

Développement d'un indicateur de brouillard en vallée de Seine

Théo Tournier (DSM/LabIA) & Florian Gibier (DSM/CS/DC)
Journée de l'IA, 08/02/2024

Remerciements

Philippe Periot (DIRN/C) qui a travaillé sur la contractualisation de l'étude et pour nous avoir partagé son lien avec le client.

Martine Baillon et **Valentine Chatel** (DSM/CS/DC) pour la réalisation de l'étude climatologique qui a précédé le développement des indicateurs.

Céline de Saint-Aubin (DSM/CS/DC/DA) pour le pilotage macro du projet.

Les prévisionnistes de la DIRN pour nous avoir conseillé dans les paramètres Arome à utiliser.

Les prévisionnistes de la DIRO pour nous faire des retours quotidiens sur la pertinence des indices.

Les prévisionnistes de la DIRO envoient chaque jour un bulletin à destination du **Pilotage de la Seine** pour **prévoir l'occurrence de brouillard** à une échéance de **+24h** (et une tendance à +36h).

La Seine de l'estuaire à Rouen

Pilotage de la Seine

Rédigé le 10 février 2022 à 16:40 lég.



🔗 Analyse et prévision jusqu'au vendredi 11 février à 09h

La perturbation de ce jour est suivie d'un temps plus sec à partir de la nuit prochaine. Le ciel se dégage, ce qui favorise le refroidissement et le rayonnement. Quelques brumes voire quelques poches de brouillard sont possibles localement dans l'intérieur en fin de nuit et en début de matinée. Le soleil domine ensuite.

🔗 Synthèse de la surveillance jusqu'au vendredi 11 février à 09h

Risques météo surveillés	De Honfleur à Port-Jérôme - Radicatel	De Port-Jérôme - Radicatel à Rouen
Brume (1000m<vis<=5000m)	 Pas de risque	 Risque faible
Brouillard (vis<=1000m)	 Pas de risque	 Risque faible

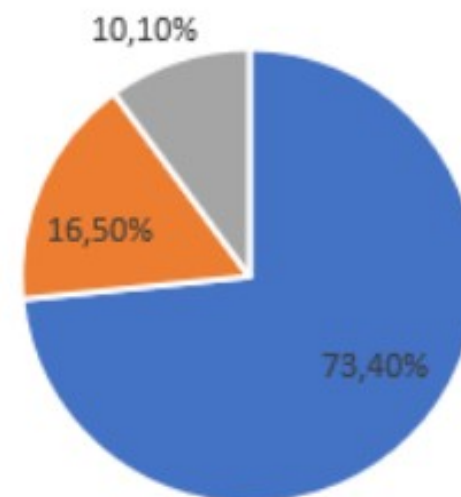
🔗 Tendance pour le vendredi 11 février de 09h00 à 23h59

Le ciel reste peu nuageux durant la période, avec cependant un voile d'altitude qui arrive par l'ouest en soirée, ce qui n'empêche pas l'atmosphère de se refroidir à nouveau durant la nuit suivante.

Le Pilotage Seine a signalé une **augmentation des cas de non détection** de brouillard dans les bulletins lors de l'hiver 2021.

Le nombre de brouillard en 2021 à augmenté par rapport aux années précédentes : +23 % par rapport à 2019.

