

**Prédiction de la qualité de l’eau en temps réel**

**Membres de l’équipe**

|  |  |  |  |
| --- | --- | --- | --- |
| Prénom | Nom | Courriel | Tel |
| Nicolas | Fortin St-Gelais | [nicolas.fstgelais@gmail.com](mailto:nicolas.fstgelais@gmail.com) | (514) 266 8735 |
| Cindy | Paquette | [cindy\_paquette@hotmail.com](mailto:cindy_paquette@hotmail.com) |  |
| Cynthia | Soued | [soueid.cynthia@courrier.uqam.ca](mailto:soueid.cynthia@courrier.uqam.ca) | (438) 875 3162 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Problématique**

*Contexte*

La ville de Montréal récolte des données de qualité de l’eau régulièrement (à chaque deux semaine) à près de 60 sites riverains. Les résultats sont rendus publics, cependant, ils sont disponibles uniquement quelques jours après l’échantillonnage étant donné le temps d’analyse nécessaire en laboratoire. De ce fait, les citoyens ne peuvent pas savoir en temps réel si la qualité de l’eau est suffisamment bonne pour se baigner ou pratiquer des activités nautiques.

*Question*

Peut-on prédire la qualité de l’eau en utilisant les variables météorologiques accessibles au public en temps réel ?

*Objectif*

1. Créer un modèle qui permettrait de prédire la qualité de l’eau (concentration des coliformes fécaux dans l’eau) selon les données météorologiques (précipitation et température) accessibles en tout temps.
2. Explorer les tendances temporelles saisonnières et annuelles de la qualité de l’eau sur 15 ans.

*Hypothèses*

1. Étant donné que les pluies abondantes engendrent des surverses dans le système d’égout, la concentration en E. coli devrait être positivement corrélée avec les précipitations du jour même ou des jours précédents l’échantillonnage.
2. La température devrait favoriser la prolifération de coliformes.
3. La qualité de l’eau pourrait être moins bonne durant les saisons les plus pluvieuses (printemps et automne).
4. Étant donné l’augmentation probable de la population sur l’île de Montréal, on s’attend à ce que la qualité de l’eau a diminué depuis les 10 dernières années.

**Méthode**

*Données utilisées*

* Nous avons utilisé les données QUALO de la ville de Montréal ainsi que les données météorologiques de l’aéroport de Montréal.
  + Température maximum
  + Précipitation dans les jours précédant la date d’échantillonnage
  + Le mois de l’année
  + L’année d’échantillonnage
* Nous avons sélectionné 9 sites ou la baignade ou autres activités nautiques étaient possibles autour de l’ile, et pour lesquels les données étaient prises de façon consistantes au cours des années.

*Analyses effectuées*

**Analyses temporelles**

* **Tendances saisonnières** : pour chaque site sélectionné, nous avons regroupé les données de concentration de *E. coli* par mois de l’année (de mai à septembre) et nous avons représenté les résultats sous forme de box plot montrant la médiane et les quartiles.
* **Évolution au cours des années** : sur un même graphique, nous avons représenté la concentration en *E. coli* de tous les points d’échantillonnage de 2004 à 2014 (chaque station étant représentée par un couleur différente). Nous avons exploré l’évolution à long termes de la station BLAP-01, pour cela nous avons regroupé les données par année pour cette station et les avons représentés sous forme de box plot.

**Modèle prédictif**

1. Nous avons utilisé un arbre de régression pour identifier la variable la plus importante pour expliquer la concentration en coliformes selon trois catégories.
   1. < 200
   2. Entre 200 et 1000
   3. >1000

Résultats et interprétation

**Tendances temporelles**

Les résultats montrent qu’il n’y a pas de tendance claire au niveau saisonnier, la qualité de l’eau donc pas nécessairement différente d’un mois à l’autre de l’année. Les graphiques montrent par contre qu’en moyenne, la qualité de l’eau est très différente d’une station à l’autre, ce qui suggère qu’il y a une influence de certains facteurs à une échelle très locale sur la qualité de l’eau d’un site. Les graphiques box plot montrent que certains sites sont systématiquement de bonne qualité (<200), d’autres sont très variables, et certains sont très souvent de mauvaise qualité (>200).

Nous avons indiqué sur une carte de Montréal les sites par code de couleur représentant une idée générale de leur qualité en moyenne au cours des 10 dernières années :

* Vert : quasiment toujours de bonne qualité
* Jaune : généralement de bonne qualité
* Orange : très variable et dépassement souvent le seuil des 200
* Rouge : très fréquemment de mauvaise qualité

L’évolution à long terme de la qualité de l’eau montre une différence claire des résultats avant versus après l’année 2006, avec une diminution des coliformes à partir de 2007. La cause de cette tendance est inconnue.

**Modèle prédictif**

Tableau1 : Variable la plus importante pour expliquer la concentration de coliformes et le r-carré associé à cette variable, pour chacune des 10 stations d’échantillonnage sélectionnées

|  |  |  |  |
| --- | --- | --- | --- |
| station | Variable 1 | Variable 2 | rsq |
| BLAP-01 | precip-1 |  | 0.19 |
| BLAP-2.5 | NA |  | NA |
| BLAP-7 | 2007.5 |  | 0.07 |
| FSL-170 | NA |  | NA |
| FSL-400 | 2006.5 | temp | 0.06 |
| IBIZ-11.5 | precip-1 |  | 0.12 |
| LSL-17 | 2006.5 | Temp | 0.12 |
| RDP-60 | precip-1 |  | 0.11 |
| RDP-250 | precip-1 |  | 0.15 |
| RDP-360 | precip-1 |  | 0.07 |

L’année de l’échantillonnage et les précipitations le jour précédant l’échantillonnage sont les deux variables sélectionnées le plus fréquemment pour expliquer les concentrations de coliformes. Pour les années où l’année est la variable principale, il faudrait retirer les années antérieures à l’année sortie dans le modèle. Ceci nous permettrait de voir par la suite quelle variable météorologique a le plus d’influence.

*Utilité et limites des résultats*

Les résultats de notre travail permettent d’avoir une idée plus claire des tendances spatiales et temporelles de la qualité de l’eau sur les rives de l’île de Montréal.

Nous avons développé un modèle prédictif embryonnaire de la qualité de l’eau basé sur les précipitations de la journée précédente, mais qui n’est pas encore assez fiable pour être utilisé comme outil fiable. Il faudrait avoir d’avantage d’information sur le système d’égout et les surverses pour pouvoir améliorer le modèle.

*Travail à venir et améliorations possibles*

* Modèle prédictif pour toutes les stations
* Inclure le vent
* Retirer les données avant 2006 pour le modèle prédictif