Laboratoire Central des Ponts et Chaussées

Réseau Scientifique et Technique de l'Equipement

1949 • 2009 LCPC 60 ans d'innovation

LES PLÉNIÈRES 2010 DU LCPC

Sciences et techniques du **Génie Civil**

JOURNÉES ACOUSTIQUE

Wissembourg – 2 et 3 JUIN 2010

Estimation robuste et analyse de données expérimentales

Application au bruit de roulement

Pierre Charbonnier (LRS - ERA 27 LCPC)
Guillaume Dutilleux (LRS - ERA 32 LCPC)
Renaud Wintzer (Stagiaire, ERA 32)

PLAN

- **Introduction**

  Problématique et état de l'art

- **M-estimateurs**

  Définition, optimisation, algorithmes

- **Application**

  Régression de courbes de bruit v.s. vitesse

- **Conclusion**

INTRODUCTION

- **Cadre général**

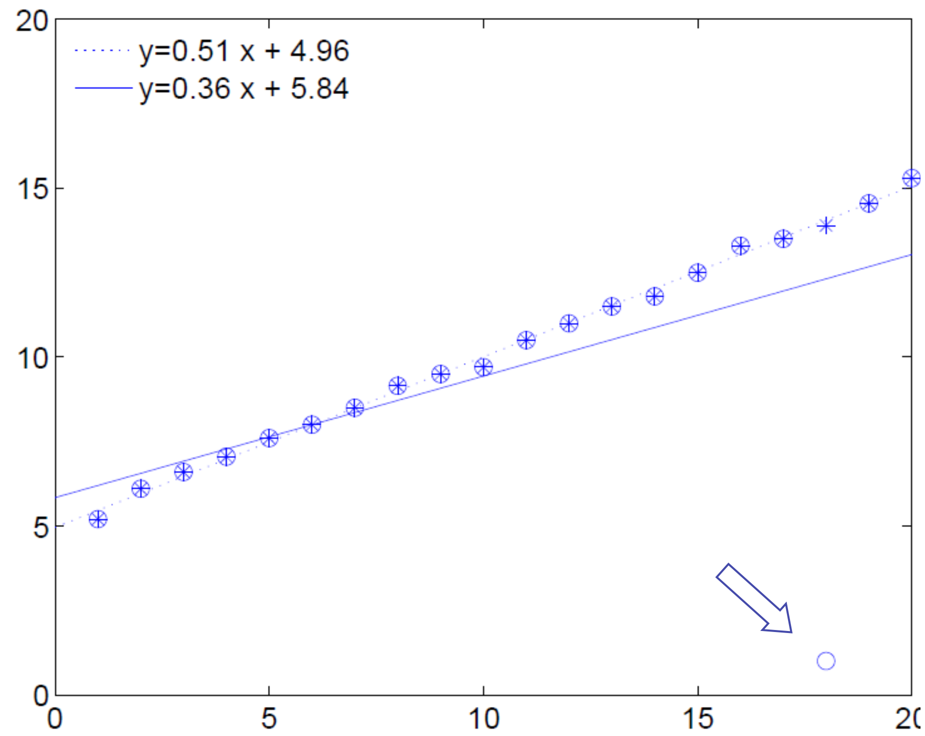
  Régression linéaire

  y=ax+b

- **Problématique**

  La régression est sensible à la présence de données erronées (ou *outliers*)…

  Comment s'en affranchir?



➔ Estimation robuste

PRINCIPALES APPROCHES

- **Diagnostic et rejet**
  Identifier les données erronées et les éliminer de l'estimation

- **Estimation statistique robuste**
  Remplacer le critère classique des moindres carrés
  M-estimateurs, R-estimateurs, MM-estimateurs…

- **Méthodes de vote**
  Quantifier l'espace des paramètres et accumuler les votes : transformée de Hough

