Departamento de Biometría, Estadística y Computación

FACULTAD DE AGRONOMIA
UNIVERSIDAD DE LA REPUBLICA

Modelos lineales mixtos en experimentación agronómica: ventajas y limitaciones en el uso de modelos espaciales en diseños experimentales

Alejandra Borges
Prof. Adjunto
Dpto. Biometría, Estadística y Computación - FAGRO

Indice

Características de la experimentación agrícola

Experimentación agrícola basada generalmente en datos de campo. Principal objetivo: comparar efectos de tratamientos.

Alta variablidad ambiental y existencia de covarianzas entre parcelas vecinas.

Condiciones similares de crecimiento (suelo, ambiente) → + ←

Competencia por recursos (suelo, agua. etc.) ← - →

Propiedades de suelo tiene patrones espaciales claros (Marriot et al., 1997)

Factores que mas inciden son la topografía (Burke et al., 1999), y el uso de la tierra (Wang et al., 2009).

Cómo trabajar con estas características?

→ Diseño experimental
→ Modelación espacial
→ Combinación de estrategias

Diseño experimental

Experimentos bien diseñados se basan en tres principios fundamentales propuestos por Fisher (1935)

- Aleatorización
- Repetición
- Control local de la variabilidad ⟵ tamaño del experimento

El principio de aleatorización es el abordaje clásico para "neutralizar" el sesgo en las estimaciones de los efectos (Fisher 1935).

Aleatorización + control local (formación de bloques) es lo que se utiliza comúnmente para solucionar el problema de errores correlacionados.

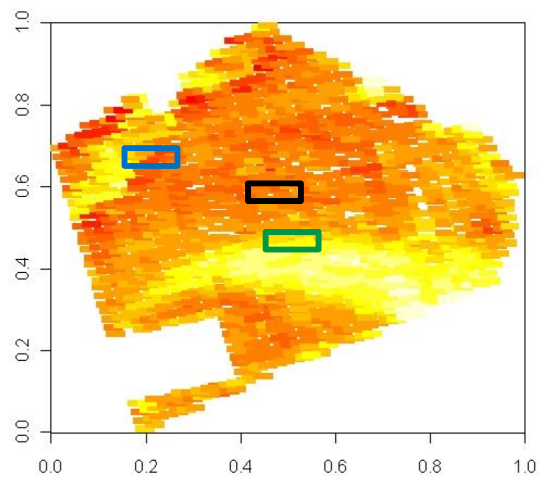
observaciones no correlacionadas (cuando los bloques son fijos)

observaciones con una correlación constante dentro de cada bloque (bloques aleatorios)

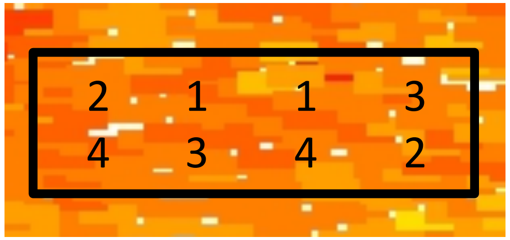
Los efectos positivos del control local utilizando bloques dependerá de una correcta evaluación de la heterogeneidad del suelo (Richther and Krowchewski 2006).

Diseño experimental

Según condiciones del terreno



DCA



$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

y_{ij}: variable de respuesta

μ: media general

τ_i: efecto del i-ésimo tratamiento

ε_{ij}: error experimental asociado cada u.e., $\varepsilon_{ij} \sim N\left(0,\sigma_\varepsilon^2\right)$

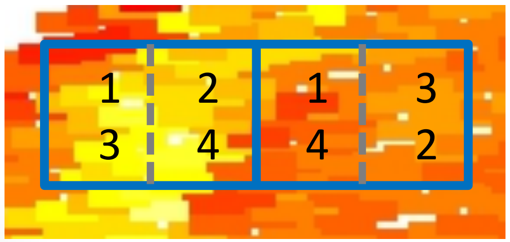
DBCA

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

y_{ij} , μ , τ_i , ε_{ij} como en DCA

β_j : efecto del j-ésimo bloque

DBI



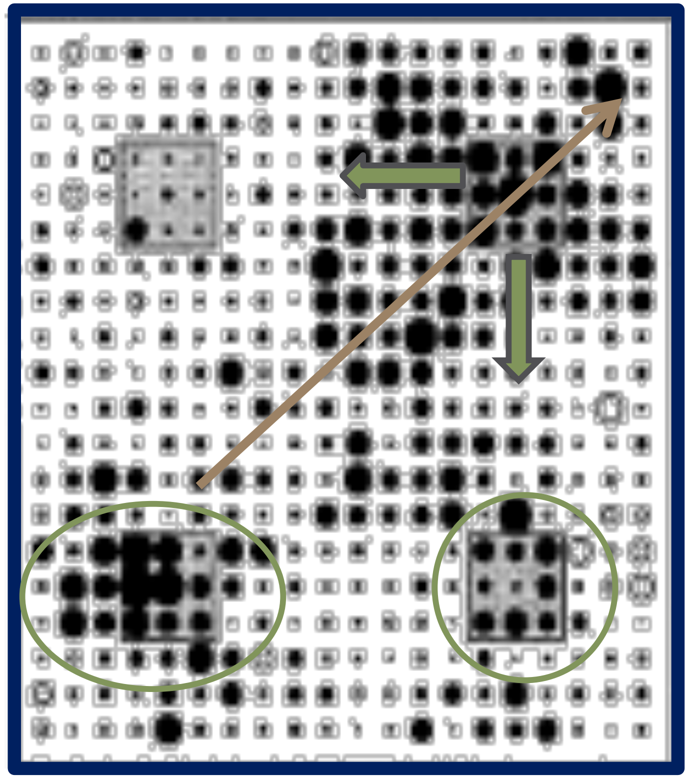
$$y_{ijk} = \mu + \tau_i + \beta_j + \gamma_{k(j)} + \varepsilon_{ijk}$$

y_{ijk} , μ , τ_i , β_j , ε_{ijk} como en DBCA

$\gamma_{k(j)}$: efecto del k-ésimo bloque incompleto anidado en el j-ésimo bloque completo

Modelos espaciales

Variabilidad espacial a menudo ocurre de forma gradual y puede no ser correctamente 'captada' por el diseño experimental (Grondona and Cressie, 1991).



