



# **Évaluation des performances des prévisions de la qualité de l'air**

**Année 2022**

**- Juin 2023 -**

# 1. SYNTHÈSE

Conformément à l'arrêté du 21 avril 2021, ce document présente le bilan des performances de prévisions d'Airparif pour l'année 2022, selon le guide harmonisé du Référentiel Technique National (RTN).

Les scores définis dans le RTN sont focalisés sur la prévision des dépassements des seuils d'information et d'alerte. De ce fait, ils concernent les prévisions des concentrations maximales en ozone sur la période estivale et des concentrations moyennes journalières en particules PM<sub>10</sub> sur l'année entière. Au-delà de la prévision des épisodes de pollution, les AASQA ont une mission d'information quotidienne de la qualité de l'air. Les scores préconisés dans le RTN ne permettent que partiellement d'évaluer la qualité de l'information délivrée quotidiennement. De ce fait, les scores ont également été calculés sur l'année entière pour l'ozone, ainsi que pour le NO<sub>2</sub> et pour les PM<sub>2.5</sub>. Les résultats de ces évaluations sont disponibles sur demande.

## 1.1. Ozone

### 1.1.1. Par rapport aux objectifs de qualité

Pour l'ozone (O<sub>3</sub>), les scores sont établis en maximum journalier, sur la période d'avril à septembre inclus. Ils sont établis sur toutes les stations de fond pour la région Île-de-France avec les objectifs de qualité suivants à respecter : RMSE : 25 µg/m<sup>3</sup>, Biais : ± 10 µg/m<sup>3</sup>, R : 0.7 (voir définitions en Partie 3).

Pour la prévision à J0 (prévision pour le jour-même), **les scores RMSE, biais et corrélation sont largement conformes aux objectifs de qualité définis par le Référentiel Technique National, sauf pour le biais moyen du modèle ESM\_JP7**, qui montre une surestimation des niveaux sur la période estivale. C'est également le cas pour l'échéance J1.

**Les meilleurs scores sont observés pour le modèle PREVI\_STAT, suivi par la chaîne de prévision analysée par le prévisionniste (ICAR), ce qui montre la capacité des prévisionnistes à faire le bon choix de modèle.**

**Les scores sont globalement homogènes sur les différentes typologies de stations, en particulier pour la prévision du jour même. L'analyse par échéance montre une bonne stabilité des scores de prévision entre J0 et J2.**

**Le calcul des scores par gamme de concentrations montre que la plupart des modèles de prévision ont tendance à surestimer en moyenne les faibles niveaux (entre +8 et +20 µg/m<sup>3</sup> quand C < 70 µg/m<sup>3</sup>), en particulier la chaîne ESM\_JP7 et à sous-estimer les plus forts niveaux (de -10 à -30 µg/m<sup>3</sup> quand C est supérieure à 150 µg/m<sup>3</sup>). Pour les concentrations supérieures à 130 µg/m<sup>3</sup>, le modèle qui présente le biais le plus faible est la chaîne ESM\_JP7, suivie de la chaîne PREVI\_STAT. Ces figures montrent des résultats très contrastés selon les concentrations, qui peuvent être masqués dans les scores moyens. Ainsi, la chaîne ESM\_JP7 ne respecte pas l'objectif fixé sur le biais moyen, du fait de la surestimation des valeurs faibles/moyennes, qui représentent plus de 60 % des données. Néanmoins, c'est le modèle qui prévoit le mieux les concentrations élevées, soit celles qui posent le plus d'enjeu en termes de prévision des épisodes de pollution.**

**La comparaison des scores entre 2021 et 2022 montre une stabilité pour le RMSE, et une légère amélioration pour le biais et les corrélations pour l'ensemble des modèles.**

La différence des scores entre la chaîne ICAR et le modèle PREVI\_STAT, qui constitue désormais la prévision de référence pour les prévisionnistes, peut s'expliquer en partie par des biais méthodologiques de

correction des cartes. **Des pistes d'amélioration ont été identifiées** (voir Partie 3) **et seront mises en œuvre dans le cadre de la refonte de la chaîne de calcul des cartes de prévision.**

### 1.1.2. Prévision des dépassements des seuils réglementaires

Au cours de l'été 2022, 3 vagues de chaleur se sont succédées en juin, juillet et août et ont entraîné 5 dépassements du seuil d'information en ozone (fixé à 180 µg/m<sup>3</sup> en moyenne horaire). Ce nombre de jours d'épisode est relativement faible au regard des températures extrêmes mesurées en Île-de-France. Ce constat s'explique par le fait que lors des journées de fortes chaleurs, les masses d'air provenaient essentiellement d'ouest, ce qui a limité l'import de polluants essentiels à la formation de l'ozone. De plus, lors de certaines journées, les températures étaient certes très élevées, jusqu'à 40,5 degrés Celsius, mais la présence de nuages a limité la formation de fortes concentrations d'ozone.

Date du dépassement	Polluant	Seuil	Prévu ou pas	Réalisation de dépassement selon les critères de l'arrêt interpréfectoral	
15/06/2022	O3	Information	NON	OUI	Détection manquée
16/06/2022	O3	Information	NON	OUI	Détection manquée
17/06/2022	O3	Information	OUI	OUI	Bonne détection
18/06/2022	O3	Information	OUI	NON	Fausse détection
13/07/2022	O3	Information	OUI	NON 179 µg/m <sup>3</sup> selon les critères AMU	Fausse détection mais prévision de dépassement justifiée
18/07/2022	O3	Information	OUI	OUI	Bonne détection
25/08/2022	O3	Information	NON	OUI	Détection manquée

Figure 1 : Ozone – Liste des évènements de pollution constatés et/ou prévus au cours de l'année 2022.

**À J1, les modèles présentent plus de détections manquées que de fausses alertes, ce qui confirme la tendance à la sous-estimation des valeurs extrêmes. La chaîne ESM\_JP7 et la prévision officielle présentent les meilleurs taux de bonnes détections (2/5). Les prévisionnistes ont réalisé une fausse alerte de plus, mais la prévision de dépassement restait pertinente d'un point de vue information, la valeur maximale ayant atteint 179 µg/m<sup>3</sup> sur la région.**

## 1.2. Particules PM<sub>10</sub>

### 1.2.1. Par rapport aux objectifs de qualité

Les scores classiques (Biais, RMSE et Corrélation) sont établis pour les particules PM<sub>10</sub>, en moyenne journalière, sur l'année entière.

**Tous les modèles respectent les objectifs de qualité fixés par le Référentiel Technique National pour les scores moyens. Les modèles présentant les meilleurs scores sont la PREVI\_STAT et la prévision expertisée par les prévisionnistes (ICAR).**

**Les scores sont globalement similaires pour toutes les typologies de stations.**

**Les scores de l'ensemble des chaînes sont globalement stables de J0 à J2, avec une légère amélioration à J0 pour les chaînes PREVI\_STAT et ICAR.**

L'analyse de scores moyens par gamme de concentrations montre que **pour les gammes de concentrations supérieures à 20 µg/m<sup>3</sup>, tous les modèles présentent une tendance à la sous-estimation des niveaux en PM10**, avec des biais moyens négatifs qui augmentent plus ou moins sensiblement selon les chaînes, notamment la chaîne ESM\_JP7 (-10 µg/m<sup>3</sup>) pour des concentrations comprises entre 40 et 50 µg/m<sup>3</sup>.

**Les scores de 2022 sont proches de l'année 2021, avec une amélioration des corrélations pour l'ensemble des modèles.** En revanche, un biais négatif apparaît pour les chaînes ESM-HR et ESM\_BR, qui peut s'expliquer par le fait que le module d'érosion n'a pas été branché entre avril et septembre (comme c'est le cas habituellement) - ce qui conduit à une sous-estimation des dusts (poussières liées à la remise en suspension), un effet accentué par la sécheresse qu'a connu la France en 2022.

### 1.2.2. Prévision des dépassements des seuils réglementaires

Bien que l'hiver ait été globalement doux, des **conditions météorologiques ponctuellement froides et peu dispersives**, survenues notamment au cours du mois de janvier, **ont entraîné 2 dépassements du seuil d'information-recommandations en PM<sub>10</sub>** (fixé à 50 µg/m<sup>3</sup> en moyenne journalière). 3 dépassements de ce seuil réglementaire ont été constatés au mois de mars 2022, du fait des conditions printanières favorables à la formation d'aérosols inorganiques secondaires, auxquels se sont ajoutées des conditions météorologiques très peu dispersives.

Date du dépassement	Polluant	Seuil	Prévu ou pas	Réalisation de dépassement selon les critères de l'arrêt inter préfectoral	
14/01/2022	PM10	Information	NON	OUI	Détection manquée
15/01/2022	PM10	Information	OUI	OUI	Bonne détection
25/01/2022	PM10	Information	OUI	NON 52 µg/m <sup>3</sup> en stations 48 µg/m <sup>3</sup> selon les critères AMU	Fausse détection mais prévision de dépassement justifiée
24/03/2022	PM10	Information	NON	OUI	Détection manquée
25/03/2022	PM10	Information	OUI	OUI	Bonne détection
26/03/2022	PM10	Information	OUI	OUI	Bonne détection
27/03/2022	PM10	Information	OUI	NON	Fausse détection
28/03/2022	PM10	Information	OUI	NON 52 µg/m <sup>3</sup> en stations 50 µg/m <sup>3</sup> selon les critères AMU	Fausse détection mais prévision de dépassement justifiée

Figure 2 : Particules PM<sub>10</sub> – Liste des évènements de pollution constatés et/ou prévus au cours de l'année 2022.

**A J1, les chaînes de prévision présentant le plus grand nombre de bonnes détections selon les critères de l'AMU sont PREVI\_STAT et la prévision officielle** (3 bonnes détections sur 5 jours d'épisodes), suivies des chaînes ESM\_HR, BR et JP7 (2/5).

**Les prévisionnistes ont également réalisé le plus faible nombre de fausses alertes (3 FA), comme la chaîne JP7.** Sur ces 3 fausses alertes, les concentrations maximales ont néanmoins atteint 52 µg/m<sup>3</sup> en station, et respectivement 48 et 50 µg/m<sup>3</sup> selon les critères de l'arrêt. **La prévision de dépassement était donc justifiée d'un point de vue information du public et n'est pas une erreur de prévision.**

**Ces scores montrent la capacité des prévisionnistes à faire le bon choix de modèle lors des prévisions d'épisode.**

## 2. PRÉSENTATION DE LA PLATEFORME DE PRÉVISION

### 2.1. Modèles de qualité de l'air

Les scores présentés concernent l'ensemble des modèles de qualité de l'air disponibles pour les prévisionnistes d'Airparif.

#### 2.1.1. Les modèles bruts :

Pour leur prévision quotidienne, les prévisionnistes ont à leur disposition les sorties de 3 modèles bruts de prévision développés par Airparif au sein de la plateforme de prévision inter-régionale ESMEALDA : ESMEALDA Haute Résolution (ESM\_HR), ESMEALDA Basse Résolution (ESM\_BR), qui délivrent des prévisions de J0 à J+2, et ESMEALDA J+7 (ESM\_JP7), qui délivre des prévisions pour les 7 jours à venir. Le tableau ci-dessous décrit les caractéristiques de ces différents modèles.

	ESM_HR	ESM_BR	ESM_JP7
Modèle	Chimie-Transport : Chimère	Chimie-Transport : Chimère	Chimie-Transport : Chimère
Résolution	3 km	15 km	15 km
Météo	MM5	MM5	MM5
Conditions aux limites	PREV' AIR_BR A défaut CONT4*	PREV' AIR_BR	CONT4

\*CONT4 : modèle continental

En complément, les prévisionnistes d'Airparif utilisent également les sorties du modèle de prévision d'ensemble (PREVI\_STAT). Ce modèle de prévision statistique utilise les modèles COPERNICUS (chimere, dehm, emep, euradim, gemaq, lotos, match, minni, mocage, monarch, silam), PREV'AIR (Haute et Basse Résolutions), PREVEST, AIRES et les modèles AIRPARIF.

Le modèle PREVAIR Basse Résolution (PREVAIR\_BR) alimente les conditions aux limites des modèles bruts ESM\_HR et ESM\_BR. Le [modèle PREVAIR Haute Résolution](#) (PREVAIR\_HR) constitue un modèle supplémentaire intéressant que peuvent utiliser les prévisionnistes d'Airparif.

#### 2.1.2. Post-traitements : la carte corrigée par le prévisionniste (ICAR)

La carte de prévision diffusée sur le site internet d'Airparif est générée par la chaîne de cartographie développée par Airparif, appelée ICAR, après analyse du prévisionniste. La carte est à une résolution de 1 km.

Lors de sa prévision, le prévisionniste a pour mission :

- de prévoir la fourchette des concentrations maximales prévues pour chacun des polluants (NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub> et PM<sub>2.5</sub>).

- de choisir, parmi les modèles présentés ci-dessus, celui qui est le plus pertinent au regard de son analyse, et qui servira au motif de base pour la carte. Ce choix peut être réalisé indépendamment pour chaque polluant et pour chaque échéance.

Si la valeur maximale prévue par le modèle choisi se trouve dans la fourchette de concentrations maximales prévue par le prévisionniste, aucune correction n'est appliquée sur la carte du modèle brut. Si la valeur maximale prévue par le modèle se situe en-dehors de la fourchette de concentrations maximales prévue par le prévisionniste, la valeur maximale modélisée est ramenée à la valeur moyenne de l'intervalle de concentration prévu.

## 2.2. Disponibilité des prévisions sur l'année

Le tableau ci-dessous présente le taux de disponibilité des prévisions de qualité de l'air fournies par les différents modèles :

Modèle	ESM_BR	ESM_HR	ESM_JP7	PREVI_STAT	ICAR
Disponibilité	100 %	100 %	96 %	97 %	100%

## 2.3. Descriptif des observations de fond

Le réseau de mesure d'Airparif compte en 2022 :

- 22 stations de mesure de l'ozone, dont 10 stations urbaines, 5 péri-urbaines et 7 stations rurales régionales ;
- 11 stations de mesure de fond en PM<sub>10</sub> dont 7 stations urbaines, 2 périurbaines et 2 stations rurales régionales.

La procédure d'information et d'alerte étant applicable à l'échelle régionale, les analyses de scores ne sont pas réalisées à l'échelle du département pour l'Île-de-France.

## 2.4. Évolutions des chaînes en 2022 pouvant impacter les prévisions

En 2022, les principales évolutions sur la chaîne ESMEALDA qui peuvent impacter les scores par rapport aux années précédentes ou qui peuvent expliquer des écarts de scores entre la chaîne JP2 et la chaîne JP7 sont liées à la mise à jour d'ESMEALDA de novembre 2021, avec 2 évolutions majeures pour la chaîne JP2 :

- Les nouvelles conditions aux limites PREVAIR : abaissement pour PREVAIR et ESMEALDA de la hauteur de la première couche et inventaires plus récents que ceux exploités dans ESMEALDA pour PREVAIR ;
- L'intégration des développements réalisés dans le cadre du projet NUAGE et validés dans le cadre du projet AMMON sur la prise en compte des émissions de NH<sub>3</sub> par l'agriculture.

La chaîne JP7, qui exploite les conditions aux limites calculées par la chaîne ESMEALDA n'a pas subi d'évolutions.

## 3. ÉVALUATION DE LA PLATEFORME DE PRÉVISION

L'évaluation des performances de prévisions de la qualité de l'air a été réalisée pour l'ozone (O<sub>3</sub>) en période estivale, ainsi que pour les particules PM<sub>10</sub> conformément au RTN. Un travail comparable a également été mené pour l'ozone sur l'année entière, les particules PM<sub>2,5</sub> et le dioxyde d'azote (NO<sub>2</sub>) afin de pouvoir évaluer les performances de la prévision quotidienne.

Trois indicateurs classiques de performances (appelés « scores ») sont considérés :

1. Le **Biais** : différence entre la valeur attendue d'un estimateur (prévision) et la vraie valeur du paramètre estimé (observation) exprimée en µg/m<sup>3</sup>. **Un estimateur avec un biais positif illustre une surestimation des niveaux prévus, alors qu'un biais négatif signifie que l'estimateur les sous-estime. Cet écart doit être le plus petit possible et tendre vers zéro (sans biais / biais nul).**
2. La racine de l'écart quadratique moyen (**Root Mean Square Error – RMSE**) correspond à l'écart-type des résidus (ou erreurs de prévision), exprimé en µg/m<sup>3</sup>. Les résidus sont la mesure de l'écart entre les données prévues versus observées avec la droite de régression linéaire. Le RMSE fournit donc une indication par rapport à la dispersion de ces résidus ou la variabilité de la qualité du modèle à reproduire les niveaux de concentrations observés. **Il a toujours une valeur positive et doit être le plus faible possible.**
3. La **corrélation r** : mesure statistique qui étudie le degré de liaison linéaire entre les observations et les prévisions. Elle donne une information sur la cohérence des variabilités temporelle et spatiale des concentrations observées et prévues. **Cette valeur r doit être la plus proche de 1.**

Leurs méthodes de calcul sont définies dans l'annexe 1 du « Document de référence du dispositif national pour la prévision de la qualité de l'air » du Référentiel Technique National.

