Data-Driven Simulation Calibration with Machine Learning

IL-Chul Moon
Department of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Il-Chul Moon

- Short Bio
  - Assistant, Associate Professor (2011-Present)
    - Department of Industrial and Systems Engineering, KAIST
  - PhD in Societal Computing (2008)
    - Institute of Software Research, School of Computer Science, Carnegie Mellon University
- Research Interest
  - Theoretic interests : ABM, M&S Formalism, Model Validation, Deep Generative Models…
  - Application domains : Defense, Recommendations, Profiling, Disaster Management…
- More information : aai.kaist.ac.kr

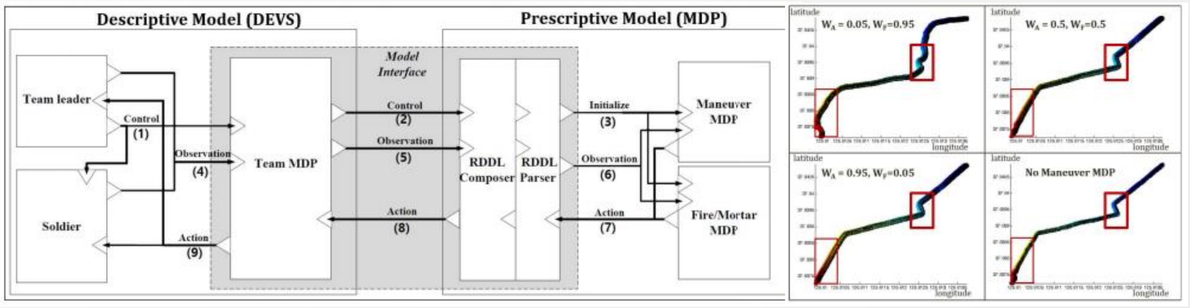
Fundamentals → Applications → Fusions

Modeling and Simulation	M&S Formalims (IEEE T-SMC 2016)	Simulation with Flattened Model Hierarchy (ACM TOMACS 2016)	Social Network Geospace ABM (AAMAS 2007)	Disaster Modeling (IEEE T-SMC 2018)
	Bayesian Nonparametric CF (IJCAI 2016)	VAE+RNN for Health Care (ICDM 2018)	VAE-CF+Tensor Fact. for Political Analysis (AAAI 2018)	Terrorists on Social Net. and Geospace (IEEE Intel. Sys. 2007)
Artificial Intelligence	Adversarial Dropout on CNN (AAAI 2018)	VAE for Collaborative Filtering (CIKM 2017)	Hierarchical Attention for Seq. Recommendation (AAAI 2019)	ABM+MDP (IEEE T-SMC Accepted)
	Adversarial Dropout on RNN (AAAI 2019)	Guided Hierarchical Topic Model (IEEE T-KDE 2017)	Social Network+Text for Email Analysis (KDD 2009)	Emotion Modeling with Brownian ABM (AAMAS 2014)

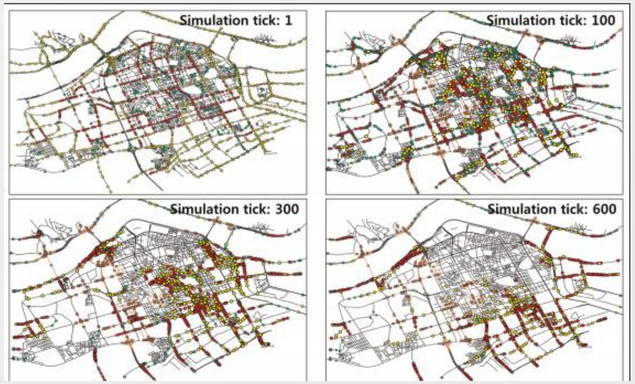
Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Modeling and Simulation

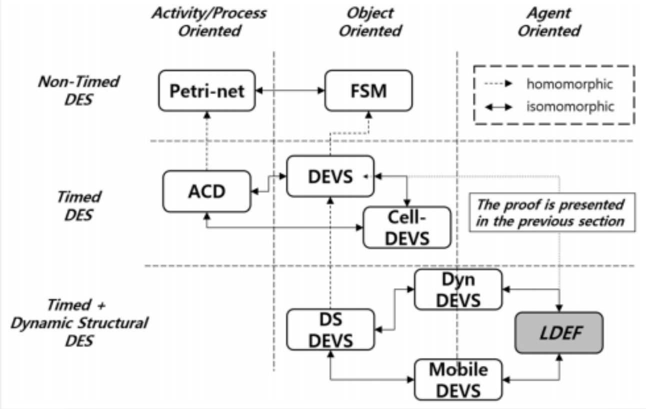
- Modeling and simulation on socio-economic problems
  - Commerce, disaster management, defense
- Theories on modeling and simulation



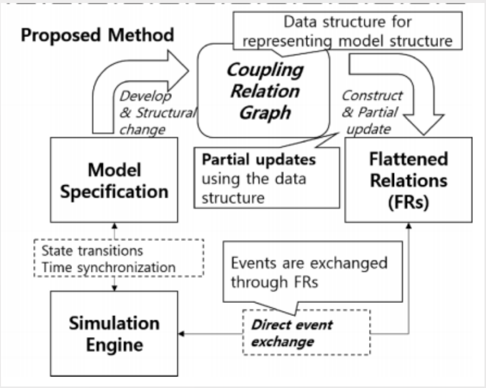
Descriptive-Prescriptive ABM with intelligent behaviors
IEEE T-SMC, Accepted



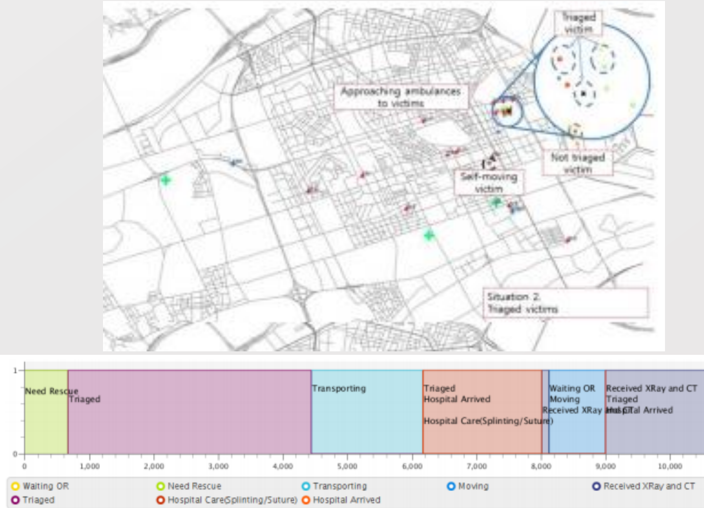
Urban Evacuation with ABM
SIMULATION, 2014



LDEF Formalism:
Formalism on ABM
IEEE T-SMC, 2016



Discrete Event Simulation
Engine Optimization
ACM TOMACS, 2016

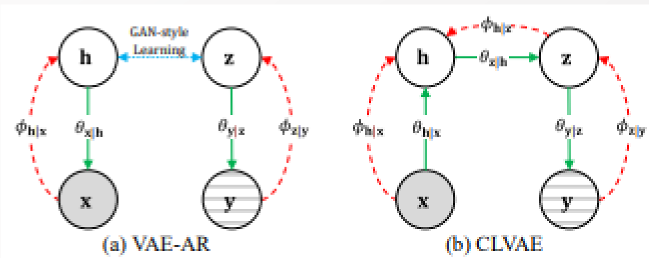


Emergency medical service with ABM
IEEE T-SMC, 2018

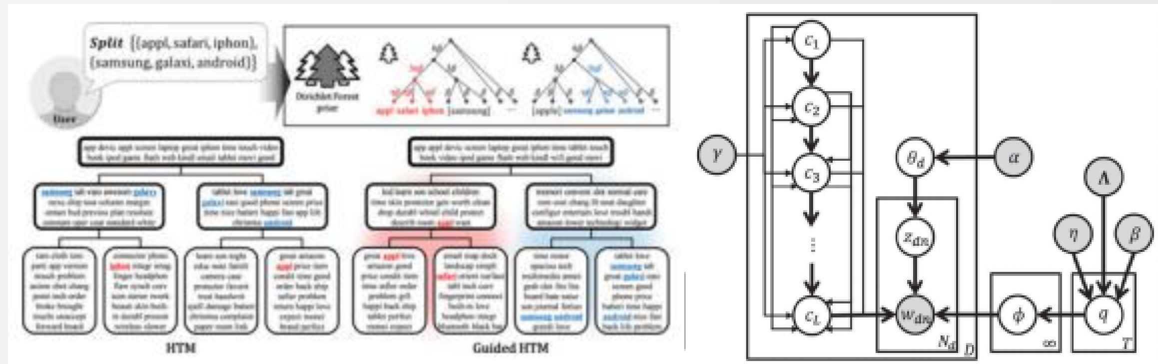
KAIST
Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Artificial Intelligence

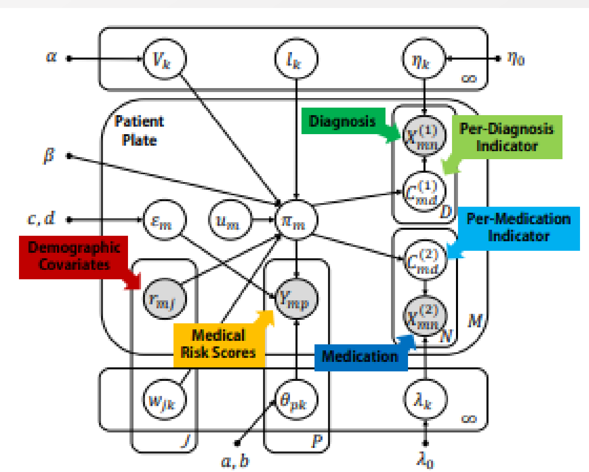
- Modeling and analyzing on socio-economic problems
  - With Probabilistic Graphical Model and Deep Generative Model
- Theories on neural network learning and deep generative models



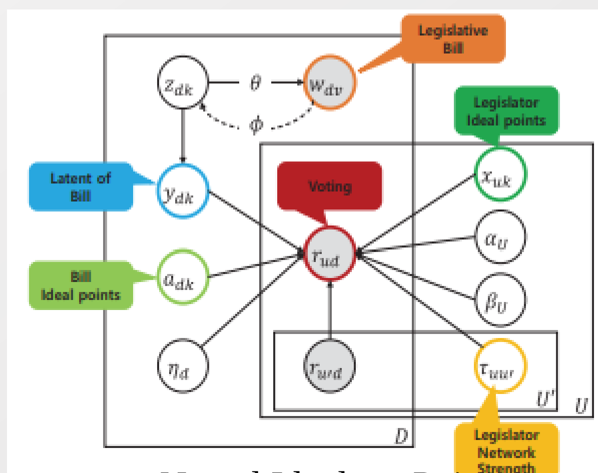
GAN-regularized, Ladder Variational Autoencoder for Collaborative Filtering, CIKM 2017



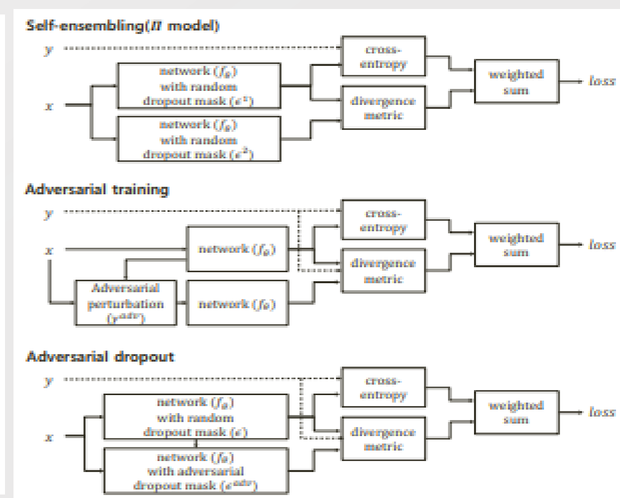
Guided hierarchical topic model, IEEE T-KDE, 2017



Bayesian Nonparametric Collaborative Topic Poisson Factorization For National Health-Care, IJCAI 2016



Neural Ideology Point Estimation Network for Law-Makers, AAAI 2018



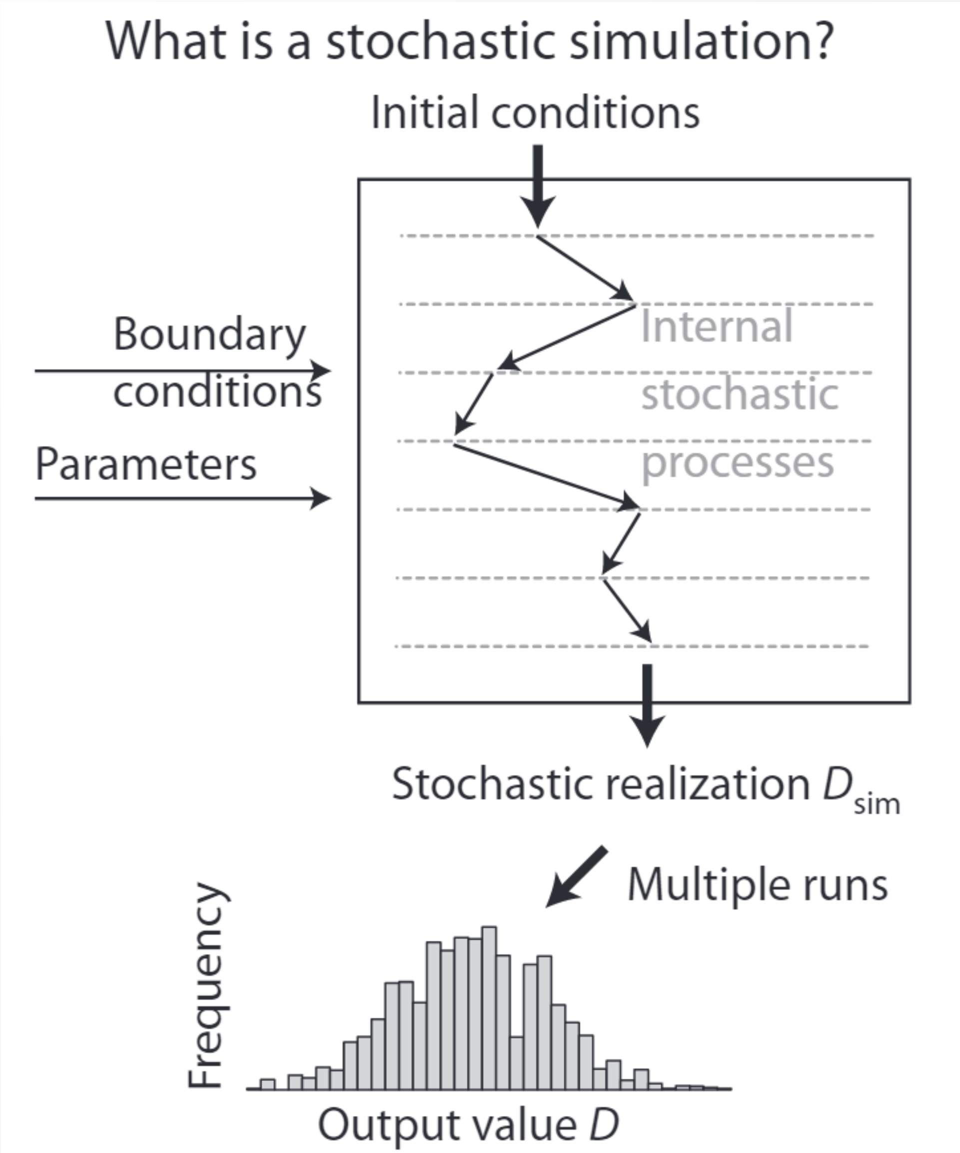
Adversarial Dropout for CNN/RNN, AAAI 2018, AAAI 2019

Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

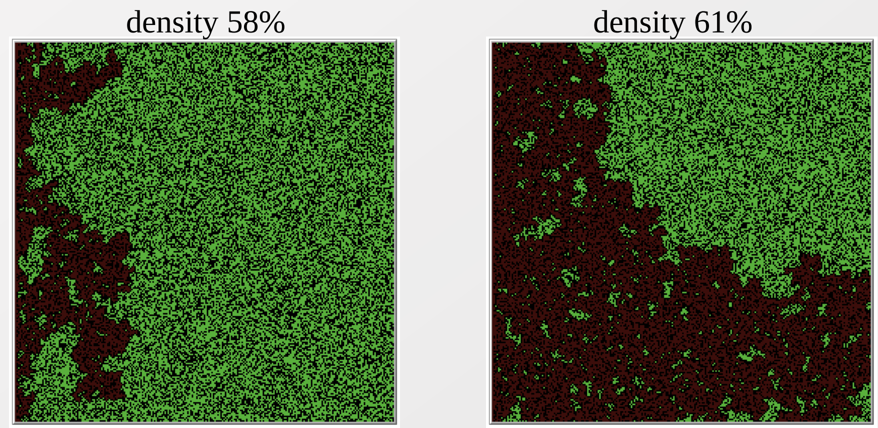
SIMULATION
AS GENERATIVE MODEL

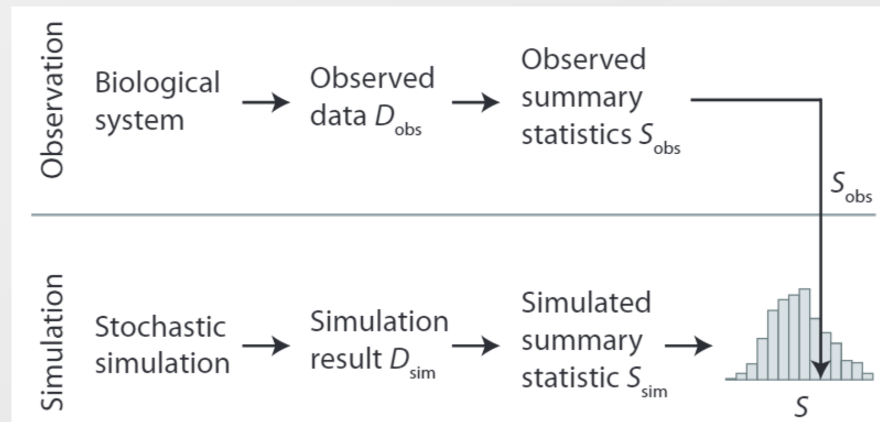
Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Simulation as Data Generation



<Fire Burning Model>

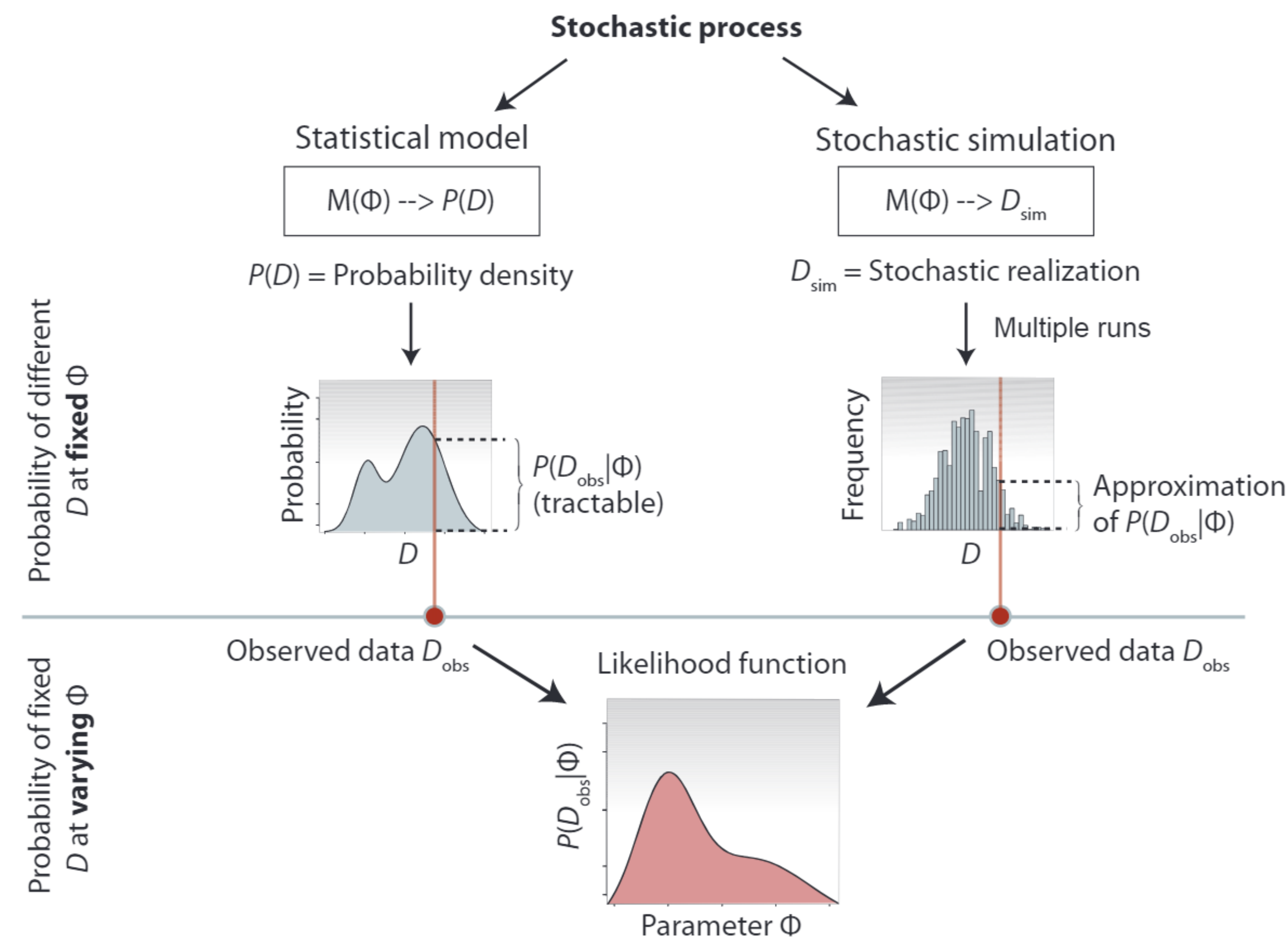




Simulation executes
Models with
Input Conditions + ***Parameters***
to generate ***Output Distribution***

Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

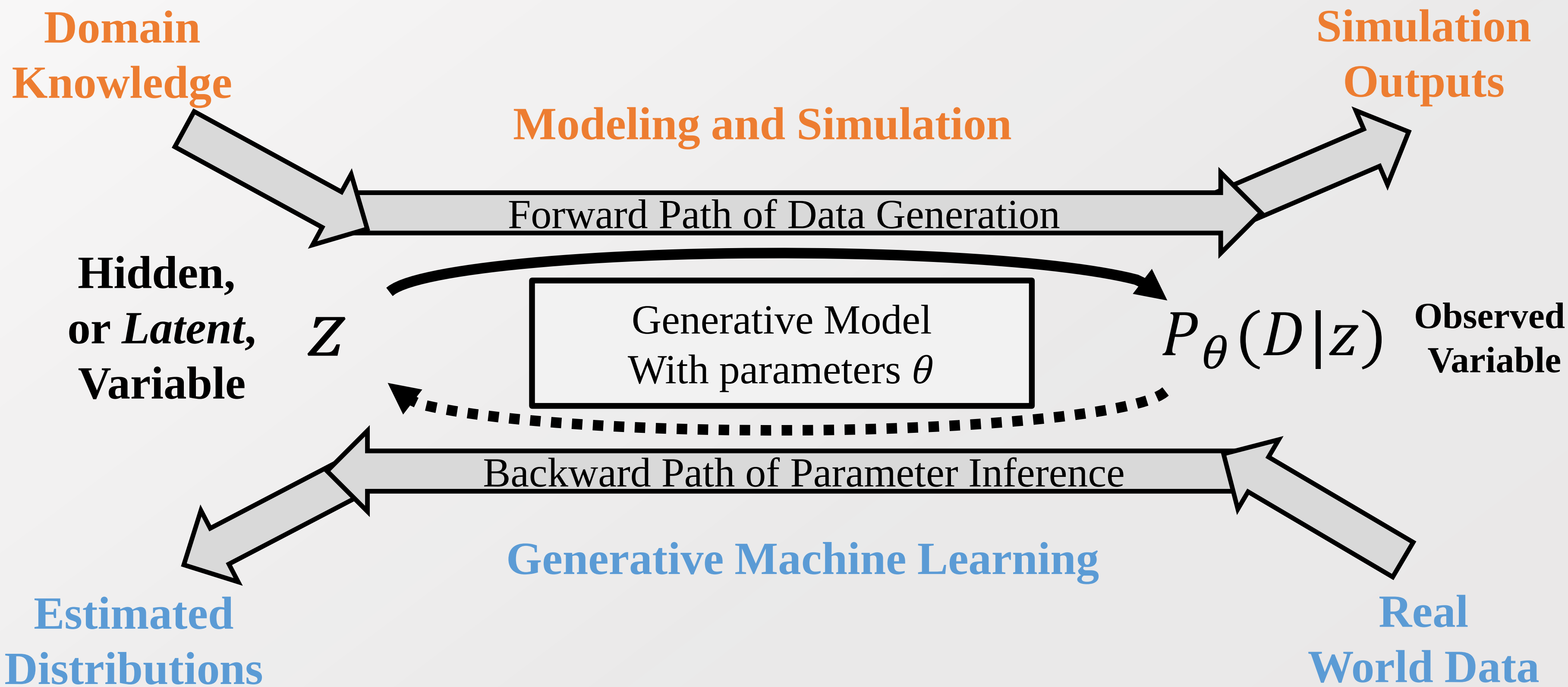
Statistical Model and Stochastic Simulation



Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Generative Model in Two Directions

- Simulation expects to generate a **realistic virtual** output
  - Validation ensures the realism of the generated output
- Not all variables in the simulation models are known
  - Some variables are selected by domain experts or modelers

