http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 13 (9): 1640-1647, 2014
ISSN 1812-5638 / DOI: 10.3923/itj.2014.1640.1647

DCSCS: A Novel Approach to Improve Data Accuracy for Low Cost Meteorological Sensor Networks

Huan Liu, Baowei Wang, Xingming Sun, Tao Li, Qi Liu and Yunbiao Guo
No. 219, Ningliu Road, Pukou, Nanjing, Jiangsu, 210044, China

Abstract: The technology of Wireless Sensor Network (WSN) has facilitated the meteorological parameter monitoring. To reduce the cost, most monitoring systems collect desired data through cheap, unprofessional meteorological equipments, so do our system. However, due to the low cost of equipments and the poor package, the data collected by unprofessional equipments is inaccurate. In this study, we propose a Data Correction Scheme for Cheap Sensor (DCSCS) to improve data accuracy. The Discrete Wavelet Transform algorithm is used to filter the collected data to clear the noise and find the data pattern. Then we use Back Propagation neural network to establish a data correction model. Finally, we use the model to correct the other data. Our experiment shows that the meteorological data corrected by our method performs a very good result both in maximum error, mean error and correlation coefficient.

Key words: Meteorological parameter monitoring, WSN, data correction, DWT

INTRODUCTION

Meteorological parameter monitoring (Kusmierek-Tomaszewska *et al*., 2012) performs an extremely important role in various fields, including weather forecast, agricultural (Ameur *et al*., 2001), traffic, etc. Traditional method of collecting meteorological parameter mainly depends on ground weather station which is equipped with professional meteorological equipment, but high cost and inconvenience to be deployed. With the advancement in wireless communications, Micro-Electro-Mechanical Systems (MEMS) technology, wireless sensor network has been considered as a promising tool for meteorological parameter monitoring (Akyildiz *et al*., 2002).

When using WSN to monitoring the meteorological parameter, there are two problems must be considered. The first is cost constrain. In general, the WSN node is cheap, it can be deployed large-scale (Vieira *et al*., 2003). However, the professional meteorological sensors equipped with the node are often very expensive and it is impractical to equip each node with professional sensors. Hence, researchers prefer to use some cheap meteorological sensors, such as on-board sensors or other non-professional sensors (Fang *et al*., 2010), to collected desired data in the applications which doesn't require high accuracy data. Benghanem (2009) focuses on the development of a Wireless Data Acquisition System (WDAS) and explains the design and implementation of WDAS in details. Gallart *et al*. (2011) deploys a low cost WSN for an environment monitoring in a local street, capable of providing sufficient data for evaluating both the T and RH gradients along the length of the street. The second is the data accuracy. When sensors have been deployed outdoor, the collected data must within a tolerable range. Otherwise, it may lead to a wrong decision. However, few researchers take this into account.

The work in this study is motivated from the finding that due to the non-professional production process and packaging technology, the cheap meteorological sensors are more sensitive to the ambient environment. The data accuracy can't be guaranteed. As shown in Fig. 1, purple dotted line denotes collected data while black line denotes standard data. The temperature data of four days collected by our monitoring system and standard temperature data acquired from national ground weather station are compared, where purple dotted line denotes collected data while black line denotes standard data. From the figure, we can find that in the four days, compared with standard data, the collected data shows some features. The first feature, the collected data is more sensitive to the ambient environment. The data curve fluctuates more obviously. The second feature, when the time earlier than about 6:00 am or later than about 18:00 pm, the two data curves have little difference but in other periods between about sunrise time and about sunset time, the collected data is higher than standard data more or less. Due to the two features, we must process the collected data before we use it.

