

# **Datos de trabajo:**

## **Variables climáticas y registros de presencia**

Variables ambientales y su selección

Fuentes de datos ambientales

Registros de presencia de especies

Extensión del área de estudio

Resolución espacial y escala de trabajo

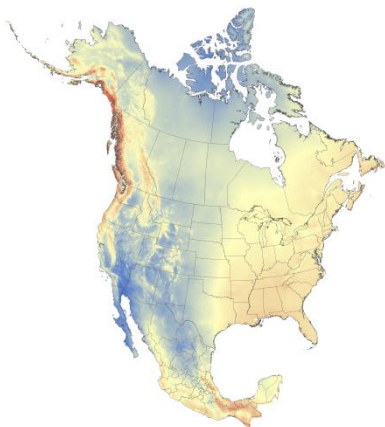
## **Variables ambientales (predictores)**

# Variables ambientales (predictores)

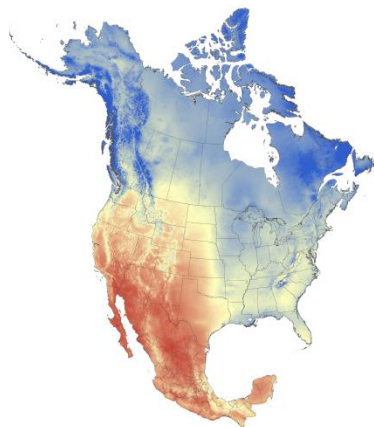
## Datos climáticos – variables bioclimáticas

*e.g., temperatura media anual, rango anual de temperatura, temperatura del mes más frío, días de crecimiento (temp. > 5°C o 10°C), número de días de helada, precipitación anual, precipitación en el mes más seco, evapotranspiración potencial, déficit hídrico, etc.*

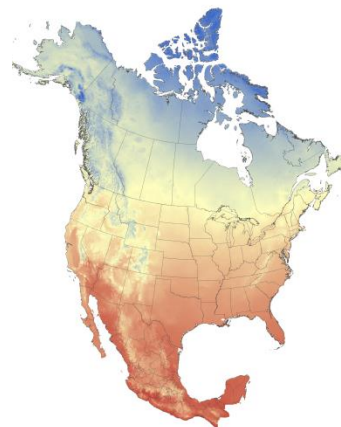
*MAP*



*CWD*



*DD5*



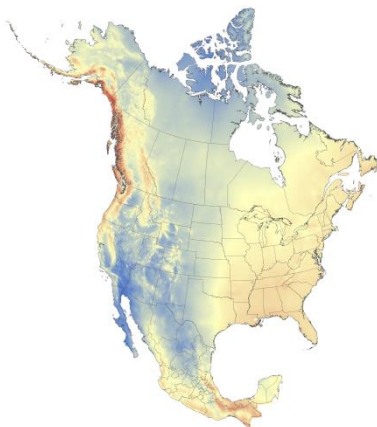
# Variables ambientales (predictores)

## Datos climáticos – variables bioclimáticas

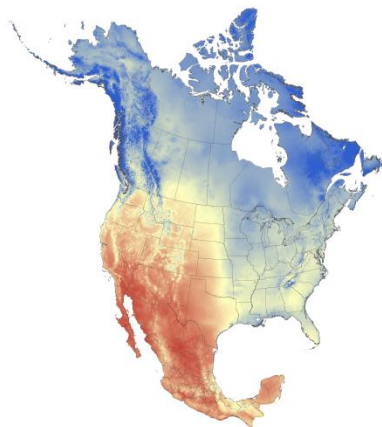
*e.g., temperatura media anual, rango anual de temperatura, temperatura del mes más frío, días de crecimiento (temp. > 5°C o 10°C), número de días de helada, precipitación anual, precipitación en el mes más seco, evapotranspiración potencial, déficit hídrico, etc.*

## Variables topográficas o de terreno (altitud, pendiente, orientación)

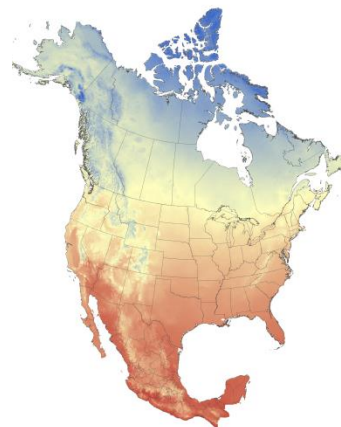
*MAP*



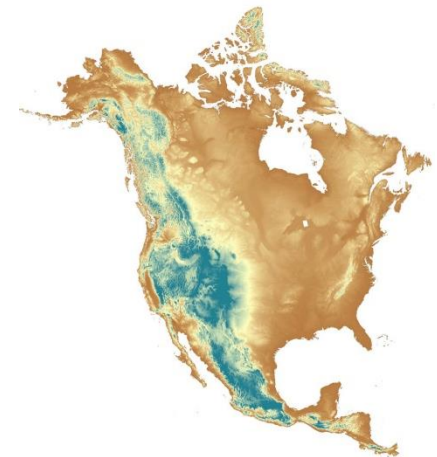
*CWD*



*DD5*



*Altitud*



# Variables ambientales (predictores)

## Datos climáticos – variables bioclimáticas

*e.g., temperatura media anual, rango anual de temperatura, temperatura del mes más frío, días de crecimiento (temp. > 5°C o 10°C), número de días de helada, precipitación anual, precipitación en el mes más seco, evapotranspiración potencial, déficit hídrico, etc.*

## Variables topográficas o de terreno (altitud, pendiente, orientación)

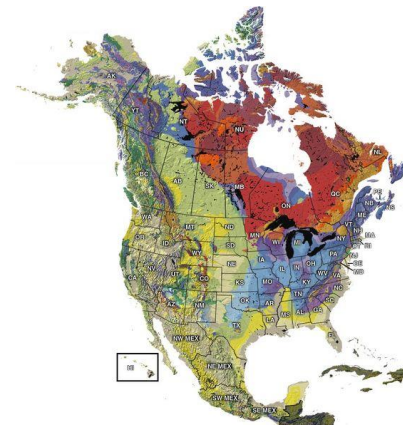
## Propiedades del suelo (textura, contenido de nutrientes)

## Vegetación o usos del suelo

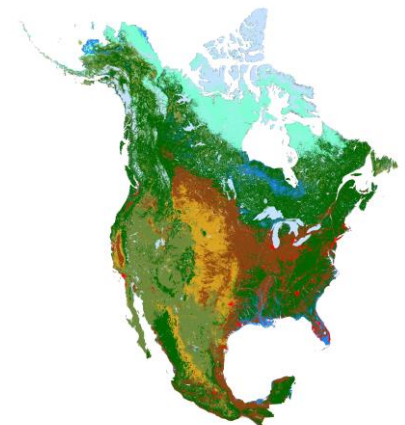
## Perturbaciones (e.g., fuego)

etc.

*Tipo de suelo*

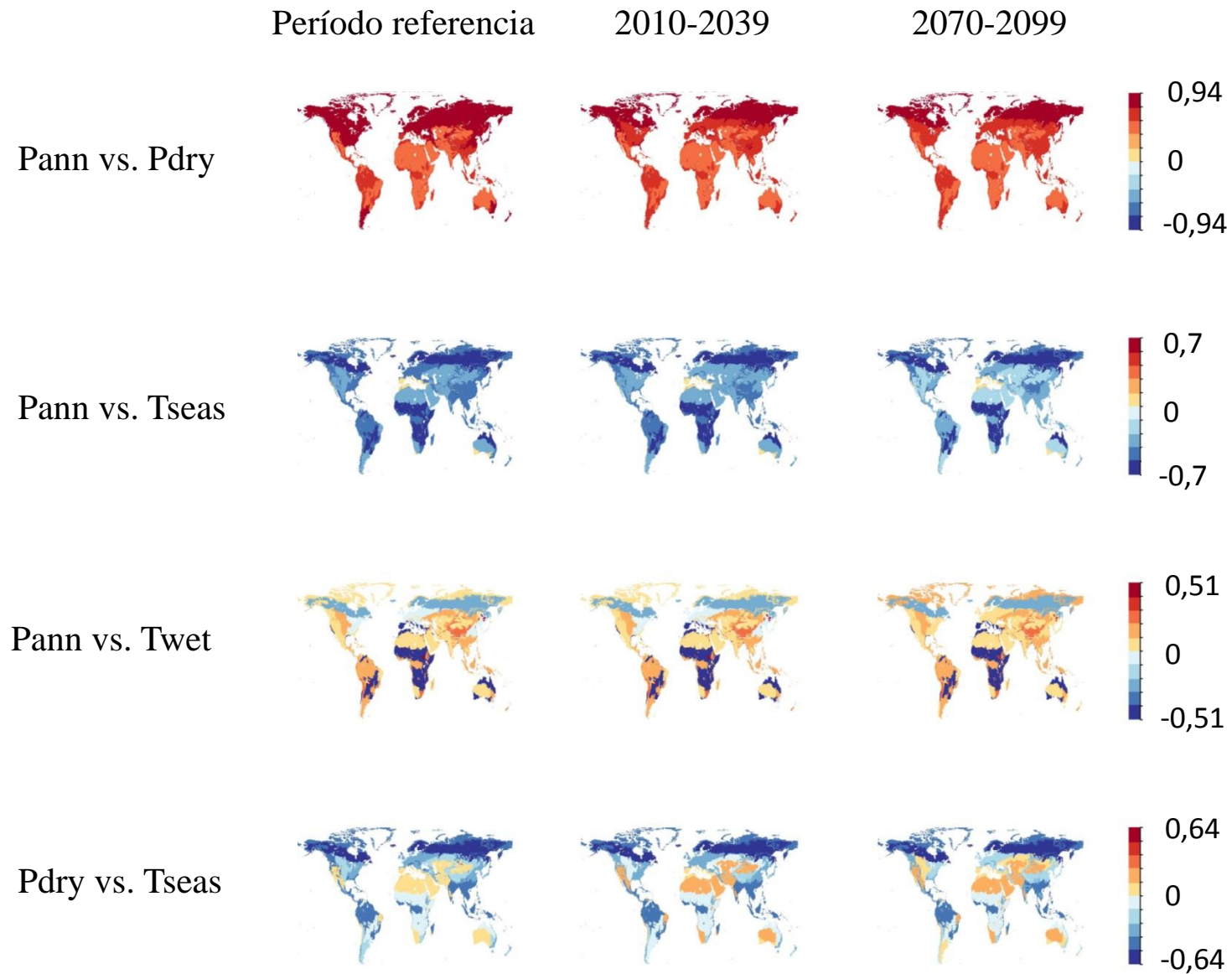


*Hábitat*



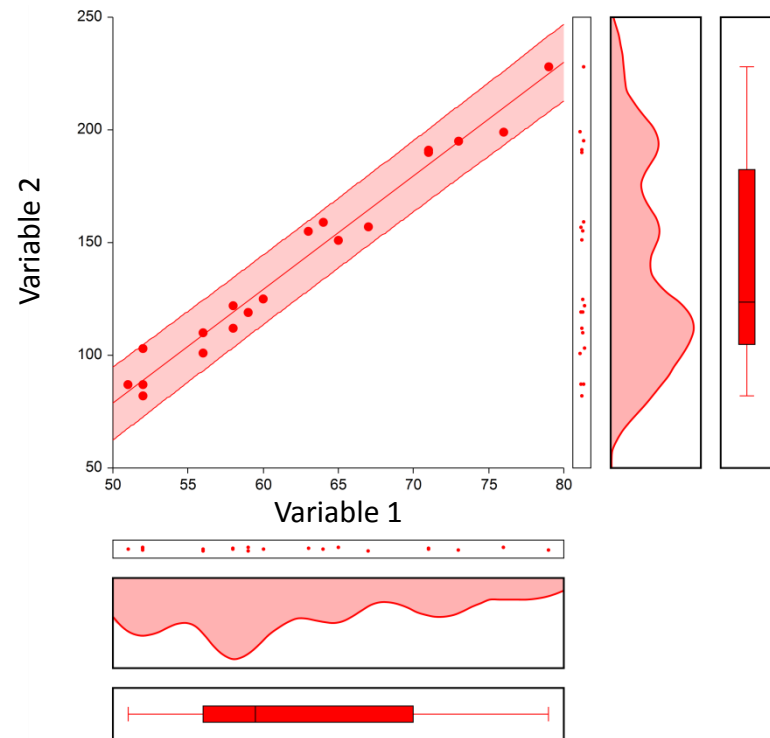
# Selección de variables - correlación y colinealidad

# Selección de variables - correlación y colinearidad



# Selección de variables - correlación y colinearidad

1. Exclusión de variables altamente correlacionadas (e.g., umbral correlación 0.8)
  - Dificultad de interpretación (funciones respuesta confusas)
  - Pueden derivar en modelos con sobre-predicción (*overprediction*)
  - Las correlaciones entre variables pueden no ser estables en el tiempo (proyecciones climáticas)





# Selección de variables - correlación y colinealidad

## 2. Factor de inflación de la varianza (*Variance inflation factor, VIF*)

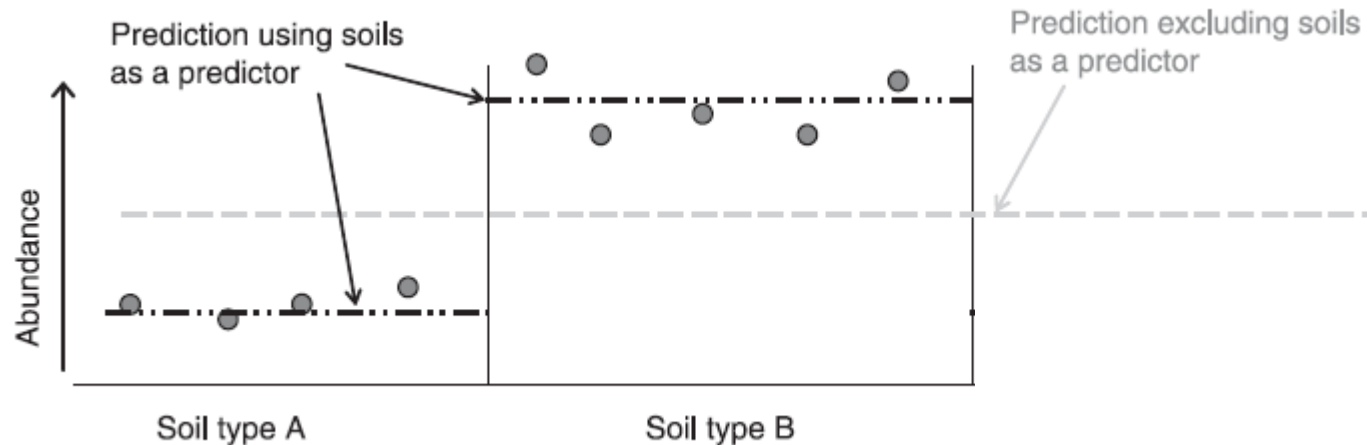
*“proporciona un índice que mide quanta parte de la varianza de un coeficiente de regresión es debida a la colinealidad”*

Valores altos de *VIF* indican que una variable predictora (*var. X*) se explica bien con una combinación lineal del resto de variables del conjunto utilizado. Por lo tanto, la inclusión de dicha variable (*var. X*) en el modelo es redundante.

Operacionalmente, valores de *VIF* por encima de 5 se suelen considerar una evidencia de colinealidad: la información proporcionada por un predictor con un  $VIF > 5$  está incluida en el resto de predictores.

