

Sélection d'échelle en utilisation des prévisions d'ensemble

François Bouttier, CNRM/GMME/PRECIP



atelier IA, Fev 2025

Comment optimiser les productions probabilistes ?

Objectifs :

- automatiser les réglages des **post-traitements des PE** (prévisions d'ensemble) comme outils de prise de décision
- les rendre plus physiques
- **explicabilité** des calculs, pour préserver l'expertise humaine (liens entre probabilités et scénarios)

Contextes :

- **projet POEME** - interprétation humaine des PE
- **projet ALPHA** - valeurs privilégiées automatisées
- **(PI probabiliste** - PIAF-PE)
- **projet IRICLIM** - PE pour la prévision des crues rapides



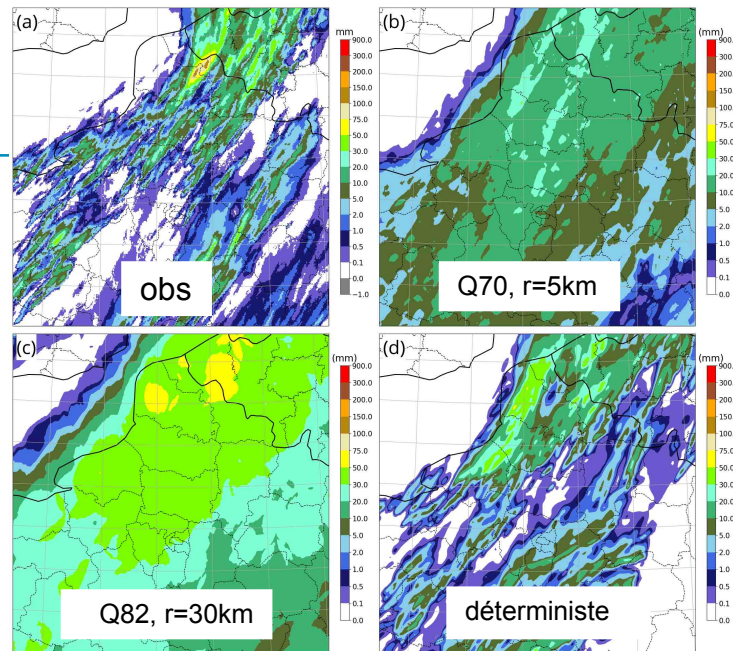
L'importance de la sélection d'échelle pour prévoir les pluies & la convection

2/10

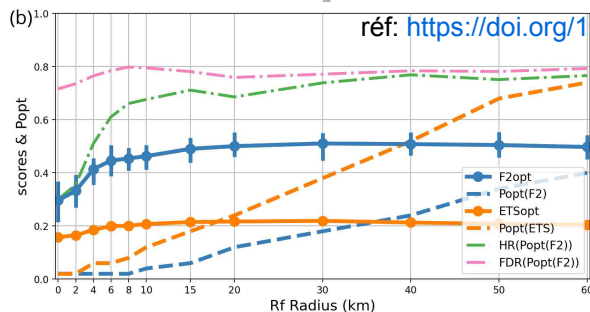
Objets avec incertitudes de position & timing >> leur dimension

- éviter de lisser les champs (les impacts proviennent des extrêmes)
- échelles de prévision optimales dépendent
 - du **public visé** (préférences fausses alarmes vs non-détections)
 - du **phénomène** (fortes pluies $\neq$ faibles)
 - des **erreurs de prévision** des systèmes PN

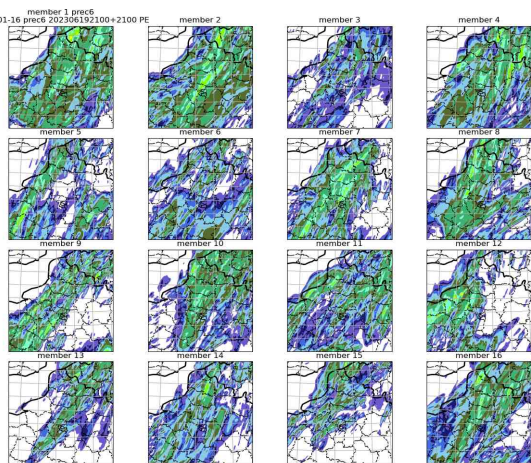
réf: <https://doi.org/10.1175/MWR-D-16-0400.1>



