Ajuste de curvas de producción de claveles y mini claveles en una empresa productora y exportadora en la Sabana de Bogotá mediante datos funcionales

Evelyn Bedoya Dorado

Universidad de Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia

2018

Ajuste de curvas de producción de claveles y mini claveles en una empresa productora y exportadora en la Sabana de Bogotá mediante datos funcionales

Evelyn Bedoya Dorado

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadística

Directores:
PhD. Javier Olaya Ochoa y MSc. Diego Alejandro Tovar Rios

Universidad de Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia

2018

Dedicatoria

A mi madre en la tierra, mi motivo
A mi padre en el cielo, mi fortaleza

Agradecimientos

Este proyecto de grado representa un trabajo arduo y hermoso en el cual participaron diferentes personas, guiándome, enseñándome, corrigiendo, opinando, acompañándome y motivándome, por lo cual quiero agradecerles en este apartado.

En primer lugar, al señor Alberto Torres de la empresa Aposentos Flores SAS, mi más sincero agradecimiento por haber confiado en mi para darle solución a su problema, por su hospitalidad, paciencia y disposición para brindarme la información necesaria para el desarrollo de este proyecto.

Al profesor Diego Alejandro Tovar, quien desde el primer momento aceptó mi propuesta, me apoyó, guió y acompañó durante este año en cada actividad del proyecto. Por compartirme sus conocimientos, experiencia, compromiso y apoyo en la elaboración de este documento.

Al profesor Javier Olaya, por su compromiso con este proyecto, por asesorarme, enseñarme y aconsejarme durante este año. Por su dedicación y esmero en que todo saliera bien y en especial por su aporte de maravillosas ideas.

Por último a mis compañeros, los que me compartieron sus conocimientos, los que tuvieron paciencia e interés en explicarme, y los que cada día me motivaron y se alegraron con cada etapa superada de este proyecto. En especial a Michael Rodriguez, por su respaldo y amistad, gracias por sus consejos, y sobretodo por animarme en este año de trabajo.

Índice general

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

Colombia tiene la mayor variedad de flores exóticas que permite estar a la vanguardia de las tendencias de consumo mundiales. Con más de 40 años de experiencia, es el segundo país exportador de flores en el mundo después de Holanda (Procolombia, 2016). Según Asocolflores (2016a), los departamentos de Cundinamarca y Antioquia son los responsables de la siembra y producción del 72 % y 27 % de las flores respectivamente. Para el año 2015; el sector floricultor contaba con 7.000 hectáreas cultivadas y dedicadas al corte de flores frescas para la exportación.

Gracias a un plan de promoción a la comercialización y sostenibilidad de las flores colombianas, apoyado por el Ministerio de Agricultura y Desarrollo Rural (de ahora en adelante MinAgricultura), se ha logrado fortalecer el posicionamiento de las flores colombianas en sus principales mercados, logrando mantener su posición a nivel mundial. Estados Unidos, durante los últimos 10 años había sido el principal destino de las flores colombianas, el país exportaba el 85 % de su producción; hoy, ese porcentaje con el desarrollo de nuevos mercados representa el 75 % y el 25 % restante llega a 89 países entre los cuales se destaca Rusia, Reino Unido, Japón y Canadá (Ministerio de Agricultura, 2015a). Durante el año 2016, se registraron exportaciones por valor de 1.312 millones de dólares y 235.000 toneladas, lo cual significó un crecimiento del 1.3 % en valor y 5.7 % en volumen con respecto al año 2015. Las principales especies exportadas fueron rosa (23.3 %), clavel (16.2 %), crisantemo (11.2 %), alstroemeria (5.9 %) y hortensia (5.5 %) (Asocolflores, 2016b). Colombia es el principal productor de claveles y es el único que se centra en la exportación. Con 1.106 hectáreas de campos de clavel, exporta el 90 % de los claveles principalmente a EE.UU (42 %), Japón (18 %), Holanda (8 %), Inglaterra y Rusia (7 %) (FloralHolland, 2016).

La empresa C.I. Flores de Aposentos S.A.S. es una empresa productora y exportadora líder de claveles y mini claveles con más de 70 hectáreas de invernaderos ubicados al norte de Bogotá. Actualmente, disponen de índices de producción por semana de vida de la planta, en cada siembra y para todas las variedades de claveles y mini claveles entre los años 2012 y 2016; con base en estos índices, la empresa ha construido de forma empírica un modelo de producción para cada variedad los cuales usan para predecir sus producciones. En este trabajo de grado, se busca ajustar las curvas de producción de la empresa y

encontrar un modelo estadístico que represente los índices de producción reales ya que gráficamente el modelo empírico no se ajusta al comportamiento de la producción. Dado que en los índices de producción no se identifica un patrón, y que éstos no se comportan de forma lineal o se ajustan a alguna distribución de probabilidad conocida, es necesario utilizar métodos no paramétricos para obtener el modelo estadístico.

Capítulo 2

Planteamiento del Problema

En los 28 municipios de la Sabana de Bogotá y Cundinamarca, existen fincas productoras de flores bajo invernaderos. La empresa C.I. Flores de Aposentos S.A.S. es una empresa líder productora con 3 fincas que producen claveles y 2 fincas que producen mini claveles, con una superficie total de más de 70 hectáreas de invernaderos. Estas fincas se dividen en bloques (recinto cerrado, cubierto y acondicionado para mantener una temperatura regular que proteja las plantas el cual está dividido por camas en las cuales se encuentran sembradas las plantas) en los cuales se siembran 48 variedades de claveles y 31 variedades de mini claveles.

Las plantas que producen estas flores tienen un ciclo de vida de aproximadamente 100 semanas, las cuales se dividen en 4 periodos:

- Pico: es el periodo de máxima producción de las flores, se alcanza aproximadamente entre la semana 23 y 29 de vida.
- Valle: periodo de receso después del Pico, aproximadamente entre la semana 30 y 45 hay una escasa producción de flores.
- Repico o pico 2: periodo en el cual hay un incremento en la producción de flores pero no alcanza el máximo, se da aproximadamente entre la semana 50 y 61.
- Permanente: en este periodo la planta empieza a reducir la producción de flores hasta que muere. Este periodo se alcanza aproximadamente después de la semana 70.

Para cada una de las variedades de claveles y mini claveles, se ha reportado un índice de producción de cada semana de vida de las plantas sembradas en los diferentes bloques y en los dos semestres (enero-junio, julio-diciembre) desde el año 2012; el índice se construye dividiendo el número de flores obtenidas durante una semana sobre el número de plantas sembradas en esa semana. A partir de estos índices, la empresa creó una curva (modelo) de manera empírica para cada variedad que mejor representa la producción en cada semana de vida. Estas curvas se usan como herramienta para poder pronosticar el cumplimiento de la

demanda en el mercado, es decir, por medio de ellas se puede proyectar cuántas flores se pueden producir en cada semana y así poder identificar si se cubrirá o no la demanda en un periodo determinado del año.

Aunque estas curvas evidencian el comportamiento de los 4 periodos de vida de las plantas, se pudo identificar en datos ofrecidos por la empresa que dichas curvas no son adecuadas, debido a que no representan los verdaderos índices de producción. En algunas variedades los errores del modelo empírico respecto a los datos reales son muy altos, se observan curvas por encima y otras por debajo de lo real. Esto implica pronósticos fallidos sobre la cantidad de plantas que se deben producir en una semana.

Los errores en los pronósticos pueden generar exceso en la producción, lo que articula pérdidas de flores y por ende de dinero. Igualmente, los errores pueden enfocar la producción hacia niveles inferiores de la demanda de los clientes, afectando la relación con estos y la rentabilidad de la empresa.

Teniendo en cuenta que Colombia está posicionado como el mayor productor mundial de claveles (Ministerio de Agricultura, 2015b), el papel de las empresas y de la gestión de su producción es fundamental. Por ello, este proyecto de grado tiene como objetivo realizar un ajuste de las curvas de producción de la empresa C.I. Flores de Aposentos S.A.S. para que dispongan de curvas precisas, teniendo en cuenta la variabilidad, para producir las cantidades requeridas por sus clientes durante todo el año y en especial en las fechas de celebraciones.

Se pudo observar que el comportamiento de los índices de producción para cada variedad no se ajusta a ninguna distribución de probabilidad conocida ni es de forma lineal para usar los modelos tradicionales de regresión. Lo anterior representa el problema estadístico el cual podría abordarse con métodos no paramétricos para la construcción de las curvas por medio de funciones, siendo una alternativa de nuestro interés.

Capítulo 3

Justificación

Las estimaciones que se hacen actualmente en la empresa C.I. Flores de Aposentos S.A.S. de la producción de claveles y mini claveles, presentan altas diferencias con respecto a los resultados reales, generando consigo ineficiencias económicas por causa de la sobre producción (más flores de las necesitadas) o por el contrario, insuficiencia en la producción. La empresa cuenta con un reporte de índices de producción para las diferentes variedades de claveles y mini claveles en cada semana de vida de las plantas. Estos índices pueden ser usados para predecir la producción de las variedades de claveles a lo largo del año, sin embargo, la empresa usa para estimar sus producciones un modelo ajustado por una curva de forma empírica, la cual no representa la información real dada por los índices.

Para C.I. Flores de Aposentos S.A.S., poder predecir la cantidad de claveles que se cosecharán en fechas específicas, representa una ventaja competitiva y una herramienta importante para identificar si la producción de la empresa puede cubrir una demanda y establecer con mayor seguridad la cantidad de claveles a ofertar. Un modelo estadístico sería esa herramienta útil, la cual sirve también como estrategia de mejora del servicio al cliente, pues por medio del pronóstico de producción se pueden realizar las entregas en las fechas acordadas con las cantidades requeridas por los clientes. Por otro lado, en fechas importantes del año como fiestas y celebraciones nacionales e internacionales, se requiere manipular el cultivo para obtener la producción deseada para cubrir la demanda de claveles en esos días, por lo cual se vuelve indispensable modelar el comportamiento de la producción de claveles y miniclaveles en la empresa.

Por lo anterior, se pretende usar un método estadístico que permita ajustar las curvas de producción de cada una de las variedades de claveles y mini claveles, con el fin de proporcionar una herramienta a la empresa, la cual permita a su vez realizar correctas estimaciones para la adecuada toma de decisiones sobre el uso óptimo de los recursos y la planeación de las diversas actividades. Se espera con esto, mejorar las relaciones con los clientes en el cumplimiento de las obligaciones pactadas, minimizando perdida por incumplimientos o excesos en la producción.

Capítulo 4

Objetivos

4.1. Objetivo General

Construir un modelo estadístico que permita ajustar las curvas de producción de la empresa C.I. Flores de Aposentos S.A.S., utilizando información registrada de la variedad, ubicación, cantidad y fecha de siembra, como también indicadores de producción semanal de las plantas.

4.2. Objetivos Específicos

- Estimar la curva de producción por medio de la construcción de datos funcionales para 20 variedades de claveles y mini claveles.
- Determinar los valores entre los cuales fluctúa la curva promedio de producción para 20 variedades de claveles y mini claveles.

Capítulo 5

Antecedentes

Modelo de proyección de la producción de rosas basado en las curvas de crecimiento de las plantas

AUTORES: Juan José Vila Arboleda. Universidad de la Salle. 2009

El modelo de proyecciones es una herramienta de ayuda para las empresas floricultoras del país ya que esta da a conocer los volúmenes de producción que puedan obtener estas empresas, conocer este volumen de producción es importante ya que las empresas pueden anticipar todas sus tareas y labores que se presentan dentro de estos cultivos. Las variedades estudiadas obtuvieron un porcentaje de acierto del 97,37 % en las producciones totales proyectadas, pero al observar el estudio semanal en las variedades éstas mostraron que tiene un margen de más o menos 17 % ya que en algunos casos y para algunas variedades se obtuvieron sobreproducciones. El modelo de proyecciones fenológico construido de forma empírica mostró un aumento de más del 15 % en las proyecciones semanales y mostró un acierto de 99 % para la producción total al finalizar las 7 semanas de evaluación. Dar a conocer a las empresas floriculturas del país una herramienta, que les permita conocer las producciones de manera anticipada para que puedan desarrollar sus estrategias administrativas y gerenciales de una forma adecuada. Esta herramienta facilita las labores administrativa, ya que se puede tener un mejor flujo en las producciones de rosas y también en obtener un excelente flujo en el manejo de inventarios (Vila Arboleda, 2009).

Ajuste de curvas mediante métodos no paramétricos para estudiar el comportamiento de contaminación del aire por material particulado PM10

AUTORES: Javier Olaya y Jhovana Reina. Revista EIA. 2012

Uno de los principales agentes contaminantes del aire es el material particulado de diámetro aerodinámico inferior a 10 micrómetros, comúnmente conocido como PM10. Su comportamiento varía de forma irregular y temporal en la atmósfera, debido a las actividades humanas, condiciones atmosféricas inestables y fenómenos meteorológicos. El propósito de este estudio es caracterizar con un modelo de suavización no paramétrica el comportamiento del PM10 en el aire a lo largo de un día, teniendo en cuenta el día de la

semana y los niveles de precipitación. El modelo propuesto se ilustra con registros sobre contaminación por PM10 y con datos de precipitación en el norte de Cali, Colombia. Se estiman curvas típicas diarias del comportamiento del PM10 usando suavizadores kernel y spline. El procesamiento se ejecutó con el software estadístico de libre distribución R. Las curvas estimadas permiten observar un comportamiento unimodal del PM10 durante las horas de la mañana, diferenciado por días de la semana y por días con lluvia y sin lluvia. Los modelos permiten caracterizar de manera robusta el comportamiento diario del PM10, teniendo en cuenta observaciones heterocedásticas bajo un escenario de múltiples respuestas por punto de diseño (Reina and Olaya, 2012).

Modelación no paramétrica de la contaminación promedio octohoraria del aire debida al Monóxido de Carbono y al Ozono Troposférico

AUTORES: Javier Olaya, Jhovana Reina, Leidy L. Torres, María C. Paz y Luz A. Pereira. Revista Ingeniería y Competitividad. 2014

Se presenta la modelación del comportamiento típico de la concentración promedio octohoraria del monóxido de carbono (CO) y ozono troposférico (O3) a lo largo de un día, mediante la implementación de un modelo de regresión no paramétrica. Los datos utilizados provienen de las mediciones registradas por el Sistema de Vigilancia de la Calidad del Aire (SVCA) de la ciudad de Santiago de Cali, para los años 2003 y 2006. Los resultados obtenidos muestran que el comportamiento de la contaminación octohoraria por O3 supera el límite máximo permitido de la Norma Colombiana en las estaciones ubicadas en la zona sur de Cali, mientras que la contaminación octohoraria por CO se aproxima al límite máximo permitido en el centro de la ciudad, coincidente con la presencia de fuentes móviles. Las bandas de variabilidad asociadas a la modelación de la contaminación octohoraria del CO, muestran indicios de que existen diferencias del comportamiento de este contaminante por tipo de día y zonas de vigilancia en los años de estudio; mientras que para el O3, el comportamiento difiere por tipo de día y zonas de vigilancia solo en el año 2006 (Olaya et al., 2014).

Capítulo 6

Marco Teórico

Este capítulo se divide en dos partes. Una parte donde se presenta el marco conceptual en el cual se define el contexto donde está situado el problema, es decir, la producción de claveles, el invernadero y el municipio donde se encuentra ubicada la finca productora. La segunda parte de este capítulo es el marco estadístico donde se presentan algunos conceptos básicos de los métodos estadísticos a usar: datos funcionales y tipos de suavización no paramétrica.

6.1. Marco Conceptual

6.1.1. El Cultivo del Clavel

El Clavel

El clavel pertenece al género *"Dianthus"* y a la familia de las *Cariofiláceas*. El *Dianthus Caryphyllus* es originario de la cuenca mediterránea. Anteriormente sólo existía el clavel silvestre, que tras multitud de hibridaciones y procesos de selección se ha convertido en la variedad actual (Rimache, 2011). Las plantas son de crecimiento suelto y vertical de modo que necesitan soporte, tanto los vástagos (rama tierna de un árbol o planta) como los tallos de las flores. Normalmente los claveles se cultivan durante dos años enteros con 18-20 meses de recolección, la técnica más antigua de cultivo de tres años no está justificada puesto que aumenta las plagas y enfermedades lo cual hace que las plantas crezcan demasiado en altura dificultando su manejo (Salinger, 1992).

Dentro de los claveles existen dos tipos (Gómez, 2010):

Clavel estándar

Es el cultivo que como característica cada tallo soporta una sola flor principal, también se conoce como monoflor.

Clavel miniatura

Son las variedades, cuya característica principal, es que cada tallo tiene un promedio de 6 a 8 botones florales.

Condiciones climáticas

Altitud

En términos generales para el clavel estándar se ha tomado como altura ideal la de la Sabana de Bogotá y algunas poblaciones de Boyacá fluctuando entre los 2200 a los 2640 metros sobre el nivel del mar. Para el clavel miniatura la altura en la que presentan mayor desarrollo las plantas está entre los 1600 y 2400 metros sobre el nivel del mar (Gómez, 2010).

Radiación solar

Tiene un importante papel en el proceso de producción del clavel, ya que interviene directamente sobre la calidad, productividad y el estado fitosanitario (prevención y curación de las enfermedades de las plantas) de las plantas. Cuando existe poco brillo solar, la planta se hace más sensible al ataque de hongos, la apertura de la flor es más lenta, la tendencia es producir hijos muy débiles, también influye en la intensidad del color de la flor y si hay exceso de luz el color original se torna pálido (Gómez, 2010).

Temperatura

Influye directamente en el crecimiento y la producción, puede acelerar o atrasar el desarrollo del cultivo. Cuando se tienen temperaturas altas se acelera el desarrollo, se producen flores con botones pequeños, tallos muy delgados, bajo número de pétalos, alargamiento de los pistilos (órgano que suele estar ubicado en la parte central de las flores y que corresponde a su faceta femenina) y follaje (conjunto de hojas y ramas de plantas) muy angosto, alto porcentaje de materia seca. Cuando se tienen temperaturas bajas se atrasa el crecimiento y éste es tan lento que produce tallos muy cortos y gruesos, flores con botones grandes pero flojos, follaje ancho. El rango de temperatura aceptables para el clavel, en general, entre el día y la noche debe ser de $8°C$ a $30°C$, con una media óptima en el día de $18°C$ y de $10°C$ en la noche (Gómez, 2010).

Humedad relativa

El rango óptimo de la humedad relativa dentro del invernadero debe ser entre 60 % y 70 %, de esta manera la transpiración y la fotosíntesis se realiza con normalidad. Si dentro del invernadero se tiene una humedad relativa muy alta (cerca del 100 %), se ve favorecido el desarrollo de hongos. Si por el contrario el porcentaje de humedad relativa es muy bajo la incidencia de plagas puede ser mayor (Gómez, 2010).

Propagación

Los claveles se propagan a partir de esquejes obtenidos de un stock de plantas madre no de lechos en floración. Los esquejes se pueden guardar en almacenes fríos hasta que se necesiten para replantación de los lechos de producción. A las plantas madres se les debe permitir producir unas pocas flores para asegurarse de que son verdaderas en color y buena forma. Los esquejes pueden almacenarse en frío hasta 6 meses de 0 a 0.25°, las plantas madre deben haberse rociado regularmente con fungicidas e insecticidas (Salinger, 1992).

El cultivo de plantas madre para la producción de esquejes, se debe considerar como una actividad independiente de la producción de flor cortada. En los últimos años en Colombia, se han montado empresas que se dedican a esta actividad para suplir de esquejes a los floricultores. Sin embargo, varias empresas grandes productoras de Clavel, han optado por desarrollar su propio proyecto de plantas. Después de sembradas las plantas madre en un sitio definitivo, en bancos elevados y debidamente desinfectados. Generalmente la planta madre se deja diez meses en total desde la siembra hasta el arranque, tiempo durante el cual produce entre 25 a 35 esquejes para producción en un período de aproximadamente 8 meses descontando el tiempo de formación de la planta (Gómez, 2010).

Ciclo de vida del Clavel

El clavel estándar, durante su desarrollo, pasa por siete estados fenológicos (en las plantas, cada una de las etapas por las que pasan a lo largo de un período vegetativo) que dependen de la variedad y las condiciones climáticas. Los estados fenológicos del clavel son: 1) fijación de la raíz después del trasplante, semana 0 a 6; 2) desarrollo de brotes laterales (crecimiento que se desarrolla formando ángulo con el tallo principal del árbol) posterior al despunte, semana 5 a 15; 3) elongación de tallos (alargamiento del tallo), semana 14 a 25; 4) desarrollo de botón principal y laterales, semana 16 a 30; 5) primera cosecha, semana 23 a 34; 6) segundo periodo vegetativo, semana 30 a 50; 7) segunda cosecha, semana 48 a 65 y producción continua, semana 64 a 104. Durante el segundo periodo vegetativo y de producción prosigue el desarrollo de brotes laterales, la elongación de los tallos y el desarrollo de los botones principales y laterales (Arevalo et al., 2007).

Invernadero hidropónico

Desde 1950, en Colombia se introdujeron a escala experimental los cultivos en sustratos (hidropónicos). A partir de 1988 el cultivo de flores se adhirió a esta técnica de cultivo y hoy en día se estiman alrededor de 2000 hectáreas cultivadas bajo este sistema (Navas, 2001).

En el cultivo de clavel en hidroponía, el sustrato está constituido por materia orgánica como cascarilla de arroz, pues se presenta como material liviano, de buen drenaje, buena aireación y buena retención de humedad. El invernadero hidropónico dispone de un sistema de regadío mediante goteros por el cual las raíces de los cultivos reciben una solución nutritiva equilibrada disuelta en agua, con todos los elementos químicos necesarios para el desarrollo de las plantas, la nutrición está perfectamente controlada mediante fertilizantes de alta pureza que se conducen hasta estos canalones (conducto que recibe y conduce sustancias líquidas). En el clavel de invernadero hidropónico se acorta el periodo vegetativo del ciclo de la planta adelantando así, en una semana en periodo productivo; además el ciclo productivo en suelo dura entre 4 a 5 semanas y en hidropónico es de 3 semanas. La productividad por planta es igual en los cultivos hidropónicos que en suelo, pero si se tiene en cuenta la producción por metro cuadrado, esta es mayor en cultivos hidropónicos debido a que la densidad de siembra es mayor a la de suelo y hay menor mortalidad de plantas.

Demarcación de la cama

En el medio floricultor se denomina cama el sitio en donde se va a realizar la siembra definitiva de las plantas, también en otros sectores se conoce con el nombre de era o surcos. En Colombia, según la

empresa, se encuentran camas con anchos desde los 60 cm hasta los 120 cm, lo mismo que la altura de la cama para la siembra, de acuerdo con el tipo de suelo va desde el mismo nivel del suelo hasta una altura de 30 cm. La longitud de la cama tiene un largo promedio de 30 a 33 metros (Gómez, 2010).

Cosecha y pos-cosecha

Los claveles, se cosechan normalmente cuando el capullo se ha abierto por completo de modo que los pétalos exteriores estén en ángulo recto con el tallo, los pétalos centrales manteniéndose relativamente apiñados en el centro. Los tallos se cortan con tijeras y el tallo se extrae a través de la red. Los ramos de las flores cortadas se depositan con frecuencia sobre las redes o alambres superiores, un método mejor es un carrito estrecho en el pasillo (Salinger, 1992). El número de flores por planta depende del número de tallos, de modo que si se pinzan los primeros brotes se estimula la formación de ramas y con ello aumenta el número de flores, aunque también se retrasa el momento de la floración (Organización de las Naciones Unidas para la Agricultura y la Alimentación, 2002). La selección de los claveles debe ser (Rimache, 2011):

- Tallos: fuerte y vigoroso, recto sin brotes laterales y largo.
- Hojas: deben ser completas y sanas.
- Flor: cáliz completo, pétalo sano color homogéneo e intenso de buen tamaño y sin deformaciones.

Los claveles pueden ser recolectados como capullos abriéndose; siendo el estadio más satisfactorio cuando los pétalos se expanden sobre el cáliz y la flor es de forma rectangular. Estas se abren por completo como una flor de larga duración si se colocan en una solución de apertura de yemas, mantenida a 20°C e introducido al conjunto en sacos de plástico para mantener la humedad (Salinger, 1992).

Normalmente el productor despacha sus flores el mismo día que realiza el corte, sin embargo, es posible almacenar el clavel en bodegas acondicionadas que permitan temperaturas de 2°C a 5°, de este modo la flor resiste hasta 3 semanas. La flor del clavel está muy adaptada a los malos tratos posteriores al corte y es mucho más resistente que una rosa. En la bodega se procede a su clasificación con arreglo a su longitud, calidad de la flor y se empaquetan. Las flores de exportación se deben tratar con sales de plata que le otorga una resistencia extraordinaria prolongando la duración de la flor (Rimache, 2011).

6.1.2. La Sabana de Bogotá

La Sabana de Bogotá, es una región con gran densidad de población ya que en ella se localiza la ciudad de Bogotá y municipios como Zipaquirá, Soacha y Facatativá, además de otros municipios. Esta zona es atravesada por vías importantes que comunican a Bogotá con diferentes poblaciones como Tunja, Zipaquirá, Chiquinquirá, Pacho, Facatativá, La Vega, Melgar, La Mesa, Sibaté, Machetá y Gachetá.

La Sabana de Bogotá, presenta precipitaciones que van desde 600 hasta 1.200 mm, lo que permite la delimitación de zonas con diferentes concentraciones de lluvia, en donde el régimen es bimodal y se caracteriza por la ocurrencia de dos épocas mayores de lluvias, separadas por dos épocas de menores (IGAC, 1992). Lo anterior, más factores como la altura y el clima determinan dos pisos térmicos: frío y de

páramo; el frío, esta caracterizado por alturas desde los 2000 hasta 3000 m.s.n.m, con temperaturas de 12 a 18ºC y el piso térmico de páramo tiene temperaturas inferiores a los 12ºC y se encuentra por encima de los 3.000 m (IGAC, 1992). La vegetación es respuesta de las altitudes y del clima; todas estas variantes se agrupa en delimitación de los pisos bioclimáticos (Montoya and Reyes, 2005).

6.2. Marco Estadístico

6.2.1. Análisis de Datos Funcionales

Hoy en día se ha empezado a trabajar en muchas áreas con grandes bases de datos, que corresponden a observaciones de una variable aleatoria tomadas a lo largo de un intervalo continuo (o discretizaciones cada vez más extensas de este intervalo continuo), generando curvas como resultado de la medición que representa a la muestra concreta que al menos se ha evaluado en una centena de puntos. A este tipo de datos se le llama datos funcionales (Febrero-Bande, 2008) .

Según Ramsay and Silverman (2005), el análisis de datos funcionales o FDA por sus siglas en ingles *Functional Data Analysis*, se refiere al análisis estadístico de muestras de datos que consisten en funciones o superficies aleatorias, donde cada función se ve como un elemento de muestra, tal como en el análisis clásico, sólo que se trabaja con funciones en vez de escalares. Las variables que se usan en el FDA son variables de tipo continuo, pero en la práctica son medidas de forma discreta en ciertos periodos de tiempo, entre los cuales no se puede tener un registro del evento pero que se sabe que existe. Las observaciones discretas forman una colección de datos que son usados para hallar la mejor función continua que representa los datos poblacionales que están siendo estudiados.

En el presente estudio, se tienen observaciones de los índices de producción medidos en un tiempo t que representa semanas, aunque entre semana no existe un registro no significa que no exista, por lo cual se requiere usar técnicas de suavización para tener conocimiento del comportamiento de los índices en esos lapsos de tiempo que no se tiene registro.

Los objetivos de análisis de datos funcionales son esencialmente los mismos que los de cualquier otra rama de la estadística (Ramsay and Silverman, 2005), entre los que se encuentra:

- Representar los datos de manera que ayudan a un análisis más detallado.
- Mostrar los datos con el fin de destacar diversas características.
- Estudiar importantes fuentes de patrones y variaciones entre los datos.
- Explicar la variación de una variable dependiente mediante el uso de información de una variable independiente.
- Comparar dos o más conjuntos de datos con respecto a ciertos tipos de variación.

Variables y Datos Funcionales

Una variable aleatoria X se dice que es una variable funcional si toma valores en un espacio dimensional infinito (espacio funcional). Un conjunto de datos funcionales $X_1, ..., X_n$ es la observación de n

variables funcionales $X_1, ..., X_n$ idénticamente distribuidas (Ferraty and Vieu, 2006).

Un modelo bien aceptado para este tipo de datos es considerarlo como trayectorias de un proceso estocástico $X = \{X_t\}_{t \in T}$ tomando valores en algún espacio de Hilbert, H, de funciones definidas en algún conjunto T. Generalmente, T representa un intervalo de tiempo, de longitudes de onda o cualquier otro subconjunto de $\mathbb{R}$.