### 3.1. Prévision ozone (O<sub>3</sub>)

#### 3.1.1. Scores classiques (Biais, RMSE et corrélation)

Ces scores sont établis pour l'ozone, en max journalier, sur la période d'avril à septembre inclus.

##### 3.1.1.1. Par rapport aux objectifs de qualité

Les scores sont établis sur toutes les stations de fond pour la région Île-de-France avec les objectifs de qualité suivants à respecter : RMSE : 25 µg/m<sup>3</sup>, Biais : ± 10 µg/m<sup>3</sup>, R : 0.7.

En Île-de-France, l'arrêté inter-préfectoral relatif à la mise en place de mesures d'urgence en cas d'épisodes de pollution concerne l'ensemble de la région et n'est pas applicable au niveau des départements. Les scores sont donc calculés à l'échelle de la région entière. Les scores sont une moyenne des scores temporels de chaque station de fond (22 au total).

La Figure 3 représente les scores pour l'échéance J0.

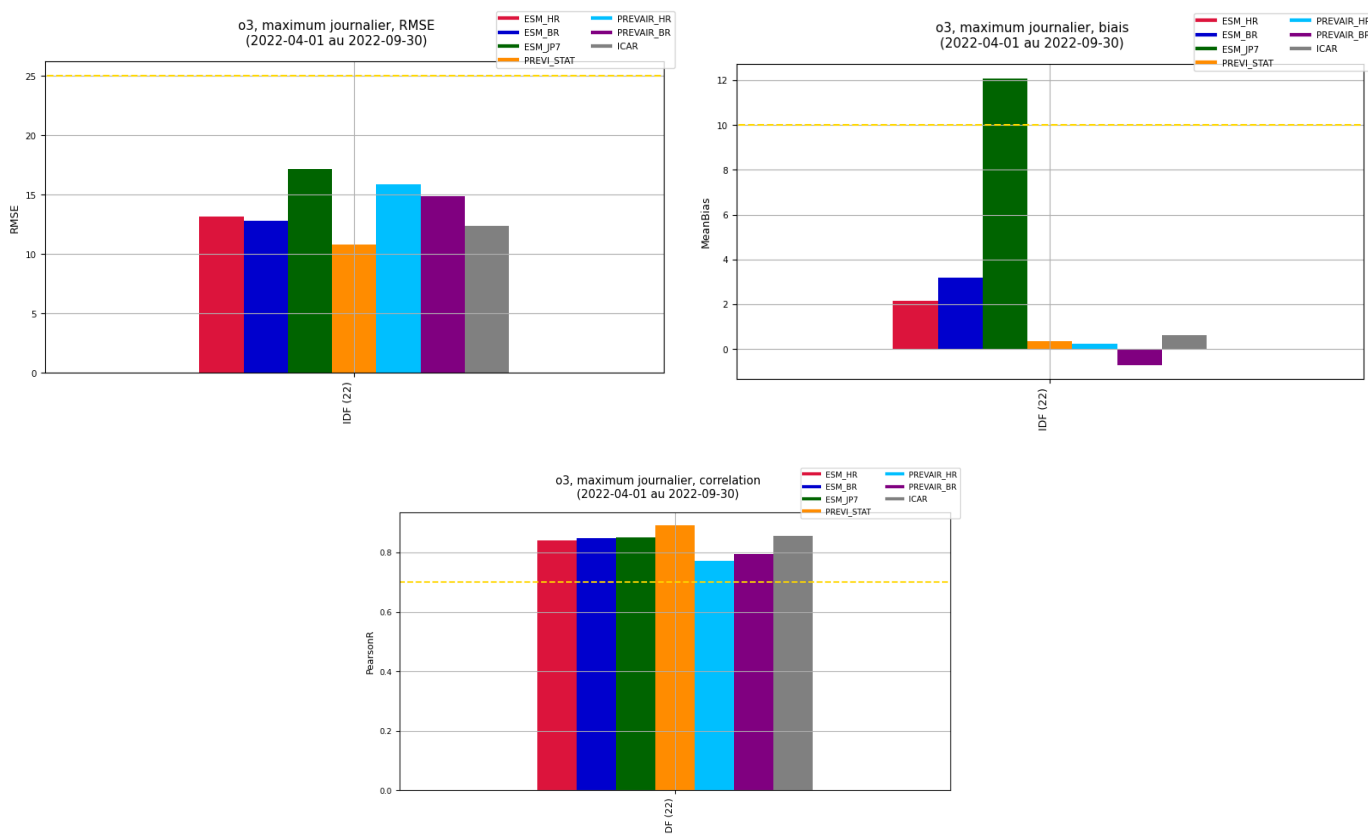


Figure 3 : Ozone – Scores de RMSE, biais et corrélation en Île-de-France des prévisions d'avril à septembre 2022 sur l'échéance J0. La ligne pointillée orange indique les objectifs de qualité à respecter.

**À l'échéance J0, l'ensemble des modèles présentent des RMSE compris entre 11 et 17 µg/m³, largement inférieures à l'objectif de qualité, fixé à 25 µg/m³.** Le modèle PREVI\_STAT présente la RMSE la plus faible. La chaîne ESM\_JP7 présente une RMSE plus élevée que les autres modèles. Cet indicateur moyen doit néanmoins être interprété avec prudence, car la chaîne ESM\_JP7 présente la RMSE la plus faible pour les concentrations élevées, qui présentent l'enjeu le plus fort en termes de prévisions des épisodes de pollution (Figure 8).

**Les modèles de prévision d'Airparif ESM\_HR, ESM\_BR, PREVI\_STAT et ICAR présentent un biais moyen compris entre 0.4 et 3 µg/m³, conforme à l'objectif de qualité défini dans le RTN (10 µg/m³).** Le modèle PREVI\_STAT présente également le biais moyen le plus faible (0,4 µg/m³), suivi de près par la prévision corrigée par les prévisionnistes ICAR (0.6 µg/m³). La chaîne de prévision ESM\_JP7 présente un biais moyen de 12 µg/m³ (non conforme à l'objectif de qualité), illustrant ainsi une surestimation moyenne des niveaux modélisés. Les modèles PREVAIR\_HR et PREVAIR\_BR présentent des biais faibles, respectivement de 0.3 et -0.7 µg/m³.

**Tous les modèles régionaux présentent de très bonnes corrélations, comprises entre 0.84 et 0.89.** La corrélation la plus forte est observée pour la chaîne PREVI\_STAT (r=0.89). Les modèles PREVAIR\_HR et PREVAIR\_BR présentent des corrélations respectives de 0.77 et 0.80. Tous les modèles respectent l'objectif de qualité.



La Figure 4 représente les mêmes scores pour l'échéance J1.

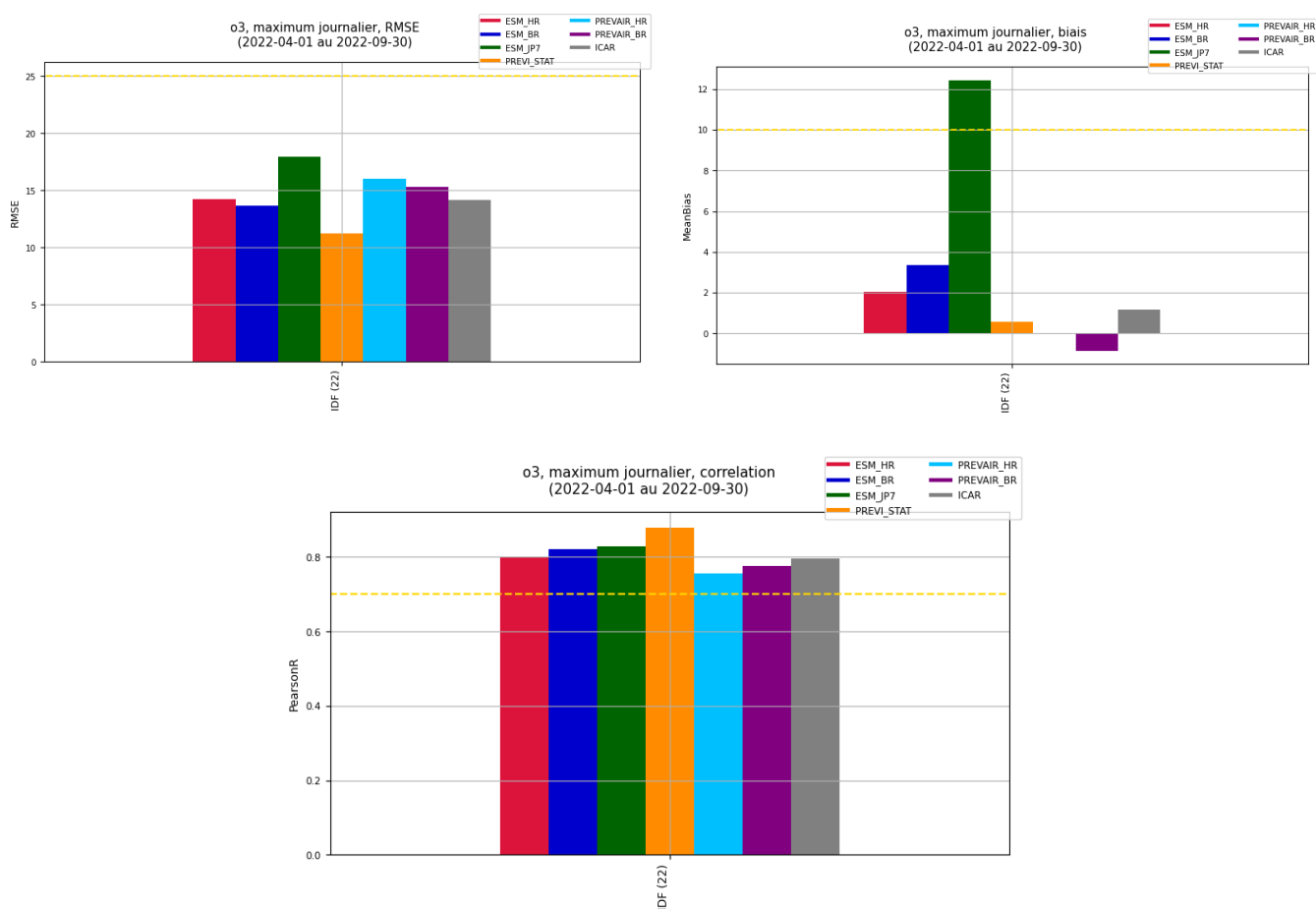


Figure 4 : Ozone – Scores de RMSE, biais et corrélation en Île-de-France des prévisions d'avril à septembre 2022 sur l'échéance J1. La ligne pointillée orange indique les objectifs de qualité à respecter.

À l'échéance J1, l'ensemble des modèles présentent des RMSE compris entre 11 et 18 µg/m³, légèrement plus élevés qu'à l'échéance J0. **Ils restent largement inférieurs à l'objectif de qualité défini dans le RTN.** Au même titre que pour J0, la chaîne PREVI\_STAT présente la RMSE la plus faible et la chaîne ESM\_JP7 présente une RMSE plus élevée que les autres modèles.

**Les modèles de prévision d'Airparif ESM\_HR, ESM\_BR, PREVI\_STAT et ICAR présentent un biais moyen proche de l'échéance J0, compris entre 0.6 et 3 µg/m³. L'objectif de qualité défini dans le RTN est respecté pour cet indicateur.** Le modèle PREVI\_STAT présente le biais moyen le plus faible (0.6 µg/m³), suivi de près par la prévision analysée ICAR (1.2 µg/m³). La chaîne de prévision ESM\_JP7 présente un biais moyen similaire à l'échéance J0 (12 µg/m³), supérieur à l'objectif de qualité. Il faut néanmoins être prudent sur les conclusions tirées de l'analyse de cet indicateur moyen, car la chaîne ESM\_JP7 présente le biais le plus faible pour les concentrations élevées, qui posent le plus d'enjeu en termes de prévision des épisodes de pollution. Les modèles PREVAIR\_HR et PREVAIR BR présentent un biais moyen quasi nul (équivalent à J0).

**L'ensemble des modèles analysés présentent des corrélations supérieures à 0.80, à l'exception de PREVAIR\_HR (0.67).** La corrélation la plus forte est observée pour la chaîne PREVI\_STAT (r = 0.88). La prévision analysée ICAR présente en moyenne la corrélation la plus faible, avec ESM\_HR. **Tous les modèles respectent l'objectif de qualité fixé pour ce paramètre (0.7).**

### 3.1.1.2. Par typologie

Les scores (RMSE, Biais et Corrélation) ont été calculés par station, puis moyennés en appliquant un filtre sur les typologies de stations de fond (urbaines, péri-urbaines et rurales).

La Figure 5 présente les scores pour l'échéance J0.

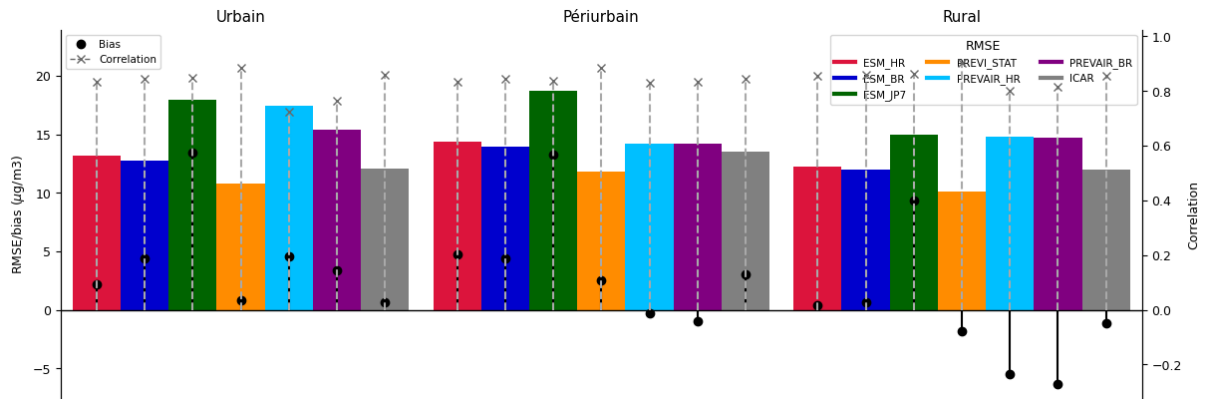


Figure 5 : Ozone – Synthèse des scores de RMSE, biais et corrélation en fonction de la typologie de station (urbaines, périurbaines, rurales) des prévisions d'avril à septembre 2022 sur l'échéance J0.

La Figure ci-dessus montre que :

- les scores sont globalement proches sur les différentes typologies de stations ;
- les RMSE sont globalement plus élevées sur les stations péri-urbaines/urbaines que sur les stations rurales, sauf pour les modèles Prev'Air ;
- les biais moyens sont positifs sur les stations urbaines et périurbaines sur l'ensemble des modèles. Pour les stations rurales, ils sont quasi nuls pour les modèles ESM\_HR et ESM\_BR et négatifs pour les autres modèles, à l'exception d'ESM\_JP7, qui conserve le biais positif le plus fort.
- les corrélations sont globalement comparables sur toutes les typologies de stations pour les modèles ESM\_HR, ESM\_BR, ESM\_JP7, PREVI\_STAT et ICAR. Concernant le modèle PREVAIR\_HR, les corrélations sont légèrement plus fortes en situation de fond rural et péri-urbain qu'en situation de fond urbain.

Pour l'échéance J1, les scores sur les trois typologies de stations sont très proches de ceux calculés à J0 (Figure 6).

