Integrated Emulators for Systems of Computer Models

D. Ming
University College London

Supervisor(s): Prof. Serge Guillas (University College London)

Ph.D. expected duration: Oct. 2016 - Sep. 2020

Adress: Department of Statistical Science, University College, Gower Street, WC1E 6BT, London

Email: deyu.ming.16@ucl.ac.uk

Abstract:

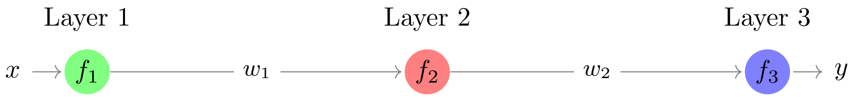
Systems of computer models constitute the new frontier of many scientific and engineering simulations. These can be multi-physics systems of computer simulators such as coupled tsunami simulators with earthquake sources [8], coupled multi-physics model of the human heart [6], and multi-disciplinary systems such as aerospace systems [1]. Other examples include climate models [2], or highly multi-disciplinary future biodiversity models [7]. The number and complexity of computer models involved can hinder the analysis of such systems. For instance, the engineering design optimization of an aerospace system typically requires hundreds of thousands of system evaluations. When the system has feed-backs across computer models, the number of simulations becomes computationally prohibitive. Therefore, building and using a surrogate model is crucial: the system outputs can be predicted at little computational cost, and subsequent sensitivity analysis, uncertainty propagation or inverse modeling can be conducted in a computationally efficient manner.

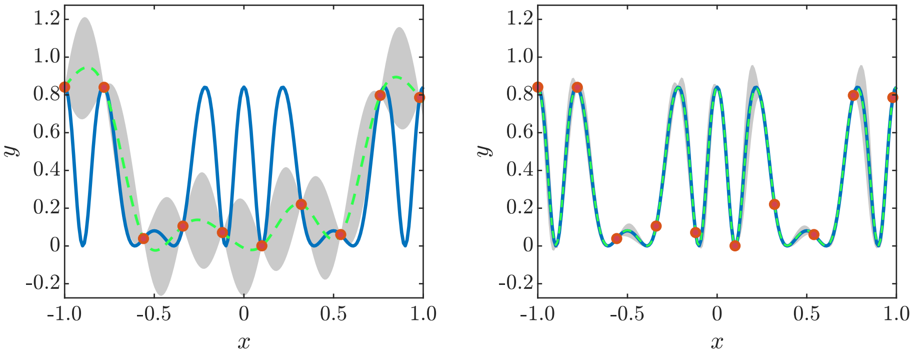
Gaussian process (GP) emulators have gained popularity as surrogate models of systems of computer models in fields including environmental science, biology and geophysics because of their attractive statistical properties. However, many studies construct global GP emulators (named as composite emulators hereinafter) of such systems based on global inputs and outputs without consideration of system structures. One major drawback of such a structural ignorance is that designing experiments can be expensive because system structures may induce high non-linearity between global inputs and outputs [5]. Furthermore, runs of the whole system are required to produce new training points, even though the overall functional complexity global inputs and outputs originates from a few computer models. This pitfall is particularly undesirable because modern engineering and physical systems can include multiple computer models.

To overcome the disadvantages of the composite emulator, two recent studies in [3] and [4] have derived a structure-informed emulator, called linked emulator [3], for a feed-forward system of two computer models under the assumption of squared exponential kernel. Inspired by the linked emulator, we generalize it to an integrated emulator for any feed-forward system of multiple computer models, under a variety of kernels (exponential, squared exponential, and two key Matérn kernels) that are essential in advanced applications. The integrated emulator combines Gaussian process emulators of individual computer models, and predicts the global output of the system using a Gaussian distribution with explicit mean and variance. By learning the system structure, our integrated emulator outperforms the composite emulator, which emulates the entire system using only global inputs and outputs. Orders of magnitude prediction improvement can be achieved for moderate-size designs. A synthetic example is presented in Figure 1 that compares the predictive performance of the integrated and composite emulators under identical design points.

Furthermore, our analytic expressions allow a fast and efficient adaptive design algorithm that allocates different runs to individual computer models based on their heterogeneous functional complexity. This design yields either significant computational gains or orders of magnitude reductions in prediction errors for moderate training sizes. We demonstrate the skills and benefits

of the integrated emulator in a series of synthetic experiments and a feed-back coupled fire-detection satellite model.





(a) Composite Emulator

(b) Integrated Emulator

Figure 1: Composite and integrated emulators of a computer system of three computer models $f_1 = \sin(\pi x)$, $f_2 = \cos(5w_1)$ and $f_3 = \sin(w_2^2)$ connected sequentially. The solid line is the true functional form between the global input and output of the system; the dashed line is the mean prediction; the shaded area represents 95% prediction interval; the filled circles are training points used to construct the emulators.

References

[1] H.R. Fazeley, H. Taei, H. Naseh, and M. Mirshams. A multi-objective, multidisciplinary design optimization methodology for the conceptual design of a spacecraft bi-propellant propulsion system. *Struct. Multidiscipl. Optim.*, 53(1):145–160, 2016.

[2] J. E. Kay et al. The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Am. Meteorol. Soc.*, 96(8):1333–1349, 2015.

[3] K. N. Kyzyurova, J. O. Berger, and R. L. Wolpert. Coupling computer models through linking their statistical emulators. *SIAM/ASA J. Uncertain. Quantif.*, 6(3):1151–1171, 2018.

[4] S. Marque-Pucheu, G. Perrin, and J. Garnier. Efficient sequential experimental design for surrogate modeling of nested codes. *ESAIM Probab. Stat.*, 23:245–270, 2019.

[5] F. Sanson, O. Le Maitre, and P. M. Congedo. Systems of Gaussian process models for directed chains of solvers. *Comput. Methods. Appl. Mech. Eng.*, 352:32–55, 2019.

[6] A. Santiago et al. Fully coupled fluid-electro-mechanical model of the human heart for supercomputers. *Int. J. Numer. Method. Biomed. Eng.*, 34(12):e3140, 2018.

[7] W. Thuiller, M. Guéguen, J. Renaud, D. N. Karger, and N. E. Zimmermann. Uncertainty in ensembles of global biodiversity scenarios. *Nat. Commun.*, 10(1):1446, 2019.

[8] T. Ulrich et al. Coupled, physics-based modeling reveals earthquake displacements are critical to the 2018 Palu, Sulawesi Tsunami. *Pure Appl. Geophys.*, 176(10):4069–4109, 2019.

Short biography – I am a PhD student in statistical science at University College London. My thesis will includes topics on spatial analysis of earthquake ground-motion model estimation and prediction, and efficient statistical surrogates for systems of computer models with their applications in uncertainty quantifications of coupled physics-based earthquake and tsunami supercomputer models. My PhD study is fully funded by the UCL-CSC joint research scholarship.