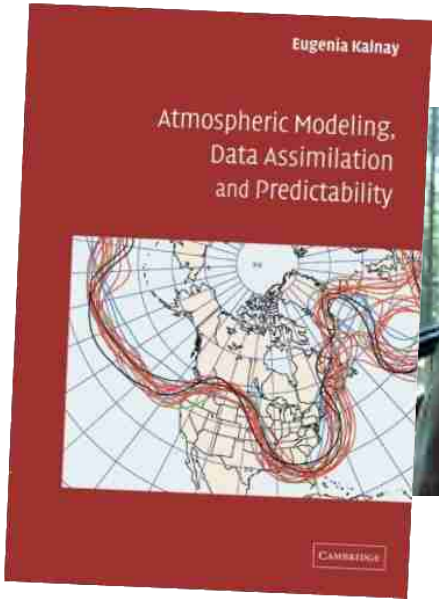
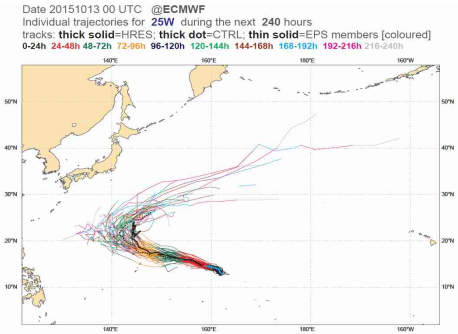
Prévision d'ensemble : motivation

- calculer les modèles d'erreurs utilisés en assimilation de données (matrice B, simplification du filtre de Kalman)
- mettre des intervalles de confiance sur les prévisions climatiques et météo
- alimenter les prises de décision météosensibles (optimisation financière, marché de l'énergie)
- gérer et exprimer les risques en temps réel (ex: vigilance, "storylines")
- comprendre lla propagation des erreurs dans les prévisions (scénarios)



Eugenia Kalnay (Argentine)



Edward Lorenz (MIT)

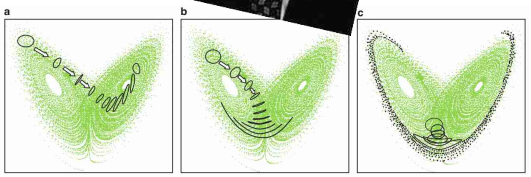
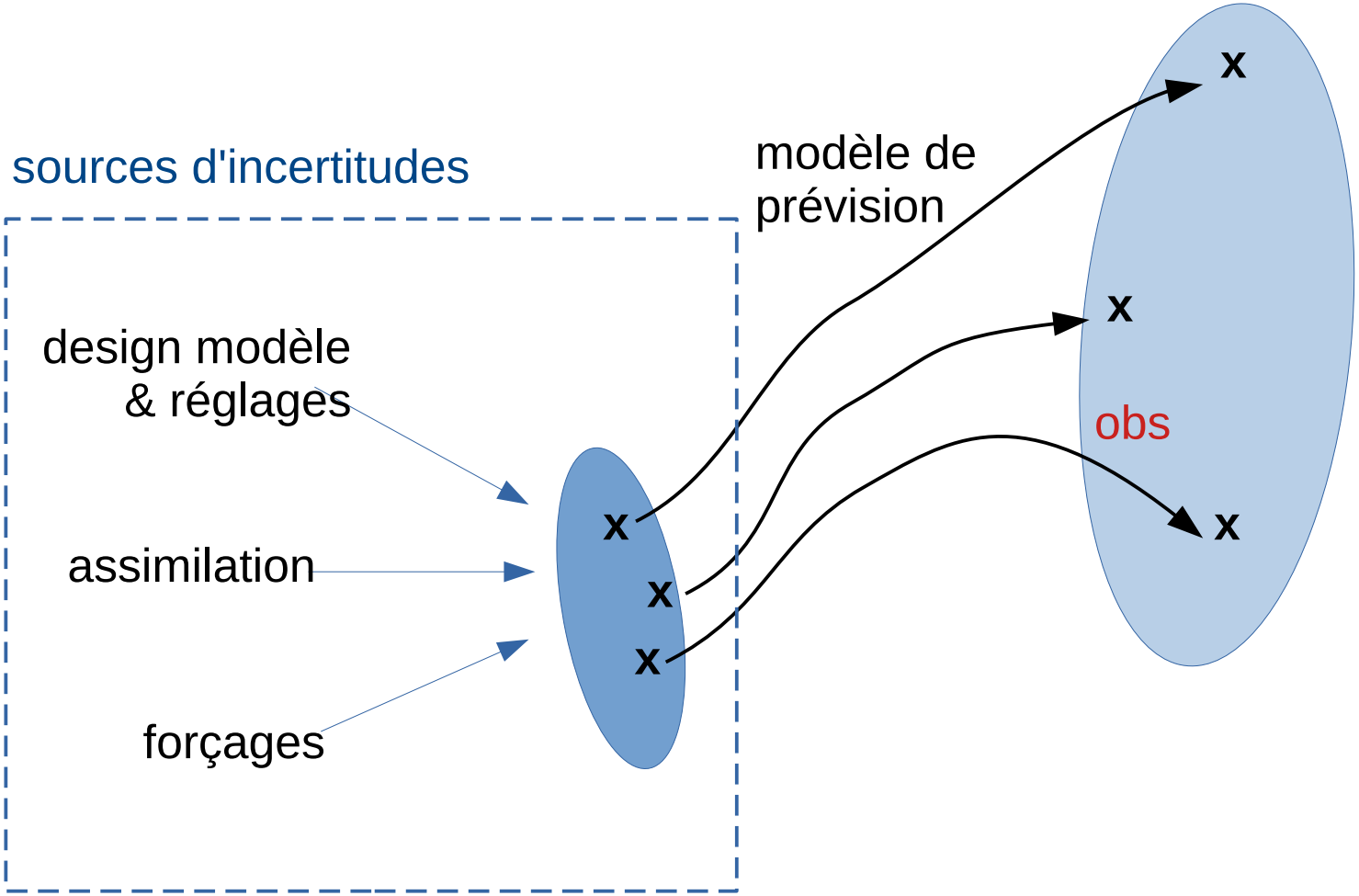


Figure 1 Scientific basis for ensemble forecasting. In a nonlinear system the growth of initial uncertainty is flow dependent – here illustrated with the Lorenz (1963) model. The set of initial conditions (black circle) is located in different regions of the attractor in (a), (b) and (c).

Prévision d'ensemble : concept

déf : un ensemble est une *représentation* ***discrète*** des ***incertitudes*** *sur* ***notre connaissance d'un système physique***



Prévision d'ensemble : définition

= méthode Monte Carlo, échantillonnage discret, quantification des incertitudes, propagation des erreurs, modélisation stochastique, prévision probabiliste, logique floue,...

déf : un ensemble est une *représentation* ***discrète*** des ***incertitudes*** *sur* ***notre connaissance d'un système physique*** (ici, un état de modèle discrétisé)

- **"discrète"** = on modélise une distribution (possiblement continue) par un 'peigne' (= somme de distributions de Dirac)
- **"incertitudes"** = distribution des états plausibles (= éventail des possibles)
- "**notre connaissance**" = l'ensemble des informations dont on dispose à un instant *t*

Attention à la différence :

- *une prévision déterministe caractérise une réalité physique (indépendante de nous)*
- *une prévision d'ensemble caractérise un système de prévision (que nous avons construit)*

Prévision d'ensemble : technique de base

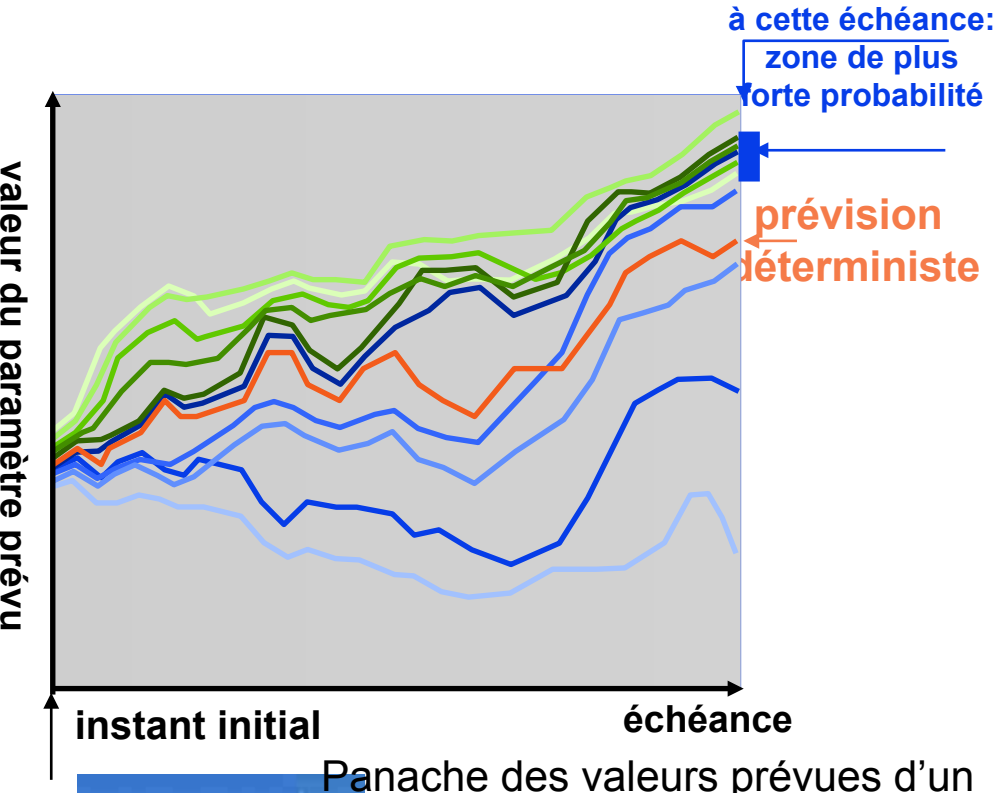
But : produire une prévision stochastique qui reflète nos incertitudes

Prévision d'ensemble: échantillonnage discret de la loi de probabilité de prévision. On veut lancer *q* prévisions perturbées ("**membres**") en parallèle (*q*~50)

Exemple: avec une matrice de covariances d'erreurs $A=L\,L^T$, si l'on tire aléatoirement *u* suivant une loi Gaussienne, alors la prévision perturbée $v=x_a + Lu$ aura la loi de probabilité impliquée par A.
Si A est la covariance d'erreurs d'analyses, alors $w_i=M(v_i)$ suivra à peu près la loi de probabilité des erreurs de prévision, et $B \sim \sum w_i\, w_i^T$

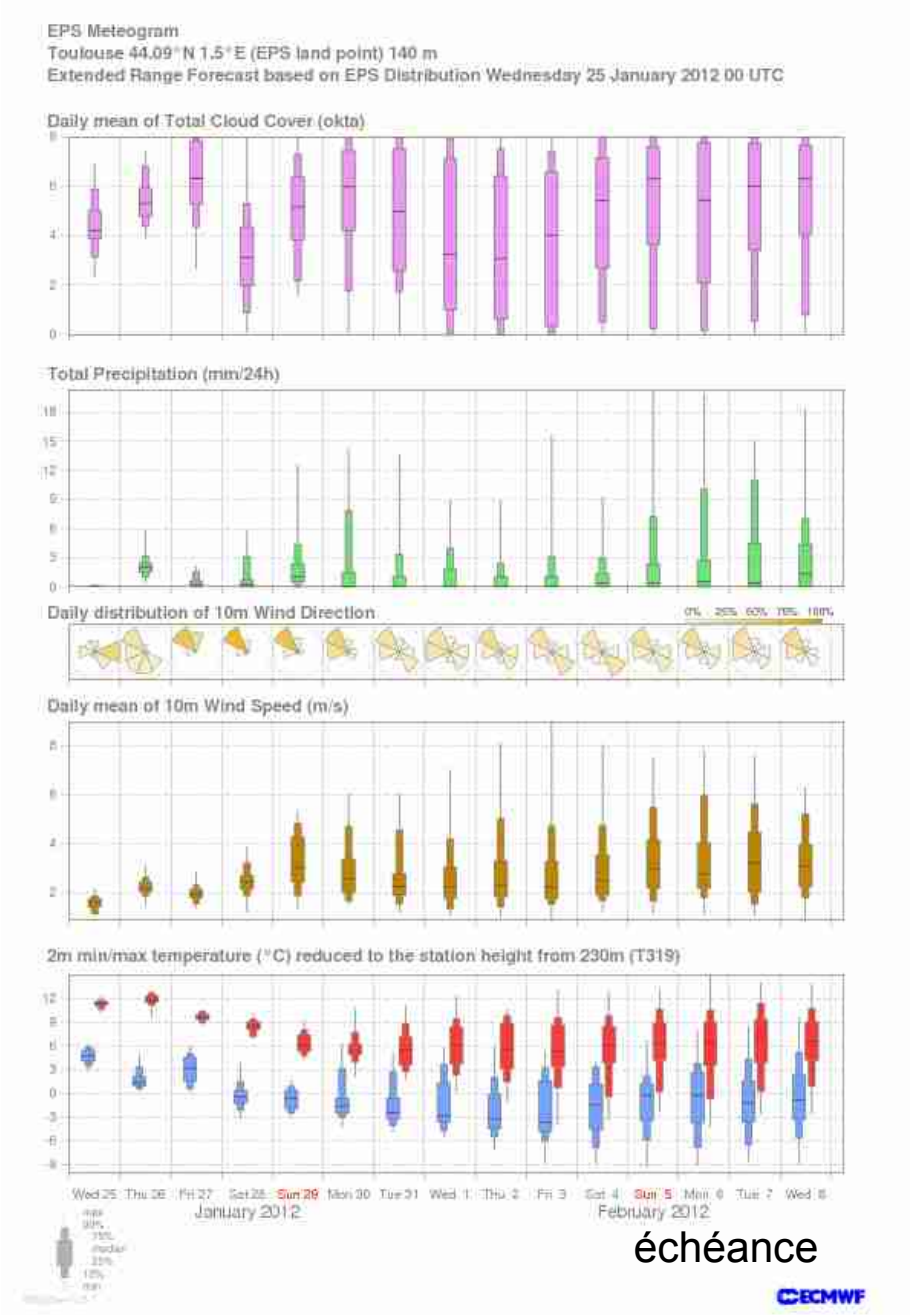
Le calcul d'un petit ensemble $M(v_i)$, *i=1...q* coûte *q* fois l'exécution d'un modèle de prévision.

