

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

APLICACIÓN DE ÍNDICES DE CAPACIDAD FUNCIONAL UNIVARIADOS PARA EVALUAR LA CAPACIDAD CLIMÁTICA PARA LA PRODUCCIÓN DE MAÍZ ANTE EL CAMBIO CLIMÁTICO POR MEDIO DE LAS VARIABLES TIEMPOS TÉRMICOS, RADIACIÓN SOLAR Y PRECIPITACIÓN

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO MATEMÁTICO

PROYECTO DE INVESTIGACIÓN

JUAN ANDRÉS JÁTIVA VIVANCO

juan.jativa@epn.edu.ec

DIRECTOR: PH.D. MIGUEL ALFONSO FLORES SÁNCHEZ

miguel.flores@epn.edu.ec

QUITO, SEPTIEMBRE 2022

DECLARACIÓN

Yo JUAN ANDRÉS JÁTIVA VIVANCO, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual, correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normatividad institucional vigente.

Juan Andrés Játiva Vivanco

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por JUAN ANDRÉS JÁTIVA VIVANCO, bajo mi supervisión.

Ph.D. Miguel Alfonso Flores Sánchez
Director del Proyecto

AGRADECIMIENTOS

A Dios y la Virgen por cuidarme, por darme las fuerzas para continuar y cumplir mis metas.

A mi madre Rita por su apoyo, su guía, su sacrificio y por su amor incondicional, por ayudarme a continuar en los momentos difíciles, todo se lo debo a ella.

A mi hermana Dayana por su apoyo, sus locuras y su compañía.

A Karen por su ayuda y consejos en los momentos difíciles.

A mi familia por sus consejos y su gran cariño.

A mis amigos les agradezco por todas las alegrías, los momentos compartidos y el apoyo brindado en esta etapa universitaria.

Finalmente agradezco a mi director de tesis Miguel Flores por su tiempo, su paciencia y su guía brindados para el desarrollo de este proyecto.

DEDICATORIA

A mi madre y hermana, por todo el cariño y apoyo.

Índice general

Índice de figuras	IX
Índice de tablas	X
Resumen	XI
Abstract	XII
1. Introducción	1
1.1. Introducción	1
1.2. Justificación	2
1.3. Objetivos	4
1.3.1. Objetivo General	4
1.3.2. Objetivos Específicos	4
2. Marco Teórico	5
2.1. Análisis de Datos Funcionales (FDA)	5
2.1.1. Definiciones	6
2.1.2. Métodos de suavizamiento	6
2.1.2.1. Bases de Fourier	8
2.1.2.2. Bases B-spline	8
2.1.2.3. Proceso de ajuste	10
2.1.2.4. Suavizamiento mediante penalización	11
2.1.2.5. Criterio de validación cruzada generalizada GCV	12
2.1.3. Profundidad de datos	13

2.1.3.1.	Funciones de profundidad funcionales	14
2.1.3.2.	P-regiones centrales funcionales	17
2.2.	Índices de capacidad de proceso	21
2.2.1.	Índices de capacidad de proceso univariados tradicionales	22
2.2.1.1.	Índice de capacidad C_p	22
2.2.1.2.	Índices de capacidad C_{pu} y C_{pl}	24
2.2.1.3.	Índice de capacidad C_{pk}	25
2.2.2.	Índices de capacidad de proceso funcionales	26
3.	Aplicación de índices de capacidad de proceso funcionales y Resultados	28
3.1.	Datos y descripción de las variables climáticas	28
3.1.1.	Datos	29
3.1.2.	Variables climáticas	29
3.2.	Límites de especificación	30
3.2.1.	Límites de especificación para la precipitación	30
3.2.2.	Límites de especificación para los tiempos térmicos	31
3.2.3.	Límites de especificación para la radiación solar fotosintéticamente activa	32
3.3.	Curvas $\chi_{0,99865}$, $\chi_{0,00135}$, $\chi_{0,5}$	33
3.3.1.	Suavizamiento de datos	33
3.3.1.1.	Suavizamiento de la precipitación	33
3.3.1.2.	Suavizamiento de los tiempos térmicos	34
3.3.1.3.	Suavizamiento de la radiación solar fotosintéticamente activa PAR	35
3.3.2.	Cálculo de la región central	35
3.3.3.	Curva más profunda	36
3.4.	Aplicación y Resultados	37
3.4.1.	Precipitación	37
3.4.2.	Tiempos Térmicos	39
3.4.3.	Radiación solar fotosintéticamente activa - PAR	40

