One-class land-cover classification using MaxEnt: the effect of modelling parameterization on classification accuracy

Ignacio C Fernández [Corresp., 1] , **Narkis S Morales** [Corresp. 1]

1 Centro de Modelación y Monitoreo de Ecosistemas, Facultad de Ciencias, Universidad Mayor, Santiago, Chile

Corresponding Authors: Ignacio C Fernández, Narkis S Morales
Email address: ignacio.fernandez@umayor.cl, narkis.morales@umayor.cl

Multiple-class land-cover classification approaches can be inefficient when the main goal is to classify only one or a few classes. Under this scenario one-class classification algorithms could be a more efficient alternative. Currently there are several algorithms that can fulfil this task, MaxEnt being one of the most promising. However, there is scarce information regarding parametrization for performing land-cover classification using MaxEnt. In this study we aimed to understand how MaxEnt parameterization affects the classification accuracy of four different land-covers (i.e. built-up, irrigated grass, evergreen trees and deciduous trees) in the city of Santiago de Chile. We also evaluated if MaxEnt manual parameterization outperforms classification results obtained when using MaxEnt default parameters setting. To accomplish our objectives, we generated a set of 25,344 classification maps (i.e. 6,336 for each assessed land-cover), which are based on all the potential combination of 12 different classes of features restrictions, four regularization multipliers, four different sample sizes, three training/testing proportions, and 11 thresholds for generating the binary maps. Our results showed that with a good parameterization, MaxEnt can effectively classify different land covers with *kappa* values ranging from 0.68 for deciduous trees to 0.89 for irrigated grass. However, the accuracy of classification results is highly influenced by the type of land-cover being classified. Simpler models produced good classification outcomes for homogenous land-covers, but not for heterogeneous covers, where complex models provided better outcomes. In general, manual parameterization improves the accuracy of classification results, but this improvement will depend on the threshold used to generate the binary map. In fact, threshold selection showed to be the most relevant factor impacting the four land-cover classification accuracies. The number of sampling points for training the model also has a positive effect on classification results. However, this effect followed a logarithmic distribution, showing an improvement of *kappa* values when increasing the sampling from 40 to 60 points, but showing only a marginal effect if more than 60 sampling points are used. In light of these results, we suggest testing different parametrization and thresholds

until satisfactory *kappa* or other accuracy metrics are achieved. Our results highlight the huge potential that MaxEnt has a as a tool for one-class classification, but a good understanding of the software settings and model parameterization is needed to obtain reliable results.

One-class land-cover classification using MaxEnt: the effect of modelling parameterization on classification accuracy

Ignacio C. Fernández[1*] & Narkis S. Morales[1*]

[1]Centro de Modelación y Monitoreo de Ecosistemas, Facultad de Ciencias, Universidad Mayor,

Santiago, Chile.

*Corresponding authors:

Ignacio C. Fernández & Narkis S. Morales

José Toribio Medina 29, Región Metropolitana, Santiago, 8340589, Chile

Email address: ignacio.fernandez@umayor.cl, narkis.morales@umayor.cl

Abstract

1   Multiple-class land-cover classification approaches can be inefficient when the main goal is to

2   classify only one or a few classes. Under this scenario one-class classification algorithms could

3   be a more efficient alternative. Currently there are several algorithms that can fulfil this task,

4   MaxEnt being one of the most promising. However, there is scarce information regarding

5   parametrization for performing land-cover classification using MaxEnt. In this study we aimed

6   to understand how MaxEnt parameterization affects the classification accuracy of four different

7   land-covers (i.e. built-up, irrigated grass, evergreen trees and deciduous trees) in the city of

8   Santiago de Chile. We also evaluated if MaxEnt manual parameterization outperforms

9   classification results obtained when using MaxEnt default parameters setting. To accomplish

10   our objectives, we generated a set of 25,344 classification maps (i.e. 6,336 for each assessed

11   land-cover), which are based on all the potential combination of 12 different classes of features

12   restrictions, four regularization multipliers, four different sample sizes, three training/testing

13   proportions, and 11 thresholds for generating the binary maps. Our results showed that with a

14   good parameterization, MaxEnt can effectively classify different land covers with *kappa* values

15   ranging from 0.68 for deciduous trees to 0.89 for irrigated grass. However, the accuracy of

16   classification results is highly influenced by the type of land-cover being classified. Simpler

17   models produce good classification outcomes for homogenous land-covers, but not for

18   heterogeneous covers, where complex models provided better outcomes. In general, manual

19   parameterization improves the accuracy of classification results, but this improvement will

20   depend on the threshold used to generate the binary map. In fact, threshold selection showed

21   to be the most relevant factor impacting the four land-cover classification accuracies. The

22   number of sampling points for training the model also has a positive effect on classification

23   results. However, this effect followed a logarithmic distribution, showing an improvement of

24   *kappa* values when increasing the sampling from 40 to 60 points, but showing only a marginal

25   effect if more than 60 sampling points are used. In light of these results, we suggest testing

26   different parametrization and thresholds until satisfactory *kappa* or other accuracy metrics are

27   achieved. Our results highlight the huge potential that MaxEnt has a as a tool for one-class

28   classification, but a good understanding of the software settings and model parameterization is

29   needed to obtain reliable results.

33   **1. Introduction**

35   Land-cover analysis has become a fundamental approach for understanding, modelling and

36   monitoring the spatio-temporal patterns of ecosystems worldwide (Foley *et al.*, 2005; Ellis *et*

37   *al.*, 2010; Pielke *et al.*, 2011). Land-cover analysis relies on a variety of land classification

38   techniques, mostly based on computational algorithms, that commonly use remote sensing

39   images as a main input for discriminating among different land-cover classes (Rogan and Chen,

40   2003; Srivastava *et al.*, 2012). These classification techniques are aimed at translating the data

41   contained in the images into usable information, such as land-cover maps. However, as in any

42   modelling procedures, the quality and usefulness of the final outcome will depend on the

43   particular approach used for discriminating among classes of interest (Lu and Weng, 2007).

44   One of the most common techniques for land-cover classification is based on the

45   Maximum Likelihood (ML) algorithm, which is a supervised parametric pixel-based algorithm

46   that is available in most of current GIS or GIS-capable platforms, such as QGIS, ArcGIS and R.

47   The ML classification assumes that pixel values from a given layer will be normally distributed,

48   therefore each pixel can be classified into a given class based on the maximum likelihood to

49   pertain to the probability density function generated by the samples from that class (Strahler,

50   1980). Nevertheless, because ML is based on parametric statics, it has shown to be

51   outperformed by newer non-parametric algorithms that do not require the assumption of any

52   *a-priori* particular distribution or number of parameters, such as Neural Network and Support

53   Vector Machines (Khatami, Mountrakis and Stehman, 2016). Also, ML-based classifications

54   often require to train the algorithm on every single class covering the study area, implying that

55   many times resources and efforts will be placed on classifying classes that are not relevant for

56   the study objectives (Sanchez-Hernandez, Boyd and Foody, 2007; Li and Guo, 2010; Deng *et al.*,

57   2018). In fact, multi-class oriented classifiers, such as ML, may attempt to optimize the

58   classification accuracy of all cover classes rather than focusing on increasing the accuracy of the

59   target class (Sanchez-Hernandez, Boyd and Foody, 2007). Therefore, the use of one-class

60   classification algorithms could provide a more efficient alternative than multi-classes

61   approaches when the objective is classifying only one or a few classes of interest (Deng *et al.*,

62   2018).

