



HAL
open science

Une méthode d'intégration de Bases de Données Géographiques pour analyser les interactions entre agricultures et ressources en eau sur le Bassin Adour-Garonne

Maud Balestrat, Olivier Therond

► To cite this version:

Maud Balestrat, Olivier Therond. Une méthode d'intégration de Bases de Données Géographiques pour analyser les interactions entre agricultures et ressources en eau sur le Bassin Adour-Garonne. Colloque International et Interdisciplinaire, " Dynamiques environnementales, politiques publiques et pratiques locales: quelles interactions? ", Jun 2013, TOULOUSE, France. hal-02744819

HAL Id: hal-02744819

<https://hal.inrae.fr/hal-02744819>

Submitted on 3 Jun 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.

Une méthode d'intégration de Bases de Données Géographiques pour analyser les interactions entre agriculture et ressources en eau sur le bassin Adour-Garonne

Maud Balestrat¹, Olivier Théron²

1 Géographe, maud.balestrat@toulouse.inra.fr Tel +33 (0) 5 61 28 57 44 Fax +33 (0) 5 61 73 55 37

2 Agronome, olivier.therond@toulouse.inra.fr Tel +33 (0) 5 61 28 50 48 Fax +33 (0) 5 61 73 55 37

INRA, UMR AGIR, Chemin de Borde-Rouge 31 326 Castanet-Tolosan

Résumé

En France, le bassin Adour-Garonne comptabilise la plus importante surface irriguée en grandes cultures. En période d'étiage estival, l'irrigation y représente plus de deux tiers des prélèvements en eau. Dans ce bassin hydrographique, qui fait face à un déficit structurel de sa ressource en eau, les gestionnaires sont régulièrement confrontés à des problèmes de répartition de l'eau entre les usagers. Avec la Loi sur l'Eau et les Milieux Aquatiques de 2006, le renforcement des réglementations ne fait qu'exacerber les conflits d'usages et les besoins des institutions en termes d'outils et procédures de gestion efficiente et durable. La modélisation des relations entre agriculture et ressources en eau peut fournir des éléments de compréhension utiles pour l'action publique. La mise à disposition croissante de données spatiales permettant de nourrir ce processus de modélisation offre de nouvelles opportunités. L'objectif de cette communication est de présenter une méthode d'intégration de Bases de Données Géographiques destinée à l'analyse des relations entre agriculture et ressources en eau. La méthodologie mise en œuvre se base à la fois sur l'hybridation de connaissances expertes en agronomie, en hydrologie et sur la structure des territoires agricoles irrigués afin d'établir des liens entre des données de nature hétérogène. Ces relations sont formalisées sous la forme de probabilités de correspondances qui sont ensuite utilisées au travers de règles d'appariement, pour lier les données entre elles. L'objectif est d'établir des relations qui respectent les grandes cohérences agronomiques, agricoles, hydrologiques et hydrauliques du territoire.

Mots clés : bassin hydrographique, étiage estival, gestion quantitative de l'eau, Base de Données Géographiques, appariement, intégration

1. Introduction : les enjeux liés à la gestion quantitative de l'eau en Adour-Garonne, des besoins en systèmes d'information intégrés

A l'échelle nationale, c'est dans le Bassin Adour-Garonne (BAG)¹ que l'on trouve les plus importantes surfaces irriguées en grandes cultures. Aquitaine, Midi-Pyrénées et, dans une moindre mesure, Poitou-Charentes arrivent en tête des régions productrices de maïs qui compose l'essentielle des surfaces irriguées du sud-ouest (MAAF 2013). Selon les chiffres moyens issus des déclarations pour les redevances, de 2003 à 2010, l'irrigation représente 41% des volumes prélevés annuels devant l'eau potable (33%) et l'industrie (26%). En période d'étiage estival (1^{er} juillet - fin octobre) la part des prélèvements agricoles passe à 70% pour l'irrigation et 15 % pour chacun des deux autres usages (AEAG 2012a). Dans ce bassin, du fait d'un déficit structurel entre ressources et demandes en eau estimé à 250 millions de m³ (AEAG 2012b), les gestionnaires sont régulièrement confrontés à des situations de pénuries d'eau et à des problèmes de répartition de l'eau entre les différents usagers. La mise en œuvre de la Loi sur l'Eau et les Milieux Aquatiques de 2006 (LEMA) devait favoriser le passage d'une logique de gestion de l'offre, centralisée, à une gestion de la demande, déconcentrée et participative. Dans les faits, les difficiles négociations entre services de l'Etat et représentants de la profession agricole ont conduit à revoir à la baisse les objectifs initiaux et au maintien d'une logique de gestion de l'offre (Debril & Therond 2010).

Pour appuyer les décisions et consolider leur discours, les gestionnaires de l'eau sont particulièrement confrontés au problème du peu d'informations structurées et homogènes permettant d'évaluer, en début et au fil de la saison estivale, les prélèvements agricoles dans chaque ressource en eau (cours d'eau, plans d'eau, barrages, nappes souterraines). La modélisation des relations entre agriculture et ressources en eau peut fournir des éléments de compréhension particulièrement utiles pour guider l'action des services publics et des gestionnaires de ressources. La production et la mise à disposition croissante de données spatiales, permettant de nourrir ce processus de modélisation, et la généralisation de l'utilisation des Systèmes d'Information Géographique (SIG) offrent de nouvelles opportunités (Moody & Ast 2012). Cependant, les Bases de Données Géographiques (BDG) disponibles sont produites par diverses institutions à différents niveaux d'organisation (national, régional,

¹ Le Bassin Adour-Garonne est l'un des 6 grands bassins hydrologiques français couvrant le quart sud-ouest du territoire métropolitain et d'une superficie d'environ 116 000 km² (soit 1/5^{ème} du territoire national). Les principaux cours d'eau qui l'alimentent sont la Garonne, l'Adour, la Charente et la Dordogne.

départemental, bassin versant, local) et suivant une logique propre à leur domaine d'action. Il en résulte une grande hétérogénéité et des contraintes d'interopérabilité entre Bases de Données (BD). L'un des enjeux actuels consiste à concevoir un Système d'Information (SI) intégré à l'échelle d'un grand territoire hydrologique qui offre la possibilité d'appréhender, voire de quantifier, les interactions entre systèmes de culture et de production, ressources en eau (de surface et souterraines) et dispositifs législatifs et réglementaires relatifs à la limitation des prélèvements en eau pour l'irrigation.

L'objectif de cette communication est de présenter une méthode d'appariement de données spatiales multi-sources permettant de représenter les relations (structurelles et annuelles) entre agriculture et eau à l'échelle du BAG et qui respecte les grandes cohérences agronomiques, agricoles, hydrologiques et hydrauliques du territoire. Nous exposons d'abord les différentes BDG mobilisées et les freins liés à leur intégration (section 2). Dans un deuxième temps, nous présentons, d'un point de vue théorique (section 3), la démarche d'appariement développée puis, sa mise en œuvre (section 4) pour produire un SI intégré à l'échelle du bassin Adour-Garonne.

2. Proposer une Base de Données Géographiques des relations eau/agriculture à l'échelle du BAG

Afin de déterminer qui prélève (exploitation agricole), en quelle quantité (volumes prélevés) et dans quelle ressource (cours d'eau, plan d'eau, nappe souterraine) plusieurs BDG peuvent être mobilisées.

2.1. Données à disposition pour caractériser la pression agricole et évaluer la disponibilité des ressources en eau

Localiser les exploitations agricoles qui irriguent et quantifier les superficies irriguées

L'un des besoins des gestionnaires de l'eau est de disposer de connaissances fines sur la nature et la distribution spatiale des systèmes de culture irrigués. L'exploitation du Référentiel Parcellaire Graphique (RPG), livré tous les ans pour la totalité du territoire national par l'Agence de Services et de Paiement (ASP), peut fournir un ensemble d'informations sur les caractéristiques des exploitations agricoles qui irriguent (structure, distribution spatiale, etc.) (Figure 1) et sur les superficies agricoles irriguées (nature de la sole cultivée, superficies des différentes cultures, distribution spatiale, etc.). En effet, chaque année, l'exploitant saisit les îlots² cultivés sur fond d'orthophotos au 1/5 000 et décrit leur occupation culturale pour toutes les surfaces de son exploitation, si au moins une d'entre elle est concernée par des mesures d'aide de la Politique Agricole Commune.

