

**UNIVERSIDAD DE LA REPÚBLICA  
FACULTAD DE AGRONOMÍA**

**ANÁLISIS DE HERRAMIENTAS ESTADÍSTICAS EN  
AGRICULTURA DE PRECISIÓN PARA EL CULTIVO DE ARROZ**

**por**

**Camila María BONILLA CEDREZ**

TESIS presentada como uno de los  
requisitos para obtener el título de  
*Magister* en Ciencias Agrarias opción  
Bioestadística

MONTEVIDEO  
URUGUAY  
mayo 2014

Tesis aprobada por el tribunal integrado por PhD. Ing. Agr. Mónica Cadenazzi, PhD. Ing. Agr. José Terra, y PhD. Richard Plant, el 21 de Abril de 2014. Autora: Ing. Agr. Camila María Bonilla Cedrez. Director PhD. Ing. Agr. Álvaro Roel, Co-director PhD. Ing. Agr. Lucía Gutiérrez.

## **AGRADECIMIENTOS**

Quisiera agradecer a la Facultad de Agronomía por seguir apostando a la educación y formación de sus egresados y darme la posibilidad de seguir avanzando en mi formación profesional y académica.

Al Instituto Nacional de Investigación Agropecuaria por brindarme la posibilidad de realizar mi trabajo de tesis y poner a mi disposición sus instalaciones y su personal.

A mis tutores los PhD. Ing. Agr. Álvaro Roel y Lucía Gutiérrez, por contribuir en mi formación profesional y personal, por guiarme, darme su apoyo y confianza durante todo el proceso.

Al personal del Laboratorio de física de suelo y riego de INIA Treinta y Tres por su disposición y colaboración en el trabajo de campo.

A la Ing. Agr. Cristina Capurro quien me ayudó y acompañó durante largas jornadas de trabajo, haciéndolas más amenas y llevaderas.

Al productor por permitir realizar mi trabajo de tesis durante dos años en su campo, y al Ingeniero Agrónomo responsable de la chacra por ayudar en las organizaciones de los trabajos.

A la Lic. Belky Mesones por su ayuda en la consulta bibliográfica.

A todos los compañeros del DBEC por darme un lugarcito para trabajar y aprender.

Finalmente a mi familia y amigos por el apoyo constante.

## TABLA DE CONTENIDO

	Página
PÁGINA DE APROBACIÓN.....	II
AGRADECIMIENTOS.....	III
RESUMEN.....	VII
SUMMARY.....	VIII
<b>1. <u>INTRODUCCIÓN</u>.....</b>	<b>1</b>
<b>1.1. AGRICULTURA DE PRECISIÓN.....</b>	<b>1</b>
<b>1.2. ANÁLISIS DE DATOS ESPACIALES: GEOESTADÍSTICA.....</b>	<b>2</b>
<b>1.3. ANÁLISIS DE CONGLOMERADOS PARA DELIMITAR         DIFERENTES ZONAS DE RENDIMIENTO.....</b>	<b>10</b>
<b>1.4. ANÁLISIS DE VARIABILIDAD ESPACIAL MULTIVARIADO         EN AGRICULTURA DE PRECISIÓN.....</b>	<b>12</b>
<b>1.5. ÁRBOLES DE DECISIÓN: UN ENFOQUE ALTERNATIVO NO-         PARAMÉTRICO.....</b>	<b>14</b>
<b>1.6. AGRICULTURA D EPRECISIÓN: SITUACIÓN NACIONAL.....</b>	<b>15</b>
<b>2. <u>ANALYSIS OF DIFFERENT STATISTICAL TOOLS FOR PRECISION     <u>AGRICULTURE IN RICE</u>.....</u></b>	<b>17</b>
<b>2.1. SUMMARY.....</b>	<b>17</b>
<b>2.2. RESUMEN.....</b>	<b>18</b>
<b>2.3. INTRODUCTION.....</b>	<b>19</b>
<b>2.4. MATERIALS AND METHODS.....</b>	<b>21</b>
<b>2.4.1. <u>Discrete sampling</u>.....</b>	<b>21</b>
<b>2.4.2. <u>Yield data</u>.....</b>	<b>22</b>
<b>2.4.3. <u>Geostatistical analysis</u>.....</b>	<b>23</b>
<b>2.4.3.1. Discrete sampling.....</b>	<b>23</b>
<b>2.4.3.2. Yield monitor data.....</b>	<b>24</b>

2.4.4. <u>Cluster analysis</u> .....	25
2.4.5. <u>Classification and regression trees (CART)</u> .....	26
2.4.6. <u>Multiple regression</u> .....	26
2.5. RESULTS.....	27
2.5.1. <u>Spatial distribution and descriptive statistics of soil properties</u> .....	27
2.5.2. <u>Cluster analysis</u> .....	28
2.5.3. <u>Spatial Multivariate Analysis</u> .....	31
2.5.4. <u>Multiple regression</u> .....	32
2.5.5. <u>Classification and regression trees analysis</u> .....	33
2.6. DISCUSSION.....	36
2.6.1. <u>Contrasting yield zones</u> .....	37
2.6.2. <u>Different statistical approaches for identify factors underlying yield</u> .....	38
2.7. CONCLUSIONS.....	40
2.8. AKNOLEDGMENTS.....	41
2.9. REFERENCES.....	41

3. <u>COSECHA LOS BENEFICIOS: AGRICULTURA DE PRECISI3N M3S QUE S3LO TECNOLOG3A</u> .....	46
3.1. SUMMARY.....	46
3.2. RESUMEN.....	47
3.3. INTRODUCCI3N.....	48
3.4. MATERIALES Y M3TODOS.....	51
3.4.1. <u>Muestreo discreto</u> .....	51
3.4.2. <u>Monitor de rendimiento</u> .....	52
3.4.3. <u>An3lisis geoestad3stico</u> .....	53
3.4.3.1. Muestreo discreto.....	53
3.4.3.2. Monitor de rendimiento.....	54
3.4.4. <u>An3lisis de conglomerado</u> .....	55

3.4.5. <u>Análisis económico</u> .....	56
3.5. RESULTADOS Y DISCUSIÓN.....	58
3.5.1. <u>Estadística descriptiva del rendimiento y las propiedades del suelo</u> .....	58
3.5.2. <u>Análisis de conglomerados: Delimitación de zonas de rendimiento</u> .....	59
3.5.3. <u>Uso práctico de las herramientas de agricultura de precisión</u> ....	63
3.6. CONCLUSIÓN.....	67
3.7. AGRADECIMIENTOS.....	68
3.8. BIBLIOGRAFÍA.....	68
4. <u>RESULTADOS Y DISCUSIÓN</u> .....	72
5. <u>CONCLUSIONES</u> .....	75
6. <u>BIBLIOGRAFÍA</u> .....	76

## RESUMEN

El objetivo del trabajo fue definir e integrar metodologías de procesamiento espacial de datos colectados a nivel predial, para generar información como punto de partida al manejo sitio-específico en una chacra comercial de arroz. La dificultad fundamental en el estudio de los factores que subyacen en la variabilidad del rendimiento intra-chacra, es la complejidad del fenómeno. Analizar los patrones espaciales del rendimiento, su estabilidad y las causas de su variación, es un punto crítico para evaluar el uso de manejo de precisión en sistema comerciales. La variabilidad espacio-temporal de datos de monitor de rendimiento relevados durante dos zafas, en una chacra comercial de arroz de 100 ha, fue cuantificada y caracterizada. El efecto de los factores de manejo agronómico y la distribución espacial de las propiedades del suelo, como causas de esa variabilidad, fue estudiado. Se realizó un análisis económico a partir de mapas de ingreso neto. El análisis de conglomerado, identificó 2 grupos de rendimiento contrastantes. Asociados a ellos, se encontraron diferencias en el contenido de arena y fósforo del suelo y diferencias en la implantación, lámina de agua y control de malezas. La zona de bajo rendimiento presentó pérdidas promedio de 361 U\$S ha<sup>-1</sup> y la zona de alto rendimiento ganancias promedio de 280 U\$S ha<sup>-1</sup>. Las tecnologías asociadas a la agricultura de precisión permitieron el estudio de la variabilidad espacial del rendimiento en chacras que a simple vista parecen uniformes. Se identificaron grupos de rendimiento, los cuales se verían beneficiadas con un manejo sitio-específico. Este estudio demostró que existen diferentes herramientas estadísticas para analizar este tipo de datos. Estas herramientas con diferentes enfoques mostraron resultados consistentes y robustos. A su vez el análisis económico permitió el uso de la información de los monitores de rendimiento, como una herramienta más para la toma de decisiones a nivel predial.

**Palabras clave:** Manejo sitio-específico, análisis de conglomerados, análisis espacial multivariado, árboles de clasificación y regresión, regresión múltiple.

# ANALYSIS OF DIFFERENT STATISTICAL TOOLS FOR PRECISION AGRICULTURE IN RICE

## SUMMARY

The objective of this study was to define and integrate methodologies for processing spatial data collected at farm level to generate information as a starting point to site-specific management in a commercial rice farm. The fundamental difficulty in the study of the factors underlying the variability of inter-farm yield is the complexity of the phenomenon. To analyze the yield spatial patterns, stability and causes of its variation, is critical to evaluate the use of accurate management in commercial system. The spatiotemporal variability of yield monitor data collected during two years in a commercial field of 100 ha. was quantified and characterized. The effect of agronomy management factors and the spatial distributions of the soil properties as the causes of the spatial variability was studied. An economic analysis was performed from net income maps. Cluster analysis identified two contrasting yield-zones. Associated to those, differences in sand content, soil phosphorus, emergence, water depth, and weed control were found. Low-yielding zone had average losses of 361 U\$ ha<sup>-1</sup> and high-yielding zone had average gains of 280 U\$ ha<sup>-1</sup>. The technologies associated with precision agriculture allowed the study of the spatial variability of yield in farms that at first glance seems uniform. Yield groups were identified which would benefit from site-specific management. This study demonstrated that there are different statistical tools to analyze this type of data. These tools with different approaches showed consistent and robust results. Also, the economic analysis allowed the use of the information on performance monitors as a tool for decision making at the farm level.

**Keywords:** Site-specific management, cluster analysis, multivariate spatial analysis, classification and regression trees, multiple regression.



## **1. INTRODUCCIÓN**

### **1.1. AGRICULTURA DE PRECISIÓN**

La combinación de suelos y la variación en sus propiedades, los atributos del terreno y las diferentes prácticas de manejo realizadas en las chacras, resultan en una alta variabilidad espacial de las condiciones ambientales que pueden afectar los rendimientos de los cultivos (Mulla y McBratney, 2002). Su identificación brinda la posibilidad de dividir a las chacras en regiones de baja, media y alta productividad potencial (Doerge, 1999).

El primer paso crítico para el control de la variabilidad, es su estudio y cuantificación. La agricultura de precisión es la aplicación de tecnologías y principios para manejar la variabilidad espacial y temporal asociada con todos los aspectos de la producción agrícola con el propósito de mejorar la performance de los cultivos y la calidad ambiental (Pierce y Nowak, 1999).

El manejo sitio-específico es el manejo de los cultivos agrícolas a escalas espaciales menores que toda la chacra. Para que el mismo sea justificado se deben cumplir 3 principios básicos: debe existir suficiente variabilidad espacial en factores que afecten el rendimiento, la/s causa/s de esa variabilidad deben ser identificadas y medidas, y la información surgida de esas medidas debe poder ser usada para modificar el manejo de los cultivos, para incrementar las ganancias o disminuir el impacto ambiental (Plant, 2001).

Para la estimación o cuantificación de la variabilidad espacial de los suelos y cultivos en las chacras, existen métodos de medición: discretos (ej: muestreo de suelos y plantas), continuos (ej: monitores de rendimiento) y remotos (ej: imágenes satelitales).

El método más tradicional para obtener información acerca de las condiciones del suelo y el estado de los cultivos, es el muestreo discreto de puntos (Plant, 2001). A partir de las muestras colectadas, se cuantifican las condiciones promedio de la chacra.

Este método se aplica por ejemplo, cuando se decide fertilizar un cultivo en función del promedio de las muestras realizadas en suelo o plantas (Roel y Terra, 2006).

La variabilidad existente en las chacras implica que los muestreos tengan un error sistemático asociado. Con el fin de minimizar este error, se utilizan distintos métodos de muestreo, los cuales se diferencian en la disposición espacial de los puntos a ser relevados. Algunos de los diseños más conocidos son: muestreo al azar simple, muestreo al azar estratificado, muestreo sistemático estratificado, muestreo sistemático estratificado desalineado, muestreo por juzgamiento, muestreo adaptativo, muestreo de búsqueda, muestreo geoestadístico, muestreo dirigido, entre otros (Mulla y McBratney, 2002).

Los monitores de rendimiento son dispositivos que periódicamente miden el flujo del material cosechado y basado en estas medidas computan un valor georreferenciado de rendimiento del cultivo (Plant, 2001). A partir de estas estimaciones, se construyen mapas de rendimiento que son una importante herramienta para inferir las características ambientales y el potencial de productividad de los suelos (Stafford et al., 1998). Con ellos, los agricultores pueden identificar las áreas dentro del campo donde los rendimientos pueden ser mejorados, o áreas donde se necesitan ajustes en los insumos para optimizar la rentabilidad de las explotaciones y la calidad del medio ambiente. Debido a que los mapas de rendimiento son una herramienta básica para el manejo de insumo y la mejora de la rentabilidad a nivel predial, los mismos son esenciales para la implementación del manejo sitio-específico (Pierce et al., 1997).

## **1.2. ANÁLISIS DE DATOS ESPACIALES: GEOESTADÍSTICA**

En agricultura de precisión, además del conocimiento de la media y la desviación estándar de una muestra poblacional, es fundamental conocer la estructura y la correlación espacial de la población muestreada. Cuando existen patrones espaciales, la asunción de que los errores son independientes e idénticamente distribuidos deja de ser válida (Bhatti et al., 1991). Los valores de rendimiento que tienen una referencia

espacial asociada, no pueden ser tratados como una muestra aleatoria debido a que los puntos de muestro son colectados en intervalos sistemáticos (Bocchi y Castrignano, 2007). Por lo tanto, se espera una relación entre los datos de rendimiento distribuidos espacialmente y la fuerza de esa relación es función de su separación espacial.

Las medidas clásicas de tendencia incluyen determinar la media, mediana y la moda. Por su parte, las medidas de distribución de la población incluyen calcular el desvío estándar y la varianza. Otras medidas de distribución incluyen el rango, los cuartiles y el coeficiente de variación (CV). En la medida que aumentan el rango, el desvío estándar y el CV, también lo hace la variabilidad de la población (Mulla y McBratney, 2002). Sin embargo, la distribución de frecuencias no provee ninguna información sobre la correlación espacial entre las muestras. La estructura espacial de una población puede ser estimada utilizando una rama de la estadística aplicada denominada Geoestadística (Mulla y McBratney, 2002).

La geoestadística es una rama aplicada de la estadística que cuantifica la dependencia y estructura espacial de una propiedad medida. A su vez, utiliza esa estructura espacial para predecir valores de esa propiedad en zonas no muestreadas. Estos 2 pasos incluyen típicamente, la modelación espacial (variografía) e interpolación espacial (kriging o krigeado) (Mulla y McBratney, 2002).

El mapeo de los datos es uno de los procedimientos de mayor importancia en agricultura de precisión. Para representar las variables de interés en dichos mapas, se deben transformar datos puntuales, y por lo tanto discretos, en variables continuas. El método que hace posible tal transformación es la interpolación de los datos. A su vez, para que la interpolación de la información sea posible, es necesario que los valores de la variable en estudio se encuentren asociados espacialmente, fenómeno que se denomina autocorrelación. Según Best y León (2006) la autocorrelación puede ser definida como la “correlación de una variable con respecto a si misma a través del espacio”. En geoestadística, los objetivos de la determinación de la autocorrelación para las variables en estudio corresponden a: 1) medir la fuerza de la autocorrelación espacial dentro de un mapa y 2) probar el supuesto de independencia o aleatoriedad en los datos.

El índice de Moran (I) se utiliza para cuantificar si una variable en un mapa, presenta o no autocorrelación espacial. Este índice se aplica a zonas o puntos con variables continuas asociadas dentro de estas zonas. Una de las particularidades de éste índice es que compara el valor de la variable en cualquier punto, con el valor de todas las otras posiciones (Best y León, 2006).

$$I = \frac{N \sum_i \sum_j W_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{ij}) \sum_j (X_i - \bar{X})^2}$$

Donde:

- N: número de casos;
- $X_i$ : valor de la variable en una posición particular;
- $X_j$ : valor de la variable en otra posición;
- $\bar{X}$ : media de la variable; y
- $W_{i,j}$ : peso aplicado a la comparación entre la posición i y la posición j.

El índice I varía entre -1 y +1. Si existe correlación espacial, entonces los valores estarán cercanos a la unidad (positiva o negativa) mientras que si no existe correlación espacial, los valores serán próximos a 0 (Best y León, 2006).

Luego que se determina la existencia de asociación espacial entre los valores de la variable en estudio, se pueden aplicar técnicas de modelaje superficial a través de la interpolación espacial (Kravchenko, 2003).

La base de la geoestadística moderna es tratar a la variable de interés como una variable aleatoria. Esto implica que en cada punto,  $\mathbf{x}$  en el espacio, existe una serie de valores para la propiedad  $Z(\mathbf{x})$ , y el valor que es observado,  $z(\mathbf{x})$ , se encuentra bajo alguna distribución de probabilidad. En  $\mathbf{x}$ ,  $Z(\mathbf{x})$  es una variable aleatoria con una media,  $\mu$ , y la varianza,  $\sigma^2$ . El conjunto de variables aleatorias,  $Z(\mathbf{x}1), Z(\mathbf{x}2), \dots$ , es un proceso aleatorio, y el valor real de  $Z$  observado, es sólo uno de las posibles realizaciones del proceso. Para describir la variación del proceso aleatorio, se asume el hecho de que en

lugares cercanos, los valores de las variables regionalizadas tienden a tener autocorrelación. Por lo tanto, para describir esta relación entre pares de puntos, se puede estimar la covarianza espacial:

$$C(x_1, x_2) = E\{[Z(x_1) - \mu(x_1)] [Z(x_2) - \mu(x_2)]\}$$

Donde  $\mu(x_1)$  y  $\mu(x_2)$  son medias de  $Z$  en  $x_1$  y  $x_2$ , y  $E$  es el valor esperado. Debido a que existe una sola realización de  $Z$  en cada punto, la solución es inviable porque la media es desconocida. Por lo tanto, se debe asumir estacionariedad (Oliver, 2010). Bajo la asunción de estacionariedad, algunos atributos del proceso aleatorio son iguales en cualquier lugar. Se asume que la media  $\mu = E[Z(x)]$  es constante para todo  $x$ , por lo que  $\mu(x_1)$  y  $\mu(x_2)$  son reemplazadas por  $\mu$ , que puede ser estimada por muestreos repetidos. Cuando  $x_1$  y  $x_2$  no coinciden, su covarianza depende de su separación y no de su posición absoluta.

$$C(x_i, x_j) = E\{[Z(x_i) - \mu] [Z(x_j) - \mu]\} = E\{[Z(x)] [Z(x+h) - \mu^2]\} = C(h)$$

Por lo tanto, la covarianza es función de la distancia y describe cuantitativamente la dependencia entre valores de  $Z$  cuando cambian las distancias (Oliver, 2010). Sin embargo, la media tiende a cambiar a través de la región y la varianza tiende a aumentar indefinidamente a medida que aumenta el área. Como consecuencia, no hay un valor de  $\mu$  que pueda ser utilizado para calcular la covarianza. Matheron (1963) estableció que a pesar que la media general no es constante, lo será a pequeñas distancias y por lo tanto se espera que las diferencias sean cero:

$$E\{Z(x) - Z(x+h)\} = 0$$

A partir de esto, las covarianzas pueden ser calculadas como:

$$E\{[Z(x) - Z(x+h)]^2\} = \text{var} [Z(x) - Z(x+h)] = 2\gamma(h)$$

El valor de  $\Upsilon(h)$  es conocido como la semivarianza a una distancia  $h$ , o como la varianza por punto cuando los puntos son considerados como pares. Al igual que para la covarianza, la semivarianza depende de la distancia entre los puntos y no de la posición absoluta de los mismos.

