Actualisation de la classification des sites de mesure

Mathieu Joly

7 février 2020

Table des matières

1 Données utilisées

- Cette version utilise un nouveau flux de l'EEA. La période a été réduite à 7 années (au lieu de 8), et comprend des données non validées pour 2019.
- Ne sont pas pris en compte les sites d'altitude supérieure à 1400 m (altitude à partir de laquelle le nombre de stations diminue fortement). En Europe, ces stations sont peu nombreuses, mais ne peuvent pas être confondues avec les sites de plaine pour l'analyse.
- Les stations renseignées comme « industrielles » ne sont pas prises en compte. La variabilité temporelle de ce type de mesure est très difficile à caractériser, et la méthode n'est pas suffisamment robuste pour appréhender le comportement potentiellement erratique des indicateurs calculés.

À partir des métadonnées, on dérive la typologie simplifiée suivante :

Type R : sites qualifiés *background* et *rural*.

Type S : sites qualifiés *background* et *suburban*.

Type U : sites qualifiés *background* et *urban*.

Type T : sites qualifiés *traffic* et *urban*.

Type O : toutes les autres stations, ainsi que les stations en dehors du sous-domaine, qui ne seront pas prises en compte pour l'Analyse Discriminante, mais qui seront classifiées *a posteriori*.

La nouveauté de cette version, c'est la disponibilité d'un certain nombre de stations en dehors du sous-domaine de l'étude. C'est en particulier le cas de la Turquie. Ces stations se retrouvent dans le « type O ».

Le CO et l'ozone font toujours figure d'anomalie, avec beaucoup de stations T dans un cas, et beaucoup de stations U et S dans l'autre. Par ailleurs, en dehors des PM dont le nombre de stations augmente significativement depuis plusieurs années, certains polluants connaissent une réduction significative du réseau de mesure.

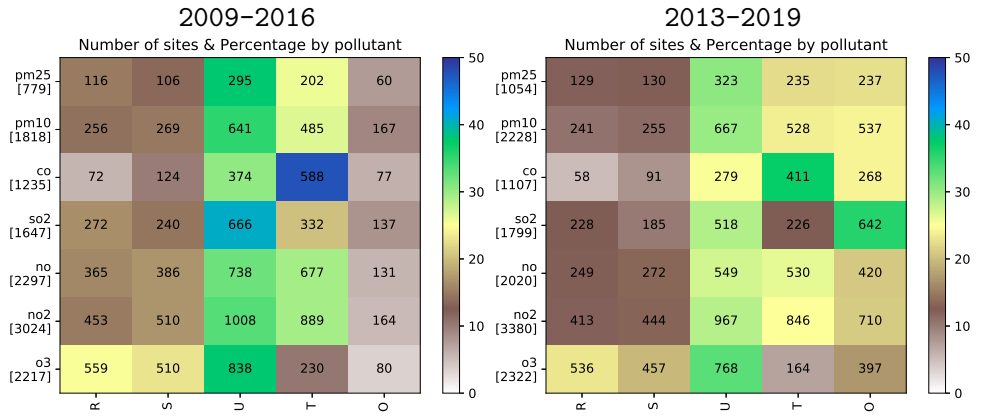

Figure 1 – Nombre de stations sélectionnées (données suffisantes), par type de métadonnée. Les couleurs correspondent au pourcentage par polluant.

2 Traitement des données

Les figures 2 et 3 montrent que les séries temporelles sur cette région, bien que suffisantes en quantité de données, ne permettent pas de calculer tous les indicateurs. C'est le cas en Italie du nord, ou en Allemagne pour le NO. Les valeurs absentes sont trop nombreuses au sein de chaque journée.

3 Cartographie du résultat

Les figures 4 et 5 illustrent la classification obtenue pour chaque polluant.

4 Validation croisée

La figure 6 compare les « validations croisées » par rapport aux types dérivés des métadonnées. La cohérence entre les classifications subjective (métadonnées) et objective est stable, sauf pour les $PM_{2.5}$, dont la performance se dégrade du fait d'un petit nombre de stations urbaines et trafic qui obtient des classes basses (1-3).

5 Étude des anomalies

Nous allons nous intéresser aux comportements marginaux de la figure 6 :

- le pourcentage des stations R qui se retrouvent dans les classes 6-10.
- le pourcentage des stations S, U et T qui se retrouvent dans les classes 1-3.

	O_3	NO_2	NO	SO_2	CO	PM_{10}	$PM_{2.5}$
R 6-10	5 → 5	2 → 2	2 → 3	26 → 28	16 → 16	18 → 16	21 → 29
S+U+T 1-3	8 → 9	3 → 3	4 → 3	10 → 12	3 → 3	7 → 6	9 → 11

Tableau 1 – Pourcentage des anomalies (cf. paragraphe ci-dessus). Évolution entre l'ancienne et la nouvelle classification (en vert pour une amélioration, en rouge pour une déterioration, et surligné de jaune quand plus de 2% des stations sont affectées).

Le tableau 1 confirme la dégration de la validation croisée pour les $PM_{2.5}$, dont les performances rejoignent celles du SO_2.

