Data-Driven Smart Buildings

Zhe Wang, PhD

Building Technology and Urban Systems Division

Smart Buildings

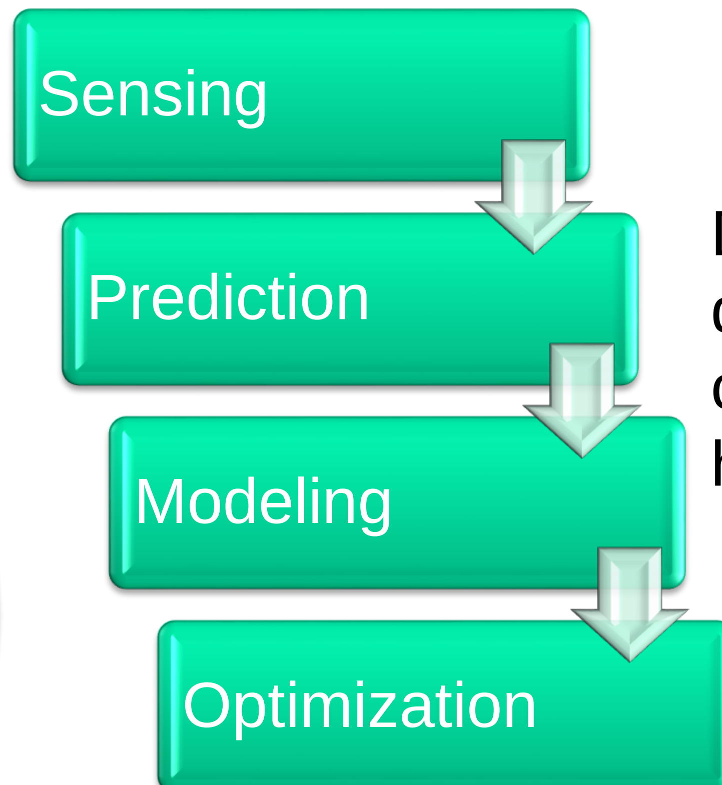
What?

- Grid interactive
  - Flexible and resilient
  - Provide grid services
- Occupant responsive
  - Human building interaction
  - Respond to individual demand
  - Minimize unnecessary waste

Smart Buildings

What?

How?



Data-driven could help

Contents

- **Previous work**
  - Sensing
  - Prediction
  - Modelling
- **Thoughts and future plans**
  - Data-driven vs. physics-based
  - How

Contents

- **Previous work**
  - Sensing
  - Prediction
  - Modelling
- **Thoughts and future plans**
  - Data-driven vs. physics-based
  - How

Sensing

You cannot manage what you cannot measure

-- Peter Drucker

Sensing

- **Conventional building**
  - Sensing *physical parameter* only

→

- **Smart building**
  - *Occupant related* sensing

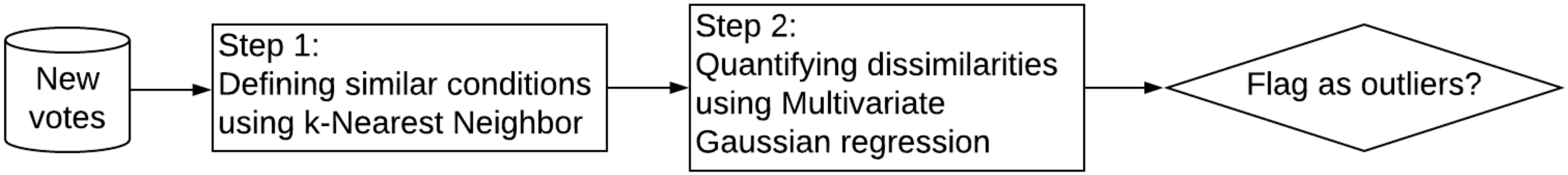
- **Our work**
  - Enhance sensing accuracy, cost effectiveness, scalability, and address other concerns (e.g. privacy)

Outlier Detection

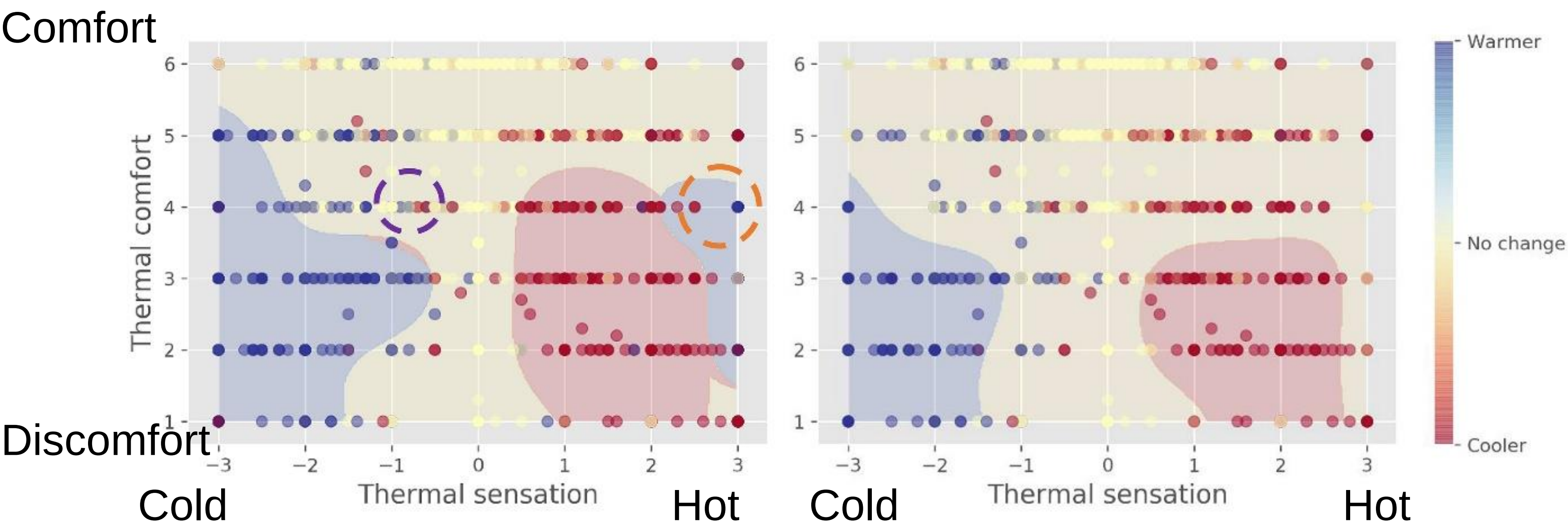
- **Motivation**
  - *Sensing occupant response*
  - Outliers in subjective comfort vote undermine *accuracy*
  - Lack of research on detecting *outliers in subjective vote data*
- **Goal**
  - Proposes an outlier detection framework to automatically flag potential outliers in subjective thermal comfort votes
  - Key challenge: individual difference vs. outliers

Outlier Detection

- **How?**



  - Test it on ASHRAE thermal comfort database



(a) Thermal preference model with outliers

(b) Thermal preference model without outliers

Outlier Detection

- **Contribution**
  - We proposed a stochastic two-step framework
    - Users could tune contextual attributes, distance measures
  - Could be used for real-time *occupant responsive control*
    - Computationally efficient
    - Active, online learning

Wang, Z., Parkinson, T., Li, P., Lin, B. and Hong, T., 2019. The Squeaky wheel: Machine learning for anomaly detection in subjective thermal comfort votes. Building and Environment, 151, pp.219-227.

