

## “LOOKING FOR BVER INTO THE GRASS” DATASET

E. Croce<sup>a</sup>, A. De Lisi<sup>a</sup>

*a. Università degli Studi di Milano, Dipartimento di Beni Ambientali e Culturali.*

Here we present the results of the modelling process discussed in the paper “*Looking for BVER into the GRASS*”, *database pubblici e modelli predittivi in archeologia: il progetto BVER in dialogo con GNA*.

The work investigates the settlement dynamics of the lower Verbano area between the Iron Age and the Middle Ages. The known archaeological sites, obtained from the public database GNA (Geoportale Nazionale per l'Archeologia), together with a series of physical characteristics of the area, were used as predictors for the calculation of an inductive predictive model. The results show the shortcomings of this methodology in the field of predictive archaeology.

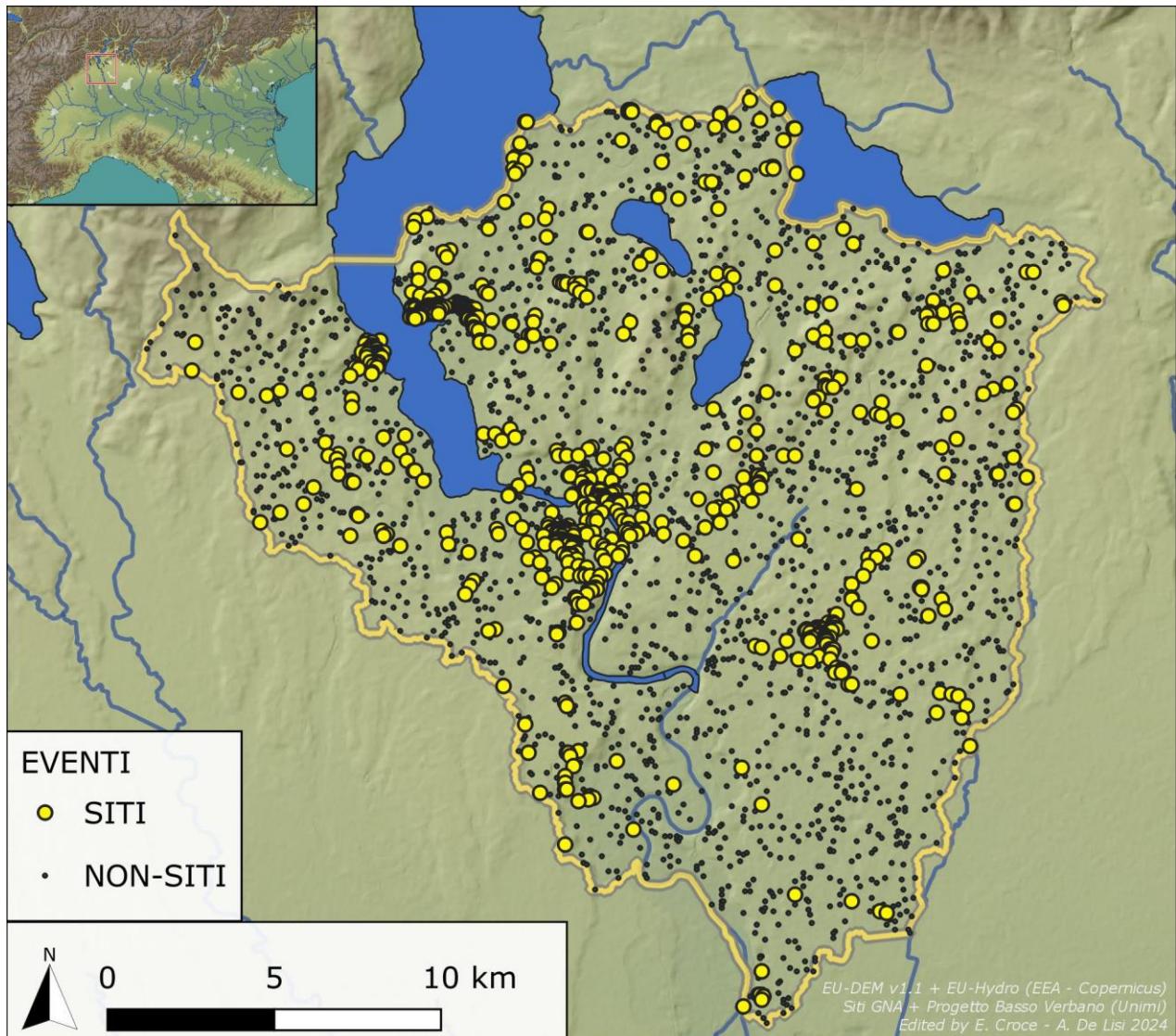
The model constructed for the general record of archaeological sites proves to be underperforming from a predictive point of view, while heterogeneous results were obtained when the site sample was chronologically sectioned. The model calculated for the Iron Age shows a high discriminatory power and, when compared with the models for subsequent periods, underlines the importance of the choice of variables for this methodological approach.

The use of public data (GNA) as variables revealed some shortcomings in the prediction of complex territorial patterns. However, it also highlighted the intrinsic potential of such a tool to study the relationship between the environment and the archaeological record. A potential we hope to develop further in the future.

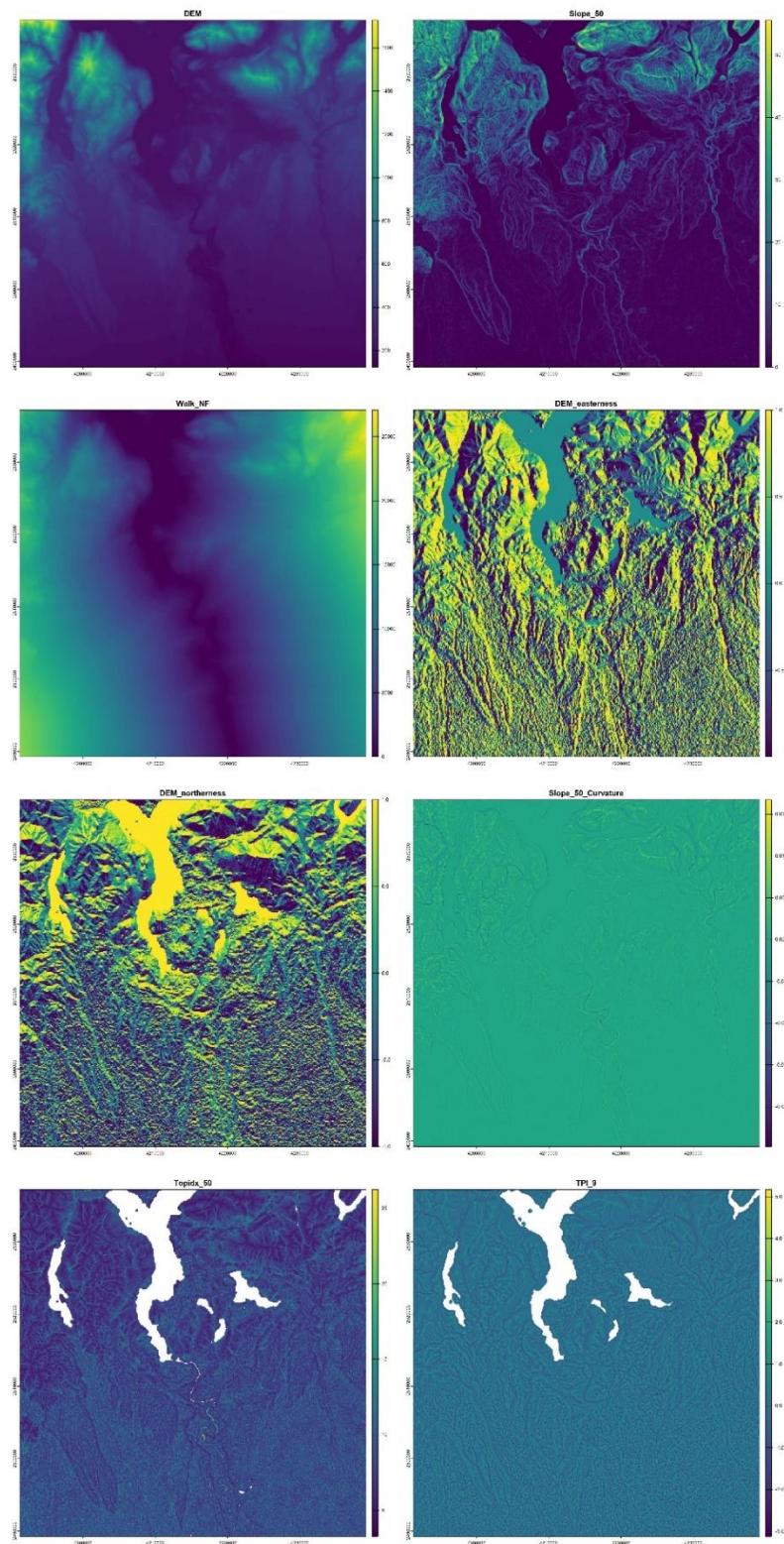
From a practical point of view, the operational protocol defined in: CROCE E., CARRER F., 2024, *SBC Predictive Model (v1.0.1)* - Dataset. <https://doi.org/10.5281/zenodo.14245171> was used, omitting the ethnoarchaeological premise that characterised it (CARRER F., 2013, *An ethnoarchaeological inductive model for predicting archaeological site location: a case-study of pastoral settlement patterns in the Val di Fiemme and Val di Sole (Trentino, Italian Alps)*, Journal of Anthropological Archaeology, 32, pp. 54-62. <https://doi.org/10.1016/j.jaa.2012.10.001>), given the need to approach strictly archaeological data, without possible local ethnographic comparisons.

## INDUCTIVE PREDICTIVE MODEL

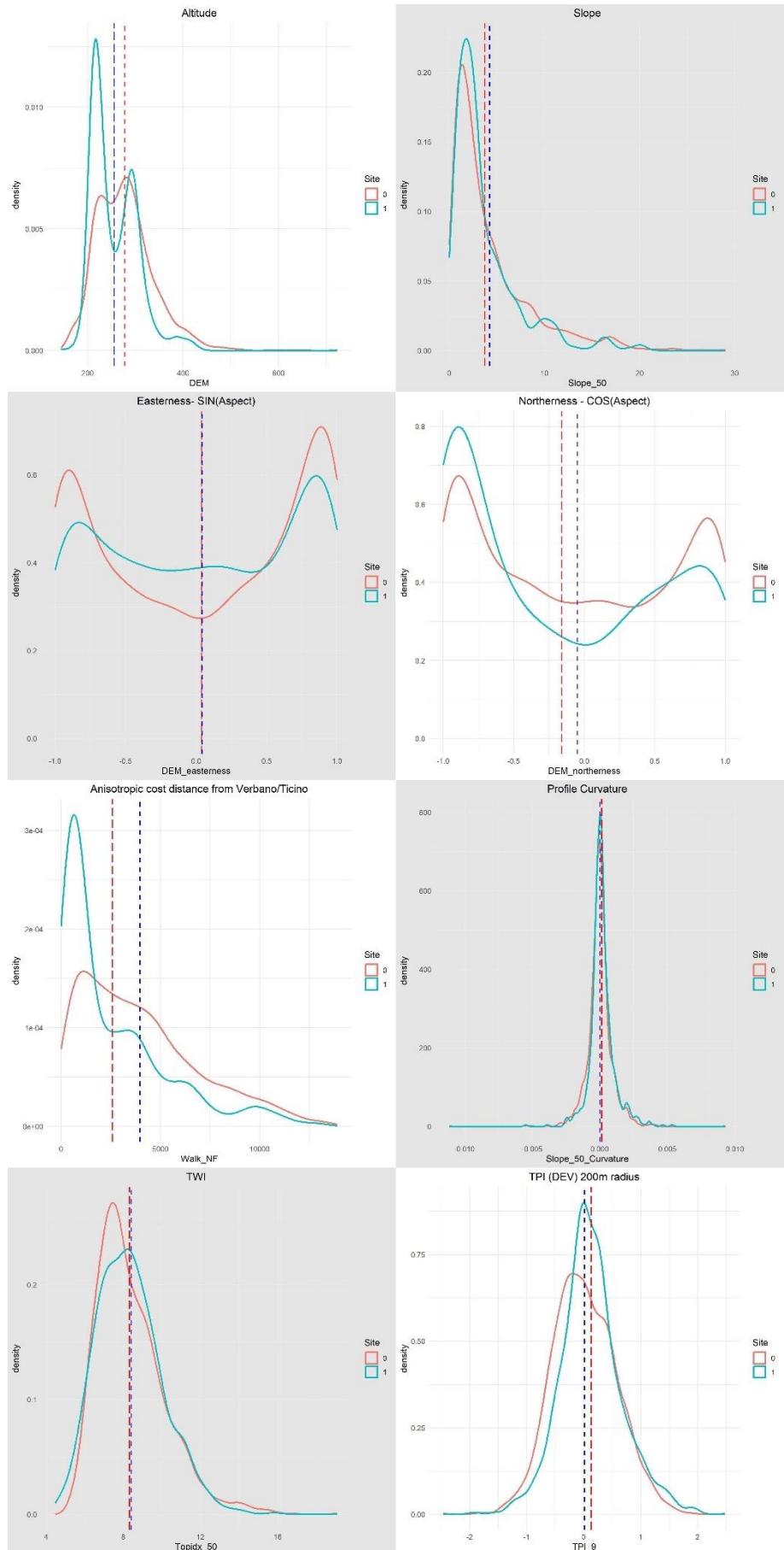
### Area, events, and predictors



Positioning of the Basso Verbano project study area and the record of events (sites and non-sites) used as the binary dependent variable for the multivariate logistic regression calculation underlying the predictive model.

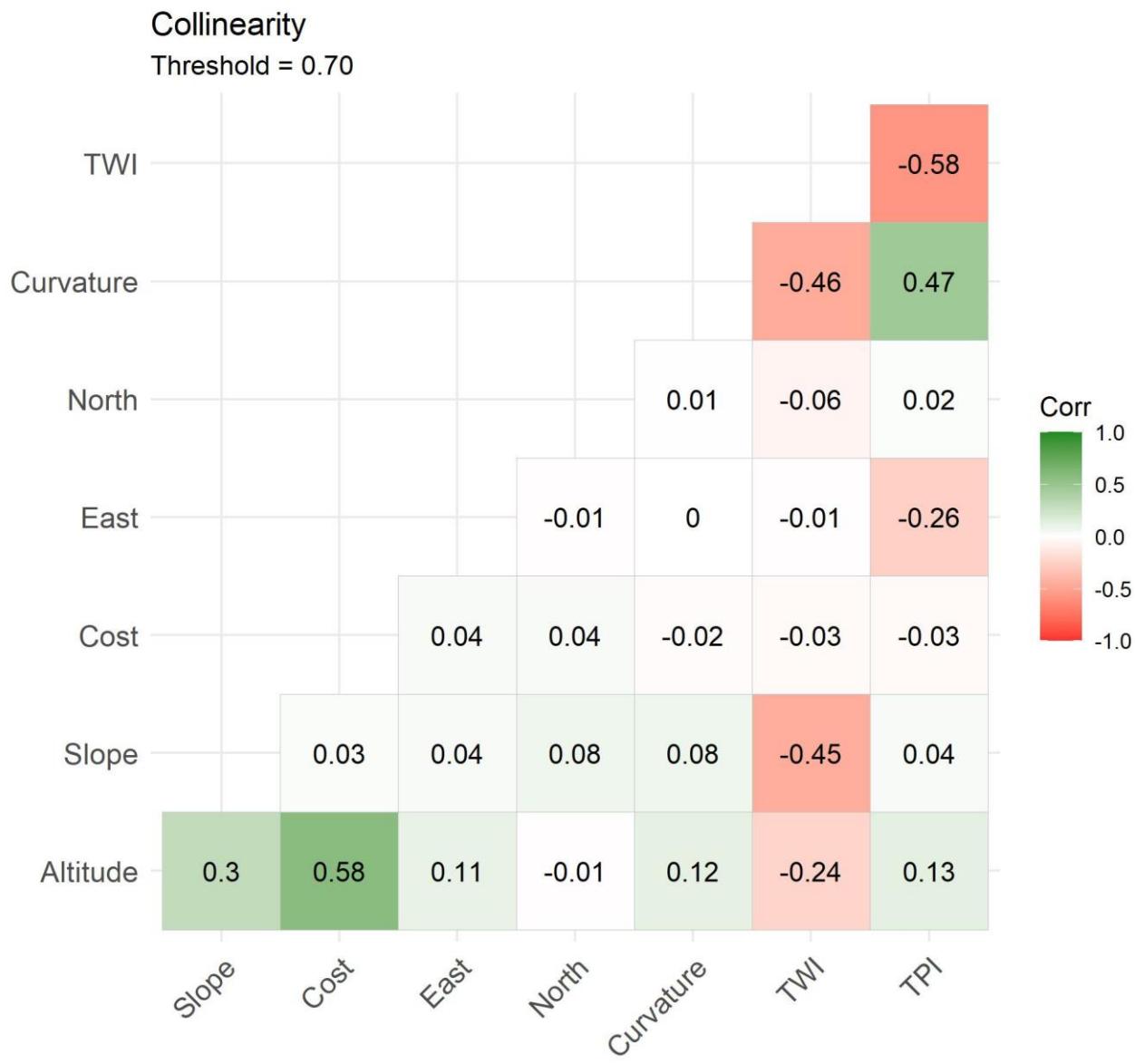


Graphical representation of the independent variables used in the calculation of the multivariate logistic regression. The raster maps were produced in the GRASS GIS environment using the European Digital Elevation Model digital terrain model (EU-DEM v1.1, European Environment Agency (EEA) - Copernicus programme (<https://land.copernicus.eu/>)).



Density plot of variable values for sites and no-sites

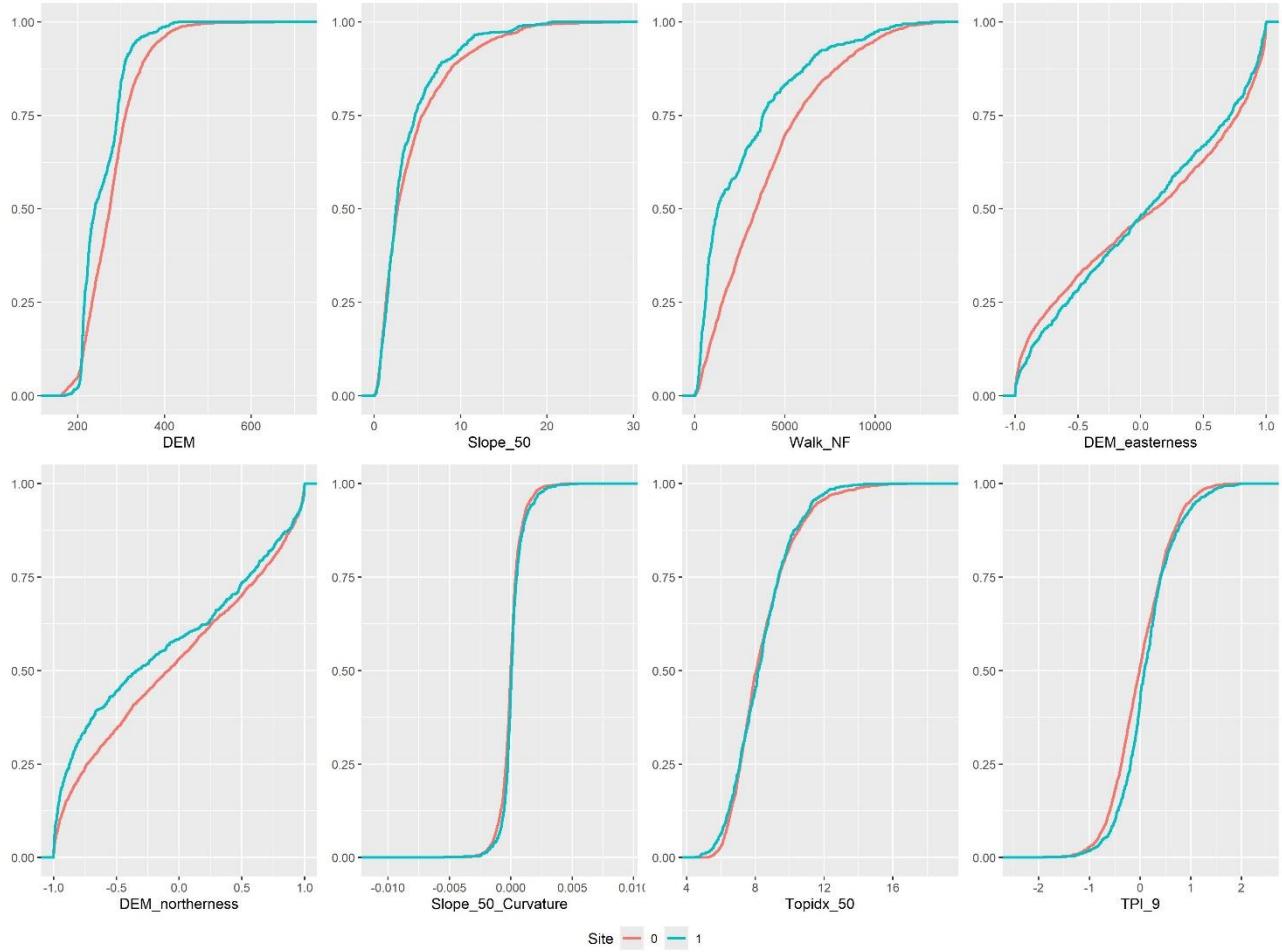
## Collinearity



Result of the collinearity calculation for the independent variables used for the multivariate logistic regression calculation. The threshold value was set at 0.70 based on: Alberti G., Grima R., Vella N., 2018, The use of geographic information system and 1860s cadastral data to model agricultural suitability before heavy mechanization. A case study from Malta, PLoS ONE, 13.

