Actualisation de la classification des sites de mesure de la qualité de l'air

Mathieu Joly, CNRM/GMGEC/PLASMA

Jeudi du Climat – 23 mars 2017

Où la qualité de l'air est-elle contrôlée ?

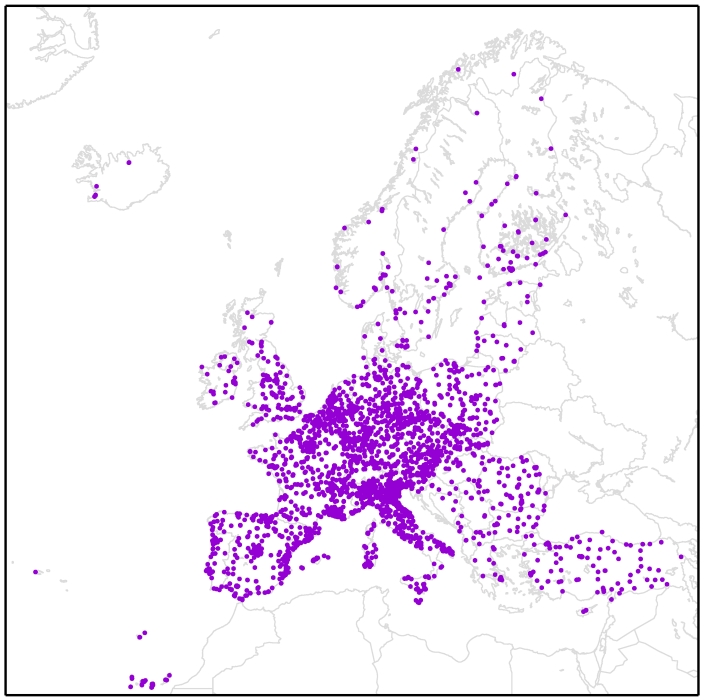
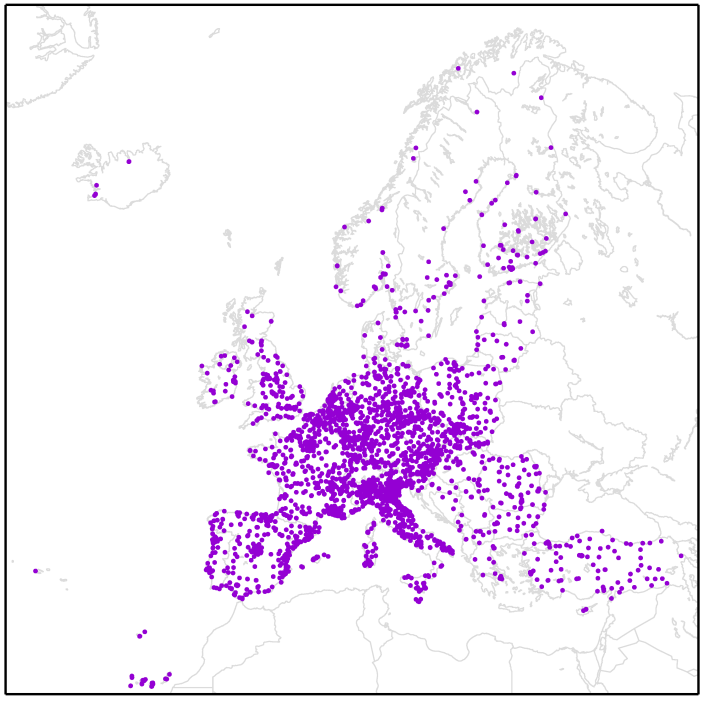


Fig. – À gauche : emplacement des 3696 sites pris en compte (tous polluants confondus). Données horaires validées de l'EEA (AQeR).

METEO FRANCE

Où la qualité de l'air est-elle contrôlée ?



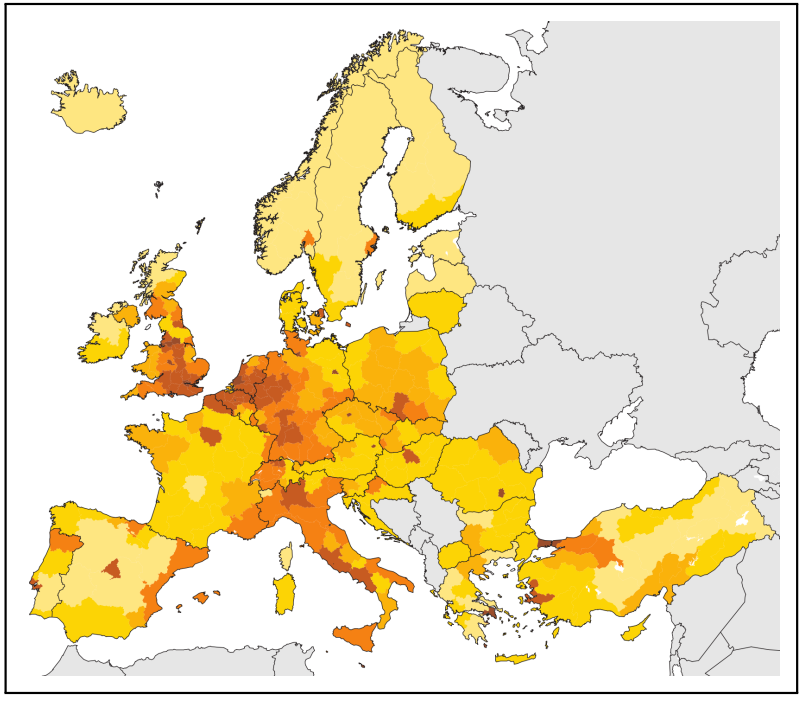


Fig. – À gauche : emplacement des 3696 sites pris en compte (tous polluants confondus). Données horaires validées de l'EEA (AQeR).
À droite : Densité de population par région en 2007 (cf. Wikipedia).