PERIODE DU 01/09 AU 05/11 21

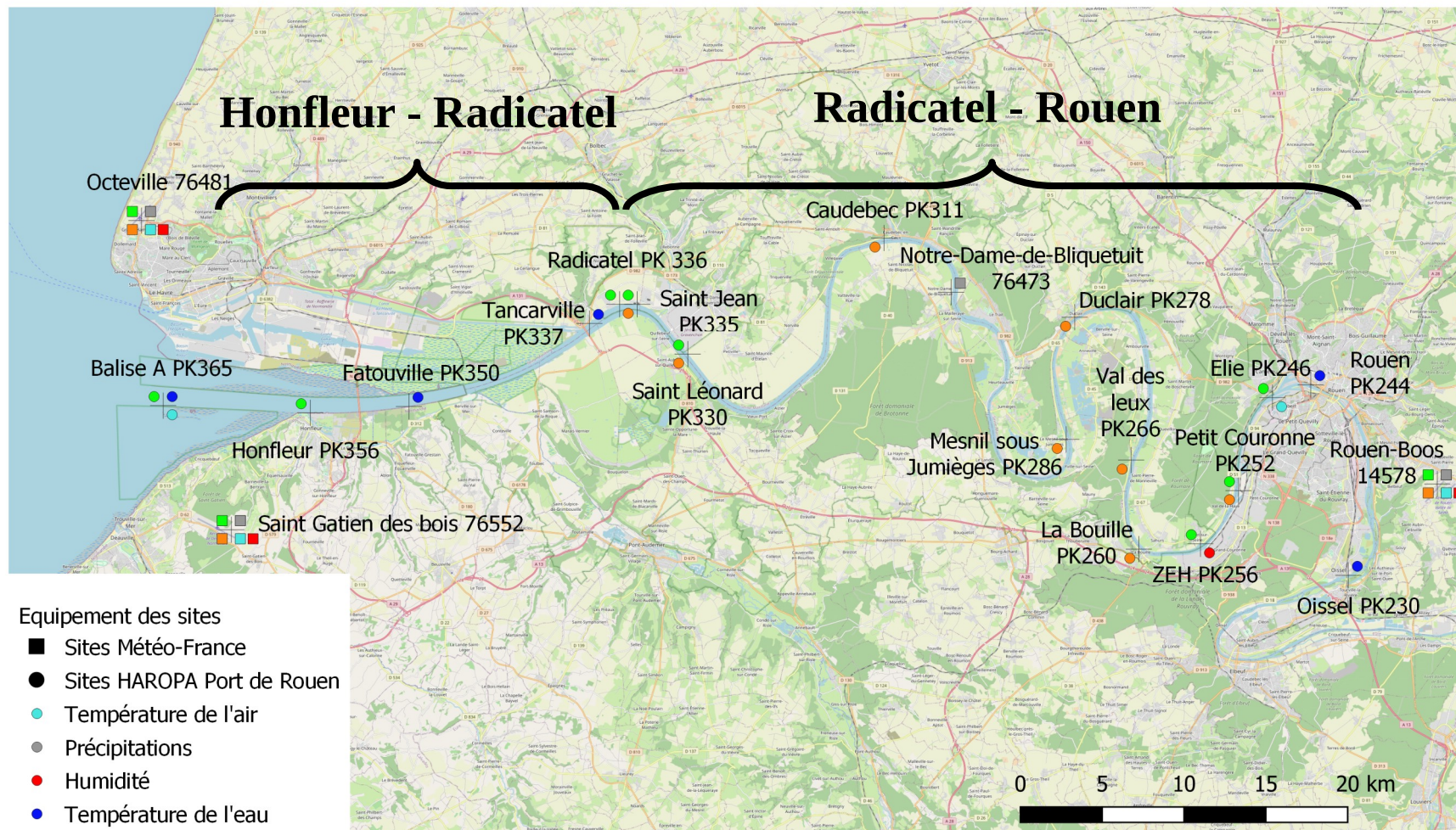


■ prévisions cohérentes ■ Non détection
■ Fausse détection

Améliorer la prévisibilité du brouillard dans la vallée de Seine :

Pour chaque heure, dire si $\text{visi} < 1000\text{m}$ ou $\text{visi} > 1000\text{m}$ et limiter les non détections!

Observations du Pilotage de la Seine



Premières observations disponibles **depuis 2006** (changement des capteurs de visi **entre 2015 et 2017**)

Pas de capteurs de visi sur le tronçon *Honfleur – Radicatel*.

Utilisation des capteurs de **visibilité + température de l'eau**.

NOM DU PARAMÈTRE	NIVEAUX (m)
FF (Force du vent)	10
HU (Humidité relative), T (Température), TD (Point de rosée)	2, 10, 20, 50
CLWC (Cloud Liquid Water Content)	10, 20, 50
LCC (Low Cloud Cover), MCC (Middle Cloud Cover)	x
MSLP (Mean Sea Level Pressure)	x
SSHF (Surface Sensible Heat Flux), SLHF (Surface Latent Heat Flux)	x
SSR (Surface Solar Radiation), SSRD (Surface Solar Radiation Downwards)	x
EVA (Évaporation)	x

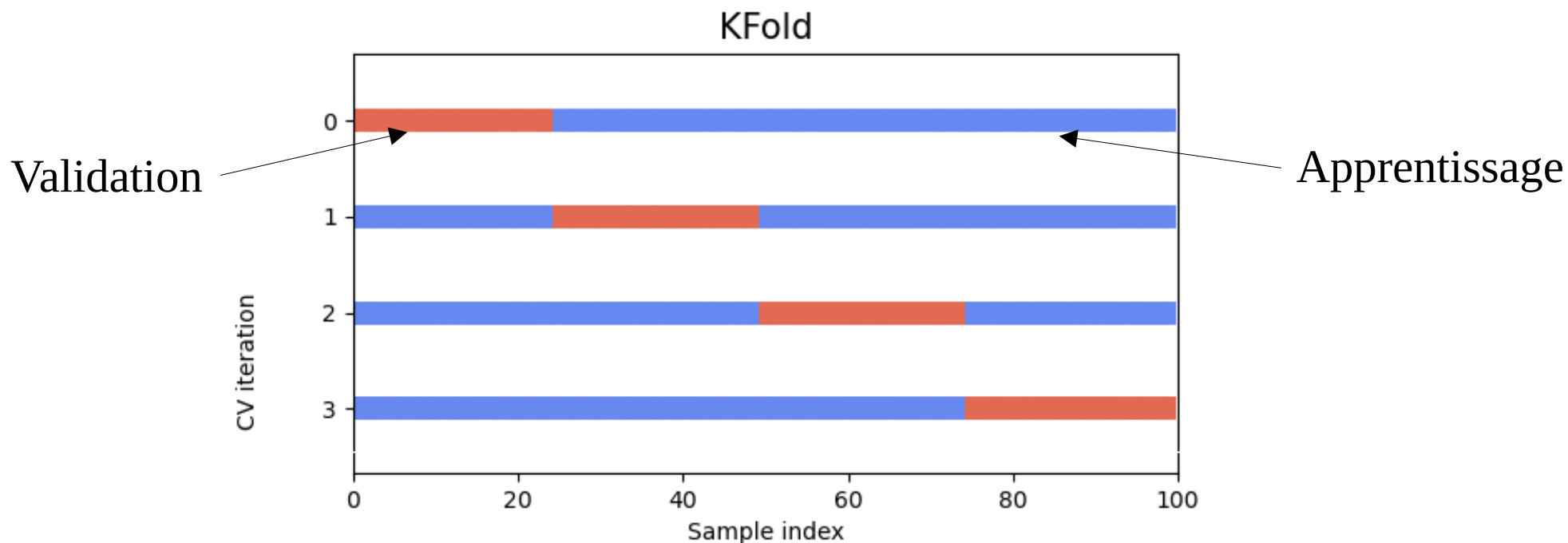
Utilisation du **run de 00h uniquement** lors de la phase d'apprentissage.

