JAST, Vol. 6, No. 1, pp 37–44

JAST Journal of Aerospace Science and Technology

Research Note

Gyroscope Random Drift Modeling, Using Neural Networks, Fuzzy Neural and Traditional Time-Series Methods

A. A. Nikkhah[1], P. Heydari[2] , H. Khaloozadeh[3], A.P. Heydari[4]

In this paper statistical and time series models are used for determining the random drift of a dynamically Tuned Gyroscope (DTG). This drift is compensated with optimal predictive transfer function. Also nonlinear neural-network and fuzzy-neural models are investigated for prediction and compensation of the random drift. Finally the different models are compared together and their advantages are discussed.

INTRODUCTION

Dynamically Tuned Gyroscope (DTG) is a kind of Two-Degree-of-Freedom mechanical gyroscopes in which a dynamically tuned flexure and gimbal mechanism both support the rotor and provide angular freedom about axes perpendicular to the spin axis. In this type of gyroscope, spinning rotor is related to gimbals by means of elastic torsion bars and rotor assay is spinning in a special speed, which is called tune speed.

In this special speed, because of the adjustment of the gimbal inertia, flexure spring rate, or the spin rate of a rotor suspension system, the dynamically induced (negative) spring rate cancels the spring rate of the flexure suspension; and makes the gyro free for small angles of rotation. In Inertial Navigation Systems (INS) based on Stable Platforms, DTG works as free gyro or Integrating gyro (open loop), and in Strapped down Inertial Navigation Systems (SINS), DTG works as Rate Gyro (with a closed- loop control of gyro). Usually DTGs are made with one or two gimbals; however, theoretically it may have three or more gimbals [18].

The dynamically tuned gyroscope (DTG) is widely employed for its high reliability, low cost and long life. However, due to the influence of the DTG's making techniques and inherent structural character, drift is an intrinsic property of the DTG, and drift error is one of the main factors which affect the precision of the DTG or even the inertial system [11]. In order to reduce the drift error, some drift models, including mathematical models and physical models, are initiated to find the law of the drift error and then compensate for it.

Mathematical modeling of gyroscope random drifts has been started with prototype production of the gyroscope simultaneously, and it is still continued. References [1-10] are several examples in this category. Models with Kalman filter and state space realization are presented in [1], [4]. Also some models have been introduced for non stationary random processes in [2] and the Reference [10] only shows statistical modeling studies.

In recent years in addition to linear traditional methods, nonlinear methods are also being used, especially neural network method, which is a powerful nonlinear tool, [6], [7]. In reference [6], a gray first order model with wavelet processing is used for random drift modeling. Fan [5] addressed a grey modeling method based on the grey theory which was employed in the gyro drift model. However, it was only roughly done without any precision.

Some of the above-mentioned models have advantages based on the type of the random drifts. For example, random drift with the long period, which is studied more has different behaviors in comparison with the short period random drift. In long periods

1. *Assistant Professor, Dept. of Aerospace Eng., K.N.Toosi Univ. of Tech., Tehran, Iran, Email:Nikkhah@kntu.ac.ir.*
2. *Faculty Member, Islamic Azad Univ., Roudehen Branch, Tehran, Iran.*
3. *Associate Professor, Dept. of Electrical Eng., K.N.Toosi Univ. of Tech., Tehran, Iran.*
4. *M.Sc. Student, Dept. of Electrical Eng., Amirkabir Univ. of Tech., Tehran, Iran.*

of time, gyro random deviation is non-stationary, and the models presented for them are no more valid and accurate for short period deviations of gyro. As a result, every produced gyroscope based on its design, and assembly process, has a different random drift, and needs its special studies.

In this paper, statistical and time series models as well as nonlinear neural-network and fuzzy neural models are used for determining the random drift of a Dynamically Tuned Gyroscope (DTG); and based on the models mentioned above, the short period random drift is compensated for with an optimal predictive transfer function.

The rest of the paper is organized as follows. The first section briefly describes the time series statistical modeling on DTG random drift time series. The second section discusses linear time series models AR (Auto Regressive), MA (Moving Average) and ARMA (Auto Regressive Moving Average), for DTG random drift modeling. The random drifts are compensated for using an optimal predictive transfer function. In the last two sections, nonlinear neural network and Fuzzy neural methods are used to predict and compensate for gyro random drifts.