Optimisation du quantile & filtre spatial (disque max)
Apprentissage sur 18 mois de PE Arome
Extrapolation des quantiles aux cas extrêmes



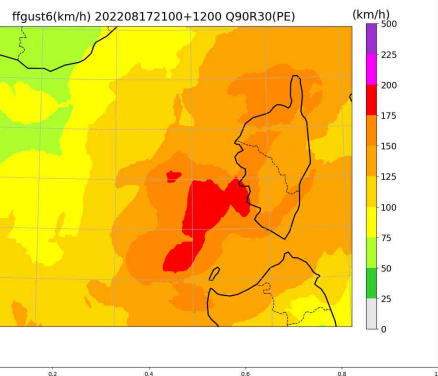
réf: <https://doi.org/10.5194/nhess-24-2793-2024>



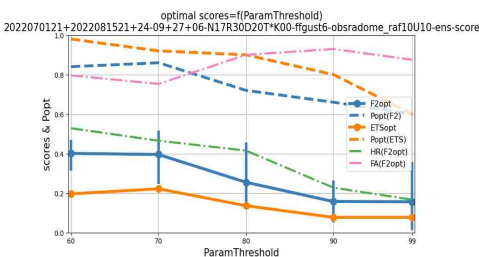
Quelques applications récentes, optimisées "à la main"

étude de prévisibilité
du Derecho corse
(cf. Prévi/GMAP)

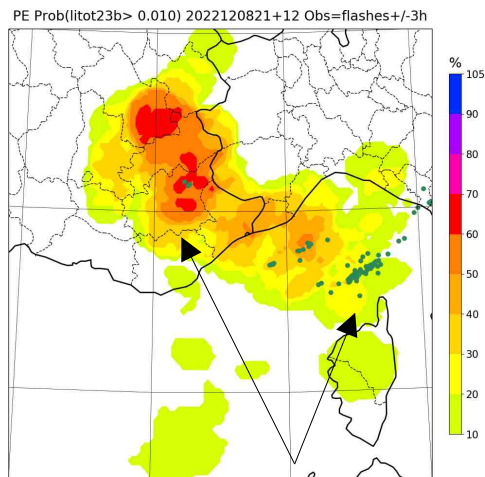
rafales Q90 ~160km/h



Quantile vigilance MF :
Qopt ~ 90km/h

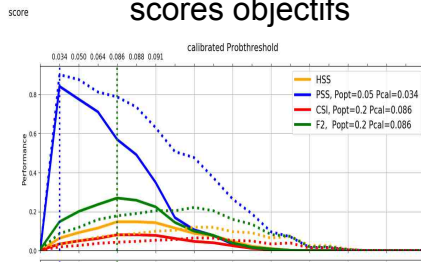


ProbOrage ALPHA
prévision de risque d'orage
(collab. GMAP)

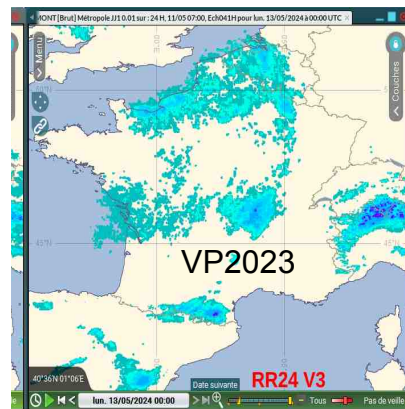


obs Météorage

scores objectifs

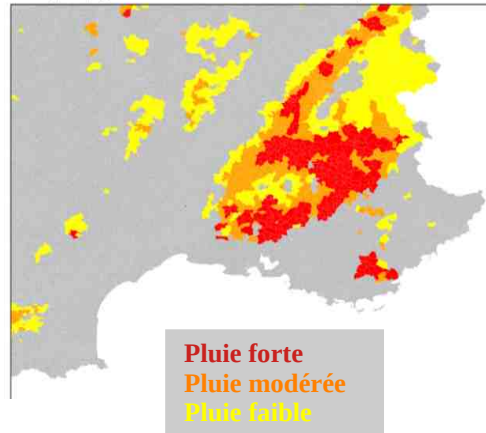


valeur privilégiée pluie ALPHA
(collab. K. Clain DirOP/MPA)

