Chapter 4

Statistical Mapping of Morbidity

In this chapter we use some statistical methods to map the disease data for subdistricts in Yala Province for 2002 and 2003. These methods can be applied more generally to any data of continuous data type recorded at specified geographical locations in a region on the earth's surface. We first use the Poisson distribution to model the incidences of the various diseases in Yala subdistricts, and assess the independence assumption using normal scores plots. Where there is over-dispersion due to failure of the independence assumption, the negative binomial distribution is more appropriate, and we calculate probabilities associated with locations where the morbidity is unusually high. The results from this method are compared and discussed, and an interpretation is given to the final results.

As described in Chapter 2, we calculate approximate p-values based on the incidence of the seven disease outcomes (Dengue haemorrhagic fever, Dengue fever, Pneumonia, Diarrhoea, Pyrexia, Haemorrhagic conjunctivitis, Malaria) in the 58 subdistricts in Yala province.

As in Chapter 3, in this chapter we again use schematic range maps for mapping such data as routinely provided by geographical information systems (GIS) software.

4.1 Statistical modelling

Table 4.1 shows the incidence rates per 1000 for the seven fever-symptomatic diseases for the 58 subdistricts in Yala Province in 2002 and 2003. The distributions of these rates for 2002 are shown in Figure 3.1.

Disease	2002 incidence	2003 incidence
Dengue Haemorrhagic Fever	2.160	0.642
Dengue Fever	3.295	0.619
Haemorrhagic Conjunctivitis	10.282	4.070
Diarrhoea	31.965	35.368
Pneumonia	5.928	6.375
Pyrexia of unknown origin	9.530	9.081
Malaria	1.941	3.773

Table 4.1: Average incidence rates/1000 of diseases in Yala subdistricts

Figures 4.1 and 4.2 show standardised residuals plotted against normal scores. These residuals are obtained by fitting Poisson distributions to the incidences for each disease in each year, and are computed from the formula

$$z_i = \frac{y_i - \beta P_i}{\sqrt{\beta P_i}}, \tag{4.1}$$

where P_i is the population at risk in subdistrict i and β is the average incidence rate for each disease and year, as given in Table 4.1. This formula is appropriate for standardising the incidences to have zero means and unit standard deviations, because if Y has a Poisson distribution with mean λ, its variance is also equal to λ, and its standard deviation is $\sqrt{\lambda}$.

All of these plots indicate that the Poisson distribution does not fit the incidence rates, suggesting that a more general model allowing for over-dispersion is more appropriate. The negative binomial distribution provides a simple generalisation of the Poisson distribution. If the mean of the negative binomial distribution is also denoted by λ, its variance is given by $\lambda(1+\lambda\alpha)$, where α is the over-dispersion parameter. Thus the standardised residual incidence after fitting the negative binomial model is

$$z_i = \frac{y_i - \beta P_i}{\sqrt{\beta P_i (1+\alpha\beta P_i)}}. \tag{4.2}$$

