

Modelos lineales y aditivos en ecología

Facundo X. Palacio

facundo_palacio@fcnym.unlp.edu.ar

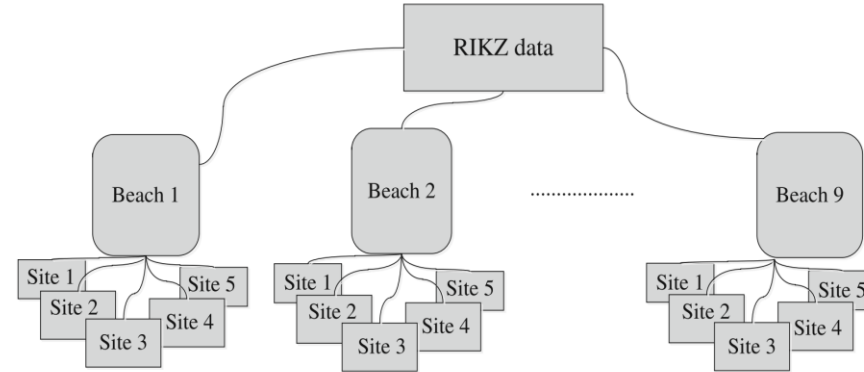


2 al 6 de mayo de 2022 – Universidad Nacional de Tucumán

Un poco de historia

- Fisher (1918) → Modelo de efectos aleatorios
- Eisenhart (1947) → Dos tipos de factores: fijos y aleatorios
- Henderson et al. (1959) → Estimaciones de efectos aleatorios
- Lindstrom y Bates (1988) → Estimación medidas repetidas
- McLean et al. (1991) → “*A unified approach to mixed linear models*”

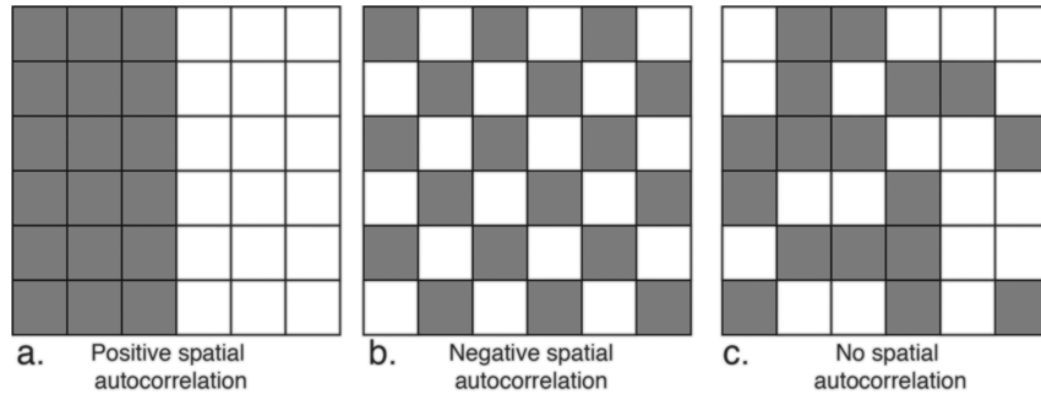
No independencia



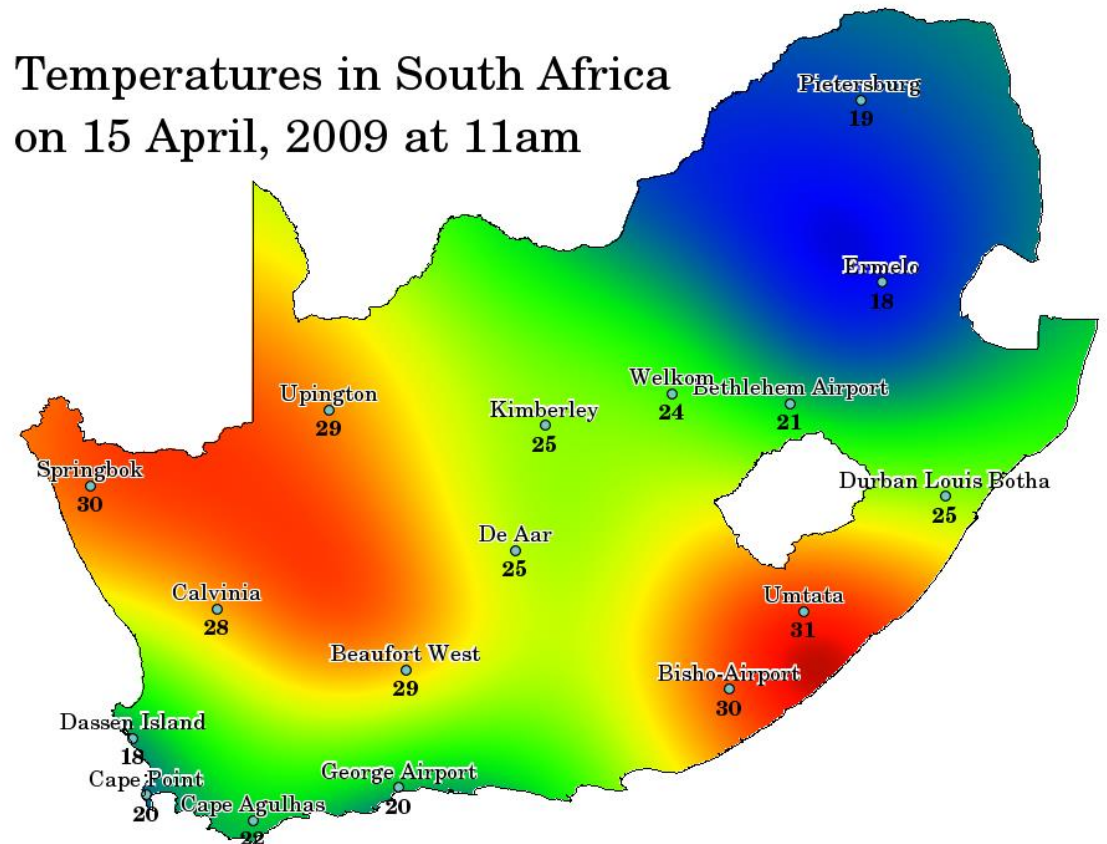
¿Cuándo aparece?

1. Estudios con estructura jerárquica
2. Estudios con diferentes escalas espaciales
3. Estudios con diferentes escalas temporales

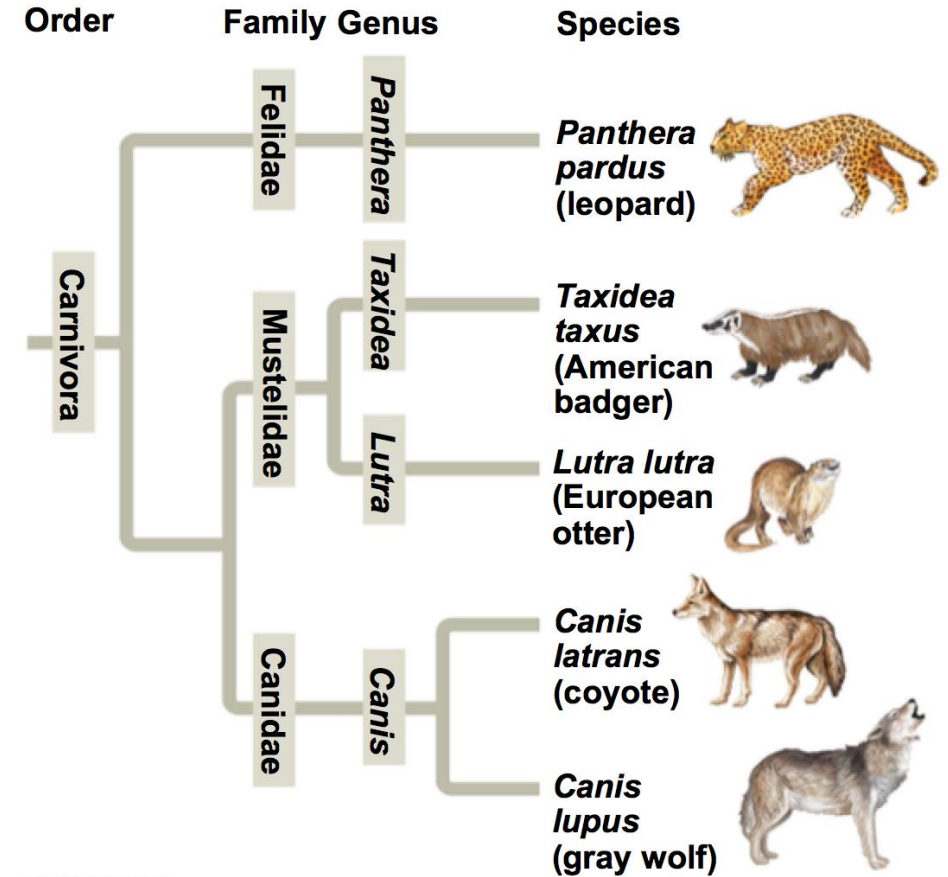
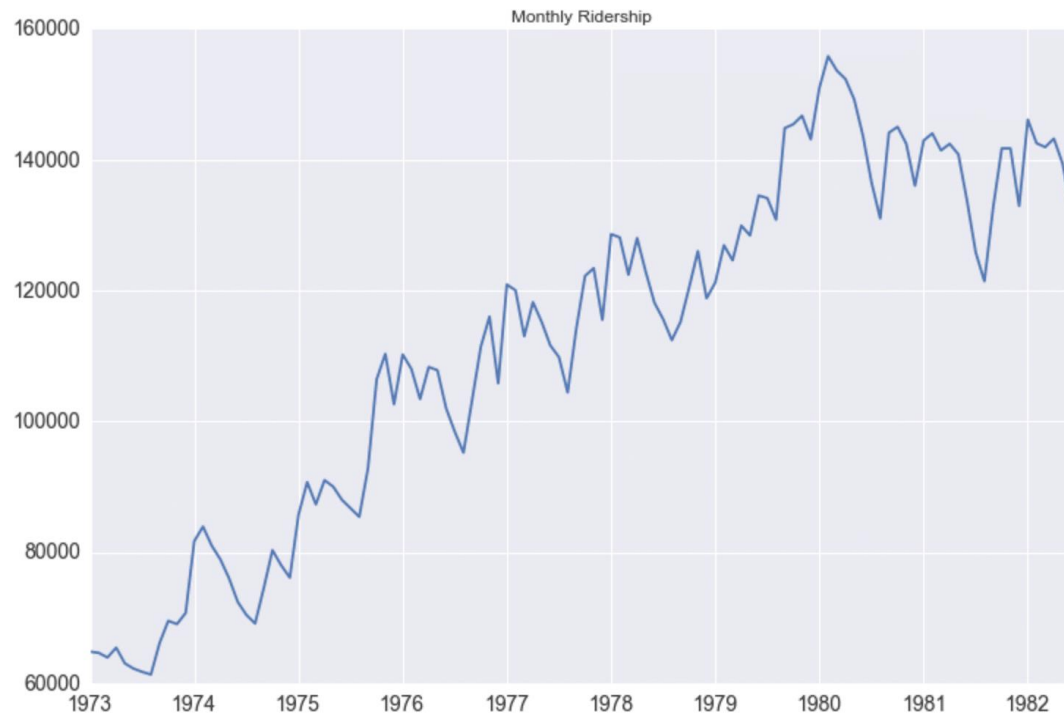
Dependencia espacial



