

ALGORITHMIQUE ET PROGRAMMATION

PROJET : ANALYSE DES SÉRIES TEMPORELLES : APPLICATION AU PROCÉDÉ DE POLISSAGE CHIMICO-MÉCANIQUE

1 Contexte et problème

Ce projet porte sur l'analyse d'un jeu de données qui a été fourni dans le cadre d'une compétition organisée en 2016 par la société PHM¹, et qui portait sur la prédiction du taux moyen d'élimination des matériaux de surface de plaques semi-conductrices (wafer) par réaction chimique. L'ensemble des données a été collecté durant le polissage chimico-mécanique (CMP) des surfaces, qui est un procédé de fabrication standard dans l'industrie des semi-conducteurs. Pour prédire le taux moyen d'élimination des matériaux de surface d'un wafer donné, 19 séries temporelles uni-variées associées à l'évolution des indicateurs de santé propre à un système CMP sont accessibles via lien suivant : <https://seafiler.emse.fr/d/d4ab64d949264f8c83a4/>

La description complète du jeu de données est disponible via le lien ci-dessous : <https://phmsociety.org/conference/annual-conference-of-the-phm-society/annual-conference-of-phm-data-challenge-4/>

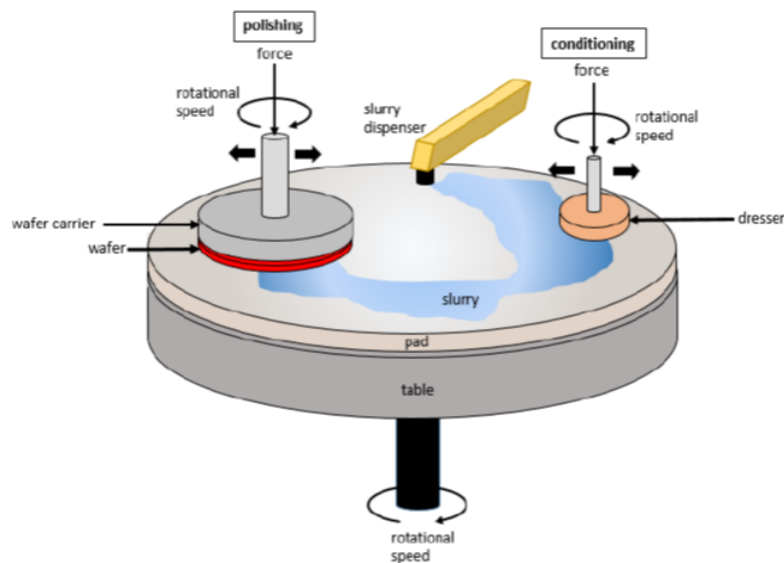


FIGURE 1 – Polissage chimico-mécanique. (Source : <https://phmsociety.org/wp-content/uploads/2016/05/PHM16DataChallengeCFP.pdf>)

L'objectif de ce projet consiste à développer une application en langage C, qui permet de pré-traiter les séries temporelles disponibles via le lien seafiler ci-dessus afin de prédire et de mieux expliquer l'évolution de la caractéristique cible (i.e., le taux moyen d'élimination du matériel de surface d'un wafer donné) en utilisant les données historiques. Il s'agit de développer les fonctionnalités suivantes :

1. <https://phmsociety.org/>

1. **Statistiques descriptives centrales et de dispersion** : Pour chaque série temporelle, calculer les statistiques descriptives suivantes : moyenne, min, max, écart-type, médiane, ensemble des quartiles. Exporter ces statistiques dans un fichier texte.
2. **Détection des observations atypiques** : Pour chaque série temporelle, détecter et supprimer les observations atypiques en appliquant la règle de $k\sigma$ (où σ représente l'écart-type), qui consiste à supprimer les observations dont la valeur est supérieure (resp. inférieure) à $k\sigma$ (resp. $-k\sigma$), $k = \{2, 3\}$.
3. **Distributions empiriques** : Pour chaque série temporelle, construire l'histogramme associé, i.e., **(i)** diviser son domaine de définition $[min, max]$ en n intervalles égaux (bins), **(ii)** agréger les observations par bin, et **(iii)** calculer la fréquence des observations dans chaque bin. Fixer un nombre de bins par défaut.
4. **Extraction des caractéristiques agrégées par wafer** : Agréger les séries temporelles par wafer. Utiliser comme critère d'agrégation la moyenne. Sauvegarder les données agrégées dans un fichier texte.
5. **Corrélation** : (Optionnel) Calculer la corrélation entre les caractéristiques agrégées extraites et le taux moyen d'élimination du matériel de surface.
6. **Prédiction** : (Optionnel) En utilisant une ou plusieurs caractéristiques agrégées extraites, prédire le taux moyen d'élimination du matériel de surface d'un wafer donné à l'aide d'une régression linéaire.

2 Livrables

Le travail se fait par binôme. Les codes sources (sous forme d'un fichier zip) et un document explicatif (optionnel) sont à déposer sur la plateforme CAMPUS au plus tard le **5 décembre 2022 à 12h00** (pour tous les groupes).

Le document explicatif doit contenir :

- ☐ le découpage en modules
- ☐ les choix techniques et des structures de données
- ☐ comment exécuter le programme et les résultats qu'il doit produire