Utilisation du **run de 18h** pour le bulletin du matin en hiver et du **run de 06h** pour le bulletin du soir **en OPER**.

Utilisation des données **entre 07/2017 et 12/2022** :

- Apprentissage : [07/2017, 12/2021]
- Test : [01/2022, 12/2022]

Mise en place d'une validation croisée (à 4 folds) sur les données d'apprentissage pour **éviter le sur-apprentissage**.



Le brouillard est un phénomène rare, même s'il est souvent présent dans la vallée de la Seine (~ 1 jour sur 5) → problème **unbalanced**.

Pour augmenter « artificiellement » le jeu de données, on utilise les **4 points de prévision Arome les plus proches des 8 stations** du Pilotage de la Seine.

Implémentation de **2 algorithmes différents** :

- **Forêt Aléatoire (RF)**
- **Réseau de neurones (DNN)**

Évaluation (uniquement!) à partir de **2 modèles différents** :

- **Arome**
 - **Arome-IFS**
-

- **Accuracy** : capacité d'un algorithme à réaliser une bonne prévision.
- **Précision** : capacité d'un algorithme à réaliser une bonne prévision parmi les cas positifs proposés par l'algorithme. Une précision de 1 signifie donc que **le nombre de fausses alarmes est nul**.

$$\text{Précision} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Positifs}}$$

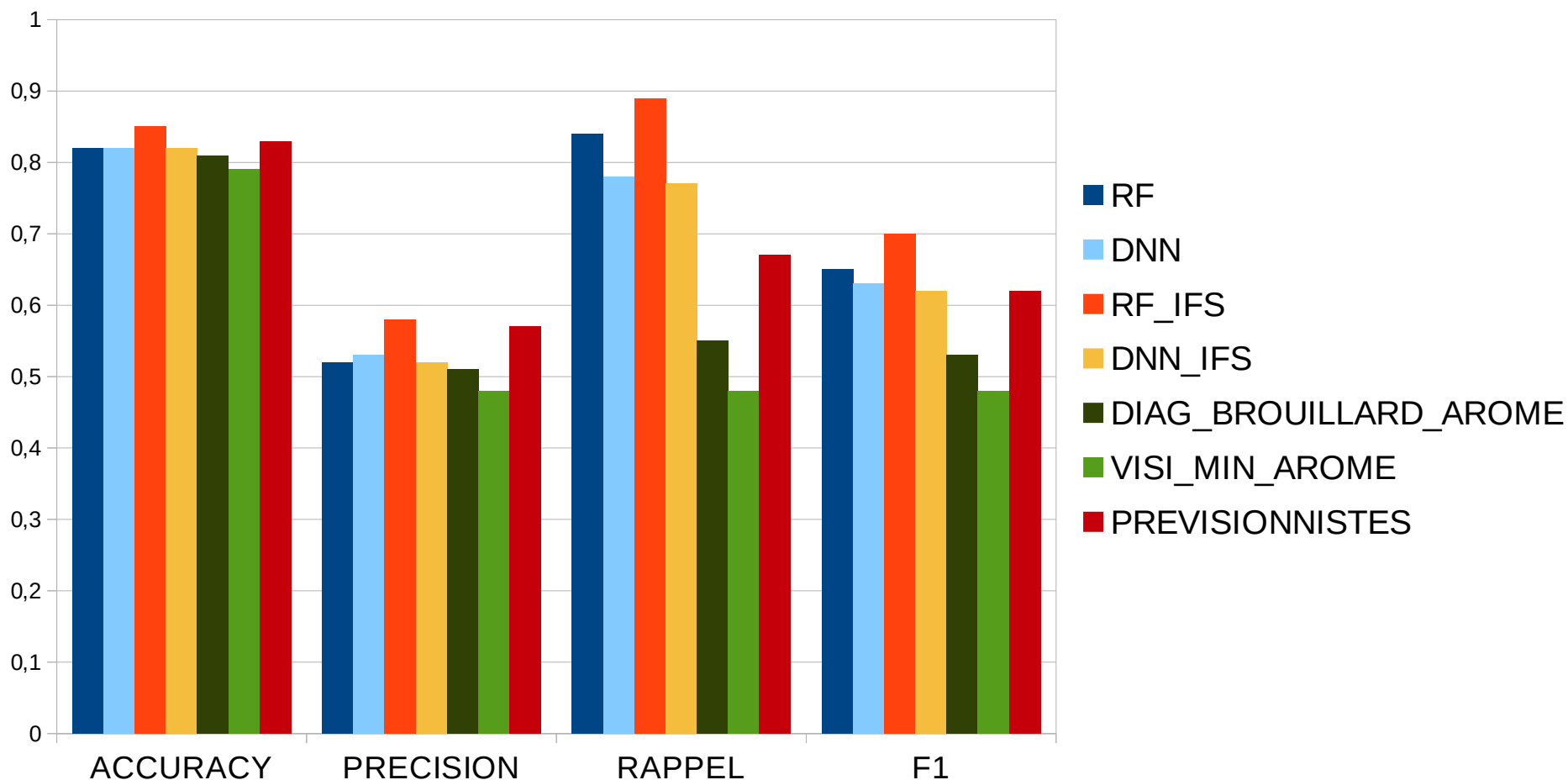
- **Recall** : capacité d'un algorithme à réaliser une bonne prévision parmi les cas réellement positifs. Un rappel de 1 signifie donc que **le nombre de non-détections est nul**.