prévisions d'ensemble point par point



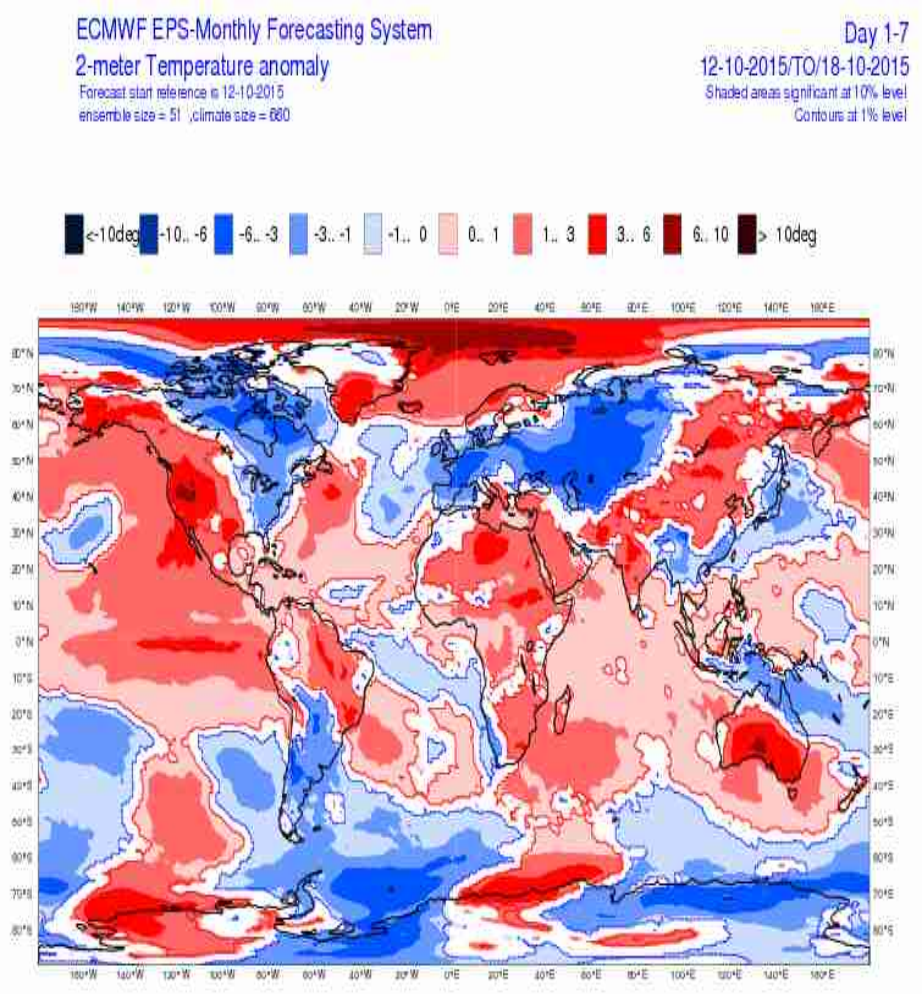
Panache des valeurs prévues d'un paramètre en un point pour différentes échéances

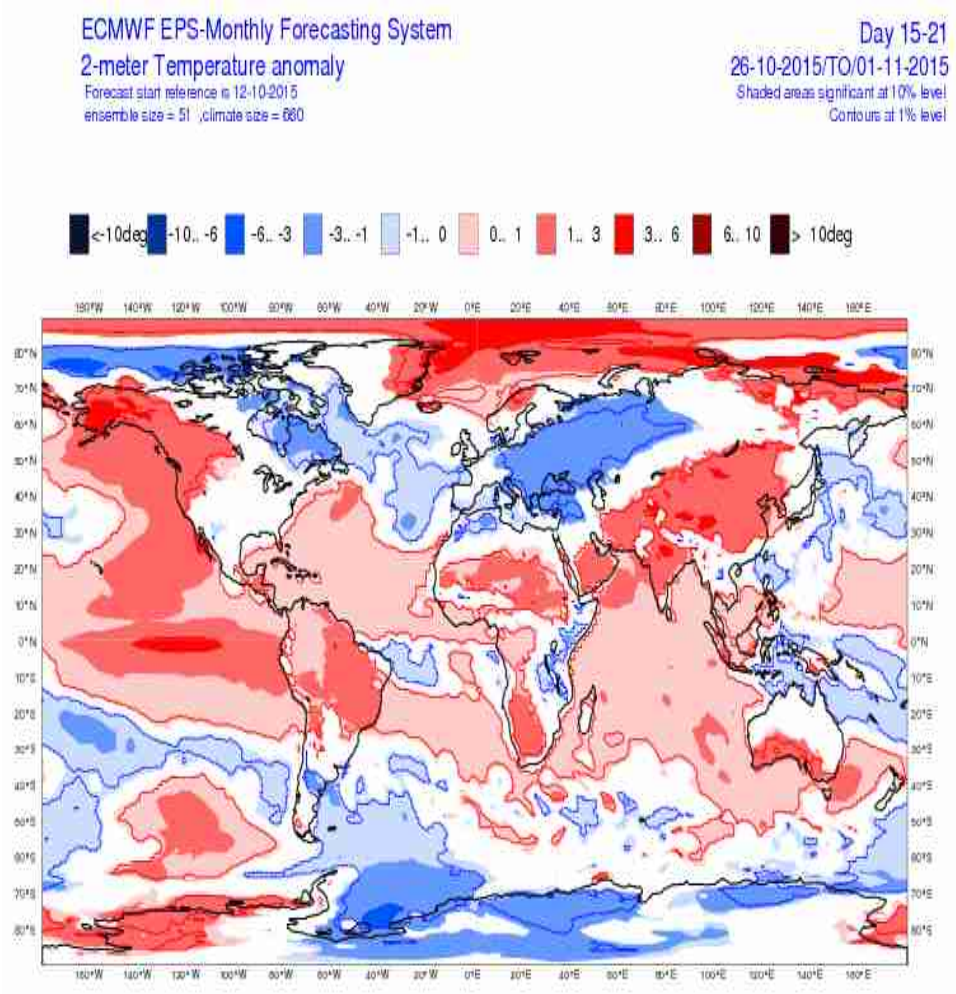
scénario privilégié pour grand public



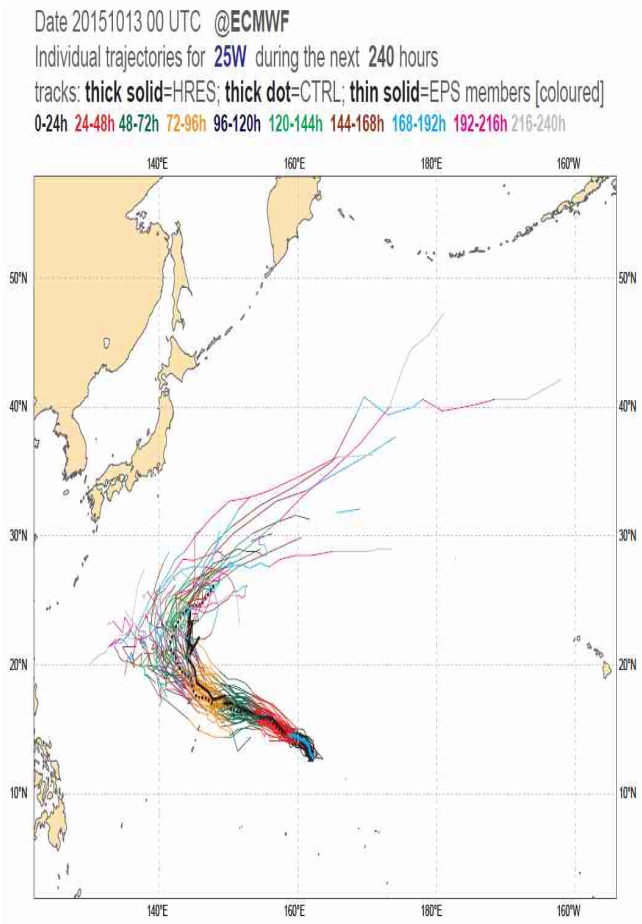
Prévision climatique (ici mensuelle) à faible signal

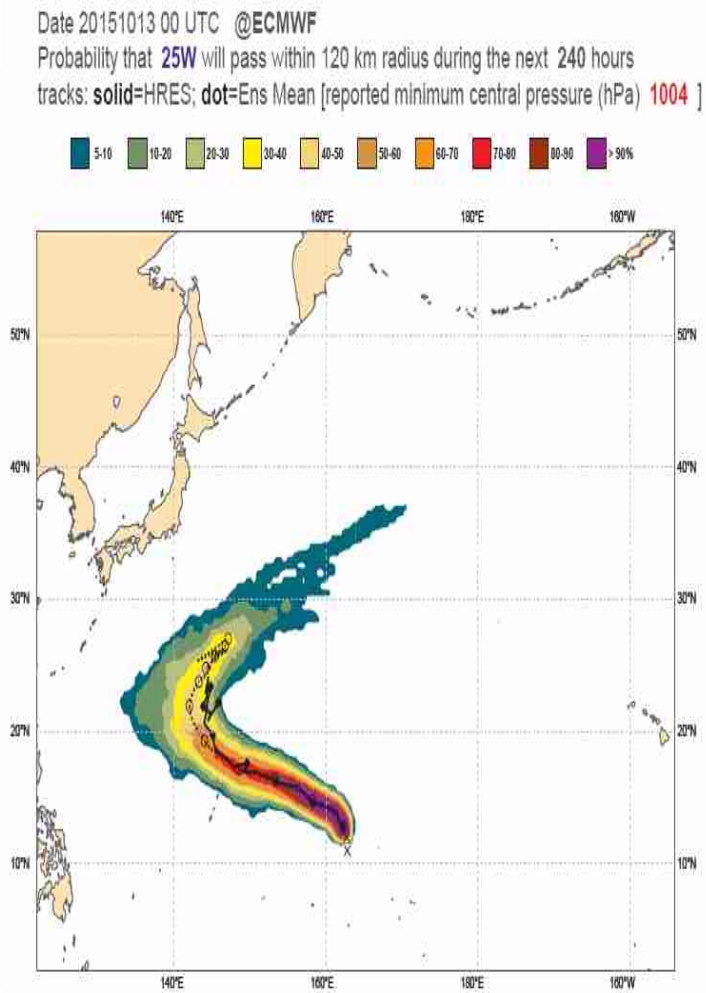
on représente en chaque point si la distribution (PDF) de l'ensemble est significativement > ou < au climat, ou pas

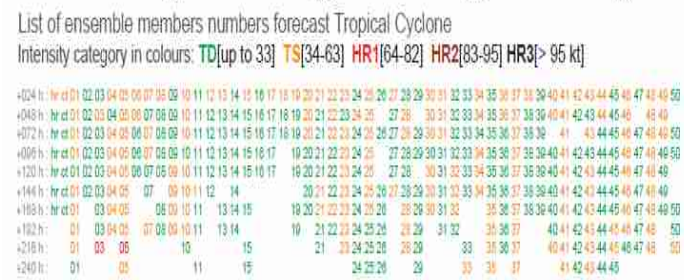


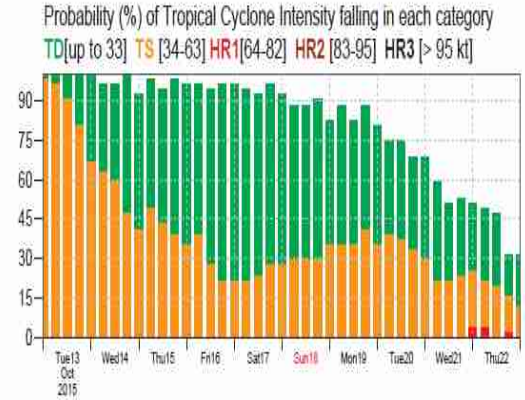


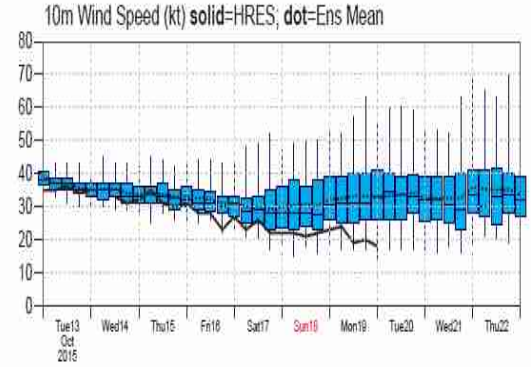
prévision d'ensemble par objet : cyclones