Dada una variable aleatoria estacionaria  $Y$  con media  $\mu$  y varianza  $\sigma^2$ , la media y la varianza son independientes de la locación, esto es,  $\mu(u)=\mu$  y la  $\sigma^2(u)=\sigma^2$  para todas las localidades o en el área de estudio. La dependencia entre los valores encontrados en función de la distancia que separa las muestras, se determina a través de la construcción de un semivariograma (Goovaerts, 1999). El variograma es una descripción matemática de la relación entre la varianza entre pares de observaciones (puntos de datos) y la distancia que separa esos puntos (Cressie, 1991). El semivariograma  $\Upsilon(h)$  es la mitad del variograma  $2\Upsilon(h)$  y es estimada por:

$$y(h) = \frac{1}{2n} \sum_{i=1}^n [z_i - z_{i+h}]^2$$

Donde:

- $Y(h)$  = semivariograma
- $h$  = distancia
- $n$  = número de observaciones pares separadas por  $h$
- $z_i z_{i+h}$  = pares de muestras observadas separadas por  $h$

El estimador clásico de semivarianza, es sensible a valores outliers debido a que un valor inusual es utilizado para calcularla, en  $N-1$  pares de datos. El estimador de Cressie-Hawkins es más robusto a estos valores outliers (Cressie, 1991) y se calcula de la siguiente manera:

$$\hat{\gamma}(h)_{CH} = \frac{0.5 [\sum_{N(h)} |Z(s_i) - Z(s_j)|^{1/2}]^4}{0.457 + \frac{0.494}{N(h)} + \frac{0.045}{N(h)^2}}$$

Si la dependencia espacial es función no sólo de la distancia pero también de la dirección, se denomina anisotrópica. Si únicamente depende de la distancia, se

denomina isotrópica. La dependencia espacial puede ser cuantificada y modelada utilizando semivariogramas (Cressie, 1991).

El ajuste de un modelo para el semivariograma observado, permite predecir el valor de la variable en estudio para los sitios no muestreados utilizando información recabada en sitios vecinos (Wollwenhaupt et al., 1997). Existen diversos modelos teóricos de semivarianza que pueden ajustarse al semivariograma experimental. En general dichos modelos pueden dividirse en no acotados (lineal, logarítmico, potencial) y acotados (esférico, exponencial, gaussiano) (Warrick et al., 1986). Los del segundo grupo garantizan que la covarianza de los incrementos es finita.

Los semivariogramas frecuentemente aumentan en valor a medida que la distancia aumenta y alcanzan una constante a separaciones a grandes distancias. En tales casos, los semivariogramas son caracterizados por 3 parámetros principales, rango, asíntota (“meseta”) y ordenada en el origen (“pepita”) (range, sill and nugget). El rango es la distancia a partir de la cual la propiedad evaluada en las muestras de suelo deja de ser espacialmente dependiente. Es por esto que de ser conocido el rango, debería ser la máxima distancia a la cual muestrear el suelo. El valor que toma el semivariograma en el rango es conocido como “meseta” (sill). Su valor es frecuentemente igual a la varianza del muestreo. A medida que las distancias se acortan, el valor del semivariograma disminuye. Sin embargo, no siempre que las distancias se hacen cero, el valor del semivariograma disminuye a cero, esto se debe al efecto “pepita” (nugget). Este efecto representa la varianza debida a la variabilidad de muestreos a menores distancias que los muestreos más cercanos, y a la variabilidad debida a errores analíticos y de muestreo (Wollwenhaupt et al., 1997).

La tasa N/S (nugget/sill) define la proporción de la variabilidad, en el corto rango, que no puede ser explicada por el modelo geoestadístico. El rango de la correlación espacial, define la distancia a lo largo de la cual los valores de la propiedad en estudio están correlacionados entre ellos. Pequeñas tasas N/S y amplios rangos usualmente indican que gran precisión puede ser lograda en el mapeo de la variable (Isaaks y Srivastava, 1989). Una tasa elevada de N/S, por ejemplo 0,6, corresponden a una débil

estructura espacial ya que significa que el 60% de la variación de los datos se debe a variaciones aleatorias inexplicables de corta distancia. Estructuras espaciales medianas y fuertes presentan valores de N/S de 0,3 y 0,1 respectivamente.

La precisión alcanzada en el mapeo de las propiedades del suelo depende de la estructura espacial. Cuanto más fuerte sea la correlación espacial, más preciso será el mapa. Propiedades del suelo con fuerte o media estructura espacial, con independencia de la variabilidad total, pueden ser mapeadas con exactitud incluso con un bajo número de muestras. Mapas precisos de propiedades con débiles estructuras espaciales pueden ser obtenidos solamente con un muestreo denso. La exactitud de los datos colectados dependerá de los criterios de aleatorización utilizados. Por lo tanto, varios conjuntos de datos serán necesarios para efectuar decisiones acerca de cuál será el plan de muestreo y de interpolación óptimos (Kravchenko, 2003).

Algunas de las técnicas de interpolación que permiten la caracterización de la distribución de los datos en un mapa son: el vecino más próximo (nearest neighbor), distancia inversa a una potencia ponderada (inverse distance to a power-weighted – IDW-), método de Shepard modificado (modified shepard's method), función de base radial (radial basis function), vecino natural ponderado (natural neighbor weighted), método de triangulación y regresión polinomial (Best y León, 2006). Los 2 métodos más comúnmente usados en prácticas agrícolas son: IDW y kriging (Kravchenko, 2003).

Cuando los parámetros del variograma son conocidos, debido a muestreos previos o gracias a información auxiliar, entonces el método de kriging es preferido con respecto al método IDW ya que su performance es mejor para los diversos tamaños de grillas y estructuras espaciales. El método de kriging con parámetros de variograma estimados se comporta de manera similar al método de IDW para conjuntos de datos con suficientes puntos. Sin embargo, el método IDW es más preciso cuando los variogramas obtenidos no son confiables debido a un número insuficiente de puntos muestreados o a largas distancias entre dos puntos. Incluso cuando la distancia entre dos puntos muestreados excede el rango de la correlación espacial, IDW continúa siendo un correcto método de interpolación sobre todo para propiedades del suelo con mediana o fuerte estructura



espacial. Es por esto, que IDW es recomendado para ser usado en pequeños conjuntos de datos para los cuales los parámetros de los variogramas no son conocidos (por lo que son estimados) y para conjunto de datos para los cuales existen grandes distancias entre dos puntos de la grilla (Kravchenko, 2003).

La palabra kriging procede del nombre del geólogo sudafricano D.G.Krige y encierra un conjunto de métodos de predicción espacial que se fundamentan en la minimización del error cuadrático medio de predicción. La técnica de interpolación de kriging se basa en la idea de que se pueden desarrollar inferencias considerando una función aleatoria  $Z(x)$ , dados los puntos  $Z(x_1), Z(x_2), \dots, Z(x_n)$  (Cressie, 1991). El kriging consiste en una combinación lineal ponderada de los valores muestreados. La diferencia que presenta respecto a otros métodos de interpolación radica en que los “pesos” asignados a cada valor muestreado, derivan de un modelo estadístico. Los mismos son determinados usando la semivarianza entre dos estaciones de muestreo (Best y León, 2006). Los pesos necesarios para el kriging son derivados de 2 criterios estadísticos. El primero requiere que la interpolación sea insesgada, por lo cual en promedio, las predicciones coinciden con lo que se trata de predecir. Es decir, los “pesos” asignados a los valores utilizados para predecir otro, deben sumar una unidad. El segundo criterio es elegir dentro de los predictores lineales e insesgados, el mejor de ellos. En sentido estadístico el mejor predictor es aquel que minimiza la varianza del error de predicción. Es por esto que el kriging es llamado el mejor, lineal e insesgado predictor (BLUP) (Wollenhaupt et al., 1997).

La formulación original del kriging, conocido como Kriging Ordinario es el método más robusto y utilizado. El mismo se basa en la asunción de que la media no es conocida. Considerando una variable aleatoria,  $Z$ , de la cual se obtienen mediciones en los puntos  $x_i, i=1,2,\dots,n$ ; se desea predecir  $Z(x_0)$  en el punto  $x_0$  donde no hubo medición. El método kriging ordinario propone que el valor de la variable puede predecirse como una combinación lineal de las  $n$  variables aleatorias:

$$Z^*(x_0) = \sum \lambda_i Z(x_i)$$

En donde los  $\lambda_i$  representan los pesos o ponderaciones de los valores originales.

Dichos pesos se calculan en función de la distancia entre los puntos muestreados y el punto donde se va a hacer la correspondiente predicción. La suma de los pesos debe ser igual a uno para que la esperanza del predictor sea igual a la esperanza de la variable. Estadísticamente la propiedad de insesgamiento se expresa como:  $E(Z^*(x_0)) = E(Z(x_0))$ . Por lo tanto, se dice que  $Z^*(x_0)$  es el mejor predictor lineal, porque los pesos se obtienen de tal manera que minimicen la varianza del error de predicción (Oliver, 2010). Esta última propiedad, es la característica distintiva de los métodos kriging sobre otros métodos de interpolación, como el de distancias inversas, que no garantizan varianza mínima de predicción.

### **1.3. ANÁLISIS DE CONGLOMERADOS PARA DELIMITAR DIFERENTES ZONAS DE RENDIMIENTO**

La interpretación de las causas de la variabilidad requiere un profundo conocimiento de las condiciones ambientales (Schroder et al., 2000). El análisis de conglomerado puede proporcionar una cuantificación objetiva de la estructura espacial de los patrones de rendimiento, así como permitir una identificación de la consistencia de estos patrones año en año (Pérez-Quezada et al., 2003). A partir de ellos se podrían definir diferentes zonas de rendimiento del cultivo, con similares capacidades de respuesta a los insumos y reducir el impacto ambiental (Bocchi y Castignano, 2007).

El reconocimiento automatizado de patrones, por medio de la agrupación multivariable, se propone como una herramienta para la interpretación de la variación temporal y espacial de rendimiento de los cultivos (Lark y Stafford, 1997). La dificultad fundamental en el estudio de los factores que influyen en el rendimiento en chacras comerciales es la complejidad de los fenómenos. Los factores ambientales que influyen en el crecimiento y desarrollo de los cultivos pueden ser relativamente permanentes, tales como las propiedades del suelo, o pueden ser transitorios, tales como un sector enmalezado. Una forma de reducir esta complejidad es tratar de organizar el campo en subregiones, a través del estudio del comportamiento espacio-temporal del rendimiento.

El agrupamiento de objetos conlleva inicialmente a una pérdida de información ya que se sitúan en una misma clase unidades que no son idénticas (sólo semejantes). Sin embargo, la síntesis de la información disponible sobre las unidades consideradas puede facilitar considerablemente la visualización de relaciones multivariadas de naturaleza compleja. Se recurre a técnicas de agrupamiento cuando no se conoce una estructura de agrupamiento de los datos “a priori” y el objetivo operacional es identificar el agrupamiento natural de las observaciones. Las técnicas de clasificación basadas en agrupamientos implican la distribución de las unidades de estudio en clases o categorías de manera tal que cada clase reúne unidades cuya similitud es máxima bajo algún criterio (Blanco, 2006).

Los algoritmos de clasificación pueden ser jerárquicos o no jerárquicos. En los primeros se genera una serie de particiones encajadas. Los grupos que se forman a un nivel de distancia comprenden grupos obtenidos a un nivel de distancia inferior. Se realiza un proceso de agrupación sucesiva cuyo resultado final es una jerarquía de unión completa en la que cada grupo se une en una determinada fase. Una secuencia ascendente de particiones de  $I$  forman una jerarquía si y solo si para cada partición  $P_q$  y  $P_s$  con  $s > q$  todo grupo de  $P_s$  está contenido en  $P_q$  (Blanco, 2006). Por otra parte, en los métodos no jerárquicos se produce una partición en un predeterminado número de clases que no se solapan y en alguno de los métodos se permite la reasignación de elementos a los grupos.

Uno de los métodos no jerárquicos es el de  $k$ -medias (Jain y Dubes, 1984). Este algoritmo, a partir de un número fijo de  $k$ -grupos, busca minimizar la varianza dentro de los grupos y maximizar la varianza entre ellos. Para ello,  $k$  puntos en el espacio de datos se seleccionan inicialmente como “semillas” para comenzar a formar los grupos. Luego las medias de cada grupo se seleccionan como el nuevo conjunto de  $k$  semillas y el proceso teóricamente se repite de forma iterativa hasta converger.

El objetivo de utilizar análisis de conglomerados con datos espacio-temporales, es desarrollar grupos que puedan ser utilizados para identificar factores que estén subyaciendo en los patrones de rendimiento observados. Esto es, los grupos deben ser

“bio-físicamente” significativos. Un criterio para comprobarlo es que los patrones sean estables, esto es; deben poder ser contruidos a partir de diferentes conjuntos de semillas iniciales. Otra forma de chequear la significancia agronómica de los grupos, es que los mismos deben estar espacialmente organizados. Para no depender de la visualización, se puede utilizar un estadístico que mide la consistencia interna de los grupos (Roel y Plant, 2004a). Suponiendo que existen  $p$  datos localizados, el set bajo consideración tiene  $n$  replicaciones y un indicador  $\beta_i$ ,  $i=1, \dots, n$ , que se define para cada ubicación  $i$ , siendo  $\beta_i=1$  si todos los miembros del set  $i$  pertenecen al mismo grupo o  $\beta_i=0$  si al menos una locación del set  $i$  pertenece a un grupo diferente.

Se obtiene así la medida de consistencia  $\gamma$  del set que está definido como:

$$\gamma = \frac{\sum_{i=1}^p \beta_i - \frac{p}{n^k}}{p - \frac{p}{n^k}}$$

El valor  $p/n^k$  es el valor esperado de celdas que pertenecen al mismo grupo por chance. Si todos los miembros del set son idénticos, entonces  $\gamma=1$ . Por otro lado, si todas las celdas de los miembros del set son asignados al azar, entonces  $\gamma=0$ . Los valores  $\gamma$  se computan para cada conjunto de grupos y luego se testea la autocorrelación espacial a partir del coeficiente de Moran como un indicador de los patrones espaciales (Cliff y Ord, 1981).

#### **1.4. ANÁLISIS DE VARIABILIDAD ESPACIAL MULTIVARIADA EN AGRICULTURA DE PRECISIÓN**

Conocer la variabilidad y co-variación espacial de variables de suelo, es crucial en agricultura de precisión para comprender y manejar la variabilidad en los rendimientos. Los métodos de análisis multivariado permiten identificar y resumir las principales estructuras de variación y co-variación en bases de datos que contienen la descripción de un número de observaciones para varias variables. Cuando hay información espacial

disponible para cada observación, pueden construirse mapas de la variación espacial multivariada. Dos objetivos principales son relevantes en el análisis de datos espaciales multivariados: resumir estructuras de correlación entre variables e identificar patrones espaciales (Dray y Jombart, 2010). Para el primero se realiza análisis de componentes principales (PCA) que permite resumir las co-variaciones. Cuando se usa un PCA con datos georreferenciados, la información espacial es incorporada a posteriori del análisis a partir de una representación gráfica de la ubicación de la observación en el espacio (Goodall, 1954). Otra alternativa es aplicar un análisis geoestadístico univariado a los componentes principales (CP) y luego utilizar técnicas de interpolación. También se puede aplicar a cada variable un estadístico de autocorrelación como el coeficiente de Moran (MC) (Moran, 1948) para medir y analizar el grado de dependencia entre observaciones de una misma variable en un contexto geográfico. La significancia del MC observado puede ser testeada a partir de un procedimiento del tipo Monte Carlo, en el que las ubicaciones son permutadas para obtener la distribución del MC bajo la hipótesis nula de distribución aleatoria. Un valor observado del MC que es más grande que el esperado por azar, indica que valores similares se encuentran agrupados en el espacio, por lo que existe una autocorrelación espacial estadísticamente significativa.

Por otra parte, existen métodos de análisis multivariado basados en la combinación del PCA y el coeficiente de Moran. Los mismos permiten considerar simultáneamente el aspecto multivariado de las observaciones y la información geográfica sobre el sitio desde el cual se releva cada observación (Dray y Jombart, 2010). El método denominado MULTISPATI-PCA estudia la relación entre dos matrices, una de ellas contiene los valores de las variables en cada sitio y la otra contiene los valores (ponderados espacialmente) de las mismas variables en los sitios vecinos de cada observación (matriz “lagged”). El análisis MULTISPATI-PCA maximiza el producto escalar entre la combinación lineal de las variables originales y la combinación lineal de variables “lag”. El resultado es una combinación lineal de las variables que maximiza el producto de la autocorrelación (una versión generalizada del coeficiente de Moran) y la varianza calculada con PCA (Dray y Jombart, 2010).

## 1.5. ÁRBOLES DE DECISIÓN: UN ENFOQUE ALTERNATIVO NO-PARAMÉTRICO

La mayoría de las herramientas estadísticas utilizadas para el análisis en ecología, son métodos paramétricos. Los mismos están basados en modelos matemáticos de los datos y por lo tanto poseen asunciones específicas. Cuando alguna de las suposiciones no es cumplida, los resultados de los análisis pueden seguir siendo válidos, pero resulta difícil confiar en ellos (Vayssières, et al., 2000).

Los datos en ecología son habitualmente complejos, desbalanceados y con valores perdidos. Las relaciones entre las variables pueden ser no lineal e involucrar interacciones de orden superior (De'ath y Fabricius, 2000). Las técnicas estadísticas y de exploración utilizadas, a menudo no logran encontrar patrones ecológicos válidos a partir de los datos. Los árboles de clasificación y regresión (Vayssières et al., 2000) son técnicas estadísticas modernas utilizadas para explorar y modelar datos.

Los árboles explican la variación de una variable de respuesta a partir de una o más variables explicativas. La variable de respuesta es por lo general categórica (árboles de clasificación) o numérica (árboles de regresión) y las variables explicativas pueden ser categóricas y/o numéricas. El árbol se construye a partir de dividir varias veces los datos por una norma basada en una sola variable explicativa. En cada división, los datos se dividen en dos grupos mutuamente excluyentes y cada uno de ellos es lo más homogéneo posible. La división es luego aplicada a cada uno de esos grupos. El objetivo es particionar la variable de respuesta en grupos homogéneos, pero a su vez se debe mantener al árbol razonablemente pequeño. Cada grupo es caracterizado por el valor de la variable de respuesta, el número de observaciones dentro del grupo y el valor de la variable exploratoria que divide al grupo (De'ath y Fabricius, 2000).

El árbol puede ser utilizado para la exploración iterativa y para la descripción y la predicción de los patrones y procesos. Algunas ventajas que esta herramienta incluye son: 1) flexibilidad para manejar una amplia gama de tipos de respuesta, incluyendo numéricas, categóricas, calificaciones y datos de supervivencia; 2) invariancia de

transformaciones monótonas de las variables explicativas; 3) facilidad y solidez de la construcción; 4) facilidad de interpretación, y 5) capacidad de gestionar los valores perdidos tanto en la variable de respuesta como en las variables explicativas (Vayssières, et al., 2000).

## **1.6. AGRICULTURA DE PRECISIÓN: SITUACIÓN NACIONAL**

Los trabajos realizados en Uruguay, han buscado identificar los principales factores que afectan la variación de los rendimientos dentro de las chacras principalmente en cultivos de secano y la posibilidad de delinear ambientes o zonas de manejo en base a diferentes capas de información.

Terra et al., (2006) y Pravia, (2009) han encontrado que los factores más relacionados con la variación del rendimiento en cultivos de secano son los atributos edáficos y topográficos relacionados con la dinámica superficial del agua y capacidad de almacenamiento de la misma en el suelo. En el cultivo de arroz, al estar bajo un sistema de inundación, se busca identificar factores permanentes que estén asociados a la variación del rendimiento. Roel et al., (2007) en chacras comerciales de arroz encontraron que factores de manejo como lámina de agua y nivel de enmalezamiento, estuvieron muy relacionados con la variación de rendimiento en el cultivo.