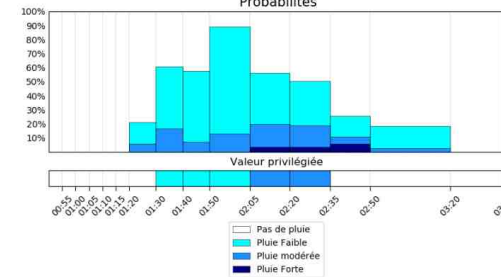


prévision immédiate d'ensemble
"pluie dans l'heure"
(collab. A. Warnan DirOP/PI)

889piafpg time=2022110917+0.25 pdh Commune:



Saint Mandé : Prévision Pluie -> 2023-03-31 00:50:00 -> Probabilités



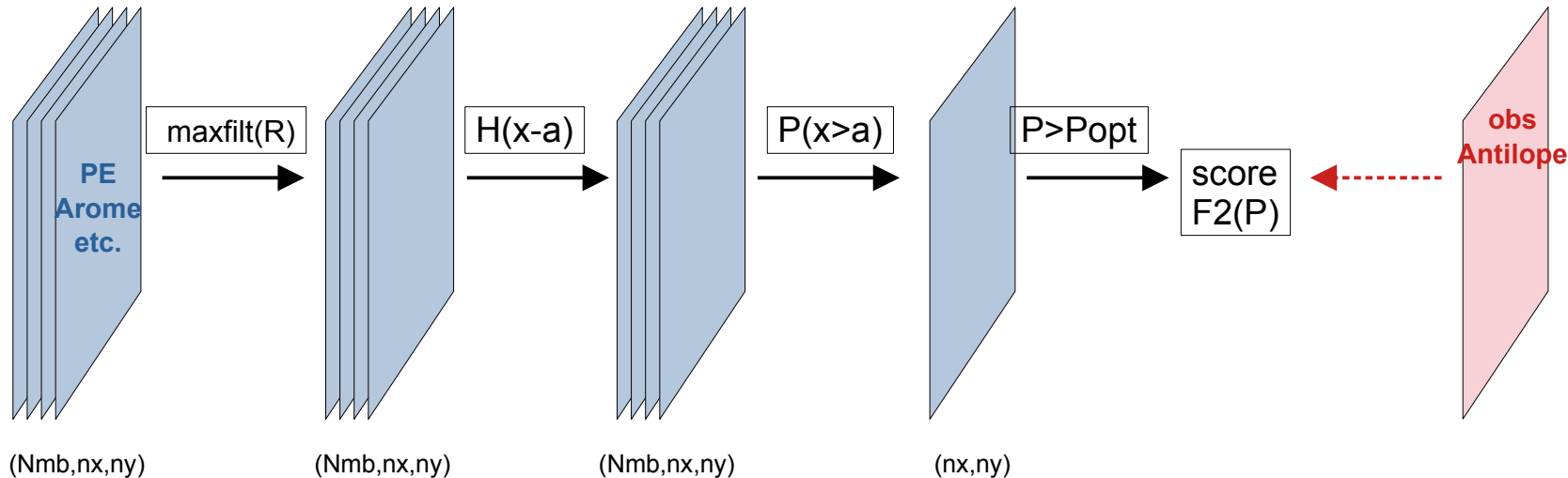
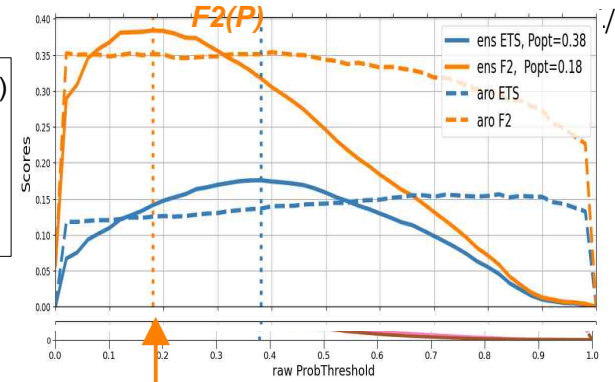
Ancien post-traitement PROBGEN