Inferring Occupant Counts

- **Motivation**
  - Sensing occupant counts
  - Current occupant sensing technologies are *expensive* or *labor-intensive*
- **Goal**
  - Propose a new approach to detect occupant counts through Wi-Fi, which is *non-intrusive, cost-effective*
  - Challenge: accuracy vs. privacy

Inferring Occupant Counts

- **How?**
  - *Feature*
    - Key idea: Cluster the devices based on connection time/duration
    - The clustering could be done locally with a simple script (*edge computing*)

Time	Shuffled Device_ID	AP_ID
...		
20180521_0000	dfd6bafb68c1cd1f1e2d9190ca9d55f0	ap135-4206w
20180521_0000	e6c1fe930c6d2c2f2e2d9d69fc0abeda	ap135-3103
...		
20180521_0000	dd464552ecc1208e94a955bffee1f749	ap135-4110
20180521_0010	dfd6bafb68c1cd1f1e2d9190ca9d55f0	ap135-4206w
20180521_0010	e6c1fe930c6d2c2f2e2d9d69fc0abeda	ap135-3103
...		

(a) Raw data collected

Time	Target zone	Device_type	Device_count
...			
20180521_0000	Zone 1	Short term (less than 1h per day)	0
...			
20180521_0000	Zone 1	Long term (more than 12h per day)	20
20180521_0000	Zone 2	Short term (less than 1h per day)	0
...			
20180521_0000	Zone 2	Long term (more than 12h per day)	15
20180521_0010	Zone 1	Short term (less than 1h per day)	0
20180521_0010	Zone 1	Long term (more than 12h per day)	21
...			

(b) data input to the machine learning algorithm

Inferring Occupant Counts

- **How?**
  - *Algorithm*
    - Random Forest outperforms the other two
    - The sequential information does not really help

	Random Forest (RF)	Neural Network (NN)	LSTM
RMSE on the training set	1.20	2.63	2.21
RMSE on the testing set	3.95	4.62	4.52
Computation time[a]	2.38s	24.86s	65.61s

Wang, Z., Hong, T., Piette, M.A. and Pritoni, M., 2019. Inferring occupant counts from Wi-Fi data in buildings through machine learning. Building and Environment, 158, pp.281-294.

Sensing: summary

- **Use *data-driven* method to**
  - *Sense occupant feedback*
    - *Accurate*
  - *Sense occupant counts*
    - *Accurate*
    - *Cost-effective*
    - *Scalable*
    - *Protect privacy*

Smart Buildings

What?

How?

Sensing

Prediction

Modeling

Optimization

Prediction

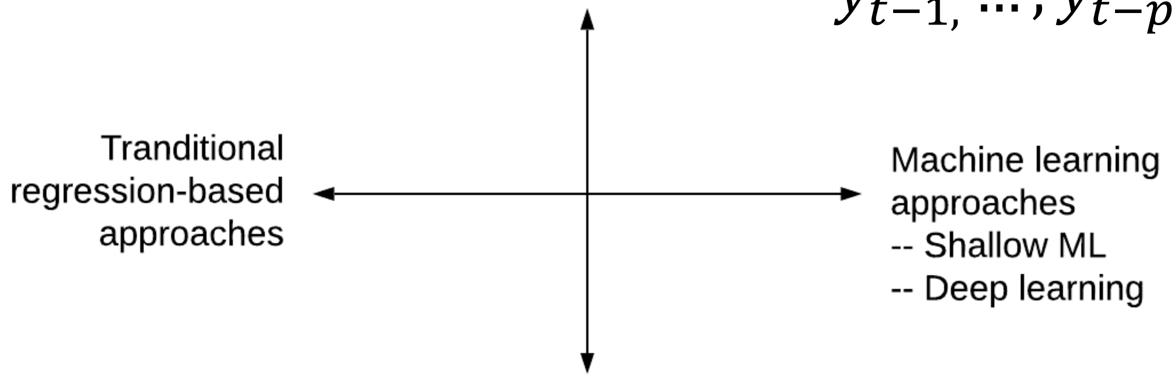
Management is prediction

-- Deming W.E.

Prediction

- **Conventional building**
  - *Schedule, fixed*
- **Smart building**
  - *Prediction, adaptive*
- **Our work**
  - Summarize and compare prediction methods

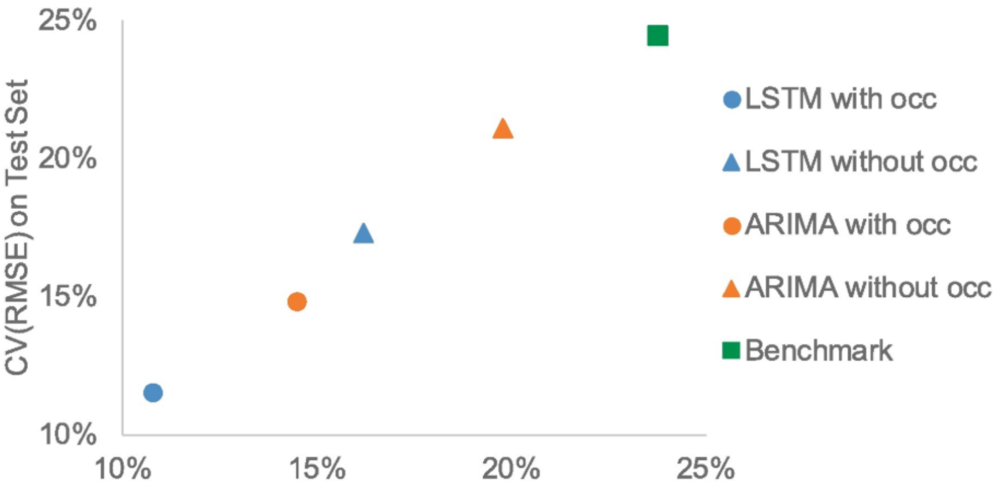
$$y_{t-1}, \dots, y_{t-p}, \boldsymbol{x_{t-1}}, \dots, \boldsymbol{x_{t-q}} \rightarrow y_t$$



$$y_{t-1}, \dots, y_{t-p} \rightarrow y_t$$

Prediction: Method

- **Case**
  - Plug load prediction
  - Comparison
    - Algorithm: *ARIMA* (Statistical) vs. *LSTM* (Machine Learning)
    - Additional feature: occupant count
  - Baseline: naïve persistent method
- **Finding**
  - *LSTM with occ* outperforms
- **Implication**
  - Select machine learning
  - Add relevant feature



