



HAL
open science

Fusion de données géoréférencées et développement de services interopérables pour l'estimation des besoins en eau à l'échelle des bassins versants

Mickaël Beaufls

► **To cite this version:**

Mickaël Beaufls. Fusion de données géoréférencées et développement de services interopérables pour l'estimation des besoins en eau à l'échelle des bassins versants. Sciences de la Terre. Conservatoire national des arts et metiers - CNAM, 2012. Français. NNT : 2012CNAM0847 . tel-00922997

HAL Id: tel-00922997

<https://theses.hal.science/tel-00922997>

Submitted on 1 Jan 2014

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.

ÉCOLE DOCTORALE Arts et Métiers
Laboratoire de Géodésie et de Géomatique

THÈSE

présentée par :

Mickaël BEUFILS

soutenue le 4 Décembre 2012

pour obtenir le grade de :

Docteur du Conservatoire National des Arts et Métiers

Discipline/ Spécialité : Géomatique

**Fusion de données géoréférencées
et développement de services interopérables
pour l'estimation des besoins en eau
à l'échelle des bassins versants**

THÈSE dirigée par :

M Laurent POLIDORI
M Jean-Michel FOLLIN

Professeur titulaire de la Chaire de Géomatique, CNAM
Maître de conférences, CNAM

RAPPORTEURS :

M Michel MAINGUENAUD
M Alain BOUJU

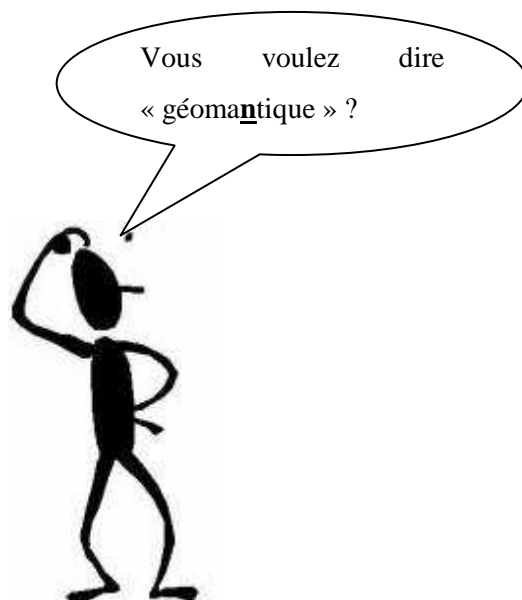
Professeur des universités. INSA de Rouen – Laboratoire LITIS
Maître de conférences HDR. Université La Rochelle – Labo. L3I

EXAMINATEURS :

Mme Yamna DJELLOULI
Mme Valérie JOUËT
M Pierre LAGARDE

Professeur des universités. Université du Maine – UMR ESO
Acatu Informatique
BRGM-STI

A celles et ceux qui œuvrent pour que « géomatique »
ne soit plus considéré comme un mot comportant une faute d'orthographe...



Remerciements

Si la réalisation d'une thèse de doctorat est avant tout un sport individuel, elle n'est pourtant pas le fruit des efforts que d'une seule personne. Aussi je tiens à exprimer ma gratitude envers celles et ceux qui ont contribué à la réalisation de ce travail de recherche.

Mes premières pensées vont à mon équipe d'encadrement. Du côté L2G, je souhaite tout d'abord remercier mon directeur de thèse, M. Laurent Polidori pour son soutien et son encouragement dès mon souhait de réaliser une thèse, ainsi que M. Jean-Michel Follin pour ses conseils et son constant encadrement tout au long de ma thèse. Je tiens également à remercier Mme Valérie Jouët, mon encadrante côté entreprise, qui a eu la difficile mission de reprendre le suivi de ma thèse, pour son aide, ses conseils et m'avoir permis de réaliser ma thèse dans les meilleures conditions possibles. Enfin je tiens aussi à remercier, M. Pierre Lagarde et Mme Sandra Lanini du BRGM, les principaux auteurs de ce sujet de recherche, pour leur suivi et pour m'avoir fait part de leurs savoirs et leurs expériences dans leurs domaines respectifs.

Je souhaite maintenant remercier toutes les personnes ayant travaillé au développement de l'IDS FUSIEAU : en particulier Mickaël Boulas, Yves Dupont, Florian Brault, Adel Helal, Laura Etienne, Alexandre Berge et Olivier Bonin côté Acatus. Je souhaite également remercier Yvan Caballero, Laure Maton et Daphné Dagneaux au BRGM pour m'avoir accordé de leur temps.

Je tiens également à remercier mes amis doctorants. Au L2G, Benoît Legru, Mhamad El-Hage, Bilel Soussi, Maryam Soheili Majd et Marcell Ferenc pour m'avoir toujours gardé une place parmi eux dans le bureau des doctorants, malgré le fait que j'étais principalement en entreprise. Mes pensées vont également à Tiana Randrianarivelo, doctorant au BRGM, dont j'aurais tellement voulu assister à la soutenance.

Plus généralement, je souhaite remercier mes collègues d'Acatus Informatique, Acatus Conseil et Atimic Informatique, ainsi que ceux du L2G et de l'ESGT pour leur sympathie et tous les bons moments passés ensemble durant ces trois années de thèse.

Enfin je tiens à remercier mon entourage personnel, ma famille, mes amis pour leur soutien et leur encouragement durant ces trois années de thèse.

Résumé

De nos jours, la préservation de l'environnement constitue un enjeu prioritaire. La compréhension des phénomènes environnementaux passe par l'étude et la combinaison d'un nombre croissant de données hétérogènes. De nombreuses initiatives internationales (INSPIRE, GEOSS) visent à encourager le partage et l'échange de ces données.

Dans ce sujet de recherche, nous traitons de l'intérêt de mettre à disposition des modèles scientifiques sur le web. Nous montrons l'intérêt d'utiliser des applications s'appuyant sur des données géoréférencées et présentons des méthodes et des moyens répondant aux exigences d'interopérabilité.

Nous illustrons notre approche par l'implémentation de modèles d'estimation des besoins en eau agricoles et domestiques fonctionnant à diverses échelles spatiales et temporelles. Un prototype basé sur une architecture entièrement orientée services web a été développé. L'outil s'appuie sur les standards Web Feature Service (WFS), Sensor Observation Service (SOS) et Web Processing Service (WPS) de l'OGC.

Enfin, la prise en compte des imperfections des données est également abordée avec l'intégration de méthodes d'analyse de sensibilité et de propagation de l'incertitude.

Mots clés : Fusion de données, Web SIG, Open Geospatial Consortium, Modèle web, Services Web, Infrastructure de Données Spatiales, Web Processing Service, capteurs virtuels, architecture orientée service, analyse de sensibilité, propagation de l'incertitude.

Abstract

Nowadays, preservation of the environment is a main priority. Understanding of environmental phenomena requires the study and the combination of an increasing number of heterogeneous data. Several international initiatives (INSPIRE, GEOSS) aims to encourage the sharing and exchange of those data.

In this thesis, the interest of making scientific models available on the web is discussed. The value of using applications based on geospatial data is demonstrated. Several methods and means that satisfy the requirements of interoperability are also purposed.

Our approach is illustrated by the implementation of models for estimating agricultural and domestic water requirements. Those models can be used at different spatial scales and temporal granularities. A prototype based on a complete web service oriented architecture was developed. The tool is based on the OGC standards Web Feature Service (WFS), Sensor Observation Service (SOS) and Web Processing Service (WPS).

Finally, taking into account the imperfections of the data is also discussed with the integration of methods for sensitivity analysis and uncertainty propagation.

Keywords : Data fusion, Web GIS, Open Geospatial Consortium, model web, web services, Spatial Data Infrastructure, Web Processing Service, virtual sensors, service oriented architecture, sensitivity analysis, uncertainty propagation.

Table des matières

Remerciements.....	3
Résumé.....	4
Abstract.....	5
Liste des tableaux.....	10
Liste des figures	11
Introduction	14
De VULCAIN à FUSIEAU	17
Plan du mémoire.....	19
Première partie - Un sujet pluridisciplinaire	20
Chapitre I : Evaluation des besoins en eau	21
I.A Le projet VULCAIN.....	21
I.A.1 Introduction.....	21
I.A.2 Modélisation hydrologique selon une approche globale.....	22
I.B Modélisation des prélèvements en eau agricole	24
I.B.1 Modèle FAO.....	24
I.B.2 Modèle VULCAIN.....	27
I.C Modélisation des prélèvements en eau domestique.....	28
I.C.1 Modèle VULCAIN.....	29
I.C.2 Modèle du SDRE de l'Isère.....	31
I.D Conclusion du chapitre.....	32
Chapitre II : IDS et services interopérables	33
II.A Infrastructures de Données Spatiales	33
II.A.1 Définition et composition	33
II.A.2 Initiatives pour la diffusion et le partage de l'information environnementale.....	34
II.A.3 Mise à disposition de modèles sur le web.....	35
II.B Services et Architectures Orientées services.....	36
II.B.1 Notions de services et d'Architectures Orientée Services	36
II.B.2 Services web dédiés à la géomatique.....	36

II.B.3 Formats standards de données.....	39
II.C Composition de services web	40
II.C.1 Orchestration et chorégraphie	40
II.C.2 Comparatif orchestration / chorégraphie.....	41
II.C.3 Outils pour l'orchestration	41
II.D Conclusion du chapitre.....	43
Chapitre III : Fusion de données et qualité de l'information géographique.....	45
III.A Fusion de données.....	45
III.A.1 Terminologie	45
III.A.2 Catégories de fusion	46
III.B De la qualité de l'information géographique	49
III.B.1 Notion de qualité de l'information géographique.....	49
III.B.2 Prise en compte des imperfections	53
III.C Conclusion du chapitre.....	56
Deuxième partie - Elaboration d'une IDS dédiée à l'évaluation des besoins en eau.....	58
Chapitre IV : Choix conceptuels et notions introduites dans FUSIEAU	59
IV.A FUSIEAU : frère jumeau de STRATEAU ?.....	59
IV.A.1 Des objectifs communs	59
IV.A.2 Des approches différentes	61
IV.B Des données aux IDS	62
IV.B.1 Simulation de multiples IDS	63
IV.B.2 Choix des formats pour la diffusion des données.....	64
IV.C Implémentation du modèle avec des services web	65
IV.C.1 Création de capteurs virtuels	65
IV.C.2 Fonctionnement général d'un algorithme associé à un calcul issue d'un modèle.....	66
IV.D Conclusion du chapitre	68
Chapitre V : Application à l'estimation des besoins en eau agricole et domestique.....	69
V.A Spatialisation du modèle FAO	69
V.A.1 Approche du modèle.....	69
V.A.2 Identification des données utilisées	71
V.A.3 Identification des traitements et de leur ordre d'enchaînement.....	76
V.B Proposition d'un modèle d'estimation des besoins en eau domestiques.....	77
V.B.1 Approche des modèles.....	77

V.B.2 Estimation des besoins en eau liés à la population permanente.....	78
V.B.3 Estimation des besoins en eaux liés au tourisme	80
V.C Evolutions de FUSIEAU.....	84
V.C.1 Vers l'évaluation des ressources en eau à l'échelle des bassins versants	84
V.C.2 Vers un modèle prévisionnel	86
V.D Conclusion du chapitre	89
 Chapitre VI : Développement d'un prototype.....	90
VI.A Une structure flexible	90
VI.A.1 Utilisation de FUSIEAU comme centre de traitement externalisé.....	92
VI.A.2 Traçabilité des résultats	93
VI.A.3 Conception modulaire	94
VI.B Implémentation des modèles	94
VI.B.1 Processus associé à la combinaison de géométries	95
VI.B.2 Processus associé à un calcul dans le modèle.....	96
VI.C Partie client	97
VI.C.1 Représentation des données et des résultats	97
VI.C.2 Construction de scénario pas à pas	98
VI.D Conclusion du chapitre	99
 Troisième partie - Vers une IDS intelligente.....	100
 Chapitre VII : Gestion des imperfections dans un cadre de fusion de données	101
VII.A FUSIEAU : frère jumeau d'UncertWeb ?.....	101
VII.A.1 Des objectifs communs	101
VII.A.2 Des approches différentes	102
VII.B Typologie des imperfections	102
VII.B.1 Imperfections liées aux données	102
VII.B.2 Imperfections liées au modèle.....	105
VII.C Prise en compte des imperfections	107
VII.C.1 Analyse de sensibilité.....	108
VII.C.2 Propagation des incertitudes	110
VII.D Conclusion du chapitre.....	112
 Chapitre VIII : Définition d'une méthode d'intégration de la prise en compte des imperfections dans l'IDS.....	113
VIII.A Interaction avec le modèle	113

VIII.A.1 Organisation des résultats avec prise en compte de l'incertitude	113
VIII.A.2 Définition des incertitudes d'entrée.....	115
VIII.A.3 Intégration de la prise en compte de l'incertitude dans les algorithmes de calcul...	116
VIII.B Intégration de la prise en compte de la qualité dans l'IHM	117
VIII.B.1 Evolutions dans la phase de préparation de la simulation	117
VIII.B.2 Evolutions dans la phase de simulation	121
VIII.B.3 Représentation des résultats.....	122
VIII.C Conclusion	129
Conclusion et perspectives.....	130
Conclusion.....	131
Perspectives	133
Bibliographie.....	135
Liste des publications	143

Liste des tableaux

Tableau III-1 : Typologie des erreurs de Fisher et de DeRunz	53
Tableau V-1 : Caractéristiques des données d'occupation du sol étudiées.	72
Tableau V-2 : Répartition des vergers du RPG (2006) dans la couche Corine Land Cover (2006).....	73
Tableau V-3 : Répartition des vergers du RPG (2006) dans la couche OCSOL (2006)	74
Tableau V-4: Liste et caractéristiques des données utilisées dans le modèle [Beaufils & al, 2011 bis]	75
Tableau V-5 : Ratios de consommation des ménages [Berge, 2012]	79
Tableau V-6 : Valeurs caractéristiques, d'après [Berge, 2012].....	87
Tableau V-7 : Caractérisation de l'année 2001 pour la maille SAFRAN n°9674 en matière de cumul de précipitations mensuelles.....	88
Tableau VI-1 : Liste des paramètres utilisés dans la modélisation.....	99
Tableau VII-1: Typologie des biais pouvant affecter la modélisation	103
Tableau VII-2 : Principaux chiffres issus de la comparaison des besoins en eau des abricotiers, cerisiers et pêchers [Beaufils & al, 2011 bis].	105
Tableau VIII-1 : Structure de la table zone d'étude.....	118
Tableau VIII-2 : Structure de la table Type_Parcelle.....	118
Tableau VIII-3: Valeurs statistiques obtenues	119
Tableau VIII-4 : Structure de la table Groupe_NomParametre.....	120
Tableau VIII-5 : Structure de la table Type_Parcelle.....	120
Tableau VIII-6 : Structure de la table Groupe_Parcelle.....	121
Tableau VIII-7 : Valeurs du 25 juillet 2003 utilisées pour les calculs	122
Tableau VIII-8 : Résultats de l'analyse de sensibilité simple	123
Tableau VIII-9 : Valeurs obtenues	125

Liste des figures

Figure 0-1 : Fonctionnalités offertes par l'IDS FUSIEAU [Beaufils & al, 2011]	18
Figure I-1: Localisation de la zone d'étude [Caballero, 2008].....	22
Figure I-2 : Phase de calibrage du modèle hydrogéologique	23
Figure I-3 : Utilisation du modèle hydrogéologique en mode prédiction	23
Figure I-4 : Détermination de l'évapotranspiration standard ET0	25
Figure I-5: Détermination de l'évapotranspiration ETc de la culture c.....	25
Figure I-6 : Exemple de variation des valeurs du Kc au cours du cycle de vie d'une culture.....	26
Figure I-7 : Détermination de l'évapotranspiration ajustée ETc adj de la culture c.....	26
Figure I-8 : Principe de la modélisation des prélèvements agricoles [Chazot, 2010]	27
Figure I-9 : Principe de la modélisation des prélèvements dédiés à l'alimentation en eau potable [Chazot, 2008]	29
Figure I-10 : Schéma synthétique du modèle AEP [Berge, 2012] d'après l'analyse de [BRL, 2009]..	30
Figure I-11 : Schéma synthétique du modèle du SDRE de l'Isère [Berge, 2012].....	31
Figure II-1 : Services web de diffusion de données	39
Figure II-2 : Schéma de principe de l'orchestration de services [Albreshne & al, 2009].....	40
Figure II-3 : Schéma de principe de la chorégraphie de services [Albreshne & al, 2009].....	41
Figure II-4 : Schéma de principe du chaînage de service centralisé	42
Figure II-5 : Schéma de principe du chaînage en cascade.....	43
Figure II-6 : Schémas de principe de la méthode 3.....	43
Figure III-1 : Niveaux de fusion de données [OGC, 2009].....	47
Figure III-2 : Image multispectrale SPOT XS prise sur Rotterdam (Pays-Bas). (Source : http://gdsc.nl/)	47
Figure III-3 : Mesures d'évapotranspiration moyennes sur la période 1970-2006 issues de SAFRAN pour le département des Pyrénées-Orientales [Caballero, 2008].....	48
Figure III-4 : Représentation schématique des concepts de qualité internet et externe [Devilleers, 2004]	50
Figure III-5 : Taxonomies de Fisher et DeRunz.....	51
Figure III-6 : Principe de sensibilité du modèle à deux variables	54
Figure III-7 : Propagation des incertitudes par le biais d'une approximation linéaire de la fonction f (en rouge) [André, 2008]	55
Figure IV-1 : Interface de STRATEAU pour la visualisation des données relatives aux prélèvements sur le bassin de la Têt. Source : [Ambassade de l'eau, 2009].....	60

Figure IV-2 : IDS mises en place dans le cadre de FUSIEAU pour le modèle d'estimation des besoins en eau agricoles	64
Figure IV-3 : Principe de fonctionnement général d'un algorithme WPS dans FUSIEAU	65
Figure IV-4 : Diagramme de séquence UML pour le calcul des ETc	67
Figure V-1 : Combinaison de données thématiques géoréférencées [Burrough, 1986]	70
Figure V-2 : Principe de croisement de deux données thématiques.....	71
Figure V-3 : Représentation schématisée de notre implémentation du modèle FAO	76
Figure V-4 : Schéma du modèle d'estimation des besoins en eau domestique issus de la population permanente, d'après [Berge, 2012].....	80
Figure V-5 : Carte de répartition des secteurs touristiques [Berge, 2012]	82
Figure V-6 : Schéma du modèle d'estimation des besoins en eau domestique issus du tourisme, d'après [Berge, 2012]	83
Figure V-7 : Unités de distribution en eau potable.....	85
Figure V-8 : Détermination des parcelles irriguées à partir du tracé des canaux et de la pente	86
Figure V-9 : Caractérisation des cumuls de précipitation enregistrés pour le mois d'août sur la période 1970-2005. Cas de la maille SAFRAN n°9674 près de Perpignan	88
Figure V-10 : Boîte à moustaches pour l'analyse des chroniques météorologiques.....	89
Figure VI-1 : Schéma de principe de l'architecture trois tiers	90
Figure VI-2 : Architecture simplifiée de l'IDS FUSIEAU	92
Figure VI-3 : Connexion de serveurs d'applications personnels à FUSIEAU.....	93
Figure VI-4 : Représentation schématique des traitements effectués pour la combinaison de géométries	95
Figure VI-5 : Représentation schématique des traitements effectués pour les calculs associés au modèle	96
Figure VI-6 : Module graphique de FUSIEAU	97
Figure VII-1 : Coefficients culturels des abricotiers, pêchers et cerisiers	104
Figure VII-2 : Comparaison des besoins journaliers en eau des abricotiers, cerisiers et pêchers [Beaufils & al, 2011 bis].....	105
Figure VII-3 : Illustration de deux modes distincts de répartition d'un même cumul de précipitation	106
Figure VII-4 : Approche de prise en compte des imperfections dans FUSIEAU	107
Figure VII-5 : Regroupement des parcelles en fonction de leurs caractéristiques	110
Figure VII-6 : Impact de l'ordre du polynôme de la fonction d'approximation.....	112
Figure VIII-1 : Schéma de principe du traitement WPS de calculs de valeurs moyennes (en blanc) et en option des incertitudes associées (en noir).....	116
Figure VIII-2 : Principe de regroupement des parcelles	117
Figure VIII-3 : Module de discrétisation.....	120
Figure VIII-4 : Possibilité de fixer la variable à une valeur constante	122

Figure VIII-5 : Diagramme de Pareto	123
Figure VIII-6: Diagramme tornade	124
Figure VIII-7 : Diagramme en toile d'araignée	124
Figure VIII-8 : Diagramme en toile d'araignée [Loucks, 2005]	125
Figure VIII-9 : Répartition de la variance du résultat	126
Figure VIII-10 : Représentation des résultats accompagnés de l'écart type mesuré	126
Figure VIII-11 : Affichage des résultats finaux pour une parcelle : valeurs calculées, écart-type estimé et sensibilité par rapport aux paramètres et variables d'entrées	127
Figure VIII-12 : Représentation de la sensibilité du résultat de l'estimation des besoins en eau de la culture	128
Figure VIII-13 : Illustration de la limite de la représentation en cas d'intervention multiple d'une variable ou d'un paramètre dans le modèle.	128

Introduction

Aujourd'hui la préservation de l'environnement est un sujet de préoccupation important. L'homme prend conscience que ses activités n'ont pas toujours des effets profitables sur l'environnement et cherche maintenant à supprimer ou tout du moins limiter les impacts néfastes de ses activités. La préservation de l'environnement constitue ainsi un des trois piliers du développement durable. C'est également un des huit objectifs pour le millénaire énoncés par l'Organisation des Nations Unies (ONU) pour le développement [ONU, 2000]. Parmi les sujets de préoccupation environnementale, il y a celui de la gestion des ressources en eau.

Bien que recouvrant près des trois quarts de la surface du globe terrestre, l'eau n'en est pas moins une denrée précieuse. En effet, seul trois pour cent de cette ressource est constitué d'eau douce, dont une infime partie est potable. Malgré cela l'eau fait l'objet de nombreux gaspillages et de pollution, mettant ainsi en péril la pérennité de cette précieuse ressource. Les changements climatiques influencent également de manière significative l'état et la disponibilité des ressources. Les tendances générales conduisent à une raréfaction des ressources en eaux. L'adoption de méthodes efficaces de gestion des ressources en eau est donc de plus en plus indispensable.

Toutefois les effets varient d'un endroit à l'autre, rendant ainsi la gestion des ressources en eau plus complexe. Ainsi selon le Groupe Intergouvernemental d'Experts sur l'évolution du Climat (GIEC), « les changements climatiques vont poser deux problèmes majeurs en Europe concernant l'eau : une augmentation du déficit en eau dans le sud de l'Europe et une augmentation des inondations dans d'autres endroits » [GIEC, 2008]. Ces changements vont entraîner la conception de nouveaux outils pour la gestion de l'eau, car « les méthodes actuelles risquent de ne pas être adéquates pour réduire les impacts négatifs des changements climatiques » [GIEC, 2008].

Compte tenu de cette variabilité spatiale importante des impacts du changement climatique, les nouveaux outils doivent s'appuyer sur des données géographiques. Cependant à l'heure actuelle l'acquisition de données géographiques reste un processus coûteux, aussi bien en temps que financièrement (pensez à la mise en place d'une campagne d'acquisition de données satellites, ou à l'organisation d'un recensement national).

Pour des problématiques aussi importantes que la gestion des ressources en eau, le partage des informations est vital. Il est donc nécessaire de mettre en place des stratégies et des directives, mais aussi des techniques dans le but de promouvoir le partage de l'information géographique.

La Directive 2007/2/CE du Parlement européen et du Conseil du 14 mars 2007 établissant une « infrastructure d'information géographique dans la Communauté européenne », dite directive INSPIRE, vise à définir des règles pour la mise en place d'une plateforme d'échange et de partage d'informations géographiques relatives à l'environnement au sein de la Communauté Européenne [INSPIRE, 2007]. On parle d'Infrastructure de Données Spatiales (IDS).

INSPIRE invite les différents fournisseurs de données publiques à portée environnementale à suivre ce modèle et donc à mettre à disposition leurs données sous des formats standards. La philosophie des IDS va au-delà du simple partage bilatéral des données. Elles permettent de passer d'un monde de récupération de données figées à un instant donné, à un accès continu et actualisé de l'information via internet. Cet accès s'appuie sur des services web standardisés permettant une interopérabilité, notamment ceux définis par l'Open Geospatial Consortium (OGC).

La définition d'IDS à portée environnementale s'intègre dans le cadre du projet GEOSS (Système mondial des systèmes d'observation de la Terre), dont l'objectif est le partage et l'échange de données et de connaissances offrant une meilleure compréhension de notre environnement [GEOSS, 2005]. Ces informations pourront servir plusieurs desseins : du simple porter à connaissance au grand public jusqu'à la définition de modèles prévisionnels pour les experts environnementaux en passant par l'aide à la prise de décision pour les décideurs.

Nos travaux de recherches œuvrent dans ce sens et nous proposons ainsi de traiter de la conception d'une IDS dédiée à l'évaluation des besoins en eau.

Pour cela, une collaboration de recherche a été mise en place. Une Convention Industrielle de Formation par la REcherche (CIFRE) mettant en scène trois partenaires a été signée : ACATUS Informatique, une Société de Services en Ingénierie Informatique (SSII), le Laboratoire de Géodésie et de Géomatique (L2G), ainsi que le Bureau de Recherches Géologiques et Minières (BRGM).

Cette thèse s'est principalement déroulée au sein des locaux d'ACATUS Informatique. Les développements informatiques mentionnés dans ce présent mémoire ont été réalisés par une équipe d'informaticiens d'ACATUS Informatique supervisée par l'auteur. Le BRGM nous a fourni son expertise à la fois dans les domaines de l'hydrogéologie (Direction Eau, Environnement & Ecotechnologies / Nouvelles Ressources et Economie de Montpellier) et des sciences et technologies de l'information (Direction des Systèmes d'Information / Information Scientifique et Technologies Numériques d'Orléans). Les données ne faisant pas l'objet d'une libre diffusion nous ont été mises à disposition par le BRGM.

Les objectifs fixés dans cette thèse sont de montrer l'intérêt de la mise à disposition de modèles environnementaux sur internet : du potentiel offert par le web comme plateforme d'échange, des avantages à s'appuyer sur des modèles ou applications cartographiques, mais également de la nécessité de suivre un formalisme rigoureux pour la diffusion des données et des traitements.

Le sujet traité est pluridisciplinaire et aborde les domaines de l'hydrologie, de l'agronomie, de la géomatique, de l'informatique et des mathématiques. Un important travail d'assimilation de

connaissances dans ces thématiques a été réalisé par l'auteur pour proposer des solutions innovantes non pas à l'intérieur d'un de ces domaines en particulier, mais plutôt au niveau de leur croisement.

Nos travaux s'inscrivent dans la continuité du projet VULCAIN qui propose lui-même une approche pluridisciplinaire liant climatologie, hydrogéologie et gestion des ressources en eau.

De VULCAIN à FUSIEAU

Le projet VULCAIN a consisté en l'étude de la vulnérabilité aux changements climatiques des ressources en eau du département des Pyrénées-Orientales aux horizons 2030, puis 2060. Les travaux de recherches ont principalement consisté en l'étude du climat (actuel et futur), la construction de modèles hydrogéologiques des bassins versants, ainsi que la création de modèles d'estimation des prélèvements en eau pour les besoins agricoles et domestiques.

La réalisation du projet, et en particulier, la modélisation des prélèvements a conduit à la collecte d'une quantité importante de données (population permanente et saisonnière, pratiques culturelles et agricoles, quantités d'eau facturées par les organismes distributeurs d'eau). Cette étape de récupération des données a occupé une partie conséquente du projet et a permis de constituer une base de données jusque là inédite. A l'issue de cet effort, les instigateurs se sont interrogés sur la réutilisabilité des travaux réalisés au cours de leur projet [Caballero, 2010].

Partant de cette problématique, nous proposons de valoriser les travaux réalisés au cours du projet VULCAIN par le biais de la création d'une IDS, baptisée FUSIEAU, dédiée à l'évaluation des besoins en eau à l'échelle des bassins versants. Elle se présente comme un portail d'information et de téléchargement, mettant à disposition les données disponibles ou recensant les données et leurs producteurs, dans le cas où elles ne sont pas accessibles gratuitement. De plus, dans nos travaux de recherches nous nous sommes également intéressés aux capacités de traitement de données sur le web, fonction rare dans les IDS.

L'objectif de nos travaux est de proposer une IDS permettant non seulement de mettre à disposition des données, mais également des traitements ou calculs. Une telle application permettrait ainsi de mettre à disposition les modèles produits et de les utiliser directement via l'interface de l'IDS. A terme, l'outil sera décliné en plusieurs versions destinées à trois types d'utilisateurs : des experts hydrologues, les décideurs de la zone d'étude et le grand public (Figure 0-1).

Pour les experts hydrologues, FUSIEAU sera avant tout un simulateur. L'utilisateur aura accès à l'intégralité des données et traitements utiles à l'évaluation des ressources en eau. Il pourra construire des scénarii d'évolution de l'occupation du sol, des pratiques de cultures et du climat en définissant ses propres paramètres, lancer les simulations, afficher les résultats et les télécharger. L'utilisateur

aura accès à de nombreux paramètres, permettant un niveau poussé de personnalisation des scénarii. Pour les décideurs, FUSIEAU sera un outil d'aide à la décision dans lequel il pourra trouver des cartes et des informations lui permettant de connaître les situations actuelles et futures des ressources en eau. Il pourra ainsi visualiser si les projets d'aménagement futurs risquent de mettre en danger la gestion des ressources en eau. Pour le grand public, l'IDS fournira des informations générales relatives à l'évaluation des besoins en eau.

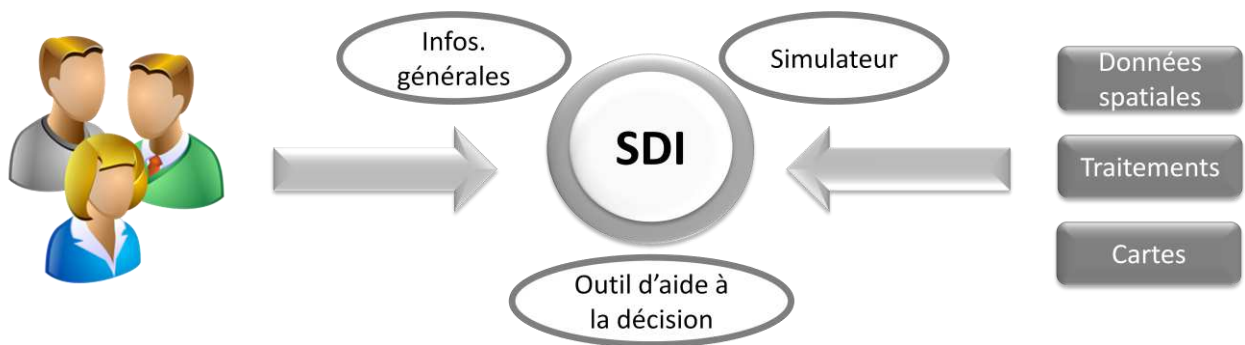


Figure 0-1 : Fonctionnalités offertes par l'IDS FUSIEAU [Beaufils & al, 2011]

La réalisation d'un tel outil nous invite à traiter plusieurs problématiques, parmi lesquelles :

- la définition de modèles génériques applicables à plusieurs zones d'études et à diverses échelles tout en étant adaptés au contexte web et à ses contraintes (quantité de données échangées limitée) ;
- l'exploitation et la combinaison de données et de connaissances hétérogènes (données à différentes échelles spatiales, mesures à divers pas de temps) ;
- la diffusion des résultats et des traitements dans des formats et protocoles permettant leur réutilisation en tout ou partie dans d'autres applications ;
- l'intégration et la propagation de la qualité des données à travers le modèle, sous la forme d'intervalles de confiance ;
- la communication des résultats par le biais de représentations adaptées aux utilisateurs.

Le traitement de ces problématiques nécessite une approche pluridisciplinaire. Les modèles doivent non seulement être pertinents d'un point de vue métier, mais également être viables d'un point de vue technique. Pour illustrer notre approche, nous nous sommes concentrés sur l'implémentation de modèles d'estimation des besoins en eaux agricoles et domestiques. Nos travaux de recherches comportent trois parties distinctes :

- un état de l'art sur les problématiques abordées par nos travaux ;
- la définition d'une approche conceptuelle pour le traitement de ces problématiques ;
- la conception d'un prototype implémentant les méthodes définies précédemment. Cet outil a été conçu pour un utilisateur expert.

Plan du mémoire

Dans une première partie consacrée à l'état de l'art, nous présentons successivement les domaines abordés lors de nos travaux de recherches. Nous commençons par introduire VULCAIN et les travaux réalisés au cours de ce projet. Puis nous nous intéressons aux caractéristiques et spécificités des IDS et de leurs principales composantes : les services web. Enfin nous abordons les notions de fusion de données, ainsi que celles de qualité de l'information géographique.

Dans la deuxième partie de ce mémoire, nous exposons notre démarche dans la conception d'une IDS dédiée à l'évaluation des besoins en eau agricoles et domestiques. Nous présentons les méthodes et concepts utilisés pour l'intégration des modèles dans l'IDS. Puis nous traitons de la conception du prototype découlant des méthodes définies précédemment.

Enfin dans une troisième partie, nous traitons de la prise en compte de la qualité des données dans notre IDS. Nous proposons des méthodes et concepts pour traiter cette problématique, et indiquons enfin les travaux à mettre en œuvre pour leur implémentation dans l'IDS existante.

Première partie

-

Un sujet pluridisciplinaire

Chapitre I : Evaluation des besoins en eau

Dans ce chapitre, nous présentons un aperçu global de VULCAIN, ainsi que de l'ensemble des travaux réalisés au cours de ce projet. Nous étudions plus particulièrement les modèles d'estimation des besoins en eau agricoles et domestiques.

I.A Le projet VULCAIN

I.A.1 Introduction

L'acronyme VULCAIN signifie « Vulnérabilité des hydrosystèmes face aux effets combinés des changements climatiques et aux activités humaines en zone méditerranéenne ». Il s'agit d'un projet de recherche débuté en 2006 et terminé en 2010. Soutenu par l'Agence Nationale pour la Recherche (ANR), le projet VULCAIN fut mené par le BRGM, accompagné de plusieurs partenaires dont le bureau d'études BRL Ingénierie, le laboratoire HydroSciences de l'Université de Montpellier, ainsi que le département recherche de Météo France.

Le projet propose une modélisation des impacts combinés des changements climatiques et socio-économiques sur les hydrosystèmes méditerranéens. Ces impacts sont étudiés à deux horizons temporels : à moyen terme (2020-2040) et à long terme (2040-2060). La prise en compte des activités humaines dans une modélisation hydrogéologique constitue la principale originalité du projet.

VULCAIN traite des impacts des changements climatiques sur les ressources en eau et les activités économiques, des effets de l'état des ressources en eau sur les activités économiques, et également des impacts des activités socio-économiques sur les ressources en eaux elles-mêmes.

Le projet a été appliqué dans le département des Pyrénées-Orientales, situé en région Languedoc Roussillon. Avec une superficie d'environ 4100 km², le département des Pyrénées-Orientales comporte trois bassins hydrographiques. Ils sont respectivement associés, du nord au sud, aux fleuves de l'Agly, de la Têt et du Tech (Figure I-1).

Cette zone d'étude a été choisie pour deux raisons majeures. La première est qu'elle comporte des hydrosystèmes typiques de la région méditerranéenne. On y retrouve ainsi « des bassins versants côtiers contrôlés par des barrages, des aquifères multi-couches côtiers sédimentaires et alluviaux, ainsi que des systèmes karstiques alimentés par les pluies mais aussi par les eaux de surface (pertes de rivières) » [Caballero, 2008]. La seconde est qu'il s'agit d'une des zones méditerranéennes où la

gestion des ressources en eaux est des plus critiques, avec un rapport besoins en eaux / ressources disponibles très élevé. La prise en compte de la demande en eaux est donc primordiale.

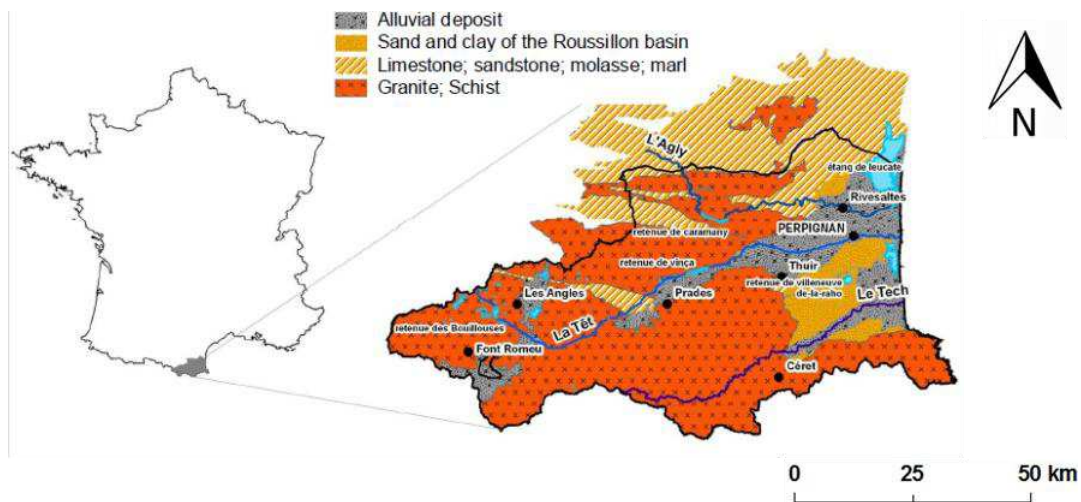


Figure I-1: Localisation de la zone d'étude [Caballero, 2008]

I.A.2 Modélisation hydrologique selon une approche globale

Chacun des trois bassins hydrographiques de la zone d'étude a été découpé en sous-bassins versants : quatre pour l'Agly, six pour la Têt, deux pour le Tech, et une zone littorale appelée Plaine du Roussillon. Nous nous intéressons plus particulièrement à la modélisation hydrologique du bassin versant amont du Tech en zone de socle. Il s'agit d'un bassin versant de surface, ce qui facilite la modélisation, d'autant que le cours d'eau ne présente pas de barrage.

Dans le cadre de la modélisation du bassin versant du Tech, une modélisation hydrologique de type pluie / débit a été retenue. Elle est réalisée au moyen du logiciel TEMPO. Il s'agit d'un outil de traitement du signal développé par le BRGM. Il est destiné au traitement et à la modélisation de séries temporelles en hydrogéologie et en hydrogéochimie [Pinault, 2001]. Dans le cadre de VULCAIN, il est utilisé pour déterminer la fonction mathématique qui retranscrit au mieux la transformation des précipitations en débit, mesuré à l'exutoire du bassin versant.

La construction du modèle s'appuie sur les chroniques historiques de précipitations et de débits observées à l'exutoire du bassin versant. Pour le bassin versant du Tech, il s'agit de la station hydrométrique de Reynès. Les mesures sont disponibles depuis 1970 tant pour la météorologie que pour l'hydrométrie. Le calage du modèle s'appuie donc sur près de trente-six années de mesures.

Pour une modélisation plus fidèle des hydrosystèmes, le projet VULCAIN prend également en compte les prélèvements en eaux (liés à l'irrigation et à l'alimentation en eau potable). Ces derniers représentent en effet jusqu'à 25% du débit du cours d'eau dans les mois estivaux, où la gestion des ressources en eau est des plus critiques. En ajoutant le débit prélevé au débit mesuré à l'exutoire du

bassin versant au débit prélevé, on obtient un débit naturalisé qui représente le débit naturel théorique obtenu en l'absence de prélèvements sur le cours d'eau (Figure I-2).

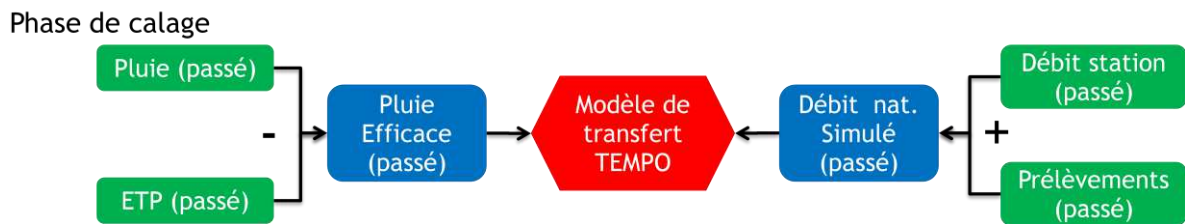


Figure I-2 : Phase de calibrage du modèle hydrogéologique

Le modèle une fois calibré permet de faire ce qu'on appelle une modélisation inverse. Il s'agit de reproduire les valeurs de débits qu'on obtiendrait en réponse à partir des chroniques météorologiques. Considérant l'hypothèse que le comportement de l'hydrosystème ne devrait pas évoluer dans le futur, la modélisation inverse réalisée dans VULCAIN permet ainsi à partir de prévisions météorologiques d'obtenir une estimation des débits futurs dans le cours d'eau.

Le projet a ainsi donné lieu à l'examen de différentes variables climatiques. Le changement climatique est d'une part simulé à partir de l'analyse des mesures climatiques effectuées au cours des dernières années, et d'autre part avec différents modèles climatiques prospectifs établis lors des dernières études du Groupe Intergouvernemental sur l'Evolution du Climat (GIEC).

Le modèle hydrogéologique permet ainsi d'estimer le débit naturalisé à l'exutoire pour ces différents scénarii climatiques. Les prélèvements futurs devront alors être soustraits pour obtenir une estimation du débit réel dans le cours d'eau tel que mesuré à la station (Figure I-3).

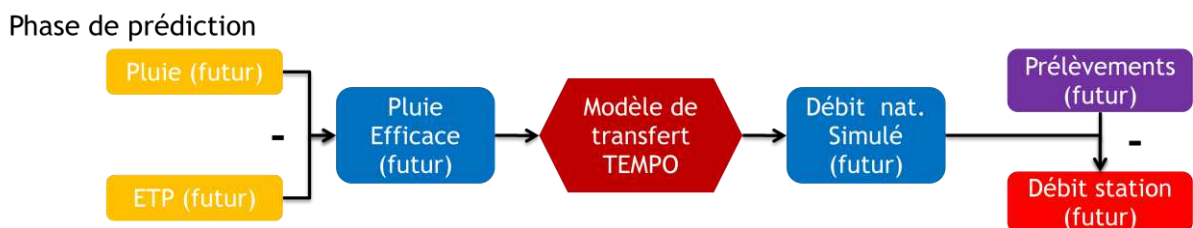


Figure I-3 : Utilisation du modèle hydrogéologique en mode prédiction

L'estimation des prélèvements futurs est construite à partir des prélèvements observés actuellement ainsi que des hypothèses ou scénarii d'évolutions. Cette modélisation des prélèvements en eau constitue une phase primordiale. En effet de ces résultats vont dépendre la construction de la modélisation hydrogéologique elle-même. VULCAIN considère les consommations d'eau destinées à un usage domestique ainsi que celles associées aux consommations d'eau agricoles. Les consommations d'eau imputables aux industries ont quant à elles été ignorées. Elles sont en effet

négligeables sur notre zone d'étude par rapport aux deux sources de consommation citées précédemment.

Les modèles retenus pour VULCAIN pour la modélisation des prélèvements reposent sur des approches pluri-disciplinaires. Ils proposent de combiner des données d'origines diverses et présentant des hétérogénéités, notamment en matière d'échelle spatiale et de granularité. L'hétérogénéité des données complique leur combinaison. Qui plus est l'indisponibilité des données sur une longue période empêche de réaliser un modèle calibré sur un grand échantillon de données. Ainsi le calage du modèle n'a été fait que sur la période de 2000 à 2005.

D'un point de vue technique, les méthodes d'estimation des prélèvements en eau développées pour VULCAIN ont été implémentées dans deux tableurs Excel séparés : un pour chacun des deux postes de consommations en eau considérés dans le projet. Ce choix technique rend la mise à jour et la réutilisabilité des modèles assez fastidieuse. Nous présentons dans les paragraphes suivants différentes méthodes, dont celles adoptées dans VULCAIN, pour estimer les prélèvements en eau agricoles et domestiques.

I.B Modélisation des prélèvements en eau agricole

Le modèle de la FAO (détaillé ci-dessous) se présente comme un candidat intéressant pour estimer les besoins en eau agricole. Il est utilisé dans le monde entier et s'appuie sur des calculs relativement simples. De plus, il n'est pas très différent de celui retenu dans le projet VULCAIN (également présenté) ce qui permet de réutiliser une partie des données utilisées dans ce modèle.

I.B.1 Modèle FAO

La méthode introduite par la Food and Agricultural Organization (FAO) de l'Organisation des Nations Unies (ONU) est très certainement l'une des plus couramment utilisées pour l'évaluation des besoins en eau des cultures. Cette approche est fondée sur le principe que le besoin en eau des cultures correspond à la quantité d'eau nécessaire pour compenser l'effet de l'évapotranspiration.

L'évapotranspiration correspond à la combinaison des pertes en eau subies par phénomènes d'évaporation et de transpiration. Ces deux phénomènes portent sur la transformation de l'eau liquide en vapeur. L'évaporation correspond à l'eau perdue par le sol, tandis que la transpiration correspond à celle perdue par les pores de la plantes. La répartition de l'évapotranspiration entre évaporation et transpiration évolue au cours du cycle de vie de la culture. Toutefois en pratique, il est difficile d'évaluer avec précision cette répartition [Allen & al, 1998].

L'évapotranspiration dépend de divers paramètres climatiques : pression atmosphérique, vitesse du vent, taux d'hygrométrie et taux d'ensoleillement. Plusieurs méthodes de calcul permettent de la

déterminer. L'approche conseillée dans le cadre d'une utilisation du modèle FAO est celle dite de Penman-Monteih. La valeur d'évapotranspiration standard ET_0 est définie comme l'eau consommée par une plante de référence : l'herbe (Figure I-4).

