



Les données environnementales : formats, usages et défis numériques

Eric Masson

Géographe Lab. TVES EA 4477

Université de Lille Sciences et Technologies

Journée données environnementales | Action ADOC - GdR MaDICS CNRS
MESHS – Lille - Vendredi 02 décembre 2016

Introduction

La donnée environnementale numérique est :

- massivement collectée (big data) par un « écosystème » multi-capteurs,
- sur des objets ponctuels (stations), linéaires (réseaux) et surfaciques (surface du globe),

elle répond à une demande :

- d'application multi-thématique et multi-scalaire,
- de suivi d'états dynamiques de l'environnement,

et présente de nombreux défis :

- scientifiques (interactions technologies-connaissances),
- d'interactions environnementales spatio-temporelles (antécédence, moment et prévision),
- Patrimoniaux car les données d'hier ne sont pas celles d'aujourd'hui ni celles de demain.

Perspective : hydrologie - topographie - occupation du sol

Plan de la présentation

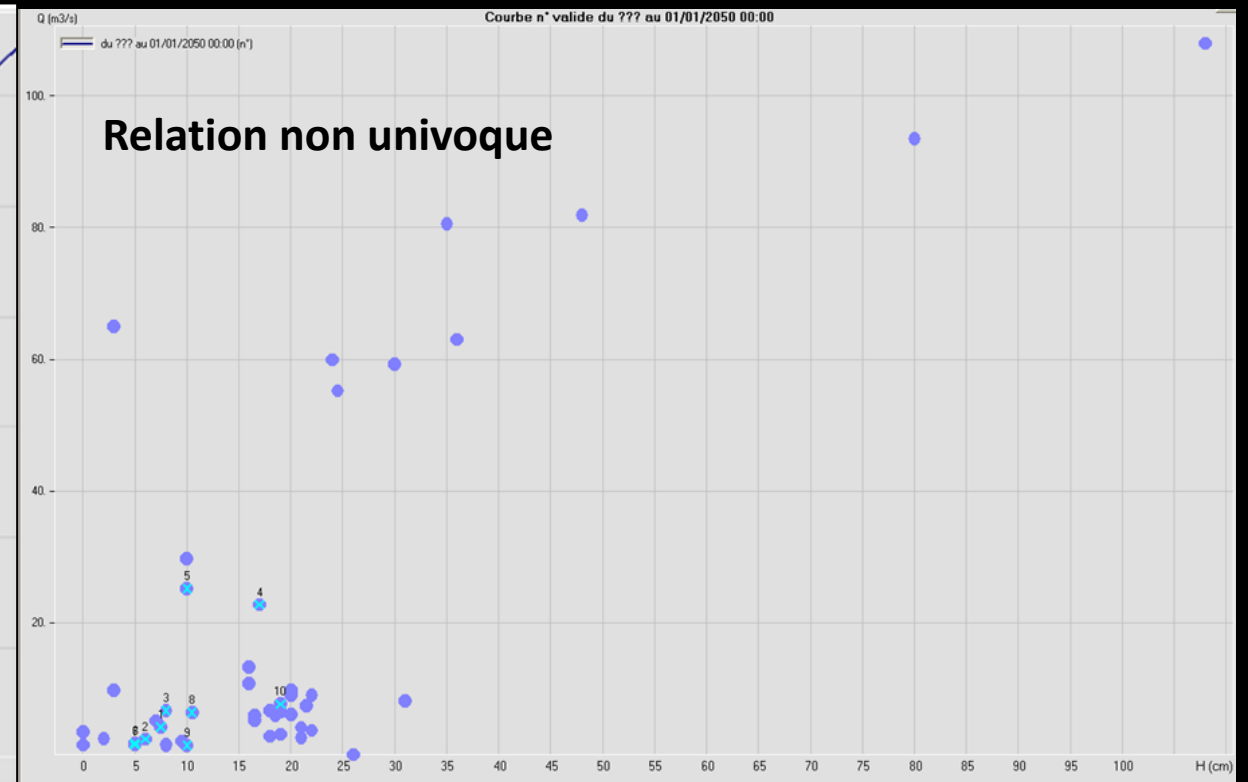
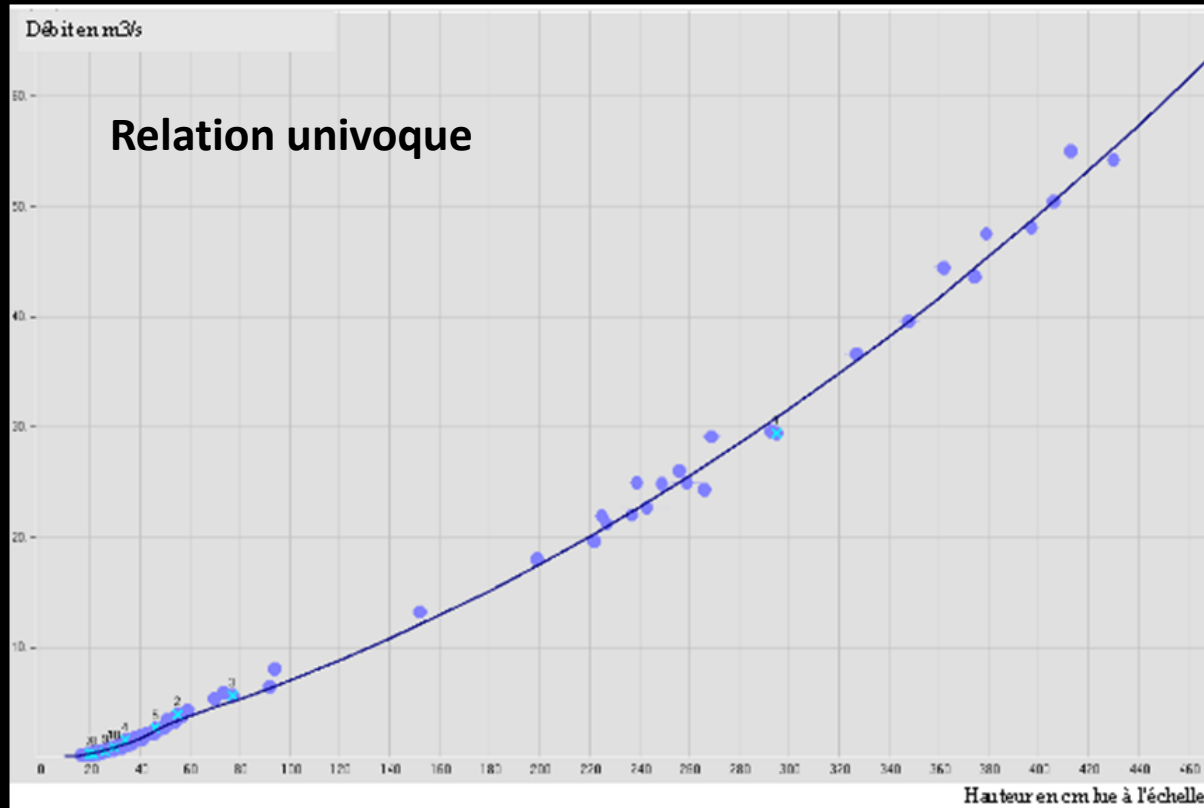
- Le big data de la donnée environnementale numérique : morceaux choisis
 - Types, formats et volumes de données (acquisition & diffusion)
 - L'exploitation des données (observatoires, SIG & télédétection)
- Défis numériques du big data environnemental
 - Echantillonnage (résolutions spatiales & temporelles) et stationnarité
 - Fusion et fouilles de données (complexité / approche orientée objet)
- Perspectives
 - Technologiques et thématiques (scientifiques ?)
 - Participatives (citoyenne ?)

