

# Prévision des pics d'ozone en Bretagne

J. Josse

<sup>1</sup>Laboratoire de Statistiques, Université Rennes 2

<sup>2</sup>Air Breizh

Date de la soutenance de stage  
Rennes, 25 septembre 2006

# Plan

- **Présentation.**
- **Problème, objectif.**
- **Les données.**
- **Méthodologie.**
- **Résultats.**
- **Conclusion, perspectives.**

## Présentation du sujet

- **Ozone phénomène complexe** :  $NO_2 + O_2 \leftrightarrow NO + O_3$ .
- **Pollution à l'ozone est un problème aux enjeux importants** : santé publique, végétation...
- **Mission d'Air breizh** : mesure, analyse, prévision  $\Rightarrow$  Envoie avant 17 heures tous les jours de l'indice ATMO du lendemain aux autorités.
- **Prev'Air** : modèle déterministe.
  - Simulation de la qualité de l'air.
  - Macro-échelle.

$\Rightarrow$  **Idée** : modèle statistique pour Rennes : local.

# Problématique

- **Nous voulons prévoir aujourd'hui avant 17 heures le pic d'ozone de demain.**
- **Prévision de la concentration maximale d'ozone à partir de données météorologiques prévues par Météo France et de données de concentration de polluant.**

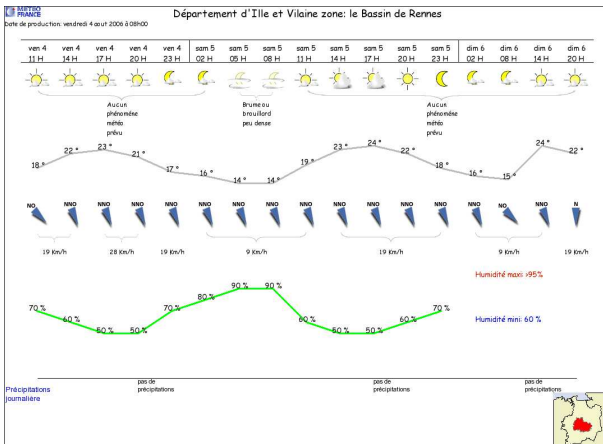
⇒ choix de la méthode.

⇒ choix des variables pertinentes.

## Les données initiales

- 1<sup>er</sup> janvier 1995 au 31 décembre 2005.
- Données météo (horaire).
  - La température (Celsius)
  - La nébulosité (0-8)
  - Le vent : vitesse et direction
  - Les précipitations (en mm)
  - La radiation, la pression et l'humidité (ne peuvent être prévus)
- Données concentration de polluant.
  - Ozone (en  $\mu\text{g}/\text{m}^3$ ) toutes les heures
- Selection du maximum d'ozone  $\Rightarrow$  variable à expliquer : maxO3.
- Création de vents projetés.
- Selection des variables à des heures ponctuelles.

# Atmogramme



## Les données de travail

- maxO3 : Concentration maximale du jour.
- maxO3v : Concentration maximale de la veille.
- T6, T9, T12, T15, T18 : Température à des heures ponctuelles 6 heures, 9 heures, 12 heures, 15 heures et 18 heures.
- Ne6, Ne9, Ne12, Ne15, Ne18 : Nébulosité à des heures ponctuelles.
- Vx6, Vx9, Vx12, Vx15, Vx18 : Projection du vent sur l'axe E-O à des heures ponctuelles.
- Vy6, Vy9, Vy12, Vy15, Vy18 : Projection du vent sur l'axe N-S à des heures ponctuelles.
- → Avril à septembre.

⇒ **Et les précipitations ?**

# Les précipitations

- Heure d'apparition des pics: 11 heures-18 heures.
- Précipitations cumulées le soir, le matin?
- Phénomène de lessivage.