☞ Le réseau est intimement lié à la présence de population.

METEO FRANCE

Un réseau de mesure très hétérogène !

- Les polluants sont mesurés dans des endroits très variés :
  - Au plus près des sources d'émission (industrielle, trafic, biogéniques, etc) ou à distance (pollution de fond).
  - Dans des environnements contrastés : urbain (surveillance) ou ruraux (grande échelle).
  - Dans des configurations topographiques toujours singulières.

METEO FRANCE

Un réseau de mesure très hétérogène !

- Les polluants sont mesurés dans des endroits très variés :
  - Au plus près des sources d'émission (industrielle, trafic, biogéniques, etc) ou à distance (pollution de fond).
  - Dans des environnements contrastés : urbain (surveillance) ou ruraux (grande échelle).
  - Dans des configurations topographiques toujours singulières.
- Le réseau de mesure diffère selon les polluants, et selon les pays, voire selon les régions (choix d'implantation arbitraires).

METEO FRANCE

Méta-données : indispensables, mais perfectibles. . .

- Méta-données : les installateurs renseignent le type (*background*, trafic, ou industriel) et l'environnement (rural, péri-urbain, ou urbain) du site.
- La description des stations dans les métadonnées est subjective, et les critères changent selon les pays.

METEO FRANCE

Méta-données : indispensables, mais perfectibles. . .

- Méta-données : les installateurs renseignent le type (*background*, trafic, ou industriel) et l'environnement (rural, péri-urbain, ou urbain) du site.
- La description des stations dans les métadonnées est subjective, et les critères changent selon les pays.
- La représentativité spatiale des mesures est hétérogène et difficile à évaluer.

METEO FRANCE

Méta-données : indispensables, mais perfectibles. . .

- Méta-données : les installateurs renseignent le type (*background*, trafic, ou industriel) et l'environnement (rural, péri-urbain, ou urbain) du site.
- La description des stations dans les métadonnées est subjective, et les critères changent selon les pays.
- La représentativité spatiale des mesures est hétérogène et difficile à évaluer.

☞ Nécessité pour la modélisation (validation & assimilation) de sous-ensembles de données ayant des caractéristiques homogènes pour un polluant donné.

METEO FRANCE

Comment ?

Joly & Peuch (2012)

- Classification objective à l'échelle de l'Europe, à partir des propriétés des séries temporelles passées,
- Classification distincte pour chaque polluant mesuré (contrairement aux métadonnées),
- Classification à partir d'Airbase+BDQA 2002-2009 pour O_3, NO_2, NO, SO_2 et PM_{10}.

METEO FRANCE

Comment ?

Joly & Peuch (2012)

- Classification objective à l'échelle de l'Europe, à partir des propriétés des séries temporelles passées,
- Classification distincte pour chaque polluant mesuré (contrairement aux métadonnées),
- Classification à partir d'Airbase+BDQA 2002-2009 pour O_3, NO_2, NO, SO_2 et PM_{10}.

Travail effectué récemment pour Copernicus

- Actualisation pour les 5 espèces initiales + CO + $PM_{2.5}$,
- Révision complète de la méthode : amélioration & robustesse,
- Utilisation des données les plus récentes : Airbase+AQeR+BDQA 2007-2014.

METEO FRANCE

Différentes étapes (en bref...)

a) Calcul d'indicateurs décrivant les séries temporelles (si les données sont suffisantes),

Différentes étapes (en bref...)

a) Calcul d'indicateurs décrivant les séries temporelles (si les données sont suffisantes),

b) Transformation, normalisation, et validité des indicateurs,

METEO FRANCE

Différentes étapes (en bref...)

a) Calcul d'indicateurs décrivant les séries temporelles (si les données sont suffisantes),

b) Transformation, normalisation, et validité des indicateurs,

c) Analyse en Composantes Principales (nouveau) : réduction du nombre de dimensions,

METEO FRANCE

Différentes étapes (en bref...)

a) Calcul d'indicateurs décrivant les séries temporelles (si les données sont suffisantes),

b) Transformation, normalisation, et validité des indicateurs,

c) Analyse en Composantes Principales (nouveau) : réduction du nombre de dimensions,

d) Analyse Linéaire Discriminante, puis projection sur le 1^{er} axe et répartition dans 10 classes.

METEO FRANCE

METEO FRANCE

Données utilisées

- Domaine : Canaries – Cap Nord & Açores – Turquie.
- Exclusion des sites d'altitude > 1400 m, et des stations déclarées industrielles.
- Filtrage des valeurs mesurées extrêmes.

METEO FRANCE

Données utilisées

- Domaine : Canaries – Cap Nord & Açores – Turquie.
- Exclusion des sites d'altitude > 1400 m, et des stations déclarées industrielles.
- Filtrage des valeurs mesurées extrêmes.
- Répartition dans 5 groupes :

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* (nouveau).

Type O : autres sites, classifiés *a posteriori*.

