Short-term forecasting of electric load based on Kalman filter with elastic net method

Y J Tian1,3, H X Shen2 and W G Gao3

1Electric Power Science Test and Research Institute, State Grid Shanghai Power Supply Company, 171 Handan Road, Shanghai 200437, China
2School of Mathematical Sciences, Fudan University, Shanghai 200433, China
E-mail: 13901712348@163.com

Abstract. The short-term forecasting model based on Kalman filter is an important class of predictive algorithm widely used. This paper discusses a self-adaptive Kalman filter. The main idea of the paper is applying elastic net method and covariance matrices estimation to parameter identification, which can select variables of state with linear regression to make the model close to reality. Finally, this paper gives a simulation experiment to prove the effectiveness and feasibility of the model in practice.

1. Introduction
The short-term forecasting of electric load is a basic requirement for reliable operation of the electric energy system. Accurate forecasting is tightly related to the economic benefits, automated operation and scientific management of power grid. The methods of short-term forecasting of electric load have been developed for a long time, and the effectiveness of various methods have been proved, such as expert system [1], wavelet transform method [2], neural network method [3] and so on.

Kalman filter theory was proposed in 1960 and it can be applied to the state-space models with low algorithm complexity, which is easy to be implemented and suitable for both stationary and non-stationary stochastic process [4]. Therefore, researchers have developed many short-term forecasting models of electric load based on Kalman filter. For instance, self-adaptive Kalman filter can be applied to estimating the noise covariance matrices to improve predictive accuracy [5], and spline smoothing is used to project the states on the elements of the basis [6]. The combination of time series analysis and Kalman filter can determine the dimension of the state and improve forecasting delays [7].

This paper is based on self-adaptive Kalman filter, and introduces elastic net method and estimation of covariance matrices to parameter identification, which can select variables of state with linear regression and improve predictive accuracy in practice.

2. Kalman filter
The forecasting of load is using previous load values and temperature to obtain the future load. Assume that we have such an equation:

\[ L(k+1) = \sum_{j=0}^{k-2} a_j(k)L(k-j) + a_{n-1}(k)T(k+1) + u(k+1), \]  

(1)

where \( L(k) \) means the load value at time \( k \), \( T(k) \) means the temperature at time \( k \), \( a_j(k) \) is
coefficient and $u(k)$ is a Gaussian random variable. In such situation, Kalman filter can be used as a predictor [4]. Considering the system in the state space form below,

$$
X(k + 1) = F(k)X(k) + w(k),
Y(k) = H(k)X(k) + v(k),
$$

(2)

where $X(k) \in \mathbb{R}^n$ is the state vector, $Y(k) \in \mathbb{R}^m$ is the observation vector, $F(k) \in \mathbb{R}^{nxn}$ is the state transition matrix, $H(k) \in \mathbb{R}^{mxn}$ is the observation matrix, $w(k) \in \mathbb{R}^n$ is the state noise and $v(k) \in \mathbb{R}^m$ is the observation noise. Both state noise and observation noise are supposed to have the following properties:

$$
E(w(k)) = 0 \quad \forall k, \quad E(w(k)w^T(l)) = \begin{cases} Q(k), & k = l \\ 0, & k \neq l \end{cases}
$$

(3)

$$
E(v(k)) = 0 \quad \forall k, \quad E(v(k)^Tv(l)) = \begin{cases} R(k), & k = l \\ 0, & k \neq l \end{cases}
$$

(3)

$$
E(w(k)v^T(l)) = 0 \quad \forall k, \forall l.
$$

At every time $k$, the following steps are reiterated:

$$
\hat{X}(k + 1 | k) = F(k)\hat{X}(k | k)
$$

$$
P(k + 1 | k) = F(k)P(k | k)F^T(k) + Q(k)
$$

$$
K(k + 1) = P(k + 1 | k)H^T(k + 1)\left[ H(k + 1)P(k + 1 | k)H^T(k + 1) + R(k) \right]^{-1}
$$

$$
\hat{X}(k + 1 | k + 1) = \hat{X}(k + 1 | k) + K(k + 1)\left[ Y(k + 1) - H(k + 1)\hat{X}(k + 1 | k) \right]
$$

$$
P(k + 1 | k + 1) = [I_n - K(k + 1)H(k + 1)]P(k + 1 | k)
$$

(4)

where $K(k)$ is the filter gain matrix and $P(k + 1 | k)$ is the prediction covariance. The initial condition is $X(0 | 0) = X_0$, $P(0 | 0) = P_0$, which is not quite important since the filter values will converge to the true values with the increasing of time $k$. Thus, we apply Kalman filter to the forecasting and let

$$
X(k) = \begin{bmatrix} L(k) & L(k - 1) & \cdots & L(k - n) & T(k) \end{bmatrix}^T,
$$