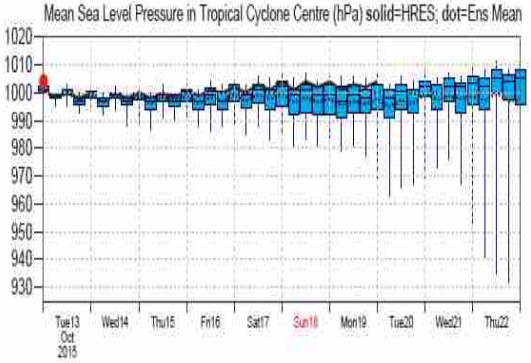








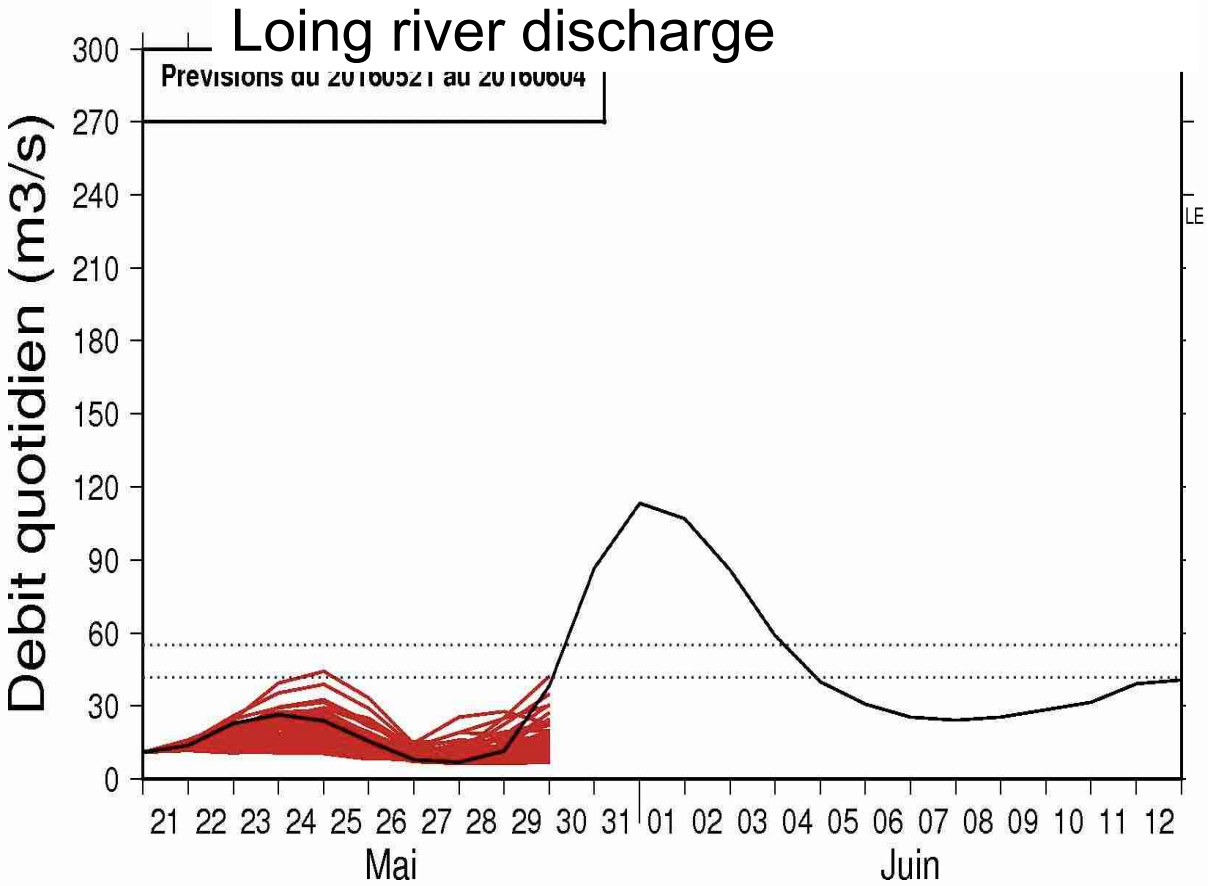




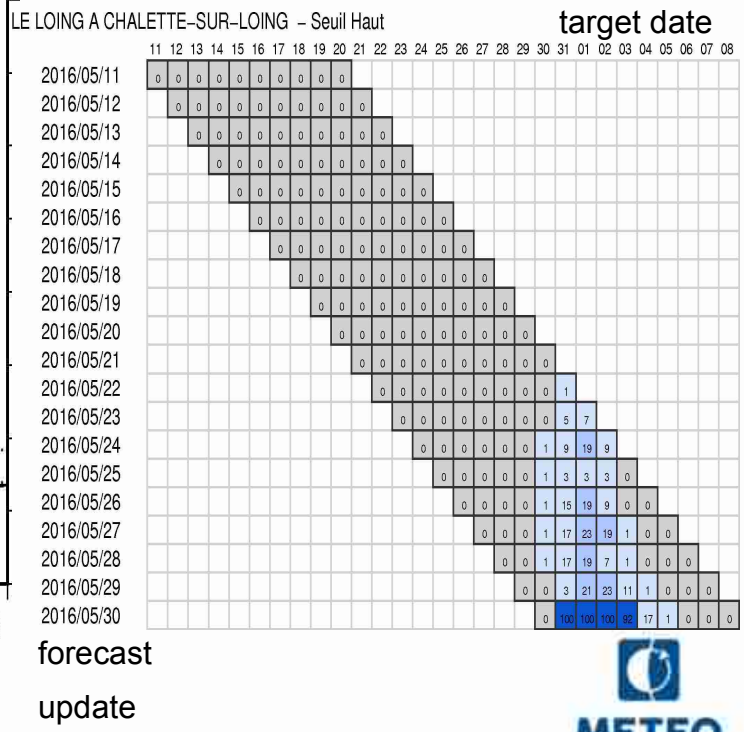
Prévision d'ensemble par scénario : crues

flood forecasts, 10 days ahead

hydrological model forced by atmospheric ensemble



Location-specific impact forecasts



Prévision d'ensemble par scénario : crues

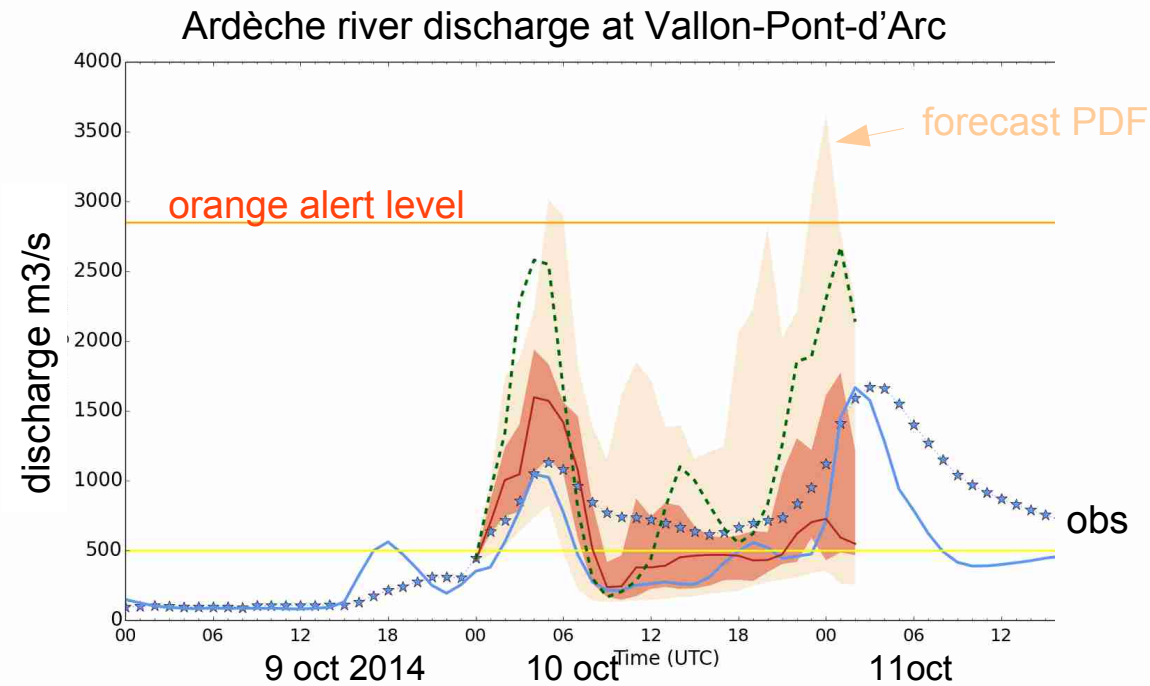
various hydrological models, forced by atmospheric ensemble

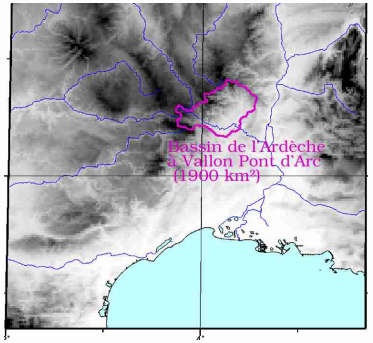
Arome-EPS
meteorological
ensemble prediction

rainfall ensemble

ISBA-TOP
surface-hydrological
flash flood
ensemble prediction

perturbations to parameters
of the hydrological model





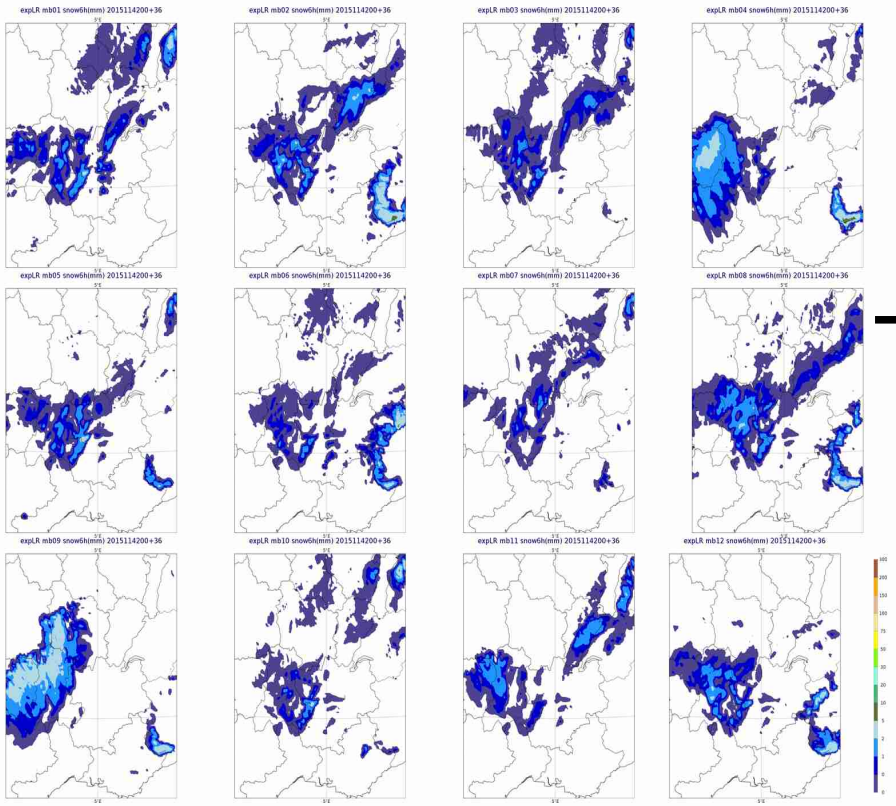
measured discharge
deterministic ISBA-TOP simulation driven by radar
deterministic ISBA-TOP driven by Arome-France
ISBA-TOP ensemble median, driven by Arome ensemble
ISBA-TOP ensemble 25-75 % quantiles
ISBA-TOP ensemble min-max

METEO FRANCE

comment résumer un ensemble de champs ?

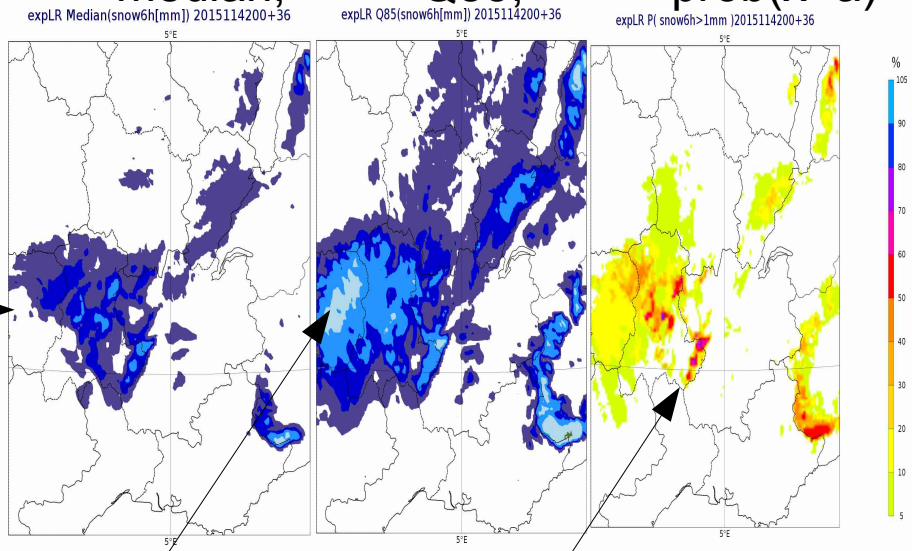
(~*O(100)* prévisions vraisemblables disponibles à chaque instant)

12 members of Arome ensemble prediction (snowfall)



probability maps :

median, Q85, prob(x>a)



