Diseños experimentales y modelos espaciales y/o temporales en sistemas productivos agrícolas, ganaderos y forestales

Alejandra Borges Mira

Doctorado en Ciencias Agrarias

Abril, 2022

Diseños experimentales y modelos espaciales y/o temporales en sistemas productivos agrícolas, ganaderos y forestales

Alejandra Borges Mira

Doctorado en Ciencias Agrarias

Abril, 2022

Tesis aprobada por el tribunal integrado por el PhD. Emilio Laca, la PhD. Mónica Balzarini y la Dra. María Virginia Pravia el 20 de octubre de 2022.

Autora: Ing. Agr. Mag. Alejandra Borges. Director: Ing. Agr. PhD. José Terra. Codirector: Biol. PhD. Philip Dixon.

AGRADECIMIENTOS

A mis tutores, José y Philip, por el apoyo y la confianza en todo el proceso y por todos los espacios de discusión compartidos, que guiaron y enriquecieron enormemente este trabajo.

A los integrantes del tribunal Mónica, Emilio y Virginia por sus aportes y preguntas tanto en la revisión y evaluación del manuscrito como en la instancia de la defensa oral de esta tesis.

A mi familia, amigxs y compañerxs de trabajo, gracias infinitas por su solidaridad, su comprensión, su paciencia y su ánimo en todo el largo proceso que llevó este doctorado.

TABLA DE CONTENIDO

RESUMEN

Uno de los principales objetivos de la experimentación agronómica es obtener datos confiables que permitan establecer comparaciones entre tratamientos y poner a prueba las hipótesis de trabajo. Para cumplir este objetivo se necesitan experimentos bien diseñados y planificados, así como una estrategia de análisis de los datos que contemple los factores de diseño e información adicional relevante, especialmente en ensayos a campo de gran tamaño evaluados por varios años, donde se pueden generan correlaciones en el espacio y/o en el tiempo en las variables de interés. El propósito de esta investigación fue evaluar la eficiencia de diseños experimentales con diferente grado de complejidad y el ajuste de modelos espaciales, temporales y espacio-temporales en experimentos con diferentes cultivos, buscando evitar sesgos en las estimaciones de varianza y un control adecuado de la tasa empírica de error de tipo 1. Los resultados más relevantes de este trabajo muestran que cuando la heterogeneidad del terreno experimental es alta y el tamaño del experimento es grande, la elección del diseño experimental se vuelve esencial para obtener estimaciones precisas y con mayor exactitud. Una vez elegido el diseño experimental, la modelización espacial de la correlación entre unidades experimentales mejora aún más el desempeño del diseño elegido. Estas mejoras pueden estar acompañadas de errores estándar insesgados y un correcto control de la tasa de error empírica de tipo 1, dependiendo en gran medida del modelo elegido para estimar la correlación espacial de la variable de respuesta. Adicionalmente, cuando el número de repeticiones es alto, se pueden obtener buenos resultados más allá del modelo espacial preferido. Por otro lado, cuando se incluyen correlaciones espaciales intraparcelarias, la ventaja del uso de modelos espaciales no es tan clara. En el caso de parcelas en ensayos forestales, la inclusión de la correlación espacial entre árboles de una misma parcela significó grandes incrementos en la precisión de las comparaciones entre medias de tratamientos, lo que logró reducciones del error estándar de la diferencia de medias de hasta 40 %. Sin embargo, en ensayos sobre campo natural con parcelas de gran tamaño, las ventajas casi no se evidenciaron. Estas diferencias entre ensayos pueden deberse a la escala a la cual se manifiestan esas correlaciones espaciales, al número de submuestras por parcela y al efecto de los tratamientos sobre la distribución espacial de las variables de interés.

Palabras clave: eficiencia de diseños experimentales, modelos mixtos, variabilidad espacio-temporal

EXPERIMENTAL DESIGNS AND SPATIAL AND/OR TEMPORAL MODELS IN AGRICULTURAL, LIVESTOCK AND FORESTRY PRODUCTION SYSTEMS

SUMMARY

One of the main objectives of agronomic experimentation is to obtain reliable data to establish comparisons between treatments and test working hypotheses. To meet this objective, it is essential to have well-designed and planned experiments, as well as a data analysis strategy that takes into account the design factors and additional relevant information, mainly in large field scale trials evaluated for several years which generate correlations in space and/or time. The purpose of this research was to evaluate the efficiency of experimental designs with different degrees of complexity and the adjustment of spatial, temporal and spatio-temporal models in experiments with different crops, seeking to avoid biases in the variance estimates, and an adequate control of the empirical type 1 error rate. The most relevant results of this work show that with high experimental field heterogeneity and large experiment size, the choice of the experimental design becomes essential to obtain accurate and precise treatment effect estimates. Once the appropriate experimental design has been chosen, spatial modeling of the correlation between experimental units further improves the design performance. These improvements can be accompanied by unbiased standard errors and a correct control of the type 1 empirical error rate if the model used to estimate the spatial correlation of the response variable is adequate. Additionally, with high number of replicates, reliable results can be obtained beyond the preferred spatial model. On the other hand, when intra-plot spatial correlations are included, the advantage of spatial modeling is not as clear. In the case of forest tillage experiments, the inclusion of spatial correlation between trees in the same plot increased the precision of comparisons between treatment means, achieving reductions in the standard error of the mean difference of up to 40 %. However, in trials conducted on permanent grasslands with large plot sizes, the advantages were not clear. These differences between trials may be due to the scale at which these spatial correlations are manifested, the number of subsamples per plot and the effect of treatments on the spatial distribution of the variables of interest.

Key words: experimental design efficiency, mixed models, spatio-temporal variability

1. INTRODUCCIÓN

1.1. CARACTERÍSTICAS DE LA EXPERIMENTACIÓN AGRÍCOLA

La investigación agronómica, en general, se basa en experimentos comparativos que implican cierta modificación de la realidad, en los cuales el objetivo global es comparar o contrastar dos o más sistemas o prácticas agrícolas que puedan tener alguna relevancia en nuestro campo de investigación científica (Casler, 2015). En muchos casos, implica ensayos de campo de gran tamaño evaluados a mediano o largo plazo, donde, si bien se trabaja bajo condiciones controladas, es un gran desafío lograr homogeneidad en las condiciones experimentales, ya sea porque se genera variabilidad en el espacio y/o en el tiempo. Esto, a su vez, puede conducir a datos desbalanceados, correlaciones a través del tiempo y el espacio, y varianzas no-homogéneas a lo largo del tiempo (Brownie et al., 2004), que debe ser considerado en el análisis estadístico de los datos.

Por estas razones, no resultan obvias las elecciones específicas que se deben hacer cuando se analizan los resultados de un experimento (Casler, 2015). El error experimental en este tipo de ensayos es sensible a varios factores, algunos de los cuales se clasifican como factores de diseño experimental, como pueden ser el tipo de diseño, el tamaño de la parcela, el tamaño del bloque y el número de repeticiones (Sripathi et al., 2017). Al realizar experimentos agronómicos comparativos, solemos seguir la teoría general de la investigación científica que se muestra en la figura 1 (Casler, 2015), en la que se comienza con preguntas e hipótesis biológicas que se traducen a modelos biológicos que, a su vez, deben traducirse en hipótesis y modelos estadísticos. El modelo estadístico se debe desarrollar conjuntamente con el diseño estadístico. El diseño estadístico de un experimento incluye tanto el diseño de tratamientos como el diseño experimental y proporciona un conjunto de reglas y procedimientos que nos permiten realizar el experimento.

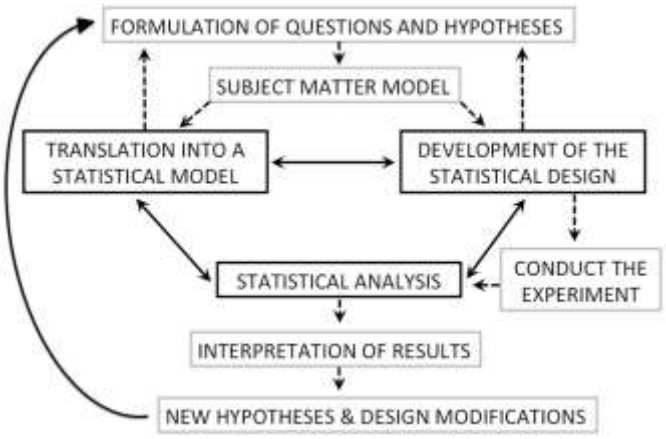


Figura 1. Diagrama de flujo de los pasos lógicos en la experimentación científica (extraído de Casler, 2015)

1.2. DISEÑOS EXPERIMENTALES

Los experimentos bien diseñados se basan en los tres principios propuestos por Fisher (1935): aleatorización, replicación y control local. Después de Fisher, la investigación agrícola se ha basado en estos tres principios que buscan controlar la variación local, asumiendo independencia entre las unidades experimentales y utilizando repeticiones para estimar el error experimental y aumentar la precisión de las estimaciones. En condiciones de campo, la variación espacial de los factores ambientales y del suelo determina, en muchos casos, la presencia de autocorrelación espacial en la respuesta del cultivo (Grondona et al., 1996, Legendre, 1993, Slaets et al., 2021). Por lo tanto, los experimentos pueden diseñarse para tener un control local de esta variabilidad espacial. Si, además, se quiere evaluar un alto número de tratamientos, la elección del diseño es clave en términos de precisión y exactitud en la estimación de los efectos de los tratamientos.

En agricultura, los diseños experimentales más utilizados son el diseño completamente al azar (CRD) y el diseño de bloques completos al azar (RCBD) (Piepho et al., 2015). En algunos casos, el bloqueo es una forma efectiva de controlar el error experimental (Casler, 2015, Cochran y Cox, 1992), pero no es suficiente en

situaciones agrícolas donde la heterogeneidad del campo y/o el tamaño del experimento son grandes. En estos casos, diseños con un mejor control local son más adecuados (Brownie et al., 1993). Algunos diseños, como los de bloques incompletos (Patterson y Williams, 1976) o los diseños *row-column* (Williams et al., 2006), se crearon específicamente para corregir los efectos de la heterogeneidad del campo en situaciones en las que se evalúa una gran cantidad de tratamientos, así como otras estructuras de tratamiento complejas (Williams et al., 2002). Más recientemente, se han propuesto diseños espaciales, que tienen restricciones adicionales en su aleatorización basadas en una estructura de correlación de dependencia para el campo que se conoce o se asume a priori (Williams y Piepho, 2013). No obstante, estos diseños no siempre se usan debido a su complejidad y dificultades prácticas que imponen ciertas restricciones sobre el número de tratamientos que se pueden incluir o simplemente porque los investigadores no están acostumbrados a elegirlos (Casler, 2015). En los ensayos de evaluación de cultivares, el número de tratamientos generalmente es alto y existe una creciente necesidad de poder diferenciar cultivares por pequeñas diferencias (Casler y Undersander, 2000), por lo que cada vez es más necesario mejorar la eficiencia de los diseños experimentales.

Sin embargo, aun con diseños de fuerte control local, como los diseños en bloques incompletos, la variabilidad espacial puede no ser correctamente controlada si ocurre a escalas más pequeñas que el tamaño de los bloques (Grondona et al., 1996). Por estas razones, la estimación de la variabilidad espacial y su inclusión en los modelos de análisis puede producir beneficios adicionales.

1.3.VARIABILIDAD ESPACIAL Y MODELOS ESPACIALES

La variabilidad espacial del suelo es ampliamente reconocida como uno de los factores más importantes que afectan las comparaciones de los efectos de tratamiento en experimentos de campo (López y Arrúe, 1995). A menudo ocurre de manera gradual y a veces no es capturada adecuadamente con el diseño experimental

(Grondona y Cressie, 1991). Una alternativa es utilizar un método de análisis que considere la variación espacial a nivel de parcela para estimar y corregir la variación espacial en la variable de respuesta. Se han utilizado varios modelos que incluyen distintas estrategias espaciales, como estimar diferentes estructuras de correlación de la matriz de varianza-covarianza de los errores (R), incluir análisis de tendencias o mezclas de ambos (Brownie et al., 1993, Casler y Undersander, 2000). Al mismo tiempo, esos procedimientos de análisis espacial pueden ser uni o bidimensionales (Cullis y Gleeson, 1991, Gilmour et al., 1997, Gleeson y Cullis, 1987, Qiao et al., 2000). En la mayoría de las situaciones analizadas en trabajos previos, se constató que los modelos que incluyen correlación espacial entre unidades experimentales son, en general, más eficientes que los que no la incluyen (Brownie et al., 1993, Kravchenko et al., 2006; Mallarino et al., 2000).

Por otra parte, cada vez es más frecuente tener múltiples observaciones por unidad experimental, sobre todo cuando el tamaño de parcela debe ser grande. Un análisis de varianza simple, que supone la independencia de las observaciones dentro de una misma parcela, puede no ser apropiado para estos datos (Slaets et al., 2021), siendo más adecuado un modelo lineal mixto que tenga en cuenta esa correlación intraparcelaria.

En ensayos forestales, el uso de la información sobre la variabilidad espacial de los componentes del suelo mejora la eficiencia en la estimación del efecto de los tratamientos en experimentos de laboreo (González Barrios et al., 2015). En este tipo de ensayos, la competencia de los vecinos y la variabilidad espacial a pequeña escala intraparcelaria pueden tener un peso relativo importante (Fonseca et al., 2011). La variación a pequeña escala tiende a crear correlaciones positivas entre los vecinos porque están creciendo en condiciones ambientales similares, mientras que la competencia tiende a crear dependencias negativas en el tamaño o el crecimiento de los individuos en la proximidad espacial (Fox et al., 2007).

En ensayos de pasturas naturales con animales en pastoreo y suelos indisturbados

también es común trabajar con parcelas grandes. Las mejoras del campo natural mediante la intersiembra de leguminosas y/o la fertilización de pastizales naturales son dos prácticas de manejo usuales para incrementar la productividad y calidad de éstas (Bondaruk et al., 2020). Para evaluar los efectos de estas mejoras del campo natural bajo pastoreo, es necesario implementar ensayos que implican el uso de parcelas de gran tamaño. Si se quiere evaluar el contenido de materia orgánica o carbono orgánico en el suelo, es necesario colectar varias muestras dentro de la parcela, donde es esperable que la variabilidad espacial entre submuestras sea importante y, además, que esta variabilidad esté afectada por los manejos diferenciales. Existe evidencia que la materia orgánica y el carbono orgánico del suelo presentan patrones espaciales claros (Marriott et al., 1997) y que la mayor variación espacial ocurre en relación con la topografía (Burke et al., 1999), así como con el uso de la tierra (Wang et al., 2009). Estos últimos autores encontraron que la variabilidad espacial del nitrógeno y el fósforo total del suelo cambian significativamente con el tipo de uso de la tierra (tierras cultivadas versus pastizales) y, por tanto, considerar esta variación espacial en este tipo de ensayos podría aumentar la precisión y predicción de los modelos.

Sin embargo, la ganancia en precisión que pueden traer los modelos espaciales puede, en parte, deberse a un sesgo en las estimaciones de varianza y, por ende, en los errores estándar de la diferencia entre medias de tratamiento (Richter y Kroschewski, 2012, Zimmerman y Harville, 1991). Según Richter y Kroschewski (2012), cuando se utilizan modelos lineales mixtos que tienen en cuenta la correlación espacial, el modelo ya no se conoce a priori como sucede en el análisis clásico de errores independientes, sino que se trabaja con un modelo basado en datos, en el que hay que estimar una estructura de covarianza. La estructura de la matriz de varianza-covarianza más adecuada a cada situación se selecciona según algunos de los criterios de ajustes más usados (Brownie et al., 2004, Richter et al., 2015, Saud et al., 2016). Los parámetros de covarianza de esas estructuras son desconocidos en la práctica y, por lo tanto, se deben usar los valores de estimación basados en los datos. En este caso, las pruebas estadísticas sobre los efectos fijos de los modelos lineales

mixtos generalmente no son exactas y los grados de libertad se determinan por aproximación usando algunas de las correcciones más extendidas (Hu y Spilke, 2009). En este marco, se puede generar un sesgo en el error estándar de la diferencia entre tratamientos, que se deriva de asumir varianzas y covarianzas estimadas como si fueran parámetros conocidos, por lo que esa incertidumbre no se toma en cuenta en las comparaciones entre tratamientos (Richter y Kroschewski, 2012). Existen antecedentes de comparación de modelos a partir de ensayos de uniformidad, donde se puede comparar la varianza observada de la diferencia de tratamientos con la estimada a partir de los diferentes modelos (Richter y Kroschewski, 2012, Wu y Dutilleul, 1999, Zimmerman y Harville, 1991). Richter y Kroschewski (2012), en un trabajo a partir de varios ensayos de uniformidad de diferentes cultivos, encontraron sesgos en el promedio de la estimación del error estándar de la diferencia entre tratamientos de los modelos espaciales, comparado con el modelo de bloques completos al azar básico (errores independientes). El sesgo fue, en algunos casos, positivo y en otros, negativo, dependiendo del cultivo, y fue de mayor magnitud en los casos donde existía una correlación espacial fuerte. Richter et al. (2015) evaluaron el efecto de incluir correlaciones espaciales en las comparaciones de tratamientos a partir de diferentes escenarios de simulación espacial y constataron que cuando la simulación espacial y el modelo de análisis no coinciden, se observan problemas de sesgo y desviaciones de la tasa de error tipo 1 nominal, por lo que la elección del criterio de selección del modelo y la elección de los modelos candidatos son cruciales. Lo que estos trabajos no evaluaron es el efecto del número de repeticiones por tratamiento sobre el sesgo de estimación y el control de la tasa de error tipo 1 empírica. En este sentido, Hu et al. (2006) se preguntan si en situaciones donde existe dependencia espacial de la variable de interés y se tiene un número bajo de repeticiones se pueden emplear modelos espaciales que tengan errores de estimación aceptables.

1.4.VARIABILIDAD TEMPORAL Y MODELOS TEMPORALES

En experimentos de largo plazo, como en ensayos forestales que implican varios años de evaluación, es un desafío modelar variables como altura o volumen porque son producto de la variabilidad temporal (Skovsgaard y Vanclay, 2013). Cuando se toma una secuencia de mediciones en cada unidad experimental, se genera una estructura temporal que puede ser explicada por variables ambientales a corto, mediano y largo plazo (Fox et al., 2001). Esto provoca que las observaciones repetidas en el tiempo en cada unidad experimental no sean independientes, sino que existe una tendencia a que dos mediciones cercanas en el tiempo sean más similares que el promedio. Esta estructura temporal de los datos longitudinales se ha modelado en diferentes disciplinas (Brownie et al., 2004, Gregoire et al., 1995) y se han utilizado, en general, dos grupos principales de estrategias para esto: i) modelos de crecimiento que utilizan enfoques de regresión (Diggle et al., 2002) y ii) modelos lineales mixtos (Saud et al., 2016). La mayoría de los modelos de regresión que se han utilizado para modelar el crecimiento de la población son logísticos o exponenciales (Tsoularis y Wallace, 2002). Por otro lado, los modelos mixtos permiten el modelado de diferentes estructuras de varianza-covarianza entre árboles individuales a lo largo del tiempo (Wolfinger, 1996), obteniéndose, en muchos casos, una mejora en la precisión de la estimación de los efectos de tratamiento (Fox et al., 2001).

1.5.MODELOS ESPACIO-TEMPORALES

En muchas situaciones, las respuestas de los cultivos bajo estudio están influidas simultáneamente por procesos temporales y espaciales, dado que se evalúan en períodos largos y en parcelas grandes. Estas condiciones pueden provocar que se generen datos desbalanceados, correlaciones a través del tiempo y el espacio, y heterogeneidad de varianzas (Brownie et al., 2004), por lo que, en algunos casos, es necesario incluir la variabilidad espacial y temporal en los modelos de análisis para

obtener estimaciones y predicciones más ajustadas a la realidad.

En los experimentos de labranza forestal, la estimación del crecimiento individual de los árboles es un proceso complejo que se ve afectado, principalmente, por la competencia de los vecinos, la variabilidad espacial a microescala, la mortalidad y los factores genéticos (Fonseca et al., 2011). Por esas razones, es relevante encontrar una estrategia de análisis óptima que combine la variabilidad espacial, la información de la competencia entre árboles y la evolución temporal. En ese sentido, existen trabajos que confirman una mejora en el desempeño de los modelos de crecimiento de cultivos que incorporan las correlaciones espaciales (Liu y Ashton, 1999), así como en modelos sobre propiedades del suelo evaluadas en el tiempo (Lee et al., 2009). En los últimos años, el desarrollo de software ha permitido comenzar a entender estos fenómenos desde una perspectiva más integral. Se han propuesto algunos enfoques espacio-temporales que se centran en las estructuras de correlación espacial y temporal, con herramientas geoestadísticas como semivariogramas espacio-temporales (Brownie et al., 2004, Gregoire et al., 1995, O'Rourke y Kelly, 2015) o más recientemente usando modelos bayesianos espacio-temporales, basados en la aproximación de Laplace anidada integrada (Blangiardo et al., 2013, Wikle et al., 2019). Sin embargo, sigue siendo escasa la información para evaluar esta variabilidad conjunta en el contexto de diseño de experimentos forestales.

1.6. DISEÑOS EXPERIMENTALES VERSUS MODELACIÓN ESPACIAL

A menudo, se debate si es mejor controlar la variabilidad espacial con diseños experimentales o con modelos de análisis espacial. No siempre el análisis espacial es necesariamente mejor que el diseño en bloques completos al azar, en particular en los casos en que la estructura espacial de la variable estudiada no está bien caracterizada (Kravchenko et al., 2006). Müller et al. (2010), en ensayos de cebada y remolacha azucarera, encontraron que el modelo base, que solo incluye un efecto bloque y

repetición, mostró mejor ajuste basado en el AIC para la mayoría de los casos que los modelos espaciales. En muchos casos, se han realizado ensayos de uniformidad con el fin de determinar el diseño experimental óptimo (Koch y Rigney, 1951). En algunos casos, se han comparado distintos tipos de diseños de bloques completos aleatorizados, donde se variaba el tamaño y la forma de la unidad experimental y de los bloques, analizados con un modelo de errores independientes y con diferentes modelos espaciales (Richter y Kroschewski, 2012).

Existen antecedentes que muestran que el desempeño de los modelos para las variables de respuesta de interés mejora cuando se combinan experimentos bien diseñados con ajustes espaciales (Gilmour et al., 1997, Qiao et al., 2000, Williams et al., 2006), lo que sugiere que se deberían considerar ambas estrategias. Algunos autores sugieren que el modelado espacial debe considerarse como una estrategia complementaria en lugar de una alternativa al diseño experimental (Piepho et al., 2015, Richter et al., 2015).

1.7. HIPÓTESIS, OBJETIVOS Y ESTRUCTURA GENERAL DE LA TESIS

Este proyecto evalúa la eficiencia de diseños experimentales con diferente grado de complejidad y el ajuste de modelos espaciales, temporales y espacio-temporales en experimentos a escala de campo con diferentes cultivos, con el objetivo de mejorar la precisión y exactitud en las estimaciones de las medias de tratamiento, tratando de evitar sesgos en las estimaciones de varianza.

1.7.1. Hipótesis de trabajo

i. En experimentos de campo grandes con un alto número de tratamientos, los diseños experimentales de mayor control local y el uso de modelos espaciales mejoran la precisión en la estimación y comparación de los efectos de tratamiento.

ii. Sin embargo, esta ganancia en precisión podría acompañarse de un sesgo en la estimación de la varianza de la diferencia de medias de tratamientos, dependiendo en gran medida de la estructura de correlación elegida en el modelo de análisis. Con un modelo espacial que dé cuenta de la variabilidad espacial subyacente, la precisión de la estimación de la diferencia del efecto del tratamiento podría lograrse sin sesgos.

iii. En la experimentación a largo plazo, como en ensayos de laboreo forestal, se necesitan parcelas de gran tamaño para evaluar el efecto del tratamiento y medidas repetidas a tiempo. Estas características provocan variabilidad espacial intraparcelaria y correlaciones en el tiempo. Por lo tanto, las variables de crecimiento se ven afectadas por procesos espacio-temporales. El modelado de estos procesos mejora la eficiencia de la estimación de efectos de tratamientos.

iv. En ensayos de largo plazo, como aquellos en campo natural bajo pastoreo, el tamaño de parcela necesariamente es muy grande, lo que lleva una alta variabilidad intraparcelaria. Esta variabilidad podría verse afectada, a su vez, por las prácticas de manejo del suelo. Los mejoramientos en cobertura de pasturas naturales mediante el agregado de leguminosas y fertilizantes fosfatados pueden provocar mayor heterogeneidad en el carbono orgánico del suelo y en el contenido de fósforo y modificar su patrón espacial. Por este motivo, el uso de modelos mixtos con correcciones espaciales mejora la eficiencia de los diseños utilizados y ayuda a entender los procesos inherentes a las propiedades del suelo.

1.7.2. Los objetivos que se plantea este trabajo son:

i. Comparar la eficiencia del diseño experimental y la exactitud en las estimaciones en experimentos de gran tamaño (es decir, ensayos de cultivares con un gran número de tratamientos) con diferentes grados de control local de la variabilidad de campo y con diferentes estrategias de modelado espacial en heterogeneidad espacial natural.

ii. Evaluar el comportamiento de diferentes modelos espaciales en diferentes escenarios (varias combinaciones de tratamientos y repeticiones) y determinar si la selección de un modelo espacial adecuado a las condiciones de variabilidad espacial natural permite obtener precisión al estimar los efectos de los tratamientos con poco o ningún sesgo en las estimaciones del error estándar de las diferencias de medias y buen control de la tasa de error de tipo 1.

iii. Comparar diferentes estrategias de análisis que incorporan la información espacial y/o temporal de la parcela para estimar variables de crecimiento en experimentos forestales a largo plazo con parcelas de gran tamaño, en términos de ajuste del modelo y precisión en las comparaciones de medias de tratamiento.

iv. Evaluar si el uso de modelos con corrección espacial mejora la eficiencia del diseño experimental y si las pasturas naturales mejoradas inducen cambios en la distribución espacial en superficie y en profundidad del carbono orgánico del suelo y del fósforo en comparación con campo natural.

1.7.3. Estructura general de la tesis

El **capítulo 1** presenta la justificación general del trabajo realizado, describiendo las principales características de la experimentación agronómica —en particular, en ensayos grandes y sostenidos en el tiempo— y los principales desafíos a la hora de diseñar y analizar esos experimentos. Se presentan los objetivos e hipótesis del trabajo.

El **capítulo 2** presenta el artículo publicado en la revista *Crop Science* titulado “Can spatial modeling substitute experimental design in agricultural experiments?”, donde se evalúa si la modelización espacial puede sustituir al diseño experimental o si su ventaja es marginal en comparación con un diseño experimental con adecuado control local. Se corresponde con el objetivo 1 de este proyecto.

En el **capítulo 3** se discute si la ganancia de precisión esperada con modelos espaciales adecuados puede lograrse sin sesgo en los errores estándar estimados y si se pueden alcanzar inferencias estadísticas válidas, logrando un control adecuado de la tasa empírica de error de tipo 1. Además, se discute cómo afecta el número de repeticiones por tratamiento y la inclusión o no de los bloques completos al sesgo y la tasa empírica de error de tipo 1. Se corresponde con el objetivo 2 de este proyecto.

El **capítulo 4** presenta el segundo artículo de la tesis publicado en *Forest Science* titulado "Spatio-temporal modelling and competition dynamics in forestry tillage experiments on early growth of Eucalyptus grandis", donde se comparan modelos espaciales, temporales y espacio-temporales en ensayos forestales. Este artículo se elaboró en igualdad de contribución con Pablo González Barrios. Se corresponde con el objetivo 3 de este proyecto.

El **capítulo 5** discute si modelar la correlación espacial dentro de parcelas de gran tamaño mejora el rendimiento del diseño experimental y si las pasturas mejoradas con leguminosas y fertilización con fósforo afectan el contenido medio de carbono orgánico y el fósforo del suelo, su variabilidad y su distribución espacial. Se corresponde con el objetivo 4 de este proyecto.

Finalmente, en el **capítulo 6** se discuten los principales resultados y aprendizajes obtenidos en la investigación y su alcance. Se plantean nuevas interrogantes emergentes del proceso.

2. CAN SPATIAL MODELING SUBSTITUTE FOR EXPERIMENTAL DESIGN IN AGRICULTURAL EXPERIMENTS?[1]

Alejandra Borges, Agustín González-Reymundez, Oswaldo Ernst, Mónica Cadenazzi, José Terra, and Lucía Gutiérrez

A. Borges, A. González†, M. Cadenazzi and L. Gutiérrez, Statistics Department, Facultad de Agronomía, Universidad de la República, Garzón 780, Montevideo, Uruguay. O. Ernst, Department of Crop Production, Facultad de Agronomía, Universidad de la República, Estación Experimental Mario Cassinoni, Paysandú, Uruguay. J. Terra, Estación Experimental Treinta y Tres, Instituto Nacional de Investigación Agropecuaria (INIA), Ruta 8, Km 281, Treinta y Tres 33000, Uruguay. †A. González-Reymundez Current address: Department of Epidemiology and Biostatistics, Michigan State University. 909 Fee Rd, B601 West Fee Hall, East Lansing, 48823 MI, USA. L. Gutierrez, Department of Agronomy, University of Wisconsin at Madison, 1575 Linden Dr., Madison, WI 53706, USA.

2.1. RESUMEN

Uno de los aspectos más críticos de la experimentación agrícola es la elección adecuada del diseño experimental para controlar la heterogeneidad del campo, especialmente en el caso de los experimentos grandes. Sin embargo, incluso con diseños experimentales complejos, es posible que la variabilidad espacial no se controle adecuadamente si se produce a escalas inferiores a la de los bloques. Por lo tanto, la modelización de la variabilidad espacial puede ser beneficiosa, y algunos estudios incluso proponen la modelización espacial en lugar del diseño experimental. Nuestro objetivo fue evaluar los efectos del diseño experimental, de la modelización espacial y de una combinación de ambos en condiciones reales de campo utilizando SIG y simulando experimentos. Los datos de rendimiento de los cultivares se

[1] Borges, A., González-Reymundez, A., Ernst, O., Cadenazzi, M., Terra, J., and Gutiérrez, L. (2019). Can spatial modeling substitute for experimental design in agricultural experiments? *Crop Science*, *59*(1). https://doi.org/10.2135/cropsci2018.03.0177

simularon utilizando la variabilidad espacial real de un gran ensayo de uniformidad con 100 localidades independientes y diferentes tamaños de experimentos para cuatro diseños experimentales: diseño completamente aleatorizado (CRD), diseño de bloques completos aleatorizados (RCBD), diseño de bloques incompletos α-lattice (ALPHA) y diseño parcialmente replicado (PREP). Cada realización se analizó utilizando diferentes niveles de corrección espacial. Los modelos se compararon según la precisión, la exactitud y la recuperación de genotipos superiores. Para tamaños de experimento moderados y grandes, ALPHA fue el mejor diseño experimental en términos de precisión y exactitud. En la mayoría de las situaciones, los modelos que incluían correlación espacial fueron mejores que los modelos sin correlación espacial, pero no superaron a los mejores diseños experimentales. Por tanto, la modelización espacial no sustituye a un buen diseño experimental.

Palabras clave: variabilidad espacial, ensayo de uniformidad, diseños experimentales, modelos espaciales, indicadores de eficiencia

Published online December 13, 2018

RESEARCH

Can Spatial Modeling Substitute for Experimental Design in Agricultural Experiments?

Alejandra Borges, Agustín González-Reymundez, Oswaldo Ernst, Mónica Cadenazzi, José Terra, and Lucía Gutiérrez*

ABSTRACT

One of the most critical aspects of agricultural experimentation is the proper choice of experimental design to control field heterogeneity, especially for large experiments. However, even with complex experimental designs, spatial variability may not be properly controlled if it occurs at scales smaller than blocks. Therefore, modeling spatial variability can be beneficial, and some studies even propose spatial modeling instead of experimental design. Our goal was to evaluate the effects of experimental design, spatial modeling, and a combination of both under real field conditions using GIS and simulating experiments. Yield data from cultivars was simulated using real spatial variability from a large uniformity trial of 100 independent locations and different sizes of experiments for four experimental designs: completely randomized design (CRD), randomized complete block design (RCBD), α-lattice incomplete block design (ALPHA), and partially replicated design (PREP). Each realization was analyzed using different levels of spatial correction. Models were compared by precision, accuracy, and the recovery of superior genotypes. For moderate and large experiment sizes, ALPHA was the best experimental design in terms of precision and accuracy. In most situations, models that included spatial correlation were better than models with no spatial correlation, but they did not outperformed better experimental designs. Therefore, spatial modeling is not a substitute for good experimental design.

A. Borges, A. González-Reymundez, M. Cadenazzi, and L. Gutiérrez, Statistics Dep., Facultad de Agronomía, Univ. de la República, Garzón 780, Montevideo, Uruguay; A. González-Reymundez, current address, Dep. of Epidemiology and Biostatistics, Michigan State Univ. 909 Fee Rd., B601 West Fee Hall, East Lansing, 48823 MI; O. Ernst, Dep. of Crop Production, Facultad de Agronomía, Univ. de la República, Estación Experimental Mario Cassinoni, Paysandú, Uruguay; J. Terra, Estación Experimental Treinta y Tres, Instituto Nacional de Investigación Agropecuaria (INIA), Ruta 8, Km 281, Treinta y Tres 33000, Uruguay; L. Gutiérrez, Dep. of Agronomy, Univ. of Wisconsin–Madison, 1575 Linden Dr., Madison, WI 53706. Received 13 Mar. 2018. Accepted 3 Oct. 2018. *Corresponding author (gutierrezcha@wisc.edu). Assigned to Associate Editor Jode Edwards.

Abbreviations: ALPHA, α-lattice incomplete block experimental design; AR(1), spatially correlated error model with one-dimensional autoregressive process; Best_Gen, proportion of times the true 15% superior genotypes are recovered; COR, Pearson's correlation coefficient between true and estimated effects; CRD, completely randomized experimental design; EXP(2), spatially correlated error model with two-dimensional exponential process; MSEP, mean square error of prediction; NSC, no spatial correction model; PREP, partially replicated experimental design; $PREP_g$, partially replicated experimental design with fixed number of genotypes; $PREP_u$, partially replicated experimental design with fixed number of experimental units; RCBD, randomized complete block experimental design; YSD, yield standard deviation.

One of the most important aspects in any experiment in agricultural research is the proper choice of the experimental design and analysis model (Casler, 2015; Piepho et al., 2015). Well-designed experiments are based on the three principles proposed by R.A. Fisher (1935): randomization, replication, and local control. After Fisher, agricultural research has been based on these three principles that seek to control local

Published in Crop Sci. 58:1–10 (2019).
doi: 10.2135/cropsci2018.03.0177

variation, assuming independence between experimental units, and using replications to estimate experimental error and increase precision of mean estimates. On the other hand, spatial variation in environmental and soil factors is common under field conditions, so spatial autocorrelation is generally present in field experiments (Legendre, 1993; Grondona et al., 1996). Therefore, experiments can be designed to have local control of this spatial variability.

In agriculture, the most commonly used experimental designs are the completely randomized design (CRD) and the randomized complete block design (RCBD) (Piepho et al., 2015). In some cases, blocking is an effective way to control experimental error (Cochran and Cox, 1957; Casler, 2015), but it is not enough in agricultural situations where field heterogeneity and/or the size of the experiment is large (Brownie et al., 1993). Furthermore, designs including blocks without considering the real spatial variation among experimental units can strongly decrease the success of an experiment (Casler, 2015).

Some designs such as α-lattice incomplete block (ALPHA) experimental designs (Patterson and Williams, 1976) or row–column designs (Williams et al., 2006) were specifically created to control field heterogeneity in situations where a large number of treatments are evaluated, as well as other complex treatment structures (Williams et al., 2002). These designs are therefore better suited for a large number of treatments where spatial variability is high (Müller et al., 2010). However, they are not always used due to their complexity and practical difficulties, or simply because researchers are not used to choosing them (Casler, 2015). In cultivar evaluation trials, where the number of treatments is always high and there is an increasing need to distinguish cultivars by smaller differences in terms of yield (Casler and Undersander, 2000), better experimental designs and analysis models are needed to control spatial variation.

Spatial variability often occurs gradually, and sometimes it is not captured well enough by the experimental design (Grondona and Cressie, 1991). Even with strong local control, as in incomplete block designs, spatial variability may not be properly controlled if it occurs at smaller scales than the size of the sub-blocks (Grondona et al., 1996). Explicitly accounting for spatial variability can be achieved in various ways such as with different structures of correlation of the R matrix, trend analysis, or a combinations of both (Brownie et al., 1993; Casler and Undersander, 2000). Additionally, one- or two-dimensional spatial analysis can be implemented according to natural field heterogeneity (Gleeson and Cullis, 1987; Cullis and Gleeson, 1991; Gilmour et al., 1997; Qiao et al., 2000).

Whether it is better to control with experimental design or spatial analysis models is often debated. Müller et al. (2010) found that in barley (*Hordeum vulgare* L.) and sugar beet (*Beta vulgaris* L.) trials, the baseline model, which only includes a block and a replicate effect, showed the best fit according to Akaike information criterion for most cases. Spatial analyses were not necessarily better than the RCBD in all possible circumstances, particularly in cases where the spatial structure of the studied variable cannot be accurately characterized (Kravchenko et al., 2006). Uniformity trials have been conducted to determine the optimal experimental design in many cases. These trials were expensive to carry out, and complex designs were not evaluated (Koch and Rigney, 1951). Studies that compare different designs and analysis models using the same database with real field spatial variability have not been reported to our knowledge. Our goal was to evaluate the effect of experimental design, spatial modeling, and a combination of both under real field conditions using GIS information and simulating experiments. One of the questions of interest was whether spatial modeling can substitute for experimental design, or whether its advantage is marginal compared with a proper experimental design.

MATERIALS AND METHODS

General Approach

Yield data from cultivars were simulated using real field variability, genotypic effects, and a number of experimental designs in different locations within a large field. Each realization was analyzed using a series of models with different levels of correction for spatial variability. Evaluation criteria for comparing designs and analysis models included precision, accuracy, and the recovery of superior genotypes.

Wheat Yield Data and Spatial Variability from Yield Monitors

A field of ~64 ha was sown with the wheat (*Triticum aestivum* L.) cultivar "Nogal" (USDA-ARS, 1992) on 20 June 2008 at a density of 120 kg ha^{-1} of seed. The field was harvested in rectangular plots of ~15-m × 5-m area to obtain 1445 yield plots (kg ha^{-1}). Each plot has geographic coordinates X (east–west) and Y (north–south) and elevation (m). An empiric variogram was calculated with the values from the yield monitor and a Matern variogram ($\kappa = 1$) was fitted. Due to the original inclination of the plots with respect to the X and Y coordinates, kriged values were used as the baseline field heterogeneity. Yield maps were created using the 'sp' package (Pebesma and Bivand, 2005) of R statistical software (R Core Team, 2016).

Genotypic Effects

Three experiment sizes were evaluated: (i) small size with 45 experimental units, (ii) moderate size with 150 experimental units, and (iii) large size with 600 experimental units. For the small experiment size, 15 genotypic effects for the simulation process were obtained from yield mean data of 15 wheat cultivars from the National Cultivar Trial Networks from 3-yr evaluation trials (2007–2009) (INIA, 2017). For moderate and large experiment size, genotypic effects (G_i) were simulated from a Gaussian distribution, assuming that $G_i \sim N(0, \sigma^2_G)$, where σ^2_G was the genotypic variance of

the 15 genotypes obtained from the National Cultivar Trial Network. For moderate size, 50 genotypic effects were simulated for CRD, RCBD, and ALPHA designs, whereas 200 genotypic effects were simulated for large-size experiments. We used two strategies for partially replicated experimental (PREP) designs: (i) fixing the number of experimental units (i.e., 150 and 600 experimental units for the moderate and large size respectively) so that 108 and 428 genotypic effects were evaluated and simulated for the medium and large experiment size, respectively; or (ii) fixing the number of genotypes evaluated. Therefore, 50 and 200 genotypes were evaluated and simulated for the moderate and large experiment sizes, respectively, using fewer experimental units than the other experimental designs (i.e., 70 experimental units for the medium-size experiment instead of 150, and 280 experimental units for the large-size experiment instead of 600), as described below.

Experimental Designs

The genotypes (treatments) were assigned to experimental units (plots of ~15-m × 5-m area) in one of four experimental designs: CRD, RCBD, ALPHA, and PREP.

The model used for the CRD experimental design with three replications was:

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

where Y_{ij} is the observed yield for the ith genotype in the jth replicate, μ is the overall mean, G_i is the ith genotypic effect, and ε_{ij} are the residual errors associated with the observation Y_{ij}, $\varepsilon_{ij} \sim N(0, \sigma_e^2)$, where σ_e^2 is the error variance, and the covariance $COV(\varepsilon_{ij}, \varepsilon_{ij'})$ was modeled according to the spatial corrections described below.

The model used for the RCBD with three replications was:

$$Y_{ij} = \mu + G_i + \beta_j + \varepsilon_{ij}$$

where Y_{ij} is the observed yield for the ith genotype in the jth block, and β_j is the jth block effect. The blocks were considered fixed effects and were located in the east–west direction.

An ALPHA design with three complete replications and a number of incomplete blocks or sub-block was used. The number of experimental units per incomplete block was three for the small size, five for the moderate size, and five [$ALPHA_{(i=3)}$] or ten [$ALPHA_{(i=10)}$] for the large experiment size. The model used for the ALPHA experiment with three full replicates was:

$$Y_{ijk} = \mu + G_i + \beta_j + \gamma_{k(j)} + \varepsilon_{ijk}$$

where Y_{ijk} is the observed yield for the ith genotype, jth replication, and kth incomplete block, β_j is the jth complete replication effect, and $\gamma_{k(j)}$ is the kth incomplete block effect nested on jth replication, and ε_{ijk} are the residual errors associated with the observation Y_{ijk}, $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$, and the covariance $COV(\varepsilon_{ijk}, \varepsilon_{ijk'})$ was modeled according to the spatial corrections described below. Incomplete blocks were considered random factors nested in each complete replication, assuming $\gamma_{k(j)} \sim N(0, \sigma_\gamma^2)$, where σ_γ^2 is the variance of the sub-blocks. Complete replications were located in the east–west direction, as in the RCBD design, and incomplete blocks were orthogonal to the complete replications.

A PREP design where repeated genotypes were randomized in a RCBD design with three replications was used. The model used for the PREP was:

$$Y_{ij} = \mu + G_i + \beta_j + \varepsilon_{ij}$$

where Y_{ij} is the observed yield of the ith genotype and jth replication. The model for G_i was $G_i = g_i + t_i$, where g_i is the effect of the ith nonreplicated genotypic effect with $i = 1, \ldots, n_g$ (where n_g is the number of nonreplicated genotypes), and t_i is the ith replicated genotypic effect (test line) with $i = n_g + 1, \ldots, n_g + n_c$ (where n_c is the number of replicated genotypes).

