University of Tehran Press

Pollution

Print ISSN: 2383-451X
Online ISSN: 2383-4501

https://jpoll.ut.ac.ir/

Thresholds Value of Soil Trace Elements for the Suitability of Eucalyptus (The Case Study of Guadiamar Green Corridor)

Blanco-Velázquez Francisco José[1] | **Anaya-Romero María**[1]✉ | **Pino-Mejías Rafael**[2]

1. Evenor-Tech "Technology-Based Company focus on Solutions for Soil Use and protection". Centro de Empresas Pabellón de Italia Business Center, 5th floor. C/ Isaac Newton, 4. Science and Technology Park Cartuja, 41092, Seville, Spain

2. University of Seville, Department of Statistics and Operational Research. Avda. Reina Mercedes s/n, 41012, Seville, Spain

Article Info

Article type:
Research Article

Article history:
Received: 11 May 2022
Revised: 15 Nov 2022
Accepted: 16 Dec 2022

Keywords:
Machine Learning
Modeling
Soil contamination
Eucalyptus
GIS

ABSTRACT

The development of suitability species models look for the availability to growth in a study area. These models can be used for different targets. In this research, a suitability model of Eucalyptus has been developed to soils contaminated by trace elements management. Guadiamar Green Corridor has been selected due to the huge data available regarding trace elements, forestry species and so on. Logistic regression (LR) and Random Forest (RF), as popular machine learning model, were applied in a geodatabase from Guadiamar Green Corridor with more of 20 years of data. This database is composed by soil physical and chemical variables, climate (temperature min and max, annual precipitation), forestry species. The results show the poor performance of LR and RF applied directly over the unbalanced training set. However, when Up-sampling or SMOTE are applied, both procedures improve its sensitivity, however, RF show more improve that LR. The methodology applied can help to determine the potential distribution of Eucalyptus in similar Mediterranean areas and extended to different areas according to Soil, Climate and Trace Elements data. Finally, the models developed under this research work can be used to reduce human and environmental health by trace elements taking into account local conditions but also climate change scenarios.

Cite this article: Blanco-Velázquez, F. J., Anaya-Romero, M., & Pino-Mejías, R. (2023). Thresholds Value of Soil Trace Elements for the Suitability of Eucalyptus (The Case Study of Guadiamar Green Corridor). *Pollution* , 9 (2), 433-444. http//doi.org/10.22059/POLL.2022.342912.1474

DOI: http//doi.org/10.22059/POLL.2022.342912.1474

INTRODUCTION

Soil degradation is a global problem in the current century decreasing soil ecosystem services by 60% between 1950 and 2010 (León et al, 2014) affecting to 33% of earth's land surface (Bini, 2009) More than 2.5 million sites have been estimated such as "potential contaminated sites", being a worldwide problem due to human activities (Panagos et al., 2013). The soil contamination has huge effects on ecosystem services as biomass production, storing and filtering, biodiversity pool and physical and cultural environment for humans (Anaya-Romero et al, 2016). According to soil contamination management, there are different strategies as soil amendments, soil remove and phytoremediation. Phytoremediation is the use of plants and associated microbes to reduce the concentrations or toxic effects of contaminants and is cost-effective, efficient, eco-friendly and has good public acceptance (Ali et al, 2013). Species as *Eucalyptus* allow stabilizing the soil and pollutants better that species as *Brassicacea* with little