METEO FRANCE

Données utilisées

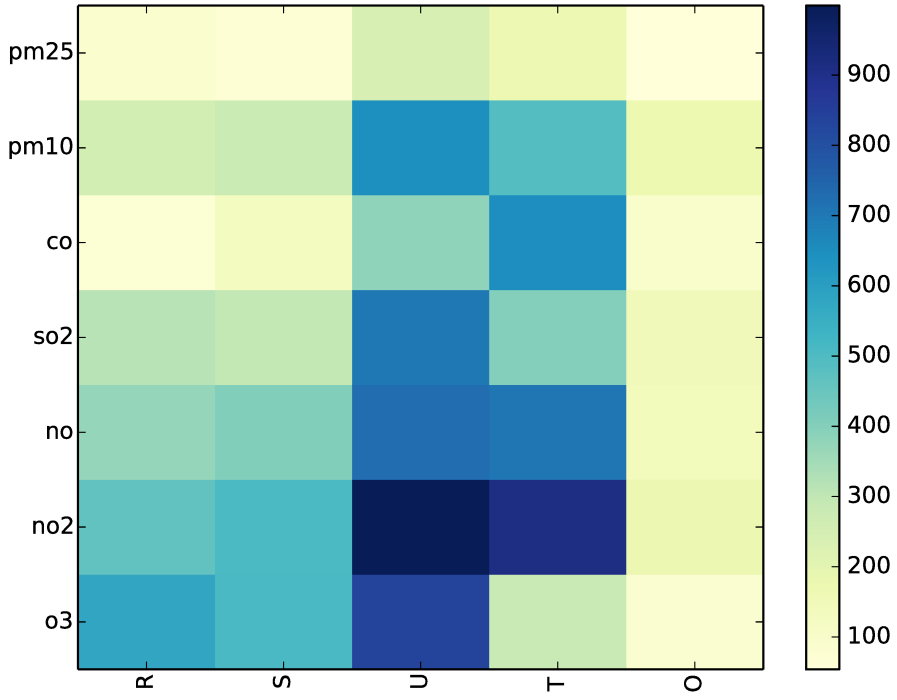


Fig. – Nombre de stations avec des données suffisantes, par groupe.

- Peu de mesures de $PM_{2.5}$,
- Déficit de stations rurales, en particulier pour le CO,
- Répartition différente selon le polluant.

METEO FRANCE

Nouveaux indicateurs

Initialement, la classification reposait sur 8 indicateurs.
De nouveaux indicateurs ont été testés :

Nouveaux indicateurs

Initialement, la classification reposait sur 8 indicateurs.
De nouveaux indicateurs ont été testés :

- Basés sur la PDF des valeurs horaires : indice de Gini (inégalité), *skewness* (asymétrie), et *kurtosis* (aplatissement).

METEO FRANCE

Nouveaux indicateurs

Initialement, la classification reposait sur 8 indicateurs.
De nouveaux indicateurs ont été testés :

- Basés sur la PDF des valeurs horaires : indice de Gini (inégalité), *skewness* (asymétrie), et *kurtosis* (aplatissement).
- Auto-corrélation des séries au $lag = 42$ (en dehors des périodicités liées au cycle diurne).

METEO FRANCE

Nouveaux indicateurs

Initialement, la classification reposait sur 8 indicateurs.
De nouveaux indicateurs ont été testés :

- Basés sur la PDF des valeurs horaires : indice de Gini (inégalité), *skewness* (asymétrie), et *kurtosis* (aplatissement).
- Auto-corrélation des séries au $lag = 42$ (en dehors des périodicités liées au cycle diurne).
- Nouveaux modes de calcul des indicateurs relatifs aux cycles diurne et annuel, et pour l'effet Weekend.

METEO FRANCE

Indicateurs retenus

Au total, 25 indicateurs ont été mis en concurrence, et 14 ont été retenus selon leur aptitude à séparer les différents types de sites (T-test de Welch & U-test de Mann–Whitney) :

Indicateurs retenus

Au total, 25 indicateurs ont été mis en concurrence, et 14 ont été retenus selon leur aptitude à séparer les différents types de sites (T-test de Welch & U-test de Mann–Whitney) :

- 3 pour caractériser la distribution des valeurs,
- 2 pour caractériser la variabilité (haute-fréquence & auto-corrélation),
- 3 pour caractériser le cycle diurne,
- 3 pour caractériser le cycle annuel,
- 3 pour caractériser l'effet Weekend.

METEO FRANCE

METEO FRANCE

Transformation, normalisation, et détection des *outliers*

a) Transformation des indicateurs dont la distribution est très asymétrique : technique automatique de Box & Cox (1964).

Transformation, normalisation, et détection des *outliers*

a) Transformation des indicateurs dont la distribution est très asymétrique : technique automatique de Box & Cox (1964).

b) Normalisation des indicateurs (indispensable avant toute technique de réduction de dimension).

METEO FRANCE

Transformation, normalisation, et détection des *outliers*

a) Transformation des indicateurs dont la distribution est très asymétrique : technique automatique de Box & Cox (1964).

b) Normalisation des indicateurs (indispensable avant toute technique de réduction de dimension).

c) Détection d'*outliers* pour chaque groupe séparément, multi-variée (méthode d'apprentissage supervisée SVM), aboutissant à environ 10% de stations par groupe (classifiées *a posteriori*).

METEO FRANCE

Analyse en Composantes Principales

Avec 14 indicateurs (au lieu de 8), se pose la question de la redondance de l'information (non souhaitable en entrée de l'Analyse Linéaire Discriminante).

METEO FRANCE

Analyse en Composantes Principales

Avec 14 indicateurs (au lieu de 8), se pose la question de la redondance de l'information (non souhaitable en entrée de l'Analyse Linéaire Discriminante).