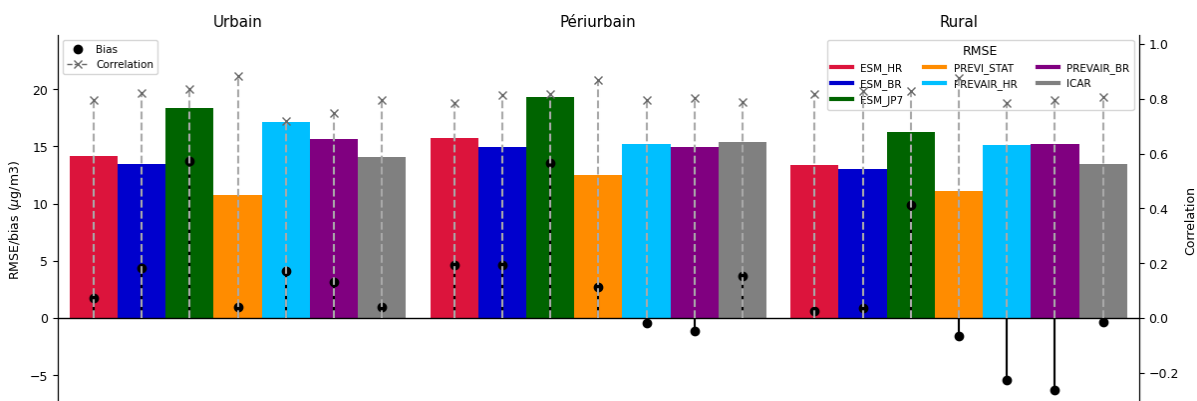


Figure 6 : Ozone – Synthèse des scores de RMSE, biais et corrélation en fonction de la typologie de station (urbaines, périurbaines, rurales) des prévisions d'avril à septembre 2022 sur l'échéance J1.

### 3.1.1.3. Par échéance de prévision

Les scores (RMSE, Biais et Corrélation) ont été calculés par station, puis moyennés sur les différents horizons de prévisions (J0 à J2).

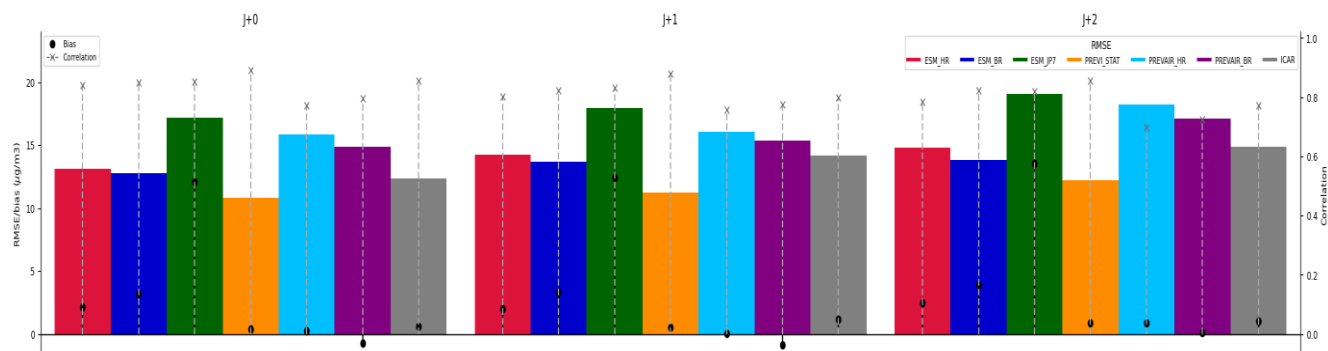


Figure 7 : Ozone – Synthèse des scores de RMSE, biais et corrélation en fonction du jour de prévision d'avril à septembre 2022.

La Figure 7 ci-dessus montre que :

- **les RMSE sont légèrement plus élevées aux échéances J1 et J2** (jusqu'à + 2  $\mu\text{g}/\text{m}^3$ ) par rapport à J0). Pour la chaîne ICAR, la RMSE diminue entre J2, J1 et J0 car les prévisionnistes réajustent les fourchettes de concentrations maximales prévues au regard des observations du matin.
- **les biais restent globalement similaires sur les trois échéances de prévision** ( $\pm 1 \mu\text{g}/\text{m}^3$ ) ;
- **les corrélations apparaissent un peu plus faibles à J2 pour l'ensemble des modèles.**

L'analyse de ces scores montre que **le modèle PREVI\_STAT présente les meilleures performances de prévision toutes échéances confondues** (les plus faibles RMSE/Biais et les plus fortes corrélations). *A contrario*, **la chaîne de modélisation ESM\_JP7 présente les moins bons scores** (plus spécifiquement pour les RMSE et Biais) **sur les trois échéances de prévision. La chaîne de prévision ICAR analysée par le prévisionniste présente des scores intermédiaires entre la PREVI\_STAT et les modèles ESM BR/ESM\_HR.**

**Ces résultats montrent également une bonne stabilité des scores de prévision selon l'échéance.**

### 3.1.1.4. Par gamme de concentration

Les données journalières sont groupées selon leur « classe de concentration » (par rapport aux observations). Pour chaque groupe, les scores sont calculés sur chaque station de fond, puis moyennés.

La Figure 8 présente les indicateurs RMSE et Biais par classes de concentration d'ozone pour l'échéance J0.

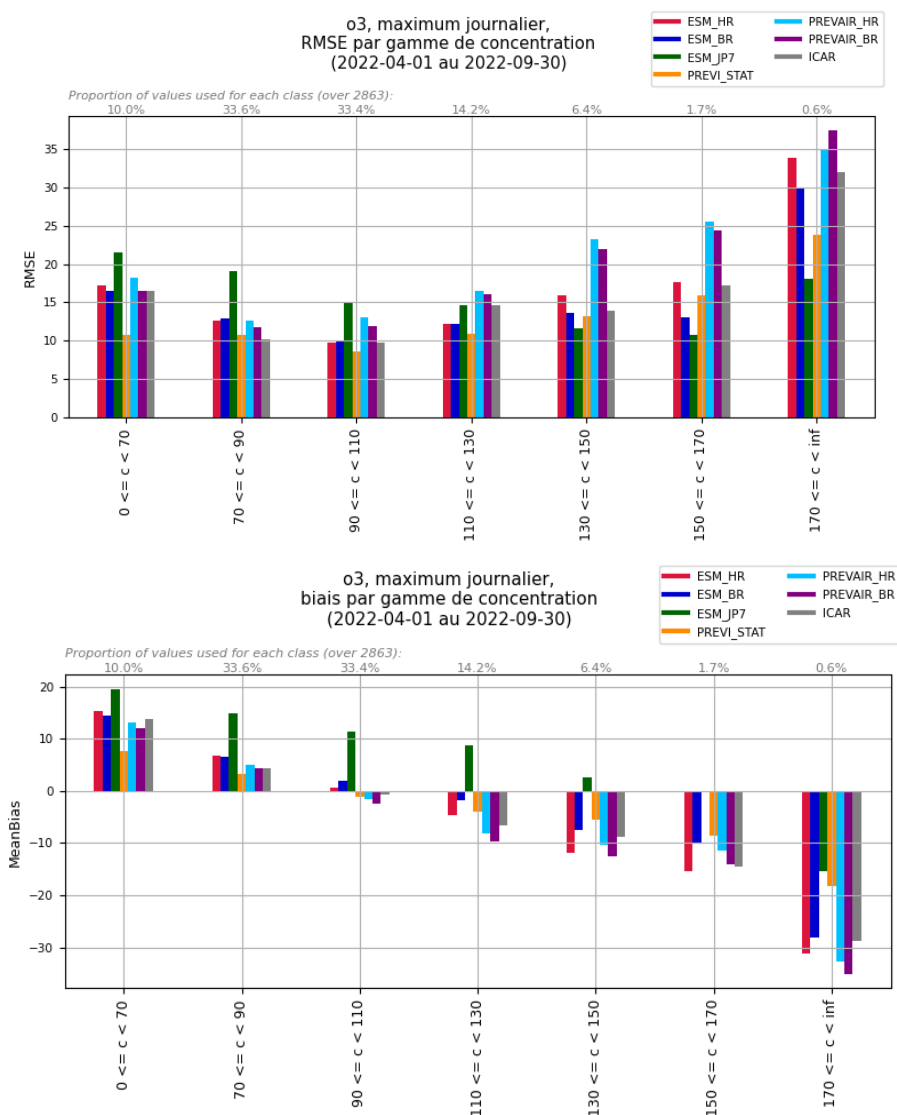


Figure 8 : Ozone – Scores de RMSE et biais par classes de concentration des prévisions d'avril à septembre 2022 sur l'échéance J0. Au-dessus de la figure, le pourcentage de données observées pour chaque classe de concentrations.

**Pour les chaînes de modélisation ESM\_HR, ESM\_BR et ICAR à l'échéance J0, les plus faibles RMSE (de l'ordre de 10-13  $\mu\text{g}/\text{m}^3$ ) concernent les concentrations en ozone comprises entre 70 et 130  $\mu\text{g}/\text{m}^3$  (soit près de 70 % des observations). Pour ces modèles, les RMSE sont légèrement plus élevées (jusqu'à 17  $\mu\text{g}/\text{m}^3$ ) pour des concentrations inférieures à 70  $\mu\text{g}/\text{m}^3$  et entre 130 et 170  $\mu\text{g}/\text{m}^3$  (soit 16 % des observations). En revanche, elles augmentent fortement pour la gamme des concentrations supérieures à 170  $\mu\text{g}/\text{m}^3$  (RMSE = 30  $\mu\text{g}/\text{m}^3$ ), mais le faible nombre d'occurrences (< 1 %) ne permet pas d'évaluer de manière pertinente ce paramètre.**

**Le modèle ESM\_JP7 semble mieux reproduire les plus fortes concentrations (130 < C < 170  $\mu\text{g}/\text{m}^3$  pour une RMSE de l'ordre de 10-12  $\mu\text{g}/\text{m}^3$ ). Pour de plus faibles niveaux (C < 110  $\mu\text{g}/\text{m}^3$ ), les RMSE peuvent être plus importantes que les autres modèles (jusqu'à 22  $\mu\text{g}/\text{m}^3$ ). Pour les modèles PREVAIR\_HR et PREVAIR\_BR, les RMSE augmentent fortement avec les concentrations à partir de 70  $\mu\text{g}/\text{m}^3$ .**

Enfin, la chaîne PREVI\_STAT présente les plus faibles RMSE (~10  $\mu\text{g}/\text{m}^3$ ) pour les gammes de concentrations en ozone inférieures à 130  $\mu\text{g}/\text{m}^3$ . Les RMSE augmentent progressivement lorsque les niveaux sont plus soutenus (jusqu'à 25  $\mu\text{g}/\text{m}^3$ ), mais restent parmi les plus faibles.

L'analyse du biais moyen par classes de concentrations montre que la plupart des modèles de prévision ont tendance à surestimer les faibles niveaux (entre +8 et +20  $\mu\text{g}/\text{m}^3$  quand  $C < 70 \mu\text{g}/\text{m}^3$ ), en particulier la chaîne ESM\_JP7 et à sous-estimer les plus forts niveaux (de -10 à -30  $\mu\text{g}/\text{m}^3$  quand  $C$  est supérieure à 150  $\mu\text{g}/\text{m}^3$ ). Pour les concentrations supérieures à 130  $\mu\text{g}/\text{m}^3$ , le modèle qui présente le biais le plus faible est la chaîne ESM\_JP7, suivie de la chaîne PREVI\_STAT.

Ces figures montrent des résultats très contrastés selon les concentrations, qui peuvent être masqués dans les scores moyens. Ainsi, la chaîne JP7 ne respecte pas l'objectif fixé sur le biais moyen (Figure 4), du fait de la surestimation des valeurs faibles/moyennes, qui représentent plus de 60 % des données. Néanmoins, c'est le modèle qui prévoit le mieux les concentrations élevées, qui posent le plus d'enjeu en termes de prévision des épisodes de pollution.

La Figure 9 présente les indicateurs RMSE et Biais par classes de concentration d'ozone pour l'échéance J1. Les enseignements sont les mêmes qu'à l'échéance J0, avec des RMSE et biais légèrement plus importants, notamment à fortes concentrations.

Ces résultats montrent qu'il faut être prudent sur l'interprétation des scores relatifs au biais moyen, qui peut comporter des effets de compensation. Ainsi, la chaîne ESM\_JP7, qui ne respecte pas l'objectif de qualité sur le biais moyen, présente pourtant les meilleurs scores pour les fortes concentrations – ces dernières présentant le plus d'enjeu en termes de prévisions.

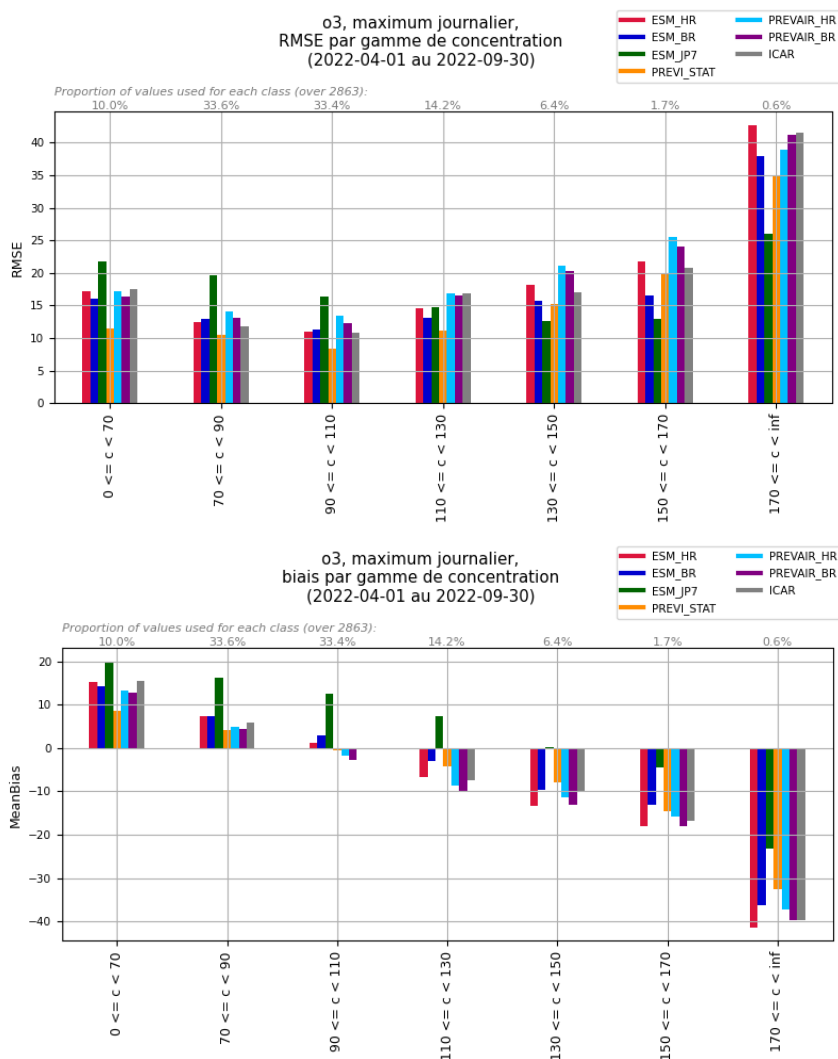


Figure 9 : Ozone – Scores de RMSE et biais par classes de concentration des prévisions d'avril à septembre 2022 sur l'échéance J1. Au-dessus des figures, le pourcentage de données observées pour chaque classe de concentrations.

### 3.1.1.5. Sur plusieurs années

Dans le cadre du Référentiel Technique National, il est demandé d'afficher les performances de prévisions des modèles de qualité de l'air sur les cinq dernières années. Pour cette édition, le graphique associé est réalisé partiellement – ne disposant pas des sorties de tous les modèles sur la période 2018-2022. À titre indicatif, la chaîne PREVI\_STAT a été mise en opérationnel à compter d'octobre 2020.

Seuls les scores de RMSE, biais moyen et corrélation pour les mêmes modèles pour les années 2021 et 2022 sont présentés. L'incrémentation des années sera réalisée au fil des bilans, l'objectif étant d'arriver à obtenir cinq années glissantes.

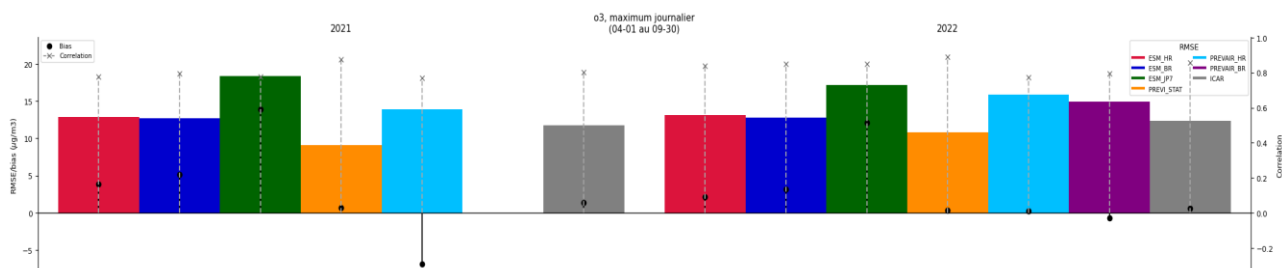


Figure 10 : Ozone – Synthèse des scores de RMSE (barres colorées), biais moyen (points noirs) et corrélation (pointillés gris) à l'échéance J0 entre avril et septembre depuis 2021

À l'échéance J0 (Figure 10), les RMSE sont globalement stables entre 2021 et 2022. Une légère augmentation est observée pour le modèle PREVI\_STAT et Prev' Air HR (+2 µg/m³). Le biais est en diminution pour l'ensemble des modèles (+0.3 à -2 µg/m³ pour la plupart des modèles, et jusqu'à -7 µg/m³ pour Prev' Air HR). La corrélation est en augmentation pour l'ensemble des modèles d'Airparif, ICAR inclus, et stable pour Prev' Air\_HR.

