



UNIVERSIDAD NACIONAL DEL NORDESTE



FACULTAD DE CIENCIAS AGRARIAS

## TRABAJO FINAL DE GRADUACIÓN

Modalidad Tesina

### **DETERMINACIÓN DE TAMAÑO DE MUESTRA Y DISTANCIA DE MUESTREO EN LAS ESTIMACIONES DE LA PRODUCCIÓN DE NARANJO SALUSTIANA EN BASE A SU VARIABILIDAD**

Autora: Nydia Alejandra Montoya

Lugar de Trabajo: Departamento de Matemática y Estadística de la Facultad de Ciencias Agrarias de la Universidad Nacional del Nordeste.

Asesora: E. E. (Dra.) Laura I. Giménez

Corrientes – Argentina

2020

## ÍNDICE

<b>ÍNDICE GENERAL</b>	I
<b>ÍNDICE DE TABLAS</b>	II
<b>ÍNDICE DE FIGURAS</b>	II
<b>RESUMEN</b>	1
<b>1. INTRODUCCIÓN</b>	2
OBJETIVO GENERAL	4
OBJETIVOS ESPECÍFICOS	4
<b>2. ANTECEDENTES</b>	5
<b>3. MATERIALES Y MÉTODOS</b>	7
Datos Experimentales	7
Estimación de Tamaño de Muestra	8
Estimación de Variabilidad “ $\sigma$ ”	8
Análisis Geoestadístico: Estimación de la distancia de muestreo	9
Modelos teóricos de Variogramas	10
Modelo lineal	10
Modelo esférico	10
Modelo exponencial	10
Modelo gaussiano	11
<b>4. RESULTADOS Y DISCUSIÓN</b>	11
Estadística Descriptiva	11
Número de Frutos por árbol al momento de cosecha	11
Estimación de Variabilidad	13
Estimación de Tamaño de Muestra	13
Análisis Geoestadístico	15
El variograma	16
<b>5. CONCLUSIONES</b>	20
<b>BIBLIOGRAFÍA</b>	21

## TABLAS

**Tabla N° 1:** Medidas descriptivas del número de frutos por árbol del lote completo y ambos lotes de naranjo Salustiana en Santa Rosa, Corrientes. Campaña 2013-2014.

**Tabla N° 2:** Medidas descriptivas del número de frutos por árbol de los lotes 1 y 2 de naranjo Salustiana en Santa Rosa, Corrientes. Campaña 2013-2014.

**Tabla N° 3:** Medias y desvío estándar (DE) de la carga a cosecha en número de frutos por árbol y kilogramos de frutos cosechados por planta, por lote.

**Tabla N° 4:** Valores de error (%) respecto al promedio de carga asociado y Tamaño de muestra (n) en número de áboles correspondiente a cada error.

**Tabla N° 5:** Parámetros de los distintos modelos de variogramas.

**Tabla N° 6:** Porcentaje de relación de fuerza y grado de dependencia espacial de carga de frutos de naranjo 'Salustiana' para la Campaña 2013-2014.

## FIGURAS

**Figura N° 1:** Imagen Satelital del lote comercial de naranjo dulce 'Salustiana', en Santa Rosa, Corrientes, Argentina.

**Figura N° 2:** Parámetros del semivariograma.

**Figura N° 3:** Histograma de frecuencias y boxplot del número de frutos cosechados por árbol.

**Figura N° 4:** Tamaño de muestra (n) asociado a diferentes valores de error expresado como % respecto al promedio de carga asociado (c).

**Figura N° 5:** Carga en la cosecha de los 81 áboles del lote completo con representación proporcional del número de frutos por árbol a cosecha.

**Figura N° 6:** Variograma empírico con estimador clásico.

**Figura N° 7:** Ajuste de los modelos a) Lineal, b) Esférico, c) Exponencial y d) Gausiano al variograma empírico.

**Figura N° 8:** Mapa de isolíneas del lote comercial de Naranjo dulce 'Salustiana' ubicado en la localidad de Santa Rosa, Corrientes, Argentina.

## RESUMEN

En Argentina se producen 1.024.918 toneladas de naranja, *Citrus sinensis* L. Osb. La provincia de Corrientes cuenta con 13.851 hectáreas de cultivo y aporta el 30% de la producción nacional. Entre las variedades producidas está Salustiana, con buena calidad de fruta. El objetivo del trabajo fue determinar el tamaño de muestra de árboles a través del análisis de la variabilidad de la carga en el momento de la cosecha de frutos de naranjo Salustiana y caracterizar la variabilidad espacial a modo de sugerir estrategias de muestreo. Los datos experimentales provienen de un lote comercial de 5,3 hectáreas de naranjo dulce *Citrus sinensis* L. Osb. var. 'Salustiana', ubicado en la localidad de Santa Rosa, Corrientes. Mediante un muestreo aleatorio simple se seleccionaron 81 árboles que fueron georreferenciados y al momento de la cosecha, se registró el número de frutos por árbol (carga). Mediante una aproximación para muestreo aleatorio simple para la estimación de carga media, se determinó número de árboles ( $n$ ) que integraron la muestra y se hizo una estimación de variabilidad. Con el análisis geoestadístico se determinó la distancia de muestreo y se obtuvo un Variograma Empírico que se ajustó a un Variograma Teórico. Se realizó también un mapa de rendimiento. Las medidas de resumen presentaron grandes diferencias entre los lotes estudiados para la Campaña 2013-2014, con mayor producción y menor variabilidad en el Lote 2, con 126,08 frutos por árbol al momento de cosecha y un coeficiente de variación del 53,47 %. Los valores de diámetro ecuatorial (DE) asociados a cada una de las medias evidenciaron la gran dispersión de la carga dentro del lote, y también se observaron diferencias entre lotes. Con respecto al tamaño de muestra, se considera apropiado un tamaño de muestra entre 22 y 26 árboles por lote, con lo cual se consigue una precisión entre el 88 y 90 % en la estimación de carga de frutos de naranjo dulce 'Salustiana'. En un gráfico denominado de "Burbujas", se observó la variabilidad del conjunto de datos y valores altos de producción en el Lote 2. Se calcularon distintos modelos de variogramas y se obtuvieron sus correspondientes parámetros. El modelo seleccionado por menor Residual SS fue el Exponencial, con un nivel moderado de dependencia espacial. Con un mapa de isolíneas se observaron valores puntuales que representan magnitudes de la producción. El Lote 2 con un  $n$  mayor, más puntos, es el de mayor precisión.

## 1. INTRODUCCIÓN

La Argentina tiene tradición citrícola, produce limones, naranjas, pomelos y mandarinas. El clima benigno y la amplia distribución geográfica, permiten abastecer la demanda de frutos cítricos del país durante casi todo el año y es muy oportuna para las exportaciones en contra-estación a los países del hemisferio Norte (Avanza, 2010).

