



Université Paris-Sud, INRIA

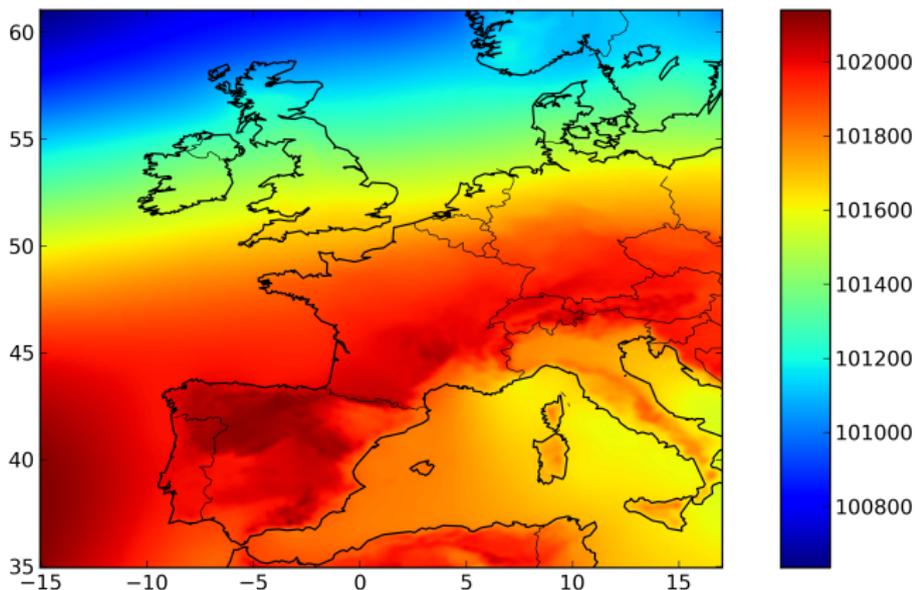
Prévision séquentielle par agrégation d'ensemble : application à des prévisions météorologiques assorties d'incertitudes

26 novembre 2015

thèse de doctorat présentée par Paul Baudin

Jury composé de

Pr Élisabeth Gassiat	Université Paris-Sud	présidente
Pr Jean-Marc Bardet	Université Panthéon-Sorbonne	rapporteur
Pr Liliane Bel	AgroParisTech	examinatrice
D ^r Laurent Descamps	Météo-France	examineur
D ^r Olivier Mestre	Météo-France	examineur
D ^r Gilles Stoltz	CNRS	co-directeur de thèse
D ^r Vivien Mallet	INRIA	co-directeur de thèse



Objectif : prévoir une telle carte de pression (en Pa) en combinant plusieurs simulations.

Domaine

La théorie des suites arbitraires, un domaine connexe à :

- la théorie des suites chronologiques
- l'apprentissage statistique (*machine learning*)
- la théorie des jeux répétés

N. CESA-BIANCHI et G. LUGOSI. *Prediction, learning, and games*. Cambridge University Press, 2006

Prévision tour après tour

1. Concevoir des stratégies de prévision automatiques à partir de prévisions existantes
2. Calculer les garanties mathématiques, très générales
3. Les appliquer à la prévision météorologique
4. Décliner ces idées à un cadre de prévision probabiliste

- 1 Introduction
- 2 Formalisation
- 3 Arbre de régression déterministe
- 4 Résultats empiriques
 - Jeux de données
 - Résultats
- 5 Prévision des incertitudes

Cadre

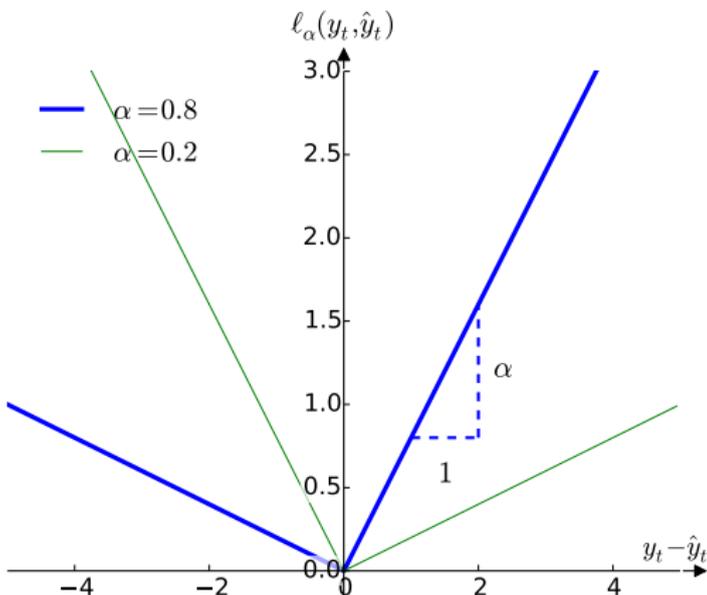
- Prévoir, avec un algorithme \mathcal{S} **automatique** et **robuste**, les $y_t \in [a, b]$, valeurs observées issues d'une **suite arbitraire** $(y_t)_t$, i.e. sans hypothèses stochastiques a priori
- Préviation proposée : $\hat{y}_t \in [a, b]$
- La **fonction de perte** : $\ell : [a, b] \times [a, b] \rightarrow \mathbb{R}$, convexe par rapport à \hat{y}_t , mesure la **proximité** de y_t et \hat{y}_t

À chaque échéance $t = 1, 2, \dots$

- 1 De l'information auxiliaire est fournie à \mathcal{S}
- 2 \mathcal{S} propose une prévision \hat{y}_t à partir de l'**information auxiliaire** et du passé
- 3 L'observation y_t est révélée et \mathcal{S} subit la **perte instantanée** : $\ell(\hat{y}_t, y_t)$
- 4 \mathcal{S} est mis à jour

Fonctions de perte

- Quadratique $\ell(\hat{y}_t, y_t) = (\hat{y}_t - y_t)^2$
- En valeur absolue $\ell(\hat{y}_t, y_t) = |\hat{y}_t - y_t|$
- *Pinball*, pour $\alpha \in]0, 1[$, $\ell_\alpha(\hat{y}_t, y_t) = \left(\alpha - \mathbb{1}_{\{y_t \geq \hat{y}_t\}}\right)(y_t - \hat{y}_t)$
- Le *continuous ranked probability score* (CRPS)



Prévisions élémentaires

Les **prévisions élémentaires** ou **experts** issus de modèles sont

$$x_{1,t}, \dots, x_{M,t} \in [a, b].$$

Algorithme : combiner les experts

Combinaison convexe ou linéaire :

$$\hat{y}_t = \sum_{m=1}^M p_{m,t} x_{m,t}$$

Notations

- Perte cumulée de l'expert m :

$$L_{m,T} = \sum_{t=1}^T \ell(x_{m,t}, y_t)$$

- **Perte cumulée** de l'algorithme d'agrégation :

$$\hat{L}_T = \sum_{t=1}^T \ell(\hat{y}_t, y_t)$$

Regret

Le regret fait intervenir la perte cumulée du meilleur des M experts (l'**oracle**) :

$$\text{Regret}_T = \widehat{L}_T - \min_{m=1,\dots,M} L_{m,T}$$

i.e.

$$\widehat{L}_T = \underbrace{\min_{m=1,\dots,M} L_{m,T}}_{\text{erreur d'approximation}} + \underbrace{\text{Regret}_T}_{\text{erreur d'estimation séquentielle}}$$

Pondération par poids exponentiels, de Littlestone & Warmuth et Vovk

$$p_{m,t+1} = \frac{e^{-\eta \ell(x_{m,t}, y_t)} p_{m,t}}{\sum_{k=1}^M e^{-\eta L_{k,t}}}$$

N. LITTLESTONE et M. K. WARMUTH. « The Weighted Majority Algorithm ». In : *Inf. Comput.* 108.2 (fév. 1994), p. 212–261

V. G. VOVK. « Aggregating Strategies. » In : *Proceedings of COLT.* 1990, p. 371–386

Théorème

Si $\ell : [a, b]^2 \rightarrow [0, B]$ est convexe,

$$\sup \left\{ \underbrace{\sum_{t=1}^T \ell(\hat{y}_t, y_t)}_{\hat{L}_T} - \min_{m=1, \dots, M} \underbrace{\sum_{t=1}^T \ell(x_{m,t}, y_t)}_{L_{m,T}} \right\} \leq \frac{\ln M}{\eta} + \eta \frac{B^2}{8} T \leq B \sqrt{\frac{T}{2} \ln M}$$

pour le choix de $\eta = (1/B) \sqrt{(8 \ln M)/T}$.

Adaptation de η par rapport à T

1. Relancer \mathcal{S} avec $\eta^{(r)} = (1/B) \sqrt{(8 \ln M)/2^r}$ à chaque échéance 2^r pour $r \in \mathbb{N}$, le **doubling trick**
2. À chaque échéance, $\eta_t = (1/B) \sqrt{(8 \ln M)/t}$
3. Calibrer η à la volée sur une **grille**

Passage à des combinaison convexes

On souhaite minimiser le regret :

$$\underbrace{\sum_{t=1}^T \ell(\hat{y}_t, y_t)}_{\hat{L}_T} - \min_{p \in \mathcal{P}} \sum_{t=1}^T \ell \left(\sum_{m=1}^M p_m x_{m,t}, y_t \right)$$

L'**astuce du gradient** (Kivinen & Warmuth) consiste à remplacer dans l'algorithme des poids exponentiels les pertes $\ell(x_{m,t}, y_t)$ par les **pseudo-pertes** :

$$\nabla \ell \left(\sum_{k=1}^M p_{k,t} x_{k,t}, y_t \right) \cdot x_{m,t}$$

pour obtenir l'algorithme *Exponentiated Gradient* (EG).

J. KIVINEN et M. K. WARMUTH. « Exponentiated Gradient Versus Gradient Descent for Linear Predictors ». In : *Information and Computation* 132.1 (1997), p. 1-63

Arbre de régression déterministe

Collaboration avec Pierre Gaillard : article « *Deterministic Regression Tree* ».

Regret

$$\text{Regret}_T = \sum_{t=d+1}^T \ell(\hat{y}_t, y_t) - \inf_{f \in \mathcal{L}_L^d} \sum_{t=d+1}^T \ell(f(y_{t-d}^{t-1}), y_t)$$

où

$$\mathcal{L}_L^d = \{f : [0, 1]^d \rightarrow [0, 1], f \text{ est } L\text{-lipschitzienne}\}$$

\mathcal{L}_L^d est non-paramétrique et plus massive que l'ensemble des combinaisons convexes d'experts.

Notation : $y_{t-d}^{t-1} = (y_{t-d}, \dots, y_{t-1})$

Survol de *Nested EG*

Nested EG est construit par strates successives :

Couche inférieure : *EG* sur des pseudo-experts valant 0 ou 1

Couche intermédiaire : *Nested EG* à d connu

Couche supérieure : Adaptation en d

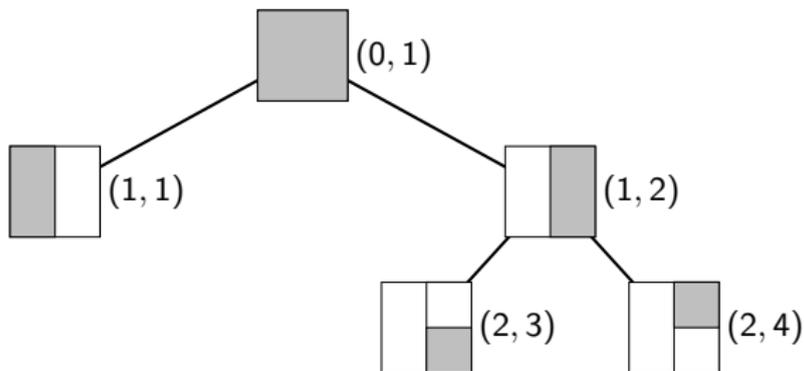


FIGURE : Un arbre de régression sur $[0, 1]^2$

EG sur des pseudo-experts

EG avec deux pseudo-experts constants $x_{1,t} = 0$ et $x_{2,t} = 1$ fournit $\hat{y}_t \in [0, 1]$.

Alors, dans toute cellule suffisamment petite ou clairsemée,

$$\begin{aligned} \sum_t \ell(\hat{y}_t, y_t) &\leq \inf_{y \in [0,1]} \sum_t \ell(y, y_t) + \mathcal{O}(\sqrt{T}) \\ &\leq \inf_{f \in \mathcal{L}_L^d} \sum_t \ell(f(y_{t-d}^{t-1}), y_t) + \mathcal{O}(\sqrt{T}) + \dots \end{aligned}$$

Nested EG

- Un EG sur pseudo-expert par feuille de l'arbre (région de $[0, 1]^d$)
- Trop d'observations vu le diamètre d'une feuille : division de la feuille et réinitialisation des EG (cf. CART).

L. BREIMAN, J. FRIEDMAN, C. J. STONE et R. A. OLSHEN. *Classification and Regression Trees*. Sous la dir. de NO. Belmont, CA : Wadsworth International Group, 1984

Lemme (couche intermédiaire)

Hypothèses : ℓ convexe et lipschitzienne.

Paramètre : d fixé.

Pour tout $y_1, \dots, y_T \in [0, 1]$,

$$\sum_{t=d+1}^T \ell(\hat{y}_t, y_t) - \inf_{f \in \mathcal{L}_L^d} \sum_{t=d+1}^T \ell(f(y_{t-d}^{t-1}), y_t) \leq \mathcal{O}(LT^{\frac{d+1}{d+2}}) = o(T)$$

Théorème (couche supérieure)

Hypothèses : ℓ convexe et lipschitzienne.

Paramètre : aucun.

Algorithme : une agrégation dénombrable (sleeping) d'algorithmes du lemme précédent.