Les mêmes tendances sont observées à J1. La corrélation est en augmentation pour l'ensemble des modèles, y compris pour Prev' Air\_HR.

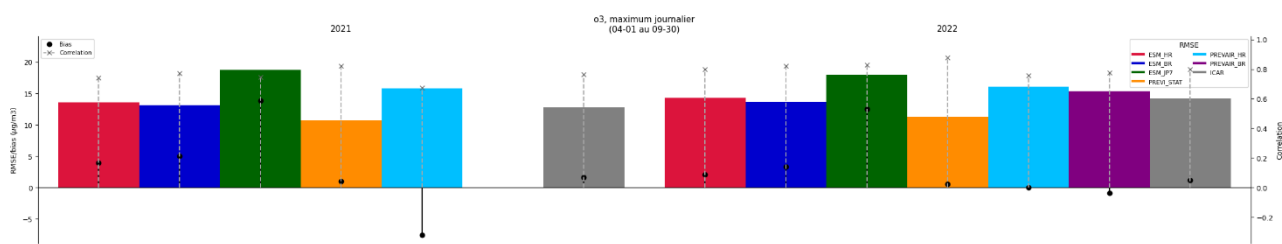


Figure 11 : Ozone – Synthèse des scores de RMSE (barres colorées), biais moyen (points noirs) et corrélation (pointillés gris) à l'échéance J1 entre avril et septembre depuis 2021

Les écarts observés entre les modèles PREVI\_STAT et ICAR sont similaires entre 2021 et 2022.

**La différence des scores entre ICAR et la chaîne PREVI\_STAT, qui constitue la prévision de référence pour les prévisionnistes peut s'expliquer en partie par la méthodologie de correction des cartes.** En effet, en 2022, la carte de prévision ICAR était, par défaut, calculée à partir de la carte ESM\_HR, afin de maintenir une cohérence avec la carte J-1, qui est calculée avec ce modèle. La méthode de correction actuelle ramène la valeur maximale modélisée au milieu de fourchette du prévisionniste, tout en conservant le motif. En améliorant la valeur maximale modélisée sous le panache, la correction peut également dégrader la qualité de la carte sur le reste de la région. Cette opération, transparente pour la carte diffusée au public, affecte en revanche l'ensemble des scores présentés précédemment.

### 3.1.2. Performances sur les dépassements des seuils réglementaires

Au cours de l'été 2022, 3 vagues de chaleur se sont succédées en juin, juillet et août et ont entraîné 5 dépassements du seuil d'information en ozone (fixé à 180 µg/m<sup>3</sup> en moyenne horaire). Ce nombre de jours d'épisode est relativement faible au regard des températures extrêmes mesurées en Île-de-France. Ce constat s'explique par le fait que lors des journées de fortes chaleurs, les masses d'air provenaient essentiellement d'ouest, ce qui a limité l'import de polluants essentiels à la formation de l'ozone. De plus, lors de certaines journées, les températures étaient certes très élevées, jusqu'à 40,5 degrés Celsius, mais la présence de nuages a limité la formation de fortes concentrations d'ozone.

Date du dépassement	Polluant	Seuil	Prévu ou pas	Réalisation de dépassement selon les critères de l'arrêt interpréfectoral	
15/06/2022	O3	Information	NON	OUI	Détection manquée
16/06/2022	O3	Information	NON	OUI	Détection manquée
17/06/2022	O3	Information	OUI	OUI	Bonne détection
18/06/2022	O3	Information	OUI	NON	Fausse détection
13/07/2022	O3	Information	OUI	NON 179 µg/m <sup>3</sup> selon les critères AMU	Fausse détection mais prévision de dépassement justifiée
18/07/2022	O3	Information	OUI	OUI	Bonne détection
25/08/2022	O3	Information	NON	OUI	Détection manquée

Figure 12 : Ozone – Liste des évènements de pollution constatés et/ou prévus au cours de l'année 2022.

#### 3.1.2.1. Scores aux stations

##### 3.1.2.1.1. Séries temporelles

Les figures suivantes représentent, pour chaque journée de la période considérée, le nombre de dépassements du seuil de 180 µg/m<sup>3</sup> en ozone observés (en pointillés gris) et prévus (en couleur) à l'échéance J0 (Figure 13) et J1 (Figure 14), cumulés sur toutes les stations de mesure de fond.

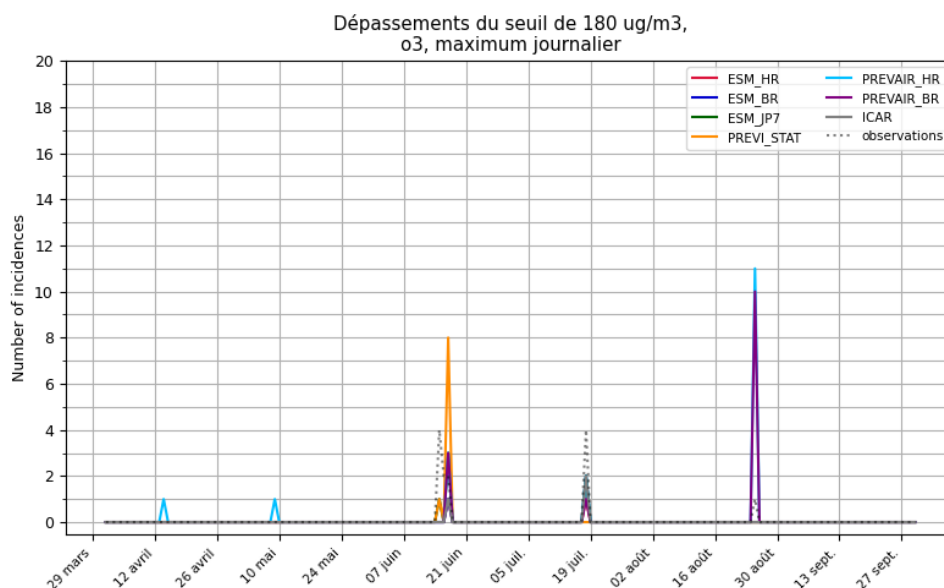


Figure 13 : superposition des dépassements du seuil de 180 µg/m<sup>3</sup> en O<sub>3</sub> enregistrés et prévus à l'échéance J0 par les différentes prévisions

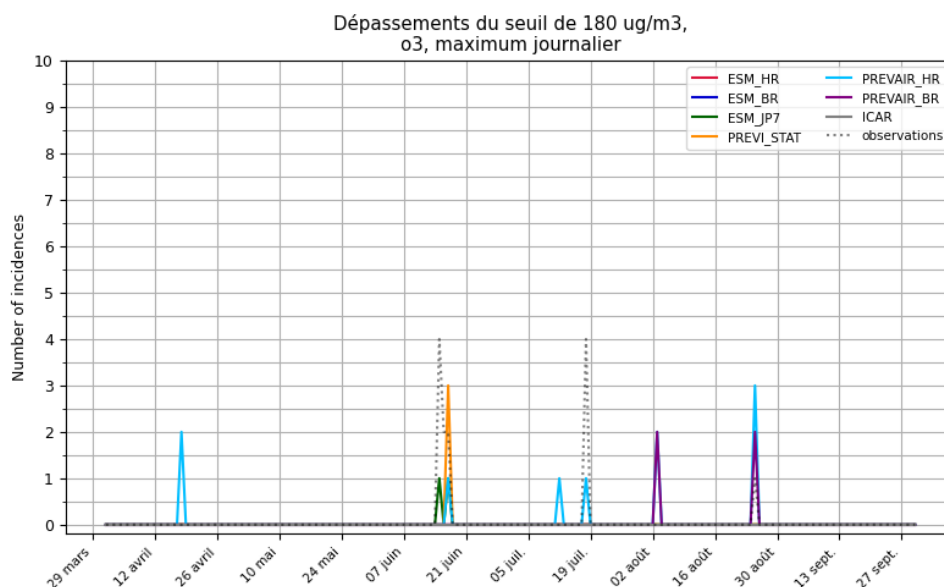


Figure 14 : superposition des dépassements du seuil de 180 µg/m³ en O<sub>3</sub> enregistrés et prévus à l'échéance J1 par les différentes prévisions

Ces périodes montrent que les périodes d'occurrence ont bien été anticipées par les modèles. Quelques fausses alertes sont relevées pour les chaînes PREVAIR en dehors des épisodes observés.

### 3.1.2.1.2. Éléments de la table de contingence

Les éléments pris en compte pour produire les figures de cette partie sont définis en Annexe 2 – Éléments de table de contingence du « Document de référence du dispositif national pour la prévision de la qualité de l'air » du Référentiel Technique National, applicable au 1<sup>er</sup> janvier 2022.

Les figures suivantes représentent les valeurs des tables de contingence (Bonnes détections, Détections manquées, Fausses détections) cumulées sur les stations de mesure de fond de la région Île-de-France prévues aux échéances J0 (Figure 15) et J1 (Figure 16).

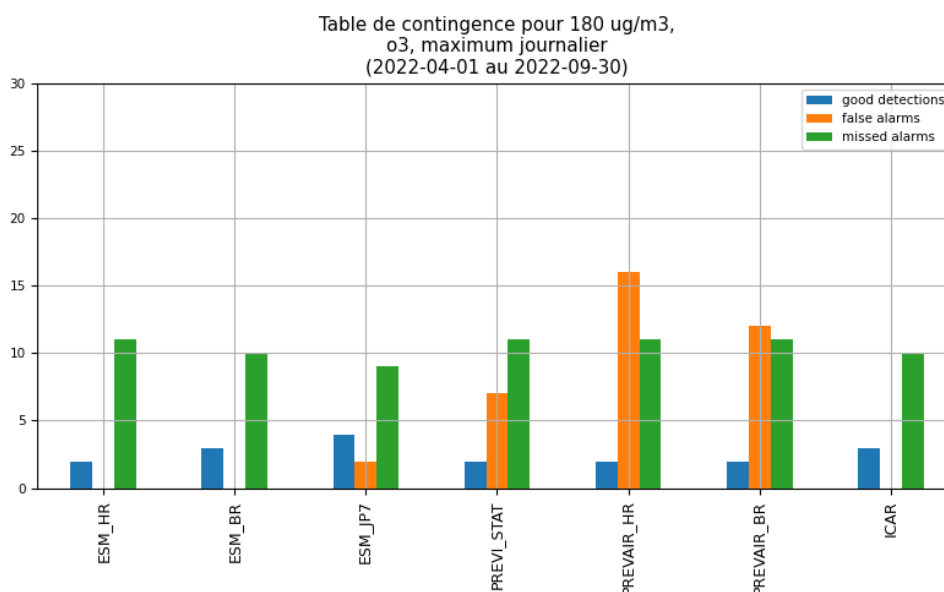


Figure 15 : Ozone – Tables de contingence sur la période d'avril à septembre 2022 à l'échéance J0 pour le seuil de 180 µg/m<sup>3</sup>. Les bonnes détections sont en bleu, les fausses détections en orange et les détections manquées en vert.



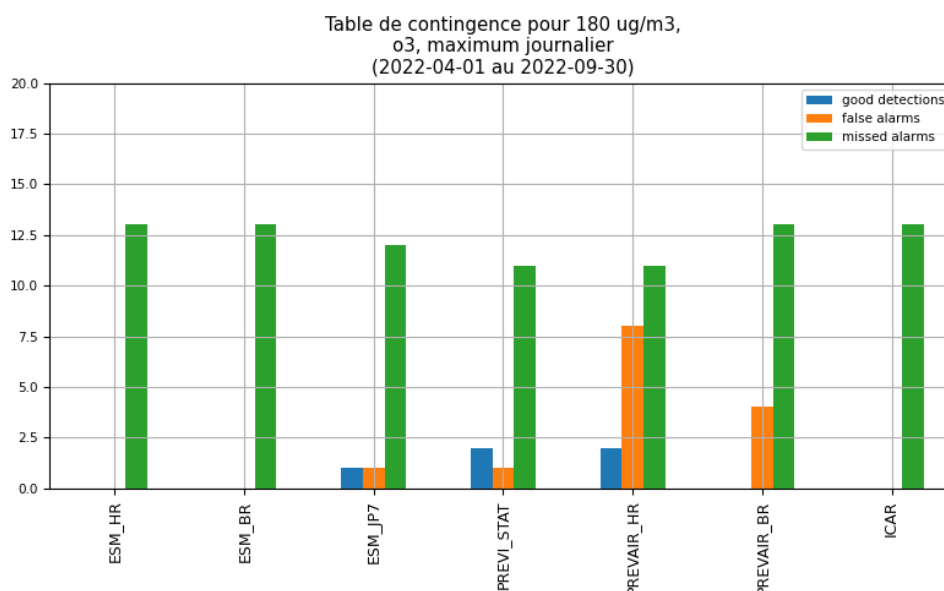


Figure 16 : Ozone – Tables de contingence sur la période d'avril à septembre 2022 à l'échéance J1 pour le seuil de 180 µg/m<sup>3</sup>. Les bonnes détections sont en bleu, les fausses détections en orange et les détections manquées en vert.

À l'échéance J1, le modèle ayant le plus de bonnes détections est la chaîne ESM\_JP7 (4), suivi de ESM\_BR et ICAR (3), les autres modèles étant à 2 bonnes détections. Le nombre de détections manquées est assez homogène sur les différents modèles. Il augmente légèrement à l'échéance J1. Les modèles PREVAIR présente le plus de fausses détections aux 2 échéances. Parmi les modèles développés par Airparif, les chaînes ESM\_JP7 et PREVI\_STAT ont réalisé respectivement 2 et 7 fausses détections à J0, et seulement 1 fausse détection à J1.

Ces éléments ne reflètent pas les scores de prévision détaillés dans la Figure 12, puisqu'ils ne font apparaître aucune bonne détection à J1 pour ICAR, alors que les prévisionnistes ont bien réalisé 2 bonnes détections. De même, les 2 fausses détections réalisées par les prévisionnistes n'apparaissent pas sur cette figure. Les raisons expliquant ces différences ont été identifiées dans l'analyse des scores au regard des critères de l'AMU (Voir sous-section 3.1.2.2).

**Ces éléments, calculés en nombres de dépassement cumulés aux stations de mesure, sont difficiles à interpréter d'un point de vue prévision opérationnelle, puisque cela revient à pondérer les épisodes en fonction du nombre de stations concernées.** Les erreurs de prévision sur les épisodes concernant plusieurs stations vont donc compter plus que des erreurs de prévision sur des épisodes isolés, ce qui n'a pas de sens d'un point de vue du déclenchement d'une procédure d'information régionale.

À partir des éléments de la table de contingence, plusieurs indicateurs (probabilité de détection, taux de détection manquée, taux de succès, biais de prévision et « Critical Success Index ») peuvent être calculés et représentés dans un diagramme de performance, qui permet de mieux évaluer les performances des prévisions des modèles par rapport à la détection des seuils réglementaires.

Ils permettent une vue complète des performances des prévisions par rapport à la détection des seuils réglementaires (l'objectif étant de se situer au plus proche du coin supérieur) :

- Axe horizontal (taux de succès) : capacité du modèle à faire de bonnes prévisions parmi les détections ;
- Axe vertical (probabilité de dépassement) : capacité du modèle à prévoir des dépassements ;
- Axe transverse : proportion de fausses alertes par rapport aux alertes manquées ;
- Axe de couleur (CSI) : capacité du modèle à faire de bonnes prévisions sans événements parasites (fausses alertes, alertes manquées).

Ces figures s'appuient sur la table de contingence (les données sont considérées d'un bloc – pas de moyenne aux stations) pour obtenir la Probabilité de dépassement en fonction du Taux de Succès.