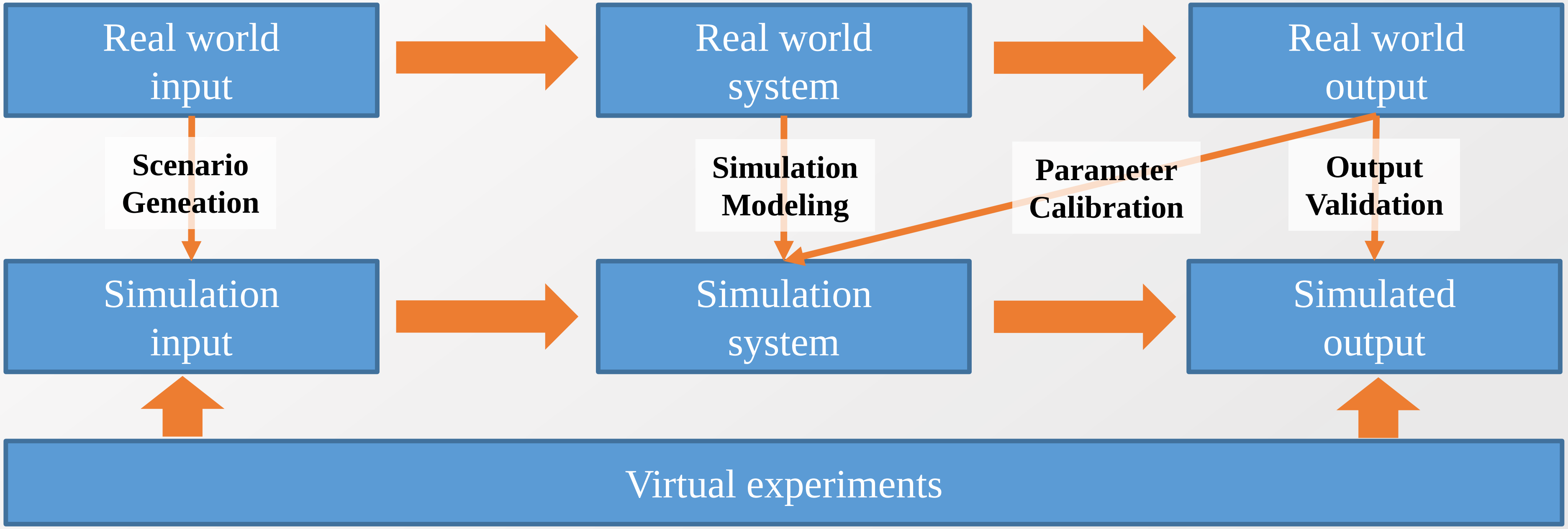


Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

SIMULATION CALIBRATION

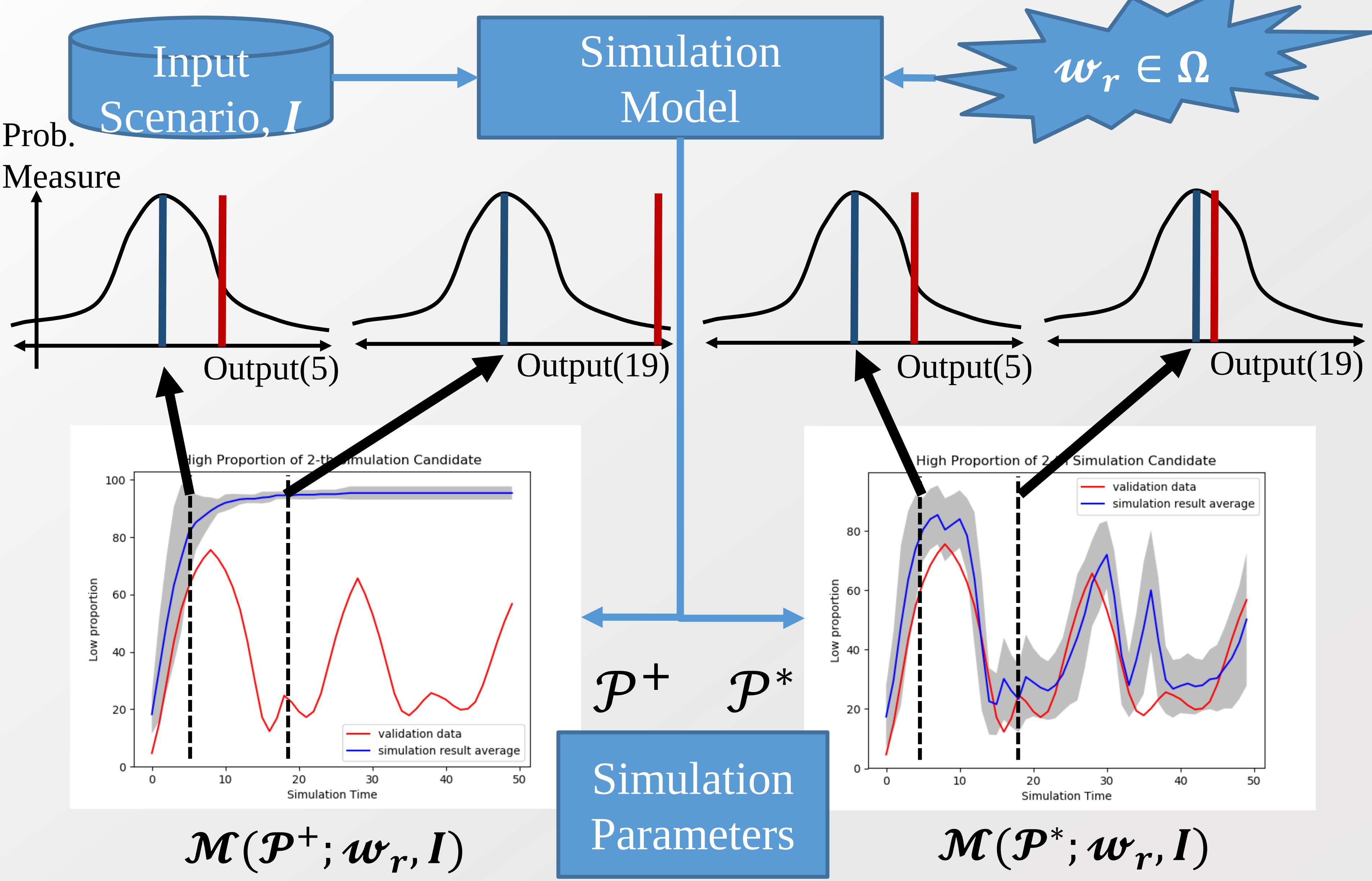
Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Simulation Validation and Calibration



- Validation requires
  - Realistic input scenario, which can be obtained from past data
  - Realistic simulation model, which can be designed by domain experts
  - Realistic simulation parameters
    - Some parameters are introduced by abstraction
    - <u>Real world abstraction inevitably introduces approximations on parameters</u>
    - How to well approximate parameters == **<u>Calibration</u>**

Concent of Calibration and Validation



Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Formal Description of Calibration

$$\mathcal{P}^* = \arg\min_{\mathcal{P}} d\left(E_R\left[\mathcal{S}\left(\mathcal{M}\left(\mathcal{P}; w_r, I\right)\right)\right], \mathcal{D}\right)$$

Modeler?
Domain Experts?
Users?
OR
Machine Learning Model?

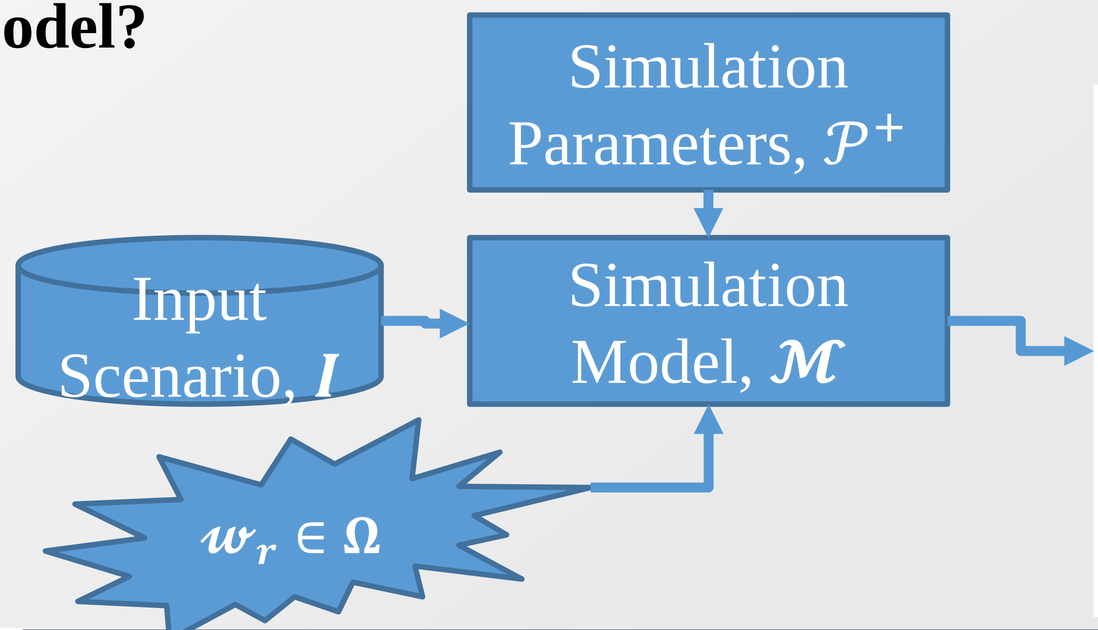
Deviation Function i.e. MAPE, MSE, Likelihood

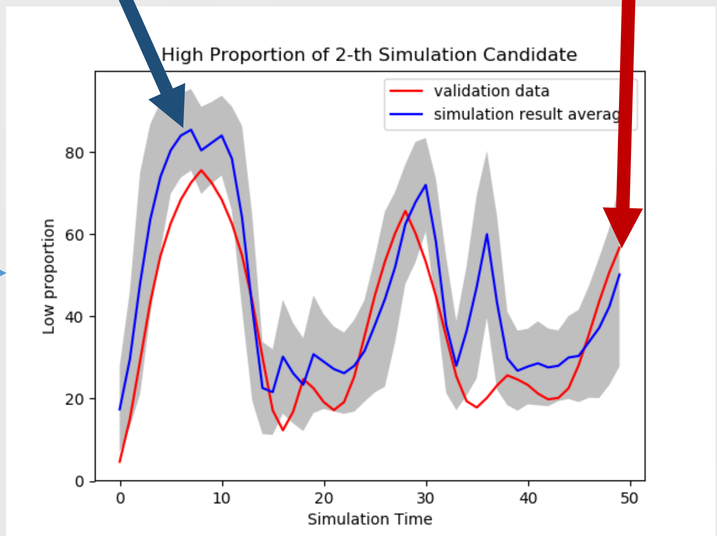
Single Simulation Exec.

Summary Statistics from Single Run

Replication for Expected Values

Real World Validation Target





Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

CALIBRATION FRAMEWORK AND PROCEDURE

Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Calibration Procedure

i.e. Agent-Based Model and Simulations

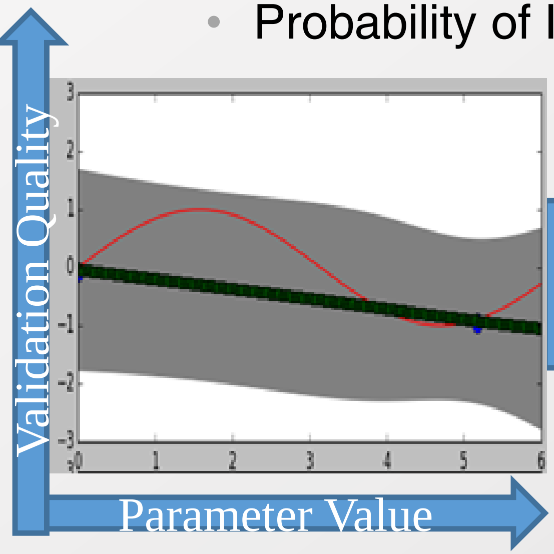
- $\mathcal{P}, w_r, I$ by simulation types
  - Initial parameter setup, $\mathcal{P}$
  - Dynamic parameter adaptation, $\mathcal{P}_{dyn}$
  - Heterogeneous parameter for instantiated sub-models, $\mathcal{P}_{het}$

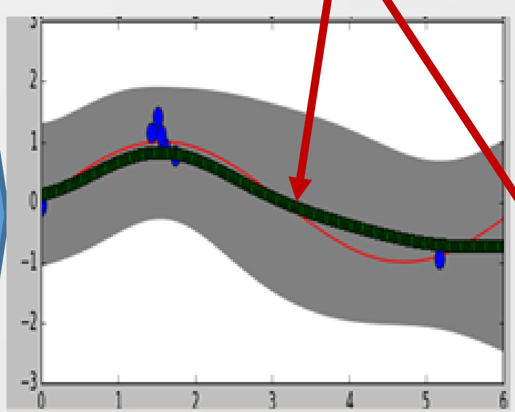
1. Set the summary statistics
   1. Collect validation data $\mathcal{D}$
   2. Set the summary statistics function $\mathcal{S}$
2. Select the simulation performance measure d
   1. d could be a likelihood, MAPE, MSE, etc.
3. Optimize the simulation parameter
   1. Static Calibration: $\mathcal{P}^* = \arg\min_{\mathcal{P}} d\left(E_R\left[\mathcal{S}\left(\mathcal{M}(\mathcal{P}; w_r, I)\right)\right], \mathcal{D}\right)$
   2. Dynamic Calibration: $\mathcal{P}_{dyn}^* = \arg\min_{\mathcal{P}_{dyn}} d_{dyn}\left(E_R\left[\mathcal{S}_{dyn}\left(\mathcal{M}(\mathcal{P}_{dyn}; w_r, I)\right)\right], \mathcal{D}_{dyn}\right)$
   3. Heterogeneous Calibration: $\mathcal{P}_{het}^* = \arg\min_{\mathcal{P}_{het}} d_{het}\left(E_R\left[\mathcal{S}_{het}\left(\mathcal{M}(\mathcal{P}_{het}; w_r, I)\right)\right], \mathcal{D}_{het}\right)$

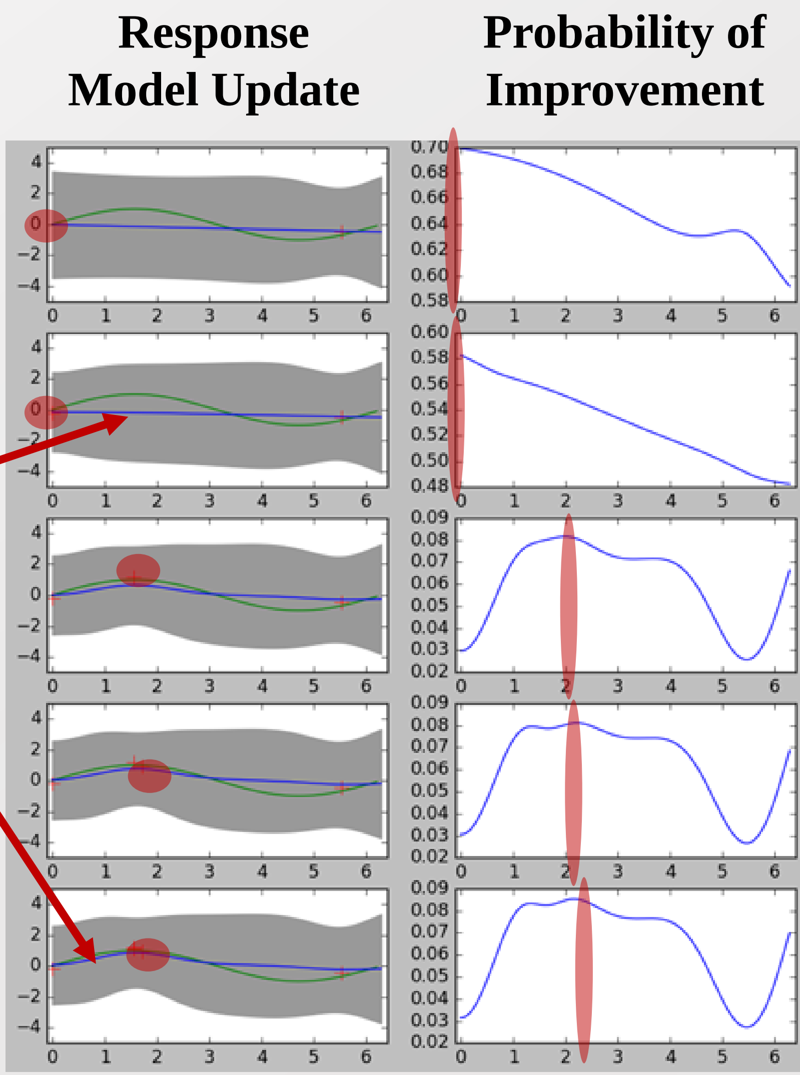
Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Concept of Calibration on $\mathcal{P}$

