Settlers located on a road had significantly greater percentages of land cleared than farmers located at a distance from a 'main' road.

Analysis of variance for the distance to mill categories (Table 4.19) shows some interesting results. Land cleared (%) was found to be significantly higher for settlers located within one mile of a grist mill, compared with those settlers located more than two miles from a mill. In terms of wheat production, farmers living closer to mills appear to have been growing significantly greater amounts of wheat than their counterparts living more than two miles from a mill. It is also notable that wheat yield was significantly greater for those farmers living less than four miles from a mill, compared to those in more remote sections of the township.

The analysis of variance results further confirm the validity of hypothesis I, that persistent settlers were successful settlers, and hypotheses IV and V, concerning settler location 'vis-à-vis' roads and mills.

4.11 Regression Analysis

The Pearson correlation coefficient (R) provides us with a measure of the strength of a linear association between two variables, but does not tell us anything about the actual line. Regression techniques may be used to calculate an equation for the line that describes the relationship between two (or more) variables, to predict the values of one variable (land cleared) from another (time settled), and to 'explain' the proportion of variability in the dependent variable (land cleared) that can be accounted for by the independent variable (time settled).

Simple regression analysis techniques have been used to further explore the relationship between land cleared (%) and time settled in Emily Township for the years 1822 to 1851. Table 4.20 summarizes the results of linear regression analysis with statistics including a regression equation, multiple R (correlation coefficient), and R-square (percentage of variability 'explained').

Table 4.20 Regression Summary

Land Cleared - Time Settled

YEAR	REGRESSION EQUATION	MULTIPLE R	R-SQUARE
1822	Y = 0.82 + 2.2(Time Settled)	.433	.187
1823	Y =-0.33 + 2.0(Time Settled)	.513	.263
1825	Y = 2.1 + 1.2(Time Settled)	.384	.148
1826	Y = 2.5 + 1.4(Time Settled)	.693	.480
1827	Y = 0.80 + 1.6(Time Settled)	.678	.459
1828	Y = -1.0 + 1.8(Time Settled)	.723	.523
1830	Y =-0.50 + 1.5(Time Settled)	.636	.404
1841	Y = -1.5 + 1.3(Time Settled)	.481	.231
1851	Y = 25.7 + 0.7(Time Settled)	.299	.089

Y= Intercept + Slope X (Independent Variable)
Dependent Variable 'Y' = Land cleared (% of farm)
Time Settled - Years

The regression equations for each census year all showed positive values for the slope of the line, indicating that, as the value of the variable 'time settled' increased, the value of the variable 'land cleared (%)' also increased. The regression equation for the years 1841 and 1851 shows the slope declining - smaller increases in land cleared with increasing time settled. The low R-square value in 1851 indicates that much less of the variability in land cleared was being explained by the number of years settled.

4.12 Multiple Regression Analysis

Multiple regression techniques may be used to study the relationship between a single dependent variable (land cleared), and several independent variables such as time settled, farm size, family size, oxen, horses, etc. Table 4.21 summarizes the results of stepwise multiple regression analysis, including multiple regression equations, multiple correlations (R), R-square percentages of variability explained, and Beta weights, which indicate the relative importance of each independent variable in explaining the variation of the dependent value.

For all census years 1822 to 1851, the multiple regression R-square is substantially higher than the R-square of the corresponding simple linear regression, indicating that the model is 'explaining' a greater amount of variability with additional variables included in the regression. Other predictor variables that are shown to have influenced percentage of land cleared are oxen, horses, total livestock, farm size, males, males over 16, and family size.