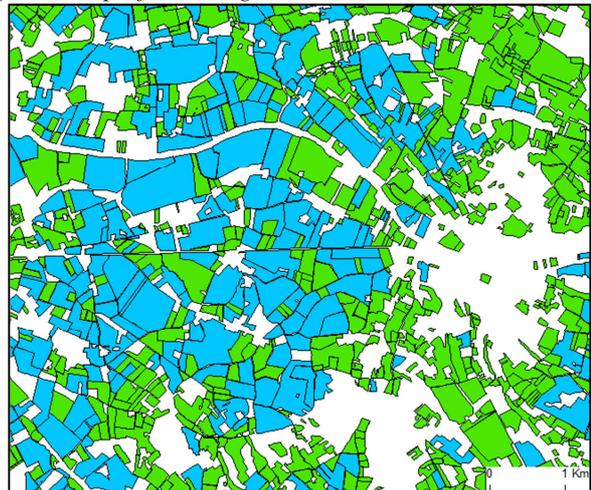


Figure 1 : Extraction du RPG d'îlots déclarés irrigués (bleus) et non irrigués (verts) (source ASP)

² Un îlot est un ensemble de parcelles culturales contiguës exploité par un même déclarant et entouré par des limites permanentes (chemin, route, ruisseau) ou par d'autres exploitations (ASP 2011)

Déterminer en quelle quantité les irrigants prélèvent et dans quel type de ressource

Chaque année, l'ensemble des prélèvements sur la ressource en eau (industriels, de production d'eau potable ou d'irrigation) au delà d'un seuil donné (e.g. 7 000 m³ en zone de répartition des eaux³) sont soumis à déclaration et redevance auprès des Agences de l'Eau (AE)⁴. A partir de ces informations, à l'échelle de chacun des 6 grands bassins hydrologiques, des données sur les prélèvements agricoles (natures et caractéristiques des irrigants, types d'ouvrages de prélèvements, modes d'irrigation, etc.) sont produites, notamment sur les volumes prélevés déclarés par type d'intervenant agricole (collectif ou individuel) en fonction de la nature des ressources en eau. Ces données sont pour la plupart mises à disposition via le Système d'Information sur l'Eau (SIE)⁵.

Localiser les différentes ressources en eau exploitées pour l'irrigation

Pour localiser et caractériser les différentes ressources en eau (cours d'eau, plans d'eau, nappes phréatiques et captives) plusieurs BDG peuvent être mobilisées. La BD CARTHAGE®, produite par l'Institut Géographique National (IGN), enrichie par le ministère en charge de l'environnement et les Agences de l'Eau (AE) et mise à disposition sur le site du SANDRE (Service d'Administration des Données et Référentiels sur l'Eau), offre une description relativement exhaustive du réseau hydrographique français. A l'échelle nationale, la BD TOPO®, produite par l'IGN, fournit de l'information sur les surfaces en eau qui peut permettre de localiser une grande partie des plans d'eau destinés à l'irrigation. Enfin, le référentiel des masses d'eau souterraines est livré dans la Base de Données de Réseau Hydrogéologique français (BDRHF®), accessible via le SIE. Elaboré par le Bureau de Recherches Géologiques et Minières (BRGM) et les AE pour les besoins de la Directive Cadre sur l'Eau (DCE), il offre une caractérisation et une délimitation précise des nappes souterraines sur la base de critères géologiques et hydrogéologiques.

La figure 2 présente de manière schématique l'objectif de croisement des couches d'information géographique, contenues dans ces trois grands types de BDG, au sein d'un SIG destiné à représenter les relations eau-agriculture à l'échelle du BAG.

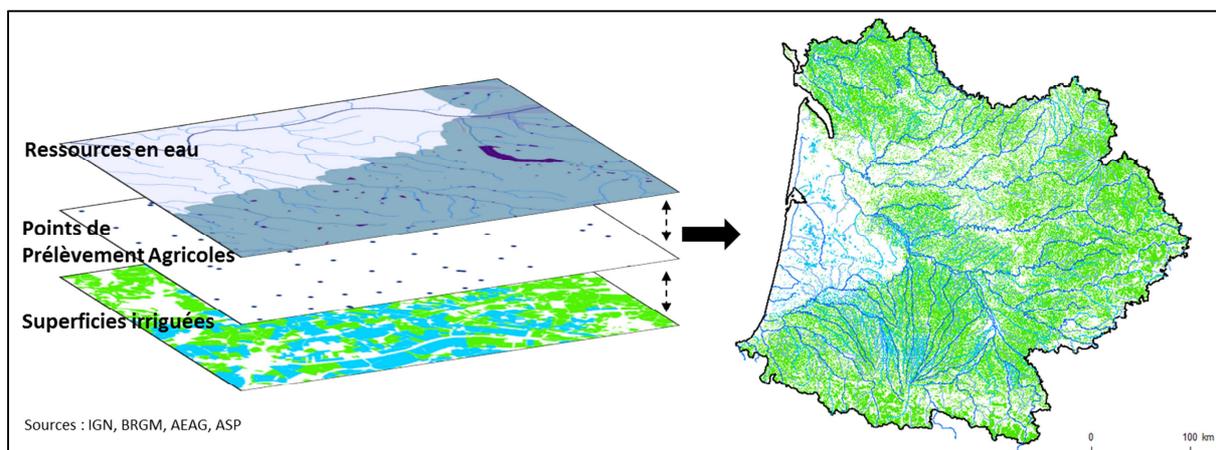


Figure 2: Couches d'information géographique mobilisées pour représenter les relations eau-agriculture à l'échelle du BAG

Pour la mise en œuvre de traitements nécessaires à l'appariement de ces bases (CF. 4.2.), un ensemble de données complémentaires ont dû être mobilisées. A titre d'exemples, la Base de Données Géographiques des Sols de France (BDGSF) au 1/1 000 000^e, mise à disposition par l'unité Infosol de l'INRA Orléans (assortie d'un ensemble de variables complémentaires estimées sur la base de règles de pédo-transfert), a été retenue pour décrire le contexte pédologique. Pour décrire le contexte climatique, ce sont les données observées journalières fournies par Météo-France et qui renseignent, sur une grille de points météorologiques d'environ 14 x 10 km (en latitude/longitude), sur la pluviométrie, la température minimale et maximale, le rayonnement global et l'évapotranspiration potentielle. La BD ALTI® a été exploitée pour décrire la topographie ou encore la BD NYME® qui correspond à la BD toponymique géographique de référence. Plusieurs zonages spatiaux ont également été intégrés : découpages administratifs issus de la BD TOPO® ; découpages physiques issus des segments Arvalis (correspondant à des petites régions agricoles homogène en terme de systèmes de culture et construites à dire d'experts), etc.

³ « Les zones de répartition des eaux sont des zones où est constatée une insuffisance, autre qu'exceptionnelle, des ressources par rapport aux besoins » (Eaufrance)

⁴ Loi n° 2006-1772 du 30 décembre 2006 (Legifrance 2006)

⁵ « Le Système d'Information sur l'Eau est un dispositif partenarial des principaux acteurs publics du domaine de l'eau qui organise la collecte, le stockage, la valorisation et la diffusion des données sur l'eau, les milieux aquatiques et leurs usages » (Eaufrance)

A un niveau local ce type de données est plus ou moins exploité, toujours partiellement, par exemple, dans le BAG, pour construire les plans de gestion des eaux qui définissent les règles de partage de l'eau entre usages et milieux pendant la période d'étiage à l'échelle d'un bassin versant (e.g. Tarn). Malgré la quantité de données à disposition, il n'existe pas d'étude dont l'objectif est d'établir des relations entre caractéristiques des systèmes agricoles irrigués, prélèvements pour l'irrigation et impacts sur les ressources en eau de manière homogène à l'échelle de grands territoires hydrologiques comme celui du bassin Adour-Garonne. Au-delà du fait que les BD manquent parfois encore de précision et d'exhaustivité, les principaux freins à l'élaboration de ce type d'analyse résident dans l'hétérogénéité des données existantes et la non interopérabilité des BD.