Corresponding Author: Xingming Sun, No. 219, Ningliu Road, Pukou, Nanjing, Jiangsu, 210044, China
Tel: 025-58731575 Fax: 025-58731575

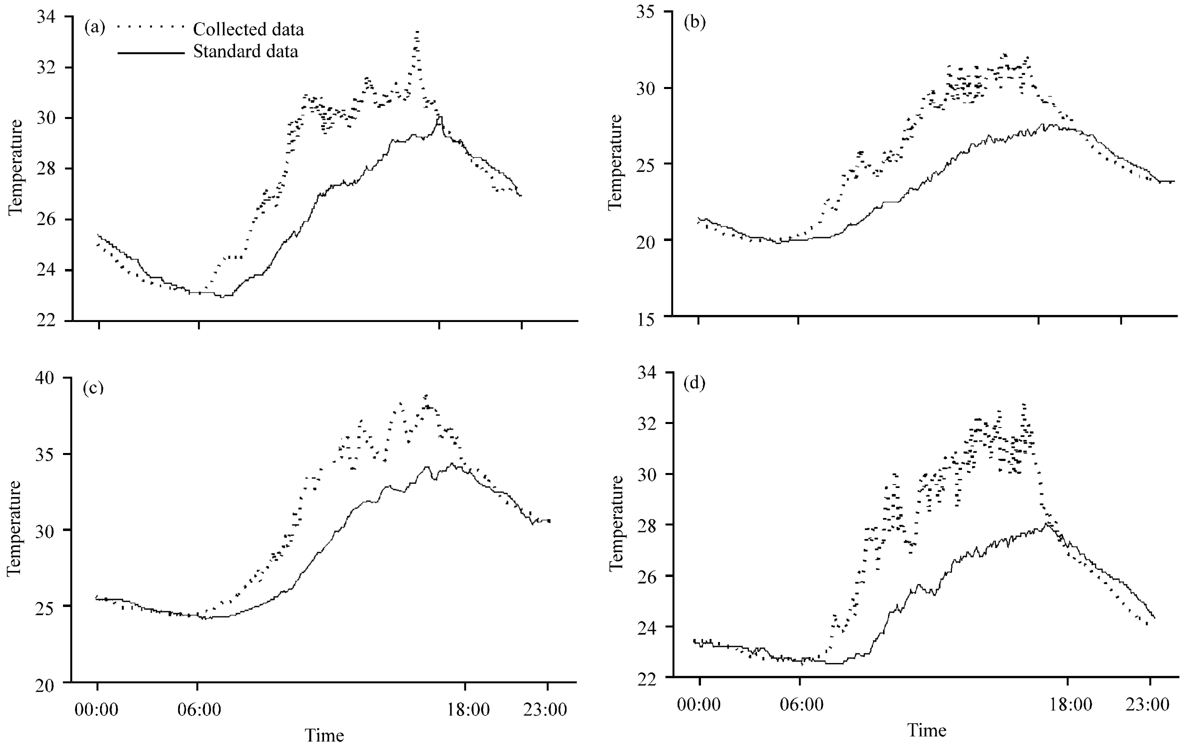

Fig. 1(a-d): Collected temperature data and standard data (a) 06-06-2013, (b) 14-06-2013, (c) 17-06-2013 and (d) 28-06-2013

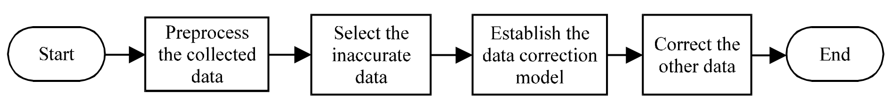

Fig. 2: Data correction scheme for cheap sensor

Following this finding, we propose the data correction scheme for cheap sensor, or DCSCS for short, as the data procession method to make the collected data closer to standard data. DCSCS contains two basis algorithms, the Discrete Wavelet Transform (DTW) (Burrus and Gopinath, 1998) and the Back Propagation (BP) neural network. The DTW algorithm is responsible for filtering the noise of the data to make the data curve smoother while the BP neuron network is used to training a data correction model to correct the collected data. We use the temperature data collected by our monitoring system which is composed of 30 MICAZ nodes, to experiment and evaluate the DCSCS through maximum error, mean error and correlation coefficient. The result shows that DCSCS not only can reduce the maximum and mean error, but also improve the correlation coefficient.

DESIGN OF DCSCS

The data correction scheme for cheap sensor runs on the database server. Each sensor node in our system transmits the data to the database server periodically. When enough data is collected, we can establish a data correction model for the node. In our experiment, it shows that when the amount of data is more than five days, the data correction results have little difference. So we adopt the data of five days to establish our model. As shown in Fig. 2, the DCSCS includes four steps: data preprocessing, training data selection, model establishing and the data correction. The four steps will be present in details in this section.

