International Journal of Forecasting 34 (2018) 89–104

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Benchmarking robustness of load forecasting models under data integrity attacks

Jian Luo [a], Tao Hong [a,b,*], Shu-Cherng Fang [c]

[a] School of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian, China
[b] Department of Systems Engineering and Engineering Management, University of North Carolina at Charlotte, NC, USA
[c] Department of Industrial & Systems Engineering, North Carolina State University, NC, USA

ARTICLE INFO

Keywords:
Cybersecurity
Data integrity attack
Electric load forecasting
Linear regression
Neural network
Support vector regression
Fuzzy regression

ABSTRACT

As the internet's footprint continues to expand, cybersecurity is becoming a major concern for both governments and the private sector. One such cybersecurity issue relates to data integrity attacks. This paper focuses on the power industry, where the forecasting processes rely heavily on the quality of the data. Data integrity attacks are expected to harm the performances of forecasting systems, which will have a major impact on both the financial bottom line of power companies and the resilience of power grids. This paper reveals the effect of data integrity attacks on the accuracy of four representative load forecasting models (multiple linear regression, support vector regression, artificial neural networks, and fuzzy interaction regression). We begin by simulating some data integrity attacks through the random injection of some multipliers that follow a normal or uniform distribution into the load series. Then, the four aforementioned load forecasting models are used to generate one-year-ahead ex post point forecasts in order to provide a comparison of their forecast errors. The results show that the support vector regression model is most robust, followed closely by the multiple linear regression model, while the fuzzy interaction regression model is the least robust of the four. Nevertheless, all four models fail to provide satisfying forecasts when the scale of the data integrity attacks becomes large. This presents a serious challenge to both load forecasters and the broader forecasting community: *the generation of accurate forecasts under data integrity attacks*. We construct our case study using the publicly-available data from Global Energy Forecasting Competition 2012. At the end, we also offer an overview of potential research topics for future studies.

1. Introduction

In the era of the internet of things, cybersecurity is of growing concern to governments, financial institutions, and many other business entities (Singer & Friedman, 2014). Among various cybersecurity issues, data integrity attacks, where hackers access supposedly protected data and inject false information, are of great importance to the forecasting community, because the quality of input data affects the forecast accuracy directly. Although the topic of "outlier detection" has been studied extensively in the literature (Akouemo & Povinelli, 2016; Hodge & Austin, 2004; Rousseeuw & Leroy, 2005; Xie & Hong, 2016), the problem of forecasting under data integrity attacks is still relatively new to the forecasting community. This paper focuses on the power industry, conducting an empirical study to test and benchmark the robustness of four representative load forecasting models under various simulated data integrity attacks using publicly available data.

The power industry operates the electric grid, one of the most complicated man-made systems in the world, to

[*] Corresponding author at: Department of Systems Engineering and Engineering Management, University of North Carolina at Charlotte, NC, USA.
E-mail address: hong@uncc.edu (T. Hong).

http://dx.doi.org/10.1016/j.ijforecast.2017.08.004
0169-2070/© 2017 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

produce and deliver electricity to over 5.6 billion people worldwide. Like many other industries, the power industry requires good forecasts of the electricity supply, demand and price in order to plan and operate the grid. Electric load forecasting has been an integral part of its business operations for over a century. Applications of load forecasting are spread across virtually every segment of the power industry (Hong, 2014). Both power companies' financial bottom lines and the resilience of power grids rely on accurate load forecasts.

Over the past several decades, researchers and practitioners have used a variety of techniques in their attempts to tackle the load forecasting problem (Hippert, Pedreira, & Souza, 2001; Hong & Fan, 2016; Weron, 2006). Some of these, such as artificial neural networks (Khotanzad & Afkhami-Rohani, 1998), semi-parametric models (Fan & Hyndman, 2012; Hyndman & Fan, 2010) and multiple linear regression models (Hong, 2010; Hong, Wilson, & Xie, 2014; Papalexopoulos & Hesterberg, 1990), have been used by power companies in the production environment, while others have excelled in notable load forecasting competitions. For instance, the support vector regression model won the EUNITE load forecasting competition in 2001 (Chen, Chang, & Lin, 2004); while regression models and gradient boosting machines took the top places in the load forecasting track of the Global Energy Forecasting Competition 2012, (GEFCom2012; see Hong, Pinson, & Fan, 2014).

The modern power grid relies heavily on communication networks and information technologies. Unfortunately, while such integration is essential for the evolution of the power grid, it also makes the grid more vulnerable to cyber-attacks by hackers around the world. For instance, a cyberattack to the supervisory control and data acquisition (SCADA) system of a Ukrainian power company disconnected seven substations for three hours (Perez, 2016). Thus, cybersecurity is an emerging field in power system research. While some researchers have studied data integrity attacks against "state estimation" (Hu & Vasilakos, 2016; Liu, Ning, & Reiter, 2011), little work has been done on data integrity attacks against "load forecasting". This paper attempts to tackle the problem of *load forecasting under data integrity attacks* by first benchmarking some of the existing models.

Addressing the issue fully requires extensive effort to be put into (1) detecting attacks, (2) identifying maliciously manipulated data, and (3) cleansing and recovering attacked data, before formally using a load forecasting system to generate forecasts. However, there could be many possible forms of malicious data integrity attacks, some of which may even be beyond our current understanding. An intelligent hacker might be able to inject false information without being detected by the state estimator (Liu et al., 2011) or load forecaster, which makes it extremely complex and technically difficult to assure that the data are attack-free all the time. The most relevant literature may be that on outlier detection and data cleansing in electric load forecasting (Xie & Hong, 2016). However, detecting and cleansing the attacked data points in load forecasting (Xie & Hong, 2016) and state estimation (Liu et al., 2011) is very difficult and expensive. Hence, it is imperative to first benchmark the performances of representative load forecasting models under possible data integrity attacks.

To establish a basic framework for benchmarking, we consider the case where hackers get hold of the historical load data and select a random set of data points (a given percentage of the whole data set) to multiply by a set of maliciously injected multipliers. The range of the multipliers is assumed to be either normally or uniformly distributed. These two types of data integrity attacks are called normally-distributed and uniformly-distributed data integrity attacks, respectively. Essentially, this simulation method can be viewed as injecting noise into the input data. It is worth noting that the "noise" in time series forecasting is natural and is usually small, while these simulated multipliers can reach large magnitudes, such as many times the original load values.

This study begins by simulating the corresponding data integrity attacks, then benchmarks the point forecast accuracy under various simulated data integrity attacks for the following four load forecasting models: multiple linear regression (MLR), artificial neural network (ANN), support vector regression (SVR), and fuzzy interaction regression (FIR). We select these four models for comparison because they are representative, in the sense that their characteristics range from black-box to non-black-box, statistical to fuzzy, and classical to emerging.

This paper contributes to the field in at least three ways: (1) it introduces an important emerging problem to the forecasting community, namely forecasting under data integrity attacks; (2) it proposes a systematic data integrity attack simulation framework for load forecasting; (3) it benchmarks and analyzes the robustness of four representative load forecasting models under data integrity attacks at various levels. Since the data used in this paper are accessible publicly, the results of this paper can be reproduced freely or used by readers directly for other benchmarking purposes.

The rest of the paper is arranged as follows. Section 2 provides an overview of the four representative load forecasting models. Section 3 introduces the settings of the benchmark study, including the GEFCom2012 data, the accuracy of the four load forecasting models without data integrity attacks, and the data integrity attack simulation framework. Section 4 presents the computational results of the four load forecasting models under two data attack scenarios and discusses the robustness of these models. Section 5 discusses several future research directions. Finally, the paper concludes in Section 6.

2. Four load forecasting models

This section introduces the four representative load forecasting models, namely MLR, ANN, SVR and FIR. Note that all of the models implemented in this study use temperature information for constructing explanatory variables. In other words, we did not use any of the well-known naïve (e.g., random walk and seasonal naïve) or time series (e.g., exponential smoothing and autoregressive integrated moving average) models for our comparisons, because load forecasting models that do not take weather inputs are of limited use in practice (Wang, Liu, & Hong, 2016).

2.1. MLR model

MLR is a technique that is used widely for forecasting. In load forecasting, the load (or some transformation thereof) is usually treated as the dependent variable, while the weather and calendar variables are treated as independent variables. The parameters of MLR models are often estimated using the ordinary least squares method (Kutner, Nachtsheim, Neter, & Li, 2004). Papalexopoulos and Hesterberg (1990) proposed a regression-based approach to load forecasting, which was implemented at the Pacific Gas and Electric Company. Ramanathan, Engle, Granger, Vahid-Araghi, and Brace (1997) further proposed the use of 24 MLR models, one for each hour, with a dynamic error structure and adaptive adjustments for correcting the forecasting errors of previous hours, which won a competition organized by the Electric Power Research Institute in the 1990s. Hong (2010) proposed a series of interaction regression models, each of which could be used for all 24 h. One of these was used as the vanilla benchmark model in GEFCom2012 (Hong, Pinson, & Fan, 2014). The one-model based approach and the 24-model based approach were later compared by Wang et al. (2016). See also GEFCom2012 for other notable regression models, such as the one developed by Charlton and Singleton (2014), who won the top place.

