



Illustration de profils squelettiques et représentation de données statistiques automatisées via un SIG (logiciel QGIS)

Céline Bemilli, Emeline Le Goff, Michel Coutureau

► To cite this version:

Céline Bemilli, Emeline Le Goff, Michel Coutureau. Illustration de profils squelettiques et représentation de données statistiques automatisées via un SIG (logiciel QGIS). Carine Carpentier; Rose-Marie Arbogast; Philippe Kuchler. Bioarchéologie : minimums méthodologiques, référentiels communs et nouvelles approches : Actes du 4e séminaire scientifique et technique de l'Inrap, 28-29 nov. 2019, Sélestat, , 7 p., 2020, 10.34692/2yx4-eq45 . hal-03102525

HAL Id: hal-03102525

<https://inrap.hal.science/hal-03102525>

Submitted on 7 Jan 2021

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Illustration de profils squelettiques et représentation de données statistiques automatisées via un SIG (logiciel QGIS)

Céline BEMILLI

Inrap - UMR 7209 AASPE
celine.bemilli@inrap.fr

Émeline LE GOFF

Inrap
emeline.le-goff@inrap.fr

Michel COUTUREAU

Inrap
michel.coutureau@inrap.fr

Résumé

L'utilisation de dessins de squelettes d'animaux numérisés pour illustrer la représentation des parties squelettiques d'une espèce donnée est devenue courante en archéologie, que ce soit en termes de présence ou d'absence, de quantité (cercles proportionnels) ou de pourcentage (par ex. densité). La quantification automatique de ces divers modes de représentation est rendue possible via le recours à un système d'information géographique (SIG). Dans ce cadre, ce n'est pas la fonctionnalité de géolocalisation qui nous intéressera ici mais la possibilité d'introduire un lien dynamique entre la projection en plan des profils squelettiques et les systèmes d'enregistrement de type tableurs ou bases de données. Les analyses statistiques et les représentations peuvent ainsi s'appuyer sur les inventaires archivés. Ces analyses peuvent être automatisées, fiables et cohérentes graphiquement, selon l'ensemble du jeu de données étudié. Le temps de mise en œuvre, plus long que pour un enregistrement classique, est très largement compensé par l'important gain de temps acquis dans le traitement de ces données. De plus, ces modèles sont immédiatement réutilisables par tout archéozoologue et dans le cadre d'études ultérieures. Le choix des classes de pourcentage réalisé en DAO, implique des calculs supplémentaires (correction des classes de valeurs) pour être statistiquement fiable. En convertissant chaque partie squelettique en entité graphique, dans un logiciel de SIG, on peut établir des relations (jointures) avec les tables de données et générer automatiquement leur remplissage. En outre, ces outils statistiques intégrés calculent directement les classes de valeur à appliquer, en fonction du jeu de données en présence. Ce projet d'utiliser un SIG, avec le logiciel *open source* QGIS, offre l'opportunité de simplifier et de systématiser l'ensemble de ces étapes de travail.

Mots clés

Archéozoologie, dessin assisté par ordinateur (DAO), squelettes vectorisés, représentation graphique, quantification automatique, logiciel QGIS, analyses statistiques, modélisateur graphique

Keywords

Archaeozoology, computer-aided design (CAD), vectorised skeletons, graphic representation, automatic quantification, QGIS software, statistical analysis, graphic modeler

Référence électronique

BEMILLI, Céline, LE GOFF, Émeline & COUTUREAU, Michel. (2020). Illustration de profils squelettiques et représentation de données statistiques automatisées via un SIG (logiciel QGIS). Dans C. Carpentier, R.-M. Arbogast & Ph. Kuchler (dir.), *Bioarchéologie : minimums méthodologiques, référentiels communs et nouvelles approches : actes du 4^e séminaire scientifique et technique de l'Inrap*, 28-29 nov. 2019, Sélestat. <<https://doi.org/10.34692/2yx4-eq45>>.