- Traditional methodology of data-driven calibration on $\mathcal{P}$
- Response surface methodology
  - Response surface model from data
    - Gaussian Process, Piece-wise Linear Regression, Meta-Modeling…
  - Experimental design
    - Latin Hypercube, Full Factorial, Daguchi method…
  - Acquisition function
    - Probability of Improvement…

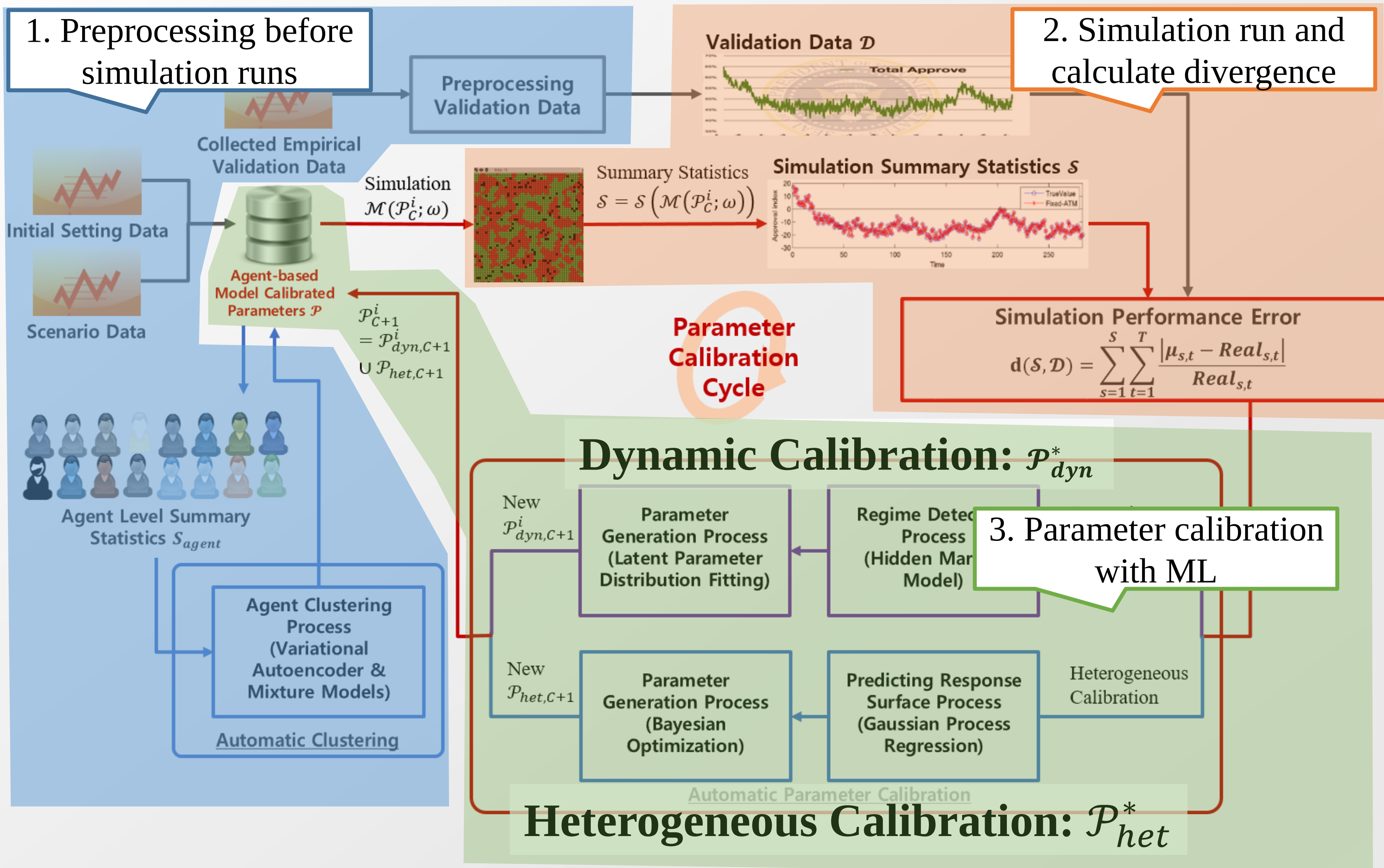






Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

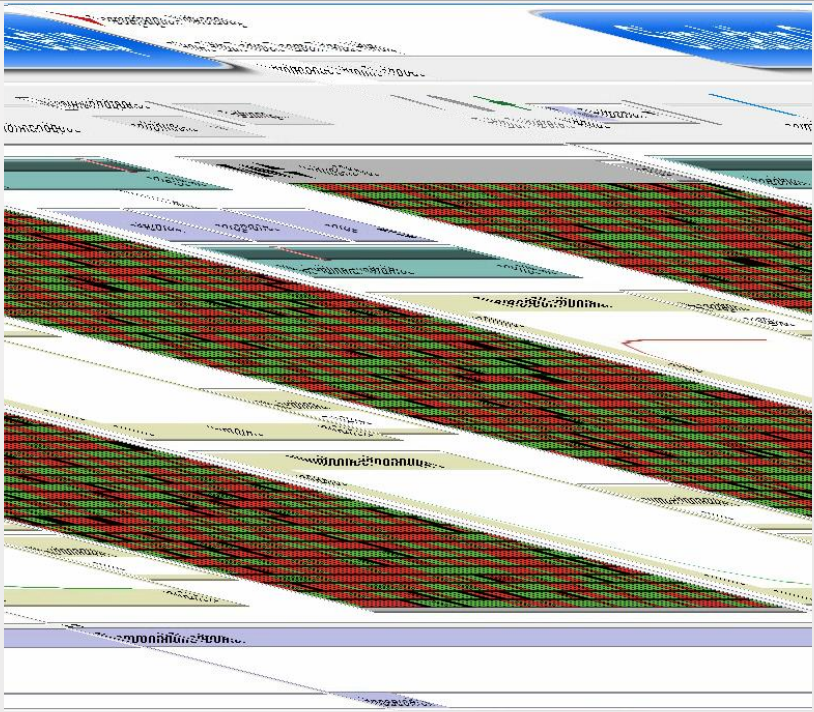
Overview on Data-Driven Simulation Calibration



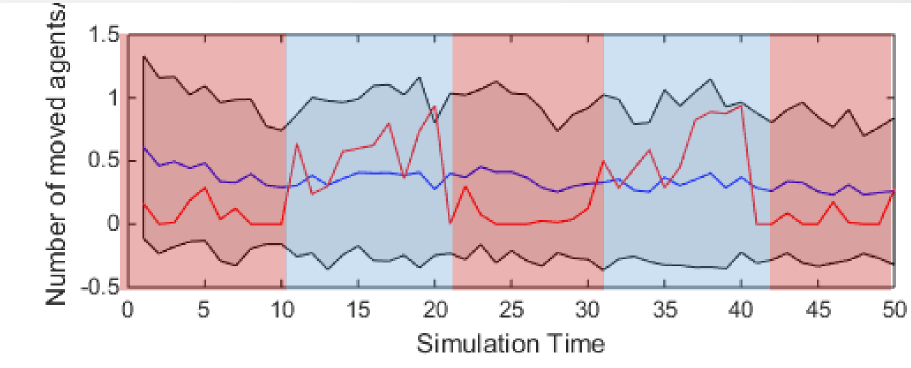
Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

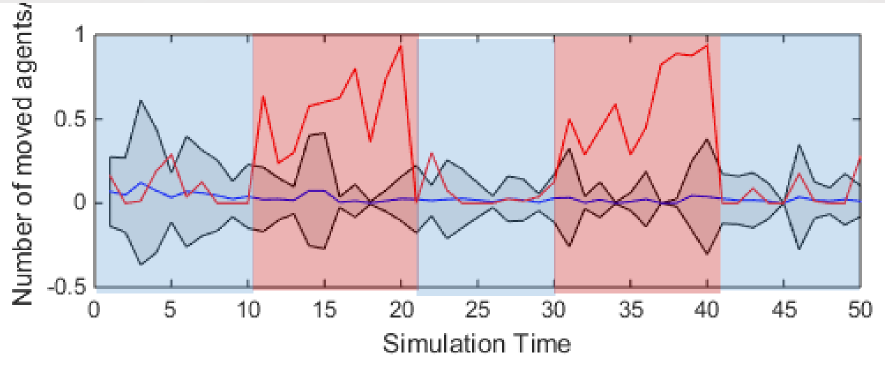
Concept of Calibration on $\mathcal{P}_{dyn}$

- $\mathcal{P}_{dyn}$ assumes parameter to be varied by t
- $t \in [1..T]$ requires too much separate setting
- Pseudo Code
  - Divide and Calibrate for cycle C
    - Suggest $\mathcal{P}_{dyn,C}$ with multiple candidates
    - Identify the **temporal regime** with better validation with a candidate
    - Selectively update $\mathcal{P}_{dyn,C+1}$ with well-fitted temporal regime
- Regime Detection
  - Hidden Markov Model….



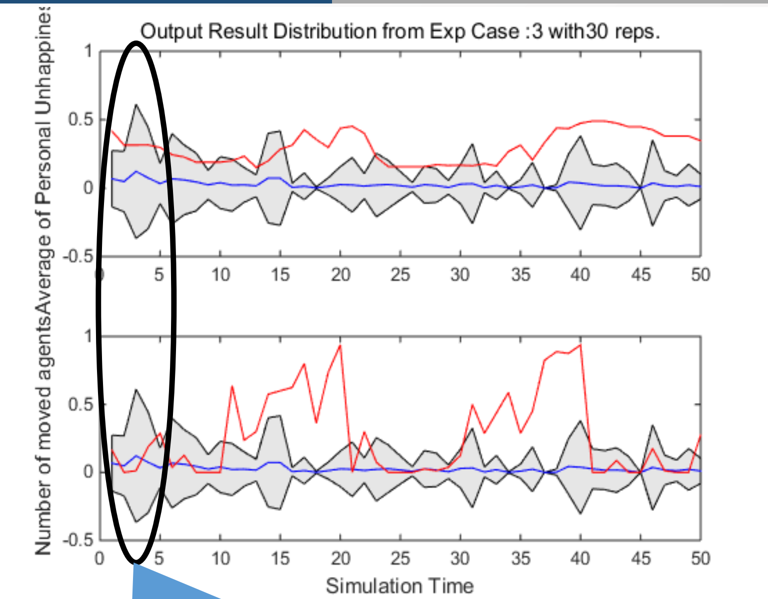
Dynamically Changed Simulation Parameter





Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

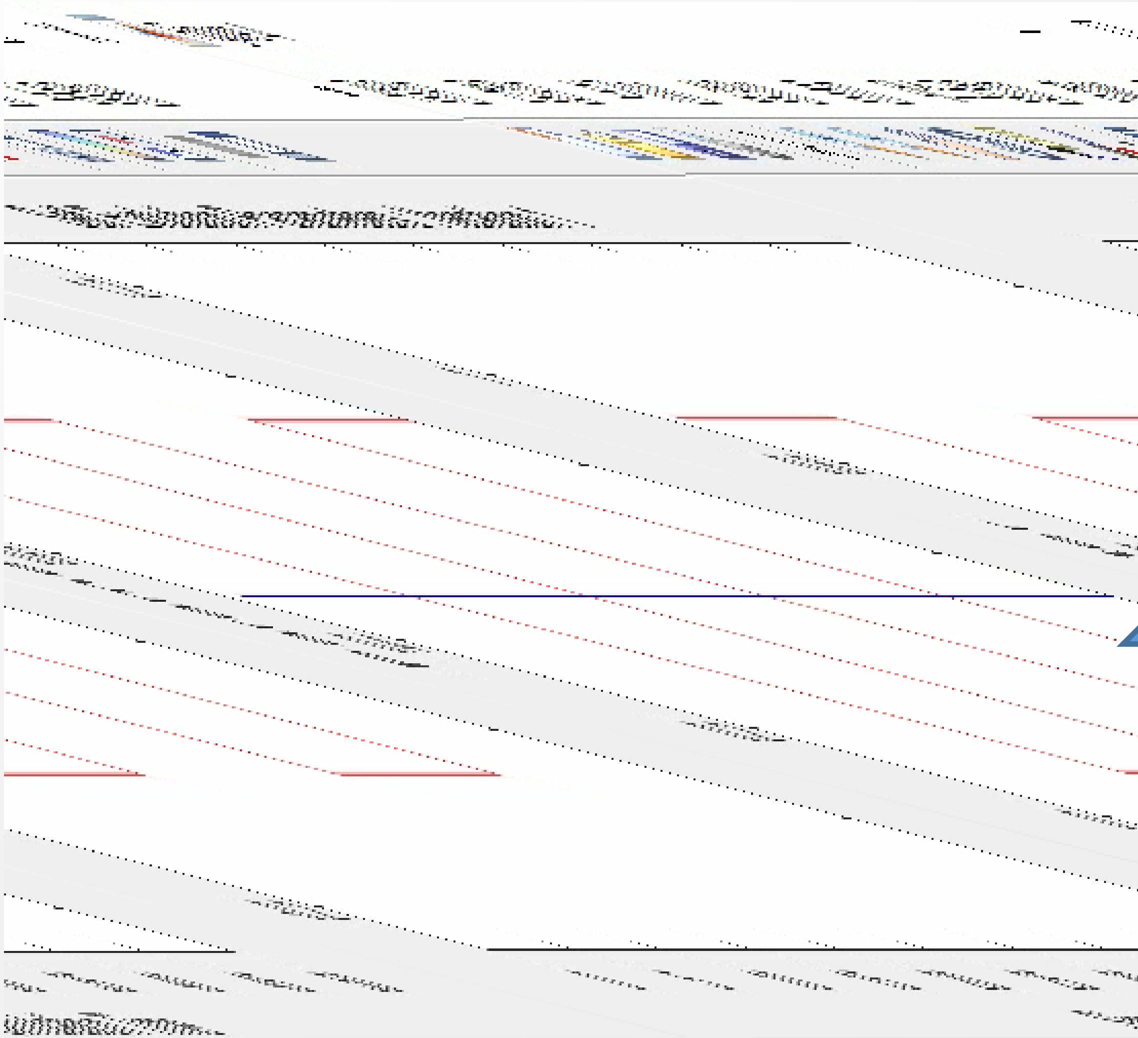
Sample Case of Temporal Regime Detection



