

# Gestion et valorisation des données géographiques et non géographiques : importance de la qualité des données et de l'information

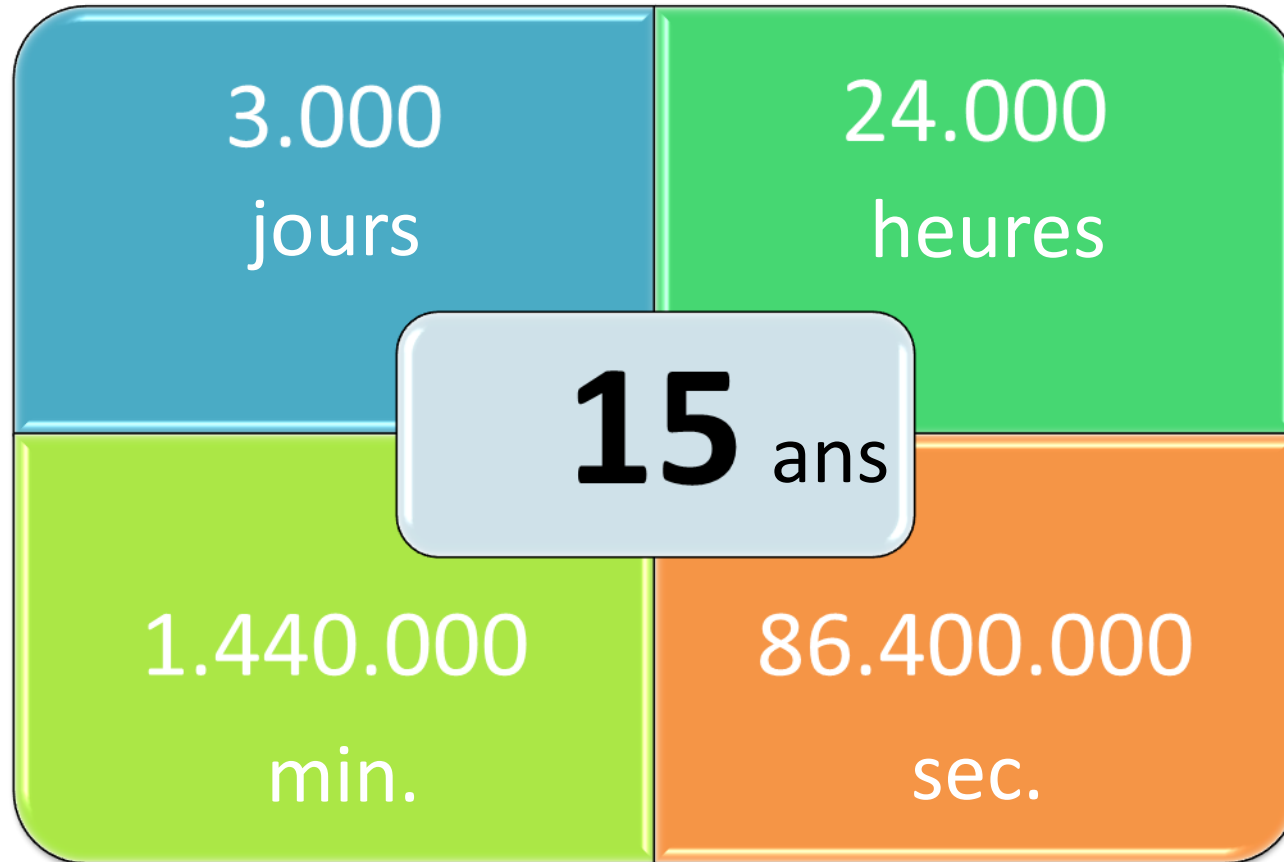
**Centre wallon de Recherches agronomiques**

**Département Agriculture et milieu naturel**

*Unité Systèmes agraires, Territoire et Technologies de l'information*

**Dominique BUFFET**

# Données, Information, Connaissance...

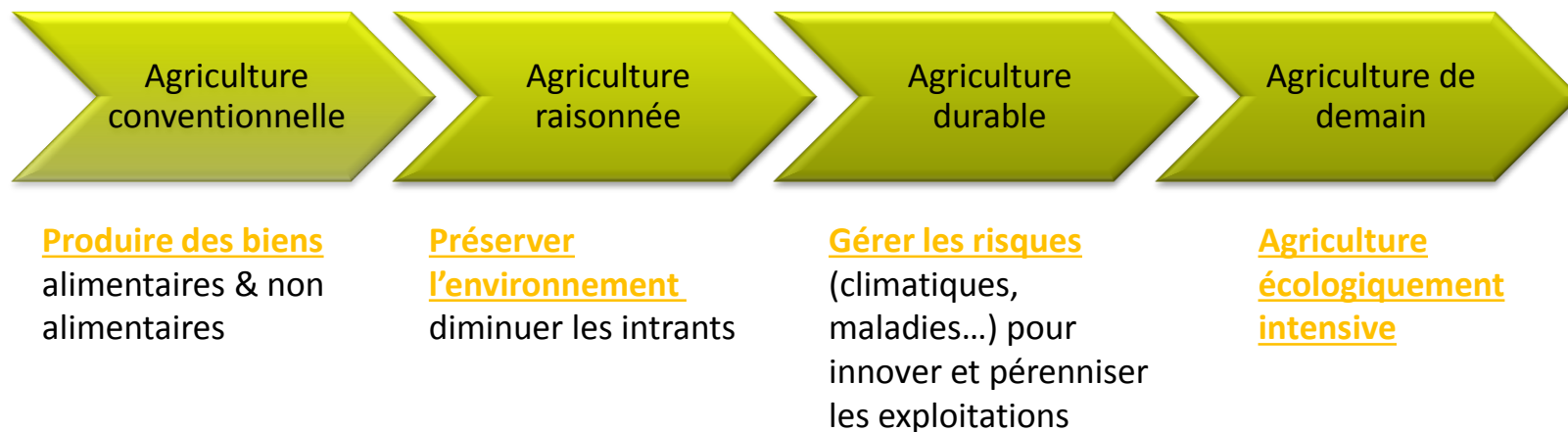


# Contexte

- La **découverte de connaissances** et la **prise de décision** à partir de données de qualité médiocre (*c-a-d contenant des erreurs, doublons, incohérences, valeurs manquantes...*) ont des conséquences directes et significatives quelques soit le domaine et l'utilisateur final (*agronomes, scientifiques, pouvoirs publiques, les citoyens...*).

