Potential distribution model of *Leontochir ovallei* using remote sensing data

Payacán, S.1, Alfaro, F.D.2,3, Pérez-Martínez, W.1,4, Briceño-de-Urbaneja, I.1,4

1 Magíster en Teledetección, Facultad de Ciencias, Universidad Mayor, José Toribio Medina 29, Santiago 8340589, Chile.
2 GEMA Centro de Genómica, Ecología y Medio Ambiente, Universidad Mayor, Camino La Pirámide, 5750, Huechuraba, Santiago 8580745, Chile.
3 Instituto de Ecología & Biodiversidad (IEB), Casilla 653, Santiago, Chile.
4 Hémera Centro de Observación de la Tierra, Facultad de Ciencias, Universidad Mayor, José Toribio Medina 29, Santiago 8340589, Chile.

Abstract: Predicting the potential distribution of short-lived species with a narrow natural distribution range is a difficult task, especially when there is limited field data. The possible distribution of *L. ovallei* was modeled using the maximum entropy approach. This species has a very restricted distribution along the hyperarid coastal desert in northern Chile. Our results showed that local and regional environmental factors define its distribution. Changes in altitude and microhabitat related to the landforms are of critical importance at the local scale, whereas cloud cover variations associated with coastal fog was the principal factor determining the presence of *L. ovallei* at the regional level. This study verified the value of the maximum entropy in understanding the factors that influence the distribution of plant species with restricted distribution ranges.

Key words: *Leontochir ovallei*, potential distribution, machine learning techniques, maximum entropy, environmental factors.
1. Introduction

*L. ovallei* is endemic to Chile, and its geographical distribution is restricted to the hyperarid coastal zone of the Atacama Desert (Muñoz-Schick and Sierra, 2006). Its distribution is associated with arid environments with scarce and unpredictable precipitation throughout the year (Muñoz, 1973). However, years with higher rainfall during the winter promote the productivity and development of this and other plant species in an event known as “desert bloom” (Errázuriz and Hanisch, 1995; Chávez et al., 2019). This species is classified as endangered (MINSEGPRES, 2008), which is partially due to its narrow distribution (Cereceda, et al., 2000). Environmental factors determining the occurrence and abundance patterns of this species are not yet completely understood. However, preliminary evaluations in the field suggest that *L. ovallei* is located along the coastal desert where cloud cover is high, particularly between Chañaral and Quebrada de Los Choros (Kottek et al., 2006; Sarricolea et al., 2017), covering an area of approximately 200 km² (Muñoz-Schick and Sierra, 2006). Squeo et al. (2008) estimated a Priority Index based on two parameters associated with i) taxonomic uniqueness and ii) the level of endemism, concluding that *L. ovallei* is the plant species that should be given the highest priority for conservation, given its endemism and taxonomic uniqueness. However, temporal variation in abundance and phenology as a response to unexpected precipitation makes estimating this species’ distribution difficult.

Research suggests that predictions derived from habitat suitability cartography for narrow distribution species is a useful tool for understanding the contribution of different environmental factors to the spatial distribution of species (Franklin, 2010; Urbina-Cardona and Flores-Villela, 2010; Marini et al., 2010; Morales, 2012; Hawk, 2017; Tang et al., 2018; Thapa et al., 2018; Barragán-Barrera et al., 2019; Sarma et al., 2018; Giannakopoulos et al., 2019). Consequently, Wang et al. (2012) recommend using predictive models to evaluate the distribution pattern of species with scarce available data. For example, the species distribution models derived from MaxEnt (Maximum Entropy, Phillips et al., 2004) show a high performance in predicting species distribution based on a small sample or species presence data (Hernández et al., 2006; Wisz et al., 2008; Kumar and Stohlgren, 2009; Baldwin, 2009; Jafari et al., 2018; Shiv et al., 2019). MaxEnt is a machine learning algorithm that uses multivariate distributions of habitat capability deducted from species presence records to generate a species occurrence probability considering the restrictions and suitability of the environmental conditions.

