# Chapter 4

## Statistical Analysis and Modelling

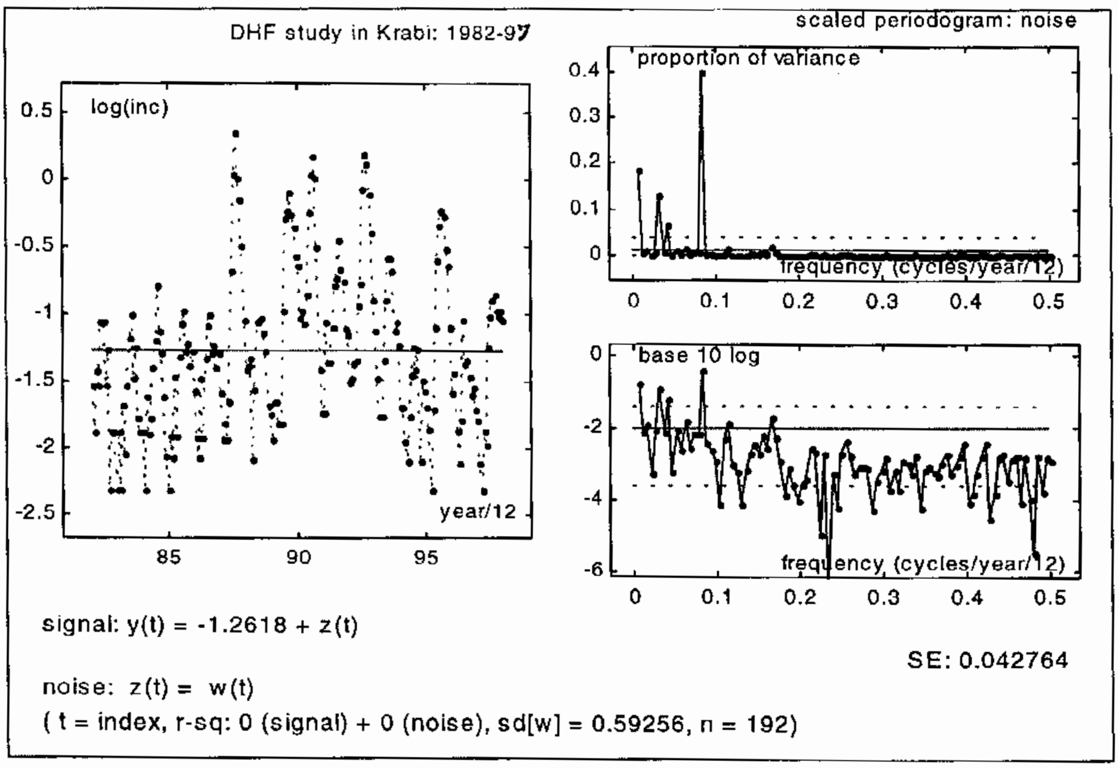
In this chapter, we further investigate the results in chapter 3 using time series analysis and fit models for predicting the data. The results may be classified as follows.

1. Univariate Time Series Analyses for DHF Incidence
2. Univariate Time Series Analyses for Rainfall
3. Time Series Analyses of DHF Incidence Using Regression Analysis with Other Variables as Determinants.

### Univariate Time Series Analyses for DHF Incidence

A preliminary time series analysis of the log-transformed monthly DHF incidence rate in Krabi is depicted in Figure 4.1.

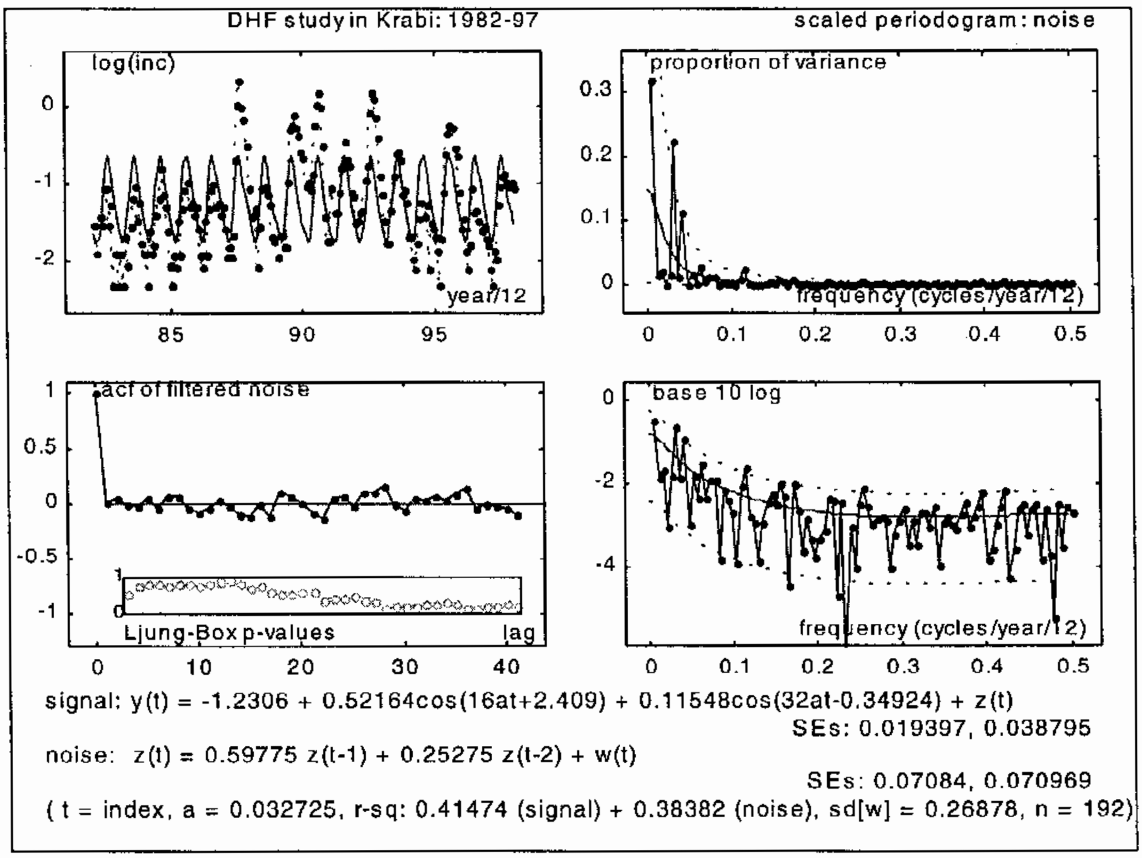
Figure 4.1: Preliminary time series analysis of DHF incidence in Krabi



It is observe that the scaled periodogram contains a spike at the 16th harmonic, which accounts for 40% of the variation in the time series. This frequency corresponds to the seasonal pattern of variation during the year, seen in Figures 3.5 and 3.6. In addition, the logarithm of the periodogram has a decreasing trend, which suggests that the residuals follow a simple autoregressive model.