Le big data de la donnée environnementale numérique : hydrologie 1/2

Types, formats et volumes de données (acquisition & diffusion) :

- Hauteurs de pluies ponctuelles (entrées)
- Jaugeage et courbe de tarage (étalonnage des écoulements)
- Débits (sorties)
- Format numérique *.CSV banque de données hydro (en France)
- Données à la minute, à l'heure, à la journée... (jusqu'à 525600 entrées par année et par station de mesure)
- Relevés manuels et visuels journaliers début 20^e Siècle
- Relevés automatiques à la minute et télétransmission fin 20^e Siècle
- Faible nombre de stations de mesure à longue série chronologique

Le big data de la donnée environnementale numérique : hydrologie 2/2



L'univocité Débit/Hauteur une relation indispensable en hydrologie

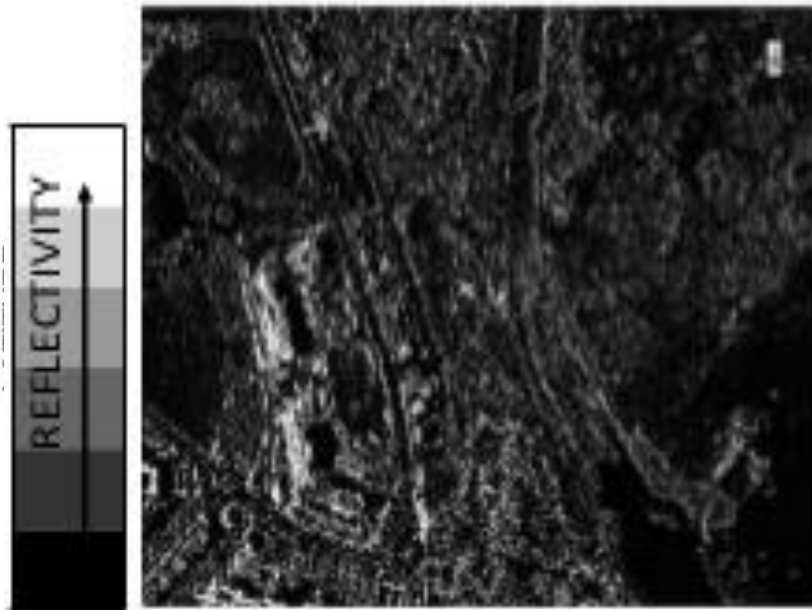
Le big data de la donnée environnementale numérique : topographie 1/2

Types, formats et volumes de données (acquisition & diffusion) :

