The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th '10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

A Time Variant Seasonal ARIMA Model for Lung Tumor Motion Prediction

K. Ichiji[1], M. Sakai[2], N. Homma[3], Y. Takai[4] and M. Yoshizawa[3]

[1] *Graduate School of Engineering, Tohoku University, 6-6-05 Aoba, Aoba-ku, Sendai 980-8579 JAPAN*
[2] *Center of Inform. Tech. Edu., Tohoku University, 41 Kawauchi, Aoba-ku, Sendai 980-8576 JAPAN*
[3] *Cyberscience Center, Tohoku University, 6-6-05 Aoba, Aoba-ku, Sendai 980-8579 JAPAN*
[4] *Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai 980-8575 JAPAN*
[1]*(Tel : 81-022-795-7130; Fax : 81-022-263-9163)*
[1]*(ichiji@yoshizawa.ecei.tohoku.ac.jp)*

Abstract: We propose a prediction method of lung tumor motion for real-time tumor following radiation therapy. An essential core of the method is a model building of time variant nature of the lung tumor motion. The method is based on a seasonal ARIMA model with an estimator of the time variant nature. The estimator provides the time variant period of the lung tumor motion by using a correlation analysis. The time variant SARIMA model can then predict complex lung motion by using the estimated period. The proposed method achieved highly accurate prediction of the average error 0.820±0.669[mm] at 0.5[sec] ahead prediction. This result is superior to other conventional methods at short- or mid-term prediction.

Keywords: Time series prediction, seasonal ARIMA, real time following radiation therapy, and lung tumor motion.

I. INTRODUCTION

To realize an effective radiation therapy, high power radiation dose must be given accurately to a limited area to enhance the effect of treatment and avoid irradiating to normal tissues to decrease adverse effects. Extracranial stereotactic radiation therapy (ESRT) is one of such advanced radiation therapy, and can irradiate accurate and sufficient dose to motionless tumor such as vesical tumor case [1]. However, lung tumor for example is not applicable target of the current ESRT because of its intra-fraction dynamic motion mainly due to the respiration.

One ideal way to give accurate and sufficient therapeutic dose to such dynamic tumor is real-time tumor-following radiation therapy (RTFT) [2]. RTFT can deliver dose to dynamic tumor continuously by moving the radiation sources or changing the shape of radiation area. For clinical use of RTFT, we need two technologies, i.e. real-time measurement technique of the tumor position and prediction technique of the tumor motion. X-ray fluoroscopy system will measure the position of the target in real-time. However, still we need to predict the position to compensate some time delays which is included in the radiotherapy instruments, such as control delay of radiation source and computational time for measurement of position and processing of time series.

Lung tumor moves mainly with patient's respiration and this motion is observed as a seasonal (i.e. cyclic) time series. There are many prediction methods for seasonal time series such as seasonal autoregressive integrated moving-average (SARIMA). However, the period of the lung tumor motion is time variant. Therefore, conventional SARIMA model is not applicable to this time series, because the SARIMA model assumes complete periodicity of the time series.

Homma *et al* developed a prediction system for lung tumor motion by using the SARIMA model [2,3]. This system converts the time variant period of target time series into a new time series with time invariant period, and achieved long-term prediction of the average error 1.05±0.99[mm] at 1[sec] ahead prediction [2].

In clinical use, we often need to predict short- or mid-term prediction at 0.1~0.5[sec] ahead. The desirable average error of short- or mid-term prediction is less than sub-millimeters. The prediction accuracy must then be more improved in such prediction.

In this paper, we propose a new prediction method of lung tumor motion for further improvement of the prediction accuracy. The basic idea of the method is to search more directly for reference points that are crucial for the prediction accuracy of the SARIMA model for the time variant seasonal time series. Simulation result by using clinical data will show that the accuracy of the

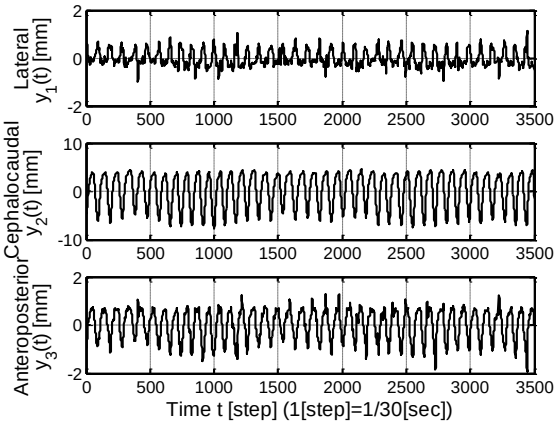

Fig.1. Three dimensional time series $\boldsymbol{Y}(t)$ of the lung tumor motion.