Unidimensional: modelos ARMA o AR1 (Gleeson & Cullis, 1987; Qiao et al., 2000)

Bidimensional: ARMA×ARMA o AR1×AR1 (Cullis & Gleeson, 1991; Qiao et al., 2000)

Análisis de tendencia: Full quadratic models (Brownie et al., 1993; Casler and Undersander, 2000)

Tendencia+ errores correlacionados: Full quadratic models + exponential structure (Brownie et al., 1993; Casler and Undersander, 2000)

Errores correlacionados: semivariogramas Brownie et al. 2004

isotrópicos

anisotrópicos

Vecino más cercano: Smith and Casler, 2004

En muchos casos, los modelos espaciales mejoran la eficiencia del experimento

Modelos mixtos

Muy flexibles para modelar la matriz de covarianzas de los residuales, con estrategias variadas, desde geoestadísticas hasta semiparametricas como el uso de splines (González Barrios et al. 2019, Richther and Krowchewski 2006).

$$Y = X\beta + Zu + \varepsilon$$

Y vector de variables de respuesta

X matriz $n \times p$ de coeficientes asociados a los efectos fijos

β vector de parámetros desconocidos de los efectos fijos

Z matriz $n \times q$ de coeficientes asociados a los efectos aleatorios

u vector $q \times 1$ de efectos aleatorios

ε vector de errores aleatorios

$$u \sim N(0, G) \longrightarrow E(u) = 0 \quad Var(u) = G$$

$$\varepsilon \sim N(0, R) \longrightarrow E(\varepsilon) = 0 \quad Var(\varepsilon) = R$$

$$Cov(\varepsilon, u) = 0$$

Modelos espaciales o temporales

Distribución conjunta de los vectores aleatorios

$$\begin{bmatrix} u \\ \varepsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right)$$

puede asumir $\sigma_\varepsilon^2 I$

$$\Rightarrow y \sim N(X\beta, \underbrace{ZGZ'+R}_{\Sigma})$$

Matriz de varianzas y covarianzas de ***Y***

Para estimar β y u :

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix}$$

En la mayoría de los casos, los componentes de varianza y covarianza de G y R son desconocidos, y se deben usar sus estimaciones (REML).

Objetivos

i. Comparar la eficiencia de varios diseños experimentales en experimentos de gran tamaño usando diferentes estrategias de modelación espacial en condiciones de heterogeneidad espacial natural.

ii. Evaluar la performance de varios modelos espaciales y determinar si la elección de un modelo espacial correcto permite una ganancia en precision, sin inducir sesgo (o muy poco) en las comparaciones de tratamientos.

Objetivo 1 – *Comparación de diseños experimentales y modelos espaciales en ensayo de uniformidad de trigo*

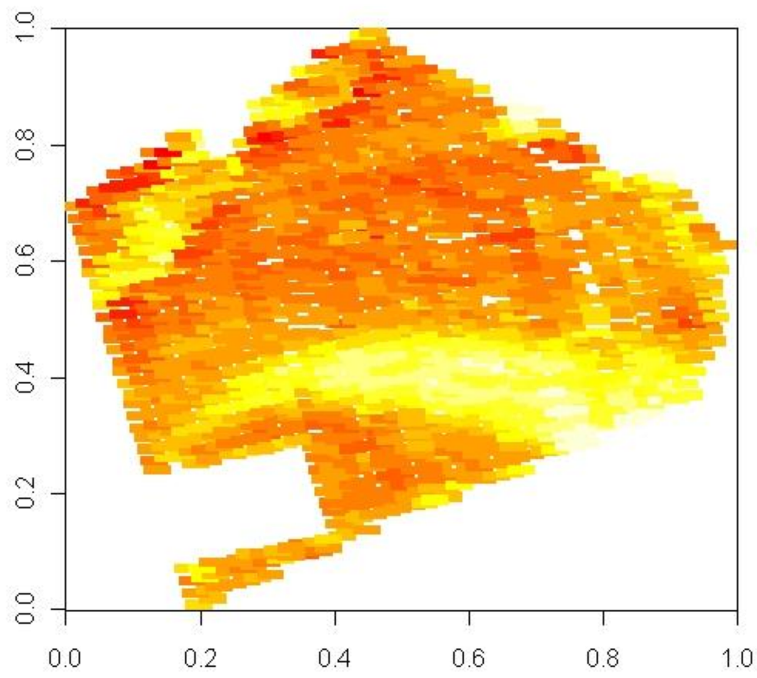
Estrategia general

Se simularon datos de rendimiento de una variedad (genotipo) de trigo y se consideró: variabilidad real de campo
efectos de diferente número de genotipos
locaciones en el terreno
diseños experimentales
modelos de correlación espacial

Datos

Ensayo de uniformidad de trigo
64 ha – Cultivar 'Nogal'

1445 datos de rendimiento de parcelas de (15m x 5m) georreferenciadas

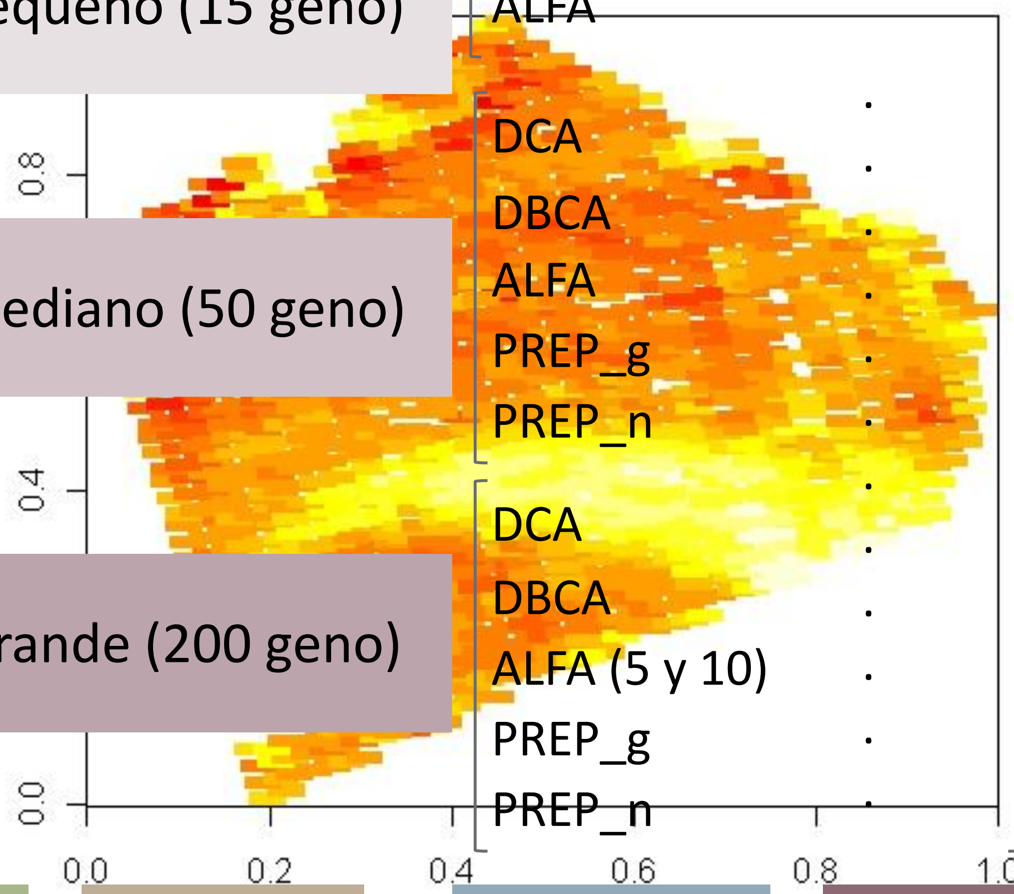


1000 realizaciones de cada combinación

Localidad
(100 puntos de partida)

'alta' y 'baja' variab. (*YSD*)

Tamaño de experim.