Pour tout $y_1, \dots, y_T \in [0, 1]$,

$$\forall d, \quad \sum_{t=d+1}^T \ell(\widehat{y}_t, y_t) - \inf_{f \in \mathcal{L}_L^d} \sum_{t=d+1}^T \ell(f(y_{t-d}^{t-1}), y_t) \leq \mathcal{O}\left(T^{\frac{d+1}{d+2}}\right)$$

Conséquence

Pour tout d , tout L ,

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \ell(\widehat{y}_t, y_t) \leq \limsup_{T \rightarrow \infty} \left(\inf_{f \in \mathcal{L}_L^d} \frac{1}{T} \sum_{t=d+1}^T \ell(f(y_{t-d}^{t-1}), y_t) \right)$$

Cadre stochastique

$Y_t \in [0, 1]$ issue d'un processus borné stationnaire et ergodique

$$\frac{\widehat{L}_T}{T} = \frac{1}{T} \sum_{t=1}^T \ell(\widehat{Y}_t, Y_t)$$

Borne inférieure

Soit $\ell : [0, 1]^2 \rightarrow [0, 1]$, continue et bornée.

Pour tout \mathcal{S} , on a (Algoet)

$$\liminf_{T \rightarrow \infty} \frac{\widehat{L}_T}{T} \geq L^* \quad \text{p.s.} \quad \text{avec} \quad L^* = \mathbb{E} \left[\inf_{f \in \mathcal{B}^\infty} \mathbb{E} \left[\ell(f(Y_{-\infty}^{-1}), Y_0) \mid Y_{-\infty}^{-1} \right] \right]$$

P. H. ALGOET. « The strong law of large numbers for sequential decisions under uncertainty ». In : *IEEE Transactions on Information Theory* 40.3 (1994), p. 609–633

Théorème (optimalité de l'arbre de régression déterministe)

Si, pour tout d , tout L ,

$$\limsup_{T \rightarrow \infty} \frac{\widehat{L}_T}{T} \leq \limsup_{T \rightarrow \infty} \left(\inf_{f \in \mathcal{L}_L^d} \frac{1}{T} \sum_{t=d+1}^T \ell(f(Y_{t-d}^{t-1}), Y_t) \right) \quad p.s.$$

alors

$$\limsup_{T \rightarrow \infty} \frac{\widehat{L}_T}{T} \leq L^* \quad p.s.$$

Idée de la preuve

$$\begin{aligned}
\limsup_{T \rightarrow \infty} \inf_{f \in \mathcal{L}_L^d} & \left(\frac{1}{T} \sum_{t=d+1}^T \ell(f(Y_{t-d}^{t-1}), Y_t) \right) \\
& \leq \inf_{f \in \mathcal{L}_L^d} \limsup_{T \rightarrow \infty} \left(\frac{1}{T} \sum_{t=d+1}^T \ell(f(Y_{t-d}^{t-1}), Y_t) \right) \\
& \leq \inf_{f \in \mathcal{L}_L^d} \mathbb{E} [\ell(f(Y_{-d}^{-1}), Y_0)] \quad (\text{Breiman}) \\
& \leq \inf_{f \in \mathcal{B}^d} \mathbb{E} [\ell(f(Y_{-d}^{-1}), Y_0)] \quad (L \rightarrow \infty) \\
& \leq \underbrace{\mathbb{E} \left[\inf_{f \in \mathcal{B}^d} \mathbb{E} [\ell(f(Y_{-d}^{-1}), Y_0) \mid Y_{-d}^{-1}] \right]}_{\triangleq Z_d}
\end{aligned}$$

Et Z_d est une sur-martingale :

$$\lim_{d \rightarrow \infty} \mathbb{E}[Z_d] = \mathbb{E} \left[\inf_{f \in \mathcal{B}^\infty} \mathbb{E} [\ell(f(Y_{-\infty}^{-1}), Y_0) \mid Y_{-\infty}^{-1}] \right] = L^*$$

Comparaison à la littérature

Résultats similaires via des méthodes de plus proches voisins pour
la perte carrée (Györfi, Lugosi, Biau & Patra)
la perte *pinball* (Biau & Patra)

Nos résultats sont plus généraux : ℓ générale (bornée, convexe et lipschitzienne).

Notre preuve en deux temps sépare les analyses déterministes et stochastiques (cf. Gerchinovitz).

L. GYÖRFI, G. LUGOSI et R. T. FARGAS. *Strategies for Sequential Prediction of Stationary Time Series*. 2001

G. BIAU et B. PATRA. « Sequential Quantile Prediction of Time Series. » In : *IEEE Transactions on Information Theory* 57.3 (2011), p. 1664–1674

S. GERCHINOVITZ. « Prediction of individual sequences and prediction in the statistical framework : some links around sparse regression and aggregation techniques. » *Thèse de doct. Université Paris-Sud 11*, 2011

Perspective

Par hypothèse, $y_t, \hat{y}_t \in [a, b] \subset \mathbb{R}$. Le cas non-borné rendrait l'algorithme plus général encore.

Résultats empiriques

Jeu de données fourni par Laurent Descamps (Météo-France).

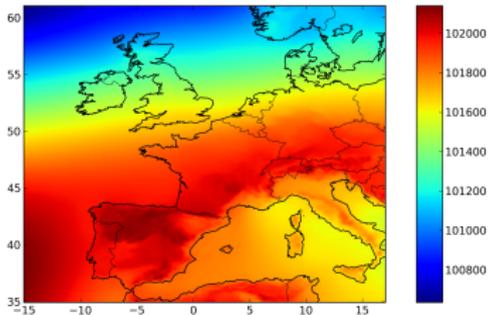
Grandeurs physiques étudiées

- La pression réduite au niveau de la mer
- La norme de la vitesse du vent à 10 mètres au-dessus du sol

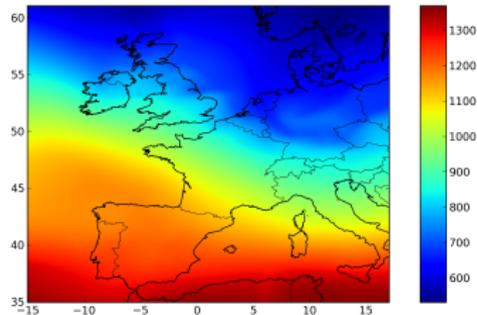
Plan d'étude (Devaine, Gaillard, Goude & Stoltz)

- 1 Présentation du jeu de données et construction des experts
- 2 Calcul des oracles et sélection de la prévision de référence
- 3 Agrégation avec paramètres fixés (les meilleurs paramètres rétrospectifs)
- 4 Agrégation avec des paramètres séquentiellement adaptatifs

M. DEVAINE, P. GAILLARD, Y. GOUDE et G. STOLTZ. « Forecasting electricity consumption by aggregating specialized experts - A review of the sequential aggregation of specialized experts, with an application to Slovakian and French country-wide one-day-ahead (half-)hourly predictions ». In : *Machine Learning* 90.2 (2013), p. 231–260, p. 12



Pression (Pa) : moyenne temporelle
de l'analyse.



Pression (Pa) : écart type
empirique, phénomène peu variable.

Cadre spatial

- Domaine $[35^\circ N, 61^\circ N] \times [15^\circ O, 17^\circ E]$, avec une résolution de $0,10^\circ$
- Un algorithme pour chaque cellule (i, j) du domaine \mathcal{N} avec $N = |\mathcal{N}|$

Cadre temporel

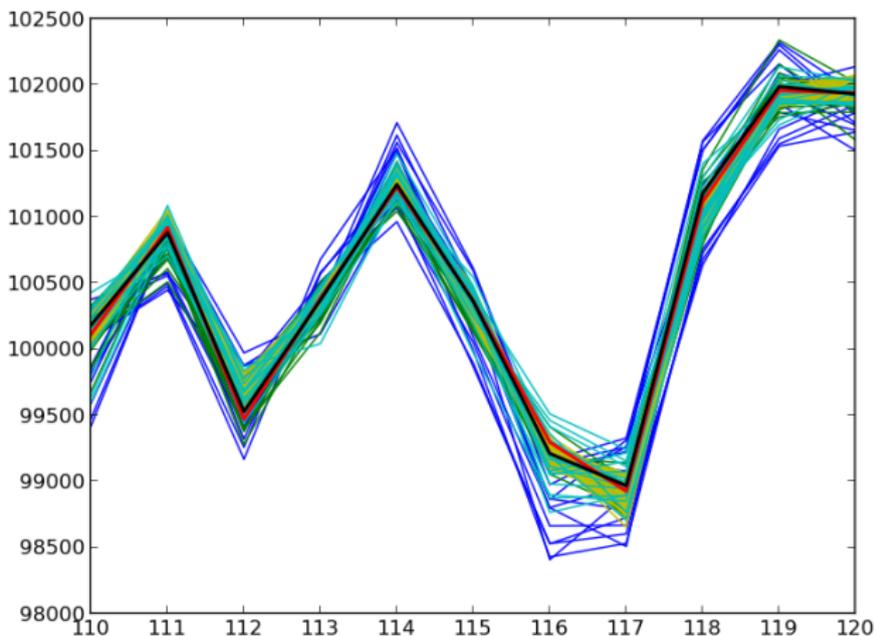
- 366 jours du 1^{er} octobre 2011 au 1^{er} octobre 2012 exclu
- L'horizon des prévisions est de 6 ou 18 heures, en partant de minuit

Prévisions élémentaires

- 150 membres (simulations) issus de divers centres météorologiques, fourni par « *TIGGE* »
- Conditions initiales et modèles d'évolution différents entre les membres
- Pour tout (i, j) , tout t , un tri préliminaire est nécessaire

Origine	Nombre de membres
Chine	14
Canada	20
Europe (dont déterministe)	51
Corée du Sud	23
États-Unis	20
Grande-Bretagne	23
Météo-France (déterministe)	1
Total (M)	152

- **Prévision déterministe** : prévision de référence simulée avec une résolution fine et un modèle standard, opérationnel
- **Analyse** y_t : estimation optimale d'une grandeur physique, i.e. le meilleur compromis entre les observations bruitées et une simulation de référence



Pression (Pa) à t et (i, j) fixés. La prévision déterministe, l'analyse, les membres de l'ensemble.

Score et perte quadratique

Performance évaluée par la *Root Mean Square Error (RMSE)* :

$$RMSE = \sqrt{\frac{1}{TN} \sum_{(i,j) \in \mathcal{N}} \sum_{t=1}^T (\hat{y}_{t,(i,j)} - y_{t,(i,j)})^2}$$

Amélioration d'une prévision A par rapport à une *prévision de référence R* :

$$\Delta\%(A, R) = \frac{RMSE_R - RMSE_A}{RMSE_R}$$

Remarques

- Un regret faible implique un $\Delta\%(A, R)$ grand
- R est le déterministe Météo-France, pour lequel un an de R&D implique quelques % d'amélioration

Paramètres et initialisation

Paramètres

- $\lambda \in \mathbb{R}_+$, le paramètre de régularisation
- $\gamma \in \mathbb{R}_+$, le paramètre d'escompte
- $\mathbf{u}_1 = (1, 0, \dots, 0)^T$, un vecteur de poids initial

Algorithme de régression ridge escompté

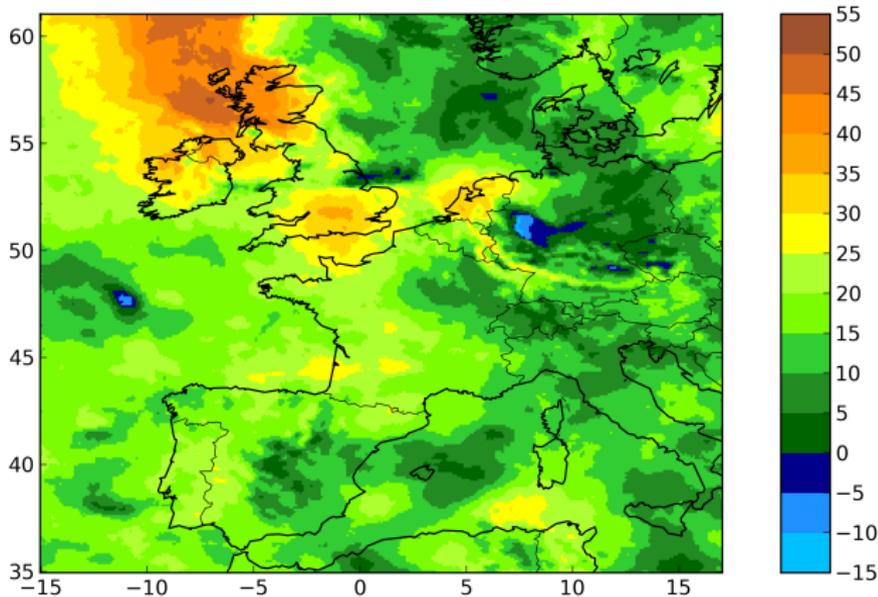
Le calcul des poids linéaires \mathbf{u}_t est déterminé en résolvant :

$$\begin{aligned} \mathbf{u}_t &= (\mathbf{u}_{1,t}, \dots, \mathbf{u}_{M,t})^T \\ &= \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^M} \left\{ \lambda \|\mathbf{u} - \mathbf{u}_1\|_2^2 + \sum_{s=1}^{t-1} \left(1 + \frac{\gamma}{(t-s)^2} \right) \left(\sum_{m=1}^M u_m x_{m,s} - y_s \right)^2 \right\} \end{aligned}$$

Pression : résultats globaux.