4. Conclusiones y Recomendaciones	43
Bibliografía	47
Anexos	52
A. Tablas de valores GCV para los pares (K, λ)	53
B. Valores de índices para escenarios considerados	57
B.1. Escenario 1: Aplicación sin suavizamiento de datos	57
B.2. Escenario 2: Suavizamiento de las curvas $\mathcal{X}_{0,99865}, \mathcal{X}_{0,00135}$	58

Índice de figuras

2.1. Splines de diferentes ordenes para ajustar la función seno	9
2.2. Banda definida por tres curvas.	16
2.3. Región central estimada $C_{n,p}$	20
2.4. Región central con $p = 0.1$	21
2.5. Región de especificación y región de proceso	23
3.1. Datos discretos y Datos funcionales para la precipitación	34
3.2. Curvas originales y ajustadas para los tiempos térmicos	34
3.3. Curvas originales y ajustadas para la PAR	35
3.4. Región Central para Precipitación, Tiempos térmicos y PAR	36
3.5. Curva media para Precipitación, Tiempos térmicos y PAR	37
3.6. Región de especificación y región del proceso para la precipitación .	38
3.7. Región de especificación y región del proceso para los tiempos térmicos	39
3.8. Región de especificación y región de proceso para la PAR y cada tipo de maíz	41

Índice de tablas

2.1. Niveles del proceso de acuerdo a los valores de C_p	24
3.1. Límites de especificación para la precipitación	31
3.2. Límites de especificación para los tiempos térmicos	32
3.3. Límites de especificación para la PAR	32
3.4. Valores de los índices $C_{p(prf)}$ y $C_{pk(prf)}$ para los métodos A y B para la precipitación	38
3.5. Valores de los índices $C_{p(prf)}$ y $C_{pk(prf)}$ para los métodos A y B para los tiempos térmicos	40
3.6. Valores de los índices $C_{p(prf)}$ y $C_{pk(prf)}$ para los métodos A y B para la PAR	42
A.1. Valores GCV para la precipitación	54
A.2. Valores GCV para los tiempos térmicos	55
A.3. Valores GCV para la PAR	56
B.1. Valores índices de capacidad proceso para métodos A y B - precipitación	57
B.2. Valores índices de capacidad proceso para métodos A y B - tiempos térmicos	57
B.3. Valores índices de capacidad proceso para métodos A y B - PAR . . .	58
B.4. Valores índices de capacidad proceso para métodos A y B - precipitación	58
B.5. Valores índices de capacidad proceso para métodos A y B - tiempos térmicos	58
B.6. Valores índices de capacidad proceso para métodos A y B - PAR . . .	58

Resumen

En el presente trabajo, se propone la aplicación de índices de capacidad de proceso funcionales univariados para evaluar la capacidad climática en la producción de maíz en La Concordia. Se consideran tres híbridos de maíz y al clima como un proceso, el cual se asume se encuentra en control. Estos índices se aplican mediante dos métodos (A y B) a variables climáticas importantes para el desarrollo del maíz como: tiempos térmicos, radiación solar fotosintéticamente activa y precipitación; las cuales se consideran como características a controlar y son tomadas semanalmente. Los límites de especificación son calculados en función de las necesidades del maíz respecto a las variables consideradas.

Previo a esta aplicación, se realiza una revisión sobre distintas técnicas del FDA como son: el ajuste de datos discretos mediante funciones suaves, medidas de profundidad funcionales y p-región central funcional; para obtener las funciones requeridas por los índices.

Finalmente, mediante esta aplicación se pudo evidenciar que el método A entrega mejores resultados que el método B. Por otro lado, se obtuvo que no existe la capacidad de precipitación para la producción de los híbridos de maíz. Se evidenció que existe la capacidad de tiempos térmicos para los tres híbridos de maíz. En el caso de la radiación fotosintéticamente activa, se observó que mediante el índice $C_{p(prf)}$, se tiene la capacidad para dos híbridos de maíz de los tres considerados; sin embargo, al considerar el índice $C_{pk(prf)}$, no se tiene la capacidad de suplir las necesidades de esta variable para la producción de los híbridos de maíz considerados.

Palabras clave: FDA, índices de capacidad de proceso, p-región central funcional, híbridos de maíz.

