Forecasting of Water Level Using Seasonal Movement Approach

Khin Mar Cho, Swe Zin Aung
University of Computer Studies (Monywa)
khinmarcho2009@gmail.com

Abstract

This system estimates water level in Chindwinn River with seasonal movement in time series. This system uses seasonal movement approach. Time-Series is a set of observations generated sequentially in time. This system uses Seasonal movements to calculate monthly data. This system use moving average to get Seasonal index. This system uses trend analysis to forecast water level. Water level is forecasted by using least square method's linear and Quadratic equation. This system gets minimum error by using mean square error method. It is being implemented using c#.

1. Introduction

A time series is a series of values of a variable at successive points in time or for successive intervals of time. The reason for analyzing a time series is that the past behavior of a time may tell the investigator something about the future and hence forecasts can be made.

Time series is a sequence of observations which are based on time factor. The value assumed by the variable at time t may or may not embody an element of random variation. Events such as: accidents, arrivals of cars at traffic lights, out breaks of epidemics, etc. may happen form time to time in a certain area and thus over a period constitute a series of events.

The use at time t of available observations from a time series to forecast it value at some future time t+1 (1 is lead time) can provide abase for

(a) economic and business planning
(b) production planning
(c) inventory and production control
(d) control and optimization of industrial process.

The study of process dynamics is made

(a) To achieve better control of existing plants and
(b) To improve the design of new plants.

In particular several methods have been proposed for estimating the transfer function of plant units from process records consisting of an input time series x_t and an output time series y_t.

Time series concerns a complex moving through time and observed at specified interval. In physical science and social sciences time series analysis has to be carried out in order to understand the underlying phenomena, with respect to time and then to forecast a head of the two approaches. Frequency domain approach has been mostly used in physical science whereas time domain approach plays a prominent role in social science especially in economic.

Sometimes the trend and cyclical components are considered together and then as given in Kendall (1976) and Kendall, Stuart and Ord (1983), the two models are,

$$Y_t = m_t + s_t + e_t \qquad (1.1)$$

Which is the additive model.

$$Y_t = m_t s_t e_t \qquad (1.2)$$

Which is the multiplicative model. They also give

$$Y t = m_t s_{1t} + s_{2t} + e_t \qquad (1.3)$$

Which is the mixed model. In these models, y_t is the time series,
m_t is the smooth component (trend and short term oscillation) of the time series, s_t, s_1t and s_2t are the seasonal components and et is the error term. The mixed model is of additive multiplicative type in which the components add and multiply.

Many time series have an important seasonal component and for such series it is important to be able to separate out the seasonal and trend components so that the overall "deseasonalized" trend can be discerned and the magnitude and pattern of seasonal variations better understood. It is important to know whether a falling value or rising value of a time series indicates the real change in the long term movement or whether it is just a temporary change due to seasons.

In other words the seasonal effects in a time series need to be examined. There are different reasons for wanting to examine seasonal effect. Some of the reasons given by Kendall (1976) are

(a) To compare the variable under study at different points of the year. In order to know the intra-year variation and to take action correspondingly.
(b) To remove seasonal effects from the series in order to study its other constituents uncontaminated by the seasonal element and

(c) To adjust the time series for seasonal effects.

2. Related work

A time series is a series of values of a variable at successive points in time or for successive intervals of time. The reason for analyzing a time series is that the past behavior of a time series may tell the investigator something about the future and hence forecasts can be made. [1] Characteristic movements of time series may be classified into four main types often called components of a time series. [1] Seasonal variation of a time series shows the within year variation of the series that repeat in a similar fashion year after year. It can only be seen in daily, weekly, monthly or quarterly data and not in yearly data. [6]. Seasonality is one of the most pervasive phenomena of economic life. Seasonality means a tendency to repeat a pattern of behavior over a seasonal period, generally one year. [2]

3. Theorybackground

A time series is a series of values of a variable at successive points in time or for successive intervals of time. In the classical approach to time series, the analysis usually begins by decomposing the time series into four components: trend, cycle, seasonal and irregular components. the two basic models usually used for analyzing a time series are

$$Y_t = T_t + C_t + S_t + R_t \quad (2.1)$$

Which is called the additive model and

$$Y_t = T_t C_t S_t R_t \quad (2.2)$$

This is called the multiplicative model.

In these models, Y_t is the time series, T_t is the trend component, C_t is the cyclical component, St is the seasonal component and R_t is the random component or the irregular component.

