

# El modelo SpATS para el análisis espacial de ensayos de campo es rápido y efectivo

Lucas Peitton<sup>1,2</sup>, Luciana Magnano<sup>2</sup>, M. Valeria Paccapelo<sup>1</sup>, M. Gabriela Borgognone<sup>1</sup>

<sup>1</sup>Department of Agriculture and Fisheries (DAF), QLD, Australia, <sup>2</sup>Universidad Nacional de Rosario (UNR), Santa Fe, Argentina



**Introducción** Los programas de mejoramiento de cultivos tienen como objetivo hallar genotipos con mejores características. Los genotipos son evaluados en ensayos a campo y su desempeño suele verse afectado por la presencia de variabilidad espacial. Modelar esta variabilidad es esencial para seleccionar los mejores genotipos de manera eficiente. El método estándar se basa en un modelo lineal mixto (MLM) que resulta muy laborioso [1]. Recientemente, se propuso un nuevo método llamado *Spatial Analysis of field Trials with Splines* (SpATS) [2] también basado en MLM, que resulta más simple y tiene un desempeño similar al del método estándar en ensayos de sorgo parcialmente replicados [3]. En este estudio se compararon los modelos estándar y SpATS en diferentes escenarios, tales como **distintos cultivos, tamaño de ensayos y diseños experimentales**.

## Materiales

➤ 14 ensayos de **porotos mung** y 17 de **garbanzo**; 90-1128 **parcelas**; 30-768 **genotipos**; diseños completamente y parcialmente replicados. Período 2008-2019.

### Modelo Estándar

El MLM siguiendo el método estándar es

$$y = X\tau + Z_g g + Z_u u + \xi + e$$

Término	Descripción	Matriz de diseño
$\tau$	Media general y efectos fijos adicionales	$X$
$g$	Efectos aleatorios de genotipo; $\sim N(0, G)$	$Z_g$
$u$	Otros efectos aleatorios (ej., fila o columna); $\sim N(0, U)$	$Z_u$
$\xi$	Errores correlacionados; $\sim N(0, R)$	—
$e$	Errores independientes; $\sim N(0, \sigma_e^2 I)$	—

- Estructura de variancia separable en procesos autorregresivos en sentido de las filas y columnas del ensayo.
- Efectos fijos y aleatorios para modelar las distintas tendencias espaciales.
- Obtención del mejor modelo espacial mediante un proceso de ajuste multi-etápico basado en gráficos de residuos y pruebas formales.
- Pueden obtenerse distintos modelos para distintos ensayos.
- Costo: Licencia de software (ASReml-R)[4].

Varios pasos para obtener el mejor modelo para un ensayo individual (ejemplo)