(Laure Berti-Equile – 2010)

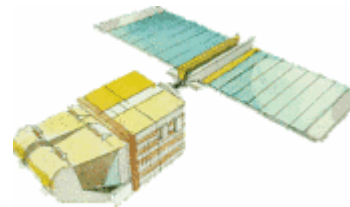
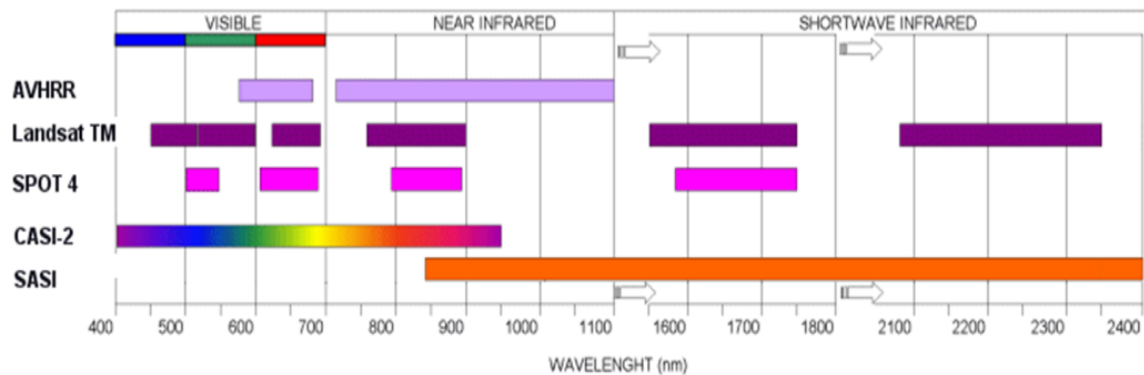
## L'évolution de l'agriculture belge et européenne



# Contexte

## L'évolution accélérée des technologies

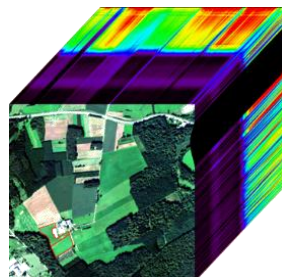
- La création des données - L'acquisition des données



SPOT-VEGETATION



SPOT 5



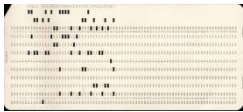
Une image satellitaire : c'est grand

Image de télévision	860 x 623 pixels
Écran de PC	1 600 x 1 200 pixels
Appareil photo	4 000 x 3 000 pixels
Image Spot 1 ... 4	6 000 x 6 000 pixels
Image SPOT 5	24 000 x 24 000 pixels
Image Quickbird	40 000 x 40 000 pixels
1 600 000 000 pixels = 1,6 Gpixels	
800 écrans de PC	



## ● L'évolution accélérée des technologies

- Le stockage des données et l'échange d'information



D'un stockage limité et peu accessible  
à des SGBD dédiés



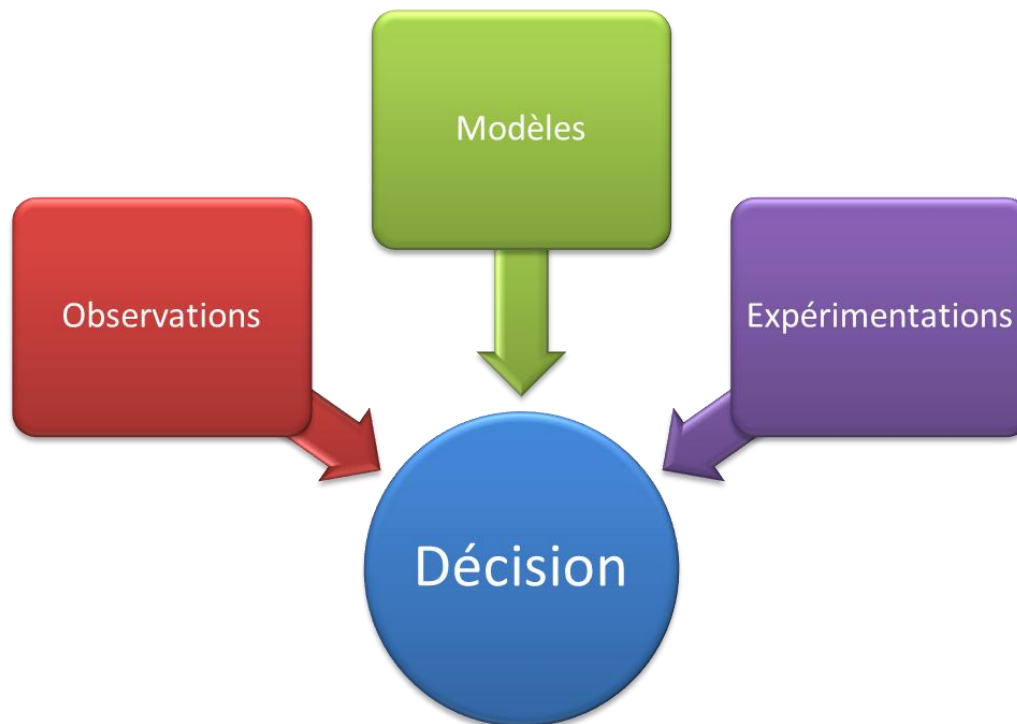
La mobilisation de compétences  
scientifiques et techniques nouvelles



Infrastructures informatiques

## ● L'évolution de la complexité des systèmes étudiés

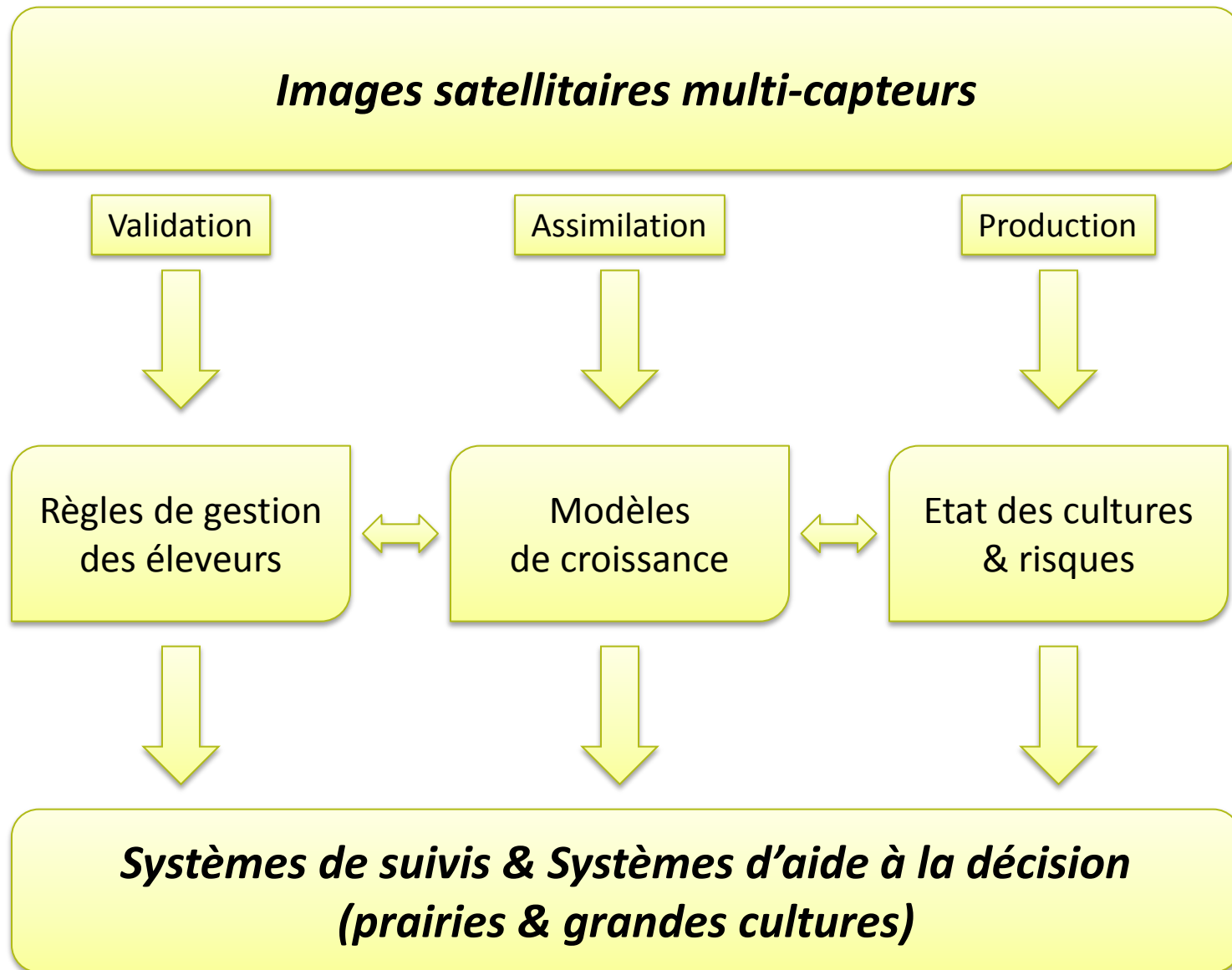
- L'intégration de données et d'information multi-sources



Le couplage de modèles et de données de différentes natures (biologiques, sociologiques, économiques...) pour concevoir, évaluer et guider les pratiques agricoles dans le contexte de changement globaux.