La principal fuente de dificultad al tratar con datos funcionales se basa en el hecho de que las observaciones se supone que pertenecen a un espacio de dimensión infinita, mientras que en la práctica sólo se han muestreado curvas observadas en un conjunto finito de puntos de observación. De hecho, es habitual que sólo tengamos observaciones discretas X_{ij} de cada trayectoria de la muestra $X_i(t)$ en un conjunto finito de nudos $\{t_{ij} : j = 1, ...m_i\}$. Debido a esto, el primer paso en el FDA es a menudo la reconstrucción de la forma funcional de los datos de las observaciones discretas (Ferraty and Vieu, 2006). Esta reconstrucción se realiza por medio de la siguiente expresión la cual se denomina expansión base (Ramsay and Silverman, 2005):

$$X(t) = \sum_{k=1}^{K} c_k \phi_k(t) \tag{6.1}$$

Donde c indica el vector de tamaño K de los coeficientes c_k y ϕ es el vector funcional de los cuales sus elementos son las funciones bases ϕ_k.

Los coeficientes del modelo (6.1) pueden ser estimados usando mínimos cuadrados. Como el objetivo es ajustar observaciones discretas y_j, $j = 1, 2, ..., n$ usando el modelo:

$$y_j = x(t_j) + \varepsilon_j \tag{6.2}$$

Se obtiene un suavizado lineal simple si se determinan los coeficientes de expansión c_k minimizando el criterio de mínimos cuadrados:

$$SCE = \sum_{j=1}^{n} (y_j - x(t_j))^2 \tag{6.3}$$

El criterio se puede expresar en forma matricial como:

$$SCE = (y - \phi c)'(y - \phi c) \tag{6.4}$$

Con la ecuación anterior, se puede encontrar la estimación de c por mínimos cuadrados ordinarios:

$$\hat{c} = (\phi'\phi)^{-1}\phi' y \tag{6.5}$$

donde el vector $\hat{y}$ de valores ajustados es:

$$\hat{y} = \phi\hat{c} = \phi(\phi'\phi)^{-1}\phi' y \tag{6.6}$$

La aproximación simple de mínimos cuadrados es apropiada en situaciones en las que suponemos que los residuos j sobre la curva verdadera son independientemente e idénticamente distribuidos con media cero y varianza constante σ^2.

Funciones Bases

Dado que en el modelo (6.1) se necesita usar una función base, es necesario escoger la que represente de manera adecuada el comportamiento de los datos. Según Ramsay and Silverman (2005), un sistema de funciones bases es un conjunto de funciones conocidas ϕ_k que son matemáticamente independiente de cada otra y que tiene la propiedad que podemos aproximar, arbitrariamente bien, cualquier función tomando una suma ponderada o combinación lineal de un número k suficientemente grande de estas funciones. Las funciones base deberían tener características que coincidan con las conocidas como pertenecientes a las funciones que se están estimando. Esto hace más fácil lograr una aproximación satisfactoria usando un número relativamente pequeño K de funciones base.

Sistema de Bases Fourier

La Base Fourier es una de las más conocidas y viene dada por:

$$\hat{x}(t) = c_0 + c_1 sen\omega t + c_2 cos\omega t + c_3 sen2\omega t + c_4 cos2\omega t + ... \tag{6.7}$$

Definida por las bases $\phi_0(t) = 1$, $\phi_{2r-1}(t) = senr\omega t$, $\phi_{2r} = cosr\omega t$

Esta base es periódica y el parámetro ω determina el periodo $2\pi/\omega$. Si los valores de t_j están igualmente espaciados en τ (intervalo de tiempo en el cual se debe aproximar la función) y el periodo es igual a la longitud del intervalo τ, entonces la base es *ortogonal*, en el sentido que la matriz del producto cruzado $\phi'\phi$ es diagonal y puede hacerse igual a la identidad, dividiendo las funciones bases por constantes adecuadas. $\sqrt{n}$ para $j = 0$, $\sqrt{n/2}$ para otro j. (Ramsay and Silverman, 2005)

Una Serie de Fourier es especialmente útil para funciones extremadamente estables, es decir, funciones donde no hay características locales fuertes y donde la curvatura tiende a ser del mismo orden en todas partes.

Sistema de Bases Spline

Las funciones de *spline* son la opción más común del sistema de aproximación para datos o parámetros funcionales no periódicos. Los *splines* combinan el cálculo rápido de polinomios con una flexibilidad sustancialmente mayor, a menudo obtenida con sólo un número modesto de funciones de base. Además, se han desarrollado sistemas de base para funciones *spline* que requieren una cantidad de cálculo proporcional a n, una propiedad vital ya que muchas aplicaciones implican miles o millones de observaciones. Los *splines* son segmentos de polinomios unidos de extremo a extremo los cuales deben ser suaves. Los puntos en los que los segmentos se unen son llamados nodos. El orden $m = n + 1$ de los segmentos de los polinomios y la ubicación de los nodos define el sistema, donde n es el grado de los polinomios.

Para construir un *spline*, se especifica un sistema de funciones de base $\phi_k(t)$ que tendrán las siguientes propiedades esenciales:

- Cada función base $\phi_k(t)$ es en si misma una función *spline* definida por un orden m y una secuencia de nudos τ.

- Puesto que un múltiplo de una función *spline* sigue siendo una función spline y como la suma y la resta de *splines* son también *splines*, cualquier combinación lineal de estas funciones base es una función *spline*.

- Cualquier función *spline* definida por m y τ puede ser expresada como una combinación lineal de estas funciones base.

- El número de bases de funciones es igual al orden m + el número de nudos interiores.

La notación $B_k(t,\tau)$ se utiliza a menudo para indicar el valor en t de la función base B-Spline definida por la secuencia de nodos. Aquí k se refiere al número del nodo más grande en o a la izquierda inmediata del valor t. Los $m-1$ nodos agregados al punto de interrupción inicial también se cuentan en este esquema. Esta notación nos da $m+L-1$ funciones de base, de acuerdo con esta notación, una función *spline* $S(t)$ con nodos interiores discretos se define como:

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t,\tau) \tag{6.8}$$

Número K de Funciones Bases

Es necesario obtener el número óptimo de funciones bases K que se van a incluir en el modelo (6.1), cuanto mayor sea el K, mejor será el ajuste a los datos, pero también se corre el riesgo de ajustar el ruido o la variación que se desea ignorar. Por otro lado, si K es demasiado pequeño, se puede perder algunos aspectos importantes de la función suavizada x que se está tratando de estimar. Lo anterior se puede expresar de otra manera. Para los grandes valores de K, el sesgo en la estimación de $x(t)$

$$Sesgo[\hat{x}(t)] = x(t) - E[\hat{x}(t)] \tag{6.9}$$

es pequeño.

Por otro lado, reducir la varianza de la estimación

$$Var[\hat{x}(t)] = E[\hat{x}(t) - E[\hat{x}(t)]]^2 \tag{6.10}$$

lleva a buscar valores más pequeños de K, pero no tan pequeños como para hacer que el sesgo sea inaceptable. Cuanto peor es la relación señal / ruido en los datos, la varianza de muestreo más reducida superará el sesgo de control. Una manera de expresar lo que realmente se quiere lograr es con el error cuadrático medio:

$$ECM[\hat{x}(t)] = Sesgo^2[\hat{x}(t)] + Var[\hat{x}(t)] \tag{6.11}$$

Lo que esta relación dice es que valdría la pena tolerar un pequeño sesgo si el resultado es una gran reducción en la varianza muestral. La validación cruzada es una técnica que permite elegir el parámetro que controla el equilibrio entre precisión y variabilidad (o entre bondad de ajuste y capacidad predictiva) (Delicado, 2008).

Suavización Penalizada

El método de rugosidad penalizada está basado en optimizar un criterio de ajuste que define que tan suaves se quieren los datos. Este método modifica el criterio de mínimos cuadrados el cual resulta ser en ocasiones torpe para el suavizado de los datos, introduciendo un término llamado penalización, el cual hace explícito lo que sacrificamos de sesgo para reducir la varianza y lograr un mejor ECM (Ramsay and Silverman, 2005). Se requiere optimizar la siguiente función:

$$PENSCE_\lambda(x|y) = [y - x(t)]'W[y - x(t)]^2 + \lambda PEN_2(x) \tag{6.12}$$

donde λ es el parámetro de suavizado que mide la tasa de cambio del ajuste a los datos, a través de la suma de cuadrados del error expresada en el primer término y la variabilidad de la función x según lo cuantificado por $PEN_2(x) = \int [D^2 x(s)]^2 ds$ en el segundo término.

6.2.2. Validación Cruzada o Método CV

Según Delicado (2008), esta técnica consiste en sacar de la muestra consecutivamente cada una de las observaciones x_i, estimar el modelo con los restantes datos (sea $\hat{x}_i(t)$ el estimador así obtenido), predecir el dato ausente con ese estimador (así se está haciendo de hecho predicción fuera de la muestra) y finalmente, comparar esa predicción con el dato real. Esto se hace con cada posible valor de λ, lo que permite construir la función:

$$VC(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{x}_i(t))^2 \tag{6.13}$$

que mide el error de predicción del estimador fuera de la muestra para cada valor de λ. El valor que minimice esa función, será el valor del parámetro de suavizado elegido. En la práctica se usa una simplificación del modelo 6.13:

$$VC(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - s_{ii}}\right)^2 \tag{6.14}$$

donde s_{ii} es un elemento de la matriz de suavizado S, la cual define Ramsay and Silverman (2005) como $S_\lambda = \Phi(\Phi' W \Phi + \lambda R)\Phi' W$, donde $R = \int D^m \phi(s) D^m \phi'(s) ds$ es una matriz que contiene integrales de productos externos de vectores de funciones.

Una medida que es popular en la literatura de suavizado spline es la validación cruzada generalizada (GCV) desarrollado por Craven y Wahba (1979). Originalmente, se desarrolló como una versión más simple del procedimiento de validación cruzada que evitó la necesidad de re-suavizar n veces. Pero también se ha encontrado que es bastante más fiable que la validación cruzada en el sentido de tener menos tendencia a suavizar (Ramsay and Silverman, 2005). Consiste en sustituir en la fórmula 6.14 los valores s_{ii} de la diagonal de S por su valor promedio (Delicado, 2008), quedando la siguiente expresión:

$$VCG(\lambda) = \frac{n^{-1} SCE}{[n^{-1} Traza(I - S_\lambda)]^2} \tag{6.15}$$

6.2.3. Estadísticas Descriptivas para Datos Funcionales

Las estadísticas descriptivas para datos funcionales son similares a las vistas en los cursos introductorios a la estadística. La función media, es la media de las funciones en función del punto a través de las repeticiones (Ramsay and Silverman, 2005).

$$\bar{x}(t) = N^{-1} \sum_{i=1}^{N} x_i(t) \tag{6.16}$$

Similarmente la función de la varianza tiene los valores:

$$var(x(t)) = (N-1)^{-1} \sum_{i=1}^{N} [x_i(t) - \bar{x}(t)]^2 \tag{6.17}$$

y la función de desviación estándar es la raíz cuadrada de la varianza.

Capítulo 7

Metodología

Para realizar el ajuste de las curvas de producción de las variedades de claveles y mini claveles, se trabajó con los datos suministrados por la empresa C.I Flores de Aposentos S.A.S, los cuales fueron originados desde una de sus cinco fincas llamada Rebaños. A continuación se presenta la metodología propuesta para alcanzar el objetivo de este trabajo.

7.1. Calidad y Selección de los Datos

Se cuenta con una base de datos suministrada por la empresa Aposentos Flores S.A., la cual está dividida en dos partes: la primera parte, corresponde al total de tallos de flores cosechados en una semana de vida de cada variedad de planta ubicada en los diferentes bloques, así como el total de plantas sembradas por bloque. En la segunda parte, los datos corresponden a los índices de producción en cada semana de vida de la planta agrupados por variedad y los índices de producción esperados que se obtienen de forma empírica según el comportamiento que se observa en los datos. Se cuenta con índices de producción registrados para 54 variedades de claveles y mini claveles, estos índices se presentan a través del tiempo, en semanas, desde la semana 0 a la semana 100 de vida de la planta. Cada variedad tiene diferentes registros de siembra desde el año 2012 al 2016. Aunque las siembras reportadas en este estudio son de diferentes periodos de tiempo (meses y años) y diferentes bloques, las condiciones del invernadero son constantes durante todo el año y en todos los bloques, lo cual se asume no debería generar ningún tipo de variabilidad; por lo tanto, solamente se tomará en cuenta el tiempo de vida y producción de las plantas.

Inicialmente se realizó una depuración de la base de datos donde se validó si los registros con valores igual a cero en semanas donde debía haber producción, correspondían a datos reales, datos faltantes o si la siembra era a 1 pico. Una vez realizado esto, se seleccionaron las 20 variedades que contaban con mayor cantidad de datos (curvas de producción) para tener mayor acercamiento al comportamiento real de la producción de cada variedad, además, se tuvo en cuenta que las siembras fueran de diferentes años con el fin de tener datos a través del tiempo. La selección de las 20 variedades se realizó para efectos del cumplimiento de los objetivos planteados para este proyecto de grado. Las 20 variedades seleccionadas

fueron: atacama, bizet, cherrio, don pedro, farida, hermes, hermes orange, kaori, kino, kirk, komachi, lizzy, mariposa, monseñor, moonlight, nahema, pomodoro, red magic, voragine y zurigo.

7.2. Modelo Empírico

Actualmente, la empresa cuenta con una curva de producción ajustada para cada variedad, la cual hemos denominado *Modelo Empírico*. Este modelo fue creado por la empresa a partir de los índices de producción observados de forma discreta sin construir una formulación matemática explícita y se usa actualmente como una herramienta para proyectar el número de flores que se pueden producir en cada semana. El modelo empírico se presenta en la base de datos como una curva de predicción de los índices de producción por cada semana de vida de la planta. Aunque estas predicciones buscan reproducir el comportamiento de los índices de producción en los cuatro periodos de vida de las plantas, éstas tienden a sobre estimar o subestimar la producción real generando grandes errores en las predicciones.

En la Figura (7.1), se presenta a modo de ilustración una parte de la base de datos de los índices de producción de la variedad Atacama en las celdas sin color y las predicciones realizadas con el modelo empírico en las celdas de color verde oscuro. En la figura (7.2) se observa el modelo empírico graficado en puntos azules y en gris los índices de producción de las siembras registradas, mostrando que es un modelo mal ajustado pues no representa el comportamiento real de la producción.

VARIEDAD	BLOQUE	PLANTAS	23	24	25	26	27	28	29	30	31
MODELO	0	0	0,3219	0,5354	1,1148	0,8981	0,6620	0,4356	0,2768	0,1931	0,1760
ATACAMA	05A	1968	0,0254	0,1524	0,3125	0,6860	0,6758	0,5386	0,4827	0,4548	0,2261
ATACAMA	09A	5904	0,0186	0,1380	0,2583	0,5420	0,5674	0,4743	0,6944	0,3853	0,2964
ATACAMA	11B	2952	0,0085	0,0593	0,1609	0,2964	0,3896	0,4048	0,4387	0,2625	0,3726
ATACAMA	14A	1968	0,0762	0,1524	0,2668	0,5462	0,5843	0,7978	0,4319	0,4040	0,4040
ATACAMA	01A	1968	0,0240	0,0356	0,0889	0,2287	0,3176	0,2033	0,3811	0,2541	0,1143
ATACAMA	03B	1968	0,0254	0,0737	0,3176	0,2287	0,3913	0,5564	0,4624	0,5030	0,6301
ATACAMA	10B	5776	0,0173	0,1697	0,2718	0,5635	0,4839	0,6882	0,5255	0,4008	0,4034

Figura 7.1: Base de Datos

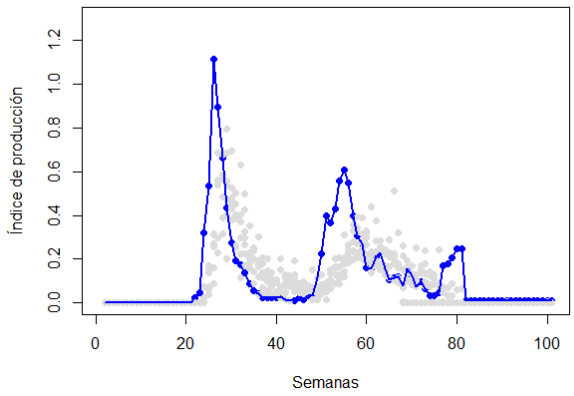


Figura 7.2: Modelo empírico variedad Atacama

7.3. Exploración de los Datos

Se realizó un análisis exploratorio de los datos con el fin de examinar los datos antes de realizar el ajuste de las curvas con la técnica de datos funcionales, de esta forma podemos entender el problema que se está presentando en la empresa con las diferencias de las curvas de producción y del modelo que están usando actualmente. Este paso permitió también identificar que habían curvas con índices de producción iguales a cero en semanas donde no debía y así poder corregirlos dado el caso en que no fuera cierto.

7.4. Construcción de las Curvas Funcionales de Producción

Una función x tiene la característica de ser de dimensión infinita, es decir, para cada valor de t existe un valor de x. Dado que las medidas discretas no pueden representar estos datos, la suavización es una manera de representar una función a priori expresada como una suma de funciones bases conocidas y que tiene el objetivo de preparar los datos para ser funciones.

Para realizar la suavización se utilizaron bases B-spline, ya que son la opción más común de sistema de aproximación para datos no periódicos cómo los índices de producción de nuestro estudio (Ramsay and Silverman, 2005), además los B-spline cuenta con ventajas computacionales de sistemas de base, debido a que no tiene en cuenta cuán grande sea el parámetro k ni el tamaño de la muestra. Otra razón para escoger como base los B-splines, es que se realizó para algunas variedades el ajuste con las bases Fourier y al comparar con las B-splines, estas últimas presentaban mejor suavización de los datos. El número de funciones bases a usar por cada variedad y el parámetro de suavización óptimo se estimaron por medio de la validación cruzada generalizada, la cual se detalló en la subsección 6.2.2.

Después de construir los datos funcionales para cada variedad de claveles, se realizó un análisis descriptivo funcional, construyendo curvas que representaran la función de la media como medida de tendencia central y como medida de dispersión se construyeron bandas de variabilidad a lo largo del tiempo de vida de la planta. Estas bandas de variabilidad expresadas como $\bar{x}(t) \mp 2\sqrt{var(x(t))}$ cubren el 95.5 % de los casos y nos permiten observar los valores entre los cuales fluctúa las siembras, así la empresa podrá contar con un margen de error para estimar la producción que tendrán en cada semana de vida de la planta. Adicional, se calculó un porcentaje de cobertura de cada modelo empírico dentro del modelo estimado, esto es, el número de veces que el modelo empírico estuvo dentro de las bandas de variabilidad del modelo estimado en cada semana de vida de la planta.

Capítulo 8

Resultados

En este capítulo se presentan los resultados obtenidos durante el estudio. En la primera sección se presenta el análisis exploratorio de los datos puros por medio de técnicas de la estadística clásica, es decir, se realizó un reconocimiento de los datos para observar el comportamiento de la producción de las variedades de claveles y poder detectar errores o registros incompletos; de igual forma la exploración permitió tener una mejor visión del problema que presentaba la empresa con el ajuste del modelo que usan actualmente. La segunda sección muestra la construcción de los datos funcionales para cada variedad así como sus estadísticos descriptivos funcionales con el objetivo de caracterizar el comportamiento de los índices de producción a lo largo de vida de la planta.

8.1. Análisis Exploratorio de Datos

La empresa Aposentos Flores S.A. actualmente exporta 54 variedades de claveles y mini claveles a diferentes países. Los datos suministrados por la empresa contiene los índices de producción de estas 54 variedades en diferentes años. Para efectos de cumplimiento de los objetivos planteados para este proyecto de grado, se trabajará con 20 variedades a las cuales se les realizará el análisis exploratorio de datos y el ajuste de las curvas de producción.

Amico Lavander	Danza	Indra	Milu	Pomodoro
Antigua	Danzica	Jera	Mizuky	Randal
Atacama	Dark Tempo	Kaori	Monseñor	Red Magic
Bizet	Don Pedro	Kino	Moonlight	Sinai
Bouzeron	Doncell	Kirk	Natalia	Solex
Breier	Farida	Komachi	Nahema	Soraya
Caribe	Fiesta Komachi	Lion King	Noblesse	Tabor
Carole	Golem	Lincy	Novia	Tiepolo
Cheryl	Harvey	Lizzy	Opera	Voragine
Cherrio	Herems	Mandalay	Orange Dandy	Zurigo
Cream Viana	Hermes Orange	Mariposa	Polimnia	

Cuadro 8.1: Variedades de flores de la empresa

A continuación se muestran las curvas de 6 variedades las cuales presentan mayor error de estimación del modelo empírico usado por la empresa, los puntos representan los índices de producción, los colores diferencian cada una de las siembras y la línea azul el modelo empírico. Las curvas de las 20 variedades se encuentran en los anexos.

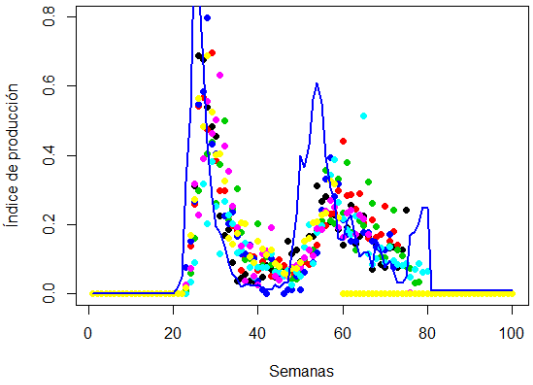


Figura 8.1: Atacama

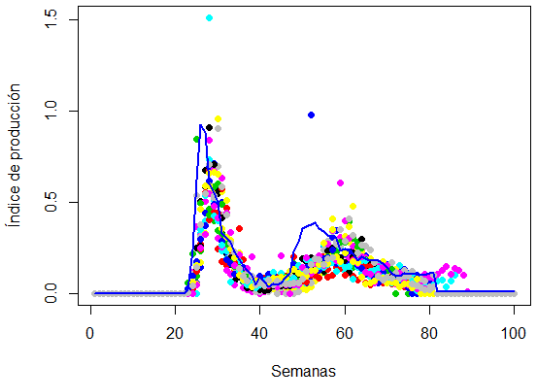


Figura 8.2: Hermes

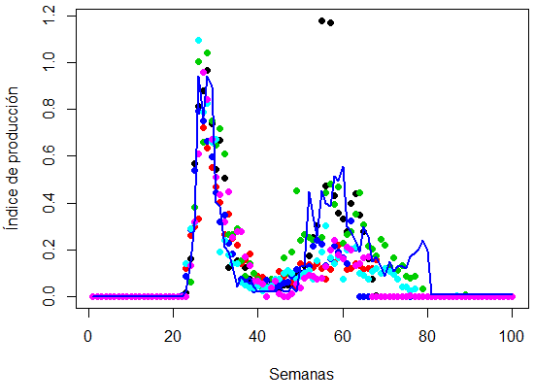


Figura 8.3: Kirk

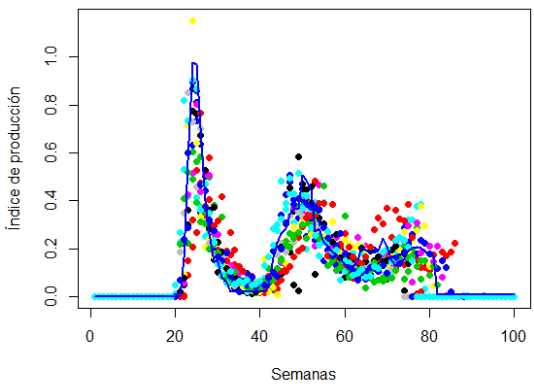


Figura 8.4: Lizzy

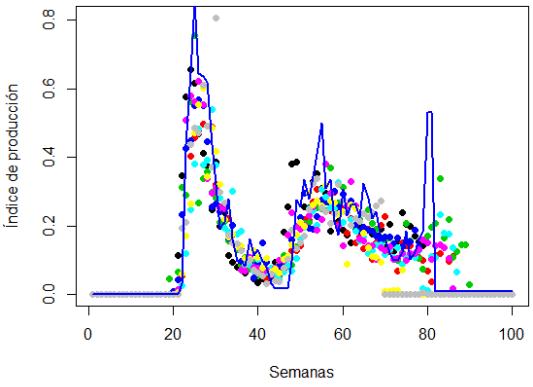


Figura 8.5: Moonlight

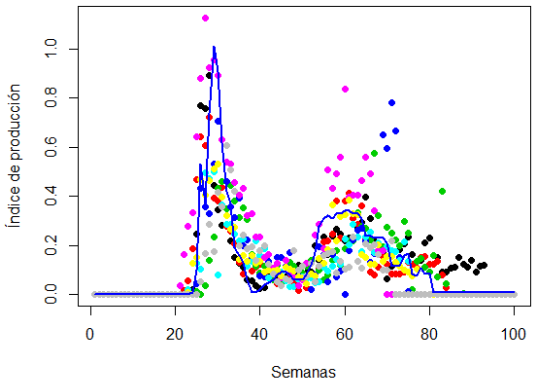


Figura 8.6: Nahema

En cada una de las gráficas, se observa cómo el modelo utilizado actualmente por la empresa (línea azul) no sigue el comportamiento de los datos. La figura 8.1 muestra la variedad Atacama la cual evidencia visualmente el mal desempeño del modelo empírico el cual tiene el primer y segundo pico antes de los datos reales y un sobreajuste durante las semanas de vida. El error estimado de este modelo para el primer pico es de 0.852 y lo alcanza en la semana 25 cuando las siembras registradas reportan su máximo entre la semana 27 y 28. Por otro lado, para el segundo pico se observa un sobreajuste en el modelo empírico, con un error estimado de 0.431 en la semana 54, las siembras registradas reportan el máximo del segundo pico entre la semana 57 y 59.

Al igual que la variedad Atacama, se observa cómo las demás variedades tienen sobreajustes de producción en el modelo empírico y alcanzan sus máximos semanas antes de lo real. El error de estimación del primer pico para la variedad Hermes es de 0.612 y de 0.242 en segundo pico. La variedad Kirk presenta mayor error en el segundo pico equivalente a 0.369 y se pueden observar varios puntos atípicos los cuales deberán verificarse. La variedad Lizzy presenta el mayor error del modelo empírico para el primer pico (0.924) alcanzado su máximo en la semana 25 cuando los datos reales reportan el máximo en la semana 28 aproximadamente. En la figura 8.5 se observa la variabilidad del modelo empírico para la variedad Moonlight, el cual además de tener un error de 0.320 para el primer pico y de 0.230 para el segundo pico, presenta un tercer pico en la semana 80 el cual no representa el comportamiento real de los datos, pues estos no presentan producción alta en esta semana. La variedad Nahema también presenta datos atípicos en el segundo pico y el modelo empírico presenta un error de 0.541 en el primer pico y de 0.190 en el segundo pico.

Se observa también, cómo muchas de las siembras tienen índice de producción cero después de llegar a la semana 80 en la cual se supone que las plantas dejan de producir. Posteriormente, se debe realizar un estudio profundo sobre esos datos y los datos atípicos, para verificar si fue algo inusual en la producción o si se debe a errores de digitación.

La Tabla 8.2 presenta los índices de producción máximos alcanzados por las variedades, así como la semana en la cual alcanzaron el máximo, para conocer el comportamiento de la producción de las plantas.

Variedad	**IP**	**Semana**	**Variedad**	**IP**	**Semana**
Atacama	0.7749	28	Komachi	1.1524	24
Bizet	1.2845	24	Lizzy	1.1128	28
Cherrio	0.9485	28	Mariposa	1.2906	25
Don Pedro	1.3231	27	Monseñor	1.4862	55
Farida	1.3694	26	Moonlight	0.8079	30
Hermes	1.5142	28	Nahema	1.1280	27
Hermes Orange	1.0924	26	Pomodoro	1.2152	23
Kaori	1.0721	26	Red Magic	1.4481	27
Kino	1.3541	25	Voragine	1.1947	24
Kirk	1.1788	57	Zurigo	1.6869	30

Cuadro 8.2: índice de producción máximo

Las variedades Kirk y Monseñor alcanzaron su máxima producción en semanas diferentes al primer pico, mostrando un comportamiento atípico con respecto a las demás variedades.

A continuación se muestran los diagrama de cajas para 2 variedades. En los gráficos se observa la variabilidad entre las siembras por cada variedad, siendo mayor en las semanas del pico y repico. También se permite observar que las variedades tienen siembras con índices de producción inusuales representados en puntos atípicos. Además, permite ver que hasta la semana 22 aproximadamente y después de la semana 70, la producción es nula. Los diagramas de cajas de las 20 variedades se encuentran en los anexos.