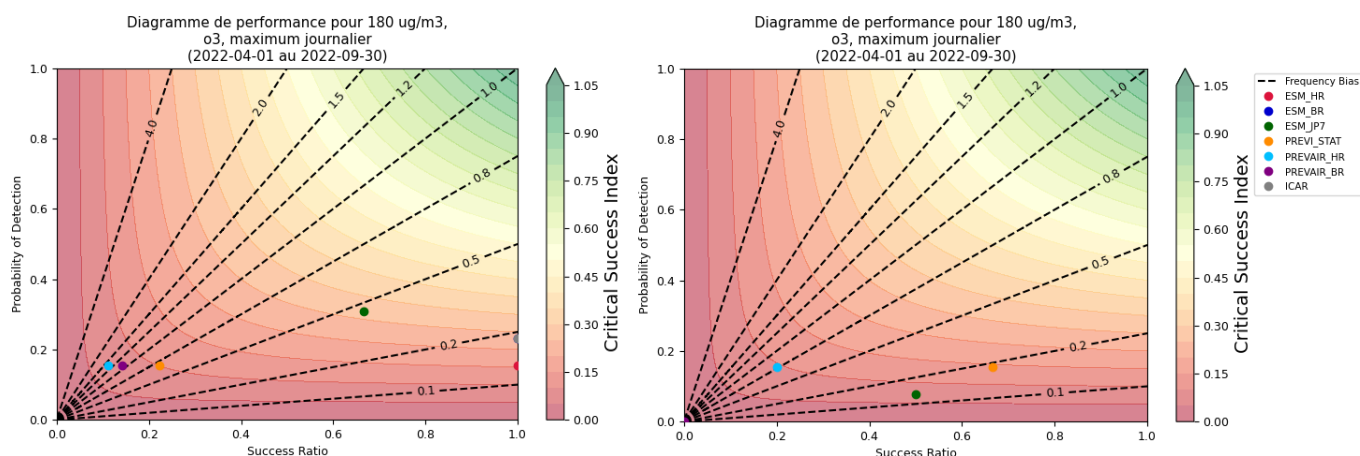


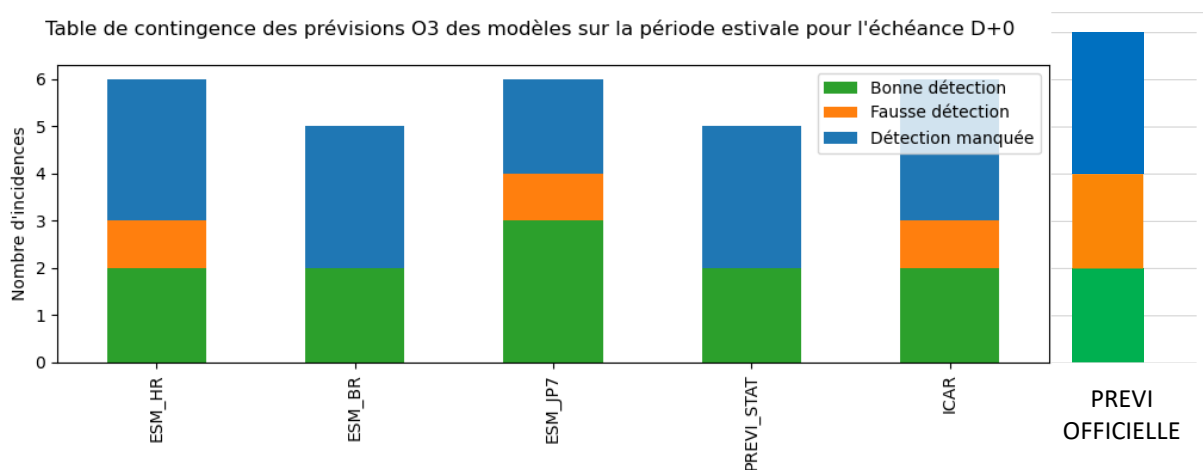
Figure 17 : Ozone - Diagramme de performance d'avril à septembre 2022 pour l'échéance J0 (gauche) et J1 (droite).

Ces diagrammes confirment que la chaîne ESM\_JP7 présente le meilleur index de succès à J0. En revanche, c'est le modèle PREVI\_STAT qui présente le meilleur score à J1, qui est l'échéance la plus critique pour la prévision. Compte-tenu des biais identifiés, ce diagramme ne permet pas d'évaluer les performances de prévision des prévisionnistes.

### 3.1.2.2. Scores en lien avec les critères AMU

Les figures suivantes représentent les éléments de la table de contingence concernant les critères des arrêtés mesure d'urgence en place en région Île-de-France aux échéances J0 et J1 (Figure 18).

Ces éléments des tables de contingence sont établis d'après le dépassement des critères surfaciques et d'exposition de la population prévue par les prévisions J0 et J1 par rapport à la situation constatée. La référence pour la validation des critères est le modèle ICAR calculé à l'échéance J-1 (issue d'une procédure d'assimilation de données pour les journées considérées). Aucune distinction n'est faite sur le critère validé (surfacique ou population exposée).



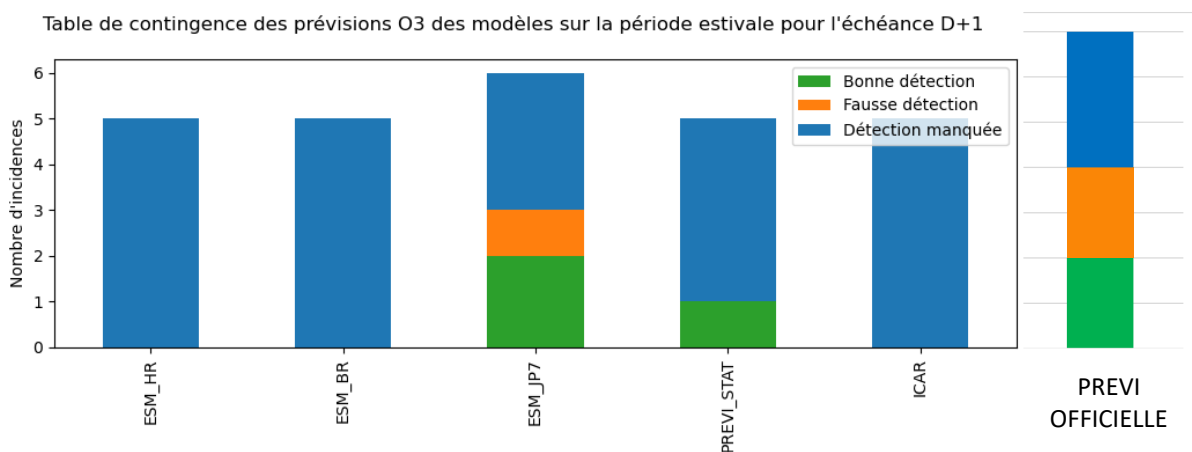


Figure 18 : Ozone – Table de contingence au regard des critères AMU (critères surfaciques et de population confondus) sur la période d'avril à septembre 2022 aux échéances J0 et J1. Les bonnes détections sont en vert, les fausses détections en orange et les détections manquées en bleu.

Les scores de l'échéance J1 présentent le plus d'importance d'un point de vue prévision des épisodes, puisque les prévisions d'épisodes ne sont généralement pas modifiées à J0, sauf en cas de changement brutal de situation météorologique remettant en cause de façon certaine les prévisions de la veille.

**Au regard des critères AMU, les chaînes de modélisation ESM\_HR, ESM\_BR et ICAR ont enregistré 5 détections manquées à l'échéance J1.** Il y a un écart entre les indicateurs ICAR et les prévisions officielles détaillées sur la **Figure 12**, et rappelées à côté du graphique. **En effet, les prévisionnistes n'ont réalisé que 3 détections manquées, soit le nombre le plus faible des modèles présentés, avec la chaîne ESM\_JP7.** L'écart concerne les journées du 17 juin et du 18 juillet 2022, pour lesquelles les prévisionnistes d'Airparif ont bien prévu un dépassement du seuil d'information-recommandations (avec des concentrations maximales prévues en ozone comprises respectivement entre 180 et 210  $\mu\text{g.m}^{-3}$  et entre 170 et 200  $\mu\text{g.m}^{-3}$ ), mais pour lequel les critères de l'arrêté inter-préfectoral ne sont pas remplis.

**La chaîne ESM\_JP7 est le modèle ayant enregistré le plus de bonnes détections à J1 (2 / 5 épisodes). Le nombre de bonnes détections réalisées par les prévisionnistes est équivalent**, mais n'apparaissent pas dans les scores de la chaîne ICAR.

**La chaîne ESM\_JP7 a également réalisé une fausse alerte, et les prévisionnistes 2 fausses alertes**, qui n'apparaissent néanmoins pas dans les scores ICAR. L'une de ces fausses alertes concerne le 13 juillet 2022. Pour cette journée, les prévisionnistes ont prévu le dépassement de seuil de 180  $\mu\text{g.m}^{-3}$  en ozone, mais les critères de l'AMU n'étaient pas remplis sur la carte ICAR. Cette fausse alerte apparaît en revanche à J0, car la carte ICAR a été produite avec ESM\_HR, qui prévoyait à cette échéance un dépassement des critères de l'AMU. L'analyse J-1 a fait état d'une concentration maximale en ozone de 179  $\mu\text{g}/\text{m}^3$  au regard des critères de l'arrêté inter-préfectoral en vigueur. La prévision était néanmoins justifiée d'un point de vue contexte général de pollution.

La seconde fausse alerte des prévisionnistes concerne le 18 juin 2023. Ce jour intervient après 2 dépassements constatés du seuil d'information et un 3<sup>ème</sup> jour de dépassement prévu (qui sera effectivement dépassé). La température annoncée est de 40 degrés Celsius, soit 3 degrés de plus que la veille. Malgré une révision à la baisse des prévisions des modèles, les prévisionnistes ont jugé que l'annonce d'une sortie d'épisode serait inintelligible pour le public. L'épisode ne sera finalement pas constaté, la concentration maximale atteinte étant de 171  $\mu\text{g}/\text{m}^3$ . Un renforcement du vent a poussé le panache en-dehors des limites de la région Île-de-France.

À J1, les modèles présentent plus de détections manquées que de fausses alertes, ce qui confirme la tendance à la sous-estimation des valeurs extrêmes. La chaîne ESM\_JP7 et la prévision officielle présentent les meilleurs taux de bonnes détections. Les prévisionnistes ont réalisé une fausse alerte de plus, mais la prévision de dépassement restait pertinente d'un point de vue information, la valeur maximale ayant atteint 179 µg/m<sup>3</sup> sur la région.

À l'échéance J0, le nombre de bonnes détections augmente pour tous les modèles (sauf pour la prévision officielle, les prévisionnistes ne révisant qu'exceptionnellement leur prévision de dépassement le jour-même). En revanche, les scores d'ICAR se rapprochent de la prévision officielle car les prévisionnistes peuvent, sans changer la prévision de dépassement, rehausser la fourchette de concentration. Dans ce cas, les critères de dépassement peuvent être remplis. On compte une fausse détection pour le modèle ESM\_HR, qui n'apparaissait pas à l'échéance J1.

Le calcul de ces scores fait apparaître des biais dans la chaîne ICAR, qui ne reflète pas les prévisions de dépassement du seuil d'information réalisées par les prévisionnistes. Ces différences sont liées à plusieurs biais méthodologiques :

- Pour produire la carte ICAR, la chaîne compare la valeur maximale modélisée avec la fourchette de concentrations maximales prévues par le prévisionniste. Si le max modélisé est en-dehors de la fourchette prévue, la valeur maximale modélisée est ramenée au milieu de la fourchette du prévisionniste. Lorsque la fourchette de concentration prévue chevauche le seuil d'information (170-200 µg/m<sup>3</sup> par exemple), la correction à 185 µg/m<sup>3</sup> peut, selon la localisation et la largeur du panache, ne pas faire apparaître de dépassement des critères de l'AMU. La fourchette 170-200 µg/m<sup>3</sup> peut également englober un maximum prévu par le modèle Esmeralda légèrement inférieur au seuil. Dans ce cas, aucune correction n'est appliquée à la carte, qui ne remplit donc pas les critères. C'est le cas du 13 juillet 2023, où le maximum prévu par la chaîne ESM\_HR de 178 µg/m<sup>3</sup>, n'a pas été corrigé par la fourchette du prévisionniste.
- La valeur maximale modélisée est actuellement considérée sur l'ensemble du domaine, et non pas sur le territoire francilien. Si cette valeur est située en-dehors de la région, la carte peut ne pas être corrigée selon la fourchette.

C'est notamment le cas du 17 juin 2022. Le modèle ESM\_HR prévoit sur l'Île-de-France une valeur maximale modélisée de 173 µg/m<sup>3</sup>. Les prévisionnistes prévoient une concentration maximale comprise entre 180 et 210 µg/m<sup>3</sup>, conformément au modèle PREVI\_STAT. Le maximum de la carte ESM\_HR aurait dû être ramené à 195 µg/m<sup>3</sup>, et donc faire apparaître un dépassement selon les critères de l'AMU. Mais le modèle a en réalité bien prévu une valeur maximale de 186 µg/m<sup>3</sup>, mais en dehors des frontières de l'Île-de-France. Cette valeur étant comprise dans la fourchette prévue par le prévisionniste, la carte ESM\_HR n'a pas été corrigée ce jour-là. La carte ICAR ne fait donc pas apparaître de dépassement sur la région.



## 3.2. Prévisions particules PM<sub>10</sub>

Les scores classiques (Biais, RMSE et Corrélation) sont établis pour les particules PM<sub>10</sub>, en moyenne journalière, sur l'année entière.

### 3.2.1. Scores classiques (Biais, RMSE et corrélation)

#### 3.2.1.1. Par rapport aux objectifs de qualité

Les scores sont établis sur toutes les stations de fond pour la région Île-de-France avec les objectifs de qualité suivants à respecter : RMSE : 15 µg/m<sup>3</sup>, Biais : ± 7.5 µg/m<sup>3</sup>, R : 0.6. Les scores suivants sont une moyenne des scores temporels de chaque station de fond (11 sites de mesure au total).

Les Figure 19 et Figure 20 représentent respectivement les scores pour les échéances J0 et J1.

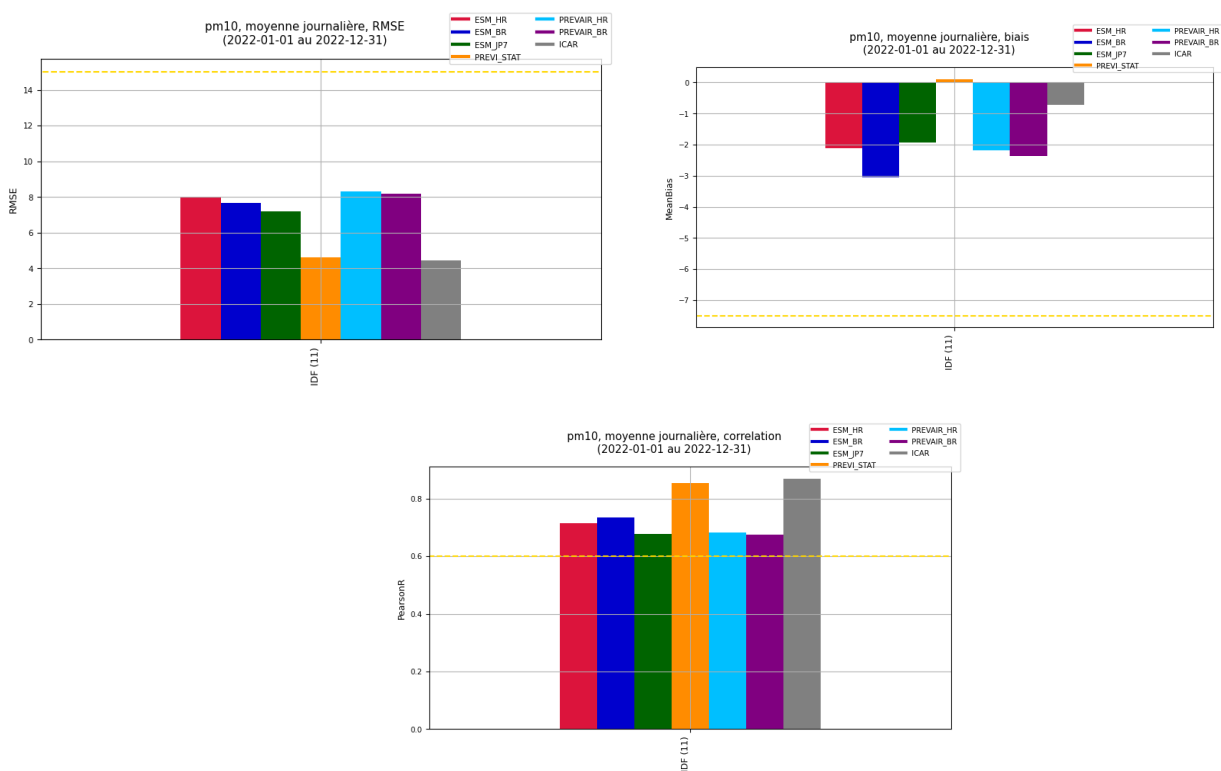


Figure 19 : Particules PM<sub>10</sub> – Scores de RMSE, biais et corrélation en Île-de-France des prévisions) sur l'année 2022 à l'échéance J0. La ligne pointillée orange indique les objectifs de qualité à respecter.