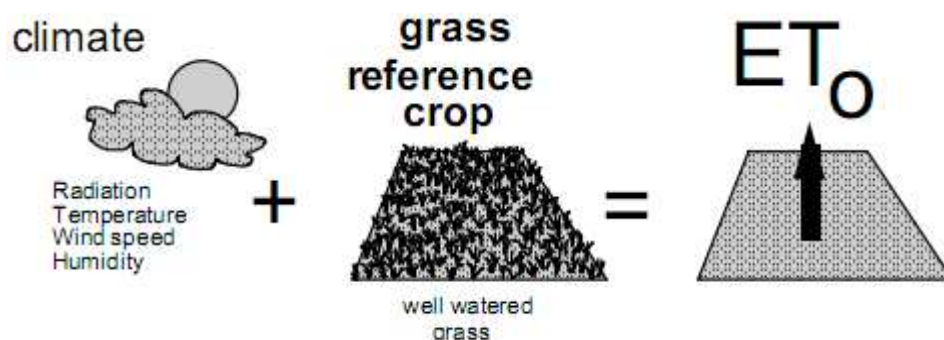


Figure I-4 : Détermination de l'évapotranspiration standard ET_0

Pour déterminer l'évapotranspiration ET_c d'une plante particulière, on multiplie cette valeur par un coefficient cultural K_c (Figure I-5). La FAO présente également une seconde approche basée sur l'utilisation de deux coefficients culturaux appelés K_e et K_{cb} . Dans ce cas on distingue les pertes en eau engendrées par évaporation (associées à K_e) et par transpiration (associées à K_{cb}) de la plante. Particulièrement adaptée à l'agriculture de précision et à la planification quotidienne de l'irrigation, l'approche à deux coefficients culturaux ne présente pas d'intérêt dans nos travaux. Nous nous intéressons donc uniquement à l'approche à un coefficient cultural.

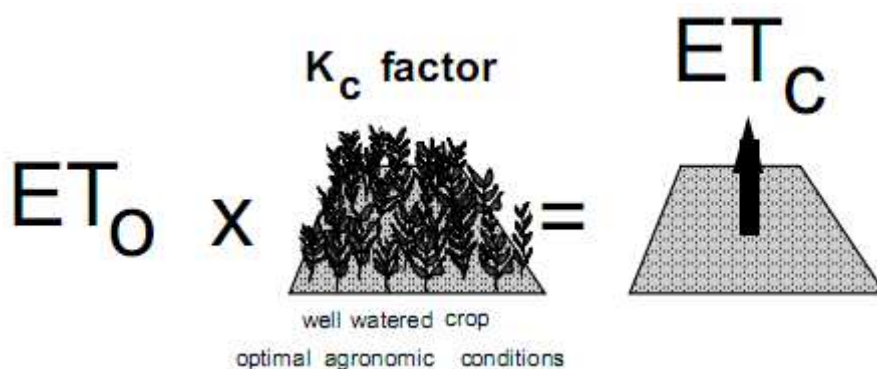


Figure I-5: Détermination de l'évapotranspiration ET_c de la culture c.

Les valeurs des coefficients culturaux prennent en compte les caractéristiques des cultures ainsi que les effets de l'évaporation dans le sol. De plus elles varient au cours du cycle de vie de la culture (Figure I-6). La FAO référence des valeurs de coefficients culturaux ainsi que les dates et périodes de chacun des cycles de vie pour diverses cultures [Doorenbos & al, 1979]. Les valeurs indiquées peuvent toutefois varier considérablement en fonction du lieu de culture, du climat et de la variété exacte cultivée. Pour obtenir une modélisation fidèle, il convient d'ajuster ces valeurs aux conditions locales de cultures.

Par ailleurs, la valeur d'ETc obtenue par l'équation décrite à la Figure I-5 correspond à des conditions optimales de culture, c'est-à-dire qu'on formule l'hypothèse d'une disponibilité de l'eau permanente qui subvient à chaque instant aux besoins de la culture. Or cette hypothèse n'est que peu souvent applicable.

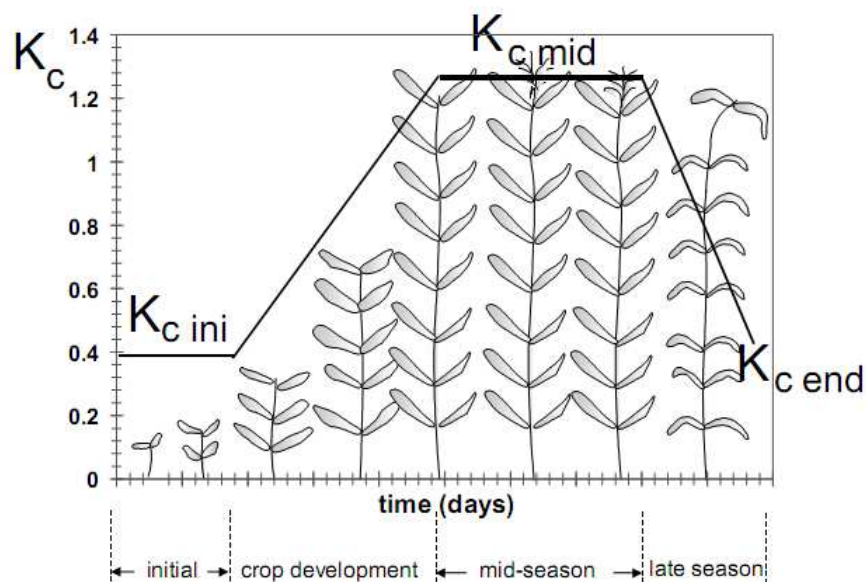


Figure I-6 : Exemple de variation des valeurs du Kc au cours du cycle de vie d'une culture.

Lorsque les besoins en eau d'une culture ne sont pas entièrement satisfaits, celle-ci subit ce que l'on appelle un stress hydrique. Dans le modèle, ce phénomène est traduit par le coefficient K_s (dont la valeur est comprise entre 0 et 1). La valeur ajustée de ETc prenant en compte K_s est alors appelée ETc Adj (Figure I-7).

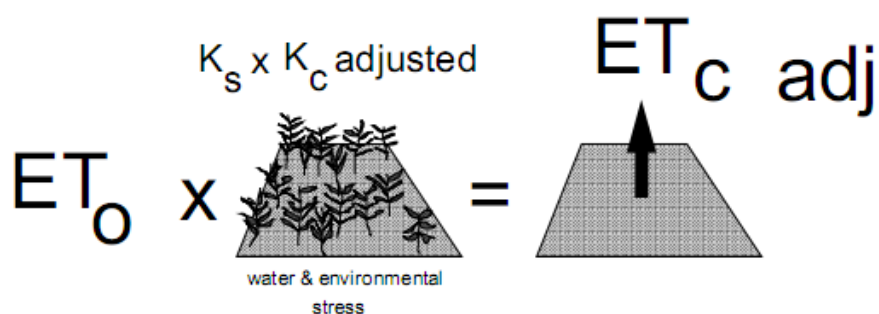


Figure I-7 : Détermination de l'évapotranspiration ajustée ETc adj de la culture c.

Le calcul du coefficient de stress hydrique K_s est principalement basé sur l'évaluation de la quantité d'eau présente dans le sol et pouvant lui servir de réserve. L'état de cette réserve varie en fonction du climat, des besoins en eau de la plante et de l'irrigation fournie par l'agriculteur. Lorsque l'eau tend à se raréfier, la plante va concentrer ses efforts dans la recherche de sources d'eau alternatives. Elle va pour cela privilégier le développement de son réseau racinaire plutôt que sa croissance propre. La

connaissance de l'état de la réserve en eau permet de savoir à quel instant la plante va commencer à avoir ce comportement.

A noter, que le modèle FAO connaît une déclinaison informatique : il s'agit du logiciel CropWat, développé par la Division Développement Terrestre et Aquatique de la FAO. Ce logiciel s'appuie sur les deux publications de la FAO en matière d'irrigation, respectivement les n°33 [Doorenbos & al, 1979] et n°56 [Allen & al, 1998].

I.B.2 Modèle VULCAIN

L'estimation des prélèvements en eaux agricoles se fait sur la base d'un calcul des besoins en irrigation des cultures définis, comme dans le modèle FAO, par la quantité d'eau à apporter à la culture pour compenser son évapotranspiration. La modélisation prend également en compte les pertes liées à l'acheminement des eaux. Ces dernières ne sont en effet pas négligeables dans cette zone d'étude où l'irrigation est encore largement assurée par des canaux. L'addition de ces deux composantes forme ce qui est appelé le prélèvement net, c'est-à-dire la quantité d'eau utilisée pour subvenir aux besoins.

Cette valeur ne doit pas être confondue avec le prélèvement brut qui correspond à la quantité d'eau effectivement prise dans le cours d'eau. Le surplus d'eau prélevé qui n'est pas consommé par les cultures ou évaporé lors de l'acheminement des eaux est considéré comme retournant au milieu, soit par ruissellement, soit par infiltration dans le sol. Elle est donc de nouveau disponible pour les zones situées en aval de son lieu de retour (Figure I-8).

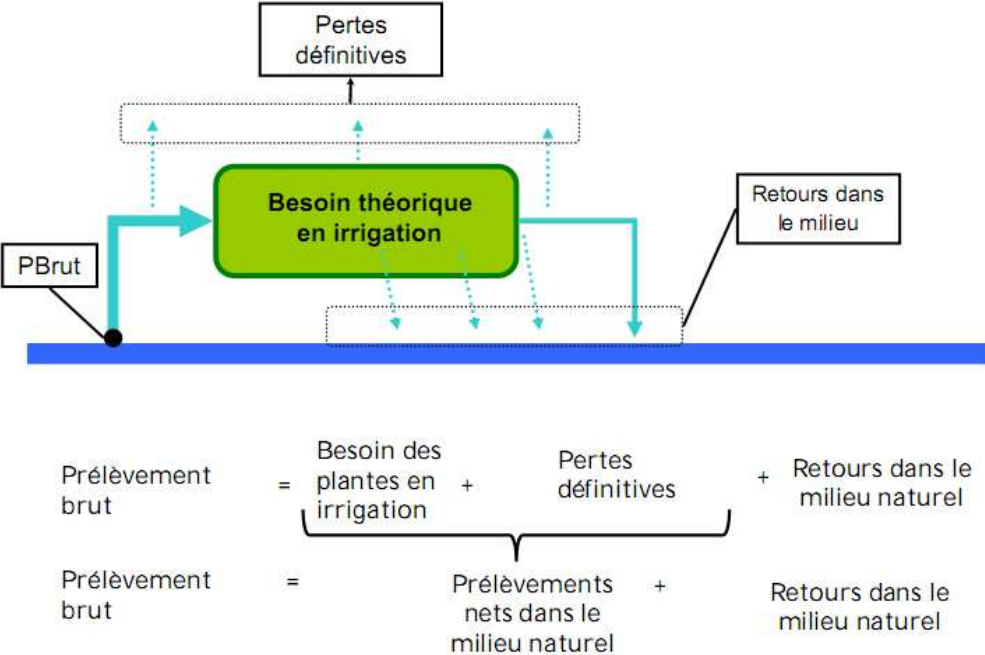


Figure I-8 : Principe de la modélisation des prélèvements agricoles [Chazot, 2010]

En ce qui concerne le calcul des besoins en irrigation des cultures, il s'agit d'estimer la quantité d'eau dont les cultures ont besoin pour vivre compte tenu de leur stade de développement et des paramètres climatiques. Le pas de temps retenu pour cette estimation est celui de la décade. Le modèle retenu pour VULCAIN est une version simplifiée de la méthode introduite par [Allen & al, 1998]. Ainsi, le besoin théorique unitaire en irrigation d'une plante i sur une zone k de surface unitaire pour la décade j est déterminé par la relation suivante :

$$B(i, j, k) = \sum_j \max [0, (Kc(k, i, j) * ETP(k, j) - P(k, j)) - RU(k, j - 1)]$$

Où :

- $Kc(k, i, j)$ correspond à la valeur du coefficient de la culture i sur la zone k au cours de la décade j (fonction du stade de développement de la culture),
- $ETP(k, j)$ correspond à l'évapotranspiration moyenne mesurée au cours de la décade j , sur la zone k (mm),
- $P(k, j)$ correspond à la pluviométrie moyenne mesurée au cours de la décade j , sur la zone k (mm),
- $RU(k, j-1)$ correspond à la réserve en eau disponible sur la zone k à la fin de la décade $j-1$ et donc au début de la décade j (mm). Cette valeur est actualisée à chaque décade telle que :

$$RU(k, j) = \max [0, RU(k, j - 1) + P(k, j) - Kc(k, i, j) * ETP(k, j)]$$

A partir de ces calculs, on en déduit la quantité d'eau à apporter par irrigation. Cette détermination se base sur l'hypothèse qu'il s'agit de la quantité d'eau manquante dans la réserve utile.

Pour obtenir une estimation de la quantité d'eau prélevée, un ratio prenant en compte les pertes dues à l'acheminement de l'eau est appliqué. Ce ratio est déterminé à dire d'expert. Enfin, pour obtenir les volumes d'eau, on multiplie les besoins unitaires par la superficie des zones de cultures. Un coefficient supplémentaire peut également être appliqué pour tenir compte de la densité culturale.

I.C Modélisation des prélèvements en eau domestique

On assimile l'eau domestique à celle issue de la cuisine (hors alimentaire), la salle de bain et les WC. Ces trois postes constituent, selon le Centre d'Information sur l'Eau, 93% de la consommation en eau d'un français à son domicile [CIEAU, 2011]. Le code de l'environnement précise ainsi à l'article R214-5 :

« Constituent un usage domestique de l'eau, au sens de l'article L. 214-2, les prélèvements et les rejets destinés exclusivement à la satisfaction des besoins des personnes physiques propriétaires ou locataires des installations et de ceux des personnes résidant habituellement sous leur toit, dans les

limites des quantités d'eau nécessaires à l'alimentation humaine, aux soins d'hygiène, au lavage et aux productions végétales ou animales réservées à la consommation familiale de ces personnes. »

Nous détaillons ci-après deux modèles visant à estimer les prélèvements en eaux domestiques : le modèle VULCAIN dans les Pyrénées-Orientales, et le modèle du SDRE de l'Isère.

I.C.1 Modèle VULCAIN

La méthode de détermination des prélèvements en eau domestiques se base sur une estimation préalable des besoins de la population. Ils sont estimés par le croisement de données démographiques et d'un ratio de consommation par habitant. Le rendement des réseaux est ensuite pris en compte pour relier les prélèvements aux besoins (Figure I-9).

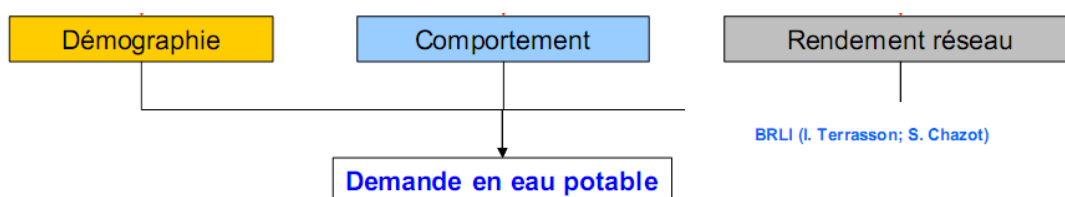


Figure I-9 : Principe de la modélisation des prélèvements dédiés à l'alimentation en eau potable
[Chazot, 2008]

Le modèle a été réalisé à l'échelle communale. Il propose de déterminer les prélèvements annuels, puis un changement d'échelle est réalisé pour estimer des valeurs mensuelles. Le modèle comporte trois parties distinctes (Figure I-10) :

- la première étape consiste à déterminer la « population équivalente ». Il s'agit de la population totale participant aux consommations en eau. Elle combine population permanente et population saisonnière. On estime ensuite cette population équivalente jusqu'en 2060 à partir d'hypothèses de variations des populations permanentes et saisonnières ;
- la seconde étape consiste à estimer le volume d'eau facturé. Cette quantité d'eau est assimilée aux besoins en eau domestiques de la population. En divisant cette valeur par la population de la commune, on en déduit des ratios de consommations. Ces derniers sont ensuite appliqués aux populations équivalentes futures pour en déduire une estimation des besoins en eau domestiques prévisionnels ;
- la troisième étape consiste à estimer les prélèvements qui découlent de ces volumes d'eau facturés. En rapprochant les volumes facturés et prélevés aux mêmes moments pour les mêmes communes, on en déduit des coefficients de rendement. On applique ces derniers aux volumes facturés prévisionnels pour en déduire une estimation des volumes prélevés prévisionnels ;

Les données sur la population des communes sont issues des bases de données sur le recensement de l'Institut National de la Statistique et des Etudes Economiques (INSEE). Celles relatives aux quantités d'eau prélevées et factures sont issues d'archives de différentes communautés de communes ou sociétés distributrices (Veolia, SAUR...), chargées de l'alimentation en eau potable des communes de la zone d'étude. Elles s'appuient également sur des données recueillies et mises à disposition par l'Agence de l'Eau Rhône Méditerranée Corse (AERMC).

Toutefois le pas de temps annuel n'étant pas suffisant pour la modélisation hydrogéologique, une reconstitution des prélèvements a été menée au pas de temps mensuel. De telles données étaient néanmoins disponibles pour certaines communes. Pour les autres communes pour lesquelles on ne disposait pas de l'information, une estimation a été réalisée sur la base d'un comportement identique à celui des communes voisines.

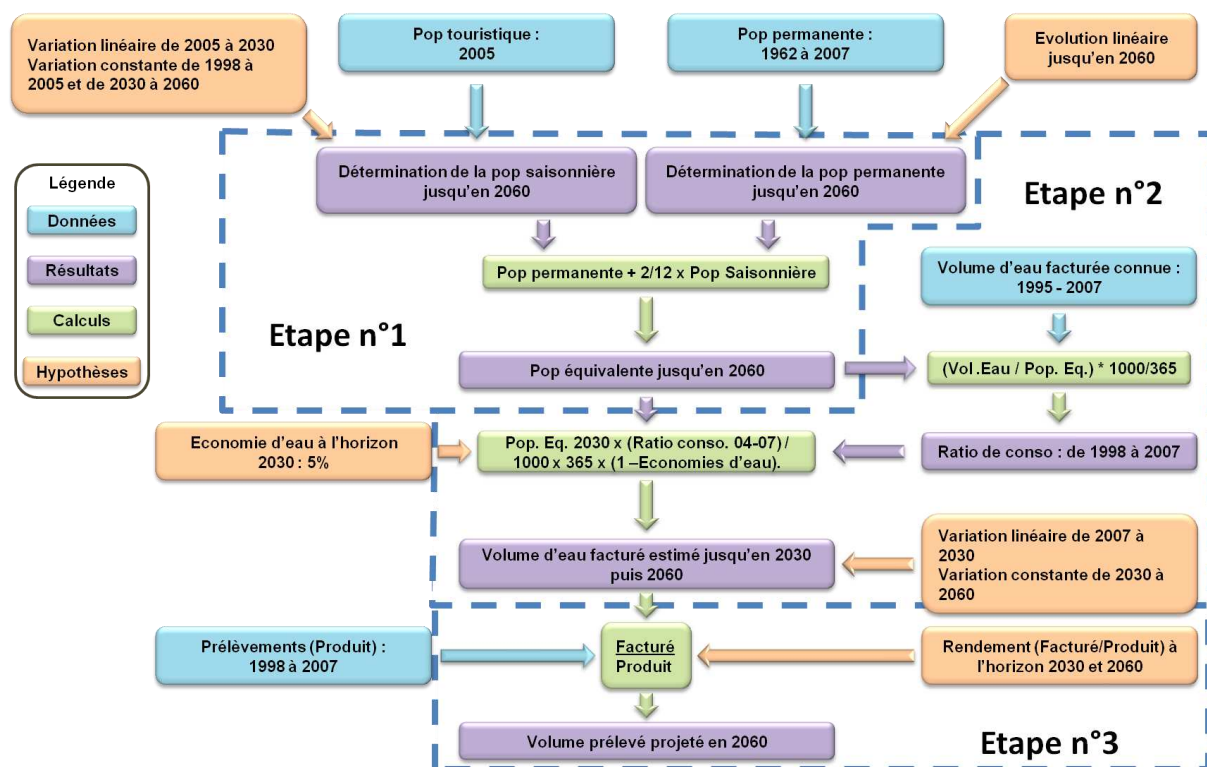


Figure I-10 : Schéma synthétique du modèle AEP [Berge, 2012] d'après l'analyse de [BRL, 2009]

Une autre tâche a également été de relier les prélèvements aux consommations. Certains prélèvements alimentent plusieurs communes, et réciproquement certaines communes sont alimentées par plusieurs sources de prélèvements. Pour effectuer ces recoupements, les instigateurs du projet VULCAIN se sont appuyés sur les informations fournies par les organismes distributeurs d'eau.

I.C.2 Modèle du SDRE de l'Isère

Situé entre la vallée du Rhône et le massif des Alpes, tout en étant assez éloigné de la mer, le département de l'Isère connaît un climat très varié ainsi que des saisons très marquées. Cette spécificité rend la gestion des ressources en eau difficile et a conduit les décideurs à la réalisation d'un Schéma Départemental de la Ressource en Eau et de ses usages (SDRE). Il s'agit actuellement du seul de France. Né de la collaboration de la Préfecture et de la Direction Départementale de l'Agriculture de la Forêt (DDAF) de l'Isère, le SDRE a deux objectifs principaux :

- améliorer la gestion quantitative de la ressource en eau ainsi que ses usages à l'horizon 2015 (échéance fixée par la Directive Cadre sur l'Eau) ;
- améliorer la gestion qualitative en contrôlant l'évolution de la qualité de l'eau et des milieux aquatiques.

Pour répondre à la première mission citée, un modèle d'estimation des prélèvements associés à la distribution publique a été réalisé (Figure I-11).

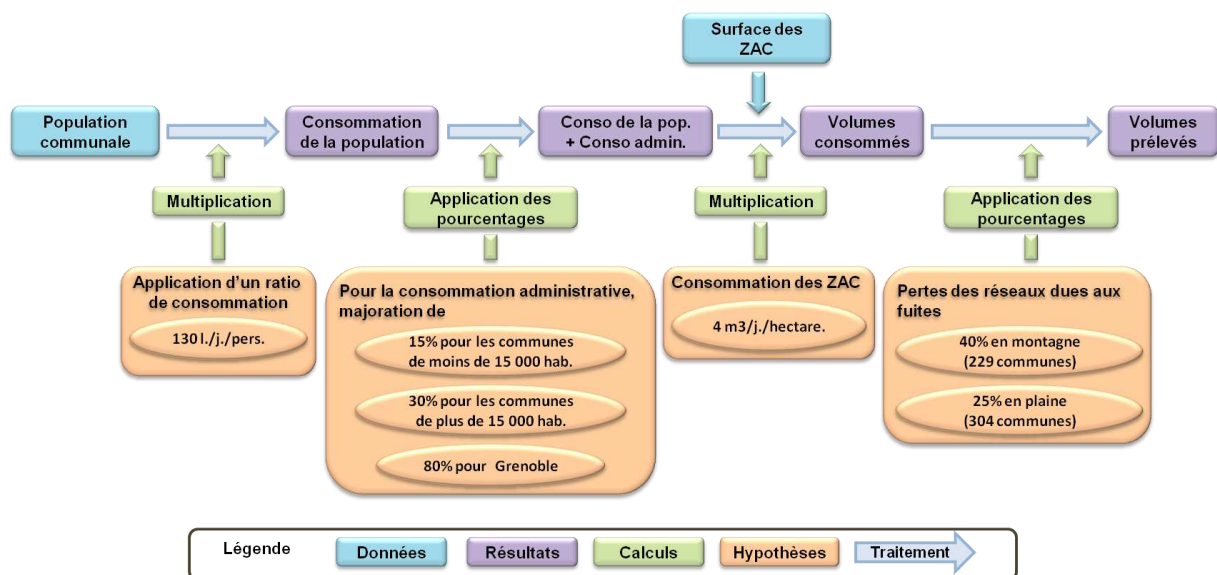


Figure I-11 : Schéma synthétique du modèle du SDRE de l'Isère [Berge, 2012]

Dans l'ensemble, le modèle proposé est assez proche de celui du projet VULCAIN :

- on commence par déterminer les besoins en eau domestiques de la population permanente en multipliant la population par un ratio de consommation (la différence avec VULCAIN est qu'on utilise ici un ratio de consommation fixe de 130 litres par jour par personne) ;
- ensuite on applique deux coefficients de majorations :
 - o un premier pour tenir compte des consommations des services publics (arrosage des espaces verts, consommation dans les bâtiments publics et administratifs) ;
 - o un second pour tenir compte des consommations des Zones d'Activités Commerciales (ZAC) ;

- enfin on applique un coefficient de rendement pour tenir compte des pertes dans les réseaux d'acheminement de l'eau. On obtient ainsi une estimation des volumes prélevés.

L'absence de prise en compte du tourisme dans le SDRE s'explique par le fait que l'afflux touristique se fait principalement en hiver, période où les ressources en eau sont plus disponibles.

L'étude des modèles existants montre une volonté des modélisateurs : celle de proposer des modèles adaptés à leur zone d'étude. La réalisation d'un document tel qu'un SDRE s'appuie sur une analyse préalable du territoire qui va permettre d'identifier les postes de consommation principaux. Ainsi nous avons pu voir que si le tourisme génère un surplus de besoin en eau domestiques non négligeable dans les Pyrénées-Orientales, il n'a pas été jugé opportun de le prendre en compte dans le département de l'Isère.

Ces choix vont entraîner la collecte des données associées à ces différentes postes de consommation. Parfois la disponibilité des données, va influencer la structure même du modèle. La disponibilité d'une donnée locale spécifique comme une enquête de terrain, va engendrer la réalisation d'un modèle spécifique, qui sera certes adapté à la zone d'étude mais ne sera bien souvent pas utilisable ailleurs. Cet aspect nuit à l'interopérabilité des modèles entre eux.

Au contraire, proposer un modèle s'appuyant sur des données disponibles sur une grande couverture territoriale (à l'échelle nationale par exemple) garantit la réutilisabilité d'un modèle. Elle sera d'autant plus forte si les données sont mises à disposition gratuitement aux utilisateurs et régulièrement mises à jour.

I.D Conclusion du chapitre

Au cours de ce chapitre, nous avons étudié des modèles traitant de problématiques proches des nôtres. Pour l'estimation des besoins en eau agricole, le modèle FAO s'avère être un bon candidat pour nos travaux. C'est un modèle simple à mettre en œuvre, dont l'approche est largement applicable. Il est ainsi utilisé dans le monde entier. Pour répondre à nos besoins, il reste toutefois à l'adapter pour qu'il s'appuie sur des données géographiques. Pour l'estimation des besoins en eau domestiques, les modèles que nous avons étudié ne correspondent pas tout à fait à nos besoins. Ils ne sont en effet applicables que dans la zone d'étude pour laquelle ils ont été développés. Notre souhait étant de proposer des modèles réutilisables, nous nous orientons vers la définition d'un nouveau modèle.

Chapitre II : IDS et services interopérables

Dans ce chapitre, nous présentons un état de l'art des technologies envisagées pour le projet. Nous passons en revue plusieurs termes techniques et rappelons leur définition. Notamment le terme d'IDS dont l'implémentation correspond à l'objectif principal de ces travaux de recherche.

II.A Infrastructures de Données Spatiales

II.A.1 Définition et composition

Le terme d'Infrastructure de Données Spatiales (IDS) a été pour la première fois introduit en 1993 par le Conseil National de Recherche des Etats-Unis d'Amérique [Masser, 2009]. Une IDS est ainsi définie comme étant un ensemble de technologies, règles et arrangements institutionnels qui ensemble facilitent la création, l'échange et l'utilisation des données géospatiales, ainsi que les données associées au travers d'un espace communautaire dédié au partage de l'information.

C'est sur cette base qu'a été érigée en 1994 la première IDS : la National Spatial Data Infrastructure (NSDI) des Etats-Unis d'Amérique, objet de l'ordre exécutif n°12906 du Président Clinton [Clinton, 1994]. Cette définition a très peu changé depuis. Toutefois les technologies et les méthodes pour les mettre en œuvre ont quant à elles bien évolué.

Aujourd'hui, de nombreuses initiatives nationales, ou transfrontalières ont suivi ce modèle et proposent leur propre IDS afin d'améliorer l'accès aux données spatiales, de promouvoir leur échange et leur réutilisabilité. Des directives telles qu'INSPIRE en Europe incitent les organismes publics à utiliser ce type d'outil pour mettre leurs données à disposition du public. Couvrant des thématiques larges, bien que souvent axées sur l'environnement, les IDS possèdent néanmoins une structure générique.

Bien qu'il n'y ait pas de règles particulières sur les composants d'une IDS, les principales fonctionnalités offertes sont :

- une interface client permettant à minima l'affichage des données spatiales, l'envoi de requêtes de sélection et au mieux la réalisation d'analyses spatiales ;
- un service de catalogage permettant de rechercher des données via leurs métadonnées,
- des services de diffusion de données ;
- des services de traitement de données ;
- des bases de données pour le stockage des informations (géographiques ou non) ;

- une interface de supervision permettant de contrôler les outils et données mises à disposition dans l'IDS.

Dans les IDS existantes, on constate que les outils de traitements de données, qui plus est par le biais de services, sont encore très rares. La mise à disposition d'outils permettant la combinaison de données géographiques présente pourtant de nombreux intérêts. Elle peut par exemple permettre à l'utilisateur de définir l'emprise des données qu'il souhaite télécharger, plutôt que de tout récupérer.

Par ailleurs afin de permettre une interopérabilité des IDS entre elles, il est indispensable que les services proposés (notamment pour la diffusion des données) emploient des formats et protocoles standardisés. Les services, notamment web, sont majoritairement adoptés dans ce cas présent.

Le développement d'IDS est encouragé par de nombreuses initiatives. Nous nous intéressons aux plus importantes d'entre elles.

II.A.2 Initiatives pour la diffusion et le partage de l'information environnementale

On dénombre trois grandes initiatives européennes pour la diffusion et le partage de l'information environnementale :

- la première initiative est la directive INSPIRE que nous avons déjà évoqué précédemment ;
- la deuxième initiative est le Système d'Information Environnemental Partagé, traduction de l'anglais « Shared Environmental Information System (SEIS) » [Saarenmaa & al, 2002] ;
- la troisième initiative est le programme de Surveillance Globale pour l'Environnement et la Sécurité, traduction de l'anglais de « Global Monitoring for Environment and Security (GMES) » [Commission Européenne, 2001].

A ces initiatives européennes s'ajoute le projet de Système Mondial des Systèmes d'Observation de la Terre (Global Earth Observation System of Systems ou GEOSS) qui vise à proposer une plateforme d'échange et de partage de données à portée environnementale à l'échelle mondiale en s'appuyant sur les réseaux d'observation existants dans les pays [Khalsa, 2009].

Les quatre projets précédemment cités présentent de nombreuses similarités aussi bien dans leurs objectifs (le partage et l'échange de l'information environnementale) que dans leurs démarches (la définition et la promotion de standards d'échanges ainsi que la mise en place d'IDS).

Aussi, afin d'assurer la cohérence et l'interopérabilité des fruits de ces initiatives, un projet européen a été lancé en 2008 : il s'agit du projet GIGAS [Klien, 2009]. L'objectif est de s'assurer que les méthodes mises en place dans le cadre des quatre projets convergent et soient compatibles entre elles, tout en n'étant pas doublons.

Toutefois ces initiatives ne se limitent pas à la diffusion de données mais s'intéressent également au partage de modèles environnementaux.

II.A.3 Mise à disposition de modèles sur le web

Le principe de mise à disposition de modèles sur le web a été plus particulièrement abordé par Geller et Turner, qui en 2007 ont introduit la notion de « Model Web » [Geller & al, 2007].

Ils définissent le « Model Web » comme étant « un concept pour un réseau de modèles numériques qui, ensemble peuvent répondre à plus de questions que des modèles isolés ne le peuvent individuellement » [Geller & al, 2007].

Obtenir un modèle environnemental pertinent nécessite en effet un long travail de réflexion, de paramétrage et d'implémentation. Or, pour une meilleure compréhension d'un phénomène, l'étude de plusieurs modèles peut s'avérer nécessaire. Dès lors, l'utilisateur voulant compléter son étude se voit confronté à trois choix possibles :

- développer un nouveau modèle ;
- adapter voire améliorer le modèle dont il dispose ;
- améliorer l'interopérabilité et l'accès à d'autres modèles.

Le « Model Web » s'intéresse à cette troisième approche. Le web constitue une plateforme idéale pour le partage de ressources telles que des modèles. Un cadre et des méthodes doivent toutefois être définis afin de mettre à disposition des modèles pleinement opérationnels.

Le développement du « Model Web » constitue ainsi une sous-tâche du projet GEOSS dont l'objectif consiste à mettre en place une infrastructure dynamique de modélisation accessible à divers types d'utilisateurs (chercheurs, gestionnaires, décisionnaires et grand public).

Le « Model Web » prévoit la possibilité d'interconnecter des modèles et de les faire travailler de manière conjointe. Plusieurs projets de recherches s'intéressent à cette approche, parmi lesquels on peut citer le projet ENVISION, qui a pour but de proposer une infrastructure de services environnementaux avec ontologies (<http://www.envision-project.eu/>), ainsi que le projet UncertWeb qui s'intéresse à la mise à disposition sur internet de modèles prenant en compte l'incertitude (<http://www.uncertweb.org/>).

Afin d'assurer cette possibilité de connexion inter-modèles, l'un des pré-requis est d'utiliser des standards de diffusion de données. L'une des actions proposées est de mettre à disposition les modèles sous la forme de services web.

II.B Services et Architectures Orientées services

II.B.1 Notions de services et d'Architectures Orientée Services

Les services consistent en des fonctions mises à disposition par un composant logiciel pour assurer une tâche particulière. Ce sont des composants autonomes, qui ne dépendent d'aucun contexte ou service externe.

Une Architecture Orientée Service (AOS), traduction du terme anglais Service Oriented Architecture (SOA), désigne une architecture basée sur l'interaction entre des services. Les caractéristiques principales d'une architecture orientée services, sont l'utilisation d'un langage d'échange commun pour les services. Les plus utilisés sont le langage XML, ainsi que ses dérivés (GML, etc...).

Un service web correspond à un composant logiciel mis à disposition sur internet. De même on parle d'Architecture Orientée Web (AOW) pour désigner une AOS basée sur des services web. L'accès et l'invocation des services web se font donc en utilisant le web comme plateforme d'échange.

Techniquement on distingue deux grandes familles de services web :

- les services web basés sur une architecture de type REST (REpresentational State Transfer), où les fonctionnalités sont mises à disposition sous la forme d'un URI (Uniform Resource Identifier) accessible par le biais du protocole HTTP (HyperText Transfer Protocol) ;
- les services web de type WS-*, utilisant le protocole SOAP (Simple Object Access Protocol) pour l'échange de messages ainsi que le WSDL (Web Service Description Language) pour la description des services web.

Les services web de l'Open Geospatial Consortium (OGC) dont il est question ci-après, comportent des spécifications essentiellement basées sur une architecture de type REST, mais progressivement les deux protocoles sont explicités dans les spécifications des nouvelles versions. Les actuels serveurs d'application les implémentant s'appuient majoritairement sur l'architecture REST.

II.B.2 Services web dédiés à la géomatique

La plupart des services web standards existants dédiés à la géomatique, et plus particulièrement à la diffusion de données géographiques sont le fruit de l'Open Geospatial Consortium (OGC).

L'OGC est une organisation internationale composée de différents partenaires ayant trait à la géomatique et plus généralement à l'information géographique. Elle comprend aujourd'hui plus de 400 membres, institutions, entreprises et organismes. Les principales missions assurées par l'OGC sont la définition de nouveaux standards, la maintenance de ces derniers, ainsi que la promotion de l'utilisation de standards ouverts, afin de garantir l'interopérabilité dans le domaine de la géomatique.

Les travaux de l'OGC ont ainsi mené à la spécification de plusieurs types de services web, ainsi que leur mode d'utilisation dont les principaux en relation avec l'objet de la thèse sont décrits ci-après.

Plusieurs de ces standards sont reconnus par l'organisme international de normalisation (ISO). C'est notamment le cas pour le Web Map Service (ISO 19128:2005) et le Web Feature Service (ISO 19142:2010).

II.B.2.a Le Web Map Service (WMS)

Le Web Map Service (WMS) [OGC, 2006] est à ce jour le service web de l'OGC le plus utilisé. Il permet de produire des cartes sous la forme d'images. Ces images sont principalement diffusées sous la forme de tuiles rasters (PNG, GIF ou JPG). Les images produites sont géoréférencées, ce qui permet leur superposition et de construire ainsi une carte. Certains serveurs de WMS permettent également la reprojection des données spatiales, c'est-à-dire la création d'images dans une projection autre que celle des données d'origine.

Le WMS a rapidement été adopté par la communauté géomatique car il permet de répondre simplement au besoin de publier des cartes, et plus généralement des informations géographiques sur le web. La personne mettant ses données à disposition via WMS contrôle le style de la couche, c'est-à-dire la représentation graphique qui en est faite. La carte est automatiquement mise à jour dès qu'une modification est effectuée sur les données. Il existe de nombreux serveurs de WMS [OSGEO, 2011]. Les plus célèbres et aussi les plus couramment utilisés sont MapServer (<http://mapserver.org/>) et GeoServer (<http://geoserver.org/>). La plupart des serveurs WMS accepte en entrée des données aux formats standards SIG (SHP, TAB) ou encore s'appuient sur des Systèmes de Gestion de Base de Données (SGBD) tels que PostGIS (<http://postgis.refractions.net/>) ou Oracle Spatial (<http://www.oracle.com/>).

II.B.2.b Le Web Feature Service (WFS) et le Web Coverage Service (WCS)

Le Web Feature Service (WFS) [OGC, 2005] et le Web Coverage Service (WCS) [OGC, 2009] sont également assez courants, bien que largement moins répandus que les WMS. Ces deux services web permettent de mettre à disposition les données elles-mêmes. Pour cette raison, le WFS et le WCS sont parfois assimilés à des services web de téléchargement.

Le WFS est dédié à la diffusion de données vecteurs tandis que le WCS est dédié à la diffusion des données rasters. Les données récupérées par le biais de WFS ou WCS peuvent être représentées dans une application cartographique. Toutefois le principal intérêt du WCS et du WFS par rapport au WMS, est que les données elles-mêmes sont mises à disposition (et pas seulement une image) et peuvent être éditées ou servir pour la réalisation d'une analyse spatiale. Le WFS-Transactionnel

(WFS-T) permet ainsi d'ajouter, de supprimer et d'éditer des entités dans la base de données directement depuis l'interface client.

II.B.2.c Le Sensor Observation Service (SOS)

Le Sensor Observation Service (SOS) [OGC, 2007] propose de récupérer des données issues de capteurs. Qu'il s'agisse de capteurs in-situ (par exemple des niveaux d'eau) ou dynamiques (par exemple des images satellites), les mesures issues de capteurs représentent une large part des données géospatiales d'aujourd'hui [Botts, 2007]. Il fallait donc un format pour diffuser de manière efficace et interopérable ce type de données. Le SOS propose ainsi une interface permettant d'interroger, filtrer et récupérer des observations (requête *GetObservation*), ainsi que des informations sur tous types de capteurs (requête *DescribeSensor*). Il s'agit donc d'un intermédiaire entre le capteur et le client. Le profil transactionnel du SOS, souvent abrégé SOS-T permet d'enregistrer des capteurs (requête *RegisterSensor*) ainsi que des mesures (requête *InsertObservation*).

II.B.2.d Le Web Processing Service (WPS)

Le Web Processing Service (WPS) [OGC, 2007 bis] sert à diffuser des traitements, processus et calculs. Il définit également comment un client peut demander l'exécution d'un traitement, ainsi que le mode de renvoi des résultats. Le WPS a été premièrement conçu pour travailler avec des données spatiales, mais peut également employer des données non spatialisées. En effet, « les spécifications n'imposent aucune restriction sur les formats de données en entrée en ou sortie » [Neun & al, 2012]. Ainsi les traitements réalisables sont très variés. Ils peuvent aller de la simple addition de deux valeurs numériques jusqu'à la réalisation d'une carte d'interpolation par krigeage à partir d'une couche de points.

L'Office Internationale de l'Eau (OIEAU) s'en sert par exemple sur le portail SANDRE pour déterminer à la demande les départements traversés par un cours d'eau (requête d'intersection entre la géométrie du cours d'eau et celles des départements). Le BRGM utilise également le WPS dans le projet DeepCity3D pour déterminer la forme des fondations des bâtiments d'une ville, ainsi que dans le projet ENVISION pour construire des cartes des risques de glissement de terrain.

Le WPS vient donc étendre les capacités des portails web cartographiques en proposant des fonctionnalités de traitements et d'analyses spatiales proches des SIG bureautiques. Tous ces services peuvent être utilisés de manière conjointe, et faire partie d'une chaîne de traitement. On parle alors de composition de services web.

II.B.3 Formats standards de données

Comme nous venons de le voir, le choix d'un service web de diffusion des données va principalement dépendre du type de données que nous avons en entrée. Le schéma ci-dessous résume les choix possibles (Figure II-1).

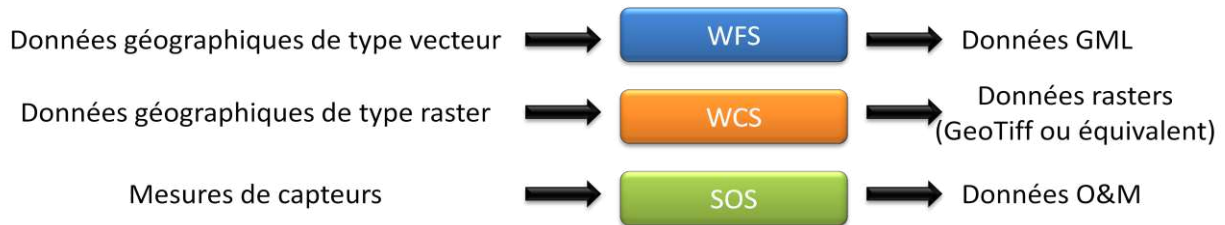


Figure II-1 : Services web de diffusion de données

L'intérêt des standards de diffusion de données est qu'ils mettent en forme les données selon des formats connus par le producteur et le consommateur du service. Pour le WCS plusieurs formats de rendus sont possibles, néanmoins pour le WFS et le SOS il s'agit respectivement du GML et de l'O&M. Nous nous intéressons plus particulièrement à ces deux formats.

II.B.3.a Geography Markup Language (GML)

Le Geography Markup Language (GML) est un standard de l'OGC [OGC, 2007] et également une norme ISO (19136:2007). Il s'agit d'un encodage basé sur le XML utilisé pour l'expression et la description d'entités géographiques. Le GML peut décrire une très grande diversité de données géographiques de une à quatre dimensions (3D + temps). Plusieurs schémas d'application s'appuient sur le GML pour décrire et échanger des informations dans des domaines spécifiques. Dans la sphère environnementale, on peut notamment citer :

- le CityGML, défini pour la modélisation d'une ville en trois dimensions. Ce schéma est d'ailleurs lui-même un standard OGC [OGC, 2008] ;
- le GeoSciML (GeoScience Markup Language), défini pour l'échange de données contenant des informations géologiques et plus généralement en rapport avec les sciences de la Terre. Il est d'ailleurs utilisé dans le cadre du projet OneGeology, visant à proposer une carte géologique du monde [Cowan, 2008] et constitue le modèle de référence pour le thème Géologie de la Directive INSPIRE [Serrano, 2006].

II.B.3.b Observations & Measurements (O&M)

Le format Observations & Measurements (O&M) est également un standard de l'OGC [O&M, 2010] et une norme ISO (19156:2011). Tout comme le GML, O&M est un encodage basé sur le XML. Il est toutefois spécifiquement dédié aux observations et mesures issues de capteurs. Le format O&M est indissociable du standard SOS. Il définit notamment des schémas XML qui spécifient le contenu des

réponses aux requêtes de type GetObservation du SOS. Tout comme le GML, des schémas d'applications ont été spécifiés pour une application dans des domaines particuliers. On peut notamment citer :

- le CSML (Climate Science Modelling Language) utilisé pour les données climatiques, atmosphériques et océanographiques [Lowe & al, 2011] ;
- le WaterML (Water Markup Language) pour la description des données hydrologiques. Ce schéma est également un standard OGC [OGC, 2010].

II.C Composition de services web

La composition de services web spécifie quels services sont invoqués, dans quel ordre et comment gérer les conditions d'exception. La norme du WPS ne préconise en aucun cas quel type d'orchestration utiliser [Cauchy, 2012]. Les architectes sont ainsi libres de chaîner les traitements comme bon leur semble. Dans la littérature, on distingue néanmoins deux grandes méthodes de composition de services web : l'orchestration et la chorégraphie de services.

II.C.1 Orchestration et chorégraphie

Dans le cas de l'orchestration de services web, on distingue un service maître. A l'image d'un chef d'orchestre, il coordonne l'intervention des autres services nécessaires à l'exécution du processus. Le service orchestrateur est le seul à connaître l'intégralité des traitements à effectuer (la composition à jouer). Les autres services se contentent de jouer leur propre partition, au moment où le service maître leur ordonne, sans se soucier de ce que font les autres processus. Le schéma suivant (Figure II-2) montre le rôle central joué par le service orchestrateur.

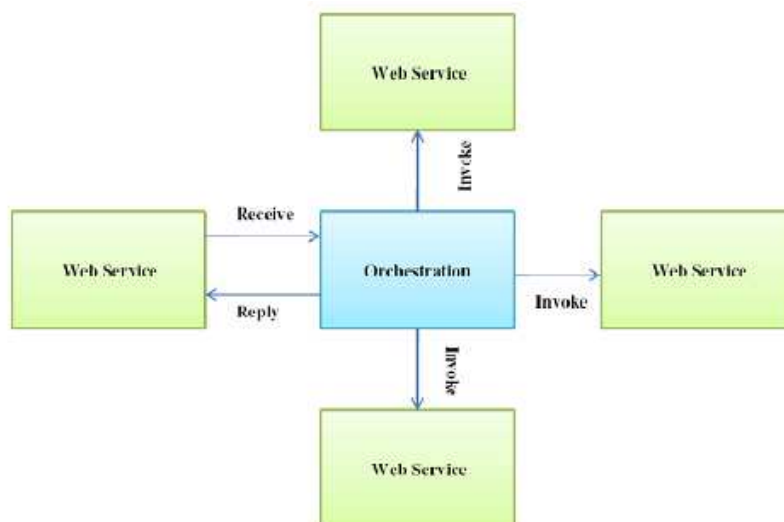


Figure II-2 : Schéma de principe de l'orchestration de services [Albreshne & al, 2009]

Dans le cas de la chorégraphie par contre, il n'y a pas de service maître. Le fonctionnement se veut plus collaboratif. Chacun des services est conscient de l'ensemble des processus effectués et sait avec

quels services il doit interagir et à quel moment il doit intervenir. Cela implique que les différents services communiquent entre eux, en informant par exemple quand ils ont terminés leur tâche. Le schéma suivant (Figure II-3) montre l'interdépendance des services web dans une chorégraphie.

