**Proceedings Paper:**
Zhang, J, Li, K orcid.org/0000-0001-6657-0522 and Chai, T (2021) Demand Forecasting of a Fused Magnesia Smelting Process Based on LSTM and FRA. In: Communications in Computer and Information Science. 6th International Conference on Life System Modeling and Simulation, LSMS 2020, and 6th International Conference on Intelligent Computing for Sustainable Energy and Environment, ICSEE 2020, 25 Oct 2020, Hangzhou, China. Springer Nature , pp. 201-215. ISBN 978-981-33-6377-9

https://doi.org/10.1007/978-981-33-6378-6_15

# Demand forecasting for a fused magnesia smelting process via LSTM and FRA

Jingwen Zhang[1], Kang LI[2], Tianyou Chai[1]

[1] State Key Laboratory of Synthetical Automation for Process Industries College of Information Science and Engineering,Northeast University,Shenyang 110819,China,
{1610277@stu.neu.edu.cn,tychai@mail.neu.edu.cn}
[2] School of Electronic and Electric Engineering,University of Leeds,Leeds , UK,
{K.Li1@leeds.ac.uk}

**Abstract.** The electric power demand of a Fused Magnesia Smelting Process(FMSP) is defined as the average of electric power consumed within a fix period of time and is used for judging the electricity cost. The power has to be turned off once the demand value exceeds one specific threshold $\bar{P}_l$. However, by taking advantage of the current controller in FMSP, the demand can be reduced hence there is no need to switch off the power supply in this case. Therefore, it is vital to forecast the power demand and its trend such that the power supply does not have to be turned off unnecessarily. In this paper, a hybrid model combining a linear model with an unknown high order function is proposed. The linear model built to capture the priori information from the domain knowledge and historic data, while the unknown information in FMSP and the error of the linear model are approximated with a high order nonlinear function. The Recursive Least Square algorithm(RLS) is used for identifying the unknown parameters in the linear model . A Long-Short Term Memory(LSTM) trained by the Fast Recursive Algorithm(FRA) is proposed to fit the unknown high-order function. Finally, the output weights of LSTM is updated by RLS online. Comparing with hybrid models like the linear model combined with Radial Basis Function Neural Network(RBF), the proposed model offers the best performance in our experimental studies.

**Keywords:** Demand forecasting, LSTM, FRA

## 1 Introduction

Fused magnesia is a key refractory matter used in many fields like the chemical industry, metallurgy and aerospace industry, etc.. To produce the fused magnesia, magnesia is smelted by a Fused Magnesium Furnace(FMF) which is an energy intensive. The energy cost takes over 60% of the total production cost. The electricity demand is used to judge the electric energy cost during the Fused Magnesia Smelting Process(FMSP) and it can not exceed a certain limit $\bar{P}_l$ namely the maximum demand which is specified by State Grid Corporation of China(SGCC), otherwise the factory will face huge fine. For this reason if the demand exceeds this limit for 4 sampling periods(28 seconds) continuously, one or more FMFs have to be turned off to ensure the total demand is below the maximum limit at the 5th sampling time. But turning off the power frequently, the product quality will reduce, leading to unexpected economic loss. Fortunately, the current controller designed for the FMF can adjust the length of the arc to reduce the current when the demand exceeds its limit, hence reducing the overall demand. But some unknown effects, such as unknown process dynamics and external disturbances, etc, may affect the performance of the current controller, hence the demand may necessarily not decrease. For these aforementioned reasons, it is vital to have an accurate forecasting of the demand and its trend to help the controller take appropriate control actions, and to determine if it is necessary to turn off FMFs. The FMF uses submerged arc mode to achieve the melting point of the raw material about 2850°C, which is much higher than 1700 °C - the temperature required by the steel-making electric arc furnace [1]. Using this mode we can only measure the electric power and demand during the smelting process. However, some unavailable information and data like the length of the arc, the height of the smelting pool and etc. also affect the model accuracy of the model. But fortunately, the past demand and electric power consumption are the major contributing factors of the model.

Over past decades time series models [2], regression models [3], Artificial Neural Networks(ANN) [4] or some hybrid models have been used to predict the electricity demand and its trend. Time series models are the simplest model used for energy demand forecasting and some statistical methods are usually used for trend analysis of the time series data. Regression models with threshold is proved to have better performance than standard linear model and it can be used for both long term and short term forecasting. In [5] ANN is used to forecast peak load planning. In [6] a hybrid model with linear and non-linear parts is proposed to forecast the electric demand and its trend. In [7] the electric demand and its peak are predicted by a hybrid model with regression models and (Artificial Neural Network) ANNs. To forecast the demand in FMSP, in [8] a Radial Basis Function Neural Network(RBFNN) has been used,and in [1, 9] the linear model with RBFNN produced by different identification methods are used. By comparing with [1, 8, 9], we find hybrid model with linear model and neural network is more suitable used for the demand forecasting in FMSP, but the demand forecasting accuracy is affected by the single neural network and the statistic method for selecting the order of the input in [1, 8, 9].