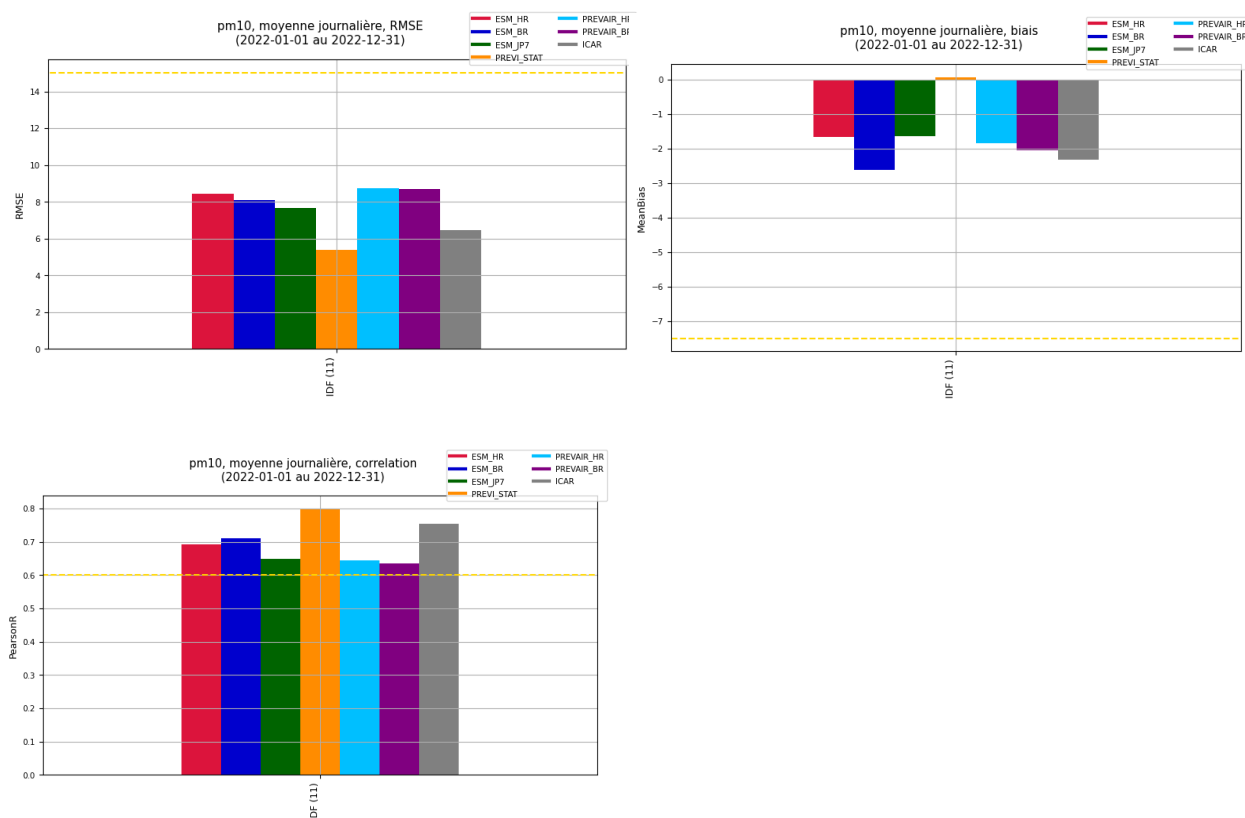


Figure 20 : Particules PM<sub>10</sub> – Scores de RMSE, biais et corrélation en Île-de-France des prévisions sur l'année 2022 à l'échéance J1. La ligne pointillée orange indique les objectifs de qualité à respecter.

Pour les échéances J0 et J1, **l'ensemble des modèles disponibles respectent largement l'objectif de qualité fixé par le RTN à 15 µg/m<sup>3</sup>** pour le paramètre RMSE. En effet, les RMSE des systèmes de modélisation sont tous globalement compris entre 4 et 8 µg/m<sup>3</sup>. **Les plus faibles RMSE concernent les modèles PREVI\_STAT et ICAR** (de l'ordre de 4-5 µg/m<sup>3</sup>), alors que **les plus fortes RMSE sont relevées pour les modèles ESM\_HR et PREVAIR\_HR/BR** (de l'ordre de 8 µg/m<sup>3</sup>). **À J0, les RMSE sont très légèrement inférieures à celles calculées à J1, plus particulièrement pour les chaînes PREVI\_STAT (-1 µg/m<sup>3</sup>) et ICAR (-2 µg/m<sup>3</sup>).**

À J1 et J0, les modèles de prévision d'Airparif ESM\_HR, ESM\_BR, ESM\_JP7 et ICAR présentent des biais moyens compris entre -1.5 et -3 µg/m<sup>3</sup>, à l'exception de la PREVI\_STAT qui présente un biais moyen quasiment nul. **L'ensemble des chaînes respectent largement l'objectif de qualité** défini dans le RTN ( $\pm 7.5$  µg/m<sup>3</sup>). Les biais moyens des différents modèles analysés sont globalement comparables pour les deux échéances de prévision, hormis pour la chaîne ICAR, qui présente un biais moyen plus faible à J0 (moins de 1 µg/m<sup>3</sup>) qu'à J1 (environ 2 µg/m<sup>3</sup>). Cela s'explique par le fait que le prévisionniste peut réajuster la fourchette de concentration maximale prévue à J0 sur la base des niveaux observés le matin.

En ce qui concerne la corrélation, pour les échéances J0 et J1, **l'ensemble des modèles respectent l'objectif de qualité** fixé par le RTN ( $r=0.6$ ). Les corrélations sont similaires entre J0 et J1, exception faite pour la chaîne ICAR. En effet, une réelle amélioration de ce paramètre est observée entre J1 et J0 ( $r$  passant de 0.75 à 0.87), qui s'explique par la réadaptation des fourchettes de concentrations en particules PM<sub>10</sub> le matin par le prévisionniste, sur la base des observations de la veille et du matin.

L'analyse de ces trois indicateurs montre **les très bons scores de la chaîne PREVI\_STAT**, ainsi que **la plus-value des prévisionnistes**. En effet, leurs interventions permettent d'améliorer significativement les scores de prévision par rapport aux chaînes brutes (ESM\_HR, ESM\_BR, ESM\_JP7).

### 3.2.1.2. Par typologie

Les scores (RMSE, Biais et Corrélation) ont été calculés par station, puis moyennés en appliquant un filtre sur les typologies de stations de fond (urbaines, péri-urbaines et rurales).

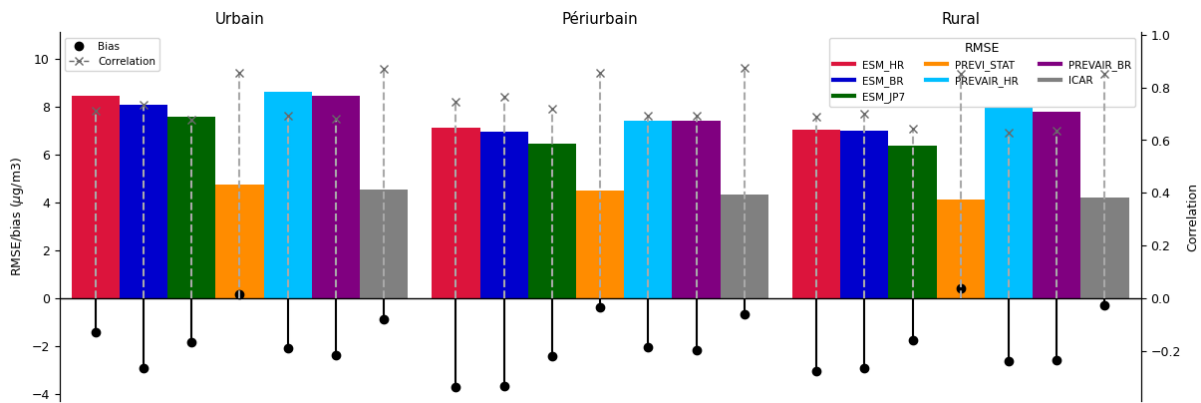


Figure 21 : Particules PM<sub>10</sub> – Synthèse des scores de RMSE, biais et corrélation en fonction de la typologie de station (urbaines, péri-urbaines, rurales) des prévisions sur l'année 2022 à l'échéance J0.

La Figure 21 ci-dessus, qui présente les scores pour l'échéance J0, montre que :

- **les RMSE calculés sur les sites de fond péri-urbain et rural sont globalement comparables** (les plus fortes RMSE étant observées pour les chaînes ESM\_HR, ESM\_BR, ESM\_JP7 et PREVAIR\_HR et les plus faibles RMSE pour les modèles PREVI\_STAT et ICAR). Elles restent néanmoins **légèrement inférieures à celles calculées sur les stations urbaines**, notamment pour les chaînes brutes ;
- **À l'exception des chaînes PREVI\_STAT et ICAR qui présentent des biais très proches sur les 3 typologies de sites, les autres modèles présentent des biais légèrement supérieurs sur les stations péri-urbaines et rurales.**
- **les corrélations calculées pour l'ensemble des modèles sont similaires pour les stations de fond urbain et péri-urbain.** Elles sont légèrement plus faibles en situation de fond rural, notamment pour les chaînes ESM\_HR, ESM\_BR, ESM\_JP7 et PREVAIR\_HR. Indépendamment de la typologie des stations, **les chaînes PREVI\_STAT et ICAR présentent les meilleures corrélations** ( $r$  supérieures à 0.8).

Pour l'échéance J1 (Figure 22), les scores sur les trois typologies de stations sont très proches de ceux calculés à J0 pour tous les modèles, excepté la chaîne ICAR. En effet, ce modèle présente **des RMSE légèrement plus élevées** (autant en situation de fond (péri)urbain que rural), alors que ces dernières étaient comparables à celles de la chaîne PREVI\_STAT à J0.

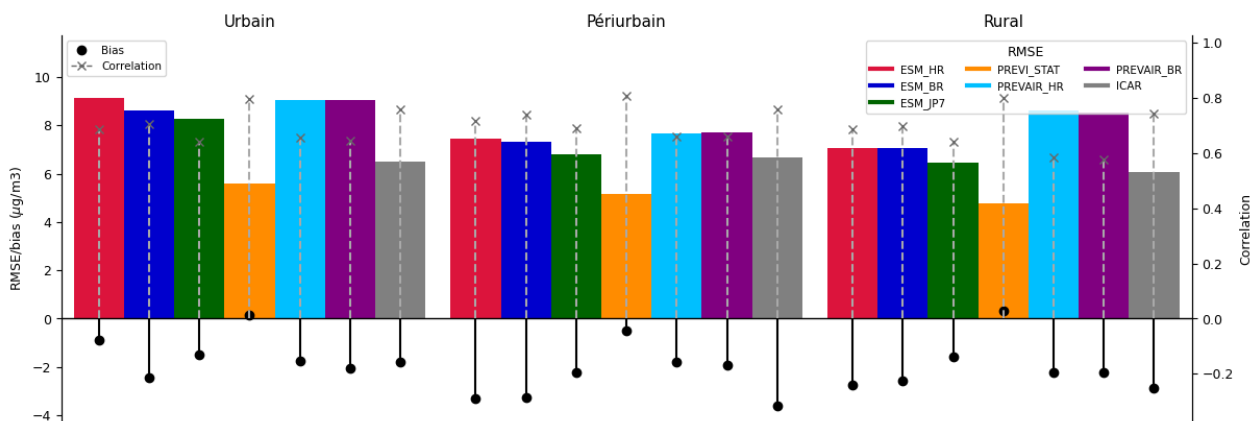


Figure 22 : Particules PM<sub>10</sub> – Synthèse des scores de RMSE, biais et corrélation en fonction de la typologie de station (urbaines, péri-urbaines, rurales) des prévisions sur l'année 2022 à l'échéance J1.

### 3.2.1.3. Par échéance de prévision

Les scores (RMSE, Biais et Corrélation) ont été calculés par station, puis moyennés sur les différents horizons de prévisions (J0 à J2).

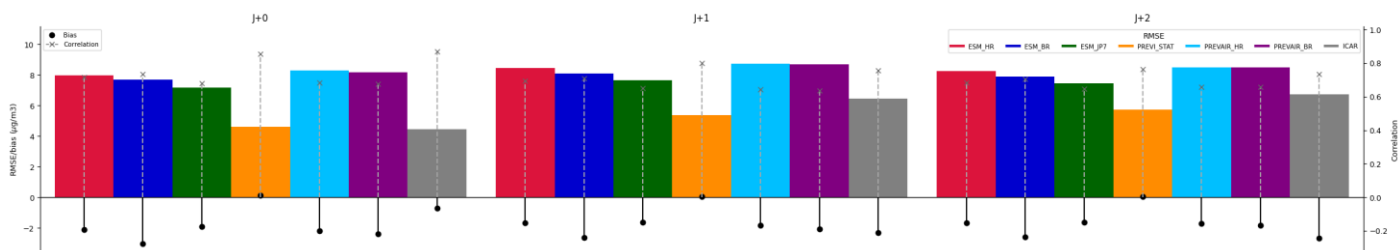


Figure 23 : Particules PM<sub>10</sub> – Synthèse des scores de RMSE, biais et corrélation en fonction du jour de prévision sur l'année 2022

La Figure 23 ci-dessus illustre que :

- les RMSE sont globalement équivalentes aux échéances J1 et J2 par rapport à J0 pour l'ensemble des modèles, à l'exception d'ICAR, dont la RMSE augmente à J1 et J2 (+2 µg/m<sup>3</sup>).
- ce constat s'applique également aux biais moyens ;
- **les corrélations des systèmes de modélisation ESM\_HR, ESM\_BR, ESM\_JP7 et PREVAIR\_HR sont globalement équivalentes pour les échéances de J0 à J2.** Les chaînes PREVI\_STAT et ICAR présentent de meilleures corrélations ( $0.7 < r < 0.8$ ), avec une amélioration de ces indicateurs à J0 par rapport à J1 et J2.

**Les scores de l'ensemble des chaînes sont globalement stables de J0 à J2, avec une légère amélioration à J0 pour les chaînes PREVI\_STAT et ICAR.**

### 3.2.1.4. Par gamme de concentration

Les données journalières sont groupées selon leur « classe de concentration » (par rapport aux observations). Pour chaque groupe, les scores sont calculés sur chaque station de fond, puis moyennés.

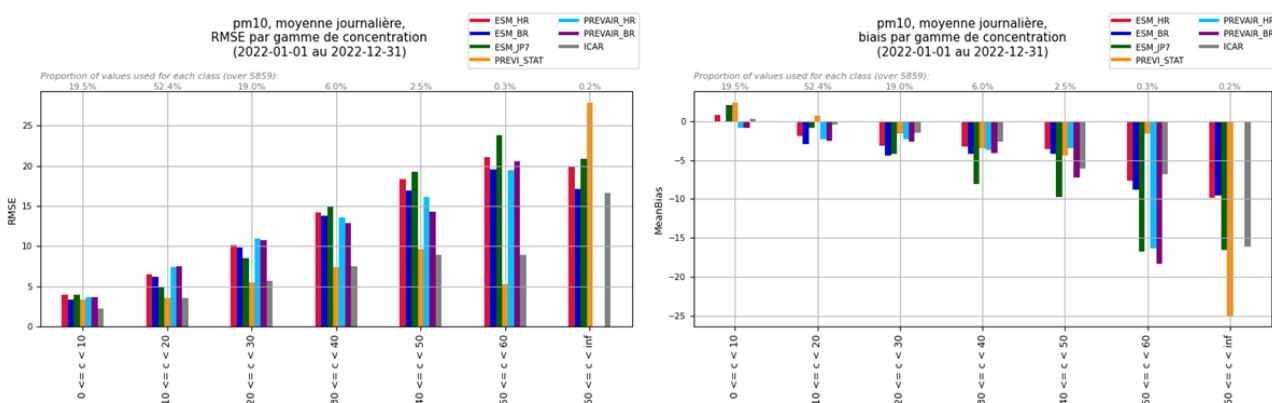


Figure 24 : Particules PM<sub>10</sub> – Scores de RMSE et biais par classes de concentration des prévisions sur l'année 2022 à l'échéance J0. Au-dessus de la figure, les taux d'observations pour chaque classe de concentrations.