# Selección de variables - correlación y colinearidad

## 3. Variables relevantes a nivel biológico



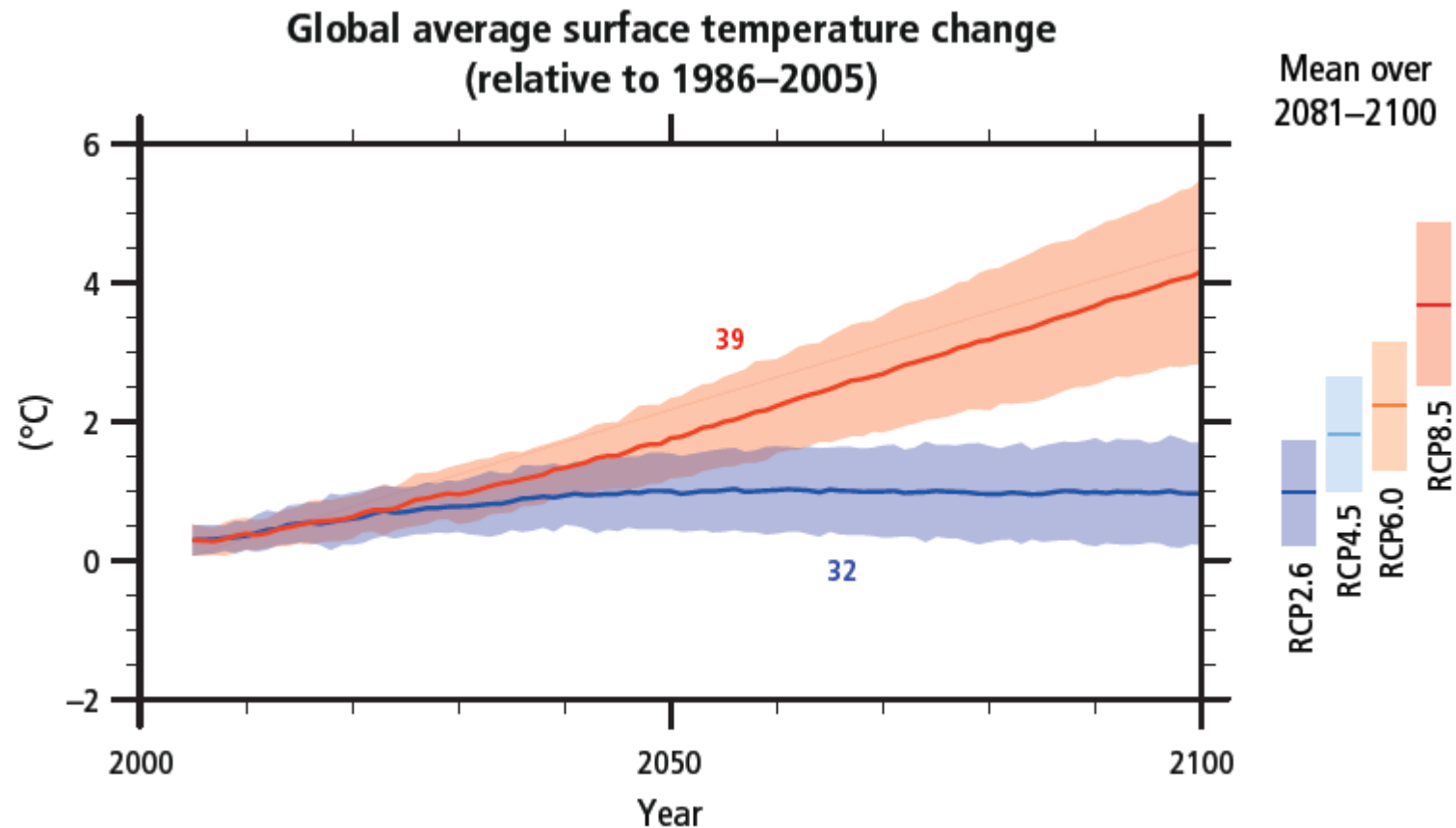
Barry & Elith 2002

### **Objetivo:**

incluir un conjunto de variables “suficiente” o representativo en vez de “exhaustivo”

# Escenarios climáticos – algunas consideraciones

# Escenarios climáticos – algunas consideraciones



# Escenarios climáticos – algunas consideraciones

Example:  
18 GCMs



## 1. All models are equally plausible.

- Average all 18 models together = ensemble



Ensemble

## 2. All models are equally plausible, but We want to plan for a range of scenarios.

- Select “bracketing” models



Warmest



Coolest

## 3. Some models perform better than others.

- Select best performing models for ensemble -

Filter 18 models based  
on performance  
criteria: trend,  
pressure, means, etc



Ensemble



Driest



Wettest



# Escenarios climáticos – algunas consideraciones

*Mean annual temperature change* for states and provinces of North America projected for the 2050s under the RCP4.5 scenario.

States and provinces are alphabetically sorted from left to right, AOGCMs are sorted by magnitude of projection for North America from top to bottom.

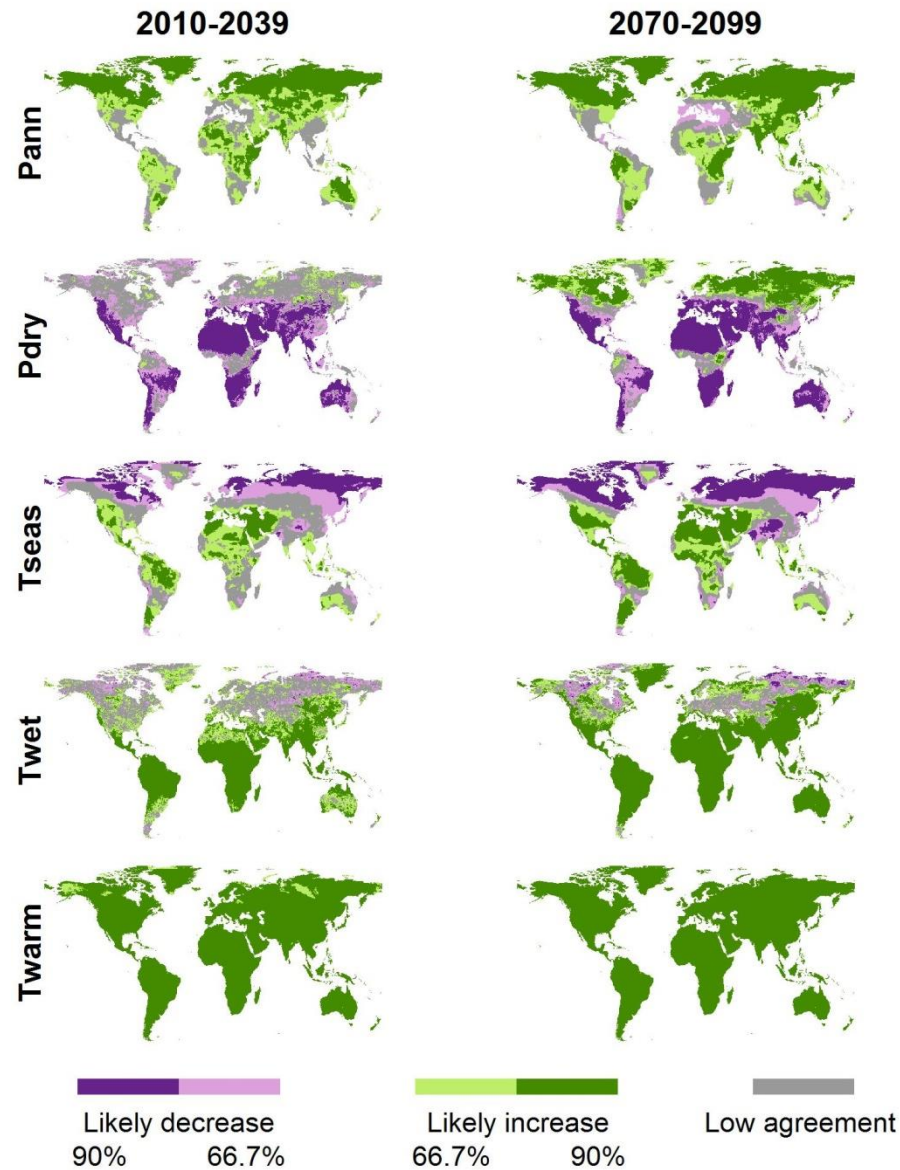
AOGCM	AB	AK	AL	AR	AZ	BC	CA	CO	CT	DC	DE	FL	GA	IA	ID	IL	IN	KS	KY	LA	MA	MB	MD	ME	MI	MN	MO	MS	MT	NB	NC	ND	NE	NH
INM-CM4	1.9	2.3	1	1.2	1.1	1.7	1.1	1.2	1.1	1.2	1	0.8	0.9	1.7	1.4	1.5	1.4	1.4	1.2	0.9	1.1	2	1.1	1.4	1.7	1.8	1.4	1	1.4	1.4	0.9	1.7	1.4	1.3
CNRM-CM5	2.3	3.1	1.7	1.9	2.3	2.1	1.9	2.4	2.2	2.1	2	1.3	1.7	2.2	2.3	2.1	2.2	2.1	2	1.7	2.3	3	2	2.4	2.6	2.5	1.9	1.7	2.4	2.5	1.8	2.6	2.3	2.4
CCSM4	2.8	3.6	2.1	2.2	2.2	2.7	1.9	2.3	2.2	2.2	2.1	1.7	2	2.7	2.7	2.3	2.3	2.3	2.3	1.9	2.3	3	2.2	2.4	2.6	2.9	2.3	2.1	2.6	2.4	2.1	2.9	2.4	2.5
MPI-ESM-LR	3	3.7	1.7	1.9	2.3	2.8	2	2.3	2.3	2.1	2	1.5	1.7	2.6	2.4	2.3	2.2	2.3	2	1.6	2.4	3.4	2.1	2.5	2.7	2.9	2.2	1.7	2.3	2.6	1.8	3	2.5	2.5
IPSL-CM5A-LR	3.2	3.6	2.5	2.7	2.9	3.2	2.7	3.1	2.9	2.8	2.7	2.1	2.4	3.3	3.3	3	2.9	2.8	2.7	2.5	2.9	3.3	2.7	3.2	3.1	3.4	2.9	2.6	3.2	3.3	2.5	3.4	3	3.2
HadGEM2-ES	3.6	4.7	2.9	3.2	2.9	3	2.6	3	3.2	3.3	2.9	2.3	2.8	3.3	3.1	3.4	3.5	3.2	3.4	2.9	3.3	4.2	3.2	3.4	3.6	3.5	3.3	3.1	3.1	3.5	2.8	3.7	3.2	3.4
GFDL-CM3	3.2	5.5	2.7	2.8	3.1	3.6	2.8	3.6	3.4	3.3	3.1	2.4	2.6	3.2	3.7	3.1	3.2	3.1	3.2	2.5	3.4	4.2	3.2	3.9	3.6	3.5	3.1	2.6	3.2	4	2.9	3.2	3.1	3.8
AOGCM	NJ	NL	NM	NS	NT	NV	NY	OH	OK	ON	OR	PA	PE	QC	RI	SC	SD	SK	TN	TX	UT	VA	VT	WA	WI	WV	WY	YT	Can	USA	ContUSA	NorAm		
INM-CM4	1.1	1.2	1.1	1.1	2	1.2	1.4	1.3	1.2	1.8	1.1	1.3	1.3	1.4	1	0.8	1.5	2	1.2	1.1	1.3	1.1	1.4	1.1	1.8	1.2	1.3	2.1	1.9	1.5	1.3	1.7		
CNRM-CM5	2.1	2.3	2.4	2.3	3.3	2.5	2.3	2.2	2	2.8	2.1	2.2	2.5	2.8	2.2	1.7	2.4	2.6	1.8	2	2.6	1.9	2.4	2	2.4	2.1	2.3	2.6	2.9	2.4	2.2	2.7		
CCSM4	2.2	2.5	2.1	2.1	3.7	2.5	2.5	2.3	2.1	2.7	2.2	2.4	2.3	2.9	2.1	2	2.7	2.9	2.3	1.9	2.7	2.2	2.5	2.3	2.8	2.3	2.6	3.2	3.3	2.7	2.3	3		
MPI-ESM-LR	2.1	3	2.2	2.4	3.8	2.2	2.4	2.2	2.1	3	2.1	2.3	2.6	3.2	2.3	1.7	2.6	3.2	1.9	2	2.4	2	2.5	2.2	2.7	2.1	2.3	3.3	3.4	2.6	2.2	3.1		
IPSL-CM5A-LR	2.8	3.3	3	2.9	3.5	3.3	3.2	2.9	2.8	3.2	2.9	2.9	3.3	3.4	2.8	2.4	3.2	3.3	2.6	2.7	3.4	2.7	3.3	2.9	3.3	2.8	3.1	3.2	3.4	3.1	3	3.3		
HadGEM2-ES	3.1	3.6	2.9	3.3	5.5	3	3.6	3.6	3	3.9	2.7	3.5	3.8	4.1	3.1	2.8	3.4	3.8	3.2	2.9	3.2	3.1	3.6	2.9	3.5	3.4	3.1	4.1	4.6	3.5	3.1	4.2		
GFDL-CM3	3.3	5	3.2	3.5	7.3	3.5	3.8	3.4	3	4.3	2.9	3.5	3.8	5.6	3.2	2.7	3.1	3.6	3	2.8	3.9	3.2	3.9	2.9	3.5	3.4	3.4	4.7	5.7	3.8	3.2	5		

*Mean annual precipitation change* for states and provinces of North America projected for the 2050s under the RCP4.5 scenario.

States and provinces are alphabetically sorted from left to right, AOGCMs are sorted by magnitude of projection for North America from top to bottom.

