ECOLOGÍA Y GESTIÓN FORESTAL SOSTENIBLE
Grupo de Investigación - UNIVERSIDAD POLITÉCNICA DE MADRID - gi.egforestal@upm.es ++34 91 336 76 60

Predictive modelling of climate suitability for *Pinus halepensis* in Spain

Aitor Gastón and Juan I. García-Viñas

ECOGESFOR-UPM
Research Group of Ecology and Sustainable Forest Management
Universidad Politécnica de Madrid
aitor.gaston@upm.es

IUFRO Joint International Meeting: Global Change and Mediterranean Pines: Alternatives for Management
10-12 February 2010, University of Valladolid at Palencia (Spain)

Predictive species distribution models and forest management

http://openmodeller.sourceforge.net/

Identify most sensitive sites to climate change

Support species selection for reforestation

Support conservation planning and reserve selection

Pinus halepensis

→ An important tree species in Mediterranean forest management

→ Previous studies on climatic suitability in Spain available:

- ⇨ Parametric autoecology of pines (Gandullo & Sánchez-Palomares)
- ⇨ Climatforest software (García-López & Allué-Camacho)

Statistical methods for species distribution models

Recent studies have shown that **regression models** generally outperform **climatic envelopes** like those used in previous studies on *Pinus halepensis* in Spain (Gandullo & Sánchez-Palomares, García-López & Allué-Camacho).

Can a logistic regression model more accurately predict the distribution of *Pinus halepensis* in Spain?

Logistic regression

- A generalization of the linear model to enable binary response variables (presence/absence)
- Logistic regression involves an equation that predict the probability of a binary outcome (e.g. species presence/absence) as a function of independent predictors (e.g. climatic variables)

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i})}}.$$

Species probability of occurrence

Estimated β_k regression coefficients

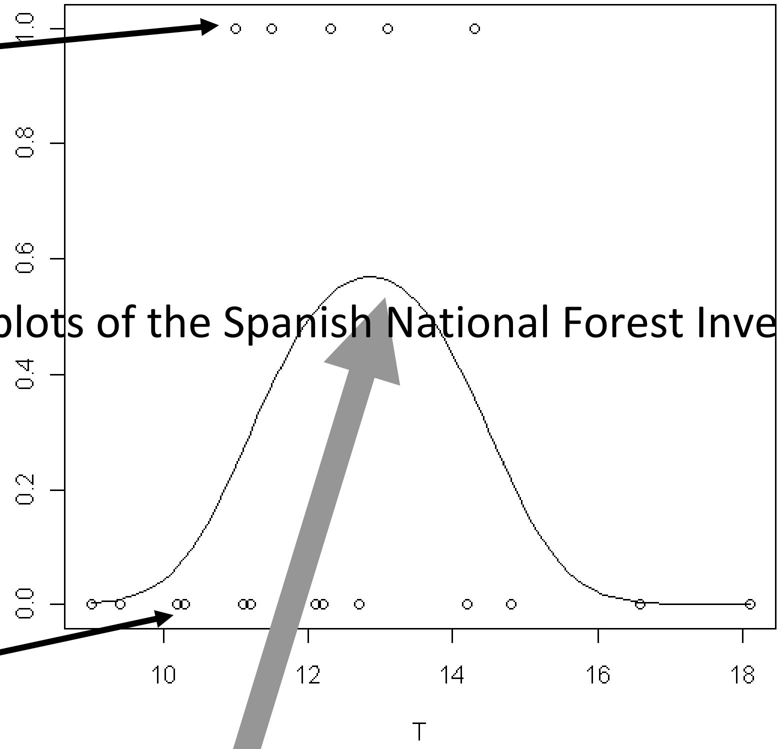
x_k independent predictors

A simple univariate example with 20 plots of the Spanish National Forest Inventory (NFI)

Presence /absence *P. halepensis*	Mean Annual Temp.
1	11.5
1	12.3
1	13.1
1	14.3
1	11
1	12.3
0	14.2
0	10.3
0	9.4
0	11.1
0	14.8
0	12.2
0	12.7
0	11.8
0	10.2
0	12.1
0	11.2
0	9
0	16.6
0	18.1