63   Among one-class classification algorithms, MaxEnt is probably one of the most used

64   nowadays. MaxEnt is a maximum entropy algorithm-based software that was originally

65   developed to be used for species distribution modelling (Phillips and Dudík, 2008; Elith *et al.*,

66   2011). Nevertheless, its built-in algorithm can be easily used for one-class classification of a

67   variety of land-use or -cover types. For example, MaxEnt has been used for mapping urban

68   land-use classes in California (Li and Guo, 2010), urban land in China (Lin *et al.*, 2014), cover of

69   invasive plant species in Colorado (Evangelista *et al.*, 2009) and California (Skowronek, Asner

70   and Feilhauer, 2017), and conservation habitats (Stenzel *et al.*, 2014) and raised bogs in

71   Southern Germany (Mack *et al.*, 2016). The increasing popularity of MaxEnt for one-class land-

72   cover classification is probably due to the fact that is a freely available software

73   (https://biodiversityinformatics.amnh.org/open_source/maxent), has shown as good or better

74   performance than other one-class classification methods (e.g. Li & Guo, 2010, Mack *et al.*,

75   2016), and comes with default parameters settings that facilitates its operation (Mack and

76   Waske, 2017).

77   Nevertheless, while the availability of predefined default parameters may have

78   promoted MaxEnt massification, the use of this feature can also result in suboptimal land-cover

79   classification outcomes. For example, recent studies from species distribution modelling show

80   that MaxEnt default parameters seldom generate optimal outcomes (Shcheglovitova and

81   Anderson, 2013; Syfert, Smith and Coomes, 2013; Morales, Fernández and Baca-González,

82   2017). In relation of using MaxEnt default parameter for land-cover classification, the few

83   published studies of which we are aware, also report that MaxEnt tend to produce better

84   classification outcomes when users select the optimal parameters (Mack and Waske, 2017;

85   Skowronek *et al.*, 2018).

86   There are several default parameters that are modifiable in MaxEnt. The two main

87   modifiable parameters are feature classes (FC) and regularization multiplier (RM). FC

88   corresponds to a set of mathematical transformation of the different covariates used in the

89   model, while RM refers to a numerical parameter that reduces or increases the smoothness of

90   the model (Elith *et al.*, 2011; Merow, Smith and Silander, 2013). Other additional settings, such

91   as sample size (SS), test percentage (TP) and binary threshold (BT), could also influence the final

92   classification outcome. SS refers to the number of samples used for building the model.

93   Although MaxEnt seems to be only slightly sensitive to sample size, larger sample sizes has

94   shown to produce better modeling results (Wisz *et al.*, 2008). TP refers to the proportion (or

95   percentage) of the sample size used for testing the model, which will be the inverse of the

96   proportion used for training. Therefore, TP can affect the model results by indirectly modifying

97   the model sample size. BT corresponds to the specific threshold value used for transforming the

98   continuous probabilistic outputs values into a binary (positive/negative) classification results.

99   The chosen threshold will largely affect the classification results, nevertheless finding an

100   optimal BT for land-cover classification is still one of the main challenges for using MaxEnt as a

101   one-class classification algorithm (Mack *et al.*, 2016; Mack and Waske, 2017).

102   Even though MaxEnt has shown promising results as a tool for one-class classification,

103   there is still a big knowledge gap regarding how classification results are affected by the

104   parameter settings used for modeling. Thus, in this work we aim to: (1) assess the effects of five

105   different parameter settings (i.e. FC, RM, SS, TP, BT) combinations on MaxEnt-based land-cover

106   classification outcomes; (2) evaluate if these effects depend on the land-cover under analysis;

107   (3) identify potential best parameter setting combination for different land-cover classes and

108   compare its classification accuracy with the outcomes of a default parameter setting

109   classification.

110

111   **2.   Material and Methods**

112

113   **Study area**

114

115   We developed our study within the metropolitan area of Santiago, Chile (33.4489° S, 70.6693°

116   W). Santiago is the capital and largest urban area of Chile, harboring an estimated population of

117   5.3 million inhabitants (Instituto Nacional de Estadísticas, 2018). The urban area covers a

118   surface of 875 km^2 (de la Barrera and Henríquez, 2017), with an elevation ranging from 450 to

119   1000 m above sea level. The climate within the study area is classified as Mediterranean, with

120   marked colder temperatures and rainy periods during winter months, and warmer and dry

121   period during summer (Luebert and Pliscoff, 2006). Historical annual precipitations average

122   ~340 mm, with more than 88% falling from May to September, and less than 2% precipitates

123   during the three warmest months (December to February). Seasonal variations in temperatures

124   and precipitation generate noticeable phenological changes in Santiago´s urban vegetation,

125   making it feasible to discriminate among different vegetation types from airborne or

126   spaceborne images (Fernandez and De la Barrera, 2018).

127

128   **Remote sensing layers**

129

130   To perform the land-cover classification we used Sentinel-2 mission satellite images with 0%

131   cloud cover, representing vegetation conditions of summer (March 06, 2016) and winter

132   (August 02, 2016) for the study area. To increase the resolution of our analysis, from the set of

133   summer and winter Sentinel-2 images, we only used bands with a native resolution of

134   10m/pixel (i.e. red, green, blue, NIR). Based on these bands, we also calculated the Normalized

135   Difference Vegetation Index (NDVI) for summer and winter. We also generated an additional

136   layer based on the arithmetic difference between the summer and winter NDVI (i.e. seasonal

137   difference of NDVI). Even though MaxEnt is regarded to be only slightly sensitive to potential

138   over-parameterization issues (Warren and Seifert, 2011), it is recommended to minimize

139   correlation among predictor variables (Merow, Smith and Silander, 2013). Therefore, we

140   performed a Pearson correlation analysis for the 11 layers (i.e. red, green, blue, and NIR, for

141   summer and winter, and the three NDVI layers), and we kept only those layers with correlations

142   smaller than 0.8. After this analysis we ended up with seven layers: Red (summer/winter), NIR

143   (summer/winter), NDVI (summer/winter) and the seasonal difference of NDVI. All remote-

144   sensing layers procedures were done in QGIS 2.18 Las Palmas (www.qgis.org).

145

146   **Land-cover classification (Maxent modelling)**

147

148   We assessed Maxent one-class classification performance by focusing on four main land-cover

149   types present in Santiago: (1) Built-up infrastructure, represented by roads and other paved

150   structures, houses, buildings and commercial infrastructure. (2) Irrigated grass, mostly

151   associated to private and public lawns, sport fields and other summer irrigated grasses. (3)

152   Evergreen trees, related to native and exotic urban shrubs and trees with all-year around leaves

153   and photosynthetic activity. (4) Deciduous trees, mostly winter deciduous exotic shrubs and

154   trees planted on residential gardens, streets and urban parks.

155   To perform the classification modelling we started by selecting 100 ground-true points

156   for each land-cover class by using very-high resolution images (<1m/pixel) available through

157   Google Earth. Because potential spatial mismatches between Google and Sentinel images due

158   to georeferencing and resolution differences, we only selected points that were within a 5 m

159   radius of the type of land-cover under analysis (i.e. minimum size 1 pixel). Each set of points

160   was doubled checked independently by each author. To assess the effect of MaxEnt

161   parametrization on land-cover classification, we generated a set of 25,344 classification maps

162   (i.e. 6,336 for each of the four assessed land-covers), which are based on all the potential

163   combination of the five analyzed parameters (Table 1). These combinations included first, 12

164   different types of restrictions ("feature class") in the modelling stage based on the combination

165 of lineal (L), hinge (H), quadratic (Q), threshold (T) and product (P). We used all the

166 computational possible combinations of these features. Second, four regularization multipliers

167 (i.e. 0.25, 1, 3 and 5). Third, four different sample sizes (i.e. 40, 60, 80 and 100). For the later,

168 samples smaller than 100 were randomly sampled from the original 100 sampling points.

169 Fourth, three different percentages of testing proportion to validate the model (i.e. 10, 30, 50).

170 Fifth, 11 binary thresholds for generating the final binary classification maps: Fixed cumulative

171 value 1 threshold (FC1); fixed cumulative value 5 threshold (FC5); fixed cumulative value 10

172 threshold (FC10); 10 percentile training presence threshold (10PTP); balance training omission,

173 predicted area and threshold value threshold (BTOPA); equate entropy of thresholded and

174 original distributions threshold (EETOD); equal test sensitivity and specificity threshold (ETeSS);

175 equal training sensitivity and specificity threshold (ETrSS); maximum test sensitivity plus

176 specificity threshold (MTeSPS); minimum training presence threshold (MTP); maximum training

177 sensitivity plus specificity threshold (MTrSPS) (Table 1). All MaxEnt models were run using the

178 random seed feature and replicated 10 times, and we used the average values of these

179 replicates as the modeling outcomes. Background points were set to default (maximum 10000

180 points). All modelling procedures were performed using MaxEnt 3.4.1

181 (https://biodiversityinformatics.amnh.org/open_source/maxent). Application of thresholds and

182 generation of final binary maps were done in R version 3.5.1 (www.r-project.org).