2.2. Une grande diversité et hétérogénéité des BDG à appairer qui contraignent le développement d'un SI intégré

Les sources d'hétérogénéités du contenu des BD sont diverses. Parent & Spaccapietra (1996) distinguent trois types d'hétérogénéité liés aux : modèles de données ; puissances d'expression ; choix de modélisation. Un *modèle (ou schéma) de données* est un ensemble de concepts et de règles de composition de ces concepts permettant de décrire des données (Gardarin 2003). La *puissance d'expression* représente la richesse sémantique de la description offerte par la BD. Enfin, les BD sont structurées selon des logiques propres aux domaines d'action respectifs mais également selon *différentes représentations du monde* propres à chaque modélisateur. Un même modèle de données, ce malgré des spécifications très strictes, peut ainsi être traduit de différentes façons en fonction des perceptions qu'en aura le concepteur en charge de la réalisation de la BD. Dans le cadre de notre étude, ceci est d'autant plus « vrai » que nous mobilisons un ensemble de données géographiques qui présentent la spécificité de concerner des informations de thèmes distincts (hydrographiques et agricoles). En outre, au sein d'une même thématique, Laurini (1996, cité par Gesbert 2005) soulève que du fait du partitionnement de l'espace, des choix de résolution spatiale, de projection et de représentations, des BDG ont de forte chance de ne pas se raccorder ou superposer précisément. Le tableau n°1 restitue différents critères d'hétérogénéité qui peuvent être identifiés pour les données à appairer dans le cadre de notre étude (nature de l'information, maille d'analyse, résolution spatiale, fréquence de mise à jour, etc.).

Thématique	Nom BD	Fournisseur	Information thématique	Précision	Temporalité	Fréquence MAJ	Type	Zonage
Système agricole	Registre Parcellaire Graphique	ASP	exploitations	1/5 000	2006-2011	annuelle		administratif
			ilots déclarés					
			superficies cultivées irriguées					
Prélèvements en eau (usage irrigation)	Système d'Information sur l'Eau	AEAG	Points de Prélèvements Agricoles (types de ressources prélevées, volumes prélevés, caractéristiques des irriguants)	communale	2003-2011	annuelle		administratif
Ressources en eau	BD CARTHAGE®	IGN	cours d'eau	1/50 000	2012	annuelle		physique
			tronçons hydrographiques					
	BD TOPO®	IGN	masses d'eau superficielles	1/10 000	2011-2012	en continu		
	BDRHF	BRGM	masses d'eau souterraines	entre 1/50 000 et 1/1 000 000	2009	/		

Tableau 1 : Tableau comparatif des principales caractéristiques des Bases de Données Géographiques à appairer

3. Une méthode d'intégration de Bases de Données Géographiques de nature hétérogène

D'après Sheeren (2002), l'appariement de données géographiques est un processus « classique » envisagé dès lors qu'il s'avère nécessaire de mettre en correspondance des objets géographiques se rapportant aux mêmes phénomènes du monde réel mais provenant de jeux de données différents. Dans la littérature on retrouve différents objectifs de l'appariement de données géographiques (Sheeren 2002, Gesbert 2005, Olteanu 2008). Il peut être réalisé afin d'analyser la qualité des données (Bel Hadj 2001), afin de détecter les incohérences entre BDG (Sheeren et al. 2004), d'automatiser les procédures de mise à jour (Gomboši et al. 2003, Badard & Lemarié 2002), de développer des BD multi-résolutions (Devogele 1997, Sheeren 2002) ou encore dans le but d'intégrer des BDG hétérogènes (Walter & Fritsch 1999). La méthode que nous proposons cherche à répondre à ce dernier objectif. Il s'agit de mettre en relation des données issues de plusieurs bases pour produire une connaissance nouvelle qui puisse appuyer une démarche de gestion intégrée des ressources en eau eu égard à la pression générée par les prélèvements agricoles en période d'étiage. La mise en œuvre de la méthode doit permettre de respecter les grandes cohérences du système socio-hydrologique et de produire une représentation réaliste de la structure des relations agriculture et eau à l'échelle du BAG et de ses sous-entités hydrologiques.

Parent & Spaccapietra (1996) identifient trois étapes principales dans le processus d'intégration souvent reprises dans la littérature et desquelles nous nous sommes inspirés pour bâtir notre méthodologie : **pré-intégration** ; **déclaration des correspondances** ; **intégration**. Celles-ci sont décrites dans la figure 3 et explicitées en suivant. Entre chacune de ces étapes des aller-retour sont parfois nécessaires pour intégrer des traitements intermédiaires complémentaires non identifiés au préalable.

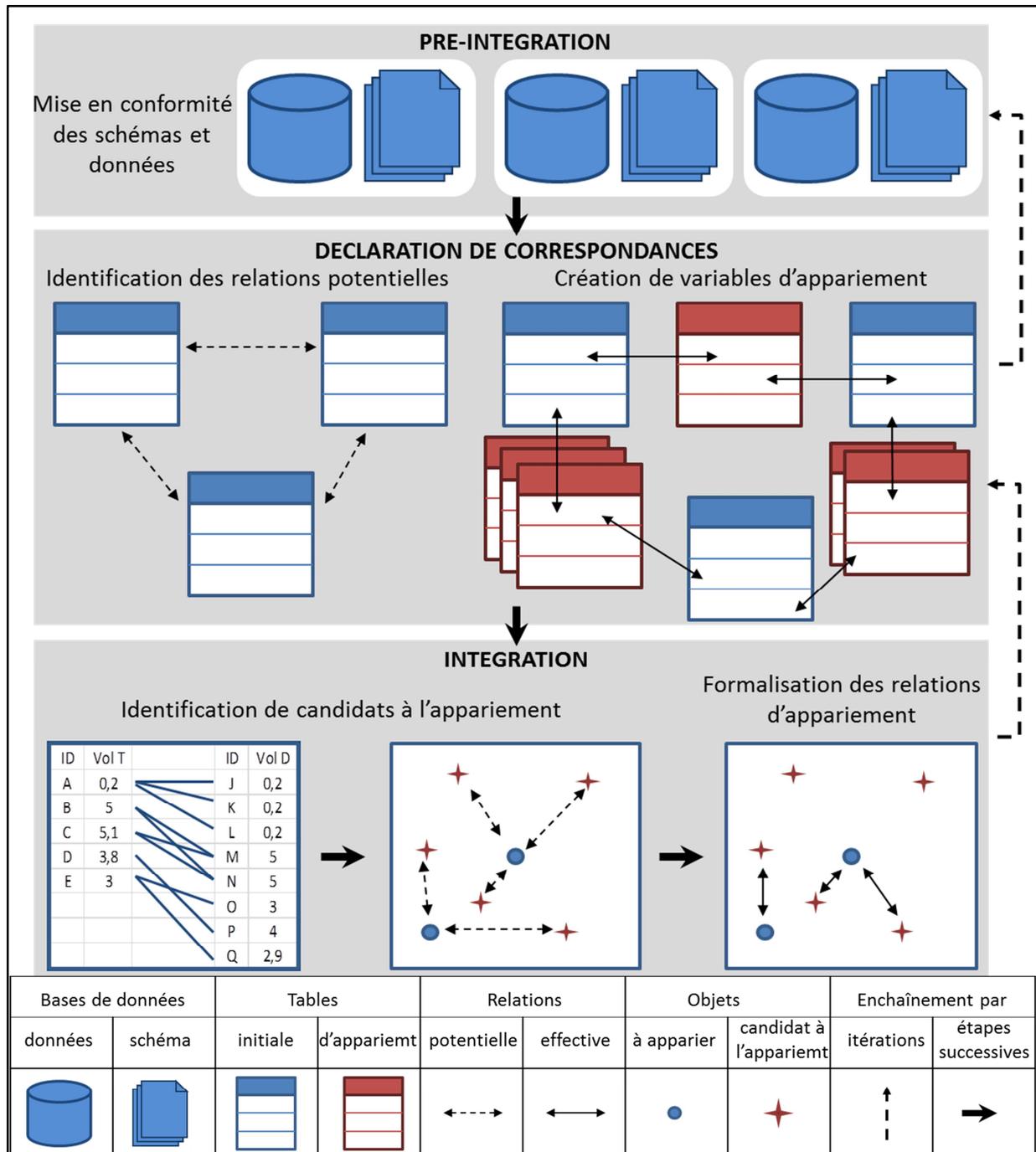


Figure 3 : Méthode générale pour l'intégration de Bases de Données Géographiques

3.1. L'étape de pré-intégration

L'étape de pré-intégration comprend toutes les phases préalables à la mise en relation : la collecte des données, la mise en conformité des schémas initiaux sur un plan sémantique et l'homogénéisation des données, tant au niveau des attributs (identifiants) que de leurs caractéristiques spatiales (projection, étendue, etc.). Cette étape ne doit pas être négligée, elle est particulièrement lourde à mettre en œuvre et conditionne ensuite l'ensemble du processus d'appariement. La mise en conformité des schémas initiaux s'appuie principalement sur

l'interprétation correcte des spécifications, directement liées à un ensemble de connaissances implicites qui reposent sur l'expertise des concepteurs de la base (Gesbert 2005). Comme le soulignent Sheeren et al. (2004) ce travail est essentiel pour interpréter les différences de représentations d'une BDG et ainsi guider l'étape suivante de recherche et de description des correspondances entre BDG.

