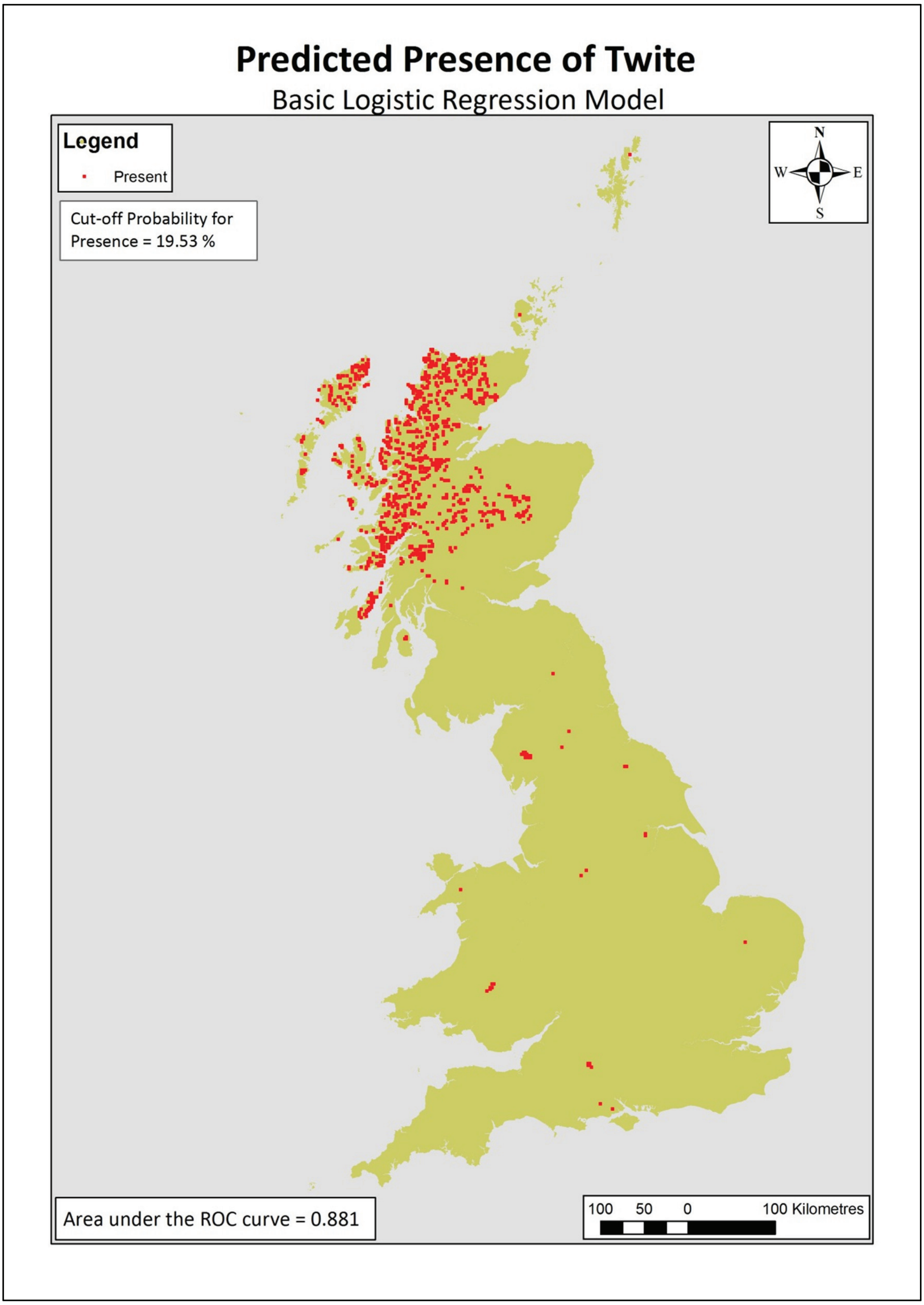

Figure 27 – Predicted distribution of twite (basic model)