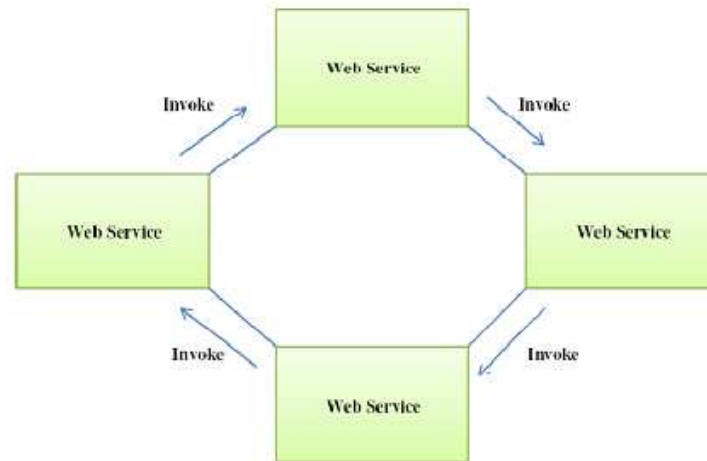


Figure II-3 : Schéma de principe de la chorégraphie de services [Albreshne & al, 2009]

II.C.2 Comparatif orchestration / chorégraphie

Le principal intérêt de l'orchestration est que l'utilisateur n'a qu'un interlocuteur : le chef d'orchestre. La demande d'exécution et le résultat passent nécessairement par lui. Par ailleurs les autres services n'ont pas besoin d'être conscients de l'action des autres services. Ce qui facilite leur réutilisabilité dans d'autres processus. Cette structure rigoureuse rend les modifications plus faciles : si un processus est défaillant, on le remplace sans se préoccuper des autres [M. Juric, 2005].

L'orchestration nécessite toutefois le développement d'un processus complémentaire : celui du chef d'orchestre. Ce processus n'est pas productif : il ne traite pas directement les données, mais organise leur production par d'autres services. Dans le cadre de l'intervention d'un faible nombre de services, ce chef d'orchestre n'est pas nécessairement indispensable et donc une chorégraphie peut tout à fait convenir.

II.C.3 Outils pour l'orchestration

II.C.3.a Un standard : le BPEL

WS-BPEL est l'acronyme de Web Service - Business Process Execution Language, souvent abrégé BPEL. Comme son nom l'indique, il s'agit d'un service web basé sur une architecture SOAP.

Le BPEL est un langage dérivé du XML, qui sert spécifiquement à définir et gérer l'orchestration de processus. Le BPEL est aujourd'hui un web service standard reconnu par l'Organisation pour

l'Avancement des Standards d'Information Structurés (OASIS) [OASIS, 2007], un consortium qui comme l'OGC s'intéresse au développement de standards. Il est soutenu par de grands acteurs de l'informatique comme IBM et Microsoft.

Le BPEL décrit deux éléments : les caractéristiques des processus, notamment les entrées et les sorties possibles, ainsi que l'ordre d'enchaînement des processus entre eux. Les implémentations du BPEL présentent généralement une interface graphique permettant de définir l'ordre ainsi que la façon dont vont s'enchaîner les divers traitements. Ce type d'interface facilite la génération du service BPEL.

II.C.3.b Une alternative : le WPS

L'utilisation du BPEL ne constitue pas l'unique moyen de réaliser l'orchestration de services web. Il existe une alternative : le WPS lui-même.

Sur le principe, l'utilisation du WPS est séduisante puisqu'elle évite l'intervention d'un service web tierce, et permet de se concentrer sur une architecture entièrement constituée de services web définis par l'OGC. Dans [Stollberg & al, 2007], les auteurs explorent les possibilités d'utiliser le standard de l'OGC pour l'orchestration de services web. Bien qu'il n'ait pas été spécialement défini pour ça à l'origine, le WPS possède les capacités pour se charger de cette tâche.

Un des reproches formulé par Stollberg quant à l'utilisation du BPEL est qu'il nécessite la définition de documents WSDL décrivant chacun des processus utilisés. La fabrication de tels WSDL ne va pas dans le sens de l'interopérabilité. En effet, elle est redondante par rapport à la fonction *DescribeProcess* définie par la spécification du WPS qui fournit déjà ces informations.

Stollberg définit trois façons de procéder à l'orchestration de services web par WPS :

- le chaînage de service centralisé qui consiste en l'invocation séparée des différents services web utiles par le service orchestrateur (Figure II-4) ;

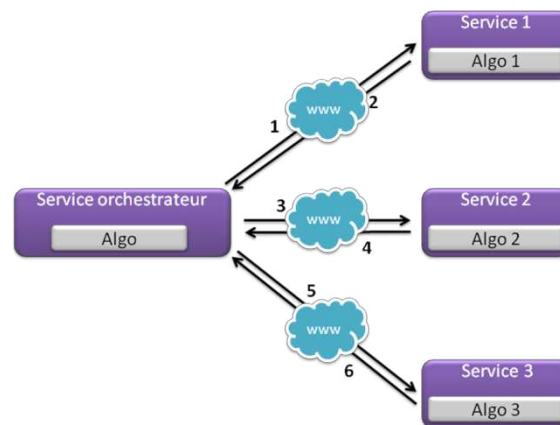


Figure II-4 : Schéma de principe du chaînage de service centralisé

- le chaînage en cascade où le service orchestrateur invoque un service web, qui lui-même invoque un autre service web et ainsi de suite (Figure II-5) ;

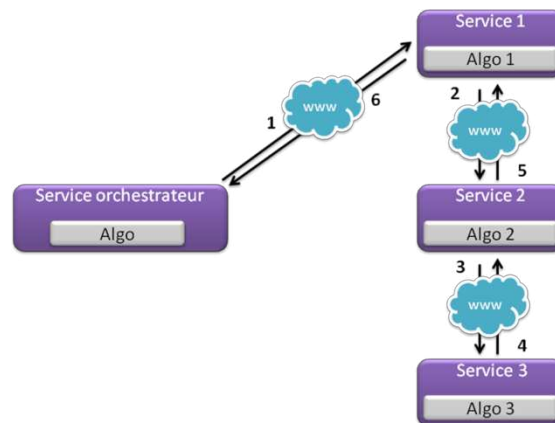


Figure II-5 : Schéma de principe du chaînage en cascade

- la constitution d'un unique WPS qui ferait directement appel aux classes et méthodes implémentés dans les autres services WPS (Figure II-6) ;

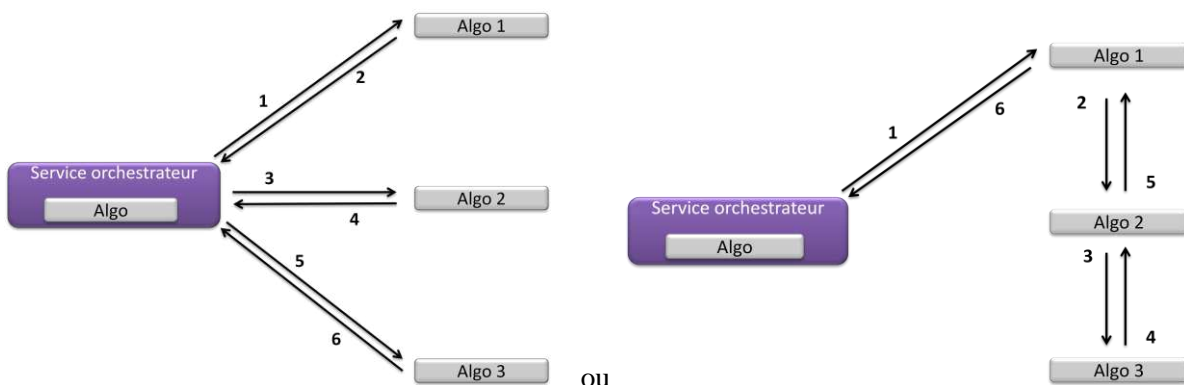


Figure II-6 : Schémas de principe de la méthode 3

Dans la troisième approche, l'invocation des services se fait sans passer par les interfaces offertes par les services web. Cette méthode est présentée comme permettant un important gain de temps puisqu'elle évite de nombreuses requêtes/réponses et de multiples échanges de données. La limite est qu'elle n'est possible que localement, c'est-à-dire si le processus orchestrateur et les autres processus utilisés sont implémentés au sein de la même machine.

II.D Conclusion du chapitre

Au cours de ce chapitre, nous avons étudié les IDS et leurs principaux constituants. La conception d'une IDS, notamment à portée environnementale, s'intègre dans le cadre de plusieurs projets européens (SEIS), voire mondiaux (GEOSS). Afin d'assurer l'interopérabilité des outils et données produits, l'utilisation de standards de diffusion est primordial. Ceux de l'OGC constituent une

référence et il semble indispensable de s'appuyer sur eux. Plusieurs d'entre eux retiennent particulièrement notre attention : les WFS, WCS et SOS qui sont adaptés à la diffusion de trois types de données différentes (respectivement des vecteurs, rasters et mesures de capteurs), ainsi que le WPS qui permet la diffusion de traitements. Nous envisageons ainsi d'utiliser ces standards dans notre IDS.

Il reste alors à organiser la composition des services. Pour cela, nous pensons nous appuyer sur une méthode d'orchestration, dont l'architecture avec un processus chef d'orchestre nous paraît à la fois plus rigoureuse, mais également plus facile à maintenir si des modifications devaient avoir lieu sur les processus.

D'un point de vue technique, le BPEL ne nous convient pas car son fonctionnement basé sur l'utilisation de SOAP et de WSDL, ne s'accorde pas parfaitement avec notre architecture de services web OGC utilisés en REST. La conception d'un WPS orchestrateur nous semble être une bonne solution, en particulier parce que nous utilisons déjà le WPS pour les autres algorithmes. C'est sans aucun doute vers cette solution, et en particulier vers le chaînage de service centralisé (Figure II-4) que nous allons nous tourner prochainement.

Néanmoins en première approche, nous n'avons pas souhaité nous lancer dans la conception d'un WPS orchestrateur et avons préféré dédier la tâche d'orchestration à la partie cliente. Ce choix ne nuit en aucun cas à l'interopérabilité des algorithmes et données qui restent accessibles par le biais de services web, et sont donc utilisables indépendamment de la partie cliente.

Chapitre III : Fusion de données et qualité de l'information géographique

Le fonctionnement des modèles repose sur le principe de la combinaison de plusieurs données, émanant généralement de différentes sources et présentant donc des hétérogénéités. La combinaison de telles données est appelée « fusion de données ». Au cours de ce chapitre, nous nous intéressons à l'étude de ce terme, de sa définition jusqu'à sa mise en œuvre. Nous nous intéressons également à la notion de qualité de l'information et à son impact dans les processus de fusion.

III.A Fusion de données

III.A.1 Terminologie

D'une manière générale, la fusion de données est une technique récente permettant le traitement d'informations issues de sources multiples. L'origine de la fusion de données est tout d'abord militaire et elle est à ce jour encore utilisée dans ce domaine.

La première définition reconnue de la fusion de données est celle fournie par le Joint Directors of Laboratories (JDL). Ce groupe de travail du Département de la Défense des Etats-Unis d'Amérique fondé en 1986 s'est intéressé à définir précisément la fusion de données.

Dans ce cadre militaire, la fusion de données est avant tout utilisée pour le repérage et l'identification de menaces potentielles, notamment par des techniques radars. La fusion de données est alors présentée comme étant « un processus à multiples facettes et à multiples niveaux, traitant de la détection automatique, de l'association, de la corrélation et la combinaison de données et informations provenant de sources multiples » [US DoD, 1991].

Cette définition a été légèrement revue en 1994, par l'Organisation Scientifique et Technologique pour la Défense des Etats-Unis d'Amérique (DSTO). Cette dernière souhaitait préciser le fait que la fusion de données pouvait également être réalisée depuis une source unique de données [DSTO, 1994].

De nos jours, les applications de la fusion de données ont largement dépassé le simple domaine militaire. La fusion de données est utilisée et trouve des applications dans les domaines de la robotique, du transport ou encore du traitement d'images.

Compte tenu de l'étendue des cadres possibles de la fusion de données, il est devenu nécessaire de présenter une définition globale et non fonctionnelle adaptée à tous les domaines dans lesquels la fusion de données peut avoir lieu.

Ainsi en 1998, Lucien Wald propose de définir la fusion de données comme étant : « un cadre formel dans lequel s'expriment les moyens et techniques permettant l'alliance des données provenant de sources diverses. Elle vise à l'obtention d'information de plus grande qualité ; la définition exacte de « plus grande qualité » dépendra de l'application » [Wald, 1998].

Le terme « alliance » contrairement à ceux employés dans les définitions précédemment citées, se veut volontairement plus générique. La fusion de données pouvant en effet porter sur toutes sortes de traitements. Cette remarque s'applique également aux données elles-mêmes, qui peuvent aussi bien être des mesures issues de capteurs, que des informations géographiques ou des signaux.

En 2009, l'OGC a lancé une étude sur les formats standards de fusion de données [OGC, 2009]. L'objectif annoncé de l'étude est de définir et de développer des standards de fusion de données. Il s'agit en fait de proposer un environnement de travail permettant à des utilisateurs d'exploiter des outils interopérables pour l'analyse, le traitement et l'exploitation de plusieurs types de données. Ces fonctionnalités (outils et données) doivent être accessibles depuis un client.

Au cours de cette étude, l'OGC a défini la fusion de données comme : « un acte ou traitement consistant en la combinaison ou l'association de données ou d'informations relatives à une ou plusieurs entités selon des méthodes issues de connaissances ou de savoir-faire, en vue d'améliorer une capacité (ou d'en produire une nouvelle) pour la détection, l'identification ou la caractérisation de cette entité » [OGC, 2010]. Cette définition a été construite à partir de celle du JDL ainsi que des propositions de membres de l'OGC ayant participé à l'étude.

III.A.2 Catégories de fusion

Dans son étude des standards pour la fusion de données, l'OGC propose également une catégorisation des processus de fusion en fonction du type de données utilisées en entrée. On distingue ainsi trois niveaux de fusion : fusion d'observations, fusion d'entités et fusion décisionnelle (Figure III-1).

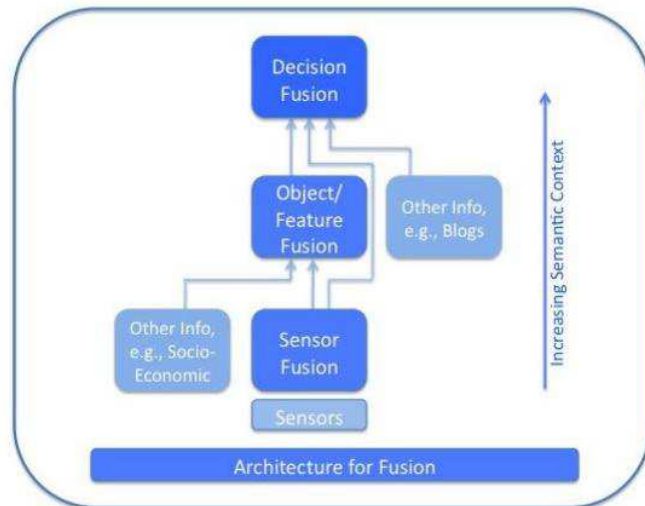


Figure III-1 : Niveaux de fusion de données [OGC, 2009]

III.A.2.a Fusion d'observations

Le premier niveau de fusion est appelé « fusion d'observations ». Il consiste en la combinaison de mesures provenant de plusieurs capteurs mais observant le même phénomène physique. Il peut s'agir par exemple de constituer une carte interpolée des températures à partir des mesures effectuées par différentes stations météo. Une telle carte permet d'avoir d'une estimation de la valeur de température dans un lieu où il n'y a pas de capteur.

Il peut également s'agir de combiner plusieurs grandeurs physiques s'intéressant au même phénomène. Le principe d'empilement ou « stacking » d'images prises dans différentes bandes spectrales peut ainsi être vu comme de la fusion d'observations. Les images satellitaires multi-spectrales SPOT (Figure III-2) ou LANDSAT utilisées pour la constitution de la base de données d'occupation du sol Corine Land Cover sont le fruit d'une fusion d'observations.



Figure III-2 : Image multispectrale SPOT XS prise sur Rotterdam (Pays-Bas).

(Source : <http://gdsc.nlr.nl/>)

III.A.2.b Fusion d'objets

Le second niveau de fusion est appelé « fusion d'objets ou d'entités ». Il propose de combiner des données et mesures associées à des phénomènes physiques différents. La fusion d'objets permet de produire des informations d'un niveau sémantique supérieur.

Les mesures d'évapotranspiration proposées dans les données SAFRAN de Météo France, par exemple, sont le fruit d'une fusion d'objets : celle de plusieurs données météorologiques (dont la température et le rayonnement) selon la formule dite de « Penmann-Monteith » [Chaouche, 2010]. De plus, les données SAFRAN sont également des données ayant subi une fusion d'observations pour être rattachées à des mailles de huit kilomètres de côté.

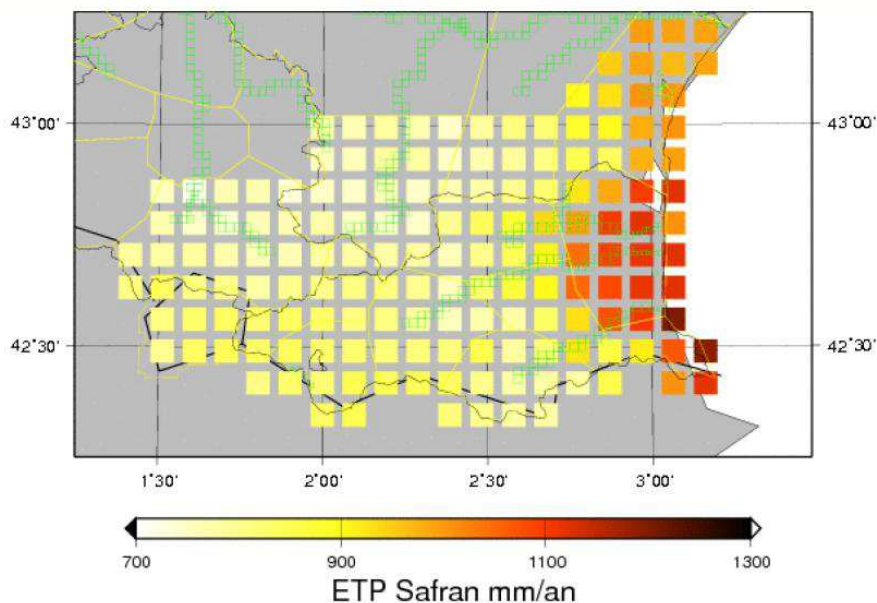


Figure III-3 : Mesures d'évapotranspiration moyennes sur la période 1970-2006 issues de SAFRAN pour le département des Pyrénées-Orientales [Caballero, 2008]

III.A.2.c Fusion décisionnelle

Le troisième niveau de fusion est appelé fusion décisionnelle. Le principe est de fournir des informations permettant à un utilisateur de prendre une décision. Les informations à produire doivent permettre de répondre à des questions que l'utilisateur se pose. Il peut s'agir par exemple de savoir si une parcelle est inondable. Les informations produites sont bien souvent concises, voire binaires : oui ou non.

Ainsi dans le cas de la fusion de données de troisième niveau, il faut prendre en compte un point essentiel : l'attente de l'utilisateur final. Une connaissance de ce même utilisateur est donc cruciale.

Dans nos modèles d'estimations des besoins en eau agricoles et domestiques, nous combinons des données associées à différents phénomènes. Pour reprendre la catégorisation de l'OGC, il s'agit plus particulièrement de fusion d'objets. De plus, les données que nous utilisons sont imprécises voire incomplètes. La connaissance des pratiques d'irrigation des agriculteurs est une information difficile à obtenir. De même nous ne savons pas exactement combien d'eau consomme chaque habitant. Nos modèles doivent ainsi s'appuyer sur des données imparfaites, et intégrer une certaine marge d'incertitude. L'étude de la prise en compte de la qualité de l'information géographique est donc essentielle.

III.B De la qualité de l'information géographique

L'utilité d'un modèle repose sur la précision et la fiabilité des résultats qu'il permet d'obtenir. Aussi il est nécessaire de s'assurer que le modèle est suffisamment représentatif du phénomène qu'il est censé décrire. La prise en compte de l'imperfection des données et des traitements est donc indispensable.

III.B.1 Notion de qualité de l'information géographique

La prise en compte de la qualité dans l'information géographique constitue un sujet d'actualité en matière de recherche. Deux conférences scientifiques internationales dédiées à ce sujet ont d'ailleurs récemment vu le jour :

- Accuracy qui a lieu toutes les années paires (<http://www.spatial-accuracy.org/Accuracy2012>) ;
- International Symposium on Spatial Data Quality (ISSDQ) qui a lieu tous les années impaires (<http://www.mat.uc.pt/issdq2011/>).

La volonté de prendre en compte la qualité des données suit une logique simple : conduire à une prise de décision éclairée, en avertissant l'utilisateur des limites des données dont il dispose [Evans, 1997].

L'ISO, définit dans sa norme ISO 9000, la qualité comme : « L'aptitude d'un ensemble de caractéristiques intrinsèques à satisfaire des exigences » [ISO, 2005].

Dans la littérature, on distingue deux définitions de la qualité :

- la qualité interne, qui désigne le niveau d'adéquation d'une donnée ou d'une mesure par rapport à une valeur de référence (généralement la valeur réelle ou physique) ;
- la qualité externe, qui porte sur l'étude de l'adéquation d'une donnée pour des usages précis.

Le terme anglais « fitness for use » est le plus souvent utilisé.

DeVillers résume la différence entre qualité interne et externe comme une différence de point de vue entre producteur et utilisateur de la donnée [DeVillers, 2004]. En géomatique, la qualité interne se place côté producteur de données, où l'on cherche à produire une donnée qui satisfait un cahier des charges et des spécifications prédéfinies. Une donnée d'excellente qualité est alors définie comme une

donnée satisfaisant ces spécifications. A contrario, la qualité externe se place plutôt du côté des utilisateurs. Une donnée de qualité sera ainsi définie comme satisfaisant les attentes des utilisateurs. Le schéma suivant (Figure III-4) montre les nuances entre les concepts de qualité interne et externe.

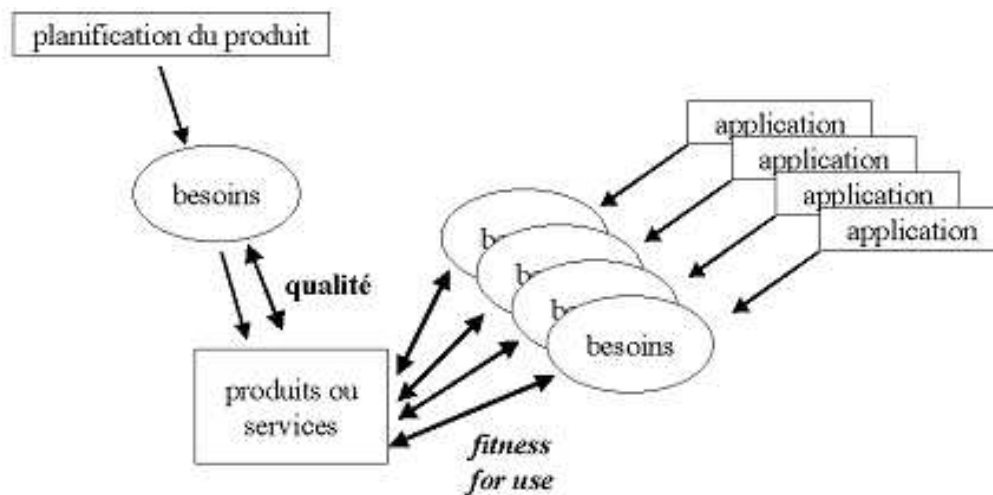


Figure III-4 : Représentation schématique des concepts de qualité interne et externe [Devilleers, 2004]

Selon la première approche, l'imperfection est un paramètre qui peut être quantifié. L'écart entre la donnée dont on dispose et la référence, est appelée « erreur ». Les travaux de recherche sur la qualité interne cherchent à connaître avec précision cette erreur, le but final étant bien entendu de la compenser. Les travaux portant sur la seconde approche sont moins nombreux que pour la première approche. Il faut dire aussi que la notion de qualité externe est assez subjective et encore récente, d'autant que la donnée peut être destinée à des usages multiples.

Les travaux de recherche sur la qualité de l'information géographique se décomposent principalement en cinq axes [Agumya & Hunter, 1997] :

1 – La modélisation de l'imperfection, qui vise à connaître de manière plus précise l'incertitude et notamment comment elle se comporte. L'analyse de sensibilité dont l'objectif est de déterminer quels paramètres ont le plus d'influence sur les résultats, est une des réponses à cette problématique.

2 – La propagation de l'imperfection, dont le principe est de déterminer l'imperfection résultant de la combinaison de données elles-mêmes incertaines. Cette entreprise se fait souvent en s'appuyant sur des approches déterministes ou probabilistes.

3 – La communication de l'imperfection, qui consiste à s'interroger sur les divers moyens de représenter le caractère imparfait des données et informations dans le but d'améliorer le processus de décision. Il peut aussi bien s'agir d'accélérer la lecture des informations ou la prise de décision, que de renforcer le degré de confiance associé à la décision prise.

4 – L'adéquation des données pour un usage donné (ou détermination de la qualité externe).

5 – La réduction de l'imperfection portant sur les données et informations, qui vise à augmenter le niveau de fiabilité des données. Cette action peut par exemple être réalisée par recoupement avec d'autres données issues de procédés d'acquisition différents.

Le choix d'une approche plutôt qu'une autre va dépendre du type d'imperfection auquel on a à faire, ainsi que de nos connaissances au sujet de ces mêmes imperfections. Nous nous intéressons donc aux différents travaux ayant menés à l'identification et la catégorisation des différentes imperfections.

III.B.1.a Typologie des imperfections de l'information géographique

Les imperfections susceptibles d'affecter l'information géographique peuvent être de plusieurs types. De nombreux travaux de recherche se sont attachés à définir une terminologie des imperfections pouvant affecter les informations géographiques et de situer chacun des termes employés par rapport aux autres. L'intérêt d'une typologie est de clarifier les termes employés (ex : incertitude, erreur, imprécision, exactitude) pour désigner les diverses imperfections pouvant qualifier une donnée. Le but est de proposer un vocabulaire universel précis dans lequel chaque type d'imperfection serait désigné par un seul et unique terme.

Nous nous intéressons ici aux typologies de Fisher et de DeRunz, définies dans [Fisher, 2005] et [DeRunz, 2007]. La typologie de DeRunz s'inspire largement de celle de Fisher. Dans sa typologie, DeRunz préfère le terme d'« imperfection » pour désigner le sommet de sa taxonomie. Il réserve le terme « incertitude » à l'étude des différentes erreurs susceptibles d'affecter un objet bien défini. DeRunz intègre également l'incomplétude comme source d'imperfection. Le schéma ci-dessous (Figure III-5) présente les similitudes et différences entre les taxonomies de Fisher et de DeRunz.

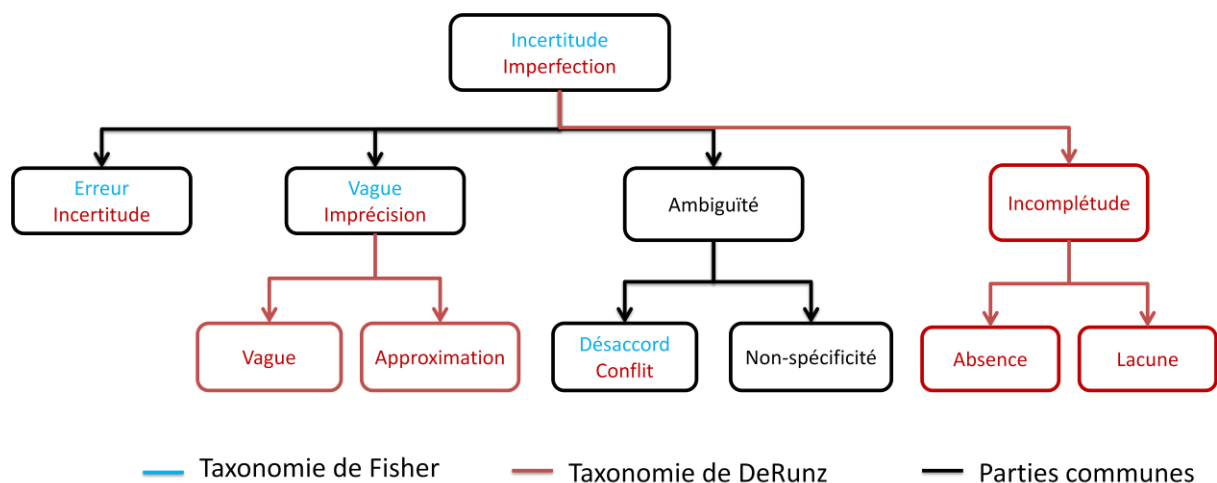


Figure III-5 : Taxonomies de Fisher et DeRunz

Dans la typologie de DeRunz, nous distinguons les termes suivants :

- l'incertitude correspond au degré de conformité de la valeur indiquée par rapport à la valeur réelle. Les biais relevant de l'incertitude sont appelés erreurs, et peuvent être estimés en se basant notamment sur des théories probabilistes.
- l'imprécision désigne une information dont la valeur est généralement fournie par un qualificatif choisi par l'utilisateur (par exemple : peu, beaucoup, etc...) et dont l'interprétation dépend en partie de son point de vue. Contrairement à Fisher, DeRunz distingue deux cas :
 - o Vague : une indication peut être interprétée d'une quantité infinie de manière, avec toutefois un degré de plausibilité décroissant (ex : A est proche de B).
 - o Approximation : les limites des catégories sont mal définies. On peut toutefois définir une limite haute et une limite basse pour identifier la catégorie (ex : A est situé en banlieue de Paris).
- l'ambiguïté, selon DeRunz, naît d'un mélange d'incertitude et d'imprécision. Il s'agit en fait d'une mauvaise classification d'un objet (donc une erreur) du fait que les catégories proposées n'étaient pas clairement définies (donc une imprécision). Fisher et DeRunz distinguent deux cas :
 - o Conflit : les libellés de plusieurs classes sont ambigus et font qu'un objet pourrait légitimement appartenir aux deux classes. C'est principalement le cas dès lors qu'on a des classes très généralistes. Dans Corine Land Cover par exemple, une parcelle recouverte de pommiers peut légitimement être rangée dans la classe 222 : Vergers, ou 242 : Systèmes cultureux et parcellaires complexes.
 - o Non-spécificité : une même indication peut être interprétée de multiples manières, et n'entre dans aucune catégorie précise. Cette notion est assez difficile à appréhender. Elle survient généralement lorsqu'une personne souhaite utiliser des données à une échelle (sémantique, spatiale ou temporelle) plus fine que celle pour laquelle elle a été créée. Par exemple, dans Corine Land Cover, le libellé de la classe 222 : Vergers, ne permet pas d'identifier précisément la culture (pêcher, abricotier, pommier, ...).
- l'incomplétude correspond au manque d'informations fournies par une source de données (par exemple : pas de mesures pour un lieu précis, ou sur une période temporelle donnée). DeRunz distingue de nouveau deux cas :
 - o Absence : La donnée est tout simplement manquante.
 - o Lacune : Une partie de l'information est manquante.

[Fisher, 2005] propose également une typologie des erreurs (ou des incertitudes pour reprendre la terminologie de DeRunz), susceptibles d'affecter une base de données géographiques. DeRunz l'adapte à la gestion d'une base de données archéologiques où la composante temporelle est au moins

aussi importante que la composante géographique. Le tableau ci-dessous (Tableau III-1) présente les différents types d'erreurs possibles selon Fisher et DeRunz.

Types d'erreurs	Causes de l'erreur
Mesure	Elle peut survenir lors de l'acquisition de l'information par des appareils de mesure (ex : positionnement d'un point par GNSS).
Classement	Suite à une erreur d'analyse, l'objet n'est pas rangé dans la bonne classe.
Regroupement de classes	La classification n'est pas adéquate et regroupe des entités possédant des différences significatives.
Généralisation spatiale	L'objet a été modifié lors de sa représentation cartographique (simplification de la géométrie, décalage).
Enregistrement	Elle peut survenir suite au transfert de la donnée depuis l'appareil de mesure vers le SGBD (ou le SIG directement).
Traitement	Elle peut survenir suite à un calcul.
Temporelle	Les données disponibles sont trop anciennes (ou trop récentes) et ne traduisent pas la réalité du terrain à l'instant souhaité.
Mesure temporelle	La donnée n'est pas associée à la bonne date.
Généralisation temporelle	La granularité de la donnée n'est pas suffisante pour le besoin exprimé (ex : entité datée au siècle près).

Légende : En bleu : Typologie de Fisher

En rouge : Typologie de DeRunz

En noir : Parties communes

Tableau III-1 : Typologie des erreurs de Fisher et de DeRunz

Maintenant que nous avons identifié les différents types d'imperfections pouvant affecter nos données, nous nous intéressons à la façon dont ces dernières peuvent affecter les résultats de nos traitements.

III.B.2 Prise en compte des imperfections

Nous nous intéressons ici à la production à la demande d'informations relatives à la qualité des résultats produits. Pour cela nous étudions deux méthodes simples de prise en compte de l'incertitude : l'analyse de sensibilité, ainsi que la propagation des incertitudes.

III.B.2.a Analyse de sensibilité

Le principe de l'analyse de sensibilité est de montrer comment se comporte le modèle face à des variations effectuées sur un ou plusieurs paramètres. Le schéma suivant (Figure III-6) présente le principe de sensibilité d'un modèle à deux variables d'entrée. Un paramètre est dit sensible à une variable lorsqu'une faible variation en entrée va engendrer une variation importante de la valeur en sortie (ici la variable B). Au contraire, un paramètre sera dit peu sensible à une variable si une faible variation sur la valeur en entrée ne va entraîner qu'une faible variation du résultat (ici la variable A).

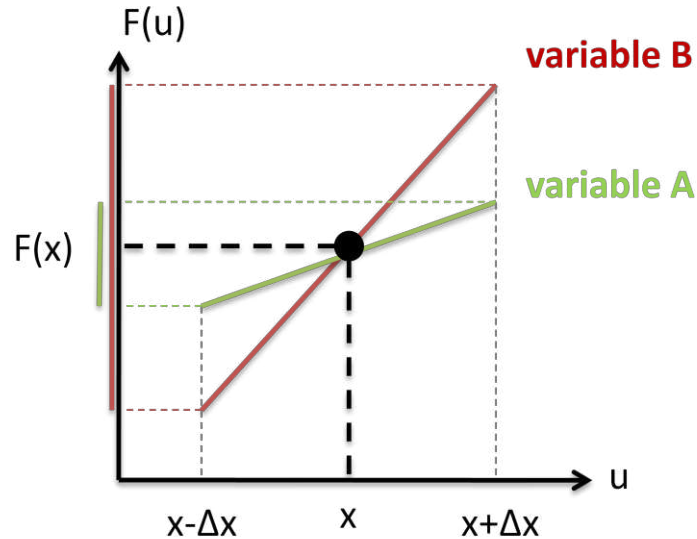


Figure III-6 : Principe de sensibilité du modèle à deux variables

La définition d'une faible variation doit être appréciée par rapport à l'étendue des valeurs possibles. On pourra par exemple se fixer sur un pourcentage de la valeur elle-même. Par exemple si une variable varie entre 20 et 30, et qu'on définit une variation comme faible dès lors qu'elle est inférieure à 5% de l'écart entre le maximum et le minimum. Alors une faible variation, correspondra à plus ou moins 0.5.

Techniquement l'appréciation de la sensibilité se fait par la détermination du « coefficient de sensibilité ». Il s'agit en fait de la dérivée partielle de la formule associée au calcul effectué par le modèle par rapport à la variable dont on souhaite étudier l'impact [GUM, 2008].

III.B.2.b Propagation des incertitudes

Le principe de propagation des imperfections vise à déterminer l'état des imperfections résultantes après divers traitements ou calculs. Ici nous nous intéressons uniquement à la propagation des incertitudes. On distingue deux grandes méthodes de propagation des incertitudes [GUM, 2008] :

- la première dite approche déterministe est analytique. Elle consiste à déterminer ou tout du moins approcher la fonction mathématique qui transforme l'incertitude en entrée en celle obtenue en sortie après calcul ;
- la seconde approche est probabiliste. Elle consiste à effectuer à plusieurs reprises les calculs avec un très grand nombre d'échantillons de données. L'incertitude résultante est déterminée par étude des résultats.

Approche déterministe

Une méthode simple pour déterminer la fonction f permettant de déterminer les incertitudes en sortie par rapport à celles en entrée, consiste à réaliser des approximations linéaires de la fonction f . A moins d'avoir une fonction parfaitement linéaire, cette approximation n'est valable que localement, c'est-à-dire seulement pour des valeurs des paramètres très proches de celles d'origine.

En pratique il s'agit d'effectuer une approximation en séries de Taylor d'ordre 1 de la fonction f . Cette méthode est recommandée par le Guide pour l'Expression des Incertitudes de Mesure (GUM), et est souvent appelée loi de propagation des incertitudes [GUM, 2008]. Lorsque plusieurs paramètres interviennent dans un calcul, l'approximation en séries de Taylor fera appel à des dérivées partielles qui traduisent l'évolution de chacune des composantes de l'incertitude. A noter que pour une approximation linéaire de la fonction, les coefficients directeurs de la fonction d'approximation sont égaux aux coefficients de sensibilité obtenus lors de l'analyse de sensibilité.

Le schéma ci-dessous (Figure III-7) présente la propagation de l'incertitude (assimilée comme la variance associée à une donnée) par le biais d'une approximation linéaire [Raphaël André, 2008].

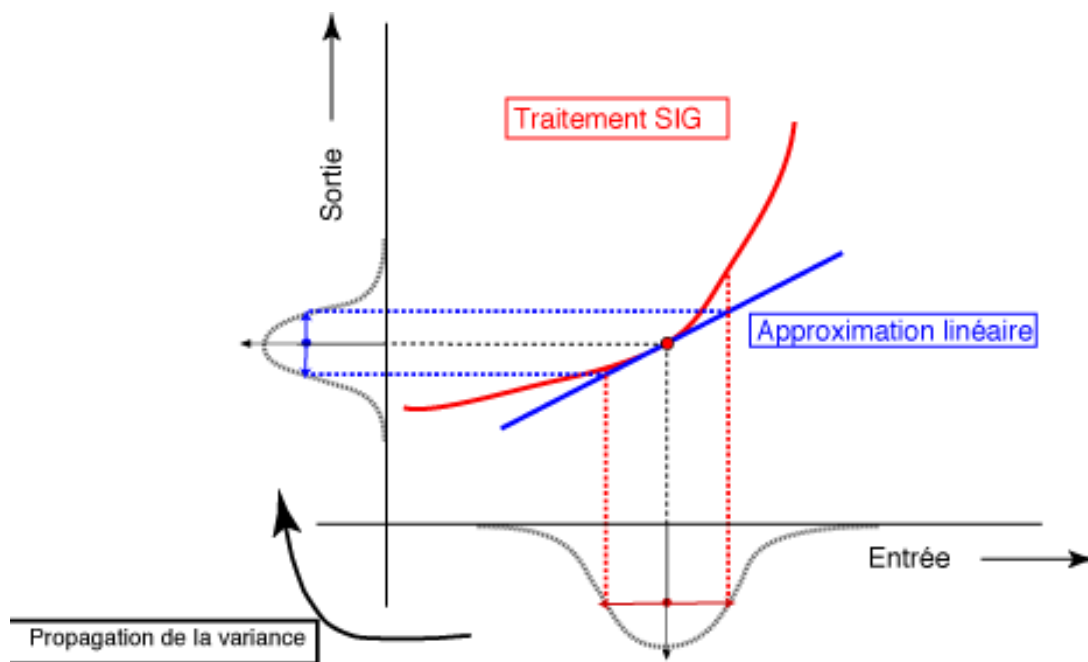


Figure III-7 : Propagation des incertitudes par le biais d'une approximation linéaire de la fonction f (en rouge) [André, 2008]

L'approche proposée par le GUM a pour avantage d'être inversible, c'est-à-dire qu'on peut déduire les incertitudes sur les données à partir de celles des résultats. Cette caractéristique peut permettre de mieux qualifier les incertitudes sur les données si on dispose de résultats vérifiés. L'utilisation de l'approximation par une série de Taylor est limitée par le fait qu'elle nécessite le calcul des différentielles. Pour obtenir ces dérivées, deux techniques sont possibles :

- la dérivation « à la main », qui requiert un effort humain conséquent pour le calcul de la dérivée, puis son implémentation pour pouvoir l'utiliser. Selon [Enciu, 2009], elle constitue la méthode la plus efficace. Néanmoins, la difficulté de calculer les dérivées augmente exponentiellement avec la complexité de la fonction ;
- la dérivation automatique, au moyen de logiciels spécialement dédiés comme ceux référencés dans [Bucker & al, 2000].

Approche probabiliste

L'approche probabiliste repose sur l'utilisation de simulations de Monte-Carlo. Il s'agit d'une méthode de simulation des résultats. Le principe de cette méthode est très simple.

Tout d'abord on constitue aléatoirement des échantillons de valeurs pour les données en entrée. Si l'on dispose des lois de distribution statistiques pour ces données, la répartition des valeurs la prendra en compte. Toutefois, si on ne dispose que d'un intervalle de confiance, les valeurs seront choisies avec une équiprobabilité dans cet intervalle. Les échantillons sont ensuite utilisés comme valeurs d'entrées dans le modèle qui fournira en retour des résultats. Ce processus est répété pour chacun des échantillons constitués. En pratique, il s'agit de relancer la simulation autant de fois que l'on dispose d'échantillons. Dès lors, on obtient un échantillon de valeurs possibles pour les résultats. Si le nombre d'échantillons est assez conséquent, on peut en déduire une loi de distribution statistique des résultats et donc l'incertitude résultante.

La méthode de Monte-Carlo permet ainsi de résoudre des problèmes complexes, dans lesquels des approches déterministes n'auraient pu être retenues. Un exemple bien connu est le problème dit de « l'aiguille de Buffon » pour l'estimation du nombre π [Aigner & al, 2006]. Néanmoins, le principal inconvénient de la méthode de Monte-Carlo réside dans le nombre de simulations à réaliser afin d'obtenir une approximation précise du résultat. Ce nombre croît en effet de manière conséquente avec le nombre de paramètres à prendre en compte au cours des simulations.

III.C Conclusion du chapitre

Au cours de ce chapitre, nous nous sommes intéressés à la fusion de données ainsi qu'à la qualité de l'information géographique. Plusieurs définitions de la fusion de données ont été données. Toutes convergent vers le fait qu'il s'agit d'un processus visant à produire une information nouvelle à partir de la combinaison de diverses données. Dans son étude des standards de fusion de données, l'OGC distingue trois niveaux de fusion en fonction du type de traitement réalisé : celles d'observations, celle d'objets et enfin la fusion décisionnelle. Comme nos modèles proposent de croiser des données issues de l'étude de différents phénomènes physiques, nous nous inscrivons plus particulièrement dans le cadre de la fusion d'objets.

Quel que soit le niveau de fusion abordé, il est nécessaire de prendre en compte les imperfections des données. Cette étape est indispensable si l'on souhaite s'assurer de la pertinence des résultats produits par le modèle. Nous nous sommes donc intéressés aux diverses imperfections pouvant affecter les données géographiques et avons fait le choix de nous focaliser sur les méthodes d'analyse de sensibilité et de propagation des incertitudes.

Il s'agit de deux méthodes simples de prise en compte de l'incertitude. La première vise à étudier l'impact des imperfections des données d'entrées sur celles en sorties. Les résultats d'une analyse de sensibilité peuvent aider à simplifier le modèle. Si on constate que le modèle est très peu sensible à une variable, ses variations pourront être négligées sans pour autant produire des résultats de moindre qualité. La seconde vise à quantifier l'incertitude associée au résultat final connaissant celles associées aux données d'entrées. Pour la propagation des incertitudes, nous nous sommes intéressés à l'approche probabiliste par simulations de Monte-Carlo, ainsi qu'à l'approche analytique d'approximation par séries de Taylor.

Les deux approches de propagation que nous avons étudiées n'abordent pas le concept d'incertitude de la même manière. Les résultats fournis par une approche par simulations de Monte-Carlo sont construits par propagation des fonctions de densité de probabilité des grandeurs d'entrée. Alors que les résultats fournis par une approche par approximation par série de Taylor s'appuient eux sur la propagation des incertitudes dans le modèle.

Les deux méthodes présentent des limites :

- celle de l'approche probabiliste, est qu'il est nécessaire de réaliser un nombre de simulations conséquent pour obtenir un résultat représentatif du phénomène réel ;
- celle de l'approche analytique réside dans notre capacité à dériver la formule associée au calcul à des ordres suffisants pour proposer une bonne approximation de la fonction.

La première limite peut-être repoussée par l'utilisation de méthodes d'échantillonnage « intelligent ». La seconde, par l'adoption de méthodes automatisées pour le calcul de dérivées. Ainsi, le choix d'une des méthodes plutôt qu'une autre va ainsi dépendre de la complexité du modèle que nous aurons à traiter. Pour des calculs simples, ou une approximation linéaire est pertinente, l'approche analytique est à privilégier. Pour les modèles plus complexes, la méthode des simulations de Monte-Carlo peut-être retenue, néanmoins compte tenu de la quantité de traitements qu'elle nécessite, cette méthode ne nous semble pas adaptée au contexte web qui vise à limiter la quantité d'échanges d'informations entre serveur et client. Dans nos travaux, nous privilégions donc l'utilisation de l'approche analytique pour la propagation des incertitudes.

Deuxième partie

-

Elaboration d'une IDS dédiée à l'évaluation des besoins en eau

Chapitre IV : Choix conceptuels et notions introduites dans FUSIEAU

Le projet FUSIEAU porte sur la conception d'une IDS dédiée à l'évaluation des besoins en eau agricoles et domestiques. Des outils visant les mêmes objectifs existent déjà. Nous présentons ici les spécificités de FUSIEAU que nous comparons à l'un de ses concurrents : STRATEAU.

IV.A FUSIEAU : frère jumeau de STRATEAU ?

IV.A.1 Des objectifs communs

STRATEAU est un outil d'aide à la décision réalisé par le bureau d'études Energies Demain suite à une commande en octobre 2009 de l'Ambassade de l'Eau, une association de type loi 1901 qui s'intéresse aux problématiques de gestion de l'eau dans les pays émergents du bassin méditerranéen.

Le projet était également soutenu par l'Office National de l'Eau et Milieux Aquatiques (ONEMA), les états français, italiens, libanais et marocains, ainsi que l'Union pour la Méditerranée, une organisation intergouvernementale visant à traiter diverses problématiques environnementales et socio-économiques affectant le bassin méditerranéen.

L'outil est mis gratuitement à disposition des pays méditerranéens. Une première version a été publiée au printemps 2010. Il est présenté comme étant destiné aux décideurs intervenant à des échelles différentes : de la commune au pays. Toutefois il est précisé qu'il s'adresse plus particulièrement aux « spécialistes de l'eau » [Ambassade de l'eau, 2012].

