

# Axe MSA

## Bilan scientifique et perspectives

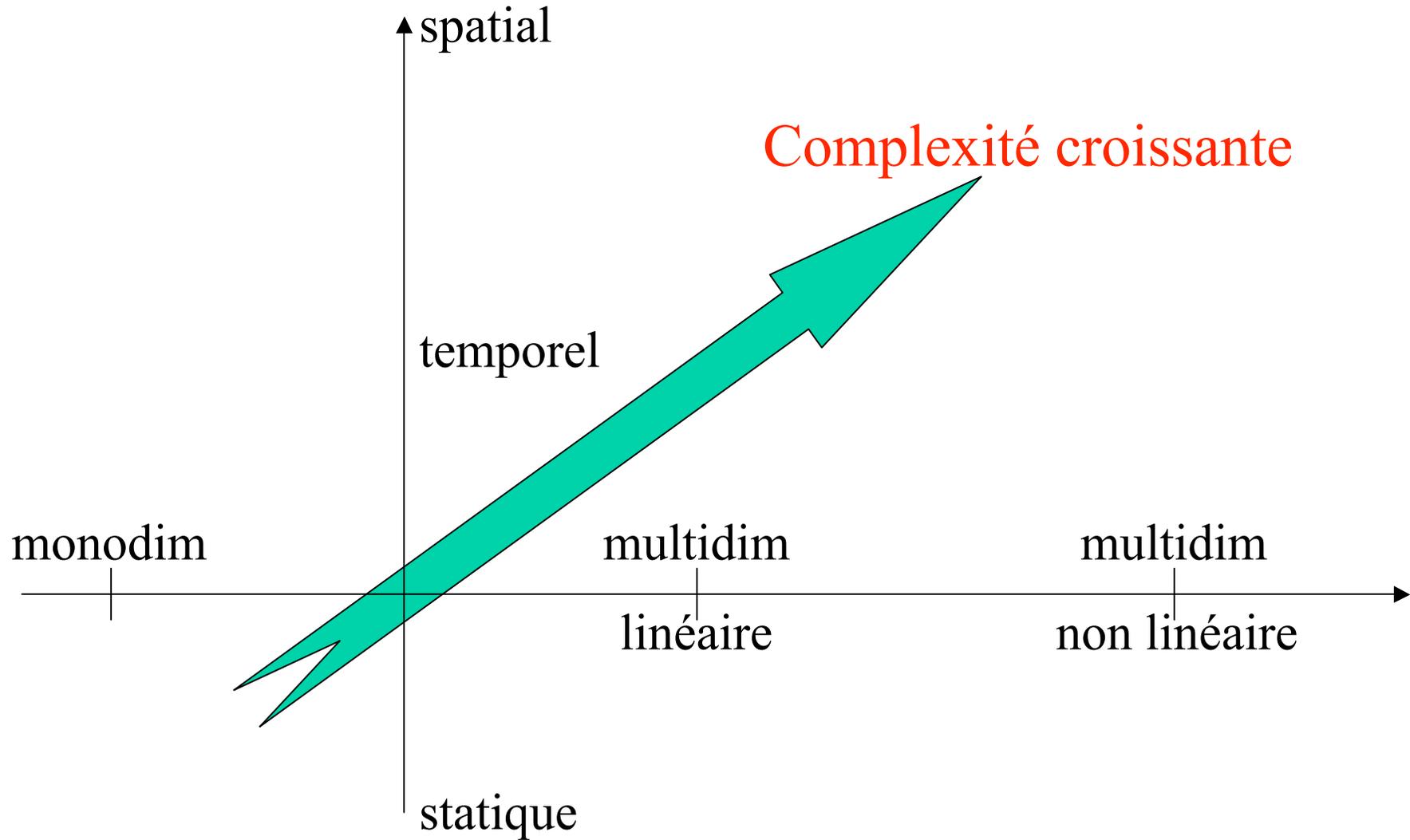
ENSM.SE

L. Carraro - 17 décembre 07

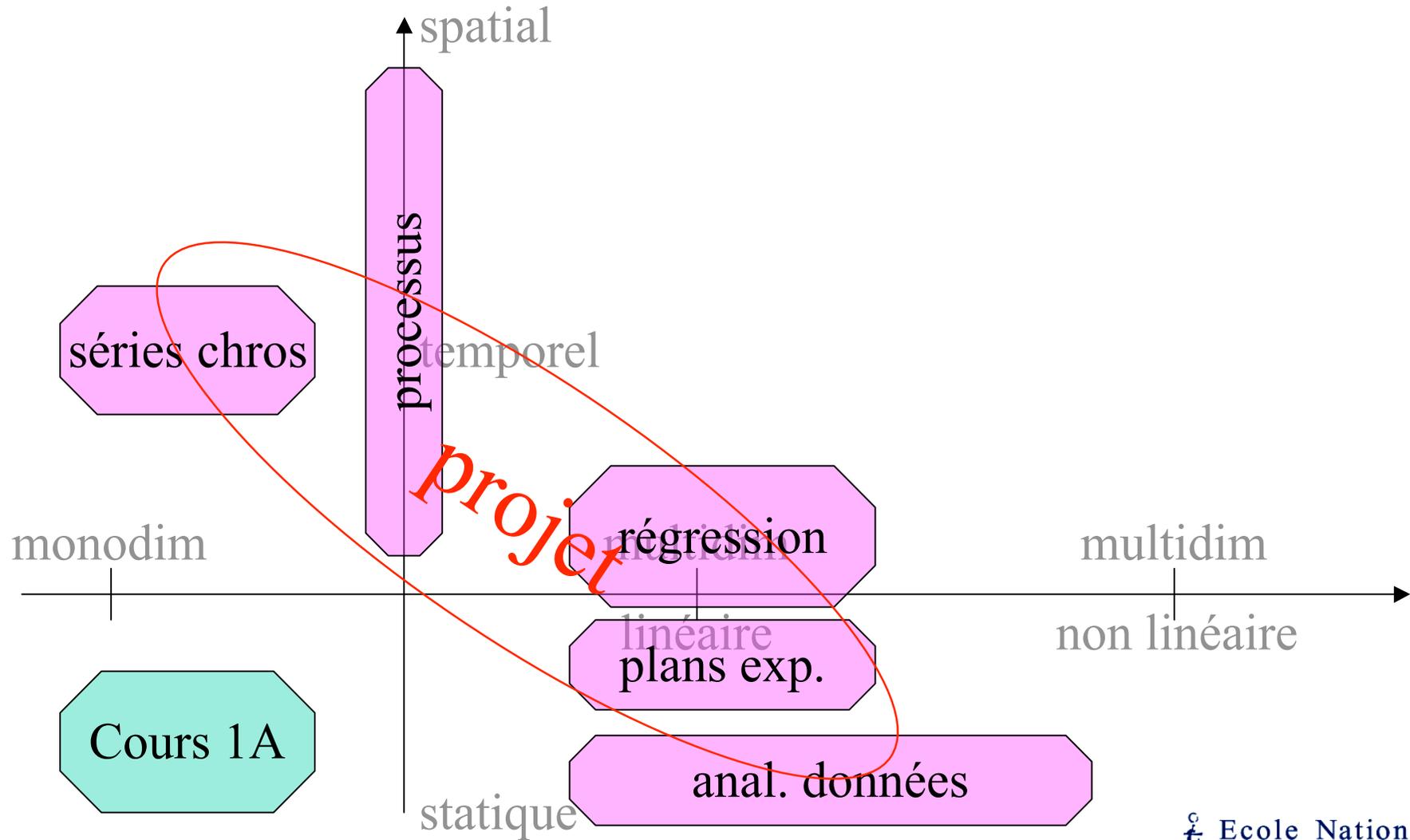
# Plan

- Compétences acquises
  - domaines scientifiques
  - compétences transverses
- Domaines ou activités accessibles :
  - immédiatement
  - au prix d'adaptations faciles
  - via des efforts conséquents
- Conclusion

# Cartographie (domaines scientifiques)



# Cartographie (domaines scientifiques)



# Cartographie

## Compétences transverses

- Questionnement sur un problème