Tamaño de experim.	Diseños
Pequeño (15 geno)	DCA, DBCA, ALFA
Mediano (50 geno)	DCA, DBCA, ALFA, PREP_g, PREP_n
Grande (200 geno)	DCA, DBCA, ALFA (5 y 10), PREP_g, PREP_n

DCA → NSC, AR(1), EXP(2)

Indicadores

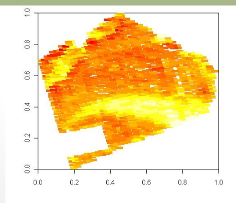
- Best_gen - 15% superior de genotipos
- SED – Error estándar de la diferencia
- COR – Coeficiente de correlación Pearson

variabilidad real de campo

efectos de genotipos

Diseños experiment.

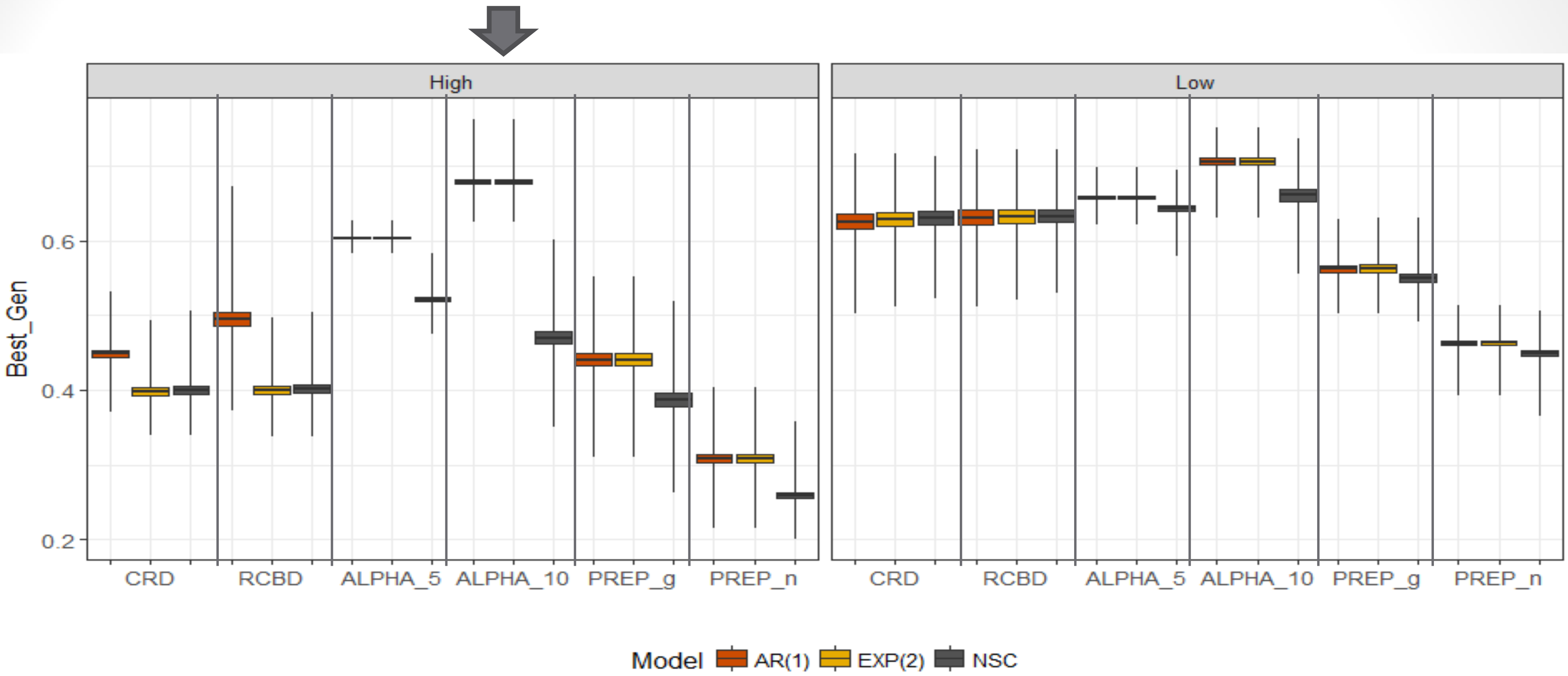
Modelos espaciales



$$G_i \sim N(0;\sigma_G^2)$$

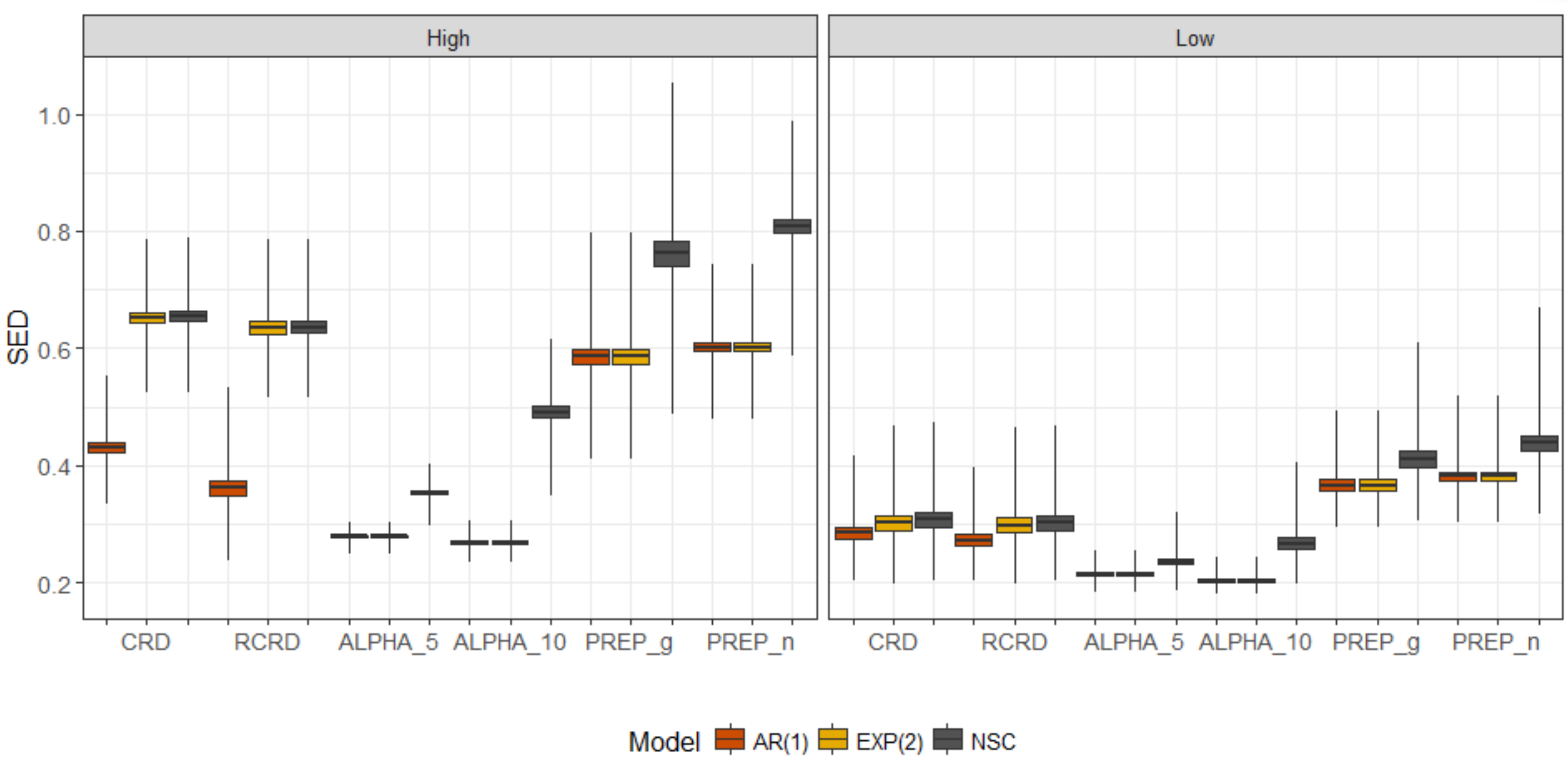
$$Y_{ij} = G_i + \varepsilon_{ij} + \delta_{ij}$$

