

Quelques exemples d'améliorations des performances de traitement des données environnementales

François Pinet*

*Université Clermont Auvergne,
INRAE, UR TSCF,
Centre Clermont-Auvergne-Rhône-Alpes,
63178 Aubière, France
francois.pinet@inrae.fr,
www.inrae.fr/tscf

Résumé. Cette communication porte sur un résumé de travaux existants autour du traitement de données agricoles et environnementales. Il s'agit ici de montrer quelques exemples d'amélioration de performances que l'on peut obtenir dans les traitements de données géoréférencées. Nous nous concentrerons sur des opérations que l'on retrouve dans les entrepôts de données.

1 Des techniques pour améliorer les performances

Dans cette communication, nous rappellerons quelques exemples d'optimisation pour les données environnementales de type raster. L'objectif est :

- de faire une synthèse des travaux de trois de nos articles déjà publiés (Kang et al., 2015; En-Nejjary et al., 2021; En-Nejjary et al., 2021) – il s'agit d'une continuité de travaux ;
- d'étendre la présentation que nous avons faite à l'atelier IoT d'Infosid 2022, en insistant sur les applications dans le domaine de l'environnement et les entrepôts de données, en partageant ainsi nos résultats sur ce thème au sein de la communauté EDA.

Les rasters sont un mode de représentation courant dans le domaine des systèmes d'information géographique. Ces données sont présentées sous la forme de grilles régulières géoréférencées où chaque cellule est associée à une valeur. Le plus souvent, nous avons à traiter et analyser une séquence temporelle (c'est-à-dire une succession) de rasters portant sur un même territoire d'étude.

Le but de la présentation sera donc, dans un premier temps, de montrer quelques techniques existantes pour améliorer les temps de traitement de ce type de données. Au cours de la présentation, nous nous concentrerons sur les opérations de sélection et d'agrégation de rasters. Il s'agit d'opérations classiques dans le cadre des entrepôts de données par exemple, mais nécessitant des adaptations sachant le format de données ciblés.

Dans une large collection de rasters portant sur un même territoire, l'opération de sélection permet de choisir automatiquement un sous-ensemble de rasters respectant une condition donnée par l'utilisateur. Ceci permet par exemple de rechercher les rasters représentant un phénomène extrême (par exemple, un pic local dans les valeurs). L'opération d'agrégation permet de résumer un sous-ensemble de rasters, afin d'avoir une vue synthétique de ce sous-ensemble.