- Cartes topographiques en hachures puis en isolignes (fin 19^e – 3/4 du 20^e Siècle)
- Acquisition topographique (géomètre) puis stéréoscopie optique (image aérienne)
- Images numériques par ordinateur (dernier 1/4 du 20^e Siècle)
- Acquisition aérienne ou satellitale (optique, radar, lidar)
- Données en sortie MNT, MNA, MNS (modèle numérique : de terrain, d'altitude, de surface)
- Format donnée image (geotiff, ecw...) ou ponctuelle (gpx, las...)
- Fréquence de mise à jour variable (plusieurs jours à plusieurs années)
- Couverture géographique variable (acquisition locale à globale)
- Base de donnée STRM (libre): 80% continents = 119.56 Mio km² > 1 To pour une résolution de 900m²
- Lidar (couvertures globale et nationale très limitées) : 1pt par m² = 1mio pt/km²

Le big data de la donnée environnementale numérique : topographie 2/2

SAR

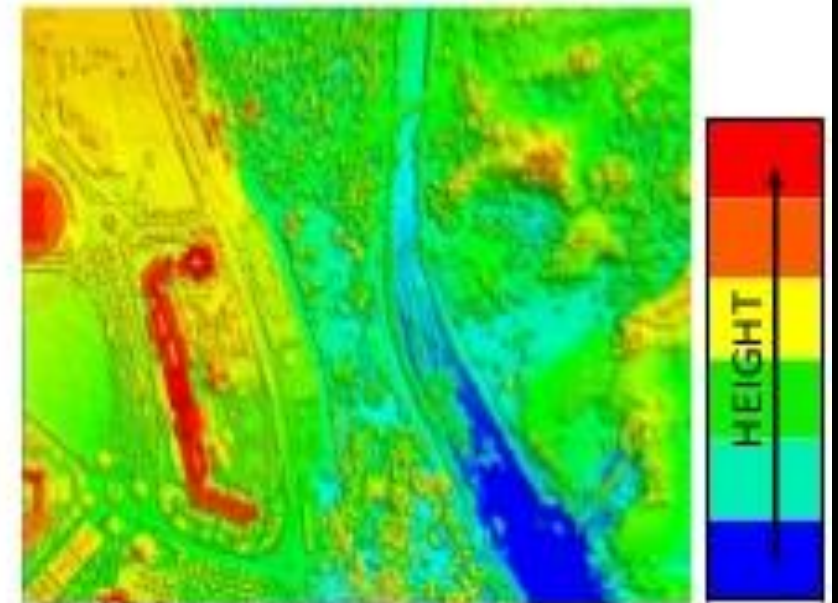


SAR image



Aerial photo (©Google)

