Data-driven Prediction of Occupant Presence and Lighting Power: A Case Study for Small Commercial Buildings

Jing Wang[1], Wangda Zuo[1], Sen Huang[2], Draguna Vrabie[2]

[1]Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, USA

[2]Pacific Northwest National Laboratory, USA

University of Colorado Boulder

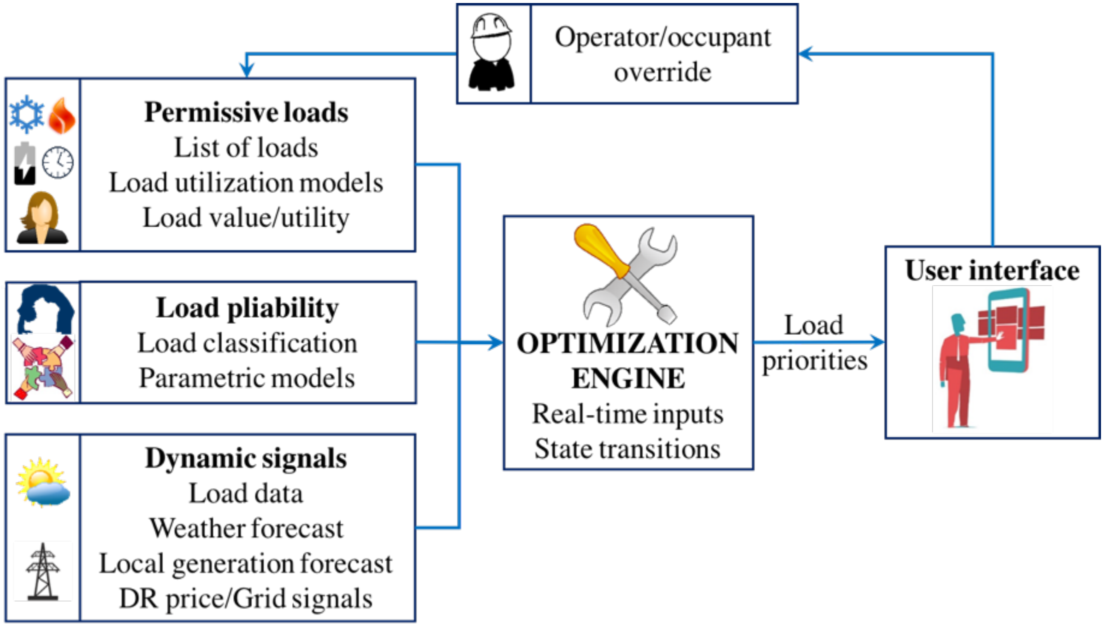
SBS LAB Sustainable Buildings and Societies Laboratory

Pacific Northwest NATIONAL LABORATORY

9/22/2020

Acknowledgement

C3PO: Comprehensive Pliant Permissive Priority Optimization (10/18-9/20), Department of Energy, collaboration with PNNL and ORNL.