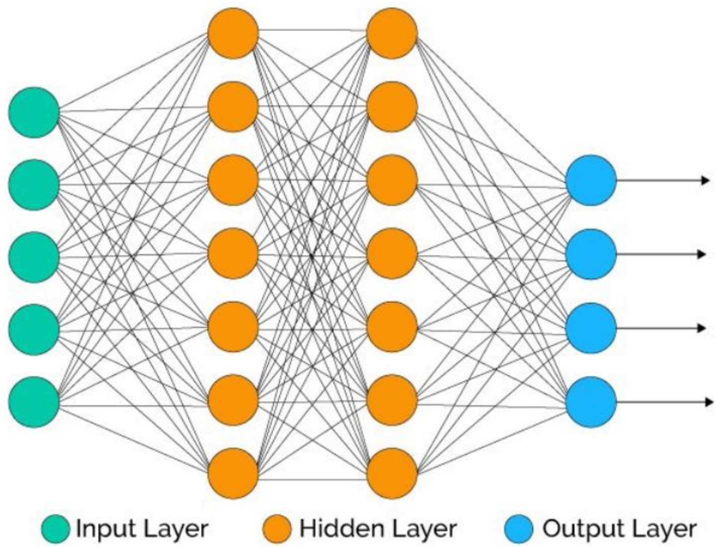
Wang, Z., Hong, T. and Piette, M.A., 2019. Predicting plug loads with occupant count data through a deep learning approach. *Energy*, 181, pp.29-42.

Prediction: Why ML outperforms

Regression based

$$y_t = \underbrace{\alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p}}_{\text{AR terms}} + \underbrace{\varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}}_{\text{MA terms}}$$

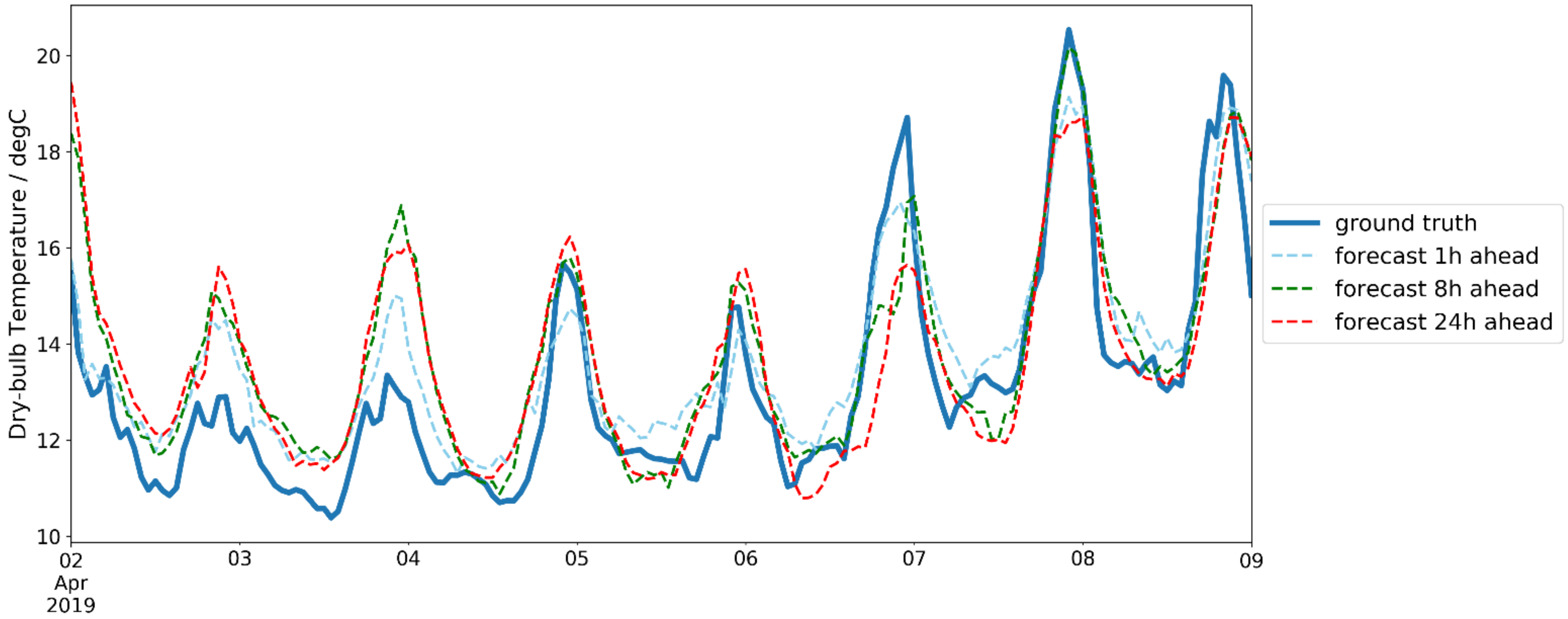
Neural network based



- Each neuron is a linear regression plus activation function
- Could capture *any form of relation* theoretically

Prediction under Uncertainty

- **Problems**
  - Adding relevant feature is helpful
  - In real-time prediction, the input feature is also predicted, which unavoidably has errors



Prediction under Uncertainty

- **Which approach is more robust to input uncertainty**

Statistical

ML

Winner!

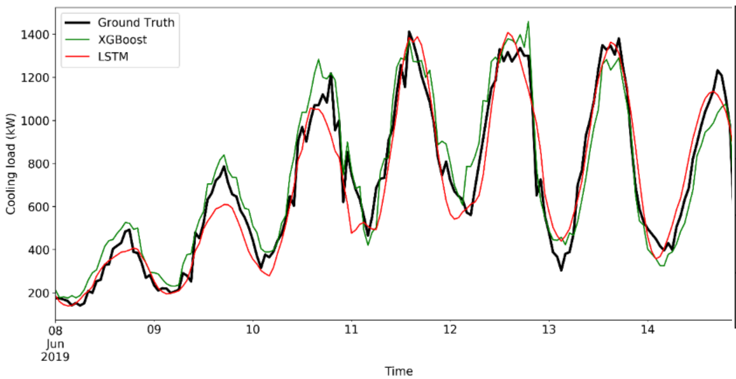
Shallow ML

VS

Deep learning

Prediction under Uncertainty

- **Case**
  - Building load prediction
  - Compare *XGBoost (shallow)* vs. *LSTM (deep)*
- **Finding**
  - Without input uncertainty: shallow model outperforms
  - With input uncertainty: deep model outperforms



