Estimating Building Simulation Parameters via Bayesian Structure Learning

Richard E. Edwards[1], Joshua R. New[2], and Lynne E. Parker[1]

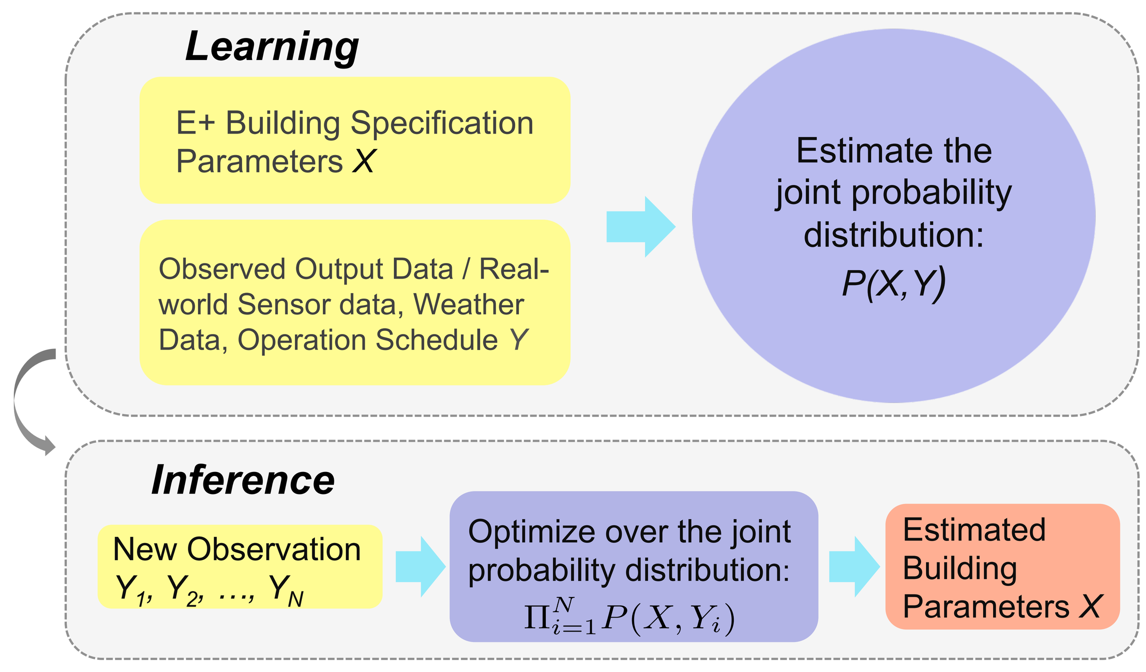
[1]University of Tennessee, Electrical Engineering and Computer Science

[2]Oak Ridge National Lab, Whole Building & Community Integration Group

Abstract

Many key building design policies are made using sophisticated computer simulations such as EnergyPlus (E+), the DOE flagship whole-building energy simulation engine. E+ and other sophisticated computer simulations have several major problems. The two main issues are 1) gaps between the simulation model and the actual structure, and 2) limitations of the modeling engine's capabilities. Currently, these problems are addressed by having an engineer manually calibrate simulation parameters to real world data or using algorithmic optimization methods to adjust the building parameters. However, some simulations engines, like E+, are computationally expensive, which makes repeatedly evaluating the simulation engine costly. This work explores addressing this issue by automatically discovering the simulation's internal input and output dependencies from ~20 Gigabytes of E+ simulation data, future extensions will use ~200 Terabytes of E+ simulation data. The model is validated by inferring building parameters for E+ simulations with ground truth building parameters. Our results indicate that the model accurately represents parameter means with some deviation from the means, but does not support inferring parameter values that exist on the distribution's tail.

Problem Formulation



Structure Learning

Gaussian Graphical Model (GGM)

$$\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}(x-\mu)^T \Omega (x-\mu))$$

where Σ is the covariance matrix, *u* is the mean, and Ω is the inverse covariance matrix or precision matrix

Dobra et. al (2004) and Li et. al (2006)

$$\Omega = (1-\Gamma)^T \Psi^{-1} (1-\Gamma)$$

where Γ represents an upper triangular matrix with zero diagonals and the non-zero elements represent the regression coefficients, and Ψ^{-1} represents a diagonal matrix containing the variances for each performed regression.