Abstract

Use of digitized animal skeletal drawings to illustrate representation of skeletal parts of a given species has become common in archaeology, whether in terms of presence or absence, quantity (proportional circles) or percentage (e.g. density). Automatic quantification of these various representation models is made possible through the use of a geographical information system (GIS). In this context, it's not the geolocation functionality that will interest us here but the possibility of introducing a dynamic link between plan projection of skeletal profiles and recording systems such as spreadsheets or databases. Statistical analyses and representations can thus be based on archived inventories. These analyses can be automated, reliable and graphically consistent, depending on dataset studied. Implementation time, which is longer than for a traditional recording, is very largely compensated by the significant time saved in data processing. Moreover, these models can be immediately reused by any archaeozoologist and for subsequent studies. Choice of percentage classes made in computer-aided design (CAD) implies additional calculations (correction of value classes) to be statistically reliable. By converting each skeletal part into a graphical entity in the GIS software, relationships (joins) can be established with data tables and their filling automatically generated. In addition, these integrated statistical tools directly calculate the value classes to be applied, depending on present dataset. This project to use a GIS, with the open source software QGIS, offers the opportunity to simplify and systematise all these working step

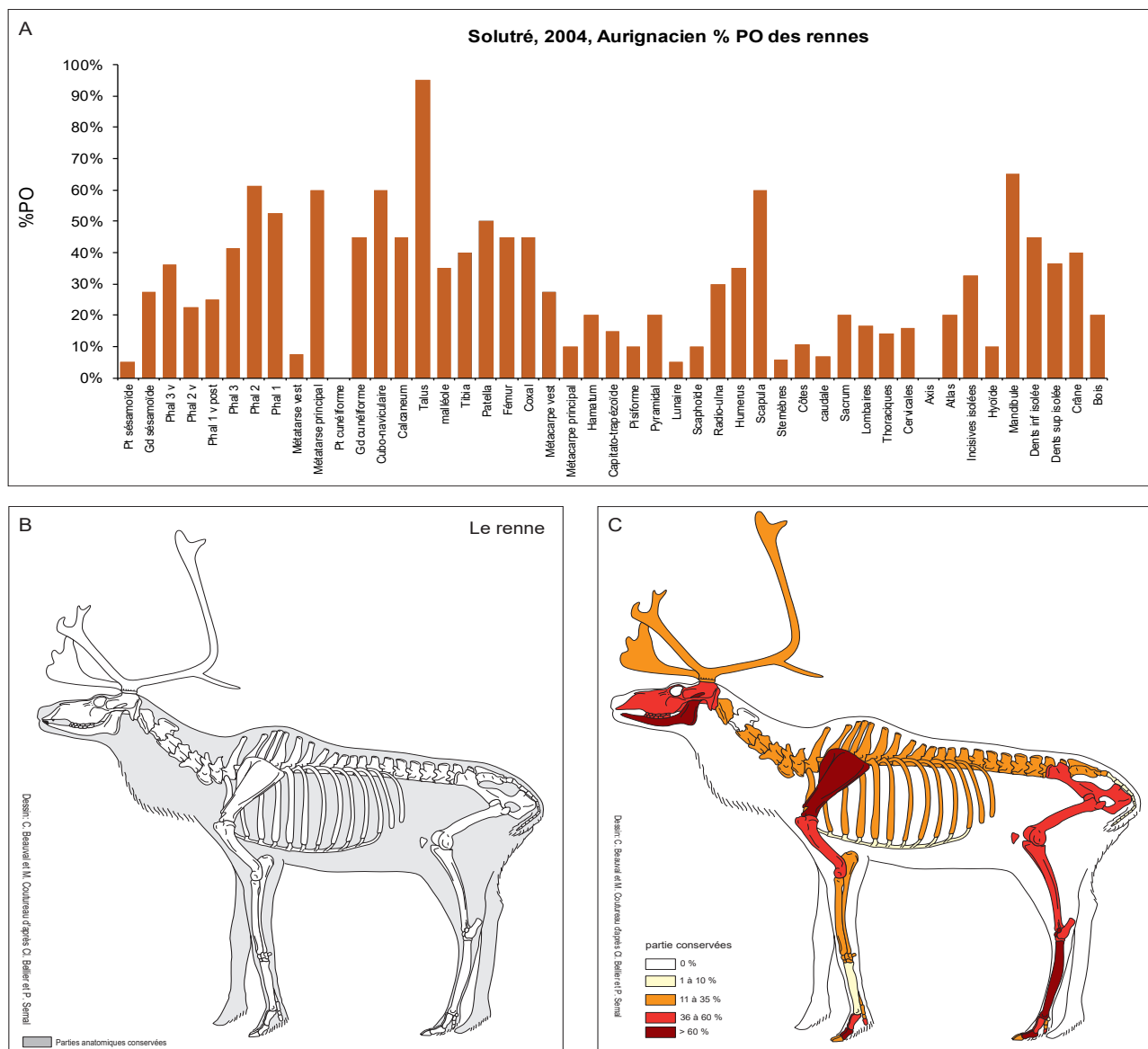


Fig. 1 - Un exemple de visualisation de données. Site de Solutré, 2004, Aurignacien (Connet et coll., 2010). A – Taux de survie (en % PO) par partie squelettique : histogramme classique ; B – Profil de squelette vectorisé ; C – Remplissage de chaque partie du squelette vectorisé figurant les taux de survie des différentes parties squelettiques. (Grégory Bayle et Céline Bemilli, Inrap)

1. Un SIG pour générer les représentations de profils squelettiques : du graphique au dessin de squelette en connexion