183

184 **Classification Accuracy**

185

186   We evaluated the accuracy of the classification result by estimating the *kappa* coefficient. To

187   calculate *kappa,* we randomly sampled 1000 points within a quadrant of 16 x 16 km, which is

188   the largest square that can be drawn within the convoluted shape of Santiago. We decided to

189   use this quadrant to avoid potential undesirable effects of choosing sampling points outside

190   Santiago´s urbanized area and to reduce the size of files to increase the efficiency of the

191   computational processes. Each of these 1000 points were visually classified into built-up,

192   irrigated grass, evergreen trees, deciduous trees and "other class" by using very-high resolution

193   images through Google Earth. If a sampling point fell on mixed land-cover (e.g. transition

194   between two or more land covers) we moved the point to the closest single-land cover surface

195   representing the dominant land-cover from the original location. We considered mixed land-

196   covers all points falling in locations with more than one land-cover within a 5 m radius. All

197   points were double checked by the authors. We then used the 1000 sampling points to build

198   four different testing layers (one for each assessed land-cover) based on positive and negative

199   labels for each land-cover. We used these testing layers to calculate the *kappa* coefficient for

200   each classification result based on the proportion of true positive, false positive, true negative

201   and false negative, following the formula shown in Fielding and Bell (1997).

202   We compared the classification accuracy of the different combinations through

203   integrated violin-boxplot graphs for the *kappa* coefficient, which facilitates the visual

204   comparison of differences on data distribution, interquartile and median values. All

205   classification accuracy procedures and graphs were done in R version 3.5.1.

206

207   **Comparison of default parameter vs best parameters settings**

208

209   We defined the best combination of parameter for each land-cover type by looking at the

210   *kappa´s* median values of the violin-boxplots of feature classes, regularization multipliers,

211   sample size and test percentage (See Figs. 1,2,3 and 4). Based on this result, we decided to use

212   a SS of 100 sampling points and a TP of 10% for all classification models (i.e. default and best

213   parameter settings), only modifying those parameters directly related to the modeling process,

214   which are the FC and RM. Therefore, for the default parameter setting models we used the

215   "autofeature" FC with a RM of 1, whereas for the best parameter setting we used the FC and

216   RM showing the highest median *kappa* for each land-cover type (See Figs. 1 and 2). Using these

217   parameters, we ran 30 replicates for each classification model in MaxEnt. To generate the final

218   binary maps in R, we did not select a specific binary threshold because we did not know the

219   effect of this parameter on the default setting models results. Instead, we compared modeling

220   results by using all the BT provided in MaxEnt. We used a conventional student t-test to

221   evaluate statistically significant differences between the default and best parameter settings

222   for each land-cover type and binary threshold.

223

224   **3.   Results**

225

226   *Parameterization effects on classification accuracy*

227

228   Results show that changing the feature classes can have an important effect on the final land-

229   cover classification accuracy, and that these effects change depending on the land-cover that is

230   being classified (Fig. 1). For example, while for built-up land cover the L and Q features

231   produced the highest *kappa* (measured as *kappa´s* median value), these same features

232   produced the poorest results for the three vegetated covers (i.e. irrigated grass, evergreen

233   trees and deciduous trees). Furthermore, the FC for which the *kappa* median was largest,

234   fluctuated for the four land-covers: L for built-up, LQPTH for irrigated grass, QP for evergreen

235   trees and QH for deciduous trees, with median *kappa´s* of 0.695, 0.693, 0.577 and 0.528,

236   respectively. There are also important differences in the variability (i.e. density distribution) of

237   *kappa* results associated to the different FC. For example, based on the differences between

238   the first and third quartile, built-up, evergreen trees and deciduous trees covers show the

239   largest dispersion of kappa values for T, Q, and TQ, but for irrigated grass all FC show very large

240   dispersion, except for L and Q (See Fig. 1).

241   Modifying the regularization multiplier could also have important effect on classification

242   results, but as our results show, these effects could considerably differ depending on the land-

243   cover type being classified (Fig. 2). While for built-up, increasing the RM from 0.25 to 1 seems

244   to slightly improve classification results, the opposite pattern is observed for the three

245   vegetated covers. In addition, while further increases (> 1) in RM do not improve classification

246   outcomes for built-up cover, they do improve the classification for vegetated covers, as is

247   shown by the consistent increase in *kappa* with larger RM. Among the three vegetated covers

248   assessed in this work, irrigated grass cover showed by far the largest response to changes in RM

249   (Fig. 2).

250   Changes in sample size show only a slight effect on classification results (Fig. 3).

251   Although there are some differences between the four assessed land-covers, there is a

252 consistent pattern showing that the lowest SS (i.e. 40 samples) generated the poorest

253 classification results. For all assessed land covers, increasing the SS from 40 to 60 improved the

254 classification. However, results show that increases in SS above 60 samples do not necessarily

255 generate large improvements on classification results (Fig. 3).

256   Changing the test percentage for validating the model has only weak effects on

257 classification results. In fact, results show that for built-up cover this setting seems to not affect

258 the median value of *kappa* nor the *kappa´s* median distribution, whereas for evergreen trees

259 and deciduous trees it is observable a slight, but consistent decrease in classification results as

260 TP increase to 50% (Fig. 4). The larger effect of TP is shown for irrigated grass cover, which is

261 consistent with the results shown for changes in the SS setting.

262   The selection of binary threshold values is the setting that had the largest effects on

263 classification results among the five assessed settings (Fig. 5). Using one BT instead of another

264 could greatly affect the classification results. For example, depending on the selected threshold,

265 *kappa´s* median for irrigated grass cover could range from 0.134 to 0.811 (Fig. 5b). The effect of

266 BT on classification shows similar patterns for irrigated grass, evergreen trees and deciduous

267 trees, but not for built-up cover. For instance, while for built-up cover the FC1 threshold

268 generated one of the best classification results (*kappa´s* median = 0.766), for the three

269 vegetated covers this same threshold produced the worst results. At the other hand, 10PTP

270 produced the best classification results for irrigated grass and evergreen trees, and the second

271 best for deciduous trees, but produced poor classification results for built-up cover (Fig. 5).

272 *Default parameter vs best parameters setting classification accuracy*

273

274   The comparison of the models produced using the default parameter setting for FC and RM

275   with those using the identified best parameters (defined as FC and RM having the largest *kappa*

276   median for each land-cover type, i.e. FC=L, RM=1 for built-up; FC=LQPTH, RM=0.25 for irrigated

277   grass; FC=QP, RM=0.25 for evergreen trees; FC=QH, RM=0.25 for deciduous trees; See Figs. 1

278   and 2) show that the identified best parameter setting does not always outperform the

279   classification accuracy of the default setting model (Fig. 6). Considering the four assessed land-

280   cover types and the 11 thresholds, the best parameters outperformed the default parameters

281   in 23 out of the 44 classification result comparisons (53,3%), while the identified best

282   parameter was outperformed by the default parameter in 11 cases (25,0%). In 10 comparisons

283   (22,7%) there were no statistical differences between both parameters settings (Fig. 6). The

284   difference in classification results between the default and identified best parameters is highly

285   dependent on the land-cover under analysis and the threshold used for generating the binary

286   maps. For example, for evergreen trees, in 9 out of 11 comparison the best parameter setting

287   outperformed the default model and for no threshold the default model resulted in better

288   classification. On the other hand, for irrigated grass the default model outperformed the

289   identified best parameter setting in five out of 11 comparisons, while the opposite results were

290   obtained 6 times (Fig. 6). In general, BT selection has a much larger effect than parameter

291   settings in classification results.