Séquence typique de traitement pour prévoir si $x > a$:

1. opérateur de voisinage max_disk (*Nmb membres)
2. seuillage physique $x > a$ (*Nmb membres)
3. conversion $\text{PDF}(x) \rightarrow P(x > a)$
4. seuillage proba $P(x > a) > P_{\text{opt}}$
5. optimisation de score $F2(P_{\text{opt}})$ sur plusieurs mois de données

$$\text{score } F2 = a / (a + 1/5b + 4/5c)$$

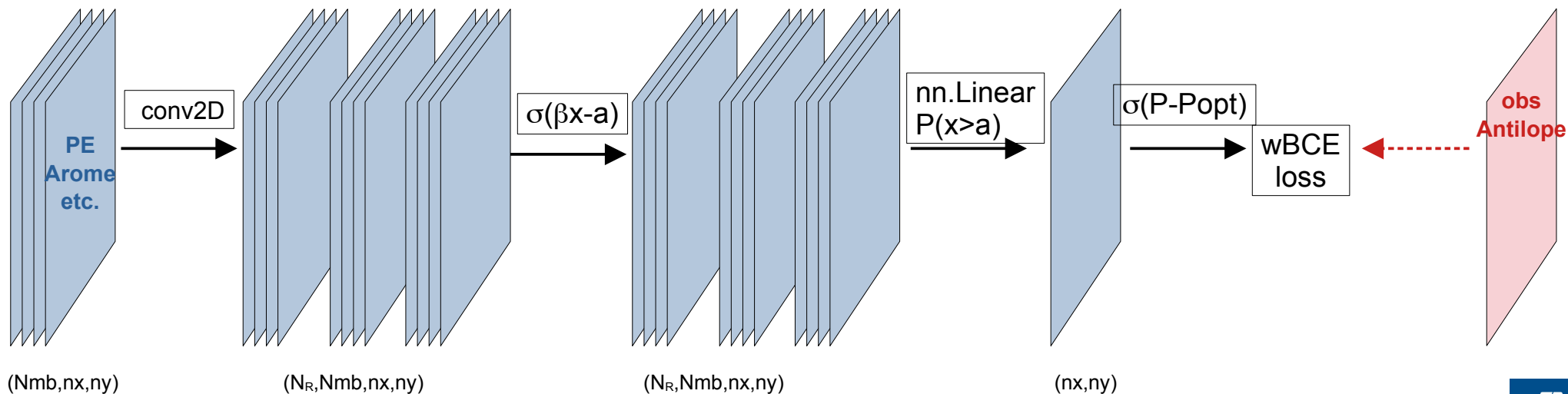
a : positifs corrects
b : FA
c : ND



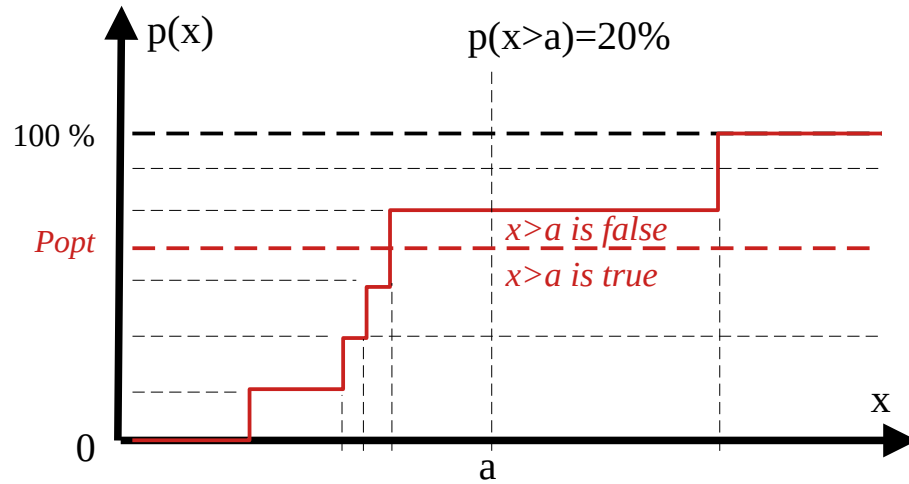
Approximation par un réseau de neurones "PROBGEN-NN"

