Chapter 3

Preliminary Data Analysis

This chapter reports on preliminary data analysis of global monthly surface temperatures in South East Asia from 1973 to 2008. Graphical and statistical analysis was carried out using R.

3.1 Study area

This study area consisted of 25 regions covering most of Southeast Asia. Each regions a 5^o by 5^o grid box covered latitudes 0^o N to 25^o N and longitudes 85^o E to 110^o E as shown in Figure 3.1. The global monthly surface temperatures cover both land and sea.

Prince of Songkla University Pattani Campus

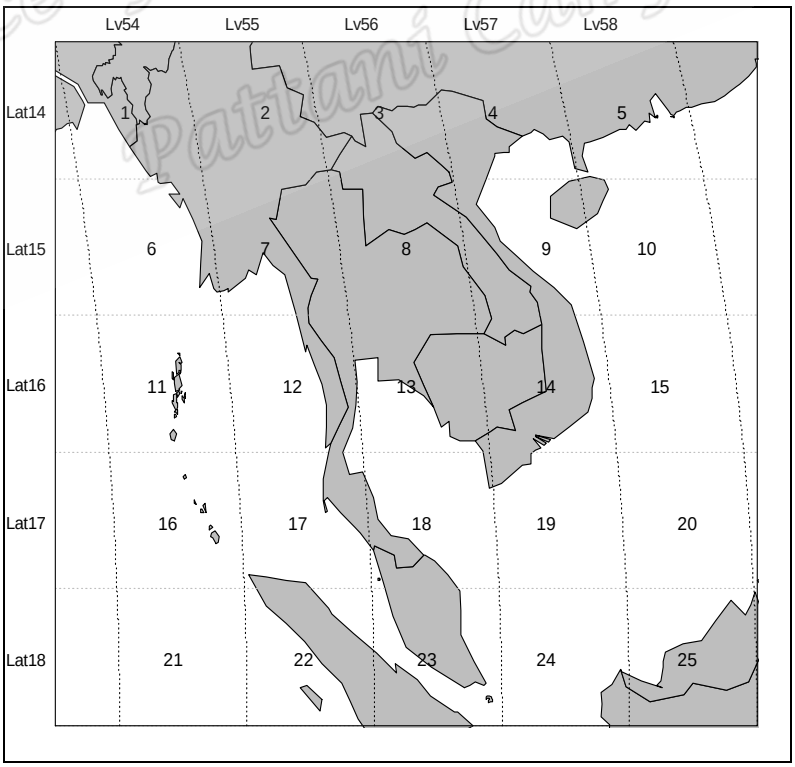

Figure 3.1 Map of study area

3.2 Summary of the temperatures data

The numerical summaries of the monthly surface temperatures for all regions are shown in Table 3.1.

Table 3.1 Data summaries of monthly raw temperature in each region

Region	Latitude band	Longitude band	n	Mean	Median	Minimum	Maximum
1	14	55	430	24.38	25.79	17.20	28.80
2	14	56	340	22.05	23.25	12.30	27.10
3	14	57	402	20.21	21.30	10.90	25.90
4	14	58	432	21.89	23.23	9.30	29.13
5	14	59	432	22.34	23.54	11.55	29.13
6	15	55	380	27.01	27.38	22.04	30.79
7	15	56	425	25.90	26.15	20.22	30.70
8	15	57	423	25.49	26.26	18.80	30.05
9	15	58	429	25.25	25.96	18.75	30.52
10	15	59	432	26.25	26.89	20.97	29.77
11	16	55	432	27.56	27.62	23.86	31.74
12	16	56	430	27.45	27.41	24.40	30.53
13	16	57	432	27.61	27.68	23.70	30.46
14	16	58	432	26.02	26.38	21.90	28.89
15	16	59	432	27.35	27.72	23.85	29.90
16	17	55	432	27.70	27.64	26.34	29.39
17	17	56	432	27.59	27.61	25.65	29.13
18	17	57	432	27.45	27.56	24.26	29.99
19	17	58	432	27.43	27.74	23.79	29.68
20	17	59	432	27.64	27.83	25.20	29.82
21	18	55	432	27.86	27.82	26.30	29.31
22	18	56	432	26.55	26.51	25.42	27.89
23	18	57	432	26.78	26.81	24.86	28.58
24	18	58	432	27.65	27.79	24.92	29.38
25	18	59	432	26.30	26.34	24.49	27.98