292

293   **4.   Discussion**

294

295   Results from our work add to previous studies showing the potential of MaxEnt to be used as a

296   one-class land-cover classification method (e.g. Li and Guo, 2010; Amici, 2011; Lin *et al.*, 2014;

297   Stenzel *et al.*, 2014), but also draw attention to the importance of selecting best parameters

298   combinations to achieve good classification accuracies (Mack and Waske, 2017). While our

299   main objective was to understand how MaxEnt parameterization may affect classification

300   results, we were also able to test the capability of MaxEnt for one-class land-cover

301   classification, which has never been tested in an entire urban area the size of Santiago.

302   Our classification accuracy shows *kappa* values above 0.6 for all the assessed land-

303   covers -and even above 0.8 for built-up and irrigated grass. These *kappa* values are comparable

304   to those obtained by Li and Guo (2010) using Maxent to classify similar land-covers in a urban

305   plot in a residential area of California, and also to those obtained by Deng et al. (2018) using

306   other state of the art classification algorithms to classify land covers in a similar urban setting in

307   California. Thus, results from our study highlight the potential of MaxEnt and its built-in

308   maximum entropy algorithm to be used for one-class classification in complex spatially

309   heterogeneous settings, such as urban areas.

310

311   *Parameterization effects on classification accuracy*

312

313   Results from our work show that MaxEnt parameterization plays an important role on land-

314   cover classification accuracy, and therefore is highly relevant for achieving good classification

315   results. However, based in our results, is not clear if there are some specific feature classes that

316   can provide higher quality classifications. On the contrary, it seems that the optimal feature

317   classes to use will depend on the intrinsic spatial variability of the layers used to build the

318   models. For example, in our study, the L and Q (i.e. lineal and quadratic) feature classes

319   generate among the best classification accuracy for built-up, but the poorest results for

320   vegetated covers. This result could be explained because the Q feature constrains the variation

321   of the predictor variable to that of the sampling points of each layer (See Elith *et al.*, 2011;

322   Merow, Smith and Silander, 2013 for an explanation of MaxEnt features). Thus, the Q feature

323   may differentially affect the specificity of the resulting model depending on the true variance of

324   the population, benefiting classification results of spatially homogenous covers, such as built-

325   up, but weakening results for more heterogeneous covers, such as irrigated grass. A similar

326   interpretation could be made for the L feature, as this feature simply uses the mean of the

327   sampling points to estimate the conditions making a pixel suitable to be classified into a given

328   class. These results suggest that MaxEnt models based on simpler parameter combination could

329   be well suited for classifying land-covers with more homogenous reflectance behaviors, such as

330   built-up areas, while models of higher complexity would be needed for vegetated or other

331   more heterogeneous areas. This could be further supported by the results of the regularization

332   multiplier, that show that for vegetated land-covers accuracy increases if regularization

333   multiplier is reduced to fit more complex models, whereas built-up classification accuracy

334   benefits from larger regularization that smooths out the model fit.

335   While sample size may not be directly considered a model parameter, it is can have a

336   direct effect on MaxEnt behavior, particularly if default features are used. In fact, MaxEnt will

337   restrict the use of potential feature classes depending of the number of sampling points (Elith

338   *et al.*, 2011; Merow, Smith and Silander, 2013), therefore models generated with a small

339   number of samples will be restricted in their complexity. This may explain why classification of

340   the three vegetated land-covers are more affected than built-up cover by the number of

341   sampling points used for training the model. Nevertheless, for the four land-covers, an effective

342   sampling size larger than ~80 seems to be optimal, which is consistent with the sampling size

343   threshold at which MaxEnt is set to use all potential FC (Elith *et al.*, 2011).

344   A key result from our study was that the selection of the binary threshold for building

345   the classified maps is perhaps the most crucial factor related to the accuracy of the final

346   classification outcome. This threshold is not a MaxEnt modeling parameter, but rather *a*

347   *posteriori* decision that needs to be taken based on modeling results. While MaxEnt provides 11

348   potential thresholds (the 11 used in this study), the final decision on what threshold to use will

349   mostly depend on the researcher or group of researchers performing the classification.

350   Previous studies have pointed out to the difficulty and lack of information for finding an optimal

351   threshold to maximize the classification accuracy using MaxEnt (Li and Guo, 2010; Mack *et al.*,

352   2016; Mack and Waske, 2017), and based in our results it seems that this quest is still open. In

353   fact, none of the tested thresholds performed well for all the assessed land-covers, implying

354   that users will need to explore a set of thresholds to find one that produce optimal

355   classification accuracy for each land-cover.

356

357   *Default parameter vs best parameters setting classification accuracy*

358

359   There is increasing evidence that using best, instead of default parameters, could improve the

360   predictability of species distribution models produced through MaxEnt (Shcheglovitova and

361   Anderson, 2013; Morales, Fernández and Baca-González, 2017), and similar results have

362   recently been found when using MaxEnt for land-cover classification (Mack and Waske, 2017).

363   However, our results show that although in general best parameters outperformed the default

364   parameters, for some land-covers and thresholds, default parameters can produce similar or

365   even better classification results than the identified best parameters. Technically this sounds

366   illogical, because a manually parameterized model can theoretically use the same parameters

367   of the default model, and therefore if best model parameters are found, classification results

368   should be at least equally good as the one obtained by the default model. Nevertheless, finding

369   the best parameters combination is still a challenging task (Muscarella *et al.*, 2014). For

370   example, to find the optimal parameters combination for our study, we run thousands of

371   models and selected each parameter based on the median value of resulting *kappas,* assuming

372   that general central tendencies could be indicative of parameters performances. However, this

373   approach ruled out potential parameters combinations that produced higher *kappas* (See

374   supplementary material), which may help explaining why in our study for some comparisons

375   the default parameter had better classification accuracies than the best model parameter.

376   An interesting finding from our study was the fact that threshold selection do not only

377   affect the accuracy of classification results (Mack and Waske, 2017), but that a same threshold

378   could differentially impact on the accuracy of different models (Fig. 6a). These findings show

379   that just using the model performance information, such as AUC (Area Under the Curve), for

380   evaluating the accuracy of a land-cover classification outcomes is not advisable.

381

382   *Best parametrization practices*

383

384   In the light of our results and other articles regarding the use of MaxEnt, we would like to

385   highlight the importance of parameterizing MaxEnt for each model individually and avoid the

386   temptation to of using default parameter or previous studies parametrizations (Merow, Smith

387   and Silander, 2013; Mack and Waske, 2017; Morales, Fernández and Baca-González, 2017). As a

388   rule of thumb, a comparison of different models with different parametrization settings must

389   be performed to determine the best model among all the evaluated parameters combinations.

390   Ideally, selection of parameters combination to be tested should not be random, but rather on

391   the specific objectives and land-cover under analysis. For example, for contrasting and

392   homogenous land-cover such as paved areas, using combinations that generate simpler models

393   could be an efficient approach. Whereas for less contrasting and heterogeneous land-covers,

394   such as irrigated grass, it would be recommended to increase the number of tested

395   combinations to include more complex models. There are several techniques available for

396   selecting MaxEnt best models, including jackknife procedures based on the corrected Akaike

397   information criterion (AICc) (Shcheglovitova and Anderson, 2013; Morales, Fernández and Baca-

398   González, 2017), correlation analysis (Syfert, Smith and Coomes, 2013), fuzzy *kappa* statistics

399   (Mestre *et al.*, 2015) and software including specific algorithms for this objective (Warren, Glor

400   and Turelli, 2010; Muscarella *et al.*, 2014). Nevertheless, once the parametrization of the model

401   is done, thresholds to generate binary maps still need to be chosen. As our results show,

402   threshold selection could be one of the most important factors influencing the final

403   classification accuracy. In this regard, we recommend not to use predefined thresholds or

404   thresholds used in previous studies, but instead try a set of different thresholds and select the

405   one that gives the best outcome. A viable option could be using a set of different thresholds

406   based on cumulative percentiles and select the generated model with the largest classification

407   accuracy (Lin *et al.*, 2014). Based in our results, percentiles between 1 and 10 seems to be a

408   good testing ground, as best classification outcomes tend to be produced by thresholds within

409   this percentile range. Nevertheless, further research is needed to better understand how

410   variation in threshold values affects the accuracy of land-cover classification produced through

411   MaxEnt.

412

413   **5.   Conclusion**

414

415   Our results showed that MaxEnt can be a useful tool for single-class land-cover classification.

416   However, the classification accuracy heavily depends on land-cover characteristics and

417   parameters used for modelling. Based in these results, we suggest avoiding the temptation of

418   using MaxEnt default parameterization features in heterogeneous layers, as manual

419   parameterization could generate better land-classification accuracy outcomes. In addition,

420   selection of optimal binary thresholds is key for increasing classification accuracy when using

421   binary maps. Nevertheless, for this is still essential to use a classification accuracy metric for

422   testing modelling results. In light of these results, we suggest testing different parametrization

423   settings and thresholds until satisfactory *kappa* or other accuracy metrics are achieved. Finally,

424   is also important to bear in mind that the reliability of the selected model will be also

425   influenced by the accuracy metric used for classification testing. In this regard, for specific

426   classification tasks, it would be recommendable to use multiple accuracy metrics against which

427   to compare the resulting models.