STATISTICAL AND TIME SERIES MODELING

Statistical modeling of the random process is as an essential step in the modeling and description of random signals. The random process which occurs in the gyroscopes is called Random Drift. Exact definition and methods of random drift testing is found in [11]. Briefly, random drift is the variance of mean drift in different periods of time in the absence of any input signal. In Figure 1, the raw data of the measured output of random drift for a DTG is shown.

A characterization of the noise and other processes in a time series of data as a function of averaging time is Allan variance. It is one half the mean value of

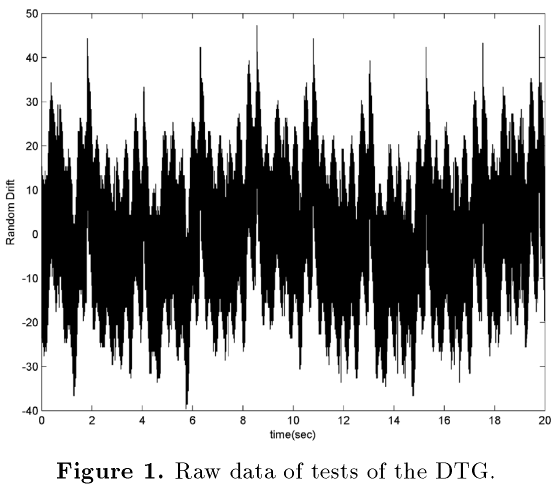


Figure 1. Raw data of tests of the DTG.

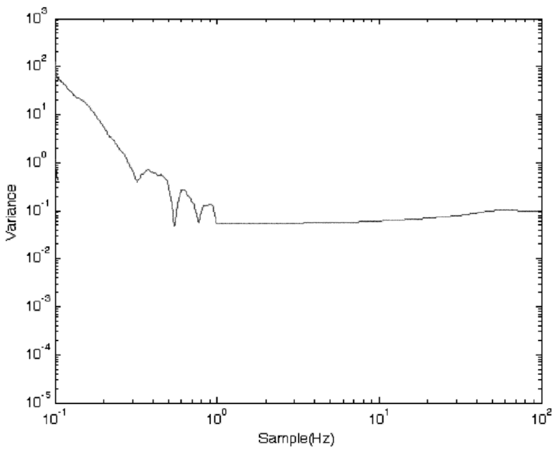


Figure 2. Variance of the random drift in different periods of time.

the square of the differences of adjacent time averages from a time series as a function of averaging time. Figure 2 shows the Allan variance of the DTG random drift. Flat curve at the end of the graph shows that the random process is stationary at the interested period of time.

Probability distribution density Function

The First and the most important characteristic of every random process is its probability distribution density function. Probability distribution density function transfers the space of input data to smaller subspaces. For example numbers between zero and twenty can be subdivided into ten equal sections, of which every one contains 2 numbers, and then using the simple relation of:

$$P(n_i) = \frac{n_i}{n} \tag{1}$$

the occurrence probability of each section can be presented as a number between zero and one. In Relation (1), n_i is number of data in each section, and n is the number of total data. Consequently, the area under the probability distribution density function will be equal 1. The simplest way to obtain probability distribution density function is to use traditional histograms in statistical methods.

In addition to deriving histograms, there are other more accurate methods where by fitting an analytic curve into the primary histogram, one can obtain more smooth results for probability function distribution density. The probability distribution density function of the raw data of Figure 1, by using the advanced methods in Matlab software, is obtained and shown in Figure 3. We can get important results from the shape of the probability distribution density function; for instance linear systems with Gaussian random input generate Gaussian output. According to Figure 3, unfortunately the random drift of the DTG does not have a Gaussian probability distribution density function. For better analysis, the periodogram of the output data is presented in the next section.

Autocorrelation Function

In random process analysis, autocorrelation function is a good criterion of randomness and data disordering of the process. For instance, in the case of white noise, which is %100 random process and is not practically predictable, autocorrelation function is equal to zero across its entire domain, except in the zero point, which has a value equal to the power noise. Autocorrelation function also is a superior criterion of predictability of a system [16], which shows the possibility of prediction and the proximity of the color noise to the white noise.