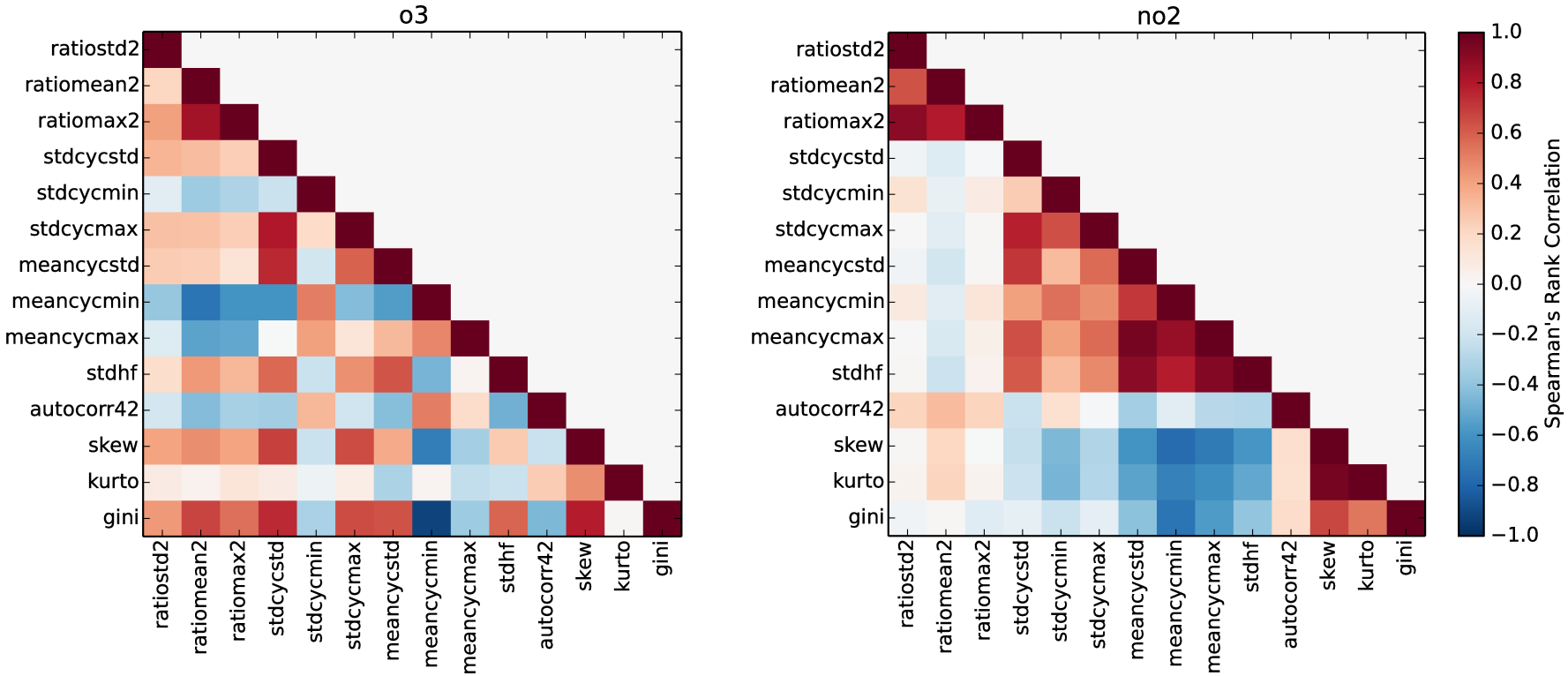


Fig. – Corrélations entre indicateurs pour l'ozone et le NO_2.

METEO FRANCE

Analyse en Composantes Principales

- L'ACP est utilisée pour réduire le nombre de dimensions en trouvant les directions qui maximisent la variance.
- Les 7 premières composantes principales permettent d'expliquer 93% à 98% de la variance totale.

METEO FRANCE

Analyse en Composantes Principales

- L'ACP est utilisée pour réduire le nombre de dimensions en trouvant les directions qui maximisent la variance.
- Les 7 premières composantes principales permettent d'expliquer 93% à 98% de la variance totale.