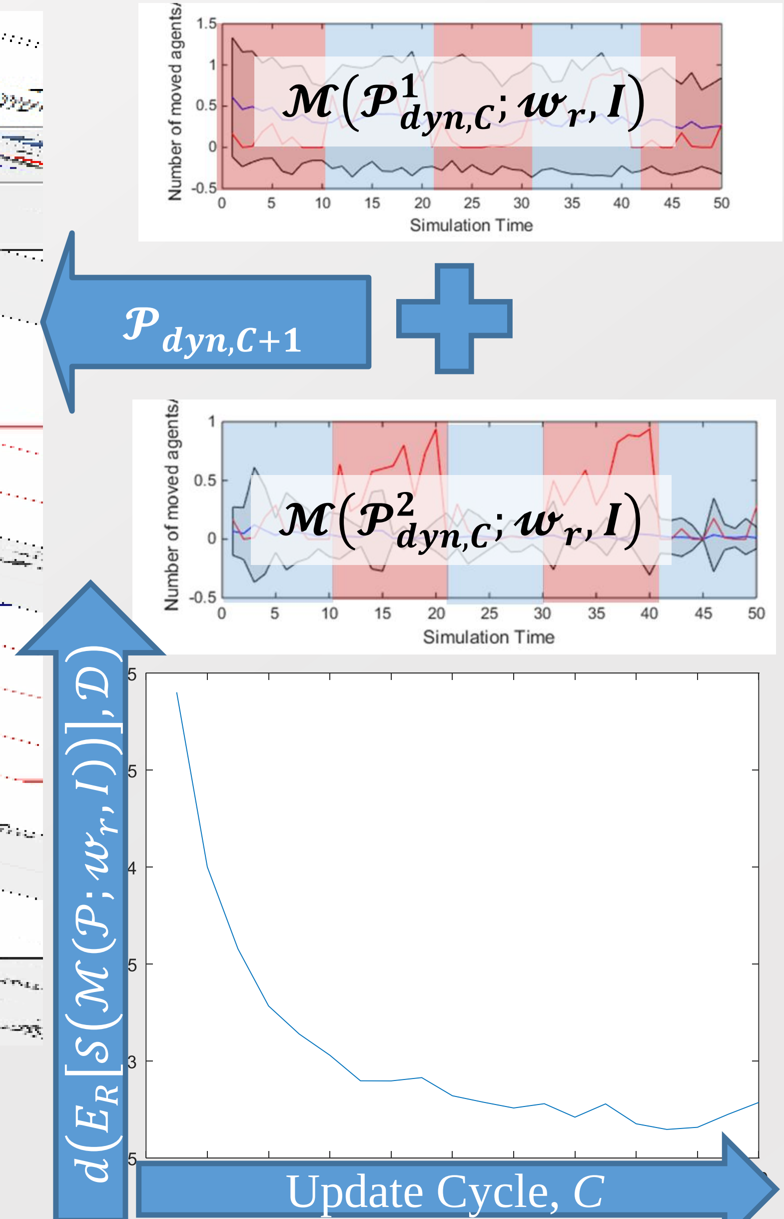
Two simulation outputs
$d\left(E_R\left[\mathcal{S}\left(\mathcal{M}\left(\mathcal{P}; w_r, I\right)\right)\right], \mathcal{D}\right)$
→ 2D Temporal Data

Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Matching Well-Fitted Regime and $\mathcal{P}_{dyn}$



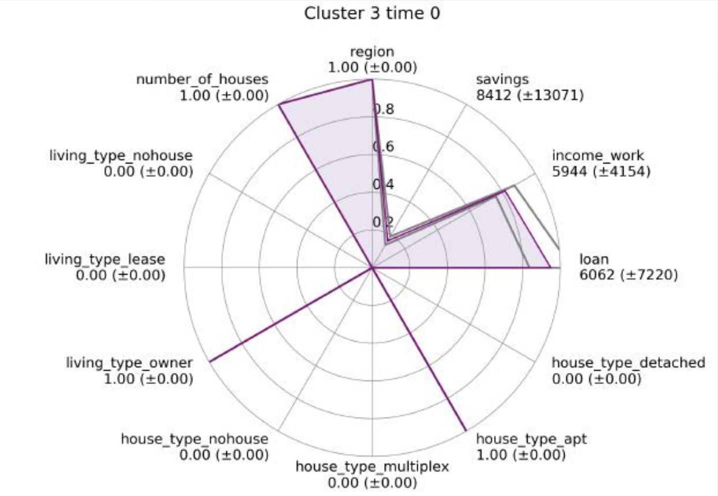
$\mathcal{P}_{dyn}$ update over the cycle, C

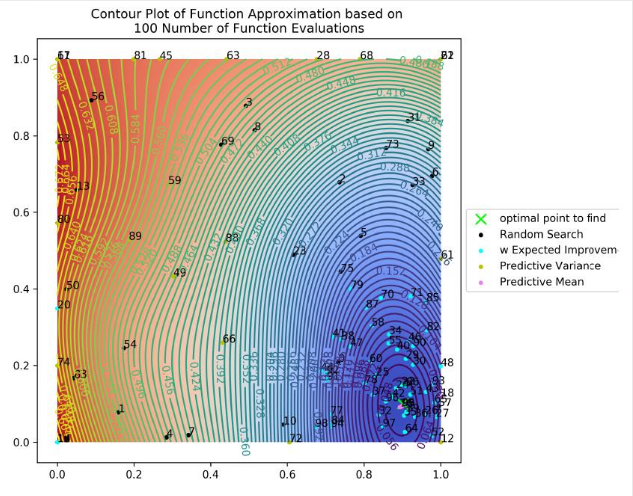


Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Concept of Calibration on $\mathcal{P}_{het}$

- $\mathcal{P}_{het}$ assumes parameter to be varied by agents
- $i \in [1..N]$ requires too much separate setting
- Pseudo Code
  - Clustering with agent demographics
  - Calibrate for cycle C
    - For each agent cluster, i
      - Update response surface curve by Gaussian Process
      - Suggest $\mathcal{P}^{i}_{het,C+1}$ with expected improvement acquisition function on GP
- Agent Cluster Detection
  - Variational Autoencoder, Gaussian Mixture Model, Gaussian Process…

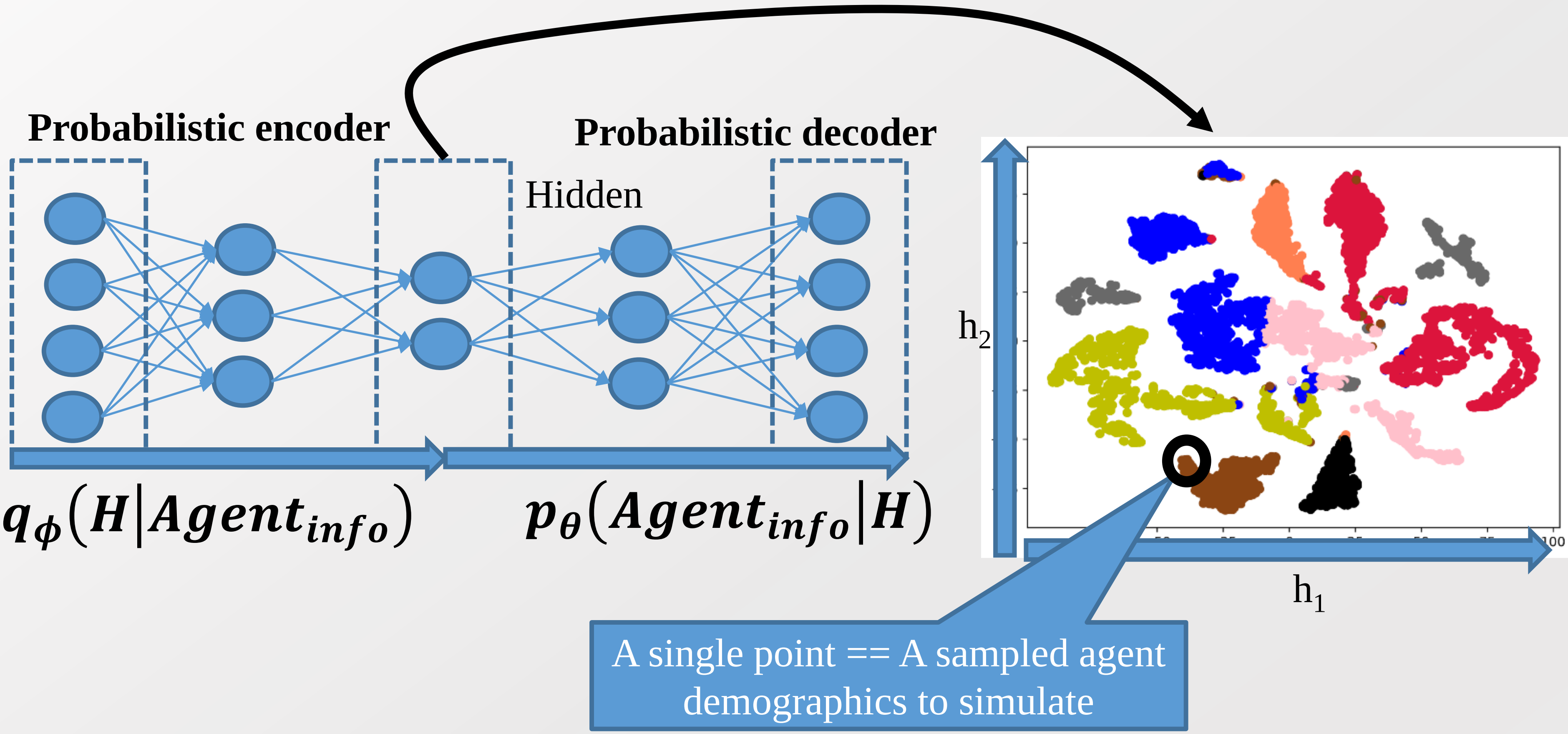




Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Agent Embedding with VAE

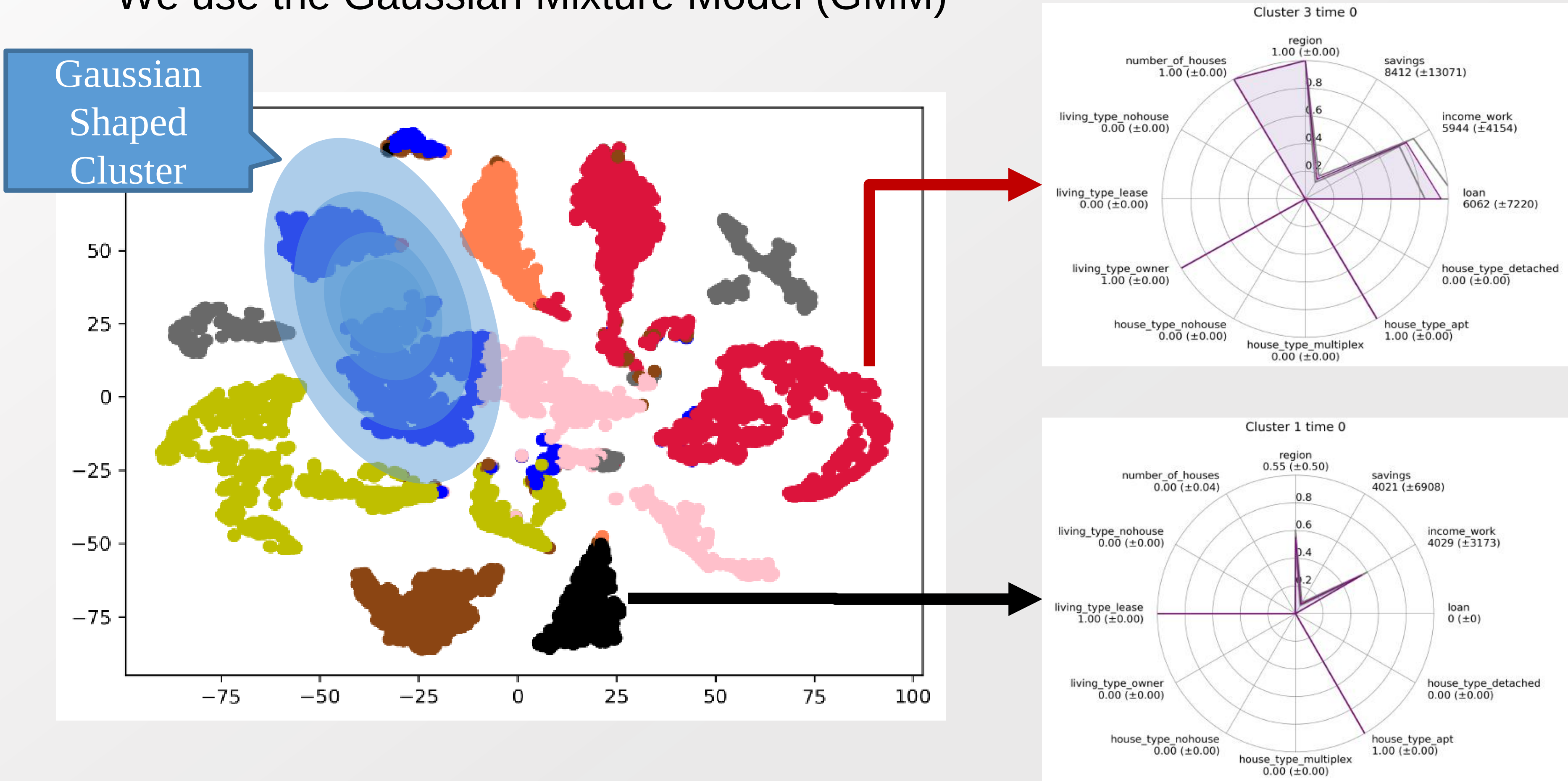
- Agent demographic information can be high-dimensional
  - Need to embed the agent information into the low dimensions
  - Use Autoencoder, and we use the variational autoencoder (VAE)
- Clustering requires further operation by Gaussian Mixture Model



Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Agent Clustering with GMM

- Low dimension agent demographics embedding
  - Closeness between two embedded points== similarity between two agents
  - Use clustering algorithm to finalize agent clusters
  - We use the Gaussian Mixture Model (GMM)

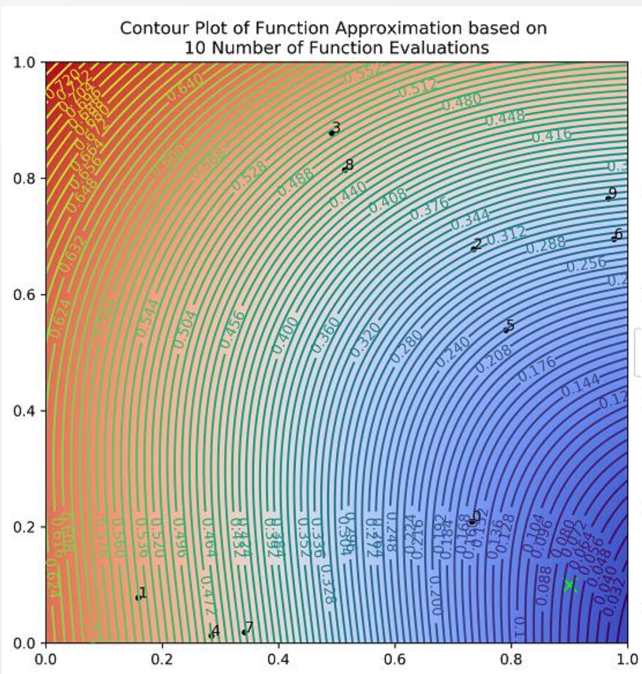


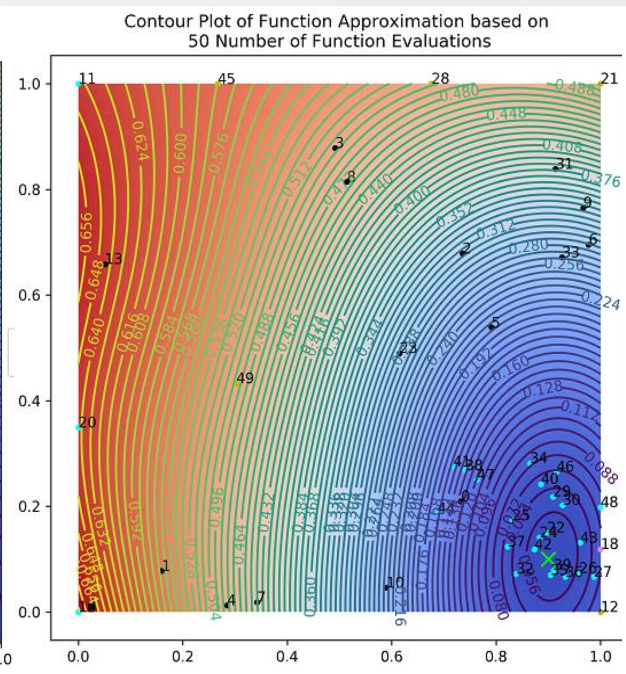
Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

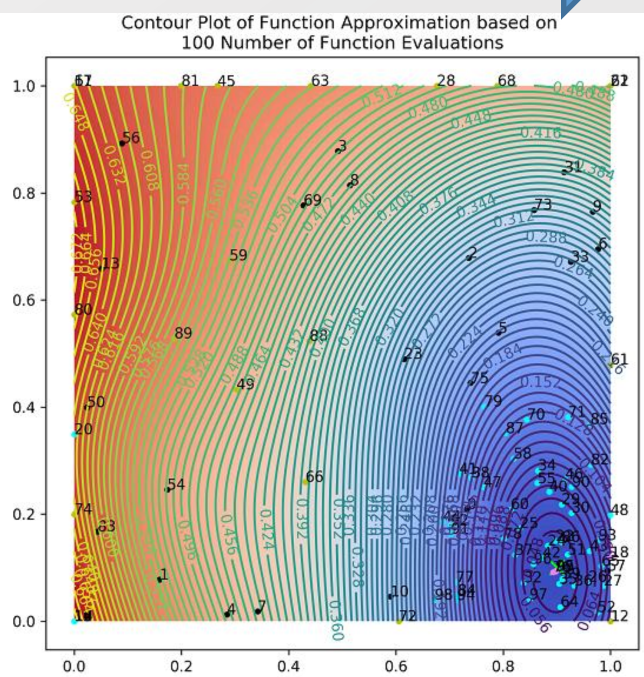
Response Surface for Each Agent Cluster

- Multiple cycle of calibration iterations
  - Multiple points of $< d_{het}\left(E_R\left[\mathcal{S}_{het}\left(\mathcal{M}\left(\mathcal{P}_{het}; w_r, I\right)\right)\right], \mathcal{D}_{het}\right), \mathcal{P}^i_{het,C} >$
  - Gaussian Process approximation on the collection of parameter points
  - Acquisition functions with expected improvement

Response surface learning with Gaussian Process

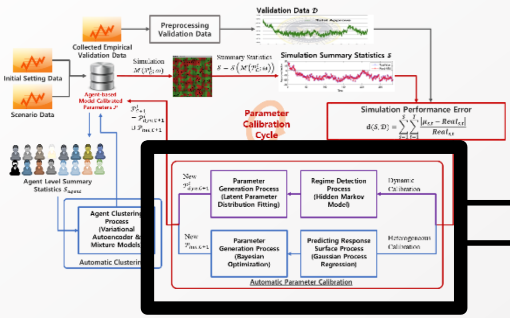


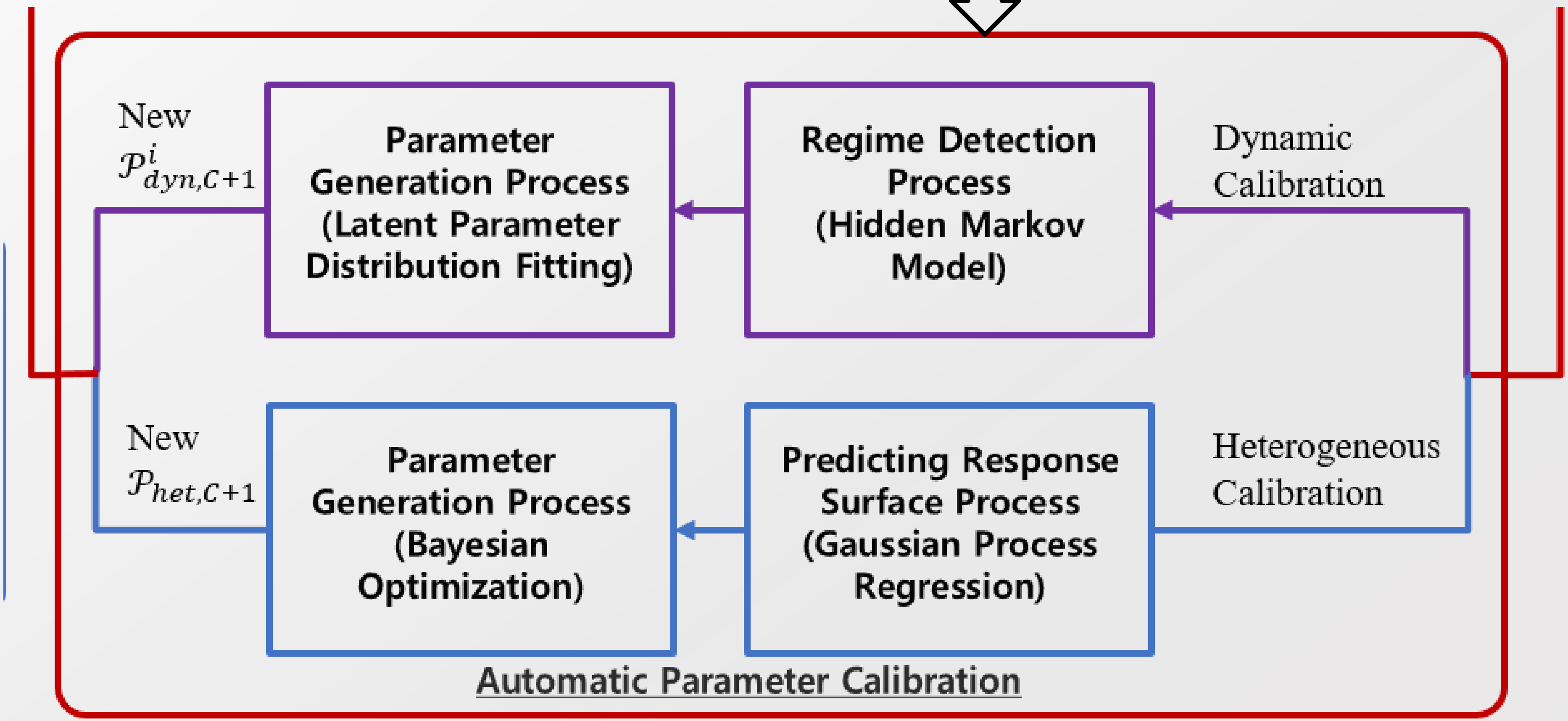




Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Highlighted on Data-Driven Simulation Calibration





Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Pseudo Code of Data-Drive Sim. Calibration

Algorithm 1: Calibration Framework Algorithm

input : Input parameter combination $\mathcal{P}^{in} = \mathcal{P}^{in}_{dyn} \cup \mathcal{P}^{in}_{het}$

output: Calibrated parameter combination $\mathcal{P}^{out} = \mathcal{P}^{out}_{dyn} \cup \mathcal{P}^{out}_{het}$

Function `CalibrationFramework`($\mathcal{P}^{in}_{dyn} \cup \mathcal{P}^{in}_{het}$):

  $\mathcal{P}_{dyn,0} = \mathcal{P}^{in}_{dyn}$

  $\mathcal{P}_{het,0}$=`AgentClustering`($\mathcal{P}^{in}_{dyn} \cup \mathcal{P}^{in}_{het}$) (see Algorithm 3)

  for c *in range*(C_{cal}) **do**

    if $0 \leq c - \left[\frac{c}{C_{dyn}+C_{het}}\right](C_{dyn} + C_{het}) < C_{dyn}$ **then**

      $\mathcal{P}_{dyn,c+1}$=DynamicCalibration($\mathcal{P}_{dyn,c} \cup \mathcal{P}_{het,c}$) (see Algorithm 2)

      $\mathcal{P}_{het,c+1} = \mathcal{P}_{het,c}$

    else if $C_{dyn} \leq c - \left[\frac{c}{C_{dyn}+C_{het}}\right](C_{dyn} + C_{het}) < C_{dyn} + C_{het}$ **then**

      $\mathcal{P}_{het,c+1}$=HeterogeneousCalibration($\mathcal{P}_{dyn,c} \cup \mathcal{P}_{het,c}$) (see Algorithm 3)

      $\mathcal{P}_{dyn,c+1} = \mathcal{P}_{dyn,c}$

  Set $\mathcal{P}^{out} = \mathcal{P}^{opt}_{dyn} \cup \mathcal{P}^{opt}_{het}$ to have the lowest simulation error

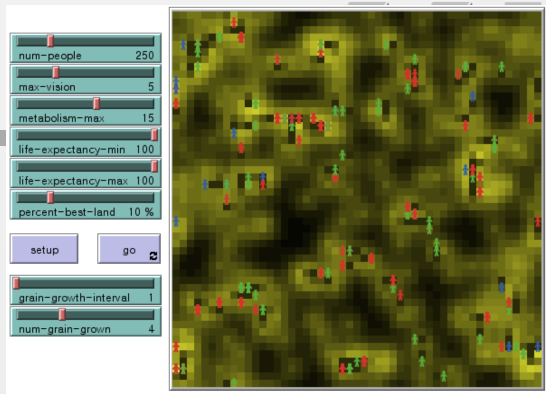
  return $\mathcal{P}^{out}_{dyn} \cup \mathcal{P}^{out}_{het}$

KAIST Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

EXPERIMENTS

Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Test Case 1



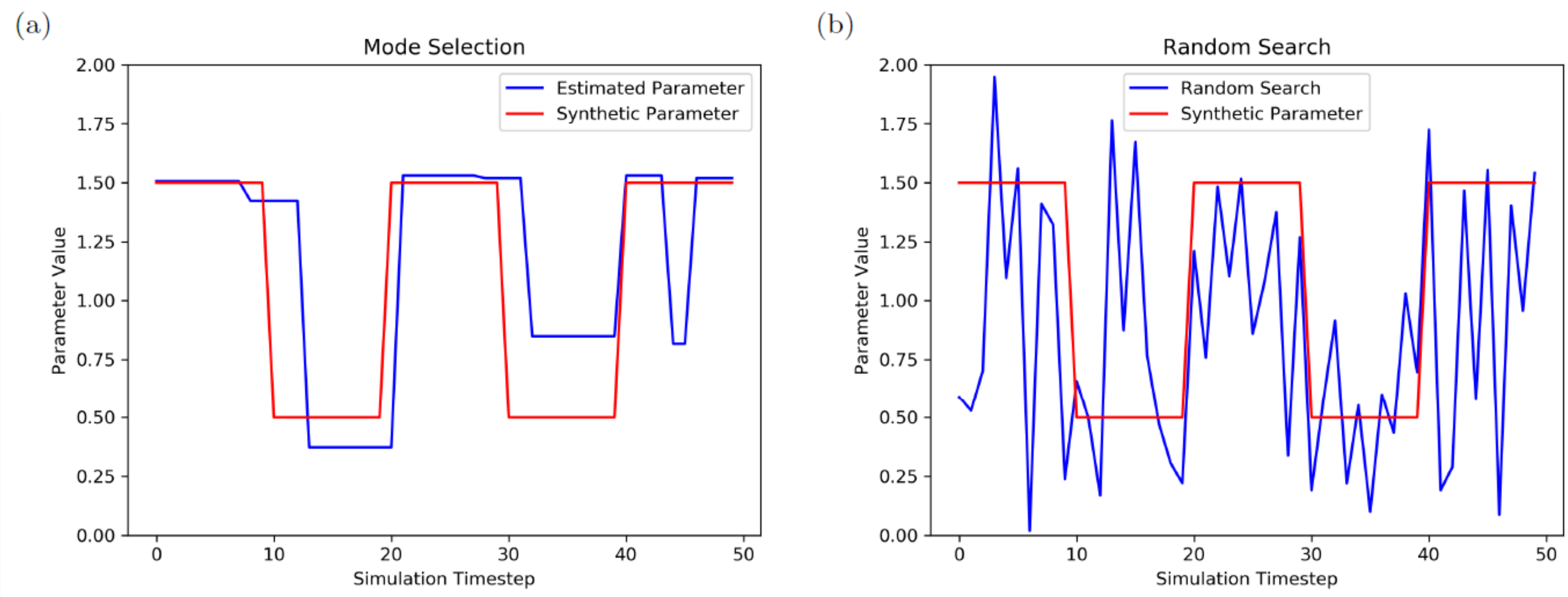
- [Model Description] Wealth Distribution ABM
  - Sugarscape Model
  - Agent seek wealth to maximize the wealth savings
  - Grid provides its wealth to the agents located at the grid
  - Agent consumes the wealth