Sin embargo, a nivel nacional existe una brecha entre la posibilidad de acceso, por parte de los productores, a tecnologías vinculadas a la agricultura de precisión (GPS, monitores de rendimiento, muestreos dirigidos georreferenciados) y la utilización de esas tecnologías para la generación de información y conocimiento.

En la cadena arrocera existe un creciente interés por la aplicación de tecnologías de agricultura de precisión con fines productivos. Poder delimitar ambientes con diferente potencial productivo dentro de chacras de arroz, con el uso de información derivada de métodos de muestreo continuo y discretos sería un buen principio para comenzar a desarrollar el manejo sitio-específico en este cultivo a nivel nacional. Como un punto de partida crítico, se encuentra el desafío de desarrollar una metodología

confiable para el procesamiento de la información disponible y su posterior utilización a nivel predial.

El objetivo general de este trabajo, fue definir e integrar metodologías de adquisición y procesamiento de datos espaciales colectados a nivel predial, para la generación de información como punto de partida a un posible manejo sitio-específico de las chacras comerciales de arroz.

Como primer objetivo específico se planteó evaluar y cuantificar la variabilidad espacial del rendimiento de una chacra comercial de arroz. Como segundo objetivo específico, se buscó mediante herramientas de agrupamiento, delinear diferentes áreas de rendimiento. El tercer objetivo específico fue analizar la relación entre las propiedades del suelo, factores de manejo y el rendimiento del cultivo, mediante la combinación de herramientas geoestadística y métodos paramétricos y no-paramétricos.

Este trabajo fue financiado con un proyecto innovagro INIA-ANII y una beca ANII de Maestría.

Este documento consta de dos artículos con formato de revista científica. El primer artículo: “Analysis of different statistical tools for precision agriculture in rice” discute la utilización de diferentes herramientas estadísticas con el fin de delinear grupos de rendimiento dentro de una chacra de arroz y las causas de esos agrupamientos como punto de partida a un posible manejo sitio-específico. El mismo tiene el formato de Precision Agriculture Journal. El segundo artículo: “Cosecha de beneficios: Agricultura de precisión más que sólo tecnología” plantea una utilización práctica con fines productivos de las tecnologías asociadas a la agricultura de precisión. El mismo tiene el formato de la Revista Agrociencia Uruguay.



## **2. ANALYSIS OF DIFFERENT STATISTICAL TOOLS FOR PRECISION AGRICULTURE IN RICE**

**Bonilla Camila, Roel Álvaro, Gutiérrez Lucía**

### **2.1. SUMMARY**

Rice farmers have a growing interest in the application of precision agriculture technologies for productive purposes. The aim of this study was to define and integrate methodologies of acquisition and processing spatial data collected at commercial field, to generate information as a starting point to a possible site-specific management of commercial rice farms. The spatiotemporal variability of yield monitor data from one rice fields growing in Rincón, Uruguay, was quantified and characterized. The role of the spatial distribution of soil properties and agronomy management, in driving yield variability was explored by different statistical approach. Using a k-means clustering method, two yield-contrasting zone could be identified. Different models for univariate semivariogram were tested for fitting the data but no model provided a good fit in all cases. Otherwise, multiple spatial analysis could summarize and describe the spatial soils properties variance of the field. Sand soil content and plant available phosphorus were identified as the underlying causes of the within-field patterns using multivariate spatial analysis and multiple regression. Agronomy management factors were identified as other underlying causes of spatial yield variability using classification and regression trees. This study demonstrated that there are many techniques and statistics tools for analyzing this type of data. Different tools with different approaches shown consistent and robust results. None of the methodologies are better or worse, but complementary. Depending on the purpose of the study and the characteristics of the data set the appropriated tools is going to be selected.

**Key words:** Cluster analysis, Multivariate spatial analysis, Classification and regression trees, Multiple Regression, Rice.

## 2.2. RESUMEN

Los productores de arroz presentan un interés creciente en la aplicación de tecnologías de agricultura de precisión con fines productivos. El objetivo de este estudio fue definir e integrar metodologías de adquisición y procesamiento de datos espaciales colectados a nivel comercial, para generar información a utilizar como punto de partida a un posible manejo sitio-específico en chacras comerciales de arroz. La variabilidad espacio-temporal de los datos de monitores de rendimiento para una chacra localizada en Rincón, Uruguay, fue cuantificada y caracterizada. El efecto de la distribución espacial de las propiedades del suelo y de factores de manejo agronómico en la variación espacial del rendimiento, fue explorado con diferentes enfoques estadísticos. Utilizando el método k-medias de conglomerado, 2 grupos de rendimiento contrastantes fue identificado. Diferentes modelos para semivariogramas univariados fueron testeados para ajustar los datos pero ningún modelo tuvo un buen ajuste. Por otro lado, el análisis espacial multivariado pudo resumir y describir la variabilidad de las propiedades del suelo en el campo. El contenido de arena en el suelo y la disponibilidad de fósforo para las plantas fueron identificados como los factores subyacentes en las causas de variabilidad dentro de la chacra utilizando análisis espacial multivariado y regresión múltiple. Factores de manejo agronómico fueron identificados como otras causas subyacentes en la variabilidad espacial del rendimiento utilizando árboles de clasificación y regresión. Este estudio demostró que existen diferentes técnicas y herramientas estadísticas para analizar este tipo de datos. Estas herramientas con diferentes enfoques mostraron resultados consistentes y robustos. Ninguna de las metodologías es mejor o peor, sino complementarias. Dependiendo del propósito del estudio y de las características del conjunto de datos se deberá seleccionar la herramienta apropiada.

**Palabras claves:** Análisis de conglomerado, Análisis especial multivariado, Árboles de clasificación y regresión, Regresión Múltiple, Arroz.

### 2.3. INTRODUCTION

Spatial heterogeneity generated by soil properties, terrain attributes, and management practices have a direct impact on crop yields (Mulla and McBratney 2002). Yield spatial variability is comprised of both permanent (i.e. soil properties) and transient (i.e. climate) components. The persistent patterns are controlled by endogenous factors, while the transient patterns are controlled by exogenous factors (Basso et al. 2007). Successful control of this variability requires an understanding of the spatial variability of biotic and abiotic factors that influence crop performance (Schroder et al. 2000). Precision agriculture is defined as the application of information technology to problems in crop production. Generally it is associated with site-specific crop management, which is the management of individual fields at a spatial and temporal scale appropriate to their natural variation (Plant 2001). Modern technologies associated with precision agriculture provide the opportunity to more precisely measure yield variability and the ecological processes underlying this variability (Pierce and Nowak 1999).

Rice (*Oryza sativa* L.) is one of the world's most important staple crops and one of the most important Uruguayan export crops. The average yields achieved in Uruguay are among the highest in the world (8500 kg ha<sup>-1</sup>) and in the last 40 years, adoption of high-yielding varieties and management practices has resulted in important yield increases. However, the increase in average yields has stalled in the last time and there is a gap between genetic yield potential and actual yield. Within-field spatial heterogeneity of soil properties can partly account for the yield gap (Mzuku et al. 2005). Land-leveling and flood-irrigation practices make irrigated rice system particularly complex due to their influence on endogenous and exogenous factors, such as soil and water chemistry and microclimate (Simmonds et al. 2013). Therefore, in Uruguay uniform management in irrigated rice fields is still use, despite the availability of technology to implement precision management at the within-field scale. Development an effective site-specific management techniques will promote a sustainable manner of production, increasing

productivity of existing rice cropping systems without environmental degradation (Cassman 1999).

Techniques associated with precision agriculture provide the opportunity to handle the spatial and temporal variability associated with agricultural production in order to improve crop performance and environmental quality (Pierce y Nowak 1999). Effective analyses of data from these measurements requires statistical methods different from those traditionally employed on data from controlled agronomic experiments. Various clustering approaches have been used to reduce the complexity of spatial phenomena in agricultural fields. Fuzzy clustering can be used to organize yield map data to combinable crops (Lark and Stafford 1997). K-means algorithms can identify yield clusters of similar spatiotemporal behavior (Pérez-Quezada et al. 2003; Jayne et al. 2003). Through the combination of different soil variables and multivariate analysis, different management zones can be identified (Córdoba et al. 2012a). Another approaches to identified different potential production areas is through multitemporal yield map analysis (Bocchi and Castrignano 2007). Dobermann et al. (2003) and Guastaferrero et al. (2010) compared several different classification methods for the delineation of management zones. Roel and Plant (2004a) found that the spatiotemporal yield variability could be characterized by spatially contiguous clusters and they assert measures to evaluate if the clusters were physically meaningful. On the other hand, alternative statistical methods can be used to explore the factors underlying yield variability. Multivariate statistics analysis founded that soil pH, microrelief, weed growth and seeding rates were driving yield variability in Russian rice fields (Dobermann 1994). Multiple regression and the boundary-line method founded that topsoil cation exchange capacity and salinity were the limits on the ability of rice to reach its yield potential (Casanova et al. 1999). Through a geostatistical and multivariate analysis of soil chemical properties yield-determining factors were evaluated (Yanai et al. 2001). Using remote sensed yield maps, differences in leaching loss of basal and top-dressed N fertilizers were founded as the likely limiting factor affecting the spatial variation of yield (Wang et al. 2012). Researchers using classification and regression

trees, reported that soil electrical conductivity and plant available phosphorus were the underlying causes of the within-field rice yield patterns (Simmonds et al. 2013).

Rice farmers have a growing interest in the application of precision agriculture technologies for productive purposes and delimitation of different potential productive zones for site-specific management in rice farms. However, as a critical starting point, the challenge is to develop a reliable methodology for processing the available information, identifying the yield underlying variables, and its subsequent use at the farm level. The aim of this study was to define and integrate methodologies of acquisition and processing spatial data collected at commercial field, to generate information as a starting point to a possible site-specific management.

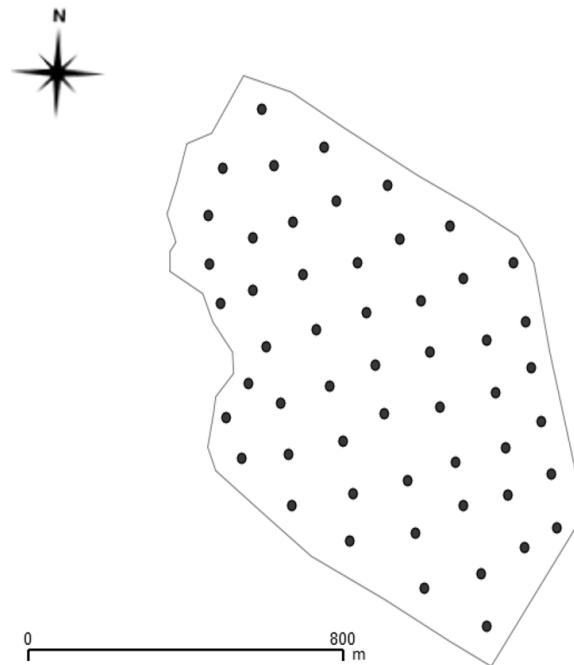
## **2.4. MATERIALS AND METHODS**

The study was performed during the course of two growing seasons 2011-2012 (season 1) and 2012-2013 (season 2) in a rice commercial field of 100 ha located near Rincón, Uruguay (32°50'12.14"S- 53°48'22.91"E). The soils of the study fields consist of Brunosoles Eutricos Luvicos and Solonetz Ocricos, silty loam, high fertility and imperfect drainage. The cultivar El Paso 144 was grown and managed by the cooperators, using standard practices for the area.

### **2.4.1. Discrete Sampling**

Fifty-three samples of soil and plant attributes were collected throughout the field (Figure 1). At each sample point, eight soil cores (0-15 cm depth) were extracted before planting from a circular region of about 5-m radius. Soil pH, Phosphorous Bray (ppm, Pb), Phosphorous citric (ppm, Pc), Potassium (meq 100 g<sup>-1</sup> of soil, K), and Soil texture (% Sand, Clay and Silt) were determined using standard methods of the National Institute Agriculture Research of Soil Analytical Laboratory. All soil properties were measured in both years except soil texture and pH, which were only measured in season

1, respectively and were assumed to be temporally stable in the timescale of the two-year study. During the crop cycle, plant height (cm), water depth (cm), biomass ( $\text{kg ha}^{-1}$ ) and level of weed infestation (subjective scale 1-5) were determined every 15 days. Also, a yield estimated value ( $\text{kg ha}^{-1}$ ) was obtained for each point with a manual harvest.



**Fig. 1:** Discrete Sampling locations

#### **2.4.2. Yield Data**

Rice was harvested using a John Deere harvester with global positioning system (GPS)-equipped commercial yield monitors, which were calibrated according to manufacturer's instructions prior each harvest. Yield map data files (geographic coordinates, yield and grain moisture) were collected and imported into the QGIS 1.8 (Open Source Geospatial Foundation, OSGeo) geographic information system for analysis. Raw yield data were cleaned in five stages prior to spatiotemporal analyses: (1)

points outside of the field boundary were deleted; (2) points with inconsistent logging intervals were deleted; (3) points within 10-m of the field boundary were omitted to remove edge effects; (4) data records with extreme yield values ( $<1500 \text{ kg ha}^{-1}$  or  $>15000 \text{ kg ha}^{-1}$ ) were omitted, and (5) the remaining dataset was passed through a Moran Scatterplot screening process from the R package `spdep` (Bivand et al. 2010) (R statistical software, version 2.10.1).

### **2.4.3. Geostatistical Analysis**

#### **2.4.3.1. Discrete Sampling**

Variography was used to analyze the spatial variability of the soil discrete sampling data. Large-scale variability (trend) and small-scale variability were separated by detrending the data using regression model technique (Cressie 1991). All sample variograms computations and model fitting were performed using R package `gstat` (Pebesma 2004). Different models were tested for fitting the data but no model provided a good fit in all cases. Therefore, no soil properties distribution maps were obtained.

Otherwise, multiple spatial analysis (MSPCA) could summarize and describe the spatial soils properties variance of the field. The first step of spatial autocorrelation analysis was to define a  $W_{n \times n}$  spatial weighting matrix. The spatial weights reflect a priori the absence ( $w_{ij}=0$ ), presence or intensity ( $w_{ij}>0$ ) of the spatial relationships between the locations concerned. Euclidean distance was used to define the spatial weighting matrix with a maximum distance of 200 m. Once the spatial weights were defined, the Moran spatial autocorrelation statistics (MC) was computed. A vector containing measurements of a quantitative variable for  $n$  spatial units ( $n \times 1$ )  $x = [x_1 \dots x_n]^T$  was considered. Then MC of spatial autocorrelation was computed:

$$MC(x) = \frac{n \sum_{(2)} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_2 w_{ij} \sum_{i=1}^n (x_i - \bar{x})} \sum_2 = \sum_{i=1}^n \sum_{j=1}^n$$

The significance of the observed value of MC was tested by a Monte Carlo procedure, in which locations are permuted to obtain a distribution of MC under the null hypothesis of random distribution. An observed value of MC greater than the expected at random indicates the clustering of similar values across space (positive spatial autocorrelation), while a significant negative value of MC indicates the neighboring values are more dissimilar than expected by chance. MSPCA maximizes the product between the variance and the spatial autocorrelation of the scores. For the analysis, principal components (PC) were analyzed with a variance and co-variance matrix for soil discrete sample data (Pb, Pc, K, Sand, Clay, Silt and pH). Then semivariogram for PC1 and PC2 were adjusted using R package gstat. Theoretical variograms models were fit to each experimental variogram. Different models were tested for fitting the data and the isotropic spherical model provided a good fit in all cases. Finally PC1 and PC2 were interpolated to a fixed 30-30 m grid using ordinary kriging and multivariate spatial variability maps were obtained.

Finally, Thiessen Polygon were used to aid in visualization of the spatial variability of the all discrete sampling variables (soil, plant and management practices). The analysis was performed using QGIS 1.8.

#### **2.4.3.2. Yield Monitor Data**

Yield point shapefile data were interpolated to a fixed 30-30-m grid using neighbor local kriging interpolation. Large-scale variability (trend) and small-scale variability were separated by detrending the data using regression model technique (Cressie 1991). A lineal trend on x and y was included in the model. For the characterization of the small-scale variation, experimental variograms were calculated using the residual yields from the regression models. Variography analyses assume the



data have a Gaussian distribution. Therefore, the distributions of the yield residuals were checked using Q-Q plots and Shapiro Test. Experimental variograms were calculated from the yield values. All sample variogram computations and model fitting were performed using R package gstat. A theoretical variogram model was fit to each experimental variogram. Different models were tested for fitting the data and the isotropic spherical model provided a good fit in all cases.

#### **2.4.4. Cluster Analysis**

In order to identify spatial-meaningful yield clusters, the k-means cluster algorithm was performed using the R package e1071 version 1.6 (Dimitriadou et al. 2010). Yield monitor data of season 1, season 2 and for combined season were used as the variables to perform the analysis. Under the assumption that physical properties of the field are spatial structured, one criteria to test is that the clusters be spatially structured. Another indicator of physical meaning is that the cluster pattern be attained starting from different sets of initial cluster seeds (Roel and Plant 2004a). For values of k from 2 to 4, data were processed in this manner 100 times. Spatial autocorrelation of cluster sets were tested by computing the Moran's statistics as an indicator of spatial pattern. Besides, a consistency of cluster sets was estimated as  $\gamma$ :

$$\gamma = \frac{\sum_{i=1}^p \beta_i - \frac{p}{n^k}}{p - \frac{p}{n^k}}$$

following Roel and Plant (2004a) where the quantity  $p/n^k$  is the expected number of cells that all belong to the same cluster by chance. If all of the members of the set are identical, then the value  $\gamma=1$ , and if each cell of each member of the set is randomly

assigned one of the  $k$  possible values, then  $\gamma=0$ . Finally, yield temporal variability was evaluated visually by examining the spatial stability of the cluster over time.

#### **2.4.5. Classification and regression trees (CART)**

A CART analysis was performed using the R package rpart version 3.1-45 (Therneau and Atkinson 2009). The yield clusters were the response variable and the soil parameters and management factors were the explanatory variables. The tree size corresponding to the lowest cost-complexity value was chosen as the optimal tree.

#### **2.4.6. Multiple Regression**

To obtain the optimum model for predicting the rice yield performance, stepwise forward multiple regression analysis was performed using the R package MASS (Ripley 2002) with AIC and Sum of Square Type III as selection criteria. Grain yield was used as a dependent variable, and  $P_c$ ,  $K$ , Sand (%), and pH were included as independent variables in the model in a stepwise forward manner.

A  $k$  partitioning cross-validation procedure was used to determine the predicting capacity of the best model. To perform this process, the original sample data was randomly split into  $k=10$  sub-samples of equal size. Of the  $k$ -subsamples, 9 sub-samples were used to construct the model while one sub-sample was used to test themes. The cross-validation process was repeated 10 times with each of the  $k$  sub-samples. Correlation between predicted and observed values was used to determine the performance of the model.

## 2.5. RESULTS

### 2.5.1. Spatial distribution and descriptive statistics of soil properties

Pb and Pc were the most variable components (CV range 48.97-76.37 %). K, Sand, Silt and Clay were moderately variable (CV range 12.56-20.02 %), while pH was the least variable (CV 5.96 %) (Table 1). The Thiessen polygon maps identified a zone with high percent of sand and high level of Pc and other zone with low values of pH (Figure 2). The mean of unhulled rice yield was 9114 kg ha<sup>-1</sup> on a dry weight basis in 2011 and 8554 kg ha<sup>-1</sup> in 2012 and the CV values were 14% and 15% suggesting a moderately within-field variability.