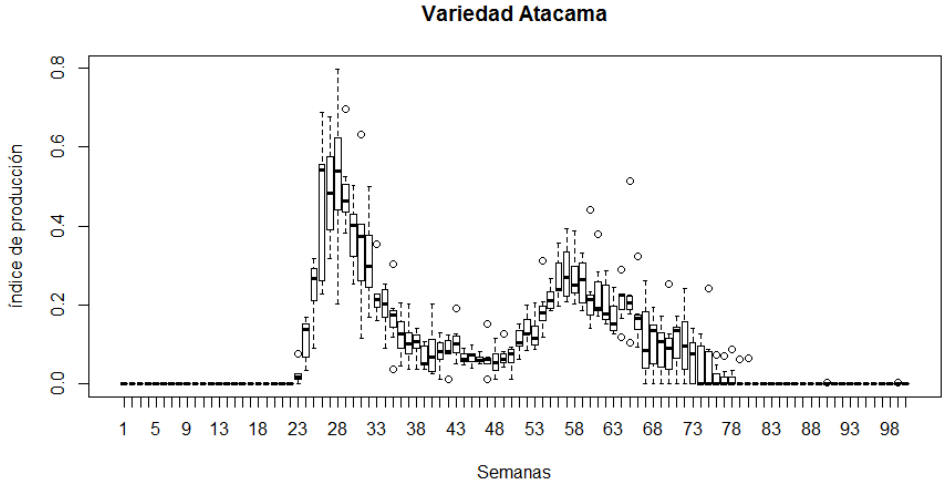


Figura 8.7: Boxplot variedad Atacama

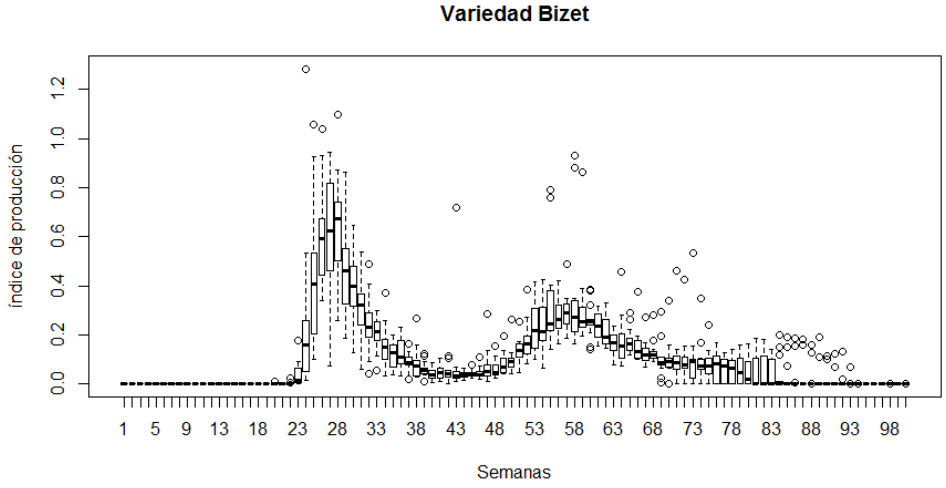


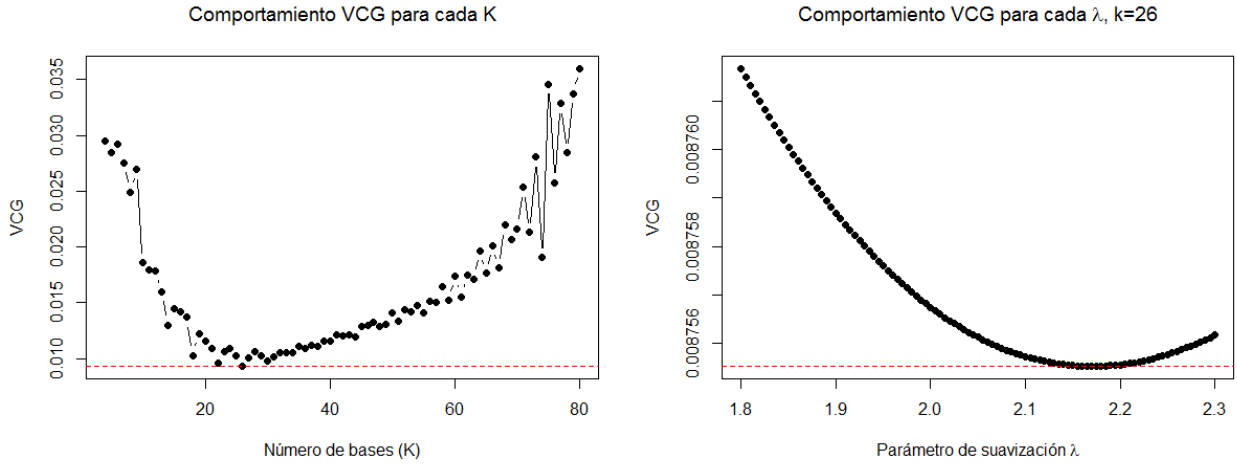
Figura 8.8: Boxplot variedad Bizet

8.2. Construcción de los Datos Funcionales

Se construyeron los datos funcionales con suavización penalizada (Sección 6.2.1) en el software estadístico R Project versión 3.2.3. Se utilizó la función `min.basis()` de la librería `fds.usc`, la cual permite obtener el número k de bases de funciones y el parámetro de suavización óptimo λ por medio del criterio de validación cruzada generalizada. Para usar esta función se define una secuencia de valores para los dos parámetros, la función devuelve los valores óptimos de ambos parámetros que hacen mínima la expresión (6.15).

A continuación se presenta la construcción detallada de los datos funcionales para la variedad Monseñor. Se utilizó el mismo procedimiento para las demás variedades.

La Figura (8.9) permite observar el comportamiento del criterio de validación cruzada generalizada para obtener el número de bases óptimo y el parámetro de suavización a utilizar. En el lado izquierdo se evalúan diferentes números de bases, dando como resultado que un $k = 26$ bases, minimiza el criterio. Por otro lado, en la gráfica del lado derecho se observa que para 26 bases, el criterio de VCG se minimiza con un parámetro de suavización $\lambda = 2{,}17$.



(a) (b)

Figura 8.9: Criterio de validación cruzada generalizada para (a) definir el número de bases y (b) definir el parámetro de suavización

Una vez obtenido los parámetros óptimos, se crearon las 26 funciones bases B-spline de orden 4, con la función `create.bspline.basis()` y luego se realizó el ajuste usando la función `smooth.basis()` la cual usa como criterio de ajuste los mínimos cuadrados y define la suavidad en términos de una penalización de rugosidad.

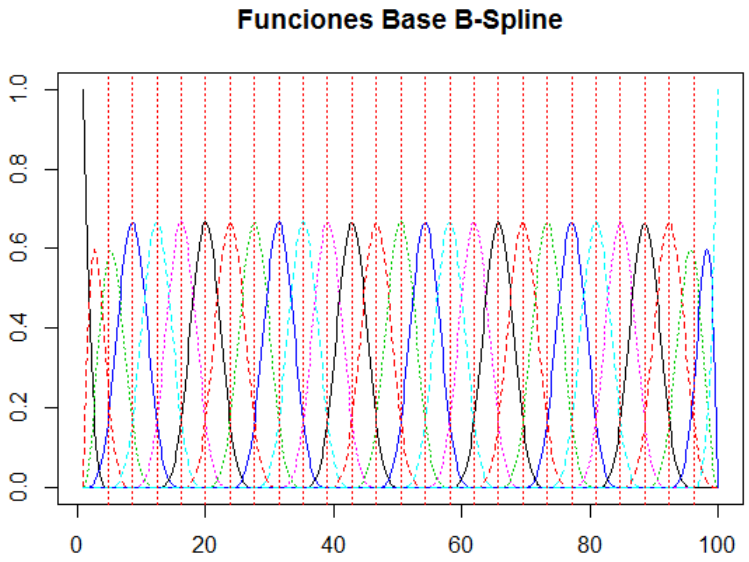


Figura 8.10: 26 Funciones base B-spline de orden 4

En la Figura (8.10) se observan las 26 funciones base de B-splines de orden 4 con 25 nodos equidistantes. El conjunto de 10 datos funcionales construidos se muestra en la Figura (8.11), representando el comportamiento de los índices de producción a lo largo de las 100 semanas de vida.

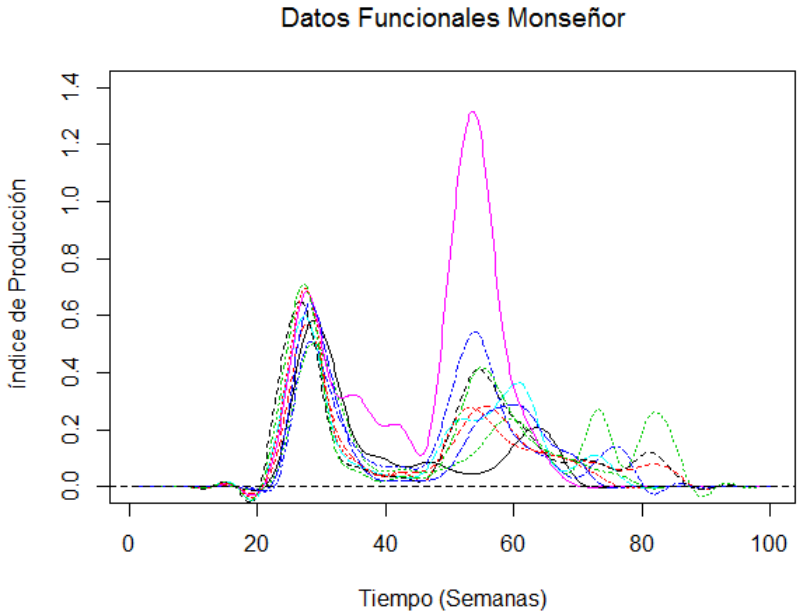


Figura 8.11: Conjunto de Datos funcionales variedad Monseñor

8.3. Análisis Exploratorio Funcional

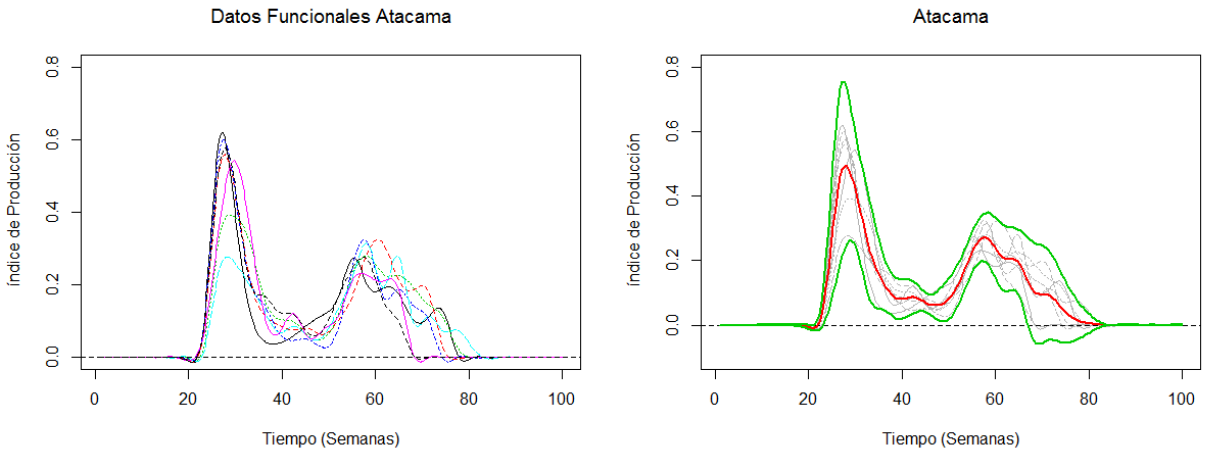
Para cada variedad, se presenta, en el lado derecho de las figuras, el comportamiento de la producción de la variedad, en el eje de ordenadas se tienen los índices de producción y en el eje de abscisas el tiempo representado en las semanas de vida de la planta. Se observa también la media funcional en color rojo, la cual nos permite tener el promedio de índice de producción a lo largo del tiempo y sus bandas de variabilidad en color verde. Los resultados se presentaron por cada variedad, ya que los índices de producción son diferentes en cada variedad.

A continuación se presentan los parámetros usados para la construcción de las curvas funcionales.

Variedad	**Número de Bases**	**Parámetro de Penalización**	**Número de Curvas**
Atacama	34	1.03	7
Bizet	33	0.63	14
Cherrio	30	0.59	5
Don Pedro	38	0.28	12
Farida	70	0.20	18
Hermes	69	0	16
Hermes Orange	69	0.24	20
Kaori	65	0	10
Kino	44	0.46	14
Kirk	34	0.96	6
Komachi	41	0.40	13
Lizzy	37	2.75	8
Mariposa	43	0.50	10
Monseñor	26	2.17	10
Moonlight	60	2.55	8
Nahema	47	0.50	8
Pomodoro	69	0	16
Red Magic	37	0.55	9
Voragine	61	0.02	9
Zurigo	36	0.57	9

Cuadro 8.3: Parámetros para la construcción de los datos funcionales

8.3.1. Atacama



(a) (b)

Figura 8.12: (a) Datos funcionales variedad Atacama. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

El valor promedio de producción más alto estimado por el modelo de datos funcionales se observó en la semana 28 correspondiente a 0.497, valor menor al usado por la empresa actualmente (1.114) el cual presenta un error de estimación de 0.852 en el pico 1 y de 0.431 en el pico 2, mostrando no sólo un sobreajuste del modelo en cuanto al índice de producción sino también una anticipación a las semanas donde alcanzan su máximo, pues como se observa en el cuadro 8.4 el modelo actual de la empresa estima el máximo en la semana 25 y el modelo estimado con un error de 0.143 lo presenta en la semana 28. El mismo comportamiento se puede observar para el pico 2. Por otro lado, se pudo calcular que el modelo empírico tenía una cobertura del 61 % dentro del modelo ajustado, porcentaje bajo, mostrando una vez más que para esta variedad resulta ser mejor utilizar el modelo estimado con datos funcionales.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	1.114	25	0.609	54
Modelo Estimado	0.497	28	0.273	58

Cuadro 8.4: Modelos de producción variedad Atacama

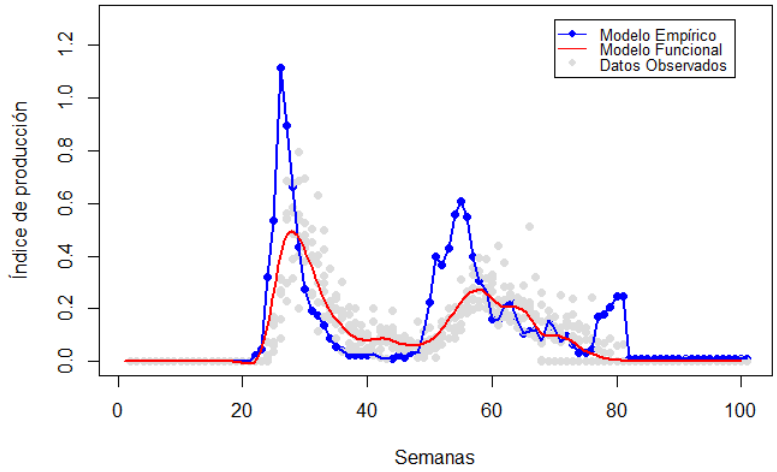
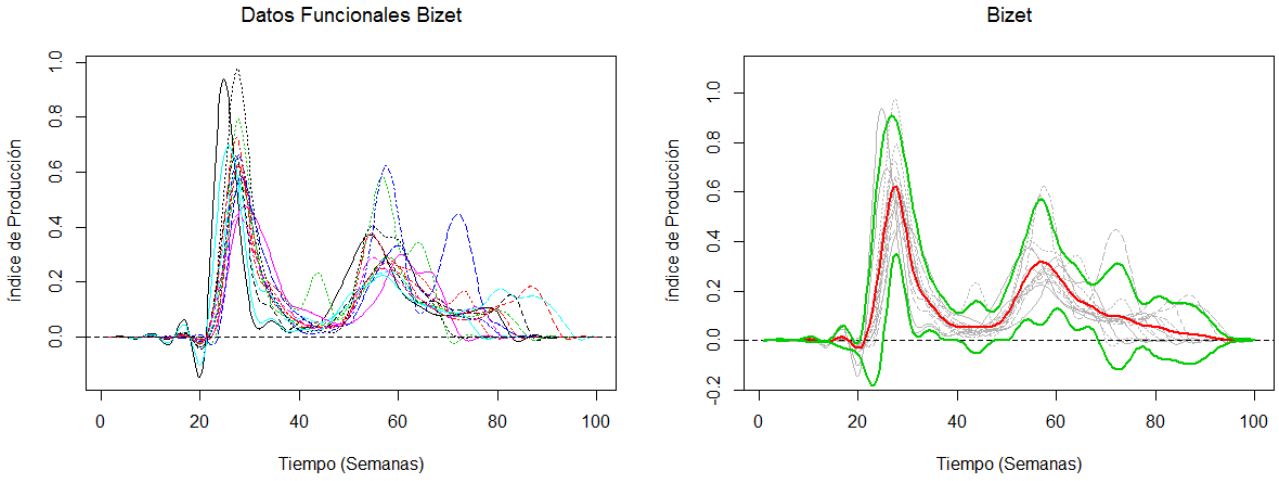


Figura 8.13: Comparación Modelo Empírico y Modelo Estimado Variedad Atacama

8.3.2. Bizet



(a) (b)

Figura 8.14: (a) Datos funcionales variedad Bizet. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

El modelo construido por medio de datos funcionales, estima que la variedad Bizet alcanzará su máximo en la semana 27 con un índice de producción de 0.616 y un error de estimación de 0.180, error bajo comparado con el error del modelo actual que es de 0.414 para el pico 1. Además, el pico 2 se alcanzará en la semana 57 y no en la semana 60 como muestra el modelo empírico. El porcentaje de cobertura del modelo empírico usado por la empresa es de 79 % dentro del modelo ajustado.

	Pico 1		Pico 2	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	1.030	27	0.557	60
Modelo Estimado	0.616	27	0.316	57

Cuadro 8.5: Modelos de producción variedad Bizet

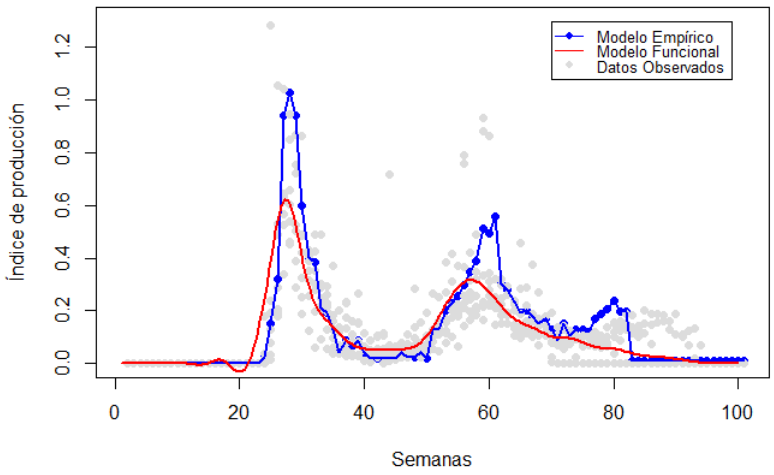
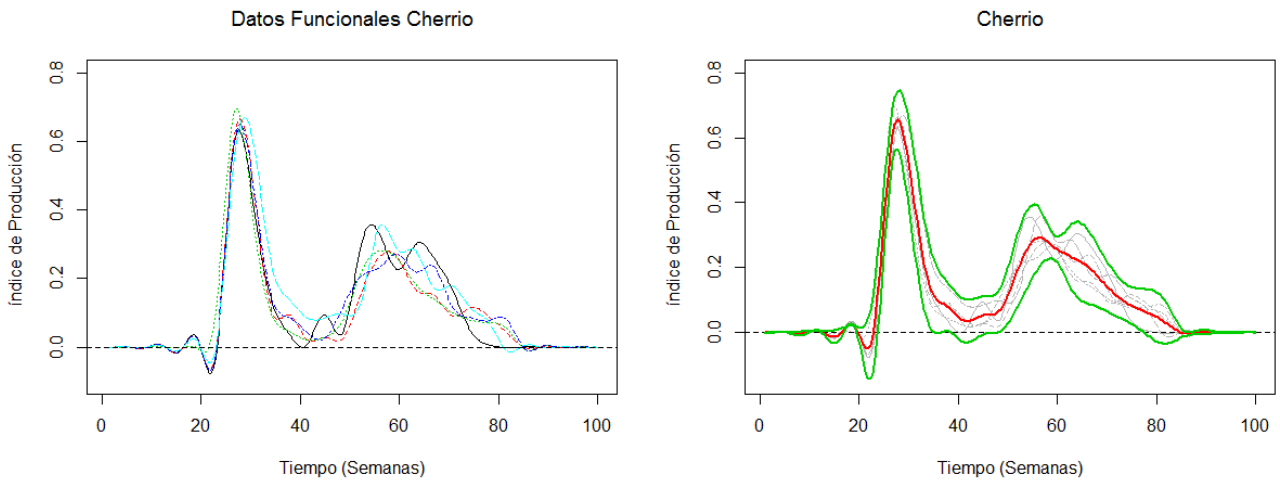


Figura 8.15: Comparación Modelo Empírico y Modelo Estimado Variedad Bizet

8.3.3. Cherrio



(a) (b)

Figura 8.16: (a) Datos funcionales variedad Cherrio. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

A continuación se presenta los valores estimados con la media funcional de los índices de producción en el pico y repico de la variedad Cherrio. El valor promedio de producción más alto se obtuvo en la semana 28 correspondiente a 0.706, este valor es menor al usado por la empresa actualmente (0.933). Por otro lado el índice de producción estimado en el re pico fue de 0.343 en la semana 58. Aunque los índices de producción del modelo empírico son mayores que los del modelo estimado tanto para el pico 1 como para el pico 2, los errores de estimación de este son bajos y similares a los del modelo de datos funcionales (0.127 modelo empírico, 0.130 modelo estimado, para el pico 1 y 0.042 modelo empírico, 0.054 modelo estimado, para el pico 2). Lo anterior muestra que para la variedad Cherrio el modelo empírico no tenía un mal desempeño y podría representar bien los datos como el modelo estimado con los datos funcionales, sin embargo, el porcentaje de cobertura del modelo empírico usado por la empresa es de 77 % dentro del modelo ajustado, razón por la cual, se decide finalmente usar el modelo estimado.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.933	28	0.343	58
Modelo Estimado	0.706	28	0.294	56

Cuadro 8.6: Modelos de producción variedad Cherrio

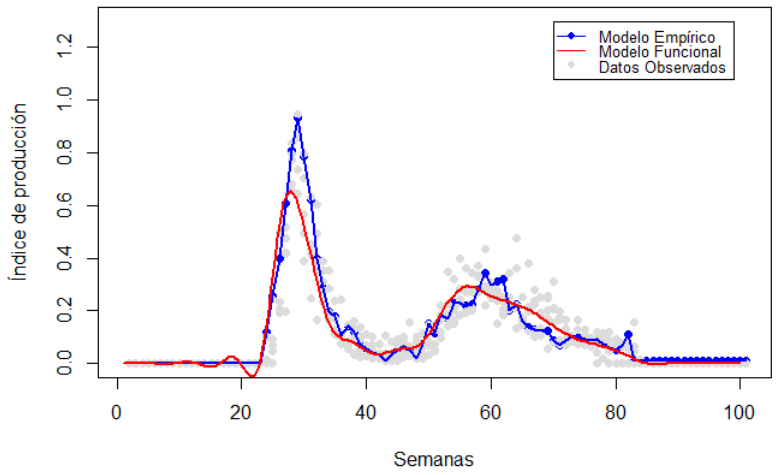
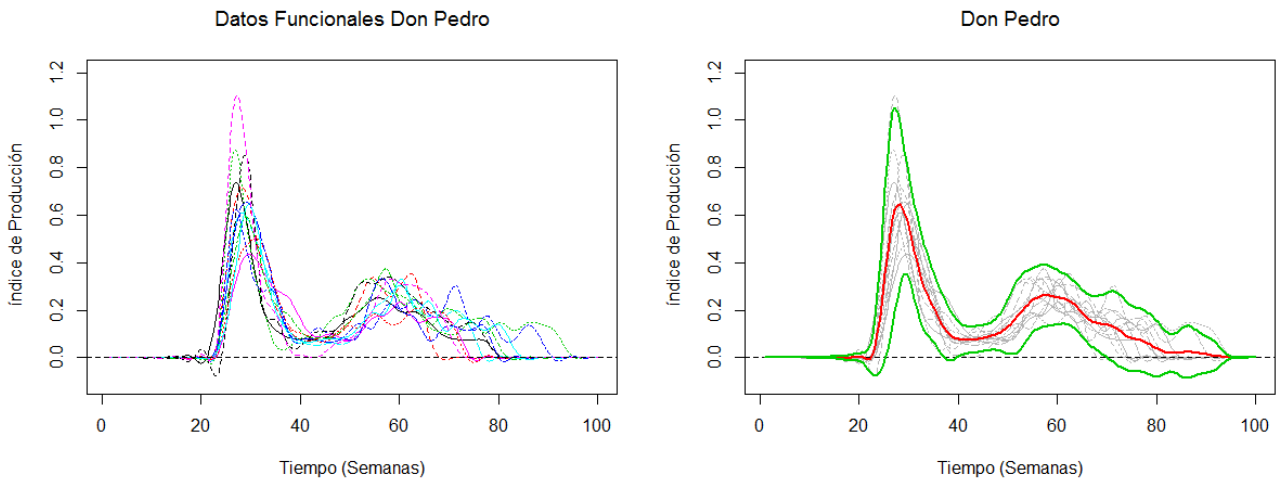


Figura 8.17: Comparación Modelo Empírico y Modelo Estimado Variedad Cherrio

8.3.4. Don Pedro

A continuación se presenta los valores estimados con la media funcional de los índices de producción en el pico y repico de la variedad Don Pedro. El valor promedio de producción más alto se obtuvo en la semana 28 correspondiente a 0.671 con un error de estimación de 0.145, este valor es menor al usado por la empresa actualmente (0.890) el cual tiene un error mayor de 0.334. Por otro lado el índice de producción estimado en el re pico fue de 0.261 en la semana 57 con un error de estimación de 0.067, valor muy bajo, lo cual sugiere que el modelo estimado puede representar muy bien el comportamiento real de



(a) (b)

Figura 8.18: (a) Datos funcionales variedad Don Pedro. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

la siembra. El porcentaje de cobertura del modelo empírico usado por la empresa es de 92 % dentro del modelo ajustado.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.890	27	0.394	58
Modelo Estimado	0.671	28	0.261	57

Cuadro 8.7: Modelos de producción variedad Don Pedro

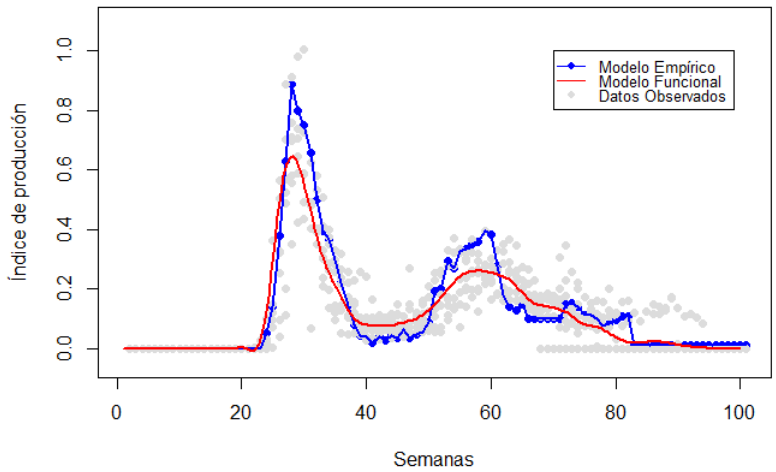
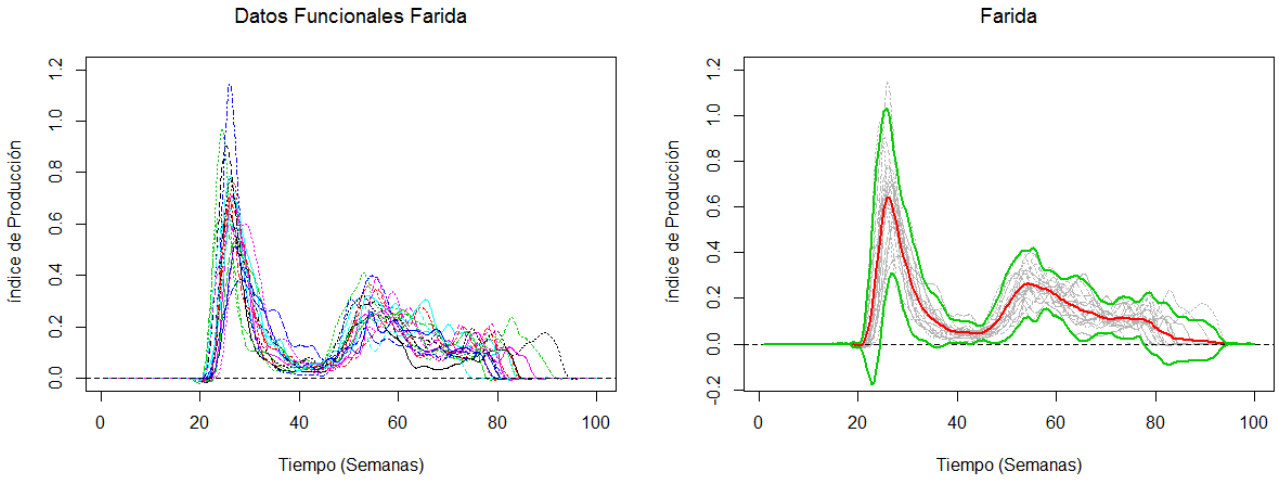