# Valorisation de l'information satellitaire à l'échelle régionale et locale

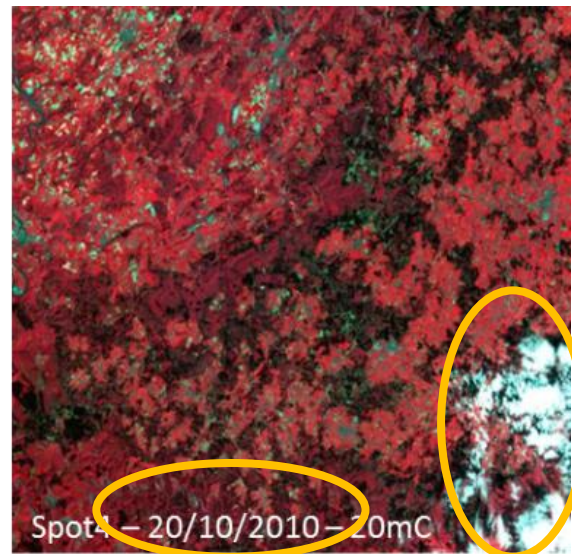
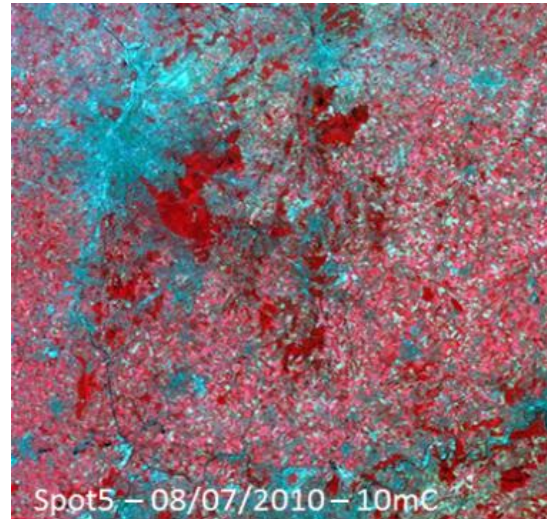


# Qualité des données

Réalité terrain



Données satellitaires



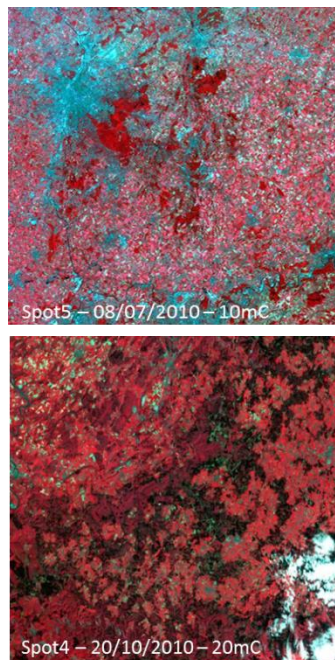


# Qualité de l'information

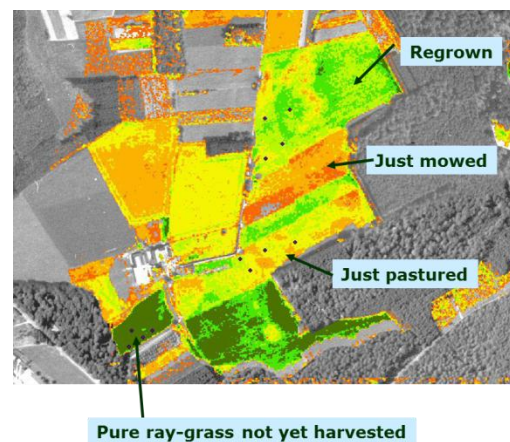
Réalité terrain



Données satellitaires



Informations dérivées  
(ex: indices, classification)



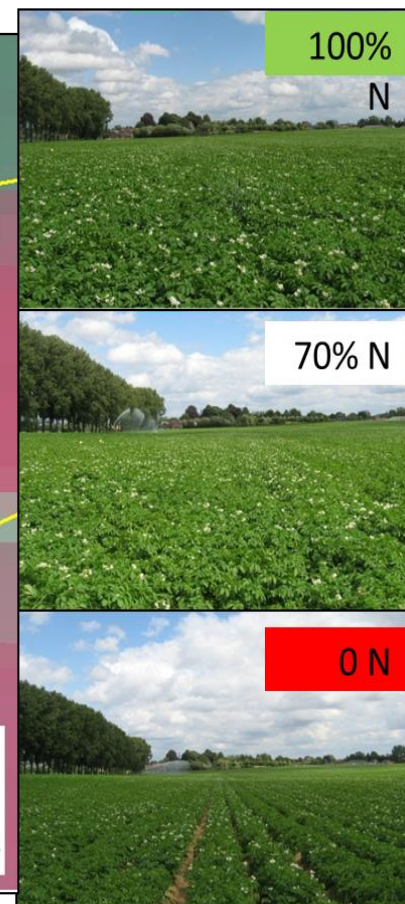
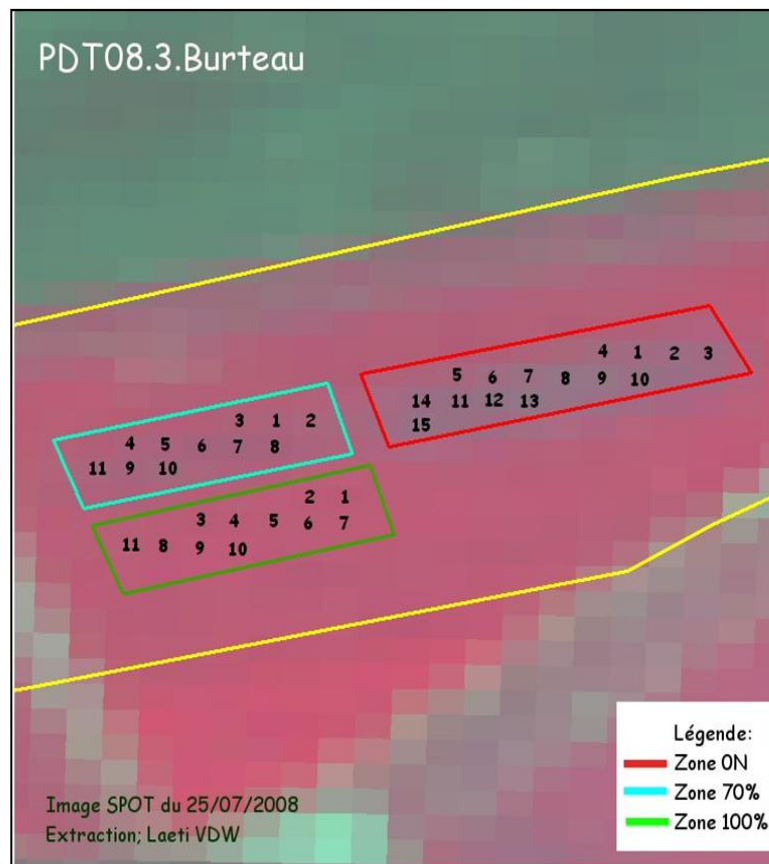
Classification results	Observed management classes		
	F	NP NF	P
F	13	0	1
NP NF	2	3	1
P	0	0	4
N total	15	3	6
N correct	13	3	4
Proportion	86,70%	100%	66,70%