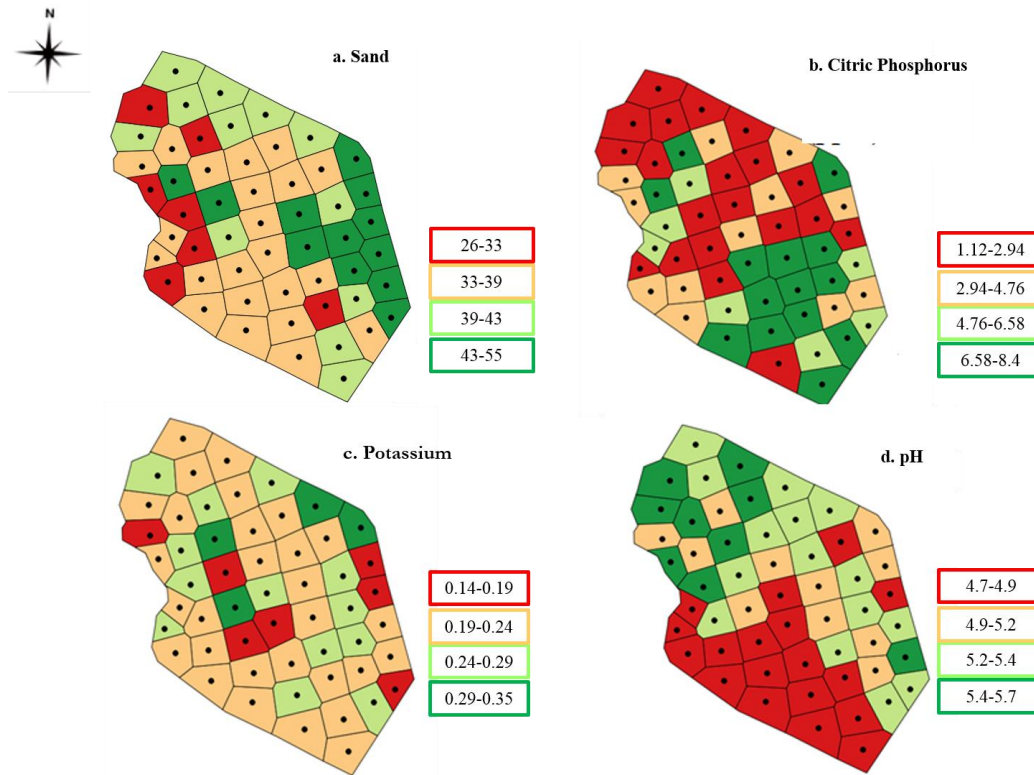
Table 1: Univariate statistics of discrete sample data in 2011-2012 and 2012-2013

Year	Statistic	pH <sup>a</sup>	P Bray (ppm)	P Citric (ppm)	K (meq 100 g <sup>-1</sup> of soil)	Sand <sup>a</sup> (%)	Silt <sup>a</sup> (%)	Clay <sup>a</sup> (%)	Yield (kg ha <sup>-1</sup> )
2011	Mean	-	2.60	4.29	0.24	40.56	32.22	27.22	9144
	SD <sup>b</sup>	-	1.27	2.20	0.04	6.63	4.05	4.69	1293
	CV <sup>c</sup>	-	48.97	50.07	17.66	16.35	12.56	17.22	14.14
2012	Mean	5.16	1.54	5.39	0.17	-	-	-	8554
	SD	0.31	1.14	4.11	0.04	-	-	-	1286
	CV	5.96	73.86	76.37	20.02	-	-	-	15.03

<sup>a</sup> Measured only in 2011

<sup>b</sup> Standard deviation

<sup>c</sup> Coefficient of variation= (SD/Mean)\*100



**Fig 2:** Thiessen polygons for **a.** Sand (%) soil content. **b.** Citric Phosphorus (ppm) soil content **c.** Potassium soil content (meq 100 g<sup>-1</sup> of soil). **d.** Soil pH.

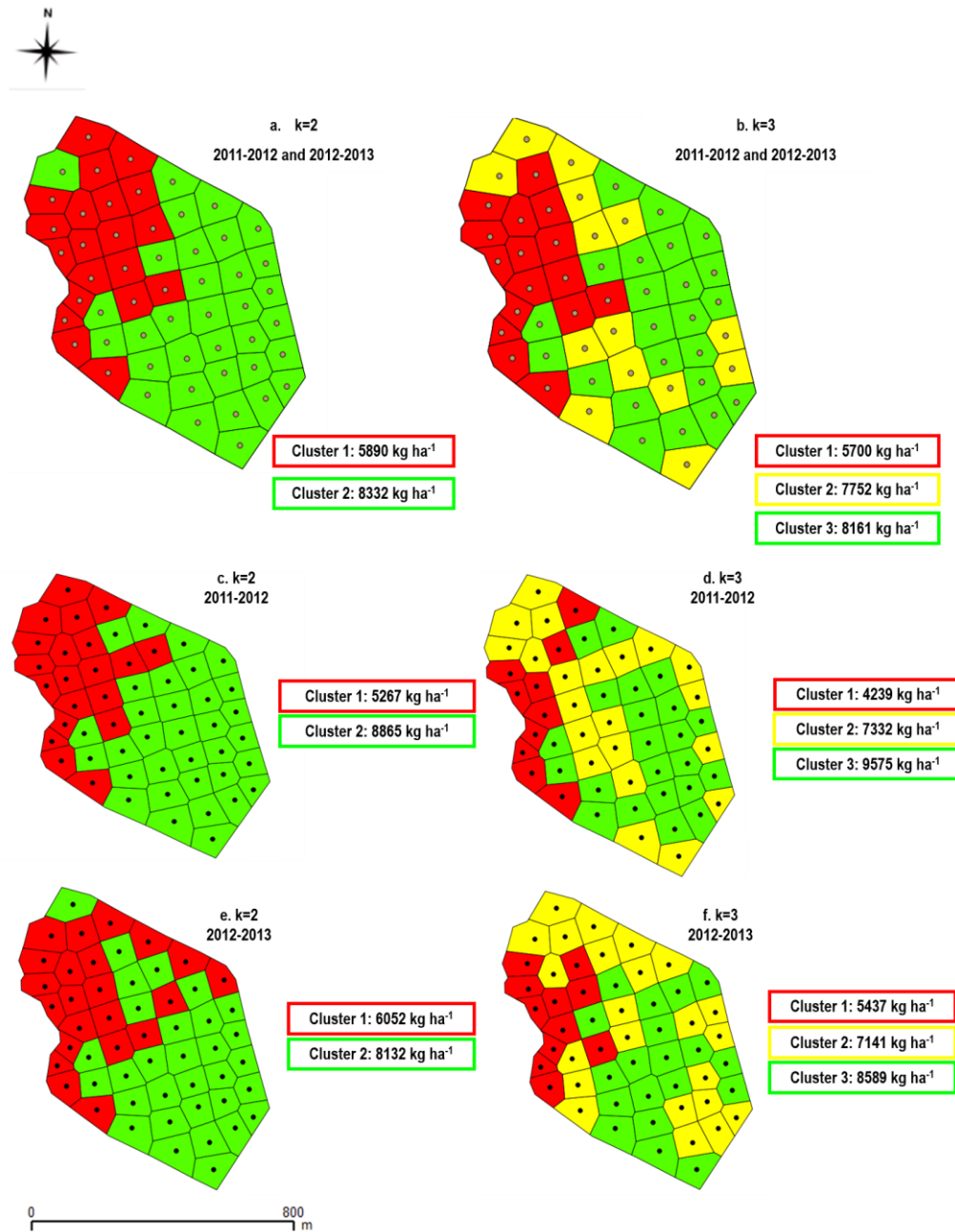
### 2.5.2. Cluster analysis

For the analysis for  $k = 2$ ,  $k = 3$  and  $k = 4$  cluster were stable to different random initial seeds and patterns remains constant with different repetitions (data not show). Additionally clusters were spatially correlated (Table 2).

Table 2: Cluster sets, number of k-means, the Moran`s coefficient and p-value.

Cluster Set	K	Moran`s coefficient
2	2	0.464 (0.001)
3.1	3	0.412 (0.001)
3.2	3	0.337 (0.001)
4.1	4	0.354 (0.001)
4.2	4	0.387 (0.001)
4.3	4	0.366 (0.001)

The degree of the spatial relationships and contiguity of the groups identified in the cluster analysis was significant (Figure 3). Two robust groups were formed with clusters algorithm for each year and for both at a time was k=2. Cluster 1 is composed of locations in the field with low yield performance (5890 kg ha<sup>-1</sup>) in both years and cluster 2 is composed of locations in the field with high yield performance (8332 kg ha<sup>-1</sup>) in both years.



**Fig. 3:** Cluster of interpolated yield data drawn using Thiessen polygons for **a.** k=2 for 2011-2012 and 2012-2013. **b.** k=3 for 2011-2012 and 2012-2013. **c.** k=2 for 2011-2012. **d.** k=3 for 2011-2012. **e.** k=2 for 2012-2013. **f.** k=3 for 2012-2013.

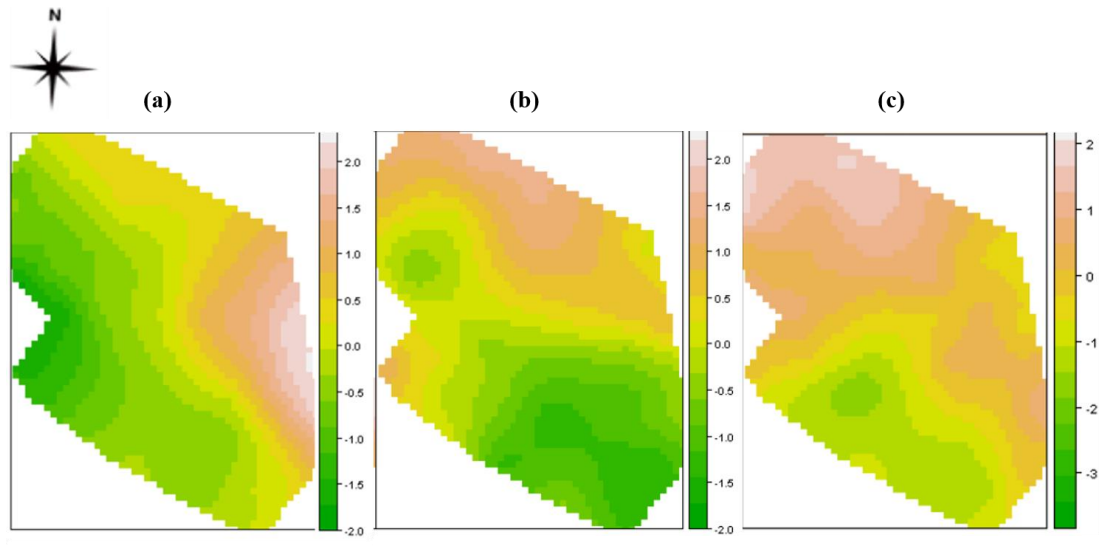
### 2.5.3. Spatial multivariate analysis

The first three principal components both in season 1 and season 2 (PC1 to PC3) accounted for more than 70% of the total variance (Table 3). The PC1 explain 36% and 34% of the total variance in 2011-2012 and 2012-2013. Since these properties were related to the texture soil, the first component was referred to as “texture soil factor” (TSF). PC2 resume the 23% and 25% of the total variance and principal components were Pb and Pc in 2011-2012 and pH in 2012-2013. In 2011-2012, since the properties of PC2 correspond to nutrient status of soil, this component was referred to as “phosphorus soil factor” (PSF). For 2012-2013 the PC2 correspond to “pH factor” (pHF). PC3 provides no extra information so it was not taken into account. In this way, variation of the soil properties was summarized into a small number of factors, which were independent of each other.

Table N°3: Component loading, eigenvalues, percentage of total variance explained for the first three principal components.

	2011-2012			2012-2013		
	PC1 (36%)	PC2 (23%)	PC3 (15%)	PC1 (34%)	PC2 (25%)	PC3 (16%)
pH	-	-	-	-0.11	0.87	0.29
P Bray (ppm)	0.004	-0.54	0.68	-0.13	-0.25	-0.23
P citric (ppm)	0.09	-0.66	-0.50	-0.09	-0.09	0.27
K	0.03	0.18	-0.27	0.10	0.17	0.07
Sand	0.58	-0.04	-0.12	-0.53	0.08	-0.11
Silt	-0.51	-0.33	0.04	0.38	-0.26	0.74
Clay	-0.41	0.35	0.13	0.45	0.12	-0.42

Figure 4 a. and b. shows the kriged maps for the multivariate variability. In TSF component, positive values represented high levels of sand, while in the PSF negative values associated with high levels of phosphorous. In pHF positive values associated with high levels of pH (Figure 4.c).



**Fig. 4:** Interpolated principal components. **a.** Principal component 1 of MULTISPATI-PCA for 2011-2012. **b.** Principal component 2 of MULTISPATI-PCA for 2011-2012. **c.** Principal component 2 of MULTISPATI-PCA for 2012-2013.

#### **2.5.4. Multiple regression**

Stepwise multiple regression analysis was performed not only to obtain the optimum model for predicting the rice yield performance also understand how the independent variables affect and interact with the dependent variable.

For 2011-2012 and for 2012-2013 the most appropriate models obtained with significance levels ( $p=0.01$  and  $p=0.0001$ ) were:

$$Y_i = \beta_0 - \beta_1 Pc - \beta_2 Ar - \beta_3 K + \beta_4 Pc*Ar + \beta_5 Pc*K + \beta_6 Ar*K - \beta_7 Pc*Ar*K$$



$$Y_i = -\beta_0 + \beta_1 \text{pH} - \beta_2 \text{Pc} + \beta_3 \text{Ar} - \beta_4 \text{pH} * \text{Ar} + \beta_5 \text{Pc} * \text{Ar}$$

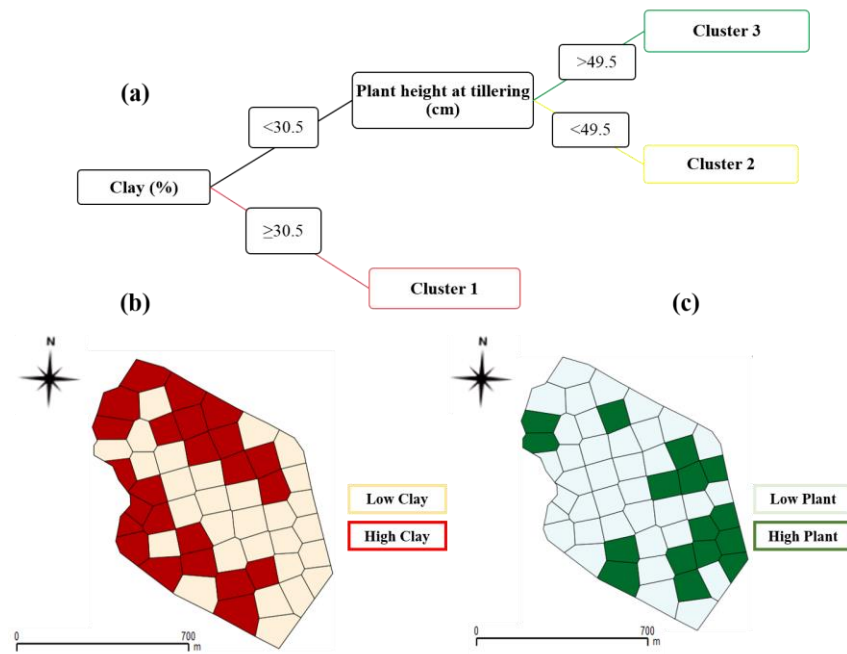
with  $R^2$  adjusted equal to 0.22 and 0.38 respectively and a correlation between predicted and observed values of 0.45 on both. While the settings of the models are not very high, they serve us as a tool to explain the factors that affect or interact with the performance. Since sand is present in both models, the texture of the soil is an important factors that affect or interact with rice yield. Also the P citric is present in both models and interact with sand in a positive contribution, suggesting the importance of the texture of soil and the availability of P for rice. Significance interaction of the models were studied (data not shown). In the first model, with low values of Pc, the yield were high with high values of sand and K and vice versa. In the second model, high values of pH and sand, and high values of Pc and sand gives high yields.

#### **2.5.5. Classification and regression tree analysis**

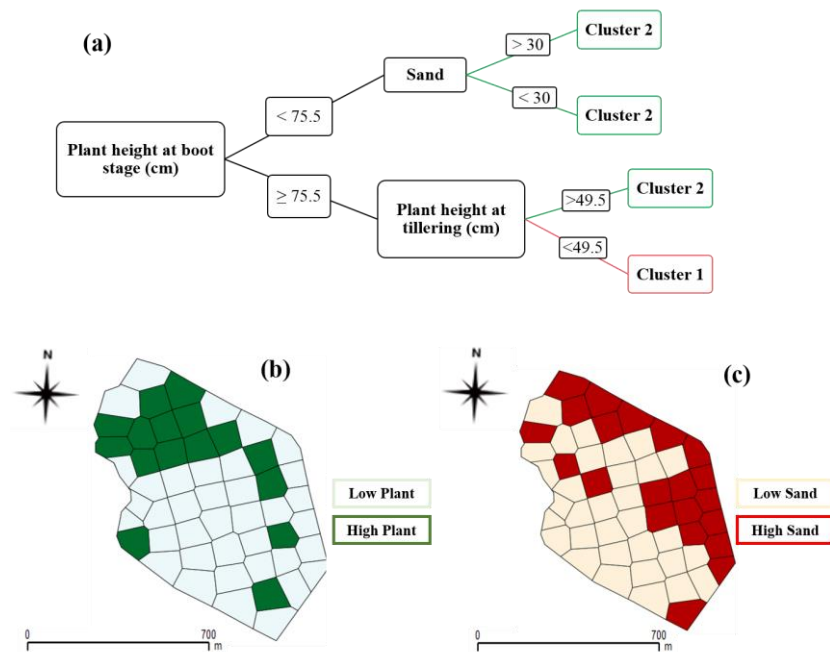
The objective of the CART analysis was to relate explanatory field-level factors to the cluster response values developed in the cluster analysis. Therefore, the stable yield clusters identified in the cluster analysis served as the categorical dependent variable in the CART analysis. Both 2011-2012 and 2012-2013 soil and crop measure data were included in the analysis.

In 2011-2012, for  $k=3$ , the first split resulted in a partition of field into points with “high Clay” and “low Clay”, corresponding to the low-yielding cluster 1 and middle and high-yielding cluster 2 and 3, respectively. Second partition variable was plant height at tillering. For  $k=2$  the tree, had 3 split dependent on plant height at boot stage, plant height at tillering and sand content (Figure 5 and 6).

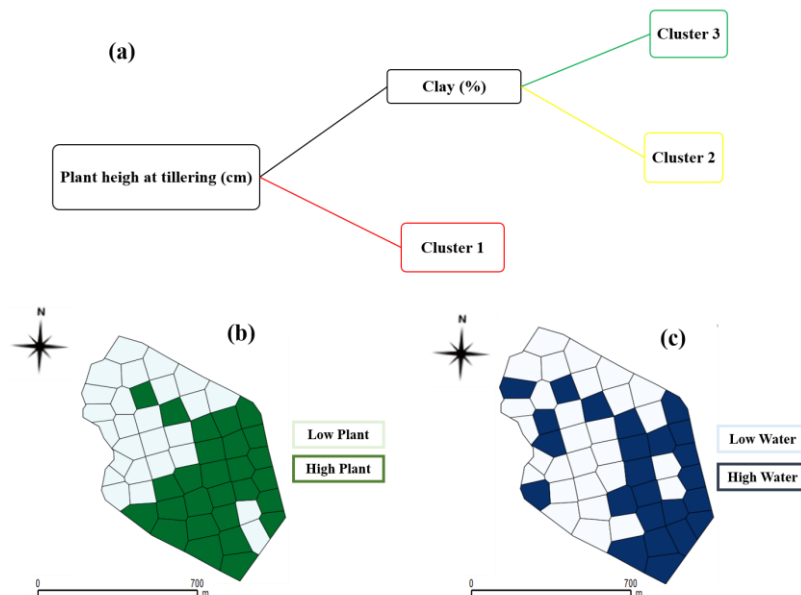
In 2012-2013, for  $k=2$  and  $k=3$ , CART analysis identified the same partition variables; plant heigh at tillering and clay content (Figure 7).