$$\text{Rappel} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Négatifs}}$$

- **F1** : combinaison entre le rappel et la précision d'un algorithme.

$$F1 = \frac{2 * (\text{Précision} * \text{Rappel})}{\text{Précision} + \text{Rappel}}$$

Comparaison des scores 2022 (bulletin après-midi)



Prévisions agrégées sur les périodes couvertes par les bulletins pour pouvoir se comparer équitablement aux prévisionnistes.

Il y a occurrence de brouillard à l'échelle du bulletin que si l'**algorithme prévoit un minimum d'heures de brouillard** (3h pour DNN et 4h pour RF)

Études de cas

Date du Bulletin	Observation	Type probable	DNN (h)	DNN_IFS (h)	RF (h)	RF_IFS (h)	MSB Brouillard (risque)	APPORT
09/01	Non Occurrence mais brume	Affaissement de stratus / précipitations	9	9	5	7	0	NON Fausse alarme mais présence de brume
16/01	Occurrence (9h)	Affaissement de stratus	4	0	10	7	0	OUI
26/01	Non Occurrence mais brume	Affaissement de stratus	1	13	2	0	1	OUI
27/01	Occurrence (8h)	Affaissement de stratus / précipitations	2	4	11	9	0	OUI
10/02	Occurrence (4h)	Rayonnement	1	0	2	3	1	NON Non Détection
12/02	Non Occurrence		1	1	6	10	0	NON Fausse alarme
13/03	Non Occurrence mais brume	Rayonnement	3	2	3	4	1	OUI Faible Risque de brouillard
20/03	Non Occurrence mais brume (faible)	Rayonnement	0	2	0	2	1	OUI
16/04	Occurrence (2h)	Rayonnement / Evaporation	0	0	0	6	1	NON Non détection majoritaire
05/05	Occurrence (3h)	Rayonnement	8	9	9	8	0	OUI

Les indices présentent un **apport 11 fois sur les 18 études de cas** menées pendant l'étude.