Figura 8.19: Comparación Modelo Empírico y Modelo Estimado Variedad Don Pedro

8.3.5. Farida



(a) (b)

Figura 8.20: (a) Datos funcionales variedad Farida. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

El modelo estimado para la variedad Farida obtuvo el promedio de producción más alto correspondiente al pico 1 en la semana 26 de 0.645 con un error de estimación equivalente a 0.173, este valor es menor al usado por la empresa actualmente (0.870) el cual tiene un error mayor de 0.249. Se espera además, con un error de 0.082, tener una producción de 0.262 en el pico 2 en la semana 54, producción menor a la que se espera con el modelo empírico el cual tiene un error de estimación de 0.158 en este pico. Se puede observar también, que para esta variedad no se presenta error en la estimación de la semana de producción, pues los dos modelos estiman las mismas semanas en las cuales se alcanzarán los picos. El porcentaje de cobertura del modelo empírico usado por la empresa es de 85 % dentro del modelo ajustado.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.870	26	0.420	54
Modelo Estimado	0.645	26	0.262	54

Cuadro 8.8: Modelos de producción variedad Farida

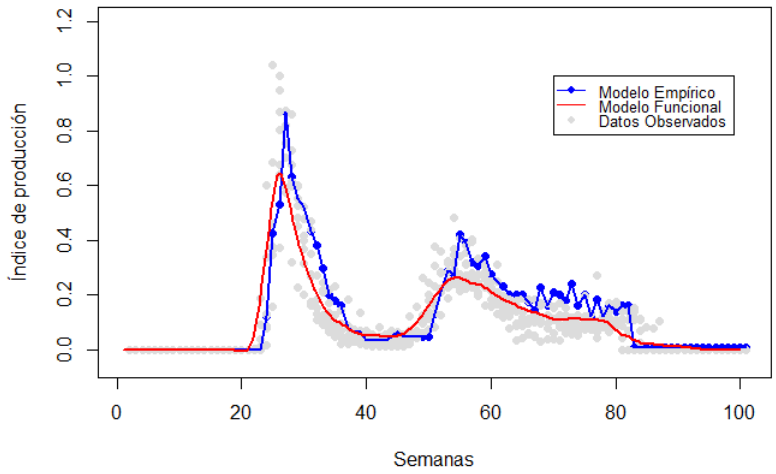
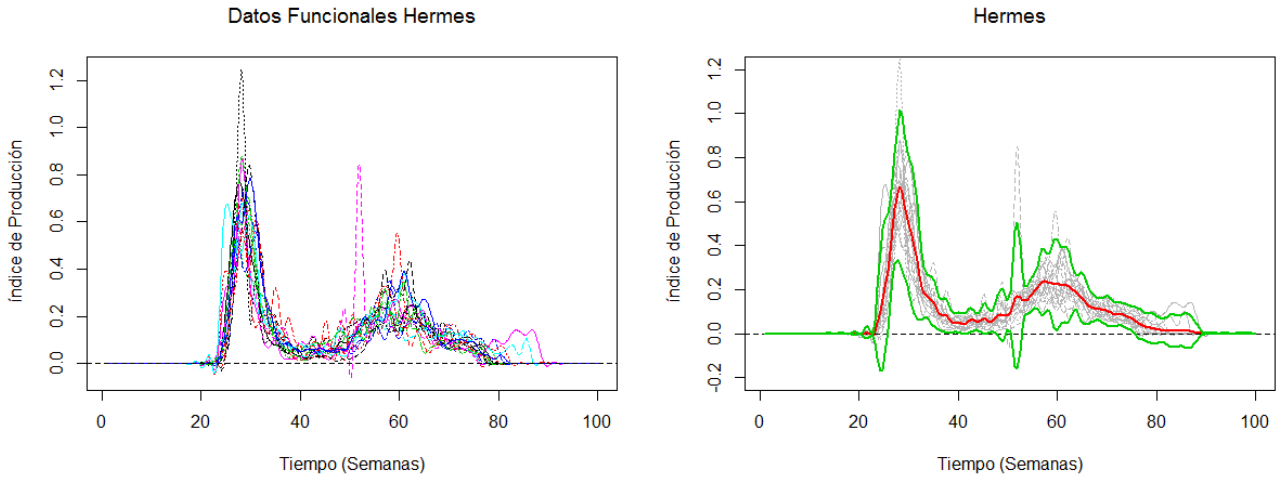


Figura 8.21: Comparación Modelo Empírico y Modelo Estimado Variedad Farida

8.3.6. Hermes



(a) (b)

Figura 8.22: (a) Datos funcionales variedad Hermes. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

Se puede observar cómo el modelo empírico usado actualmente por la empresa presenta índices de producción mayores al estimado por medio de los datos funcionales, este sobreajuste se puede medir mejor con el error de estimación presentado por el modelo empírico, el cual equivale a 0.612 en el pico 1 y de 0.242 en el pico 2. Adicional al error presentado en los índices de producción, se observa cómo el modelo empírico estima los picos anticipadamente con una diferencia de 6 semanas en el pico 2. El porcentaje de cobertura del modelo empírico usado por la empresa es de 67 % dentro del modelo ajustado, porcentaje bajo, por lo cual se recomienda usar el modelo estimado.

	Pico 1		Pico 2	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.927	26	0.391	53
Modelo Estimado	0.695	28	0.242	59

Cuadro 8.9: Modelos de producción variedad Hermes

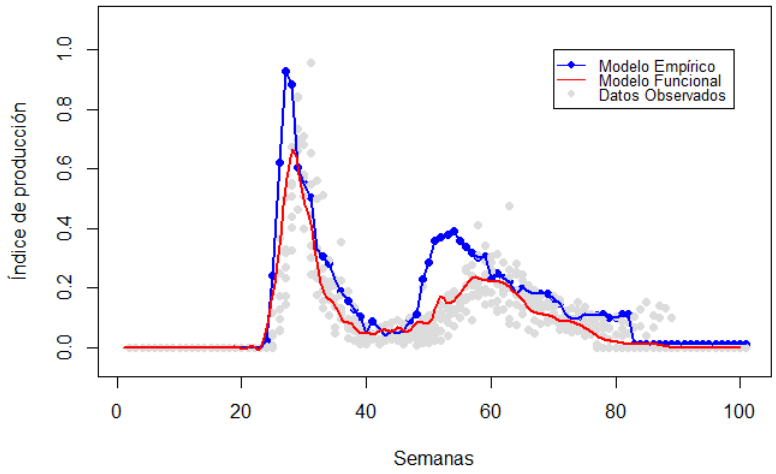
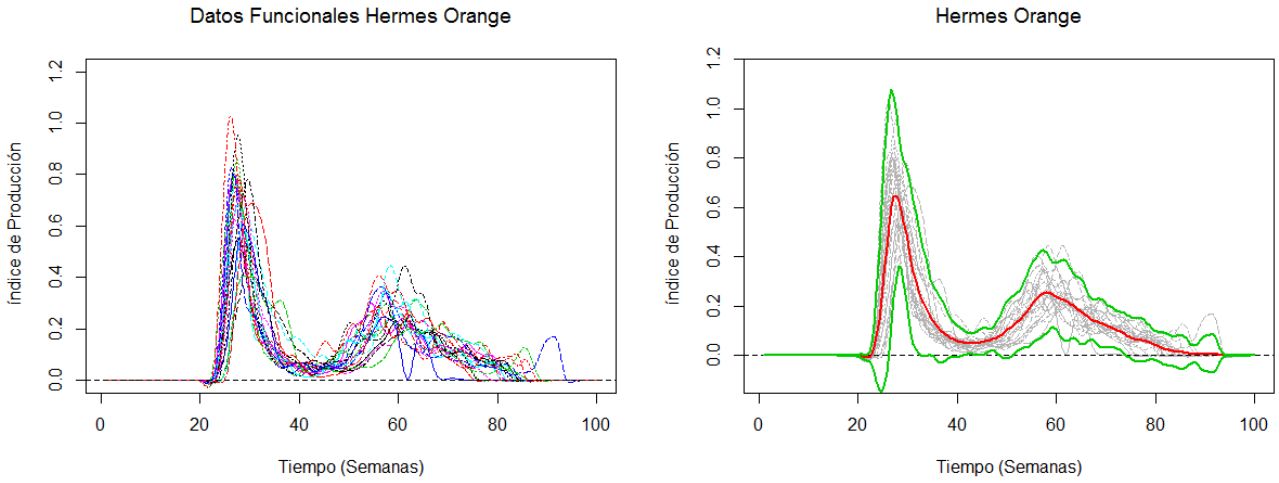


Figura 8.23: Comparación Modelo Empírico y Modelo Estimado Variedad Hermes

8.3.7. Hermes Orange



(a) (b)

Figura 8.24: (a) Datos funcionales variedad Hermes Orange. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

A continuación se presenta los valores estimados con la media funcional de los índices de producción

en el pico y repico de la variedad Hermes Orange. El valor promedio de producción más alto se obtuvo en la semana 28 correspondiente a 0.640 con un error de estimación de 0.117, error de estimación menor al del modelo empírico usado actualmente que equivale a 0.498. Por otro lado, para el pico 2, aunque los modelos estiman los máximos en diferentes semanas, los índices de producción estimados no son muy diferentes y los errores de cada uno son cercanos y bajos (0.098 modelo empírico, 0.072 modelo estimado), lo cual sugiere que el problema de estimación del índice de producción se presenta sólo en el primer pico. El porcentaje de cobertura del modelo empírico usado por la empresa es de 88 % dentro del modelo ajustado.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.988	26	0.343	60
Modelo Estimado	0.640	28	0.254	58

Cuadro 8.10: Modelos de producción variedad Hermes Orange

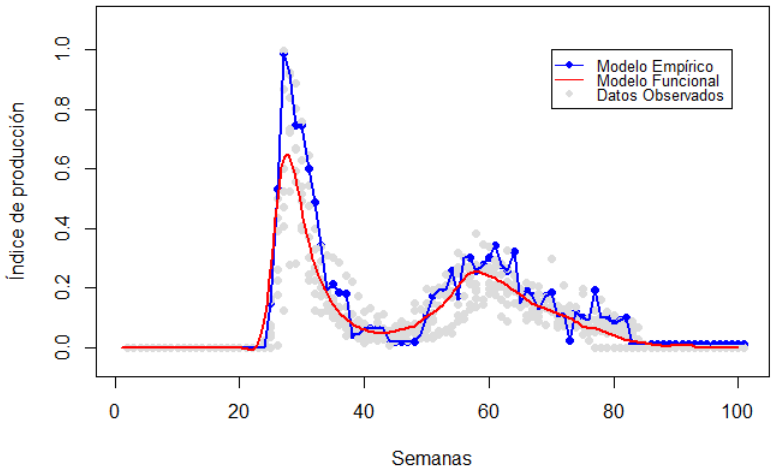
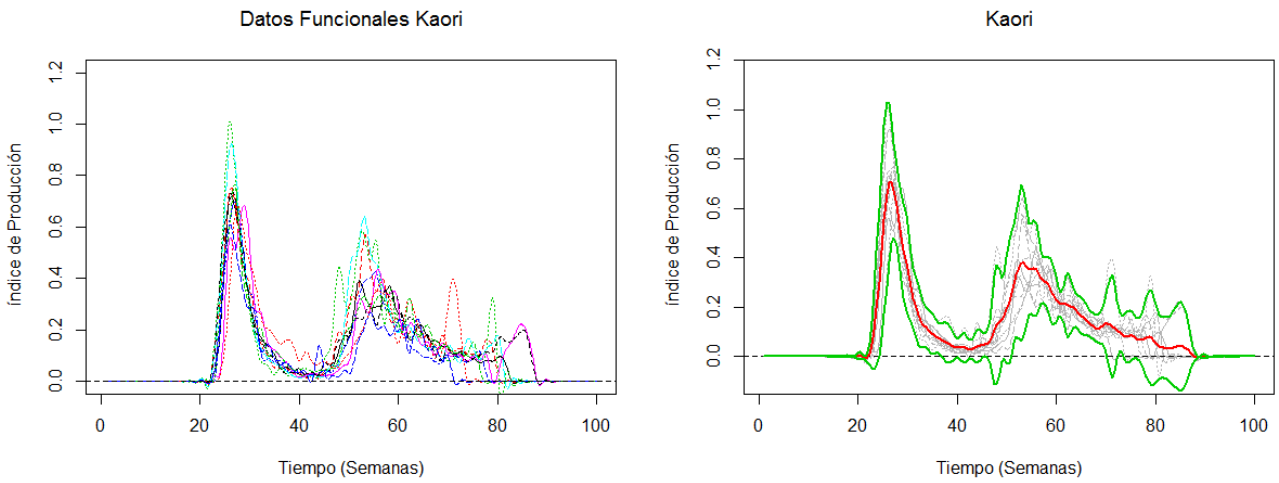


Figura 8.25: Comparación Modelo Empírico y Modelo Estimado Variedad Hermes Orange

8.3.8. Kaori

Para la creación de los datos funcionales de la variedad Kaori, se eligieron 79 bases de funciones B-spline de orden 4, no fue necesaria una penalización por rugosidad. En la parte izquierda de la figura 8.26 se muestra los 10 datos suavizados que corresponden a 10 siembras de la variedad Kaori en diferentes espacios de tiempo (años) y diferentes bloques (tierra). El porcentaje de cobertura del modelo empírico usado por la empresa es de 82 % dentro del modelo ajustado.

Para esta variedad, las diferencias en los índices de producción de los dos modelos equivalen a 0.199 en el pico 1 y a 0.149 en el pico 2, estando por encima las estimaciones del modelo empírico en el pico 1 y pico 2 con errores equivalentes a 0.212 y 0.205 respectivamente. El modelo estimado por medio de datos funcionales estima una producción de 0.700 para el primer pico con un error de estimación de 0.151



(a) (b)

Figura 8.26: (a) Datos funcionales variedad Kaori. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

y para el pico 2 de 0.380 con un error de estimación de 0.122, lo cual sugiere que el modelo estimado representa mejor el comportamiento real de la producción de esta variedad.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.899	27	0.529	52
Modelo Estimado	0.700	26	0.380	53

Cuadro 8.11: Modelos de producción variedad Kaori

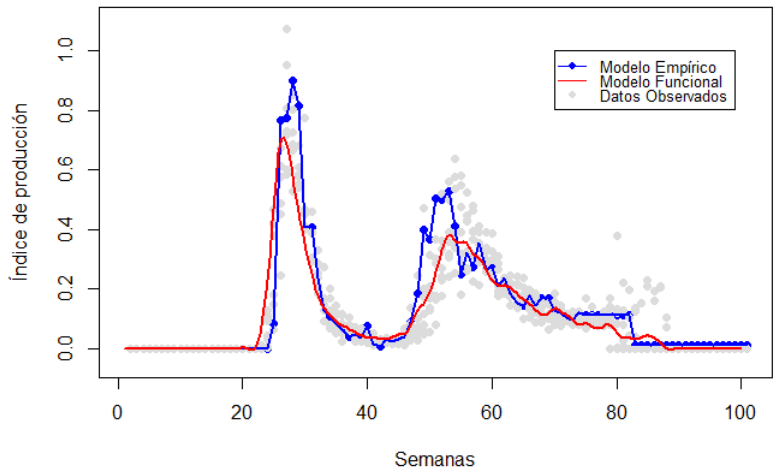
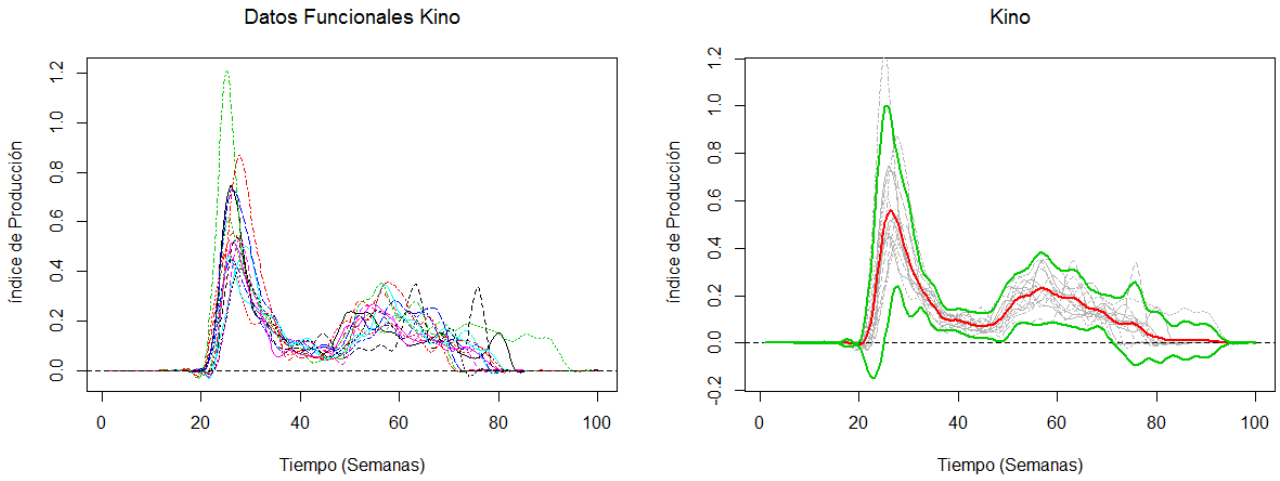


Figura 8.27: Comparación Modelo Empírico y Modelo Estimado Variedad Kaori

8.3.9. Kino



(a) (b)

Figura 8.28: (a) Datos funcionales variedad Kino. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

El valor promedio de producción más alto obtenido por el modelo estimado con datos funcionales, se obtuvo en la semana 26 correspondiente a 0.562 , este valor es menor al usado por la empresa actualmente (0.937) el cual presenta un error de estimación de 0.359. Para el segundo pico, la diferencia de los índices de producción entre los dos modelos es baja, equivalente a 0.071, sin embargo, el modelo estimado presenta un error menor (0.072) al modelo empírico (0.110) debido a que el modelo actual estima la producción del segundo pico 4 semanas antes de lo real. El porcentaje de cobertura del modelo empírico usado por la empresa es de 85 % dentro del modelo ajustado.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.937	25	0.303	53
Modelo Estimado	0.562	26	0.232	57

Cuadro 8.12: Modelos de producción variedad Kino

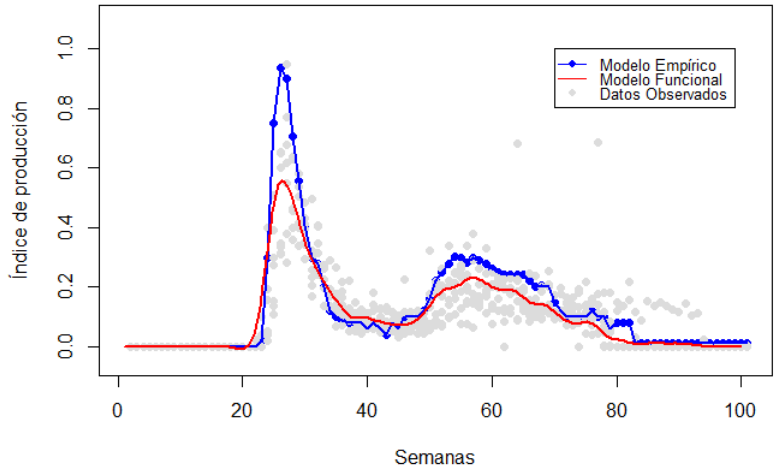
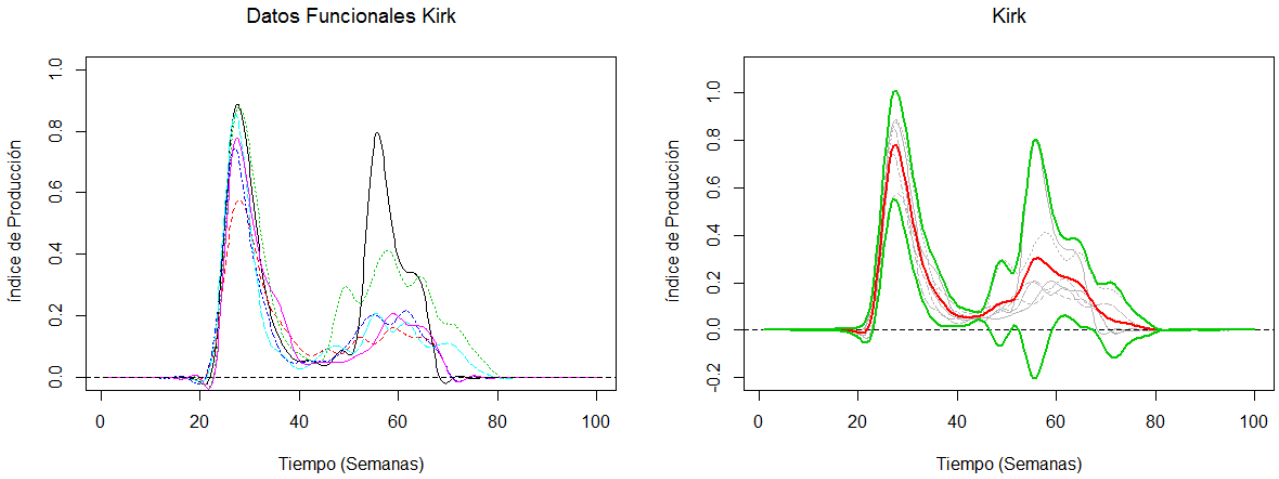


Figura 8.29: Comparación Modelo Empírico y Modelo Estimado Variedad Kino

8.3.10. Kirk



(a) (b)

Figura 8.30: (a) Datos funcionales variedad Kirk. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

A continuación se presenta los valores estimados con la media funcional de los índices de producción en el pico y repico de la variedad Kirk. El valor promedio de producción más alto se obtuvo en la semana 27 correspondiente a 0.775, índice menor al usado por la empresa actualmente (0.937). En cuanto a los errores de estimación de los modelos en el primer pico, el modelo estimado presenta un error bajo (0.082) menor al del modelo actual (0.153). Por otro lado la diferencia entre los índices de producción estimados en el segundo pico es mayor a la del primero, y sus errores son más grandes, equivalentes a 0.369 para el modelo empírico y 0.273 para el modelo estimado. El porcentaje de cobertura del modelo empírico usado por la empresa es de 70 % dentro del modelo ajustado.

	Pico 1		Pico 2	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.941	28	0.557	60
Modelo Estimado	0.775	27	0.315	57

Cuadro 8.13: Modelos de producción variedad Kirk

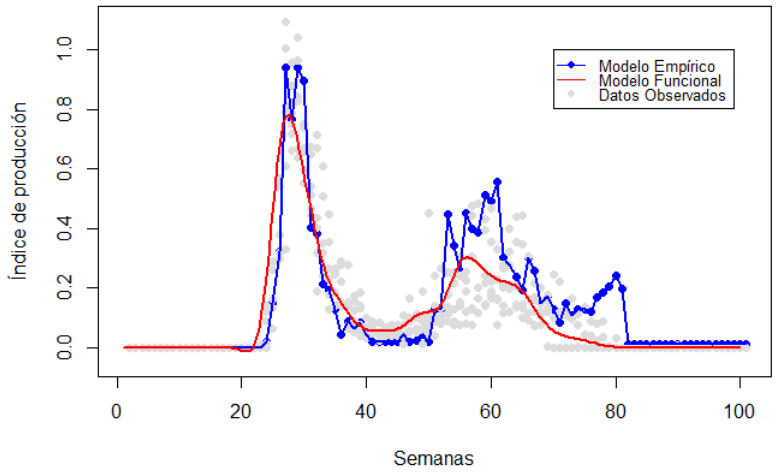
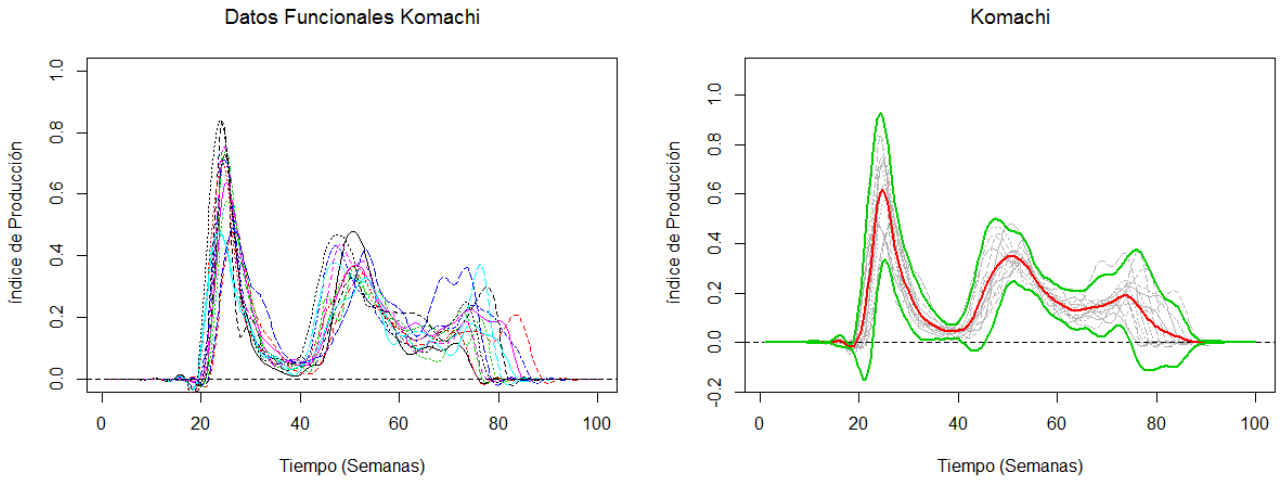


Figura 8.31: Comparación Modelo Empírico y Modelo Estimado Variedad Kirk

8.3.11. Komachi



(a) (b)

Figura 8.32: (a) Datos funcionales variedad Komachi. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

En esta variedad, los modelos estimados presentan diferencias altas en los índices de producción, en el pico 1 la diferencia es de 0.378 y en el pico 2 de 0.157, mostrando un sobreajuste en el modelo empírico con errores de 0.339 y 0.150 en cada pico. El modelo estimado por medio de datos funcionales presenta un error bajo en la estimación del pico 2 equivalente a 0.062, mostrando su buen ajuste a los datos reales. Se observa además, que esta variedad alcanza el pico 2 en semanas tempranas comparada con las demás variedades. El porcentaje de cobertura del modelo empírico usado por la empresa es de 75 % dentro del modelo ajustado.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.975	24	0.508	50
Modelo Estimado	0.597	25	0.351	51

Cuadro 8.14: Modelos de producción variedad Komachi

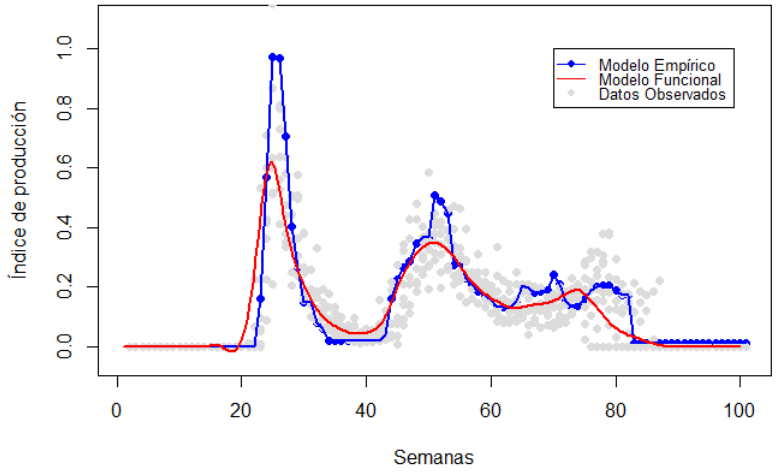
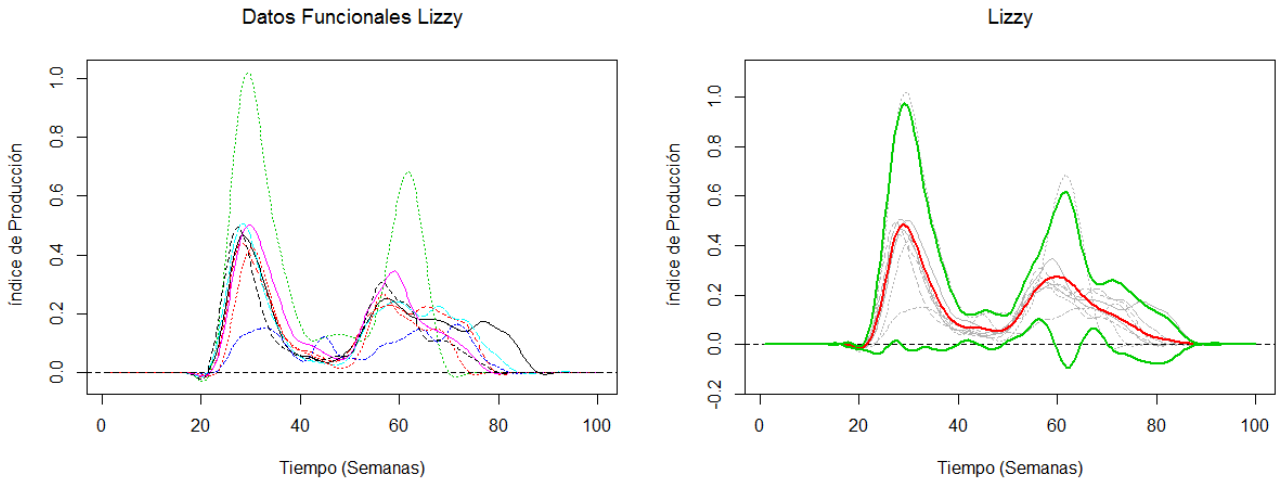


Figura 8.33: Comparación Modelo Empírico y Modelo Estimado Variedad Komachi

8.3.12. Lizzy

La variedad Lizzy es la variedad que presenta mayor variabilidad en las curvas de producción y un mayor sobreajuste del modelo empírico el cual alcanza un máximo de 1.115 en el pico 1 con un error de estimación de 0.924 y un índice de producción de 0.997 en el segundo pico con un error de 0.699. Los errores reportados muestran el mal desempeño del modelo que está usando la empresa para estimar los índices de producción y las semanas en las cuales se alcanzan los picos. El modelo estimado por medio de datos funcionales se ajusta bien al comportamiento real de la producción, con un error de estimación de 0.098 en el primer pico y de 0.112 en el segundo pico. El porcentaje de cobertura del modelo empírico usado por la empresa es de 63 % dentro del modelo ajustado, porcentaje bajo lo cual sugiere que el modelo



(a) (b)

Figura 8.34: (a) Datos funcionales variedad Lizzy. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

empírico está en muchos puntos por encima de las bandas del modelo estimado y no sería bueno utilizarlo.