En Argentina se producen 1.024.918 toneladas de naranja, *Citrus sinensis* L. Osb., de las cuales 240.014 son para industria, 487.792 para consumo interno y 76.957 para exportación en fresco. La provincia de Corrientes cuenta con 13.851 hectáreas de cultivo y aporta el 30% de la producción nacional (FEDERCITRUS, 2018). Las variedades producidas a nivel nacional son Salustiana, Washington Navel, Navel Late, Valencia Seedless y Valencia late, de las cuales se encuentran en expansión las variedades sin semilla como la Navelina (Danza, 2017).

La naranja 'Salustiana' fue originada probablemente en Valencia, España, por mutación espontánea de la variedad común a finales de la década de 1940. Se trata de un árbol vigoroso, de tamaño medio a grande y con cierta tendencia a la formación de ramas verticales que lo distinguen de las otras variedades (Agustí, 2003). Es de cosecha tardía que puede mantener frutos maduros en el árbol durante largos períodos sin perder sus cualidades. Sin embargo, presenta años de alternancia en la producción y tamaño pequeño que disminuyen su productividad, con una regular producción promedio pero con buena calidad de la fruta; es de madurez intermedia (Orduz *et al.*, 2009).

La previsión de la producción de los cultivos resulta de gran interés para agricultores e instituciones agrarias públicas, puesto que una estimación precoz de la cosecha final ayuda en gran medida a planificar y ajustar los subsidios gubernamentales (Sinclair y Seligman, 2000), del mismo modo que supone una importante fuente de información de cara a la planificación de los recursos humanos y económicos necesarios para la recogida del fruto además de optimizar las estrategias de comercio y distribución internacional (Galán *et al.*, 2008). Un pronóstico de rendimiento preciso que resulte en previsiones de producción de fruto con una considerable antelación a la recolección podría predecir la seguridad alimentaria regional y el abastecimiento a los mercados, mejorando la toma de decisiones a nivel empresarial, tanto en términos de calendarios de cosecha y

equipos de procesamiento, como en el marketing o necesidades de almacenamiento. A nivel gubernamental regional o nacional, este conocimiento permite formular políticas y planificar estrategias tendientes a la apertura de mercados sobre una base cierta de la producción que se puede ofrecer (Folwell *et al.*, 1994).

Es por ello importante lograr la mayor exactitud posible y anticipación de los pronósticos de cosecha de frutas, con el fin de incrementar la base de información para la toma de decisiones de varios sectores: administradores de empresas frutícolas, empaques, frigoríficos, industrias de jugos de fruta y secaderos, que sufren el impacto de los cambios en los niveles de aprovisionamiento de productos (Avanza, 2010).

La representatividad de una muestra, permite extrapolar y por ende generalizar los resultados observados en ésta, a la población accesible y a partir de ésta, a la población blanco. En consecuencia, una muestra será representativa o no; sólo si fue seleccionada al azar y por otro lado, si el número de sujetos seleccionados representan numéricamente a la población que le dio origen respecto de la distribución de la variable en estudio en la población (Otzen & Manterola, 2017). Es así como el análisis de una muestra permite realizar inferencias, extrapolar o generalizar conclusiones a la población blanco con un alto grado de certeza (Dieterich, 1996).

A través de la estimación de la variabilidad de la producción con un adecuado tamaño muestral, se podrían realizar pronósticos de cosecha confiables. La precisión puede establecerse por el recíproco de la estimación de varianza, mientras que los costos se basan principalmente en el tiempo empleado en la toma de mediciones. Dicha precisión puede ser prefijada por el investigador, mientras que el costo en tiempo puede disminuirse tomando aquel tamaño de muestra mínimo que proporcione la precisión deseada (Cochran, 1982; Scheaffer *et al.*, 2007). En este contexto, el estudio de la variabilidad espacial de la producción, permitiría la identificación de sitios representativos, la asignación de recursos de acuerdo al sistema agrícola o agroecosistema y la interpretación de los efectos del sitio en resultados experimentales. Esto ayudaría a establecer sistemas de producción más sostenibles y eficientes (Valbuena Calderón *et al.*, 2008; Rochas *et al.*, 2016; Mestas *et al.*, 2010).

Giménez *et al.*, 2012 han utilizado, herramientas estadísticas, geoestadísticas y software en conjunto para mejorar los métodos de predicción de cosecha de naranja “Valencia late”, adoptando la variabilidad espacial y aplicando las herramientas del análisis geoestadístico para describir la autocorrelación en los datos. La metodología de la geoestadística reside en reconocer, una variación a gran escala o tendencia y otra a pequeña escala o correlación espacial y expresar estas variaciones en un modelo espacial (Dibiasi, 2001). Sin embargo, son escasos los trabajos acerca del uso de la geoestadística y de la variabilidad como herramientas para definir estrategias de muestreo y así realizar pronósticos de cosecha.

## **OBJETIVO GENERAL**

El presente trabajo tuvo como objetivo determinar el tamaño de muestra de árboles a través del análisis de la variabilidad de la carga en el momento de la cosecha de frutos de naranjo Salustiana y caracterizar la variabilidad espacial a modo de sugerir estrategias de muestreo.

## **OBJETIVOS ESPECÍFICOS**

- Determinar el tamaño de muestra para la estimación de carga a cosecha.
- Describir y cuantificar la dependencia espacial de la producción por árbol de naranjo Salustiana.
- Determinar la distancia a partir de la cual la producción por árbol no tiene dependencia espacial.

## **2. ANTECEDENTES**

La producción de cítricos se caracteriza por la gran heterogeneidad en las plantaciones, la cual difiere en el tiempo (alternancia entre años) y fundamentalmente en el espacio, ya que depende de factores ambientales y de manejo (Avanza, 2010; Reeb, 2003). El conocimiento anticipado (previo a la cosecha) de la producción por árbol, se torna en una ventaja estratégica para el productor citrícola, ya que le permite definir estrategias de gestión y de comercialización de sus productos.

Los estudios basados en observaciones objetivas de campo requieren de la decisión acerca del tamaño de muestra, ya que se debe equilibrar entre los recursos disponibles y la precisión deseada. Una muestra demasiado grande implica un desperdicio de recursos, y una demasiado pequeña disminuye la precisión de las estimaciones. Dicha precisión puede ser prefijada por el investigador, mientras que el costo en tiempo puede disminuirse tomando aquel tamaño de muestra mínimo que proporcione la precisión deseada (Gomez & Gomez, 1984; Scheaffer *et al.*, 2007). Redons *et al.* (2008) han determinado el tamaño óptimo de muestra (número de árboles) teniendo en cuenta los errores de estimación y la varianza de la proporción de árboles.

Avanza (2010) en la definición de una metodología de predicción de cosecha en naranjo Valencia late, determinó que para estimar la carga al momento de cosecha, con una precisión del 90 % aproximadamente, se requiere un tamaño de muestra de entre 18 y 21 árboles por huerto.

