




PRÉDICTION QUANTITATIVE ET SPATIALE DE LA NATURE DES FONDS MARINS

ALOHA 12/12/2018 A. GAILLOT, S. BERMELL, IFREMER/GM/CTDI


Prédiction quantitative et spatiale de la nature des fonds marins

Personnes impliquées dans l'étude



Bermell Sylvain

02 98 22 42 45 (4245)
Sylvain.Bermell@ifremer.fr



Simplet Laure

02 98 22 46 25 (4625)
Laure.Simplet@ifremer.fr



Johan Saout
 Technicien Océanographe
 Monaco | Génie civil

Poste actuel Hydroconsult SARL
 Postes précédents IFREMER, DCNS ENERGIES,
 CREOCEAN
 Formation CNAM-INTECMER



Marie ECHAZAR
 Perpignan, Languedoc-Roussillon, France
 | Études/recherche

Formation Université de Perpignan Via Domitia

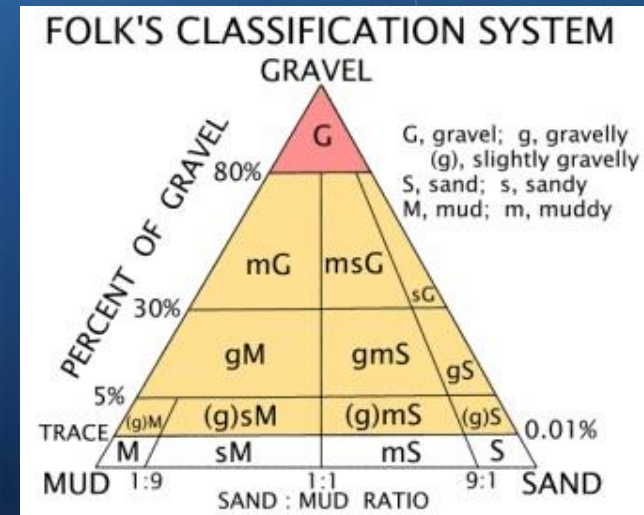
Prédiction quantitative et spatiale de la nature des fonds marins

Objectif :

Produire des modèles prédictifs, ici de classes granulométriques de type folk, en % de sable/vase/gravier

Besoins :

- **géosciences marines** : brique de base pour comprendre les systèmes sédimentaires
- **biologie marine** : base pour définir les typologies d'habitats
- **océanographie physique** : contraintes des modèles
- **porteurs de projets** liés au développement d'activités ou d'infrastructures en mer : information nécessaire à l'état de référence.

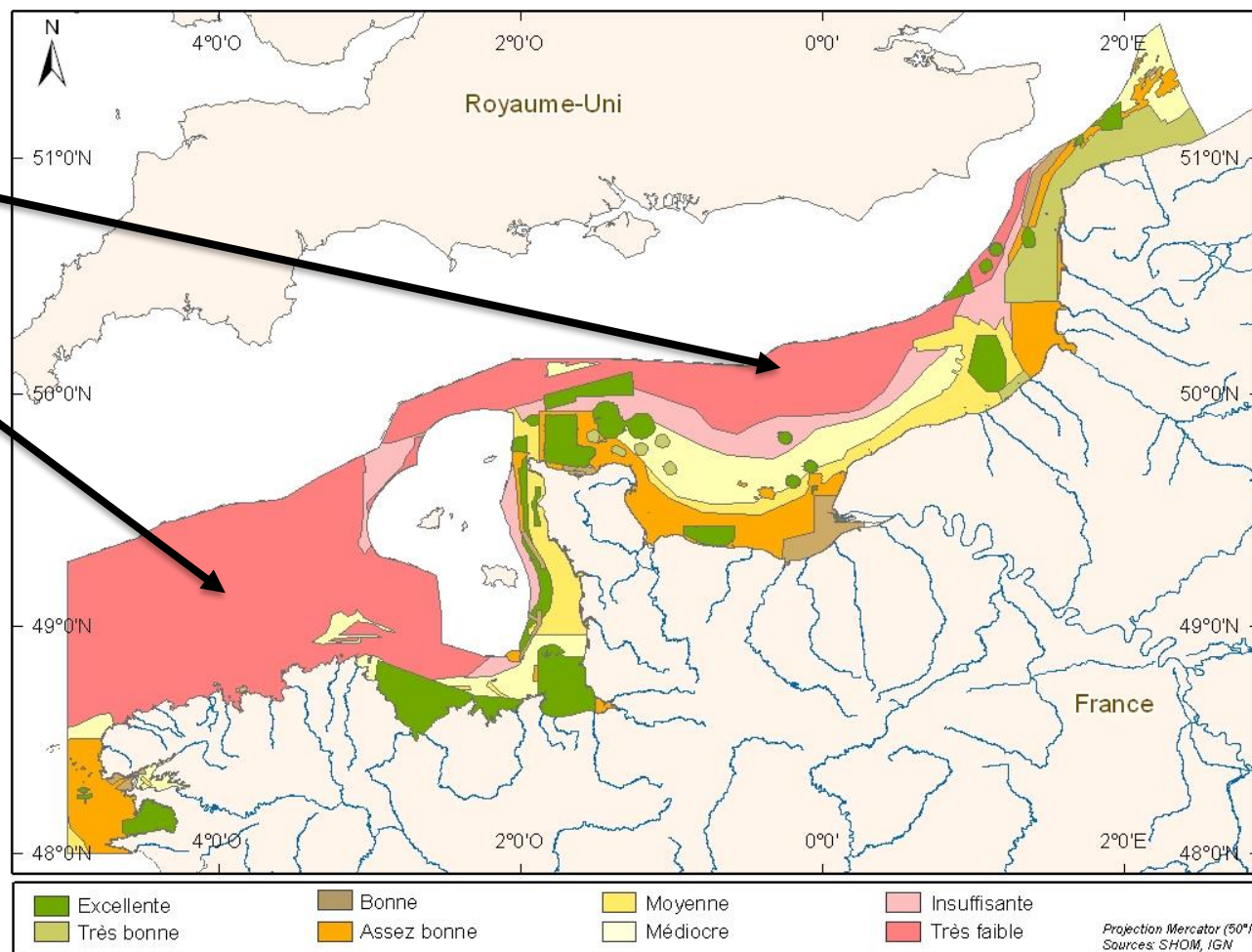


Etat de l'art sur le plateau continental de métropole

La directive-cadre stratégie pour le milieu marin (DCSMM) 2008

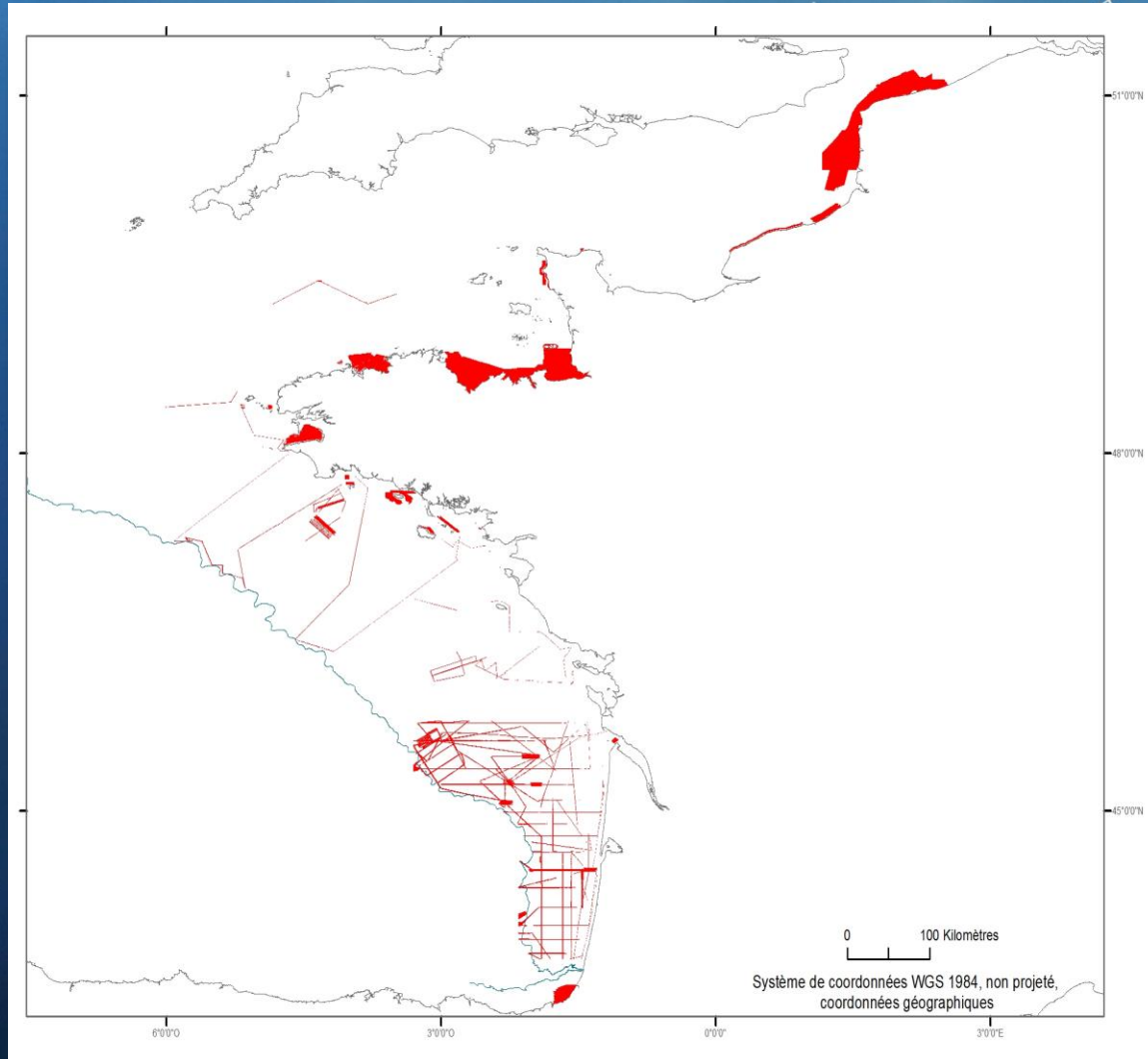
Nature des fonds marins - Etat de référence fourni par le SHOM