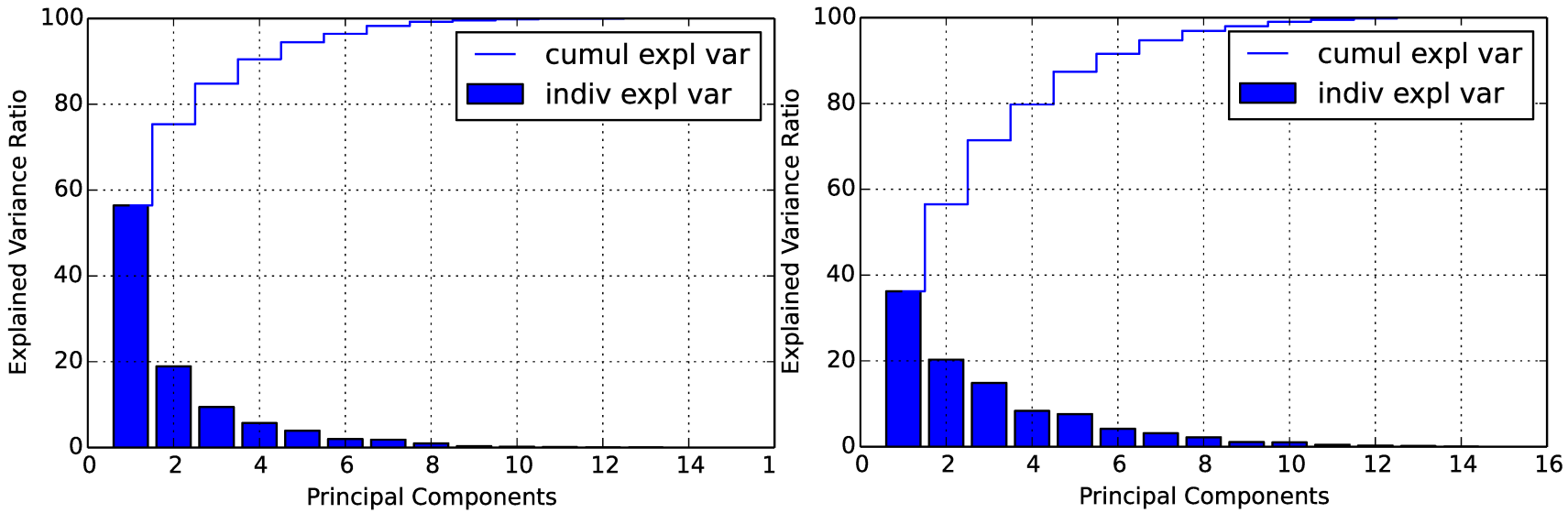


Fig. – Variance expliquée par les composantes du NO et des $PM_{2.5}$.

METEO FRANCE

Analyse Linéaire Discriminante

ACP : privilégie les directions qui maximisent la variance.

ALD : privilégie les directions qui séparent le mieux différents groupes (algorithme dit « supervisé »).

METEO FRANCE

Analyse Linéaire Discriminante

ACP : privilégie les directions qui maximisent la variance.
ALD : privilégie les directions qui séparent le mieux différents groupes (algorithme dit « supervisé »).

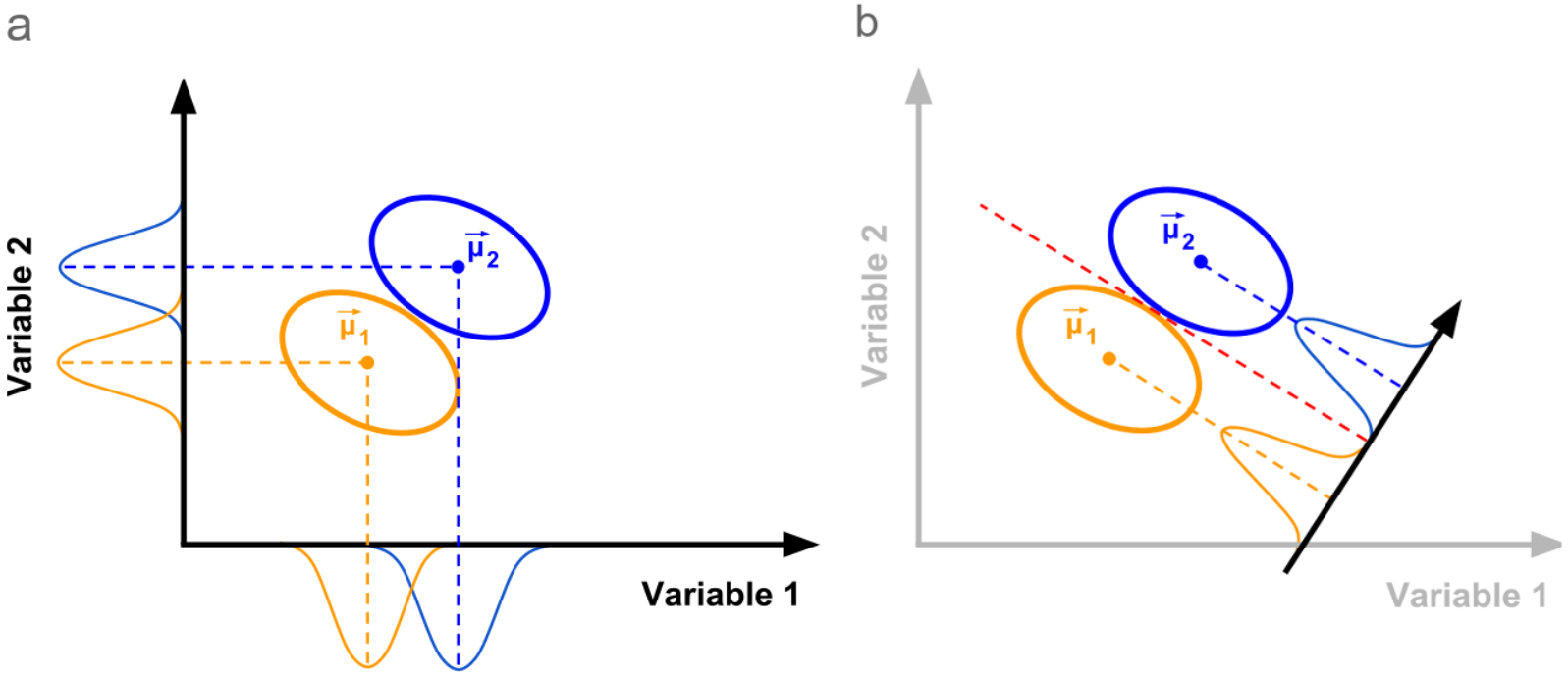


Fig. – Principe de l'Analyse Linéaire Discriminante : maximiser la distance inter-groupes & minimiser la variance intra-groupe.

METEO FRANCE

Analyse Linéaire Discriminante

Initialement, l'ALD était réalisée entre les 2 groupes : R & U+T.
On utilise désormais une ALD multiple pour séparer les 4 groupes R, S, U, et T.

Analyse Linéaire Discriminante

Initialement, l'ALD était réalisée entre les 2 groupes : R & U+T.
On utilise désormais une ALD multiple pour séparer les 4 groupes R, S, U, et T.

