Methods of the Initial Data Transformation for GM(1,1) Grey Modeling

Wenzhan Dai, Qiumei Chen

College of Information and Electronic Engineering，
Zhejiang Gongshang University
310033, HangZhou, P.R.China

dwzhan@zist.edu.cn
xiaoqiu@pop.hzic.edu.cn

Abstract

The grey model's precision depends greatly on the smooth degree of data series. In this paper, several methods for improving smooth degree of data series are analyzed, and a more general theorem for enhancing smooth degree of initial data series is put forward .Then two new methods based on the function transformation $y \cdot a^{-k} (a > 1)$ and $y^{a} (a < 0)$ are represented. The practical models by using these two methods are built respectively and the results show the effectiveness and superiority of this method.
Keyword: Data transformation, Grey modeling, GM(1,1).

I. Introduction

Grey system theory has already been widely used in many fields since 1982. Although lots of achievements have been made in the grey system theory, there still exist some failures in building grey models. Because sometimes the precision of grey method by means of AGO (accumulated generation operation) and IAGO (inverse accumulated generation operation) can not meet the requirement of actual forecasting, much research in theory and application has been done. The results show that the smooth degree of initial data series is the one of key factors for the model's precision. The higher the smooth degree of data series is, the higher the model's precision is. Therefore, improving the smooth degree of initial data series has great signification for increasing grey model's precision. In order to enhance the smooth degree of data series, logarithm function transformation, exponent function transformation and power function transformation were put forward in papers [2-4] respectively, and a great achievement has been made in practical application.

In this paper, a general method for improving smooth degree of data series is put forward on the basis of summarizing several kinds of ways to improve smooth degree of data series .It is proved that logarithm function transformation, exponent function transformation and power function transformation all accord with the general theorem proposed in this paper. According to the theorem, two new transformations are put forward and practical applications show the effectiveness of these methods.

II. Some Methods of the Initial Data Transformation

Although lots of achievements have been made in the grey system theory, there still exist some failures in building grey models. Theoretically and practically it is proved that the smooth degree of data series is key factor of grey model's precision. Therefore, improving the smooth degree of initial data series has great signification for increasing grey model's precision.

Definition2.1[2] Let $\left\{x^{(0)}(k), k=1,2,\cdots n\right\}$ be non-negative data series, for $\forall \varepsilon > 0$, if there exists k_0, when $k > k_0$, the follow equation(1) is held,

$$\frac{x^{(0)}(k)}{\sum_{i=1}^{k-1} x^{(0)}(i)} = \frac{x^{(0)}(k)}{x^{(1)}(k-1)} < \varepsilon \tag{1}$$

Then series $\left\{x^{(0)}(k), k=1,2,\cdots n\right\}$ is defined as smooth data series.

Lemma 2.1[2] The necessary and sufficient condition for series $\left\{x^{(0)}(k), k=1,2,\cdots n\right\}$ to be smooth is that the function $\frac{x^{(0)}(k)}{\sum_{i=1}^{k-1} x^{(0)}(i)}$ is decrement with k.

In order to enhance the smooth degree of data series, logarithm function transformation, exponent function transformation and power function transformation are put forward in papers [2～4] respectively, and a great achievement has been made in practical application.

Theorem 2.1[2] If $x^{(0)}(k)$ is increment series with k and $x^{(0)}(1) > e$ comes into existence, then $\frac{\ln x^{(0)}(k)}{\sum_{i=1}^{k-1} \ln x^{(0)}(i)} \le \frac{x^{(0)}(k)}{\sum_{i=1}^{k-1} x^{(0)}(i)}$ is attainable.

Theorem 2.2[3] If $x^{(0)}(k)$ is increment series with k and $x^{(0)}(1) \ge 1, T > 1$ comes into existence, then $\frac{(x^{(0)}(k))^{\frac{1}{T}}}{\sum_{i=1}^{k-1} (x^{(0)}(i))^{\frac{1}{T}}} \le \frac{x^{(0)}(k)}{\sum_{i=1}^{k-1} x^{(0)}(i)}$ is attainable.