# Exemple de valorisation

## Gestion des apports d'engrais azotés (Projet MIMOSA)

### Méthodes classiques



(J.P Goffart CRA-W)

# Outil d'aide à l'échantillonnage

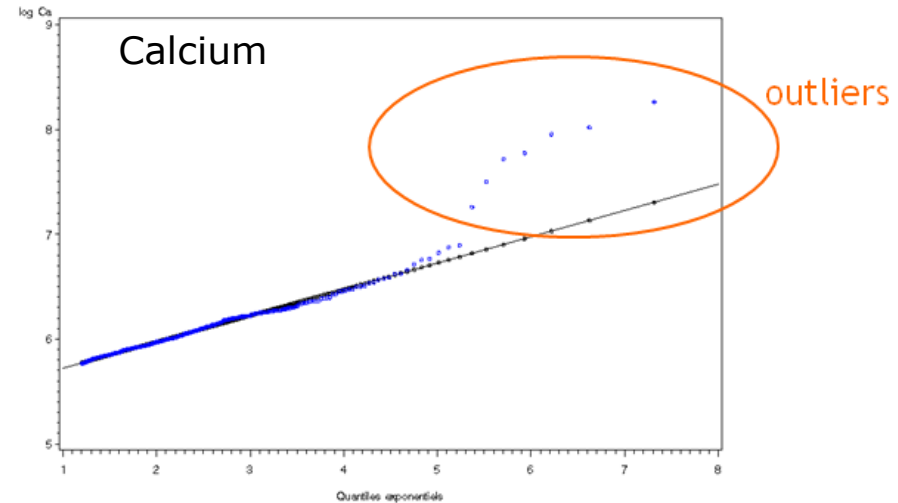
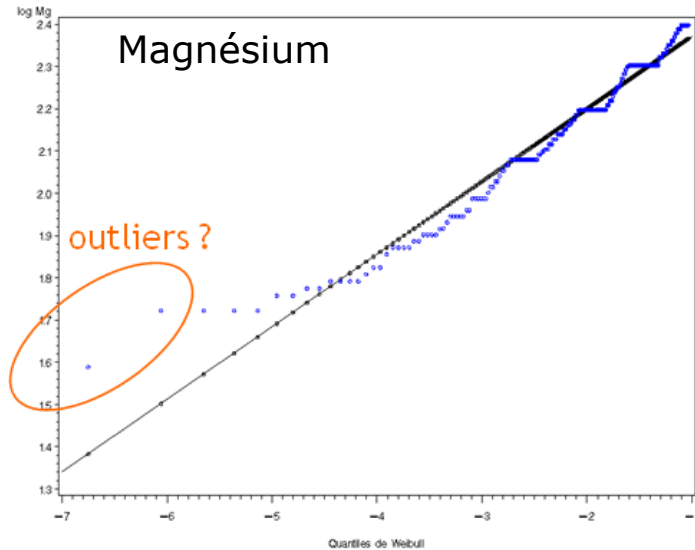
## ● Base de données échantillons de terre

- 300.000 échantillons de terre (>20.000 échantillons/an)
- Outil d'aide à l'échantillonnage => Géoportail REQUACARTO
- Données de référence => Géoportail REQUASOL





## Données aberrantes (outliers)



Grands ensembles de données :

- Présence de valeurs aberrantes (extrêmes, suspectes, contaminantes)
- Estimations biaisées des paramètres des populations .
- Interprétation erronée des résultats.

## ● Représentativité/reproductivité des données (échantillonnage)

Zone de prélèvement	pHKCl	COT	NT	CEC	Ca disp	Mg disp	K disp	P disp
		%	g/kg	cmol+/kg		mg/100g		
Parcelle	7.1 – 7.2	1.15 – 1.35	0.12 – 0.13	13.0 – 13.2	<b>273 – 306</b>	10.7 – 11.8	23.5 – 27.0	<b>9.1 – 10.0</b>
17739107-B	7.0 – 7.0	1.15 – 1.12	0.12 – 0.13	12.9 – 12.9	284 – 273	11.1 – 11.7	25.5 – 27.0	9.1 – 9.0
17739107-C	<b>7.6 – 7.6</b>	<b>0.90 – 1.05</b>	0.08 – 0.09	<b>9.0 – 9.3</b>	<b>920 – 935</b>	14.1 – 14.6	18.5 – 19.5	8.2 – 8.3
17739107-D	<b>7.4 – 7.3</b>	<b>1.00 – 1.15</b>	0.10 – 0.11	<b>10.6 – 11.3</b>	319 – 265	9.8 – 10.3	20.5 – 22.0	<b>6.1 – 6.7</b>

Résultats des analyses de terre pour la parcelle 17739107 illustrant :

- la faible variabilité des analyses intra-zone de prélèvement et
- la forte variabilité entre zones de prélèvement qui se trouvent gommées si on réalise un seul échantillon pour l'ensemble de la parcelle.



Normes en matière d'échantillonnage  
Outil pour déterminer les Zones de prélèvement homogènes,  
objectives d'une façon uniformisée



## Données incohérentes ou contradictoires

9, Rue de Liroux, 5030 Gembloux

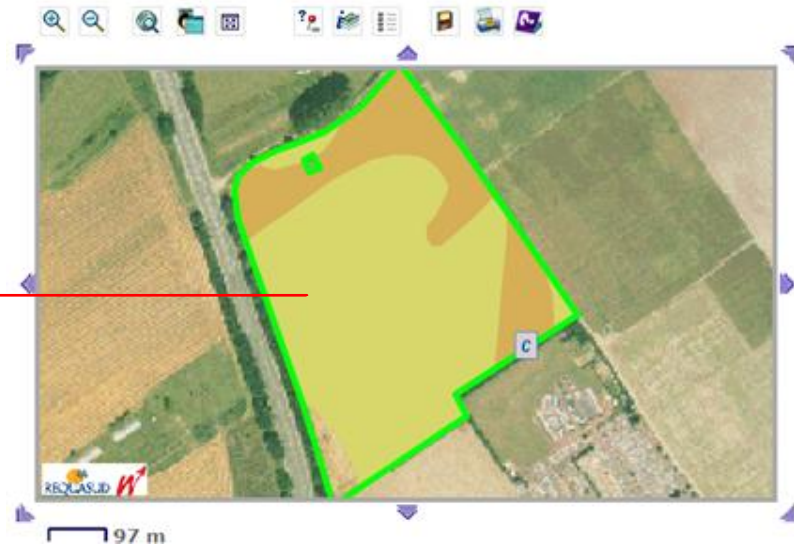


Tableau d'interprétation (cliquer sur l'icône pour télécharger le fichier [XML](#)).