In the case that the number of experimental units was fixed ($PREP_u$), 20% of the genotypes were replicated, whereas the remaining genotypes were unreplicated. In the case that the number of genotypes was fixed ($PREP_g$), 15% of the genotypes were replicated, containing either 70 or 280 plots for moderate and large experiment size respectively (47% of the number of plots in the CRD, RCBD, and ALPHA designs).

Locations

One hundred randomly selected locations within the field were used to conduct all simulations. The selected location was used as the upper left corner of each experimental design, and the shape was always rectangular with longer east–west dimension for all locations to follow the direction of smallest variance. The shape of all locations was the same for a given experimental size. For each experimental size, the same experimental units were used for all experimental designs, except for $PREP_g$, where fewer experimental units were used. Each location was characterized by their yield SD (YSD), expressed as tons per hectare. For each experimental size, the mean YSD of all 100 locations was calculated. Locations were then classified as either with high or low variability (Fig. 1) according to this criterion: locations with YSD lower than the mean YSD were considered as low-variability locations, and those with greater YSD than the mean YSD were considered as high-variability locations. Results are reported as averages for all high- and low-variability locations for each experimental size.

Simulation Procedure

Yield was obtained for each plot according to the procedure below. First, treatments were assigned to plots according to one of the three experimental designs described above. Second, yield of each plot was obtained using the equation:

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

where Y_{ij} is the yield plot data simulated corresponding to the ith cultivar and the jth replication, G_i is the ith genotypic effect corresponding to the randomly assigned treatment to the plot, ε_{ij} is the field experimental error that represents the spatial heterogeneity of the field and was obtained from the yield monitor, and δ_{ij} is a repeatability error. To avoid using a deterministic model, the δ_{ij} values were assumed as independent random variables with $\delta_{ij} \sim N(0, \sigma_r^2)$, where σ_r^2 is a random noise or repeatability. We conducted the simulations using two values for σ_r^2. First, we used a value of $\sigma_r^2 = 0.07$ that represents 5% of the total field heterogeneity and a yield heritability of ~0.5 when a simple experimental design is used. The second value, $\sigma_r^2 = 0.2$, targeted a lower yield heritability.

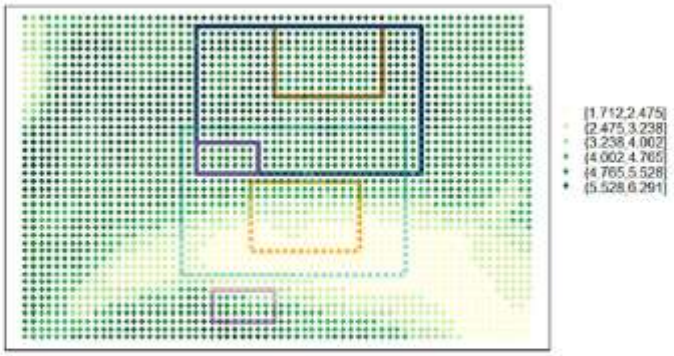


Fig. 1. Wheat yield map (tons ha^{-1}), according to east–west oriented coordinates (*Y* coordinates) and north–south oriented coordinates (*X* coordinates). Darker colors indicate higher yield values. Localization of the experimental units for two contrasting locations and three experimental design sizes are shown: for small experiment size (15 genotypes), light violet lines represent high-variability location and dark violet lines represent for low-variability ones; for moderate experiment size (50 genotypes), light orange lines represent high-variability locations and dark orange lines represent low-variability ones; for large experiment size (200 genotypes), light blue lines represent high-variability locations and dark blue lines represent low-variability locations. Two locations out of 100 field locations that were evaluated for each experimental design size are shown.

For each experimental design and location, 1000 simulations were run, performing an independent randomization for each simulation. The 'agricolae' package (de Mendiburu, 2012) of R (R Core Team, 2016) was used for experimental design randomization. The simulations and statistical analysis of the data were performed with the 'nlme' package (Pinheiro et al., 2013) and personal code of R (R Core Team, 2016).

Analysis Models

Each vector of phenotypic yield was analyzed according to the following three models:

1. No spatial correction (NSC), where experimental errors are assumed independent (uncorrelated), $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$, where $COV(\varepsilon_{ij}, \varepsilon_{ij'}) = 0 \; \forall \; \varepsilon_{ij} \neq \varepsilon_{ij'}$.
2. A spatially correlated error model with one-dimensional autoregressive process [AR(1)], $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$, where $COV(\varepsilon_{ij}, \varepsilon_{ij'}) = \sigma^2_\varepsilon \rho^k$. The correlation function corresponding to an autoregressive model of order 1 decreases in absolute value with every unit of distance within columns: $h(k, \rho) = \rho^k$, $k = 0, 1, \ldots$, where ρ is the correlation parameter to be estimated, and k is the distance unit between rows (i.e., the direction of maximum variance in our uniformity trial).
3. Spatially correlated error model with two-dimensional exponential spatial correlation structure [EXP(2)] , $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$, where $COV(\varepsilon_{ij}, \varepsilon_{ij'})$ was modeled according to an isotropic exponential model. With d being the range, the correlation between two observations at r distance was $\exp(-r/d)$. The EXP(2) model was fitted for rows and columns with and without a nugget variance.

Estimation Method and Evaluation Criteria for Model Comparison

The residual maximum likelihood method was used to estimate parameters. Models were compared by precision and accuracy statistics and by their ability to recover superior genotypes. For all statistics, we calculated the mean and the SD over the 1000 realizations of the simulation for each experiment size, type of location, experimental design, and spatial correction.

The SE of the difference between cultivar means (SED) was used as a precision statistic to compare models. The lower the value, the better the precision.

The recovery of the best genotypes (Best_Gen), the Pearson's correlation coefficient between true and estimated genotypic effects (COR), and the mean square error of prediction (MSEP) were used as an accuracy statistic to compare models. The Best_Gen was calculated as the proportion of times 15% of the true superior genotypes were recovered. The MSEP was calculated according to Gauch et al. (2003) as follows:

$$MSEP = \Sigma(X_n - Y_n)2/N$$

where X_n and Y_n are the model-based and true genotypic values, and N is the number of genotypes, where summation is over $n = 1, 2, \ldots, N$.

RESULTS

The experimental designs and spatial correction models performed similarly for the small experiment size in both low- and high-variability locations, and with low and high yield heritability (Supplemental Tables S1 and S2).

For moderate experiment size in high-variability conditions, the ALPHA design had the best performance for most statistics (Table 1), although RCBD with AR(1) performed similar to the ALPHA with NSC. For CRD and RCBD, only the AR(1) model was better than the NSC (Table 1). Therefore, there is some performance compensation by

spatial modeling, in the case of AR(1). The PREP had a poor performance over all designs, and spatial corrections did not improve the performance. Even with $PREP_n$ with the same number of experimental units as the others experimental designs, performance was poorer.

For moderate experiment size with low variability, no differences in the performance of experimental design or spatial corrections were found among CRD, RCBD, and ALPHA for most statistics (Table 1). The PREP design in these circumstances underperformed for all statistics in both cases ($PREP_g$ and $PREP_n$, Table 1).

For large experiment size under high variability, the results were similar to those of the moderate size, but the advantages of a more complex experimental design were more noticeable. The ALPHA design outperformed all other experimental designs, especially when it was combined with AR(1) or EXP(2) using either 5 or 10 experimental units per incomplete block (Table 2, Fig. 2). The AR(1) spatial correction improved model performance of CRD and RCBD but did not outperform experimental design in general (Table 2), although RCBD with AR(1) obtained similar values to $ALPHA_{(r=10)}$ with no spatial correction (Fig. 2). For this size of experiment, the performance of the $PREP_g$ was closer to the simple designs for most of the statistics (Table 2).

For large experiment size with low variability, the differences among experimental designs and spatial corrections were small, although larger than for moderate size designs. The $ALPHA_{(r=10)}$ experimental design with some spatial correction (Table 2, Fig. 2) obtained the best

Table 1. Mean and SD (in parentheses) of the recovery of 15% of superior genotypes (Best_Gen), the SE for mean differences (SED), the Pearson's correlation coefficient between observed and estimated genotypic effects (COR), and the mean square error of prediction (MSEP), for four experimental designs and three spatial corrections in two types of field locations (high and low spatial heterogeneity). This table shows results for the moderate experiment size (50 genotypes) and high heritability.

Variability	YSD†	Locations	Design‡	Model§	Best_Gen	SED	COR	MSEP
	tons ha^{-1}	no.				tons ha^{-1}		tons ha^{-1}
High	0.80	46	CRD	NSC	0.45 (0.15)	0.49 (0.02)	0.63 (0.08)	0.25 (0.04)
			CRD	AR(1)	0.49 (0.15)	0.38 (0.03)	0.68 (0.07)	0.18 (0.03)
			CRD	EXP(2)	0.45 (0.15)	0.49 (0.02)	0.64 (0.07)	0.29 (0.06)
			RCBD	NSC	0.46 (0.15)	0.47 (0.02)	0.65 (0.07)	0.32 (0.04)
			RCBD	AR(1)	0.52 (0.15)	0.33 (0.03)	0.72 (0.06)	0.24 (0.03)
			RCBD	EXP(2)	0.46 (0.15)	0.47 (0.02)	0.65 (0.07)	0.34 (0.06)
			ALPHA	NSC	0.55 (0.14)	0.36 (0.02)	0.75 (0.05)	0.21 (0.03)
			ALPHA	AR(1)	0.64 (0.13)	0.29 (0.01)	0.85 (0.04)	0.16 (0.03)
			ALPHA	EXP(2)	0.64 (0.13)	0.29 (0.01)	0.85 (0.04)	0.16 (0.03)
			$PREP_g$	NSC	0.38 (0.15)	0.60 (0.10)	0.53 (0.10)	0.52 (0.16)
			$PREP_g$	AR(1)	0.39 (0.15)	0.56 (0.09)	0.55 (0.10)	0.49 (0.18)
			$PREP_g$	EXP(2)	0.40 (0.15)	0.57 (0.09)	0.56 (0.10)	0.47 (0.15)
			$PREP_n$	NSC	0.25 (0.14)	0.62 (0.07)	0.48 (0.07)	0.54 (0.09)
			$PREP_n$	AR(1)	0.26 (0.14)	0.58 (0.06)	0.50 (0.07)	0.49 (0.10)
			$PREP_n$	EXP(2)	0.26 (0.14)	0.58 (0.06)	0.51 (0.07)	0.48 (0.09)
Low	0.25	53	CRD	NSC	0.68 (0.13)	0.21 (0.01)	0.88 (0.03)	0.05 (0.01)
			CRD	AR(1)	0.68 (0.13)	0.21 (0.01)	0.88 (0.03)	0.05 (0.01)
			CRD	EXP(2)	0.69 (0.13)	0.21 (0.01)	0.88 (0.02)	0.05 (0.01)
			RCBD	NSC	0.69 (0.12)	0.20 (0.01)	0.89 (0.02)	0.06 (0.01)
			RCBD	AR(1)	0.69 (0.13)	0.19 (0.01)	0.89 (0.02)	0.06 (0.01)
			RCBD	EXP(2)	0.69 (0.12)	0.21 (0.27)	0.89 (0.02)	0.06 (0.01)
			ALPHA	NSC	0.70 (0.12)	0.19 (0.01)	0.90 (0.02)	0.06 (0.01)
			ALPHA	AR(1)	0.71 (0.12)	0.18 (0.01)	0.90 (0.02)	0.06 (0.01)
			ALPHA	EXP(2)	0.71 (0.12)	0.18 (0.01)	0.90 (0.02)	0.06 (0.01)
			$PREP_g$	NSC	0.56 (0.14)	0.30 (0.05)	0.76 (0.05)	0.13 (0.04)
			$PREP_g$	AR(1)	0.55 (0.14)	0.29 (0.05)	0.75 (0.06)	0.14 (0.04)
			$PREP_g$	EXP(2)	0.56 (0.14)	0.29 (0.05)	0.76 (0.05)	0.13 (0.04)
			$PREP_n$	NSC	0.45 (0.15)	0.31 (0.04)	0.73 (0.04)	0.13 (0.02)
			$PREP_n$	AR(1)	0.45 (0.15)	0.30 (0.04)	0.73 (0.04)	0.13 (0.03)
			$PREP_n$	EXP(2)	0.46 (0.15)	0.30 (0.04)	0.74 (0.04)	0.13 (0.02)

† YSD, yield SD.

‡ CRD, completely randomized design; RCBD, randomized complete block design; ALPHA, incomplete blocks and α-lattice design; $PREP_g$, partially replicated design with 50 genotypes (this experiment preserved the number of genotypes and thus used fewer experimental units); $PREP_n$, partially replicated design with 108 genotypes (this experiment preserved the number of experimental units and thus evaluated more genotypes).

§ NSC, no spatial correction model; AR(1), spatial correlated error model with one-dimensional autoregressive process; EXP(2), spatial correlated error model with two-dimensional exponential spatial correlation structure without a nugget variance.

results. The PREP design had no advantages in these conditions either.

Similar relative performance of experimental designs and spatial corrections was observed when a low heritability was simulated for both, medium (Supplemental Table S3) and large experiment sizes (Supplemental Table S4). In both cases, all the experiments and spatial corrections performed poorer than with high heritability.

The use of a nugget variance in the EXP(2) spatial correction increased the mean and the variability in the mean square error estimation of the CRD and RCBD experiments and did not improve any other statistic of any experimental design (data not shown); therefore, it is not further discussed.

DISCUSSION

Experiment size and field heterogeneity clearly affected the performance of the experimental designs and spatial correction. With large experiment sizes under high-variability conditions, the choice of the experimental design is essential to obtain good results in terms of precision and accuracy in the estimation of genotypic effects.

The ALPHA design had the best performance across design and spatial correction. This design has been shown

Table 2. Mean and SD (in parentheses) of the recovery of 15% of superior genotypes (Best_Gen), the SE for mean differences (SED), the Pearson's correlation coefficient between observed and estimated genotypic effects (COR), and the mean square error of prediction (MSEP) for four experimental designs and three spatial corrections in two types of field locations (high and low spatial heterogeneity). This table shows results for the large experiment size (200 genotypes) and high heritability.

Variability	YSD†	Locations	Design‡	Model§	Best_Gen	SED	COR	MSEP
	tons ha^{-1}	no.				tons ha^{-1}		tons ha^{-1}
High	1.10	58	CRD	NSC	0.40 (0.07)	0.66 (0.01)	0.51 (0.05)	0.43 (0.04)
			CRD	AR(1)	0.45 (0.07)	0.43 (0.02)	0.60 (0.04)	0.27 (0.02)
			CRD	EXP(2)	0.40 (0.07)	0.65 (0.01)	0.51 (0.05)	0.43 (0.04)
			RCBD	NSC	0.40 (0.07)	0.64 (0.01)	0.52 (0.05)	0.52 (0.04)
			RCBD	AR(1)	0.50 (0.07)	0.36 (0.02)	0.67 (0.03)	0.31 (0.02)
			RCBD	EXP(2)	0.40 (0.07)	0.64 (0.01)	0.52 (0.05)	0.52 (0.04)
			$ALPHA_{k=5}$	NSC	0.52 (0.07)	0.35 (0.01)	0.71 (0.03)	0.23 (0.02)
			$ALPHA_{k=5}$	AR(1)	0.60 (0.07)	0.28 (0.01)	0.80 (0.03)	0.18 (0.02)
			$ALPHA_{k=5}$	EXP(2)	0.60 (0.07)	0.28 (0.01)	0.80 (0.03)	0.18 (0.02)
			$ALPHA_{k=10}$	NSC	0.47 (0.07)	0.49 (0.01)	0.63 (0.04)	0.35 (0.02)
			$ALPHA_{k=10}$	AR(1)	0.68 (0.06)	0.27 (0.01)	0.87 (0.02)	0.19 (0.02)
			$ALPHA_{k=10}$	EXP(2)	0.68 (0.06)	0.27 (0.01)	0.87 (0.02)	0.19 (0.02)
			$PREP_g$	NSC	0.39 (0.07)	0.76 (0.06)	0.46 (0.05)	0.65 (0.07)
			$PREP_g$	AR(1)	0.44 (0.07)	0.59 (0.04)	0.57 (0.05)	0.36 (0.07)
			$PREP_g$	EXP(2)	0.44 (0.07)	0.59 (0.04)	0.57 (0.05)	0.36 (0.07)
			$PREP_n$	NSC	0.26 (0.07)	0.81 (0.04)	0.40 (0.04)	0.80 (0.06)
			$PREP_n$	AR(1)	0.31 (0.07)	0.60 (0.02)	0.52 (0.04)	0.47 (0.06)
			$PREP_n$	EXP(2)	0.31 (0.07)	0.60 (0.02)	0.52 (0.04)	0.47 (0.06)
Low	0.46	35	CRD	NSC	0.63 (0.06)	0.31 (0.01)	0.78 (0.02)	0.10 (0.01)
			CRD	AR(1)	0.63 (0.06)	0.28 (0.01)	0.79 (0.02)	0.10 (0.01)
			CRD	EXP(2)	0.63 (0.06)	0.30 (0.01)	0.79 (0.02)	0.10 (0.01)
			RCBD	NSC	0.63 (0.06)	0.30 (0.01)	0.79 (0.02)	0.11 (0.01)
			RCBD	AR(1)	0.63 (0.06)	0.27 (0.01)	0.80 (0.02)	0.10 (0.01)
			RCBD	EXP(2)	0.63 (0.06)	0.30 (0.01)	0.79 (0.02)	0.11 (0.01)
			$ALPHA_{k=5}$	NSC	0.64 (0.06)	0.24 (0.01)	0.83 (0.02)	0.07 (0.01)
			$ALPHA_{k=5}$	AR(1)	0.66 (0.06)	0.21 (0.01)	0.85 (0.02)	0.06 (0.01)
			$ALPHA_{k=5}$	EXP(2)	0.66 (0.06)	0.21 (0.01)	0.85 (0.02)	0.06 (0.01)
			$ALPHA_{k=10}$	NSC	0.66 (0.06)	0.27 (0.01)	0.83 (0.02)	0.09 (0.01)
			$ALPHA_{k=10}$	AR(1)	0.71 (0.06)	0.20 (0.01)	0.88 (0.01)	0.05 (0.01)
			$ALPHA_{k=10}$	EXP(2)	0.71 (0.06)	0.20 (0.01)	0.88 (0.01)	0.05 (0.01)
			$PREP_g$	NSC	0.55 (0.07)	0.41 (0.04)	0.68 (0.03)	0.20 (0.02)
			$PREP_g$	AR(1)	0.56 (07)	0.37 (0.03)	0.71 (0.04)	0.17 (0.03)
			$PREP_g$	EXP(2)	0.56 (0.07)	0.37 (0.03)	0.71 (0.04)	0.17 (0.03)
			$PREP_n$	NSC	0.45 (0.07)	0.44 (0.03)	0.63 (0.03)	0.22 (0.02)
			$PREP_n$	AR(1)	0.46 (0.07)	0.38 (0.02)	0.66 (0.03)	0.18 (0.02)
			$PREP_n$	EXP(2)	0.46 (0.07)	0.38 (0.02)	0.66 (0.03)	0.18 (0.02)

† YSD, yield SD.

‡ CRD, completely randomized design; RCBD, randomized complete block design; $ALPHA_{k=5}$, incomplete blocks and α-lattice design with a block size of 5; $ALPHA_{k=10}$, incomplete blocks and α-lattice design with a block size of 10; $PREP_g$, partially replicated design with 200 genotypes (this experiment preserved the number of genotypes and thus used fewer experimental units); $PREP_n$, partially replicated design with 428 genotypes (this experiment preserved the number of experimental units and thus evaluated more genotypes).

§ NSC, no spatial correction model; AR(1), spatial correlated error model with one-dimensional autoregressive process; EXP(2), spatial correlated error model with two-dimensional exponential spatial correlation structure without a nugget variance.

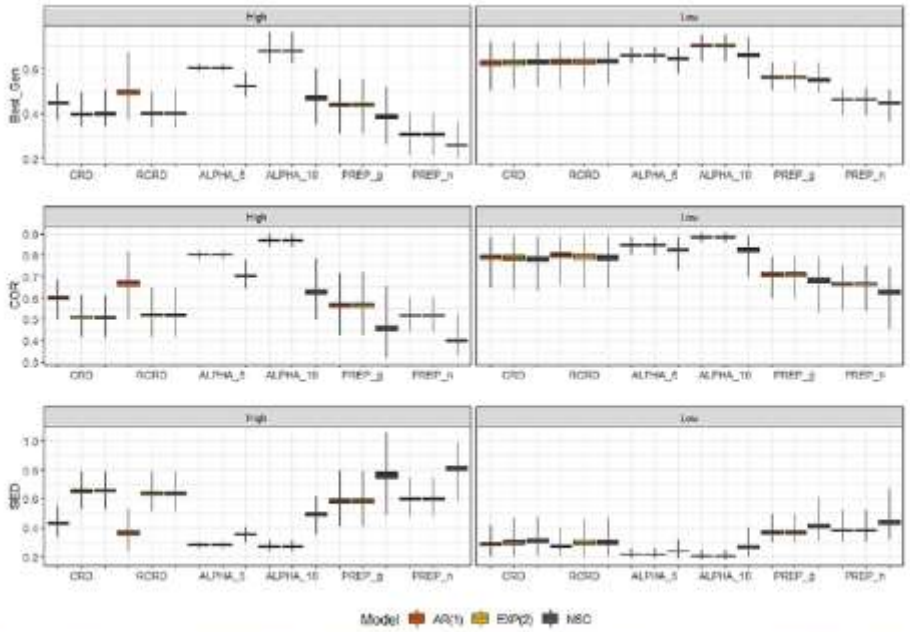


Fig. 2. Recovery of the best genotypes (Best_Gen), Pearson's correlation coefficient between true and estimated genotypic effects (COR), and the SE of the difference between cultivar means (SED) for six experimental designs: completely randomized design (CRD), randomized complete block design (RCBD) incomplete blocks and α-lattice design with block size of 5 (ALPHA_5), incomplete blocks and α-lattice design with block size of 10 (ALPHA_10), partially replicated design with 200 genotypes ($PREP_p$, this experiment preserved the number of genotypes and thus used fewer experimental units), and a partially replicated design with 428 genotypes (PREPn, this experiment preserved the number of experimental units and thus evaluated more genotypes). Each experimental design was analyzed with three spatial models: no spatial correction model (NSC), spatial correlated error model with one-dimensional autoregressive process [AR(1)], spatial correlated error model with two-dimensional exponential spatial correlation structure [EXP(2)], in high- (left) and low-variability (right) locations.

as superior to RCBD (Masood et al., 2008; Gonçalves et al., 2010) due to stronger local control. The experimental units' layout and blocking orientation in our study might have favored CRD designs over all other experimental designs, with RCBD being the most unfavorable design. Experimental designs that exerted some local control on the spatial variability, like ALPHA or row–column designs (Müller et al., 2010; Sripathi et al., 2017), generally increase the power of the ANOVA (Legendre et al., 2004). Additionally, we found that the ALPHA design was the best in its ability to recover superior genotypes in terms of yield and the accuracy measured by the correlation between true and expected genotypic effects. Although we expect unbiased estimations in all models due to the randomization process (Fisher, 1935), we expect models taking into account the underlying variability in the field to be more efficient (Lehmann and Casella, 1998). The larger variability not controlled with some experimental designs possibly induced more noise in the estimation of the genotypic effect, and therefore smaller values of Best_Gen and COR were obtained. The low efficiency of simpler designs in controlling field heterogeneity should reduce the accuracy in recovering superior genotypes. Stroup et al. (1994) suggested that the presence of spatial correlation in the data inflates the estimation of experimental error and could lead to inaccurate estimates of treatment means using classical RCBD analysis.

Even with low field heterogeneity present, the ALPHA design was slightly better when a large number of genotypes are used, mainly with some spatial correction. The efficiency of an experimental design depends on its complexity, and it is regulated by experiment size and the nature of the spatial variability. The larger the number of treatments, the smaller the efficiency of simpler designs in general (Casler, 2015).

Model performance improves when combining well-designed experiments with spatial adjustments (Gilmour et al., 1997; Qiao et al., 2000; Williams et al., 2006), suggesting that both, design information and spatial modeling should be considered. We found that under high

field heterogeneity, spatial models improved the performance of all experimental designs. On the other hand, spatial modeling did not outperform experimental design in any of our situations. Therefore, spatial modeling should be considered as a supplemental strategy rather than an alternative to experimental design, as was also suggested by Richter and Kroschewski (2012) and Piepho et al. (2015). Under high field variability, CRD or RCBD with spatial corrections were similar to ALPHA without spatial correction for precision.

One approach that was beyond the scope of this paper but is worth mentioning is the use of spatial designs. Our paper focused on answering the question as to whether spatial analysis can be a substitute for a good classical experimental design, and we found that there is no good substitute for a well-planned experimental design. However, we did not evaluate the use of a priori spatial information in the optimization of classical experimental designs, which have been called spatial designs (Williams et al., 2006). There is a growing interest in the literature for these spatial designs (Piepho and Williams, 2010; Williams and Piepho, 2013; Piepho et al., 2016). However, one of the most challenging aspects of this approach is demonstrating the validity of the restrictions in the randomization and the presence of error variance bias in these designs (Williams and Piepho, 2018). Furthermore, even if soil spatial variation was previously well characterized and experiments are designed accordingly, there is still variation that cannot be predicted a priori (Cooper et al., 2014), making it even more challenging to produce good randomizations.

The best spatial model seems to be, most likely, case specific (Cullis and Gleeson, 1991; Richter and Kroschewski, 2012; Moehring et al., 2014). We found that the spatial correlation structure modeled with AR(1) seemed to be more suitable to describe the spatial pattern present in our uniformity trial, where the strongest spatial variability gradient was given in the direction of the columns (Fig. 1). We used the direction of maximum spatial variability known a priori to model the AR(1) process. This could theoretically favor the AR(1) model. However, because in the absence of prior information on spatial variability it is a common practice to evaluate the spatial processes in both directions, we do not believe our results are biased. The exponential structure in two dimensions gave the same weights to spatial variability in rows and columns that did not correspond with the pattern observed in the field at most of the locations. However, a two-dimensional AR(1) process was better than a one-dimensional one in several studies (Cullis and Gleeson, 1991; Moehring et al., 2014) because it does not assume isotropy. One of the limitations of our study is that only an isotropic spatial correction model was considered. Given that our uniformity trial had higher variability in the north–south direction than in the east–west direction, an anisotropic model might have provided better spatial control. In that situation, an AR(1) × AR(1) model might outperform the AR(1) and EXP(2) models (Moehring et al., 2014).

Because of the nature of PREP, a choice between fixing the number of treatments (and therefore having fewer experimental units) or fixing the number of experimental units (and having more treatments) had to be made. We evaluated both strategies. First, we compared all experimental designs for a fixed set of treatments; therefore, $PREP_g$ used fewer experimental units and was at a disadvantage compared with the other designs. Under these circumstances, $PREP_g$ did not perform as well as the other experimental designs. With a moderate size of experiment, it was the worst design. However, under large experiment sizes, the differences with simpler designs were smaller, while $PREP_g$ used fewer resources. Second, we evaluated the performance of the PREP experimental designs, fixing the resources by using the same number of experimental units and therefore evaluating a larger number of treatments ($PREP_n$). This situation was not better in terms of precision and accuracy than the $PREP_g$. Other studies did find that for a fixed experiment size, PREP designs are efficient and can outperform replicated designs in multiple-environment experiments (Cullis et al., 2006, Moehring et al., 2014). It can be expected that in PREP designs, some precision and accuracy are sacrificed because there are fewer replications. Endelman et al. (2014) pointed out that in preliminary yield trials in multiple locations, allocating additional plots per entry increases accuracy. What we are not considering in this work is the impact of evaluating larger population sizes with more genotypes. In our $PREP_n$, 428 genotypes were evaluated, vs. the 200 that other designs evaluated using the same number of experimental units. Moehring et al. (2014) found that an augmented design outperformed replicated and classical augmented designs in terms of prediction accuracy, providing a better sample of the genotype × environment interaction. Another important consideration is the allocation of repeated plots. There are different ways in which this can be accomplished, mainly by increasing the number of repeated genotypes, or increasing the number of replications. However, the most important decision to make is the total number of repeated plots used and not the number of repeated genotypes, because that number determines the total degrees of freedom that affect design performance (Clarke and Stefanova, 2011). Optimizing this might slightly improve the performance of our PREP. The use of larger population sizes would have other advantages not reflected in our study, such as larger population sizes for genomic studies (i.e., mapping or genomic selection). Larger population sizes would result in larger selection intensities that could increase the selection response in breeding populations.

Trait heritability did not change the conclusions of this work (Supplemental Tables S2–S4). The relative performance of the experimental designs and spatial correction models was the same regardless of the heritability. The experiments and spatial models all performed more poorly with lower heritability. Therefore, we believe that our conclusions can be generalized to a large number of situations.

In summary, under high field heterogeneity and with a large number of treatments, spatial modeling without local control does not outperform local control with proper experimental designs. Once you have chosen the proper experimental design, spatial modeling can further improve its performance, as also seen in Qiao et al. (2000). This is especially the case in moderate- to large-sized experiments and under both high and low field variability. Spatial patterns using precision agriculture technology could therefore be used to better design experiments and to find an adequate spatial correction for each experiment (Cooper et al., 2014).

Conflict of Interest

The authors declare that there is no conflict of interest.

Supplemental Data Available

Supplemental material is available online for this article.

Acknowledgments

A. Borges was funded by a young scientist award from the Comisión Sectorial de Investigación Científica (CSIC-ID 714) to conduct this study. Funding was also provided by a research grant from the Fondo María Viñas (FMV_2009_1_2496). We would like to thank two anonymous reviewers for their careful reading and constructive comments.

References

Brownie, C., D.T. Bowman, and J.W. Burton. 1993. Estimating spatial variation in analysis of data from yield trials: A comparison of methods. Agron. J. 85:1244–1253. doi:10.2134/agronj1993.00021962008500060028x

Casler, M.D. 2015. Fundamentals of experimental design: Guidelines for designing successful experiments. Agron. J. 107:692–705. doi:10.2134/agronj2013.0114

Casler, M.D., and D.J. Undersander. 2000. Forage yield precision, experimental design, and cultivar mean separation for alfalfa cultivar trials. Agron. J. 92:1064–1071. doi:10.2134/agronj2000.9261064x

Clarke, G.P., and K. Stefanova. 2011. Optimal design for early-generation plant-breeding trials with unreplicated or partially replicated test lines. Aust. N. Z. J. Stat. 53:461–480. doi:10.1111/j.1467-842X.2011.00642.x

Cochran, W.G., and G.M. Cox. 1957. Experimental designs. 2nd ed. John Wiley & Sons, New York.

Cooper, M., C.D. Messina, D. Podlich, L.R. Totir, A. Baumgarten, N.J. Hausmann, et al. 2014. Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. Crop Pasture Sci. 65:311–336. doi:10.1071/CP14007

Cullis, B.R., and A.C. Gleeson. 1991. Spatial analysis of field experiments-an extension to two dimensions. Biometrics 47:1449–1460. doi:10.2307/2532398

Cullis, B.R., A.B. Smith, and N.E. Coombes. 2006. On the design of early generation variety trials with correlated data. J. Agric. Biol. Environ. Stat. 11:381–393. doi:10.1198/108571106X154443

de Mendiburu, F. 2012. Agricolae: Statistical procedures for agricultural research. R package version 1.1–3. R Found. Stat. Comput., Vienna. http://CRAN.R-project.org/package=agricolae (accessed 13 Nov. 2018).

Endelman, J.B., G.N. Atlin, Y. Beyene, K. Semagn, X. Zhang, M.E. Sorrells, and J.L. Jannink. 2014. Optimal design of preliminary yield trials with genome-wide markers. Crop Sci. 54:48–59. doi:10.2135/cropsci2013.03.0154

Fisher, R.A. 1935. The design of experiments. Oliver and Boyd, Oxford, UK.

Gauch, H.G., J.T.G. Hwang, and G.W. Fick. 2003. Model evaluation by comparison of model-based predictions and measured values. Agron. J. 95:1442–1446. doi:10.2134/agronj2003.1442

Gilmour, A.R., B.R. Cullis, and A.P. Verbyla. 1997. Accounting for natural and extraneous variation in the analysis of field experiments. J. Agric. Biol. Environ. Stat. 2:269–293. doi:10.2307/1400446

Gleeson, A.C., and B.R. Cullis. 1987. Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. Biometrics 43:277–287. doi:10.2307/2531812

Gonçalves, E., A. St Aubyn, and A. Martins. 2010. Experimental designs for evaluation of genetic variability and selection of ancient grapevine varieties: A simulation study. Heredity 104:552–562. doi:10.1038/hdy.2009.153

Grondona, M.O., and N. Cressie. 1991. Using spatial considerations in the analysis of experiments. Technometrics 33:381–392. doi:10.1080/00401706.1991.10484867

Grondona, M.O., J. Crossa, P.N. Fox, and W.H. Pfeiffer. 1996. Analysis of variety yield trials using two-dimensional separable ARIMA processes. Biometrics 52:763–770. doi:10.2307/2532916

INIA. 2017. Evaluación de cultivares. Inst. Nacl. Invest. Agropecu. http://www.inia.uy/productos-y-servicios/servicios-t%C3%A9cnicos/Evaluación-de-Cultivares (accessed 13 Nov. 2018).

Koch, E., and J. Rigney. 1951. A method of estimating optimum plot size from experimental data. Agron. J. 43:17–21. doi:10.2134/agronj1951.00021962004300010005x

Kravchenko, A.N., G.R. Robertson, S.S. Snap, and A.J.M. Smucker. 2006. Using Information about spatial variability to improve estimates of total soil carbon. Agron. J. 98:823–829. doi:10.2134/agronj2005.0305

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? Ecology 74:1659–1673. doi:10.2307/1939924

Legendre, P., M.R.T. Dale, M.J. Fortin, P. Casgrain, and J. Gurevitch. 2004. Effects of spatial structures on the results of field experiments. Ecology 85:3202–3214. doi:10.1890/03-0677

Lehmann, E.L., and G. Casella. 1998. Theory of point estimation. 2nd ed. Springer, New York.

Masood, A.M., K. Farooq, Y. Mujahid, and M.Z. Anwar. 2008. Improvement in precision of agricultural field experiments through design and analysis. Pak. J. Life Social Sci. 6:89–91.

Moehring, J., E.R. Williams, and H.P. Piepho. 2014. Efficiency of augmented p-rep designs in multi-environmental trials. Theor. Appl. Genet. 127:1049–1060. doi:10.1007/s00122-014-2278-y

Müller, B.U., K. Kleinknecht, J. Möhring, and H.P. Piepho. 2010. Comparison of spatial models for sugar beet and barley trials. Crop Sci. 50:794–802. doi:10.2135/cropsci2009.03.0153

Patterson, H.D., and E.R. Williams. 1976. A new class of resolvable incomplete block designs. Biometrika 63:83–92. doi:10.1093/biomet/63.1.83

Pebesma, E.J., and R.S. Bivand. 2005. Classes and methods for spatial data in R. R News 5(2):9–13. https://cran.r-project.org/doc/Rnews/Rnews_2005-2.pdf (accessed 3 Nov. 2018).

Piepho, H.P., and E.R. Williams. 2010. Linear variance models for plant breeding trials. Plant Breed. 129:1–8.

Piepho, H.P., E.R. Williams, and V. Michel. 2015. Beyond Latin squares: A brief tour of row-column designs. Agron. J. 107:2263–2270. doi:10.2134/agronj15.0144

Piepho, H.P., E.R. Williams, and V. Michel. 2016. Nonresolvable row–column designs with an even distribution of treatment replications. J. Agric. Biol. Environ. Stat. 21:227–242. doi:10.1007/s13253-015-0241-2

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, S. Heisterkamp, B. Van Willigen, R Core Development Team. 2013. nlme: Linear and nonlinear mixed effects models. R package version 3.1-109. R Found. Stat. Comput., Vienna.

Qiao, C.G., K.E. Basford, I.H. DeLacy, and M. Cooper. 2000. Evaluation of experimental designs and spatial analyses in wheat breeding trials. Theor. Appl. Genet. 100:9–16. doi:10.1007/s001220050002

R Core Team. 2016. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna.

Richter, C., and B. Kroschewski. 2012. Geostatistical models in agricultural field experiments: Investigations based on uniformity trials. Agron. J. 104:91–105. doi:10.2134/agronj2011.0100

Sripathi, R., P. Conaghan, D. Grogan, and M.D. Casler. 2017. Field design factors affecting the precision or ryegrass forage yield estimation. Agron. J. 109:858–869. doi:10.2134/agronj2016.07.0397

Stroup, W.W., P.S. Baenzinger, and D.K. Mulitze. 1994. Removing spatial variation from wheat yield trials: A comparison of methods. Crop Sci. 34:62–66. doi:10.2135/cropsci1994.0011183X003400010011x

USDA-ARS. 1992. Germplasm Resources Information Network (GRIN) database. *Triticum aestivum* L. subsp. *aestivum*. POACEAE 'Nogal'. Natl. Germplasm Resour. Lab., Natl. Genet. Resour. Progr., Beltsville, MD. https://npgsweb.ars-grin.gov/gringlobal/accessiondetail.aspx?id=1457618 (accessed 1 Oct. 1992).

Williams, E.R., J.A. John, and D. Whitaker. 2006. Construction of resolvable spatial row–column designs. Biometrics 62:103–108. doi:10.1111/j.1541-0420.2005.00393.x

Williams, E.R., A.C. Matheson, and C.E. Harwood. 2002. Experimental design and analysis for tree improvement. CSIRO Publ., Clayton, VIC.

Williams, E.R., and H.P. Piepho. 2013. A comparison of spatial designs for field variety trials. Aust. N. Z. J. Stat. 55:253–258. doi:10.1111/anzs.12034

Williams, E.R., and H.P. Piepho. 2018. An evaluation of error variance bias in spatial designs. J. Agric. Biol. Environ. Stat. 23:83–91. doi:10.1007/s13253-017-0309-2

2.7. SUPPLEMENTAL MATERIAL

Supplemental Table S1. Mean and standard deviation (in parenthesis) of the recovery of 15% of superior genotypes (Best_Gen), the standard error for mean differences (SED), the Pearson correlation coefficient between observed and estimated genotypic effects (COR), and the mean square error of prediction (MSEP), for three experimental designs (CRD, RCBD and ALPHA) and three spatial corrections (NSC, AR(1) and EXP(2)), in two types of field locations (Loc: high and low spatial heterogeneity). Small experiment size (15 genotypes) and high heritability.

Loc†	Design‡	Model§	Best_Gen	SED	COR	MSEP
High variability	CRD	NSC	0.53 (0.29)	0.35 (0.03)	0.73 (0.11)	0.14 (0.04)
$YSD = 0.55$	CRD	AR(1)	0.52 (0.29)	0.34 (0.03)	0.73 (0.11)	0.13 (0.04)
40 locations	CRD	EXP(2)	0.56 (0.29)	359.59 (4507.16)	0.76 (0.09)	0.16 (0.08)
	RCBD	NSC	0.57 (0.29)	0.31 (0.03)	0.77 (0.09)	0.28 (0.05)
	RCBD	AR(1)	0.56 (0.29)	0.30 (0.03)	0.77 (0.09)	0.28 (0.06)
	RCBD	EXP(2)	0.57 (0.29)	65.83 (1886.22)	0.77 (0.09)	0.28 (0.07)
	ALPHA	NSC	0.56 (0.29)	0.30 (0.03)	0.77 (0.09)	0.28 (0.06)
	ALPHA	AR(1)	0.56 (0.29)	0.30 (0.03)	0.76 (0.10)	0.28 (0.06)
	ALPHA	EXP(2)	0.56 (0.29)	0.30 (0.03)	0.77 (0.09)	0.28 (0.06)
Low variability	CRD	NSC	0.70 (0.26)	0.18 (0.02)	0.89 (0.04)	0.04 (0.01)
$YSD = 0.18$	CRD	AR(1)	0.69 (0.26)	0.18 (0.02)	0.89 (0.05)	0.04 (0.01)
59 locations	CRD	EXP(2)	0.70 (0.26)	17.12 (90.34)	0.90 (0.04)	0.04 (0.01)
	RCBD	NSC	0.72 (0.25)	0.17 (0.02)	0.90 (0.04)	0.05 (0.02)
	RCBD	AR(1)	0.71 (0.26)	0.17 (0.02)	0.90 (0.04)	0.05 (0.02)
	RCBD	EXP(2)	0.71 (0.25)	2.56 (36.5)	0.90 (0.04)	0.05 (0.02)
	ALPHA	NSC	0.71 (0.26)	0.17 (0.02)	0.90 (0.04)	0.05 (0.02)
	ALPHA	AR(1)	0.70 (0.26)	0.17 (0.02)	0.89 (0.05)	0.05 (0.02)
	ALPHA	EXP(2)	0.71 (0.26)	0.17 (0.02)	0.90 (0.04)	0.05 (0.02)

† YSD: yield standard deviation (MgMg ha)
‡CRD: Completely Randomized Design; RCBD: Randomized Complete Block Design; ALPHA: Incomplete Blocks - Alpha Lattice Design
§NSC: No spatial correction model; AR(1): spatial correlated error model with one-dimensional auto-regressive process; EXP(2): spatial correlated error model with two-dimensional exponential spatial correlation structure

Supplemental Table S2. Mean and standard deviation (in parenthesis) of the recovery of 15% of superior genotypes (Best_Gen), the standard error for mean differences (SED), the Pearson correlation coefficient between observed and estimated genotypic effects (COR), and the mean square error of prediction (MSEP), for three experimental designs (CRD, RCBD and ALPHA) and three spatial corrections (NSC, AR(1) and EXP(2)), in two types of field locations (Loc: high and low spatial heterogeneity). Small experiment size (15 genotypes) and low heritability.