Conclusion & Perspectives

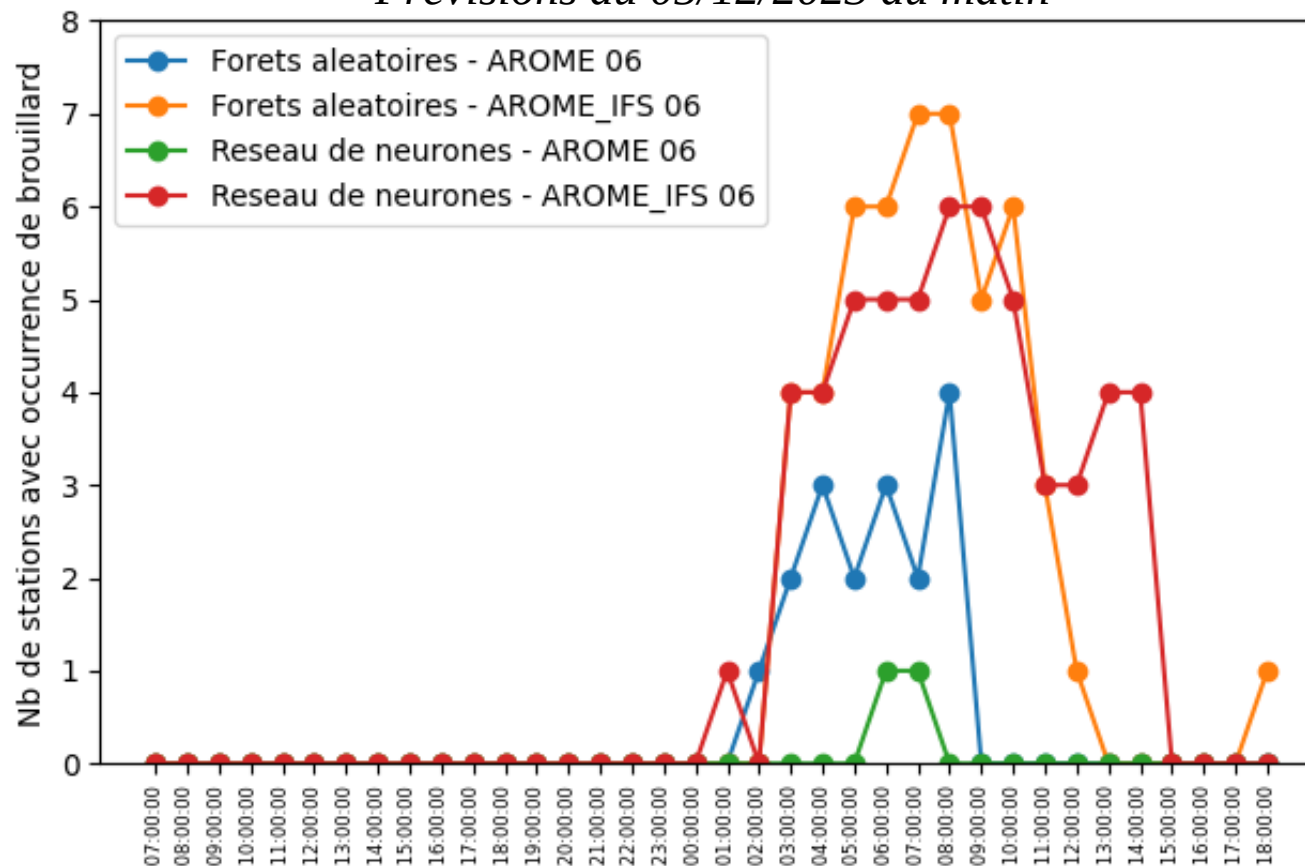
- Le modèle basé sur des **Forêts Aléatoire alimenté par Arome-IFS est le meilleur.**
 - **Information probabiliste** donnée au prévisionniste → nombre de stations ou les indices prévoient du brouillard.
 - Statistiquement, les algorithmes régissent souvent à tort (environ 1 jour sur deux) → **une séquence consécutive de plusieurs heures est indicative.**
 - Suite aux études de cas, **des recommandations ont été faites aux prévisionnistes** sur le comportement des différents algorithmes.
 - Réaliser un **apprentissage à partir des données utilisées en OPER** fournirait très probablement des résultats encore plus performants.
 - **Les prévisionnistes de la DIRO remplissent un document chaque jour** sur l'apport (ou non) des algorithmes.
-

Un produit aujourd'hui opérationnel

Envoi **biquotidien par mail des prévision sous forme d'image**, pour le bulletin MSB du matin ou celui de l'après-midi :

- à 8h32 loc. pour le bulletin du matin
- à 16h32 loc. pour le bulletin de l'après-midi