L'ensemble des modèles présentent des RMSE qui augmentent jusqu'à 20  $\mu\text{g}/\text{m}^3$  avec la concentration en particules, à l'exception des modèles PREVI\_STAT et ICAR, qui restent inférieures à 10  $\mu\text{g}/\text{m}^3$ . **Les RMSE les plus faibles concernent les modèles PREVI\_STAT et ICAR**, sauf pour la PREVI\_STAT aux concentrations supérieures à 60  $\mu\text{g}/\text{m}^3$ , mais ces concentrations représentent moins de 0,2 % des données. Ce paramètre n'est donc pas représentatif.

Concernant l'analyse des biais moyens, il est important de noter que **tous les modèles de qualité de l'air ESM surestiment très légèrement les faibles concentrations** ( $C < 10 \mu\text{g}/\text{m}^3$ ), entre -1 et +2  $\mu\text{g}/\text{m}^3$ . Pour une telle gamme de concentrations, **la chaîne ICAR présente des biais moyens quasi-nuls, comme PREVAIR HR et BR**. Pour des niveaux en  $\text{PM}_{10}$  compris entre 10 et 20  $\mu\text{g}/\text{m}^3$  (soit plus de 50 % des observations), les biais moyens sont compris entre -3 et moins de 1  $\mu\text{g}/\text{m}^3$  pour l'ensemble des chaînes.

**Pour les gammes de concentrations supérieures à 20  $\mu\text{g}/\text{m}^3$ , tous les modèles présentent une tendance à la sous-estimation des niveaux moyens en  $\text{PM}_{10}$ , avec des biais moyens négatifs qui augmentent plus ou moins sensiblement selon les chaînes, notamment la chaîne ESM\_JP7 (-10  $\mu\text{g}/\text{m}^3$ ) pour des concentrations de 40 à 50  $\mu\text{g}/\text{m}^3$ .** Les biais restent inférieurs à -5  $\mu\text{g}/\text{m}^3$  pour les autres chaînes. Indépendamment des modèles, les plus forts biais moyens sont observés lorsque les concentrations en  $\text{PM}_{10}$  sont supérieures à 50  $\mu\text{g}/\text{m}^3$ , mais le peu d'occurrences (soit 1 % des observations) ne permet pas de prendre en compte de manière pertinente cet indicateur dans l'analyse. Il faut également être prudents sur l'interprétation de ces scores moyens. En effet, ils montrent une sous-estimation des valeurs fortes de  $\text{PM}_{10}$ , mais dans le même temps, la Figure 36 montre que les modèles font également un nombre élevé de fausses alertes, ce qui indique donc parfois une surestimation des résultats. Les scores moyens peuvent ainsi conduire à de mauvaises interprétations en termes de décision pour les prévisionnistes.

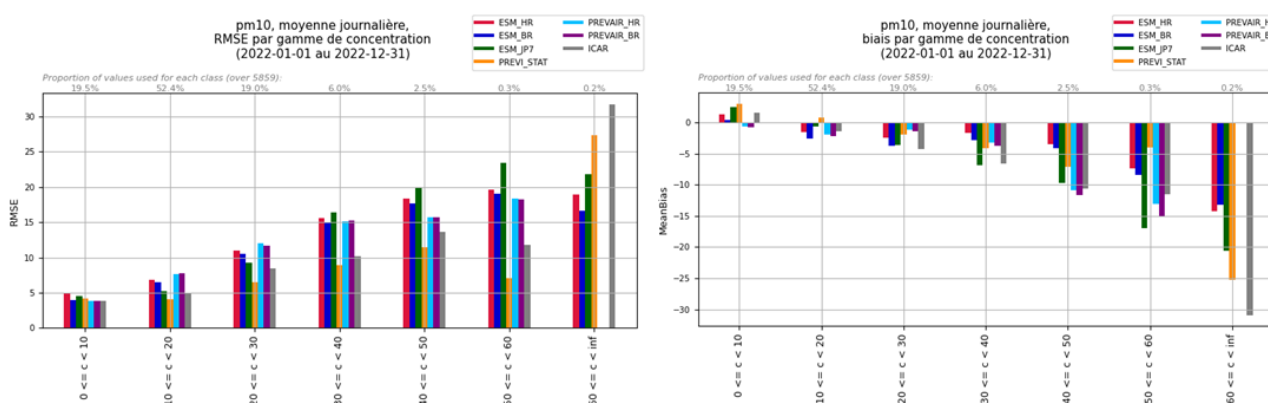


Figure 25 : Particules  $\text{PM}_{10}$  – Scores de RMSE et biais par classes de concentration des prévisions sur l'année 2022 à l'échéance J1. Sous les figures, les taux d'observations pour chaque classe de concentrations.

Les scores à l'échéance J1 sont proches de ceux de l'échéance J0. La chaîne ICAR présente néanmoins des biais sensiblement plus importants pour toutes les gammes à partir de 20  $\mu\text{g}/\text{m}^3$ .

### 3.2.1.5. Sur plusieurs années

L'objectif étant d'arriver à obtenir cinq années glissantes. Pour cette édition, le graphique associé est réalisé partiellement – ne disposant pas des sorties de tous les modèles sur la période 2018-2022. À titre indicatif, la chaîne PREVI\_STAT a été mise en opérationnel à compter d'octobre 2020.

Seuls les scores de RMSE, biais moyens et corrélation pour les mêmes modèles pour les années 2021 et 2022 sont présentés. L'incrémentation des années sera réalisée au fil des éditions. Les Figure 26 et Figure 27 présente la moyenne des scores temporels par station en 2021 et 2022 aux échéances J0 et J1.

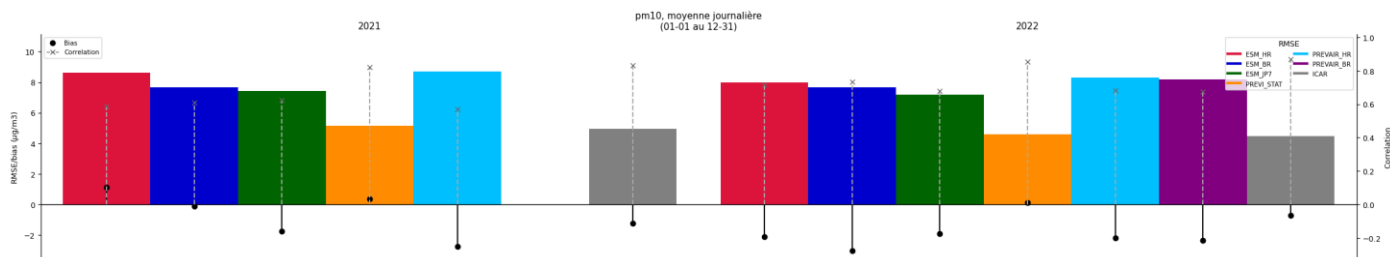


Figure 26 : Particules PM<sub>10</sub> – Synthèse des scores de RMSE (barres colorées), biais moyen (points noirs) et corrélation (pointillés gris) à l'échéance J0 sur l'année entière depuis 2021.

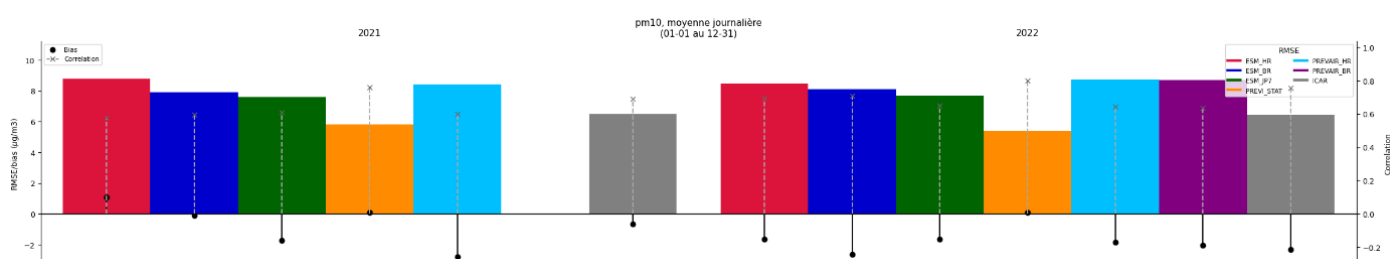


Figure 27 : Particules PM<sub>10</sub> – Synthèse des scores de RMSE (barres colorées), biais moyen (points noirs) et corrélation (pointillés gris) à l'échéance J1 sur l'année entière depuis 2021.

Entre 2021 et 2022, les RMSE sont sensiblement équivalentes pour les 2 échéances et l'ensemble des modèles. Les biais sont stables pour les chaînes ESM\_JP7, PREVI\_STAT et PREVAIR\_HR aux 2 échéances. En revanche, les chaînes ESM\_BR et ESM\_HR présentent des biais plus importants (-1 à -3 µg/m<sup>3</sup>) en 2022. Cette dégradation des scores peut s'expliquer par le fait que le module d'érosion n'a pas été branché entre avril et septembre, comme c'est le cas habituellement, ce qui conduit à une sous-estimation des dusts (poussières liées à la remise en suspension), un effet accentué par la sécheresse qu'a connu la France en 2022.

La chaîne ICAR montre une légère amélioration du biais à l'échéance J1, mais une dégradation à J0. Cela peut s'expliquer par le fait que le modèle utilisé pour calculer la carte a été en grande majorité la carte ESM\_HR. Comme pour l'ozone, **ces scores montrent qu'il est plus pertinent de choisir le modèle le plus proche de la prévision du prévisionniste plutôt que d'avoir recours à la correction par la fourchette de concentrations maximales prévues.**

**Les scores relatifs aux corrélations sont en amélioration sur les deux échéances pour l'ensemble des modèles.**

### 3.2.2. Performances sur les dépassements de seuils réglementaires

Bien que l'hiver ait été globalement doux, des **conditions météorologiques ponctuellement froides et peu dispersives**, survenues notamment au cours du mois de janvier, **ont entraîné 2 dépassements du seuil d'information-recommandations en PM<sub>10</sub>** (fixé à 50 µg/m<sup>3</sup> en moyenne journalière). 3 dépassements de ce seuil réglementaire ont été constatés au mois de mars 2022, du fait des conditions printanières favorables à la formation d'aérosols inorganiques secondaires, auxquels se sont ajoutées des conditions météorologiques très peu dispersives.

Date du dépassement	Polluant	Seuil	Prévu ou pas	Réalisation de dépassement selon les critères de l'arrêt inter préfectoral	
14/01/2022	PM10	Information	NON	OUI	Détection manquée
15/01/2022	PM10	Information	OUI	OUI	Bonne détection
25/01/2022	PM10	Information	OUI	NON 52 µg/m <sup>3</sup> en stations 48 µg/m <sup>3</sup> selon les critères AMU	Fausse détection mais prévision de dépassement justifiée
24/03/2022	PM10	Information	NON	OUI	Détection manquée
25/03/2022	PM10	Information	OUI	OUI	Bonne détection
26/03/2022	PM10	Information	OUI	OUI	Bonne détection
27/03/2022	PM10	Information	OUI	NON	Fausse détection
28/03/2022	PM10	Information	OUI	NON 52 µg/m <sup>3</sup> en stations 50 µg/m <sup>3</sup> selon les critères AMU	Fausse détection mais prévision de dépassement justifiée

Figure 28 : Particules PM<sub>10</sub> – Liste des évènements de pollution constatés et/ou prévus au cours de l'année 2022.

### 3.2.2.1. Scores aux stations

#### 3.2.2.1.1. Séries temporelles

Les figures suivantes représentent, pour chaque journée de la période considérée, le nombre de dépassements du seuil de 50 µg/m<sup>3</sup> en particules PM<sub>10</sub> observés (en pointillés gris) et prévus (en couleur) à l'échéance J0 (Figure 29) et J1 (Figure 30), cumulés sur toutes les stations de mesure de fond.

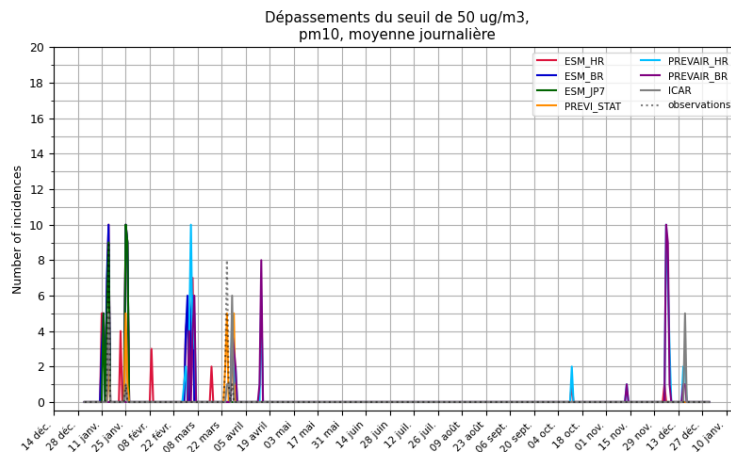


Figure 29 : superposition des dépassements du seuil de 50 µg/m<sup>3</sup> en particules PM<sub>10</sub> enregistrés et prévus en 2022 à l'échéance J0 par les différentes prévisions.

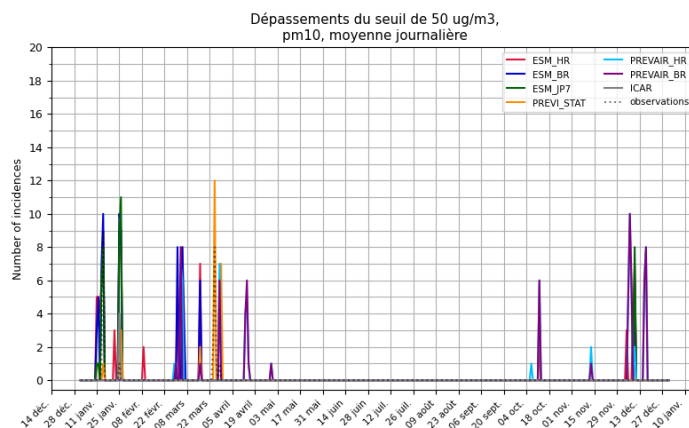


Figure 30 : superposition des dépassements du seuil de 50 µg/m<sup>3</sup> en particules PM<sub>10</sub> enregistrés et prévus en 2022 à l'échéance J1 par les différentes prévisions

Ce cumul sur l'ensemble des stations de mesure permet d'évaluer si les occurrences des épisodes ont bien été prévues par les modèles. Ces figures montrent que les périodes d'épisode ont globalement bien été prévues en période hivernale et printanière. Elles mettent en évidence des fausses détections, en particulier par les modèles PREVAIR, en période printanière, qui a notamment été touchée par des épisodes de type saharien, difficiles à prévoir, et en fin d'année.

### 3.2.2.1.2. Éléments de la table de contingence

Les éléments pris en compte pour produire les figures de cette partie sont définis en Annexe 2 – Éléments de table de contingence, du RTN Évaluation des performances de la qualité de l'air.

Les figures suivantes représentent les valeurs des tables de contingence (Bonne détections, Détections manquées, Fausses détections) cumulées sur les stations de fond de la région Île-de-France prévues aux échéances J0 et J1 (Figure 31) et (Figure 32).

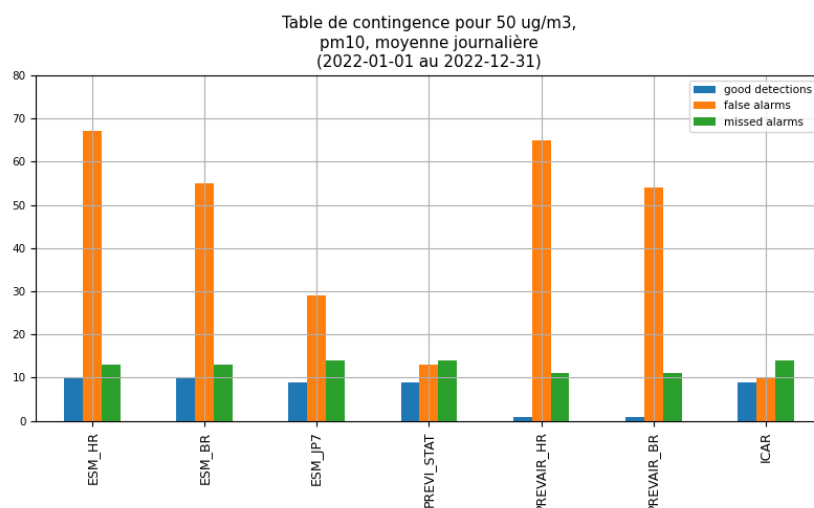


Figure 31 : Particules PM<sub>10</sub> – Tables de contingence sur la période sur l'année 2022 à l'échéance J0 pour le seuil de 50 µg/m<sup>3</sup>. Les bonnes détections sont en bleu, les fausses détections en orange et les détections manquées en vert.