In [1, 9] a mechanism model which combines with a linear model and an unknown high-order function is proposed to help capture the dynamics of FMSP. The linear model uses known past electric power as its inputs, the unknown information and the error of linear model in the smelting process can be written as an unknown high-order function with the past modelling errors and electric power consumptions as the model inputs. The biggest challenge for forecasting the demand and its trend is that there are overwhelmingly unknown information during the FMSP because of the submerged arc mode. The single RBFNN structure also limits the accuracy of the model. So there is a need to build a more accurate demand neural network with better forecasting performance.

In this paper we propose a hybrid model with a linear model and an unknown high order function to forecast the demand value and its trend. The parameters of the linear model will be identified by Recursive Least Squares(RLS) [12], and the Long-Short Term Memory(LSTM) [10] is chosen to fit the unknown high order function. A Fast Recursive Algorithm(FRA) [11] is used to build LSTM for identifying the output weights. The RLS is used for updating the output weights online. Comparing with some ANNs like RBFNN, the hybrid model with LSTM proposed in this paper has the best performance in our experimental studies using the real data from the FMSP.

## 2 Demand forecasting model

### 2.1 Introduction to FMSP

Figure 1 gives a brief illustration of the FMSP, which comprises a power supply system, current control systems, power systems, multiple FMFs and a demand monitoring system. Due to the high melting point of the raw materials, the FMF has to use the submerged arc mode (the electrodes are embedded into the raw material). The current control system will adjust the distance between the three-phase electrodes and the surface of the molten pool for generating the arc to melt the raw material and form a magnesium oxide bath. The smelting process is with adding raw material until the furnace is full.

There are a number of uncertainties associated with the FMSP which stop the current controller acting appropriately to adjust the current to the set point timely. For the cases when the current controller does not perform, we can turn off the FMF manually to stop the demand exceeding its limit. The power consumptions of FMFs can be collected by the power transducer and the demand will be calculated by a demand calculation device which calculates the current demand every seven seconds. With these information, the demand and its predictive value can be obtained. If the demand exceeds its limitation for 4 successive sampling points, the power of one or more FMFs will be cut off. Vice versa, the FMFs can be turned on when the demand is below the limit.

However if the power of FMFs are switched off and on frequently, the quality of the fused magnesia will be significantly decreased, and the energy waste can be huge as well.Therefore it is necessary to to model the demand accurately for the current controller such that it can action properly to avoid unnecessary power cut-off and economic losses.

### 2.2 Demand forecasting model

In FMSP, each FMF may have different current, smelting state and other state values, therefore ideally the demand of each FMF is modelled separately. But there exist numerous unavailable information in FMSP like the length of arc,

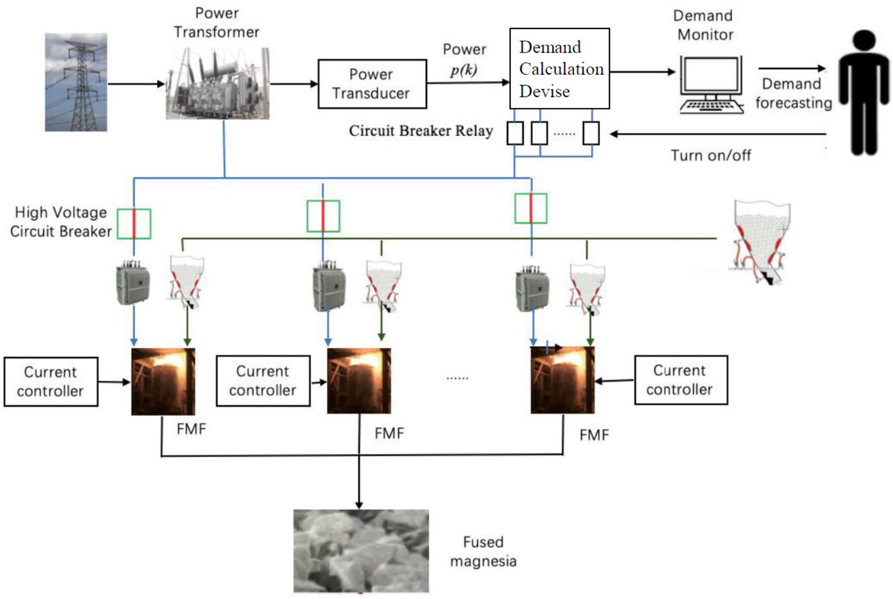


Fig. 1: A sketch of demand monitoring for FMSP.