This paper uses the variables of the vanilla benchmark model as the backbones of the MLR, FIR and SVR models. The vanilla model can be written as:

$$\begin{aligned} y = {} & \beta_0 + \beta_1 x_T + \beta_2 x_M + \beta_3 x_H x_W + \beta_4 x_t x_H + \beta_5 x_t^2 x_H \\ & + \beta_6 x_t^3 x_H + \beta_7 x_t x_M + \beta_8 x_t^2 x_M + \beta_9 x_t^3 x_M, \end{aligned} \tag{1}$$

where x_T is an increasing natural number representing a linear trend; x_H is a class variable that is equivalent to 24 indicator variables representing the 24 h of a day; x_W is a class variable representing seven days of a week; x_M is a class variable representing 12 months of a year; and x_t is a quantitative variable representing the temperature. For ease of presentation, we use β_j to denote the coefficients. However, it should be noted that, for a quantitative variable, β_j is a number, whereas for a class variable or an interaction including a class variable it is a vector of multiple coefficients. This vanilla model requires a total of 289 coefficients to be estimated. For parameter estimation, we augment the original data set of n records with the 289-dimensional training data set $\left\{\left(x^i, y^i\right), i = 1, \ldots, n\right\}$ using the temperature, calendar and load information, where $x^i \in \mathbb{R}^{289}$ indicates all 289 listed features on the right side of Eq. (1) for the ith observation, and $y^i \in \mathbb{R}$ indicates the actual i th observation of the load. Let the vectors x and β be the vectors of independent variables and their corresponding coefficients, respectively. Then, the MLR is to find the parameter vector (β, β_0) $\epsilon\mathbb{R}^{290}$ of a fitting hyperplane $y = \beta^T x + \beta_0$ of the training data set by minimizing the sum of squares as follows:

$$min_{\beta\in\mathbb{R}^{289}, \beta_0\in\mathbb{R}} \sum_{i=1}^{n} \left(y^i - \left(\beta^T x^i + \beta_0\right)\right)^2. \tag{2}$$

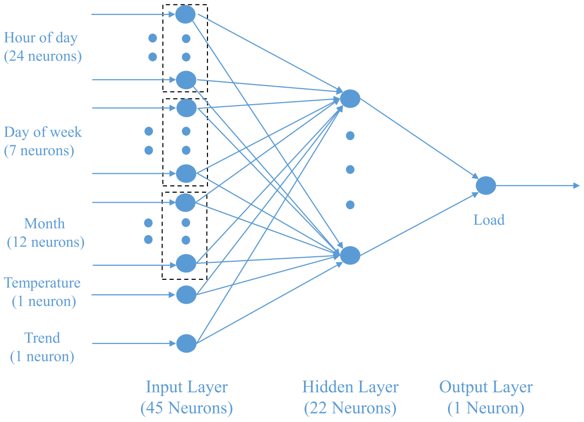

Fig. 1. The architecture of the ANN model.

2.2. ANN model

ANNs have been applied extensively to load forecasting since the 1980s, with varying degree of success. Being a black-box technique, ANN models do not require the forecaster to specify the functional form between the input and output variables (Hagan, Demuth, Beale, & De Jesús, 2014). By learning from the historical data, they can approximate the relationship between the input variables and the electric load, then use it for forecasting. Hippert et al. (2001) provided both a critical review of ANN-based load forecasting and a high-level methodology for developing ANN models for load forecasting. The best-known implementation of ANNs for load forecasting to date is the ANN short-term load forecaster (Khotanzad & Afkhami-Rohani, 1998), which was commercialized and is used by many power companies.

This paper adopts a three-layer feed-forward neural network, such as that implemented by Hong (2010). Fig. 1 depicts the architecture of this ANN model, which has 45 input neurons (representing the *Hour*, *Month*, *Weekday*, *Temperature*, and *Trend*) and one output neuron (*Load*). Here, the number of neurons in the hidden layer is set to be 22 (about half the number of input neurons). Every neuron from the input layer is connected to every neuron in the hidden layer, while the transfer functions for the hidden and output layers are selected to be 'logsig' and 'purelin' (Beale, Hagan, & Demuth, 2016), respectively. The training function for the ANN model uses the Levenberg–Marquardt back-propagation algorithm.

2.3. SVR model

A notable SVR model in the load forecasting field was developed by Chen et al. (2004), who won the load forecasting competition organized by the EUNITE network using it. However, the literature reporting successful implementations of SVR models for load forecasting in practice is very limited.

For the training data set $\left\{\left(x^i, y^i\right), i = 1, \ldots, n\right\}$, the least squares SVR model is used to find the parameter

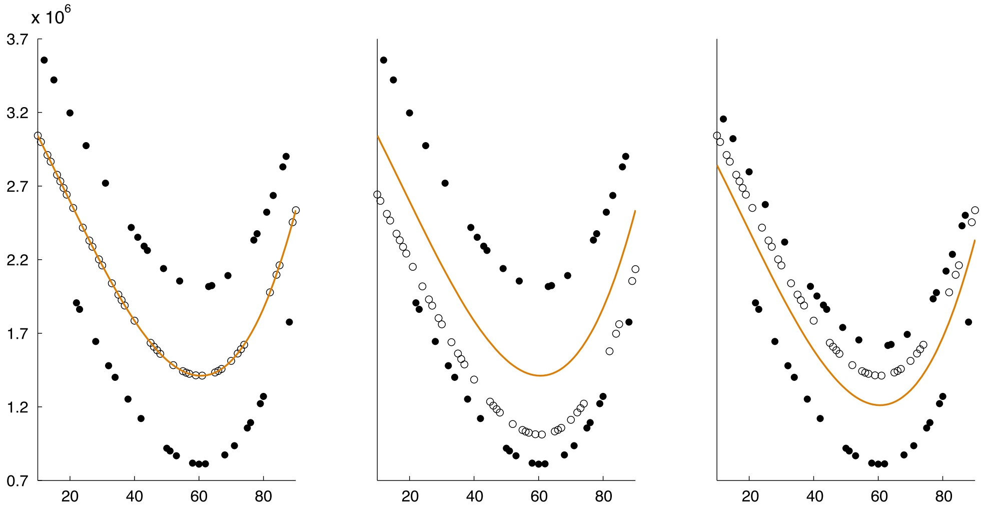

Fig. 2. Boundary observations determine the predictions of FIR models.

vector $(\beta, \beta_0)\ \epsilon\mathbb{R}^{290}$ of a fitting hyperplane $y = \beta^T x + \beta_0$ by solving the following problem:

$$\min_{\beta\in\mathbb{R}^{289},\beta_0\in\mathbb{R},\varepsilon\in\mathbb{R}^n} \frac{1}{2}\beta^T\beta + C\sum_{i=1}^{n}\varepsilon_i^2$$

$$\text{s.t.}\quad \delta + \varepsilon_i \geq y^i - \left(\beta^T x^i + \beta_0\right) \geq -\delta - \varepsilon_i, \varepsilon_i \geq 0, \quad i = 1, \ldots, n, \qquad (3)$$

where $\delta, C > 0$ are given parameters. This problem can be solved efficiently using a primal–dual interior-point method, especially for large-scaled cases.

2.4. FIR model

The fuzzy regression model was initially formulated by Tanaka, Uejima, and Asai (1982). The fundamental difference between the assumptions of fuzzy and linear regression models lies in the deviations between the observed and estimated values. The MLR model assumes these values to be errors in measurement (or observations), while the fuzzy regression model assumes that they depend on the indefiniteness of the system structure. A recent development in fuzzy regression for load forecasting was made by Hong and Wang (2014), who discovered the importance of the underlying linear model for the accuracy of a fuzzy regression model. The FIR model that they proposed also used the vanilla model as the underlying linear model.