**Fig. 5:** **a.** Classification tree obtained for  $k=3$  in 2011-2012. **b. and c.** Maps of for the primary splitting variable, Clay (%) and the secondary splitting variable, Plant height at tillering (cm) drawn for each grid point using Thiessen polygons.



**Fig. 6:** **a.** Classification tree obtained for  $k=2$  in 2011-2012. **b. and c.** Maps the primary splitting variable, Plant height at boot stage (cm) and the secondary splitting variable, Sand (%) drawn for each grid point using Thiessen polygons.



**Fig. 7:** **a.** Classification tree obtained for  $k=3$  in 2012-2013. **b. and c.** Maps of the primary splitting variable, Plant height at tillering (cm) and the top surrogate splitting variable, height of the layer of water (cm) drawn for each grid point using Thiessen polygons.

## 2.6. DISCUSSION

Precision agriculture, site-specific soil and crop management (SSCM), are farming system that considers spatial and temporal variability in soil properties and crop productivity (Mulla and Schepers 1997). This farming system is expected to improve the efficiency and efficacy of inputs, which in turn may reduce environmental pollution and improve profit (Pierce and Nowak 1999). The concept of SSCM promotes the delineation of management zones that are similar based on certain quantitative measures and the application of appropriate management practices (Cassman et al. 2002). Wehlan and Mc Bratney (2000) proposed the Null Hypothesis of Precision Agriculture (NHPA) to test whether reducing the management scale of a given area would improve a set of predefined variables. For precision management to be beneficial, spatial variability must

be identified, persist temporally and its underlying causes be understood in order to be effectively treated (Plant 2001). The objective of this study was to use multivariate and exploratory statistical methods to identify yield spatial variability and factor underlying. The data collected in this research are observational, and are not appropriate for confirmatory statistics. However, this dataset allows to interpret how complex patterns of yield and edaphic properties change in space and interact with the crop performance.

### **2.6.1. Contrasting yield zones**

The field investigated exhibited degrees of in-field variability of yield with a high variation on yield among grid points (i.e. CV of grid point in 2011-2012= 14% and 15% in 2012-2013). Simmonds et al. (2013) reported a range of CV of 7.09% to 10.48% in yield variation within irrigated rice fields. Also others researches, but in smaller field sizes, reported a CV of 39% in yield variation (Dobermann 1994), CV of 5.89% (Yanai et al. 2001).

The algorithm k-means cluster analysis could identify two contrast and consistent zones of yield. Based on the cluster analysis results, the field exhibited a stable and spatially contiguous cluster yield pattern. The consistency in the possible cluster sets for k=2 indicates that the spatial patterns in this field have been influenced in large part by persistent endogenous factors. Zones with high-yielding level were the zones with high level of sand and P. In this field, areas with higher sand content, allowed a better and early emergence of the crop in both years. In addition, sandy soils have their nutrients more available to the crop. In irrigated rice, water is generally not a major growth-limiting factor and the availability of nutrients like P is one of the most important factors that determine yield concretion (Dobermann and Cassman 2002). On the other side, the spatial and temporal inconsistency in the cluster set k=3 and k=4 indicates that the some spatial patterns have been influenced by transient exogenous factors (weeds, microclimate). This was corroborated with the observations made during the crop cycle. Higher weed density occurring in an area with deeper water and poor emergence of the

crop. Also other zones had instable layer of water during all the crop cycle. Roel et al. (2007) and Simmonds et al. (2013), reported that in rice fields, management factors like weed density and water management had significant effect on rice yield variability.

### **2.6.2. Different statistical approaches for identified factors underlying yield**

New precision agriculture technologies allow collecting information from several variables at many georeferenced locations within crop fields and this information could be used to predict the variables in unsampled sites. However, univariate geostatistical analysis could not predict and represent the spatial variability of the soil properties, probably due to the set distance between grid points and the density of them. Doberman (1994) and Yanai et al. (2001) could describe the spatial variability of soil properties but with an intensive grid point (1 sample per 0.5 ha) in small sizes fields (1 ha). In rice production system this intensity sampling is impractical due to time, economic implications and the sizes of the fields (75-100 ha). Technologies of continue sampling like remote sensing or electrical conductivity sensors can be used to relieve a higher density of data and thus describe the spatial variability of soil. However, the spatial covariation of soil properties could be evaluated by multivariate spatial analysis (MSPCA). Nevertheless, MSPCA has not been explicitly developed for spatial data as other multivariate descriptive methods, and for rice crop system, this kind of analysis are not reported. The multivariate analysis could summarized covariation structures and identified spatial patterns. For both years, about 60% of soil variability was resume in two principal component. First component (TSF) was characterized by soil texture, and the zone with high sand content coincided with high-yielding cluster. Also this result match with CART analysis, which identified texture soil as primary partitioning variable. Experimental variogram and universal kriging of TSF, PSF and pHF components, allowed to make predictions of the variable at unsampled sites considering the spatial correlation structure. Through multivariate spatial variability maps, homogenous areas within the field were identified, key to the potential SSCM. High-

yielding cluster zone matched with a high-sand and P soil content. As Córdoba et al. (2012b) reported the multivariate spatial analysis is a promising statistics tool for the delimitation of homogeneous zones within the field.

In CART analysis, for both years and  $k=2$  and  $k=3$ , texture soil was identified as primary splitting variables, competitor and/or surrogate splitting variables. The key yield-limiting edaphic factors explaining the stable yield patterns found in both years; low and high yielding zones were the same. The lowest content of clay, allows a better and early emergence, resulting in a better development of the crop and in a highest harvest index. On the other side, transient factors like plant height in different growth stages and irrigated management indicated differences between low and high yield clusters. Highest plant height at tillering was identified as a primary splitting partitioning variable, denoting the importance of early emergence and good crop development to achieve good harvest index. However, highest plant height at boot stages corresponded to low yield cluster. This highest height was the result of a later emergence and a rapid development, culminating in an exuberant biomass production which is ultimately not translated into a high performance. As a surrogate and competitor variable, irrigated management was identified (Figure 8). Similar to finding Roel et al. 2005, yield variability may already be affected and exacerbated by water management. Low yielding zone had an instable layer of water throughout all crop cycle, mainly during boot stages where the yield is concretized. In Uruguay irrigation system are formed by low dykes, called *taipas* that are used to promote more even flooding. Thus, the success in the irrigated management depends indirectly on the land-leveling. As Walker et al. (2003) and Roel and Plant (2004b) observed, CART identified land-leveling effected on rice yield variability through the identification of layer water as a surrogate and competitor variable. Low-yielding cluster had high density of *taipas* that produce low potential yield (due to a worse weed management and low availability of nutrient, mainly nitrogen) and also affect the water management make it instable.

Multiple regression analysis identified the same variables as Multivariate spatial analysis and CART analysis identified. Significance interactions between soil variables,

indicate the complexity of the phenomena and the difficulty in studying yield-influencing factors. Through this statistical tools not only obtained the optimum model for predicting rice yield performance also we understand how the independents variables affect and interact with crop performance.

## **2.7. CONCLUSIONS**

The objective of this study was to define and integrate methodologies of processing spatial data collected at commercial field, to generate information as a starting point to a possible site-specific management of commercial rice farms. The fundamental difficulty in studying yield-influencing factors in commercial fields is the complexity of the phenomena. Analyzing the in-field spatial yield patterns, their stability and underlying causes is a critical first step in evaluating precision management in rice systems.

This study demonstrated that there are many techniques and statistics tools for analyzing this type of data (spatial, auto-correlated, and large volumes data). This different tools with different approaches shown consistent and robust results. None of the methodologies are better or worse, but complementary. Depending on the purpose of the study and the characteristics of the data set the appropriated tools is going to be selected.

The k-means algorithms carries out a successful randomization process and identify two different yield zones that would benefit from a site-specific management. Through the use of different statistics tools we arrived to similar and concordant results about the possible factors underlying rice yield. At the farm level through the use of yield monitors, consistent yield areas can be delineate and then they can be used for targeted sampling and site-specific management practices.



## **2.8. ACKNOWLEDGEMENT**

This work was funded by an Innovagro Project INIA-ANII and a Schollarsip awarded to C. Bonilla by the National Research and Innovation Agency. To the Farmers and Engineer of the field, who allows us to perform our work during two years. To Ing. Agr. Cristina Capurro and the staff of the Laboratory of soil physics and irrigation INIA-Treinta y Tres for their willingness and cooperation in the field work. To Ing. Agr. Mónica Cadenazzi, Ing. Agr. José Terra and Professor Richard Plant for their valuable contributions.

## **2.9. REFERENCES**

Basso, B., Bertocco, M., Sartori, L., & Martin, E. C. (2007). Analyzing the effects of climate variability on spatial pattern of yield in a maize-wheat-soybean rotation. *European Journal of Agronomy*, 26(2), 82-91.

Bivand, R., Altman, M., Anselin, L., Assuncao, R., Berke, O., & Bernat, A. (2010). Spdep: spatial dependence: weighting schemes, statistics and models. R package version 0.5-21. <http://CRAN.R-project.org/package=spdep>. Accessed 01 jul 2013.

Bocchi, S., & Castrignano, A. (2007). Identification of different potential production areas for corn in Italy through multitemporal yield map analysis. *ScienceDirect*, 102, 185-197.

Casanova, D., Goudriaan, J., Bouma, J., & Epema, G. F. (1999). Yield gap analysis in relation to soil properties in direct-seeded flooded rice. *Geoderma*, 91(3-4), 191-216.

Cassman, K. G. (1999). Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. *Proceedings of the National Academy of Sciences, USA*, 96(11), 5952–5959.

Cassman, k. g., Dobermann, A., & Walters, D. T. (2002). Agroecosystems, nitrogen-use efficiency, and nitrogen management, *Ambio*, 31, 132-140.

Córdoba, M., Balzarini, M., Bruno, C., & Costa, J.L. (2012a). Identificación de zonas de manejo sitio-específico a partir de la combinación de variables de suelo. *Corpoica*. 13(1): 47-54.

Córdoba, M., Balzarini, M., Bruno, C., & Costa, J.L. (2012b). Análisis de componentes principales con datos georreferenciados. *Facultad Uncuyo*. 44(1): 27-39.

Cressie, N.S.C. (1991). *Statistics for spatial data*. Nueva York: Wiley.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., & Weingessel, A. (2010). e1071: Misc functions of the department of statistics (e1071), TU Wien. R package version 1.5-24. <http://CRAN.R-project.org/package=e1071>. Accessed 01 Jul 2013.

Dobermann, A. (1994). Factors causing field variation of direct-seeded flooded rice. *Geoderma*, 62(1–3), 125–150.

Dobermann, A., & Cassman, K.G. (2002). Plant nutrient management for enhanced productivity in intensive grain production systems of the United States and Asia. *Plant and Soil*, 247, 153-175.

Dobermann A., Ping J.L., Adamchuk, V.I., Simbahan, G.C., & Ferguson, R.B. (2003). Classification of crop yield variability in irrigated production fields. *Agronomy Journal*, 95, 1105-1120.

Guastaferro, F., Castrignano, A., De Benedetto, D., Sollitto, D., Troccoli, A., & Cafarelli, B. (2010). A comparison of different algorithms for the delineation of management zones. *Precision Agriculture*, 11(6), 600–620.

Jayne, D.B., Kaspar, T.C., Colvin, T.S., & James, D.E. (2003). Cluster analysis of spatiotemporal corn yield patterns in an Iowa fields. *Agronomy Journal*, 95, 574-586.

Lark, R. M., & Stafford, J. V. (1997). Classification as a first step in the interpretation of temporal and spatial variation of crop yield. *Annals of Applied Biology*, 130(1), 111–121.

Mulla, D.J., & McBratney, A.B. (2002). Soil spatial variability. In: Warrick, A.W. (Ed.). *Soil physics companion*. (pp. 343-373). Boca Ratón: CRC.

Mulla, D. J., & Schepers, J. S. (1997). Key processes and properties for site-specific soil and crop management. In F. J. Pierce & E. J. Sadler (Ed.). *The state of site-specific management for agriculture* (pp. 211-244). CSSA, SSSA, Madison, WI: ASA.

Mzuku, M., Khosla, R., Reich, R., Inman, D., Smith, F., & MacDonald, L. (2005). Spatial variability of measured soil properties across site-specific management zones. *Soil Science Society of America Journal*, 69(5), 1572-1579.

Perez-Quezada, J. F., Pettygrove, G. S., & Plant, R. E. (2003). Spatial-temporal analysis of yield and the influence of soil factors in two four-crop-rotation fields in the Sacramento Valley, California. *Agronomy Journal*, 95(3), 676–687.

Pierce, F.J., & Nowak, P., (1999). Aspects of precision agriculture. In: Sparks, D.L. (Eds.). *Advances in agronomy*. Newark, Delaware, Academic Press. Pp. 1-85.

Plant, R.E. (2001). Site-specific management: the application of information technology to crop production. *Computers and Electronics in Agriculture*, 30, 9-29.

Ripley, B (2002), MASS: Support functions and datasets for Venables and Ripley's MASS. R package version 7.3-29. <http://CRAN.R-project.org/package=MASS>. Accessed 01 jul 2013.

Roel, A., Mutters, R. G., Eckert, J. W., & Plant, R. E. (2005). Effect of low water temperature on rice yield in California. *Agronomy Journal*, 97(3), 943–948.

Roel, A., & Plant, R. E. (2004a). Spatiotemporal analysis of rice yield variability in two California fields. *Agronomy Journal*, 96(1), 77–90.

Roel, A., & Plant, R. E. (2004b). Factors underlying yield variability in two California rice fields. *Agronomy Journal*, 96(5), 1481–1494.

Roel A., Firpo H., & Plant, R. E. (2007). Why do some farmers get higher yields? Multivariate analysis of a group of Uruguayan rice farmers. *Science Direct*, 58, 78-92.

Simmonds, M. B., Plant, R. E., Peña-Barragán J. M., van Kessel, C., Hill, J., & Linquist B. A. (2013). Underlying causes of yield spatial variability and potential for precision management in rice systems. *Precision Agriculture*, 14, 512-540.

Schroder, J.J., Neeteson, J.J., Oenema, O., & Struik, P.C. (2000). Does the crop or the soil indicate how to save nitrogen in maize production? Reviewing the state of the art. *Field Crop Research*, 66, 151-164.

Therneau, T. M., & Atkinson, B. (2009). R port by brian ripley. rpart: recursive partitioning. R package version 3.1-45. <http://CRAN.R-project.org/package=rpart>. Accessed 01 Jul 2013.

Walker, T. W., Kingery, W. L., Street, J., Cox, M. S., Oldham, J. L., Gerard, P. D., et al. (2003). Rice yield and soil chemical properties as affected by precision land-leveling in alluvial soils. *Agronomy Journal*, 95(6), 1483–1488.

Wang, Y., Shou-Hung, C., Kuo-Wei, C., Shen, Y. (2012). Identifying and characterizing yield limiting factors in paddy rice using remote sensing yield maps. *Precision Agriculture*, doi 10.1007/s1111901292665.

Wehlan, B. M., & McBratney, A. B. (2000). The “null hypothesis” of precision agriculture management. *Precision Agriculture*, 2(3), 265–279.

Yanai, J., Lee, C. K., Kaho, T., Iida, M., Matsui, T., Umeda, M., et al. (2001). Geostatistical analysis of soil chemical properties and rice yield in a paddy field and application to the analysis of yield-determining factors. *Soil Science and Plant Nutrition*, 47(2), 291–301.

### **3. COSECHA LOS BENEFICIOS: AGRICULTURA DE PRECISIÓN MÁS QUE SÓLO TECNOLOGÍA**

**Bonilla Camila, Roel Álvaro, Gutiérrez Lucía**

#### **3.1. SUMMARY**

Rice farmers have a growing interest in the application of precision agriculture technologies for productive purposes. The objective of this study was to develop a methodology for processing information relayed with yield monitors and its subsequent use as a starting point to a site-specific management. The spatiotemporal variability of the yield monitor data relayed for two years, in a commercial rice field of 100 ha. was quantified and characterized. The effect of agronomy management factors and the spatial distribution of soil properties, as the causes of the yield spatial variability was explored. An economic analysis with maps of net income was performed. The cluster analysis identify two zones of contrasting yield. Associated with them, differences in sand content, soil phosphorus, emergence, water depth, and weed control were found. Low-yield zone present average losses of 361 U\$S ha<sup>-1</sup> and high yield zone had average earnings of 280 U\$S ha<sup>-1</sup>. Our study demonstrate that the technologies associated with precision agriculture allow the study of the yield spatial variability in fields that a first glance seem uniforms. Two yield zones, which would be benefit from a site-specific management, were identified. Economic analysis allow the use of yield monitor data, as another tool for making decision on the farm.

**Key words:** yield monitors, cluster analysis, site-specific management, economic analysis, rice.

### **3.2. RESUMEN**

En la cadena arrocerera existe interés por la aplicación de tecnologías de agricultura de precisión con fines productivos. El objetivo del trabajo, fue desarrollar una metodología para el procesamiento de información relevada con monitores de rendimiento y su posterior utilización como punto de partida a un manejo sitio-específico. La variabilidad espacio-temporal de datos de monitor de rendimiento relevados durante dos zafras, en una chacra comercial de arroz de 100 ha, fue cuantificada y caracterizada. El efecto de los factores de manejo agronómico y la distribución espacial de las propiedades del suelo, como causas de la variabilidad espacial del rendimiento fue estudiado. Se realizó un análisis económico a partir de mapas de ingreso neto. El análisis de conglomerado, identificó 2 grupos de rendimiento contrastantes. Asociados a ellos, se encontraron diferencias en el contenido de arena y fósforo del suelo y diferencias en la implantación, lámina de agua y control de malezas. La zona de bajo rendimiento presentó pérdidas promedio de 361 U\$S ha<sup>-1</sup> y la zona de alto rendimiento ganancias promedio de 280 U\$S ha<sup>-1</sup>. Nuestro estudio demostró que las tecnologías asociadas a la agricultura de precisión permiten el estudio de la variabilidad espacial del rendimiento en chacras que a simple vista parecen uniformes. Se identificaron grupos de rendimiento, los cuales se verían beneficiadas con manejo sitio-

específico. El análisis económico permitió el uso de la información de los monitores de rendimiento, como una herramienta más para la toma de decisiones a nivel predial.

**Palabras claves:** monitores de rendimiento, análisis de conglomerados, manejo sitio-específico, análisis económico, arroz.

### **3.3. INTRODUCCIÓN**

La agricultura de precisión es la aplicación de tecnologías y principios para manejar la variabilidad espacial y temporal asociada a la producción agrícola con el propósito de mejorar la performance de los cultivos y la calidad ambiental (Pierce y Nowak, 1999). Por su parte, el manejo sitio-específico es el manejo de los cultivos agrícolas a escalas espaciales menores que toda la chacra. Para que el mismo sea justificado se deben cumplir tres principios básicos: debe existir suficiente variabilidad espacial en factores que afecten el rendimiento, la/s causa/s de esa variabilidad deben ser identificadas y medidas, y la información surgida de esas medidas debe poder ser usada para modificar el manejo de los cultivos, para incrementar las ganancias o disminuir el impacto ambiental (Plant, 2001). La alta variabilidad espacial de las condiciones ambientales pueden afectar los rendimientos de los cultivos (Mulla y McBratney, 2002) por lo que su identificación brinda la posibilidad de dividir a las chacras en regiones de baja, media y alta productividad potencial (Doerge, 1999). Para la estimación o



cuantificación de la variabilidad espacial de los suelos y cultivos en las chacras, existen métodos de medición discretos (Ej. muestreo de suelos y plantas), continuos (Ej. monitores de rendimiento) y remotos (Ej. imágenes satelitales). Los monitores de rendimiento son dispositivos que periódicamente miden el flujo del material cosechado y computan un valor georreferenciado de rendimiento del cultivo (Plant, 2001). A partir de estas estimaciones, se construyen mapas de rendimiento que son una importante herramienta para inferir las características ambientales y el potencial de productividad de los suelos (Stafford et al., 1998). Con ellos, los agricultores pueden identificar las áreas dentro del campo donde los rendimientos pueden ser mejorados, o áreas donde se necesitan ajustes en los insumos para optimizar la rentabilidad de las explotaciones y la calidad del medio ambiente (Pierce et al., 1997). Debido a que los mapas de rendimiento son una herramienta básica para el manejo de insumo y la mejora de la rentabilidad a nivel predial, los mismos son esenciales para la implementación del manejo sitio-específico (Roel y Plant, 2001a).