3.2. Identifier les correspondances entre BDG

La deuxième étape consiste à identifier les éléments communs ou similaires (classes, objets, attributs, etc.) qui peuvent permettre de faire le lien entre les BDG. Il s'agit ici de s'intéresser précisément à ce qui est représentés et dans quelle mesure il y a des similarités entre deux BD que l'on souhaite relier (Devogele 1997). L'identification des correspondances entre bases doit aboutir à l'élaboration d'un modèle de données commun avec une définition précise des cardinalités (1-1, 1-n, n-n) entre les éléments à appairer. Celles-ci déterminent le niveau de complication de la procédure d'affectation des objets entre eux.

Dans la configuration idéale, certains objets géographiques vont avoir un homologue direct et facilement identifiable (e.g. avec le même identifiant) dans la BD à relier. Dans de nombreux cas, il n'existe pas de variable commune permettant de faire le lien. Il va donc être nécessaire de construire des **variables d'appariement complémentaires**. Devogele (1997) et Olteanu (2008) différencient ainsi trois types d'appariement qui peuvent être combinés entre eux (Figure 4) :

- *sémantique* (ou *attributaire*), caractéristique de l'intégration de BD classiques, basé sur des critères attributaires, qu'ils soient quantitatifs (e.g. superficie, volume, etc.) ou qualitatifs (e.g. nom, nature, identifiant, etc.) ;
- *géométrique* qui consiste à appairer les données géographiques par leur localisation et leur forme ;
- *topologique* basé sur les relations spatiales des objets entre eux.

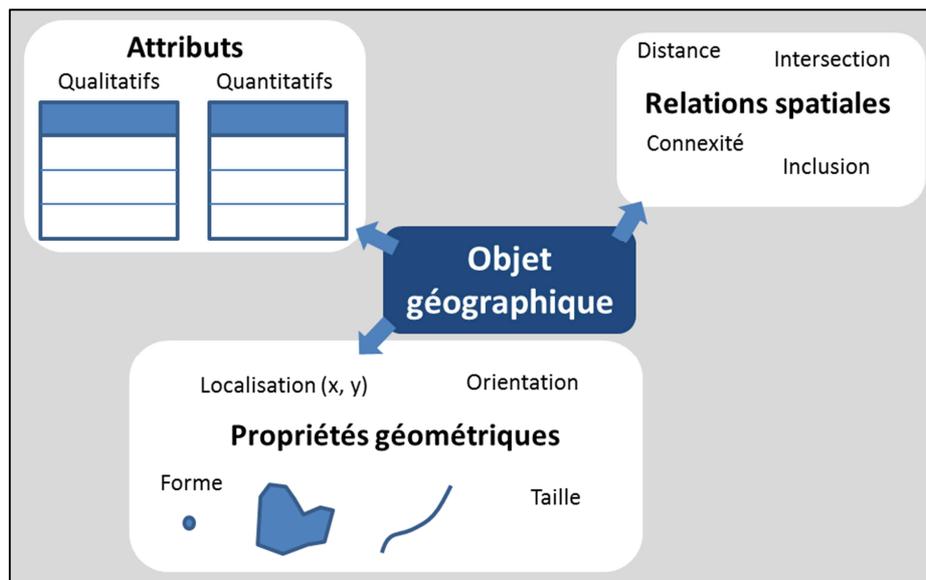


Figure 4 : Critères permettant la caractérisation d'un objet géographique

Pour construire ces variables complémentaires nous faisons appel à des connaissances expertes dans les différents domaines qui nous intéressent (agronomie, hydrologie, etc.).

3.3. Intégration : valider et formaliser les relations d'appariement

La dernière étape consiste à déterminer des règles d'affectation des différents objets géographiques entre eux. En l'absence d'homologues direct, le processus d'intégration se base nécessairement sur des hypothèses dont l'analyse et la vérification vont permettre de déterminer les « meilleurs » candidats à l'appariement (Olteanu 2008). Ces hypothèses d'appariement entre les objets vont être analysées soit par des mesures de distance entre des valeurs quantitatives soit par des mesures de similarité entre des chaînes de caractères. Ces deux types de distances peuvent être traduits en probabilités de correspondances (tel objet de la BD1 à tant de chance de correspondre à tel objet de la BD2).

Plus les concepts que l'on veut appairer sont différents (distance conceptuelle importante) plus le nombre de candidats à l'appariement, et donc de relations potentielles entre eux, risque d'être important. Dans une telle situation, il est nécessaire de concevoir et de quantifier plusieurs variables d'appariement complémentaires afin

de créer un faisceau d'hypothèses de relations qui permettent de réduire fortement la multiplicité des relations potentielles entre candidats.

Pour réaliser l'appariement l'application des hypothèses et donc les calculs de distance peuvent s'enchaîner de façon successive ou combinées. Dans le premier cas, les candidats à l'appariement sont filtrés au fur et à mesure par application successive d'opération de minimisation des distances sur les différents critères. Dans le second cas, une mesure de distance globale, basée sur les différents critères retenus pour réaliser l'appariement, est estimée. Elle permet de ne pas définir d'ordre de mobilisation des critères d'appariement mais peut nécessiter de d'affecter des valeurs de pondération pour chaque critère.

Plusieurs types de distances peuvent être exploités afin de comparer deux variables quantitatives et de déterminer les plus proches voisins (euclidienne, de Mahalanobis, basée sur le score de propension, sur le pourcentage de différence, etc.) (Fougère, 2000, Gu & Rosenbaum, 2013, Rosenbaum & Rubin, 1983). Quelle que soit la méthode retenue, sa mise en œuvre nécessite généralement une étape préalable de normalisation des données pour permettre leur comparaison. Pour la mesure de similarité entre des chaînes de caractères il existe également différentes méthodes possibles telles que la distance de Levenshtein (Levenshtein 1965) ou encore de Damerau (Damerau 1964) souvent combinées entre elles. Plusieurs de ces méthodes sont testées, comparées et combinées afin d'estimer et d'améliorer la qualité de l'appariement. L'évaluation finale du résultat de l'intégration se base ensuite sur la vérification du respect des grandes cohérences via des analyses géostatistiques.

4. Mise en œuvre de la méthode d'intégration à l'échelle du BAG pour déterminer qui prélève, en quelle quantité et dans quelle ressource

Dans cette partie nous présentons l'état d'avancement de la méthode d'appariement (Figure 3) mise en œuvre pour caractériser et analyser les interactions entre systèmes de culture irrigués, prélèvements agricoles et ressources en eau à l'échelle du BAG.

4.1. L'étape de pré-intégration

Les spécifications décrivant les BD sont parfois peu explicites (Sheeren 2002) et ne permettent pas de connaître le sens d'une valeur saisie pour caractériser un objet. Afin de préciser la signification des différents attributs et d'enrichir les schémas de bases de données initiaux, de nombreux échanges avec les organismes de diffusion ont été nécessaires. A partir de l'analyse d'un exemplaire d'une fiche de déclaration des prélèvements en eau remplies par les irrigants et des échanges conduits avec l'Agence de l'Eau Adour-Garonne (AEAG) nous avons, par exemple, pu identifier et récupérer des données clés pour notre étude mais non mises à disposition via le SIE (e.g. tables sur les capacités des plans d'eau, les caractéristiques des adhérents à des structures d'irrigation collectives, etc.). Ce travail a permis de formaliser et d'enrichir chacun des modèles de données initiaux des BDG à appairier présentées dans la partie 2 (Figure 5).