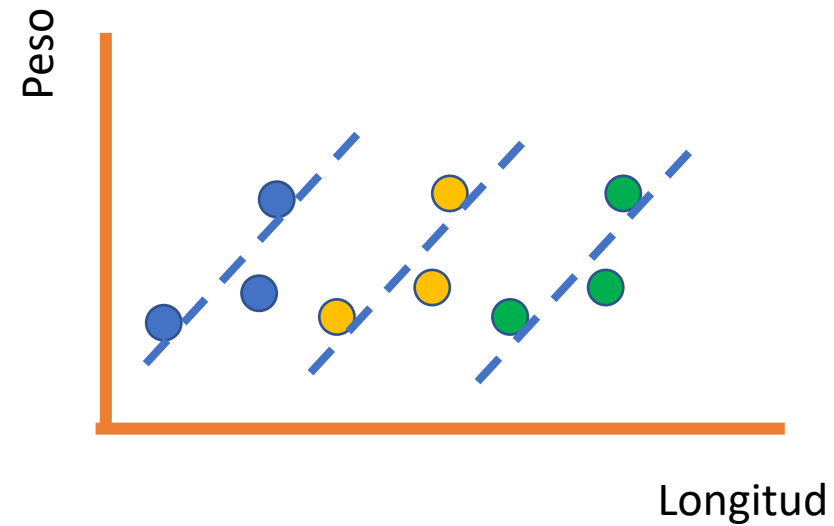
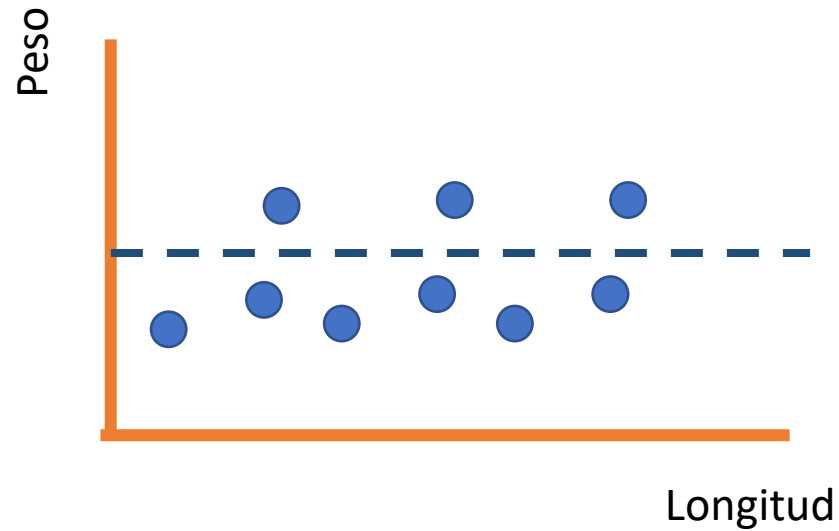
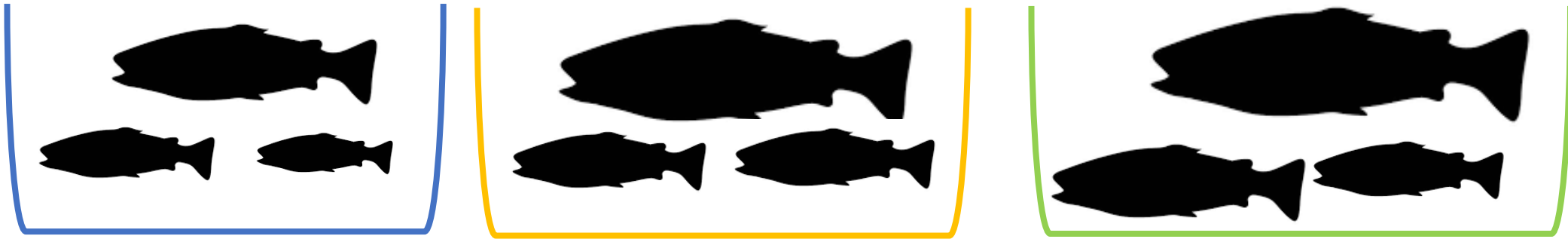
Temperatures in South Africa
on 15 April, 2009 at 11am



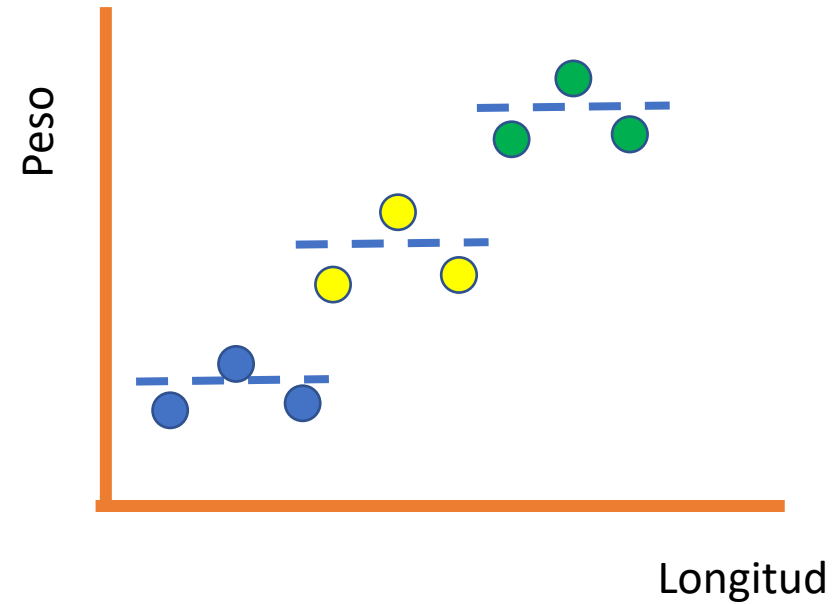
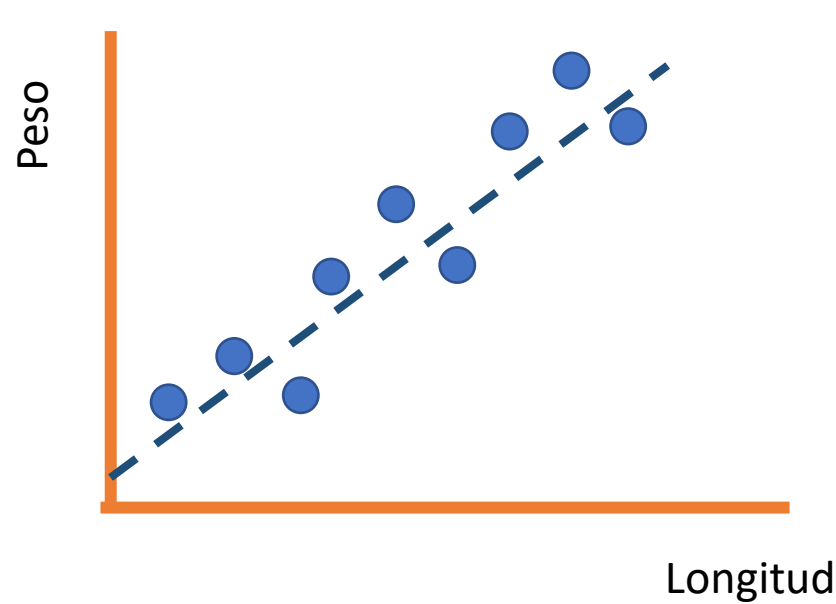
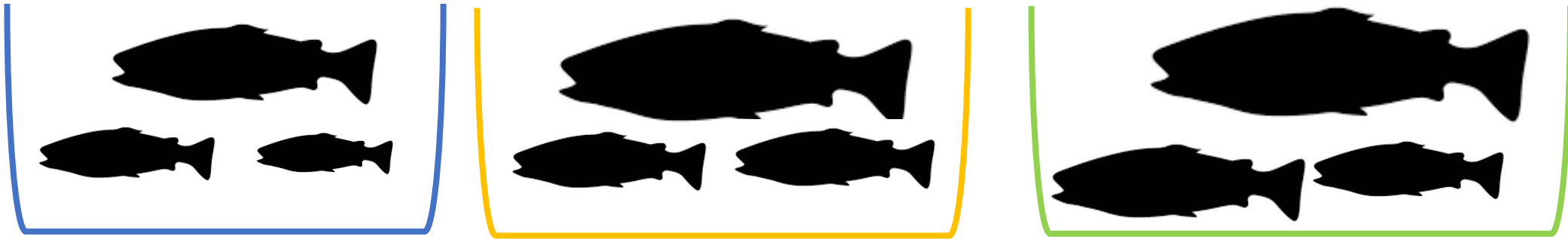
Dependencia temporal