Figure 4.2 shows the result of fitting a model comprising a sinusoidal signal with period 12 months and modelling the residuals (i.e., the noise) as a second-order autoregression.

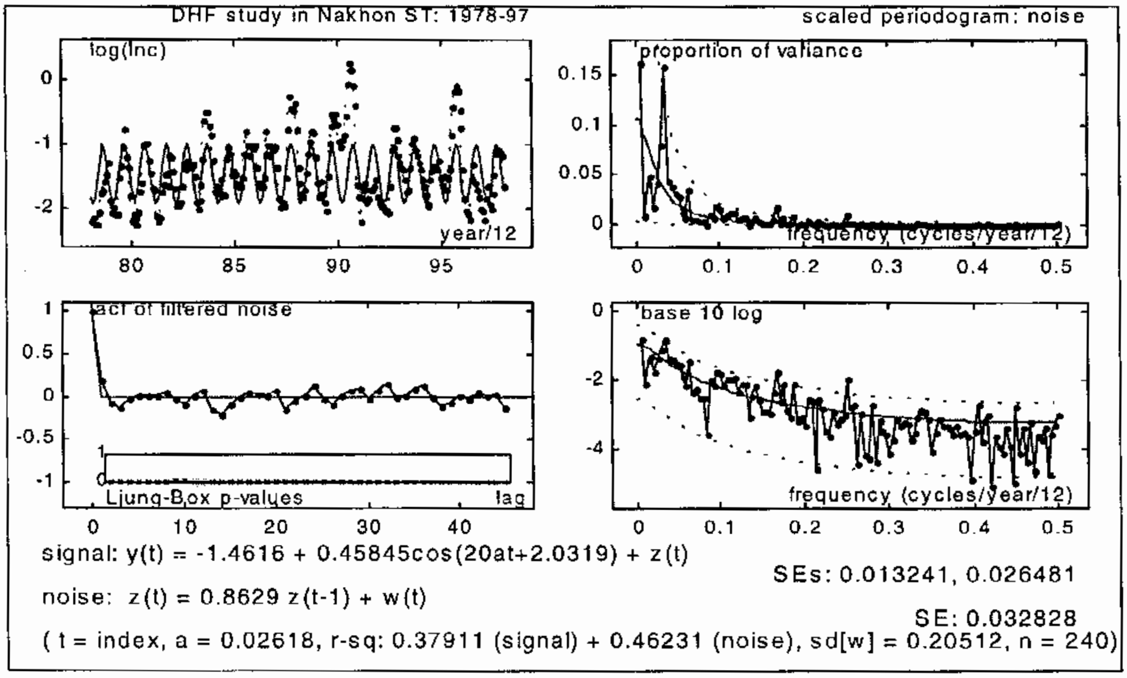
Figure 4.2: Further time series analysis of DHF incidence in Krabi




For this model, most of the periodogram values are inside the 95% confidence band, and the Ljung-Box p-values for testing that the filtered noise is white are all non-significant. So we concludedthat this model provides a satisfactory fit to the monthly DHF incidences in Krabi. The model says that, on a logarithmic scale, the DHF incidence follows a sinusoidal seasonal signal with residuals that follow a simple random walk process with parameter 0.6. The signal accounts for 41.4% of the variation, and the autoregressive parameter accounts for a further 38.4%, giving a total goodness-of-fit equal to 80%.

Figure 4.3 shows the result of fitting the same model to the data for Nakhon Si Thammarat.

Figure 4.3: Initial time series analysis of DHF incidence in Nakhon Si Thammarat




From Figure 4.3, it is clear that the model does not provide a satisfactory fit. Even though the signal explains 37.9% of the variation and the autoregressive parameter accounts for a further 46.2%, most of the Ljung-Box p-values are statistically significant.

Figure 4.4 shows a further time series analysis of the DHF incidence in Nakhon Si Thammarat, in which the number of parameters fitted to the noise increased to two. The result shows a much more satisfactory fit for the residual noise: the Ljung-Box p-values up to lag 20 months are all non-significant. The fitted parameters in the second-order autoregression are $a_1 = 1.07$ and $a_2 = -0.24$. The final model has a total goodness-of-fit of 84.9%, comprising 38.5% for the signal and 46.4% for the noise.

The result of fitting the same model to the monthly DHF incidence series for Trang over the 20-year period is shown in Figure 4.5.

