Journal of Hazardous Materials 454 (2023) 131482

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

Research Article

Statistical analysis, machine learning modeling, and text analytics of aggregation attachment efficiency: Mono and binary particle systems

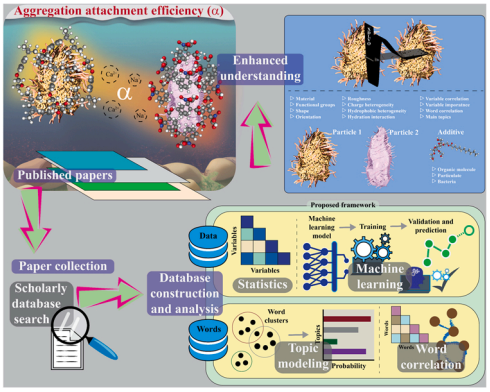
Allan Gomez-Flores [a], Scott A. Bradford [b], Gilsang Hong [a], Hyunjung Kim [a,*]

[a] Department of Earth Resources and Environmental Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea
[b] USDA, ARS, Sustainable Agricultural Water Systems Unit, 239 Hopkins Road, Davis, CA 95616, USA

HIGHLIGHTS

- Databases on aggregation in mono and binary particle systems were analyzed.
- Statistics, machine learning, and text analytics were used for analyses.
- Variable correlations and importance were identified.
- Random forest and neural network were used after missing data imputation.
- Topic modeling and word correlations were performed on the literature.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Dr. R Teresa

Keywords:
Missing data imputation
Machine learning
Text analytics
Topic modeling
Word correlation

ABSTRACT

The aggregation attachment efficiency (α) is the fraction of particle–particle collisions resulting in aggregation. Despite significant research, α predictions have not accounted for the full complexity of systems due to constraints imposed by particle types, dispersed matter, water chemistry, quantification methods, and modeling. Experimental α values are often case–specific, and simplified systems are used to rule out complexity. To address these challenges, statistical analysis was performed on α databases to identify gaps in current knowledge, and machine learning (ML) was used to predict α under various particle types and conditions. Moreover, text analytics was employed to support knowledge from statistics and ML, as well as gain insight into the ideas communicated by current literature. Most studies investigated α in mono–particle systems, but binary or higher systems require more investigation. Furthermore, our work highlights that numerous variables, interactions, and mechanisms influence α behavior, making its investigation complex and difficult for both experiments and modeling. Consequently, future research should incorporate more particle types, shapes, coatings, and surface heterogeneities, and aim to address overlooked variables and conditions. Therefore, building a comprehensive α database can enable the development of more accurate empirical models for prediction.

* Corresponding author.
E-mail address: kshjkim@hanyang.ac.kr (H. Kim).

https://doi.org/10.1016/j.jhazmat.2023.131482
Received 16 March 2023; Received in revised form 11 April 2023; Accepted 22 April 2023
Available online 24 April 2023