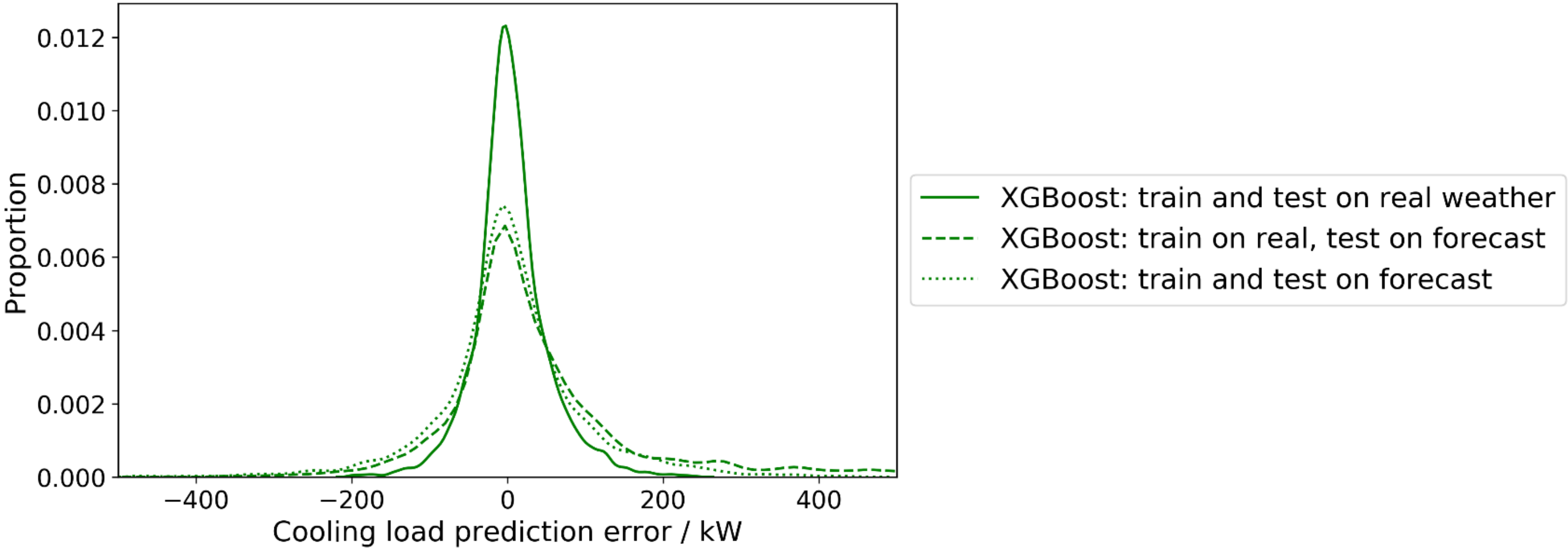
Prediction result

	Weather forecast uncertainty not considered	Weather forecast uncertainty considered	
		Train on real, test on forecast weather	Train on forecast, test on forecast weather
XGBoost	15.2%	27.7%	23.3%
LSTM	20.2%	21.2%	21.0%
Best Baseline	29.9%	29.9%	29.9%

Error: CV(RMSE)

Prediction under Uncertainty

- **Implications**
  - Uncertainty needs to be considered
  - Deep learning is recommended
  - The model is recommended to be trained using the forecasted (uncertain) weather data



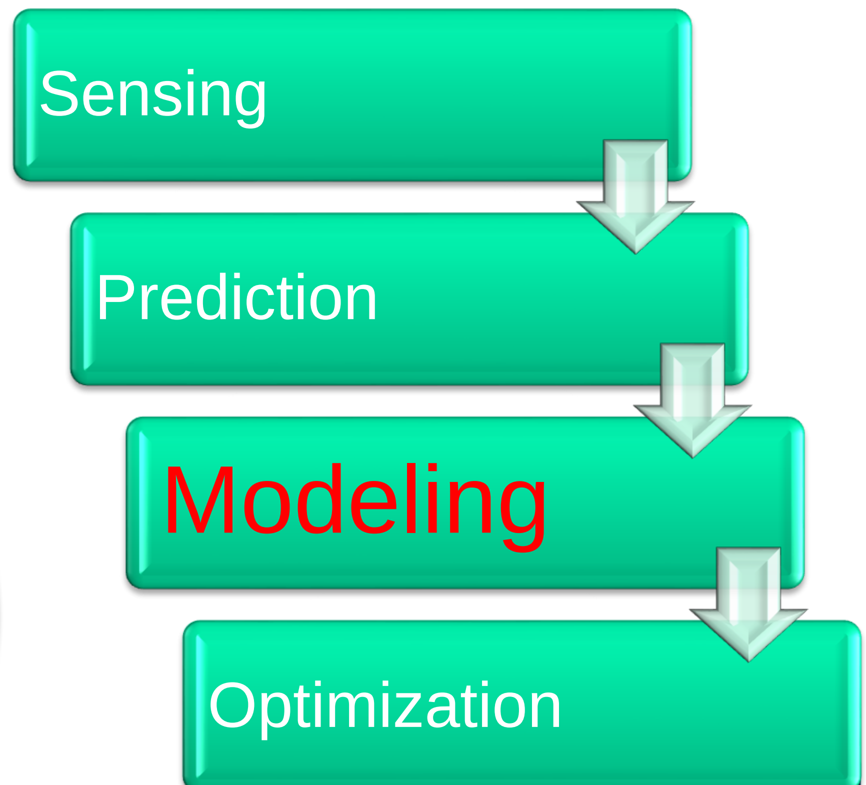
Prediction: Summary

- **Accuracy**
  - ML outperforms statistical approach
- **Uncertainty**
  - Deep model is more robust to input uncertainty
  - Expose model to uncertainty during training stage

Smart Buildings

What?

How?



Modelling

All models are wrong, but some are useful

-- George Box

Modeling

- **Conventional building**
  - *Feedback* control
- **Smart building**
  - *Model-based* *Feedforward* control
- **Problem**
  - Gap between model and reality
- **Our work**
  - Improve modelling accuracy: consider occupant behaviors
  - Data-driven approach