STRATEAU permet sur un territoire donné (unité de gestion, bassin versant) de réaliser un bilan hydrique en confrontant la ressource en eau disponible avec la demande en eau associée à divers usages. L'échelle temporelle utilisée dans l'outil est le mois et l'échelle spatiale est variable (de la commune au pays) [Prétot, 2010].

L'outil prend en compte six types de consommation :

- usage agricole (pour l'irrigation des cultures) ;
- usage énergétique (pour le refroidissement des centrales nucléaires et thermiques) ;
- usage industriel (pour les industries consommatrices d'eau) ;
- usage domestique (pour les consommations des habitants) ;

- usages administratifs et commerciaux (tourisme, administrations, commerces, ...);
- usages en milieu naturel (forêts, ...).

Le développement et la validation du logiciel s’est appuyé sur trois bassins versants situés sur le pourtour méditerranéen : Nahr El Kalb au Liban, Sebou au Maroc et la Têt en France.

STRATEAU permet de faire des scénarii en actionnant différents leviers comme les modifications des pratiques agricoles (irrigation, choix variétal), le climat, la démographie, les usages industriels, les ressources alternatives, la construction d’un barrage, etc., tout en incluant une demande environnementale. La capture d’écran ci-dessous (Figure IV-1) montre la représentation cartographique des données relatives aux prélèvements en eau.

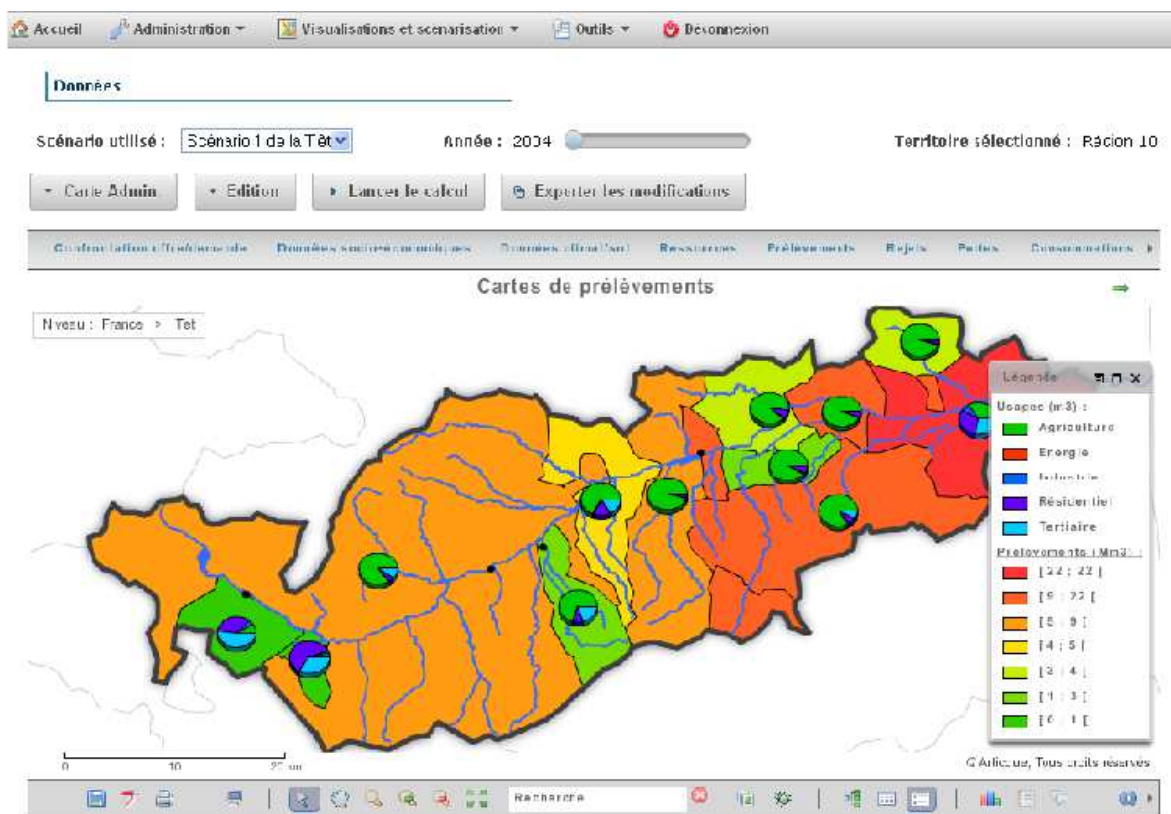


Figure IV-1 : Interface de STRATEAU pour la visualisation des données relatives aux prélèvements sur le bassin de la Têt. Source : [Ambassade de l’eau, 2009]

L’objectif est d’explorer la soutenabilité de différents scénarii de changement global et d’aménagement du territoire. STRATEAU est présenté comme une solution au service de la gestion quantitative de l’eau en France et plus largement dans le bassin méditerranéen dont les pays connaissent un stress hydrique fort et chronique.

IV.A.2 Des approches différentes

Bien que STRATEAU et FUSIEAU aient des objectifs communs, les outils qui en découlent diffèrent dans leur conception, en particulier sur deux points : la façon d'utiliser les données spatiales, ainsi que l'interopérabilité des modèles et des résultats.

STRATEAU n'est pas une IDS, il ne s'agit pas non plus d'une application de type SIG. Il s'agit plutôt de la mise à disposition sous forme informatique de modèles, tout comme CropWat est la version informatique du modèle FAO. Même si l'outil propose de produire des cartes : il ne s'agit ni plus ni moins que d'une façon de représenter les données. De même STRATEAU ne gère pas vraiment l'aspect temporel et propose seulement une succession d'instantanés.

L'application n'exploite pas la position des points de prélèvements ou de consommation comme une donnée. Or compte tenu de l'existence de nombreuses données spatiales à portée environnementale, et des actuels mouvements visant à soutenir et accroître la diffusion de ces données (INSPIRE, GEOSS), il nous paraît judicieux de développer des modèles et des outils capables d'exploiter ce potentiel d'information.

FUSIEAU quant à lui est une application SIG, permettant de manipuler des données géographiques, auquel nous avons souhaité ajouter des fonctionnalités de traitement de données propres à l'évaluation des besoins en eau.

Les modèles utilisés dans FUSIEAU s'appuient sur l'utilisation de données géoréférencées en entrée. Elles possèdent une composante géométrique, voire également une composante temporelle. Cette propriété présente de nombreux avantages. La composante géométrique permet par exemple d'indiquer de manière non équivoque à quel territoire la donnée est rattachée. Cela est particulièrement important dès lors que l'on travaille à des échelles telles que celles d'un bassin versant. En effet, un même bassin versant pourra avoir plusieurs représentations cartographiques selon l'outil et les paramètres qui auront été choisis pour les dessiner. Les écarts entre les tracés des bassins versants peuvent en effet être importants [El Hage, 2012]. Le fait de disposer des géométries évite toute ambiguïté sur l'étendue et les limites des zones étudiées.

De même, connaître la position des points de prélèvements ainsi que les lieux de consommation de l'eau prélevée facilite la réalisation du bilan hydrologique et permet d'affecter plus facilement un point de prélèvement à un ensemble de consommateurs. Le fait qu'une commune soit située dans un bassin versant, ne garantit pas forcément que ce soit ce dernier qui l'alimente. Seul le lieu du point de prélèvement permet de s'en assurer.

L'utilisation de données spatiales est également importante pour les changements d'échelle. Si pour les données administratives, l'imbrication est bien connue (commune > canton > département >

région) et peut éventuellement se faire de manière attributaire, pour les bassins versants la composante géographique prime car les différents organismes entendent parfois les limites des bassins versants de manières différentes.

IV.B Des données aux IDS

Pour pouvoir fonctionner, un modèle doit être alimenté par des données et des paramètres que l'utilisateur doit fournir. Même si les intitulés des données demandées sont clairs, il peut parfois être difficile pour l'utilisateur de les obtenir. L'utilisateur devra par exemple indiquer la superficie agricole dédiée à une culture X sur un territoire identifié. Or, comme il a pu être constaté lors du projet VULCAIN, cette donnée n'est pas forcément connue des gestionnaires et décideurs.

L'obtention de la donnée manquante et utile au modèle passera au minimum par une collecte de données et sa mise en forme pour correspondre au format réclamé par l'application, si tant est que la donnée nécessaire existe. Dans le cas contraire, il pourra être nécessaire d'effectuer une analyse voire un traitement de données ou une mise à l'échelle (spatiale ou temporelle). Cette étape de prétraitement peut toutefois être simplifiée par l'exploitation de données interopérables.

Les mouvements actuels de diffusion des données environnementales prônent l'utilisation de formats standards d'échange de données notamment sous la forme de services web. En considérant que les producteurs mettent à disposition les données sous la forme de formats standards tel que suggéré par des directives comme INSPIRE, la phase de mise en forme des données s'en retrouve considérablement simplifiée dans une application.

Dans FUSIEAU, nous nous sommes ainsi focalisés sur l'exploitation et l'intégration de données sous la forme de services web. Dans la même logique, nous avons souhaité que les données intermédiaires de notre modèle respectent également ces formats. Nous avons en effet anticipé que les résultats de nos modèles pourraient très bien devenir les données d'un autre modèle, et avons travaillé pour que leur future intégration en soit facilitée.

Toutefois, notre effort sur l'interopérabilité ne se limite pas aux formats d'échanges de données et résultats. Nous nous sommes également penchés sur les traitements et calculs réalisés au cours de notre simulation.

Dans de nombreuses applications, les modèles et l'outil les mettant à disposition ne font qu'un. De ce fait, une personne souhaitant exploiter un modèle devra forcément s'adapter à l'outil qui l'implémente, avec ses règles et contraintes. La conception de FUSIEAU s'inscrit, elle, dans une démarche de réutilisabilité des ressources. Nous avons ainsi souhaité proposer une application qui soit en accord

avec le « Model web ». C'est-à-dire qui permette de mettre à disposition des algorithmes de calcul, et plus important encore que leur utilisation soit indépendante d'une quelconque application cliente.

IV.B.1 Simulation de multiples IDS

L'intérêt de disposer d'une IDS utilisant des formats d'échange standards est renforcé lorsque les autres fournisseurs de données disposent eux-mêmes d'une IDS. Dans notre approche, nous formulons l'hypothèse que comme dans le projet GEOSS chacun des organismes détenteurs des données et informations dispose également de sa propre IDS, à partir de laquelle ils mettraient à disposition leurs données. Les informations seraient ainsi diffusées par le biais de formats standards interopérables tels que des services web OGC.

Pour l'évaluation des besoins en eau agricoles par exemple, nous avons donc simulé la présence de cinq IDS en plus de celle de FUSIEAU (Figure IV-2) :

- celle de Météo France pour la fourniture des données climatiques ;
- celle de l'INRA pour la fourniture des données agronomiques et pédologiques ;
- celle de l'Agence de l'Eau du bassin Rhône-Méditerranée Corse qui fournirait des informations relatives aux prélèvements en eaux ;
- celle de l'Agence de Services et de Paiement pour la fourniture des données relatives au Registre Parcellaire Graphique ;
- celle de l'Agence Environnementale Européenne qui fournirait les données Corine Land Cover relatives à l'occupation du sol.

Nous simulons également le fait que les données seraient mises à disposition sous la forme de services web standards de l'OGC (ce qui aujourd'hui n'est pas encore disponible).

Pour la découverte des données, chacune des IDS devrait disposer d'un service de catalogage permettant de rechercher et de récupérer des données via leurs métadonnées. De même, une interface client permettant de consulter les données est également fortement conseillée. Ces deux aspects ne sont pas traités dans nos travaux.

Nous nous sommes plus particulièrement concentrés sur la partie traitement et combinaison de données. Pour les IDS autres que FUSIEAU, seules les composantes de stockage et de mise à disposition de données ont été développées. Nous accédons ainsi directement aux données en interrogeant les serveurs les mettant à disposition. L'interface de FUSIEAU permet de visualiser et consulter les données issues des autres IDS.

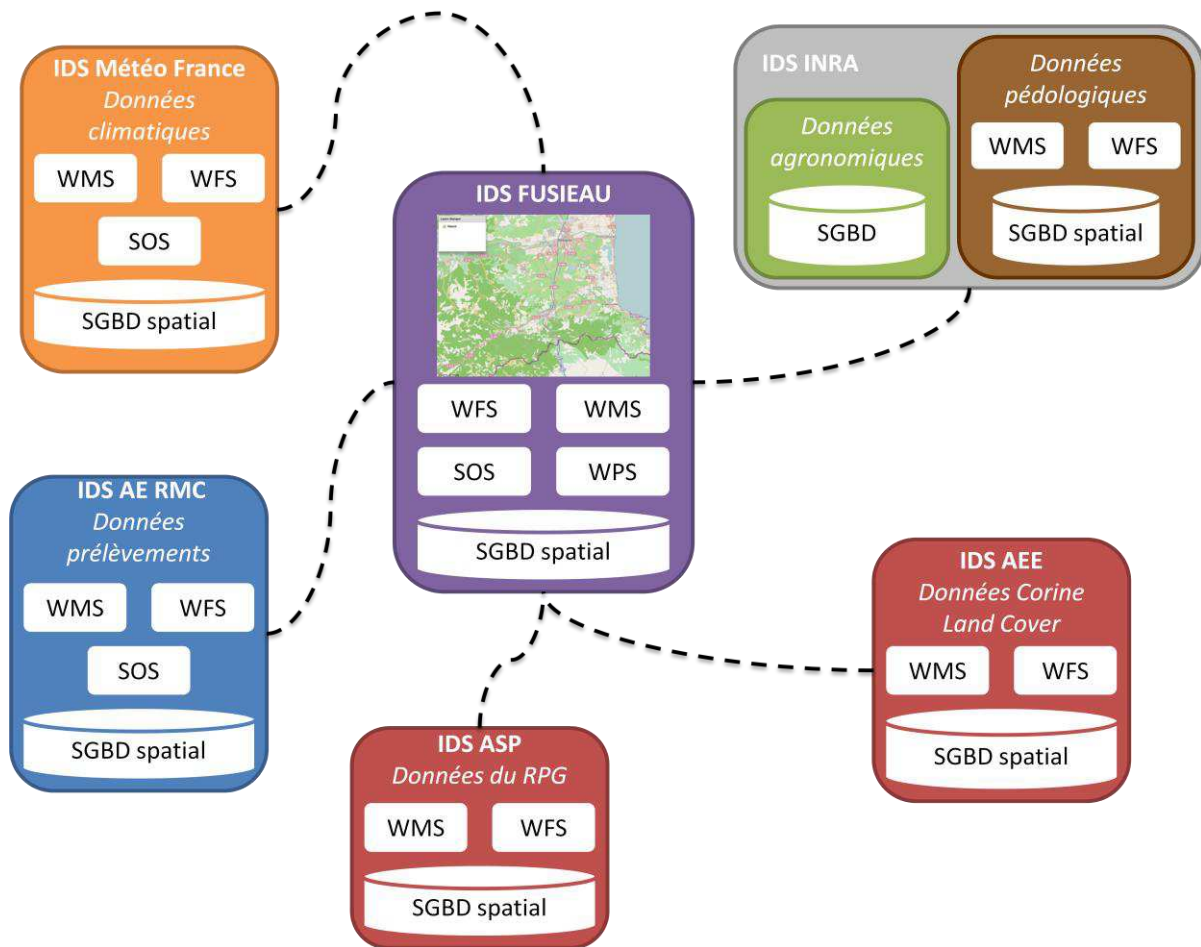


Figure IV-2 : IDS mises en place dans le cadre de FUSIEAU pour le modèle d'estimation des besoins en eau agricoles

IV.B.2 Choix des formats pour la diffusion des données

Compte tenu des caractéristiques des données utilisées dans le modèle FAO, nous avons eu recours à deux types de services web pour leur mise à disposition :

- pour les données qui correspondent à des mesures régulières et portant sur des entités immuables, nous avons retenu le standard SOS. Ce standard est en effet particulièrement adapté pour les séries temporelles. Nous utilisons ainsi le SOS pour la diffusion des données météorologiques, ainsi que celles relatives aux prélèvements ;
- pour les données dont nous disposons de peu d'échantillons temporels et qui présentent des variations dans leur géométrie, nous avons retenu le standard WFS. Nous l'utilisons ainsi pour la diffusion des données relatives à la pédologie, ainsi qu'à l'occupation du sol.

Le service WMS n'est pas utilisé comme donnée possible en entrée car il se contente de diffuser une représentation géographique des données et non les données elles-mêmes. Or c'est précisément de cela dont nous avons besoin pour pouvoir effectuer nos traitements de données.

IV.C Implémentation du modèle avec des services web

Notre approche basée sur l'utilisation exclusive de services, nous a tout naturellement mené à nous intéresser au WPS pour l'exécution des traitements sur les données. Pour aller au bout de la logique du « Model web », nous avons souhaité que les résultats des traitements WPS soit eux-mêmes diffusés sous la forme de services web.

Les services WPS développés permettent alors de produire des résultats soit sous la forme de SOS, soit sous la forme de WFS (Figure IV-3). Toutefois les services web ne constituent pas un système de stockage, mais seulement de diffusion des données. Les données produites doivent donc être préalablement stockées afin d'être diffusées par service web. Nous avons opté pour une sauvegarde des données dans un SGBD spatial. Le recours à un tel outil simplifie l'accès et la gestion des données, en centralisant tout en un espace unique.

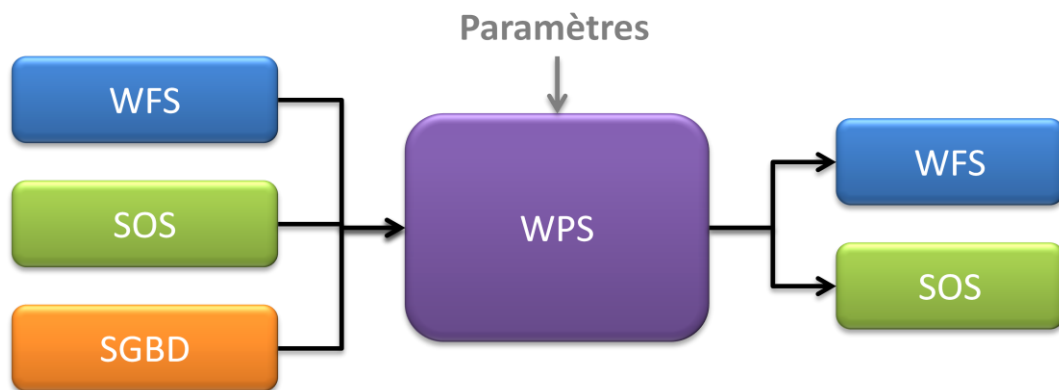


Figure IV-3 : Principe de fonctionnement général d'un algorithme WPS dans FUSIEAU

IV.C.1 Création de capteurs virtuels

Le principe de fonctionnement du modèle FAO s'appuie sur des calculs effectués à un pas de temps régulier (généralement journalier, décadaire ou mensuel). Les calculs sont effectués pour des cycles de culture d'une année. Compte tenu de la granularité des résultats, nous avons choisi de publier les résultats de ces traitements sous la forme de services web de type SOS.

Comme nous l'avons expliqué précédemment, le SOS est un standard particulièrement adapté à la diffusion de données issues de séries temporelles. Il a été pensé pour la mise à disposition des mesures effectuées par des capteurs. Nous avons repris cette approche capteur, ainsi que les fonctionnalités offertes par le standard SOS et l'avons adaptée à notre cas d'étude.

A l'image d'un capteur programmé pour prendre des mesures à des dates régulières, nos calculs sont également effectués pour des dates précises avec un pas de temps fixe. Aussi nous avons considéré

que les résultats de nos calculs effectués par le modèle FAO pourraient être assimilés à des mesures de capteurs.

La seule différence est qu'au lieu d'être le fruit de l'observation réelle de phénomènes environnementaux, les valeurs de nos mesures sont calculées selon le modèle FAO. Nous associons alors ces observations à des capteurs virtuels. Ces derniers présentent quasiment toutes les caractéristiques des capteurs classiques : ils effectuent des mesures, associées à un phénomène (physiquement observable ou provenant d'un modèle), ils s'intéressent à une zone d'étude (Feature Of Interest ou FOI en anglais).

La différence majeure est que les capteurs virtuels ne correspondent pas à une réalité physique. Ils ne sont pas donc pas associés à un dispositif de mesure qui serait placé dans un endroit précis. Le standard SensorML autorise la définition de capteurs associés à des mesures non-physique [OGC, 2007]. Le paramètre de localisation du capteur est ainsi optionnel.

Dans le cas où un capteur n'aurait pas de coordonnées géographiques associées, le lien entre les mesures et des entités géographiques peut quand même se faire par la géométrie de la zone d'étude couverte par les mesures. Sur l'interface utilisateur par exemple, un clic sur la maille SAFRAN (unité spatiale de base pour les données climatologiques) permet d'accéder aux mesures météorologiques. De même un clic, sur une parcelle permettra d'obtenir les données qui lui sont relatives.

Pour les capteurs associés à une même zone d'étude et se basant sur les mêmes hypothèses, nous utilisons des capteurs dits composites. Il s'agit en fait de capteurs observant diverses variables mais situés au même endroit. Une station météorologique mesurant la pluviométrie et la température peut ainsi être assimilée à un capteur composite. L'intérêt de cette approche est que cela permet d'affecter au même capteur tous les résultats d'un même scénario. Cet aspect est particulièrement important car compte tenu du grand nombre de paramètres pouvant être ajustés pour les simulations, plusieurs capteurs couvrant la même parcelle et la même période temporelle peuvent exister.

IV.C.2 Fonctionnement général d'un algorithme associé à un calcul issu d'un modèle

A l'échelle d'un algorithme, le fonctionnement s'apparente à une orchestration. Nous avons un service de type WPS qui assure l'interface avec l'utilisateur :

- l'utilisateur indique les paramètres et données à utiliser pour le traitement ;
- le WPS récupère (si nécessaire) les données sur les serveurs de diffusion de données par le biais de requêtes de type *GetFeature* pour un WFS, ou *GetObservation* pour un SOS, pour récupérer les données nécessaires ;

- le WPS se charge ensuite de combiner les données et de mettre en forme les résultats, puis de les renvoyer au client ou à l'utilisateur.

Nous décrivons ici le fonctionnement détaillé d'un algorithme WPS pour le calcul des valeurs d'évapotranspiration de la plante ETc. Le diagramme suivant (Figure IV-4) montre les différentes étapes réalisées lors de l'exécution du processus décrit plus haut.

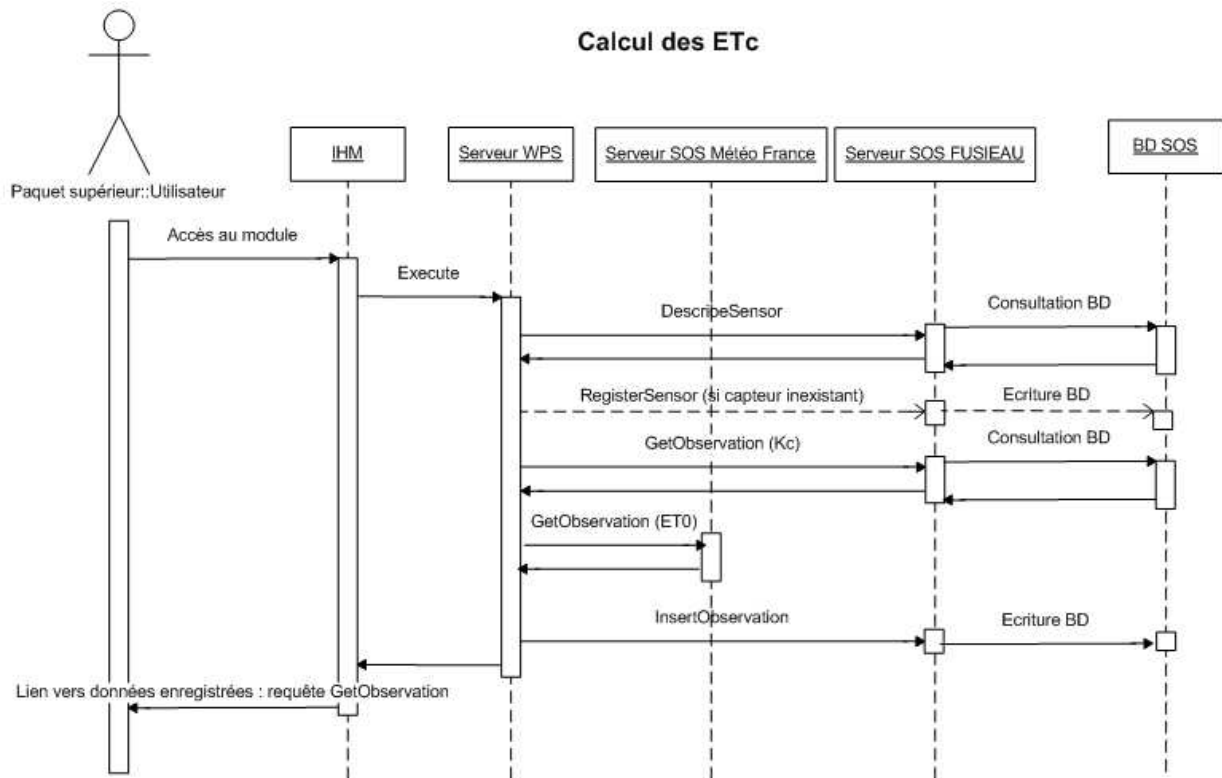


Figure IV-4 : Diagramme de séquence UML pour le calcul des ETc

Ces valeurs sont déterminées à partir de deux mesures : celle des coefficients cultureux (Kc), et celle de l'évapotranspiration de référence (ETo).

1 - L'utilisateur accède à l'application et plus précisément à la partie permettant de paramétrer une simulation par le biais de l'IHM. Il sélectionne l'algorithme (ici CalculETc) et précise les paramètres de sa simulation. Il lance alors le traitement, ce qui déclenche l'envoi d'une requête de type *Execute* au serveur WPS.

2 - L'application va ensuite envoyer une requête de type *DescribeSensor* pour vérifier si un capteur associé à la simulation existe (par comparaison des noms des capteurs). Si le capteur n'existe pas, on le crée via une requête de type *RegisterSensor*. Sinon on passe directement à l'étape 3.

3 – On récupère les données nécessaires au calcul. Comme il s’agit ici de données mises à disposition par le biais d’un standard SOS, on envoie donc deux requêtes de type *GetObservation* (une pour les mesures Kc, l’autre pour les mesures ETo).

4 – Les valeurs d’ETc sont calculées et on alimente le capteur associé (créé à l’étape 2 ou antérieurement) par le biais d’une requête de type *InsertObservation*.

5 – On renvoie à l’utilisateur l’URL permettant d’accéder aux données du capteur nouvellement alimenté.

IV.D Conclusion du chapitre

Dans ce chapitre, nous avons présenté le projet STRATEAU, dont les objectifs sont proches des nôtres : permettre à divers utilisateurs, notamment des experts hydrologues et décisionnaires, de réaliser des bilans ressources / demandes en eau à diverses échelles. Bien que l’outil STRATEAU soit mis à disposition sur le web, sa conception ne s’inscrit pas dans le cadre de l’interopérabilité, et plus particulièrement dans la philosophie du « Model web ». L’outil et les modèles implémentés forment un tout indissociable, de sorte que l’utilisateur souhaitant exploiter les modèles doit nécessairement passer par l’interface produite.

Dans le cadre du projet FUSIEAU, nous nous sommes focalisés sur l’exploitation et l’intégration des données existantes. Les modèles sont conçus de sorte à pouvoir fonctionner avec des données diffusées sous la forme de services web. Le respect de cette hypothèse s’inscrit dans les pré-requis des directives telles qu’INSPIRE pour la diffusion des informations géographiques à portée environnementale.

Nos modèles s’appuient principalement sur des données fournies par des organismes nationaux (MétéoFrance, Agence de Services et de Paiements, INRA) voire internationaux (Agence Environnementale Européenne). Cette caractéristique garantit la pérennité des données et assure la reproductibilité des modèles sur d’autres zones d’études du territoire français. Par ailleurs, les données employées sont régulièrement mises à jour par les organismes détenteurs. Ce service peut néanmoins être payant (cas des données Météo France).

En considérant que les organismes détenteurs de données continuent de les diffuser et de les mettre à jour sans en changer radicalement la structure, les modèles peuvent continuer d’exploiter les nouvelles données sans qu’il soit nécessaire de prévoir des développements complémentaires.

Chapitre V : Application à l'estimation des besoins en eau agricole et domestique

Pour les besoins de notre projet, nous avons travaillé à la fourniture de modèles dédiés à l'estimation des besoins en eau agricoles et domestiques. Si pour le premier nous sommes repartis sur un modèle existant, pour le second il s'agit de la conception d'un modèle inédit. Nous détaillons ici les travaux menés pour la conception de ces modèles.

V.A Spatialisation du modèle FAO

L'approche retenue dans VULCAIN pour la modélisation des besoins agricoles correspond à une simplification et une restriction du modèle FAO. Les concepteurs du projet ont fait des choix simplificateurs dans leur modélisation compte tenu des caractéristiques de leur cas d'étude. Ces restrictions font que le modèle n'est valable que sur la zone d'étude. Afin de proposer une modélisation plus générique, qui puisse être réutilisée plus largement, nous nous sommes intéressés au modèle FAO dans sa version complète.

Le modèle FAO s'appuie sur de nombreuses données environnementales : caractéristiques des cultures, du sol, du climat. De telles données existent sous la forme de données géoréférencées. De plus, les directives telles qu'INSPIRE incitent également à accroître leur diffusion. Aussi afin de proposer un outil dans la philosophie de GEOSS et celle du « Model web » nous avons entrepris de spatialiser le modèle FAO.

V.A.1 Approche du modèle

A l'échelle d'une parcelle l'estimation des besoins en eau n'est pas problématique, car les paramètres ont des valeurs assez proches d'un point à l'autre. Toutefois, dès lors qu'on s'intéresse à plusieurs parcelles la variabilité des paramètres peut être problématique [Er-Raki, 2007].

L'utilisation de données spatiales permet de rendre compte des caractéristiques d'un territoire, ainsi que de la variabilité spatiale des phénomènes étudiés : une carte pédologique montre les différentes natures du sol, une carte d'occupation du sol montre les différents usages de la terre, etc... De telles données existent et il semble judicieux d'utiliser le potentiel offert par ces sources de données dans notre modélisation.

Le modèle FAO est adapté au traitement de parcelles possédant des propriétés homogènes : une seule culture pratiquée, sur un type de sol, soumis aux mêmes conditions météorologiques. Afin d'estimer

les besoins en eau d'une zone d'étude présentant des propriétés hétérogènes, il est nécessaire de la découper préalablement en parcelles possédant des propriétés homogènes. Les besoins en eau de la zone d'étude se résumeront alors en la somme des besoins en eau des parcelles qui la composent.

Afin de réaliser ce découpage, notre approche consiste à fusionner les données associées aux couches des différentes thématiques abordées par le modèle. Cette méthode est classique des traitements SIG depuis [Burrough, 1986]. Le schéma suivant (Figure V-1) montre la détermination des terres agricoles les plus adaptées en fonction de divers critères.

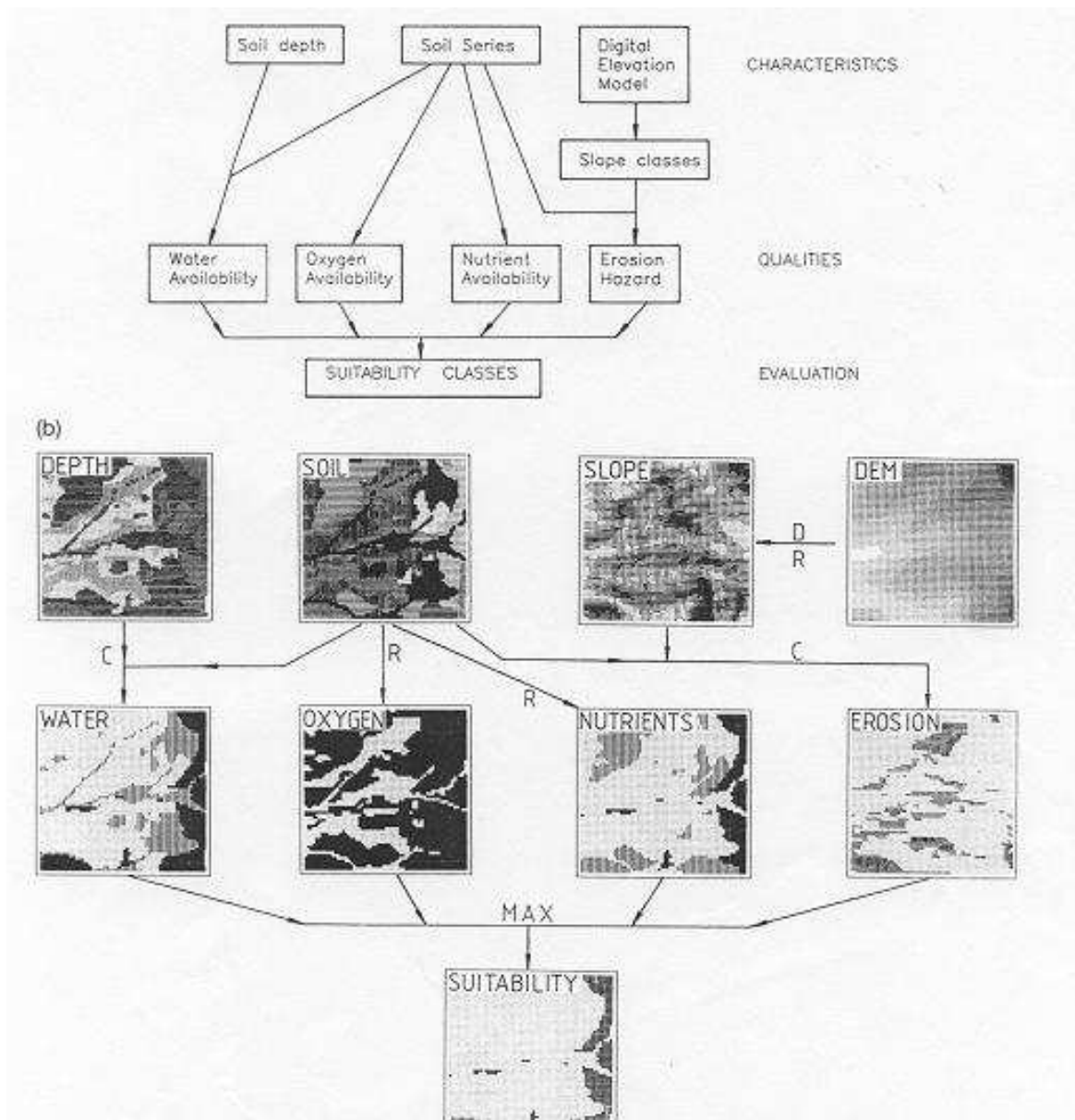


Figure V-1 : Combinaison de données thématiques géoréférencées [Burrough, 1986]

[Burrough, 1986] utilisait des données raster tandis que notre approche s'appuie sur des données de type vecteur. De ce fait, la superposition des différentes géométries constituant chacune des couches n'est pas parfaite.

Afin de pouvoir combiner les caractéristiques des données entre elles, nous devons préalablement combiner les géométries des couches. Le croisement de ces données forme un nouveau découpage dans lequel les propriétés ou caractéristiques pédologiques, météorologiques et phénologiques sont uniques et homogènes (Figure V-2). Pour reprendre la catégorisation des fusions décrite au Chapitre II, ce traitement s'inscrit dans le cadre de la fusion d'objets.

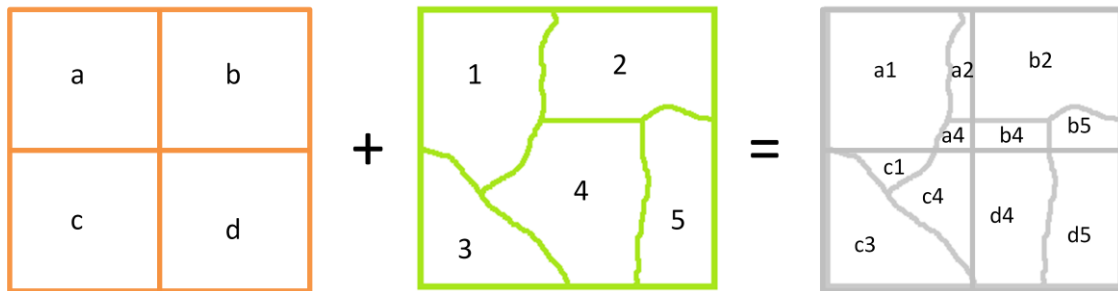


Figure V-2 : Principe de croisement de deux données thématiques

Afin de pouvoir réaliser le découpage, l'utilisateur sélectionne les couches associées aux trois thématiques décrites précédemment. Il indique ensuite le nom de la couche qui sera générée suite à la combinaison des données. On s'attachera ensuite à identifier les besoins en eau de chacune des parcelles créées.

V.A.2 Identification des données utilisées

Pour spatialiser le modèle, il est indispensable de disposer de données qui rendent compte des caractéristiques de chacune des parcelles. Il est ainsi notamment nécessaire de connaître :

- la nature des cultures pratiquées et leurs caractéristiques ;
- le type de climat auquel est soumise la culture ;
- la nature des sols et leurs caractéristiques (en particulier sur les propriétés hydriques) ;
- des informations (si elles existent) sur les méthodes d'irrigation.

Une des données majeures pour notre modèle est celle de l'occupation du sol. Il s'agit de la couverture biophysique de la surface des terres émergées [FAO, 1998]. L'occupation du sol ne doit pas être confondue avec l'usage du sol, qui comme son nom l'indique traite de la destination des sols. Ainsi un terrain de football pourrait être caractérisé par l'occupation du sol « herbe », mais par l'usage « équipement sportif ». La différence est toutefois moins significative dans le cas des parcelles agricoles où la couverture traduit généralement l'usage.

Afin d'identifier les données les plus pertinentes pour notre modèle, nous nous sommes intéressés à l'étude et à la comparaison de trois sources de données :

- la BD Corine Land Cover (CLC) de l'Agence Environnementale Européenne ;

- la BD OCSOL mise à disposition par l'Association régionale SIG Languedoc Roussillon ;
- le Registre Parcellaire Graphique (RPG) mis à disposition par l'Agence de Services et de Paiement (ASP).

Le tableau suivant (Tableau V-1) synthétise les principales caractéristiques des données :

Nom	Origine	Echelle spatiale équivalente	PPUC*	Couverture	Périodicité de mise à jour
CLC	Photo-interprétation (images SPOT + Landsat)	1/100'000	25 ha	Européenne	Trois réalisations : 1990, 2000, 2006 (prochaine en 2012)
OCSOL	Photo-interprétation (images Landsat)	1/50'000	2.5 ha	Région LR + PACA	Deux réalisations : 2000 et 2006
RPG	Déclaration sur BD Ortho IGN	1/5'000	Parcelle	France	Annuelle depuis 2006

* : PPUC = Plus petite unité cartographiée

Tableau V-1 : Caractéristiques des données d'occupation du sol étudiées.

Les deux premières données ont été produites par photo-interprétation d'images satellite (Landsat ou SPOT selon l'année de la campagne d'acquisition pour CLC, et Landsat pour OCSOL). Par ailleurs, si CLC possède une couverture européenne, celle d'OCSOL correspond à une commande de la région Languedoc Roussillon et est donc limitée au territoire de la région. Toutefois, une commande similaire et utilisant les mêmes procédés d'acquisition et d'interprétation a été précédemment effectuée par la région Provence Alpes Cote d'Azur. La donnée OCSOL couvre ainsi les territoires des régions Languedoc Roussillon et PACA. La principale motivation des décideurs pour la constitution de la BD OCSOL était d'avoir une délimitation plus fine du bâti diffus.

Les données du Registre Parcellaire Graphique ont une origine toute autre. Cette base de données regroupe les déclarations faites par les agriculteurs pour bénéficier des aides de la Politique Agricole Commune. Les cultivateurs étant invités à dessiner les contours de leurs parcelles sur des orthophotographies et d'y indiquer les cultures pratiquées. Toutefois, compte tenu de son mode d'acquisition, basée sur le volontariat (même si la possibilité d'obtenir des subventions est quelque peu motivante) cette base de données souffre d'incomplétude, mais également d'autres imperfections liées au procédé de saisie manuelle, en particulier des approximations dans les contours des parcelles (qu'elles soient volontaires ou non).

Pour compenser l'incomplétude des données du RPG, nous souhaitons utiliser une autre couche de données : CLC ou OCSOL. A défaut de nous indiquer précisément la culture pratiquée, elle nous donnera au moins une information sur le type de culture. La donnée utilisée en complément du RPG ne doit pas montrer de désaccord avec ce dernier.

Pour vérifier la correspondance des données d'occupation du sol, nous avons établi des matrices de confusion afin d'identifier dans quelle catégorie une parcelle clairement identifiée dans le RPG se retrouve classée dans les données OCSOL et CLC.

Nous avons pour cela réalisé deux combinaisons de couches : entre la couche du RPG ainsi que celles des données OCSOL et CLC. Les parcelles situées sur deux zones ou plus ont été découpées en autant d'entités. L'exemple ci-dessous montre la correspondance des parcelles déclarées comme vergers dans le RPG, respectivement avec CLC (Tableau V-2), ainsi que OCSOL (Tableau V-3).

CODE	Désignation	Nb entités	Superficie totale (ha)	Superficie moyenne (ha)	Part en %
112	Tissu urbain discontinu	13	5,02	0,39	0,33
121	Zones industrielles et commerciales	1	0,80	0,80	0,05
131	Extraction de matériaux	1	0,28	0,28	0,02
133	Chantiers	2	2,20	1,10	0,15
142	Equipements sportifs et de loisirs	1	0,71	0,71	0,05
211	Terres arables hors périmètres d'irrigation	4	2,28	0,57	0,15
221	Vignobles	15	236,30	15,75	15,75
222	Vergers et petits fruits	25	877,34	35,09	58,49
242	Systèmes culturaux et parcellaires complexes	25	218,84	8,75	14,59
243	Surfaces essentiellement agricoles, interrompues par des espaces naturels importants	26	106,76	4,11	7,12
311	Forêts de feuillus	13	26,58	2,04	1,77
312	Forêts de conifères	1	0,86	0,86	0,06
321	Pelouses et pâturages naturels	3	1,49	0,50	0,10
322	Landes et broussailles	1	1,12	1,12	0,07
323	Végétation sclérophylle	7	4,05	0,58	0,27
324	Forêt et végétation arbustive en mutation	4	12,19	3,05	0,81
411	Marais intérieurs	1	0,80	0,80	0,05
421	Marais maritimes	1	0,27	0,27	0,02
512	Plans d'eau	1	2,07	2,07	0,14
Total			1500 ha		100,00

Tableau V-2 : Répartition des vergers du RPG (2006) dans la couche Corine Land Cover (2006)

Les résultats avec CLC montrent qu'une majorité (58.49%) des vergers sont identifiés dans la bonne zone (222 : Vergers et petits fruits). De même, la présence de vergers dans les catégories 242 (Systèmes culturaux et parcellaires complexes) et 243 (Surfaces essentiellement agricoles, interrompues par des espaces naturels importants) ne constitue pas une erreur tant le libellé de la classe est générique.

La présence de vergers dans la catégorie 221 (Vignobles) reste mineure et nous avons également constaté que la superficie des vergers concernés ne dépassait jamais le seuil de la plus petite unité cartographiée dans CLC (soit 25 ha). Cela explique pourquoi ils n'ont pas été représentés. Enfin, pour

ce qui est des autres classes, elles s'expliquent par une imprécision dans la délimitation des zones. On note que la superficie des zones concernées est toujours très faible.

Pour OCSOL, les résultats sont beaucoup moins concluants (Tableau V-3). La répartition des vergers se fait équitablement (environ 25% pour chaque) entre trois catégories : 221 (Vignobles), 222 (Vergers et petits fruits) et 242 (Systèmes culturaux et parcellaires complexes).

Code	Désignation	Nb entités	Superficie totale (ha)	Superficie moyenne (ha)	Part en %
112	Tissu urbain discontinu	31	24,30	0,78	1,62
113	Bâti diffus	107	47,71	0,45	3,18
121	Zones industrielles et commerciales	5	1,52	0,30	0,10
133	Chantiers	1	2,20	2,20	0,15
212	Périmètres irrigués en permanence	9	6,28	0,70	0,42
214	Zones à forte densité de serres	12	5,41	0,45	0,36
221	Vignobles	46	411,25	8,94	27,41
222	Vergers et petits fruits	58	338,14	5,83	22,54
242	Systèmes culturaux et parcellaires complexes	87	410,30	4,72	27,35
243	Surfaces essentiellement agricoles, interrompues par des espaces naturels importants	6	13,05	2,17	0,87
311	Forêts de feuillus	30	17,18	0,57	1,14
312	Forêts de conifères	12	10,89	0,91	0,73
313	Forêts mélangées	1	0,14	0,14	0,01
321	Pelouses et pâturages naturels	25	24,93	1,00	1,66
323	Végétation sclérophylle	48	57,00	1,19	3,80
324	Forêt et végétation arbustive en mutation	36	32,07	0,89	2,14
325	Landes	32	72,29	2,26	4,82
333	Végétation clairsemée	12	13,39	1,12	0,89
421	Marais maritimes	1	10,63	10,63	0,71
511	Cours et voie d'eau	5	1,53	0,31	0,10
512	Plans d'eau	2	0,06	0,03	0,00
Total			1500 ha		100,00

Tableau V-3 : Répartition des vergers du RPG (2006) dans la couche OCSOL (2006)

Comme la plus petite unité cartographiée est indiquée comme étant de 2.5 ha, nous nous attendions à retrouver la majorité des vergers supérieurs à 2.5 ha dans la bonne catégorie (222) ou à défaut dans une catégorie plus englobante (242 ou 243).

Cette étude confirme le rapport sur la qualité de la BD OCSOL [Bourbier, Barret] qui indiquait de nombreuses confusions entre vignobles et vergers. Nous supposons que pour la réalisation de la BD OCSOL, la photo-interprétation des images a été orientée de manière à privilégier une meilleure identification du bâti diffus, au détriment des autres catégories. Il faut également noter que les données

CLC ont été vérifiées par superposition d'orthophotographies puis corrigées. Ce qui n'est pas le cas pour les données OCSOL.

Ainsi même si CLC offre une résolution inférieure à OCSOL, la BD est beaucoup plus cohérente avec le RPG et donc plus efficace pour l'identification précise des pratiques culturales. Or comme il s'agit de notre objectif, nous avons pris la décision d'écarter l'utilisation d'OCSOL. Qui plus est, comme la donnée ne couvre que le littoral méditerranéen français, son intégration aurait nuit à la réutilisabilité du modèle en dehors de cette zone. Nous utilisons les informations fournies par Corine Land Cover pour nous aider à compléter les données fournies par le RPG.