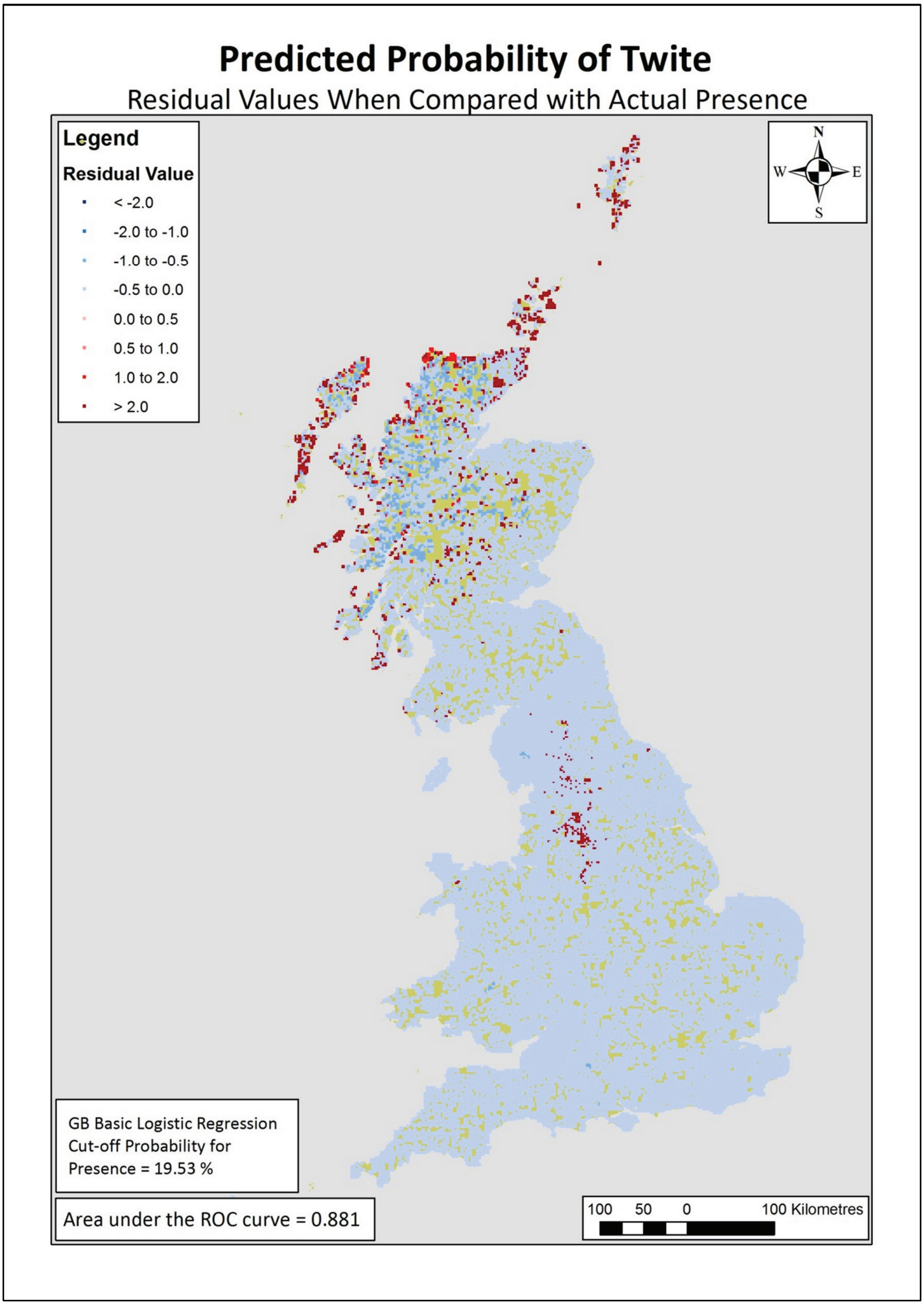

Figure 28 – Map showing residuals for predicted twite probability versus actual distribution