*Corresponding Author Email: *fj.blanco@evenor-tech.com*

biomass produced (Landberg and Greger, 1996; Pyatt, 2001). In addition, *Eucalyptus* species have a great tolerance to trace elements because is able to develop arbuscular mycorrhizal and ectomycorrhizal symbiosis (Arriagada et al., 2004; Pereira, 1998). Species as *E. camaldulensis* has a low accumulation of trace elements in leaves in soil contaminated areas. These skills (low accumulation in leaves and elevated biomass production) are key characteristics to be a good specie in phytostabilization planning (Madejón et al, 2017). The eucalyptus, although originally from Australia, was introduced in Europe through of the Royal Botanic Gardens, Kew (Kew Gardens) in 1774 (Silva-Pando, 2016). Currently, a total of 1.3 million hectares of forested area are covered by *Eucalyptus globulus* in Europe which more than 80% in the Iberian Peninsula, followed of France and Italy (Iglesias-Trabado and Wilstermann, 2008). The trees develop well in the Iberian Peninsula due to soil and climate conditions (Cerasoli et al, 2016;). According to the national forest inventory of Spain (DGDRPF, 2012), common eucalyptus (*E. globulus*) covers the 2% of all producer-type reforestation, occupying 3% of the forest area in Spain. *E. globulus* is the most important species of forest production, with about 40%, ahead of *Pinus pinaster* with 30% (Granda-García, 2015). In many cases the needs of environmental researchers are different to the modelers due to the application of empirical models (such as black box) and mechanistic models for diverse environmental applications (Krapu and Borsuk, 2019) There are many statistical analyses and models in different areas for predicting Eucalyptus distribution (Anaya-Romero, 2004; Austin et al, 1990; De la Rosa et al, 2009). The models were developed applying different environmental variables and statistical analyses as Generalized Linear Modelling, Logistic Regression, Artificial Neuronal Network, Tree Decision and others. Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Logistical Regression (LR) are potential tools for building a classification rule. This rule could be used to assign new points to their correct class in a set of multivariate vectors of measurements (presence/absence of Eucalyptus). In addition, Classification Trees (CTs), Random Forests (RF), Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) have been successfully applied to data from different fields. RF offers many practical advantages and we have some experience with its good performance (Pino-Mejías et al., 2010), so we have considered only two models in our study: LR and RF. The previous models mentioned are freely available in the R system (R Development Core Team, 2019) and provides the user with a powerful statistical programming language. Data mining models have offered new and valid tools for classification and regression problems and for clustering in different fields (Medicine, Econometric, Image Analysis and Environmental sciences). Spatial models of vegetation dynamics could be developed through data mining techniques (Herguido-Sevillano et al, 2018, Blanco-Velázquez et al, 2020). In the present research it was considered data from Corine Land Cover in order to develop and compare data mining models for predicting the potential habitat for the Eucalyptyus (*Eucalyptus globulus* and *Eucalyptus camaldulensis*) forest type in soils contaminated. The case study selected for the present research is the Guadiamar Green Corridor, in southern Spain, with more than 20 years of monitoring data after the mine accident. The results can be extrapolated to others Mediterranean climate areas.

MATERIAL AND METHODS

The Guadiamar River Basin is located in southern Spain (**Figure 1**). In 2003, due to its environmental context, the area was declared a Protected Landscape under the name "Guadiamar Green Corridor ", which occupies some 4500 ha and runs between the Sierra Morena Mountains, in the north, to the Doñana Natural Area, in the south.

The climate in the area is Mediterranean, characterized by an accentuated summer drought, with mildly wet winters and springs with high temperatures (Zinng 2014). The average annual rainfall is 700mm, where 85% of the rain takes place between October and April. The minimum annual temperature is 17ºC, with maximum in July (35ºC) and minimum in January (5ºC).

The Study area is located on the south-eastern edge of the Iberian Pyrite Belt where transgresses Miocene sediments cover Paleozoic materials. The main alluvial deposits of the Guadiamar fluvial system are silt, sand and gravel. Due to highly variable lithology and mesoclimate, the soils are very heterogeneous and corresponding to the Mediterranean edaphic zone. In the head of basin, Cambisol soils are the main type while Fluvisols and Luvisols are in the middle section. Also, Solonchaks can be found in the south of the area (Ministry of Agriculture and Fisheries, 1999; Borja et al, 2001).

In April 1998, the Aznalcóllar mine accident occurred, caused by the rupture of the reservoir dam of the mine that caused the discharge of about 6 hm 3 to the Guadiamar river of sludge with a high concentration of trace elements. The affected surface was evaluated in 4.630 ha. After the incident, the sludge was mechanically removed and different types of amendments and tree afforestation were applied in order to immobilize as many trace elements as possible (Madejón et al. 2009). Since then, multiple research groups have been taking data from different parameters in order to assess the extent of the impact of the spill in the area, so that there is a large database available to observe how the different compartments have evolved environmental issues since the discharge.

Non-irrigated arable land, agro-forestry areas, broad-leaved forests and natural grasslands are the most representative land uses of our study area.