Framework to dynamically value and classify building loads

Our Team

University of Colorado Boulder

Dr. Wangda Zuo
Jing Wang

Pacific Northwest National Laboratory

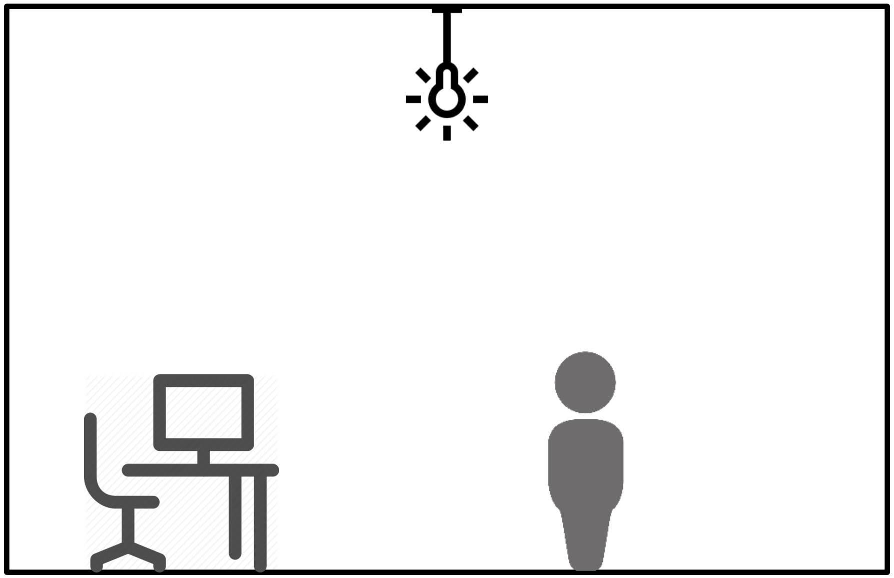
Dr. Draguna Vrabie
Dr. Sen Huang

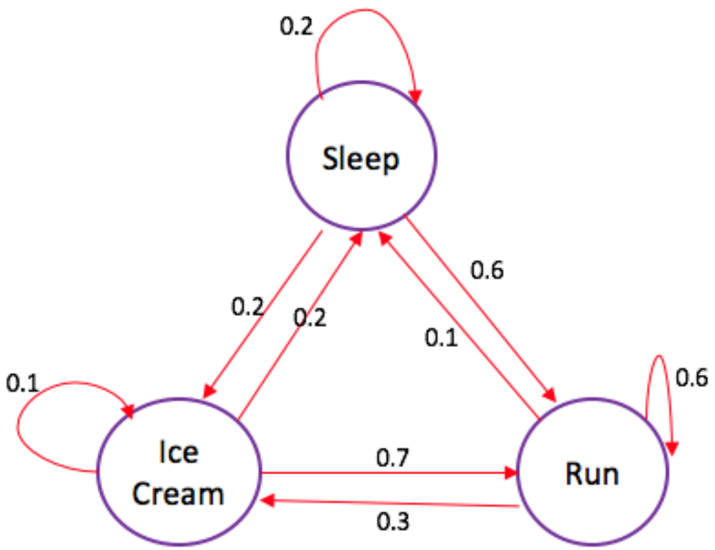
Oak Ridge National Laboratory

Dr. Piljae Im
Dr. Yeonjin Bae
Dr. Jian Sun

Challenges

- ❑ Uncertainty in building load prediction
- ❑ Occupant behavior stochasticity
- ❑ Static hourly schedules in building energy simulation tools
- ❑ Occupant sensor data often unavailable

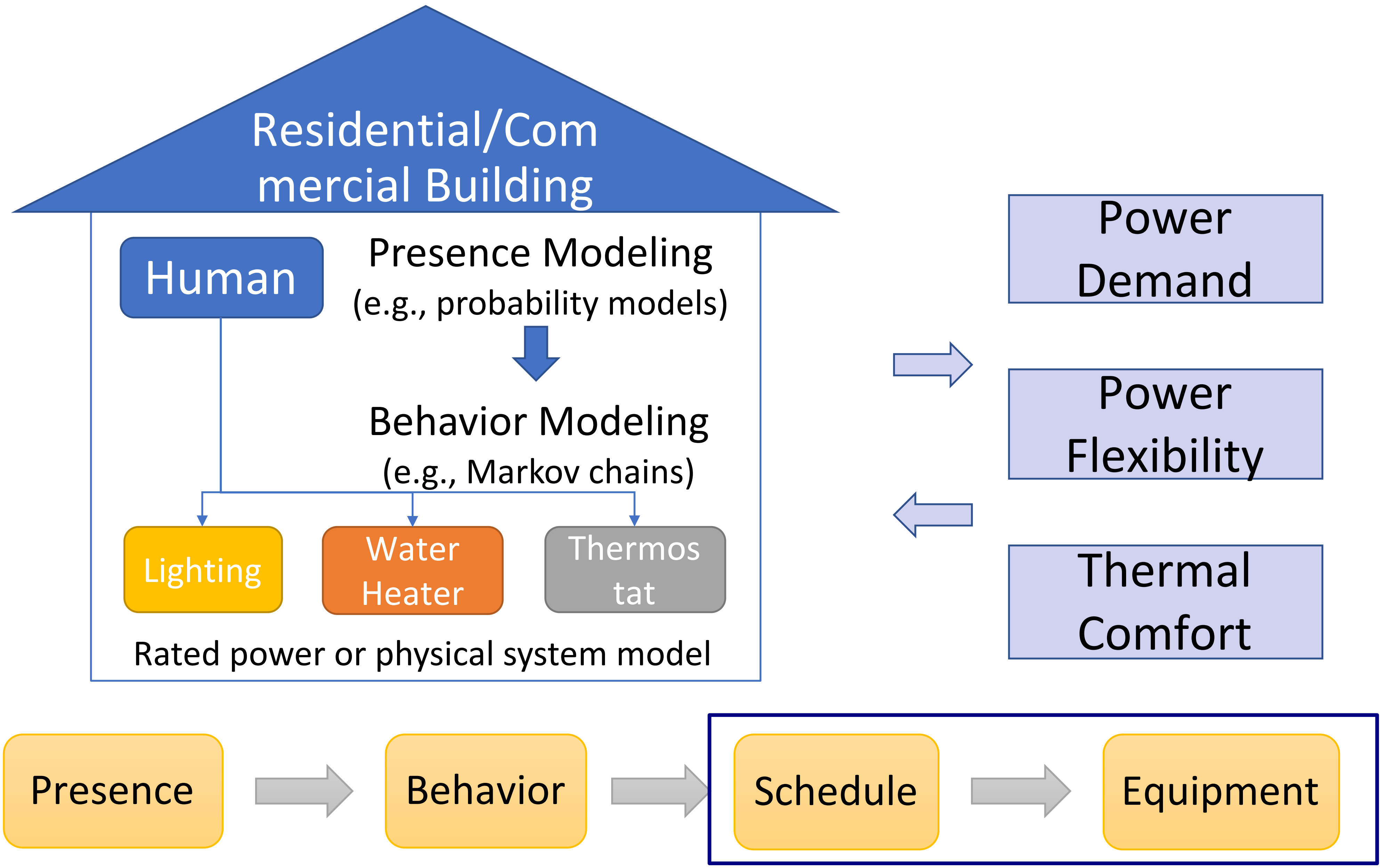




Research Question

How to predict building occupancy and power demand on a sub-hourly basis without occupant sensor data ?