Figure 4.4: Further time series analysis of DHF incidence in Nakhon Si Thammarat

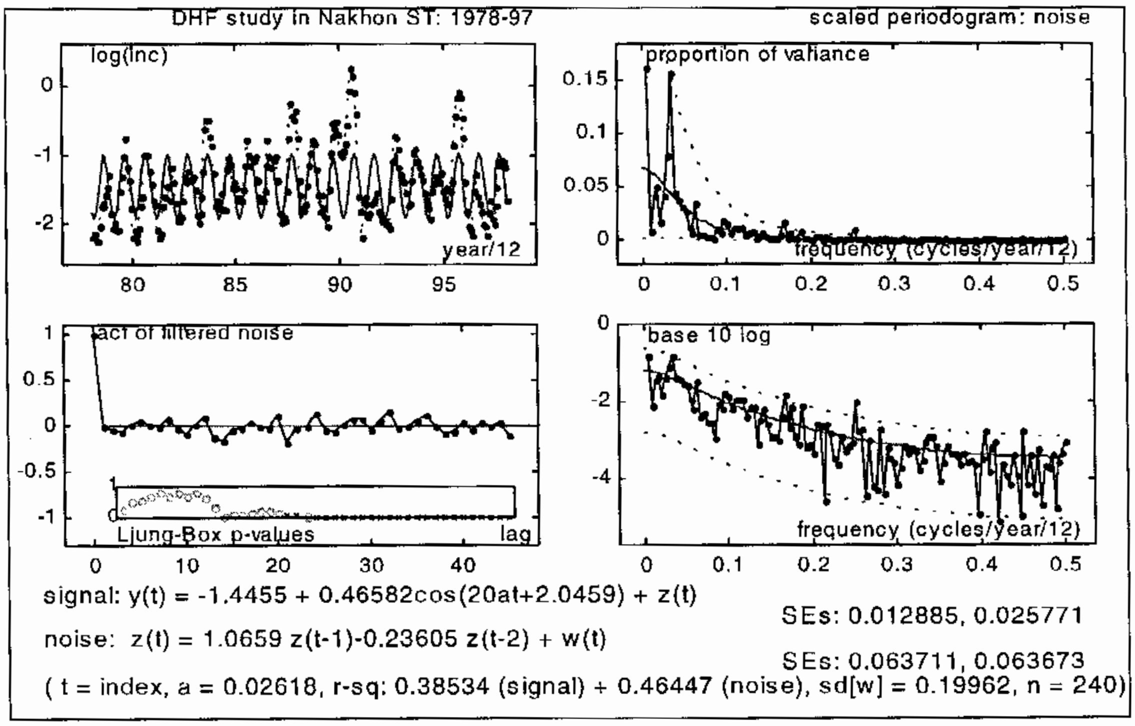
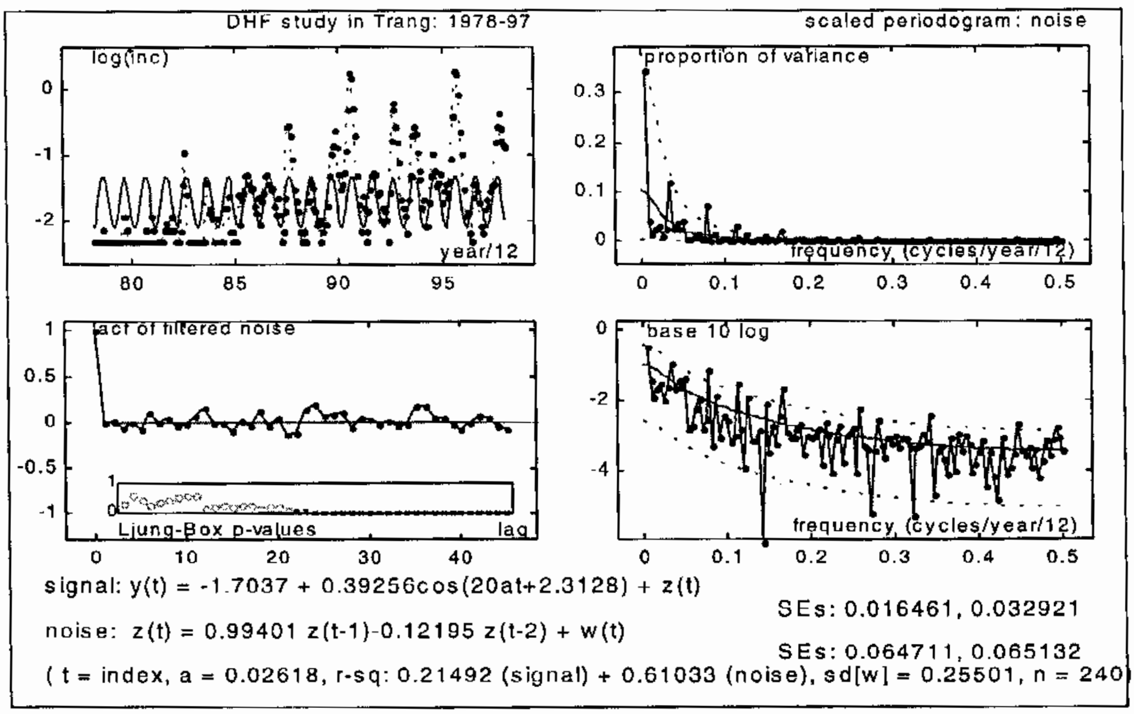
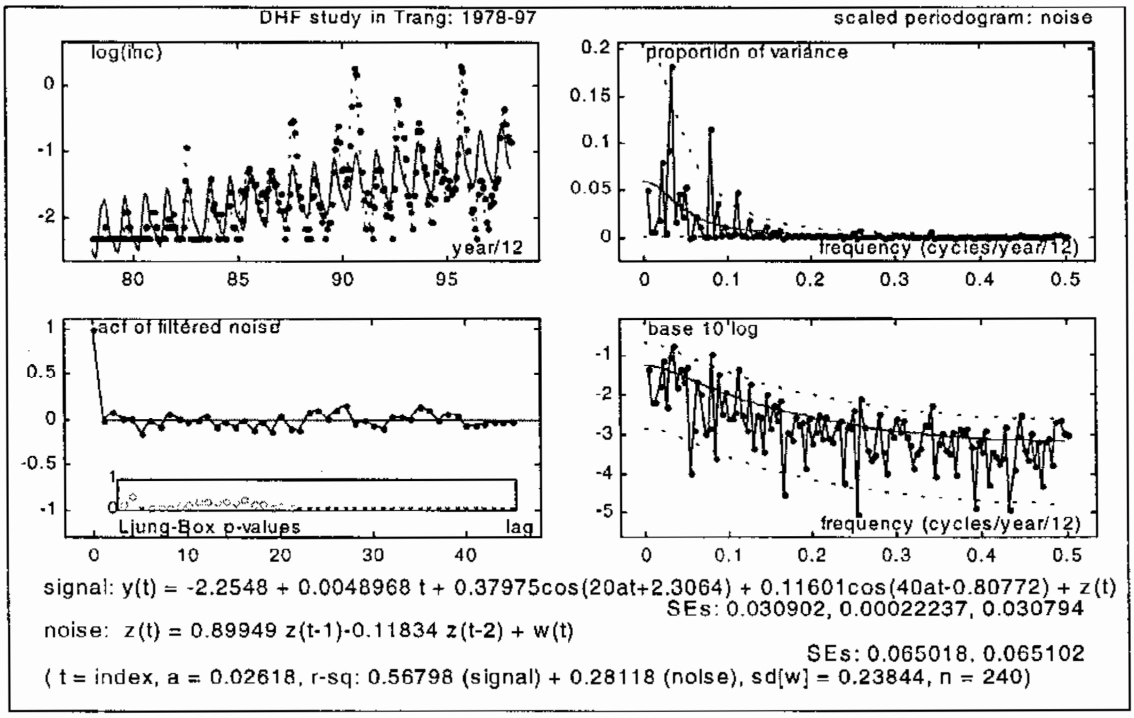



Figure 4.5: Initial time series analysis of DHF incidence in Trang



It is clear from Figure 4.5 that there is an increasing trend in the DHF incidence at Trang. Figure 4.6 shows the result after fitting a signal containing both a linear trend and the seasonal periodic harmonic components.

