Data Cleansing
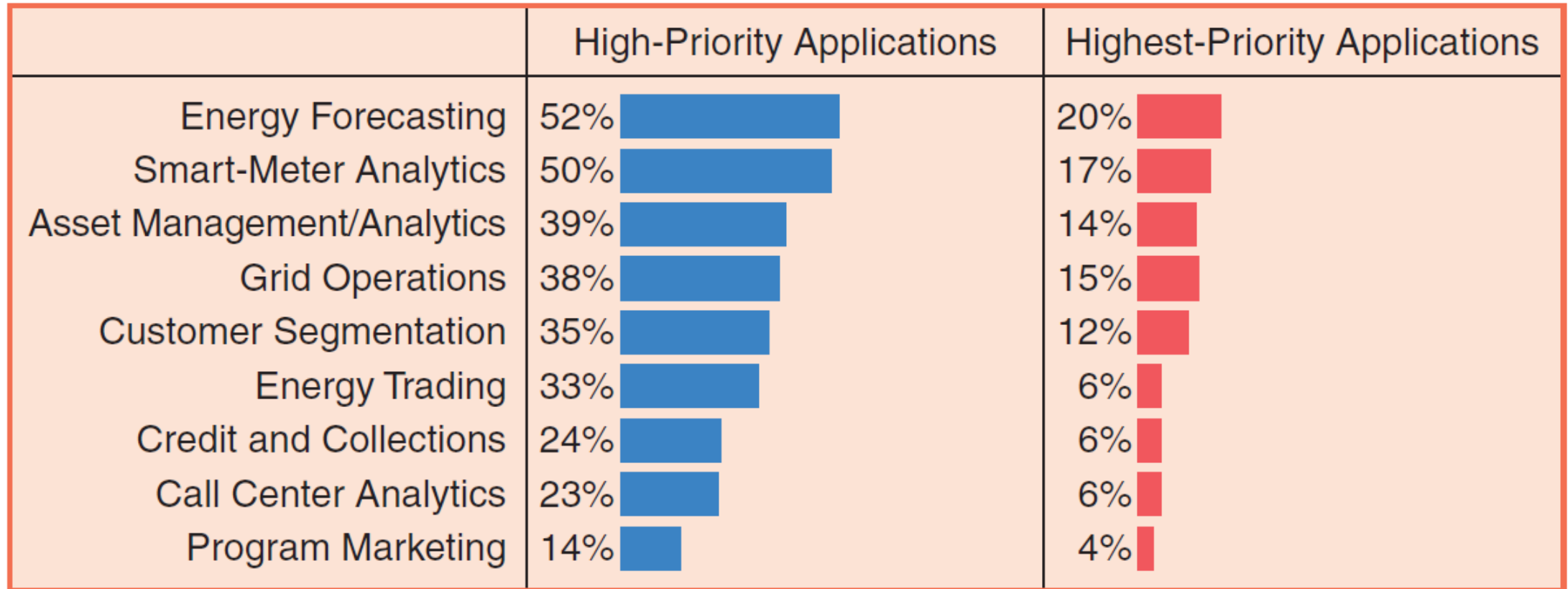## Three Case Studies and Beyond

Tao Hong

Director of Big Data Energy Analytics Laboratory (BigDEAL)

Associate Professor and NCEMC Faculty Fellow of Energy Analytics

Department of Systems Engineering and Engineering Management

University of North Carolina at Charlotte
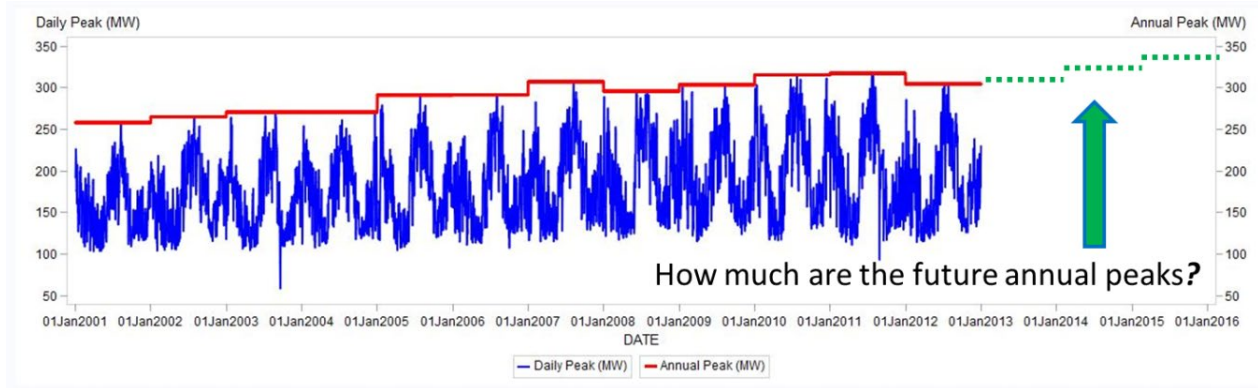
www.drhongtao.com