La interpretación de las causas de la variabilidad requiere un profundo conocimiento de las condiciones ambientales (Schroder et al., 2000). El análisis de conglomerados puede proporcionar una cuantificación objetiva de la estructura espacial de los patrones de rendimiento, así como permitir una identificación de la consistencia de estos patrones año en año (Perez-Quezada et al., 2003). A partir de ellos se podrían definir diferentes zonas de rendimiento del cultivo, con similares capacidades de respuesta a los insumos y reducir el impacto ambiental (Bocchi y Castignano, 2007). El reconocimiento automatizado de patrones, por medio de la agrupación multivariable, se

propone como una herramienta para la interpretación de la variación temporal y espacial de rendimiento de los cultivos (Lark y Stafford, 1997).

Los trabajos realizados en Uruguay, han buscado identificar los principales factores que afectan la variación de los rendimientos dentro de las chacras principalmente en cultivos de secano y la posibilidad de delinear ambientes o zonas de manejo en base a diferentes capas de información (Terra et al., 2006; Pravia, 2009). Sin embargo, a nivel nacional existe una brecha entre la posibilidad de acceso, por parte de los productores, a tecnologías vinculadas a la agricultura de precisión (GPS, monitores de rendimiento, muestreos dirigidos georreferenciados) y la utilización de esas tecnologías para la generación de información y conocimiento. En la cadena arrocera existe un creciente interés por la aplicación de tecnologías de agricultura de precisión con fines productivos. Poder delimitar ambientes con diferente potencial productivo dentro de chacras de arroz, sería un buen principio para comenzar a desarrollar el manejo sitio-específico en este cultivo a nivel nacional.

El objetivo general de este trabajo, fue desarrollar una metodología confiable para el procesamiento de información relevada con monitores de rendimiento y su posterior utilización como punto de partida a un posible manejo sitio-específico en chacras comerciales de arroz. Como primer objetivo específico se planteó evaluar y cuantificar la variabilidad espacial del rendimiento de una chacra comercial de arroz. Como segundo objetivo específico, se buscó mediante herramientas de agrupamiento, delinear diferentes áreas de rendimiento. El tercer objetivo específico fue plantear un uso práctico

de la información brindada por los mapas de rendimiento, a través de un análisis económico de las zonas de rendimiento delineadas.

### **3.4. MATERIALES Y MÉTODOS**

El estudio fue llevado a cabo durante las zafas de 2011-2012 y 2012-2013 en una chacra comercial de arroz de 100 ha ubicada en la localidad de Rincón, Uruguay (32°50'12.14" latitud sur- 53°48'22.91"longitud este). Los suelos que componen la chacra pertenecen a la Unidad de Rincón de Ramírez y Río Branco. La variedad cultivada fue El Paso 144 y se utilizaron medidas estándar de manejo del cultivo (guía de buenas prácticas en el cultivo de arroz en Uruguay).

#### **3.4.1. Muestro discreto**

Dentro de la chacra se implementó una grilla de muestreo sistemática no alineada, compuesta por 53 puntos (Figura 1) donde se realizaron mediciones en pre-siembra, durante el ciclo del cultivo y en cosecha. En pre-siembra se realizó un muestreo de suelo, donde cada muestra estuvo compuesta por 6 submuestras (0-15 cm de profundidad) y los análisis determinados fueron, para los 2 años de estudio: Fósforo Bray (ppm, Pb), Fósforo cítrico (ppm, Pc), Potasio intercambiable (meq 100 g<sup>-1</sup> de suelo, K) y textura de suelo (% Arena, Arcilla y Limo) y en la zafra 2012-2013 también se

analizó el pH. Para una visión global de la evolución del cultivo y la variabilidad del mismo dentro de la chacra, cada 15 días se evaluó, en cada uno de los puntos, implantación, altura del cultivo (cm), altura de la lámina de agua (cm) y se describió visualmente el nivel de enmalezamiento. A fin de ciclo del cultivo, se realizó una cosecha manual y se obtuvo un valor de rendimiento estimado ( $\text{kg ha}^{-1}$ ) para cada punto. El área cosechada fue de  $2,04 \text{ m}^2$  y las muestras fueron trilladas, pesadas y a su vez se midió la humedad (%).

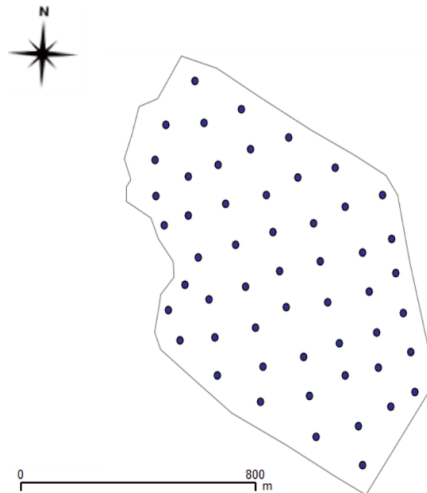


Figura 1: Ubicación puntos de muestreo discreto

### **3.4.2. Monitor de rendimiento**

El arroz fue cosechado con una cosechadora (John Deere) equipada con monitor de rendimiento Ag Leader con sistema de posicionamiento global diferencial (DGPS). Antes de la cosecha el monitor de rendimiento fue calibrado según las instrucciones de

fabricación y durante la misma, la cosechadora siguió un patrón concéntrico de cosecha. El rango de velocidad fue entre 1 a 2 m s<sup>-1</sup>, el ancho de cosecha fue de 5,5 m, y el flujo medido en kg s<sup>-1</sup> fue medido cada 2 segundos.

Los archivos obtenidos con los datos de rendimiento (coordenadas geográficas, velocidad, flujo, humedad, rendimiento) fueron importados al sistema de información geográfico QGis 1.8. (Open Source Geospatial Foundation, OSGeo) como un archivo de puntos (cada punto representaba un valor de rendimiento). Cada valor de rendimiento fue llevado a un rendimiento con 13% de humedad. El conjunto de datos posee entre 50000 a 90000 valores de rendimiento. Con el objetivo de quitar el mayor “ruido” posible en los datos recolectados por el monitor, los mismos fueron analizados siguiendo la metodología propuesta por Ping y Dobermann, 2005. Los puntos que se encontraban cercanos a los bordes y que estaban asociados a arranques y paradas de la cosechadora, fueron borrados para eliminar el efecto borde. Datos de rendimientos extremos (<1000 kg ha<sup>-1</sup> >15000 kg ha<sup>-1</sup>) fueron omitidos. Los datos de rendimiento resultantes fueron analizados con un análisis de dispersión de Moran utilizando el paquete de R spdep (Bivand et al., 2010) (Software estadístico R, versión 2.10.1).

### **3.4.3. Análisis geoestadístico**

#### **3.4.3.1. Muestreo discreto**

Se utilizaron semivariogramas para analizar la variabilidad espacial de las propiedades de suelo del muestreo discreto en forma individual. La tendencia espacial de

gran escala fue separada de los datos utilizando modelos de regresión (Cressie, 1991). Todos los variogramas y los modelos a ajustar fueron realizados con el Paquete de R gstat (Pebesma, 2004). Diferentes modelos de ajuste a los datos fueron testeados pero ningún modelo pudo ser ajustado. Por lo tanto, no se pudieron obtener mapas de variabilidad espacial para las propiedades del suelo a través de la interpolación. Por otro lado se realizaron Polígonos de Thiessen para tener una visualización de la distribución espacial de todas las variables muestreadas en el muestreo discreto (suelo, planta y prácticas de manejo). Este análisis fue realizado con QGis 1.8. (Open Source Geospatial Foundation, OSGeo).

#### **3.4.3.2. Monitor de rendimiento**

La capa de puntos de rendimiento fue interpolada a una grilla fija de 30-30 m utilizando kriging del vecino más próximo. La tendencia de gran escala y pequeña escala fue removida utilizando modelos de regresión lineal (Cressie, 1991). Una tendencia lineal en x e y fue incluida en el modelo. Para la caracterización de la variación a pequeña escala, los variogramas experimentales fueron calculados utilizando los residuales de rendimiento obtenidos del modelo de regresión. Se asumió una distribución normal de los datos y se chequeó utilizando Q-Q plots y Test de Shapiro. Un variograma teórico fue ajustado a cada variograma experimental y diferentes modelos de ajuste fueron testados. El modelo esférico isotrópico fue el de mejor ajuste

en los 2 casos. Todos los variogramas computados y los modelos de ajuste fueron realizados en el Paquete de R gstat.

#### **3.4.4. Análisis de conglomerado**

Con el objetivo de delinear zonas con diferente potencial productivo, se realizaron análisis de conglomerados de los datos de rendimiento del monitor resultantes de la interpolación. El método utilizado fue k-medias. En este algoritmo, k puntos son seleccionados como semillas de los grupos y luego cada grupo es formado asignando todos los demás puntos a la semilla más cercana. La media de cada aglomeramiento es luego seleccionada como el nuevo set de k semillas. El proceso se repite iterativamente. Se realizaron grupos para dos set de datos; uno que contenía a los 2 años evaluados y otro set compuesto por los datos de rendimiento para cada año por separado. Para comprobar que los patrones fueran espacio-temporalmente estables, los grupos fueron contruidos a partir de diferentes conjuntos de semillas iniciales y se hicieron 100 réplicas para cada k-grupos. A su vez para chequear que los grupos estuvieran espacialmente organizados y no depender de la visualización, se utilizó el estadístico  $\gamma$ , que mide la consistencia interna de los grupos:

$$\gamma = \frac{\sum_{i=1}^p \beta_i - \frac{p}{n^k}}{p - \frac{p}{n^k}}$$

Suponiendo que existen  $p$  datos localizados, el set bajo consideración tiene  $n$  replicaciones y un indicador  $\beta_i$ ,  $i=1, \dots, n$  que se define para cada ubicación  $i$ , siendo  $\beta_i=1$  si todos los miembros del set  $i$  pertenecen al mismo grupo ó  $\beta_i=0$  si al menos una locación del set  $i$  pertenece a un grupo diferente. El valor  $p/n^k$  es el valor esperado de celdas que pertenecen al mismo grupo por chance. Si todos los miembros del set son idénticos, entonces  $\gamma=1$ , por otro lado si todas las celdas de los miembros del set son asignados al azar, entonces  $\gamma=0$ . Los valores  $\gamma$  se computaron para cada grupo y luego se testeó la autocorrelación espacial con el coeficiente de Moran como un indicador de los patrones espaciales.

Con el set de datos compuesto por los rendimientos para cada año y para el conjunto de los 2 años, se realizaron mapas de Polígonos de Thiessen para una mejor interpretación visual de los resultados.

### **3.4.5. Análisis económico**

Se realizó un análisis económico de la siguiente manera. En primer lugar, se realizaron mapas de ingreso neto (U\$S  $ha^{-1}$ ), donde los valores del monitor de rendimiento se multiplicaron por el precio de la bolsa de arroz (U\$S  $bolsa^{-1}$ ) para los respectivos años, y a este valor se le restaron los costos de producción (U\$S  $ha^{-1}$ ) (Cuadro 1). Esto se realizó para las 2 zafas estudiadas. En segundo lugar, se realizó el mismo mapa resaltando los valores de ingresos positivos y negativos, utilizando para esto el rendimiento equilibrio en el cual los ingresos y egresos se hacen 0. Por último, se



realizó un mapa económico representando el valor promedio de producción de los grupos de rendimientos identificados por el análisis de conglomerados. Se utilizaron Polígonos de Thiessen para representar los mapas realizados en QGis 1.8.

Cuadro 1: Costos de producción de arroz (U\$S ha<sup>-1</sup>) y precio de la bolsa de arroz (U\$S bolsa<sup>-1</sup>) para las zafras 2011-2012 y 2012-2013

Rubro	2011-2012 U\$S/ha	2012-2013 U\$S/ha
Glifosato en pre-siembra	25	28
Gasoil preparación de tierras y siembra	162	167
Fertilizante basal	90	92
Semilla	82	86
Agrimensor	10	12
Herbicidas post-emergentes	66	78
Urea	60	56
Funguicida	32	34
Aplicaciones aéreas	83	89
Gasoil cosecha	90	93
Fletes	145	145
Lubricantes	25	28
Personal	145	164
Reparaciones de maquinaria	98	107
Secado	115	122
Imeba-Inia-Mevir	14	15
Rentas (agua y tierra)	427	439
Administración	88	96
Interés de financiación	34	41
Imprevistos	59	68
Total gastos	1850	1960
Precio arroz 50 kg (U\$S bolsa <sup>-1</sup> )	12,93	13,3
Rendimiento promedio (kg ha <sup>-1</sup> )	7643	7590
Rendimiento equilibrio (kg ha <sup>-1</sup> )	7170	7259

### 3.5. RESULTADOS Y DISCUSIÓN

#### 3.5.1. Estadística descriptiva del rendimiento y las propiedades de suelo

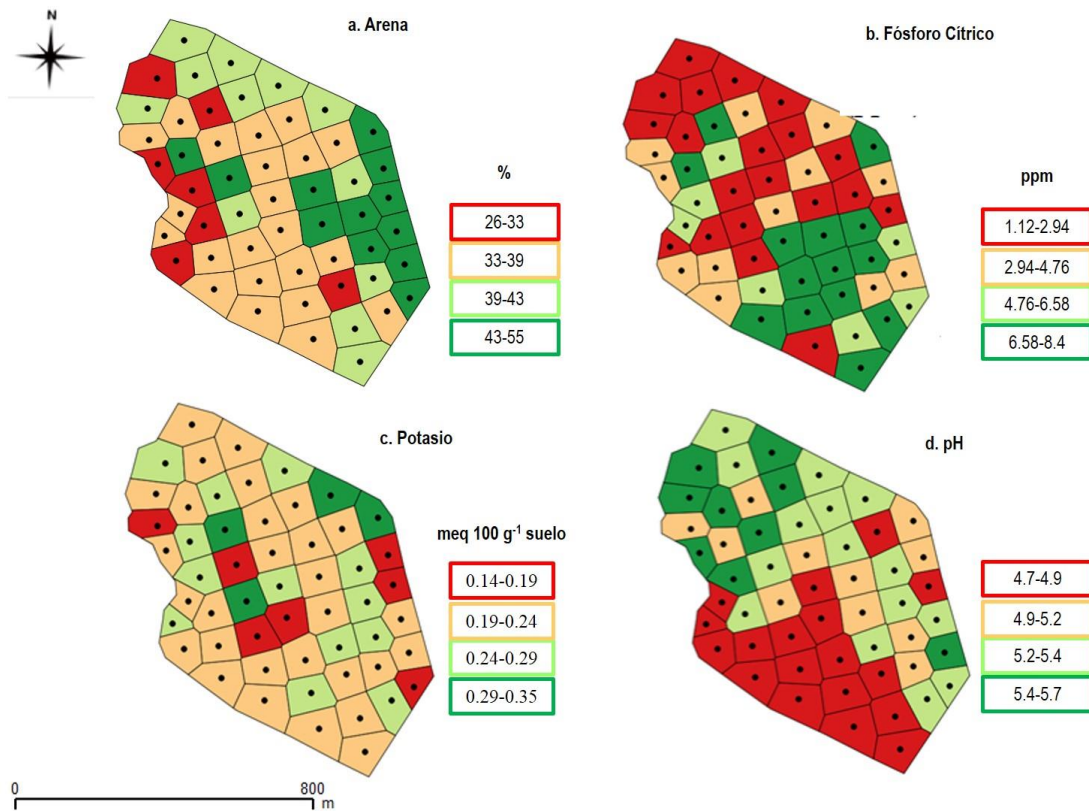
La chacra exhibió un grado de variación de rendimiento alto (i.e. CV de la grilla 2011-2012= 14% y 15% en 2012-2013) y un promedio cercano a la media nacional (Cuadro 2). Simmonds et al., (2013) reportaron un rango de CV de 7,09% a 10,48% en la variación del rendimiento de un cultivo de arroz irrigado. Otros investigadores, en chacras de menor tamaño, reportaron un CV de 39% (Dobermann, 1994), CV de 5,89% (Yanai et al., 2001).

Cuadro 2: Estadística descriptiva de rendimiento (kg ha<sup>-1</sup>) para 2011-2012 y 2012-2013

Año	Rendimiento en kg ha <sup>-1</sup>				
	Mínimo	1er Cuartil	Media	3er Cuartil	Máximo
2011-2012	963	5511	7643	9771	13500
2012-2013	1160	6400	7370	8510	13090

En cuanto a las propiedades del suelo, Pb y Pc fueron las más variables (CV 48,97-76,37 %). K, Arena, Limo y Arcilla exhibieron una variación moderada (CV 12,56-20,02 %), mientras que el pH fue el menos variable (CV 5,96 %). Los mapas

resaltan una zona de alto porcentaje de arena y alto Pc y otra zona con bajos valores de pH (Figura 1).



**Figura 1:** Polígonos de Thiessen de **a.** Contenido de arena del suelo (**%**) **b.** Fósforo cítrico (ppm) en suelo **c.** Potasio en suelo (meq 100 g<sup>-1</sup> de suelo). **d.** pH del suelo.

### **3.5.2. Análisis de conglomerados: Delimitación de zonas de rendimiento**

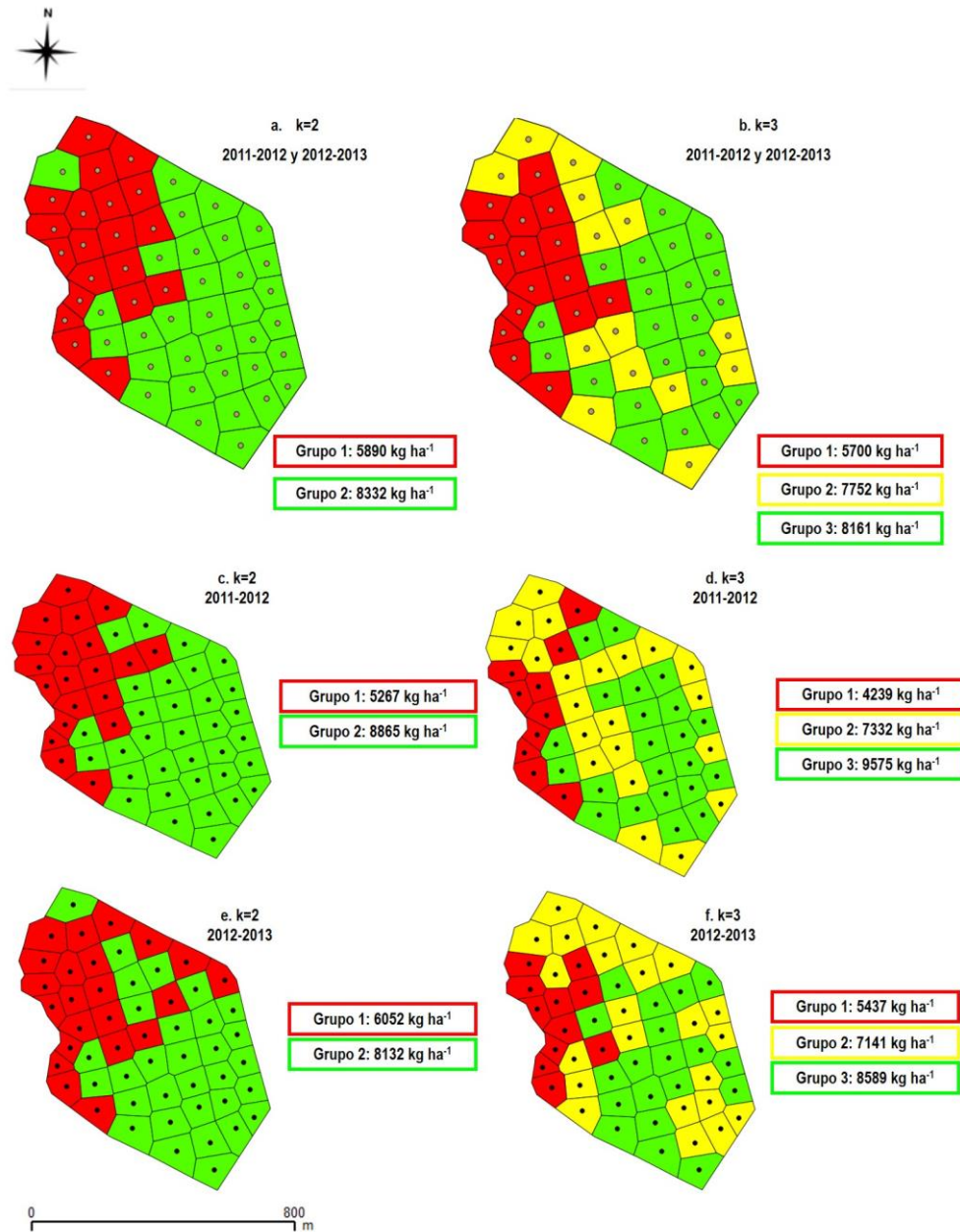
El análisis de conglomerados, presentó estabilidad al aleatorizar las semillas iniciales de cada cluster y los patrones permanecieron constantes, luego de realizar 100

réplicas. A su vez, los conglomerados presentaron correlación espacial significativa, evaluada a través del Índice de Moran (Cuadro 3). Los mismos podrían entonces ser analizados como zonas de diferente rendimiento.