Table 3.1 shows that there are 432 observations in each region. Region 1, 2, 3, 6, 7, 8, 9 and 12 have some missing values. The range of monthly surface temperature varied from 9.30^{o} C in region 4 to 31.74^{o} C in region 11. The time series plot of these monthly surface temperatures is displayed in Figure 3.2 for region 1 (latitude 20 ^{o} N

Prince of Songkla University Pattani Campus

to 25^{o} N and longitude 85^{o} E to 90^{o}E) and region 6 (latitude 15^{o} N to 20^{o} N and 85^{o} E to 90^{o}E**)**

In Figure 3.2, each dot denotes the monthly surface temperature, and the dots are joined to show the pattern of change over the period. The graph clearly shows that the pattern is seasonally affected, that is, the monthly temperatures vary with time in a pattern that matches the seasons.

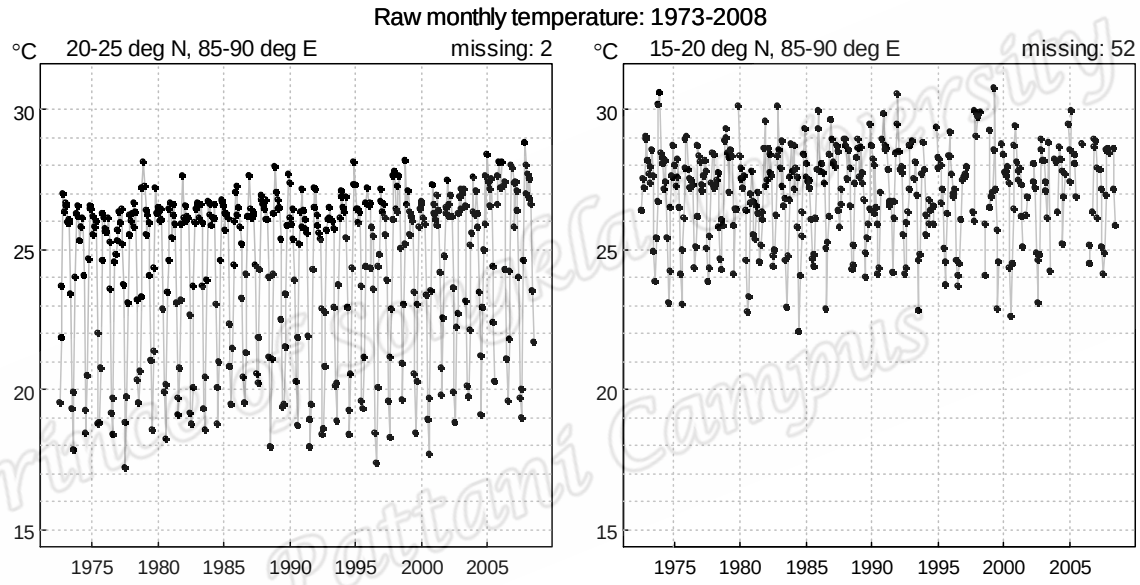

Figure 3.2 Time series plot of raw temperatures for regions 1 and 6

Next, the data were seasonally adjusted to remove the variation of monthly surface temperatures by subtracting the monthly averages of raw surface temperature and then adding back the overall mean temperatures. Figure 3.3 shows time series plots of seasonally adjusted temperatures of the 2 regions above.

In Figure 3.3, the trend can be easily revealed and compared to the graph without adjusting for the seasonal effect. However, the change of overall trend along period not clear.