The autocorrelation function estimation for a discrete system is determined as:

$$\hat{r}(\tau) = \frac{1}{N}\sum_{t=1}^{N-\tau} x(t+\tau)x(t) \tag{2}$$

where N is number of data, $x(t)$ is data at time t. Most often the normalized autocorrelation function is used as:

$$x_r = \frac{\hat{r}(\tau)}{\hat{r}(0)} \tag{3}$$

Normalized Autocorrelation function for raw data of Figure 3 is sketched in Figure 4.

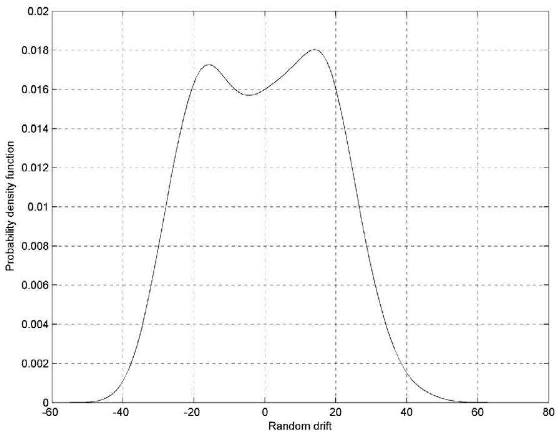


Figure 3. Probability distribution density function of raw data.

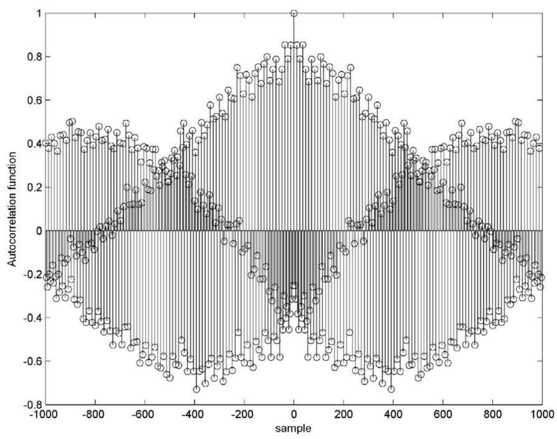


Figure 4. Normal autocorrelation function of raw data.

Power Spectral Density Function

To investigate frequency spectral in a random signal, power spectral density function is the best tool. Power spectral density function determines what would the frequency content of the signal be. For instance, if there is a systematic noise source in a specific frequency, one can omit it by using a filter or finding and eliminating the noise source.

For discrete data power, spectral density function can be obtained as:

$$\Phi_{xx}(\omega) = \frac{1}{2\pi}\sigma\hat{r}(\tau)e^{-i\tau\omega} \tag{4}$$

where $\hat{r}(\tau)$ is the autocorrelation function. In this work we use the following suitable form of power spectral density function estimation, which is called Periodogram:

$$\hat{P}_{xx} = \frac{|\hat{\Phi}_{xx}(\omega)|^2}{\omega_s} \tag{5}$$

The ω_s is the sample frequency. Figure 5 shows the periodogram of raw data of random outputs of the DTG. Because of using a wind-up anti aliasing filter, for data acquisition, there is no frequency more then 50 Hz. In fact, the high frequency noises are omitted by bandwidth of system and low pass filter. The other problem is in the frequency near 32 Hz. In this frequency, occurs an unusual jump occurs in the amplitude of the power spectral density function. The main reason for existence of this output signal is the magnetic effect of motor current which has a frequency component at 167 Hz. This frequency jump occurs because of the data sampling frequency (f_s) of system, and is equal to 100 Hz. Based on the theory of data acquisition, frequencies greater than 50 Hz are aliased in lower frequencies, and alias frequency can be obtained by the following relation [17]:

$$f_a = |n \times f_s - f| \tag{6}$$

From the Relation (6) we can find that the frequency 167 is aliased in frequency 33 Hz.