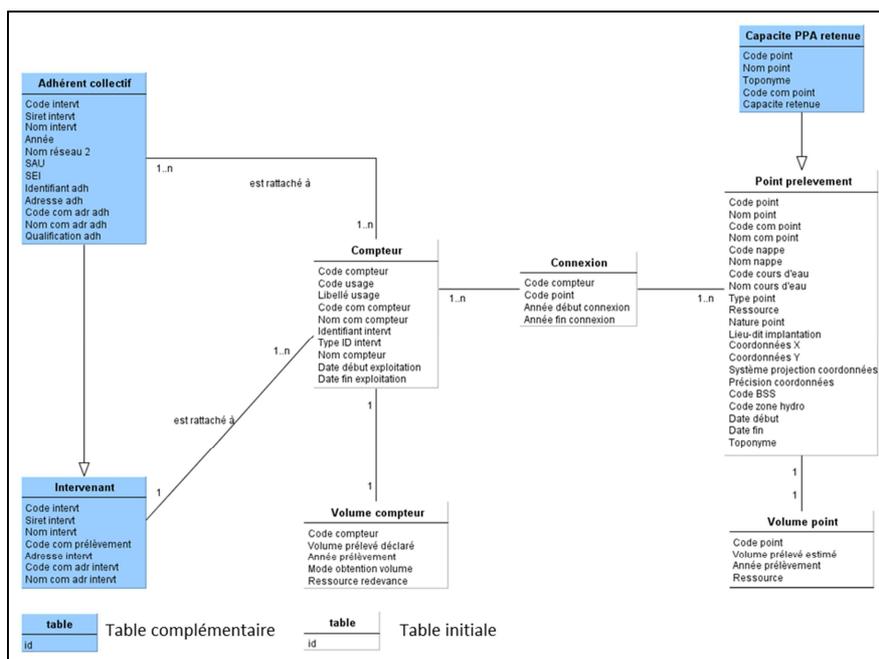


Figure 5 : Modèle de données des tables initiales et complémentaires issues de la BDG sur les prélèvements agricoles (source AEAG, réalisation INRA)

4.2. Déclaration des correspondances entre BDG et intégration

L'analyse approfondie des spécifications permet ensuite d'identifier les relations potentielles entre les BDG à appairer et les variables complémentaires à produire pour leur intégration. Les étapes de déclaration des correspondances et d'intégration ont été volontairement imbriquées. En effet, dans certains cas, l'appariement entre BD a dû être réalisé en amont pour permettre la création de variables d'appariement complémentaires. C'est par exemple le cas de l'appariement entre Points de Prélèvement Agricoles (PPA) et ressources en eau nécessaire pour affiner l'appariement des îlots irrigués aux PPA et différentes ressources.

4.2.1. Identification des relations potentielles et démarche d'intégration

La figure n°6 illustre les principales relations que nous cherchons à établir, cardinalités à respecter, variables d'appariement à mobiliser pour relier les exploitations agricoles du RPG aux PPA du SIE et aux différentes ressources en eau (nappes, cours d'eau et plans d'eau). Chacune de ces relations sont ensuite explicitées et détaillées via des tables de correspondances permettant de décrire les variables intermédiaires et d'appariement à créer. Les procédures de création de ces dernières sont présentées en détail dans la sous-section 4.2.2. L'ordre général d'appariement proposé dans la figure ne permet pas de saisir l'enchaînement précis des différentes étapes car comme nous l'avons déjà mentionné (CF. 3.) des aller-retour sont parfois nécessaires entre celles-ci.

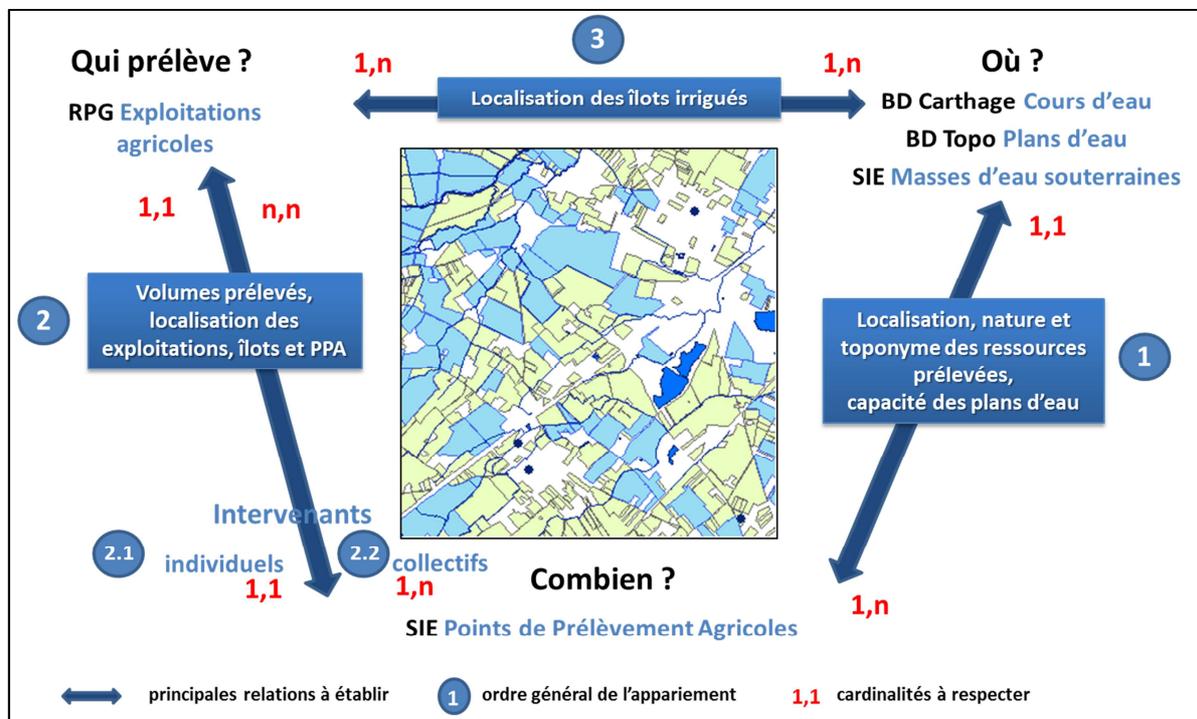


Figure 6 : Principales relations que nous cherchons à établir pour relier exploitations, points de prélèvement agricoles et ressources en eau

Variables d'appariement à mobiliser pour appairer ressources en eau et Points de Prélèvement Agricoles du SIE
 Dans la grande majorité des cas, les Points de Prélèvement Agricoles n'ont pas de géoréférencement précis, ils sont localisés au centroïde de la commune. En revanche, les PPA rattachés à des cours d'eau (ce), à des nappes phréatiques (np) ou captives (nc) sont cependant respectivement caractérisés par le code et le nom de l'entité hydrographique, le code et le nom de l'entité hydrogéologique de rattachement qui permettent de faire le lien avec les identifiants et libellés des entités de la BD CARTHAGE® et de la BDRHF®. Ces champs ne sont néanmoins pas renseignés de manière exhaustive (il manque l'information pour 21% des PPA rattachés à un ce et 50% des PPA rattachés à une np ou nc). De même, pour les plans d'eau nous ne disposons d'aucune information dans le SIE permettant de faire le lien direct avec les plans d'eau de la BD TOPO®. Dans ces deux types de cas, i.e. en l'absence d'information sur la ressource ou d'homonyme, nous utilisons la BD NYME® pour spatialiser les PPA à partir de variables correspondantes dans le SIE ("code commune", "libelle du PPA" correspondant au nom d'usage de l'installation et "toponyme" de la localisation du point) fournissant des informations complémentaires sur le lieu d'implantation du PPA. L'appariement entre le SIE et la BD NYME® est réalisé par comparaison des chaînes de caractères. Pour affiner l'appariement des plans d'eau aux PPA correspondants, l'estimation de la capacité de ces plans d'eau peut être croisée au volume prélevé déclarée. La

combinaison de ces variables permet à la fois d'apparier ressources en eau et PPA (Tableau 2) et de générer une variable d'appariement (localisation du PPA apparié) nécessaire aux autres étapes d'intégration (Tableau 3).