Lidar



Lidar point cloud



TanDEM-X DEM of Death Valley with data from SRTM. ©DLR

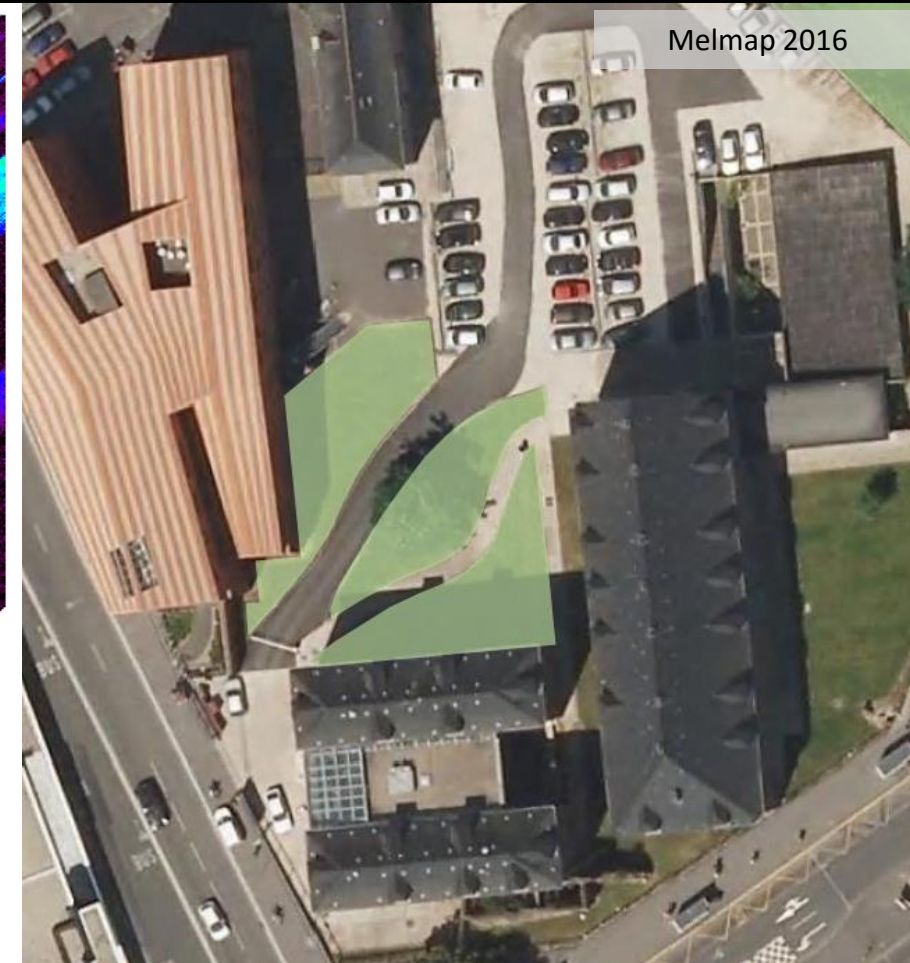
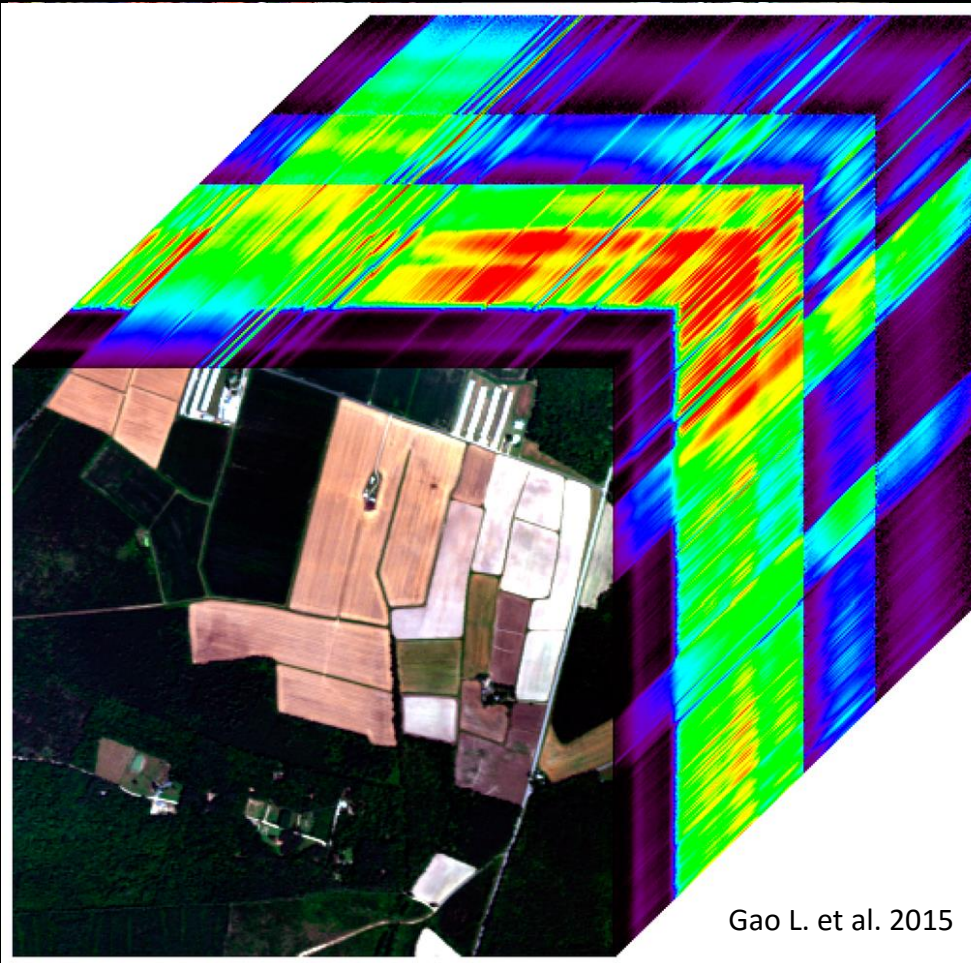


Le big data de la donnée environnementale numérique : occupation du sol (1/2)

Types, formats et volumes de données (acquisition & diffusion) :

- Cartes topographiques (fin 18^e – 3/4 du 20^e Siècle) par levé topographique puis photographie aérienne
- Données numérique d'occupation du sol (dernier 1/4 du 20^e Siècle) par imagerie satellitale
- Format raster (grille, image) ou format vectoriel
- Echelles d'utilisation variables : cadastrale (1/2000) à globale
- Fréquence de mise à jour variable (plusieurs mois à plusieurs années)
- Couverture géographique variable (acquisition locale à globale)
- Corine Land Cover (1990, 2000, 2006, 2012) 39 pays Européens pour la dernière version

Le big data de la donnée environnementale numérique : occupation du sol (2/2)



Worldview 3
 $680000\text{km}^2 / \text{jour}$
 $= 3\text{mio pixels} / \text{km}^2$
 $= 27 \times 10^6 \text{ données} / \text{km}^2$

Landsat 4-5
 $2,6 \text{ mio images } 1982\text{-}2012$
 $= 6,1 \text{ mio données} / \text{image}$
 $= 15,783 \times 10^{12} \text{ données}$

Aviris (hyperspectral)
résolution spatiale 5m &
193 bandes spectrales

Orthophoto MEL 2011
Résolution spatiale 8cm
 $3/4 \text{ bandes spectrales}$

Le big data de la donnée environnementale numérique : L'exploitation des données (1/2)

Les observatoires de l'environnement

- Aménagement du territoire
- Risques
- Biodiversité
- Pollution...

amènent une multiplication des :

- Sources et capteurs (réseaux)
- Données et thématiques (acteurs de l'environnement)
- Besoins de suivi temporel

Le big data de la donnée environnementale numérique : L'exploitation des données (2/2)

- La problématique des outils
 - Pérennité de la chaine d'acquisition-traitement (monitoring-modélisation)
 - Systèmes d'information géographique & Traitement numérique d'image
 - Fusion des formats vecteur (point, ligne, surface) et raster (résolutions, mesures, interpolations)
 - Couplage multi-sources : données météo-hydro-occupation du sol- topographie pour la prévention/prévision du risque hydrologique (par exemple)
 - Serveurs de données à références spatiales
 - Interopérabilité...