AOGCM	AB	AK	AL	AR	AZ	BC	CA	CO	CT	DC	DE	FL	GA	IA	ID	IL	IN	KS	KY	LA	MA	MB	MD	ME	MI	MN	MO	MS	MT	NB	NC	ND	NE	NH
INM-CM4	6	10	-4	1	-4	5	-6	-2	-1	-2	-3	1	-2	-2	-6	1	1	2	-1	-2	0	3	-2	3	1	3	2	-2	-4	3	-3	0	1	1
IPSL-CM5A-LR	4	13	-2	-1	-11	5	0	0	3	3	4	-1	-1	1	3	3	1	-4	1	-6	3	7	3	6	1	1	1	-3	-1	5	1	0	-1	4
CCSM4	4	17	5	2	-3	1	0	5	7	8	8	5	10	3	-1	0	1	1	2	4	6	3	7	6	2	1	0	4	3	7	10	1	4	5
CNRM-CM5	7	15	2	2	1	5	10	4	11	8	7	3	1	12	4	10	9	9	8	-1	10	8	8	10	10	10	11	2	3	9	6	5	6	10
MPI-ESM-LR	5	19	5	6	0	11	4	3	1	6	5	3	4	7	0	6	7	2	8	3	0	7	5	3	6	5	6	5	2	5	5	2	5	0
HadGEM2-ES	7	24	0	-2	-5	4	-1	4	11	4	5	0	4	14	-3	8	5	6	0	-7	13	11	4	13	10	12	6	-5	3	13	5	12	12	13
GFDL-CM3	24	27	9	6	4	15	-5	8	13	14	15	13	11	9	4	9	11	3	8	9	12	9	14	19	13	14	5	8	10	19	11	15	9	17
AOGCM	NJ	NL	NM	NS	NT	NV	NY	OH	OK	ON	OR	PA	PE	QC	RI	SC	SD	SK	TN	TX	UT	VA	VT	WA	WI	WV	WY	YT	Can	USA	ContUSA	NorAm		
INM-CM4	-4	3	-5	1	9	-5	0	0	1	6	-6	-2	1	5	-1	-4	0	5	-4	-4	-3	-2	2	-5	2	0	-1	6	6.8	1	-2.1	4.5		
CNRM-CM5	4	9	-11	3	14	4	5	2	-5	9	-1	3	2	12	3	0	-1	4	0	-11	-1	2	5	-2	1	2	5	10	11	2.2	-1.4	7.5		
CCSM4	7	5	1	7	15	4	3	1	0	2	-2	3	6	6	7	11	5	3	3	0	0	8	4	-4	1	4	7	14	9.5	6.1	2.1	8.2		
MPI-ESM-LR	9	6	1	10	15	7	10	10	5	9	1	8	10	10	11	4	9	10	6	-2	5	6	10	-2	10	8	5	9	11	7.7	5.2	9.9		
IPSL-CM5A-LR	3	10	2	4	19	8	3	6	5	8	-2	4	6	12	1	5	4	5	6	1	-2	6	1	-2	6	5	2	18	14	7.2	3.2	11.4		
HadGEM2-ES	7	14	-5	12	27	3	8	1	0	12	-4	4	9	15	12	6	13	11	-2	-6	1	4	12	-3	12	1	4	20	19	8	2.6	14.6		
GFDL-CM3	15	15	5	13	36	-2	16	12	3	12	-4	14	17	24	11	13	13	13	8	2	2	11	18	1	11	9	14	30	27	12	6.9	20.9		

# Escenarios climáticos – algunas consideraciones



# Fuentes de datos ambientales

## Escala global

- **WorldClim** – datos bioclimáticos, varias resoluciones (<http://www.worldclim.org/>)
- **CRU TS** datos bioclimáticos (<https://crudata.uea.ac.uk/cru/data/hrg/>)
- **SPEI** (*standardized precipitation-evapotranspiration index*) (<http://spei.csic.es/map/maps.html#months=1#month=11#year=2016>)
- **CGIAR-CSI Consortium for spatial information** (<http://www.cgiar-csi.org/data>)
- **SEDAC Socioeconomic data and application center** (<http://sedac.ciesin.columbia.edu/data/sets/browse>)

## Escala continental

- **Adaptwest** – datos climáticos Norte América resolución 1 km (<https://adaptwest.databasin.org/pages/adaptwest-climatena>)

## Escala regional

- **EEA European Environment Agency** (<http://www.eea.europa.eu/data-and-maps>)
- **SEDAC Socioeconomic data and application center** (<http://sedac.ciesin.columbia.edu/data/sets/browse>)



# Datos de presencia de especies


# Datos de presencia de especies

- Muestreo (e.g., parcelas, transectos)
- Datos de inventarios
- Atlas de distribución
- Global biodiversity information facility (GBIF)

# Datos de presencia de especies

- Muestreo (e.g., parcelas, transectos)
- Datos de inventarios
- Atlas de distribución
- Global biodiversity information facility (GBIF)

## Propiedades básicas:

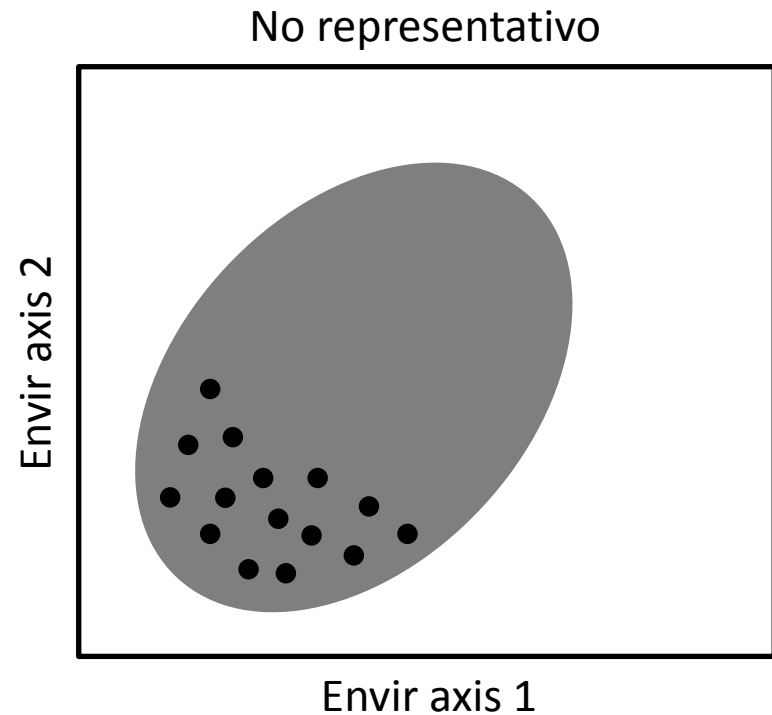
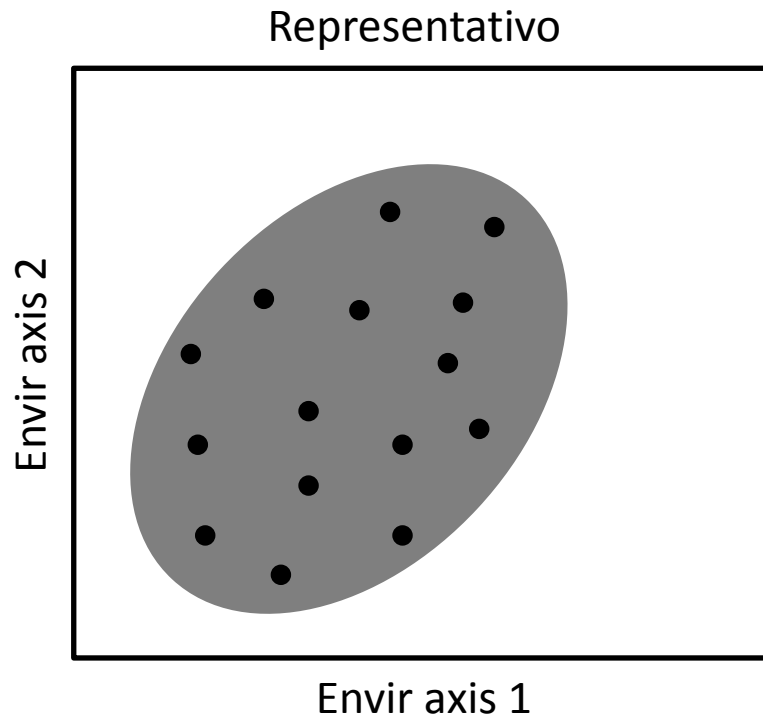
- “Grano” (observación simple, punto de recuento, parcela) y extensión espacial (tamaño del área de estudio)
  - Densidad de muestreo =  $\text{total área muestreada} / \text{extensión área estudio}$
  - Escala temporal de las observaciones (estación año, duración muestreo)
-  *en la mayoría de aproximaciones de modelización se asume que la distribución de las especies es estática en el tiempo y en el espacio*

# Diseño espacial de muestreo

# Diseño espacial de muestreo

El objetivo del muestreo (des de una perspectiva de modelización) es la selección de unidades del total de la población para estimar la extensión o rango biogeográfico de una variable en el espacio, o para interpolar su valor en localidades no muestreadas.

1. **Tamaño de muestra** – número de observaciones. Objetivo: recoger el mínimo conjunto de datos (optimización de muestreo) para obtener estimaciones o interpolaciones precisas.

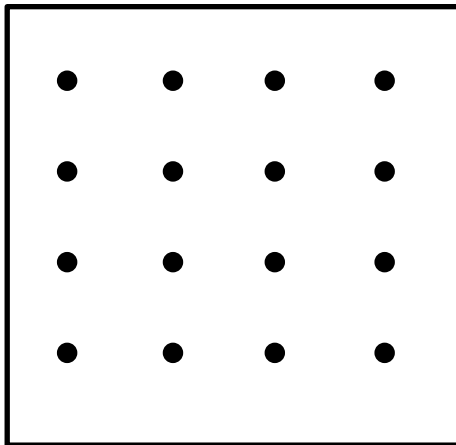


# Diseño espacial de muestreo

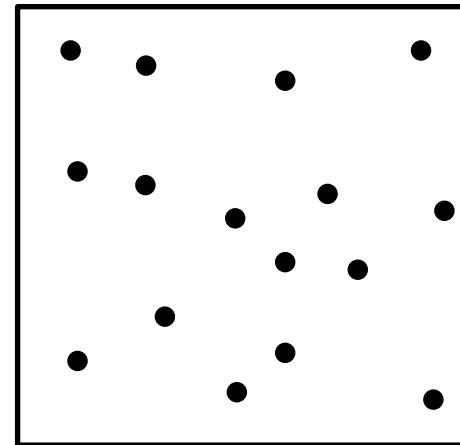
El objetivo del muestreo (des de una perspectiva de modelización) es la selección de unidades del total de la población para estimar la extensión o rango biogeográfico de una variable en el espacio, o para interpolar su valor en localidades no muestreadas.

1. **Tamaño de muestra** – número de observaciones. Objetivo: recoger el mínimo conjunto de datos (optimización de muestreo) para obtener estimaciones o interpolaciones precisas.
2. **Diseño de muestreo** – sistemático, aleatorio, aleatorio con estratificación

Sistemático



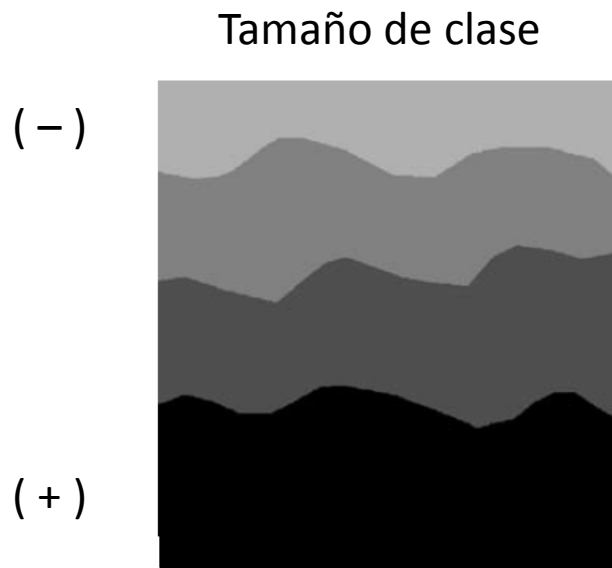
Aleatorio



# Diseño espacial de muestreo

## Muestreo aleatorio con estratificación

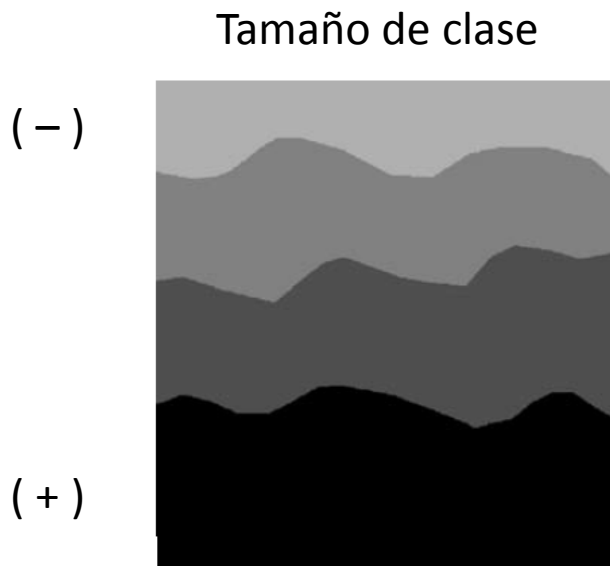
Área de estudio hipotética con 4 clases con distintas propiedades (e.g., altitud, temperatura, densidad de población,...)



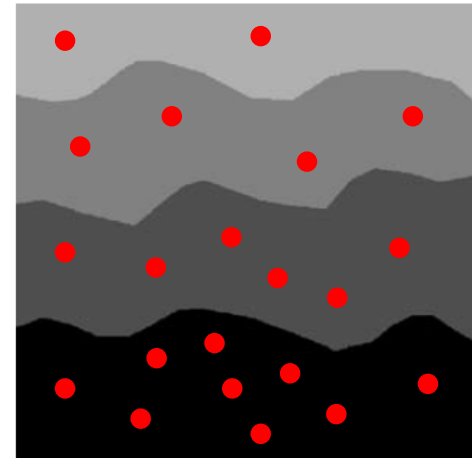
# Diseño espacial de muestreo

## Muestreo aleatorio con estratificación

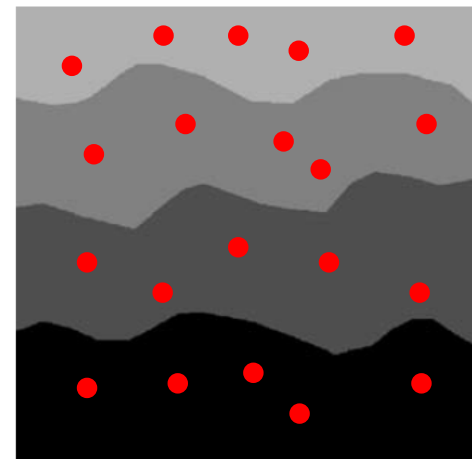
Área de estudio hipotética con 4 clases con distintas propiedades (e.g., altitud, temperatura, densidad de población,...)



## Muestreo proporcional



## Muestreo "equitativo"

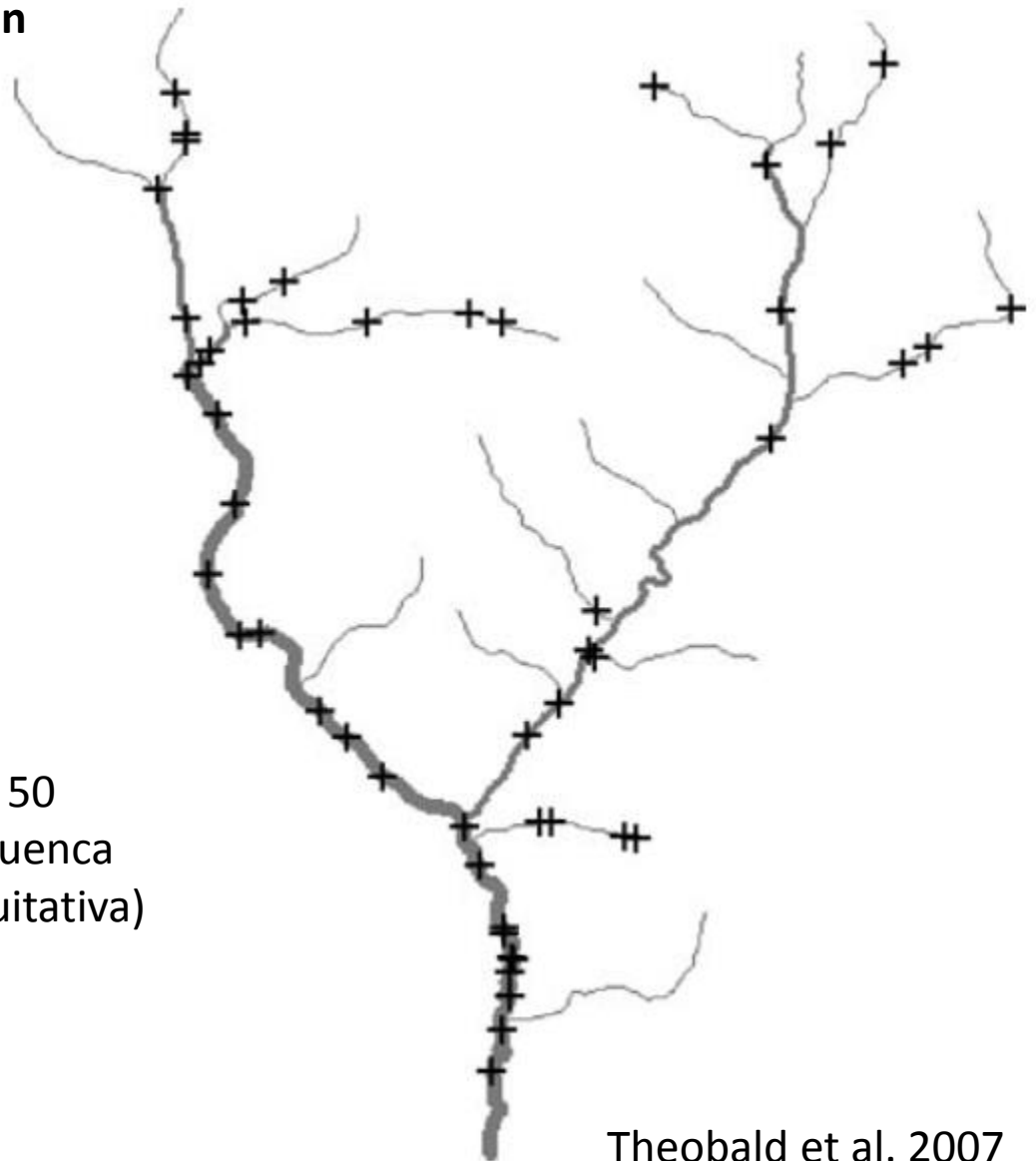




# Diseño espacial de muestreo

## Muestreo aleatorio con estratificación

Ejemplo de una muestra aleatoria de 50 puntos de muestreo a lo largo de la cuenca fluvial (probabilidad de muestreo equitativa)