The values for slope are positive for time settled, oxen, horses, family size, males, and males over 16, indicating that as these variable values increased, a corresponding

Table 4.21 Multiple Regression Summary 1822-1851

YEAR	REGRESSION EQUATION	MULTIPLE R	R SQUARE	BETA WEIGHTS
1822	Y= 1.9 + 2.7(Time)+ 2.3(Oxen) - 0.04(Farm Size)	.613	.376	Time .542 Oxen .346 Farm Size -.264
1823	Y=-0.1 + 1.4(Time) + 0.6(Livestock)	.593	.352	Time .360 Livestock .335
1825	Y= 4.4+0.9(Livestock)	.482	.232	Livestock .482
1826	Y= 4.3 + 1.1(Time) + 1.6(Oxen) + 0.4(Family) - 0.04(Farm Size)	.770	.593	Time .518 Oxen .271 Family .233 Farm Size -.193
1827	Y= 6.8+1.1(Livestock)- 0.1(Farm Size) + 0.7(Time) + 0.4(Males)	.818	.669	Livestock .410 Farm Size -.329 Time .308 Males .131
1828	Y= 5.1 + 1.1(Time) + 1.2(Livestock)- 0.1(Farm Size) + 1.1(Male over 16)	.848	.718	Time .430 Livestock .373 Farm Size -.310 Male over 16 .131
1830	Y= 7.7+0.9(Livestock)- 0.1(Farm Size) + 0.8(Time) + 1.8(Oxen)+ 2.2(Horses)	.823	.678	Livestock .294 Farm Size -.320 Time .326 Oxen .246 Horses .122
1841	Y = 6.1 + 8.3(Horses)+ 0.7(Time) - 0.1(Farm Size)+ 1.7(Oxen)+1.0(Males)	.703	.494	Horses .548 Time .269 Farm Size -.301 Oxen .129 Males .109
1851	Y= 34.7 + 0.7(Time) - 0.2(Farm Size) + 0.4(Livestock)	.545	.297	Time .273 Farm Size -.463 Livestock .424

Dependent variable 'Y' = Land Cleared (% of farm)
Time = Time Settled (Years) ; Farm Size (acres)

Family = Family Size ; Males = Total males in family;

increase in land cleared (%) could be expected. The negative slope of the variable 'farm size' indicates that farms with greater percentages of land cleared tended to be smaller.

For the years 1822, 1823 (note that time settled is the only significant variable in 1823), 1826, and 1828, the varaible time settled (per Beta weights) was the most important in explaining variation in percentage land cleared; livestock (another indicator of relative success) is most important in 1825 and 1827, while oxen, horses, and farm size are most important in 1830, 1841, and 1851, respectively. It is interesting to note that, by 1841, horses appeared to be playing a greater role in land clearance than oxen; by 1851 oxen had all but disappeared from the township.

Overall, the results of regression analysis show the close links that existed among three aspects of frontier agriculture - land cleared, livestock, and time settled. In particular, the regression analysis offers more support for hypothesis I.