distance between the three-phase electrodes and the surface of the molten pool and the height of the pool , etc., therefore a practical approach is to assume that all the FMFs have same state and we only need to predict the total demand of all FMFs. The demand $\bar{P}(k)$ of all FMFs at time instant $k$ is defined as the average value of electric power consumptions $p(k), p(k-1), \cdots, p(k-n+1)$ in equation(1).

$$\bar{P}(k) = \frac{1}{n}\sum_{i=1}^{n} p(k-i+1) \tag{1}$$

where $p(k) = \sum_{i=1}^{m} \sqrt{3}UI_i(k)\cos\varphi$ is the electric power of m FMFs at time instant $k$, $U$ is the voltage of each FMF, $I_i$ is the current of $i$th FMF, and $\cos\varphi$ is the power factor. According to [1], the electric power forecasting model can be formulated as in equation (2),

$$p(k+1) = \boldsymbol{\psi}(k)\boldsymbol{\theta} + v(k) \tag{2}$$

where $\boldsymbol{\psi}(k) = (p(k), p(k-1), p(k-2), p^*)$,$p^* = m\sqrt{3}UI^*\cos\varphi$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^T$ , $I^*$ is the set value of current and $v(k)$ is an unknown high order function which will be fitted by LSTM trained with FRA. We can rewrite equation (1) as follows

$$\bar{P}(k+1) = \bar{P}(k) + \frac{p(k+1) - p(k+1-n)}{n} \tag{3}$$

We can see that if we want to predict the demand $\bar{P}(k+1)$ at $k+1$, we have to predict the electric power $p(k+1)$ at $k+1$. So our target in this paper is to forecast the electric power $\hat{p}(k+1)$ firstly and then calculate the forecasting demand $\hat{\hat{P}}(k+1)$ using equation (3).

## 3 Demand forecasting algorithm

In this section we use Recursive Least Square(RLS) algorithm to identify the parameters $\boldsymbol{\theta}$ of linear model in (2) and the unknown high order nonlinear function $v(k)$ is fitted by LSTM trained with FRA. The initial value of $\boldsymbol{\theta}$ is can be identified by the Least Square(LS) algorithm. According to equation (2), we write linear model as equation (4) and we set $v(k) = 0$ for $k = 1, 2, \cdots$ at first. Then the parameter $\boldsymbol{\theta}$ will be identified by RLS until it is convergent and then it will be fixed as $\hat{\boldsymbol{\theta}}$. Afterwards, we fit $v(k), k = 1, 2, \cdots$ by LSTM trained with FRA and its value can be calculated by equation (5). Finally, an online parameter update algorithm will be given.

$$p(k+1) - \hat{v}(k) = \boldsymbol{\psi}(k)\boldsymbol{\theta} \tag{4}$$

$$v(k) = p(k+1) - \boldsymbol{\psi}(k)\hat{\boldsymbol{\theta}} \tag{5}$$

### 3.1 Offline parameter identification for linear model by RLS

We need to identify linear model parameter $\boldsymbol{\theta}$ in equation (4) with Least Square(LS) [12] algorithm firstly and update it by Recursive Least Square(RLS) [12] until it is convergent. For offline training, we rewrite equation (4) in matrix form as equation (6), where $t$ is training time, $Y_p = (p(1), p(5), \cdots, p(n_{train}))^T, \Psi = (\boldsymbol{\psi}(0); \boldsymbol{\psi}(1); \cdots; \boldsymbol{\psi}(n_{train} - 1))$ and $n_{train}$ is number of training data, $p(k) = 0$ for $k \leqslant 0$. $Y(t) = (v(1), v(2), \cdots, v(n_{train}))^T$ can be calculated by equation (7).

$$Y_p - Y(t) = \Psi\boldsymbol{\theta}(t) \tag{6}$$

$$Y(t) = Y_p - \Psi\hat{\boldsymbol{\theta}}(t) \tag{7}$$

The offline identification algorithm for parameter $\boldsymbol{\theta}$ is given as follows:
(1)Set $t = 0, Y(0) = \mathbf{0}$;
(2)Calculate the initial value of parameter $\boldsymbol{\theta}$ by LS algorithm $\hat{\boldsymbol{\theta}}(0) = \left(\Psi^T\Psi\right)^{-1}\Psi^T Y_p$;
(3)Calculate $Y(t)$ by equation (7). Update the parameter $\hat{\boldsymbol{\theta}}(t)$ using equation (8)