Figure 4.6: Further time series analysis of DHF incidence in Trang

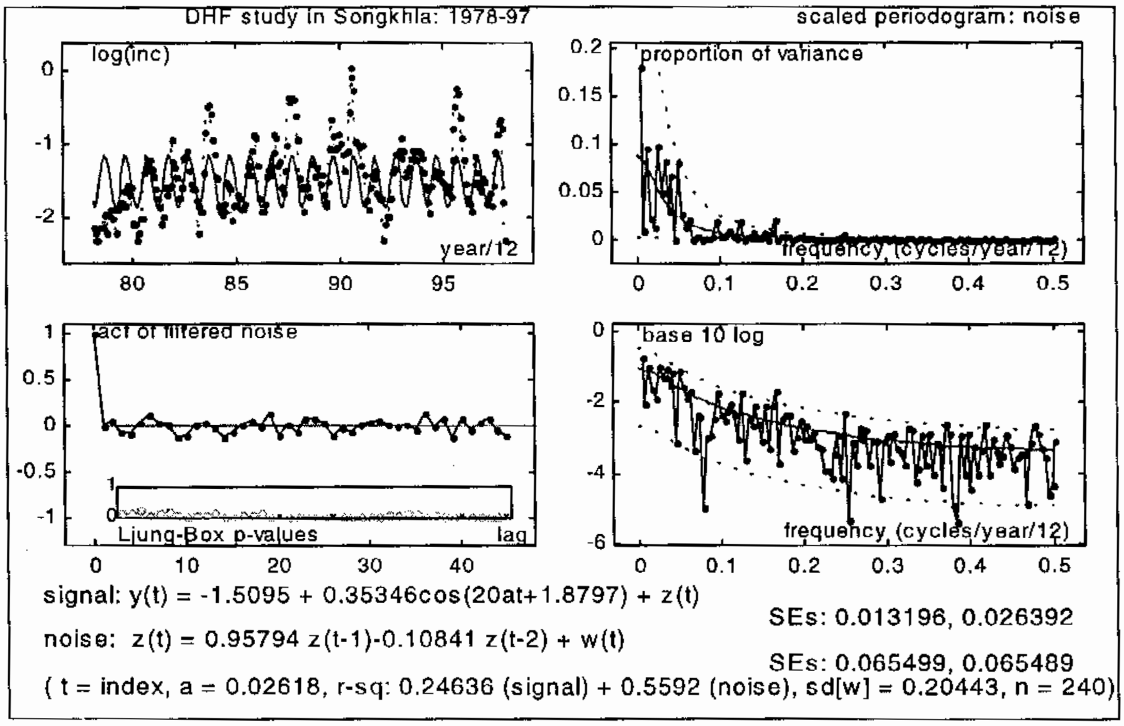



The trend has a slope of 0.0048 per month, so the model says that the DHF incidence at Trang increased by just over 5% per annum over the 20-year period. In this model, the signal explains 56.8% of the variation, and the noise accounts for a further 28.12%, giving a total goodness-of-fit of 84.92%.

Finally, Figure 4.7 shows the time series analysis of the DHF incidence in Songkla province, using the same model as developed for Nakhon Si Thammarat.

For this model, it is clear that the model provides a satisfactory fit. The signal explains 24.6% of the variation and the autoregressive parameter accounts for 55.9%, giving a total goodness-of-fit equal to 80.5%.

Figure 4.7:Time series analysis of DHF incidence in Songkhla



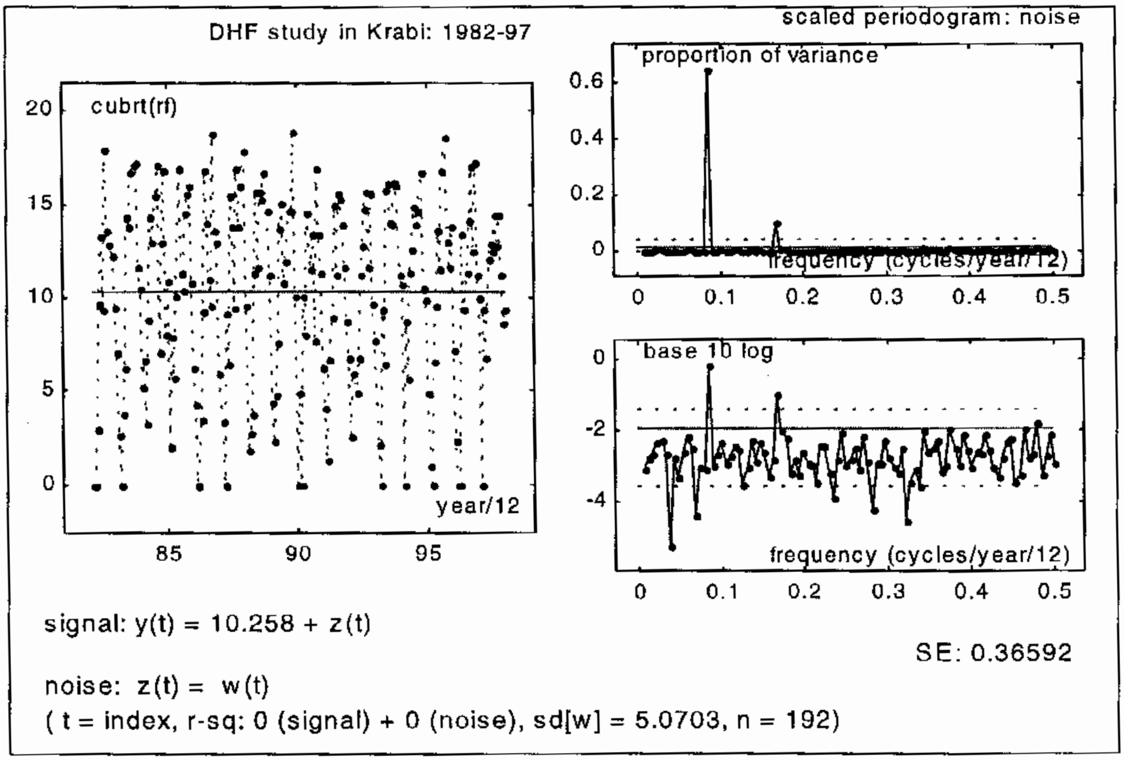

## Univariate Time Series Analyses for Rainfall

In this section we apply the methods used in the preceding section to the monthly rainfall series.

Figure 4.8 shows a preliminary time series analysis of the cube root transformed monthly rainfall in Krabi. It was found that the scaled periodogram contains a large spike at the 16$^{th}$ harmonic and a smaller one at the 32$^{nd}$ harmonic, which account for approximately 65% and 10%, respectvely, of the variation in the time series. These frequencies correspond to the seasonal pattern of variation during the year, seen in Figures 3.9 and 3.10. We also observe that the logarithm of the periodogram is flat, (apart from the two spikes) which suggests that the residuals follow a white noise process.

The result of fitting a model comprising a periodic signal having the two components is depicted in Figure 4.9. The presence of the second harmonic indicates that the seasonal pattern is slightly more complex than just a simple sine wave. This can also be seen from the left panel of Figure 3.10.

Figure 4.8: Preliminary time series analysis of rainfall in Krabi




For this model, most of the periodogram values are inside the 95% confidence band. But the Ljung-Box p-values for testing the whiteness of the filtered noise are statistically significant at lags 1 and 2 months, so this model does not provide a statistically satisfactory fit to the monthly rainfall in Krabi. The situation is not improved by fitting a higher order autoregressive model. The estimated autoregressive parameter $a_1 = -0.152$ is barely statistically significant ($z = -0.152/0.0725 = -2.10$, p-value = 0.036). The signal accounts for a high proportion (75%) of the variation, but the autoregressive model explains less than 1%. This means that we can predict the rainfall at Krabi quite well in the long term, but not in the short term. Since the autoregressive parameter is negative, it means that after allowing for the seasonal pattern there is slightly negative correlation between the rainfall in successive months. The explanation of this could be that a delay is needed after a storm before another one occurs.