  - Compréhension
  - Prise en compte du contexte, des contraintes
  - Rapport coût/qualité
- Communication
  - écrite
  - orale
- Travail en groupe

# Domaines ou activités accessibles immédiatement

- Etudes de données expérimentales
  - laboratoires (industrie, santé...)
  - mise en évidence de facteurs importants
  - service qualité
- Prévisions
  - prévisions de vente ou de consommation
  - finance, assurances
  - prévisions de coûts

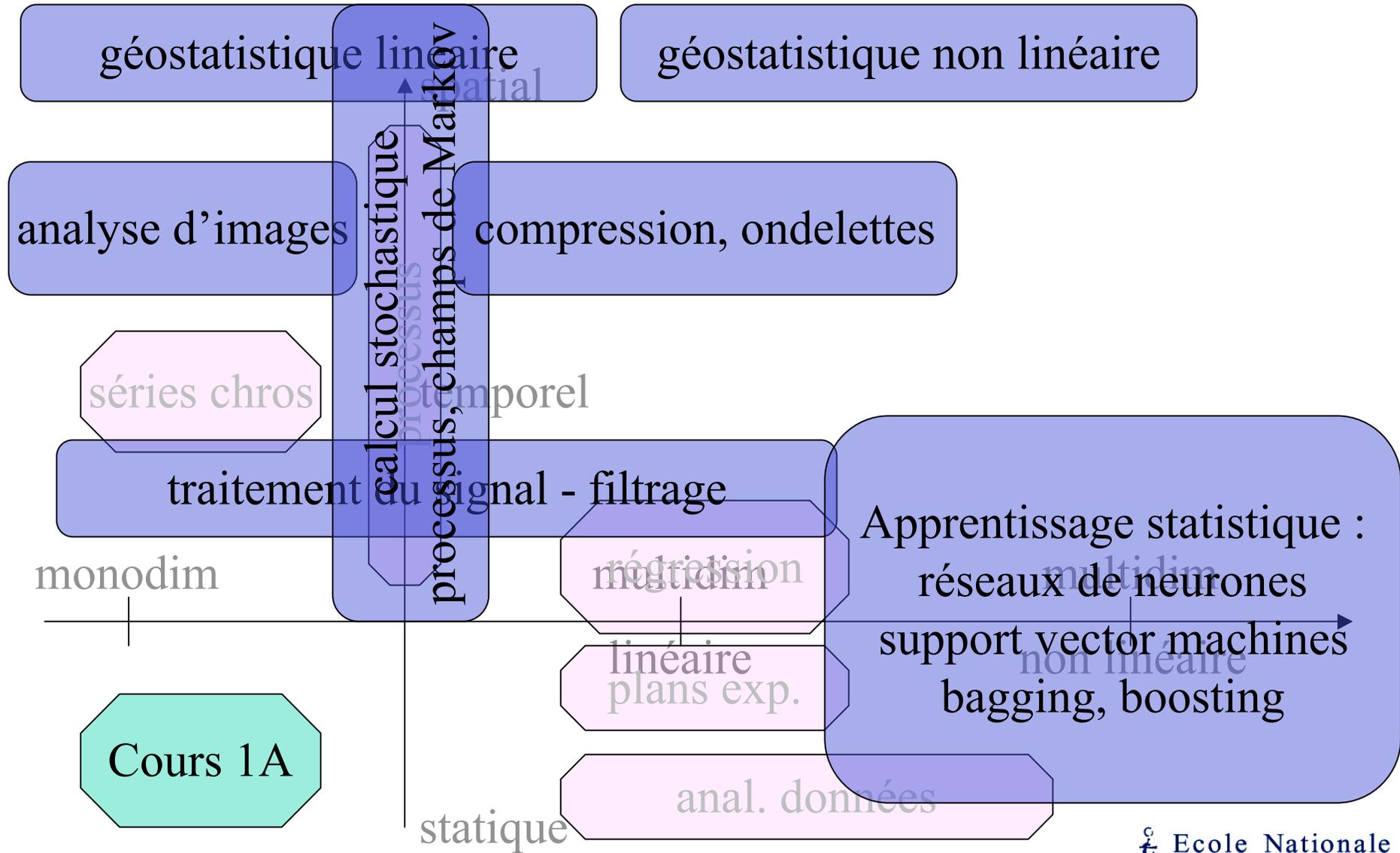
# Domaines ou activités accessibles immédiatement