Connaissances
insuffisantes



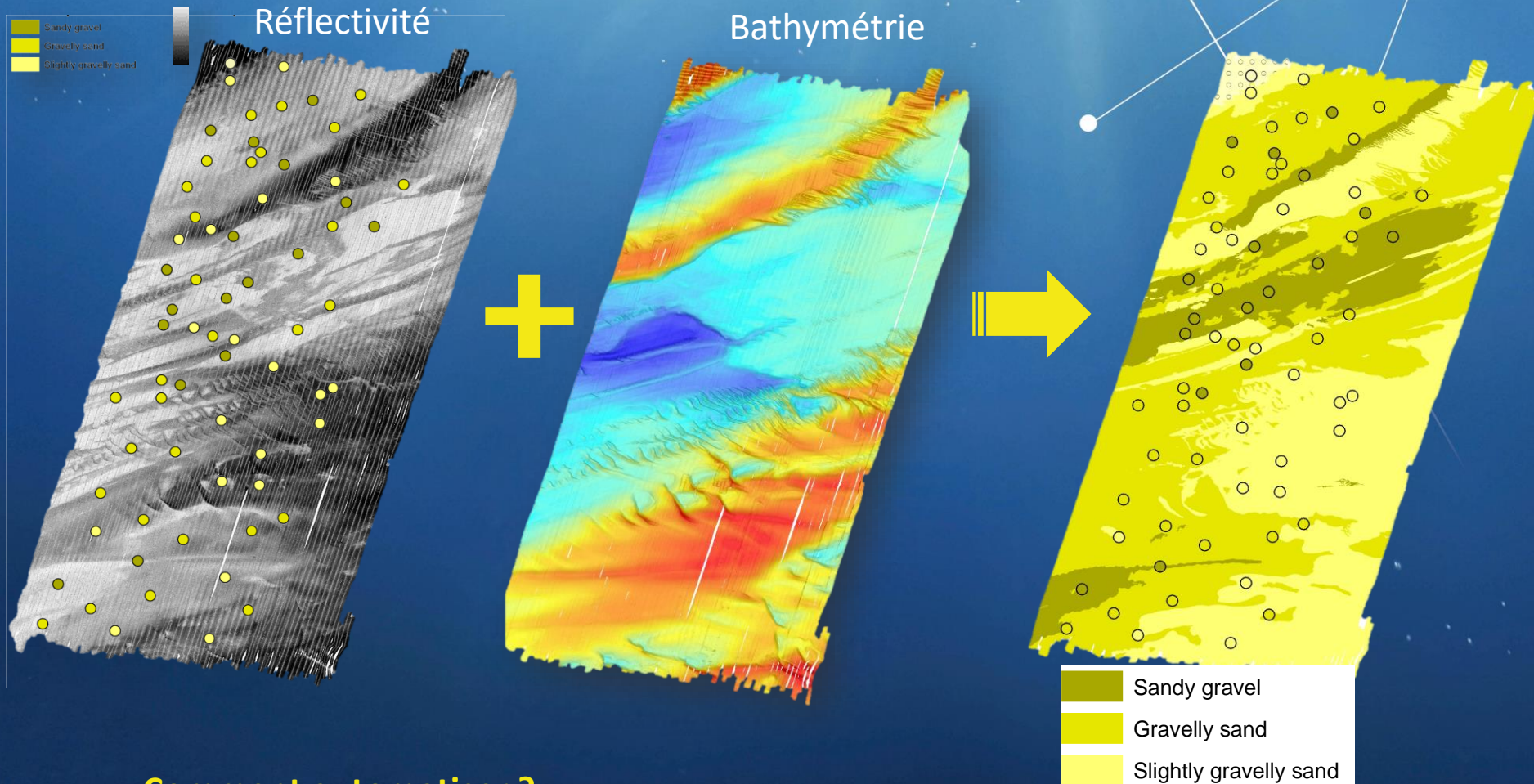
Etat de l'art sur le plateau continental de métropole

Cartes de nature des fonds issues des travaux
d'interprétation menés à ifremer/GM



Approche classique :

Classification supervisée humaine : Prélèvements + données acoustiques



Comment automatiser ?

Comment prendre en compte d'autres paramètres (courant, houle, etc.) ?

Comment garantir la reproductibilité ?

Méthode explorée :

Prédiction supervisée :

- Random forest

Données d'apprentissage :

- Echantillons sédimentaires
- Bathymétrie (+ dérivées)
- Réflectivité acoustique (+ dérivées)
- Modèle océanographie physique (courant, houle, etc.)