<https://doi.org/10.1371/journal.pone.0192039>

## ECD & Kolmogorov-Smirnov Test



Empirical Cumulative Distribution of each predictor (sites and no-sites).

## Kolmogorov-Smirnov Test results (Bootstrapping 10000)

Threshold p-value: 0.05

### Altitude

\$DEM  
Bootstrap p-value: **< 2.22e-16**  
Naive p-value: 2.4643e-21  
Full Sample Statistic: 0.21398

### Northerness

\$DEM\_northerness  
Bootstrap p-value: **< 2.22e-16**  
Naive p-value: 1.7945e-06  
Full Sample Statistic: 0.11507

### Slope

\$Slope\_50  
Bootstrap p-value: **0.0026**  
Naive p-value: 0.0033409  
Full Sample Statistic: 0.077984

### Profile Curvature

\$Slope\_50\_Curvature  
Bootstrap p-value: **4e-04**  
Naive p-value: 0.00049009  
Full Sample Statistic: 0.088921

### Walk Distance From Ticino

\$Walk\_NF  
Bootstrap p-value: **< 2.22e-16**  
Naive p-value: 6.2616e-42  
Full Sample Statistic: 0.30147

### Topographic Wetness Index

\$Topidx\_50  
Bootstrap p-value: **0.1784**  
Naive p-value: 0.18541  
Full Sample Statistic: 0.047551

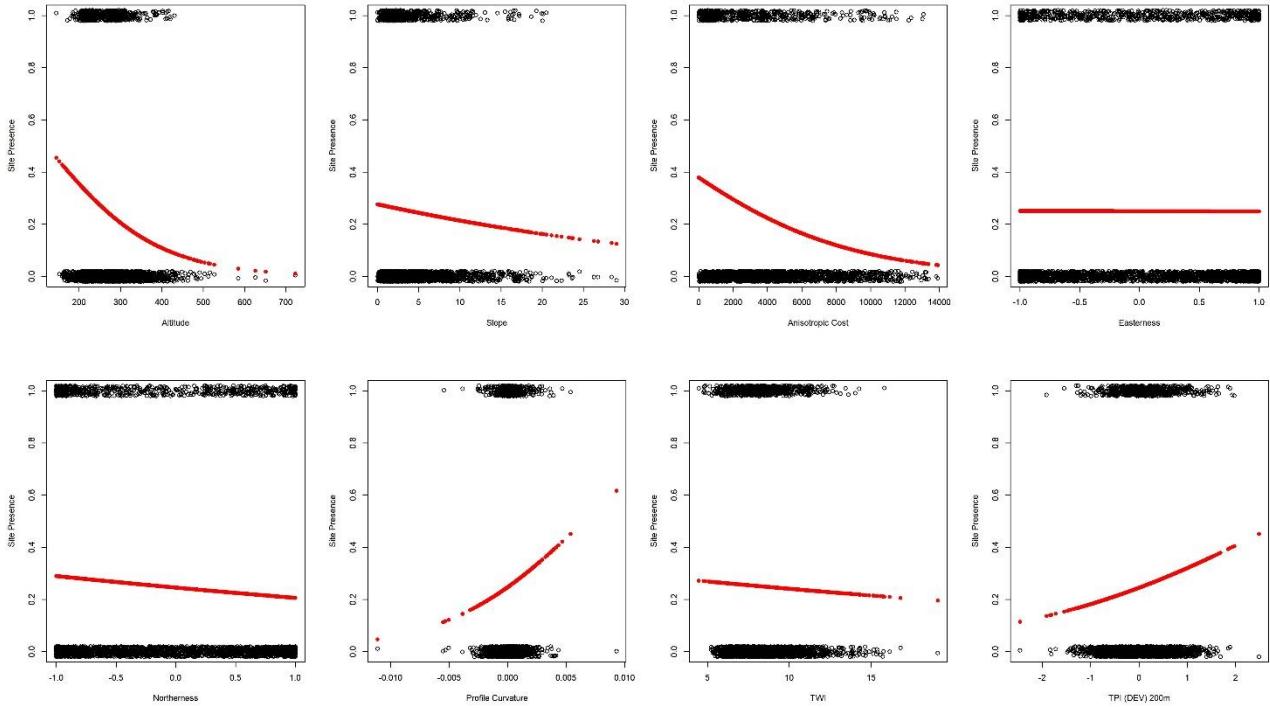
### Easterness

\$DEM\_easterness  
Bootstrap p-value: **0.1357**  
Naive p-value: 0.13826  
Full Sample Statistic: 0.050404

### Topographic Position Index

\$TPI\_9  
Bootstrap p-value: **< 2.22e-16**  
Naive p-value: 5.4522e-09  
Full Sample Statistic: 0.13695

## Univariate Logistic Regression



Univariate logistic regression plot for each predictor. The univariate approach was usually used to better understand the relationship between the dependent variable and each covariate (Carrer 2013), in order to avoid bias due to non-linear relationships. As we did not find such problems in the current case study, we decided to remove this step from the workflow presented in the paper.

## Multivariate Logistic Regression

### Standard

```
Call:  
glm(formula = name ~ DEM + slope_50 + walk_NF + DEM_northerness +  
    slope_50_Curvature + TPI_9, family = binomial(logit), data =  
    tab)  
  
Coefficients:  
              Estimate Std. Error z value Pr(>|z|)  
(Intercept) 2.807e-01 2.531e-01 1.109 0.267516  
DEM         -3.409e-03 1.145e-03 -2.979 0.002896 **  
Slope_50     -1.906e-02 1.239e-02 -1.539 0.123771  
Walk_NF      -1.394e-04 2.316e-05 -6.017 1.77e-09 ***  
DEM_northerness -2.196e-01 6.462e-02 -3.398 0.000679 ***  
Slope_50_Curvature 1.101e+02 5.601e+01 1.966 0.049304 *  
TPI_9        3.280e-01 9.100e-02 3.605 0.000312 ***  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
Null deviance: 3153.6 on 2803 degrees of freedom  
Residual deviance: 2976.3 on 2797 degrees of freedom  
AIC: 2990.3  
  
Number of Fisher Scoring iterations: 4
```

### Backward Stepwise Selection (AIC)

```
Call:  
glm(formula = name ~ DEM + Slope_50 + Walk_NF + DEM_northerness +  
    Slope_50_Curvature + TPI_9, family = binomial(logit), data =  
    tab)  
  
Coefficients:  
              Estimate Std. Error z value Pr(>|z|)  
(Intercept) 2.807e-01 2.531e-01 1.109 0.267516  
DEM         -3.409e-03 1.145e-03 -2.979 0.002896 **  
Slope_50     -1.906e-02 1.239e-02 -1.539 0.123771  
Walk_NF      -1.394e-04 2.316e-05 -6.017 1.77e-09 ***  
DEM_northerness -2.196e-01 6.462e-02 -3.398 0.000679 ***  
Slope_50_Curvature 1.101e+02 5.601e+01 1.966 0.049304 *  
TPI_9        3.280e-01 9.100e-02 3.605 0.000312 ***  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
Null deviance: 3153.6 on 2803 degrees of freedom  
Residual deviance: 2976.3 on 2797 degrees of freedom  
AIC: 2990.3
```

Number of Fisher Scoring iterations: 4

### Backward Stepwise Selection (Schwartz's AIC - BIC)

```
Call:  
glm(formula = name ~ DEM + Walk_NF + DEM_northerness + TPI_9,  
    family = binomial(logit), data = tab)  
  
Coefficients:  
              Estimate Std. Error z value Pr(>|z|)  
(Intercept) 2.976e-01 2.499e-01 1.191 0.233685  
DEM         -3.802e-03 1.078e-03 -3.526 0.000422 ***  
Walk_NF     -1.350e-04 2.257e-05 -5.980 2.23e-09 ***  
DEM_northerness -2.270e-01 6.431e-02 -3.529 0.000417 ***  
TPI_9        4.136e-01 8.171e-02 5.062 4.15e-07 ***  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
Null deviance: 3153.6 on 2803 degrees of freedom  
Residual deviance: 2982.2 on 2799 degrees of freedom  
AIC: 2992.2
```

Number of Fisher Scoring iterations: 4

### **Standardized Coefficients**

DEM	Cost-Walk	Northerness	Profile Curvature
-0.04495796	-0.08121656	-0.03222255	0.04666688

### **Variance inflation Factor**

#### Variance Inflation Factor (Standard)

DEM	Slope_50	walk_NF	DEM_northerness
1.833781	1.143792	1.674733	1.010768

Slope_50_Curvature	TPI_9
1.295337	1.292550

#### Variance Inflation Factor (AIC)

DEM	Slope_50	walk_NF	DEM_northerness
1.833781	1.143792	1.674733	1.010768

Slope_50_Curvature	TPI_9
1.295337	1.292550

#### Variance Inflation Factor (BIC)

DEM	walk_NF	DEM_northerness	TPI_9
1.627025	1.600141	1.001526	1.032011

## Area under the ROC curve (AUC)

Calculated only for the BIC (Schwartz's AIC) Regression

Call:

```
roc.default(response = BIC_glm$y, predictor =  
BIC_glm$fitted.values)
```

Data: BIC\_glm\$fitted.values in 2103 controls (BIC\_glm\$y 0) < 701  
cases (BIC\_glm\$y 1).

Area under the curve: 0.6661

Discriminatory Ability

(HOSMER et. al. 2013 - *Applied Logistic Regression* - 3rd Ed.)

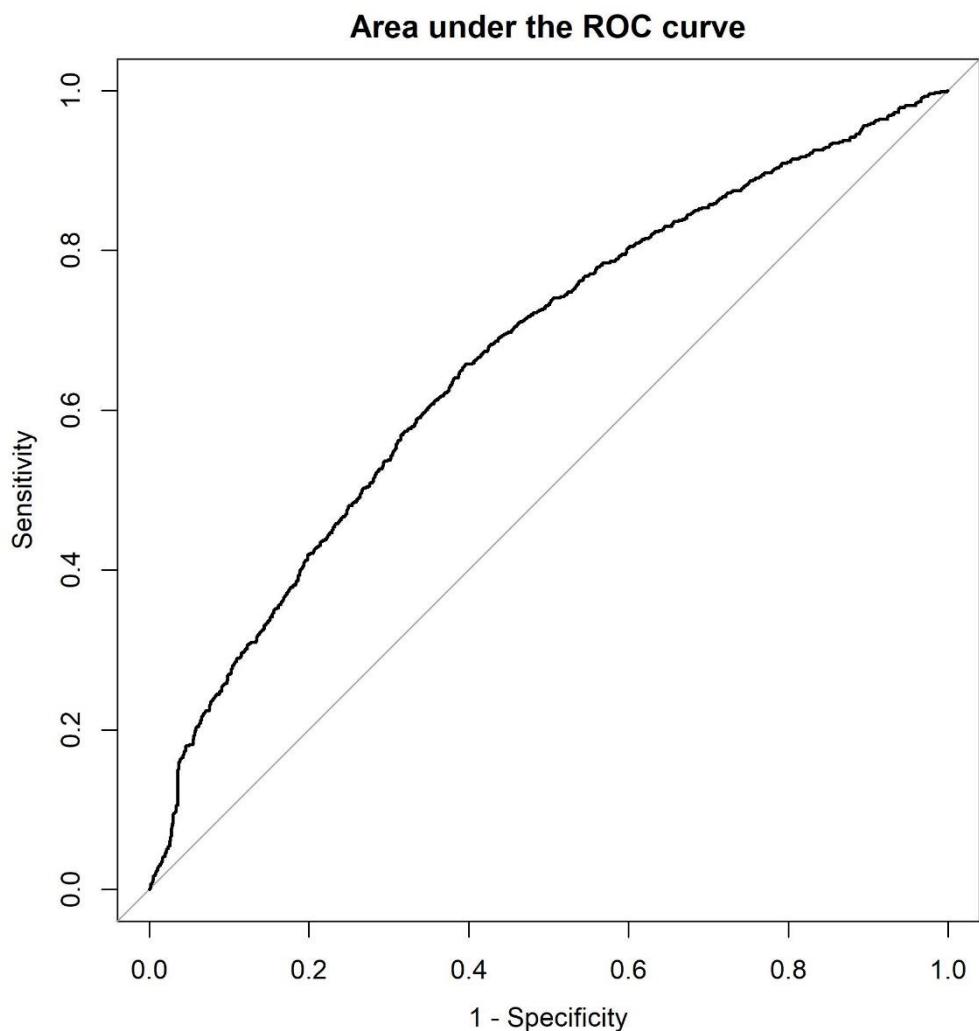
0.5 = No better than chance

0.5-0.7 = Poor

0.7-0.8 = Acceptable

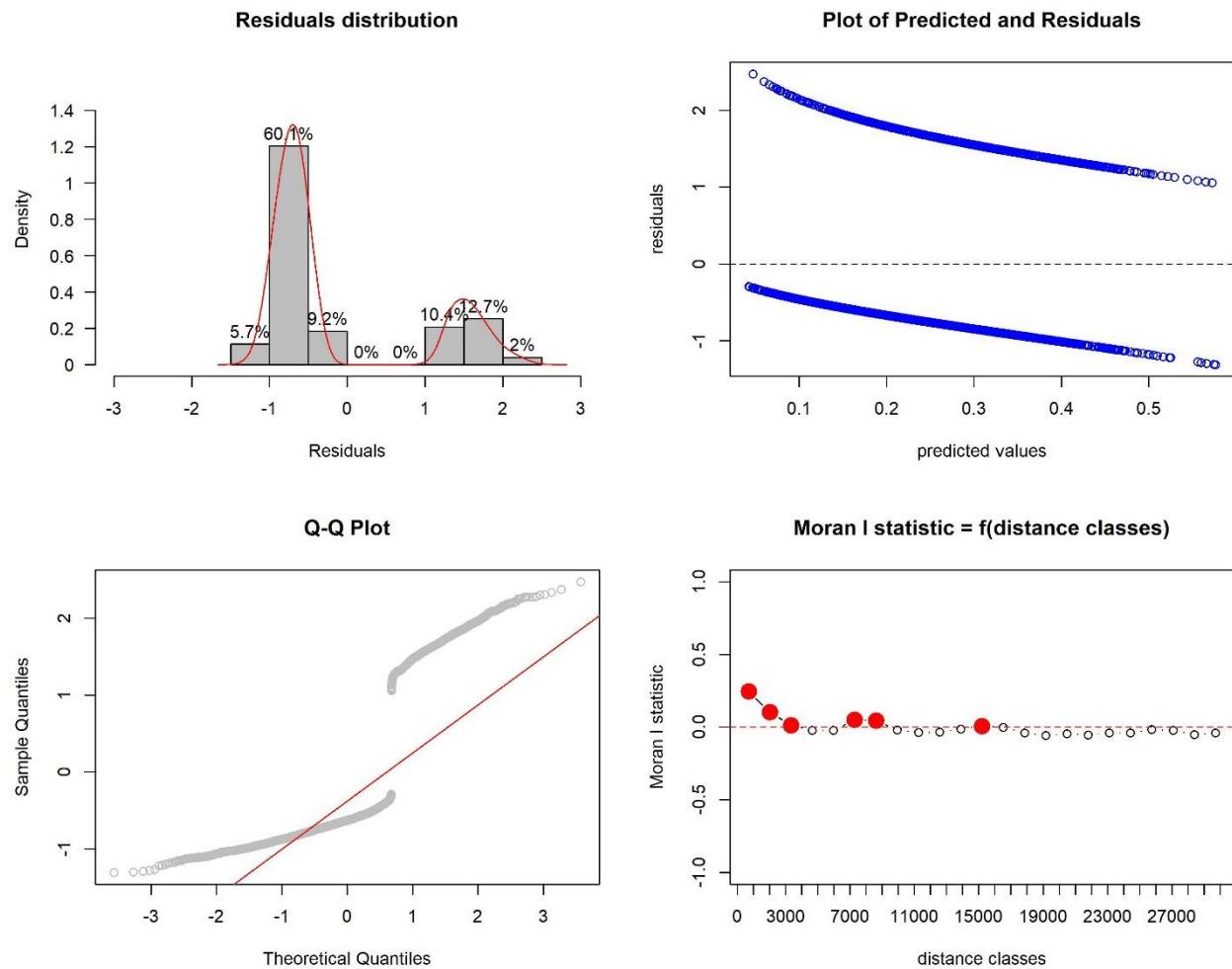
0.8-0.9 = Excellent

0.9-1.0 = Outstanding

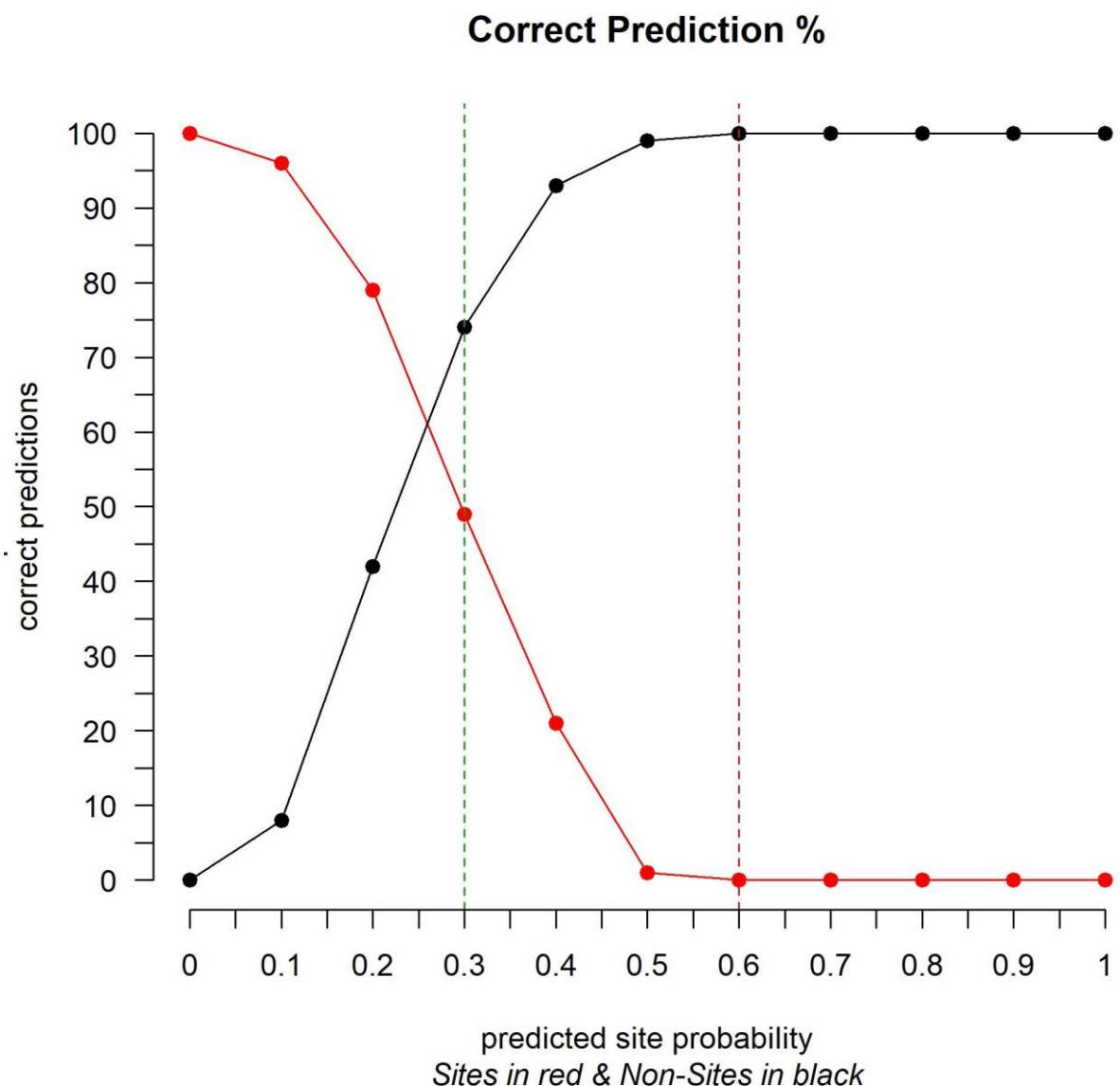


## Residuals

Multivariate Regression (Stepwise BIC) Residuals



### Correct prediction threshold



## Model assessment

### Kolmogorov-Smirnov test results for the predictive values of Sites vs No-Sites

Bootstrap p-value: < 2.22e-16  
Naive p-value: 8.9624e-32  
Full Sample Statistic: 0.26201

