

# Prédiction du Rayonnement Solaire par Apprentissage Automatique

Pierrick Bruneau, Philippe Pinheiro, Yoann Didry

LIST, L-4362 Esch-sur-Alzette  
prenom.nom@list.lu,  
<http://www.list.lu>

**Résumé.** Cet article décrit une approche flexible pour la prédiction à court terme de variables météorologiques. En particulier, nous nous intéressons à la prédiction du rayonnement solaire à une heure. Cette tâche est d'une grande importance pratique dans l'optique d'optimiser les ressources énergétiques solaires. Comme le défi EGC 2018 nous fournit des données météorologiques enregistrées sur cinq sites géographiques de l'île de la Réunion, nous utilisons ces données historiques comme base pour créer des modèles de prédiction, et nous testons la performance de ces modèles selon le site considéré. Après avoir décrit notre méthode de nettoyage de données et de normalisation, nous combinons une méthode de sélection de variables basée sur les modèles ARIMA (*AutoRegressive Integrated Moving Average*) à l'utilisation de méthodes de régression génériques, telles que les arbres de régression et les réseaux de neurones.

## 1 Introduction

Dans cet article, nous traitons de la prédiction des valeurs futures d'une série temporelle d'intérêt. Nous considérons un horizon de prédiction arbitraire, et un contexte multivarié, où les valeurs historiques de plusieurs séries temporelles sont disponibles en entrée. Vu le contexte particulier du Défi EGC 2018, nous nous intéressons à la prédiction du rayonnement solaire. La prédiction des valeurs futures de variables météorologiques a notamment un intérêt pratique quand il s'agit d'optimiser des sources d'énergie renouvelables (Barbounis et al., 2006).

L'approche classique à la prédiction météorologique utilise des simulations physiques initialisées par des relevés sur le terrain (Lynch, 2008). De manière alternative, dans cet article nous adoptons une approche basée sur l'apprentissage automatique, complètement agnostique de la dimension physique. Plus précisément, la tâche de prédiction est vue comme un problème de régression, avec pour variable cible le rayonnement solaire à un horizon de prédiction donné. Le vecteur d'entrée peut potentiellement utiliser l'ensemble des valeurs historiques (i.e. observées jusqu'à l'instant présent). Ce choix engendre les sous-problèmes suivants :

- *pré-traitement* : les valeurs manquantes affectent généralement les modèles d'apprentissage. L'inspection et la correction préalable des données d'apprentissage est nécessaire.