3.1 Classification of time series Movements

Characteristic movements of time series may be classified into four main types often called components of a time series:

1. Long-term or secular movements
2. Cyclical movements or cyclical variations
3. Seasonal movements or seasonal variations
4. Irregular or random movements

3.1.1 Seasonal movements of seasonal Variations

It refers to the identical, or almost identical, patterns which a time series appears to follow during corresponding months of successive years. Such movements are due to recurring events which take place annually, as for instance the sudden increase of department store sales before Christmas. Although seasonal movements in general refer to annual periodicity in business or economic theory, the ideas involved can be extended to include periodicity in business or economic theory, the ideas involved periodicity over any interval of time such as daily, hourly, weekly, etc., depending on the type of data available.

The seasonal variation represents a type of periodic movement, where the period is not longer than one year. A lot of business activity in influenced by changing seasons of the year. Seasonal fluctuations are patterns that reoccur regularly during the year. Since seasonal fluctuations occur within the year, annual data will not capture or reflect seasonal changes, quarterly, monthly or weekly data.

3.2 Types of seasonal variation

Seasonal variation of a time series shows the within year variation of the series that repeat in a similar fashion year after year. It can only be seen in daily, weekly, monthly or quarterly data and not in yearly data. Types of seasonal variation are

1. Constant seasonal variation
2. Gradually changing seasonal variation,
3. Osscillating seasonal variation and
4. Abruptly changing seasonal.

3.3 Seasonality

Seasonality is one of the most pervasive phenomena of economic life. Seasonality means a tendency to repeat a pattern of behavior over a seasonal period, generally one year. Large business concerns need indices of seasonal variation for purpose of analysis and control.

Many economic time series are available both seasonally unadjusted and adjusted. The adjusted series permits one to examine variations unimpaired by seasonal movements. Many of the reported statistical series on economic activity are published in seasonally adjusted form. Seasonal adjustment refers to any smoothing procedure applied to raw data.

The objective is to obtain a good statistical description, of the seasonal variation,

1. For use in planning, production scheduling, and so on,
2. To aid in the calculation of statistical normal.
3. To use in elimination of the seasonal variation from time series data so that the trend and cycle effects may be seen uncluttered by the seasonal influence.

3.4 Methods for finding seasonal indexes

There exist different methods for measuring the seasonal variation of estimating seasonals and the assumed models of the time series. These methods are

1. Average percentage method
2. The percentage trend or ratio to trend method
3. Ratio to moving average method
4. Lind relatives method

3.4.1 Ration to moving average method

In this method a 12 month moving average is computed. Results obtained between successive months instead of in the middle of the month as for the original data. This method computes a 2 month moving average of this 12 month moving average. The result is called a centered 12 month moving average.

3.5 Deasonalization (trend) of data

If the original monthly data are divided by the corresponding seasonal index numbers, the resulting data are said to be deseasonalized or adjusted for seasonal variation. Such data will include trend, cyclical and irregular movements.

3.6Estimation of trend

Estimation of trend can be achieved in several possible ways. There are

1. The method of least square
2. The freehand method
3. The moving average method
4. The method of semi-averages

3.7The least square method

The least squares parameter estimation method is a variation of the probability plotting methodology in which one mathematically fits a straight line to a set of points in an attempt to estimate the parameters. The method of least square requires that a straight line be fitted to a set of data points such that the sum of the squares of the vertical deviations from the points to the line is minimized, if the regression is on Y, or the line be fitted to a set of data point such that the sum of the squares of the horizontal deviations from the points to the line is minimized, if the regression is on X.

The linear equation is

$Y_t = b_0 + b_1t + e_t$

$$\hat{b}_1 = \frac{n\sum ty_t - \left(\sum t\right)\left(\sum y_t\right)}{n\sum t^2 - \left(\sum t\right)^2}$$

$$\hat{b}_0 = \frac{1}{n}\left(\sum y_t - \hat{b}_1\sum t\right)$$

The quadratic equation is

$$\sum Y_t = nb_0 + b_2\sum t^2$$

$$\sum t^2 Y_t = b_0\sum t^2 + b_2\sum t^4$$

Y_t = the observed value at time t

b_0, b_1= unknown constant

e_t= error term

3.8 Mean square error

$$\text{MSE} = \frac{\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2}{n}$$

Y_t = actual value

$\hat{Y}_t$ = forecast value

3.9 System flow diagram

This system flow diagram, the user stores the water level data for 8 years. And then, he/she extracts the data to forecast the water level. First, this system calculates trend analysis by using these data. Finally, the result is executed by using two methods: Linear and Quadratic equations. This system estimate error using minimum mean square error. This system shows error results. In this prediction, Linear equation has less error than Quadratic equation.

