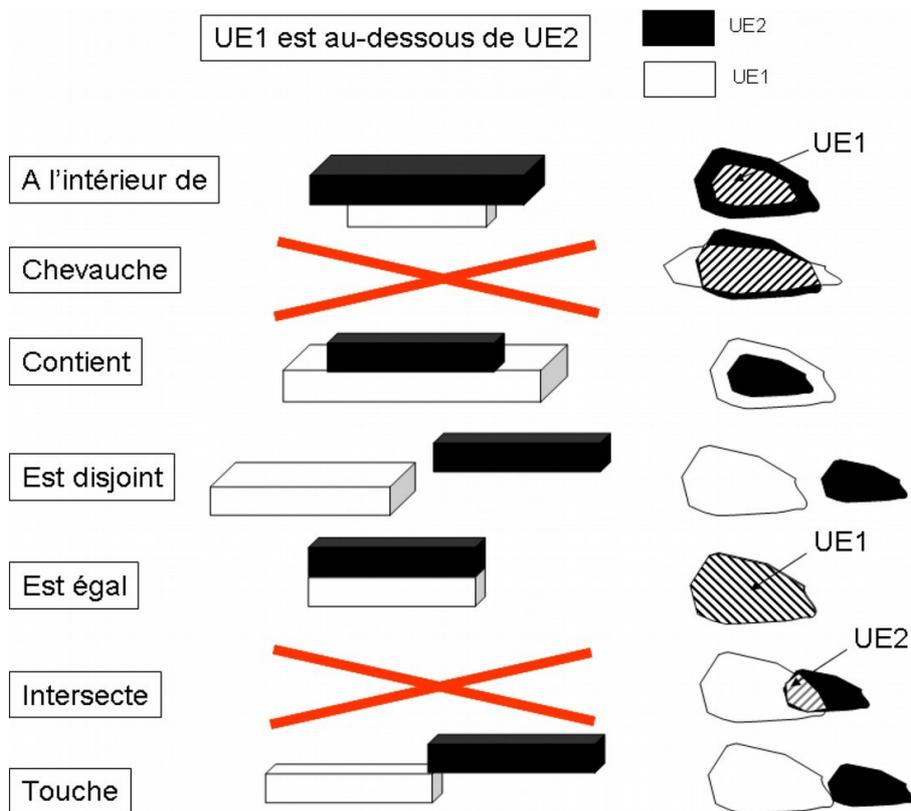


Figure 7

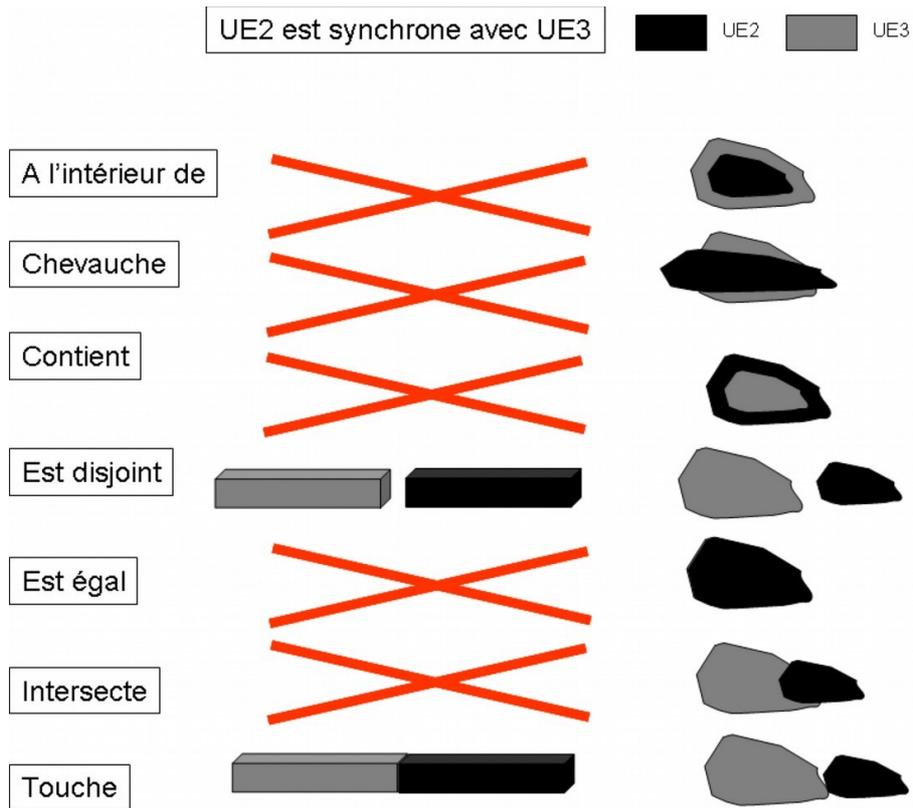


Figure 8

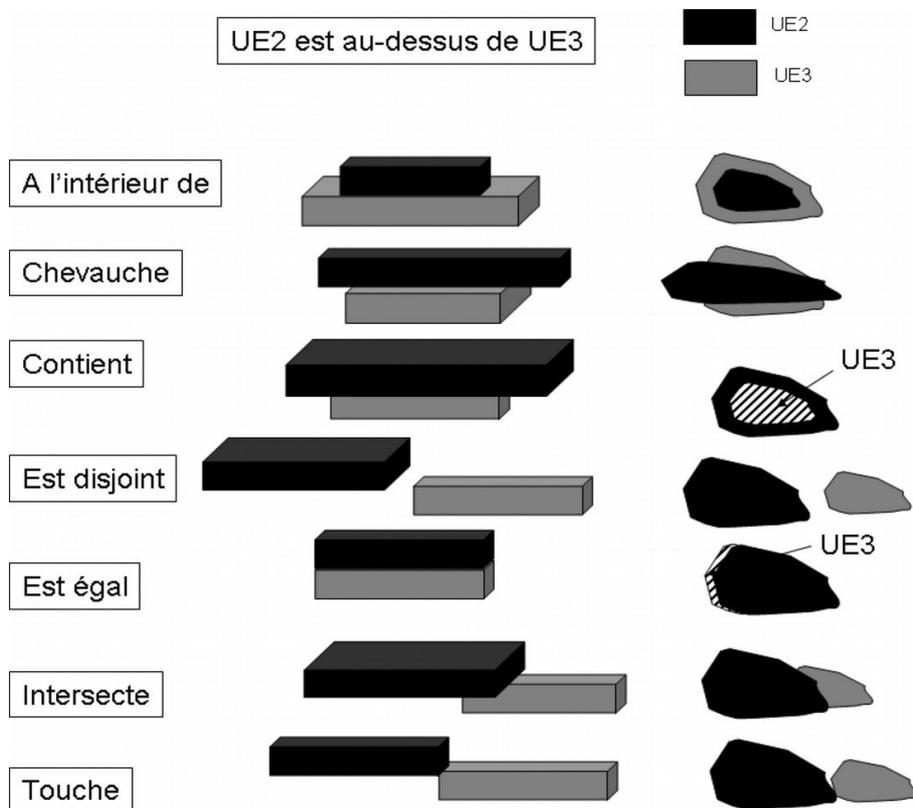


Figure 9

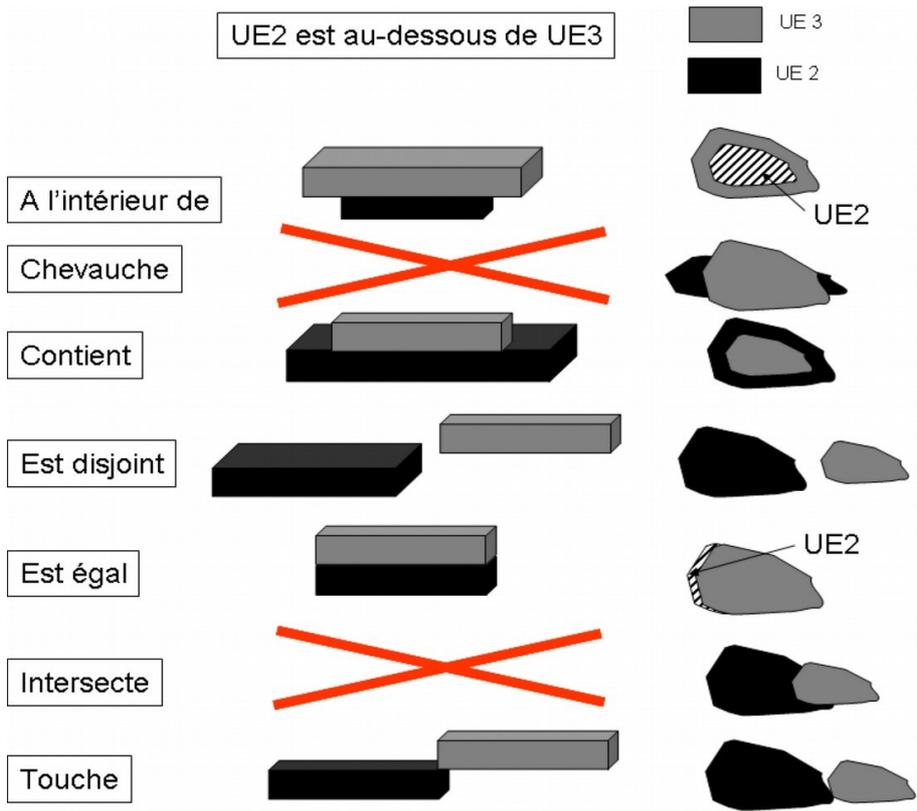


Figure 10

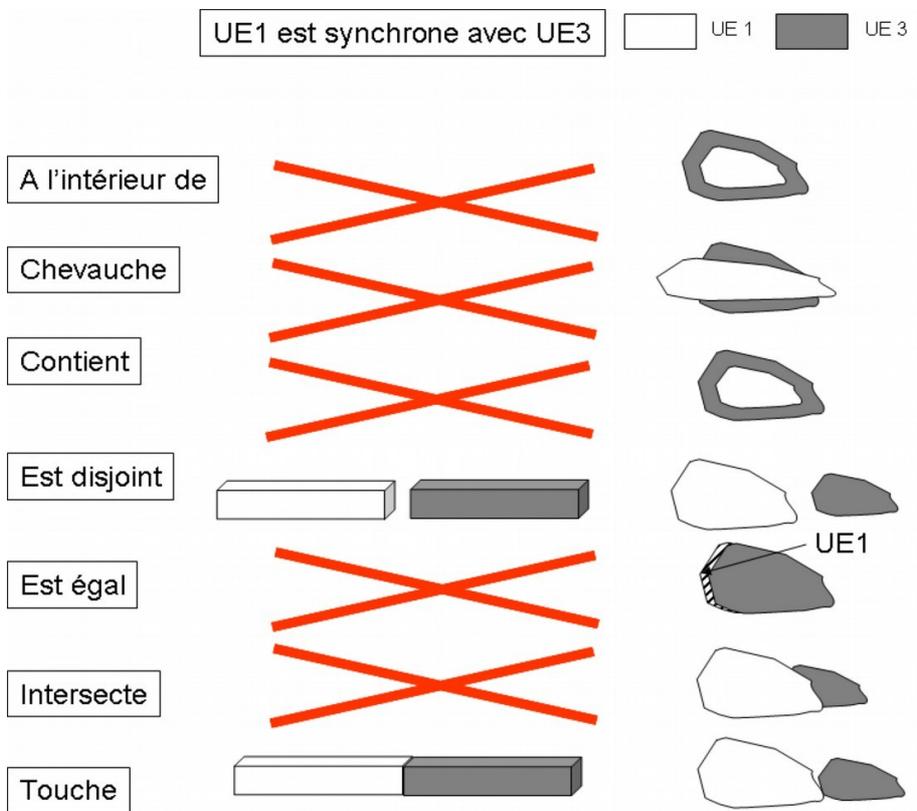


Figure 11

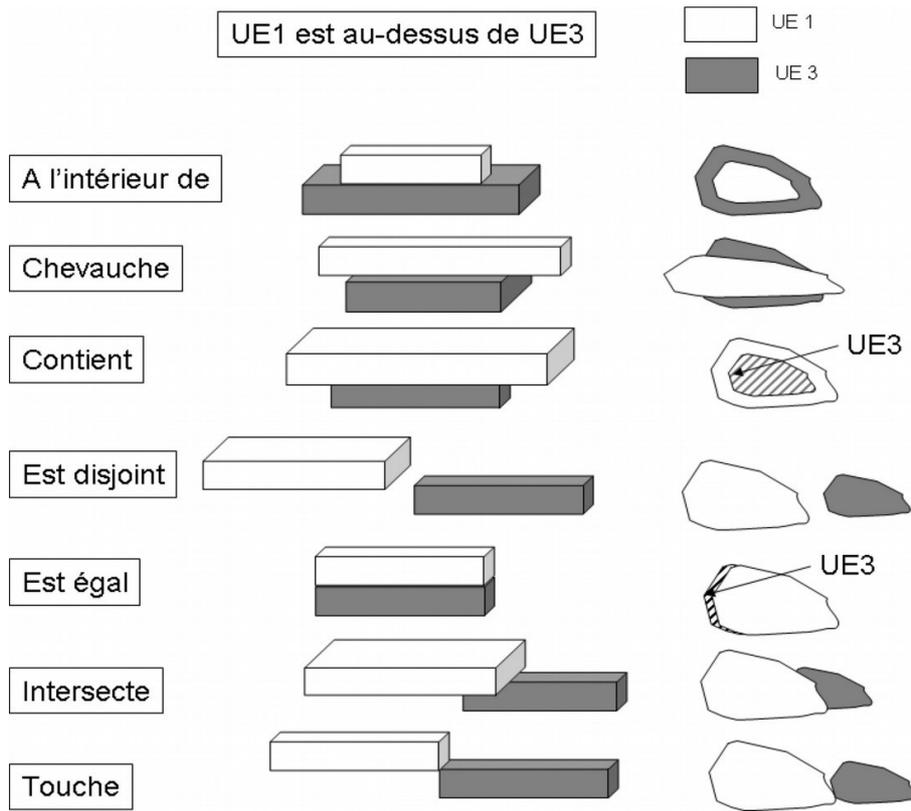


Figure 12

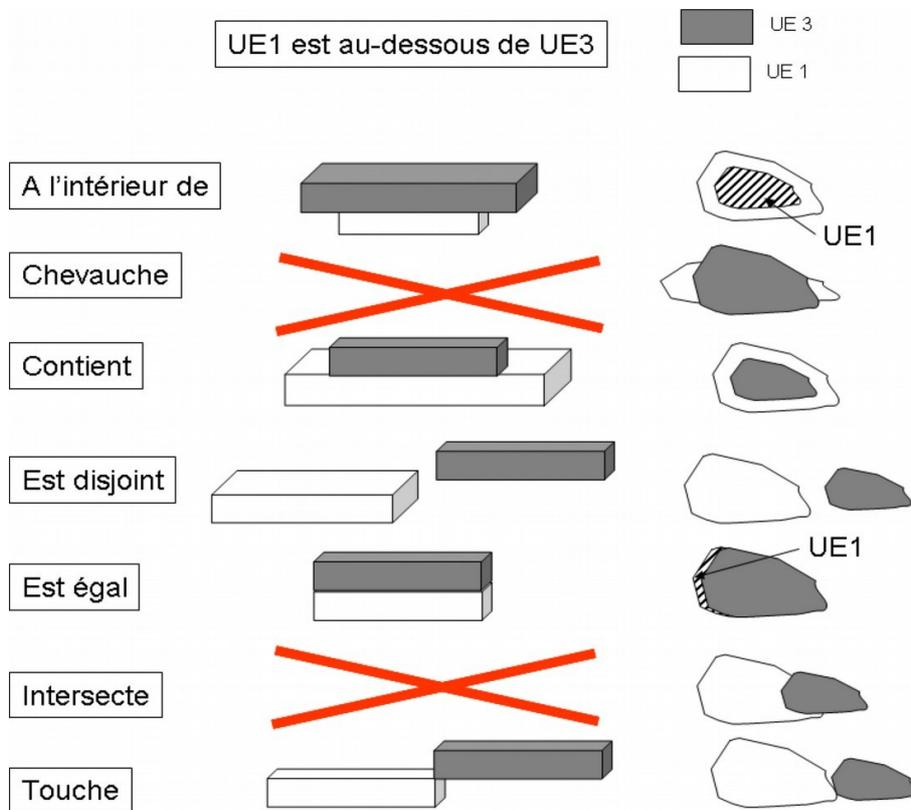


Figure 13

suite et développements postérieurs

Une séance de l'atelier revenant sur ces questions et intégrant les retours et remarques suscités par ces propositions aura lieu en dans la saison 2016-2017

références

Bibliographie très partielle :

Allen 1991

ALLEN J. F., « Time and Time Again: The Many ways to Represent Time », *Int. J. Intell. Syst.*, 6, 4, pp. 341-355.