Presences

Absences



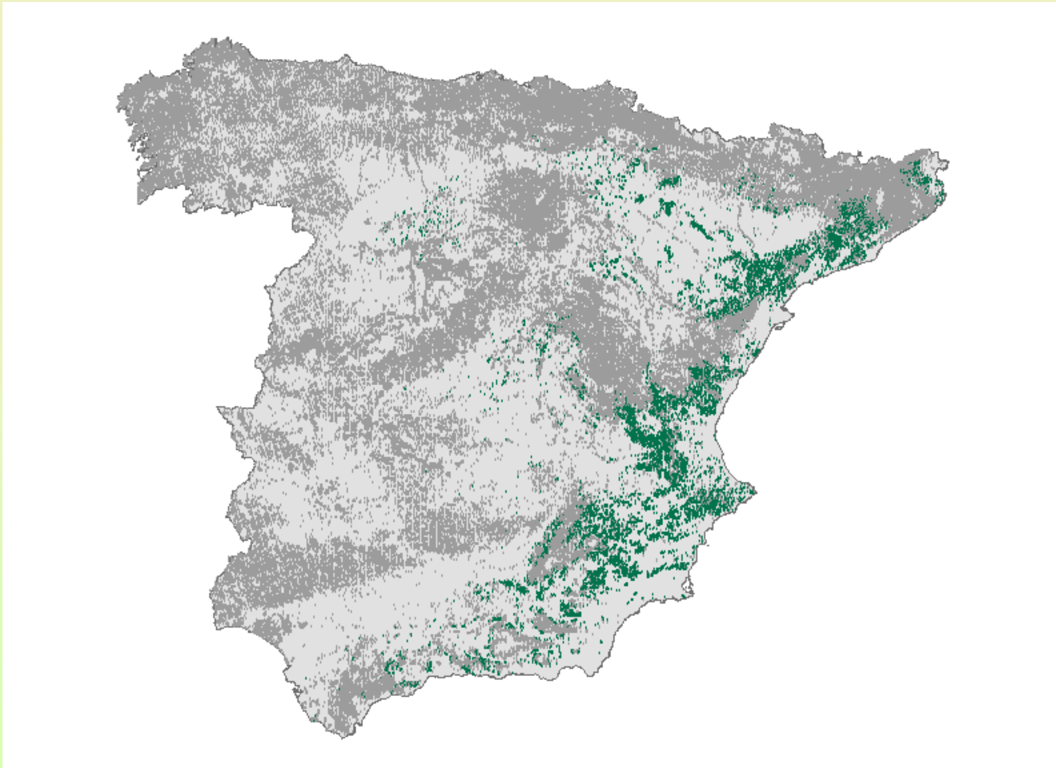
$$p = 1/(1+e^{-[-67.9+10.6 \cdot T-0.41 \cdot T^2]})$$

Probability of occurrence for *P.halepensis* for a new site with T=15:

$$p = 1/(1+e^{-[-67.9+10.6 \cdot 15-0.41 \cdot 15^2]}) = 0.16$$

Response variable

Presence or absence of *Pinus halepensis* dominated stands in the Third National Forest Inventory (91,939 plots in continental Spain)



Independent variables (predictors)

ESTCLIMA
climatic modelling software
(Sánchez-Palomares et al. 1999)

17 climatic variables

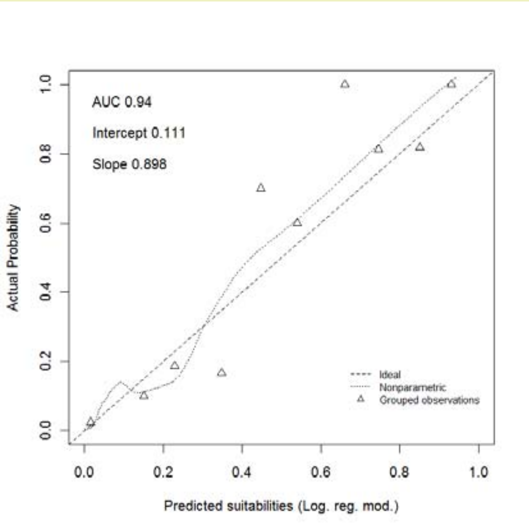
- WiR, mean winter rainfall
- SpR, mean spring rainfall
- SuR, mean summer rainfall
- AuR, mean autumn rainfall
- R, mean annual rainfall
- WiT, mean winter temperature
- SpT, mean spring temperature
- SuT, mean summer temperature
- AuT, mean autumn temperature
- T, mean annual temperature
- Tw, mean of maximum temperatures of the warmest month
- Tc, mean of minimum temperatures of the coldest month
- DSL, dry season length
- DSI, dry season intensity
- PET, mean annual potential evapotranspiration
- WS, mean annual water surplus
- WD, mean annual water deficit

