

NOTE MÉTHODOLOGIQUE :**MÉTHODE D'ANALYSE DES TENDANCES APPLIQUÉE AUX DONNÉES DE L'OBSERVATOIRE MERA****1. Méthode d'analyse des tendances**

Plusieurs **tests statistiques non-paramétriques** existent afin de déterminer les **tendances monotones à long terme**. Le principal objectif de l'analyse des tendances est de savoir objectivement s'il existe un **changement systématique significatif** dans une série chronologique.

Similairement à l'analyse de tendances nationales menée par le LCSQA*, les tendances sont caractérisées par deux tests statistiques non paramétriques dans le cadre des données collectées sur les stations MERA: **Mann-Kendall** pour déterminer l'**existence** et la **significativité** d'une **tendance globale et linéaire** au sein d'une série chronologique et **Sen-Theil** pour le calcul de la **pente** de cette tendance. Ces méthodes sont fondées sur l'étude de la **distribution des évolutions par paires**.

L'idée du test Mann-Kendall est d'évaluer jusqu'à quel point les concentrations (x) en une espèce donnée d'une série chronologique de n années de mesure (ou saisons, mois ...) corrélient temporellement. La corrélation est mesurée à l'aide de la statistique S de Mann-Kendall, obtenue en comparant deux à deux l'observation de l'année k à toute donnée d'année postérieure j (j > k). La statistique S s'obtient de la manière suivante :

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k) \quad \text{où} \quad \begin{cases} \text{sgn}(x_j - x_k) = 1 \text{ si } x_j - x_k > 0 \\ \text{sgn}(x_j - x_k) = 0 \text{ si } x_j - x_k = 0 \\ \text{sgn}(x_j - x_k) = -1 \text{ si } x_j - x_k < 0 \end{cases}$$

La statistique S est finalement la différence entre le nombre de différences positives et négatives. Plus la valeur de S est importante, plus la tendance est significative. A l'inverse, une valeur de S proche de 0 indique que les données ne montrent aucun lien avec le temps. Quand n est suffisamment important (**n > 10**), la statistique S peut être distribuée comme **une loi normale de moyenne nulle** et de variance calculée selon la formule suivante :

$$\text{Var}(S) = \frac{1}{18} \left[n(n-1)(2n+5) - \sum_{p=1}^q t_p(t_p-1)(2t_p+5) \right]$$

Avec t_p le nombre de fois qu'un groupe de p observations identiques est rencontré dans une série chronologique et q le plus grand nombre d'observations incluses dans un groupe d'observations identiques. La **présence d'une tendance statistiquement signifiante** peut être évaluée en étudiant la **distribution réduite Z** :

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{var}(S)}} & \text{si } S > 0 \\ 0 & \text{si } S = 0 \\ \frac{S+1}{\sqrt{\text{var}(S)}} & \text{si } S < 0 \end{cases} \rightarrow \begin{cases} Z > 0 & \text{tendance monotone croissante} \\ Z < 0 & \text{tendance monotone décroissante} \end{cases}$$

* Rapport d'étude LCSQA, Analyse de tendances nationales en matière de la qualité de l'air, Programme 2016, url : <https://www.lcsqa.org/fr/rapport/2016/ineris-imt-lille-douai/analyse-tendances-nationales-matiere-qualite-air>

La tendance est considérée comme **significative** en fonction d'un **seuil de risque α** donné (correspondant au risque de conclure à tort à l'existence d'une tendance). Lorsque $|Z|$ est supérieure à $Z_{\alpha/2}$, Z est jugée compatible avec l'hypothèse d'existence d'une tendance. Une autre façon de réaliser le test est de calculer la **valeur-p** de la statistique Z qui correspond à la **probabilité** d'obtenir une valeur de la statistique Z inférieure ou égale à celle observée dans la série. Plus faible est la valeur-p, plus on peut conclure avec assurance à l'existence d'une tendance à long terme.

En faisant l'hypothèse d'une **tendance linéaire**, la concentration peut se décomposer comme suit :

$C_t = B + Q.t + \varepsilon_t$, avec C_t la concentration estimée pour l'année t (ou saison, mois ...), B l'ordonnée à l'origine, Q la pente et ε_t le résidu de moyenne nulle. Dans le test de Sen, **un ensemble de pentes** Q_{jk} est estimé pour chacune des $n(n-1)/2$ paires de données de la manière suivante :

$$\forall k \in [1, n-1], \forall j \in [k+1, n], Q_{jk} = \frac{x_j - x_k}{j - k} \quad (j > k)$$

La **pente de Sen** est alors la **médiane** de l'ensemble des pentes Q_{jk} estimées. Elle s'exprime en unité de concentration par an ou, après transformation (rapport EMEP de 2016 **), en pourcentage par an. Elle peut être assortie d'un intervalle de confiance (en général 95 %).

2. Application de la méthode d'analyse aux données MERA

Dans le cas de l'analyse des tendances des **concentrations moyennes annuelles** des polluants mesurés aux stations MERA, les tests non-paramétriques ont été réalisés sous R avec la fonction **TheilSen** du pack **OpenAir** en considérant un **α fixé à 5 %**. Les tendances ont été jugées suffisamment significatives si les **valeurs-p** associées étaient **inférieures à 5 %**. Pour certains polluants, nous disposons à l'heure actuelle de séries chronologiques de moins de 10 années de mesure, les analyses des tendances ont tout de même été initiées dans les fiches de résultats synthétiques concernées mais les tendances doivent être considérées avec précaution et devront être consolidées dans les années à venir.

Toute modification fera l'objet de la mise à jour de la note technique

** EMEP/CCC Report 1/2016 : Air pollution trends in the EMEP region between 1990 and 2012, url : https://www.unece.org/fileadmin/DAM/env/documents/2016/AIR/Publications/Air_pollution_trends_in_the_EMEP_region.pdf