Agrégation séquentielle pour la prévision de la qualité de l'air

Vivien Mallet[1,2], Bruno Sportisse[1,2],
Gilles Stoltz[3,4], Boris Mauricette,
Sébastien Gerchinovitz[3,1]

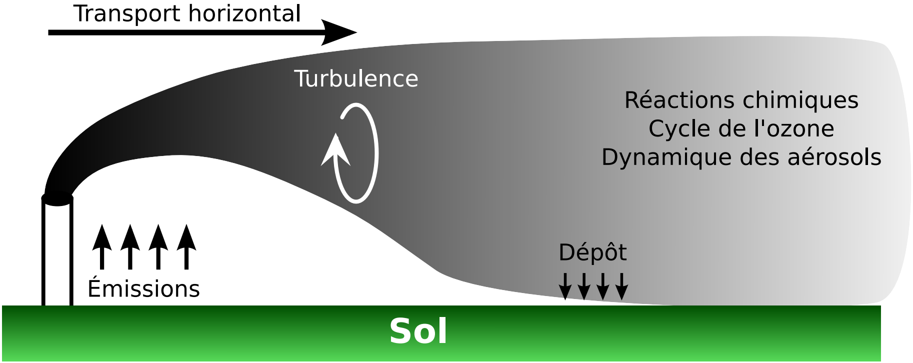
Journées MAS de la SMAI
29 août 2008

[1] INRIA
[2] CEREA, laboratoire commun ENPC - EDF R&D, Université Paris-Est
[3] CNRS & DMA, École normale supérieure
[4] HEC Paris

Projet ANR ATLAS, « From Applications to Theory in Learning and Adaptive Statistics »

INRIA

Modèles de chimie-transport pour la qualité de l'air

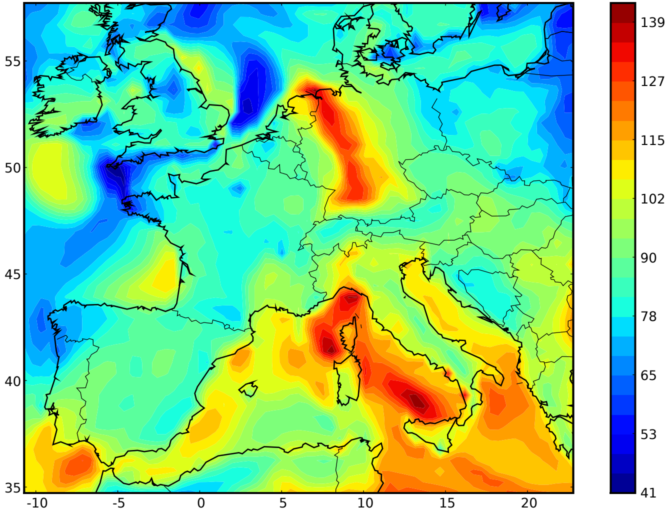


$$\forall i \quad \frac{\partial c_i}{\partial t} = \underbrace{-\mathrm{div}(Vc_i)}_{\text{advection}} + \underbrace{\mathrm{div}\left(\rho K \nabla \frac{c_i}{\rho}\right)}_{\text{diffusion}} + \underbrace{\chi_i(c,t)}_{\text{chimie}} + \underbrace{S_i - P_i}_{\text{sources and pertes}}$$

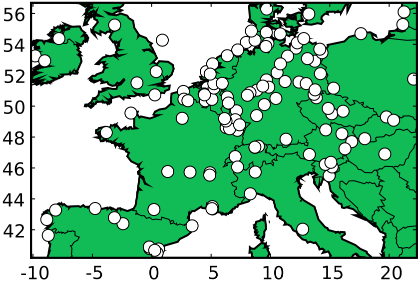
$c_i(t, x, y, z)$: concentration du polluant i