Outline

- Utility analytics survey
- Electric load forecasting
- Load data cleansing
- Detecting load transfers
- Weather data cleansing
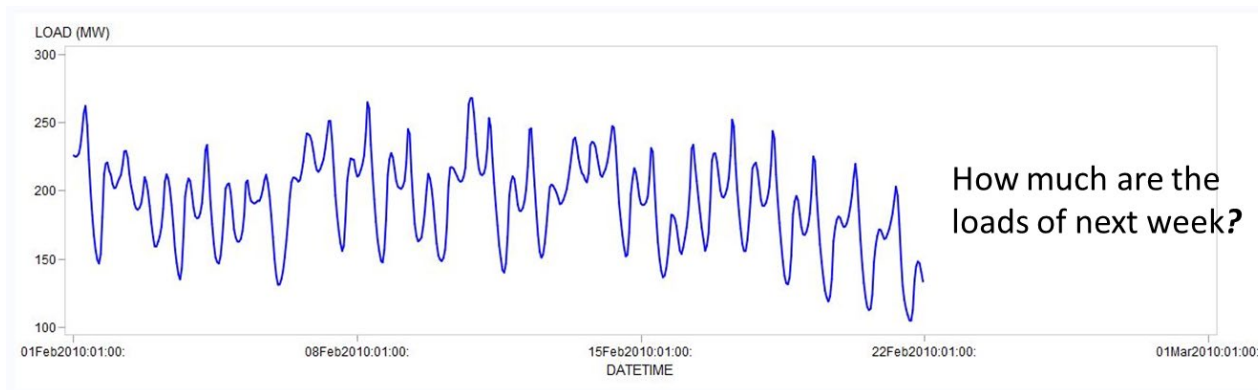- Garbage in, ~~garbage~~ valuable out
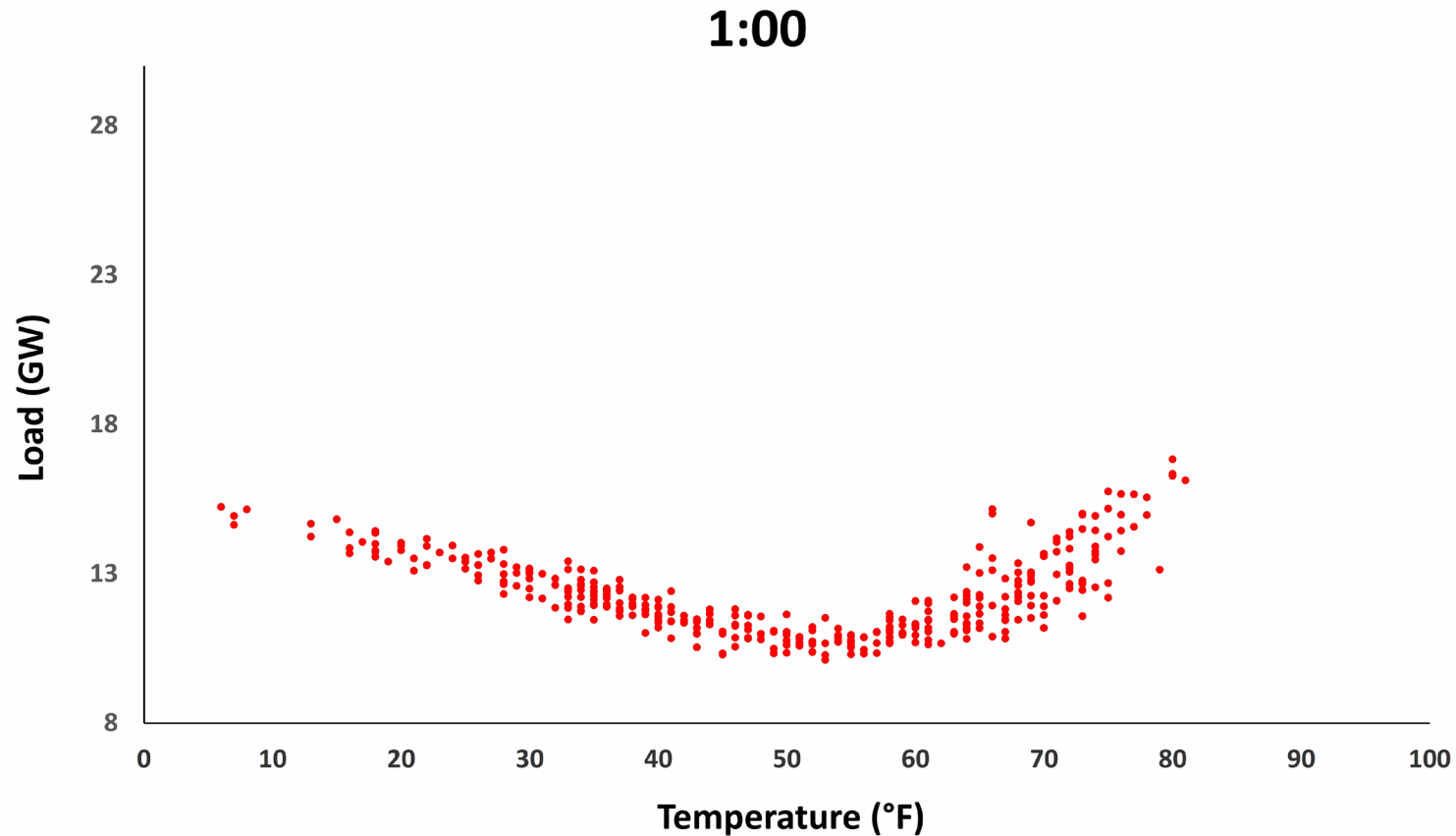
From the industry ...