L'emploi d'illustrations dans les études archéozoologiques est essentiel pour en faciliter la compréhension, permettant une visualisation instantanée des données. Le développement des outils informatiques a progressivement permis de passer des simples histogrammes [fig. 1] à l'utilisation de profils squelettiques numérisés en DAO (dessin assisté par ordinateur).

Cette utilisation de profils de squelettes d'animaux s'est généralisée grâce à leur vectorisation réalisée par Michel Coutureau et collaborateurs, notamment dans le cadre de l'*Archéozoothèque*¹ et du GDR 3644 BioArchéoDat (atelier « Dessins de squelettes vectorisés »). Depuis le milieu des années 1990, des squelettes numérisés de la plupart des espèces rencontrées en contexte archéologique ont été mis progressivement à disposition de la communauté. Ce sont aujourd'hui plus de 200 squelettes numérisés de mammifères, d'oiseaux, de poissons, de reptiles et d'amphibiens qui sont disponibles. Ce corpus ne cesse de s'enrichir au grès des besoins spécifiques des archéozoologues.

1 - Banque d'images et de ressources en archéozoologie proposée par le site ArchéoZoo.org : <https://www.archeozoo.org/archeozootheque/>

En utilisant un logiciel de DAO, on peut ainsi figurer la représentation des parties anatomiques (présence, absence ou taux de survie), les zones brûlées, la localisation des traces de découpe ou des impacts de percussion. Ce travail requiert un remplissage manuel de chaque partie squelettique et, en ce qui concerne les taux de survie, des calculs préalables de correction des valeurs de classe. Ces différentes étapes s'avèrent très chronophages.

2. Le recours au système d'information géographique (SIG) : une idée simple

Le recours au SIG pour la quantification automatique de ces modes de représentations, bien qu'introduit depuis déjà une décennie par David Orton (2010), n'est pas encore répandu en France. Cet outil permet pourtant une utilisation à la fois graphique et statistique automatisée, fiable et cohérente, grâce au lien dynamique existant avec les systèmes d'enregistrement (tableurs ou bases de données). C'est sur la base de ce constat que nous avons lancé ce projet.