Centre for Environment
Fisheries & Aquaculture
Science



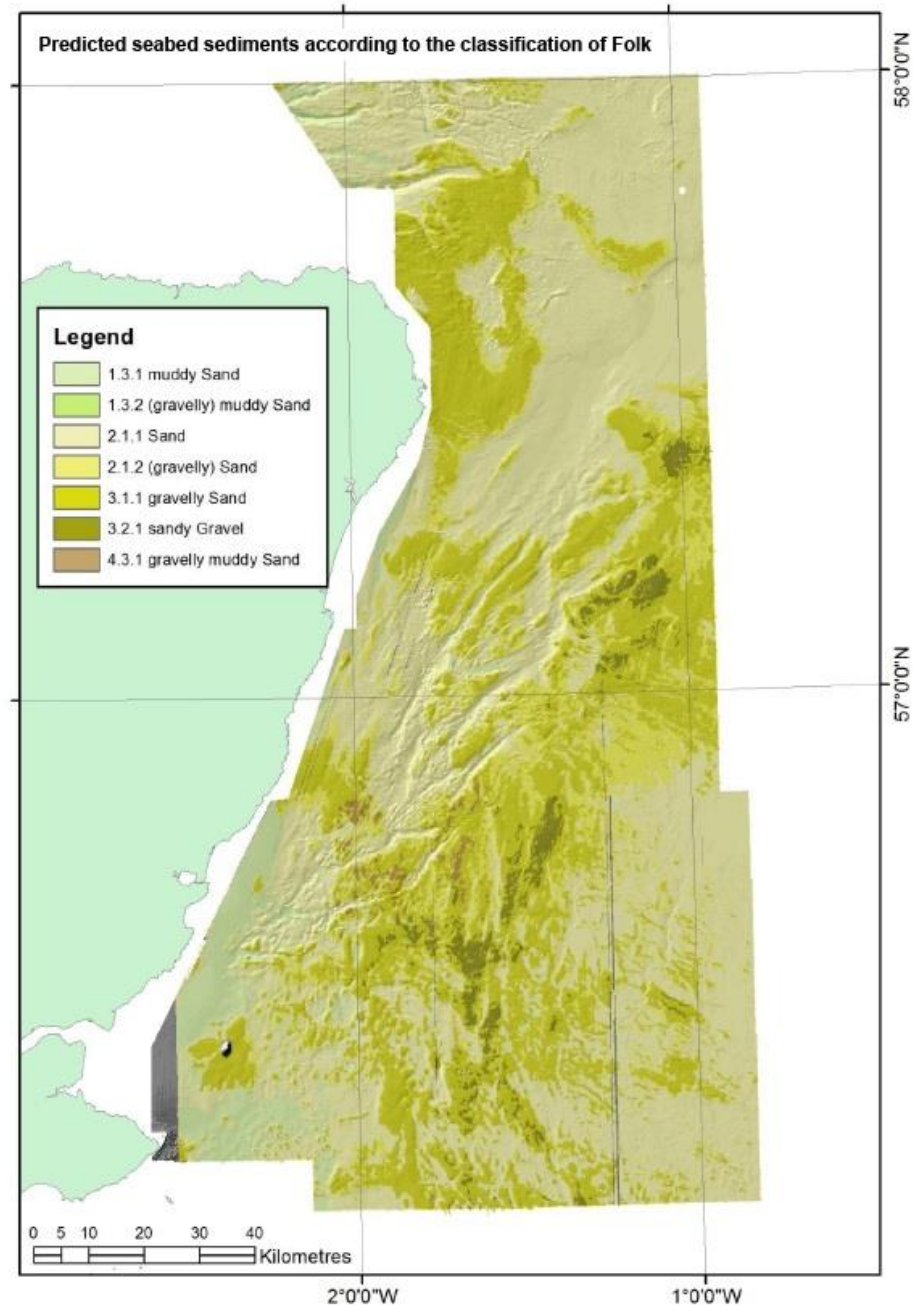
Cefas

C5818

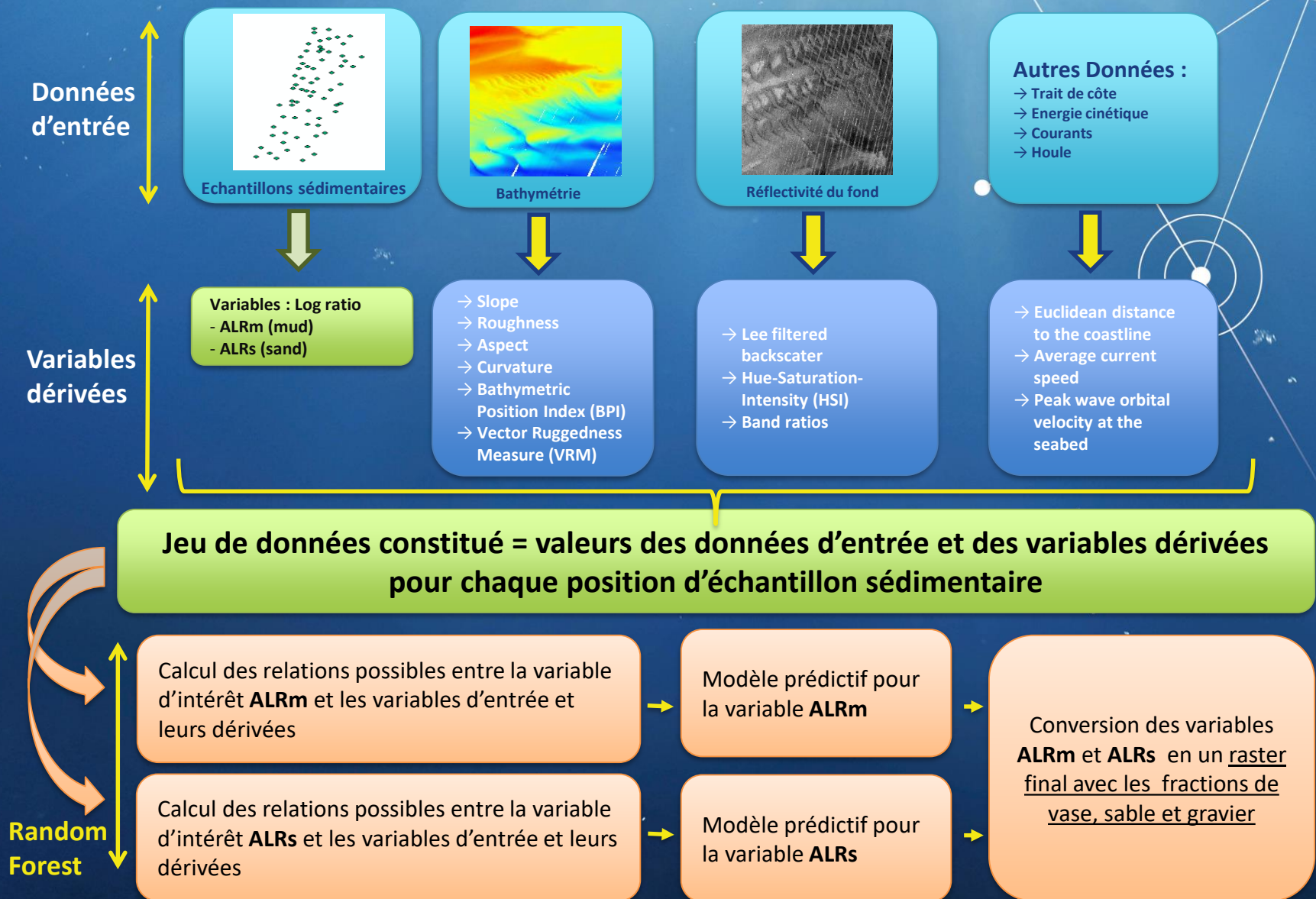
EMODnet-Geology Phase II

Case study: Quantitative spatial prediction of seabed sediment composition

Markus Diesing
7th October 2015



Grandes étapes de la méthode :



Random forest

Random Forest (Breiman 2001) : agrégation d'arbres de décision pour produire le modèle prédictif le plus précis possible.

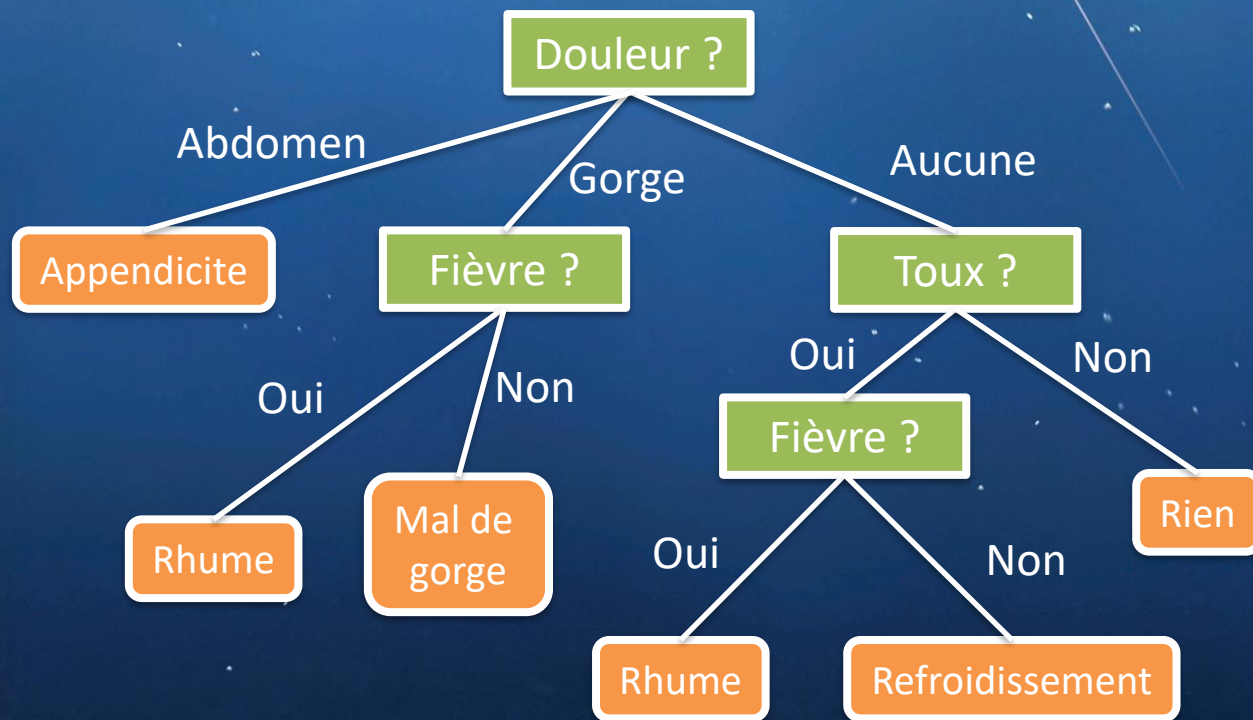
Arbre de décision, analogie médicale :

Les variables explicatives sont "Fièvre", "Douleur" et "Toux".