Data preprocessing by discrete wavelet transform: The first feature of collected data is mentioned in first section. The collected data is more sensitive to the ambient environment. To find the data pattern, we use Discrete Wavelet Transform algorithm to filter the noise of collected data.

The Discrete Wavelet Transform (DWT) is a popular signal processing technique suitable for several application areas. It decomposes its input signal components into deferent frequency bands. The Inverse

Discrete Wavelet Transform (IDWT) reconstructs a signal from this decomposition. Mallat formalized the DWT and showed it to be a computationally efficient process (Shensa, 1992). The decomposition algorithm is expressed by the equations Eq. 1 and 2:

$$x_{\alpha,L}[n] = \sum_{k=0}^{K-1} x_{\alpha-1,L}[2n-k]g[k] \tag{1}$$

$$x_{\alpha,H}[n] = \sum_{k=0}^{K-1} x_{\alpha-1,L}[2n-k]h[k] \tag{2}$$

where, $X_{\alpha,L}[n]$ is the nth approximation coefficient at the αth stage, $x_{\alpha,H}[n]$ is the nth detail coefficients at the αth stage, $x[n]$ is the input discrete signal, $g[n]$, $h[n]$ are the dilation coefficients corresponding to the approximation functions and detail functions (Lindsay *et al*., 1996), respectively. K is the length of $x_{\alpha,L}[n]$ or $x_{\alpha,H}[n]$ and if the length of $x[n]$ is N, then K equals $N/2^{\alpha}$.

In this study, we are only concerned with the results of de-noising. Through a lot of experiments, we select the 'db1' basic wavelets to decomposition the input signal. Furthermore, this decomposition is repeated to further increase the frequency resolution. The approximation coefficients is decomposed by high and low pass filters and then down-sampled. This is represented as a binary tree with nodes representing a sub-space with different time-frequency localization. At each stage, the signal is decomposed into low and high frequencies. The three-stage DWT decomposition filter bank used in our experiment is shown in Fig. 3.

Select the inaccurate data which need to be corrected: After the preprocessing of collected data, we need to determine which data need to be corrected. First, we give some definition of the symbols which appear in the Fig. 4:

t_r	=	The sunrise time, in different season, it has different value
δ_r	=	The buffer time between the sunrise time and the time when the sensor is affected by sunlight in the morning. It changes with the weather conditions and the season
t_s	=	The sunset time, in different season, it has different value
δ_s	=	The buffer time between the sunset time and the time when the sensor returns to normal status in the afternoon. It changes with the weather conditions and the season too

In the Fig. 4, the purple dotted line denotes the filtered data while the black line denotes the standard