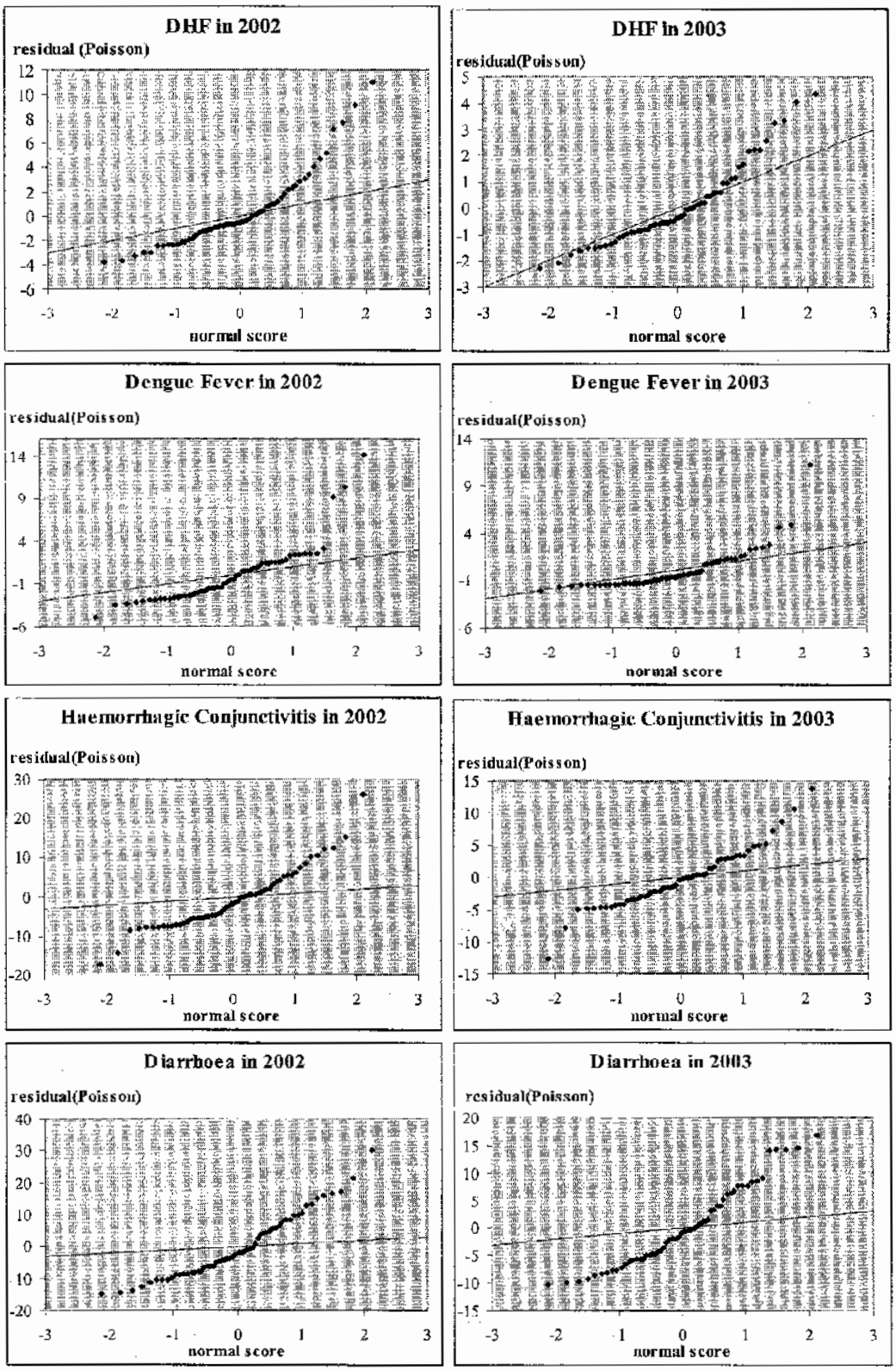

Figure 4.1: Normal scores plots of standardised residuals for four diseases in Yala