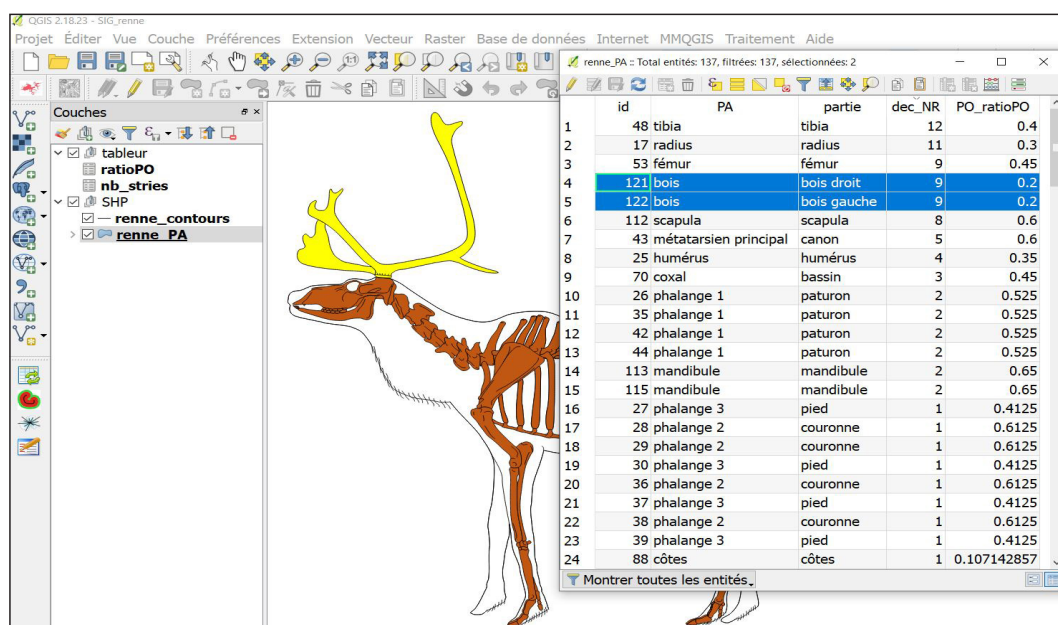
Parallèlement, on peut citer d'autres initiatives comme la création de l'outil de gestion, d'analyse et d'illustration en archéozoologie (*OGAIA*) (Creuzieux & Hofmann, 2020) et l'outil d'enregistrement et d'analyse des données archéozoologiques en Préhistoire (*TIPZOO*)² (Discamps, 2020). Ce dernier est principalement dédié aux méthodes d'enregistrement des faunes paléolithiques mais comporte également un volet de représentation graphique via un SIG.

2 - TIPZOO est accessible en ligne sur <<http://tipzoo.cnrs.fr/>>.

3. De la table de donnée à l'illustration simultanée

La mise en œuvre du projet nécessite donc l'utilisation de profils de squelette numérisés et la transformation de chaque partie squelettique en entité graphique distincte dans le logiciel QGIS, pour être ensuite enregistrée au sein d'une couche vectorielle (ou *shapefile*) par espèce. Ces parties squelettiques sont identifiées et individualisées par leur dénomination, au sein de la table attributaire (tableur) de chaque couche vectorielle. On réalise ensuite des relations (ou jointures) avec les tables de données d'inventaire, sur la base du champ contenant la codification de chaque partie anatomique [fig. 2].

Fig. 2 - Lien entre entités graphiques et système d'enregistrement. (Émeline Le Goff, Inrap)