La variable à prédire est "Maladie"

Règles de classification :

Fievre	Douleur	Toux	Maladie
oui	Abdomen	non	Appendicite
non	Abdomen	oui	Appendicite
oui	gorge	non	rhume
oui	gorge	oui	rhume
non	gorge	oui	mal de gorge
oui	non	non	aucune
oui	non	oui	rhume
non	non	oui	refroidissement
non	non	non	aucune

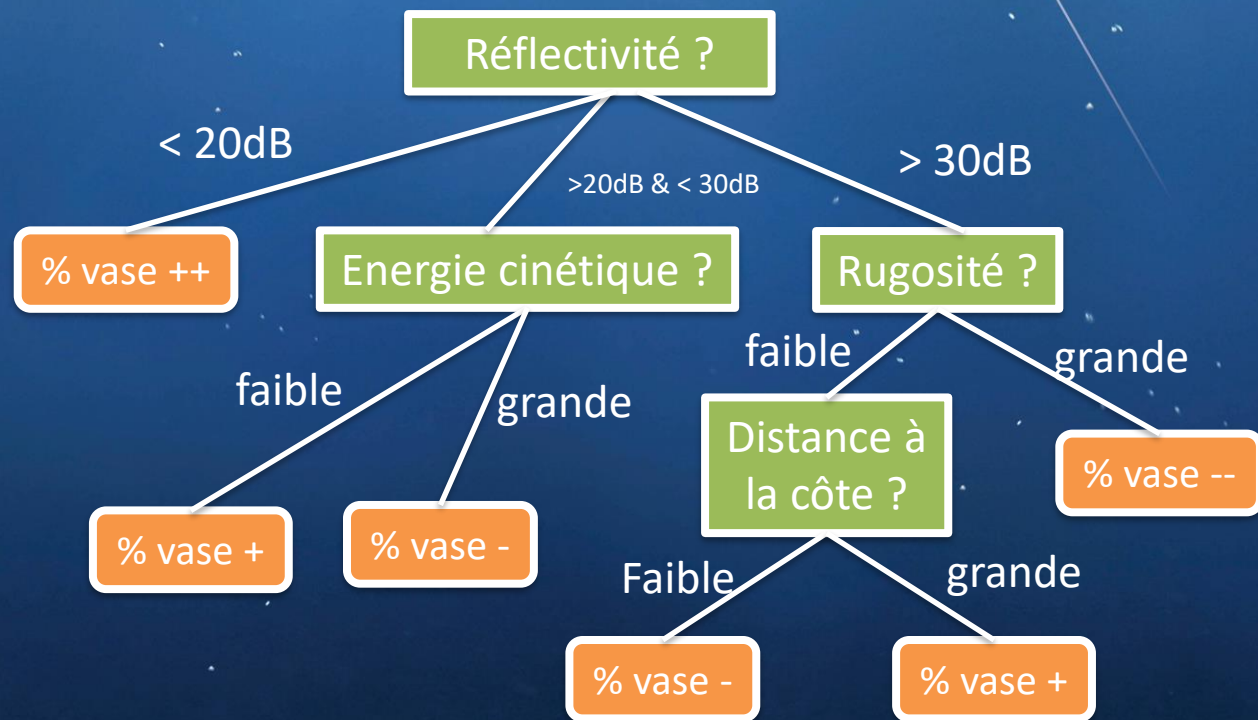


Random forest

Random Forest (Breiman 2001) : agrégation d'arbres de décision pour produire le modèle prédictif le plus précis possible.

Arbre de décision, notre cas d'étude :

Les variables explicatives sont « Réflectivité », « Bathylmétrie », etc...
La variable à prédire est « % vase »



Random forest

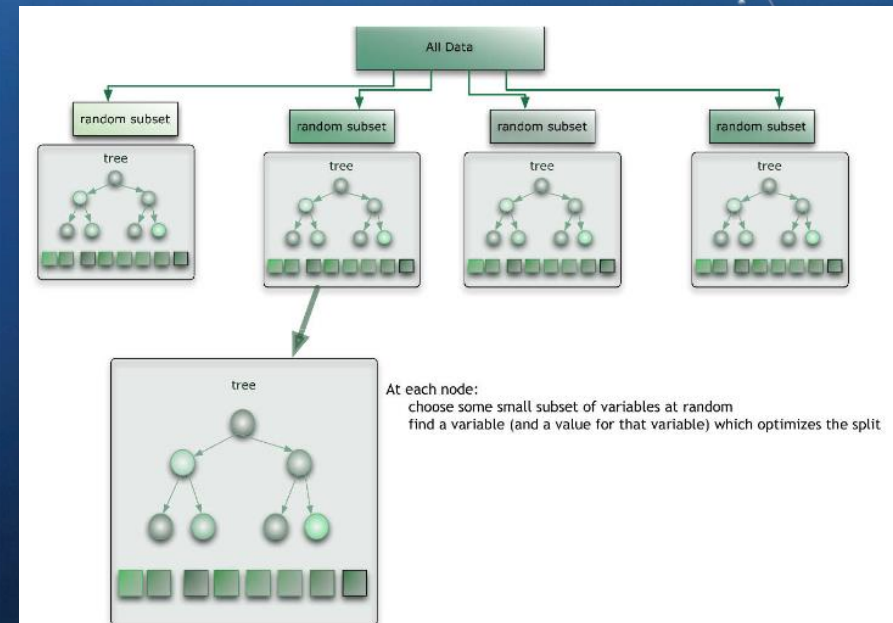
Random Forest (Breiman 2001) : agrégation d'arbres de décision pour produire le modèle prédictif le plus précis possible.

Principe = l'union fait la force

- apprendre grand nombre T (\sim qlq 10aines ou 100aines) d'arbres simples
- utilisation par vote des arbres (classe majoritaire, voire probabilités des classes par % des votes) si classification, ou **moyenne des arbres** si régression

But = obtenir des arbres les + **décorrelés possible**

- **chaque arbre appris sur un sous-ensemble** ($\sim 2/3$) aléatoire différent des exemples d'apprentissage
- **chaque nœud** de chaque arbre choisi comme « split » optimal parmi k **variables tirées aléatoirement** dans les entrées (avec $k \ll d$ la dim des entrées)



Outils :

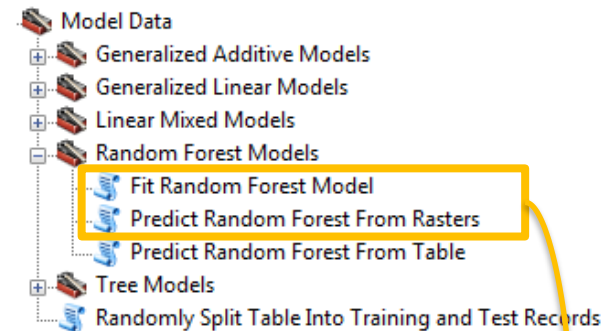
SIG



- Données géoréférencées
- Production des variables dérivées
- Prédiction basée sur des rasters



Toolbox MGET



<http://mgel.env.duke.edu/mget>

R randomForest package (Liaw and Wiener, 2002) Breiman's classic algorithm



Equivalent python (non testé)