Figure 4.9: Further time series analysis of rainfall in Krabi

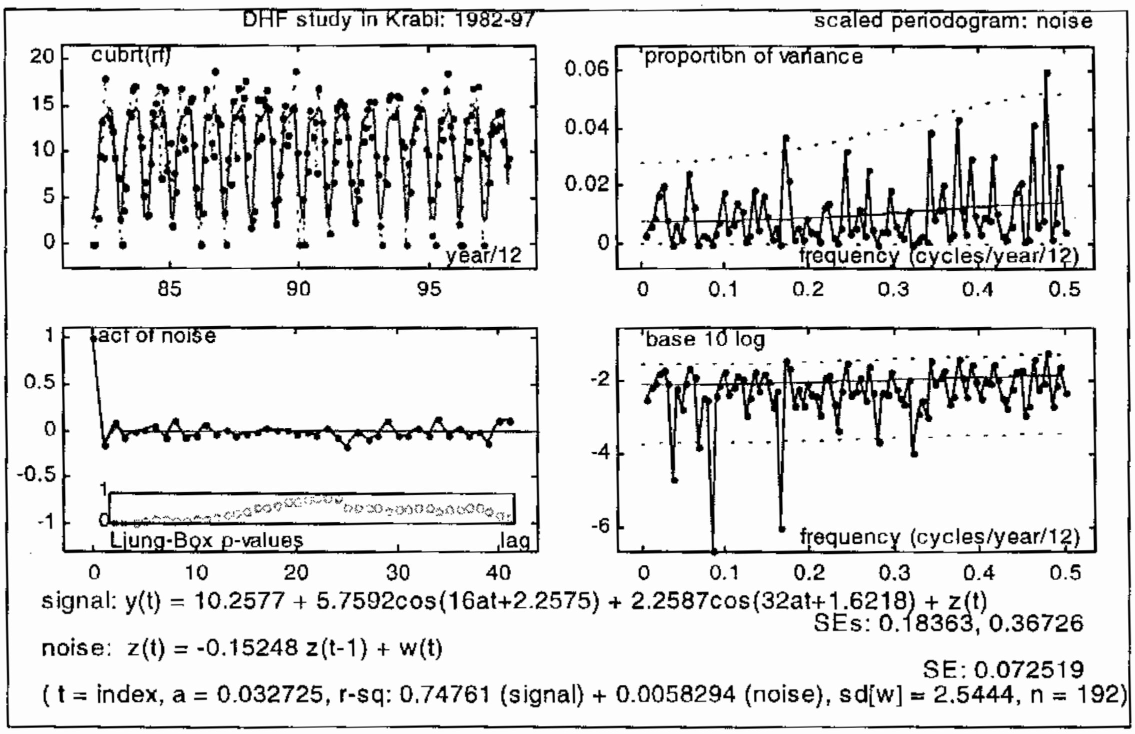
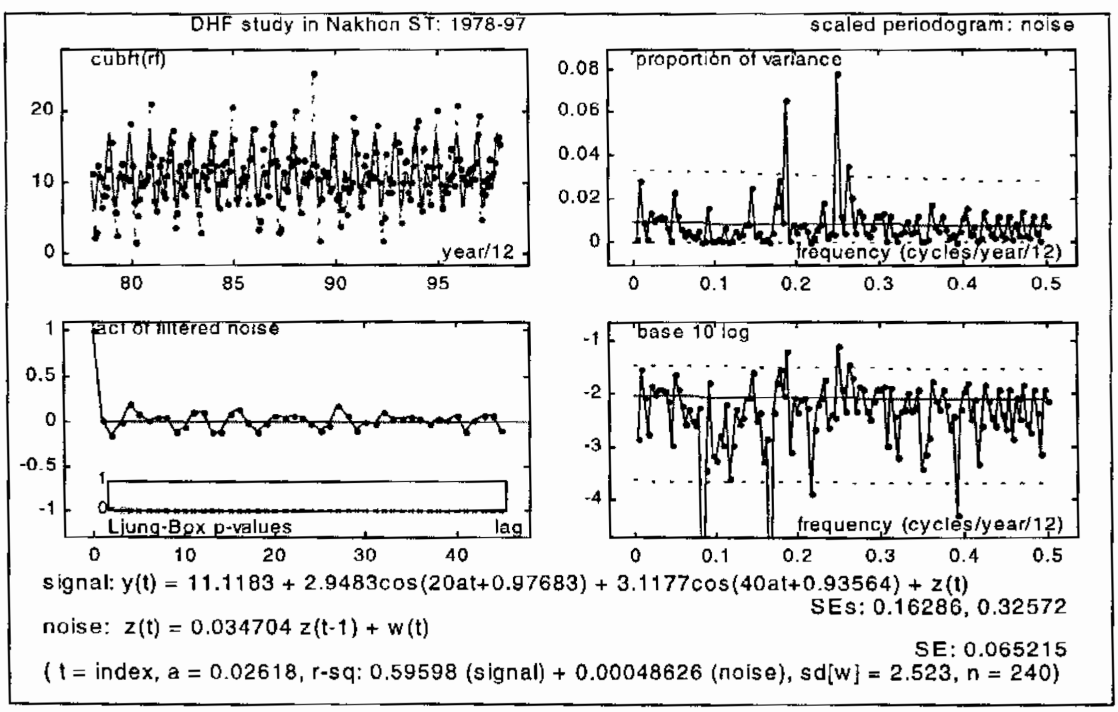



Figure 4.10 shows the result of fitting the same model to the data for Nakhon Si Thammarat.

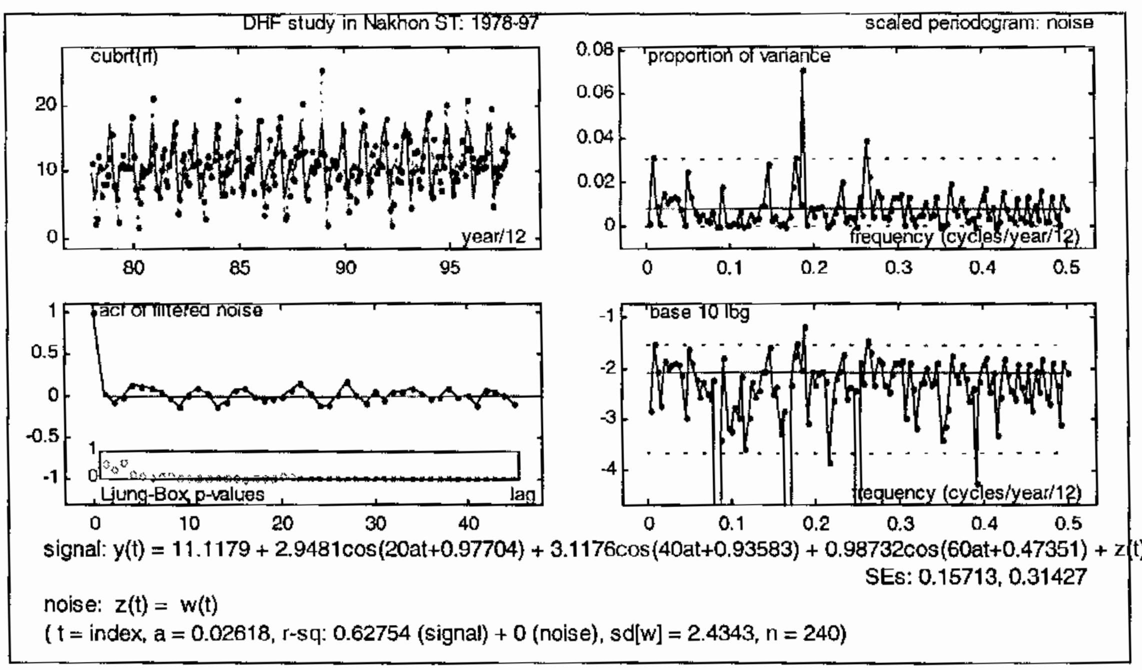
Figure 4.10: Initial time series analysis of rainfall in Nakhon Si Thammarat