Following Hong and Wang (2014), a possibilistic linear function can be defined as $y = A^T x + A_0$, where $x = (x_1, x_2, \ldots, x_{289})^T$ is non-fuzzy, $A = (A_1, A_2, \ldots, A_{289})^T$, and $A_i, i = 0, 1, \ldots, 289$ is a symmetric triangular fuzzy number that is characterized by a fuzzy membership function $u_{A_i}(a_i) = max\left(0, 1 - |a_i - \beta_i| / c_i\right), i = 0, 1, \ldots, 289$, with β_i and $c_i > 0$ as the center and the spread, respectively. For the training data set $\left\{\left(x^i, y^i\right), i = 1, \ldots, n\right\}$, the following FIR model is solved to find the centered fitting hyperplane $y = \beta^T x + \beta_0$ for load forecasting:

$$\min_{\beta\in\mathbb{R}^{289},\beta_0\in\mathbb{R},c\in\mathbb{R}^{289},c_0\in\mathbb{R}} \sum_{i=1}^{n}\left(c^T\left|x^i\right| + c_0\right)$$

$$\begin{aligned}\text{s.t. } |1-h|\left(c^T\left|x^i\right| + c_0\right) &\geq y^i - \left(\beta^T x^i + \beta_0\right)\\ &\geq -|1-h|\left(c^T\left|x^i\right| + c_0\right), \qquad (4)\end{aligned}$$

$$i = 1, \ldots, n, \quad c_j \geq 0, \quad j = 0, 1, \ldots, 289,$$

where $0 \leq h < 1$, $\left|x^i\right| = \left(\left|x_1^i\right|, \left|x_2^i\right|, \ldots, \left|x_{289}^i\right|\right)^T$ and $c = (c_1, c_2, \ldots, c_{289})^T$. This problem can also be solved efficiently by an interior-point method. In the numerical experiments of this paper, h is set to be 0.1, without loss of generality.

Note that the regression line provided by the FIR model is determined primarily by these observations on the boundary of the input data range. Fig. 2 depicts an example where the x-axis represents the temperature and the y-axis represents the load. The three subfigures all have the same scales on each axis. In each subfigure, the orange curve depicts the FIR function, estimated using the observations in both circles and dots. The dots represent the boundary observations that affect the FIR estimation process, while the circles represent the interior observations that are ignored by the FIR estimation process. The left panel shows the original data. The middle panel shows downward-shifted inner observations, which do not change the orange curve at all. The right panel shows downward-shifted upper boundary observations, which actually lead to a downward shift in the orange curve.

3. Benchmarking framework and setup

This section introduces the framework and setup of our benchmarking case study, including the GEFCom2012 data, the accuracy of load forecasting models using the original data, and the simulation of data integrity attacks.

Table 1
MAPEs (%) of hourly load forecasts in 2007 without data integrity attacks.

	Zone	MLR	ANN	SVR	FIR
Aggregated zone	21	**5.22**	5.69	5.23	5.54
Regular zone	1	**7.01**	8.88	7.02	8.14
	2	5.62	5.99	**5.61**	6.36
	3	5.62	6.19	**5.61**	6.36
	5	**9.88**	10.80	9.93	13.11
	6	**5.55**	6.34	**5.55**	6.20
	7	5.62	6.15	**5.61**	6.36
	8	7.50	8.57	**7.47**	8.40
	10	**6.70**	7.39	6.75	7.80
	11	**7.70**	9.46	7.75	8.05
	12	**6.78**	8.45	6.88	7.77
	13	**7.39**	9.46	7.40	8.35
	14	**9.38**	11.08	9.48	10.76
	15	**7.44**	9.36	7.47	8.27
	16	**8.12**	9.74	8.24	9.65
	17	**5.26**	6.41	5.27	5.83
	18	**6.72**	7.79	6.77	7.27
	19	**7.90**	10.28	7.96	8.78
	20	**5.74**	6.67	5.75	6.45
Special zone	4	16.08	17.72	**16.06**	19.72
	9	139.16	128.82	140.04	**110.66**

3.1. GEFCom2012 data

The dataset used in this study is from the load forecasting track of GEFCom2012 (Hong, Pinson et al., 2014), which has been used widely in the forecasting community (Ben Taieb & Hyndman, 2014; Charlton & Singleton, 2014; Høverstad, Tidemann, Langseth, & Öztürk, 2015; Lloyd, 2014; Nedellec, Cugliari, & Goude, 2014; Wang et al., 2016). The dataset includes 4.5 years of hourly load and temperature information for a US utility with 21 zones ($Z_1, \ldots, Z_{21}$), where the load in Z_{21} is the sum of the other 20 zones ($Z_1, \ldots, Z_{20}$). In this paper, we take three full calendar years (2005–2007) for our empirical study. The first two years (2005 and 2006) are used as training data for estimating the parameters of the aforementioned four models, while the third year (2007) is used as test data for comparing the forecast accuracies. The temperature for each zone is the simple average of the weather stations selected using the methodology proposed by Hong, Wang, and White (2015). The time series plot of the load series is provided by Wang et al. (2016).

3.2. Benchmarking performance without data integrity attacks

For each zone, we test the four models and calculate their forecast errors. Table 1 shows the mean absolute percentage errors (MAPE) of the four models for the testing period (year 2007) in a normal situation (no data integrity attacks). Here, MAPE is specified as:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|, \tag{5}$$

where A_t and F_t are the actual and forecasted hourly loads at time t, respectively. A smaller MAPE value indicates that the corresponding model produces more accurate forecasts. The tables in this paper display the smallest MAPE values of the best models in bold, while cells with MAPE values of 10% or higher are filled in gray. We use 10% as the cut-off here for the sake of both simplicity and practicality, since a 10% MAPE at the aggregated level for the hourly load forecasts in one year would typically be considered to be too inaccurate for practical use. Table 1 shows that MLR is the most accurate overall, followed by SVR, FIR, and finally ANN. The rest of this paper focuses on Z_{21} in order to avoid verbose presentation, because similar observations are obtained for the other zones.

3.3. Framework for simulating data integrity attacks

We simulate a normally-distributed (or uniformly-distributed) data integrity attack on the training dataset by randomly selecting k% of all data points and altering their loads by multiplying them by $1 + p\%$ to make them anomalies, where p is generated by a normal distribution N (or uniform distribution U) with mean μ and standard deviation σ. For example, Figs. 3 and 4 depict the hourly load profiles under normally-distributed and uniformly-distributed data integrity attacks, respectively, for one week in the summer of 2006, with $k = 30$, $\mu = 30$, and $\sigma = 50$.

Without loss of generality, the next section tests the impacts of data integrity attacks with $\mu \geq 0$. Similar observations can be obtained under data integrity attacks with $\mu < 0$, though they are not shown in this paper in order to save space. When $\mu \geq 0$, the load forecasts provided by the forecasting models will probably be higher than the nominal. These over-forecasts may cause power companies unnecessary expenses for the upgrade and maintenance of infrastructure.

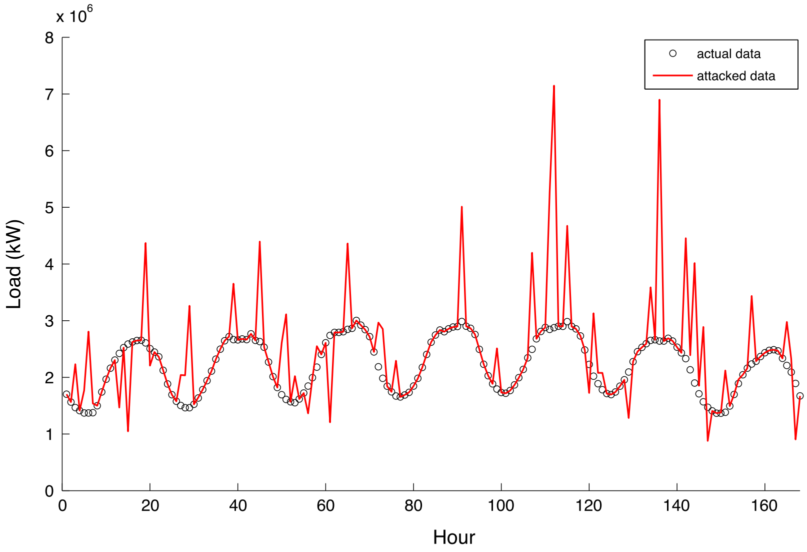

Fig. 3. The hourly load profile (2006/7/30–2006/8/5) under normally-distributed data integrity attacks.

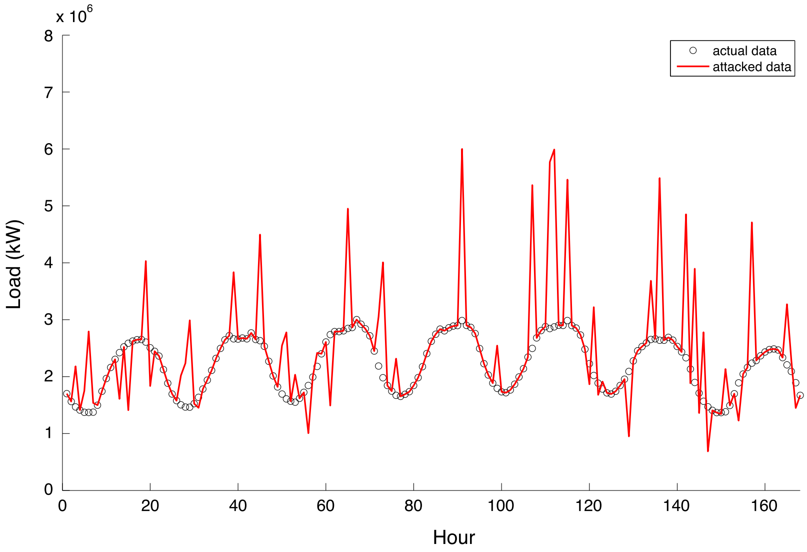

Fig. 4. The hourly load profile (2006/7/30–2006/8/5) under uniformly-distributed data integrity attacks.