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + Y(t)G \tag{8}$$

where

$$G = \left(\Psi^T\Psi\right)^{-1}\Psi\left[I + \Psi\left(\Psi^T\Psi\right)^{-1}\Psi\right]^{-1}$$

and $I$ is an identity matrix;
(4) If $\hat{\boldsymbol{\theta}}(t) - \hat{\boldsymbol{\theta}}(t-1) = \xi, \xi \to 0$ end the training for linear model, else set $t = t+1$ and return to step (3).

We choose $\hat{\boldsymbol{\theta}}(t)$ as the parameter of linear model and calculate new error $v(k)$ by equation (7) as the target values for LSTM training .

### 3.2 LSTM trained with FRA for unknown high order nonlinear function

Traditional Recurrent Neural Network(RNN)(figure 2, left) in equation (9),(10) usually used for short-term time series trained by backpropagation through time(BPTT) [10], but error signals flowing backward in time tend to blow up or vanish with long-term time series.

$$h_t = g\left(W_{xh}x_t + W_{hh}h_{t-1} + b_h\right) \tag{9}$$

$$z_t = g\left(W_{hz}h_t + b_z\right) \tag{10}$$

where $x_t$ is input.

The LSTM (figure 2, right) is a variant of RNN with gates can be used for both long-term and short-term time series [10]. The LSTM updates for timestep $t$ given inputs $x_t, h_t, c_{t-1}$ are :

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + b_i\right) \tag{11}$$

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right) \tag{12}$$

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + b_o\right) \tag{13}$$

$$g_t = \phi\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \tag{14}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{15}$$

$$h_t = o_t \odot \phi\left(c_t\right) \tag{16}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is a sigmoid function in range $[0,1]$ , $\phi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ is a tanh function in range $[-1,1]$.

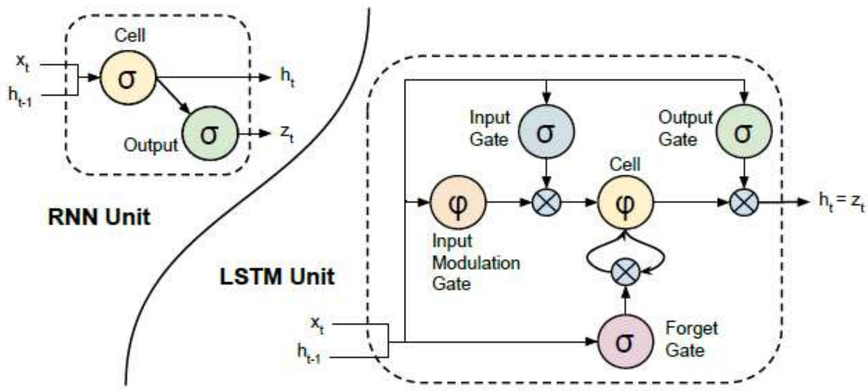


Fig. 2: A diagram of a basic RNN cell (left) and an LSTM memory cell (right) [13].

The LSTM with gates can avoid gradient blow up or vanish with over 1000 time steps, therefore the LSTM is chosen for fit unknown high order nonlinear function. The electric demand and electric power updates ever 7 seconds, so we can't update all parameters in LSTM online. To solve this, weights link the hidden notes of the last layer and the output trained by FRA is proposed in this paper and the weights can be updated by RLS online, so that the electric demand can be forecast in 7 seconds.

The inputs at $k$ of the unknown high order nonlinear function is $\boldsymbol{x}(k) = (p(k), \cdots, p(k-n_p), v(k-1), \cdots, v(k-n_v))$, the output $v(k)$ can be written as

$$v(k) = \boldsymbol{\beta}\boldsymbol{h}_t(k) + b \tag{17}$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m)$ is trained by FRA, $\boldsymbol{h}_t(k) = \left(h_t^1(k), \cdots, h_t^m(k)\right)^T$, $h_t^i(k) = tanh\left(h_{t-1}^i(k)\right)$, for $i = 1, \cdots, m$, $m$ is the number of the hidden notes of the last layer, $t$ is the number of the hidden layers. The hidden notes in $1, \cdots, t-1$ layers are LSTM units.

### 3.3 A Fast Recursive Algorithm(FRA)

A Fast Recursive Algorithm(FRA) designed in [11] is an alternative algorithm for orthogonal least square(OLS) algorithm to avoid Gram–Schmidt process in nonlinear system identification. We use FRA instead of LS to avoid calculating the pseudo inverse of the information matrix and FRA is described as follows.