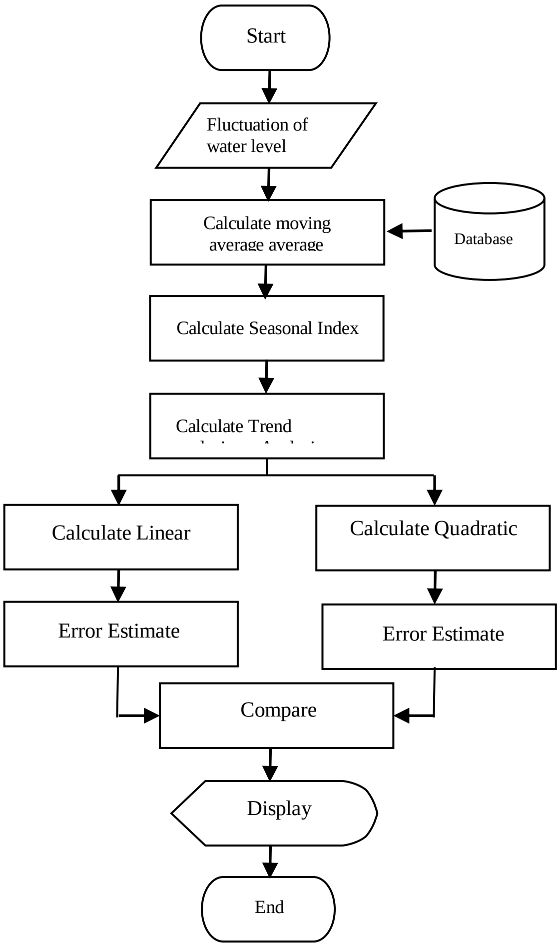
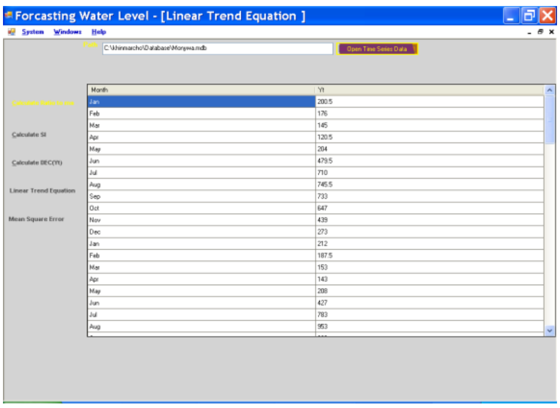


Figure 1 System flow diagram

5. Result

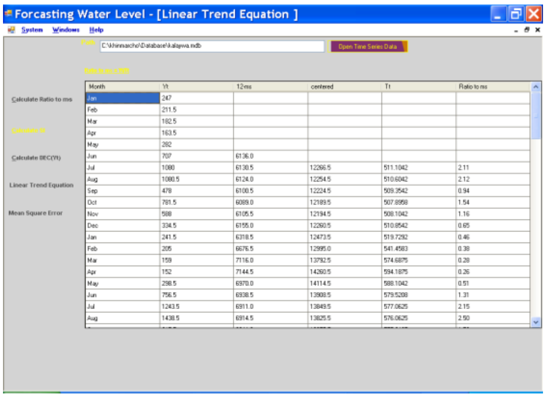
Table 1 is forecasted actual data with linear equation and quadratic equation.

Table 1 original table



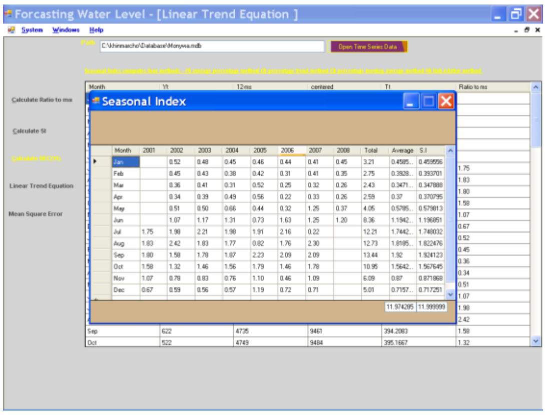
Month	Yt
Jan	200.5
Feb	176
Mar	145
Apr	120.5
May	204
Jun	479.5
Jul	710
Aug	745.5
Sep	733
Oct	647
Nov	439
Dec	273
Jan	212
Feb	187.5
Mar	153
Apr	143
May	208
Jun	427
Jul	783
Aug	953