Theobald et al. 2007

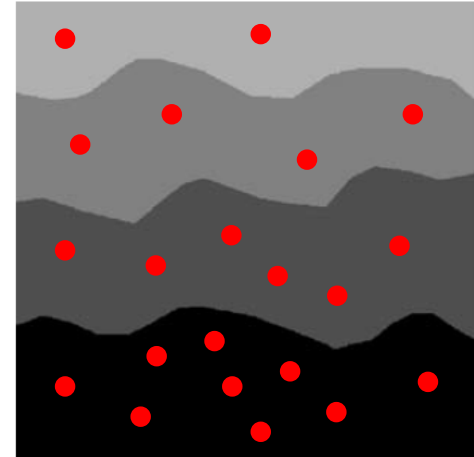
# Diseño espacial de muestreo

## Muestreo aleatorio con estratificación

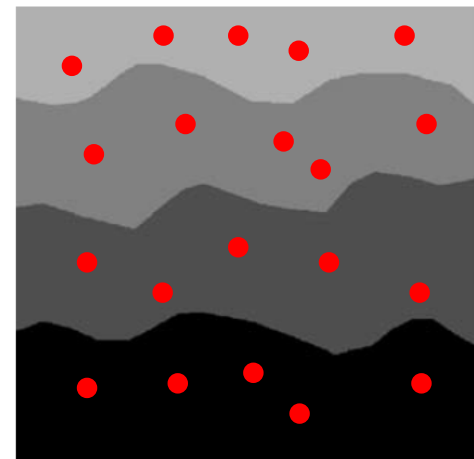
La estratificación puede ser definida en base a:

- Variables ambientales relacionadas con la distribución de la especie de interés
- Rango geográfico (para asegurar que todas las regiones son muestreadas).

### Muestreo proporcional



### Muestreo “equitativo”



# Diseño espacial de muestreo

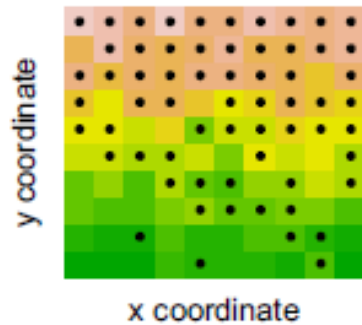
El objetivo del muestreo (des de una perspectiva de modelización) es la selección de unidades del total de la población para estimar la extensión o rango biogeográfico de una variable en el espacio, o para interpolar su valor en localidades no muestreadas.

1. **Tamaño de muestra** – número de observaciones. Objetivo: recoger el mínimo conjunto de datos (optimización de muestreo) para obtener estimaciones o interpolaciones precisas.
2. **Diseño de muestreo** – sistemático, aleatorio, aleatorio con estratificación
3. **Prevalencia** – frecuencia (o proporción) de datos de presencia en relación a las ausencias. En general – la capacidad predictora de los SDM mejora cuando la prevalencia de las especies es cercana al 50%.

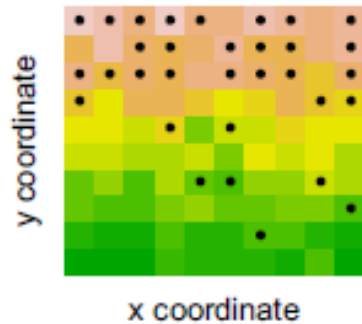
# Diseño espacial de muestreo

## Efectos de la prevalencia en el modelo

**Alta**  
prevalencia



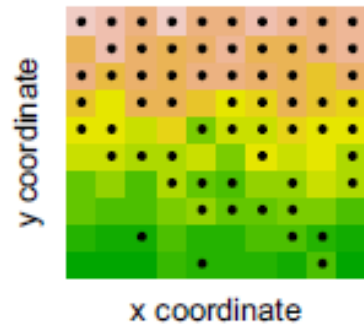
**Baja**  
prevalencia



# Diseño espacial de muestreo

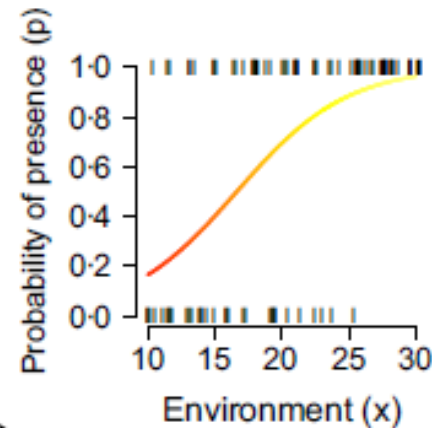
## Efectos de la prevalencia en el modelo

**Alta**  
prevalencia

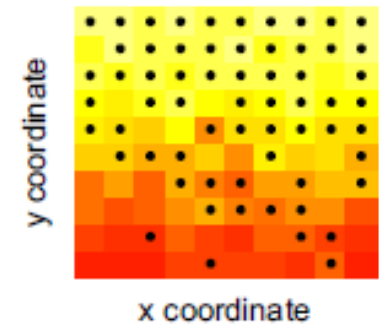


Model

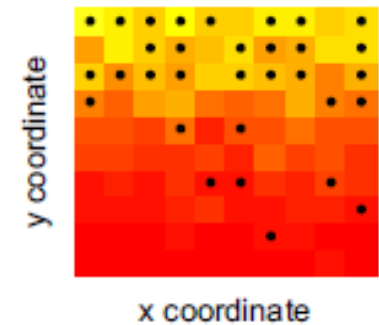
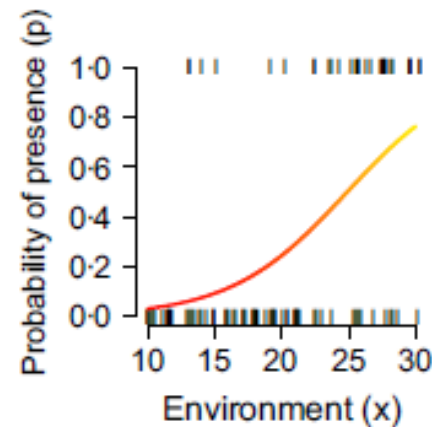
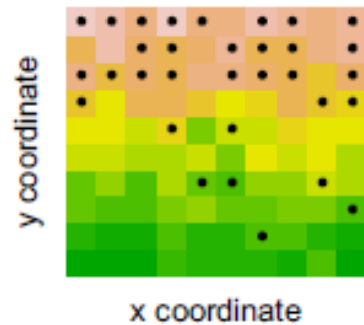
Nicho "climático"



Distribución  
geográfica



**Baja**  
prevalencia



# Diseño espacial de muestreo

El objetivo del muestreo (des de una perspectiva de modelización) es la selección de unidades del total de la población para estimar la extensión o rango biogeográfico de una variable en el espacio, o para interpolar su valor en localidades no muestreadas.

1. **Tamaño de muestra** – número de observaciones. Objetivo: recoger el mínimo conjunto de datos (optimización de muestreo) para obtener estimaciones o interpolaciones precisas.
2. **Diseño de muestreo** – sistemático, aleatorio, aleatorio con estratificación
3. **Prevalencia** – frecuencia (o proporción) de datos de presencia en relación a las ausencias. En general – la capacidad predictora de los SDM mejora cuando la prevalencia de las especies es cercana al 50%.
4. **Resolución de muestreo** – en general los registros de las especies (coord. x-y) son más precisos que la resolución de las variables ambientales. Pero en algunos casos (e.g., modelización áreas de forrajeo) los registros de la especie pueden estar caracterizados por más de una ‘observación’ de la variable ambiental.

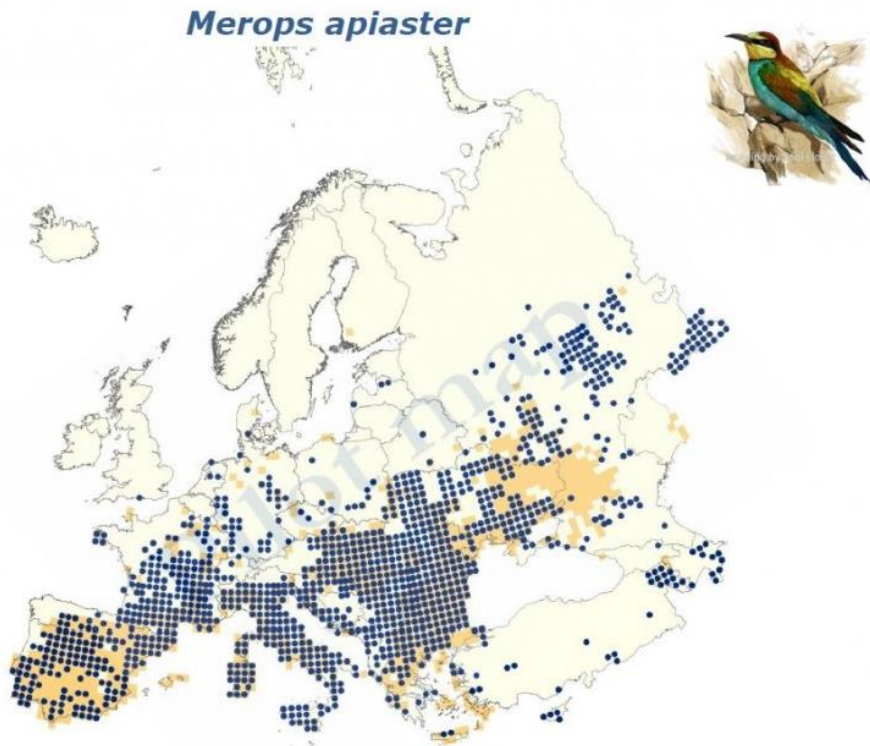
# Diseño espacial de muestreo

El objetivo del muestreo (des de una perspectiva de modelización) es la selección de unidades del total de la población para estimar la extensión o rango biogeográfico de una variable en el espacio, o para interpolar su valor en localidades no muestreadas.

1. **Tamaño de muestra** – número de observaciones. Objetivo: recoger el mínimo conjunto de datos (optimización de muestreo) para obtener estimaciones o interpolaciones precisas.
2. **Diseño de muestreo** – sistemático, aleatorio, aleatorio con estratificación
3. **Prevalencia** – frecuencia (o proporción) de datos de presencia en relación a las ausencias. En general – la capacidad predictora de los SDM mejora cuando la prevalencia de las especies es cercana al 50%.
4. **Resolución de muestreo** – en general los registros de las especies (coord. x-y) son más precisos que la resolución de las variables ambientales. Pero en algunos casos (e.g., modelización áreas de forrajeo) los registros de la especie pueden estar caracterizados por más de una ‘observación’ de la variable ambiental.
5. **Extensión del área de estudio** – afectará la relación estimada entre la distribución de la especie y los predictores ambientales (i.e., descripción del nicho).

# Diseño espacial de muestreo

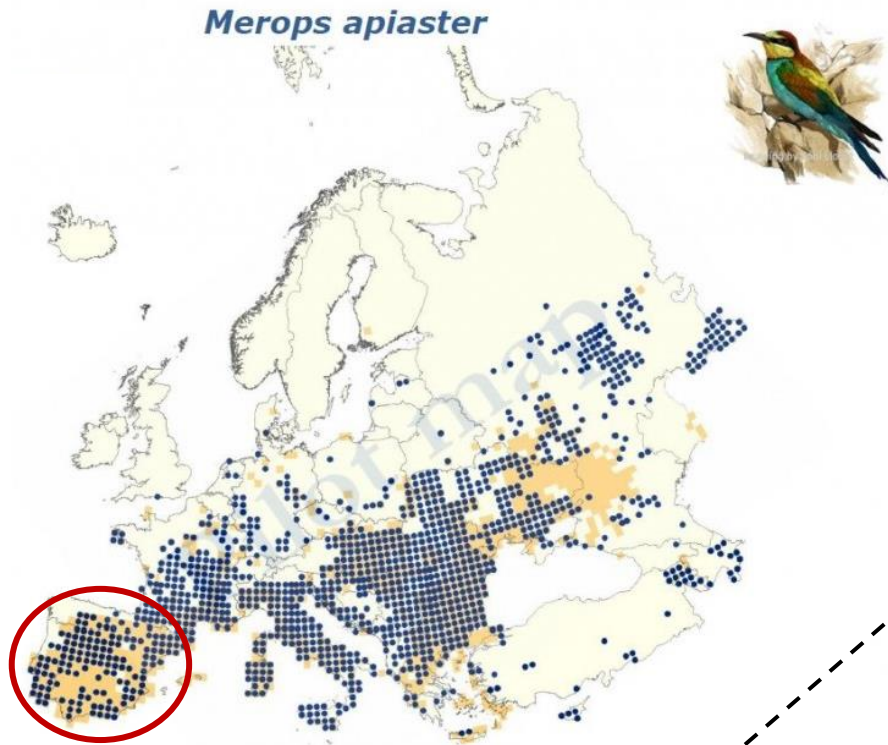
## Extensión del área de estudio





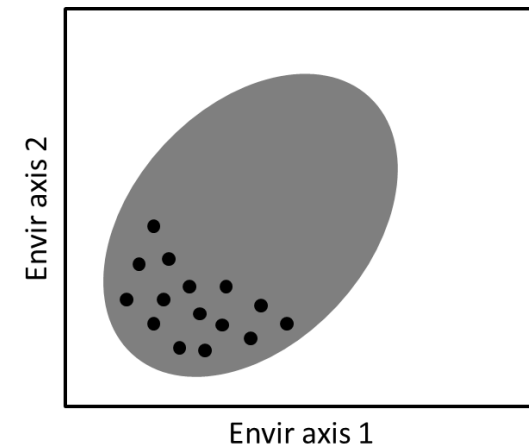
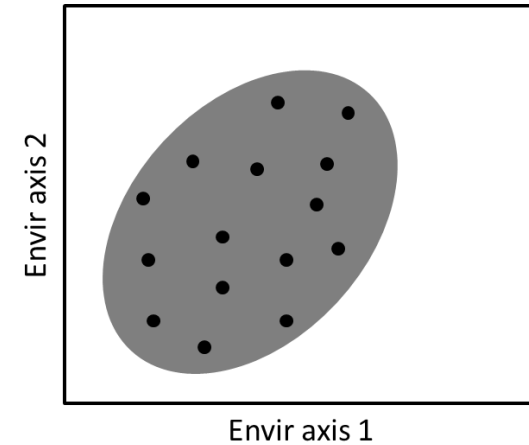
# Diseño espacial de muestreo

