Emulation numerical experiment

- Methods: GPy, PCGPwM, and PCGPwMatComp
- Functions: borehole, OTL circuit, Wing weight, and piston
- Number of locations: 15
- Number of training parameters, n:
  25, 50, 100, 250, 500, 1000, 1500, 2500. $(15 \times 2500 = 37500)$
- Failures are {random, structured} $\times$ {25%, 75%} levels.
- 1 hour timeout for every run
- x are sampled uniformly in $[0,1]^{d_x}$, fixed across n.
- θ are sampled from latin hypercube sampling in $[0,1]^{d_\theta}$
- 1000 test parameters are sampled uniformly in $[0,1]^{d_\theta}$
- Comparison metrics:
  - RMSE, MAE, median absolute error, mean error (bias)
  - 90% coverage, 90% prediction interval width
  - interval score, CRPS