In order to reduce the noise with a frequency of 30 Hz, and simultaneously maintain the bandwidth of the system, we can use a notch filter. Notch filter is a kind of filter which can omit a very narrow band of frequency. In order to eliminate this noise, an appropriate notch filter near 33 Hz has been employed.

The periodogram of filtered data is shown in Figure 5. Figure 6 shows the power spectral density function of data, which passes through the notch filter. Also Figure 7 briefly shows that the probability distribution density function of the DTG random drift is very close to the Gaussian form. Hence we conclude that the DTG random drift is affected by a linear system; thus, linear models can describe it very well. One important demerit in using notch filter is omission of all the data in the frequency of 33 Hz, and the weakening of signals near this same frequency. In this case, by suitable insulation of input and output wiring, or use of suitable data acquisition rate higher than system bandwidth and omission of resultant noise in the frequency greater than the bandwidth, we can reduce the effects of this noise on output. Also one can conclude that this noise has nonrandom sources; thus, by using linear modeling and identification methods without using a notch filter, the drift can be reduced.

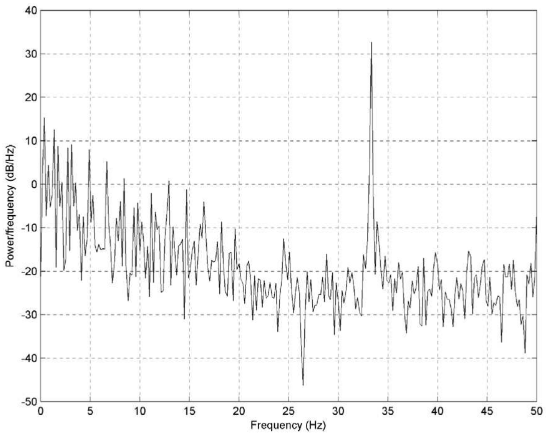


Figure 5. Periodogram of raw data.

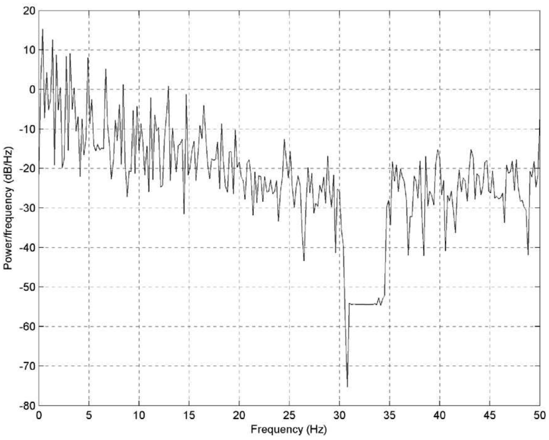


Figure 6. Periodogram of filtered data.

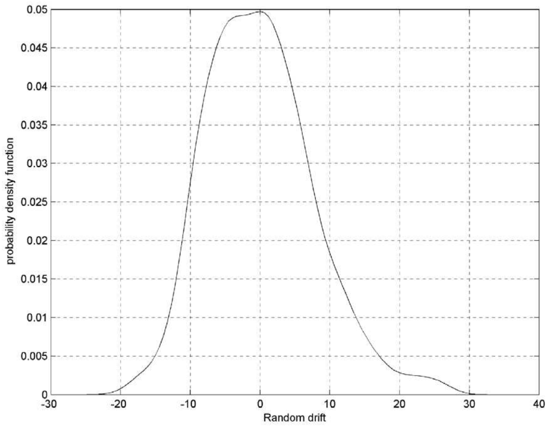


Figure 7. Probability distribution density function of filtered data.

LINEAR MODELS

In this section, we present AR (Autoregressive) and ARMA (Autoregressive Moving Average) models to describe time series by means of the output error method. Time series models have been obtained for raw data of DTG random drift (Figure 1). Table 1 shows the values of widely used AIC (Akaike Information Criteria) for various AR model orders from 1 to 9. The utilized form of AIC is the following:

$$AIC = N \log V_N(\theta_N) + 2p \quad (7)$$

where N is the number of data, and $V_N(\theta_N)$ is root mean square error of predictive model.