Tamaño de experimento grande (200 genotipos)



Para COR (correlaciones entre observados y predichos) se obtuvo una respuesta muy similar

Tamaño de experimento grande (200 genotipos)



RESUMEN 1

- Efecto del diseño experimental mayor que modelos espaciales, sobre todo en experimentos de gran tamaño y locaciones de alta variabilidad
- Una vez elegido un diseño experimental adecuado, el modelado espacial puede mejorar la precisión y la exactitud en las estimaciones.
- Utilizando información que brindan tecnologías como la agricultura de precisión, se podrían lograr experimentos mejor diseñados y con una corrección espacial adecuada para cada situación.

Objetivos

i. Comparar la eficiencia de varios diseños experimentales en experimentos de gran tamaño usando diferentes estrategias de modelación espacial en condiciones de heterogeneidad espacial natural.

ii. Evaluar la performance de varios modelos espaciales y determinar si la elección de un modelo espacial correcto , permite una ganancia en precision, sin inducir sesgo (o muy poco) en las comparaciones de tratamientos.

Modelos mixtos – modelos espaciales

Modelo ya no se conoce a priori, usan un enfoque basado en datos.
Se debe estimar alguna estructura de covarianza (Richter y Kroschewski, 2012).

Desconocido, se usan los valores estimados ($\hat{\rho}$)

$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Selección mediante algún criterio de ajuste (Brownie et al., 2004, Richter et al, 2015, Saud et al., 2016).

$$\downarrow$$

$$\text{CS}\begin{bmatrix} \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 \end{bmatrix} \quad \text{AR}\begin{bmatrix} \sigma^2 & \rho & \rho^2 & \rho^3 \\ \rho & \sigma^2 & \rho & \rho^2 \\ \rho^2 & \rho & \sigma^2 & \rho \\ \rho^3 & \rho^2 & \rho & \sigma^2 \end{bmatrix} \quad \text{TOEP}\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

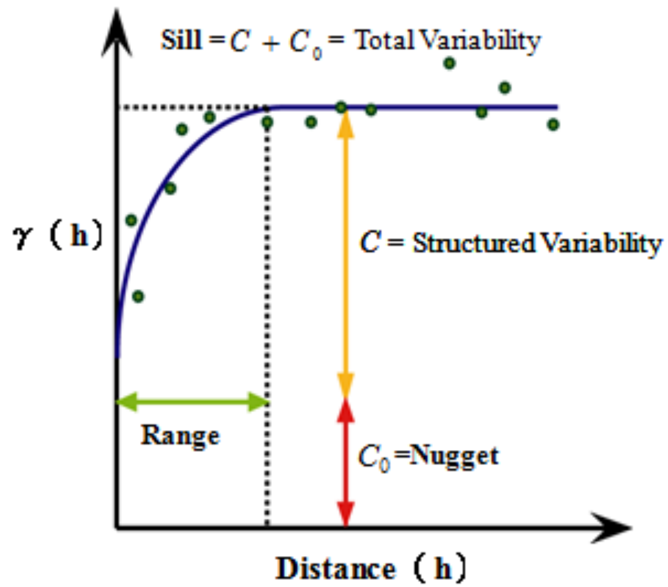
$$\text{CSH}\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix} \quad \text{ARH}\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix} \quad \text{ANTE}\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_3\sigma_1\rho_1\rho_2 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 \\ \sigma_3\sigma_1\rho_2\rho_1 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 \end{bmatrix}$$

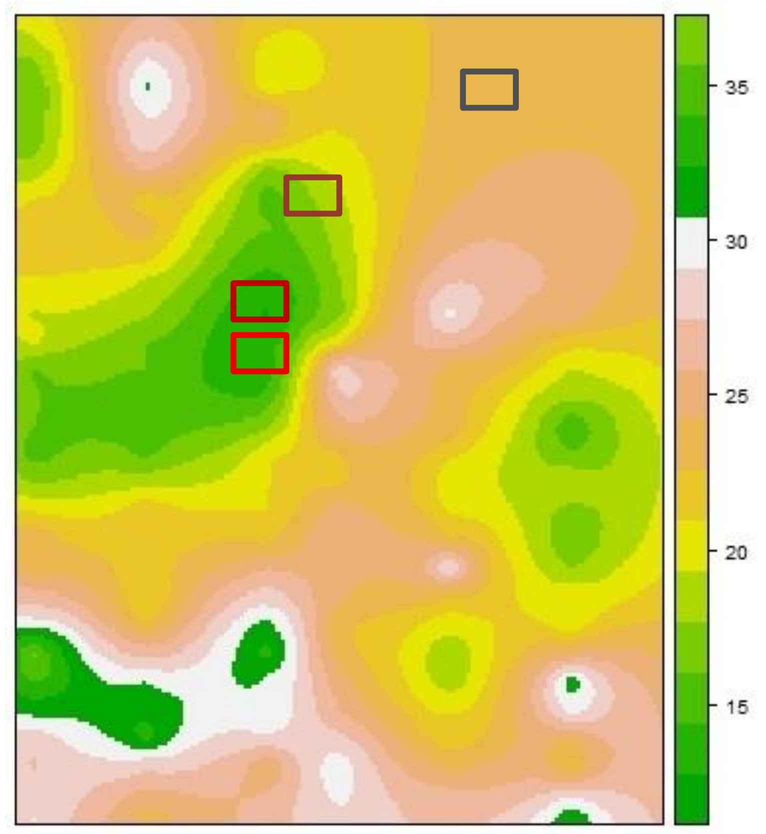
Varias estructuras de covarianza o correlación para modelar correlaciones en el espacio (y en el tiempo).