Défis numériques du big data environnemental : l'évolution du contexte spatial en hydrologie (1/2)

- La relation Hauteur/Débit dépend :
 - des conditions topographiques locales (du lit majeur / vallée)
 - de l'occupation du sol dans un voisinage (un environnement) proche
- Toute évolution temporelle de ces conditions entraîne :
 - Une modification de la section mouillée (géométrie / surface)
 - Une modification des vitesses d'écoulement (rugosité / concentration des flux)
- Conséquences pour les séries chronologiques de débits :
 - Problème de stationnarité
 - Nécessité de redresser les données
 - Abandon de site de mesure
- Upscaling/downscaling des modèles de prévision hydrologique ???
- Changement climatique ??? Tendence de fond ??? Connaissance des événements extrêmes ???

Défis numériques du big data environnemental : l'évolution du contexte spatial en hydrologie (2/2)

- Suivi de l'occupation du sol et de la géométrie (topographie) du lit majeur est donc indispensable
 - Défis numériques :
 - résolution spatiale (cadence d'échantillonnage)
 - résolution numérique (précision d'encodage de la mesure)
 - résolution temporelle (fréquence de revisite)
- L'actualisation des courbes de tarage et nécessaire
 - Défis technologiques :
 - jaugeage en situation extrêmes
 - intégration des données historiques
 - reconstitution des chroniques hydrométriques

Défis numériques du big data environnemental : fusion des sources et données

- Fusion et fouille de données (complexité / approche orienté objet)
 - Fusion de données multicapteurs (lidar / optique, radar / optique)
 - Fusion de données multi-temporelles (pas de temps d'acquisition variables)
 - Interopérabilité des données (formats numériques...)
 - Intégration des chaines de production depuis l'acquisition jusqu'à la diffusion sur webserveur
 - Outils d'extraction orienté objet Vs pixel à pixel Vs mixtes...

Défi de la très haute résolution spatiale multi-temporelle à l'échelle régionale

= $124,14 \times 10^9$ pixels x 4 bandes spectrales



Données aériennes PPIGE région NPDC (site de Metaleurop)
Composition colorée fusion IR-2009 (20cm) / RVB-2005 (50cm)

Défi de l'extraction orientée objet = fouille spatialement contextualisée dans les données images