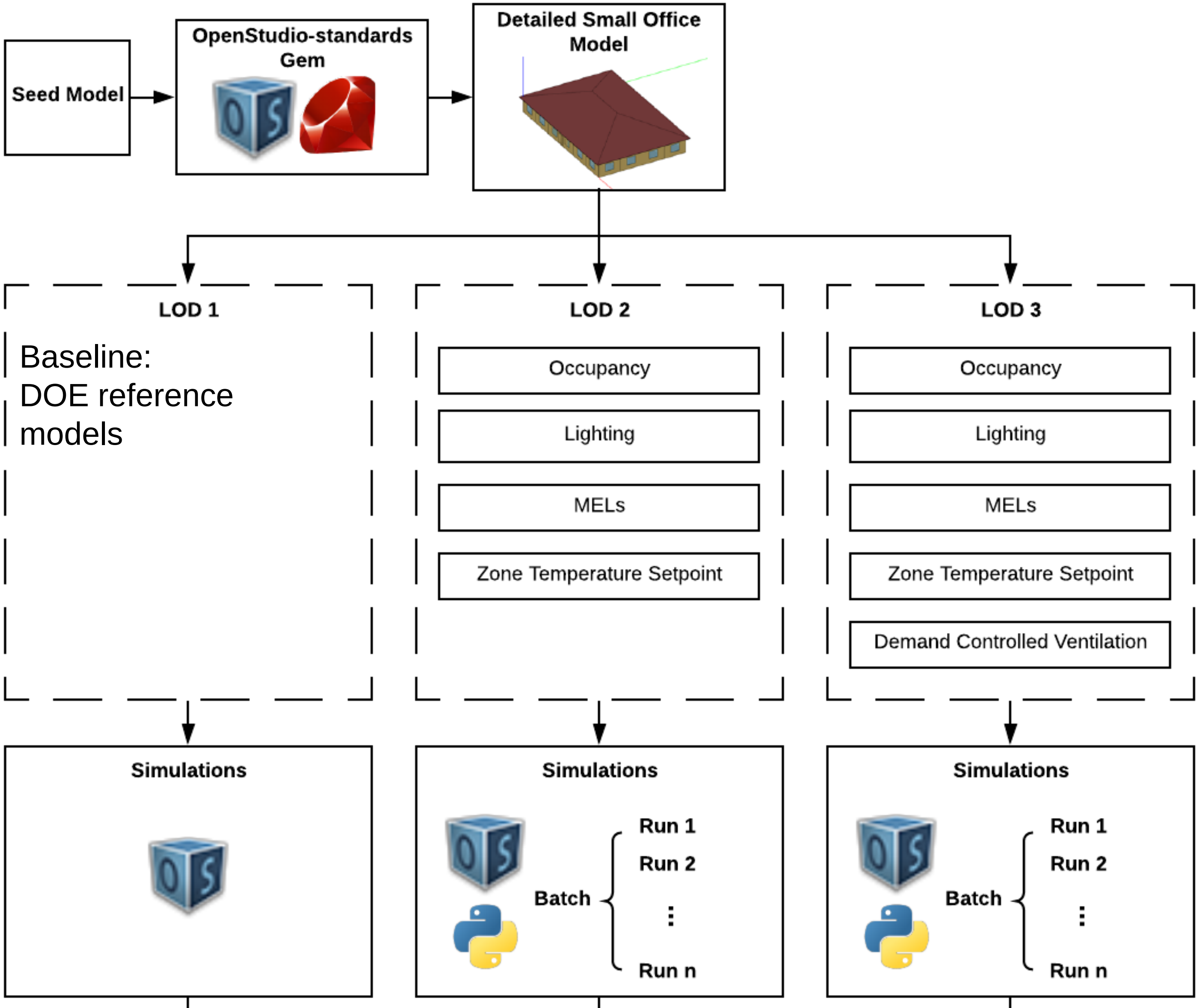
Modeling Occupant Behaviors (OB)

- **Motivation**
  - OB: major source for performance gap
    - Fixed schedule vs. dynamic, stochastic
  - Need new OB modelling tools
- **Tool developments**
  - Developed *Buildings.Occupants,* and open-sourced with *Modelica Buildings Library*
    - Simulate *occupancy*, *lighting*, *windows*, *blinds*, *heating* and *thermostat* behaviors in *office* and *residential* buildings
    - Include 34 models in the current version

- Wang, Z., Hong, T. and Jia, R., 2019. Buildings. Occupants: a Modelica package for modelling occupant behaviour in buildings. *Journal of Building Performance Simulation*, *12*(4), pp.433-444.

Modeling Occupant Behaviors

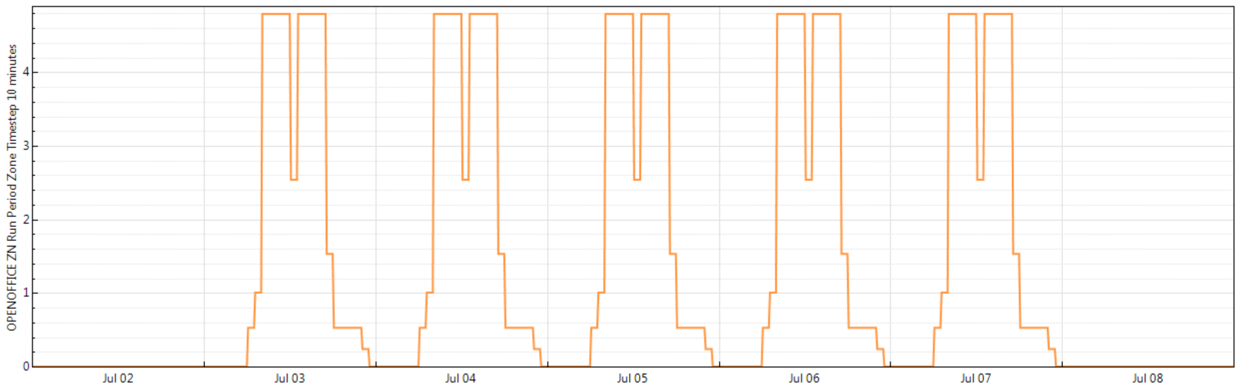
- **Application**
  - Realistic building load curve