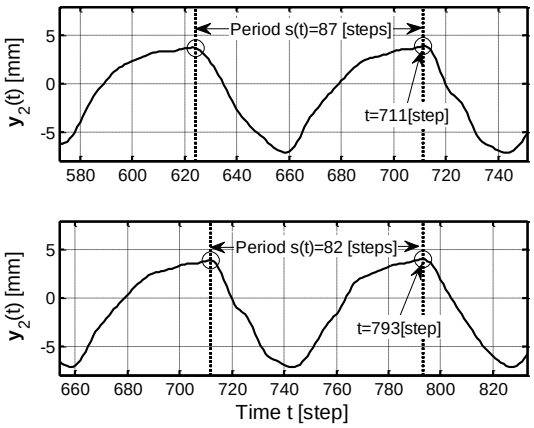

Fig.2. Two example of time variant periods as interval between peak and peak.

short- or mid-term prediction is improved by the proposed method.

II. TARGET TIME SERIES

In this study, we used a real data of lung tumor motion as target time series. We will show the time variant nature of the tumor motion.

1. Lung tumor motion

Three-dimensional time series of the motion of the marker implanted into the lung tumor was observed by a fluoroscopy system at Hokkaido University Hospital [4]. The sampling rate of time series was 30[Hz]. The time series was smoothed by using statistical filter and the Kalman filter for reduction of noise. In addition, the average value of the time series was removed.

The time series of lung tumor motion is expressed as follows.

$$\boldsymbol{Y}(t) = [\, \boldsymbol{y}_1(t),\, \boldsymbol{y}_2(t), \boldsymbol{y}_3(t)\,]. \quad (1)$$

where $\boldsymbol{y}_i(t),\ i = 1,2,3$, are lateral, cephalocaudal, and anteroposterior coordinates of the target tumor at time t[step] (1[step] = 0.033[sec]), respectively. For simplify, we use a one dimensional time series as $\boldsymbol{y}(t) = \{y(t), y(t-1), y(t-2), \ldots, y(2), y(1)\}$ for explanation of prediction methods.

Fig.1 shows the time series of the lung tumor motion. The lung tumor motion is quasi cyclic due to patient's respiration and cardiac motion. Period of the cyclic motion is approximately 90[steps] (=3[sec]).

2. Period of the tumor motion

Period of the lung tumor motion changes with time evolution. For example, time intervals between peak and peak are time variant as shown in Fig.2 Peak to peak periods are 87[steps] at time t=711[step] and 82[steps] at time t=793[step] respectively.

There are many prediction methods for seasonal time series such as SARIMA model and Holt-Winters Seasonal [5]. However, they assume that cyclic period of their target time series is constant. Therefore, we need to modify those seasonal methods for accurate prediction of the lung tumor motion.

III. PREDICTION METHOD

The prediction method proposed in this paper is composed of a prediction model and a period estimator as shown in Fig.3. The period estimator provides the cyclic period of the tumor motion estimated at each time. Then the prediction model will generate predicted values by using the estimated periods.

1. Prediction model: Seasonal ARIMA model

The general SARIMA model can be given as follows.

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D y(t) = \theta(B)\Theta(B^s)e(t) \quad (2)$$

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \quad (3)$$

$$\Phi(z) = 1 - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_P z^P \quad (4)$$

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \quad (5)$$

$$\Theta(z) = 1 + \Theta_1 z + \Theta_2 z^2 + \cdots + \Theta_Q z^Q \quad (6)$$

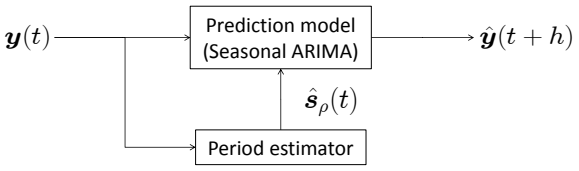

Fig.3. Schematic diagram of the proposed prediction system.