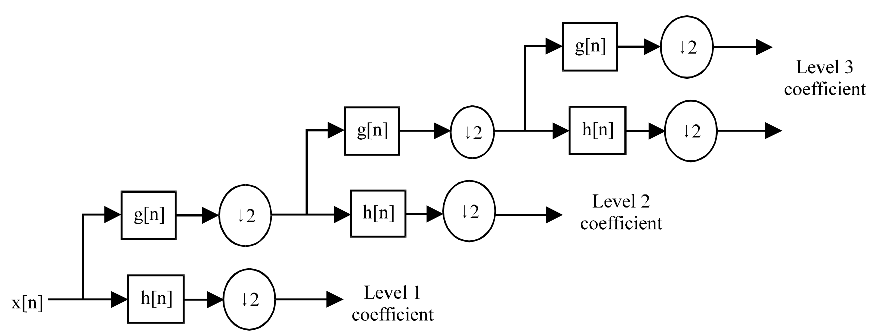

Fig. 3: Three-stage DWT decomposition filter bank

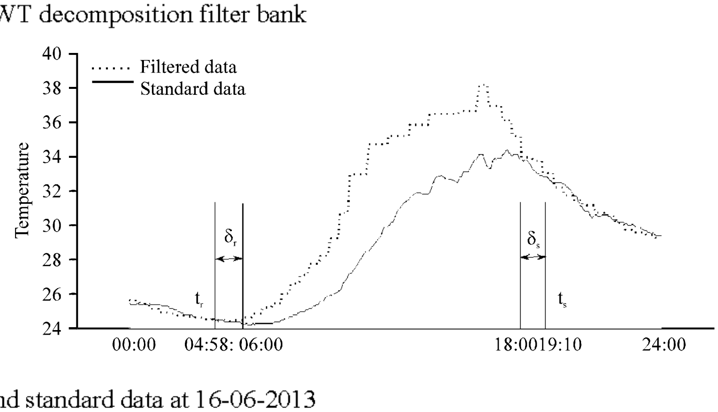

Fig. 4: Filtered data and standard data at 16-06-2013

data. To find the moment when difference begins to expand or reduce, we use the sunrise time t_r and the buffer time δ_r to determine the start time of the data which need to be corrected. In the same way, we use t_s and δ_s to determine the end time. From the figure, we can find the data which is collected between $t_r+\delta_r$ and $t_s-\delta_s$, is the inaccuracy data. So we can get our final data by the following Eq. 3:

$$y(t)=\begin{cases} x_{dwt}(t) & t \le t_r+\delta_r \\ f(x_{dwt}(t)) & t_r+\delta_r < t < t_s-\delta_s \\ x_{dwt}(t) & t \ge t_s-\delta_s \end{cases} \tag{3}$$

where, $X_{dwt}(t)$ denote the filtered value at time t, f(x) is the data correction model function which will be elaborated in this section.

Data correction model establishing based on back propagation neural network: To make the collected data closer to the standard data, we use BP neuron network to establish data correction model. The model will be used to correct the other data.

BP neural network is a backward propagation of errors and multilayered feed-forward neural network (Han, 2006). It contains one input layer, one or more hidden layers and one output layer. The topology of BP neural network with one hidden layers is shown in Fig. 5.

The training result is decided by many factors, including input and out data, network layer, node number and transition function, etc. From the input and output data aspect, in our study, both the input and output are the temperature data. The output data is acquired from national ground weather station (NO.58238, Nanjing), the input data is the inaccuracy data selected in 2.2 section.

From the network layer aspect, it has been proved that if the hidden layer has enough neurons quantity, the BP neural network can realize any nonlinear mapping. However if the number of neurons is too large, the calculation efficiency will declines intensively (Ding and Xu, 2004). In this study, the temperature data is chosen to train a correct model, so the number of neurons in input and output layer is 1 and 1, respectively. The number of neurons in the single hidden layer can be calculated by Eq. 4:

$$n_1=\sqrt{n+m}+\alpha \tag{4}$$

where, n, m and n_1 denote the number of input, output and hidden neurons, respectively and α means a constant value among 1 to 10. Generally speaking, the value of input and output is fixed. We can get the different number of neurons of hidden layer by changing the value of α. Through a large number of experimental comparisons, we set α as 6 in this study.

Beside the factors mentioned above, the choice of the hidden layer function and output layer function has a great influence on the prediction accuracy of BP neural network. In normal case, the transition function of node in the hidden layer is Logsig function or Tansig function and the transition function of node in the output layer is Tansig function or Purelin function (Shi *et al*., 2011).

Data correction: After the establishing of data correction model, we have got our correction function f(x). Then we

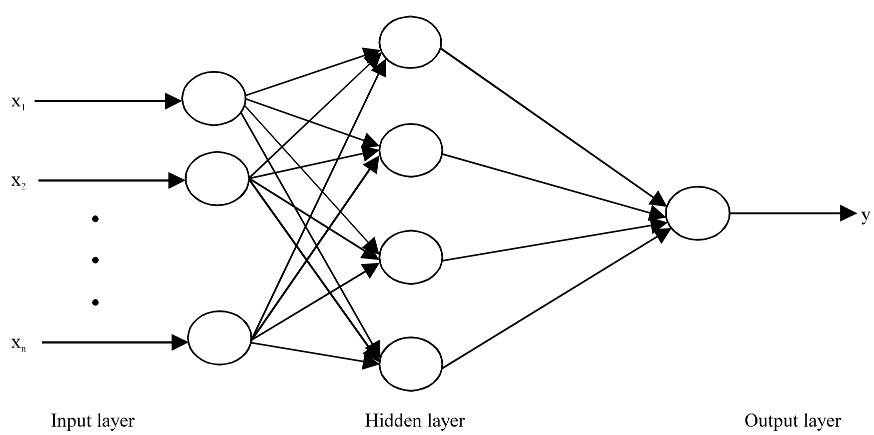

Fig. 5: BP neural network

use the Eq. 3 to get our final data. Notice that, a correction model can not fit all seasons or weather conditions. Even in the same month, we should build different data correction model corresponding to the weather condition.