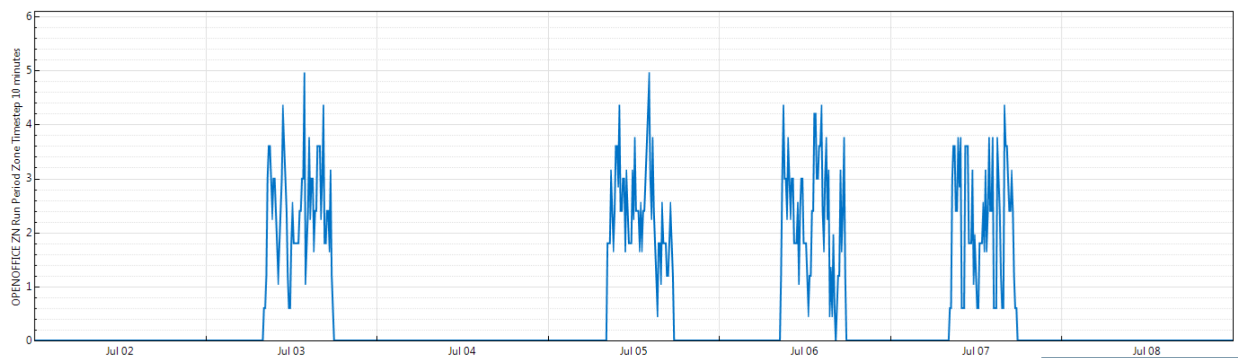
Modeling Occupant Behaviors

- **Application**
  - Realistic building load curve

DOE Reference model



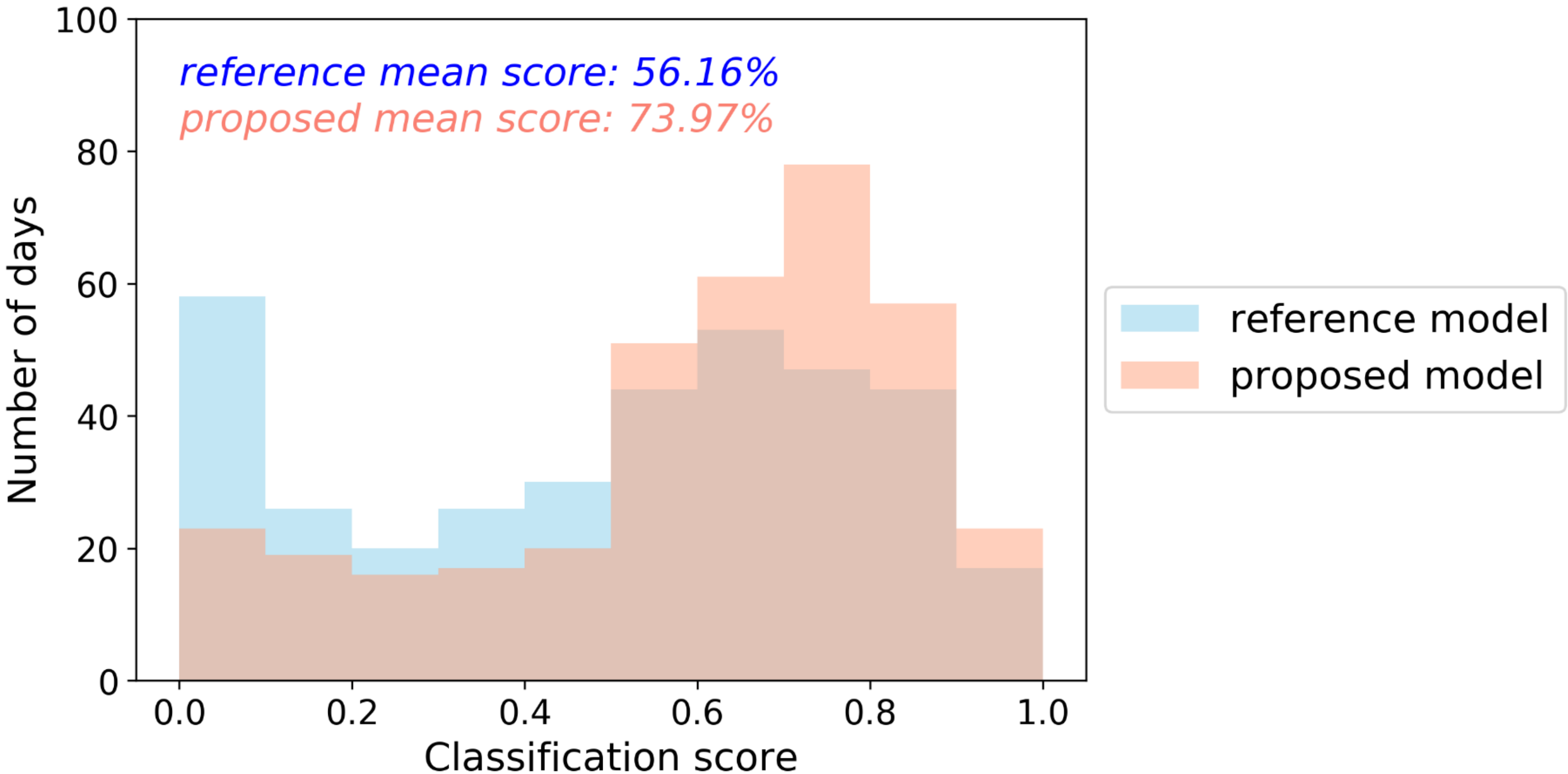
Our proposed model



Which one is more realistic?

Modeling Occupant Behaviors

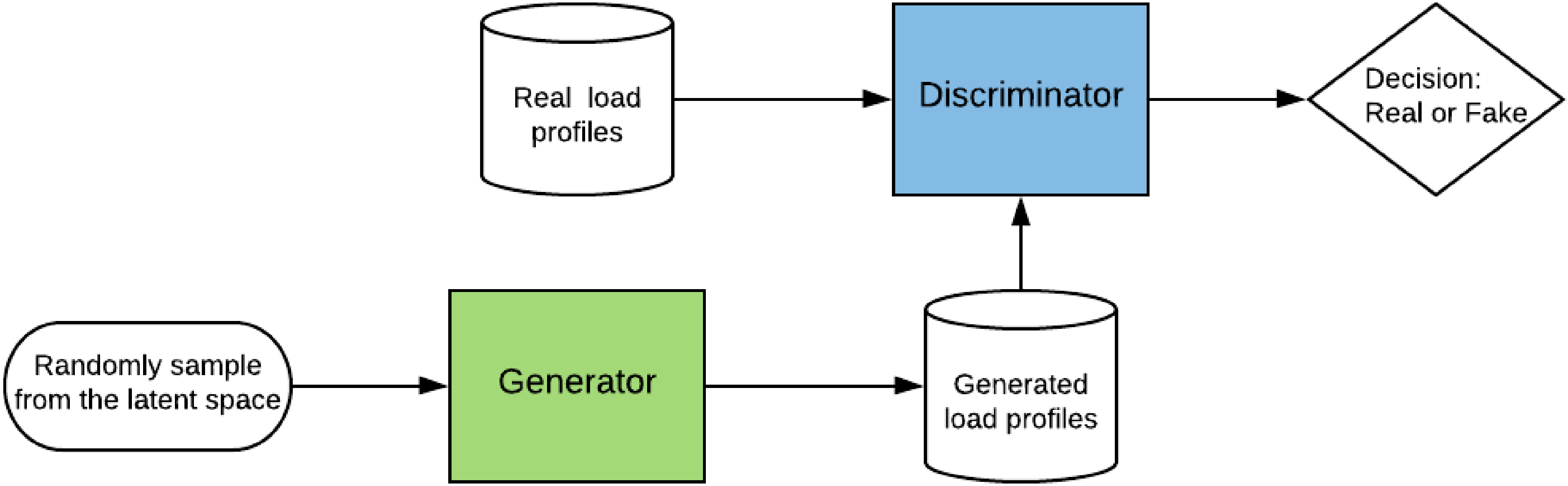
Results



- Classification score: output of the sigmoid function
  - Approaching to 1: more similar to the measured data
  - Approaching to 0: less similar to the measured data

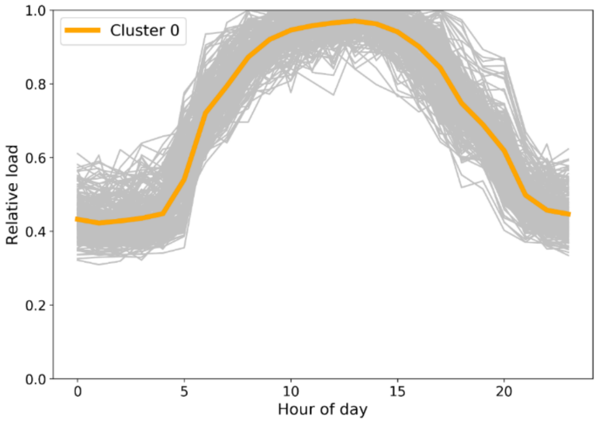
Data-driven Model-free Approach

- **Motivation**
  - Physics-based model: too many assumptions and inputs
- **Solution**
  - Explored data-driven approach
  - Generative Adversarial Network (GAN)