From Figure 4.10, it is clear that the model does not provide a satisfactory fit. Even though the signal explains 60% of the variation, most of the Ljung-Box p-values are statistically significant, and the autoregressive parameter is not statistically different from 0. After the annual and biannual harmonic components have been removed, the scaled periodogram of the residuals has two more substantial spikes, at frequencies corresponding to 45 and 60 cycles in the 20-year period. It is difficult to explain the first of these periodic components, but the second one corresponds to a triannual cycle (that is, three times a year).

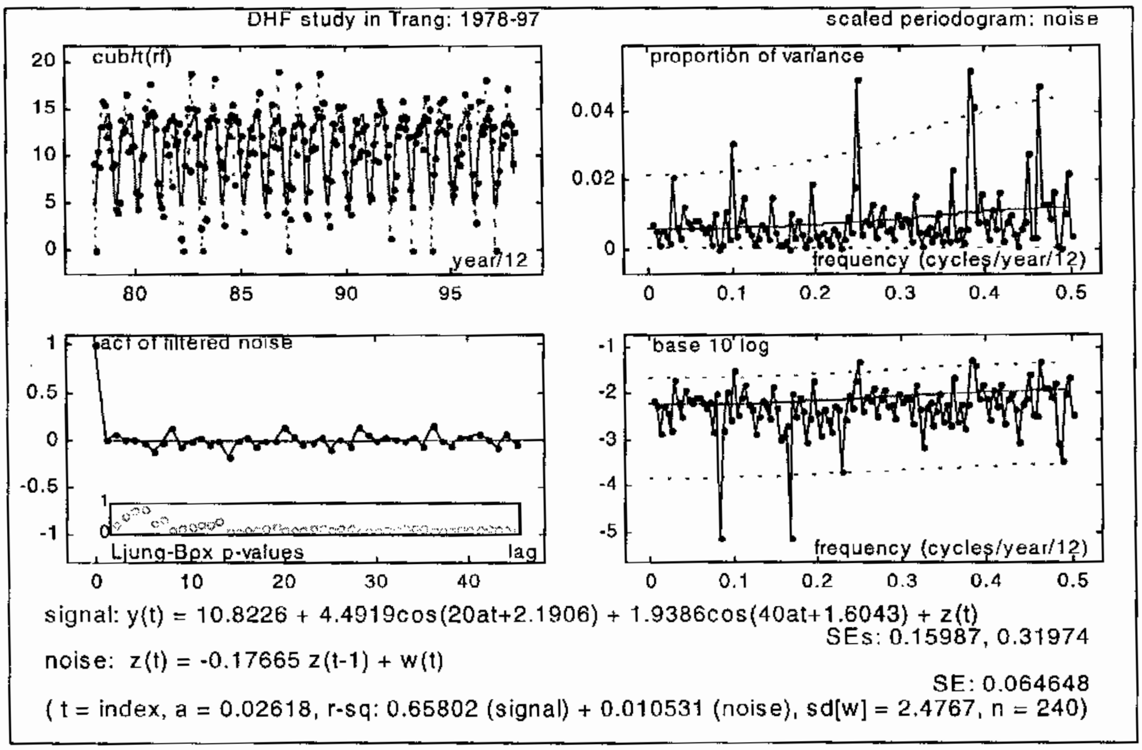
Figure 4.11 shows a further time series analysis of the rainfall in Nakhon Si Thammarat, in which the redundant autoregressive parameter is omitted and the fitted signal comprises a combination of the annual seasonal components and its two harmonics. The result shows a satisfactory fit for the residual noise: the Ljung-Box p-values at most lags up to 24 months are non-significant. The final model has a goodness-of-fit of 62.8%.

Figure 4.11: Further time series analysis of rainfall in Nakhon Si Thammarat



From Figure 3.10, we know that the seasonal patterns of rainfall in Krabi and Trang are very similar. Figure 4.12 shows the result of fitting the same model to the monthly rainfall series for Trang as we did for Krabi.

Figure 4.12: Initial time series analysis of rainfall in Trang

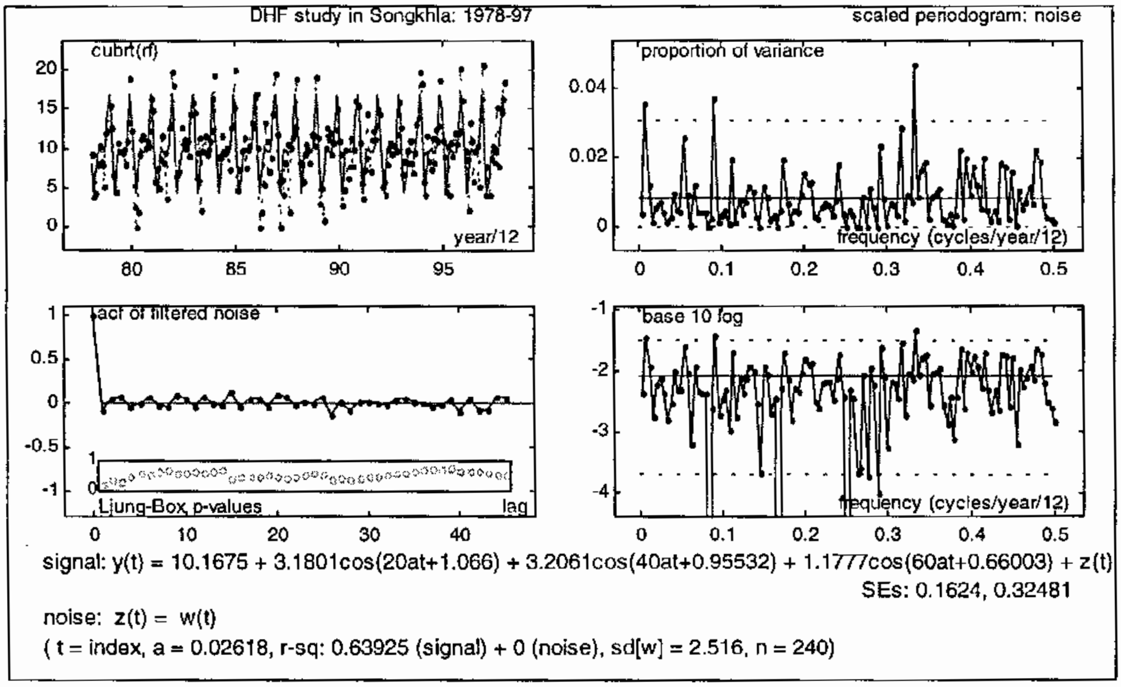



It is clear from Figure 4.12 that the fitted model is satisfactory. The signal explains 65.8% of the variation, but the model for the noise accounts for only 1%. As for the Krabi ranfall, the autoregressive parameter is negative.