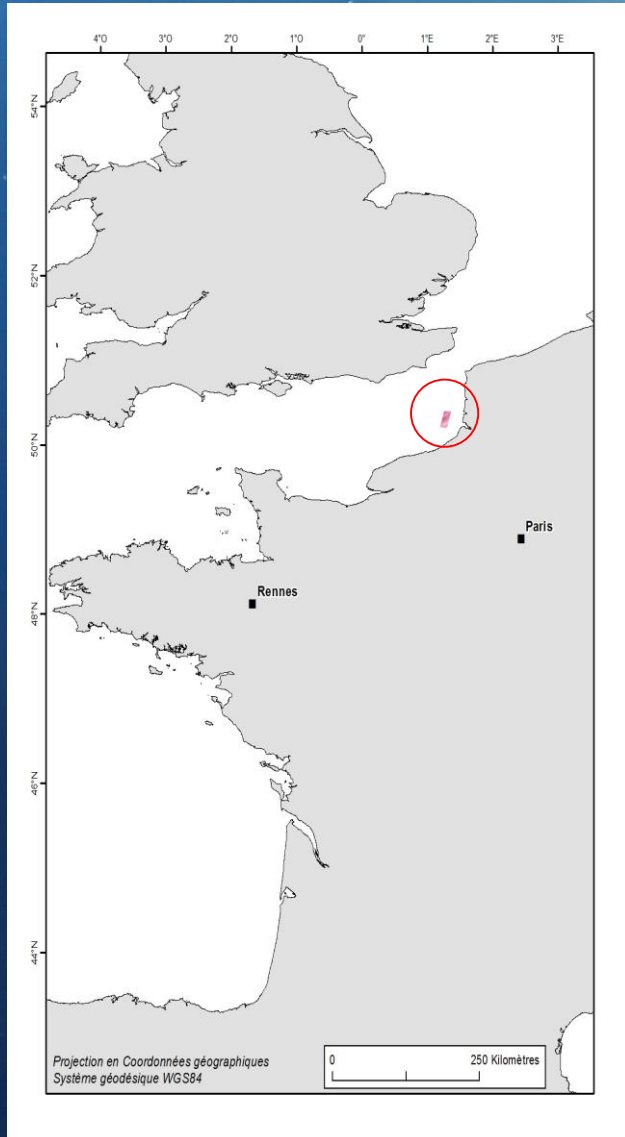
[sklearn.ensemble](#).RandomForestClassifier

Examples

```
>>> from sklearn.ensemble import RandomForestClassifier
>>> from sklearn.datasets import make_classification

>>> X, y = make_classification(n_samples=1000, n_features=4,
...                           n_informative=2, n_redundant=0,
...                           random_state=0, shuffle=False)
>>> clf = RandomForestClassifier(n_estimators=100, max_depth=2,
...                              random_state=0)
>>> clf.fit(X, y)
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
                       max_depth=2, max_features='auto', max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=100, n_jobs=None,
                       oob_score=False, random_state=0, verbose=0, warm_start=False)
>>> print(clf.feature_importances_)
[0.14205973 0.76664038 0.0282433  0.06305659]
>>> print(clf.predict([[0, 0, 0, 0]]))
[1]
```

Zone d'apprentissage :



RECOSOM 2 RECUP

Type Campagne océanographique

Navire [Thalia](#)

Propriétaire navire Ifremer

Dates 02/04/2013 - 14/04/2013

Chef(s) de mission [SIMPLET Laure](#)

GM-GEOSCIENCES MARINES
IFREMER Centre de Bretagne
ZI Pointe du diable
CS 10070
29280 PLOUZANE
☎ +33(0)2 98.22.42.24
✉ geosciences.marines@ifremer.fr
🌐 <http://www.ifremer.fr/gm>

DOI [10.17600/13070060](https://doi.org/10.17600/13070060)

Objectif Levé bathymétrique (SMF EM2040) d'une zone précédemment reconnue lors de la mission RECOSOM 2. Le projet de rattachement est "Systèmes et évolutions sédimentaires".



Citer cette campagne

SIMPLET Laure (2013)
RECOSOM 2 RECUP cruise, RV
Thalia, <https://doi.org/10.17600/13070060>

<https://doi.org/10.17600/13070060>

Variables dérivées de la bathymétrie :

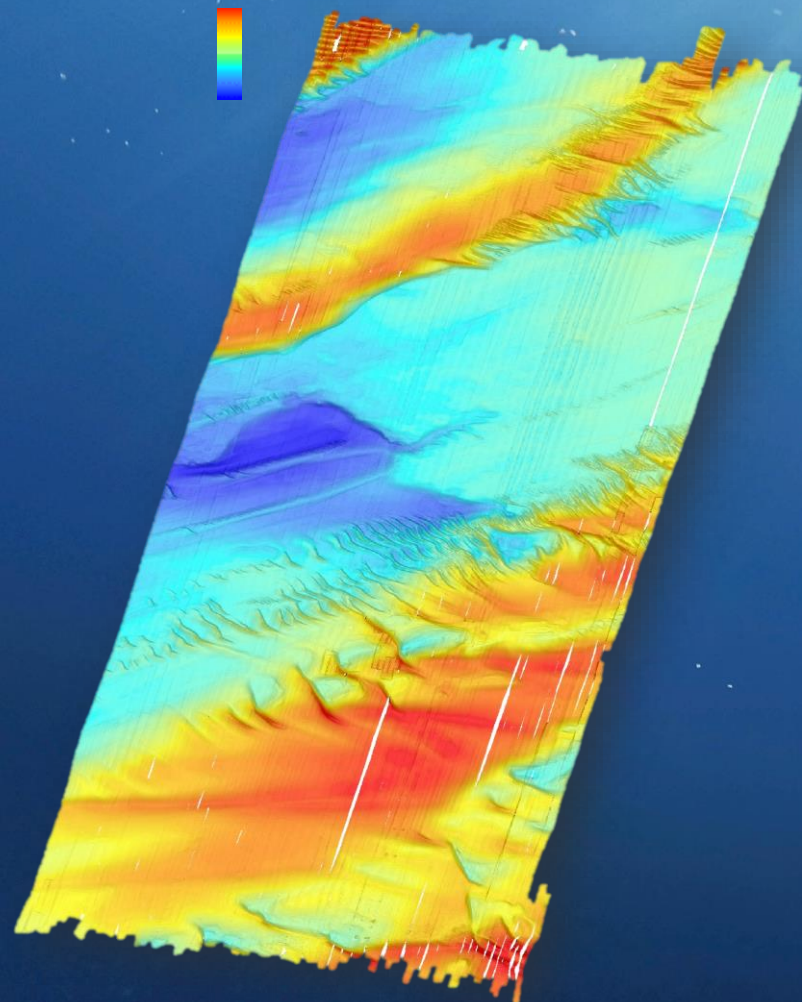


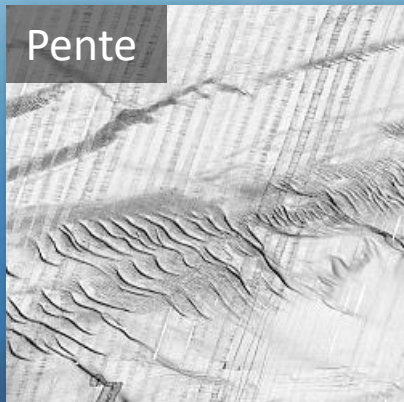
Table 5: Predictor variables derived from bathymetry data (MBFP).

Variable	Description	Unit	Name	Reference
Slope	The maximum slope gradient	degree	slope	Wilson et al. (2007)
Roughness	The difference between minimum and maximum of cell and its 8 neighbours.	m	rgh	Wilson et al. (2007)
Aspect	Direction of steepest slope, expressed as eastness (sine of aspect) and northness (cosine of aspect)		eastness northness	Wilson et al. (2007)
Curvature	Rate of change of slope. Profile curvature (curvPR) is measured parallel to maximum slope; plan curvature (curvPL) is measured perpendicular to slope.		curvPL curvPR curv	Wilson et al. (2007)
Bathymetric Position Index (BPI)	Vertical position of cell relative to neighbourhood (identifies topographic peaks and troughs). Radii of 3, 5, 10, 25, 30, 40, 50, 60, 70, 80, 90 and 100 pixels were used.	m	BPI3, BPI5, BPI10, BPI25, BPI30, BPI40, BPI50, BPI60, BPI70, BPI80, BPI90, BPI100	Lundblad et al. (2006)
Vector ruggedness measure (VRM)	Based on a geomorphological method for measuring vector dispersion.		VRM3, VRM5, VRM7, VRM9, VRM11	Sappington et al. (2005)

Variables dérivées de la bathymétrie :