4. Numerical results

The four load forecasting models (MLR, ANN, SVR and FIR) are compared using simulated training datasets under various levels of data integrity attacks, as described in the previous section. The original test dataset is used to calculate the MAPE values for benchmarking the accuracy. All of the following experiment results are based on Z_{21} only.

The four models are implemented using MATLAB (R2014a) on a personal computer equipped with Intel Core i5 2.40 GHz CPU, 4 GB usable RAM and Microsoft Windows 8 Professional. Table 2 lists the modules that are used to implement each model. When implementing the ordinary least squares estimator for MLR, the “wfun” input of the “robustfit” module was set to “ols”.

4.1. Normally-distributed data integrity attacks

This subsection investigates the performances of the four load forecasting models under data integrity attacks with normally distributed distortions.

4.1.1. Varying the mean and standard deviation of data integrity attacks

We first set the percentage of observations attacked to be 30%, i.e., $k = 30$. We then set the values of the mean μ of distortion to 0, 20 and 40, and the standard deviation σ

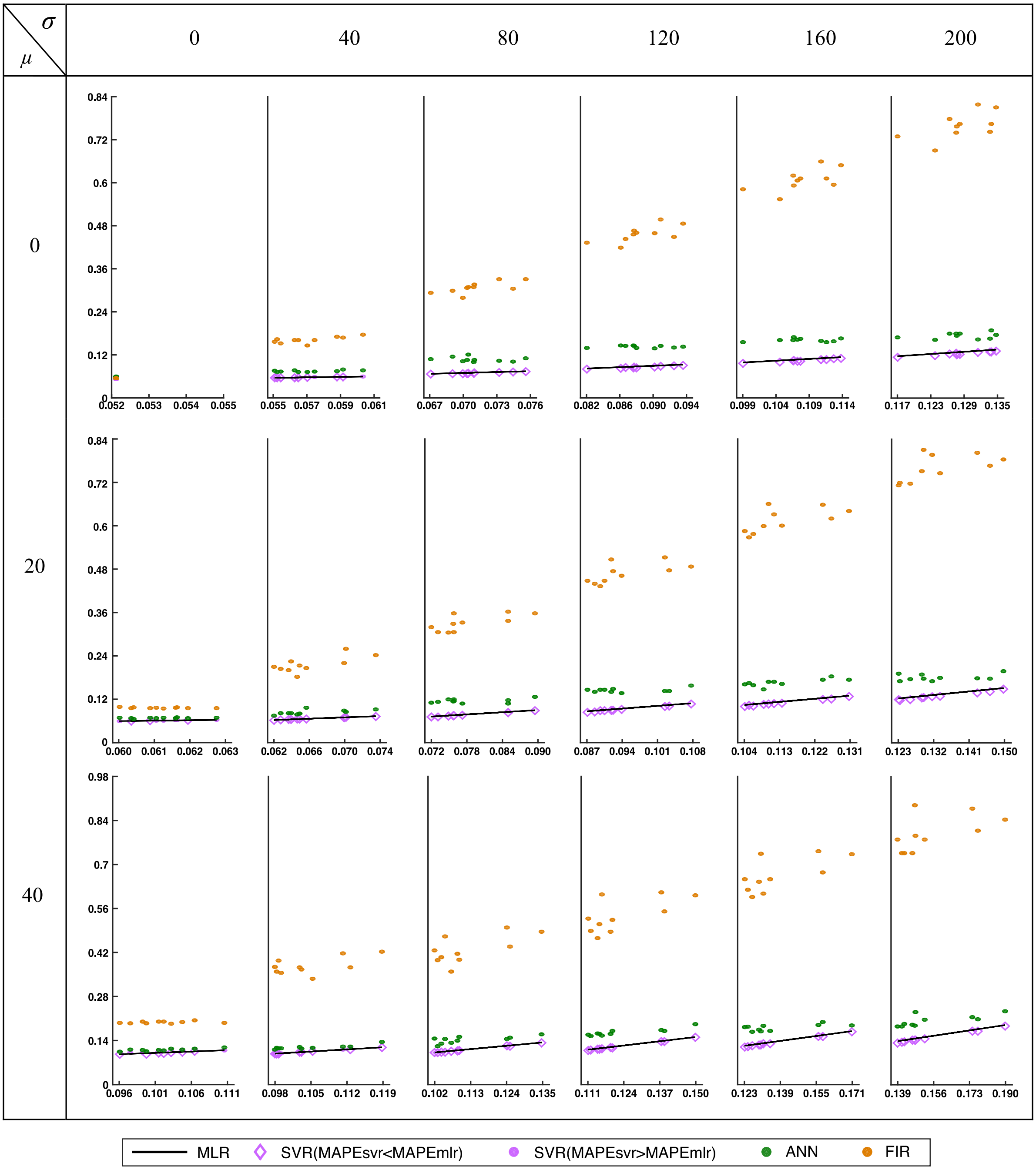

Fig. 5. Scatterplots of MAPE values under normally-distributed attacks with varying means and standard deviations.

to be between 0 and 200 by increments of 40, making 18 scenarios for each model. Within each scenario, we create 10 test cases by increasing the magnitude of randomly-selected load values by p %, where p is generated randomly by $N(\mu,\sigma^2)$. For all 18 scenarios, Table 3 shows the average MAPE values of the 10 test cases from each model. Fig. 5 depicts the MAPE values of all test cases in a panel of scatterplots with three rows (for $\mu = 0, 20, 40$) and six columns (for $\sigma = 0, 40, 80, 120, 160, 200$). Each scatterplot shows the relationship between the MAPE values of the four models (vertical axis) and those of the MLR model (horizontal axis). The MAPEs are the forecast errors of Z_{21} over the testing period (year 2007). Each marker represents one of the 10 corresponding test cases. The six scatterplots in each row all share the same vertical axis. The MLR results are shown with black lines. Since the MAPE values of SVR

Table 2
MATLAB modules used in the implementation.

Model	MATLAB module
MLR	robustfit
ANN	Neural network toolbox
SVR	quadprog
FIR	linprog

are quite close to those of MLR, we use the purple diamond markers to represent the cases where SVR outperforms MLR, and the purple dots to indicate otherwise. The MAPE values of all test cases are also recorded in a spreadsheet that is provided as supplementary material to this paper.

Based on Table 3 and Fig. 5, we make the following observations:

(1) Among all of our models, the SVR model demonstrates the strongest robustness to increases in μ and σ, as it has the slowest increase in MAPE values, followed closely by the MLR model. The ANN and FIR models rank third and fourth, respectively.
(2) None of the four models are robust enough to provide accurate load forecasts as μ or σ becomes large (such as $\mu = 40$ or $\sigma = 160$).
(3) Increases in σ have less of an impact on the performances of all models than increases in μ. In other words, the models are more sensitive to the average increase in magnitude than to the standard deviation of the increase.

4.1.2. Varying the percentage of injected data and the mean of data integrity attacks

We investigate the effect of the number of injected data points on the performances of our load forecasting models by varying k from 10 to 50, with increments of 10. Meanwhile, we vary μ from 0 to 40, with increments of 10, where σ is 50 or 100. This leads to 50 scenarios for each model. Similarly to the settings in Section 4.1.1, 10 test cases are created within each scenario. Table 4 shows the average MAPE values, while Figs. 5 and 6 depict the MAPE values of all test cases for $\sigma = 50$ and 100, respectively. The MAPE values of all test cases are also recorded in the supplementary material. Based on Table 4 and Figs. 6 and 7, we make the following observations:

(1) As the number of injected data points increases, the SVR model leads to the smallest MAPE values in most cases, meaning that it is the most robust of these four models, followed closely by the MLR model. The ANN and FIR models again rank third and fourth, respectively.
(2) Given a fixed value of k , the MAPE values of the FIR model increase the most significantly as either μ or σ increases. In other words, the accuracy of the FIR model deteriorates much more rapidly than those of the other three models as μ or σ increases. On the other hand, for a fixed (μ,σ) pair, the accuracy of the FIR model does not change much as k varies. The main reason for this is that the FIR model considers only data points that are on the boundary of the input data range (see Fig. 2 and the discussion in Section 2.4), which is driven mainly by the (μ,σ) pair.
(3) Given a fixed σ (or μ), the performances of MLR, ANN and SVR models deteriorate more rapidly with respect to the increase of k for large values of μ (or σ) than for small values of μ (or σ).
(4) For small values of k (such as 10 or 20), the forecast errors of MLR and SVR models decrease as μ increases from 0 to 10 (or 20). This is because there is a trend of increasing electric loads from 2005 to 2007. The minor data integrity attacks (such as $\mu = 10$, $\sigma = 50$) on the load data (of the years 2005–2006) help to offset the bias in the load forecast for the year 2007.
(5) None of the four models are robust enough to provide accurate load forecasts for large values of μ or k (such as $\mu = 30$ or $k = 40$).