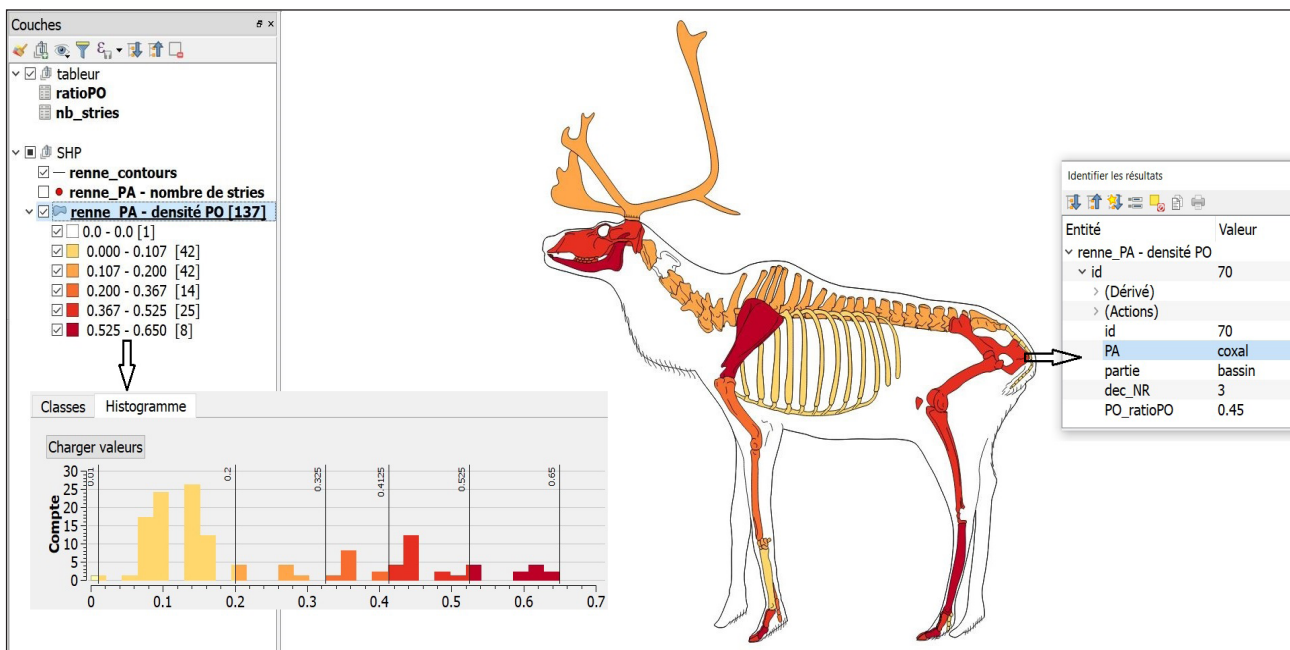
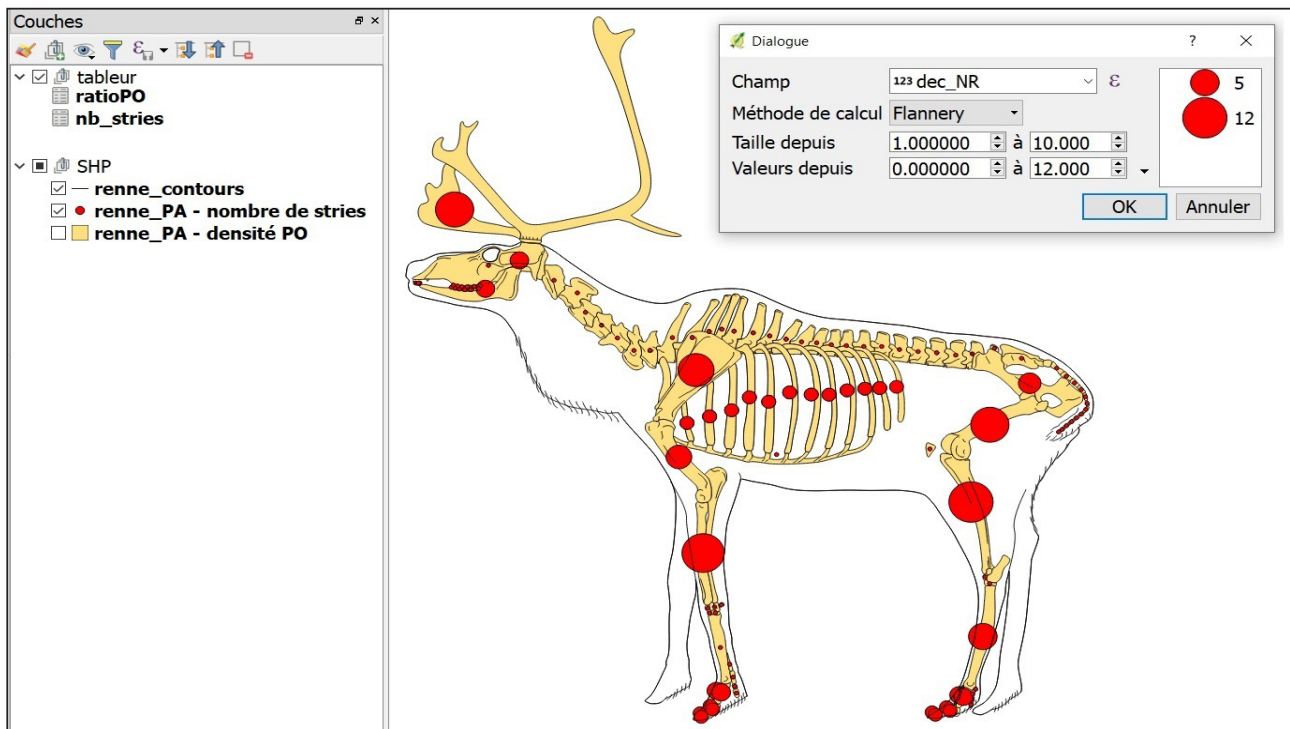