4.13 Residual Analysis

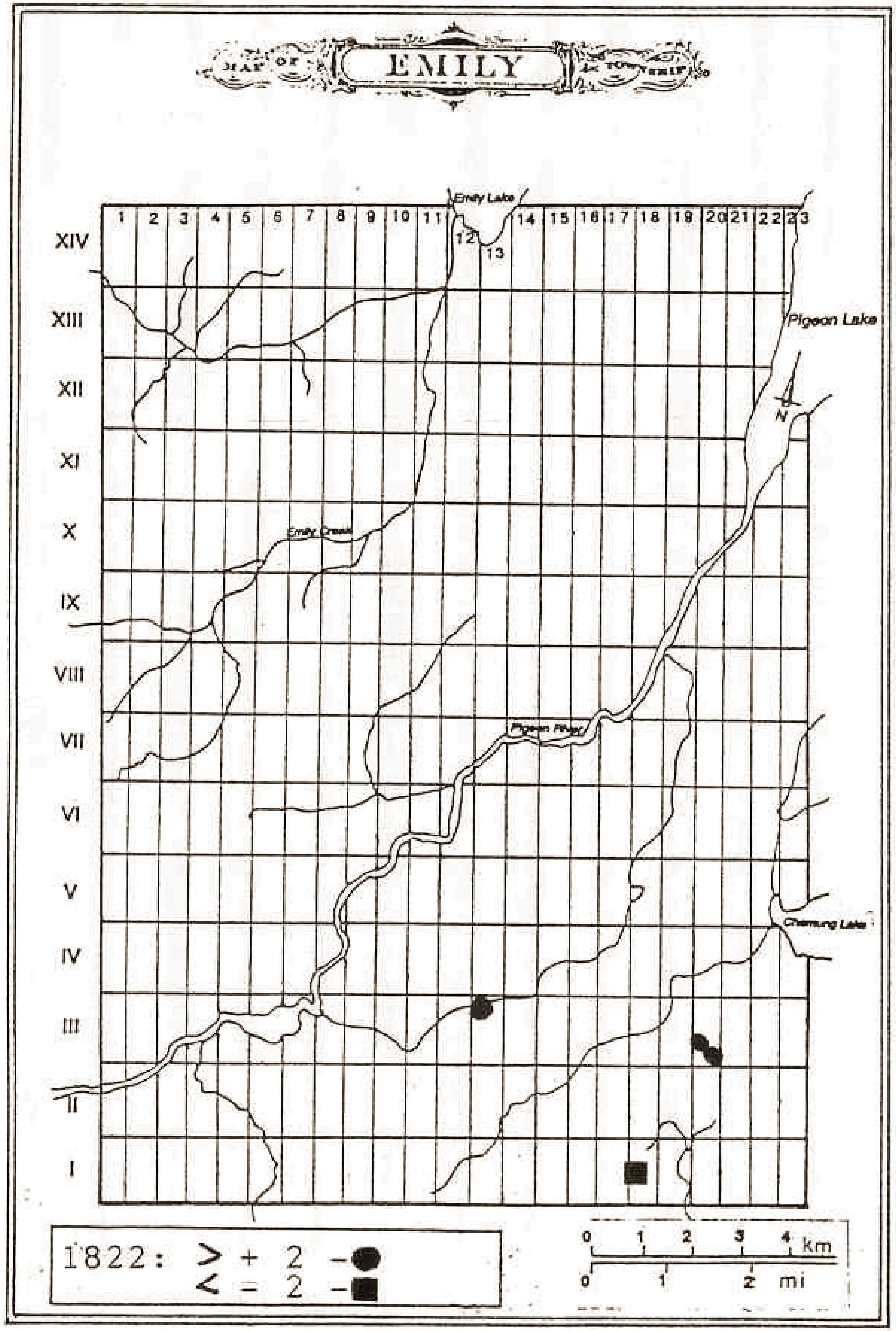
In a linear regression, the residual is the difference between the observed and predicted values of the dependent variable (Norušis 1988,p.351). 'Outliers' (major departures from the pattern of residuals), may point out errors in measurement or data transcription, but may also provide revealing insights about the phenomenon under study (Clark and Hosking 1986,p.373). Norcliffe (1982 p.222) suggests that in geographical analysis large residuals (outliers) should be mapped. Figures 4.48, 4.49, 4.50, and 4.51 illustrate the location of standardized residuals larger than +2, and smaller than -2, for the simple regression of land cleared with time settled.

Positive residuals indicate cases where settlers had cleared more land than the model predicted, while negative residuals represent cases where the model has predicted too much land cleared for settlers with little land cleared.

Time constraints do not allow for a detailed look at all large residuals for each year, but a brief examination of several cases may illustrate the value of residual analysis in identifying other variables which may have influenced an individual settler's land clearance.

For 1822, the model has predicted that William Lee at Con.1, Lot 18 would have cleared 7.3% of his land. Lee's actual land clearance is reported as none. A check of the data reveals that he is located beside his son, Nathan, who has cleared 4.5 acres. It is possible that the elder Lee has been busy helping his son clear his lot, but another possibility exists - errors in the assessment. By 1823, Lee had cleared five acres of his lot, and listed two additional adult males and one adult female as family members.