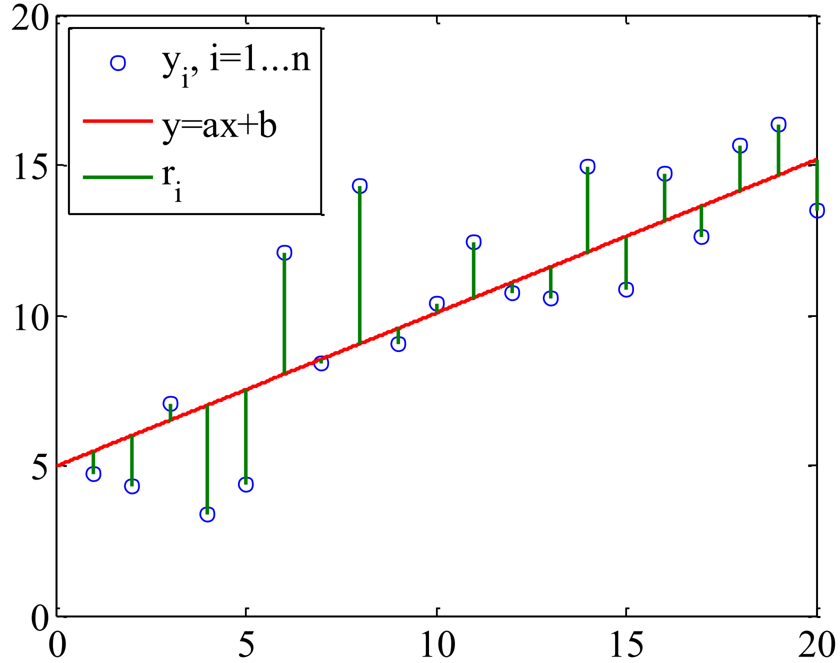
- **Méthodes stochastiques**
  Agglomérer autant de données que possible autour d'un ajustement minimal : RANSAC

NOTATIONS

- **Modèle**

$$\begin{cases} y_1 = ax_1 + b \\ \vdots \\ y_i = ax_i + b \\ \vdots \\ y_n = ax_n + b \end{cases} \quad \text{soit} \quad \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_i & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}}_{} . \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\mathbf{y} = \mathbf{R.f}$$



ESTIMATION DES MOINDRES CARRES

- **Notion de résidu**

  Écart (aléatoire) entre modèle et observation $\mathbf{y} = \mathbf{R}.\mathbf{f} + \mathbf{r}$

- **Moindres carrés**

  Minimiser la norme quadratique du résidu $e_{LS}(\mathbf{r}) = \sum_{i=1}^{n} \mathbf{r}_i^2 = \|\mathbf{y} - \mathbf{R}\mathbf{f}\|^2$

  Solution :

$$\boxed{\mathbf{f}_{LS} = \left(\mathbf{R}^T\mathbf{R}\right)^{-1}\mathbf{R}^T\mathbf{y}}$$

- **Raison du manque de robustesse ?**

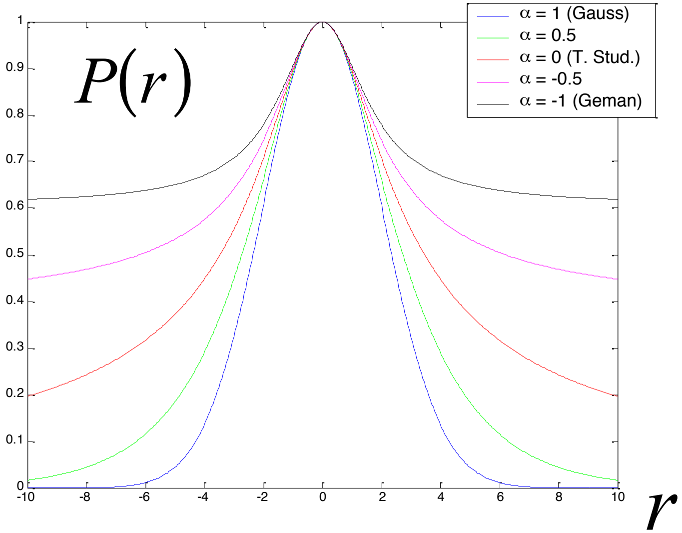
  Sous-jacent : modèle gaussien du résidu

  Fort résidu $\Rightarrow$ probabilité $\approx$ nulle !