Fig. 8 shows the forecasted hourly load profiles for one representative week in the winter of year 2007, under the situation of data integrity attacks with $k = 30$, $\mu = 30$ and $\sigma = 50$. We can see that, on average, all models over-forecast the actual load for the testing period. Moreover, the forecasts provided by the MLR model almost overlap with those provided by the SVR model. Overall, they are much closer to the actual loads than the forecasts provided by the ANN and FIR models are.

4.2. Uniformly-distributed data integrity attacks

We now generate the performances of the four models using simulated data integrity attacks with a uniformly-distributed distortion U. To be consistent with the numerical experiments in Section 4.1.2, we consider $U(-0.87,0.87)$, $U(-0.77,0.97)$, $U(-0.67,1.07)$, $U(-0.57,1.17)$, and $U(-0.47,1.27)$, corresponding to the normally-distributed data integrity attacks of $N(0,0.5^2)$, $N(0.1,0.5^2)$, $N(0.2, 0.5^2)$, $N(0.3,0.5^2)$ and $N(0.4, 0.5^2)$, respectively, in Section 4.1.2.

Following Section 4.1.2, we vary k from 10 to 50 with increments of 10. Fig. 9 depicts the MAPE values of all test cases in the 5 × 5 panel of scatterplots. From top to bottom, the five rows are for $U(-0.87,0.87)$, $U(-0.77,0.97)$, $U(-0.67,1.07)$, $U(-0.57,1.17)$ and $U(-0.47,1.27)$. From left to right, the five columns are for $k = 10$, 20, 30, 40 and 50. The MAPE values of all test cases are also recorded in the spreadsheet that is provided as supplementary material to this paper. Table 5 shows the average MAPE values. Similar results can be observed in Section 4.1.2. The ranking of model robustness from strong to weak is SVR, MLR, ANN and FIR. Fig. 10 shows the results for the same week as in Fig. 8, where the uniformly-distributed data attack $U(-0.57,1.17)$ distorts 30% of the training data. While all models over-forecast on average, the forecasts provided by the MLR and SVR models, which are almost identical, are much closer overall to the actual load. Again, the forecasts provided by the FIR model are farthest away from the actual loads.

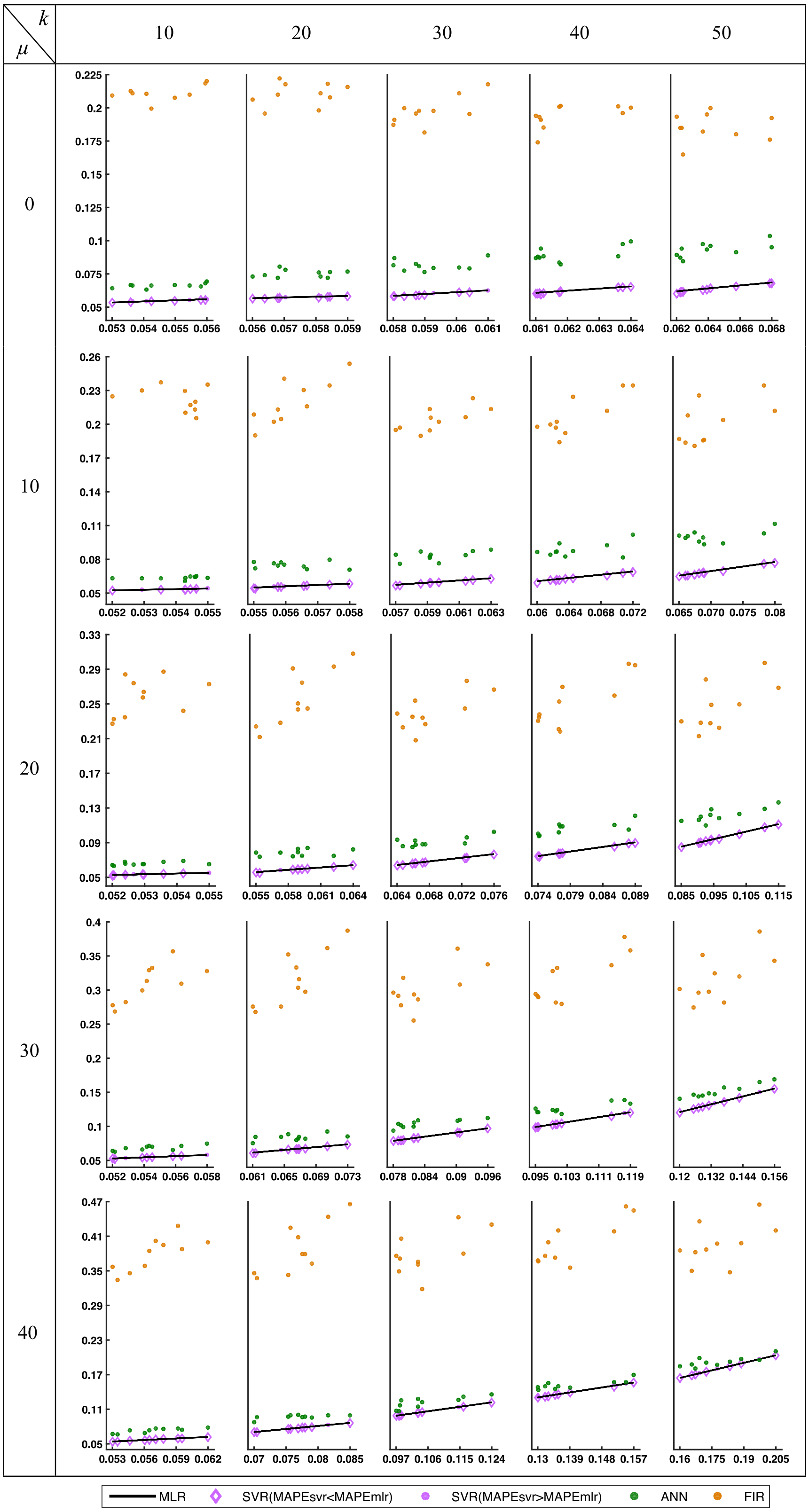

Fig. 6. Scatterplots of MAPE values under normally-distributed attacks with varying amounts of injected data ($\sigma = 50$).