EXPERIMENT AND ANALYSIS

In our meteorological parameter monitoring WSN, we select the MICAZ node and MDA300 data acquisition board as our sensor station. Both of them are cheap and convenience. Furthermore, the MDA300 is an extremely versatile data acquisition board that also includes an onboard temperature/humidity sensor. After a proper calibration an accuracy of 0.2°C can be achieved by the temperature sensor. The humidity sensor, SHT15, provides a digital and fully calibrated output which is easy to integrate. The packaged MDA300 and MICAZ node are deployed on the proof of our lab. There are 30 nodes totally, they communicate with each other wirelessly. The details of node package and the standard package in the weather station are shown in Fig. 6; the location of nodes is shown in Fig. 7.

The collected temperature data, from 1st-10th July, is adopted as our experiment data. In the forty days, there is different weather condition such as cloudy, sunny, rainy. In our experiment, we choose cloudy day which is the most common weather in June and July, to train our data correction model and for further analysis.

First, we load the temperature data from 12th-16th June and filter the noise using DWT algorithm. For each day, the data is collected at June, the sunrise

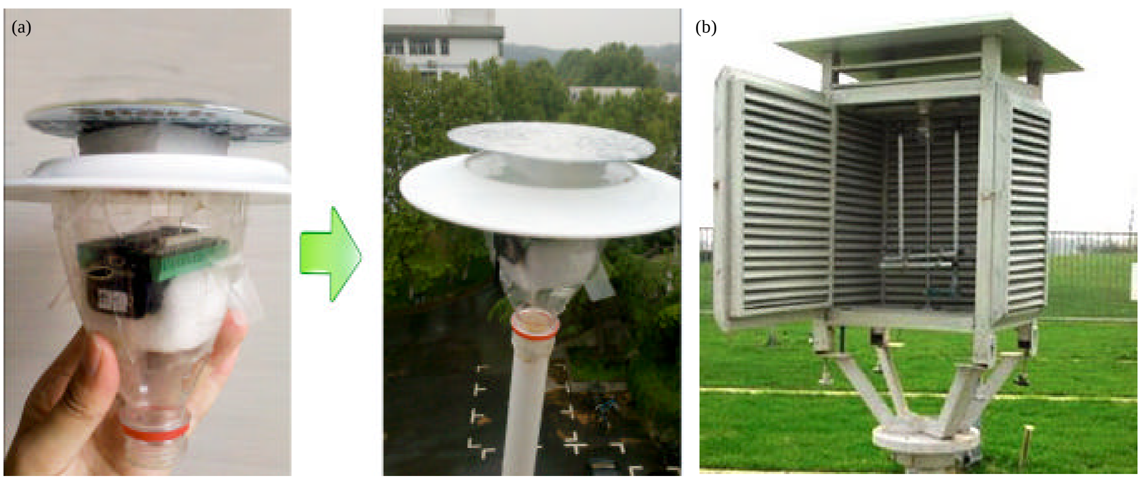

Fig. 6(a-b): (a) Package of our node and (b) Standard package

Fig. 7: Location of nodes

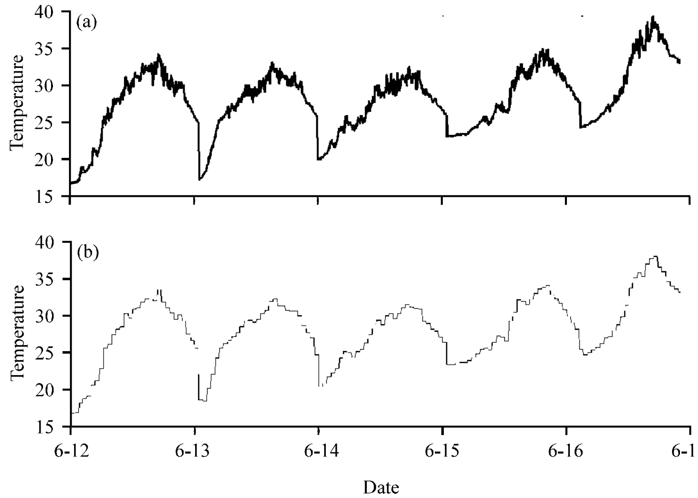