Occupant Behavior Modeling

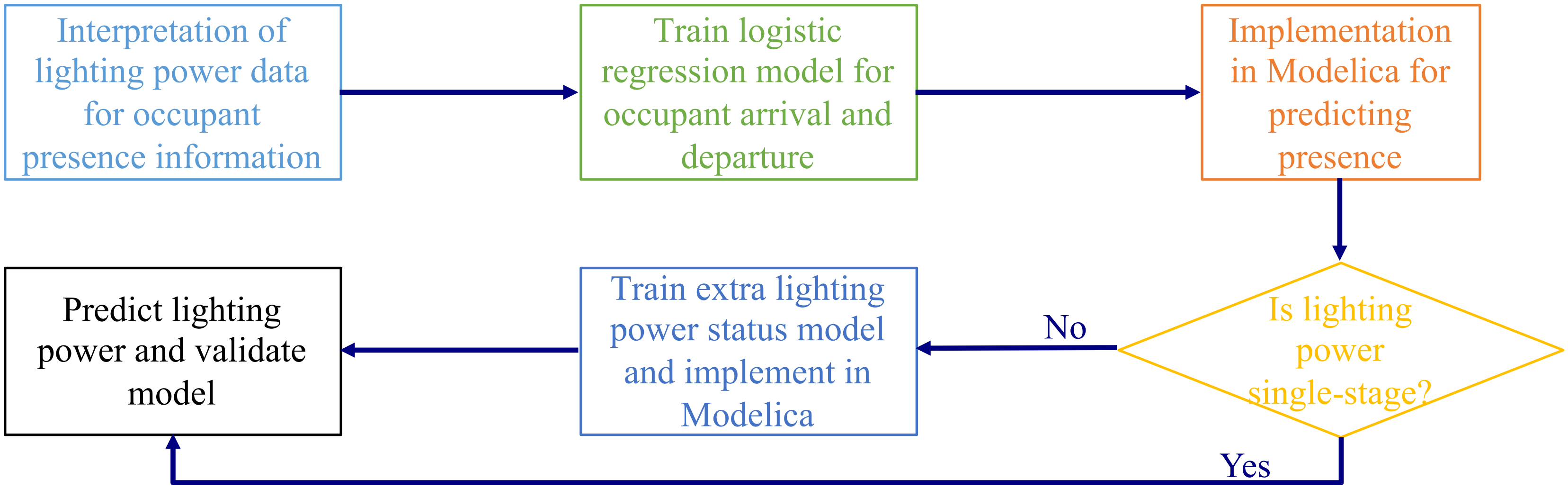


Methodology

❑ Correlation between occupant presence and light switching

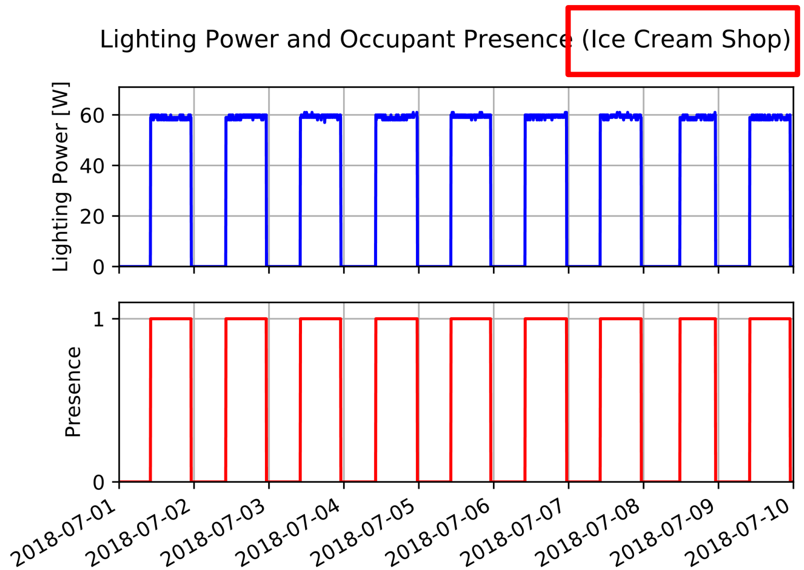
Finding	Reference
Switching mainly takes place when entering or vacating a space.	Hunt 1979
The switch-on probability on arrival exhibits a strong correlation with minimum daylighting illuminance in the working area.	Hunt 1979
The manual switch-off probability of the lights strongly relates to the expected length of absence.	Pigg 1998

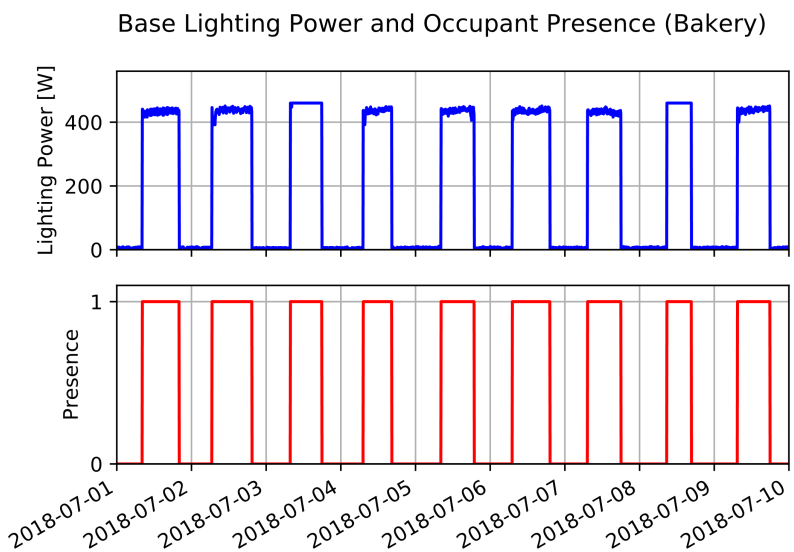
❑ Extracting presence information from lighting power data