Parameters	Parameter Type	Parameter Range	Synthetic Parameter Setting	
			Value	Time or Cluster
Wealth Income	Dynamic	0-2	1.5	1-10,21-30,41-50
			0.5	11-20,31-40
Wealth Consumption	Heterogeneous	0-1	0.9	Top 50% in Initial Wealth
			0.1	Bottom 50% in Initial Wealth

Type of Summary Statistics	Name of Summary Statistics	Variable Description
Validation Summary Statistics	High Class Wealth Average	Average wealth of top 1/3 agents
	Middle Class Wealth Average	Average wealth of middle 1/3 agents
	Low Class Wealth Average	Average wealth of bottom 1/3 agents
	Gini Index	The area ratio of the Lorenz curve to measure the wealth inequality

KAIST
Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

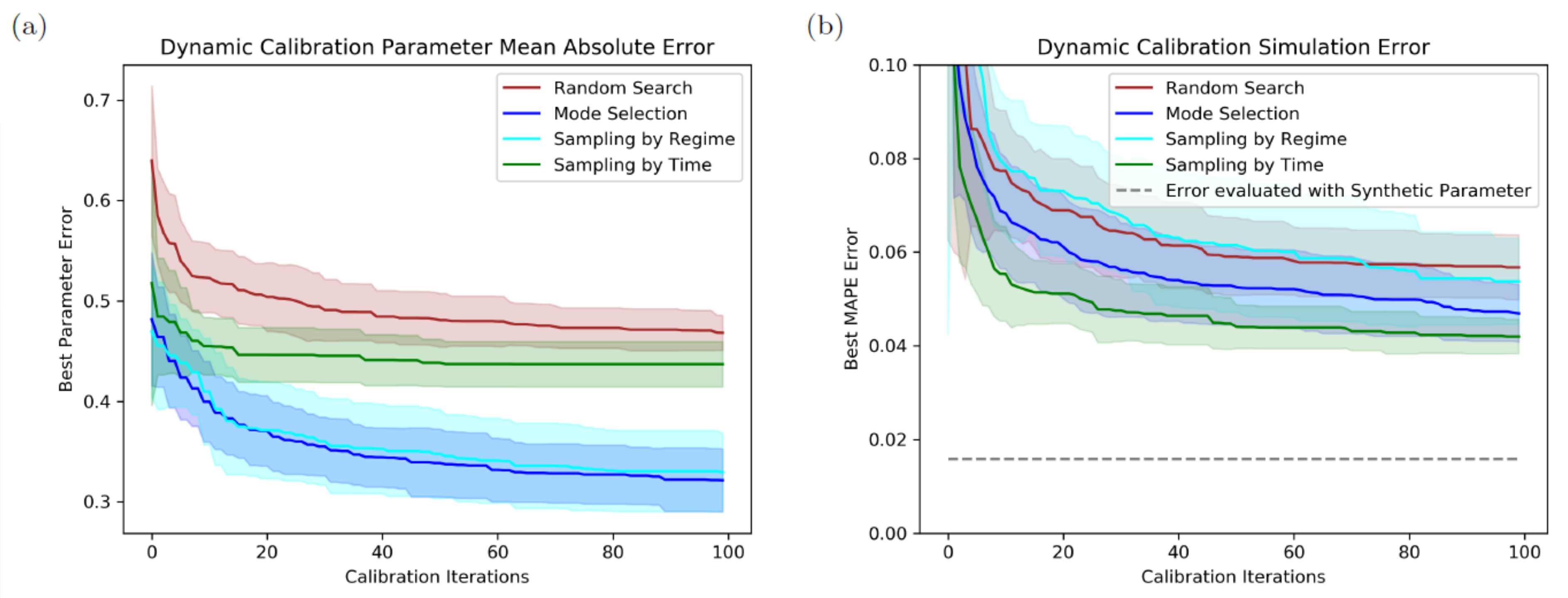
Ablation Study – Calibrated Dynamic Parameter



- Dynamic calibration finds a dynamic parameter by regime to avoid overfitting.
  - Red line is the synthetic parameter
  - Blue line is the estimated dynamic parameter.
- Random Search finds an overfitted dynamic parameter, which only fits the given validation data without matching with the given synthetic parameter.

Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

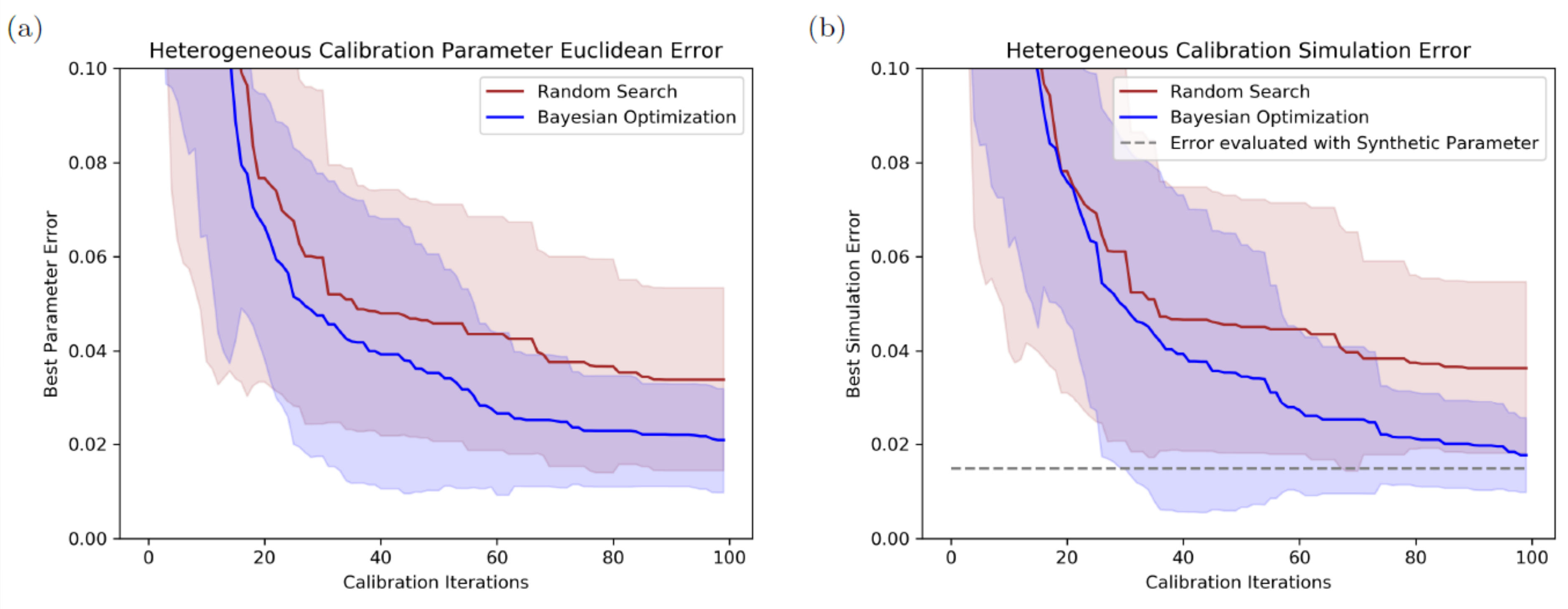
Ablation Study – Error Plots of Dynamic Calibration



- (a) Parameter Mean Absolute Error of dynamic calibration
  - Parameter generation methods Sampling by Regime and Mode Selection find parameters closer to the synthetic parameter than the other methods.
- (b) Dynamic calibration simulation error
  - Parameter generation method Sampling by Time performs the best in terms of simulation MAPE.

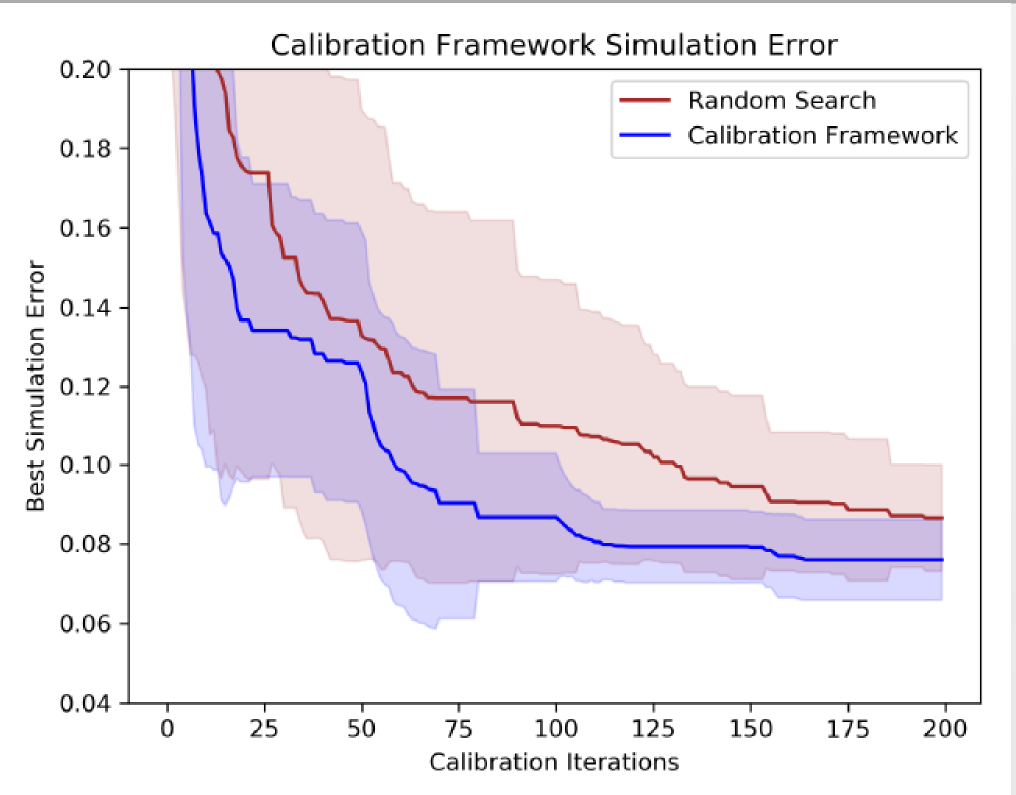
Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Ablation Study – Calibrated Heterogeneous Parameter



- (a) Parameter Euclidean Error of Heterogeneous calibration
- (b) Heterogeneous calibration simulation MAPE
  - Suggested Bayesian optimization converges to the optimal lower bound, which is not 0 because of the stochasticity.

Calibration Framework Simulation Error



- The suggested calibration framework simulation MAPE
  - Four cycles of the calibration framework is replicated
    - Each cycle includes
      - Dynamic calibration for the first 20 iterations
      - Heterogeneous calibration for the next 30 iterations.

Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Test Case 2

- [Model Description] Real Estate Market Agent-based Model
  - Agent buy and sell/lease either house/apartment/condo
  - House price is increased when the same typed houses are popular
  - House price is decreased if the house is not sold, while the house is listed in the housing market
- Housing Transaction Number
  - Market Participation Rate
  - Willing to Pay
  - Purchase Rate
- Housing Price
  - Market Price Increase Rate
  - Market Price Decrease Rate

[Dynamic] Demand change due to the up and down of the economic trends is modeled in the dynamic parameter

[Heterogeneity] Household investment portfolio is modeled in the heterogeneous parameter