In spite of the good results for reduction of variance, AR models still have some problems. In Figure 8, the coefficient of 25 lags for model AR (25) has been shown. This figure shows that by increasing the lags (delays), the amplitude of the autocorrelation coefficients doesn't decrease. This diagram, which is called partial autocorrelation function [14], is an important representative feature of the time series. In the AR model, it is assumed that there is a white noise in the system. If this assumption is not valid, then AR model can not describe the system exactly. According to Figure 8, since the AR models autocorrelation coefficients are not reduced by increasing the lags, these models are poor for modeling the DTG random drift time series.

To address this problem, the ARMA has been used to model noise signal as a colored noise. Table 2 shows variance of the DTG random drift. According to this data, the best selection for term MA in the ARMA model depends on the selected AR model. For example, in a model with AR (4), obtain the best MA

Table 1. Different autoregressive models and AIC criterion for them.

AR	1	2	3	4	5	6	7	8	9
AIC	5.71	5.62	2.63	1.58	1.52	1.44	1.04	0.99	0.98

term selection as MA(2). Also when noise is a second order Markov process, with AR (5), the best MA term selection is MA (5).

Resultant transfer function for model ARMA (4, 2) is obtained as below:

$$\theta(z)=\frac{z^4+0.8534z^3+0.926z^2}{z^4-0.353z^3+0.03929z^2-0.9666z+0.384} \quad (8)$$

According to the obtained model, an optimal predictable transfer function for prediction and reduction of the random drift of the DTG is proposed. However, existence of some small nonlinear factors, which can be found in all systems, leads us to the use of nonlinear models.

Also, the result of compensation of output random drift of the gyroscope by one of the introduced methods (ARMA (4, 2)) is shown in Figure 9. As it can be seen in Figure 9, the DTG's random drift is reduced to 1/10 of its primary (raw) value.

NEURAL NETWORK MODEL

The purpose of this section is to give an artificial neural network (ANN) for nonlinear modeling of DTG random drift. An artificial neural network (ANN) is a mathematical model consisting of a number of highly interconnected processing elements organized into layers, geometry and functionality, which have been likened to that of the human brain. The network has one input

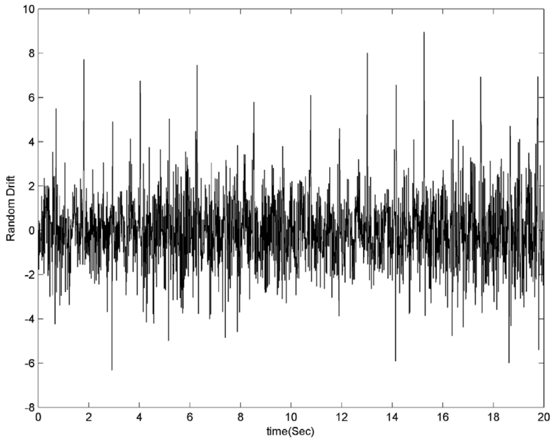


Figure 9. Random drift of the gyro after prediction and processing of raw data.

layer, one hidden layer and one output layer. Among the various kinds of ANN approaches that exist, the Multi Layer Perceptron (MLP) architecture with back propagation learning algorithm has been chosen in this study. It has been proven that MLP neural network with only one hidden layer of neurons and a specific type of activation function (e.g. sigmoid function or hyperbolic tangent) can approximate any functional relation arbitrarily well, provided that enough hidden neurons are available [12].A schematic description of the MLP is given in Figure 2. The tapped delay line (TDL) has been utilized for making the input layer with signal at the current time and the three previous input signals.