Concernant les données météorologiques, nous reprenons les données utilisées dans le cadre du projet VULCAIN. C'est-à-dire les mesures fournies par le système SAFRAN de Météo France. La production de ces données s'appuie sur la collecte de plusieurs mesures météorologiques : température, taux d'humidité, vitesse du vent, précipitations pluviales et neigeuses, rayonnement solaire et infrarouge incident. Ces données sont ensuite analysées par pas de 300 m d'altitude, puis interpolées sur une grille de calcul de dimensions 8 x 8 km (en projection Lambert II cartographique). Ces données nous fournissent des mesures quotidiennes de température, d'évapotranspiration, ainsi que de précipitations pluviales et neigeuses. Les valeurs d'évapotranspiration sont calculées selon la méthode de Penman-Monteith, comme il est recommandé par la FAO.

En matière de données pédologiques, nous nous appuyons sur des données fournies par l'INRA. Celles-ci proviennent d'une étude du sol menée sur la région Languedoc Roussillon en 1999. Elles cartographient les différents types de sol dans notre zone d'étude et précisent leurs caractéristiques, en particulier leurs propriétés hydriques. Compte tenu de l'échelle de travail à laquelle ont été créées les données, l'étude ne prend pas en compte les spécificités locales des sols, c'est-à-dire d'une parcelle à l'autre. Ainsi, dans le cadre du projet FUSIEAU nous utilisons les données suivantes (Tableau V-4) :

Thématique	Description	Fournisseur	Echelle spatiale équivalente	Information sur l'échelle temporelle
Phénologie	Durées des cycles de vie et valeurs des Kc	FAO-56	Sans	Estimation journalière
Pédologie	BD DONESOL	INRA	1/250 000	Une réalisation en 1999
Occupation du sol	Corine Land Cover	Agence Environnementale Européenne	1/100000	Trois réalisations : 1990, 2000 et 2006 (prochaine en 2012).
	Registre parcellaire graphique	Agence de Services et de Paiement (ASP)	1/5000	Tous les ans depuis 2006
Climatologie	Pluviométrie, nivométrie, ETP et température	Météo France	Grille SAFRAN 8km x 8km	Mesures journalières depuis 1970
Prélèvements en eau	Mesures de prélèvements	Agence de l'eau Rhône Méditerranée Corse	Données ponctuelles	Mesures annuelles depuis 1987

Tableau V-4: Liste et caractéristiques des données utilisées dans le modèle [Beaufils & al, 2011 bis]

Ces données sont hétérogènes. Les données phénologiques ne sont pas des données spatiales. Les données météorologiques quant à elles sont avant tout une collection de mesures, qui ne peuvent être reliées à des entités spatiales que par la zone d'étude qu'elles couvrent. Les autres présentent quant à elles des échelles spatiales variées, ainsi que des dates de disponibilité différentes.

V.A.3 Identification des traitements et de leur ordre d'enchaînement

L'évaluation des besoins en eaux telle qu'elle est réalisée par le biais du modèle FAO correspond à des fusions de deuxième niveau. Il s'agit en effet de combiner des données attachées à des phénomènes différents. Suite à l'étude du modèle FAO, nous avons identifié les calculs à effectuer et l'ordre d'enchaînement de ces calculs. Chacun des processus propose la fusion de différents types de données. La combinaison de deux données ou plus issues de thématiques distinctes conduit à la production d'une donnée hybride.

Le schéma ci-dessous (Figure V-3) présente les calculs réalisés et l'ordre dans lequel ils sont effectués. Des données issues de différentes thématiques sont combinées afin de produire de nouvelles informations hybrides.

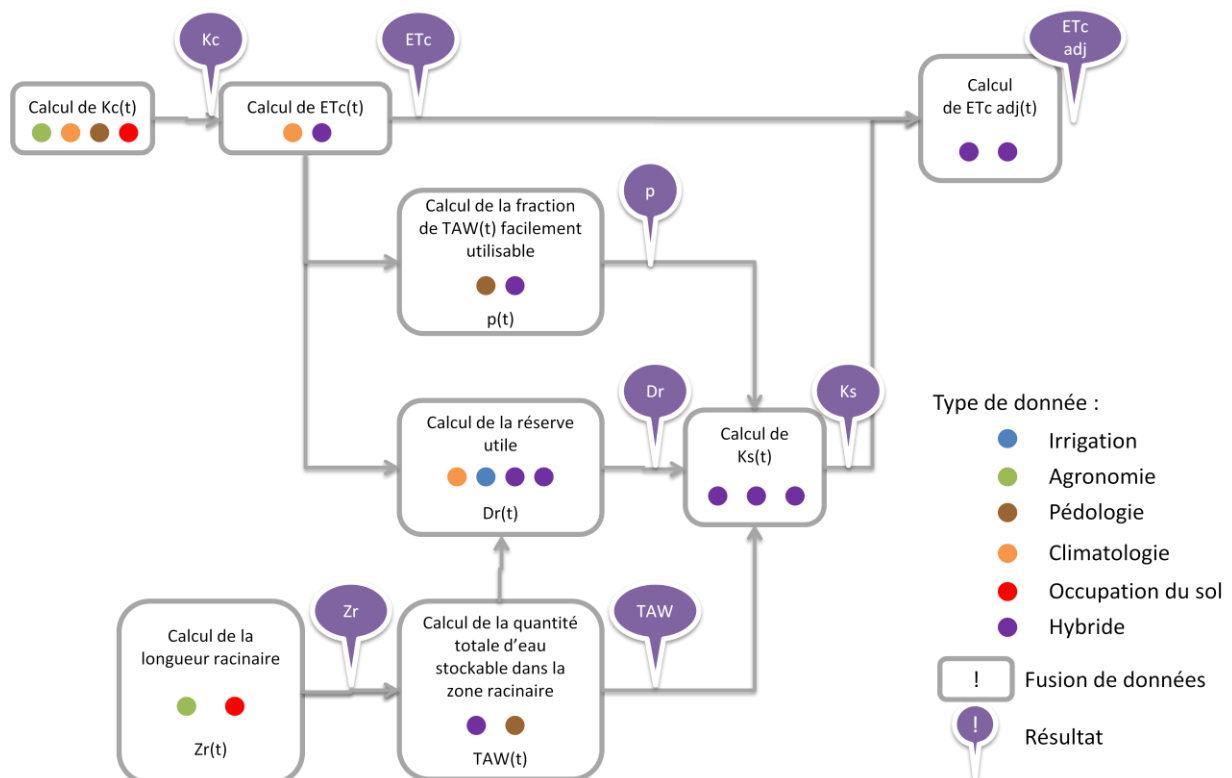


Figure V-3 : Représentation schématisée de notre implémentation du modèle FAO

Si on reprend la formule de calcul du besoin en eau de la culture ETC_{Adj} (Figure I-7), on note que le modèle repose principalement sur la détermination de deux paramètres : K_c et K_s , dont les valeurs évoluent au cours du cycle de vie de la plante. On parle alors de $K_c(t)$ et $K_s(t)$.

La première étape du modèle consiste à déterminer $K_c(t)$. Les valeurs dépendent principalement du type de la culture (caractéristique agronomique), toutefois les conditions de cultures (climat, type de sol) interviennent également dans le calcul. Les données d'occupation du sol permettent de situer la culture géographiquement et donc d'identifier les conditions de cultures. Une fois $K_c(t)$ déterminé, on en déduit $ET_c(t)$ l'évapotranspiration de la plante en supposant des conditions optimales de culture par la formule indiquée à la Figure I-6.

La seconde partie du modèle porte sur le calcul de $K_s(t)$. Son calcul repose sur la réalisation d'un bilan hydrique de la réserve en eau. On commence par calculer la longueur des racines $Z_r(t)$, si celle-ci évolue. Puis on détermine la quantité totale d'eau (Total Available Water) $TAW(t)$ qui peut être stockée dans la réserve en eau (c'est-à-dire la partie de la terre dans laquelle la plante peut puiser de l'eau en cas de besoin).

Après cela on calcule la fraction $p(t)$ de la réserve en eau facilement utilisable (souvent abrégée RFU). Puis on détermine l'évolution du remplissage de cette réserve en eau $Dr(t)$, en faisant un bilan hydrique. Quand il pleut ou que l'on irrigue la réserve en eau se remplit, sinon elle se vide au gré des besoins de la plante. L'évolution de la réserve en eau nous permet ensuite de déterminer le stress hydrique $K_s(t)$ subi par la plante. La détermination de la valeur finale $ET_c Adj(t)$ est alors triviale.

V.B Proposition d'un modèle d'estimation des besoins en eau domestiques

La définition de modèles d'estimation des besoins en eau domestique a fait l'objet d'un travail de fin d'études d'école d'ingénieur, encadré et dirigé par l'auteur. La présente section résume les principaux travaux et résultats de ce travail de recherche. L'ensemble des travaux réalisés figure dans [Berge, 2012].

V.B.1 Approche des modèles

La corrélation entre la consommation en eau d'une commune et sa population est évidente. De ce fait, il semble inévitable de proposer un modèle permettant d'estimer les besoins en eau de la population permanente. Pour de nombreuses communes, la consommation en eau de la population permanente sera entièrement liée à la consommation totale en eaux domestiques de la commune. Un modèle estimant les besoins en eau de la population permanente sera alors suffisant pour traduire la consommation en eau de la zone d'étude.

Toutefois, le projet VULCAIN a montré que compte tenu du caractère hautement touristique du département des Pyrénées-Orientales, il est nécessaire de prendre également en compte la

consommation en eau engendrée par la population saisonnière. Nous avons donc entrepris de concevoir deux modèles distincts :

- un modèle estimant les besoins en eau de la population permanente ;
- un modèle estimant les besoins en eau liés au tourisme.

Dans la suite nous nous sommes attachés à produire des modèles raisonnant à l'échelle communale et au pas de temps mensuel.

Le choix de l'échelle communale s'est imposé comme une évidence. La politique de gestion de l'eau est définie à l'échelle communale (voire intercommunale), aussi il nous a paru logique de raisonner à cette échelle. Utiliser ce découpage permet également de passer facilement à de plus petites échelles : intercommunalité, canton, département, ... Il suffit en effet de faire la somme des besoins de chaque commune constituant l'entité souhaitée.

Le choix du pas de temps mensuel fut plus délicat. Il s'agit en fait d'un compromis par rapport aux données que nous avons à disposition. Un raisonnement au pas de temps décadaire aurait permis de rendre compte des variations de consommation liées aux vacances non estivales, un raisonnement à l'échelle journalière aurait permis de rendre compte des variations de consommations liées aux week-ends. Les données dont nous disposions étant le plus souvent exprimées au pas de temps annuel, et au mieux mensuel, il était délicat de passer à un pas de temps plus fin.

V.B.2 Estimation des besoins en eau liés à la population permanente

Pour l'estimation des besoins en eaux liés à la population permanente, nous sommes dans un premier temps partis sur l'exploitation des données d'occupation du sol (en particulier celles de la BD OCSOL, décrites pour le modèle des besoins agricole). L'idée était de déterminer une relation entre la superficie des zones du territoire dédiées à l'habitation (code 111, 112 et 113 dans la BD OCSOL) et la population permanente.

Menée sur l'ensemble des communes du département des Pyrénées Orientales, ainsi que sur des groupes de communes, les résultats de cette approche ne se sont pas avérés concluants. Nous sommes donc partis sur une autre voie et avons donc proposé un modèle classique basé sur l'application de ratios de consommations à la population permanente.

La principale originalité du modèle est qu'il s'appuie sur la décomposition de la population selon les ménages. Plusieurs études [Barraque & Nercissian, 2008 ; Souriau, 2011 ; Mirad, 2004], ont en effet montré que les ratios de consommation individuels n'étaient pas les mêmes que l'on habite seul ou à plusieurs. Ces constatations ont motivé le choix d'un tel modèle. Le tableau suivant (Tableau V-5) expose les résultats obtenus suite aux deux études évoquées précédemment :

<i>Nb pers. par ménage</i>	<i>[Haniquaut, 2005]</i>		<i>[Barraqué & Nercissian, 2008]</i>	
	<i>Echelle nationale, constats</i>		<i>Paris, sondage</i>	
	<i>Conso (en m³ par an)</i>	<i>Ratio (en L./j./pers.)</i>	<i>Conso (en m³ par an)</i>	<i>Ratio (en L./j./pers.)</i>
<i>1</i>	55	151	52,4	144
<i>2</i>	100	137	90,4	124
<i>3</i>	120	110	119,7	109
<i>4</i>	140	96	147,2	101
<i>5</i>	160	88	169,1	93
<i>6</i>	180	82	186,9	85

Tableau V-5 : Ratios de consommation des ménages [Berge, 2012]

Les données relatives au recensement des ménages proviennent des bases de données de l'INSEE. Elles ont l'avantage d'être gratuites et mises à jour régulièrement. De plus, les données sont disponibles pour l'ensemble des communes du territoire national. Cette caractéristique rend le modèle applicable à l'ensemble des départements français.

L'INSEE met à disposition plusieurs données relatives aux ménages :

- découpage selon la composition du ménage (personne seule, couple, couple avec enfants, famille monoparentale) ;
- découpage selon la taille du ménage et les caractéristiques du chef de famille (âge et sexe, ou profession) ;
- découpage selon la taille du ménage et le type d'habitation (appartement ou maison).

Le modèle s'appuie principalement sur les données issues de la troisième répartition. Les ratios de consommation appliqués sont ajustés en fonction de plusieurs critères :

- le type de logement : on applique un coefficient de majoration aux maisons individuelles (surconsommation liée à l'entretien d'un jardin, au remplissage d'une piscine) ;
- la météorologie, et en particulier la température qui peuvent modifier les habitudes de consommation de la population (surconsommation en période de canicule).

Le schéma ci-dessous (Figure V-4) présente les différents traitements réalisés pour l'estimation des besoins en eau de la population permanente, leur ordre d'enchaînement ainsi que les services web choisis pour la diffusion des données.

Le point de départ du modèle est une couche d'informations géographiques comportant les données de l'INSEE relatives aux ménages (idéalement cette couche de données serait diffusée par l'INSEE). Après une intervention de l'utilisateur visant à sélectionner les communes, et la période qu'il souhaite étudier, nous créons des capteurs virtuels pour chacune des communes sélectionnées et leur affectons plusieurs mesures : d'abord le nombre de ménage selon le logement par extraction des données de la couche de l'INSEE, puis selon le principe mis en place pour le modèle d'estimation des besoins en eau agricoles (Partie IV.C.2), les résultats des différents calculs proposés dans le modèle.

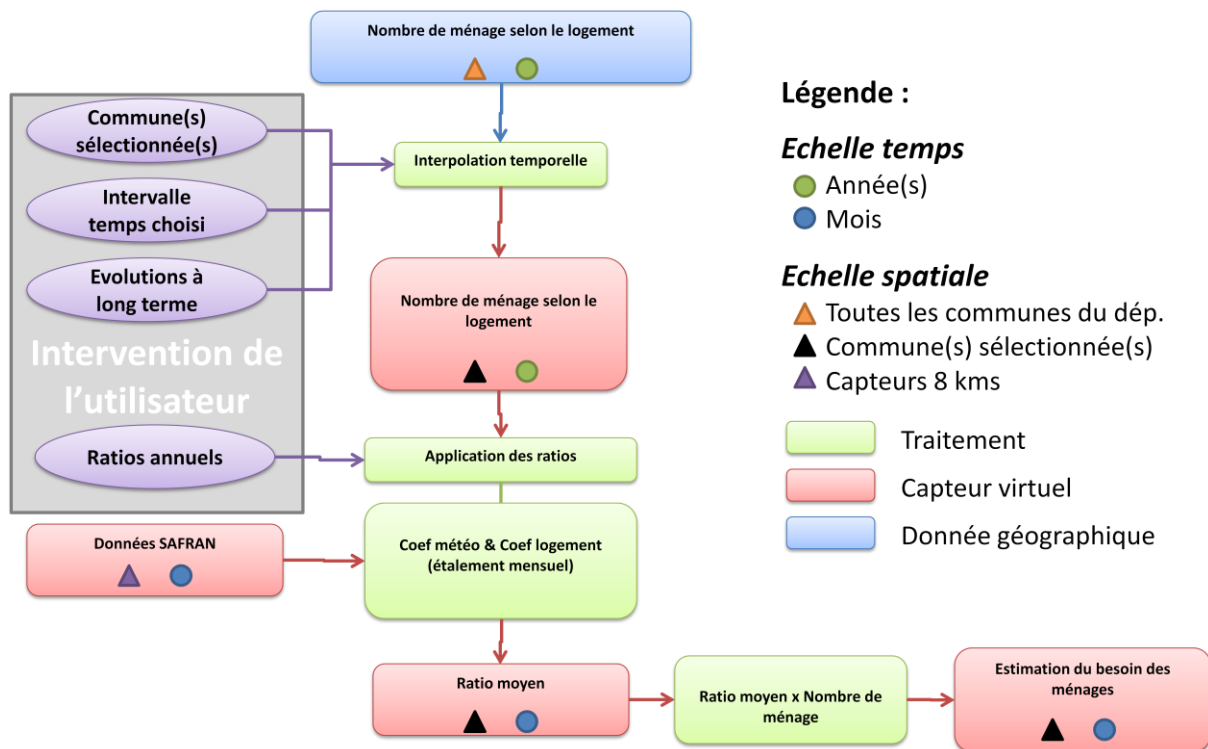


Figure V-4 : Schéma du modèle d'estimation des besoins en eau domestique issus de la population permanente, d'après [Berge, 2012]

Pour consulter alors les résultats intermédiaires et finaux du modèle, il suffit d'interroger les capteurs associés à chacune des communes étudiées.

V.B.3 Estimation des besoins en eaux liés au tourisme

Comme nous avons pu le voir dans notre état de l'art (Partie I.C) peu de modèles s'intéressant à l'estimation des besoins en eau du tourisme sont disponibles. L'approche retenue dans le projet VULCAIN est assez simpliste et consiste à affecter l'ensemble du tourisme annuel sur les mois de juillet et août. Cette technique est appliquée à l'ensemble des communes. Or certaines communes ont une répartition temporelle bien différente du tourisme. Les communes situées par exemple en région montagneuse connaissent trois principales périodes d'afflux touristiques : durant les vacances de Noël, durant les vacances d'hiver, ainsi qu'en été. Aussi, affecter l'ensemble des touristes à la période estivale ne correspond donc en rien à la réalité.

De plus, des études menées par l'Agence de l'Eau Loire-Bretagne [AELB, 2005] ont montré que la consommation en eau des touristes varie selon le type d'hébergement, et même selon son niveau de standing. Les écarts peuvent être conséquents : on estime ainsi que les consommations en eau s'élèvent à 140 litres par nuitée dans un camping « moyennement équipé », tandis qu'elles s'élèvent à 580 litres par nuitée dans un hôtel 4 étoiles [Berge, 2012].

Le modèle proposé prend en compte ces deux facteurs et propose ainsi de déterminer le nombre de nuitées passées dans chaque type d'hébergement offerts par la commune. L'approche se fait ainsi en deux étapes :

- répartir les touristes du département dans les différentes communes. Nous proposons pour cela de réaliser une typologie des communes en matière de tourisme. Le principe est tout d'abord de catégoriser les différentes communes, en fonction de leur attractivité. La réalisation de cette entreprise passe par une analyse du territoire, et en l'établissement de règles. La catégorisation va se faire en fonction de critères attributaires (présence d'une station thermale dans la commune, présence d'une station de ski, ...), ou spatiaux (proximité de la mer) ;
- identifier le nombre d'hébergements de chaque type offerts par la commune et répartir les touristes dans ces hébergements. Cette répartition varie en fonction de la saison (pas de camping en hiver), mais également de la zone touristique dans laquelle se situe la commune. En première approche nous supposons une équi-répartition des touristes dans les différents hébergements proposés.

Le modèle d'estimation des besoins en eau liés au tourisme, s'appuie sur deux sources de données :

- les bases de données détaillées de l'INSEE nous fournissent des informations relatives aux capacités d'accueil des communes en distinguant plusieurs types d'hébergements (hôtels, camping, résidences secondaires) ;
- les données de l'Observatoire du Tourisme du département des Pyrénées-Orientales nous fournissent des mesures de fréquentation des hébergements touristiques.

Si les données de l'INSEE s'appuient bien entendu sur une échelle communale, celles de l'Observatoire sont diffusées à une échelle particulière : celle des zones touristiques (Figure V-5). Le département est ainsi découpé en quatre secteurs:

- le secteur Littoral, situé sur la côte méditerranéenne. Cette zone représente 60% des nuitées dans le département ;
- le secteur Montagne, situé à l'est du département. Ce secteur profite de l'attrait des Pyrénées, dont la fréquentation touristique se fait aussi bien en hiver qu'en été ;
- le secteur Passage, qui contient la ville de Perpignan ainsi que sa banlieue. Sa fréquentation est fortement liée aux événements ayant lieu à Perpignan ;
- le secteur Arrière-pays situé entre le littoral et la montagne, qui couvre la moitié des communes du département. Il comporte aussi bien des communes très attractives (présence de station thermales) que non touristiques. L'une des difficultés pour le modèle va être de tenir compte de ces hétérogénéités en matière de fréquentation touristique.

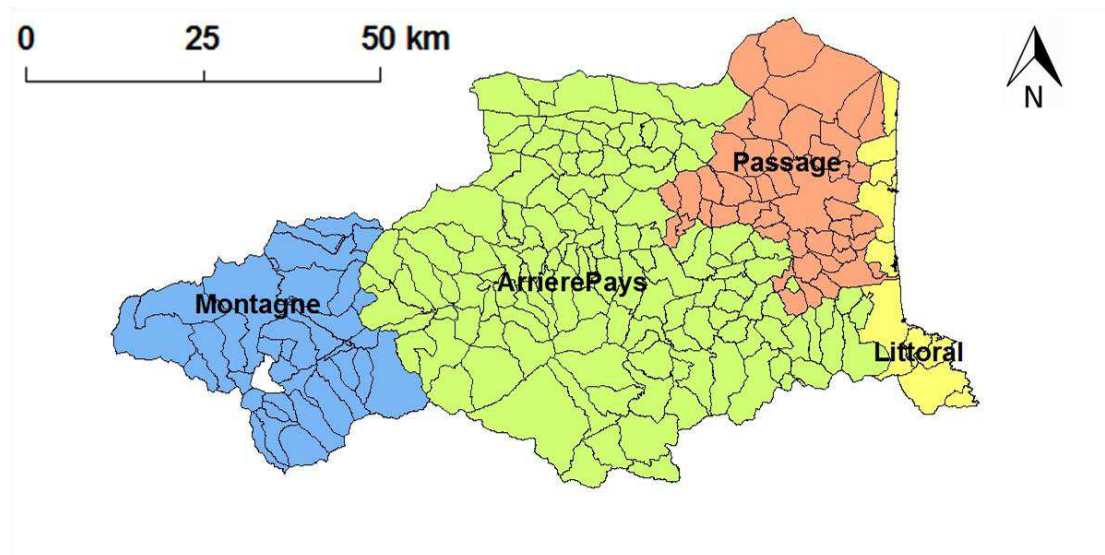


Figure V-5 : Carte de répartition des secteurs touristiques [Berge, 2012]

Le schéma suivant (Figure V-6) présente les différents traitements réalisés pour l'estimation des besoins en eau liés aux touristes ainsi que leur ordre d'enchaînement.

Tout comme le modèle d'estimation des besoins en eau de la population permanente, l'utilisateur est invité à sélectionner les communes et périodes qu'il souhaite étudier.

Le modèle comporte ensuite une étape de pré-traitement qui vise à définir le type de chacune des communes en matière d'attractivité touristique. Actuellement cette typologie est réalisée en fonction des attributs de la commune (nombre de campings, présence d'une station thermale, etc...). Néanmoins dans le futur cette typologie pourrait être construite par le biais de requêtes d'analyse spatiale (croisement d'une couche contenant les stations thermales avec celle des communes, de même avec les emplacements de camping). L'étape de pré-traitement n'a pas la vocation d'être réalisée pour chacune des simulations (sauf si l'utilisateur modifie la catégorisation des communes).

Partant ensuite des données de l'INSEE relatives à l'offre d'hébergement des communes, nous dotons chacune des communes d'un capteur virtuel et y insérons les données de l'INSEE. Puis par croisement des données relatives à la fréquentation et des capacités d'hébergement nous procédons au remplissage des différents lieux d'accueil. En appliquant ensuite un ratio de consommation propre à chaque type d'hébergement, on en déduit une estimation des besoins en eaux liés à l'afflux touristique pour chacune des communes.

De même que pour le modèle pour les besoins en eau de la population permanente, il suffit d'interroger les capteurs associés à chacune des communes étudiées pour consulter les résultats intermédiaires et finaux du modèle.

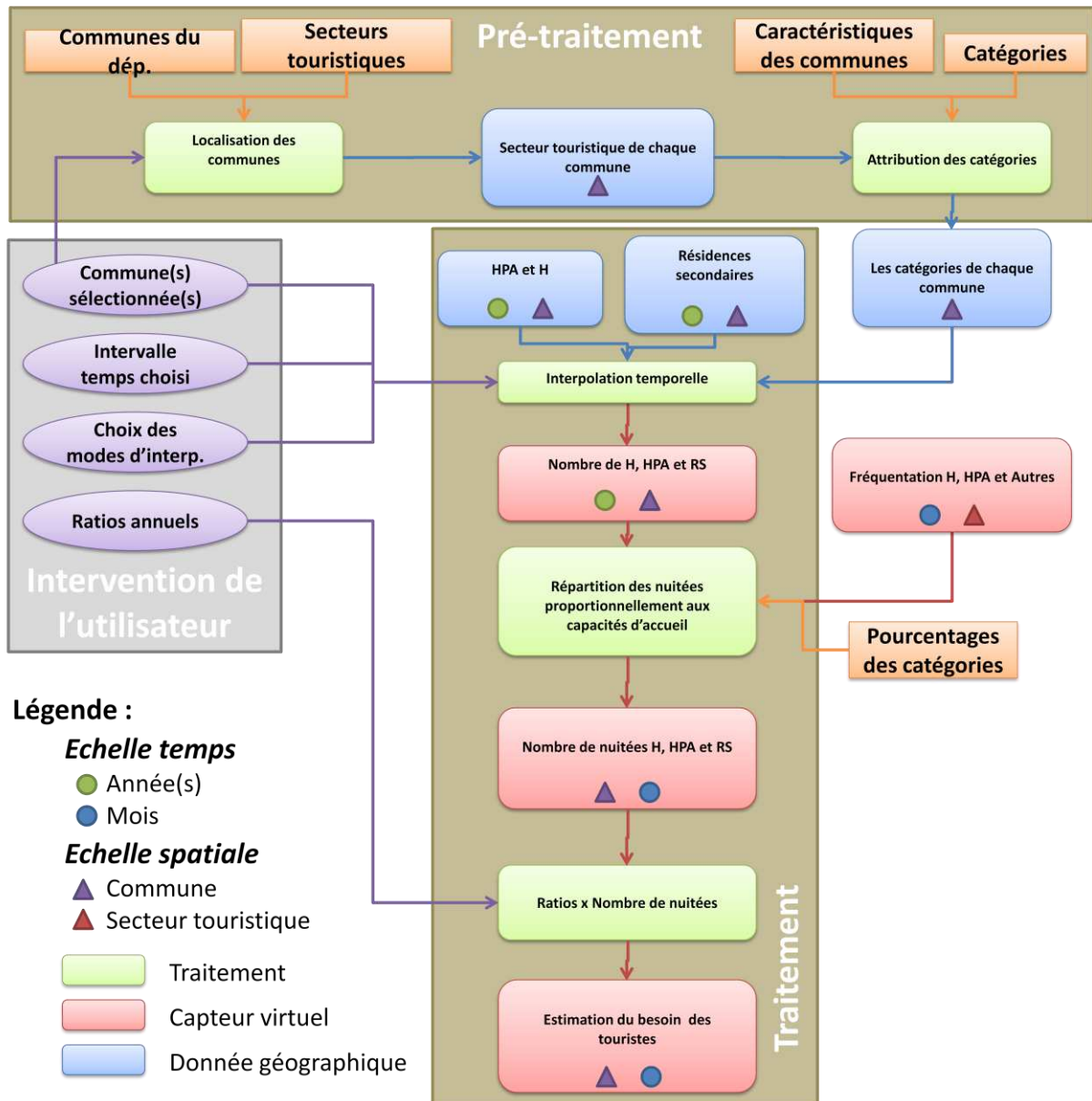


Figure V-6 : Schéma du modèle d'estimation des besoins en eau domestique issus du tourisme, d'après [Berge, 2012]

Une limite de ce modèle est sa transposabilité sur d'autres zones d'études. En effet, même si les départements français disposent d'un Comité Départemental du Tourisme qui comptabilise la fréquentation touristique, la façon dont ces organismes diffusent leurs données est hétérogène. Il est ainsi rare de constater des distinctions en zones touristiques à l'intérieur même des départements.

A noter également que certaines communes (principalement les chefs-lieux de département et les communes les plus attractives) disposent également de leurs propres chiffres sur la fréquentation touristique. L'exploitation de ces données permettrait de mieux répartir la fréquentation touristique départementale.

V.C Evolutions de FUSIEAU

L'évaluation des besoins en eau constitue une première étape pour la réalisation d'un bilan hydrologique, c'est-à-dire l'étude de l'adéquation entre ressources et demandes en eau. Nous étudions donc les étapes supplémentaires à mettre en œuvre pour permettre la réalisation de ce bilan hydrologique.

Une autre évolution intéressante de FUSIEAU est son utilisation dans un cadre prédictif. La principale limite de cette approche est l'obtention de données futures. Nous traitons donc d'une méthode basée sur l'utilisation de données historiques pour une exploitation dans un but prédictif.

V.C.1 Vers l'évaluation des ressources en eau à l'échelle des bassins versants

L'échelle du bassin versant est pertinente pour un expert hydrologue car elle permet, en particulier pour un bassin versant de socle, d'appliquer le principe de conservation de la ressource en eau. En effet selon cette approche, toute goutte d'eau tombée dans un bassin versant est soit consommée (pour un usage spécifique ou en tant que perte), soit se retrouve à l'exutoire du bassin versant. L'échelle du bassin versant permet ainsi de réaliser ce que l'on appelle un bilan hydrologique, c'est-à-dire la comparaison des entrées et des sorties en eau.

Pour pouvoir réaliser un bilan hydrologique à l'échelle d'un bassin versant, il est nécessaire de passer des besoins en eau théoriques, obtenus via nos modèles aux quantités réellement prélevées. Le passage de l'un à l'autre doit prendre en considération trois points importants :

- le volume d'eau prélevé à la source est toujours supérieur à celui effectivement distribué aux entités consommatrices (parcelle irriguée, commune) à cause des pertes liées à l'acheminement en eau. Comme nous l'avons vu dans notre état de l'art sur les modèles (Chapitre I), l'approche souvent retenue est d'appliquer un coefficient de perte ;
- un point de prélèvement peut alimenter plusieurs entités, et réciproquement plusieurs points de prélèvements peuvent venir alimenter une seule et même entité consommatrice (Figure V-7). Le projet VULCAIN a mis en évidence la difficulté de relier les points de prélèvement aux entités consommatrices (groupes de population, parcelles agricoles) ;
- le lieu où l'eau est prélevée et l'endroit où elle est consommée (la parcelle irriguée, la commune alimentée en eau potable) diffèrent et peuvent parfois se trouver dans des bassins versants différents, compliquant ainsi la réalisation du bilan hydrologique.

Des mesures des quantités d'eau prélevées sont réalisées régulièrement et stockées dans des bases de données. La disponibilité de ces données et notamment la fréquence des mesures réalisées varient selon les organismes gestionnaires. Les données fournies par les Agences de l'Eau indiquent la

quantité d'eau prélevée annuellement. La position des points de prélèvement est indiquée dans ces données. Cette information peut dans les cas simples (Figure V-7), suffire pour relier les prélèvements aux besoins : c'est-à-dire permettre de savoir que les besoins d'une commune, ou d'un groupe de parcelles sont satisfaits par un prélèvement précis. Néanmoins dans les cas complexes, des données des organismes distributeurs d'eau ou des associations syndicales autorisées d'irrigation relatives à la répartition des eaux pompées sont indispensables.

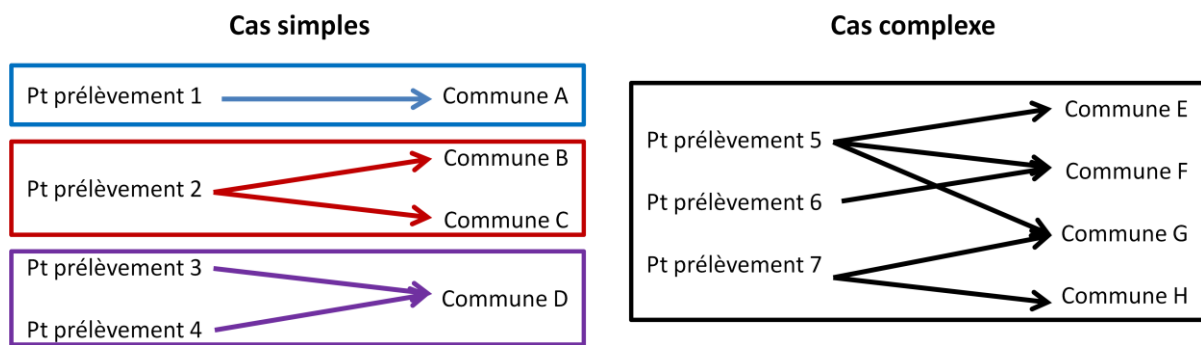


Figure V-7 : Unités de distribution en eau potable

En l'absence de ces informations, on peut toutefois utiliser le potentiel offert par les données géoréférencées dont nous disposons pour réaliser des postulats d'association : point de prélèvement / entité consommatrice.

Par exemple, dans le cadre de l'irrigation par canaux (majoritaire dans le département des Pyrénées-Orientales), le tracé des canaux peut nous aider à identifier les parcelles irriguées. A l'exception des canaux souterrains, les principaux canaux sont cartographiés dans la BD TOPO de l'IGN.

L'hypothèse d'une irrigation gravitaire (par écoulement selon la pente du terrain), nous indique que les parcelles pouvant être irriguées de cette manière se trouvent en aval du tracé du canal. Aussi en procédant à une analyse spatiale, il est possible d'estimer les parcelles susceptibles d'être irriguées par méthode gravitaire.

Le schéma suivant (Figure V-8) montre la détermination du point de prélèvement associé à la parcelle (en vert). Un canal d'irrigation (en orange) passe près de la parcelle. Si la parcelle est située en aval du canal, alors il y a de grande chance que le canal participe à l'irrigation de la parcelle. Le point de prélèvement associé est alors celui situé à l'entrée du canal.

L'utilisation d'un modèle numérique de terrain (MNT), nous permettrait de connaître facilement les zones situées en aval d'un point choisi. La BD ALTI de l'IGN qui propose un MNT au pas de 25m, offrant une précision altimétrique d'environ 1m, permettrait de définir précisément l'écoulement des eaux. Cette méthode reste toutefois théorique et mériterait d'être expérimentée.

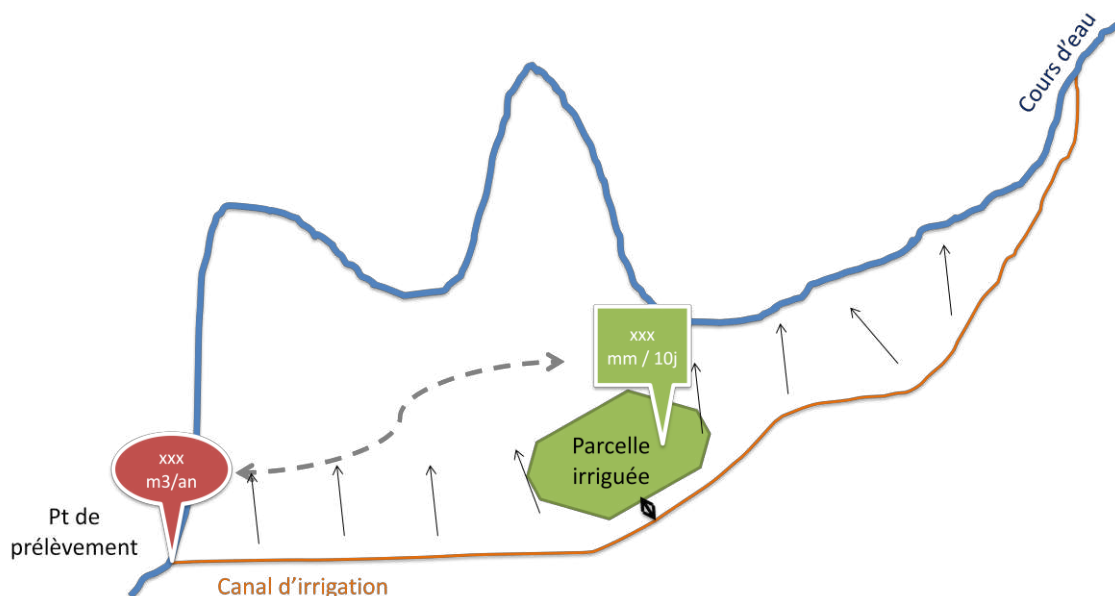


Figure V-8 : Détermination des parcelles irriguées à partir du tracé des canaux et de la pente

Cet exemple illustre l'intérêt ainsi que le potentiel offert par l'utilisation d'une application cartographique plutôt qu'un outil basé uniquement sur l'application de ratios et de coefficients.

V.C.2 Vers un modèle prévisionnel

L'utilisation des modèles développés dans un cadre prédictif constitue un objectif majeur de FUSIEAU. La réalisation d'une telle entreprise passe par l'exécution du modèle avec des données que l'on pense être représentatives du futur. L'obtention de données prévisionnelles fiables est très délicate. Les prévisions des météorologistes s'appuient sur des tendances globales, par exemple le réchauffement climatique depuis le début du siècle. Toutefois les données issues de ces prévisions ne possèdent pas une granularité temporelle assez importante pour permettre une évaluation des besoins en eau décadaire voire mensuelle.

FUSIEAU n'intègre pas (pour le moment) de modèles permettant de créer ses propres chroniques météorologiques. Toutefois, nous proposons de réutiliser les données historiques dont nous disposons : traitement statistique des données passées pour évaluer les besoins en eau futurs. Pour rappel, les données météorologiques dont nous disposons sont issues du système SAFRAN de Météo France. Elles nous fournissent des mesures journalières de température, de précipitations pluvieuses et neigeuses, ainsi que de l'évapotranspiration. Nous accédons ainsi à 36 années de ces mesures : depuis 1970 jusqu'à 2006.

Notre approche consiste à caractériser chacune des années par rapport à des valeurs statistiques obtenues sur les 36 années. L'idée est de proposer à l'utilisateur de déterminer les besoins en eau pour une année caractéristique : par exemple une année quinquennale sèche.

La première étape a donc été de caractériser les données dont nous disposions. Nous avons alors procédé à une analyse statistique des mesures et observations. Nous avons ainsi cherché à déterminer des valeurs statistiques de référence pertinentes pour l'utilisateur. Notre attention s'est portée sur la détermination des valeurs suivantes (Tableau V-6) :

Nom de la valeur	Caractérisation
Moyenne	Valeur moyenne recensée sur les 36 années
Ecart-type	Répartition des valeurs
Valeur minimum	Plus faible valeur relevée
Valeur maximum	Plus forte valeur relevée
Quinquennale sèche (ou froide dans le cas de la température)	Niveau le plus bas pouvant être atteint une année sur cinq
Quinquennale humide (ou chaude dans le cas de la température)	Niveau le plus haut pouvant être atteint une année sur cinq
Décennale sèche (ou froide dans le cas de la température)	Niveau le plus bas pouvant être atteint une année sur dix
Décennale humide (ou chaude dans le cas de la température)	Niveau le plus haut pouvant être atteint une année sur dix

Tableau V-6 : Valeurs caractéristiques, d'après [Berge, 2012]

Pour établir ces valeurs, nous nous sommes appuyés sur des lois statistiques usuelles. Les températures étant réparties de façon symétriques autour de la moyenne, il est pertinent d'établir les quinquennales et les décennales à partir de la Loi Normale (ou loi de Gauss). La valeur statistique souhaitée est déterminée au travers de la Loi Normale centrée réduite Inverse en fonction de la probabilité, la moyenne et l'écart-type.

Dans le cas des précipitations toutefois, l'Observatoire Gravitaire Géo Azur a mis en avant dans une de ses études [Berolo, 2010] que : « à l'échelle mensuelle, la distribution des événements pluviométriques étant fortement dissymétrique par rapport à la valeur moyenne, la racine carrée de la variable est gaussienne ». La loi Racine Normale se présente comme la solution pour établir les quinquennales et décennales souhaitées. Elle fonctionne de la même façon que la loi Normale. La seule différence est que la moyenne et l'écart-type sont calculés à partir de la racine carrée de chaque valeur.

Le schéma suivant (Figure V-9) présente les résultats obtenus pour le mois d'Août pour la maille SAFRAN n°9764, située à l'est de Perpignan. On remarque que sur la période de 1970 à 2005, l'année 2001 a présenté le mois d'août le plus sec, tandis que celui de 1974 fut de loin le plus humide. Ces données sont bien entendu locales et valables uniquement pour la maille SAFRAN sur laquelle a été faite l'analyse.

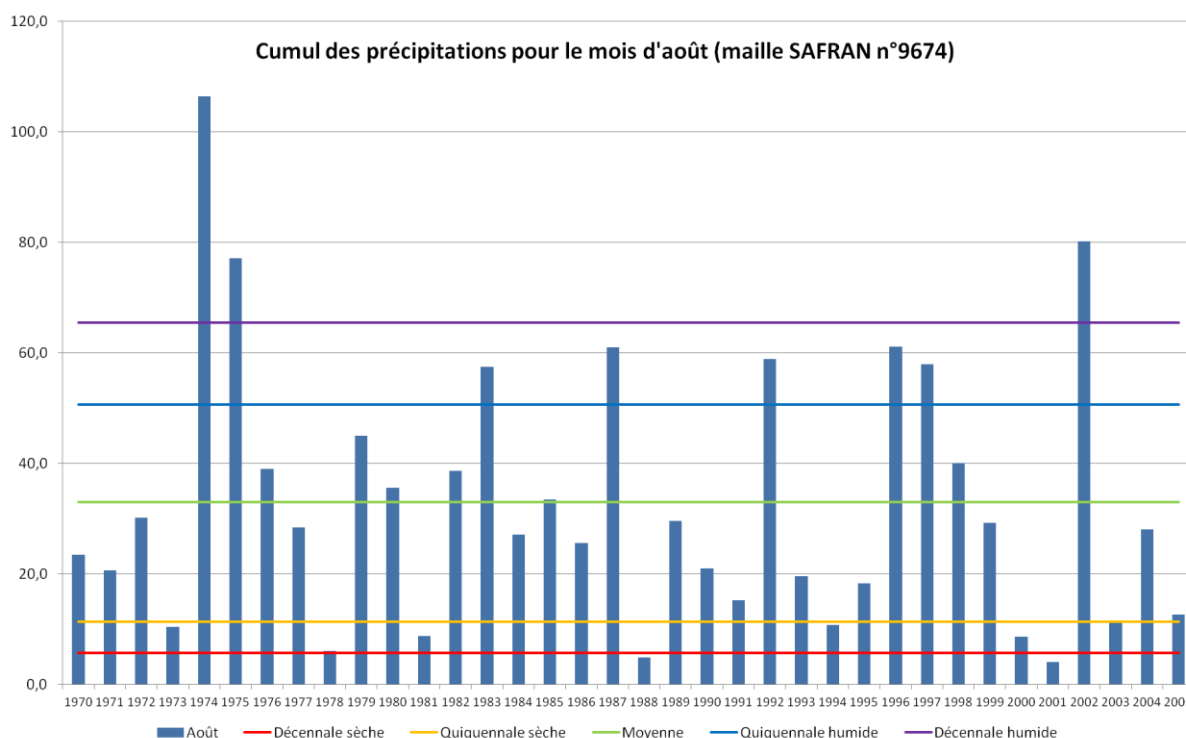


Figure V-9 : Caractérisation des cumuls de précipitation enregistrés pour le mois d’août sur la période 1970-2005. Cas de la maille SAFRAN n°9674 près de Perpignan

Les résultats de ces analyses permettent de caractériser la météorologie des différentes années dont nous disposons en indiquant les spécificités de chacun des mois qui la compose. Le tableau suivant (Tableau V-7) présente les résultats obtenus pour l’année 2001 en matière de cumul mensuel de précipitation. Il permet de positionner de manière simple l’année étudiée par rapport aux valeurs de référence définies dans le Tableau V-6. On remarque que cette année a été marquée par une saison estivale particulière. C’est en effet au cours de cette année qu’ont été enregistrés les records de précipitation maximale (en juillet) mais également minimale (en août). Les autres saisons présentent des précipitations proches de la moyenne, à l’exception du mois de décembre qui fut très sec.

2001	Min		DSec		QSec		Moy		QHu		DHu		Max
Janvier	0,3		3,3		12,8		64,7	103,2	107,9		148,1		220,9
Février	1,6		3,4		11,0	13,2	48,4		79,9		108,3		229,1
Mars	0,1		5,1		12,0		40,7	43,1	65,4		86,4		166,7
Avril	7,4		8,2		17,4	23,4	53,0		84,0		109,6		217,6
Mai	5,6		8,7		17,9	21,9	53,2		83,9		109,2		261,1
Juin	1,7		4,8	8,1	10,3		31,8		50,4		65,9		104,0
Juillet	0,3		1,9		4,5		14,9		23,9		31,6	51,1	51,1
Août	4,1	4,1	5,8		11,4		33,0		50,6		65,5		106,4
Septembre	0,0		2,7		10,0	44,2	48,1		79,9		109,2		269,8
Octobre	0,6		5,2		17,4	29,1	83,7		129,9		176,6		293,5
Novembre	3,0		2,2		10,5		59,3	95,7	99,0		137,0		273,6
Décembre	0,2	0,4	3,3		12,5		62,9		102,1		139,7		263,8

Tableau V-7 : Caractérisation de l'année 2001 pour la maille SAFRAN n°9674 en matière de cumul de précipitations mensuelles

Les résultats contenus dans ce même tableau pourront également être présentés de manière graphique sous la forme de boîte à moustaches (Figure V-10). Pour une meilleure lisibilité, nous utilisons une échelle logarithmique.