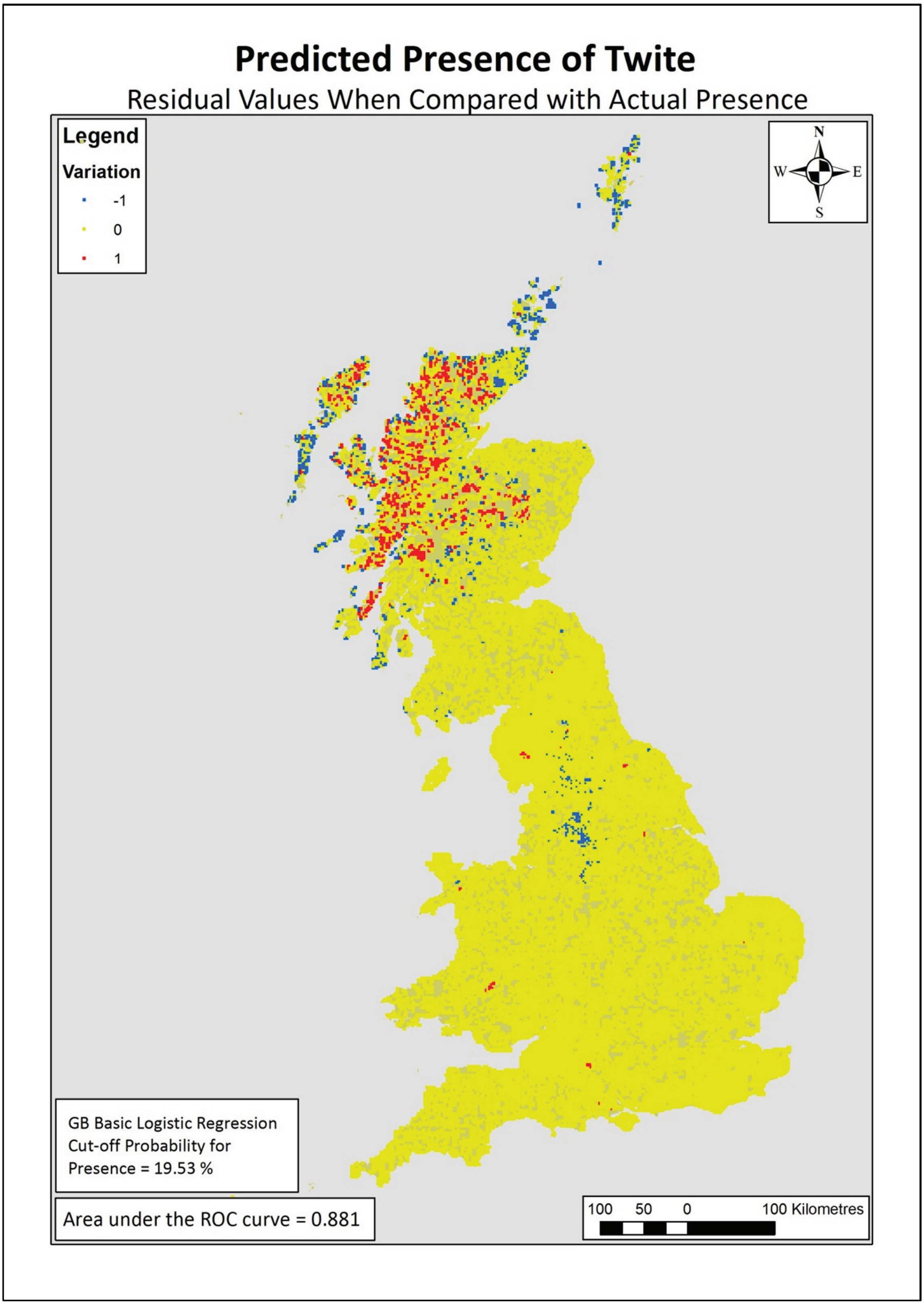

Figure 29 – Map showing residuals for predicted twite distribution versus actual distribution

3.4 Basic Model Plus Elevation

This result suggests that there are varying relationships between twite and land cover and land use across Britain. In order to try and address this, another attempt was made at developing a model but elevation data was included in the logistic regression. This analysis produced a model represented by equation 3.2

$$P_{DEM} = \frac{1}{1 + e^{-\left(\begin{array}{c}-4.132 + 4.058 \times M - 0.020 \times C + 4.093 \times Pt + 4.241 \times NG \\ + 2.485 \times P - 0.001 \times Ca - 0.001 \times Sh - 0.005 \times E\end{array}\right)}} \qquad (3.2)$$

where, the M = % cover of moor & heathland (CLC90), C = area of crops (ha) per tetrad, Pt = % cover of peatland (CLC90), NG = % cover of natural grassland (CLC90), P = % cover of pasture (CLC90), Ca = number of cattle per tetrad (AgC88), Sh = number of sheep per tetrad (AgC88), E = elevation (DEM) and the constant = -4.132. This model produces the goodness of fit results shown in Table 9. The score for the area under the ROC curve was 0.901 which indicates an improved performance compared with the basic model, statistically at least.