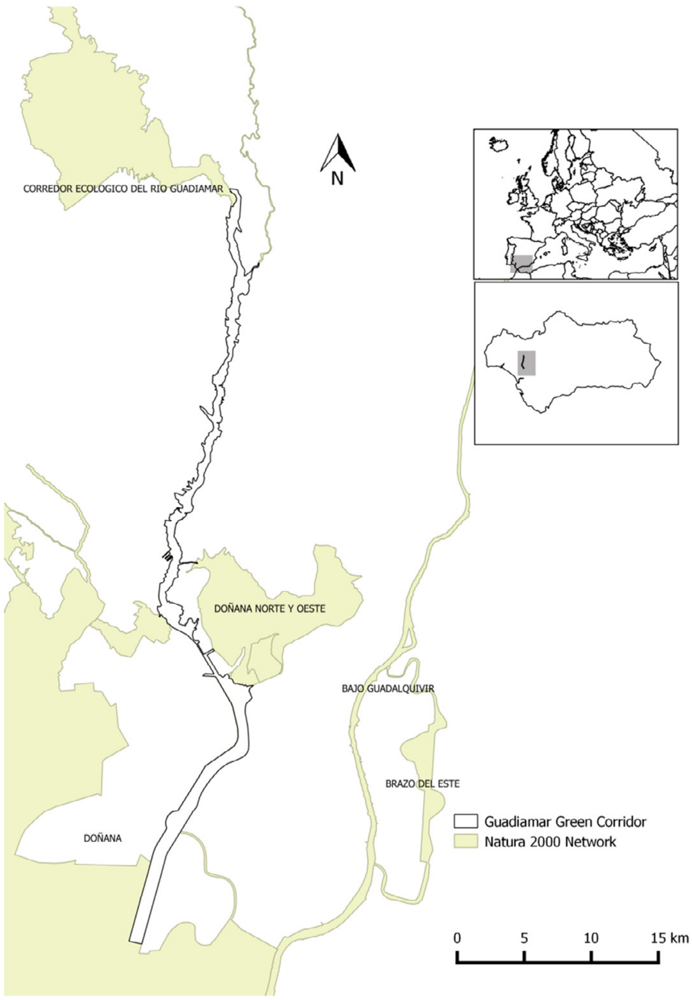

Fig. 1. Guadiamar Case Study location

A set of tree species were used in the afforestation planning (i.e. *Eucaliptus camaldulensis, Eucalyptus globulus, Fraxinus angustifolia, Olea europaea, Pinus pinea, Populus alba, Quercus ilex and Quercus suber*) distributed along of the Guadiamar Green Corridor.

Soil, site and climate data were compiled from Environmental Information Network (Environmental regional administration) between 1998 and 2016.

The map of distribution of the eucalyptus forest type at the study area was extracted from the map of land use and land cover map of Andalusia (LULCMA; Moreira. 2007) (scale 1:25.000) for the year 2003 and 2007, and Distribution cartographic of tree species from land use map of Andalusia. Currently, the Eucalyptus distribution in Guadiamar Green Corridor is 757981 m^2.

Quantum Gis Software were applied in the elaboration of Eucalyptus distribution map through the intersection tool between the case study selected and the different geodatabases from Brus et al. 2011(Figure 2). A total of 55 Gbytes of data were compilated and quality study and review were applied following the previous work (Blanco-Velázquez et al, 2019).

Harmonization and standardization processes were applied and selected the information related to Eucalyptus sites from Guadiamar Green Corridor. Predictor variables were selected from Sierra model and key variables from standardized geodatabase developed for afforestation planning sustainable in soil contaminated. The variables selected were: soil (useful depth, texture, trace element level, drainage and pH), site (latitude, altitude and physiographic position) and climate (monthly minimum temperature, monthly maximum temperature and annual precipitation).

First, a systematic sampling of the area of study was made according to the distribution of the space. Furthermore, the spatial distribution of error data and missing was verified on a DTM. This allows analysing the effect of dropping the absent data. The key variables were selected and NA was assigned to values 0 from our database. Due to the values of Cd obtained will change values of Cd to Cd log.

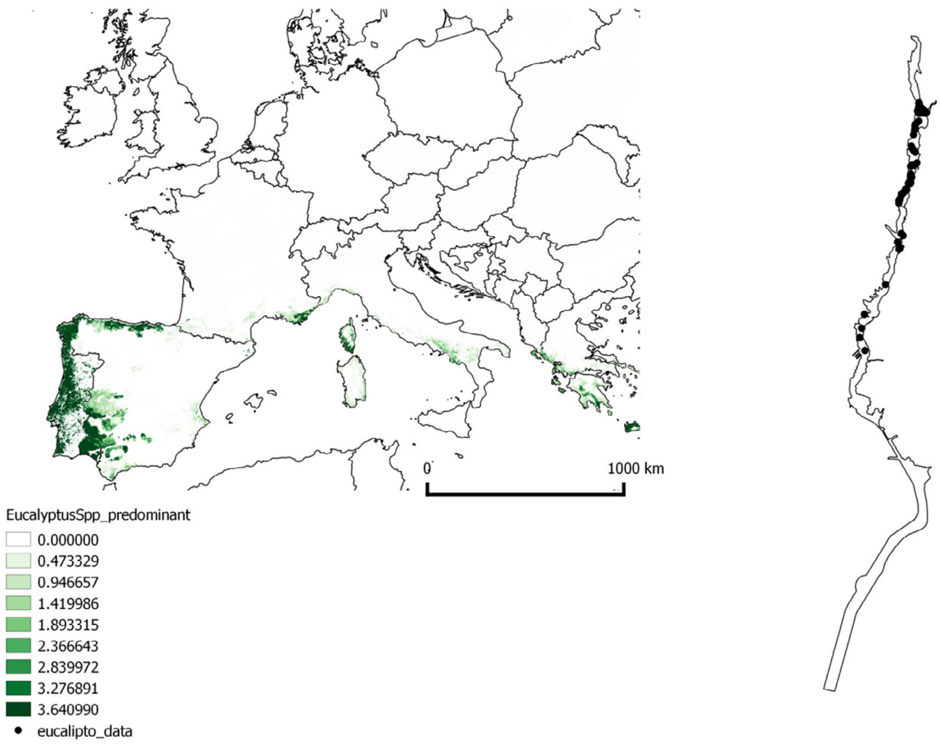

Fig. 2. Eucalyptus distribution (source Brus et al. 2011)