- voisinage : convolutions max_disk (rayon R, 2R, 4R, ...)
- seuillage physique : fonction sigmoïde $\sigma(\beta x - a)$ = habillage de PDF
- conversion PDF(x) : couche linéaire
- seuillage proba : fonction sigmoïde
- score F2 approché par fonction-perte entropie wBCE

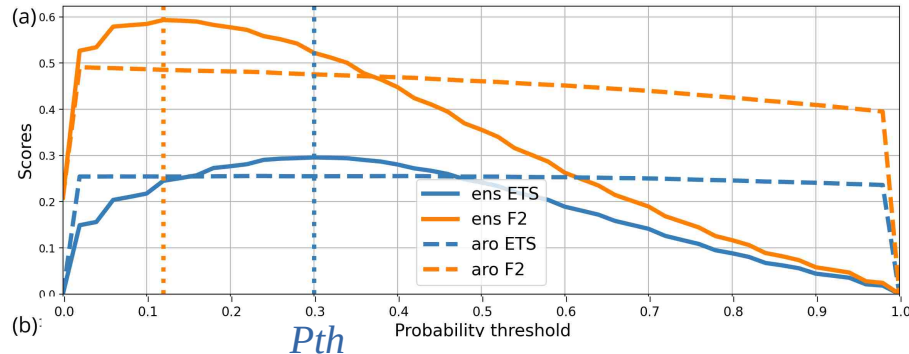
Tous les paramètres réglables deviennent des poids appris du modèle, invariants par permutation des membres



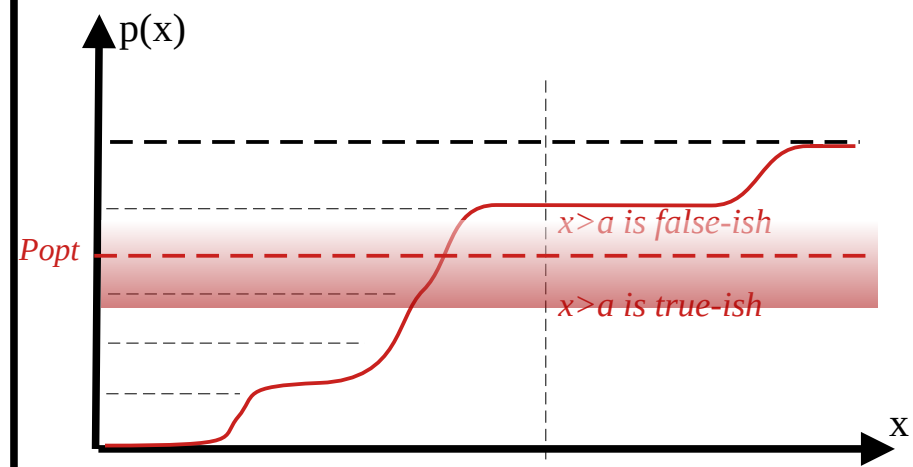
original algo : Heaviside CDFs & hard thresholds



P_{opt} optimizes $F2 = a / (a + 1/5b + 4/5c)$
of the confusion matrix (or F1, ETS, HSS etc)



neural network : sigmoid CDFs & smooth thresholds



P_{th} optimizes $wBCE = a + b + c + d$

	obs=yes	obs=no
fcast=yes	$a=0$	$b = -\text{Log}(1-p)$
fcast=no	$c = -w\text{Log}(p)$	$d=0$