Si bien el mencionado trabajo, ha contribuido al conocimiento relativo al muestreo, es necesario plantearse el estudio de otras variedades y condiciones ambientales para explicar con más detalle la variabilidad encontrada.

Por otro lado, la facilidad para la obtención de datos georreferenciados en los últimos años ha favorecido el desarrollo de software, como sistemas de información geográficos, específicos de geoestadística y procesamiento de imágenes para el conocimiento de factores de producción (Cabrera Brunetti *et al.*, 2010).

En Colombia, la producción de naranja de variedad 'Salustiana' con cinco años de edad, estimada para una distancia de siembra 8m x 6m y 208 árboles hectárea<sup>-1</sup> tiene un promedio de 45 frutos por árbol, con un peso promedio del fruto de 200 g y 9 kg de fruta árbol<sup>-1</sup>, lo que lleva a una producción de 1,87 kg hectárea<sup>-1</sup>.

Los resultados de producción de estas variedades son preliminares puesto que se refieren a la primera cosecha y su comportamiento y producción en las mismas condiciones pueden modificarse al tener un número mayor de cosechas en los años futuros (Orduz *et al.*, 2009).

La producción de fruta cítrica para exportación, requiere la obtención de un producto de alta calidad y de regularidad en la oferta año con año. La alternancia productiva, caracterizada por ciclos de alta cosecha, seguidos de años de poca o nula fructificación, además de un origen genético, puede ser inducida por problemas climáticos, por falta de adaptación de una variedad a una determinada zona o por manejos inadecuados de la plantación (Gravina *et al.*, 2000).

La naranja ‘Salustiana’, originada por mutación de la variedad ‘Comuna’ en la provincia de Valencia (Bono *et al.*, 1985), se cosecha en media estación, presentando frutos de buena coloración, excelente calidad interna y sin semillas. Sin embargo, bajo determinadas condiciones de cultivo, ha sido reportada como una variedad con alta tendencia a la alternancia de cosechas (Agustí *et al.*, 1985), lo que disminuye su potencial comercial. En Uruguay, esta variedad es la tercera en importancia dentro de las naranjas (Patiño, 1998) y es considerada de producción irregular, aunque no existe ningún estudio que permita confirmar una tendencia al comportamiento alternante (Gravina *et al.*, 2000).

### 3. MATERIALES Y MÉTODOS

#### Datos Experimentales

La información de campo provino de un lote comercial ubicado en la localidad de Santa Rosa, Corrientes, donde predominan suelos arenosos, del tipo Udisamamment. Según el sistema de Köppen, la clasificación climática, al igual que para todas las localidades de la provincia, es la de templado húmedo, sin estación seca definida, con precipitaciones máximas en otoño y primavera, veranos muy cálidos con temperaturas superiores a los 22° C y media anual superior a 18° C. Dicho lote cuenta con 5,3 hectáreas de naranjo dulce *Citrus sinensis* L. Osb. var. 'Salustiana', de cinco años injertado sobre trifolio *Poncirus trifoliata*, en un marco de plantación de 7 metros por 3,5 metros. Durante la campaña 2013-2014 se seleccionaron 81 árboles mediante un muestreo aleatorio simple. En dichos árboles, al momento de la cosecha, en el mes de Mayo, se registró el número de frutos por árbol (carga). Los árboles fueron georreferenciados con el GPS (Sistema de Posicionamiento Global con Sistema de Navegación por Tiempo y Distancia) obteniendo así las coordenadas geográficas (grados, minutos y segundos).

La **Figura N° 1** muestra la distribución de los puntos en el espacio, la línea blanca representa una división realizada por el productor (cortina de eucaliptos) dividiendo en dos al lote descripto, resultando a la derecha el Lote 1 y a la izquierda el Lote 2.



**Figura N° 1:** Imagen Satelital del lote comercial de naranjo dulce 'Salustiana', Santa Rosa, Corrientes, Argentina (Google Earth Pro, 2019).

### Estimación de Tamaño de Muestra

Mediante una aproximación para muestreo aleatorio simple para la estimación de carga media, basado en la distribución normal, con una confianza y una precisión dada, se determinó número de árboles ( $n$ ) que integraron la muestra. La aproximación utilizada fue la siguiente (Di Rienzo *et al.*, 2000):

$$n \geq \left( \frac{2Z_{(1-\alpha/2)} \cdot \sigma}{c} \right)^2$$

Donde:

$n$ : Tamaño de muestra (Nº de árboles por huerto);

$Z_{(1-\alpha/2)}$ : Valor de la distribución normal estándar  $(1-\alpha/2) = 0,975$ ;

$\sigma$ : Desviación estándar;

$c$ : Error admitido expresado como un porcentaje (%) del promedio de la carga a cosecha.

### Estimación de Variabilidad “ $\sigma$ ”

Para estimar el desvío estándar  $\sigma$  necesario en la determinación del tamaño de muestra se planteó el siguiente modelo:

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

Donde:

$y_{oj}$ : carga real a cosecha del árbol  $j$ , lote  $i$ ;

$\mu$ : media general;

$H_i$ : efecto del lote  $i$ , con  $i = 1, 2$ ;

$\varepsilon_{ij}$ : error experimental, dado por el árbol  $j$ , lote  $i$ , con  $j = 1, 2, \dots, 50$  para el Lote 1 y  $j = 1, 2, \dots, 40$  para el Lote 2.

Para la estimación del modelo propuesto se utilizó el paquete 'anova' del programa R (R Development Core Team, 2018).

Luego el tamaño de muestra “ $n$ ” se calculó con la fórmula antes mencionada; el procedimiento consistió en ir variando los errores y determinando el  $n$  para cada uno de ellos, y luego se graficaron valores de  $n$  para diferentes errores ( $c$ ) expresados en porcentaje (%), en un eje de coordenadas cartesianas. Finalmente, la forma en que

se determinó el tamaño de muestra apropiado, consistió en seleccionar el tamaño de muestra de árboles visualizando en el gráfico el punto en el cual se logra un error de estimación inferior al 10% (Avanza, 2010).

### Análisis Geoestadístico: Estimación de la distancia de muestreo

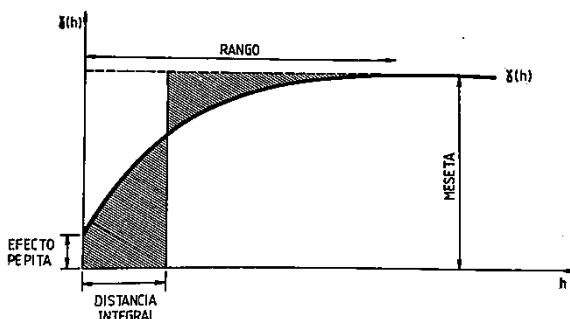
Mediante el análisis geoestadístico se determinó la distancia de muestreo. El primer paso se denomina análisis estructural y consiste en determinar la continuidad espacial de la carga por medio de un Variograma Experimental o Variograma Empírico, el cual se describe como:

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(s_i) - Z(s_j))^2$$

donde:  $\hat{\gamma}(h)$  es la semivarianza para todas las muestras localizadas en el espacio separado por el intervalo de distancia  $h$ ,  $Z(s_i)$  y  $Z(s_j)$  son los valores de la variable observados en los puntos de muestreo  $s_i$  y  $s_j=(s_i + h)$  respectivamente,  $h$  es la separación entre puntos de muestreo y  $N(h)$  es el número total de pares de muestras separados por un intervalo de distancia  $h$  (Maestre *et al.*, 2014).