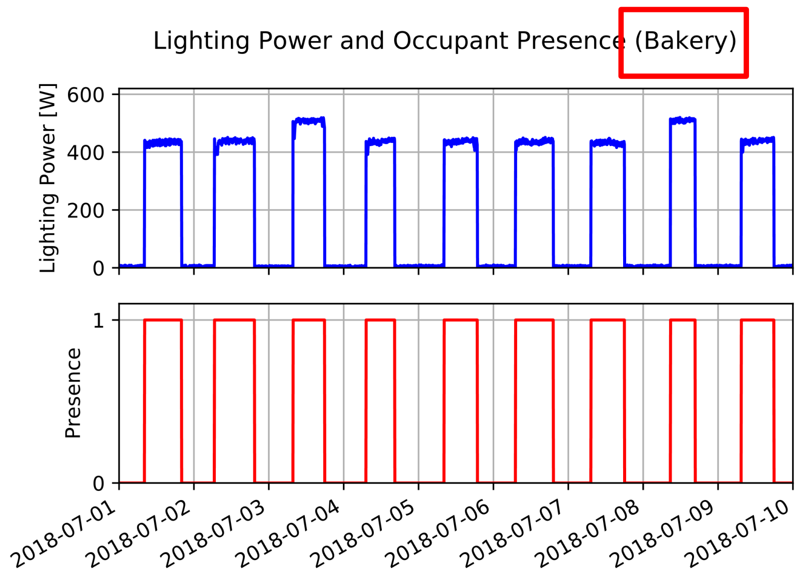


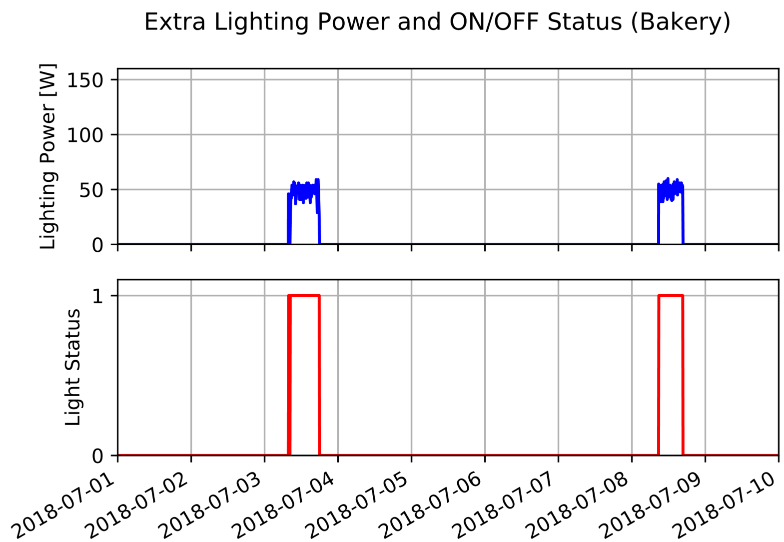
Methodology – Occupant Presence

❑ Lighting power shapes and interpreted occupant presence







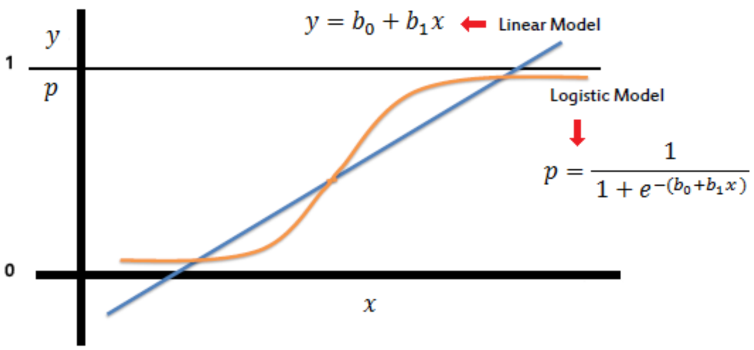


Methodology – Logistic Regression

❑ Logistic regression model

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m)}}$$

- p – probability of occupant present or extra lights on;
- x – independent variables (e.g. **time of day** for occupant presence)



❑ Why logistic regression?

- It is a linear classifier and is easy to train
- It can reach the same level of accuracy as non-linear classifiers;
- It is easy to implement in Modelica.

Figure source: https://www.saedsayad.com/logistic_regression.htm

Methodology – Lighting Power

❑ Training data

- Data of summer 2018, June, July for training, August for testing

$$Accuracy = \frac{No. of\ correctly\ classified\ points}{No. of\ total\ data\ points}$$

	Predicted No	Predicted Yes
Actual No	3693	44
Actual Yes	31	624

❑ Multi-stage lighting power description

$$P(t) = a_0(t)P_{base} + a_1(t)P_{extr,1} + a_2(t)P_{extr,2} + \cdots + a_{n-1}(t)P_{extr,n-1}$$