	High-Priority Applications	Highest-Priority Applications
Energy Forecasting	52%	20%
Smart-Meter Analytics	50%	17%
Asset Management/Analytics	39%	14%
Grid Operations	38%	15%
Customer Segmentation	35%	12%
Energy Trading	33%	6%
Credit and Collections	24%	6%
Call Center Analytics	23%	6%
Program Marketing	14%	4%

(Hong, et al., P&E Magazine 2018)

Electric load forecasting

$$Load = f(x_1, x_2, x_3, \ldots) + e$$

(Hong, Foresight 2014)

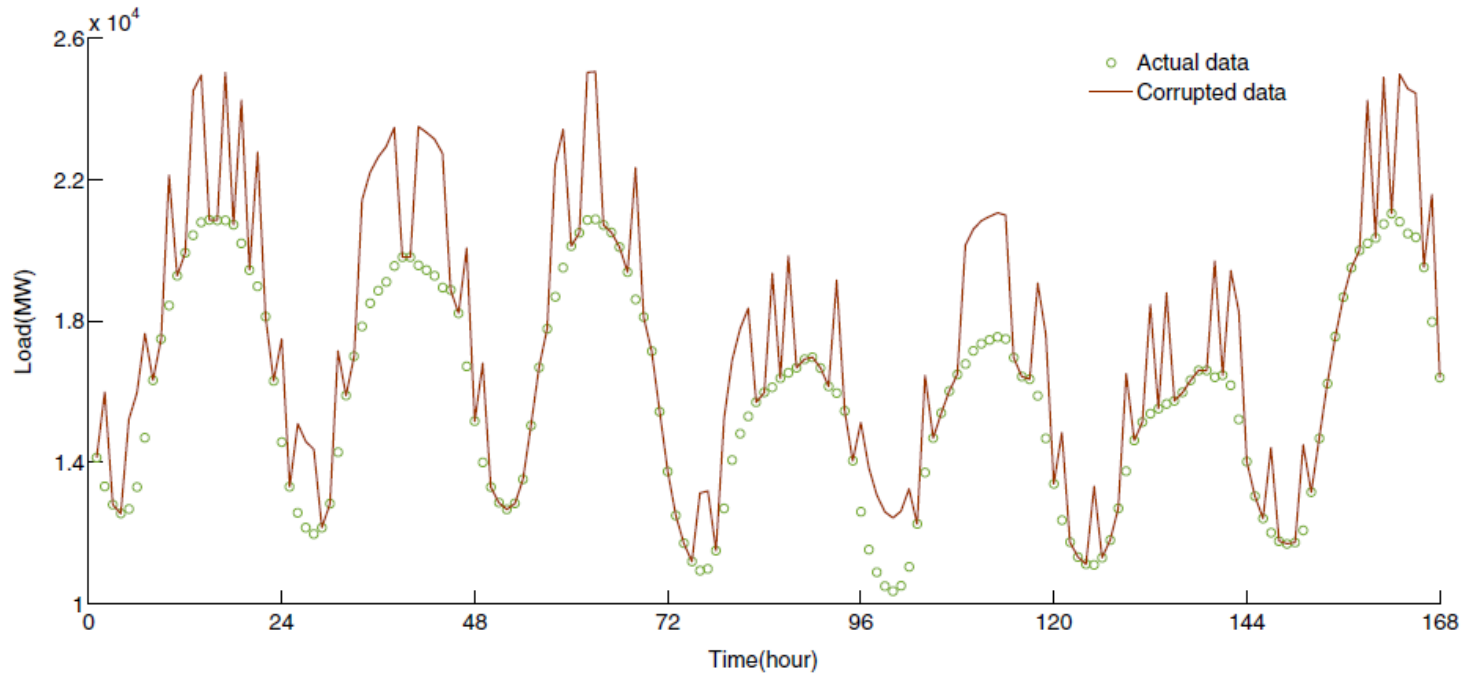# Load vs. temperature

1:00

$$L = \beta_0 + \beta_1 \text{Trend} + \boldsymbol{\beta_2}M + \boldsymbol{\beta_3}H + \boldsymbol{\beta_4}W + \boldsymbol{\beta_5}HW + g(T)$$

(Hong, Pinson & Fan, IJF 2014)

Load data cleansing
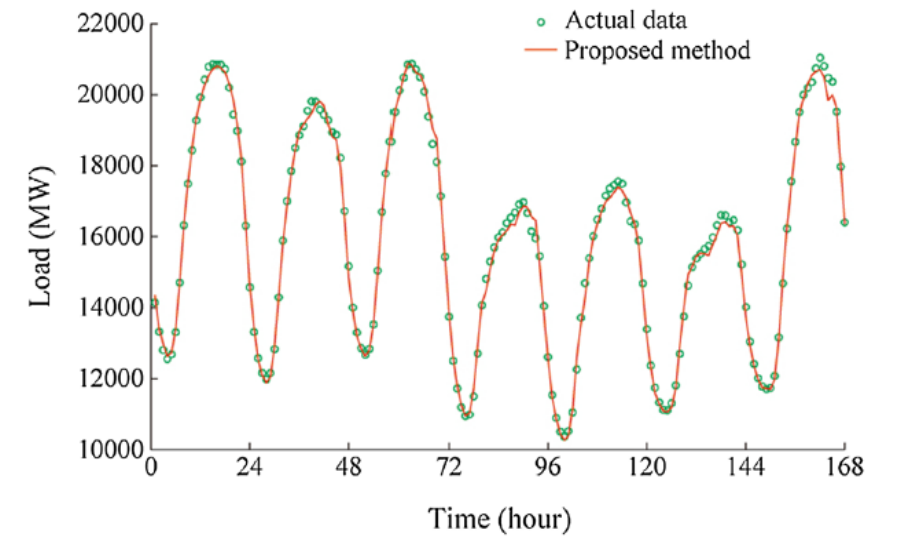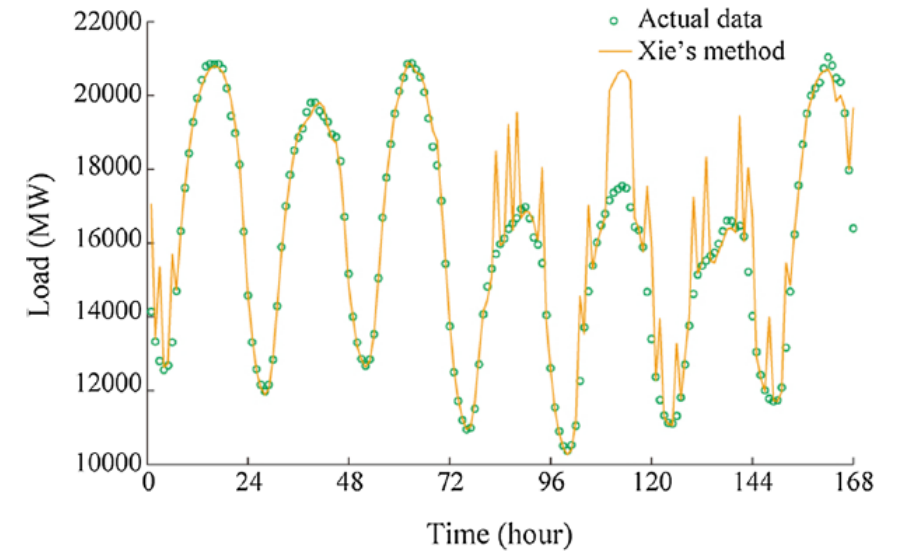
(Xie & Hong, IJF 2016)
(Luo, Hong, & Yue, MPCE 2018)