⇒ **Création Base pluie et base sans pluie.**

## Deux grandes familles d'approches

$$Y = r(x) + \varepsilon$$

- Deux grandes familles d'approches
  - Approche paramétrique
    - Hypothèse restrictive sur la fonction de régression.
    - Nombre fini de paramètres à estimer.
  - Approche non paramétrique
    - Aucune hypothèse à priori sur l'allure de la fonction de régression.
    - Pas de paramètre à estimer.
- Principales méthodes de regression non paramétriques :  
noyaux, splines...

## Modèles additifs

- On va simplifier la forme de la fonction de régression  $r$  en supposant que cette fonction est additive:

$$Y = \alpha_0 + \sum_{j=1}^p r_j(X_j) + \varepsilon$$

- Généralise les modèles de régression linéaire.
- Permet une interprétation de l'effet marginal de chacune des variables sur la fonction de régression.
- **Méthode d'estimation** : Backfitting, MCO.

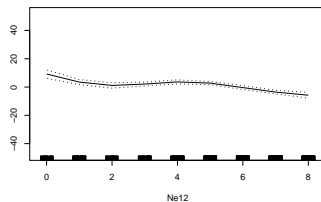
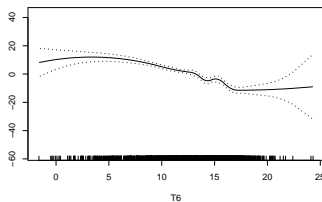
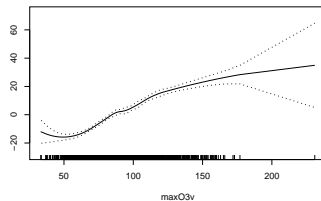
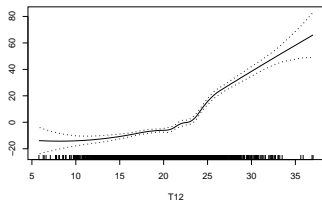
## Choix des variables

- Apprentissage (été 1995-2004)/ Validation (été 2005).
- Choix d'un critère d'erreur:  $MAPE = 100 * \frac{|PRED - OBS|}{|OBS|}$ .
- Choix d'un modèle.
- Modèle sans pluie:

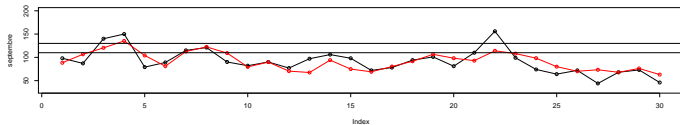
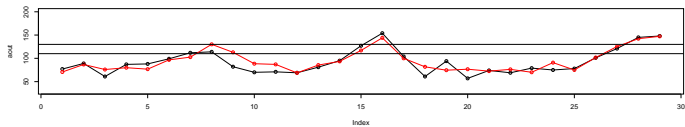
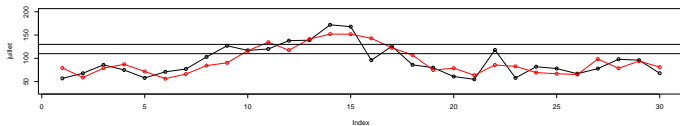
$$MaxO3 = s_1(T12) + s_2(maxO3v) + s_3(T6) + s_4(Ne12) + \varepsilon$$

- erreur : 11 %, 9.8%, 9.8%.

# Estimations des fonctions



# Résultats 2005



## Prévision en temps réel

- **Erreur:** 12.74 % le soir.
- erreur  $>30 \mu\text{g}/\text{m}^3$  : 5 %.
- erreur prev air  $>30 \mu\text{g}/\text{m}^3$  : 17 %.

⇒ **Création de AirBreizhStat.**

# Bilan

- Elaboration d'une méthodologie pour la prévision des pics d'ozone à Rennes.
- Modèle additif, Erreur été 2005  $< 12\%$ .
- Phase opérationnelle : prévision en temps réel + interface graphique.
- Modèle performant avec peu de variables.

# Perspectives

- Modèle pour Lorient.
- Combiner plusieurs méthodes.
- Incorporer au fur et à mesure les nouvelles observations.
- Nouvelles variables.

# Malediction de la dimension

- Il est difficile d'étendre à plusieurs dimensions les régressions non paramétriques.