(5)

$$
H(k) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix},
$$

(6)

$$
F(k) = \begin{bmatrix}
    a_0(k) & a_1(k) & \cdots & a_{n-1}(k) & a_{n-2}(k) & a_{n-3}(k) \\
    1 & 0 & \cdots & 0 & 0 & 0 \\
    0 & 1 & \cdots & 0 & 0 & 0 \\
    0 & 0 & \cdots & 0 & 0 & 0 \\
    0 & 0 & \cdots & 1 & 0 & 0 \\
    0 & 0 & \cdots & 0 & 0 & 1 
\end{bmatrix}.
$$

(7)

Due to the parameters such as $a_j(k), Q(k)$ and $R(k)$ unknown, we have to identify the parameters by using the observations $Y(k)$ in order to construct the Kalman filter.
3. Parameter identification
The parameter identification in our model can be divided into two steps. First, we apply linear regression method to estimating the state transition matrix $F(k)$. Then, we use self-adaptive Kalman filter to modify the noise covariance matrices $Q(k)$ and $R(k)$.

3.1. Elastic net method
Elastic net method is based on LASSO (least absolute shrinkage and selection operator). LASSO is a regression analysis method that performs both variable selection and regularization [8]. The basic form is to solve the optimization problem

$$
\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \| \beta \|_1,
$$

where $x_i$ and $y_i$ are the input and output of the $i$th sample, $\beta$ and $\beta_0$ are the coefficients of the input and intercept, and $\lambda > 0$ is a parameter which is usually calculated by cross-validation. The $l_1$ regularization can efficiently reduce the order of the predictive model. In other words, LASSO can select which variables are useful in the model during regression process [9]. Therefore, we apply LASSO to the estimate of coefficients $a_j(k)$ to reduce over-fitting and to clarify the essential variables [13].

Moreover, Ridge regularization can be often added to the form of LASSO (6) in practice. The method is called elastic net regularization, which can stabilize the $l_1$ regularization path [14]. The new form to find $F^*(k)$ is that

$$
F^*(k) = \arg \min_{F(k)} \frac{1}{N} \sum_{i=1}^{N} (L(k + 1 - i) - H(k)F(k)X(k - i))^2
$$

$$
+ \lambda [\alpha \| H(k)F(k) \|_1 + (1 - \alpha) \| H(k)F(k) \|_2],
$$

where $\alpha \in (0,1]$.

3.2. Noise covariance matrices
The estimation methods of noise covariance matrices $Q(k)$ and $R(k)$ in state-space models have been developed for a long time [10]. In our model, we combine two techniques to give an estimation at every time $k$. By using ICM (indirect correlation method) [11], we have the equation

$$
e(k) = Y(k) - H(k)\hat{X}(k | k - 1),
$$

$$
\hat{R}(k) = E[e(k)e^T(k)] - H(k)P(k | k - 1)H^T(k).
$$

Applying CMM (covariance matching method) [12] to the estimation of $Q(k)$, we have

$$
\hat{Q}(k) = (1 - z(k))\hat{Q}(k - 1) + z(k) [K(k)e(k)e^T(k)K^T(k)
$$

$$
+ P(k | k) - F(k - 1)P(k - 1 | k - 1)F^T(k - 1)],
$$

where $z(k) \in [0,1]$ is a parameter which determine the rate of update of $\hat{Q}(k)$.

The combination of those two techniques gives an unbiased estimation. However, if the dimension of $X(k)$ is high, the filter values may diverge, while $\hat{Q}(k)$ and $\hat{R}(k)$ are not positive semidefinite [5]. Therefore, we give a biased estimation form,
\[ \hat{R}(k) = \frac{1}{k} \sum_{i=1}^{k} e(i) e^T(i), \]
\[ \hat{Q}(k) = (1 - z(k)) \hat{Q}(k-1) + z(k) \left[ K(k) e(k) e^T(k) K^T(k) + P(k | k) \right]. \] (12)

4. Simulation

In order to test the forecasting model based on Kalman filter, we select 80 days between Nov. 2017 and Feb. 2018 and apply our algorithm to a station load data at every 30 minutes in Shanghai. There are 32 time points in each day after dropping some time points at which some of the load data is missing. We use our model to forecast the load in last day with the previous load data at same time point, and the relative error of forecasting at each time point is calculated with the form,

\[ \varepsilon(k) = \frac{(Y(k) - \hat{Y}(k))}{Y(k)} \times 100\% \] (13)

where \( Y(k) \) is the real load at time \( k \) in the selected day and \( \hat{Y}(k) \) is the forecasting of the load. Moreover, MAPE (mean absolute percentage error) is calculated with the form,

\[ MAPE = \frac{1}{N} \sum_{j=1}^{N} |\varepsilon(k)| \] (14)

**Figure 1.** Forecasting results and the real values.  **Figure 2.** Relative errors \( \varepsilon(k) \) by elastic net.