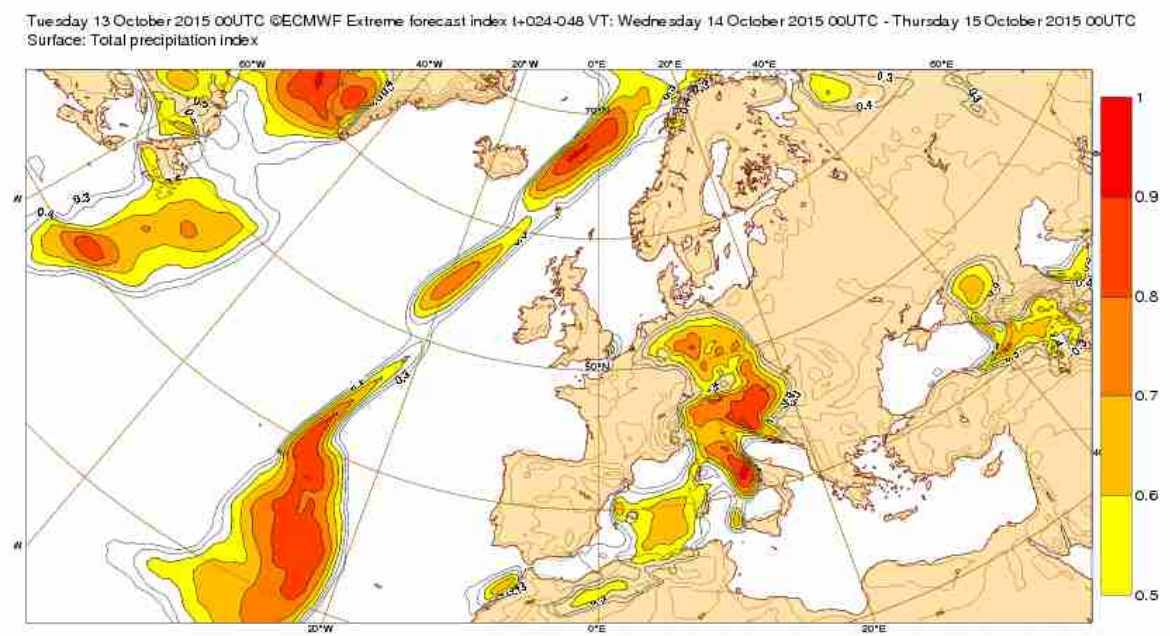
small probability of heavy snow

high probability of light snow

résumé des risques d'événements extrêmes :

probabilité que la prévision soit dans un quantile extrême du climat en chaque point

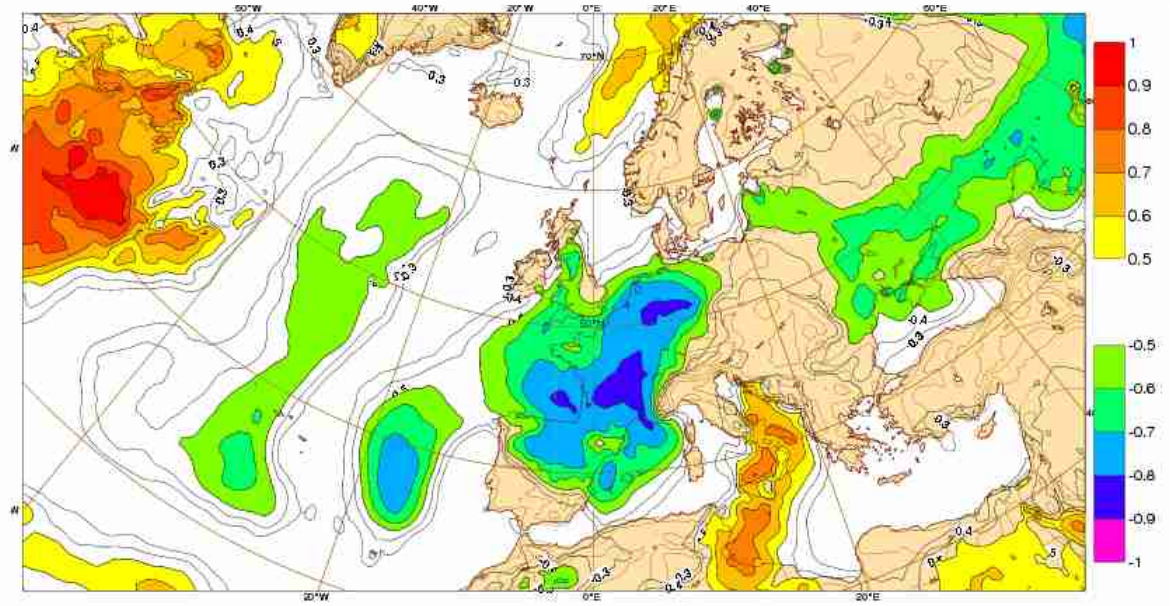
pluies



"extreme forecast index"

for human interpretation

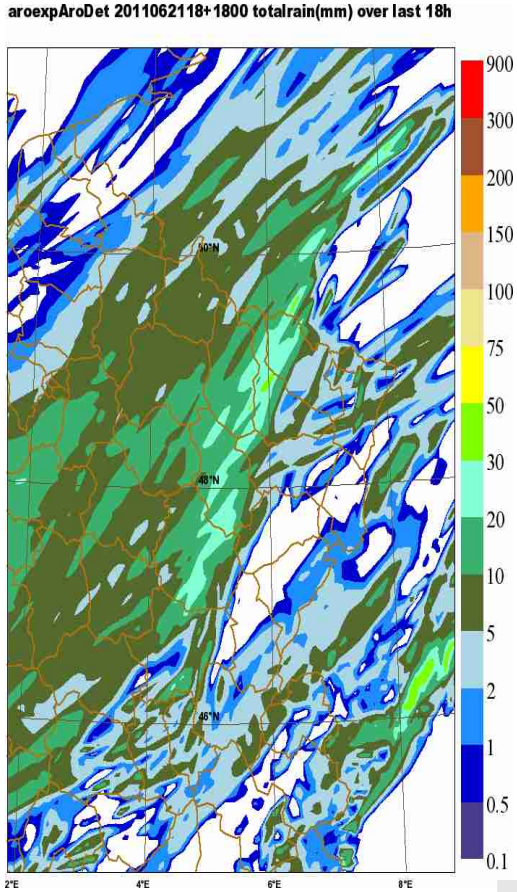
température



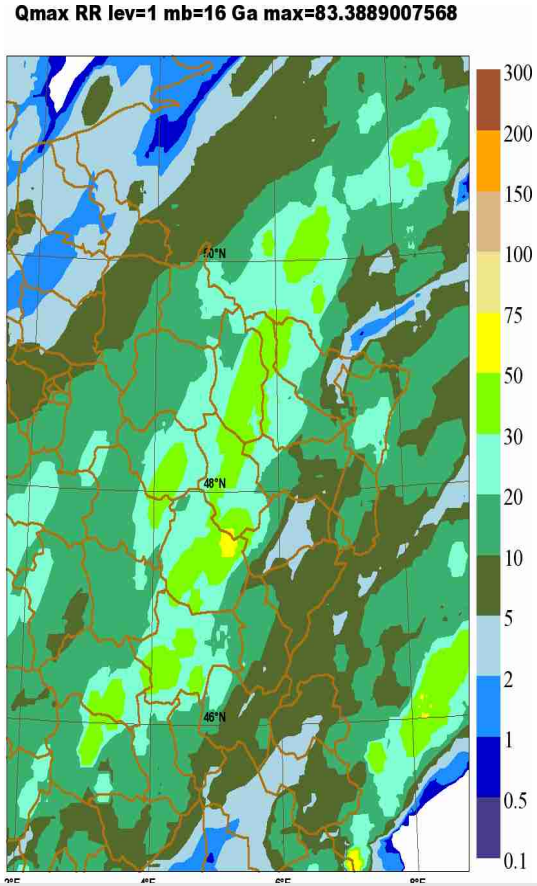
Les limites de la représentation des incertitudes par probabilités

(événements intenses à localisation & timing incertains)

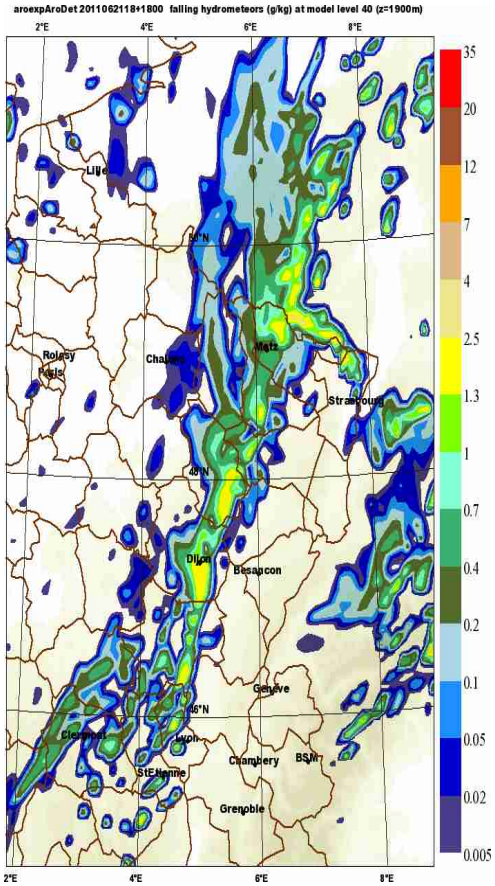
cumul sur 18h déterministe



cumul sur 18h proba>10%



précip instantanée



quantile Q90 de la distribution de la prévision d'ensemble

quel est le meilleur quantile qui résume cette prévision d'ensemble ?

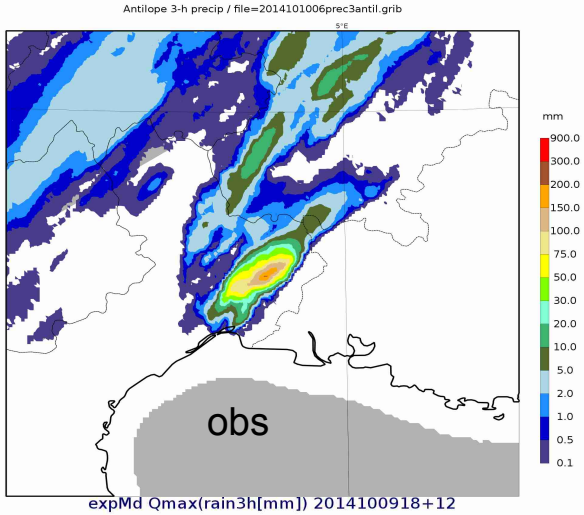
case study :

small-scale intense raining cells with high spread

location/timing uncertainties can hide probabilities

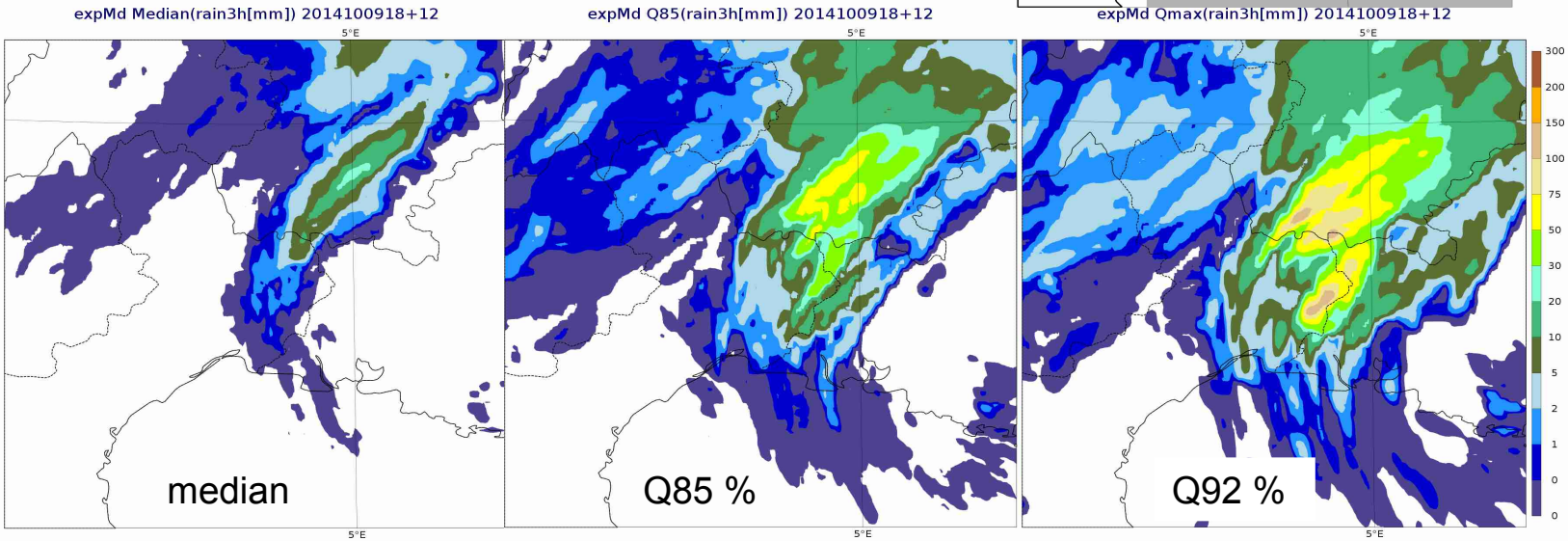
decision-making needs to choose a cost/loss ratio (~false alarm tolerance)

the most likely forecast is not the most useful !