- Aide à la décision
  - construction d'indicateurs
  - probabilisation de décisions
  - sociétés de conseil (finance, risques...)
- Appui méthodologique à un autre domaine
  - mécanique, matériaux
  - procédés, environnement
  - santé

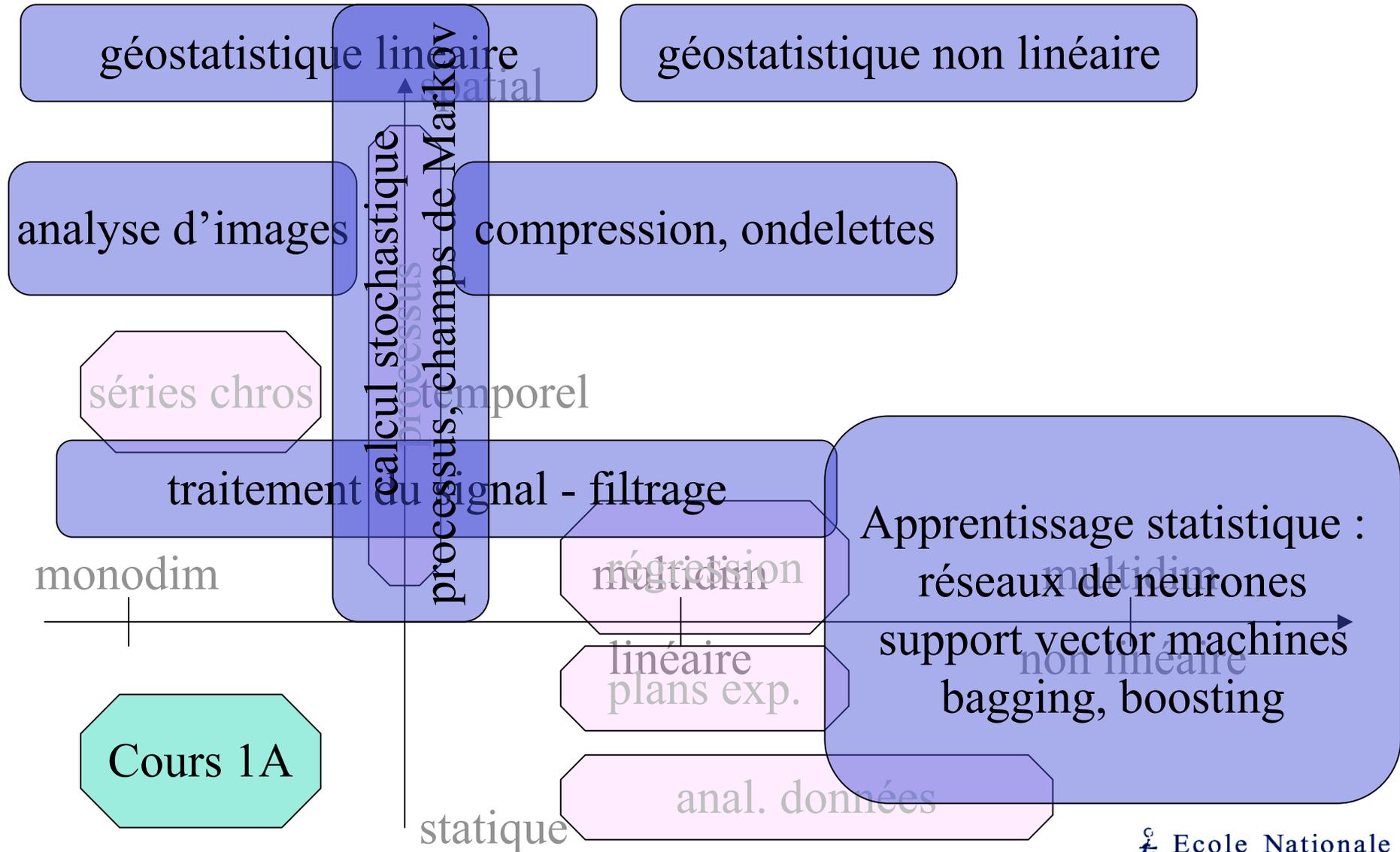
# Domaines ou activités accessibles facilement

- Statistical Process Control.
- Traitement statistique du signal.
- Séries temporelles multidimensionnelles, économétrie.
- Régression non linéaire.
- Méthodes non paramétriques en petite dimension.

# Aller plus loin ?



# Aller plus loin ?



# Filtrage de Kalman et modèles d'état

➤ Modèle d'état :

$$X_n = F X_{n-1} + G E_n \quad \text{équation d'état}$$

$$Y_n = H X_n + F_n \quad \text{équation d'observation}$$

$Y_n$  : vecteur  $p \times 1$  observé au temps  $n$

$X_n$  : vecteur  $q \times 1$  décrivant le système au temps  $n$

$E_n, F_n$  : vecteurs de résidus aléatoires

➤ Modèle discret d'équation différentielle  
bruitée, avec mesure.