Herramientas geoestadísticas (semivariogramas).

Idea general:

- Semivariograma modela correlaciones espaciales
- Basado en cálculo de semivarianzas

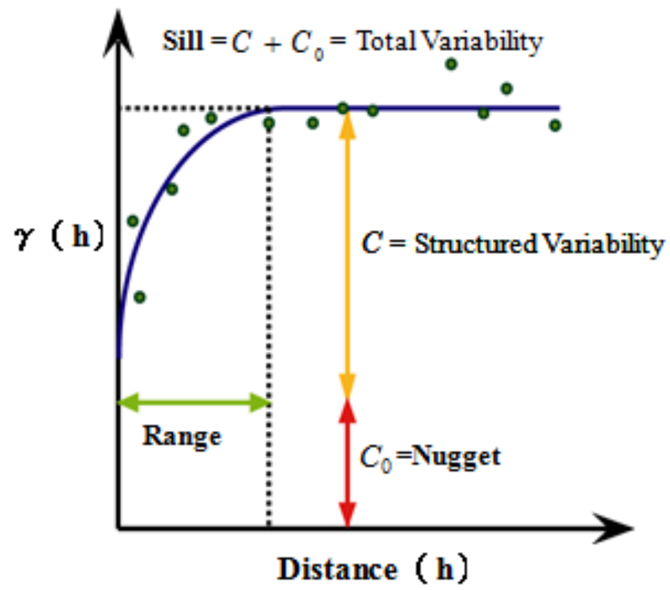




$$\gamma(\mathrm{h}) = \frac{1}{2} Var(\text{dif.entre pares de obervaciones separadas } h \text{ unidades})$$

Ejemplo: Modelo esférico

$$\gamma(h)=\begin{cases} C_0+C_1\left[\dfrac{3h}{2a}-\dfrac{1}{2}\left(\dfrac{h}{a}\right)^3\right] & 0<h\le a \quad ;a>0 \\ C & h>a \end{cases}$$



Relación de la semivarianza con la covarianza ($C(\mathrm{h})$):

$$C(\mathrm{h})=C(0)-\gamma(\mathrm{h})$$ ← estacionariedad de segundo orden

$$R=\begin{pmatrix} C(0) & C(h) & \cdots & C(h) \\ C(h) & C(0) & \cdots & \vdots \\ \vdots & & \ddots & \vdots \\ C(h) & \cdots & \cdots & C(0) \end{pmatrix}$$

Modelos mixtos – modelos espaciales

Desconocido, se usan los valores estimados ($\hat{\rho}$)

$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Esta incertidumbre no se tiene en cuenta en las comparaciones entre tratamientos → puede inducir sesgo (Richter and Kroschewski, 2012).

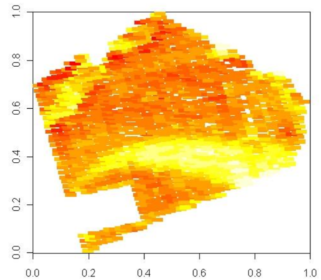
Kenward-Roger method

Este método de corrección asintótica de los grados de libertad y del error estándar estimado de los efectos fijos

Si el modelo no es el óptimo, pueden persistir sesgos en las estimaciones y desvíos en la tasa de error de tipo I esperada (Hu and Spilke, 2009; Richter and Kroschewski, 2012).

Objetivo 2 - *Estimación del sesgo en las estimaciones de varianza de modelos espaciales en un ensayo de uniformidad de trigo*

Datos – Ensayo de uniformidad de trigo

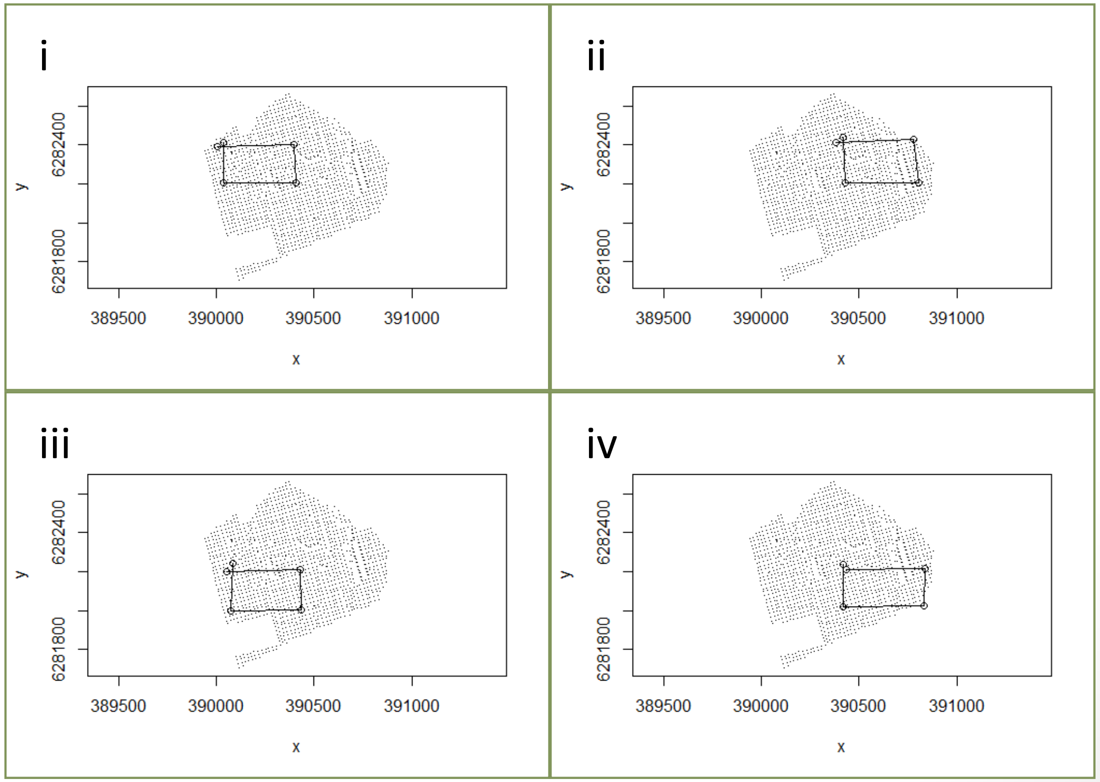


Proceso

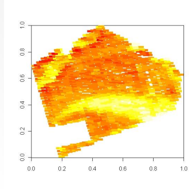
Se marcaron cuatro cuadrantes en el terreo y en cada uno se marcó una zona

En cada zona, se seleccionaron 144 parcelas.

Diferentes combinaciones de # de tratamientos y # de repeticiones.



Procedimiento estadístico



Efectos de tratamientos simulados: $\tau_i \sim N(0\ ;\ \sigma_\tau^2)$, siendo τ_i el efecto del i-*th* tratamiento y σ_τ^2 la varianza de los tratamientos.

