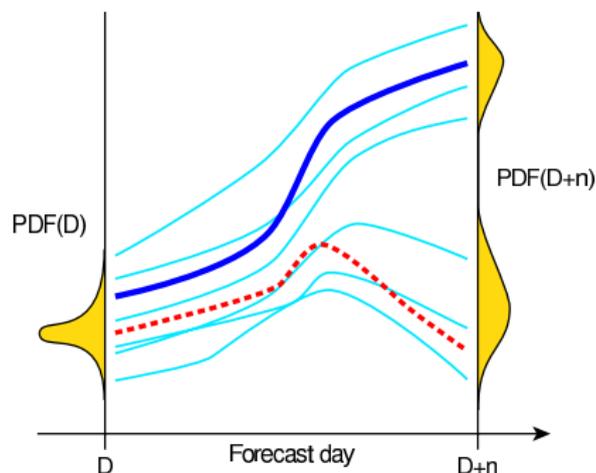


Validation probabiliste d'un Système de Prévision d'Ensemble

Guillem Candille, janvier 2006



Système de Prévision d'Ensemble (EPS) (*ECMWF Newsletter 90, 2001*)

Plan

- 1 Critères de validation probabiliste d'un système de prévision d'ensemble (EPS)
- 2 Mesures probabilistes
- 3 Comparaison "objective" de 2 EPS : méthodes bootstrap
- 4 Conclusions et perspectives

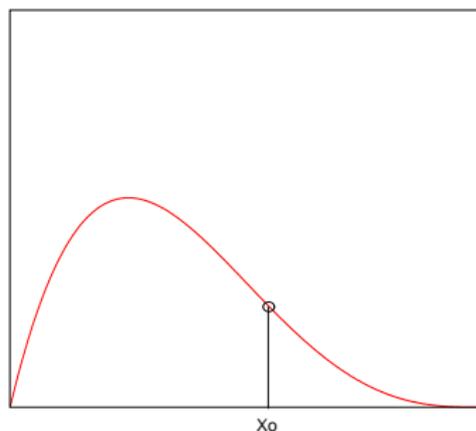
Principes probabilistes de validation d'un EPS

- "Oublier" les concepts de validation déterministe

- 1 prévision = distribution de probabilités (PDF)

vs.

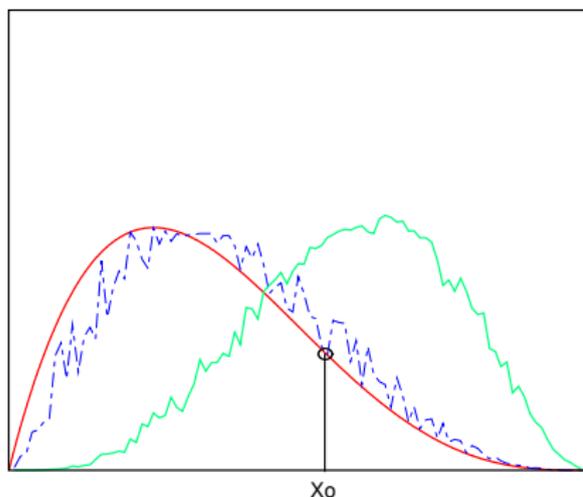
- 1 vérification = 1 valeur numérique



- Validation des EPS par **accumulation statistique**.
- 2 critères : la **fiabilité** et la **résolution**.

Critère d'évaluation : **Fiabilité**

- Cohérence statistique entre les prévisions et les observations correspondantes.

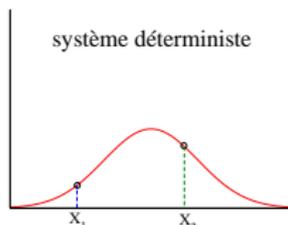
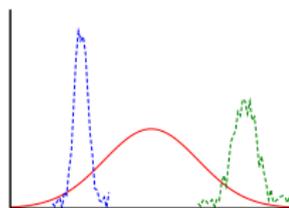
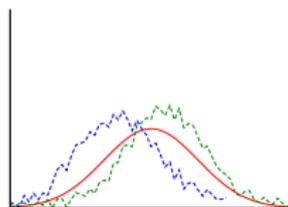
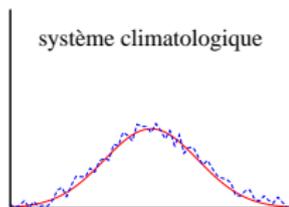


- f est la PDF prévue.
 f'_1 et f'_2 sont 2 distributions de x_0 quand f est prévue.
 L'EPS est parfaitement fiable si et seulement si $f = f'$.
- tout système peut être rendu fiable par une calibration statistique *a posteriori*.