Abstract

In the present work, the application of univariate functional process capability indices is proposed to evaluate the climatic capability in corn production at La Concordia. Three corn hybrids are studied, and the weather is considered as a process, which is assumed to be in control. These indexes are applied by two methods (A and B) to relevant climatic variables for maize development such as: thermal times, photosynthetically active solar radiation, and precipitation, which are considered as characteristics to be controlled and are measured on a weekly basis. The specification limits are calculated according to the needs of the corn in relation to the variables.

Prior to this application, it was done a review of different FDA techniques, such as: adjustment of discrete data by smooth functions, functional depth measures and functional central p-region; in order to obtain the functions required by the indexes.

Finally, through this application, it could be proved that method A yields better results than method B. On the other hand, it was determined that there is no precipitation capacity for the production of corn hybrids. It was evidenced that the capacity of thermal times exists for the three corn hybrids. In the case of photosynthetically active radiation, it was observed that through the index $C_{p(prf)}$, there is capacity for two corn hybrids of the three considered; nevertheless, when considering the index $C_{pk(prf)}$, there is no capacity to supply the needs of this variable for the production of the corn hybrids considered.

Key words. FDA, process capability indices, functional central p-region, corn hybrids.

Capítulo 1

Introducción

1.1. Introducción

La temperatura del planeta es afectada por el incremento de las emisiones de gases que ocasiona el derretimiento de glaciares, el aumento de las precipitaciones, la frecuencia de eventos meteorológicos extremos y modificaciones en las estaciones del clima. Al ser la agricultura extremadamente vulnerable al cambio climático, se ha podido evidenciar impactos directos que repercuten en el rendimiento de los cultivos y en los ciclos de crecimiento de las especies agrícolas. Por ejemplo, la variación en la temperatura y radiación solar ocasionan una mayor presencia de: insectos y plagas; los cuales son perjudiciales para los cultivos, afectando su desarrollo. Por otro lado, la variación de pluviosidad ocasiona incrementos o decrementos en los volúmenes de la precipitación en ciertos períodos del año, provocando inundaciones o sequías; aumentando la probabilidad de fracaso de las cosechas a corto plazo y la producción en el largo plazo [41]. Por estas variaciones, el sector agrícola ha tomado medidas que permiten adaptarse a estos cambios, por ejemplo, el desarrollo de cultivos más tolerantes a sequías, a plagas, etc., y métodos de producción más eficaces y sostenibles [33].

El estudio del análisis de la capacidad del proceso tuvo inicio como una comparación entre la distribución estable del proceso y las tolerancias fijadas para el producto. Este estudio estuvo basado inicialmente en el histograma de la distribución del proceso y los límites de especificación de las características de calidad; sin embargo, las herramientas gráficas tenían muchas limitaciones al momento de evaluar la capacidad del proceso. Hubo una alta demanda de resumir esta capacidad, con un número; lo que dio paso a los índices de capacidad de proceso [28]. En la

actualidad, existen diversos tipos de índices que ayudan a medir la capacidad del proceso, los cuales pueden ser: univariantes, si se está evaluando una sola característica o multivariantes, cuando se estudia en conjunto varias características del producto; siendo estos últimos como los más utilizados.

El análisis de datos funcionales (FDA por sus siglas en inglés) es una herramienta estadística novedosa y que posee un gran alcance. Tiene como finalidad simplificar los estudios de variables que dependen de parámetros continuos, puesto que, los datos son representados generalmente a través de funciones. Esta técnica permite solventar problemas de alta dimensionalidad; ya que, los datos que se observan inicialmente poseen una naturaleza discreta. Mediante el uso de técnicas de suavizamiento, se puede explorar de manera fluida los procesos descritos por estos datos; es decir, esta técnica permite obtener información adicional que se encuentra contenida en la función y sus derivadas [19].

En el presente trabajo, se plantea evaluar la capacidad climática para la producción de tres tipos de híbrido de maíz duro seco mediante la aplicación de índices de capacidad de proceso funcionales univariados considerando tres variables: precipitación, tiempos térmicos y radiación solar. Estas variables climáticas afectan directamente en el crecimiento del maíz duro seco; y por lo tanto, en su rendimiento.

1.2. Justificación

Los modelos climáticos prevén cambios drásticos en muchas regiones del mundo, incluyendo variación de temperatura, precipitación, incremento en la frecuencia y severidad de eventos extremos como sequías e inundaciones.