Theorem 2.3[4] If $x^{(0)}(k)$ is increment series with k and $x^{(0)}(1) > e, a > 1$ comes into existence, then $\frac{a^{-x^{(0)}(k)}}{\sum_{i=1}^{k-1} a^{-x^{(0)}(i)}} \le \frac{x^{(0)}(k)}{\sum_{i=1}^{k-1} x^{(0)}(i)}$ is attainable.

The following theorem 2.4 which was proposed in paper [4] can be gotten based on the above three theorems.

Theorem 2.4[4] If $x^{(0)}(k)$ is increment series with k and $x^{(0)}(1) \ge e, T \ge 1, a > 1$ comes into existence, then the following three equations are attainable.

(1) $\frac{a^{-x^{(0)}(k)}}{\sum_{i=1}^{k-1} a^{-x^{(0)}(i)}} \le \frac{\ln x^{(0)}(k)}{\sum_{i=1}^{k-1} \ln x^{(0)}(i)} \le \frac{x^{(0)}(k)}{\sum_{i=1}^{k-1} x^{(0)}(i)}$;

(2) $\frac{[a^{-x^{(0)}(k)}]^{1/T}}{\sum_{i=1}^{k-1} [a^{-x^{(0)}(i)}]^{1/T}} \le \frac{[\ln x^{(0)}(k)]^{1/T}}{\sum_{i=1}^{k-1} [\ln x^{(0)}(i)]^{1/T}} \le \frac{\ln x^{(0)}(k)}{\sum_{i=1}^{k-1} \ln x^{(0)}(i)}$;

（3） $\frac{[a^{-x^{(0)}(k)}]^{1/T}}{\sum_{i=1}^{k-1}[a^{-x^{(0)}(i)}]^{1/T}} \le \frac{a^{-x^{(0)}(k)}}{\sum_{i=1}^{k-1}a^{-x^{(0)}(i)}} \le \frac{\ln x^{(0)}(k)}{\sum_{i=1}^{k-1}\ln x^{(0)}(i)}$;

III. A General Method of Improving Smooth Degree of Data Series

Theorem 3.1 Let the non-negative original data series be denoted by $x^{(0)}(k)$, if there exist non-negative and decrement function $f(x^{(0)}(k),k)$ with k and the transfer function F that can be written in the following form

$$F(x^{(0)}(k)) = x^{(0)}(k)\cdot f(x^{(0)}(k),k) \tag{2}$$

Then the smooth degree of data series $F(x^{(0)}(k))$ is better than that of data series $x^{(0)}(k)$.

Proof Because $f(x^{(0)}(k),k)$ is non-negative and decrement strictly with k, so

$$0 < f(x^{(0)}(k),k) < f(x^{(0)}(i),i),\ i = 1,2,\cdots,k-1$$

According to the non-negative transfer $F(x^{(0)}(k)) = x^{(0)}(k)\cdot f(x^{(0)}(k),k)$, we can obtain

$F(x^{(0)}(1)) = x^{(0)}(1)\cdot f(x^{(0)}(1),1) > x^{(0)}(1)\cdot f(x^{(0)}(k),k)$

$F(x^{(0)}(2)) = x^{(0)}(2)\cdot f(x^{(0)}(2),2) > x^{(0)}(2)\cdot f(x^{(0)}(k),k)$

…

$F(x^{(0)}(k-1)) = x^{(0)}(k-1)\cdot f(x^{(0)}(k-1),(k-1)) > x^{(0)}(k-1)\cdot f(x^{(0)}(k),k)$

Therefore

$$\sum_{i=1}^{k-1}F(x^{(0)}(k)) = \sum_{i=1}^{k-1}(x^{(0)}(i)\cdot f(x^{(0)}(i),i)) > f(x^{(0)}(k),k)\cdot\sum_{i=1}^{k-1}x^{(0)}(i)$$

while $F(x^{(0)}(k)) = x^{(0)}(k)\cdot f(x^{(0)}(k),k)$

Then $\frac{F(x^{(0)}(k))}{\sum_{i=1}^{k-1}F(x^{(0)}(i))} < \frac{x^{(0)}(k)\cdot f(x^{(0)}(k),k)}{f(x^{(0)}(k),k)\cdot\sum_{i=1}^{k-1}x^{(0)}(i)} = \frac{x^{(0)}(k)}{\sum_{i=1}^{k-1}x^{(0)}(i)}$

By deducting the two above equations, the theorem is easily proved completely.