In our neural network simulation, we have used a neural network with a hidden layer containing 20 neurons. Additionally, the nodes of the input layer just propagate input values to the nodes of the first hidden layer. The input–output relationship between each node of the hidden layers is given by:

$$Y=\phi\left(\sum_{j=1}^{4}W1(j,1)x(t-j)+B1(j)\right) \quad (9)$$

where $x(t-j)$ is the output from the jth node of the Input layer, $W1(j,1)$ the weight of the connection between the jth node and the current node, and $B1$ is

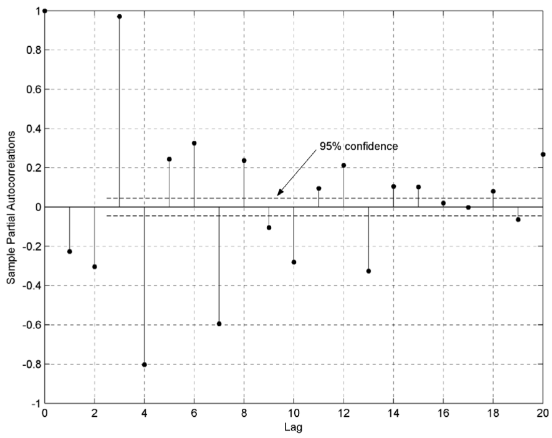


Figure 8. Partial autocorrelation function of data.

Table 2. Values of AIC criterion for different types of ARMA models.

MA	1	2	3	4	5
1	5.5053				
2	4.8434	3.9065			
3	2.1855	1.1226	0.9963		
4	1.7828	0.9237	1.7036	1.6720	
5	1.7693	1.7149	1.6898	1.6937	0.7981

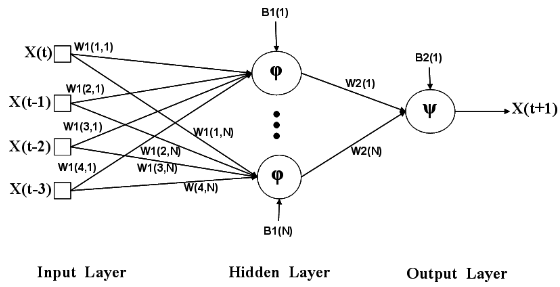


Figure 10. Schematic description of the MLP.

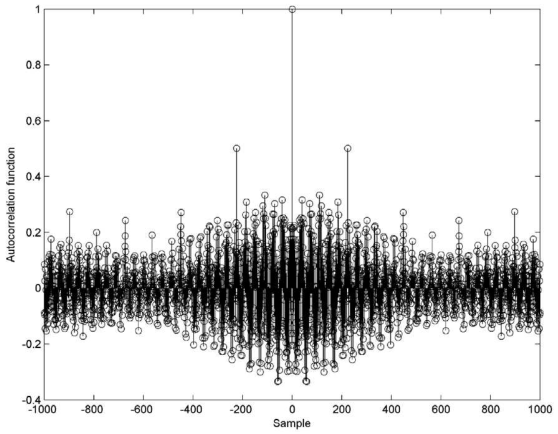


Figure 11. Autocorrelation function result after training process of proposed ANN.

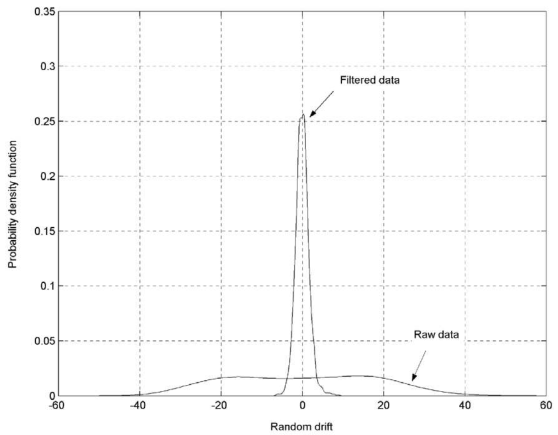


Figure 12. Probability density function of ANN modeling in comparison with the raw data.

the bias of the current node. ϕ is a function that can be nonlinear, *e.g.* hyperbolic tangent Eq. (9):

$$\phi(z) = \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)} \tag{10}$$

We have used ψ as a linear transfer function. $W2(j)$ are the weight of the connection between the jth node and the output node, and the $B2$ is the bias of output nod. During the training process, the weights and biases in the network are adjusted to minimize the error in obtaining high performance in the solution. At the end and during the training error, mean squared error is computed between predicted outputs and targeted outputs. There are various training algorithms used in neural network applications. In this study, the used training algorithms were Levenberg–Marquardt (LM) methods.