### Kvamme's Gain calculation results

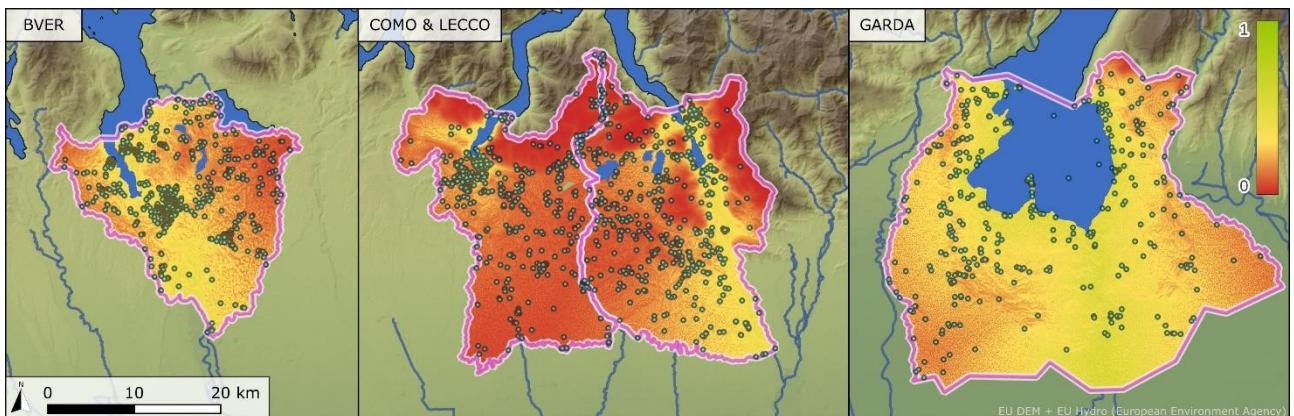
```
-- Kvamme's Gain - Sites in predicted area --
      >0.1 >0.2 >0.3 >0.4 >0.5 >0.6 >0.7 >0.8 >0.9
% Area      0.92 0.60 0.27 0.06 0.01    0    0    0    0
% Siti       0.96 0.79 0.49 0.21 0.01    0    0    0    0
Kvamme's Gain 0.04 0.24 0.45 0.71 0.00   NaN   NaN   NaN   NaN

-- Kvamme's Gain - No Sites in predicted area --
      >0.1 >0.2 >0.3 >0.4 >0.5 >0.6 >0.7 >0.8 >0.9
% Area      0.92 0.60 0.27 0.06 0.01    0    0    0    0
% No Sites   0.92 0.58 0.26 0.07 0.01    0    0    0    0
Kvamme's Gain 0.00 -0.03 -0.04 0.14 0.00   NaN   NaN   NaN   NaN

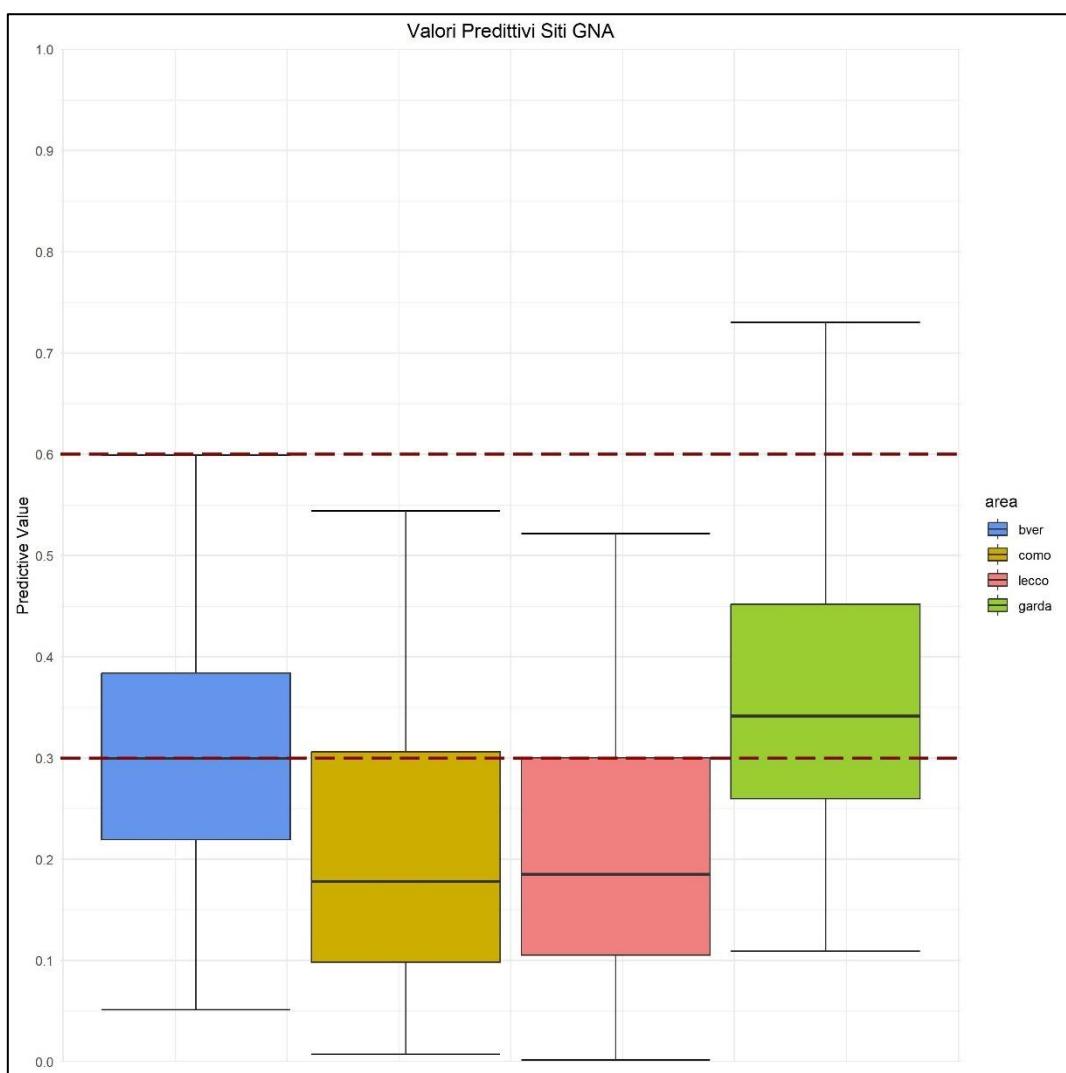
-- Kvamme's Gain - No Sites in unpredicted area --
      >0.1 >0.2 >0.3 >0.4 >0.5 >0.6 >0.7 >0.8 >0.9
% Area Unpred 0.08 0.40 0.73 0.94 0.99    1    1    1    1
% No Sites    0.08 0.42 0.74 0.93 0.99    1    1    1    1
Kvamme's Gain 0.00 0.05 0.01 -0.01 0.00    0    0    0    0
```

## External area test

Testing of the model on external areas, using archaeological sites from the GNA dataset for the area around Como, Lecco, and the lower Garda.



Comparison of BVER and control area predictive surfaces



Comparative plot of BVER and control area predictive values at site locations

-- Kolmogorov-Smirnov test --

h0 = how likely the samples are from same distribution

**BVER/Lecco**

Bootstrap p-value: < 2.22e-16  
Naive p-value: 1.8167e-40  
Full Sample Statistic: 0.37179

**Como/Lecco**

Bootstrap p-value: 0.5991  
Naive p-value: 0.63135  
Full Sample Statistic: 0.043254

**BVER/Como**

Bootstrap p-value: < 2.22e-16  
Naive p-value: 1.4144e-51  
Full Sample Statistic: 0.40144

**Como/Garda**

Bootstrap p-value: < 2.22e-16  
Naive p-value: 4.0998e-52  
Full Sample Statistic: 0.49399

**BVER/Garda**

Bootstrap p-value: < 2.22e-16  
Naive p-value: 1.8072e-09  
Full Sample Statistic: 0.19762

**Lecco/Garda**

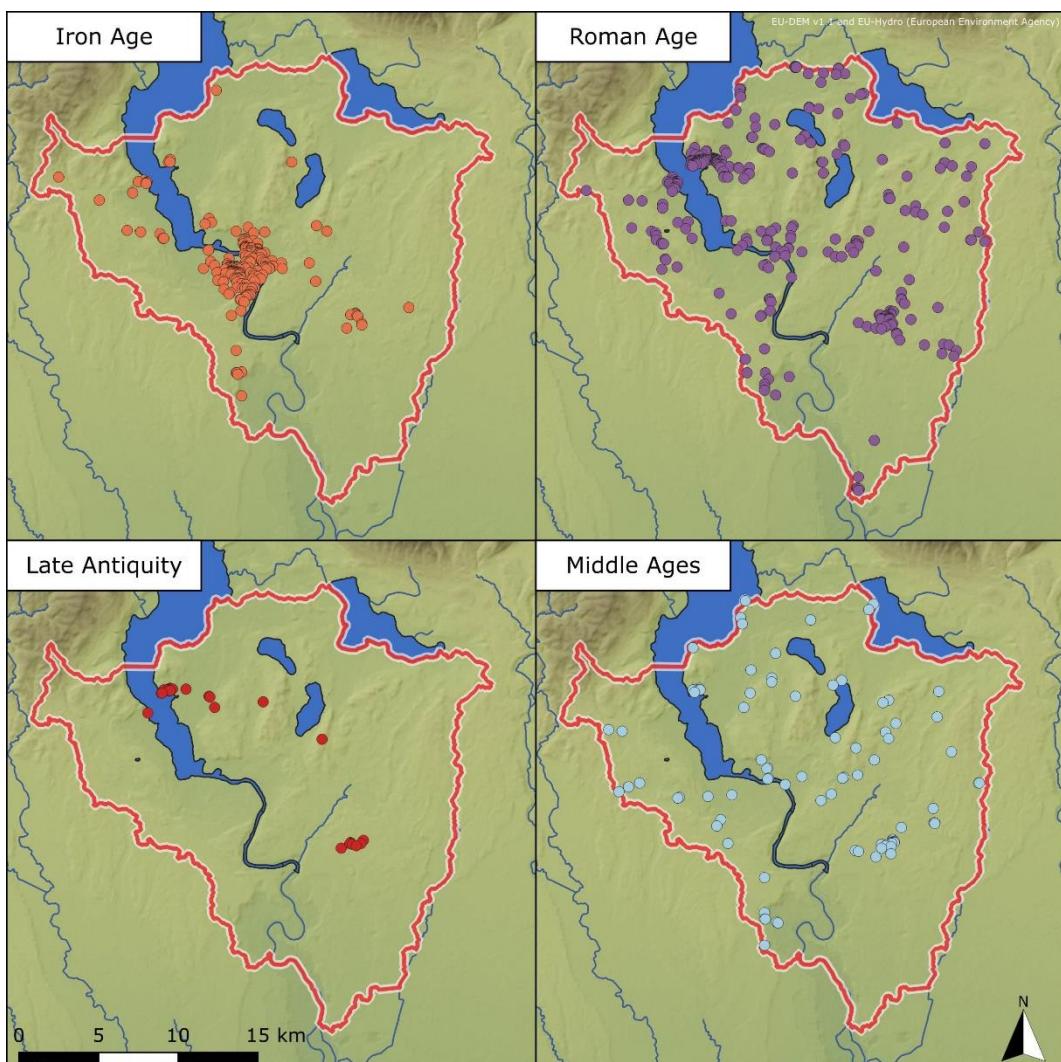
Bootstrap p-value: < 2.22e-16  
Naive p-value: 2.2232e-43  
Full Sample Statistic: 0.46448

## CHRONOLOGICAL SUBSET

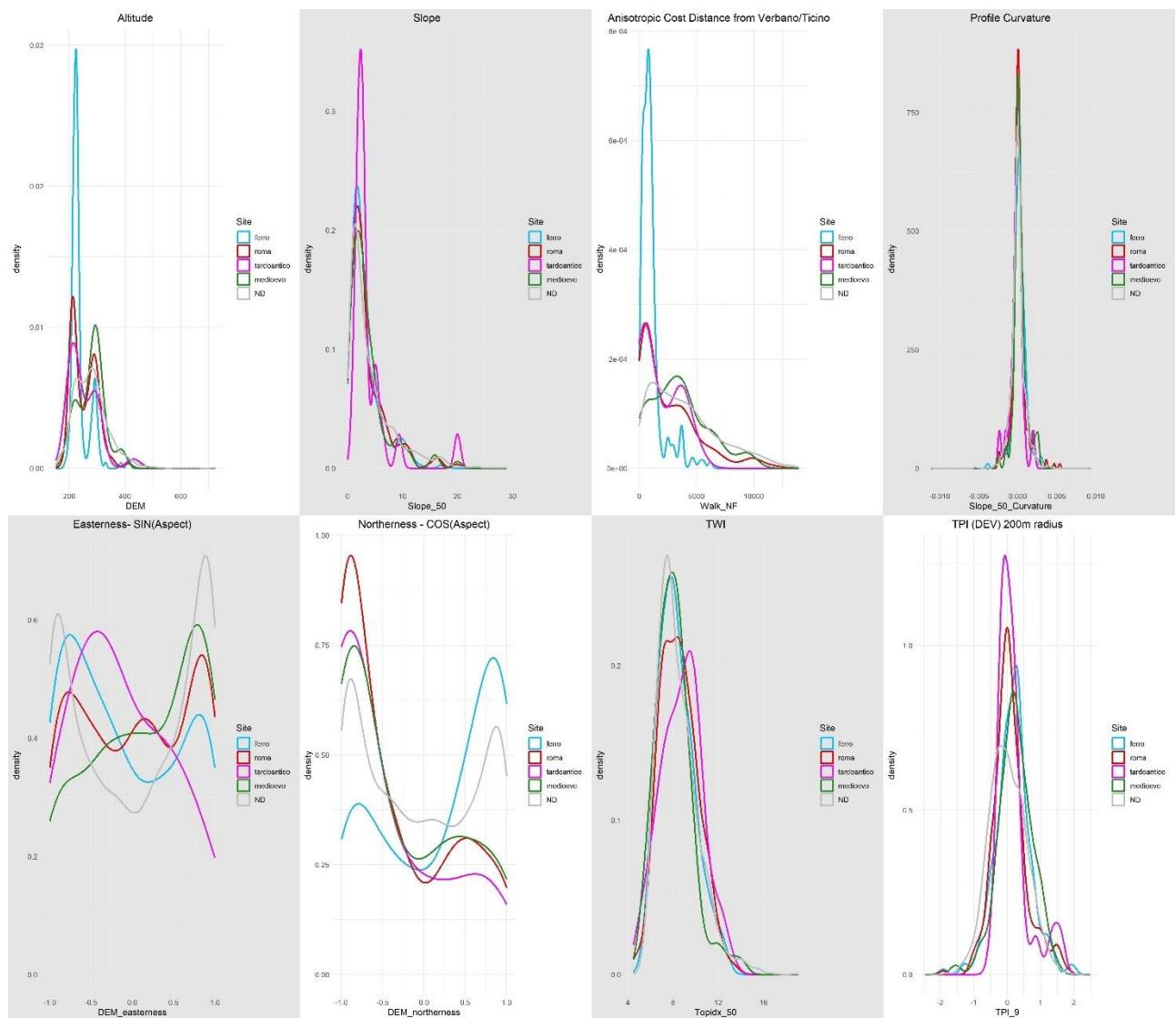
Given the low level performance demonstrated by the model calculated on the archaeological sites of Basso Verbano, it was decided to further analyse the data in order to identify any problems related to the nature of the archaeological record. The dataset used is composed of elements that are heterogeneous in terms of data collection mode, structural type and chronological location. Analysing all the sub-categories that can be generated by cross-filtering the data was beyond the scope of the current research, which only intends to carry out a preliminary analysis of the data and methods that can be used. As a subset criterion, therefore, only chronology was used.