Frequently, habitat suitability can be deducted from climate data such as WorldClim (Hijmans et al., 2005). However, it is recommended to use this database with caution, particularly in Chilean regions where climatic information is scarce (Pliscoff and Fuentes-Castillo, 2011). On the other hand, it has been suggested that information on soil, geology, elevation (and other topographic variables) retrieved using remote sensing improves predictions of species distribution models. For instance, several studies highlight the use of information derived from remote sensing as predictive variables of models using MaxEnt (Kumar and Stohlgren, 2009; Zimmermann et al., 2007; Kafley et al., 2009; González et al., 2013; Ćorović et al., 2018). Similarly, Austin (2007) suggested that the response of species to changes in local environmental conditions expressed as adaptations in physiology and ecology has a high predictive value for distribution models. Therefore, understanding the relative importance of local factors on species’ distribution can contribute to improving rare plant species’ recovery plans (Bartel and Sexton, 2009; Haughian et al., 2018).

In this work, we evaluated the effect of environmental variables derived from ALOS PALSAR, LANDSAT 8, and MODIS data on the potential distribution of an endemic plant with limited field data via machine learning techniques and spatially explicit information. We specifically evaluated the effect of environmental parameters such as cloud cover, topography, and earth surface reflectance to predict the potential distribution of species along the arid coastal zone of the Atacama Desert.

2. Material and methods

2.1. Study area

The study area was located in the coastal desert in the northern region of the Atacama, Chile (Figure 1). Along the hyperarid coastal desert, the...
2.2. Records on species’ presence and environmental variables

Information on the species’ existence was gathered from GBIF (Global Biodiversity Information Facility) and studies performed by the Chilean Ministry of the Environment. The environmental variables or environmental layers corresponded to single-temporal (DEM image and Landsat image) and multi-temporal (MODIS) satellite sensor products. Specifically, ecological variables used as predictor variables were obtained from the digital elevation model (ALOS PALSAR DEM image), bands, and spectral indices of the OLI/TIRS sensor of Landsat 8. Cloud cover multi-temporal data were derived from the Monthly Cloud Fraction product of MODIS (Moderate Resolution Imaging Spectroradiometer). DEM spatial resolution was resampled from 12.5 m to 30 m, using the bilinear interpolation algorithm.

The Landsat image (Path 001/Row 079), with an acquisition date of November 19th, 2015, was retrieved from the U.S. Geological Survey (http://earthexplorer.usgs.gov/) as a GeoTIFF file Level L1TP and was processed using the QGIS software (version 2.14.18) and the Semi-Automatic Classification Plugin (version 5.3.8). The original values of the images, digital number (DN) were calibrated to values of radiance using gain (Gb) and bias values (Bb) for each band obtained from the files of the header of the images by Equation 1.

\[ L_e = \text{DN}_i \times G_b + B_b \]  

Where \( L_e \) is the measured spectral radiance value in values from Wm.²sr⁻¹μm⁻¹; \( \text{DN}_i \) is the digital number of each pixel per band; \( G_b \) is the gain for each particular band (gain), and \( B_b \) is the offset for each particular band (bias). Images in radiance (L_e) can be converted to Top Of Atmosphere (TOA) Reflectance (Chavez, 1996). The Normalized Difference Vegetation Index (NDVI) was calculated, as well as the ratio between SWIR1 and SWIR2 (band 6/ band 7) and SWIR2. The model used eight spectral bands and the thermal infrared (band 11).
MODIS data covered 15 years of cloud cover observations from the MOD09, available from the global database at 1 km spatial resolution (Wilson and Jetz, 2016). Monthly values from February 2000 to March 2014 were adjusted to monthly averages, creating 12 raster files resampled to 30 m spatial resolution using the bilinear interpolation algorithm. Each variable was delimited to a 2675 km² polygon (Figure 1).

The spatial relationship between the 28 environmental predictor variables in the study area was analyzed (collinearity diagnostic). After calculating Spearman’s correlation between each pair of variables, those with correlation values greater than 0.7 were removed from further analysis (Dormann et al., 2013). The total number of selected variables was 10 (see Table 1 for further detail).

### 2.3. Statistical model, validation, and software

The MaxEnt algorithm was applied to a target probability distribution by finding the probability distribution of maximum entropy (Phillips et al., 2006). The model performance was evaluated by computing the area under the curve (AUC), and the Jackknife test verified the contribution of the environmental variables. The record on species’ presence, raster data, and final cartography were processed in QGIS (Quantum Source Geographic Information System) Desktop 2.14.3 (QGIS Development Team, 2017) and R 3.4.1 (R Core Team, 2017), using the UTM coordinate system and datum WGS-84 19S. Each step of this analysis is described in Figure 2.

