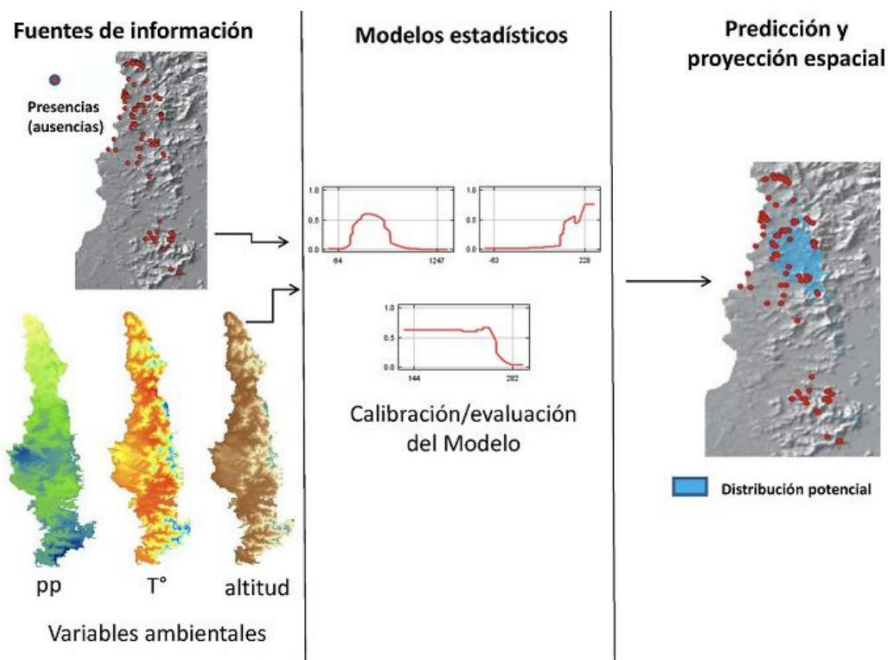




Fuente de datos

- Los **modelos de distribución** se generan usando dos tipos de fuentes de información:
 - Los datos de **presencia o ausencia** de las especies/ecosistemas que se quiere modelizar.
 - Las **variables descriptoras** que definirán el **espacio ambiental**, donde se distribuirá en el espacio geográfico.
- La **técnica de modelación** seleccionada establecerá una **relación entre la posición geográfica de la información de presencia o ausencia y el rango de valores del conjunto de variables donde se ubican estos puntos**.

Pasos metodológicos para la generación de modelos de distribución de especies/ecosistemas



Fuente: Modificado de Guisan *et al.*, 2002.



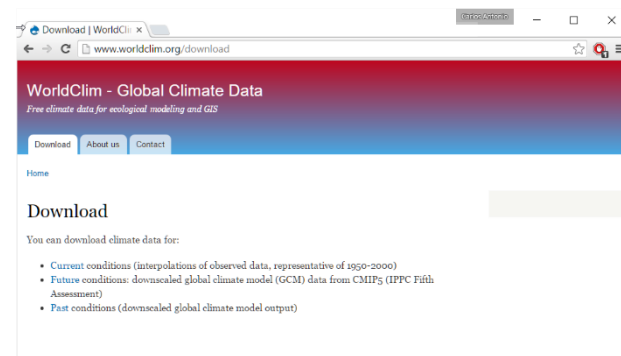
Datos de especies/ecosistemas

- La forma más básica de obtener los datos de presencia/ausencia, es **generando la información en terreno mediante GPS** u otros tipos de herramienta de adquisición de datos.
- También se puede recopilar información de forma indirecta mediante el uso de **bases de datos de libre acceso** ([Inventario Forestal Nacional](#)), que cada vez adquieren mayor importancia.
- En general, independiente de la escala de trabajo definida se deben seguir **criterios de heterogeneidad en la selección de sitios de muestreos** (privilegiar el muestreo en distintos ambientes) y de **homogeneidad en la recolección** (definir distancias regulares de muestreo con el fin de evitar la posterior correlación espacial de las colectas).