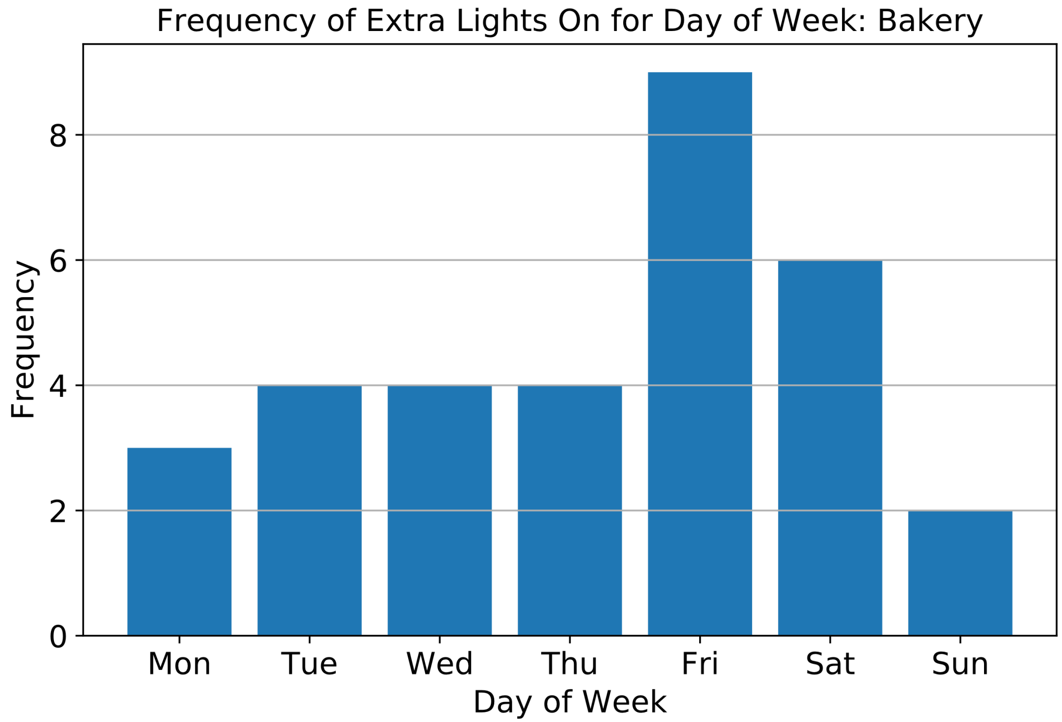
- P – total lighting power;
- a_i – binary variable indicating the status of base or extra lighting power;

$$a_i(t) = Bool(Random\ Number < p)$$

- n – number of stages.

Methodology – Logistic Regression

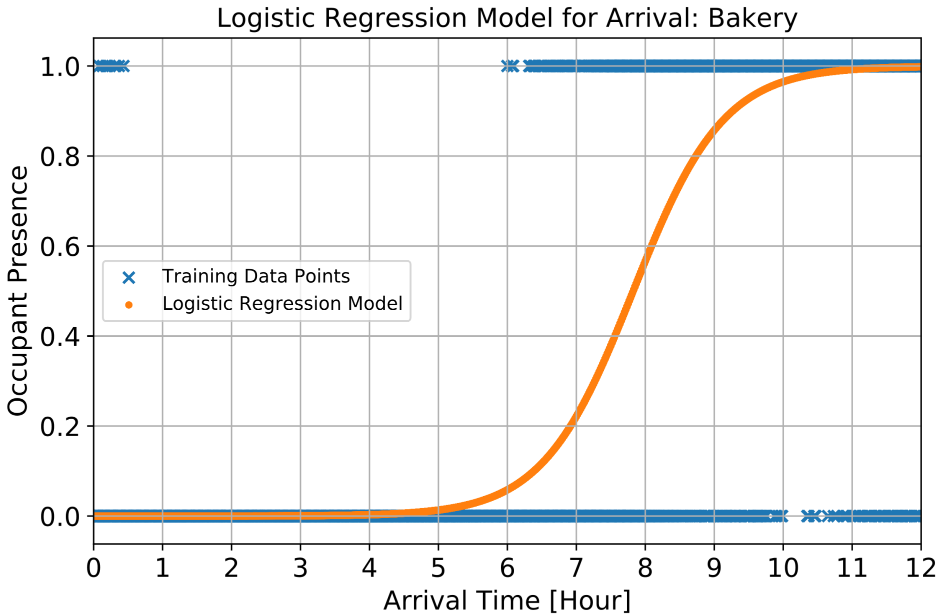
- **Extra lighting model in the bakery**

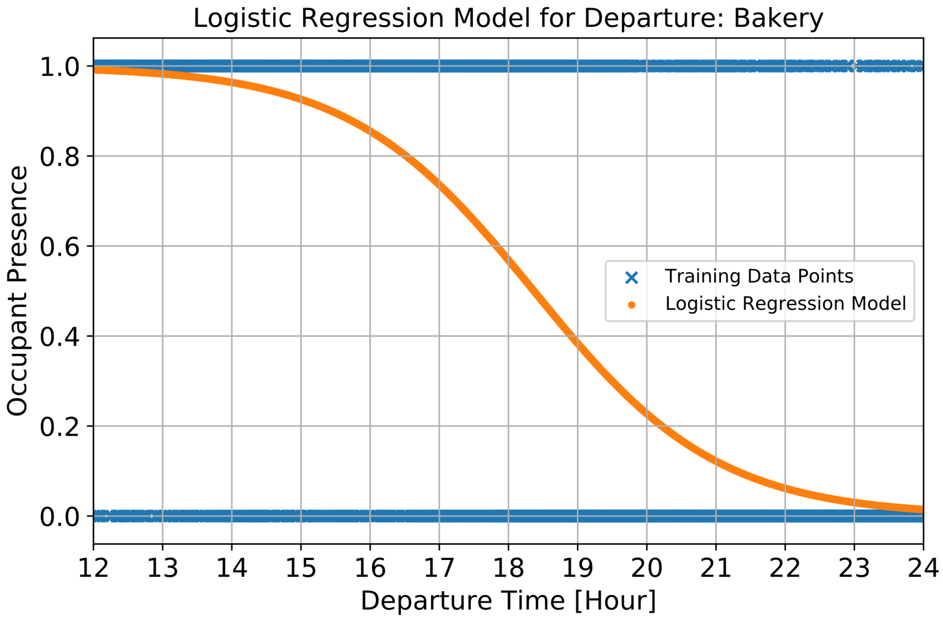


- The status of the extra lighting has a correlation with day of week
- The total frequency of extra lights on in 2018 is only 8.8%
- To deal with the imbalance in the training dataset, we adopted the Synthetic Minority Over-sampling Technique (SMOTE)