# Cas d'un ARMA(2,1)

$$X_n - \phi_1 X_{n-1} - \phi_2 X_{n-2} = \varepsilon_n - \theta_1 \varepsilon_{n-1}$$

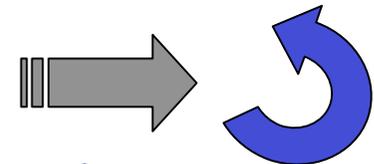
$$X_n = \begin{bmatrix} X_n \\ X_{n-1} \end{bmatrix}, E_n = \begin{bmatrix} \varepsilon_n \\ \varepsilon_{n-1} \end{bmatrix}, Y_n = X_n$$

$$X_n = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} X_{n-1} + \begin{bmatrix} 1 & -\theta_1 \\ 0 & 0 \end{bmatrix} E_n$$

$$Y_n = [1 \quad 0] X_n$$

# Problèmes traités

- Filtrage :  
 $E(X_n/Y_1, \dots, Y_n) = ?$
- Prédiction :  
 $E(X_{n+h}/Y_1, \dots, Y_n) = ?$   
 $\text{Var}(X_{n+h}/Y_1, \dots, Y_n) = ?$   
Loi de  $X_{n+h}/Y_1, \dots, Y_n) = ?$
- Algorithme de Kalman, valable en temps continu (ED stochastiques), non stationnaire.
- Applications innombrables aux pbs temps réel (cf. aéronautique)
- Vers le contrôle...



# Géostatistique

- Historique : exploitation minière (or)
- D. Krige, G. Matheron
- Modélisation de phénomènes spatiaux :
  - Géométrie en 2D ou 3D (ou ND !)
  - Observations non régulières
- Observation d'une fonction  $y(x)$  en certains points  $x$  de l'espace.
- Objectif = prévoir  $y(x)$  pour des  $x$  non observés ou inobservables.

# Modèle probabiliste

$y(x)$  est supposé être la réalisation d'un processus gaussien  $Y(x)$  tel que :

$$E[Y(x)] = m$$

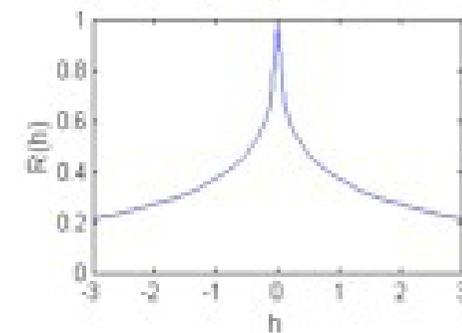
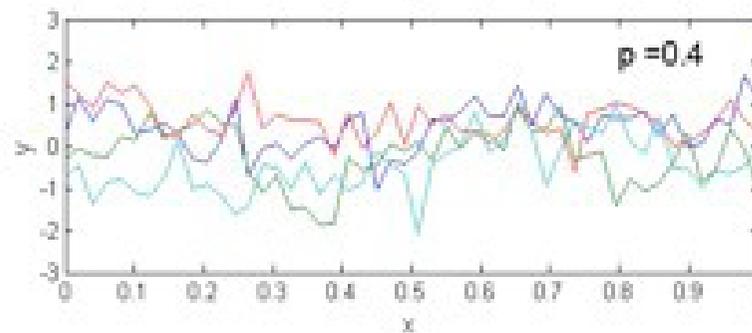
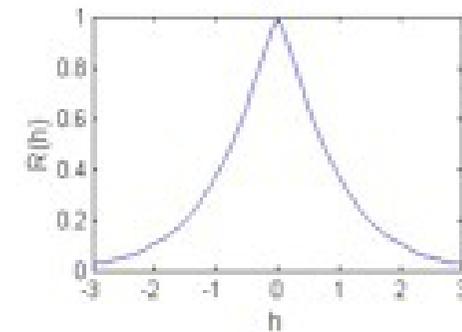
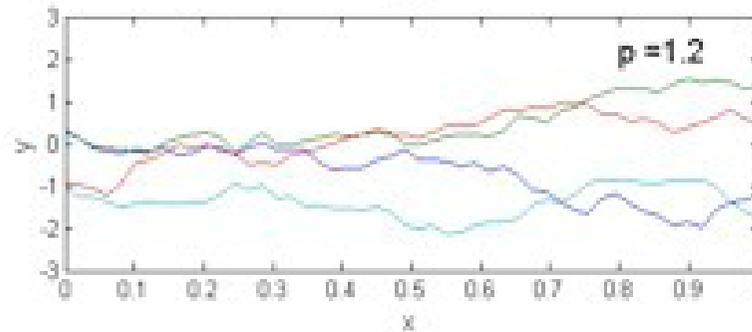
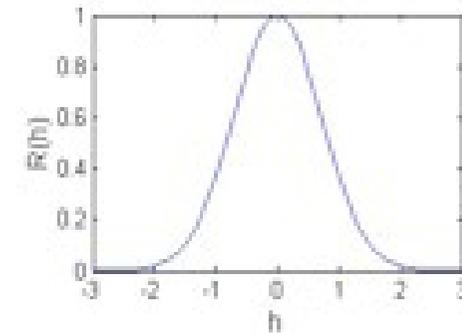
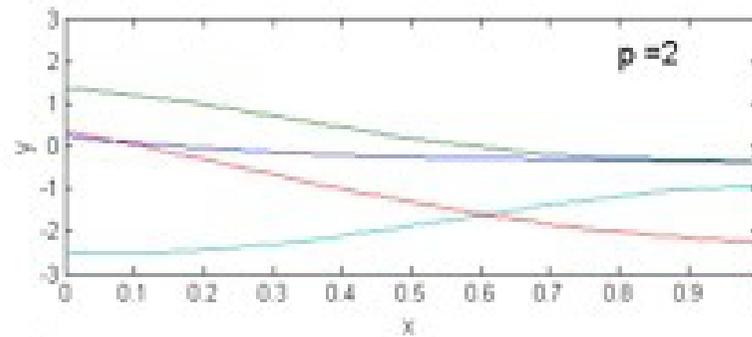
$$E[(Y(x+h) - Y(x))^2] = 2\gamma(h)$$

$\gamma$  est le (demi-) **variogramme** du processus

Si  $\text{Var}(Y(x)) = \sigma^2$  :