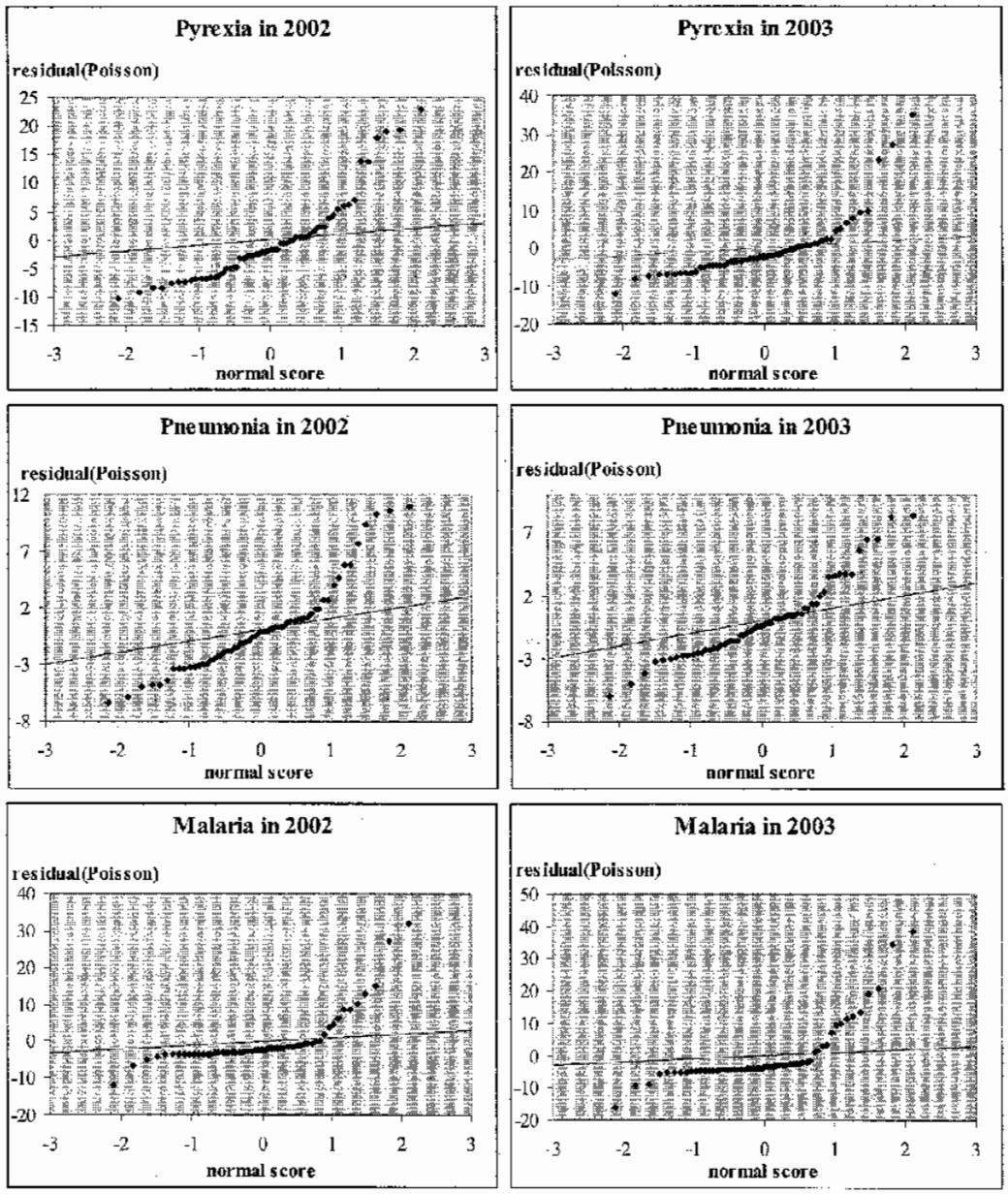

Figure 4.2: Normal scores plots of standardised residuals for three diseases in Yala

Table 4.2 shows the parameter estimates and 95% confidence intervals after fitting negative binomial distributions to the disease incidences (using the R statistical package). Note that the estimates of the parameter β are similar to the values given in Table 4.1, because this parameter is the mean of the distribution. The slight differences

are due to the fact that the estimates in Table 4.2 are obtained using maximum likelihood instead of the method of moments.

Disease	2002		2003	
	β	α	β	α
DHF	2.183 (1.82-2.62)	0.422 (0.27-0.65)	0.644 (0.54-0.77)	0.245 (0.11-0.56)
Dengue	3.224 (2.69-3.87)	0.434 (0.29-0.65)	0.617 (0.49-0.78)	0.522 (0.30-0.92)
Conjunctivitis	10.234 (7.34-14.22)	1.604 (1.14-2.26)	4.046 (3.14-5.21)	0.917 (0.63-1.34)
Diarrhoea	31.937 (25.71-39.68)	0.704 (0.50-0.99)	35.331 (30.82-40.50)	0.276 (0.19-0.39)
Pneumonia	5.929 (4.91-7.17)	0.509 (0.34-0.75)	6.371 (5.61-7.24)	0.215 (0.14-0.33)
Pyrexia	9.495 (7.39-12.20)	0.922 (0.66-1.30)	9.050 (7.21-11.35)	0.753 (0.54-1.06)
Malaria	1.944 (1.24-3.05)	2.942 (2.02-4.28)	3.773 (2.29-6.21)	3.700 (2.57-5.32)

Table 4.2: Estimated parameters in negative binomial model for disease incidences

The Poisson distribution arises as the special case of the negative binomial distribution when α 0. Since all the confidence intervals for α are substantially greater than 0, the negative binomial distribution should be a better model than the Poisson distribution.