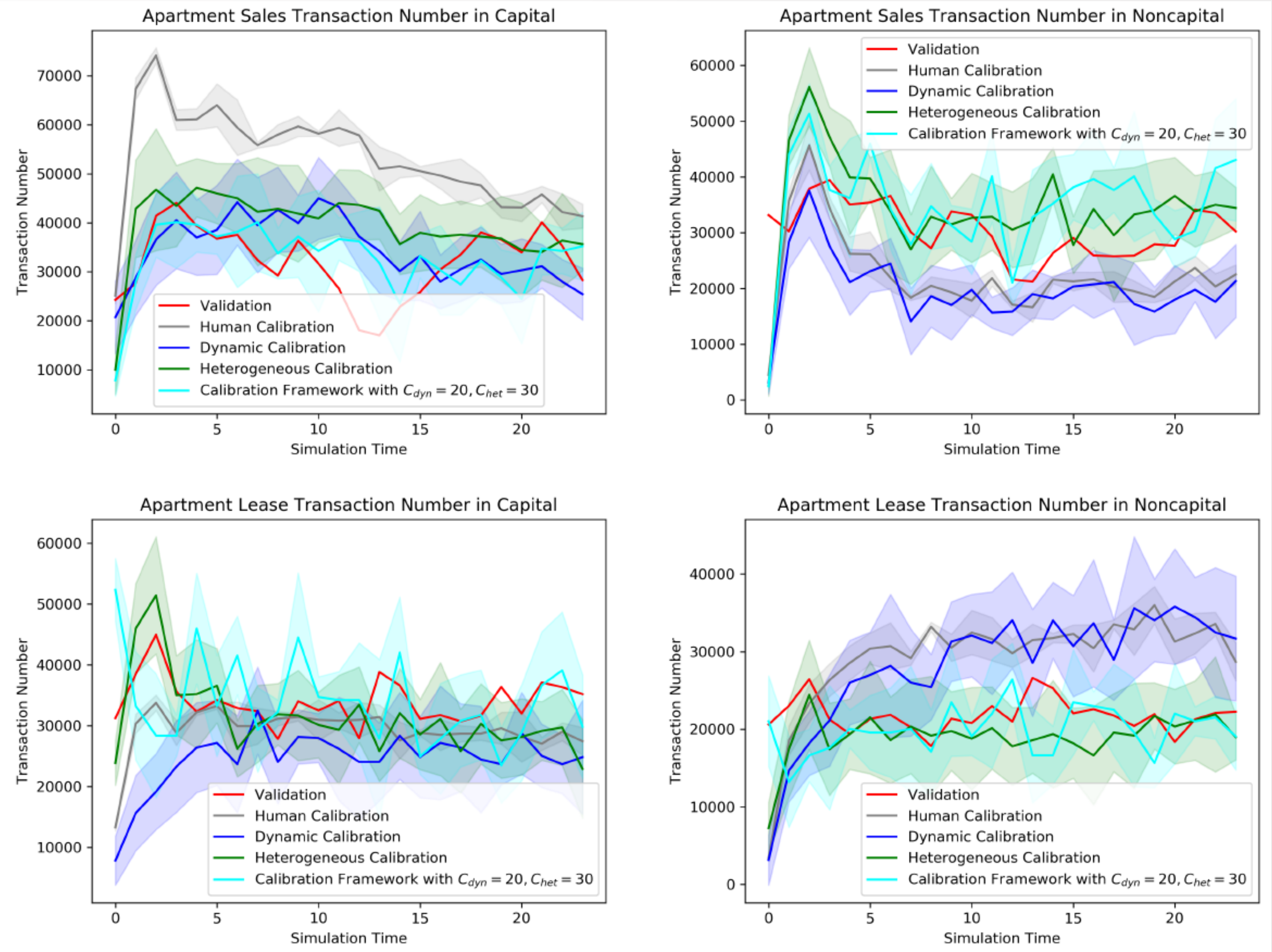
Parameter	Parameter Type	Parameter Range
Market Participation Rate	Dynamic	0-0.05
Market Price Increase Rate	Dynamic	0-0.1
Market Price Decrease Rate	Dynamic	0-0.1
Willing to Pay	Heterogeneous	0.3-0.9
Purchase Rate	Heterogeneous	0.3-0.9

Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Summary Statistics

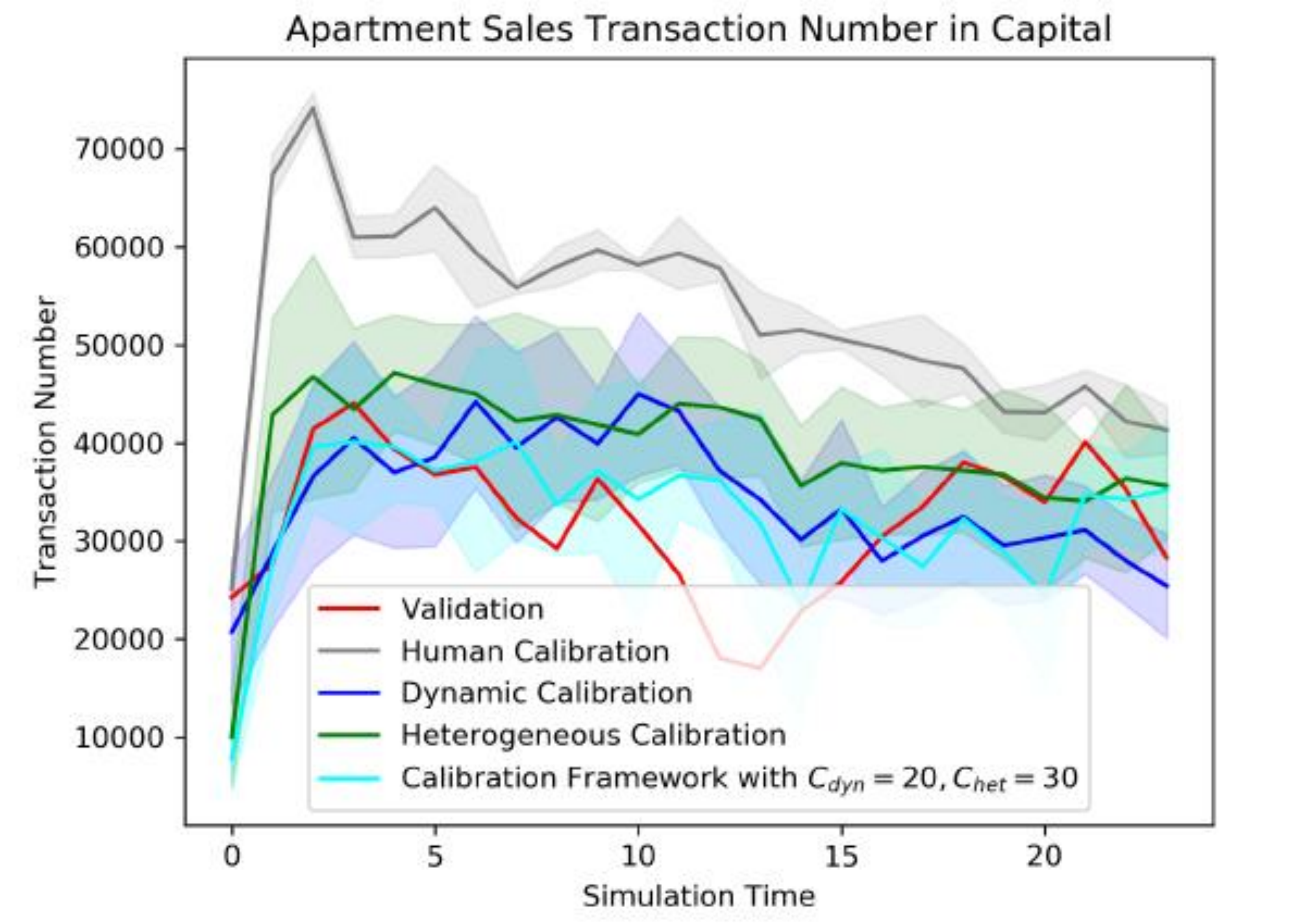
Types of Summary Statistics	Name of Summary Statistics	Variable Description	Variable Value
Validation-level Summary Statistics	Apartment Sales Price Index in Capital	Jevons price index of Apartment sales price	Housing Price is converted into a percentage, with base value as 100 at the initial timestep.
	Apartment Sales Price Index in Noncapital		
	Apartment Lease Price Index in Capital	Jevons price index of Apartment lease price.	
	Apartment Lease Price Index in Noncapital		
	Apartment Sales Transaction Number in Capital	Transaction numbers of Apartment sales.	Simulation transaction number is scaled up to be compatible with the validation transaction number.
	Apartment Sales Transaction Number in Noncapital		
	Apartment Lease Transaction Number in Capital	Transaction numbers of Apartment lease.	
	Apartment Lease Transaction Number in Noncapital		
Agent-level Summary Statistics	Living Region	Agent living region between capital/noncapital area.	1: Capital, 0: Noncapital
	Savings	Total savings.	1 unit/1000 KRW
	Income	Sum of the labor income and transfer income.	1 unit/1000 KRW
	Loan	Total amount of money agent have borrowed from bank.	1 unit/1000 KRW
	House Type	Type of house where an agent lives.	1: Detached House, 2: Apartment, 3: Multiplex House, 4: No House
	Living Type	Type of living where an agent lives	1: Owner, 2: Lease, 3: No House
	Number of Own Houses	Number of houses agent owns	1 unit/1 House

Calibrated Apartment Transaction Numbers



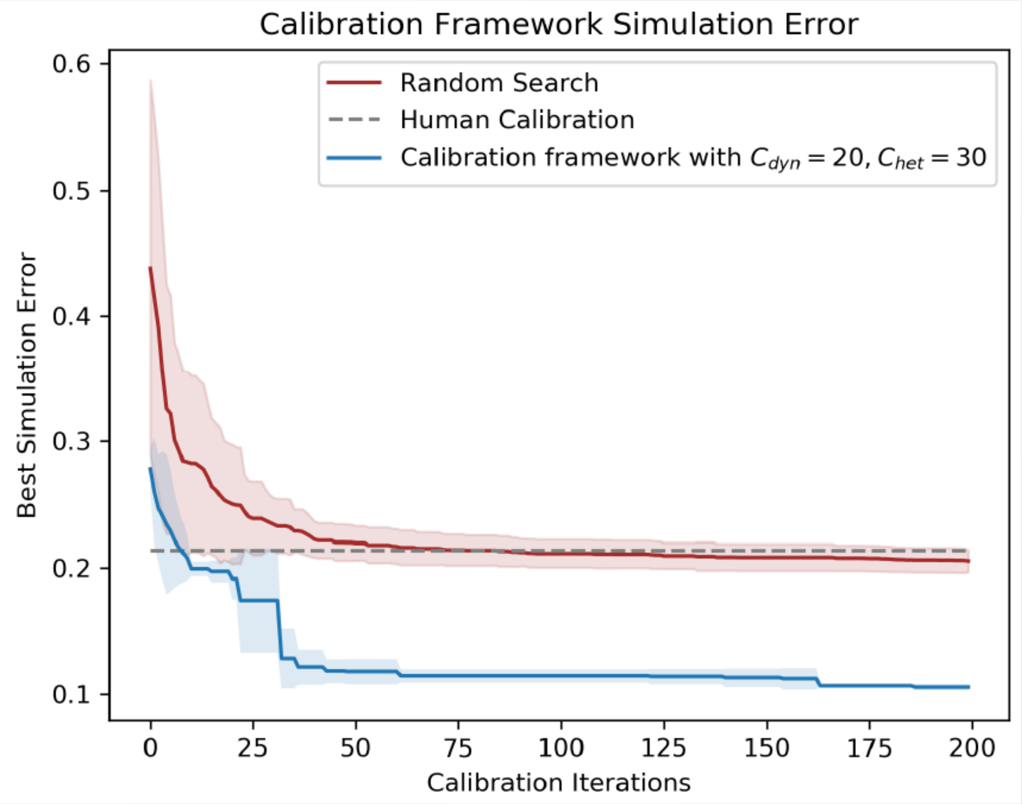
Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Calibrated Apartment Transaction Numbers



Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Calibration Framework Simulation Error

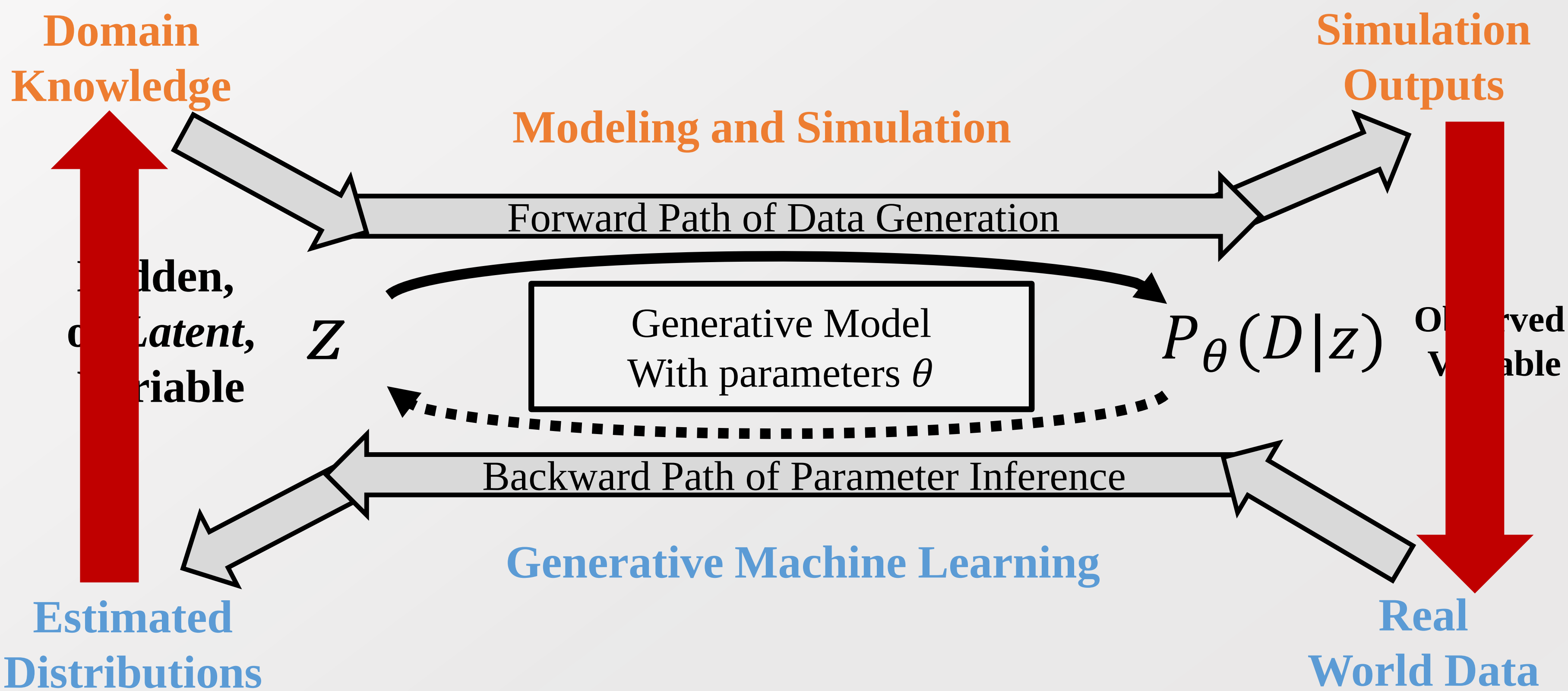


- Suggested calibration framework simulation MAPE
  - Blue line is a experimental result where the calibration framework is executed with $C_{dyn} = 20$ and $C_{het} = 30$.
  - Red line is a random search experimental result.
  - Dotted line is the human calibration result.

Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST

Lessons Learned

- We propose the new calibration framework, using Dynamic Calibration, and Heterogeneous Calibration.
  - Dynamic Calibration estimates an optimal set of dynamic parameters, using two components
  - Heterogeneous Calibration estimates an optimal set of heterogeneous parameters using three components



Copyright © 2018 by Il-Chul Moon, Dept. of Industrial and Systems Engineering, KAIST