

Exploration et analyses multi-objectifs de séries temporelles de données météorologiques

Yelen PER*, Kevin DALLEAU**, Malika SMAIL-TABBONE***

*LORIA UMR 7503, CNRS, yelen.per@loria.fr

**LORIA UMR 7503, CNRS, kevin.dalleau@loria.fr

***LORIA UMR 7503, Université de Lorraine, malika.smail@loria.fr

Résumé. Cet article présente les investigations menées sur les données mesurées par des capteurs positionnés dans cinq villes de l'île de la Réunion. Des analyses exploratoires préalables permettent de comparer les caractéristiques statistiques des villes considérées relativement aux différentes variables météorologiques mesurées (flux solaires diffus et global, pression atmosphérique, humidité, température, force et direction du vent). Nous appliquons diverses transformations sur les données avant d'analyser les séries univariées ou multivariées agrégées au pas de l'heure ou de la journée afin de construire des modèles de prédiction. Une approche classique de clustering de séries temporelles est testée. Deux algorithmes de biclustering appliqués successivement ont permis de grouper les journées d'observations partageant des paramètres météorologiques horaires. Une caractérisation des biclusters, une visualisation calendaire de leur succession ainsi qu'une recherche de séquences fréquentes permettent d'exploiter les résultats et de faciliter leur interprétation.

1 Introduction

Dans le cadre du défi EGC 2018, deux années de mesures - au pas de la minute - de variables météorologiques dans cinq villes de l'île de la Réunion ont été mises à disposition en vue de leur analyse. Le projet à l'origine de ces données s'inscrit dans le cadre de la politique de développement vers l'autonomie énergétique de l'île. Nous avons choisi d'une part de construire, sur la base des séries de données multivariées, des modèles de prédiction ou *prévision* de l'indice de fraction directe et d'autre part, d'appliquer des méthodes de classification non-supervisée (clustering) sur des données univariées puis multivariées décrivant les journées. Un outil de visualisation permet de voir comment des groupes de journées à profil météorologique similaire se succèdent sur un calendrier.

Les éléments de l'analyse exploratoire des données sont présentés avec les principales transformations des données brutes dans la section 2. Les méthodes de régression utilisées et quelques résultats obtenus sont présentés dans la section 3. Les sections 4 et 5 exposent les deux expériences de clustering ainsi que l'outil de visualisation réalisé pour l'occasion.