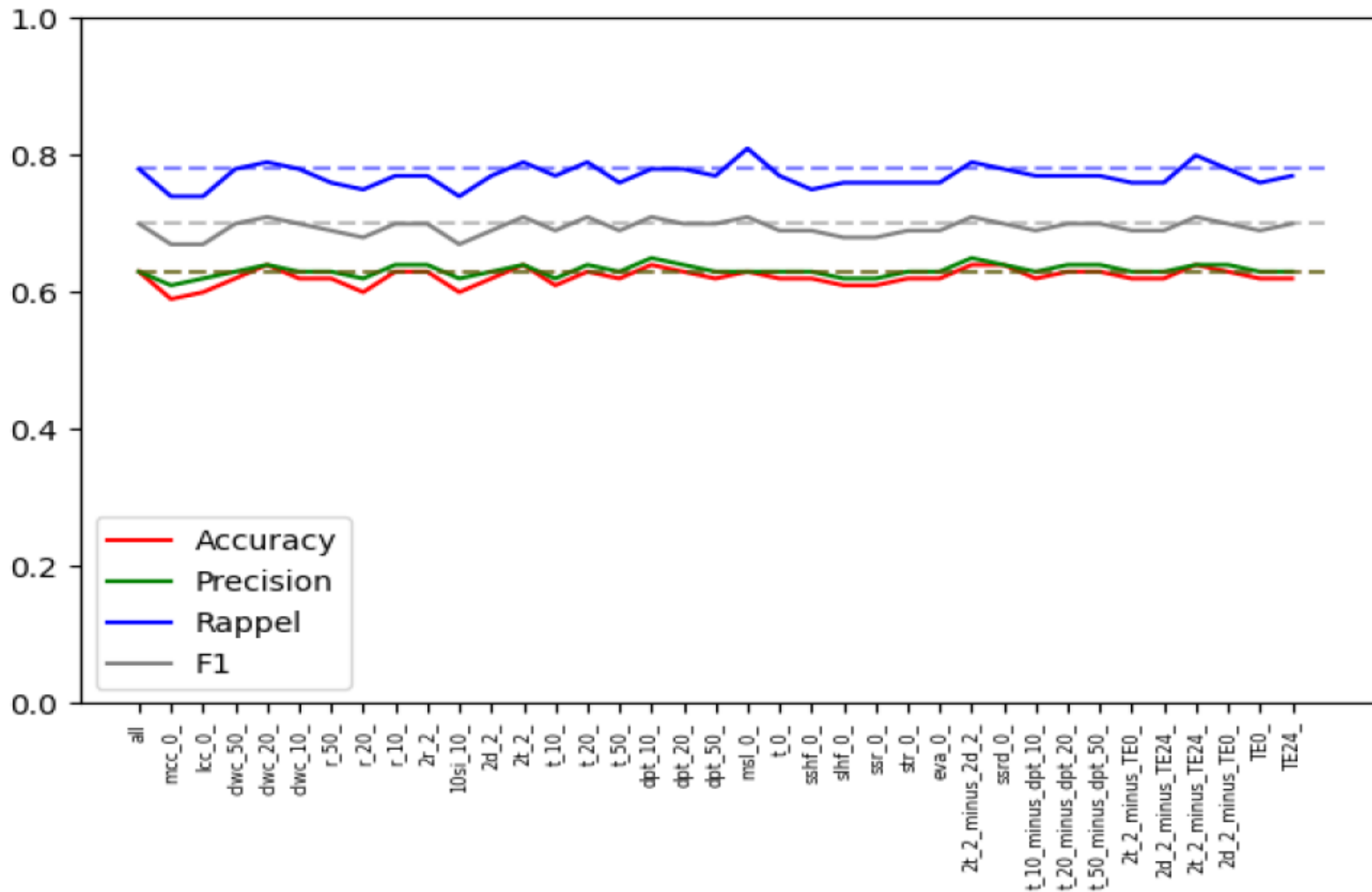
Prévisions du 05/12/2023 au matin



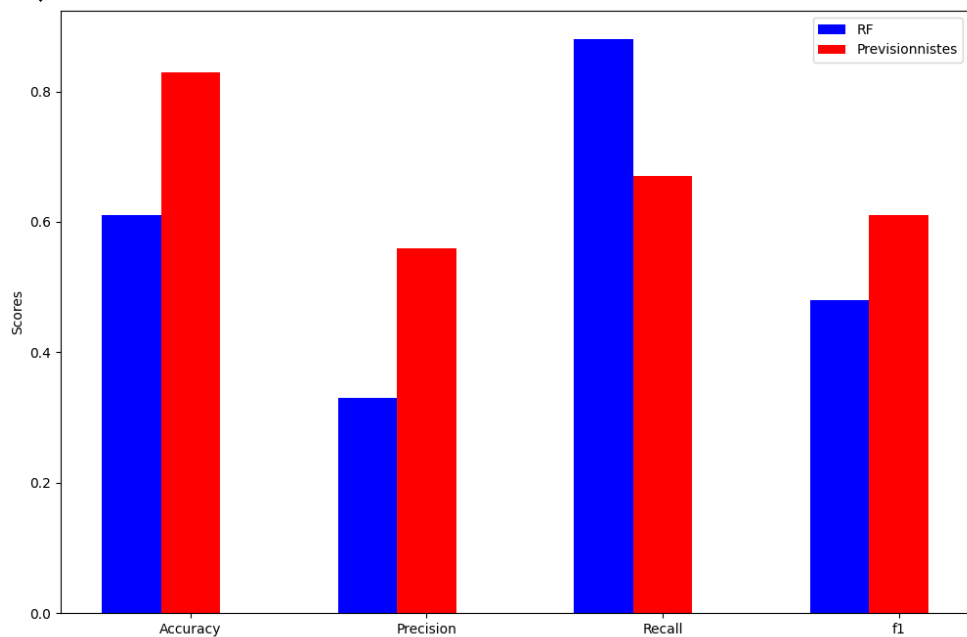
Merci !