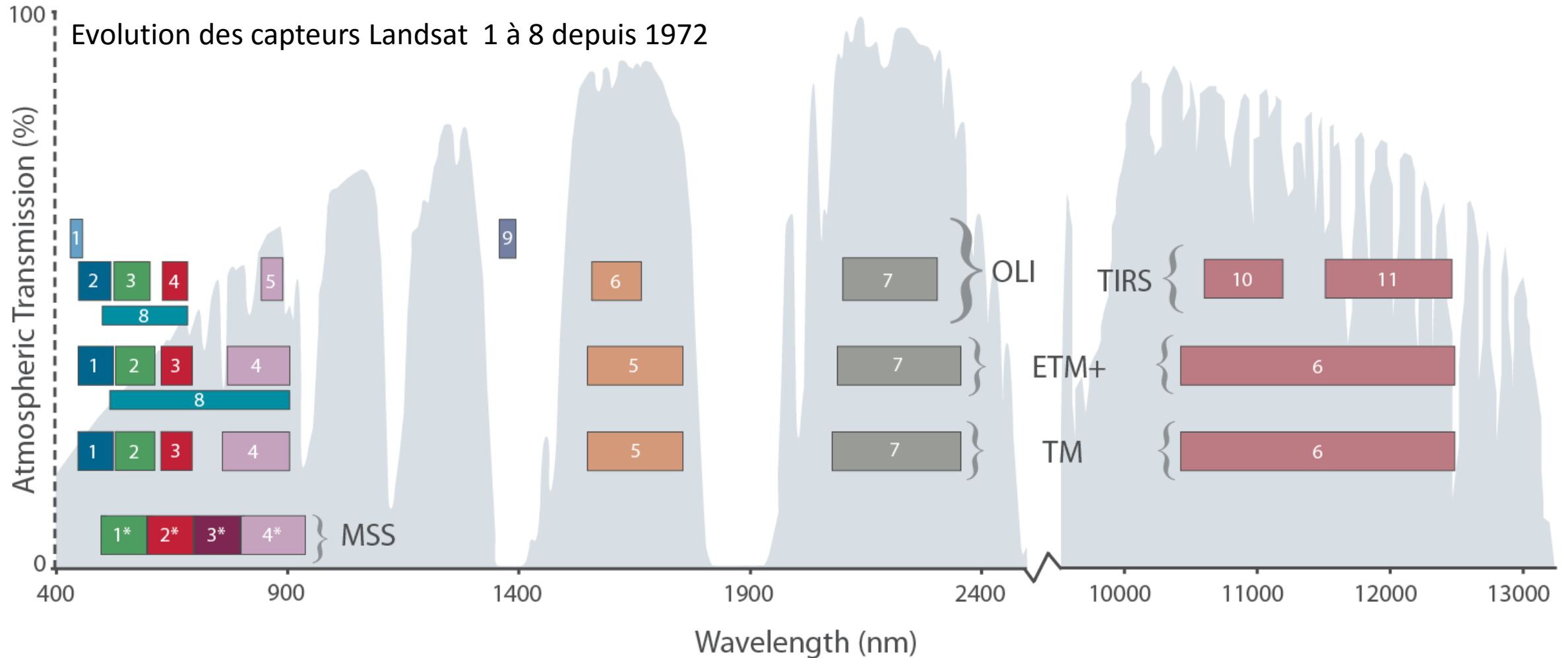


Image Object Information	
Feature	Value
Layer Values	
Mean	
Brightness	72.95
B2005	51.51
IR2009	123.27
Max. diff.	1.081
R2005	81.86
R2009	44.39
V2005	83.32
V2009	53.36
Layer Values	
Standard devi	
B2005	12.75
IR2009	19.75
R2005	19.33
R2009	14.56
V2005	18.11
V2009	15.46
Geometry	
Extent	
Area	20858 Pxl
Border length	1248 Pxl
Length	187.03 Pxl
Length/Width	1.110
Number of pixels	20858
Volume	20858 Pxl
Width	168.44 Pxl
Geometry	
Shape	
Asymmetry	0.021384
Border index	2.144
Compactness	1.510
Density	2.426
Elliptic Fit	0.8494
Main direction	139.62
Radius of largest enclosed ellipse	0.7048
Radius of smallest enclosing ellipse	1.194
Rectangular Fit	0.8866
Roundness	0.4889
Shape index	2.160
GLCM Homogeneity	
All directions	
GLCM Homogeneity (all dir.)	0.091627
IR2009	0.060720
GLCM Contrast	
All directions	
GLCM Contrast (all dir.)	134.94
GLCM Dissimilarity	
All directions	
GLCM Dissimilarity (all dir.)	9.164
GLCM Entropy	
All directions	
GLCM Entropy (all dir.)	7.643
GLDV Entropy	
All directions	
GLDV Entropy (all dir.)	3.183

9,969 Objects

Quel niveau de segmentation /extraction choisir ?

Défi de la dérive temporelle des technologies et des archives numériques environnementales



Landsat Science, 2016

Discontinuité temporelle/qualité des données environnementales : Landsat 7 ETM +

1999-2003

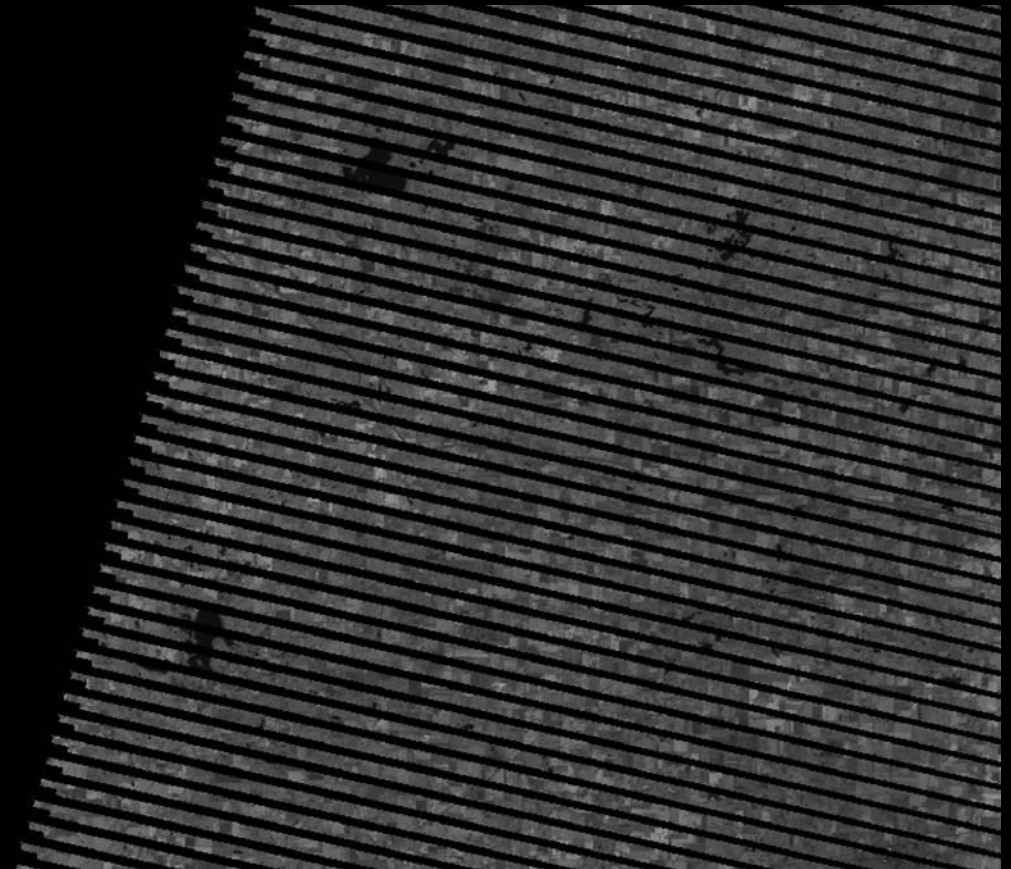
Satellite complètement opérationnel

2003-...