Model fitting

• Predictors were transformed using Restricted Cubic Splines (4 knots) to allow modelling of asymmetric unimodal responses

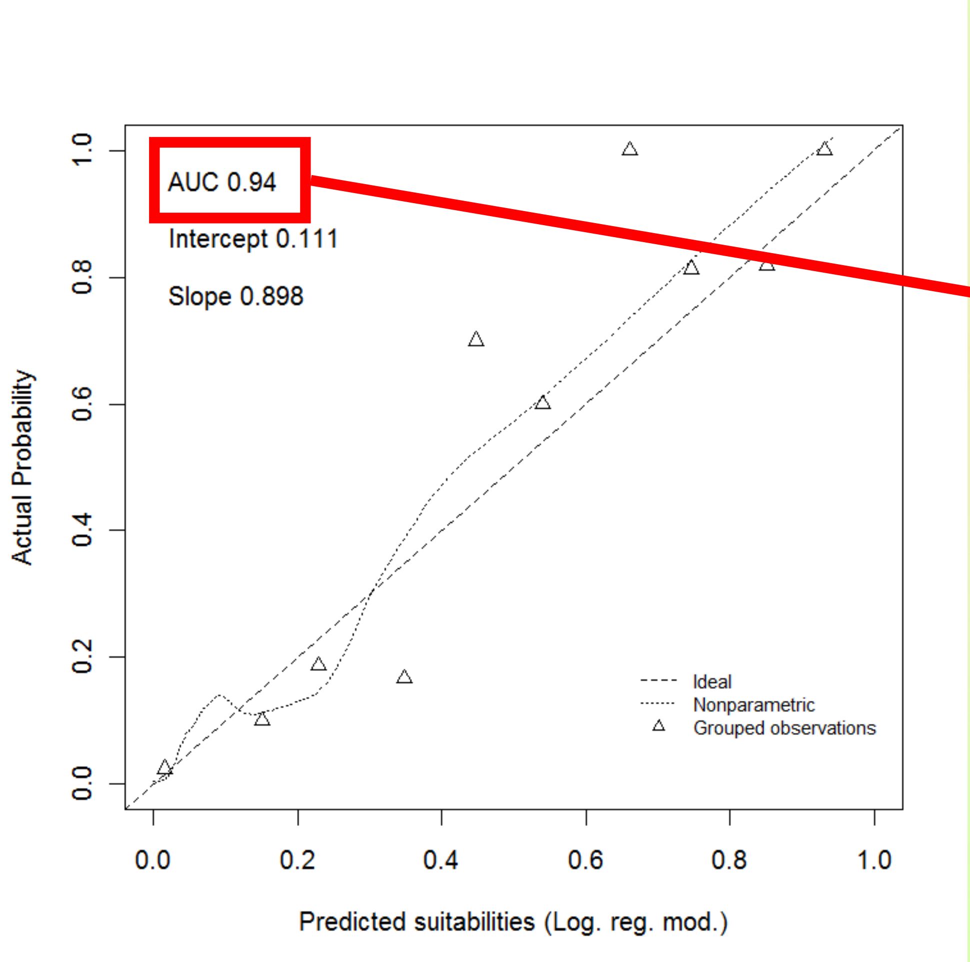
• Model parameters were estimated using standard maximum likelihood

Model validation



1. Generate suitability predictions for *Pinus halepensis* in an independent sample (444 ICP-Forests Level I plots).
2. Evaluate the level of agreement between suitability predictions and observed presence/absence.
   - ✓ **Discrimination**: the ability of the model to discriminate sites by suitability, i.e., does the model predicts higher suitability for an occupied site than for an unoccupied one?
   - ✓ **Calibration**: the resemblance of the suitability predictions to observed frequencies of occurrence.