Pente



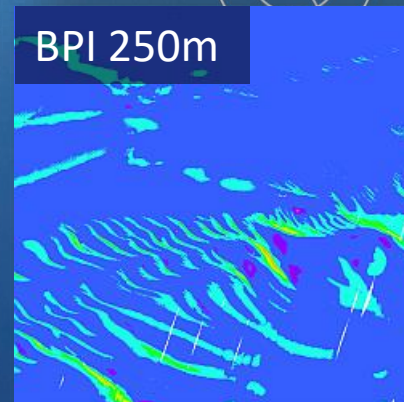
Rugosité



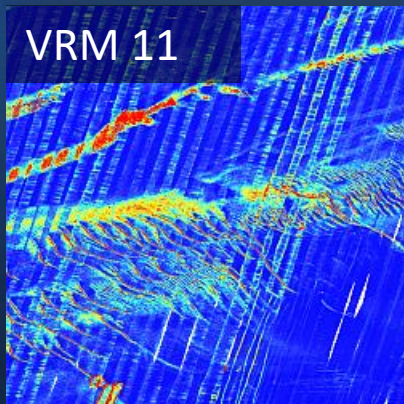
Courbure



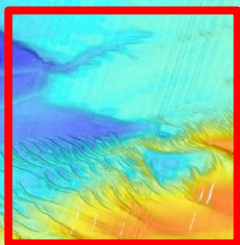
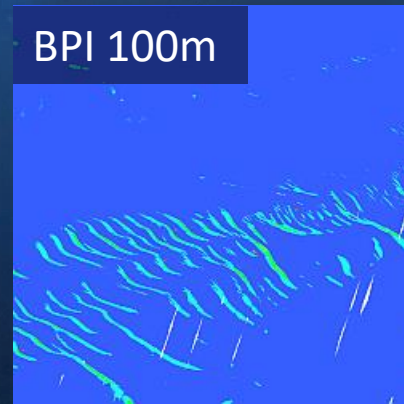
BPI 250m



VRM 11



BPI 100m



Variables dérivées de la réflectivité :

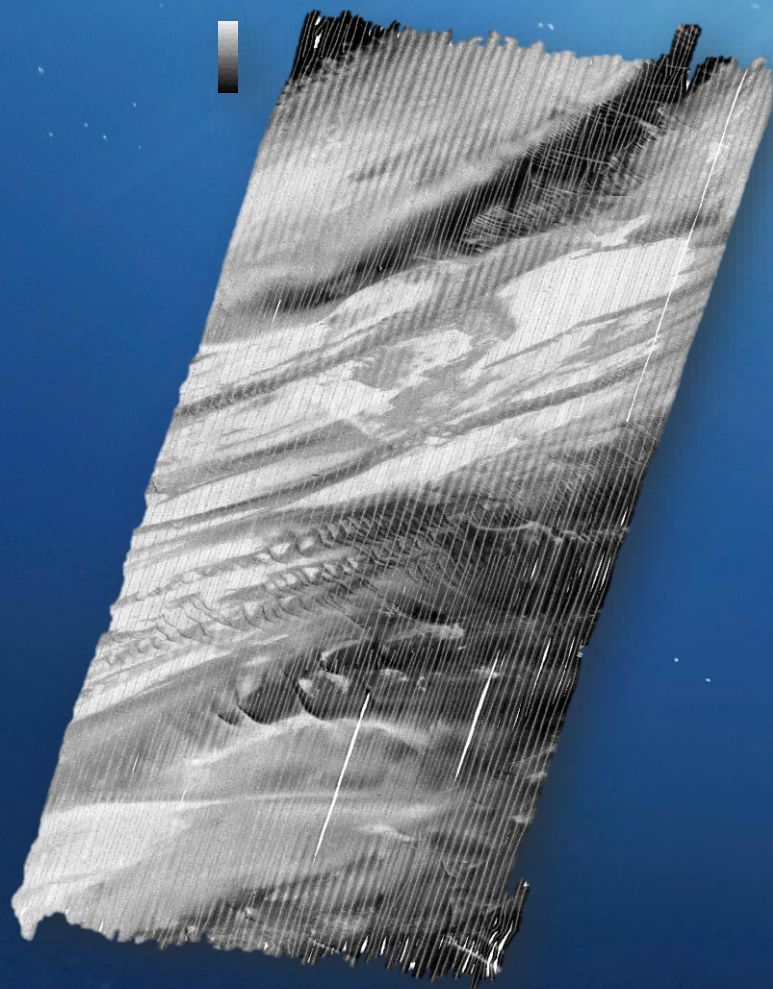
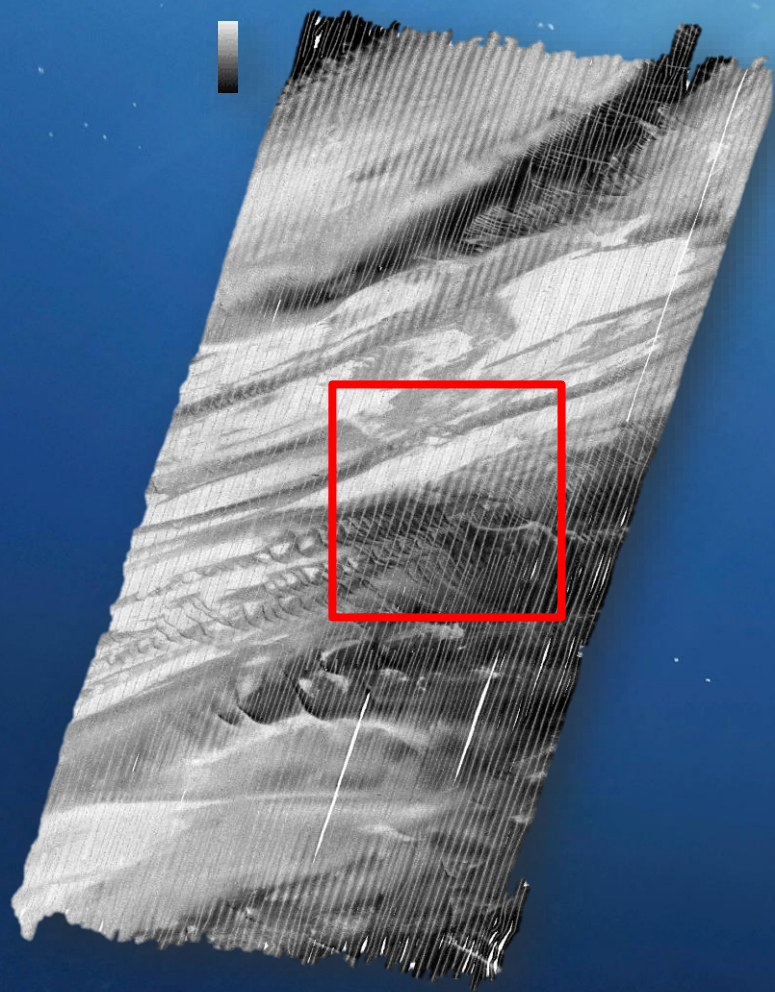


Table 6: Predictor variables derived from backscatter data (BSFP).

Variable	Description	Unit	Name	Reference
Lee-filtered backscatter	Reduces the speckle noise by applying a spatial filter to each pixel in an image, which filters the data based on local statistics calculated within a square window. The value of the centre pixel is replaced by a value calculated using the neighbouring pixels. A neighbourhood size of 5 pixels was chosen here after trials.	dB	BSFP_Lee5	
HSI	A synthetic colour image derived by applying high and low pass filters in order to separate high and low frequency information. High and low frequency information is then mapped to hue (chromatic) and intensity (achromatic) respectively with a fixed saturation value. These HSI values are then transformed to red, green, blue (RGB) colour space.		HSI_R HSI_G HSI_B	Daily (1983)
Band ratios	Band ratios of the three HSI layers were derived by simple division: Ratio_RG = HSI_R/HSI_G Ratio_RB = HSI_R/HSI_B Ratio_GB = HSI_G/HSI_B		Ratio_RG Ratio_RB Ratio_GB	

Variables dérivées de la réflectivité :