Dependencia jerárquica



Dependencia jerárquica



Modelo de intercepto aleatorio

$$\hat{y} = 1.19 + 1.73x$$

Componente fijo

$$\hat{y} = b_0 + b_1 \times \text{longitud pedicelo} + a_i$$

Componente fijo Componente aleatorio

$$y_5 = 1.19 + 1.73x + 1.35$$

$$Si^2 = 1.05$$

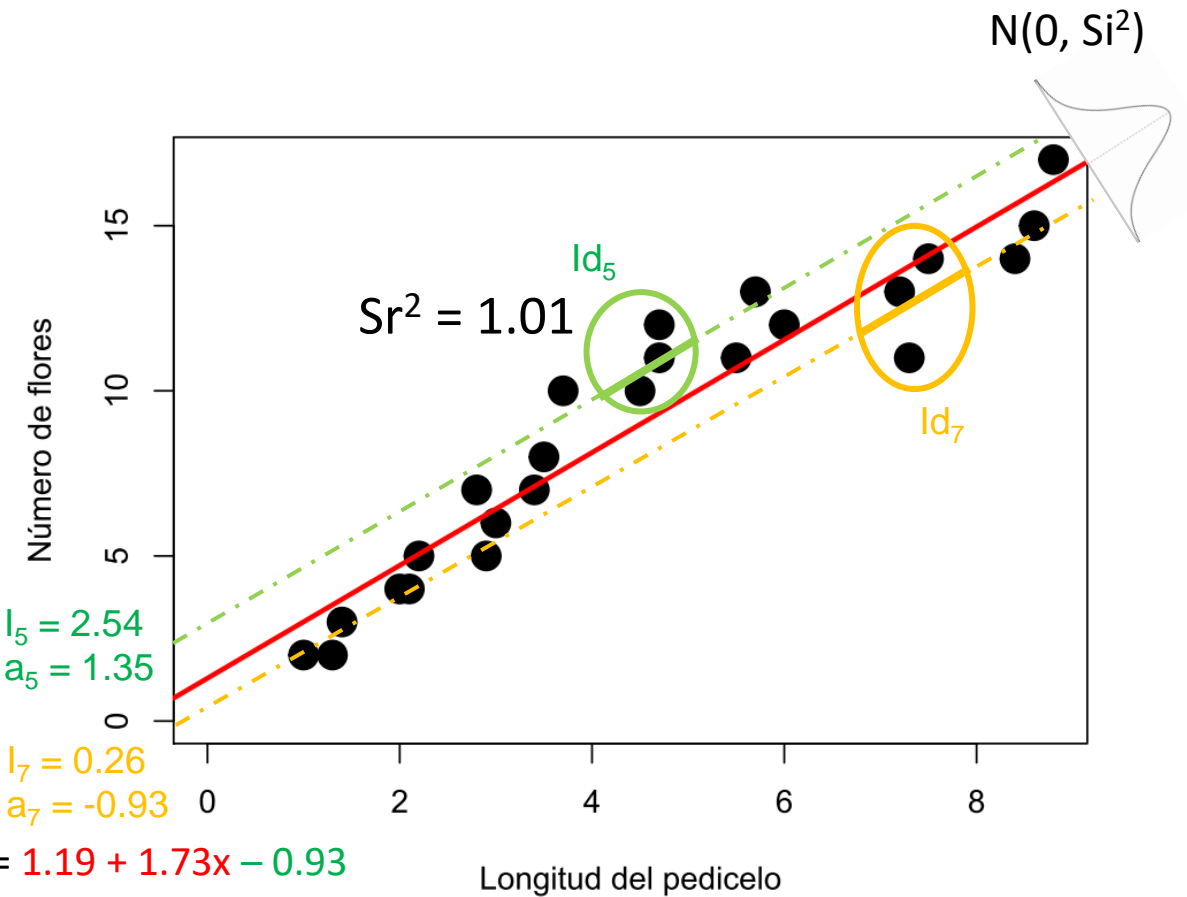
$$I_5 = 2.54$$

$$a_5 = 1.35$$

$$I_7 = 0.26$$

$$a_7 = -0.93$$

$$y_7 = 1.19 + 1.73x - 0.93$$



Modelo de intercepto y pendiente aleatorios

$$\hat{y} = 0.04 + 2.08x + a_i + z_i x$$

Componente fijo

Componente aleatorio

$$\hat{y} = b_0 + b_1 \times \text{longitud pedicelo} + a_i + z_i \times \text{longitud pedicelo}$$