428

429   **References**

430   Amici, V. (2011) 'Dealing with vagueness in complex forest landscapes: A soft classification
431   approach through a niche-based distribution model', *Ecological Informatics*, pp. 371–383. doi:
432   10.1016/j.ecoinf.2011.07.001.
433   Deng, X., Li, W., Liu, X., Guo, Q. and Newsam, S. (2018) 'One-class remote sensing classification:
434   One-class vs. Binary classifiers', *International Journal of Remote Sensing*. Taylor & Francis, 39(6),
435   pp. 1890–1910. doi: 10.1080/01431161.2017.1416697.
436   Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. and Yates, C. J. (2011) 'A statistical
437   explanation of MaxEnt for ecologists', *Diversity and Distributions*, 17(1), pp. 43–57. doi:
438   10.1111/j.1472-4642.2010.00725.x.
439   Ellis, E. C., Klein Goldewijk, K., Siebert, S., Lightman, D. and Ramankutty, N. (2010)
440   'Anthropogenic transformation of the biomes, 1700 to 2000', *Global Ecology and Biogeography*,
441   19, pp. 589–606. doi: 10.1111/j.1466-8238.2010.00540.x.
442   Evangelista, P. H., Stohlgren, T. J., Morisette, J. T. and Kumar, S. (2009) 'Mapping invasive
443   tamarisk (Tamarix): A comparison of single-scene and time-series analyses of remotely sensed
444   data', *Remote Sensing*, pp. 519–533. doi: 10.3390/rs1030519.
445   Fernandez, I. C. and De la Barrera, F. (2018) 'Biodiversidad urbana , servicios ecosistémicos y
446   planificación ecológica : un enfoque desde la ecología del paisaje', in Figueroa, J. and Lazzoni, I.
447   (eds) *Biodiversidad Urbana en Chile: Estado del arte y los desafíos futuros*. First. Santiago, Chile:
448   Universidad Central de Chile, pp. 113–146.
449   Fielding, A. H. and Bell, J. F. (1997) 'A review of methods for the assessment of prediction errors
450   in conservation presence/absence models', *Environmental Conservation*, 24, pp. 38–49.
451   Foley, J. A., Defries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe,
452   M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. a, Kucharik, C. J.,
453   Monfreda, C., Patz, J. a, Prentice, I. C., Ramankutty, N. and Snyder, P. K. (2005) 'Global
454   consequences of land use.', *Science*, 309(570), pp. 570–574. doi: 10.1126/science.1111772.
455   Instituto Nacional de Estadísticas (2018) *Población por sexo y grupo de edad*, *Censo de
456   Población y Vivienda 2017*. Available at: https://redatam-
457   ine.ine.cl/redbin/RpWebEngine.exe/Portal?BASE=CENSO_2017&lang=esp (Accessed: 9
458   September 2018).
459   Khatami, R., Mountrakis, G. and Stehman, S. V. (2016) 'A meta-analysis of remote sensing
460   research on supervised pixel-based land-cover image classification processes: General
461   guidelines for practitioners and future research', *Remote Sensing of Environment*. Elsevier Inc.,
462   177, pp. 89–100. doi: 10.1016/j.rse.2016.02.028.
463   de la Barrera, F. and Henríquez, C. (2017) 'Vegetation cover change in growing urban
464   agglomerations in Chile', *Ecological Indicators*. Elsevier, 81, pp. 265–273. doi:
465   10.1016/j.ecolind.2017.05.067.

Li, W. and Guo, Q. (2010) 'A maximum entropy approach to one-class classification of remote sensing imagery', *International Journal of Remote Sensing*, 31(8), pp. 2227–2235. doi: 10.1080/01431161003702245.

Lin, J., Liu, X., Li, K. and Li, X. (2014) 'A maximum entropy method to extract urban land by combining MODIS reflectance, MODIS NDVI, and DMSP-OLS data', *International Journal of Remote Sensing*, 35(18), pp. 6708–6727. doi: 10.1080/01431161.2014.960623.

Lu, D. and Weng, Q. (2007) 'A survey of image classification methods and techniques for improving classification performance', *International Journal of Remote Sensing*, 28(5), pp. 823–870. doi: 10.1080/01431160600746456.

Luebert, F. and Pliscoff, P. (2006) *Sinopsis bioclimática y vegetacional de Chile*, *Biodiversidad*. Santiago, Chile: Editorial Universitaria.

Mack, B., Roscher, R., Stenzel, S., Feilhauer, H., Schmidtlein, S. and Waske, B. (2016) 'Mapping raised bogs with an iterative one-class classification approach', *ISPRS Journal of Photogrammetry and Remote Sensing*, pp. 53–64. doi: 10.1016/j.isprsjprs.2016.07.008.

Mack, B. and Waske, B. (2017) 'In-depth comparisons of MaxEnt, biased SVM and one-class SVM for one-class classification of remote sensing data', *Remote Sensing Letters*. Taylor & Francis, 8(3), pp. 290–299. doi: 10.1080/2150704X.2016.1265689.

Merow, C., Smith, M. J. and Silander, J. A. (2013) 'A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter', *Ecography*, 36(10), pp. 1058–1069. doi: 10.1111/j.1600-0587.2013.07872.x.

Mestre, F., Pita, R., Paupério, J., Martins, F. M. S., Alves, P. C., Mira, A. and Beja, P. (2015) 'Combining distribution modelling and non-invasive genetics to improve range shift forecasting', *Ecological Modelling*. Elsevier B.V., 297, pp. 171–179. doi: 10.1016/j.ecolmodel.2014.11.018.

Morales, N. S., Fernández, I. C. and Baca-González, V. (2017) 'MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review', *PeerJ*, 5, p. e3093. doi: 10.7717/peerj.3093.

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M. and Anderson, R. P. (2014) 'ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for <scp>Maxent</scp> ecological niche models', *Methods in Ecology and Evolution*, 5(11), pp. 1198–1205. doi: 10.1111/2041-210X.12261.

Phillips, S. and Dudík, M. (2008) 'Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation', *Ecography*, (31), pp. 161–175. doi: 10.1111/j.2007.0906-7590.05203.x.

Pielke, R. A., Pitman, A., Niyogi, D., Mahmood, R., McAlpine, C., Hossain, F., Goldewijk, K. K., Nair, U., Betts, R., Fall, S., Reichstein, M., Kabat, P. and de Noblet, N. (2011) 'Land use/land cover changes and climate: Modeling analysis and observational evidence', *Wiley Interdisciplinary Reviews: Climate Change*, 2(6), pp. 828–850. doi: 10.1002/wcc.144.

Rogan, J. and Chen, D. (2003) 'Remote sensing technology for mapping and monitoring land-cover and land-use change', *Progress in Planning*, 61(61), pp. 301–325. doi: 10.1016/S0305-9006(03)00066-7.

Sanchez-Hernandez, C., Boyd, D. S. and Foody, G. M. (2007) 'One-class classification for mapping a specific land-cover class: SVDD classification of fenland', *IEEE Transactions on*

510   *Geoscience and Remote Sensing*, pp. 1061–1072. doi: 10.1109/TGRS.2006.890414.