Methodology – Logistic Regression

❑ Regression results (Bakery)

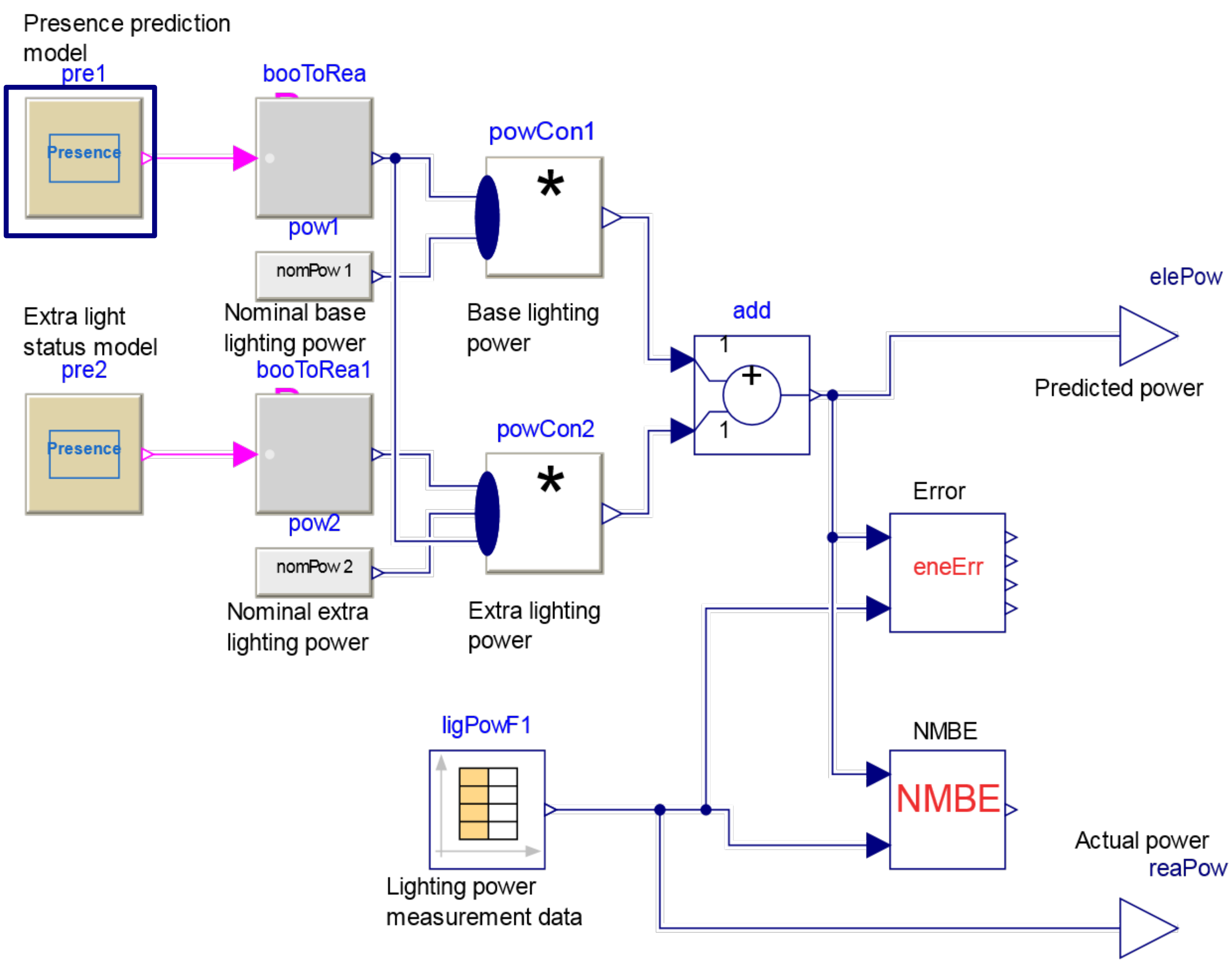




		Accuracy
Ice Cream Shop	Arrival	0.98
	Departure	0.97
Bakery	Arrival	0.94
	Departure	0.88
	Extra	0.84