Pros: Scalable for a large number of random variables

Problem for the E+ Data Set

E+ Building Parameters *X*

Sample ID	P_1	P_2	…	P_{156}
1	Min(V_{P1})	avg		avg
2	Max(V_{P1})	avg		avg
…	…			
299	avg	avg		Max(V_{P156})

generate

One Observed Output Y_1

Row ID	V_1	V_2	…	V_{90}
1				
2				
…				
35,040				

How to sample X to cover the whole space of building parameters for all possible real values?

Bayesian Regression GGM Learning

Hyper-parameter θ_i for *T*

- $P(\theta_i) = \frac{\gamma}{2} exp(\frac{-\gamma\theta_i}{2})$

Wishart prior for *Ω*

- $W(\delta, T)$

Solved by *LASSO* with *ADMM*

Maximum a posterior (MAP) distribution for all *β* and all ψ^{-1}

- $P(\beta_i|\psi_i, D) \propto exp(\frac{\sum_{n=1}^{D}(x_{ni} - \sum_{j=i+1}^{N}\beta_{ij}x_{nj})^2 + \sqrt{\gamma\Psi_i}\sum_{j=i+1}^{N}|\beta_{ij}|}{-\psi_i})$

Maximize $P(\psi_i^{-1}|\theta_i, \beta_i, D)$ with respect to ψ^{-1}

- $P(\psi_i^{-1}|\theta_i, \beta_i, D) \propto P(D|\Psi_i^{-1}, \beta_i, \theta_i)P(\beta_i|\Psi_i^{-1}, \theta_i)P(\psi_i^{-1}|\theta_i)$

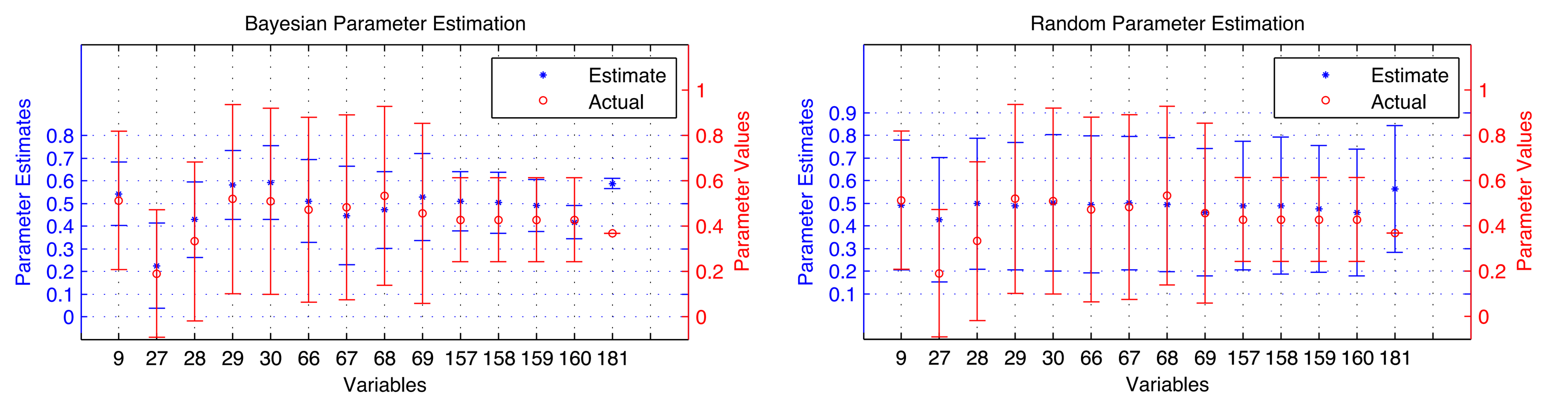
$$\sim Gamma(\frac{\delta + 1 + N - 2i + |D|}{2}, \frac{\sum_{j=i+1}^{N}\beta_{ij}^2\theta_i^{-1} + \theta_i^{-1} + \sum_{n=1}^{D}(x_{ni} - \sum_{j=i+1}^{N}\beta_j x_{nj})^2}{2})$$

1. Given a fixed θ^{-1}, we estimated a ψ^{-1} sample by computing the maximum likelihood estimate (MLE)
$$\psi_i^{-1} = \frac{\delta - 1 + N - 2i + |D|}{\sum_{j=i+1}^{N}\beta_{ij}^2\theta_i^{-1} + \theta_i^{-1} + \sum_{n=1}^{D}(x_{ni} - \sum_{j=i+1}^{N}\beta_{ij}x_{nj})^2}$$
2. Computed multiple MLE samples and estimated $E[\psi^{-1}|\theta_i, \beta_i, D]$ using weighted sampling
3. Sampled $P(\theta_j)$ using its CDF and a uniform distribution on the interval [0,1] to compute $\psi_i^{-1} = \frac{1}{M}\sum_{j=1}^{M}\hat{\psi}_j^{-1}P(\theta_j)$