Once the data set is built, the construction of models was proceeding. Previously, a partition of the data has been made, thus obtaining a training set and a test set, with respective sizes 206 and 88 (70% and 30%). This partition will be used to evaluate the predictive capacity when applying strategies for the treatment of unbalanced data

The reading of the data file reflects the spatial distribution of the studied species (Figure 3)

Logistic Regression: For a binary response and p quantitative predictors $x_1,\ldots,x_p$, (some of them may be dummy variables for coding qualitative variables), the LR model assumes that the probability of the target response is

$$\pi\left(x_1,\ldots,x_p\right)=\frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}$$

The *glm* function in R (Venables and Ripley, 2002) tries to compute the maximum likelihood estimators of the p+1 parameters by an iterative weighted least squares (IWLS) algorithm. There are several inferential procedures to test the statistical significance of the whole model and the individual significance of each variable. The model may also be interpreted and a great family of diagnostics and criteria are available to identify influential and outlying observations.

Random Forests: RF was proposed by Breiman (2001) as a way to combine many different trees. A classification tree (CT) is a set of logical "if-then" conditions which drive each case to a final decision. These conditions can be easily plotted helping us to understand the model. A binary CT is grown by binary recursive partitioning using the response in the specified formula and choosing splits from the set of predictor variables. The split which maximizes the reduction in impurity (a measure of diversity for the outcome in a specific set of nodes) is chosen, the data set is split and the process is repeated. Splitting continues until the terminal nodes are too small to be split.

In RF, a number of trees are constructed. Each one is grown over a bootstrap sample of the training data set, and a random selection of variables is considered to choose splits in each

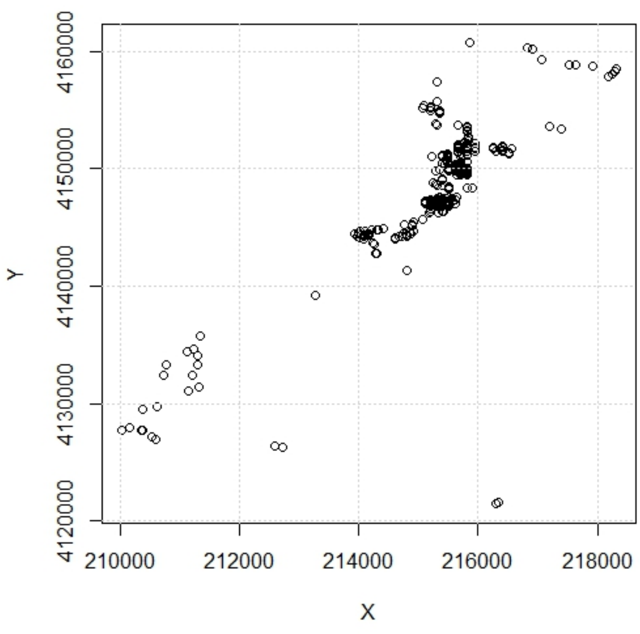

Fig. 3. Distribution of eucalyptus data

node. The predictions of the different trees are combined by majority voting. One important feature of this ensemble method is the availability of some measures to asess the importance of each variable. Breiman (2001) claims that RF does not generally overfit, and he has shown that Bayes consistency is achieved with a simple version of RF. Moreover, it runs efficiently on large data bases, being able to handle thousands of input variables. We have used the R package *randomForest* (Liaw and Wiener, 2002) where the Gini index is the default impurity measure.

The classification problem addressed in this work has a clear imbalance nature. An imbalance occurs when one or more classes have very low proportions in the training data as compared to the other class. Absence of *Eucaliptus* is observed in 13% of the locations, with a clear impact in the performance of LR and RF. There exist several techniques trying to improve the predictive capability of the trained models.

A first procedure is to modify the threshold value in a probabilistic classification rule. In our binary classification rules, the default probability value at which a prediction is either classified as presence or absence is 0.5. To improve performance, it can be helpful to optimize this probability threshold. We have used 10-fold cross-validation to select the new threshold both in LR and RF maximizing the success rate in presence class.

A second idea is to modify the training set to achieve a balanced data set. Our study has considered two of these methods: Up-Sampling and SMOTE. In Up-sampling, cases from the minority classes are sampled with replacement until each class has approximately the same number. The synthetic minority over-sampling technique (SMOTE), described by Chawla et al. (2002), is a data sampling procedure that uses both up-sampling and down-sampling (in the majority class), and has three operational parameters: the amount of up-sampling, the amount of down-sampling, and the number of neighbors that are used to impute new cases. To up-sample for the minority class, SMOTE synthesizes new cases. To do this, a data point is randomly selected from the minority class and its K-nearest neighbours (KNNs) are determined. Chawla et al. (2002) used five neighbours in their analyses, but diferent values can be used depending on the data. The new synthetic data point is a random combination of the predictors of the randomly selected data point and one randomly selected neighbour. Therefore, the SMOTE algorithm adds new samples to the minority class via up-sampling, but it can also down-sample cases from the majority class via random sampling in order to balance the training set.