Annexes

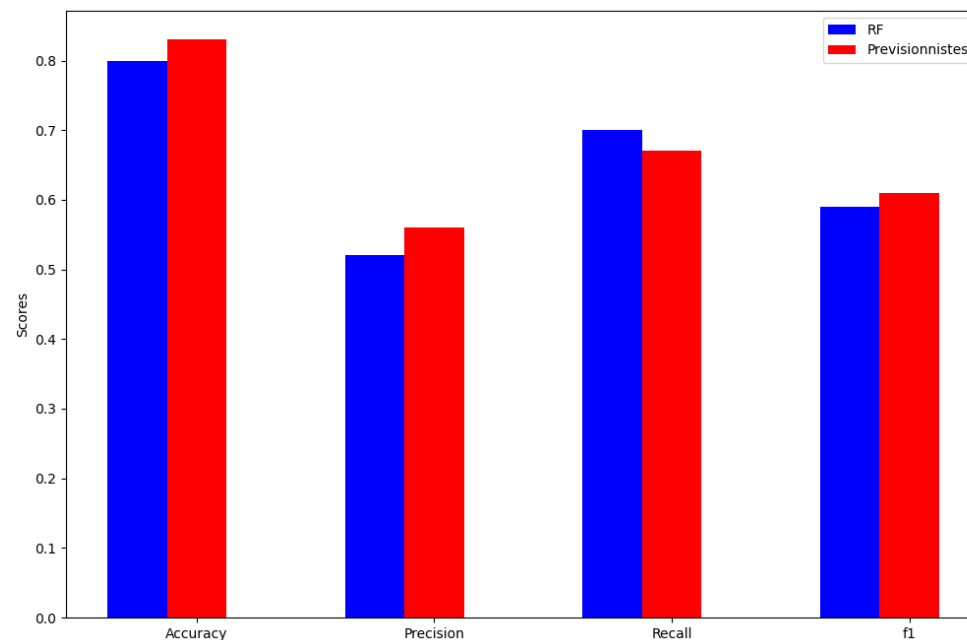
Choix des prédicteurs (Forêt Aléatoire)