LES M-ESTIMATEURS (HUBER, 1964)

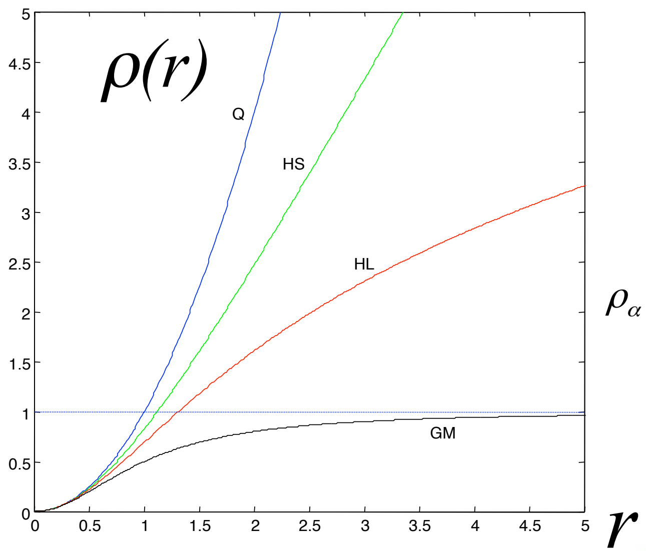
- En probabilités

$$P(\mathbf{y}|\mathbf{f}) = P(\mathbf{r}) \propto \exp\left[-\sum_{i=1}^{N} \rho(\mathbf{r}_i)\right]$$



- En énergie

$$e_M(\mathbf{r}) = \sum_{i=1}^{N} \rho(\mathbf{r}_i)$$



$$\rho_\alpha(r) = \frac{1}{\alpha}\left[\left(1 + r^2\right)^\alpha - 1\right]$$

CHOIX DE LA FONCTION ρ ?

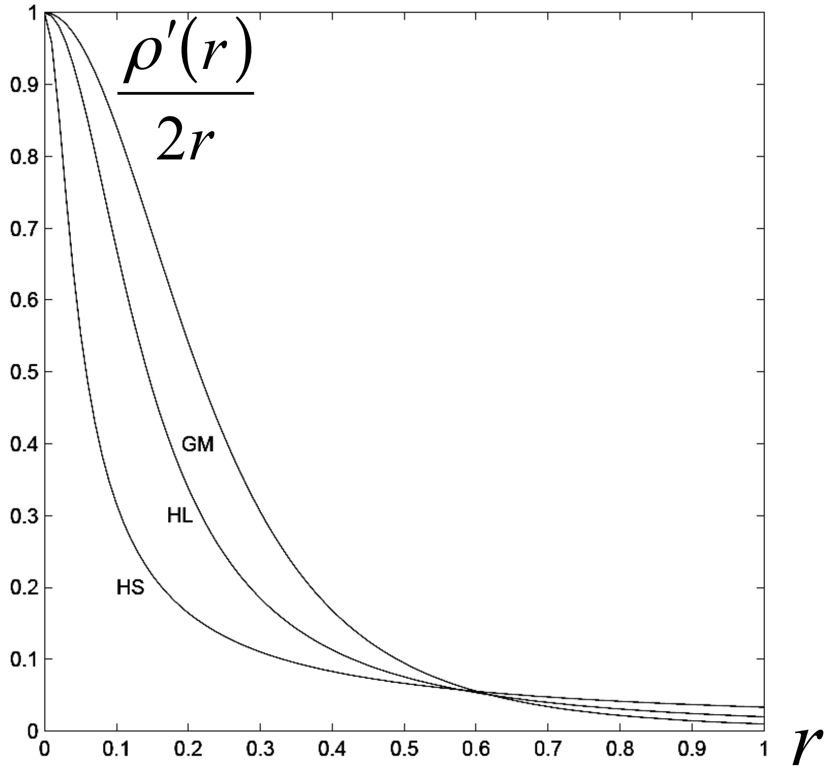
- **Formes variées**

Critères : robustesse, efficacité, convexité…

- **Conditions unifiées**

Fonction *de pondération*

$$(*)\quad \begin{cases} \lim_{r\to 0} \dfrac{\rho'(r)}{2r} = 1 \\ \lim_{r\to +\infty} \dfrac{\rho'(r)}{2r} = 0 \\ \dfrac{\rho'(r)}{2r} \quad \text{strictement décroissante} \end{cases}$$