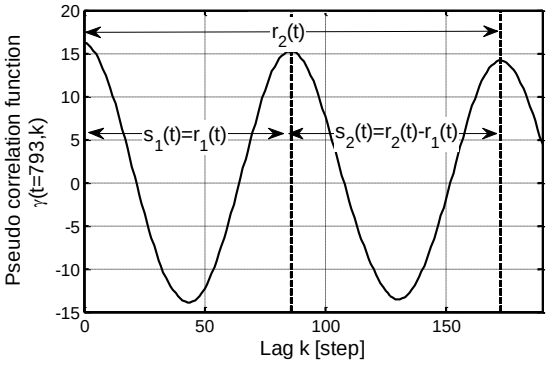

Fig.4. Pseudo correlation function $\gamma(t,k)$ at time t=793[step] and relationship between reference point $r_\rho(t)$ and estimated period $\hat{s}_\rho(t)$.

$$B^k x(t) = x(t-k) \tag{7}$$

where $e(t)$ is a Gaussian noise. Parameters d, D, p, P, q and Q imply differences, seasonal differences, autoregressive, seasonal autoregressive (SAR), moving average and seasonal moving average of the SARIMA model, respectively.

The SARIMA can express a wide range of time series by designing the parameters. In this study, we simplify the SARIMA model to avoid the over-fitting problem as $d = D = q = Q = p = 0$. The designed model is thus composed by SAR component only.

Then the prediction equation of the model can be expressed as follows.

$$\hat{y}(t+h) = \sum_{\rho=1}^{P} \Phi_\rho \cdot y(t - \rho \cdot s + h) \tag{8}$$

where $\hat{y}(t+h)$ is the prediction value at time $t+h$ of h[steps] forward. $\Phi_\rho, \rho = 1,2,\dots,P$ are weight coefficients of SAR model, P is the order of SAR model and s is a constant period (time invariant). In this case, the prediction value is a function of the past values at the corresponding phase.

2. Period estimator by using correlation analysis

To predict a quasi cyclic time series using by SARIMA model, we need to know a period of the lung tumor motion. In the followings, we will show how to estimate the time variant period.

A. Correlation function

We used a kind of correlation analysis to estimate the period of the lung tumor motion.

The pseudo correlation function we used is calculated by using two subsets of time series, e.g. $\boldsymbol{y}_t = \{y(t), y(t-1), y(t-2), \dots, y(t-w-1)\}$ and $\boldsymbol{y}_{t-k} = \{y(t-k), y(t-k-1), y(t-k-2), \dots, y(t-k-w+1)\}$. Subset $\boldsymbol{y}_t$ is a time series within a window width w[steps] from past time $t-w-1$ to current time t and $\boldsymbol{y}_{t-k}$ is a k[steps] delayed subset time series. We used window length w=90[steps] in this study.

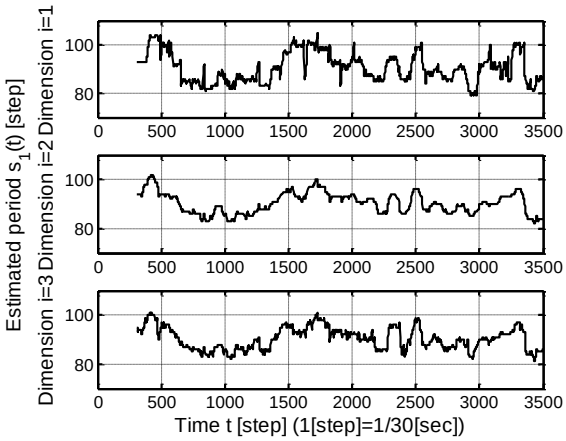

Fig.5. Time series of estimated period $\hat{s}_\rho(t)$ on each dimension i.

Then the pseudo correlation function as a function of time t and lag k will be given as follows.

$$\gamma(t,k) = \frac{1}{w}\sum_{j=0}^{w-1} y(t-j) \cdot y(t-k-j) \tag{9}$$

An example of the calculated correlation function at time t=793[step] is shown in Fig.4.

B. Estimation of time variant period

To estimate a period, we first search reference points as peak points of the correlation function as follows.

$$r_\rho(t) = \arg \max_{r_\rho(t-1)-l \le k \le r_\rho(t-1)+l} \gamma(t,k) \tag{10}$$

where l is the parameter to define the search area for the peak point.

The estimated period $\hat{s}_\rho(t)$ will then be able to calculate from the reference point $r_\rho(t)$ by using the following equation (Fig.4).

$$\hat{s}_\rho(t) = \begin{cases} r_\rho(t) - r_{\rho-1}(t), & \text{if } \rho > 1 \\ r_\rho(t), & \text{otherwise} \end{cases} \tag{11}$$

The initial value of the reference point is initialized by

$$r_\rho(1) = \rho \times \bar{s} \tag{12}$$

where $\bar{s}$ is the average value of the pre-estimated periods and is set $\bar{s}$=90[steps] in this study.

Fig.5 shows an example of the estimated period from the time series of the lung tumor motion.