Table 2 ratio to moving average table



Month	Yt	12-ms	centered	Tt	Ratio to ms
Jan	247				
Feb	211.5				
Mar	182.5				
Apr	163.5				
May	282				
Jun	707	6136.0			
Jul	1080	6130.5	12266.5	511.1042	2.11
Aug	1080.5	6124.0	12254.5	510.6042	2.12
Sep	478	6100.5	12224.5	509.3542	0.94
Oct	781.5	6089.0	12189.5	507.8958	1.54
Nov	588	6105.5	12194.5	508.1042	1.16
Dec	334.5	6155.0	12260.5	510.8542	0.65
Jan	241.5	6318.5	12473.5	519.7292	0.46
Feb	205	6676.5	12995.0	541.4583	0.38
Mar	159	7116.0	13792.5	574.6875	0.28
Apr	152	7144.5	14260.5	594.1875	0.26
May	298.5	6970.0	14114.5	588.1042	0.51
Jun	756.5	6938.5	13908.5	579.5208	1.31
Jul	1243.5	6911.0	13849.5	577.0625	2.15
Aug	1438.5	6914.5	13825.5	576.0625	2.50

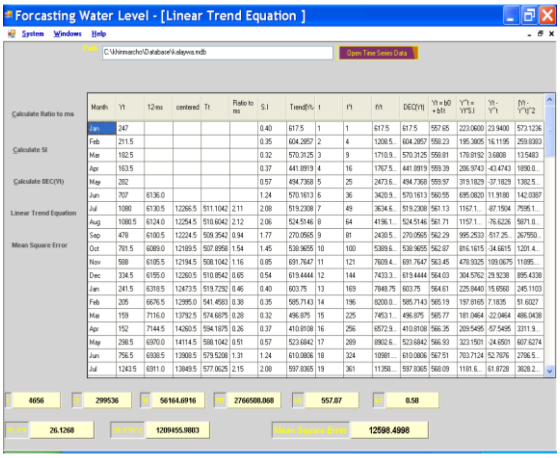
Table 3 seasonal index table



Month	2001	2002	2003	2004	2005	2006	2007	2008	Total	Average	S.I
Jan		0.52	0.48	0.45	0.46	0.44	0.41	0.45	3.21	0.4585...	0.459556
Feb		0.45	0.43	0.38	0.42	0.31	0.41	0.35	2.75	0.3928...	0.393701
Mar		0.36	0.41	0.31	0.52	0.25	0.32	0.26	2.43	0.3471...	0.347888
Apr		0.34	0.39	0.49	0.56	0.22	0.33	0.26	2.59	0.37	0.370795
May		0.51	0.50	0.66	0.44	0.32	1.25	0.37	4.05	0.5785...	0.579813
Jun		1.07	1.17	1.31	0.73	1.63	1.25	1.20	8.36	1.1942...	1.196851
Jul	1.75	1.98	2.21	1.98	1.91	2.16	0.22		12.21	1.7442...	1.748032
Aug	1.83	2.42	1.83	1.77	0.82	1.76	2.30		12.73	1.8185...	1.822476
Sep	1.80	1.58	1.78	1.87	2.23	2.09	2.09		13.44	1.92	1.924123
Oct	1.58	1.32	1.46	1.56	1.79	1.46	1.78		10.95	1.5642...	1.567645
Nov	1.07	0.78	0.83	0.76	1.10	0.46	1.09		6.09	0.87	0.871868
Dec	0.67	0.59	0.56	0.57	1.19	0.72	0.71		5.01	0.7157...	0.717251
										11.974285	11.999999

Month	Yt	12-ms	centered	Tt	Ratio to ms
Sep	622	4735	9461	394.2083	1.58
Oct	522	4749	9484	395.1667	1.32