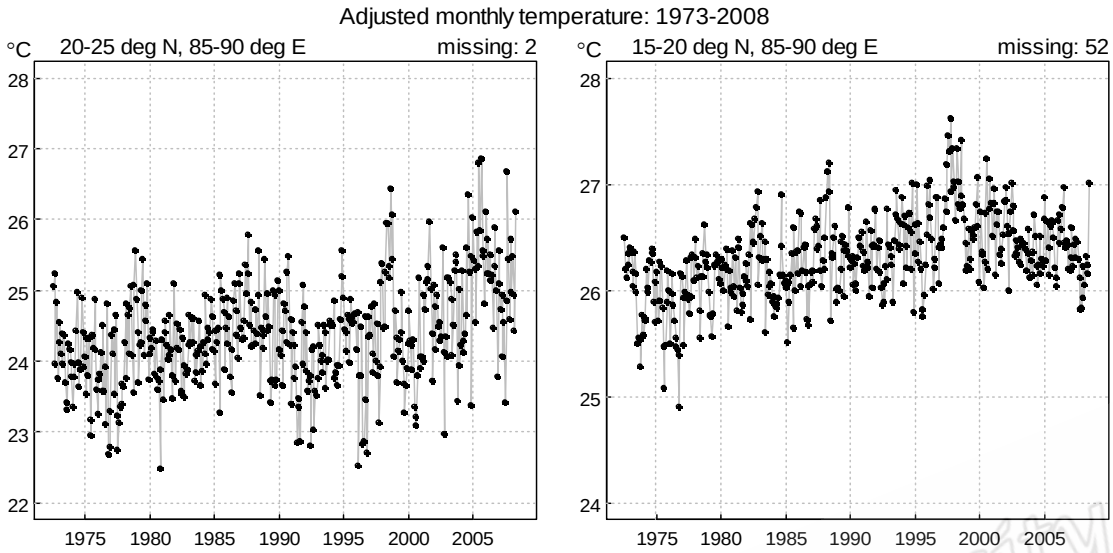

Figure 3.3 Time series plot of seasonally adjusted temperatures for regions 1 and 6

3.3 Autoregressive process and simple linear regression model

To see the surface temperature trend more clearly, a simple linear regression model was fit to these two regions. The fitted line from the model is shown in Figure 3.4. The temperature increases about 0.26^{o} per decade for region 1 (the region on the left panel) and 0.17^{o} per decade for region 6 (the region on the right panel). In time-series analysis the correlation of residuals from the fitted model is assumed to be stationary. Figure 3.5 shows autocorrelations of the residuals.