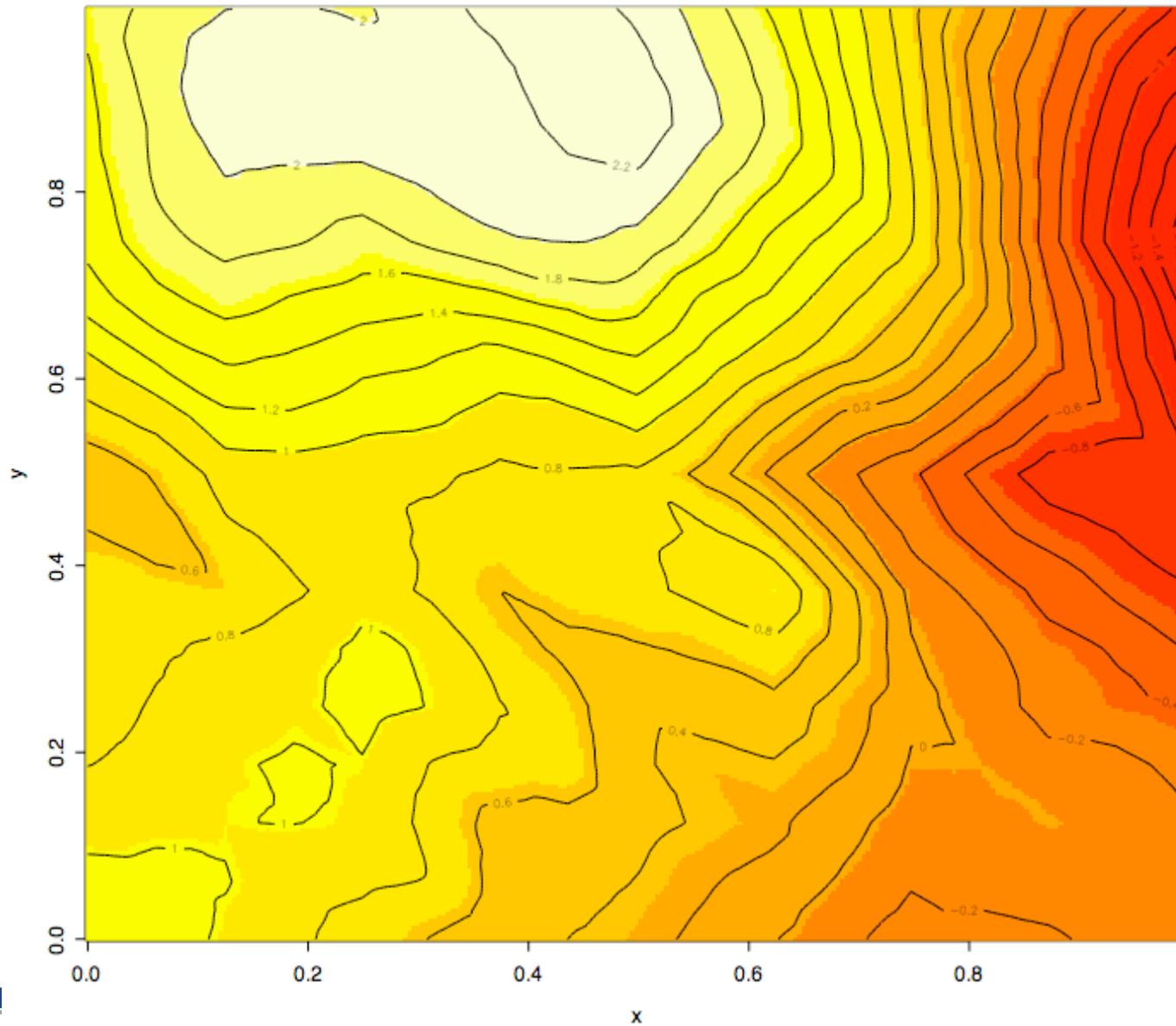
$$\text{cov}(Y(x), Y(x+h)) = \sigma^2 - \gamma(h)$$

