# Leveraging AI to generate fast and accurate surrogate models of bacteria to predict processes in terrestrial systems

Juan D. Tibocha-Bonilla, Rodrigo Santibáñez-Palominos, Maxwell Neal, Karsten Zengler
*University of California, San Diego*

**Topic:** (1) Novel Algorithms, (2) Multiscale/Multimodal Modeling

**Challenge:** Metabolism and gene expression models (ME-models) represent the most advanced genome-scale models to date as they describe and predict the proteome allocation in bacteria. Thus, ME-models provide a powerful tool for understanding microbial fitness and interactions in microbial communities (**Figure 1a**)[1]. However, ME-models are computationally intensive, caused by a ~10-fold increase in the number of reactions compared to standard genome-scale metabolic models[2] (GEMs or M-models) (**Figure 1b**). Thus, their application is currently limited for the study of complex microbial consortia and communities, which usually include hundreds or thousands of microbial members[3]. Solving ME-models for complex systems with multiple interacting species requires computational resources beyond typical capacities, hindering real-time and high-throughput simulations critical to advancing the study of microbe-microbe and plant-microbe interactions in terrestrial systems.

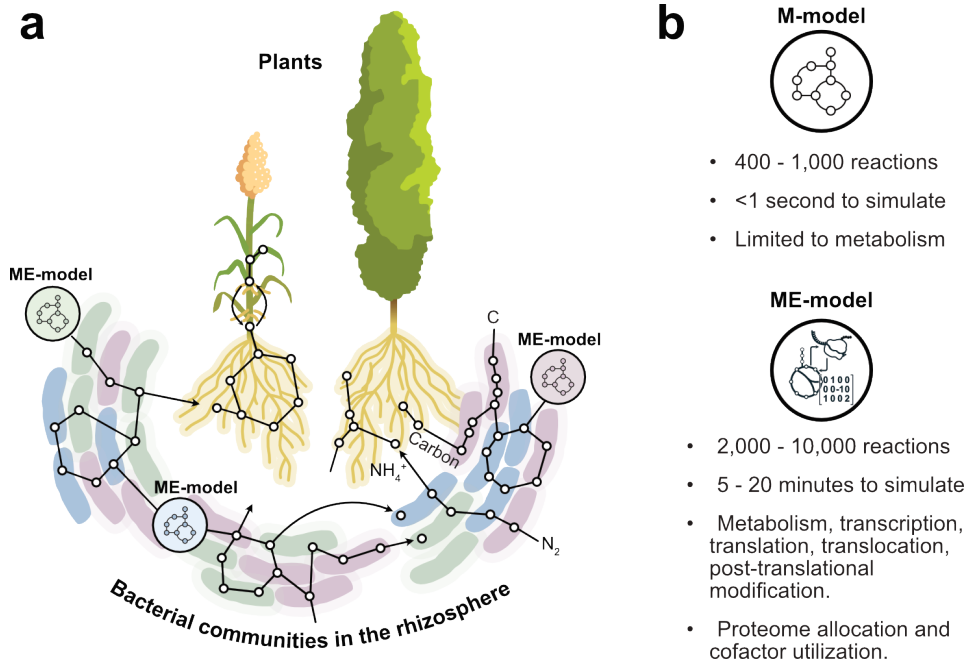


**Figure 1. The complexity of plant-microbe and microbe-microbe interactions in terrestrial systems. a,** Plant-microbe and microbe-microbe interactions in the rhizosphere. **b,** Increased resolution at the cost of higher computational cost in ME-models vs. M-models.

**Opportunity:** The generation of surrogate models using AI offers a solution to this computational resource constraint by generating surrogates that efficiently and accurately reproduce ME-model predictions. The key idea is to use a machine learning (ML) model to interpolate between ME-model simulations rapidly[4]. This will require the use of the ME-models to generate thousands of datasets simulating diverse extracellular conditions, e.g. nutrient replete, nutrient deplete, and supplementation of individual metabolites, among others. The ML model would then receive enough datasets to capture complex metabolic traits, such as resource allocation, overflow metabolism, and proteome limitation. The resulting surrogates, trained on high-resolution predictions by ME-models, can then approximate the metabolic and gene expression mechanisms of complex microbial communities while maintaining predictive accuracy with only a fraction of the computational cost. Thus, this approach has great potential to allow the application of ME-models to understand plant-microbe and microbe-microbe interactions in terrestrial systems, which could revolutionize how we describe nutrient exchange in the rhizosphere.

**Timeliness or Maturity:** The rapid advancement in the application of AI to generate surrogate models makes this proposal especially timely[5–7]. Recent breakthroughs in neural networks and generative modeling to address biological questions mechanistically[4,8], along with the increasing accessibility to high-performance computing, allow for the first time the generation of robust and reliable surrogates for ME-models. Along with integration of multi-omics data, the implementation of this proposed solution would allow scientists to harness the outstanding detail from ME-model predictions across various microbial communities and systems of interest to DOE. There are currently numerous efforts advanced by the DOE that could benefit from this advancement. Some examples are nutrient cycling, microbial contributions to the rhizosphere, the effect of microbial interactions in plant diseases and drought tolerance, and the design of interventions to mitigate them.

**References**


1. Kumar, M., Ji, B., Zengler, K. & Nielsen, J. Modelling approaches for studying the microbiome. *Nat. Microbiol.* **4**, 1253–1267 (2019).
2. Lloyd, C. J. *et al.* COBRAme: A computational framework for genome-scale models of metabolism and gene expression. *PLOS Comput. Biol.* **14**, e1006302 (2018).
3. Daniel, R. The metagenomics of soil. *Nat. Rev. Microbiol.* **3**, 470–478 (2005).
4. Gherman, I. M. *et al.* Bridging the gap between mechanistic biological models and machine learning surrogates. *PLOS Comput. Biol.* **19**, e1010988 (2023).
5. Alizadeh, R., Allen, J. K. & Mistree, F. Managing computational complexity using surrogate models: a critical review. *Res. Eng. Des.* **31**, 275–298 (2020).
6. Elwy, I. & Hagishima, A. The artificial intelligence reformation of sustainable building design approach: A systematic review on building design optimization methods using surrogate models. *Energy Build.* **323**, 114769 (2024).
7. Wang, L. *et al.* A critical review on intelligent optimization algorithms and surrogate models for conventional and unconventional reservoir production optimization. *Fuel* **350**, 128826 (2023).
8. Choudhury, S., Narayanan, B., Moret, M., Hatzimanikatis, V. & Miskovic, L. Generative machine learning produces kinetic models that accurately characterize intracellular metabolic states. *Nat. Catal.* **7**, 1086–1098 (2024).