Les cartes 7 et 8 cartographient les anomalies du tableau 1. L'analyse est difficile, car il faudrait regarder localement la configuration de chacun de ces sites « douteux », et les sources de pollution environnantes. On notera néanmoins que pour les PM, les stations problématiques (cf. § ci-dessus) se trouvent majoritairement en France et en Belgique.

6 Comparaison à la précédente version

Pour les stations en commun dans les deux classifications, la figure 9 compare les classes obtenues. C'est pour le SO_2 que les deux versions sont le plus différentes.

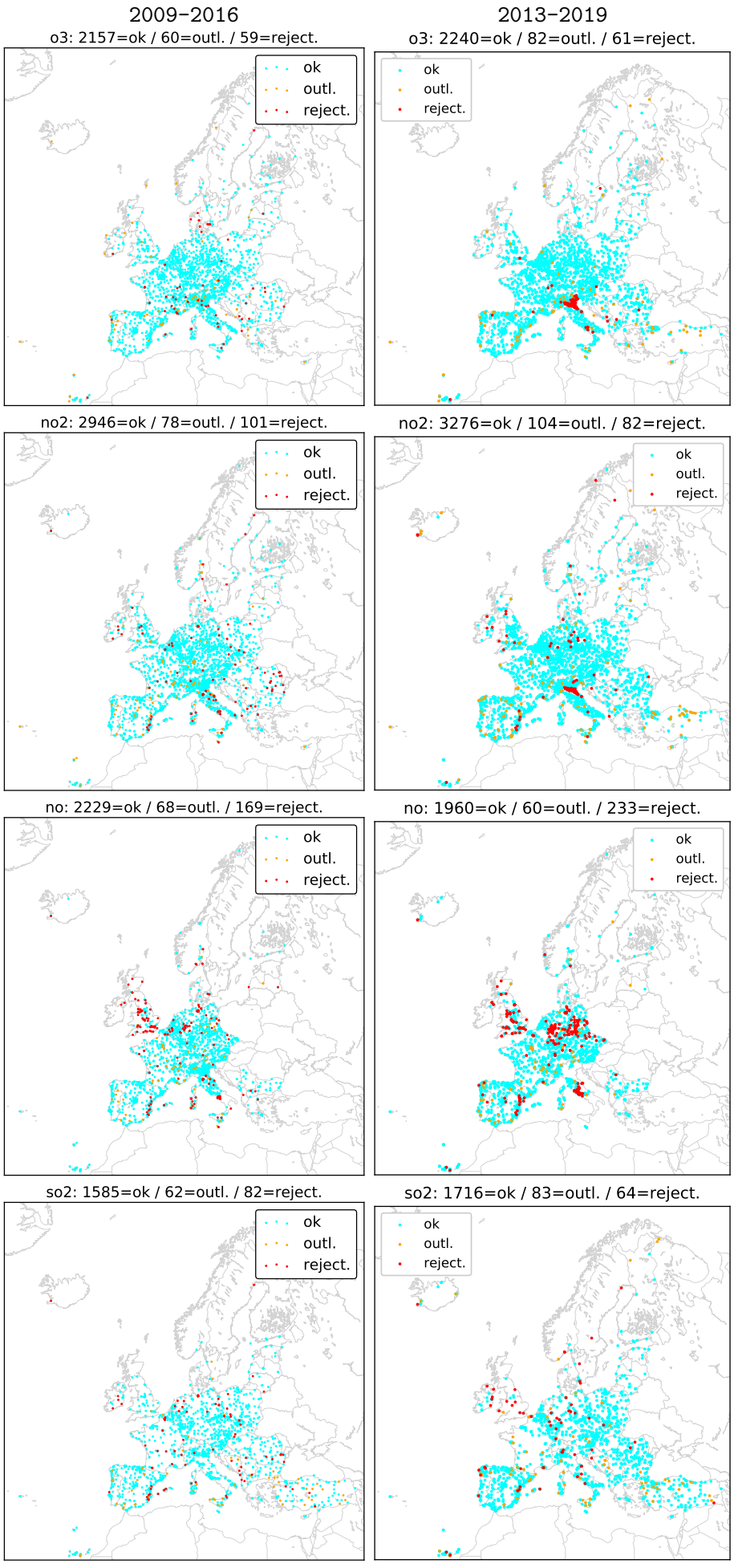

Figure 2 – Localisation des stations rejetées lors du calcul des indicateurs (*rejected*), ou lors de l'analyse (*outliers*). À gauche, pour la précédente classification ; et à droite, pour la nouvelle version.