Prince of Songkla University Pattani Campus

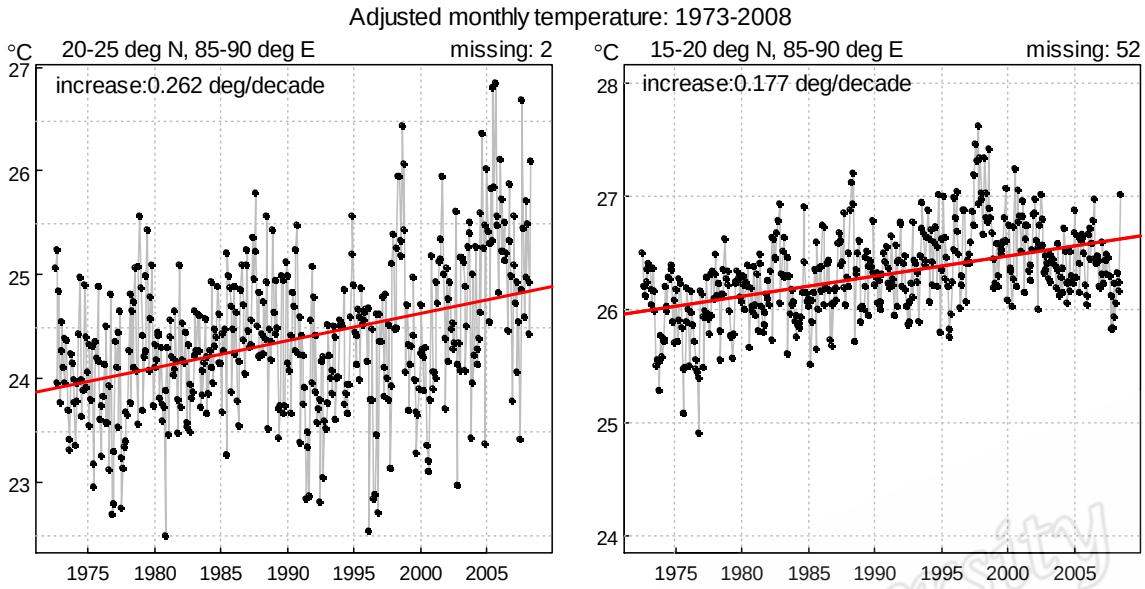

Figure 3.4: Seasonally adjusted temperatures with linear trend superimpposed

Figure 3.5 shows autocorrelations of the residuals before filtering for region 1 and 6. The autocorrelation function (ACF) plots suggest that the residuals are not stationary. To adjust the residual to be stationary, an autoregressive process will be fit to the residuals.

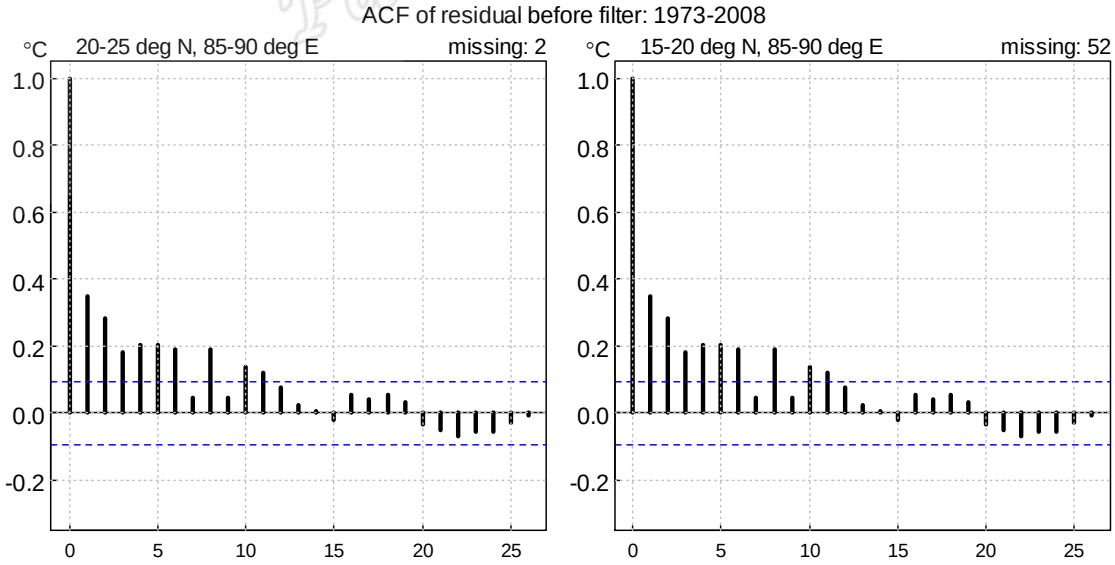

Figure 3.5: ACF of residuals before filtering for regions 1 and 6

Prince of Songkla University Pattani Campus

Figure 3.6 shows autocorrelations of the residuals after filter the AR(1) process. The plot shows some autocorrelations were removed, then we fitted an AR(2) to remove the autocorrelations of the second order.