Assume that there are $m$ hidden nodes in the last hidden layer, define $R_h$ as equation (18), where $\phi = (\phi_1, \cdots, \phi_m)$ is the information matrix and $I$ is an identity matrix.

$$R_h = \begin{cases} I, & h = 0 \\ I - \phi\left[(\phi)^T \phi\right]^{-1}(\phi)^T, & 1 < h < H_l \end{cases} \tag{18}$$

According to [11] $R_h$ can update iteratively by (19) for $h = 1, 2, \cdots, m-1$ and it avoids calculating the pseudo inverse of $\left[(\phi)^T \phi\right]$.

$$R_{h+1} = R_h - \frac{R_h\phi_{h+1}\left(R_h\phi_{h+1}\right)^T}{\left(\phi_{h+1}\right)^T R_h\phi_{h+1}} \tag{19}$$

In [11] there are some propositions of $R_h$ listed as (20)-(23) and the proofs can be found in [11].

$$(R_h)^T = R_h \tag{20}$$

$$(R_h)^2 = R_h \tag{21}$$

$$R_h R_j = R_j R_h = R_h, h \geq j \tag{22}$$

$$R_h\phi_i = 0, \forall i \in \{1, 2, 3, \cdots, m\} \tag{23}$$

As for traditional LS algorithm the parameters can be obtained by (24). Time $\phi$ on both sides of equation (24) and we will get equation (25)

$$\hat{\boldsymbol{\beta}}_i = \left[(\phi)^T \phi\right]^{-1} \phi Y \tag{24}$$

$$\phi\hat{\boldsymbol{\beta}} = \phi\left[(\phi)^T\phi\right]^{-1}\phi Y$$
$$= (I - R_h)\,Y \tag{25}$$

where $Y = (v(1), \cdots, v(k))$.

And then define an upper triangular matrix as $A = [a_{h,p}]_{h\times H_i}$ as equation (26), where $h = 1, 2, \cdots, m, p = h, h+1, \cdots, m$.

$$a_{h,p} = (\phi_h)^T\phi_{-}\sum_{j=1}^{h-1}\frac{a_{j,h}a_{j,p}}{a_{j,j}} \tag{26}$$

And $a_{h,y}$ as equation (27)

$$a_{h,y} = (\phi_h)^T Y - \sum_{j=1}^{h-1}\frac{a_{j,h}a_{j,y}}{a_{j,j}} \tag{27}$$

Multiply $(\phi)^T$ to both sides of equation (25) is equation (28).

$$(\phi)^T\phi\hat{\boldsymbol{\beta}} = (\phi)^T(I - R_h)\,Y \tag{28}$$

According to proposition (23), we can renew equation (25) as equation (29).

$$a_{h,h}\beta_h + \sum_{p=h+1}^{m}a_{j,p}\beta_p = a_{h,y} \tag{29}$$

By equation (29), the output weights $\beta_h$ can be calculated as equation (30), which can avoid calculating the pseudo inverse of the information matrix .

$$\hat{\beta}_h = \frac{a_{h,y} - \sum_{p=h+1}^{m}a_{h,p}\hat{\beta}_p}{a_{h,h}} \tag{30}$$

### 3.4 Demand forecasting in FMSP online

The $v(k)$ has to be written as follow for online forecasting

$$v(k) = \boldsymbol{\beta}(k)\boldsymbol{h}_t(k) + b(k) \tag{31}$$

We complex online forecasting by updating $\boldsymbol{\beta}(k), b(k)$ online.

When new electric power $p(k)$ and demand $\bar{P}(k)$ is measured at $k$, we fix linear model parameter $\boldsymbol{\theta}$ and update the output $v(k-1)$ of LSTM by equation (5). Output weights $\boldsymbol{\beta}$ of LSTM can be updated by RLS as followings.
(1)Set a sliding window with width $n_{train}$ ;
(2)Measure a new electric power $p(k)$ and demand $\bar{P}(k)$;
(3)Renew input matrix as

$$X(k) = [\boldsymbol{x}(k - n_{train} + 1), \cdots, \boldsymbol{x}(k)]\,, l = 1, 2, \cdots, L$$

output of LSTM as

$$Y(k) = [v(k - n_{train}), \cdots, v(k-1)]^T$$

(4)Update the information matrix $\Phi(k) = (\boldsymbol{h}_t(k - n_{train} + 1), \cdots, \boldsymbol{h}_t(k))$ of LSTM with $X$ ;
(5)Update $\boldsymbol{\beta}(k)$ and $b(k)$ by RLS as equation (32)