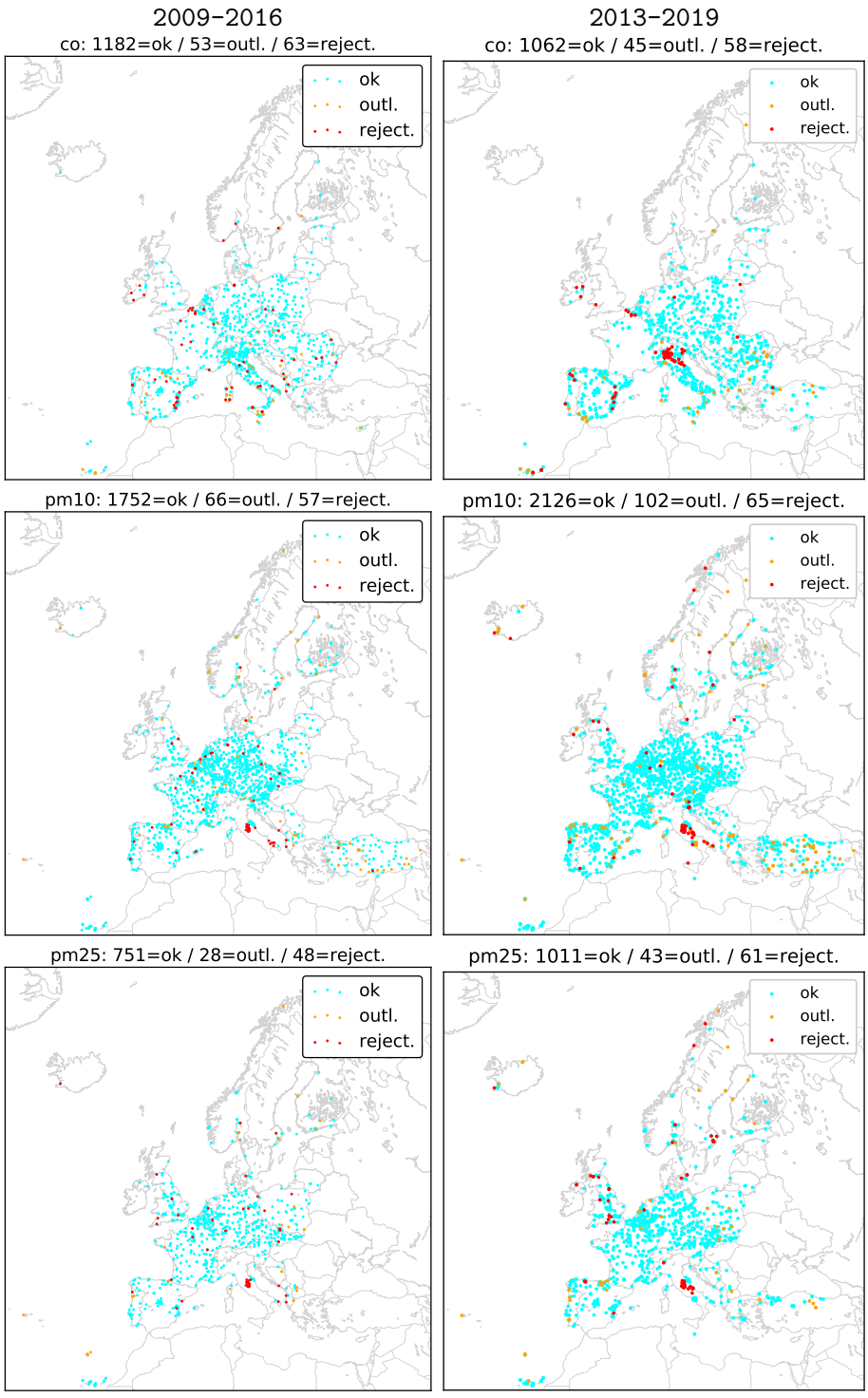

Figure 3 – Localisation des stations rejetées lors du calcul des indicateurs (*rejected*), ou lors de l'analyse (*outliers*). À gauche, pour la précédente classification ; et à droite, pour la nouvelle version.