In addition, the initial training set (206 cases), two balanced training sets were founded..The first one arose from Up-sampling and this new training set comprises all absence observations, and a sample with replacement extracted from locations with presence of eucalyptus. This first balanced training set has 174 absence cases and 174 presence cases. The first balanced training set were called "train_B1" and the second balanced training set was obtained with the SMOTE function in DMwR library. Default options were maintained, providing our second balanced training set, train_B2, with 303 cases: 116 absences and 187 presences. For both balanced training data sets, cross-validation process was applied to select an appropriate threshold value.

Given the high number of lost values, we have first determined which predictive variables are significantly associated with presence / absence. To do this, a t-test of means comparison was calculated and the p-value was saved, obtaining

CD	Y	CU	PB	ZN	X	sand	AS	clay	pH
0.000	0.003	0.009	0.027	0.036	0.037	0.207	0.339	0.801	0.866

Those variables with p-value <0.05 are chosen. Pearson's Chi-squared test with Yates' continuity correction was also performed to study the association between Affection and presence/absence, obtaining a p-value < 0.001. Therefore, Affection was also included in the set of predictors. This way, after selecting only complete cases for these variables, we obtained 294 cases, where 294 locations were presence, and 45 absences. The predictors were: X, Y, CD, CU, PB and ZN. Moreover, the graphical representation of CD showed the convenience of taking logarithm to obtain a more symmetric distribution, so the models were built with the variable log_CD.

For the final development of the model, QGIS graphical modeller was used (http://qgis.osgeo.org/). Using the geoprocessing tools and identifying the minimum and maximum threshold values of the selected variables (Table 1), the algorithm was developed for the resulting model.

The algorithm developed is based on a series of conditional rules with the CASE function for each of the variables to be analyzed and returns a positive response in the event that the study area is suitable for the species (Figure 4).

RESULTS AND DISCUSSION

In the present research, an extensive set of models were applied. First, base statistical techniques were applied: LR and RF. Each one was fitted to the initial training set, and its performance was computed with the default 0.5 threshold probability and with the cross-valiation selected value. These four combinations were also generated with train_B1 and train_B2 data sets, and therefore 12 classification rules were obtained, each one was applied to the test set. The performance of each rule is contained in Table 1.

Four measures are presented in Table 2.

Accuracy: percent test sets correctly classified.

Sensitivity: percent presence cases in test set that are correctly classified.

Specificity: percent absence cases in test set that are correctly classified.

AUC: area under the Receiving Operating Characteristic Curve.

Fist column identifies the model, LR or Rf. When it is accompanied by (CV_u) it is showed decision threshold computed with Cross Validation. Up denotes model built in UpSampling training set (train_B1), while SM refers to smoted training set (train_B2).

Table 2 shows the poor performance of both LR and RF when they are built directly over the unbalanced training set. Moreover, the VC selection of the thresholds is ineffective for LR, while only RF improves its sensitivity. LR classification rules have greater sensitivity values and AUC when Upsampling or SMOTE are applied. However, specificity has not a similar improvement.

Table 1. Variables included in the algorithm (Thresholds values from PHY, DEPT and DRAI belong to Sierra model (De la Rosa, 2009)

Acronym	Mean	Value
ALT	Altitude	Numeric
PHY	Physiography	1=Valley; 2= Hillside; 3= Top; 4= Terrace
DEPT	Depth	1= High; 2= Moderate
DRAI	Drainage	1= Good; 2= Moderate; 3= Excessive; 4= Poor
TMAX	Annual temperature maximum	Numeric
TMIN	Annual temperature minimum	Numeric
PREC	Annual precipitation	Numeric
CDPT	Cadmium pseudo total value	Numeric
ASPT	Arsenic pseudo total value	Numeric
CUPT	Cupper pseudo total value	Numeric
PBPT	Plumb pseudo total value	Numeric
ZNPT	Zinc pseudo total value	Numeric
PH	pH value	Numeric

Table 2. performance of classification rules

CLASSIFICATION RULE	%ACCURACY	%SENSITIVITY	%SPECIFICITY	AUC
LR	84.09	0	100.00	0.508
LR (CV_U=0.45)	84.09	0	100.00	0.508
RF	82.95	21.42	94.59	0.742
RF (CV_U =0.38)	80.68	35.71	89.19	0.742
LR UP	56.82	87.50	50.00	0.697
LR UP (CV_U =0.02)	21.59	100.00	4.17	0.697
RF UP	81.50	50.00	95.83	0.884
RF UP (CV_U =0.66)	84.09	18.75	98.61	0.884
LR SM	65.91	43.75	70.83	0.738
LR SM (CV_U =0.01)	23.86	100.00	6.94	0.738
RF SM	94.32	100.00	93.05	0.999
RF SM (CV_U =0.66)	98.86	100.00	98.61	0.999

Table 3. Results from algorithm.