Three points of view

on the same rain event :

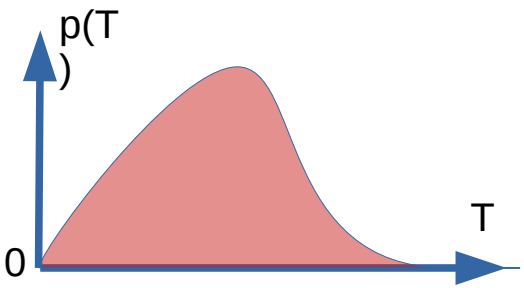


Représentation pratique des incertitudes

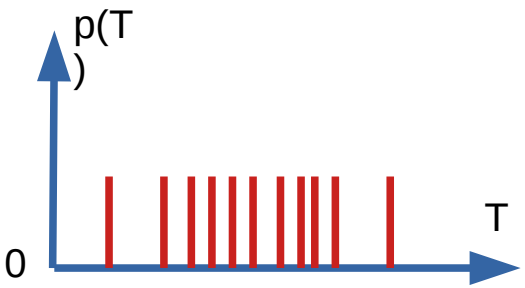
- statistiques de base : PDF d'un réel
- variables binaires, réelles, scénarios
- pièges dans l'utilisation des probabilités
- techniques de perturbation
- validation et réglage des ensembles

Représentation pratique des incertitudes

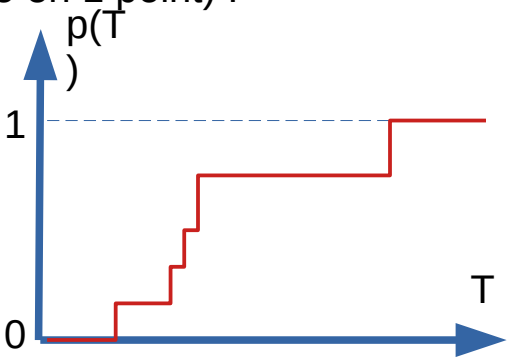
Exemples : incertitude d'un **paramètre réel** (ex : température en 1 point) :



une fonction de distribution continue (PDF : probability density function)



une PDF discrète



une PDF discrète représentée par son intégrale (CDF : cumulative density function)

1 marche par membre

Quelques "statistiques" càd des manières de résumer une PDF :

- moyenne
- min, max
- quantile Qx : valeur de T telle que x % de la distribution est en-dessous
- médiane Q50 : valeur de T telle que 50 % est en-dessous
- probabilité de dépasser un seuil t : P(T>t) = intégrale (surface) de la PDF à droite de t

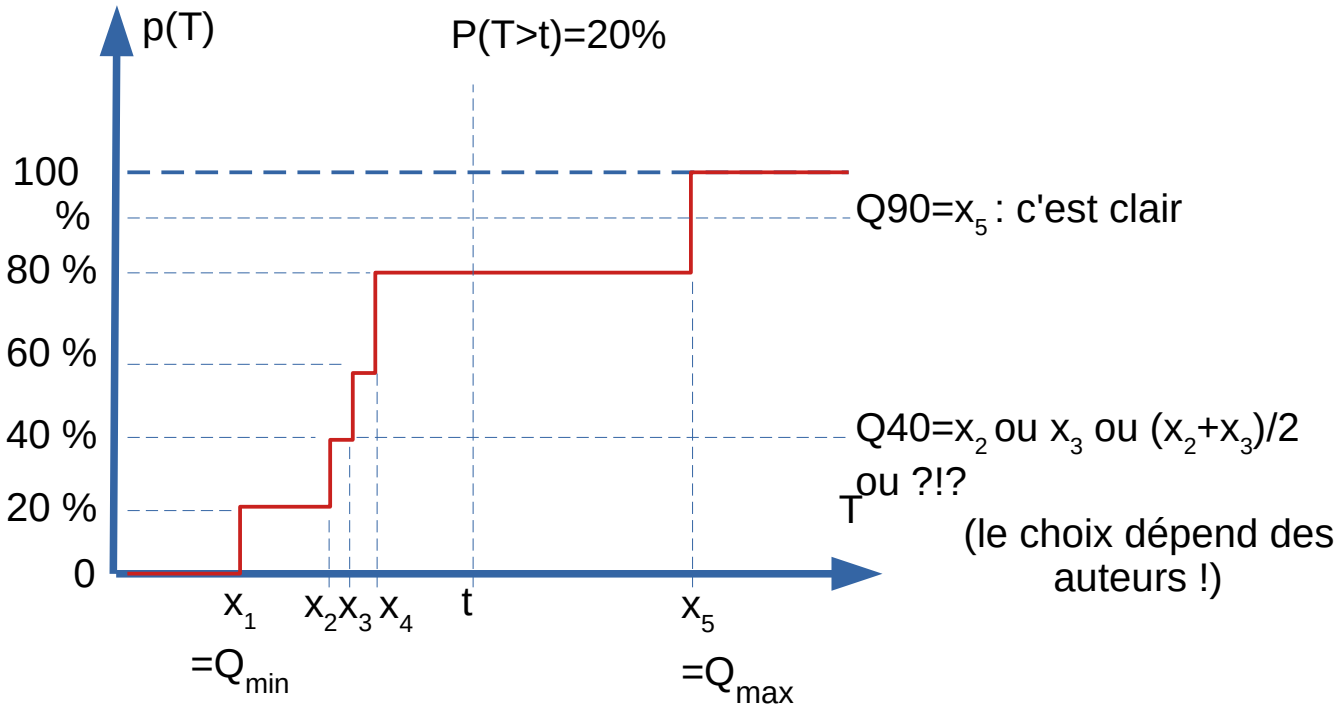
La proba P(T>Qx) de dépasser un quantile Qx est donc 1-x : 10 % de chances d'être au-dessus du Q90.

Représentations pratiques des incertitudes

Toujours pour un **paramètre réel** : repérage des statistiques sur un graphique de CDF ensembliste

exemple d'ensemble à n membres : (pas forcément par ordre croissant)

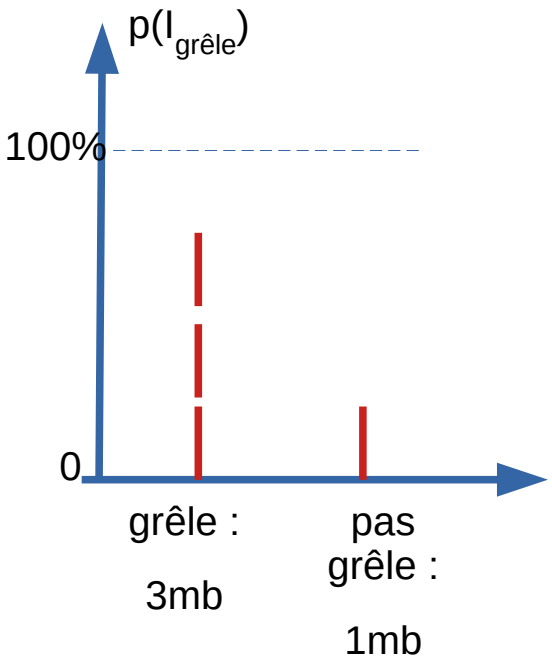
- moyenne $m(x) = (x_1 + ... + x_5)/n$
- écart-type = $\mathrm{sqrt}\{ 1/(n-1) \ \mathrm{sum}[(x - m(x))^2] \}$
- certain(e)s définissent les quantiles différemment



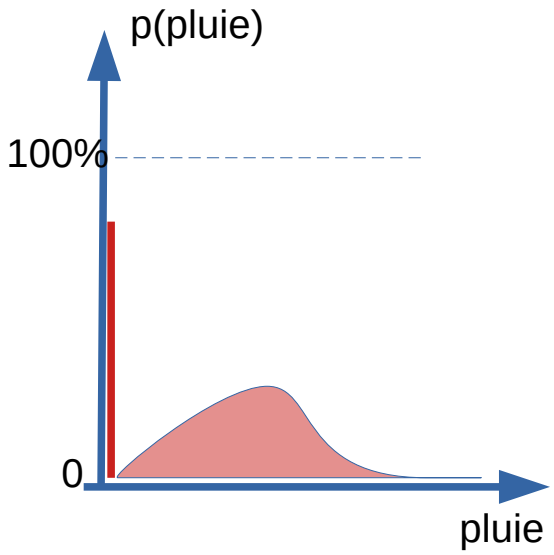
Représentations pratiques des incertitudes :

On a illustré un paramètre réel et continu (ex: température en un point)

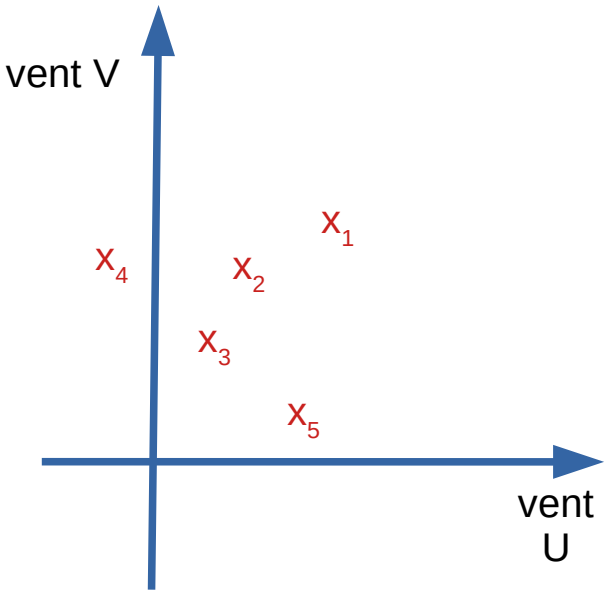
Mais songer aussi à :



un paramètre binaire



une distribution hybride et bornée



un vecteur à 2 dimensions

(moyenne ok, mais pas de proba ni de quantile...)

...voire **un champ** :
vecteur à n dimensions

Trois pièges courant des probabilités

1) **Erreur d'échantillonnage :** pour estimer (à peu près) précisément

- une moyenne, il faut *n>>10* réalisations (= données) indépendantes
- un écart-type, *n>>30*
- un quantile, *n>>30* au-dessus et en dessous du quantile visé

n doit inclure des points et instants aux erreurs non corrélées.

Faire des tests (ex: **bootstrap**), se méfier des stats sur un seul cas d'étude

2) **The curse of dimensionality :** pour estimer avec un ensemble une densité de probabilité (notamment, l'état le plus probable) d'un vecteur d'état de dimension m, il faut *>>10m* membres

- **rarement ok pour des champs météo ou des scénarios** (m= plusieurs millions/milliards)
- mais on peut tenter sur des diagnostics de dimension 1 ou 2, maximum

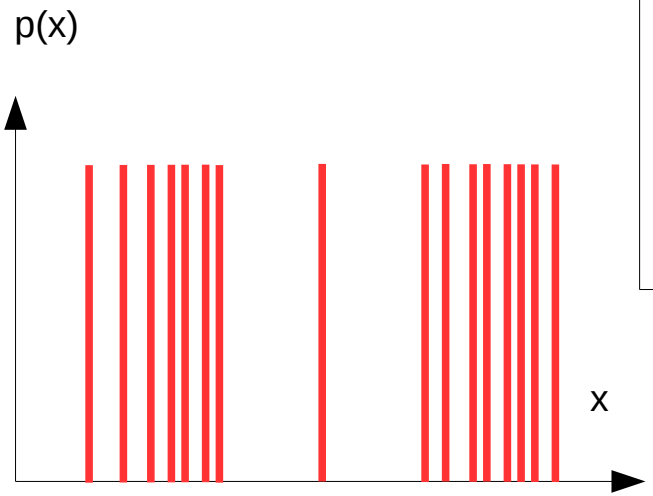
3) **Non-additivité des probas et quantiles** : *"la proba de la somme n'est pas la somme des probas"*

- si j'ai 1/6 de chance de faire un 6 avec un dé, quelle est la proba d'avoir cinq 6 avec 5 dés ?

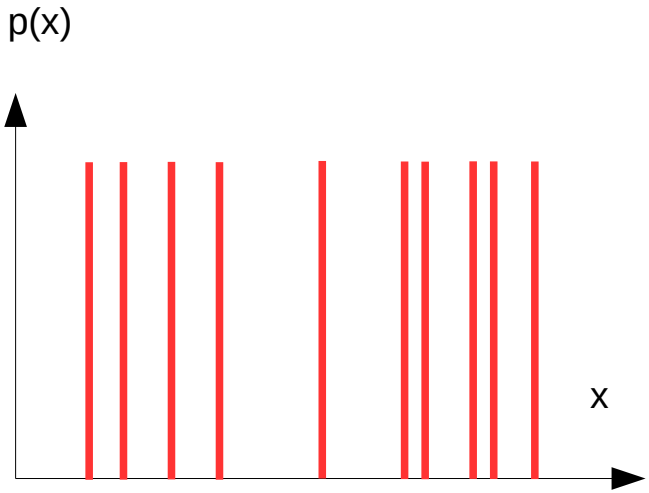
- de même, si on a 5 % de chances d'avoir de la pluie pendant chaque heure entre 0h et 6h, quelle est la proba d'avoir de la pluie>0 entre 0h et 6h ? ***on ne peut pas faire d'arithmétique sur les quantiles et probas. Il faut les recalculer pour chaque variable physique.***

Erreurs d'échantillonnage dans les ensembles

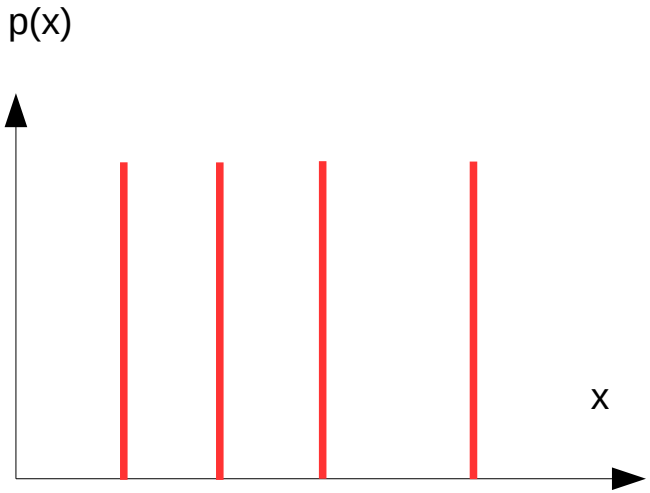
Cette distribution est-elle bimodale ?