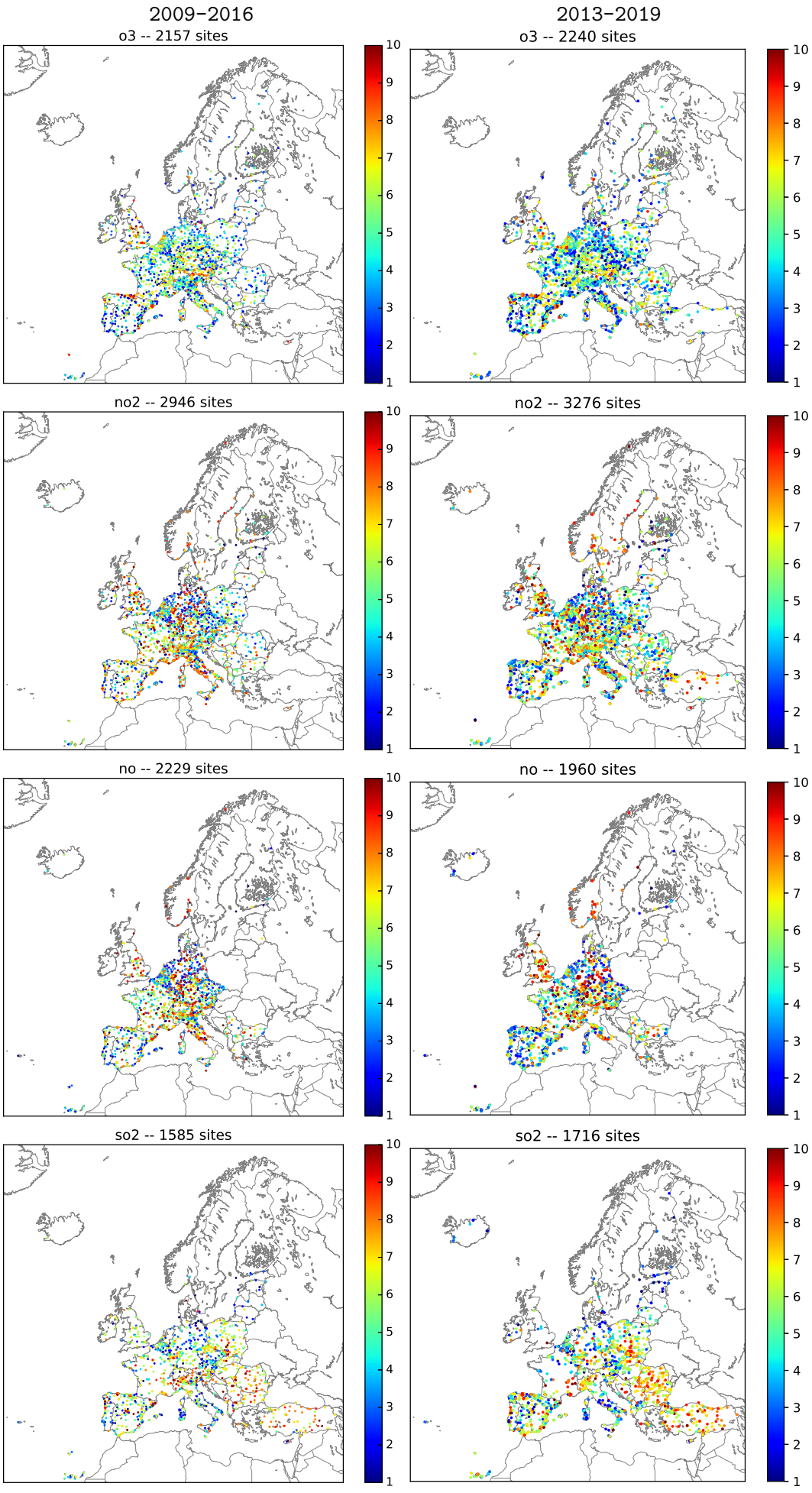

Figure 4 – Cartographie de la classification obtenue.

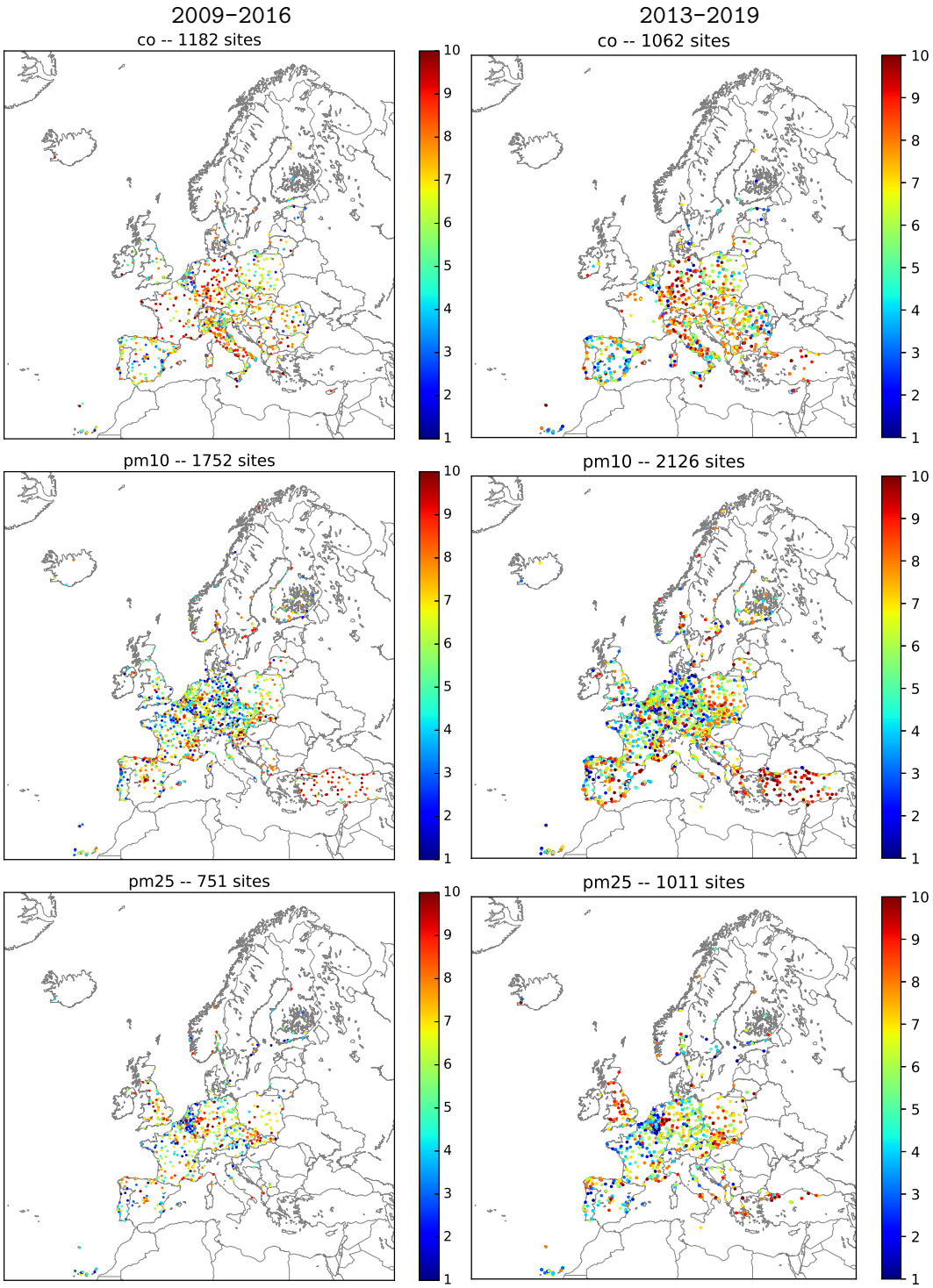

Figure 5 – Cartographie de la classification obtenue. À gauche, pour la précédente classification ; et à droite, pour la nouvelle version.