511   Shcheglovitova, M. and Anderson, R. P. (2013) 'Estimating optimal complexity for ecological

512   niche models: A jackknife approach for species with small sample sizes', *Ecological Modelling*.

513   Elsevier B.V., 269, pp. 9–17. doi: 10.1016/j.ecolmodel.2013.08.011.

514   Skowronek, S., Asner, G. P. and Feilhauer, H. (2017) 'Performance of one-class classifiers for

515   invasive species mapping using airborne imaging spectroscopy', *Ecological Informatics*. Elsevier

516   B.V., 37, pp. 66–76. doi: 10.1016/j.ecoinf.2016.11.005.

517   Skowronek, S., Van De Kerchove, R., Rombouts, B., Aerts, R., Ewald, M., Warrie, J., Schiefer, F.,

518   Garzon-Lopez, C., Hattab, T., Honnay, O., Lenoir, J., Rocchini, D., Schmidtlein, S., Somers, B. and

519   Feilhauer, H. (2018) 'Transferability of species distribution models for the detection of an

520   invasive alien bryophyte using imaging spectroscopy data', *International Journal of Applied*

521   *Earth Observation and Geoinformation*. Elsevier, 68(February), pp. 61–72. doi:

522   10.1016/j.jag.2018.02.001.

523   Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M. and Islam, T. (2012) 'Selection of

524   classification techniques for land use/land cover change investigation', *Advances in Space*

525   *Research*. COSPAR, 50(9), pp. 1250–1265. doi: 10.1016/j.asr.2012.06.032.

526   Stenzel, S., Feilhauer, H., Mack, B., Metz, A. and Schmidtlein, S. (2014) 'Remote sensing of

527   scattered natura 2000 habitats using a one-class classifier', *International Journal of Applied*

528   *Earth Observation and Geoinformation*, pp. 211–217. doi: 10.1016/j.jag.2014.05.012.

529   Strahler, A. H. (1980) 'The use of prior probabilities in maximum likelihood classification of

530   remotely sensed data', *Remote Sensing of Environment*, 10(2), pp. 135–163. doi:

531   10.1002/erv.2405.

532   Syfert, M. M., Smith, M. J. and Coomes, D. A. (2013) 'The Effects of Sampling Bias and Model

533   Complexity on the Predictive Performance of MaxEnt Species Distribution Models', *PLoS ONE*,

534   8(2). doi: 10.1371/journal.pone.0055158.

535   Warren, D. L., Glor, R. E. and Turelli, M. (2010) 'ENMTools: a toolbox for comparative studies of

536   environmental niche models', *Ecography*, 1(September 2009), pp. 607–611. doi:

537   10.1111/j.1600-0587.2009.06142.x.

538   Warren, D. L. and Seifert, S. N. (2011) 'Ecological niche modeling in Maxent : the importance of

539   model complexity and the performance of model selection criteria', *Ecological Applications*,

540   21(2), pp. 335–342. doi: 10.2307/29779663.

541   Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., Elith, J., Dudík, M.,

542   Ferrier, S., Huettmann, F., Leathwick, J. R., Lehmann, A., Lohmann, L., Loiselle, B. A., Manion, G.,

543   Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M. C., Phillips, S. J., Richardson, K. S.,

544   Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S. E. and Zimmermann, N. E. (2008)

545   'Effects of sample size on the performance of species distribution models', *Diversity and*

546   *Distributions*, 14(5), pp. 763–773. doi: 10.1111/j.1472-4642.2008.00482.x.

547

Parameters used to generate the land-cover classification results

See methodological section for references to the acronym used in FC and BT columns.

Feature Class (FC)	Regul. Multiplier (RM)	Sample Size (SS)	Test Percentage (TP)	Binary Threshold (BT)
L	5.00	100	50	FC1
H	3.00	80	30	FC5
Q	1.00	60	10	FC10
T	0.25	40		10PTP
LQ				BTOPA
PT				EETOD
QH				ETeSS
QP				ETrSS
TQ				MTeSPS
LQP				MTP
LQPT				MTrSPS
LQPTH				

Table 1. Parameters used to generate the land-cover classification results. See methodological section for references to the acronym used in FC and BT columns.

Figure 1

Violinplots of the *kappa* values obtained from the single land-cover maps generated through MaxEnt using the feature classes (FC) shown in the x-axis.

Each violinplot includes the combination of all the other tested parameters (See Table 1 for reference). Red dashed line is the median *kappa* value for all combinations.
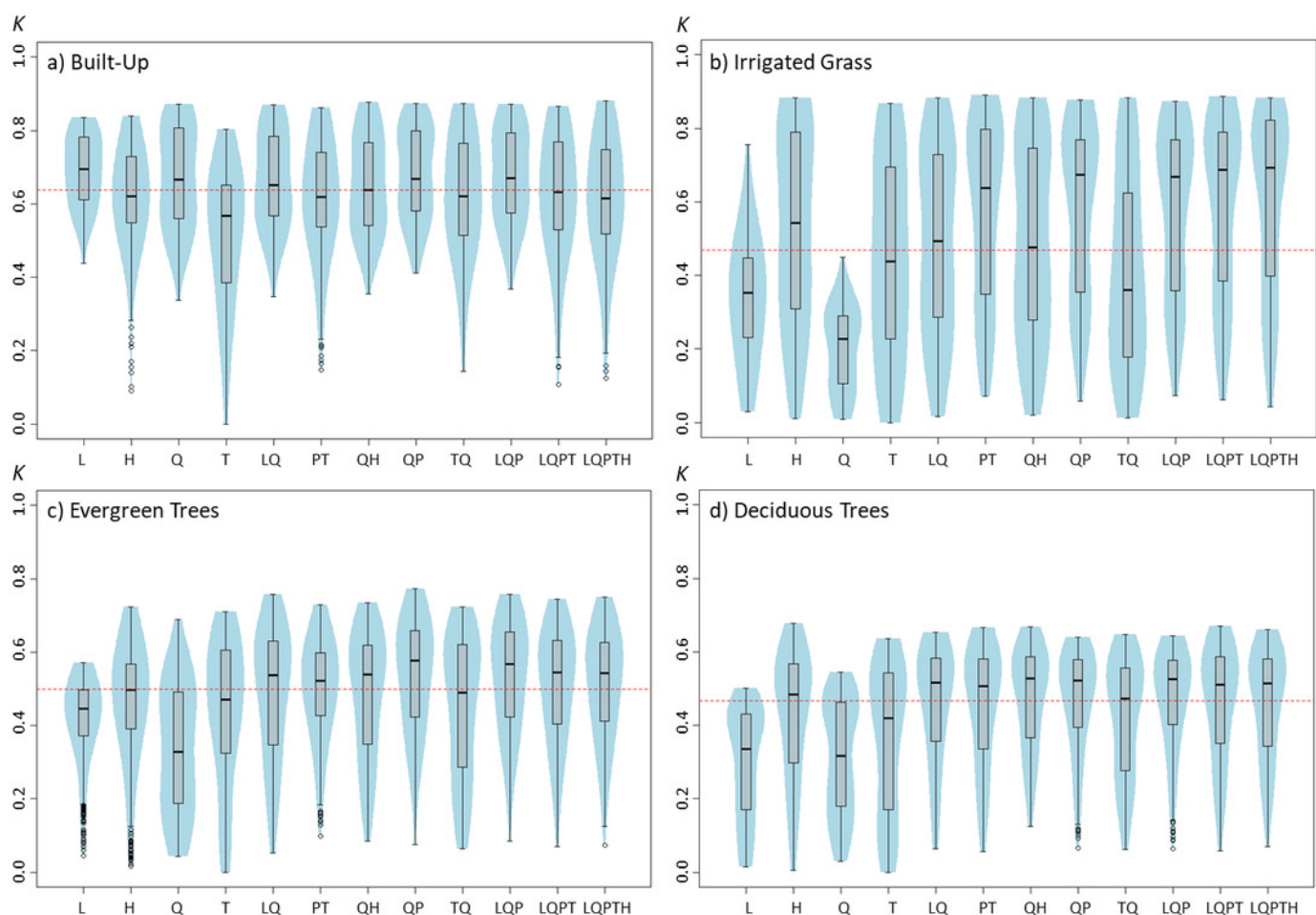
Figure 2

Violinplots of the *kappa* values obtained from the single land-cover maps generated through MaxEnt using the regularization multipliers (RM) shown in the x-axis.