La agricultura en particular se enfrenta a grandes retos en el siglo XXI. Por un lado, se espera que a mitad del siglo la demanda de alimentos y fibras a nivel mundial aumente aproximadamente un 70 %. Se estima que los rendimientos de los granos básicos como: el maíz duro seco, el arroz y el trigo; disminuyan significativamente a nivel mundial. Además, la agricultura tendrá que competir por el uso de la tierra y el agua; debido al aumento de asentamientos urbanos. Así mismo, tendrá que hacer frente al cambio climático. Tendrá que adaptarse y contribuir simultáneamente a la mitigación del mismo. Finalmente, tendrá el reto de ayudar a la preservación de hábitats naturales y la conservación de su biodiversidad [25].

Por esta razón, el sector agropecuario necesitará nuevas tecnologías que contribuyan a la producción y ayuden al uso reducido de tierras y mano de obra [57].

Los índices de capacidad miden la relación que existe entre la dispersión del proceso y la de especificación. Estos índices son valores que no poseen unidades, y se utilizan para comparar la capacidad de distintos procesos. Por esta razón, resulta interesante considerar al clima como un proceso. Mediante sensores y estaciones meteorológicas se puede obtener una gran cantidad de registros dependiendo del período en el que se lo mida. Con estos datos, se pueden aplicar índices para evaluar la capacidad climática, los cuales a su vez, pueden ser una de herramienta para que el agricultor pueda ajustar a tiempo las prácticas de cultivo, protegiéndose de condiciones climáticas adversas o aprovechando condiciones ventajosas. Además, mediante el análisis de datos funcionales (FDA), se puede reducir la alta dimensionalidad de estos datos, y utilizar la información adicional contenidas en las curvas, obteniendo así un mejor entendimiento de esta variabilidad.

De acuerdo con el Banco Central del Ecuador (BCE) en 2019, en el Ecuador, el sector ganadero, agrícola, silvicultura y pesquero aportó el 7,7 % del producto interno bruto (PIB), convirtiéndose en el cuarto sector económico de mayor relevancia en el país. El sector agrícola, por si sólo, representa el 0,1 %. Por otro lado, entre los años 2014 y 2018, el 11,6 % del uso del suelo ecuatoriano lo ocuparon cultivos permanentes (caña de azúcar, banano, palma africana), mientras el 15,76 % lo ocuparon cultivos transitorios (arroz, maíz duro seco, papa). El maíz duro seco fue el segundo cultivo transitorio con mayor producción, teniendo una representación del 32,68 %, solo por detrás del arroz con un 36,87 % [24].

Por esta razón, resulta importante equipar al sector agrícola con herramientas técnicas que permitan hacer frente a los efectos del cambio climático y de esa manera, maximizar la producción de los cultivos.

1.3. Objetivos

1.3.1. Objetivo General

Evaluar la capacidad climática para producción de maíz en La Concordia, mediante las variables de tiempos térmicos, radiación solar y precipitación.

1.3.2. Objetivos Específicos

- Construir una base de datos de variables climáticas (tiempos térmicos, radiación solar, precipitación) de La Concordia.
- Analizar series climáticas utilizando técnicas descriptivas de FDA.
- Aplicar índices de capacidad funcional univariados considerando límites de tolerancia para las variables consideradas.
- Determinar la capacidad climática a partir de los valores obtenidos.

Capítulo 2

Marco Teórico

En este capítulo se describe los fundamentos que se utilizaron en la elaboración de los índices de capacidad funcional definidos por Guevara y Vargas [22]. Para entender estos índices inicialmente se realizará una revisión del análisis de datos funcionales (FDA) y de las distintas técnicas usadas como: curva más profunda, suavizamiento, región central, entre otros. Adicionalmente, se presenta los conceptos del control estadístico de la calidad, en especial aquellos relacionados a índices de capacidad univariados tradicionales y la variación propuesta por Clements [8], basada en el uso de percentiles no normales para cualquier tipo de distribución.

2.1. Análisis de Datos Funcionales (FDA)

El análisis de datos funcionales es un área de investigación estadística en la cual los datos utilizados son funciones. Actualmente, es considerado un campo activo, tanto en la parte práctica como teórica. En efecto, con el avance tecnológico de los últimos años se ha obtenido un gran desarrollo sobre nuevos procedimientos de medición, los cuales son más rápidos y precisos, permitiendo obtener una mayor cantidad de datos que pueden ser tratados como funciones [16].