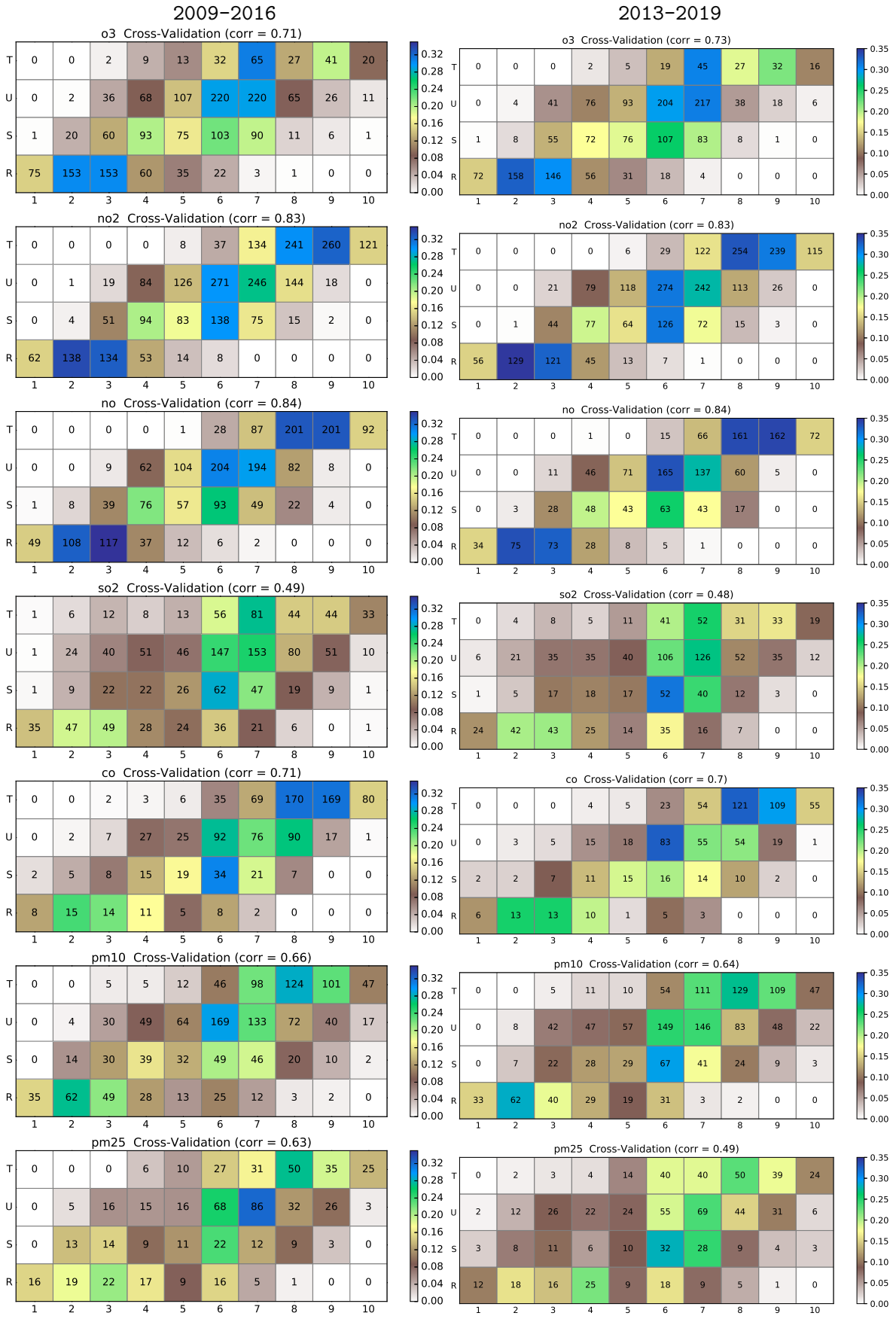

Figure 6 – Validation croisée : nombre et pourcentage (en couleur) dans chaque classe pour chaque type de station. À gauche, pour la précédente classification ; et à droite, pour la nouvelle version.