Detecting load transfers
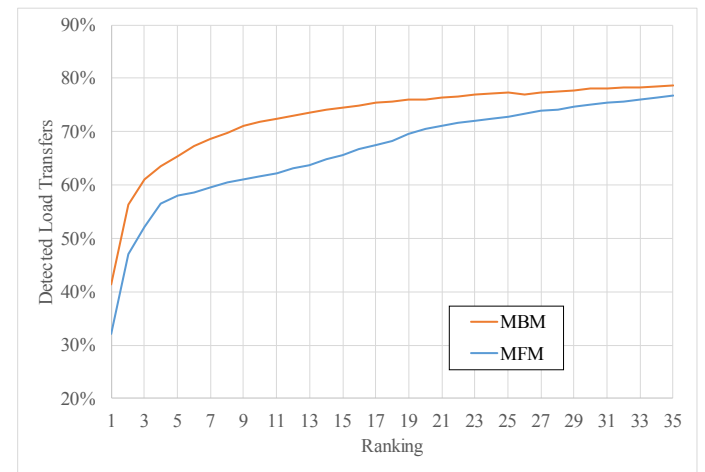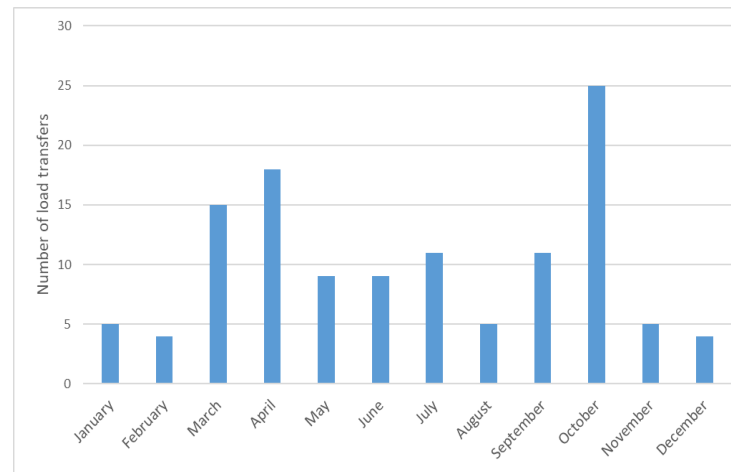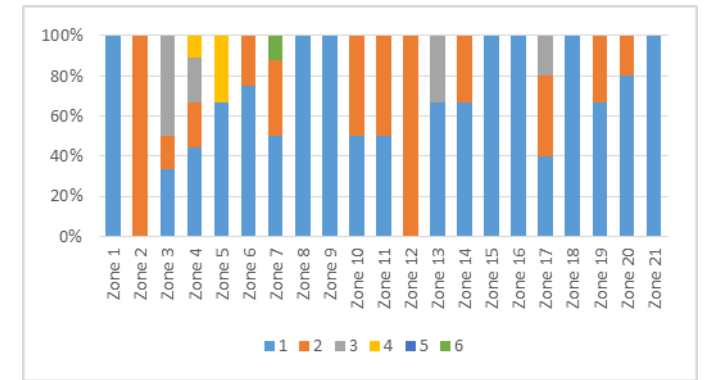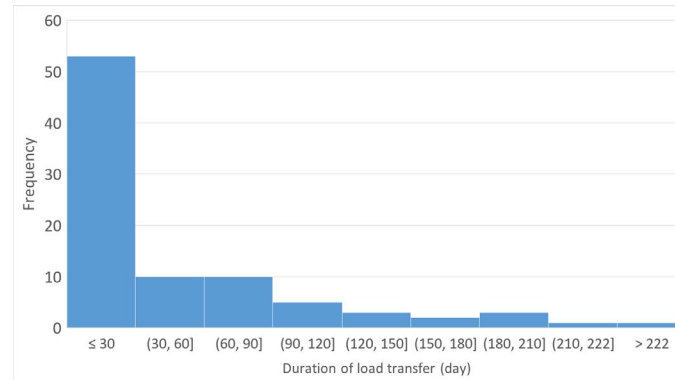
Weather stations