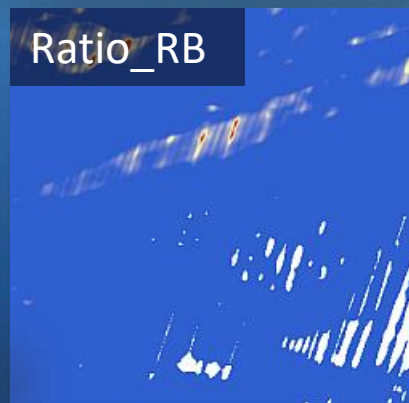
BSP_lee5



HSI_B



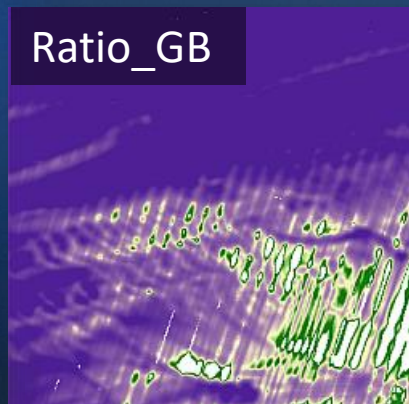
Ratio_RB



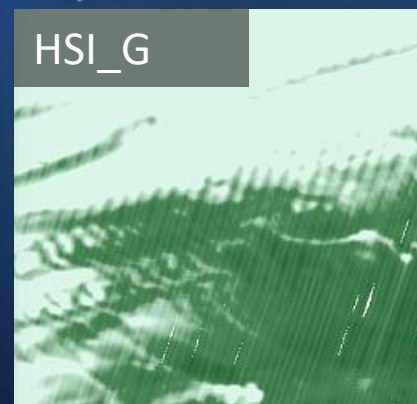
HSI_R



Ratio_GB

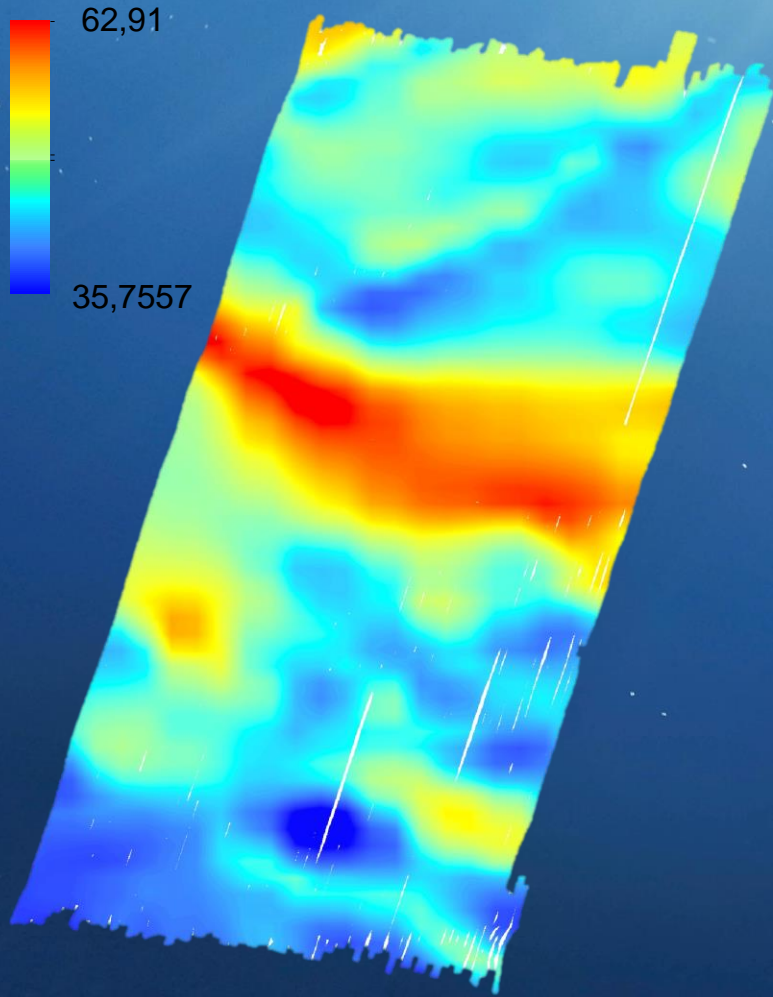


HSI_G



Autres données :

Kinetic mean bottom (kmb15)



Energie cinétique sur le fond N/m^2

Table 7: Other predictor variables.

Variable	Description	Unit	Name	Reference
Euclidean distance to coastline	The straight-line distance to the nearest coastline.	m	EucDist	
Average current speed	Depth mean tidal and wind-driven currents calculated using the POLCOMS model (Holt & James, 2001) with a grid resolution of approximately 11 km.	m/s	Av_current	Aldridge et al. (2015)
Peak wave orbital velocity at the seabed	The WAM spectral wave model was used to provide significant wave height and zero-crossing wave period with a grid resolution of approximately 11 km over the period 2000 to 2008. These were used in conjunction with a bathymetric model at a resolution of 200 m to derive statistics of wave orbital velocity at the seabed. These included the maximum (Mx), mean (Mn) and standard deviation (SD).	m/s	PkOrbVelMx PkOrbVelMn PkOrbVelSD	Aldridge et al. (2015)

Données de faible résolution

Résultats

Réfectivité

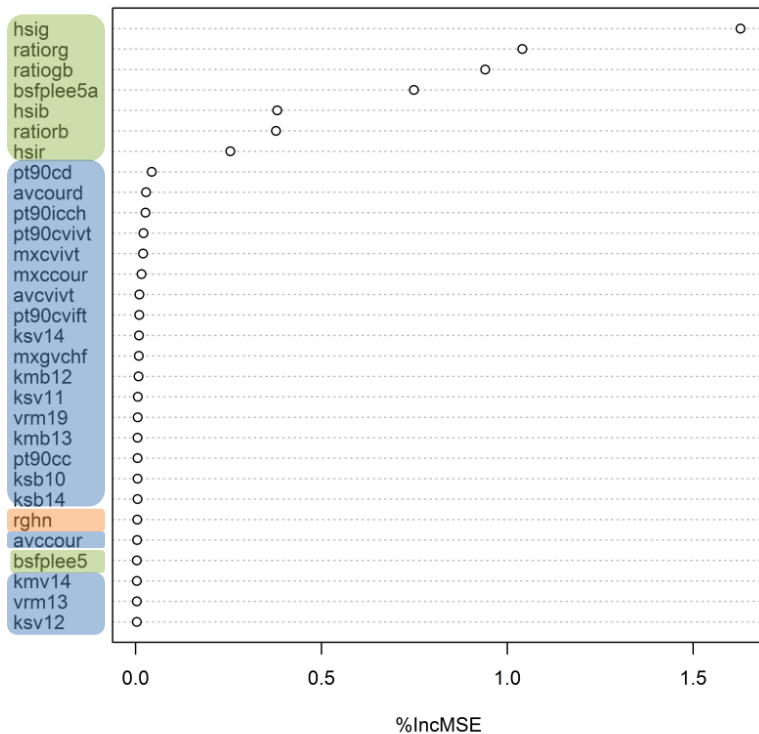
Bathy

Océano

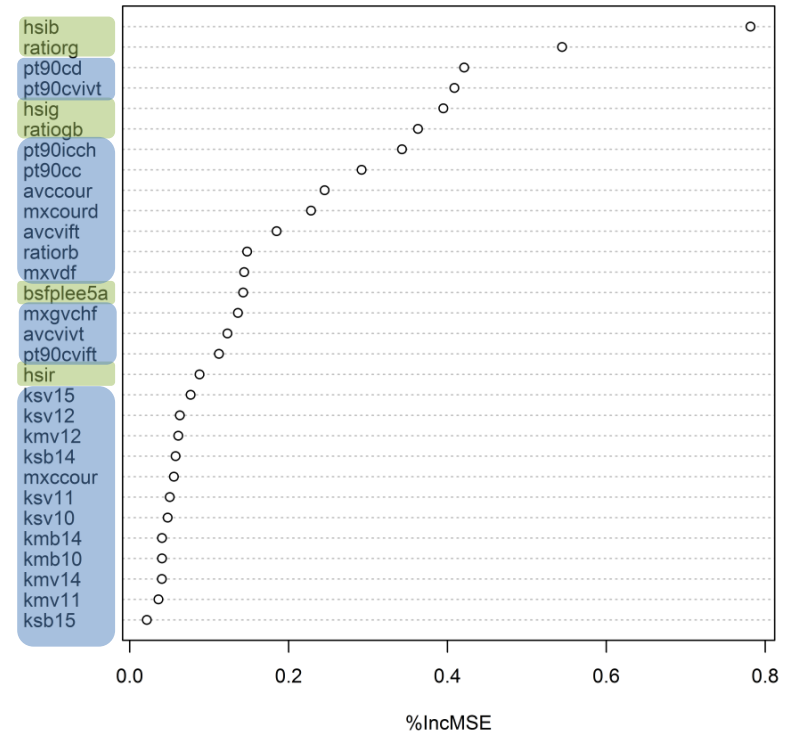
SAND (ALRs)