Spp	**Training**			**Test**		
	Available area	**Unavailable area**	**Thresholds variables**	**Available area**	**Unavailable area**	**Thresholds variables**
Eucalyptus	55	1	pH error	15	0	
O. europaea	168	10	pH, CDPT, CUPT, ZNPT	64	6	pH, ASPT, CDPT, CUPT, PBPT
F. angustifolia	29	3	pH, CUPT	9	0	
P. alba	38	11	pH, ASPT, CDPT, CUPT, PBPT	14	3	CDPT
Q. ilex	62	4	pH, ASPT, CDPT, CUPT, PBPT	15	0	
Q. suber	8	4	pH, CDPT	3	2	CDPT
P. pinea	12	3	pH, CDPT	4	3	CDPT
U. minor	28	0		7	1	pH
	400	36		131	15	

RF trained over smoted data set is clearly the best procedure, particularly when the threshold value is selected with Cross-Validation. As it can be seen in last row in Table 1, RF offers very satisfying measures, with a percent of test correctly classified cases near to 100%, and an AUC almost equal to 1.

The model developed has been tested with the complete database compiled by Blanco et al, 2019 giving positive data in the known areas in which the *Eucalyptus* is present, as well as in other areas where it could be suitable for future necessary reforestation.

The model was subjected to a training and testing test (75-25) for the validation of the selected variables and thresholds. The results obtained are summarized in the following table (Table 3)

With a success rate of almost 100% on areas suitable for Eucalyptus, the model describes areas suitable for the development of Eucalyptus in contaminated soils. Additionally, it highlights the aptitude of the species to be present in areas with other species that are more dominant but that, in the event of their decline due to other factors (climate change, pests, etc.), the Eucalyptus could occupy those areas.

Similar methodology was developed by Hainz-Renetzeder et al, 2015, developing maps of vegetation types in GIS and ecosystem services provided. In this research work, we only focus on the improve of quality and health of the soils attending to the resilience of the Eucalyptus to trace elements levels.

Currently, Olea europaea is the species with the greatest presence in the study area, having threshold values greater than those of the Eucalyptus. However, it is also one of the species of which Lorite et al, (2018) stands out for its possible decline due to the impact of climate change mainly decrease of precipitation and increase of extreme temperatures (IPCC, 2007).

CONCLUSIONS

The methodology proposed allowing an appropriate model to predict the potential Eucalyptus distribution under soils contaminated by trace elements.

The diversity of classification models currently available enriches statistical practice in the environmental framework must be remarked, offering different alternatives for the analyst, although the number of decisions to be made is clearly increased.

The R system offer the free implementation of classical and modern classification models. The use of balanced training sets allow obtaining more accuracy results in classical and modern classification models.

The results of the proposed methodology could help the recovery of degraded areas through free statistical tools to determine the potential distribution of Eucalyptus in similar Mediterranean areas.

The useful data from the results of the model provides the potential area of Eucalyptus in the study area in case of decline of others species.

GRANT SUPPORT DETAILS

This work has been funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreements n° 603498 (RECARE), Spanish Ministry of Economy and Competitiveness under grant number CGL2014-52858-R (RESTECO) and the Spanish Ministry of Science and Innovation that granted INDUSTRIAL DOCTORATE Projects to Evenor-Tech (DI-15-07873) co-financed by the European Social Fund.

CONFLICT OF INTEREST

The authors declare that there is not any conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/ or falsification, double publication and/or submission, and redundancy has been completely observed by the authors.

LIFE SCIENCE REPORTING

No life science threat was practiced in this research.

REFERENCES

Ali, H., Khan, E. and Sajad, M.A. (2013). Phytoremediation of heavy metals – Concepts and applications. Chemosphere 91 (7). DOI: 10.1016/j.chemosphere.2013.01.075

Anaya-Romero, M. (2004). Modelo de distribución potencial de usos forestales basado en parámetros edáficos, geomorfológicos, climáticos y topográficos. PhD. University of Seville, Seville, Spain

Anaya-Romero, M., Muñoz-Rojas, M., Ibáñez, B. and Marañón, T. (2016). Evaluation of forest ecosystem services in Mediterranean areas. Ecosystem services, 20, 82-90. DOI: 10.1016/j.ecoser.2016.07.002

Arriagada, C.A., Herrera, M.A., Garcia-Romera, I. and Ocampo, J.A. (2004). Tolerance to Cd of soybean (Glycine max) and eucalyptus (Eucalyptus globulus) inoculated with arbuscular mycorrhizal and saprobe fungi. Symbiosis 36, 285–299.

Austin, M.P., Nicholls, N.O. and Margules, C.R. (1990). Measurement of the Realized Qualitative Niche: Environmental Niches of Five Eucalyptus Species. Ecological Society of America, 60(2), 161-177.

Bini C. (2009). Soil: a Precious Natural Resource. Conservation of Natural Resources. Venezia, Nova Science Publisher: 1–48.