THEORIE SEMI-QUADRATIQUE

- **Introduction d'une variable auxiliaire, λ**

  Aide à linéariser et marque les données erronées

- **Théorème (Geman92, Charbonnier95)**

  Si ρ satisfait les conditions (*)

  Il existe ξ strictement convexe et décroissante, telle que

$$\rho(r) = \inf_{0 \leq \lambda \leq 1} \left(\lambda r^2 + \xi(\lambda) \right)$$

  Pour tout r ≥ 0 fixé, l'inf est atteint pour une valeur unique, λ_r

$$\lambda_r = \frac{\rho'(r)}{2r}$$

ALGORITHME SEMI-QUADRATIQUE

- **Critère augmenté**

$$e^*(\mathbf{r},\boldsymbol{\lambda}) = \sum_{i=1}^{n} \left\{ \lambda_i \mathbf{r}_i^2 + \xi(\lambda_i) \right\} = \left\| \mathbf{y} - \mathbf{R}\mathbf{f} \right\|_B^2 + Cte(\boldsymbol{\lambda}) \qquad \text{où} \qquad B = diag\left\{ \rho'(r_i)/2r_i \right\}_{i=1\ldots n}$$

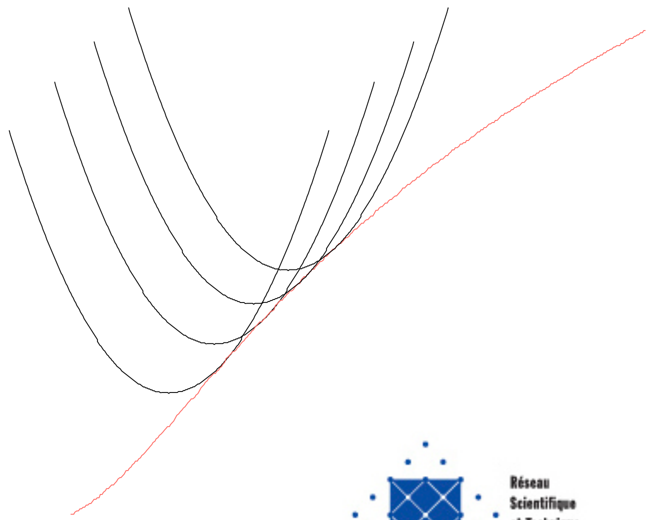
- **Optimisations alternées**

Série d'optimisations quadratiques

- **Moindres carrés pondérés itératifs**

Itération *IRLS* :

```
r=(y-R*f);
B=diag((1+r.^2).^(alpha-1));
f=(R'*B*R)^(-1)*R'*B*y;
```



LA QUESTION DE L'ECHELLE

• Le véritable modèle

$$e_M(\mathbf{r}) = \sum_{i=1}^{n} \rho(\mathbf{r}_i/\sigma) + n\ln(\sigma)$$

• Comment gérer l'échelle ?

- Utiliser un estimateur invariant / échelle, ex: norme L^p $(1 \leq p < 2)$
- Fixer l'échelle arbitrairement
  - Étude des résidus
- Estimer l'échelle à partir des données
  - En préalable à l'algorithme
  - À chaque étape de l'algorithme ? ➔ à éviter