	Pico 1		Pico 2	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	1.115	25	0.997	57
Modelo Estimado	0.483	29	0.273	60

Cuadro 8.15: Modelos de producción variedad Lizzy

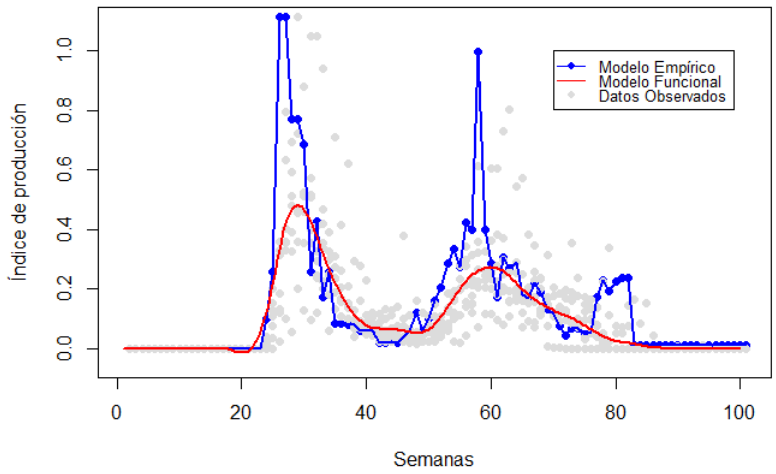
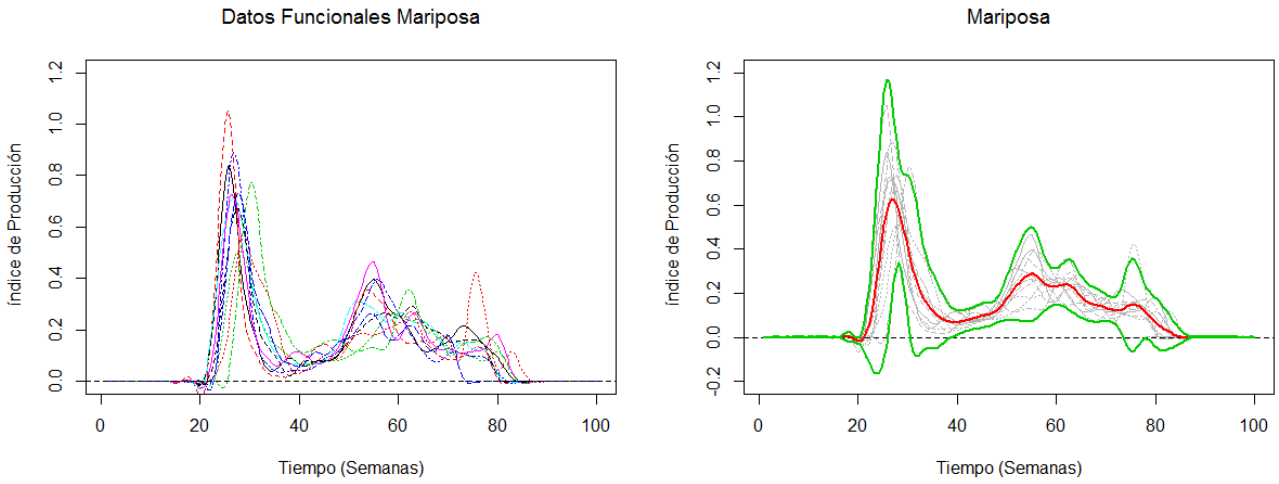


Figura 8.35: Comparación Modelo Empírico y Modelo Estimado Variedad Lizzy

8.3.13. Mariposa



(a) (b)

Figura 8.36: (a) Datos funcionales variedad Mariposa. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

A continuación se presenta los valores estimados con la media funcional de los índices de producción en el pico y repico de la variedad Mariposa. El valor promedio de producción más alto se obtuvo en la semana 27 correspondiente a 0.622 con un error de estimación de 0.195, este valor es menor al usado por la empresa actualmente (0.946) el cual tiene un error de estimación mayor (0.393). La diferencia entre los índices de producción de ambos modelos es de 0.324 y 0.131 para el pico 1 y 2 respectivamente, respecto a la semana estimada de producción se observa mayor diferencias en el pico 2 con un error de 0.173 para el modelo empírico. El porcentaje de cobertura del modelo empírico usado por la empresa es de 73 % dentro del modelo ajustado.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.946	26	0.413	60
Modelo Estimado	0.622	27	0.282	55

Cuadro 8.16: Modelos de producción variedad Mariposa

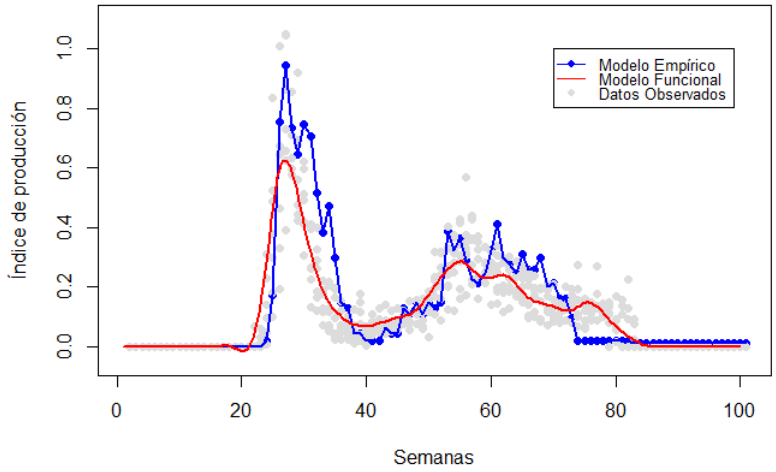
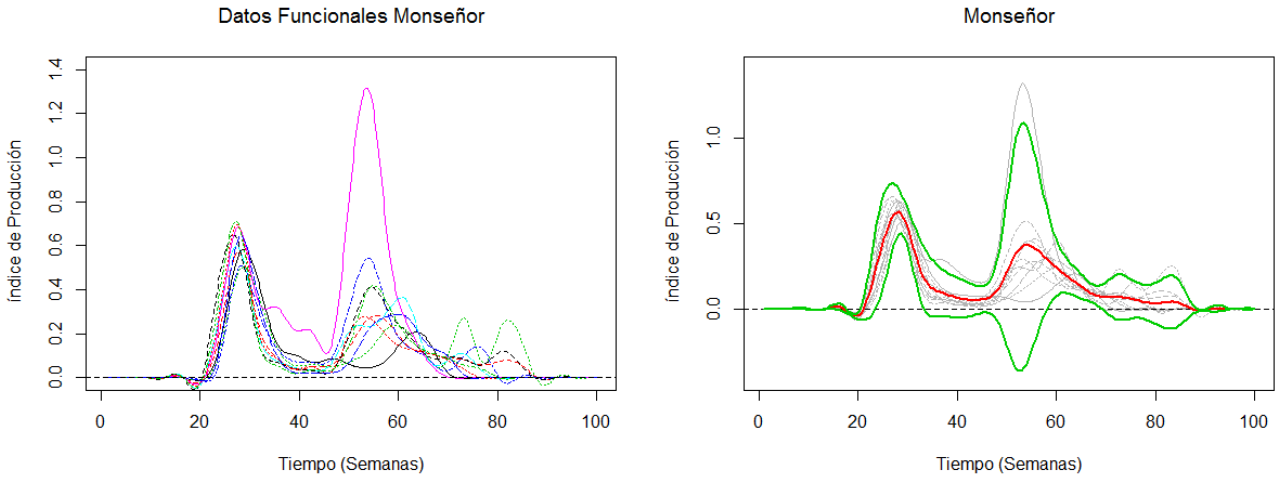


Figura 8.37: Comparación Modelo Empírico y Modelo Estimado Variedad Mariposa

8.3.14. Monseñor



(a) (b)

Figura 8.38: (a) Datos funcionales variedad Monseñor. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

Esta variedad presenta una diferencia muy alta entre los índices de producción estimados del modelo empírico y el modelo construido con datos funcionales para el pico 1, correspondiente a 0.632. El error del modelo empírico para este pico es de 0.570, error alto comparado con el del modelo estimado el cual es de 0.122. En el segundo pico, se observa una mayor diferencia en la estimación de la semana para alcanzar el pico, teniendo el modelo empírico un retraso de 7 semanas y un error de 0.401. El porcentaje de cobertura del modelo empírico usado por la empresa es de 76 % dentro del modelo ajustado.

	Pico 1		Pico 2	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	1.212	27	0.588	61
Modelo Estimado	0.580	27	0.381	54

Cuadro 8.17: Modelos de producción variedad Monseñor

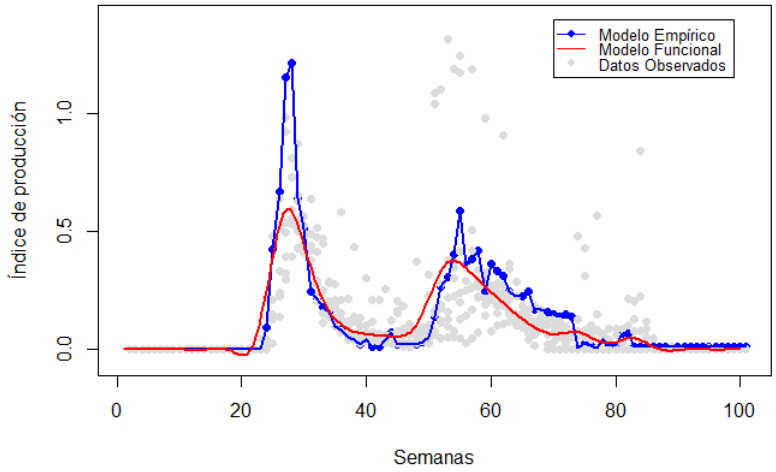
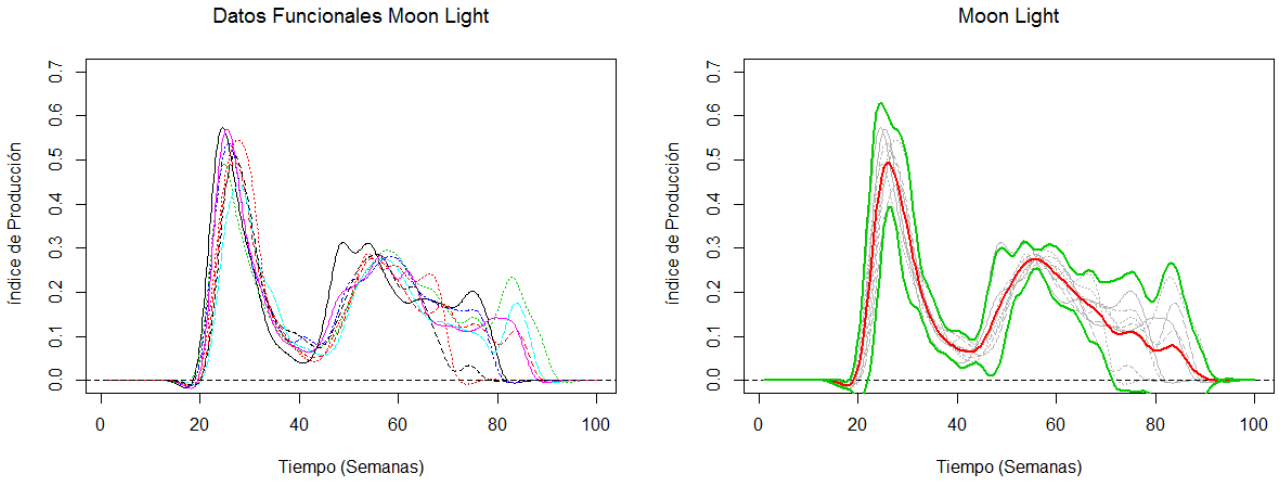


Figura 8.39: Comparación Modelo Empírico y Modelo Estimado Variedad Monseñor

8.3.15. Moonlight



(a) (b)

Figura 8.40: (a) Datos funcionales variedad Moonlight. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

El modelo construido por medio de los datos funcionales para esta variedad, se ajusta bien al compor-

tamiento de la producción real, pues tiene un error de estimación de 0.069 para el pico 1 y de 0.032 para el pico 2. Estos resultados nos permiten concluir que el modelo estimado podría tener un buen desempeño y que la producción real no es tan alta como el modelo empírico lo sugiere. Los errores de estimación del modelo empírico encontrados equivalen a 0.320 y 0.230 para el pico 1 y 2 respectivamente. El porcentaje de cobertura del modelo empírico usado por la empresa es de 66 % dentro del modelo ajustado.

	Pico 1		Pico 2	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.855	25	0.500	55
Modelo Estimado	0.494	26	0.274	56

Cuadro 8.18: Modelos de producción variedad Moonlight

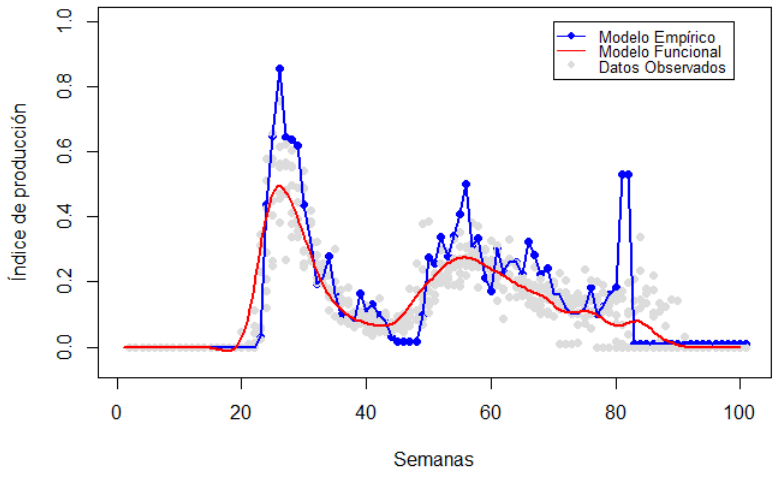
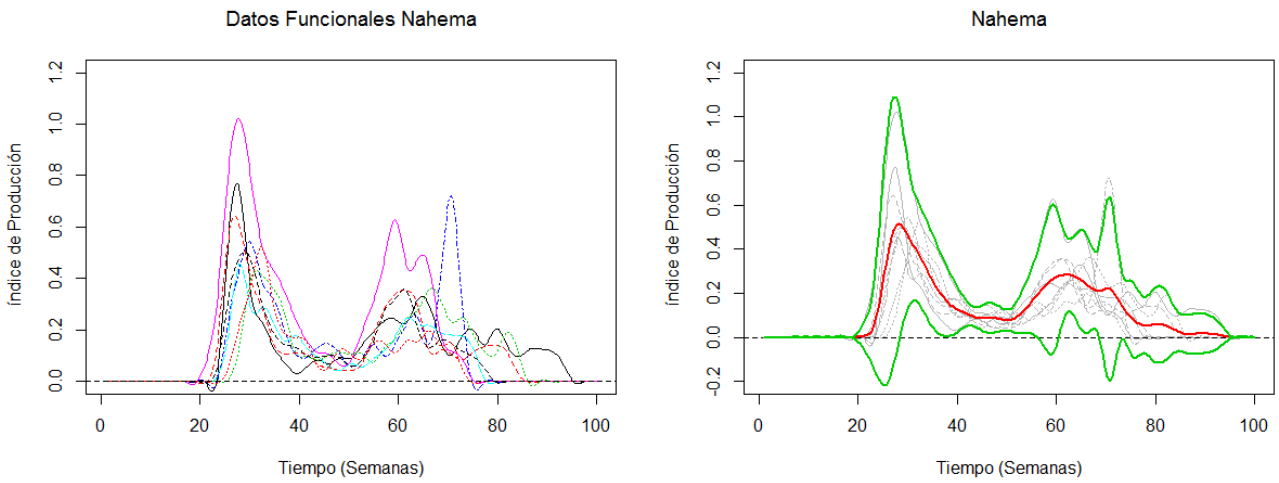


Figura 8.41: Comparación Modelo Empírico y Modelo Estimado Variedad Moonlight

8.3.16. Nahema

El modelo empírico que usa la empresa actualmente, estima que la producción para el pico 1 es de 1.010 en la semana 29, esta estimación tiene una diferencia de 0.498 respecto al índice que estima el modelo construido con datos funcionales y presenta un error de 0.541. Lo anterior sugiere, que el modelo empírico no tiene un buen desempeño para estimar la producción real y que el modelo estimado podría resultar mejor ya que tiene un error más bajo para el pico 1 (0.233). En cuanto al segundo pico, los modelos no presentan diferencias grandes en los índices de producción estimados ni en los errores encontrados (0.190 modelo empírico, 0.102 modelo estimado). El porcentaje de cobertura del modelo empírico usado por la empresa es de 85 % dentro del modelo ajustado.



(a) (b)

Figura 8.42: (a) Datos funcionales variedad Nahema. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

	Pico 1		Pico 2	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	1.010	29	0.340	60
Modelo Estimado	0.512	28	0.283	61

Cuadro 8.19: Modelos de producción variedad Nahema

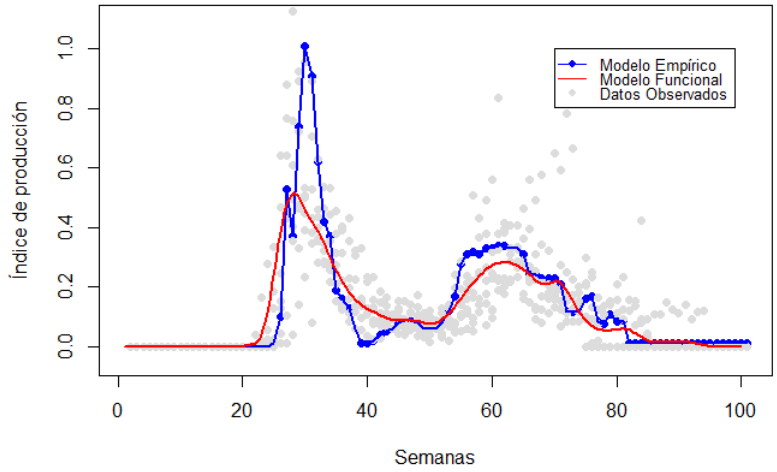
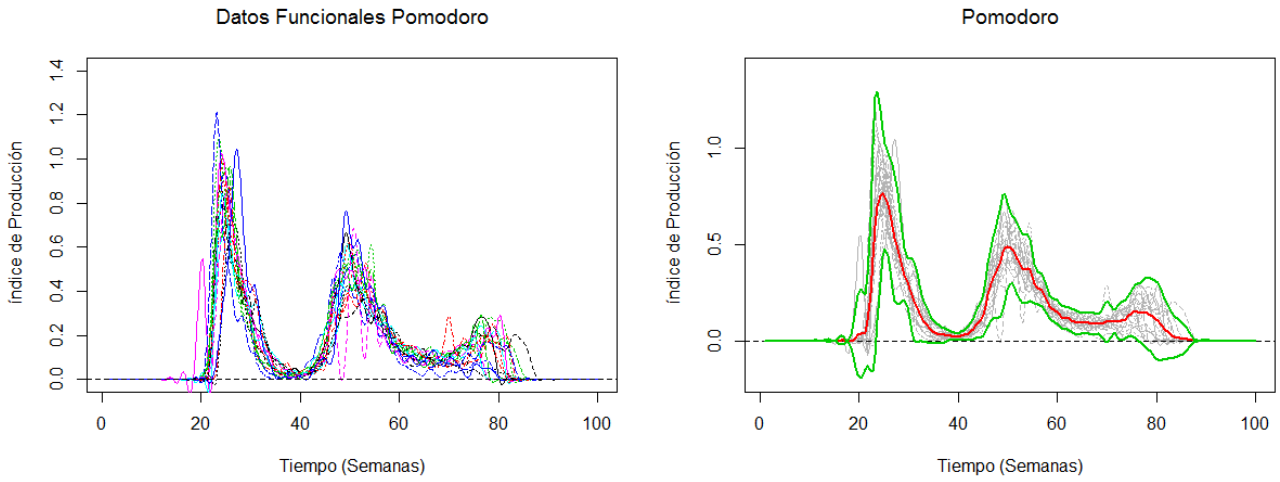


Figura 8.43: Comparación Modelo Empírico y Modelo Estimado Variedad Nahema

8.3.17. Pomodoro



(a) (b)

Figura 8.44: (a) Datos funcionales variedad Pomodoro. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

Como se observa en la gráfica anterior, esta variedad presenta alta variabilidad entre las curvas de producción y un sobreajuste en el modelo empírico usado por la empresa. La diferencia entre el modelo actual y el modelo estimado es de 0.531 y 0.305 para el pico 1 y 2 respectivamente, presentando el modelo empírico un error de estimación de 0.323 en el primer pico y de 0.190 en el segundo. Por otro lado, el modelo estimado con datos funcionales, muestra un mejor desempeño en la estimación con errores de 0.133 y 0.163 en cada pico. El porcentaje de cobertura del modelo empírico usado por la empresa es de 82 % dentro del modelo ajustado.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	1.058	25	0.615	52
Modelo Estimado	0.527	28	0.310	60

Cuadro 8.20: Modelos de producción variedad Pomodoro

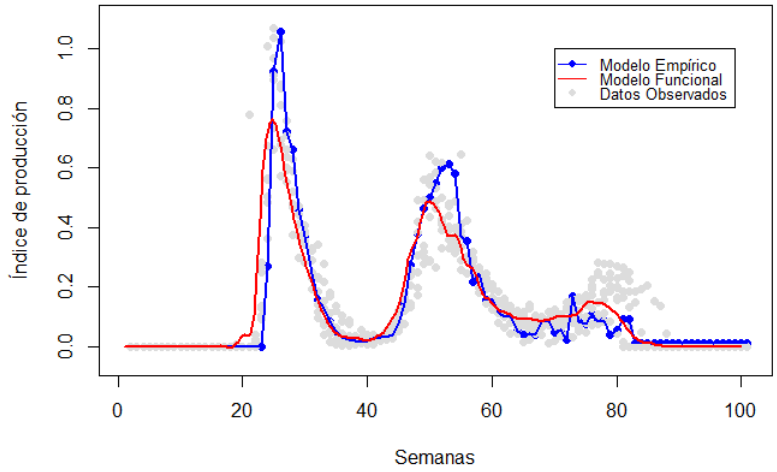
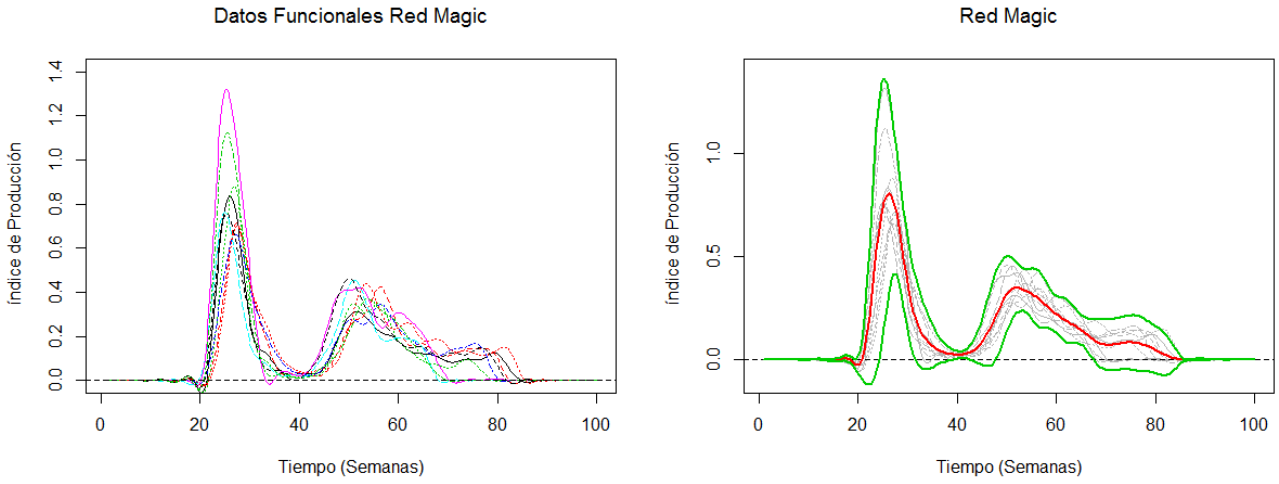


Figura 8.45: Comparación Modelo Empírico y Modelo Estimado Variedad Pomodoro

8.3.18. Red Magic



(a) (b)

Figura 8.46: (a) Datos funcionales variedad Red Magic. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

Para esta variedad no se observan diferencias altas entre los índices de producción de los dos modelos, sin embargo sigue siendo mayor la producción estimada por el modelo empírico con errores de 0.185 y 0.089 para el pico 1 y 2 respectivamente, comparado con los errores del modelo estimado equivalentes a 0.157 y 0.094. Lo anterior sugiere que aunque el modelo empírico no presenta mal desempeño, el modelo estimado podría ser mejor para representar el comportamiento de la producción real. El porcentaje de cobertura del modelo empírico usado por la empresa es de 93 % dentro del modelo ajustado.