## Extensión del área de estudio



Área geográfica de estudio

Área de estudio en el espacio “ambiental”



# Resolución espacial y escala de trabajo

# Resolución espacial y escala de trabajo

## Ajuste resolución variables predictoras

Predictor 1

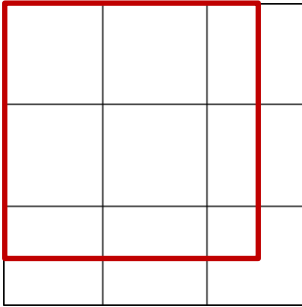

Predictor 2


Predictor 3

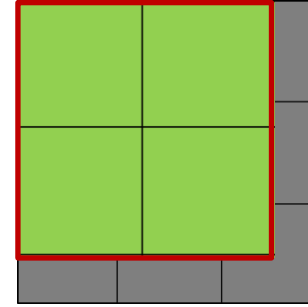

# Resolución espacial y escala de trabajo

## Ajuste resolución variables predictoras

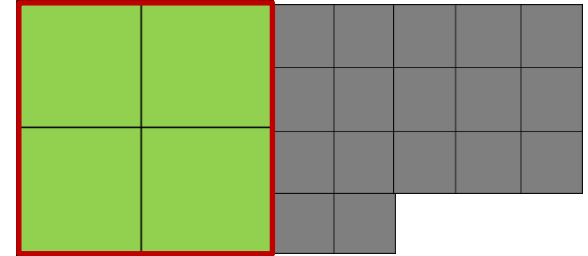
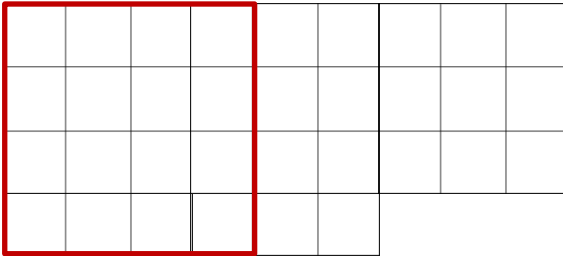
Predictor 1



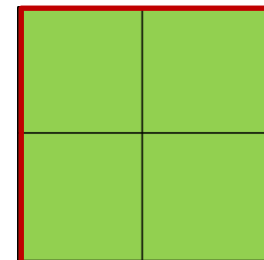
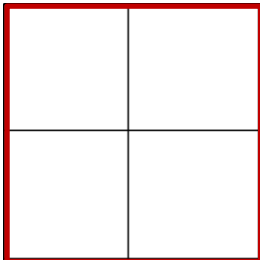
Ajuste resolución  
y extensión



Predictor 2



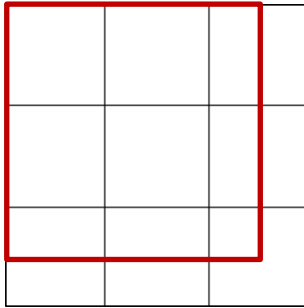
Predictor 3



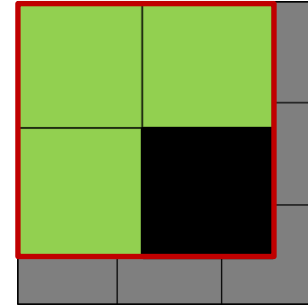
# Resolución espacial y escala de trabajo

## Ajuste resolución variables predictoras

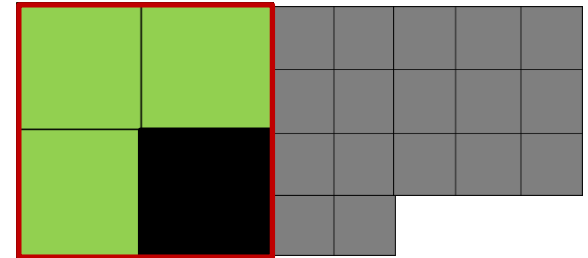
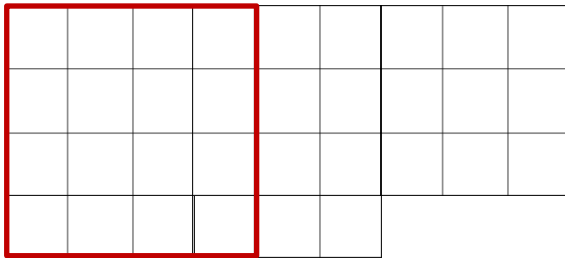
Predictor 1



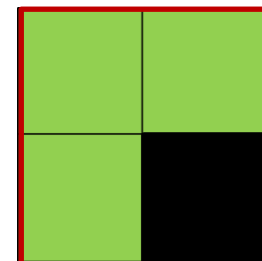
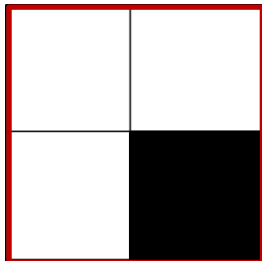
Ajuste resolución  
y extensión



Predictor 2



Predictor 3



Registros nulos  
(*missing data*)

# Resolución espacial y escala de trabajo

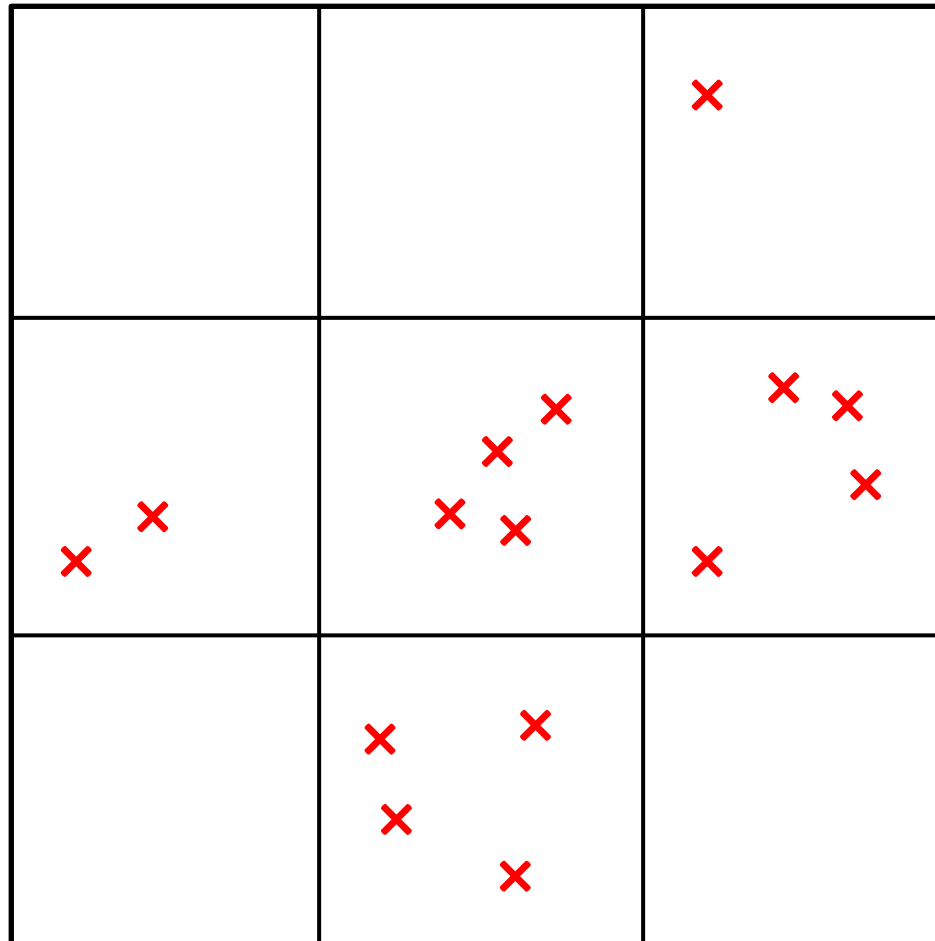
## Ajuste resolución variables predictoras y registros presencia especies

Variable ambiental


# Resolución espacial y escala de trabajo

## Ajuste resolución variables predictoras y registros presencia especie

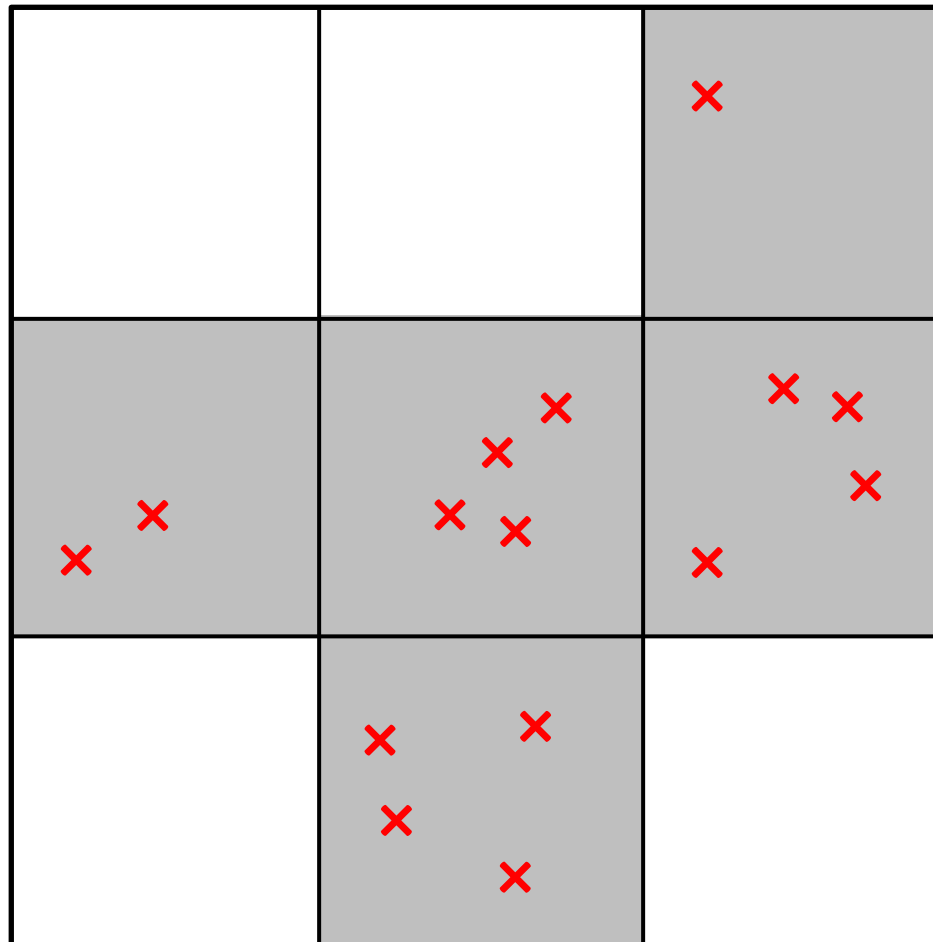
Variable ambiental + observaciones especie (x)



# Resolución espacial y escala de trabajo

## Ajuste resolución variables predictoras y registros presencia especies

Grid de “presencia/ausencia”

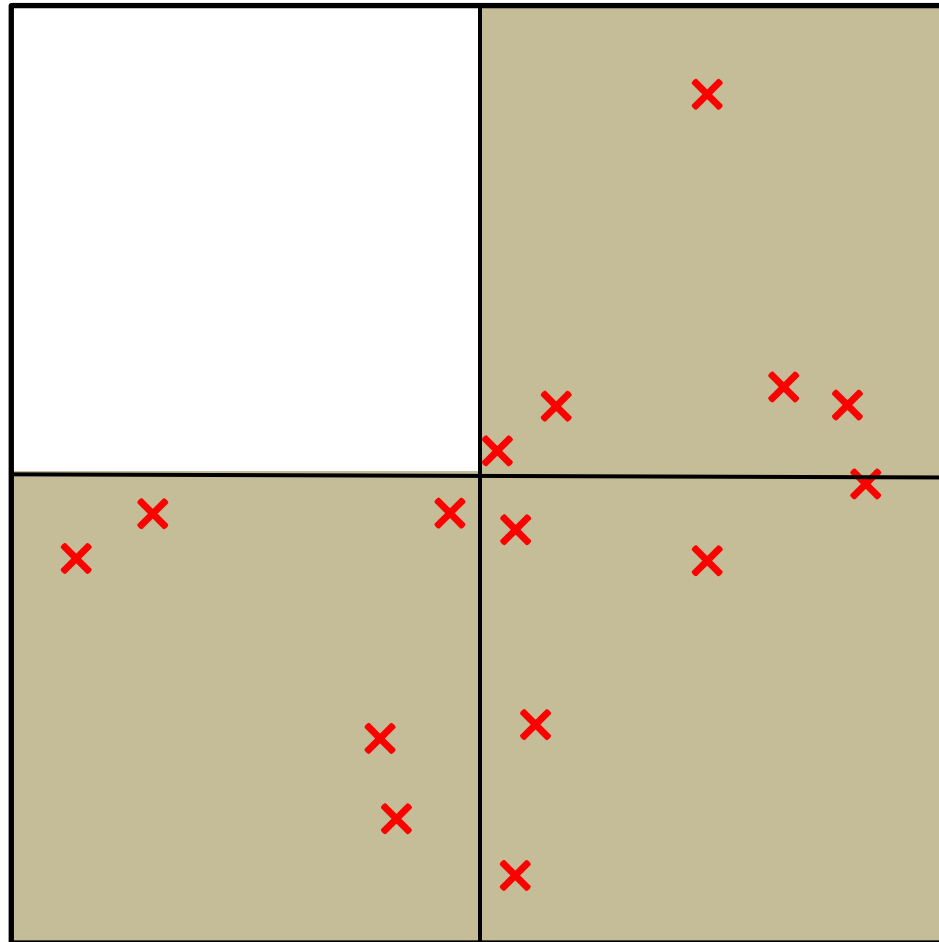




# Resolución espacial y escala de trabajo

## Ajuste resolución variables predictoras y registros presencia especies

Grid de “presencia/ausencia” (resolución menor)



# Resolución espacial y escala de trabajo

## Ajuste resolución variables predictoras y registros presencia especies

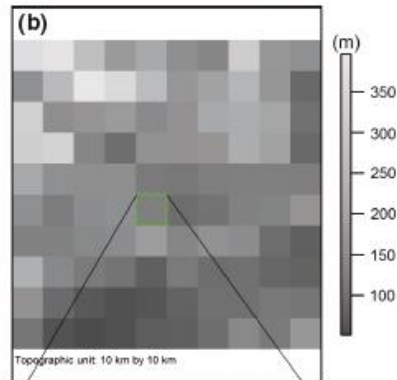
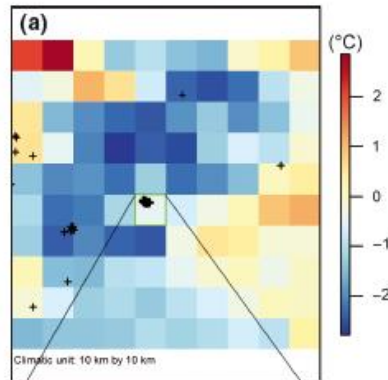
- La agregación espacial de las observaciones aumenta la fiabilidad de las “presencias”
- Toda la variabilidad (a pequeña escala) se pierde a nivel de pixel
- Una baja resolución (i.e., tamaño de pixel grande) de las variables ambientales puede no reflejar correctamente las condiciones ambientales a las que realmente están sujetas la especies.

