View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Okayama University Scientific Achievement Repository

Memoirs of the Faculty of Engineering, Okayama University, Vol.40, pp.9-15, January, 2006

Statistical Estimation Model for Product Quality of Petroleum

Takashi Nukina *
Dept. of Electrical and Electronic Engineering
Okayama University
3-1-1, Tsushima-Naka, Okayama, 700-8530

Masami Konishi
Dept. of Electrical and Electronic Engineering
Division of Industrial Innovation Sciences
The Graduate School of
Natural Science and Technology
Okayama University
3-1-1, Tsushima-Naka, Okayama, 700-8530

Tatsushi Nishi
Dept. of Electrical and Electronic Engineering
Division of Industrial Innovation Sciences
The Graduate School of
Natural Science and Technology
Okayama University
3-1-1, Tsushima-Naka, Okayama, 700-8530

Jun Imai
Dept. of Electrical and Electronic Engineering
Division of Industrial Innovation Sciences
The Graduate School of
Natural Science and Technology
Okayama University
3-1-1, Tsushima-Naka, Okayama, 700-8530

(Received January 27, 2006)

Controls of the temperature, pressure and flowing quantity are important for the stable operation of the product quality in the distillation tower. The usual measuring way of product quality estimation is made by the off-line analysis. In this paper, online estimation method of product quality is studied for improving the product quality. The estimation method based on stochastic analysis was developed for online estimation. In this paper, the data of temperature, pressure and flow volum in the distillation tower are treated.

As the estimation models, RNN (Recurrent Neural Net Work) and PLS (Partial Least Square Regression Method) were adopted. The actual plant data were used in the analysis. Both PLS and RNN models could compensate each other to improve the accuracy in estimation.

1 Introduction

Controls of temperature, pressure, and flow volume are inevitable for stable operation and assurance of the product quality in the distillation tower. The usual measuring way of product quality was the off-line analysis of the product. To improve the product quality, it is necessary develop a new method to estimate it for online use. In the present time, an expensive high technological quality sensor is used for online use. However, to adjust the sensor, expert technology is needed. To solve the above problem, acculate estimation technology for product qualities is needed. To attain the purpose, signal processing method is studied using the data of temperature, pressure and flow volume in the

*nukina@cntr.elec.okayama-u.ac.jp

This work is subjected to copyright.
All rights reserved by this author/authors.

distillation tower.

The methods of estimating data are the methods of using a physical model and a statistical model. However, the achievement of the physical model is difficult because the influence of disturbance deteriorates in accuracy. So, it is difficult to estimate accurate qualities only by a physical model. PLS analysis and RNN analysis are adopted because the multivariate analysis that is a statistical analytical method is suitable for the data analysis in the distillation tower. But the deterioration in the estimation accuracy is a problem for online use of the method for the distillation tower.

In this research, sequential update method was adopted for a method of achieving high estimation accuracy by a statistical method. On the while, the estimation accuracy changes by the period of estimation or the volume of used data. In this research, there are a number of training data, a number of latent number and a the number of hiddenlayers as the estimation of a statistical model. The method to modelate these parameters was developed to utilize it. The effectiveness of the proposed method was confirmed by the application test of the estimation model using actual plant data.

Next, the analysis method using good points of RNN and PLS model was proposed. This estimation method proposed here, consists of a basic characteristic model by PLS and correction model by RNN. The effectiveness of this method was confirmed by the application test of estimation using the actual plant data.

2 Distillation tower and its data

The distillation tower is a plant that separates the crude oil to diesel oil, heavy oil, and coal oil, etc. by using the difference of the boiling point. The conditions of the product quality in the distillation tower are forecast by using a variety of analysis methods. In the following, the data used for the quality estimation in the distillation tower are explained.

2.1 Distillation tower

A view of the distillation tower is shown in Fig.1.The distillation tower separates from crude oil to various elements by using the difference of the boiling point. It is not possible to separate to each element completely in one distillation tower. So, product purity is raised by separating through plural distillation towers. There are a lot of devices that measure temperatures, pressures and follow volums in the distillation tower. In the estimation, a lot of data measured in the distillation tower by many measuring devices are adopted as input data. And the temperature of the product in the distillation tower is adopted as output data. A statistical model is made from these input and output datas, and the temperature of the product is estimated. After that, the way how the product qualities change by disturbance can be estimated based on a statistical model.