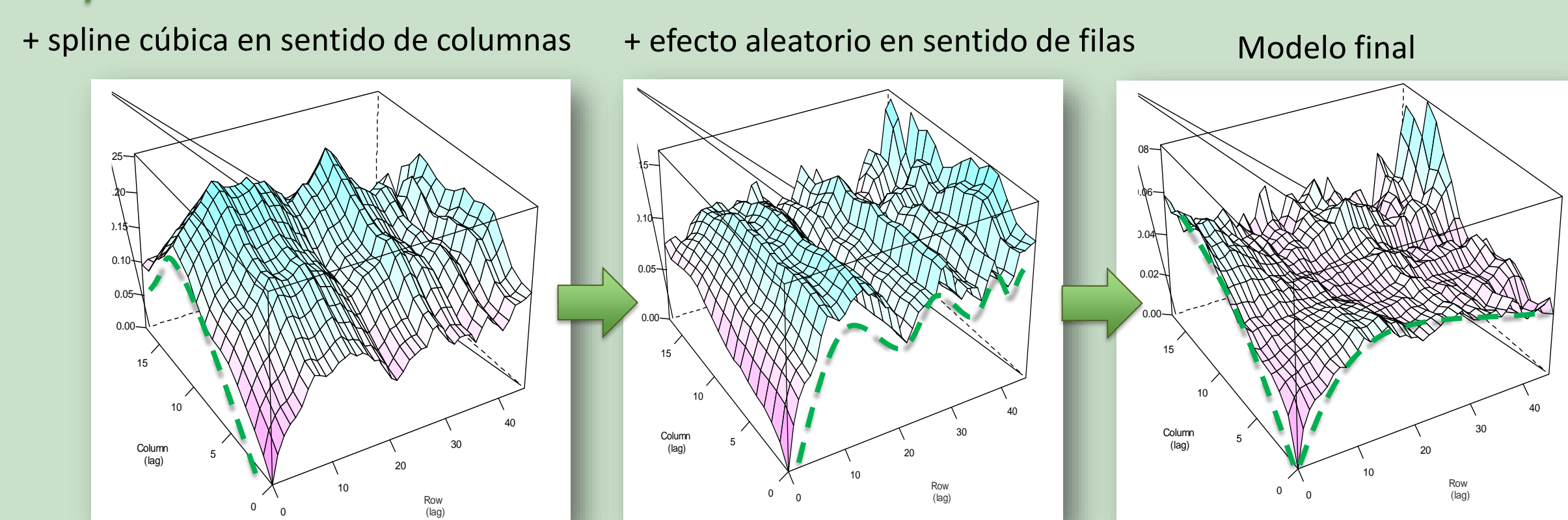


Fig. 1. Variogramas muestrales para rendimiento en grano en ensayo de garbanzo en Hermitage, Australia. En cada paso el investigador ajusta un efecto (fijo o aleatorio) dependiendo de las tendencias que observe en los residuos y el resultado de las pruebas formales.