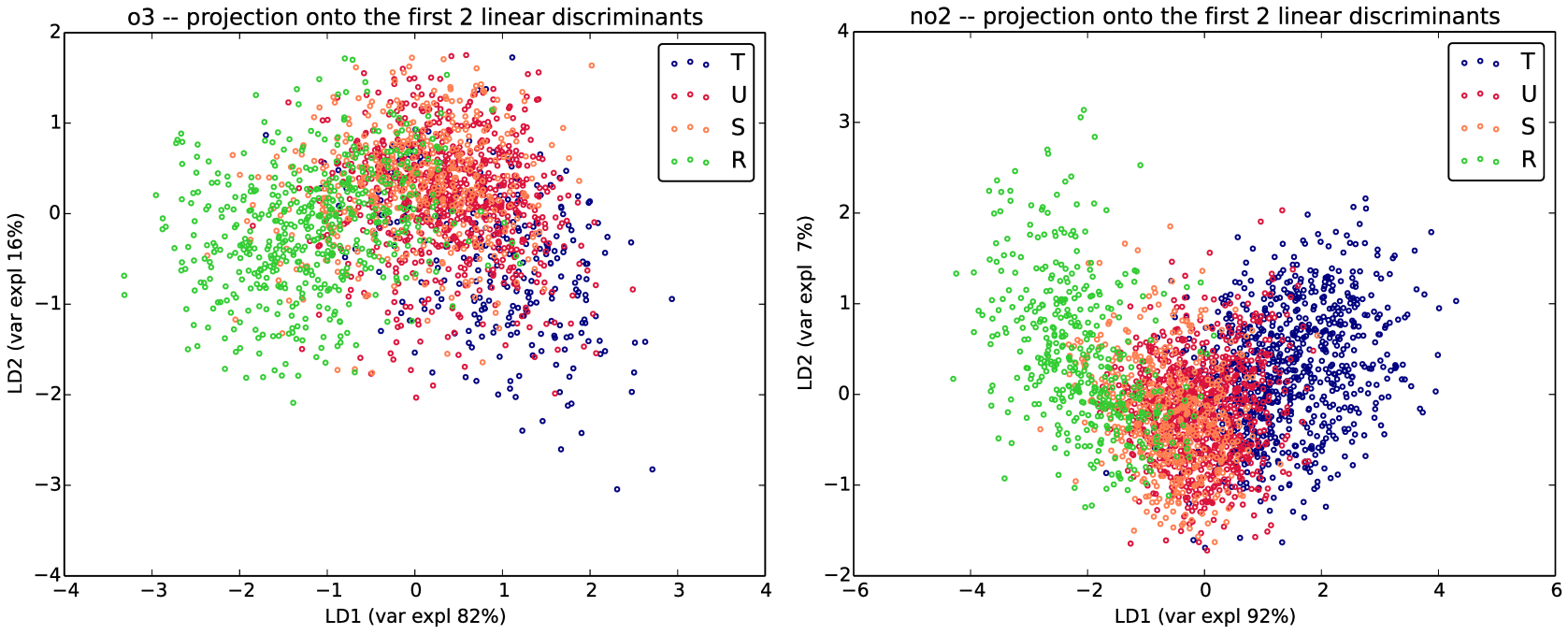


Fig. – Projection sur les 2 premiers axes de l'ALD multiple pour l'ozone et le NO_2.

METEO FRANCE

Détermination des classes

- Initialement, la projection sur l'axe de Fisher était « découpée » selon les 9 déciles des valeurs.

  Avantage : classes de même effectif.

  Inconvénient : comparaison impossible entre les polluants.

METEO FRANCE

Détermination des classes

- Initialement, la projection sur l'axe de Fisher était « découpée » selon les 9 déciles des valeurs.

  Avantage : classes de même effectif.

  Inconvénient : comparaison impossible entre les polluants.
- Désormais, on tient compte de la proportion des différents types de stations.

METEO FRANCE

Détermination des classes

- Initialement, la projection sur l'axe de Fisher était « découpée » selon les 9 déciles des valeurs.

  Avantage : classes de même effectif.

  Inconvénient : comparaison impossible entre les polluants.
- Désormais, on tient compte de la proportion des différents types de stations.

☞ Si la classification et les métadonnées étaient parfaitement cohérentes, on aurait pour chaque polluant :

$$R = 1\text{-}3,\ S = 4\text{-}5,\ U = 6\text{-}7,\ \text{et}\ T = 8\text{-}10.$$

METEO FRANCE

Détermination des classes

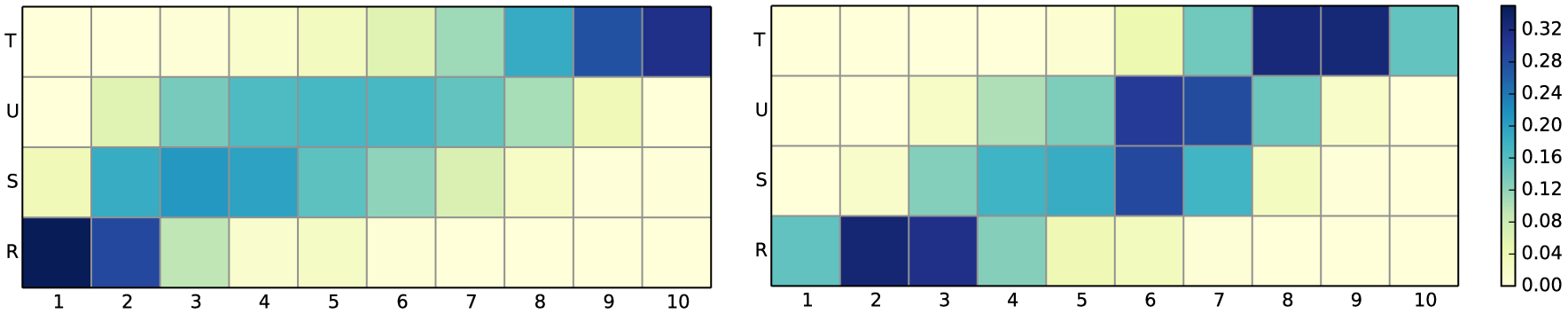


Fig. – NO_2 : pourcentage dans chaque classe pour chaque type de station.
À gauche : en déterminant les bornes à partir des déciles.
À droite : avec la nouvelle méthode.