From Figure 3.10, we also know that the seasonal patterns of rainfall in Nakhon Si Thammarat and Songkhla are very similar. Figure 4.13 shows the result of fitting the same model to the monthly rainfall series for Songkhla as we did for Nakhon Si Thammarat.

It is clear from Figure 4.13 that the fitted model is satisfactory. The signal explains 63.9% of the variation. Note that the mysterious spike in the periodogram at frequency 45 cycles per 20 years we saw for Nakhon Si Thammarat did not occur for Songkhla.

Figure 4.13: Time series analysis of rainfall in Songkhla




**Time Series Analyses of DHF Incidence Using Regression Analysis**

In this section we apply the time series regression methods described in Chapter 2 to the monthly DHF incidence and other variables including rainfall, rain days, maximum temperature and minimum temperature and humidity.

Figure 4.14 shows the model for DHF incidence in Krabi and the most appropriate predictor variable, which is maximum temperature (labeled $x_1$). This model is obtained by starting with the univariate model obtained in Section 1 of this chapter, and adding maximum temperature as a predictor variable.

From the model of DHF incidence in Figure 4.2, when applying each variable in the model we found that the model provides the most satisfactory fit when maximum temperature is included with the regression coefficient of 0.0055 (p-value 0.05). However, rainfall, rain days, minimum temperature and humidity are not significant, having regression coefficients of –0.0067, 0.0033, -0.0003 and 0.0073, respectively. The annual seasonal effect has two harmonics, the 16th harmonic (having period equal to 12 months) and the 32nd harmonic, with period equal to 6 months. Comparing Figure 4.2 and this figure, it was found that the annual seasonal values at the 16th and 32nd harmonics increased from 0.52 to 0.58 and from 0.12 to 0.14, respectively. Most of the Ljung-Box p-value are non-significant. The signal explains 44% of the variation, and the autoregressive parameter accounts for 36%, giving a total goodness-of-fit equal to 80%.

Figure 4.14: Time series analysis of DHF incidence with max temp in Krabi

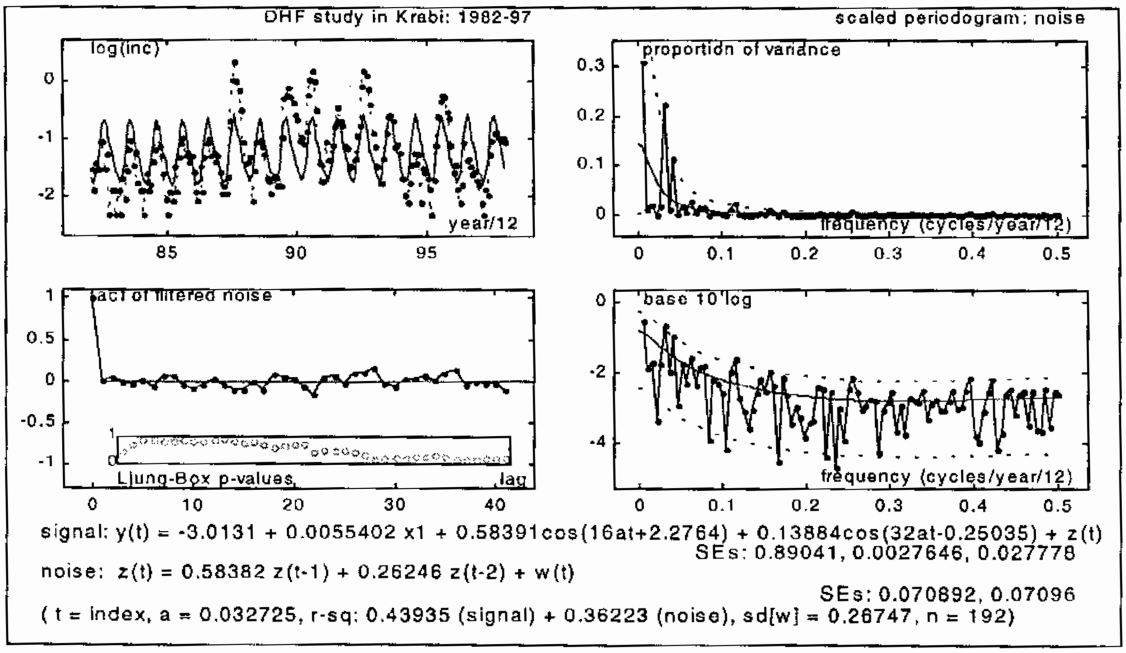
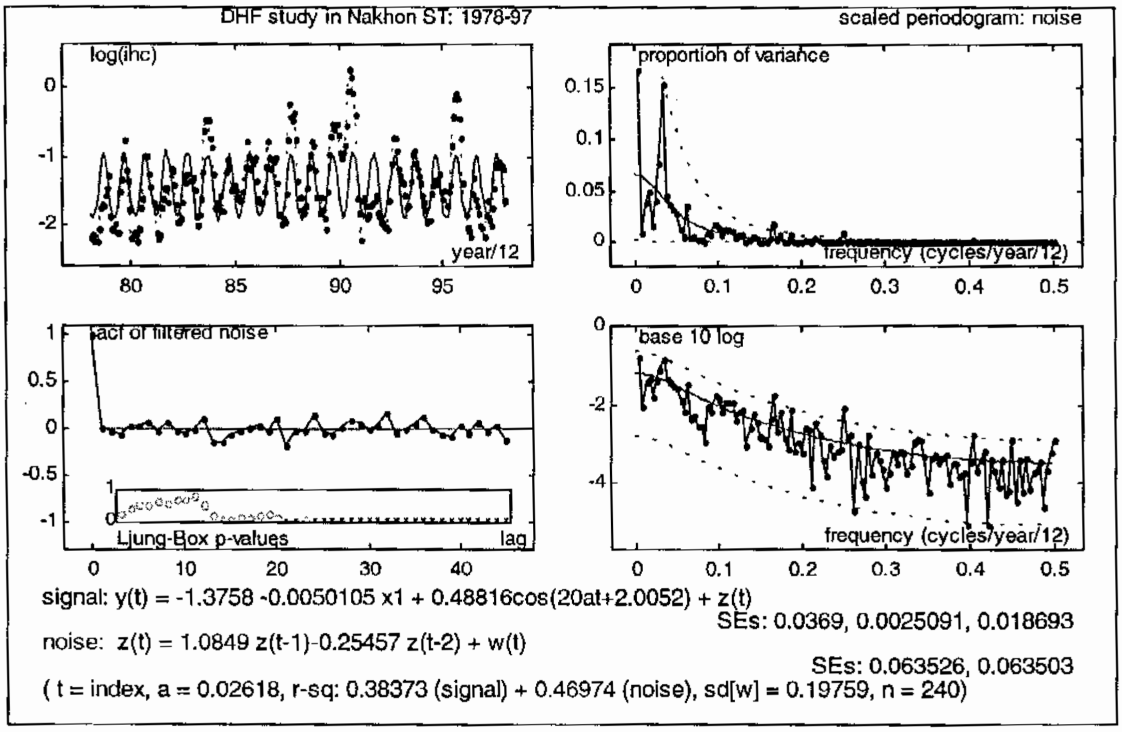



Figure 4.15 shows the model for DHF incidence in Nakhon Si Thammarat and the most appropriate predictor variable, which is rain days (labeled $x_1$).