Table 4 Forecast linear equation

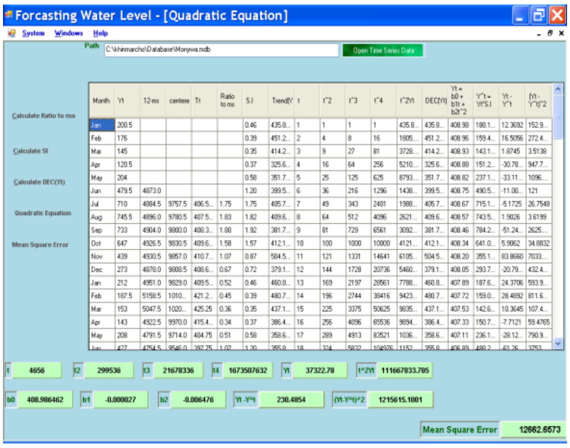


Month	Yt	12-ms	centered	Tt	Ratio to ms	S.I	Trend(Yt/...	t	t*t	f*t	DEC(Yt)	Yt = b0 + b1t	Y^t = Yt*S.I	Yt - Y^t	(Yt - Y^t)^2
Jan	247					0.40	617.5	1	1	617.5	617.5	557.65	223.0600	23.9400	573.1236
Feb	211.5					0.35	604.2857	2	4	1208.5...	604.2857	558.23	195.3805	16.1195	259.8383
Mar	182.5					0.32	570.3125	3	9	1710.9...	570.3125	558.81	178.8192	3.6808	13.5483
Apr	163.5					0.37	441.8919	4	16	1767.5...	441.8919	559.39	206.9743	-43.4743	1890.0...
May	282					0.57	494.7368	5	25	2473.6...	494.7368	559.97	319.1829	-37.1829	1382.5...
Jun	707	6136.0				1.24	570.1613	6	36	3420.9...	570.1613	560.55	695.0820	11.9180	142.0387
Jul	1080	6130.5	12266.5	511.1042	2.11	2.08	519.2308	7	49	3634.6...	519.2308	561.13	1167.1...	-87.1504	7595.1...
Aug	1080.5	6124.0	12254.5	510.6042	2.12	2.06	524.5146	8	64	4196.1...	524.5146	561.71	1157.1...	-76.6226	5871.0...
Sep	478	6100.5	12224.5	509.3542	0.94	1.77	270.0565	9	81	2430.5...	270.0565	562.29	995.2533	-517.25...	267550...
Oct	781.5	6089.0	12189.5	507.8958	1.54	1.45	538.9655	10	100	5389.6...	538.9655	562.87	816.1615	-34.6615	1201.4...
Nov	588	6105.5	12194.5	508.1042	1.16	0.85	691.7647	11	121	7609.4...	691.7647	563.45	478.9325	109.0675	11895...
Dec	334.5	6155.0	12260.5	510.8542	0.65	0.54	619.4444	12	144	7433.3...	619.4444	564.03	304.5762	29.9238	895.4338
Jan	241.5	6318.5	12473.5	519.7292	0.46	0.40	603.75	13	169	7848.75	603.75	564.61	225.8440	15.6560	245.1103
Feb	205	6676.5	12995.0	541.4583	0.38	0.35	585.7143	14	196	8200.0...	585.7143	565.19	197.8165	7.1835	51.6027
Mar	159	7116.0	13792.5	574.6875	0.28	0.32	496.875	15	225	7453.1...	496.875	565.77	181.0464	-22.0464	486.0438
Apr	152	7144.5	14260.5	594.1875	0.26	0.37	410.8108	16	256	6572.9...	410.8108	566.35	209.5495	-57.5495	3311.9...
May	298.5	6970.0	14114.5	588.1042	0.51	0.57	523.6842	17	289	8902.6...	523.6842	566.93	323.1501	-24.6501	607.6274
Jun	756.5	6938.5	13908.5	579.5208	1.31	1.24	610.0806	18	324	10981....	610.0806	567.51	703.7124	52.7876	2786.5...
Jul	1243.5	6911.0	13849.5	577.0625	2.15	2.08	597.8365	19	361	11358....	597.8365	568.09	1181.6...	61.8728	3828.2...

4656 | 299536 | 56164.6916 | 2766508.068 | b0 557.07 | b1 0.58

26.1268 | (Yt-Y^t)^2 1209455.9803 | Mean Square Error 12598.4998

Table 3 Forecast quadratic equation



5. Conclusion

The system forecasts Water level of Chindwinn River with seasonal movement in time series. In this thesis, the system uses monthly data to calculate least square's Linear equation and Quadratic equation. This system calculates minimum error by using mean square error method. In this prediction, Linear equation has less error than Quadratic equation.

10. References

1. Anderson, T.W, " The Statistical Analysis of Time Series", John- Wiley, New York, 1971.
2. Durbin, J," Trend elimination for the purpose of estimating seasonal and periodic components of time series ", John-Wiley, New York, 1963.
3. Hannan, E.J, " The Estimation for Seasonal Variation in Economic time series", Jour-Amer-Stat. Assoc.58.No.301, 31-44, 1963.
4. Kendall, M.G, " Time Series", 2nd ed, Charles Griffin, London, 1976.
5. Nerlove, Marc, " Spectral Analysis of Seasonal Adjustment procedures ", Econometrica,32, No.3, 241-286, 1964.
6. Steiner, P.O, " An Introduction to the Analysis of Time Series", Preliminary ed, Rinehart, New York, 1956.