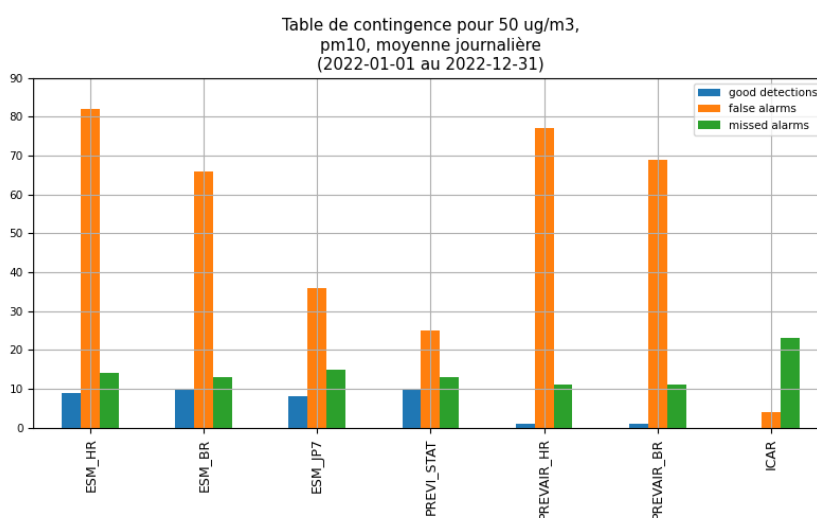


Figure 32 : Particules PM<sub>10</sub> – Tables de contingence sur la période sur l'année 2022 à l'échéance J1 pour le seuil de 50 µg/m<sup>3</sup>. Les bonnes détections sont en bleu, les fausses détections en orange et les détections manquées en vert.

Ces éléments, calculés en nombres de dépassement cumulés aux stations de mesure, sont difficiles à interpréter d'un point de vue prévision opérationnelle, puisque cela revient à pondérer les épisodes en fonction du nombre de stations concernées. Les erreurs de prévision sur les épisodes concernant plusieurs stations vont donc compter plus que des erreurs de prévision sur des épisodes isolés, ce qui n'a pas de sens d'un point de vue du déclenchement d'une procédure d'information régionale, les déclenchements étant validés à partir des cartes sur l'ensemble du territoire.

Pour le modèle ICAR, **les scores présentés sur la Figure 32 ne reflètent pas les prévisions de dépassement réalisées par les prévisionnistes détaillées dans la Figure 28**. En effet, aucune bonne détection n'apparaît à J1 pour la chaîne ICAR alors que les prévisionnistes ont bien prévu 3 dépassements du seuil d'information-recommandations. Cela s'explique par le fait que le 15 janvier, le 25 mars et le 26 mars, les épisodes ont bien été prévus par les prévisionnistes, mais avec des concentrations maximales prévues comprises entre 45 et 55  $\mu\text{g}/\text{m}^3$ , afin de faire apparaître les incertitudes liées à cette prévision de dépassement. Dans tous ces cas, la concentration maximale prévue par le modèle n'était pas comprise dans cette fourchette. Elle a donc été réhaussée au milieu de la fourchette, soit à 50  $\mu\text{g}/\text{m}^3$ , ce qui ne permet pas de faire apparaître un dépassement sur la carte ICAR.

#### **De cette analyse, deux pistes d'amélioration sont identifiées :**

- Dissocier la fourchette de concentrations prévues à des fins de communication de la valeur utilisée pour corriger la carte ;
- Privilégier l'utilisation d'un modèle cohérent avec l'expertise du prévisionniste pour avoir recours le moins possible à la correction via la fourchette.

**Une légère amélioration des scores (augmentation des bonnes détections et diminution des fausses détections et des détections manquées) est observée à J0 par rapport à J1 pour la plupart des modèles, en particulier pour ICAR.** Cela s'explique par le rehaussement des fourchettes réalisées le jour même par les prévisionnistes, qui permet de faire apparaître un dépassement sur la carte ICAR, contrairement à J0..

Les figures suivantes représentent, pour les échéances J0 (Figure 33) et J1 (Figure 34), une vue complète des performances des prévisions par rapport à la détection des seuils réglementaires (l'objectif étant de se situer au plus proche du coin supérieur) :

- Axe horizontal (taux de succès) : capacité du modèle à faire de bonnes prévisions parmi les détections ;
- Axe vertical (probabilité de dépassement) : capacité du modèle à prévoir des dépassements ;
- Axe transverse : proportion de fausses alertes par rapport aux alertes manquées ;
- Axe de couleur (CSI) : capacité du modèle à faire de bonnes prévisions sans événements parasites (fausses alertes, alertes manquées).

Ces figures s'appuient sur la table de contingence (les données sont considérées d'un bloc – pas de moyenne aux stations) pour obtenir la Probabilité de dépassement en fonction du Taux de Succès.

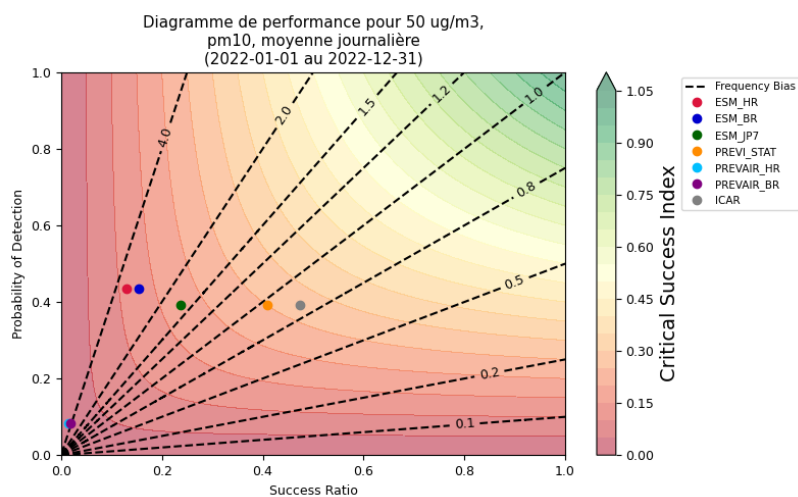


Figure 33 : Particules PM<sub>10</sub> – Diagramme de performance pour les prévisions sur l'année 2022 pour l'échéance J0.

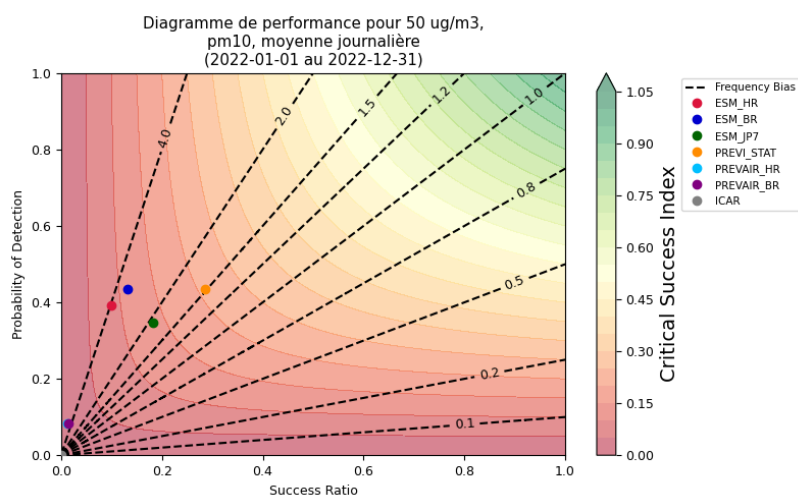


Figure 34 : Particules PM<sub>10</sub> - Diagramme de performance pour les prévisions sur l'année 2022 pour l'échéance J1.

Les modèles ICAR et PREVI\_STAT présentent le meilleur taux de succès critique par rapport aux modèles bruts à l'échéance J0. Néanmoins, le fait de considérer le nombre de dépassements cumulés aux stations limite l'intérêt de cette figure, qui pourrait permettre de mieux évaluer la plus-value des prévisionnistes. En effet, là encore, les scores calculés en cumulant les dépassements aux stations reviennent à pondérer les épisodes.

À l'échéance J1, la PREVI\_STAT présente les meilleurs scores de prévision de dépassement. Les biais identifiés sur la carte ICAR faussent les scores de cette chaîne, qui n'apparaît pas sur le diagramme.

### 3.2.2.2. Scores en lien avec les critères AMU

Les figures suivantes représentent les éléments de la table de contingence concernant les critères des arrêtés mesure d'urgence en place en Île-de-France à l'échéance J0 (Figure 35) et J1 (Figure 36).

Ces éléments des tables de contingence sont établis d'après la cohérence de la validité des critères surfaciques et d'exposition de la population prévue par les prévisions J0 et J1 par rapport à l'évaluation de ces mêmes critères par l'analyse (issue d'une procédure d'assimilation de donnée pour les journées considérées). La référence pour la validation des critères est le modèle ICAR calculé à l'échéance J-1. Aucune distinction n'est faite sur le critère validé (surfacique ou population exposée).

Table de contingence des prévisions PM10 des modèles - 2022 pour l'échéance D+0

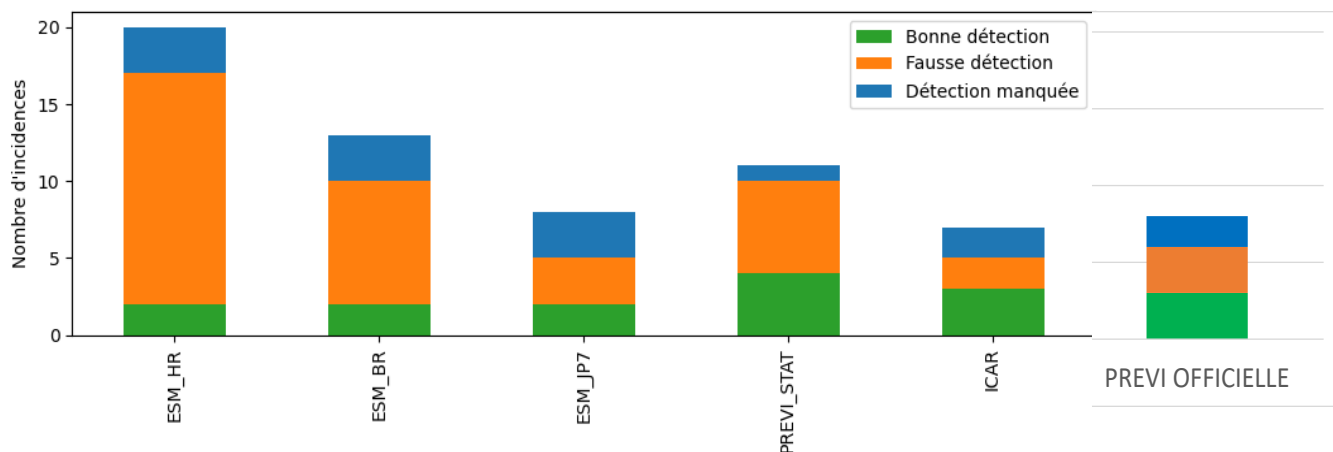


Figure 35 : Particules PM<sub>10</sub> – Table de contingence au regard des critères AMU (critères surfaciques et de population confondus) sur l'année 2022 à l'échéance J0. Les bonnes détections sont en vert, les fausses détections en orange et les détections manquées en bleu.

Table de contingence des prévisions PM10 des modèles - 2022 pour l'échéance D+1

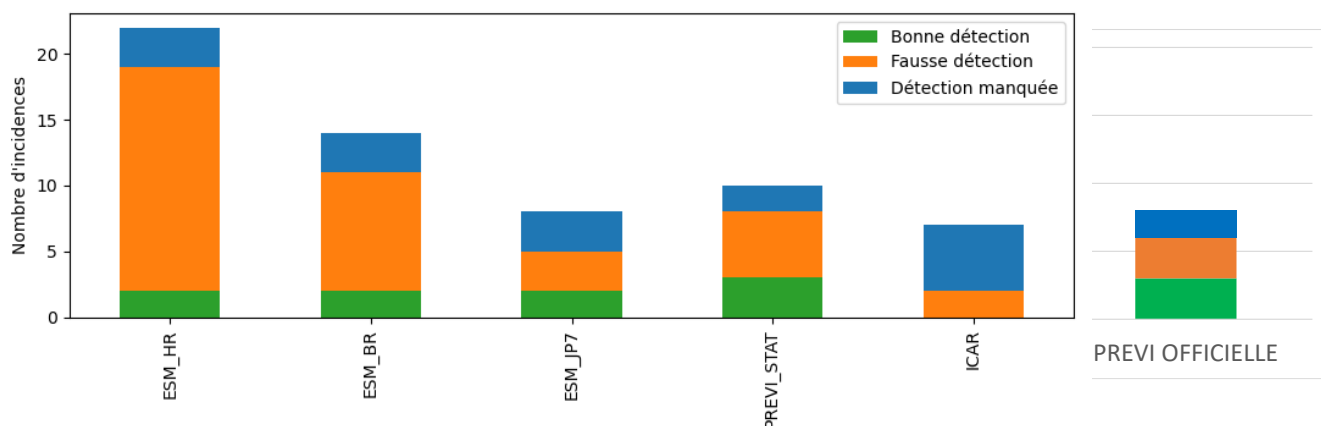


Figure 36 : Particules PM<sub>10</sub> – Table de contingence au regard des critères AMU (critères surfaciques et de population confondus) sur l'année 2022 à l'échéance J1. Les bonnes détections sont en vert, les fausses détections en orange et les détections manquées en bleu.

Comme pour la Figure 32, la Figure 36 ne reflète pas, pour les scores ICAR, les déclenchements réels de la procédure d'information et d'alerte présentés en Figure 28. La prévision officielle a donc été ajoutée aux graphiques pour mieux évaluer la plus-value des prévisionnistes.

**A J1, les chaînes de prévision présentant le plus grand nombre de bonnes détections sont PREVI\_STAT et la prévision officielle** (3 bonnes détections sur 5 jours d'épisodes), suivies des chaînes ESM\_HR, BR et JP7 (2/5).

**Les prévisionnistes ont également réalisé le plus faible nombre de fausses alertes (3 FA), comme la chaîne JP7.** Sur ces 3 fausses alertes, les concentrations maximales ont néanmoins atteint 52 µg/m<sup>3</sup> en station, et respectivement 48 et 50 µg/m<sup>3</sup> selon les critères de l'arrêt. **La prévision de dépassement était donc justifiée d'un point de vue information du public et n'est pas une erreur de prévision.** Les chaînes ESM\_HR et ESM\_BR enregistrent un nombre de fausses alertes plus important.

**Ces scores montrent la capacité des prévisionnistes à faire le bon choix de modèle lors des prévisions d'épisode.**

Les scores à J0 sont en amélioration pour la plupart des modèles. Les scores ICAR se rapprochent de la prévision officielle en raison d'un rehaussement de la fourchette de concentration à J0, qui permet de faire apparaître un dépassement sur la carte ICAR, contrairement à J1.

## 4. PISTES D'AMÉLIORATION

L'analyse fine des scores a fait apparaître des biais méthodologiques dans la chaîne ICAR et ont permis d'identifier des améliorations à prendre en compte dans la refonte prévue en 2023 :

- **La PREVI\_STAT étant actuellement le modèle de référence des prévisionnistes, au regard de ses excellents scores en 2021 et 2022, c'est ce modèle qui est utilisé en 2023 par défaut dans la feuille de synthèse.**
- La consigne a été donnée aux prévisionnistes de **privilégier l'utilisation d'une carte dont la valeur maximum prévue se situe dans la fourchette prévue plutôt que d'avoir recours à la correction de la carte.** Lorsque la valeur maximale prévue par la PREVI\_STAT se situe en dehors de la plage de concentrations prévue par les prévisionnistes, les prévisionnistes doivent utiliser, dans la mesure du possible, un modèle plus cohérent.
- Il est également envisagé de dissocier la fourchette de concentration maximale prévue pour des questions de communication de la valeur maximale à laquelle le prévisionniste souhaite corriger la carte.
- **Une réflexion est en cours pour faire évoluer la méthodologie de correction, en corrigeant uniquement les valeurs maximales, et non pas l'ensemble de la carte.**
- **La valeur maximale prévue par le modèle sera, dans la prochaine version de la chaîne, considérée sur le territoire francilien et non pas sur l'ensemble du domaine modélisé.**