Variables			
SIE	d'appariement	intermédiaires	ressources
code et nom de l'entité hydrologique du PPA (si ils sont connus)			identifiant et libellé du cours d'eau
code de l'entité hydrogéologique du PPA (si ils sont connus)			identifiant et libellé de la masse d'eau souterraine
localisation du PPA (commune)	localisation de la ressource (commune)		localisation de la ressource (x,y)
libellé du PPA	localisation du PPA (x,y) affinée par croisement au toponyme		localisation de la ressource (x,y)
toponyme du PPA	localisation du PPA (x,y) affinée par croisement au toponyme		localisation de la ressource (x,y)
volume prélevé déclaré du PPA	capacité du plan d'eau	localisation, superficie, forme et profondeur des plans d'eau à usage irrigation	localisation et description des masses d'eau superficielles

Tableau 2 : Table de correspondances entre PPA du SIE et ressources en eau

Variables d'appariement à mobiliser pour apparier îlots du RPG et Points de Prélèvement Agricoles du SIE

L'appariement entre îlots agricoles et PPA peut être réalisé au travers de la mise en relation des exploitations agricoles du RPG⁶ et des intervenants du SIE. Le caractère anonyme des données du RPG prive d'une variable, le numéro SIRET de l'exploitation, qui permettrait un appariement direct avec un intervenant du SIE⁷. Pour réaliser cet appariement nous cherchons à rapprocher les volumes prélevés déclarés par les intervenants aux besoins théoriques en eau des exploitations. Deux types d'intervenant sont considérés. Les intervenants individuels (en moyenne 72% des volumes prélevés déclarés) et les intervenants collectifs. Les intervenants individuels sont traités en amont car leur relation est d'ordre 1 à 1 (un intervenant pour une exploitation RPG). En revanche, pour les collectifs la tâche est plus complexe puisque la relation est d'ordre n à n avec les données du RPG (un à n intervenants collectifs pour n exploitations du RPG).

Variables			
RPG	intermédiaires	d'appariement	SIE
superficies irriguées	superficies irriguées reconstituées	volumes prélevés théoriques	volumes prélevés déclarés
département siège de l'exploitation	commune siège supposée de l'exploitation	commune d'implantation potentielle de l'exploitation	commune de l'intervenant (adresse déclaration redevance)
superficie de l'exploitation (via îlots) par commune			
localisation de l'îlot (commune)		communes d'approvisionnement potentielle de l'îlot	localisation du PPA (commune)
localisation de l'îlot (x,y)		localisation du PPA (x,y) apparié à la ressource	localisation du PPA (commune)
identifiant exploitation	nombre d'exploitations appariées	nombre d'adhérents potentiel	nombre d'adhérents à un collectif

Tableau 3 : Table de correspondances entre exploitations du RPG et intervenants du SIE

⁶ Etant donné que dans le RPG le caractère irrigué de l'îlot est seulement connu sur trois années, de 2007 à 2009, l'ensemble des traitements présentés ci-après ont été réalisés sur ces années

⁷ Dans les données du SIE sur les prélèvements agricoles le code SIRET est fourni pour tous les salariés dont l'agriculture est l'activité principale soit 93% des intervenants déclarant

- *Appariement des exploitations du RPG avec les intervenants individuels du SIE*
Les volumes prélevés théoriques par commune et par exploitation (du RPG) sont comparés aux volumes déclarés par commune et par intervenant (du SIE) afin de déterminer, parmi les îlots agricoles (via les exploitations) les candidats à l'appariement avec chacun des PPA (via les intervenants). Des variables d'appariement permettent ensuite de filtrer et d'affiner ces résultats (commune d'implantation potentielle de l'exploitation, communes d'approvisionnement potentielles de l'îlot, localisation du PPA apparié à la ressource) ;
- *Appariement des exploitations du RPG avec les intervenants collectifs du SIE*
La table des adhérents à des collectifs du SIE renseigne sur le nombre d'exploitants rattachés à ces structures. En revanche, il n'est pas possible de savoir si ces exploitants (intervenants) sont également des intervenants individuels. Pour établir des relations entre les exploitations du RPG et intervenants collectifs du SIE nous procédons de nouveau à la comparaison des volumes théoriques et déclarés, en deux temps. Sont traitées en priorité les exploitations du RPG non rattachées à des intervenants individuels (dans la phase d'appariement précédente) puis les exploitations agricoles déjà rattachées à des individuels mais dont les volumes prélevés théoriques sont supérieurs aux volumes déclarés et dont la distance de rattachement au PPA semble cohérente. Comme pour les individuels, des critères complémentaires permettent ensuite de filtrer et d'affiner ces résultats (commune d'implantation potentielle de l'exploitation, communes d'approvisionnement potentielles de l'îlot, localisation du PPA apparié à la ressource). Dans un dernier temps, il est possible de comparer le nombre d'adhérents potentiel issu de ces résultats avec le nombre d'adhérents renseignés dans le SIE.

Variables d'appariement à mobiliser pour appairer îlots du RPG et ressources en eau

La relation entre îlots irrigués et ressources en eau est déterminée par la localisation respective de ces différents objets. Plusieurs matrices de distance (pour chaque type de ressource) sont donc générées pour déterminer les ressources d'approvisionnement potentielles des îlots. Pour affiner la qualité de l'appariement des îlots aux différentes ressources en eau les résultats issus des de la mise en relation des autres BDG sont mobilisés. En effet, une fois le PPA d'approvisionnement de l'îlot connu il est par exemple possible de restreindre à une seule ressource la recherche de candidats à l'appariement (à condition toutefois que l'îlot ne soit associé exclusivement à des PPA rattachés à un seul type de ressource). Dans le cas des îlots rattaché à un PPA de type 'retenue', le volume du plan d'eau peut être rapporté au volume prélevé déclaré théorique.

Variables					
RPG	intermédiaires	d'appariement	d'appariement	intermédiaires	ressources
localisation de l'îlots (x,y)					localisation de la ressource (x,y)
superficies irriguées	superficies irriguées reconstituées	volumes prélevés théoriques	capacité du plan d'eau	superficie, forme et profondeur des plans d'eau à usage irrigation	localisation et description des masses d'eau superficielles

Tableau 4 : Table de correspondances entre îlots du RPG et ressources en eau

4.2.2. *Création de variables complémentaires (intermédiaires et d'appariement)*

Proposer une typologie des plans d'eau pour extraire les plans d'eau à usage irrigation et en évaluer la capacité
La couche surface en eau du thème hydrographie de la BD TOPO® peut permettre de localiser les plans d'eau à usage irrigation. Pour cela il est nécessaire de caractériser les différents types de plans d'eau existants en nous appuyant sur un ensemble de critères spatiaux (géométrie, distribution spatiale par rapport aux cours d'eau, position pédo-morphologique et topographique, etc.), d'usage et de fonctionnement hydrologique par croisement avec d'autres sources d'informations contenues dans d'autres BDG (ouvrages de stockage au soutien d'étiage, grands lacs destinées à la production d'hydroélectricité, nappe d'accompagnement, zones humides, étangs littoraux, îlots irrigués etc.) et intégration de connaissances expertes. L'estimation de la capacité des plans d'eau est réalisée par application d'une formule empirique basée sur les caractéristiques des retenues (superficie, forme/ profondeur).

Affiner la localisation des Points de Prélèvement Agricoles

Nous avons reconstitué une couche des PPA dont le géoréférencement a été affiné par appariement à la ressource correspondante. Lorsque les données attributaires le permettent, le PPA est automatiquement repositionné au centroïde de l'objet (cours d'eau, plan d'eau, nappe souterraine) au sein de sa commune d'affectation (Figure 7).