Statistic	Value
N_p (Number of parameters in the model)	9
-2LL	6609
Cox & Snell R Square	0.078
Nagelkerke R Square	0.313
Hosmer & Lemeshow Chi-square	96.123
AIC (-2LL + $2N_p$)	6627
Area Under ROC	**0.901**

Table 9 - Goodness of Fit statistics for the Basic Model plus Elevation (Britain)

Maps showing the predicted probability and distribution of twite tetrads across Britain generated by the ‘Basic plus Elevation’ model are presented in Appendix 6. These maps show a similar broad pattern of distribution although the range of probabilities is higher, with a maximum of 45.3 %. In this case a revised cut-off value of 24.58 % was applied to match the actual number of twite tetrads, producing the predicted distribution shown in Appendix 6. This map appears to contradict the statistical improvement of the model as represented by the ROC values. The revised model predicts almost no tetrads with twite in England and Wales, and those it does predict are in unexpected locations. Also, there are very few twite predicted in the central Highlands of Scotland. The residual values for Basic plus Elevation' model, which emphasise these inaccuracies, are also presented at Appendix 6.

These results suggest there must be different relationships between twite distribution and elevation in different parts of Britain, and taking the whole of Britain together and applying a simple logistic regression model has resulted in the model predicting twite in coastal and low-lying areas, probably because the majority of twite are found in such locations in the north and west of the country. Hence, it was considered that a more in depth approach is obviously required, perhaps involving climatic data as described in the following section.

3.5 Basic Model Plus Elevation & Climate

This analysis produced a model represented by equation 3.3

$$P_{E\&C} = \frac{1}{1 + e^{-\left(\begin{array}{c} 7.233 + 1.284 \text{ x } M - 0.804 \text{ x } T_{Jul} + 3.219 \text{ x } C_{Ave} \\ - 0.870 \text{ x } NG_{Ave} - 0.412 \text{ x } Pt_{Ave} - 0.013 \text{ x } R_{Ave} - 0.007 \text{ x } E \end{array}\right)}} \qquad (3.3)$$

where, the M = % cover of moor & heathland (CLC90), T_{Jul} = LTA Temperature for July, C_{Ave} = Area of crops above or below British average, NG_{Ave} = % cover of natural grassland above or below British average (CLC90), Pt_{Ave} = % cover of peatland above or below British average (CLC90), R_{Ave} = LTA Monthly Rainfall averaged over the breeding season, E = elevation (DEM) and the constant = 7.233. This model produces the goodness of fit results shown in Table 10. The score for the area under the ROC curve was 0.909 indicating a further enhancement to the performance of the model compared with the basic model.

Statistic	Value
N_p (Number of parameters in the model)	9
-2LL	6309
Cox & Snell R Square	0.086
Nagelkerke R Square	0.348
Hosmer & Lemeshow Chi-square	26.438
AIC (-2LL + $2N_p$)	6320
Area Under ROC	**0.909**

Table 10 - Goodness of Fit statistics for the Basic Model plus Elevation & Climate (Britain)

Maps showing the predicted probabilities and distribution of twite as generated by the 'Basic plus Elevation & Climate' model are presented at Appendix 6. The predicted probability of twite is visibly low for the Pennines but looks better for Scotland. The predicted distribution supports this observation and emphasises once again the lack of twite tetrads being predicted in England. The residual values for the 'Basic plus Elevation & Climate' model (Appendix 6) further emphasise the areas where this model performs poorly.

Overall this does look like an improvement on the 'Basic plus Elevation' model but it is obviously still not predicting distribution of twite very well for the whole country. With this in mind, it was decided to divide the country into geographic zones and develop individual models for each zone, then combine them to derive another, hopefully improved, predicted distribution. It was anticipated that by applying a geographically zoned approach to the model apparent differences in relationships between twite distribution and some parameters might be addressed.