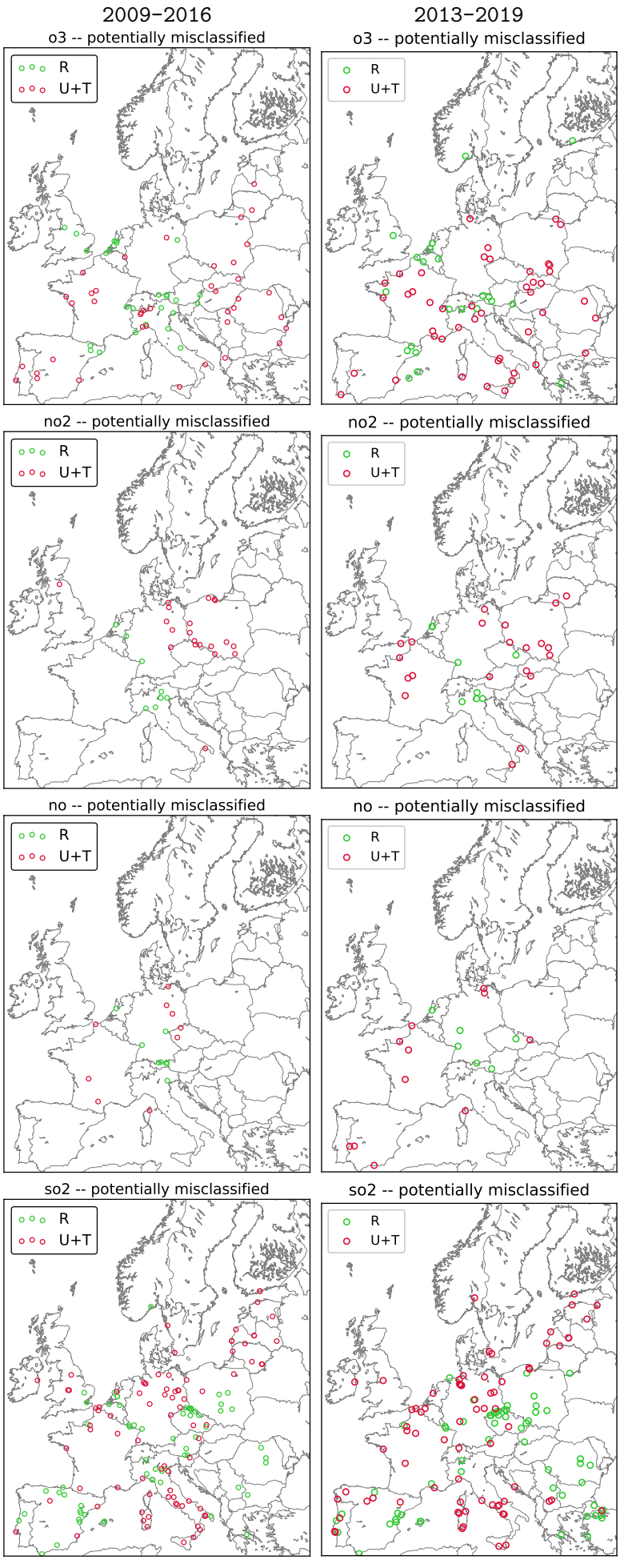

Figure 7 – Stations R qui se retrouvent dans les classes 6-10, et stations U et T qui se retrouvent dans les classes 1-3. À gauche, pour la précédente classification ; et à droite, pour la nouvelle version.