Methodology – Implementation in Modelica

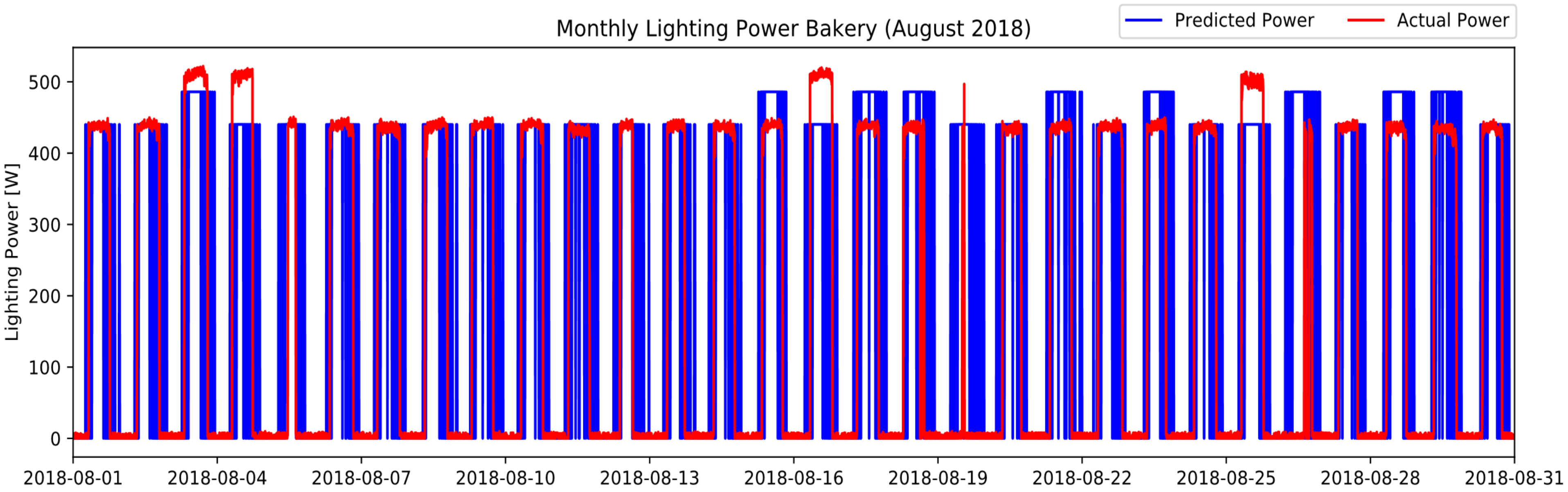
❑ Modelica model for Bakery



- Stochastic simulation model
- Every two minutes, a binary variable generator randomly generates a binary number.
- The probability of this number being 1 equals the probability at that time of day based on the logistic regression model.

Methodology – Implementation in Modelica

❑ Monthly predicted and actual lighting power (Bakery)



❑ Evaluation Metrics

- Root mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(x_{f,i}-x_{o,i}\right)^2}{N}}$$

- Coefficient of Variation of RMSE (CVRMSE)

$$CVRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_{f,i}-x_{o,i}\right)^2}}{\overline{x_o}}$$

- Relative Error (RE) of Peak Power

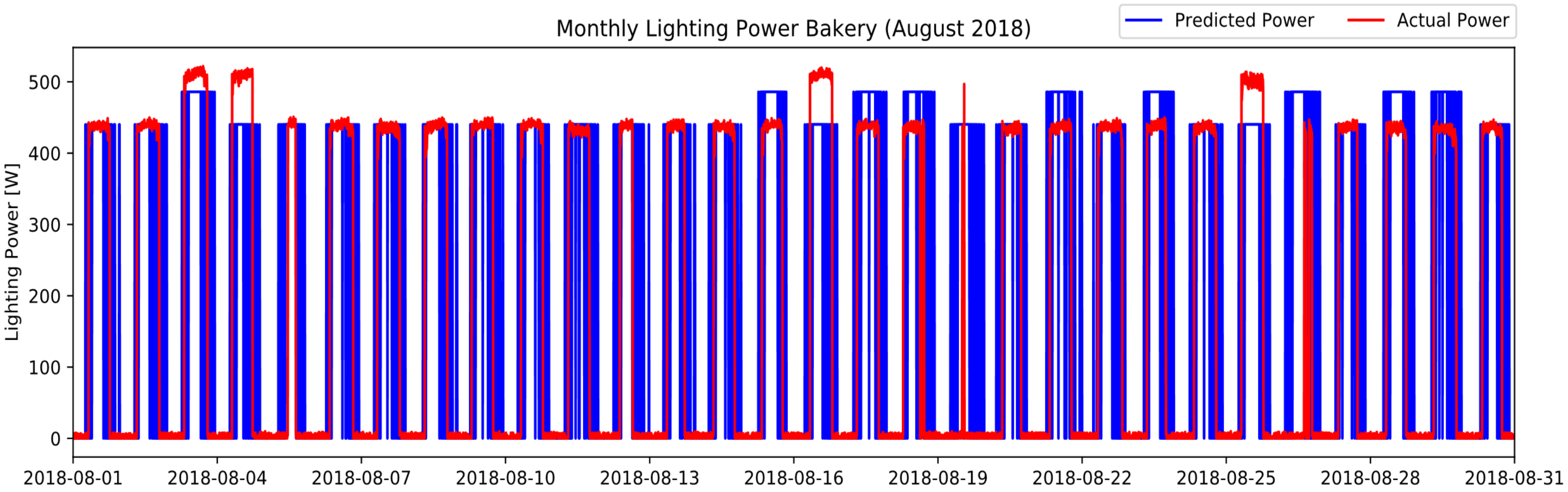
$$RE = \frac{|x_{f,i}-x_{o,i}|}{x_{o,i}}$$

- Normalized Mean Biased Error (NMBE)

$$NMBE = \frac{\sum_{i=1}^{N}\left(x_{f,i}-x_{o,i}\right)}{N\times\overline{x_o}}$$