Los datos funcionales son de gran importancia, cada vez con una mayor frecuencia, se convierten en la forma natural de representar los resultados observados en distintas áreas del conocimiento. Por ejemplo, en economía, es posible considerar curvas de cotizaciones de acciones intradía; en estudios ambientales, se pueden encontrar mediciones continuas de las redes de vigilancia atmosférica; y también en el reconocimiento de imágenes o en información espacio-temporal. Otras áreas donde también resulta útil el análisis de datos funcionales son: biología, física, medicina e

ingeniería [18, 34].

En la práctica, inicialmente se trabaja con observaciones discretas de las curvas muestrales, a partir de las cuales, se reconstruyen la forma funcional. Esto permite mantener las propiedades de las variables continuas y por tanto, se pierde menos información. Entre las ventajas del FDA sobre técnicas como series de tiempo se tienen que: no es necesario contar con observaciones equiespaciadas o eliminar la tendencia de la serie.

2.1.1. Definiciones

Con la finalidad de desarrollar la metodología usada, es necesario revisar algunas definiciones importantes, las cuales son tomadas de Ferraty y Vieu [19].

Definición 1. *Una variable aleatoria funcional X , es una variable aleatoria que toma valores en un espacio infinito dimensional o espacio funcional. Una observación \mathcal{X} de X se conoce como dato funcional.*

Definición 2. *Un conjunto de datos funcionales $\{\mathcal{X}_1 \dots \mathcal{X}_n\}$ es una muestra tomada de una variable funcional X .*

Definición 3. *Sea $L^2(T)$ el espacio de Hilbert separable definido por las funciones cuadrado integrable en el intervalo $T = [a, b] \subset \mathbb{R}$, notado por:*

$$L^2(T) = \left\{ \mathcal{X} : T \rightarrow \mathbb{R}; \int_T \mathcal{X}(t)^2 dt < \infty \right\}, \quad (2.1)$$

y cuyo producto interno se define por:

$$\langle \mathcal{X}(t), \mathcal{Y}(t) \rangle = \int_T \mathcal{X}(t)\mathcal{Y}(t) dt.$$

El producto interno definido, es equivalente al producto interno de vectores en \mathbb{R}^n ; de esta manera el FDA utiliza funciones definidas en $L^2(T)$, ya que se puede extender técnicas del análisis multivariado a datos funcionales.

2.1.2. Métodos de suavizamiento

Como se menciona en [46], cuando se está estudiando algún fenómeno, se vuelve imposible observar un conjunto de n funciones de manera continua a lo largo del tiempo. Una posible solución es tomar mediciones de dichas funciones en deter-

minados momentos de tiempo (segundos, minutos, horas), donde estos momentos estarían definidos por la naturaleza del fenómeno estudiado.

Estas observaciones son almacenadas como pares de tipo tiempo-observación, de la siguiente manera:

$$\mathcal{X}_i = \{(t_j, x_j)\}_{j=1}^{J_i}, \quad T_{\min} \leq t_1 < \dots < t_{J_i} \leq T_{\max}, \quad i = 1, \dots, n,$$

donde, la información de cada función \mathcal{X}_i se almacena de forma discreta, con J_i observaciones en el intervalo $[T_{\min}, T_{\max}]$.

El objetivo es reconstruir la forma funcional de las trayectorias muestrales $\mathcal{X}_i, i \in \{1, \dots, n\}$, que describen el comportamiento de los datos discretizados, lo cual se lo puede realizar mediante una combinación lineal de funciones base, obteniendo mediante esta reconstrucción, los datos funcionales. Como se observa, cada dato i puede estar discretizado bajo distintos tiempos t_j , ya que, $j \in \{1, \dots, J_i\}$. Para el desarrollo de este trabajo, cada dato se encuentra discretizado bajo los mismos tiempos, es decir, para cada i se tienen los tiempos t_j con $j \in \{1, \dots, J\}$.