	Pico 1		Pico 2	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	0.962	26	0.262	57
Modelo Estimado	0.815	26	0.358	58

Cuadro 8.21: Modelos de producción variedad Red Magic

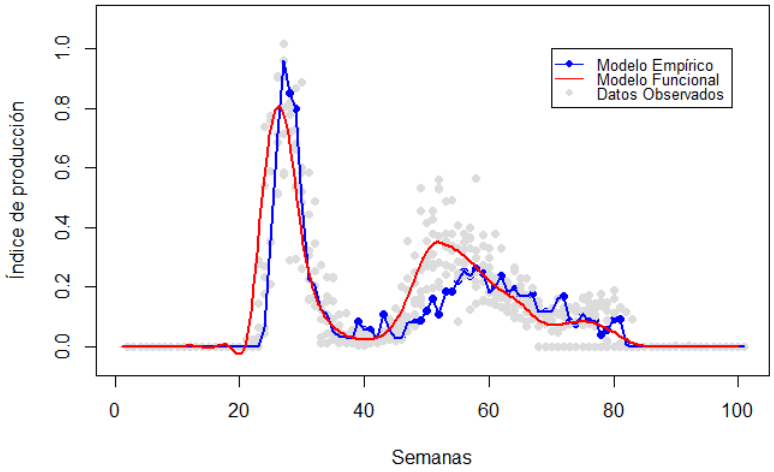
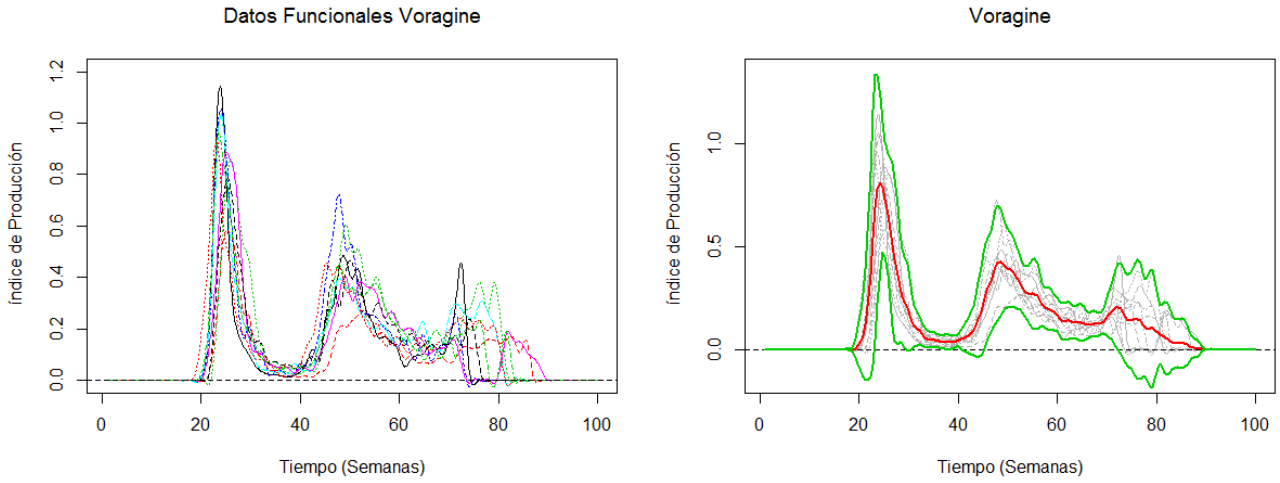


Figura 8.47: Comparación Modelo Empírico y Modelo Estimado Variedad Red Magic

8.3.19. Voragine



(a) (b)

Figura 8.48: (a) Datos funcionales variedad Voragine. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

El modelo empírico usado por la empresa para estimar la producción de esta variedad, presenta un

sobreajuste, mostrando una diferencia en el primer pico respecto al modelo estimado de 0.583 con un error de 0.571. Para el pico 2 el error presentado por el modelo empírico equivale a 0.138. Estos errores de estimación se deben al sobreajuste del modelo en los índices de producción y no a la estimación de las semanas en las cuales alcanzan los picos. El modelo estimado por medio de datos funcionales, presenta el promedio de producción más alto en la semana 24 correspondiente a 0.817 con un error de 0.222. El porcentaje de cobertura del modelo empírico usado por la empresa es de 86 % dentro del modelo ajustado.

	Pico 1		**Pico 2**	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	1.400	24	0.507	50
Modelo Estimado	0.817	24	0.401	50

Cuadro 8.22: Modelos de producción variedad Voragine

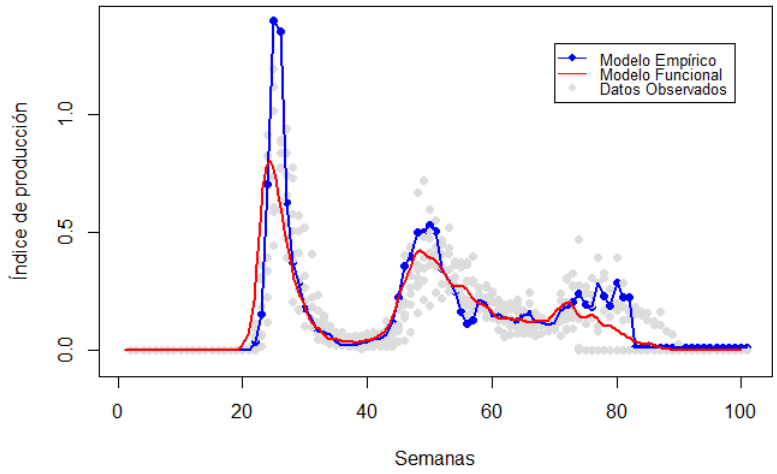
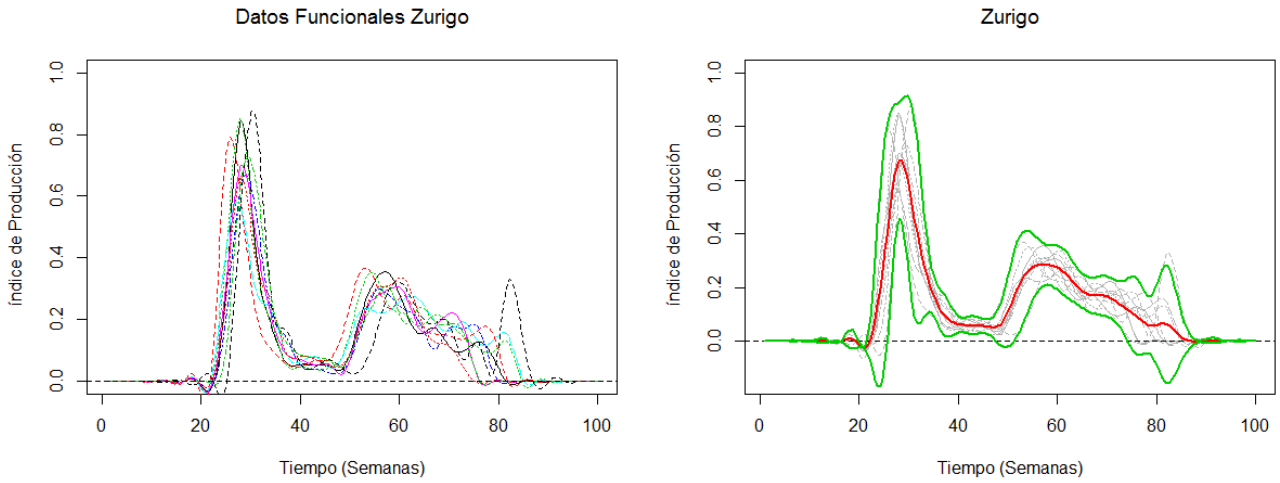


Figura 8.49: Comparación Modelo Empírico y Modelo Estimado Variedad Voragine

8.3.20. Zurigo

Para esta variedad, el modelo construido con datos funcionales presenta una estimación para la producción en el pico 1 menor a la del modelo empírico, con una diferencia de 0.439 por debajo y un error de estimación de 0.132, lo cual sugiere que tiene un buen desempeño además de que el error de estimación del modelo empírico equivale a 0.433. En el pico 2, el error del modelo estimado también es menor al del modelo empírico (0.049 y 0.284 respectivamente). El porcentaje de cobertura del modelo empírico usado por la empresa es de 63 % dentro del modelo ajustado.



(a) (b)

Figura 8.50: (a) Datos funcionales variedad Zurigo. (b) Función de la media (rojo) y bandas de variabilidad funcional (verde)

	Pico 1		Pico 2	
	IP	**Semana**	**IP**	**Semana**
Modelo Empírico	1.100	28	0.536	54
Modelo Estimado	0.661	28	0.284	58

Cuadro 8.23: Modelos de producción variedad Zurigo

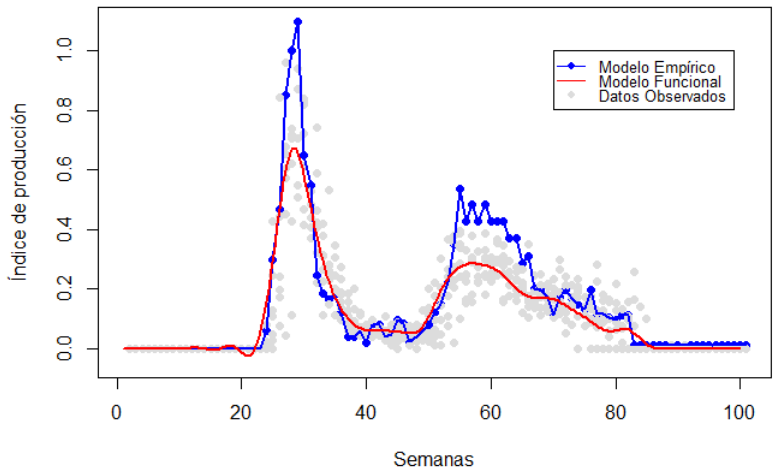


Figura 8.51: Comparación Modelo Empírico y Modelo Estimado Variedad Zurigo

8.4. Resumen de los modelos

En la siguiente tabla, se presenta el resumen de los resultados del modelo empírico y el modelo estimado para cada variedad en el pico 1, permitiendo comparar los índices de producción, las semanas en las cuales alcanza el pico y los errores de estimación.

Variedad	**Modelo Empírico**			**Modelo Ajustado**				**Δ Error** %
	IP	Semana	Error	IP	Bandas de Var.	Semana	Error	
Atacama	1.114	25	0.852	0.497	0.278;0.750	28	0.143	83.22
Bizet	1.030	27	0.414	0.616	0.330;0.962	27	0.180	56.52
Cherrio	0.933	28	0.127	0.706	0.652;0.759	28	0.130	2.31
Don Pedro	0.890	27	0.334	0.671	0.334;1.059	28	0.145	56.59
Farida	0.870	26	0.249	0.645	0.310;1.023	26	0.173	30.52
Hermes	0.927	26	0.612	0.695	0.349;1.144	28	0.164	73.20
Hermes Orange	0.988	26	0.498	0.640	0.351;1.072	28	0.117	76.51
Kaori	0.899	27	0.212	0.700	0.466;1.081	26	0.151	28.77
Kino	0.937	26	0.359	0.562	0.224;0.973	26	0.220	38.72
Kirk	0.941	28	0.153	0.775	0.549;1.003	27	0.082	46.41
Komachi	0.975	24	0.339	0.597	0.318;0.957	25	0.173	48.97
Lizzy	1.115	25	0.924	0.483	0.012;0.613	29	0.098	89.39
Mariposa	0.946	26	0.393	0.622	0.330;1.151	27	0.195	50.38
Monseñor	1.212	27	0.570	0.580	0.403;0.802	27	0.122	78.60
Moonlight	0.855	25	0.320	0.494	0.395;0.629	26	0.069	78.44
Nahema	1.010	29	0.541	0.512	0.166;1.106	28	0.233	56.93
Pomodoro	1.058	25	0.323	0.527	0.211;1.182	28	0.133	58.82
Red Magic	0.962	26	0.185	0.815	0.433;1.369	26	0.157	15.14
Voragine	1.400	24	0.571	0.817	0.487;1.330	24	0.222	61.12
Zurigo	1.100	28	0.433	0.661	0.384;0.978	28	0.132	69.52

Cuadro 8.24: Comparación modelos para el primer pico

Se observa para todas las variedades que el modelo empírico tiene índices de producción altos que varían entre 0.855 y 1.400 alcanzados entre las semanas 24 y 29, comparados con los de los modelos estimados los cuales fluctúan entre 0.483 y 0.817 en las mismas semanas. Las variedades con los índices de producción promedio estimados más altos son Voragine, Red Magic y Kirk, y las variedades con los índices de producción más bajo son Lizzy, Moonlight y Atacama. Los modelos empíricos que presentaron mayor error de estimación o ajuste fueron los de las variedades Lizzy y Atacama, curvas que lograron ajustarse con los modelos estimados con datos funcionales los cuales presentaron errores menores de 0.098 y 0.143 para cada variedad respectivamente.

Los cambios en los errores porcentuales mayores se observaron en la variedad Lizzy y Atacama con un cambio porcentual de 89.38 % y 83.22 % respectivamente, mostrando que eran modelos empíricos con muy mal desempeño pero que con la técnica estadística aplicada se pudo lograr un buen ajuste al comportamiento real de la producción.

La tabla que se presenta a continuación, muestra el resumen de los resultados del modelo empírico y el modelo estimado para cada variedad en el pico 2, permitiendo comparar los índices de producción, las semanas en las cuales alcanza el pico y los errores de estimación.

Variedad	**Modelo Actual**			**Modelo Ajustado**				**Δ Error** %
	IP	Semana	Error	IP	Bandas de Var.	Semana	Error	
Atacama	0.609	54	0.431	0.273	0.196;0.359	58	0.049	88.63
Bizet	0.557	60	0.299	0.316	0.117;0.565	58	0.145	51.51
Cherrio	0.343	58	0.042	0.294	0.229:0.403	56	0.054	22.22
Don Pedro	0.394	58	0.136	0.261	0.145;0.402	57	0.067	50.74
Farida	0.420	54	0.158	0.262	0.154;0.411	54	0.082	48.10
Hermes	0.391	53	0.242	0.242	0.108;0.561	59	0.080	66.94
Hermes Orange	0.343	60	0.098	0.254	0.109;0.426	58	0.072	26.53
Kaori	0.529	52	0.205	0.380	0.212;0.692	53	0.122	40.49
Kino	0.303	53	0.110	0.232	0.087;0.391	57	0.078	29.09
Kirk	0.557	60	0.369	0.315	0.076;0.825	57	0.273	26.02
Komachi	0.508	50	0.150	0.351	0.259;0.468	51	0.062	58.67
Lizzy	0.997	57	0.699	0.273	0.106;0.613	60	0.112	83.98
Mariposa	0.413	60	0.173	0.286	0.147;0.495	55	0.100	42.40
Monseñor	0.588	61	0.401	0.381	0.081;1.092	54	0.294	26.68
Moonlight	0.500	55	0.230	0.274	0.255;0.317	56	0.032	86.09
Nahema	0.340	60	0.190	0.283	0.135;0.596	61	0.102	46.32
Pomodoro	0.615	52	0.190	0.310	0.148;0.743	60	0.163	14.21
Red Magic	0.262	57	0.089	0.358	0.233;0.504	58	0.094	5.32
Voragine	0.507	50	0.138	0.401	0.208;0.605	50	0.073	47.10
Zurigo	0.536	54	0.300	0.284	0.211;0.426	58	0.049	83.67

Cuadro 8.25: Comparación modelos para el segundo pico

Para el segundo pico, los índices de producción de los modelos empíricos variaron entre 0.262 y 0.977 con errores de ajuste entre 0.042 y 0.699, índices de producción altos comparados con los de los modelos estimados con datos funcionales los cuales variaron entre 0.232 y 0.401 con errores de estimación de 0.032 a 0.29a, ambos modelos alcanzaron el pico 2 entre la semana 50 y 61. Las variedades con mayores índices de producción promedio fueron Voragine, Monseñor y Kaori y los que reportaron índices más bajos fueron Kino, Hermes y Hermes Orange. Los modelos empíricos que presentaron mayores errores de estimación o ajuste fueron Lizzy y Atacama, errores que disminuyen con los nuevos modelos estimados los cuales fueron de 0.112 y 0.049 para cada variedad respectivamente. Por otro lado, los modelos empíricos con menor error de estimación fueron Cherrio y Hermes Orange, los cuales podrían sugerir que representaban bien el comportamiento de los datos, ya que fueron similares a los reportados por los modelos estimados.

Los cambios en los errores porcentuales mayores se observaron en la variedad Atacama y Moonlight con un cambio porcentual de 88.63 % y 86..09 % respectivamente, mostrando que eran modelos empíricos con muy mal desempeño en el segundo pico pero que con la técnica estadística aplicada se pudo lograr un buen ajuste al comportamiento real de la producción.

Capítulo 9

Conclusiones

Este trabajo propone y desarrolla una metodología para ajustar las curvas de producción de la empresa C.I Flores de Aposentos S.A.S registradas desde el año 2012 hasta el 2016 y estimar un modelo que permita ayudarles con la toma de decisiones administrativas y de producción. A continuación se presentan las conclusiones obtenidas a partir de los resultados del estudio y los futuros trabajos.

9.1. Conclusiones Generales

Para ajustar las curvas de producción de esta empresa para cada una de las 20 variedades, se usó la técnica estadística de datos funcionales la cual permitió tener una representación de la naturaleza continua de la producción de flores por cada variedad. Las suavizaciones realizadas por medio de los B-Spline se ajustaron bien al comportamiento real de los datos en todas las variedades, logrando tener la tendencia de los picos y re picos de producción en las semanas que se esperaban. Esto demuestra que esta técnica fue apropiada para resumir nuestra información.

En el análisis exploratorio clásico se pudo observar que aunque las 20 variedades tenían problemas de ajuste del modelo usado actualmente por la empresa, presentando errores de estimación en el primer pico entre 0.127 y 0.924, habían variedades en donde el mal ajuste era muy evidente gráficamente como las variedades Atacama, Hermes, Lizzy, Monseñor, Nahema y Zurigo en las cuales se pudo comprobar por medio de los errores de estimación que el modelo definitivamente no seguía el comportamiento real de la producción, mostrando sobre ajuste en el pico y repico. Al construir la media funcional, se obtuvo un modelo coherente con la producción registrada por la empresa, en cada variedad se pudo observar cómo la media funcional seguía el comportamiento de los datos en cada semana de vida de la planta con errores de estimación bajos entre 0.069 y 0.233, dando como resultados índices de producción esperados y por debajo de los indicados en el modelo actual.

Las medias funcionales fueron presentadas con sus respectivas bandas de variabilidad, estas bandas nos permitieron observar entre cuales valores fluctuaban los índices de producción medios en cada semana de vida de la planta, cubriendo el 95.5 % de las siembras registradas por cada variedad. En algunas variedades como Bizet, Hermes, Kaori, Kirk, Lizzy, Mariposa, Monseñor, Nahema y Pomodoro se observó bandas de variabilidad muy anchas en algunos segmentos de tiempo, esto pudo ser causado por la variabilidad que presentaban estas variedades entre las siembras y por algunas siembras que podrían considerarse como atípicas dado que se salían del comportamiento usual de producción.

Las coberturas de los modelos empíricos dentro de los modelos estimados se calcularon con el fin de saber si un modelo empírico aunque tuviera un error de estimación alto podría tener un comportamiento de producción dentro de las bandas de variabilidad del modelo estimado, pues indicaría que aunque no es la producción que se esperaría tener siempre en esa variedad, si es un comportamiento que se puede dar. Para el primer pico, los modelos empíricos que tuvieron menor error de estimación y coberturas altas que podrían acercarse al comportamiento real de la producción son Cherrio, Kirk y Red Magic con porcentajes de cobertura de 77 %, 70 % y 93 % respectivamente. Para el segundo pico, los modelos empíricos de las variedades Cherrio y Hermes Orange tuvieron errores de estimación bajos con porcentajes de cobertura altos de 77 % y 88 % cada uno. Lo anterior sugiere, que el modelo empírico de la variedad Cherrio puede tener un buen desempeño para representar la producción real de los datos ya que tiene buena cobertura y bajos errores de estimación en ambos picos.

En general, los modelos que tuvieron mejor cobertura fueron los de las variedades Red Magic (93 %), Don Pedro (92 %) y Hermes Orange (88 %). Red Magic tuvo alta cobertura y errores de estimación bajos en ambos picos (0.185 y 0.089), Don Pedro tuvo alta cobertura y errores bajos en ambos picos (0.334 y 0.136) y Hermes Orange tuvo cobertura alta y error medio en el primer pico (0.498) y en el segundo pico error bajo (0.098), por lo cual el modelo empírico de la variedad Red Magic y de Don Pedro, también podrían estar representando bien el comportamiento de los datos.

Por otro lado, los modelos empíricos que tuvieron menor cobertura fueron los de las variedades Atacama (61 %), Lizzy (63 %), Zurigo (62 %), Moonlight (66 %) y Hermes (67 %). Lizzy tuvo un porcentaje de cobertura bajo y errores de estimación altos em ambos picos (0.924 y 0.699) y Zurigo, tuvo porcentaje de cobertura bajo y errores de estimación medios en ambos picos(0.433 y 0.300) los cual demuestra que el modelo empírico de la variedad Lizzy es el modelo con peor desempeño para representar el comportamiento de la producción real en esa variedad.

El haber construido estos nuevos modelos representados con la media funcional para cada variedad, permite a la empresa tener una herramienta útil con estimaciones adecuadas sobre la producción de claveles y mini claveles y así, evitar ineficiencias económicas o en la producción. Los modelos ajustados permitirán identificar si la producción de la empresa puede cubrir una demanda, al igual que tener seguridad a la hora de ofertar sus flores, ya que cuenta con índices de producción promedios esperados con un margen de error durante cada semana. De este modo, la empresa podrá manipular sus cultivos para obtener la producción deseada en las fechas que requieran.

9.2. Trabajos Futuros

Realizar el ajuste de las curvas de producción para las 34 variedades que no se trabajaron en este proyecto por motivos de limitación al objetivo, podría ser beneficioso para la empresa, ya que el problema del ajuste del modelo usado actualmente se presenta en todas las variedades.

Se recomienda el uso de otros estadísticos descriptivos funcionales como la mediana funcional para estimar el modelo y evaluar el ajuste comparado con el estimado en este estudio (media funcional).

Bibliografía

Arevalo, G. A., Ibarra, D., and Flórez, V. J. (2007). Desbotone en diferentes estadios de desarrollo del botón floral en clavel estándar (dianthus caryophyllus l.) var. nelson. *Agronomía Colombiana*, 25:73 – 82.

Asocolflores (2015). Boletin estadístico dirección económica y logística. *Asociación Colombiana de Exportadores de Flores*.

Asocolflores (2016a). Cifras estadísticas. `https://www.asocolflores.org/servicios/cifras-estadisticas/36`. [Consultada: 2017-10-16].

Asocolflores (2016b). Informe de gestión 2016. `http://www.asocolflores.org/comunicaciones/noticias/conozca-el-impacto-social,-economico-y-ambiental-de-la-gestion-2016-de-asocolflores/160/1`. [Consultada: 2017-10-16].

Benhenni, K., Ferraty, F., Rachdi, M., and Vieu, P. (2007). Local smoothing regression with functional data. *Computational Statistics*, 22:353–369.

Bi, J. and Kuesten, C. (2013). Using functional data analysis (fda) methodology and the r package fda"for sensory time-intensity evaluation. *Journal of Sensory Studies*, 28:474–482.

DANE (2010). Informe de resultados, censo de fincas productoras de flores en 28 municipios de la sabana de bogotá y cundinamarca 2009.

Delicado, P. (2008). Curso de modelos no paramétricos. `http://www-eio.upc.es/~delicado/docencia/Apuntes_Models_No_Parametrics.pdf`.

Febrero-Bande, M. (2008). A present overview on functional data analysis. *Boletín de estadística e investigación operativa, Universidad de Santiago de Compostela*, 24:6–12.

Febrero-Bande, M. and Manuel, O. d. l. F. (2012). Statistical computing in functional data analysis: The r package fda.usc. *Journal of Statistical Software*, 51(4).

Ferraty, F. (2011). *Recent Advances in Functional Data Analysis and Related Topics*. Springer.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer.

FloralHolland, R. (2016). Carnation production in colombia is growing. https://www.royalfloraholland.com/en/speciale-paginas/search-in-news/v44446/carnation-production-in-colombia-is-growing. [Consultada: 2017-10-17].

Giraldo, R. (2007). Análisis exploratorio de variables regionalizadas con métodos funcionales. *Revista Colombiana de Estadística*, 30:115–127.

Gómez, I. (2010). *Floricultura Colombiana*. Universidad Santo Tomás.

Hooker, G. (2010). Functional data analysis, a short course.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Chapman and Hall/CRC Press.

Jacques, J. and Preda, C. (2013). *Functional data clustering: a survey*. Springer.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. Chapman and Hall/CRC Press.

Ministerio de Agricultura, C. (2015a). Colombia se consolida como segundo exportador de flores del mundo y se expande a nuevos mercados. https://www.minagricultura.gov.co/noticias/Paginas/Colombia-se-expande-a-nuevos-mercados-flores.aspx. [Consultada: 2016-10-21].

Ministerio de Agricultura, C. (2015b). Esperamos exportar 500 millones de flores a estados unidos: Minagricultura. https://www.minagricultura.gov.co/noticias/Paginas/Esperamos-exportar-500-millones-de-flores-.aspx. [Consultada: 2016-10-22].

Montoya, D. M. and Reyes, G. A. (2005). *Geología de la Sabana de Bogotá*. Instituto Colombiano de Geología y Minería INGEOMINAS.

Müller, H.-G. (2011). *Functional Data Analysis*. Springer Berlin Heidelberg.

Navas, M. C. N. (2001). Comparación de cultivos hidropónicos y en suelo de snapdragon y goderia en la finca rosas sabanilla. Master's thesis, Universidad de la Sabana.

Olaya, J., Reina, J., Torres, L. L., Paz, M. C., and Pereira, L. A. (2014). Modelación no paramétrica de la contaminación promedio octohoraria del aire debida al monóxido de carbono y al ozono troposférico. *Revista Ingeniería y Competitividad*, 16(1).

Organización de las Naciones Unidas para la Agricultura y la Alimentación, F. (2002). *El cultivo protegido en clima mediterráneo*. FAO.

Patiño, V. and Martínez, J. (2013). Modelo de regresión funcional simple para estudiar la contaminación promedio octohoraria de monóxido de carbono a partir de una variable clima. Master's thesis, Universidad del Valle.

Portaforlio, D. (2016). Las exportaciones de flores colombianas cayeron 5,7%. http://www.portafolio.co/negocios/exportaciones-flores-colombianas-febrero-2016-491086. [Consultada: 2016-10-21].

Procolombia (2016). Oportunidades de negocio en el sector flores y plantas vivas. http://www.procolombia.co/node/1242. [Consultada: 2017-10-17].

Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53:539–572.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer.

Reina, J. and Olaya, J. (2012). Ajuste de curvas mediante métodos no paramétricos para estudiar el comportamiento de contaminación del aire por material particulado pm10. *Revista EIA*, pages 19 – 31.

Rimache, M. (2011). *Floricultura, cultivo y comercialización*. Ediciones de la U.

Salinger, J. P. (1992). *Comercial Flower Growing*. Inkata Press.

Universidad Nacional Abierta y a Distancia, U. (2013). Lección 11. cultivo del clavel. http://datateca.unad.edu.co. [Consultada: 2016-11-07].

Vila Arboleda, J. J. (2009). *Modelo de Proyección de la Producción de Rosas, Basado en las Curvas de Crecimiento de las Plantas*. Universidad de la Salle.