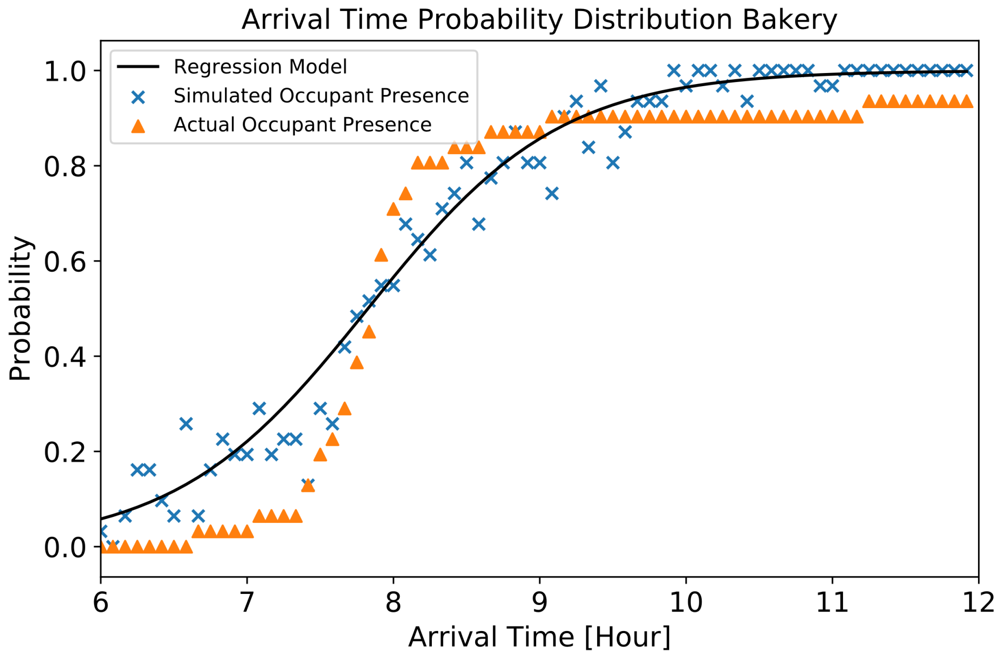
Methodology – Implementation in Modelica

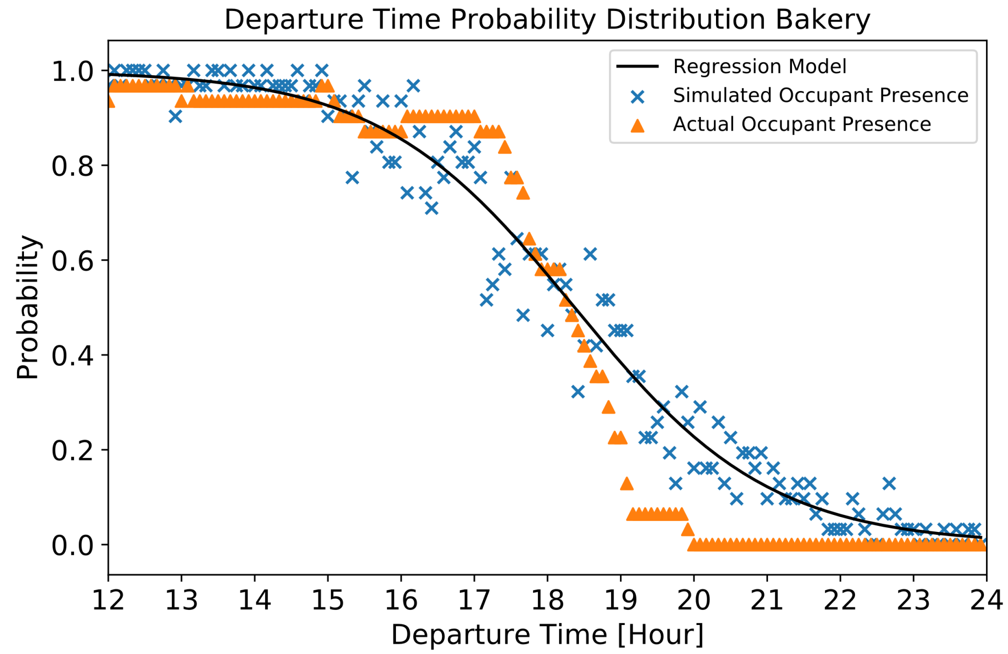
❑ Monthly predicted and actual lighting power (Bakery)



- ➢ Noticed oscillations in predicted power profile.
- ➢ 9 times of predicted extra lights on in a month but only 4 times of actual extra lights on.

Results and discussions





➢ Presence Prediction Performance

	C2	F1	
	Presence	Presence	Extra Lights
RMSE	0.108	0.101	0.153
CVRMSE	20.9%	25.0%	125%

➢ Probability of Extra Lights On

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Simulated	0	0.29	0.29	0.14	0.29	0.29	0.14
Actual	0	0	0	0.14	0.29	0.29	0.14

Results and discussions

➢ Peak lighting power prediction (avg. RE)

	Monthly Peak Power	Weekly Peak Power	Daily Peak Power
C2	2.36%	2.36%	1.99%
F1	6.90%	5.34%	2.42%

➢ Lighting power prediction (avg. NMBE)

		Baseline	Model
Monthly NMBE	C2	0.061%	3.92%
	F1	-0.55%	8.28%
Weekly NMBE	C2	0.060%	4.07%
	F1	-0.68	7.92%
Daily NMBE	C2	0.057%	4.03%
	F1	0.39%	44.1%

❑ Discussions

1. Low accuracy for extra lights on prediction.
2. Simulated and actual probability of extra lights on deviate on Tuesday and Wednesday.
3. Lighting power two-stage prediction has larger errors. The errors stay below 6.9%.
4. Better prediction performance in longer prediction horizons.

Conclusion

- A method for occupant presence learning and reproducing based on lighting power data is proposed and validated.
- The proposed models can predict daily lighting peak power within 2.42% relative error.
- Stochastic models can be very accurate for longer-term predictions. However, they cannot predict uncommon events, and this leads to larger short-term prediction errors.
- **Limitation:** not having the ground truth data for occupant presence.
- **Future work:** cross validation of the occupant presence with other appliance usage data.

Q&A

Thank You!

Jing Wang

Jing.Wang@colorado.edu