Fig. 3 - Cartographie des traces de découpe (nombre de stries). (Émeline Le Goff, Inrap)

Fig. 4 - Cartographie du taux de survie (densité en % PO), pour le renne. (Émeline Le Goff, Inrap)

Cela permet de récupérer les données d'inventaires mises à jour, puis de générer automatiquement leur remplissage : cercles proportionnels représentant des effectifs (par ex. nombre de découpe recensées par partie osseuse) [fig. 3] ou densités de couleur traduisant des ratios (par ex. taux de survie des parties squelettiques correspondant au nombre de restes décomptés rapportés à l'effectif total attendu) [fig. 4].

Ces représentations graphiques sont rendues possibles au moyen d'outils statistiques intégrés effectuant des cercles réellement proportionnels les uns aux autres, selon les bornes du jeu de données ou calculant directement les classes de valeur à appliquer. Le choix des classes créées se base sur la méthode de Jenks dite de « ruptures naturelles » qui tient compte de la distribution du jeu de données (Beguín & Pumain, 2000).

4. Déroulement du projet

Le projet est donc de convertir progressivement les figures numérisées par DAO de chaque squelette en entités graphiques dans QGIS, créant ainsi autant de couches vectorielles que d'animaux. Cette étape, longue à mettre en œuvre, servira ensuite à la création d'une base de donnée graphique, accessible et utilisable par la communauté. Elle offrira, à terme, un gain de temps en simplifiant et systématisant l'ensemble de ces étapes de travail et contribuant ainsi, efficacement, à la diffusion et la valorisation des travaux de recherche.

La qualité du dessin vectoriel en DAO est très importante. Il est notamment nécessaire d'avoir des formes fermées pour les parties osseuses afin de pouvoir appliquer des fonds colorés et de cartographier, par exemple, les taux de survie (ici donnés en %PO)³. Ces formes fermées seront intégrées dans une couche vecteur (*shape*) de type polygone. Les lignes dessinées en DAO, pour le contour des animaux ou les détails, sont intégrées dans une couche vecteur de ligne mais ne sont pas intégrées dans les analyses statistiques ; elles servent juste de « cadre » pour le rendu graphique des animaux.

3 - Le %PO est le pourcentage des parties observées. Il est égal au % des parties anatomiques observées divisé par le % des parties attendues, soit le nombre d'éléments squelettiques pour l'espèce multiplié par le nombre minimum d'individus (NMI).

4.1. Intégration du SIG

L'intégration peut se faire avec n'importe quel logiciel de SIG. À l'Inrap, c'est le logiciel QGIS qui a été choisi car libre de droit (gratuit et évolutif par l'ajout d'extensions) et très performant. Les dessins des animaux réalisés en DAO ont été enregistrés en format DXF (format interopérable pouvant être lu par un logiciel de SIG) puis intégrés dans QGIS sous la forme de deux couches : « polygonale » pour les formes osseuses fermées et « ligne » pour le contour des animaux.

Dans QGIS, l'interface géographique permettant de manipuler des systèmes de coordonnées de référence (composante géoréférencée) n'est pas du tout utilisée. C'est bien les aspects de représentation graphique automatique et d'analyse statistique qui nous intéressent ici. Pour lier les dessins de formes polygonales aux inventaires (bases de données ou tableurs), il est nécessaire d'avoir un champ commun entre l'identifiant de chaque forme et sa ligne d'inventaire dans le système d'enregistrement. Habituellement, cet identifiant est un nombre entier de 1 à n. Un identifiant numérique unique pourra, à terme, être mise en place, correspond toujours au même os.