Here we report data on the application of the predictive protocol to sites belonging to the following chronological classes:

- Iron Age (197 sites, 591 no-sites)
- Roman Age (320 sites, 960 no-sites)
- Late Antiquity (24 sites, 72 no-sites)
- Middle Ages (88 sites, 264 no-sites)



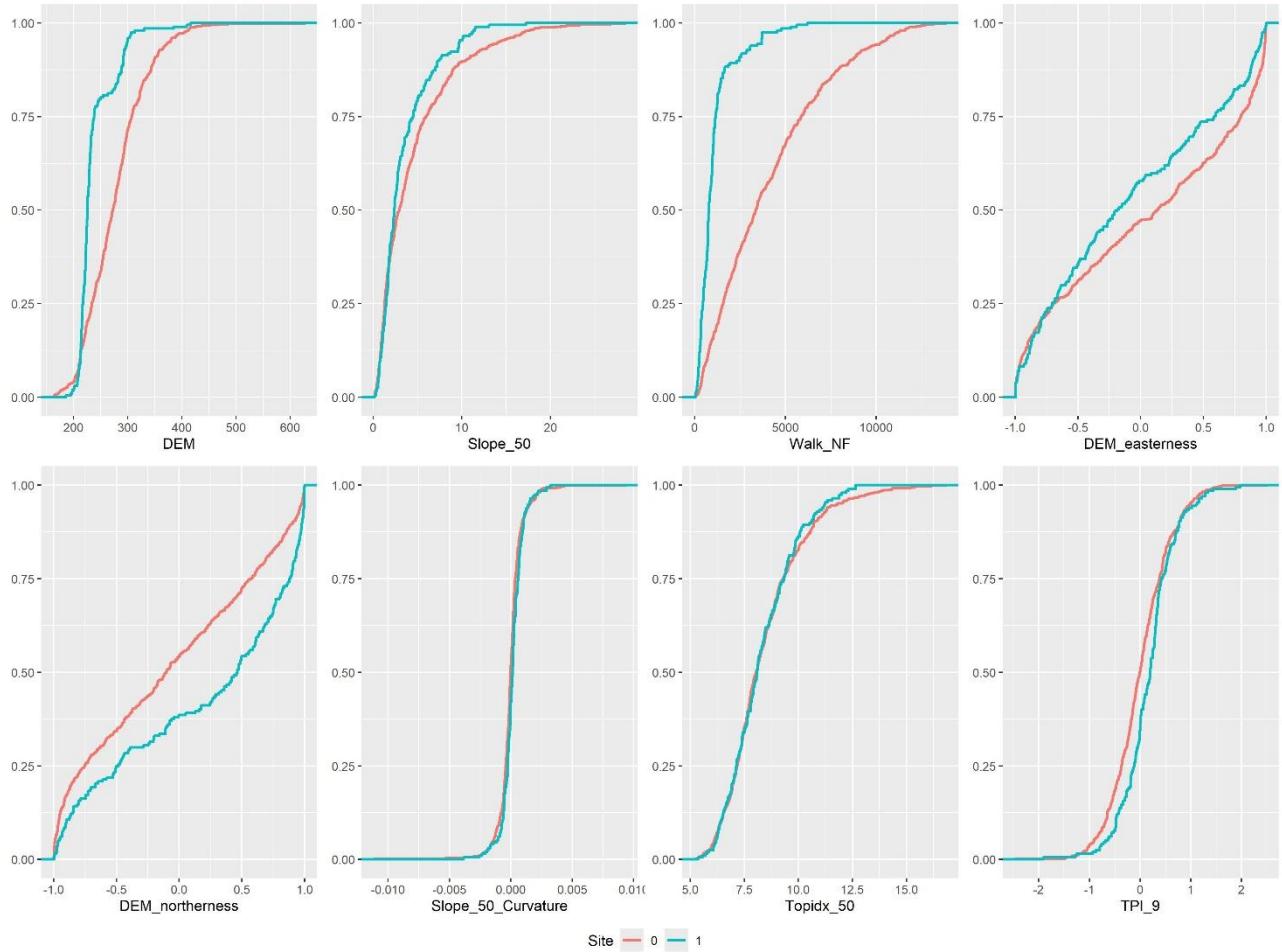
## Predictors



Density plot of variable values for each of the chronological subset

## IRON AGE

### ECD & Kolmogorov-Smirnov test



### Kolmogorov-Smirnov Test (Bootstrapping 10000)

**\$DEM**  
 Bootstrap p-value: < 2.22e-16  
 Naive p-value: 9.3981e-32  
 Full Sample Statistic: 0.49408

**\$Slope\_50**  
 Bootstrap p-value: 0.0121  
 Naive p-value: 0.013261  
 Full Sample Statistic: 0.13029

**\$Walk\_NF**  
 Bootstrap p-value: < 2.22e-16  
 Naive p-value: 6.5033e-50  
 Full Sample Statistic: 0.62098

**\$DEM\_easterness**  
 Bootstrap p-value: 0.0208  
 Naive p-value: 0.022031  
 Full Sample Statistic: 0.12352

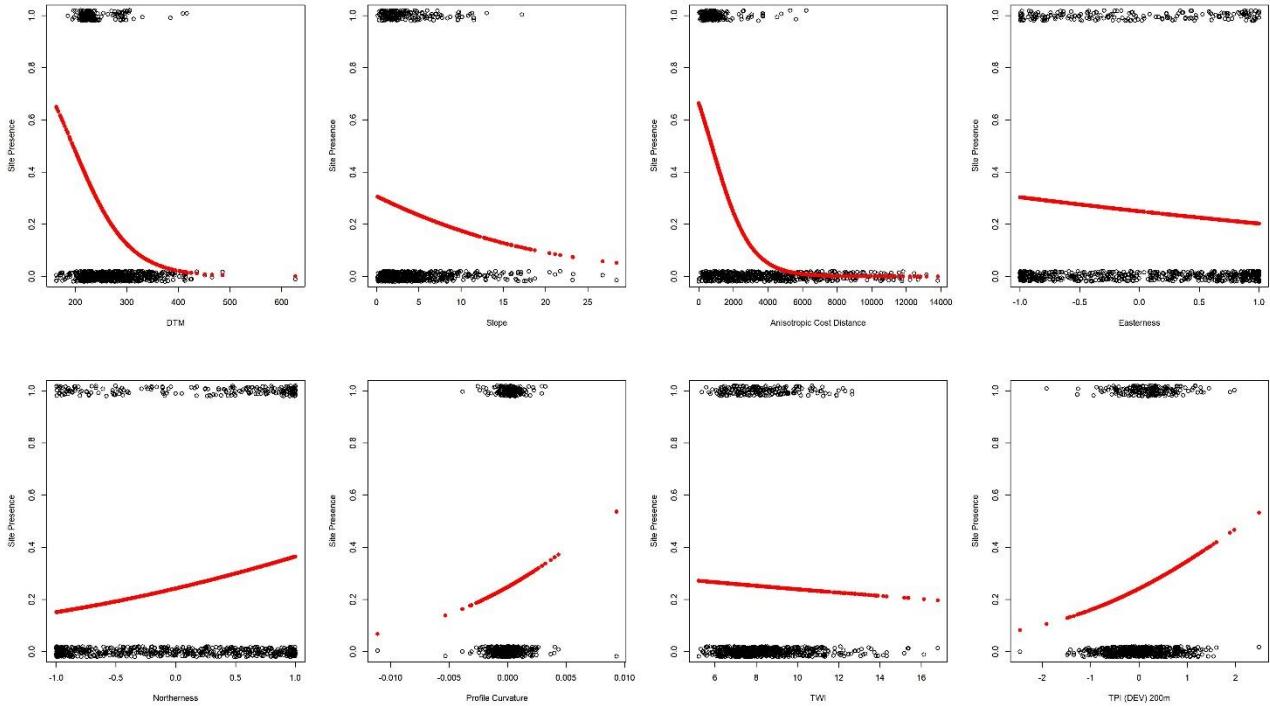
**\$DEM\_northerness**  
 Bootstrap p-value: < 2.22e-16  
 Naive p-value: 1.5372e-06  
 Full Sample Statistic: 0.21827

**\$slope\_50\_Curvature**  
 Bootstrap p-value: 0.0086  
 Naive p-value: 0.0089027  
 Full Sample Statistic: 0.13536

**\$Topidx\_50**  
 Bootstrap p-value: 0.856  
 Naive p-value: 0.86895  
 Full Sample Statistic: 0.049069

**\$TPI\_9**  
 Bootstrap p-value: 1e-04  
 Naive p-value: 0.00012428  
 Full Sample Statistic: 0.18105

## Univariate Logistic Regression



## Multivariate Logistic Regression

### Standard

Call:

```
glm(formula = name ~ DEM + Slope_50 + walk_NF + DEM_easterness +
    DEM_northerness + Slope_50_Curvature + TPI_9, family =
    binomial(logit),
    data = tab_fer)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.134e+00	6.941e-01	-1.633	0.1024
DEM	1.088e-02	3.485e-03	3.122	0.0018 **
Slope_50	-1.251e-01	3.069e-02	-4.077	4.57e-05 ***
Walk_NF	-1.234e-03	1.496e-04	-8.245	< 2e-16 ***
DEM_easterness	-8.239e-02	1.532e-01	-0.538	0.5908
DEM_northerness	6.583e-01	1.432e-01	4.599	4.25e-06 ***
Slope_50_Curvature	7.004e+01	1.240e+02	0.565	0.5721
TPI_9	4.096e-01	2.021e-01	2.026	0.0427 *

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 886.24 on 787 degrees of freedom  
 Residual deviance: 590.38 on 780 degrees of freedom

AIC: 606.38

Number of Fisher Scoring iterations: 7

### Backward Stepwise Selection (AIC)

Call:  
glm(formula = name ~ DEM + slope\_50 + walk\_NF + DEM\_northerness +  
TPI\_9, family = binomial(logit), data = tab\_fer)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.1502253	0.6883411	-1.671	0.09472 .
DEM	0.0109844	0.0034657	3.169	0.00153 **
Slope_50	-0.1232231	0.0304685	-4.044	5.25e-05 ***
Walk_NF	-0.0012443	0.0001494	-8.330	< 2e-16 ***
DEM_northerness	0.6561684	0.1427957	4.595	4.32e-06 ***
TPI_9	0.4766682	0.1787869	2.666	0.00767 **
---				

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 886.24 on 787 degrees of freedom  
Residual deviance: 590.89 on 782 degrees of freedom

AIC: 602.89

Number of Fisher Scoring iterations: 7

### Backward Stepwise Selection (Schwartz's AIC - BIC)

Call:  
glm(formula = name ~ DEM + slope\_50 + walk\_NF + DEM\_northerness +  
TPI\_9, family = binomial(logit), data = tab\_fer)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.1502253	0.6883411	-1.671	0.09472 .
DEM	0.0109844	0.0034657	3.169	0.00153 **
Slope_50	-0.1232231	0.0304685	-4.044	5.25e-05 ***
Walk_NF	-0.0012443	0.0001494	-8.330	< 2e-16 ***
DEM_northerness	0.6561684	0.1427957	4.595	4.32e-06 ***
TPI_9	0.4766682	0.1787869	2.666	0.00767 **
---				

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 886.24 on 787 degrees of freedom  
Residual deviance: 590.89 on 782 degrees of freedom

AIC: 602.89

Number of Fisher Scoring iterations: 7

### Standardized Coefficients

DEM	Slope	walk	Northerness	TPI
0.6381362	-0.5485152	-3.950018	0.4905073	0.289093

### Variance Inflation Factor

#### Variance Inflation Factor (Multivariate)

DEM	Slope_50	walk_NF	DEM_easterness
2.248972	1.211864	2.148395	1.112221

DEM_northerness	Slope_50_Curvature	TPI_9
1.031436	1.273825	1.303330

#### Variance Inflation Factor (AIC)

DEM	Slope_50	walk_NF	DEM_northerness	TPI_9
2.237725	1.191186	2.130355	1.026266	1.018798

#### Variance Inflation Factor (BIC)

DEM	Slope_50	walk_NF	DEM_northerness	TPI_9
2.237725	1.191186	2.130355	1.026266	1.018798

### Area under the ROC curve (AUC)

Call:

```
roc.default(response = fr_BIC_glm$y, predictor =  
fr_BIC_glm$fitted.values)
```

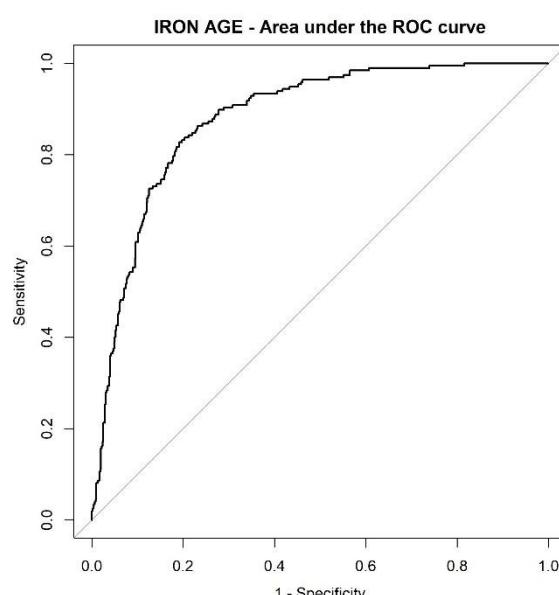
Data: fr\_BIC\_glm\$fitted.values in 591 controls (fr\_BIC\_glm\$y 0) < 197 cases (fr\_BIC\_glm\$y 1).

Area under the curve: 0.8793

### Discriminatory Ability

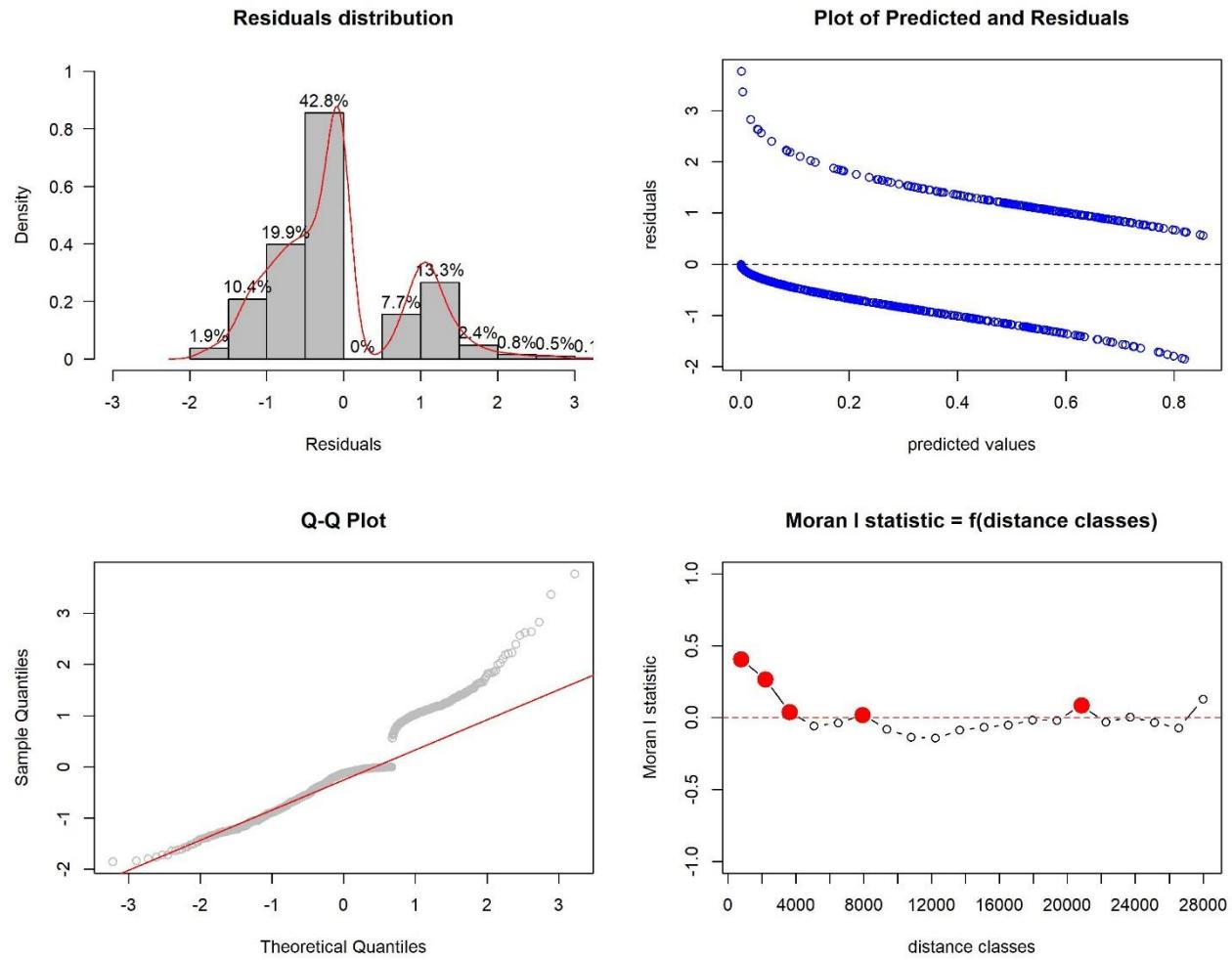
(HOSMER et. al. 2013 - Applied Logistic Regression - 3rd Ed.)

0.5 = No better than chance  
0.5-0.7 = Poor  
0.7-0.8 = Acceptable  
0.8-0.9 = Excellent  
0.9-1.0 = Outstanding

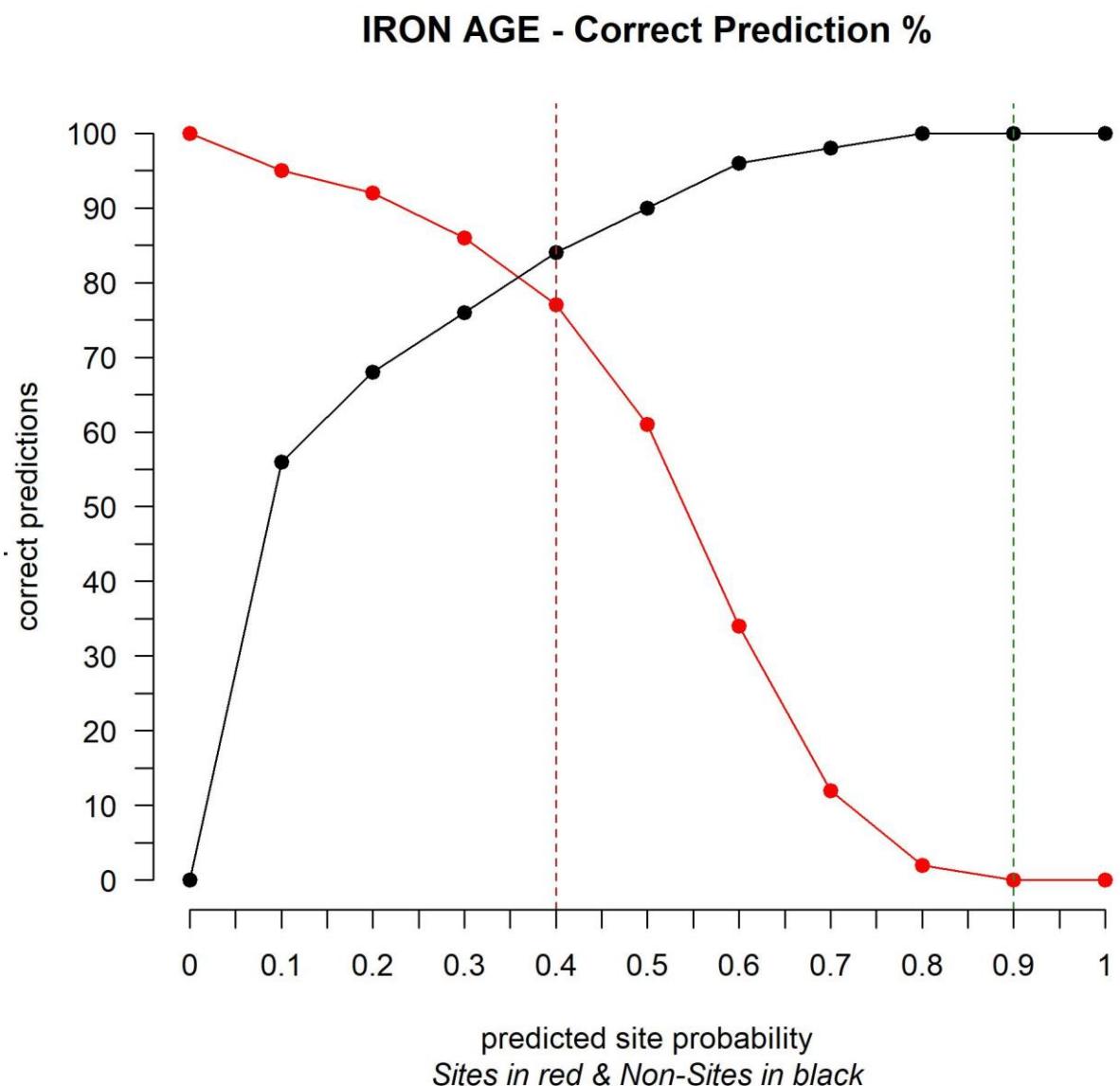


## Residuals

### IRON AGE - Multivariate Regression (Stepwise BIC) Residuals



**Correct prediction threshold**



## Model Assessment

-- Kolmogorov-Smirnov test --

Prediction values Sites/No-Sites

Bootstrap p-value: < 2.22e-16  
Naive p-value: 4.2866e-52  
Full Sample Statistic: 0.63452

-- Kvamme's Gain - Sites --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area	0.46	0.34	0.26	0.19	0.12	0.07	0.03	0.01	0
% Siti	0.95	0.92	0.86	0.77	0.61	0.34	0.12	0.02	0
Kvamme's Gain	0.52	0.63	0.70	0.75	0.80	0.79	0.75	0.50	NaN

-- Kvamme's Gain - No Sites in predicted area --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area	0.46	0.34	0.26	0.19	0.12	0.07	0.03	0.01	0
% No Sites	0.44	0.32	0.24	0.16	0.10	0.04	0.02	0.00	0
Kvamme's Gain	-0.05	-0.06	-0.08	-0.19	-0.20	-0.75	-0.50	-Inf	NaN

-- Kvamme's Gain - No sites in unpredicted area --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area Unpred	0.54	0.66	0.74	0.81	0.88	0.93	0.97	0.99	1
% No Sites	0.56	0.68	0.76	0.84	0.90	0.96	0.98	1.00	1
Kvamme's Gain	0.04	0.03	0.03	0.04	0.02	0.03	0.01	0.01	0

## External Area test

-- Kolmogorov-Smirnov test --

h0 = how likely the samples are from same distribution

### BVER/Lecco

Bootstrap p-value: < 2.22e-16  
Naive p-value: 4.32e-10  
Full Sample Statistic: 0.72665

### Como/Lecco

Bootstrap p-value: 2e-04  
Naive p-value: 0.0004909  
Full Sample Statistic: 0.47736

### BVER/Como

Bootstrap p-value: < 2.22e-16  
Naive p-value: 4.4965e-34  
Full Sample Statistic: 0.74974

### Como/Garda

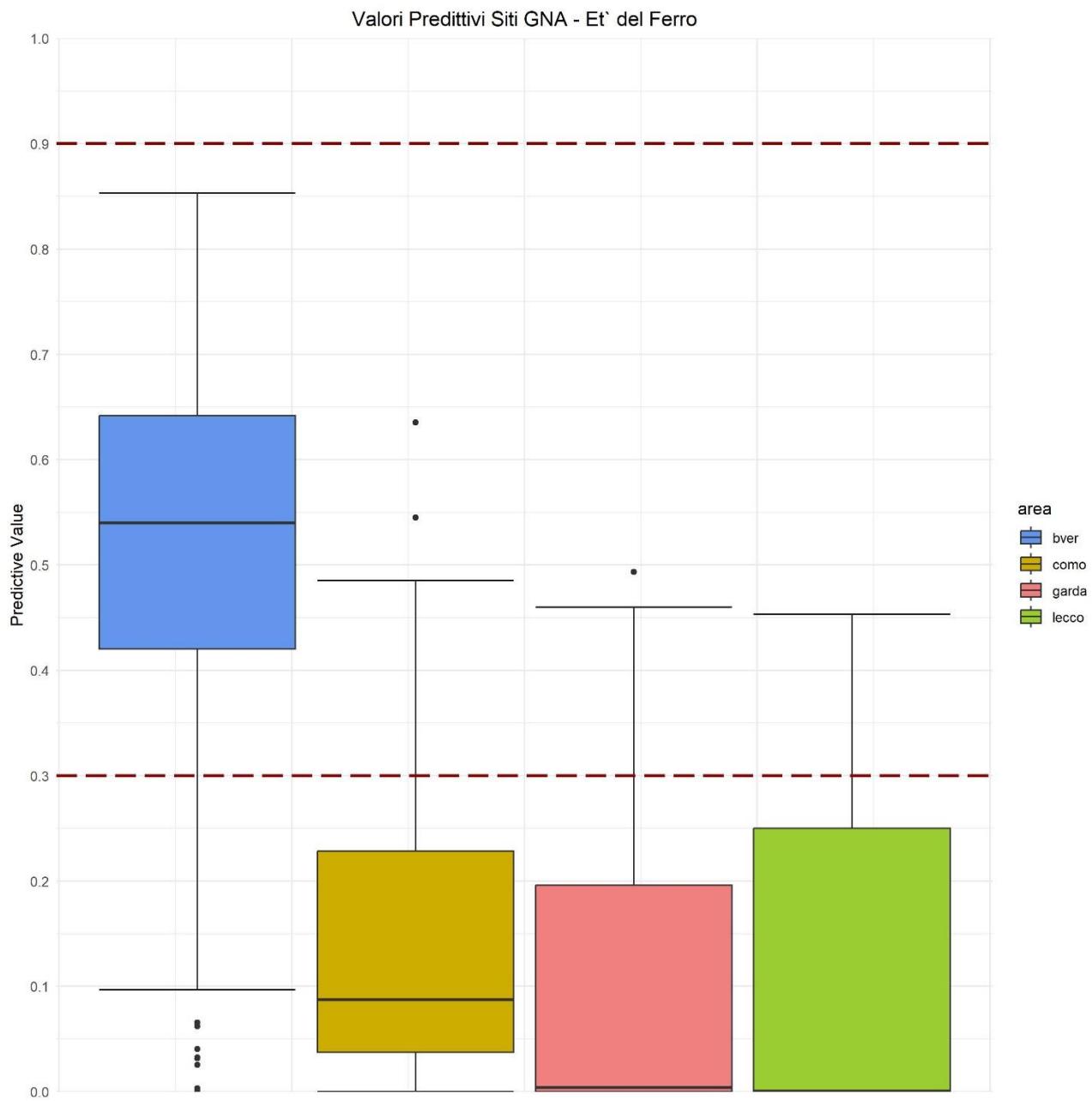
Bootstrap p-value: 0.0115  
Naive p-value: 0.01039  
Full Sample Statistic: 0.40677

### BVER/Garda

Bootstrap p-value: < 2.22e-16  
Naive p-value: 3.2254e-08  
Full Sample Statistic: 0.70469

### Lecco/Garda

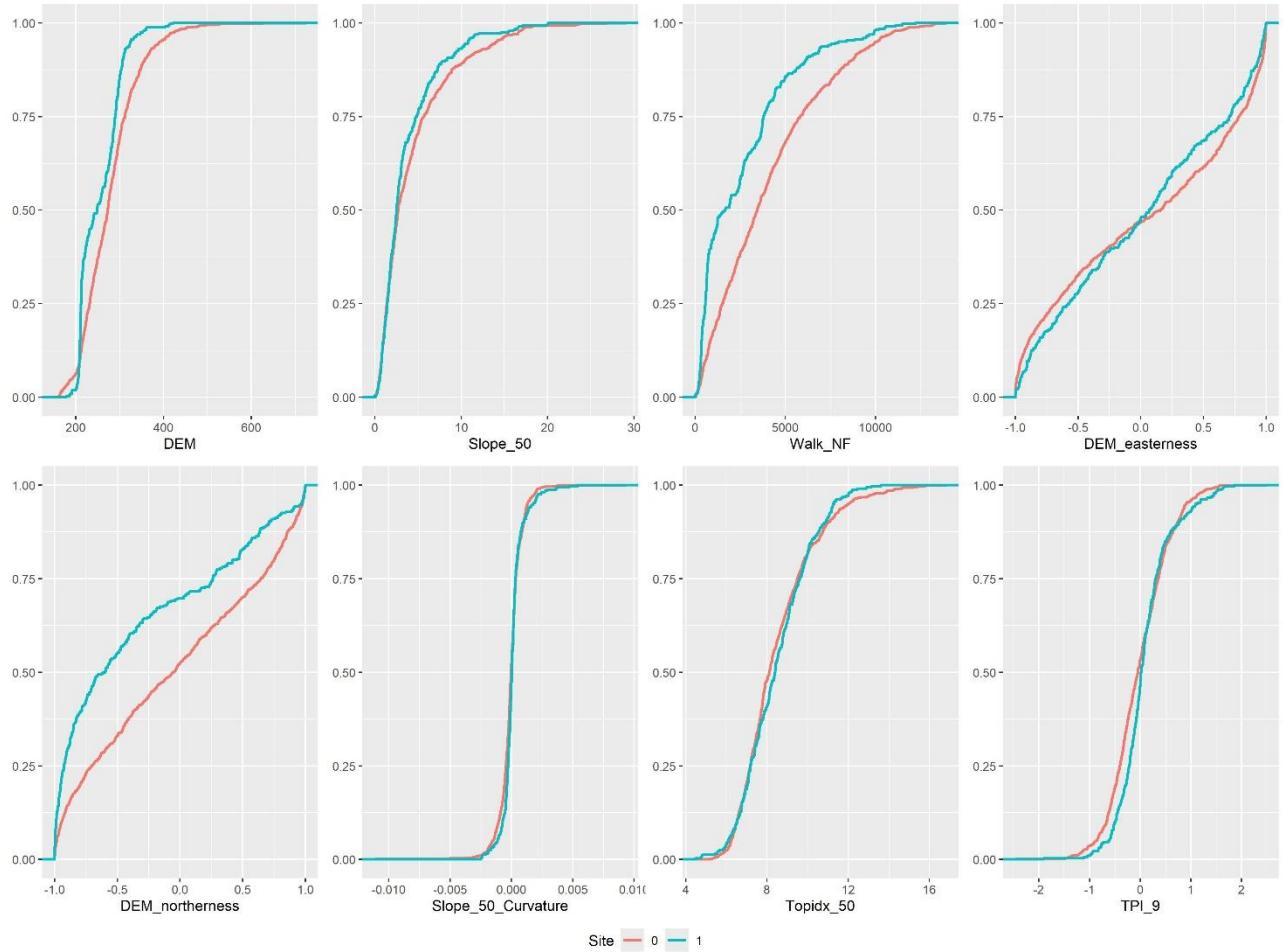
Bootstrap p-value: 0.6203  
Naive p-value: 0.68859  
Full Sample Statistic: 0.21471



Predictive values of the Iron Age sites in the BVER and control areas

# ROMAN AGE

## ECD & Kolmogorov-Smirnov test



## Kolmogorov-Smirnov Test (Bootstrapping 10000)

**\$DEM**  
 Bootstrap p-value: < 2.22e-16  
 Naive p-value: 1.1786e-09  
 Full Sample Statistic: 0.21042

**\$Slope\_50**  
 Bootstrap p-value: 0.0222  
 Naive p-value: 0.022114  
 Full Sample Statistic: 0.096875

**\$Walk\_NF**  
 Bootstrap p-value: < 2.22e-16  
 Naive p-value: 5.9443e-16  
 Full Sample Statistic: 0.27292

**\$DEM\_easterness**  
 Bootstrap p-value: 0.11  
 Naive p-value: 0.11543  
 Full Sample Statistic: 0.077083

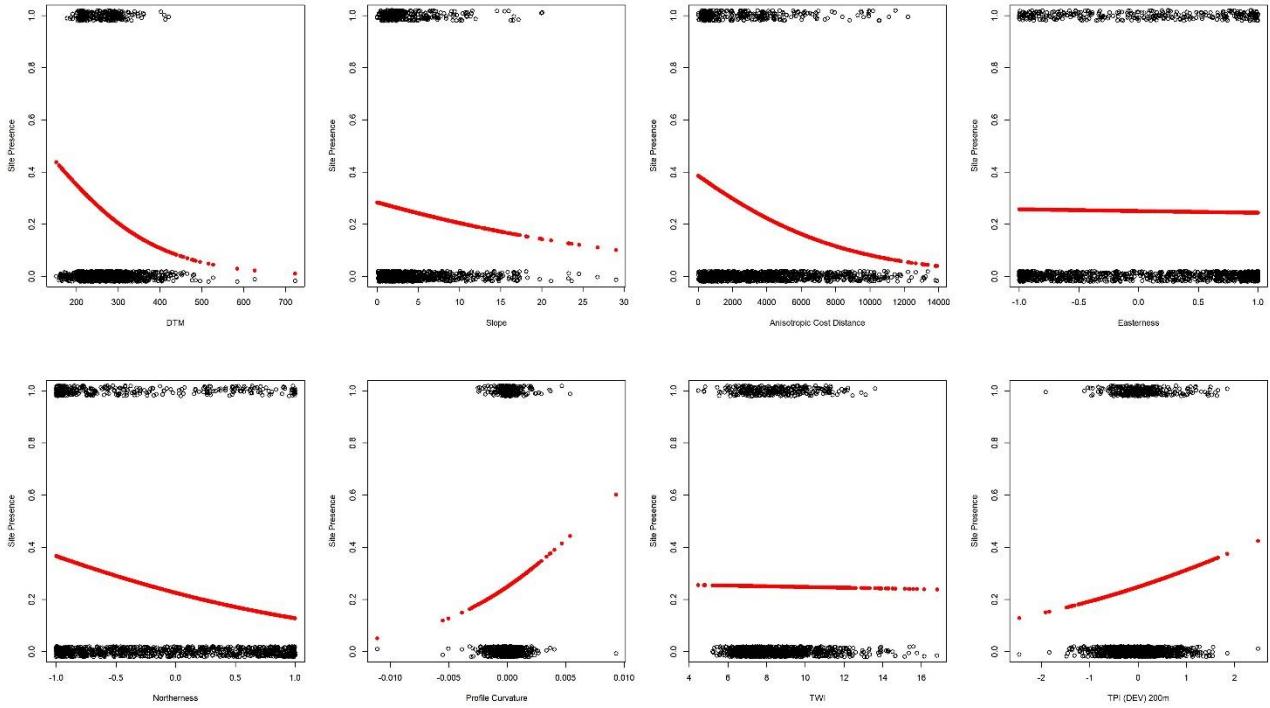
**\$DEM\_northerness**  
 Bootstrap p-value: < 2.22e-16  
 Naive p-value: 2.2552e-11  
 Full Sample Statistic: 0.22917

**\$slope\_50\_Curvature**  
 Bootstrap p-value: 0.0132  
 Naive p-value: 0.014885  
 Full Sample Statistic: 0.10104

**\$Topidx\_50**  
 Bootstrap p-value: 0.1173  
 Naive p-value: 0.1246  
 Full Sample Statistic: 0.076042

**\$TPI\_9**  
 Bootstrap p-value: 1e-04  
 Naive p-value: 4.0799e-05  
 Full Sample Statistic: 0.15

## Univariate Logistic Regression



## Multivariate Logistic Regression

### Standard

Call:

```
glm(formula = name ~ DEM + Slope_50 + walk_NF + DEM_northerness +
  Slope_50_Curvature + TPI_9, family = binomial(logit), data =
  tab_rom)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	7.147e-02	3.654e-01	0.196	0.8449
DEM	-2.894e-03	1.667e-03	-1.737	0.0825
Slope_50	-1.995e-02	1.859e-02	-1.073	0.2833
Walk_NF	-1.396e-04	3.498e-05	-3.992	6.56e-05 ***
DEM_northerness	-6.097e-01	1.032e-01	-5.906	3.51e-09 ***
Slope_50_Curvature	1.123e+02	7.989e+01	1.406	0.1597
TPI_9	2.808e-01	1.322e-01	2.125	0.0336 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 1439.6 on 1279 degrees of freedom
Residual deviance: 1322.7 on 1273 degrees of freedom
AIC: 1336.7
```

Number of Fisher Scoring iterations: 4

### Backward Stepwise Selection (AIC)

Call:  
glm(formula = name ~ DEM + walk\_NF + DEM\_northerness + TPI\_9,  
family = binomial(logit), data = tab\_rom)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	6.371e-02	3.598e-01	0.177	0.85945
DEM	-3.205e-03	1.563e-03	-2.050	0.04034 *
walk_NF	-1.367e-04	3.409e-05	-4.011	6.05e-05 ***
DEM_northerness	-6.240e-01	1.026e-01	-6.081	1.19e-09 ***
TPI_9	3.609e-01	1.217e-01	2.966	0.00301 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1439.6 on 1279 degrees of freedom  
Residual deviance: 1325.7 on 1275 degrees of freedom  
AIC: 1335.7

Number of Fisher Scoring iterations: 4

### Backward Stepwise Selection (Schwartz's AIC - BIC)

Call:  
glm(formula = name ~ walk\_NF + DEM\_northerness, family =  
binomial(logit),  
data = tab\_rom)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-6.265e-01	1.072e-01	-5.844	5.10e-09 ***
walk_NF	-1.837e-04	2.721e-05	-6.749	1.49e-11 ***
DEM_northerness	-6.182e-01	1.023e-01	-6.044	1.50e-09 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1439.6 on 1279 degrees of freedom  
Residual deviance: 1337.2 on 1277 degrees of freedom  
AIC: 1343.2

Number of Fisher Scoring iterations: 4

## Standardized Coefficients

walk      Northerness  
-0.14508 -0.1122316

## Variance Inflation Factor

Variance Inflation Factor (Standard)  
DEM                  Slope\_50                  walk\_NF                  DEM\_northerness  
1.882931            1.149960            1.719892            1.015733

Slope\_50\_Curvature    TPI\_9  
1.243282            1.225888

Variance Inflation Factor (AIC)  
DEM                  walk\_NF                  DEM\_northerness          TPI\_9  
1.659770            1.641090            1.003960            1.028199

Variance Inflation Factor (BIC)  
walk\_NF            DEM\_northerness  
1.002365           1.002365

## Area under the ROC curve (AUC)

Call:

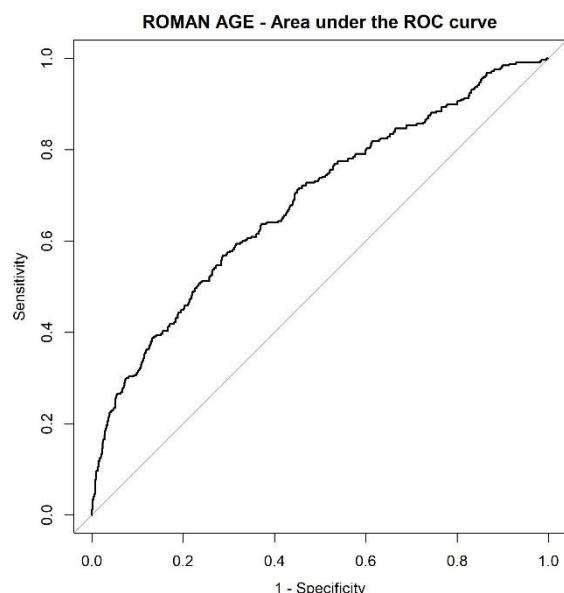
```
roc.default(response = rm_BIC_glm$y, predictor =  
rm_BIC_glm$fitted.values)
```

Data: rm\_BIC\_glm\$fitted.values in 960 controls (rm\_BIC\_glm\$y 0) <  
320 cases (rm\_BIC\_glm\$y 1).

Area under the curve: 0.6848

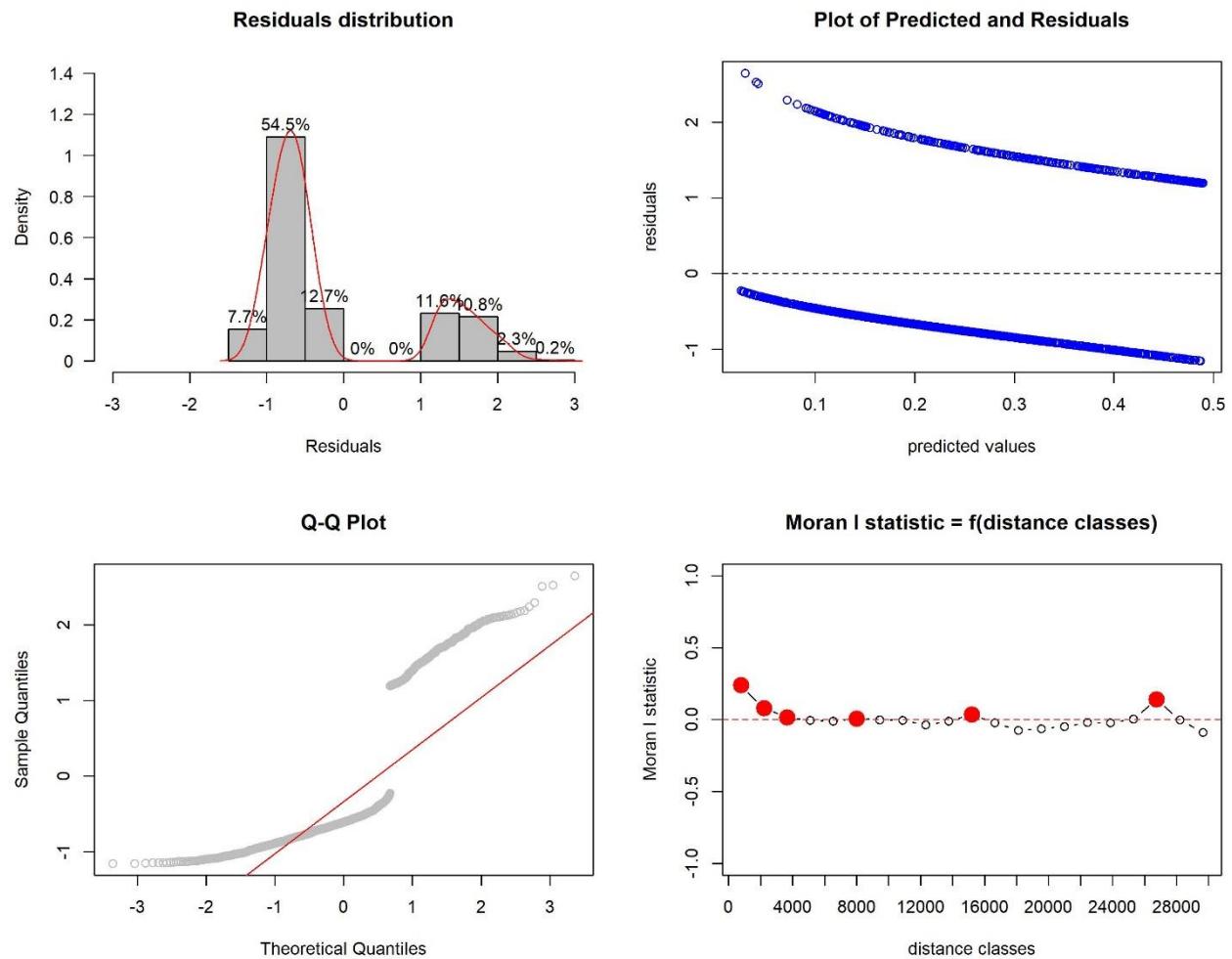
Discriminatory Ability  
(HOSMER et. al. 2013 - Applied  
Logistic Regression - 3rd Ed.)

0.5 = No better than chance  
0.5-0.7 = Poor  
0.7-0.8 = Acceptable  
0.8-0.9 = Excellent  
0.9-1.0 = Outstanding

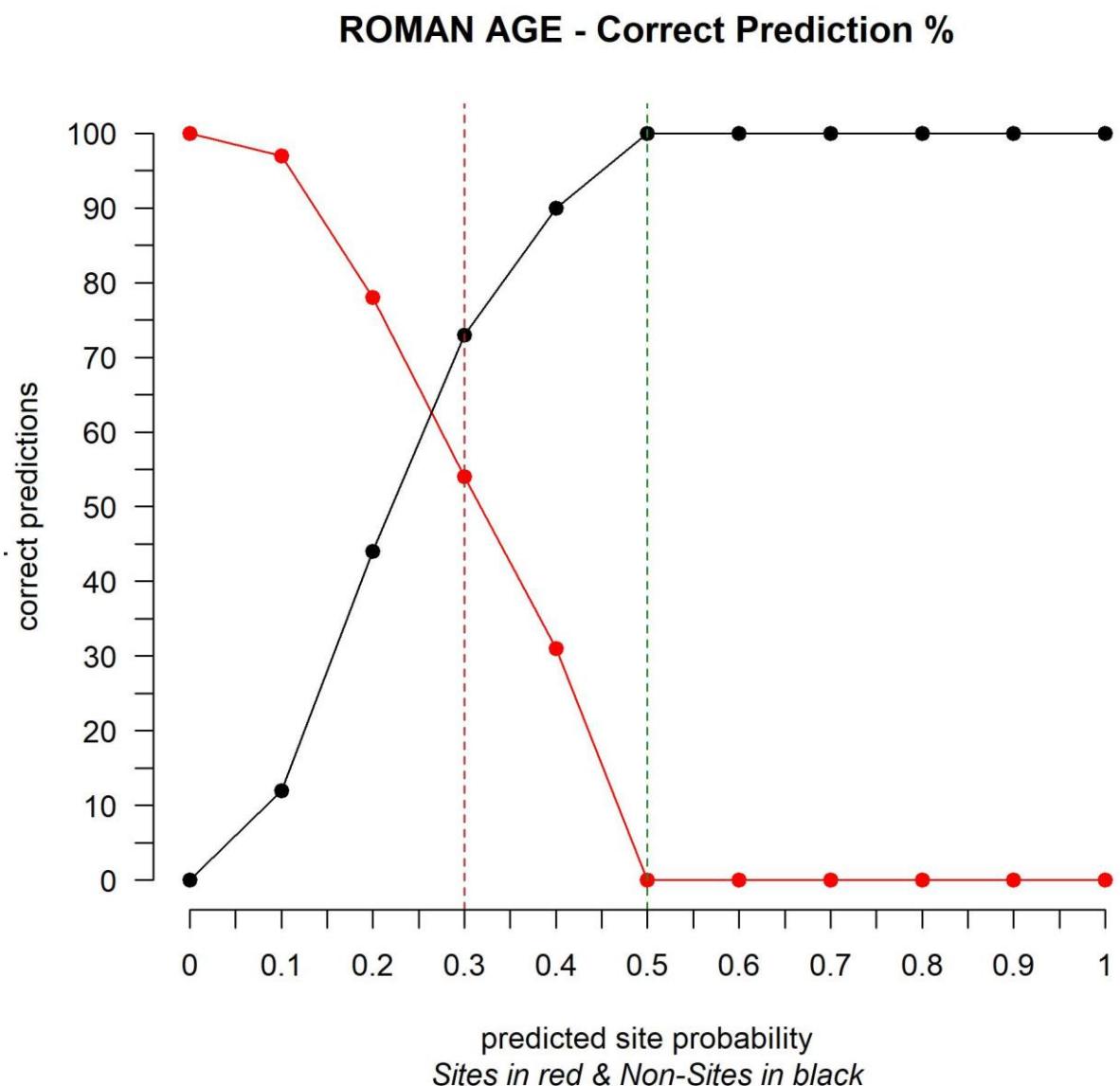


## Residuals

ROMAN AGE - Multivariate Regression (Stepwise BIC) Residuals



### Correct prediction threshold



## Model Assessment

-- Kolmogorov-Smirnov test --

Prediction values Sites/No-Sites

Bootstrap p-value: < 2.22e-16  
Naive p-value: 4.887e-17  
Full Sample Statistic: 0.28229

-- Kvamme's Gain - Sites --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area	0.90	0.59	0.27	0.09	0	0	0	0	0
% Siti	0.97	0.78	0.54	0.31	0	0	0	0	0
Kvamme's Gain	0.07	0.24	0.50	0.71	NaN	NaN	NaN	NaN	NaN

-- Kvamme's Gain - No Sites in predicted area --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area	0.90	0.59	0.27	0.09	0	0	0	0	0
% No Sites	0.88	0.56	0.27	0.10	0	0	0	0	0
Kvamme's Gain	-0.02	-0.05	0.00	0.10	NaN	NaN	NaN	NaN	NaN

-- Kvamme's Gain - No sites in unpredicted area --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area Unpred	0.10	0.41	0.73	0.91	1	1	1	1	1
% No Sites	0.12	0.44	0.73	0.90	1	1	1	1	1
Kvamme's Gain	0.17	0.07	0.00	-0.01	0	0	0	0	0

## External Area test

-- Kolmogorov-Smirnov test --

h0 = how likely the samples are from same distribution

### BVER/Lecco

Bootstrap p-value: < 2.22e-16  
Naive p-value: 3.8726e-07  
Full Sample Statistic: 0.26042

### Como/Lecco

Bootstrap p-value: 0.2284  
Naive p-value: 0.26253  
Full Sample Statistic: 0.096504

### BVER/Como

Bootstrap p-value: < 2.22e-16  
Naive p-value: 7.1724e-13  
Full Sample Statistic: 0.30888

### Como/Garda

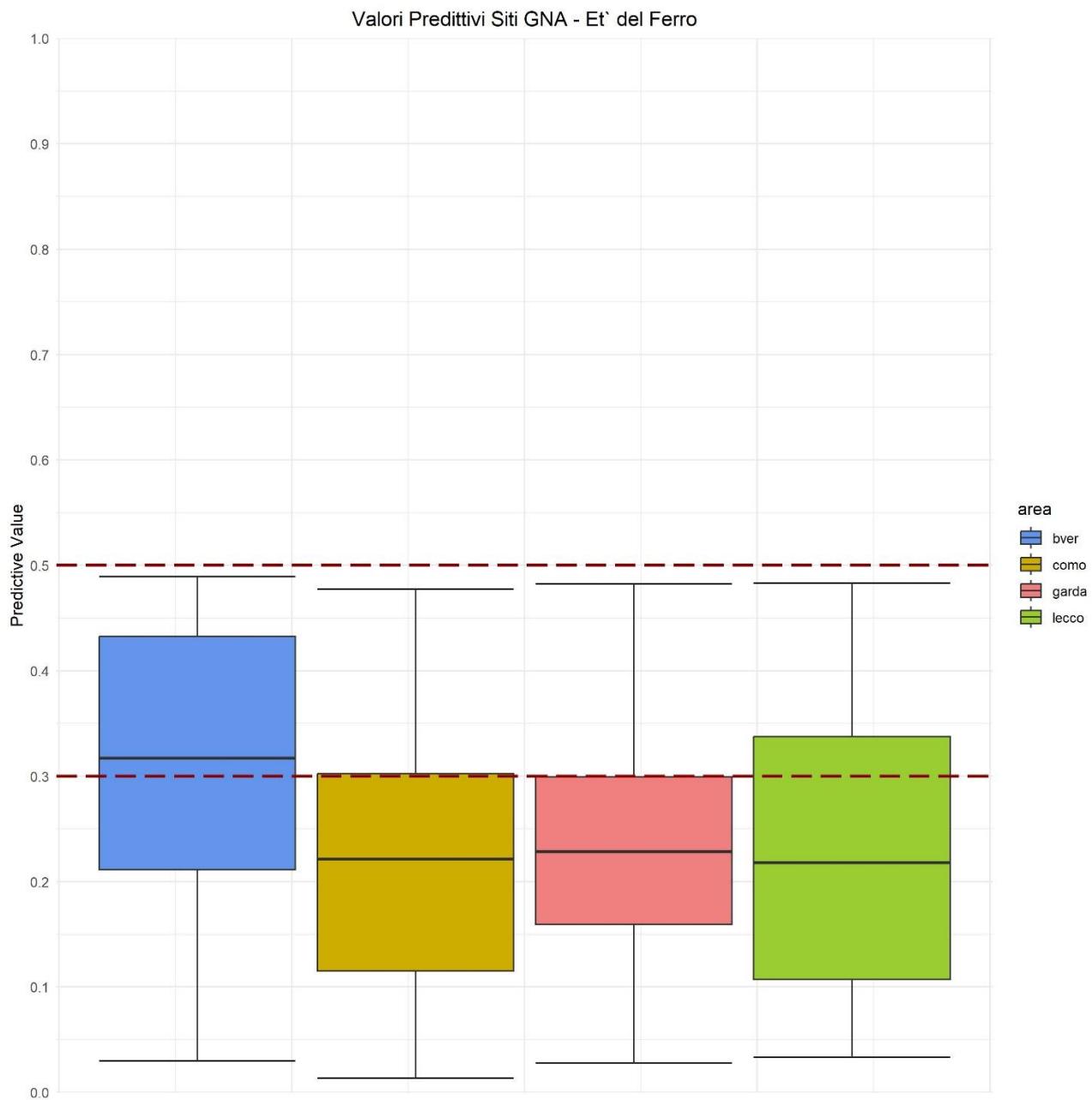
Bootstrap p-value: 0.012  
Naive p-value: 0.016074  
Full Sample Statistic: 0.13854

### BVER/Garda

Bootstrap p-value: < 2.22e-16  
Naive p-value: 1.111e-10  
Full Sample Statistic: 0.29856

### Lecco/Garda

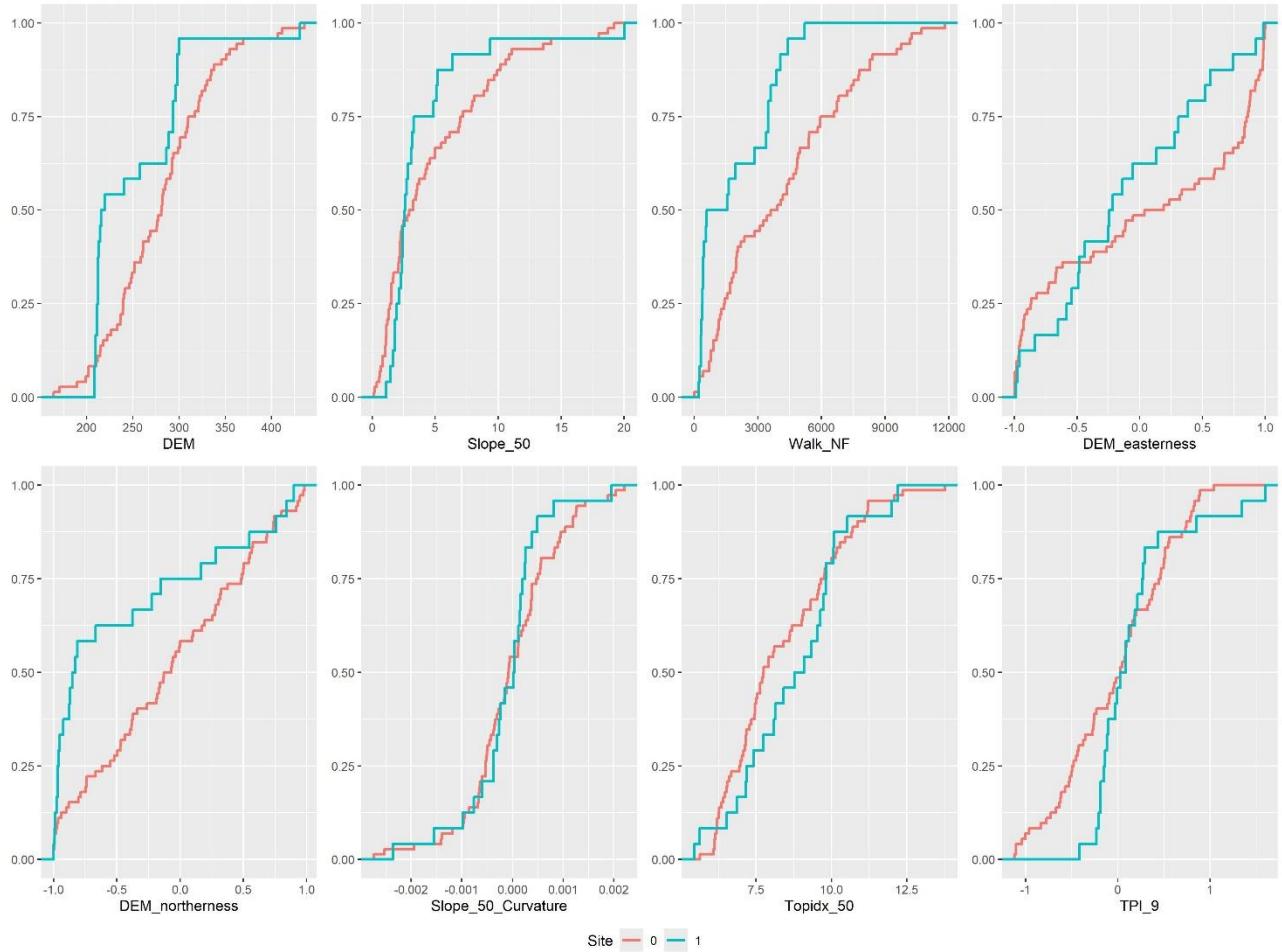
Bootstrap p-value: 0.0026  
Naive p-value: 0.0034704  
Full Sample Statistic: 0.17894



Predictive values of the Roman sites in the BVER and control areas

## LATE ANTIQUITY

### ECD & Kolmogorov-Smirnov test



### Kolmogorov-Smirnov Test (Bootstrapping 10000)

\$DEM  
Bootstrap p-value: 0.0043  
Naive p-value: 0.0069819  
Full Sample Statistic: 0.38889

\$Slope\_50  
Bootstrap p-value: 0.2265  
Naive p-value: 0.24974  
Full Sample Statistic: 0.23611

\$Walk\_NF  
Bootstrap p-value: 0.0013  
Naive p-value: 0.0018861  
Full Sample Statistic: 0.43056

\$DEM\_easterness  
Bootstrap p-value: 0.0779  
Naive p-value: 0.084277  
Full Sample Statistic: 0.29167

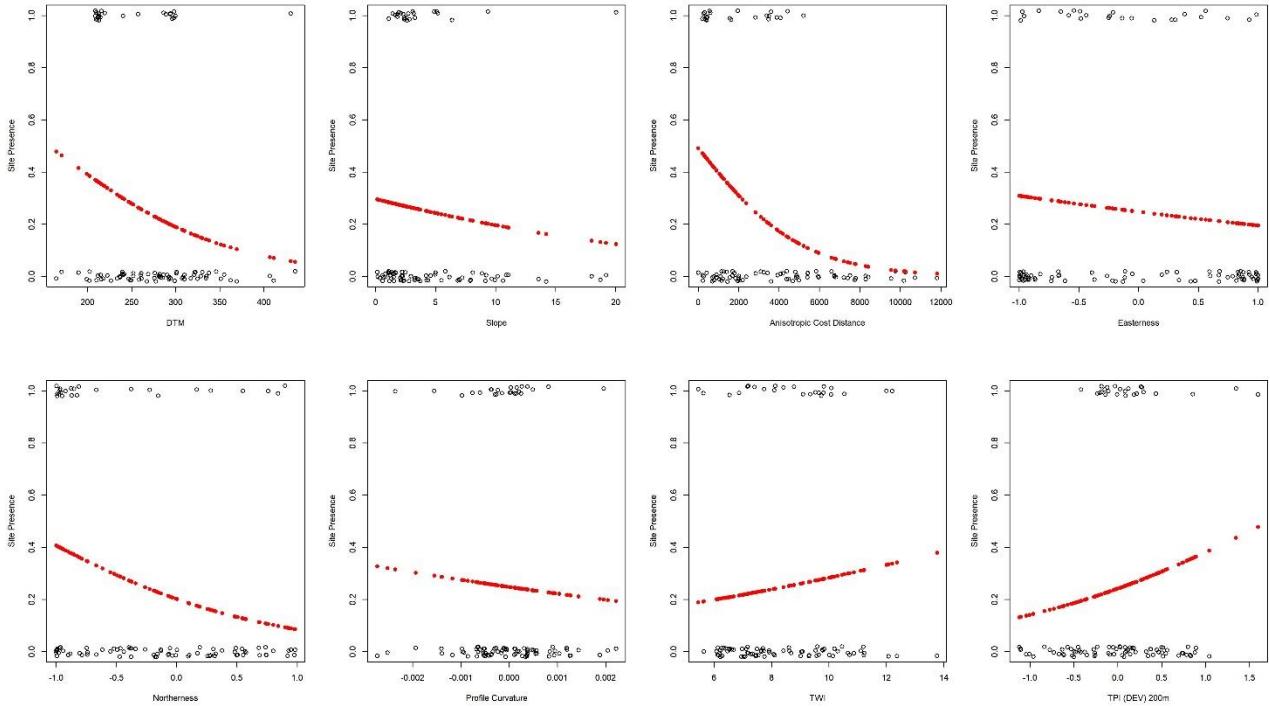
\$DEM\_northerness  
Bootstrap p-value: 0.0021  
Naive p-value: 0.0018861  
Full Sample Statistic: 0.43056

\$slope\_50\_Curvature  
Bootstrap p-value: 0.361  
Naive p-value: 0.39196  
Full Sample Statistic: 0.20833

\$Topidx\_50  
Bootstrap p-value: 0.2924  
Naive p-value: 0.31538  
Full Sample Statistic: 0.22222

\$TPI\_9  
Bootstrap p-value: 0.0196  
Naive p-value: 0.022308  
Full Sample Statistic: 0.34722

## Univariate Logistic Regression



## Multivariate Logistic Regression

### Standard

Call:

```
glm(formula = name ~ DEM + walk_NF + DEM_northerness + TPI_9,
    family = binomial(logit), data = tab_tar)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.7051443	1.5949680	-0.442	0.6584
DEM	0.0013115	0.0070723	0.185	0.8529
Walk_NF	-0.0003646	0.0001833	-1.989	0.0467 *
DEM_northerness	-0.7838041	0.4336749	-1.807	0.0707 .
TPI_9	0.6669812	0.4922287	1.355	0.1754

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 107.97 on 95 degrees of freedom
Residual deviance: 89.57 on 91 degrees of freedom
AIC: 99.57
```

Number of Fisher Scoring iterations: 5

### **Backward Stepwise Selection (AIC)**

Call:  
glm(formula = name ~ walk\_NF + DEM\_northerness + TPI\_9, family = binomial(logit),  
data = tab\_tar)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.4207163	0.4373635	-0.962	0.33608
walk_NF	-0.0003412	0.0001293	-2.640	0.00829 **
DEM_northerness	-0.7862450	0.4340619	-1.811	0.07008 .
TPI_9	0.6810080	0.4859153	1.401	0.16107
---				

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 107.968 on 95 degrees of freedom  
Residual deviance: 89.604 on 92 degrees of freedom  
AIC: 97.604

Number of Fisher Scoring iterations: 5

### **Backward Stepwise Selection (Schwartz's AIC - BIC)**

Call:  
glm(formula = name ~ walk\_NF, family = binomial(logit), data = tab\_tar)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.0340337	0.3723704	-0.091	0.92718
walk_NF	-0.0003826	0.0001266	-3.022	0.00251 **
---				

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 107.968 on 95 degrees of freedom  
Residual deviance: 94.753 on 94 degrees of freedom  
AIC: 98.753

Number of Fisher Scoring iterations: 5

## Standardized Coefficients

walk\_NF  
-0.4927506

## Variance Inflation Factor

Variance Inflation Factor (Standard)  
DEM walk\_NF DEM\_northerness TPI\_9  
1.997966 1.992560 1.041547 1.052639

Variance Inflation Factor (AIC)  
Walk\_NF DEM\_northerness TPI\_9  
1.012909 1.039097 1.026246

## Area under the ROC curve (AUC)

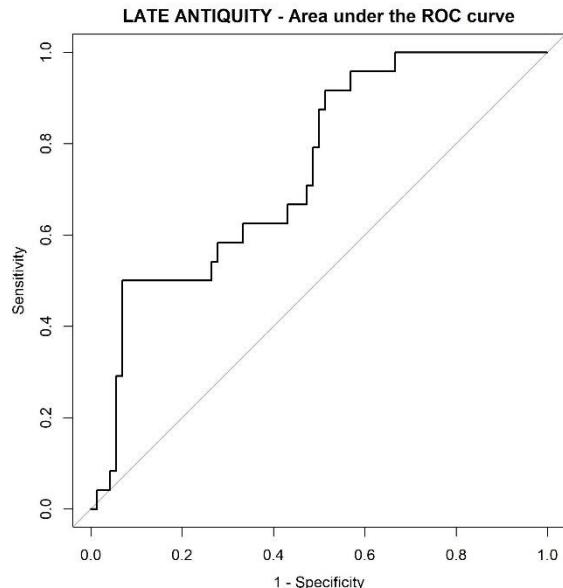
Call:

```
roc.default(response = ta_BIC_glm$y, predictor =  
ta_BIC_glm$fitted.values)
```

Data: ta\_BIC\_glm\$fitted.values in 72 controls (ta\_BIC\_glm\$y 0) <  
24 cases (ta\_BIC\_glm\$y 1).  
Area under the curve: 0.7425

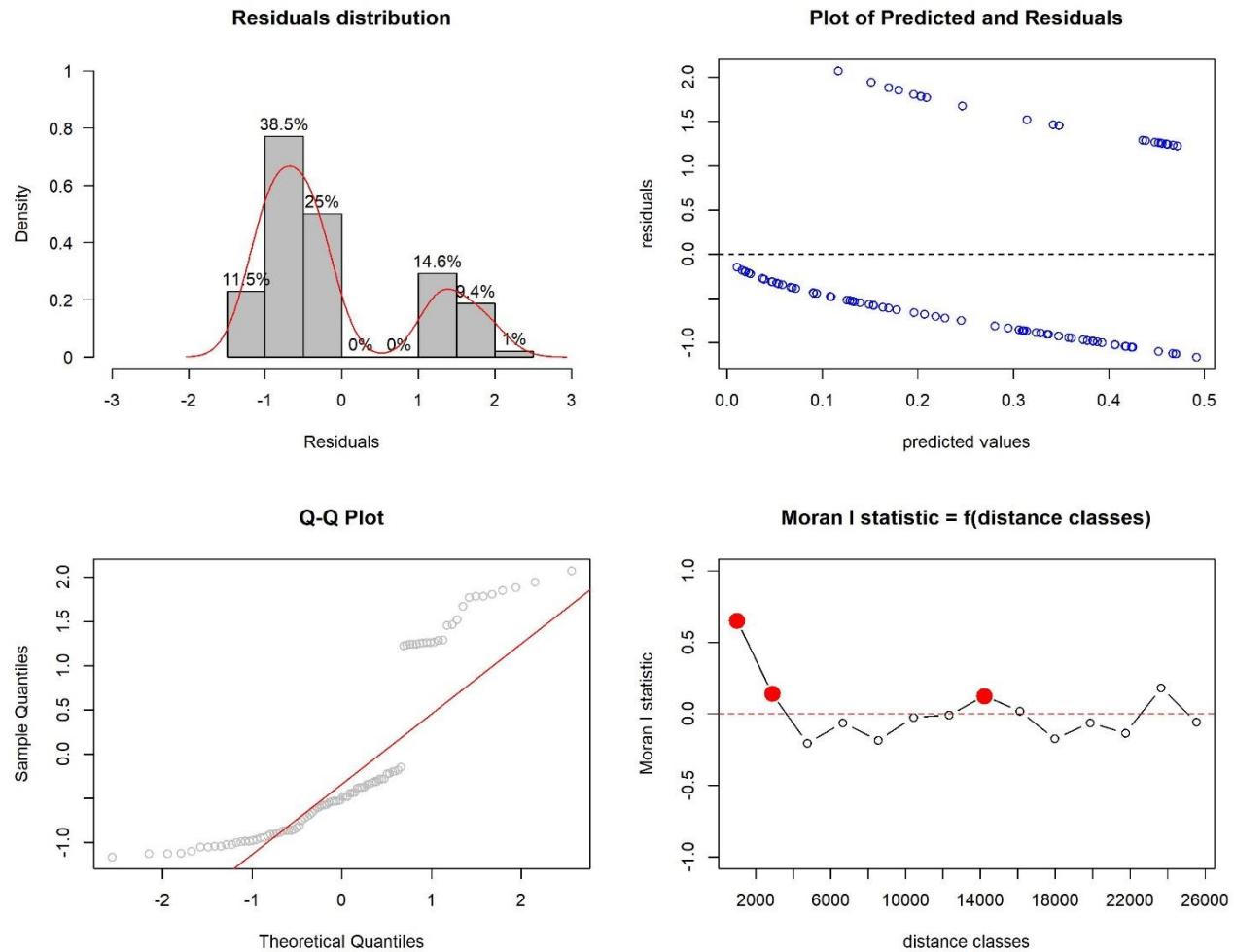
Discriminatory Ability  
(HOSMER et. al. 2013 - Applied  
Logistic Regression - 3rd Ed.)

0.5 = No better than chance  
0.5-0.7 = Poor  
0.7-0.8 = Acceptable  
0.8-0.9 = Excellent  
0.9-1.0 = Outstanding

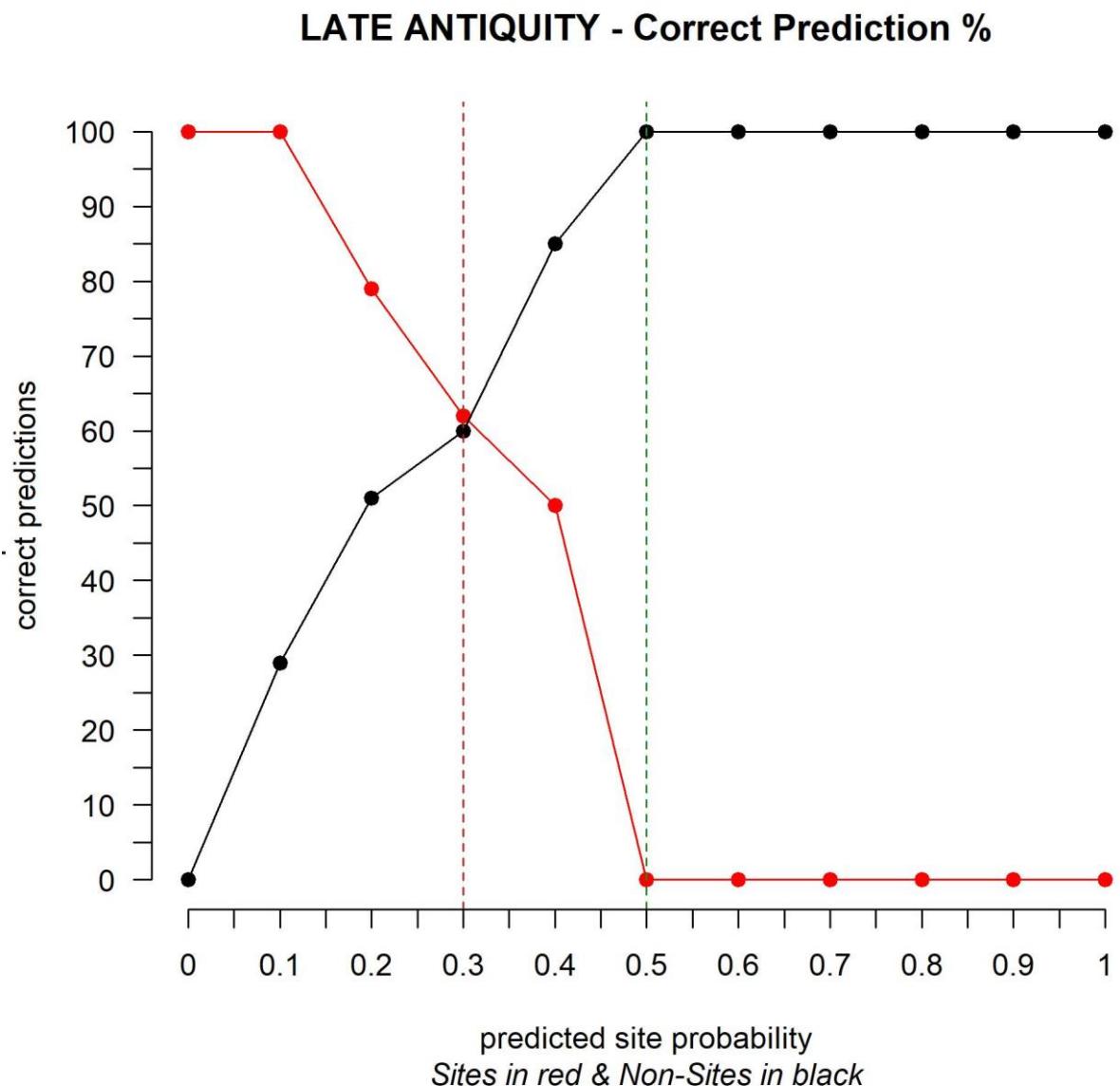


## Residuals

### LATE ANTIQUITY - Multivariate Regression (Stepwise BIC) Residuals



### Correct prediction threshold



## Model Assessment

-- Kolmogorov-Smirnov test --

Prediction values Sites/No-Sites

Bootstrap p-value: 0.0026  
Naive p-value: 0.0029669  
Full Sample Statistic: 0.41667

-- Kvamme's Gain - Sites --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area	0.76	0.55	0.37	0.19	0	0	0	0	0
% Siti	1.00	0.79	0.62	0.50	0	0	0	0	0
Kvamme's Gain	0.24	0.30	0.40	0.62	NaN	NaN	NaN	NaN	NaN

-- Kvamme's Gain - No Sites in predicted area --

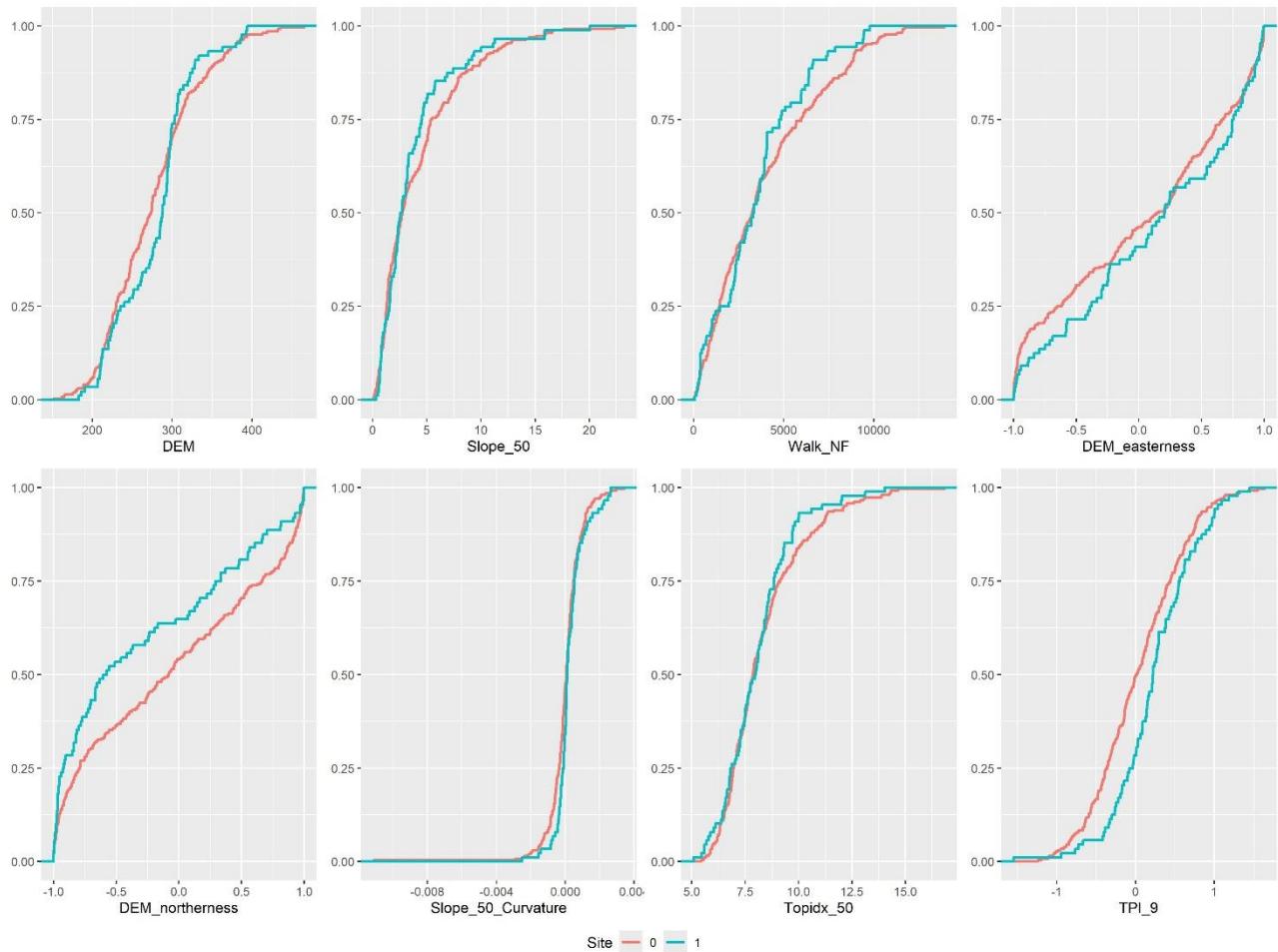
	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area	0.76	0.55	0.37	0.19	0	0	0	0	0
% No Sites	0.71	0.49	0.40	0.15	0	0	0	0	0
Kvamme's Gain	-0.07	-0.12	0.08	-0.27	NaN	NaN	NaN	NaN	NaN

-- Kvamme's Gain - No Sites in unpredicted area --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area Unpred	0.24	0.45	0.63	0.81	1	1	1	1	1
% No Sites	0.29	0.51	0.60	0.85	1	1	1	1	1
Kvamme's Gain	0.17	0.12	-0.05	0.05	0	0	0	0	0

## MIDDLE AGES

### ECD & Kolmogorov-Smirnov test



### Kolmogorov-Smirnov Test (Bootstrapping 10000)

**\$DEM**  
 Bootstrap p-value: 0.0446  
 Naive p-value: 0.051122  
 Full Sample Statistic: 0.16667

**\$Slope\_50**  
 Bootstrap p-value: 0.2298  
 Naive p-value: 0.25375  
 Full Sample Statistic: 0.125

**\$Walk\_NF**  
 Bootstrap p-value: 0.2895  
 Naive p-value: 0.32266  
 Full Sample Statistic: 0.11742

**\$DEM\_easterness**  
 Bootstrap p-value: 0.3394  
 Naive p-value: 0.36153  
 Full Sample Statistic: 0.11364

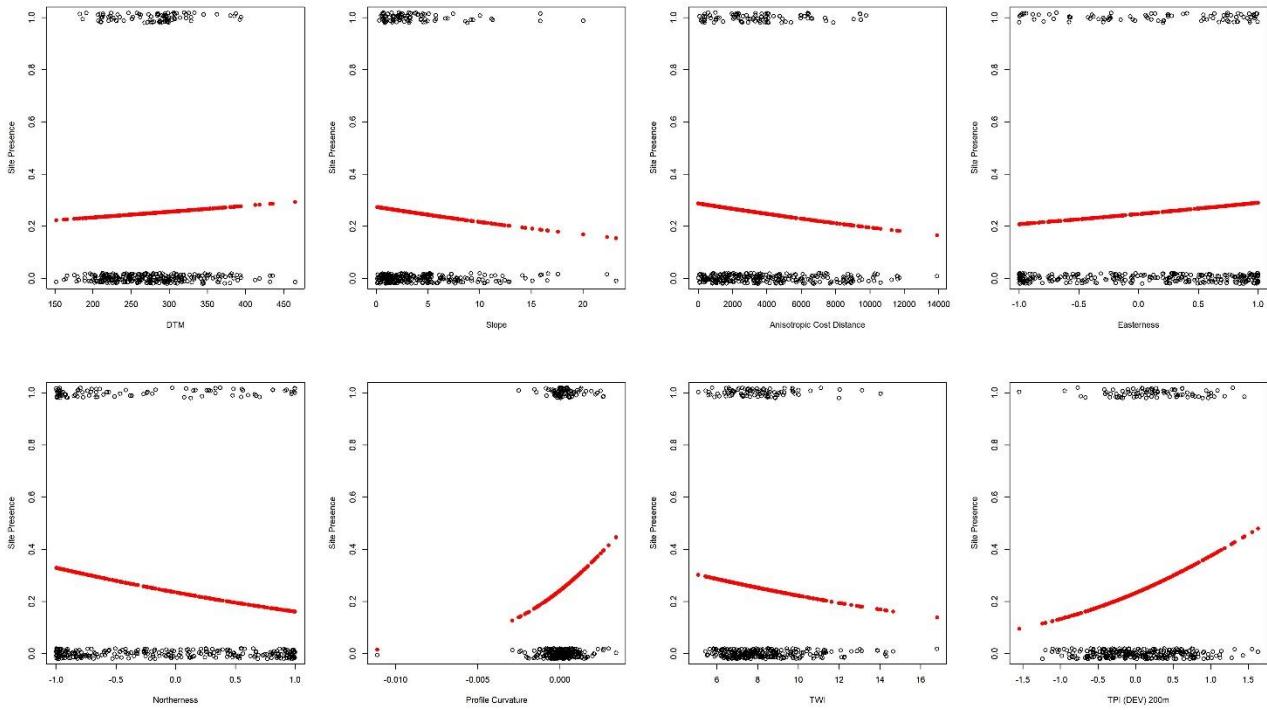
**\$DEM\_northerness**  
 Bootstrap p-value: 0.0297  
 Naive p-value: 0.036355  
 Full Sample Statistic: 0.17424

**\$slope\_50\_Curvature**  
 Bootstrap p-value: 0.0889  
 Naive p-value: 0.096591  
 Full Sample Statistic: 0.15152

**\$Topidx\_50**  
 Bootstrap p-value: 0.6081  
 Naive p-value: 0.64647  
 Full Sample Statistic: 0.090909

**\$TPI\_9**  
 Bootstrap p-value: 0.0039  
 Naive p-value: 0.0042524  
 Full Sample Statistic: 0.21591

## Univariate Logistic Regression



## Multivariate Logistic Regression

### Standard

Call:

```
glm(formula = name ~ DEM + DEM_northerness + TPI_9, family =
binomial(logit),
      data = tab_med)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.133900	0.646201	-1.755	0.07931 .
DEM	-0.000504	0.002308	-0.218	0.82712
DEM_northerness	-0.464284	0.184141	-2.521	0.01169 *
TPI_9	0.675968	0.234553	2.882	0.00395 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 395.88 on 351 degrees of freedom
Residual deviance: 380.28 on 348 degrees of freedom
AIC: 388.28
```

Number of Fisher Scoring iterations: 4

### **Backward Stepwise Selection (AIC)**

Call:

```
glm(formula = name ~ DEM_northerness + TPI_9, family =  
binomial(logit),  
data = tab_med)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.2719	0.1397	-9.102	< 2e-16 ***
DEM_northerness	-0.4590	0.1826	-2.514	0.01193 *
TPI_9	0.6698	0.2330	2.874	0.00405 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 395.88 on 351 degrees of freedom  
Residual deviance: 380.33 on 349 degrees of freedom  
AIC: 386.33

Number of Fisher Scoring iterations: 4

### **Backward Stepwise Selection (Schwartz's AIC - BIC)**

Call:

```
glm(formula = name ~ DEM_northerness + TPI_9, family =  
binomial(logit),  
data = tab_med)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.2719	0.1397	-9.102	< 2e-16 ***
DEM_northerness	-0.4590	0.1826	-2.514	0.01193 *
TPI_9	0.6698	0.2330	2.874	0.00405 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 395.88 on 351 degrees of freedom  
Residual deviance: 380.33 on 349 degrees of freedom  
AIC: 386.33

Number of Fisher Scoring iterations: 4

## Standardized Coefficients

North	TPI
-0.04853772	0.0544747

## Variance Inflation Factor

### Variance Inflation Factor (Multivariate)

DEM	DEM_northerness	TPI_9
1.032377	1.017676	1.014832

### Variance Inflation Factor (AIC)

DEM_northerness	TPI_9
1.000006	1.000006

### Variance Inflation Factor (BIC)

DEM_northerness	TPI_9
1.000006	1.000006

## Area under the ROC curve (AUC)

Call:

```
roc.default(response = me_BIC_glm$y, predictor =  
me_BIC_glm$fitted.values)
```

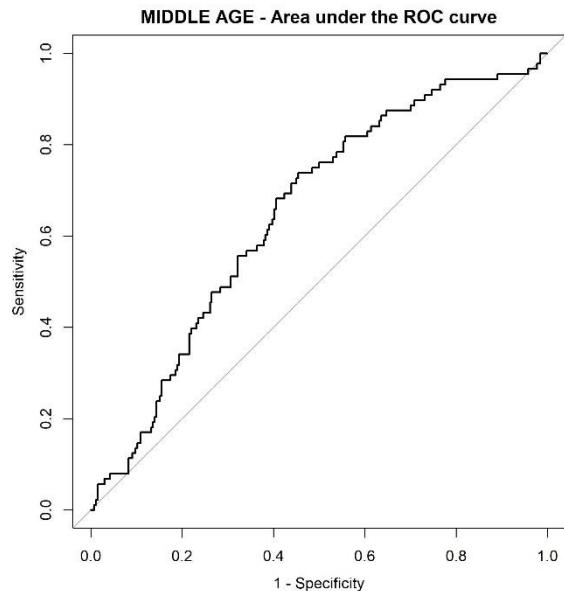
Data: me\_BIC\_glm\$fitted.values in 264 controls (me\_BIC\_glm\$y 0) <  
88 cases (me\_BIC\_glm\$y 1).

Area under the curve: 0.6493

## Discriminatory Ability

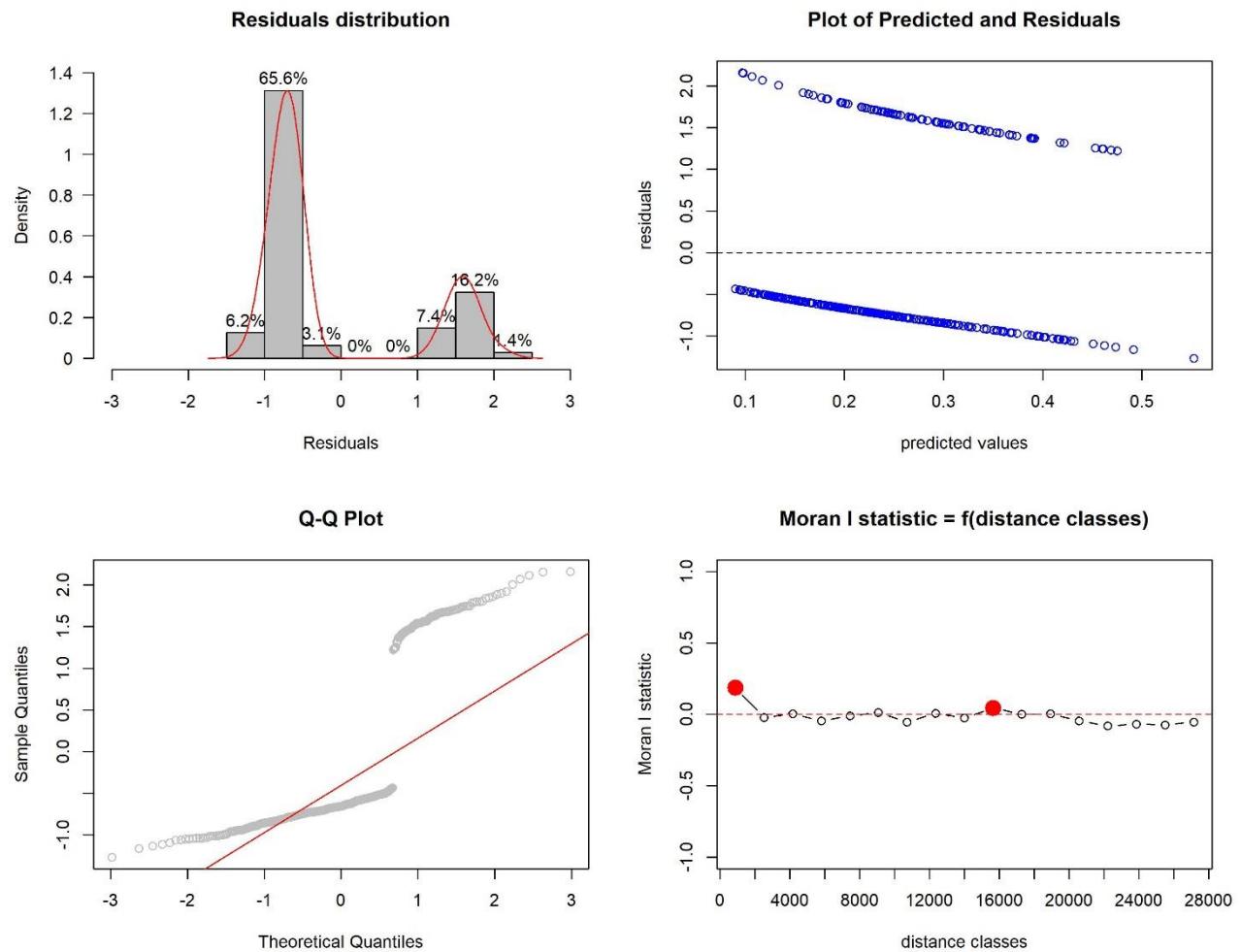
(HOSMER et. al. 2013 - Applied  
Logistic Regression - 3rd Ed.)

0.5 = No better than chance  
0.5-0.7 = Poor  
0.7-0.8 = Acceptable  
0.8-0.9 = Excellent  
0.9-1.0 = Outstanding

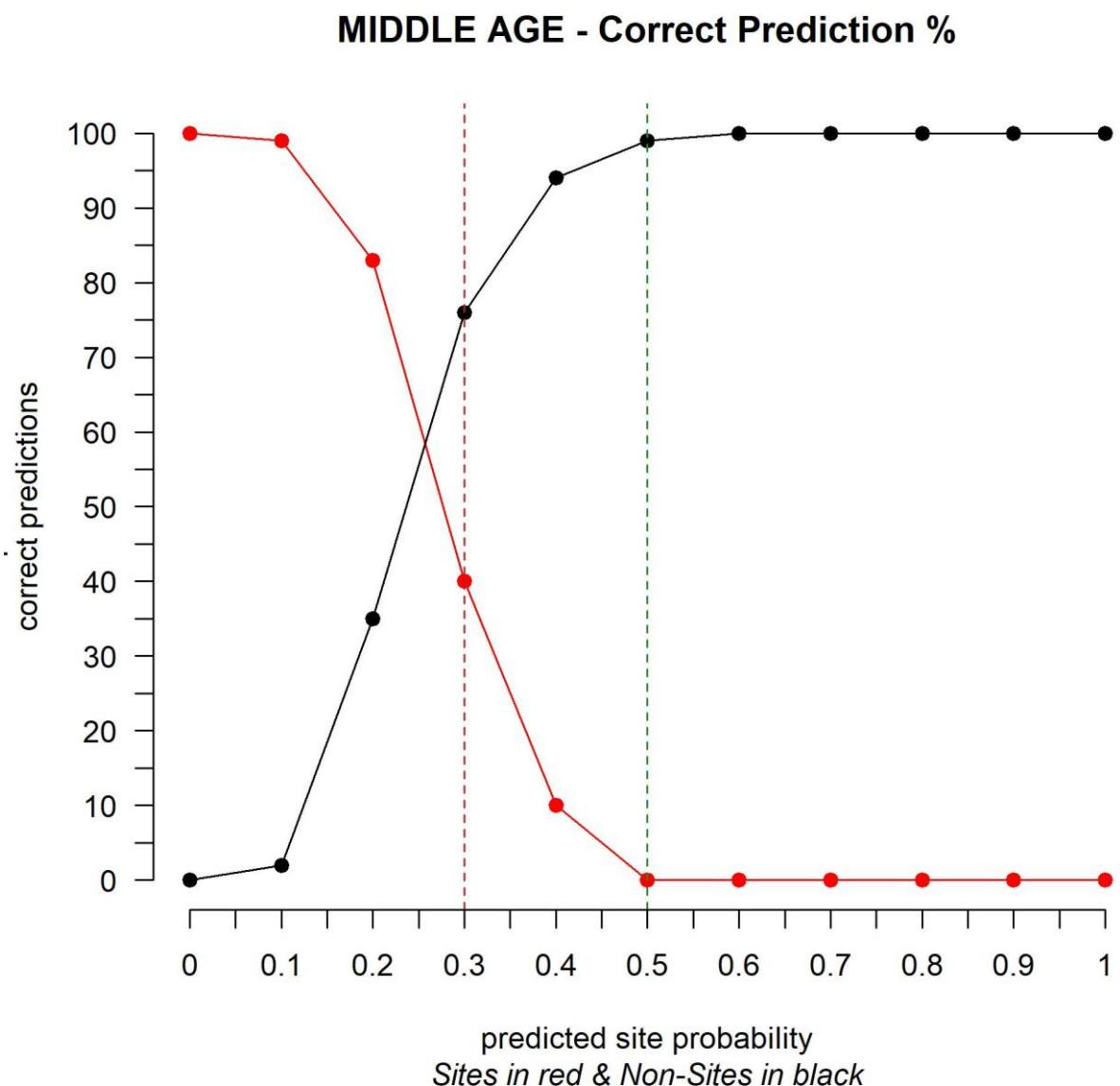


## Residuals

MIDDLE AGE - Multivariate Regression (Stepwise BIC) Residuals



**Correct prediction threshold**



## Model Assessment

-- Kolmogorov-Smirnov test --

Prediction values Sites/No-Sites

Bootstrap p-value: < 2.22e-16  
Naive p-value: 4.3116e-06  
Full Sample Statistic: 0.31439

-- Kvamme's Gain - Sites --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area	0.97	0.58	0.22	0.06	0.01	0	0	0	0
% Siti	0.99	0.83	0.40	0.10	0.00	0	0	0	0
Kvamme's Gain	0.02	0.30	0.45	0.40	-Inf	NaN	NaN	NaN	NaN

-- Kvamme's Gain - No Sites in predicted area --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area	0.97	0.58	0.22	0.06	0.01	0	0	0	0
% No Sites	0.98	0.65	0.24	0.06	0.01	0	0	0	0
Kvamme's Gain	0.01	0.11	0.08	0.00	0.00	NaN	NaN	NaN	NaN

-- Kvamme's Gain - No Sites in unpredicted area --

	>0.1	>0.2	>0.3	>0.4	>0.5	>0.6	>0.7	>0.8	>0.9
% Area Unpred	0.03	0.42	0.78	0.94	0.99	1	1	1	1
% No Sites	0.02	0.35	0.76	0.94	0.99	1	1	1	1
Kvamme's Gain	-0.50	-0.20	-0.03	0.00	0.00	0	0	0	0

## External Area test

-- Kolmogorov-Smirnov test --

h0 = how likely the samples are from same distribution

### BVER/Lecco

Bootstrap p-value: 0.4571  
Naive p-value: 0.51926  
Full Sample Statistic: 0.1131

### Como/Lecco

Bootstrap p-value: 0.0019  
Naive p-value: 0.0025949  
Full Sample Statistic: 0.2424

### BVER/Como

Bootstrap p-value: 1e-04  
Naive p-value: 0.00036346  
Full Sample Statistic: 0.2959

### Como/Garda

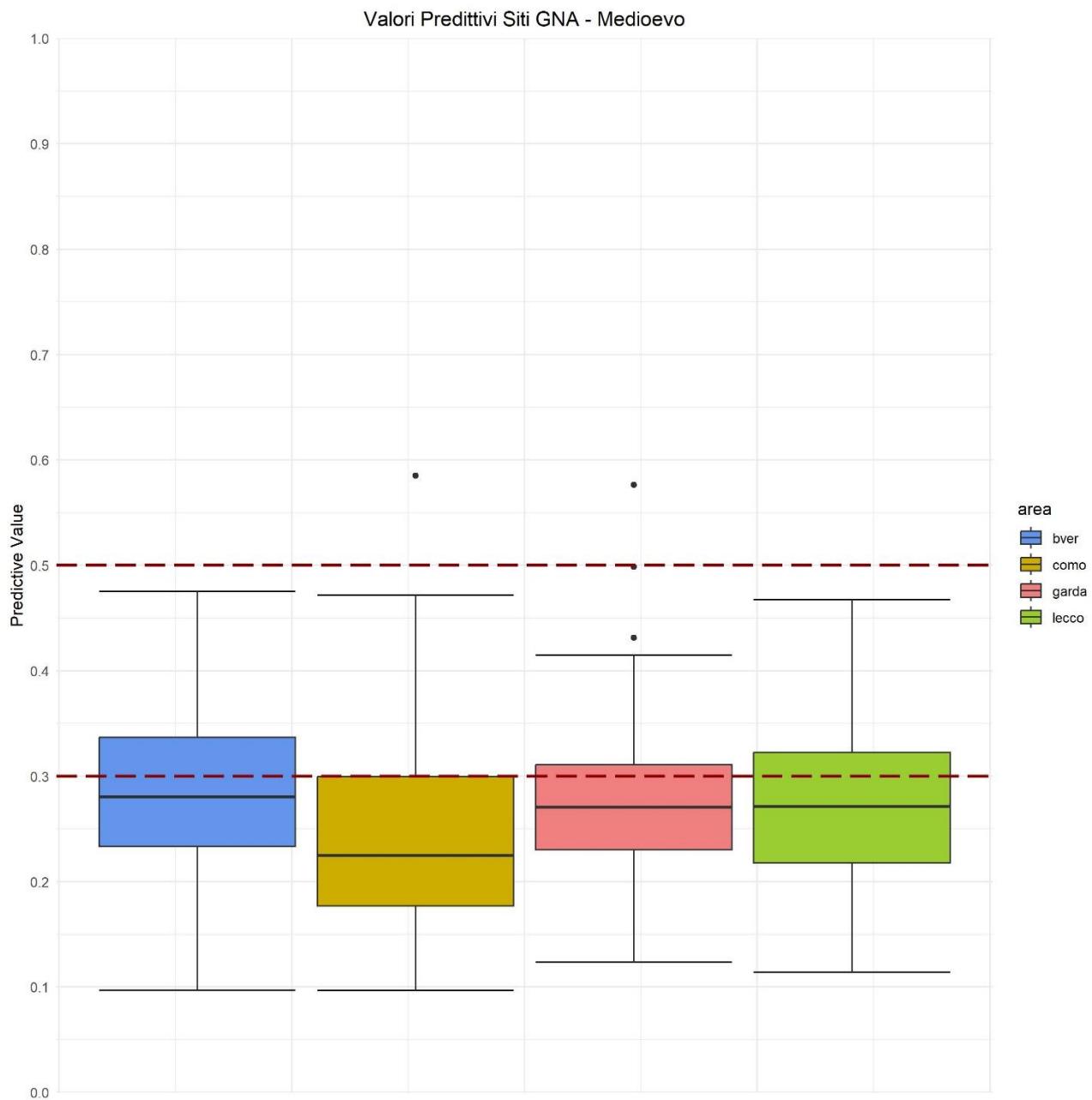
Bootstrap p-value: 0.005  
Naive p-value: 0.0045371  
Full Sample Statistic: 0.29592

### BVER/Garda

Bootstrap p-value: 0.6925  
Naive p-value: 0.72771  
Full Sample Statistic: 0.11688

### Lecco/Garda

Bootstrap p-value: 0.6784  
Naive p-value: 0.70847  
Full Sample Statistic: 0.11265



Predictive values of the Medieval sites in the BVER and control areas