- La fiabilité est un critère nécessaire mais non suffisant :
cf. e.g. un système climatologique.

Critère d'évaluation : Résolution

- Capacité de séparer les prévisions menant à des distributions d'observations correspondantes (DOC) bien distinctes.



- Exemples DOC allant d'une résolution minimale nulle pour le système climatologique à une résolution maximale pour le système déterministe.

(la courbe **rouge** représente ici la distribution de toutes les observations : distribution climatologique)

Critères probabilistes : résumé

- La fiabilité et la résolution sont 2 propriétés qui déterminent la qualité et l'utilité d'un système de prévision probabiliste.
- Un EPS doit être capable de séparer *a priori* les prévisions en classes suffisamment variées afin que les observations correspondantes représentent des situations bien distinctes.

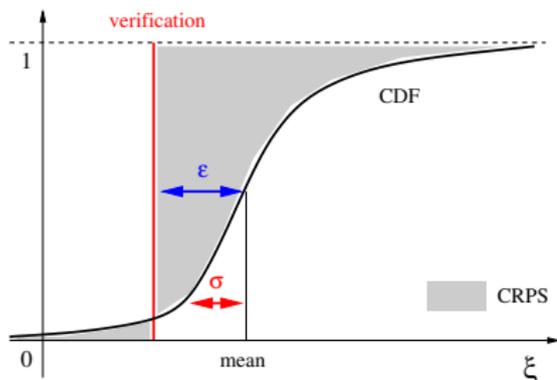
Continuous Ranked Probability Score (CRPS)

- Le CRPS mesure la qualité globale d'un EPS :

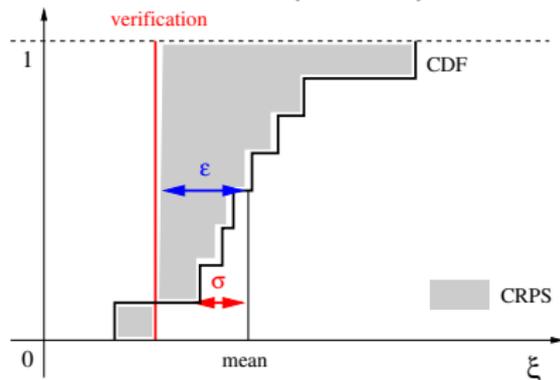
$$CRPS = E \left[\int_{\Omega} (F_p(\xi) - H(\xi - x_o))^2 d\xi \right] \quad (1)$$

F_p est la distribution cumulative de probabilités (CDF) associée à l'ensemble prévu et H est la distribution de *Heaviside* associée à la vérification.

cas continu



cas réel (discret)



Décomposition du CRPS

- Le CRPS se décompose en 2 composantes permettant de mesurer la fiabilité et la résolution.
- On utilise ici la décomposition proposée par H. Hersbach¹ :

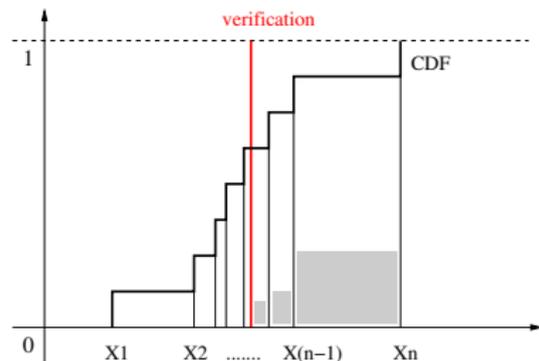
$$CRPS = Reli + CRPS_{pot} \quad (2)$$

où *Reli* mesure la fiabilité et $CRPS_{pot}$ mesure la résolution.