Prévision de l'ozone à l'échelle européenne

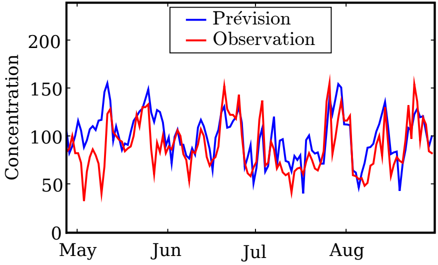
Carte d'ozone ($\mu g\, m^{-3}$, 24 septembre 2006, 16h UT)



Performance des modèles pour les pics d'ozone

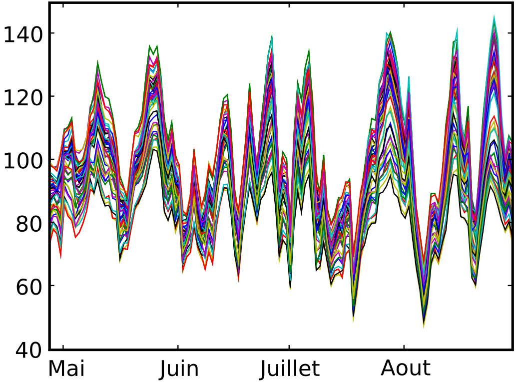


Un réseau d'observation européen



Performance moyenne
(station à Saint Nazaire)

Ensemble de prévisions



Ensemble de prévisions de pics d'ozone (station Saint Nazaire)
48 modèles physiques, soit 48 experts

Agrégation séquentielle

Notations

Prédiction de l'expert m à l'échéance t et à la station s : $x_{m,t}^{s}$

Observation : y_t^s

Prédiction agrégée : $\widehat{y}_t^s = \sum_{m=1}^{N} \alpha_{m,t} x_{m,t}^{s}$

Vecteurs $\mathbf{y}_t$, $\mathbf{x}_{m,t}$ et $\widehat{\mathbf{y}}_t$ des valeurs aux S stations

Notes

1. Combinaison convexe ou non
2. Les poids ne dépendent pas de la station, pour autoriser une agrégation autour des stations d'observation
3. Les poids $\alpha_{m,t}$ peuvent dépendre de $\mathbf{y}_1, \ldots, \mathbf{y}_{t-2}, \mathbf{y}_{t-1}$ et $\mathbf{x}_{m',1}, \ldots, \mathbf{x}_{m',t-1}, \mathbf{x}_{m',t}$ (pour tout m')
4. Aucune hypothèse stochastique

Agrégation séquentielle

Notations

- Poids : $\alpha_{m,t} = \mathcal{F}_{m,t}\left(\mathbf{y}_1, \ldots, \mathbf{y}_{t-2}, \mathbf{y}_{t-1}, \mathbf{x}_{m',1}, \ldots, \mathbf{x}_{m',t-1}, \mathbf{x}_{m',t}\right)$
- Combinaison : $\widehat{y}_t^s = \sum_m \alpha_{m,t} x_{m,t}^s$
- Perte instantanée (à t) : $L(\widehat{\mathbf{y}}_t, \mathbf{y}_t) = \sum_s \left(\widehat{y}_t^s - y_t^s\right)^2$

Principe de l'agrégation séquentielle

- Borner le regret

$$\sum_{t \leq n} L(\widehat{\mathbf{y}}_t, \mathbf{y}_t) - \min_{\beta} \sum_{t \leq n} L\Big(\sum_m \beta_m \mathbf{x}_{m,t}, \mathbf{y}_t\Big)$$

- Si le regret est en $o(n)$:

$$\frac{1}{Sn} \sum_{t \leq n} L(\widehat{\mathbf{y}}_t, \mathbf{y}_t) - \frac{1}{Sn} \min_{\beta} \sum_{t \leq n} L\Big(\sum_m \beta_m \mathbf{x}_{m,t}, \mathbf{y}_t\Big) \longrightarrow_{n \to \infty} 0$$