# Lien variogramme-processus



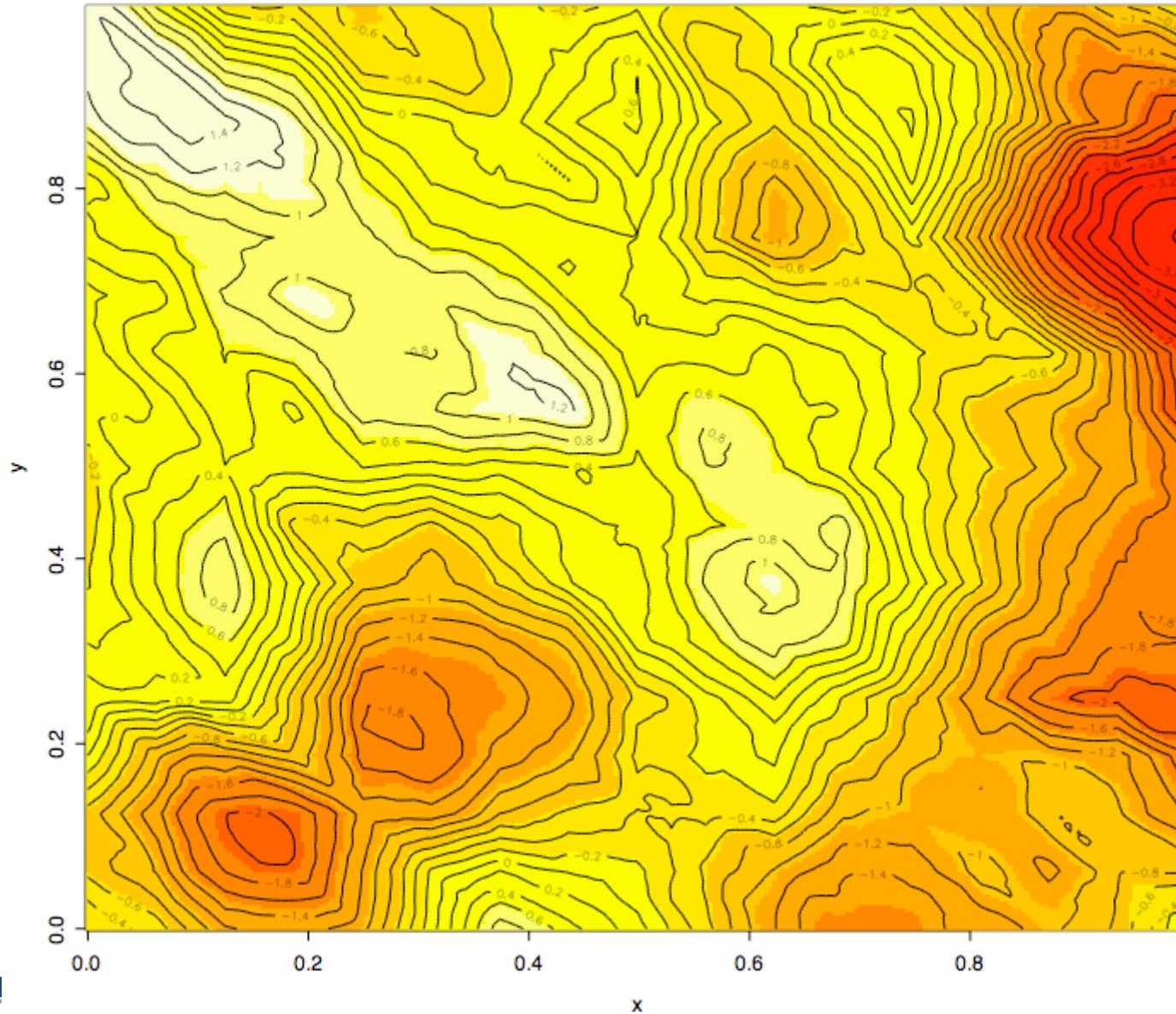
# Simulations en 2D

champ gaussien de covariance isotrope gaussienne  
portée = 0.5



# Simulations en 2D

champ gaussien de covariance isotrope gaussienne  
portée = 0.2



# Applications

Tous les domaines qui mettent en œuvre des variables spatiales non totalement accessibles :

- Mines, pétrole, gaz
- Environnement
- Halieutique
- Computer experiments
- ...

# Estimation de la longueur moyenne du hareng écossais

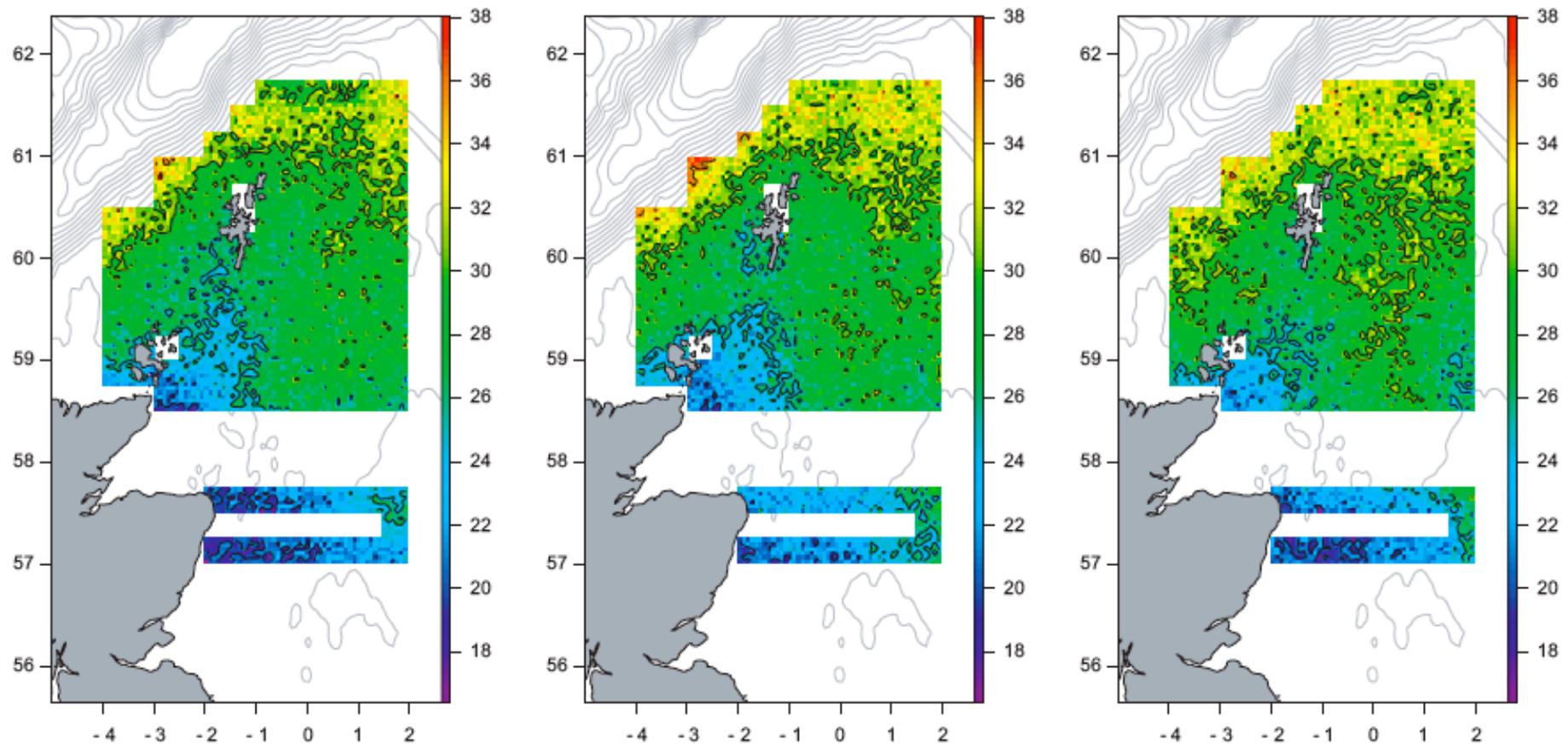
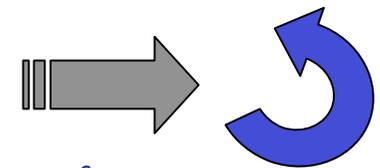
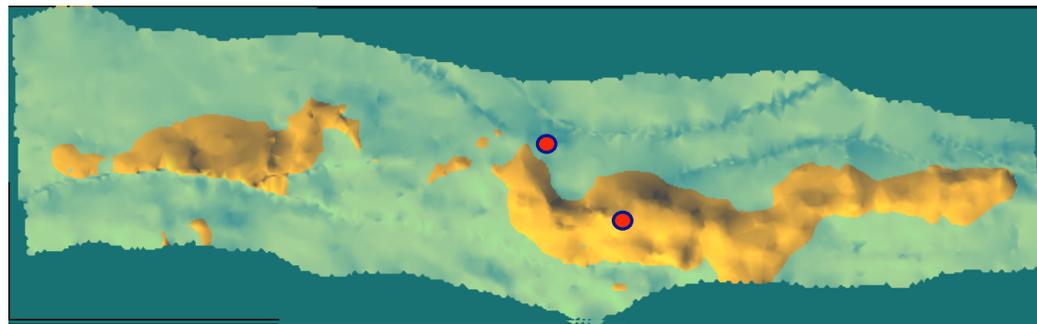