Armand 2012

ARMAND C., « Quels outils numériques pour représenter le temps en histoire ? » [en ligne], URL : <http://advertisinghistory.hypotheses.org/79> [lien valide au 19 août 2016].

Berger et al. 2005

BERGER J.-F., BERTONCELLO F., BRAEMER F., DAVTIAN G., GAZENBEEK M., *Temps et espaces de l'Homme en société. Analyses et modèles spatiaux en archéologie*, Textes des communications en français ou en anglais. - APDCA = Association pour la promotion et la diffusion des connaissances archéologiques. - Le CD-Rom contient tous les textes et les illustrations de l'ouvrage avec un système de recherche lié au sommaire pour Mac et PC, Antibes, Antibes : APDCA, Rencontres internationales d'archéologie et d'histoire d'Antibes.

Bertin 1967

BERTIN J., *Sémiologie graphique: les diagrammes, les réseaux, les cartes*, notamment présent à Michelet, Paris : Gauthier-Villard.

Boissinot 2011

BOISSINOT P., *L'archéologie comme discipline ?*, Paris : Seuil.

Boissinot 2015

BOISSINOT P., *Qu'est-ce qu'un fait archéologique ?*, Paris : Éditions EHESS.

Bordin 2006

BORDIN P., *Méthode d'observation multi-niveaux pour le suivi de phénomènes géographiques avec un SIG* [en ligne], (s.l., Université de Marne-La-Vallée, Thèse de doctorat sous la direction de L. Sanders), URL : http://recherche.ign.fr/labos/cogit/pdf/THESES/BORDIN/these_bordin10.04.07.pdf.