$$y_5 = 0.04 + 2.08x - 0.86 + 0.41x$$

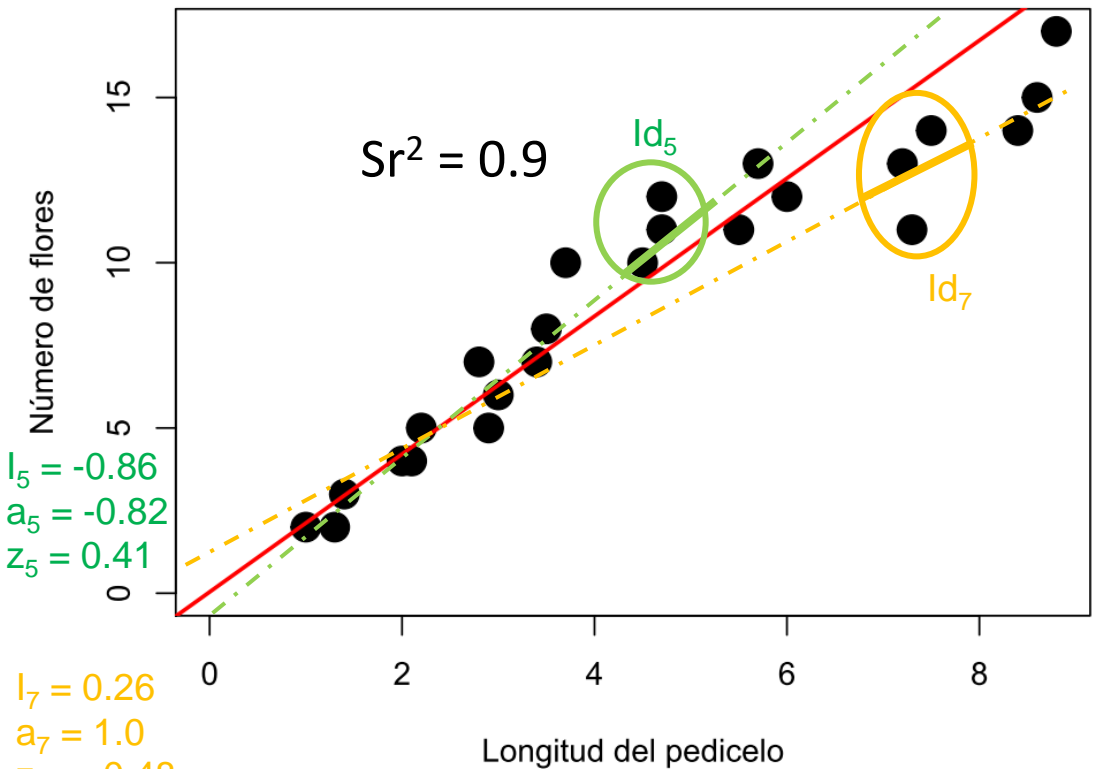
$$S_i^2 = 0.82$$

$$S_z^2 = 0.19$$

$$\begin{aligned} l_5 &= -0.86 \\ a_5 &= -0.82 \\ z_5 &= 0.41 \end{aligned}$$

$$\begin{aligned} l_7 &= 0.26 \\ a_7 &= 1.0 \\ z_7 &= -0.48 \end{aligned}$$

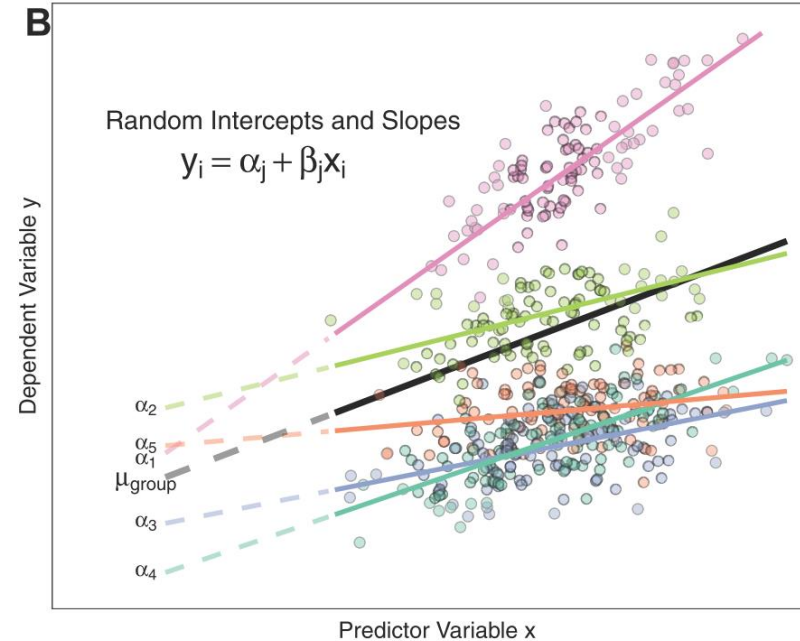
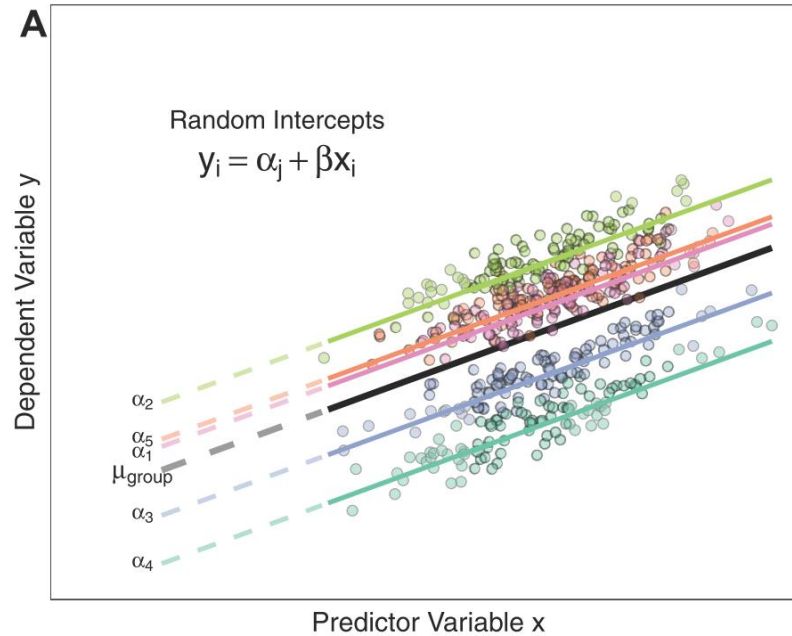
$$y_7 = 0.04 + 2.08x + 1.0 - 0.48x$$





A brief introduction to mixed effects modelling and multi-model inference in ecology

Xavier A. Harrison¹, Lynda Donaldson^{2,3}, Maria Eugenia Correa-Cano², Julian Evans^{4,5}, David N. Fisher^{4,6}, Cecily E.D. Goodwin², Beth S. Robinson^{2,7}, David J. Hodgson⁴ and Richard Inger^{2,4}



Efectos fijos vs aleatorios

- **Efectos fijos**
 - Entidad fija que se quiere estimar
 - Determinado por el investigador
 - Influencia la media de y
- **Efectos aleatorios**
 - Variable aleatoria
 - Factor que agrupa observaciones
 - Influencia la varianza de Y .
 - Proviene de una gran población

Efectos fijos vs aleatorios

Reglas...



Recomendaciones (Harrison et al. 2018)

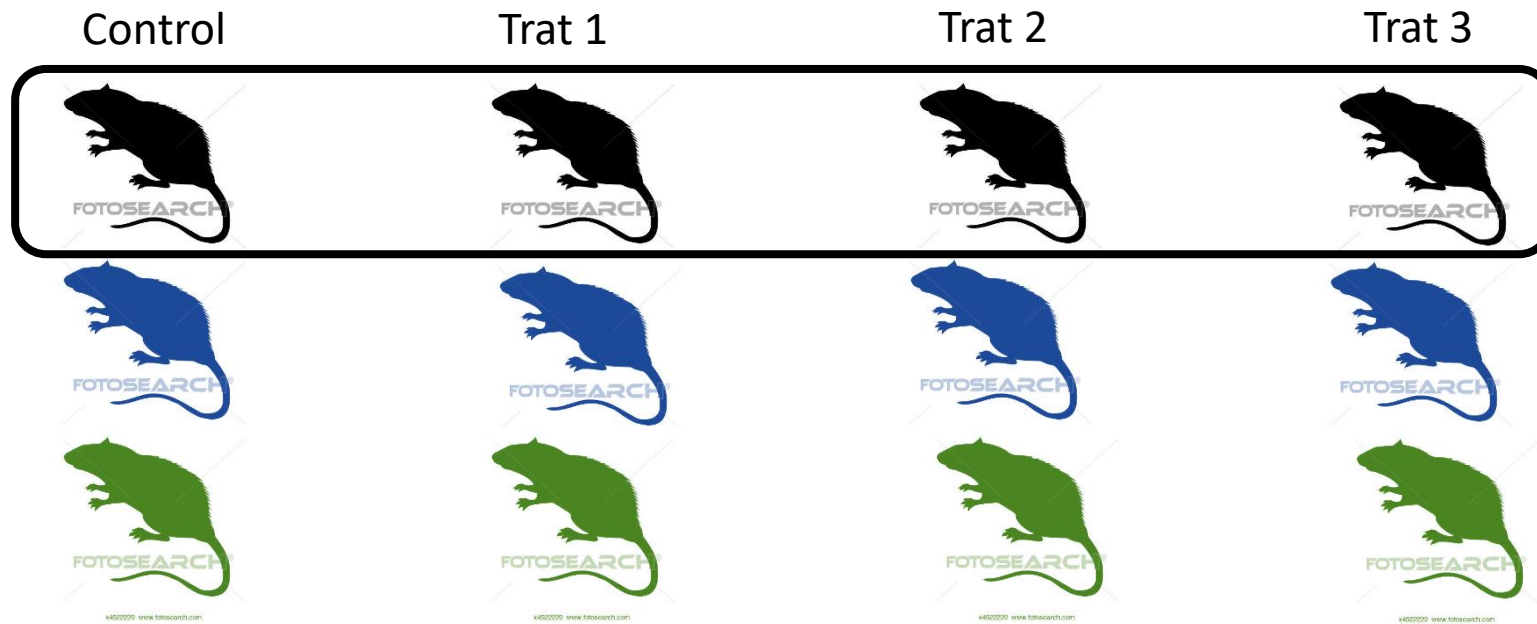
- ✓ Las variables continuas NO pueden considerarse efectos aleatorios.
- ✓ NO considerar un factor con < 5 niveles, varianzas imprecisas.