3.6 Geographic Zonation Model

This model comprised three separate parts that were combined to produce predicted probability and distribution of twite across Britain. The three geographic zone models can be described by equations 3.4 -3.6 for N&W Scotland, S&C Scotland and England & Wales respectively.

$$P_{NWSco} = \frac{1}{1 + e^{-\left(\begin{array}{c} 4.775 - 0.309 \text{ x } T_{Aug} - 0.019 \text{ x } R_{Aug} \\ - 0.007 \text{ x } E + 0.068 \text{ x } F - 1.478 \text{ x } Con \end{array}\right)}} \qquad (3.4)$$

where, the T_{Aug} = LTA Temperature for August, R_{Aug} = LTA Monthly Rainfall for August, E = elevation (DEM), F = area (ha) of bare fallow per tetrad (AgC88), Con = % cover of coniferous forest (CLC90), and the constant = 4,775.

$$P_{SCSco} = \frac{1}{1 + e^{-(-3.790 - 0.999 \text{ x } M_{TetAve} - 0.497 \text{ x } CLCAgC_{Ave})}} \qquad (3.5)$$

where, the M_{TetAve} = % cover of moor & heathland above or below the twite tetrad average (CLC90), $CLCAgC_{Ave}$ = Sum of the average comparisons of the eight CLC90 and AgC88 variables, and the constant = -3.790.

$$P_{E\&W} = \frac{1}{1 + e^{-(-5.421 + 2.747 \times M - 2.805 \times C_{TetAve} + 3.243 \times NG)}} \quad (3.6)$$

where, the M = % cover of moor & heathland (CLC90), C_{TetAve} = Area of crops above or below twite tetrad average, NG = % cover of natural grassland (CLC90), and the constant = -5.421.

This model produces the goodness of fit results shown in Table 11. The scores vary considerably between the different zones. The model appears to perform noticeably better for England & Wales than it does for the two Scottish zones. The combined model performed relatively well with an area under the ROC curve was 0.932 which indicates yet another enhancement to the performance compared with the previous models. Figure 30 show the predicted probability and predicted presence of twite based on the Geographic Zones model.

Statistic	N&W Sco	S&C Sco	E & W
N_p (Number of parameters in the model)	3	3	4
-2LL	3120	925	1342
Cox & Snell R Square	0.128	0.028	0.022
Nagelkerke R Square	0.203	0.138	0.288
Hosmer & Lemeshow Chi-square	7.763	4.459	10.155
AIC (-2LL + $2N_p$)	3126	931	1350
Area Under ROC	**0.757**	**0.795**	**0.933**

Table 11 - Goodness of Fit statistics for the Geographic Zones Model (Britain)

The predicted distribution for England & Wales is much improved with this model, but there are some clear gaps in Scotland, most notably the north-central Highlands. This is reflected in the area under the ROC curve values for the component zones in the model. Overall, the results produce an area under the ROC curve value of 0.938 for the predicted probability of twite. When the ROC value is determined for the predicted distribution it is not so impressive, being only 0.649, but this compares favourably with the other models. The residual values for this model are presented in Figure 32 illustrating clearly the improvement in the model performance as well as the areas where the model if deficient. There is an over prediction in Shetland, Orkney and Caithness, and in parts of South Uist, as well as to the north-west of the actual Pennines population. At the same time there are under-estimates along the north coast of Scotland, in the central Highlands and in the South Pennines. Figure 32 shows numerous areas where the standardized residual values are in excess of +/- 2.0. This suggests that these predictions are outliers, i.e. significantly different from the actual value.

The main concern about this approach to modelling twite distribution is that the boundaries of the geographic zones is to some extent arbitrary, being based on District boundaries. As mentioned earlier this does raise concern over the MAUP issue: are the relationships being modelled genuine relationships between twite and their environment, or merely artefacts of the arbitrary boundaries. This is not a serious issue perhaps due to the selection of zones being partly based on output from previous models, but it was decided to try elevation zones as an alternative approach.

3.7 Elevation Zonation Model

This model comprised two separate parts that were combined to produce predicted probability and distribution of twite across Britain. The two elevation zone models can be described by equations 3.7 - 3.8 for areas under 200 m above sea level and areas greater than or equal to 200 m above sea level respectively.

$$P_{LT200} = \frac{1}{1 + e^{-\left(\begin{array}{c}-9.506 + 1.832 \text{ x } M - 0.964 \text{ x } T_{Jul} + 3.051 \text{ x } Pt \\ + 1.841 \text{ x } NG - 0,015 \text{ x } C + 2.280 \text{ x } P - 0.001 \text{ x } Sh\end{array}\right)}} \qquad (3.7)$$

where, the M = % cover of moor & heathland per tetrad (CLC90), T_{Jul} = LTA Temperature for Jul, Pt = % cover of peatlands per tetrad (CLC90), NG = % cover of natural grasslands per tetrad (CLC90), C = area (ha) of crops per tetrad (AgC88), P = % cover of pasture (CLC90), Sh = number of sheep per tetrad (AgC88), and the constant = -9.506.