Combinaciones:

i) t = 48 × r = 3
ii) t = 36 × r = 4
iii) t = 4 × r = 36

144 e.u.

Efectos de tratamientos fijos (incluye 2 tratamientos de igual efecto).

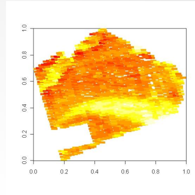
Diseños experimentales:

48 y 36 trt → CRD, RCBD, ALPHA

4 trt → CRD, RCBD

Modelos: i) Null (errores independientes)
ii) Gau (modelo Gaussiano)
iii) Sph (modelo esférico).

Procedimiento estadístico



1000 realizaciones en cada escenario: región, diseño, combinación (trat × r), modelo

Kenward-Rogers approximation (primer orden) para todos los casos

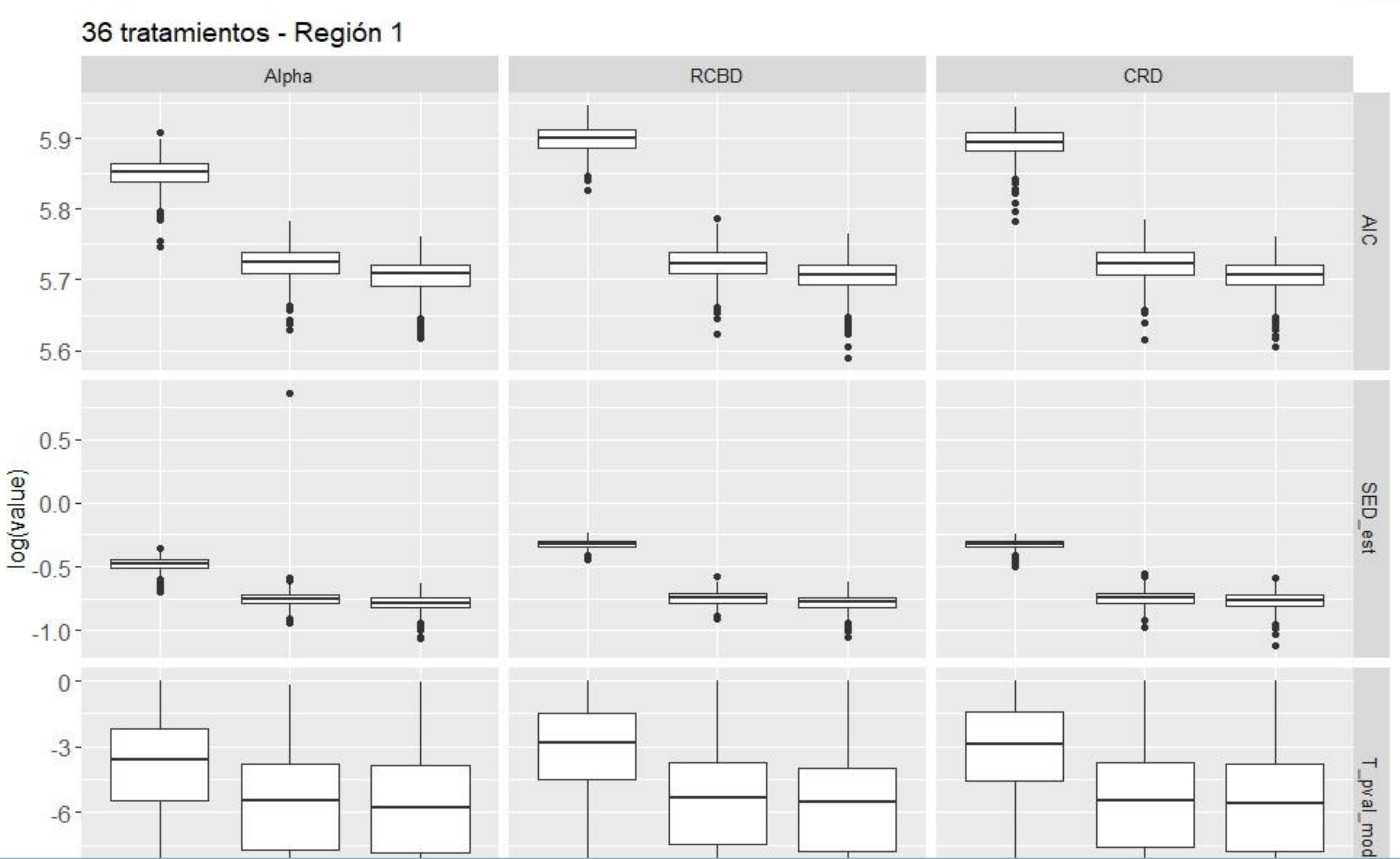
Indicadores de comparación:

- **Akaike information criterion (AIC)**
- Probabilidad de rechazar H_0 dada la muestra (p-value)
- **Número de veces H_0 es rechazada (RH_0)**
- **Error estándar estimado de la dif (SED_est)**
- Error estándar observado de las dif (SED_obs)
- **Ratio: SED_obs / SED_est**

T-test

Sin diferencia
‘Moderada’ diferencia
‘Gran’ diferencia

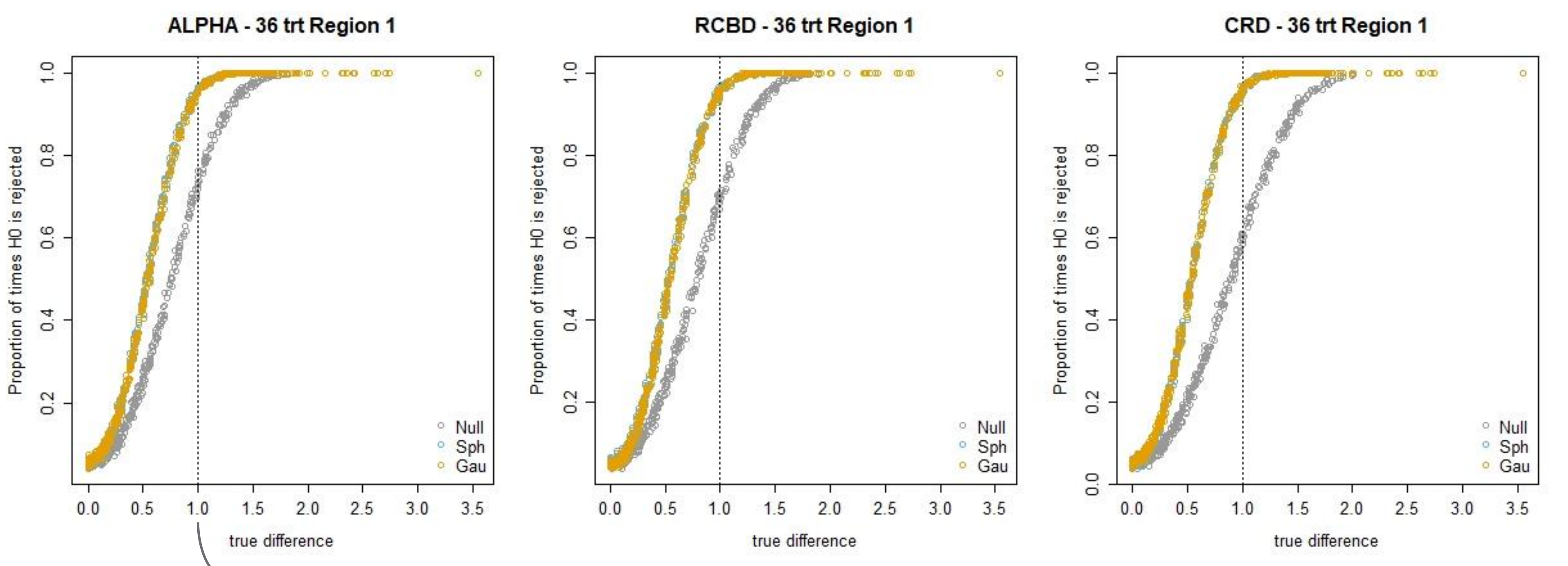
AIC, SED_est y p-value de la prueba T para tratamientos con moderada diferencia de efectos