### 3. Results

The estimated potential distribution of *L. ovallei* is limited to the coastal desert between the coastal area of Vallenar (328830 E and 6838772 N) and Copiapó (369806 E and 6970418 N). The probability of occurrence decreased when elevation increased in inland valleys (Figure 3).

The model calibration test for *L. ovallei* yielded satisfactory results according to the area under the receiver operating characteristic curve or ROC curve (Figure 4).

A summary of the contribution of each variable in the distribution model is shown in Table 2. The environmental variables with higher relative contribution were the altitude, cloud cover in November, topographic index, the normalized burn index, and the band 7 (SWIR 2) of Landsat 8. Likewise, omitting the DEM within the model had the highest impact on the model gain. The response curves generated by the model (Figure 5) suggest that the relative importance of elevation decreases with the increase of this variable. In contrast, the relative importance of cloud cover increases with higher percentages of cloud cover. The relative importance of the Index of

<table>
<thead>
<tr>
<th>Code of variable</th>
<th>Description</th>
<th>Units</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM</td>
<td>Digital Elevation Model</td>
<td>m</td>
<td>ALOS PALSAR</td>
</tr>
<tr>
<td>SLOPE_EW</td>
<td>First order partial derivative (of EW slope)</td>
<td>-1 to 1 (adimensional)</td>
<td></td>
</tr>
<tr>
<td>SLOPE_NS</td>
<td>First order partial derivative (of NS slope)</td>
<td>-1 to 1 (adimensional)</td>
<td></td>
</tr>
<tr>
<td>TPI</td>
<td>Topographic Position Index</td>
<td>adimensional</td>
<td></td>
</tr>
<tr>
<td>VRM</td>
<td>Vector Ruggedness Measure</td>
<td>0 to 1</td>
<td></td>
</tr>
<tr>
<td>SWIR2</td>
<td>Short-Wave Infrared</td>
<td>0 to 10000</td>
<td>Landsat 8</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
<td>-1 to 1 (adimensional)</td>
<td></td>
</tr>
<tr>
<td>SWIR1divSWIR2</td>
<td>Band ratio SWIR1/SWIR2</td>
<td>adimensional</td>
<td></td>
</tr>
<tr>
<td>NBR</td>
<td>Normalized Burn Ratio</td>
<td>-1 to 1 (adimensional)</td>
<td></td>
</tr>
<tr>
<td>CF11</td>
<td>Average cloud cover in November</td>
<td>0 to 1</td>
<td>MODIS</td>
</tr>
</tbody>
</table>
Topographic Position (TPI) is high with negative values of this variable (this suggests that the cell is at or near the bottom of a valley).

The Jackknife test indicated that the environmental variable with the highest gain, when used in isolation, is elevation (DEM), followed by cloud cover in November (CF11), which therefore appears to have the most useful information by itself. In turn,

---

**Figure 2.** Workflow showing the used methodology.

**Figure 3.** Potential spatial distribution of *L. ovallei*.

**Figure 4.** AUC (area under the receiver operating characteristic curve) of the model. Y-axis shows the sensitivity of prediction to data and the x-axis is predicted area to data. AUC values are more significant than the random forecast of 0.5 (training AUC= 0.966). The red line (training) shows how well the model fits the training data.
The environmental variable that decreases the gain the most when it is omitted is elevation, followed by cloud cover in November, which indicates that they have the most information that is not present in the other variables (Figure 5).

The response curves of environmental variables (Figure 6) show how the probability of predicted presence changes as each environmental variable varies, keeping all other ecological variables at their average sampled value. The meaning of the y-axis corresponds to an estimate between 0 and 1 of the presence probability. Based on the response curves, the suitable elevation (A) ranged from 0 m to 400 m, which demonstrated that the optimal altitude for the growth of *L. ovallei*’s was low. The response curves of cloud cover in November (B) showed that the probability of presence tends to increase as cloud cover approaches values of 0.7, values greater than 0.7 are infrequent in the area, what explains why the possibility of occurrence is not increased.