Loc†	Design‡	Model§	Best_Gen	SED	COR	MSEP
High variability	CRD	NSC	0.47 (0.30)	0.41 (0.05)	0.68 (0.13)	0.17 (0.06)
$YSD = 0.55$	CRD	AR(1)	0.47 (0.30)	0.40 (0.05)	0.67 (0.13)	0.17 (0.06)
40 locations	CRD	EXP(2)	0.49 (0.30)	124.63 (740.99)	0.69 (0.12)	0.20 (0.10)
	RCBD	NSC	0.50 (0.30)	0.37 (0.04)	0.71 (0.12)	0.34 (0.09)
	RCBD	AR(1)	0.49 (0.30)	0.36 (0.05)	0.70 (0.12)	0.35 (0.09)
	RCBD	EXP(2)	0.49 (0.30)	7.45 (106.13)	0.70 (0.12)	0.34 (0.11)
	ALPHA	NSC	0.49 (0.30)	0.37 (0.04)	0.70 (0.12)	0.35 (0.09)
	ALPHA	AR(1)	0.48 (0.30)	0.36 (0.05)	0.69 (0.13)	0.36 (0.10)
	ALPHA	EXP(2)	0.49 (0.30)	0.36 (0.04)	0.70 (0.12)	0.35 (0.09)
Low variability	CRD	NSC	0.56 (0.29)	0.28 (0.04)	0.80 (0.08)	0.08 (0.03)
$YSD = 0.18$	CRD	AR(1)	0.55 (0.29)	0.27 (0.04)	0.80 (0.09)	0.08 (0.03)
59 locations	CRD	EXP(2)	0.56 (0.29)	51.07 (1228.16)	0.8 (0.08)	0.08 (0.03)
	RCBD	NSC	0.57 (0.29)	0.27 (0.04)	0.81 (0.08)	0.10 (0.04)
	RCBD	AR(1)	0.56 (0.29)	0.26 (0.04)	0.80 (0.08)	0.10 (0.04)
	RCBD	EXP(2)	0.56 (0.29)	3.61 (38.07)	0.81 (0.08)	0.10 (0.05)
	ALPHA	NSC	0.56 (0.29)	0.27 (0.04)	0.81 (0.08)	0.10 (0.04)
	ALPHA	AR(1)	0.55 (0.29)	0.26 (0.04)	0.80 (0.09)	0.10 (0.04)
	ALPHA	EXP(2)	0.56 (0.29)	0.26 (0.04)	0.81 (0.08)	0.10 (0.04)

† YSD: yield standard deviation (MgMg ha)
‡CRD: Completely Randomized Design; RCBD: Randomized Complete Block Design; ALPHA: Incomplete Blocks - Alpha Lattice Design
§NSC: No spatial correction model; AR(1): spatial correlated error model with one-dimensional auto-regressive process; EXP(2): spatial correlated error model with two-dimensional exponential spatial correlation structure

Supplemental Table S3. Mean and standard deviation (in parenthesis) of the recovery of 15% of superior genotypes (Best_Gen), the standard error for mean differences (SED), the Pearson correlation coefficient between observed and estimated genotypic effects (COR), and the mean square error of prediction (MSEP), for four experimental designs (CRD, RCBD, ALPHA and PREP) and three spatial corrections (NSC, AR(1) and EXP(2)), in two types of field locations (Loc: high and low spatial heterogeneity). Moderate experiment size (50 genotypes) and low heritability

Loc†	Design‡	Model§	Best_Gen	SED	COR	MSEP
High variability	CRD	NSC	0.42 (0.15)	0.53 (0.03)	0.58 (0.08)	0.29 (0.05)
YSD=0.80	CRD	AR(1)	0.44 (0.15)	0.44 (0.04)	0.62 (0.08)	0.23 (0.04)
46 locations	CRD	EXP(2)	0.42 (0.15)	17.2 (493.22)	0.59 (0.08)	0.30 (0.07)
	RCBD	NSC	0.42 (0.15)	0.51 (0.03)	0.60 (0.08)	0.38 (0.06)
	RCBD	AR(1)	0.46 (0.15)	0.40 (0.04)	0.65 (0.07)	0.31 (0.05)
	RCBD	EXP(2)	0.42 (0.15)	2.28 (20.09)	0.6 (0.08)	0.38 (0.07)
	ALPHA	NSC	0.48 (0.15)	0.42 (0.03)	0.68 (0.07)	0.28 (0.05)
	ALPHA	AR(1)	0.52 (0.15)	0.37 (0.03)	0.73 (0.06)	0.24 (0.05)
	ALPHA	EXP(2)	0.52 (0.15)	0.37 (0.03)	0.73 (0.06)	0.24 (0.05)
	$PREP_r$	NSC	0.33 (0.15)	0.70 (0.12)	0.46 (0.11)	0.70 (0.20)
	$PREP_r$	AR(1)	0.34 (0.15)	0.66 (0.12)	0.46 (0.11)	0.70 (0.23)
	$PREP_r$	EXP(2)	0.34 (0.15)	0.67 (0.11)	0.47 (0.11)	0.66 (0.20)
	$PREP_c$	NSC	0.22 (0.14)	0.72 (0.08)	0.42 (0.08)	0.69 (0.12)
	$PREP_c$	AR(1)	0.22 (0.14)	0.68 (0.08)	0.43 (0.08)	0.66 (0.13)
	$PREP_c$	EXP(2)	0.22 (0.14)	0.69 (0.08)	0.44 (0.08)	0.65 (0.12)
Low	CRD	NSC	0.59 (0.14)	0.30 (0.02)	0.79 (0.05)	0.09 (0.02)
YSD=0.25	CRD	AR(1)	0.58 (0.14)	0.29 (0.02)	0.79 (0.05)	0.09 (0.02)
53 locations	CRD	EXP(2)	0.59 (0.14)	0.29 (0.02)	0.79 (0.05)	0.09 (0.02)
	RCBD	NSC	0.59 (0.14)	0.29 (0.02)	0.80 (0.04)	0.11 (0.02)
	RCBD	AR(1)	0.59 (0.14)	0.28 (0.02)	0.79 (0.04)	0.11 (0.02)
	RCBD	EXP(2)	0.59 (0.14)	0.29 (0.02)	0.79 (0.04)	0.11 (0.02)
	ALPHA	NSC	0.60 (0.14)	0.28 (0.02)	0.80 (0.04)	0.11 (0.02)
	ALPHA	AR(1)	0.60 (0.14)	0.28 (0.02)	0.80 (0.04)	0.11 (0.02)
	ALPHA	EXP(2)	0.60 (0.14)	0.28 (0.02)	0.80 (0.04)	0.11 (0.02)
	$PREP_r$	NSC	0.45 (0.15)	0.45 (0.08)	0.62 (0.08)	0.27 (0.07)
	$PREP_r$	AR(1)	0.44 (0.15)	0.42 (0.08)	0.60 (0.09)	0.29 (0.09)
	$PREP_r$	EXP(2)	0.45 (0.15)	0.44 (0.08)	0.62 (0.08)	0.27 (0.08)
	$PREP_c$	NSC	0.33 (0.15)	0.45 (0.05)	0.59 (0.06)	0.26 (0.04)
	$PREP_c$	AR(1)	0.32 (0.15)	0.44 (0.05)	0.58 (0.06)	0.27 (0.05)
	$PREP_c$	EXP(2)	0.33 (0.15)	0.45 (0.05)	0.59 (0.06)	0.26 (0.04)

† YSD: yield standard deviation (MgMg ha)

‡†CRD: Completely Randomized Design; RCBD: Randomized Complete Block Design; ALPHA: Incomplete Blocks - Alpha Lattice Design; $PREP_r$: Partially replicated design with 50 genotypes (this experiment preserved the number of genotypes and in consequence used fewer experimental units); $PREP_c$: Partially replicated design with 108 genotypes (this experiment preserved the number of experimental units and in consequence evaluated more genotypes).

§NSC: No spatial correction model; AR(1): spatial correlated error model with one-dimensional auto-regressive process; EXP(2): spatial correlated error model with two-dimensional exponential spatial correlation structure

Supplemental Table S4. Mean and standard deviation (in parenthesis) of the recovery of 15% of superior genotypes (Best_Gen), the standard error for mean differences (SED), the Pearson correlation coefficient between observed and estimated genotypic effects (COR), and the mean square error of prediction (MSEP), for four experimental designs (CRD, RCBD, ALPHA and PREP) and three spatial corrections (NSC, AR(1) and EXP(2)), in two types of field locations (Loc: high and low spatial heterogeneity). Large experiment size (200 genotypes) and low heritability.

Loc†	Design‡	Model§	Best_Gen	SED	COR	MSEP
High variability	CRD	NSC	0.39 (0.07)	0.69 (0.02)	0.49 (0.05)	0.48 (0.04)
YSD=1.10	CRD	AR(1)	0.43 (0.07)	0.49 (0.02)	0.56 (0.04)	0.32 (0.03)
58 locations	CRD	EXP(2)	0.39 (0.07)	0.68 (0.02)	0.49 (0.05)	0.47 (0.04)
	RCBD	NSC	0.39 (0.07)	0.67 (0.02)	0.50 (0.05)	0.56 (0.04)
	RCBD	AR(1)	0.46 (0.07)	0.43 (0.02)	0.61 (0.04)	0.37 (0.03)
	RCBD	EXP(2)	0.39 (0.07)	0.67 (0.02)	0.50 (0.05)	0.56 (0.04)
	$ALPHA_{(5)}$	NSC	0.46 (0.07)	0.43 (0.01)	0.63 (0.04)	0.29 (0.03)
	$ALPHA_{(5)}$	AR(1)	0.49 (0.07)	0.38 (0.01)	0.67 (0.04)	0.25 (0.03)
	$ALPHA_{(5)}$	EXP(2)	0.49 (0.07)	0.38 (0.01)	0.67 (0.04)	0.25 (0.03)
	$ALPHA_{(10)}$	NSC	0.44 (0.07)	0.54 (0.02)	0.59 (0.04)	0.40 (0.03)
	$ALPHA_{(10)}$	AR(1)	0.56 (0.07)	0.36 (0.01)	0.74 (0.03)	0.25 (0.03)
	$ALPHA_{(10)}$	EXP(2)	0.56 (0.07)	0.36 (0.01)	0.74 (0.03)	0.25 (0.03)
	$PREP_g$	NSC	0.35 (0.07)	0.84 (0.06)	0.42 (0.06)	0.78 (0.09)
	$PREP_g$	AR(1)	0.38 (0.07)	0.71 (0.05)	0.48 (0.06)	0.55 (0.09)
	$PREP_g$	EXP(2)	0.38 (0.07)	0.71 (0.05)	0.48 (0.06)	0.55 (0.09)
	$PREP_e$	NSC	0.23 (0.07)	0.89 (0.05)	0.37 (0.04)	0.93 (0.07)
	$PREP_e$	AR(1)	0.26 (0.07)	0.72 (0.04)	0.44 (0.04)	0.65 (0.07)
	$PREP_e$	EXP(2)	0.26 (0.07)	0.72 (0.04)	0.44 (0.04)	0.65 (0.07)
Low variability	CRD	NSC	0.56 (0.07)	0.37 (0.01)	0.72 (0.03)	0.14 (0.01)
YSD=0.46	CRD	AR(1)	0.56 (0.07)	0.36 (0.01)	0.73 (0.03)	0.14 (0.01)
35 locations	CRD	EXP(2)	0.56 (0.07)	0.37 (0.01)	0.73 (0.03)	0.14 (0.01)
	RCBD	NSC	0.57 (0.07)	0.37 (0.01)	0.73 (0.03)	0.16 (0.01)
	RCBD	AR(1)	0.56 (0.07)	0.35 (0.01)	0.74 (0.03)	0.15 (0.01)
	RCBD	EXP(2)	0.56 (0.07)	0.37 (0.01)	0.73 (0.03)	0.15 (0.01)
	$ALPHA_{(5)}$	NSC	0.54 (0.07)	0.32 (0.01)	0.72 (0.03)	0.12 (0.01)
	$ALPHA_{(5)}$	AR(1)	0.55 (0.07)	0.32 (0.01)	0.73 (0.03)	0.11 (0.01)
	$ALPHA_{(5)}$	EXP(2)	0.55 (0.07)	0.32 (0.01)	0.73 (0.03)	0.11 (0.01)
	$ALPHA_{(10)}$	NSC	0.58 (0.07)	0.34 (0.01)	0.75 (0.03)	0.13 (0.01)
	$ALPHA_{(10)}$	AR(1)	0.59 (0.07)	0.31 (0.01)	0.78 (0.02)	0.11 (0.01)
	$ALPHA_{(10)}$	EXP(2)	0.59 (0.07)	0.31 (0.01)	0.78 (0.02)	0.11 (0.01)
	$PREP_g$	NSC	0.46 (0.07)	0.54 (0.04)	0.58 (0.04)	0.33 (0.04)
	$PREP_g$	AR(1)	0.46 (0.07)	0.51 (0.04)	0.59 (0.05)	0.31 (0.04)
	$PREP_g$	EXP(2)	0.46 (0.07)	0.51 (0.04)	0.59 (0.04)	0.31 (0.04)
	$PREP_e$	NSC	0.34 (0.07)	0.56 (0.03)	0.53 (0.03)	0.35 (0.03)
	$PREP_e$	AR(1)	0.34 (0.07)	0.53 (0.03)	0.54 (0.03)	0.32 (0.03)
	$PREP_e$	EXP(2)	0.34 (0.07)	0.53 (0.03)	0.54 (0.03)	0.32 (0.03)

† *YSD*: yield standard deviation (MgMg ha)
‡CRD: Completely Randomized Design; RCBD: Randomized Complete Block Design; $ALPHA_{(5)}$: Incomplete Blocks - Alpha Lattice Design with block size = 5; $ALPHA_{(10)}$: Incomplete Blocks - Alpha Lattice Design with block size = 10; $PREP_g$: Partially replicated design with 200 genotypes (this experiment preserved the number of genotypes and in consequence used fewer experimental units); $PREP_e$: Partially replicated design with 428 genotypes (this experiment preserved the number of experimental units and in consequence evaluated more genotypes).
§NSC: No spatial correction model; AR(1): spatial correlated error model with one-dimensional auto-regressive process; EXP(2): spatial correlated error model with two-dimensional exponential spatial correlation structure.

3. EFFECT OF SPATIAL COVARIANCES ON THE BIAS OF VARIANCE ESTIMATES, CONTROL OF TYPE 1 ERROR RATE AND POWER

3.1. RESUMEN

Los experimentos agronómicos pueden diseñarse para tener un control local de la variabilidad espacial y, en función de las condiciones del campo y del material experimental, pueden aplicarse diseños experimentales adecuados para diversas situaciones. Sin embargo, incluso con diseños de fuerte control local, la variabilidad espacial puede no estar correctamente controlada. Por estas razones, se pueden obtener beneficios adicionales en la precisión y exactitud de los resultados de un experimento si se incluye en los modelos de análisis la estimación de la variabilidad espacial. Sin embargo, más allá de estas ventajas de los modelos espaciales en situaciones prácticas, existen algunos problemas a los que hay que enfrentarse, como la selección de un modelo de covarianza adecuado para el análisis o de un método de aproximación que corrija los grados de libertad y la matriz de covarianza estimada de los efectos fijos. Incluso en situaciones donde se utilizan cualquiera de los métodos de corrección disponibles, no siempre se logran corregir problemas de sesgo en las varianzas de los efectos fijos y de falta de control de la tasa de error empírica de tipo 1 (ET1), cuestionando la validez de las inferencias realizadas. Los principales objetivos de este trabajo fueron: i) evaluar si la ganancia esperada de precisión de los modelos espaciales puede lograrse sin sesgo en los errores estándar estimados y con adecuado control de la ET1 y ii) valorar cómo el número de réplicas por tratamiento y la inclusión o no del efecto de bloque completo pueden afectar al rendimiento de los modelos de análisis. En cuatro sitios de un ensayo de uniformidad de trigo, se simularon los efectos de diferente número de tratamientos, aleatorizados según tres diseños experimentales y analizados con tres modelos diferentes (dos modelos espaciales y un modelo de error independiente). La comparación de los modelos se realizó con base en el criterio de información de Akaike (AIC), al sesgo en el error estándar de la diferencia media (SED), el control de la ET1 y la potencia de las

pruebas de comparación de medias. Los modelos espaciales disminuyeron el error estándar de la diferencia en todos los casos. En casi todos los escenarios, hubo algún modelo espacial que aumentó la precisión sin sesgo en el SED y con un control adecuado de la ET1. El modelo elegido dependió del lugar, del diseño experimental y del número de réplicas por tratamiento. En las situaciones donde el número de réplicas fue elevado (36), no se observaron los problemas mencionados, independientemente del modelo de análisis. Por último, cuando no se incluyeron bloques completos en el análisis, al menos uno de los modelos espaciales mejoró el rendimiento del modelo de error independiente, minimizando los efectos de sesgo y logrando un control adecuado de la ET1.

3.2.ABSTRACT

Agronomic experiments can be designed to have local control of the spatial variability, and, depending on the field and experimental material conditions, suitable experimental designs can be implemented for a variety of situations. However, even with strong local control designs, spatial variability may not be correctly controlled. For these reasons, the estimation of spatial variability and its inclusion in the analysis models can produce additional benefits, improving the estimation accuracy and precision. Nevertheless, beyond these advantages of spatial models in practical situations, there are some problems to be faced, like the selection of a suitable covariance model for the analysis or an approximation method that corrects the degrees of freedom and the estimated covariance matrix for fixed effects. However, using any of the correction methods available, problems of bias and validity of statistical tests due to lack of control of empirical type 1 error rate (ET1) are not always corrected. The main objectives of this work were to evaluate if the expected gain in precision of spatial models can be achieved without bias in the estimated standard errors and lack of control of ET1 and to assess how the number of replicates per treatment and the inclusion or not of complete block effect can affect model performance. At four sites of a wheat uniformity trial, effects of different number of treatments were simulated, randomized according to three experimental designs and

analyzed with three different models (two spatial models and one independent error model). Model comparison was performed on the basis of Akaike information criterion (AIC), bias in the standard error of the mean difference (SED), control of ET1 and power of mean comparison tests. The spatial models decreased the standard error of the difference in all cases. In almost all scenarios, there was some spatial model that increased precision without bias in the SED and with adequate control of ET1. The appropriate model depended on the site, the experimental design, and the number of replicates per treatment. When the number of replicates was high (36), the aforementioned problems were not observed, regardless of the analysis model. Finally, when complete blocks were not included in the analysis, at least one of the spatial models improved the performance of the independent error model, minimizing bias effects and achieving adequate control of ET1.

3.3.INTRODUCTION

Agronomic research is generally based on comparative experiments, whose main objective is to compare or contrast two or more agricultural systems or practices, which may have some relevance in our field of scientific research (Casler, 2015). In many cases, it involves large field trials, where it is a great challenge to achieve homogeneity among experimental units in space. For these reasons, the specific choices that should be made in the analysis are not obvious.

To deal with field variability and correlated errors, two strategies have been proposed alone or combined: effective experimental designs and/or spatial modeling. It is often debated whether it is better to control spatial variability with experimental designs or with spatial models. Spatial analysis is not always necessarily better than the basic randomized complete block design analysis, particularly in cases in which the spatial structure of the variable studied is not well characterized (Kravchenko et al., 2006). There are previous works that confirm that model performance improves when well-designed experiments are combined with spatial adjustments (Gilmour et al., 1997, Qiao et al., 2000, Williams et al., 2006, Borges et al., 2019), which suggests that both

strategies should be considered. Some authors suggest that spatial modeling should be considered as a complementary strategy instead of an alternative to experimental design (Piepho et al., 2015, Richter and Kroschewski, 2012).

The well-designed experiments are based on the three principles proposed by Fisher (1935): randomization, replication and local control. After Fisher, agricultural research has been based on these principles that seek to control local variation, assuming independence among the experimental units and using replications to estimate the experimental error. In field conditions, the spatial variation of environmental and soil factors determines in many cases the presence of spatial autocorrelation in the crop response (Grondona et al., 1996, Legendre, 1993). Therefore, the experiments can be designed to have local control of this spatial variability.

In agriculture, the most commonly used experimental designs are completely randomized design (CRD) and randomized complete block design (RCBD) (Piepho et al., 2015). In some cases, blocking is an effective way to control experimental error (Cochran and Cox, 1992, Casler, 2015), but it is not always enough depending on the heterogeneity of the field and/or the size of the experiment. In these cases, designs with better local control are more suitable (Brownie et al., 1993). Some designs, such as incomplete block designs (Patterson and Williams, 1976) or 'row-column' designs (Williams et al., 2006), are recommended to control field heterogeneity where a large number of treatments are considered, as well as other complex treatment structures (Williams et al., 2002). More recently, spatial designs have been proposed, which have additional restrictions on their randomization based on a dependence correlation structure for the field that is either known or assumed a priori (Williams and Piepho, 2013).

However, even with strong local control designs, spatial variability may not be correctly controlled if it occurs at scales smaller than the size of the blocks (Grondona et al., 1996). For these reasons, the estimation of spatial variability and its

inclusion in the analysis models can produce additional benefits. Soil spatial variability is widely recognized as one of the most important factors affecting the comparisons of treatment effects in field experiments (López and Arrúe, 1995). To deal with this situation, the spatial variation at the plot level can be estimated and included in the response variable model. Several models that include different spatial strategies have been used, such as estimating different correlation structures of the R matrix, including trend analysis or mixtures of both (Brownie et al., 1993, Casler and Undersander, 2000), with one or two-dimensional approach can be implemented according to natural field heterogeneity (Cullis and Gleeson, 1991, Gilmour et al., 1997, Gleeson and Cullis, 1987, Qiao et al., 2000). In many situations, it was found that models that include spatial correlation are generally more efficient than those that do not include it (Brownie et al., 1993, Kravchenko et al., 2006, Mallarino et al., 2000).

Nevertheless, beyond these advantages of spatial models in practical situations, there are some points to attend. The first one is the selection of a suitable covariance model for the analysis. The most appropriate structure for each situation is commonly selected according to a model fit criterion (Brownie et al., 2004, Richter et al., 2015, Saud et al., 2016). According to Richter and Kroschewski (2012), when models take spatial correlation into account, the model is no longer known a priori as it happens in the classical analysis of independent errors. Instead, the choice of model is data-driven because the covariance structure has to be chosen. From a frequentist approach, the uncertainty related to the estimated parameters is not taken into account in the comparisons and, therefore, bias of standard errors of difference and lack of control of the empirical type 1 error could appear (Richter and Kroschewski, 2012).

In this scenario, the statistical tests for the fixed effects of the mixed linear models are generally not exact and the degrees of freedom are determined by approximation, using some of the most widespread corrections (Hu and Spilke, 2009). Among others, the Kenward-Roger approximation method corrects downward the degrees of

freedom and also adjusts standard errors and test-statistics (Guerin and Stroup, 2000), providing an adjusted estimator of the covariance matrix of the fixed effects (Kenward and Roger, 1997). However, using any of the correction methods available, problems of bias and validity of statistical tests due to lack of control of empirical type 1 error rate are not always corrected.

There are previous works that compare spatial models using uniformity trial data, where the observed variance of the treatment differences can be compared with the estimated variance of the differences (Richter and Kroschewski, 2012, Wu and Dutilleul, 1999, Zimmerman and Harville, 1991). Richter and Kroschewski (2012), based on several uniformity trials of different crops, found biases in the estimation of the standard error of the difference of spatial models, compared with the basic randomized complete block model (independent errors). This bias was, in some cases, positive and, in others, negative, depending on the crop considered, and it was of greater magnitude in cases of strong spatial correlation. More recently, Richter et al. (2015) evaluated the effect of accounting for spatial correlations on treatment comparisons based on different spatial simulated scenarios and found that when spatial simulation and analysis model do not match, problems of bias and departures from nominal type 1 error rate are observed, pointing that the choice of model selection criterion and the choice of candidate models are crucial. However, none of these works evaluate the role of the number of replications per treatment on estimation bias and control of the empirical type 1 error rate. In this sense, Hu et al. (2006) wonder if with a low number of replicates it is possible to use spatial models that have acceptable estimation errors, despite the existing spatial dependencies.
The main objectives of this work were to evaluate in an Uruguayan field example: 1) whether the expected gain in precision with adequate spatial models can be achieved without bias in the estimated standard errors and if valid statistical inferences can be reached; 2) how the number of replicates and its trade-off with the number of treatments can affect model performances, and 3) whether or not spatial models can substitute for the complete block effects, mitigating bias and empirical type 1 error rate problems.

3.4. MATERIALS AND METHODS

3.4.1. Experimental field

A uniformity trial of approximately 64 ha was sown with the "Nogal" wheat cultivar (USDA-ARS, 1992) on June 20, 2008, at a density of 120 kg ha^{-1} of seed. The field was harvested in rectangular plots of approximately 15 x 5 m area to obtain 1445 yield plots (Mg ha^{-1}) that were recorded via a yield monitor. Each plot had the geographic coordinates associated: UTM X (East-West) and Y (North-South) and elevation (m).

3.4.2. Sites and spatial variability

From yield monitor data, a map of yield data was obtained and the observed trend with respect to the geographical coordinates was removed with regression splines. With the corrected data (after removing trend), a new map was obtained and divided in four sites that suggested a different underlying spatial pattern. In each site, we selected 144 plots that were spatially arranged in a rectangular shape, occupying an area of approximately 225 x 50 meters (figure S1). Three variogram models were fitted for each site (Exponential, Gaussian and Spherical covariance structures) to describe the pre-spatial variability (prior to simulations). The best model was selected according to the weighted residual sum of squares (lower is better). In the case of parity between two models, predictions were obtained by ordinary kriging and the best model selected was the one with the lowest mean square prediction error (PE) and mean square deviation ratio (MSDR), according to Oliver and Webster (2014).

$$PE = \frac{1}{N}\sum_{i=1}^{N}\left\{z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i)\right\}^2 \qquad [1]$$

$$MSDR = \frac{1}{N}\sum_{i=1}^{N}\frac{\{z(\mathbf{x}_i)-\hat{Z}(\mathbf{x}_i)\}^2}{\hat{\sigma}_K^2(\mathbf{x}_i)} \qquad [2]$$

Where $z(\mathbf{x}_i)$ is the ith datum at $\mathbf{x}_i$, $\hat{Z}(\mathbf{x}_i)$ is the kriged prediction at $\mathbf{x}_i$, and $\hat{\sigma}_K^2(\mathbf{x}_i)$ is the kriging variance

The proportion of spatial structure (PSS) was computed as follows in Eq. [3]:

$$PSS = \sigma_0^2 / (\sigma_0^2 + \sigma_N^2) \qquad [3]$$

Where σ_0^2 is the partial sill and σ_N^2 is the nugget variance

Yield maps were created using the *sp* package (Pebesma and Bivand, 2005), regression splines were performed using *mgvc* package (Wood, 2011) and experimental variograms and kriging predictions were obtained using the *gstat* package (Pebesma, 2004) of the R software (R Core Team, 2020), using the *gstat* default binning procedure.

3.4.3. Treatments and replications

Different combinations of number of treatments (t) and number of replications (r) were used, totaling 144 plots for each site: i) t = 48 with r = 3; ii) t = 36 with r = 4 and iii) t = 4 with r = 36.

Treatment effects were simulated from a Gaussian distribution, assuming that $\tau_i \sim N(0\,;\,\sigma_\tau^2)$, being τ_i the effect of the i-th treatment and σ_τ^2 the variance of treatments. For each number of treatments, the effects were always the same. To estimate bias of the standard errors of mean differences and empirical type 1 error rate, we included 4 treatments with equal effect for t = 48 and t = 36 and 2 for t = 4, while the remaining treatments had different effects to estimate power.

3.4.4. Experimental design

Each combination of t and r was randomized according to different experimental designs. For 48 and 36 treatments, the experimental designs used were completely randomized design (CRD), randomized complete block design (RCBD) and alpha-lattice resolvable incomplete block experimental design (ALPHA). For 4 treatments, only CRD and RCBD were used.

In the CRD design, treatments were assigned to plots without any restrictions on the

randomization. In the RCBD design, treatments were randomly assigned to plots for each block independently, such that each treatment occurs once within each block. For ALPHA design, first incomplete blocks were randomized within each replication, separately for each complete replication. After that, treatments are randomly assigned to experimental units within each incomplete block, separately for each one. Since the ALPHA design is resolvable, each treatment occurs once in a complete replication.

3.4.5. Data generation procedure

A vector of yields was obtained according to the following procedure: first, treatments were assigned to plots according to one of the three experimental designs described above. Second, yield of each plot was obtained according to Eq. [4]:

$$Y_{ij} = \tau_i + \varepsilon_{ij} \qquad [4]$$

where Y_{ij} is the yield plot data simulated corresponding to the *i*th treatment and the *j*th replication, τ_i is the *i*th treatment effect corresponding to the randomly assigned treatment to the plot, ε_{ij} is the field experimental error that represents the yield spatial heterogeneity of the field, obtained from the yield monitor.

In each site, 1,000 iterations were run for each experimental design, analysis model and number of replications, performing an independent randomization for each iteration. The 'agricolae' package (de Mendiburu and Yaseen, 2020) of the R Statistical Software (R Core Team, 2020) was used for experimental design randomization.

3.4.6. Analysis model

The model used for the CRD experimental design is shown in Eq. [5]:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad [5]$$

where Y_{ij} is the observed yield for the *i*th treatment in the *j*th replicate, μ is the overall mean, τ_i is the *i*th treatment effect and ε_{ij} are the residual errors associated with the observation Y_{ij}. $\varepsilon_{ij} \sim N(0 , \sigma^2_{\varepsilon})$, and the covariance $COV(\varepsilon_{ij,} \varepsilon_{ij'})$ was modeled

according to one of the analysis models described below.

The model used for the RCBD is shown in Eq. [6]:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad [6]$$

where Y_{ij}, μ, τ_i are as previously defined, β_j is the *j*th block effect and ε_{ij} are the residual errors associated with the observation Y_{ij}. $\varepsilon_{ij} \sim N(0 , \sigma^2_{\varepsilon})$, and the covariance $COV(\varepsilon_{ij,} \varepsilon_{ij'})$ was modeled according to one of the analysis model described below. The blocks were considered fixed effects.

The model used for ALPHA is shown in Eq. [7]:

$$Y_{ijk} = \mu + \tau_i + \beta_j + \gamma_{k(j)} + \varepsilon_{ijk} \qquad [7]$$

where Y_{ijk} , μ, $\tau_{i,}$ and β_j are as previously defined, $\gamma_{k(j)}$ is the *k*th incomplete block effect nested on *j*th replication, and ε_{ijk} are the residual errors associated with the observation Y_{ijk}. $\varepsilon_{ijk} \sim N(0 , \sigma^2_{\varepsilon})$, and the covariance $COV(\varepsilon_{ijk,} \varepsilon_{ijk'})$ was modeled according to one of the analysis model described below. Incomplete blocks with four plots per block were considered random factors nested in each complete replication, assuming $\gamma_{k(j)} \sim N(0 , \sigma^2_s)$, where σ^2_s is the variance of the subblocks and the covariance $COV(\gamma_{k(j),} \gamma_{k(j)'})$ is σ^2_R the variance of the complete replications.

For each design, all analysis models considered can take the general form of linear mixed models:

$$y = X\beta + Zu + e \qquad [8]$$

Where y is the vector of observations, X is a matrix of constants associated with the fixed effects of the design, β is a vector of unknown fixed effects (treatments and blocks when corresponds), Z is a matrix of constants associated with the random effects (incomplete blocks when corresponds), u is a vector of random effects (incomplete blocks when corresponds) and e is a vector of random residual errors. The random effect vector is assumed to be distributed as $u \sim MN(0 , G)$ and the residual vector is assumed to be distributed as $e \sim MN(0 , R)$, where MN denotes a multivariate normal distribution, G is the covariance matrix among random effects

and R is the covariance matrix among the random residual errors.

Three covariance structures were used:

i) null model with independent errors (NULL)

ii) spatial model with Gaussian covariance function (GAU)

iii) spatial model with Spherical covariance function (SPH)

For NULL model, $R = I\sigma_e^2$, where I is an identity matrix and σ_e^2 is the variance component of residual random errors. For the spatial models, the covariances in R depend on spatial distances among plots, assuming isotropy (spatial correlation does not depend on the direction) and always including a nugget effect (σ_N^2). Observations from two plots with Euclidean distance (d) d > 0 (measured between plot centers) have covariance $Cov(d) = \sigma_N^2 + \sigma_e^2 f(d)$, where $f(d)$ takes different forms depending on the analysis model:

$$\text{GAU: } f(d) = exp^{-\left(\frac{d}{A}\right)^2} \qquad [9]$$

$$\text{SPH: } f(d) = \begin{cases} 1 - \frac{3d}{2A} + \frac{d^3}{2A^3}, & d < A \\ 0, & d > A \end{cases} \qquad [10]$$

where *A* is the range parameter.

Models were run including complete block effects or not, resulting in 6 analysis models for ALPHA and RCBD. In the case of the CRD, since there are no blocks, only 3 analysis models were evaluated. Table S1 shows all the situations analyzed in each site. All models were run by PROC MIXED of SAS/STAT® software version 9.4 (SAS Institute Inc., 2013). In all cases the improved Kenward-Rogers correction (first order) was used (Kenward and Roger, 2009), recommended by Richter and Kroschewski (2012) as they found substantial reduction in bias of the standard errors in the presence of nonlinear covariance structures. The MIXED statement for a SPH model with ALPHA design is presented in appendix 1 as an example.

3.4.7. Evaluation criteria for model comparisons

AIC: Best fitting model for each combination of number of treatments, design, and

site was selected according to the Akaike Information Criterion (AIC). For that purpose, the number of times each analysis model was selected according to AIC (smaller is better), AIC mean and variance were obtained for each situation. The AIC value was obtained with REML method as follows:

$$AIC = -2lnL + 2q \qquad [11]$$

Where *lnL* is the log of maximum likelihood of the model and q is the number of parameters of the variance-covariance structure.

Standard error of the difference (SED): The estimated standard error of mean differences (SED_est) and observed standard error of mean differences (SED_obs) for equal effect comparisons were obtained. The SED_est was calculated as the square root of the mean estimated variances of treatment differences and SED_obs as the square root of the mean variance of all estimated treatment differences. To relate these standard errors, the relative bias of SED_{est} (RB%) was computed according to Richter and Kroschewski (2012) and the ratio between the two standard errors was obtained as follows:

$$RB(\%) = \frac{(SED_{est} - SED_{obs})}{SED_{obs}} \times 100 \qquad [12]$$

$$Ratio = \frac{SED_{est}}{SED_{obs}} \qquad [13]$$

Empirical Type I error rate (ET1): For each model and design, we obtained p-values for t-test where the null hypothesis was true (equal treatment means). In order to assess if ET1 was controlled, the critical values to assess if ET1 is different from 5 % were obtained according to Richter et al. (2015):

$$CV_{ET1}(95\%) = 0.05 \pm 1.96 \times \sqrt{\frac{0.05 \times (1 - 0.05)}{\#iterations}} \qquad [14]$$

where *#iterations* refers to the effective number of runs for each model (for 36 and 48 treatments the maximum effective runs were 6000, while for 4 treatments was

1000). If ET1 was contained within the critical values, control of ET1 was considered acceptable. For those cases that ET1 was not contained in CV(95 %), ET1 was considered conservative if it was below the lower limit and liberal when it was above the upper limit.

Power: To assess model power (P[Reject $H_0|H_0$ false]), the proportion of times the null hypothesis was correctly rejected was obtained for each of the true mean difference (Δ) and plotted together for all situations.

3.5. RESULTS

3.5.1. Spatial variability of the uniformity trial

In each site, yield spatial correlation was verified. After removing large scale trend, in all cases this spatial correlation was well described with one of the semivariogram model considered (figure 1). In table 1, estimated parameters of best model are presented. The proportion of spatial structure indicates that in all cases the correlation was strong.

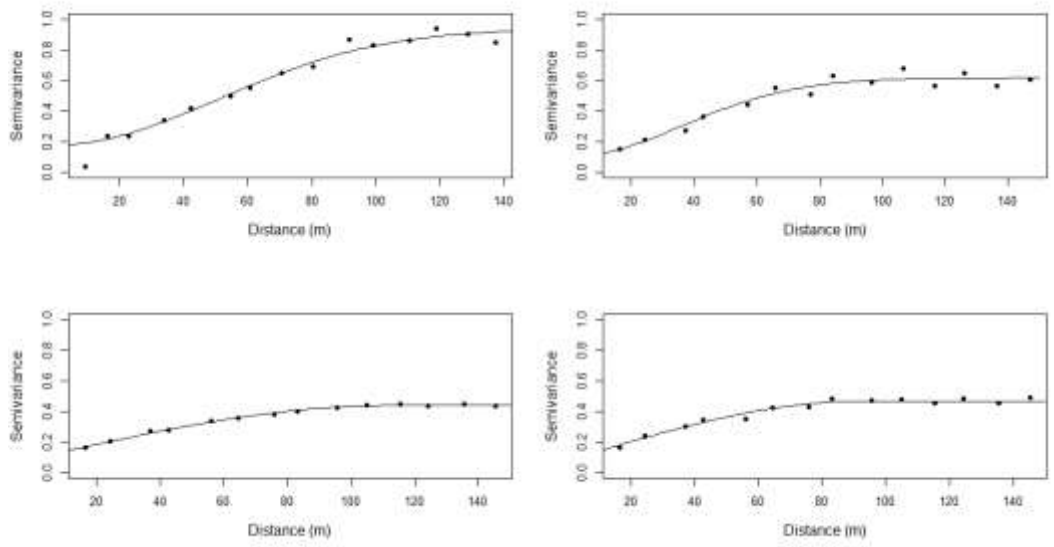


Figure 1. Empirical semivariance and best-fitted semivariogram model for each site. Site 1: Gaussian model (top left), site 2: Gaussian model (top right), site 3: Spherical

model (bottom left) and site 4: Spherical model (bottom right).

Table 1. Effective range (m), partial sill, nugget variance and proportion of spatial structure (PSS) of best model for each site.

Site	Best Model	Effective Range	Partial sill	Nugget	PSS
1	Gau	124.5	0.761	0.178	0.81
2	Gau	88.7	0.518	0.098	0.84
3	Sph	111.9	0.348	0.092	0.79
4	Sph	92.6	0.393	0.076	0.84

3.5.2. Analysis models

Based on AIC, spatial models outperformed the null model (table 2). This advantage of spatial models was expected, given the strong spatial correlation found in the uniformity trial, prior to adding treatment effects in the simulations. Differences in median AIC value between SPH and GAU were much smaller than the differences between either of these spatial models and the NULL for all regions (e.g. for t = 36 in figure S2).

An agreement was found between the semivariogram model chosen in each site of the uniformity trial and the best fitting model according to AIC after the simulation process (table 1 and table 2). In site 2, SPH model had serious convergence problems, converging on average in 23 % of the cases, except for RCBD with 4 treatments (see details of maximum iterations and optimization technique used in appendix 1). Beyond this, in the cases where both spatial models converged, the proportion of times that the SPH model was selected never exceeded 36 %. Similar behavior was observed when blocks were not included for ALPHA and RCBD design (data not shown).

Table 2. Number of times each model including complete block was the best according to AIC (# times) for all sites (1, 2, 3 and 4). The maximum #times for each design and number of treatments (t) is indicated in bold type.

			#times			
t	Design	Model	1	2	3	4
48	Alpha	Null	0	1	0	0
		Sph	252	25	**582**	**545**
		Gau	**748**	**974**	418	455
	RCBD	Null	0	0	332	0
		Sph	263	32	**337**	**822**
		Gau	**737**	**968**	331	178
	CRD	Null	1	1	0	0
		Sph	259	43	**519**	**648**
		Gau	**740**	**956**	481	352
36	Alpha	Null	0	0	0	0
		Sph	208	11	**553**	**551**
		Gau	**792**	**989**	447	449
	RCBD	Null	0	0	0	0
		Sph	234	21	**626**	**812**
		Gau	**766**	**979**	374	188
	CRD	Null	0	1	0	0
		Sph	278	30	**540**	**661**
		Gau	**722**	**969**	460	339
4	RCBD	Null	0	0	0	0
		Sph	0	0	**924**	**570**
		Gau	**1000**	**1000**	76	430
	CRD	Null	0	0	0	0
		Sph	88	0[†]	**952**	**1000**
		Gau	**912**	**1000**	48	0

[†] Only one iteration converged

3.5.3. Bias of standard errors and control of empirical type 1 error rate

In figure 2, the ratio between SED_est and SED_obs are presented for all models that included complete blocks. Some deviations from the expected value 1 can be observed in some cases, but the most noticeable deviations are clearly in site 2, for the SPH model combined with CRD and RCBD, where values significantly larger than 1 can be observed (mostly for t = 36). Ratio values greater than 1 correspond to a positive relative bias (table 3). For this model in other sites, RB never exceeded 11 %.

In general, the NULL model including blocks did not show bias of SED_est as expected. The GAU model had fewer bias problems than the SPH model. Also, differences among sites can be observed, with sites that in general all models had smaller bias (site 4, table 3) and sites where biases were larger (site 1 or 2, table 3). The lack of control of ET1 is observed fundamentally in those cases where the biases are high in absolute value, although there are some exceptions with low biases, where ET1 was not well-controlled.

The number of replications affected the SED_est bias, since there were no ET1 control problems with a high number of replications (r = 36), except for one case (table 3).

For those cases where there was lack of control of ET1, the GAU model tended to lead to more liberal inferences (above upper limit of CV(95 %), while SPH depends on the site. For site 1 and 2, SPH model had always conservative type I error rates corresponding to high and positive bias. For site 3 and 4 where SPH was the best fitting model according to AIC, ET1 tended to be liberal (data not shown).

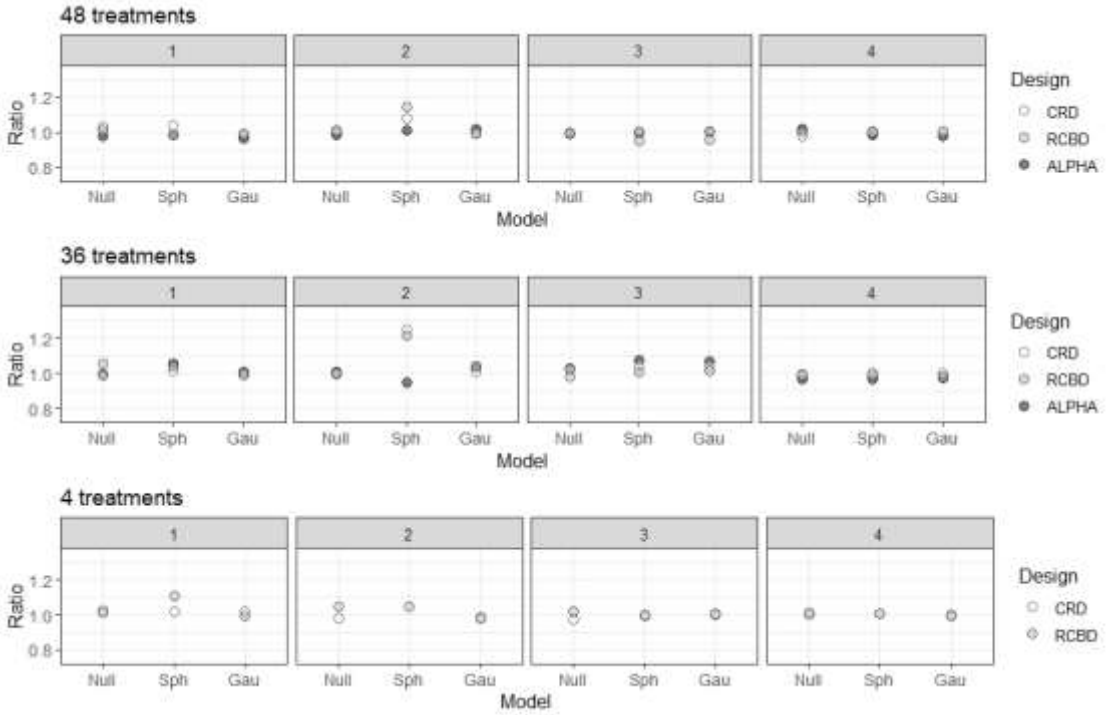


Figure 2. Ratio between mean estimated standard error of differences (SED_est) and mean observed standard error of differences (SED_obs) for each design and analysis model for 48 treatments (top), 36 treatments (middle) and 4 treatments (bottom). Each panel represents one of the four sites (1, 2, 3, 4).