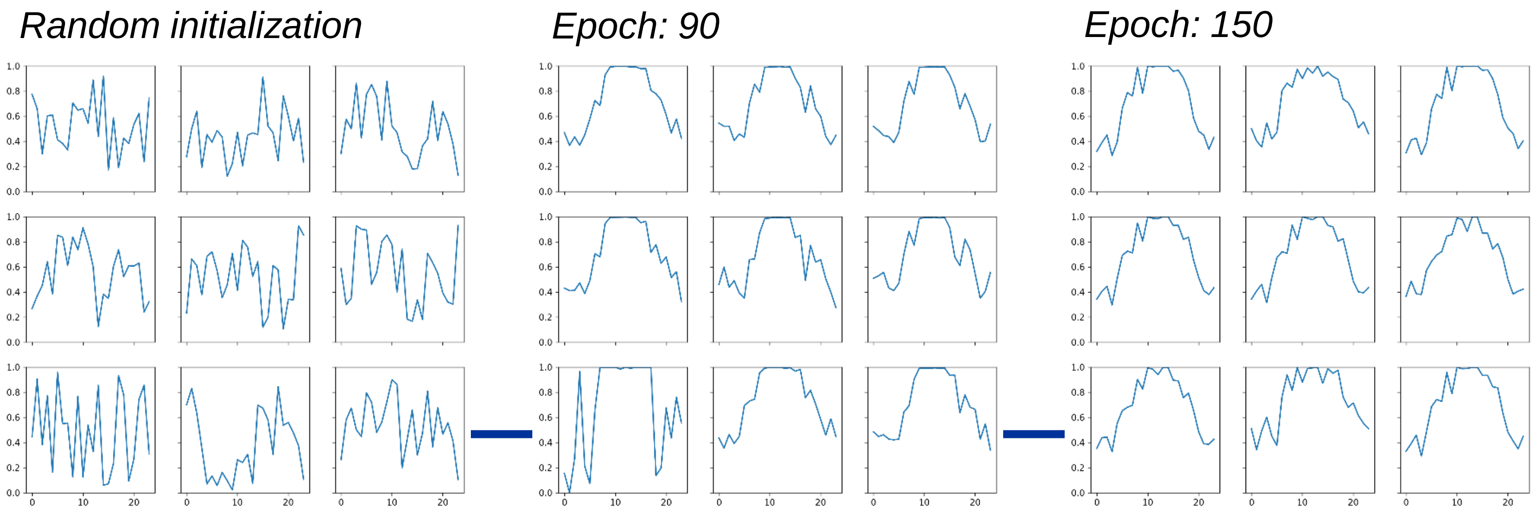


Data-driven Model-free Approach

- **Smart meter data**



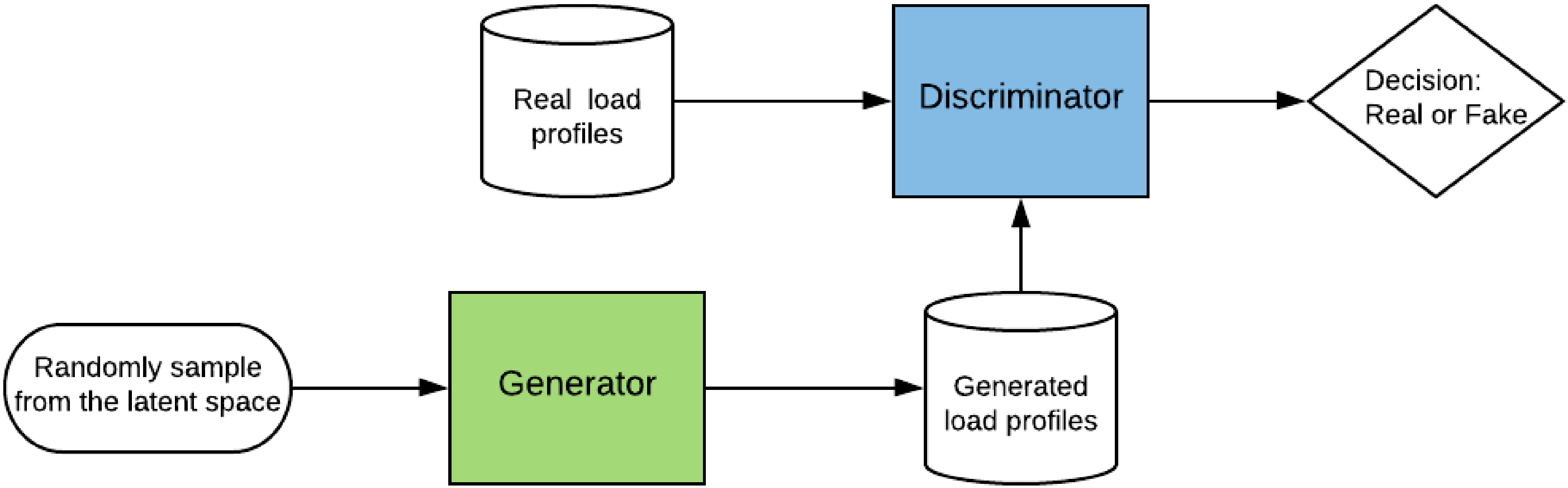
- **Generated by GAN**



Data-driven Model-free Approach

- **Application**

- **Validation**
- **Alert**



- **Anonymize**
- **Forecast**

Modelling: summary

- **Accuracy**
  - Modelling occupant behavior to reduce performance gap
- **Data-driven**
  - Use GAN to generate building load

Contents

- **Previous work**
  - Sensing
  - Prediction
  - Modeling
- **Thoughts and future plans**
  - Data-driven vs. physics-based
  - How