Efectos fijos vs aleatorios

Pregunta	Fijo	Aleatorio
¿Los niveles del factor son informativos? ¿Queremos saber por qué difieren? ¿Queremos estimar la magnitud de estos efectos?	Sí e.j. macho, hembra	No e.j. categorías A, B, C e.j. categorías con etiquetas numéricas
¿El factor representa: 1. Autocorrelación espacial 2. Autocorrelación temporal 3. Medidas repetidas 4. Una estructura jerárquica/anidada? 5. Historia evolutiva	No	Sí 1. e.j. ubicación a lo largo de una transecta. 2. e.j. tiempo 3. e.j. medidas repetidas de un mismo individuo 4. estructura jerárquica del diseño, e.g. huevos en una serie de nidos
¿Los niveles del factor son una muestra aleatoria de una población mucho más grande?	No e.j. macho, hembra e.j. muerto, vivo	Sí e.j. campo 1, campo 2, campo 3 son una muestra aleatoria de todos los campos e.j. familia 1, familia 2, familia 3 son una muestra aleatoria de todas las familias
¿Hay suficientes niveles del factor en los datos, como para estimar una varianza?	No < 5	Sí ≥ 5
¿El factor influye sólo la media de Y o también la varianza de Y?	Sólo media	Sólo varianza

Ejemplos de diseños con efectos aleatorios

Medidas repetidas o diseños longitudinales



Ejemplos de diseños con efectos aleatorios

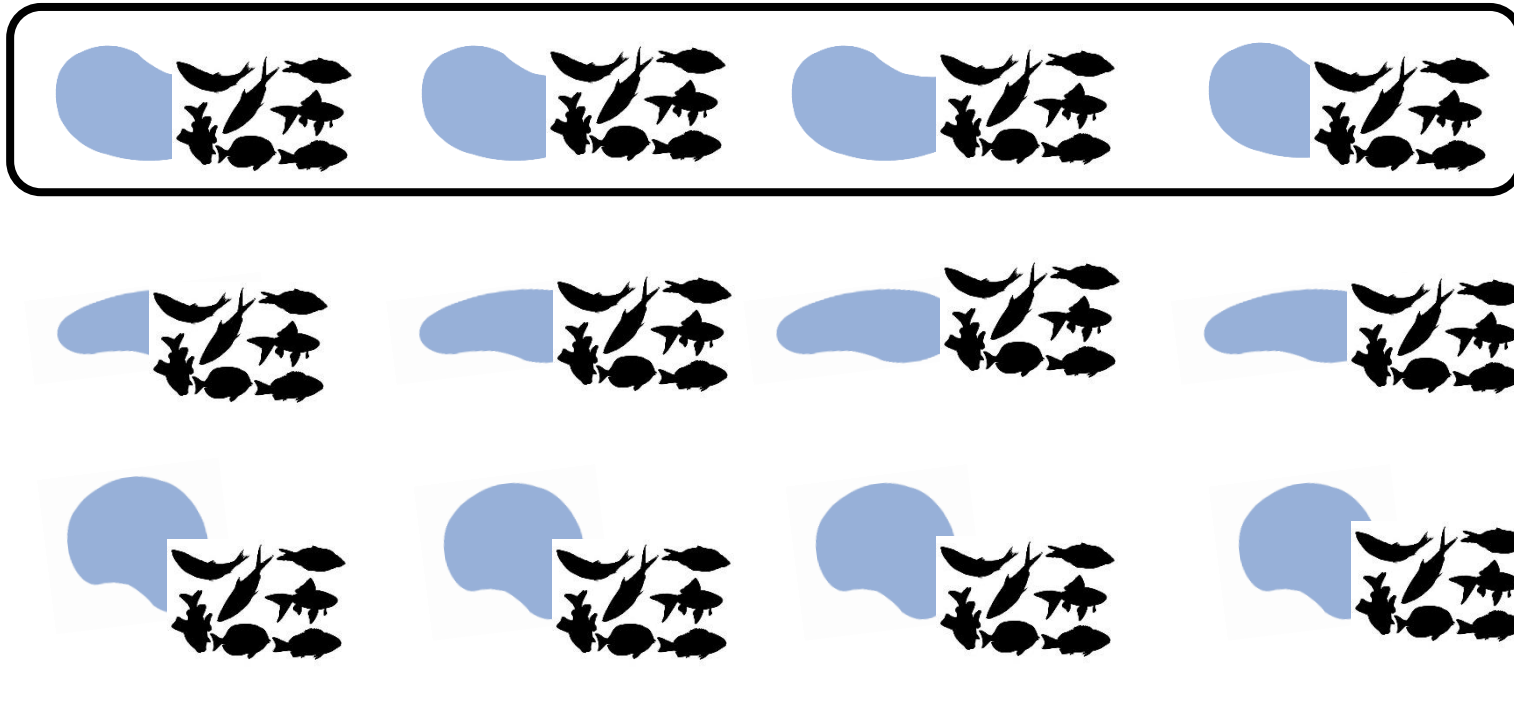
Medidas repetidas o diseños longitudinales

