**Title**

Gaussian process emulator for upscaling complex multi-scale stochastic biological models

**Author Name\*, Oluwole, K. Oyebamiji**

\* School of Mathematics & Statistics, Newcastle University, NE1 7RU, UK., wolemi2@yahoo.com

\*\* second author’s address, email address

**Keywords:** Gaussian process, biofilms; floc, kriging

**Introduction**

One of the crucial aspects of engineering biology approach in wastewater treatment study is to run a high complex simulation of biological floc or biofilm models. The models have the ability to scale from one level to another, to better understand how to effectively manage real systems with minimal physical experimentation. It is generally considered that to identify crucial features and model water treatment plant on a large scale, there is a need to understand the interactions of microbes at fine resolution based models that could provide the best available representation of micro scale responses. The challenge then becomes how we can transfer this small-scale information to the macroscale process via a mesoscale in a computationally efficient and sufficiently accurate way, and to also probably quantified the associated risk or error in the process.

In addition, simulation of open biological systems is difficult because it involves a large number of bacteria that ranges from 10^12 to 10^18 individual particles and are physically and genetically complex systems. The models are computationally expensive and due to computing constraints, limited set of scenarios are often possible. This problem can be eradicated by using a statistical approximation of the complex models which will help in reducing the computational burden. Our aim in this work is to build a cheaper surrogate models (called metamodels) from simulations of the LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator), a classical molecular dynamical model for biological particle simulation.

* The maximal length of your abstract is two A4 pages of text including figures, tables, and references.

The Introduction section should clearly state the background, relevance and the objectives of your work, without extensive literature review.

**Material and Methods**

Our approach is to condense the massive long time series outputs of particle of various species from LAMMPS models by spatially aggregating to produce the most relevant outputs in the form of floc aggregates or biofilm. The data compression has the benefit of suppressing or reducing some of the nonlinear response features, simplifying the construction of the metamodels. Some of highly interested properties at the mesoscale level like the size, shape and structure of biofilm and floc are characterized. For instance, we approximate the floc size using an equivalent diameter. This strategy enable us to treat the floc as a ball of a sphere. We then emulate the diameter of a sphere that circumscribes its boundary or outline. The center of the sphere will be equivalent to the center of mass of the component particles. See Figure 1.

We use the Gaussian process emulation in the form of kriging metamodels where output data can be decomposed into a mixture of deterministic (non-random trend) and a residual random variation. Our approach combines the two stage technique proposed in \citet{l5,l7} as a single step and also similar to \citet{l6} who combine GP emulation with a basis representation for calibration of computer models with high dimensional outputs.

is to highlight what we have done so far, our broad plan and strategy for the upscaling high-level summary from the simulation

**Results and Conclusions**

* This section should summarise clearly the main experimental results or the major outcomes of the study, supported by graphics or tables.

The review of your submission will be based on the material provided and not on expected future results.

Figures and tables should appear in numerical order, be described in the body of the text and be positioned close to where they are first cited. Make sure all figures and tables fit inside the text area.

Please ensure that all figures are clearly labelled and easily readable in black‑and‑white printing.

**Table 1.1** This is a style for Table Titles. “Table 1.1, 1.2 etc” should be in bold. Table captions should appear above tables.

Please include your table here.

Please include your figure here.

**Figure 1.1** This is a style for Figure legends. “Figure 1.1, 1.2 etc” should be in bold. Figure legends should appear **below** figures.

**References**

* For citation in the main text, use surname of author and year of publication: Jones (2002) or (Jones, 2002). Insert initials only if there are two different authors with the same surname and same year of publication. The abbreviation "et al." should be used in the text when there are more than two co-authors.

References should be listed alphabetically at the end of the abstract. Although "et al." is preferable in the text, in the list of references all authors should be listed.

Andrews, J. F. (1993), Modeling and simulation of wastewater treatment processes. *Wat. Sci. Tech.,* **28**(11/12), 141–150.

Billing, A. E. (1987), Modelling techniques for biological systems. M.Sc. thesis, Dept Chem. Eng., Univ. of Cape Town, Rondebosch 7700, South Africa.