In 1825, William Holroyd (the first township clerk), had cleared 19% of his land, well above the 8.2% predicted by the model. Other factors that may be taken into consideration include Holroyd's large family (including five sons), and his situation, located in an area of good soils, and



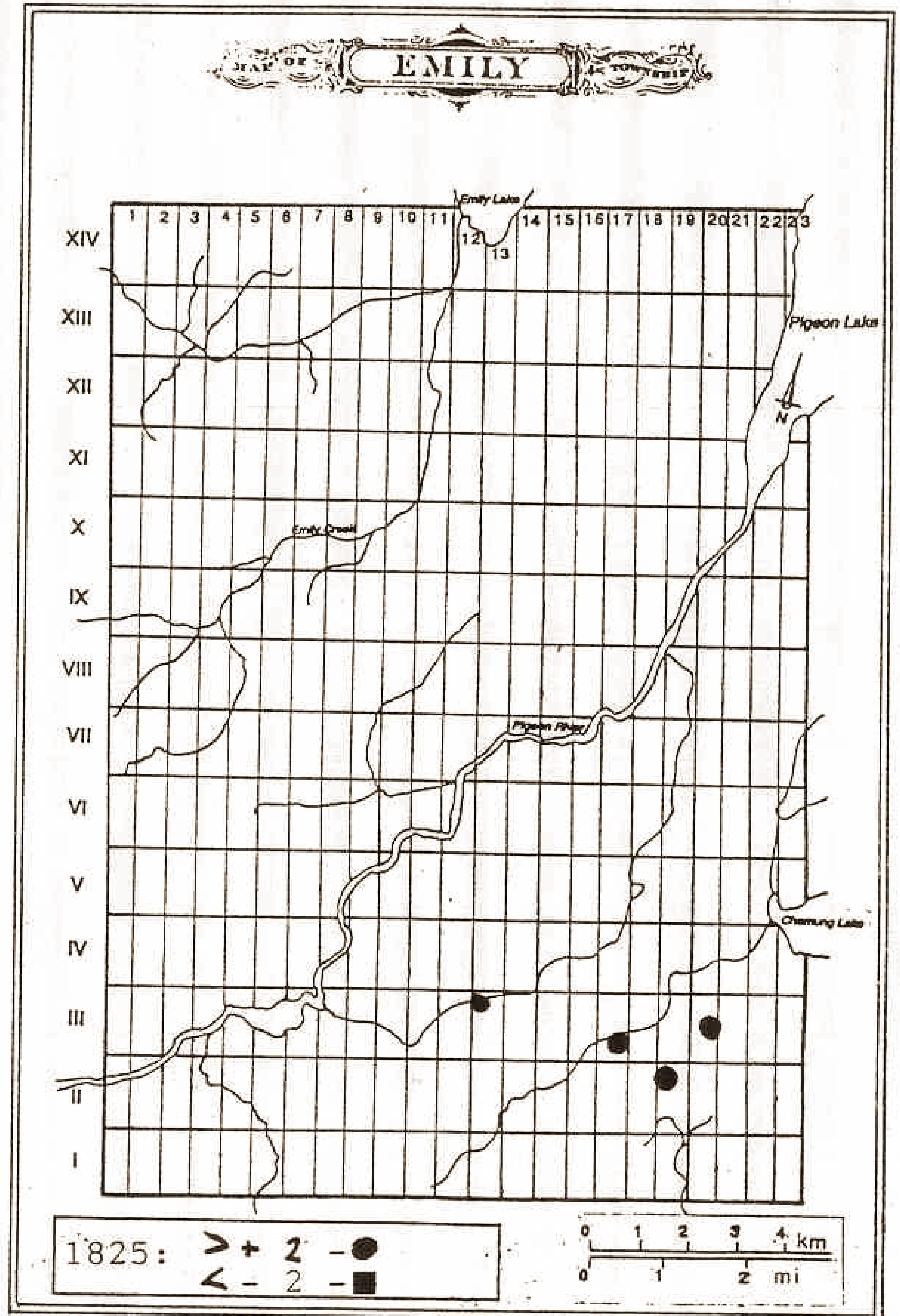


Figure 4.48 : Location of Outliers - Standardized Residuals 1822 and 1825