Figure 11 shows the autocorrelation function result after the training process of proposed ANN for comparison with Figure 4. The results show the autocorrelation function of the proposed ANN model is so close to white noise. The probability density function of ANN modeling in comparison with the raw data is given in Figure 12.

FUZZY NEURAL MODEL

In this section a class of fuzzy neural model referred to as ANFIS is introduced. The acronym ANFIS derives its name from adaptive neuro-fuzzy inference system. Figure 13 illustrates that the reason mechanism for the two if-then rules linear Sugeno fuzzy model under consideration has two inputs of x and y, and one output of z. The corresponding equivalent ANFIS architecture is as shown in Figure 13b, where each layer is described below.

Layer1: In this layer A_i and B_i are fuzzy sets and outputs are the membership function values of the premise part. Here, the membership function for A_i and B_i are characterized by the generalized bell function:

$$\mu(x) = \frac{1}{1 + \left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}} \tag{11}$$

where $\{a_i, b_i, c_i\}$ is the parameter set.

Layer2: In this layer, every node labeled $\prod$ used (x) as fuzzy AND. Outputs the layer are:

$$W_i = \mu_{A_i}(x) \times \mu_{B_i}(y) \tag{12}$$

Layer3: Output of this layer is called normalized firing strength:

$$\bar{W_i} = \frac{W_i}{W_1 + W_2} \tag{13}$$

Layer4: In this layer, every node function is:

$$\bar{W_i} f_i = \bar{W_i}(p_i x + q_i y + r_i) \tag{14}$$

where $\{p_i, q_i, r_i\}$ is parameter set. The parameter in this layer will be referred to as consequent parameter.

Layer5: the single node in this layer computes the overall output as summation of all incoming signals:

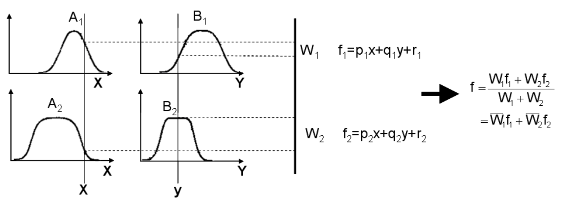
$$\sum \bar{W_i} f_i = \frac{\sum_i W_i f_i}{\sum_i W_i} \tag{15}$$

In the fuzzy model for prediction of $x(t+1)$, four inputs from time t to $t-3$ are used like in the neural network model. Each input has two Gaussian transfer functions; therefore, the total number of rules is 16. The learning of ANFIS models is performed by method of error back propagation.

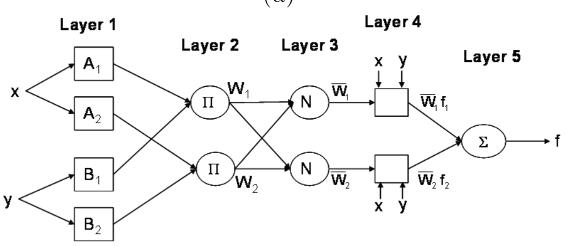
In Figure 14, periodogram of raw data (up curve) and processed data (down curve) on an ANFIS model is shown.

CONCLUSION

Physically un-modeled drifts of dynamically tuned gyroscopes, which are errors with random sources (disturbances and noises), has a serious effect on sensor accuracy. This paper investigates the use of linear and nonlinear methods. Based on the theories of Neural-Network, Fuzzy logic and Time series methods, actual test data about random drift of a DTG has been processed, and the random drift is compensated for and is reduced. According to the obtained results, the DTG's random drift is reduced to 1/10 of its primary (raw) value. In Table 3, qualification and efficiency of different linear and non-linear models are compared. This comparison shows that, with adequate qualification and simplicity in realization, the time series are the best solution for DTG random drift analysis and compensation.



(a)



(b)

Figure 13. Two if-then rules linear Sugeno fuzzy model with two inputs x and y, and one output z.