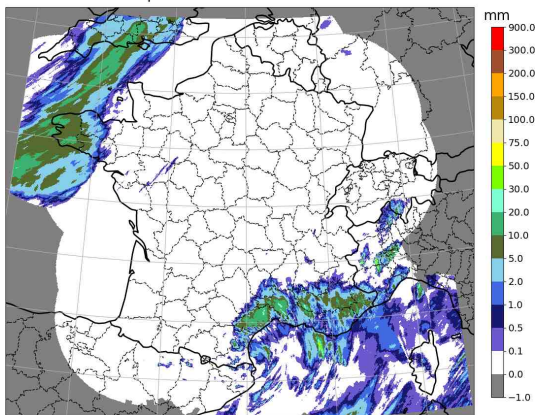
big if p small

big if p big

Test 1 : optimisation directe des prévisions binaires de dépassement de seuil

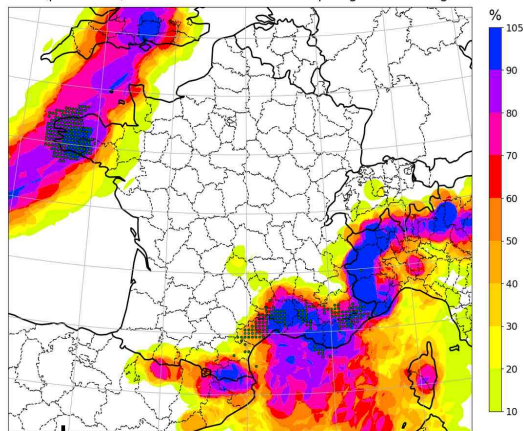
obs Antilope

Antilope obs rr3h end=202309201500



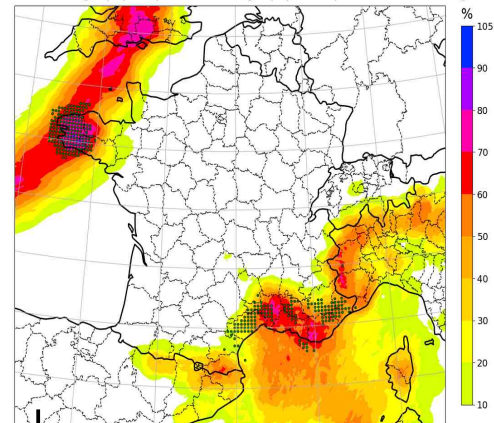
Probgen-ref : proba $P(x>a)$

pearo $P(\text{prec3}>5.0)$ obs>0.95 2023091921+18 probgen5.0R30.0.gt.0.95



Probgen-NN : proba

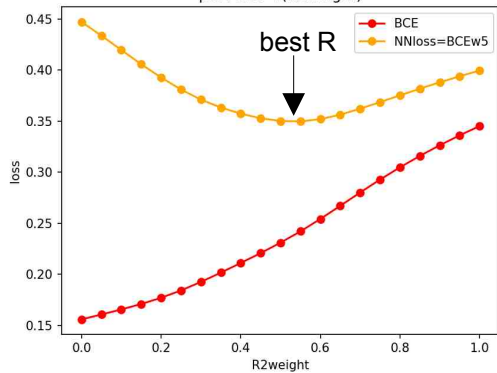
pearo $P(\text{prec3 } 2023091921+18 \text{ neigh2popt30-postP5mm neigh2popt30 } P_{\text{opt}}=0.533872 \text{ R2weight=learned1.359}$



fonction perte = $f(\text{rayon } R)$

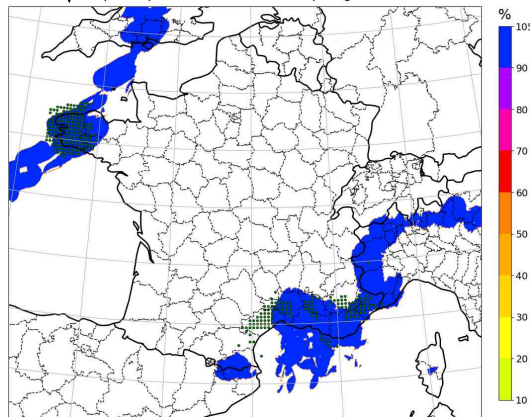
post loss= $f(R2weight)$

best R



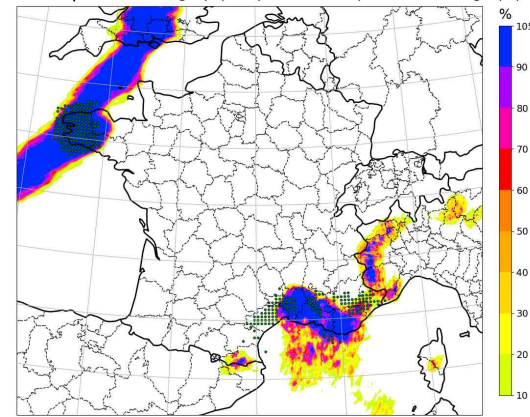
Probgen-ref : prévi binaire

pearo $\text{prec3 } 2023091921+18 \text{ probgenP5}$



Probgen-NN : prévi "binaire"