Cheyland 2007

CHEYLAND J.-P., « Les processus spatio-temporels : quelques notions et concepts préalables à leur représentation », *M@ppemonde* [en ligne], 87, 3, URL : <http://mappemonde.mgm.fr/num15/articles/art07303.html>.

Collectif 2016

COLLECTIF, *Du silex au gobelet en plastique : réflexions sur les limites chronologiques de l'archéologie*, Bordeaux : Fedora.

Desachy 2008

DESACHY B., *De la formalisation du traitement des données stratigraphiques en archéologie de terrain* [en ligne], (s.l., Université Paris I - Panthéon-Sorbonne, Thèse de doctorat sous la direction de J. Burnouf), URL : <https://tel.archives-ouvertes.fr/tel-00406241/document>.

Desachy 2012a

DESACHY B., « Formaliser le raisonnement chronologique et son incertitude en archéologie de terrain », *CyberGeo Eur. J. Geogr.* [en ligne], 597, URL : <http://cybergeog.revues.org/25233> [lien valide au 28 février 2013].

Desachy 2012b

DESACHY B., « Systèmes d'information archéologique de terrain et fondamentaux de l'enregistrement archéologique. Quelques remarques à propos de l'application Stratibase », *Archeol. e Calc. Suppl.* 3 [en ligne], 61, URL : http://soi.cnr.it/archcalc/indice/Suppl_3/05-desachy.pdf.

Hägerstrand 1970

HÄGERSTRAND T., « What about people in regional science? », *Reg. Sci. Assoc. Pap.*, 24, 1, pp. 7-24.

Langran 1993

LANGRAN G., *Time in geographic information systems*, Londres : Taylor et Francis.

Le Goff et al. 2015

LE GOFF É., MARLET O., RODIER X., CURET S., HUSI P., « Interoperability of the ArSol (Archives du Sol) database based on the CIDOC-CRM ontology », in *CAA2014, 21st Century Archaeology, 42nd International Conference on Computer Application and Quantitative Methods in Archaeology* [en ligne], Oxford : Archaeopress, pp. 179-186, URL : <http://www.archaeopress.com/ArchaeopressShop/Public/displayProductDetail.asp?id=%7BE35F9954-5653-493D-884B-4A7D2DE66610%7D>.

Mathian, Sanders 2014

MATHIAN H., SANDERS L., *Objets géographiques et processus de changement : approches spatio-temporelles*, Londres : ISTE Editions.

Paque 2004

PAQUE D., « Gestion de l'historicité et méthodes de mise à jour dans les SIG », *CyberGeo Eur. J. Geogr.* [en ligne], 278, URL : <http://cybergeog.revues.org/2500?lang=en>.

Peuquet 2002

PEUQUET D., *Representations of space and time*, New-York : Guilford.

Rodier, Saligny 2010

RODIER X., SALIGNY L., « Modélisation des objets historiques selon la fonction, l'espace et le temps pour l'étude des dynamiques urbaines dans la longue durée », *CyberGeo Eur. J. Geogr.* [en ligne], 502, URL : <http://cybergeog.revues.org/pdf/23175> [lien valide au 2 mai 2013].

Testart 2012

TESTART A., *Avant l'histoire : l'évolution des sociétés, de Lascaux à Carnac*, Paris : Gallimard.

Présentation de projet : traiter statistiquement les phénomènes de cycles et répétitions de pas variable : le cas des dérivées de sigillée d'Argonne décorées à la molette (P. Ciezar)



type de projet :

Programme *Corpus de la sigillée d'Argonne au Bas-Empire* de l'équipe GAMA de l'UMR 7041 ArScAn conduit depuis plus de 20 ans par Paul Van Ossel (univ. Paris Ouest), Lothar Bakker (conservateur au musée d'Augsbourg - Allemagne) et Wim Dijkmann (Archéologue municipal de la ville de Maastricht - Pays-Bas) et soutenu par la Römisch-Germanische Kommission (RGK) de Frankfurt.

Le travail entrepris par les porteurs du projet consiste en un inventaire systématique et une relecture et révision rigoureuse de la documentation existante.

thématiques de l'atelier concernées par la présentation

3 (sériations et typologies des éléments matériels)

contenu de la présentation et de la discussion

L'objet de cette présentation est d'évoquer pour la première fois en cet Atelier un projet en cours, dont les traitements sont encore au stade de l'exploration et de la préparation des données.

Les dimensions du corpus

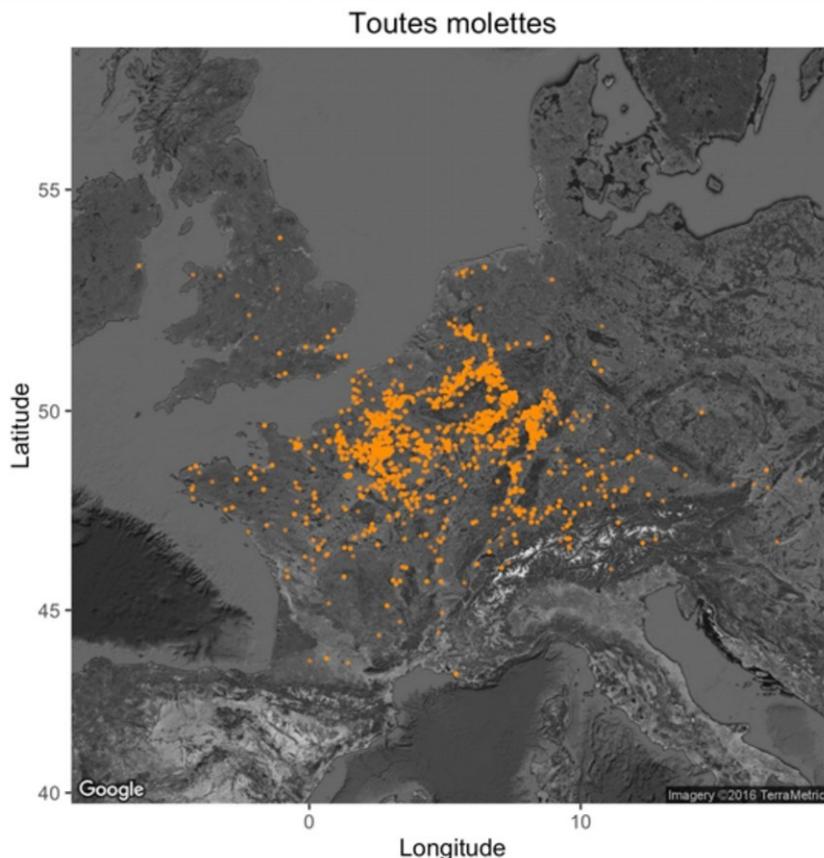
Environ 604 motifs de décor à la molette sont individualisés. Ils sont inventoriés, illustrés et identifiés par une codification numérique. Leurs associations forment plus de 900 séquences originales (ou "variantes originales") documentées et décrites. Environ 18 000 tessons illustrent ces molettes. Ils sont identifiés, décrits, mesurés et commentés. Leurs lieux de découverte, lorsque cela est possible, sont géoréférencés. Plus de 2 600 lieux sont enregistrés. Ils sont répartis en au moins 11 pays européens. La documentation permet de constituer environ 2 600 contextes d'association de découvertes dont 1 600 peuvent être considérés comme "fiables"¹ et seront exploités pour l'analyse sérielle.