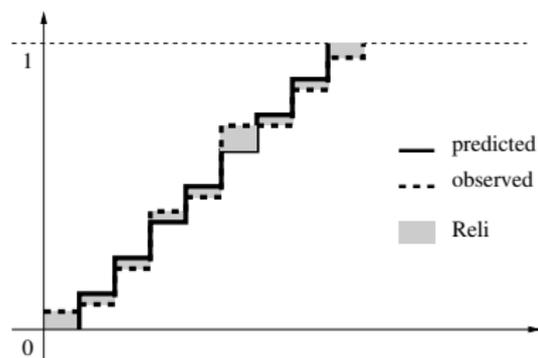
¹Hersbach 2000 : *Decomposition of the continuous ranked probability score for ensemble prediction systems*, in *Weather and Forecasting* (15), pp 559–570.

Composante fiabilité du CRPS : *Reli*

- A chaque réalisation du système, on définit des valeurs pour chaque intervalle $[x_i, x_{i+1}]$ selon la position de la vérification et la taille de l'intervalle.

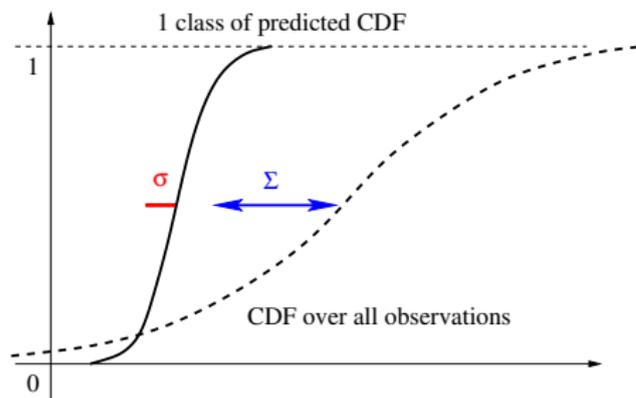


- Après accumulation sur l'ensemble des réalisations du système, on peut définir la CDF observée correspondante et la comparer à la moyenne des CDF prévues.



Composante résolution du CRPS : $CRPS_{pot}$

- $CRPS_{pot}$ serait la valeur potentielle du CRPS si l'EPS était fiable, *i.e.* $Reli = 0$.
- $CRPS_{pot} = uncertainty - Resol$



- $CRPS_{pot} \propto \mathcal{F}(\sigma)$
- $uncertainty \propto \mathcal{F}(\Sigma)$
- $Resol \propto \mathcal{F}(\Sigma - \sigma)$

- Plus $\sigma \ll \Sigma$, meilleure la résolution est.

Reduced Centered Random Variable (RCRV)

- La RCRV permet de détailler les caractéristiques de la fiabilité en terme de biais (b) et dispersion (d).
- A chaque réalisation, on considère la moyenne m et la dispersion σ de l'ensemble, ainsi que l'observation o et l'erreur d'observation σ_o . La RCRV se définit alors :

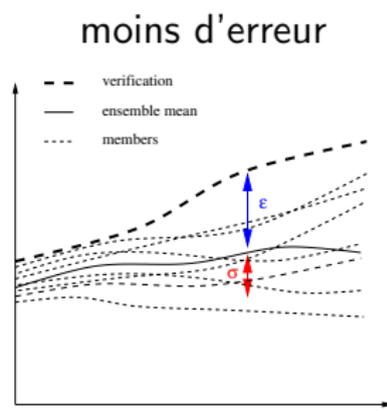
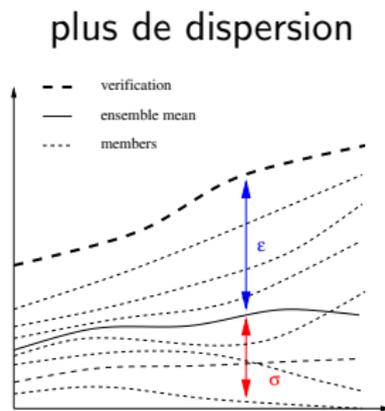
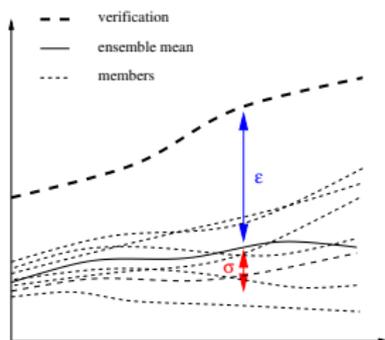
$$y = \frac{o - m}{\sqrt{\sigma^2 + \sigma_o^2}} \quad (3)$$