Table 3. Relative bias (RB%) of mean estimated standard error of differences for different combinations of number of treatments (t) and replications (r), design and analysis model with complete blocks in all sites (1, 2, 3, 4). Lack of control of the nominal type I error rate is indicated in bold type.

				RB (%)			
t	r	Design	Model	1	2	3	4
48	3	Alpha	Null	-2.003	-1.202	-0.853	1.758
			Sph	-1.646	**1.199**	**-0.124**	-1.760
			Gau	**-3.589**	1.729	**0.179**	-2.051
		RCBD	Null	1.182	0.176	0.027	0.201
			Sph	-1.329	**14.192**	**-4.656**	0.077
			Gau	**-1.091**	-0.401	**-4.271**	0.403
		CRD	Null	**3.425**	1.450	-0.888	-2.523
			Sph	3.932	**7.854**	0.367	0.771
			Gau	**-1.741**	-1.050	0.505	0.455
36	4	Alpha	Null	-0.318	1.011	3.159	**-3.402**
			Sph	5.415	**-5.103**	7.469	**-2.891**
			Gau	1.115	3.336	**7.120**	**-2.100**
		RCBD	Null	5.400	0.357	-1.497	-0.935
			Sph	**4.066**	**21.812**	1.138	**-1.097**
			Gau	0.352	4.539	1.388	**-1.822**
		CRD	Null	-1.321	-0.717	2.010	-0.595
			Sph	1.767	**24.830**	3.992	0.213
			Gau	-0.982	1.088	3.983	0.213
4	36	RCBD	Null	1.116	4.666	2.140	1.472
			Sph	**10.935**	4.666	0.054	0.484
			Gau	-0.412	-2.110	0.451	-0.316
		CRD	Null	2.599	-1.900	-2.498	-0.025
			Sph	2.271	. [†]	-0.677	0.750
			Gau	2.068	-1.295	0.052	0.325

[†] Since only one iteration converged, no SED_obs and bias could be computed.

If complete blocks of RCBD and ALPHA design were not included in the analysis, the performance of the NULL model got worse, specially for RCBD design (table 4). For sites 3 and 4, no matter the number of treatment or replications, NULL model without blocks had problems of SED bias and also control of ET1. For 36 replications, these problems were observed in all sites. In all of those cases, at least one of the spatial models improved the RCBD performances, achieving good control of ET1 and eliminating the bias problem in almost any cases. For the ALPHA design, the NULL model without complete blocks had almost no bias or ET1 issues (table S3).

Table 4. Relative bias (RB%) of mean estimated standard error of differences for all analysis model, number of treatments (t) and replications (r), for RCBD design **without blocks** in all sites (1,2,3,4). Lack of control of the nominal type I error rate is indicated in bold type.

			RB%			
t	r	Model	1	2	3	4
48	3	Null	1.182	1.004	**17.647**	**4.954**
		Sph	-2.113	**64.639**	**-4.776**	0.118
		Gau	**-1.148**	-0.243	**-4.464**	0.117
36	4	Null	4.921	1.329	**15.060**	**2.162**
		Sph	**4.710**	**23.982**	0.939	-1.230
		Gau	1.297	4.403	1.878	-1.058
4	36	Null	**33.244**	**14.808**	**29.117**	**28.575**
		Sph	0.214	**14.808**	0.261	3.918
		Gau	0.907	2.849	-0.313	**8.068**

3.5.4. Power of t-test

In figure 3, empirical power curves for all design and models in site 4 are presented (for t = 48 and t = 36). For t = 4 we did not include power curves because there are only 6 possible comparisons. The general behavior of the three models were similar

for the other sites (data not shown), but we decided to display site 4, since it was the one that had the least control problems of ET1. The advantages of spatial models over NULL are clearly observed. The vertical dotted line represents a value of $\Delta = 1$ (true mean difference of 1 Mkg ha^{-1}) and was added to aid in the comparison of models at a certain value of Δ that represents a significant yield difference to be declared different from 0. For this certain difference, spatial models reach in some cases 50 % more power than the NULL one. No clear difference between GAU and SPH could be observed in any case. Comparing only NULL model performance among designs, it can be observed that the more complex the design (greater local control), the greater power it has. Additionally, it can be confirmed the effect of higher error degrees of freedom with t = 36 respect to t = 48, resulting in more power (for example, 105 vs. 94 degrees of freedom for RCBD, respectively).

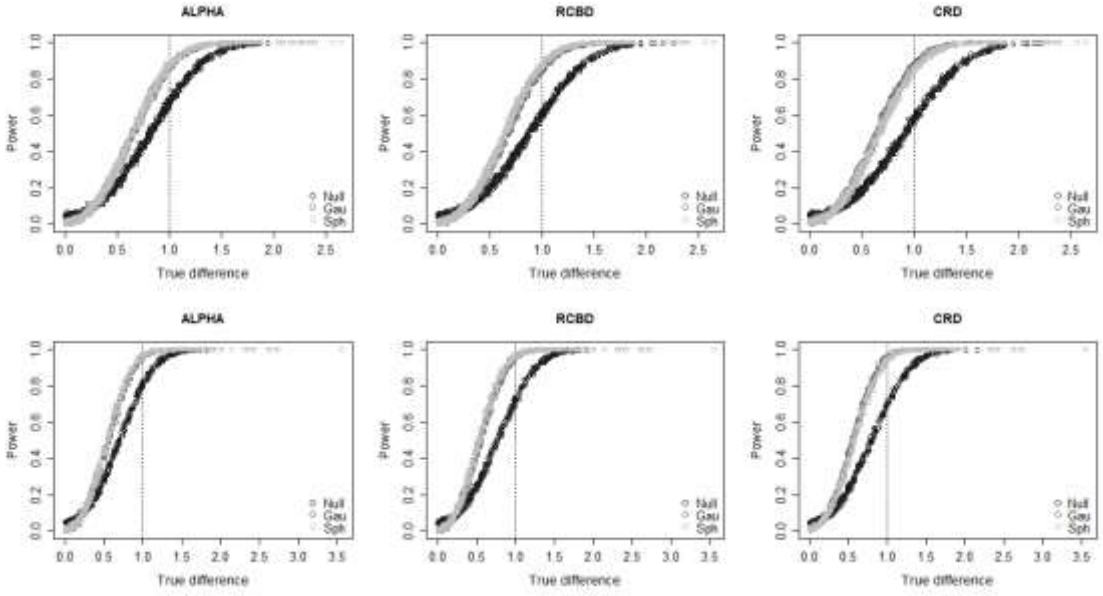


Figure 3. Power (P[Reject H0|H0 false]) for each design and analysis model with 48 treatments (top) and 36 treatments (bottom) in Site 4. Vertical dotted line in 1 Mkg ha^{-1} true mean differences is added to highlight power for a certain difference

3.6. DISCUSSION

3.6.1. Pre and post spatial variability

A strong spatial correlation of wheat yield of the uniformity trial was confirmed. The range parameter for all models and sites was always above 88 m. In this context, it was expected that the spatial models could outperform the independent error model in terms of precision and model fit. Although differences among model residual sum of squares were small, in site 1 and 2 the GAU model was preferred, while for site 3 and 4, the preferred model was the SPH. After simulation process, the structure of spatial variability did not change and the models selected for each site were the same, based on the number of times each model had the lowest AIC value. Considering the scenario with 36 replications per treatment (t = 4), it can be assumed that the real spatial pattern is well captured or estimated due to the high number of replications. In our work, for all number of replications considered, the same spatial model was selected based on AIC criterion. Also, it is noteworthy that the different spatial arrangements posed by the different experimental designs did not change the spatial pattern detected prior to the simulations. We also computed the Bayesian Information Criteria (BIC) and best fitting models were the same for each site. We preferred AIC as a model selection criterion, recommended by Guerin and Stroup (2000). In the context of repeated measures analysis, they found that AIC tends to select a more complex model but with better type I error rate control than BIC.

3.6.2. Bias of standard error and empirical type I error rate

Considering only those treatments with equal effects, the estimated mean differences were always very close to zero (data not shown), as it was expected, because in all cases treatments were randomly assigned and the EBLUE (empirical best linear unbiased estimation) guaranteed unbiasedness of fixed effect estimations (Kackar and Harville, 1984). Beyond the presence or absence of bias, it is noticeable the increasing precision of the spatial models for any of the designs (table S2), having always lower SED_obs and SED_est. Also, comparing the NULL models for all experimental designs, the ALPHA has the smallest standard errors of the differences.

Similar results were showed and discussed in chapter 2 (Borges et al., 2019), where the advantages in precision of the more complex designs and/or spatial models were observed for moderate and big experiment size. When a spatial model was used, the ALPHA was not much better than the other designs, considering this size of experiment (144 plots).

The bias of SED_est for the NULL model was almost non-existent, when blocks were included (RCBD and ALPHA). This was to be expected since proper randomization was used (Piepho et al., 2013). As they stated, "the main advantages of randomization is that it provides protection against all types of underlying experimental structure". However, Hu et al. (2006) found significant bias in RCBD basic model (including blocks) in the presence of important spatial covariances, leading to considerable lack of control of ET1. Starting from different simulated scenarios, in those cases with strong spatial correlation (high values of PSS) a high relative bias were observed for those pair of treatments that were located at very small or high distances, since NULL model assumes the same estimation error for all contrasts, leading to overestimation or underestimation of SED (for small and high distances, respectively). On the other hand, when the blocks were not considered in the analysis, the NULL model presented some problems of bias, in particular for RCBD design (table 4). Regarding the ALPHA design, the bias and control problems of ET1 in the NULL model were almost non-existent (table S3). Since it still has incomplete block adjustments implictly present in the analysis model, it seems that these blocks are soaking up the variability between the complete blocks, pooling all the variability among main blocks into the variability among incomplete blocks.

Regarding to the spatial models, they showed different performances. The GAU model did not present severe problems of bias, never exceeding 7 %, even in those sites where it was not the preferred model. Although the magnitude of bias was relatively small, in a few cases it did not control ET1 correctly. In those cases, ET1 tended to be liberal (6 out of 8). Other works observed the same liberal trend of ET1 for Gaussian model (Richter et al., 2015, Richter and Kroschewski, 2012).

On the other hand, the SPH showed very different responses depending on the site considered. In site 2, where it also had serious convergence problems, it had large positive relative bias of SED_est. These high positive biases corresponded to a lack of control of ET1, which led to more conservative inferences. For the remaining sites, there were some combinations of design and number of treatments that showed some moderate bias, always below 11%. In those situations where a lack of control of ET1 was verified, the inferences were liberal in sites where it was the preferred model (sites 3 and 4) and conservative in those sites where it was not (site 1 and 2). Starting from simulated scenarios, Richter et al. (2015) found that the degree and sign of bias depends on the combination of simulation and analysis model and on the range parameter value used for simulations. When SPH model was both the simulation and analysis model and spatial correlation was strong (range parameter greater than 20 m), this model had minimal negative bias and/or problems with control ET1. On the other hand, when GAU was the true model, the use of SPH as analysis model led to severe positive bias for range parameter values greater than 4 m. According to these authors, in this scenario the SPH model is not able to reproduce the sigmoidal shape of the covariance function, leading to high positive biases. In our case, we do not know what the true covariance structure is, but taking into account that the preferred model for the uniformity trial data was GAU for some sites, possibly what these authors found is what could explain the high positive biases of the SPH model in site 2. It is important to keep in mind that some of the treatments involved in these comparisons may have been located in adjacent plots, at a short distance from each other, and it is precisely at short distances where SPH and GAU differ the most. Therefore, the contributions that these nearby plots make to the SED_est calculations could be part of the explanation for the discrepancy in SED_est when working with the wrong model.

Considering the cases that complete blocks were not included in the analysis, at least one of the spatial models improved the RCBD performance, achieving good control of ET1 and eliminating the bias problem. The spatial model that improved this

behavior depended on the site (in general GAU for sites 1 and 2 and SPH for sites 3 and 4). Therefore, the effect of reducing bias and adequately control ET1 depends on the quality of the spatial model. Nevertheless, spatial models can not be relied on for solving all the problems when blocks are not included in the analysis, but it is still recommended to include them if they are part of the experimental design, considering not only these problems of bias but also the gain in efficiency, reducing the SED_{est} when blocks are included.

The number of replications had a marked effect on SED_est bias and lack of control of ET1. With 36 repetitions per treatment (t = 4), the bias was minimal and ET1 was well-controlled, while with 3 or 4 replicates those problems appeared, even with the preferred model (although the preferred model had less problems than the other spatial model). It is important to notice that in the independent error model with few replications, there is a substantial negative correlation between residuals in the same treatment because of the restriction that residuals of the same treatments must sum 0. This negative correlation is much smaller when the number of replicates is high. With spatially correlated error models, the computation of covariance between residuals of the same treatment became much harder, and this negative correlation may interfere with the estimation of the real spatial covariance. Hu et al. (2006) wondered whether spatial models can be employed with databases with few replications per treatment, with an acceptable estimation of errors in the presence of spatial correlation. We show that the answer is generally no, at least for these fields. To be useful in practice, the frequently selected correlation models should control ET1. This is not consistently true for all sites, with the exception of site 2 where the model selected by AIC (GAU) does not present problems for any combination of number of treatments and repetitions. Although it is also true for site 4 with 3 replications, it is not for that site with 4 replications. With 36 replications, a much more precise estimate of treatment means and the standard errors of the difference could be obtained, therefore less induced correlation.

Considering the approximation method for mixed models, the pros and cons of each

have been reported in various works previously and they depend on several factors, such as data covariance structure, sample size, degree of imbalance and experimental design among others (Luke, 2017, Spilke et al., 2005). Evaluating significance in linear mixed-effects models and control of empirical type 1 error rate, Luke (2017) found that using Kenward-Roger or Satterthwaite approximations using REML as estimation method produced consistent type 1 error rate (neither conservative nor liberal), even with small sample sizes. However, this author warns that situations with more complex covariance structures combined with small sample sizes can lead to inflate type 1 error rates. The smallest sample size used in his work was 144 observations, which is consistent with the sample size used in our work, but the main difference is that in our work the covariance structure is more complex. Spilke et al. (2005) also report advantages in control of ET1 using KR over other correction methods, with unbalanced data using different experimental designs, but the authors warns that this method should be used with particular caution when the covariance structure is more complex. Other authors found that using KR1 (Kenward-Rogers first order correction) as an estimation method in spatial models, control of ET1 was satisfactory for t-test and results obtained were considered reliable (Richter et al., 2015, Richter and Kroschewski, 2012).

The KR approximation method (no matter which version of KR is considered) has two components: one has to do with the correction of the degrees of freedom and the other with an asymptotic correction of the variance-covariances matrix (Kenward and Roger, 1997, Kenward and Roger, 2009). Both components affect ET1, while only the second affects the SED bias. Since we do not know which is the real spatial model that describes the field heterogeneity of the uniformity trial, for those scenarios with problems of bias and control of ET1 we cannot conclude if it they arise because the spatial model is wrong or if it is something systematic of the KR approximation of the variance-covariance matrix of fixed effects.

3.6.3. Power

The potential advantages of spatial models over NULL are clear in all sites and for all experimental designs. Mainly in moderate mean differences is where these advantages can be seen more clearly. This superiority in power was observed in previous work (Hu et al., 2006, Hu and Spilke, 2009, Richter et al., 2015). However, it must be kept in mind that there are cases where ET1 was not adequately controlled, so the power obtained in those cases would not be statistically valid. Anyway, Hu et al. (2006) and Richter et al. (2015) computed 'corrected' power using significant thresholds obtained from simulations under the null hypothesis and they found that the best fitting spatial model had always more power than classical RCBD analysis or incomplete-block analysis.

We did not compute corrected power, but if we choose a scenario where there were no problems of lack of control of ET1 for any model, such as site 4 with 48 treatments (figure 3, top panel), it can be observed, in addition to the advantage of the two spatial models already mentioned, the advantage of experimental designs with greater local control. If we pick the NULL model of this scenario for the three experimental designs, we see that for Δ=1 Mkg ha^{-1}, the ALPHA has the highest power, followed by the RCBD and, finally, CRD. The power advantages of ALPHA over RCBD were also observed by Richter et al. (2015) for scenarios with strong spatial correlation (range parameter greater than 20 m). Additionally, it can be noticed that the gains in power of spatial models over the NULL are more remarkable in the simplest design (CRD) than in the ALPHA.

As it was mentioned before, we only used KR first order correction method, but there are previous studies that evaluate the effect of using different correction method over power when using mixed models (Hu and Spilke, 2009, Luke, 2017, Richter et al., 2015, Spilke et al., 2004, Spilke et al., 2005). In these works, KR methods (KR, KR1 or KR2) had a good performance, controlling ET1 without decreasing power.

3.6.4. Conclusion remarks

In summary, the gain in precision that can be obtained with the use of spatial models in field trials is not always accompanied by valid inferences, with adequate control of ET1 and reliable power. This depends on being able to adequately estimate the underlying spatial pattern (that is to choose the correct variance-covariance structure), but also on other factors that must be taken into account when analyzing the data. As Cox (2009) argues: "the reality of any apparent gain in precision depends on the adequacy of the assumed model". If the underlying spatial pattern of the field is not known (most common situation in agricultural experiments), it has to be estimated using different techniques and selection criteria. In making decisions about how to analyze data, there are a number of key steps to minimize the risks of not being able to draw valid conclusions that depend fundamentally on the structure of the data, the experimental design and the number of replications per treatment. If spatial correlation of the response variable is presumed to be strong, then the election of an adequate experimental design that accounts for this heterogeneity will be crucial. In addition, the use of spatial models can lead to more power without problems of bias or lack of control of ET1 if the covariance structure of the model adequately describes the spatial pattern of the data. As we mentioned, since the true spatial structure of the data is unknown, the best spatial model has to be chosen according to some model fit criteria. In our case, the AIC seems to work well, since the best model chosen for each site according to this criterion had much less problems of bias or control of ET1, even when complete blocks were not included in the analysis. Finally, the number of repetitions per treatment is also a very important factor to consider when using spatial models. With high number of replications, no spatial model had the aforementioned problems even if it was not the preferred spatial model, because a much more precise estimate of treatment means and standard errors of mean differences could be obtained.

3.7. REFERENCES

Borges A, González-Reymundez A, Ernst O, Cadenazzi M, Terra J, Gutiérrez L. 2019. Can spatial modeling substitute for experimental design in agricultural experiments?. Crop Science. *59*(1): 44-53. https://doi.org/10.2135/cropsci2018.03.0177

Brownie C, Bowman DT, Burton JW. 1993. Estimating Spatial Variation in Analysis of Data from Yield Trials: A Comparison of Methods. Agronomy Journal. *85*: 1244–1253.

Brownie C, King LD, Dube TJ. 2004. Longitudinal and Spatial Analyses Applied to Corn Yield Data from a Long-Term Rotation Trial. Institute of Statistics Mimeo Series. *2559*: 1–27.

Casler MD. 2015. Fundamentals of experimental design: Guidelines for designing successful experiments. Agronomy Journal. *107*(2): 692–705. https://doi.org/10.2134/agronj2013.0114

Casler MD, Undersander DJ. 2000. Forage yield precision, experimental design, and cultivar mean separation for alfalfa cultivar trials. Agronomy Journal. *92*(6): 1064–1071. https://doi.org/10.2134/agronj2000.9261064x

Cochran WG, Cox GM. 1992. Completely Randomized, Randomized Block, and Latin SquareDesigns. En: Wiley J. (Ed.). Experimental Designs (2nd ed.). New York: Wiley (Wiley Classics Library). 95-147

Cox DR. 2009. Randomization in the design of experiments. International Statistical Review. *77*(3): 415–429. https://doi.org/10.1111/j.1751-5823.2009.00084.x

Cullis, ABR, Gleeson AC. 1991. Spatial Analysis of Field Experiments-An Extension to Two Dimensions. Biometrics. *47*(4): 1449–1460.

de Mendiburu F, Yaseen M. 2020. *agricolae: Statistical Procedures for Agricultural Research* (R package version 1.4.0). https://myaseen208.github.io/agricolae/https://cran.r-project.org/package=agricolae.

Fisher RA. 1935. The principles of experimentation, illustrated by a pshico-physical experiment. En: Oliver and Boyd (Eds). *The Design of Experiments.* New York:

Hafner Press, 11-26.

Gilmour, AR, Cullis BR, Verbyla AP. 1997. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. Journal of Agricultural, Biological, and Environmental Statistics. *2*(3): 269–293.

Gleeson AC, Cullis BR. 1987. Residual Maximum Likelihood (REML) Estimation of a Neighbour Model for Field Experiments. Biometrics. *43*(2): 277–287.

Grondona M, Crossa J, Fox P, Pfeiffer W. 1996. Analysis of variety yield trials using two-dimensional separable ARIMA processes. Biometrics. *52*(2): 763–770. http://www.jstor.org/stable/10.2307/2532916

Guerin L, Stroup WW. 2000. A Simulation Study To Evaluate Proc Mixed Analysis of Repeated Measures Data. En: Conference on Applied Statistics in Agriculture (12th, 2000, Manhattan, Kansas). Proceedings. Kansas State University Libraries. 170–203. https://doi.org/10.4148/2475-7772.1249

Hu X, Spilke J. 2009. Comparison of various spatial models for the analysis of cultivar trials. New Zealand Journal of Agricultural Research. *52*(3): 277–287. https://doi.org/10.1080/00288230909510512

Hu X, Spilke J, Richter C. 2006. The influence of spatial covariances on the type I error and the power for different evaluation models. Biometrical Letters. *43*(1): 19–37. papers3://publication/uuid/CA1A021D-3BE5-4AE9-989A-8047B3A480BD

Kackar RN, Harville DA. 1984. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. Journal of the American Statistical Association. *79*(388): 853–862. https://doi.org/10.1080/01621459.1984.10477102

Kenward MG, Roger JH. 1997. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics. *53*(3): 983–997.

Kenward MG, Roger JH. 2009. An improved approximation to the precision of fixed effects from restricted maximum likelihood. Computational Statistics & Data Analysis. *53*(7): 2583–2595. https://doi.org/10.1016/j.csda.2008.12.013

Kravchenko AN, Robertson GR, Snap SS, Smucker AJM. 2006. Using Information about Spatial Variability to Improve Estimates of Total Soil Carbon. Agronomy

Journal. *98*(3): 823–829. https://doi.org/10.2134/agronj2005.0305

Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology. *74*(6): 1659–1673. http://www.esajournals.org/doi/abs/10.2307/1939924

Lenth R. 2020. emmeans: Estimated Marginal Means, aka Least-Squares Means (R package version 1.4.5.). https://cran.r-project.org/package=emmeans

López MV, Arrúe JL. 1995. Efficiency of an Incomplete Block Design Based on Geostatistics for Tillage Experiments. Soil Science Society of America Journal. *59*(4): 1104–1111. https://doi.org/10.2136/sssaj1995.03615995005900040023x

Luke SG. 2017. Evaluating significance in linear mixed-effects models in R. Behavior Research Methods. *49*(4): 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Mallarino A, Bermudez M, Wittry D, Hinz P. 2000. Alternative Data Managements and Intepretations. En: International Conference on Site-Specific Management for Agricultural Systems (5th, 2000, Madison, Wisconsin). Proceedings. Robert PC et al (ed) . CD-ROM. July 16- 19. Bloomington, MN. ASA, CSSA, SSSA, Madison, WI

Oliver MA, Webster R. 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena. *113*: 56–69. https://doi.org/10.1016/j.catena.2013.09.006

Patterson HD, Williams ER. 1976. A new class of resolvable incomplete block designs. Biometrika. *1963*(1): 83–92. https://doi.org/https://doi.org/10.1093/biomet/63.1.83

Pebesma EJ, Bivand RS. 2005. Classes and methods for spatial data in R. R News 5 (2), https://cran.r-project.org/doc/Rnews/.

Pebesma EJ. 2004. Multivariable geostatistics in S: The gstat package. Computers and Geosciences. *30*(7): 683–691. https://doi.org/10.1016/j.cageo.2004.03.012

5thexperiments? Journal of Agronomy and Crop Science. *199*(5): 374–383. https://doi.org/10.1111/jac.12026

Piepho HP, Williams ER, Michel V. 2015. Beyond Latin squares: A brief tour of row-column designs. Agronomy Journal. *107*(6): 2263–2270. https://doi.org/10.2134/agronj15.0144

Qiao CG, Basford KE, DeLacy IH, Cooper M. 2000. Evaluation of experimental designs and spatial analyses in wheat breeding trials. Theoretical and Applied Genetics. *100*: 9–16.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.

Richter C, Kroschewski B. 2012. Geostatistical models in agricultural field experiments: Investigations based on uniformity trials. Agronomy Journal. *104*(1): 91–105. https://doi.org/10.2134/agronj2011.0100

Richter C, Kroschewski B, Piepho HP, Spilke J. 2015. Treatment comparisons in agricultural field trials accounting for spatial correlation. Journal of Agricultural Science. *153*(7): 1187–1207. https://doi.org/10.1017/S0021859614000823

SAS Institute. 2013. Base SAS® 9.4 Procedures Guide: Statistical Procedures. Second edition. SAS Institute Inc. Cary, NC, USA. 550 p.

Saud P, Lynch TB, Anup KC, Guldin JM. 2016. Using quadratic mean diameter and relative spacing index to enhance height-diameter and crown ratio models fitted to longitudinal data. Forestry. *89*(2): 215–229. https://doi.org/10.1093/forestry/cpw004

Spilke J, Piepho HP, Hu X. 2005. A simulation study on tests of hypotheses and confidence intervals for fixed effects in mixed models for blocked experiments with missing data. Journal of Agricultural, Biological, and Environmental Statistics. *10*(3): 374–389. https://doi.org/10.1198/108571105X58199

Spilke J, Piepho HP, Meyer U. 2004. Approximating the degrees of freedom for contrasts of genotypes laid out as subplots in an alpha-design in a split-plot experiment. Plant Breeding. *123*(2): 193–197. https://doi.org/10.1046/j.1439-0523.2003.00964.x

Williams ER, John JA, Whitaker D. 2006. Construction of resolvable spatial row-column designs. Biometrics. *62*(1): 103–108. https://doi.org/10.1111/j.1541-0420.2005.00393.x

Williams ER, Matheson AC, Harwood CE. 2002. Experimental Design and Analysis for Tree Improvement. En: CSIRO Publishing. Experimental Design and

Analysis for Tree Improvement. Melbourne: CSIRO. 214

Williams ER, Piepho HP. 2013. A comparison of spatial designs for field variety trials. Australian and New Zealand Journal of Statistics. *55*(3): 253–258. https://doi.org/10.1111/anzs.12034

Wood SN. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Methodological). *73*(1): 3–36.

Wu T, Dutilleul P. 1999. Validity and efficiency of neighbor analyses in comparison with classical complete and incomplete block analyses of field experiments. Agronomy Journal. *91*(4): 721-731. https://doi.org/10.2134/agronj1999.914721x

Zimmerman DL, Harville DA. 1991. A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments. Biometrics. *47*(1): 223-239. https://doi.org/10.2307/2532508

3.8. SUPPLEMENTAL MATERIAL

Table S1. List of scenarios analyzed for each site, resulting from different combinations of number of treatments (t), replications (r), experimental design (Design), analysis model (Model) and inclusion of complete blocks (+ including, - not including, --- not applicable).

t	r	Design	Model	Blocks
48	3	ALPHA	NULL	+
				-
			GAU	+
				-
			SPH	+
				-
		RCBD	NULL	+
				-
			GAU	+
				-
			SPH	+
				-
		CRD	NULL	---
			GAU	---
			SPH	---
36	4	ALPHA	NULL	+
				-
			GAU	+
				-
			SPH	+
				-
		RCBD	NULL	+
				-
			GAU	+
				-
			SPH	+
				-
		CRD	NULL	---
			GAU	---
			SPH	---
4	36	RCBD	NULL	+
				-
			GAU	+
				-
			SPH	+
				-
		CRD	NULL	---
			GAU	---
			SPH	---

Table S2. Mean observed standard error of differences (_obs) and mean estimated standard error of differences (_est), for different combinations of number of treatments (t) and replications (r), design and analysis model with complete blocks for all sites (1, 2, 3, 4).

				1		2		3		4	
t	r	Design	Model	_obs	_est	_obs	_est	_obs	_est	_obs	_est
48	3	Alpha	Null	0.742	0.727	0.559	0.552	0.449	0.445	0.402	0.409
			Sph	0.574	0.565	0.330	0.334	0.319	0.319	0.322	0.316
			Gau	0.570	0.549	0.310	0.315	0.321	0.322	0.328	0.321
		RCBD	Null	0.831	0.840	0.558	0.559	0.317	0.317	0.443	0.444
			Sph	0.572	0.564	0.291	0.333	0.333	0.317	0.316	0.316
			Gau	0.555	0.549	0.318	0.317	0.335	0.320	0.326	0.327
		CRD	Null	0.808	0.836	0.553	0.561	0.520	0.515	0.470	0.458
			Sph	0.546	0.568	0.307	0.331	0.314	0.315	0.313	0.316
			Gau	0.559	0.549	0.320	0.317	0.316	0.317	0.324	0.326
36	4	Alpha	Null	0.622	0.620	0.472	0.477	0.370	0.382	0.362	0.349
			Sph	0.453	0.477	0.293	0.278	0.252	0.271	0.271	0.263
			Gau	0.452	0.457	0.256	0.265	0.255	0.274	0.276	0.270
		RCBD	Null	0.689	0.726	0.481	0.483	0.400	0.394	0.392	0.389
			Sph	0.455	0.473	0.227	0.277	0.267	0.270	0.266	0.263
			Gau	0.458	0.460	0.258	0.269	0.269	0.273	0.277	0.272
		CRD	Null	0.733	0.723	0.490	0.486	0.437	0.445	0.400	0.398
			Sph	0.467	0.475	0.223	0.278	0.259	0.269	0.270	0.271
			Gau	0.471	0.466	0.267	0.270	0.261	0.271	0.281	0.281
4	36	RCBD	Null	0.182	0.184	0.141	0.148	0.115	0.118	0.103	0.105
			Sph	0.136	0.151	0.141	0.148	0.088	0.088	0.083	0.083
			Gau	0.148	0.148	0.084	0.082	0.088	0.089	0.084	0.084
		CRD	Null	0.235	0.241	0.165	0.162	0.152	0.148	0.132	0.132
			Sph	0.148	0.152	.[†]	0.086	0.088	0.087	0.087	0.087
			Gau	0.146	0.149	0.088	0.086	0.088	0.088	0.092	0.092

[†] Since only one iteration converged, no SED_obs could be computed.

Table S3. Relative bias (RB%) of mean estimated standard error of differences for all analysis model, number of treatments (t) and replications (r), for ALPHA design **without blocks** in all sites (1,2,3,4). Lack of control of the nominal type I error rate is indicated in bold type.

			RB%			
t	r	Model	1	2	3	4
48	3	Null	-0.443	-2.045	-0.588	0.298
		Sph	-1.407	**7.741**	**0.027**	-2.238
		Gau	**-3.272**	1.793	0.662	-2.921
36	4	Null	-3.737	0.189	**2.362**	-1.375
		Sph	**3.104**	**-1.478**	7.503	**-3.223**
		Gau	1.307	3.172	**6.876**	**-1.838**

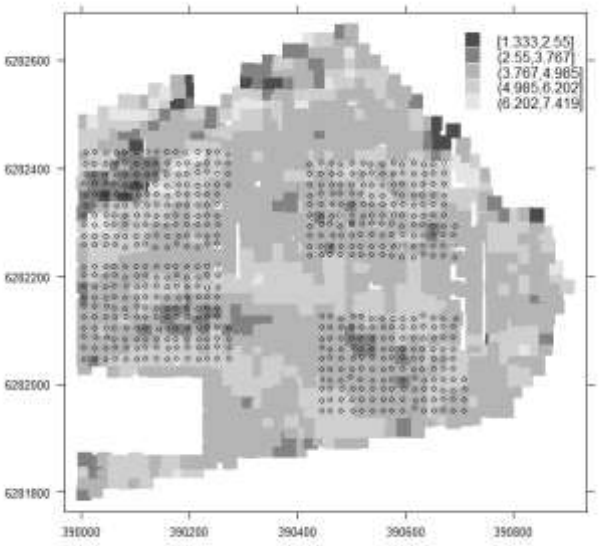


Figure S1. Spatial variability of corrected yield (yield data minus regression splines predictions) and geographic location of the experimental units for all sites (from site 1 upper left to site 4 counterclockwise).

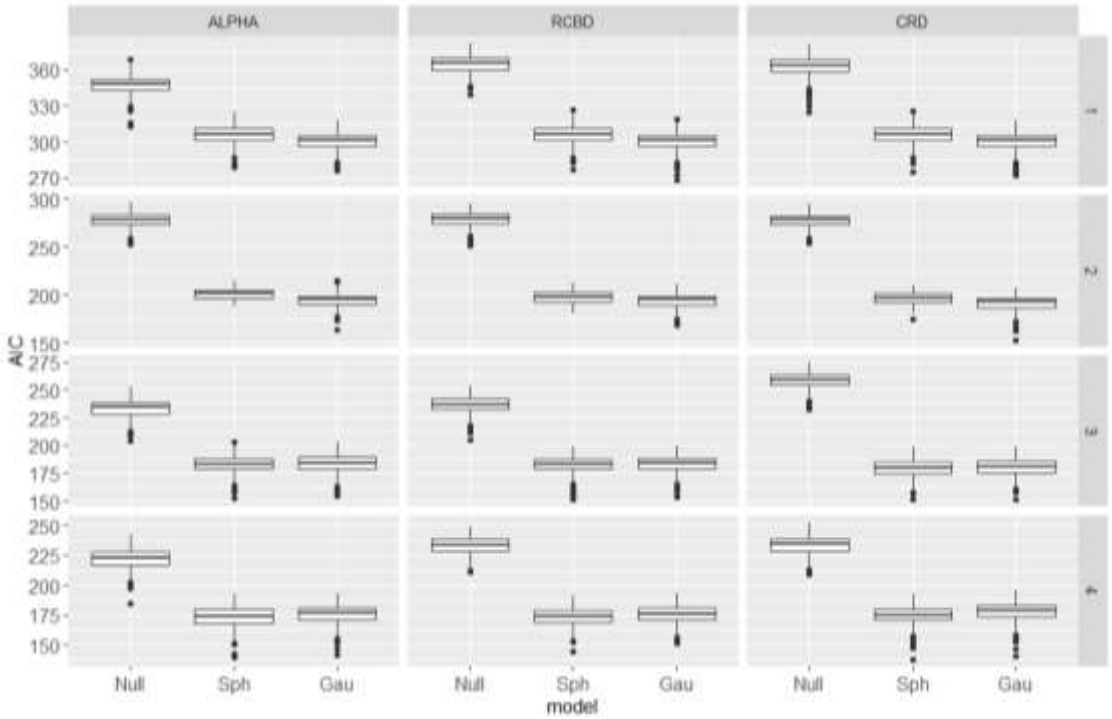


Figure S2. Boxplot of AIC values for all models and experimental design for each site (1, 2, 3, 4), for 36 treatments and 4 replications.

Appendix 1

Proc MIXED statement for Spherical model with ALPHA design, 48 treatments with 3 complete replications, being k the number of iteration. The maximum number of iterations was set in 5000 and the optimization method implemented in PROC MIXED is a ridge-stabilized Newton-Raphson algorithm:

```
proc mixed data=alpha48 maxiter=5000 ; by k;
class rep trat block;
model yield= rep trat /ddfm=kr(firstorder);
random block(rep) ;
repeated /subject=intercept type=sp(sph)(x y) local rcorr;
lsmeans d/  pdiff  ;
ods output CovParms=covp diffs=pp fitstatistics=fit tests3=anova ;
run;
```

4. SPATIO-TEMPORAL MODELING AND COMPETITION DYNAMICS IN FOREST TILLAGE EXPERIMENTS ON EARLY GROWTH OF *EUCALYPTUS GRANDIS* L.[2]

Pablo González Barrios†, Alejandra Borges†, José Terra, Mario Pérez Bidegain, and Lucía Gutiérrez

†The first two authors made equal contributions to the manuscript.

Pablo González Barrios (pgonzalez6@wisc.edu), Alejandra Borges (aborges@fagro.edu.uy), and Lucía Gutiérrez (gutierrezcha@wisc.edu), Statistics Department, Facultad de Agronomía, Universidad de la República, Uruguay. José Terra (jterra@tyt.inia.org.uy), National Agricultural Research Institute (INIA), Uruguay. Mario Pérez Bidegain (mperezb@fagro.edu.uy), Soils Department, Facultad de Agronomía, Universidad de la República, Uruguay. Lucía Gutiérrez (gutierrezcha@wisc.edu), Department of Agronomy, University of Wisconsin, Madison.

4.1. RESUMEN

Los experimentos de labranza forestal utilizan regularmente evaluaciones a largo plazo en grandes parcelas que generan correlaciones temporales y/o espaciales entre las observaciones. No modelizar estas correlaciones podría comprometer las comparaciones entre tratamientos. El objetivo de este estudio fue evaluar el efecto de la modelización de la variabilidad espacio-temporal (ST) en los experimentos de laboreo forestal, así como evaluar el efecto de la mortalidad intraparcelaria y la dinámica de la competencia entre árboles. Utilizamos diferentes estrategias que incorporan correlaciones espaciales y/o temporales en la evaluación del efecto de la

[2] González-Barrios, P., Borges, A., Terra, J., Bidegain, M. P., and Gutiérrez, L. (2020). Spatio-temporal modeling and competition dynamics in forest tillage experiments on early growth of eucalyptus grandis L. *Forest Science*, *66*(5). https://doi.org/10.1093/forsci/fxaa007

intensidad de laboreo en el crecimiento inicial de *Eucalyptus*. Se compararon el efecto de tres intensidades de laboreo en dos condiciones de suelo contrastantes sobre la altura del árbol y el volumen de madera. Además, se comparó el uso de tres curvas de crecimiento individuales para la altura de la planta con el fin de evaluar el tiempo necesario para alcanzar 2 m de altura (T_{2m}). Modelamos la correlación espacial de T_{2m} utilizando modelos mixtos. En ambos sitios, los modelos ST fueron superiores para la altura de la planta y el volumen de madera por hectárea, mientras que para el volumen de madera de los árboles individuales, los modelos temporales fueron superiores. La plantación al pozo siempre tuvo un rendimiento inferior al de la excéntrica y el subsolador, que se comportaron de forma similar. La dinámica de la competencia dentro de la parcela debido a la mortalidad de los árboles se vio afectada por los tratamientos y el sitio. La modelización de la variabilidad espacio-temporal es clave para mejorar las comparaciones de los tratamientos en los experimentos forestales.

Palabras clave: preparación del sitio, intensidad del laboreo, variabilidad espacio-temporal, curvas de crecimiento

RESEARCH ARTICLE

For. Sci. XX(XX):1–11
doi: 10.1093/forsci/fxaa007

biometrics

Spatio-Temporal Modeling and Competition Dynamics in Forest Tillage Experiments on Early Growth of *Eucalyptus grandis* L.

Pablo González Barrios†, Alejandra Borges†, José Terra, Mario Pérez Bidegain, and Lucía Gutiérrez

Forest tillage experiments regularly use long-term evaluations of large plots creating temporal and/or spatial correlations among observations. Not modeling these correlations could compromise treatment comparisons. The aim of this study was to evaluate the effect of modeling spatio-temporal (ST) variability in forest tillage experiments. We used different strategies that incorporate spatial and/or temporal correlations in the evaluation of tillage intensity effect in initial *Eucalyptus* growth as well as evaluate the effect of intraplot mortality and competition dynamics. Three tillage intensities in two contrasting soil conditions were compared for tree height and wood volume. Additionally, we compared the use of three individual growth curves for plant height to evaluate the time needed to reach 2 m in height (T_{2m}). We modeled the spatial correlation of T_{2m} using mixed models. In both sites, ST models were superior for plant height and wood volume per hectare, whereas for individual-tree wood volume, temporal models were superior. Pit planting always had a lower performance than disk harrowing and subsoiler, which behaved similarly. The competition dynamics within the plot because of tree mortality was affected by treatments and site. Modeling ST variability is key to improving treatment comparisons in forest experiments.

Study Implications: The management of tillage practices is a key aspect in the productivity of *Eucalyptus* plantations. The intensity of tillage directly affects the soil physical properties generating changes in the growth and mortality dynamics of trees. Our results indicate that pit planting showed the lower performance in growth variables for all evaluation moments and in both sites. Disk harrowing improves the growth conditions in the first year after planting compared to subsoiler under limiting soil conditions. However, at 30 months after planting, there were no significant differences between disk harrowing and subsoiler treatments. Because disk harrowing is a practice with lower operating costs and environmental impact, it appears to be the superior alternative in terms of productivity and sustainability of forest systems.

Keywords: site preparation, tillage intensity, spatio-temporal variability, growth curves

Commercial plantations of *Eucalyptus* species play an important role in worldwide forest production systems (Grattapaglia and Kirst 2008), representing the largest source of cellulose for the paper industry (Overbeek et al. 2012). The main advantages of these species are high wood yield, adaptability to a wide range of environmental conditions, and diversified areas of final products (Verhaegen and Plomion 1996). The increasing demand for *Eucalyptus* products has caused an expansion of afforestation areas in several regions of the world, leading to strong changes in production systems and land use (Kröger 2012). The expansion of new forested areas is mainly located on marginal or low-productivity soils, where, in many cases, a strong spatial variability structure is present. Furthermore, most forest experiments require large experimental units, evaluated for long periods of time, with increased spatial and temporal variability that challenge their power and precision.

Manuscript received July 30, 2019; accepted January 30, 2020; published online March 28, 2020

Affiliations: *Pablo Gonzalez Barrios (pgonzalez6@wisc.edu), Alejandra Borges (aborges@fagro.edu.uy), and Lucía Gutiérrez (gutierrezcha@wisc.edu), Statistics Department, Facultad de Agronomía, Universidad de la Republica, Uruguay. José Terra (jterra@eyt.inia.org.uy), National Agricultural Research Institute (INIA), Uruguay. Mario Pérez Bidegain (mperezb@fagro.edu.uy), Soils Department, Facultad de Agronomía, Universidad de la Republica, Uruguay. Lucía Gutiérrez (gutierrezcha@wisc.edu), Department of Agronomy, University of Wisconsin, Madison.*

†The first two authors made equal contributions to the manuscript.

