**Emulator as a vital tool in a multiscale modelling of wastewater treatment plants**

**Introduction**

We shall describe briefly in this report how an emulation approach can be integrated as a part of a multiscale modelling strategy of wastewater treatment plants. Water is crucial for life on earth and is valuable also for its supporting role in ecosystem functions. Water that is safe for drinking is scarce partly due to an increase in wastewater. Biofilm technology is being deployed in the management and treatment of wastewater. Wastewater treatment plants are open systems that depend on many species of bacteria to form a microbial community for the transformation of waste into biomass and other substances. A model is required that describes the individual processes in the wastewater treatment system.

**Objectives**

We describe how to use an emulator as an effective tool for studying and incorporating microscale processes in a computationally efficient way into macroscale models. The emulator produced will be incorporated into the

NUFEB model routine at mesoscale to produce a more refined NUFEB models that are much computationally efficient for providing information on a large scale.

To identify crucial features and model water treatment plants (WWTP) on a large scale, there is a need to understand the interactions of microbes at fine resolution using models that provide the best possible representation of micro-scale responses.

The challenge then becomes how we can transfer this small-scale information to the engineered macroscale process in a computationally efficient and sufficiently accurate way. A multiscale modelling strategy is required to relate the microscopic microbial actions to the macroscopic bulk WWTP operation thus making the simulation from the microscale model a computationally expensive task. One useful approach for addressing this problem is using statistical models often called “emulators”. We are using statistical emulators for modelling biofilms and flocs morphological behaviours and passing aggregate information from one level to the other.

Our approach is to efficiently transfer emerging properties to macroscale from the microbial interactions occurring in microscale. However, the wide separation between the micro and macro scales is massive, and therefore we need a mesoscale to transfer properties between scales more efficiently. We transfer microscale emerging properties to mesoscale and then perform individual-based (IB) simulations again at a mesoscale level to calculate new emerging properties and then these properties are transferred to macroscale. This technique is especially convenient in the case of activated sludge reactors, since the floc is the natural elemental unit for mesoscale modelling.

The knowledge of open biological simulation is crucial in the design and management of wastewater treatment systems. The key aspect of an engineering biology approach to wastewater treatment study is the ability to simulate complex biological communities. However, the simulation of open biological systems is challenging because they often involve many bacteria cells. We know that biofilm growth simulation is computationally demanding. Statistical emulators can be applied to overcome this problem.

The fundamental issue here is the development of statistical algorithms for the emulation and calibration of complex multiscale biological models (eg **NUFEB 1.1**) and to use the emulator produced as a tool for upscaling NUFEBmodel outputs. This involves transferring of information from one spatial or temporal scale to another one. The purpose of building an emulator is to facilitate other calculations that would be too computationally demanding using the simulator.



Figure 1: Schematic diagram of different length scales for multiscale modelling of an

activated sludge process based WWTP. The scale transition from a bacterium or cellular levels (microscale - < micrometer size) to the floc and biofilm aggregates (mesoscale - millimeter size) to the macroscopic bulk WWTP operation as well as floc and biofilm interactions (macroscale - metre size). The emulator is linking the microscopic (bacterium/cell) to the mesoscopic (biofilm/floc) to the macroscopic bulk operational parameters.

**Statistical emulator**

Emulation is a statistical technique for simplifying models that lead to reduced-form representations of complex models which are computationally much faster to run. In other words, is an attempt to imitate the internal design of a computer model or simulator statistically. The emulator is based on the multivariate Gaussian process regression. A Gaussian process is a distribution over functions, and it is fully specified by a mean function ***m*** and a covariance function ***k*** such that ***f ∼ GP (m,k)***.

The emulator can enhance the understanding of the common frameworks for theoretical modelling of multiscale problems.

The emulator is designed to pass microscale emerging properties to macroscale. The floc size, composition, and environmental conditions (eg nutrients, pH, flow field, temperature etc.) can be considered as the inputs for the emulator, and then the emulators can be constructed that predict the growth rate of each species in the floc or biofilm, fractal dimension of the floc etc. as outputs. The simulator (eg **NUFEB 1.1**) can be represented by a linear function ***fs*** that gives a vector of outputs ***Y*** with an input matrix ***X*** such that ***Y = fs(X)***. We wish to design an emulator that can predict the values of ***Y*** as accurately as possible:

***Ỹ= fₑ(X),***

where ***fₑ(·)*** is the emulator. To achieve this, we can choose a function ***fₑ= H(.)B***

where ***B=[β1… βp]*** is a vector of unknown regression coefficients ***and H(x)= [h1(x),… hm(x)]*** is a matrix of regression functions. See further details in (Oyebamiji et al. 2017; Conti et al. 2009).

**Overview**

We highlight what we have done so far and plan for the tasks ahead.

* Evaluate the strategy for emulating high-level summary from the

Individual-based simulation of microbial communities using the NUFEB 1.1 model.

* Measure aggregated characteristics on the microscale simulation

data to reduce the dimensionality of the problem and develop novel surrogate based techniques for incorporating microscale processes in a computationally efficient way into engineered macroscale models of the wastewater plant.

* Apply a dynamic Bayesian surrogate model (GP emulator) for modelling the interesting morphological characteristics that are essential for the design and performance of wastewater reactor. We also performed Bayesian sensitivity analysis to identify the most influential input parameters.
* Quantify and model detachment patterns of a biofilm in response to hydrodynamic shear stress.

**Future plans**

We have made considerable progress on the parameter calibration using the recently released **NUFEB 2.0** version of the model which has an improved chemistry such as explicitly modelled chemical reactions like anaerobic/aerobic respiration and digestion, nitrification, multiple-species and PH. We plan to continue with the NUFEB 2.0 parameter calibration with IdynoMICS and experimental data. Secondly, we plan to properly investigate the use of emulators as an upscaling strategy.

**References**

* Oyebamiji OK, Wilkinson DJ, Jayathilake PG, Curtis TP, Rushton SP, Li B, Gupta P. [**Gaussian process emulation of an individual-based model simulation of microbial communities**](http://www.ncl.ac.uk/maths-physics/staff/profile/oluwoleoyebamiji.html#231356). *Computational Science* 2017. In Press.
* Jayathilake PG, Gupta P, Li B, Madsen C, Oyebamiji O, Gonzalez-Cabaleiro R, Rushton S, Bridgens B, Swailes D, Allen B, McGough AS, Zuliani P, Ofiteru ID, Wilkinson DJ, Chen J, Curtis TP. [**A mechanistic individual-based model of microbial communities**](http://www.ncl.ac.uk/maths-physics/staff/profile/oluwoleoyebamiji.html#240439). *PLoS One* 2017, **12**(8), e0181965.
* Oyebamiji O, Wilkinson DJ, Pahala-Gedara J, Rushton SP, Li B, Bridgens B, Zuliani P. [**A surrogate-based approach to modelling the impact of hydrodynamic shear stress on biofilm deformation**](http://www.ncl.ac.uk/maths-physics/staff/profile/oluwoleoyebamiji.html#240440). *To be s*ubmitted.
* Oyebamiji OK, Wilkinson DJ. [**Statistical emulation as a tool for analysing complex multiscale stochastic biological model outputs**](http://www.ncl.ac.uk/maths-physics/staff/profile/oluwoleoyebamiji.html#231357). In: VIII International Conference on Sensitivity Analysis of Model Output. 2016, The University of Reunion Island, Le Tampon.
* Conti, S., Gosling, J.P., Oakley, J.E., & O'Hagan, A. (2009). **Gaussian process emulation of dynamic computer codes**. Biometrika, 96(3); 663-676.