- La partie rurale est mise en valeur.
- La partie urbaine est plus concentrée.

METEO FRANCE

Cartographie des résultats

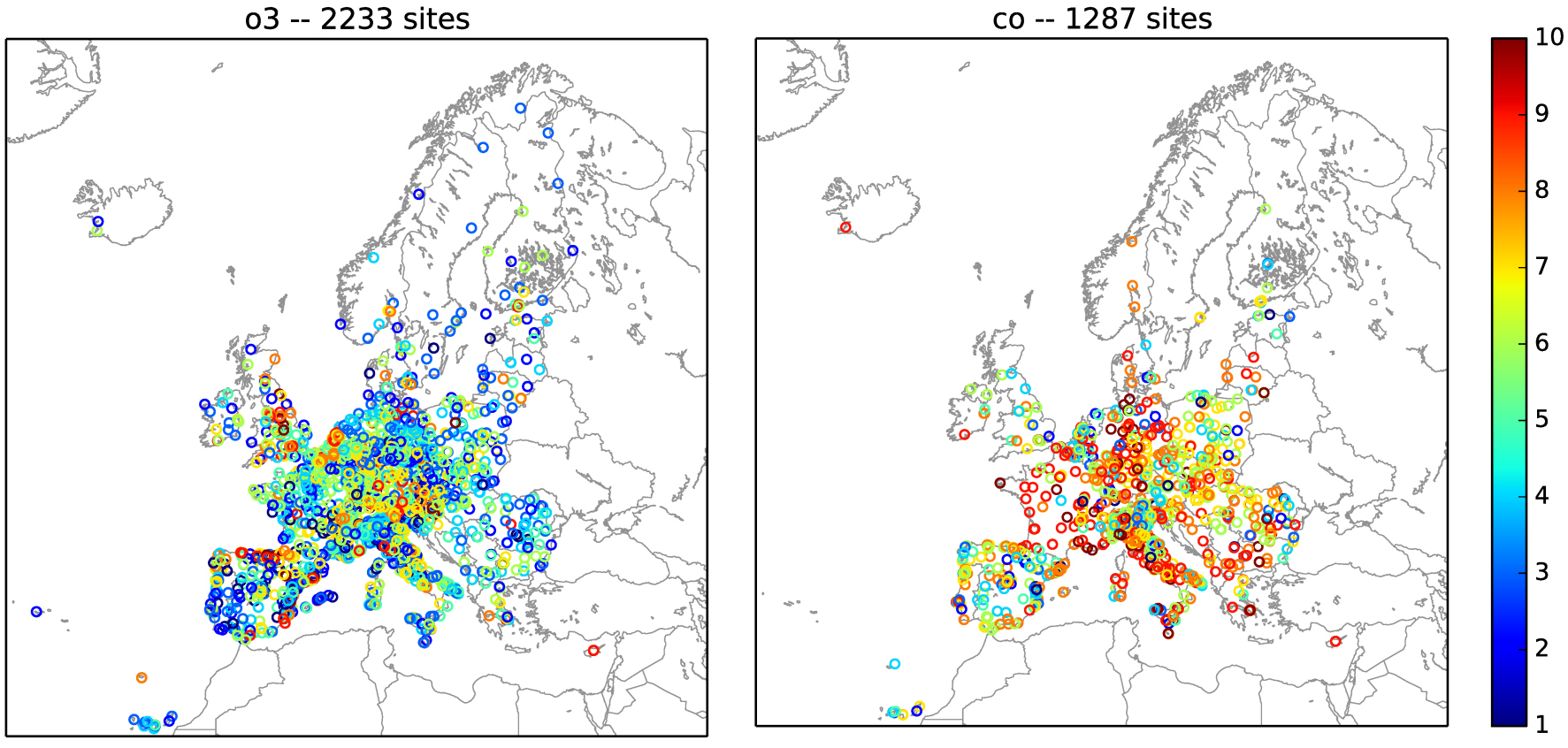


Fig. – Cartographie de la classification de l'ozone et du CO.

☞ Les caractéristiques du réseau de mesure sont plus intelligibles avec cette nouvelle façon de définir les classes.

METEO FRANCE

METEO FRANCE

Comparaison à la précédente classification

L'algorithme de Joly & Peuch (2012) a été appliqué aux nouvelles données (2007 à 2014), avec un minimum de modifications (limites géographique, exclusion des sites d'altitude, T = *traffic*+*urban*).

METEO FRANCE

Comparaison à la précédente classification

L'algorithme de Joly & Peuch (2012) a été appliqué aux nouvelles données (2007 à 2014), avec un minimum de modifications (limites géographique, exclusion des sites d'altitude, T = *traffic+urban*).