# Resolución espacial y escala de trabajo

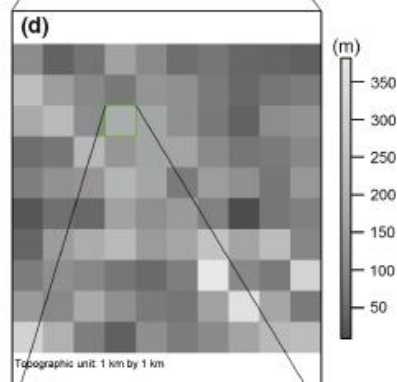
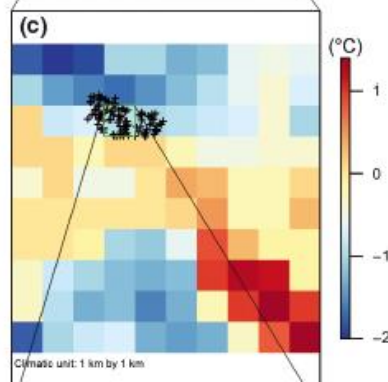
Clima

Topografía

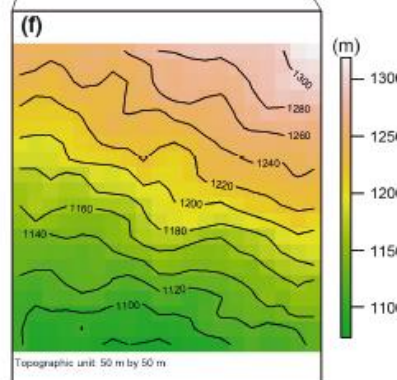
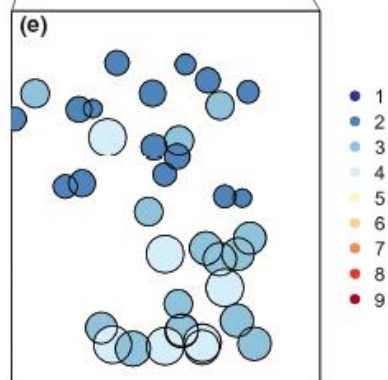
resolución  
10 km



resolución  
1 km



resolución  
50 m



( - )

Variabilidad topo-climática

( + )

Lenoir et al. 2013 GCB

# Resolución espacial y escala de trabajo

## Ajuste resolución variables predictoras y registros presencia especies

- La agregación espacial de las observaciones aumenta la fiabilidad de las “presencias”
- Toda la variabilidad (a pequeña escala) se pierde a nivel de pixel
- Una baja resolución (i.e., tamaño de pixel grande) de las variables ambientales puede no reflejar correctamente las condiciones ambientales a las que realmente están sujetas la especies.



En muchos casos, la distribución de las especies es conocida – las inferencias en base a SDM son útiles para entender la relación entre la distribución de las especies y las variables ambientales a escala biogeográfica, y para predecir cambios temporales de distribución bajo el cambio climático.

**“Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)**

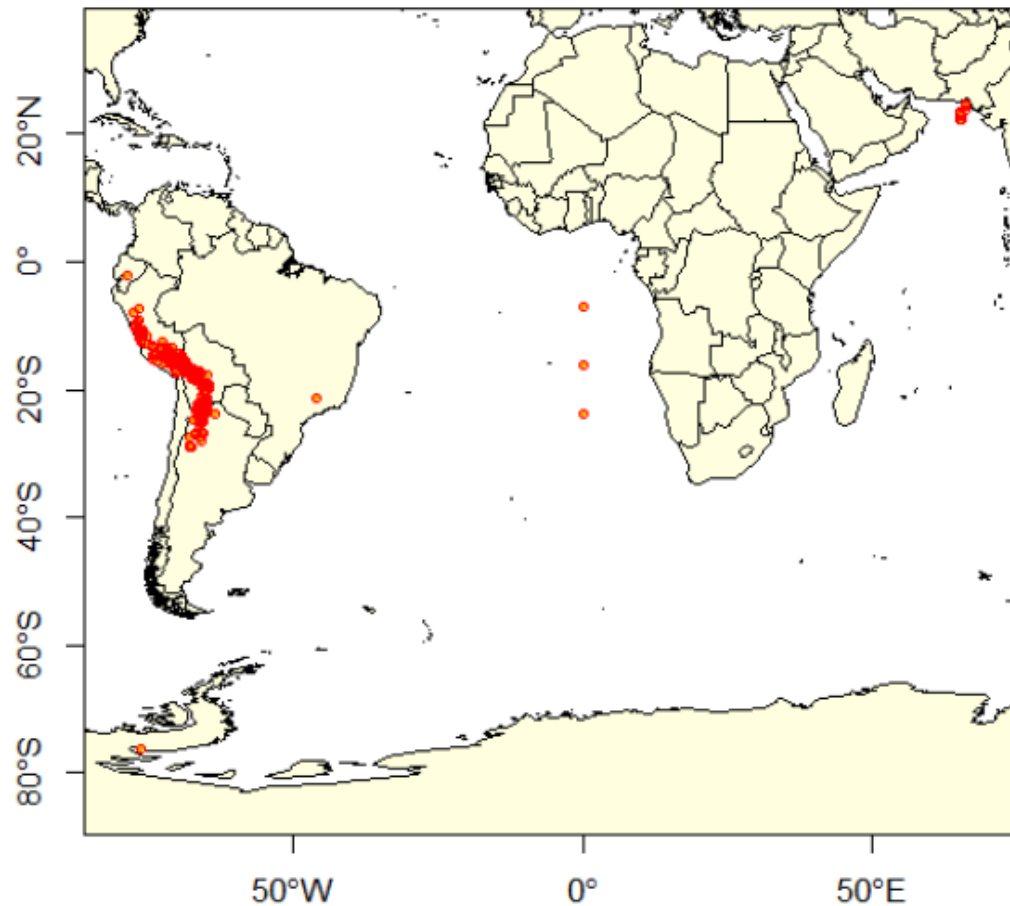
# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

## 1. Limpieza de datos (datos GBIF)

- Suprimir las observaciones con coordenadas incorrectas
- Suprimir las observaciones con una precisión baja en las coordenadas de observación (e.g., datos de principios de siglo XX)
- Suprimir observaciones en sitios artificiales (e.g., jardines botánicos)
- Selección taxonómica (e.g., subespecies) y en base al identificador específico

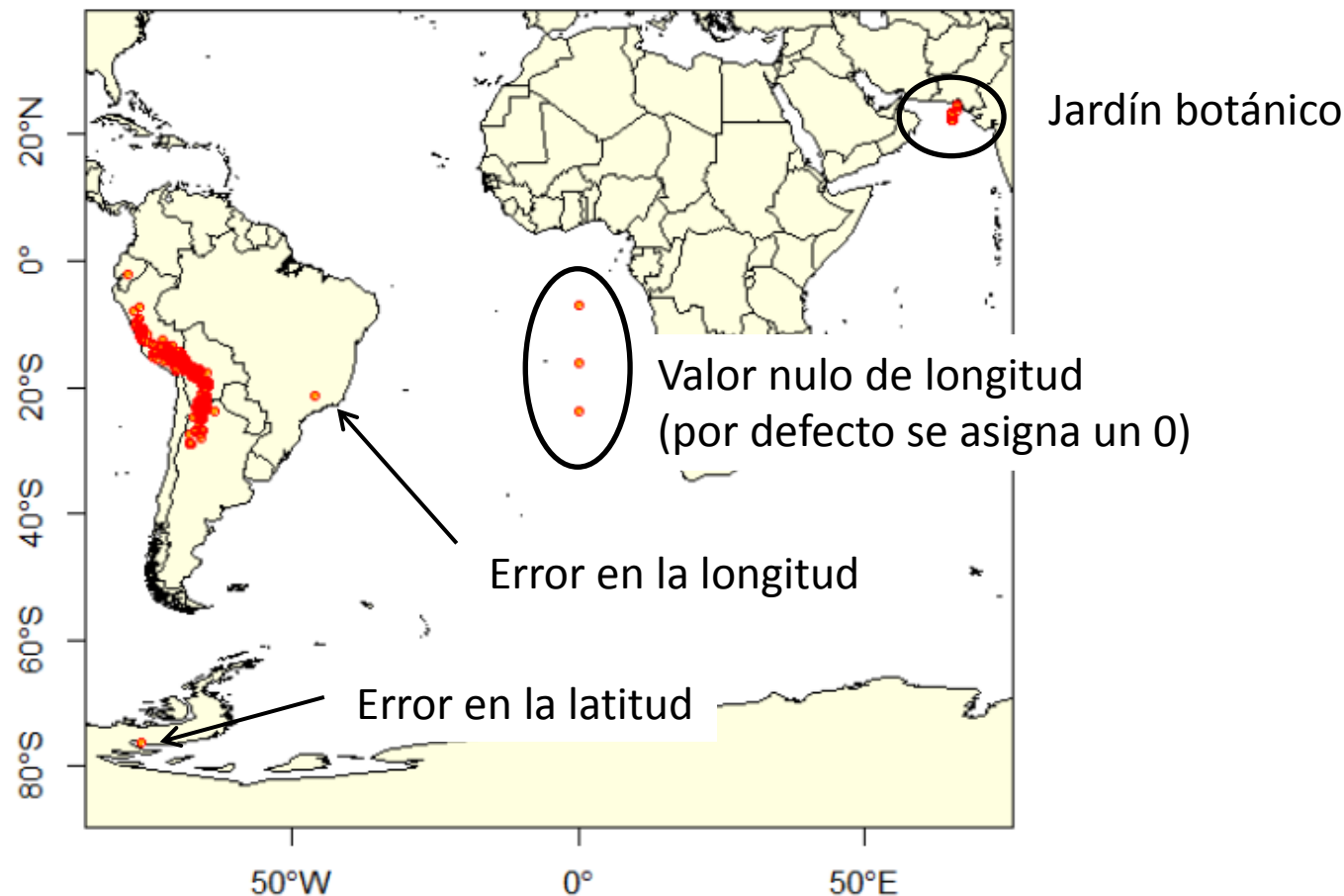
# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

## 1. Limpieza de datos (datos GBIF)



# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

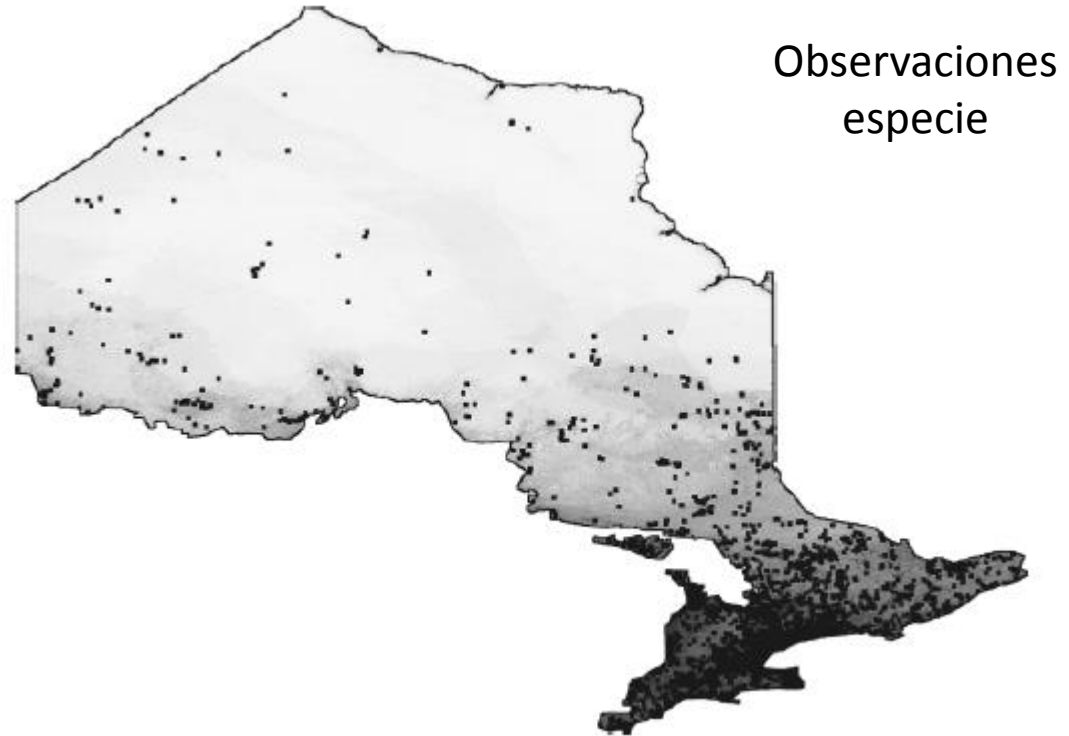
## 1. Limpieza de datos (datos GBIF)





# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

1. Limpieza de datos (datos GBIF)
2. Datos sesgados (bias)



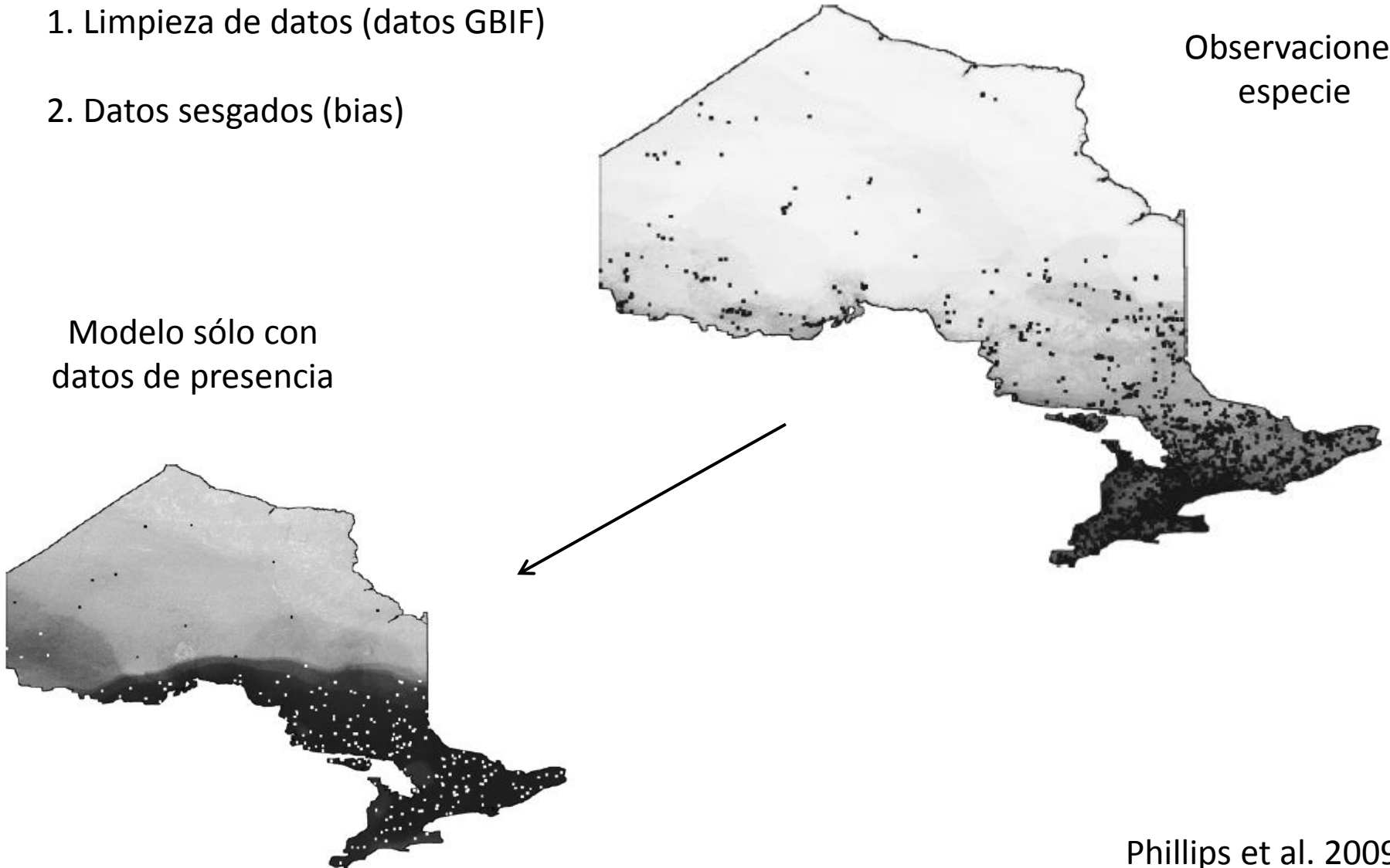
# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