Datos ambientales: Bioclimáticos ([Worldclim](https://worldclim.org/))

- Es un conjunto de capas climáticas globales (19 variables en formato ráster) con una resolución espacial de aproximadamente 1 kilómetro cuadrado (30" de arco).
- Tipos de datos de descarga:
 - **Condiciones actuales:** interpolaciones de datos mensuales observados (temperaturas media, mínima y máxima y precipitación media), representativos de l periodo1950-2000.
 - **Condiciones futuras:** datos correspondientes al modelo climático global (GCM) y derivados del proyecto CMIP5 (IPPC Fifth Assessment).
 - **Condiciones pasadas:** modelo climático global (GCM).
- Variables derivadas de los datos mensuales:
 - 19 variables bioclimáticas.





Datos ambientales: Bioclimáticos (Moscow Forestry Sciences Laboratory)

- A Spline Model of Climate for the Western United States ([Rehfeldt G.E., 2006](#)).
- Es un conjunto de capas climáticas globales (formato ráster) con una resolución espacial de aproximadamente 1 kilómetro cuadrado (30" de arco).
- Tipos de datos de descarga:
 - **Condiciones actuales:** interpolaciones de datos mensuales observados (temperaturas media, mínima y máxima y precipitación media), representativos del periodo 1961-1990 y procedentes de 3.000 estaciones climáticas ubicadas en el oeste de Estados Unidos y suroeste de Canada.
 - **Condiciones futuras:** datos correspondientes al modelo climático global (GCM) y derivados del proyecto CMIP5 (IPPC Fifth Assessment).
- Variables derivadas de los datos mensuales:
 - 18 variables bioclimáticas.





Datos del terreno (MDE)

- El [Modelo Digital de Elevaciones](#) (MDE) es el conjunto de:
 - **Modelos Digitales del Terreno** (MDT), a nivel del suelo.
 - **Modelos Digitales de Superficies** (MDS), incluye copa de los árboles, viaductos, puentes, ...etc.
- Los Modelos Digitales del Terreno (MDT) se obtienen mediante interpolación de [modelos digitales del terreno de 5 m de paso de malla](#) procedentes del Plan Nacional de Ortofotografía Aérea (PNOA).



Otro tipo de variables

- Geología y Edafología.
- Coberturas de uso del suelo:
 - Mapa forestal de España.
 - Corinne Land Cover.
- Sensores remotos.- se pueden derivar coberturas de uso del suelo.
 - Resolución espectral.- Multiespectral, Hiperespectral.
 - Resolución espacial.- baja, media y alta.
 - Resolución temporal.- series temporales.
- Patrones de Paisaje.- directamente relacionado con sustentabilidad de hábitats.
- Distribución de otras especies.- interacciones con la distribución de otras especies.
- Especialmente en el caso de las topo-edáficas, se ha probado que su uso en forma conjunta con los datos climáticos, puede mejorar el resultado de los modelos a escalas espaciales de mayor detalle (Randin et al., 2009b).