The general method to improve smooth degree of data series has been put forward according to the theorem. In fact, the methods proposed in papers [2～4] can be considered as examples of the theorem. For example:

(1) Logarithm function transfer [2].

Let $x^{(0)}(k)$ be an increment data series with k and $x^{(0)}(1) > e$

Make transfer $F(x^{(0)}(k)) = \ln[x^{(0)}(k)]$, it is easy to prove that $f(x^{(0)}(k),k)$ is a non-negative and decrement function strictly with k, when $f(x^{(0)}(k),k) = \frac{\ln[x^{(0)}(k)]}{x^{(0)}(k)}$.

Proof Because $x^{(0)}(k)$ is an increment series with $x^{(0)}(1) > e$, we can get $f(x^{(0)}(k),k) = \frac{\ln[x^{(0)}(k)]}{x^{(0)}(k)} > 0$. It is to say, $f(x^{(0)}(k),k)$ is non-negative.

Because $(f(x^{(0)}(k),k))' = (\frac{\ln[x^{(0)}(k)]}{x^{(0)}(k)})' = \frac{1-\ln[x^{(0)}(k)]}{(x^{(0)}(k))^2}\cdot(x^{(0)}(k))' < 0$

then $f(x^{(0)}(k),k)$ is decreased strictly.

(2) Exponent function transfer [3].

Let $x^{(0)}(k)$ be an increment series and $x^{(0)}(1)>0$, Make transfer $F(x^{(0)}(k))=x^{(0)}(k)^{\frac{1}{T}}$ ($T>1$), it is easy to prove that $f(x^{(0)}(k),k)$ is a non-negative and decrement function with $f(x^{(0)}(k),k)=\frac{x^{(0)}(k)^{\frac{1}{T}}}{x^{(0)}(k)}=x^{(0)}(k)^{\frac{1}{T}-1}$. The proof is omitted.

(3) Power function transfer [4].

Let $x^{(0)}(k)$ be an increment series and $x^{(0)}(1)>0$ Make transfer $F(x^{(0)}(k))=a^{-x^{(0)}(k)}$ ($a>1$), it is easy to prove that $f(x^{(0)}(k),k)$ is a non-negative and decrement function with $f(x^{(0)}(k),k)=\frac{a^{-x^{(0)}(k)}}{x^{(0)}(k)}$.
The proof is omitted.

Only three kinds of transformation functions for improving smooth degree of data series have been put out above. It also has been proved that all these transfer functions accord with theorem proposed in this paper. Here, the author puts forward another two kinds of new transfer functions.

(1) $F(x^{(0)}(k))=x^{(0)}(k)\cdot a^{-k}(a>1)$, it is easy to prove that $f(x^{(0)}(k),k)$ is a non-negative and decrement function with $f(x^{(0)}(k),k)=a^{-k}$.

(2) $F(x^{(0)}(k))=(x^{(0)}(k))^{a}(a<0)$, it is easy to prove that $f(x^{(0)}(k),k)$ is a non-negative and decrement function with $f(x^{(0)}(k),k)=(x^{(0)}(k))^{a-1}$.

Ⅳ. The Grey modeling mechanism with function transformation

$GM(1,1)$ is the most frequently used grey model. It is formed by a first order differential equation with a single variable. The steps of modeling with functions transformation $y\cdot a^{-k}(a>1)$ or $y^{a}(a<0)$ are as follows:

(1) Let non-negative and increment original series be denoted by

$$Y^{(0)}=\{y^{(0)}(1),y^{(0)}(2),...,y^{(0)}(n)\} \quad (3)$$

where $y^{(0)}(i)>1, i=1,2,...,n$.

(2) When the original data series are transformed by $y\cdot a^{-k}(a>1)$, $X^{(0)}(k)$ is defined as:

$$X^{(0)}=\{x^{(0)}(1),x^{(0)}(2),...,x^{(0)}(n)\} \quad (4)$$

where $x^{(0)}(k)=[y^{(0)}(k)]\cdot a^{-k}, i=1,2,...,n$

While the original data series are transformed by $y^{a}(a<0)$, $X^{(0)}(k)$ is defined as:

$$X^{(0)}=\{x^{(0)}(1),x^{(0)}(2),...,x^{(0)}(n)\} \quad (5)$$

where $x^{(0)}(k)=[y^{(0)}(k)]^{a}, i=1,2,...,n$.