MUD (ALRm)

Variable Importance



Variable Importance



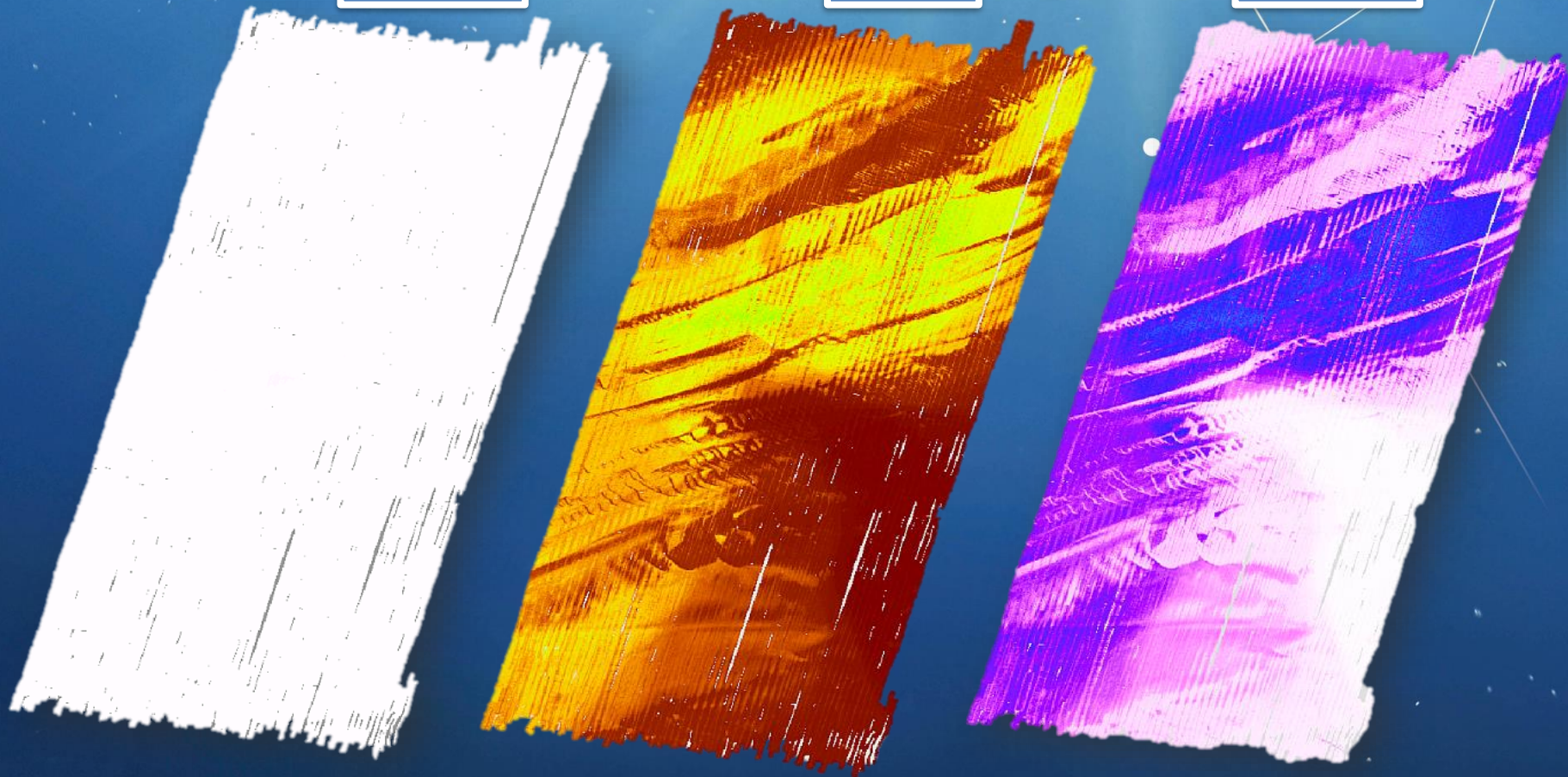
Nbre d'arbres : 500

Résultats :

Vase

Sable

Gravier



0%

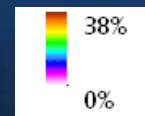
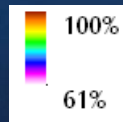
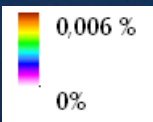
100%

Résultats :

Vase

Sable

Gravier



Résultats :

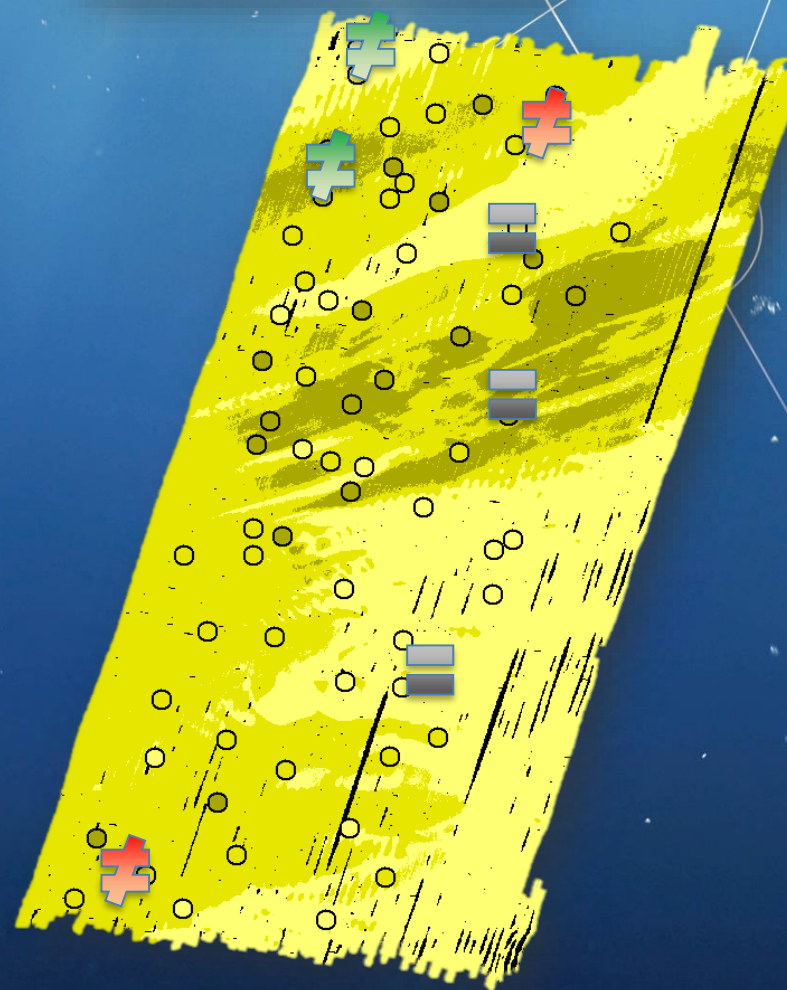
Traduction en classification Folk de la carte
FaciesSedimentaires_BaiedeSomme_Gautier_2017

Sandy gravel
Gravelly sand
Slightly gravelly sand
Sand with trace gravel

Référence

Carte de prédiction avec réflectivité (classification Folk)

Sandy gravel
Gravelly sand
Slightly gravelly sand



Résultats : sans réflectivité

Traduction en classification Folk de la carte
FaciesSedimentaires_BaiedeSomme_Gautier_2017

Sandy gravel
Gravelly sand
Slightly gravelly sand
Sand with trace gravel

Référence



Carte de prédiction sans réflectivité (classification Folk)

Sandy gravel
Gravelly sand
Slightly gravelly sand



Perspectives

1. Stabiliser l'algorithmie statistique du modèle de Diesing
2. Proposer des démarches interprétatives répétables et transférables, d'un jeu de données à un autre, d'un utilisateur à un autre
3. Rejouer les données historiques d'échantillons granulométriques au regard des nouvelles variables « modernes »
4. Produire des données utiles à d'autres thématiques (études hydrodynamiques)
5. Utiliser les SIG à des fins de modélisation