Cartographie

Une projection cartographique rapide des 2 600 lieux de découverte permet de visualiser l'aire de diffusion des molettes d'Argonne.

¹ - ne sont retenus parmi les contextes "fiables" que ceux dont la contemporanéité peut être envisagée, en écartant là où les molettes paraissent intrusives au vu du mobilier associé

Cette carte² reprend l'ensemble des points de découverte identifiés dans la base de données. Très regroupée dans le nord de l'Europe, la diffusion de ces objets montre, à l'échelle représentée, des



anomalies qu'il sera utile d'éclairer et d'explorer (biais de la recherche ou phénomènes réels).

Les motifs

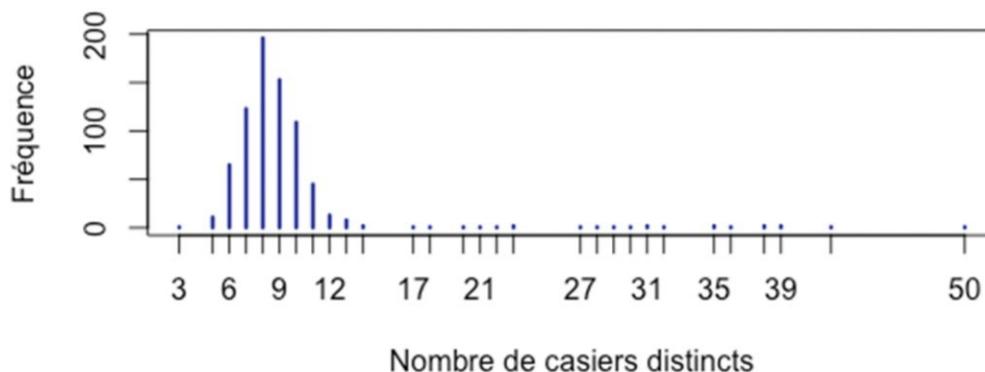
Le motif est l'élément unitaire de la description. Il est identifié visuellement et codifié numériquement. Il s'agit de formes géométriques simples ou complexes (hachures, croix, grilles, etc.), de motifs figuratifs (animaux, personnages, grappes de raisin, calices), de motifs très chargés symboliquement (croix, chrismes,...) et même des inscriptions ("Concordia"...). Certains sont très fréquents et peuvent être qualifiés de motifs de "remplissage", d'autres sont moins fréquents mais bien présents, certains, la plupart, sont rares, voire uniques. Enfin, à ce cortège de motifs il faut ajouter l'identification de séparateurs simples, double ou triples. Dans la codification utilisée une place est laissée pour les lacunes (motif illisible, partie manquante) ou les parties inconnues d'une séquence (fragment de séquence très courte).

Décors à la molette

C'est l'association d'un nombre fini de motifs qui crée une séquence. Les séquences sont ainsi composées par des séries de casiers allant de 3 à 50 casiers.

2 - carte réalisée grâce au package RgoogleMaps

Variation du nombre de casiers par molette



Elles constituent des "familles" pouvant présenter de nombreuses associations de motifs.

- décors géométriques ;
- décors figuratifs ou symboliques;
- décors avec "inscriptions"

La base de données

Ces données sont assemblées dans une base de données relationnelle sous Filemaker Pro par choix "historique" et pratique. Cet outil a évolué pour faciliter les identifications, l'enregistrement des tessons et des lieux de découverte et compléter ainsi les informations de diffusion des différentes molettes. C'est un support informatique qui facilite l'édition de l'ouvrage, le corpus des sigillées d'Argonne décorées à la molette, en cours de rédaction. Les données recueillies et enregistrées dans cette base serviront aussi de support à des traitements cartographiques et statistiques. La Base de données et sa structure, brièvement présentées dans l'atelier du 25/06/2016, mériteraient une présentation plus détaillée lors d'un prochain atelier. D'ores et déjà, elle bénéficiera des conseils et remarques émis lors de cette séance d'atelier.

Le tableau ci-dessous illustre les principales tables de la base :

| Table | Objet |
|-----------|--|
| Motif | Table des motifs individualisés |
| Molette | Table des combinaisons de motifs identifiées |
| Diffusion | Table des tessons des molettes |
| Lieux | Table des lieux de découverte |
| Contexte | Table des contextes d'association des molettes |

Les formulaires de saisie associent informations descriptives, codées, prédéfinies ou libres, et des images. Une optimisation des rubriques image mérite d'être engagée à la lumière des conseils émis lors du séminaire.

Le module d'aide à l'identification permet, par la saisie des codes de quelques motifs reconnus de faciliter l'attribution à une molette existante ou, en cas de d'éléments très partiels, d'obtenir les proximités immédiates d'associations.