Otoño

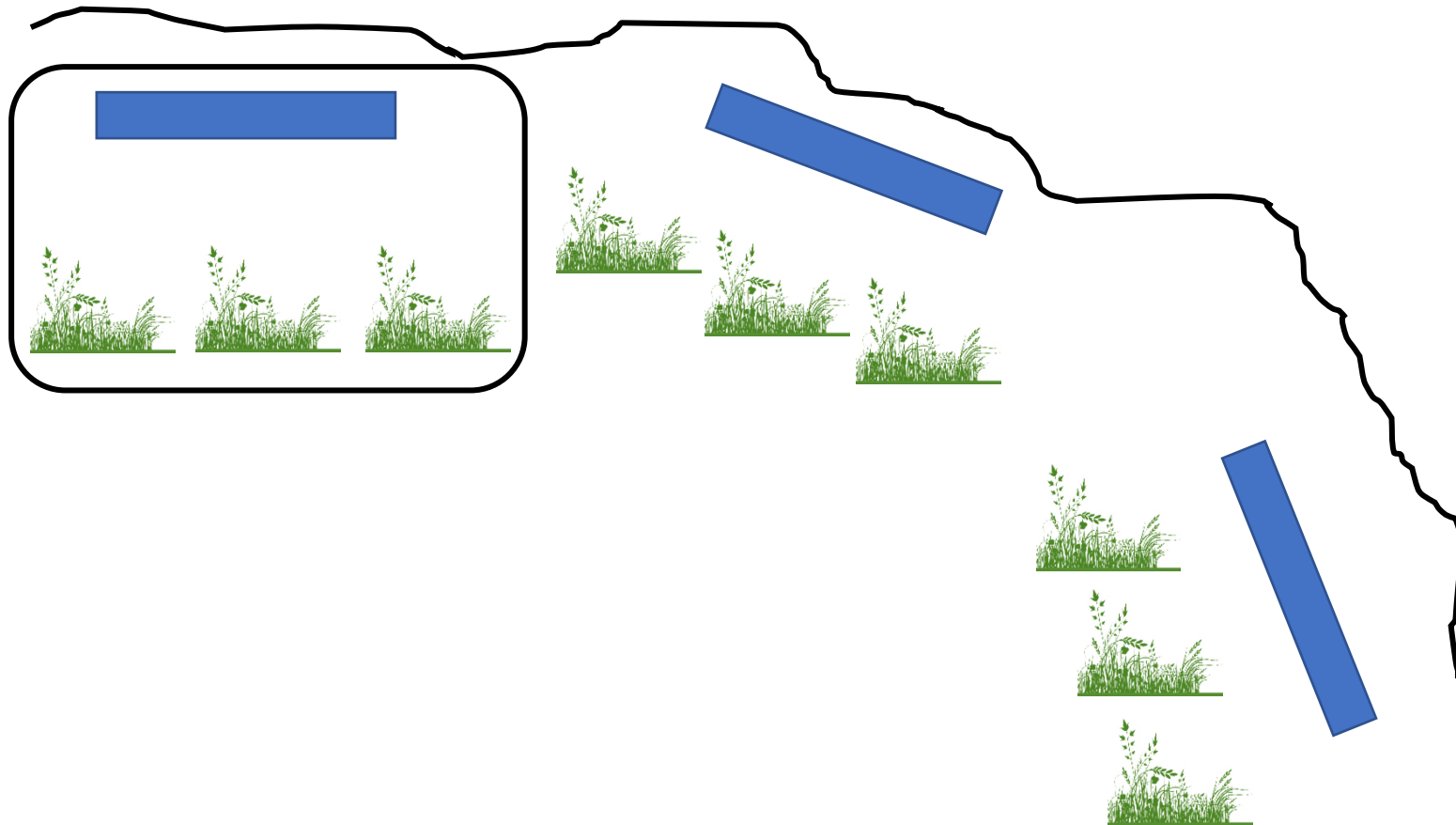
Invierno

Primavera

Verano



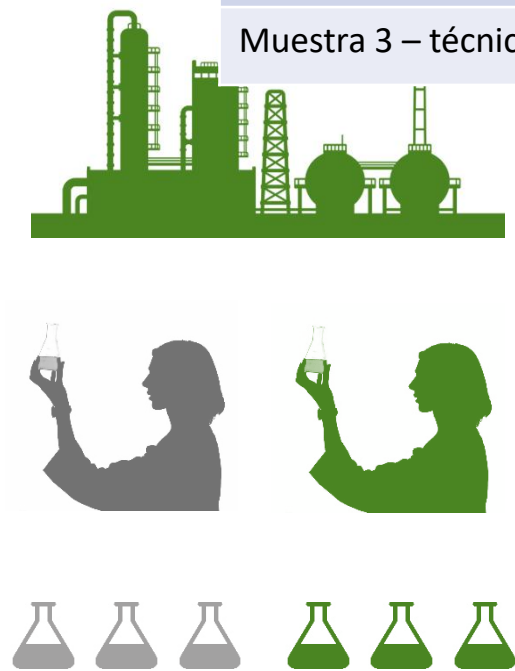
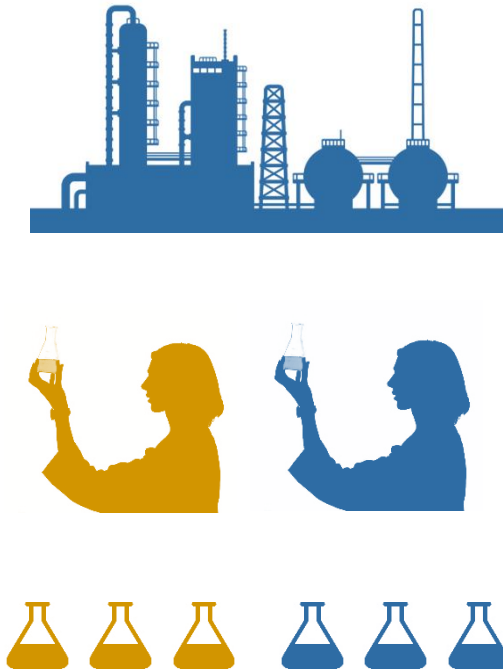
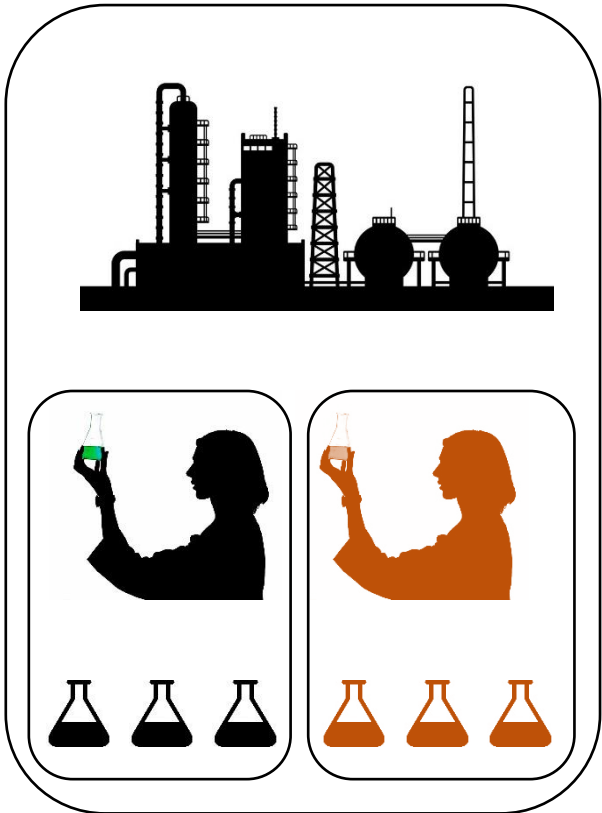
Ejemplos de diseños con efectos aleatorios



Ejemplos de diseños con efectos aleatorios

Factores anidados

Cada nivel del primer factor está presente en un solo nivel del segundo factor



Combinación de niveles	N
Muestra 1 – técnica 1	1
Muestra 2 – técnica 1	1
Muestra 3 – técnica 1	1
Muestra 1 – técnica 2	0
Muestra 2 – técnica 2	0
Muestra 3 – técnica 2	0

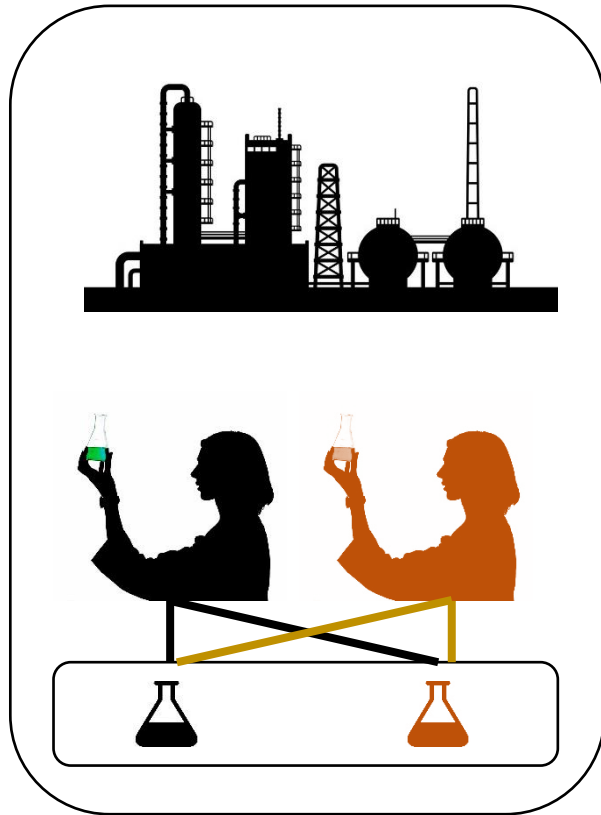
Ejemplos de diseños con efectos aleatorios

Factores cruzados

Al menos una observación por cada combinación de niveles

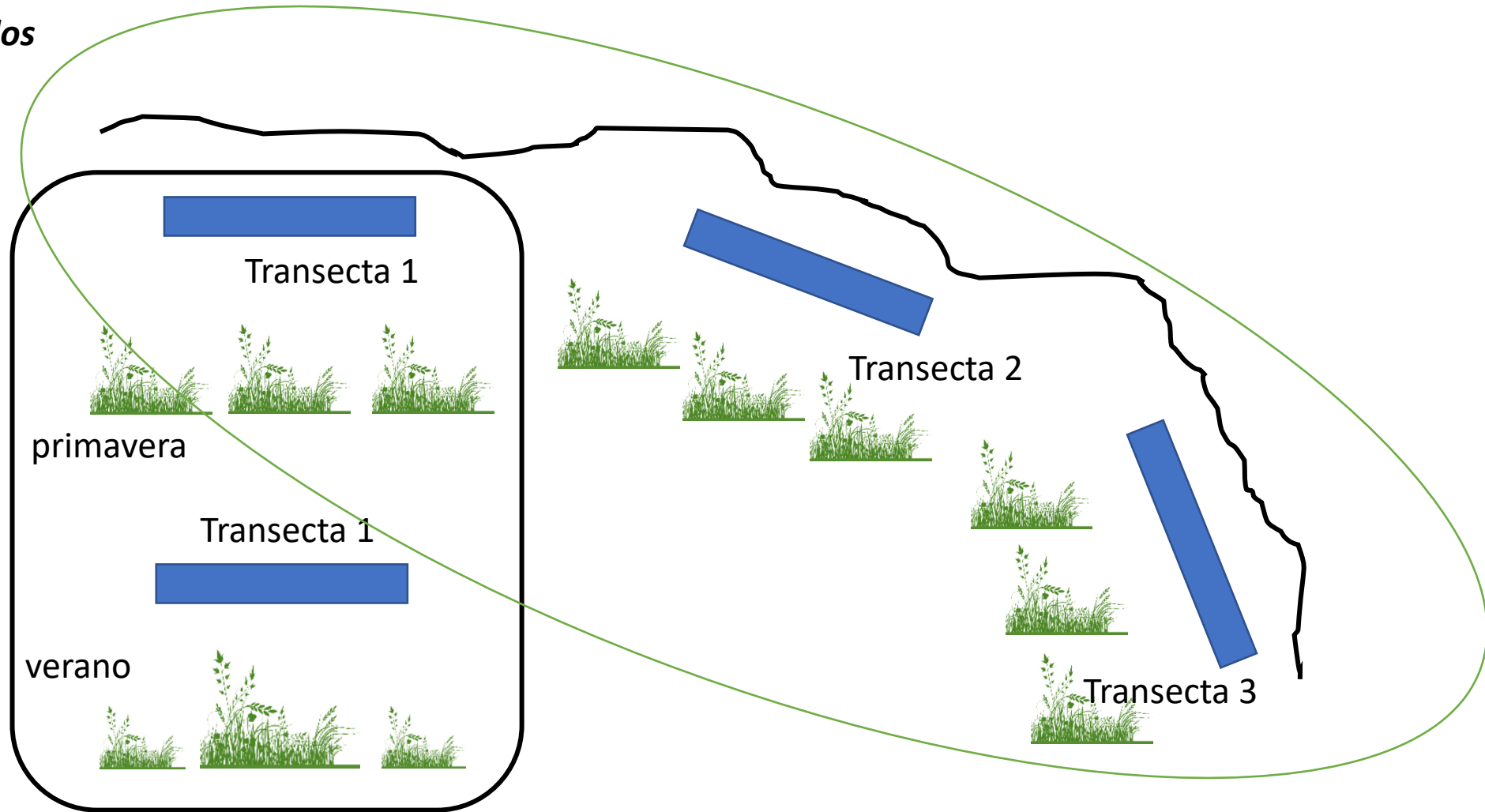
(= cada nivel del primer factor se combina con cada nivel del segundo factor)

Combinación de niveles	N
Muestra 1 – técnica 1	1
Muestra 2 – técnica 1	1
Muestra 1 – técnica 2	1
Muestra 2 – técnica 2	1