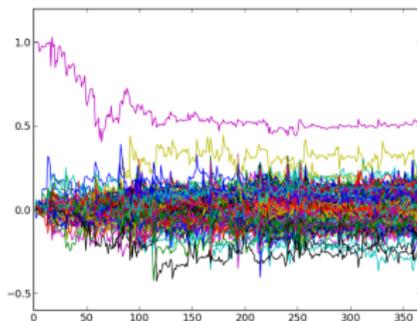
Type de prévision	RMSE (Pa)	$\Delta\%$ (A, R) (%)
Moyenne d'ensemble	52,4	-73,9
Déterministe	30,1	0,0
Oracle convexe	24,3	19,4
Oracle linéaire	18,0	40,2
Agrégation ridge (paramètres optimaux rétrospectifs)	24,9	18,1
Agrégation ridge (paramètres locaux optimaux rétrospectifs)	24,4	18,9
Agrégation ridge (adaptation locale en ligne sur une grille de paramètres)	24,8	17,7

- Conclusion : amélioration forte par rapport à la référence
- Perspective : passage en opérationnel

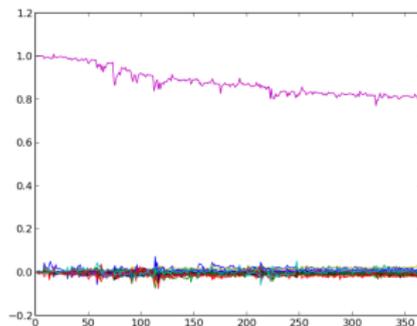


Pression (%) : résultats locaux en

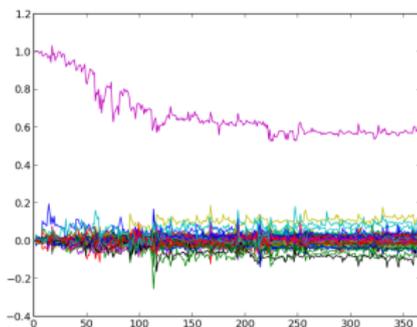
$$\Delta_{\%}(A, R) = \frac{RMSE_R - RMSE_A}{RMSE_R}$$



$$\lambda = 6 \times 10^3$$



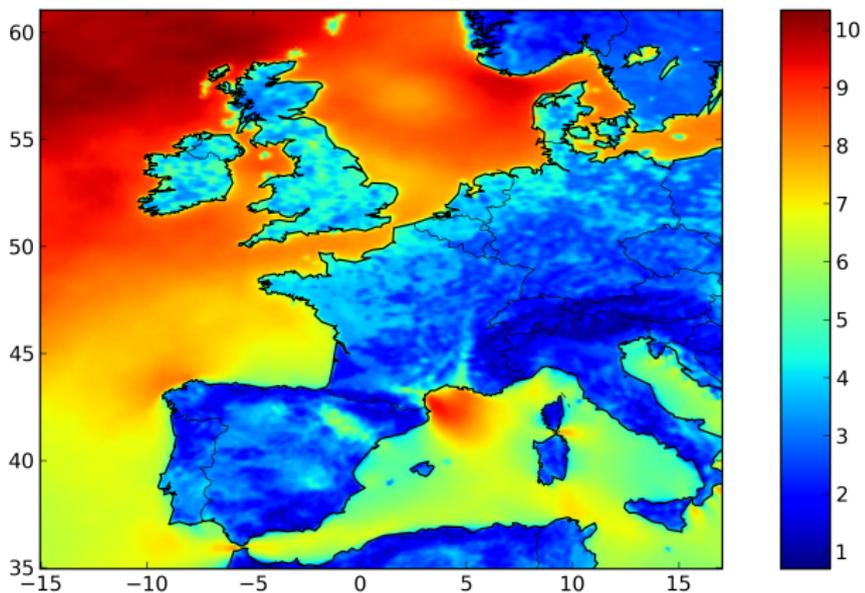
$$\lambda = 6 \times 10^5$$



$$\lambda = 6 \times 10^4$$

Pression, sensibilité des poids à la régularisation λ (Pa).

Remarque : $\sum_{m=1}^M p_{m,t} \approx 1$.

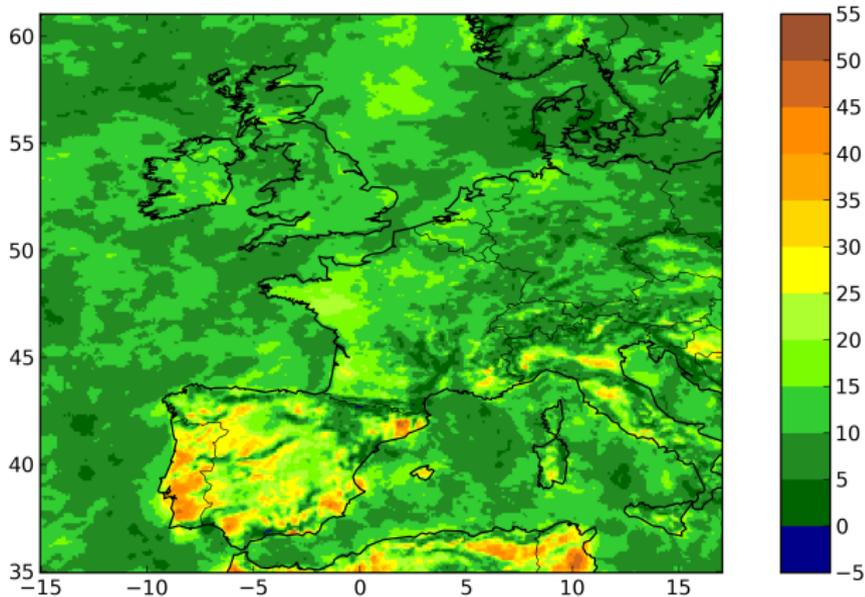


Jeu de données de vent (m s^{-1}) : moyenne temporelle de l'analyse.

Vent : résultats globaux.

Type de prévision	<i>RMSE</i> (m s^{-1})	Différence relative (%)
Moyenne d'ensemble	2,32	-47,0
Déterministe	1,58	0,0
Oracle convexe	1,43	9,4
Oracle linéaire	1,38	12,5
Agrégation ridge (paramètres optimaux rétrospectifs)	1,43	9,0
Agrégation ridge (paramètres locaux optimaux rétrospectifs)	1,47	6,5

- Conclusion : amélioration forte par rapport à la référence
- Perspective : passage en opérationnel



Vent (%) : résultats locaux en

$$\Delta_{\%}(A, R) = \frac{RMSE_R - RMSE_A}{RMSE_R}$$

Prévision des incertitudes

Collaboration avec Jean Thorey.

Objectif : prévoir des fonctions de répartition.

Ensemble et observation

Soit $x \in \mathbb{R}$, la **fonction échelon de Heaviside** H_x est définie comme $\mathbb{1}_{[x, +\infty[}$.
Agrégation convexe \mathbf{p}_s des prévisions élémentaires $H_{x_{m,s}}$:

$$\sum_{m=1}^M p_{m,s} H_{x_{m,s}}$$

Continuous Ranked Probability Score

$$CRPS(\mathbf{p}_s) = \int_{-\infty}^{\infty} \left(H_{y_s}(z) - \sum_{m=1}^M p_{m,s} H_{x_{m,s}}(z) \right)^2 dz$$

Cas binaire.

Score de Brier

- $y_s \in \{0, 1\}$, observation binaire
- $p_s \in [0, 1]$, probabilité d'occurrence de 1

$$\frac{1}{t} \sum_{s=1}^t (p_s - y_s)^2$$

Généralisation à K classes.

Ranked probability score

Soit $r_{k,s}^o$ la probabilité cumulée d'occurrence de $k \in \{1, \dots, K\}$ et $r_{k,s}^p$, sa prévision.

$$\frac{1}{t} \sum_{s=1}^t \frac{1}{K-1} \sum_{k=1}^K (r_{k,s}^p - r_{k,s}^o)^2$$

Continuous Ranked Probability Score

$$CRPS(\mathbf{p}_s) = \int_{-\infty}^{\infty} \left(H_{y_s}(z) - \sum_{m=1}^M p_{m,s} H_{x_{m,s}}(z) \right)^2 dz$$

est convexe. C'est une fonction de perte standard en météorologie.

Algorithmes convexes

- *Exponentiated Gradient*
- **ML-poly-grad** (Gaillard, Stoltz & van Erven)

P. GAILLARD, G. STOLTZ et T. VAN ERVEN. « A Second-order Bound with Excess Losses ». In : *Proceedings of COLT*. 2014

Protagonistes

- Ensemble : 152 échelons issues des membres de l'ensemble
- Prévision de référence : échelon dont le seuil est la prévision déterministe de Météo-France, **qui est très précise**
- Autres benchmarks : fonctions de répartition uniformes associées aux prévisions d'*ECMWF* ou de *TIGGE*

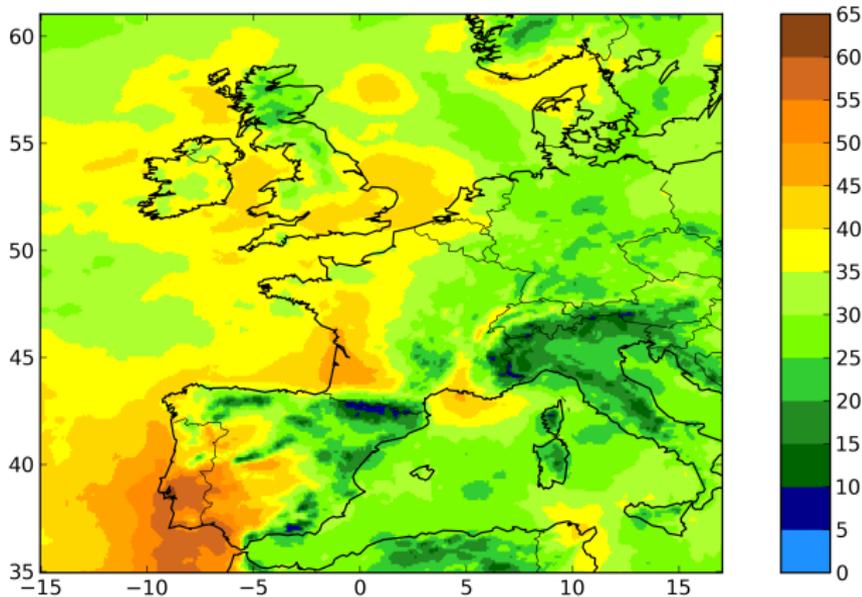
La différence relative est adaptée ici au CRPS :

$$\Delta_{\%}(A, R) = \frac{CRPS_R - CRPS_A}{CRPS_R}$$

Pression : résultats globaux.

Type de prévision	CRPS (Pa)	différence relative (%)
Ensemble <i>TIGGE</i>	28,0	-19,7
Ensemble <i>ECMWF</i>	32,8	-40,1
Échelon déterministe	23,4	0,0
Poids exponentiels (avec paramètre optimal a posteriori)	15,6	33,4
ML-poly-grad	19,3	17,6

- Conclusion : résultats novateurs. Quantiles
- Perspective : l'incertitude sur les observations est ignorée. Alternatives : fonctions rampes ou fonctions de répartition de loi normale sous-jacente



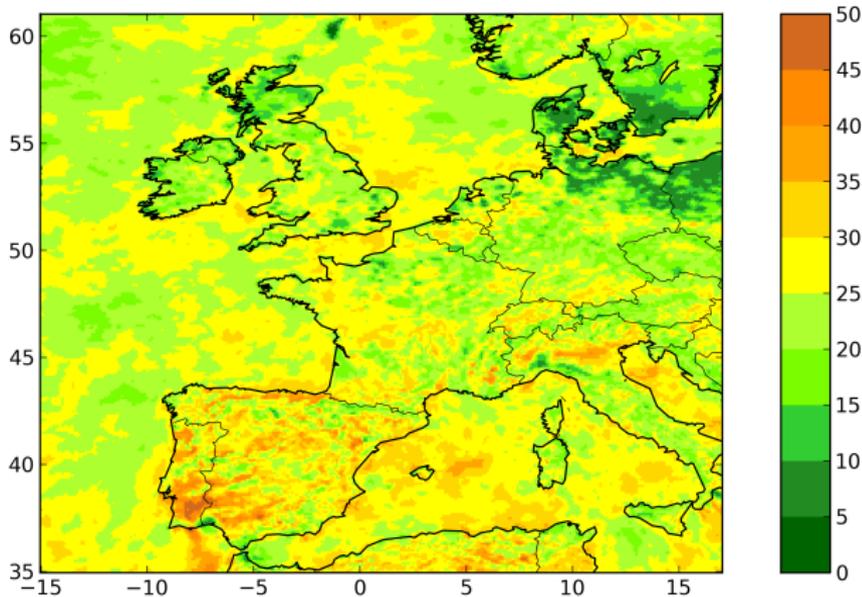
Pression (%) : résultats locaux en

$$\Delta_{\%}(A, R) = \frac{CRPS_R - CRPS_A}{CRPS_R}$$

Vent : résultats globaux.