Esta función representa la variación en pequeña escala, midiendo la relación entre 2 puntos separados a una distancia  $h$ , expresada como la varianza de la diferencia entre tales pares.

Las partes que conforman el variograma son: el efecto pepita ( $C_0$ ), que representa el salto o discontinuidad en el origen; la meseta o umbral ( $C_0+C_1$ ), punto en el cual las distancias se estabilizan; y el rango ( $a$ ) que es el valor de distancia  $h$  donde el variograma alcanza la meseta y la correlación entre variables es nula.



**Figura N° 2:** Parámetros del semivariograma (Samper Calvete *et al.*, 1996).

Una vez obtenido el Variograma Empírico se ajustó a un Variograma Teórico, el cual representa la tasa media de cambio de la carga con la distancia.

### Modelos teóricos de Variogramas

El semivariograma teórico que no es más que una función matemática con una expresión analítica, que se emplea frecuentemente para representar semivariogramas reales (Samper Calvete *et al.*, 1996).

A continuación se presentan los diferentes modelos de variograma calculados con el Sofware GS+:

$$\gamma(h, \theta) = \begin{cases} 0 & \text{si } \|h\| = h = 0 \\ c_0 + c_1 \|h\| & \text{si } \|h\| = h \geq 0 \end{cases} \quad \text{donde } \theta = (c_0, c_1) \text{ si } c_i \geq 0 \text{ para } i=0,1$$

la función cuando  $\|h\| \rightarrow \infty$ , y así no corresponde a un proceso estacionario.

$$\gamma(h, \theta) = \begin{cases} 0 & \text{si } \|h\| = h = 0 \\ c_0 + c_s \left( \frac{3\|h\|}{2a_s} \right) - \frac{1}{2} \left( \frac{\|h\|}{a_s} \right)^3 & \text{si } 0 < h \leq a_s \\ c_0 + c_s & \text{si } h > a_s \end{cases}$$

**Modelo esférico:**

donde  $\theta = (c_0, c_s, a_s)$  y  $c_i \geq 0$  para  $i = 0, s$  y  $a_s \geq 0$ . En este modelo,  $c_0 + c_s$  es la meseta y  $a_s$  es el rango. Su comportamiento en el origen es lineal y es uno de los más usados en la práctica debido a su facilidad para ajustarse a los datos.

$$\gamma(h, \theta) = \begin{cases} 0 & \text{si } \|h\| = h = 0 \\ c_0 + c_e \left( 1 - \exp \left( - \frac{\|h\|}{a_e} \right) \right) & \text{si } \|h\| = h > 0 \end{cases}$$

**Modelo exponencial:**

donde  $\theta = (c_0, c_e, a_e)$  y  $c_i \geq 0$  para  $i = 0, e$  y  $a_e \geq 0$ . El valor de  $c_0+c_e$  es la meseta que se alcanza en forma asintótica, siendo  $\sqrt{3}a_e$  su rango. Se caracteriza por su comportamiento parabólico en el origen.

**Modelo gaussiano:**  $\gamma(h, \theta) = \begin{cases} 0 & \text{si } \|h\| = h = 0 \\ c_0 + c_1 \left( 1 - \exp \left( -\frac{\|h\|^2}{a} \right) \right) & \text{si } \|h\| = h > 0 \end{cases}$

donde  $\theta = (c_0, c_1, a)$  t y ci  $\geq 0$  para i = 0, 1 y a  $\geq 0$ . El valor de  $c_0 + c_1$  es la meseta que se alcanza en forma asintótica. Se caracteriza por su comportamiento parabólico en el origen.

Una vez que el modelo ha sido identificado, el rango, es decir  $a_r$ , representa la distancia a partir de la cual no hay correlación entre las cargas de los árboles y de esta manera se puede sugerir un esquema de muestreo que tenga en cuenta esta distancia.

Se realizó también un mapa de rendimiento, donde la posición de los puntos se calculó con la expresión denominada 'Posición real' (Hernández Alonso *et al.*, 2006),

$$\text{Posición real} = \text{grados} + (\text{minutos}/60) + (\text{segundos}/3600)$$

Los análisis geoestadísticos se realizaron con paquetes específicos como 'GeoR' del programa RStudio (R Development Core Team, 2018) y GS+ (Gamma Design Software, 2004. Versión de prueba).

El grado de dependencia espacial se analizó con la "relación Efecto Pepita/Meseta" para la carga de frutos de 'Salustiana' en los puntos muestreados para la campaña en estudio. Esta relación, a modo descriptivo, interpreta cuanto es la dependencia con respecto a su relación de fuerza expresando el resultado en porcentaje.

## 4. RESULTADOS Y DISCUSIÓN

### Estadística Descriptiva

#### Número de Frutos por árbol al momento de cosecha

En las tablas 1 y 2 se presentan medidas resumen de la carga en todo el lote comercial y luego se tratan por separado los lotes L1 y L2, como se mencionó en materiales y métodos.

**Tabla N° 1:** Medidas descriptivas del número de frutos por árbol en el lote comercial de naranjo Salustiana en Santa Rosa, Corrientes. Campaña 2013-2014.

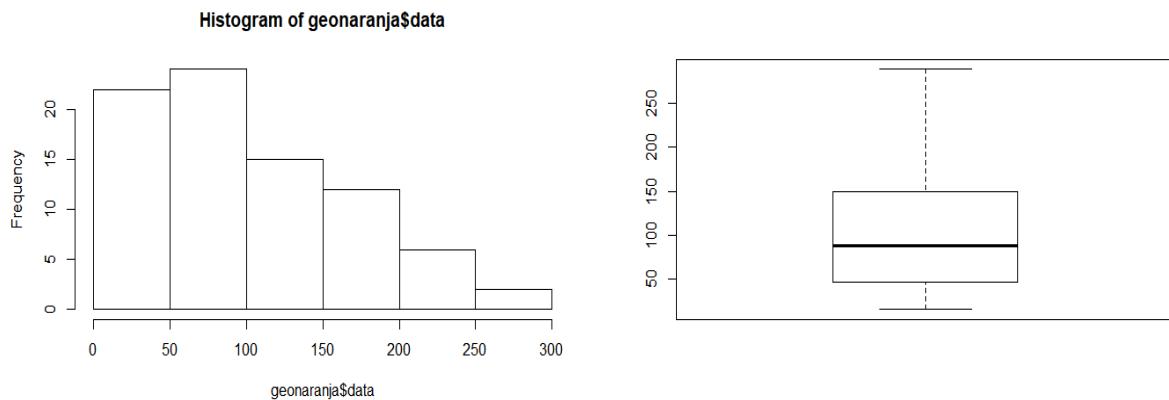
Variable	N	Media	D.E.	Var(n-1)	CV (%)	Mín	Máx	Asimetría	Kurtosis
Carga	81	102,31	67,47	4551,59	65,94	15	289	0,68	-0,30

**Tabla N° 2:** Medidas descriptivas de la Carga (número de frutos por árbol) de naranjo Salustiana por lote en Santa Rosa, Corrientes. Campaña 2013-2014.