Objectifs du traitement statistique envisagé :

en l'état actuel de la réflexion, les objectifs du traitement de données sont :

- d'obtenir un tableau de distances entre les différentes molettes permettant de déterminer le degré de "familiarité" des unes avec les autres ;
- d'effectuer une sériation des molettes en exploitant les associations dans les contextes ;
- de tenir compte d'une caractéristique importante des molettes qui est la position des casiers dans la succession, qui se traduit par un ordre incluant des répétitions cycliques (sachant que l'enregistrement de la molette impose nécessairement mais artificiellement un début et une fin dans cette succession cyclique de casiers) ;

premiers essais : approche classique d'analyse des données multidimensionnelle (AFC et CAH sur tableau des fréquences ou présences de motifs par molettes)

Le processus suivant est en cours d'expérimentation :

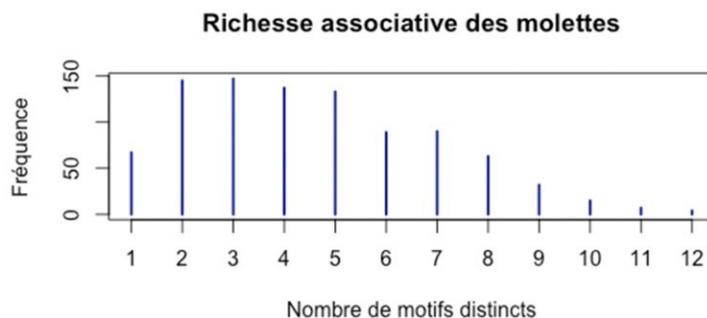
- création d'un tableau où chaque séquence est décrite par une succession de motifs, suivant deux options complémentaires :
 - présence / absence (les motifs qui se répètent dans une même séquence ne sont comptabilisés qu'une seule fois) ;
 - fréquences des motifs (les colonnes indiquent le nombre d'occurrences d'un motif dans la séquence) ;
- réduction(s) du tableau de x lignes par y colonnes en fonction des effectifs (lignes et colonnes) ;
- AFC sur tableau consolidé puis sélection des variables et individus actifs les plus pertinents par répétition du traitement en fonction des contributions relatives des individus et des variables ;
- CAH des séquences à partir des coordonnées sur les premiers axes factoriels.

L'outil de traitement choisi est le logiciel R, en utilisant quelques "packages" spécifiques. Les tableaux de données sont extraits de la base, au format CVS ou XLS.

```
# librairies principalement utilisées
#library(xlsx)
library(dplyr)
library(ade4)
library(ggplot2)
library(ggdendro)
library(rmarkdown)
```

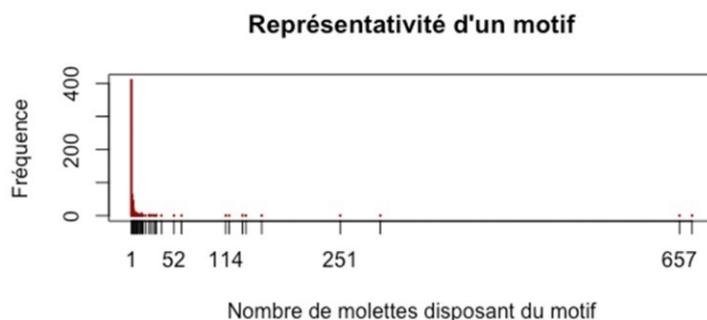
Les tableaux "Molette" et "Motifs" sont constitués à partir de la base de données (ci-dessous extrait du tableau d'inventaire des molettes) et importés dans R en format CSV.

| Numéro | État | N° Hübener | Formule | Long. |
|--------|------|------------|---|-----------|
| 1447 | C | 3 | 116 116 115 116 115 116 115 | 7 cm |
| 1577 | C | 1 | 101 101 101 101 101 101 101 101 | 7,5 |
| 3064 | C | 5 | 162 116 100 115 100 116 100 115 136 116 115 | env. 8 cm |



- représentation des motifs (nombre de molettes utilisant le motif) :

```
## effcol01
## 1 2 3 4 5 6 7 8 9 10 12 13 14 15 18 22 24 27
## 410 63 45 19 10 8 5 7 4 3 1 7 2 2 1 2 2 2
## 29 31 37 52 61 114 118 134 138 157 251 299 657 672
## 1 2 1 1 1 1 1 2 1 1 1 1 1 1
```



L'étape suivante consiste à réduire le tableau ; dans son état d'origine (929 lignes par 609 colonnes) il présente en effet des caractéristiques peu propices à un traitement statistique multidimensionnel basé sur la recherche des associations : codes de « motif » correspondant en fait à l'indication de lacune, individus (molettes) « orphelins » correspondant à un seul motif, réciproquement variables (motifs) « orphelines » représentées par un seul individu... Après retrait de ces variables et individus problématiques, le tableau "nettoyé" est composé de 927 lignes par 263 colonnes.

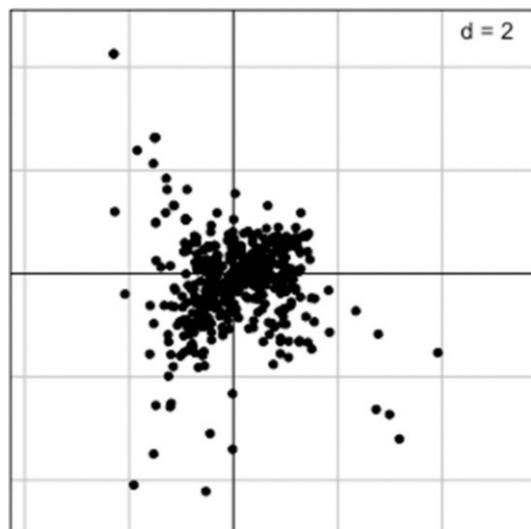
Le tableau est soumis au traitement sous R, converti en "data.frame", puis traité par analyse factorielle des correspondances (AFC) à l'aide des fonctions des bibliothèques *ade4* ou *FactoMineR*.

```
## [1] "itération n° 1"
## [1] "réduction de 28 lignes"
## [1] "réduction de 19 colonnes"
## [1] "itération n° 2"
## [1] "réduction de 10 lignes"
## [1] "réduction de 11 colonnes"
## [1] "itération n° 3"
## [1] "réduction de 84 lignes"
## [1] "réduction de 12 colonnes"
## [1] "itération n° 4"
## [1] "réduction de 33 lignes"
## [1] "réduction de 5 colonnes"
## [1] "itération n° 5"
## [1] "réduction de 12 lignes"
## [1] "réduction de 5 colonnes"
## [1] "itération n° 6"
## [1] "réduction de 25 lignes"
```