Agrégation séquentielle

EG : exponentielle des gradients (Vovk, 1990)

- Combinaison EG : $\widehat{y}_t^s = \sum_m \alpha_{m,t} x_{m,t}^s$
- Perte instantanée (à t) : $L(\widehat{\mathbf{y}}_t, \mathbf{y}_t) = \sum_s \left(\widehat{y}_t^s - y_t^s\right)^2$
- Poids initiaux ($t = 1$) : $\forall m \quad \alpha_{m,1} = \frac{1}{N}$
- Poids à $t > 1$:

$$\forall m \quad \alpha_{m,t} = \exp\left(-\eta \sum_{t'<t} \frac{\partial L(\widehat{\mathbf{y}}_{t'}, \mathbf{y}_{t'})}{\partial \alpha_{m,t'}}\right) \times \text{normalisation}$$

- Borne sur le regret

$$\sum_{t\leq n} L(\widehat{\mathbf{y}}_t, \mathbf{y}_t) - \min_{\beta} \sum_{t\leq n} L\Big(\sum_m \beta_m \mathbf{x}_{m,t}, \mathbf{y}_t\Big) \leq \gamma S \sqrt{n \ln N}$$

où $\quad \forall m, t \quad \gamma \geq \left|\frac{\partial L(\widehat{\mathbf{y}}_t, \mathbf{y}_t)}{\partial \alpha_{m,t}}\right|$

Agrégation séquentielle

Régression ridge (Vovk, 2001, et Azoury et Warmuth, 2001)

- Combinaison RR : $\widehat{y}_t^s = \sum_m \alpha_{m,t} x_{m,t}^s$
- Perte instantanée (à t) : $L(\widehat{\mathbf{y}}_t, \mathbf{y}_t) = \sum_s \left(\widehat{y}_t^s - y_t^s\right)^2$
- Calcul des poids :

$$\alpha_{\cdot,t} = \text{argmin}_\beta \left[\sum_{t'<t} \sum_s \left(\sum_m \beta_m x_{m,t'}^s - y_{t'}^s \right)^2 + \lambda \|\beta\|_2^2 \right]$$

- Borne sur le regret

$$\sum_{t \le n} L(\widehat{\mathbf{y}}_t, \mathbf{y}_t) - \min_\beta \sum_{t \le n} L\Big(\sum_m \beta_m \mathbf{x}_{m,t}, \mathbf{y}_t\Big) \lesssim O(\ln n)$$

Cesa-Bianchi & Lugosi, *Prediction, Learning, and Games, 2006*

Aggrégation séquentielle : variantes

Exemple de la régression ridge

$$\alpha_{\cdot,t} = \text{argmin}_\beta \left[\lambda \|\beta\|_2^2 + \sum_{t'<t} \sum_s \left(\sum_m \beta_m x_{m,t'}^s - y_{t'}^s \right)^2 \right]$$

Escompte

Idée : atténuer l'influence du passé lointain

$$\alpha_{\cdot,t} = \text{argmin}_\beta \left[\lambda \|\beta\|_2^2 + \sum_{t'<t} \sum_s (1 + \psi_{t-t'}) \left(\sum_m \beta_m x_{m,t'}^s - y_{t'}^s \right)^2 \right]$$

où ψ_i est une suite décroissante, par exemple $\psi_i = \dfrac{100}{i^2}$

Méta-apprentissage

Application (score : RMSE, $\mu g\, m^{-3}$)

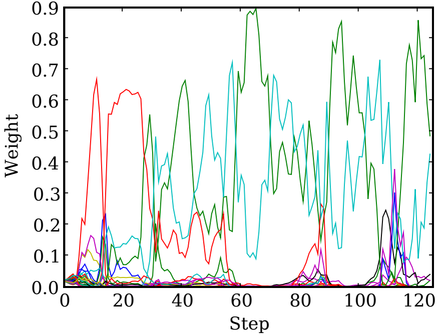
Pics d'ozone

	Réseau 1	Réseau 2	Réseau 3
Meilleur expert	22.43	21.90	23.87
EG	21.47	21.05	24.12
RR escomptée	**19.45**	**18.12**	**20.88**
C_α	19.24	18.16	20.26

