**Pôle DATA – Test technique**

Décorrélation des effets environnementaux sur les mesures d’un capteur

**Contexte :**

Dans le cadre de ses activités de monitoring, Cementys instrumente les ouvrages d’art (ponts, plateformes pétrolières, barrages) pour suivre leur vieillissement ainsi que des structures en construction (tunnels, gares). De nombreux capteurs sont ainsi installés et des données sont collectées en continu avec pour objectif de qualifier l’état de la structure. La base de données de l’entreprise intègre donc à la fois des données de capteurs (déplacement, contrainte, déformation, angle, …) mais également des données environnementales (température, humidité, ensoleillement, …) issues notamment des stations météorologiques.

Or, l’environnement a un impact important et réversible sur le comportement des structures. Par conséquent, on cherche à supprimer les effets de l’environnement en décorrélant les variables environnementales des mesures des capteurs. Les mesures décorrélées donnent une information directe sur le vieillissement de l’ouvrage.

**Problématique :**

Le fichier *dataset.dat* contient les mesures d’un capteur de déplacement en fonction du temps ainsi que les mesures de température et d’ensoleillement associées. L’objectif de l’exercice est de fournir les valeurs de déplacement décorrélées (c’est-à-dire ne dépendant plus de la température et de l’ensoleillement) aux mêmes pas de temps.

**Remarques :**

Le sujet est un problème ouvert. **Il n’est pas demandé au candidat une solution complète**. Le candidat pourra se pencher sur les questions suivantes :

* Lecture, nettoyage et affichage des données.
* Étude qualitative et quantitative des corrélations.
* Choix et implémentation d’un algorithme de décorrélation.
* Affichage et enregistrement des résultats.
* Automatisation et généralisation de la solution.

Les pistes suivantes sont proposées pour l’algorithme de décorrélation : modèle linéaire simple, modèle HST, réseau de neurones …

La solution prendra la forme d’un code Python commenté éventuellement accompagné de graphes et de datasets. Il est recommandé au candidat de ne pas passer plus d’une demi-journée sur le problème.