Label	Geo-ID	Occupation	Texture	Drainage	Charge		Substrat		Divers	Coord. Centroïde	
		(%)			Nature	Abondance	Nature	Prof. sol		X (m)	Y (m)
A	115407-A	100	-	-	-	-	-	-	Parcelle	200872	109839
B	115407-B	54.3	-	Excessif à imparfait	Psammite	-	Argile ou argile sableuse	-	-	200852	109795
C	115407-C	27.9	-	Excessif à imparfait	Psammite	-	-	-	-	200964	109781
D	115407-D	13.7	-	Excessif à imparfait	Psammite	-	Argile ou argile sableuse	-	-	200956	109920

- Incohérence (spatiale, temporelle...)
- Contradiction des informations.

# Suivi de l'état des cultures



## BD Agrométéorologique

- Plus de **7Gb** de données dupliquées sur 2 serveurs
- **25 années** de données journalières spatialisées pour toute la Belgique (maillage de 10x10Km)
- **3.293.370 lignes** d'enregistrement pour différents paramètres (Tmin, Tmax, Préc., Vitesse du vent, Pression de vapeur, ESO, ET0)
- 4.177 unités de sol cartographiques
- 7.219 EMU = **180.475 valeurs** de rendements estimés



# Qualité des données

Valeur incorrecte  
(pb de conformité)

Doublons  
(pb d'unicité)

Valeur incohérente par rapport au champ  
(pb de cohérence)

Champ inutile

GRID_NO	DAY	MAXIMUM_TEMPERATURE	MINIMUM_TEMPERATURE	AVERAGE_TEMPERATURE	VAPOUR_PRESSURE	WINDSPEED	RAINFALL	EO	ESO	ETO
1	01-janv-12	12,6	9,2	10,91	11,60	6,0000000	4,1	0,99	0,99	1,07
1	02-janv-12	13,8	7,6	10,71	10,12	4,8000002	16,0	0,55	0,50	0,62
500	03-janv-12	5,5	-0,6	2,43	7,86	8,8999996	3,6	1,19	1,16	1,28
1	04-janv-12	11,0	32,0	21,50	8,21	7,3000002	12,2	1,21	1,17	1,34
1	04-janv-12	10,7	6,8	8,76	8,09	10,8000002	9,2	1,55	1,52	1,65
1	06-janv-12	10,5	5,0		8,17	7,0000000	4,2	0,93	0,89	1,02
1	2012-01-08	8,8	4,9	6,86	7,32	7,0999999	-99999,0	0,75	0,70	0,81

Typos-format  
(pb de conformité)

Valeur manquante  
(pb d'intégrité  
référentielle)

Hors nomenclature  
(pb de conformité et d'intégrité)

Valeur manquante  
(pb de complétude)

# Qualité des informations

Pomme de terre - Bintje							
Région agricole	2011 (INS)	Model1	Model2	Model1/2011 (%)	Model2/2011 (%)	Model1/moy (%)	Model2/moy (%)
3 Région sablonneuse	528,3	510,1	528,7	-3,5	0,1	6,4	10,3
4 Campine	451,3	460,1	442,8	2,0	-1,9	1,7	-2,1
5 Région sablo-limoneuse	538,3	518,1	524,8	-3,8	-2,5	8,1	9,5
6 Région Limoneuse	535,2	515,6	522,3	-3,7	-2,4	6,8	8,2
7 Campine Hennuyère	530,5	511,8	512,4	-3,5	-3,4	7,4	7,5
8 Condroz	531,8	533,0	539,8	0,2	1,5	8,8	10,2
9 Région herbagère	542,0	606,6	607,5	11,9	12,1	13,9	14,1
10 Région herbagère (Fagne)	500,0	535,9	542,8	7,2	8,6	4,1	5,4
11 Famenne	528,0	466,1	460,7	-11,7	-12,7	4,6	3,4
12 Ardenne	567,5	350,3	356,5	-38,3	-37,2	-0,4	1,4
13 Région Jurassique	255,0	302,1	302,2	18,5	18,5	-1,0	-1,0
14 Haute Ardenne	470,0	458,5	459,1	-2,4	-2,3	1,0	1,1
15 Dunes&Polders	503,6	500,6	507,3	-0,6	0,7	10,8	12,3
<b>Belgique</b>	<b>499</b>	<b>482</b>	<b>485</b>	<b>-2</b>	<b>-2</b>	<b>6</b>	<b>6</b>

Model1	CGMS-TREND	WLYB_24	TREND_98_11	
Model2	CGMS-RS-TREND	WLYB_24	OX1624	TREND_98_11

RMSE	R2 aj
54,6	0,57
54,8	0,56



## Analyse de l'incertitude

# Exemple de valorisation

## Evaluer les risques climatiques et suivre les impacts sur l'agriculture

### Bulletin Spécial Sécheresse Situation au 26 mai 2011

#### Résumé

Depuis le début du mois de mars, les précipitations sont déficitaires sur l'ensemble du royaume de Belgique. La situation est tout à fait exceptionnelle et fait ressembler l'année 2011 à l'année 1996, considérée comme l'année la plus sèche depuis 1988, début de la période considérée par l'équipe qui réalise ce bulletin à l'échelle du pays. Ce bulletin fournit certaines informations chiffrées permettant d'objectiver le niveau de la sécheresse actuelle. Il ne fait aucun doute que les prairies et les cultures vont être affectées par ces conditions climatiques particulièrement rares. L'ampleur des dégâts causés par cette sécheresse va dépendre de sa durée.

#### Objectifs

Ce bulletin spécial est réalisé suite aux conditions exceptionnelles de sécheresse observées sur l'ensemble du pays. Il vise à donner quelques informations chiffrées sur l'état de la situation agrométéorologique. Les variables explicatives utilisées proviennent de trois sources d'information : données météo, données agrométéorologiques issues du modèle B-CGMS (Belgian Crop Growth Monitoring System) et imageries satellitaires. Il est réalisé à la demande des cabinets des ministères de l'agriculture flamande et wallonne.

#### Situation météorologique

Depuis le début du printemps météorologique (1<sup>er</sup> mars), la Belgique subit de plein fouet une vague de sécheresse exceptionnelle. Lors de notre dernier bulletin du 10 mai, nous avions déjà annoncé le risque de sécheresse pour les cultures. La prolongation de la sécheresse jusqu'au 31 mai permet cette fois d'affirmer le caractère exceptionnel de la situation.

Si l'on se réfère aux sommes des précipitations depuis le début du printemps météorologique qui coïncide avec la reprise de la végétation, les volumes précipités sont très largement inférieurs à la situation de référence que nous estimons à partir des précipitations moyennes sur la période 1988-2010. En moyenne sur les 15 premières décades de l'année, les précipitations correspondent à la moitié (52%) des précipitations observées habituellement. Les précipitations observées à l'échelle du pays en mars, avril et mai, soit depuis la reprise de la végétation, ne représentent que 30% de la situation normale (49 mm pour 161 mm en moyenne) ce qui va inévitablement affecter les rendements des cultures et la production de biomasse prairiale.