$$\left(\hat{\boldsymbol{\beta}}(k), \hat{b}(k)\right) = \left(\hat{\boldsymbol{\beta}}(k-1), \hat{b}(k-1)\right) + Y(k)G(k) \tag{32}$$

where

$$G(k) = P(k-1)\Phi(k)\left[I + \Phi^T(k)P(k-1)\Phi(k)\right]^{-1}$$
$$P(k) = P(k-1) - G(k)\Phi^T(k)P(k-1)$$

and $P(0) = \left(\Phi^T(k)\Phi(k)\right)^{-1}$;
(6)Forecasting $\hat{v}(k)$ by LSTM with $\hat{\boldsymbol{\beta}(k)}, \hat{b}(k)$ as equation (31), calculate $\hat{p}(k+1)$ by (2) and then calculate the predictive value of demand $\hat{\bar{P}}(k+1)$ by (3);
(7)Set $k = k+1$ at $k+1$ and return to step (2).

To sum up, we can predict the demand $\hat{\bar{P}}(k+1)$ in FMSP and the next electric power $\hat{p}(k+1)$ at $k$ by the aforementioned method. Furthermore, we can analyse the trend of demand for FMSP and evaluate the accuracy of the prediction by next section.

## 4 Accuracy and trend of demand forecasting analysis

We use Root Mean Square Error(RMSE) which is used in [1] to evaluate the accuracy of demand by our method in this paper. RMSE is described by equation (33), where $N$ is the number of observations.

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\bar{P}(k) - \hat{\bar{P}}(k+1)\right)^2} \tag{33}$$

We choose Mean Absolute Scaled Error(MASE) [?] in equation (34) as another evaluation for demand forecasting accuracy.

$$MASE = \frac{1}{N}\sum_{k=1}^{N}\frac{2\left|\bar{P}(k) - \hat{\bar{P}}(k+1)\right|}{\bar{P}(k) + \hat{\bar{P}}(k+1)} \tag{34}$$

The Direction Symmetry(DS) in equation (35) is used for analysing the trend of demand.

$$DS = \frac{1}{N}\sum_{k=1}^{N} d(k) \tag{35}$$

where $d(k)$ is as follows.

$$d(k)=\begin{cases}1, & \left[\bar{P}(k)-\bar{P}(k-1)\right]\left[\hat{\bar{P}}(k)-\hat{\bar{P}}(k-1)\right]\geq 0\\ 0, & \left[\bar{P}(k)-\bar{P}(k-1)\right]\left[\hat{\bar{P}}(k)-\hat{\bar{P}}(k-1)\right]<0\end{cases}$$

From DS we can't figure out the upward or downhill tendency of demand in a continuous period, and then True Positive Rate(TPR) in equation (36) and True Negative Rate(TNR) in equation (37) are used to solve that.

$$TPR=\frac{TP}{TP+FN} \tag{36}$$

$$TNR=\frac{TN}{TN+FP} \tag{37}$$

The relationship among parameters in equation (36) and (37) are listed in table 1.

Table 1: Relationship table

| $\bar{P}$ \ $\hat{\bar{P}}$ | Up | Down |
|---|---|---|
| Up | TP | FP |
| Down | TN | FN |

## 5 Simulations

We use real data from a fused magnesia factory within a day to test the demand forecasting algorithm in this paper. The limitation of demand is $\bar{P}_l = 22100$kVA and there are 4 FMFs working together that day. $U = 190$V is the voltage of each FMF, $I_i = 15000$A is set value of $i$th FMF'current and $\cos\varphi = 0.92$ is the power factor. We choose continuous electric power $p(1), \cdots, p(N)$ and demand $\bar{P}(1), \cdots, \bar{P}(N)$ samples for training and testing, where the total number of samples is $N = 4000$, top $n_{train} = 2000$ samples are for training offline and the rest is for testing online.

### 5.1 Identify linear model parameters

As discussed above, we first set $v(k) = 0$ for $k = 1, 2, \cdots, n_{train}$ and then identifying parameters $\boldsymbol{\theta} = (0.0234, 0.1586, -0.4169, 1.2344)^T$ by LS. Update $\boldsymbol{\theta}$ by RLS in equation (8) until $\boldsymbol{\theta}$ is convergent to $(-1.2082, 0.3686, -0.0702, 0.3003)^T$ . Then we fix parameters $\boldsymbol{\theta} = (-1.2082, 0.3686, -0.0702, 0.3003)^T$ and fit the unknown high order function in equation (7) by LSTM with FRA.