- Mejor ajuste (menor AIC) de modelos espaciales frente a RCBD+’Null’
- Menor error estándar estimado
- Mayor número de test significativos en la prueba T

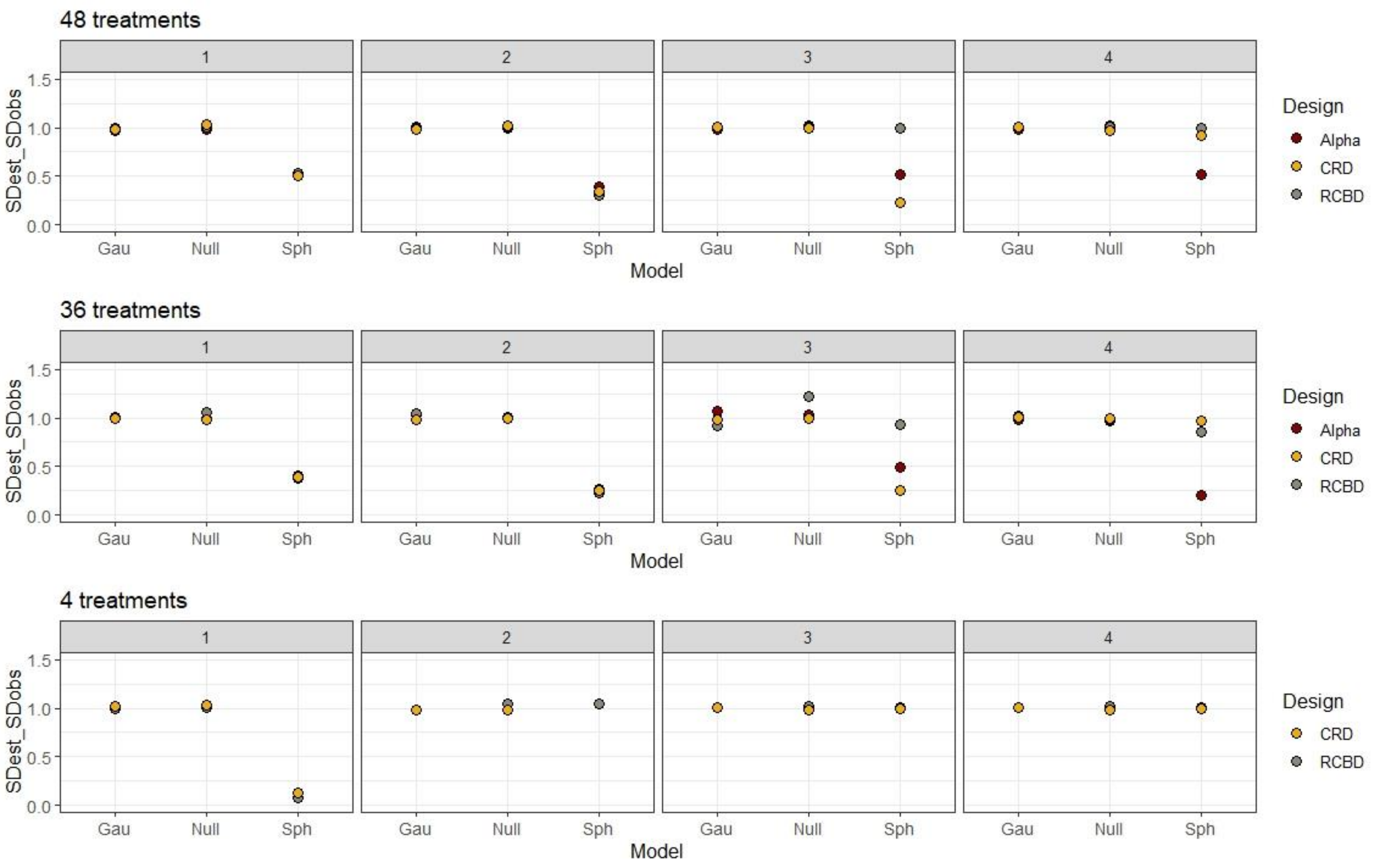
(Richter and Kroschewski, 2012, Hu and Spilke 2009)

Proporción de veces que se rechaza H0 (Región 1)