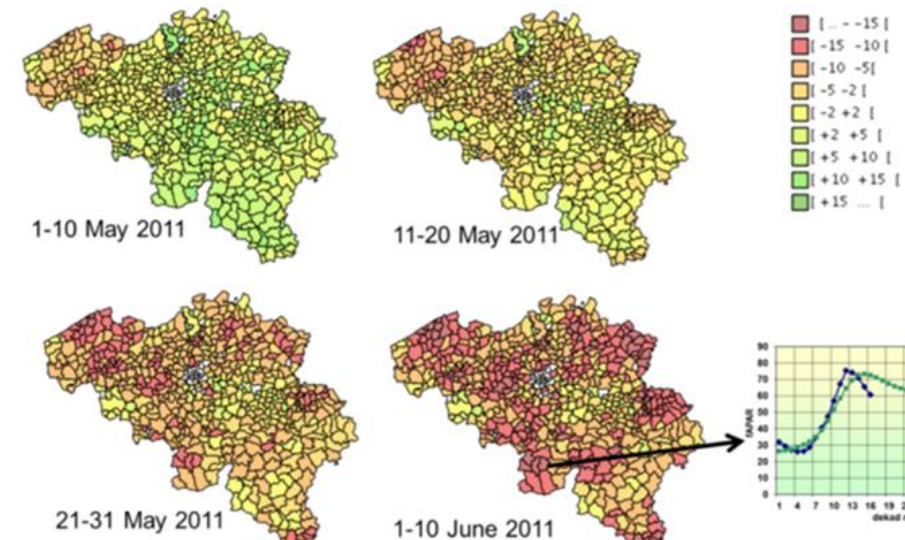
La première carte (Figure 1a) montre les variabilités spatiales de la baisse du total des précipitations sur la période de mars à mai exprimée en pourcentage par rapport à la moyenne calculée sur la période 1988-2010. Les baisses sont très importantes partout mais le Hainaut, la Flandre Occidentale et la Flandre Orientale sont les plus touchées. La seconde carte (Figure 1b) exprime les mêmes résultats mais en termes de déficits pluviométriques absolus (mm). Des déficits de l'ordre de 150 à 200 mm sont observés dans l'est et le sud du pays sur cette période de 3 mois, ce qui est très exceptionnel.

L'analyse temporelle sur les 15 premières décades montrent la dégradation de la situation à partir de mars pour deux régions agricoles (région herbagère et région limoneuse) en valeurs absolues (figure 1c et d) et cumulées (figure 1e et f). Les autres régions agricoles montrent des comportements similaires disponibles en annexe 1

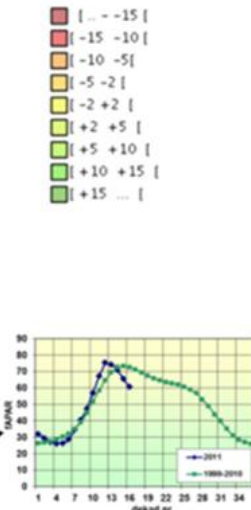
## Anomaly maps

Comparison of actual indicators with long term average.

### "fAPAR relative differences maps"



### Grassland 2011





# Exemple de valorisation

## Une plateforme dédiée



Situation au 1er septembre 2012

### PARAMÈTRES AGROMÉTÉOROLOGIQUES

Les mois de juillet et août furent contrastés au niveau des précipitations. Juillet fut anormalement pluvieux, août anormalement sec, l'un compensant l'autre. Les conditions de températures au cours de ces deux mois furent par contre normales. La courte période chaude et sèche...

### INFORMATION SATELLITAIRE

Comme lors du bulletin précédent, l'état de la végétation dérivé de la télédétection montre globalement une année très normale en termes de développement de la végétation. Les écarts de la fraction du rayonnement photosynthétiquement actif absorbé (FAPAR) par rapport à la période de référence (1999-2011) restent quasi partout entre +5 et -5%...

### SITUATION DES CULTURES

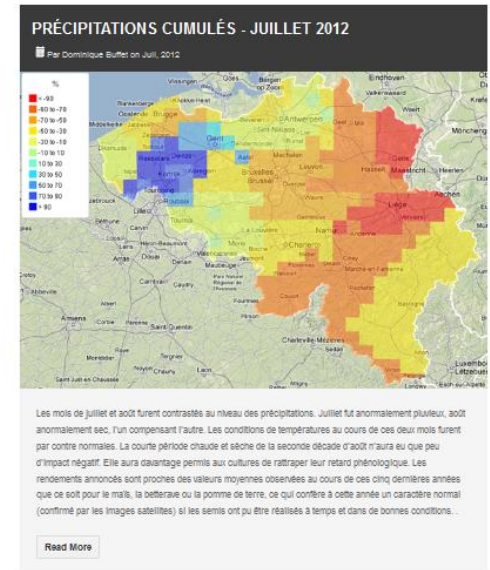
D'une façon générale, on observe un retard de développement plus ou moins contrasté en fonction des dates des semis qui ont été perturbés par les précipitations supérieures à la normale du mois d'août. À ceci vient s'ajouter les températures exceptionnellement basses du mois de mai qui ont provoqué des levées tardives...

### PRÉVISIONS DES RENDEMENTS

Pour la culture de betterave, les prévisions basées sur des dates de semis normales sont annoncées légèrement supérieures aux rendements moyens en biomasse des cinq dernières années (cf. Tableau 1). Nous nous situons cependant très loin de l'année 2011. Pour ce qui est du rendement en sucre et des conditions de récolte, les conditions météorologiques à venir seront déterminantes.

#### ⚠ AVERTISSEMENTS

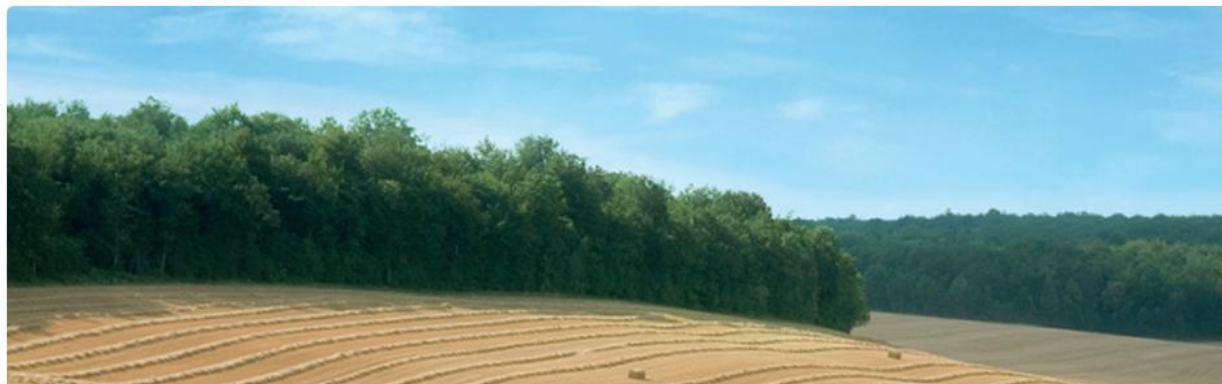
ble et La Belgique est ainsi (avec le Nord/Pas-de-Calais) une des « ré les plus « tardives » en plantation de pomme de terre cette année — Source Phvap



- Pollution de l'eau par les pesticides (projet PESTEAX)
  - Système d'Information à Référence Spatiale de **15 Gb** de données
  - Couche SOLESUEROS contenant **51.851.300 objets spatiaux** avec un système d'indexation spatial faisant 3.519 Mb



 Pesteaux



Evaluer le risque de pollution par les pesticides

L'outil PESTEAX vous permet rapidement d'estimer le risque de pollution des eaux pour une zone déterminée.