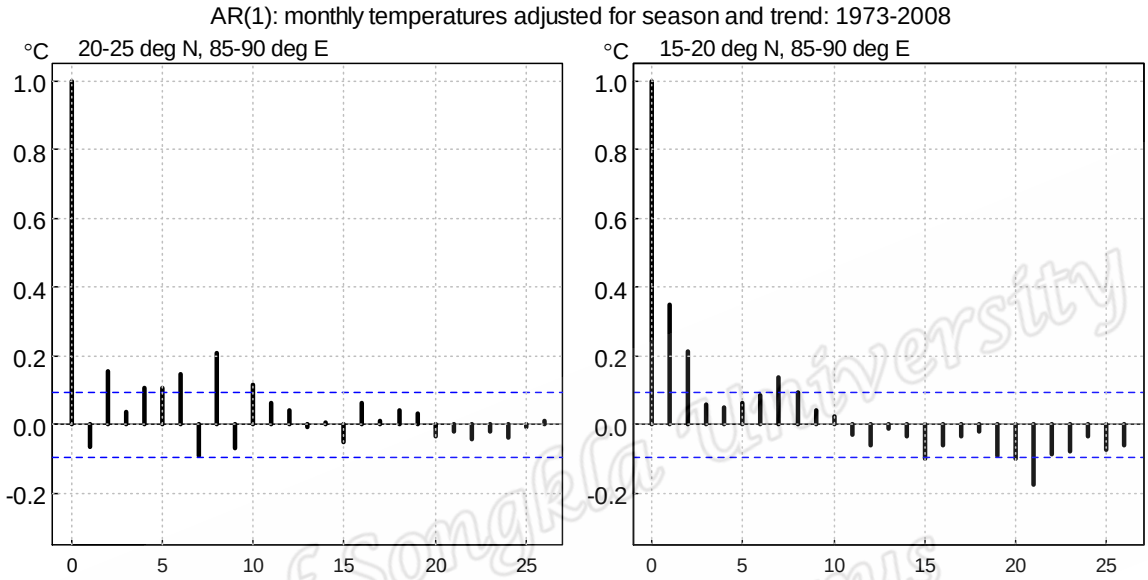

Figure 3.6 Autocorrelations of the residuals corresponding to region 1 and 6

Figure 3.7 shows the autocorrelation of the residuals after filter AR(2) process in region 1 and 6. The autocorrelation function (ACF) plots reveal that most of all autocorrelations are within the 95% confidence interval.