Ejemplos de diseños con efectos aleatorios

Factores cruzados



Dependencia temporal

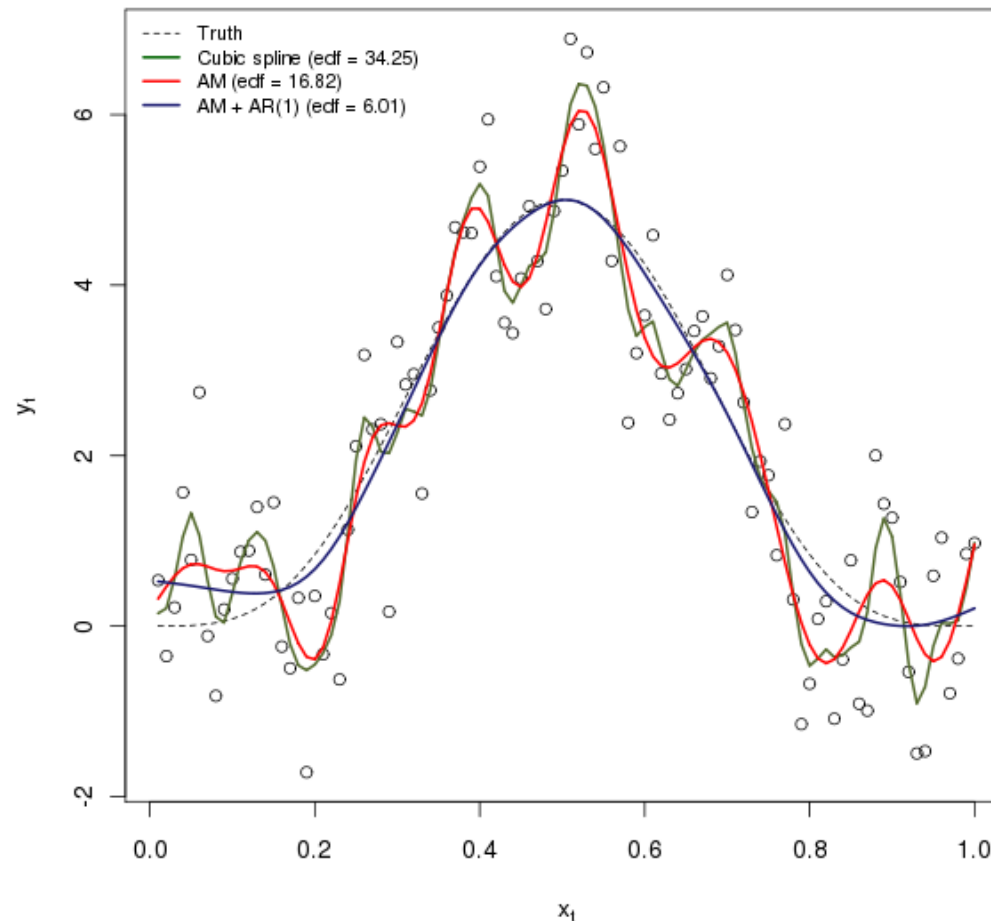
$$r(e_i, e_j) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{de otra forma} \end{cases}$$

$$r(e_i, e_j) = \begin{cases} 1 & \text{si } i = j \\ \rho & \text{de otra forma} \end{cases}$$

$$e_i = \rho e_{i-1} + \eta_i$$

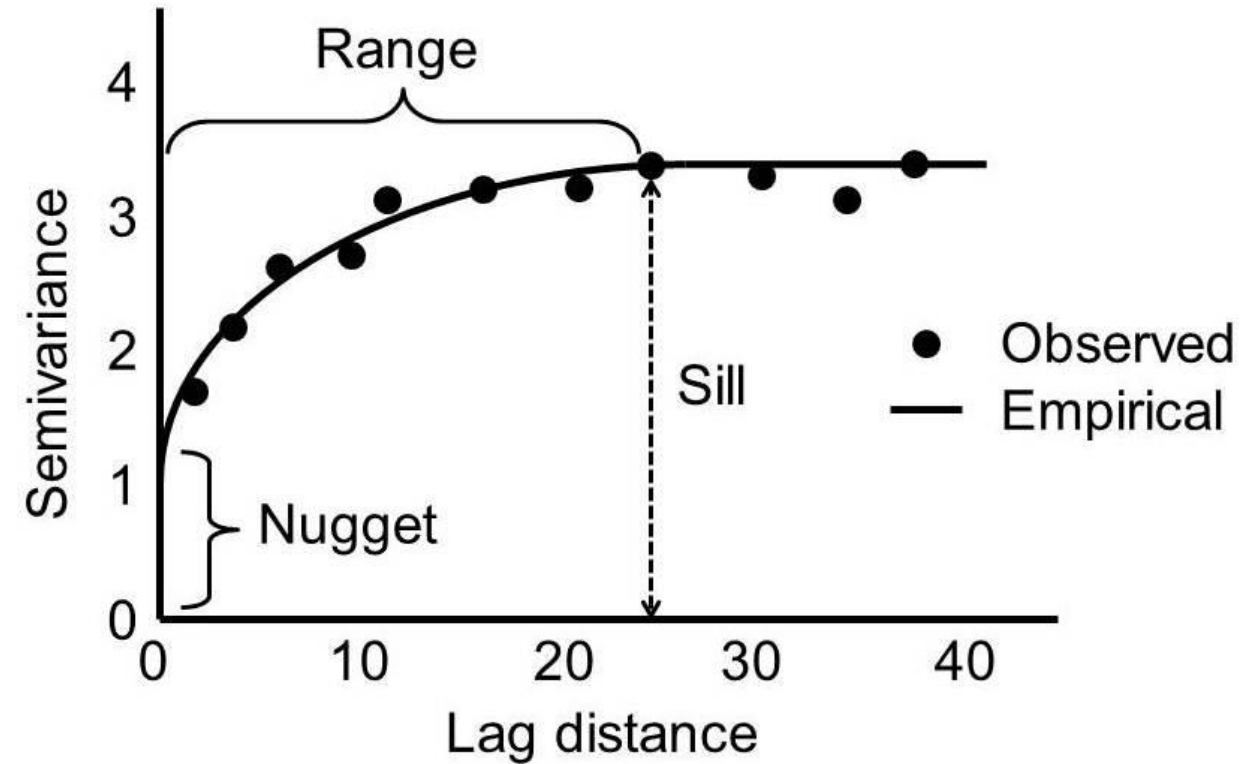
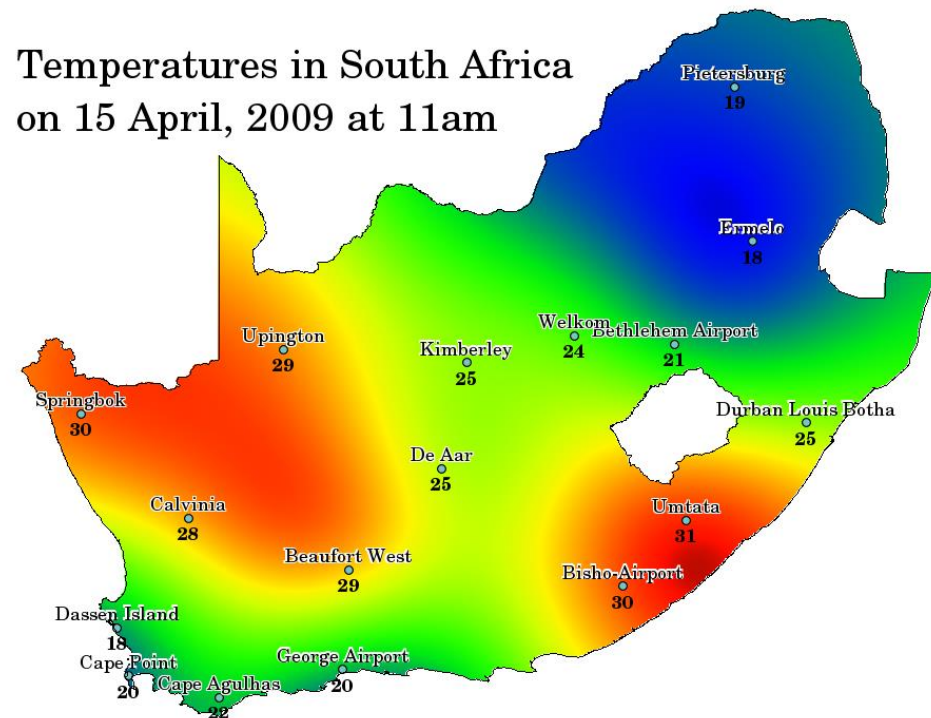
$$r(e_i, e_j) = \begin{cases} 1 & \text{si } i = j \\ \rho^d & \text{de otra forma} \end{cases}$$

Modelo autorregresivo (orden 1)