Cuadro 3: Número de k-medias y Coeficiente de Moran y p-valor de significancia.

K-medias	Coeficiente de Moran
2	0,464 (0,001)
3	0,412 (0,001)
4	0,366 (0,001)

Los mapas de Polígonos de Thiessen que representan los grupos para cada año, resaltan la relación y contigüidad espacial entre sus miembros (Figura 2). El óptimo valor de k-medias para cada año y para los dos años evaluados a la vez, fue k=2. El grupo 1 está compuesto por zonas de bajo rendimiento promedio (5890 kg ha<sup>-1</sup>) y los integrantes del grupo 2 son zonas de alto rendimiento promedio (8332 kg ha<sup>-1</sup>) en los 2 años.



**Figura 2:** Conglomerados obtenidos para los datos de rendimiento interpolados representados con Polígonos de Thiessen para: **a.** k=2 para 2011-2012 y 2012-2013 **b.** k=3 para 2011-2012 y 2012-2013 **c.** k=2 para 2011-2012 **d.** k=3 para 2011-2012 **e.** k=2 para 2012-2013 **f.** k=3 para 2012-2013.

Varias técnicas de aglomeramiento han sido utilizadas con el fin de reducir la complejidad espacial del fenómeno. Lark y Stafford (1997) utilizaron fuzzy-cluster para organizar datos de mapas de rendimiento de diferentes cultivos. Empleando el algoritmo de k-medias se han podido identificar zonas de rendimiento contrastantes con similar comportamiento espacio-temporal (Jayne et al., 2003; Perez-Quezada et al., 2003). A través de un análisis de contigüidad espacial se ha observado la variabilidad espacio-temporal del rendimiento (Roel y Plant, 2004a). En este estudio, el algoritmo k-medias pudo identificar 2 zonas contrastantes y consistentes de rendimiento para los 2 años. Basado en los resultados del análisis de conglomerados, la chacra exhibió un patrón de rendimiento estable y espacialmente correlacionado. La consistencia encontrada indica que los patrones de rendimiento en la chacra están influenciados principalmente por factores persistentes endógenos de la chacra (tipo y textura de suelo, nutrientes). Las zonas de alto rendimiento fueron las zonas con mayor contenido de Pc y arena. A su vez los suelos de esta zona pertenecen a la Unidad de suelo de Rincón de Ramírez, mientras que las zonas de bajo rendimiento pertenecen a la Unidad Río Branco. En esta chacra, las zonas de suelo arenoso permitieron una mayor y más temprana implantación del cultivo en los 2 años. A su vez, estos suelos con mayor contenido de arena tienen sus nutrientes más disponibles para el cultivo. En sistemas de arroz irrigados, el agua no es un limitante por lo que la disponibilidad de nutrientes como el P es uno de los factores más importantes para la concreción de altos rendimientos (Dobermann y Cassman, 2002). Por otro lado, las inconsistencias tempo-espaciales asociadas a los grupos con  $k=3$  y 4, indican que en algunos puntos de la chacra, los patrones de rendimiento

también estuvieron influenciados por factores exógenos transitorios como enmalezamiento y láminas de riego inestables. No solo la altura de la lámina de agua es un factor clave para concretar altos rendimientos, sino que también el movimiento del agua en la chacra y el microclima que se genera en los cuadros de riego. La zona de bajo rendimiento se encuentra en el lado opuesto al canal principal de riego, lo que indica que el agua llegaba a esta zona de la chacra en última instancia. Roel et al., (2007) y Simmonds et al. (2013), reportaron que en chacras de arroz, los factores de manejo tienen un efecto significativo en la variación espacial del rendimiento.

### **3.5.3. Uso práctico de las herramientas de agricultura de precisión**

Técnicas asociadas a la agricultura de precisión proveen la oportunidad de mapear la variabilidad espacial y temporal asociada a la producción agrícola de manera de mejorar la performance de los cultivos, reducir el impacto ambiental y/o mejorar la rentabilidad del sistema productivo (Pierce y Nowak, 1999). El resultado de los mapas de rendimiento y su posterior aglomeración en zonas de rendimientos contrastantes para los 2 años, remarcó 3 ideas importantes. La primera es que una chacra de arroz que parece uniforme a simple vista, puede tener una variabilidad de rendimiento importante a lo largo de toda la chacra. Evento que ha sido reportado en otras investigaciones (Doberman, 1994; Yanai, 2001; Roel y Plant, 2004; Simmonds et al., 2013). La segunda es que mapear la variabilidad puede ayudar a analizar e interpretar los factores bióticos o abióticos que subyacen en la variabilidad del rendimiento. Terra et al., (2006), para

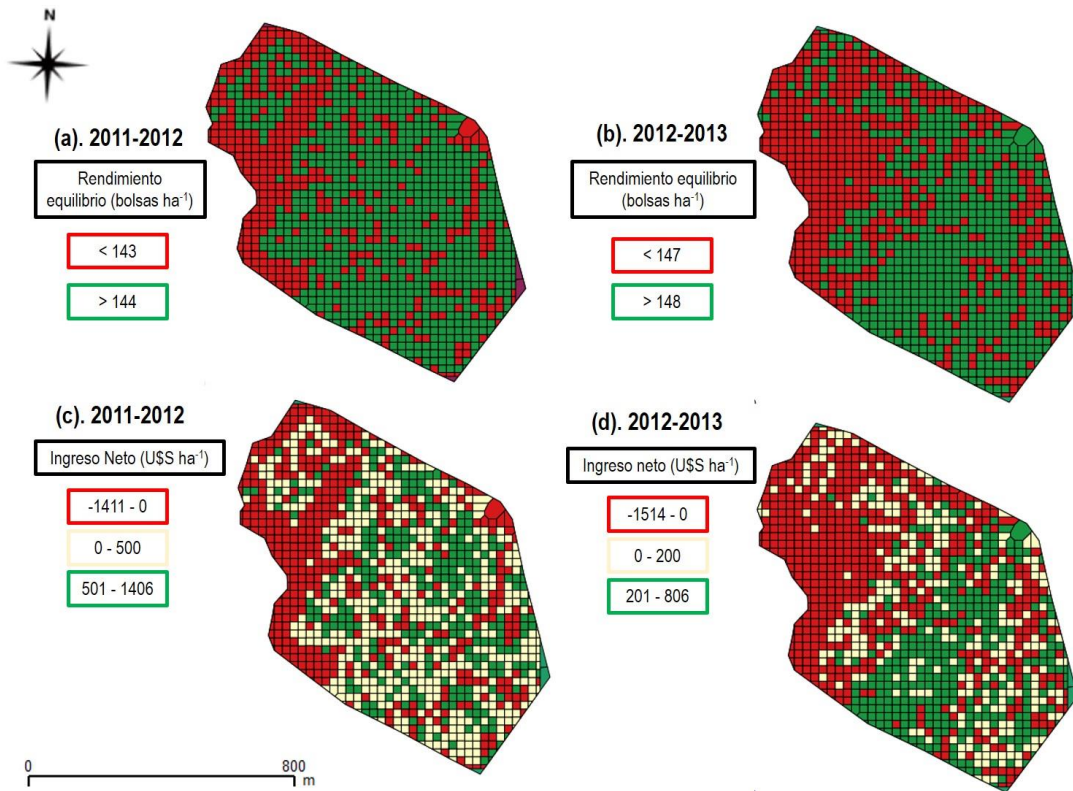
cultivos de secano encontraron que los factores asociados a la variabilidad del rendimiento fueron los atributos topográficos que están relacionadas con la dinámica y capacidad de almacenamiento de agua en el suelo. La tercera es que esta información puede ser utilizada para delinear zonas que se beneficiarían con un manejo sitio-específico, sería una fuente de información para delinear muestreos dirigidos y/o realizar un análisis económico como una herramienta más en la toma de decisiones.

Una de las mejores maneras de analizar los datos que proporcionan los monitores de rendimiento, es a través de los sistemas de información geográficos, que pueden ser utilizados no solo para una interpretación visual de los mapas, sino también estadística. A pesar que el rendimiento promedio de la chacra se aproxima al promedio nacional, existe una zona dentro de la misma, constante para los 2 años, con bajos rendimientos y representa aproximadamente el 50% del área total de producción (Figura 2). Asociada a esta zona de bajo rendimiento se encontraron suelos con mayor contenido de arcilla y menor  $P_c$  disponible. Durante las 2 zafas, esta zona de la chacra presentó una implantación despareja y más tardía que el resto de la chacra repercutiendo luego en el desarrollo del cultivo durante todo el ciclo. Esta área presentó un mayor enmalezamiento, lámina inestable de agua y planta con desarrollo excesivo sin concreción en el rendimiento final. Por lo tanto, si examinamos las causas asociadas a la variabilidad del rendimiento, concluimos que el principal factor es un factor permanente y endógeno de la chacra, como la textura del suelo que repercute luego en los factores transitorios mencionados. Las 2 zonas con variabilidad de rendimiento, en conjunto con los factores asociados a esa variabilidad podrían ser candidatas a un manejo sitio-



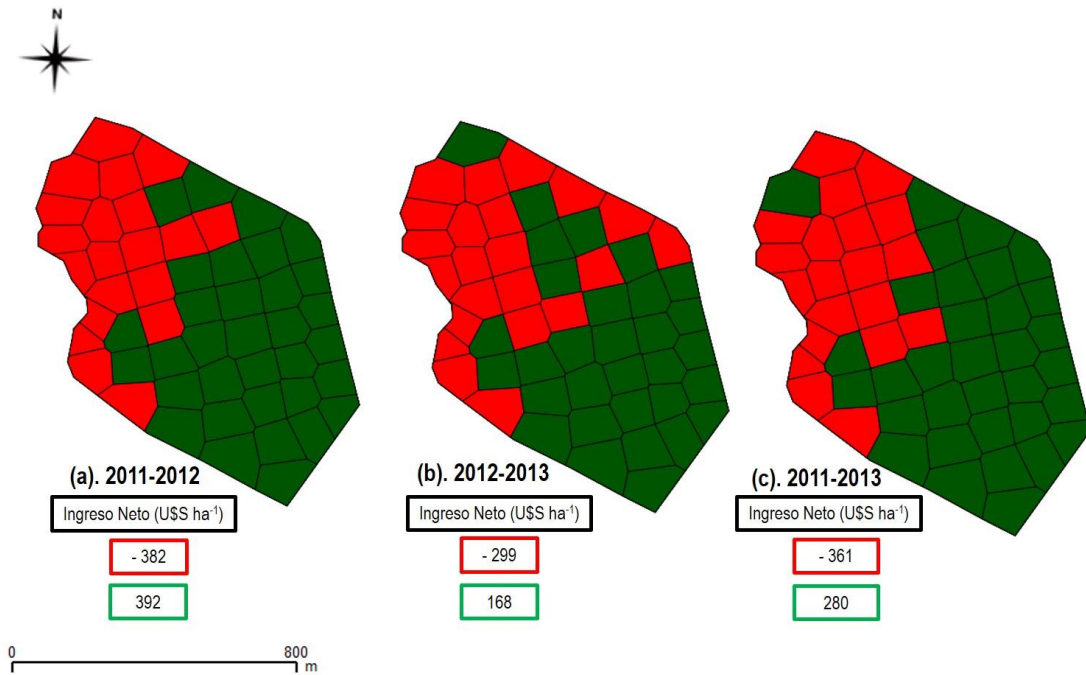
especifico diferencial. Uno de problemas identificados en el área de bajo rendimiento fue la mala y tardía implantación del cultivo para los 2 años, producto de una siembra sobre suelo arcilloso muy húmedo. Se podría pensar, para esta zona (50 ha) la utilización de variedades de ciclo más corto (INIA Tacuarí) que permitan mayor flexibilidad en la fecha de siembra. También podría evaluarse un manejo diferencial de nutrientes, sobretodo de P debido a que se describieron zonas con niveles de P diferenciales y que a su vez coincidieron con las zonas de rendimiento diferenciadas. Se deberían entonces realizar ensayos en fajas de respuesta a nutrientes teniendo en cuenta el suelo como factor de bloqueo para verificar estas hipótesis.

Por otro lado, el factor por el cual finalmente un productor toma las decisiones es la ganancia monetaria que le deja la actividad. Analizando los datos de los monitores de rendimiento se observa que en los 2 años de estudio, la mayor concentración de puntos con rendimientos que determinaron márgenes negativos, se encuentra en la zona de bajos rendimientos identificada por el análisis de conglomerados (Figura 3). Estos puntos expresan valores de pérdidas de hasta 1500 U\$S ha<sup>-1</sup>.



**Figura 3:** Polígonos de Thiessen para **a.** Rendimiento equilibrio (bolsas  $\text{ha}^{-1}$ ) 2011-2012. **b.** Rendimiento equilibrio (bolsas  $\text{ha}^{-1}$ ) 2012-2013. **c.** Ingreso Neto (US\$  $\text{ha}^{-1}$ ) para 2011-2012. **d.** Ingreso Neto (US\$  $\text{ha}^{-1}$ ) para 2012-2013.

Al resumir la información a las zonas identificadas por el análisis de conglomerado, observamos que la zona de bajo rendimiento tuvo pérdidas de 382 y 299 US\$  $\text{ha}^{-1}$  contra ganancias de 392 y 168 US\$  $\text{ha}^{-1}$  para 2011-2012 y 2012-2013, respectivamente. Si tomamos en cuenta los 2 años en conjunto, la zona de bajo rendimiento representó pérdidas de 361 US\$  $\text{ha}^{-1}$  contra ganancias de 280 US\$  $\text{ha}^{-1}$  encontradas en las zona de alto rendimiento (Figura 4).



**Figura 4:** Polígonos de Thiessen para grupos de rendimiento **a.** Ingreso Neto (US\$ ha<sup>-1</sup>) 2011-2012. **b.** Ingreso Neto (US\$ ha<sup>-1</sup>) 2012-2013. **c.** Ingreso Neto (US\$ ha<sup>-1</sup>) 2011-2012 y 2012-2013.

### 3.6. CONCLUSIÓN

Las tecnologías asociadas a la agricultura de precisión permiten el estudio de la variabilidad espacial del rendimiento en chacras que a simple vista parecen uniformes.

Mediante el análisis de la variabilidad espacial y su agrupamiento, se pueden identificar zonas de diferentes rendimientos dentro de la chacra, las cuales se verían beneficiadas con un posible manejo sitio-específico. Estas zonas pueden ser a su vez asociadas con factores bióticos y abióticos del ambiente y del manejo agronómico que causan esa variabilidad.

Un análisis económico sencillo, con mapas de ingreso neto por zonas, permite poner en práctica el uso de la información que brindan los monitores de rendimiento como una herramienta más para la toma de decisiones a nivel predial.

### 3.7. AGRADECIMIENTOS

Este trabajo fue financiado con un proyecto innovagro INIA-ANII: Manejo sitio-específico en cultivos agrícolas y una beca ANII de Maestría. Al productor y al Ing. Agr. Responsable de la chacra, por dejarnos realizar nuestro trabajo en su campo durante dos años. A Ana García y al Ing. Agr. Raúl Úraga por proporcionarnos información referente al rubro.

### 3.8. BIBLIOGRAFÍA

- Bocchi S, Castrignano A. 2007. Identification of different potential production areas for corn in Italy through multitemporal yield map analysis. *ScienceDirect*, 102: 185-197.
- Cressie NSC. 1991. Statistics for spatial data. Wiley: Nueva York. 928p.
- Doerge TA. 1999. Yield map interpretation. *Production Agriculture*. 12:54-61.
- Dobermann A. 1994. Factors causing field variation of direct-seeded flooded rice. *Geoderma*, 62(1-3): 125-150.
- Dobermann A, Cassman KG. 2002. Plant nutrient management for enhanced productivity in intensive grain production systems of the United States and Asia. *Plant and Soil*, 247: 153-175.
- Dobermann A, Ping JL, Adamchuk VI, Simbahan GC, Ferguson RB. 2003. Classification of crop yield variability in irrigated production fields. *Agronomy Journal*, 95: 1105-1120.

- Jayne DB, Kaspar TC, Colvin TS, James DE. 2003. Cluster analysis of spatiotemporal corn yield patterns in an Iowa fields. *Agronomy Journal*, 95: 574-586.
- Lark RM, Stafford JV. 1997. Classification as a first step in the interpretation of temporal and spatial variation of crop yield. *Annals of Applied Biology*, 130(1): 111–121.
- Mulla DJ, McBratney A.B. 2002. Soil spatial variability. En: Warrick, A.W. (Eds.). *Soil physics companion*. Boca Ratón. (CRC). pp. 343-373.
- Pebesma EJ. 2004. Multivariate geostatistics in S: the gstat package. *Computers & Geosciences*. 30: 683-691.
- Perez-Quezada JF, Pettygrove GS, Plant RE. 2003. Spatial-temporal analysis of yield and the influence of soil factors in two four-crop-rotation fields in the Sacramento Valley, California. *Agronomy Journal*, 95(3): 676–687.
- Pierce FJ, Nowak P. 1999. Aspects of precision agriculture. En: Sparks, D.L. (Eds.). *Advances in agronomy*. Newark. (Academic Press). pp. 1-85.
- Pierce FJ, Anderson NW, Colvin TS, Schuler JK, Humburg DS, McLaughlin NB. 1997. Yield mapping. En: Pierce FJ, Sadler EJ. [Eds.]. *The state of site-specific management for agriculture*. Madison: ASA, CSSA, SSSA. pp. 211- 244.
- Ping JL, Dobermann A. 2005. Processing of yield map data. *Precision Agriculture*, 6: 193-212.
- Plant RE 2001. Site-specific management: the application of information technology to crop production. *Computers and Electronics in Agriculture*. 30: 9-29.