APPLICATION BRUIT DE ROULEMENT

• Caractérisation acoustique des revêtements

Mesures de vitesse et de bruit

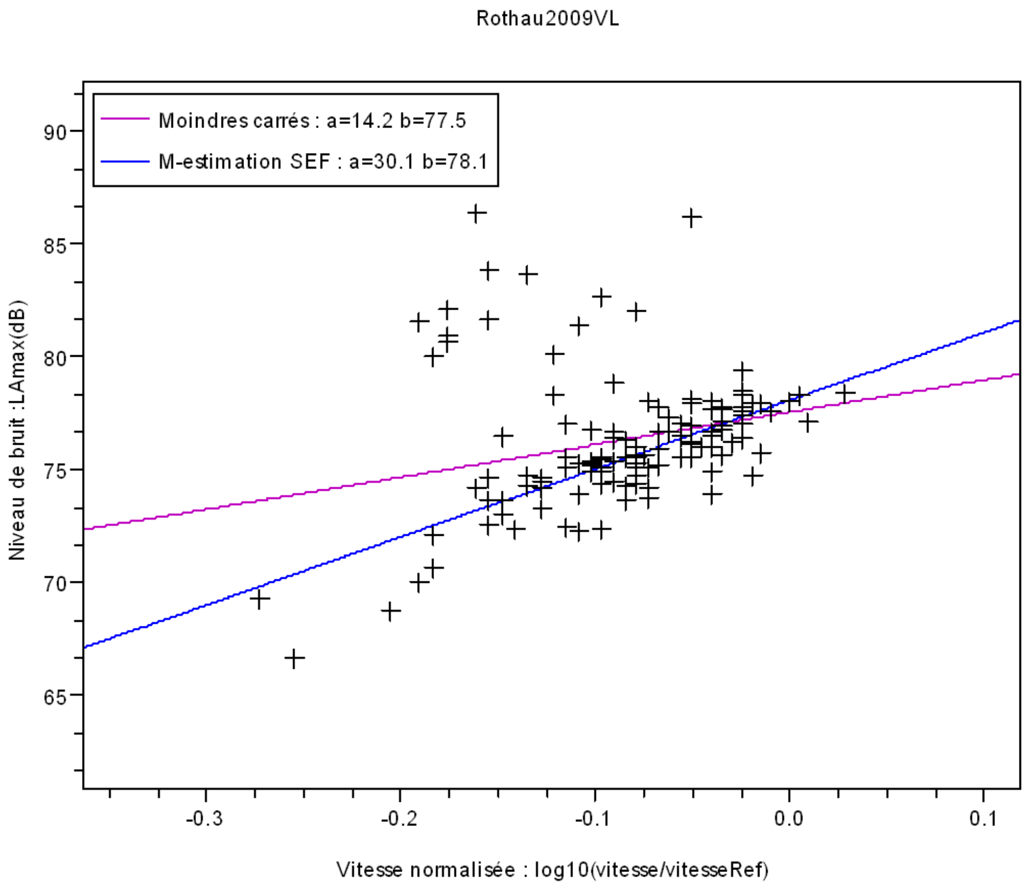
• Dispositif

Un microphone et un cinémomètre en bord de voie

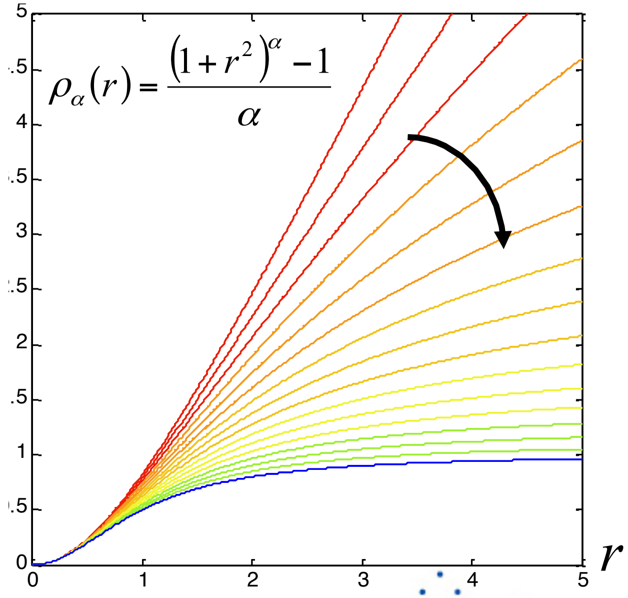
• Traitement

- Contrôle et filtrage des mesures
- Régression linéaire
- Calcul du niveau de bruit « type » à une vitesse de référence

