

BV-RAGE

« Bureaux de Vote : Recensement, Agrégation, Géométrie, Échelles »

Responsables : *Christèle Marchand & Jessica Sainty*, LBNC

Laboratoires impliqués : LBNC, ESPACE, LIA

Résumé du projet (max 1000 caractères)

Le projet BV-RAGE fait suite aux projets ABSTER et CIVIDEN. Il vise à étudier les Bureaux de Vote comme entité socio-spatiale généralisée (big data) de référence pour l'analyse des pratiques de vote. Il se décline en plusieurs volets :

- élaboration de méthodes de croisement d'informations géographiques mêlant contraintes spatiales et données sociales en vue d'évaluer et d'enrichir le bureau de vote (BV) comme entité statistique,
- test et validation d'algorithmes de construction automatique des partitions spatiales de BV, en phase avec le découpage électoral administratif ou guidés par des critères d'optimalité socio-spatiale statistique,
- mise à l'épreuve des méthodes et des algorithmes élaborés sur deux corpus : Avignon (élaboration) et Val de Marne (application), dans le cadre du partenariat avec le Ministère de l'Intérieur,
- étude de l'impact des BV sur les pratiques de vote par analyse factorielle des questionnaires sortie des urnes croisée avec nos représentations des BV.

Contexte, positionnement et objectifs

Le projet BV-RAGE fait suite aux travaux fructueux interdisciplinaires (cf. publications co-signées par les trois laboratoires en annexe) réalisés au sein d'Agor@ntic dans le cadre :

- du projet ABSTER, où il était question de la spatialisation de l'abstention et des modes de recensement et d'analyse de l'information individualisée sur le vote,
- du projet CIVIDEN, où l'échelle pertinente de mesure du vote en lien avec les effets de distance ainsi que les profils des abstentionnistes ont été notamment appréhendés,
- du stage de Soumaya Yahiaoui, qui a débouché sur un outil automatique de délimitation géométrique des bureaux de vote à partir des textes juridiques.

En lien avec ces trois actions, et plus particulièrement avec la dernière, une nouvelle opportunité se présente aujourd'hui, grâce à une évolution du contexte national, plaçant le projet BV-RAGE en parfaite adéquation avec l'appel à projet d'Agorantic, mais également du point de vue de la demande nationale. En effet, le Ministère de l'Intérieur a récemment lancé une **action collaborative** (un *data camp* avec un *Github* du cadastre électoral mis à disposition par la Préfecture du Val de Marne) associant l'administration électorale, des entreprises privées et des

laboratoires de recherche pour chercher à simplifier la cartographie des bureaux de vote : <https://www.etalab.gouv.fr/datacamp-cadelect>. L'idée est à terme de généraliser et d'automatiser la gestion et la cartographie des bureaux de vote à l'échelle nationale, en commençant sur un exemple concret, le Val de Marne. Vue l'ampleur de l'information cadastrale et réticulaire au niveau national, ce projet relève des **données massives géographiques**, à la fois du point de vue de la quantité de données, que de la nécessité de leur traitement par des systèmes d'information géographiques robustes, voire des systèmes de gestion, de désambiguïsation ou d'apprentissage des textes. Pour des raisons de disponibilité et de budget, nous n'avons pu nous rendre à cette réunion à Paris, mais nous avons établi la connexion avec Etalab, dont les animateurs se montrent particulièrement intéressés par nos recherches. L'obtention du financement d'Agor@ntic nous permettrait, entre autres choses, de participer à ce projet, avant de rechercher d'autres financements auprès de ces partenaires. Forts de nos échanges interdisciplinaires, nous ajoutons à ce projet méthodologique une **dimension problématique plus socio-spatiale** que nous développons maintenant.

Résultats attendus et caractère innovant de la recherche

L'entité spatiale « Bureau de Vote » revêt deux caractéristiques majeures :

- elle constitue le grain le plus fin de découpage administratif qui fait foi pour récolter les votes lors des élections ;
- elle donne accès aux données nominatives par électeur (nom et prénom, adresse, date de naissance et participation aux scrutins) ce qui permet une analyse fine des comportements d'abstention.

En ce qui concerne le premier aspect de la problématique, le partitionnement se fait en lien avec la géographie des villes et des communes et leur capacité à organiser les élections aux lieux de vote. Il est donc le résultat d'un choix purement arbitraire, en fonction de considérations administratives et peut d'ailleurs être utilisé à des fins politiques (Richomme, 2013). L'ensemble des BV à l'échelle nationale doit constituer une partition complète, sans trou ni superposition, de façon à ce que chaque citoyen puisse voter une seule fois. Comme nous l'avons montré dans nos publications, la cartographie des BV est actuellement très incomplète, notamment pour les communes de taille moyenne. Elle est aussi imprécise car très délicate à mettre à jour. Il persiste une quantité très importante d'erreurs sémantiques ou topologiques. D'où le premier objectif très pratique du Ministère : automatiser la construction des découpages électoraux pour réduire les coûts et rationaliser le processus de partitionnement à l'échelle nationale. Possédant déjà une expertise avancée sur cette question (Yahiaoui et al., 2014, 2015, Josselin et al., 2016) et un *plugin* en *Python* opérationnel dans l'environnement libre *QGIS*, nous sommes à même de **tester notre outil sur les données mises à disposition par la Préfecture du Val de Marne et rencontrer ces partenaires pour leur faire part de nos résultats**.