(3) The AGO (accumulated generation operation) of original data series is defined as:

$$X^{(1)}=\{x^{(1)}(1),x^{(1)}(2),...,x^{(1)}(n)\} \quad (6)$$

where $x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i), k=1,2,...,n$.

(4) The grey model can be constructed by establishing a first order differential equation as following:

$$\frac{dx^{(1)}(t)}{dt}+ax^{(1)}(t)=u \quad (7)$$

The difference equation of GM (1,1) model:

$$x^{(0)}(k)+az^{(1)}(k)=u, k=2,3,...,n \quad (8)$$

Unfolding equation（7）, we can obtain:

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \times \begin{bmatrix} a \\ u \end{bmatrix} \tag{9}$$

Let $Y = \left[x^{(0)}(2), x^{(0)}(3), ..., x^{(0)}(n)\right]^T$, $\Phi = \left[a \quad u\right]^T$,

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \tag{10}$$

The background $z^{(1)}(k)$ in equation（7）is defined as:

$$z^{(1)}(k+1) = \frac{1}{2}[x^{(1)}(k+1) + x^{(1)}(k)],\ k = 1,2,...,n-1$$

(5) The estimation value of parameter Φ by using least squares is

$$\hat{\Phi} = (B^T B)^{-1} B^T Y$$

(6) The discrete solution of equation (6):

$$\hat{x}^{(1)}(k+1) = [x^{(1)}(1) - \frac{u}{a}]e^{-ak} + \frac{u}{a} \tag{11}$$

(7) When the original data series are transformed by $y \cdot a^{-k}\,(a > 1)$, revert $\hat{x}^{(1)}(k+1)$ into initial data series:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)$$

$$= (1 - e^{a})[x^{(1)}(1) - \frac{u}{a}]e^{-ak};$$

$$\hat{y}^{(0)}(k) = \hat{x}^{(0)}(k) * a^{k}$$

While the original data series are transformed by $y^{a}\,(a < 0)$, revert $\hat{x}^{(1)}(k+1)$ into initial data series:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)$$

$$= (1 - e^{a})[x^{(1)}(1) - \frac{u}{a}]e^{-ak};$$

$$\hat{y}^{(0)}(k) = (\hat{x}^{(0)}(k))^{\frac{1}{a}}$$

Ⅴ. Example

5.1 example 1

The national finance expenditure is an important index of national economic strength. To build the model of national finance expenditure by means of function $y \cdot a^{-k}\,(a > 1)$ and forecast its trend will have great significance. Now let build the model based on《Statistic almanac of China-2005》 from 1991 to 2002 and predict national finance expenditure of 2003 and 2004.

The traditional $GM(1,1)$ model of these data is as following:

$$\begin{aligned} &\hat{x}^{0}(k+1) = 3313.14e^{0.1724k} \\ &x^{0}(1) = 3386.62 \end{aligned} \qquad k > 1$$

The model by using function $y \cdot a^{-k}(a>1)$ is as following:

$$\hat{y}^{0}(k+1) = 3087.96e^{0.0960k} * a^{k}$$

$$a = 1.08, k > 1, y^{0}(1) = 3386.62$$

Table 1 gives comparison of two modeling methods. Figure 1 is the fitted curve.