- Pravia MV. 2009. Uso de los elementos de agricultura de precisión y modelos de simulación para la incorporación de la dimensión espacio-temporal en la investigación de cultivos agrícolas: A) Impacto de prácticas de manejo de suelos y atributos del terreno en la productividad de sorgo a escala de chacra. B) Simulación de la producción de arroz en Uruguay utilizando el modelo DSSAT v4 CERES-Rice. Tesis Magíster en Ciencias Agrarias. Montevideo, Uruguay. Facultad de Agronomía. 102 p.
- Roel A., Firpo H., & Plant, R. E. 2007. Why do some farmers get higher yields? Multivariate analysis of a group of Uruguayan rice farmers. *Science Direct*, 58: 78-92.
- Roel A, Plant R E. 2004a. Spatiotemporal analysis of rice yield variability in two California fields. *Agronomy Journal*, 96(1): 77–90.
- b. Factors underlying yield variability in two California rice fields. *Agronomy Journal*, 96(5): 1481–1494.
- Roel A, Plant R E. 2001. Using GIS to improve Rice Yields. *Rice Farming*, 35(4) p.:
- Schroder JJ, Neeteson JJ, Oenema O, Struik PC. 2000. Does the crop or the soil indicate how to save nitrogen in maize production? Reviewing the state of the art. *Field Crop Research*, 66: 151-164.
- Simmonds, M. B., Plant, R. E., Peña-Barragán J. M., van Kessel, C., Hill, J., & Linqvist B. A. 2013. Underlying causes of yield spatial variability and potential for precision management in rice systems. *Precision Agriculture*, 14: 512-540.

- Stafford JV, Murray LR, Bolam HC. 1998. Using yield maps to regionalize fields into potential management units. En: Proceeding of the international conference on precision agriculture. St. Paul: ASA, CSSA, SSSA. pp. 225-237.
- 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: 98-107.
- Yanai J, Lee CK, Kaho T, Lida M, Matsui T, Umeda M, et al. 2001. Geostatistical analysis of soil chemical properties and rice yield in a paddy field and application to the analysis of yield-determining factors. *Soil Science and Plant Nutrition*, 47(2): 291–301.

#### **4. RESULTADOS Y DISCUSIÓN**

La Agricultura de Precisión es definida como la aplicación de tecnologías y principios para la resolución de problemas en la producción agrícola. Generalmente está asociada con el manejo sitio-específico de los cultivos, que es el manejo de las chacras a una escala espacio-temporal apropiada a la variación natural (Plant, 2001). Las tecnologías asociadas a la agricultura de precisión, proveen la oportunidad de realizar medidas más precisas de la variabilidad del rendimiento y de los procesos ecológicos que subyacen en esta variabilidad (Roel y Plant, 2004a). Análisis efectivos de los datos provenientes de estas mediciones, requieren métodos estadísticos diferentes a los tradicionalmente empleados en los experimentos agronómicos controlados.

Dentro de la chacra el P Bray (Pb) y el P cítrico (Pc) fueron los que más variaron (CV 48,97-76,37 %). El Potasio (K) y la textura del suelo presentaron una variabilidad moderada (CV 12,56-20,02 %), mientras que el pH fue el de menor variación (CV 5,96 %). El rendimiento de arroz promedio fue 9114 kg ha<sup>-1</sup> en 2011 y 8554 kg ha<sup>-1</sup> en 2012 y el CV fue 14% y 15%, respectivamente; sugiriendo una variabilidad inter-chacra moderada a alta. Simmonds et al. (2013) reportaron un rango de CV de rendimiento de 7,09% a 10,48% en campos de arroz irrigados. Otro autores, en chacras más pequeñas (0,5-1 ha) reportaron un CV de 39% (Dobermann 1994), y 5,89% (Yanai et al. 2001).

El algoritmo k-medias de análisis de conglomerado pudo identificar dos zonas de rendimiento contrastantes. Basándose en los resultados del análisis, la chacra exhibe un patrón de rendimiento estable y contiguo para cada conglomerado. La consistencia en el agrupamiento para k=2 indica que el patrón espacial de rendimiento en el campo, fue influenciado en su mayoría por factores persistentes y endógenos del campo, como por ejemplo las propiedades del suelo. Las zonas identificadas con altos rendimientos, coincidieron con zonas de mayor contenido de P y arena. Durante las 2 zafras estudiadas, las zonas de suelo arenoso, permitieron una mejor y más temprana implantación del cultivo. Adicionalmente, en los suelos arenosos, los nutrientes están más accesibles para los cultivos. En sistemas de arroz irrigado, el agua no es un factor



limitante. Sin embargo, la disponibilidad de nutrientes, como el P es uno de los factores más determinantes en la concreción del rendimiento (Dobermann y Cassman, 2002). Por otro lado, la inconsistencia en la estabilidad espacial de los grupos  $k=3$  y  $k=4$ , indica que algunos factores transitorios exógenos de la chacra (malezas, microclima), influyeron sobre la variabilidad espacial del rendimiento. En las observaciones realizadas durante el ciclo del cultivo, para los 2 años, se encontraron puntos (en las zonas de bajo rendimiento) con alta densidad de malezas, asociadas con pobre implantación del cultivo y altas y/o inestables láminas de agua. Roel et al., (2007) y Simmonds et al., (2013) reportaron en campo de arroz irrigado, que factores de manejo como la densidad de malezas y la estabilidad de la lámina de agua tuvieron un efecto en la variabilidad espacial del rendimiento inter-chacra.

Las nuevas tecnologías asociadas a la agricultura de precisión, permiten coleccionar información georreferenciada de muchas variables en el campo. Mediante análisis geoestadísticos univariados no se pudo mapear la variabilidad espacial de las propiedades del suelo de la chacra. Sin embargo, mediante un análisis espacial multivariado se pudo evaluar la covariación espacial de las propiedades del suelo a partir del análisis de componentes principales (ACP). El análisis multivariado pudo resumir las estructuras de covariación e identificar patrones espaciales. Para los 2 años de estudio, aproximadamente un 60% de la variabilidad total fue resumida en 2 componentes principales. El primer componente (CP1) fue caracterizado por la textura del suelo, coincidiendo la zona de alto contenido de arena con las zonas de altos rendimientos identificadas por el análisis de conglomerado. A su vez, estos resultados coincidieron con el análisis de CART, el cual identificó a la textura del suelo como una variable primaria y/o competidora de partición. Para 2011-2012, la variación para el segundo componente fue impulsada por los fósforos (P Bray y P cítrico) y para el 2012-2013 la variación fue impulsada por el pH. Córdoba et al., (2012) reportaron para campos de soja, que la electro-conductividad covariaba con el rendimiento, marcando a MULTISPATI-PCA como una herramienta estadística prometedora para delimitar zonas de manejo homogéneo dentro de la chacra.

El análisis de CART, para los 2 años de estudio y con  $k=2$  y  $k=3$ , la textura del suelo (% de arena y limo) fue identificado como una variables primaria y competidora de partición. Los factores edáficos que limitaron el rendimiento, explican el patrón estable encontrado para los 2 años; las zonas de alto y bajo rendimiento fueron las mismas. Sin embargo, factores de manejo como lámina de agua inestable y diferencias de altura de planta, asociadas a la implantación, indicaron algunas diferencias entre zonas para cada año. Roel et al. 2005, encontraron variaciones en el rendimiento producto de diferencias en el manejo del agua. Zonas de bajo rendimiento presentaron láminas de riego inestables durante todo el ciclo del cultivo, con marcada importancia en la etapa de primordio, donde el rendimiento es concretado. Debido al sistema de riego con taipas, el éxito en el buen manejo del agua depende de una buena sistematización y nivelación del terreno. Walker et al., (2003) y Roel y Plant (2004b) observaron efectos de la nivelación del terreno en la variabilidad espacial del rendimiento. El análisis de CART identificó como variables competidoras y sugeridas a la altura de la lámina de agua como variables de partición entre grupos.

El análisis de regresión múltiple identificó las mismas variables que el análisis espacial multivariado y el análisis de CART. Las interacciones significativas entre las variables del suelo, indican la complejidad del fenómeno y la dificultad que conlleva el estudio de los factores que afectan el rendimiento. A través del uso de esta herramienta, no solo se obtuvo el mejor modelo para redecir el rendimiento sino que también se puede entender como interaccionan las variables independientes con el rendimiento.

Mediante un análisis económico simple de la información obtenida de los monitores de rendimiento, se pudo identificar una zona que presentó para los 2 años pérdidas de alrededor de 300 U\$S  $ha^{-1}$ . La siembra en los 2 años de esta zona, que representa la mitad de la chacra, produjo un balance negativo de U\$S -4000, sugiriendo que la no siembra de esa zona cuando las condiciones óptimas de siembra no están presentes, podría ser una medida de manejo sitio-específico sencilla.

## **5. CONCLUSIONES**

Se obtuvieron resultados consistentes y robustos a partir del uso de diversas herramientas estadísticas que permitieron diferentes abordajes a la situación.

Se identificaron 2 zonas consistentes de rendimiento que podrían ser beneficiadas con un manejo diferencial.

Mediante un análisis económico sencillo se utilizó la información que brindan las tecnologías asociadas a la AP, y se construyó una herramienta válida en la toma de decisiones.

A nivel predial se podría pensar en utilizar los datos de monitores de rendimiento para delinear zonas de manejo, muestreo dirigido y/o aplicar manejo sitio-específico.

## **6. BIBLIOGRAFÍA**

- Basso B, Bertocco M, Sartori L, Martin EC. 2007. Analyzing the effects of climate variability on spatial pattern of yield in a maize-wheat-soybean rotation. *European Journal of Agronomy*, 26(2): 82-91.
- Best S, León L. 2006. Análisis de datos. En: Bongiovanni R, Chartuni E, Best S, Roel A. [Eds.]. *Agricultura de precisión; integrando conocimientos para una agricultura moderna y sustentable*. Montevideo: PROCISUR. pp. 145-170.
- Bhatti AU, Mulla DJ, Frazier BE. 1991. Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and Thematic Mapper images. *Remote Sensing Environment*. 37: 181-191.
- Bivand R, Altman M, Anselin L, Assuncao R, Berke O, Bernat A, et al. 2010. Spdep: spatial dependence: weighting schemes, statistics and models. R package version 0.5-21. <http://CRAN.R-project.org/package=spdep>. Accessed 01 jul 2013.
- Blanco J. 2006. Análisis de grupos. En: Camaño G, Nalbarte L, Alvarez R, Castrillejo A. [Eds.]. *Introducción al análisis multivariado*. Montevideo: Universidad de la República. pp 9-35.
- Bocchi S, Castrignano A. 2007. Identification of different potential production areas for corn in Italy through multitemporal yield map analysis. *ScienceDirect*, 102: 185-197.
- Casanova D, Goudriaan J, Bouma J, Epema GF. 1999. Yield gap analysis in relation to soil properties in direct-seeded flooded rice. *Geoderma*, 91(3-4): 191-216.
- Cassman KG. 1999. Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. *Proceedings of the National Academy of Sciences, USA*, 96(11): 5952-5959.
- Cassman KG, Dobermann A, Walters DT 2002. Agroecosystems, nitrogen-use efficiency, and nitrogen management, *Ambio*, 31: 132-140.

- Cliff AD, Ord KK. 1981. *Spatial Processes: Models and Applications*. Londres: Pion Limited. 266p.
- Córdoba M, Balzarini M, Bruno C, Costa JL. 2012. Identificación de zonas de manejo sitio-específico a partir de la combinación de variables de suelo. *Corpoica*. 13(1): 47-54.
- Córdoba M, Balzarini M, Bruno Cecilia, Costa JL. 2012. Análisis de componentes principales con datos georreferenciados. *Facultad Uncuyo*. 44(1): 27-39.
- Cressie NSC. 1991. *Statistics for spatial data*. Wiley: Nueva York. 928p.
- De'ath G, Fabricius KE. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. 81: 3178-3192.
- Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A. 2010. e1071: Misc functions of the department of statistics (e1071), TU Wien. R package version 1.5-24. <http://CRAN.R-project.org/package=e1071>. Accessed 01 Jul 2013.
- Dobermann, A. 1994. Factors causing field variation of direct-seeded flooded rice. *Geoderma*, 62(1-3): 125-150.
- Dobermann A, Cassman KG. 2002. Plant nutrient management for enhanced productivity in intensive grain production systems of the United States and Asia. *Plant and Soil*, 247: 153-175.
- Dobermann A, Ping JL, Adamchuk VI, Simbahan GC, Ferguson RB. 2003. Classification of crop yield variability in irrigated production fields. *Agronomy Journal*, 95: 1105-1120.
- Doerge TA. 1999. Yield map interpretation. *Production Agriculture*. 12:54-61.
- Dray S, Jombart T. 2010. Revisiting Guerry's data: introducing spatial constraints in multivariate analysis. *The Annuals of Applied Statistics*, 5(4): 2278-2299.
- Goodall DW. 1954. Objective methods for the classification of vegetation III. An essay on the use of factor analysis. *Australian Journal of Botany*, 2: 304-324.
- Goovaerts P. 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, 89: 1-45.

- Guastaferrero F, Castrignano A, De Benedetto D, Sollitto D, Troccoli A, Cafarelli B. 2010. A comparison of different algorithms for the delineation of management zones. *Precision Agriculture*, 11(6): 600–620.
- Isaaks EH, Srivastava RM. 1989. *An introduction to applied geostatistics*. Nueva York: Oxford University Press. 553p.
- Jain AK, Dubes RC. 1984. *Algorithms for clustering data*. Nueva Jersey: Prentice-Hall. 304p.
- Jayne DB, Kaspar TC, Colvin TS, James DE. 2003. Cluster analysis of spatiotemporal corn yield patterns in an Iowa fields. *Agronomy Journal*, 95: 574-586.
- Kravchenko AN. 2003. Influence of spatial structure on accuracy of interpolation methods. *Soil Science Society of America Journal*, 67: 1564-1571.
- Lark RM, Stafford JV. 1997. Classification as a first step in the interpretation of temporal and spatial variation of crop yield. *Annals of Applied Biology*, 130(1): 111–121.
- Matheron G. 1963. Principles of geostatistics. *Economic Geology*, 58: 1246-1266.
- Moran PAP. 1948. The interpretation of statistical maps. *Royal Statistical Society*, 10(2): 243-251.
- Mulla DJ, McBratney A.B. 2002. Soil spatial variability. En: Warrick AW. [Eds.]. *Soil physics companion*. Boca Ratón: CRC. pp. 343-373.
- Mulla, DJ, Schepers, JS. 1997. Key processes and properties for site-specific soil and crop management. En: Pierce FJ, Sadler EJ (Eds.), *The state of site-specific management for agriculture*. Madison: ASA, CSSA, SSSA. pp. 211-244.
- Mzuku M, Khosla R, Reich R, Inman D, Smith F, MacDonald L. 2005. Spatial variability of measured soil properties across site-specific management zones. *Soil Science Society of America Journal*, 69(5): 1572-1579.
- Oliver MA. 2010. An overview of geostatistics and precision agriculture. En: Oliver MA [Eds.]. *Geostatistical applications for precision agriculture*. Nueva York: Springer. pp. 1-32.

- Pebesma EJ. 2004. Multivariate geostatistics in S: the gstat package. *Computers & Geosciences*. 30: 683-691.
- Perez-Quezada JF, Pettygrove GS, Plant RE. 2003. Spatial-temporal analysis of yield and the influence of soil factors in two four-crop-rotation fields in the Sacramento Valley, California. *Agronomy Journal*, 95(3): 676–687.
- Pierce FJ, Nowak P, 1999. Aspects of precision agriculture. En: Sparks DL. [Eds.]. *Advances in agronomy*. Newark: Academic Press. pp. 1-85.
- Pierce FJ, Anderson NW, Colvin TS, Schuler JK, Humburg DS, McLaughlin NB. 1997. Yield mapping. En: Pierce FJ, Sadler EJ. [Eds.]. *The state of site-specific management for agriculture*. Madison: ASA, CSSA, SSSA. pp. 211- 244.
- Ping JL, Dobermann A. 2005. Processing of yield map data. *Precision Agriculture*, 6: 193-212.
- Plant, RE. 2001. Site-specific management: the application of information technology to crop production. *Computers and Electronics in Agriculture*, 30: 9-29.
- Pravia MV. 2009. Uso de los elementos de agricultura de precisión y modelos de simulación para la incorporación de la dimensión espacio-temporal en la investigación de cultivos agrícolas: A) Impacto de prácticas de manejo de suelos y atributos del terreno en la productividad de sorgo a escala de chacra. B) Simulación de la producción de arroz en Uruguay utilizando el modelo DSSAT v4 CERES-Rice. Tesis Magíster en Ciencias Agrarias. Montevideo, Uruguay. Facultad de Agronomía. 102p.
- Ripley, B 2002, MASS: Support functions and datasets for Venables and Ripley`s MASS. R package version 7.3-29. <http://CRAN.R-project.org/package=MASS>. Accessed 01 Jul 2013.
- Roel A, Firpo H, Plant RE. 2007. Why do some farmers get higher yields? Multivariate analysis of a group of Uruguayan rice farmers. *Science Direct*, 58: 78-92.
- Roel A, Terra J. 2006. Muestreo de suelos y factores limitantes del rendimiento. En: Bongiovanni R, Chartuni E, Best S, Roel A. [Eds.]. *Agricultura de precisión*;

- integrando conocimientos para una agricultura moderna y sustentable. Montevideo, Uruguay, PROCISUR. pp. 65-80.
- Roel A, Mutters RG, Eckert JW, Plant RE. 2005. Effect of low water temperature on rice yield in California. *Agronomy Journal*, 97(3): 943–948.
- Roel A, Plant RE 2004a. Spatiotemporal analysis of rice yield variability in two California fields. *Agronomy Journal*, 96(1): 77–90.
- Roel A, Plant RE. 2004b. Factors underlying yield variability in two California rice fields. *Agronomy Journal*, 96(5): 1481–1494.
- Schroder JJ, Neeteson JJ, Oenema O, Struik PC. 2000. Does the crop or the soil indicate how to save nitrogen in maize production? Reviewing the state of the art. *Field Crop Research*, 66: 151-164.
- Simmonds MB, Plant RE, Peña-Barragán JM, van Kessel C, Hill J, Linqvist BA. 2013. Underlying causes of yield spatial variability and potential for precision management in rice systems. *Precision Agriculture*, 14: 512-540.
- Stafford JV, Murray LR, Bolam HC. 1998. Using yield maps to regionalize fields into potential management units. En: *Proceeding of the international conference on precision agriculture*. St. Paul: ASA, CSSA, SSSA. pp. 225-237.
- 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: 98-107.
- Therneau TM, Atkinson B. 2009. R port by brian ripley. rpart: recursive partitioning. R package version 3.1-45. <http://CRAN.R-project.org/package=rpart>. Accessed 01 Jul 2013.
- Vayssières MP, Plant RE, Allen-Diaz BH. 2000. Classification trees: An alternative non-parametric approach for predicting species distributions. *Vegetation Science*, 11: 679-694.
- Walker TW, Kingery WL, Cox MS, Oldham JL, Gerard PD, Han FX. 2003. Rice yield and soil chemical properties as affected by precision land-leveling in alluvial soils. *Agronomy Journal*, 95(6): 1483–1488.



- Wang Y, Shou-Hung C, Kuo-Wei C, Shen Y. 2012. Identifying and characterizing yield limiting factors in paddy rice using remote sensing yield maps. *Precision Agriculture*, doi 10.1007/s1111901292665.
- Warrick AW, Myers DE, Nielsen DR. 1986. Geostatistical methods applied to soil science. A. Klute. [Eds.]. *Methods of soil analysis: Part I, physical and mineralogical methods, agronomy monograph*. Madison: ASA, CSSA, SSSA. pp. 53-82.
- Wehlan, BM, McBratney AB. 2000. The “null hypothesis” of precision agriculture management. *Precision Agriculture*, 2(3): 265–279.
- Wollenhaupt NC, Mulla DJ, Gotway Crawford CA 1997. Soil sampling and interpolation techniques for mapping spatial variability of soil properties. En: Pierce FJ, Sadler EJ. [Eds.]. *The state of site-specific management for agriculture*. Madison: ASA, CSSA, SSSA. pp. 19-53.
- Yanai J, Lee CK, Kaho T, Iida M, Matsui T, Umeda M, Kosaki T. 2001. Geostatistical analysis of soil chemical properties and rice yield in a paddy field and application to the analysis of yield-determining factors. *Soil Science and Plant Nutrition*, 47(2): 291–301.