Définition du seuil

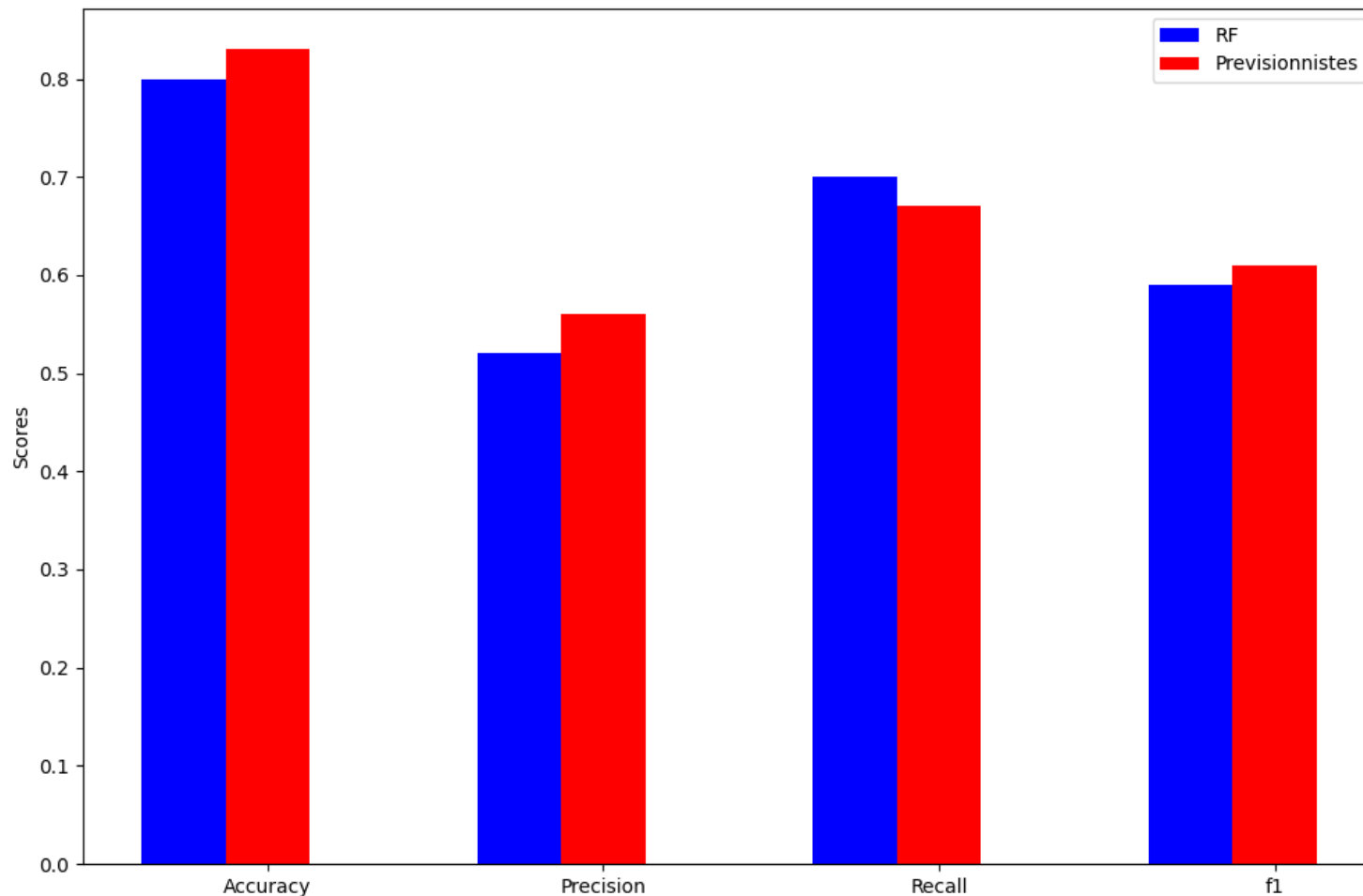


$y_pred_sum \leq 2 \rightarrow 0$



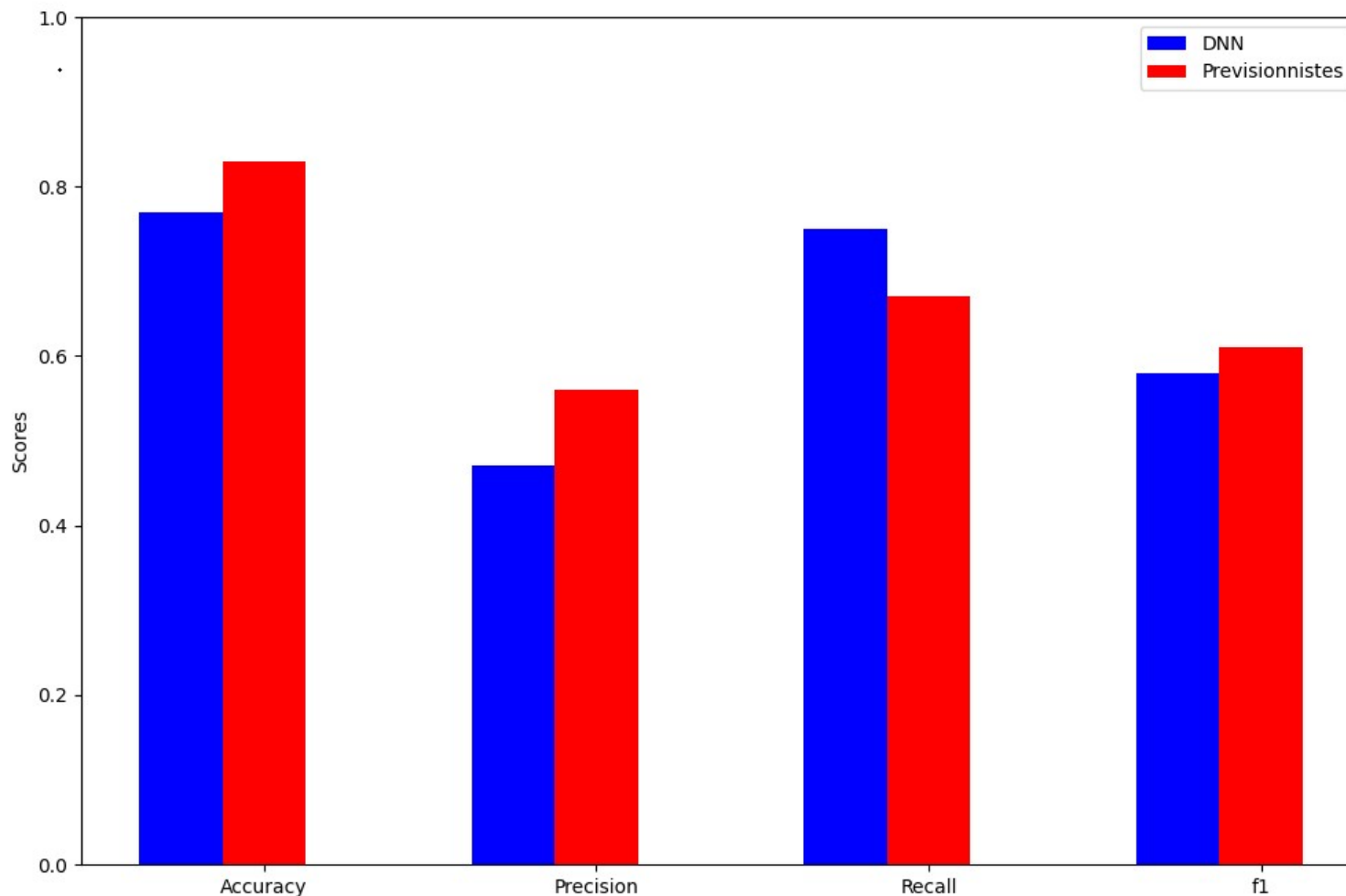
Mise en évidence du phénomène de sur-réaction des indices. On recommande donc aux prévisionnistes d'observer une prévision sur plusieurs heures consécutives pour éviter le nombre de fausses alarmes.

Forêt Aléatoire



Prédicteurs non pris en compte : msl, clwc_20, 2t_2, t_20, dpt_10, 2t_2_minus_2d_2, 2t_2_minus_TE24 et tous les paramètres liés à TE0

Réseaux de neurones



Prédicteurs non pris en compte : 2d_2_minus_TE24, 2t_2_minus_TE24, t_20_minus_dpt_20, 2t_2_minus_2d_2, TE24, eva_0, str_0, ssr_0, msl_0, dpt_50, dpt_10, t_50, t_10, clwc_50, r_50, r_10, et tous les paramètres liés à TE0