(n=16) oui, probablement



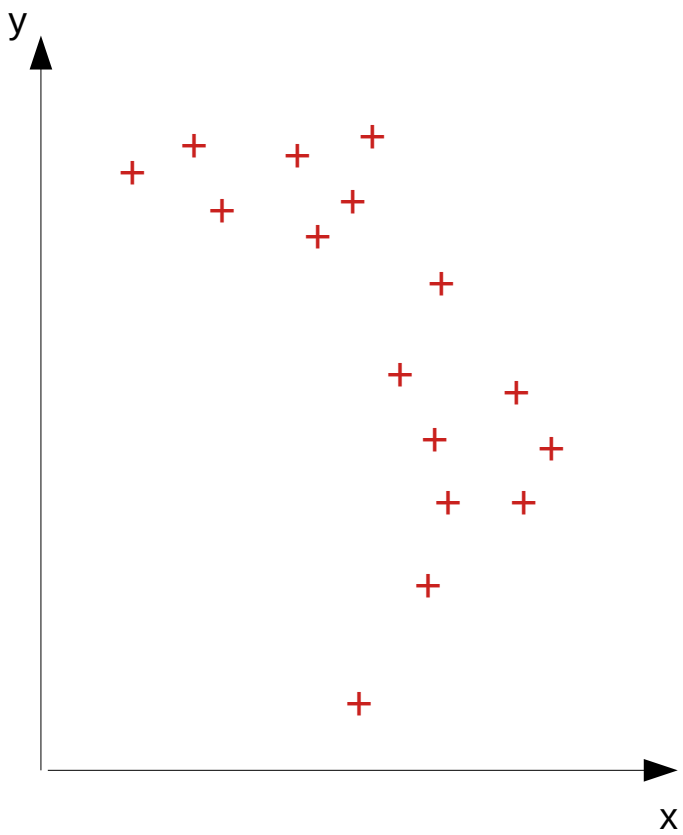
(n=10) c'est moins clair....

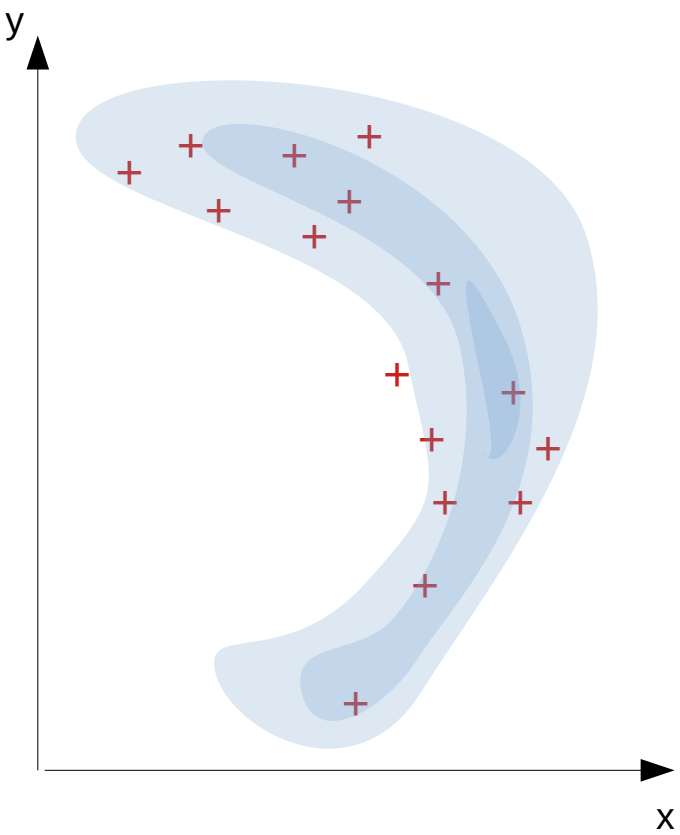


(n=4) euh ?!?

The curse of dimensionality

Cette distribution est-elle bimodale ? quel est l'état le plus probable ?

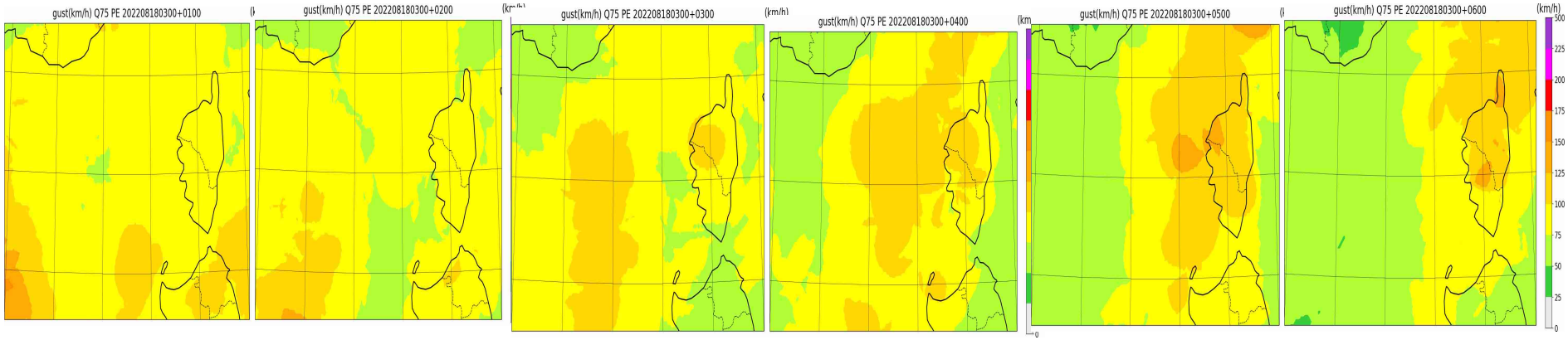




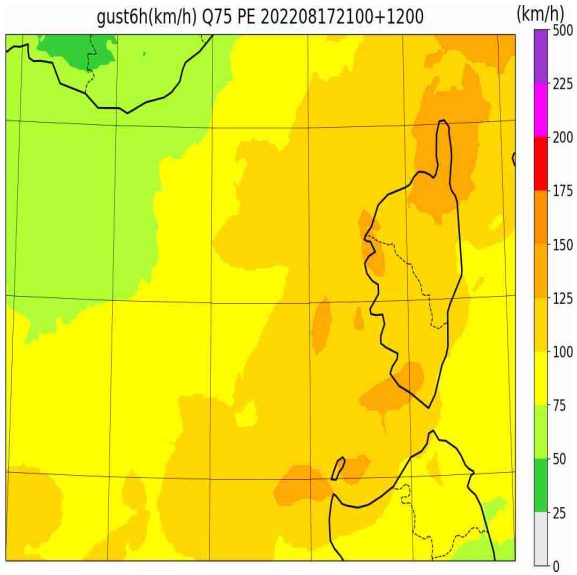
Non-additivité des probas et quantiles

(tempête corse d'août 2022)

quantile Q75 de vent horaire prévu, durant 6h consécutives :



quantile Q75 du vent sur ces 6h :



C'est plus que le max des Q75 horaires : il est très probable qu'il y aura un orage quelque part durant la période, mais on ne sait pas quand

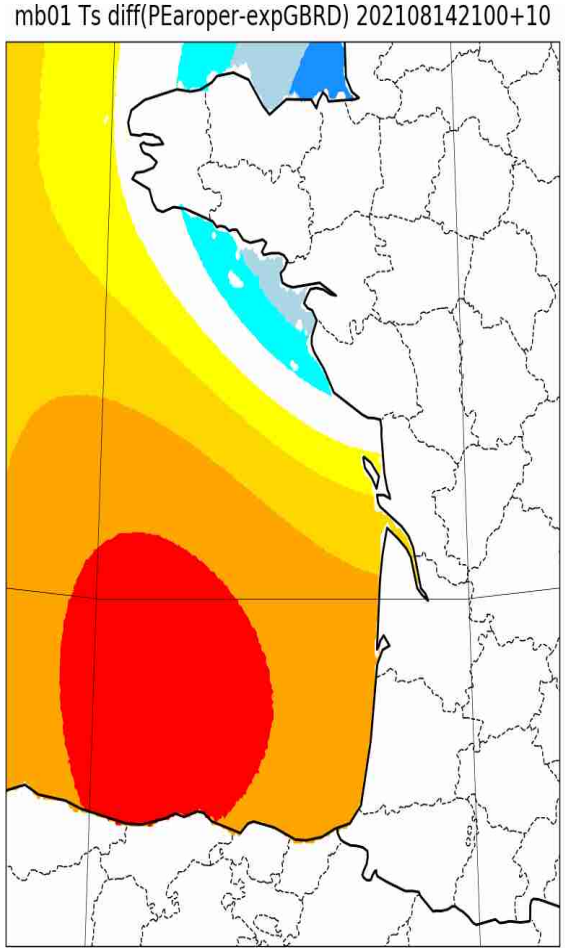
Génération des ensembles: techniques de perturbation

- identifier les **paramètres sensibles** (= processus incertains ET impactants sur les prévisions):
  - conditions initiales
  - forçages (latéraux, surfaces, physiques)
  - incertitude structurelle du modèle (paramétrisation, erreurs de discrétisation)
- les modifier par de**s** **perturbations aléatoires** de chaque membre - leurs distributions (amplitudes et corrélations) doivent
  - ressembler à celles des erreurs de x
  - ne pas provoquer de modification non-physique du modèle
- optimiser par validation statistique des sorties de l'ensemble (**scores**)

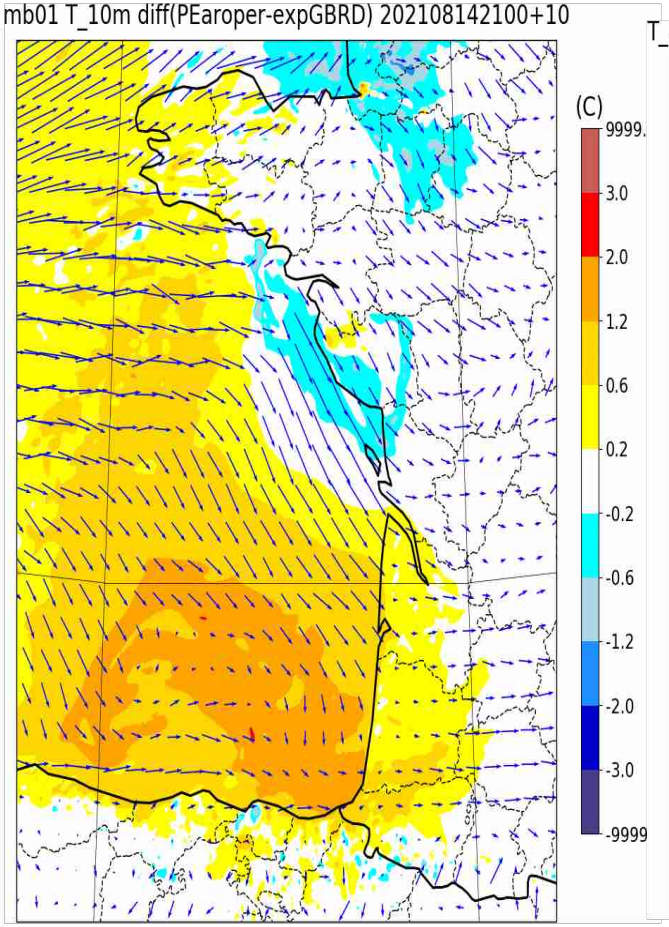
Techniques de perturbation : choisir ce que l'on perturbe : identifier les champs importants

ex de champ impactant : la Tmer (thèse M. Marquillie)

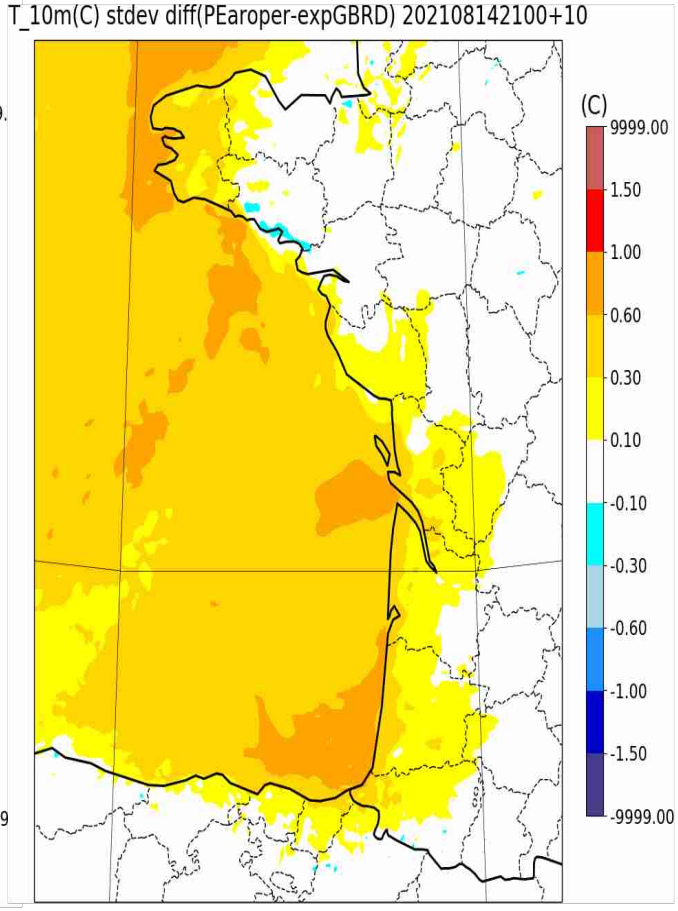
perturbation de Tmer



réponse de T10m



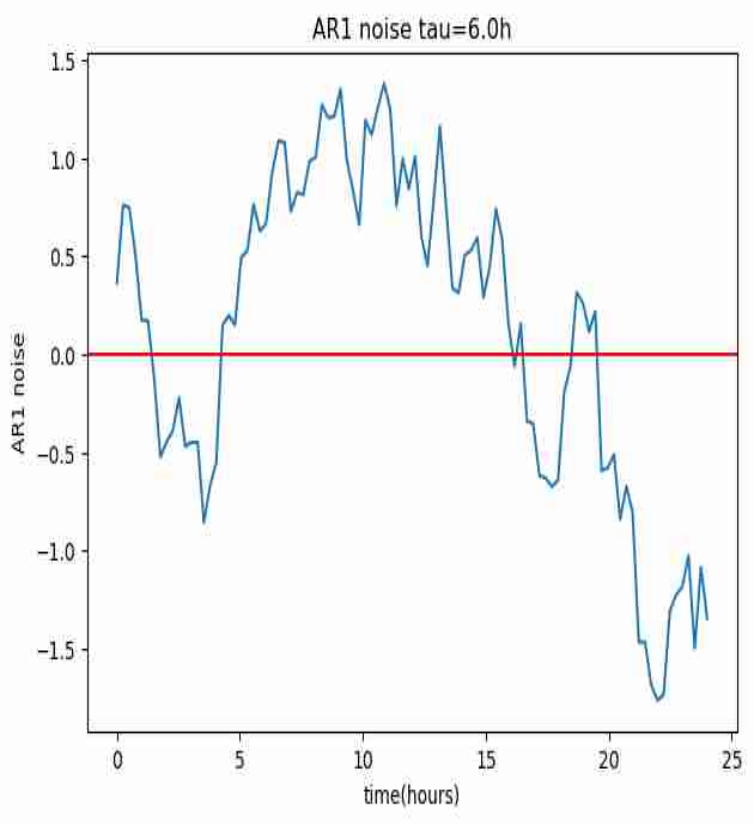
impact sur l'écart-type de T10m après 10h (16 membres)