The residual deviances obtained from fitting the Poisson and negative binomial models are given in Table 4.3. The deviance is used to globally assess the goodness-of-fit of a model, by computing the probability that a chi-squared random variable with $n-1$ degrees of freedom (57 in this case) exceeds it.

Based on the values in Table 4.3, the negative binomial distribution provides a satisfactory fit in all cases, because the smallest such probability is 0.124, corresponding to haemorrhagic conjunctivitis in 2002.

Disease	2002		2003	
	Poisson	Neg Bin	Poisson	Neg Bin
DHF	344.321	63.110	112.607	68.431
Dengue	471.040	62.415	171.569	61.934
Conjunctivitis	2950.139	69.464	1015.511	68.131
Diarrhoea	5812.382	66.036	2845.134	60.937
Pneumonia	895.916	66.668	505.962	61.008
Pyrexia	2818.782	66.315	2682.209	63.895
Malaria	1889.359	64.423	3553.866	65.097

Table 4.3: Deviance from Poisson and negative binomial models

A more detailed picture of the goodness-of-fit of a model is provided by plotting standardised residuals against appropriate expected values. For the Poisson distribution it is appropriate to use normal scores, because the Poisson distribution is asymptotically normal in the limit as its mean approaches infinity. However, as stated in Chapter 2, the corresponding limit for the standardised negative binomial distribution is a shifted gamma distribution with mean 0, scale parameter $1/\sqrt{\alpha}$, and shape parameter $1/\alpha$.

Figures 4.3 and 4.4 show the standardised residuals from the negative binomial model plotted against these gamma scores for the seven diseases in 2002 and 2003.

Comparing Figures 4.3 and 4.4 with the corresponding graphs based on the Poisson distribution shown in Figures 4.1 and 4.2, we see that the fits are much better.