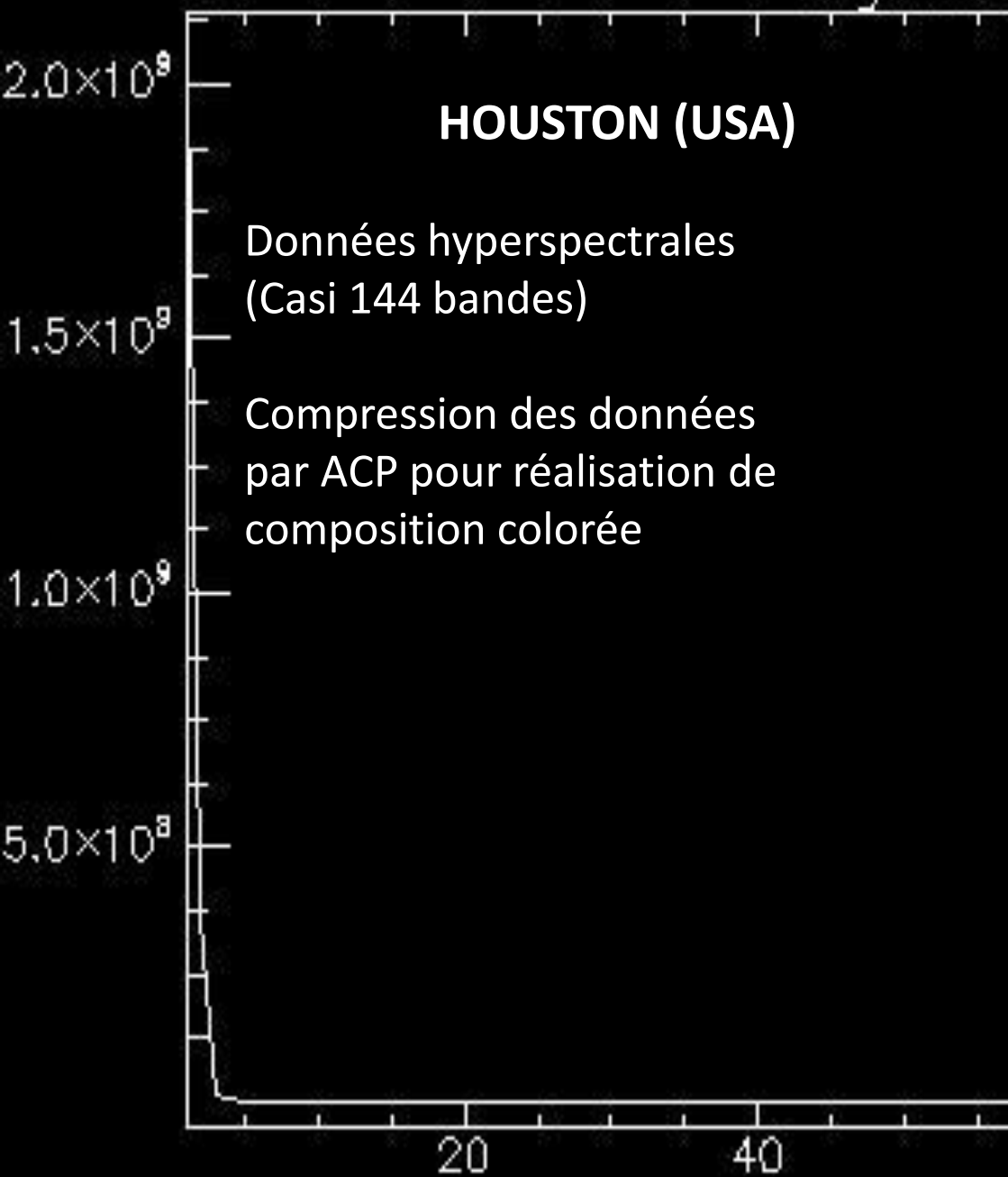
Satellite endommagé par débris en orbite