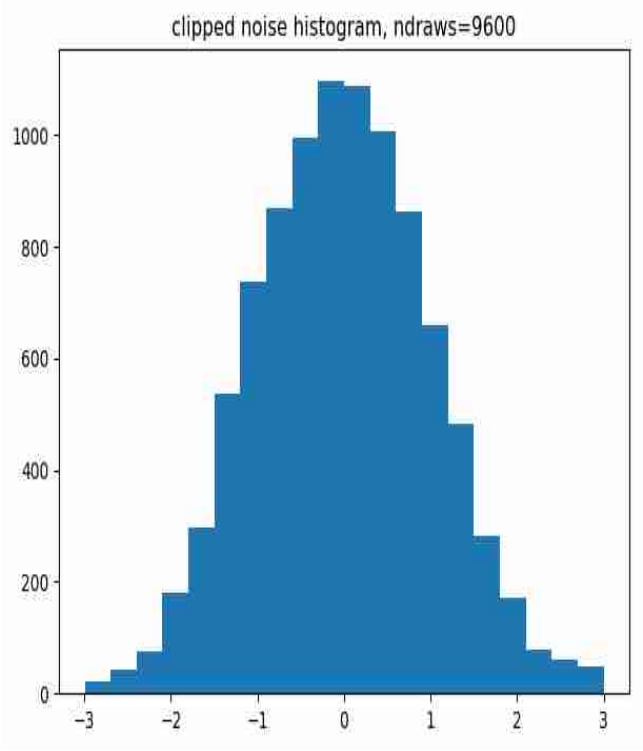
Techniques de perturbation : générateur de bruit

bruit temporel avec processus aléatoire AR1 :

série temporelle



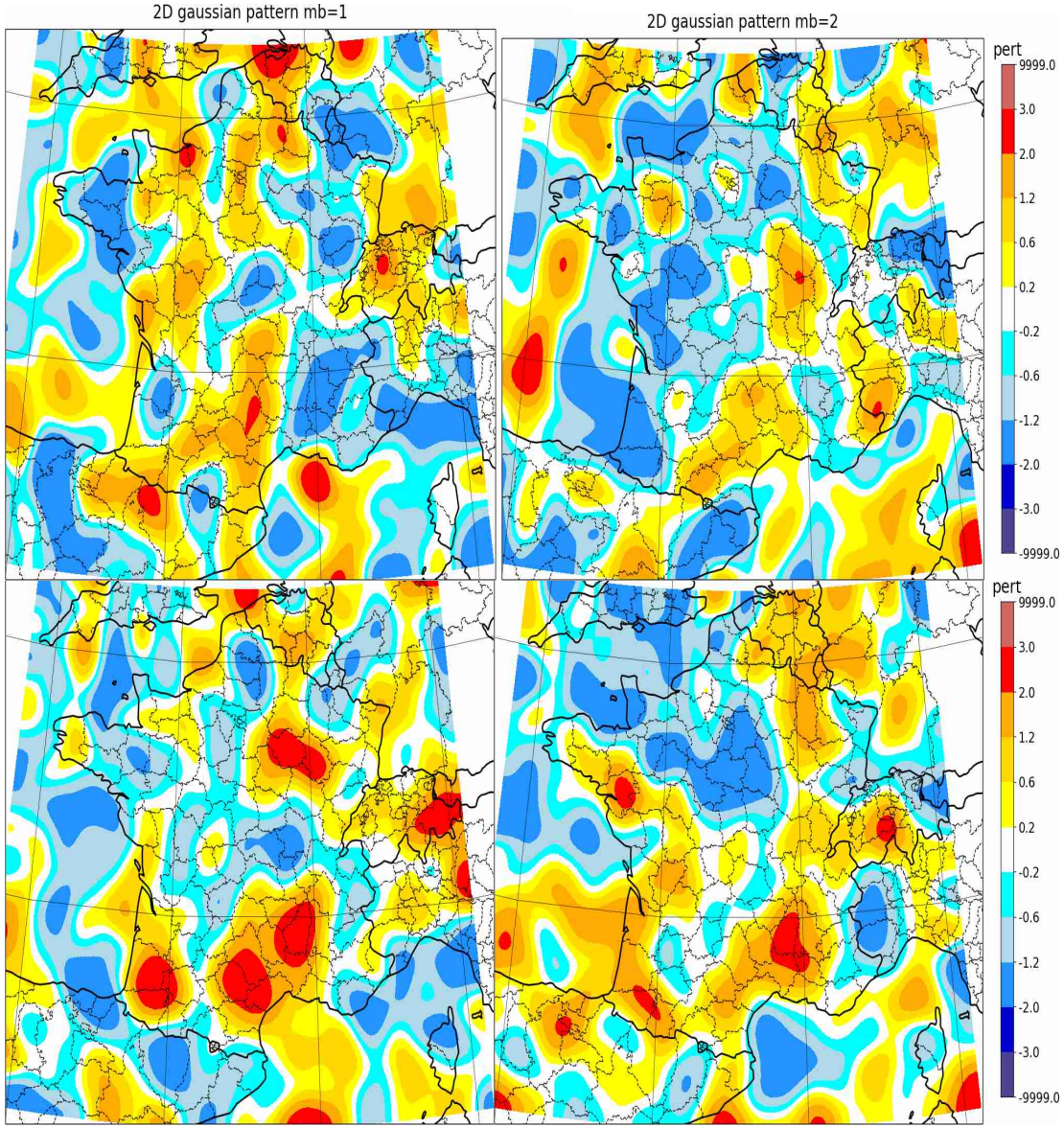
distribution sur une longue période



Techniques de perturbation : générateur de bruit

champ de bruit 2D : par lissage spatial d'un bruit blanc (corrélation résultante quasi-gaussienne)

4 réalisations



Principaux scores des prévisions d'ensemble
(cf fonctions de perte et de validation en IA)
pour des paramètres binaires, réels ou catégoriels

Dispersion :

- erreurs des membres > erreur d'une prévision non perturbée
- variance de dispersion ~ erreur rms de la moyenne de l'ensemble
- -> ratio dispersion/erreur_moyenne ~1 (+ erreur d'observation)

Fiabilité statistique des probabilités :

- si un événement est prévu avec une probabilité p%, il doit se produire p% du temps.

Résolution statistique :

- les probabilités doivent différer le + possible de leur moyenne climatologique
- **scores synthétiques fiabilité+dispersion:** Brier, CRPS, entropie

Utilité : de décisions binaires basées sur les probabilités

- ratio de bonnes prévisions / mauvaises prévisions
- attention au ratio entre non-détections et fausses alarmes ! dépend des applications
- ex: Jaccard, CSI, ETS, diagrammes ROC et precision-recall

Score : le diagramme de fiabilité

Principe : lorsque l'ensemble prévoit un événement binaire avec x % de probabilité, est-il observé x % du temps ?

C'est un **calcul de biais des probabilités** qui **stratifie les prévisions par classe de probabilités prévues** :

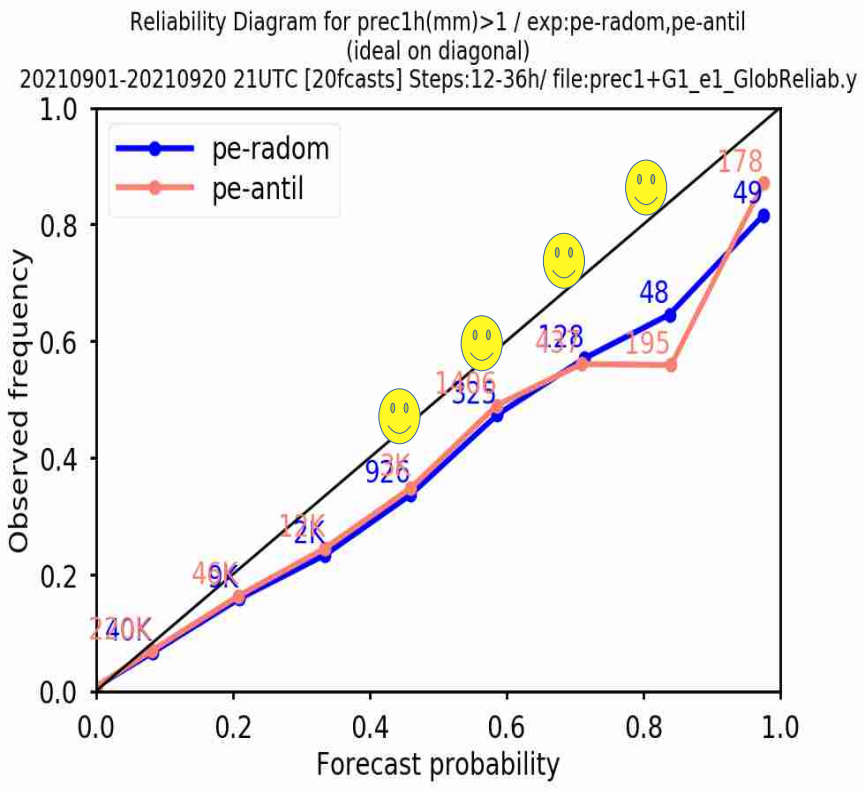
1) on fait un histogramme des probabilités prévues

2) on regarde dans chaque classe x à quelle fréquence y l'événement est observé : idéalement y=x

3) courbe trop plate = ensemble trop confiant (probas prévues trop proches de 1 ou de 0)

pièges : variations des populations de classes, et présence de classes peu peuplées

C'est un "sanity check" visuel bon pour la com, mais pas une mesure de qualité (utiliser plutôt le score de Brier)



Score : le diagramme de rang

Principe : peut-on distinguer les distributions des valeurs prévues et observées ?

C'est un histogramme des rangs relatifs des prévisions et des observations :

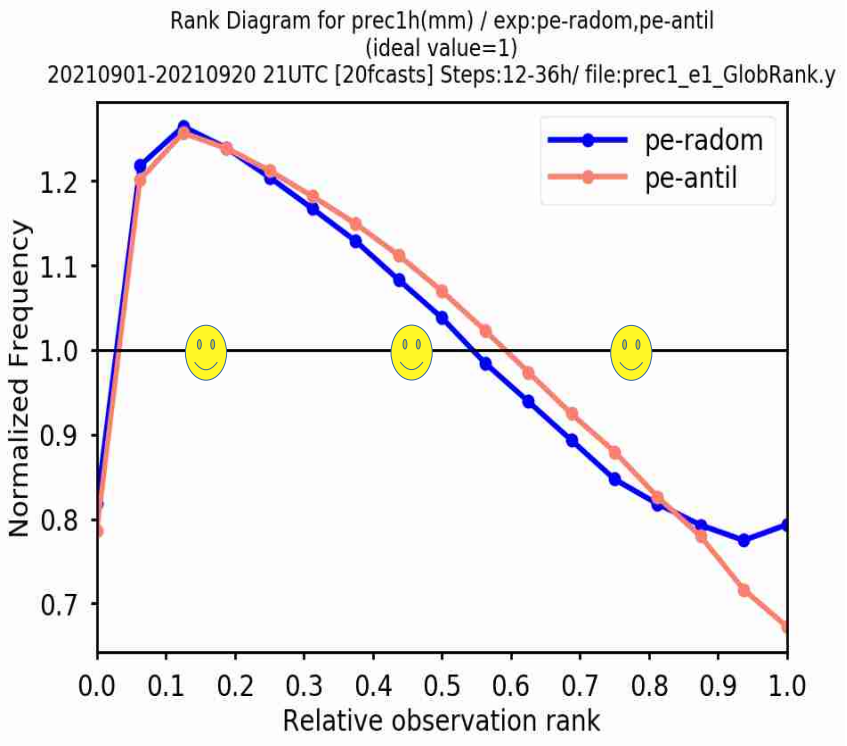
ordonner (x1,x2,...xn) et regarder le rang i de l'observation y (n+1 rangs possibles)

faire un histogramme des rangs i

en cas d'ex aequo, il faut répartir entre tous les rangs possibles (ex : si tous les x et y sont nuls)

optimum : lorsque la fréquence de chacune des classes est 1/(n+1)

piège : mélange les fortes et les faibles valeurs



Score : le diagramme ROC (relative operating characteristic)

Défini pour un événement binaire :

on fixe le seuil de valeur y

on fait varier le seuil de proba z --> table de contingence paramétrée : (a,b,c,d)=f(z) que l'on résume par les points

- **abscisse FAR(z) = b/(b+d)** false alarm rate (variante : certains utilisent b/(b+a))
- **ordonnée POD(z) = a/(a+c)** probability of detection

le ROC est la courbe paramétrique [FAR(z),POD(z)] pour z dans [0,1]

Lecture de la courbe ROC :

chaque point correspond à un niveau de probabilité z (pas toujours montré!)

si z~0, on prévoit presque toujours -> POD et FAR sont grands : point (1,1)

si z~1, on ne prévoit qu'à coup quasi-sûr -> POD et FAR sont petits : point (0,0)

prévision aléatoire : sur la diagonale (FAR=POD)

prévision parfaite : si (FAR,POD)=(0,1) -> **PE meilleure si plus près de (0,1)**