Each violinplot includes the combination of all the other tested parameters (See Table 1 for reference). Red dashed line is the median *kappa* value for all combinations.
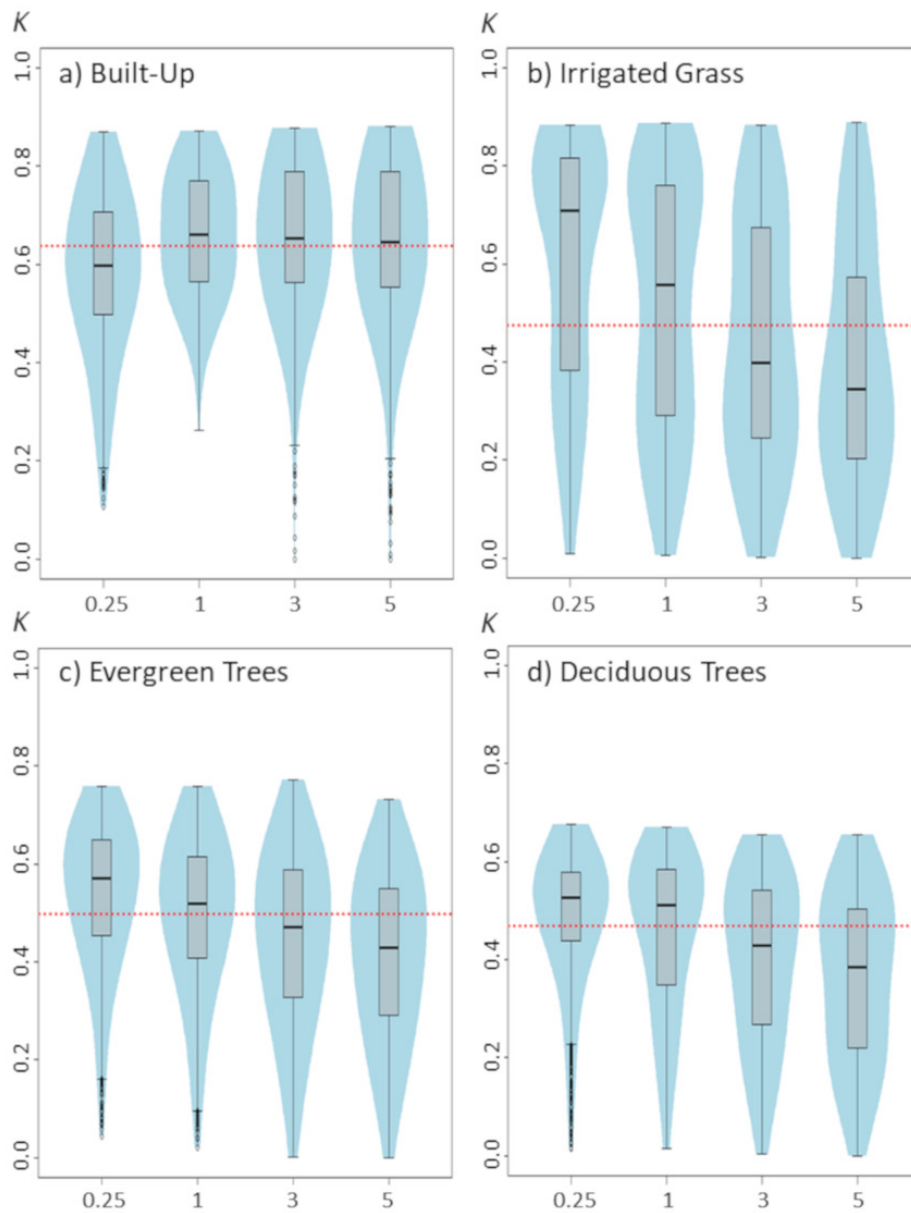
Figure 3

Violinplots of the *kappa* values obtained from the single land-cover maps generated through MaxEnt using the sample size (SS) shown in the x-axis.

Each violinplot includes the combination of all the other tested parameters (See Table 1 for reference). Red dashed line is the median *kappa* value for all combinations.
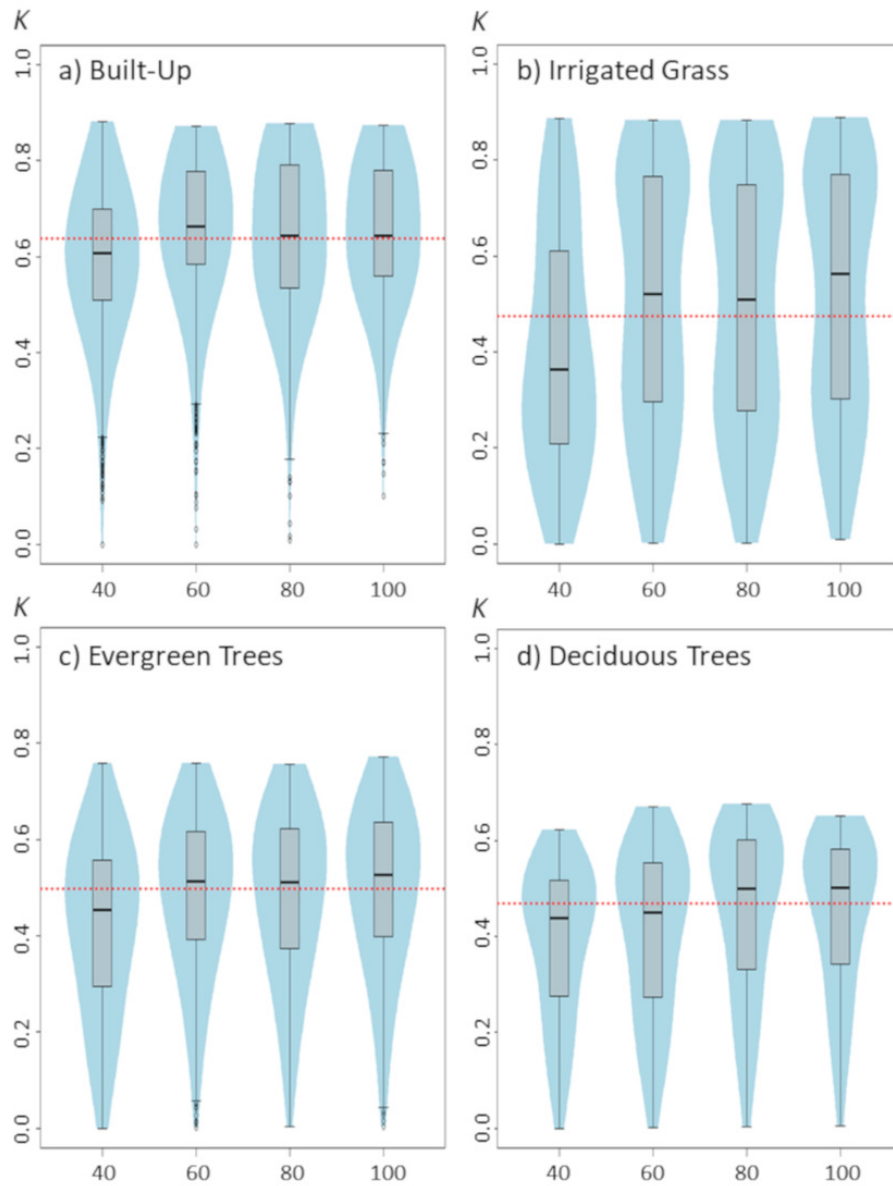
Figure 4

Violinplots of the *kappa* values obtained from the single land-cover maps generated through MaxEnt using the test percentage (TP) shown in the x-axis.