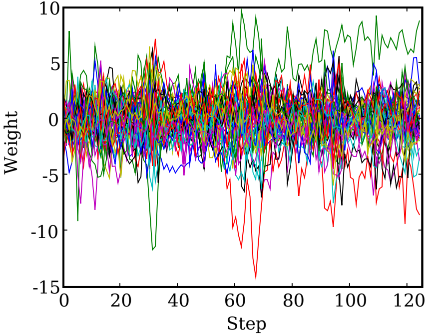
Concentrations horaires d'ozone

	Réseau 1	Réseau 2	Réseau 3
Meilleur expert	26.68	25.98	28.45
EG	24.31	24.67	26.01
RR escomptée	**22.02**	**22.82**	**22.27**
C_α	22.80	23.52	23.19

Évolution temporelle des poids



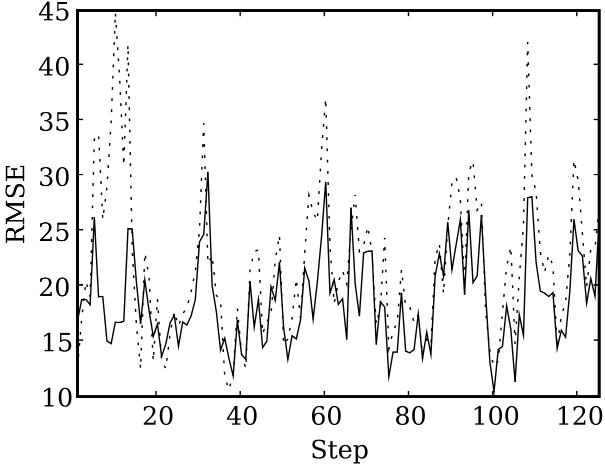
EG



RR escomptée

Application

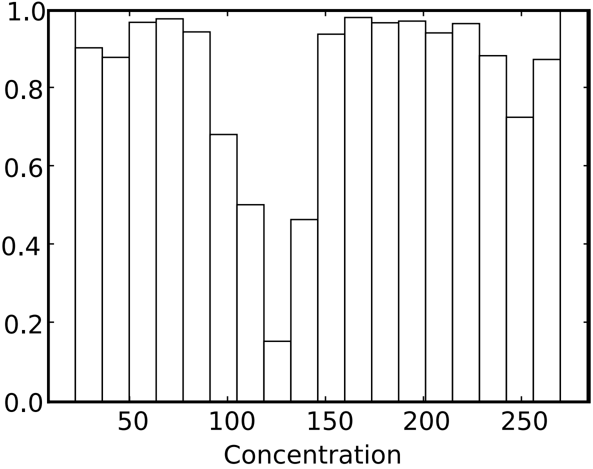
Robustesse



RMSE ($\mu g\, m^{-3}$) fonction du temps

Application

Comportement pour les événements extrêmes



Fréquence avec laquelle RR escomptée prévoit mieux que le meilleur modèle

Un mot sur l'assimilation de données « classique »

Exemple : le filtre de Kalman

- Vecteur d'état à l'échéance t : $\mathbf{e}_t^f$
- Correction du filtre (analyse) : $\mathbf{e}_t^a = \mathbf{e}_t^f + K_t(\mathbf{o}_t - H_t\mathbf{e}_t^f)$
- Prévision $\mathbf{e}_{t+1}^f = \mathcal{M}_{t\to t+1}(\mathbf{e}_t^a)$, et propagation *approchée* de la matrice de covariance d'erreur