D'un autre côté, le BV, contrairement aux IRIS ou aux communes, porte une information électorale individualisée sur la participation d'un intérêt indéniable. Cette information est accessible aux partis politiques qui l'intègrent dans leurs campagnes de proximité notamment entre les deux tours d'une élection. Ainsi, se baser exclusivement sur une approche mixant structures géométriques et organisation des élections peut sembler très réducteur, car trop fonctionnelle. D'autre part, considérer le BV comme un découpage « absolu » peut largement

éluder l'usage réel qui en est fait lors des campagnes électorales de terrain. Cet usage est complexifié par l'hétérogénéité des effectifs de votants par BV et, comme nous l'avons montré, une proportion non négligeable de votants n'habitant pas dans le périmètre de leur BV. Autant de facteurs qui rendent incomparables les indicateurs statistiques qui découlent des BV avec les autres entités (Louvet et al., 2016).

Ce problème est connu sous les termes d'*upscaling*, d'*effet écologique d'agrégation* en sociologie (Robinson, 1950, King, 1997), de *paradoxe de Simpson* (1951) en économie, de *MAUP* (Modifiable Areal Unit Problem, Openshaw, 1984) en géographie ou de *COSP* (Change of Support Problem, King, 2004) plus généralement dans les sciences de l'environnement. Si l'on cherche également à désagréger des informations disponibles à un niveau plus général comme l'IRIS, se pose alors le problème encore plus délicat du *downscaling* (Josselin & Poss, 2016), à savoir, l'application d'hypothèses statistiques pour estimer des valeurs d'indicateurs sociaux au niveau plus fin qu'est le BV. D'autres croisements d'informations spatialisées sont possibles, avec le carroyage national de l'INSEE par exemple.

Pour étudier l'éventuel impact du découpage administratif des BV sur le déroulement des campagnes électorales et en particulier sur la mobilisation des électeurs, la qualité du découpage cartographique est cruciale.

Nous proposons donc **d'étudier la partition officielle administrative du bureau de vote** en la questionnant à différents niveaux :

- *Quelle est la variabilité des BV (taille, forme) ? Quel impact a la partition électorale administrative sur l'analyse des votes ? En agrégeant l'information, observe-t-on des différences significatives si l'on fait des enquêtes ou des hypothèses sur les votes et l'abstention ?*
- *Avec la connaissance plus fine des quartiers dont nous disposons grâce en particulier aux enquêtes locales menées par le LBNC, les informations officielles des BV sont-elles pertinentes ? Le BV est-il une entité administrative neutre pour la compréhension des comportements des électeurs en fonction de leurs profils sociaux ou, au contraire, existe-t-il des effets de bord ?*
- *Pourrait-on se baser sur des informations socio-spatiales (IRIS, carroyage INSEE, enquêtes sur les quartiers, par exemple) pour enrichir les informations à l'échelle des BV ? Quelles méthodes de désagrégation et d'intersection spatiales nous permettent d'inférer une connaissance fiable à l'échelle de ces nouveaux BV ? Comment évaluer ces estimations ?*
- *Au delà, peut-on évaluer l'impact d'une redéfinition de l'emprise des bureaux de vote, et de quelle façon ? Existe-t-il une partition qui serait plus optimale du point de vue de l'homogénéité des BV susceptible de faciliter la participation ? Quel en serait l'impact sur le vote par correspondance ou l'éventuelle introduction du vote électronique ?*

Pour travailler efficacement sur ce second volet du projet, nous ne pouvons d'emblée considérer le corpus d'Avignon, sur lequel nous disposons de nombreuses données depuis plusieurs années. Notons que dans le cadre de l'ANR A/CoV – à laquelle sont associées Christèle Marchand et Jessica Sainty, il est prévu que le LBNC réalise des enquêtes au moment des élections présidentielles et législatives 2017 qui compléteront les données disponibles. Pour l'évaluation de l'impact d'un redécoupage en BV, il s'agit de développer des méthodes d'extrapolation pour typer les BV à partir de quelques indicateurs sociaux (types d'habitats, caractéristiques sociales des populations qui y vivent et leur hétérogénéité) ou spatiaux (distance aux facilités urbaines : centre-ville, service public, lieu de convivialité...). L'objectif est de

proposer des **analyses différentielles des résultats électoraux selon les types de BV** (sorte d'indicateur d'incertitude ou de non vote plus ou moins significatif selon les BV).

En prolongement, si l'on parvient à croiser des données sociales à différentes échelles et à identifier des relations entre ces profils et les comportements face au vote, ces analyses devraient nous permettre d'appréhender la **question de la mixité sociale et son effet sur la participation aux élections** et plus généralement sur les pratiques de vote.