Type de prévision	CRPS (m s^{-1})	différence relative (%)
Ensemble <i>TIGGE</i>	1,15	-36,6
Ensemble <i>ECMWF</i>	1,41	-66,9
Échelon déterministe	0,84	0,0
Poids exponentiels (avec paramètre optimal a posteriori)	0,64	24,1
ML-poly-grad	0,73	13,1

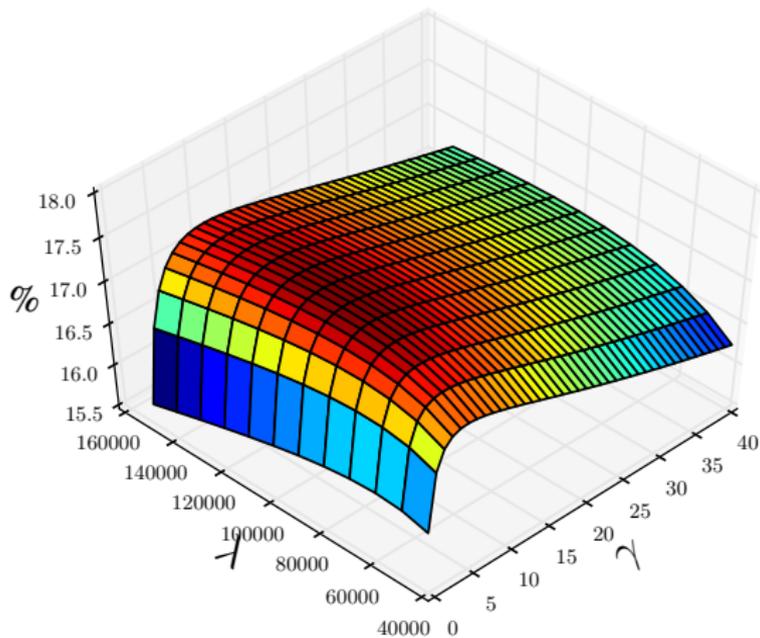
- Conclusion : résultats novateurs. Quantiles
- Perspective : l'incertitude sur les observations est ignorée. Alternatives : fonctions rampes ou fonctions de répartition de loi normale sous-jacente



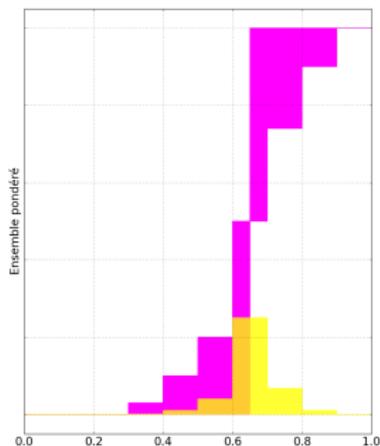
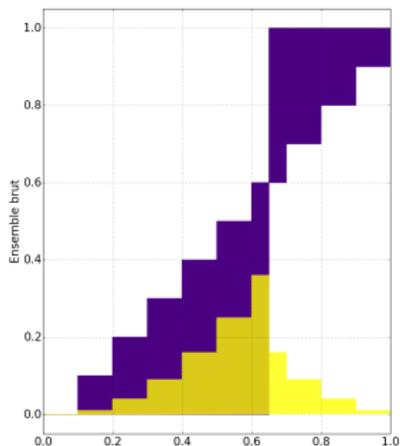
Vent (%) : résultats locaux en

$$\Delta_{\%}(A, R) = \frac{CRPS_R - CRPS_A}{CRPS_R}$$

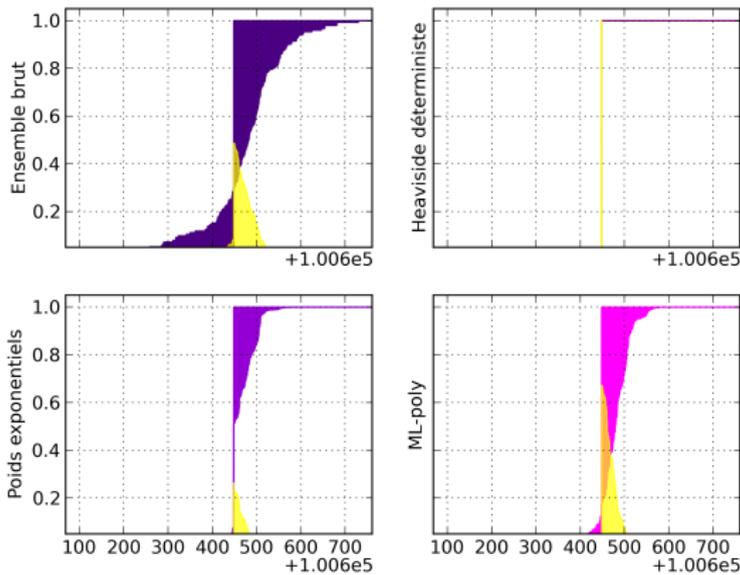
Merci



Pression (%) : sensibilité de $\Delta\%$ (A, R) aux paramètres.



$H_{y_s}(z) = \sum_{m=1}^M p_{m,s} H_{x_{m,s}}(z)$, avec des poids uniformes (gauche) ou non (droite) et leurs CRPS (en jaune).



Pression. À (i, j) et t fixés, prévisions probabilistes et leurs CRPS (en jaune).

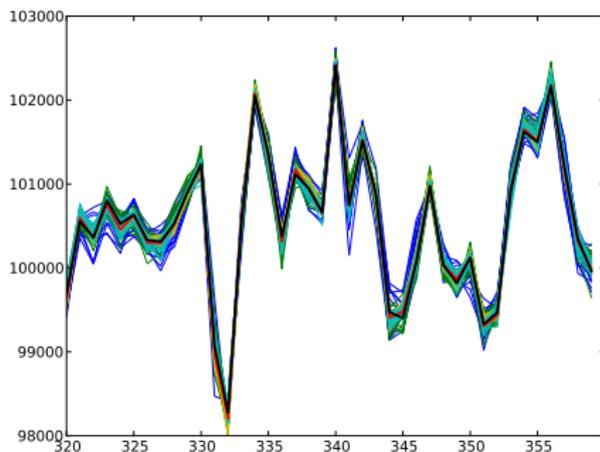


FIGURE : Données de pression réduite au niveau de la mer dans une cellule typique (de coordonnées $(-13^\circ, 37^\circ)$), entre les échéances 320 (16 août 2012) et 360 (25 septembre 2012) de la période temporelle considérée. Un échantillon de 100 membres de l'ensemble parmi les 150 est représenté. En rouge, la prévision déterministe, $x_{t,(i,j)}^{\text{det}}$; en noir, l'analyse, $y_{t,(i,j)}$; le reste des couleurs est dédié aux membres de l'ensemble, $x_{t,(i,j)}^m$.

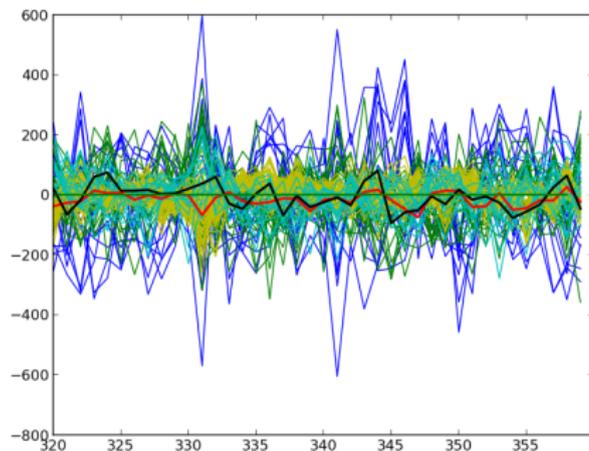


FIGURE : Représentation graphique supplémentaire des données de 3, ici centrées autour de la moyenne d'ensemble, $\bar{x}_{t,(i,j)}$, représentée par la ligne horizontale en vert foncé et d'ordonnée 0, afin d'assurer une plus grande visibilité de la variabilité. Un échantillon de 100 membres de l'ensemble parmi les 150 est représenté. En rouge, la prévision déterministe centrée, $x_{t,(i,j)}^{\text{det}} - \bar{x}_{t,(i,j)}$; en noir, l'analyse centrée, $y_{t,(i,j)} - \bar{x}_{t,(i,j)}$; le reste des couleurs est dédié aux membres de l'ensemble, $x_{t,(i,j)}^m - \bar{x}_{t,(i,j)}$.

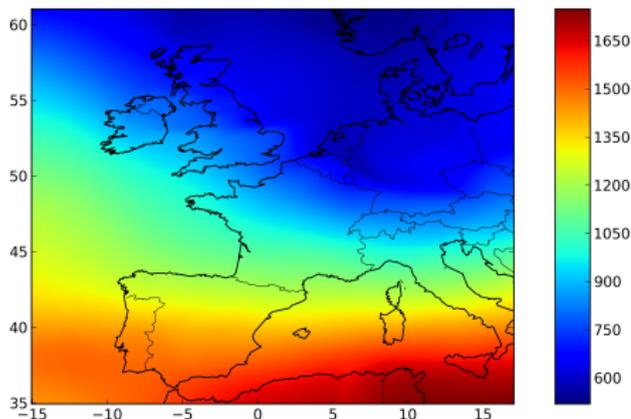


FIGURE : Écart type de la moyenne temporelle de l'ensemble. Dans chaque cellule de coordonnées (i, j) est représentée la grandeur $\frac{1}{T} \sum_{t=1}^T \sqrt{\frac{1}{M} \sum_{m=1}^M (x_{t,(i,j)}^m - \bar{x}_{t,(i,j)})^2}$. L'étendue sur toute la zone considérée de cet écart type est de 1200 Pa. On constate que les régions dans lesquelles se concentre la dispersion de l'ensemble (sud de la zone étudiée, couleurs chaudes) se trouvent être les régions dans lesquelles la pression moyenne est relativement basse (couleurs froides).

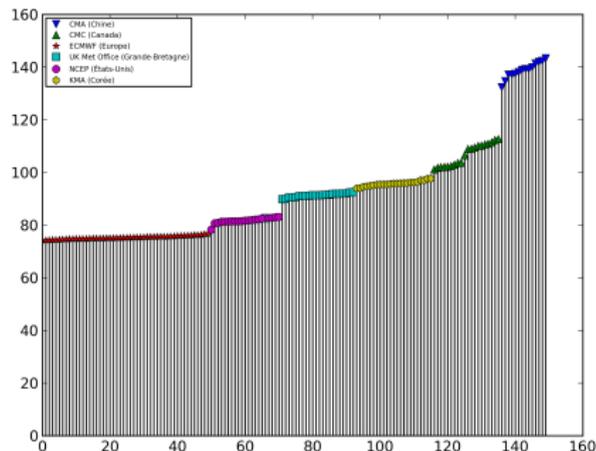


FIGURE : Représentation graphique de la performance moyenne de chacun des membres de l'ensemble issus de TIGGE, triés par ordre croissant de $RMSE$, selon la formule $\frac{1}{N_x \times N_y} \sum_{(i,j) \in \text{Carte}} \sqrt{\frac{1}{T} \sum_{t=1}^T (x_{(i,j),t}^m - y_{t,(i,j)})^2}$. Les prévisions sont réalisées à une échéance de 6 heures. Le symbole en haut de chacune des barres verticales indique l'origine (centre météorologique) du membre correspondant. Il s'agit de l'ensemble à 150 membres, ne comprenant pas les prévisions déterministes de Météo-France ou d'ECMWF. Remarquons que ce sont les prévisions du centre européen qui réalisent les erreurs les plus basses.

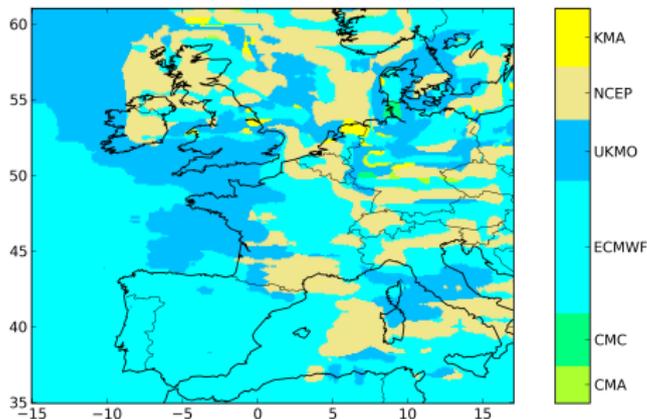


FIGURE : Représentation graphique dans chaque cellule (i, j) de la carte de $m^*(i, j)$, du meilleur membre annuel de l'ensemble issu de TIGGE (constitué de $M = 150$ membres) pour les prévisions à horizon de six heures. Ce meilleur membre est celui qui possède la *RMSE* la plus basse dans la cellule (i, j) . Les centres de prévisions ont les origines suivantes : CMA - Chine, CMC - Canada, ECMWF - Europe, UKMO - Grande-Bretagne, NCEP - États-Unis, KMA - Corée. Les meilleurs modèles sont issus d'ECMWF (50% de la carte), UKMO (25%) et NCEP (20%).

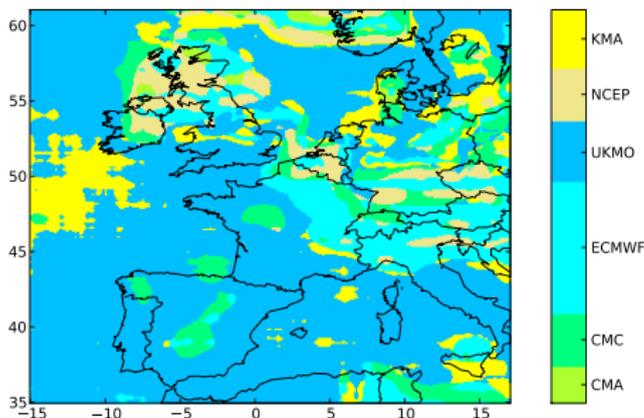


FIGURE : Représentation graphique dans chaque cellule (i, j) de la carte de $e^*(i, j)$, de la meilleure moyenne d'ensemble parmi les moyennes d'ensemble des centres régionaux de TIGGE (constitué de $M = 150$ membres, répartis dans 6 centres régionaux), lors du premier jour de la période considérée, pour les prévisions à horizon de six heures. Cette meilleure moyenne d'ensemble est celle qui possède la *RMSE* la plus basse dans la cellule (i, j) . Les centres de prévision ont les origines suivantes : CMA - Chine, CMC - Canada, ECMWF - Europe, UKMO - Grande-Bretagne, NCEP - États-Unis, KMA - Corée.

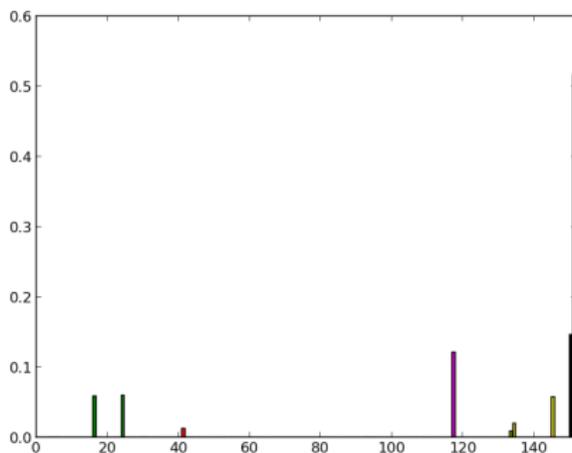


FIGURE : Représentation graphique des poids de l'oracle convexe pour l'ensemble complet (152 membres). La meilleure combinaison convexe constante tire parti uniquement de quelques membres de l'ensemble : seuls les poids d'une dizaine de membres diffèrent de zéro à 10^{-6} près. Notons que ces poids sont bien tous compris entre 0 et 1 et leur somme est égale à 1. Notons que les deux derniers poids sont affectés aux prévisions déterministes de l'ECMWF et de Météo-France, dans cet ordre.