Let \( n = 7, \alpha = 0.5 \) and we obtain the results in figures 1 and 2. Figure 1 presents the real load and predicted value at time point \( k \) in the selected day. Figure 2 shows the relative error \( \varepsilon(k) \). From the results, we can find that the relative errors at most of time are less than 7% and \( MAPE_{EN} = 5.27\% \), which can be accepted in practice. Then, we apply least square method instead of elastic net method to the same data to estimate \( F(k) \), and the relative errors are in figure 3. From the results, we can find that some of the relative errors are so high that we cannot accept the forecasting error though there exist some time points where the relative errors of least square method is smaller than those of elastic net method, and \( MAPE_{LS} = 5.82\% \). Compared with our model, we can find \( a_j(k) \) estimated by elastic net method has many zero values but those estimated by least square method are not.
Moreover, we construct a LSTM (long short term memory) neural network to forecast the load with the past 7 days load data [15]. The results are in figures 1 and 4. The MAPE of LSTM is 8.48%, which is higher than our Kalman filter method with elastic net method. It shows that our model is more effective dealing with non-stationary series like the electric load data we used. Therefore, applying elastic net method to parameter identification may be a good choice.

5. Conclusion
This paper has presented an application of Kalman filter to the short-term forecasting of the load shape with the station load data in Shanghai. The main idea is parameter identification of the state-space model, and the paper has introduced two methods into the forecasting model.

- Elastic net method is applied to obtaining the state transition matrix $F(k)$. The algorithm can efficiently select the main variables during regression process and always reduce the dimension of state $X(k)$, which means that it may save computing resources during filtering.
- ICM and MCM are combined to estimate the noise covariance matrices $Q(k)$ and $R(k)$. The modified self-adaptive estimation can improve predictive accuracy and avoid divergence of filtering.

Finally, the paper uses the simulation to prove effectiveness and feasibility of our model based on Kalman filter.

Acknowledgments
The research of this work is supported by Science and Technology Commission of Shanghai Municipality 16JC1402600 and State Grid Shanghai Electric Power Company 52094017001X.

References
[1] Saifur R and Bhatnagar R 1988 An expert system based algorithm for short term load forecast IEEE Transactions on Power Systems 3 392-9
[2] Tai N L, Hou Z J, Li T et al 2003 New principle based on wavelet transform for power system short-term load forecasting Proceedings of the CSEE 1 45-50
[3] Hippert, Henrique S, Carlos E and Reinaldo C 2001 Neural networks for short-term load forecasting: A review and evaluation IEEE Transactions on Power Systems 16 44-55
[4] Harvey and Andrew C 1990 Forecasting, Structural Time Series Models and the Kalman Filter (Cambridge: Cambridge University Press)
[5] Ma J B and Yang H G 2005 Application of adaptive Kalman filter in power system short-term
load forecasting *Power System Technology* **29** 75-9

[6] Komi N, Jairo C and Julien J 2018 Short-term electricity demand forecasting using a functional state space model *Energies* **11** 1-24

[7] Shi W Q, Wu K Y, Wang D X *et al* 2018 Eclectic power system short-term load forecasting model based on time series analysis and Kalman filter algorithm *Techniques of Automation and Applications* **37** 9-12+23

[8] Tibshirani R 1996 Regression shrinkage and selection via the lasso *Journal of the Royal Statistical Society B* **58** 267-88

[9] Tibshirani R, Martin W and Trevor H 2015 *Statistical Learning with Sparsity: The Lasso and Generalizations* (New York: Chapman and Hall/CRC)

[10] Duník J *et al* 2017 Noise covariance matrices in state-space models: A survey and comparison of estimation methods—Part I *International Journal of Adaptive Control and Signal Processing* **31** 1505-43

[11] Mehra and Raman 1972 Approaches to adaptive filtering *IEEE Transactions on Automatic Control* **17** 693-8

[12] Sage A and Husa G 1969 Adaptive filtering with unknown prior statistics *Joint Automatic Control Conference* **7** 760-9

[13] Hisashi T, Yoshiyasu T and Seisho S 2016 Using the ensemble Kalman filter for electricity load forecasting and analysis *Energy* **104** 184-98

[14] Hui Z and Trevor H 2005 Regularization and variable selection via the elastic net *Journal of the Royal Statistical Society B* **67** 301-20

[15] Chen Z and Sun L X 2018 Short-term electrical load forecasting based on deep learning LSTM networks *Electronic Technology* **47** 39-41