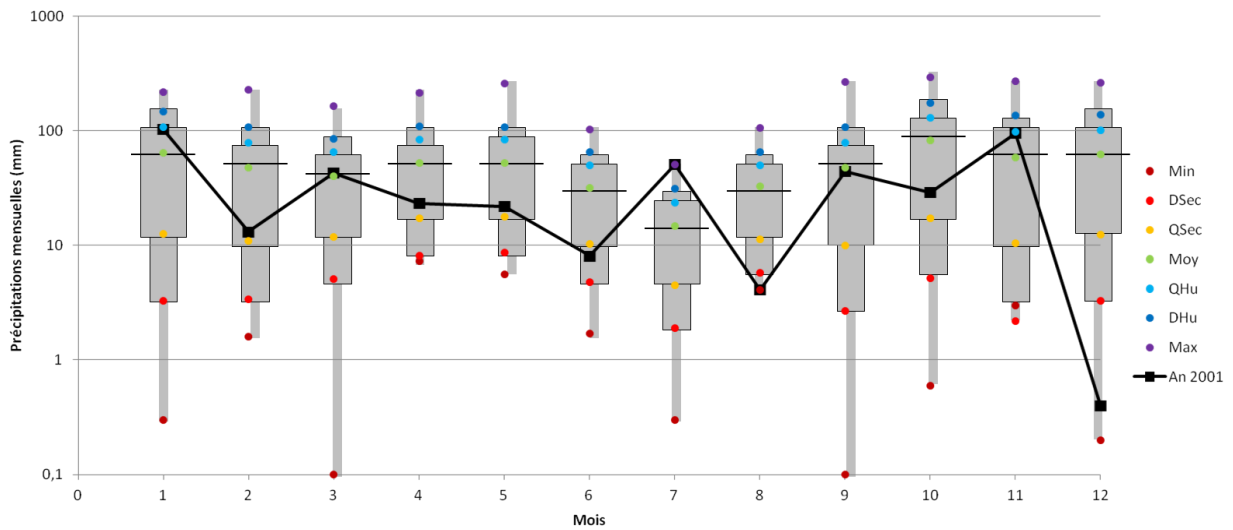


Figure V-10 : Boîte à moustaches pour l'analyse des chroniques météorologiques

V.D Conclusion du chapitre

Dans ce chapitre, nous avons proposé plusieurs modèles permettant de réaliser une évaluation des besoins en eau à la fois agricoles et domestiques. Ces modèles s'appuient sur des traitements simples de combinaison de données géométriques et de calculs mathématiques.

Le fait que les modèles adoptent la même architecture montre la généricité de notre approche. De plus, à l'exception des données météorologiques et pédologiques les données que nous utilisons font l'objet d'une libre diffusion sur internet. Elles sont par ailleurs souvent disponibles pour d'autres zones d'études, ce qui renforce le caractère générique de nos modèles.

Pour aller au-delà de l'estimation des besoins en eau, les modèles proposés peuvent permettre de réaliser un bilan hydrologique à l'échelle d'un bassin versant. Nous avons d'ailleurs indiqué les actions à mener pour réaliser cette tâche, et avons montré que l'exploitation du caractère géographique des données pouvait largement contribuer à comprendre l'affectation des quantités d'eaux prélevées aux entités consommatrices (parcelles irriguées, commune alimentée en eau potable).

De même nous avons introduit une méthode visant à caractériser les chroniques météorologiques. A défaut de disposer de données futures, elles permettent à un utilisateur d'identifier les réponses du modèle dans le cadre de conditions météorologiques caractéristiques (année quinquennale sèche).

Chapitre VI : Développement d'un prototype

Afin d'illustrer les concepts introduits, nous avons procédé au développement d'un prototype. Le présent chapitre détaille les fonctionnalités et caractéristiques de l'outil produit, ainsi que son mode de fonctionnement. Nous avons commencé par implémenter le modèle d'estimation des besoins en eau agricoles. L'ensemble des développements informatiques du prototype a été réalisé par une équipe d'ingénieurs d'ACATUS Informatique supervisée par l'auteur.

VI.A Une structure flexible

L'IDS FUSIEAU adopte une très classique architecture de type trois-tiers (Figure VI-1) comprenant :

- une partie client qui assure l'interface avec l'utilisateur et lui permet d'accéder aux ressources proposées par les serveurs ;
- une partie application qui fournit l'accès aux données ainsi qu'aux traitements ;
- une partie base de données qui contient l'ensemble des données, mesures et informations.

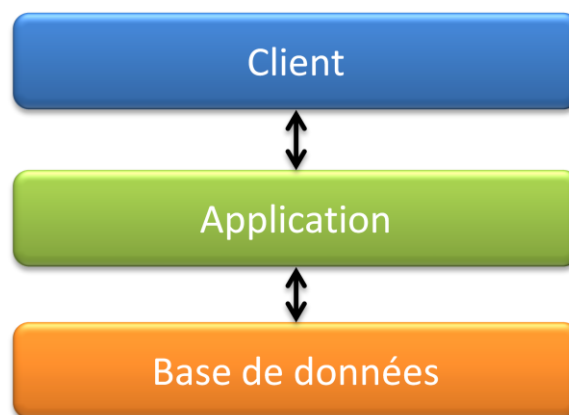


Figure VI-1 : Schéma de principe de l'architecture trois tiers

Pour le stockage des données et résultats nous nous appuyons sur le SGBD PostgreSQL (<http://www.postgresql.org/>), ainsi que sa cartouche spatiale PostGIS (<http://postgis.refractor.net/>).

Pour la partie client, une Interface Homme Machine (IHM) en Flex et Actionscript a été développée. Plusieurs bibliothèques ont été utilisées pour la conception de cette IHM :

- Openscales pour la partie cartographique (<http://openscales.org/>) ;
- AmCharts pour la représentation des données et résultats sous la forme de graphiques (<http://www.amcharts.com/>) ;

- Alivepdf pour l'export de données ou de résultats sous la forme de pdf (<http://alivepdf.byterray.org/>).

La partie application est composée de trois applications serveurs implémentant des standards de l'OGC :

GeoServer (<http://geoserver.org/>), une application serveur implémentant les standards WMS, WFS et WCS de l'OGC pour distribuer les données géographiques sous la forme de services web. Elle offre toutes les opérations obligatoires associées aux normes WMS, WFS et WCS. Nous détaillons ici celles du WFS, que nous utilisons plus particulièrement dans FUSIEAU :

- *GetCapabilities* pour demander l'autodescription du service ;
- *DescribeFeatureType* pour demander la description d'une donnée (nom des champs, emprise, projection utilisée) ;
- *GetFeature* pour consulter les données géographiques. Les résultats sont rendus sous la forme d'un document GML.

GeoServer implémente également le profil transactionnel du WFS (souvent abrégé WFS-T). Il permet au client de modifier les données. Trois fonctions sont disponibles :

- *Insert* pour ajouter une entité géographique ;
- *Update* pour modifier une propriété de la donnée (géométrie ou attribut) ;
- *Delete* pour supprimer une entité géographique.

52nSOS, une application serveur implémentant le standard SOS de l'OGC. Elle fonctionne conjointement avec une base de données PostgreSQL / PostGIS, et offre les trois opérations obligatoires de la norme SOS :

- *GetCapabilities* pour demander l'autodescription du service, comportant notamment la liste des capteurs enregistrés ;
- *DescribeSensor* pour demander la description d'un capteur (nom, position géographique, type de mesures enregistrées, période couverte par les mesures). Les résultats sont alors rendus sous la forme d'un document encodé en Sensor Model Language (SensorML) ;
- *GetObservation* pour consulter les mesures enregistrées par un capteur. Les résultats sont alors renvoyés sous la forme d'un document encodé en Observations&Measurements (O&M).

L'application implémente également le profil transactionnel (optionnel dans la norme SOS). Les fonctions suivantes sont alors disponibles :

- *RegisterSensor* pour inscrire de nouveaux capteurs dans la base de données ;
- *InsertObservations* pour alimenter des capteurs et enregistrer leurs mesures dans la base de données.

52nWPS, une application serveur implémentant la norme WPS. Elle permet la mise à disposition de services de traitements sur le web. 52nWPS offre les trois opérations obligatoires de la norme WPS :

- *GetCapabilities* pour demander l'auto-description du service, comportant notamment la liste des processus disponibles ;
- *DescribeProcess* pour demander la description d'un processus (nom, données et paramètres nécessaires pour lancer le traitement, formats autorisés en entrée, formats disponibles en sortie) ;
- *Execute* pour demander au serveur d'exécuter le traitement. Les résultats sont renvoyés dans le format spécifié par l'utilisateur dans sa requête d'exécution.

Les algorithmes associés aux processus mis à disposition sont des classes JAVA compilées et stockées dans un dossier auquel accède l'application. Plusieurs bibliothèques d'algorithmes « prêt à l'emploi » existent, l'une d'elles se nomme Sextante (<http://www.sextantegis.com/>). Nous l'utilisons pour réaliser le découpage de la zone d'étude décrit dans la partie V.A.1 de ce mémoire.

Le schéma suivant montre l'architecture interne simplifiée de l'IDS FUSIEAU (Figure VI-2) :

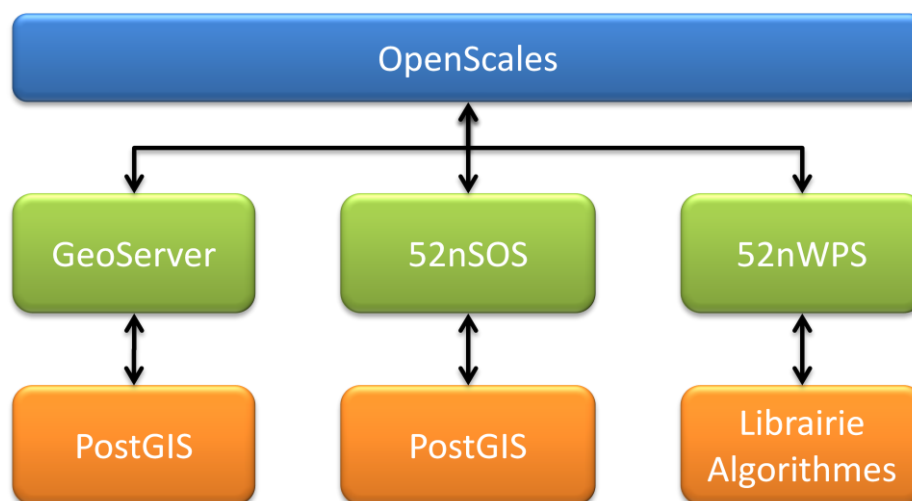


Figure VI-2 : Architecture simplifiée de l'IDS FUSIEAU

VI.A.1 Utilisation de FUSIEAU comme centre de traitement externalisé

Le principal intérêt de l'architecture trois tiers est sa flexibilité. Comme elle dissocie client et serveurs, différentes combinaisons sont envisageables.

Pour FUSIEAU, nous avons travaillé sur la possibilité de se connecter à d'autres serveurs d'applications que ceux implémentés dans notre IDS. L'utilisateur pourra par exemple connecter ses propres serveur WFS et SOS à FUSIEAU et faire que les résultats produits par les modèles soient stockés dans sa base de données personnelle (Figure VI-3).

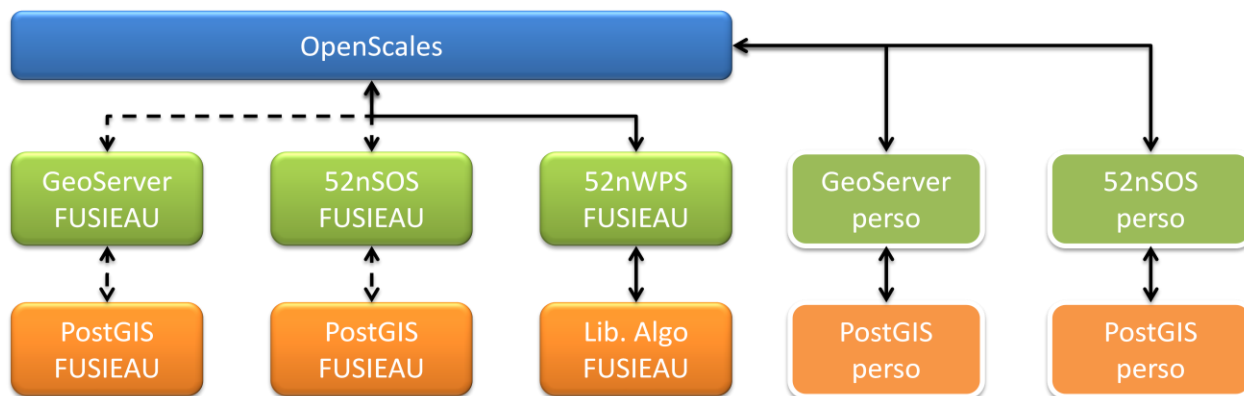


Figure VI-3 : Connexion de serveurs d'applications personnels à FUSIEAU

Cette approche constitue le principal avantage de l'architecture SOA, nous y voyons deux intérêts principaux :

- l'utilisateur maîtrise mieux les données qu'il produit : il est libre de les modifier et de les diffuser. Il peut les réutiliser dans ses propres applications sans devoir passer par FUSIEAU ;
- cela évite une saturation de la base de données associée à FUSIEAU.

Dans l'IHM, l'utilisateur peut ainsi préciser les serveurs WFS et SOS qu'il souhaite utiliser. Il doit également fournir les paramètres de connexion (login et mot de passe) nécessaires à la création et l'édition de données sur le serveur, ainsi que les paramètres de connexion à sa base de données PostGIS pour la sauvegarde des données dans son SGBD personnel.

Techniquement il est également possible de substituer le serveur WPS par un autre. Néanmoins cette possibilité est actuellement fortement contrainte par le fait que les deux algorithmes (celui d'origine sur le serveur WPS FUSIEAU et son remplaçant sur l'autre serveur) doivent utiliser les mêmes données en entrées, et proposer le même formalisme pour les données en sortie pour que le modèle puisse fonctionner.

Un autre intérêt d'avoir choisi une architecture orientée services est que les modèles sont indépendants du client. Les services peuvent être appelés individuellement et intervenir dans des chaînes de traitements autres que celles prévues par FUSIEAU.

VI.A.2 Traçabilité des résultats

Nous avons également fait le choix d'accompagner la réalisation d'une simulation par la production de métadonnées. Ces dernières indiquent les paramètres retenus par l'utilisateur pour sa réalisation. La génération de ces métadonnées est indispensable à la réutilisation des traitements et résultats produits par les divers utilisateurs. Elles indiquent la zone d'étude traitée, les données utilisées, les hypothèses formulées pour les simulations. Ces métadonnées sont actuellement stockées dans le SGBD et

uniquement accessible via l'IHM. Toutefois nous avons pour projet de les associer aux services produits et mis à disposition.

L'utilisateur peut consulter les paramètres choisis par les autres utilisateurs de l'application et vérifier si la simulation qu'il souhaite effectuer a déjà été faite. Si c'est effectivement le cas, il pourra alors directement en consulter les résultats, évitant ainsi l'attente liée au calcul des valeurs.

Pour les données stockées sur le SGBD associé à FUSIEAU, nous avons également intégré un paramètre de confidentialité sur les résultats. L'utilisateur peut alors choisir que les résultats de ses simulations soient rendus publics, et puissent donc être réutilisés par les autres utilisateurs ou bien les laisser confidentiels. Il sera alors le seul à pouvoir les réutiliser. Enfin pour chacun des résultats des simulations, une trace des paramètres ayant permis d'obtenir ces résultats est conservée et stockée dans la base de données.

VI.A.3 Conception modulaire

La conception de l'IDS FUSIEAU suit un développement de type modulaire tant du côté serveur, que du côté client. Les différents éléments conçus séparément fonctionnent ensuite ensemble comme un tout homogène.

La spécification OGC du WPS se prête bien à une conception modulaire puisqu'il s'agit seulement d'une interface pour l'accès aux traitements. Chacun de nos modèles de calcul a été décomposé en processus élémentaires. Nous distinguons deux catégories de processus :

- les processus dits « métiers » qui sont propres à un modèle, dans lesquels on trouve tous les algorithmes chargés de calculer les résultats ;
- les processus dits « techniques » qui servent à la manipulation de certains types de données. On y trouve des briques permettant la lecture d'un type de données particulier et de la récupération des informations utiles aux traitements ou encore des processus permettant la publication des résultats dans un format prédéterminé.

Du côté client, chacune des fonctionnalités de l'IHM est rattachée à un module. Le chargement de ce module engendre automatiquement l'ajout de la fonctionnalité à l'IHM. Le client permet l'exécution de l'ensemble des traitements qui compose le modèle et permet également de réaliser une simulation complète. L'IHM propose également différents types de représentations des données et résultats de sorte à en faciliter la lecture et l'analyse.

VI.B Implémentation des modèles

Les traitements développés dans le cadre de FUSIEAU sont de deux types :

- combinaison de géométries : regroupement ou dissociation en fonction de leurs attributs ;

- détermination de valeurs résultantes selon un calcul issu d'un des modèles implémentés.

VI.B.1 Processus associé à la combinaison de géométries

Les traitements combinant des géométries exploitent des données fournies sous la forme de standards WFS. Actuellement ces traitements sont spécifiques au modèle d'estimation des besoins en eau agricoles. Le modèle d'estimation des besoins en eau domestiques fonctionne quant à lui à l'échelle communale. Pour combiner les géométries, nous utilisons des algorithmes issus de la librairie d'algorithme Sextante :

L'algorithme implémenté en WPS réalise successivement les quatre actions suivantes (Figure VI-4) :

- le découpage des zones, selon le principe évoqué à la Figure V-2. Ce traitement engendre :
 - o la génération de nouvelles géométries à partir de celles des couches d'origine ;
 - o l'affectation des attributs. Les géométries filles héritent des propriétés de leurs deux couches génitrices ;
- le stockage des données produites dans PostGIS. Cette étape va permettre de pérenniser la donnée en la stockant dans un SGBD ;
- la publication sous la forme de WFS dans GeoServer. Il s'agit là de rendre les données accessibles. On fournit à GeoServer les paramètres utiles et nécessaires pour accéder à la base de données contenant les résultats.

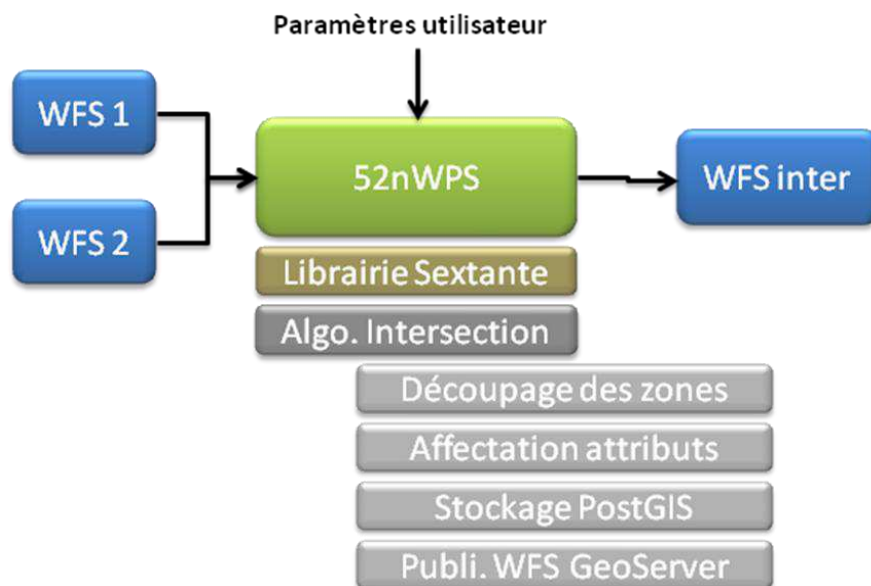


Figure VI-4 : Représentation schématique des traitements effectués pour la combinaison de géométries

L'envoi d'une requête de type *GetFeature* au serveur WFS chargé de publier la donnée permettra alors d'obtenir les géométries nouvellement créées.

Une fois que la zone d'étude et les mailles qui la constituent ont été créées, les modèles proposent d'estimer les besoins en eau (agricoles ou domestiques) pour ce territoire.

VI.B.2 Processus associé à un calcul dans le modèle

Le deuxième type de traitement porte sur le calcul des valeurs via le modèle. Les algorithmes ont été conçus pour lire des mesures exprimées sous le format O&M. Les données fournies à des dates communes peuvent être combinées et interviennent alors dans le calcul spécifique.

Selon le principe de création de capteurs virtuels évoqué dans la partie IV.C.1, les algorithmes de ce type comportent trois parties distinctes (Figure VI-5) :

- La création d'un capteur par le biais d'une requête SOS-Transactionnelle de type *RegisterSensor*. La zone d'intérêt se voit dotée d'un capteur dans lequel seront stockées toutes les mesures la concernant. Cette étape comprend également l'affectation de la géométrie de la zone d'intérêt comme attribut du capteur.
- Le calcul des résultats. Les valeurs sont calculées à partir des données et des paramètres choisis, et selon la formule associée au calcul.
- L'alimentation des capteurs par le biais d'une requête SOS-Transactionnelle de type *InsertObservation*. Les résultats du calcul sont affectés au capteur sous la forme de mesures.

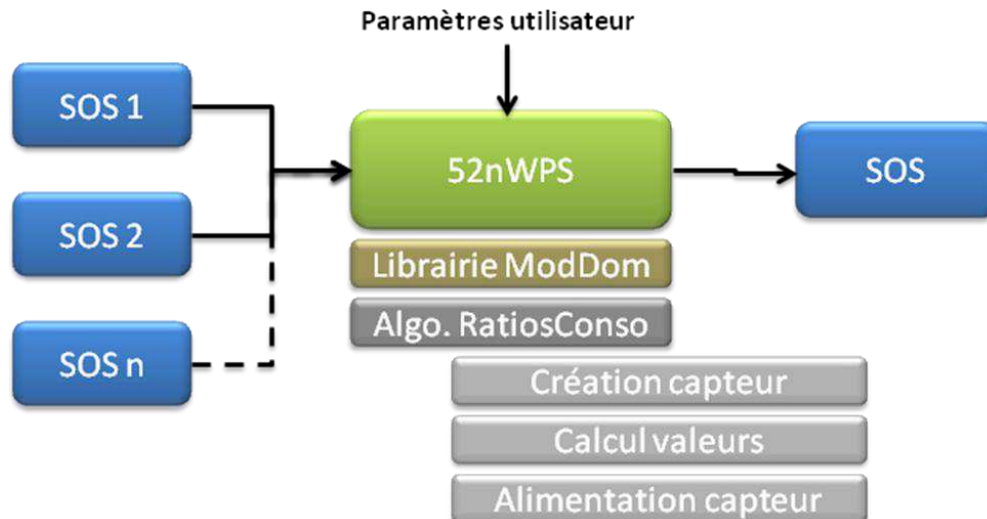


Figure VI-5 : Représentation schématique des traitements effectués pour les calculs associés au modèle

La consultation des résultats revient alors à l'interrogation du capteur auquel ont été affectées les mesures. Les géométries des zones couvertes par le capteur étant connues, on peut par la réalisation de requêtes appropriées, déterminer les capteurs présents.

VI.C Partie client

VI.C.1 Représentation des données et des résultats

L'interface cartographique que nous avons développée pour l'IDS FUSIEAU permet d'afficher des données issues de services WMS et WFS.

Nous avons développé un petit module qui permet d'interroger un serveur WMS ou WFS en exploitant les requêtes spécifiées par l'OGC : l'utilisateur commence par indiquer l'adresse d'un serveur WFS. Une requête *GetCapabilities* est alors envoyée au serveur qui en retour renvoie la liste des couches disponibles. L'utilisateur choisit ensuite une couche dans la liste. Une requête *GetFeature* est alors envoyée au serveur et la couche s'affiche sur la carte.

Par ailleurs, nous avons également souhaité représenter les données issues des serveurs SOS. S'agissant de séries temporelles, l'utilisation de graphiques s'est imposée comme une évidence. Ce format permet en effet de visualiser facilement l'évolution des valeurs en fonction du temps. La capture d'écran ci-dessous présente la représentation graphique des données météorologiques associées à une maille SAFRAN (Figure VI-6). Pour améliorer l'ergonomie de FUSIEAU, nous avons souhaité créer une interaction entre l'interface cartographique et ces graphiques. L'idée était d'afficher les chroniques associées à une zone d'étude en sélectionnant sa géométrie sur la carte.

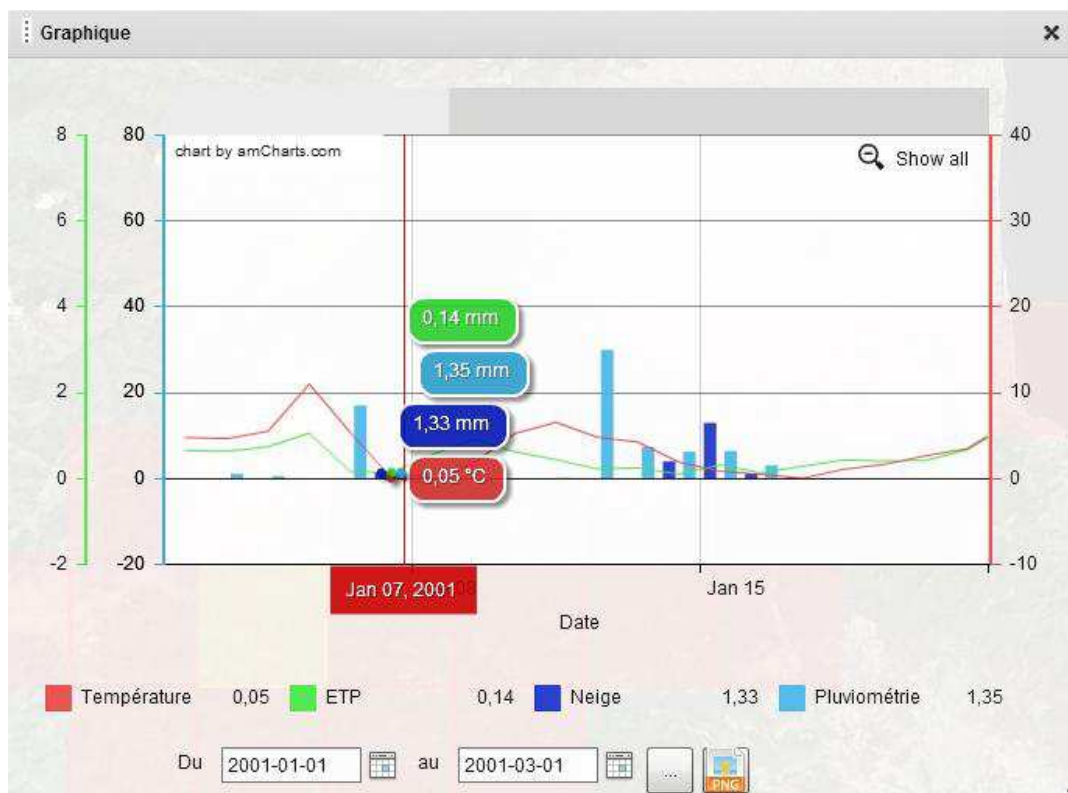


Figure VI-6 : Module graphique de FUSIEAU

Nous avons dès lors entrepris de coupler un service SOS avec un service WFS. Cette approche a été suggérée auparavant par [Jirka, 2010]. Pour obtenir ces graphiques nous avons simplement sélectionné une des mailles SAFRAN à partir de l'interface cartographique et choisi d'afficher les chroniques associées à cette zone d'étude.

A ce moment, l'application va parcourir les attributs de la couche WFS représentant les stations météo et récupérer l'identifiant du capteur SOS associé à la station météo. Puis dans un second temps, elle va interroger le serveur SOS pour récupérer les mesures associées au capteur par le biais d'une requête de type *GetObservation*. La suite n'est que la représentation graphique des données fournies par le SOS.

VI.C.2 Construction de scénario pas à pas

L'évaluation des besoins en eau agricoles selon le modèle FAO s'appuie sur la combinaison de multiples paramètres. La connaissance des caractéristiques de la zone d'étude retenue, permet de déterminer une grande partie de ces informations.

En effet, le remplissage des paramètres utiles au calcul d'un élément fonctionne selon un principe d'héritage depuis les algorithmes / fonctions situés en amont dans la chaîne de traitement. Le tableau suivant récapitule les paramètres utilisés pour le calcul de chacun des paramètres, ainsi que l'élément dont ils héritent (Tableau VI-1).

Par exemple, l'algorithme « CalculETc » nécessite deux couples variables et paramètres pour fonctionner :

- *UrlKc* correspond à la requête *GetObservation* qui permet d'obtenir les mesures relatives aux coefficients culturaux ;
- *UrlKcPosValue* indique le numéro de la mesure à récupérer (dans le cas de capteurs composites effectuant plusieurs types de mesures) ;
- *UrlETo*, correspond à la requête *GetObservation* qui permet d'obtenir les mesures relatives à l'évapotranspiration de référence ;
- *UrlEToPosValue* indique le numéro de la mesure à récupérer (dans le cas de capteurs composites effectuant plusieurs types de mesures).

Même si les champs sont déjà pré-remplis, l'utilisateur a la possibilité de les modifier manuellement. Les résultats et paramètres générés par l'utilisateur sont sauvegardés dans les préférences de l'utilisateur. Si l'utilisateur a décidé de laisser le scénario et ses résultats publics, ils seront également accessibles aux autres utilisateurs.

Nom paramètre	Hérité de	CalculKc	CalculEtc	CalculZr	CalculTaw	CalculPt	CalculDr	CalculKs	CalculEtcAdj
NameFOI	Zone d'étude	x							
FOI	Zone d'étude	x							
EPSG	Zone d'étude	x							
Id	Zone d'étude	x							
Inf1	Couche pédologique	x							
Inf1nd	Couche OCSOL	x							
Inf1ni2	Couche OCSOL	x							
Inf1mid	Couche OCSOL	x							
Inf1end	Couche OCSOL	x							
U2	Couche météo	x							
Rh1min	Couche météo	x							
H	Couche OCSOL	x							
Date Plantation	Couche OCSOL	x							
Date De bDev	Couche OCSOL	x							
Date De bGénescence	Couche OCSOL	x							
Date FinGénescence	Couche OCSOL	x							
Ur1ict	CalculKc	Résultat	x						
Ur1ictPosValue	CalculKc	Résultat	x						
Ur1Etp	Couche météo		x						
Ur1EtpPosValue	Couche météo		x						
Ur1Etc	CalculEtc		Résultat			x	x		x
Ur1EtcPosValue	CalculEtc		Résultat			x	x		x
Zr1ni	Couche OCSOL			x					
Zr1inter1	Couche OCSOL			x					
Zr1inter2	Couche OCSOL			x					
Zr1fin	Couche OCSOL			x					
Date 1ni	Couche OCSOL			x					
Date 1inter1	Couche OCSOL			x					
Date 1inter2	Couche OCSOL			x					
Date 1fin	Couche OCSOL			x					
Ur1Zr	CalculZr			Résultat	x				
Ur1ZrPosValue	CalculZr			Résultat	x				
ThetaFc	Couche pédologique				x				
ThetaWp	Couche pédologique				x				
Ur1Taw	CalculTaw				Résultat		x	x	
Ur1TawPosValue	CalculTaw				Résultat		x	x	
Pt	Couche OCSOL					x			
Ur1Pt	CalculPt					Résultat		x	
Ur1PtPosValue	CalculPt					Résultat		x	
Ur1P	Couche météo						x		
Ur1PPosValue	Couche météo						x		
Ur1i	Politique irrigation						x		
Ur1iPosValue	Politique irrigation						x		
Ur1Drt	CalculDr						Résultat	x	
Ur1DrtPosValue	CalculDr						Résultat	x	
Ur1Is	CalculKs							Résultat	x
Ur1IsPosValue	CalculKs							Résultat	x

Tableau VI-1 : Liste des paramètres utilisés dans la modélisation

VI.D Conclusion du chapitre

Au cours de ce chapitre, nous avons présenté l'implémentation du modèle d'estimation des besoins en eau agricoles défini au Chapitre V, selon le principe du « Model Web » défini au Chapitre II, ainsi que les concepts introduits au cours du Chapitre IV.

Le prototype d'IDS réalisé s'appuie sur une architecture trois tiers et utilise des serveurs d'applications implémentant les services WFS, SOS et WPS de l'OGC. Nous exploitons le potentiel offert par chacun de ces services web pour reproduire le modèle de la FAO.

Une interface cartographique visant à faciliter l'utilisation du modèle a été développée. Elle ne constitue toutefois pas le seul biais d'utiliser le modèle, dont les traitements et les données peuvent être utilisés indépendamment et par d'autres applications.

Troisième partie

-

Vers une IDS intelligente

Chapitre VII : Gestion des imperfections dans un cadre de fusion de données

Dans la deuxième partie de ce mémoire, nous avons traité de la mise à disposition de modèles d'estimations des besoins en eau selon le principe du « Model Web », évoqué au Chapitre II. Toutefois, dans le cadre de l'utilisation de modèles, nous avons vu dans le Chapitre III que la prise en compte des imperfections est importante. Aussi dans cette troisième partie, nous traitons de la gestion des imperfections dans notre modèle et plus particulièrement de son intégration dans l'IDS développée dans la deuxième partie du mémoire. Une approche similaire est déjà abordée par le projet UncertWeb, aussi nous proposons d'étudier plus en détail ce projet.

VII.A FUSIEAU : frère jumeau d'UncertWeb ?

VII.A.1 Des objectifs communs

En février 2010, a commencé le projet de recherche européen UncertWeb (<http://www.uncertweb.org/>), porté par un consortium de laboratoires et d'organismes de recherche, dont l'Université d'Aston (Angleterre) et l'Université de Munster (Allemagne).

UncertWeb fait suite à un autre projet européen, nommé INTAMAP (<http://www.intamap.org/>) achevé en 2009. Ce projet avait conduit à la définition du format UncertML, candidat pour devenir un standard de diffusion de l'information liée à l'incertitude [Williams & al, 2008]. A l'image du SensorML pour la description des capteurs, l'UncertML est un schéma XML, mais dédié à la description de l'incertitude [OGC, 2008].

Les objectifs du projet UncertWeb sont de proposer un modèle d'infrastructure répondant aux objectifs et contraintes du model web, et qui soit en plus capable de prendre en compte l'incertitude et la qualité des données en général [Bastin & al, 2012].

Les objectifs sont identiques à ceux de FUSIEAU, et techniquement les outils qui en découlent présentent logiquement de fortes similitudes, à savoir une architecture de type SOA et même WOA basée sur l'exploitation de services web (principalement ceux spécifiés par l'OGC). En particulier le SOS et le WFS pour la diffusion des données, ainsi que le WPS pour celle des modèles.

VII.A.2 Des approches différentes

Même si les objectifs d'UncertWeb, ainsi que l'architecture globale du système, sont très proches de FUSIEAU, on note tout de même des points de différence dans la réalisation.

La principale différence réside dans l'approche retenue pour la méthode de calcul et de propagation des incertitudes. UncertWeb aborde l'incertitude selon une approche probabiliste. Le projet s'intéresse aux probabilités d'occurrence d'un évènement au regard de probabilités d'occurrence d'autres évènements. Selon cette approche, le calcul de l'incertitude des résultats se fait par la propagation des fonctions de répartition des densités de probabilité des valeurs d'entrée. Pour se faire UncertWeb s'appuie sur le principe des simulations de Monte-Carlo.

D'autre part, UncertWeb traite de l'utilisation d'un format dédié : l'UncertML, pour l'expression de l'incertitude. La promotion du format UncertML et l'établissement de ce format comme standard constituent même une des principales missions du projet UncertWeb. Dans FUSIEAU, nous n'avons pas jugé nécessaire de recourir à un format particulier pour l'expression de l'incertitude. Nous considérons que le résultat du calcul d'incertitude constitue une observation à part entière, tout comme celle de la mesure principale. Résultat et incertitude sont ainsi implémentés de la même manière.

D'un point de vue technique ensuite, la solution développée pour UncertWeb utilise le langage BPEL pour le chaînage des services. Or nous avons expliqué à l'issue du Chapitre II, notre souhait de ne pas s'appuyer sur cette technologie.

De plus, au chaînage entièrement automatisé géré par BPEL, nous préférons une approche pas-à-pas où l'utilisateur vérifie les résultats obtenus à l'issue de chaque traitement et valide le passage à l'étape suivante. Cette méthode évite que l'utilisateur ait à attendre la fin du chaînage de tous les processus s'il se rend compte rapidement d'une erreur dans un traitement.

VII.B Typologie des imperfections

VII.B.1 Imperfections liées aux données

En nous appuyant sur la typologie des imperfections introduite par [De Runz, 2008], ainsi que la typologie des erreurs introduite par [Fisher, 2005] nous avons commencé par identifier et caractériser l'ensemble des imperfections pouvant affecter notre jeu de données. Cette étape est importante car les méthodes à mettre en place pour prendre en compte les diverses imperfections vont différer en fonction du type d'imperfection qu'on aura à traiter.

Le tableau ci-dessous (Tableau VII-1) présente les différentes imperfections sur les données pouvant affecter le modèle d'estimation des besoins en eau agricole.

Type de données	Biais	Raison	
Climatiques	Erreur de généralisation spatiale	La mesure associée à la maille SAFRAN est différente de la mesure locale	
	Erreur de mesure	La mesure faite par l'appareil est fautive	
	Erreur de traitement	Erreur lors du calcul (détermination d'ETO)	
Phénologiques	Fidélité du modèle	Les valeurs des coefficients Kc de la plante ne correspondent pas à ceux du modèle	
	Fidélité du modèle	Le cycle de développement de la plante ne correspond pas à celui du modèle	
Pédologiques	Erreur de généralisation spatiale	L'échelle de représentation des données ne permet pas de rendre compte de la pédologie locale	
	Erreur de généralisation spatiale	Les limites des zones sont imprécises	
	Erreur temporelle	La texture du sol a changé depuis la date du relevé	
	Erreur de classification	Le sol n'a pas été classé dans la bonne catégorie	
	Erreur de mesure	Les mesures faites pour classer le sol sont erronées	
Occupation du sol	Erreur de généralisation spatiale	L'échelle de représentation des données ne permet pas de rendre compte de l'occupation du sol locale	
	(CLC)		
	(CLC & RPG)	Erreur de généralisation spatiale	Les limites des zones sont imprécises
	(CLC & RPG)	Erreur de mesure	Erreur lors de l'acquisition des données
	(CLC & RPG)	Erreur de classification	L'occupation du sol n'a pas été classée dans la bonne catégorie
	(CLC & RPG)	Erreur temporelle	L'occupation du sol a changé depuis la date du relevé
	(CLC & RPG)	Ambiguïté	Non-spécificité de la nomenclature (ex : Cultures complexes)
	(RPG)	Incomplétude	Parcelles agricoles absentes du RPG
Prélèvements en eau	Granularité temporelle trop faible	La fréquence des mesures n'est pas suffisante (pas de temps annuel)	
	Erreur de mesure	Erreur lors de l'acquisition des données	
Politique d'irrigation	Fidélité du modèle	La politique d'irrigation simulée ne correspond pas à la réalité	

Tableau VII-1: Typologie des biais pouvant affecter la modélisation

Pour illustrer cette approche, nous détaillons ici les résultats de l'influence de la non-spécificité de la nomenclature sur la détermination des besoins en eau d'une plante.

Les données d'occupation du sol utilisées dans FUSIEAU (Corine Land Cover, Registre Parcellaire Graphique) catégorisent les différents usages du sol. La nomenclature de ces données distingue les types de cultures, toutefois elle ne va pas toujours jusqu'à identifier précisément la culture pratiquée. C'est en particulier le cas pour la catégorie « vergers » qui regroupe l'ensemble des arbres fruitiers. Dans le département des Pyrénées-Orientales où les vergers représentent une part importante de l'agriculture, nous avons souhaité évaluer l'impact de la méconnaissance des fruits cultivés sur une parcelle.

Notre connaissance de la zone d'étude nous a amené à réduire les hypothèses de fruits cultivés à trois possibilités : pêches, abricots ou cerises. Les pêchers, abricotiers et cerisiers n'ont pas les mêmes besoins en eau. Surtout leurs cycles de développement sont décalés. Cela se traduit par des coefficients culturaux différents. Le graphique ci-dessous (Figure VII-1) présente les courbes de coefficients culturaux des abricotiers, pêchers et cerisiers. Les valeurs sont issues de la publication 56 de la FAO [Allen & al, 1998].

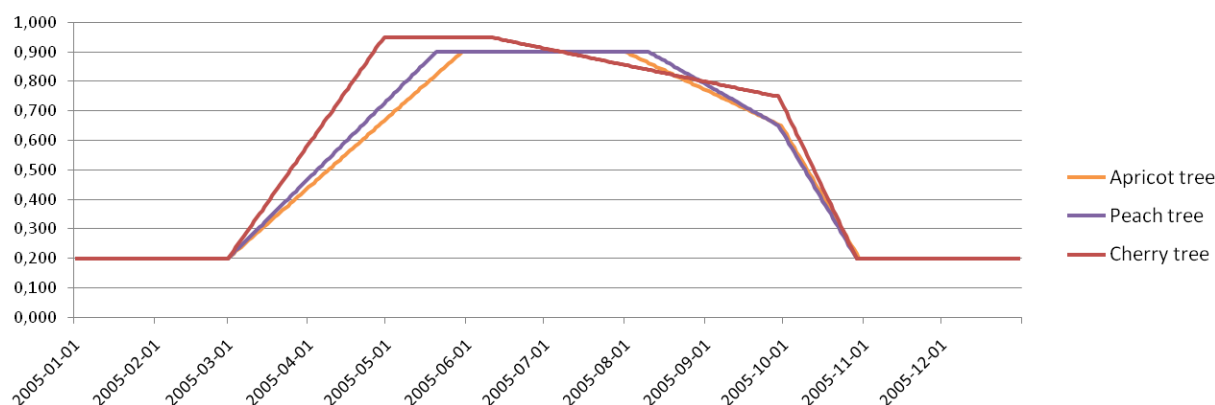


Figure VII-1 : Coefficients culturaux des abricotiers, pêchers et cerisiers

Nous remarquons que la période de forte demande en eau du cerisier intervient plus tôt dans l'année que celle des pêchers et abricotiers. Toutefois ses fruits arrivent également à maturité plus tôt. A contrario, les besoins en eau des pêchers et abricotiers se manifestent principalement pendant l'été.

Comme les vergers sont des cultures irriguées systématiquement, nous nous intéressons ici à la comparaison des besoins en eau dans les conditions optimales de culture (ETc). Nous nous appuyons sur les données météo de l'année 2005, qui s'apparente à une année quinquennale sèche pour le département des Pyrénées-Orientales, ainsi que sur les coefficients culturaux indiqués par la FAO. La zone d'étude retenue se trouve à proximité de Céret, sur le bassin du Tech.

Le graphique ci-dessous (Figure VII-2) présente la comparaison jour par jour des besoins en eau des abricotiers, cerisiers et pêchers. Le tableau suivant (Tableau VII-2) synthétise les principaux chiffres à retenir de cette comparaison.

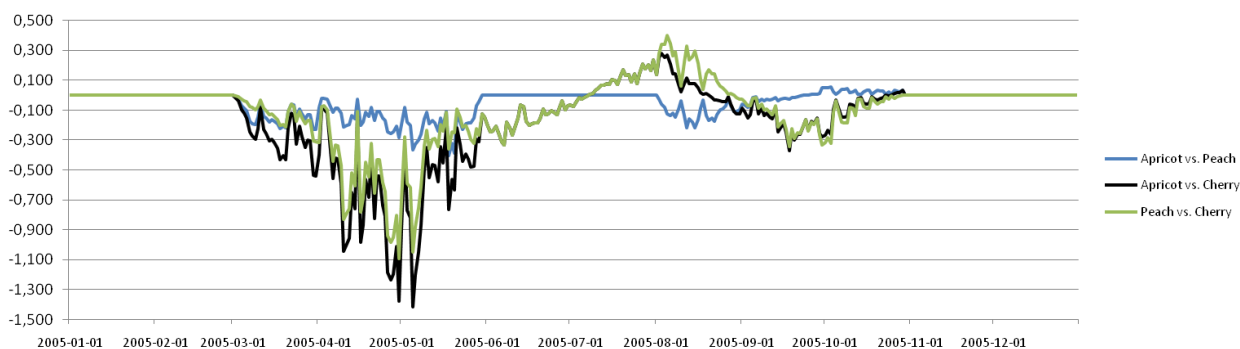


Figure VII-2 : Comparaison des besoins journaliers en eau des abricotiers, cerisiers et pêchers
[Beaufils & al, 2011 bis]

	Abricotier	Pêcher	Cerisier	Abricotier vs. Pêcher	Abricotier vs. Cerisier	Pêcher vs. Cerisier
Besoins en eau annuel (mm)	668.8	656.4	722.4	-17.6 (2.7%)	-53.6 (7.4%)	-36.0 (5.0%)
Parcelles étudiées (m3)	127072	124716	137256	-3344	-10184	-6840
Besoins en eau journalier moyen (mm)	1.83	1.88	1.98	-0.05 (2.7%)	-0.15 (7.4%)	-0.10 (5.0%)
Parcelles étudiées (m3)	347.7	357.2	376.2	-9.5	-28.5	-19
Besoins en eau journalier maximum (mm)	4.092	4.459	5.510	-0.367 (8.2%)	-1.418 (25.7%)	-1.051 (19.1%)
Parcelles étudiées (m3)	777.48	847.21	1046.9	-69.73	-269.42	-199.69

Tableau VII-2 : Principaux chiffres issus de la comparaison des besoins en eau des abricotiers, cerisiers et pêchers [Beaufils & al, 2011 bis].

Cette étude montre que sur une année entière, les écarts de valeurs sont faibles (aux alentours de 5%). En effet à l'échelle annuelle les différences de consommation des arbres fruitiers se compensent. Toutefois à des pas de temps plus courts, des différences sont nettement visibles. Ainsi au pas de temps journalier, l'écart constaté peut être important et représenter jusqu'à 25% d'erreur. La connaissance précise des cultures pratiquées sur le terrain est donc primordiale pour l'estimation des besoins en eau d'une zone d'étude donnée.

VII.B.2 Imperfections liées au modèle

Le fonctionnement de nos modèles s'appuie sur l'exploitation de données diffusées à différentes échelles temporelles et spatiales. Ces hétérogénéités sont autant de contraintes qui doivent être prises en compte dans nos modèles.

VII.B.2.a 1 - Influence de la granularité temporelle

Le modèle FAO peut être utilisé pour estimer les besoins en eau à plusieurs fréquences : journalière, décadaire et mensuelle. Le choix de la granularité temporelle du modèle va influencer les résultats.

D'une manière générale, les résultats du modèle seront toujours exprimés à un pas de temps supérieur ou égal à celui des données. Par exemple, si nous ne disposons que de données météorologiques mensuelles il ne sera pas possible d'obtenir une estimation des besoins en eau décennaux, et encore moins quotidiens. Nous aurons au mieux une estimation des besoins mensuels, voire annuels.

Par ailleurs, plus le pas de temps de raisonnement est faible, plus l'application sera apte à retranscrire finement les besoins en eau des cultures. Le graphique ci-dessous (Figure VII-3) montre deux méthodes de répartition d'une même quantité de précipitation mensuelle. Le scénario 1 montre une répartition équilibrée des précipitations, tandis que le scénario 2 montre une période de sécheresse marquée par un unique épisode pluvieux mais violent.