Différences avec l'agrégation séquentielle

	Assimilation classique	Apprentissage
Paramètres	grand nombre	très peu
Théorie & application	fortes limitations	compatibles
Spatialisation	naturelle	hors cadre

Conclusion

Stratégie

- Ensemble de prédictions de modèles physiques ou experts
- Agrégation séquentielle
- Performances de la meilleure combinaison linéaire d'experts

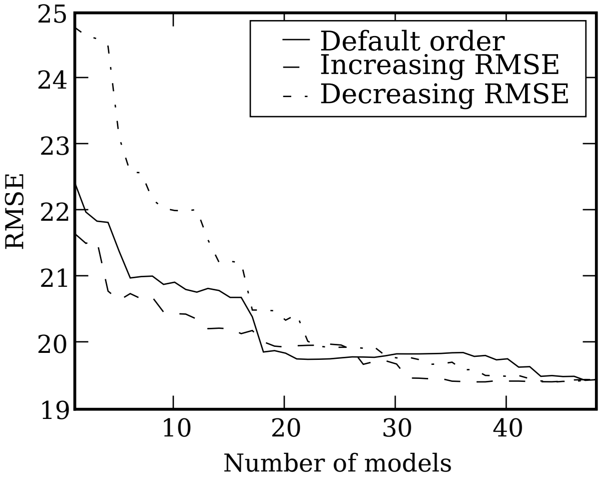
Positionnement en terme de modélisation

- Statistiques pures / agrégation séquentielle / assimilation classique

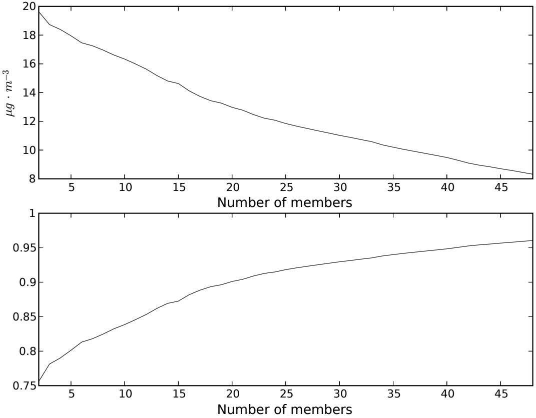
La suite...

- Seuils sur les poids, LASSO
- Sélection des experts (modèles coûteux)
- Spatialisation

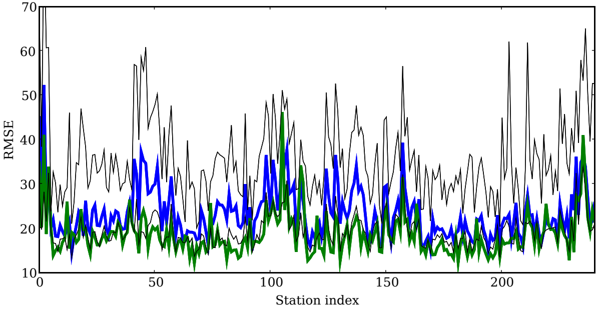
Number of Models (*RR* with Discount)



Number of Models (*ELS* a posteriori)

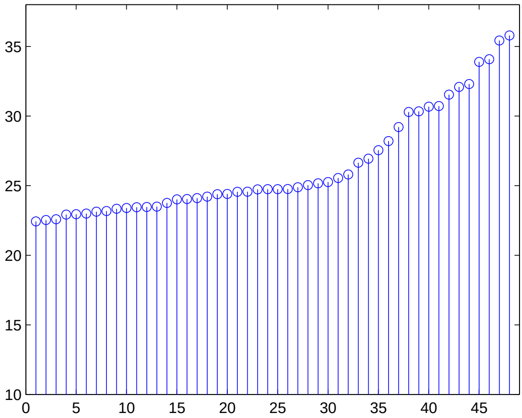


Per Station



INRIA

Models RMSEs



Ensemble Spread