Fig. 1. View of distillation tower

2.2 Process data of distillation tower

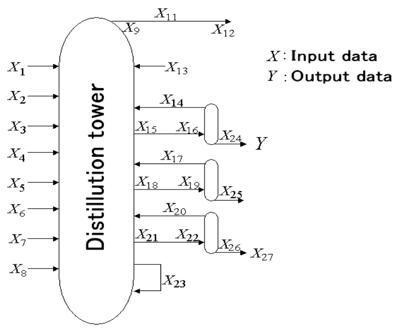

Fig. 2. Schemaic model of distillation tower

A schemaic model of the distillation tower is shown in Fig.1. As shown in Fig.2 there are 28 measuring devices set at each places in the distillation tower and tower datas are measured by these devices in every 15 minutes simultaneously. The data measured by a high-tech sensor is set as output data and the data of other measuring devices at 27 places of the distillation tower are set as input data.

There are 1266 sampled data for one output and 27 inputs measured in every 15 minutes.

3 Description of analysis method

In this chapter, analysis method by PLS, RNN and the evaluation index will be explained.

3.1 PLS analysis

Construction of PLS analysis is shown in Fig.3.

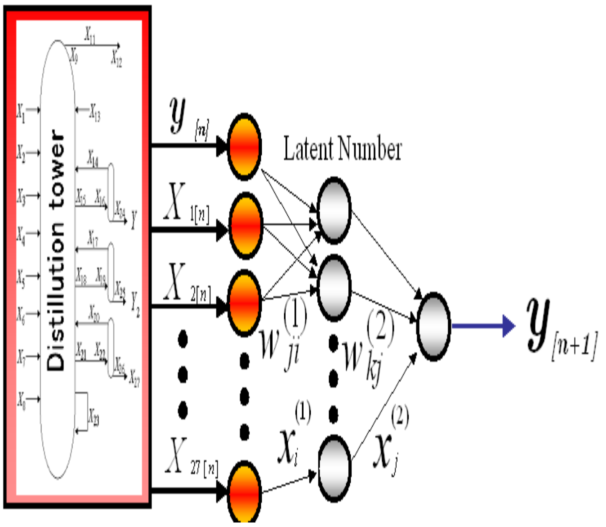

Fig. 3. PLS analysis

There is a MRA(multiple regression analysis) in multivariate analysis techniques. MRA is a least square method of N dimensions. And MRA is able to estimate process characteristics in good accuracy. But MRA has a problem of multicollinearity. Multicollinearity is linear relation between variables. If there is linear relation between variables, the estimation accuracy deteriorates because unnecessary variables enlarges estimation error. The PLS analysis can solve a problem of multicollinearity of MRA. PLS analysis generates new variables made from original variables avoiding a problem of multicollinearity. In addition, PLS analysis makes new variables that have a strong correlation to an output variable. So, PLS can estimate better accuracy than MRA. The PLS analysis is adopted because there are correlation between the datas measured by each sensor in the distillation tower.

Here, the mathematical description of PLS analysis will be introduced. There are P kinds of input variables X that have each N samples, and there is one output variable Y that have N samples. Input datas(measured by each sensor) are measured by sensor. PLS makes the first latent variable that inner product becomes the maximum from input variables. If coupling coefficient of latent variable is w_1, the value of t_{n1} of the first latent variable z_1 corresponded nth samples x_n is as follows.

$$t_{n1} = \sum_{p=1}^{p} w_{p1} x_{np} = x_n w_1 \tag{1}$$

The value t_{n1} of this first latent variable z_1 in matrix form is as follows.

$$t_1 = \begin{pmatrix} w_{11}x_{11} + w_{21}x_{12} + \cdots + w_{p1}x_{1p} \\ w_{11}x_{21} + w_{21}x_{22} + \cdots + w_{p1}x_{2p} \\ \vdots \\ w_{11}x_{N1} + w_{21}x_{N2} + \cdots + w_{p1}x_{Np} \end{pmatrix} \tag{2}$$

$$= X\omega_1 \tag{3}$$

So, inner product ϕ_1 of the output variable y and the latent variable z_1 is given as follows.

$$\phi_1 = y^T X \omega_1 = \sum_{n=1}^{N} w_{p1} y_n t_{n1} \tag{4}$$

To derive the latent variable z_1, the coupling coefficient ω_1 that makes inner product ϕ_1, maximum is to be derived. Consequently, coupling coefficient ω_1 is as follows.

$$\omega_1 = \frac{1}{\phi_1} X^T y \tag{5}$$

Thus, the coupling coefficient ω_1 can be obtained.