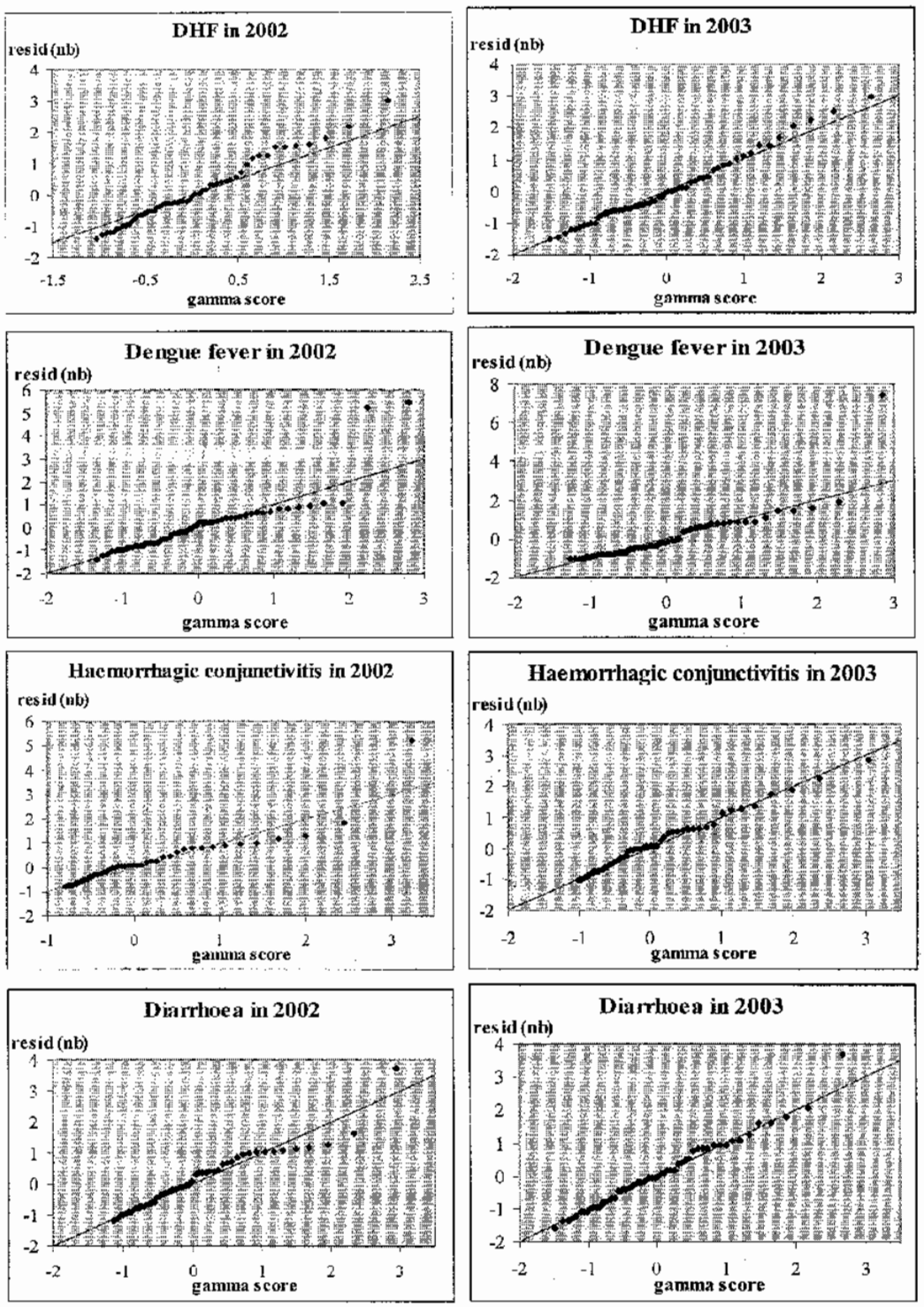

Figure 4.3: Gamma scores plots of standardised residuals for four diseases in Yala