Acknowledgments: *The research that provided the results presented in this study received funding from the Agencia Nacional de Innovación e Investigación (ANII) under the code POS_NAC_2012_1_8747, through a fellowship to PGB, Comisión Sectorial de Investigación Científica (CSIC) through an internship to AB and PGB and through an industry partnership research project with Forestal Oriental S.A. (FOSA). The authors would like to thank two anonymous reviewers and the Associate Editor for substantial revision that improved the manuscript.*

Declaration on conflicts of interest: *The authors declare that they have no conflict of interest.*

Soil properties and dynamics are directly affected by soil type and site preparation practices, and therefore the choice of an adequate tillage management becomes crucial (Strudley et al. 2008, Albaugh et al. 2015). In particular, tillage practices modify soil density, decreasing root penetration resistance and increasing water-retention levels (Horn et al. 1995, Worrell and Hampson 1997). Intensive tillage practices such as subsoiling have shown reductions in bulk density and increased porosity of deep horizons, promoting the development of deeper and uniform root systems, and reducing the possibility of plant stress because of long periods of drought (Morris and Lowery 1988, Wheeler et al. 2002). The use of intensive soil practices such as subsoiling has been widespread in recent decades under certain growth conditions including some regions in Uruguay (Garcia-Prechac 2001, Resquin et al. 2019), Argentina (Currie et al. 2004), Brazil (Ferreira et al. 2018), South Africa (du Toit et al. 2010), the Mediterranean region (Fonseca et al. 2011), and some areas of the United States (Carlson et al. 2006). However, there is little information about the impact of these treatments on plant growth (Garcia Prechac et al. 2001, González et al. 2012).

Individual tree growth is a complex process affected mainly by competition from neighbors, spatial micro-site variability, mortality, and genetic factors (Matérn 1960, Fonseca et al. 2011). Micro-site variation tends to create positive correlations among neighbors because they are growing under similar environmental conditions, whereas competition tends to create negative dependences in size or growth of individuals in spatial proximity (Fox et al. 2007a). It is therefore challenging to identify an optimal analysis strategy that allows the integration of spatial variability and competition information among trees, combined with the temporal evolution over time. In experiments with large plot sizes, where planting density is similar among plots, the mortality dynamics over time is probably one of the most important factors in tree competition (Law et al. 2003) affecting productivity performance and competition dynamics (Fonseca et al. 2011). Therefore, understanding the phenomena that generate spatial and temporal variability in large forest experiments is critical to obtaining a greater efficiency in treatment-effects estimations and adopting site-specific management practices.

Modeling growth variables in long-term experiments, such as those in forest research, can be challenging because they are the product of temporal variability (Skovsgaard and Vanclay 2013). This temporal structure has been modeled on longitudinal data in different disciplines (Gregoire et al. 1995, Brownie et al. 2004) in two main groups of strategies: growth models using regression approaches (Diggle et al. 2002) and modeling correlation structures with linear mixed models (Saud et al. 2016). Most of the regression models that have been used to model population growth are logistic or exponential (Tsoularis and Wallace 2002). Additionally, different variance–covariance structures at an individual tree level can be modeled with mixed models (Wolfinger 1996) improving the efficiency of treatment means estimation (Hewitt et al. 2001).

On the other hand, soil spatial variability is widely recognized as one of the most important factors affecting treatment effect comparisons in field experiments (López and Arrué 1995). Modeling spatial correlations in growth models improved model performance (Liu and Ashton 1999, Lee and Wong 2001) and efficiency in treatment effect estimation (González Barrios et al. 2015) in forest tillage experiments.

More recently, spatio-temporal (ST) approaches focusing on both spatial and temporal correlation structures have been proposed (O'Rourke and Kelly 2015), and its use reported significant improvement on treatment effect estimations (Gregoire et al. 1995, Brownie et al. 2004). In forest experiments, an ST regression approach improved the prediction accuracy of end products, when the predictions included previous measurements on the same trees (O'Rourke et al. 2016). Therefore, modeling the spatial and temporal correlations as well as individual monitoring of each tree could provide more efficient models for comparing tillage practices.

The aims of this study were: (1) to compare spatial, temporal, and ST models using both growth curves and mixed models approaches to estimate plant height, wood volume and initial growth rate; (2) to evaluate tillage treatment effects on intraplot mortality, and competition dynamics; and (3) to use the superior ST model to evaluate the effect of different tillage intensities in initial *Eucalyptus* growth.

Materials and Methods

Experimental Site and Design

The experiment was conducted at two adjacent sites (same climate), with differences in the soil properties (S_1 and S_2) 400 m apart in Rio Negro, Uruguay (32°37′49″S; 57°10′07″W). Dominant soils in the S_1 experiment were Lithic Dystrudepts with a 10 cm A horizon of sandy texture (71 percent sand), 5.6 pH (H_2O), 6.8 cmolc kg^{-1} of cation exchange capacity (CEC), 2 percent soil organic matter (SOM), and an R horizon below the A horizon. Lithic soils are characterized by the presence of rock fragments or outcrop. Dominant soils in the S_2 experiment were Typic Argiudoll with a 7 cm sandy A horizon (67 percent sand), 5.5 pH (H_2O), 7.9 cmolc kg^{-1} CEC, 4.6 percent SOM, and an AB horizon with 53 percent sand and 21 percent clay and less organic matter (Table S1). The area has a temperate climate and isohydric precipitation regime (Figure 1).

Both experiments were planted with an *E. grandis* commercial clone (X2334) from seedlings with homogeneous height and excellent sanitary conditions in a randomized complete block experiment with four replications in S_1 and five in S_2. Previous land use was natural grassland on undisturbed soils. Tillage treatments on the plantation row were pit planting (P), disk harrowing (D), and subsoiler (S). The P treatment consisted of a pit of 20 cm × 20 cm dug up with a hand shovel on March 23 at a spacing of 2.5 m, followed by the use of 2.4 kg ha^{-1} of isopropylamine salt of glyphosate applied on the row on March 29 to reduce weed competition, and hand planting on April 4. The D treatment consisted of the use of a 1-m-width heavy disk harrow with notched 91.44 cm disks on the row on January 14, followed by a 1-m-width light disk harrow with smooth 40–50 cm disks on the row on February 10, and hand planting on April 4. The goal of the disk harrowing was to achieve an effective depth of 20–25 cm. The S treatment consisted of the use of a subsoiler shank at 50 cm depth on the row on February 10, followed by hand planting on April 4. The effective soil impact on soil penetration resistance can be seen on Figure S1. The use of a subsoiler is a widely adopted practice in this geographical area of production under

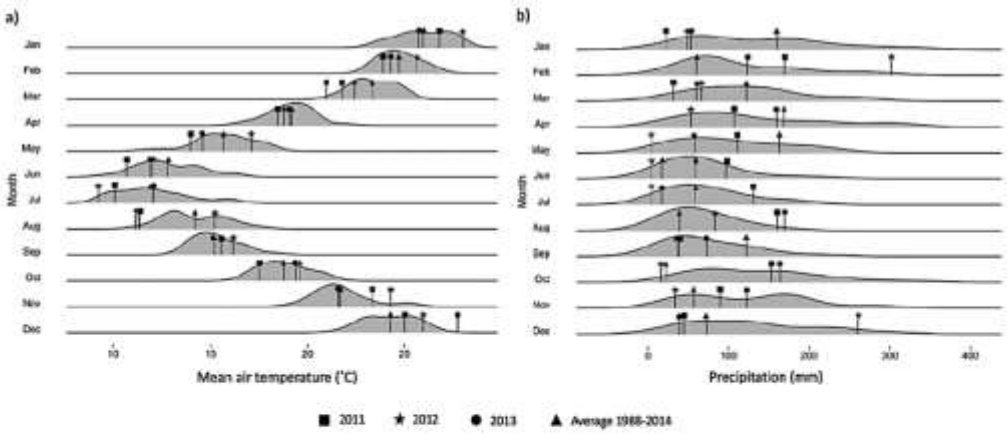


Figure 1. Distribution of mean air temperature (°C) (left panel) and cumulative monthly rainfall (right panel) in Mellizos for the period 1988–2014 (density plots) and averages for the period 1988–2014, and for the years 2011, 2012, and 2013 (indicated by different symbols in the graph). The experiments were planted in April 2011, and the last measurement date at 30 months corresponds with October 2013.

limiting soil conditions (García-Préchac et al., 2001, Currie et al. 2004, Ferreira et al. 2018, Resquin et al., 2019) regardless of the higher energy cost and environmental consequences of more aggressive tillage. Disk harrowing and pit planting were chosen as alternatives in terms of energy cost and soil erosion-risk reduction. Experimental units consisted of three rows with 12–15 trees per row with 3.5 m between rows and 2.5 m between trees, for a density of 1,150 trees per hectare. There was a buffer area of 5 m between blocks. After plantation, 2 kg ha^{-1} of Lampo™ (Fipronil 0.003 percent weight) distributed in 400 points over the field at each site was applied for weed control. During the remaining experimental period, grass control consisted of periodical herbicide application in the inter-row spacing for all treatments.

Measurements

Plant height (h) and diameter at breast height (dbh) for each tree were measured for the first 30 months after out planting. Plant height measurements were taken at 7, 12, 16, 20, 25, and 30 months postplanting, and dbh was measured at 20, 25, and 30 months postplanting. Individual-tree volume (cm^3 $tree^{-1}$) was calculated as:

$$\text{Volume} = \text{basal area}\left(m^2\right) \times h(m) \times f = \pi \times \left(\frac{\text{dbh}}{2}\right)^2 \times h \times f \quad (1)$$

where basal area was calculated using the dbh value, h is the plant height in meters, and f is a form factor for *Eucalyptus*. We followed the definition of the International Union of Forest Research Organizations to estimate individual wood volume (from ground to tip of the tree). This definition is recommended by FAO (1981), when only dbh and h values are available. We used a form factor of 0.45 as proposed by Perdomo et al. (2007) to estimate wood volume for *Eucalyptus* in Uruguay. We also conducted a sensitivity analysis to evaluate the impact of the form factor on standard errors and mean comparisons. Treatment means were compared by a Tukey test. The total wood volume per hectare (m^3 ha^{-1}) was calculated for each tree multiplying the individual-tree volume by the number of live trees in each plot for each measurement date, and then scaling it by the plot area to get volume per hectare.

Statistical Analysis

Different strategies to model the spatial and temporal variability for plant growth variables were used: spatial modeling of growth curves for initial growth rate and ST mixed model approach for plant height and wood volume.

Strategy 1: Spatial Modeling of Growth Curves

To evaluate the effect of tillage intensities on initial tree growth, the performance of different plant growth curves was analyzed in both sites. For each tree, three different regression models were fitted for plant height: Logistic, Gompertz, and Cubic. The mathematical functions for each model are presented below.

$$\text{Logistic}: Y_{ij} = \frac{\alpha_j}{1 + \beta_j e^{-\gamma_j X_i}} \quad (2)$$

where Y_{ij} is the plant height of the jth plant in the ith month after planting, α_j, β_j and γ_j are parameters of the growth curve for each tree, and X_i is the time in months.

$$\text{Gompertz}: Y_{ij} = \alpha_j e^{-\beta_j e^{-\gamma_j X_i}} \quad (3)$$

where Y_{ij} is the plant height of the jth plant in the ith month after planting, α_j, β_j and γ_j are parameters of the growth curve for each tree, and X_i is the time in months.

$$\text{Cubic}: Y_{ij} = \beta_{0j} + \beta_{1j} X_i + \beta_{2j} X_i^2 + \beta_{3j} X_i^3 \quad (4)$$

where Y_{ij} is the plant height of the jth plant in the ith month after planting, β_{0j}, β_{1j}, β_{2j}, and β_{3j} are parameters of the growth curve for each tree, and X_i is the time in months.

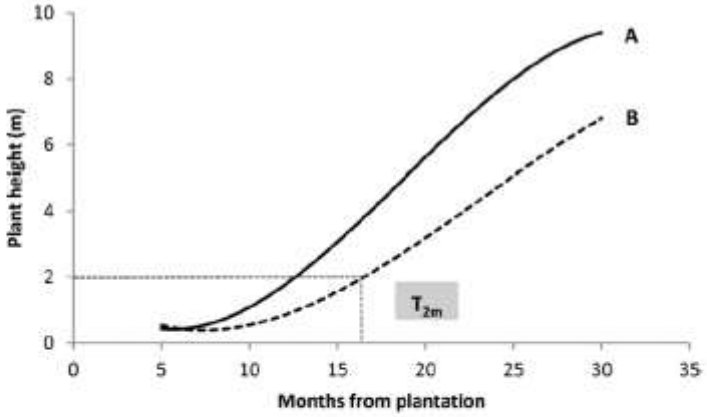


Figure 2. Theoretical cubic curves that differ in the time to reach 2 m of height (T_{2m}) for two theoretical trees: A and B.

Model selection was based on two metrics, the root mean squared error of the prediction (RMSEP) and the coefficient of determination (R^2) between observed and predicted plant height values. For each model and metric, we calculated the median for all trees and the number of times the model had the lowest RMSEP and highest R^2.

For the selected model, the time needed to reach a predicted height of 2 m (T_{2m}) was extracted from each curve as a measure of initial tree growth rate (Figure 2). The values of T_{2m} were considered as realizations of a random variable in each tree and were analyzed by mixed models according to the experimental design and spatial information for each individual tree. We used geographically weighted regressions (Zhang and Haijin 2004) to compare T_{2m} among tillage treatments with the following model:

$$Y_{ijk} = \mu + T_i + B_j + \varepsilon_{ij} + \delta_{ijk} \quad (5)$$

where Y_{ijk} is the response variable (T_{2m}) of the kth plant in the ith tillage treatment, and the jth block; μ is the overall mean; T_i is the ith tillage treatment effect; B_j is the jth block effect, ε_{ij} is the residual error with $\varepsilon_{ij} \sim N(0, \sigma^2_e)$ and δ_{ijk} is the subsampling error with $\delta_{ijk} \sim N(0, \sigma^2_s \mathbf{R})$, where σ^2_e is the residual error variance, σ^2_s the subsampling error variance, and $\mathbf{R}$ the variance–covariance matrix among trees. Four variogram models were used to estimate spatial correlations between trees to model the $\mathbf{R}$ matrix (SAS Institute Inc 2011): exponential (EXP), Gaussian (GAU), spherical (SPH), and power (POW). The model was selected based on the Bayesian information criterion (BIC) (small is better) and the average standard error of the difference between two treatment means (standard error of mean differences [SED]; small is more precise). Best linear unbiased estimators were obtained for T_{2m} for each treatment based on the superior model and compared by a Tukey test when significant differences among treatments were detected.

Strategy 2: ST Modeling of Productive Variables

To address the spatial and temporal variability of the productive variables, we used linear mixed models with different temporal, spatial, and ST correlation structures. This allowed us to model intraplot (i.e., among trees) and temporal (i.e., repeated measurements in the same tree) correlations. All models used the same structure of fixed effects. The basic model can be written as follows:

$$Y_{ijkl} = \mu + T_i + B_j + \varepsilon_{ij} + D_k + TD_{ik} + \delta_{ijkl} \quad (6)$$

where Y_{ijkl} is the response variable (i.e., plant height, tree wood volume, or total wood volume) of the lth tree in the ith tillage treatment, in the jth block and the kth time; μ is the overall mean; T_i is the ith tillage treatment effect; B_j is the jth block effect; ε_{ij} is the residual error with $\varepsilon_{ij} \sim N(0, \sigma^2_e \boldsymbol{\Sigma}_s)$ with σ^2_e the residual error variance and $\boldsymbol{\Sigma}_s$ the spatial variance–covariance matrix among plots; D_k is the kth measurement time effect; TD_{ik} is the interaction between tillage treatment and measurement time and δ_{ijkl} is the subsampling error with $\delta_{ijkl} \sim N(0, \sigma^2_s \mathbf{R})$ with σ^2_s the subsampling error variance and $\mathbf{R}$ a block diagonal matrix that models among time variance–covariance relations ($\boldsymbol{\Sigma}_t$) and the spatial variance–covariance among trees ($\boldsymbol{\Sigma}_s$).

Null Model

The null model assumes that $\boldsymbol{\Sigma}_s$ and $\boldsymbol{\Sigma}_t$ matrices are identity matrices ($\boldsymbol{\Sigma}_s = \mathbf{I}$ and $\boldsymbol{\Sigma}_t = \mathbf{I}$), assuming no correlation between plots, trees, or time measurement.

Temporal Modeling

This group of models considers only the temporal relation between measures on the same tree. The $\boldsymbol{\Sigma}_t$ matrix was modeled with six different correlation structures, three assuming homogeneous variances among times, i.e., compound symmetry (CS), autoregressive of order 1 (AR1), Toeplitz (TOEP), and three assuming heterogeneous variances, i.e., heterogeneous compound symmetry (CSH), heterogeneous autoregressive of order 1 (ARH1), and antedependence (ANTE).

Spatial Models

Four spatial correlations models were compared ($\boldsymbol{\Sigma}_s$): Gaussian (GAU), exponential (EXP), power (POW), and spherical (SPH). Spatial structures were modeled as an average spatial structure across all measurement times for each tree and plot.

ST Models

Combined models consider simultaneously the temporal ($\boldsymbol{\Sigma}_t$) and spatial ($\boldsymbol{\Sigma}_s$) correlation structures as described above.

Table 1. Model fit (BIC) and SED for the variable time to reach 2 m of tree height, evaluated in months.

	S_1		S_2	
Structure	BIC	SED	BIC	SED
Null model	1,267.3	0.322	1,579.5	0.282
Gaussian	*1,264.2*	*0.339*	*1,552.9*	*0.261*
Exponential	1,264.0	0.391	NC	NC
Spherical	1,270.7	0.366	1,566.6	0.355
Power	1,264.0	0.357	NC	NC

Note: BIC, Bayesian information criterion; NC, did not converge; SED, standard error of the mean difference among tillage treatments. Different spatial structures at two experimental sites (S1 and S2) are compared. The model with the smallest BIC or SED for each site is shown in italics.

Table 2. Adjusted means and standard errors of time to reach 2 m of tree height evaluated in months for three tillage treatments at two experimental sites (S_1 and S_2).

Site	Tillage	Time to reach 2 m of tree height (months)
S_1	Pit planting	13.1 ± (0.2)^a
	Disk harrowing	10.2 ± (0.3)^c
	Subsoiler	11.3 ± (0.2)^b
S_2	Pit planting	12.2 ± (0.2)^a
	Disk harrowing	9.8 ± (0.2)^b
	Subsoiler	9.2 ± (0.2)^b

Note: Different letters indicate significant differences ($P < .05$).

Descriptive Analysis

A plant-height map was performed for each of the six measurement times to illustrate the spatial variability pattern changes over time with the "sp" package of R Statistical Software (R Core Team 2018).

Plant Survival and Competition Dynamics

Plant survival rates during the experimental period were estimated for each plot and for each measurement time as a mortality index. This variable was calculated as the ratio of the remaining number of trees in each plot divided by the initial number of trees of the plot. To evaluate and compare levels of intraplot competition, two indices were calculated for each plot: height competition index (HCI) and wood volume competition index (VCI):

$$\mathrm{HCI} = \frac{\sum_{i=1}^{n} h}{n} \tag{7}$$

$$\mathrm{VCI} = \frac{\sum_{i=1}^{n} \mathrm{vol}}{n} \tag{8}$$

where h and vol are the plant height and wood volume of the ith tree respectively, and n is the initial number of trees for each plot.

Both, plant survival rate and the two competition indices were treated as response variables and were analyzed by measurement date using the following model:

$$Y_{ij} = \mu + T_i + B_j + \varepsilon_{ij} \tag{9}$$

where Y_{ij} is the plant survival rate (HCI or VCI) for the ith tillage treatment and the jth block, μ is the overall mean, T_i is the effect of the ith tillage treatment, B_j is the effect of the jth block, and ε_{ij} is the residual error associated with the ith tillage treatment and the jth block. The ε_{ij} are independent random variables with $\sim N(0, \sigma^2_\varepsilon)$, where σ^2_ε is the residual error variance.

All models used residual maximum likelihood to estimate variance components. The Kenward–Roger method to correct the degrees of freedom and the variance–covariance estimation was used for the spatial, temporal, and ST models. Models were compared based on the BIC and the SED. All analyses were performed with SAS [Version 9.2] (SAS Institute 2011).

Results

Plant Height Descriptive Analysis

The plant heights for each tree and for all measurement times in S_1 are shown in Figure S2. Spatial variability among plots of the same treatment (between-plot variability) and among trees within a plot (within-plot variability) was observed as well as treatment differences.

Strategy 1: Spatial Model Comparison for Growth Curves

Cubic curves had the lowest RMSEP and the highest R^2 values for plant height in the initial 30 months after planting (Table S2) for both sites. The use of spatial modeling for T_{2m} reduced the BIC values compared to the null model (Table 1), with the GAU model having the lowest values in both experimental sites.

There were significant differences between tillage intensities for T_{2m} (Table 2). In both S_1 and S_2, trees growing in the disk harrowing or subsoiler treatments reached an average height of 2 m earlier than those growing in the pit planting treatments. Additionally, in S_1, disk harrowing reached T_{2m} earlier than subsoiler.

Strategy 2: Spatio-, Temporal, and ST Modeling of Productive Variables

For all variables considered (i.e., plant height, individual-tree volume, and wood volume per hectare) and sites (S_1 and S_2), the use of any temporal, spatial, or ST variance–covariance structure had lower BIC values than with the null model (Table 3). For plant height and wood volume per hectare, the model with smallest BIC values for both sites was an ST, with the ANTE temporal structure and SPH spatial structure for plant height, and CSH temporal and GAU spatial structure for wood volume per hectare. For individual-tree wood volume, the model with the smallest BIC was obtained with the CSH temporal structure at both sites. In terms of precision, models that assumed some correlation structure had a smaller SED than the null model, except for individual-tree wood volume in S_2 (Table 3).

There were significant differences between tillage treatments for plant height in S_1 in the initial 12 months of growth; disk harrowing had the tallest trees, and pit planting the shortest (Table 4). However, no statistical differences in plant height were found between disk harrowing and subsoiler after 12 months and until the end of the evaluation period, but both were taller than with pit planting. The superiority of more intensive tillage than with pit planting was constant throughout all the measurements. For S_2, disk harrowing was never different from subsoiler for tree height, except on the first evaluation date (7 months). Finally, no significant differences between disk harrowing and subsoiler were found for wood volume per tree and

Table 3. Model fit (BIC) and average of the SED of plant height (Height), individual-tree wood volume (Tree volume), and wood volume (Area volume) for different spatial, temporal, and spatio-temporal structures at both experimental sites (S_1 and S_2).

Strategy	Temporal structure	Spatial structure	S_1						S_2					
			Height (cm)		Tree volume (cm^3 $tree^{-1}$)		Area volume (m^3 ha^{-1})		Height (cm)		Tree volume (cm^3 $tree^{-1}$)		Area volume (m^3 ha^{-1})	
			BIC	SED	BIC	SED	BIC	SED	BIC	SED	BIC	SED	BIC	SED
Null	Null	Null	24,045	10.2	23,240	2,766	7,018	0.98	3,849	10.5	29,981	1,232	9,337	0.96
Temporal	CS		21,563	10.3	22,778	2,682	6,555	0.88	2,677	11.2	29,398	1,898	8,749	0.65
	CSH		20,682	7.3	*22,288*	*2,655*	6,055	0.62	–	–	*28,624*	*1,858*	7,952	0.77
	AR		21,073	10.1	22,748	2,634	6,529	0.88	2,371	8.0	29,302	1,877	8,652	0.64
	ARH		21,044	9.8	22,339	2,709	6,111	0.66	1,586	7.4	28,687	1,835	8,018	0.76
	ANTE		21,073	10.1	22,349	2,710	6,113	0.66	1,278	7.1	28,695	1,837	8,019	0.75
	TOEP		21,044	9.7	22,756	2,641	6,530	0.88	2,357	7.5	29,312	1,866	8,652	0.63
Spatial		GAU	21,564	10.2	22,781	2,517	6,558	0.74	2,677	10.3	29,398	1,898	8,752	0.56
		EXP	21,562	10.2	22,778	2,683	6,557	0.70	2,676	10.2	29,398	1,898	8,749	0.57
		SPH	21,652	10.2	22,885	2,555	–	–	–	–	29,538	5,190	–	–
		POW	21,562	10.9	22,781	2,666	6,558	0.69	2,676	10.1	29,401	1,845	8,751	0.56
Spatio–Temporal	CS	GAU	21,565	10.9	22,783	2,684	6,560	0.79	2,679	10.6	29,404	1,846	8,754	0.63
		SPH	21,567	10.8	22,783	–	6,568	0.64	2,681	11.3	29,403	2,067	–	–
		EXP	21,564	11.2	–	–	6,560	0.88	2,678	15.0	29,401	–	8,751	0.63
		POW	–	–	–	–	–	–	–	–	29,404	–	–	–
	CSH	GAU	20,507	9.4	22,293	2,656	*6,006*	*0.81*	1,620	7.1	28,629	1,858	*7,909*	*0.81*
		SPH	20,510	9.4	22,293	2,656	–	–	1,626	7.1	28,629	1,858	7,953	0.77
		EXP	20,507	9.4	22,293	2,657	6,060	0.90	1,619	7.1	28,629	1,858	–	–
		POW	20,510	9.4	22,293	2,656	6,010	0.62	1,626	7.1	28,629	1,858	7,953	0.77
	AR(1)	GAU	21,216	10.4	22,751	2,632	6,532	0.84	2,372	8.6	29,305	1,818	8,654	0.81
		SPH	21,217	10.4	22,752	–	6,534	0.84	2,373	8.0	29,307	2,075	8,656	0.89
		EXP	21,215	10.4	22,749	2,631	6,530	0.88	2,372	10.2	29,305	1,877	8,654	0.81
		POW	21,217	10.4	22,752	–	6,534	0.84	2,373	8.0	29,307	2,076	8,656	0.89
	ARH(1)	GAU	20,108	–	22,842	2,567	6,099	0.74	1,270	6.9	29,395	2,182	8,001	0.96
		SPH	21,328	10.5	22,830	2,605	–	–	–	–	29,414	1,772	8,004	1.07
		EXP	20,108	–	22,829	2,619	6,095	0.71	1,269	7.1	29,392	2,256	8,021	1.00
		POW	20,110	7.3	22,829	3,055	6,095	0.67	1,264	–	29,412	2,344	8,003	1.09
	ANTE(1)	GAU	21,020	11.3	22,847	2,763	–	–	–	–	29,384	2,170	8,008	0.96
		SPH	*20,042*	*6.2*	22,830	2,601	6,105	0.96	*1,264*	*6.8*	29,385	1,760	8,007	1.07
		EXP	21,017	12.3	22,829	2,458	–	–	–	–	29,381	2,254	8,019	1.00
		POW	20,042	13.1	22,829	2,949	–	–	–	–	29,384	2,096	–	–
	TOEP	GAU	21,017	12.2	22,759	2,642	6,532	1.18	2,354	8.1	29,314	1,866	8,655	0.80
		SPH	21,049	10.6	22,759	2,642	6,532	1.18	2,359	7.5	29,314	1,866	8,655	0.80
		EXP	21,049	10.6	22,761	2,703	6,534	1.25	2,353	9.4	29,314	1,866	8,655	0.80
		POW	21,049	10.6	22,759	2,642	6,532	1.18	2,359	7.4	29,314	1,866	8,655	0.80

Note: ANTE, antedependence; AR1, autoregressive; ARH1, heterogeneous autoregressive; BIC, Bayesian information criterion; CS, compound symmetry; CSH, heterogeneous compound symmetry; EXP, exponential; GAU, Gaussian; Null, Null model; POW, power; SED, standard error of the mean difference among tillage treatments; SPC, spherical; TOEP, Toeplitz. The model with the smallest BIC or SED for each variable and site is shown in italics. "–" indicates that the model did not converge.

wood volume per hectare in any measurement date and were higher than pit planting.

Plant Survival and Competition Dynamics

Under more restrictive conditions for root development such as S_1, no significant differences in the mean number of surviving trees among tillage treatments were detected until 16 months (Figure 3A). Subsoiler had a higher plant survival proportion than with disk harrowing and pit planting from 20 to 30 months. This was expected because the subsoiler effect on soil is most noticeable at 40–50 cm depth, and it would take the plants several months to reach that root exploration depth. Under less restrictive soil conditions as in S_2, no significant differences in plant survival among tillage treatments were found (Figure 3B).

The competition dynamics in each plot was affected differently according to the site and the competition index (Figure 4). No differences among treatments were found in either site for HCI at the end of the evaluation period. However, at early growth stages, differences among treatments depended on the site considered. For S_1, differences appeared from 12 to 16 months, where D had the highest values (Figure 4A). For S_2, differences among treatments appear from the beginning until 16 months, where D and S had higher values than P (Figure 4B).

In the case of VCI, differences in competition among treatments were more evident in later stages of growth (30 months for both sites), where D and S showed a better performance than P (Figure 4C and D respectively).

Discussion

This study evaluated different modeling strategies of spatial and temporal correlations in forest tillage experiments identifying alternatives that increased the precision of treatment mean comparisons when large plots were evaluated over time. Spatial and/or temporal models were better than the null model both in terms of model fit and precision in treatment comparisons. Mixed models are useful in forest experiments with large plots (Fox et al. 2007b) because they can model the variability at different levels (Brownie et al. 2004, Piepho et al. 2004) including among plot and among trees within plots variability, and appropriate error terms

Table 4. Adjusted means and standard errors for plant height, wood volume, and wood volume at 7, 12, 16, 20, 25, and 30 months postplanting for both experimental sites.

Site	Tillage	Measurement time					
		7	12	16	20	25	30
		Plant height (cm)					
S_1	P	33.7 ± 2.0^b	191.5 ± 4.9^b	277.3 ± 5.4^b	488.6 ± 5.3^b	752.0 ± 6.2^b	846.7 ± 6.3^b
	D	42.4 ± 2.0^a	299.3 ± 4.9^a	398.0 ± 5.5^a	592.9 ± 5.4^a	836.1 ± 6.4^a	907.5 ± 6.5^a
	S	29.8 ± 2.0^b	260.6 ± 4.8^a	357.2 ± 5.4^a	577.1 ± 5.3^a	828.6 ± 6.2^a	909.1 ± 6.4^a
S_2	P	43.1 ± 2.3^b	201.0 ± 6.2^c	286.7 ± 7.1^b	498.1 ± 7.3^b	762.6 ± 7.2^b	855.1 ± 7.0^b
	D	54.8 ± 2.3^a	314.3 ± 6.2^a	412.4 ± 6.7^a	606.8 ± 7.3^a	850.4 ± 6.8^a	921.4 ± 7.0^a
	S	42.1 ± 2.3^b	273.1 ± 5.9^b	369.4 ± 6.7^a	590.3 ± 7.0^a	841.3 ± 7.1^a	921.9 ± 6.7^a
		Wood volume (cm^{-3} $tree^{-1}$)					
S_1	P	–	–	–	12,000 ± 1,800^b	33,000 ± 1,700^b	49,000 ± 1,800^b
	D	–	–	–	24,000 ± 1,800^a	53,000 ± 1,700^a	67,000 ± 1,800^a
	S	–	–	–	15,000 ± 1,800^b	49,000 ± 1,700^a	64,000 ± 1,800^a
S_2	P	–	–	–	17,000 ± 1,500^b	37,000 ± 1,700^b	64,000 ± 1,800^b
	D	–	–	–	31,000 ± 1,600^a	67,000 ± 1,700^a	85,000 ± 1,800^a
	S	–	–	–	31,000 ± 1,500^a	67,000 ± 1,700^a	85,000 ± 1,800^a
		Wood volume (m^{-3} ha^{-1})					
S_1	P	–	–	–	4.2 ± 0.6^b	13.8 ± 0.9^b	21.0 ± 0.9^b
	D	–	–	–	9.3 ± 0.6^a	22.3 ± 0.9^a	28.6 ± 0.9^a
	S	–	–	–	7.8 ± 0.6^a	21.6 ± 0.9^a	28.4 ± 0.9^a
S_2	P	–	–	–	7.0 ± 0.6^b	19.0 ± 0.8^b	27.7 ± 1.0^b
	D	–	–	–	11.7 ± 0.6^a	26.5 ± 0.9^a	34.5 ± 1.0^a
	S	–	–	–	12.3 ± 0.6^a	27.7 ± 0.8^a	35.3 ± 1.0^a

Note: For each variable and site, different letters within columns indicate significant differences ($P < .05$). Tillage treatments: D, disk harrowing; P, pit planting; S, subsoiler. Wood volume per tree was calculated using a form factor of 0.45.

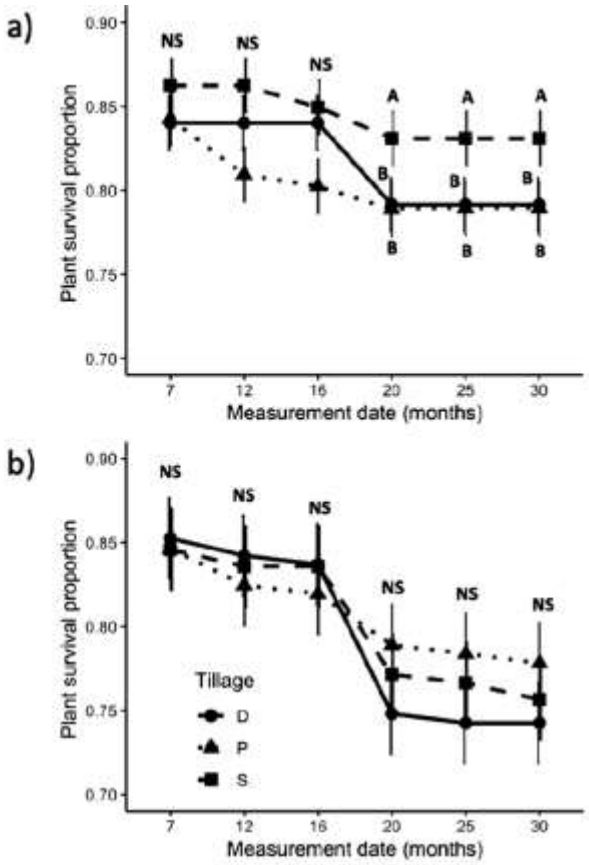
for effect testing are used. Furthermore, mixed models are versatile enough to model spatial (isotropic or anisotropic; Dutkowski et al. 2006), temporal (Piepho et al. 2004, Hong et al. 2005, Fortin et al. 2007), and ST (Brownie et al. 2004, O'Rourke and Kelly, 2015, O'Rourke et al. 2016) correlation structures (Piepho et al. 2011).

Strategy 1: Spatial Model Comparison for Growth Curves

All growth curves used in our study (linear and nonlinear) to model plant height had a good fit measured as the RMSEP and the square of the correlation between observed and predicted plant height values (R^2) (Table S2). The reverse curvature at the beginning of the evaluation period shown in all trees ("sigmoid" shape) led us to choose these three specific models for comparison. This strategy has been proven successful to model *Eucalyptus* growth in other studies (Tsoularis and Wallace 2002, Calegario et al. 2005). The cubic model had the lowest RMSEP and highest R^2, and was a useful and simple model that required less computational time than the other two models. We found spatial variability in tree growth that affected the time needed to reach 2 m. Reaching 2 m height in a shorter period could improve the probability of plant survival because of the avoidance of critical drought periods or other stresses during implantation (Graciano et al. 2005). This could in turn increase volume production. The differences observed in T_{2m} between sites indicate that the initial growth rate is site-specific, as was also found in other studies (Calegario et al. 2005). Fox et al. (2001) confirmed the existence of spatial dependence among individual-tree attributes, but different magnitudes of positive and negative spatial dependence have been identified. If the dominant spatial mechanism is competition, observations in spatial proximity tend to be more dissimilar than average, whereas if the dominant mechanism is spatial micro-site effects, tree attributes tend to be more similar than average. This dominant mechanism could be related to the developmental stage. The spatial micro-site effects seem to be the dominant mechanism at initial stages, whereas competition seems dominant at later stages (O'Rourke et al. 2016). Furthermore, Liu and Burkhart (1994) suggested that final tree height is less sensitive to competition than dbh in the juvenile period of loblolly pine stands. We used the estimated time required to reach 2 m of tree height as a response variable; however, we did not account for error in the estimation of this random variable, and it would have been interesting to include it in the spatial models in order to separate and quantify two sources of variability, the micro-site variation (nugget variance) and the variation associated with the parameter estimations.

Strategy 2: Spatio-, Temporal, and ST Modeling of Productive Variables

We found spatial and/or temporal correlations for plant height and wood volume. The specific variance–covariance structure selected was different depending on the variable considered. Beyond this, the selected models for any variable and site included both temporal correlation between measurements of the same tree and a heterogeneous variance component among measurement times. The selection of superior models was primarily based on BIC, but all selected models showed a lower SED value in comparison with the Null model (except for tree volume in S_2). In particular, the reduction in terms of SED between the selected and the null model was 4 percent and 17 percent for plant height and 35 percent and 40 percent for individual-tree wood volume, for S_1 and S_2 respectively. This indicates that by incorporating spatial and/or temporal variability into our model, we increased the experimental precision. However, the selection of a suitable covariance model is not trivial, given that covariance parameters are unknown, and their estimations must be used when comparing fixed effects (Hu and Spilke 2009). This leads to statistical tests that may not be exact, and the estimated standard error could be biased if the spatial model is not the adequate (Richter and Kroschewski 2012). We used Kenward–Roger approximation in



† Different letters indicate significant differences ($P < 0.05$). NS: no significant differences.

Figure 3. Plant survival proportion (i.e., ratio of the number of surviving plants and the initial number of plants per plots) for each tillage and measurement time in (a) site 1 and (b) site 2. Different letters within each measurement date indicate significant differences ($P < .05$). Tillage treatments: D, disk harrowing; P, pit planting; S, subsoiler.

all cases to improve the reliability of the test, but we are aware that there may still be some bias in the SED estimates. That is why we chose BIC as the primary indicator for model comparisons and left the SED as a secondary one. For plant height and wood volume per hectare, the selected models had ST variance–covariance structures. For plant height, we used a model combining ANTE structure for temporal correlations and SPH for spatial variability, whereas for wood volume per hectare, we used CSH for temporal and GAU for spatial correlations. Other studies found evidence of space–time interaction in dbh (O'Rourke et al. 2016). In both sites, we confirmed the presence of spatial structure in plant-height growth, similar to Raimundo et al.'s (2017) findings. However, for individual-tree wood volume, the selected model only included temporal correlations, modeled by a CSH structure. This indicates that there might be a compensatory effect of competition among trees for wood volume. The individual growth of trees is affected by multiple factors that interact in complex ways, especially competition and spatial micro-variability (Fox et al. 2007a). Wood volume was evaluated only in the last three measurement dates of our study, not including the initial tree growth where spatial micro-site effects could be more important than competition (Fox et al. 2001). Additionally, the variance in the number of remaining trees in each plot reinforced the compensatory effect of competition in wood volume, possibly leading to no spatial correlations among trees. However, when trees were modeled using the wood volume per hectare as a response variable, the lowest BIC value was obtained with an ST model. The use of intraplot mortality information to calculate this variable seems to provide additional information about growth dynamics. These differences in the mortality levels among plots could be indicating the presence of spatial patterns of mortality in the experimental area that are captured by the use of ST structures.

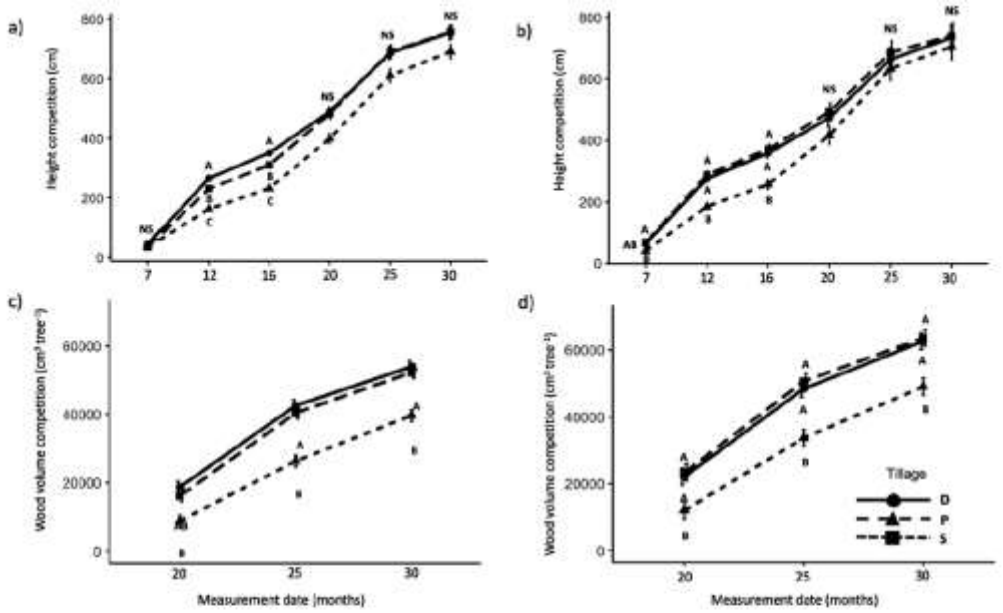


Figure 4. Means and standard error bars for height competition index (HCI) and wood volume competition index (VCI), in site 1 (left panel, a and c) and site 2 (right panel, b and d) for each measurement date (months postplanting). Different letters within each measurement date indicate significant differences ($P < .05$). Tillage treatments: D, disk harrowing; P, pit planting; S, subsoiler.

Tillage Intensity Effect on Initial Growth

We used different strategies to compare tillage treatment effects on *Eucalyptus* early growth. We were able to detect difference among tree heights of less than 1 m, less than 70,000 cm^3 $tree^{-1}$ for tree volume, and less than 5 m^3 ha^{-1} for wood volume per hectare. Although individual wood volume estimations may vary depending on how they are calculated, in comparative terms we were able to verify that higher volumes were found in the more intense treatments (i.e., disk harrowing, subsoiler) than with pit planting. This was found for a form factor of 0.45 (Table 4) and confirmed for a sensitivity analysis evaluating a range of form factors including 0.35 and 0.41 (data not shown). This method assumes that all treatments would create trees of the same shape.