(Hong, Wang, & White, IJF 2015)

New England
2015

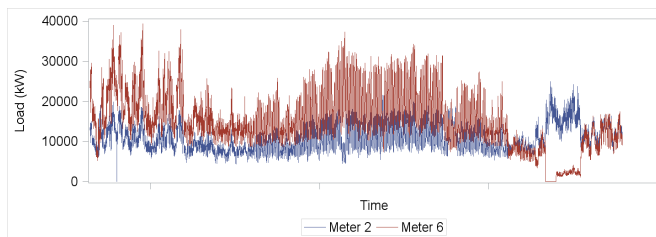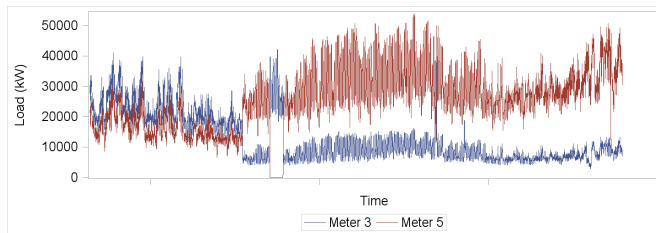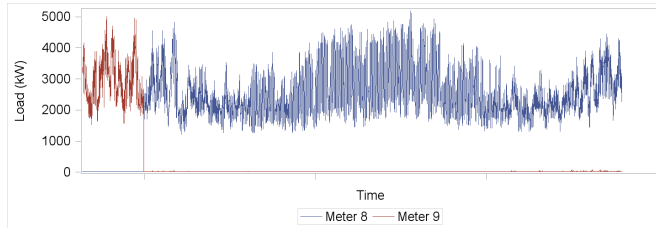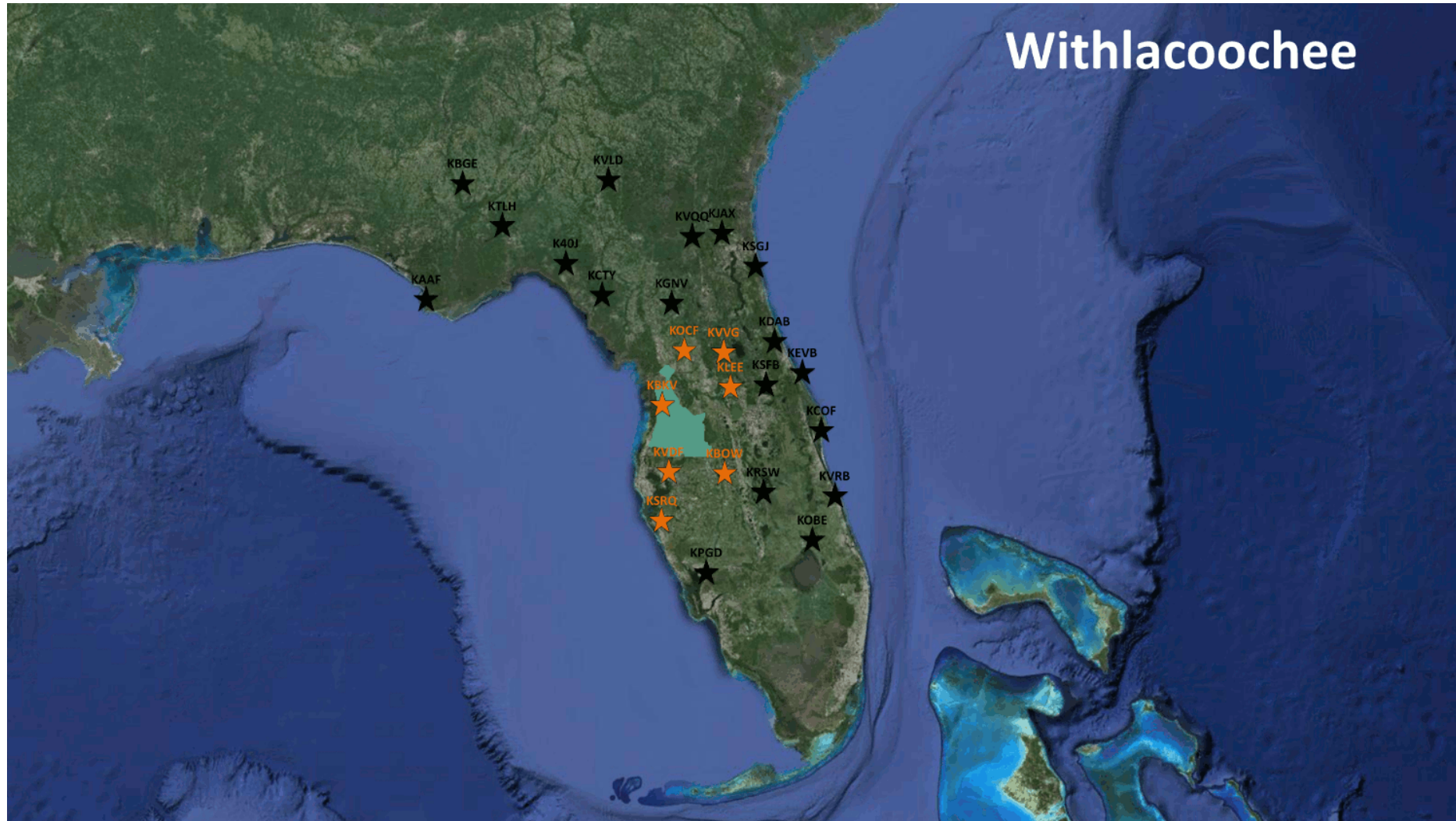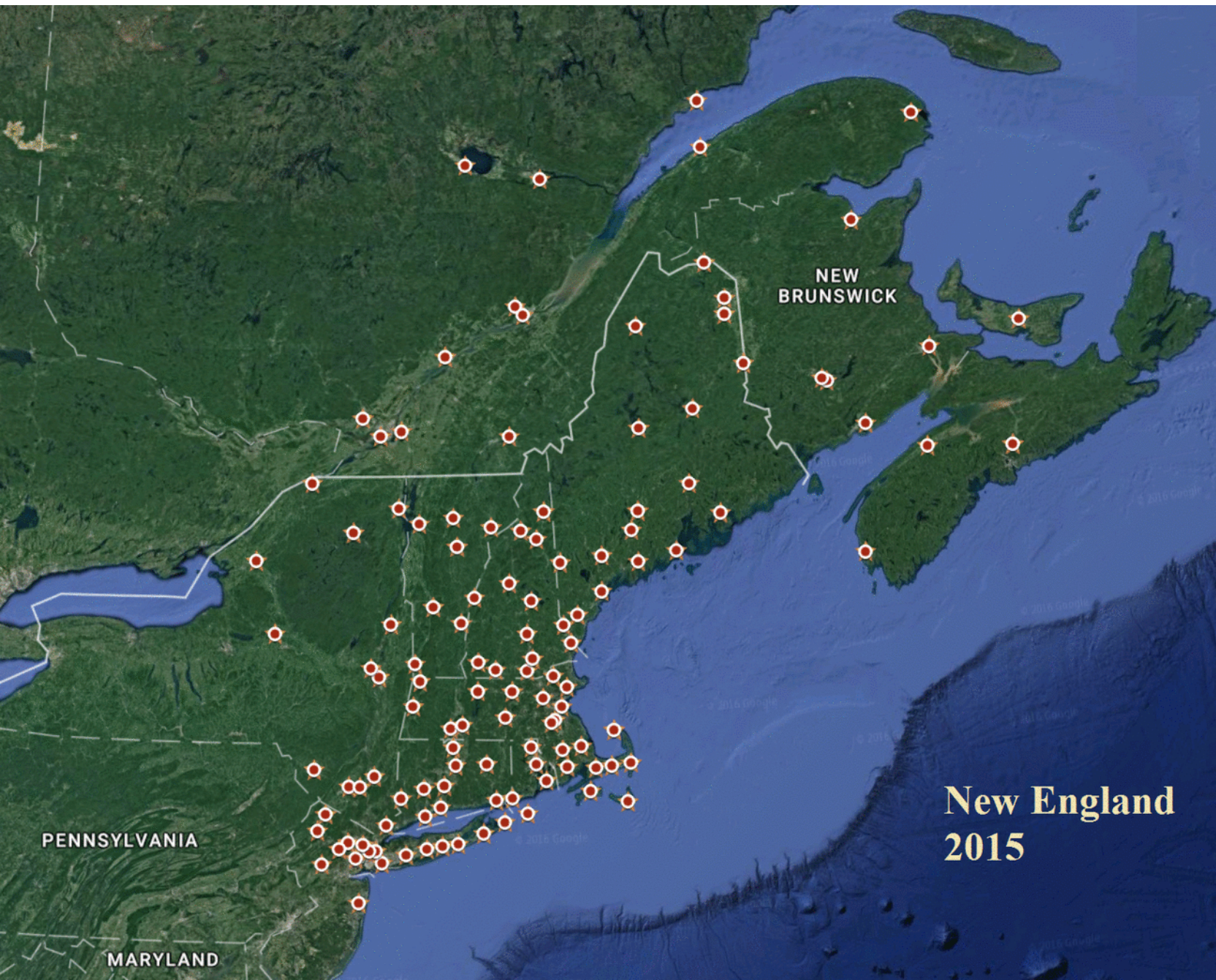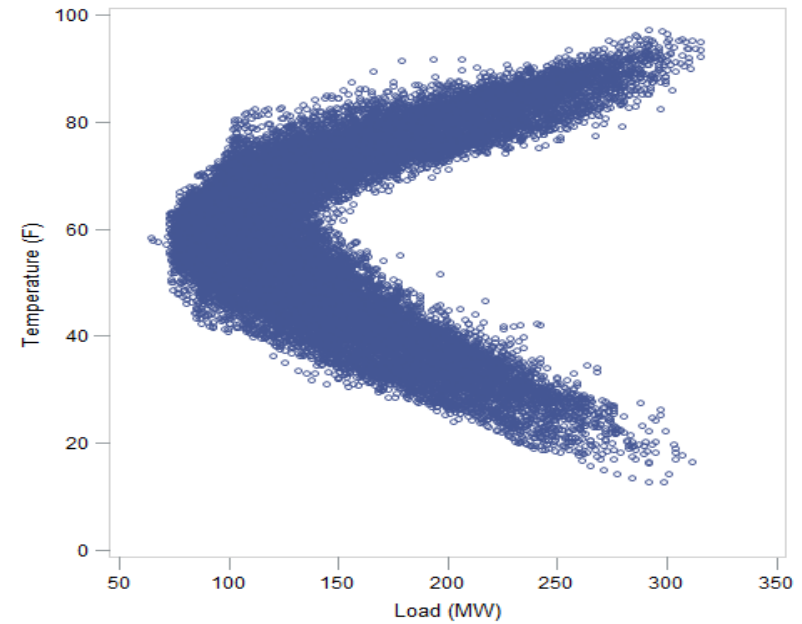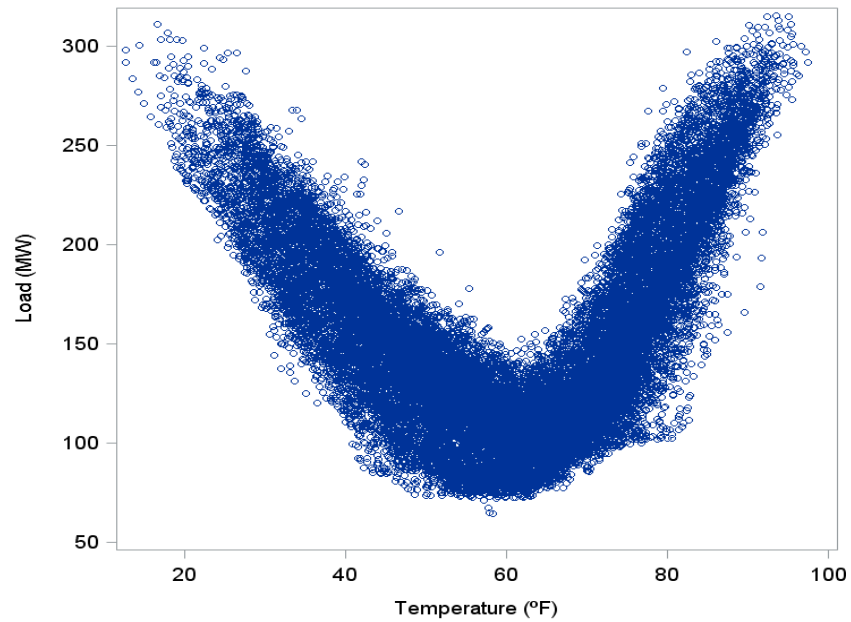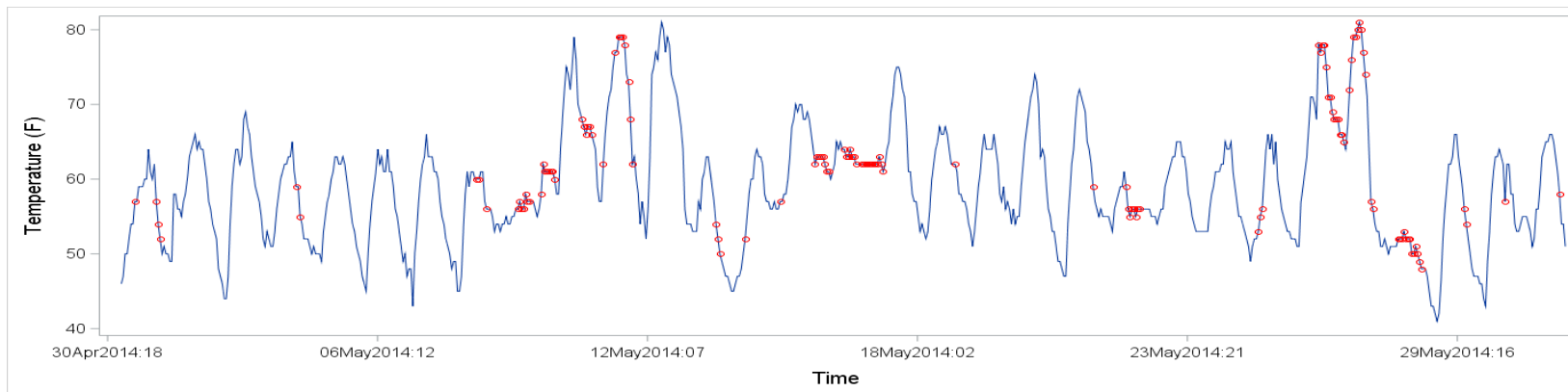|				