4.2. Exploitation avec QGIS

Une fois les deux couches vectorielles créées (*shapefiles* de polygone et de ligne), il faut pouvoir identifier la typologie de chaque os de la couche polygone (par ex. bois, crâne...). Cette étape est essentielle afin de ne pas tronquer le lien avec le tableur d'inventaire et donc les objets graphiques. Une fois cette dénomination effectuée, on procède à la liaison (jointure attributaire) entre la couche de type polygone et le système d'enregistrement, sur la base d'un champ commun qui est, ici, le nom des os. Les colonnes du tableur de données sont alors jointes (liées) au *shapefile* dans QGIS, ce qui permet de générer automatiquement des cartes.

Le logiciel tient compte du jeu de données (bornes minimale et maximale, répartition des classes) pour procéder à la création de cartes :

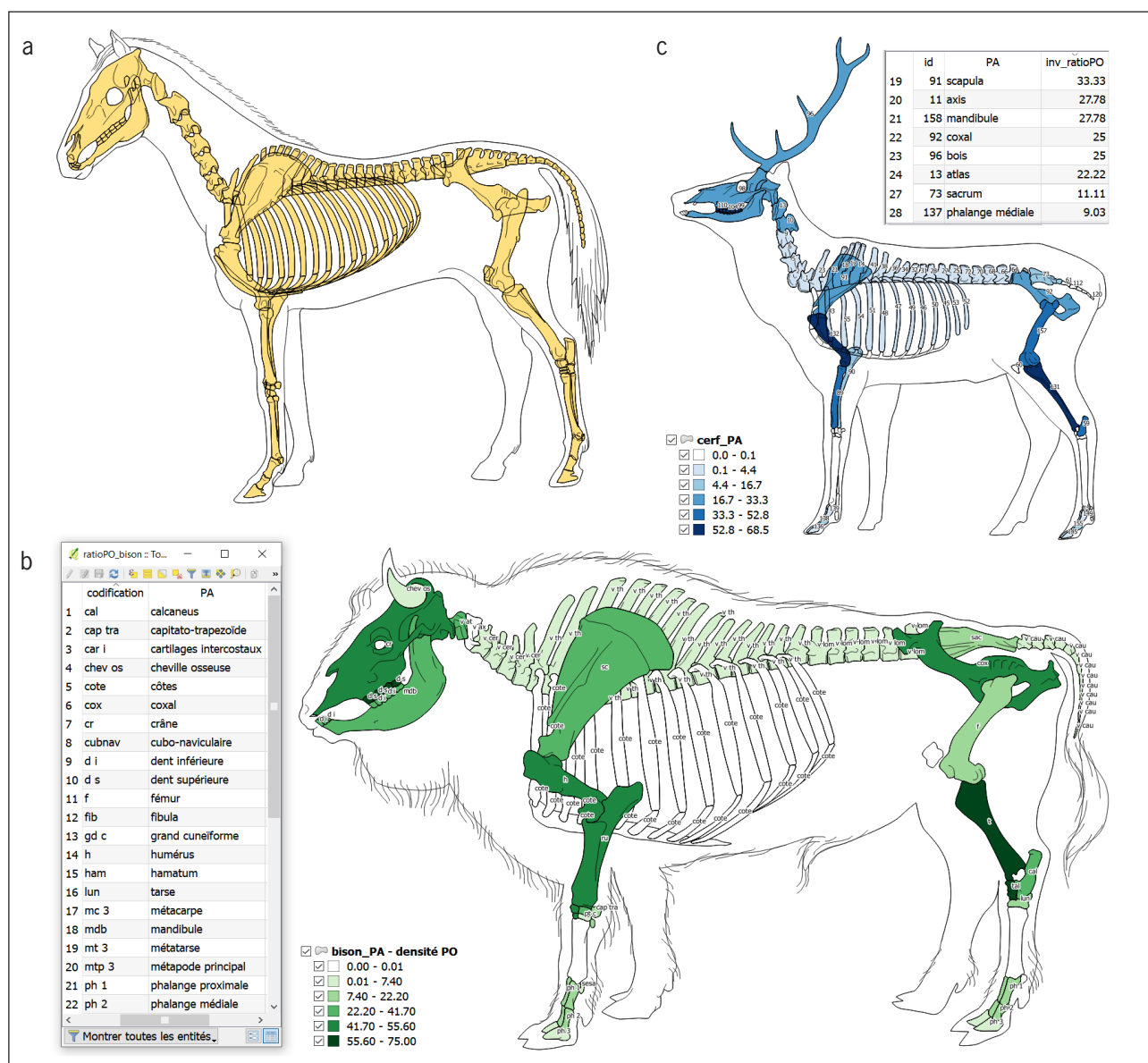
- en cercles proportionnels ; on choisit de cartographier le champ comportant des nombres en valeurs absolues (par ex. nombre de stries de découpe) ;
- de densité ; on choisit de cartographier le champ comportant une densité, c'est-à-dire des proportions et non des nombres (par ex. ratio de %PO par os).