If the change of output variable can not be estimated enough by the latent variable z_1, the number of latent variables are increased until the change of output variable can be estimated. The second latent variable is made according to the same process as the first latent variable. $X^{(1)}$ and $y^{(1)}$ that could not be expressed by the first latent variable z_1 are assumed to X and y, and the the second latent variable z_2 is made by the same process as the first latent variable. Also the third latent variable z_3 is made by the same process.

The PLS analysis improves a problem of multicollinearity of MRA like above. So, PLS analysis is a better analysis than MRA because it can make new variables that have a strong correlation with the output. New variables made from original variables can solve a problem of multicollinearity.

3.2 RNN analysis

The method of RNN analysis is shown in Fig.4.

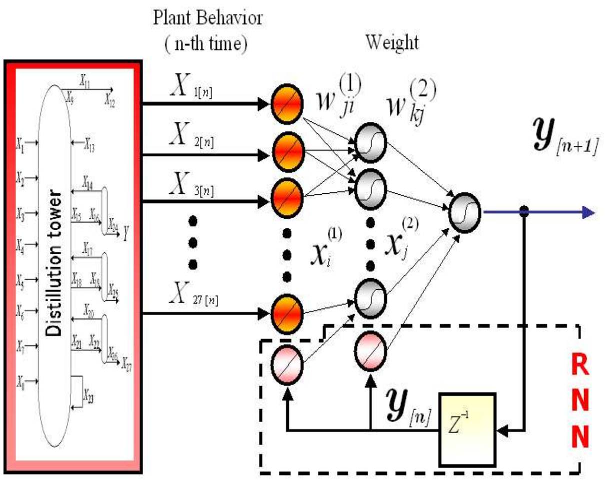

Fig. 4. RNN analysis

Usually, NN(neural network) analysis is not used for the quality estimation of the distillation tower. NN analysis is an excellent analysis method for a nonlinear element. Further, RNN(recurrent newral network) analysis can estimate an excellent estimation accuracy in the plant such as the distillation towers which is suffered by frequent turbulances. Moreover, qualities of the distillation tower are affected by past influences. The neural network used in this research is a type of hierarchical neural network with three layers, and it has the feedback of the estimation of output. So the neural net work used in this research is RNN. The learning way of RNN is back propagation method.

3.3 Evaluation Index

Each variable is shown as follows.

y_i is actual data
$\bar{y}$ is mean of actual data
$\hat{y}_i$ is value of estimation
$e_i = \hat{y}_i - \bar{y}$
$\bar{e} = y_i - \bar{y}$

Two methods are used for the evaluation index showing estimation accuracy.

- RMSE(Root Mean Square Error)

$$RMSE = \sqrt{\frac{\sum e_i^2}{N}} \tag{6}$$

This evaluation index shows the square mean of the error with the actual value.

- INDEX(Index of mean square error)

$$INDEX = 100 \times \frac{\sum e_i^2}{(y_i - \bar{y})^2} \tag{7}$$

This evaluation index showing difference between estimation valuesand actual data.

4 Analytical Method

In this chapter, the cause of residual error by PLS analysis and RNN analysis is tried to be solved. In the following, sequential and update functions of these models will be explained. And, these effects are checked based on actual data.

4.1 Sequential Modeling

The sequential modeling is a method of reconstraction of the model when new data is obtained. The sequential modeling is shown in Fig.5.

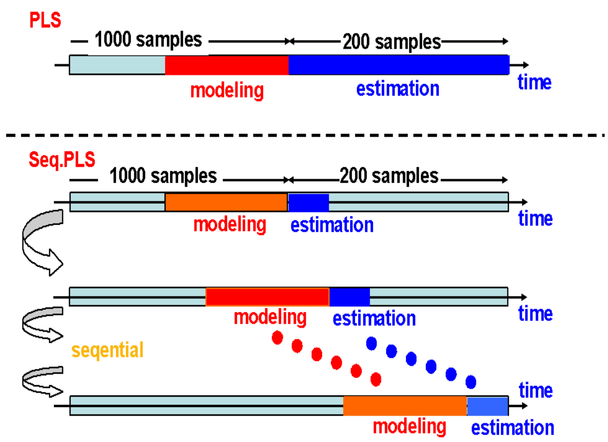

Fig. 5. Sequential Modeling

4.2 Model Update Function

As show in Fig.6, it is necessary to decide three parameters, the number of the latent variables, the number of training datas and delay time when a statistical model is applied. In this reseach, parameters of PLS model is analyzed by past data. Appropriate values of these three variables are decided. This procedure defined as Model Update Function.