Data Sets

1. **Fine Grain (FG)**: contains building simulations built from small brute force increments across ***14*** building parameters. ***250*** simulations were sampled without replacement, and ***150*** test simulations were used to evaluate the GGM model.
2. **Model Order 1 (MO1)**: contains ***299*** building simulations and ***~150*** building parameters. The building parameters were independently set to their min or max value, while other parameters were set to their average value. This results in two simulations per building parameter. The entire MO1 data set was used to build GGM model.
3. **Model Order 2 (MO2)**: contains pairwise min and max combinations across the same building parameter set. ***300*** sampled MO2 simulations were used to evaluate the GGM model built from the entire MO1 data set.

Results



(a) Bayesian Estimates (b) Random Estimates

Figure 1. Compares Bayesian GGM's error against a [0, 1] uniform distribution's error on estimating FG building parameters. Additionally, it shows how estimates align with test building parameter values.

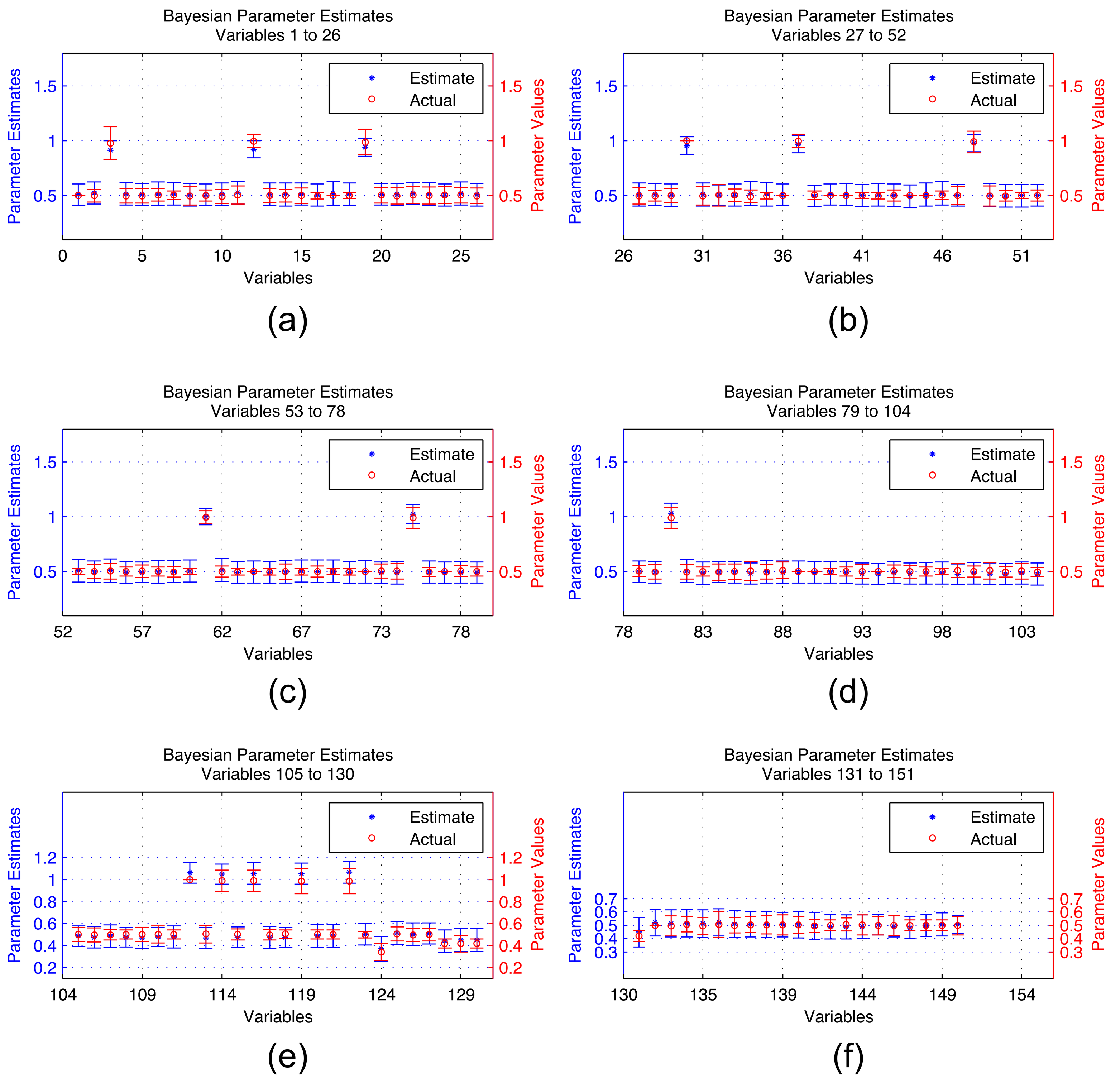
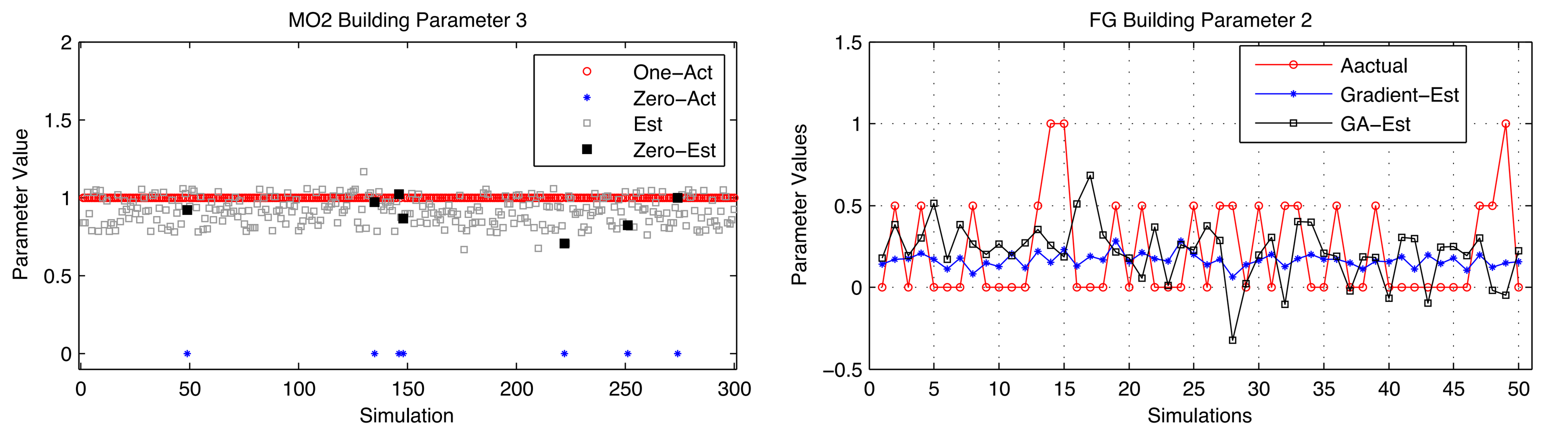