The required time to reach 2 m of height was different among tillage treatments. Intensive tillage like disk-harrowing and subsoiler allowed trees to growth faster than pit planting. No differences were found in growth speed between disk harrowing and subsoiler in nonlimiting conditions for root development (like in S_2). However, disk harrowing allowed faster growth than subsoiler in limiting soil conditions, probably because of a better weed control of this tillage at plantation time. Under these conditions, the disk-harrowing treatment reaches similar values of tree height 3 months before pit planting. The use of tillage systems that allow trees to grow faster and reach key developmental stages early could differentiate better survival rates and consequently improve the wood volume produced per unit of area. In our work, this was not evident in all cases. Under nonlimiting soil conditions like S_2, there were no differences in plant survival among tillage systems. However, under more restrictive conditions for root exploration and development as shown in S_1, the use of subsoiler reached higher tree survival rates than with disk harrowing or pit planting systems. The "baseline" plant height needed to significantly improve the survival rate could vary according to the location, mainly affected by the environmental conditions. We consider this "baseline" height as an arbitrary value of 2 m for both sites, but other values could have been used. Fonseca et al. (2011) compared different site-preparation techniques and found a higher tree survival with more intensive tillage systems.

It is well documented that tillage system could affect nutrient and water-resource availability, and its effect can be short, medium, or long term (Albaugh et al. 2015). However, Smith et al. (2001) indicated that the growth responses to the subsoiler were erratic, given that the effectiveness of these operations is dependent on the soil type, water availability, and time of tillage. In particular, subsoiler has shown reductions in the soil bulk density and increased in soil porosity of deep horizons, promoting the development of deeper and uniform root systems, which reduce the probability of water stress by plants over long periods of drought (Morris and Lowery 1988, Wheeler et al. 2002). In our results, this was evidenced by a lower soil penetration resistance generated by the use of the subsoiler in the first 50 cm of soil depth (Figure S2), giving the plants the ability to withstand longer periods of water deficit. In our study, monthly accumulated precipitations were far below the historical average for the region between 16 and 20 months after plantation (Figure 1). Considering soils with low water retention such as S_1, this water deficit could have explained the lower plant survival in plots corresponding to disk-harrowing treatment, which is a moderate-intensity tillage. In contrast, the subsoiler could have

favored root growth and thus increase its ability to capture water at a greater depth, decreasing intraplot mortality.

Competition indices were affected by treatments, where more intense tillage managements showed better results. These indices summarize how competition affects plant growth for each plot in each measurement date, because they are affected simultaneously by plant height or volume and survival rate. For HCI, there was a clear advantage of more intense tillage treatment in early growth stages, allowing trees to reach a greater height in less time in both sites. For S_1, disk harrowing presented the highest HCI values until 20 months after plantation, where the subsoiler reached similar values. As we mentioned above, between 16 and 20 months after planting, disk harrowing showed higher plant mortality levels than the subsoiler, which was possibly compensated by higher height values that disk harrowing brought from previous measurement dates. This compensatory effect could have explained the similar VCI values observed for these two treatments.

In summary, the use of tillage systems in forest-site preparation can improve conditions in many soil types, affecting positively the installation of plants (Berry 1979). These beneficial contributions of more intensive tillage systems are associated with increased drainage, improved micro-environmental conditions (i.e., nutrients, aeration, temperature), improved root development, and reduced competition (Haines et al. 1975). However, the high costs of site preparation, using mechanical methods, such as tillage, and inconsistent results cast doubt on the use of these practices (Lincoln et al. 2006). On the other hand, the application of fertilizer or herbicide at planting time is less expensive and can in certain situations favor the availability of resources at this critical stage. That said, under these particular environmental production conditions, the use of disk harrowing proves to be the optimal tillage method. Even in more limiting situations for tree development, the rapid initial growth generated by this system could compensate for a lower survival rate in stress situations, reaching similar wood yields than more intense managements. Additionally, it is a practice that involves lower operating costs. Beyond this, the use of a subsoiler could be useful for site-specific management when soil conditions are limiting for root exploration and produce survival issues.

Supplementary Materials

Supplementary data are available at *Forest Science* online.

Supplement 1. Table S1. Descriptive information of soil physical, chemical, and water-content characteristics for each horizon in both experiments.

Supplement 2. Figure S1. Soil-penetration resistance (kPa) profile for each treatment between 0 and 50 cm of depth for site 2.

Supplement 3. Figure S2. Diagram of the experimental design used in S1 (shades of gray) and individual plant height (cm) for the six measurement times (colors) during the experimental period (7, 12, 16, 20, 25, and 30 months postplanting).

Supplement 4. Table S2. Number of times each model was superior (i.e., smallest root mean square error of the prediction [RMSEP] or largest R^2, number of times) and median RMSEP and the coefficient of determination (R^2) for three regression models (Cubic, Gompertz, Logistic) fitted for each individual tree in two sites (S_1 and S_2).

Literature Cited

Albaugh, T., J.J. Alvarez, R.A. Rubilar, T.R. Fox, H.L. Allen, J.L. Stape, and O. Mardones. 2015. Long-term *Pinus radiata* productivity gains from tillage, vegetation control, and fertilization. *For. Sci.* 61(4):800–808.

Berry, C.R. 1979. *Subsoiling improves growth of pine on a Georgia Piedmont site*. Res. Note SE-284. USDA Forest Service, Asheville, NC. 3 p.

Brownie, C., L. King, and T. Dube. 2004. *Longitudinal and spatial analyses applied to corn yield data from a long-term rotation trial*. NCSU Institute of Statistics Mimeo Series # 2559. North Carolina State University, Raleigh, NC.

Calegario, N., R.F. Daniels, R. Maestri, and R. Neiva. 2005. Modeling dominant height growth based on nonlinear mixed-effects model: A clonal *Eucalyptus* plantation case study. *For. Ecol. Manag.* 204:11–20.

Carlson, C.A., S.R. Colbert, D.L. Kelting, H.L. Allen, and T.J. Albaugh. 2006. Growth and survival of *Pinus taeda* in response to surface and subsurface tillage in the southeastern United States. *For. Ecol. Manag.* 234(1–3):209–217.

Currie, H.M., R. Rujana, and L. Bartra Vasquez. 2004. Influencia del subsolado en suelos forestales de la provincia de Corrientes. *Rev. Cienc. Agropecu.* 8(2):7–14.

Diggle, P.J., P.J. Heagerty, K.Y. Liang, and S.L. Zeger. 2002. *Analysis of longitudinal data*. 2nd ed. Oxford University Press, Oxford, UK. 400 p.

du Toit, B., C.W. Smith, K.M. Little, G. Boreham, and R.N. Pallett. 2010. Intensive, site-specific silviculture: Manipulating resource availability at establishment for improved stand productivity. A review of South African research. *For. Ecol. Manag.* 259:1836–1845.

Dutkowski, G.W., J. Costa e Silva, A. Gilmour, H. Wellendorf, and A. Aguiar. 2006. Spatial analysis enhances modeling of a wide variety of traits in forest genetics trials. *Can. J. For. Res.* 36:1851–1870.

FAO. 1981. *Eucalyptus for planting*. Forestry Service N° 11. Food and Agriculture Organizations of the United Nations, Rome. 677 p.

Ferreira, G., F. Oliveira, L. Souza, J. Soares, E. Araújo, and I. Silva. 2018. Nitrogen alters initial growth, fine-root biomass and soil organic matter properties of a *Eucalyptus dunnii* Maiden plantation in a recently afforested grassland in Southern Brazil. *Forests* 9(2):62.

Fonseca, F., T. Figueiredo, and A. Martins. 2011. Survival and early growth of mixed forest stands installed in a Mediterranean Region: Effects of site preparation intensity. *For. Ecol. Manag.* 262:1905–1912.

Fortin, M., G. Daigle, C. Ung, J. Begin, and L. Archambault. 2007. A variance–covariance structure to take into account repeated measurements and heteroscedasticity in growth modeling. *Eur. J. Forest. Res.* 126:573–585.

Fox, J.C., P.K. Ades, and H. Bi. 2001. Stochastic structure and individual-tree growth models. *For. Ecol. Manag.* 154:261–276.

Fox, J., H. Bi, and P. Ades. 2007a. Spatial dependence and individual-tree growth models. I. Characterizing spatial dependence. *For. Ecol. Manag.* 245:10–19.

Fox, J., H. Bi, and P. Ades. 2007b. Spatial dependence and individual-tree growth models. I. Modelling spatial dependence. *For. Ecol. Manag.* 245:20–30.

Garcia Prechac, F., M. Pérez Bidegain, S. Christie, and P. Santini. 2001. Efecto de la intensidad de laboreo en el crecimiento aéreo y radicular de Eucalyptus dunnii y sobre algunas propiedades físicas y químicas del suelo. *Agrociencia Uruguay.* 5(1):1–9.

González, P., M. Pérez Bidegain, L. Gutiérrez, L. Martinez, and F. Garcia Prechac. 2012. Evaluation of different tillage intensities on *Eucalyptus grandis* on a typic Hapludult of Uruguay. *Agrociencia Uruguay.* 16(3):302–305.

González Barrios, P., M. Pérez Bidegain, and L. Gutiérrez. 2015. Effects of tillage intensities on spatial soil variability and site-specific

management in early growth of Eucalyptus grandis. *For. Ecol. Manag.* 346:41–50.

Graciano, C., J. Guiamet, and J. Goya. 2005. Impact of nitrogen and phosphorus fertilization on drought responses in *Eucalyptus grandis* seedlings. *For. Ecol. Manag.* 212:40–49.

Grattapaglia, D., and M. Kirst. 2008. Eucalyptus applied genomics: From gene sequences to breeding tools. *New Phytol.* 179:911–929.

Gregoire, T.G., O. Schabenberger, and J.P. Barrett. 1995. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent plot measurements. *Can. J. For. Res.* 25:137–156.

Haines, L.W., T.E. Maki, and S.G. Sanderford. 1975. The effect of mechanical site preparation treatments on soil productivity and tree (*Pinus taeda* L. and *P. elliottii* Engelm. var. elliottii) growth. P. 379–395 in *Forest soils and forest land management*, Bernier, B., and C.H. Winget (eds.). Les Presses de l'University Laval, Quebec, Canada.

Hewitt, J.E., S.E. Trush, and V.J. Cummings. 2001. Assessing environmental impacts: Effects of spatial and temporal variability at likely impact scales. *Ecol. Appl.* 11(5):1502–1516.

Hong, N., J.G. White, M.L. Gumpertz, and R. Weisz. 2005. Spatial analysis of precision agriculture treatments in randomized complete blocks: Guidelines for covariance model selection. *Agron. J.* 97:1082–1096.

Horn, R., H. Domzal, A. Slowinska-Jurkiewicz, and C. Ouwerkerk. 1995. Soil compaction processes and their effects on the structure of arable ansoils and the environment. *Soil Till. Res.* 35:23–36.

Hu, X., and J. Spilke. 2009. Comparison of various spatial models for the analysis of cultivar trials. *New Zeal. J. Agr. Res.* 52(3):277–287.

Kröger, M. 2012. *Global tree plantation expansion: A review.* ICAS Review Paper Series No 3. Initiatives in Critical Agrarian Studies (ICAS), Rotterdam.

Law, R., D.J. Murrell, and U. Dieckmann. 2003. Population growth in space and time: Spatial logistic equations. *Ecology* 84(1):252–262.

Lee, J., and D.W.S. Wong. 2001. *Statistical analysis with ArcView GIS.* John Wiley and Sons, New York. 192 p.

Lincoln, M.C., R.E. Will, J.R. Britt, E.A. Cater, and L.A. Morris. 2006. Relationship between tillage intensity and initial growth of loblolly pine seedlings. P. 191–194 in *Proceedings of the 13th Biennial Southern Silviculture Conference*, Connor, K.F. (ed.). Gen. Tech. Rep. SRS-92, USDA Forest Service, Asheville, NC.

Liu, J., and H.E. Burkhart. 1994. Spatial characteristics of diameter and total height in Juvenile Loblolly Pine (*Pinus taeda* L.) plantations. *For. Sci.* 40(4):774–786.

Liu, J., and P.S. Ashton. 1999. Simulating effects of landscape contexts and timber harvest on tree species diversity. *Ecol. Appl.* 9:186–201.

López, M.V., and J.L. Arrue. 1995. Efficiency of an incomplete block design based on geostatistics for tillage experiments. *Soil Sci. Soc. Am. J.* 59:1104–1111.

Matérn, B. 1960. *Spatial variation.* Technical report, Statens Skogsforsningsinstitut, Stockholm. 144 p.

Morris, L., and R. Lowery. 1988. Influence of site preparation on soil conditions affecting stand establishment and tree growth. *South. J. Appl. For.* 12(3):170–178.

O'Rourke, S., and G. Kelly. 2015. Spatio-temporal modelling of forest growth spanning 50 years—The effects of different thinning strategies. *Procedia Environ. Sci.* 26:101–104.

O'Rourke, S., M.M. Siúrtáin, and G. Kelly. 2016. Characterizing dependence of Irish Sitka spruce stands using spatio-temporal sum-metric models. *For. Sci.* 62(5):490–502.

Overbeek, W., M. Kröger, and J.F. Gerber. 2012. *An overview of industrial tree plantation conflicts in the global South. Conflicts, trends, and resistance struggles.* EJOLT Report No. 3, Environmental Justice Organisations, Liabilities and Trade. 100 p.

Perdomo, C., J. Durán, and P. Llovet. 2007. Soil and plant indices for predicting eucalypt response to nitrogen in Uruguay. *Soil Sci. Soc. Am. J.* 71:1708–1718.

Piepho, H.-P., A. Büchse, and C. Richter. 2004. A mixed modelling approach for randomized experiments with repeated measures. *J. Agron. Crop Sci.* 190(4):230–247.

Piepho, H.-P., C. Richter, J. Spilke, K. Hartung, A. Kunick, and H. Thöle. 2011. Statistical aspects of on-farm experimentation. *Crop Pasture Sci.* 62:721–735.

R Core Team. 2018. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Raimundo, M.R., H. F. Scolforo, J.M. de Mello, J.R.S. Scolforo, J.P. McTague, and A.A. dos Reis. 2017. Geostatistics applied to growth estimates in continuous forest inventories. *For. Sci.* 63(1): 29–38.

Richter, C., and B. Kroschewski. 2012. Geostatistical models in agricultural field experiments: Investigations based on uniformity trials. *Agron. J.* 104:91–105.

Resquin, F., R.M. Navarro-Cerrillo, L. Carrasco-Letelier, and C.R. Casnati. 2019. Influence of contrasting stocking densities on the dynamics of above-ground biomass and wood density of *Eucalyptus benthamii*, *Eucalyptus dunnii*, and *Eucalyptus grandis* for bioenergy in Uruguay. *For. Ecol. Manag.* 438:63–74.

SAS Institute 2011. *SAS® Version 9.2.* SAS Institute, Cary, NC.

Saud, P., T. Lynch, K.C. Anup, and J.M. Guldin. 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.

Skovsgaard, A.P., and J.K. Vanclay. 2013. Forest site productivity: A review of spatial and temporal variability in natural site conditions. *Forests* 86:305–315.

Smith, C.W., K.M. Little, and C.H. Norris. 2001. The effect of land preparation at re-establishment on the productivity of fast growing hardwoods in South Africa. *Aust. For.* 64(3):165–174.

Strudley, M., T. Green, and J. Ascough. 2008. Tillage effects on soil hydraulic properties in space and time: State of the science. *Soil Till. Res.* 99:4–48.

Tsoularis, A., and L. Wallace. 2002. Analysis of logistic growth models. *Math. Biosci.* 179:21–55.

Verhaegen, D., and C. Plomion. 1996. Genetic mapping in *Eucalyptus urophylla* and *E. grandis.* RAPD markers. *Genome* 39:1051–1061.

Wheeler, M.J., R.E. Will, D.B. Markewitz, M.A. Jacobson, and A.M. Shirley. 2002. Early loblolly pine stand response to tillage on the piedmont and upper coastal plain of Georgia: Mortality, stand uniformity, and second and third year growth. *South. J. Appl. For.* 26(4):181–189.

Wolfinger, R.D. 1996. Heterogeneous variance–covariance structures for repeated measures. *J. Agr. Biol. Environ. Stat.* 1:205–230.

Worrell, R., and A. Hampson. 1997. The influence of some forest operations on the sustainable management of forest soils—A review. *Forestry* 70(1):61–85.

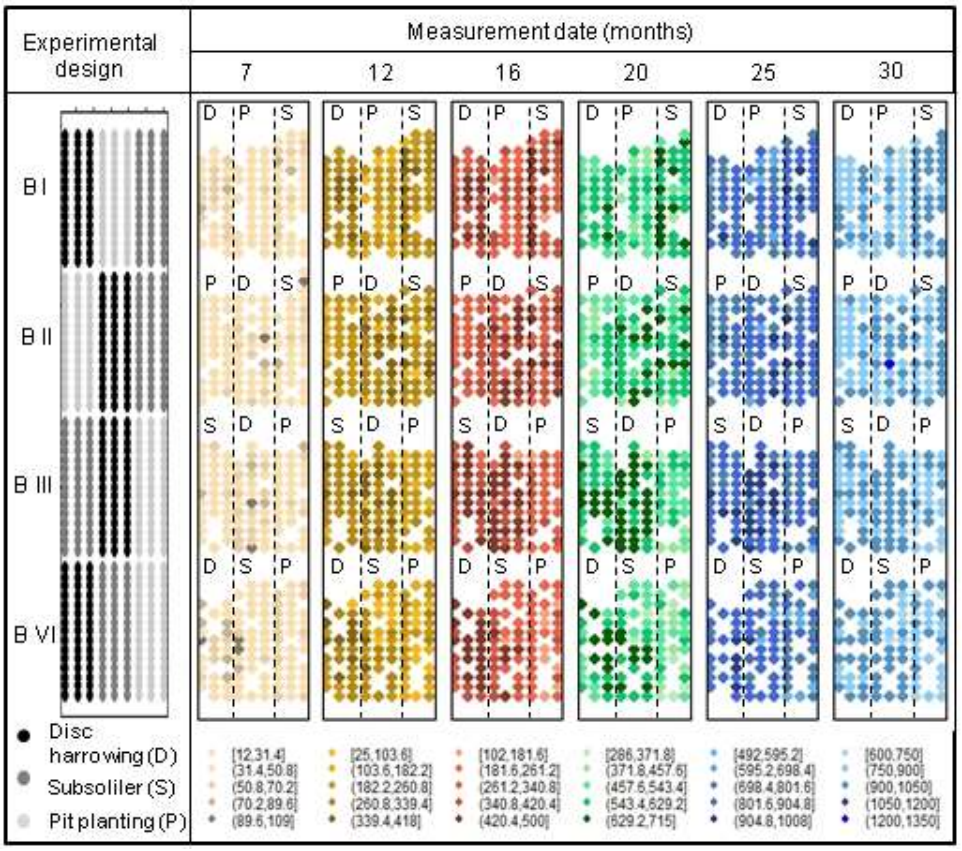
Zhang, L., and S. Haijin. 2004. Local modeling of tree growth by geographically weighted regression. *For. Sci.* 50:225–244.

4.7. SUPPLEMENTAL MATERIAL

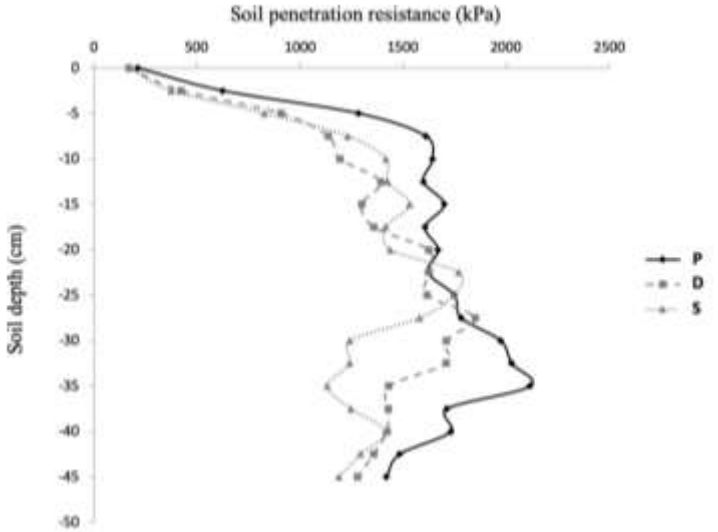
Supplement 1. Table 1. Descriptive information of soil physical, chemical and water content characteristics for each horizon in both experiments.

Site	Horizon	Depth (cm)	OM (%)	Sand (%)	Silt (%)	Clay (%)	FC (-10 KPa)	PWP (-1500KPa)	BD (g/cc)	PAW (%)
1	A	10.0	2.0	71.0	15.0	14.0	22.2	8.9	1.42	13.3
	R	-	-	-	-	-	-	-	-	-
2	A	7.0	4.6	67.7	19.9	16.4	29.3	12.9	1.19	16.5
	Bt	36.0	2.1	63.8	14.7	21.5	29.4	13.1	1.41	16.3

OM: organic matter (%); FC: Field capacity (% of volumetric water content); Permanent wilting point (% of volumetric water content); BD: Bulk density; PAW: Plant available water content.



Supplemental figure 1. Diagram of the experimental design used in S_1 (shades of gray) and individual plant height (cm) for the six measurement times (colors) during the experimental period (7, 12, 16, 20, 25, and 30 months post-planting).



Supplemental figure 2. Soil penetration resistance (kPa) profile for each treatment between 0 and 50 cm of depth for site 2.

Supplement 3. Table 2. Number of times each model is selected (# times) and median of root mean square error of prediction (RMSEP) and coefficient of determination (R^2) for three regression models (Cubic, Gompertz, Logistic), fitted for each tree in both sites.

Site	Model	RMSEP		R2	
		# times	Median	# times	Median
S1	Cubic	328	27.625	288	0.992
	Gompertz	6	33.585	17	0.989
	Logistic	12	43.307	41	0.983
S2	Cubic	221	38.940	139	0.985
	Gompertz	0	38.940	0	0.985
	Logistic	130	44.712	215	0.982

5. SPATIAL VARIABILITY OF SOIL ORGANIC CARBON AND PHOSPHORUS IN NATURAL GRASSLANDS

5.1. RESUMEN

Los pastizales naturales almacenan grandes cantidades de carbono en sus suelos y su manejo determina su dinámica. La resiembra de leguminosas junto con los fertilizantes fosforados suele utilizarse para promover su productividad y calidad nutricional. Para evaluar el efecto de esta tecnología sobre el carbono orgánico del suelo (SOC), deben considerarse ensayos que incluyan parcelas de gran tamaño y varios muestreos dentro de éstas. La alta variabilidad espacial del SOC relacionada con los cambios en los tipos de suelo, la topografía y el uso anterior de la tierra es habitual en los suelos agrícolas y debe tenerse en cuenta en el análisis estadístico de los impactos de las prácticas de gestión en los experimentos a escala de campo. Se evaluó si el uso de modelos espaciales para la correlación dentro de la parcela mejora el rendimiento del diseño experimental y si la siembra de leguminosas en cobertura y las fertilizaciones con fósforo (P) inducen cambios en la distribución espacial de SOC y P después de 9 años. Los tratamientos fueron: campo natural (NG), campo natural con *Trifolium repens* L. y *Lotus corniculatus* L. fertilizados anualmente con 30 kg de P_2O_5 ha^{-1} $año^{-1}$ (OP30) y con 60 kg de P_2O_5 ha^{-1} $año^{-1}$ (OP60). Se utilizaron once puntos georreferenciados en cada parcela para el muestreo de SOC y P lábil a dos profundidades. Se utilizaron diferentes modelos mixtos que incluían o no la correlación espacial dentro de la parcela para evaluar el efecto del tratamiento. Para la comparación de los modelos, se calcularon el AIC, la varianza de parcela y residual y los errores estándar de las medias. Además, se caracterizó la distribución espacial de SOC y P en cada tratamiento mediante semivariogramas y se utilizó una prueba F para determinar si los tratamientos inducían cambios en la distribución espacial de estas variables. Las ventajas de incluir la correlación espacial dentro de la parcela no fueron evidentes y sólo se eligió un modelo espacial en un solo caso. Esto no significa que no haya correlación espacial entre las observaciones dentro de la

parcela, ya que la correlación espacial por tratamiento se confirmó en la mayoría de los casos. Esta aparente discrepancia podría deberse principalmente a que se observaron diferentes patrones espaciales por tratamiento, confirmado con la prueba F, por lo que es razonable pensar que un modelo que proponga una estructura de correlación común para todos los tratamientos no mejorará el ajuste del modelo. En resumen, la fertilización con fósforo durante nueve años consecutivos cambió la distribución espacial del contenido de SOC y P en el suelo y aumentó la varianza del fósforo en el suelo de los tratamientos OP. Sería aconsejable aumentar la intensidad de muestreo en futuros estudios para entender mejor cómo estas tecnologías afectan al contenido y la distribución espacial de SOC y P.

Palabras clave: campo natural, fertilización fosfatada, variogramas, modelos mixtos

5.2. SUMMARY

Natural grasslands store large amounts of carbon in their soils and their management determines its dynamics. The overseeding of legumes joint with phosphorus fertilizers is usually used to promote their productivity and nutritional quality. To evaluate the effect of this technology on soil organic carbon (SOC), trials involving large plot sizes and several within-plot samplings should be considered. High SOC spatial variability related to changes in soil types, topography and past land use is usual in agricultural soils and should be considered in the statistical analysis of management practices impacts in field scale experiments. We evaluated if the use of spatial models for within plot correlation improves the performance of the experimental design, and if overseeded legumes and P-fertilized treatments induce changes in the spatial distribution of SOC and phosphorus (P) after 9 years. Treatments were: natural grassland (NG), natural grassland overseeded with *Trifolium repens* L. and *Lotus corniculatus* L. fertilized annually with 30 kg of P_2O_5 ha^{-1} $year^{-1}$ (OP30) and with 60 kg of P_2O_5 ha^{-1} $year^{-1}$ (OP60). Eleven georeferenced points in each plot were used for SOC and labil P sampling at two depths. Different mixed models including or not the spatial correlation within the plot were used to evaluate treatment effect. For model comparison, AIC, plot and residual variance

components and standard errors were computed for all models. Additionally, the spatial distribution of SOC and P in each treatment was characterized using semivariograms and an F-test was used to determine whether the treatments induced changes in the spatial distribution of these variables. The advantages of including within-plot spatial correlation were not evident and only a spatial model was chosen for one case. This does not mean that there is no spatial correlation between the within-plot observations, since spatial correlation by treatment was confirmed in most cases. This apparent discrepancy could be mainly due to the fact that different spatial patterns were observed depending on the treatment, as confirmed by the F-test for model comparison, so it is reasonable to think that a model proposing a common correlation structure for all treatments will not improve the model fit. In summary, fertilization with phosphorus for nine consecutive years changed the spatial distribution of SOC and P soil content and increased the variance of phosphorus of OP treatments. It would be advisable to increase the sampling intensity in future studies to better understand how these technologies impacts SOC and P content and their spatial distribution.

Keywords: natural grasslands, phosphate fertilization, variograms, mixed models

5.3. INTRODUCTION

The South American Campos is an ecological region covering central-eastern Argentina, Uruguay, and southern Brazil, dominated by spatially heterogeneous temperate and subtropical natural grasslands conformed by a complex mosaic of species (Jaurena et al., 2021). These ecosystems are mainly used for extensive livestock production and provide a range of valuable ecosystem services, including water supply and flow regulation, carbon storage, erosion control and climate mitigation, among others (Bengtsson et al., 2019). However, the sustainability of these grasslands may be threatened by changes in land use and intensification processes based on inputs.

The overseeding of legumes with phosphorus fertilizers is usually used to improve the productivity and the nutritional quality of natural grasslands (Jaurena et al., 2016) and restore nutrients losses caused by grazing (Bondaruk et al., 2020). Initially, phosphorus fertilizers promote symbiotic nitrogen fixation by rhizobium of introduced legumes, increasing native grasses production (Rodríguez et al., 2007) and C and N turnover in soils (Poeplau et al., 2018). However, repeated phosphorus applications could negatively affect the diversity of species adapted to low P soils content (Jaurena et al. 2016) and increase the risk of P losses in runoff with the potential to generate eutrophication in water bodies (Li et al., 2020, Nze Memiaghe et al., 2020). Thus, these technologies should be managed with care since they could lead to reductions of ecosystems services (Jaurena et al., 2016).

Soils under permanent grasslands store large amounts of carbon and their management plays a role as potential sinks in the global carbon cycle (Bengtsson et al., 2019). Soil organic carbon is one of the most widespread indicators of soil quality (due to its important role in many soil dynamics, such as soil erosion and crop water availability) as well as an energy source for microbial processes (Husein et al., 2019, Reeves, 1997). From a sustainable perspective, agricultural practices that preserve or increase soil organic carbon are critical to improve source/sink balance of

C and mitigate climate change (Dingkuhn et al., 2020).

To correctly assess the field scale impact of management practices on soil organic carbon and nutrients it is important to consider their spatial distribution. Understanding the spatial variability of SOC is essential to assess soil fertility, productivity and carbon source-sink potential, and establish sustainable management practices (Zhang et al., 2020). There is strong evidence that soil organic carbon has clear spatial patterns (Marriott et al., 1997, Terra et al., 2004, Zhang et al., 2020) and that the greatest spatial variation occurs in relation to topography (Burke et al., 1999, Kravchenko et al., 2006), as well as to land use (Wang et al., 2009). The latter authors found that total soil nitrogen and phosphorus spatial variability are significantly affected by land use type (cultivated land versus natural grassland).

SOC response to changes in land use and management practices is generally slow and gradual, thus, it is difficult to detect short term differences between strategies (Kravchenko et al., 2006). To evaluate the effect of these input-based technologies on soil organic carbon, trials involving large plot sizes and several samples within the plot should be considered. It is expected that closer subsamples are correlated and this spatial correlation can be affected by management practices. Soil spatial variability is widely recognized as one of the most important factors affecting the comparisons of treatment effects in field scale experiments (López and Arrúe, 1995), because it often occurs gradually and sometimes it is not efficiently captured by the experimental design (Grondona and Cressie, 1991). To deal with this situation, the spatial variation at the plot level can be estimated and included in the response variable model. Several models that include different spatial strategies have been used (Borges et al., 2019, Brownie et al., 1993, Casler and Undersander, 2000). In some situations, models that include spatial correlation are generally more efficient than those that do not include it (Brownie et al., 1993, González-Barrios et al., 2020, Kravchenko et al., 2006, Mallarino et al., 2000).

In addition to understanding the effects of overseeded legumes and phosphorus

fertilizers on soil organic carbon content and its spatial distribution in natural grasslands, it is important to also evaluate their effects on the accumulation of this nutrient in soil and its spatial variability. Periodic phosphorus fertilizations can affect its spatial heterogeneity, even with broadcast fertilizations. Nze Memiaghe et al. (2020) observed soil phosphorus accumulation in an old grassland in Canada (a permanent pasture for 10 years) because of long-term manure applications. They found that repeated applications of organic fertilizers with high P concentration in permanent grasslands increased soil P content but decreased soil P variability and spatial dependence in the long-term.

Understanding spatial variability of P can help improve crop management and the profitability and sustainability of agricultural businesses through the site specific use and management of P fertilizers, reducing P losses by runoff, and protecting the environment (Li et al., 2020, Nze Memiaghe et al., 2020).

The hypothesis that we proposed were: 1) accounting for any existent spatial correlation as an add-on model component may improve the precision of estimated treatment means, 2) overseeding legumes coupled with consecutive annual phosphorus fertilization (OP) applied to natural grasslands increase SOC and P content in functionally relevant amount and 3) OP treatments changed the spatial distribution of soil organic carbon and phosphorus content after 9 years.

For this reason, the aims of this work are to evaluate: 1) if the use of spatial models for within plot correlation improves the performance of the experimental design, 2) if overseeded and P-fertilized (OP) grasslands change SOC and P mean content and 3) if OP treatments induce changes in the spatial distribution of soil organic carbon and phosphorus after 9 years.

5.4. MATERIALS AND METHODS

5.4.1. Experimental field

The 30 ha trial was installed in the fall of 1996 in the Agricultural Research Station of 'Palo a Pique' of the National Institute of Agricultural Research (INIA), in Treinta y Tres, Uruguay (33 °: 15'36 "S, 54 °: 29 ' 26 "W, 60-m elevation). According to the USDA-National Cooperative Soil Survey (National Cooperative Soil Survey, 2014), soils at the experimental site were classified as Abruptic Argiaquolls and Oxyaquic Vertic Argiudoll (fine, smectitic and thermic). The main soil characteristic (0-15 cm depth) at the site in 2005 were: pH (1:1 soil:water) = 5.5; extractable K = 0.27 meq.100 g^{-1}; clay = 180 $g.kg^{-1}$; sand = 510 $g.kg^{-1}$.

A randomized complete block design with 5 replications (2 ha each experimental unit) was used. Three treatments were evaluated: i) natural grassland (NG), ii) natural grassland overseeded with *Trifolium repens* L. and *Lotus corniculatus* L. fertilized at sowing with 45 kg of P_2O_5 ha^{-1} and re-fertilized annually with 30 kg of P_2O_5 ha^{-1} $year^{-1}$ (OP30) and iii) natural grassland overseeded with *Trifolium repens* L. and *Lotus corniculatus* L. fertilized at sowing with 90 kg of P_2O_5 ha^{-1} and re-fertilized with 60 kg of P_2O_5 ha^{-1} $year^{-1}$ (OP60). During the first 3 years, superphosphate was used (21 % P_2O_5 soluble – 23 % P_2O_5 total). Later, rock phosphate (10 % P_2O_5 soluble – 29 % P_2O_5 total) was used. The treatments were applied to the same experimental units every year, for 9 consecutive years. All 2 ha plots (experimental unit) were rotationally grazed by yearling steers and heifers with similar moderate grazing intensities among treatments.

In the spring of 2005, 11 sampling points located in a relatively regular grid in each experimental unit of 2 ha were georeferenced (figure S1). Eight 2.5 cm diameter sampling cores to a depth of 0-5 and 5-15 cm were taken and composited within a 2.5 m radius from the center of each sampling point. Soil samples were dried at 50 °C for 72 hours, ground and passed through sieves of 2000 μm. Soil organic C was

analyzed according to Walkley-Black method (Walkley and Black, 1934). Extractable phosphorus content was measured adapting the citric acid method proposed by Thompson (1995). Fifty milliliters of citric acid (0.5 %) were added to 5 g of soil and stirred for 30 minutes. The extract was filtered through a Whatman No. 2 filter paper. A 5-mL aliquot of filtered extract was removed and mixed with 8 mL of colorimetric reagent (Murphy and Riley, 1962) and then diluted to 50 mL. The blue color development was measured after 45 minutes by absorbance at 882 nm with a UV spectrometer (Thermo Spectronic, Genesys 10S) and compared with known reference concentrations. In each point, values of soil organic carbon (SOC) and citric phosphorus (P) were obtained at both soil depts.

5.4.2. Treatment effects on SOC and P

The analysis model to compare the effects of legume-overseeded grasslands versus natural grasslands on soil organic carbon (SOC) and phosphorus (P) for each depth was selected according to the following procedure:

First, a null model was fitted, which assumes independence of errors, normality, and homoscedasticity, and also that block and treatment effects are additive. Departure from normality and homoscedasticity may indicate the need for data transformation or removal of outlying observations (Schützenmeister et al., 2012). To assess the validity of these assumptions, informal procedures were performed: quantile-quantile plot, that draws the correlation between a given sample and the normal distribution to test for normality, boxplot for residuals per treatment and scatter plots of residuals vs fitted values to test for homoscedasticity, and spatial plot of residuals were obtained to identify spatial correlation patterns. Also, to answer if treatments induce different variability, the residual variances between treatments were compared by Levene's test.

Thereafter, several analysis models that include spatial correlation of errors and/or heterogeneity of variances were adjusted for each variable in each soil stratum.

Evaluated models are detailed in table 1.

Table 1. List of analysis model for soil organic carbon and phosphorus content

Model	**Code**
Independent errors among plots (null model)	NULL
Independent errors among plots + heterogeneous variances per treatment	NULL_H
AR(1) spatial structure for covariance matrix	AR1
AR(1) covariance structure + heterogeneous variances per treatment	AR1_H
Spherical covariance structure	SPH
Spherical covariance structure + heterogeneous variances per treatment	SPH_H
Exponential covariance structure	EXP
Exponential covariance structure + heterogeneous variances per treatment	EXP_H
Gaussian covariance structure	GAU
Gaussian covariance structure + heterogeneous variances per treatment	GAU_H

All these models can take the general form of linear mixed models:

$$y = X\beta + Zu + e \qquad [1]$$

Where y is the vector of observations, X is a matrix of constants associated with the fixed effects of the design, β is a vector of unknown fixed effects (treatments and blocks), Z is a matrix of constants associated with the random effects, u is a vector of random effects (block per treatment interaction, that is, plot effect) and e is a vector of random residual errors (subsampling errors). The random effect vector is assumed to be distributed as $u \sim MN(0, G)$ and the residual vector is assumed to be distributed as $e \sim MN(0, R)$, where MN denotes a multivariate normal distribution, G is the covariance matrix among random effects and R is the covariance matrix among the random residual errors.

For NULL model, R is a block-diagonal matrix of 165 × 165, that has σ_e^2 for all

diagonal elements and zeroes for the off-diagonal elements.

For the spatial models including random plot effects, R structure is more complex: the covariances depend on spatial distances among observations within plots, assuming isotropy (spatial correlation does not depend on the direction). The correlation is zero for two observations from different plots. Observations from two subsamples of the same plot with Euclidean distance (d) d > 0 (measured between subsample centers) have covariance $Cov(d) = \sigma_e^2 f(d)$.

AR1: $f(d) = \rho^d$ [2]

GAU: $f(d) = exp^{-\left(\frac{d}{A}\right)^2}$ [3]

EXP: $f(d) = exp^{-\left(\frac{d}{A}\right)}$ [4]

SPH: $f(d) = \begin{cases} 1 - \frac{3d}{2A} + \frac{d^3}{2A^3}, & d < A \\ 0, & d > A \end{cases}$ [5]

where ρ is a correlation parameter and A is the range parameter.

For models with heterogeneous variances ('_H'), variances along the diagonal of R matrix are different for each treatment.

For each of these models, the Akaike Information Criterion (AIC) was computed. Considering the AIC criteria to select a model (smaller value is better), the minimum difference required to choose between two models was two points of AIC (Arnold, 2010). If two models differed by less than two points, parsimony was prioritized and the simplest model was chosen. Additionally, plot and residual variance components (σ_p^2 and σ_ε^2 respectively) and standard error for treatment means (SE) were extracted for all models in order to evaluate their performance on estimation precision. For heterogeneous variance models, the reported values of σ_ε^2 and SE were those corresponding to treatment OP30.

In all cases, the estimation method of restricted maximum likelihood (REML) was

used. All analysis models were run using the *nlme* package (Pinheiro et al., 2020) and Tukey tests for multiple mean comparisons were performed using *emmeans* package (Lenth, 2020) of the R software (R Core Team, 2020).

5.4.3. Spatial variability of SOC and P

To characterize the spatial pattern of organic carbon for each treatment, the spatial variability of SOC and P was studied for each depth, using all sampling points in each treatment (55). For this purpose, residuals of null model were obtained after removing treatment and block effects, and the overall mean of SOC for each depth was added to residuals.

Generally, residuals of SOC distributions were positively skewed; therefore, log-transformations were performed for SOC at both depths to make them comparable. In the case of P, log-transformations were only performed for the deepest layer (5-15 cm). For each depth and treatment, Spherical, Exponential and Gaussian models were fitted and compared by weighted residual sum of squares (WRSS, smaller is better). Finally, the SPH model was chosen for all cases because it had the smallest WRSS value in 5 out of 6 times. In order to determine whether the treatments changed the spatial distribution of the variables of interest, we fit two models to each response and depths. One of them allowed any of the 3 semivariogram parameters (sill, range and nugget) to differ between treatments (full model) and the second one assumed that all 3 treatments had the same variogram parameters (reduced model). We used the weighted residual sum of squares to compute a model comparison F-test (null vs. reduced model).

Partial sill, range and nugget estimations of each semivariogram model were obtained and the proportion of spatial structure (PSS) was computed as follows:

$$PSS = \sigma_0^2 / (\sigma_0^2 + \sigma_R^2) \qquad [6]$$

Where σ_0^2 is the partial sill and σ_R^2 is the nugget variance.

Additionally, using the predictions from ordinary kriging, two statistics from cross-validation were obtained: mean squared deviation ratio (MSDR) as in Oliver and Webster (2014) and the Pearson correlation coefficient between observed and predicted values (r_op).

$$MSDR = \frac{1}{N}\sum_{i=1}^{N}\frac{\{z(\mathbf{x}_i)-\hat{Z}(\mathbf{x}_i)\}^2}{\hat{\sigma}_K^2(\mathbf{x}_i)} \qquad [7]$$

Where $z(\mathbf{x}_i)$ is the i*th* observation at location $\mathbf{x}_i$, $\hat{Z}(\mathbf{x}_i)$ is the kriged prediction at $\mathbf{x}_i$, and $\hat{\sigma}_K^2(\mathbf{x}_i)$ is the kriging variance.

5.5. RESULTS

5.5.1. Model performance

No departures from model assumptions were observed for soil organic carbon at the surface layer ($SOC_{0\text{-}5}$), and no differences among treatment variances were found in the Levene's test (p-value = 0.5983). Nevertheless, all models detailed in table 1 were fitted for $SOC_{0\text{-}5}$. AIC, plot and residual variance and standard error values for all models are shown in table S1. Since most of the models differ from less than two points of AIC, the NULL model was selected for treatment mean comparisons. The variance between plots was slightly lower in the spatial models and the residual variance was slightly higher, with the SE remaining almost the same for all models (table S1).

For $SOC_{5\text{-}15}$, the assumptions of normality and homocedasticity were fulfilled according to the exploratory analysis and Levene´s test (p-value = 0.11, but not the independence of residuals. A certain spatial pattern of the residuals of NULL model was observed (figure S2).

The EXP model had the lowest AIC value (table S2), obtaining a substantial difference with any other model, confirming the spatial correlation observed in figure

S2. In general, the spatial models had lower inter-plot variance, especially the EXP model, while the opposite was true for the residual variance. No differences in SE values were observed among the models considered (table S2).

Strong heteroscedasticity among treatments was observed for phosphorus content at the top layer ($P_{0\text{-}5}$) (figure S3), also confirmed by Levene´s test (p-value = 1.18×10^{-5}). Although this could be partially corrected with a model that assumes heterogeneous variances among treatments, this figure also suggests lack of additivity between block and treatment effects. For this reason, a log-transformation of $P_{0\text{-}5}$ ($\log(P_{0\text{-}5})$ was carried out to simultaneously lift these two restrictions on model assumptions. After log-transformation, homocedasticity was achieved, confirmed by Levene´s test (p-value = 0.31). No clear spatial pattern was observed for residuals (data not shown).