Fig. 8(a-b): Collected inaccurate data and filtered inaccurate data (a) Collected inaccurate data and (b) Filtered inaccurate data

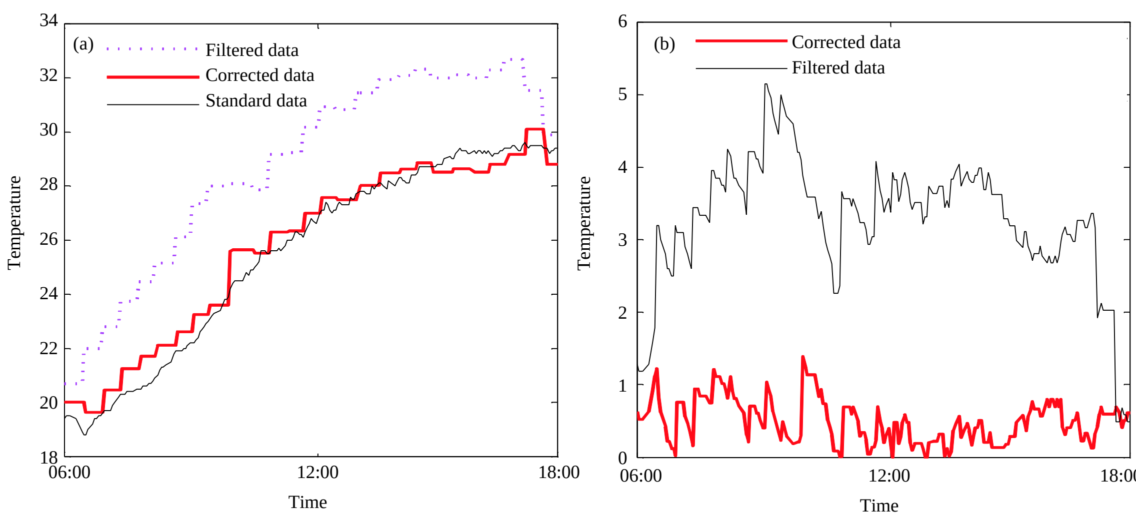

Fig. 9(a-b): Result of 4th June (a) Correction result and (b) Error statistics

time t_r is about 04:58 am and the sunset time t_s is about 19:10 pm. By analysis, we set δ_r equals 60 min while δ_s equals 70 min. Then, we can get the time range if inaccurate data, from 06:00 am to 18:00 pm. Figure 8 shows the collected inaccurate data and filtered inaccurate data of the five days.

From the Fig. 8, we can find that the DWT algorithm can both remove the noise and retain the sharp details of the signal.

Second, the filtered data and the standard temperature data are used as input data and output data for BP neuron network. After the training processing, a data correction model can be established.

The last step is to correct the previous and future data by the model established in the second step. The data of 4th, 17th and 28th June is used to test our data correction model. The result is show in Fig. 9-11. In the left of the Fig. 9, the purple dotted line denotes the filtered data by DWT; the red line denotes the scorrected data by model; the black line denotes the standard data. In the right of Fig. 9, the red line denotes the absolute error between filtered data and standard data, the black line denotes the absolute error between corrected data and standard data. Figure 10 and 11 are the same.