1. Limpieza de datos (datos GBIF)

2. Datos sesgados (bias)

Observaciones  
especie

Modelo sólo con  
datos de presencia

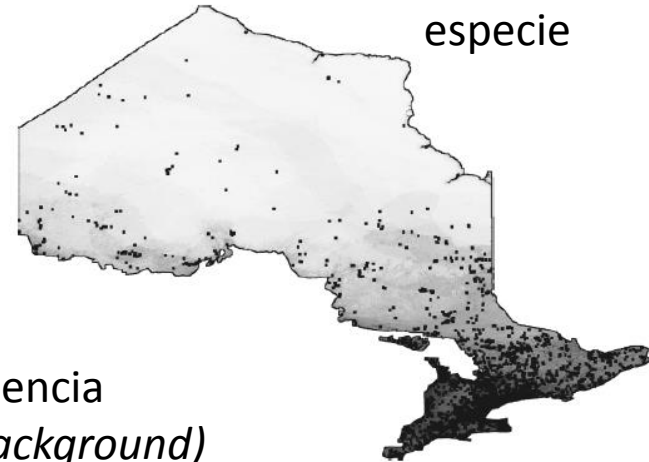


# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

1. Limpieza de datos (datos GBIF)

2. Datos sesgados (bias)

Observaciones  
especie



Modelo sólo con  
datos de presencia



Modelo con presencia  
y datos de “fondo” (*background*)



Phillips et al. 2009

# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

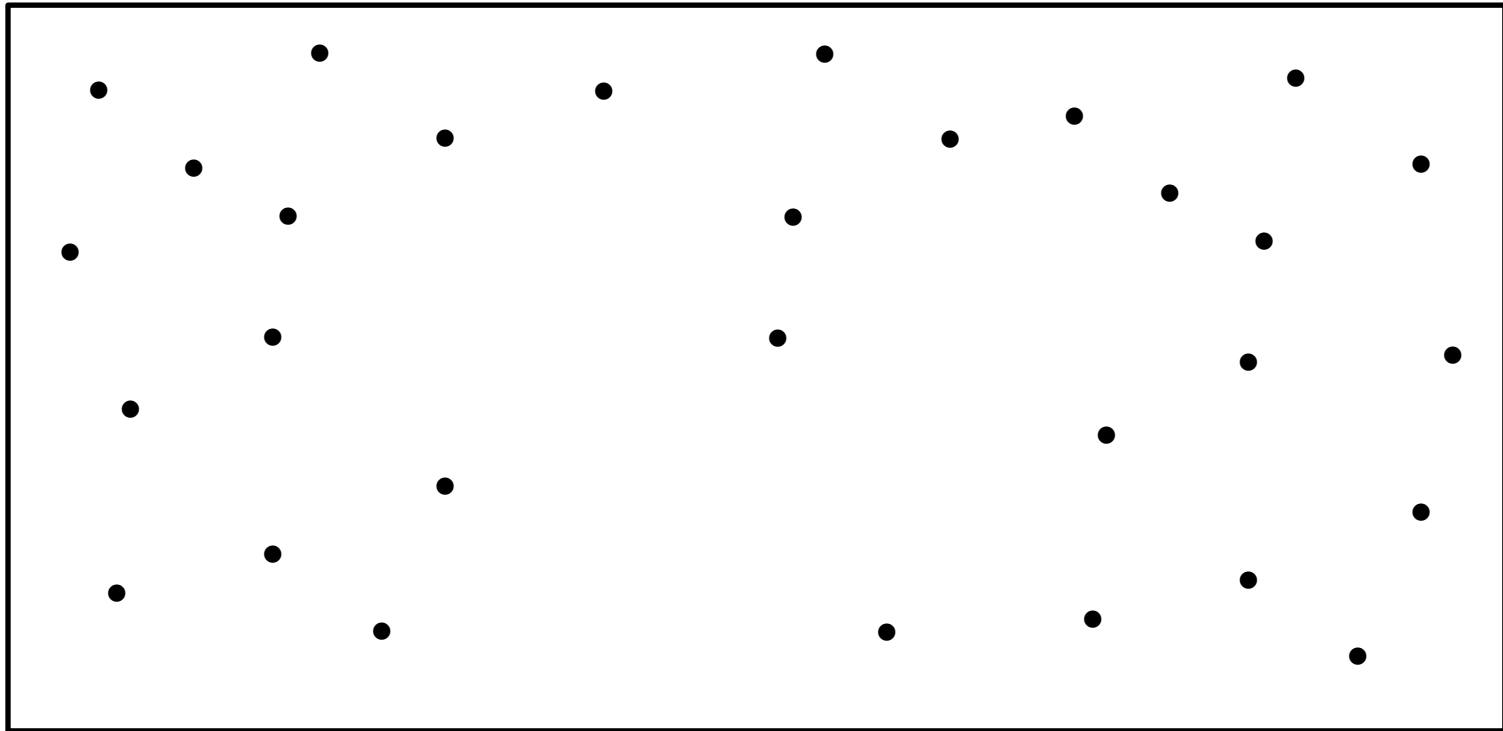
1. Limpieza de datos (datos GBIF)
2. Datos sesgados (bias)
3. Datos presencia y datos de “fondo” (*background*) o pseudo-ausencias

Los modelos en base a datos de presencia y datos *background* estiman la idoneidad de las especies en relación al rango de las variables ambientales utilizadas. La información del espacio ambiental potencialmente disponible para la especie proviene de los datos *background*, sin asunciones adicionales.

En el caso de las pseudo-ausencias, se asume que la especie realmente no está presente en los puntos asignados como pseudo-ausencias.

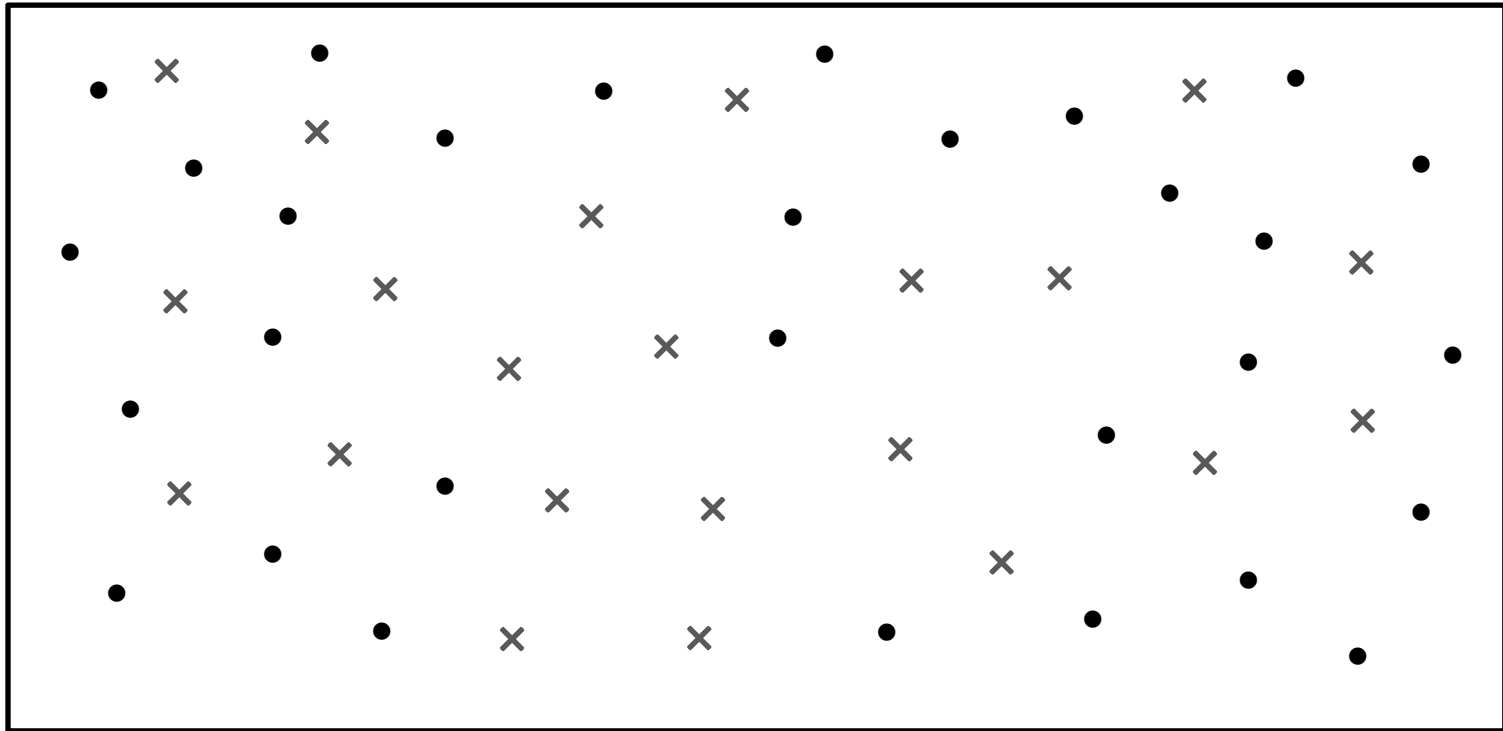
# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

- Observaciones de la especie



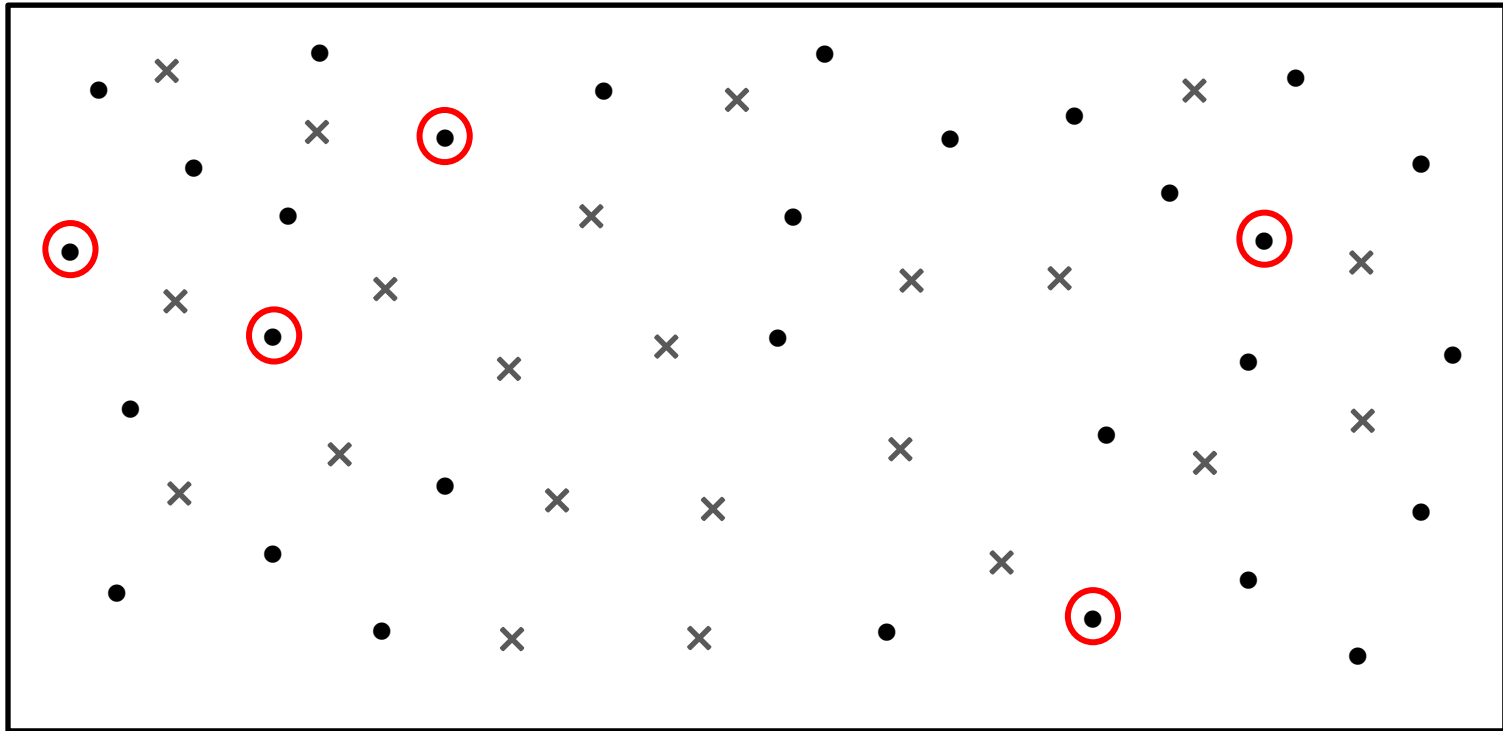
# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

- Observaciones de la especie
- × Datos *background* (o pseudo-ausencias)



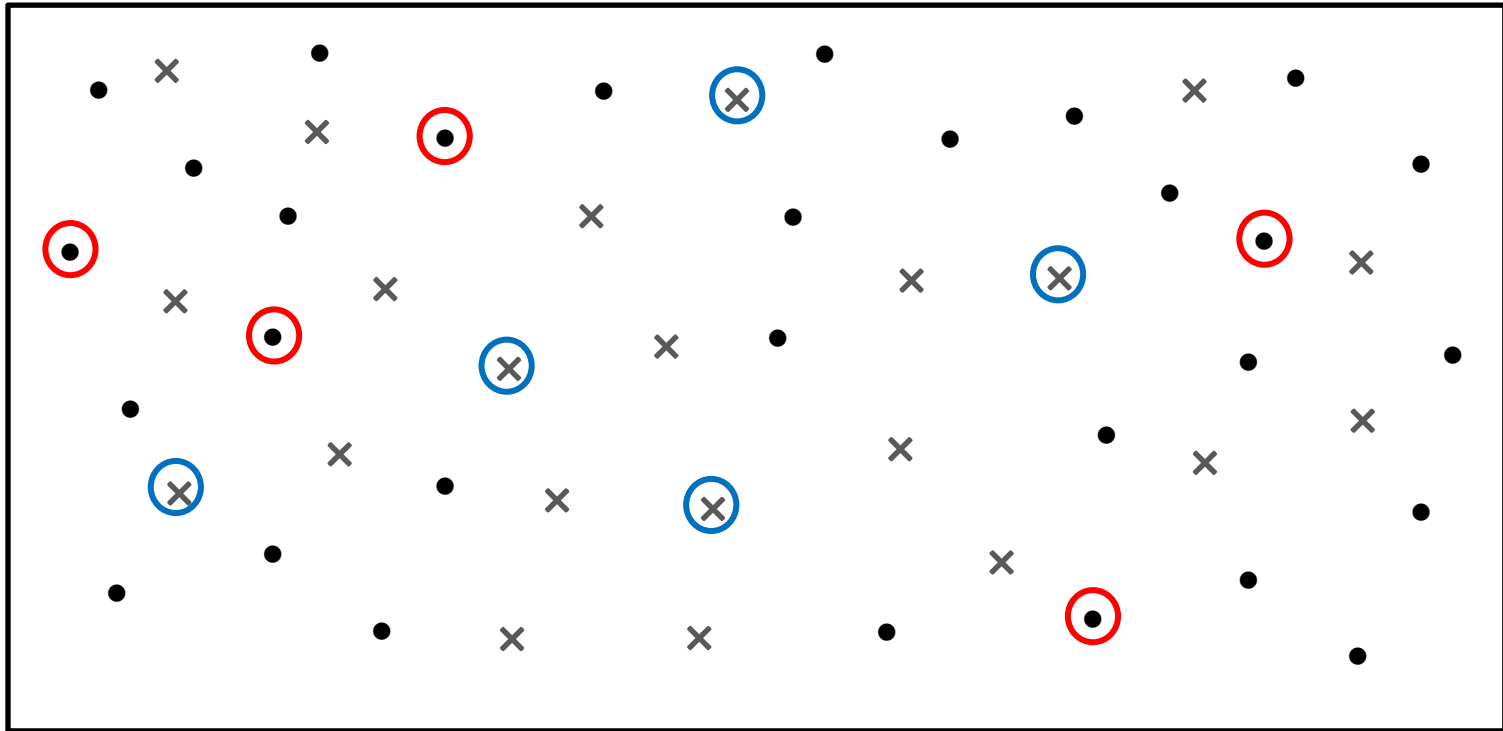
# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