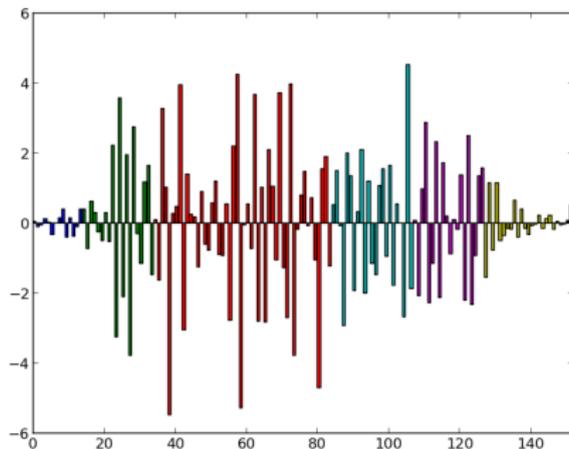


FIGURE : Représentation graphique des poids de l'oracle linéaire pour l'ensemble complet (152 membres). La meilleure combinaison linéaire constante tire parti de tous les membres de l'ensemble : ils sont tous visiblement différents de zéro. Notons que ces poids varient dans \mathbb{R} . La somme de ces poids est proche de 1 (à 10^{-4} près), et cela n'est pas nécessairement attendu par la théorie dans le cas des poids linéaires.

origine	<i>RMSE</i> (Pa)
moyenne d'ensemble	52,4
déterministe ECMWF	34,2
déterministe Météo-France	30,1

TABLE : Performances des trois prévisions candidates au rôle de prévision de référence dans le cadre non-asymptotique.

Type de prévision	RMSE (Pa)	$\Delta_{\%}(A, R)$ (%)
Moyenne d'ensemble	53,1	-74,8
Déterministe	30,4	0,0
Oracle convexe	24,4	19,6
Oracle linéaire	18,1	40,4
Agrégation ridge (avec paramètres optimaux rétrospectifs)	24,9	18,1

TABLE : *RMSE* : Scores et différences relatives par rapport à la prévision déterministe de Météo-France pour une prévision agrégée à un horizon de 6 heures pour des paramètres optimisés au préalable : $(\lambda; \gamma) = (60\ 000; 15)$. Le poids initial est une mesure de Dirac sur le déterministe Météo-France. La période d'entraînement est de 100 jours.

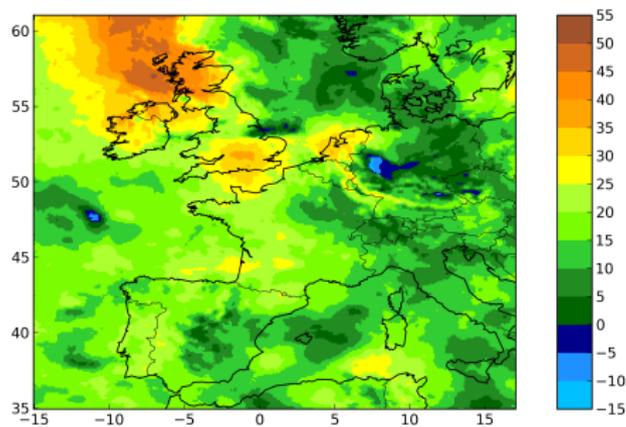


FIGURE : Représentation graphique des différences relatives de *RMSE* entre l'agrégé avec paramètres optimaux rétrospectifs et le déterministe Météo-France à un horizon de 6 heures. Dans chaque cellule (i, j) est représentée la différence relative : $(r_{(i,j)}^{\text{det}} - r_{(i,j)}^{\text{agr}}) / r_{(i,j)}^{\text{det}}$ où $r_{(i,j)}^{\text{agr}}$ (respectivement $r_{(i,j)}^{\text{det}}$) est la *RMSE* moyenne de la prévision agrégée (respectivement déterministe) dans la cellule (i, j) . Les paramètres sont optimaux et fixés à l'avance, l'ensemble compte 152 membres (les déterministes d'ECMWF et de Météo-France sont compris). Le poids initial est une mesure de Dirac sur le déterministe Météo-France. La période d'entraînement est de 100 jours.

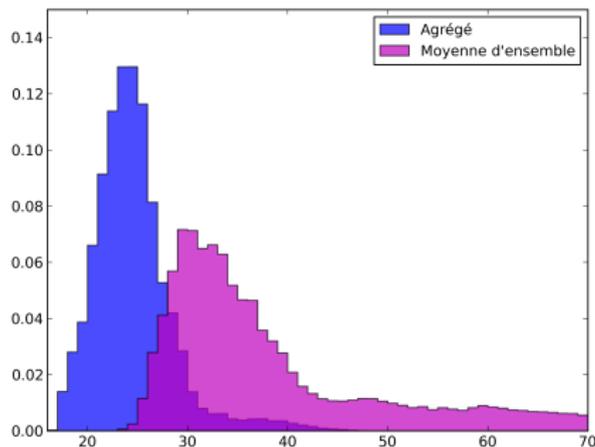


FIGURE : Distribution des *RMSE* de l'agrégé (en bleu) et de la moyenne d'ensemble (en magenta). Les individus représentés sont les *RMSE* pour l'agrégé $r_{(i,j)}^{agr}$ et les *RMSE* pour la moyenne d'ensemble $r_{(i,j)}^{ens}$ des cellules (i, j) de la carte dans son intégralité. Les prévisions sont réalisées à une échéance de 6 heures, pour des paramètres optimaux de la régression ridge. L'ensemble fait 152 membres (le déterministe Météo-France et le déterministe ECMWF sont inclus). La *RMSE* est prise par rapport à l'analyse Météo-France. Le poids initial est une mesure de Dirac sur le déterministe Météo-France.

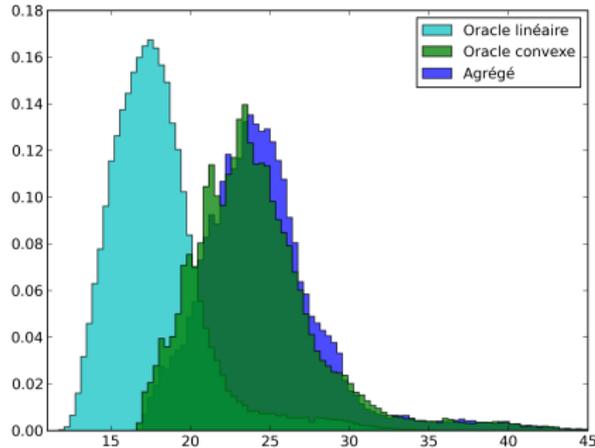


FIGURE : Histogramme des $RMSE$ de l'oracle linéaire (en cyan), de l'oracle convexe (en jaune) et de l'agrégé (en bleu). Les individus représentés sont les $RMSE$ pour l'oracle linéaire $r_{(i,j)}^{\text{lin}}$, les $RMSE$ pour l'oracle convexe $r_{(i,j)}^{\text{cvx}}$ et les $RMSE$ pour l'agrégé $r_{(i,j)}^{\text{agr}}$ des cellules (i,j) de la carte dans son intégralité. Les prévisions sont réalisées à une échéance de 6 heures, pour des paramètres optimaux. L'ensemble fait 152 membres (le déterministe Météo-France et le déterministe ECMWF sont inclus). La $RMSE$ est prise par rapport à l'analyse Météo-France. Le poids initial est une mesure de Dirac sur le déterministe Météo-France.

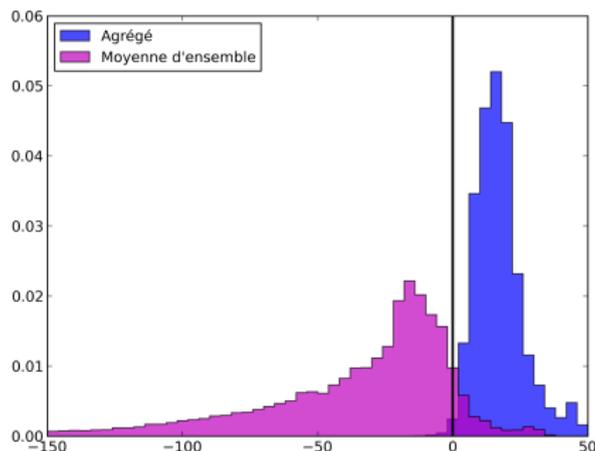


FIGURE : Histogramme de la différence relative par rapport au déterministe de la *RMSE* de la moyenne d'ensemble (en magenta) et de l'agrégé (en bleu). Les individus représentés sont les différences relatives en *RMSE* pour l'agrégé et pour la moyenne d'ensemble. Les prévisions sont à une échéance de 6 heures, pour des paramètres optimaux. L'ensemble fait 152 membres (le déterministe Météo-France et le déterministe ECMWF sont inclus). La *RMSE* est prise par rapport à l'analyse Météo-France. Le poids initial est une mesure de Dirac sur le déterministe Météo-France.

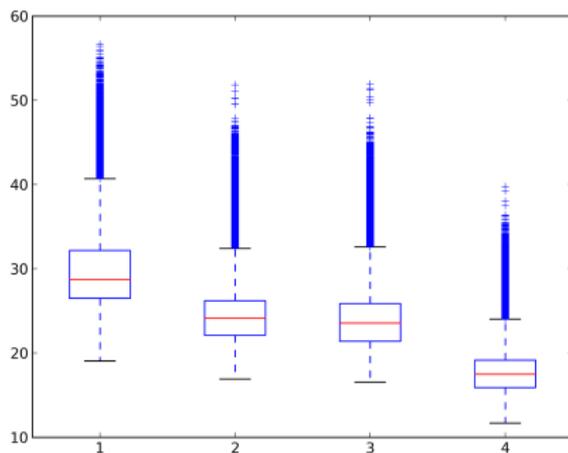


FIGURE : Diagramme en boîte des *RMSE* à horizon de 6 heures entre l'analyse Météo-France et respectivement, le déterministe (1), l'agrégé issu de la régression ridge (2), l'oracle convexe (3) et l'oracle linéaire (4). Les individus représentés sont les *RMSE* pour l'agrégé $r_{(i,j)}^{agr}$, les *RMSE* pour la moyenne d'ensemble $r_{(i,j)}^{ens}$, les *RMSE* pour l'oracle convexe $r_{(i,j)}^{cvx}$ et les *RMSE* pour l'oracle linéaire $r_{(i,j)}^{lin}$ des cellules (i, j) de la carte dans son intégralité. À un individu plus bas correspond une *RMSE* plus basse donc de meilleures performances. Les paramètres sont optimaux et fixés à l'avance, l'ensemble compte 152 membres (les déterministes d'ECMWF et de Météo-France sont compris). Le poids initial est une mesure de Dirac sur le déterministe Météo-France.

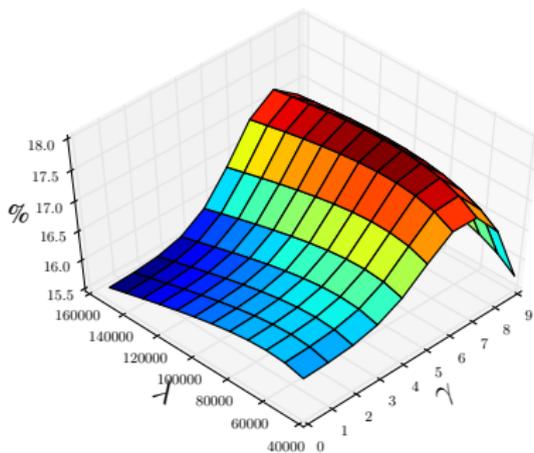


FIGURE : Représentation graphique de l'influence des paramètres sur les performances de l'algorithme de régression ridge escompté pour la prévision à horizon 6 heures de la pression réduite au niveau de la mer. La cote représente la différence relative (en pourcentage), entre d'une part, la prévision agrégée avec les paramètres en abscisse et ordonnée et, d'autre part, la prévision du déterministe Météo-France. Le facteur d'escompte γ varie en puissance de 10 entre 10^{-7} et 10^4 et le facteur de régularisation λ varie par pas linéaires de 10000 dans le segment $[40000, 160000]$. Le maximum sur cette plage de paramètre a pour valeur 17,3% et a pour antécédant le couple $(\lambda, \gamma) = (100\ 000, 100)$.

Type de prévision	Biais (Pa)	Différence relative (%)
Moyenne d'ensemble	35,6	-53,7
Déterministe	23,2	0,0
Oracle convexe	18,5	20,0
Oracle linéaire	14,1	39,2
Agrégation ridge (avec adaptation locale rétrospective des paramètres)	18,7	19,3
Agrégation ridge (avec paramètres optimaux rétrospectifs)	18,8	18,8

TABLE : Stratégie de régression ridge avec optimisation locale des paramètres (excepté pour la dernière ligne). Biais : scores et différences relatives par rapport à la prévision déterministe de Météo-France pour une prévision agrégée à un horizon de 6 heures pour un jeu de paramètres appartenant à une grille prédéfinie. 1000 cellules sont sélectionnées aléatoirement. Le poids initial est une mesure de Dirac sur le déterministe Météo-France. La période d'entraînement est de 100 jours.

Type de prévision	Biais (Pa)	Différence relative (%)
Moyenne d'ensemble	35,6	-53,7
Déterministe	23,2	0,0
Oracle convexe	18,5	20,0
Oracle linéaire	14,1	39,2
Agrégation ridge (paramètres optimaux rétrospectifs)	18,8	18,8
Agrégation ridge (adaptation locale rétrospective des paramètres)	18,7	19,3
Agrégation ridge (adaptation locale en ligne sur une grille de paramètres)	18,7	19,4

TABLE : Stratégie partiellement automatique. Biais : scores et différences relatives par rapport à la prévision déterministe de Météo-France pour une prévision agrégée à un horizon de 6 heures pour des paramètres pouvant varier tous les $t_{\text{bascule}} = 30$ jours sur une grille prédéfinir. 1000 cellules sont sélectionnées aléatoirement. Le poids initial est une mesure de Dirac sur le déterministe Météo-France. La période d'entraînement est de 100 jours.