- On cherche ainsi à vérifier si l'observation peut être considérée comme une réalisation indépendante de la loi de probabilité définie par l'ensemble
→ la moyenne de y (b) mesure le biais pondéré de l'ensemble et la variance de y (d) mesure la dispersion comparée des ensembles et des observations.

Comment améliorer la fiabilité d'un EPS ?

- Pour un EPS fiable, on a : $b = 0$ et $d = 1$
- On considère un EPS non fiable sous-dispersif et 2 corrections possibles :

$$d \equiv \text{Var} \left(\frac{\epsilon}{\sigma} \right) > 1$$



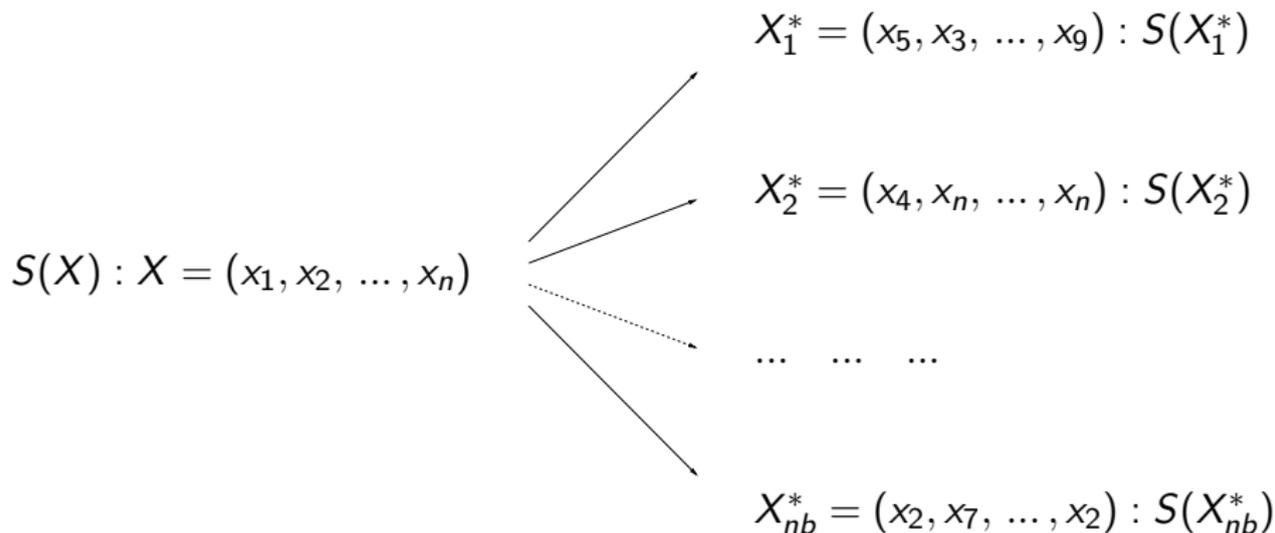
Méthode de comparaison

- Vérifications effectuées avec des observations de radio-sondages réparties sur tout le globe (Chantal Côté).
- Comparaison des performances de 2 EPS mesurée par les différences en terme de CRPS et RCRV.
- Limitation de la validation due à la taille de l'échantillon sur lequel les statistiques sont calculées
→ estimation d'intervalle de confiance par des techniques de ré-échantillonnage *bootstrap*².