From the Fig. 9-11, it is obviously that our model performs very well in data correction. We analysis the maximum error, mean error, correlation coefficient of filtered data and corrected data with standard data. The result is shown in Table 1. Both the maximum error and

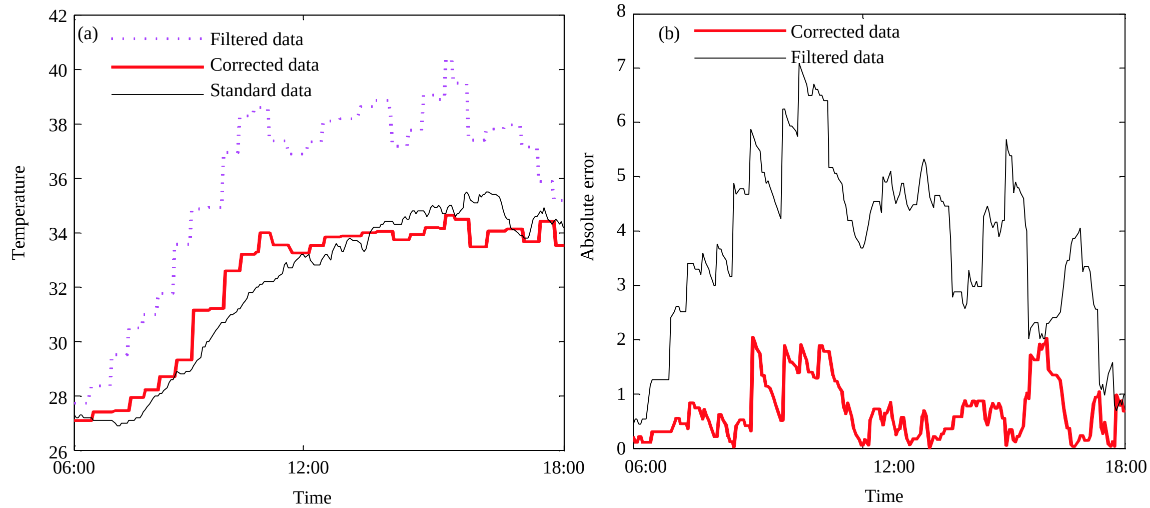

Fig. 10(a-b): Result of 17th June (a) Correction result and (b) Error statistics

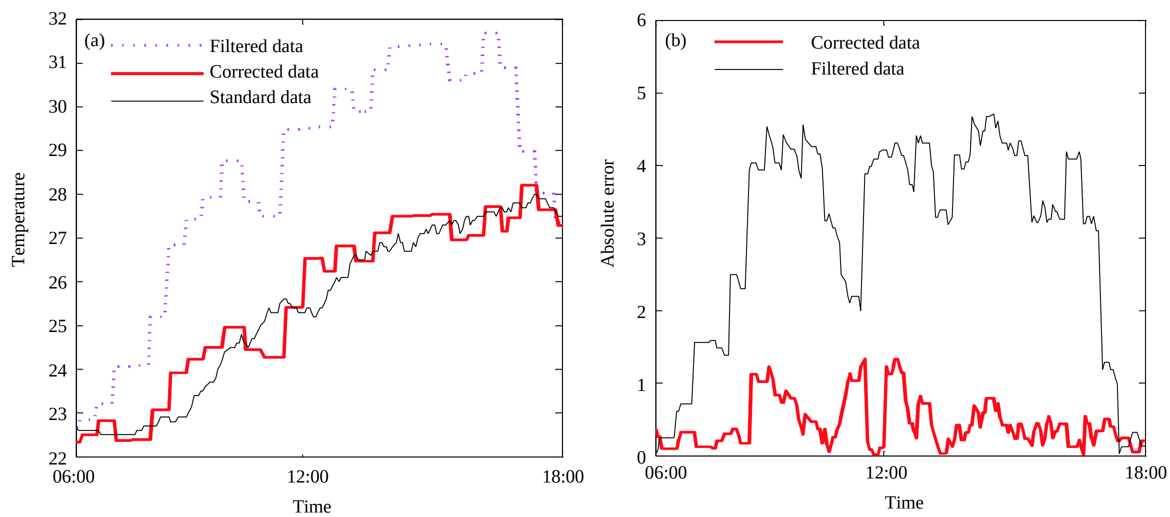

Fig. 11(a-b): Result of 28th June (a) Correction result and (b) Error statistics

Table 1: Correction results

	Maximum error		Mean error		Correlation coefficient	
Time	Filtered	Corrected	Filtered	Corrected	Filtered	Corrected
2013-06-04	5.15	1.39	3.19	0.57	0.9292	0.9667
2013-06-17	7.08	1.34	3.80	0.58	0.8849	0.9379
2013-06-28	4.17	2.04	3.02	0.45	0.8697	0.9594

mean error have been greatly reduced. The correlation coefficient is higher than 0.93, it means that the corrected data and standard data are strongly positive correlation. The three parameters reflect the accuracy of the data correction model.