Model validation



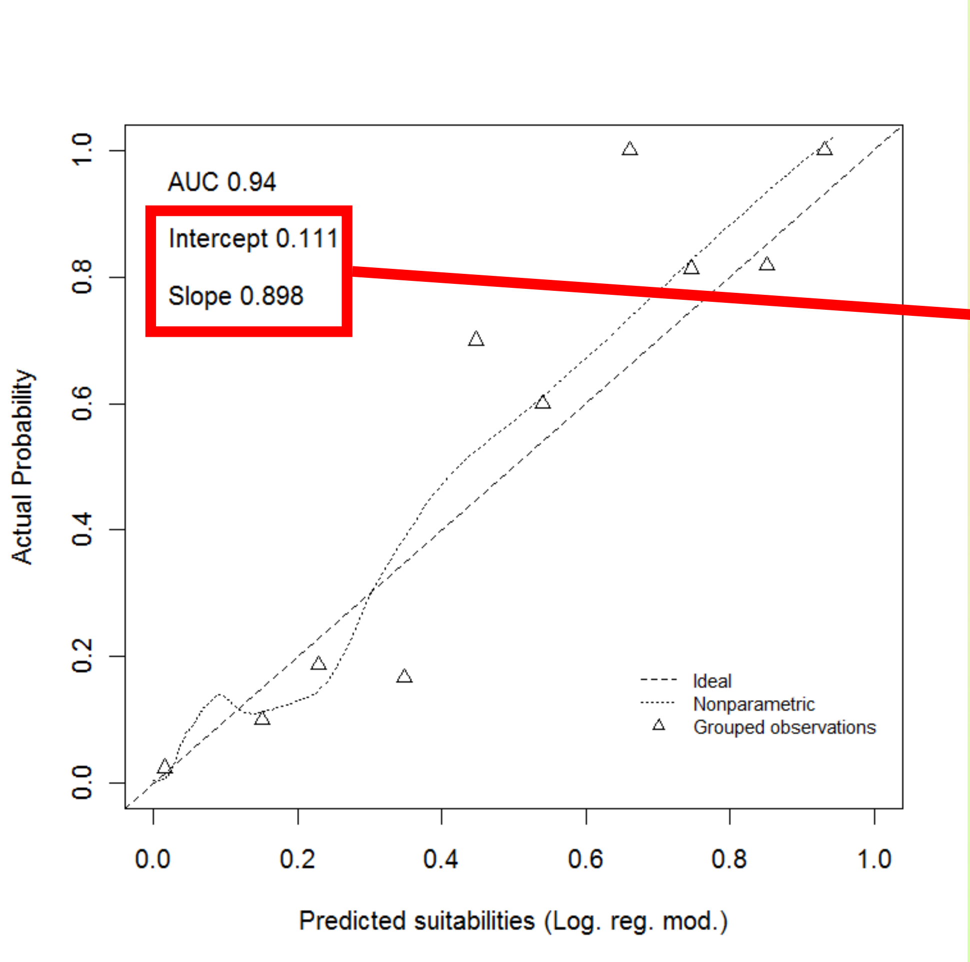
Discrimination

AUC is equal to the probability that a model will rank a randomly chosen occupied site higher than a randomly chosen unoccupied one

0.5 Models not better than random
>0.7 Acceptable discrimination
1.0 Perfect discrimination

Important issue if sites will be ranked according to species suitability (e.g., for reserve selection or mapping suitable sites for species reintroduction).