Capítulo 10

Anexos

10.1. Gráficas Exploratorias de las Variedades

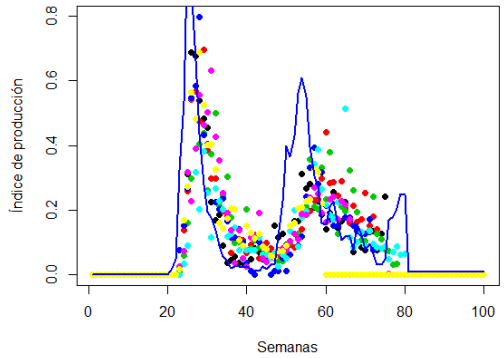


Figura 10.1: Atacama

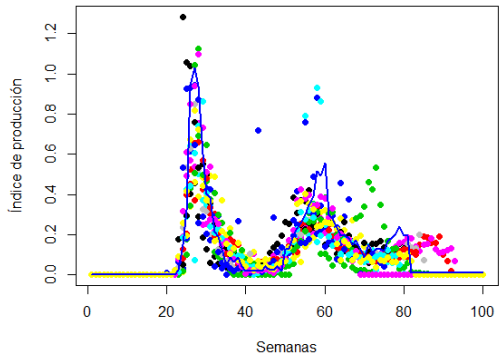


Figura 10.2: Bizet

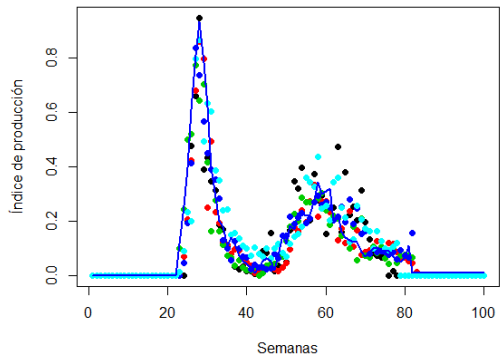


Figura 10.3: Cherrio

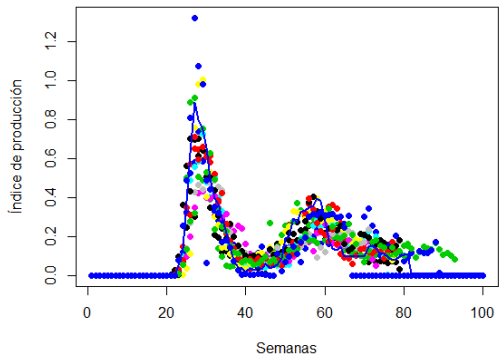


Figura 10.4: Don Pedro

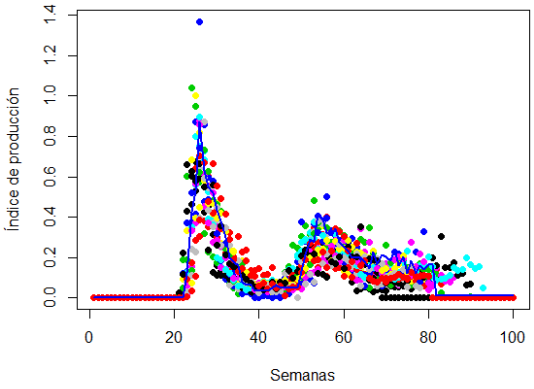


Figura 10.5: Farida

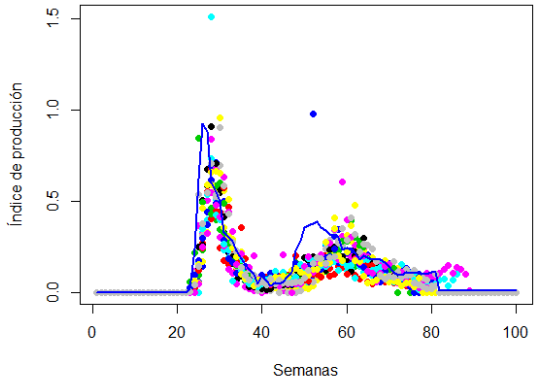


Figura 10.6: Hermes

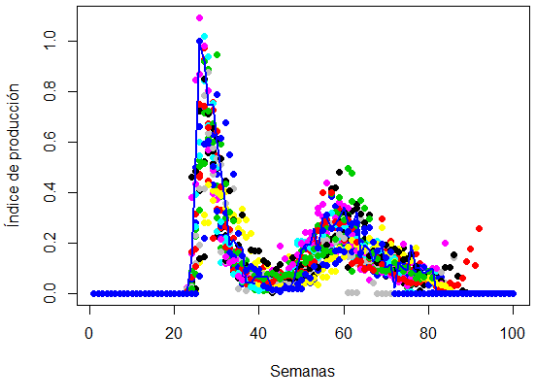


Figura 10.7: Hermes Orange

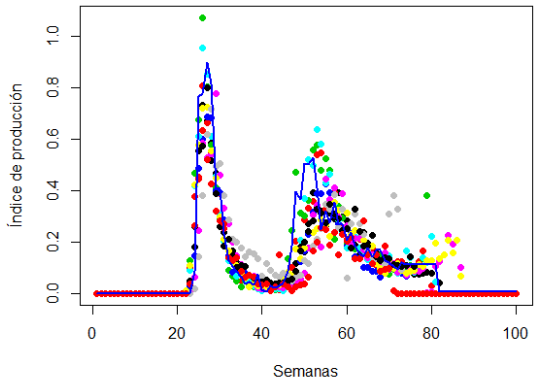


Figura 10.8: Kaori

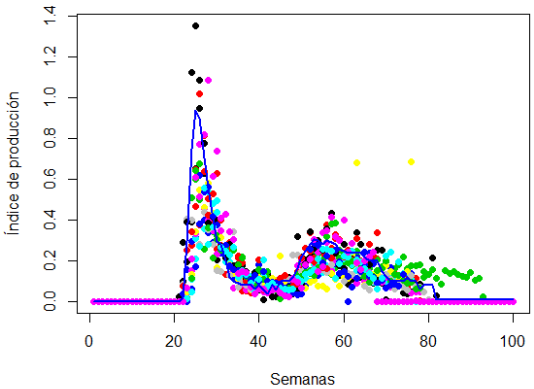


Figura 10.9: Kino

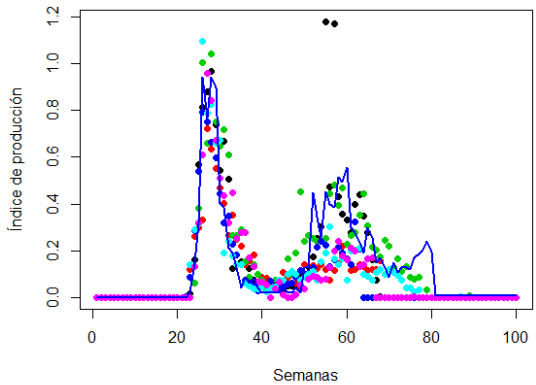


Figura 10.10: Kirk

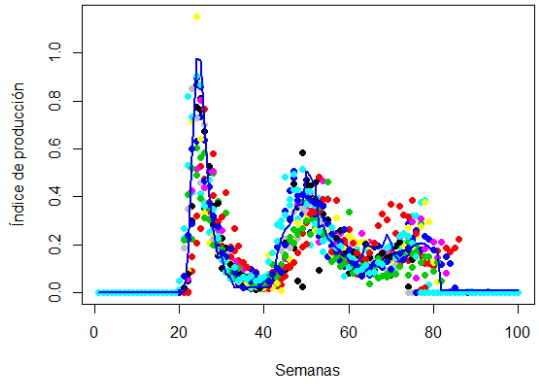


Figura 10.11: Komachi

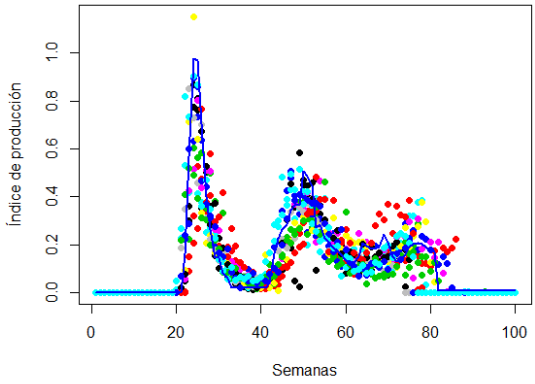


Figura 10.12: Lizzy

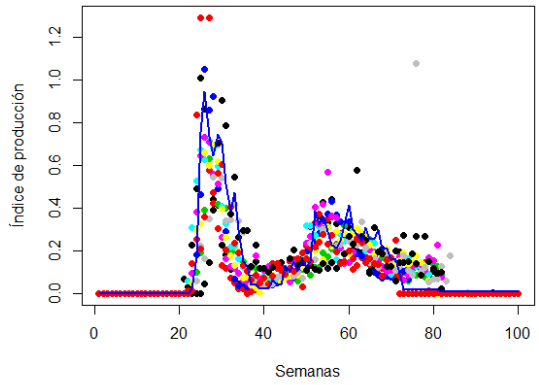


Figura 10.13: Mariposa

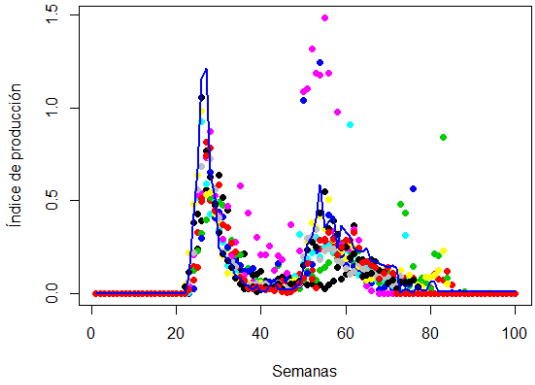


Figura 10.14: Monseñor

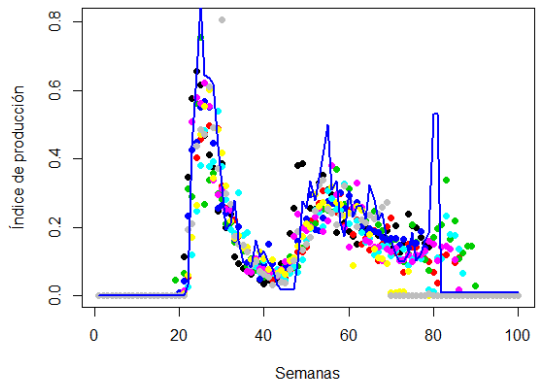


Figura 10.15: Moonlight

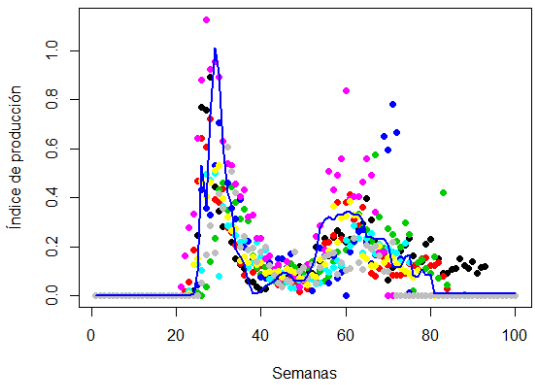


Figura 10.16: Nahema

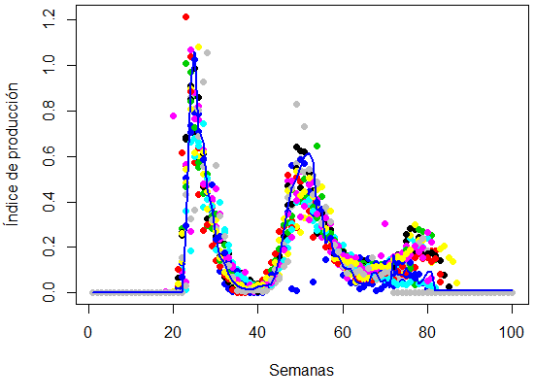


Figura 10.17: Pomodoro

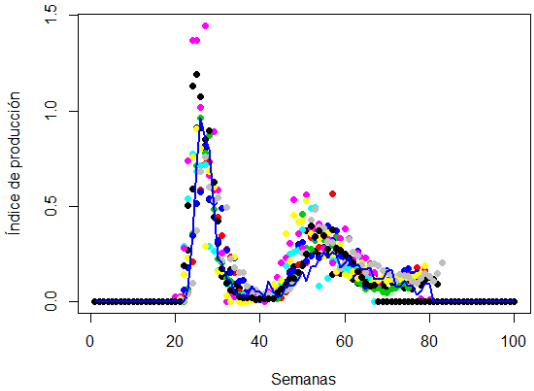


Figura 10.18: Red Magic

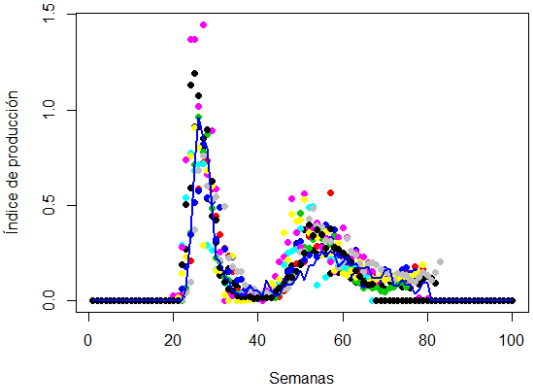


Figura 10.19: Voragine

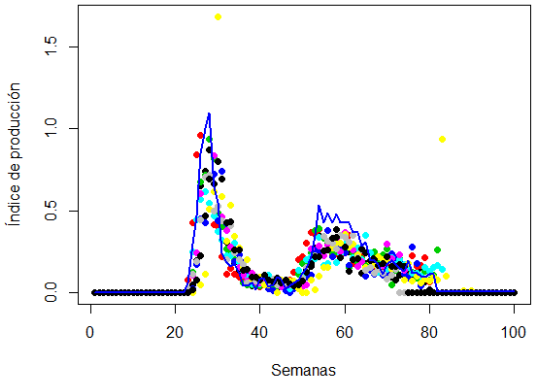


Figura 10.20: Zurigo

10.2. Diagrama de Cajas de las Variedades

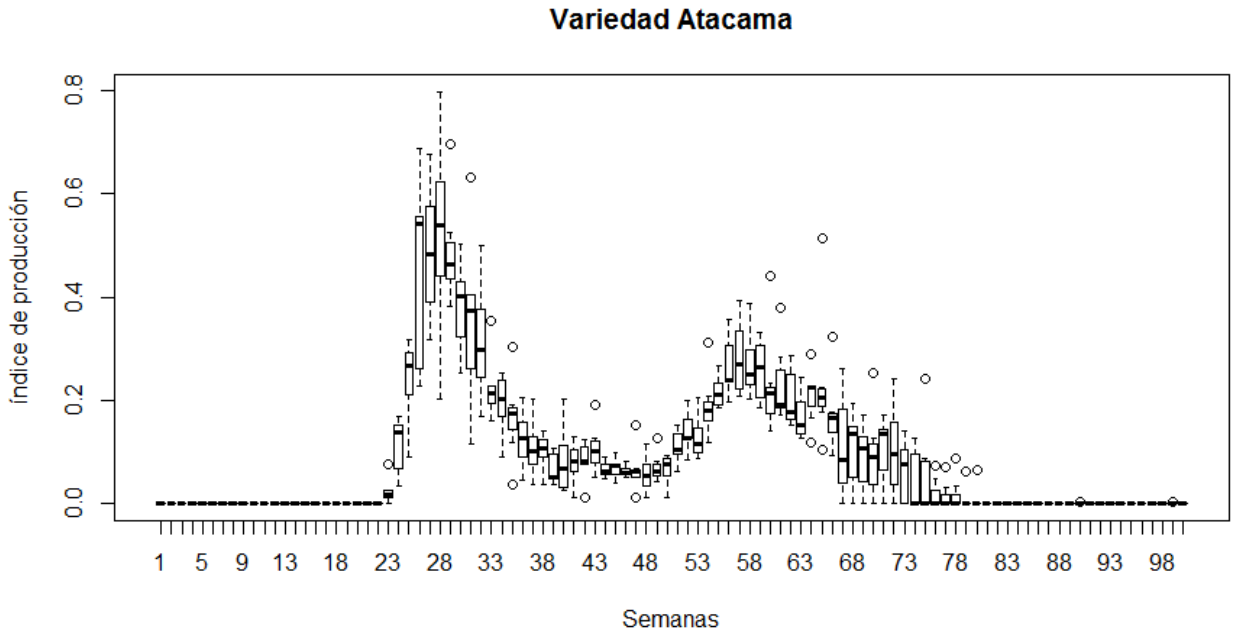