Une fois ces méthodes testées sur Avignon, nous proposons de les **appliquer aux données du Val de Marne et d'évaluer leur niveau de généricité** et les **contraintes et coûts éventuels induits par leur mise en œuvre nationale**, dans un contexte de *big data*.

Dimension interdisciplinaire (champs disciplinaires associés) et cohérence par rapport à l'axe identitaire « Culture, Patrimoines, Sociétés numériques »

Dans ce projet, l'interdisciplinarité va de soi. Elle est pratiquée depuis plusieurs années et a donné de nombreuses publications communes et co-signées. Nous souhaitons continuer à travailler de concert et co-encadrer deux stagiaires sur ce sujet, l'un(e) en développement SIG, l'autre en sociologie du vote.

Le projet BV-RAGE s'inscrit dans [Agor@ntic](#) par la généralisation de l'approche à tous les bureaux de vote français et par l'approche collaborative prônée par le Ministère de l'Intérieur (Etalab).

Voici la liste des collègues impliqué(e)s dans ce projet :

LBNC : *Marchand-Lagier Christèle, Sainty Jessica*

ESPACE : *Genre-Grandpierre Cyrille, Josselin Didier*

LIA : *SanJuan Eric*

Partenariats extérieurs envisagés (en cours et à venir)

Le réseau national Etalab du gouvernement pour la question plus particulière des bureaux de vote est le partenaire privilégié du projet : <https://www.etalab.gouv.fr/>

La rencontre avec d'autres partenaires du *Data Camp* (entreprises, autres laboratoires) nous permettra d'envisager des projets de plus grande ampleur dans un second temps.

Budget prévisionnel (dont financements complémentaires)

6000 € demandés, répartis comme suit :

- 1500 € de frais de mission (environ 5 missions sur Paris dans le réseau Etalab)
- 2250 € de stage (1 stage de 4 mois en science politique) ; le stage consistera à rechercher les critères sociaux qui vont guider l'évaluation et la construction de partitionnement optimal
- 2250 € de stage (1 stage de 4 mois en géomatique) ; le but du stage est de peaufiner et mettre en œuvre les méthodes automatiques déjà développées sur les données du Val de Marne notamment.

Publications interdisciplinaires des équipes dans le cadre d'Agor@ntic

- Bernard Laura, Marchand-Lagier Christèle, Josselin Didier, Louvet Romain. 2016, Mal-inscription de proximité et effets sur la participation électorale en région PACA. *Pôle Sud - Revue de science politique de l'Europe méridionale*, 2016. <hal-01338069>
- Josselin Didier, Poss Michael. 2016, Exploring Accurate Spatial Downscaling using Optimization. *Spatial Accuracy'2016*, Jul 2016, Montpellier, France. Proceedings of *Spatial Accuracy'2016*. <hal-01338085>
- Josselin Didier, Yahiaoui Soumaya, Eric Sanjuan, Marchand-Lagier Christèle. 2016, Automatic Delineation of Polling Station Areas using Legal Texts. *AGILE'2016*, Jun 2016, Helsinki, Finland. Proceeding of *AGILE'2016* short papers, 2016. <hal-01338079>
- Louvet Romain, Aryal Jagannath, Josselin Didier, Marchand-Lagier Christèle, Genre-Grandpierre Cyrille. 2016, The polygon overlay problem in electoral geography. *Spatial Accuracy'2016*, Jul 2016, Montpellier, France. Proceedings of *Spatial Accuracy'2016*. <hal-01338084>
- Yahiaoui Soumaya, Josselin Didier, San Juan Eric, Marchand-Lagier Christèle. 2015, Délimitation géométrique semi-automatique des bureaux de vote à partir de textes juridiques et d'information géographique numérique : enjeux et difficultés. *Congrès AFSP 2015*, Jun 2015, Aix en Provence, France. 14 p., <<http://www.congres-afsp.fr/>>. <hal-01251317>
- Yahiaoui Soumaya, Josselin Didier, Marchand-Lagier Christèle, Douvinet Johnny. 2014, Vérification et (re)construction automatiques des limites des bureaux de vote par l'étude des textes juridiques. *SAGEO 2014*, Nov 2014, Grenoble, France. 2014. <hal-01100504>

Autres références

- King G. (1997). A solution to the ecological inference problem. reconstructing individual behaviour from aggregate data. Princeton University Press.
- King G., Rosen O., Tanner A. M. (Eds.). (2004). Ecological inference. new methodological strategies. Cambridge University Press.
- Openshaw S. (1984). The modifiable areal unit problem. Norwich: Geo Books, CAT-MOG 38.
- Richomme O. (2013). A permanent republican majority? consequences of the 2011 gerrymandering. *Hérodote*, vol. 149, p. 129-148.
- Robinson W. (1950). Ecological correlations and the behaviour of individuals. *American Sociological Review*, vol. 15, p. 351-357.
- Simpson E. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society - Series B (Methodological)*, vol. 13-2, p. 238-241.