²Efron and Tibshirani, 1993 : An introduction to the bootstrap 

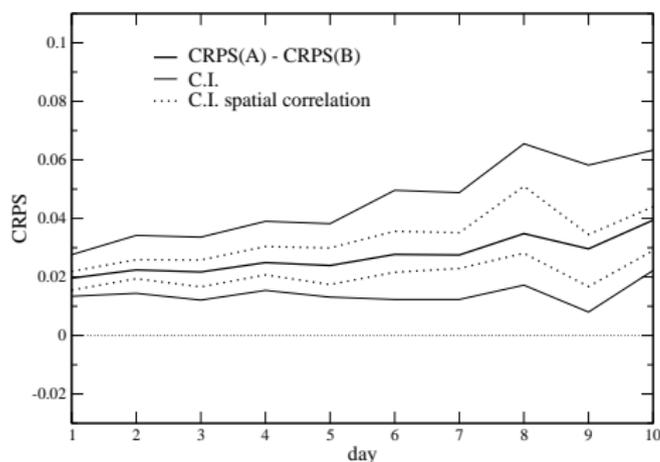
Technique bootstrap

- S : score CRPS ou RCRV
- X : échantillon de données



Intervalle de confiance : propriétés (1)

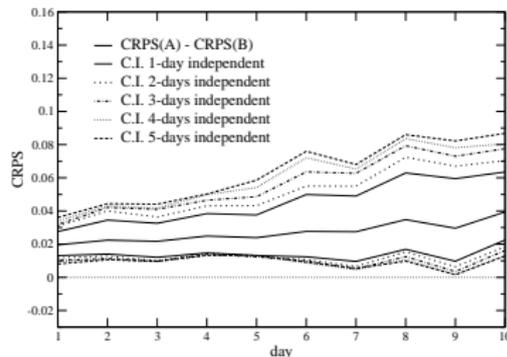
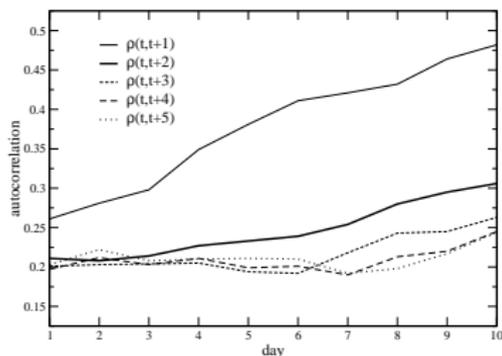
- Le ré-échantillonnage s'effectue seulement parmi les jours de prévision et non aussi parmi les différentes stations d'observations
 → cela évite de donner trop de poids à des situations "extrêmes" journalières et d'ainsi sous-évaluer la taille de l'intervalle de confiance.



- Température à 850 hPa pour la période du 12 septembre au 5 novembre 2005
- EPS A = EPS opérationnel
- EPS B = EPS parallèle

Intervalle de confiance : propriétés (2)

- Risque de sous-évaluation de la taille des intervalles de confiance à cause de l'auto-corrélation temporelle des erreurs de prévision, surtout à moyenne échéance.

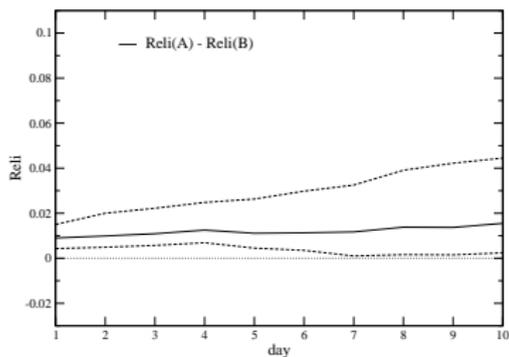


- Les intervalles de confiance traduisent les incertitudes liées aux auto-corrélations des erreurs de prévision.
- Calcul des intervalles de confiance en ré-échantillonnant par séquences de 3 jours.

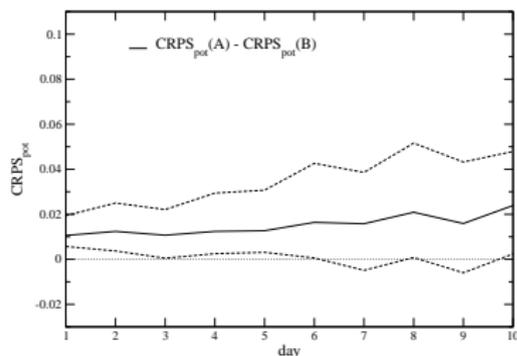
Comparaison : décomposition du CRPS

- Amélioration globale (CRPS) pour l'EPS parallèle.