Lote	N	Media	D.E.	Var(n-1)	CV (%)	Mín	Máx	Asimetría	Kurtosis
Lote 1	32	65,91	49,47	2447,18	75,06	15	214	1,26	1,25
Lote 2	49	126,08	67,42	4545,16	53,47	18	289	0,38	-0,56

Los resultados expresados en las tablas 1 y 2 muestran que el análisis del lote completo representa un promedio de los lotes 1 y 2, el cual no sería representativo ya que ambos lotes presentan grandes diferencias. En el Lote 1 se observa mayor variabilidad y menor producción, 65,91 frutos por árbol, mientras que en el Lote 2 la media es prácticamente el doble 126,08 frutos por árbol con un menor coeficiente de variación (53,47 %). Ambos lotes presentan gran variabilidad y una distribución platicúrtica indicada por el valor de la kurtosis la cual es menor que 3.

En la **Figura N° 3** se muestran el histograma y el gráfico de cajas de la carga en el lote completo, teniendo en cuenta que si bien muestran un promedio poco representativo de los lotes 1 y 2, para representar la distribución del número de frutos por árbol, se considera apropiado trabajar con un  $n=81$  sin particionar.



**Figura N° 3:** Histograma de frecuencias y gráfico de cajas del número de frutos cosechados por árbol.

### Estimación de Variabilidad

Utilizando el paquete 'anova' del programa R (R Development Core Team, 2018) se realizó la estimación del modelo de análisis propuesto para estimar la variabilidad de la carga, en función de las esperanzas de los distintos cuadrados medios involucrados.

**Tabla N° 3:** Medias y desvío estándar (DE) de la carga a cosecha en número de frutos por árbol y kilogramos de frutos cosechados por planta, por lote.

Lote	Carga media	
	Nº frutos árbol <sup>-1</sup> ± DE	Kg planta <sup>-1</sup> ± DE
1	65,91 ± 49,87	12,72 ± 9,44
2	126,08 ± 67,42	36,47 ± 18,84

En la **Tabla N° 3**, se presentan las medias y desvíos estándar (DE) de la carga a cosecha por lote. Los valores de DE asociados a cada una de las medias evidencian la gran dispersión de la carga dentro del lote, y también entre lotes. Ésto implica que para una misma temporada, los lotes y los árboles produzcan cantidades diferentes de frutos por árbol. Es necesario continuar el estudio de otros factores durante más ciclos productivos que permitan explicar la variabilidad encontrada.

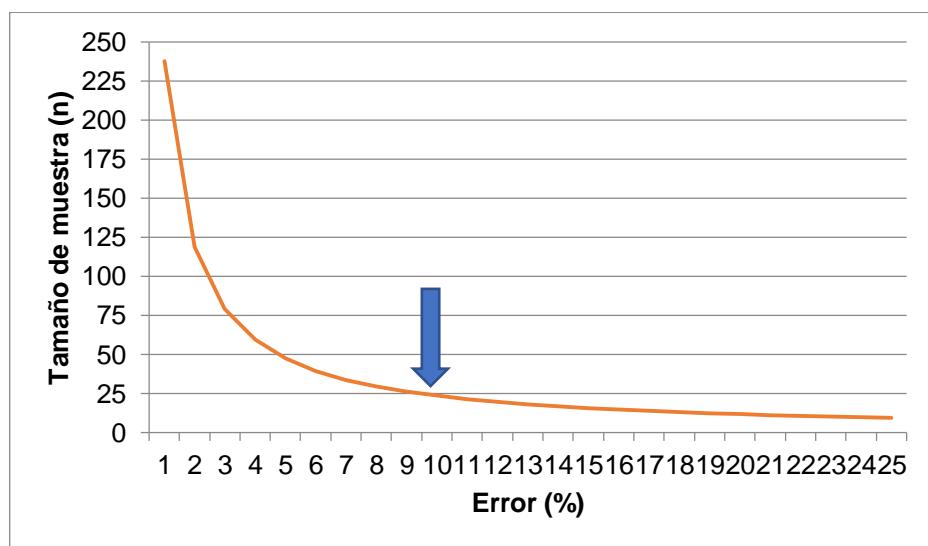
### Estimación de Tamaño de Muestra

Para el cálculo del tamaño de muestra en función a los diferentes porcentajes de error, se utilizó el mayor valor de variabilidad hallado, correspondiente al Lote 2 y se determinó número de árboles ( $n$ ) que integraron la muestra, mediante una aproximación para muestreo aleatorio simple para la estimación de carga media.

Para obtener la **Figura N° 4** se realizaron cálculos del  $n$  variando los errores y se graficaron valores de  $n$  para diferentes errores ( $c$ ) expresados en porcentaje (%). Los mismos se presentan en la siguiente **Tabla N° 4**.

**Tabla N° 4:** Valores de error (%) respecto al promedio de carga asociado y Tamaño de muestra ( $n$ ) en número de árboles correspondiente a cada error.

Error (%)	Tamaño de muestra (n)
1	264
2	132
3	88
4	66
5	53
6	44
7	38
8	33
9	29
10	26
11	24
12	22
13	20
14	19
15	18
16	17
17	16
18	15
19	14
20	13
21	13
22	12
23	11
24	11
25	11



**Figura N° 4:** Tamaño de muestra ( $n$ ) asociado a diferentes valores de error expresado como % respecto al promedio de carga asociado ( $c$ ).

En la **Tabla N° 4** y su correspondiente **Figura N° 4** se puede observar que si quisieramos que el error no fuera mayor al 5% es necesario tomar una muestra de 53 árboles por lote, o más. Considerando márgenes de error del orden del 10% se podrían seleccionar 25 árboles o más. Estas situaciones demandan una gran cantidad de tiempo para realizar las evaluaciones de carga, por lo cual desde el punto de vista económico resultan poco prácticas y costosas. Se debe buscar un equilibrio entre la precisión y los márgenes de errores posibles. Habría que analizar si el costo de obtener una muestra más está justificado por la ganancia en la precisión.