### 5.2 Results of demand forecasting

We train LSTM by the method mentioned earlier in this paper with training data set and then test online with test data set. In figure 3 we can see that the error between actual demand and its prediction is in $[-200, 400]$kVA, and we can forecast the accurate trend when the actual demand exceeds its limitation via the method mentioned in this paper. Figure 4 shows the electric power prediction results and it can be accepted. Both the errors of demand and electric power are Gaussian white noise.

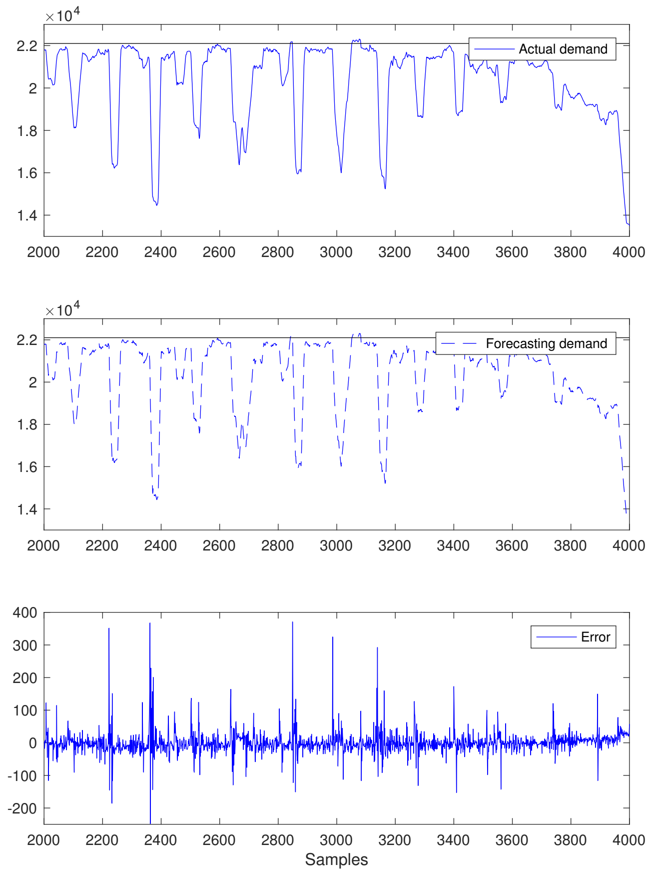


Fig. 3: Results of demand forecasting

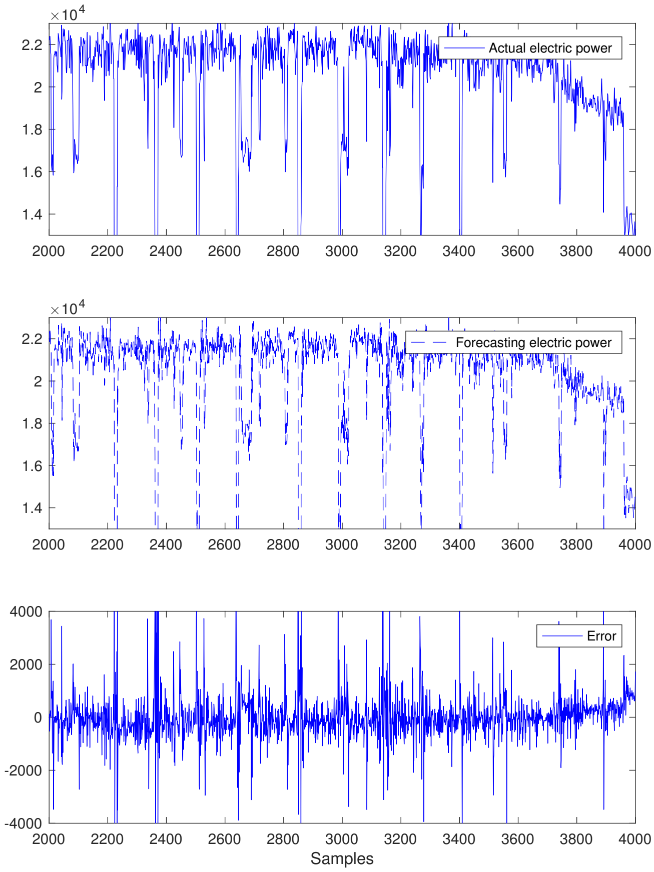


Fig. 4: Results of power forecasting

Results of RMSE, MASE, DS, TPR and TNR with LSTM and RBF as unknown high order function are shown in table 2. From table 2 we can see that the unknown high order function $v(k)$ fitted by LSTM with FRA mentioned earlier is with the best performance for demand forecasting in this paper. By LSTM the trend forecasting accuracy is up to almost 90% and we can judge whether it is necessary to turn off or turn on the FMF by it.