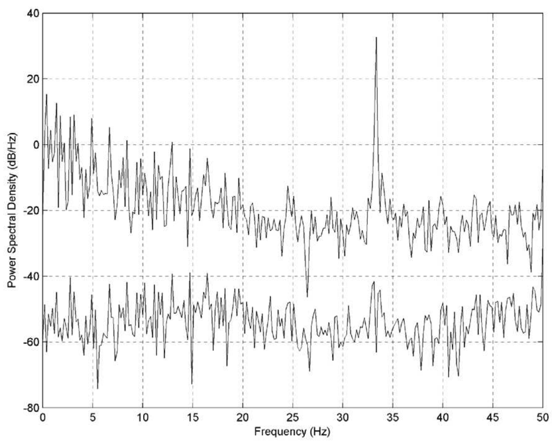


Figure 14. Periodogram of raw data (up curve) and filtered data (down curve).

Table 3. Comparison between linear and nonlinear models.

Models	Standard deviation	AIC criterion	Number of parameters
AR(4)	2.1972	1.5817	4
ARMA(4,2)	1.5769	0.9237	6
ANFIS	1.8436	1.3568	96
TD-MLP NN	1.8953	1.4068	100
Raw data	17.8608	—	—

REFERENCES

1. Xiyuan C., "Modeling Random Gyro Drift by Time Series Neural Network and by Traditional Method", *IEEE int. Conf., Networks & Signal proceeding*, Nanjing, China, PP 810-813(2003).
2. Oraetz A.S., Sandberg H.J., "Stationary and Nan Stationary Characteristics of Gyro Drift Rate", *AIAA Journal*, **8**(10), PP 1765-1772(1970).
3. Sudhaker M.P., Zhang W., "Modeling Random Gyro Drift Rate by Data Dependent Systems", *IEEE Transactions on Aerospace and Electronic System*, **22**(4), PP 455-459(1986).
4. Hong J., Yang Y., "state Space Modeling of Random Drift Rate in High-Precision Gyro", *IEEE Trans. on Aero and Elect. Sys.*, **32**(3), PP 1138-1142(1996).
5. Fan C., Jin Z. and Tian W., "A Novel Hybrid Grey-Based Strategy for Improving the Model Precision of a Dynamically Tuned Gyroscope", *Measurement Science and Technology*, **14**, PP 759-765(2003).
6. Zhu R., Zhang Y. and Bao Q., "A Novel Intelligent Strategy for Improving Measurement Precision of FOG", *IEEE Trans. on Instr. and Meas.*, **49**(6), PP 1183-1188(2000).
7. Fan C., Jin G.Z., "Bayesian Neural Networks for Life Modeling and Prediction of Dynamically Tuned Gyroscope ", *Advances in Neural Networks*, PP 749-754(2004).
8. Xu, G., Tian W. and Jin Z., "An AGO-SVM drift Modeling Method for a Dynamically Tuned Gyroscope", *Meas. Sci. Techno.*, **17**(1), PP 161-167(2006).
9. Akcayir Y., Ozkazanc Y., "Gyroscope Drift Estimation Analysis in Land Navigation Systems", *Proceedings of 2003 IEEE Conference on Control Applications*, Istanbul, (2003).
10. Cooper J.R., "Statistical Analysis of Gyro Drift Test Data", MS.c. Thesis, Massachusetts Tnstitute of Technology, (1965).
11. IEEE Std., "IEEE Specification Format Guide and Test Procedure for Two Degree of Freedom Dynamically Tuned Gyros", *ANSI/IEEE STD 813*, (1989).
12. Cichocki A., Cichoki A, Unbehauen R., *Neural Networks for Optimization and Signal Processing*, John WILEY and Sons, (1995).
13. Haykin H., *Neural Networks*, Prentice-Hall Inc., (1994).
14. Soderstrom T. and Stoica P., *System Identification*, Prentice-Hall Inc., (1989).

15. Jung J. and Sun C., "Neuro-Fuzzy Modeling and Control", *Proceedings of the IEEE*, PP 378-406(1995).

16. Vaseghi S.V., *Advanced Digital Signal Processing and Noise Reduction*, John WILEY and Sons, (2000).

17. Oppenheim A.V., Willsky A.S., Nawab S.H., *Signals and Systems*, Prentice-Hall, (1997).

18. Brozgool L.E., *Dynamically Tuned Gyroscopes*, Maschinostroenie Publishing, (1989).