## Resultados

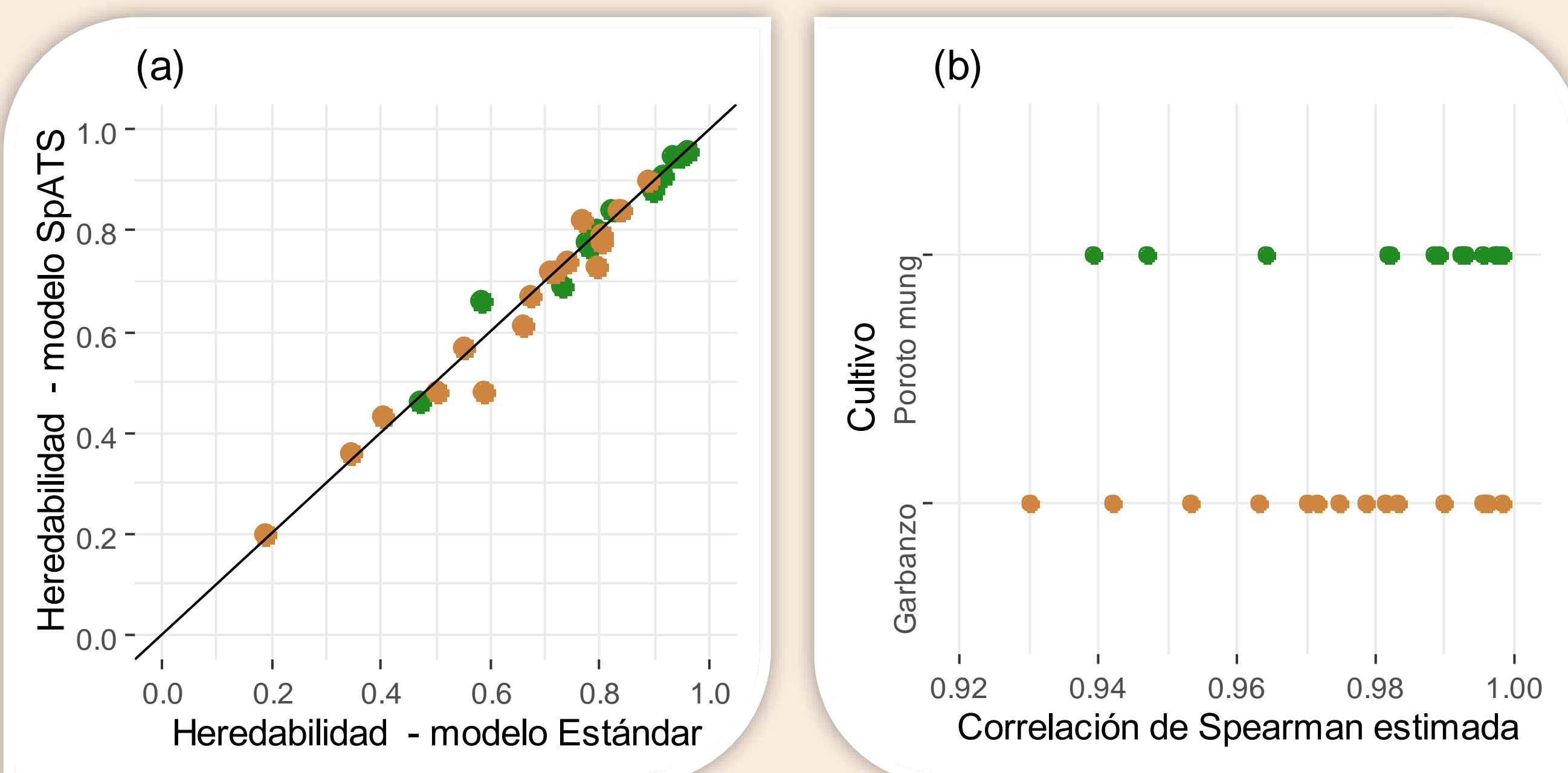


Fig 3. (a) La heredabilidad calculada fue prácticamente la misma (los puntos están muy cerca de la línea de identidad) para ambos cultivos (verde, poroto mung; naranja, garbanzo), reflejando consistencia en la precisión de los ensayos. (b) La correlación entre rankings de predicciones de efectos genotípicos de ambos modelos varió entre 0.93 y 0.99 (para ambos cultivos), mostrando un fuerte acuerdo en el ranking de predicciones.

## Más investigación para el modelo SpATS...

- Aplicación en el análisis de datos de ensayos en ambientes múltiples (MET) para comprender la interacción Genotipo-Ambiente.
- Incorporación de información genética disponible (e.g. pedigrí).

**Agradecimientos** Los autores quisieran agradecer a Grains Research and Development Corporation of Australia por financiar el proyecto DAQ1905-003RTX.



### Referencias

- [1] Gilmour, Cullis & Verbyla (1997) J Agric Bio Environ Stat 2, 269-293. [2] Rodríguez-Álvarez, Boer, van Eeuwijk & Eilers (2016) [arXiv:1607.08255](https://arxiv.org/abs/1607.08255) [stat.ME]. [3] Velazco, Rodríguez-Álvarez, Boer, Jordan, Eilers, Malosetti & van Eeuwijk (2017) Theor Appl Genet 130, 1375-1392. [4] Butler, Cullis, Gilmour and Gogel (2009) ASReml-R Reference Manual. Technical Report 3, Queensland Department of Agriculture, Fisheries & Forestry. [5] Lee, Durbán & Eilers (2013) Computational Statistics and Data Analysis 61, 22-37. [6] Rodríguez-Álvarez, Boer, Eilers & van Eeuwijk (2016) SpATS R-package. [7] Patterson & Thompson (1971) Biometrika 31, 100-109. [8] Eilers, Marx & Durbán (2015) SORT 30, 149-186.

### Modelo SpATS

El MLM siguiendo el nuevo método es

$$y = X\beta + X_s \beta_s + Z_s s + Z_g g + Z_u u + e$$

*Superficie espacial ajustada*

- Función de suavizado penalizada bajo la parametrización PS-ANOVA [5], descomponiendo los efectos espaciales  $s$  en cinco componentes espaciales mutuamente independientes.
- Los parámetros de suavizado ajustan automáticamente el grado de penalización.
- Sólo se considerará distinto número de nodos para ajustar la superficie espacial.
- Obtención del modelo en un único paso.
- Modelo general para todos los ensayos.
- Paquete de R gratuito (SpATS) [6].

Término	Descripción	Matriz de diseño
$\beta$	Media general y efectos fijos adicionales	$X$
$\beta_s$	Efectos espaciales fijos (no penalizados)	$X_s$
$s$	Efectos espaciales aleatorios (penalizados); $\sim N(0, S)$	$Z_s$
$g$	Efectos aleatorios de genotipo; $\sim N(0, \sigma_g^2 I)$	$Z_g$
$u$	Efectos aleatorios de bloque, fila y columna; $\sim N(0, U)$	$Z_u$
$e$	Errores independientes; $\sim N(0, \sigma_e^2 I)$	—