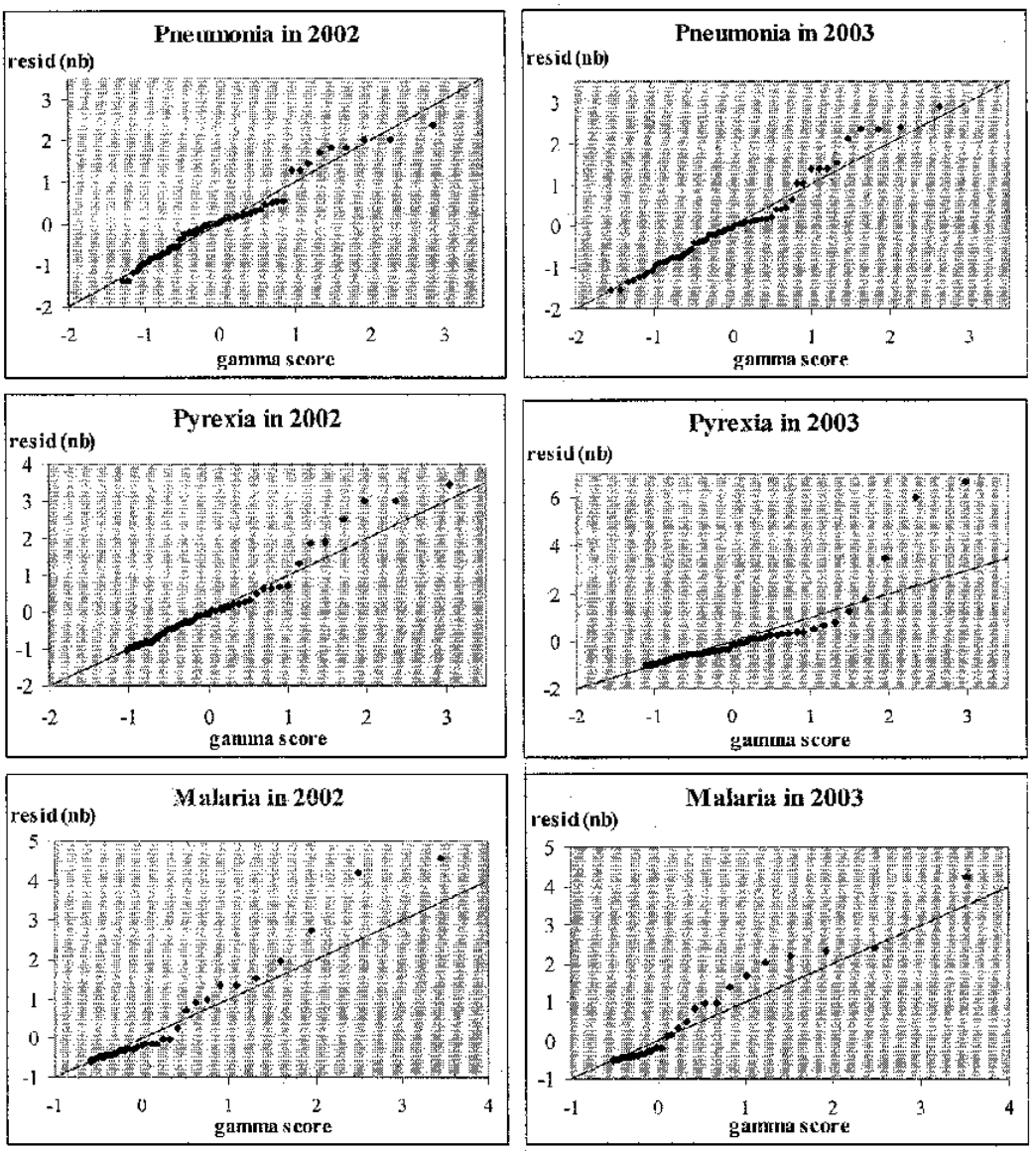

Figure 4.4: Gamma scores plots of standardised residuals for three diseases in Yala

4.2 Alert maps of disease incidence

Table 4.3 shows the colouring system used for creating maps indicating alert levels. In this case just three colours are used to code the subdistricts, based on p-values calculated from the tail of the distribution. For simplicity we use the limiting gamma distribution to compute the p-value associated with a subdistrict. Assuming that the

whole region contains n subregions to be mapped using a range map with these colours, the p-values need to be divided by n to allow for multiplicity. In this case $n = 58$.

Colour	P-value	Alert level
Light blue	$> 0.05/n$	None
Orange	$0.01/n - 0.05/n$	Moderate
Red	$< 0.01/n$	High

Table 4.4: Alert levels based on p-values adjusted for multiplicity

Applying this definition, we find that only five alerts are recorded for any disease in 2002 or 2003. There are orange alerts for Ka Yu Bo Ko ($p = 0.0296$) and KuTaBaRu ($p = 0.0422$) for dengue fever in 2002, and for Kerikat ($p = 0.0172$) and Than To ($p = 0.0358$) for pyrexia in 2003. There is just one red alert, for Ka Yu Bo Ko ($p = 0.0032$) in 2003. Figures 4.5 and 4.6 show the corresponding maps for dengue fever and pyrexia.