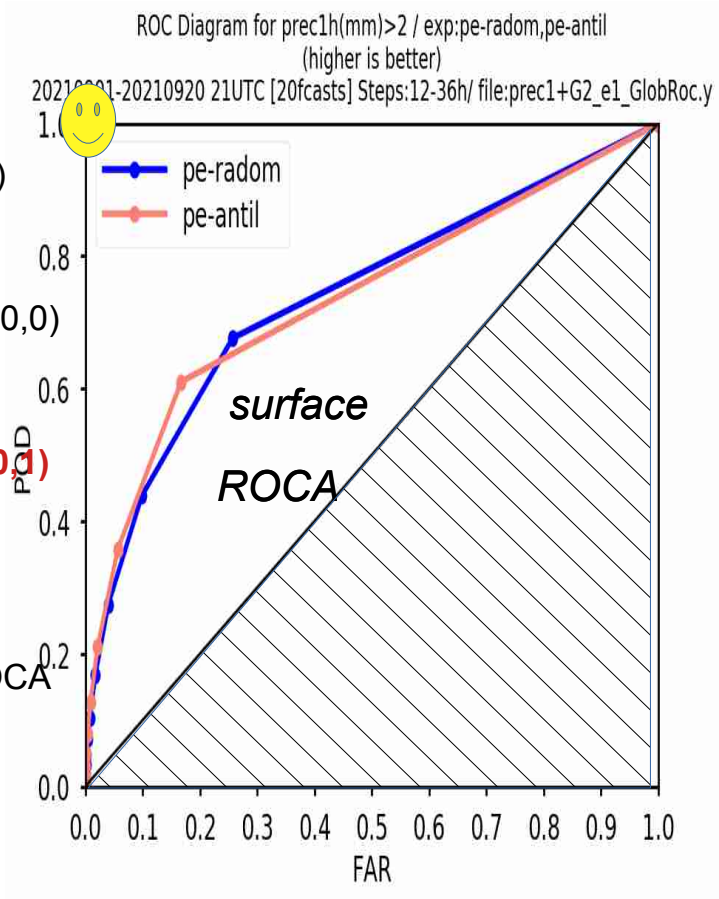
Interprétation :

mesure un contenu en information mais cache les pbs de fiabilité.

ambigu : plusieurs manières de définir "plus près de (0,1)", par ex score ROCA

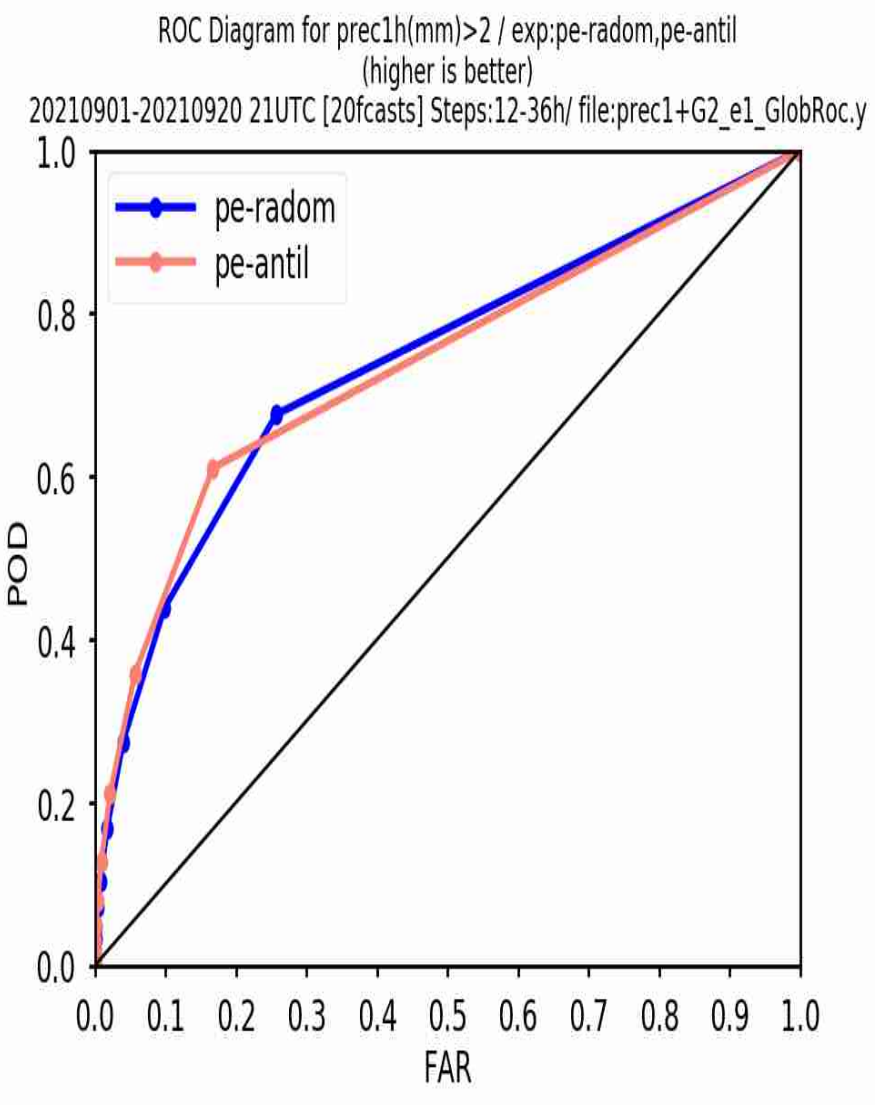
mélange qualité des PDFs et qualité du modèle de prévi

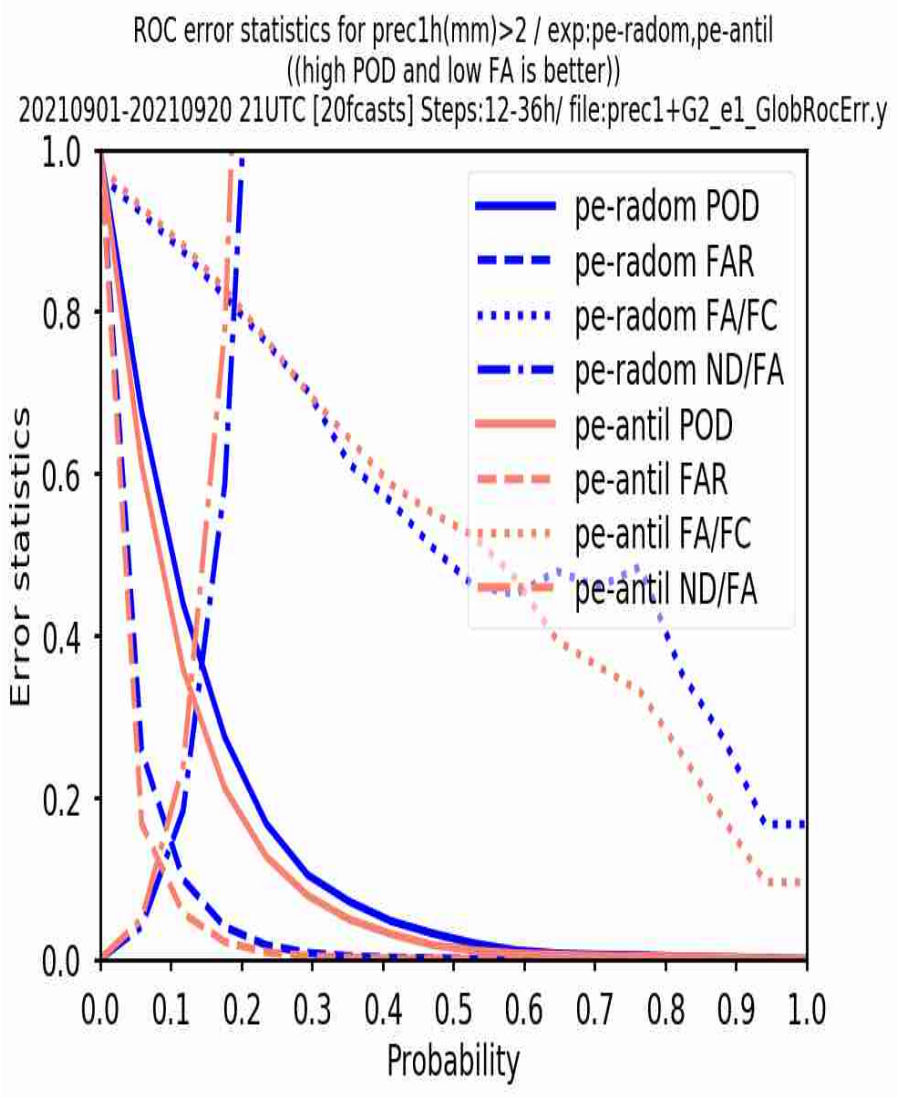
difficile à utiliser pour les événements rares



Score : le diagramme ROC

on peut "déplier" la courbe ROC en fonction des probabilités prévues (-> cf diagramme de fiabilité pour voir si elles sont bien calées)





Scores et évaluation des prévisions d'ensemble

CRPS : un score intégral pour les prévisions réelles *"Continuous Ranked Probability Score"*

Déf : moy de CRPS = $\int [\, cdf(x_i) - H(obs)\,]^2\, dx$ a

$cdf(x_i)$ la distribution prévue par l'ensemble x_i

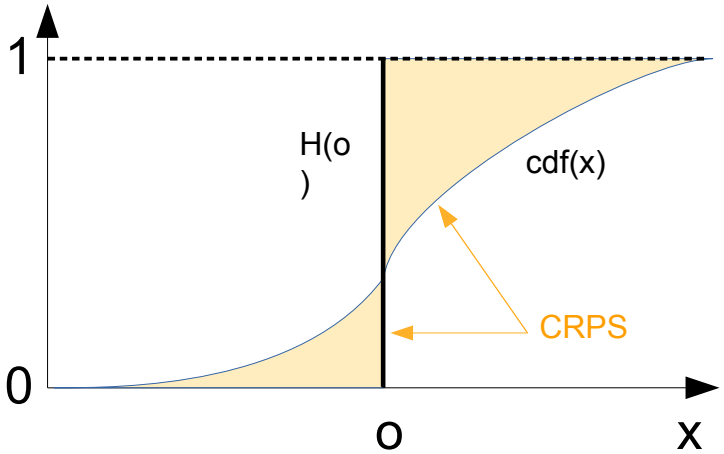
H(o) la fonction de Heaviside de l'obs o

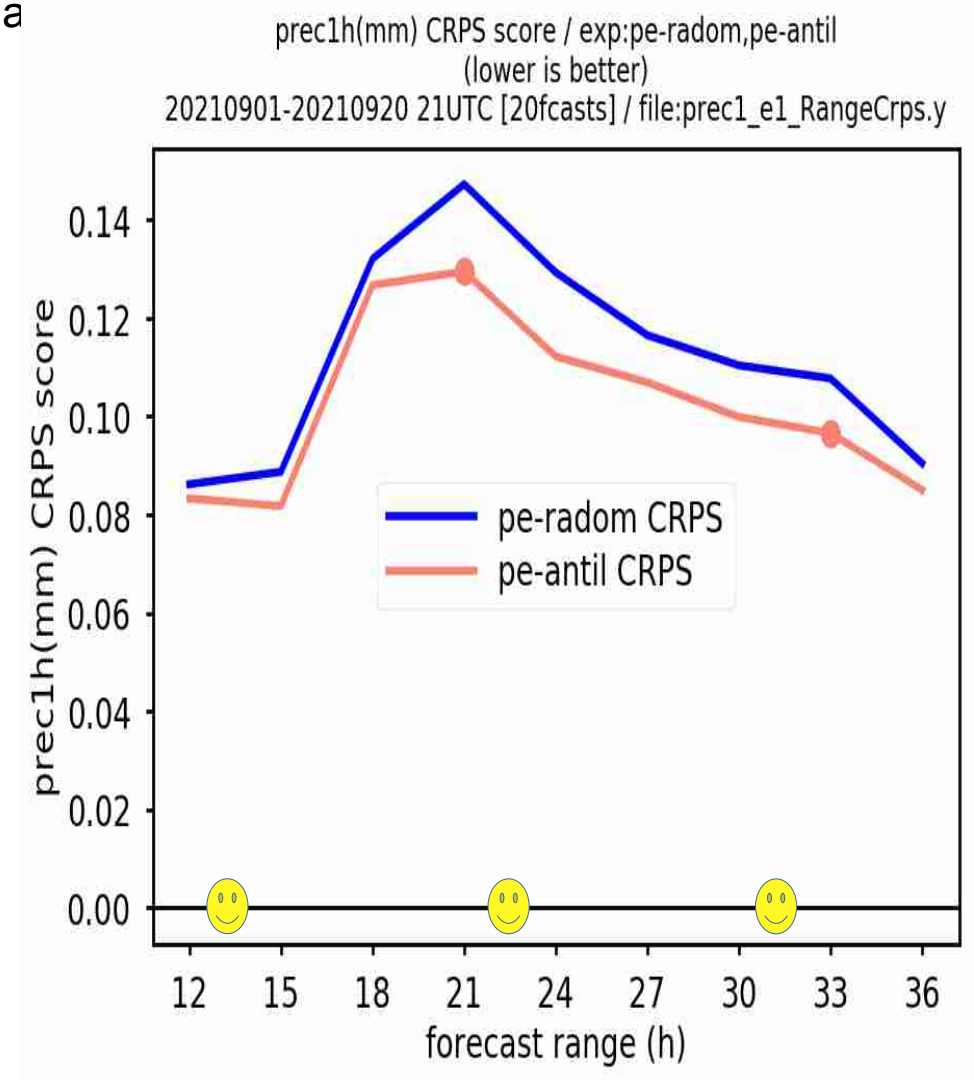
Remarques :

CRPS= ∫(Brier sur tous les seuils y) dy

pour une prévi déterministe, CRPS(x)=|x-o|

(c'est la MAE : mean absolute error)





Merci pour votre attention