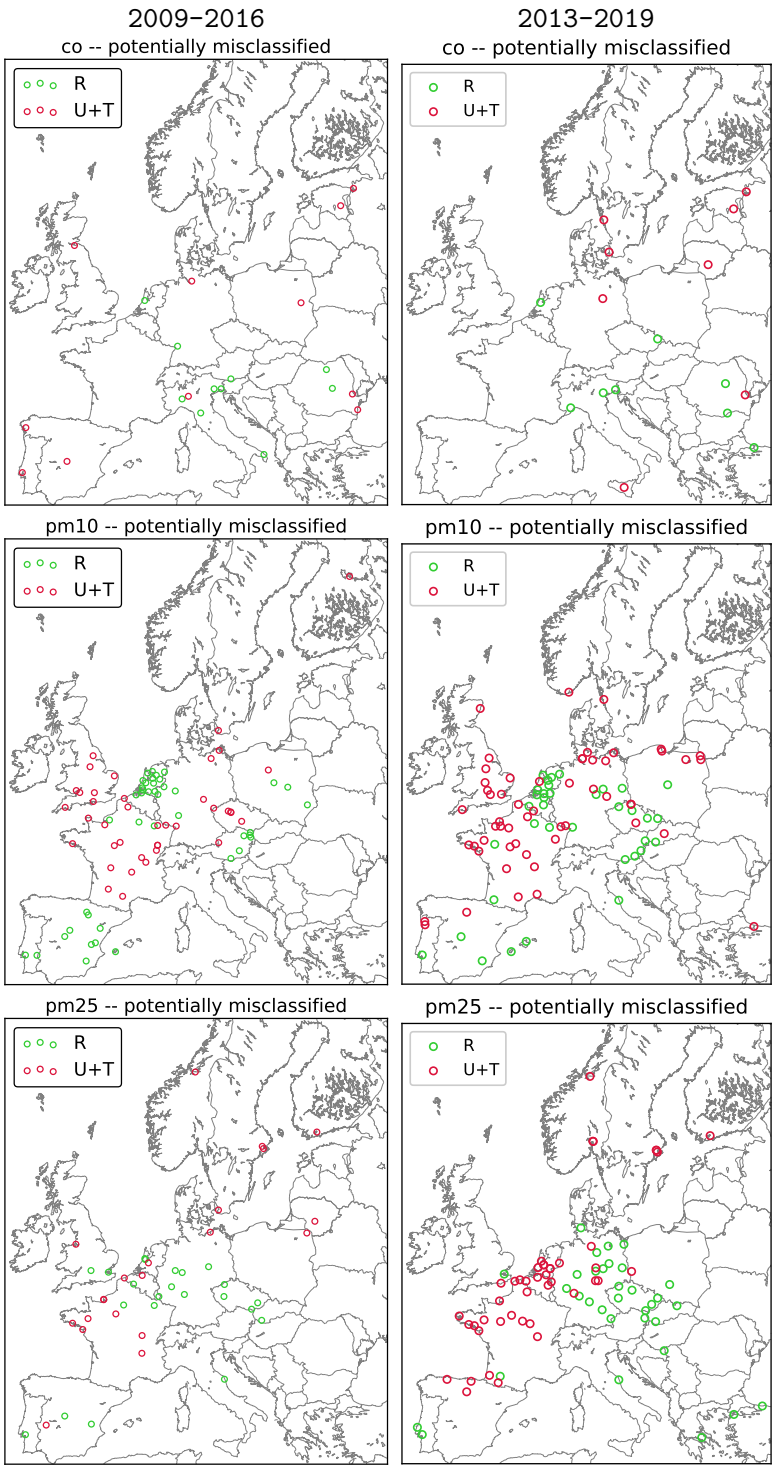

Figure 8 – Stations R qui se retrouvent dans les classes 6-10, et stations U et T qui se retrouvent dans les classes 1-3. À gauche, pour la précédente classification ; et à droite, pour la nouvelle version.

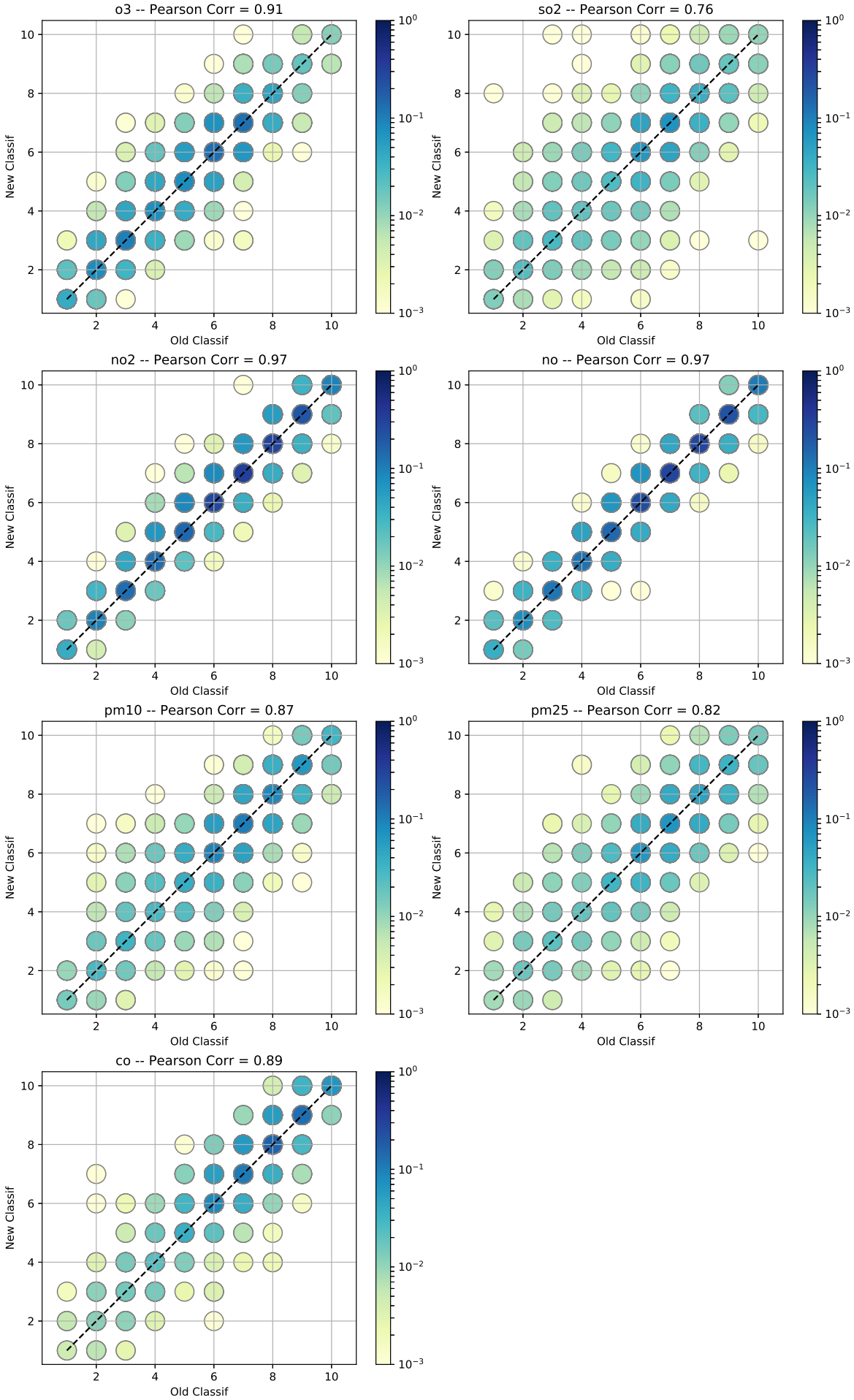

Figure 9 – Scatter Plot des classes obtenues avec l'ancienne et la nouvelle classification. La couleur indique la fréquence d'occurence.