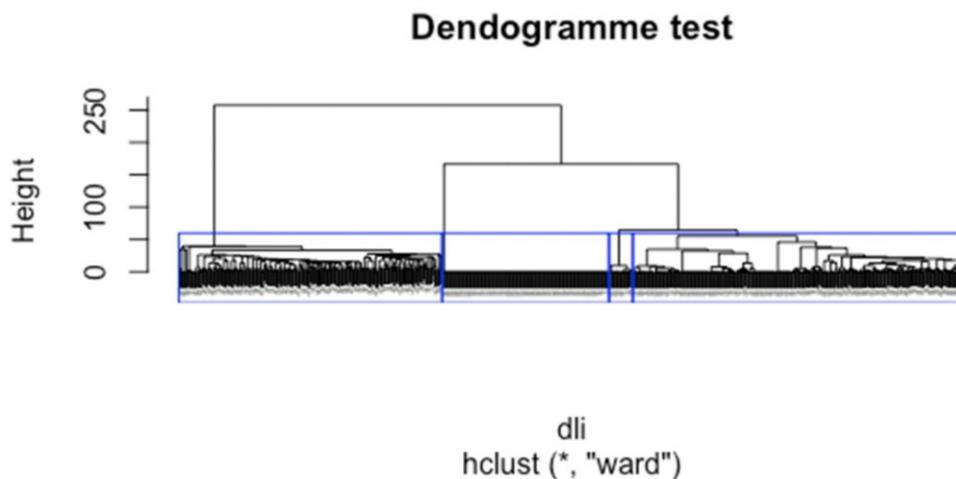
```
## [1] "réduction de 14 colonnes"  
## [1] "itération n° 7"  
## [1] "réduction de 7 lignes"  
## [1] "réduction de 8 colonnes"  
## [1] "itération n° 8"  
## [1] "réduction de 2 lignes"  
## [1] "réduction de 2 colonnes"  
## [1] "itération n° 9"  
## [1] "réduction de 1 lignes"  
## [1] "réduction de 1 colonnes"  
## [1] "itération n° 10"  
## [1] "réduction de 0 lignes"
```

Après 10 itérations on obtient la projection sur le premier plan factoriel. Au préalable, à chaque étape, on conserve les variables (motifs) et molettes (individus) qui sont "retirés" du tableau analysé afin de procéder à l'explication de leur caractère "discordant" : unicité des associations (même motif dans toute la molette, n'apparaissant que dans une seule molette), exclusivité des associations, etc.

Les résultats doivent encore être explorés à l'aide des outils d'aide à l'interprétation proposés par les différents packages de R utilisés.



Une classification peut ensuite être opérée sur les coordonnées des axes factoriels, selon le processus habituel.



Cependant, par rapport aux objectifs de traitement posés plus haut, cette démarche a pour limite d'analyser les associations en fonction des seules fréquences de motifs, sans tenir compte de l'ordre des motifs dans les molettes.

Autre piste de traitement : approche par appariements des motifs

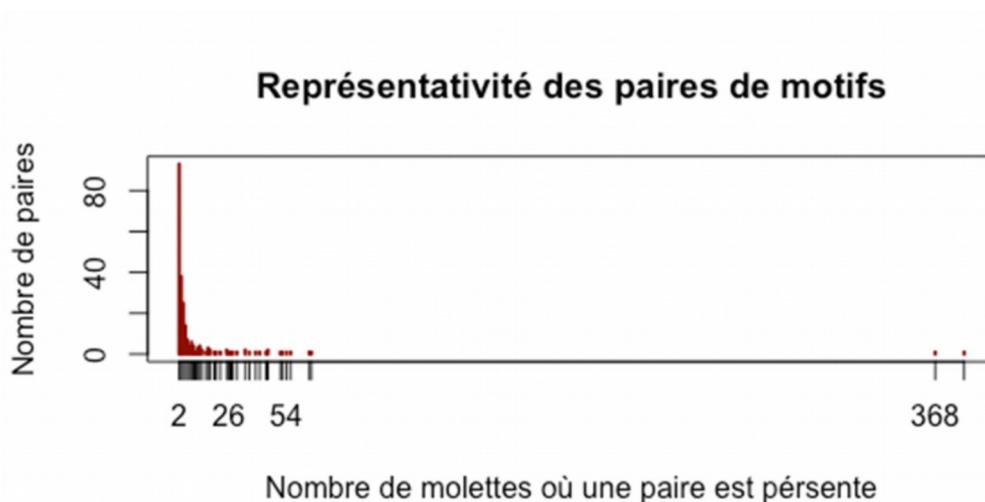
pour tenir compte (au moins partiellement) de cet ordonnancement des motifs dans les molettes, une approche complémentaire (suggérée par B. Desachy au cours de la discussion préalable à la présentation) a fait l'objet de premiers essais. Elle consiste à identifier et prendre comme variables les paires ordonnées de motifs (motif précédent – motif suivant) présentes dans les molettes.

Un tableau comprenant les paires (hors séparateurs et lacunes) a été créé ;

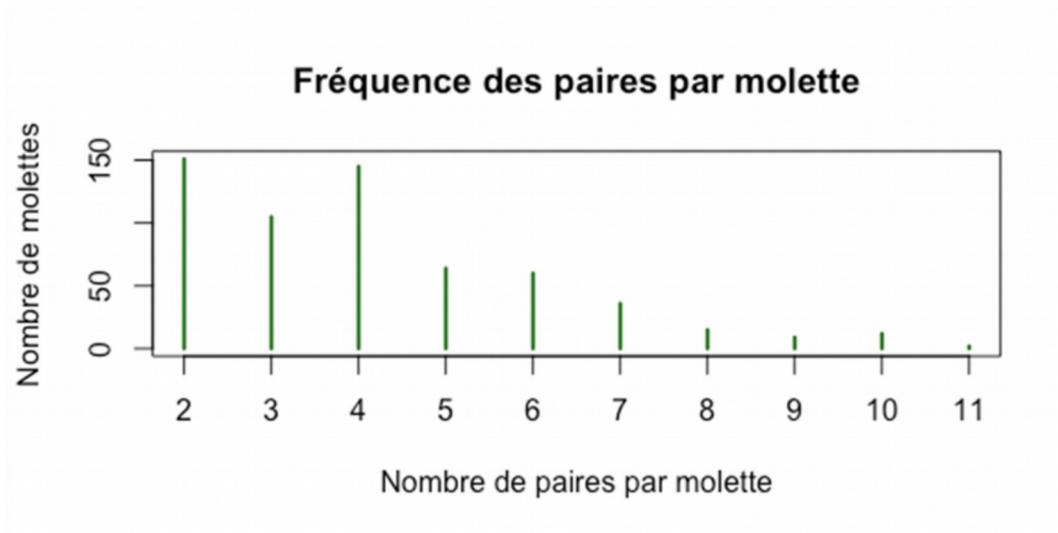
```
## [1] "Nombre de paires : 1278"
## [1] 0.81
## [1] 599 233
```

Le nettoyage et la réduction du tableau initial (élimination des paires non réellement représentées) ont conduit à un nouveau tableau de 599 lignes par 233.