Model validation



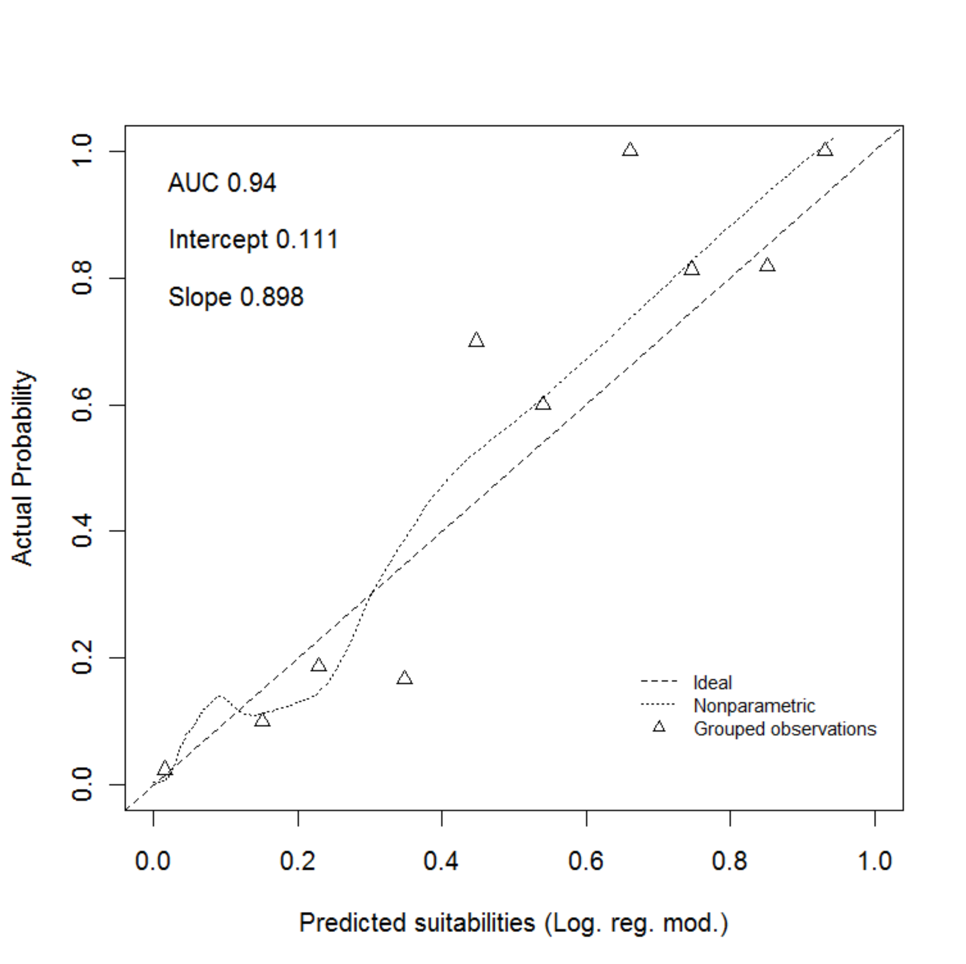
Calibration

The calibration intercept measures the difference between average predicted suitability and observed frequency **(should be near 0)**.

The calibration slope measures the concordance of predicted probabilities and observed frequencies **(should be near 1)**.

Important issue if reliable estimates of occurrence likelihood are required (e.g., for species selection in ecological restoration).

Model validation of logistic regression model



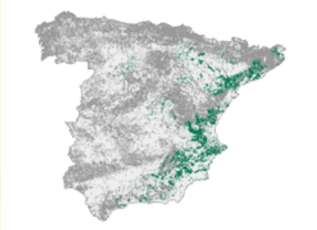
Excellent discrimination

Acceptable calibration

Does logistic regression outperform previously available models?

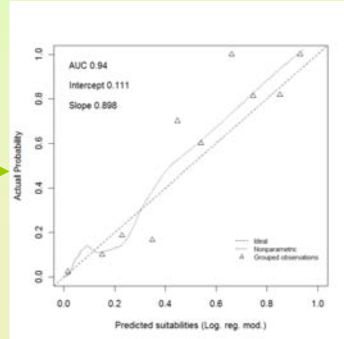
Comparison with previous models

Training sample NFI

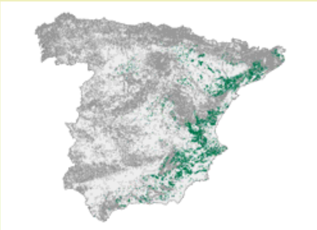


Logistic regression model

Validation sample ICP-F

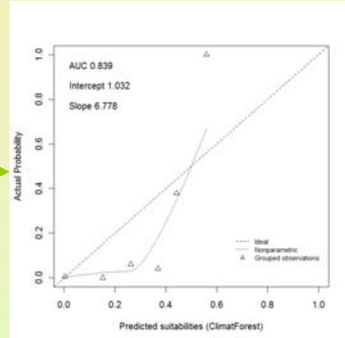


Training sample NFI

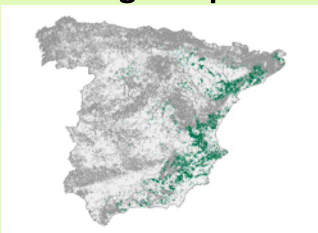


Parametric autoecology (Gandullo & Sánchez-Palomares)