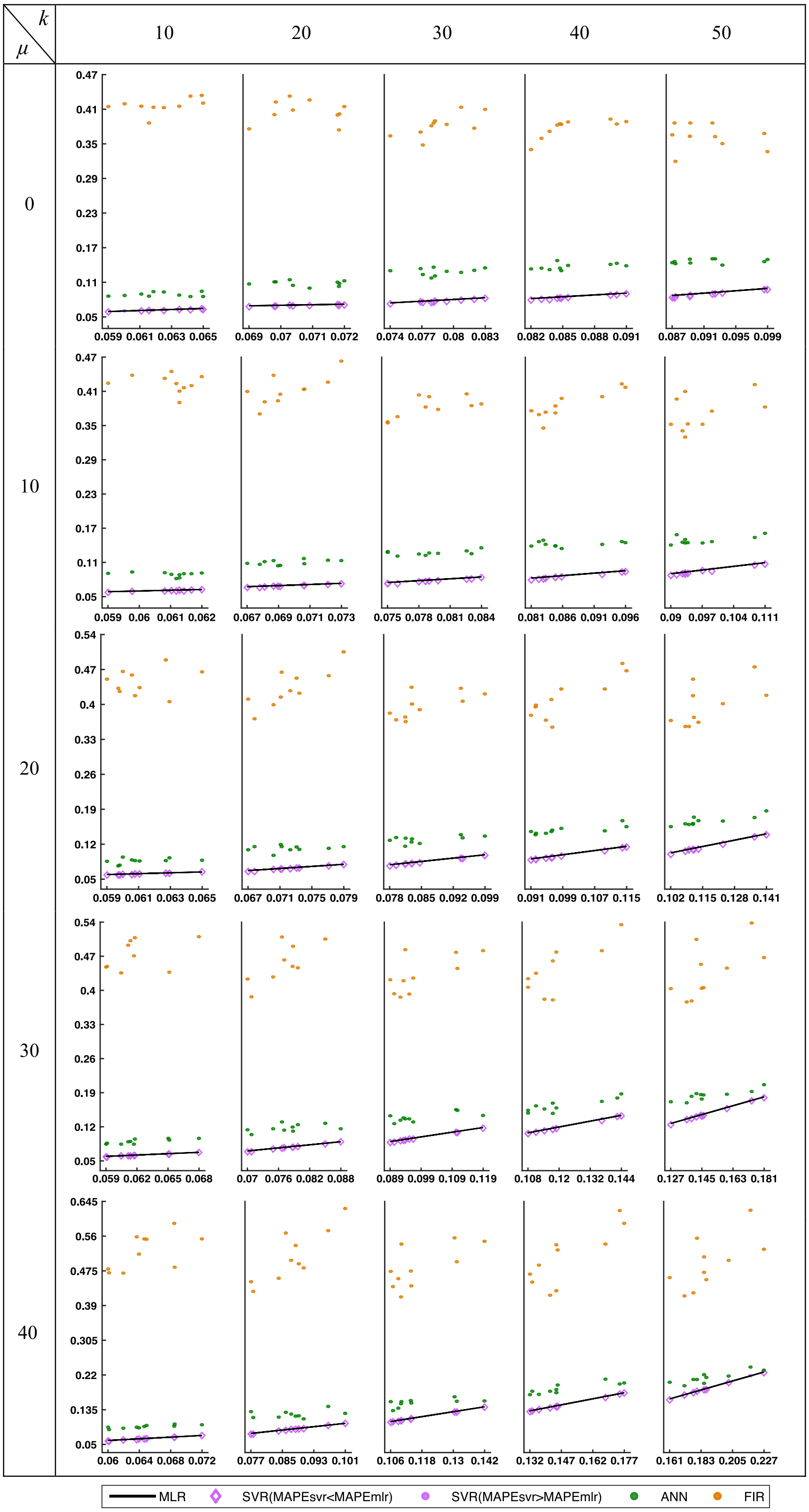

Fig. 7. Scatterplots of MAPE values under normally-distributed attacks with varying amounts of injected data ($\sigma = 100$).

Table 3
Average MAPEs (%) under normally-distributed data attacks with varying means and standard deviations.

	μ \ σ	0	40	80	120	160	200
MLR	0	**5.22**	5.74	7.06	8.81	10.78	12.86
ANN		5.69	7.46	10.72	14.27	16.11	17.37
SVR		5.23	**5.72**	**6.94**	**8.57**	**10.42**	**12.39**
FIR		5.54	16.16	30.75	45.69	60.79	73.99
MLR	20	**6.09**	6.62	7.84	9.48	11.38	13.41
ANN		6.80	8.39	11.52	14.47	16.62	18.09
SVR		6.10	**6.59**	**7.71**	**9.24**	**11.00**	**12.91**
FIR		9.61	21.63	33.10	46.87	61.40	76.06
MLR	40	**10.17**	10.48	11.21	12.36	13.83	15.53
ANN		11.06	11.85	14.21	16.59	18.13	20.38
SVR		**10.17**	**10.45**	**11.10**	**12.15**	**13.50**	**15.08**
FIR		19.78	37.70	42.93	53.71	65.30	79.83

Table 4
Average MAPEs (%) under normally-distributed data attacks with varying amounts of injected data.

		$\sigma = 50$					$\sigma = 100$				
	μ \ k	10	20	30	40	50	10	20	30	40	50
MLR	0	5.49	5.76	6.01	6.24	6.44	6.23	7.09	7.90	8.59	9.18
ANN		6.62	7.52	8.13	8.96	9.32	8.88	10.82	12.67	13.79	14.62
SVR		**5.48**	**5.73**	**5.97**	**6.18**	**6.36**	**6.16**	**6.97**	**7.72**	**8.37**	**8.91**
FIR		21.25	21.02	19.74	19.37	18.53	41.63	40.58	38.24	37.77	36.24
MLR	10	5.36	5.63	5.99	6.41	6.96	6.12	7.00	7.92	8.76	9.67
ANN		6.36	7.48	8.31	8.84	10.02	8.90	11.00	12.55	13.96	14.89
SVR		**5.34**	**5.60**	**5.94**	**6.34**	**6.88**	**6.05**	**6.87**	**7.73**	**8.52**	**9.40**
FIR		22.22	21.94	20.40	20.77	20.06	42.37	41.25	38.20	38.60	37.15
MLR	20	5.36	5.95	6.87	8.03	9.60	6.12	7.27	8.62	10.00	11.70
ANN		6.59	7.82	9.08	10.63	12.18	8.66	11.16	12.99	14.77	16.65
SVR		**5.34**	**5.92**	**6.82**	**7.96**	**9.54**	**6.04**	**7.14**	**8.44**	**9.77**	**11.47**
FIR		25.76	25.70	24.07	25.16	24.65	44.46	43.27	39.75	41.16	39.81
MLR	30	5.49	6.69	8.49	10.70	13.54	6.23	7.88	9.93	12.14	14.94
ANN		6.83	8.38	10.42	12.66	15.18	8.99	11.75	13.91	16.41	18.46
SVR		**5.47**	**6.65**	**8.43**	**10.65**	**13.50**	**6.15**	**7.75**	**9.76**	**11.95**	**14.77**
FIR		30.98	31.70	30.25	31.69	31.78	47.97	46.67	43.25	45.24	43.77
MLR	40	5.75	7.78	10.62	13.99	18.05	6.46	8.80	11.73	14.96	18.96
ANN		7.31	9.68	12.15	15.23	19.25	9.33	12.36	15.25	18.28	21.36
SVR		**5.72**	**7.74**	**10.57**	**13.94**	**18.03**	**6.37**	**8.67**	**11.58**	**14.80**	**18.85**
FIR		37.91	38.89	37.99	39.94	39.67	52.27	51.11	48.32	50.65	49.34

4.3. Performance analysis for robustness

In the aforementioned test cases, the forecasting performances of the four models under normally-distributed and uniformly-distributed data integrity attacks are compared with those for the GEFCom2012 load data in three dimensions: the percentage (k %) of the load data being perturbed maliciously, and the mean (μ%) and standard deviation (σ%) of the magnitude of the perturbation. The overall performances of the four representative models, with and without data integrity attacks, are ranked in Table 6, with 1 being the most robust and 4 the least robust.

Several observations can be made from Table 6:

(1) It is important to note that the most accurate forecasting model for clean data usage is not necessarily the most robust model under data integrity attacks. See for example MLR vs. SVR, or FIR vs. ANN.

(2) In general, the SVR and MLR models are more robust than the other two. The SVR model outperforms MLR model slightly, largely because SVR has the additional objective of maximizing the "confidence margin" of the fitted hyperplane (i.e., minimizing the term $\frac{1}{2}\beta^T\beta$ in its objective function of Eq. (3) in Section 2.3), as well as minimizing the squared fitting errors of all training points.

(3) The FIR model is much less robust than the other three models, because the regression line provided by the FIR model is determined primarily by the observations sitting on the boundary of the input data range, as was discussed in Section 2.4. This is the region that is the most likely to be affected by data integrity attacks. In contrast, the other three models consider all training data points, of which only some are attacked. At the same time, the ANN model also fails to work well because it soon faces the problem of overfitting.

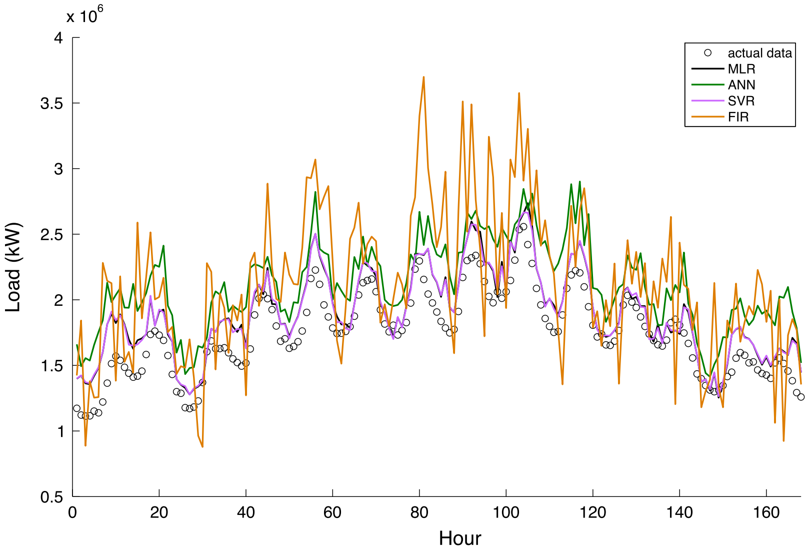

Fig. 8. Hourly load forecasts (2007/1/7–2007/1/13) under normally-distributed data integrity attacks.

Table 5
Forecast error in MAPEs (%) under uniformly-distributed data attacks.