3. Prediction model: Time variant Seasonal ARIMA model

We now propose a modified SARIMA model for time variant periodic time series by using the estimated period.

Modifying the equation (8), the prediction equation of the proposed SARIMA model can be given as follows.

$$\hat{y}(t+h) = \sum_{\rho=1}^{P} \Phi_\rho \cdot y(t - r_\rho(t) + h) \qquad (13)$$

Note that there is no estimated period $\hat{s}_\rho(t)$ in equation (13). We can use reference point $r_\rho(t)$ instead of estimated period as $r_\rho(t) = \sum_{\rho=1}^{P} \hat{s}_\rho(t)$.

IV. RESULTS and DISCUSSIONS

We have evaluated prediction performance of the proposed method by comparing with these of conventional SARIMA and our previous SARIMA model [2]. For this evaluation, we used the clinical data of the lung tumor motion introduced in section 2 as test data. The parameters of the proposed method and the conventional methods were designed as P=2 and Φ_ρ=$1/P$.

We calculated a prediction error as the Euclidean distance between real and predicted positions as follows.

$$e(t+h,h) = \sqrt{\sum_{i=1}^{3} (\hat{y}_i(t+h) - y_i(t+h))^2} \qquad (14)$$

Note that the prediction error $e(t+h,h)$ is a function of the prediction interval h.

Then mean absolute error (MAE) is calculated as prediction performance with the prediction interval h as follows.

$$\mathrm{MAE}(h) = \frac{1}{N} \sum_{n=1}^{N} e(n,h) \qquad (15)$$

where $N = t_{\mathrm{end}} - t_{\mathrm{start}}$ =2500[steps] is time interval for this evaluation.

The evaluation results of the MAE and standard deviation (SD) of the error at h=15[steps] (0.5[sec]) are shown in Table 1. The proposed method demonstrated the least MAE and the least SD in comparison another methods. There is a difference of about 0.1[mm] against other methods on MAE. Fig.6 shows MAE as functions of the prediction interval h. As is clear from this figure, the prediction error of the proposed method is smaller than these of the other two SARIMA methods at any intervals $h \leq 30$.

Table 1. The prediction performances of each prediction methods at 0.5[sec](h=15[step]) ahead.

Prediction method	MAE±SD[mm]
Conventional SARIMA[2]	0.9603±0.8775
Modified SARIMA[2]	0.9408±0.8239
Proposed SARIMA	**0.8204±0.6693**

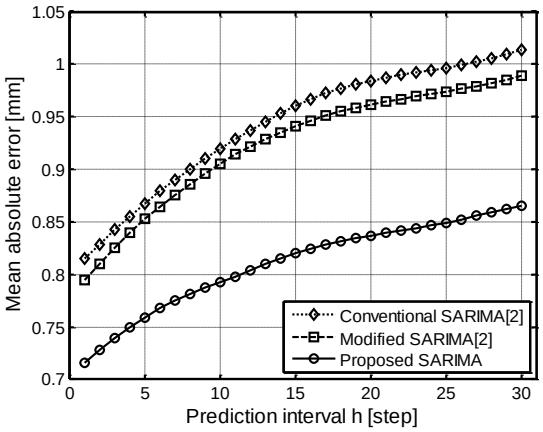

Fig.6. Comparison among three methods of mean absolute errors (MAE) as functions of the prediction interval h.

V. CONCLUSION

In this paper, we have developed a prediction method of time series for lung tumor motion. The proposed method is composed of the period estimator and the time variant SARIMA model. Simulation results showed the proposed method can achieve highly accurate prediction and has superiority against conventional methods in short- or mid-term prediction.

REFERENCES

[1] H. Ohnishi, T. Araki, T. Shirato, *et al*. (2004), Stereotactic hypofractionated high-doseirradiation for stage I non small cell lung carcinoma. Cancer 101(7):1623-1631

[2] N. Homma, M. Sakai, H. Endo, *et al*. (2009), A New Motion Management Method for Lung Tumor Tracking Radiation Therapy. IEEE WSEAS Transactions on Systems 8(4):471-480

[3] M. Sakai, N. Homma and Y. Takai. (2008), Respiratory motion prediction model of lung tumor for tracking radiation therapy. 51st the Japan Joint Automatic Control Conference 51.96-99 (in Japanese)

[4] Hokkaido University Hospital, Sapporo, Japan. http://www.huhp.hokudai.ac.jp/

[5] P.J. Brockwell and R.A. Davis (2002), Introduction to Time Series and Forecasting, 2nd edition, Springer Texts in Statistics, Springer