Physics-based

- **Pain-point**
  - Every building is unique

Case by case

VS

Photo courtesy: twincities.com

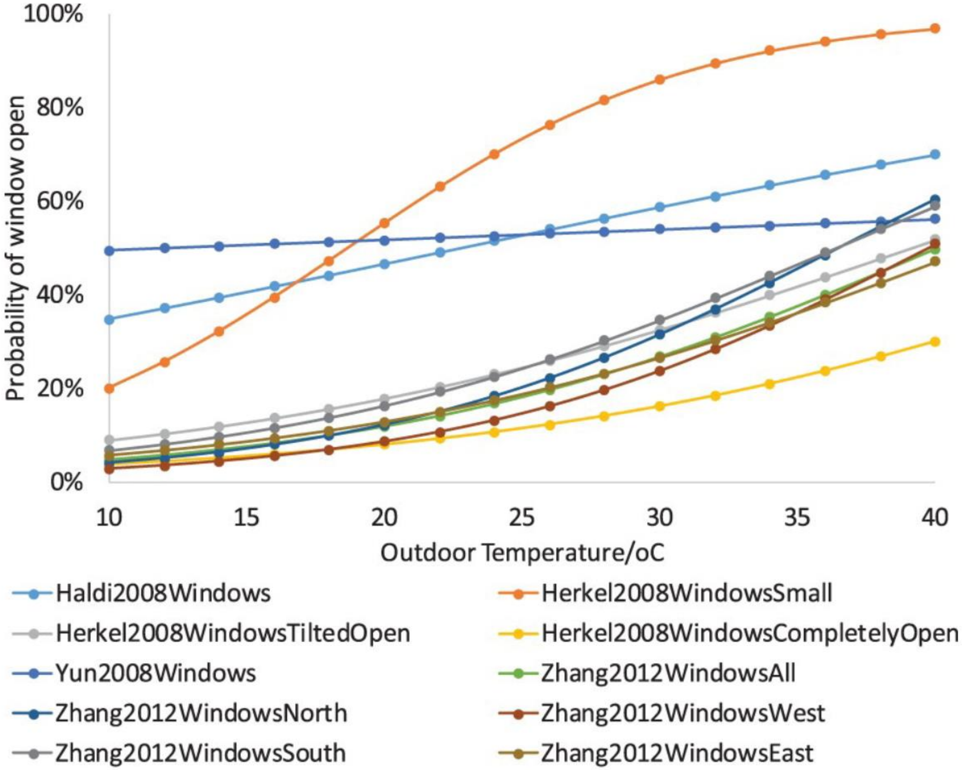
Photo courtesy: BusinessInsider

Standardized mass production

Physics-based

- **Pain-point**
  - Every building is unique
    - A detailed model might not make economic sense
  - Make simplifications
    - Difficult assumptions
    - Uncertainty/error propagation
  - Interact with non-deterministic humans

Source: Wang et al., 2019



Data-driven

- **Architecture**
  - Reflect the nature of the problem/building
    - Customize the *structure*, encoding the nature of physics
    - Customize the *objective function*, based on the problem to be solved
- **Parameters**
  - Learned from the data
    - No assumptions are needed
    - Let the data speak

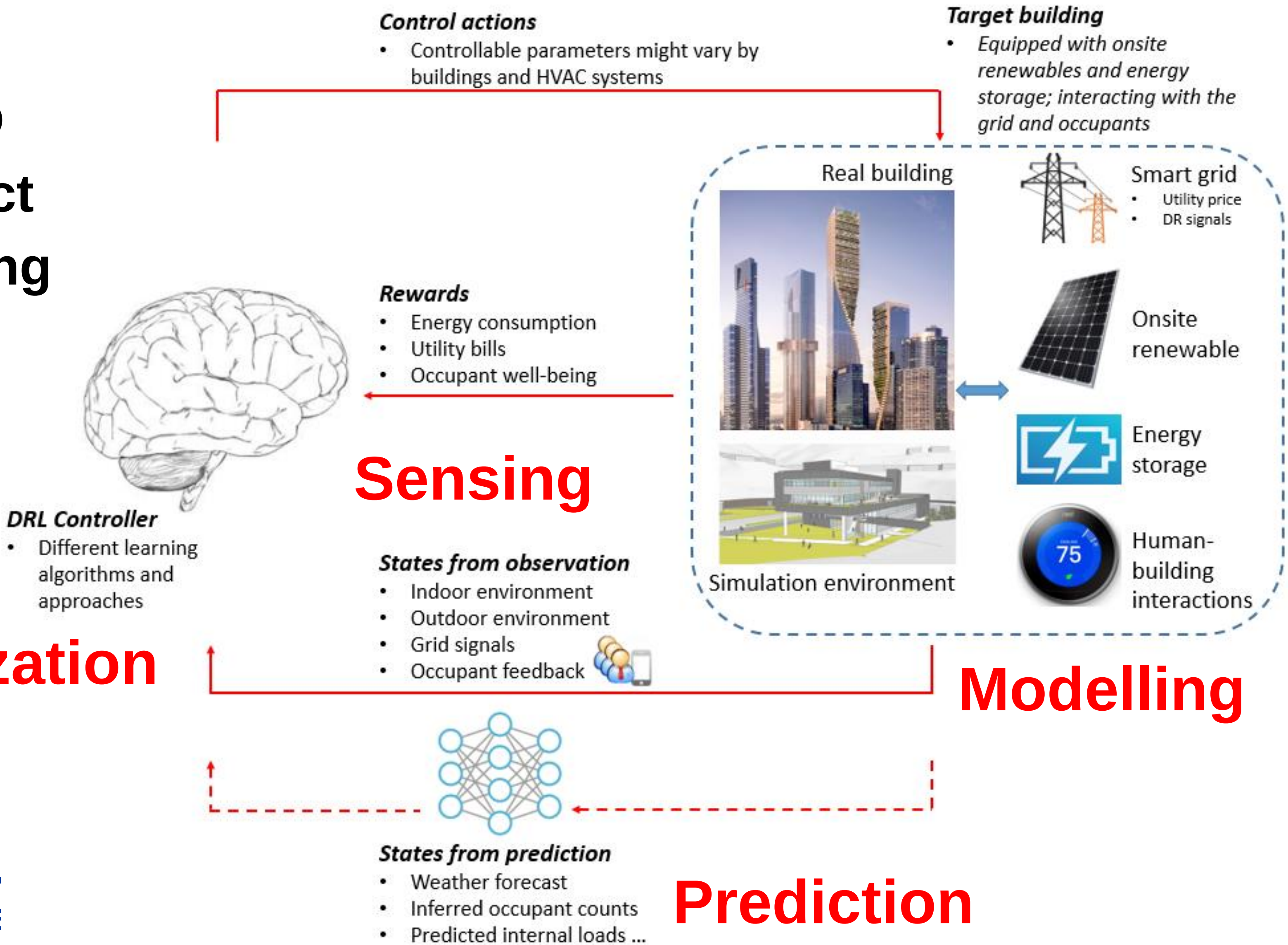
Data-driven

- **Benefits and improvements**
  - More *flexible model architecture*
  - More *powerful parameter identification*
- **Has improved sensing, prediction, and modelling**
- **What’s next?**

Plan

- **Occupant responsive**
- **Grid interactive**

- **LDRD project starting soon**



Sensing

Optimization

Modelling

Prediction

Thanks for your attention

About myself

- Education

2011	Bachelor	Civil Engineering	Tsinghua
2014	Master	Energy Technology	University of Cambridge
2017	Ph.D.	Civil Engineering	Tsinghua

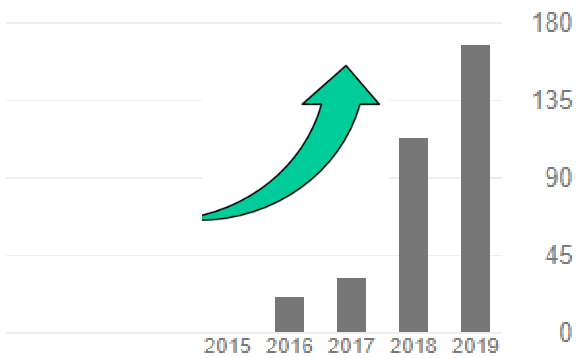
- Working

2016-2018	Energy consultant	World Bank
2017-2018	Postdoc	UC Berkeley
2018-2019	Postdoc	LBNL

Google Scholar

Cited by

	All	Since 2014
Citations	335	335
h-index	11	11
i10-index	11	11



BERKELEY LAB