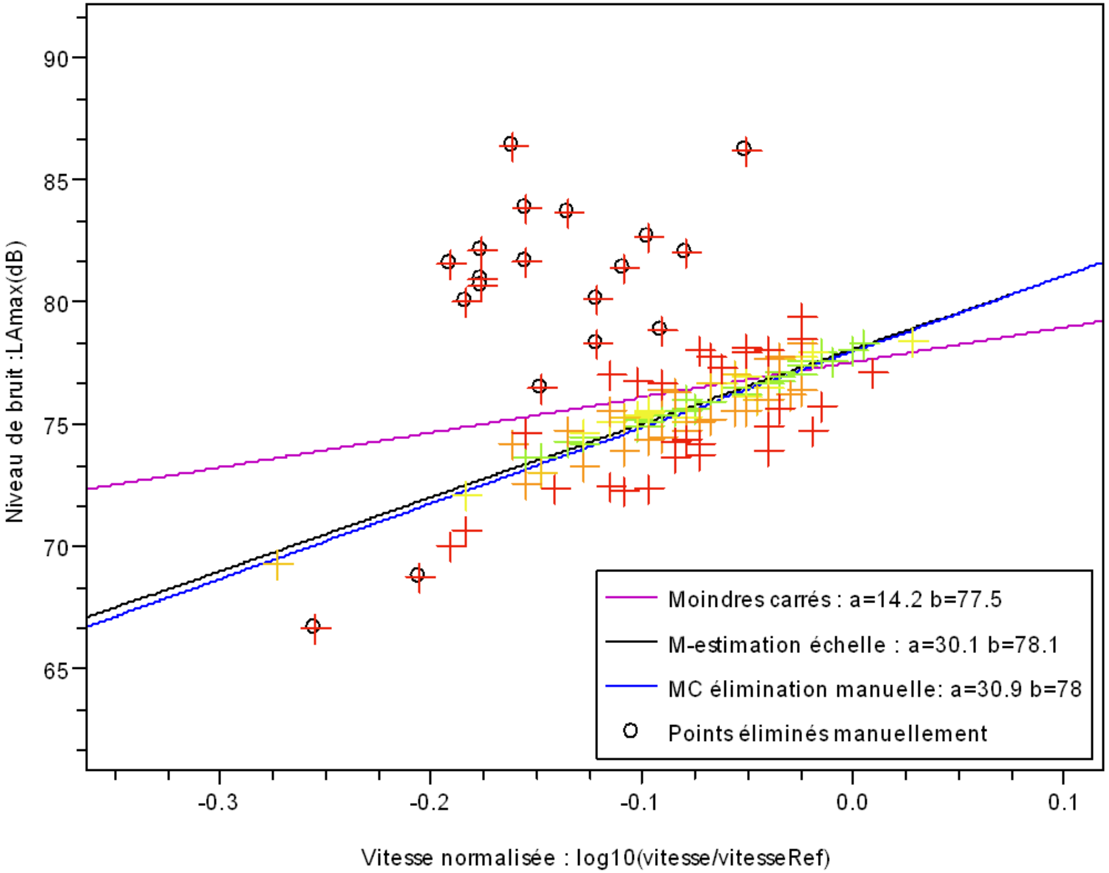
FONCTIONNEMENT ALGORITHME



- Estimation échelle
- Estimation en continuation



VISUALISATION PONDERATIONS



- **Mesures**

  Secteur de Rothau (67)

  Véhicules légers

- **Code couleur**
  - + $0,0 < \lambda_i \leq 0,2$
  - + $0,2 < \lambda_i \leq 0,4$
  - + $0,4 < \lambda_i \leq 0,6$
  - + $0,6 < \lambda_i \leq 0,8$
  - + $0,8 < \lambda_i \leq 1,0$

CONCLUSION

- ## Estimation robuste

  Un moyen « automatique » de s'affranchir des données erronées

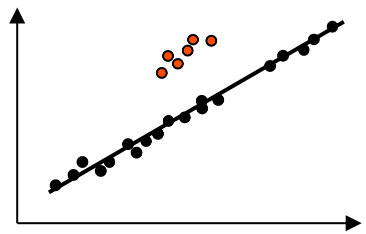
  Algorithmique simple (M.C. pondérés)

- ## Application au bruit de roulement

  Premiers résultats encourageants…

  Modèle de base : question du bruit sur les vitesses

  Que nous disent les données erronées ?