Prince of Songkla University Pattani Campus

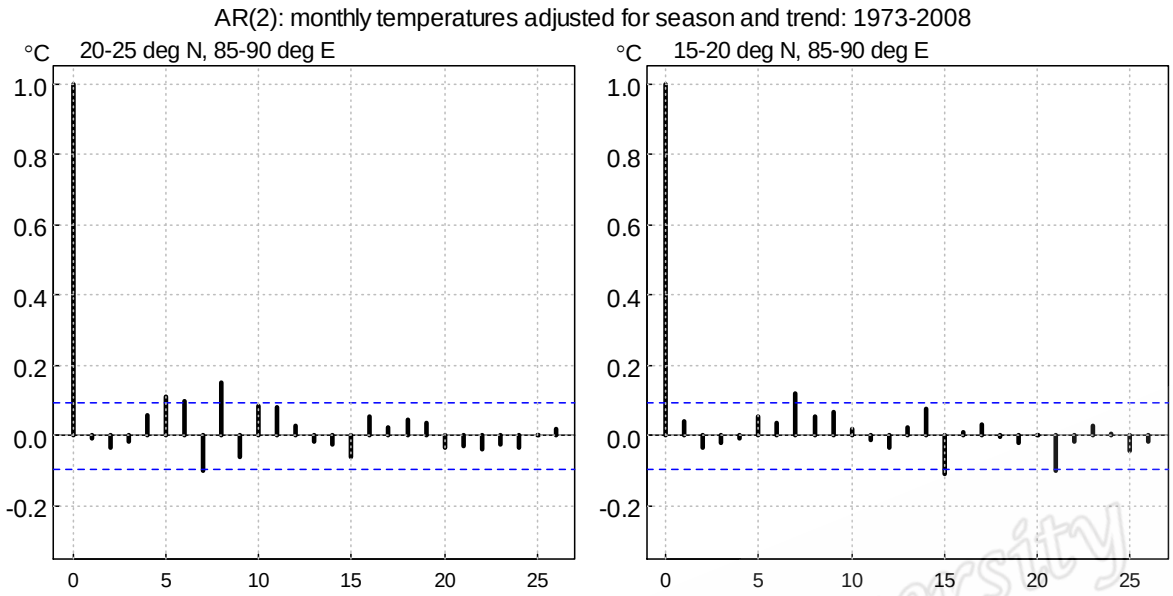

Figure 3.7 AR(2) corresponding to 2 example regions

Then fit the autocorrelation of the residuals order two in all 25 regions. After filtering all regions then linear regression models were fitted to the data. The model indicates that a statistically significant increase in monthly temperature but there are some spatial correlations between regions. Figure 3.9 shows the linear fit for all 25 regions.

Prince of Songkla University Pattani Campus

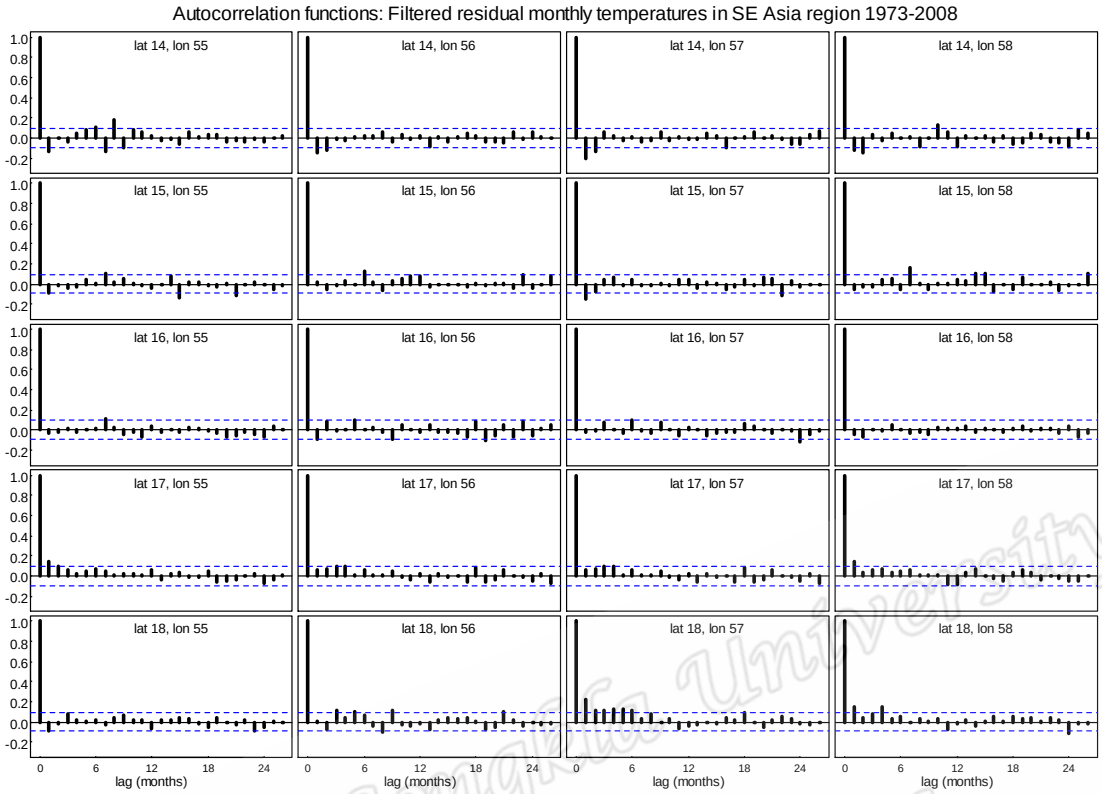

Figure 3.8 AR(2) corresponding to all 25 regions

Figure 3.9 shows the time series of seasonally adjusted data with fitted lines superimposed in each region. Separate linear models were fitted to the data of the 25 grid-box regions. The average temperatures were different in each region. The graphs show that the temperature in each region varies with time and the trend steadily increases by about 0.08^{o}C - 0.48^{o}C over the past three decades.