Table 1. Comparison of two modeling methods (Unit: billion RBM)

Year	No.	National FinanceExpenditure	Traditional GM (1,1)		Method proposed in this paper	
			Model values	Relative error(%)	Model values	Relative error (%)
1991	1	3386.62	3386.62	0	3386.62	0
1992	2	3742.20	3936. 64	-5.20	3935.36	-4.16
1993	3	4642.30	4677.46	-0.76	4678.60	-0.78
1994	4	5792.62	5557.70	4.06	5562.20	3.97
1995	5	6823.72	6603.59	3.23	6612.68	3.10
1996	6	7937.55	7846.30	1.15	7861.55	0.95
1997	7	9233.56	9322.87	-0.97	9346.29	-1.22
1998	8	10798.18	11077.32	-2.59	11111.44	-2.90
1999	9	13187.67	13161.93	0.20	13209.96	-0.16
2000	10	15886.50	15638.83	1.56	15704.80	1.14
2001	11	18902.58	18581.86	1.70	18670.83	1.22
2002	12	22053.15	22078.73	-0.12	22197.01	-0.65
2003*	13	24649.95	26233.67	-6.42	24818.27	0.68
2004*	14	28360.79	31170.51	-9.91	30342.75	6.98

(* Forecasting value)

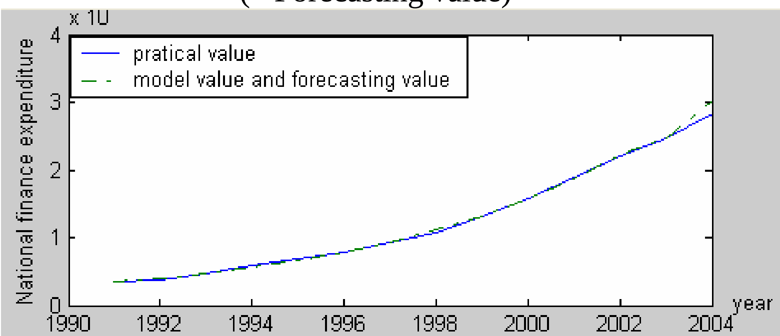

Fig.1 The fitted curve

From table 1, it is easy to see that the absolute value of the relative error of the model proposed by this paper is very lower. The error inspection of post-sample method can be used to inspect quantified approach. The post-sample error $c = S_1 / S_0$ of the model proposed by this paper is 0.0292(where S_1 is variation value of the error and S_0 is variation value of the original series), while the post-sample error of traditional GM (1,1) model is 0.0303. The probability of the small error $p = \{|e^{(0)}(i) - e^{-(0)}|\} < 0.6745S_0 = 1$. The post-sample error of this model is far less than 0.3640. In a word, it is obvious that the model using function $y \cdot a^{-k}(a>1)$ has improved the fitted precision and prediction precision.

5.2 example 2

The per capita power consumption can greatly reflect the industry developing level. Now let build the model based on 《Statistic almanac of China-2002》 from 1987 to 1999 and predict per capita power consumption of 2000 and 2001.

The traditional $GM(1,1)$ model of per capita power consumption is as following:

$$\begin{cases} \hat{x}^0(k+1) = 472.5931e^{0.0651k} \\ x^0(1) = 458.75 \end{cases} \quad k > 1$$

The model by using function $y^a(a<0)$ is as following:

$$\hat{y}^0(k+1) = (0.9405e^{-0.0007k})^{\frac{1}{a}}$$

$$a = -0.01, k > 1, \hat{y}^0(1) = 458.75$$

Table 2 is Comparison of two modeling methods. Figure.2 is the fitted curves.

Table2. Comparison of two modeling methods (Unit: Kilogram)

Year	No.	Per Capita output of steel	Traditional GM (1,1)		Method proposed in this paper	
			Model values	Relative error(%)	Model values	Relative error(%)
1987	1	458.75	458.75	0	458.75	0
1988	2	494.9	504.37	-1.91	494.36	0.11
1988	3	522.78	538.29	-2.97	529.17	-1.22
1990	4	547.22	574.49	-4.98	566.44	-3.51
1991	5	588.7	613.12	-4.15	606.33	-2.99
1992	6	647.18	654.35	-1.11	649.03	-0.28
1993	7	712.34	698.35	1.96	694.75	2.47
1994	8	778.32	745.31	4.24	743.67	4.45
1995	9	835.31	795.42	4.77	796.05	4.70
1996	10	888.1	848.91	4.41	852.11	4.05
1997	11	923.16	906.00	1.86	912.12	1.19
1998	12	939.48	966.92	-2.92	976.36	-3.92
1999	13	988.60	1031.94	-4.38	1045.13	-4.71
2000*	14	1140.54	1101.33	3.44	1146.42	-0.52
2001*	15	1223.31	1175.39	3.92	1229.51	-0.51