Validation sample ICP-F

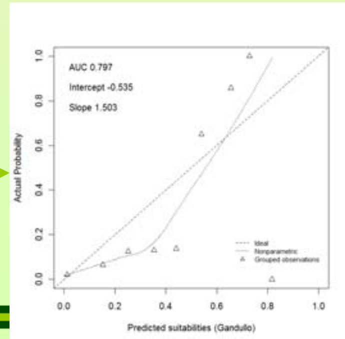


Training sample NFI

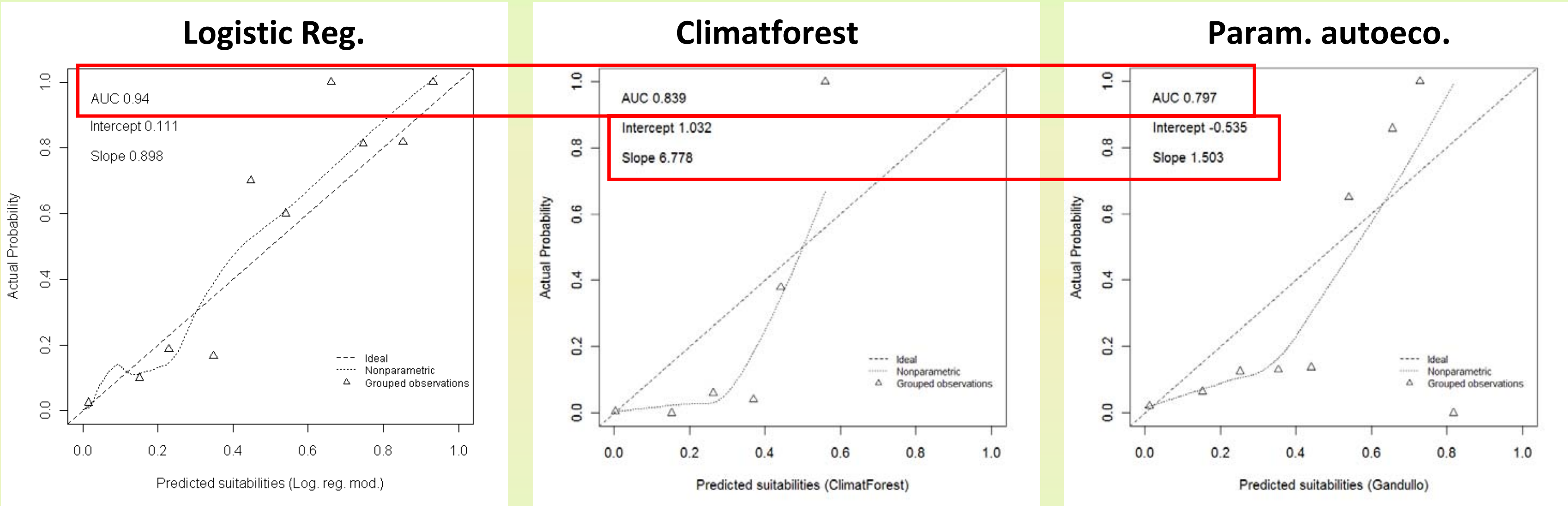


ClimatForest software (García-López & Allué-Camacho)

Validation sample ICP-F



Comparison with previous models



Previous models are not probabilistic, therefore good calibration can not be expected.

Discrimination ability of previous models is acceptable but lower than that of the logistic regression model.

Conclusions

Results show good predictive performance for the logistic regression model, considering both discrimination and calibration.

The logistic regression model outperformed other models previously available for *Pinus halepensis* in Spain.

Thanks for your attention