CYBC	KACK	KDXR	KIZG	KPOU
CYBG	KAFN	KEEN	KJFK	KPQI
CYCX	KALB	KEWB	KLCI	KPSF
CYFC	KAQW	KEWR	KLEB	KPSM
CYGP	KASH	KFIT	KLEW	KPVC
CYHU	KAUG	KFMH	KLGA	KPVD
CYHZ	KBAF	KFOK	KLWM	KPWM
CYMX	KBDL	KFRG	KMGJ	KPYM
CYOY	KBDR	KFVE	KMHT	KRKD
CYQB	KBED	KFWN	KMLT	KRME
CYQI	KBGR	KGFL	KMMK	KRUT
CYQM	KBHB	KGNR	KMMU	KSCH
CYRJ	KBID	KGON	KMPV	KSFM
CYRQ	KBLM	KGTB	KMQE	KSFZ
CYSC	KBML	KHFD	KMSS	KSLK
CYSJ	KBOS	KHIE	KMSV	KSMQ
CYUL	KBTV	KHPN	KMTP	KSWF
CYYG	KBVY	KHTO	KMVL	KTAN
CYYY	KCAR	KHUL	KMVY	KTEB
CYZX	KCDW	KHVN	KMWN	KUUU
CZBF	KCEF	KHWV	KNYC	KVSF
K1P1	KCON	KHYA	KORE	KWST
K1V4	KCQX	KIJD	KORH	KWVL
K12N	KDAW	KISP	KOWD	KPBG
K40B	KDDH	KIWI	KOXC	