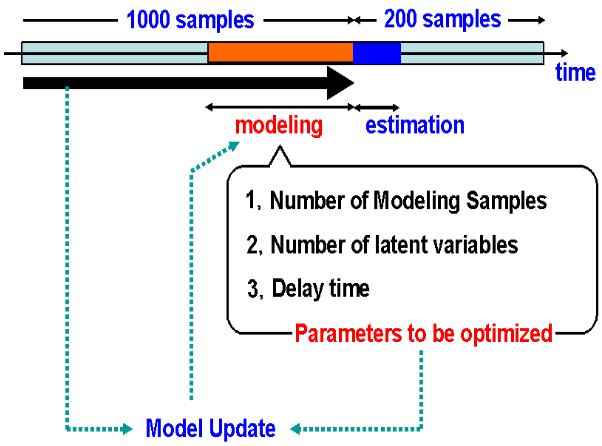

Fig. 6. Model Update Function

To examine the effectiveness of the model update function, output datas are analyzed by two kinds of PLS models. One is analyzed by PLS without model update function and the other is analyzed by PLS with model update function.

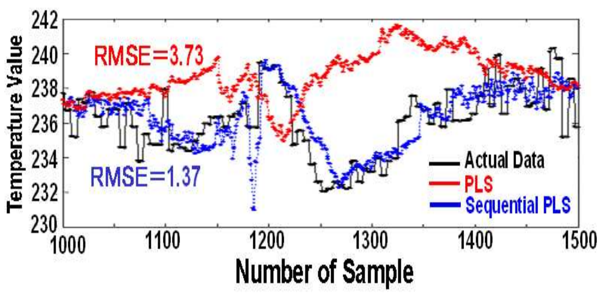

Fig. 7. Seq.PLS without model update function

The result in Fif.8 by the model update function can be confirmed that the estimation accuracy has improved more than the result in Fig.7. Moreover, it is possible to operate the plant for a long time stably by use of the model update function.

As shown in Fig.9, it can be confirmed that the estimation accuracy has improved by using the model

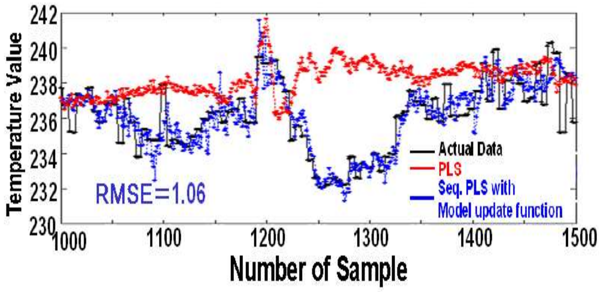

Fig. 8. Seq.PLS with model update function

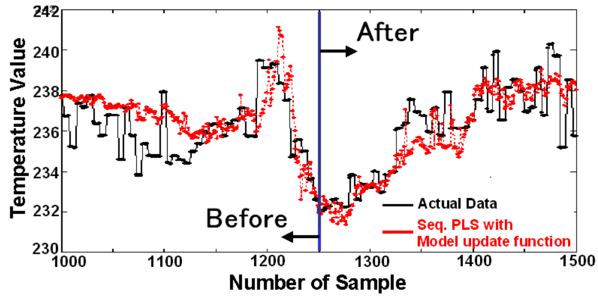

Fig. 9. Effectiveness of Model Update Function for a long time

update function for a long time test.

4.3 Model with delay time

There exists delay times between measure devices in the distillation tower. For instance, when it rains, its effects difference in the height position of distillation tower. That is, the temperature of the tower bottom part is not changing, but the temperature of the top part of the distillation tower is changing. There are places of measure devices that detects a change at first after a certain delay time. For the above reasons, there are delay times between all input and output data. As for reflection of delay time to estimation model, Akaike's method is adopted.