Reli



CRPS_{pot}

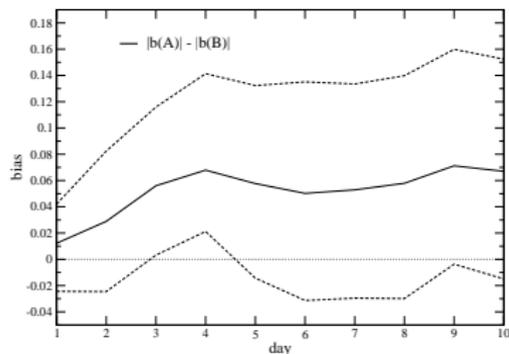


- Amélioration significative de la fiabilité.
- Amélioration significative de la résolution jusqu'au jour 6.

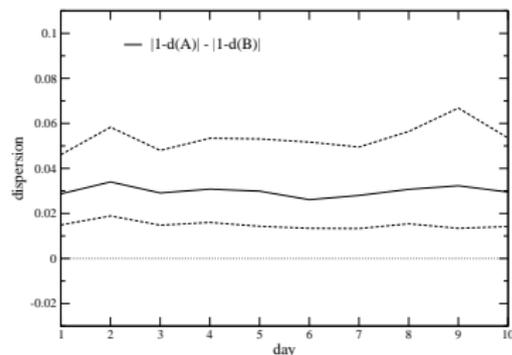
Comparaison : décomposition de la fiabilité

- Amélioration significative de la fiabilité pour l'EPS parallèle.

biais



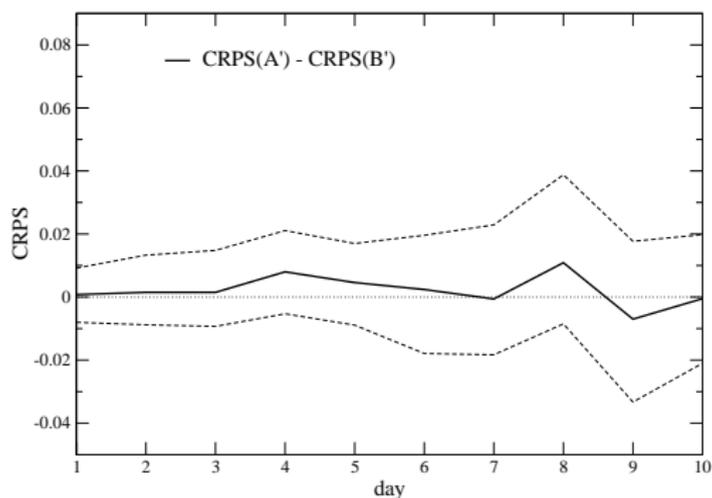
dispersion



- Amélioration non significative du biais.
- Amélioration significative de la dispersion.

Comparaison : changement neutre

- Construction de 2 EPS thériquement équivalents :
 - EPS A' = EPS oper. jours pairs + EPS para. jours impairs
 - EPS B' = EPS oper. jours impairs + EPS para. jours pairs



- Pas de différence significative.

Résumé

- Développement d'outils de validation probabiliste "objective" pour la comparaison de la performance de 2 EPS :
 - vérification contre observations.
 - mesure globale CRPS avec décomposition fiabilité-résolution et caractérisation de la fiabilité par la RCRV.
 - définition d'intervalle de confiance par des techniques de ré-échantillonnage bootstrap.
- Amélioration significative avec l'EPS parallèle, surtout en terme de fiabilité, qui est devenu l'EPS opérationnel le 12 décembre 2005.

La suite ...

- Transfert des outils de validation/comparaison probabiliste en mode opérationnel
 - inclus dans une "boite à outils" de vérification plus générale (Barbara Casati).
- Développer les techniques de ré-échantillonnage pour d'autres mesures probabilistes telles que le score de Brier.
- Préviation déterministe et préviation d'ensemble :
 - liens entre les incertitudes de préviation déterministe et la dispersion des ensembles ?
 - liens entre la discrimination (courbe ROC, valeur économique) et la fiabilité-résolution ?

Merci

Questions ?