

Stage (M1)

Transcription de scripts de traitements statistiques de SAS sous R

Caracteristiques

Domaine professionnel : Informatique, statistiques et sciences de l'environnement

Durée du stage : 4 mois, à partir d'avril ou mai 2018

Rémunération : stage gratifié selon la réglementation en vigueur

Structure d'accueil

Service de la donnée et des études statistiques / sous-direction de l'information environnementale (*)

5, route d'Olivet, 45100 Orléans

Contact service d'accueil :

02 38 79 78 51

Irénée JOASSARD – adjoint au sous directeur de l'information environnementale

irenee.joassard@developpement-durable.gouv.fr

Tuteurs :

pascal.irz@developpement-durable.gouv.fr

frederique.janvier@developpement-durable.gouv.fr

didier.eumont@developpement-durable.gouv.fr

lubomira.guzmova@developpement-durable.gouv.fr

Contexte et enjeux

Problématique

L'eau est devenue, avec l'accroissement de l'activité anthropique, le récepteur et le vecteur de nombreuses substances polluantes d'origine industrielle, agricole ou domestique.

Chaque année, les réseaux de surveillance officiels produisent des millions de données sur la qualité physico-chimique des eaux superficielles (cours d'eau, lacs) et des eaux souterraines (nappes) afin d'évaluer l'efficacité des politiques publiques sur l'eau et d'orienter les plans de gestion agro-écologique de ces masses d'eau. Ces données sont complexes : données distribuées dans l'espace, données variables dans le temps, données qualitatives (données nominales, intervalles de données numériques, données géolocalisées) ou quantitatives.

Le service de la donnée et des études statistiques du ministère en charge de l'environnement exploite ces multiples jeux de données provenant de plusieurs acteurs territoriaux, l'objectif étant de mettre à disposition du public et des décideurs des éléments d'information synthétiques sur l'état de contamination de ces masses d'eau par les

polluants (indices d'évolution...).

La réalisation de ces analyses impose au préalable un "nettoyage" des bases de données brutes (suppression des doublons, traitement des valeurs manquantes et abérantes...), la concaténation des données territoriales portant sur la même thématique.

Pour satisfaire cette étape, le service a développé de nombreux programmes de pré-traitement de données sous SAS. Au fil des années, ces programmes ont été révisés.

Dans le cadre de sa démarche qualité, le service a pour ambition de consolider ses outils de traitement, des les optimiser et des les transcrire sous R.

Objectifs du stage :

Le travail consistera à traduire sous R les programmes SAS suivants :

1. prétraitement des données de qualité des eaux superficielles (macro-polluants, pesticides et autres micropolluants) ;
2. prétraitement des données de qualité des eaux souterraines (nitrates, pesticides et autres micropolluants) ;
3. prétraitement des données de qualité des sédiments (pesticides et autres micropolluants).

Un outil de cartographie dynamique permettant de visualiser les données sera également développé (paramètre année, nombre de polluants détectés, concentrations...).

L'ensemble des programmes sous R devront être documentés.

Données et méthodes

Les données mobilisées sont issues des réseaux de surveillance de la qualité des eaux (eaux souterraines, superficielles, sédiments) gérés par les agences de l'eau.

Les programmes devront respectés les prescriptions techniques définies dans les fiches de traitement précisées par les chargés de mission "qualité de l'eau".

Formations et compétences attendues

Les travaux devront être développés sous R. Le stagiaire devra connaître cet environnement de programmation ainsi que les techniques usuelles de modélisation statistique, et se montrer intéressé par les questions environnementales.

Un connaissance de SAS est également nécessaire.

() Le service de la donnée et des études statistiques fait partie du Commissariat général au développement durable au sein du ministère de la Transition écologique et solidaire. Il a notamment pour mission d'organiser le système d'observation et statistique en matière d'environnement et de développement durable, en liaison avec les institutions nationales, européennes et internationales intéressées.*

Au sein de ce service, la sous-direction de l'information environnementale apporte la connaissance scientifique des mécanismes en jeu dans les domaines environnementaux en définissant les données pertinentes et utilisables à collecter, en proposant les traitements à réaliser sur ces données et en analysant les résultats ainsi obtenus afin de produire des diagnostics.