(* Forecasting value)

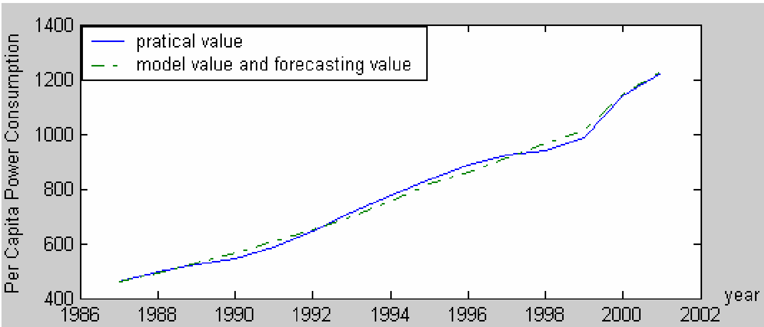

Fig.2 The fitted curve

From table 2, it is obvious that the absolute value of the relative error of the model proposed by this paper is less than 5 %. The error inspection of post-sample method can be used to inspect quantified approach. The post-sample error $c = S_1 / S_0$ of the model proposed by this paper is 0.1529(where S_1 is variation value of the error and S_0 is variation value of the original series). The probability of the small error $p = \{| e^{(0)}(i) - e^{-(0)} |\} < 0.6745S_0 = 1$. Then we can come to conclusion that the method using function $y^a(a<0)$ has enhanced the fitted precision and prediction precision. From the two examples above, we can get the conclusion that the methods proposed in this paper have increased the smooth degree of data series. Therefore, the methods improve the fitted precision and prediction precision of models greatly.

Ⅵ. Conclusion

In this paper, the more general method of improving smooth degree of data series is put forward on the basis of summarizing several kinds of methods of the initial data transformation, and two new transformations are represented. At last, the models by means of these methods are built respectively and the results show the effectiveness and superiority of this method.

Ⅶ. Acknowlegement

This paper is supported by the Natural Science Foundation of Zhejiang Province, P.R .China(NO：602016)

References

[1] S. F. Liu, T. B. Guo, Y. G. Dang, "*Theory and Application of Grey System,*" BeiJing, Science Publishing Company, 1999.
[2] T. J. Chen, "An Expansion of Grey Prediction Model, System Engineering," Systems Engineering—Theory & Practice, vol. 7, 1990, pp. 50-52.
[3] Q. Li, "The Further Expansion of Grey Prediction Model," Systems Engineering-Theory & Practice, vol. 1, 1993, pp. 64~66.
[4] B. Bin, Q. Meng, "Research on Methods of Expending Grey Prediction Model," Systems Engineering-Theory & Practic, vol. 9, 2002, pp. 138~141.
[5] Y. L. Xiang, "Research on GIM (1) Model for Investment on Environmental Protection," Environmental Protection Science, vol.2, 1995, pp. 72~76.
[6] Y. L. Xiang, "GIM(1) Model of Predicting Regional Ambient Noise," Si Chuan Environment, vol. 1, 1996, pp. 68—71.
[7] Y. K. Chen, X. R. Tan, "Grey Relation Analysis on Serum Makers of Liver Fribrosis," The Journal of Grey System, vol. 1, 1995, pp. 63~68.
[8] S. F. Liu, J. L.Deng, "The Range Suitable for GM(1,1)," The Journal of Grey System, vol. 1, 1999, pp. 131~138.
[9] J. Kendrick, A. David, *Stochastic Control for Economic Models*, New York, McGraw-Hill, 1981.

Wenzhan Dai, Professor, Corresponding author, Vice-President of Zhejiang Sci-Tech University, P.R.C. He received his B.S. at Zhejiang University ,P.R.C in 1982 and his M.S. at East China University of Science and Technology in1987. His research interests include systems engineering, modeling & control. He has published more than 100joural, conference, and symposium papers.

Qiumei Chen received her B.S. in Information & Electronic Engineering at Zhejiang Gongshan University, P.R. China. She is going for her M.S. degree in signal & information procession from Zhejiang Gongshan University. Her research interests include systems engineering, modeling & control.