Se debe considerar la presencia de ruido en los datos, por lo que, se supone que un dato funcional \mathcal{X}_i con $i \in \{1, \dots, n\}$ se lo puede observar mediante el modelo:

$$x_{ij} = \mathcal{X}_i(t_j) + \varepsilon_{ij}, \quad j \in \{1, \dots, J\}, \quad (2.2)$$

y de forma matricial se puede ver como:

$$\mathbf{x}_i = \mathcal{X}_i(\mathbf{t}) + \varepsilon_i,$$

donde los residuos ε_i son independientes de $\mathcal{X}_i(\mathbf{t})$. De esta manera, la reconstrucción se basa en reducir o eliminar este error y obtener una estimación lo más cercana a la función \mathcal{X}_i .

Para llevar a cabo la representación de las funciones, se necesita elegir un conjunto conocido $\{\phi_k\}_{k \in \mathbb{N}}$ de funciones base; y construir un vector de coeficientes que permita aproximar la función mediante una expansión de las funciones $\{\phi_k\}_{k \in \mathbb{N}}$.

Ramsay y Silverman [49], mencionan que las funciones $\{\phi_k\}_{k \in \mathbb{N}}$ son independientes entre ellas y que poseen la propiedad de aproximar arbitrariamente cualquier función mediante una combinación lineal de las mismas.

En la práctica, se debe seleccionar un número K suficientemente grande pero finito de estas funciones, para que los errores sean pequeños. De esta manera, las

funciones muestrales se pueden representar mediante la expresión:

$$\mathcal{X}_i(t) = \sum_{k=1}^K c_{ik} \phi_k(t) + \varepsilon_i(t), \quad i = 1, \dots, n, \quad (2.3)$$

Para la selección del conjunto $\{\phi_k\}_{k \in \mathbb{N}}$, existen varios tipos de bases, como las bases exponenciales o las bases Wavelets. Se realizará una revisión de las bases más utilizadas en el FDA, las bases de Fourier y las bases B-spline, las cuales se utilizan para suavizar datos periódicos y no periódicos, respectivamente [49].

2.1.2.1. Bases de Fourier

Las bases de Fourier están formadas por las funciones seno y coseno de frecuencia creciente, teniendo los siguientes elementos:

$$1, \sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t), \dots, \sin(\rho\omega t), \cos(\rho\omega t)$$

con $\rho \in \mathbb{N}$ y $\omega = \frac{2\pi}{\mathcal{T}}$, donde \mathcal{T} representa el período de oscilación.

Si los t_j , donde se toman las observaciones, son equiespaciados en \mathcal{T} , para todo $h \geq 1$, se define la siguiente base de funciones ortonormales:

$$\phi_0(t) = \frac{1}{\sqrt{\mathcal{T}}}, \quad \phi_{2h-1} = \frac{1}{\sqrt{\mathcal{T}/2}} \sin(h\omega t), \quad \phi_{2h} = \frac{1}{\sqrt{\mathcal{T}/2}} \cos(h\omega t) \quad (2.4)$$

Como se observa en (2.4), se tiene una secuencia alternada de senos y cosenos luego del término inicial ϕ_0 ; su argumento se multiplica por $1, 2, \dots$ hasta un cierto valor ρ , por lo que se recomienda tomar $K = 1 + 2\rho$ funciones bases. [49].

2.1.2.2. Bases B-spline

Estas funciones son las más utilizadas en el caso de aproximar series de datos no periódicos. Para entender las bases B-spline, es importante primero entender las funciones spline.

Un spline es un conjunto de polinomios por partes de orden m , los cuales se encuentran definidos sobre L subintervalos, de tal manera que la parte del polinomio que se encuentra al final de un subintervalo coincide con la parte inicial del polinomio del siguiente subintervalo.

El orden m de un polinomio $p(x)$ está dado por el número de constantes requeridas para construirlo, es decir, es el grado de $p(x)$ más uno. Los L subintervalos se

obtienen al dividir el intervalo sobre el cual se va a representar las curvas muestrales y están separados por los valores τ_l con $l = 1, \dots, L - 1$ que comúnmente son llamados nodos, de tal manera que la parte del polinomio que se encuentra al final de un subintervalo coincide con la parte inicial del polinomio del siguiente subintervalo [49].