Origine (acronyme)	RMSE (Pa)	Différence relative (%)
Ridge escompté		
Chine (CMA)	24,9	17,3
Canada (CMC)	24,7	18,0
Europe (ECMWF)	24,9	17,4
Corée (KMA)	24,7	18,3
États-Unis (NCEP)	24,2	19,7
Grande-Bretagne (UKMO)	24,6	17,9
Ensemble complet	24,7	18,0
Moyenne d'ensemble		
Chine (CMA)	67,2	-123,0
Canada (CMC)	65,0	-115,7
Europe (ECMWF)	55,8	-85,3
Corée (KMA)	70,0	-130,9
États-Unis (NCEP)	70,7	-134,8
Grande-Bretagne (UKMO)	69,6	-132,2
Ensemble complet	52,4	-73,9

TABLE : RMSE : scores et différences relatives par rapport à la prévision déterministe de Météo-France pour différentes prévisions agrégées à un horizon de 6 heures pour des paramètres optimisés au préalable : $(\lambda; \gamma) = (100\ 000; 50)$. 1000 cellules sont sélectionnées aléatoirement. Le poids initial est une mesure de Dirac sur le déterministe Météo-France. La période d'entraînement est de 100 jours.

Type de prévision	<i>RMSE</i> (Pa)	Différence relative (%)
Moyenne d'ensemble	29,8	-0,1
Déterministe	27,9	0,0
Oracle convexe	21,3	23,6
Oracle linéaire	14,4	48,2
Paramètre de régularisation		
6×10^4	22,5	19,3
6×10^5	22,1	20,6
6×10^6	24,0	14,0

TABLE : Scores et différences relatives par rapport à la prévision déterministe de Météo-France en fonction de la régularisation.

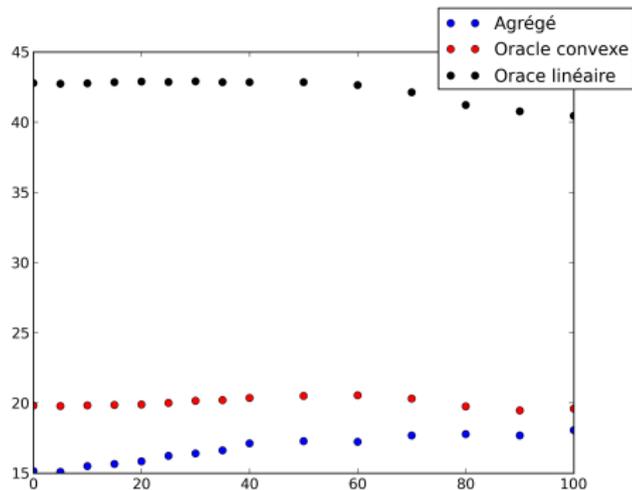


FIGURE : Évolution de de la différence relative de $RMSE$ entre la prévision déterministe d'une part et, d'autre part, la prévision agrégée (en bleu), l'oracle convexe (en rouge) et l'oracle linéaire (en noir) en fonction de la période d'entraînement sur un échantillon de 1000 cellules.

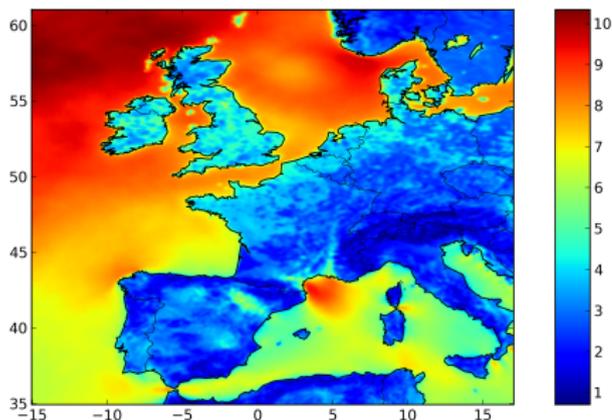


FIGURE : Représentation graphique de la moyenne temporelle de l'analyse, dans le cas de la vitesse du vent à 10 mètres au-dessus du sol. Dans chaque cellule de coordonnées (i, j) est représentée la moyenne $\frac{1}{T} \sum_{t=1}^T y_{t,(i,j)}$. L'unité est le m s^{-1} . La moyenne de ces valeurs est de $5,6 \text{ m s}^{-1}$, et l'écart type de ces valeurs moyennes est de $2,62 \text{ m s}^{-1}$.

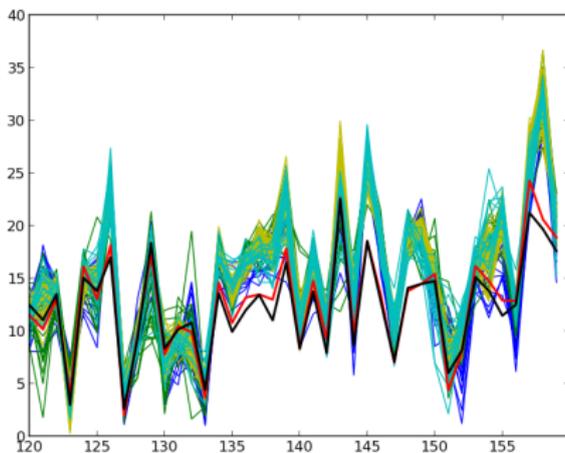


FIGURE : Données de la vitesse du vent à 10 mètres au-dessus du sol dans une cellule typique (de coordonnées $(-13^\circ, 37^\circ)$), entre les échéances 320 (16 août 2012) et 360 (25 septembre 2012) de la période temporelle considérée. Un échantillon de 100 membres de l'ensemble parmi les 150 est représenté. En rouge, la prévision déterministe, $x_{t,(i,j)}^{\text{det}}$; en noir, l'analyse, $y_{t,(i,j)}$; le reste des couleurs est dédié aux membres de l'ensemble, $x_{t,(i,j)}^m$.

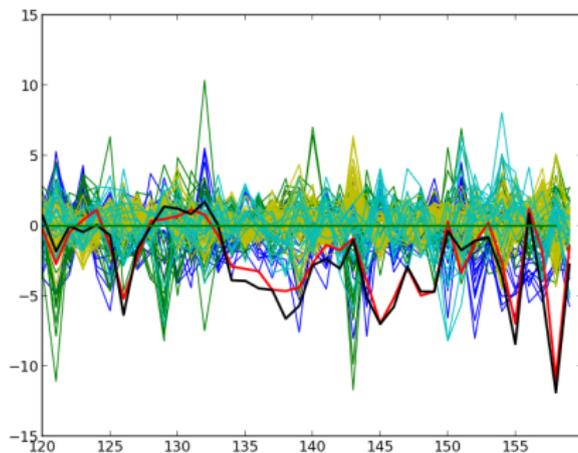


FIGURE : Représentation graphique alternative des données de 19, ici centrées autour de la moyenne d'ensemble, $\bar{x}_{t,(i,j)}$, représentée par la ligne horizontale en vert foncé et d'ordonnée 0, afin d'assurer une plus grande visibilité de la variabilité. Un échantillon de 100 membres de l'ensemble parmi les 150 est représenté. En rouge, la prévision déterministe centrée, $x_{t,(i,j)}^{\text{det}} - \bar{x}_{t,(i,j)}$; en noir, l'analyse centrée, $y_{t,(i,j)} - \bar{x}_{t,(i,j)}$; le reste des couleurs est dédié aux membres de l'ensemble, $x_{t,(i,j)}^m - \bar{x}_{t,(i,j)}$.

Type de prévision	$RMSE$ ($m s^{-1}$)
Moyenne d'ensemble	3,1
Déterministe ECMWF	2,8
Déterministe Météo France	2,6
Oracle convexe	2,3
Oracle linéaire	1,6

TABLE : $RMSE$: Score à un horizon de 6 heures avec une période d'entraînement de 100 jours.

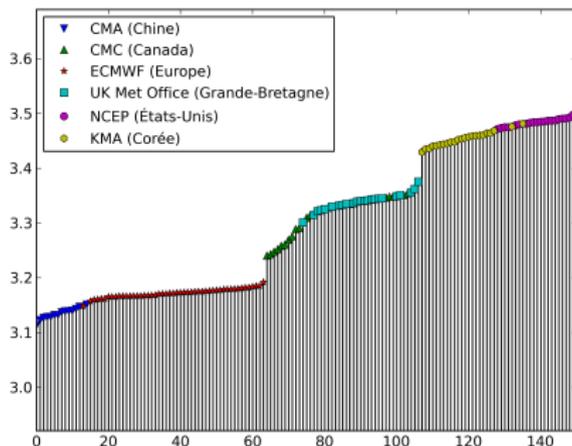


FIGURE : Représentation graphique de la performance moyenne de chacun des membres de l'ensemble issus de TIGGE, triés par ordre croissant de $RMSE$, selon la formule $\frac{1}{N_x \times N_y} \sum_{(i,j) \in \text{Carte}} \sqrt{\frac{1}{T} \sum_{t=1}^T (x_{(i,j),t}^m - y_{t,(i,j)})^2}$. Les prévisions sont réalisées à une échéance de 6 heures. Le symbole en haut de chacune des barres verticales indique l'origine (centre météorologique) du membre correspondant. Il s'agit de l'ensemble à 150 membres, ne comprenant pas les prévisions déterministes de Météo France ou d'ECMWF. Remarquons que ce sont les prévisions du centre européen qui réalisent les erreurs les plus basses.

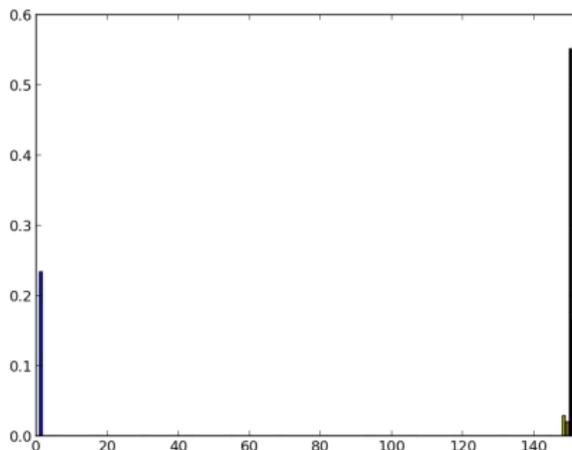


FIGURE : Représentation graphique des poids de l'oracle convexe pour l'ensemble complet (152 membres). La meilleure combinaison convexe constante tire parti uniquement de quelques membres de l'ensemble : seuls les poids d'une dizaine de membres diffèrent de zéro à 10^{-6} près. Notons que ces poids sont bien tous compris entre 0 et 1 et leur somme est égale à 1 (à 10^{-11} près). Notons que les deux derniers poids sont affectés aux prévisions déterministes de l'ECMWF et de Météo-France, dans cet ordre.

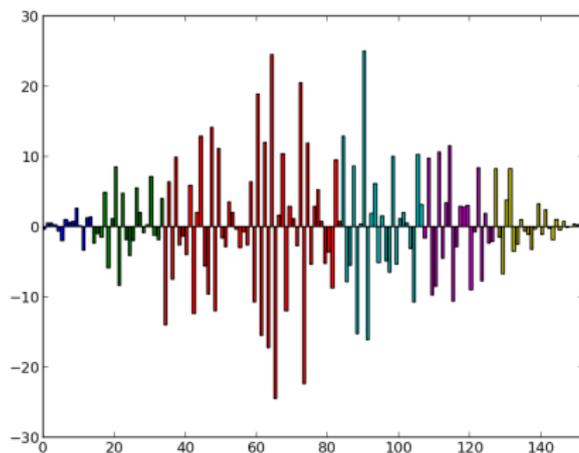


FIGURE : Représentation graphique des poids de l'oracle linéaire pour l'ensemble complet (152 membres). La meilleure combinaison linéaire constante tire parti de tous les membres de l'ensemble : ils sont tous visiblement différents de zéro. Notons que ces poids varient dans \mathbb{R} . La somme de ces poids est égale à 0,9. Nous pouvons interpréter cet écart à la valeur 1 (qui est bien autorisé par la théorie) comme la conséquence d'un débiaisement de l'ensemble.

Type de prévision	$RMSE$ ($m s^{-1}$)	Gain relatif (%)
Moyenne d'ensemble	3,1	-18,7
Déterministe	2,6	0,0
Oracle convexe	2,3	9,8
Oracle linéaire	1,6	38,1
Agrégation ridge (avec paramètres optimaux rétrospectifs)	2,3	10,9

TABLE : $RMSE$: scores et gains relatifs par rapport à la prévision déterministe de Météo-France pour une prévision agrégée à un horizon de 6 heures pour des paramètres optimisés au préalable : $(\lambda; \gamma) = (25; 0,2)$. La période d'entraînement est de 100 jours.

Type de prévision	Biais (m s^{-1})	Gain relatif (%)
Moyenne d'ensemble	2,3	-29,2
Déterministe	1,8	0,0
Oracle convexe	1,6	8,9
Oracle linéaire	1,2	33,9
Agrégation ridge (avec paramètres optimaux rétrospectifs)	1,6	11,2

TABLE : Biais : scores et gains relatifs par rapport à la prévision déterministe de Météo-France pour une prévision agrégée à un horizon de 6 heures pour des paramètres optimisés au préalable : $(\lambda; \gamma) = (25; 0,2)$. La période d'entraînement est de 100 jours.