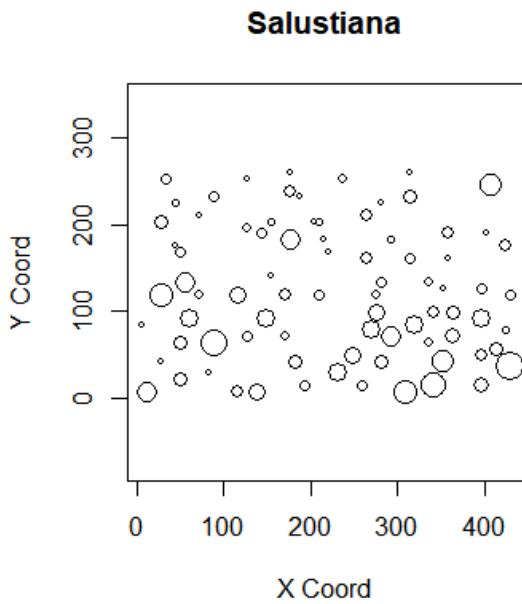
Con valores de  $n=25$  se podría obtener un error del 10,5%, correspondiente a  $\pm 10$  frutos respecto a la carga promedio. A partir de dicho punto ya se observa que, la reducción del tamaño de muestra significa menor esfuerzo pero con una diminución importante en la precisión.

Avanza (2010), encontró como adecuado un tamaño de muestra entre 18 y 21 árboles por huerto para naranjo 'Valencia late' en Corrientes, con lo cual se consigue una precisión entre el 89 y 90 % en la estimación de carga de frutos de naranjo de esa variedad.

Por lo expuesto anteriormente, se considera apropiado un tamaño de muestra entre 22 y 26 árboles por lote, con lo cual se consigue una precisión entre el 88 y 90 % en la estimación de carga de frutos de naranjo dulce 'Salustiana'.

### Análisis Geoestadístico

La **Figura N° 5** es un gráfico denominado de "Burbujas", se representan los puntos de acuerdo a la posición, la cual está indicada por las coordenadas X e Y, cada punto queda representado en dicho espacio de coordenadas y cada círculo tiene tamaño proporcional al valor de la variable. Teniendo en cuenta que el análisis geoestadístico implica ajustar modelos a una estimación de varianza, se decidió realizar el análisis geoestadístico del lote completo, con la totalidad de las muestras, de manera tal de realizar las estimaciones con mayor la cantidad de datos.



**Figura N° 5:** Carga en la cosecha de los 81 árboles del lote completo con representación proporcional del número de frutos por árbol a cosecha.

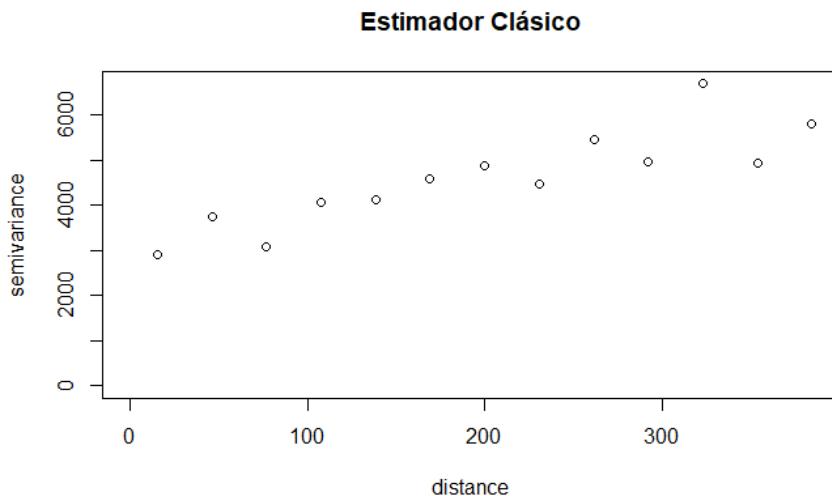
Se puede observar la gran variabilidad en el lote, vale aclarar que los valores altos en las coordenadas de Y representan el Lote 2.

Las distancias mínima y máxima fueron 5,50 y 462,02 metros respectivamente. Se observa mayor producción y homogeneidad en el sector Sur-Oeste, que corresponde al mencionado lote.

Luego se estudió la continuidad espacial de la carga calculándose un variograma experimental.

### El variograma

En la **Figura N° 6** se muestra la estimación del variograma, obtenida bajo el supuesto de constancia de la media mediante el estimador del método "de los momentos" o estimador de Matheron, que recibe el nombre de estimador clásico de variograma. La misma considera a los puntos de muestreo dispuestos sobre una grilla regular y con una distribución normal.



**Figura N° 6:** Variogramma empírico con estimador clásico.

La búsqueda de un variograma que recoja correctamente la variación espacial y la interdependencia del proceso aleatorio, está íntimamente relacionada con la variación espacial de la muestra ( $Z(s_1)$ ,  $Z(s_2)$ , . . .  $Z(s_n)$ ). El espacio de todas las posibles elecciones de variogramas es muy amplio, y en la práctica se restringe a una familia paramétrica en la que se procura hallar unos parámetros que mejor se acoplen a la muestra.