5 Proposed method

While the chemical plant is being operated, qualities of the chemical plant change gradually. PLS analysis is used because the correlations exist among variables. But, because the disturbance induces serious influence, it is necessary to consider the nonlinear element for the robustness of the model.

In this chapter, qualities of the distillation tower are analyzed by using PLS analysis, and the error of the estimation value by disturbance etc. is corrected by RNN. This is proposed method. Moreover, the model evaluation I_j are added as the input to RNN. Configuration of proposed model is shown in Fig.10.

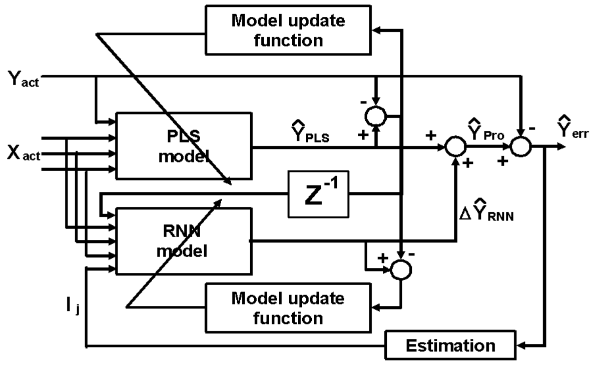

Fig. 10. Configuration of Proposed Model

6 Numerical experiments

In this chapter, the sequential PLS analysis with model update function named proposed method and the simple sequential RNN analysis named conventional method are compared. Numerical experiments of conventional method and proposed method were made. In numerical experiment, 27 kinds of input data and one kind of output data are used. Each operation data of distillation tower has 1200 samples. The conditions of the sequence PLS analysis with model update function are 36 samples for training data, delay time of 5 samples, and 5 latent variables. And the conditions of the sequence RNN analysis are 30 samples for training data and 10 inner layers. The conditions of the fixed PLS analysis of proposed method are 200 samples for training data, delay time of 5 samples and 5 latent variables. The conditions of the sequence RNN analysis of proposed method are 3 samples for training data, and 10 inner layers. 501-600 samples are estimated by using these three method of two conventional methods and a proposed method.

As shown in Fig.11,13 and Table1, PLS+RNN of proposed method is better than sequential PLS both in RMSE and INDEX indicies. As shown in Fig.12,13 and Table1, PLS+RNN of proposed method is better than sequential RNN both in RMSE and INDEX indicies. Next as shown in Fig.13 and 14, the result of the fixed PLS is corrected by the sequence RNN. In addition, the results of applying the proposed method to 601-1000 samples is shown in Fig.15 revealing follow up ability of proposed estimation model to new output data.

7 Conclusion

In this paper, the method of estimation for product quality in the distillation tower was studied. In this paper, the distillation tower is analyzed by combining two statistical models, PLS analysis and RNN analysis. Further, the idea of sequential model, model update function, and the delay time are applied to PLS and RNN analysis. Then the effectiveness of these methods was examined. For the actual use, a new estimation method of product qualities in the distillation tower is proposed in consideration of the feature of PLS and RNN respectively. The proposed method estimates an physical phenomena in the distillation tower by PLS analysis, and securing the effect of disturbance by RNN analysis. High accuracy in the estimation by the sequential PLS analysis with model update function by RNN analysis is demonstrated.

Bibliography

[1] G. E. P. Box and G. M. Genkins: TIME SERIES ANALYASIS forecasting and Control, HOLDEN-DAY (1976)

[2] I. M. Mujitaba and M. A. Hussain: Application of Neural Networks and Other Learning Technologies in Process Engineering, Imperial College Press (2001)

[3] S. Joe Qin: Recursive PLS Algorithms for Adaptive Data-modeling, Computers, chem.. Engng, Vol. 22 (1998)

[4] P. Vijaysi,R. D. Gudi and S. Lakshminarayanan: A New Recursive PLS Structure for Rapid Identification on Demand, PRE-PRINTS of 6th IFAC SYMPOSIUM ON DYNAMICS AND CONTROL OF PROCESS SYSTEMS (2001)

Table 1 Numerical Result

Method	Model Condition				Estimation Range	RMSE	INDEX
	Training data	Model	Time delay	latent number			
Seq.PLS	36	Seq	5	5	501-600	1.9419	109.2709
Seq.RNN	30	Seq	5	5		1.74621	88.3607
PLS +	200	Fix	5	5		1.63638	77.5945
RNN	3	Seq	0	0	601-1000	2.1647	59.4820