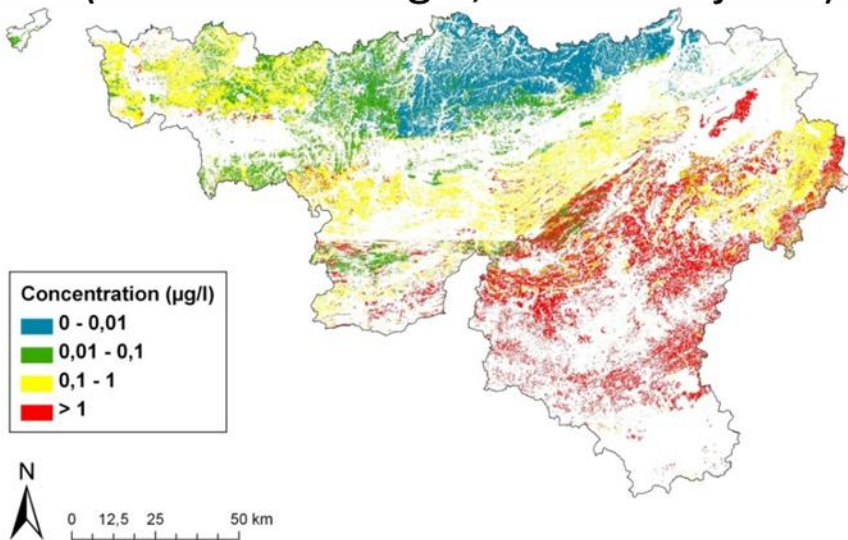
Evaluation du risque de pollution diffuse des eaux par l'usage des pesticides à l'échelle de la parcelle ou d'un territoire

## ☐ Sensibilité des sols aux transferts des PPP

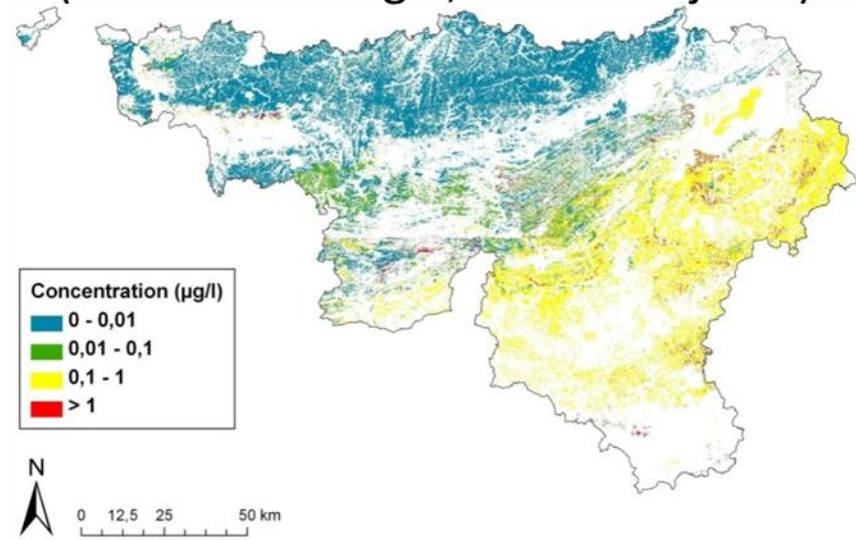
### Eaux souterraines - Approche mécaniste

#### PPP lixivié à 1 m [ $\mu\text{g.l}^{-1}$ ] - Exemples

Pesticide B – Automne  
(Koc :  $17 \text{ dm}^3.\text{kg}^{-1}$  ; DT50 : 20 jours)



Pesticide B – Printemps  
(Koc :  $17 \text{ dm}^3.\text{kg}^{-1}$  ; DT50 : 20 jours)

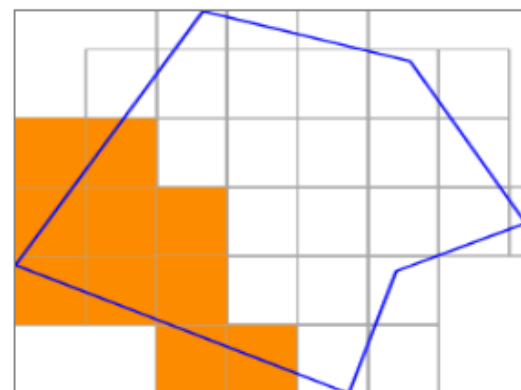


# Qualité de l'information

Risque pour les eaux  
souterraines

Pratiques 0,7331651

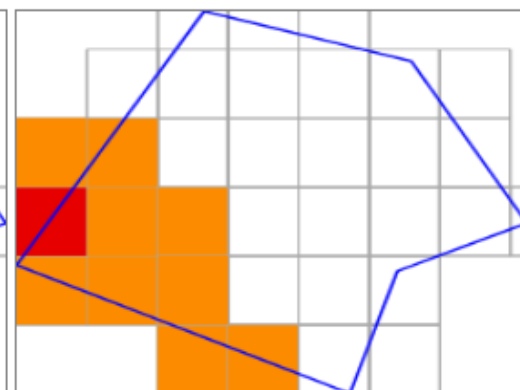
Pesticide 0,6863636



Risque pour les eaux  
de surface (érosion)

Pratiques 0,6581652

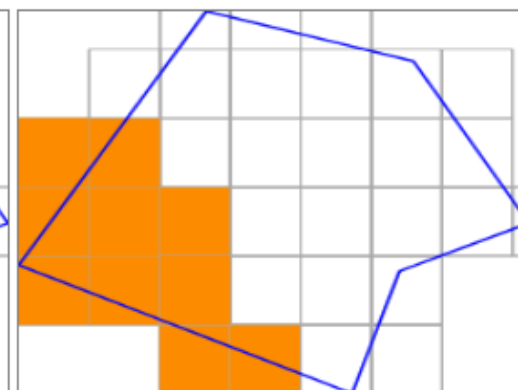
Pesticide 0,185641



Risque pour les eaux  
de surface (ruissellement)