**Table 2.** The relative contribution of the environmental variables to the model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Contribution (%)</th>
<th>Permutation importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital Elevation Model (DEM)</td>
<td>66.1</td>
<td>84.2</td>
</tr>
<tr>
<td>Average cloud cover in November (CF11)</td>
<td>16</td>
<td>2.4</td>
</tr>
<tr>
<td>Topographic Position Index (TPI)</td>
<td>6.9</td>
<td>1.2</td>
</tr>
<tr>
<td>Normalized Burn Ratio (NBR)</td>
<td>5.9</td>
<td>9.9</td>
</tr>
<tr>
<td>Short-Wave Infrared (SWIR2)</td>
<td>2.6</td>
<td>1.3</td>
</tr>
<tr>
<td>Vector Ruggedness Measure (VRM)</td>
<td>1.6</td>
<td>0</td>
</tr>
<tr>
<td>Band ratio SWIR1/SWIR2 (SWIR1divWIR2)</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>First order partial derivative (SLOPE_EW)</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>Normalized Difference Vegetation Index (NDVI)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>First order partial derivative (SLOPE_NS)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure 5.** Jackknife test with the contribution of environmental variables.
Figure 6. Response curves of environmental variables. A: Response to DEM (Digital Elevation Model in meters); B: Response to CF11 (Average cloud cover in November); C: Response to TPI (Topographic Position Index); D: NBR (Normalized Burn Ratio); E: SWIR2 (Short-Wave Infrared); F: VRM (Vector Ruggedness Measure); G: SWIR1 div SWIR2 (Band ratio SWIR1/SWIR2); H: SLOPE_EW (First-order partial derivative of EW slope).
4. Discussion

The results suggest that the use of local environmental variables to develop a model of the spatial distribution of *L. ovallei* species is useful, particularly in the case of species with limited data and those that grow in desert areas, where environmental factors influencing distribution are relatively accessible for evaluation. The greater importance of cloud cover may be related to the presence of coastal fog along the Chilean coastal desert, which affects the distribution and abundance of different plant communities (Rundel *et al.*, 1991; Garreaud *et al.*, 2002; Thompson *et al.*, 2003).

On the Chilean coast and over large spatial scales, the intensity and frequency of fog cover show a clear seasonal pattern (Garreaud *et al.*, 2008), with fog cover highly correlated with cloud cover in the study area. Therefore, its relative impact on plant distribution could be different among diverse spatial scales (Larrain *et al.*, 2002; Latorre *et al.*, 2011). The cloud cover along the coastal desert promotes the increase of relative humidity to nearly 66% during summers and around 78% during winters (Juliá *et al.*, 2008), thus reducing the evapotranspiration and hydric stress of plants (Carvajal *et al.*, 2014). The higher probability of the presence of *L. ovallei* in valleys may be related to the fog coming from the coastal desert to these coastal areas, promoting growth and development of individuals, which would be similar to patterns reported for other plant species (Garreaud *et al.*, 2008).

Our model predicts that the discrete and sectorial distribution of *L. ovallei* is influenced by environmental factors that operate over different spatial scales. Cloud cover and fog affect the spatial distribution at the regional scale, while local variations in elevation are the principal factor determining the presence and abundance of this endemic species.

5. Conclusions

This study demonstrates the importance of local variation in elevation, microhabitat, and regional factors such as cloud cover on the distribution of *L. ovallei*. From a methodological perspective, our results suggest that analyses based on machine learning can predict species’ distribution, particularly those of a narrow distribution and with limited associated data. From a conservation perspective, our model facilitates the prediction of potential variations of *L. ovallei* under global change scenarios, with foreseen changes in cloud cover as a reaction to the rise of upwelling throughout the Chilean coastal desert. Given the scarcity of records reporting the presence of the species *L. ovallei*, it would be necessary to discover new sites outside the currently known range, especially in areas where the model shows highly suitable environmental conditions for development for this species without current known records.

Acknowledgments

The Master supported this work in Remote Sensing program, Earth Observation Center Hémera and Centre for Genomics, Ecology and Environment (GEMA), Faculty of Sciences, Universidad Mayor.

References