- ## Autres exemples d'application des méthodes robustes et semi-quadratiques

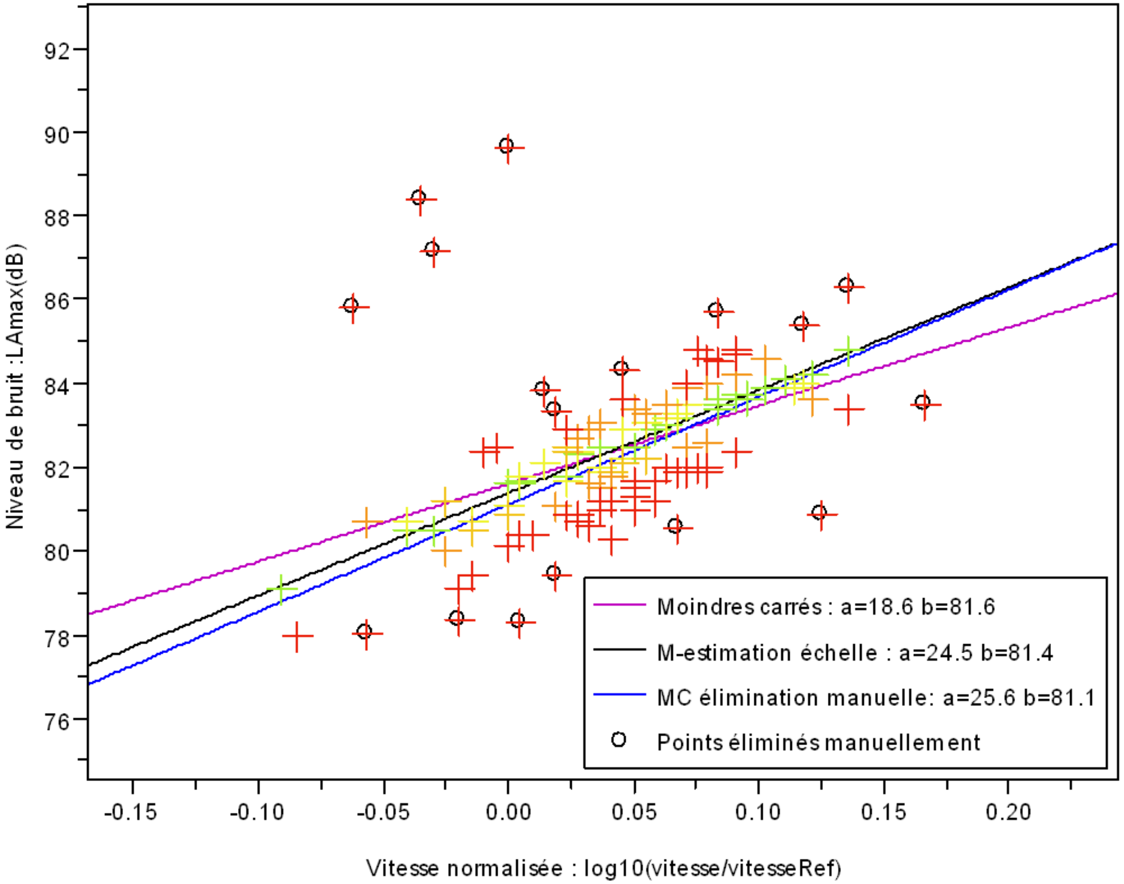
  Avec LIVIC, LEPSiS (JP. Tarel) : détection et suivi des marquages routiers

  LRS Détection et reconnaissance des panneaux de signalisation

  Régularisation des problèmes inverses

"Any question?"

COMPARAISON



- **Mesures**

  Secteur Haguenau (67)

  Véhicules légers

- **Code couleur**

  + $0,0 < \lambda_i \leq 0,2$
  + $0,2 < \lambda_i \leq 0,4$
  + $0,4 < \lambda_i \leq 0,6$
  + $0,6 < \lambda_i \leq 0,8$
  + $0,8 < \lambda_i \leq 1,0$