

## Description des algorithmes importants

### Informations

Auteurs : Fabien GELUS, Baptiste PAULETTO, Louis UNG, Mengxin ZHANG  
Version courante : V0.3  
Date de création : 01/04/2019  
Dernière modification : 02/05/2019

Dans le cadre de notre projet, nous distinguons 5 types d'algorithmes importants :

- La récupération des données depuis les fichiers csv
- Le calcul des valeurs caractéristiques d'un lieu pour chaque type de mesure
- Le calcul de la qualité moyenne de l'air
- Repérage de capteurs ayant un comportement similaire
- Repérage de capteurs ayant un dysfonctionnement

### Récupération des données

Tout d'abord, les données de mesure seront stockées en local dans des fichiers CSV.

Lorsque le programme sera lancé, il parcourra chaque fichier CSV et extraira les données dans des structures de données comme des listes (cela dépendra du langage de programmation utilisé).

La structure de données devra favoriser une recherche rapide des éléments car souvent on aura besoin d'extraire une partie des éléments de la liste pour faire des calculs selon des critères géographiques et temporels.

Comme il y a 3 types de fichiers, nous aurons 3 structures de données.

- Pour les types de mesures, le choix de la structure de données aura peu d'impact car il n'y aura que 4 éléments dans la liste. Nous avons décidé d'utiliser un vector.
- Pour les capteurs, nous aurons besoin de pouvoir insérer rapidement à la fin dans la structure cependant nous ne pouvons pas trier ces capteurs par coordonnées GPS, il faudrait définir par rapport à un point. Nous avons donc décidé d'utiliser ici aussi un vector.
- Pour les mesures, nous aurons besoin de les insérer rapidement à la fin, d'autoriser les doublons et de les parcourir rapidement en les triant au préalable par ordre chronologique. C'est pourquoi nous avons décidé de prendre la structure de données multiset.

Ces structures de données auront une durée de vie aussi longue que celle du programme.

### Calcul des valeurs caractéristiques

Les valeurs caractéristiques d'un lieu correspondent à des moyennes pondérées de chaque type de mesure. Les types de mesures sont au nombre de 4 : SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub> et les particules en suspension. Les valeurs caractéristiques correspondent donc à 4 moyennes pondérées.

A chaque fois que nous calculerons des valeurs caractéristiques, l'utilisateur rentrera en paramètre les coordonnées du lieu et une période de temps dans laquelle les mesures auront été faites. Il y aura donc des sous-structure de données qui vont parcourir les structures de données créées à l'initialisation du programme pour extraire les mesures faites pendant l'intervalle de temps indiqué.

En ce qui concerne le lieu, l'utilisateur doit choisir un point précis.

Nous chercherons alors les mesures des 3 capteurs les plus proches dans un rayon de 10 kilomètres. Ensuite, à chaque capteur sera attribué un coefficient proportionnel à la distance du capteur par rapport au point précis donné. Plus le capteur sera proche, plus le coefficient sera grand. Nous avons arbitrairement décidé que le coefficient sera calculé selon la formule suivante :

$$\alpha = 10 - d$$

$\alpha$  : coefficient de distance

$d$  : distance par rapport au point donné en km

Par exemple, un capteur situé à 1 km aura un coefficient de 10 et un capteur situé à 10 km aura un coefficient de 1.

Finalement, pour chaque capteur nous calculerons la moyenne de chaque type de mesure et on fera 4 moyennes générales (1 pour chaque type de mesure) à partir de ces moyennes en associant à chaque capteur le coefficient de distance mentionné précédemment.

### Calcul de la qualité moyenne de l'air

L'utilisateur choisira une zone (un cercle en précisant un rayon) et une période de temps, nous récupérerons alors les données de mesure de tous les capteurs situés dans la zone et l'intervalle de temps donnés. Puis pour chaque type de mesure, indépendamment du capteur d'où proviennent les mesures, nous ferons la moyenne de toutes les mesures de ce type. Enfin à partir de ces 4 moyennes générales, nous déterminerons le score ATMO à l'aide du tableau suivant (toutes les mesures sont en  $\mu\text{g}/\text{m}^3$ , la même unité que dans le tableau ci-dessous) :

| Indice ATMO | O <sub>3</sub> | SO <sub>2</sub> | NO <sub>2</sub> | PM10    | Niveau   |
|-------------|----------------|-----------------|-----------------|---------|----------|
| 1           | 0 à 29         | 0 à 39          | 0 à 29          | 0 à 6   | Très bon |
| 2           | 30 à 54        | 40 à 79         | 30 à 54         | 7 à 13  | Très bon |
| 3           | 55 à 79        | 80 à 119        | 55 à 84         | 14 à 20 | Bon      |
| 4           | 80 à 104       | 120 à 159       | 85 à 109        | 21 à 27 | Bon      |
| 5           | 105 à 129      | 160 à 199       | 110 à 134       | 28 à 34 | Moyen    |
| 6           | 130 à 149      | 200 à 249       | 135 à 164       | 35 à 41 | Médiocre |
| 7           | 150 à 179      | 250 à 299       | 165 à 199       | 42 à 49 | Médiocre |

|    |           |           |           |         |              |
|----|-----------|-----------|-----------|---------|--------------|
| 8  | 180 à 209 | 300 à 399 | 200 à 274 | 50 à 64 | Mauvais      |
| 9  | 210 à 239 | 400 à 499 | 275 à 399 | 65 à 79 | Mauvais      |
| 10 | ≥ 240     | ≥ 500     | ≥ 400     | ≥ 80    | Très mauvais |

### Repérage des capteurs ayant un comportement similaire

D'abord, l'utilisateur précisera une période de temps. Nous récupérerons alors tous les capteurs ayant réalisé des mesures pendant cette période de temps et calculerons le score ATMO de chaque capteur. Tous les capteurs qui auront alors le même score ATMO seront considérés comme ayant un comportement similaire.

### Repérage des capteurs ayant un dysfonctionnement

En premier lieu, l'utilisateur précisera une période de temps. Nous examinerons ensuite les mesures effectuées lors de cette période de temps par chaque capteur et un capteur sera déclaré « dysfonctionnant » si :

- il a réalisé des mesures dont la valeur est négative
- il a réalisé des mesures trop grandes par rapport à ce qui est défini par le score ATMO ( $\geq 5000 \mu\text{g}/\text{m}^3$ )
- il a arrêté de réaliser des mesures pendant cette période de temps (on vérifie que la fréquence de prise de mesures est bien respectée)
- il présente des valeurs trop incohérentes (valeurs qui oscillent beaucoup continuellement entre 0 et 200 par exemple)
- ses mesures restent constantes dans le temps sur une période trop longue ( $>12\text{h}$ )