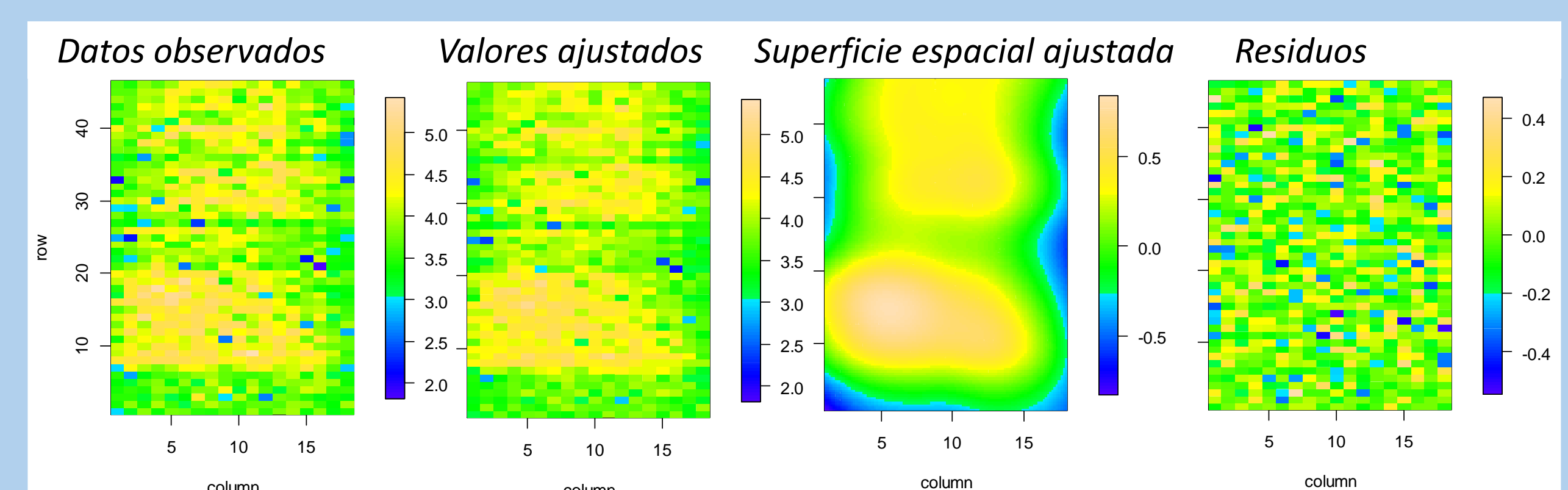


Fig. 2. Áreas de mayor rendimiento en el medio y menor rendimiento en los lados del ensayo debido a tendencias espaciales (datos observados). Los valores ajustados son precisos debido a la eficiencia del ajuste de la superficie espacial. No pueden identificarse tendencias sin modelar en los residuos.

### Número de parámetros de la superficie espacial ajustada y penalización

- Bajo la parametrización PS-ANOVA,  $s = \text{vec}(s_k)$   $k = 1, \dots, 5$ .
- Los parámetros de suavizado ajustan el grado de suavizado y son estimados vía REML [7];  
 $\lambda_{s_k} = \sigma_e^2 / \sigma_{s_k}^2 = \text{variancia residual} / k\text{-ésima variancia espacial}$ 
  - Tendencias espaciales complejas ( $\sigma_{s_k}^2 \rightarrow \infty$ ),  $\lambda_{s_k} \rightarrow 0 \Rightarrow$  baja penalización  $\Rightarrow$  más parámetros.
  - Tendencias espaciales suaves ( $\sigma_{s_k}^2 \rightarrow 0$ ),  $\lambda_{s_k} \rightarrow \infty \Rightarrow$  alta penalización  $\Rightarrow$  menos parámetros.
- El grado de flexibilidad de la función de suavizado depende del número de nodos, el cual depende del número de filas y columnas del ensayo;
  - Si el número de nodos > requerido  $\Rightarrow \lambda_{s_k}$  previenen el sobre ajuste.
  - Si el número de nodos < requerido  $\Rightarrow \lambda_{s_k}$  no proveen flexibilidad extra.
- Es preferible elegir un número de nodos que exceda la cantidad requerida para darle a la función la flexibilidad necesaria [8]. Esto no genera mayor carga computacional para el análisis de ensayos de campo.

## Conclusiones

- SpATS demostró tener resultados consistentes con los del modelo Estándar en diversos escenarios (Fig.3a). No hubo diferencias respecto al diseño ni tamaño de ensayo.
- Más flexible.** La correlación de Spearman para ensayos de garbanzo no fue tan fuerte como en los ensayos de poroto mung (Fig. 3b) debido a tendencias más complejas en estos que SpATS modeló de manera más precisa que el modelo Estándar.
- Mejor aprovechamiento del tiempo.** Se obtiene un modelo general en cuestión de segundos dado que sólo se elige el número de nodos.
- El proceso de modelado en un único paso no es subjetivo.** La superficie espacial ajustada es suficientemente flexible para modelar cualquier tendencia espacial sin importar su naturaleza y la penalización es automática.
- El paquete de R, SpATS, es de acceso gratuito** mientras que la licencia de la metodología estándar (ASReml-R [4]) es costosa.
- Todos pueden usar SpATS.** No se necesita un alto grado de conocimiento sobre MLM para ajustar el modelo y obtener las predicciones.