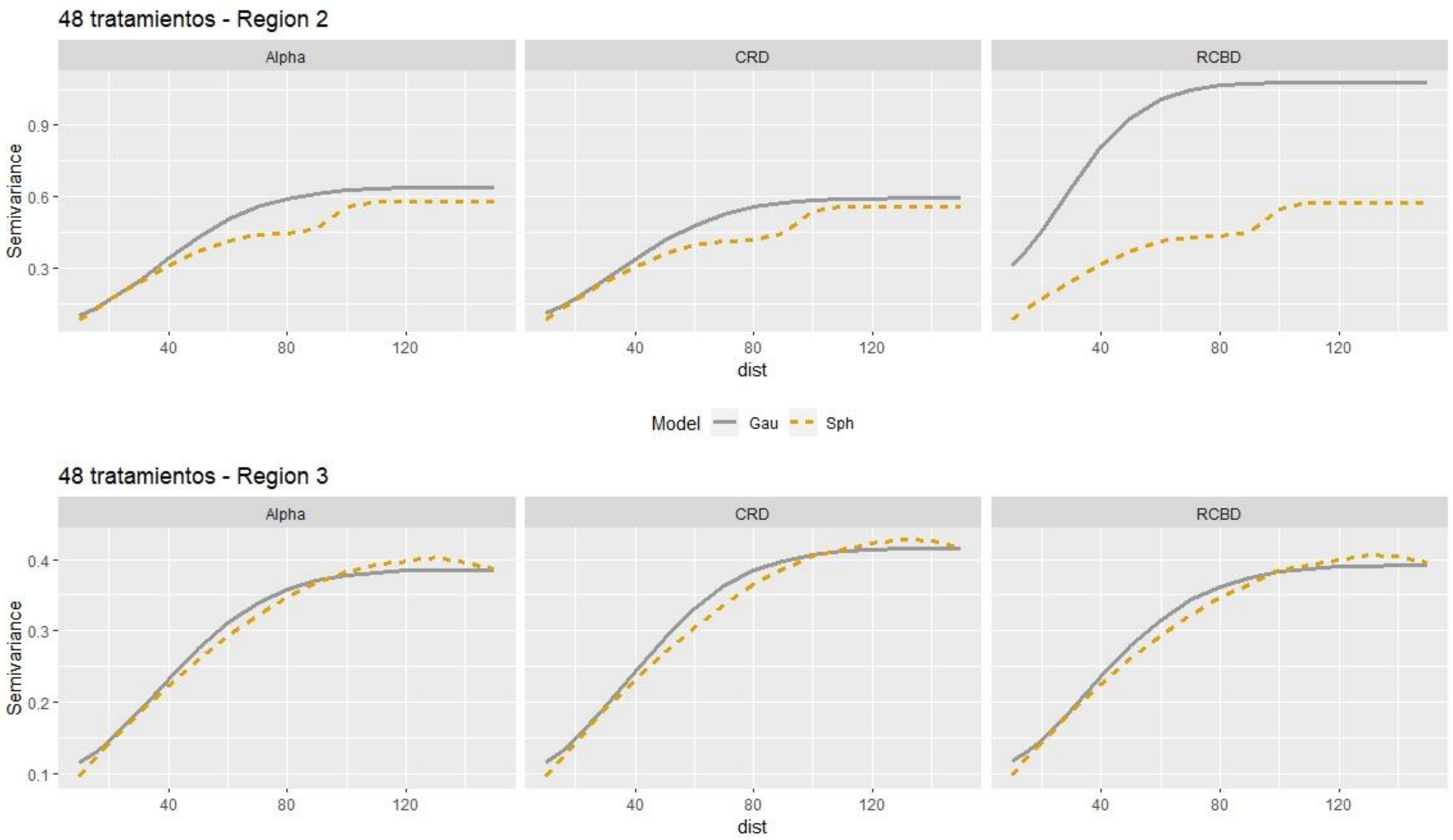


1 ton/ha de diferencia

Ratio = SDest / SDobs



Semivarianzas estimadas para modelo Gaussiano y Esférico, en dos regiones



Modelos con forma ‘sigmoide’ ajustan mejor cuando datos provienen de parcelas (‘regularized data’) (Richter and Kroschewski, 2006)

Problemas de ajuste modelo esférico con KR (Hu and Spilke 2009)

RESUMEN 2

- Los modelos espaciales mejoran la precisión y el ajuste con respecto al nulo.
- Para una diferencia dada entre tratamientos, se logra aumentar el número de veces en que la hipotesis nula es rechazada.
- Si bien estas ganancias se logran con los dos modelos evaluados con uno de ellos (Sph) se obtienen sesgos importantes en las estimaciones.

Ventajas del uso de modelos mixtos

- Permite la modelación de diseños experimentales complejos (ALPHA).
- El uso de diseños experimentales complejos es la mejor estrategia para controlar la variabilidad subyacente.
- La modelación de la variabilidad espacial en la matriz R mejora la precisión de las estimaciones y el ajuste del modelo.
- Estas ventajas son más evidentes en experimentos de mayor tamaño y bajo condiciones de alta variabilidad.

Limitaciones del uso de modelos mixtos

- Cuando no se conoce la heterogeneidad espacial del terreno, los efectos positivos del diseño experimental pueden reducirse significativamente

- El uso de modelos espaciales (modelos mixtos) puede inducir sesgo en los errores estándar de las diferencias, si el modelo no el adecuado.

- La selección del mejor modelo espacial debe basarse en más de un criterio de selección

- Brownie, C., Bowman, D.T., y J.W. Burton. 1993. Estimating spatial variation in analysis of data from yield trials: a comparison of methods. Agron. J. 85:1244–1253.
- Brownie, C., King, L., Dube, T. 2004. Longitudinal and Spatial Analyses Applied to Corn Yield Data from a Long-Term Rotation Trial. NCSU Institute of Statistics Mimeo Series # 2559
- Burke, I.C., Lauenroth, W.K., Riggle, R., Brannen, P., Madigan, B. y S. Beard. 1999. Spatial variability of soil properties in the shortgrass steppe: the relative importance of topography, grazing, microsite, and plant species in controlling spatial patterns. Ecosystems, 2(5):422-438
- Casler, M.D., y D.J. Undersander. 2000. Forage yield precision, experimental design, and cultivar mean separation for alfalfa cultivar trials. Agron. J. 92:1064–1071.
- Cullis, B.R., y A.C. Gleeson. 1991. Spatial analysis of field experiments-an extension to two dimensions. Biometrics 47(4):1449–1460.
- Fisher, R.A. 1935. The design of experiments (8th ed., 1966) New York Hafner Press
- Fox, J., Bi, H., Ades. P. 2007. Spatial dependence and individual-tree growth models. I. Characterizing spatial dependence. Forest Ecology and Management. 245:10-19.
- Gleeson, A.C., y B.R. Cullis. 1987. Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. Biometrics 43(2):277–287.
- González Barrios, P., Pérez Bidegain, M., Gutiérrez, L. 2015. Effects of tillage intensities on spatial soil variability and site-specific management in early growth of Eucalyptus grandis. Forest Ecology and Management. 346:41-50.
- Grondona, M.O., y N. Cressie. 1991. Using spatial considerations in the analysis of experiments. Technometrics 33(4):381–392.
- Hu, X. y J. Spilke. 2009. Comparison of various spatial models for the analysis of cultivar trials, New Zealand Journal of Agricultural Research, 52:3, 277-287
- Marriott, C.A., Hudson, G., Hamilton, D., Neilson, R., Boag, B., Handley, L.L., Wishart, J., Scrimgeour, C.M. y D. Robinson. 1997. Spatial variability of soil total C and N and their stable isotopes in an upland Scottish grassland. Plant and Soil 196: 151–162

Referencias

- Qiao, C.G., Basford, K.E., DeLacy, I.H., y M. Cooper. 2000. Evaluation of experimental designs and spatial analyses in wheat breeding trials. Theor. Appl. Genet. 100:9–16.
- Richter, C., y B. Kroschewski. 2012. Geostatistical models in agricultural field experiments: Investigations based on uniformity trials. Agron. J. 104:91-105.
- Richter, C, Kroschewski, B., Piepho, H.P., y J. Spilke. 2015. Treatment comparisons in agricultural field trials accounting for spatial correlation. J. Agric. Sci. 153:1187–1207.
- Saud, P., Lynch, T., Anup, K.C., Guldin, J.M. 2016. Using quadratic mean diameter and relative spacing index to enhance height–diameter and crown ratio models fitted to longitudinal data. Forestry. 89:215–229.
- Wang, Y., Zhang, X. y C. Huang. 2009. Spatial variability of soil total nitrogen and soil total phosphorus under different land uses in a small watershed on the Loess Plateau, China. Geoderma 150:141–149