Prince of Songkla University Pattani Campus

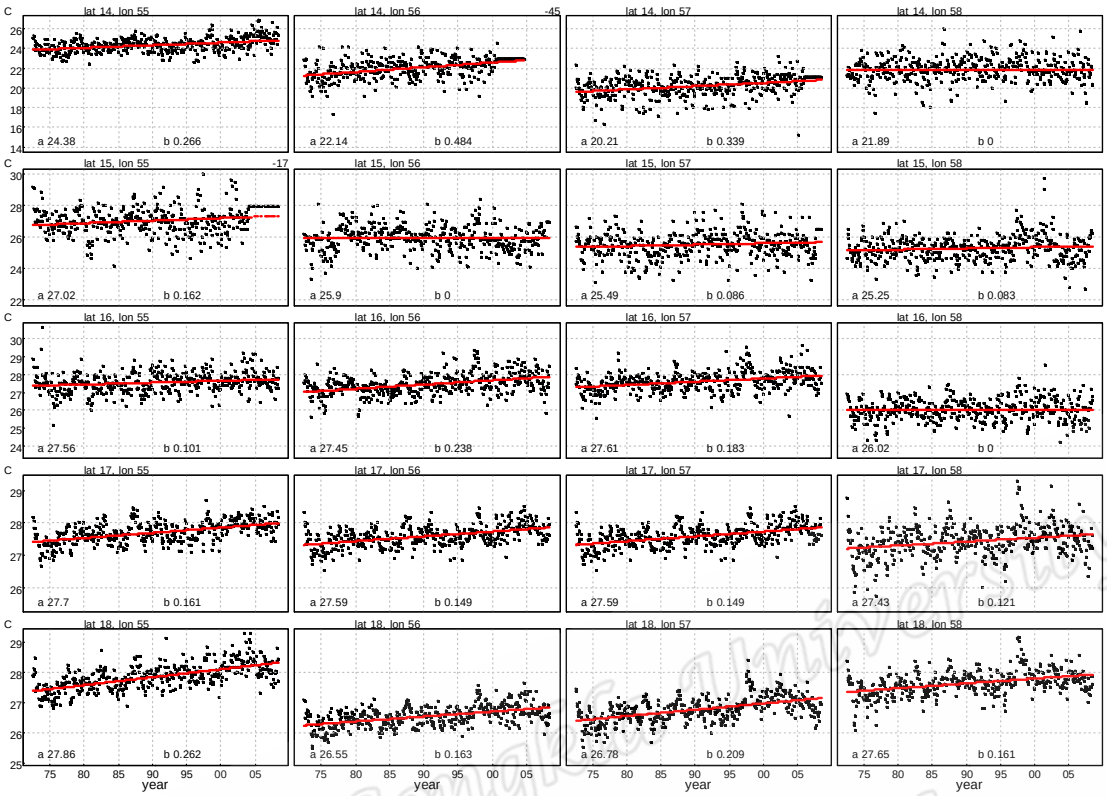

Figure 3.9 Seasonally adjusted temperatures with fitted line in each region.

Linear regression models may not be an appropriate model to fit this data because the data exhibit is more complicated. The data not always increase or decrease whole time but exhibit up and down through time period. The long term trend does not steadily increase or decrease over time. The seasonally adjusted temperatures appear to increase and decrease. In the next section, linear splines are fit to the temperature data in order to obtain a better fit.