124 weather stations

9

Weather data cleansing

(Sobhani, Hong, & Martin, IJF 2019)

Weather data cleansing

(Sobhani, Hong, & Martin, IJF 2019)

Robust forecasting models

(Luo, Hong, & Fang, TSG 2019)
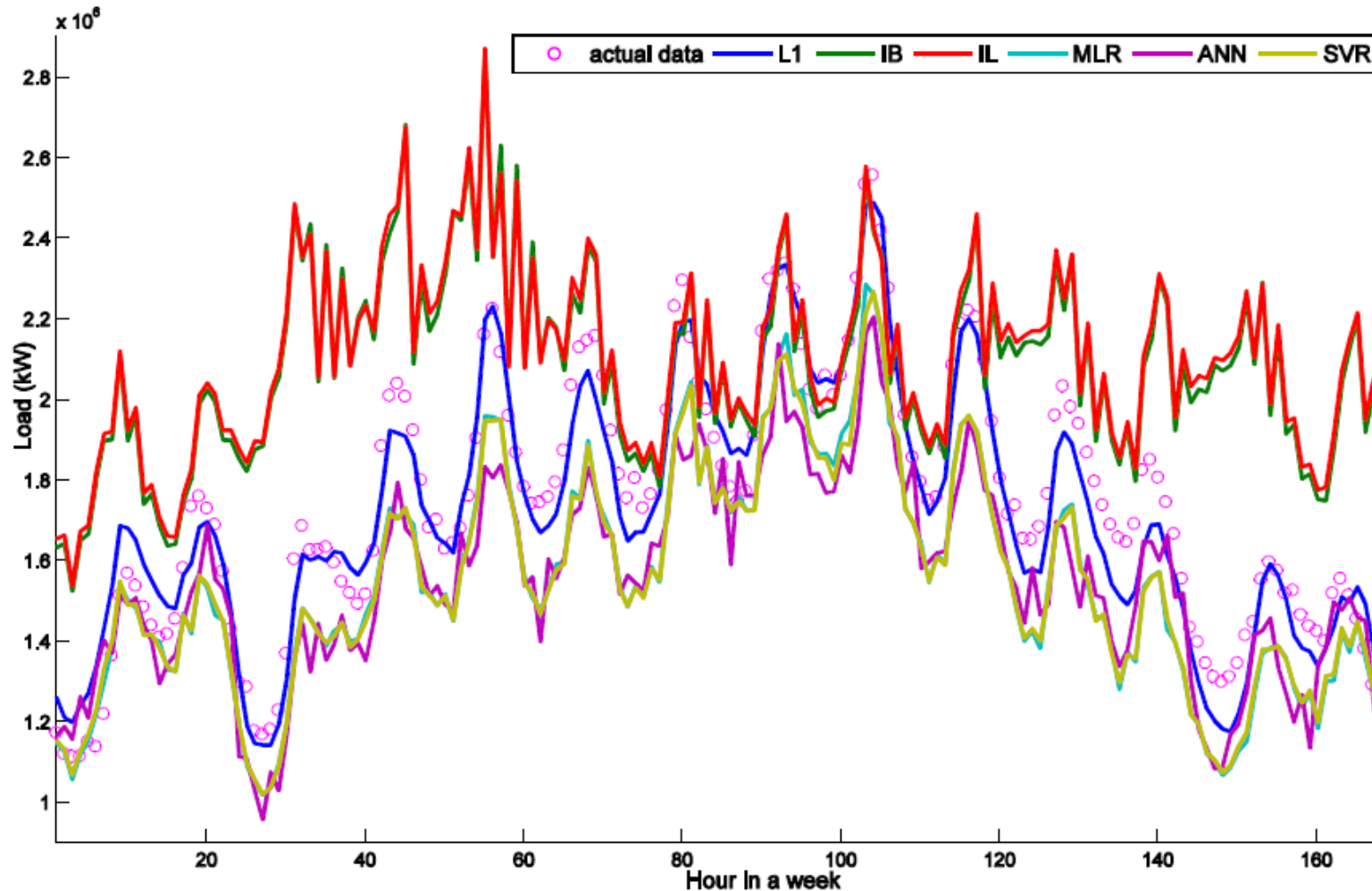
References

1. Tao Hong, "Energy forecasting: past, present and future," Foresight: The International Journal of Applied Forecasting, issue 32, pp. 43-48, Winter 2014.

2. Tao Hong, Pierre Pinson and Shu Fan, "Global energy forecasting competition 2012," *International Journal of Forecasting*, vol.30, no.2, pp 357-363, April-June 2014.

3. Tao Hong, Pu Wang and Laura White, "Weather station selection for electric load forecasting," *International Journal of Forecasting*, vol.31, no.2, pp 286-295, April-June 2015.

4. Jingrui Xie and Tao Hong, "GEFCom2014 probabilistic electric load forecasting: an integrated solution with forecast combination and residual simulation," *International Journal of Forecasting*, vol.32, no.3, pp 1012-1016, July-September 2016.

5. Jian Luo, Tao Hong and Meng Yue, "Real-time anomaly detection for very short-term load forecasting," *Journal of Modern Power Systems and Clean Energy*, vol.6, no.2, pp 235-243, March 2018.

6. Tao Hong, David Wenzhong Gao, Tom Laing, Dale Kruchten, and Jorge Calzada, "Training energy data scientists: universities and industry need to work together to bridge the talent gap," IEEE Power and Energy Magazine, vol.16, no.3, pp 66-73, May-June 2018.

7. Masoud Sobhani, Tao Hong, and Claude Martin, "Temperature data cleansing for electric load forecasting," *International Journal of Forecasting*, in press.

8. Jian Luo, Tao Hong, and Shu-Cherng Fang, "Robust regression models for load forecasting," *IEEE Transactions on Smart Grid*, in press.

9. Masoud Sobhani and Tao Hong, "Detecting load transfers," submitted to *IEEE Transactions on Smart Grid*.