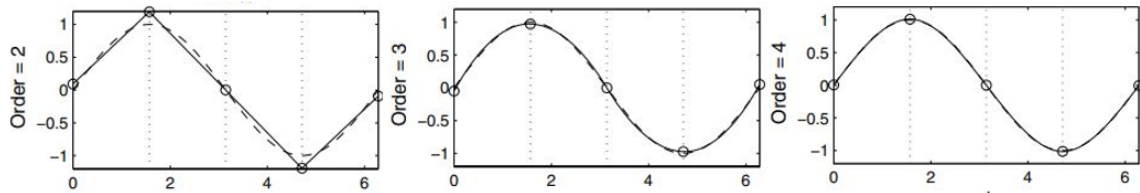


Figura 2.1: Splines de diferentes ordenes para ajustar la función seno. **Fuente:** Ramsay, J. and Silverman, B.(2005) [49].

La figura 2.1 es un ejemplo de lo que se ha descrito anteriormente. La línea entrecortada representa la función seno definida en el intervalo $[0, 2\pi]$, mientras que la línea continua representa los splines para diferentes ordenes que ajustan la función seno. Finalmente, las líneas verticales que están punteadas representan los nodos que dividen al intervalo $[0, 2\pi]$, que en el caso de los tres gráficos se tienen tres nodos internos y dividen dicho intervalo en 4 subintervalos, que se los puede observar al incluir los puntos extremos.

En el primer gráfico, se tiene un ajuste con polinomios de orden dos, es decir, polinomios lineales a trozos; en el segundo gráfico, se realiza un ajuste con polinomios cuadráticos a trozos; mientras para el último gráfico se ajusta con polinomios cúbicos a trozos. Se observa que entre más se aumenta el orden de los polinomios, se tiene un mejor ajuste.

Un spline se construye mediante el uso de un sistema de funciones ϕ_k , las cuales deben cumplir con las siguientes propiedades:

- Cada función ϕ_k debe ser un spline que se define por un orden m y un conjunto de nodos τ_l .
- Cualquier spline que sea definido por un orden m y un conjunto de nodos τ_l , se puede expresar por una combinación lineal de estas funciones.

Existe diversas maneras de poder construir estos sistemas de tal manera que cumplan con estas propiedades, siendo los más populares los B-spline desarrollados por de Boor [11]. La fórmula recursiva de Cox-De Boor, la cual es numéricamente estable, ayuda definir las funciones B-spline de la siguiente manera:

Sea τ_l con $l = 1, \dots, L - 1$, un conjunto de nodos en el intervalo $[0, L]$, se tiene:

$$B_{k,0}(t) = \begin{cases} 1 & \text{sit } \in [\tau_k, \tau_{k+1}) \\ 0 & \text{en otros casos,} \end{cases}$$

$$B_{k,m} = \frac{t - \tau_k}{\tau_{k+m} - \tau_k} B_{k,m-1}(t) + \frac{\tau_{k+m+1} - t}{\tau_{k+m+1} - \tau_{k+1}} B_{k+1,m-1}(t).$$

2.1.2.3. Proceso de ajuste

Otro de los pasos importantes a seguir cuando se está suavizando, es estimar los coeficientes c_{ik} de la expresión (2.3), que se obtiene utilizando el método de mínimos cuadrados ordinarios. En [49], se menciona que regularmente el ajuste de los datos se lo define mediante el criterio de minimización de la suma de errores cuadráticos, teniendo así la expresión:

$$SSE_i = \sum_{j=1}^J (x_{ij} - \mathcal{X}_i(t_j))^2 = \sum_{j=1}^J (x_{ij} - \mathbf{c}_i^\top \boldsymbol{\phi}(t_j))^2, \quad (2.5)$$

donde $\mathbf{c}_i \in \mathbb{R}^k$ representan los coeficientes de la combinación lineal (2.3) y $\boldsymbol{\phi}$ es el vector que contiene las K funciones base $\{\phi_k\}$.

Esta expresión es posible verla de manera matricial de la siguiente forma:

$$SSE_i = (\mathbf{x}_i - \boldsymbol{\Phi} \mathbf{c}_i)^\top (\mathbf{x}_i - \boldsymbol{\Phi} \mathbf{c}_i),$$

teniendo que \mathbf{x}_i es el vector de observaciones para el i-ésimo dato funcional y $\boldsymbol{\Phi}$ es la matriz de orden $m \times K$ que contiene los valores $\phi_k(t_j)$. De esta manera, si se toma la derivada respecto a \mathbf{c}_i , se tiene la siguiente estimación para \mathbf{c}_i :

$$\widehat{\mathbf{c}}_i = (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^\top \mathbf{x}_i, \quad (2.6)$$

por tanto, el ajuste del vector de datos $x_{ij} \approx \widehat{\mathcal{X}}_i(t_j)$ está dado por la expresión:

$$\widehat{\mathcal{X}}_i = \boldsymbol{\Phi} \widehat{\mathbf{c}}_i = \mathbf{S} \mathbf{x}_i, \quad (2.7)$$

donde S se considera como la matriz de suavizamiento de orden $m \times m$ y se la calcula mediante la expresión $\mathbf{S} = \boldsymbol{\Phi} (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^\top$.