Figure 2. Bayesian GGM's parameter estimates compared against the actual parameter values on 300 randomly sampled MO2 simulations.



(a) (b)

Figure 3. Figure 3(a) highlights that building parameter three has values that occasionally differ greatly from the parameter's mean, and Figure 3(b) compares parameters estimated via a genetic algorithm and gradient optimization.

Conclusions

- Our FG and MO2 experimental results indicated that the GGM performs well at estimating building parameters.
- Overall, the Bayesian models built using the FG and full MO1 data sets are statistically better than the uniform distribution, which is expected.
- However, our current GGMs have difficulty estimating parameters that deviate significantly from the mean. This implies we need to explore different hyper-parameter settings, which may induce more estimation variance, or possibly a mixture model approach, which will allow more variable means.

Acknowledgments

This work was funded by field work proposal CEBT105 under the Department of Energy Building Technology Activity Number BT0305000. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. Our work has been enabled and supported by NSF funded RDAV Center of the University of Tennessee, Knoxville (NSF grant no. ARRA-NSF-OCI-0906324 and NSF-OCI-1136246).

References

A. Dobra, C. Hans, B. Jones, J. Nevins, G. Yao, and M. West. Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis*, 90(1):196–212, 2004.

F. Li, Y. Yang, and E. Xing. From lasso regression to feature vector machine. *Advances in Neural Information Processing Systems*, 18:779, 2006.