2016 ETM+ toujours en activité...

Qualité données Vs maintien observation



Gismap, 2016



ACP 1

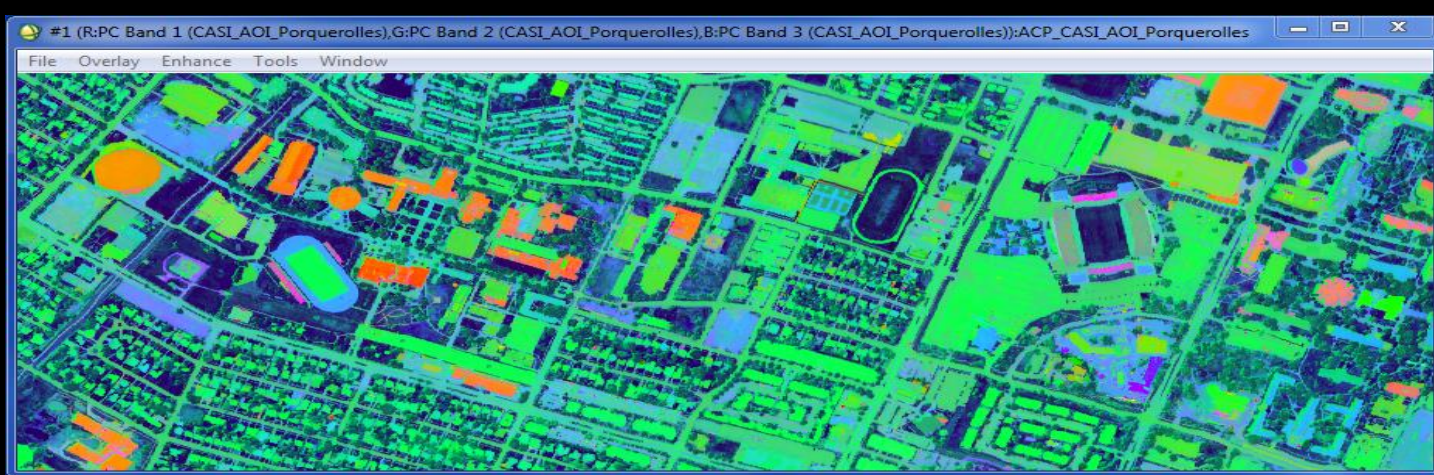


ACP 2

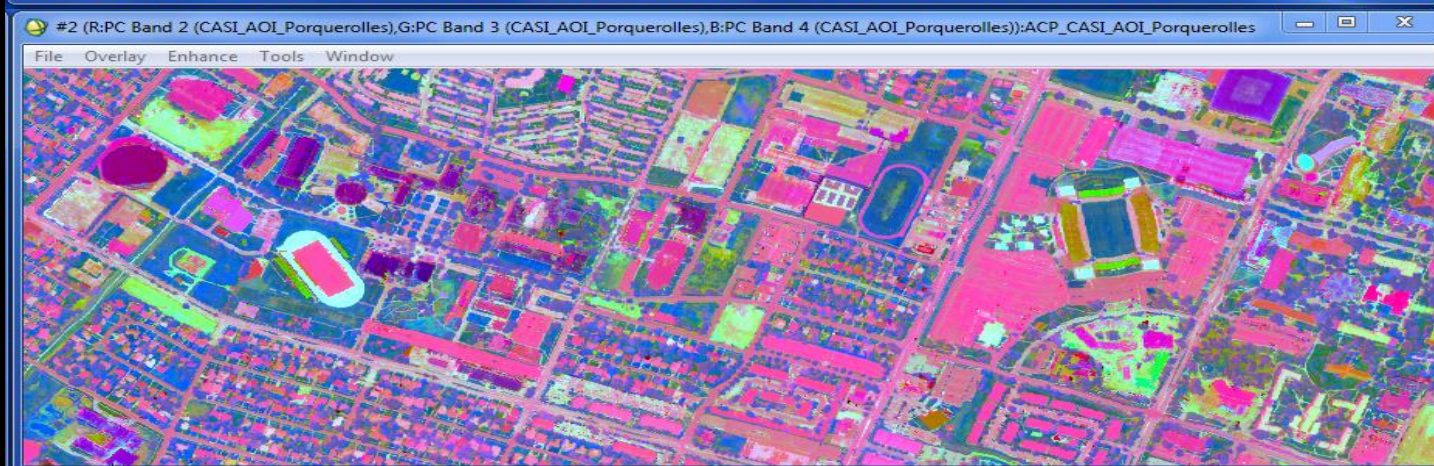


ACP 3





Visualisation complexes-expertes HOUSTON (USA)



Données hyperspectrales
(Casi 144 bandes)

Combinaisons de composantes principales
pour l'interprétation et l'analyse
thématique des données images



De haut en bas :
Composantes 123
Composantes 234
Composantes 456

Perspectives technologiques et thématiques (sciences ?)

- Quelques enjeux technologiques du big data environnemental :
 - Accéder, Explorer Extraire Analyser (BAU)
 - Puissance de calcul et de stockage en réseau
 - Discontinuités spatiales et temporelles : « boucher les trous »
 - Antécédences-moments-prévision : la complexité environnementale
- Quelques enjeux thématiques du big data environnemental :
 - Gestion des risques naturels
 - Gestion des ressources naturelles
 - Gestion de l'occupation du sol

Perspectives participatives (Citoyennes ?)

- Population connectée ...
 - Géolocalisation des données individuelles
 - Géolocalisation des parcours
 - Intégration des données multi-réseaux sociaux
 - La plateforme « humaine » vecteur de capteurs environnementaux
- ... Population engagée ?
 - Participation citoyenne (citizens as sensors, M.F. Goodchild 2007) ?
 - Géo activisme environnemental par/pour un big data militant ?
 - Big data collectif Vs Big data individuel ?

Les données d'hier ne sont pas celles
d'aujourd'hui ni celles de demain...

Merci pour votre attention