	μ \ k	10	20	30	40	50
MLR	0	5.45	5.67	5.93	6.19	6.43
ANN		6.64	7.37	8.16	9.23	9.77
SVR		**5.45**	**5.64**	**5.88**	**6.11**	**6.34**
FIR		11.25	12.03	10.09	10.16	10.98
MLR	10	5.34	5.54	5.86	6.32	6.82
ANN		6.63	7.59	8.30	9.05	10.44
SVR		**5.34**	**5.51**	**5.78**	**6.23**	**6.71**
FIR		13.85	14.09	13.10	13.10	13.40
MLR	20	5.37	5.88	6.70	7.92	9.37
ANN		6.73	7.79	9.00	10.34	12.37
SVR		**5.36**	**5.83**	**6.61**	**7.84**	**9.28**
FIR		19.61	20.46	20.61	21.26	21.91
MLR	30	5.53	6.64	8.31	10.61	13.27
ANN		6.87	8.66	10.59	12.86	15.16
SVR		**5.49**	**6.59**	**8.23**	**10.55**	**13.22**
FIR		28.73	29.41	29.92	30.76	31.56
MLR	40	5.81	7.75	10.45	13.90	17.77
ANN		7.29	9.79	12.26	15.44	18.90
SVR		**5.79**	**7.70**	**10.39**	**13.86**	**17.69**
FIR		36.69	39.02	39.75	40.60	41.51

As any one of the three factors (i.e., k % of data, μ% and σ% of actual load) involved in data integrity attacks increases, all four of these representative load forecasting models may fail to provide satisfying forecasts. Hence, it becomes imperative to develop robust anti-attacking models for accurate forecasting.

5. Directions for future research

As the first of its kind, this paper may inspire new research ideas for future study. This section provides some immediate next steps for consideration, with the hope that more researchers may get involved in this field to make further contributions.

This paper has modeled data integrity attacks as randomly injected shocks to the loads in form of a normal or uniform distribution for benchmarking the performances of four representative load forecasting models. Although the results clearly point out the shortfalls of these models for providing robust forecasts, obtaining a more thorough understanding of the problem would require a more extensive simulation framework that allows other types of data integrity attacks, such as attacks on the peak loads, the ramping periods, and so forth.

Our benchmark results have shown that all four of the representative load forecasting models fail to generate robust forecasts under severe data integrity attacks. One would naturally look for the development of a more robust methodology under data integrity attacks. Although this paper does not provide a direct answer, it does shed some

Table 6
Rankings of the overall accuracy of the forecasting models.

	Without data integrity attacks	Under data integrity attacks
MLR	1	2
ANN	4	3
SVR	2	1
FIR	3	4

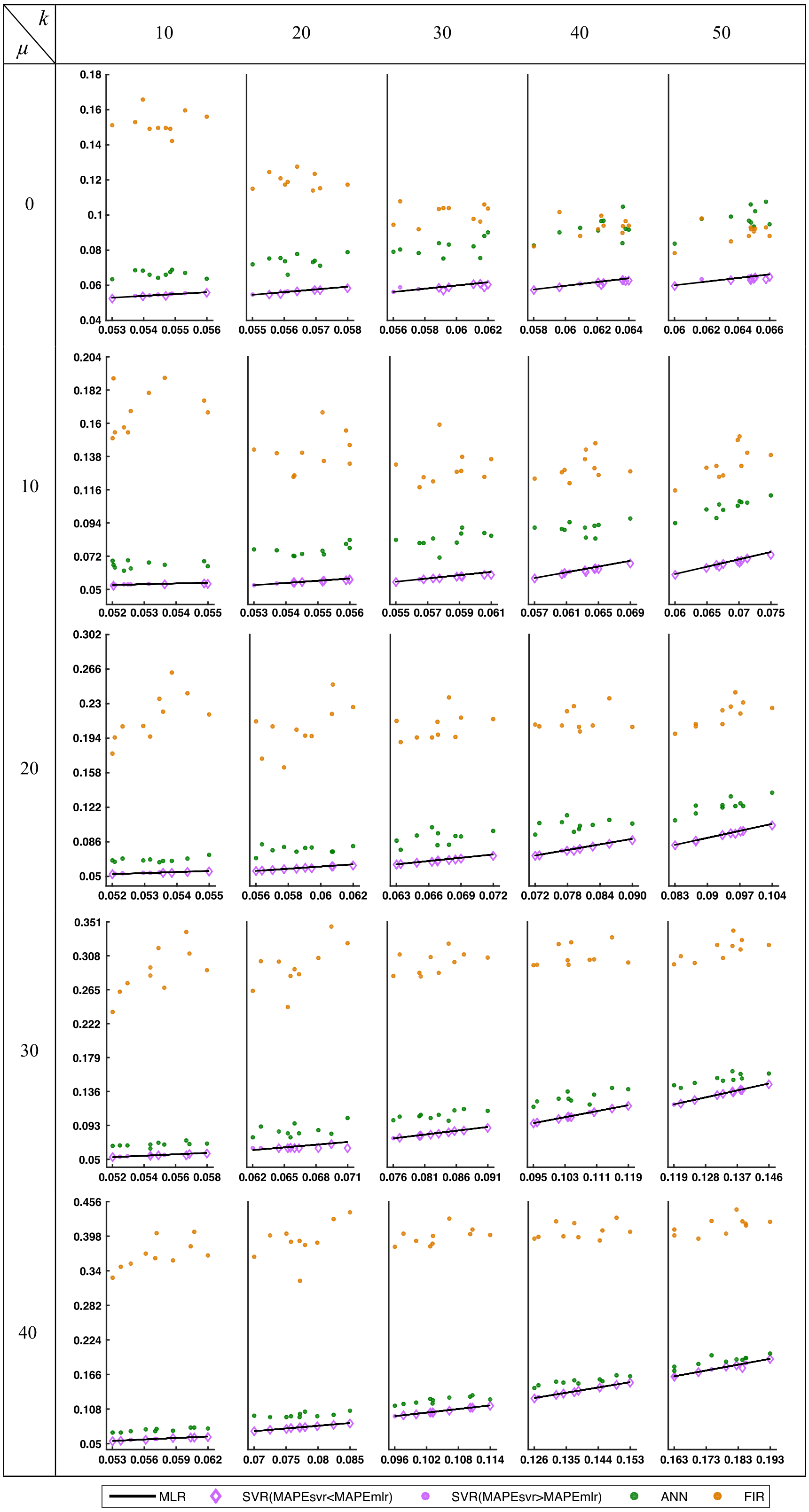

Fig. 9. Scatterplots of MAPE values under uniformly-distributed attacks.

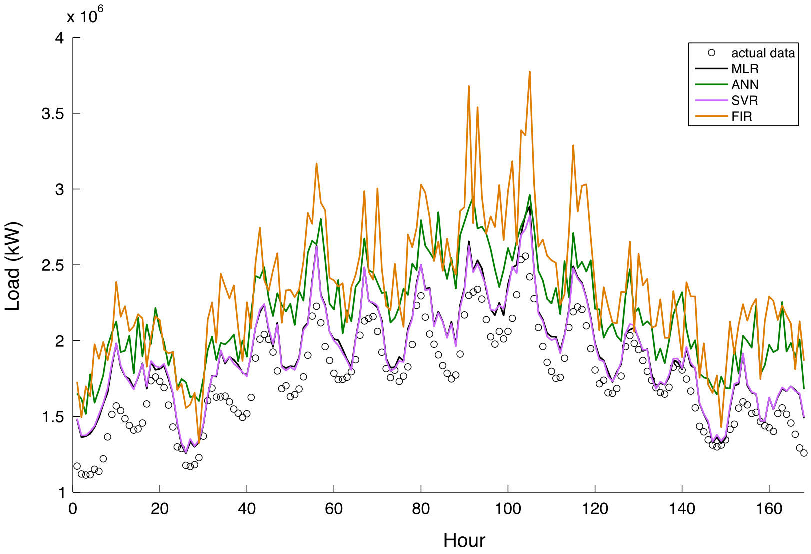

Fig. 10. Hourly load forecasts (2007/1/7–2007/1/13) under uniformly-distributed data integrity attacks.

light on the development of a potential solution. As was discussed in Section 4.2, the addition of the term $\frac{1}{2}\beta^T\beta$ to the objective function of the SVR model helps it to outperform the MLR model. In other words, an alternative objective function could lead to a more robust model. One branch of the family of regression analysis is robust regression (Rousseeuw & Leroy, 2005). Instead of taking the ordinary least square estimator, an alternative is to use the least absolute deviation (LAD), known as LAD regression (or L_1 regression, from minimizing l_1-norm). Note that L_1 regression is essentially the quantile regression on the 50th percentile, and quantile regression is not new to the load forecasting literature, especially probabilistic load forecasting (Hong & Fan, 2016). Other alternatives such as the M-estimator, the least trimmed squares (LTS) estimator, the S estimator, and the MM estimator are also good candidates.

Although the data integrity attacks in this paper have been applied to the load data, it is possible that the predictors, such as weather data, could also come under attack. Since weather is a key driving factor of electricity demand, an attack on the weather data is very likely to lead to significant forecast errors. Hence, another possible direction for future research would be to study data integrity attacks on the predictors. One potential solution could be to treat such values as measurement errors. Researchers in statistics and econometrics have studied measurement error models extensively (Carroll, Ruppert, Stefanski, & Crainiceanu, 2006; Fuller, 2006). Nevertheless, it is not clear whether these methods would help load forecasting models to stay robust to data integrity attacks.