Each violinplot includes the combination of all the other tested parameters (See Table 1 for reference). Red dashed line is the median *kappa* value for all combinations.
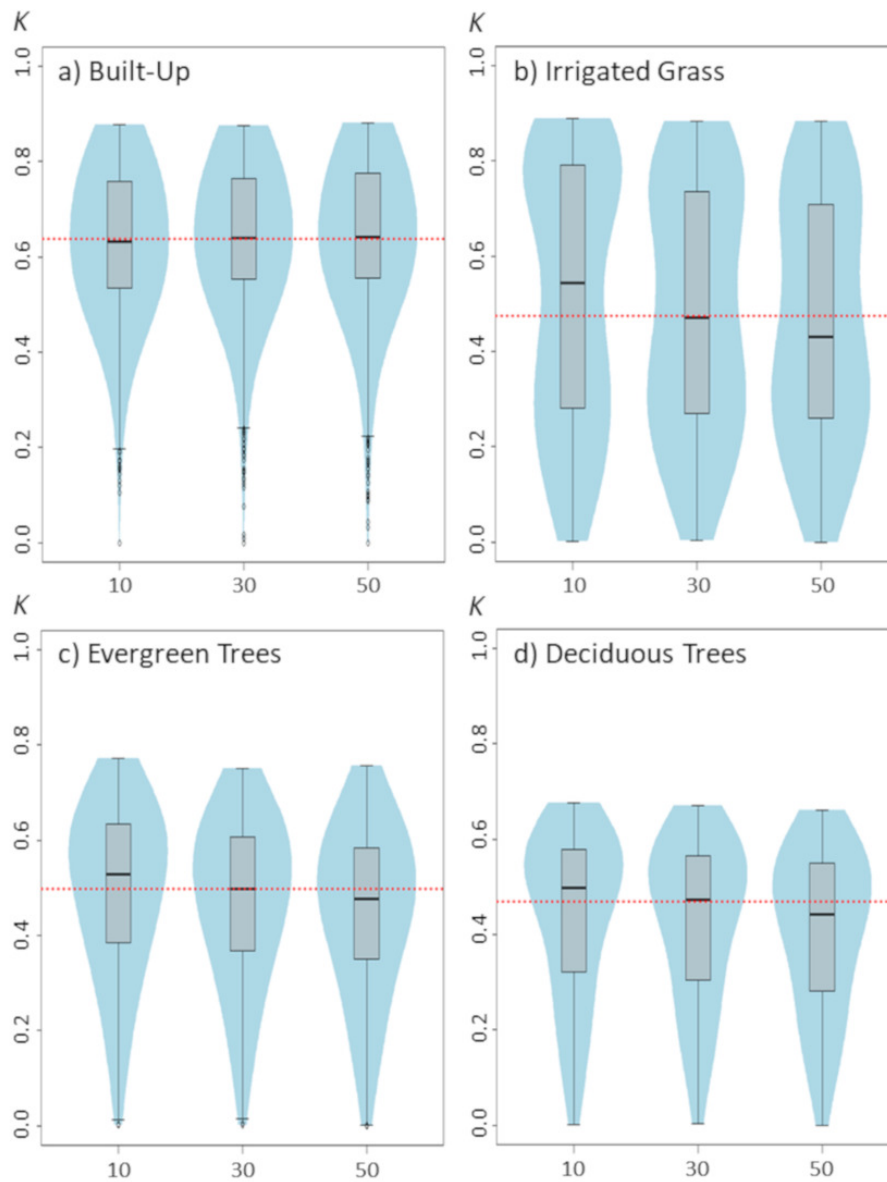
Figure 5

Violinplots of the *kappa* values obtained from the single land-cover maps generated through MaxEnt using the binary threshold (BT) shown in the x-axis.

Each violinplot includes the combination of all the other tested parameters (See Table 1 for reference). Red dashed line is the median *kappa* value for all combinations.
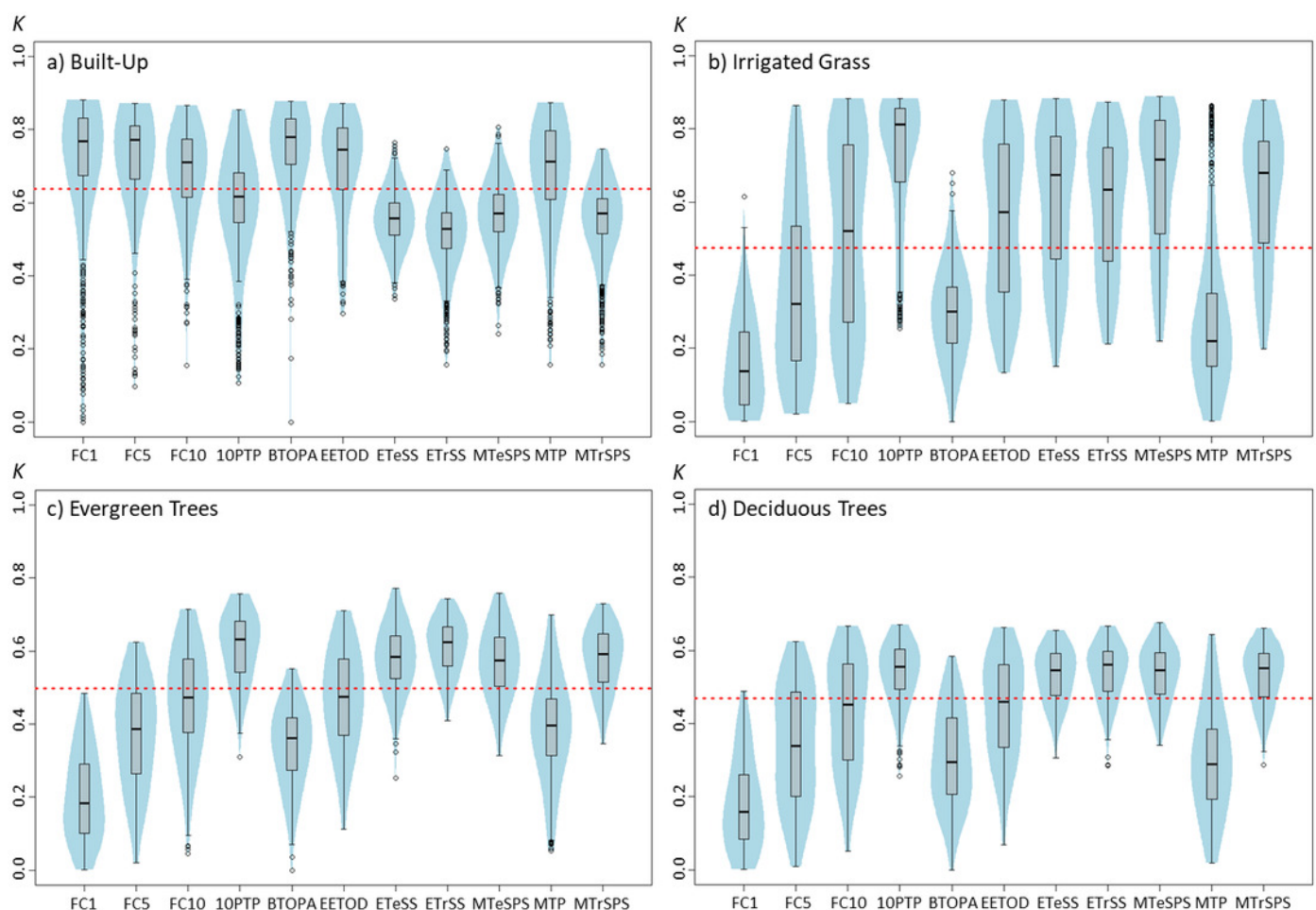
Figure 6

Barplots comparing the average *kappa* values obtained from the single land-cover maps generated through MaxEnt using the default vs best parameters settings.

Bars show the mean (±SE) *kappa* value of 30 replicates after applying the 11 BT. All models were built using a SS of 100 and a TP of 10%. The best FC and RM parameter setting used for each land-cover type is shown in the figure legend. Statistically significant values are represented by p-values of *(<0.05), **(<0.01), ***(<0.001), ****(<0.0001), after applying a two-sided t-test for each comparison (i.e. default vs best parameter setting).