Blanco-Velázquez, FJ., Muñoz-Vallés, S. and Anaya-Romero, M. (2019). Assessment of sugar beet lime measure efficiency for soil contamination in a Mediterranean Ecosystem. The case study of Guadiamar Green Corridor (SW Spain). Catena, 178, 163-171. https://doi.org/10.1016/j.catena.2019.03.014

Blanco-Velázquez FJ, Pino-Mejías R. and Anaya-Romero M. (2020). Evaluating the provision of ecosystem services to support phytoremediation measures for countering soil contamination. A case-study of the Guadiamar Green Corridor (SW Spain). LAND DEGRADATION & DEVELOPMENT. https://doi.org/10.1002/ldr.3608

Borja, F., López-Geta, J.A., Martín-Machuca, M., Mantecón, R., Mediavilla, C., del Olmo, P., Palancar, M. and Vives, R. (2001). Marco geográfico, geológico e hidrológico regional de la cuenca del Guadiamar. Boletín geológico y Minero. Special issue, 13-34. ISSN 0366-0176

Breiman, L. (2001). Random Forests. *Machine Learning* **45**(1): 5-32.

Brus, D.J., Hengeveld, G.M., Walvoort, D.J.J., Goedhart, P.W., Heidema, A.H., Nabuurs, G.J. and Gunia, K. (2011). Statistical mapping of tree species over Europe. European Journal of Forest Research, 131(1), 145-157.

Cerasoli, S., Caldeira, M. C., Pereira, J. S., Caudullo, G. and de Rigo, D. (2016). Eucalyptus globulus and other eucalypts in Europe:distribution, habitat, usage and threats. In: San-Miguel-Ayanz, J.,de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. (Eds.), European Atlas of Forest Tree Species. Publ. Off. EU, Luxembourg, pp. e01b5bb+

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal Of Artificial Intelligence Research, 16, 321-357.

De la Rosa, D., Anaya-Romero, M., Diaz-Pereira, E., Heredia, N. and Shahbazi, F. (2009). Soil specific agro-ecological strategies for sustainable land use. A case study by using MicroLEIS DSS in Sevilla Province (Spain). Land Use Policy 26: 4. DOI: http://dx.doi.org/10.1016/j.landusepol.2009.01.004

DGDRPF (2012). Cuarto Inventario Forestal Nacional. Inf. Tec., Ministerio de Agricultura, Alimentación y Medio Ambiente. http://www.magrama.gob.es/es/biodiversidad/temas/inventarios-nacionales/inventario-forestal-nacional/

Díaz, S. and Cabido, M. (1997) Plant Functional Types and Ecosystem Function in Relation to Global Change. Journal of Vegetation Science 8(4), 463-474. DOI: 10.2307/3237198

FAO and ITPS (2015). Status of the World's Soil Resources (SWSR) – Main Report. Chapter 4 Soils and Humans. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy. ISBN 978-92-5-109004-6

Fayad, U., Piatetsky-Shapiro, G. and Smith, P. (1996). From data mining to knowledge discovery in databases (a survey). AI Magazine 3(17), 37-54.

Feranec, J., Hazeub, G., Christensenc, S. and Jaffraind, G. (2007). Corine land cover change detection in Europe (case studies of the Netherlands and Slovakia). Land Use Policy, 24, 234–247.

Granda-García, V. (2015). Severe drought tolerance of Eucalyptus globullus (Labill): from physiology to genetics. Universidad de Oviedo. Doctoral Thesis. http://digibuo.uniovi.es/dspace/bitstream/10651/33796/1/TD_VictorGrandaGarcia.pdf

Hainz-Renetzeder, C., Scheineidergruber, A., Kuttner, M. and Wrbka, T. (2015). Assessing the potential supply of landscape services to support ecological restoration of degraded landscapes: A case study in the Austrian-Hungarian trans-boundary region of Lake Neusiedl. Ecological modelling, 295, 196-206

Herguido-Sevillano, E., Lavado-Contador, J.F., Schnabel, S., Pulido, M. and Ibáñez, J. (2018). Using spatial models of temporal tree dynamics to evaluate the implementation of EU afforestation policies in rangelands of SW Spain. Land Use Policy, 78, 166-175. https://doi.org/10.1016/j.landusepol.2018.06.054

Hosmer, D.W. and Lemeshow, S. (1989). Applied Logistic Regression. Wiley: New York.

Iglesias-Trabado, G. and Wilstermann, D. (2008). Eucalyptus universalis, Global cultivated eucalypt forests map. Version 1.0.1 (GIT Forestry Consulting's Eucalyptologics, www.git-forestry.com)

IPCC, 2007: Climate Change 2007: Synthesis Report. Geneva: IPCC. ISBN 2-9169-122-4

Jacobs, M.R. (1981). Eucalyptus for planting, vol. 11 of FAO Forestry Series (Food & Agriculture Organization of the United Nations, Rome.