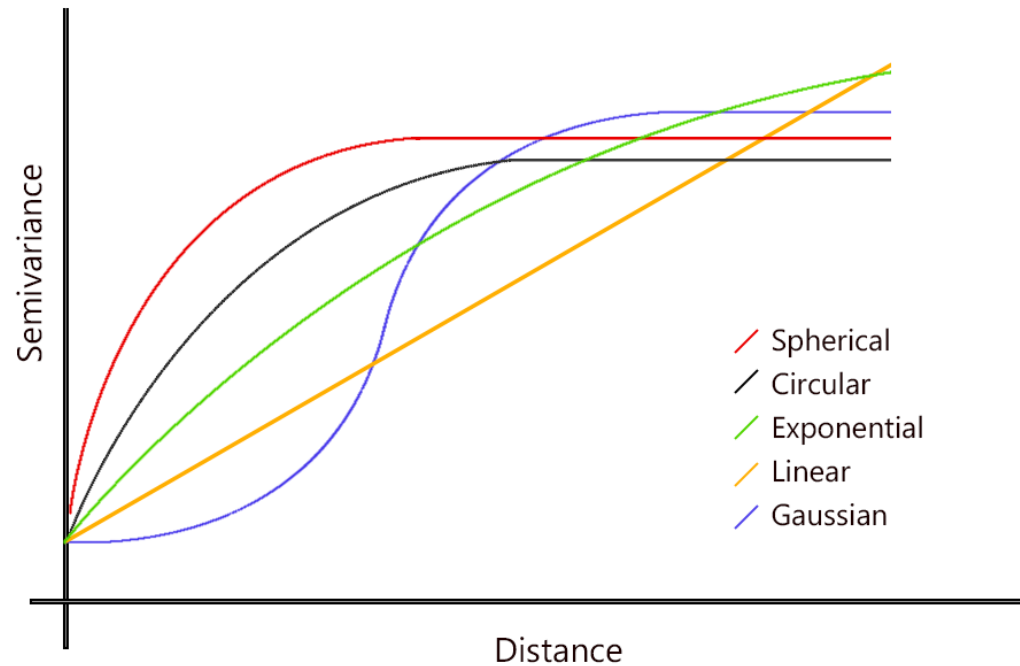
Dependencia espacial

$$r(e_i, e_j) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{de otra forma} \end{cases}$$



$$r(e_i, e_j) = \begin{cases} 1 & \text{si } i = j \\ 1 - e^{-\frac{d}{\rho}} & \text{de otra forma} \end{cases}$$

Dependencia espacial



$$r(e_i, e_j) = \begin{cases} 1 & \text{si } i = j \\ f(d, \rho) & \text{de otra forma} \end{cases}$$

¿Qué tipos de modelos son mixtos?

- GLMMs
- GAMMs

LM \rightarrow GLM \rightarrow GLMM \rightarrow GAMM



Supuestos

- Residuos con distribución normal (modelo lineal general mixto).
- Los residuos dentro de grupos son independientes con media 0 y varianza σ^2 .
- Los efectos aleatorios tienen distribución normal con media 0 y varianza σ^2 .

Validación

- Gráfico de residuos normalizados vs valores predichos.
- Gráfico de residuos normalizados vs cada variable explicatoria.

Ajuste e inferencia

- R^2_{GLMM} (Nakagawa y Schielzsch 2013)

R^2 marginal \rightarrow efectos fijos

R^2 condicional \rightarrow efectos fijos + efectos aleatorios

Parámetros

Test de razón de verosimilitud

AIC y variantes

REML

- ML es sesgado

$$Y \sim N(\alpha + \beta X, \sigma^2)$$

- REML es insesgado

$$A^T Y \sim N(0, A^T \sigma^2 A)$$

→ ML



- Si el número de variables explicatorias es pequeño con respecto al número de observaciones, no hay mucha diferencia entre ML y REML

Selección de modelos

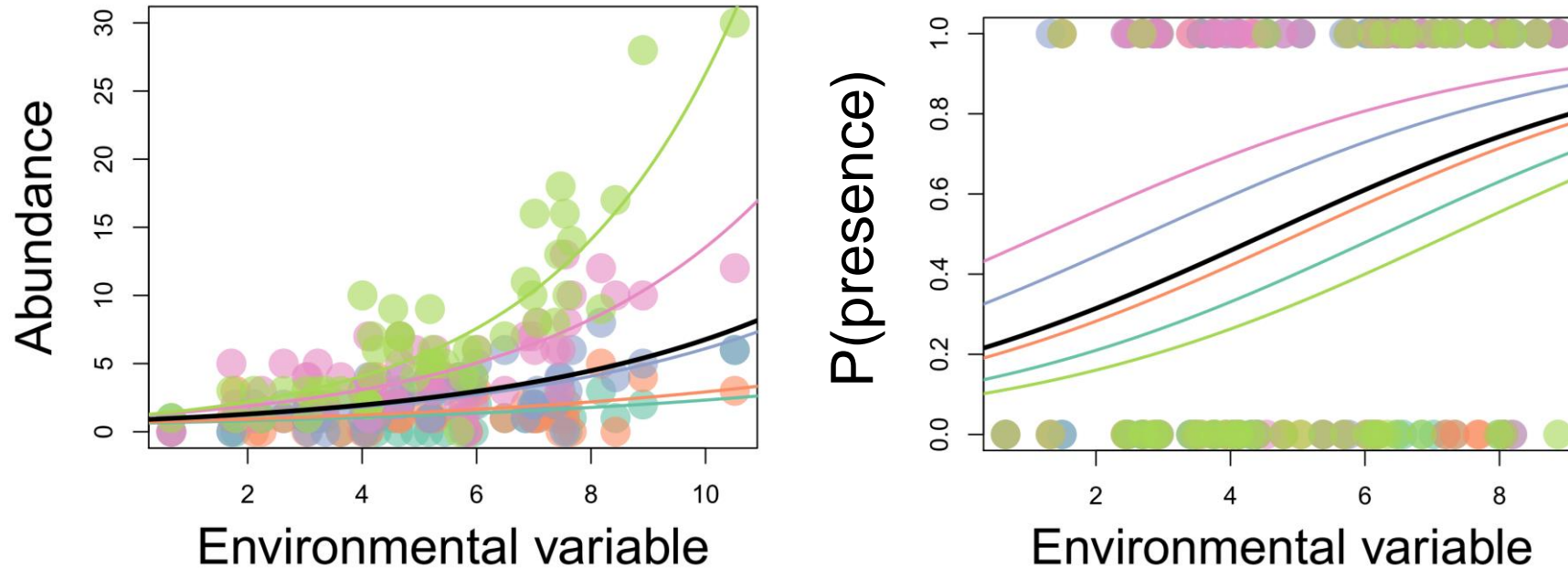
- **Protocolo de Diggle et al. (2002) y Zuur et al. (2009)**
 1. Modelo con el componente fijo que incluye todas las variables explicatorias y tantas interacciones como sea posible. REML.
 2. Identificar la estructura óptima del componente aleatorio.
 3. Identificar la estructura óptima del componente fijo. ML.
 4. Modelo final. REML.

Desventajas de los modelos mixtos

- “Hambrientos de datos” → al menos 5 niveles.
- Pueden ser inestables cuando hay gran desbalanceo.
- Dificultad en evaluar la importancia o significancia de la varianza de los efectos aleatorios.
- Problema de identificar la estructura apropiada de efectos aleatorios.

Modelos lineales generalizados mixtos

- Otras distribuciones además de la normal



Revista Brasileira de Ornitologia 26(2): 90–104.
June 2018

ARTICLE

**Advocating better habitat use and selection
models in bird ecology**

Facundo Xavier Palacio^{1,2}



<http://glmm.wikidot.com>

Paquetes:

- lme4
- nlme
- MASS
- glmmADMB
- glmmTMB
- MCMCglmm
- ASreml-R
- glmm
- glmmML
- glmmfields
- **bmrs**
- **mgcv**
- **gamm4**

Modelos mixtos en R

	lme4	nlme
Tipos de modelos	GLMM	LMM
Efectos aleatorios	$\text{lmer}(y \sim x1 + (1 \text{factor}), \text{data})$ $\text{glmer}(y \sim x1 + (1 \text{factor}), \text{family}, \text{data})$	$\text{lme}(y \sim x1, \text{random} = \sim 1 \text{factor}, \text{data})$
Diseños	Anidados y cruzados	Medidas repetidas (datos correlacionados en espacio y tiempo)
Valores de P	Sí	No
Heteroscedasticidad	No	Sí

Modelos lineales generalizados mixtos

Formula	Alternative	Meaning
$(1 \mid g)$	$1 + (1 \mid g)$	Random intercept with fixed mean.
$0 + \text{offset}(o) + (1 \mid g)$	$-1 + \text{offset}(o) + (1 \mid g)$	Random intercept with <i>a priori</i> means.
$(1 \mid g1/g2)$	$(1 \mid g1) + (1 \mid g1:g2)$	Intercept varying among $g1$ and $g2$ within $g1$.
$(1 \mid g1) + (1 \mid g2)$	$1 + (1 \mid g1) + (1 \mid g2).$	Intercept varying among $g1$ and $g2$.
$x + (x \mid g)$	$1 + x + (1 + x \mid g)$	Correlated random intercept and slope.
$x + (x \parallel g)$	$1 + x + (1 \mid g) + (0 + x \mid g)$	Uncorrelated random intercept and slope.

Table 2: Examples of the right-hand-sides of mixed-effects model formulas. The names of grouping factors are denoted g , $g1$, and $g2$, and covariates and *a priori* known offsets as x and o .

Modelos aditivos generalizados mixtos