For $\log(P_{0\text{-}5})$, most of the models differ from each other less than two points of the AIC value, therefore, the NULL model was selected (table S3). Very few differences in SE were observed between the analysis models. There was a minimal trend for spatial models to have lower σ_p^2 and higher σ_ε^2.

For $P_{5\text{-}15}$, heterocedasticity among treatments was confirmed with residual plots (figure S4) and Levene´s test (p-value = 0.003). In this case, since a lack of additivity was not verified between blocks and treatments, no transformation was performed. The AIC values for all fitted models confirm that the model that admits heterogeneity of variances (NULL_H) was the best model (table S4). All models exhibited almost the same values of σ_p^2, σ_ε^2 and SE.

5.5.2. Treatment effects on SOC and P

Soil organic Carbon (SOC)

As it was mentioned before, the NULL model was selected for treatment mean

comparisons at surface and the EXP model for the deeper soil layer (table 2).

Table 2. Treatment means and standard errors of soil organic carbon at 0 to 5 (SOC_{0-5}) and 5 to 15 cm soil layer (SOC_{5-15}), using NULL model for SOC_{0-5} and EXP model for SOC_{5-15}

Treatment	SOC_{0-5}	SOC_{5-15}
NG	2.89 ± (0.11) b†	1.40 ± (0.06) a†
OP30	3.34 ± (0.11) a	1.29 ± (0.06) a
OP60	3.18 ± (0.11) ab	1.27 ± (0.06) a

†Different letters within columns indicate significant differences ($P < .05$)

Natural grassland overseeded with *Trifolium repens* L. and *Lotus corniculatus* L. fertilized annually with 30 units of P_2O_5 ha^{-1} $year^{-1}$ (IP30) presented the highest SOC_{0-5}, being statistically different from NG. Overseeded and P-fertilized grasslands with 60 units of P_2O_5 ha^{-1} $year^{-1}$ (OP60) did not differ from NG. For SOC_{5-15}, besides not finding significant differences among treatment, mean values of SOC_{5-15} for all treatments were similar (table 2).

Phosphorus (P)

For $log(P_{0-5})$, after rejecting the null hypothesis of ANOVA, treatment mean comparisons with the NULL model are shown in table 3.

Table 3. Treatment means and standard errors of log-transformed phosphorus at 0 to 5 cm depth ($log(P_{0-5})$) and soil phosphorus content at 5 to 15 cm depth (P_{5-15}), using NULL model for $log(P_{0-5})$ and NULL_H model for P_{5-15}

Treatment	$log(P_{0-5})$	P_{5-15}
NG	1.60 ± (0.07) c†	1.29 ± (0.26) b†
OP30	3.31 ± (0.07) b	1.85 ± (0.27) ab

OP60	3.94 ± (0.07) a	2.42 ± (0.26) a

†Different letters within columns indicate significant differences ($P < .05$)

As it was expected, treatment with higher phosphorus application had higher soil phosphorus content (OP60 vs. OP30) and both overseeded and P-fertilized treatments differed in phosphorus content from the natural grassland. Back-transformed $P_{0\text{-}5}$ means were 4.95, 27.39 and 51.42 $g.kg^{-1}$ for NG, OP30 and OP60, respectively.

As we mentioned before, for $P_{5\text{-}15}$, the NULL_H model was selected. Although no differences in $P_{5\text{-}15}$ means between OP30 and OP60 were found (table 3), a clear gradient was observed from NG to OP60, being these two treatments significantly different.

5.5.3. Spatial variability of SOC and P

The spatial pattern of SOC appears to be affected by treatments at both depths. As it can be seen in figure 1, the semivariograms for each treatment differ in sill, range or nugget values, and when all semivariances values are considered together (right pannels), a large variability of semivariances values is observed at any distance.

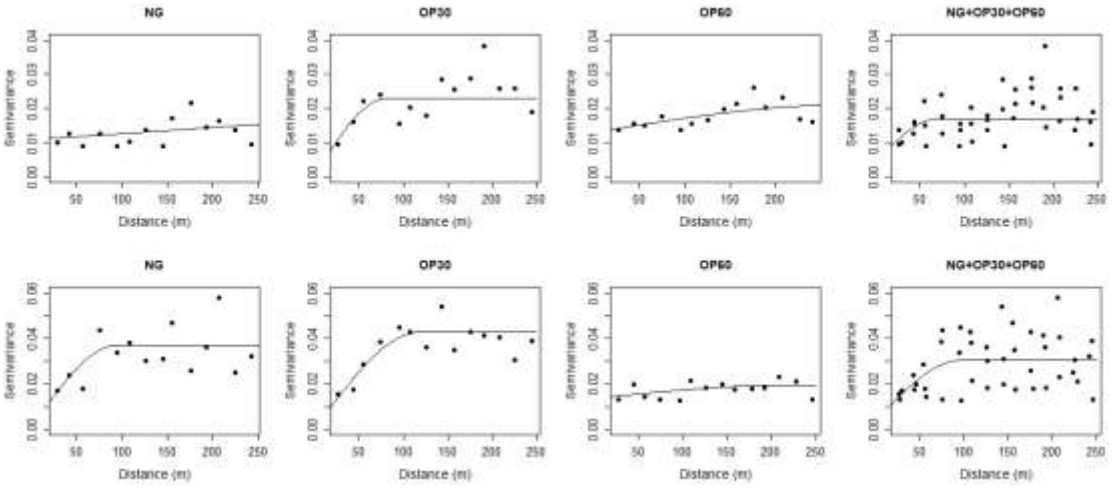


Figure 1. Isotropic sample semivariances versus distance and semivariogram line for

Spherical model, for log (SOC) at 0-5 cm (top) and 5-15 cm (bottom) layer for each treatment (NG, OP30 and OP60) and all treatments together (NG + OP30 + OP60)

The null hypothesis (reduced = full model) of the F-test was rejected for both depths (8.40E-08 and 7.04E-07 p-values for 0-5 and 5-15 cm layer, respectively), indicating that it is preferable to use a model that considers different estimates of semivariogram parameter for each treatment.

In the case of phosphorus, the effect of treatments on the spatial variability of P was more noticeable, as it can be seen in the different variograms in figure 2.

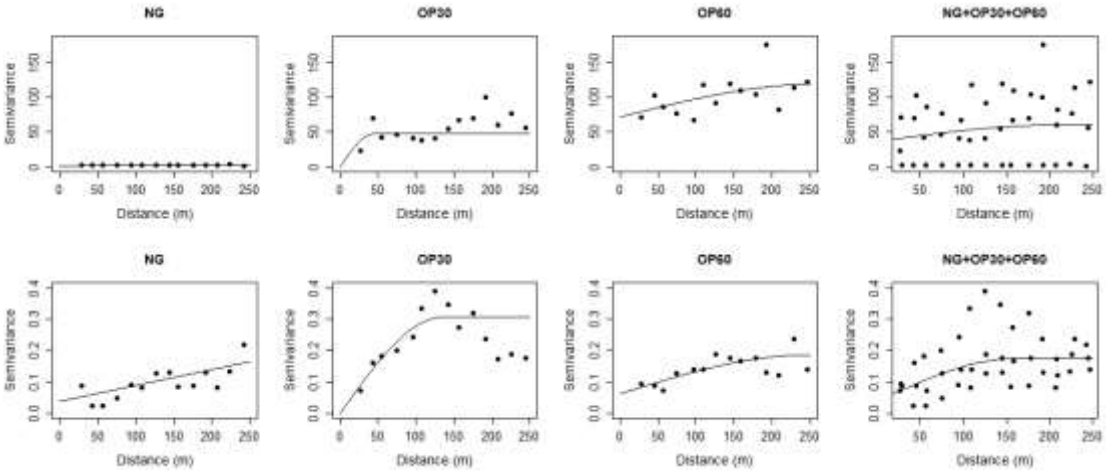


Figure 2. Isotropic sample semivariances versus distance and semivariogram line for Spherical model, for P at 0-5 cm (top) and log(P) 5-15 cm (bottom) layer for each treatment (NG, OP30 and OP60) and all treatments together (NG + OP30 + OP60).

The variograms for P differed markedly in the values of the estimates obtained for the model parameters (range, sill and nugget) for both layers, being more noticeable in the surface layer. The PSS values were 0.5, 1 and 0.4 at surface and 0.9, 1 and 0.67 at the deeper layer, for NG, OP30 and OP60, respectively.

As for SOC, the null hypothesis of the F-test was rejected for both depths (2.38E-13 and 8.88E-16 p-values for 0-5 and 5-15 cm layer, respectively) confirming that it is

more appropriate to allow parameter estimates to vary according to treatment.

In table 4 and 5, cross-validation statistics are presented for SOC and P, respectively. For each treatment separately, MSDR and r_op were obtained using only the 55 locations of sampling points that belong to each treatment, while for all treatments together, these statistics were calculate using all sampling points (165). Considering both (MSDR and r_op), the reduced model was never the best option, both for SOC and P.

Table 4. Mean square deviation ratio (MSDR) and Pearson correlation coefficient between observed and predicted values (r_op) of SOC for each treatment separately (NG, OP30 and OP60) and all treatments together (NG + OP30 + OP60) at both depths.

Depth	**Model**	**MSDR**	**r_op**
0-5 cm	NG	1.09	0.02
	OP30	1.53	0.27
	OP60	1.28	0.19
	NG+OP30+OP60	1.40	0.09
5-15 cm	NG	1.24	0.47
	OP30	1.49	0.47
	OP60	1.28	0.19
	NG+OP30+OP60	1.36	0.42
	Ideal values	1	1

For SOC, at 0-5 cm depth, MSDR values were well above 1, except for the NG model, and the correlation between predicted and observed was very low. At 5-15 cm

depth, MSDR values were all high (although for NG was still the lowest) and correlation values were higher for the topsoil layer for all models, except for OP60 (table 4).

Table 5. Mean square deviation ratio (MSDR) and Pearson correlation coefficient between observed and predicted values (r_op) of P for each treatment separately (NG, OP30 and OP60) and all treatments together (NG + OP30 + OP60) at both depths.

Depth	Model	MSDR	r_op
0-5 cm	NG	1.07	0.71
	OP30	1.31	0.26
	OP60	1.30	0.18
	NG+OP30+OP60	1.41	0.13
5-15 cm	NG	0.87	0.61
	OP30	1.84	0.38
	OP60	0.94	0.60
	NG+OP30+OP60	1.47	0.48
	Ideal values	1	1

In the case of P, at 0-5 cm depth, the model for NG treatment was the best overall (MSDR closer to 1 and r_op higher), while at 5-15 cm depth, the best results for both indicators were obtained with models for NG and OP60 treatments (table 5).

5.6. DISCUSSION

To obtain accurate estimates of the impact of input-based technologies on soil

organic carbon and its spatial variability are key steps to improve the sustainability of agricultural systems. Also, if these technologies imply broadcast fertilization with phosphorus, characterizing the spatial variability of P becomes crucial to design potential site-specific fertilization plans. Responses of soil properties to these managements are gradual and slow, so it is difficult to evidence significant differences in the short term. Moreover, many times large plot sizes need various subsamples per plot that may be spatially correlated. Therefore, the use of the spatial information to increase efficiency in treatment effect comparisons and to design sampling strategies could bring additional advantages.

5.6.1. Model performance

The spatial models in general did not perform better than the model assuming independent errors, except for SOC in the deepest soil layer, although spatial patterns were observed for both variables. The advantages of including both inter-plot and intra-plot spatial correlations have been previously reported, although there is much more precedent for including the former (Borges et al., 2019, Brownie et al., 1993, González-Barrios et al., 2020, Kravchenko et al., 2006, Mallarino et al., 2000) than for models that include the latter (Slaets et al., 2021).

For SOC, at 0-5 cm layer, AIC values were similar for all models; therefore, the null model was chosen for the data analysis. However, when studying its spatial distribution, spatial correlation of SOC was observed in some treatments. This discrepancy could be due to at least two causes. First, different spatial patterns between treatments were observed, so it is reasonable to think that a model proposing a common correlation structure to all treatments would not improve the model fit. Second, the number of sampling points per plot in our study (11 subsamples per plot) may not have been large enough to obtain an accurate estimate of spatial structure. It is important to keep in mind that when a random effect is specified for plots in *nlme* package, then plots are considered as subjects and spatial correlation of two observations will be modeled only within plots but not between plots (Slaets et al.,

2021). Although the spatial pattern is estimated by pooling information across all plots, maybe the weakness is that we do not have information about squared difference for large distances, because within plot distances were all relatively short.
For the deepest soil layer (5-15 cm), more noticeable changes of the spatial models were observed in relation to the null model, especially for the EXP model, which obtained a much lower AIC value, a lower plot variance and higher residual variance than the NULL model. With the exception of the AR1 model, which always behaved very similar to the NULL model, the ratio $\sigma_p^2/\sigma_\varepsilon^2$ decreased from approximately 0.28 to 0.13 for the spatial models and, in the case of the EXP model, it dropped almost to 0. However, this decrease did not translate into a reduction in SE, with all models presenting very similar values. It appears that spatial models alter the estimated (within-plot) error variance, which induces changes in the estimate of the plot variance. Both variance components affect the standard error of the treatment mean, although the weight of each component is not clear.

In the case of phosphorus, the most noticeable feature was the great heterogeneity of variances observed between treatments. Overseeding and P-fertilization induced changes in the spatial variances of phosphorus at both depths. At the surface layer, the estimated variance before transformation was equal to 1.93, 57.78 and 106.45 for NG, OP30 and OP60, respectively. These increments in phosphorus content variance were expected for the last two treatments, but mainly at the top layer, since phosphorus was applied on surface and this nutrient has little mobility, mainly in clayey soils (Azevedo et al., 2018).

5.6.2. Treatment effects on SOC and P

As it was expected for all treatments, SOC mean content was higher in surface than in deeper soil layer (Husein et al., 2019, Zhang et al., 2020). Treatment effects on SOC was confirmed for the surface layer (OP30 had higher SOC mean content than the NG treatment), while, for the deepest layer (5-15 cm), treatment means did not reach significant differences. Bondaruk et al. (2020), comparing natural grasslands

and natural grasslands overseeded with legumes and fertilized with phosphorous in several farms in Uruguay, showed both increases and decreases in SOC content, depending on the farm considered. They suggested that carbon sequestration can be achieved after legumes introduction in grazed natural grasslands, but this achievement will depend on how farms apply this practices and the grazing management.

Beyond the statistical differences, within each treatment, the SOC content difference between the two soil layers (Δ) was more noticeable for the overseeded and P-fertilized treatments than for natural grassland, being Δ = 1.49 for NG, 2.05 for OP30 and 1.91 for OP60. Bondaruk et al. (2020) also observed changes in the vertical distribution of carbon, where overseeded with legumes and P-fertilized treatments showed higher differences of SOC content between 0-5 and 5-10 cm layer than in natural grasslands. Legumes introduction accompanied with phosphorus fertilization for a long period of time can cause significant changes in native community and invasion of exotic species in the long-term (Jaurena et al., 2016). Based on our same field trial, these authors found that in the overseeded and P-fertilized grasslands, exotic grasses displaced native species and overseeded legumes. Legumes tend to have shallower roots than native species (Formoso, 2011). Also, exotic grasses as *Cynodon dactylon* usually concentrate the highest density of roots in the most superficial layers, especially in heavy soils (Horowitz, 1996). Therefore, for these OP treatments, it was expected to find higher SOC content in the top layer, due to higher root density and microbial activity. On the other hand, the higher species richness and higher proportion of native species in natural grasslands can cause a less heterogeneous distribution of SOC content in the first 15 cm of soil. This could partly explain why the differences between SOC content of 0-5 and 5-15 layers is larger in overseeded and P-fertilized grasslands.

As expected, overseeded and P-fertilized treatments had higher soil P content in both layers (although not statistically difference was found between OP30 and NG at 5-15 layer). Dominant soils in the region are deficient in P, so it is a common practice to add P to improve grassland performances (Bondaruk et al., 2020, Jaurena et al.,

2016).

5.6.3. Spatial variability of SOC

The spatial variability of SOC was affected by treatments, showing different spatial patterns between them. At the surface layer, for NG treatment, no clear spatial structure was observed, resembling a "pure nugget effect" model, where the range exceeds 3000 meters. On the other hand, for overseeded and P-fertilized treatments, a clearer spatial correlation was observed, reaching ranges of 80 and 278 for OP30 and OP60 respectively. At a deeper soil layer (5-15 cm), a spatial correlation of SOC was observed for all treatments; although the shape of the semivariograms varied (e.g. the range was 95, 120 and 225 m for NG, OP30 and OP60, respectively). The existence of SOC spatial correlation has already been widely reported previously for different soils and land use (Simbahan et al., 2006, Terra et al., 2006, Wang et al., 2009, Zhang et al., 2020).

The stronger spatial dependence of deeper layers compared to the top layer was previously reported for other regions (Zhang et al., 2020). These authors propose that spatial variability of SOC at surface layer is affected not only by soil texture, parent material and topography, but also by land use, fertilization and other management practices and that the impact of these human activities may be gradually reduced across the soil profile.

Depending on whether the main purpose is to estimate or to predict, the goodness of fit of a model can be evaluated using different criteria. An estimator seeks to know a property of the true state of nature, while a prediction seeks to guess the outcome of a random variable (Kiefer, 1987). If the main purpose is to estimate the spatial pattern properties, it is clear that the strategy of the full model was better than the reduced model one, confirmed by the F-test, based on the weighted residual sum of squares. On the other hand, if the main purpose is to obtain better predictions, according to Oliver and Webster (2014), one of the most telling prediction quality criterion is the

MSDR, the mean ratio between the squared prediction error and the kriging variance. If the model for the variogram is accurate, then the MSE should equal the kriging variance, and so the MSDR should be 1. Considering the MSDR criterion, the reduced model was never the best option. Beyond this, except for the variogram model of NG treatment at 0-5 cm depth (MSDR very close to 1), all models had values greater than 1, indicating that the kriging variance underestimated the squared prediction errors. Regarding the correlation coefficient between observed and predicted (r_op), the values obtained by all models were very low at surface and at 5-15 cm depth were very similar to each other (around 0.5).

5.6.4. Spatial variability of P

The spatial distribution of phosphorus was also affected by the treatments, even to a greater extent than what was observed for SOC, as can be seen in figure 2, where the variograms showed very different behaviors among treatments. These greater differences between treatments were to be expected, given that two of them included consecutive broadcast fertilizations, which supports our hypothesis that P-fertilization for several consecutive years changed the spatial correlation of soil phosphorus content.

At both layers, the semivariogram parameter estimates differed greatly between treatments. Mostly at 0-5 cm depth, large differences in sill values (variance) had already been observed for when comparing treatment effects on phosphorus content (objective 1). The proportion of spatial structure was also affected, due to these large differences of sill values and also of nugget values, being the OP60 the one with less value of PSS, indicating a moderate spatial dependence. Differences were also observed for ranges values, being 176, 44 and 261 m for NG, OP30 and OP60, respectively. In terms of prediction ability, for both depths, the reduced model was always the worst option in terms of MSDR and r_op, reaffirming the results obtained with the F test. For the 0-5 cm soil layer, the best predictions were obtained for the NG treatment, with MSDR values very close to 1 and a correlation coefficient of

0.71 between observed and predicted values. For the deeper soil layer, the best predictions were obtained for the NG and OP30 treatments.

The effect of P-fertilizations on P spatial variability was previously reported by other works. Sauer and Meek (2003), studying the spatial variation of plant-available P in two pastures with different poultry litter management (limited and intense poultry litter application), found that for intense applications (with a significant P-input) the spatial correlation of soil P was significantly reduced. These findings coincide with those of Nze Memiaghe et al. (2020), who found that repeated applications of organic fertilizers with high P content can impact the long-term accumulation of soil P, decreasing P spatial dependence in permanent grasslands. Our results were not as clear as these previous works, but we did observe less spatial dependence in plots with the higher amount of phosphorus added (OP60) and, coinciding with previous reports, we did find greater variability in soil phosphorus content as P added to the soil increased (Ciarlo et al., 2020, Corazza et al., 2003). However, Nze Memiaghe et al. (2020) reported a decreasing in soil P variability as the amount of P-fertilizations increased. These differences could be possibly due to the different fertilization procedures that both works employed.

In any case, in order to have a more accurate estimate of the spatial distribution of phosphorus in our conditions, it would be advisable to have more sampling points in each treatment and to have, at least for a subset of the sampling points, smaller distances between them to better understand the spatial correlation at short distances (in our work, the minimum distances were approximately 25 m). Banded application of phosphorus fertilizer with seeding machines may have resulted in a strong spatial dependence in the range of a few meters, at a shorter distance than that used between soil samples and, therefore, it could have increased the so-called nugget variation, absorbed by the sampling distance used (Ciarlo et al., 2020). Understanding spatial variability of P can help improve crop habitats and fertilizer utilization and reduce agricultural pollution (Li et al., 2020).

5.6.5. Conclusion remarks

The improvements in precision of treatment comparisons of spatial models in the presence of strong spatial correlation among experimental units have been widely reported. However, if treatments that affect the structure of the spatial correlation of soil properties, such as organic carbon or phosphorus content, are applied, these improvements may no longer be achieved. In addition, since the data are at the subsampling level, plots must be added to the model as a random effect. Using the *nlme* package for mixed model with plots specified as independent subjects, the spatial correlation is only estimated within each plot, so maybe although correlation is pooled across plots, the weakness is that there is no information about squared difference for large distances.

The difference in SOC content between both soil layers was more evident for the overseeded and P-fertilized treatments than for NG. It would be very interesting to extend this study to deeper soil layers, where the carbon stock also plays an important role.

Fertilization with phosphorus for nine consecutive years changed the spatial distribution of SOC and P soil content. Furthermore, it seems that this periodic P-input increased the variance of soil phosphorus among observations of OP treatments, mostly at the surface layer. Under this situation, the implementation of site-specific fertilizations may be challenging. To better understand how these technologies impact SOC and P content, more information is needed. It would be advisable to evaluate the spatial variability of SOC and P in each treatment with greater sampling intensity, in order to have a more precise estimate of the semivariances in each situation that allows defining the relevance of using site-specific fertilization strategies. Also, an important aspect that would have been relevant to evaluate is the temporal process of SOC and P content over the years since its installation, but that information was not available. Beyond that, this work contributes to the understanding of the impacts of these frequently applied

technologies on the content and spatial distribution of SOC and P in the soil. This understanding is key when making decisions about how to increase pasture productivity without affecting the sustainability of the system or negatively impacting the ecosystem services that permanent pasture can provide.

5.7. REFERENCES

Arnold TW. 2010. Uninformative Parameters and Model Selection Using Akaike's Information Criterion. Journal of Wildlife Management. *74*(6): 1175–1178. https://doi.org/10.2193/2009-367

Azevedo RP, Salcedo IH, Lima PA, da Silva FragaV, Lana RMQ. 2018. Mobility of phosphorus from organic and inorganic source materials in a sandy soil. International Journal of Recycling of Organic Waste in Agriculture. *7*(2): 153–163. https://doi.org/10.1007/s40093-018-0201-2

Bengtsson J, Bullock JM, Egoh B, Everson C, Everson T, O'Connor T, O'Farrell PJ, Smith HG, Lindborg R. 2019. Grasslands—more important for ecosystem services than you might think. Ecosphere. *10*(2): 1-20. https://doi.org/10.1002/ecs2.2582

Bondaruk V, Lezama F, del Pino A, Piñeiro G. 2020. Overseeding legumes in natural grasslands: Impacts on root biomass and soil organic matter of commercial farms. Science of the Total Environment, *743*: 1-9. https://doi.org/10.1016/j.scitotenv.2020.140771

Borges A, González-Reymundez A, Ernst O, Cadenazzi M, Terra J, Gutiérrez L. 2019. Can spatial modeling substitute for experimental design in agricultural experiments? Crop Science. *59*(1): 44-53. https://doi.org/10.2135/cropsci2018.03.0177

Brownie C, Bowman DT, Burton JW. 1993. Estimating Spatial Variation in Analysis of Data from Yield Trials: A Comparison of Methods. Agronomy Journal. *85*: 1244–1253.

Burke IC, Lauenroth WK, Riggle R, Brannen P, Madigan B, Beard S. 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. https://doi.org/10.1007/s100219900091

Casler MD, Undersander DJ. 2000. Forage yield precision, experimental design, and cultivar mean separation for alfalfa cultivar trials. Agronomy Journal. *92*(6): 1064–1071. https://doi.org/10.2134/agronj2000.9261064x

Ciarlo EA, Muschietti MDP, Peralta N, Comparín M, Gregorini F, Cipriotti PA, Giuffre L. 2020. Spatial variability of soil properties: Effect of land use and type. Ciencia Del Suelo. *38*(2): 249–261.

Corazza EJ, Brossard M, MuraokaI T, Coelho Filho MA. 2003. Spatial variability of soil phosphorus of a low productivity Brachiaria brizantha pasture. Scientia Agricola. *60*(3): 559–564. https://doi.org/10.1590/s0103-90162003000300022

Dingkuhn M, Luquet D, Fabre D, Muller B, Yin X, Paul MJ. 2020. The case for improving crop carbon sink strength or plasticity for a CO2-rich future. Current Opinion in Plant Biology. *56*: 259–272. https://doi.org/10.1016/j.pbi.2020.05.012

Formoso F. 2011. Producción de semillas de especies forrajeras. INIA Serie Técnica N° 190. 231p.

González-Barrios P, Borges A, Terra J, Bidegain MP, Gutiérrez L. 2020. Spatio-temporal modeling and competition dynamics in forest tillage experiments on early growth of eucalyptus grandis L. Forest Science. *66*(5): 526-536. https://doi.org/10.1093/forsci/fxaa007

Grondona MO, Cressie N. 1991. Using Spatial Considerations in the Analysis of Experiments. Technometrics. *33*(4): 381–392.

Horowitz M. 1996. Bermudagrass (Cynodon dactylon): A history of the weed and its control in Israel. Phytoparasitica. *24*(4): 305–320. https://doi.org/10.1007/BF02981413

Husein HH, Mousa M, Sahwan W, Bäumler R, Lucke B. 2019. Spatial Distribution of Soil Organic Matter and Soil Organic Carbon Stocks in Semi-Arid Area of Northeastern Syria. Natural Resources. *10*(12): 415–432. https://doi.org/10.4236/nr.2019.1012028

Jaurena M, Durante M, Devincenzi T, Savian JV, Bendersky D, Moojen FG, Pereira M, Soca P, Quadros FLF, Pizzio R, Nabinger C, Carvalho PCF, Lattanzi FA. 2021. Native Grasslands at the Core: A New Paradigm of Intensification for the Campos of Southern South America to Increase Economic and Environmental Sustainability. Frontiers in Sustainable Food Systems, *5*: 1-15. https://doi.org/10.3389/fsufs.2021.547834

Jaurena M, Lezama F, Salvo L, Cardozo G, Ayala W, Terra J, Nabinger C. 2016. The Dilemma of Improving Native Grasslands by Overseeding Legumes: Production Intensification or Diversity Conservation. Rangeland Ecology and Management, *69*(1): 35–42. https://doi.org/10.1016/j.rama.2015.10.006

Kiefer JC. 1987. Classifications of Statistical Problems. En: Lorden G. (eds) Introduction to Statistical Inference. New York: Springer Texts in Statistics. Springer. 22-30. https://doi.org/10.1007/978-1-4613-9578-2_3

Koch E, Rigney J. 1951. A method of estimating optimum plot size from experimental data. Agronomy Journal. *43*: 17–21. https://doi.org/10.2134/agronj1951.00021962004300010005x

Kravchenko AN, Robertson GR, Snap SS, Smucker AJM. 2006. Using Information about Spatial Variability to Improve Estimates of Total Soil Carbon. Agronomy Journal. *98*(3): 823–829. https://doi.org/10.2134/agronj2005.0305

Lenth R. 2020. emmeans: Estimated Marginal Means, aka Least-Squares Means (R package version 1.4.5.). https://cran.r-project.org/package=emmeans

Li Z, Zhang R, Liu C, Zhang R, Chen F, Liu Y. 2020. Phosphorus spatial distribution and pollution risk assessment in agricultural soil around the Danjiangkou reservoir, China. Science of the Total Environment. *699*: 1-9. https://doi.org/10.1016/j.scitotenv.2019.134417

López MV, Arrúe JL. 1995. Efficiency of an Incomplete Block Design Based on Geostatistics for Tillage Experiments. Soil Science Society of America Journal. *59*(4): 1104–1111. https://doi.org/10.2136/sssaj1995.03615995005900040023x

Mallarino A, Bermudez M, Wittry D, Hinz P. 2000. Alternative Data Managements and Intepretations. En: International Conference on Site-Specific Management for Agricultural Systems (5th, 2000, Madison, Wisconsin). Proceedings. Robert

PC et al (ed) . CD-ROM. July 16- 19. Bloomington, MN. ASA, CSSA, SSSA, Madison, WI

Marriott CA, Hudson G, Hamilton D, Neilson R, Boag B, Handley LL, Wishart J, Scrimgeour CM, Robinson D. 1997. Spatial variability of soil total C and N and their stable isotopes in an upland Scottish grassland. Plant and Soil, *196*(1): 151–162. https://doi.org/10.1023/A:1004288610550

Murphy J, Riley J. 1962. A modified single solution method for the determination of phosphate in natural waters. Analytica Chimica Acta. *27*: 31–36.

National Cooperative Soil Survey. 2014. National Cooperative Soil Survey Soil Characterization Database [En línea]. Disponible en: http://ncsslabdatamart.sc.egov.usda.gov. Último acceso: 23 de octubre de 2021.

Nze Memiaghe JD, Cambouris AN, Ziadi N, Karam A, Perron I. 2020. Spatial Variability of Soil Phosphorus Indices under Two Contrasting Grassland Fields in Eastern Canada. Agronomy, *11*(1): 1-18. https://doi.org/10.3390/agronomy11010024

Oliver MA, Webster R. 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena. *113*: 56–69. https://doi.org/10.1016/j.catena.2013.09.006

Pinheiro J, Bates D, DebRoy S, Sarkar D. 2020. nlme: Linear and Nonlinear Mixed Effects Models (R package version 3.1-144). https://cran.r-project.org/package=nlme

Poeplau C, Zopf D, Greiner B, Geerts R, Korvaar H, Thumm U, Don A, Heidkamp A, Flessa H. 2018. Why does mineral fertilization increase soil carbon stocks in temperate grasslands? Agriculture, Ecosystems and Environment. *265*: 144–155. https://doi.org/10.1016/j.agee.2018.06.003

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.

Reeves DW. 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. Soil and Tillage Research. *43*(1–2): 131–167. https://doi.org/10.1016/S0167-1987(97)00038-X

Rodríguez AM, Jacobo EJ, Scardaoni P, Deregibus VA. 2007. Effect of phosphate fertilization on flooding Pampa grasslands (Argentina). Rangeland Ecology and Management. *60*(5): 471–478. https://doi.org/10.2111/1551-5028(2007)60[471:EOPFOF]2.0.CO;2

Sauer TJ, Meek DW. 2003. Spatial Variation of Plant-Available Phosphorus in Pastures with Contrasting Management. Soil Science Society of America Journal. *67*(3): 826–836. https://doi.org/10.2136/sssaj2003.8260

Schützenmeister A, Jensen U, Piepho HP. 2012. Checking normality and homoscedasticity in the general linear model using diagnostic plots. Communications in Statistics - Simulation and Computation. *41*: 141–154. https://doi.org/10.1080/03610918.2011.582560

Simbahan GC, Dobermann A, Goovaerts P, Ping J, Haddix ML. 2006. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. Geoderma. *132*(3–4): 471–489. https://doi.org/10.1016/j.geoderma.2005.07.001

Slaets JIF, Boeddinghaus RS, Piepho HP. 2021. Linear mixed models and geostatistics for designed experiments in soil science: Two entirely different methods or two sides of the same coin? European Journal of Soil Science. *72*(1): 47–68. https://doi.org/10.1111/ejss.12976

Terra JA, Shaw JN, Reeves DW, Raper RL, van Santen E, Schwab EB, Mask PL. 2006. Soil Management and Landscape Variability Affects Field-Scale Cotton Productivity. Soil Science Society of America Journal. *70*(1): 98-107. https://doi.org/10.2136/sssaj2005.0179

Terra JA, Shaw JN, Reeves DW, Raper RL, van Santen E, Mask PL. 2004. Soil Carbon Relationships With Terrain Attributes, Electrical Conductivity, and a Soil Survey in a Coastal Plain Landscape. Soil Science. *169*(12): 819–831. https://doi.org/10.1097/01.ss.0000149368.71009.6f

Thompson GR. 1995. The direct determination of phosphorus in citric acid soil extracts by colorimetry and direct-current plasma emission spectroscopy. South African Journal of Plant and Soil. *12*(4): 152–157. https://doi.org/10.1080/02571862.1995.10634356

Walkley A, Black IA. 1934. An examination of the Degtjareff Method for

Determining Soil Organic Matter, and a proposed Modification of the Chromic Acid Titration Method. Soil Science. *37*(1): 29–38.

Wang Y, Zhang X, Huang C. 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*(1–2): 141–149. https://doi.org/10.1016/j.geoderma.2009.01.021

Zhang Y, Zhen Q, Li P, Cui Y, Xin J, Yuan Y, Wu Z, Zhang X. 2020. Storage of soil organic carbon and its spatial variability in an agro-pastoral ecotone of northern China. Sustainability (Switzerland). *12*(6): 1–13. https://doi.org/10.3390/su12062259

5.8.SUPPLEMENTAL MATERIAL

Table S1. AIC, plot variance (σ_p^2), residual variance (σ_ε^2) and standard error (SE) values for each model fitted for soil organic carbon content at depth 0 to 5 cm

Model	**AIC**	$\boldsymbol{\sigma_p^2}$	$\boldsymbol{\sigma_\varepsilon^{2\dagger}}$	**SE**†
NULL	238.23	0.0415	0.1901	0.108
AR1	240.23	0.0415	0.1901	0.108
SPH	239.22	0.0391	0.1924	0.108
EXP	238.98	0.0338	0.2006	0.109
GAU	239.34	0.0385	0.1930	0.108
NULL_H	236.87	0.0393	0.2576	0.112
AR1_H	238.27	0.0393	0.2576	0.112
SPH_H	237.73	0.0380	0.2545	0.112
EXP_H	238.23	0.0386	0.2565	0.112
GAU_H	237.96	0.0379	0.2549	0.112

†For models with heterogeneous variances ('_H') values correspond to OP30 treatment

Table S2. AIC, plot variance (σ_p^2), residual variance (σ_ε^2) and standard error (SE) values for each model fitted for soil organic carbon content at depth 5 to 15 cm. Best model is underlined

Model	**AIC**	$\boldsymbol{\sigma_p^2}$	$\boldsymbol{\sigma_\varepsilon^{2\dagger}}$	**SE**†
NULL	19.43	0.0133	0.0471	0.0593
AR1	21.43	0.0133	0.0471	0.0593
SPH	7.99	0.0069	0.0527	0.0552
EXP	<u>0.21</u>	0.0000	0.0614	0.0596
GAU	8.61	0.0072	0.0517	0.0542
NULL_H	23.23	0.0133	0.0474	0.0594
AR1_H	25.23	0.0133	0.0474	0.0594
SPH_H	11.97	0.0068	0.0514	0.0538
EXP_H	4.04	0.0000	0.0575	0.0563
GAU_H	12.60	0.0072	0.0513	0.0533

†For models with heterogeneous variances ('_H') values correspond to OP30 treatment

Table S3. AIC, plot variance (σ_p^2), residual variance (σ_ε^2) and standard error (SE) values for each model fitted for log of Phosphorus content at depth 0 to 5 cm (log($P_{0\text{-}5}$))

Model	**AIC**	$\boldsymbol{\sigma_p^2}$	$\boldsymbol{\sigma_\varepsilon^{2\dagger}}$	**SE**†
NULL	69.07	0.0155	0.0652	0.0660
AR1	71.07	0.0155	0.0652	0.0660
SPH	68.41	0.0141	0.0671	0.0662
EXP	69.23	0.0134	0.0682	0.0668
GAU	68.29	0.0139	0.0678	0.0665
NULL_H	71.03	0.0155	0.0730	0.0671
AR1_H	73.03	0.0155	0.0730	0.0671
SPH_H	70.57	0.0143	0.0742	0.0678
EXP_H	71.15	0.0135	0.0764	0.0687
GAU_H	70.52	0.0142	0.0746	0.0680

†For models with heterogeneous variances ('_H') values correspond to OP30 treatment

Table S4. AIC, plot variance (σ_p^2), residual variance (σ_ε^2) and standard error (SE) values for each model fitted for Phosphorus content at depth 5 to 15 cm ($P_{5\text{-}15}$). Best model is underlined

Model	**AIC**	$\boldsymbol{\sigma_p^2}$	$\boldsymbol{\sigma_\varepsilon^{2\dagger}}$	**SE**†
NULL	329.04	0.3219	0.3323	0.266
AR1	331.04	0.3219	0.3323	0.266
SPH	330.05	0.3199	0.3351	0.266
EXP	330.82	0.3209	0.3340	0.266
GAU	330.16	0.3201	0.3351	0.266
NULL_H	<u>310.94</u>	0.3170	0.5923	0.274
AR1_H	312.94	0.3170	0.5923	0.274
SPH_H	312.62	0.3164	0.5903	0.273
EXP_H	312.94	0.3169	0.5922	0.274
GAU_H	312.86	0.3165	0.5909	0.274

†For models with heterogeneous variances ('_H') values correspond to OP30 treatment

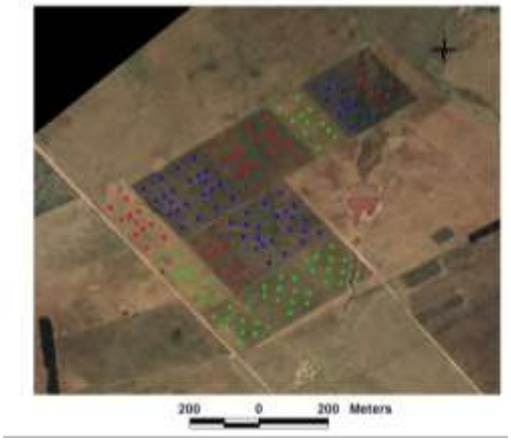


Figure S1. Field trial and sampling points (dots) for NG (green), OP30 (red) and OP60 (blue). NG: natural grassland; OP30: overseeded and P-fertilized with 30 units of P_2O_5 ha^{-1} $year^{-1}$; OP60: overseeded and P-fertilized with 60 units of P_2O_5 ha^{-1} $year^{-1}$

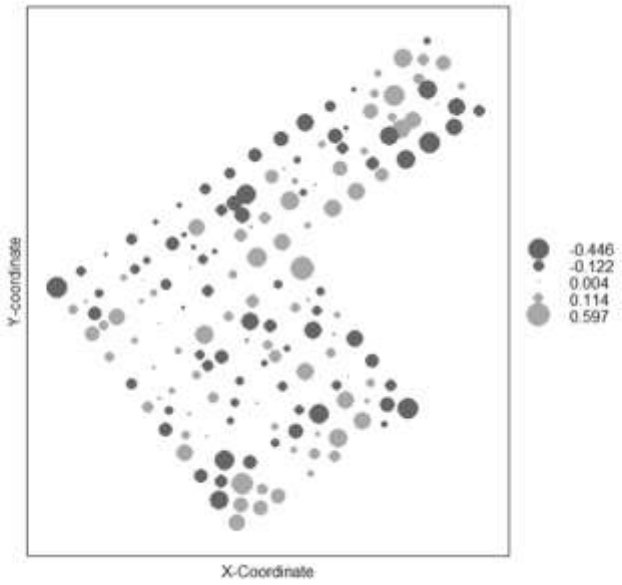


Figure S2. Spatial distribution of residuals of $SOC_{5\text{-}15}$ for NULL model

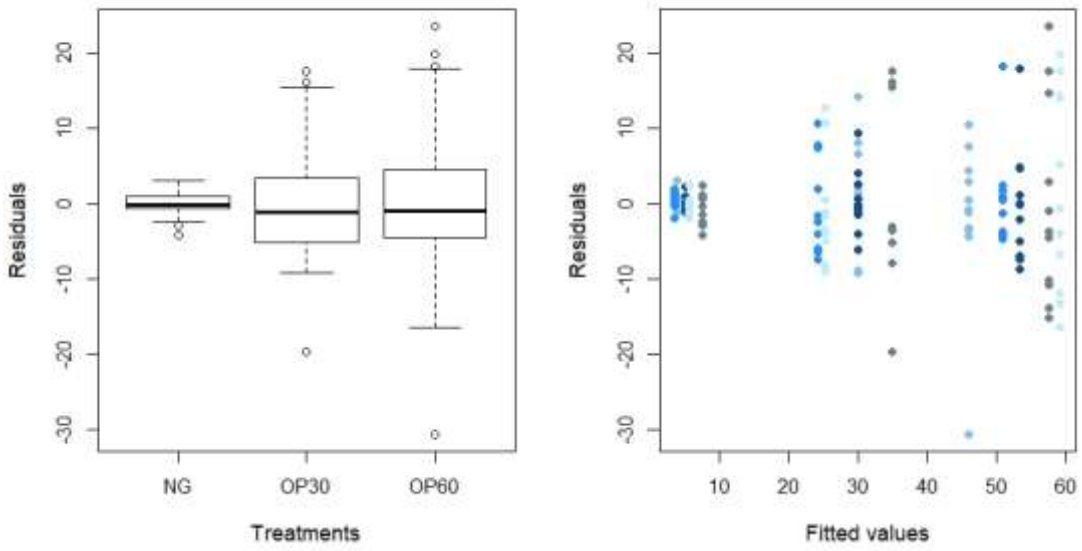


Figure S3. Boxplot for residuals of each treatment (left) and residuals versus fitted values (right) of the NULL model for $P_{0\text{-}5}$ (different colors indicate different blocks). NG: natural grassland; OP30: overseeded and P-fertilized with 30 units of P_2O_5 ha^{-1} $year^{-1}$; OP60: overseeded and P-fertilized with 60 units of P_2O_5 ha^{-1} $year^{-1}$

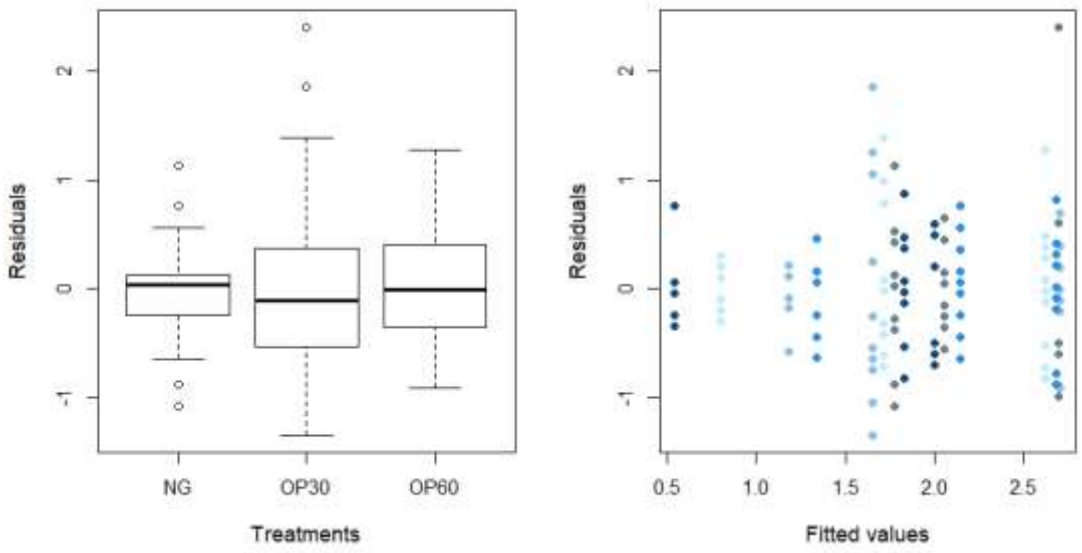


Figure S4. Boxplot of residuals for each treatment (left) and residuals versus fitted values (right) of the NULL model for $P_{5\text{-}15}$ (different colors indicate different blocks). NG: natural grassland; OP30: overseeded and P-fertilized with 30 units of P_2O_5 ha^{-1} $year^{-1}$; OP60: overseeded and P-fertilized with 60 units of P_2O_5 ha^{-1} $year^{-1}$

6. DISCUSIÓN GENERAL

El propósito de esta investigación fue evaluar la eficiencia de diseños experimentales con diferente grado de complejidad y el ajuste de modelos espaciales, temporales y espacio-temporales en experimentos con diferentes cultivos, con el objetivo de mejorar la precisión y exactitud en las estimaciones de las medias de tratamiento, tratando de evitar sesgos en las estimaciones de varianza, con un control adecuado de la tasa empírica de error de tipo 1.