Krapu, C. and Borsuk, M. (2019). Probabilistic programming: A review for environmental modellers. Environmental Modelling & Software, 40-48. http://dx.doi.org/10.1016/j.envsoft.2019.01.014

Landberg, T. and Greger, M. (1996). Differences in uptake and tolerance to heavy metals in Salix from unpolluted and polluted areas. Appl. Geochem. 11, 175–180

Liaw, A and Wiener, M. (2002). Classification and Regression by random Forest. R News 2(3), 18--22.

Lorite, I. J., Gabaldón-Leal, C., Ruiz-Ramos, M., Belaj, A., De La Rosa, R., León, L. and Santos, C. (2018). Evaluation of olive response and adaptation strategies to climate change under semi-arid conditions. Agricultural Water Management, 204, 247-261.

Madejon, P., Maranon, T., Navarro-Fernandez, C. M., Dominguez, M. T., Alegre, J. M., Robinson, B. and Murillo, J. M. (2017) Potential of *Eucalyptus camaldulensis* for phytostabilization and biomonitoring of trace-element contaminated soils. PLoS ONE 12(6): e0180240. https://doi.org/10.1371/journal.pone.0180240

Madejón, E., Madejón, P., Burgos, P., Perez de Mora, F, and Cabrera, F. (2009). Trace elements, pH and organic matter evolution in contaminated soils under assisted natural remediation: A 4-year field study. Journal of Hazardous Materials 162, 931-938

Millennium Ecosystem Assessment. (2005) Ecosystems and Human Well-Being: Synthesis. Island Press, Washington, DC

Ministry of Agriculture and Fisheries. Junta de Andalucía (1999). Report on the work carried out by the Ministry of Agriculture and Fisheries to characterize the impact of agricultural land by the discharges of the raft from the Aznalcóllar mine. Seville.

Moreira, J.M. (2007). Guía técnica del Mapa de Usos y Coberturas Vegetales del Suelo de Andalucía 1:25.000.

Panagos, P., Liedekerke, M.V., Yigini, Y. and Montanarella, L. (2013). Contaminated sites in Europe: review of the current situation based on data collected through a European network. J. Environ. Public Health Artic. ID 158764

Pereira, G.E. (1998). Efecto de las Micorrizas Vesiculo Arbusculares en Plántulas de Eucalyptus globulus (Labill.) y E. camaldulensis (Dehnh.) en Relación a la Tolerancia de Sustancias Fitotoxicas. Dissertation, University of Córdoba. Spain.

Pielke, R. A., T. Stohlgren, L. Schell, W. Parton, N. Doesken, K. Redmond, J. Moeny, McKee, T. and Kittel, T. G. F. (2002). Problems in evaluating regional and local trends in temperature: An example from eastern Colorado, USA, Int. J. Climatol. , 22, 421–434. DOI: 10.1002/joc.706

Pino-Mejías, R., Cubiles-de-la-Vega, M.D., Anaya-Romero, M., Pascual-Acosta, A., Jordan-Lopez, A. and Bellinfante-Crocci, N. (2010). Predicting the potential habitat of oaks with data mining models and the R System. *Environmental Modelling & Software*, *25*(7), 826-836.

Pyatt, F.B. (2001). Copper and lead bioacumullation by Acacia retinoides and Eucalyptus torquata in sites contaminated as a consequence of extensive ancient mining activities in Cyprus. Ecotoxicology and Environmental Safety, 50, 60-64. DOI: 10.1006/eesa.2001.2087

QGIS Development Team (2019). QGIS Geogrpahical Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Qiu, J. and Turner, M.G. (2013). Spatial interactions among ecosystem services in an urbanizing agricultural watershed. Proceedings of the National Academy of Sciences, 110 (29), 12149-12154. DOI:10.1073/pnas.1310539110

R Development Core Team (2019). R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rodríguez, J. P., Beard Jr, T. D., Bennett, E. M., Cumming, G. S., Cork, S. J., Agard, J., ... & Peterson, G. D. (2006). Trade-offs across space, time, and ecosystem services. *Ecology and society*, *11*(1). http://www.ecologyandsociety.org/vol11/iss1/art28/

Shahzad, L., Tahir, A., Sharif, F., Haq, I.U. and Mukhtar, H. (2019). Assessing the impacts of changing climate on forest ecosystem services and livelihood of Bakalot mountainous communities. Pakistan Journal of Botany. 51 (4) **DOI:**10.30848/PJB2019-4(1

Silva-Pando, F.J. and Pino-Pérez, R. (2016). Introduction of *Eucalyptus* into Europe, Australian Forestry, 79:4, 283-291, DOI: 10.1080/00049158.2016.1242369

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. New York: Springer.

Zinng, F. (2014). Evaluate Long-Term Fate of Metal Contamination after Mine Spill; Assessing Contaminant Changes in Soil. The Guadiamar Case Study; Southern Spain. MSc Thesis. Wagenigen University,57 pp.