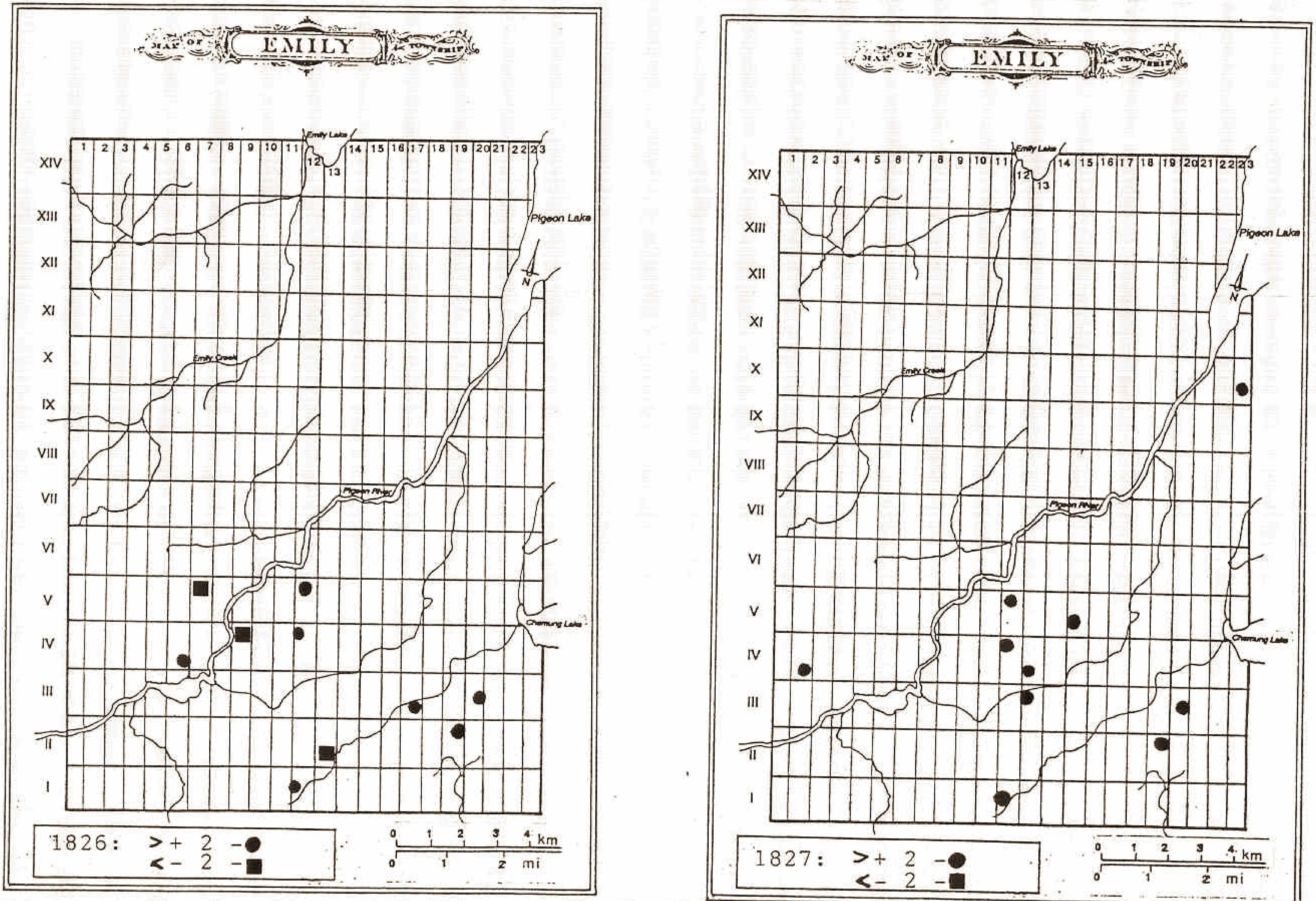


Figure 4.49 : Location of Outliers - Standardized Residuals 1826 and 1827

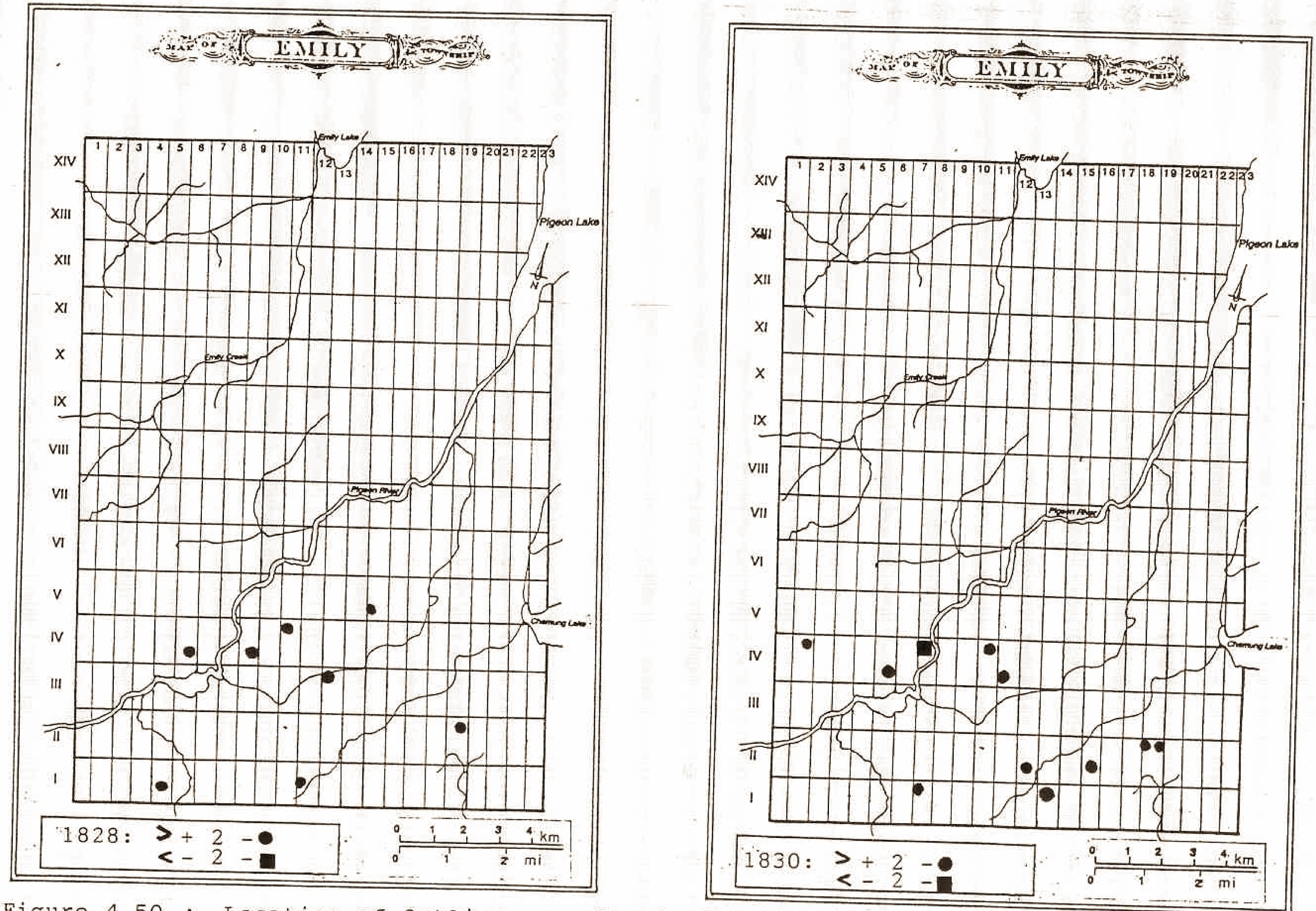