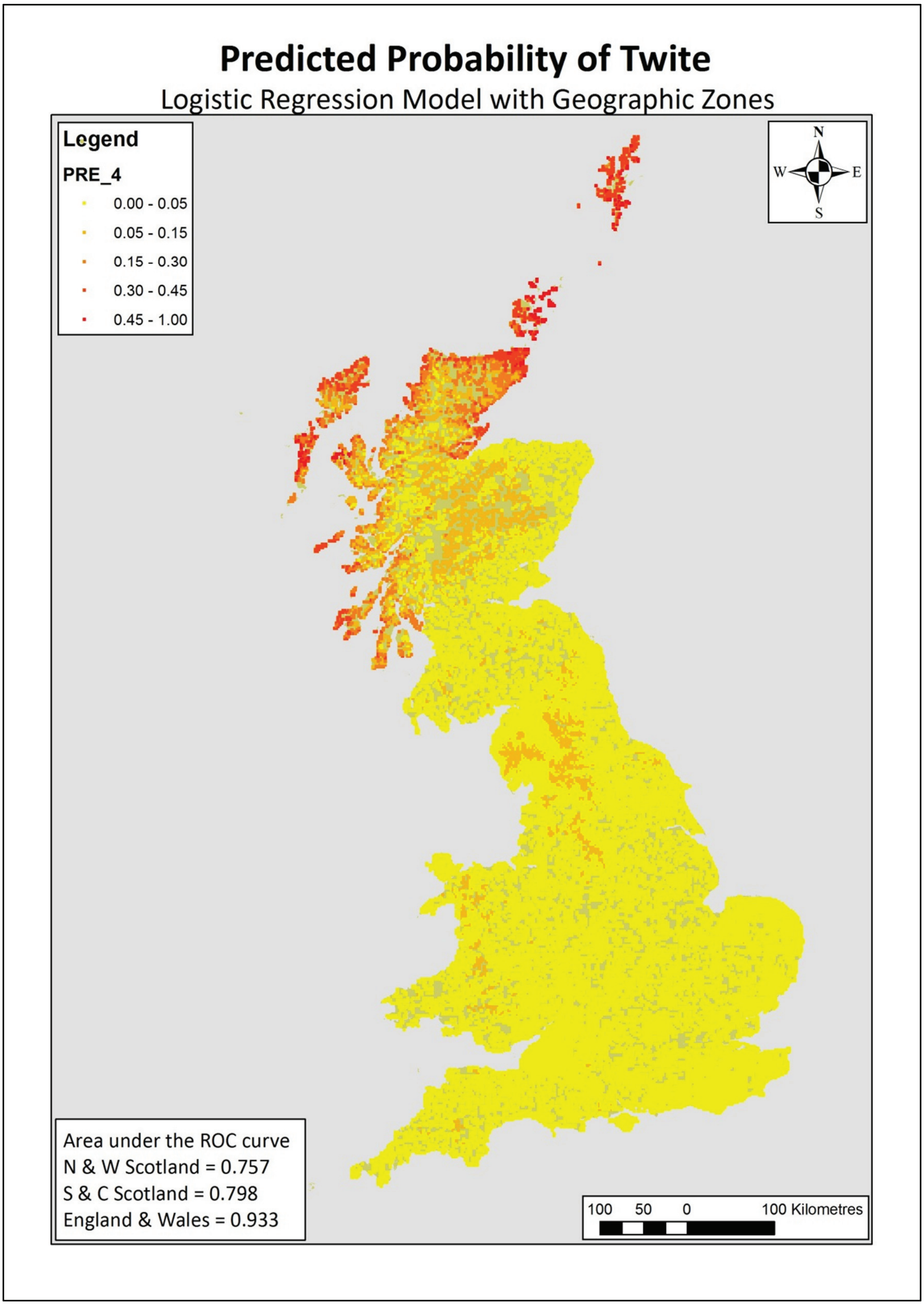

Figure 30 - Predicted probability of twite presence (geographic zones model)