5. Quelles suites pour le projet ?

Il est envisagé de créer un *shapefile* pour chaque dessin vectoriel d'animal déposé sur le site de l'ArchéoZoothèque [fig. 5a]. Dans chacune des tables attributaires de ces couches vectorielles, l'identification de chaque os sera validée en fonction de la nomenclature préalablement définie [fig. 5b]. À terme, une classification numérique pourra leur être associée [fig. 5c]. Enfin, il pourra être envisagé, selon les besoins et les attentes de la communauté des archéozoologues, de réaliser des scripts via le modèleur graphique de QGIS et le gestionnaire de bases de données pour automatiser la création de couches personnalisées.

Fig. 5 - Exemples de possibles manipulations à réaliser. a - Couche intégrant l'entité graphique correspondant au cheval ; b - Cartographie du taux de survie (densité en % PO), pour le bison. Ici, une codification renvoyant à un identifiant nominatif est associée à chaque os ; c - Cartographie du taux de survie (densité en % PO), pour le cerf. Ici, un identifiant unique numérique est associé à chaque os. (Émeline Le Goff, Inrap)

Par ailleurs, à l'image de TIPZOO-QGIS, il sera possible de faire le lien entre QGIS et le logiciel R, pour corréler cartographie et graphiques statistiques. Tout ceci sera mis à disposition de la communauté scientifique, qui devra, pour faire le lien avec leurs bases de données, reprendre soit le nom du type d'os, sa codification (identifiant nominatif) ou le numéro correspondant (identifiant numérique). Un protocole d'utilisation sera également proposé sous la forme d'un tutoriel pour faire le lien entre les dessins et les futurs systèmes d'enregistrement (tableurs, ...), établir des cartes de densité et générer des cercles proportionnels.



Bibliographie

BÉGUIN, Michèle & PUMAIN, Denise. (2000). *La représentation des données géographiques. Statistique et cartographie* (2^e édition). Paris : Armand Colin. 192 p.

CONNET, Nelly, BAYLE, Grégory, BEMILLI, Céline, KERVAZO, Bertrand, KONIK, Stéphane, JULIEN, Michèle, VANHAEREN, Marian, BOITARD et coll. (2012). Stratégies de subsistance des Aurignaciens de Solutré (Saône-et-Loire). Les apports de la fouille préventive de 2004. *Gallia Préhistoire*, 54, 33-65. <<https://doi.org/10.3406/galip.2012.2493>>.

CREUZIEUX, Aurélien & HOFMANN, Etienne. (2020). OGAIA : Outil de Gestion, d'Analyse et d'illustration en Archéozoologie. Dans C. Carpentier, R.-M. Arbogast & Ph. Kuchler (dir.), *Bioarchéologie : minimums méthodologiques, référentiels communs et nouvelles approches : actes du 4^e séminaire scientifique et technique de l'Inrap, 28-29 nov. 2019, Sélestat*. <<https://doi.org/10.34692/yshn-nk24>>.

DISCAMPS, Emmanuel. (2020). *TIPZOO*: a Touchscreen Interface for Palaeolithic Zooarchaeology. Making data entry and analysis easier, faster, and more reliable. 15 p. <<https://doi.org/10.31219/osf.io/aew5c>>.

ORTON, David C. (2010). A New Tool for Zooarchaeological Analysis: ArcGIS Skeletal Templates for Some Common Mammalian Species. *Internet Archaeology*, 28. <<https://doi.org/10.11141/ia.28.4>>.