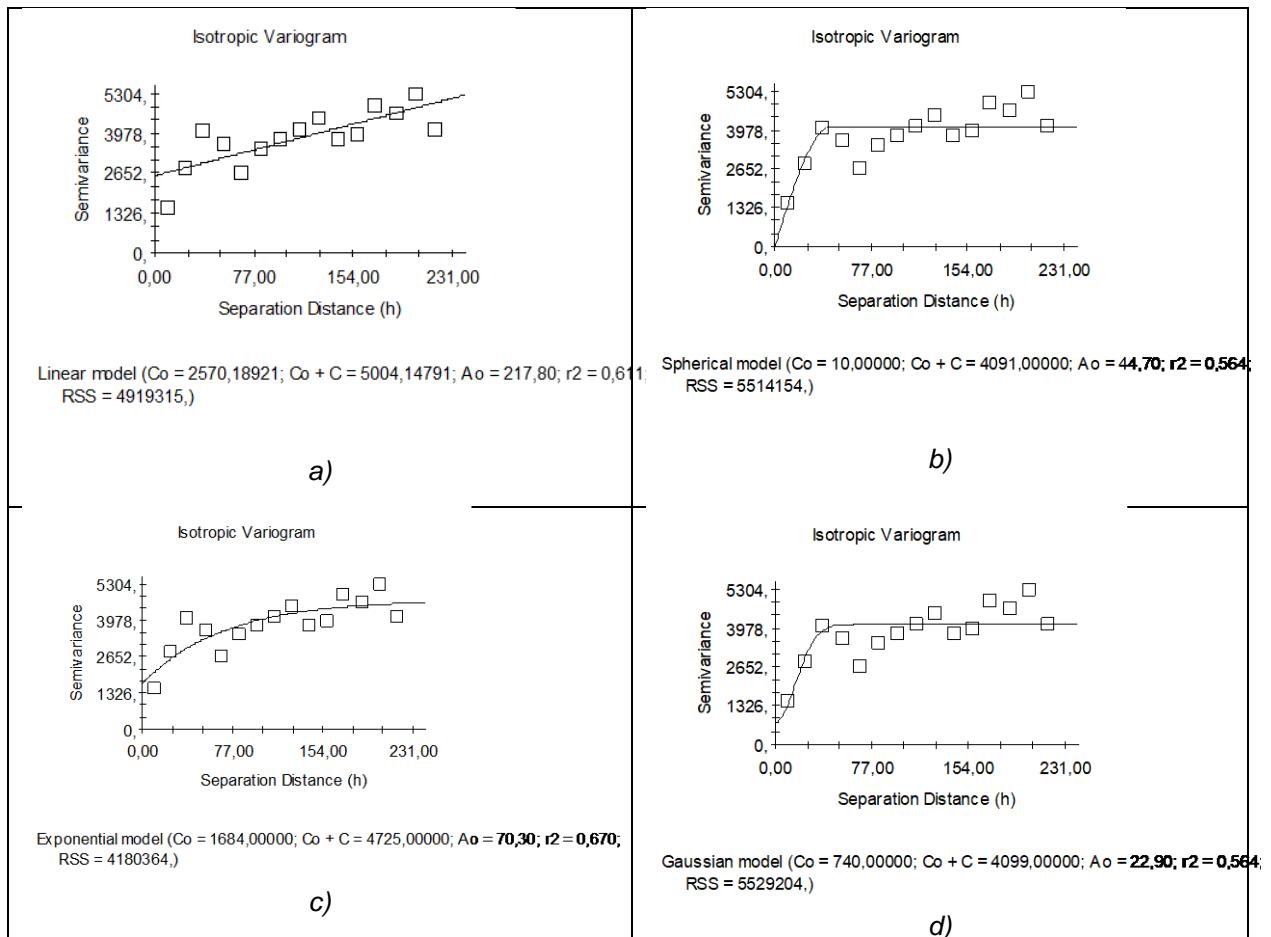
La **Tabla N° 5** muestra los parámetros de los distintos modelos de variogramas calculados con el software GS+.

**Tabla N° 5:** Parámetros de los distintos modelos de variogramas.

Modelo de Variograma Tipo	Efecto Pepita (Co)	Meseta (Co+C)	Rango (A)	Residual SS
Lineal	2570,19	5004,15	217,80	4919315
Esférico	10	4091	44,70	5514154
Exponencial	1684	4725	210,90	4180364
Gaussiano	740	4099	39,66	5529204

Se utilizó como criterio de selección Residual SS o suma de cuadrados residual. Los resultados obtenidos muestran que el ajuste con menores errores cuadráticos es el exponencial por ser el de menor suma de cuadrados residual, menor variación atribuida al error comparándola con los demás variogramas calculados.

En la **Figura N° 7** se presentan los gráficos de ajuste al variograma empírico de los modelos lineal, esférico, exponencial y gaussiano.



**Figura N° 7:** Ajuste de los modelos a) Lineal, b) Esférico, c) Exponencial y d) Gaussiano al variograma empírico.

El modelo que representa la varianza espacial fue el Exponencial que define un rango de 210,90 metros y una varianza total de 4725 (**Tabla N° 5**). El rango representa la máxima distancia a partir de la cual no existe dependencia espacial. Esto indica que, los muestreos deberían realizarse a distancias superiores a los 210,9 metros. En este caso que el marco de plantación es de 7 metros por 3,5 metros los árboles vecinos tienen dependencia espacial en el número de frutos por árbol.

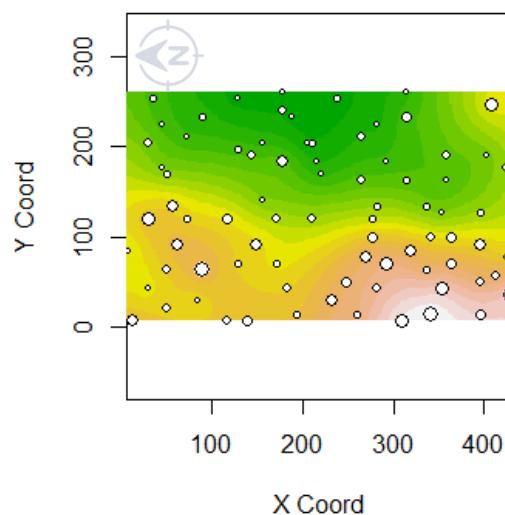
Antes de la realización de los mapas, a modo descriptivo, se calculó el nivel de dependencia espacial de los datos, dividiendo el valor del efecto “nugget” y el “sill”. Con este cociente se interpreta cuanto es la dependencia con respecto a su relación

de fuerza expresando el resultado en porcentaje. Si el resultado es inferior al 25%, el nivel de dependencia espacial es elevado, si es entre 26 y 75%, el nivel de dependencia espacial es moderado, y si es superior al 76%, el nivel de dependencia espacial es bajo (Bolfe *et al.*, 2010, Sánchez-Pale *et al.*, 2011, citados por Cabrera Brunetti, 2019 *in litt.*).

**Tabla N° 6:** Porcentaje de relación de fuerza y grado de dependencia espacial de carga de frutos de naranjo 'Salustiana' para la Campaña 2013-2014.

Modelo de Variograma Tipo	Efecto Pepita (Co)	Meseta (Co+C)	Pepita/Meseta (%)	Grado de dependencia espacial
Lineal	2570,19	5004,15	51,36	MODERADO
Esférico	10	4091	0,24	ELEVADO
Exponencial	1684	4725	35,64	MODERADO
Gaussiano	740	4099	18,05	ELEVADO

Del análisis de la dependencia espacial “relación Efecto Pepita/Meseta” para carga de frutos de 'Salustiana', se observa en la **Tabla N° 6** que el grado de dependencia espacial, en casi todos los casos fue de elevado a moderado. Estos grados de dependencia alcanzados apuntan a la eficiencia en el muestreo realizado reflejada en la distribución espacial.



**Figura N° 8:** Mapa de isolíneas del lote comercial de Naranjo dulce 'Salustiana' ubicado en la localidad de Santa Rosa, Corrientes, Argentina.

Examinando resultados en el mapa de isolíneas **Figura N° 8** se observan valores puntuales que representan magnitudes de la producción.

Cabe destacar que los árboles representaron una grilla irregular de muestras ya que se obtuvieron de un muestreo aleatorio.

Entre las magnitudes de producción relevadas, existen manchones en blanco donde se desconoce el valor real y hay que estimar, ésto se refleja en la **Figura N° 8**.

En el sector del campo donde se tomaron más puntos de muestreo, hay más precisión y ésto corresponde al Lote 2.

## 5. CONCLUSIONES

Se determinó que con un tamaño de muestra entre 22 y 26 árboles por lote, se consigue una precisión entre el 88 y 90 % en la estimación de carga de frutos de naranjo dulce 'Salustiana'.