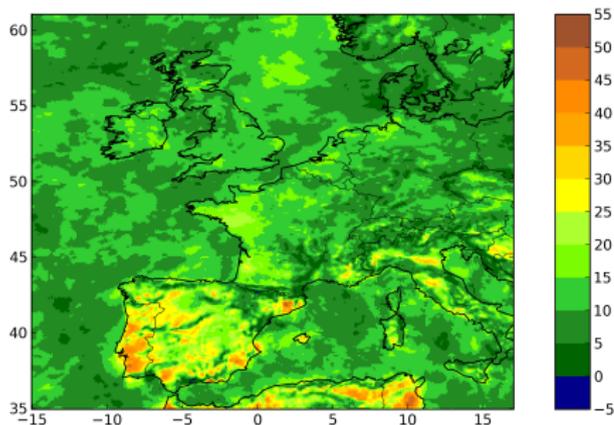


FIGURE : Représentation graphique des différences relatives en $RMSE$ entre l'agrégé avec paramètres optimaux rétrospectifs et le déterministe Météo-France à un horizon de 6 heures. Dans chaque cellule (i, j) est représentée la différence relative : $(RMSE_{(i,j)}^{det} - RMSE_{(i,j)}^{agr}) / RMSE_{(i,j)}^{det}$ où $r_{(i,j)}^{agr}$ (respectivement $r_{(i,j)}^{det}$) est la $RMSE$ moyenne de la prévision agrégée (respectivement déterministe) dans la cellule (i, j) . Les paramètres sont optimaux et fixés à l'avance, l'ensemble compte 152 membres (les déterministes d'ECMWF et de Météo-France sont compris). Le poids initial est une mesure de Dirac sur le déterministe Météo-France. La période d'entraînement est de 100 jours.

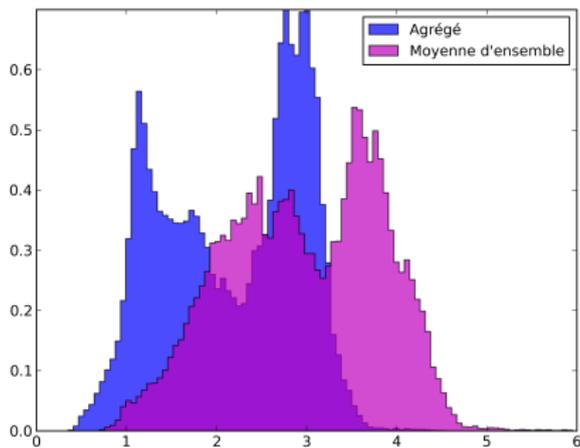


FIGURE : Histogramme des $RMSE$ de l'agrégé (en bleu) et de la moyenne d'ensemble (en magenta). Les individus représentés sont les $RMSE$ pour l'agrégé $r_{(i,j)}^{agr}$ et les $RMSE$ pour la moyenne d'ensemble $r_{(i,j)}^{ens}$ des cellules (i, j) de la carte dans son intégralité. Les prévisions sont réalisées à une échéance de 6 heures, pour des paramètres optimaux de la régression ridge. L'ensemble fait 152 membres (le déterministe Météo-France et le déterministe ECMWF sont inclus). La $RMSE$ est prise par rapport à l'analyse Météo-France. Le poids initial est une mesure de Dirac sur le déterministe Météo-France.

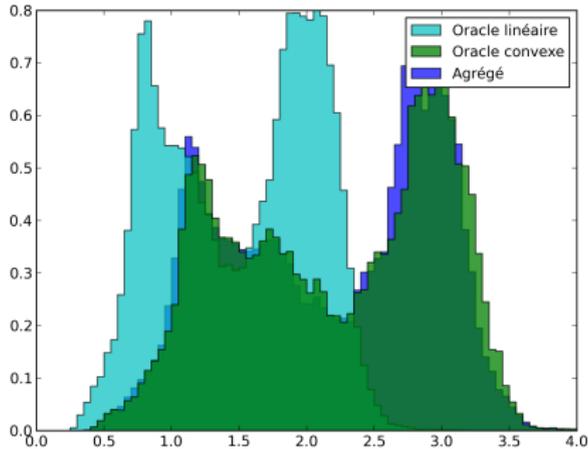


FIGURE : Histogramme des $RMSE$ de l'oracle linéaire (en cyan), de l'oracle convexe (en vert) et de l'agrégé (en bleu). Les individus représentés sont les $RMSE$ pour l'oracle linéaire $r_{(i,j)}^{lin}$, les $RMSE$ pour l'oracle convexe $r_{(i,j)}^{cvx}$ et les $RMSE$ pour l'agrégé $r_{(i,j)}^{agr}$ des cellules (i,j) de la carte dans son intégralité. Les prévisions sont réalisées à une échéance de 6 heures, pour des paramètres optimaux. L'ensemble fait 152 membres (le déterministe Météo-France et le déterministe ECMWF sont inclus). La $RMSE$ est prise par rapport à l'analyse Météo-France. Le poids initial est une mesure de Dirac sur le déterministe Météo-France.

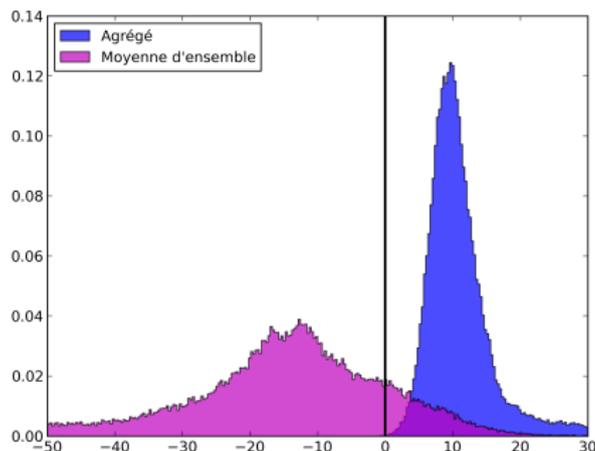


FIGURE : Histogramme de la différence relative par rapport au déterministe de la *RMSE* de la moyenne d'ensemble (en magenta) et de l'agrégé (en bleu). Les individus représentés sont les différences relatives en *RMSE* pour l'agrégé et pour la moyenne d'ensemble. Les prévisions sont à une échéance de 6 heures, pour des paramètres optimaux. L'ensemble fait 152 membres (le déterministe Météo-France et le déterministe ECMWF sont inclus). La *RMSE* est prise par rapport à l'analyse Météo-France. Le poids initial est une mesure de Dirac sur le déterministe Météo-France.

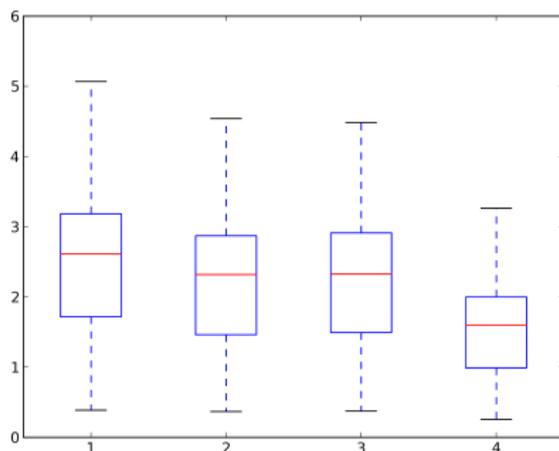


FIGURE : Diagramme en boîte des *RMSE* à horizon de 6 heures entre l'analyse Météo-France et respectivement, le déterministe (1), l'agrégé issu de la régression ridge (2), l'oracle convexe (3) et l'oracle linéaire (4). Les individus représentés sont les *RMSE* pour l'agrégé $r_{(i,j)}^{agr}$, les *RMSE* pour la moyenne d'ensemble $r_{(i,j)}^{ens}$, les *RMSE* pour l'oracle convexe $r_{(i,j)}^{cvx}$ et les *RMSE* pour l'oracle linéaire $r_{(i,j)}^{lin}$ des cellules (i, j) de la carte dans son intégralité. À un individu plus bas correspond une *RMSE* plus basse donc de meilleures performances. Les paramètres sont optimaux et fixés à l'avance, l'ensemble compte 152 membres (les déterministes d'ECMWF et de Météo-France sont compris). Le poids initial est une mesure de Dirac sur le déterministe Météo-France.

Type de prévision	$RMSE$ ($m s^{-1}$)	Différence relative (%)
Moyenne d'ensemble	2,3	-47,0
Déterministe	1,6	0,0
Oracle convexe	1,4	9,4
Oracle linéaire	1,4	12,5
Agrégation ridge (paramètres optimaux rétrospectifs)	1,4	9,0
Agrégation ridge (adaptation locale en ligne sur une grille de paramètres)	1,5	6,5

TABLE : Stratégie partiellement automatique. $RMSE$: scores et différences relatives par rapport à la prévision déterministe de Météo-France pour une prévision agrégée à un horizon de 6 heures pour des paramètres pouvant varier tous les $t_{\text{bascule}} = 30$ jours sur une grille prédéfinie. 10000 cellules sont sélectionnées aléatoirement. Le poids initial est une mesure de Dirac sur le déterministe Météo-France. La période d'entraînement est de 100 jours.

Type de prévision	Biais (m s^{-1})	Différence relative (%)
Moyenne d'ensemble	1,7	-66,6
Déterministe	1,0	0,0
Oracle convexe	1,0	6,5
Oracle linéaire	0,9	8,5
Agrégation ridge (paramètres optimaux rétrospectifs)	1,0	7,3
Agrégation ridge (adaptation locale en ligne sur une grille de paramètres)	1,0	6,4

TABLE : Stratégie partiellement automatique. Biais : scores et différences relatives par rapport à la prévision déterministe de Météo-France pour une prévision agrégée à un horizon de 6 heures pour des paramètres pouvant varier tous les $t_{\text{bascule}} = 30$ jours sur la grille prédéfinie. 10000 cellules sont sélectionnées aléatoirement. Le poids initial est une mesure de Dirac sur le déterministe Météo-France. La période d'entraînement est de 100 jours.

Initialisation : $R_0 = (0, \dots, 0)$ est le vecteur de regret initial.

À chaque échéance $t = 1, 2, \dots$

1. on calcule les composantes du vecteur de paramètres d'apprentissage η_{t-1} en suivant la règle de décision :

$$\eta_{m,t-1} = \frac{1}{1 + \sum_{s=1}^{t-1} (\widehat{\ell}_s - \ell_{m,s})^2}$$

2. on crée le vecteur de poids p_t défini composante par composante par

$$p_{m,t} = \eta_{m,t-1} (R_{m,t-1})_+ / \eta_{t-1} (R_{t-1})_+$$

où x_+ représente le vecteur constitué des parties positives de chaque composante de x ;

3. on observe le vecteur de perte ℓ_t , de composantes

$$\ell_{m,t} = \int_{\gamma}^{\Gamma} (H_{y_t}(z) - H_{x_{m,t}}(z))^2 dz$$

et on calcule $\widehat{\ell}_t = \sum_{m=1}^M p_{m,t} \ell_{m,t}$;

4. on met à jour le regret pour chaque expert $R_{m,t} = R_{m,t-1} + \widehat{\ell}_t - \ell_{m,t}$.

<i>CRPS</i>	MSLP (Pa)	Vitesse du vent (m s^{-1})
Ensemble TIGGE	28,0	1,2
Ensemble ECMWF	32,8	1,4
Échelon déterministe	23,4	0,8

TABLE : Comparaison des valeurs de *CRPS* entre trois types de prévisions probabilistes dans le cas de la pression réduite au niveau de la mer et de la norme de la vitesse du vent. Les prévisions sont effectuées à un horizon de 6 heures. Ces scores sont évalués sur l'intégralité du domaine et de la période.

η	CRPS (Pa)	différence relative (%)
0.0001	23,4	-3,0
0.001	16,7	26,7
0.01	15,4	32,6
0.1	21,7	4,5

TABLE : Performance du *CRPS* en fonction au paramètre η de la prévision de pression réduite au niveau de la mer. Scores et différences relatives par rapport à la prévision déterministe de Météo-France pour une prévision probabiliste agrégée à un horizon de 6 heures. Le poids initial est uniforme. La période d'entraînement est de 100 jours et 100 cellules sont sélectionnées aléatoirement pour évaluer ces scores.

Type de prévision	CRPS (Pa)	différence relative (%)
Ensemble TIGGE	28,0	-19,7
Ensemble ECMWF	32,8	-40,1
Échelon déterministe	23,4	0,0
Poids exponentiels (avec paramètre optimal a posteriori)	15,6	33,4
ML-poly-grad	19,3	17,6

TABLE : CRPS : Scores et différences relatives par rapport à la prévision déterministe de Météo-France pour une prévision probabiliste agrégée à un horizon de 6 heures. Le poids initial est uniforme chaque ensemble considéré. La période d'entraînement est de 100 jours. Ces scores sont évalués sur l'intégralité de la carte.

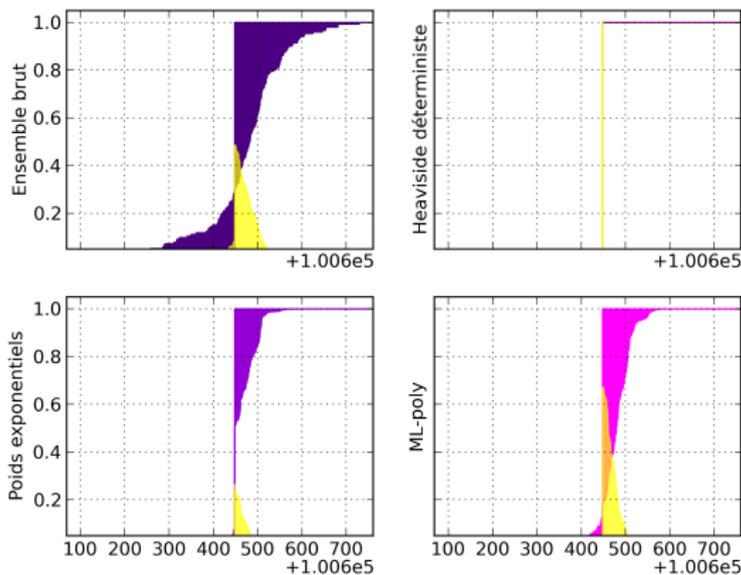


FIGURE : Représentation de plusieurs prévisions probabilistes et visualisation de leurs CRPS respectifs, dans une cellule typique (de coordonnées $(-13^\circ, 37^\circ)$), à l'échéance 320 (16 août 2012). Les fonctions de répartition de haut en bas, de gauche à droite sont la pondération uniforme de l'ensemble, l'échelon déterministe, $\mathcal{E}_\eta^{\text{grad}}$ et ML-poly-grad.