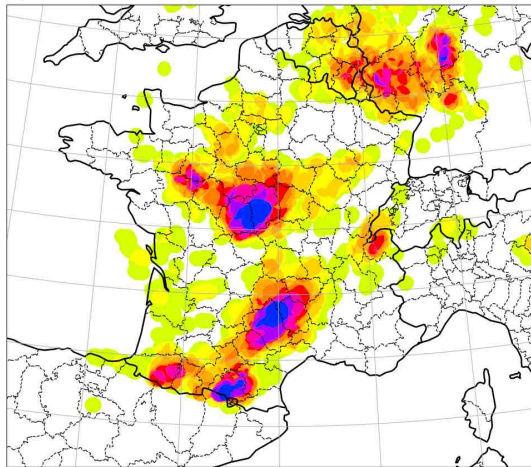
pearo $\text{prec3 } 2023091921+18 \text{ neigh2popt30-postP5mm } >P_{\text{opt}}=0.533872 \text{ neigh2popt30}$



Test 2 : optimisation des probabilités prévues avec un modèle un peu + complexe

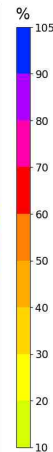
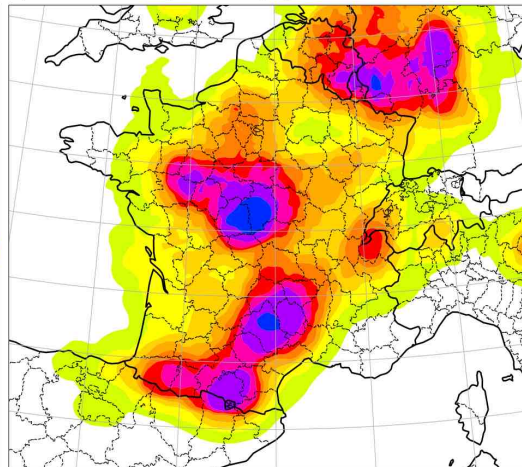
Probgen-ref : $P(x>a)$