El modelo seleccionado es el exponencial, que explica el crecimiento de la dependencia espacial con respecto a la distancia entre las observaciones.

La distancia a partir de la cual no hay dependencia espacial es de 210,90 metros que corresponde al valor del rango y que es igual a la distancia para la cual el semivariograma toma un valor igual al 95% de la meseta.

El rango del modelo teórico ajustado al semivariograma seleccionado es un indicador de fuerte dependencia espacial para el caso considerado puesto que supera los 210 metros, lo cual resulta relativamente alto, teniendo en cuenta que la distancia entre los extremos del lote estudiado es de 254,28 metros por 283,43 metros.

## BIBLIOGRAFÍA

- AGUSTÍ, M. 2003.** Citricultura. Ediciones Mundi-Prensa, Barcelona, pp 39-49.
- AVANZA, M. M. 2010.** Desarrollo de una propuesta metodológica para la predicción de cosecha de naranjo dulce (*Citrus sinensis* L. Osbeck) var. 'Valencia late' en la provincia de Corrientes, Argentina. Tesis Doctoral, Universidad Nacional del Nordeste, Corrientes, 2010. 226.
- BRUNETTI P. C., C. BRUNO, R. ZAPATA, L. TORRES, G. BARBOZA & M. OJEDA. 2014.** Sample size estimation for in situ morphological variability studies of *Lippia integrifolia* (Verbenaceae) in Parque Nacional Talampaya, La Rioja (Argentina). *Bonplandia* 23(1): 15-23.
- BOLFE, E. L.; GOMES, J. B. V. & ROLLEMBERG, H. 2010.** Geoestadística en la producción integrada de frutas en el nordeste de Brasil. *Revista UD y la Geomática*, 2010, v. 2, p. 15-27.
- CABRERA BRUNETTI, S. C., GIMÉNEZ, L. I. & AVANZA, M. M. 2010.** Uso del lenguaje 'R' para el análisis de la variación espacial de la producción de naranjo 'Valencia late'. XXIº Reunión de Comunicaciones Científicas, Técnicas y de Extensión - UNNE.
- CRAVIOTTI, C. 2010.** La producción familiar en la globalización agroalimentaria: La diferenciación social en la citricultura del noreste argentino. *Revista Europea de Estudios Latinoamericanos y del Caribe/European Review of Latin American and Caribbean Studies*, 65-85.
- COCHRAN, W.G. 1982.** Técnicas de muestreo. México: Compañía Editorial Continental, 1982. Pp. 513.
- DANZA, A. 2017.** Perfil de Mercado de Cítricos Dulces 2017. Ministerio de Agroindustria. Presidencia de la Nación. Subsecretaría de Mercados Agropecuarios. Pp. 16.
- DIBLASI, A. & BOWMAN, A. W. 2001.** On the use of the variogram in checking for independence in spatial data. *Biometrics*, 57(1), 211-218.

**FEDERCITRUS 2018.** Federación Argentina del Citrus. La Actividad Cítrcola Argentina. Bs.As. Argentina. [en linea]. <http://www.federcitrus.org/> wp-content/uploads/2018/05/Actividad-Citricola-2018.pdf (Consulta: Noviembre de 2019).

**FOLWELL, R. J., SANTOS, D. E., SPAYD, S. E., PORTER, L. H. & WELLS, D. S. 1994.** Statistical technique for forecasting Concord grape production. American journal of enology and viticulture, 45(1), 63-70.

**GAMMA DESIGN SOFTWARE 2004.** GS+: Geostatistics for the environmental sciences.

**GIMENEZ, L. I., PETROPOULOS, C. & PIPERIGOU, V. E. 2013.** The use of spatial statistics in Orange tree production. Greek Statistical Institute, Proceddings of the 25<sup>th</sup> Panhellenic Statistics Conference, pp.285-294.

**GOMEZ, K.A. & GOMEZ, A.A. 1984.** Statistical procedures for agricultural research. 2nd ed. Ames: John Wiley & Sons, 1984. 680.

**Google Earth Pro 7.3.2.5776 (64-bit).**

**GRAVINA, A., ARBIZA, H., FERENCZI, A., ORLANDO, L., SEVERINO, V., GAMBETTA, G., ... & AGUSTÍ, M. 2000.** Estudio del comportamiento alternante de la naranja "Salustiana" [*Citrus sinensis* (L.) osb.] en diferentes condiciones productivas. Agrociencia-Sitio en Reparación, 4(1), 17-22.

**MAESTRE GIL, F. T. & ESCUDERO ALCANTARA, A. 2014.** Introducción al Análisis Espacial de Datos en Ecología y Ciencias Ambientales: Métodos de aplicación, Cap. 6, Madrid. ISBN 9788490318263. Editorial Dykinson, S.L. 850 p.

**OTZEN, T. & MANTEROLA, C. 2017.** Técnicas de Muestreo sobre una Población a Estudio. International Journal of Morphology, 35(1), 227-232. <https://dx.doi.org/10.4067/S0717-95022017000100037>.

**ORDUZ R., J.; LEÓN, G. & ARANGO W., L. 2009.** Evaluación de variedades de naranjas para la Altillanura colombiana. Corpoica. 2009. 16 págs.

**PALACIOS, J. 2005.** Citricultura. 520 págs. ALFA BETA. ISBN: 987-4-8326-7. Argentina.

**REDONS, J. R.; ARCILA, P.J. & MONTOYA, R.E.C. 2008.** Estimación de la producción de café con base de los registros de floración. Cenicafé 59 (3) pp 28-259.

**REEB, P.; BRAMARDI, S. & ALVAREZ, O. 2003.** Estudio de la variabilidad de la producción de manzanas red delicious en los montes frutales del alto valle de río negro, argentina. Agro sur, ene. 2003, vol.31, no.1, p.21-26. ISSN 0304-8802.

**SAMPER CALVETE, F. J. & CARRERA-RAMÍREZ, J. 1996.** Geoestadística. Aplicaciones a la hidrología subterránea. 2<sup>a</sup>. Edición. Centro Internacional de Métodos Numéricos en Ingeniería, Barcelona.

**SANCHEZ-PALE, J. R.; RAMIREZ-DAVILA, J. F.; GONZALEZ-HUERTA, A. & DE LEON, C. 2011.** Distribución espacial del carbón de la espiga (*Sporisorium reilianum*) del maíz en México. Cien. Inv. Agr. , 2011, v. 38, n. 2, p. 253-263.

**SCHEAFFER, R.L.; MENDENHALL III, W. & LYMAN OTT, R. 2007.** Elementos de muestreo. 6ta Edición. International Thomson Editors Spain, 455.

**SINCLAIR, T. R. & SELIGMAN, N. A. 2000.** Criteria for publishing papers on crop modeling. Field Crops Research, 68(3), 165-172.