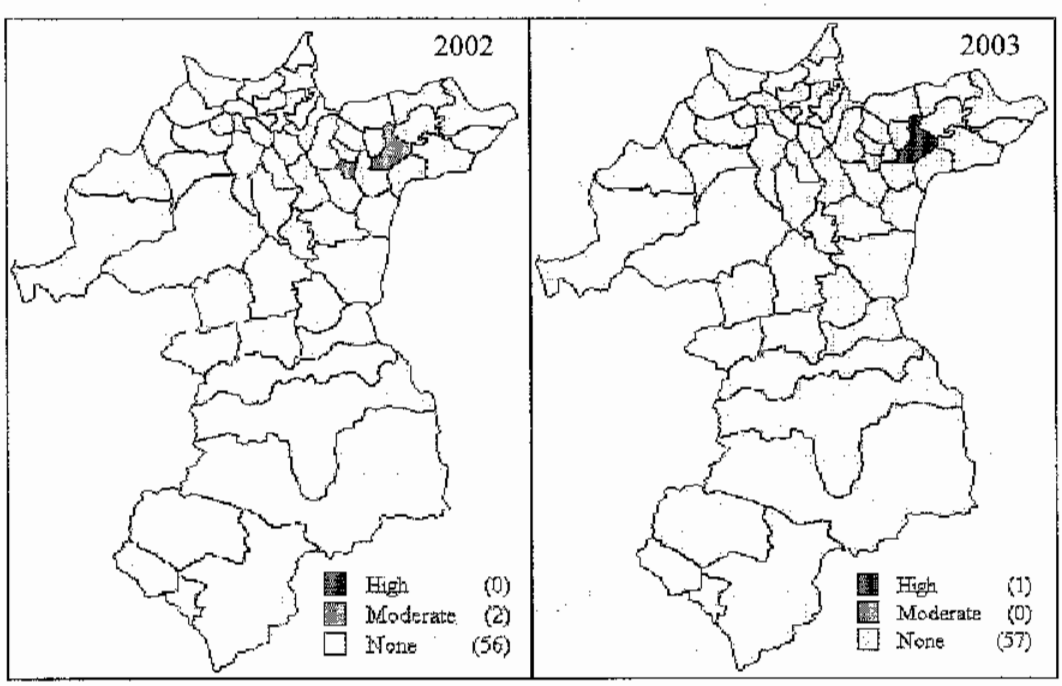

Figure 4.5: Alert maps of Dengue fever in 2002 and 2003

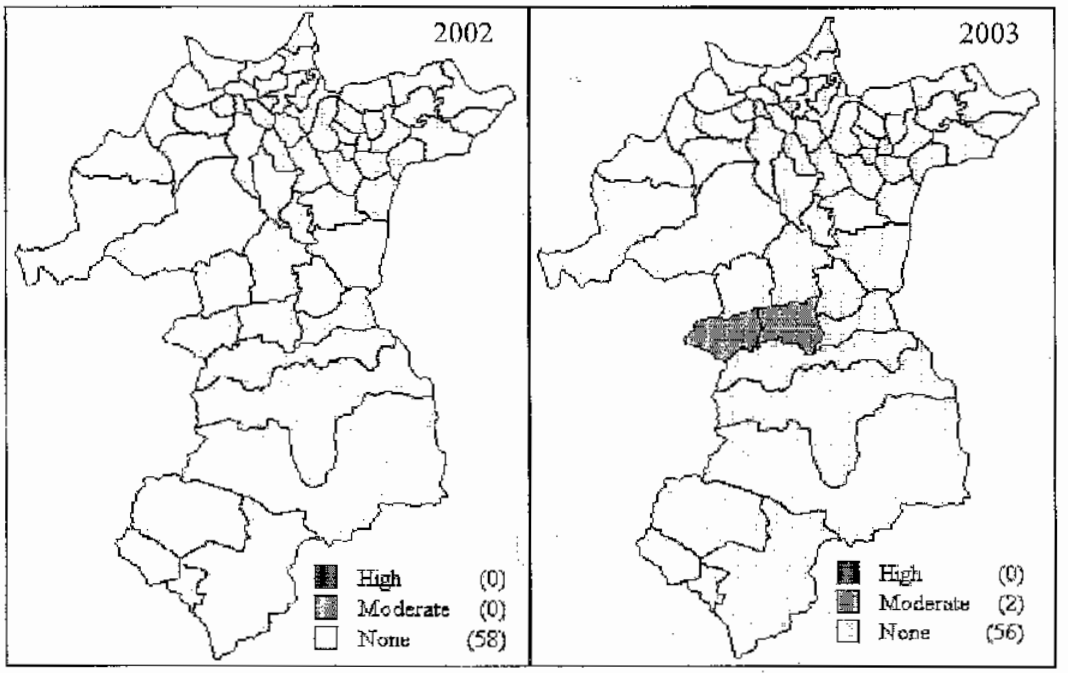

Figure 4.5: Alert maps of Pyrexia in 2002 and 2003