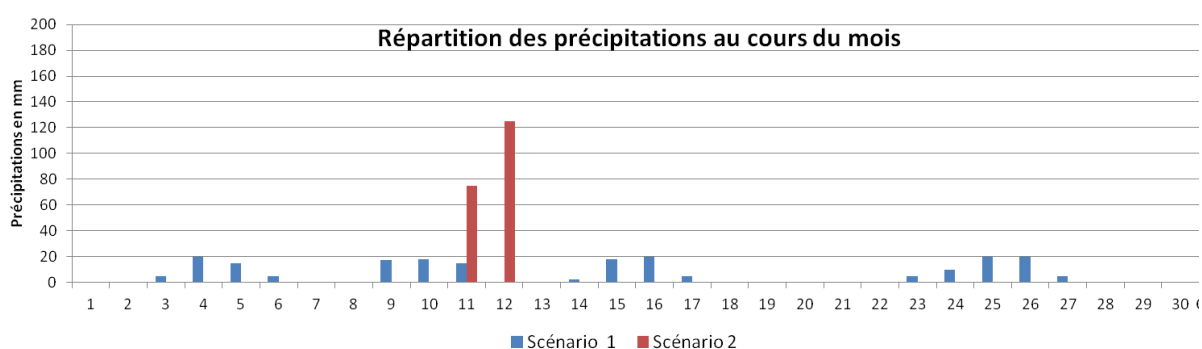


Figure VII-3 : Illustration de deux modes distincts de répartition d'un même cumul de précipitation

Une application raisonnant à l'échelle mensuelle ne fera pas la distinction entre ces deux cas de figure. Or les besoins en eau qui en résultent diffèrent totalement. Le scénario 1 correspond à un apport d'eau régulier qui va contribuer efficacement à satisfaire les besoins en eau des cultures, tandis que le scénario 2 risque d'entraîner la saturation de la réserve en eau des cultures sans pour autant assurer une bonne irrigation.

Pour ces raisons, il est toujours souhaitable de travailler avec le plus faible pas de temps proposé par les données.

VII.B.2.b 2 - Influence de la taille des mailles spatiales dans le calcul

Pour l'estimation des besoins en eau des cultures, le modèle FAO prend en compte les différentes combinaisons possibles des caractéristiques pédologiques, météorologiques et également d'occupation du sol.

Tout comme pour la granularité temporelle, les résultats du modèle seront toujours exprimés à une échelle inférieure ou égale à celle des données. Dans notre approche pour spatialiser le modèle FAO, nous avons introduit la détermination des mailles de calculs par combinaison des géométries des couches thématiques.

La simplification de la zone d'étude est un moyen efficace pour réduire le nombre de cas à traiter. Cette action doit toutefois être réalisée avec précaution afin de ne pas trop s'éloigner de la réalité. La pratique préalable d'une analyse de sensibilité est un bon moyen pour déterminer ce qui peut être simplifié ou non.

VII.C Prise en compte des imperfections

Dès lors que nous avons identifié les imperfections pouvant affecter un modèle, trois possibilités s'offrent à nous [Martin, 2005] :

- la première consiste bien sûr à chercher à éliminer ces imperfections ;
- la deuxième consiste à les tolérer ;
- enfin la troisième consiste à les modéliser.

Chercher à éliminer les imperfections est bien entendu la solution la plus tentante. Toutefois cela présuppose que les erreurs soient bien connues. Et a priori si cela avait été le cas, les données auraient d'ores et déjà été corrigées. Tolérer les imperfections, n'est pas aussi simple qu'il puisse y paraître. Il ne s'agit absolument pas de les ignorer. Au contraire, tolérer des imperfections nécessite une bonne connaissance de ces dernières. L'idée est de s'assurer que leur portée est suffisamment faible pour pouvoir être absorbée par le modèle. Pour identifier les imperfections susceptibles d'être négligées dans un modèle, on utilise des méthodes d'analyse de sensibilité comme décrit au Chapitre III.

Enfin lorsque les imperfections n'ont pu être corrigées, et ne peuvent être absorbées par le modèle, le dernier recours consiste à les modéliser. Ensuite pour connaître les imperfections des résultats, il est nécessaire de propager celles des données au travers des algorithmes. Dans FUSIEAU, nous appliquons les deux dernières techniques. Dans un premier temps, nous réalisons une analyse de sensibilité sur notre modèle. Puis dans un second temps, nous propageons uniquement les imperfections qui ont un impact non-négligeable sur la modélisation (Figure VII-4).

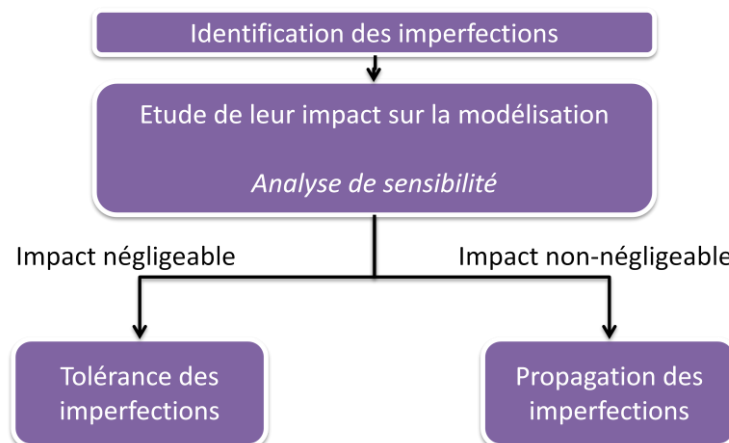


Figure VII-4 : Approche de prise en compte des imperfections dans FUSIEAU

VII.C.1 Analyse de sensibilité

Comme évoqué dans la partie III.B.2.a, pour la réalisation d'une analyse de sensibilité la connaissance des incertitudes sur les variables d'entrée n'est pas indispensable. L'utilisateur peut définir lui-même une perturbation à appliquer à chacune des entrées, et analyser leurs impacts sur les résultats.

Les résultats d'une analyse de sensibilité peuvent être utilisés dans plusieurs optiques. Dans FUSIEAU, nous l'utilisons pour simplifier le modèle. Cette opération se fait en deux temps :

- par réduction du nombre de variables à traiter ;
- par regroupement des parcelles ayant des propriétés proches.

Le but visé est principalement de limiter les échanges de données entre les clients et les serveurs, contrainte que nous impose le contexte web.

VII.C.1.a 1 – Réduction du nombre de variables à traiter

Si l'on se rend compte que certains paramètres ont un impact négligeable sur les résultats (les limites de ce qui est négligeable ou ne l'est pas sont à définir au cours de l'étude), alors on pourra envisager de ne pas prendre leur variabilité en compte et simplifier ainsi le modèle, sans pour autant engendrer un biais trop préjudiciable. Nous aurons ainsi moins de variables à gérer lors de nos calculs, ce qui ne peut avoir qu'un effet bénéfique sur les temps de traitement.

Afin de nous rendre compte de cette possibilité de simplification, une analyse de variabilité est menée. Connaissant la plage des valeurs possibles pour chacun des paramètres d'entrée, on cherche à en déduire la plage possible du résultat. Sa mise en œuvre nécessite donc de disposer des valeurs extrêmes des variables d'entrée. Nous détaillons l'implémentation de cette méthode dans le Chapitre VIII.

Ensuite pour pouvoir comparer les impacts de la variabilité des différentes entrées il est nécessaire de normaliser leur expression. L'incertitude obtenue est rapportée à la valeur de la variable. On parle alors d'incertitude relative.

VII.C.1.b 2 – Regroupement des parcelles ayant des propriétés proches

Une autre approche, qui convient particulièrement à notre modèle d'estimation des besoins en eau agricoles est de regrouper les cas présentant des caractéristiques proches. Le principe retenu dans ce modèle repose en effet sur un calcul des besoins en eau pour chacune des parcelles obtenues par le croisement des couches pédologiques, météorologiques et d'occupation du sol. Cela fait donc un nombre conséquent de paramètres qui interviennent dans les calculs.

Dans FUSIEAU, les parcelles présentant exactement les mêmes propriétés sont regroupées car selon les hypothèses formulées dans notre modèle, elles auront les mêmes besoins en eau. Cette méthode permet de réduire le nombre de cas à traiter, néanmoins nous souhaitons aller plus loin.

Notre idée est d'étendre le principe de regroupement aux parcelles qui à défaut de posséder exactement les mêmes paramètres, présentent de fortes similarités. Nous formulons ainsi l'hypothèse que les comportements de ces parcelles agricoles en termes de besoins en eau seront très proches. Cette méthode nous permet de discrétiser notre échantillon.

Afin d'appliquer cette technique de regroupement, il est nécessaire de connaître l'ensemble des cas possibles. Pour reprendre des termes statistiques, il s'agit de définir l'univers des possibles. Imaginons que nous ayons trois types de cultures (C1, C2, C3), et deux types de sol (S1, S2) sur notre zone d'étude, il y a donc six cas possibles : (C1S1, C1S2, C2S1, C2S2, C3S1, C3S2). Des informations sur le nombre d'occurrence de chacun des cas possibles dans notre zone d'étude peuvent également nous permettre de déterminer quels regroupements réaliser. Pour cela nous effectuons un simple traitement de comptage des occurrences dans la couche.

L'identification des parcelles pouvant être regroupées se fait à partir de méthodes d'analyse de sensibilité. Nous ne regroupons en effet que les parcelles dont les différences ne portent pas sur des paramètres sensibles. On aura toutefois une résolution sémantique plus faible.

L'exemple ci-dessous (Figure VII-5) présente les possibilités de regroupement de trois parcelles :

- s'il s'avère que les pêchers cultivés sur un sol sablo-argileux ont un comportement identique à ceux cultivés sur sol limoneux-argileux, alors on pourra regrouper les parcelles A et C dans une classe « sol argileux » ;
- s'il s'avère que les abricotiers et pêchers cultivés sur un sol sablo-argileux ont un comportement identique, alors on pourra regrouper les parcelles B et C dans une classe « abricotiers et pêchers sur sol sablo-argileux » ;
- s'il s'avère que les deux hypothèses précédentes sont vraies, on pourra regrouper les trois parcelles ensembles dans une classe « abricotiers et pêchers sur sol argileux ».

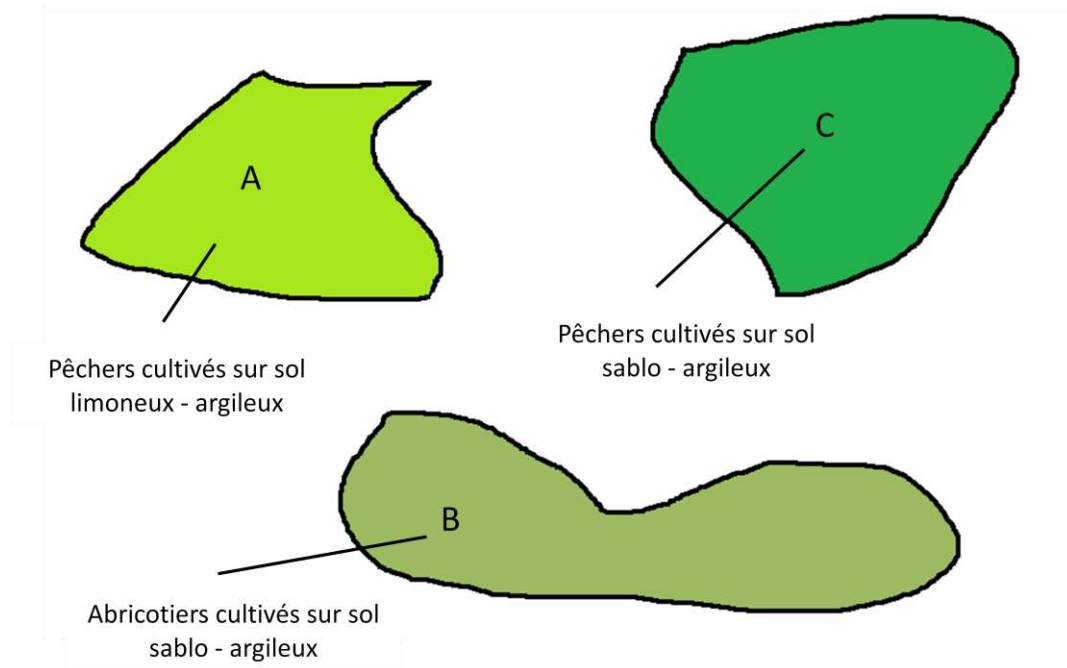


Figure VII-5 : Regroupement des parcelles en fonction de leurs caractéristiques

L'analyse de sensibilité permet à l'utilisateur d'orienter ses choix de regroupement. Réduire le nombre de cas à traiter est bien entendu le but visé par cette démarche, mais l'utilisateur doit néanmoins s'assurer de ne pas trop simplifier son modèle, sans quoi il ne sera plus fidèle aux phénomènes réels.

VII.C.2 Propagation des incertitudes

VII.C.2.a 1 – Le choix de l'approche déterministe

Dans FUSIEAU, nous avons retenu l'approche déterministe pour la propagation des incertitudes, plutôt que la méthode dite de Monte-Carlo. Notre choix s'appuie sur trois principaux arguments :

Le premier argument est que les calculs réalisés dans nos modèles sont relativement simples. De ce fait l'obtention des coefficients de sensibilité n'est pas problématique car les fonctions intervenant dans nos algorithmes sont facilement dérivables. Les fonctions de propagation obtenues sont rattachées au modèle et ne nécessitent pas d'être recalculées tant que le modèle reste inchangé.

Le deuxième argument concerne le contexte de notre projet. S'agissant d'une application web, il est nécessaire de limiter les transferts de données et d'informations entre serveurs et clients. La méthode de Monte-Carlo nécessite plusieurs centaines de simulations pour pouvoir fournir un résultat fiable. La réalisation de ces traitements demande une forte puissance de calcul, et ne nous semble pas adaptée au contexte web où l'application est rendue accessible à plusieurs utilisateurs à la fois.

Enfin le troisième argument porte sur la disponibilité des informations relatives à la qualité des données. En effet, bien que les organismes diffuseurs de données soient conscients de l'importance de

diffuser des informations relatives à la qualité de leurs données, force est de constater qu'aujourd'hui, il est très difficile d'accéder à ces ressources. Or disposer de ces informations est primordial pour pouvoir prétendre obtenir une estimation de l'incertitude associée aux résultats. Selon la méthode de propagation retenue, les besoins diffèrent au niveau de la nature des informations à fournir. L'approche déterministe ne nécessite que de simples informations : la plage de variation de chacune des variables d'entrée. Tandis que l'approche probabiliste requiert des informations plus pointues telles que les fonctions de distribution des probabilités. Or l'obtention de ces informations peut se révéler problématique. Notre choix se porte ainsi sur la méthode la moins exigeante.

Aussi compte tenu des connaissances limitées sur l'incertitude caractérisant les données, il nous paraît plus judicieux de s'appuyer sur une méthode d'approximation.

VII.C.2.b 2 - Mise en œuvre de l'approche déterministe

Si on assimile le calcul à une fonction mathématique f , le principe de la décomposition en séries de Taylor est d'approximer la fonction f par une fonction polynomiale qui localement constitue une bonne approximation du résultat. D'une manière générale, plus l'ordre de la fonction polynomiale augmente, plus l'approximation sera proche de la courbe réelle. De même, plus l'intervalle de validité de l'approximation sera étendu.

Le graphique ci-dessous (Figure VII-6) montre l'approximation de la courbe représentative du calcul (en pointillés noir). On note que pour cette fonction l'approximation linéaire, c'est-à-dire l'approximation de Taylor d'ordre 1 (en bleu) n'est valable qu'au très proche voisinage du point A. Au contraire, la courbe représentative de l'approximation d'ordre 9 épouse la courbe originale sur un intervalle de valeurs bien plus grand.

Comme nous l'avons vu au Chapitre III, pour une approximation linéaire de la fonction, les coefficients directeurs de la fonction d'approximation sont égaux aux coefficients de sensibilité obtenus lors de l'analyse de sensibilité.

Aux ordres supérieurs, il est nécessaire de dériver à nouveau les résultats obtenus à l'ordre précédent et ainsi de suite jusqu'à obtenir l'ordre souhaité. A partir de l'ordre 2 se pose la question de la corrélation des paramètres entre eux et de leur signification.

Par ailleurs, comme les calculs s'appuient sur des mesures prises à des pas de temps réguliers, il n'y a pas de corrélation entre les événements de la date t et de la date $t+1$ (à l'exception de celui du calcul de la réserve en eau quotidienne disponible). Notre approche consiste donc à estimer la propagation de la variance à chaque instant t .

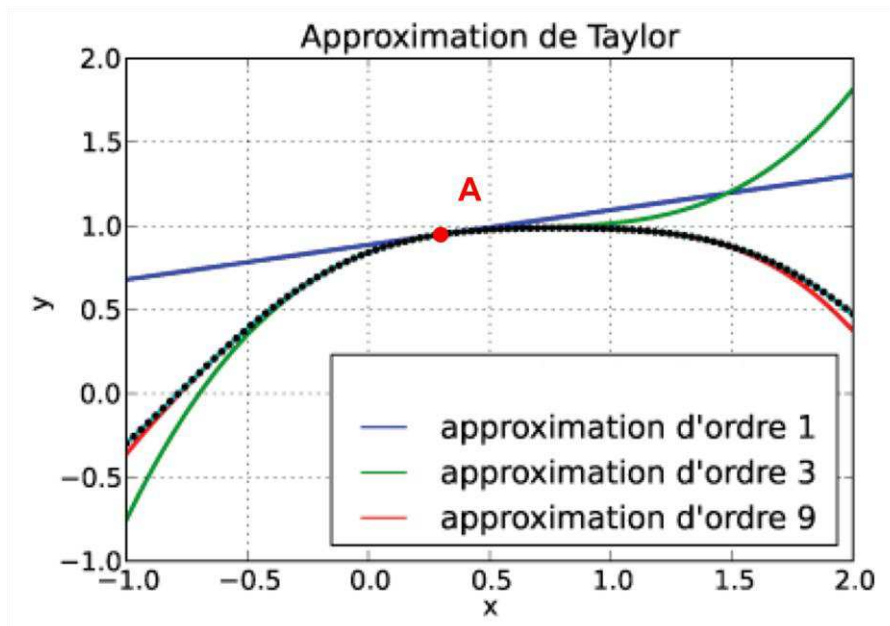


Figure VII-6 : Impact de l'ordre du polynôme de la fonction d'approximation

Dans le cadre de FUSIEAU, nous nous contentons d'approximations linéaires. De plus, comme les calculs réalisés dans nos algorithmes sont simples, nous avons pris l'initiative de dériver les fonctions à la main et les avons ensuite implémentées de manière informatique. Cette méthode ne saurait toutefois être retenue pour des modèles plus complexes, aussi la possibilité d'exploiter une méthode de dérivation automatisée serait intéressante et permettrait d'étendre le champ des algorithmes pouvant être pris en compte.

VII.D Conclusion du chapitre

Dans ce chapitre, nous avons présenté notre approche quant à la prise en compte de l'incertitude dans nos modèles et l'avons confrontée à celle retenue dans le cadre du projet UncertWeb.

Nous avons tout d'abord étudié l'impact d'une des imperfections affectant nos données, et avons ainsi illustré l'intérêt d'étudier la prise en compte les imperfections dans nos modèles.

Nous avons ensuite introduit une méthode basée sur des fonctionnalités d'analyse de sensibilité pour procéder à des groupements de parcelles et ainsi réduire le nombre de parcelles dont il faut estimer le besoin en eau.

De plus, nous avons également proposé de propager l'incertitude associée aux données et paramètres par le biais d'une méthode analytique basée sur la décomposition en séries de Taylor. Cette méthode a été retenue car elle s'intègre bien dans le contexte web, contrairement à la méthode dite de Monte-Carlo qui requiert que les calculs soient effectués un très grand nombre de fois pour produire une estimation de l'incertitude associée au résultat.

Chapitre VIII : Définition d'une méthode d'intégration de la prise en compte des imperfections dans l'IDS

Ce chapitre détaille les modifications à apporter pour la prise en compte des imperfections au sein de l'IDS. Nous nous intéressons tout particulièrement à l'intégration de ces modules spécifiques dans l'architecture du prototype. Deux problématiques se posent alors : l'interaction des informations relatives à l'incertitude avec le modèle, ainsi que les modifications à apporter à l'IHM actuelle pour sa prise en compte.

VIII.A Interaction avec le modèle

La prise en compte de l'incertitude dans le modèle impacte à la fois les traitements et les résultats produits. Cela nous amène à étudier ces deux cas :

- la façon dont doivent être stockées les informations relatives à l'incertitude des données ;
- la façon d'intégrer la prise en compte de l'incertitude dans le modèle sans engendrer de profondes altérations sur les algorithmes précédemment développés.

La prise en compte de l'incertitude doit être un « plus » et non une contrainte pour l'utilisateur.

VIII.A.1 Organisation des résultats avec prise en compte de l'incertitude

Les coefficients de sensibilité expriment la sensibilité du résultat au regard des différents paramètres. Leurs valeurs dépendent de celles des autres paramètres, de ce fait les coefficients de sensibilité sont propres à une simulation. Il en va de même pour les valeurs des incertitudes qui sont elles aussi associées à une mesure.

Les valeurs d'incertitudes et les coefficients de sensibilité sont ainsi corrélés aux valeurs calculées. En effet, si une mesure non assortie d'indications sur sa qualité n'a que peu de pertinence, une valeur d'incertitude seule n'a quant à elle aucun intérêt si elle n'est pas rattachée à un contexte.

Aussi, compte tenu du lien étroit entre les coefficients de sensibilité, les valeurs d'incertitude et les mesures elles-mêmes il nous apparaît indispensable de ne pas dissocier mesures et informations relatives à la qualité de la mesure.

Par conséquent, notre approche vise à stocker toutes ces informations dans le même capteur. Le capteur se voit ainsi doté de plusieurs types de mesure. La juxtaposition des mesures, valeurs

d'incertitudes et coefficients de sensibilité met en valeur le lien fort entre ces données. Le couplage des valeurs d'incertitudes et des coefficients de sensibilité avec les valeurs mesurées permet également d'éviter des mauvaises manipulations de données : utilisation des données d'incertitude relatives à une simulation dans une autre simulation.

L'encadré ci-dessous présente la partie résultat brute de la réponse à une requête *GetObservation* portant sur le capteur dans lequel sont stockées les mesures calculées. L'exemple porte sur un capteur contenant les valeurs d'évapotranspiration d'une plante (ETc), calculées à partir de la formule indiquée à la Figure I-5. Pour chacune des dates de la simulation, on retrouve dans l'ordre :

- la date à laquelle a été faite la mesure (en noir) ;
- l'identifiant de la parcelle à laquelle est rattachée la mesure (en violet) ;
- la valeur calculée, ici ETc (en rouge) ;
- le coefficient de sensibilité du résultat par rapport à la première variable, ici Kc (en bleu) ;
- le coefficient de sensibilité du résultat par rapport à la seconde variable, ici ETp (en orange) ;
- l'écart-type associée à la valeur calculée (en vert).

```

...
<om:result>
  <swe:DataArray>
    <swe:elementCount>
      <swe:Count>
        <swe:value>230</swe:value>
      </swe:Count>
    </swe:elementCount>
    <swe:elementType name="Components">
      <swe:DataRecord>
        <swe:field name="SamplingTime">
          <swe:Time
definition="http://www.opengis.net/def/property/OGC/0/SamplingTime">
            <swe:uom
xlink:href="http://www.opengis.net/def/uom/ISO-8601/0/Gregorian"/>
              </swe:Time>
          </swe:field>
          <swe:field name="FeatureOfInterest">
            <swe:Text
definition="http://www.opengis.net/def/property/OGC/0/FeatureOfInterest"/>
              </swe:field>
          <swe:field name="Etc">
            <swe:Quantity definition="EvapoTranspiration:Etc">
              <swe:uom code="mm"/>
            </swe:Quantity>
          </swe:field>
          <swe:field name="S_Etc-Kc">
            <swe:Quantity definition="CoefSensibilité:S_Etc-Kc">
              <swe:uom code=" "/>
            </swe:Quantity>
          </swe:field>
          <swe:field name="S_Etc-Etp">
            <swe:Quantity definition="CoefSensibilité:S_Etc-Etp">
              <swe:uom code=" "/>
            </swe:Quantity>
          </swe:field>
          <swe:field name="Ecart-type_Etc">
            <swe:Quantity
definition="Ecart_Type_EvapoTranspiration:I_Etc">
              <swe:uom code="mm"/>
            </swe:Quantity>
          </swe:field>
        </swe:DataRecord>
      </swe:elementType>
    </swe:encoding>

```

```

...
    <swe:TextBlock decimalSeparator="." tokenSeparator=","
blockSeparator="@@" />
    </swe:encoding>
    <swe:values>
        ...
        2003-05-27T07:00:00.000+01:00,parcette.568,3.58,4.06,0.88,0.21@@
        2003-05-28T07:00:00.000+01:00,parcette.568,3.38,3.81,0.89,0.19@@
        2003-15-29T07:00:00.000+01:00,parcette.568,4.25,4.75,0.89,0.24@@
        ...
    </swe:values>
</swe:DataArray>
</om:result>
...

```

L'ordre des paramètres est lié à l'ordre dans lequel les traitements sont effectués : les résultats des derniers traitements effectués sont mis à la suite des précédents. Les coefficients de sensibilité sont déterminés au cours de l'analyse de sensibilité. Ils sont nécessaires au calcul de l'incertitude. De ce fait, les coefficients de sensibilité apparaissent logiquement avant les valeurs d'incertitude.

VIII.A.2 Définition des incertitudes d'entrée

L'approche de Taylor vise à propager les variances des données d'entrées afin d'en déduire celles des résultats. Or dans de nombreux cas, l'utilisateur ne connaît pas l'écart-type ou la variance associé à ses données. Au mieux, il peut savoir que la valeur vraie de sa mesure se trouve dans un intervalle de confiance. Nous proposons ici un moyen simple de déterminer l'écart-type associé à chacune des mesures à partir des valeurs minimales et maximales possibles de la mesure.

La première étape consiste en la détermination des valeurs limites de l'intervalle de confiance. Notre approche admet qu'il puisse être défini de l'une des deux manières suivantes :

- à partir d'un écart ΔX constant par rapport à la valeur moyenne X_{moy} . Par exemple, le coefficient cultural de la plante est estimé à plus ou moins 0.025. Dans ce cas on a :
 - o $X_{min} = X_{moy} - \Delta X$
 - o $X_{max} = X_{moy} + \Delta X$
- à partir d'un écart ΔV proportionnel à la valeur moyenne X_{moy} . Par exemple, la mesure d'évapotranspiration est estimée à plus ou moins 5%. Dans ce cas on a :
 - o $X_{min} = (1 - \Delta V) * X_{moy}$
 - o $X_{max} = (1 + \Delta V) * X_{moy}$

En faisant ensuite l'hypothèse d'une répartition normale des erreurs, et en assimilant les valeurs minimales et maximales calculées aux valeurs seuils d'un intervalle de confiance de 95%. On en déduit l'écart-type par la formule suivante :

$$\text{Ecart type} = (X_{max} - X_{min}) / (2 \times 1.645)$$

En effet dans les conditions de Gauss, la distance entre le 5^{ème} et le 95^{ème} percentile est de 1.645 fois celle de l'écart-type de chaque côté de la valeur moyenne [Loucks, 2005]. Pour des résultats représentatifs, il est bien sûr conseillé à l'utilisateur de s'appuyer sur les valeurs communiquées par les organismes diffuseurs de données.

VIII.A.3 Intégration de la prise en compte de l'incertitude dans les algorithmes de calcul

Notre objectif est de limiter l'ampleur de l'intervention visant à intégrer la prise en compte de l'incertitude dans les algorithmes de calcul développés pour notre modèle. La prise en compte de l'incertitude doit rester optionnelle et ne pas entraver le fonctionnement du modèle.

D'un point de vue technique, chacun des algorithmes liés à un modèle est « amélioré », afin de pouvoir prendre en compte l'incertitude des paramètres d'entrée. Les calculs d'analyse de sensibilité et de propagation des incertitudes sont optionnels (Figure VIII-1), et ne seront réalisés que si l'utilisateur fournit les données nécessaires sur l'incertitude des valeurs des paramètres. Elles font par ailleurs l'objet de classes distinctes dans leur développement.

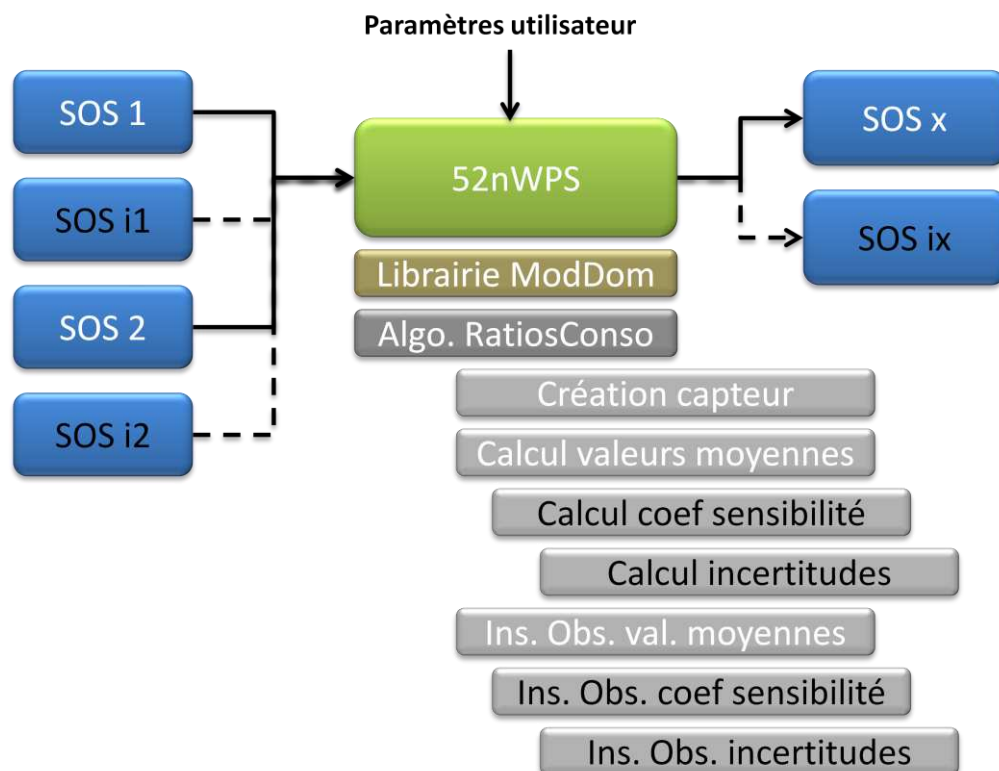


Figure VIII-1 : Schéma de principe du traitement WPS de calculs de valeurs moyennes (en blanc) et en option des incertitudes associées (en noir)

La fonction de propagation des incertitudes est connue et propre au calcul. Elle est prédéterminée et implémentée au sein de l'algorithme.

VIII.B Intégration de la prise en compte de la qualité dans l'IHM

La prise en compte de l'incertitude impacte l'IHM. On distingue deux modifications majeures :

- les analyses de la zone d'étude et de la sensibilité visant à simplifier le modèle sont proposées dans la phase de préparation de la simulation ;
- la propagation de l'incertitude et l'analyse de sensibilité finale des résultats sont proposées dans l'actuel module de simulation.

VIII.B.1 Evolutions dans la phase de préparation de la simulation

Comme nous l'avons évoqué au Chapitre V, l'estimation des besoins en eau pour notre modèle est réalisée à l'échelle de la parcelle. Pour un ensemble de parcelles, on réalise donc la somme des besoins en eau de chacune des parcelles constituant l'ensemble. Pour réduire le nombre de cas à traiter, et donc le nombre de fois que les calculs devront être lancés, nous avons proposé au Chapitre VII de procéder à des regroupements de parcelles. Nous présentons ici l'implémentation des méthodes d'analyse et de regroupement des parcelles selon leurs propriétés.

Les regroupements se font en deux temps (Figure VIII-2) :

- après analyse de la zone d'étude et des caractéristiques des parcelles qu'elle comporte, on constitue des groupes constitués de parcelles ayant des propriétés communes ;
- après analyse de la sensibilité à la variation spatiale et temporelle des mesures, mais également de la variation des paramètres, on procède au regroupement des parcelles ayant des propriétés proches sur des points où la sensibilité est faible.

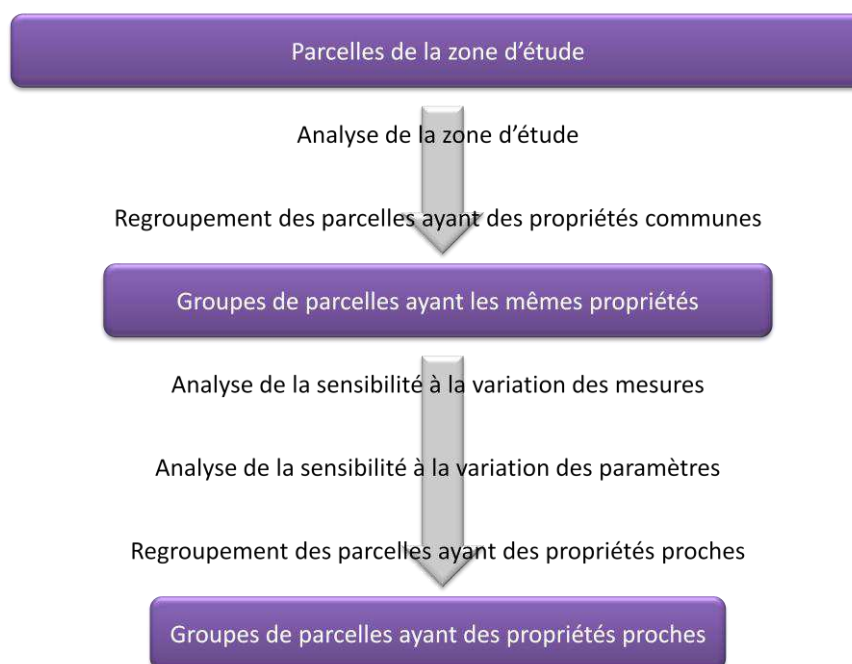


Figure VIII-2 : Principe de regroupement des parcelles

VIII.B.1.a Analyse de la zone d'étude et regroupement des parcelles ayant les mêmes propriétés

Au cours de cette étape nous analysons la couche obtenue suite au traitement évoqué à la Figure V-2. Nous appelons cette couche « Zone d'étude ». L'analyse menée consiste à parcourir l'ensemble de la couche et d'identifier les parcelles ayant strictement les mêmes propriétés. Ce traitement va entraîner deux actions :

- l'ajout dans la couche « Zone d'étude » d'un champ « Type_parcelle » qui indique de quel type de parcelle il s'agit compte tenu de ses caractéristiques, ainsi qu'un champ « Superficie » dans lequel est calculée la superficie de la parcelle compte tenu de sa géométrie. On note dans l'exemple ci-dessous (Tableau VIII-1) que les parcelles 1234 et 1237 possèdent les mêmes caractéristiques et sont donc rattachées au même type de parcelle ;

Id_parcelle	Param1	Param2	Param3	Param4	Superficie	Type_parcelle
1234	X	5	AA	0.40	68.6	1
1235	Z	6	AA	0.60	4.2	5
1236	X	8	BB	1.30	125.1	3
1237	X	5	AA	0.40	21.1	1

Tableau VIII-1 : Structure de la table zone d'étude

- la création d'une table « Type_Parcelle », qui précise pour chaque type de parcelles les valeurs de chacun des paramètres. De plus, on indique le nombre d'occurrences de chaque type de parcelles dans la zone d'étude, ainsi que la superficie totale occupée par ces différents types (Tableau VIII-2).

Id_type	Param1	Param2	Param3	Param4	Occurrence	Superf_tot
1	X	5	AA	0.40	123	12256.2
2	X	5	AA	0.60	3	48.5
3	X	8	BB	1.30	226	32655.9
4	Y	5	AA	0.40	22	777.4

Tableau VIII-2 : Structure de la table Type_Parcelle

Une fois ce traitement réalisé, l'utilisateur peut décider d'estimer les besoins en eau de chaque type de parcelle référencé ou bien procéder à une analyse de sensibilité dans le but de réduire le nombre de cas à traiter. Si on s'intéresse à estimer les besoins en eau sur une grande zone d'étude, on peut également choisir de ne pas traiter certains cas rares dont la superficie couverte par les parcelles est très faible par rapport à la taille de la zone d'étude.

VIII.B.1.b Analyse de l'importance de la variabilité spatiale et temporelle des mesures

La première étape consiste à parcourir l'ensemble des mesures associées aux variables et d'identifier des valeurs caractéristiques : moyenne, valeur minimale, valeur maximale et variance (ou écart type). Si on s'intéresse par exemple à l'étude de la variabilité temporelle des mesures d'évapotranspiration (ETo) sur la période du 1^{er} juillet au 31 août 2003 pour la maille SAFRAN n°9759 (située dans la région de Céret), on obtient les valeurs suivantes (Tableau VIII-3) :

Moyenne	Ecart-type	Minimum	Maximum
4.48 mm	1.40 mm	0.91 mm	7.68 mm

Tableau VIII-3: Valeurs statistiques obtenues

Une étude similaire sur la variabilité spatiale peut être menée : dans ce cas on comparera les mesures obtenues pour chacune des mailles SAFRAN concernées par la zone d'étude à des dates données.

On effectue ensuite une analyse de sensibilité en considérant ces valeurs. Si on constate que les variations des variables ont peu d'impact sur le modèle, alors on pourra procéder à une simplification du modèle en fixant la valeur de la variable à la valeur moyenne. Le potentiel de simplification du modèle sera d'autant plus grand que l'étude est restreinte (aussi bien spatialement, que temporellement). La prise en compte des résultats de cette analyse se fait au niveau du paramétrage de la simulation.

VIII.B.1.c Analyse de l'importance de la variabilité des valeurs des paramètres

A l'image du traitement précédent (VIII.B.1.b), la première étape consiste à identifier l'ensemble des valeurs possibles pour chaque paramètre, et d'en déduire la moyenne, l'écart type ainsi que les valeurs extrêmes. On effectue alors une analyse de sensibilité avec les valeurs obtenues. Si on constate que les variations de certains paramètres ont peu d'impact sur le modèle, alors on pourra envisager de regrouper les types de parcelles qui ne se différencient que par ce paramètre.

VIII.B.1.d Regroupement des parcelles ayant des propriétés proches

Suite à l'analyse de sensibilité précédente, l'utilisateur aura la possibilité de regrouper certains types de parcelles. Ses choix seront guidés par les résultats obtenus au cours des analyses de sensibilité précédentes. Les regroupements se font selon des méthodes d'échantillonnage basées sur des analyses unicritères. Dans un premier temps, nous proposons d'implémenter deux approches (Figure VIII-3) :

- une méthode proposant d'utiliser des intervalles équivalents, où les seuils sont déterminés à partir des valeurs extrêmes et du nombre de classes souhaité ;
- une méthode de discrétisation manuelle où l'utilisateur définit lui-même le nombre de classe, ainsi que les seuils.

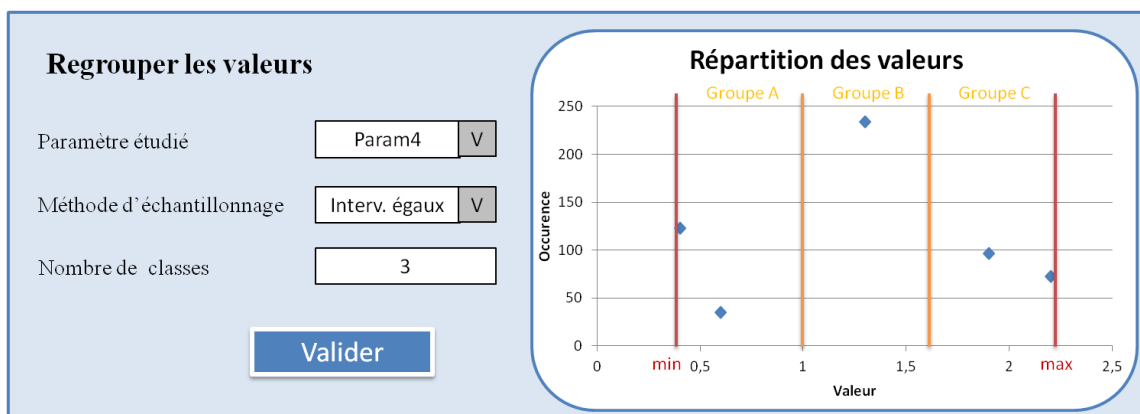


Figure VIII-3 : Module de discrétisation

D'autres méthodes d'échantillonnage, notamment probabilistes, pourront être envisagées. Une méthode appréciable serait une classification automatique en fonction des résultats de l'analyse de sensibilité. Une fois que les options de regroupement ont été déterminées, un traitement de généralisation est lancé. Ce traitement va entraîner les actions suivantes :

- le calcul des valeurs moyennes pour chacun des groupes définis pour les paramètres discrétisés, et la création de tables nommées « Groupe_NomParametre » dans lesquelles les données seront stockées. Pour reprendre l'exemple de la Figure VIII-3, on aurait (Tableau VIII-4) :

Id_Groupe_Param4	Groupe_Param4	Valeur moyenne
1	A	0.7
2	B	1.3
3	C	1.9

Tableau VIII-4 : Structure de la table Groupe_NomParametre

- l'ajout dans la table « Type_Parcelle », obtenue suite au traitement évoqué en VIII.B.1.a, de champs intitulés « Id_Groupe_NomParametre », dans lesquels seront indiqués les identifiants des groupes associés pour chacun des paramètres, ainsi que l'ajout d'un champ « Groupe_Parcelle » dans lequel indique dans quel groupe est mis le type de parcelle compte tenu de ses caractéristiques (Tableau VIII-5) ;

Id_type	...	Param4	...	Occurrence_type	Superf_tot	Id_Groupe_Param4	Id_Groupe_Parcelle
1		0.40		123	12256.2	1	1
2		0.60		3	48.5	2	2
3		1.30		226	32655.9	1	1
4		0.40		22	777.4	2	3

Tableau VIII-5 : Structure de la table Type_Parcelle

- la création d'une table « Groupe_Parcelle », qui précise pour chaque groupe identifié les valeurs à retenir pour chacun des paramètres. De plus, on indique le nombre d'occurrences de chaque groupe de parcelles dans la zone d'étude, ainsi que la superficie totale occupée par ces différents groupes (Tableau VIII-6).

Id_groupe	...	Param4	...	Occurrence_groupe	Superf_tot
1		0.7		4	52656.4
2		1.3		7	44648.5
3		1.9		2	23757.3

Tableau VIII-6 : Structure de la table Groupe_Parcelle

Une fois ce traitement réalisé, l'utilisateur peut se concentrer sur l'exécution du modèle pour chacun des groupes de cas référencés.

VIII.B.2 Evolutions dans la phase de simulation

Au sein de ce module de calcul, l'utilisateur dispose de quatre choix possibles :

1. Calculer les valeurs moyennes (sans prise en compte de l'incertitude) : un capteur virtuel sera créé et alimenté par les mesures calculées,
2. Tout calculer (valeurs moyennes, coefficients de sensibilité et incertitudes associées aux valeurs calculées) : un capteur virtuel sera créé et les mesures moyennes ainsi que celles des coefficients de sensibilité et des incertitudes lui seront affectées,
3. Calculer uniquement les coefficients de sensibilité : l'utilisateur précise quel capteur sera alimenté, les coefficients de sensibilité sont calculés et sont aussitôt associées au capteur,
4. Calculer uniquement les incertitudes associées au calcul : de même l'utilisateur précise quel capteur sera alimenté, les mesures d'incertitude sont effectuées et sont aussitôt associées au capteur.

Selon le choix de l'utilisateur, des paramètres et données complémentaires peuvent lui être demandés (notamment les incertitudes sur les données entrantes).

Les options 3 et 4 permettent à l'utilisateur, s'étant dans un premier temps contenté de calculer les valeurs moyennes seules, de venir compléter sa simulation en prenant en compte l'impact de l'incertitude des données sur ses résultats. Pour les variables, l'utilisateur aura la possibilité de définir une valeur constante à prendre en compte plutôt que la variable (Figure VIII-4).

Vitesse moyenne du vent

Utiliser les données d'un capteur

Fixer la valeur

m.s⁻¹

Figure VIII-4 : Possibilité de fixer la variable à une valeur constante

Cette option est intéressante si l'utilisateur a identifié que la variabilité d'une variable avait peu d'impact lors de l'analyse de sensibilité. Elle l'est également bien sûr si l'utilisateur n'a pas accès à des capteurs diffusant cette information. L'utilisateur ayant fait une analyse de sensibilité est conscient de l'approximation engendrée par ce choix.

VIII.B.3 Représentation des résultats

L'incertitude associée à un résultat est une information essentielle pour toute personne devant s'appuyer sur cette donnée pour prendre une décision.

La communication de l'incertitude doit être aussi soignée que celle du résultat lui-même. Une représentation sous forme graphique semble la plus appropriée pour communiquer ces informations. Dans cette partie nous traitons de méthodes de communication de la sensibilité et de l'incertitude.

Nous les illustrons dans le cadre du calcul de l'évapotranspiration d'une culture (ETc), selon la formule introduite par la Figure I-5. Les données utilisées sont :

- les mesures d'évapotranspiration ETo de la maille SAFRAN n°9759 (dans la région de Céret) pour l'année 2003, avec un intervalle de confiance de plus ou moins 0.05 mm pour chacune des mesures ;
- les coefficients culturaux Kc pour des abricotiers définis dans la Figure VII-1, avec également un intervalle de confiance de plus ou moins 0.05 à chaque instant.

On s'intéresse plus particulièrement à la représentation de la sensibilité et de l'incertitude associée à une mesure : celle du 25 juillet 2003, pour laquelle on a les valeurs suivantes (Tableau VIII-7) :

Paramètre	Valeur moyenne	Ecart	Valeur min	Valeur max
ETo	4.35 mm	+ / - 0.05 mm	4.30 mm	4.40 mm
Kc	0.90	+ / - 0.05	0.85	0.95

Tableau VIII-7 : Valeurs du 25 juillet 2003 utilisées pour les calculs

VIII.B.3.a Résultats de l'analyse de sensibilité

La sensibilité d'un résultat à plusieurs variables peut-être représentée de manière graphique. [Loucks, 2005] présente trois représentations possibles.