Table 2: Results of demand forecasting with LSTM and RBF

| | RMSE | MASE | DS | TPR | TNR |
|---|---|---|---|---|---|
| LSTM | 26.1457 | 0.016 | 93.76 | 88.69 | 91.75 |
| RBF | 33.9494 | 0.028 | 89.39 | 84.61 | 85.21 |

In figure 5 the probability distribution for the predictive error of electric power and auto-correlation coefficient curve of the predictive error of demand are shown. From the probability distribution curve we can see that the error equals to zero in proportion as 2.5% and auto-correlation coefficient curve of the predictive error of demand shows the errors are less auto-correlation.

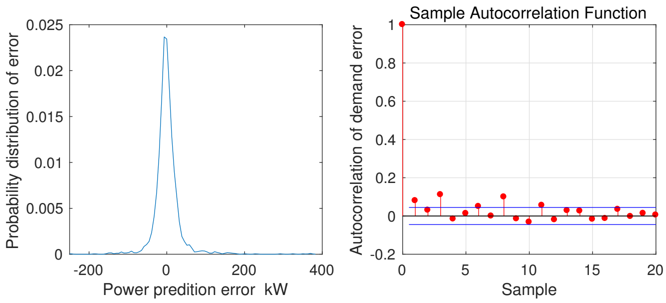


Fig. 5: Results of demand forecasting

## 6 Conclusion

For demand and its tend forecasting the most difficulty is there is too munch unknown information in FMSP, and we solve this by adding an unknown high order function $v(k)$ to a linear model. The main work in this paper is to find a method to improve the predictive accuracy of demand and its tend, and we use LSTM with FRA finally for 2 reasons:

(1)LSTM is a kind of deep network and it can improve the accuracy by this structure;

(2)FRA can accelerate the network training speed and avoid calculating the pseudo-inverse of information matrix in network.

We can see that the method mentioned in this paper has the best performance by comparing with RBF generally. We may need a new model which can describe the FMSP better in the future. Maybe a multi-steps prediction model and algorithm will be used for improving the predictive accuracy of demand and its tend.

## References


1. Jie Yang, Tianyou Chai, Chaomin Luo, and Wen Yu. Intelligent demand forecasting of smelting process using data-driven and mechanism model. *IEEE Transactions on Industrial Electronics*, 2018.
2. Martin T Hagan and Suzanne M Behr. The time series approach to short term load forecasting. *IEEE Transactions on Power Systems*, 2(3):785–791, 1987.
3. Ibrahim Moghram and Saifur Rahman. Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transactions on power systems*, 4(4):1484–1491, 1989.
4. HM Al-Hamadi and SA Soliman. Long-term/mid-term electric load forecasting based on short-term correlation and annual growth. *Electric power systems research*, 74(3):353–361, 2005.
5. Che-Chiang Hsu and Chia-Yon Chen. Regional load forecasting in taiwan—-applications of artificial neural networks. *Energy conversion and Management*, 44(12):1941–1949, 2003.
6. Hsiao-Tien Pao. Comparing linear and nonlinear forecasts for taiwan's electricity consumption. *Energy*, 31(12):2129–2141, 2006.
7. Andre Luiz S Maia, Francisco de AT de Carvalho, and Teresa B Ludermir. Symbolic interval time series forecasting using a hybrid model. In *2006 Ninth Brazilian Symposium on Neural Networks (SBRN'06)*, pages 202–207. IEEE, 2006.
8. Jie Yang and Tianyou Chai. Data-driven demand forecasting method for fused magnesium furnaces. In *2016 12th World Congress on Intelligent Control and Automation (WCICA)*, pages 2015–2022. IEEE, 2016.
9. Jie Yang, Tianyou Chai, Yajun Zhang, and Zhiwei Wu. Data and model driven demand forecasting method for fused magnesium furnaces. *Acta Autom. Sin.*, 44(8):1460–1474, Aug 2018.
10. Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
11. Kang Li, Jian-Xun Peng, and George W Irwin. A fast nonlinear model identification method. *IEEE Transactions on Automatic Control*, 50(8):1211–1216, 2005.
12. Manoj Badoni, Alka Singh, and Bhim Singh. Variable forgetting factor recursive least square control algorithm for dstatcom. *IEEE Transactions on Power Delivery*, 30(5):2353–2361, 2015.
13. Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634, 2015.