- Observaciones de la especie ○ *Training points*
- × Datos *background* (o pseudo-ausencias)



# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

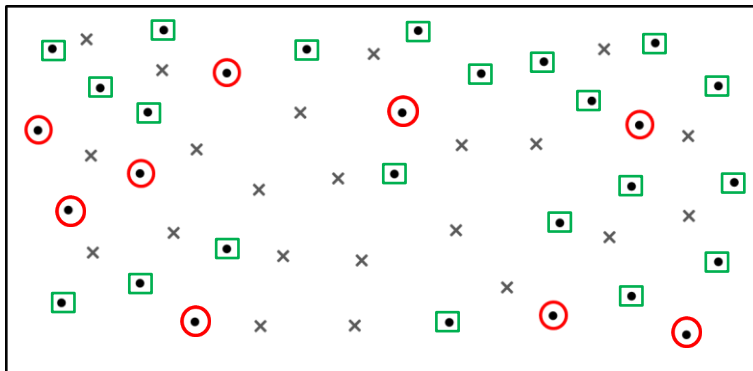
- Observaciones de la especie ○ *Training points (50%)*
- × Datos *background* (o pseudo-ausencias) ○ *Training points (50%)*





# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

## 1. Selección de una **fracción** de las observaciones

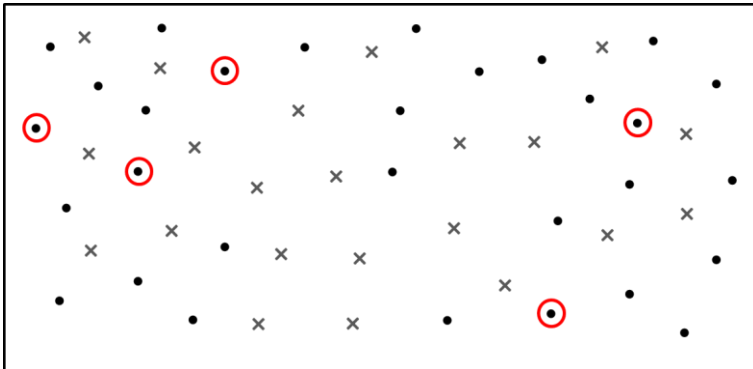


○ *Training points (30%)*

□ *Testing points (70%)*

# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

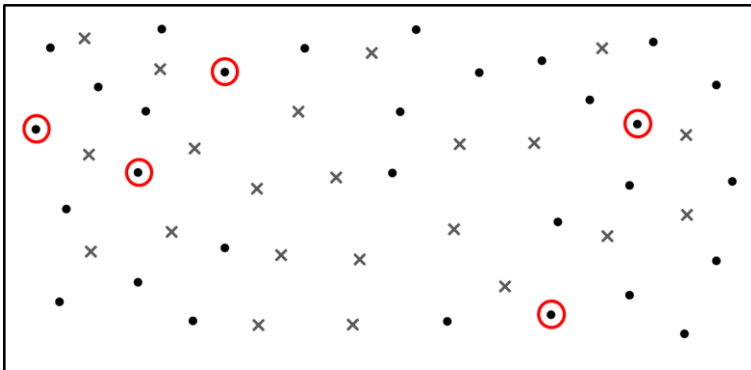
1. Selección de una **fracción** de las observaciones
2. Selección **aleatoria** estratificada



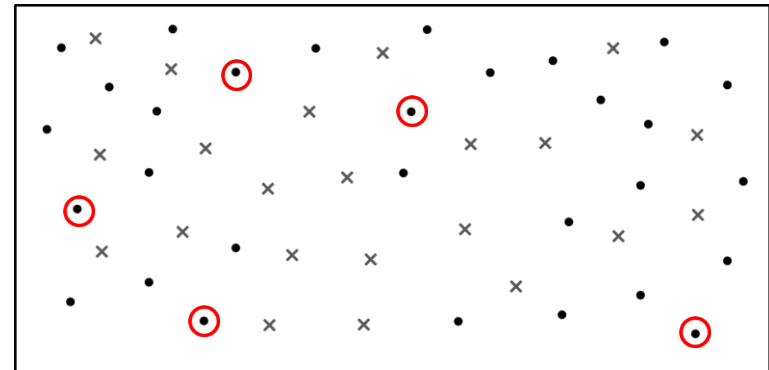
# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

1. Selección de una **fracción** de las observaciones
2. Selección **aleatoria** estratificada – algunas observaciones no se usan y otras se usan repetidamente

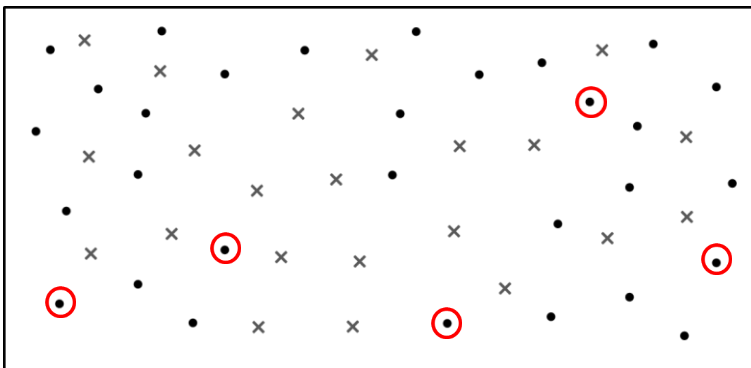
Replica 1



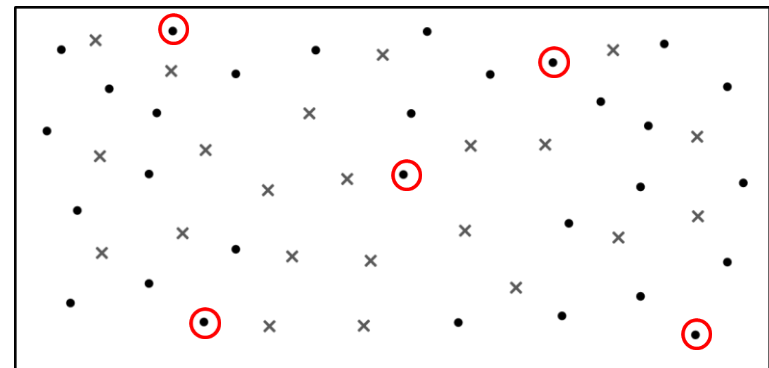
Replica 2



Replica 3

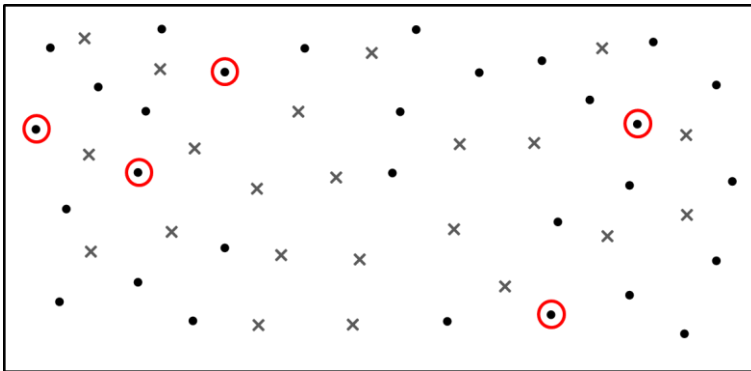


.... Replica N



# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

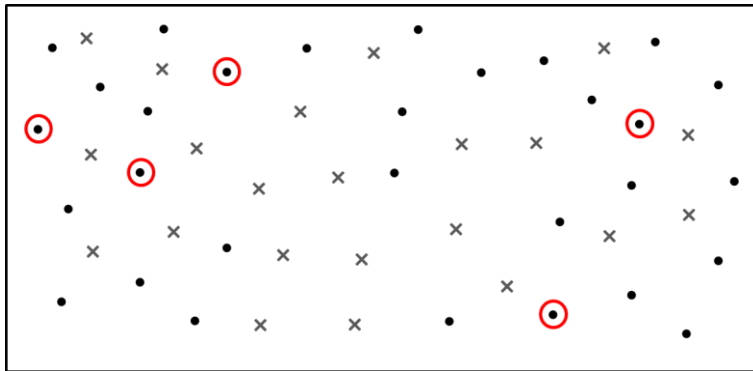
1. Selección de una **fracción** de las observaciones
2. Selección **aleatoria** estratificada – algunas observaciones no se usan y otras se usan repetidamente
3. Selección por ‘bloques’ (***K-fold cross validation***)



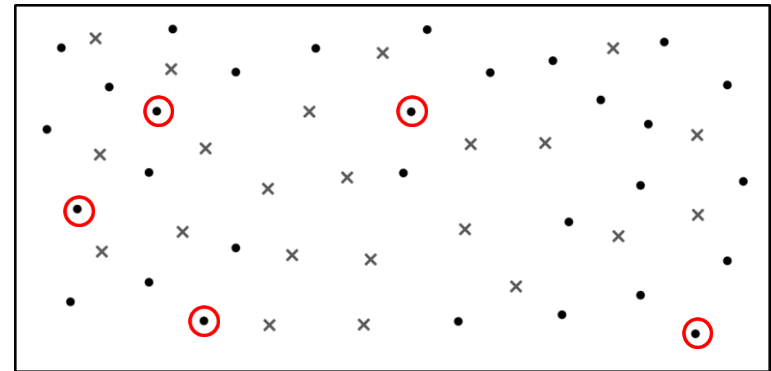
‘bloque’ 1

# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

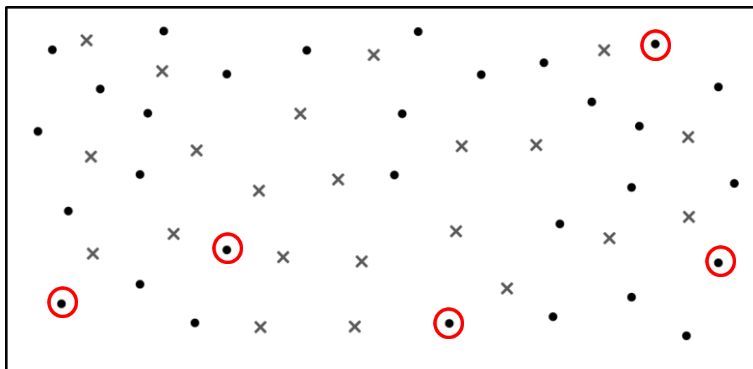
1. Selección de una **fracción** de las observaciones
2. Selección **aleatoria** estratificada – algunas observaciones no se usan y otras se usan repetidamente
3. Selección por ‘bloques’ (***K-fold cross validation***) – uso sistemático de todas las observaciones



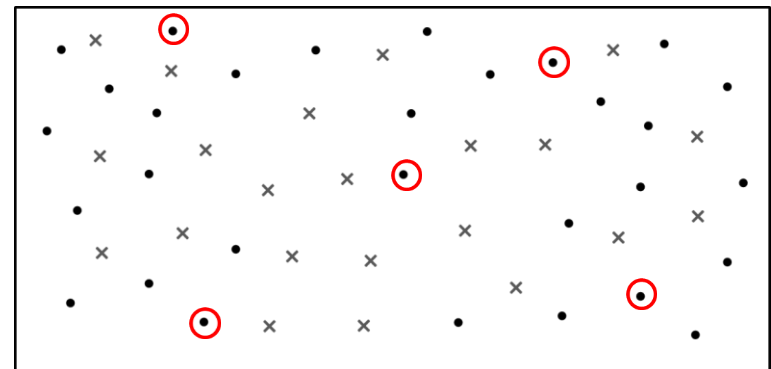
‘bloque’ 1



‘bloque’ 2



‘bloque’ 3



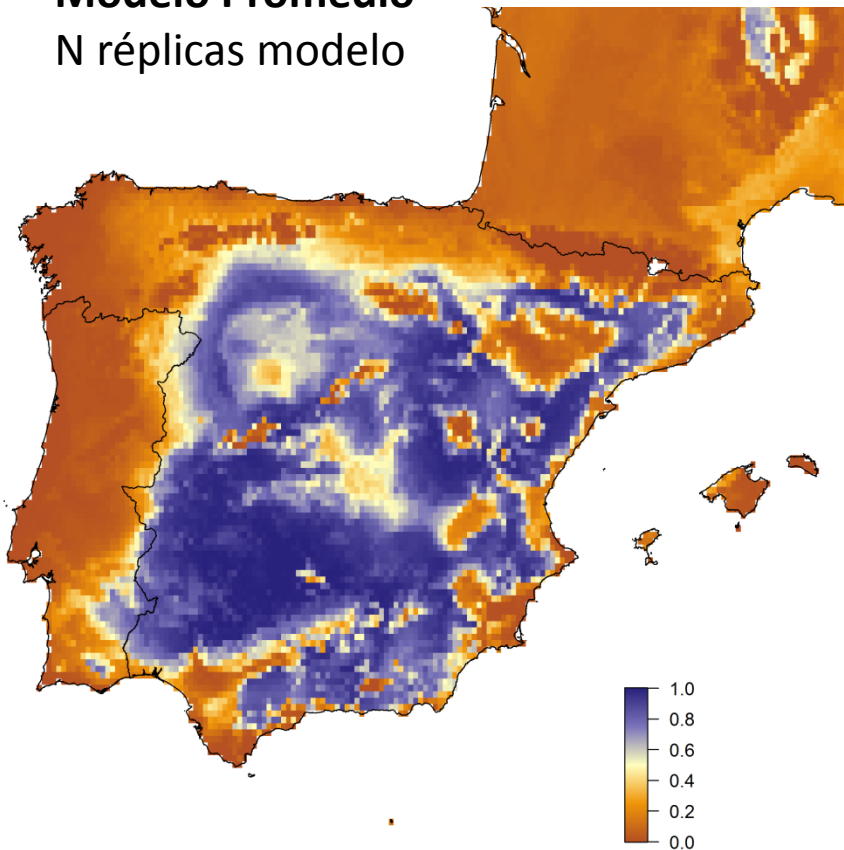
.... ‘bloque’ N

# “Limpieza” de datos y selección de observaciones para desarrollar y validar los SDM (*training and testing points*)

1. Selección de una **fracción** de las observaciones
2. Selección **aleatoria** estratificada – algunas observaciones no se usa y otras se usan repetidamente
3. Selección por ‘bloques’ (***K-fold cross validation***) – uso sistemático de todas las observaciones

## Modelo Promedio

N réplicas modelo



## Desviación estándar

N réplicas modelo