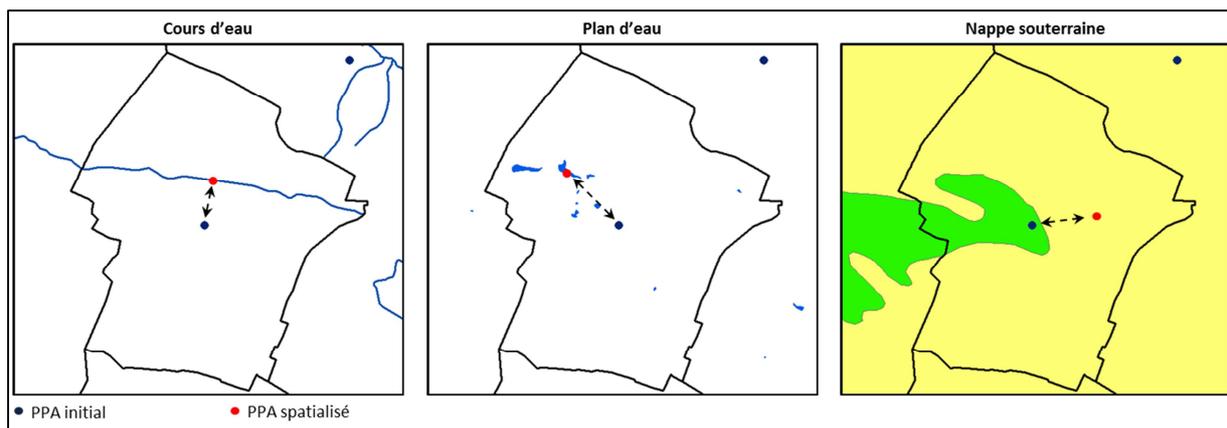


Figure 7 : Appariement entre PPA et ressource en eau

Définir des volumes prélevés théoriques

L'estimation des volumes prélevés théoriques a été réalisée en deux grandes étapes : (i) estimation des superficies effectivement irriguées sur une campagne puis (ii) estimation des besoins en eau d'irrigation des cultures et des volumes prélevés théoriques. Pour réaliser la première étape nous avons associé à chaque culture irriguée une probabilité d'irrigation pouvant être conditionnée à la présence d'autres cultures dans l'îlot. Ces probabilités ont été déterminées par dire d'expert et analyse des statistiques agricoles. Ainsi nous considérons, que dans un îlot irrigué du RPG, "maïs grain et ensilage", "autres oléagineux", "protéagineux", "semences", "vergers" et "autres cultures industrielles" (cultures irriguées principales) sont irriguées et que "fourrage", "prairies temporaires" et "prairies permanentes" (culture irriguées secondaires) sont irriguées si une culture du premier groupe n'est pas présente sur l'îlot. Toutes les autres cultures ne sont jamais irriguées soit "gel", "estives et landes", "vignes", (etc.) sauf si elles sont présentes dans un îlot irrigué sans aucune autre culture des deux premiers groupes (Figures 8 et 9).

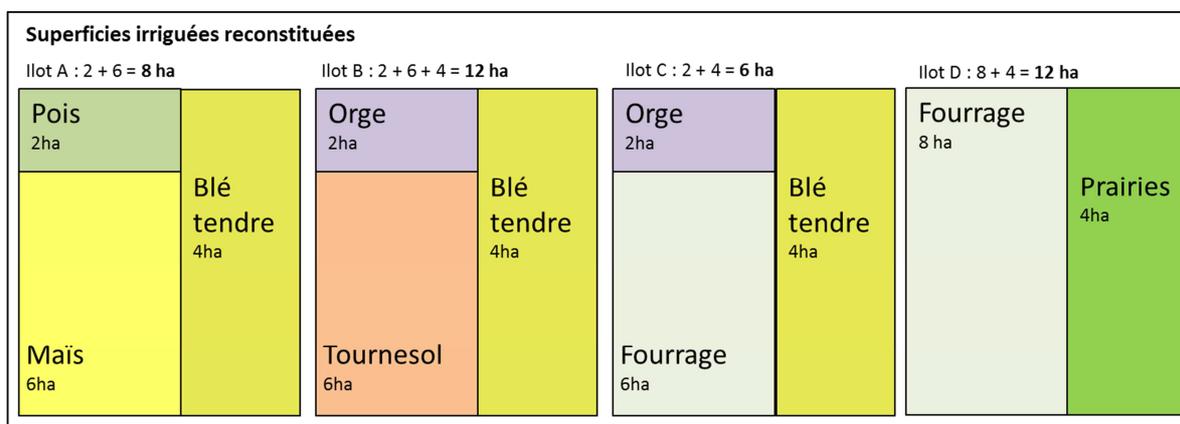


Figure 8 : Reconstitution des superficies irriguées via les données du RPG

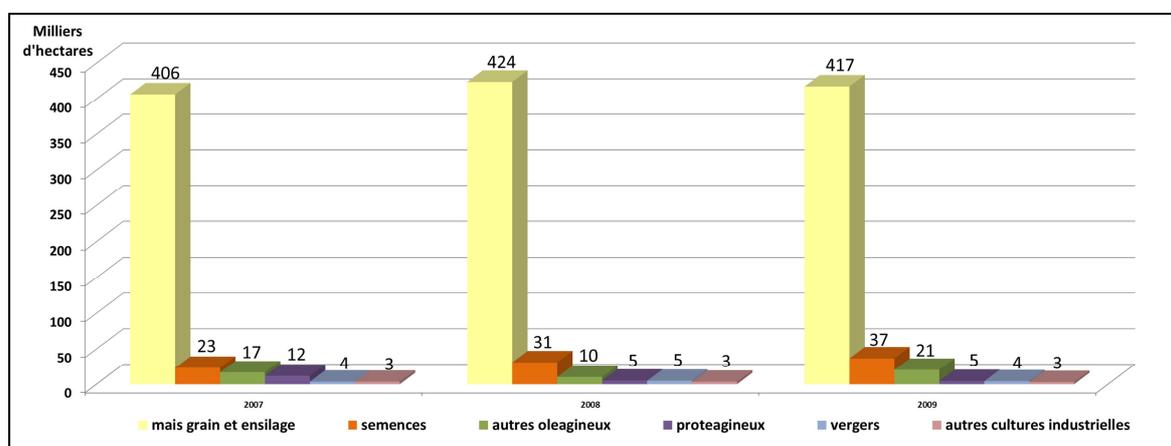


Figure 9 : Répartition et évolution des principales superficies irriguées après reconstitution à l'échelle du BAG

Deuxième étape, nous avons estimé la demande théorique en eau d'irrigation considérant un taux de couverture des besoins de la plante déterminé par l'agriculteur en fonction de ses objectifs et de ses contraintes en termes de matériel d'irrigation et de ressources en eau (Teyssier 2006). Plusieurs méthodes ont été testées et comparées :

- DEau1 - on applique un apport d'eau en fonction d'un nombre de tours d'eau moyen pour toutes les cultures irriguées ;
- DEau2 - pour les classes "maïs", "semences" et "autres oléagineux" on affine la demande en eau par la formule suivante $ETP \times Kc \text{ moyen} \times TCBE$. Avec $ETP = \text{EvapoTranspiration Potentielle}$ ⁸ cumulée sur la période potentielle d'irrigation (du 15 juin au 31 août), Kc moyen le coefficient cultural⁹ moyen par culture qui tient compte du stade de développement de la culture sur la période 15 juin – 31 août et $TCBE$ le taux de couverture des besoins en eau des cultures estimés à 75 % (Teyssier 2006) ;
- DEau3 - on affine la demande en eau avec un modèle de croissance des plantes empirique développé par l'UMR AGIR à Toulouse (Nolot & Debaeke 2003) prenant en compte la teneur en argile et la réserve utile du sol.

Les deux dernières méthodes tiennent compte de la variabilité du sol et du climat à l'échelle du BAG. Pour cela, nous avons attribué à chaque îlot la série climatique du point météo le plus proche et les caractéristiques pédologiques par intersection avec les unités cartographiques de sol correspondantes¹⁰ (CF. section 2.1.).