Del concepto de Nicho Ecológico a los Modelos de Distribución

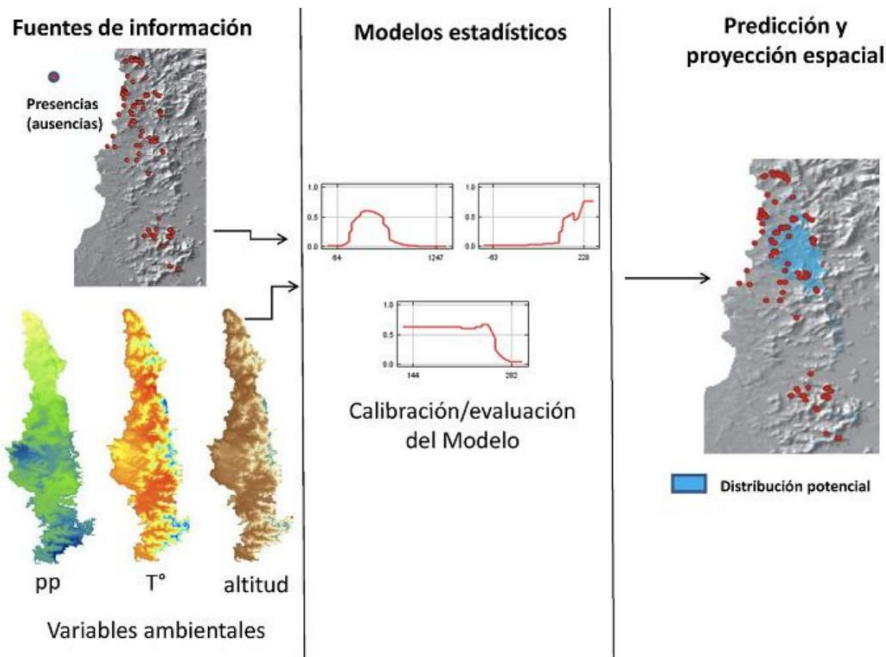
- Con el inicio de la masividad de los **Sistemas de Información Geográfica** en los años 90 y la mayor disponibilidad y manejo de grandes bases de datos, se dispone de nuevas herramientas para proyectar en el espacio ecológico y geográfico los modelos estadísticos de nicho. Es así como comienzan a surgir distintas aproximaciones estadísticas y programas computacionales que permiten determinar la distribución espacial de las especies y ecosistemas, basándose en datos de presencia/ausencia (Guisan y Zimmermann, 2000; Guisan y Thuiller, 2005; Elith y Leathwick, 2009).



Inferencia espacial y predicción

- Una vez definida la información de las presencias y variables, se debe seleccionar la **técnica de modelación más adecuada** para las preguntas que se quieran responder y proceder a los pasos básicos que involucra el trabajo con modelos.
- Procedimiento de inferencia espacial y predicción:
 - Calibración
 - Evaluación
 - Predicción
 - Proyección espacial.

Pasos metodológicos para la generación de modelos de distribución de especies/ecosistemas



Fuente: Modificado de Guisan *et al.*, 2002.



Calibración del modelo

- La calibración del modelo se refiere a los **análisis estadísticos asociados a buscar el ajuste del modelo** que posteriormente se proyectará en el espacio geográfico.
- El tipo de calibración dependerá de la técnica que se esté utilizando.

Comparación de algunas de las principales técnicas para la generación de modelos de distribución

Técnicas (BM = presente en <i>software</i> BIOMOD)	Nombre completo	Referencia	Estudios comparativos					
			A	B	C	D	E	F
ANN (BM)	<i>Artificial Neural Networks</i>	Lek <i>et al.</i> (1996)					+	
BIOCLIM (BM = SRE)	<i>Bioclimatic Envelope Algorithm</i>	Busby (1991)	-	-				±
BRT (BM)	<i>Boosting Regression Trees</i>	Friedman (2001)	+			+		
CART (BM)	<i>Classification And Regression Trees</i>	Vayssieres <i>et al.</i> (2000)			-	-	±	
ENFA	<i>Ecological Niche Factor Analysis</i>	Hirzel <i>et al.</i> (2002)					-	-
GAM (BM)	<i>Generalized Additive Models</i>	Hastie & Tibshirani (1990)	±	+	+		±	
GARP	<i>Genetic Algorithm for Rule-set Production</i>	Stockwell & Peters (1999)	±		-			+
GDM	<i>Generalised Dissimilarity Modelling</i>	Ferrier <i>et al.</i> (2007)	+					
GLM (BM)	<i>Generalized Linear Models</i>	McCullagh & Nelder (1989)	±		+		±	
MARS (BM)	<i>Multivariate Adaptive Regression Splines</i>	Friedman (1991)	±			-		
MAXENT	<i>Maximum Entropy</i>	Phillips <i>et al.</i> (2006)	+	+				
RF (BM)	<i>Random Forest</i>	Breiman (2001)				+		

Referencias para los estudios comparativos: A = Elith *et al.* (2006), B = Hijmans & Graham (2006), C = Meynard & Quinn (2007), D = Prasad *et al.* (2006), E = Segurado & Araujo (2004), F = Tsoar *et al.* (2007). La evaluación de las técnicas es indicada como: -- (muy mal), - (mal), ± (promedio) o + (buena).