Nous montrons ici les résultats d'une analyse de sensibilité simple, basée sur l'étude de l'impact d'un paramètre à la fois. Ainsi pour l'exemple et les valeurs indiquées dans le Tableau VIII-7, on obtient les résultats suivants (Tableau VIII-8) :

Etude	Paramètres	Valeur calculée pour ETc	Ecart min - max
Variation de Kc	(Kc min, ETo moy)	3.70 mm	0.43 mm
	(Kc max, ETo moy)	4.13 mm	
Variation d'ETo	(Kc moy, ETo min)	3.87 mm	0.09 mm
	(Kc moy, ETo max)	3.96 mm	

Tableau VIII-8 : Résultats de l'analyse de sensibilité simple

Le diagramme de Pareto (Figure VIII-5) offre une présentation classique sous la forme d'histogrammes. Cette représentation met l'accent sur l'étendue des résultats fournis au regard de l'incertitude associées aux valeurs d'entrées. Les colonnes sont disposées selon un ordre décroissant. On remarque ici que l'intervalle de résultat obtenu compte tenu de la variabilité de la valeur de Kc est bien supérieur à celui associé à la variabilité de la valeur d'ETo.

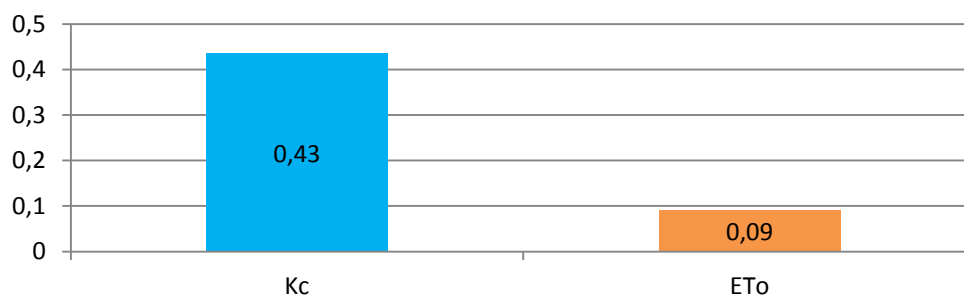


Figure VIII-5 : Diagramme de Pareto

Dans le diagramme tornade (Figure VIII-6), les paramètres sont classés de manière à ce que les plus grandes étendues de valeurs soit en haut, et les plus faibles en bas, ce qui donne au diagramme un aspect de tornade.

Le diagramme tornade montre la plage des valeurs possibles compte tenu de l'incertitude sur les paramètres. D'un point de vue technique, la représentation des diagrammes tornade est plus complexe et n'est proposée que par peu d'outils.

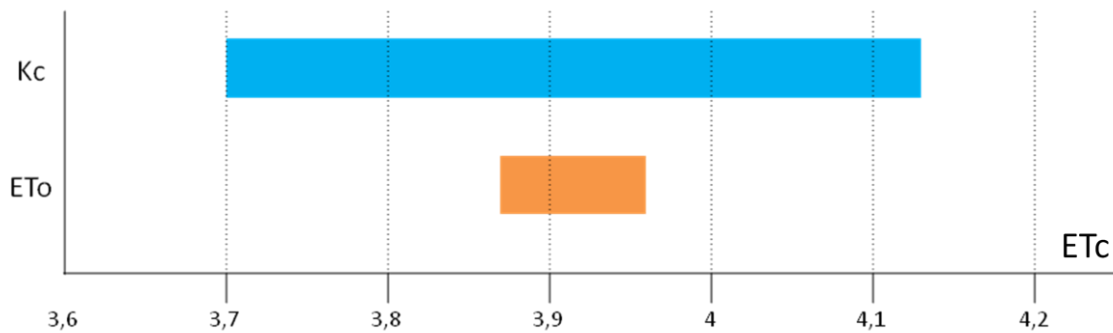


Figure VIII-6: Diagramme tornade

Le troisième schéma est appelé diagramme en toile d'araignée. Par rapport aux deux premiers diagrammes, il montre en plus comment varie le résultat par rapport à une variation sur un paramètre.

Pour notre exemple, cette représentation présente peu d'intérêt car une variation de X % sur n'importe laquelle des deux variables entraîne une variation de X % sur le résultat. Les deux courbes sont ainsi superposées (Figure VIII-7).

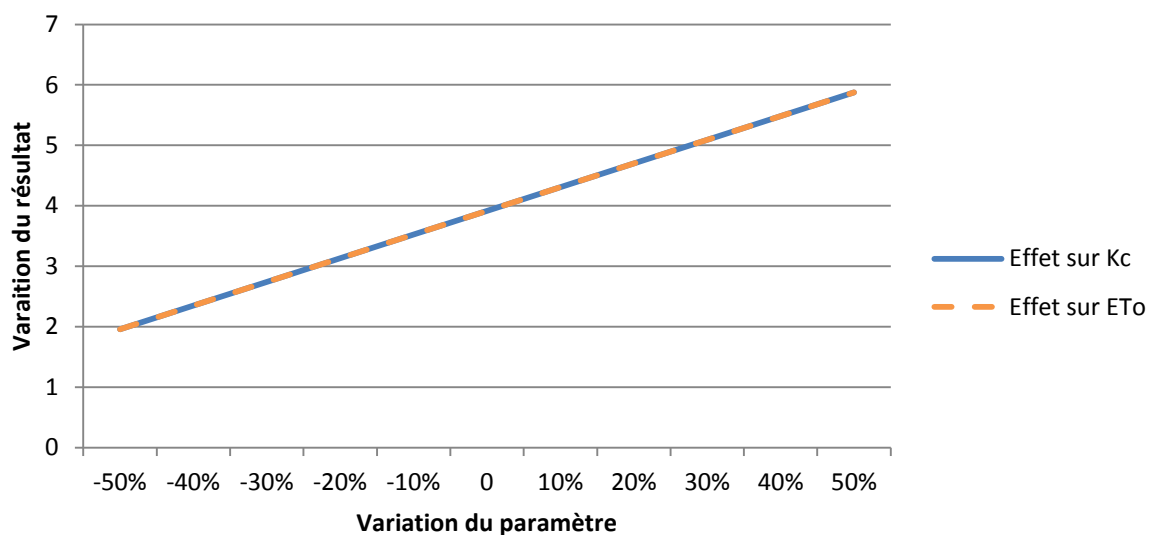


Figure VIII-7 : Diagramme en toile d'araignée

Pour mieux montrer l'intérêt de cette représentation, nous présentons (Figure VIII-8) des résultats issus de [Loucks, 2005] pour le calcul de la sensibilité de la concentration en phosphore (P) d'un lac par rapport à trois paramètres : le taux annuel de phosphore (L), le débit annuel (q), ainsi que la profondeur moyenne de l'eau (z), selon la formule suivante :

$$P = (L/q) / [1 + \left(\frac{z}{q}\right)^{0.5}]$$

Le graphique montre des différences de comportement du résultat au regard de variations sur les variables du calcul : on note la prépondérance des variables L et Q sur z dont la sensibilité est faible.

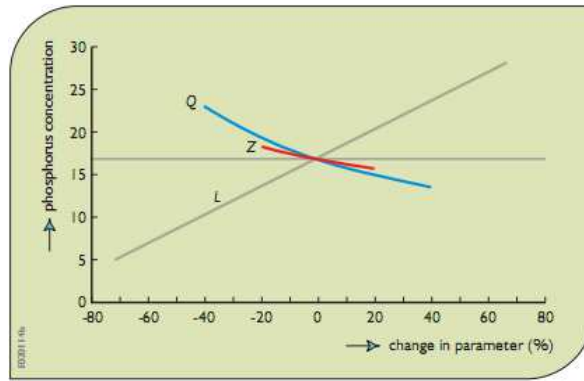


Figure VIII-8 : Diagramme en toile d'araignée [Loucks, 2005]

Une alternative aux trois méthodes évoquées ci-dessus est un classique diagramme de type camembert. Ce type de représentation met l'accent sur les contributions des incertitudes associées à chacune des valeurs d'entrée dans l'incertitude associée au résultat final.

Pour illustrer cette représentation, nous réalisons une analyse de sensibilité de premier ordre, telle que celle introduite par [Loucks, 2005].

- nous déterminons d'abord les coefficients de sensibilité $\frac{\delta ETc}{\delta X}$, respectivement $\frac{\delta ETc(t)}{\delta ETo}$ pour $ETo(t)$ et $\frac{\delta ETc(t)}{\delta Kc}$ pour $Kc(t)$. D'après la formule utilisée pour calculer ETc (Figure I-5), nous avons :

$$\circ \quad \frac{\delta ETc(t)}{\delta ETo} = Kc(t) \quad \text{ainsi que} \quad : \quad \frac{\delta ETc(t)}{\delta Kc} = ETo(t) ;$$

- nous calculons ensuite les écart-types sur les deux variables à partir de la formule introduite dans la partie VIII.A.2. ;
- en réalisant un développement en séries de Taylor de la formule introduite Figure I-5, et en supposant les variables ETo et Kc indépendantes, nous avons :

$$Var(ETc(t)) = \left[\left(\frac{\delta ETc(t)}{\delta ETo} \right)^2 * (Ecart - type ETo(t))^2 \right] + \left[\left(\frac{\delta ETc(t)}{\delta Kc} \right)^2 * (Ecart - type Kc(t))^2 \right]$$

- on en déduit la répartition de la variance du résultat.

En reprenant les valeurs utilisées pour VIII.B.3, on obtient ces résultats (Tableau VIII-9) :

Variable	Unité	$\frac{\delta ETc}{\delta X}$	Ecart-type [X]	$(\frac{\delta ETc}{\delta X})^2 * Var[X]$	%
ETo	mm	0,90	0,03	7,4.10 ⁻⁴	4
Kc	-	4,35	0,03	1,7.10 ⁻²	96

Tableau VIII-9 : Valeurs obtenues

Le diagramme suivant montre la répartition de la variance des résultats (Figure VIII-9).

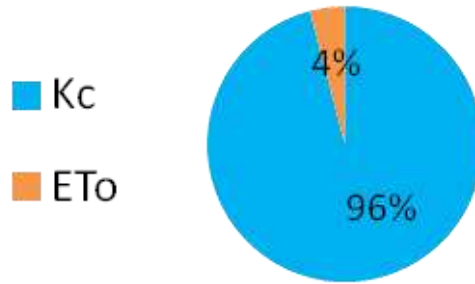


Figure VIII-9 : Répartition de la variance du résultat

Il s'agit de la représentation que nous avons retenue.

VIII.B.3.b Résultats de l'incertitude propagée

Dans le Chapitre VI, nous avons présenté notre approche de représentation des données issues de capteurs virtuels, basée sur la présentation des résultats sous la forme de chroniques. Dans le Chapitre VII, nous avons également exprimé notre souhait de ne pas dissocier les valeurs d'incertitude des mesures auxquelles elles sont rattachées.

Afin de nous inscrire dans cette continuité, nous avons conservé une représentation sous la forme de courbes. Nous proposons ainsi de représenter des intervalles de confiance. Les seuils de ces intervalles sont déterminés en présupposant une répartition normale des erreurs à partir des écarts types que nous avons estimés.

Le graphique ci-dessous (Figure VIII-10) présente la courbe des valeurs calculées pour ETc (en rouge) à partir des valeurs indiquées dans la partie VIII.B.3, avec un intervalle de confiance de 95% déterminé par l'utilisateur (en vert).

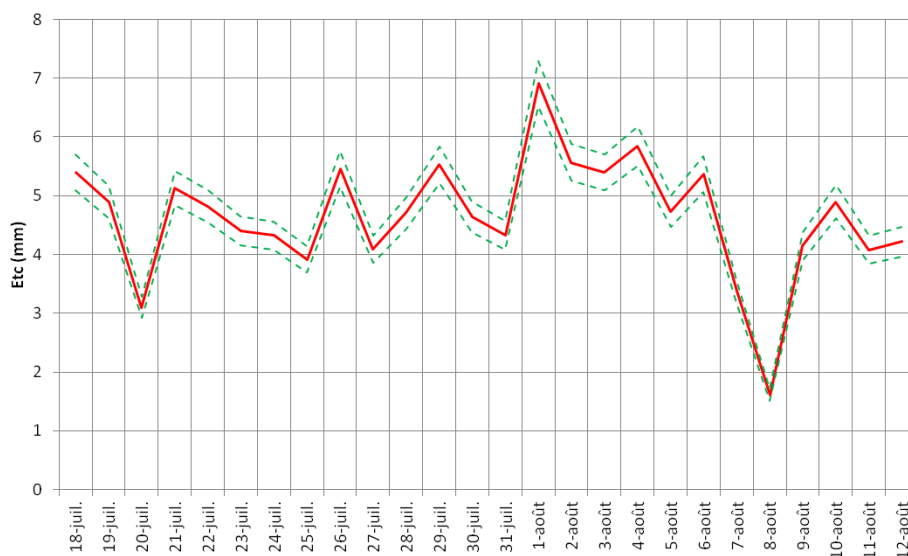


Figure VIII-10 : Représentation des résultats accompagnés de l'écart type mesuré

VIII.B.3.c Consultation des résultats finaux

Les résultats d'une simulation sont accessibles par un clic sur la parcelle (ou la zone d'étude) qui nous intéresse. Tout comme pour la consultation des données météorologiques, les chroniques sont affichées. Les valeurs calculées sont représentées sous la forme d'une courbe. Un intervalle de confiance de 95% est représenté autour de la valeur estimée. Un clic sur une mesure permet d'obtenir des informations complémentaires : la valeur estimée, l'écart type estimé ainsi que la sensibilité du résultat aux paramètres et variables ayant directement permis son calcul (Figure VIII-11).

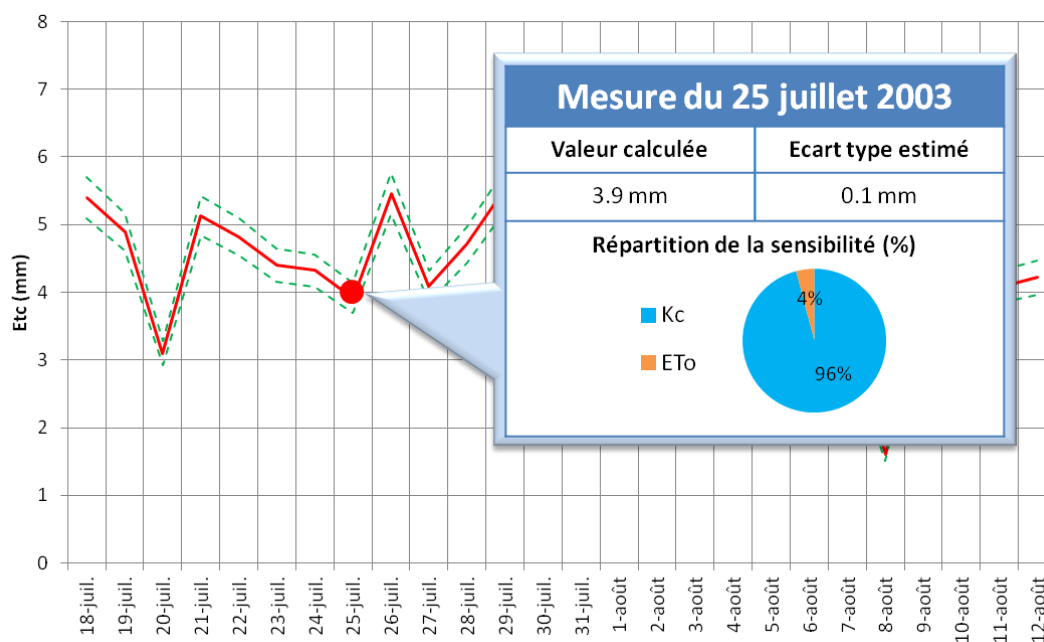


Figure VIII-11 : Affichage des résultats finaux pour une parcelle : valeurs calculées, écart-type estimé et sensibilité par rapport aux paramètres et variables d'entrées

Ce type de représentation concentre les informations et permet à l'utilisateur de visualiser simplement l'ensemble des informations relatives à la qualité de la mesure indiquée.

VIII.B.3.d Réflexion sur la communication des incertitudes propagées

Dans le cadre de modèles faisant appel à une succession de traitements dans lesquels les résultats d'un traitement sont réutilisés par un voire plusieurs autres, une information intéressante serait une représentation graphique de l'impact d'une variable sur le résultat final.

Pour reprendre l'exemple évoqué à la Figure VIII-9, si les valeurs de la variable 1 sont déterminées via un traitement faisant intervenir d'autres variables. Il serait intéressant de pouvoir visualiser d'un seul coup d'œil l'impact de ces dernières sur le résultat final.

Pour rendre compte de la sensibilité du résultat final par rapport à l'ensemble des paramètres et variables intervenant dans sa construction, nous proposons d'utiliser un diagramme de type « donut

multi-séries ». Un tel diagramme se lit de l'intérieur vers l'extérieur, du plus général vers le plus spécifique.

En prenant l'exemple de notre modèle d'estimation des besoins en eau agricole, on retrouve le résultat final au centre, puis en s'éloignant du centre les résultats intermédiaire avec une représentation de leur « poids » dans la détermination du résultat. Le diagramme ci-dessous (Figure VIII-12) présente les résultats (fictifs) de la part des variables intervenant dans le calcul du besoin en eau de la plante. La représentation est ici limitée à deux niveaux.

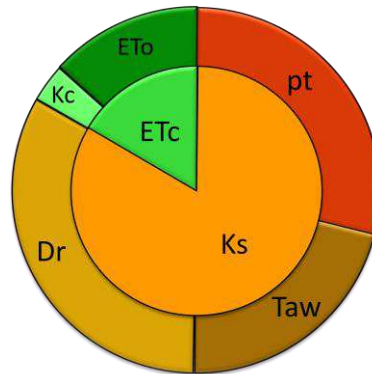


Figure VIII-12 : Représentation de la sensibilité du résultat de l'estimation des besoins en eau de la culture

En remontant à trois niveaux (Figure VIII-13), on note dans l'exemple ci-dessous que le paramètre ET_c intervient deux fois dans l'estimation du besoin en eau de la plante. L'impact d'une perturbation sur ce paramètre ne peut plus être directement quantifié.

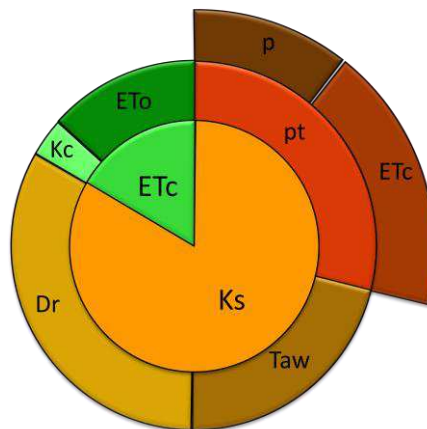


Figure VIII-13 : Illustration de la limite de la représentation en cas d'intervention multiple d'une variable ou d'un paramètre dans le modèle.

Cette remarque nous montre la nécessité d'aborder le modèle selon deux niveaux : traitement par traitement, ou de manière globale. La propagation des incertitudes du modèle global nécessiterait :

- de formaliser l'expression du résultat final sous la forme d'une unique équation faisant intervenir tous les paramètres et toutes les variables intervenant consécutivement dans le calcul ;
- puis de déterminer les différentielles associées à chacun des paramètres à partir de cette même formule.

C'est dans cette dernière tâche que réside le gros de l'effort à fournir. L'utilisation de méthode de dérivations automatisées serait ici des plus utiles.

VIII.C Conclusion

Dans ce chapitre, nous avons présenté les actions à mener pour l'implémentation des méthodes et techniques de prise en compte de l'incertitude introduites dans le Chapitre VII dans le prototype d'IDS présenté au Chapitre VI.

L'impact porte sur deux éléments principaux de notre IDS :

- au niveau du modèle lui-même, les algorithmes de calcul implémentés dans le prototype sont modifiés de sorte à comporter des traitements supplémentaires relatifs à la propagation de l'incertitude et l'analyse de sensibilité du résultat. Ces calculs optionnels sont effectués uniquement si l'utilisateur fournit des informations relatives aux incertitudes des valeurs et paramètres d'entrée. Pour le modèle d'estimation des besoins en eau implémenté dans le prototype d'IDS, les données en sortie du calcul, les incertitudes associées, ainsi que les résultats de l'analyse de sensibilité sont tous affectés au même capteur ;
- au niveau de la préparation de la zone d'étude, des traitements permettant de réaliser le regroupement des parcelles sont mis en place. Ils s'appuient sur l'analyse des caractéristiques des parcelles constituant la zone d'étude, ainsi qu'en leur regroupement selon des critères définis par l'utilisateur. Des fonctionnalités d'analyse de sensibilité sont décrites afin de guider l'utilisateur au niveau des possibilités de regroupement. Ces techniques mériteraient d'être automatisées de sorte à proposer dynamiquement les regroupements à mettre en place.

Nous avons enfin présenté et illustré des solutions relatives à la représentation de l'incertitude et des résultats de l'analyse de sensibilité dans notre IDS.

Conclusion et perspectives

Conclusion

Au cours de ce travail de recherche, nous avons souhaité montrer l'intérêt de mettre à disposition des modèles environnementaux sur le web. La constitution de modèles fiables et pertinents est un processus long et couteux, aussi la possibilité de réutiliser directement des travaux déjà réalisés permettrait de faciliter l'avancée des chercheurs et environnementalistes quant à la compréhension de notre environnement.

Le principe de la mise à disposition de modèles sur internet s'inscrit comme une suite logique aux travaux entrepris dans le cadre des différentes initiatives mises en place pour le partage des informations relatives à l'environnement (Directive INSPIRE, projets GEOSS, GMES et SEIS).

Dans ce mémoire, nous avons présenté nos travaux de recherche et de développement dans la conception d'une Infrastructure de Données Spatiales (IDS). De tels outils permettent à des organismes de diffuser et mettre à disposition de ressources sous une plateforme d'échange mutualisé.

Notre IDS, baptisée FUSIEAU, s'intéresse à la mise à disposition de modèles d'estimation des besoins en eau agricoles et domestiques. Les résultats de tels modèles peuvent être utilisés dans le cadre de la modélisation hydrologique d'un bassin versant, tel qu'il a été réalisé dans le cadre du projet VULCAIN, mené par le BRGM.

Les méthodes d'estimation des besoins en eau utilisées dans FUSIEAU, reposent sur le principe de la fusion de données. Pour reprendre la catégorisation indiquée par l'Open Geospatial Consortium (OGC) à ce sujet, les algorithmes développés font appel au principe de la fusion de données de deuxième niveau. Il s'agit en effet de combiner des données brutes issues de thématiques diverses pour produire des indicateurs hybrides.

La principale problématique du projet porte sur la conception de modèles adaptés au contexte web. Outre le fait que le modèle doit reproduire fidèlement le phénomène environnemental qu'il modélise, les contraintes techniques du web ont réduit le champ des possibles. Les traitements effectués dans les modèles doivent ainsi être relativement simples, pour offrir un temps de réponse raisonnable.

Les modèles s'appuient sur une approche générique qui permet de les appliquer à d'autres zones d'études et à d'autres échelles de travail. Afin de favoriser l'interopérabilité, les modèles s'appuient sur des données disponibles sur le web. Selon cette logique, nous avons travaillé pour que les résultats soient diffusés dans des formats standards. Il en va de même pour les traitements eux-mêmes qui peuvent ainsi être réutilisés dans d'autres applications.

Le fonctionnement de FUSIEAU repose ainsi sur l'utilisation de services web, aussi bien pour la diffusion des données d'entrées, que pour celle des résultats (données en sorties). Cette approche vaut également pour la mise à disposition des fonctionnalités de traitement de données.

Nous utilisons ainsi les standards WFS et SOS pour les données, ainsi que le WPS pour les traitements. Les services WPS travaillent de manière autonome et assurent l'interface avec le client (de la récupération des données, jusqu'à la fourniture des résultats) selon un principe d'orchestration. Ces WPS peuvent ensuite être chaînés selon le principe d'orchestration ou de chorégraphie de services web, afin de produire une évaluation complète des besoins en eau « en un clic ».

Les résultats intermédiaires et finaux (le besoin en eau des cultures) sont sauvegardés sous la forme de services SOS. Nous produisons ainsi des capteurs virtuels, dont les mesures sont en fait alimentées par les calculs effectués au cours de notre modèle. Ces capteurs sont associés à une zone d'étude, et donc rattachés à une entité géographique. Un couplage entre SOS et WFS a pour cela été mis en place afin de visualiser les capteurs associés à une zone d'étude, et réciproquement pour visualiser la zone d'étude couverte par un capteur.

La conception de FUSIEAU s'accompagne également d'une interface cartographique, qui permet de visualiser les informations spatiales. Elle permet aussi d'utiliser les fonctionnalités de traitement de données : créer ses propres scénarii pour l'estimation des besoins en eau, les lancer et visualiser les résultats. Les mesures des SOS (données ou résultats) peuvent être visualisées sous la forme de graphiques.

Nous avons enfin exploré les possibilités d'introduire la notion d'imperfection des données. La prise en compte de cet aspect est en effet inhérente à toute entreprise de modélisation, aussi nous avons souhaité l'inclure dans notre IDS. La gestion de la composante qualité requiert quelques modifications : en particulier sur les données et sur les traitements.

Nous avons introduit des méthodes permettant de prendre en compte l'imperfection et étudié son intégration dans notre prototype. Du stockage des informations qualifiant l'imperfection, jusqu'à leur représentation graphique en passant par les traitements à mettre en place pour l'exploiter.

A terme, FUSIEAU proposera des fonctionnalités d'analyse de sensibilité permettant de caractériser le comportement du résultat vis-à-vis de variations apportées sur les données. De même une méthode de propagation des incertitudes, basée sur des approximations de Taylor a été retenue. Elle s'avère plus adaptée au contexte web qu'une approche probabiliste telle que les simulations de Monte-Carlo. L'ensemble des traitements inhérents à la prise en compte de la qualité des données sera implémenté sous la forme de WPS. Le but est que ces processus puissent être utilisés dans d'autres modèles.

Perspectives

Nous voyons plusieurs axes de recherches et perspectives d'évolution pour FUSIEAU. Parmi nos premiers projets, il y a bien sûr celui de continuer le développement du prototype et finir d'intégrer les concepts que nous avons définis précédemment. Ainsi dans un futur très proche nous pensons travailler sur :

- l'intégration dans le prototype des méthodes de prise en compte de l'imperfection des données introduites dans le Chapitre VII ;
- la mise en place du modèle d'estimation des besoins en eau domestiques présenté dans le Chapitre V.

De nombreuses améliorations des processus et outils déjà développés ont également été suggérés dans les chapitres précédents (automatisation du regroupement des parcelles en fonction de leurs caractéristiques, utilisation de méthodes automatisées pour le calcul des dérivées partielles). La réalisation des évolutions précitées devrait ensuite permettre d'envisager l'industrialisation de l'outil.

D'un point de vue technique, des solutions visant à optimiser les temps de traitement peuvent également être mises en place. Nous pensons notamment à la parallélisation des calculs, telle qu'évoquée par [Neun & al, 2012]. Sa mise en œuvre serait facilitée par le fait que notre modèle a été décomposé en tâches élémentaires. Nous voyons deux méthodes d'implémentation de cette approche :

- au niveau du modèle, répartir les traitements à réaliser entre différentes machines de calcul. Cette action est uniquement réalisable pour les processus indépendants (ne nécessitant pas les résultats d'un autre calcul),
- au niveau de la zone d'étude, une approche particulièrement intéressante pour le modèle agricole, qui consiste à traiter plusieurs parcelles à la fois.

Une autre étude à mener est celle de la validation de l'outil proposé auprès de divers utilisateurs, et ainsi mesurer la qualité externe de notre solution. Parmi les points à étudier nous voyons :

- pour le modèle lui-même : l'intérêt des traitements réalisés et des résultats produits, la facilité d'utilisation (notamment sa capacité à s'appuyer sur des données disponibles) ;
- pour l'interface : l'ergonomie, la pertinence des rendus pour les résultats.

Néanmoins, le réel intérêt de FUSIEAU ne pourra être mesuré que lorsque les organismes producteurs de données mettront à disposition leurs informations sur internet selon des formats interopérables de diffusion, tel que suggéré par la Directive INSPIRE. La conception et l'architecture de FUSIEAU repose en effet sur ce postulat fort de partage et d'échanges de ressources sous la formes de services web standards spécifiés par l'OGC.

D'une manière générale, le principe de « Model web » doit être soutenu et encouragé par la mise à disposition de nombreux modèles environnementaux qui constituent la meilleure démonstration de l'intérêt du concept. Des procédés visant à simplifier la mise à disposition de traitements sur internet devraient également être mis en place pour accélérer le processus.

Bibliographie

[AELB, 2005] Agence de l'Eau Loire-Bretagne. *Rapport n°2 Consommations d'eau potable et potentiels d'économies*. Septembre 2005.

[Agumya & Hunter, 1997] AGUMYA A., Hunter G.J., *Determining fitness for use of geographic information*, ITC Journal, vol. 2, n° 1, 1997, p. 109-113.

[Aigner & al., 2006] AIGNER, M ; ZIEGLER, G M. (trad. Nicolas Puech), *Raisonnements divins : quelques démonstrations mathématiques particulièrement élégantes*, Springer Verlag, 2^e éd., 2006, 270 p.

[Albreshne & al., 2009] ALBRESHNE, A; FUHRER, P ; PASQUIER, J. *Web Services Orchestration and Composition : Case Study of Web services composition*. September 2009.

<http://diuf.unifr.ch/drupal/softeng/sites/diuf.unifr.ch.drupal.softeng/files/file/publications/internal/WP09-03.pdf>

[Allen & al, 1998] ALLEN R. G., PEREIRA L. S., RAES D., and SMITH, M. *Crop evapotranspiration: Guidelines for computing crop water requirements*. 1998. FAO Irrigation and Drainage Paper No. 56, Roma, Italy.

[Ambassade de l'eau, 2012] *Présentation de STRATEAU sur le site de l'Ambassade de l'Eau*. Consulté le 26 septembre 2012. <http://www.ambassade-eau.com/strateau-presentation>

[Ambassade de l'eau, 2009] *Guide utilisateur de STRATEAU*. Consulté le 26 septembre 2012. http://www.ambassade-eau.com/sites/default/files/docs/strateau/guideSTRATEAU_fr.pdf

[André, 2008] ANDRE R. *Sur les incertitudes de mesure*. 2008. www.math.u-bordeaux1.fr/ecaq/cours_pdf/EcAq2008_20_R.Andre.pdf

[Barraqué & al, 2008] BARRAQUÉ, B. ; NERCESSIAN A. *Mieux Comprendre Comment Évolue la Consommation d'Eau à Paris (Rapport Final)*. Décembre 2008.

[Bastin, 2012] BASTIN, L.; CORNFORD, D.; JONES, R.; HEUVELINK, G.B.M.; PEBESMA, E.; STASCH, C.; NATIVI, S.; MAZZETTI, P. and WILLIAMS, M. *Managing uncertainty in integrated environmental modelling : the UncertWeb framework*. Environmental Modelling & Software, (2012).

[Beaufils & al, 2011] BEAUFILS M, FOLLIN J-M. *Geospatial Data Fusion for Water Resources Assessment*. Poster + résumé pour la conference AGILE 2011. 18-21 Avril, Utrecht, Pays-Bas.

[Beaufils & al, 2011 bis] BEAUFILS M, FOLLIN J-M. *Toward a quality aware SDI to assess agricultural water consumption*. Article long pour la conférence ENVROINFO 2011. 5-7 Octobre, JRC-Ispra, Italie.

[Berge, 2012] BERGE A. *Maquettage d'un Web-SIG dédié à la modélisation des besoins en eau domestique*. Mémoire de Travail de Fin d'Etudes de l'ESGT. Soutenu le 6 juillet 2012. ESGT. 50 pages.

[Berolo, 2010] BEROLO W., Géoazur, UNS, Nice, France. *Estimation des précipitations mensuelles décennales et centennales*. 11 mai 2010. <http://gravitaire.oca.eu/spip.php?article187>

[Botts, 2007] BOTTS M. *OGC White Paper – OGC Sensor Web Enablement: Overview and High Level Architecture*.- Version 3. OGC 07-165. (rev.: 28.12.2007).
<http://www.opengeospatial.org/pressroom/papers>

[BRL, 2009] *Prélèvements mensuels actuels estimés pour l'Alimentation en Eau Potable sur la zone d'étude*. 20 septembre 2009. Fichier excel zippé.
http://agire.brgm.fr/to%20download/Pr%C3%A9l%C3%A8vements%20pour%20l%27AEP_VULCAIN.zip

[Bucker, 2000] BUCKER, M. ; HOVLAND, P. Automatic differentiation. <http://www.autodiff.org/>, 2000, (consulté en 2012).

[Burrough, 1986] BURROUGH, P.A. *Principles of Geographical Information Systems for Land Resources Assessment*. 1986. Oxford, Oxford University Press, 193 p.

[Caballero & al., 2010] CABALLERO Y. *Présentation des résultats finaux du projet à la réunion de clôture*. 9 novembre 2010. Collioure. <http://www.conferenciasequia.es/upload/Caballero.pdf>

[Caballero, 2008] CABALLERO Y. *Vulnerability of Mediterranean hydrosystems to climate changes and human activities: the VULCAIN project*. 13th IWRA World Water Congress 2008, September 1-4, Montpellier, France.
http://agire.brgm.fr/to%20download/Vulcain_WWC_Caballero%20et%20al_Septembre2008.pdf

[Cauchy, 2012] CAUCHY A, *Présentation du standard WPS, implémentations disponibles et exemple d'application au traitement d'images spatiales*, Présentation à Rencontres SIG la Lettre 2012.

[Chaouche, 2010] CHAOUICHE, K. ; NEPPEL, L. ; DIEULIN C. ; PUJOL N. ; LADOUICHE B. ; MARTIN E. ; SALAS Y MELIA, D. ; CABALLERO, Y. (2010), *Analyses of precipitation, temperature and evapotranspiration in a French Mediterranean region in the context of climate change*, Comptes Rendus Geosciences, 342(3), 234-243, doi:10.1016/j.crte.2010.02.001

[Chazot & al., 2010] CHAZOT, S ; TERRASSON, I. Dans [Caballero & al, 2010]

[Chazot & al., 2008] CHAZOT, S ; TERRASSON, I. Dans [Caballero & al, 2008]

[CIEAU, 2011] *Utiliser : Le bon usage de l'eau*. Brochure du CIEAU. Novembre 2011. http://www.cieau.com/launch_pdf.php?uri=images/brochures/pdf/utiliser.pdf

[Clinton, 1994] CLINTON, W J. *Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure*. Executive Order 12906. 11 Avril 1994.

[De Runz, 2008] DE RUNZ, C. 2008. *Imperfection, temps et espace: modélisation, analyse et visualisation dans un SIG archéologique*. Mémoire de thèse. Université de Reims Champagne-Ardenne, France. 2008.

[Commission Européenne, 2001] Commission Européenne. "*Global Monitoring for Environment and Security (GMES) Outline GMES EC Action Plan (Initial Period: 2001 – 2003)*". COM(2001)609 final. 23 Octobre 2001.

[Doorenbos & al, 1979] DOORENBOS, J. ; KASSAM, A. 1979. Yield response to water. *FAO Irrigation and Drainage Paper 33*. Rome.

[DSTO, 1994] DSTO (Defence Science and Technology Organization) Data Fusion Special Interest Group "*Data fusion lexicon*". . 21 September 1994. Department of Defence, Australia, 7 p.

[El Hage, 2012] EL HAGE M. *Etude de la qualité géomorphologique de modèles numériques de terrain issus de l'imagerie spatiale (photogrammétrie et interférométrie radar)*. Mémoire de thèse de doctorat. Soutenue le 12 novembre 2012.

[Er-Raki, 2007] ER-RAKI S. *Estimation des besoins en eau des cultures dans la région de Tensift AL Haouz : Modélisation, Expérimentation et Télédétection*. Mémoire de thèse. Soutenue le 2 mai 2007.

[Geller & al., 2007] GELLER, G. N. ; TURNER, W. (2007), *The model web: a concept for ecological forecasting*, Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, 2469 – 2472, 23-28 July 2007.

[GEOSS, 2005] *The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan*. Février 2005. <http://www.earthobservations.org/documents/10-Year%20Implementation%20Plan.pdf>

[GIEC, 2008] BATES, B.C.; KUNDZEWICZ, Z.W. ; WU, S.; PALUTIKOF, J.P., Eds. *Climate Change and Water*. Technical Paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva, 210p. 2008.

[GUM, 2008] JCGM. *Guide to the expression of uncertainty in measurement (GUM)*. Edition de Septembre 2008.
http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_F.pdf

[Haniquaut, 2005] HANIQUAUT, C. *Les économies d'eau au niveau des usages de type domestique, dans le cadre individuel ou collectif*. Rapport de stage. Juillet 2005.

[INSPIRE, 2007] *DIRECTIVE INSPIRE*. Journal Officiel de l'Union Européenne. 14 mars 2007.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:fr:PDF>

[ISO, 2005] ISO 9000:2005. *Systèmes de management de la qualité – Principes essentiels et vocabulaire*. 2005.
http://www.iso.org/iso/fr/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=42180

[Jackson, 2008] JACKSON, I. *OneGeology – An overview*. Présentation lors de la réunion de lancement. Juillet 2008. Brighton, UK.

[Jirka, 2010] JIRKA, S. *Coupling 52° North's SOS and Geoserver's WFS*, Presentation at the FOSS4G 2010 conference, September 6th-9th, Barcelona, Spain.

[Juric, 2005]. JURIC, M. *Business Process Execution Language for Web Services*. 2006. 372 pages.

[Khalsa, 2009] KHALSA, S.J., NATIVI, S., GELLER, G. (2009). *The GEOSS Interoperability Process Pilot Project (IP3)*. IEEE Transactions on Geoscience and Remote Sensing, Special issue on data archiving and distribution, 47(1):80-91

[Klien, 2009] KLIEN, E ; ANNONI, A ; MARCHETTI, P G. *The GIGAS project – an action in support to GEOSS, INSPIRE, and GMES* Dans : J. Hřebíček, J. Hradec, E. Pelikán, O. Mírovský, W. Pillmann, I. Holoubek, T. Bandholtz (eds.): Proceedings of TOWARDS eENVIRONMENT: Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe. 2009.

[Loucks & al., 2005] LOUCKS, D.P. ; VAN BEEK, E. *Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications*, 2005, UNESCO, Paris.

[Lowe & al, 2011] LOWE, D. ; WOOLF, A. *CSML3 : Climate Science Modelling Language – MetOcean DWG*. Présentation lors du 76ème Comité technique de l'OGC. 1-4 Mars 2011. Bonn, Allemagne.

[Martin, 2005] MARTIN, A. *La fusion d'informations*. Polycopié de cours ENSIETA. Janvier 2005. <http://martin.iutlan.univ-rennes1.fr/Doc/fusion.pdf>

[Masser, 2009] MASSER I. *GIS Worlds: Creating Spatial Data Infrastructures*. 2005 Redlands: ESRI Press.

[Mirad, 2004] MIRAD, B. *La consommation et les usages domestiques de l'eau dans les Pays-Coeur-Entre-deux-Mers*. Travail de fin d'étude. Avril-Août 2004.

[Neun & al, 2012] NEUN, M., REGNAULD, N. & WEIBEL, R.: *WebGen: des services web pour partager des outils de généralisation cartographique*. Dans : Bucher, B. & Le Ber, F. (eds.): *Développements logiciels en géomatique – innovations et mutualisation*. Cachan, France: Lavoisier, 239-260. (2012).

[NRC, 1994] NRC (National Research Council). 1994. *Science and judgment in risk assessment*. Washington, D.C., National Academy Press.

[OASIS, 2007] *Web Services Business Process Execution Language Version 2.0*. 11 avril 2007. <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.pdf>

[OGC, 2005] *Web Feature Service Implementation Specification (v1.1 04-094)*. 131 pages.

[OGC, 2006] *Web Map Server Implementation Specification* (v1.3.0 06-042). 85 pages.

[OGC, 2007] *Sensor Observation Service Implementation Specification* (v1.0 06-009r6). 104 pages.

[OGC, 2007] *OpenGIS Geography Markup Language (GML) Encoding Standard Specifications* (v3.2.1 07-036). 437 pages.

[OGC, 2007 bis] *Web Processing Service Implement Specification* (v1.0.0 05-007r7). 87 pages.

[OGC, 2008] *Uncertainty Markup Language (UnCertML) Discussion Paper* (v0.6 08-122r2). 61 pages.

[OGC, 2008] *OpenGIS® City Geography Markup Language (CityGML) Encoding Standard Specification* (v1.0 08-007r1). 234 pages.

[OGC, 2009] *Web Coverage Service Implement Specification* (v2.0.0 09-110r3). 53 pages.

[OGC, 2009] OGC. *Request for Information (RFI) for OGC Fusion Standards Study*. Juillet 2009.
http://portal.opengeospatial.org/files/?artifact_id=34909&format=pdf

[OGC, 2010] PERCIVALL, G. *Fusion standards study engineering report*. OGC. 21 mars 2010.

[OGC, 2010] *Observations and Measurements - XML Implementation Specification* (v2.0 10-025r1). 2010. 76 pages.

[OGC, 2010] *OGC® WaterML 2.0: Part 1- Timeseries* (v2.0 10-126r3). 149 pages.

[ONU, 2000] Assemblée Générale des Nations Unies. *Déclaration du Millénaire*. 13 Septembre 2000.
<http://unpan1.un.org/intradoc/groups/public/documents/un/unpan004147.pdf>

[OSGEO, 2011] OSGEO. *WMS Benchmarking 2011*. Conférence FOSS4G 2011. Denver.
<http://www.slideshare.net/gatewaygeomatics.com/wms-performance-shootout-2011>

[Pinault, 2001] PINAULT, J.L. *Manuel utilisateur de TEMPO. Logiciel de traitement et de modélisation des séries temporelles en hydrogéologie et en hydrogéochimie*. 2001. Rapport BRGM/RP-51459-FR.

[Prétot, 2010] PRETOT, J. *STRATEAU : Outil de modélisation Demande / ressources EAU*. Présentation lors de la 3^{ème} Semaine de l'eau de Beyrouth. 24-27 octobre 2010.

[Saarenmaa & al, 2002] SAARENMAA, H., MARTIN, J., JENSEN, S., PEIFER, H. & MCINNES, G. 2002. *Development of common tools and a shared information infrastructure for the European Environmental Information System*. Rapport technique. Vol. 83, 36 p. Agence Européenne pour l'Environnement, Copenhague.

[Serrano & al, 2006] SERRANO, J-J ; ROBIDA, F. *Contribution of GeoScienceML to the INSPIRE data harmonisation process for Geology*. Papier court. International Association for Mathematical Geology XIth International Congress. 3-8 septembre 2006. Liège, Belgique.

[Souriau, 2011] SOURIAU, J. *Les principaux déterminants de la consommation d'eau à Paris : état des lieux*. Doctorat Eau de Paris – CIRED. Soutenue en Juillet 2011.

[Stollberg & al, 2007] STOLLBERG, B ; et ZIPF A, i3mainz - Institute for Spatial Information and Surveying Technology. "*OGC Web Processing Service Interface for Web Service Orchestration: Aggregating geo-processing services in a bomb threat scenario*", "Web and wireless geographical information systems: 7th international symposium", By J. Mark Ware, George E. Taylor, 2007.

[US DoD, 1991] United States Department of Defence, Data Fusion Subpanel of the Joint Directors of Laboratories, Technical Panel for C3, 1991, "Data fusion lexicon".

[Wald, 1998] WALD, L. "A European proposal for terms of reference in data fusion". International Archives of Photogrammetry and Remote Sensing, Vo.. XXXII, Part 7, pp. 651-654, 1998.

[Williams & al, 2008] WILLIAMS, M., CORNFORD, D., BASTIN, L. and INGRAM, B. R., 2008. *UncertML: an XML schema for exchanging uncertainty*. Présentation lors de GISRUK2008, Manchester, UK 2-4 April.

Liste des publications

[Beaufils & al, 2011] BEAUFILS Mickaël, FOLLIN Jean-Michel. *Geospatial Data Fusion for Water Resources Assessment*. Poster + résumé pour la conference AGILE 2011. 18-21 Avril, Utrecht, Pays-Bas.

[Beaufils & al, 2011 bis] BEAUFILS Mickaël, FOLLIN Jean-Michel. *Toward a quality aware SDI to assess agricultural water consumption*. Article long pour la conférence ENVROINFO 2011. 5-7 Octobre, JRC-Ispra, Italie.

Mickaël BEAUFILS

Fusion de données géoréférencées et développement de services interopérables pour l'estimation des besoins en eau à l'échelle des bassins versants

Résumé

De nos jours, la préservation de l'environnement constitue un enjeu prioritaire. La compréhension des phénomènes environnementaux passe par l'étude et la combinaison d'un nombre croissant de données hétérogènes. De nombreuses initiatives internationales (INSPIRE, GEOSS) visent à encourager le partage et l'échange de ces données.

Dans ce sujet de recherche, nous traitons de l'intérêt de mettre à disposition des modèles scientifiques sur le web. Nous montrons l'intérêt d'utiliser des applications s'appuyant sur des données géoréférencées, et présentons des méthodes et des moyens répondant aux exigences d'interopérabilité.

Nous illustrons notre approche par l'implémentation de modèles d'estimation des besoins en eau agricoles et domestiques fonctionnant à diverses échelles spatiales et temporelles. Un prototype basé sur une architecture entièrement orientée services web a été développé. L'outil s'appuie sur les standards Web Feature Service (WFS), Sensor Observation Service (SOS) et Web Processing Service (WPS) de l'OGC.

Enfin, la prise en compte des imperfections des données est également abordée avec l'intégration de méthodes d'analyse de sensibilité et de propagation de l'incertitude.

Mots clés : Fusion de données, Web SIG, Open Geospatial Consortium, Modèle web, Services Web, Infrastructure de Données Spatiales, Web Processing Service, capteurs virtuels, architecture orientée service, analyse de sensibilité, propagation de l'incertitude.

Résumé en anglais

Nowadays, preservation of the environment is a main priority. Understanding of environmental phenomena requires the study and the combination of an increasing number of heterogeneous data. Several international initiatives (INSPIRE, GEOSS) aims to encourage the sharing and exchange of those data.

In this thesis, the interest of making scientific models available on the web is discussed. The value of using applications based on geospatial data is demonstrated. Several methods and means that satisfy the requirements of interoperability are also purposed.

Our approach is illustrated by the implementation of models for estimating agricultural and domestic water requirements. Those models can be used at different spatial scales and temporal granularities. A prototype based on a complete web service oriented architecture was developed. The tool is based on the OGC standards Web Feature Service (WFS), Sensor Observation Service (SOS) and Web Processing Service (WPS).

Finally, taking into account the imperfections of the data is also discussed with the integration of methods for sensitivity analysis and uncertainty propagation.

Keywords : Data fusion, Web GIS, Open Geospatial Consortium, model web, web services, Spatial Data Infrastructure, Web Processing Service, virtual sensors, service oriented architecture, sensitivity analysis, uncertainty propagation.