FIG. 3.16 – Trois réalisations (faible, moyenne, forte) de la simulation conditionnelle de la variable longueur moyenne (cm).

# Carte au toit d'un champ pétrolier



# Vers l'apprentissage statistique

## ➤ Arbres de régression :

### Principe :

- Première séparation :

Parmi tous les prédicteurs et tous les niveaux, trouver ceux qui séparent la réponse en deux groupes les plus hétérogènes (déviante)

- Séparations suivantes :

Recommencer dans chaque sous-groupe

- Elagage de l'arbre

Remonter dans l'arbre et couper les branches homogènes.

# Exemple : données de spam

- 1 réponse qualitative : spam ou pas spam
- 57 prédicteurs : fréquence d'apparition de caractères ou de mots.
- 4601 observations

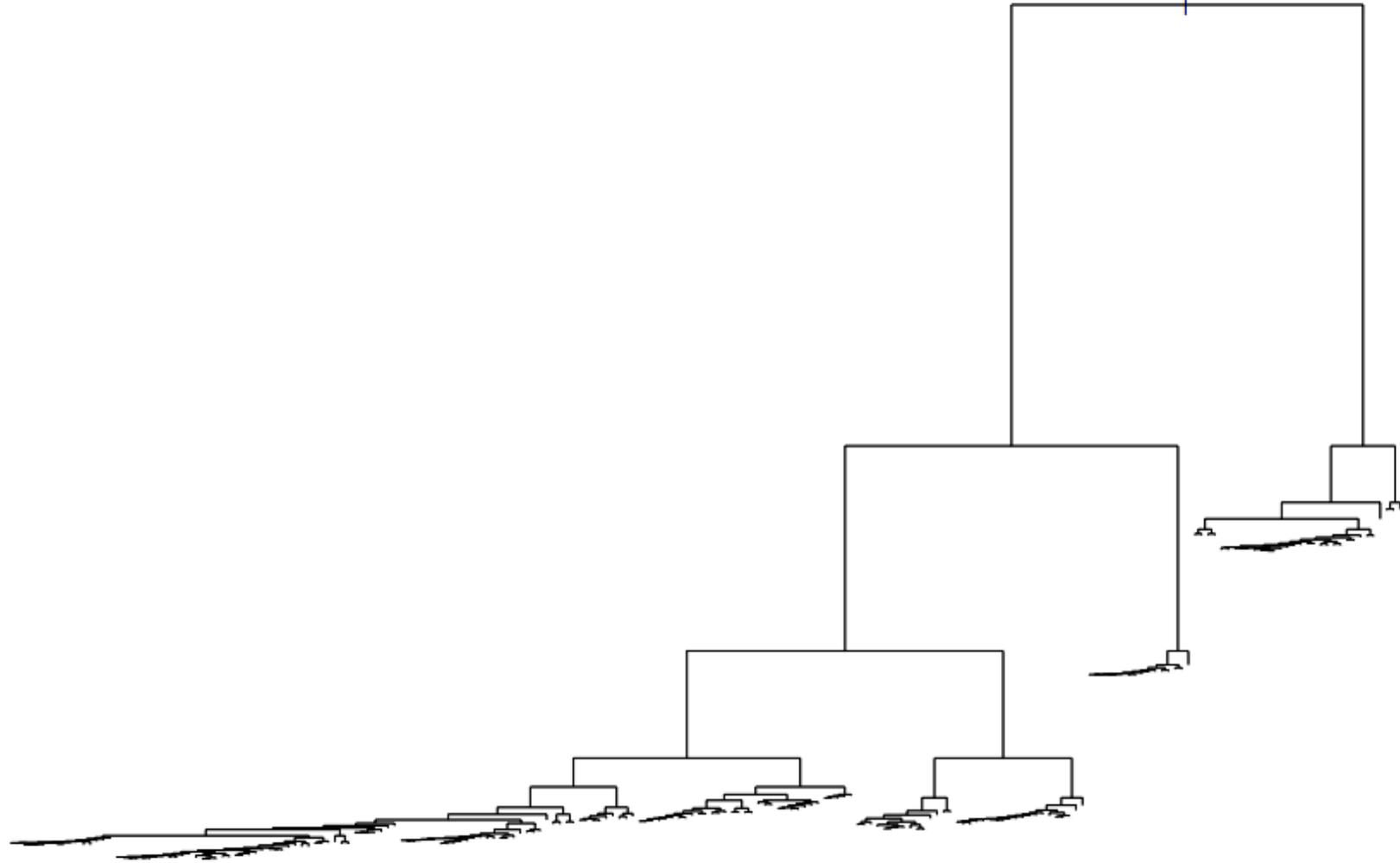
NB :