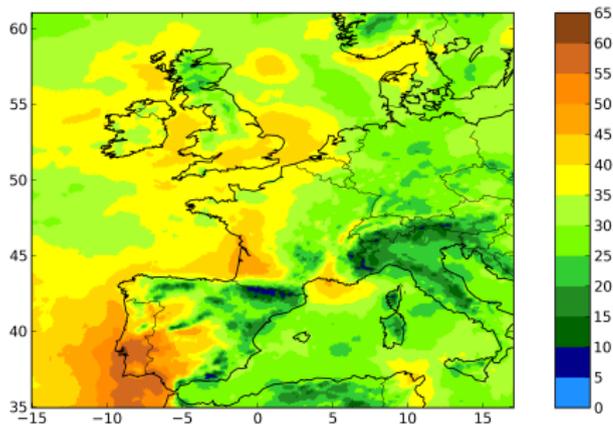


FIGURE : Représentation graphique des différences relatives du *CRPS* entre la fonction de répartition agrégée par les poids exponentiels $\mathcal{E}_\eta^{\text{grad}}$ avec paramètre optimal rétrospectif et l'échelon déterministe Météo-France à un horizon de 6 heures. Dans chaque cellule (i, j) est représentée la différence relative : $(r_{(i,j)}^{\text{det}} - r_{(i,j)}^{\text{agr}}) / r_{(i,j)}^{\text{det}}$. Les paramètres sont optimaux et fixés à l'avance, l'ensemble compte 152 membres (les déterministes d'ECMWF et de Météo-France sont compris). Les poids initiaux sont uniformes. La période d'entraînement est de 100 jours.

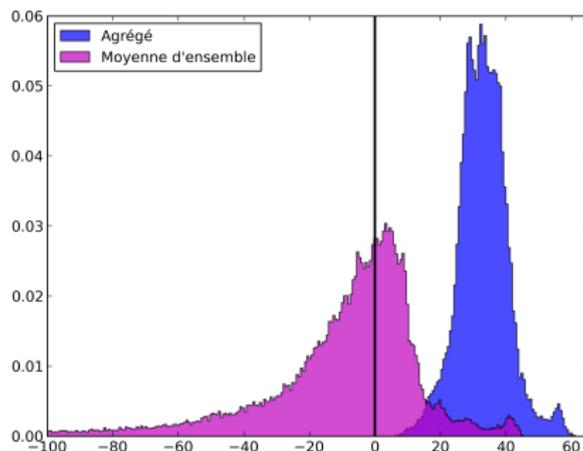


FIGURE : Histogramme de la différence relative par rapport à l'échelon déterministe du *CRPS* entre la fonction de répartition uniforme issue de l'ensemble TIGGE (en magenta) et de l'agrégé de ML-poly-grad (en bleu). Les individus représentés sont les différences relatives du *CRPS* pour l'agrégé et pour la moyenne d'ensemble. Les prévisions sont à une échéance de 6 heures, pour des paramètres optimaux. L'ensemble fait 152 membres (le déterministe Météo-France et le déterministe ECMWF sont inclus). Le *CRPS* est pris par rapport à l'échelon d'analyse Météo-France. Les poids initiaux sont uniformes.

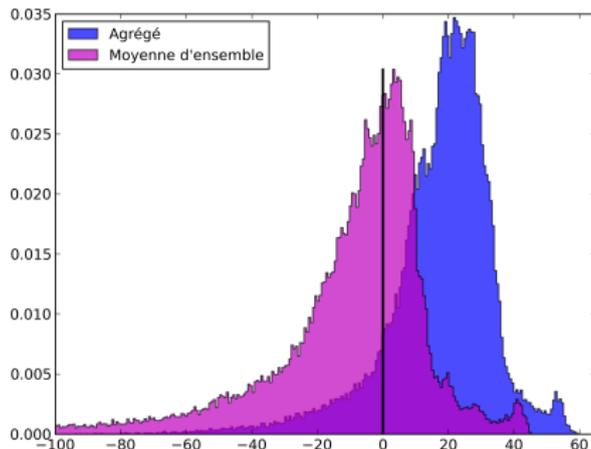


FIGURE : Histogramme de la différence relative par rapport à l'échelon déterministe du *CRPS* entre la fonction de répartition uniforme issue de l'ensemble TIGGE (en magenta) et de l'agrégé ML-poly-grad (en bleu). Les individus représentés sont les différences relatives du *CRPS* pour l'agrégé pour la moyenne d'ensemble. Les prévisions sont à une échéance de 6 heures, pour des paramètres optimaux. L'ensemble fait 152 membres (le déterministe Météo-France et le déterministe ECMWF sont inclus). Le *CRPS* est pris par rapport à l'échelon d'analyse Météo-France. Le poids initial est uniforme.

η	CRPS (m s^{-1})	différence relative (%)
0.001	1,3	-20,0
0.01	0,9	12,6
0.1	0,8	25,5
1	1,0	6,7
10	1,3	-24,1

TABLE : Performance du CRPS en fonction du paramètre η : scores et différences relatives par rapport à la prévision déterministe de Météo-France pour une prévision probabiliste agrégée à un horizon de 6 heures. Le poids initial est uniforme chaque ensemble considéré. La période d'entraînement est de 100 jours. 100 cellules sont sélectionnées aléatoirement pour évaluer ces scores.

Type de prévision	CRPS (m s^{-1})	différence relative (%)
Ensemble TIGGE	1,2	-36,6
Ensemble ECMWF	1,4	-66,9
Échelon déterministe	0,8	0,0
Poids exponentiels (avec paramètre optimal a posteriori)	0,6	24,1
ML-poly-grad	0,7	13,1

TABLE : CRPS : Scores et différences relatives par rapport à la prévision déterministe de Météo-France pour une prévision probabiliste agrégée à un horizon de 6 heures. Le poids initial est uniforme chaque ensemble considéré. La période d'entraînement est de 100 jours. 1000 cellules sont sélectionnées aléatoirement pour évaluer ces scores.

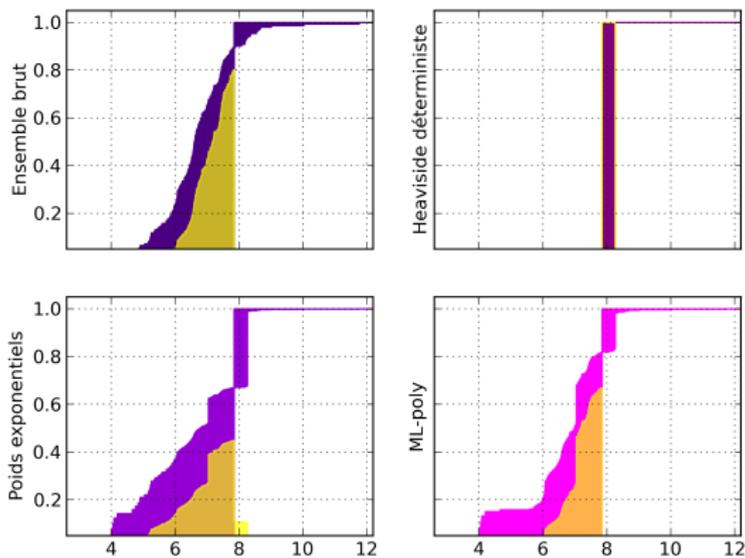


FIGURE : Représentation de plusieurs prévisions probabilistes et visualisation de leurs CRPS respectifs, dans une cellule typique (de coordonnées $(-13^\circ, 37^\circ)$), à l'échéance 320 (16 août 2012). Les fonctions de répartition de haut en bas, de gauche à droite sont la pondération uniforme de l'ensemble, l'échelon déterministe, $\mathcal{E}_\eta^{\text{grad}}$ et ML-poly-grad.

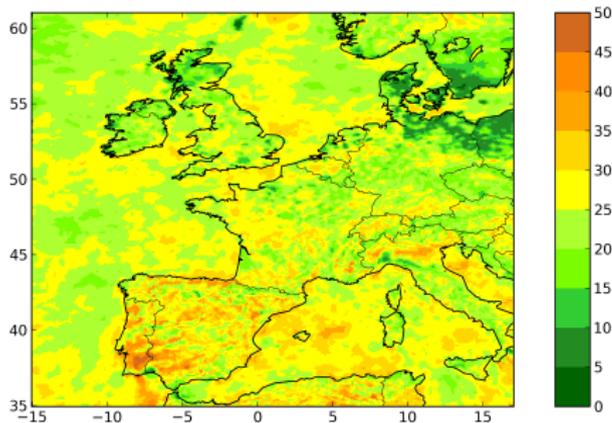


FIGURE : Représentation graphique des différences relatives du *CRPS* entre la fonction de répartition agrégée par les poids exponentiels $\mathcal{E}_\eta^{\text{grad}}$ avec paramètre optimal rétrospectif et l'échelon déterministe Météo-France à un horizon de 6 heures. Dans chaque cellule (i, j) est représentée la différence relative : $(r_{(i,j)}^{\text{det}} - r_{(i,j)}^{\text{agr}}) / r_{(i,j)}^{\text{det}}$ où $r_{(i,j)}^{\text{agr}}$ (respectivement $r_{(i,j)}^{\text{det}}$) est le *CRPS* moyen de la prévision agrégée (respectivement déterministe) dans la cellule (i, j) . Les paramètres sont optimaux et fixés à l'avance, l'ensemble compte 152 membres (les déterministes d'ECMWF et de Météo-France sont compris). Le poids initial est uniforme. La période d'entraînement est de 100 jours.

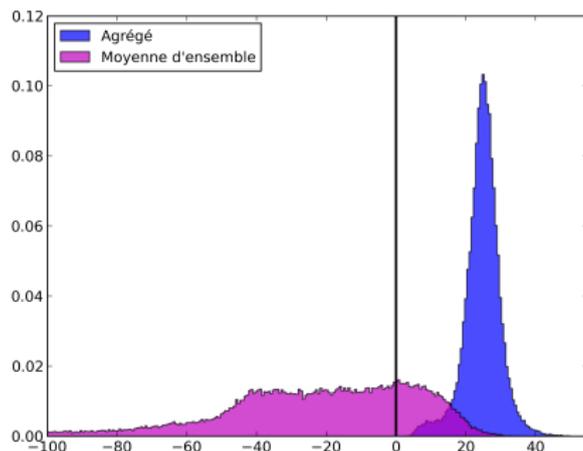


FIGURE : Histogramme de la différence relative par rapport à l'échelon déterministe du *CRPS* entre la fonction de répartition uniforme issue de l'ensemble TIGGE (en magenta) et de l'agrégé de ML-poly-grad (en bleu). Les individus représentés sont les différences relatives du *CRPS* pour l'agrégé et pour la moyenne d'ensemble. Les prévisions sont à une échéance de 6 heures, pour des paramètres optimaux. L'ensemble fait 152 membres (le déterministe Météo-France et le déterministe ECMWF sont inclus). Le *CRPS* est pris par rapport à l'échelon d'analyse Météo-France. Le poids initial est uniforme.

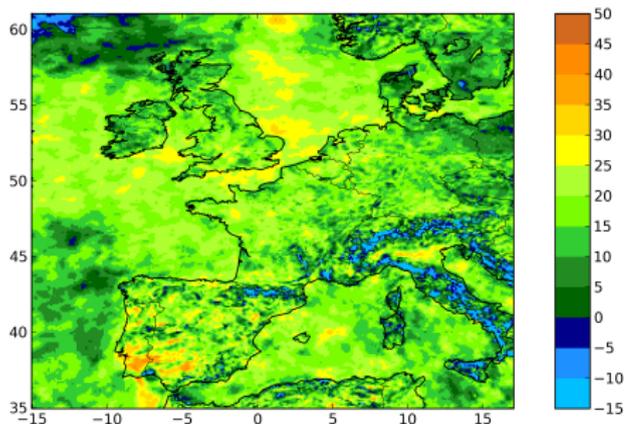


FIGURE : Représentation graphique des différences relatives du *CRPS* entre la fonction de répartition agrégée par ML-poly-grad et l'échelon déterministe Météo-France à un horizon de 6 heures. Dans chaque cellule (i, j) est représentée la différence relative. L'ensemble compte 152 membres (les déterministes d'ECMWF et de Météo-France sont compris). Le poids initial est uniforme. La période d'entraînement est de 100 jours.

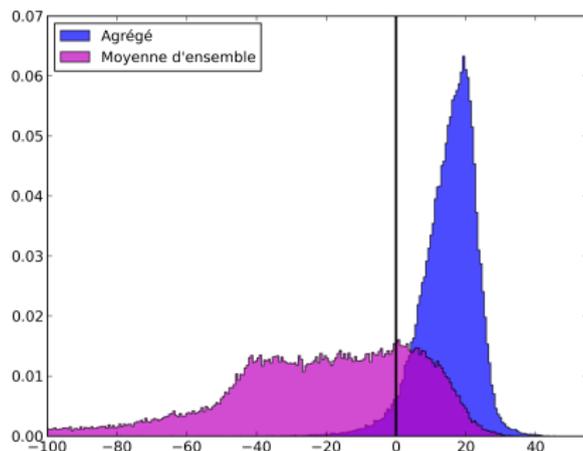


FIGURE : Histogramme de la différence relative par rapport à l'échelon déterministe du *CRPS* entre la fonction de répartition uniforme issue de l'ensemble TIGGE (en magenta) et de l'agrégé ML-poly-grad (en bleu). Les individus représentés sont les différences relatives du *CRPS* pour l'agrégé et pour la moyenne d'ensemble. Les prévisions sont à une échéance de 6 heures, pour des paramètres optimaux. L'ensemble fait 152 membres (le déterministe Météo-France et le déterministe ECMWF sont inclus). Le *CRPS* est pris par rapport à l'échelon d'analyse Météo-France. Le poids initial est uniforme.