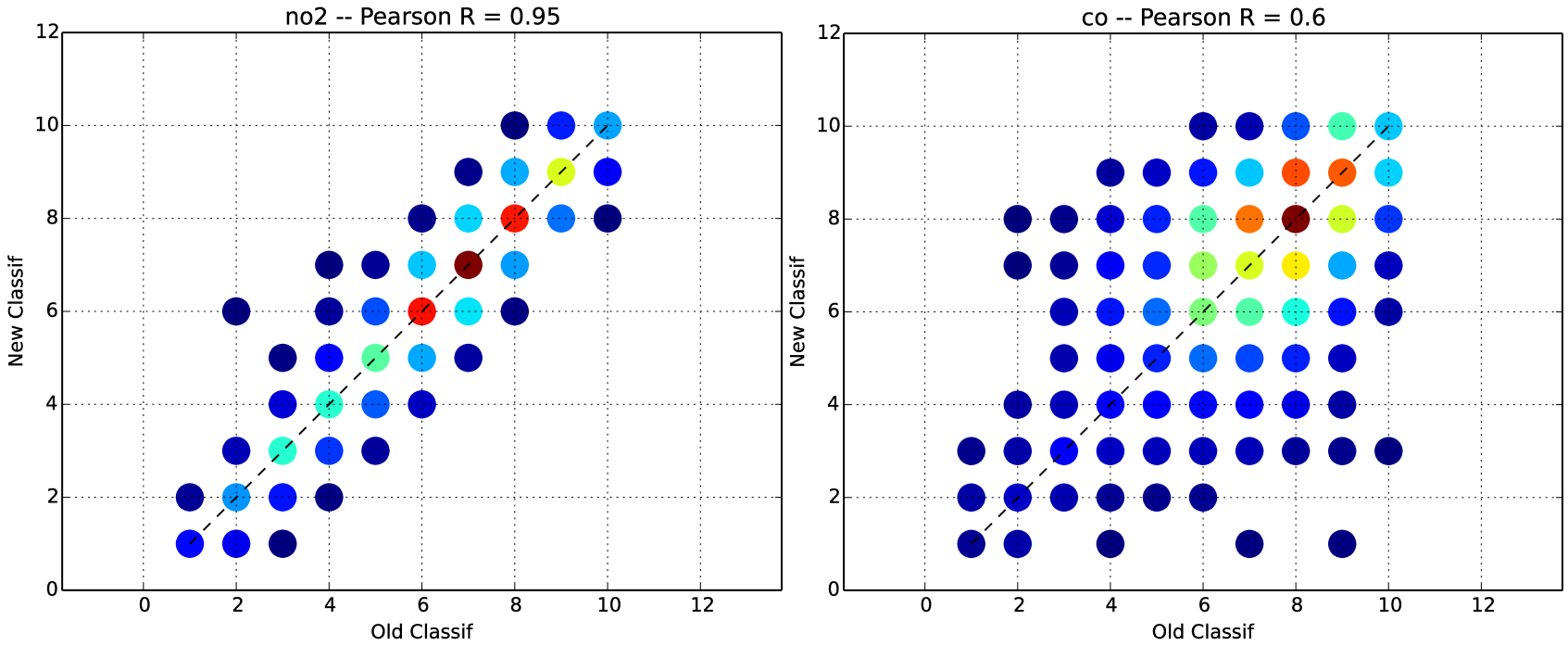


Fig. – NO_2 & CO : comparaison des classes obtenues avec l'ancien et le nouvel algorithme. La couleur indique la fréquence d'occurence.

METEO FRANCE

Quantification des anomalies

Si on s'intéresse aux comportements marginaux :

	O_3	NO_2	NO	SO_2
R 6-10	12 → 6	5 → 4	4 → 4	30 → 31
S+U+T 1-3	12 → 10	4 → 4	5 → 4	10 → 11

	CO	PM_{10}	$PM_{2.5}$
R 6-10	24 → 14	11 → 17	29 → 26
S+U+T 1-3	2 → 2	6 → 6	9 → 8

Tab. – Évolution entre l'ancien et le nouvel algorithme. Pourcentage des anomalies.

- Amélioration pour la plupart des polluants,
- Déterioration pour les stations rurales des PM_{10} (petit nombre de sites passant de la classe 4 à la classe 6).

METEO FRANCE

Conclusion

- Mise à jour pour Copernicus de la classification des sites de mesure pour O_3, NO_2, NO, SO_2 et PM_{10},

Conclusion

- Mise à jour pour Copernicus de la classification des sites de mesure pour O_3, NO_2, NO, SO_2 et PM_{10},
- Extension réussie à CO et $PM_{2.5}$ (malgré certaines carences du réseau),

METEO FRANCE

Conclusion

- Mise à jour pour Copernicus de la classification des sites de mesure pour O_3, NO_2, NO, SO_2 et PM_{10},
- Extension réussie à CO et $PM_{2.5}$ (malgré certaines carences du réseau),
- Un algorithme *a priori* plus complet, plus performant, plus robuste, et automatisé pour faciliter les futures mises à jour,

METEO FRANCE

Conclusion

- Mise à jour pour Copernicus de la classification des sites de mesure pour O_3, NO_2, NO, SO_2 et PM_{10},
- Extension réussie à CO et $PM_{2.5}$ (malgré certaines carences du réseau),
- Un algorithme *a priori* plus complet, plus performant, plus robuste, et automatisé pour faciliter les futures mises à jour,
- Précision et limites de la méthode : testées et discutées dans Joly & Peuch (2012).

METEO FRANCE

Conclusion

- Mise à jour pour Copernicus de la classification des sites de mesure pour O_3, NO_2, NO, SO_2 et PM_{10},
- Extension réussie à CO et $PM_{2.5}$ (malgré certaines carences du réseau),
- Un algorithme *a priori* plus complet, plus performant, plus robuste, et automatisé pour faciliter les futures mises à jour,
- Précision et limites de la méthode : testées et discutées dans Joly & Peuch (2012).

☞ Pour Copernicus, évaluation de l'impact des nouvelles classes sur les analyses & la vérification (scores, figures, etc).

METEO FRANCE