Figura 10.21: Boxplot variedad Atacama

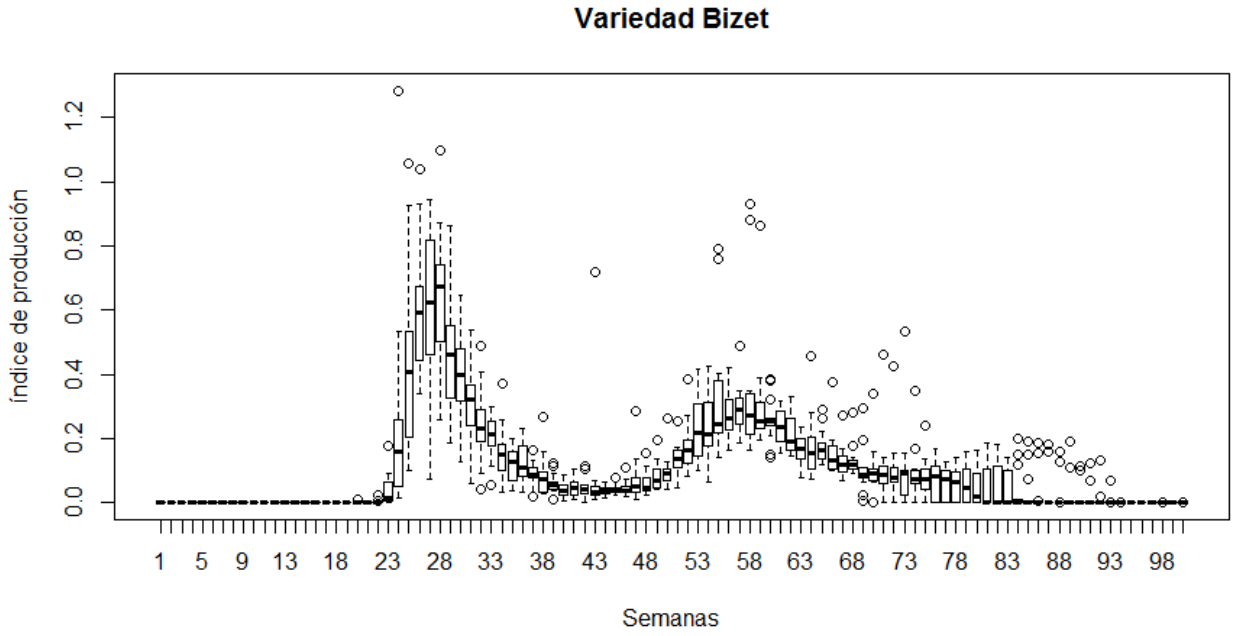


Figura 10.22: Boxplot variedad Bizet

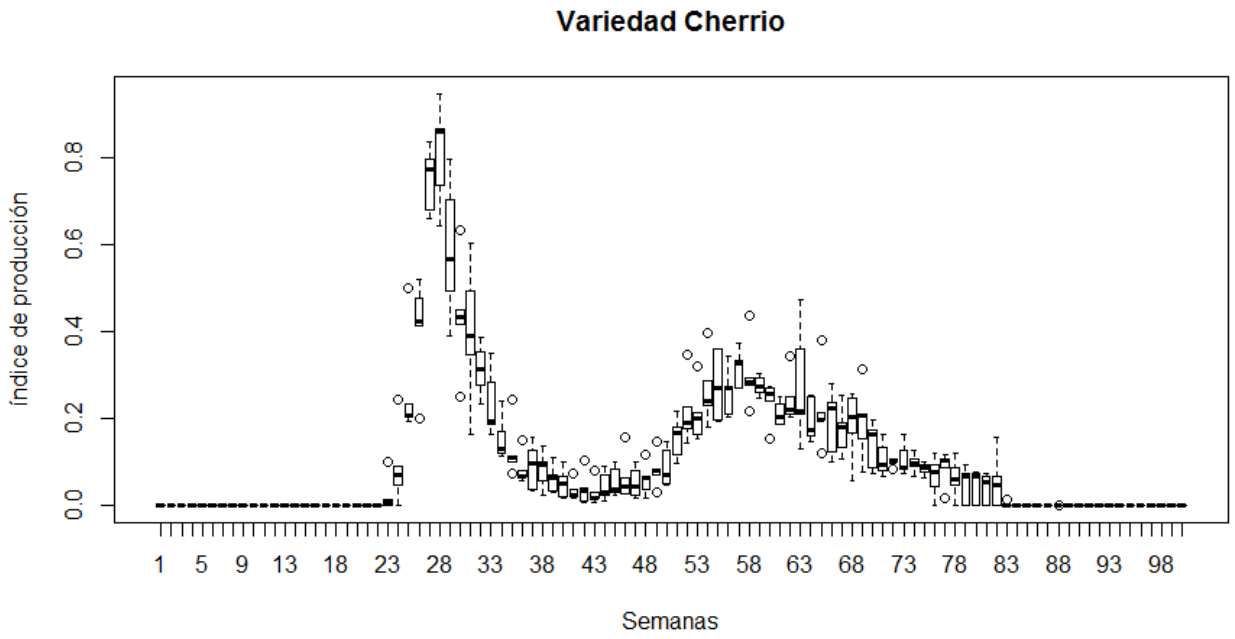


Figura 10.23: Boxplot variedad Cherrio

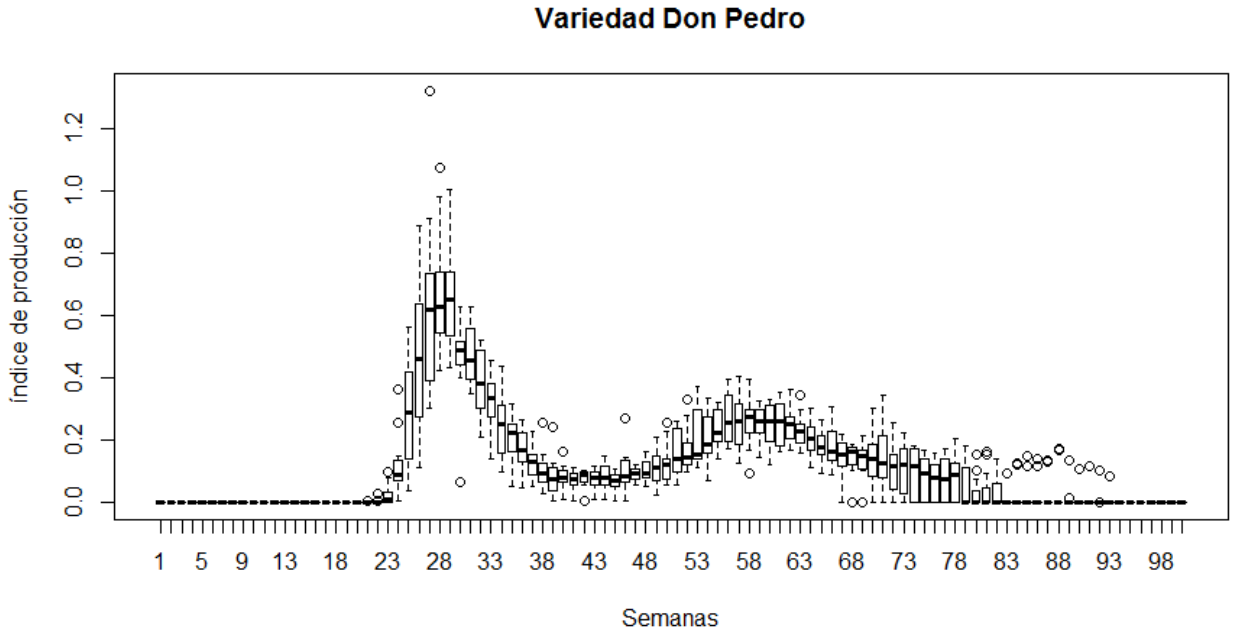


Figura 10.24: Boxplot variedad Don Pedro

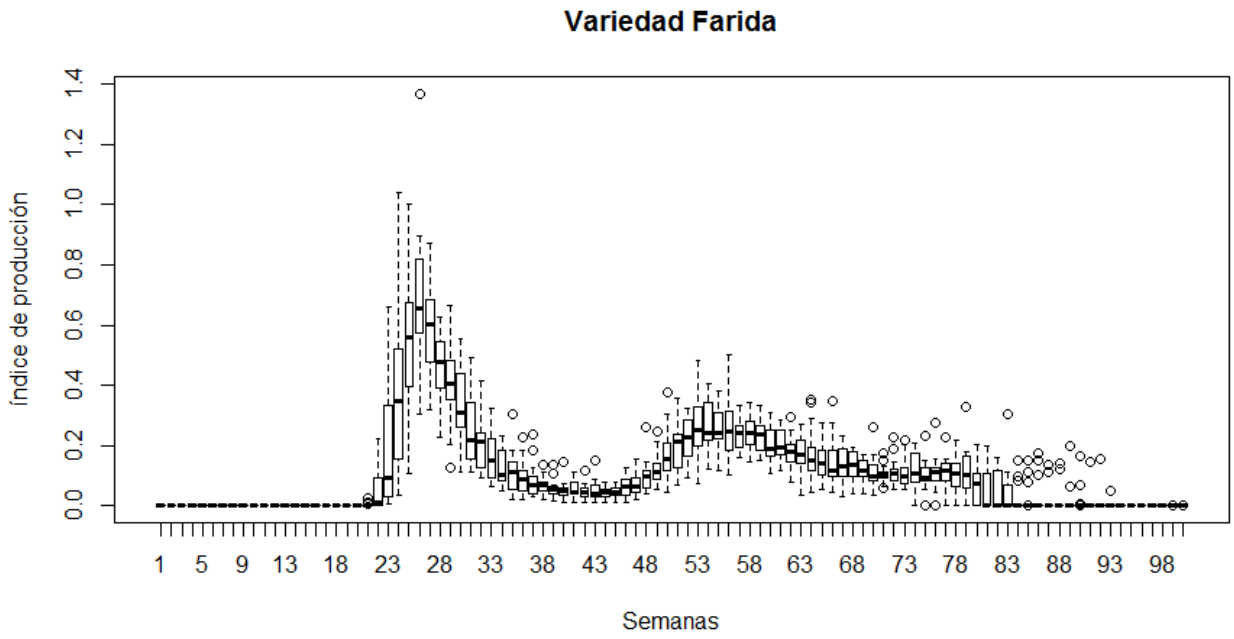


Figura 10.25: Boxplot variedad Farida

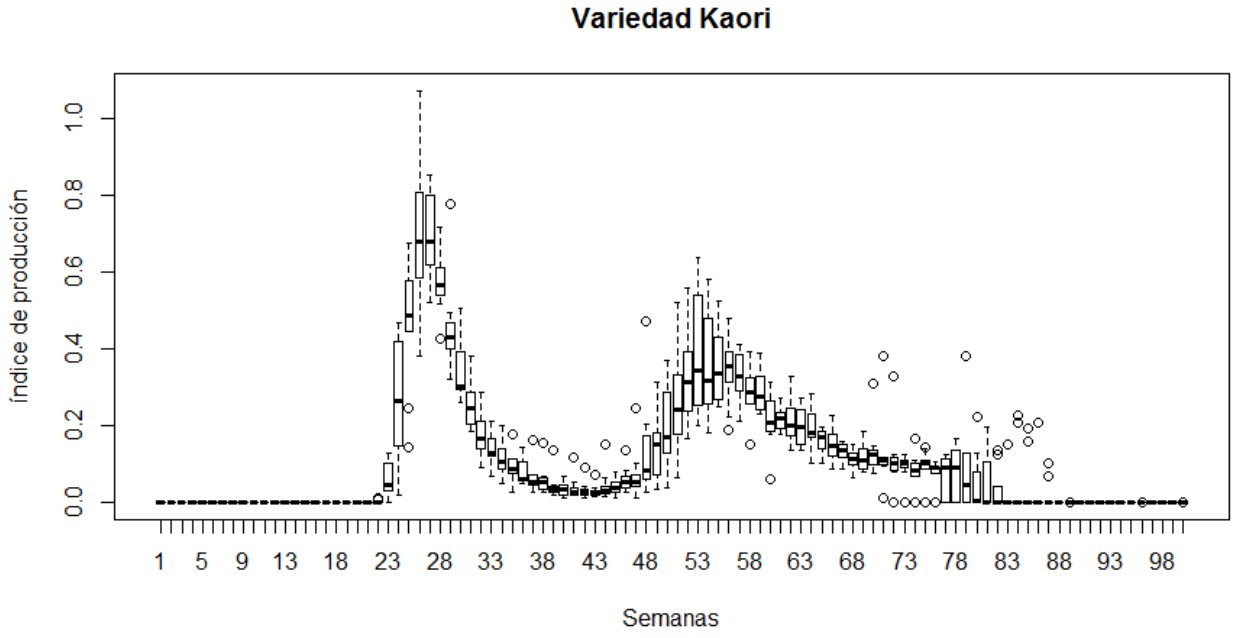


Figura 10.26: Boxplot variedad Kaori

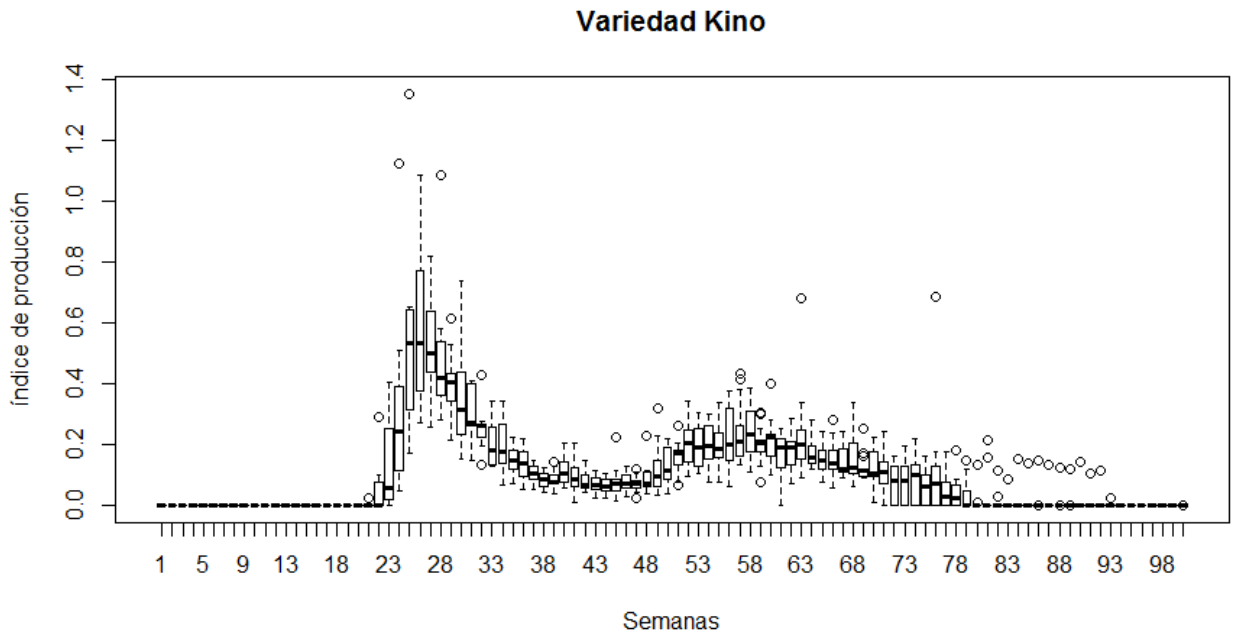


Figura 10.27: Boxplot variedad Kino

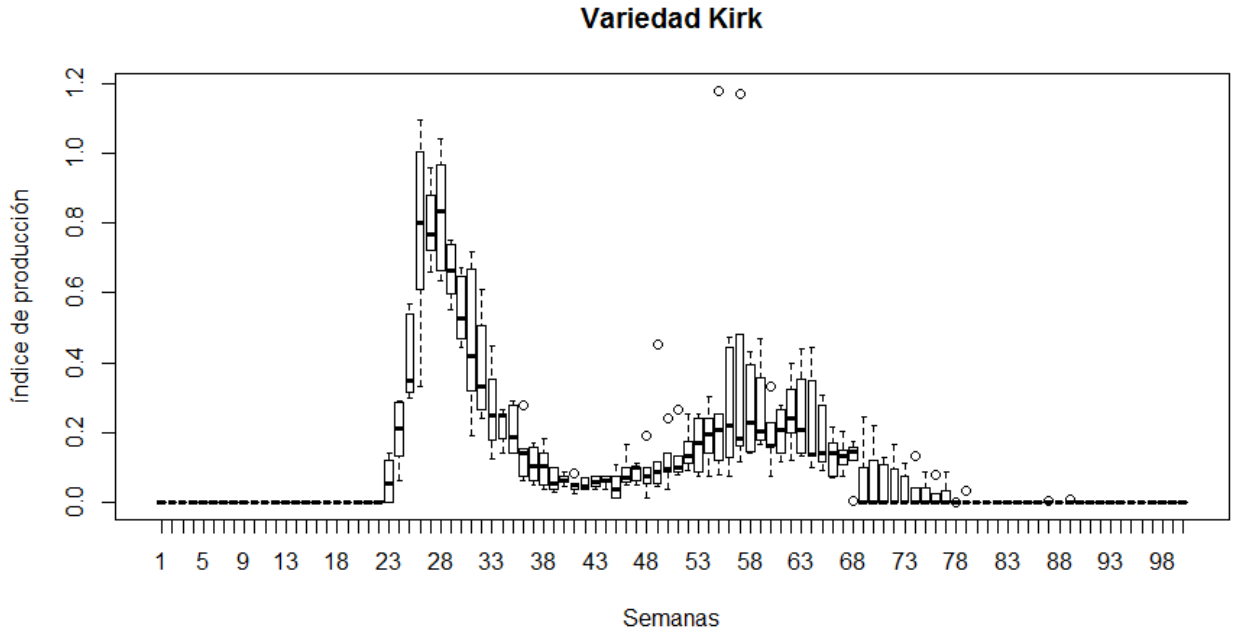


Figura 10.28: Boxplot variedad Kirk

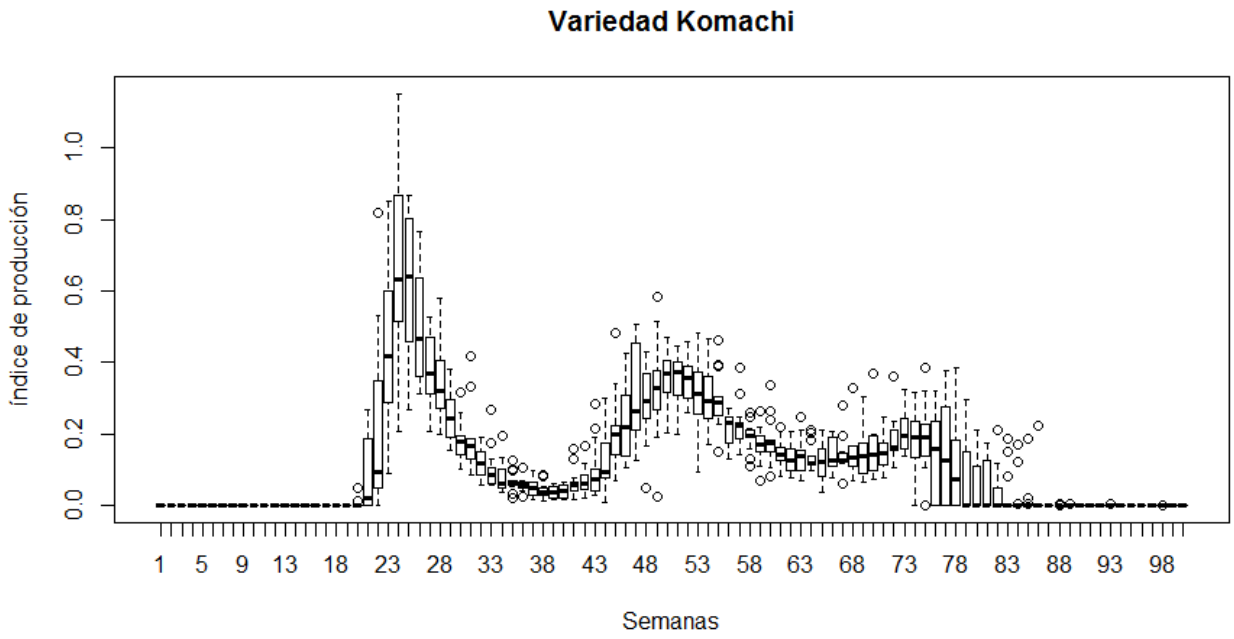


Figura 10.29: Boxplot variedad Komachi

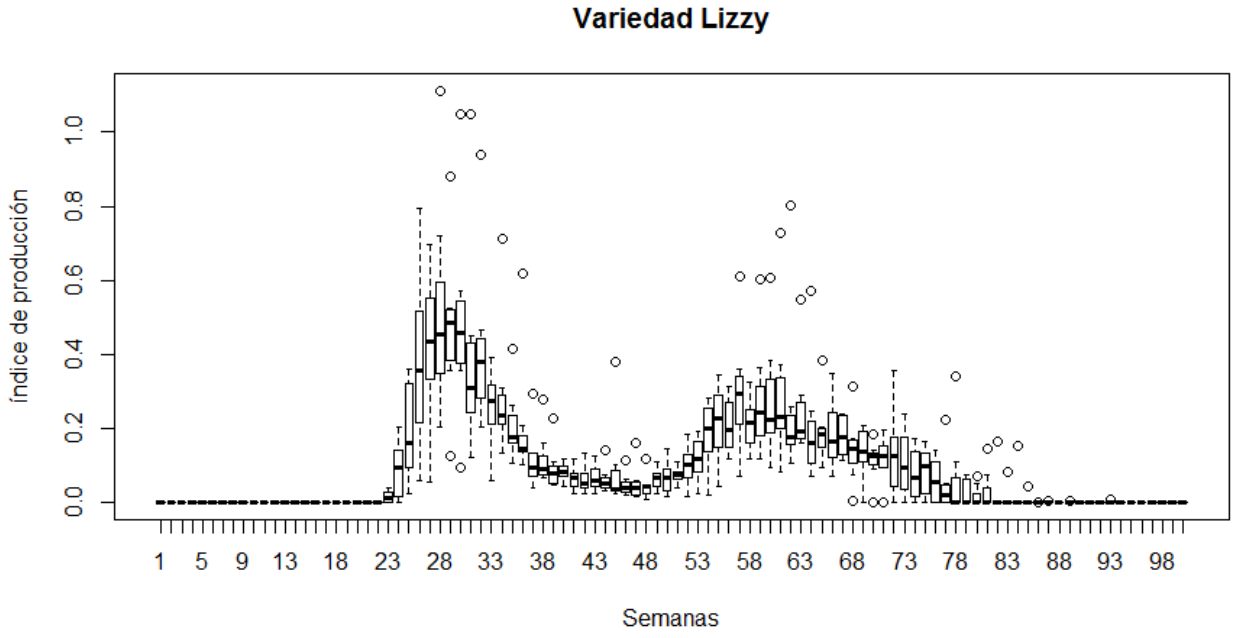


Figura 10.30: Boxplot variedad Lizzy

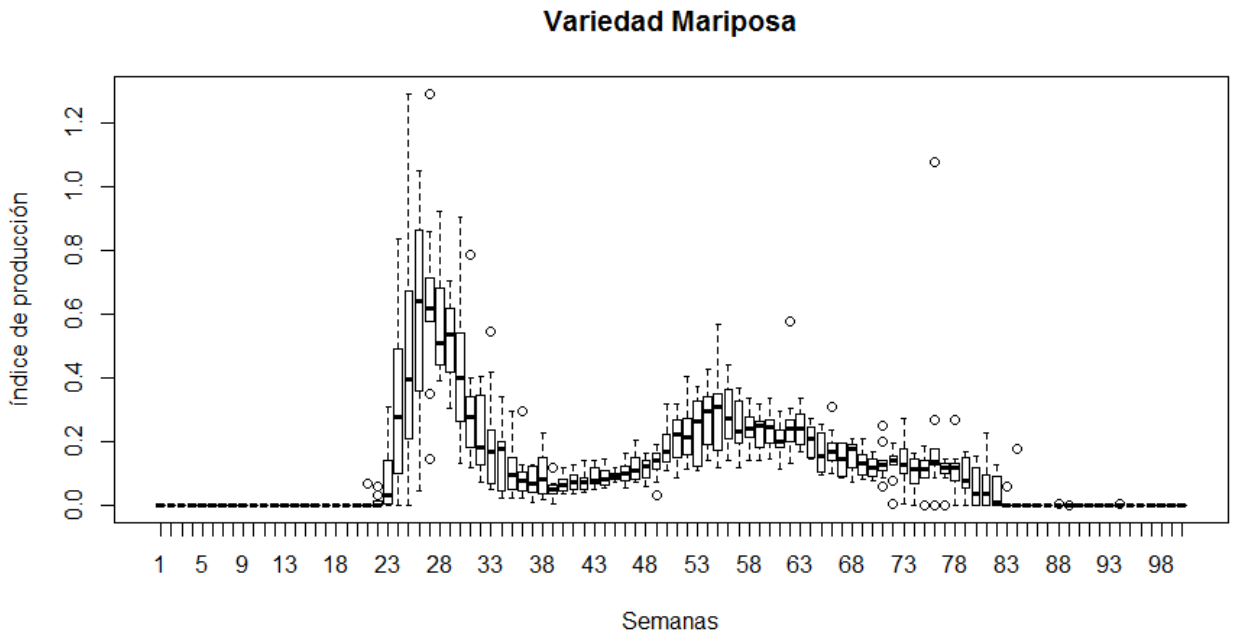


Figura 10.31: Boxplot variedad Mariposa

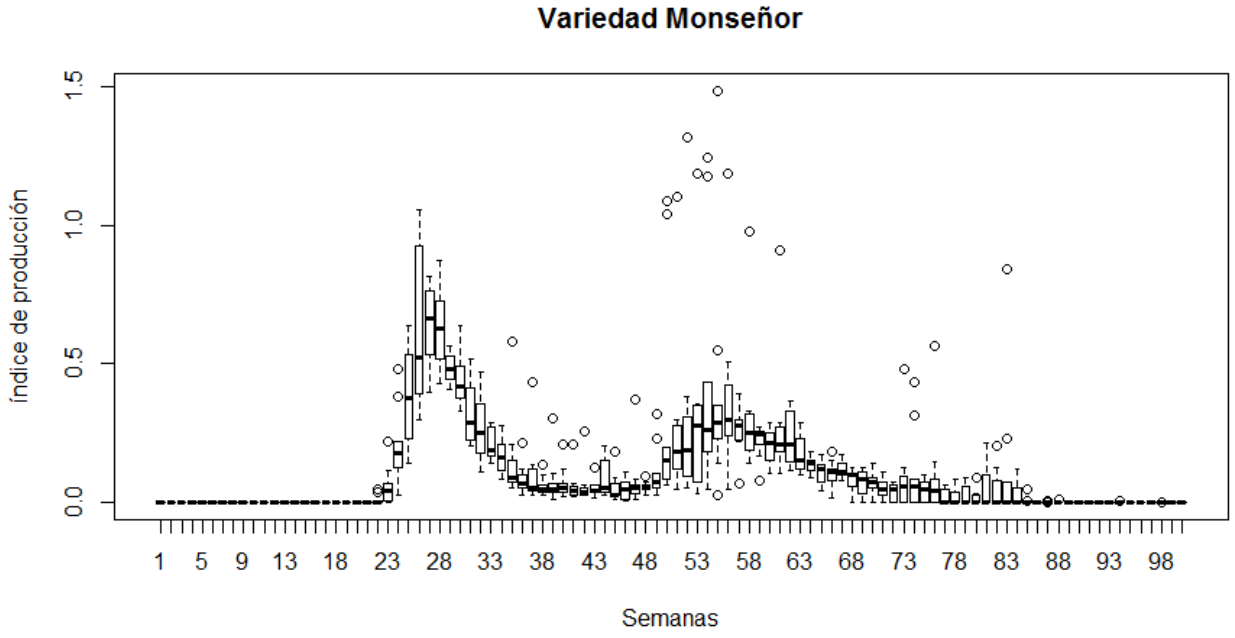


Figura 10.32: Boxplot variedad Monseñor

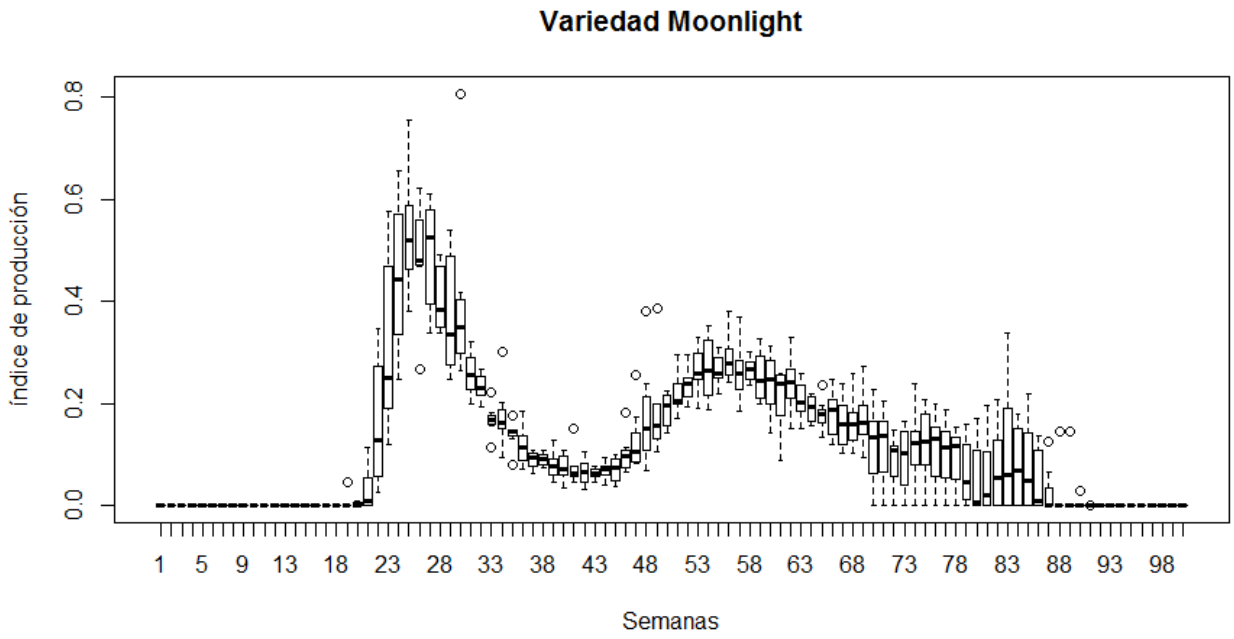


Figura 10.33: Boxplot variedad Moonlight

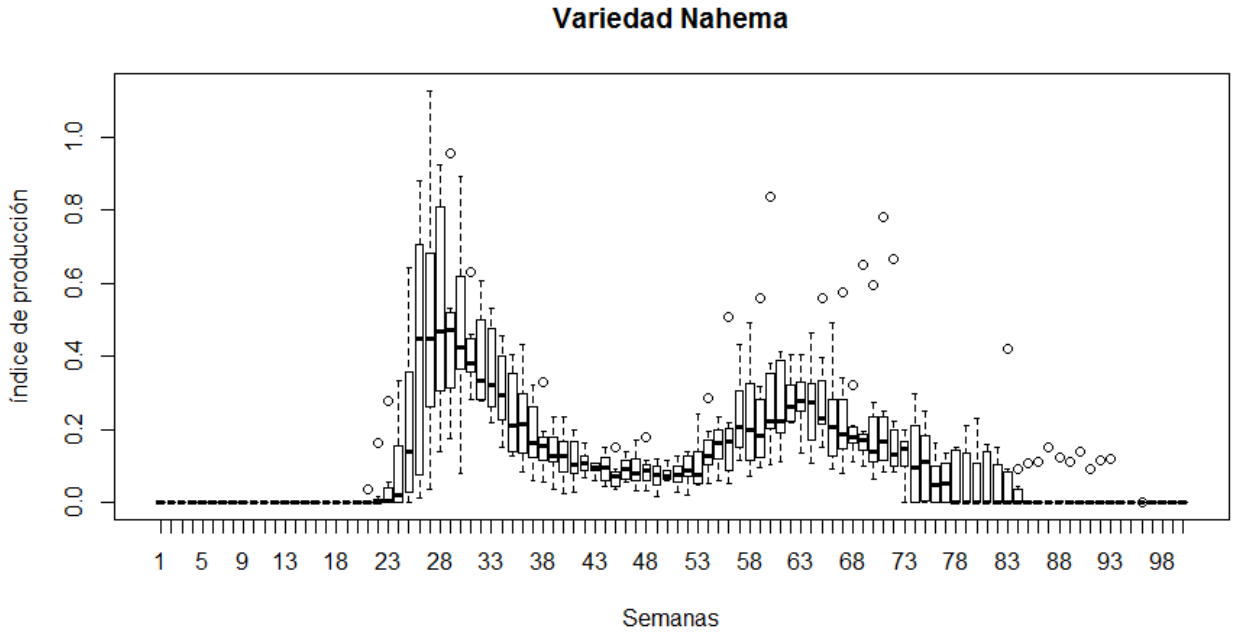


Figura 10.34: Boxplot variedad Nahema

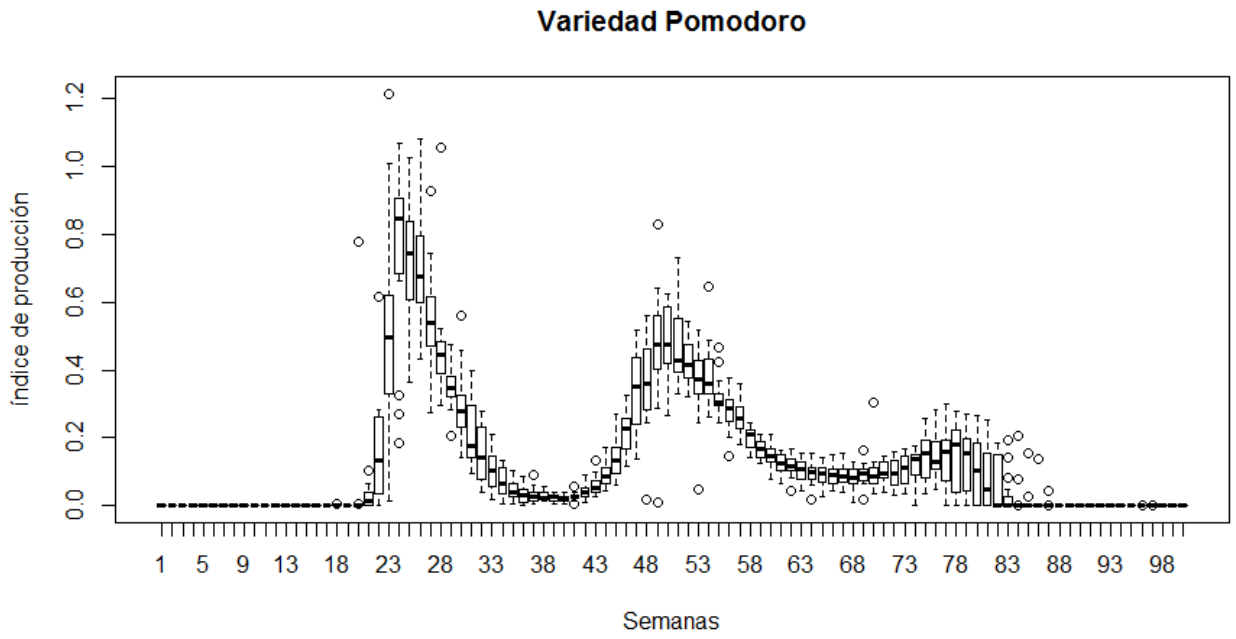


Figura 10.35: Boxplot variedad Pomodoro

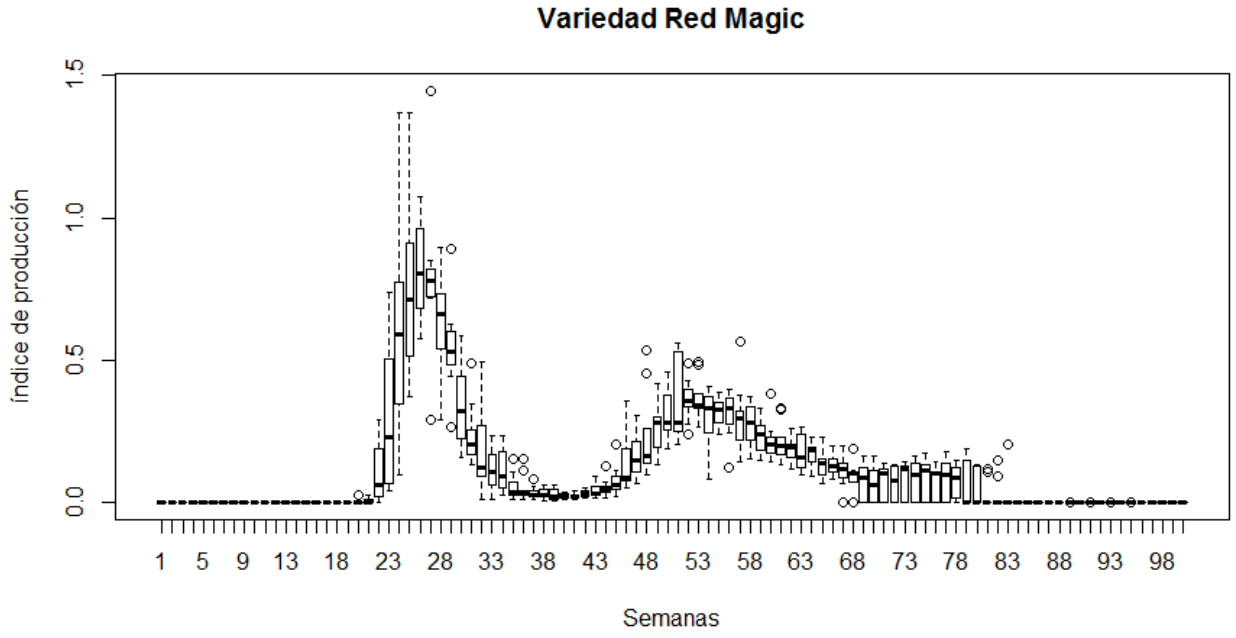


Figura 10.36: Boxplot variedad Red Magic

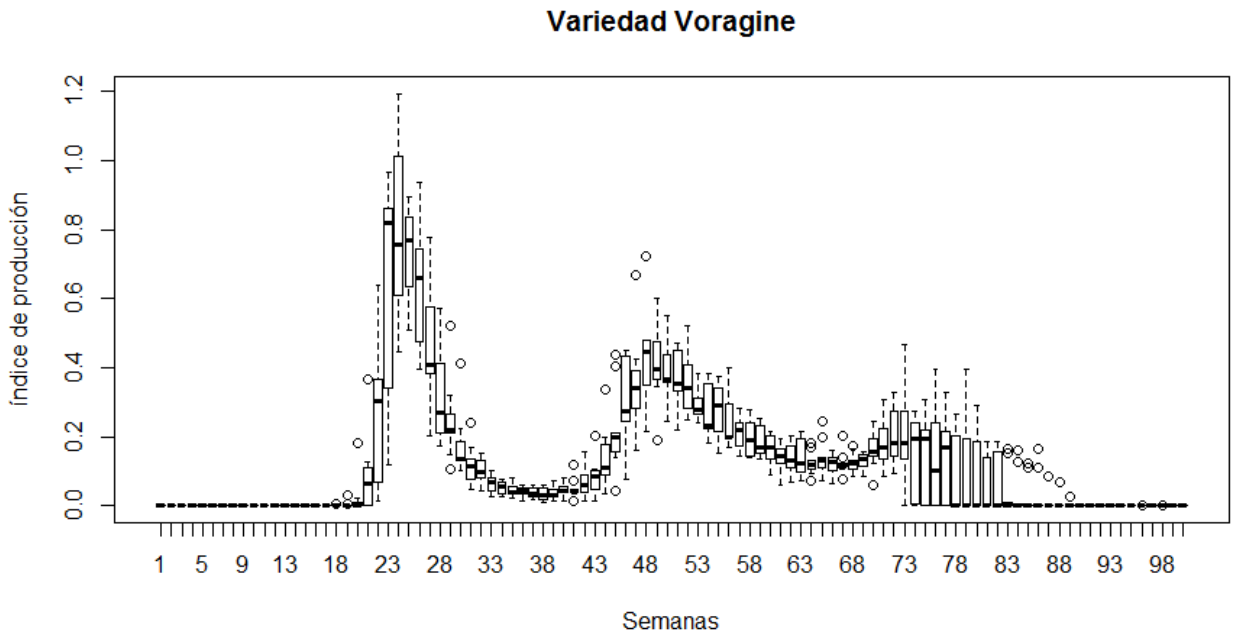


Figura 10.37: Boxplot variedad Voragine

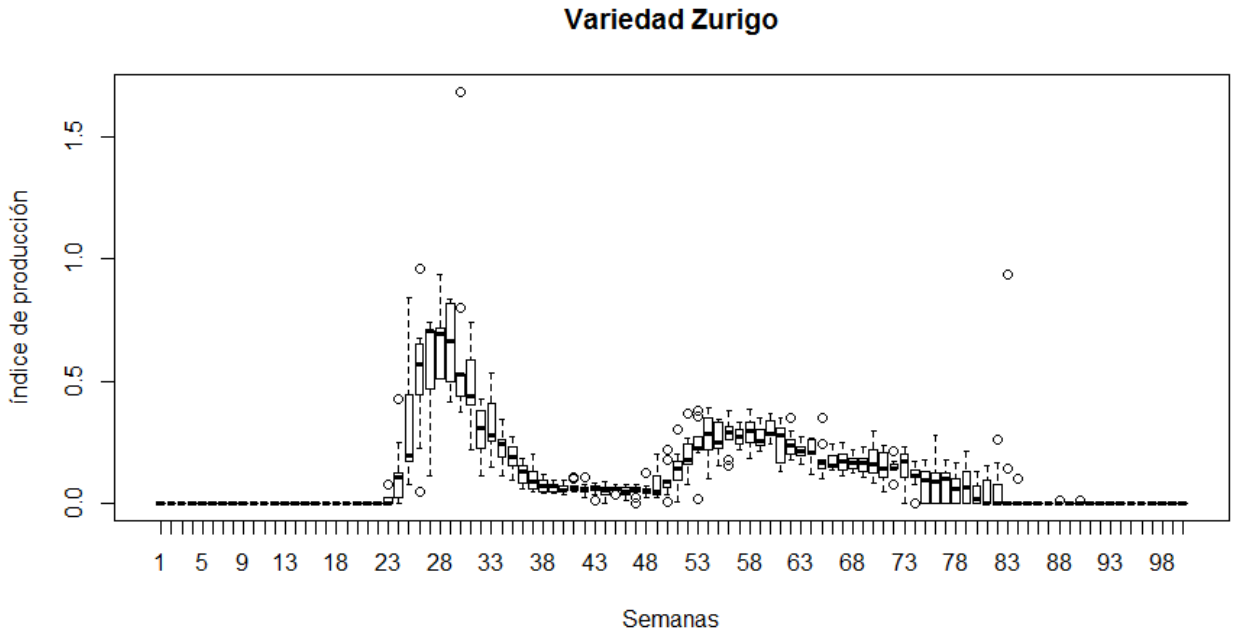


Figura 10.38: Boxplot variedad Zurigo