Figure 4.50 : Location of Outliers - Standardized Residuals 1828 and 1830

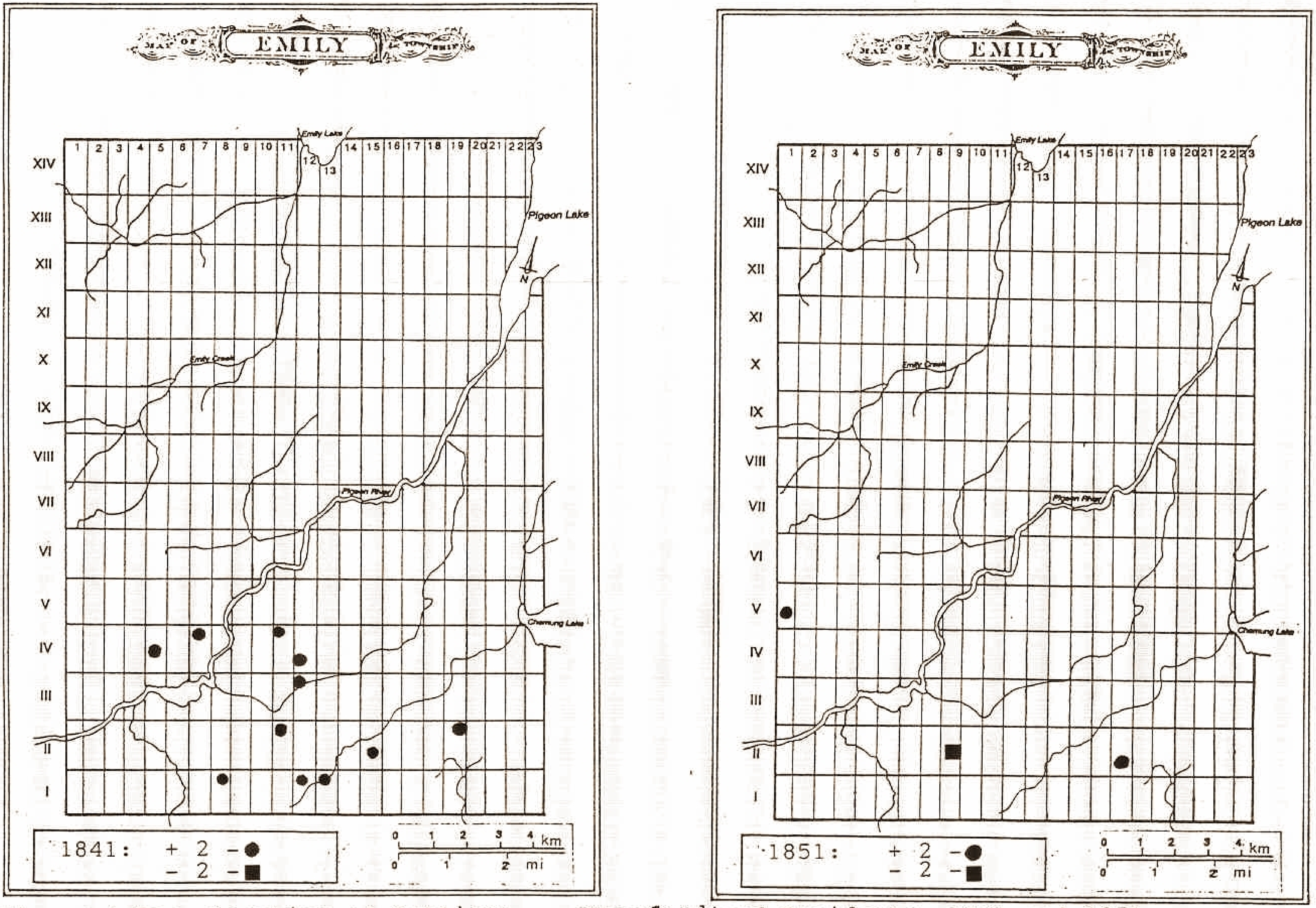


Figure 4.51 : Location of Outliers - Standardized Residuals 1841 and 1851