Générer des matrices de distance géométrique euclidienne

Afin de générer des variables décrivant l'agencement spatial des objets (PPA, îlots, exploitations, ressources) les uns par rapport aux autres il faut procéder au calcul de plusieurs matrices de distance géométrique :

- entre la commune siège supposée de l'exploitation agricole¹¹ et la commune d'implantation des intervenants renseignée dans le champ adresse du SIE. Pour qualifier la qualité de la relation représentée par cette distance, et déterminer une commune d'implantation potentielle de l'exploitation, nous associons une note de 1 à 3 (correspondant respectivement à des communes « jumelles », contiguës ou aux communes proches mais à une distance inférieure à 100 km) ;
- entre le centroïde des communes et celui des îlots irrigués pour attribuer à chaque îlot irrigués des communes d'approvisionnement potentielles.

Pour finaliser l'intégration, une dernière étape consiste à formaliser les règles d'affectation des exploitations agricoles aux points de prélèvement et aux différentes ressources en eau au sein d'une BDG intégrée et enrichie.

Discussion-Conclusion

D'après les conclusions récentes du projet Garonne 2050, la fragilité de la ressource en eau devrait se renforcer sous le double effet conjugué de la croissance démographique et du changement climatique (évapotranspiration potentielle accrue, influence réduite de la neige sur l'hydrologie, etc.) (AEAG 2012c). La mise en place de la Loi sur l'Eau et les Milieux Aquatiques de 2006, se traduisant par la mise en œuvre de nouveaux Volumes Prélevables (VP) par l'agriculture définis par unité de gestion, impose à l'Organisme Unique, nouvelle institution en charge de la gestion de ces volumes, de disposer d'une connaissance fine des interactions entre les différents éléments et dynamiques du système socio-hydrographique. Pour répondre à cet objectif, de répartition des VP entre les usagers agricoles et de mise en œuvre de politiques de gestion cohérentes, l'interopérabilité entre BDG, permettant de décrire le territoire de façon intégrée semble indispensable. La méthode d'appariement proposée dans cette communication s'attache à répondre en partie à ces exigences en proposant des règles d'affectation des données sur l'irrigation (RPG), sur les prélèvements en eau (SIE) et les différentes ressources disponibles (BD CARTHAGE®, BD TOPO®, BDRH®). Cette étude montre la difficulté de proposer des procédures automatiques d'intégration de données géographiques hétérogènes construites pour des besoins divers et selon des représentations du monde distinctes. La méthode d'appariement présentée ici est assez lourde à mettre en œuvre mais présente l'avantage d'être reproductible aux autres grands bassins versants hydrologiques français puisque toutes les données mobilisées sont disponibles au niveau national. Malgré sa lourdeur, ce type de travaux sur l'interopérabilité des BDG est indispensable pour proposer des outils qui permettent une analyse intégrée de la structure de la gestion quantitative de l'eau à l'échelle d'un grand bassin versant, i.e. des relations entre systèmes de culture et de production irrigués et ressources en eau.

⁸ L'ETP représente la quantité d'eau évaporée par un couvert végétal sur un sol en condition normale (correspond à l'ETP d'un gazon bien approvisionné en eau)

⁹ $Kc \times ETP = \text{EvapoTranspiration Maximale (ETM)}$ qui représente la quantité d'eau consommée par une culture donnée.

¹⁰ Les unités cartographiques de sol permettent de spatialiser les grands ensembles morpho-paysagers. Elles sont composées de plusieurs types de sol (ou unités typologiques de sol) qui se caractérisent par un ensemble de variables pédologiques.

¹¹ La commune siège supposée de l'exploitation est obtenue en croisant plusieurs informations contenues dans les données brutes du RPG, il s'agit de la commune contenant la plus grande superficie cultivée déclarée au sein du département siège de l'exploitation.

Remerciements

Ce travail a été réalisé dans le cadre du projet MAELIA (Multi-agent for Environmental Norms impact Assessment) financé par le RTRA Sciences et Technologies pour l'Aéronautique et l'Espace.

Références

- AEAG (2012a) Sécheresse : origines , conséquences et actions de l'agence de l'eau Adour-Garonne. 16.
- AEAG (2012b) Etude prospective sur les besoins et les ressources en eau, à l'échelle du bassin de la Garonne. Journée d'échanges "Garonne 2050". 21.
- AEAG (2012c) Garonne 2050. Etude prospective sur les besoins et les ressources en eau, à l'échelle du bassin de la Garonne. <http://www.garonne2050.fr/>. Accessed 23 Apr 2013
- ASP (2011) Mise à disposition du registre parcellaire graphique anonyme. <http://www.asp-public.fr/?q=node/856>. Accessed 7 Mar 2013
- Badard T, Lemarié C (2002) Associer des données : l'appariement. Ruas, A (éd.), Généralisation et représentation multiple, Lavoisier. Paris, pp 163–183
- Bel Hadj A (2001) Qualité géométrique des entités géographiques. Écarts, surfaciques - Application à l'appariement et définition d'une typologie des géométries. 420.
- Damerau F (1964) A technique for computer detection and correction of spelling errors. *Communications of the ACM* 7:659–664.
- Debril T, Therond O (2010) Les difficultés associées à la gestion quantitative de l'eau et à la mise en œuvre de la réforme des volumes prélevables : le cas du bassin Adour-Garonne. *Agronomie environnement et sociétés* 2:127–137.
- Devogele T (1997) Processus d'intégration et d'appariement de Bases de Données Géographiques. Application à une base de données routières multi-échelles. 209.
- Eaufrance Le système d'information sur l'eau (SIE). Http://www.eaufrance.fr/comprendre/les-donnees-sur-l-eau/?id_article=833. Accessed 16 Jan 2013
- Fougère D (2000) Les méthodes économétriques d'évaluation. *RFAS* 105–128.
- Gardarin G (2003) Bases de données, Gpe Eyroll. 788.
- Gesbert N (2005) Étude de la formalisation des spécifications de bases de données géographiques en vue de leur intégration. 134.
- Gomboši M, Žalik B, Krivograd S (2003) Comparing two sets of polygons. *International Journal of Geographical Information Science* 17:431–443. doi: 10.1080/1365881031000072627
- Gu XS, Rosenbaum PR (2013) Comparison of Multivariate Matching Methods : Structures, Distances, and Algorithms. *Journal of Computational and Graphical Statistics* 2:405–420.
- Legifrance (2006) LOI n° 2006-1772 du 30 décembre 2006 sur l'eau et les milieux aquatiques. <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000649171&dateTexte=&categorieLien=id>. Accessed 23 Apr 2013
- Levenshtein V (1965) Binary codes capable of correctiong deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR* 163:845–848.

- MAAF (2013) Agreste La statistique, l'évaluation et la prospective agricole. <http://www.agreste.agriculture.gouv.fr/page-d-accueil/article/donnees-en-ligne>.
- Moody R, Ast JA van (2012) Implementation of GIS-Based Applications in Water Governance. *Water Resources Management* 26:517–529. doi: 10.1007/s11269-011-9929-4
- Nolot J-M, Debaeke P (2003) Principes et outils de conception, conduite et évaluation de systèmes de culture. *Cahiers Agricultures* 387–400.
- Olteanu A (2008) Fusion de connaissances imparfaites pour l'appariement de données géographiques Proposition d'une approche s'appuyant sur la théorie des fonctions de croyance. 268.
- Parent C, Spaccapietra S (1996) Intégration de bases de données : Panorama des problèmes et des approches. *Ingénierie des systèmes d'information* 4:1–18.
- Rosenbaum PR, Rubin DB (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70:41–55.
- Sheeren D (2002) L'appariement pour la constitution de bases de données géographiques multi-résolutions. *Revue internationale de géomatique* 10:19.
- Sheeren D, Mustiere S, Zucker J-D (2004) How to Integrate Heterogeneous Spatial Databases in a Consistent Way? Proceedings of the Conference on Advanced Databases and Information Systems (ADBIS). Budapest, septembre 2004, pp 364–378
- Teyssier F (2006) Les consommations d'eau pour l'irrigation en Midi-Pyrénées. Surfaces irriguées et volumes d'eau consommés. Rapport avec le climat, la PAC, les ressources, sur la période 2001-2005. 42.
- Walter V, Fritsch D (1999) Matching spatial data sets: a statistical approach. *International Journal of Geographical Information Science* 13:445–473.