From the model of DHF incidence in Figure 4.4, when applying each variable in the model, we found that the number of rain days is most appropriate predictor, having regression coefficient of -0.005 (p-value 0.10). The rainfall, maximum temperature, minimum temperature and humidity are not significant, having regression coefficients of – 0.0058, 0.0018, – 0.00163 and – 0.005, respectively. For rain days, the annual seasonal effect has one harmonic at the 20th harmonic (having period equal to 12 months). Comparing Figure 4.4 and this figure, it was found that the annual seasonal value at the 20th harmonic increases from 0.46582 to 0.0.48816. Most of the Ljung-Box p-values are non-significant. The signal explains 38% of the variation, and the autoregressive parameters account for 46.9%, giving a total goodness-of-fit equal to 84.9%.

Figure 4.15: Time series analysis of DHF incidence with rain days in NST

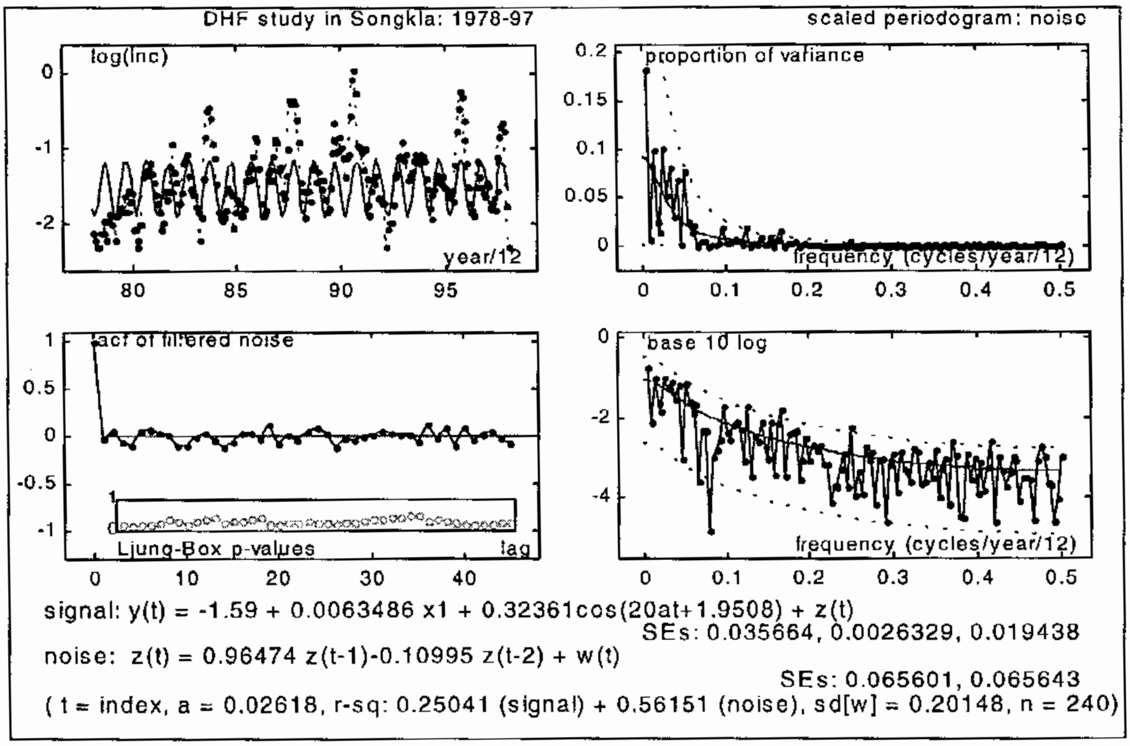



For Trang province, when applying each variable in the model we found that all of the other variables, rainfall, rain days, maximum and minimum temperature and humidity are not significant with the regression coefficients of –0.0072, -0.0052, 0.0001, -0.0024 and –0.0086, respectively (p-values > 0.05).

Figure 4.16 shows the model for DHF incidence in Songkhla with the most appropriate predictor variable, which is rain days (labeled $x_1$).

From the model of DHF incidence in Figure 4.7, when applying each variable in the model, we found that rain days is most appropriate, having regression coefficient of 0.006 (p-value <0.10). The rainfall, maximum temperature, minimum temperature and humidity are not significant, having regression coefficients of 0.004, 0.0034, - 0.0013 and - 0.01, respectively. For rain days, the annual seasonal effect has one harmonic, the 20th harmonic (having period equal to 12 months). Comparing Figure 4.7 and this figure, it was found that the annual seasonal value at the 20th harmonic decreased from 0.35 to 0.32. Most of the Ljung-Box p-values are non-significant. The signal explains 25% of the variation, and the autoregressive parameters for 56.2%, giving a total goodness-of-fit equal to 81.2%.

Figure 4.16: Time series analysis of DHF incidence with rain days in Songkhla




## Summary of Results

Table 4.1 presents the models for predicting DHF incidence with variables in each province (the result in section 1 of this chapter).

Table 4.1: The full model of time series analysis of DHF incidence

| Province | Trend | Signal | Noise | Rain days | Max temp | Total $r^2$ |
|---|---|---|---|---|---|---|
| Krabi | - | (0.58,0.14) | (0.58,0.26) | - | 0.0055 | 0.44+0.36 |
| Trang | 0.0049 | (0.38,0.12) | (0.89,-0.12) | - | - | 0.58+0.30 |
| NST | - | (0.5,0) | (1.08,-0.25) | -0.005 | - | 0.38+0.47 |
| Songkhla | - | (0.32,0) | (0.96,-0.11) | 0.006 | - | 0.25+0.56 |

From Table 4.1, it shows the number of rain days and maximum temperature are not significant related with DHF incidence (p-value 0.05), so the model can be reduced to be that shown in Table 4.2

Table 4.2: The final model of time series analysis of DHF incidence

| Province | Seasonal | Noise | Total $r^2$ |
|---|---|---|---|
| Krabi | (0.52,0.12) | (0.59,0.25) | 0.41+0.38 |
| Trang | (0.37,0.12) | (0.89,-0.12) | 0.56+0.28 |
| NST | (0.46,0) | (1.06,-0.23) | 0.38+0.46 |
| Songkhla | (0.35,0) | (0.95,-0.11) | 0.25+0.56 |