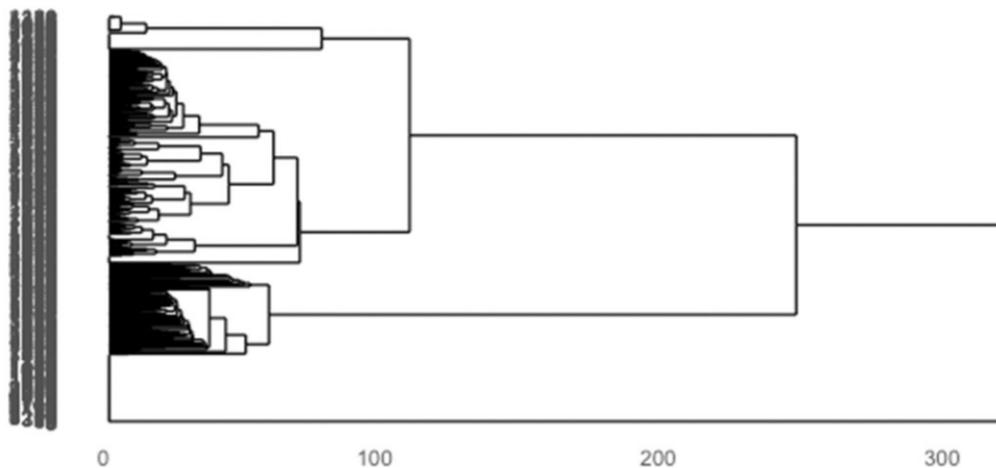
Le graphe ci-dessous montre que la plupart des molettes ne contiennent que peu d'associations de paires :



Le graphe suivant montre, réciproquement, qu'il existe un petit nombre d'associations de paires très fréquentes :



L'obtention d'un dendrogramme de CAH a aussi été testée :



Grâce aux possibilités de programmation de R, il est possible de poursuivre dans cette voie d'étude des appariements et ordonnancements des motifs, en traitant des triplets (et non plus des paires) ordonnés, voire plus (quadruplets, etc.). Mais ce type de traitements présente l'inconvénient de tendre vers une augmentation exponentielle du nombre de variables (suivant toutes les combinaisons possibles de motifs par 3, par 4, etc.), donc vers des tableaux gigantesques (même en ne retenant que les combinaisons de motifs réellement représentées).

Encore d'autres pistes ?

Les bibliothèques de R comprennent des outils dont l'application au présent problème pourrait être tentée, par exemple de recherche par chaînes de caractères : `library(stringdist)` ;

suite et développements postérieurs

les éléments présentés ici ne sont qu'exploratoires ; la poursuite de ce projet de traitement de données donnera lieu à de futures discussions dans le cadre de l'atelier. Il faut d'ores et déjà souligner le grand apport des « boîtes à outils » sur R présentées par J. Gravier lors de précédentes séances, qui ont ouvert de nouvelles perspectives au projet.

Index des projets :

Travaux universitaires :

- Ville et mouvements, le rôle des flux de circulation pour l'évolution du tissu urbain parisien aux périodes médiévale et moderne ; (thèse, Léa Hermenault) : p. 7 ; 16 ;
- Caractérisation spatio-temporelle de la ville sur le temps long. Approche diachronique de l'articulation du système intra-urbain de Noyon et de son système de villes (1^{er}-21^e s.) ; (thèse, Julie Gravier) : p. 39
- le site de Saint-Pierre-Lentin (Orléans, Loiret) : reprise des données et nouvelle synthèse ; (master 2, Laurine Guyot) p.48 ;
- Économie du bois et gestion des forêts à Naachtun (Guatemala), site maya de la période Classique (150-950 apr. J.-C.) (thèse, Lydie Dussol) p.58

Autres projets de recherche discutés dans l'atelier :

- La Plaine de Troyes : évolution d'un territoire rural des premiers agriculteurs au premier réseau villageois ; (PCR coordonné par V. Riquier, INRAP Grand-Est-Nord, SRA, UMR 8215, UMR 6289, UMR 7209) : p. 13
- Développement de l'outil de traitement de données stratigraphiques et chronologiques *Le Stratifiant* (projet actuellement mené dans le cadre de l'UMR 7041 ArcScAn, programme Archéologie du Bassin Parisien) : p. 23, 28
- programme « Corpus de la sigillée d'Argonne au Bas-Empire » équipe GAMA - UMR 7041 ArcScAn dirigé par P. Van Ossel (univ. Paris Ouest), L. Bakker (musée d'Augsbourg - Allemagne) et W. Dijkmann (ville de Maastricht – Pays-Bas) p.74

plan-type de notice de présentation de projet

titre de la présentation

intitulé du projet dans lequel s'inscrit la présentation

type de projet :

mémoire universitaire – master, doctorat, autre (préciser), projet professionnel (préciser)

porteur(s) / auteur(s) du projet :

noms, organisme de rattachement, coordonnées ; préciser auteur(s) de la présentation dans l'atelier (si pas tous les porteurs)

thématiques de l'atelier concernées par la présentation

(1, 2, 3 ou 4)

précédente présentation dans l'atelier du projet :

(si il y a lieu)

contenu de la présentation et de la discussion

commentaire ou résumé plus ou moins développé (quelques lignes au minimum) sur l'objet de la présentation, les questions ou problème à discuter ; rubrique à compléter après la présentation et la discussion, avec les principaux points et résultats de cette discussion, ainsi que les pièces jointes liées à la présentation (diaporama ou autre)

suite et développements postérieurs

par exemple : soutenance de thèse, contribution de colloque, publication... rubrique à compléter ultérieurement (pour compte rendu finalisé de fin d'année)

références

autres références liées au projet (bibliographie, sitographie...)