Seq.PLS:Sequential PLS with Model Update Function

Seq.RNN:Sequential RNN

PLS+RNN:PLS Corrected by RNN

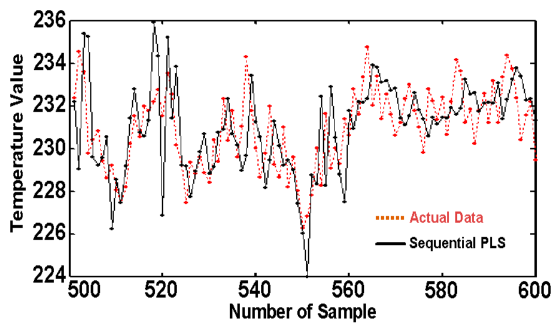

Fig. 11. Results of Sequential PLS

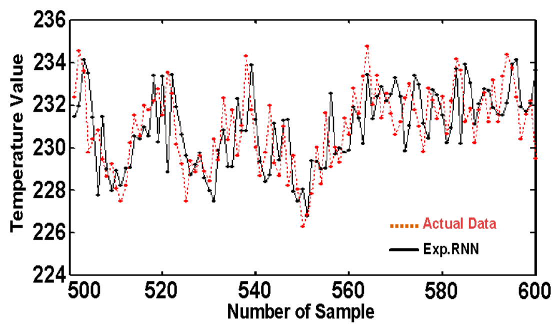

Fig. 12. Results of sequential RNN

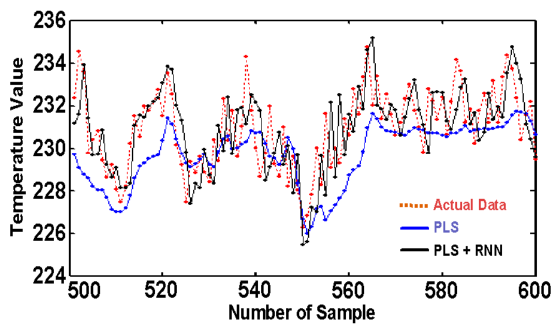

Fig. 13. Results of PLS+RNN

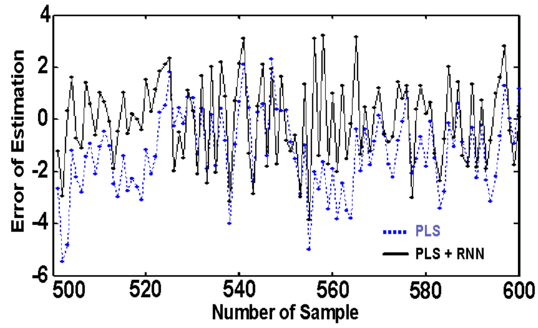

Fig. 14. Errors by PLS and PLS+RNN

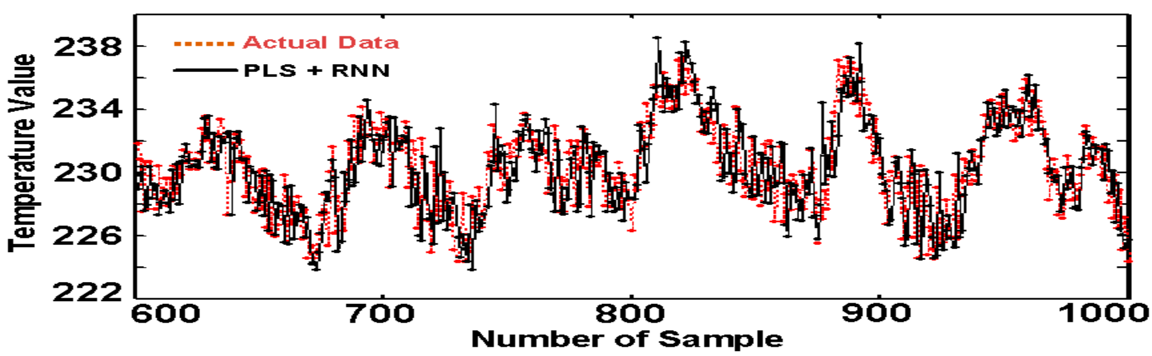

Fig. 15. Results of PLS+RNN(601-1000 samples)