All of the case studies presented in this paper consider point (or single-valued) load forecasting. As probabilistic load forecasting is gaining an increasing amount of attention from both the research community and industry (Hong & Fan, 2016), it would be worthwhile to investigate the robustness of these models for probabilistic load forecasting. Xie and Hong (2017) recently attempted to connect the point and probabilistic forecasting performances of load forecasting models by comparing two variable selection methods. Thus, the next research question could be: can a robust point load forecasting model also be robust for probabilistic load forecasting?

This paper focuses on the robustness aspect. In other words, we are seeking models that are prone to attacks. An alternative approach to addressing data integrity attacks in load forecasting would involve trying to mitigate the attacks. As was mentioned in Section 1, we need to be able to detect attacks, identify maliciously manipulated data, and cleanse and recover attacked data. However, like computer viruses, which evolve over time, data integrity attacks to the grid data may evolve as such mitigation methods are improved.

In addition to the data integrity attacks discussed in this paper, other cyberattacks that target privacy and data availability (Gu, Yang, Jirutitijaroen, Walsh, & Reindl, 2014) would also be worth studying. Similarly, we can add the security layer of research to other energy forecasting problems beyond load forecasting, and even to other forecasting fields beyond energy forecasting.

6. Conclusion

This paper has introduced the issue of data integrity attacks to the forecasting community. We begin by proposing a framework for simulating some data integrity attack scenarios, in which hackers may maliciously inject multipliers to the historical load data. By testing four representative load forecasting models using the GEFCom2012 data under different levels of simulated data integrity attacks, we found that the SVR model is the most robust with respect to the point forecast accuracy, closely followed by the MLR model, while the FIR model is the least robust.

It is important to note that all four models fail to provide accurate load forecasts when the data points are severely contaminated. The forecast MAPEs provided by the four models all exceed 10% if 40% of the historical load data points are deliberately increased by a large magnitude. This finding presents a serious question for the forecasting community: *How can robust forecasts be generated under data integrity attacks*?

As the first attempt to tackle load forecasting under data integrity attacks, this paper offers a set of benchmark results for the interest of readers. Further investigations into model performances under different data integrity attack scenarios (such as data integrity attacks on the weather and calendar data) and into the design of more robust load forecasting models are of particular importance to the field. Moreover, this study may open the door to a consideration of the problem of data integrity attacks in other similar forecasting fields such as renewable energy, finance, and retail forecasting.

Acknowledgments

The authors would like to express their appreciation to the editors and reviewers for helping to improve the quality of the manuscript during the peer review process.

The work was supported in part by the National Natural Science Foundation of China Grant #71701035, the US Department of Energy, Cybersecurity for Energy Delivery Systems (CEDS) Program under Work Order Number M616000124, and US Army Research Office Grant Number W911NF-15-1-0223.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2017.08.004.

References

Akouemo, H. N., & Povinelli, R. J. (2016). Probabilistic anomaly detection in natural gas time series data. *International Journal of Forecasting*, *32*(3), 948–956.

Beale, M. H., Hagan, M. T., & Demuth, H. B. (2016). *Neural network toolbox user's guide*. Mathworks.

Ben Taieb, S., & Hyndman, R. J. (2014). A gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting*, *30*(2), 382–394.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective*. CRC Press.

Charlton, N., & Singleton, C. (2014). A refined parametric model for short term load forecasting. *International Journal of Forecasting*, *30*(2), 364–368.

Chen, B.-J., Chang, M.-W., & Lin, C.-J. (2004). Load forecasting using support vector machines: A study on EUNITE competition 2001. *IEEE Transactions on Power Systems*, *19*(4), 1821–1830.

Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, *27*(1), 134–141.

Fuller, W. A. (2006). *Measurement error models*. John Wiley & Sons.

Gu, C., Yang, D., Jirutitijaroen, P., Walsh, W. M., & Reindl, T. (2014). Spatial load forecasting with communication failure using time-forward kriging. *IEEE Transactions on Power Systems*, *29*(6), 2875–2882.

Hagan, M.T., Demuth, H.B., Beale M.H., & De Jesús, O. (2014). Neural network design. (2nd ed.) Martin Hagan.

Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, *16*(1), 44–55.

Hodge, V. J., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, *22*(2), 85–126.

Hong, T. (2010). *Short term electric load forecasting*. North Carolina State University, (Ph.D. thesis).

Hong, T. (2014). Energy forecasting: past, present, and future. *Foresight: The International Journal of Applied Forecasting*, *32*, 43–48.

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, *32*(3), 914–938.

Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, *30*(2), 357–363.

Hong, T., & Wang, P. (2014). Fuzzy interaction regression for short term load forecasting. *Fuzzy Optimization and Decision Making*, *13*(1), 91–103.

Hong, T., Wang, P., & White, L. (2015). Weather station selection for electric load forecasting. *International Journal of Forecasting*, *31*(2), 286–295.

Hong, T., Wilson, J., & Xie, J. (2014). Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid*, *5*(1), 456–462.

Høverstad, B. A., Tidemann, A., Langseth, H., & Öztürk, P. (2015). Short-term load forecasting with seasonal decomposition using evolution for parameter tuning. *IEEE Transactions on Smart Grid*, *6*(4), 1904–1913.

Hu, J., & Vasilakos, A. V. (2016). Energy big data analytics and security: Challenges and opportunities. *IEEE Transactions on Smart Grid*, *7*(5), 2423–2436.

Hyndman, R. J., & Fan, S. (2010). Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems*, *25*(2), 1142–1153.

Khotanzad, A., & Afkhami-Rohani, R. (1998). ANNSTLF-artificial neural network short-term load forecaster generation three. *IEEE Transactions on Power Systems*, *13*(4), 1413–1422.

Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2004). *Applied linear statistical models*. McGraw-Hill/Irwin.

Liu, Y., Ning, P., & Reiter, M. K. (2011). False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security*, *14*(1), 1–33.

Lloyd, J. R. (2014). GEFCom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes. *International Journal of Forecasting*, *30*(2), 369–374.

Nedellec, R., Cugliari, J., & Goude, Y. (2014). GEFCom2012: Electric load forecasting and backcasting with semi-parametric models. *International Journal of Forecasting*, *30*(2), 375–381.

Papalexopoulos, A. D., & Hesterberg, T. C. (1990). A regression-based approach to short-term system load forecasting. *IEEE Transactions on Power Systems*, *5*(4), 1535–1547.

Perez, E. (2016). First on CNN: U.S. investigators find proof of cyberattack on Ukraine power grid. Retrieved from http://www.cnn.com/2016/02/03/politics/cyberattack-ukraine-power-grid/.

Ramanathan, R., Engle, R., Granger, C. W. J., Vahid-Araghi, F., & Brace, C. (1997). Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, *13*(2), 161–174.

Rousseeuw, P., & Leroy, A. (2005). *Robust regression and outlier detection*. John Wiley & Sons.

Singer, P. W., & Friedman, A. (2014). *Cybersecurity and cyberwar: What everyone needs to know*. Springer.

Tanaka, H., Uejima, S., & Asai, K. (1982). Linear regression analysis with fuzzy model. *IEEE Transactions on Systems Man and Cybernetics*, *12*(6), 903–907.

Wang, P., Liu, B., & Hong, T. (2016). Electric load forecasting with recency effect: A big data approach. *International Journal of Forecasting*, *32*(3), 585–597.

Weron, R. (2006). *Modeling and forecasting electricity loads and prices: A statistical approach*. John Wiley & Sons.

Xie, J., & Hong, T. (2016). GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation. *International Journal of Forecasting*, *32*(3), 1012–1016.

Xie, J., & Hong, T. (2017). Variable selection methods for probabilistic load forecasting: Empirical evidence from seven states of the United States. *IEEE Transactions on Smart Grid*. http://dx.doi.org/10.1109/TSG.2017.2702751.

Jian Luo is an assistant professor of management science and engineering at Dongbei University of Finance and Economics. His research interests include machine learning techniques with applications, electric load forecasting, and fuzzy and nonlinear optimization. He received his Bachelors and Masters degrees in applied maths from Wuhan University in China in 2007 and 2009, and his Ph.D. in Industrial Engineering from North Carolina State University in USA in 2014.

Tao Hong is an associate professor of systems engineering and engineering management and NCEMC Faculty Fellow of Energy Analytics at the University of North Carolina at Charlotte. His research interests include forecasting and its industrial applications, such as energy, healthcare, sports and transportation. He received his Bachelor of Engineering in Automation from Tsinghua University in 2005 and his Ph.D. in operations research and electrical engineering from North Carolina State University in 2010.

Shu-Cherng Fang is the Walter Clark Chair and Alumni Distinguished Graduate Professor in the Department of Industrial & Systems Engineering and the Gradute Program in Operations Research at NC State University. His key research interests are in optimization thery and algorithms with real life applications, such as intelligent human-machine decision support systems, terrain data representation, logistics and supply chain management, telecommunications and bio-informatics. Prof. Fang received his Bachelors degree from the National Tsing Hua University in Taiwan and his Ph.D. from Northwestern University in the US.