Fuente: Modificado de Van Strien, 2008.



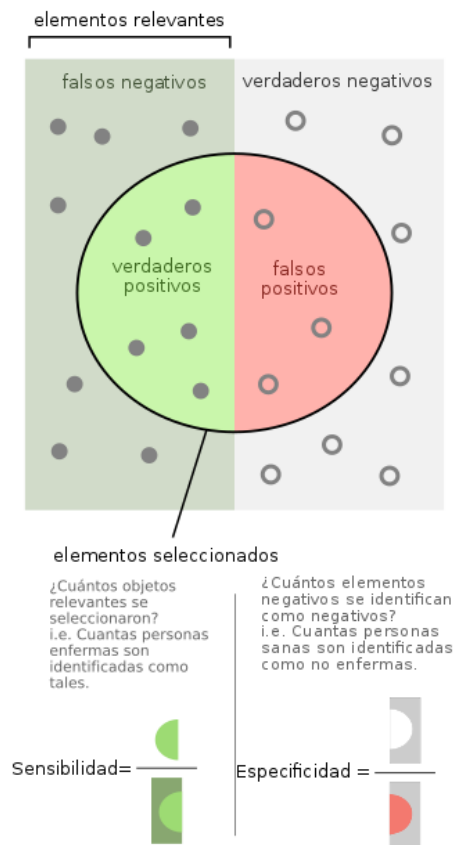
Evaluación del modelo

- La evaluación o validación se enfoca en la **caracterización de la confiabilidad de los resultados de los modelos**, más que en la definición de uno como correcto o incorrecto (Allouche et al., 2006).
- La forma correcta de evaluar el resultado sería **comparándolo con un conjunto de presencias independientes** al que se está usando, lo que es difícil de realizar porque se tienen datos de presencia limitados. Otro enfoque consiste en dividir el conjunto de datos en dos sets: uno de **calibración** (training) y otro de **evaluación** (test) (Guisan & Zimmermann, 2000).



Evaluación del modelo

- La forma más básica de evaluar un modelo consiste en generar una **matriz de confusión o error**, donde las ausencias y presencias predichas se comparan con las ausencias y presencias reales. Esto permite calcular indicadores para evaluar los modelos, como la **sensibilidad** (porcentaje de positivos verdaderos correctamente predichos) y la **especificidad** (porcentaje de negativos verdaderos correctamente predichos) (Hirzel et al., 2006).





Evaluación del modelo

- Existen **análisis estadísticos llamados umbral-independiente**, esto quiere decir que se aplican para evaluar la precisión del modelo sobre el mapa de valores continuos, sin la necesidad de calcular un valor de corte o umbral (Manel et al., 2001). Los de mayor utilización son **AUC**, **Kappa** y **TSS**. El método umbral-independiente más conocido es el **AUC** (Area Under the Curve) (Dodd & Pepe, 2003), que se obtiene comparando la **proporción de falsos positivos (1-especificidad) y verdaderos positivos (sensibilidad)**.
- La **precisión de un modelo** es mayor cuando la proporción de positivos verdaderos es mayor a la de positivos falsos; se estima que valores sobre 0,9 de AUC dan cuenta de modelos con alta precisión y valores entre 0,7 y 0,9 indican modelos con buena precisión.



Predicción del modelo y Proyección espacial

- El paso final en la generación de un modelo de distribución, es la **proyección en el espacio geográfico del modelo que ha sido calibrado y evaluado**. En este paso los **valores continuos de probabilidad**, o los **valores binarios de presencia/ausencia** obtenidos de un umbral, son transferidos al espacio geográfico, por lo que se pueden representar en forma cartográfica.
- El modelo de distribución se puede **proyectar en el espacio actual**, utilizando variables descriptoras actuales, pero también se puede **proyectar en el futuro o en el pasado**, utilizando variables descriptoras (modelos climáticos globales, tanto para escenarios futuros y pasados) que correspondan a los tiempos que se quieran analizar.