CONCLUSION

In this study, a data correction scheme for cheap meteorological monitoring system is studied based on Discrete Wavelet Transform and BP neuron network. We first use DWT algorithm to filter the data collected by WSN to find the pattern. Then the filtered data is used as the input while the standard date collected by national ground weather station as the output of the BP neuron network to get our data correction model. Finally, we use this model to correct the other data. The experiment result shows that the corrected data performs very well, both the maximum error and mean error get greatly reduced, the correlation coefficient is also improved about 4-9% points.

In the future, we will use our low cost meteorological monitoring system to collect more meteorological parameters, such as the solar radiation, to quantitatively analyze the relationship between temperature and the other meteorological parameters. Furth more, with the increase of meteorological data, the scope of data correction model will be expanded.

ACKNOWLEDGMENTS

This study is supported by the NSFC (61232016, 61173141, 61173142, 61173136, 61103215, 61373132, 61373133, 61300238), GYHY201206033, 201301030, 2013DFG12860, BC2013012, BK20131004 and PAPD fund.

REFERENCES

Akyildiz, I.F., W. Su, Y. Sankarasubramaniam and E. Cayirci, 2002. Wireless sensor networks: A survey. Comput. Networks, 38: 393-422.

Ameur, S., M. Laghrouche and A. Adane, 2001. Monitoring a greenhouse using a Microcontroller-based meteorological Data-acquisition system. Renewable Energy, 24: 19-30.

Benghanem, M., 2009. Measurement of meteorological data based on wireless data acquisition system monitoring. Applied Energy, 86: 2651-2660.

Burrus, C.S. and R.A. Gopinath, 1998. Introduction to Wavelets and Wavelet Transforms: A Primer. 1st Edn., Prentice Hall, Upper Saddle River, New Jersey, USA.

Ding, H. and B. Xu, 2004. Investigation of speech separation of cross talk resistant adaptive noise canceller. Acta Acustica, 1: 37-42.

Fang, Z., Z. Zhao, X. Cui, L. Du and D. Geng *et al*., 2010. Micro sensor network node design for meteorological parameter monitoring. Proceedings of the IET International Conference on Wireless Sensor Network, November 15-17, 2010, Beijing, China, pp: 123-128.

Gallart, V., S. Felici-Castell, M. Delamo, A. Foster and J.J. Perez, 2011. Evaluation of a real, low cost, urban WSN deployment for accurate environmental monitoring. Proceedings of the IEEE 8th International Conference on Mobile Adhoc and Sensor Systems, October 17-22, 2011, Valenica, Spain, pp: 634-639.

Han, L.Q., 2006. Artificial Neural Network Tutorial. Beijing University of Posts and Telecommunications Press, Beijing, China, pp: 22-30.

Kusmierek-Tomaszewska, R., J. Zarski and S. Dudek, 2012. Meteorological automated weather station data application for plant water requirements estimation. Comput. Electron. Agric., 88: 44-51.

Lindsay, R.W., D.B. Percival and D. Rothrock, 1996. The discrete wavelet transform and the scale analysis of the surface properties of sea ice. IEEE Trans. Geosci. Remote Sens., 34: 771-787.

Shensa, M.J., 1992. The discrete wavelet transform: Wedding the a trous and Mallat algorithms. IEEE Trans. Signal Process., 40: 2464-2482.

Shi, F., X.C. Wang and L. Yu, 2011. Analysis of Thirty Cases about MATLAB Neural Network. BeiHang University Press, Beijing, China.

Vieira, M.A.M., C.N. Coelho Jr., D.C. da Silva Jr. and J.M. da Mata, 2003. Survey on wireless sensor network devices. Proceedings of the IEEE Conference on Emerging Technologies and Factory Automation, Volume 1, September 16-19, 2003, Lisbon, Portugal, pp: 537-544.