Los resultados más relevantes de la tesis muestran que cuando la heterogeneidad del terreno experimental es alta y el tamaño del experimento es de mediano a grande (mayor o igual a 50 tratamientos), la elección del diseño experimental se vuelve esencial para obtener estimaciones precisas y de mayor exactitud. Una vez elegido el diseño experimental adecuado, la modelización espacial de la correlación entre unidades experimentales puede mejorar aún más el desempeño del diseño elegido.

Estas mejoras en precisión y exactitud en las estimaciones pueden estar acompañadas de errores estándar insesgados y un correcto control de la tasa de error empírica de tipo 1 si el modelo espacial para estimar la correlación espacial de la variable de respuesta es el adecuado. Sumado a esto, cuando el número de repeticiones por tratamiento es alto, se pueden obtener buenos resultados más allá del modelo espacial preferido.

En los casos en que se trabaja con parcelas grandes, donde la variable de interés necesariamente tiene que ser relevada a nivel de submuestras dentro de la parcela, los resultados no son tan contundentes y la ventaja de la modelización espacial no es tan clara. Eso puede depender de la variable de interés, de la escala a la cual se manifiestan esas correlaciones espaciales y del número de submuestras por parcela. En el caso de parcelas en ensayos forestales de laboreo, en especial para la variable altura de árbol, la inclusión de la correlación espacial entre árboles de una misma parcela produjo mejoras significativas en la precisión

de las comparaciones entre medias de tratamientos y logró reducciones del error estándar de la diferencia de medias de hasta 40 %. Sin embargo, en un experimento de largo plazo sobre campo natural bajo pastoreo con diferentes manejos de fertilización, con parcelas de gran tamaño (2 ha) y relativamente pocos puntos de muestreo por parcela (11), no se evidenciaron ventajas de considerar la correlación espacial para comparar los efectos sobre algunas propiedades del suelo (contenido de C y P lábil).

6.1. DISEÑOS EXPERIMENTALES Y MODELOS ESPACIALES

El tamaño del experimento y la heterogeneidad del campo afectaron claramente al desempeño de los diseños experimentales y de las correcciones espaciales. Cuanto mayor es la variabilidad del terreno y el tamaño del experimento, la elección del diseño experimental se vuelve más importante para obtener estimaciones precisas y exactas de las medias de tratamiento. En estas condiciones (alta heterogeneidad y tamaño de experimento grande), el diseño de bloques incompletos Alfa-látice (ALPHA) es el que obtuvo el mejor desempeño. Más allá de la variabilidad del terreno, en experimentos que implican la comparación de un numero alto de tratamientos, es ventajoso usar un diseño más complejo como el ALPHA, reafirmando lo que plantea Casler (2015) sobre la baja eficiencia de los diseños simples para comparar muchos tratamientos. La ventaja de diseños experimentales con mayor control local de la variabilidad espacial como los diseños ALPHA ha sido reportada previamente en varios trabajos (Gonçalves et al., 2010, Müller et al., 2010, Sripathi et al., 2017). En nuestro trabajo, las ventajas de este diseño sobre los otros no sólo se observó en la precisión del experimento, cuando obtuvo menores errores estándar de la diferencia de medias, sino también en la exactitud de las estimaciones, siendo el diseño que presentó mayores valores de correlación entre los efectos verdaderos y estimados de los tratamientos, así como en la capacidad de recuperar los mejores genotipos. Stroup et al. (1994) plantean que si se utiliza un análisis clásico de diseño de bloques completos al azar (RCBD) en presencia de

correlación espacial de los datos, incrementa la estimación del error experimental, lo que puede conducir a estimaciones inexactas de las medias de los tratamientos.

En el mismo sentido, en el capítulo 3 se evaluó la potencia de las pruebas T (P[Rechazar H0|H0 falsa]) de los diferentes diseños experimentales evaluados, siendo ALPHA el diseño con mayor potencia para una misma diferencia de medias, seguido por RCBD y, por último, por el diseño completamente al azar (CRD). En escenarios donde existe una correlación espacial fuerte, Richter et al. (2015) también observaron ganancias en la potencia de la prueba del diseño ALPHA respecto a RCBD.

Cuando la variabilidad espacial es alta, las correcciones espaciales combinadas con experimentos bien diseñados mejoran el desempeño del diseño experimental (estimaciones más precisas, predicciones más exactas), lo que sugiere que se debe tener en cuenta tanto la información de diseño como el modelado espacial (Gilmour et al., 1997, Qiao et al., 2000, Williams et al., 2006). Nuestros resultados presentados en el capítulo 2 y 3 confirman lo planteado por estos autores. En nuestro trabajo, en situaciones donde el número de tratamientos fue moderado ($\leq$ 50) y la variabilidad del terreno no fue alta, se obtuvieron resultados similares utilizando un diseño ALPHA y un modelo de errores independientes que con un diseño RCBD analizado con un modelo de errores correlacionados espacialmente.

Sin embargo, independientemente de la variabilidad espacial, cuando el número de tratamientos es muy alto (200), el uso de modelos espaciales no sustituyó el efecto del diseño experimental en el control de la variabilidad local presente. Por lo tanto, particularmente en ensayos de gran tamaño, la modelización espacial debe considerarse como una estrategia complementaria y no como una alternativa al diseño experimental.

En aquellas situaciones donde el uso modelos espaciales mejoró la eficiencia del experimento y logró mayor precisión en las comparaciones de medias, en

algunos casos esa mejora vino acompañada de problemas de sesgo en los errores estándar de la diferencia de medias y de control de la tasa empírica de error de tipo 1 (ET1). En los cuatro sitios marcados en el ensayo de uniformidad de trigo utilizado en el capítulo 3, algún modelo espacial mejoró el ajuste en relación al modelo nulo de errores independientes (menor AIC) para todos los diseños utilizados. Además de mejor ajuste, también se observaron valores menores del error estándar de diferencias de medias estimado (SED_est) y observado (SED_obs), lo que incrementa la precisión de las comparaciones. Sin embargo, dependiendo del sitio y del modelo de análisis, se constataron problemas de sesgo en el SED_est y una falta de control de ET1 cuando se usaron modelos espaciales. Los modelos espaciales considerados en este capítulo fueron un modelo con función de covarianza gaussiana (GAU) y función de covarianza esférica (SPH). Previo a la simulación de los efectos de los tratamientos, se estudió para cada sitio la variabilidad espacial del ensayo de uniformidad. En dos sitios, el modelo que mejor describió el patrón espacial fue el modelo GAU, mientras que, en los dos sitios restantes, el modelo elegido fue el SPH. Posterior a la simulación, ambos modelos presentaron diferentes desempeños dependiendo del sitio. El modelo GAU presentó menos problemas de sesgo (menos casos y menor ratio) que el modelo SPH y menos casos de falta control de ET1, tendiendo a tasas más liberales (mayor tasa de rechazo de H0 que lo esperado), coincidiendo con trabajos previos (Richter et al., 2015, Richter y Kroschewski, 2012). El modelo SPH en los sitios donde no fue el modelo preferido según AIC presentó serios problemas de sesgo (superó el 20 % de ratio en uno de los sitios) y de control de ET1. De acuerdo al trabajo de Richter et al. (2015), a partir de escenarios simulados con diferentes modelos espaciales, cuando el modelo usado para la simulación era un modelo GAU, el uso de SPH como modelo de análisis condujo a un severo sesgo positivo. Según estos autores, en este escenario, el modelo SPH no es capaz de reproducir la forma sigmoidal de la función de covarianza a distancias cortas, dando lugar a altos sesgos positivos. Si bien nosotros no conocemos cuál es la verdadera estructura de covarianza presente en cada sitio, esto podría ser una explicación para el mal desempeño del modelo

esférico, particularmente en el sitio 2.

Para analizar los datos luego de realizado el ensayo, es clave establecer los criterios de selección de modelos y los métodos de corrección o de aproximación para modelos mixtos. En relación con el criterio de selección de modelos, nosotros usamos el criterio de información de Akaike (AIC), recomendado por Guerin y Stroup (2000), porque tiende a seleccionar modelos más complejos, pero con mejor control de ET1 que otros. De todas formas, también se computó el criterio de información bayesiano (BIC) y se obtuvieron los mismos resultados (datos no mostrados). En relación al método de aproximación, sería interesante para futuros trabajos evaluar el efecto de diferentes métodos para modelos mixtos. En el capítulo 3, el método de aproximación utilizado fue Kenward-Rogers de primer orden (KR1). La elección estuvo basada en trabajos previos que reportan mejores resultados de este método frente a otros cuando se trabaja con modelos espaciales (Richter et al., 2015, Richter y Kroschewski, 2012). Este método tiene dos componentes, uno tiene que ver con la corrección de los grados de libertad y el otro, con una corrección asintótica de la matriz de varianza-covarianza (Kenward y Roger, 2009, Kenward y Roger, 1997). Ambos componentes afectan la ET1 y no se pudo concluir si los problemas encontrados tuvieron que ver con una mala elección del modelo o con algo sistemático que este método pueda tener cuando realiza la aproximación de la matriz de varianza-covarianza de los efectos fijos.

La inclusión o no de los bloques completos en el modelo de análisis para los diseños ALPHA y RCBD tuvo también un efecto importante en el sesgo del SED_est y control de ET1. En particular, el diseño RCBD combinado con el modelo que asume errores independientes (NULL), presentó problemas en todos los sitios analizados con diferente número de tratamientos. Estos problemas fueron levantados por al menos uno de los dos modelos espaciales considerados en casi todas las combinaciones de sitio y número de tratamientos, en los que se obtuvieron estimaciones del SED insesgadas y con control adecuado de ET1. Para el diseño ALPHA analizado con el modelo NULL, los problemas de sesgo y

control de ET1 fueron casi inexistentes. De todas formas, en todos los sitios y para todos los números de tratamientos evaluados, siempre existió también un modelo espacial que no presentara estos problemas. El mejor desempeño de ALPHA puede deberse a que el modelo de análisis todavía tiene ajustes de la variabilidad local por los bloques incompletos, que podrían estar absorbiendo la variabilidad entre los bloques completos.

Por otro lado, el número de repeticiones por tratamiento tuvo un efecto notorio en el control de los problemas de sesgo y ET1. Con un alto número de repeticiones (r = 36), a excepción de un caso, para todos de los modelos de análisis combinados con cualquiera de los tres diseños experimentales evaluados, el sesgo del SED_est fue casi inexistente y la tasa empírica de error de tipo 1 fue bien controlada. Al disminuir el número de repeticiones a 3 o 4, los problemas mencionados aparecieron incluso con el modelo preferido para cada sitio. Hu et al. (2006) se preguntaban si los modelos espaciales pueden emplearse en bases de datos con pocas réplicas por tratamiento, con una estimación aceptable de los errores en presencia de correlación espacial. En nuestro trabajo, encontramos que esto no fue posible en la mayoría de los casos, al menos para estos sitios del ensayo de uniformidad. Para ser útiles en la práctica, los modelos de correlación frecuentemente seleccionados deberían controlar la ET1 para asegurarse de obtener inferencias válidas. Esto no es consistentemente cierto para todos los sitios, con la excepción del sitio 2 donde el modelo seleccionado por AIC (GAU) no presentó problemas para ninguna combinación de número de tratamientos y repeticiones. Con 36 repeticiones se puede obtener una estimación mucho más precisa de las medias de los tratamientos y de los errores estándar de la diferencia, por lo que los problemas de sesgo y falta de control de la tasa de error de tipo 1 se podrían evitar. Hubiera sido deseable probar diferentes números intermedios de repeticiones entre 3 o 4 y 36 para saber, en esas condiciones, cuál sería el número de repeticiones a partir del cual se obtienen estimaciones e inferencias válidas.

En relación con la potencia de los modelos espaciales evaluados en el capítulo 3,

las ventajas de los modelos espaciales sobre el modelo nulo son claras en todos los sitios y para todos los diseños experimentales, coincidiendo con trabajos anteriores (Hu et al., 2006, Hu y Spilke, 2009, Richter et al., 2015). En nuestro trabajo, con diferencias de medias de 1 Mg ha^{-1}de de rendimiento en trigo, estas ventajas se pueden ver más claramente. Sin embargo, hay que tener en cuenta que existieron varios casos en los que no se controló adecuadamente la ET1, por lo que la potencia obtenida en éstos no sería estadísticamente válida. Pero si elegimos uno de los escenarios donde para ningún diseño ni modelo de análisis hubo falta de control de ET1, los modelos espaciales tuvieron mayor potencia que el modelo nulo y estas ganancias en potencia fueron más evidentes en el diseño más simple (CRD).

En resumen, la ganancia de precisión que puede obtenerse con el uso de modelos espaciales en los ensayos de campo no siempre va acompañada de inferencias válidas, con un control adecuado de la ET1, estimaciones insesgadas y una potencia fiable. Esto depende de que se pueda estimar adecuadamente el patrón espacial subyacente, es decir, que se elija la estructura de varianza-covarianza correcta. Al tomar decisiones sobre cómo analizar los datos, hay una serie de pasos clave para minimizar los riesgos que dependen fundamentalmente de la estructura de los datos, el diseño experimental y el número de réplicas por tratamiento. Si se presume que la correlación espacial de la variable de respuesta es fuerte, la elección de un diseño experimental adecuado que dé cuenta de esta heterogeneidad será crucial, así como la elección a través de algún indicador de ajuste del modelo de la estructura de la matriz de varianza-covarianza que describa adecuadamente el patrón espacial. Por último, el número de repeticiones por tratamiento es otro factor importante a considerar cuando se desea utilizar modelos espaciales, dado que permite una estimación más precisa e insesgada de los errores estándar de las diferencias, con un control adecuado de la tasa de error de tipo 1.

6.2. MODELOS ESPACIALES PARA ESTIMAR CORRELACIONES INTRAPARCELARIAS

En la experimentación agronómica, cada vez es más frecuente tener múltiples observaciones por unidad experimental, sobre todo cuando el tamaño de parcela debe ser grande. En esos casos, sería más adecuado utilizar un modelo mixto que permita modelar la correlación entre observaciones dentro de una misma parcela o unidad experimental que asumir independencia de las observaciones (Slaets et al., 2021).

En esta tesis se trabajó con dos experimentos diferentes donde la variable de interés se registró en submuestras adentro de la unidad experimental. En el primer caso, era un ensayo forestal donde se comparaban los efectos de tratamientos de laboreo sobre el crecimiento inicial de los árboles. Las parcelas tenían una dimensión entre 315 y 394 m^2, dependiendo del sitio, con 36 y 45 árboles por parcela, respectivamente. En cada árbol se registraron varias variables de crecimiento en diferentes momentos durante 30 meses (altura, DAP, volumen, tiempo en alcanzar 2 metros de altura), que se analizaron según varios modelos mixtos. El segundo caso se trataba de un experimento donde se evaluaban los efectos de tratamientos de mejora del campo natural a través del agregado de leguminosas y fertilización con fósforo sobre el contenido de carbono orgánico del suelo (SOC) y el contenido de fósforo lábil (P). Las parcelas tenían una superficie de 2 ha y en 11 puntos de muestreo dentro de cada parcela se relevaron las dos variables de respuesta a dos profundidades (0 a 5 y 5 a 15 cm). Los efectos de los tratamientos se compararon mediante diferentes modelos mixtos que incluían o no la correlación espacial dentro de la parcela.

En el ensayo forestal se evaluaron dos estrategias de análisis. La primera estrategia consistió en dos etapas: en la primera, se utilizaron modelos de regresión para modelar la altura de árbol en función de los meses de vida, donde se obtuvo de cada curva el tiempo en meses en que se alcanzan los 2 m de altura (T_{2m}) y, en una segunda etapa, se compararon los tratamientos para la variable T_{2m} utilizando modelos mixtos que incluían o no la correlación espacial entre árboles. La segunda

estrategia se utilizó para la variable altura, volumen de madera individual y volumen de madera por hectárea. En este caso, se utilizaron varios modelos mixtos que podían incluir la correlación espacial y/o temporal entre árboles. En ambos casos, los modelos se compararon mediante el criterio de información bayesiano (BIC) y, en una segunda etapa, por el error estándar de la diferencia de medias (SED). En ambas estrategias, los modelos que incluían la correlación espacial entre árboles tuvieron mejor ajuste (menor BIC) que el modelo de errores independientes, salvo para el caso de volumen individual de madera, donde el mejor modelo solamente incluía la correlación temporal. Las ventajas más grandes en términos de precisión se observaron para la variable altura usando un modelo espacio-temporal, donde se observaron reducciones del SED entre 35 y 40 %, dependiendo del sitio. Existen trabajos anteriores donde se reporta la variable altura como indicadora de sitio, ya que presenta, en general, una correlación espacial fuerte, por lo que los modelos que incluyen esta correlación mejoran la precisión de las estimaciones (González Barrios et al., 2015) e incluso pueden permitir reducir la intensidad del muestreo (Raimundo et al., 2017).

Sin embargo, en el experimento de mejoramiento del campo natural, las ventajas de incluir la correlación espacial dentro de la parcela no fueron tan claras y solamente se eligió un modelo espacial en un solo caso. Esto no significa que no exista correlación espacial entre las observaciones dentro de la parcela para ambas variables de estudio en las dos profundidades muestreadas. De hecho, cuando se estudió la variabilidad espacial por tratamiento para ambas variables y profundidades, en la mayoría de los casos, se constató la existencia de correlación espacial. Esta aparente discrepancia entre ambos resultados pudo deberse, principalmente, a dos factores. En primer lugar, se observaron diferentes patrones espaciales según tratamiento confirmado a través de la prueba F para comparación de modelos, por lo que es razonable pensar que un modelo que proponga una estructura de correlación común a todos los tratamientos no va a mejorar el ajuste del modelo. En segundo lugar, el número de puntos de muestreo por parcela en nuestro estudio puede no haber sido suficiente como para obtener una estimación precisa de la estructura espacial. Es importante

tener en cuenta que cuando se especifica un efecto aleatorio para las parcelas, en el paquete nlme, las parcelas se consideran sujetos y la correlación espacial de dos observaciones se modelará sólo dentro de las parcelas, pero no entre ellas (Slaets et al., 2021). Aunque el patrón espacial se estima agrupando la información de todas las parcelas, quizás el punto débil es que no tenemos información sobre la semivarianza para distancias grandes, ya que esas distancias no se alcanzan dentro de las parcelas.
En el único caso donde el modelo elegido fue un modelo espacial (SOC a 5-15 cm de profundidad), al analizar el efecto de los modelos sobre los componentes de varianza, se observó que los modelos espaciales y, especialmente, el modelo elegido reducen de forma importante la varianza entre parcelas y aumentan la varianza residual. Sin embargo, esta disminución de la varianza entre parcelas no se tradujo en una reducción del error estándar de una media, dado que todos los modelos presentaron valores muy similares. Parecería que los modelos espaciales alteran la varianza de error estimada (dentro de la parcela), lo que induce cambios en la estimación de la varianza de la parcela. Ambos componentes de la varianza afectan al error estándar de la media del tratamiento, aunque el peso de cada componente no está claro.

6.3. MODELOS TEMPORALES Y ESPACIO-TEMPORALES

En este trabajo, solo se evaluaron las correlaciones temporales entre individuos para el ensayo forestal descripto anteriormente (capítulo 4), dado que fue en el único caso en que las variables de respuesta se registraron repetidamente a lo largo de un período de tiempo. Estas correlaciones se incluyeron en modelos donde sólo se consideraba la correlación temporal entre árboles y en modelos donde se consideraba conjuntamente la correlación espacial y temporal entre individuos.

Más allá del sitio o de la variable considerada, siempre se constató la existencia de una correlación temporal entre observaciones de un mismo árbol y varianzas heterogéneas en el tiempo, aunque las estructuras de covarianza no fueran las mismas, dependiendo de la variable de interés. Para algunas variables, los mejores

resultados se obtuvieron con modelos que también incluían la variabilidad espacial, como altura o volumen de madera por hectárea. Por ejemplo, para la variable altura, los modelos espacio-temporales fueron los de mejor desempeño. Si pensamos en escalas pequeñas, como la que puede existir en una parcela forestal, se pueden dar, simultáneamente, correlaciones positivas (por condiciones ambientales similares) y negativas (por competencia entre árboles) que influyen en el crecimiento de los árboles (Fox et al., 2007) y el peso de cada una varía con el tiempo. Posiblemente, a etapas iniciales de crecimiento la competencia tenga poco o nulo efecto y los que predominen sean los efectos espaciales. El desafío está en identificar una estrategia de análisis óptima que permita la integración de la variabilidad espacial y la información de la competencia entre árboles combinada con la evolución temporal. En nuestro trabajo, ambos procesos se incluyeron de forma independiente; sería interesante evaluar modelos que incluyan la interacción del espacio y el tiempo.

6.4.CONCLUSIONES Y CONSIDERACIONES FINALES

Se evaluó la eficiencia de diseños experimentales con diferente grado de complejidad y el ajuste de modelos espaciales, temporales y espacio-temporales en experimentos con diferentes cultivos, buscando una mejora en la precisión y exactitud en las estimaciones de las medias de tratamiento, evitando sesgos en las estimaciones de varianza, con un control adecuado de la tasa empírica de error de tipo 1.

En condiciones de alta heterogeneidad espacial y con un gran número de tratamientos, la modelización espacial sin un adecuado control local no supera en eficiencia al control local que realizan los diseños experimentales. Una vez elegido el diseño experimental, la modelación espacial puede mejorar su rendimiento. Incluso en condiciones de baja variabilidad, si el número de tratamientos es grande, la decisión más importante pasa por elegir un diseño experimental que permita obtener estimaciones precisas y con mayor exactitud.

Para tener más información a la hora de elegir un diseño experimental, se podría utilizar la estimación de los patrones espaciales que se pueden obtener a través de la tecnología de la agricultura de precisión y así poder diseñar mejor los experimentos y encontrar una descripción espacial adecuada para cada uno (Cooper et al., 2014). En esta tesis, no evaluamos el uso de información espacial *a priori* en la optimización de los diseños experimentales clásicos, que se han denominado diseños espaciales (Williams et al., 2006). Sin embargo, uno de los aspectos más difíciles de este enfoque es demostrar la validez de las restricciones en la aleatorización y la presencia de un sesgo de varianza de error en estos diseños (Williams y Piepho, 2018). En cualquier caso, incluso si la variación espacial del suelo está bien caracterizada a través de herramientas de agricultura de precisión y los experimentos se diseñan en consecuencia, hay variación que no se puede predecir de antemano (Cooper et al., 2014), lo que hace aún más difícil producir buenas aleatorizaciones.

En relación con los modelos espaciales evaluados, en la mayoría de los casos, alguno mostró un mejor desempeño en relación al modelo de errores independientes. El modelo espacial elegido varió con el sitio, la variable de interés, la escala y el cultivo. Un aspecto a resaltar es que en todos los capítulos de esta tesis se evaluaron modelos isotrópicos, es decir, modelos que consideran que la correlación espacial solamente depende de la distancia y no de la dirección. La heterogeneidad y la dependencia espacial pueden estar más acentuadas en una dirección por factores edáficos, topográficos, cercanía a cursos de agua, etc., y, por lo tanto, los modelos espaciales anisotrópicos podrían describir de forma más precisa el patrón espacial de la variable de interés.

Por estas razones, lograr estimar adecuadamente ese patrón (elegir la estructura de varianza-covarianza correcta) es un aspecto clave para sacar conclusiones fiables, sin sesgo ni problemas de control de la tasa de error. Como sostiene Cox (2009), "la realidad de cualquier ganancia aparente de precisión depende de la adecuación del modelo asumido". A la hora de tomar decisiones sobre cómo analizar los datos, hay una serie de pasos clave para minimizar los riesgos mencionados que dependen en

gran parte de la estructura de los datos, el diseño experimental y el número de repeticiones por tratamiento. En este sentido, poder contar con una buena caracterización a priori de los patrones espaciales del fenómeno bajo estudio, por ejemplo, por medio de monitores de rendimiento, es una herramienta muy útil para poder seleccionar la estructura de covarianza adecuada y, de esa forma, llegar a conclusiones válidas y precisas. Además, permiten diseñar, por ejemplo, manejos específicos para cada sitio, como podría ser el diseño de la fertilización, sin que ello afecte a la sostenibilidad del sistema o repercuta negativamente en los servicios ecosistémicos que pueden proporcionar los cultivos o el campo natural (Jaurena et al., 2016).

Por último, en los experimentos de larga duración, las respuestas de los cultivos bajo estudio están influidas por procesos temporales y espaciales. Esto provoca desbalance de los datos, correlaciones a través del tiempo y el espacio y varianzas heterogéneas a lo largo del tiempo (Brownie et al., 2004), por lo que, en muchos casos, se hace necesario incluir en el análisis la variabilidad espacial y temporal para obtener mejor ajuste y poder conseguir comparaciones más precisas entre tratamientos. Es importante identificar una estrategia de análisis que modele en forma conjunta la variabilidad espacial y la evolución temporal, permitiendo la inclusión de la interacción de ambos procesos.

En este trabajo, nos planteamos como línea general de investigación evaluar la eficiencia de diferentes diseños experimentales y las ventajas y desventajas de incluir en los modelos de análisis la correlación espacial y/o temporal entre observaciones en diferentes sistemas productivos. En los casos en que se trabajó con un ensayo de uniformidad de trigo (capítulos 2 y 3), donde se contaba previamente con una buena descripción de la variabilidad del terreno, se pudieron exponer ventajas y desventajas de los diseños evaluados y de incluir la correlación espacial de los residuales. Se pudieron identificar algunos elementos que afectan ese balance (tamaño del experimento, heterogeneidad espacial, número de repeticiones por tratamiento, elección del modelo de covarianza adecuado e inclusión o no de los bloques

completos). Sin embargo, quedaron algunas preguntas sin responder, como el efecto de diferentes métodos de aproximación para modelos mixtos según las estructuras de covarianza propuestas o el número de repeticiones necesario para minimizar problemas de sesgo o de control de la tasa de error de tipo 1. En el capítulo 4, analizando variables de crecimiento inicial en ensayos forestales, se pudieron observar las ventajas de incluir las correlaciones espaciales y temporales entre árboles de una misma parcela en evaluaciones repetidas en el tiempo, donde se contaba con muchos puntos de muestreo y se tenía una buena cobertura de la superficie de la parcela. Hubiera sido interesante evaluar posibles interacciones de los procesos espacial y temporal en esas variables de crecimiento, porque es razonable pensar que las correlaciones espaciales cambian con el tiempo, dado que los fenómenos de competencia cobran importancia a medida que los árboles crecen, lo que afecta el balance entre correlaciones positivas y negativas entre ellos. Por último, en el capítulo 5, se pudo corroborar la hipótesis de que la siembra directa con leguminosas y fertilizaciones anuales con P sobre el campo natural, modifica las correlaciones espaciales del carbono orgánico y el fósforo del suelo. Principalmente por estas razones es que, en la mayoría de los casos, los modelos de análisis que incluyeron una estructura de covarianza espacial común para todos los tratamientos no presentaron mejor ajuste ni mayor precisión que el modelo que asumía independencia de residuales. Sería recomendable, a futuro, contar con mayor información espacial a través de aumentar la intensidad de muestreo, como información temporal realizando evaluaciones repetidas de las variables de interés. Esta información podría ser muy útil para entender cómo ocurren estos procesos y poder diseñar medidas de manejo más eficientes y, sobre todo, más sustentables. En este sentido, la descripción de la variabilidad espacial obtenida en varios ciclos de cultivo a través de la tecnología de agricultura de precisión podría utilizarse para hacer un uso eficiente de los recursos agregados al sistema, para diseñar mejor los experimentos y encontrar una corrección espacial adecuada para cada uno de ellos.

7. BIBLIOGRAFÍA GENERAL

Blangiardo M, Cameletti M, Baio G, Rue H. 2013. Spatial and spatio-temporal models with R-INLA. Spatial and Spatio-temporal Epidemiology, *4*: 33-49. https://doi.org/10.1016/j.sste.2012.12.001.

Bondaruk V, Lezama F, del Pino A, Piñeiro G. 2020. Overseeding legumes in natural grasslands: Impacts on root biomass and soil organic matter of commercial farms. Science of the Total Environment, *743*: 1-9. https://doi.org/10.1016/j.scitotenv.2020.140771

Brownie C, Bowman DT, Burton JW. 1993. Estimating Spatial Variation in Analysis of Data from Yield Trials: A Comparison of Methods. Agronomy Journal. *85*: 1244–1253.

Brownie C, King LD, Dube TJ. 2004. Longitudinal and Spatial Analyses Applied to Corn Yield Data from a Long-Term Rotation Trial. Institute of Statistics Mimeo Series. *2559*: 1–27.

Burke IC, Lauenroth WK, Riggle R, Brannen P, Madigan B, Beard S. 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. https://doi.org/10.1007/s100219900091

Casler MD. 2015. Fundamentals of experimental design: Guidelines for designing successful experiments. Agronomy Journal. *107*(2): 692–705. https://doi.org/10.2134/agronj2013.0114

Casler MD, Undersander DJ. 2000. Forage yield precision, experimental design, and cultivar mean separation for alfalfa cultivar trials. Agronomy Journal. *92*(6): 1064–1071. https://doi.org/10.2134/agronj2000.9261064x

Cochran WG, Cox GM. 1992. Completely Randomized, Randomized Blocks and Latin Square Designs. En: J. Wiley and Sons. (Eds.). Experimental Designs (2nd ed.). New York. Wiley (Wiley Classics Library). 95-147.

Cooper M, Messina CD, Podlich, D, Totir, LR, Baumgarten A, Hausmann NJ, Wright D, Graham G. 2014. Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. Crop and Pasture

Science. *65*(4): 311–336. https://doi.org/10.1071/CP14007

Cox DR. 2009. Randomization in the design of experiments. International Statistical Review. *77*(3): 415–429. https://doi.org/10.1111/j.1751-5823.2009.00084.x

Cullis ABR, Gleeson AC. 1991. Spatial Analysis of Field Experiments-An Extension to Two Dimensions. Biometrics. *47*(4): 1449–1460.

Diggle PJ, Heagerty PJ, Liang KY, Zeger SL. 2002. Analysis of longitudinal data (2nd. ed.). En: Atkinson AC, Carroll RJ, Hand DJ, Wang JL. Oxford: Oxford University Press (Oxford Statistical Inference Series). 379 p.

Fisher RA. 1935. The principles of experimentation, illustrated by a pshico-physical experiment. En: Oliver and Boyd (Eds). The Design of Experiments. New York: Hafner Press, 11-26.

Fonseca F, de Figueiredo T, Martins A. 2011. Survival and early growth of mixed forest stands installed in a Mediterranean Region: Effects of site preparation intensity. Forest Ecology and Management. *262*(10): 1905–1912. https://doi.org/10.1016/j.foreco.2011.01.040

Fox JC, Ades PK, Bi H. 2001. Stochastic structure and individual-tree growth models. Forest Ecology and Management. *154*(1–2): 261–276. https://doi.org/10.1016/S0378-1127(00)00632-0

Fox JC, Bi H, Ades PK. 2007. Spatial dependence and individual-tree growth models. I. Characterising spatial dependence. Forest Ecology and Management. *245*(1–3): 10–19. https://doi.org/10.1016/j.foreco.2007.04.025

Gilmour AR, Cullis BR, Verbyla AP. 1997. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. Journal of Agricultural, Biological, and Environmental Statistics. *2*(3): 269–293.

Gleeson AC, Cullis BR. 1987. Residual Maximum Likelihood (REML) Estimation of a Neighbour Model for Field Experiments. Biometrics. *43*(2): 277–287.

Gonçalves E, St.aubyn A, Martins A. 2010. Experimental designs for evaluation of genetic variability and selection of ancient grapevine varieties: A simulation study. Heredity. *104*(6): 552–562. https://doi.org/10.1038/hdy.2009.153

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. https://doi.org/10.1016/j.foreco.2015.02.031

Gregoire TG, Schabenberger O, Barrett JP. 1995. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent plot measurements. Canadian Journal of Forest Research. *25*: 137–156.

Grondona M, Crossa J, Fox P, Pfeiffer W. 1996. Analysis of variety yield trials using two-dimensional separable ARIMA processes. Biometrics. *52*(2): 763–770. http://www.jstor.org/stable/10.2307/2532916

Grondona MO, Cressie N. 1991. Using Spatial Considerations in the Analysis of Experiments. Technometrics. *33*(4): 381–392.

Guerin L, Stroup WW. 2000. A Simulation Study To Evaluate Proc Mixed Analysis of Repeated Measures Data. En: Conference on Applied Statistics in Agriculture (12th, 2000, Manhattan, Kansas). Proceedings. Kansas State University Libraries. 170–203. https://doi.org/10.4148/2475-7772.1249

Hu X, Spilke J. 2009. Comparison of various spatial models for the analysis of cultivar trials. New Zealand Journal of Agricultural Research. *52*(3): 277–287. https://doi.org/10.1080/00288230909510512

Hu X, Spilke J, Richter C. 2006. The influence of spatial covariances on the type I error and the power for different evaluation models. Biometrical Letters. *43*(1): 19–37. papers3://publication/uuid/CA1A021D-3BE5-4AE9-989A-8047B3A480BD

Jaurena M, Lezama F, Salvo L, Cardozo G, Ayala W, Terra J, Nabinger C. 2016. The Dilemma of Improving Native Grasslands by Overseeding Legumes: Production Intensification or Diversity Conservation. Rangeland Ecology and Management, *69*(1): 35–42. https://doi.org/10.1016/j.rama.2015.10.006

Kenward MG, Roger JH. 1997. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics. *53*(3): 983–997.

Kenward MG, Roger JH. 2009. An improved approximation to the precision of fixed effects from restricted maximum likelihood. Computational Statistics & Data Analysis. *53*(7): 2583–2595. https://doi.org/10.1016/j.csda.2008.12.013

Koch E, Rigney J. 1951. A method of estimating optimum plot size from

experimental data. Agronomy Journal. *43*: 17–21. https://doi.org/10.2134/agronj1951.00021962004300010005x

Kravchenko AN, Robertson GR, Snap SS, Smucker AJM. 2006. Using Information about Spatial Variability to Improve Estimates of Total Soil Carbon. Agronomy Journal. *98*(3): 823–829. https://doi.org/10.2134/agronj2005.0305

Lee J, Laca EA, van Kessel C, Rolston DE, Hopmans JW, Six J. 2009. Tillage Effects on Spatiotemporal Variability of Particulate Organic Matter. Applied and Environmental Soil Science. *2009*: 1–14. https://doi.org/10.1155/2009/219379

Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology. *74*(6): 1659–1673. http://www.esajournals.org/doi/abs/10.2307/1939924

Liu J, Ashton PS. 1999. Simulating effects of landscape contexts and timber harvest on tree species diversity. Ecological Applications. *9*: 186–201. https://doi.org/https://doi.org/10.1890/1051-0761(1999)009[0186:SEOLCA]2.0.CO;2

López MV, Arrúe JL. 1995. Efficiency of an Incomplete Block Design Based on Geostatistics for Tillage Experiments. Soil Science Society of America Journal. *59*(4): 1104–1111. https://doi.org/10.2136/sssaj1995.03615995005900040023x

Mallarino A, Bermudez M, Wittry D, Hinz P. 2000. Alternative Data Managements and Intepretations. En: International Conference on Site-Specific Management for Agricultural Systems (5th, 2000, Madison, Wisconsin). Proceedings. Robert PC et al (ed) . CD-ROM. July 16- 19. Bloomington, MN. ASA, CSSA, SSSA, Madison, WI

Marriott CA, Hudson G, Hamilton D, Neilson R, Boag B, Handley LL, Wishart J, Scrimgeour CM, Robinson D. 1997. Spatial variability of soil total C and N and their stable isotopes in an upland Scottish grassland. Plant and Soil, *196*(1): 151–162. https://doi.org/10.1023/A:1004288610550

Matérn B. 1960. Spatial variation, Meddelanden fran Statens Skogsforskningsinstitut, 49, 5. Second ed. (1986), Lecture Notes in Statistics 36, New York: Springer.

Müller BU, Kleinknecht K, Möhring J, Piepho HP. 2010. Comparison of spatial models for sugar beet and barley trials. Crop Science. *50*(3): 794–802.

https://doi.org/10.2135/cropsci2009.03.0153

O'Rourke S, Kelly GE. 2015. Spatio-temporal Modelling of Forest Growth Spanning 50 Years – The Effects of Different Thinning Strategies. Procedia Environmental Sciences. *26*: 101–104. https://doi.org/10.1016/j.proenv.2015.05.008

Patterson HD, Williams ER. 1976. A new class of resolvable incomplete block designs. Biometrika. *1963*(1): 83–92. https://doi.org/https://doi.org/10.1093/biomet/63.1.83

Piepho HP, Williams ER, Michel V. 2015. Beyond Latin squares: A brief tour of row-column designs. Agronomy Journal. *107*(6): 2263–2270. https://doi.org/10.2134/agronj15.0144

Qiao CG, Basford KE, DeLacy IH, Cooper M. 2000. Evaluation of experimental designs and spatial analyses in wheat breeding trials. Theoretical and Applied Genetics. *100*: 9–16.

Raimundo MR, Scolforo HF, de Mello JM, Scolforo JRS, McTague JP, dos Reis AA. 2017. Geostatistics applied to growth estimates in continuous forest inventories. Forest Science. *63*(1): 29–38.

Richter C, Kroschewski B. 2012. Geostatistical models in agricultural field experiments: Investigations based on uniformity trials. Agronomy Journal. *104*(1): 91–105. https://doi.org/10.2134/agronj2011.0100

Richter C, Kroschewski B, Piepho HP, Spilke J. 2015. Treatment comparisons in agricultural field trials accounting for spatial correlation. Journal of Agricultural Science. *153*(7): 1187–1207. https://doi.org/10.1017/S0021859614000823

Saud P, Lynch TB, Anup KC, Guldin JM. 2016. Using quadratic mean diameter and relative spacing index to enhance height-diameter and crown ratio models fitted to longitudinal data. Forestry. *89*(2): 215–229. https://doi.org/10.1093/forestry/cpw004

Skovsgaard JP, Vanclay JK. 2013. Forest site productivity: A review of spatial and temporal variability in natural site conditions. Forestry. *86*(3): 305–315. https://doi.org/10.1093/forestry/cpt010

Slaets JIF, Boeddinghaus RS, Piepho HP. 2021. Linear mixed models and

geostatistics for designed experiments in soil science: Two entirely different methods or two sides of the same coin? European Journal of Soil Science. *72*(1): 47–68. https://doi.org/10.1111/ejss.12976

Sripathi R, Conaghan P, Grogan D, Casler MD. 2017. Field design factors affecting the precision or ryegrass forage yield estimation. Agronomy Journal. *109*: 858–869. https://doi.org/10.2134/ agronj2016.07.0397

Stroup WW, Baenziger PS, Multize DK. 1994. Removing spatial variation from wheat yield trials: A comparison of methods. Crop Science. *34*(1): 62–66. https://doi.org/10.2135/cropsci1994.0011183X003400010011x

Tsoularis A, Wallace J. 2002. Analysis of logistic growth models. Mathematical Biosciences *179*: 21–55.

Wang Y, Zhang X, Huang C. 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*(1–2): 141–149. https://doi.org/10.1016/j.geoderma.2009.01.021

Wikle CK, Zammit-Mangion A, Cressie N. 2019. Spatio-Temporal Statistics with R. En: Chambers JM, Hothorn T, Temple Lang D, Wickham H. Boca Raton, FL: Chapman & Hall/CRC. https://spacetimewithr.org. 380 p.

Williams ER, John JA, Whitaker D. 2006. Construction of resolvable spatial row-column designs. Biometrics. *62*(1): 103–108. https://doi.org/10.1111/j.1541-0420.2005.00393.x

Williams ER, Matheson AC, Harwood CE. 2002. Experimental Design and Analysis for Tree Improvement. En: CSIRO Publishing. Experimental Design and Analysis for Tree Improvement. Melbourne: CSIRO. 214 p.

Williams ER, Piepho HP. 2013. A comparison of spatial designs for field variety trials. Australian and New Zealand Journal of Statistics. *55*(3): 253–258. https://doi.org/10.1111/anzs.12034

Williams ER, Piepho HP. 2018. An Evaluation of Error Variance Bias in Spatial Designs. Journal of Agricultural, Biological, and Environmental Statistics. *23*(1): 83–91. https://doi.org/10.1007/s13253-017-0309-2

Wolfinger RD. 1996. Heterogeneous variance-covariance structures for repeated

measures. Journal of Agricultural, Biological, and Environmental Statistics, *1*(2): 205–230. https://doi.org/10.2307/1400366

Wu T, Dutilleul P. 1999. Validity and efficiency of neighbor analyses in comparison with classical complete and incomplete block analyses of field experiments. Agronomy Journal. *91*(4): 721-731. https://doi.org/10.2134/agronj1999.914721x

Zimmerman DL, Harville DA. 1991. A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments. Biometrics. *47*(1): 223-239. https://doi.org/10.2307/2532508