pearom0m8s6s6 P(prec3>10.0) algo:maxneigh 2024072021+06



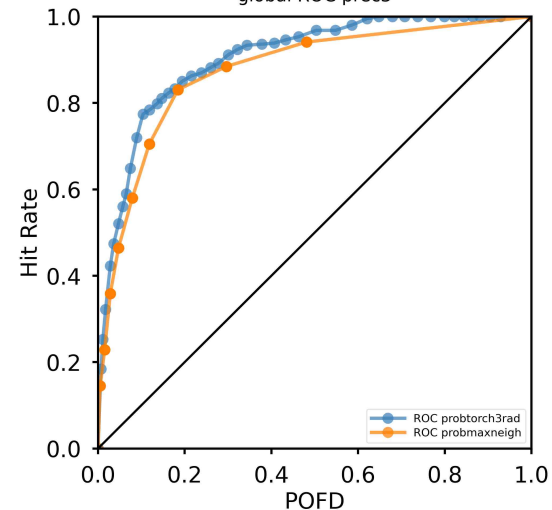
Probgen-NN : $P(x>a)$

pearom0m8s6s6 P(prec3>10.0) algo:torch3rad 2024072021+06

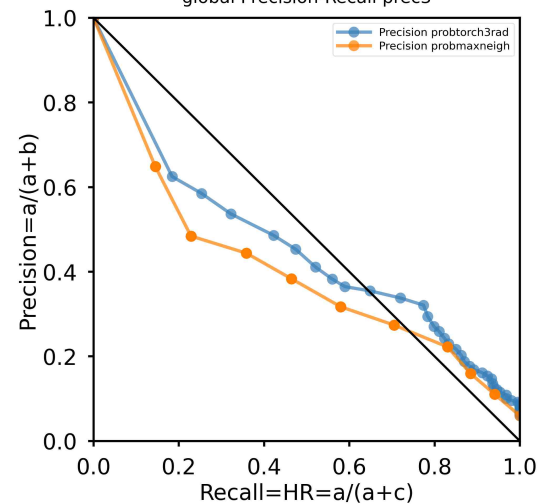


scores: NN vs REF

global ROC prec3



global Precision-Recall prec3

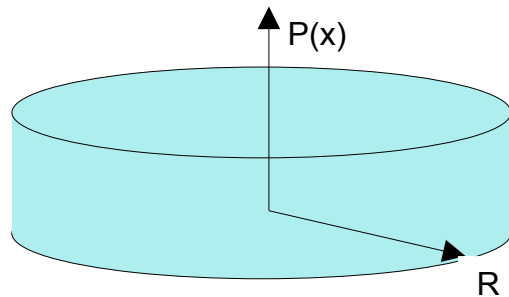


Prochaine étape : améliorer l'opérateur de convolution spatiale

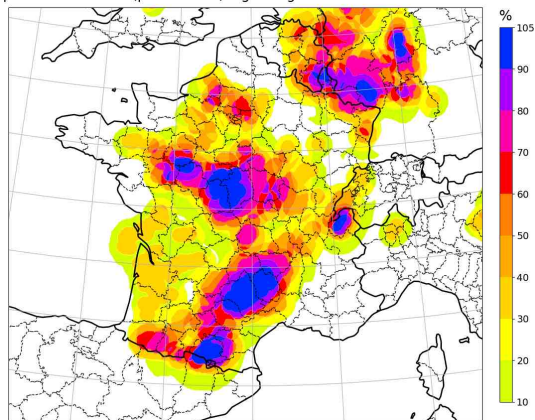
voisinage max_disk:

- erreur de déplacement uniforme dans chaque disque
- isotrope

-> prévisions "granuleuses"



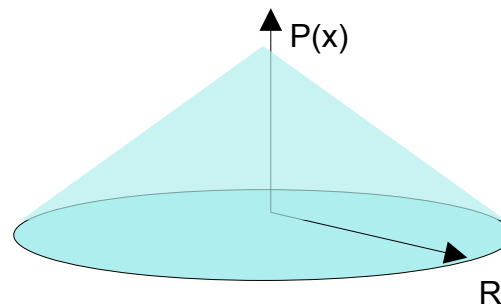
pearom0m2s6s6 P(prec3>10.0) algo:weigh3radmax 2024072021+06



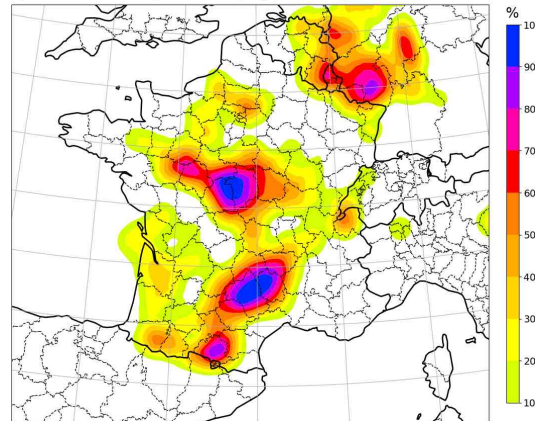
voisinage conique:

- les déplacements plus petits sont plus probables
- anisotropie possible sans surcoût

-> plus physique



pearom0m2s6s6 P(prec3>10.0) algo:conicneigh 2024072021+06



Conclusions

Fait :

- conversion du post-traitement ensembliste PROBGEN en réseau de neurone équivalent
- automatisation totale des réglages par apprentissage machine
- amélioration des scores de prévision, à complexité égale
- (réduction des coûts de calcul des voisinages)
- modèle qui permet l'interprétation par scénario (en remontant des probas aux membres de la PE)

En cours :

- généraliser l'opérateur de voisinage
- apprendre la dépendance à l'orographie (voisinages anisotropes) et au type de temps (fronts...)

A faire :

- démonstrateur temps réel
- adaptation à VP ALPHA
- application au projet PEPR IRICLIM
- perspectives comme post-traitement d'AROME hectométrique et AROME-IA