7 Évolution du jeu de stations classifiées

La figure 10 permet de suivre l'évolution du jeux de données classifiées. On notera l'apparition de nouvelles stations en Turquie (PM_{10} et SO_2). Par contre, le réseau perdun grand nombre de stations en France et en Angleterre pour SO_2 et CO.

8 Conclusion

Cette version utilise un nouveau flux de l'EEA. La période a été réduite à 7 années (au lieu de 8), et comprend des données non validées pour 2019.

- Certains polluants connaissent une réduction significative du réseau de mesure (en France et en Angleterre pour SO_2 et CO). Par contre, de nouvelles stations apparaissent, en particulier en Turquie (PM_{10} et SO_2).
- La qualité des séries temporelles est insuffisante en certaines régions d'Italie, ou en Allemagne pour le NO. Les valeurs absentes sont trop nombreuses au sein de chaque journée.
- La cohérence entre les métadonnées et la classification objective se dégrade pour les $PM_{2.5}$, dont les performances rejoignent celles du SO_2.

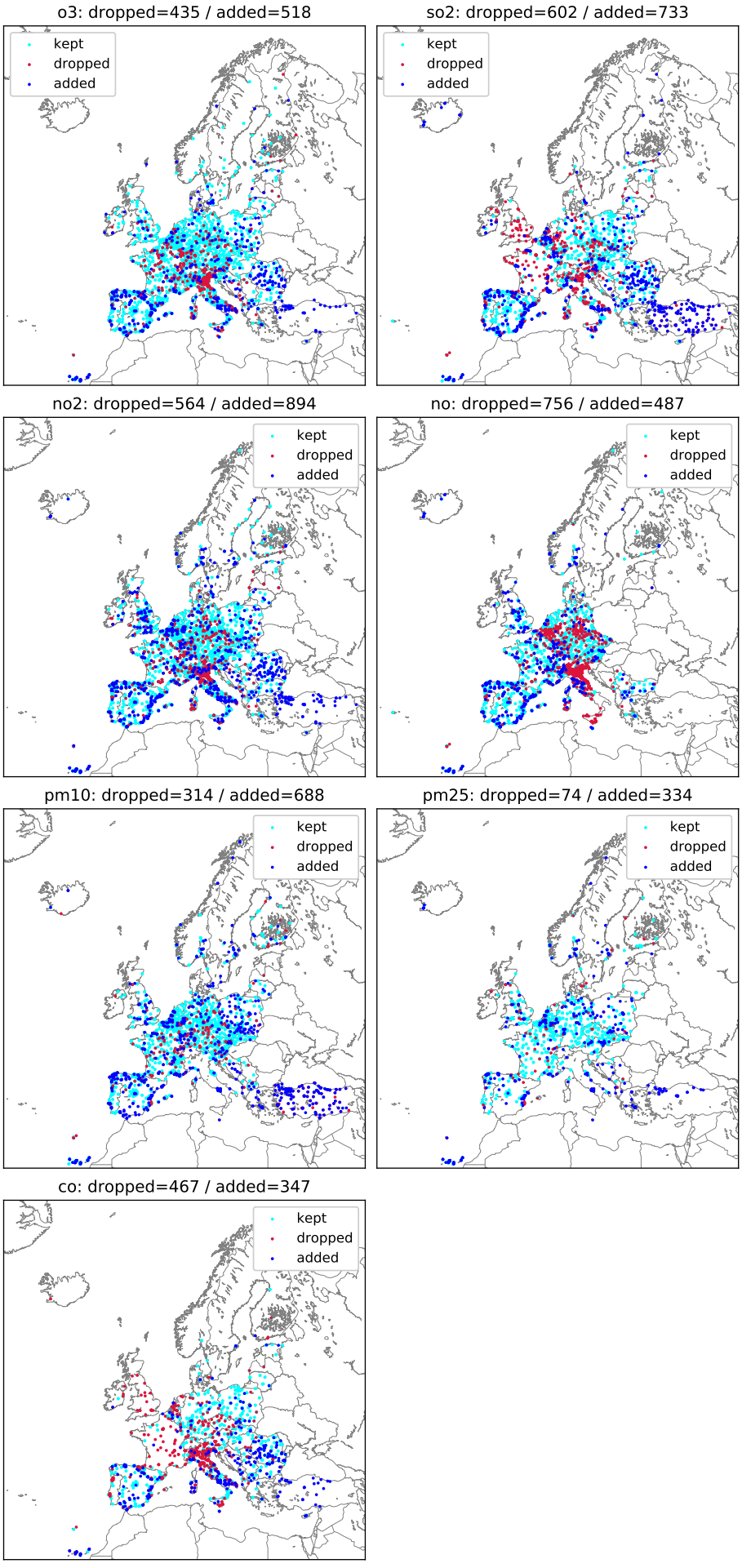

Figure 10 – Stations qui disparaissent (rouge), ou qui apparaissent (bleu) dans la nouvelle version.