X. = ;            X..3=!

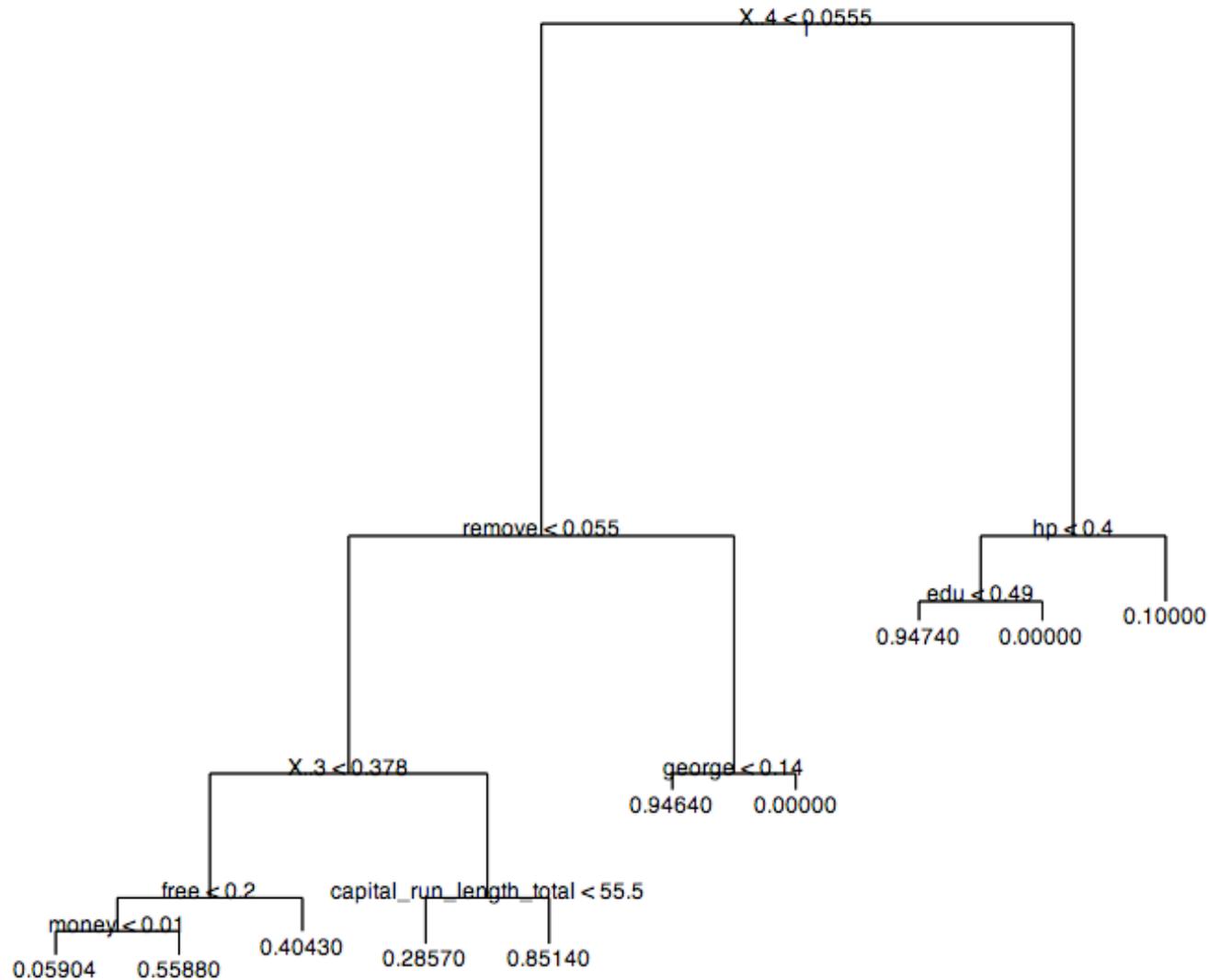
X..1=(            X..4=\$

X..2=[            X..5=#

# Arbre avant élagage



# Arbre après élagage





# Modèles de ruine

- Origine : assurances sinistres, conception de barrages, fils d'attente...

Réserve de risque  $Z(t)$  au temps  $t$  :

$$Z(t) = x + \beta t - X(t)$$

–  $x$  = capital initial

–  $\beta$  = taux (constant) de primes d'assurances entrantes

–  $X(t)$  = montant des sinistres cumulé au temps  $t$

$$X(t) = \sum_{k=1}^{N_t} X_k$$

# Hypothèses et premiers éléments

- $(N_t)_t$  est un processus de Poisson homogène d'intensité  $\lambda$
- Les variables  $X_k$  sont i.i.d. d'espérance  $\mu$

$$E(X(t)/N_t) = \mu N_t$$

Et par suite :

$$E(Z(t)) = x + \beta t - \lambda \mu t$$

Donc on doit avoir

$$\beta > \lambda \mu$$

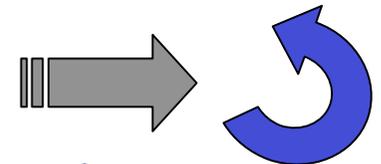
# Vers la théorie de la ruine

Explicitation du lien entre  $x$  et  $P(T < +\infty)$

$$T = \text{instant de ruine} = \text{Inf}\{t, Z(t) < 0\}$$

Difficulté du problème :

- Théorique (voir martingales)
- Numérique (processus à sauts)
- Introduction de la réassurance...



# Conclusion

- Axe MSA = des bases solides
- Compléments méthodologiques :
  - option maths applis
  - finance, assurance
  - autres...
- Compléments applicatifs :
  - tous domaines