Pratiques 0,6581652

Pesticide 0,6758975



Homogénéité

de la  
parcelle

Elevée

Elevée

Elevée

Sensibilité

de la  
parcelle

modérément

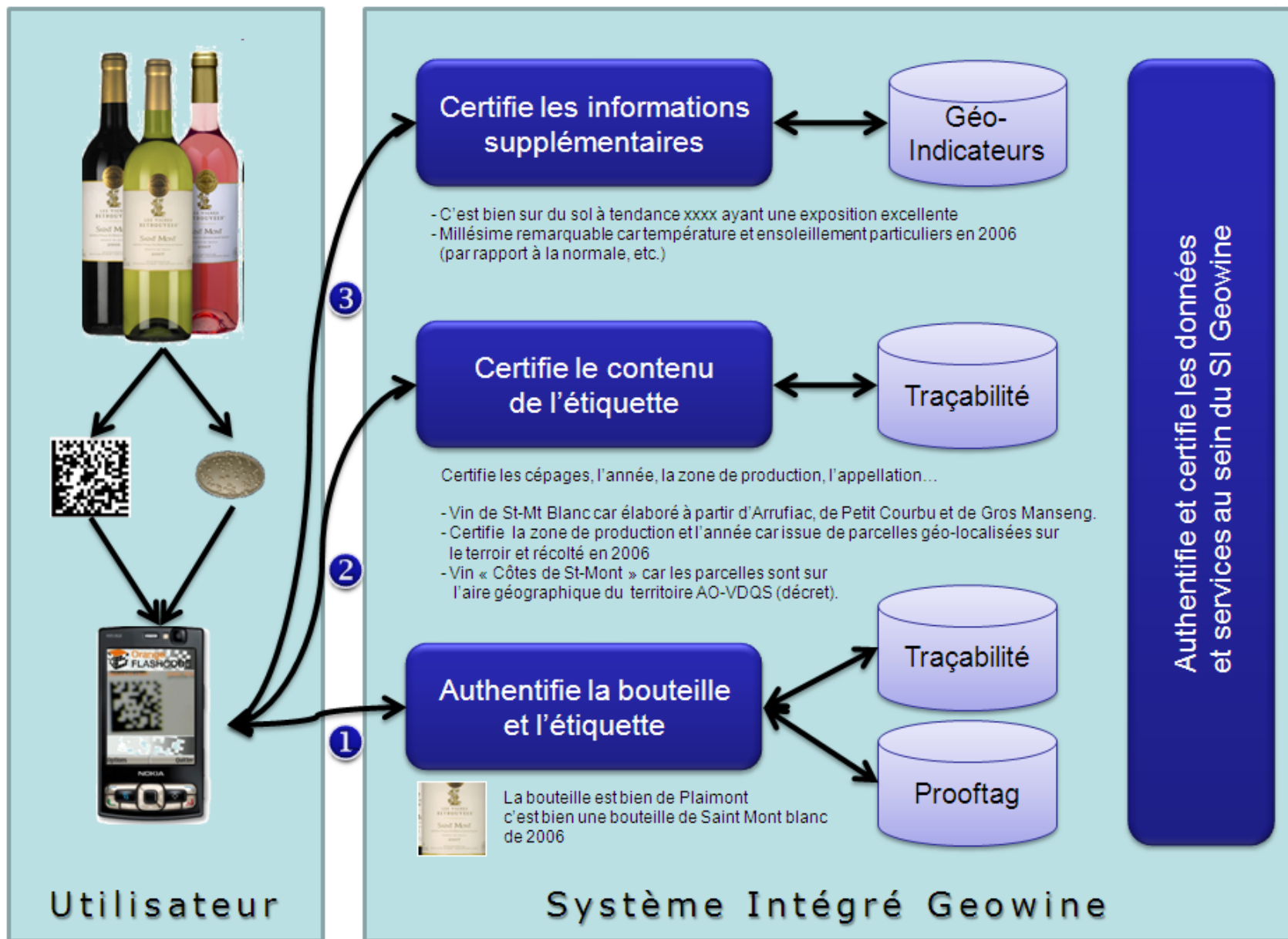
modérément

modérément

Grille de 98 x 98 m



# Infrastructures de données







## Sécurité informatique

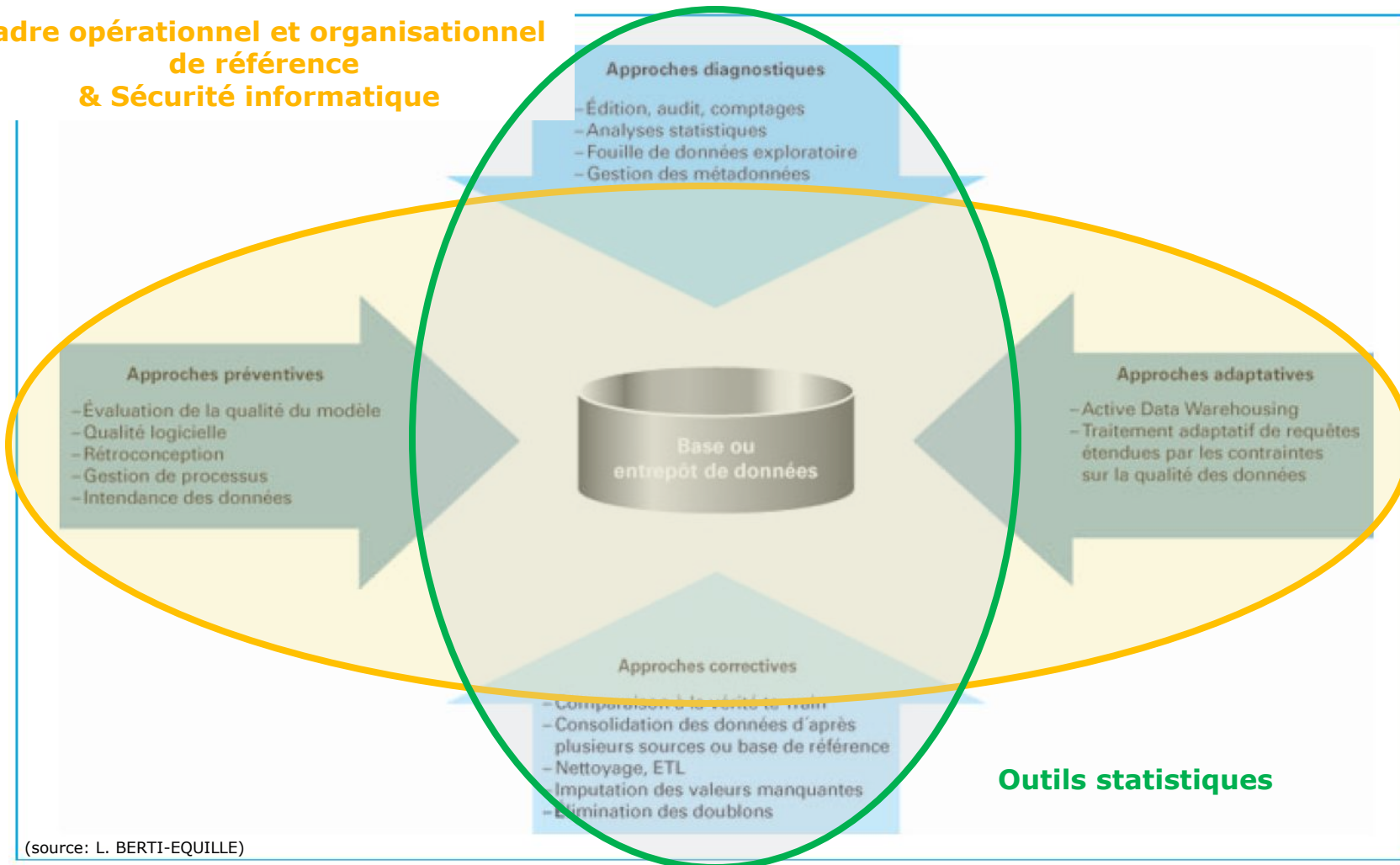
visait généralement 3 principaux objectifs

- **L'intégrité** : confiance que l'on peut avoir dans une information ou une ressource (données altérées durant la communication de manière fortuite ou intentionnelle)
- La **non répudiation** de l'information: garantir qu'une transaction ne peut être niée
- **L'authentification** : assurer l'identité d'un utilisateur

Mise en place de services de sécurité qui vont permettre de vérifier **l'intégrité** d'une donnée, d'**authentifier** ou non l'émetteur des données reçues, et garantir la **non-répudiation** des données émises et reçues.

# Approches pour l'évaluation et le contrôle de qualité des données

Cadre opérationnel et organisationnel  
de référence  
& Sécurité informatique



(source: L. BERTI-EQUILLE)

Et demain...

favoriser les services rendus par l'agriculture