Con esta metodología se puede suavizar los datos; sin embargo, no se necesita

que el ajuste sea perfecto porque ocasionaría que la función sea muy rugosa o por lo contrario sea incorrecta, ya que se tendría un comportamiento desorientado de los datos, por esta razón, se puede utilizar una penalización por rugosidad.

2.1.2.4. Suavizamiento mediante penalización

Otro de los criterios para poder lograr la suavidad de una función, es mediante el uso del cuadrado de la segunda derivada de \mathcal{X} en el tiempo t , que a menudo se la relaciona con la curvatura en t ; esta denotada por $[D^2\mathbf{X}(t)]^2$. Para medir la rugosidad de una función, se integra esta expresión respecto a t , es decir:

$$PEN_2(\mathcal{X}) = \int [D^2\mathbf{X}(t)]^2 dt. \quad (2.8)$$

Así, la suma penalizada de los errores cuadráticos está dada por:

$$PENSSE_\lambda(\mathbf{x}_i|\mathbf{c}_i) = [\mathbf{x}_i - \Phi\mathbf{x}_i]^\top [\mathbf{x}_i - \Phi\mathbf{x}_i] + \lambda PEN_2(\mathcal{X}_i), \quad (2.9)$$

la cual se la minimiza, de forma que ayude a regular la suavidad respecto al ajuste de la curva descrita por la función. El parámetro λ cuantifica la tasa de intercambio entre el ajuste y la variación de la función. La expresión (2.8), puede ser reescrita de la siguiente manera:

$$\begin{aligned} PEN_2(\mathcal{X}) &= \int [D^2\mathbf{X}(t)]^2 dt \\ &= \int [D^2\mathbf{c}^\top \boldsymbol{\phi}(t)]^2 dt \\ &= \int [\mathbf{c}^\top D^2 \boldsymbol{\phi}(t)] [D^2 \boldsymbol{\phi}(t)^\top \mathbf{c}] dt \\ &= \mathbf{c}^\top R \mathbf{c}, \end{aligned}$$

donde $R \in \mathbb{R}^{K \times K}$ es la matriz de penalización por rugosidad, en la cual se encuentran los productos entre cada par de funciones base, es decir:

$$\begin{aligned}
R &= \int D^2 \boldsymbol{\phi}(t) D^2 \boldsymbol{\phi}(t)^\top \\
&= \begin{pmatrix} \int [D^2 \phi_1(t)]^2 dt & \int D^2 \phi_1(t) D^2 \phi_2(t) dt & \dots & \int D^2 \phi_1(t) D^2 \phi_K(t) dt \\ \int D^2 \phi_2(t) D^2 \phi_1(t) dt & \int [D^2 \phi_2(t)]^2 dt & \dots & \int D^2 \phi_2(t) D^2 \phi_K(t) dt \\ \vdots & \vdots & \ddots & \vdots \\ \int D^2 \phi_K(t) D^2 \phi_1(t) dt & \int D^2 \phi_K(t) D^2 \phi_2(t) dt & \dots & \int [D^2 \phi_K(t)]^2 dt \end{pmatrix}
\end{aligned}$$

De esta manera, para estimar \mathbf{c}_i y \mathcal{X}_i , se reescriben las expresiones (2.6) y (2.7) de la siguiente manera:

$$\widehat{\mathbf{c}}_i = (\boldsymbol{\Phi}^\top \boldsymbol{\Phi} + \lambda R)^{-1} \boldsymbol{\Phi}^\top \mathbf{x}_i, \quad (2.10)$$

$$\widehat{\mathcal{X}}_i = \boldsymbol{\Phi} (\boldsymbol{\Phi}^\top \boldsymbol{\Phi} + \lambda R)^{-1} \boldsymbol{\Phi}^\top \mathbf{x}_i = \mathbf{S}_\lambda \mathbf{x}_i. \quad (2.11)$$

Es importante tener las siguientes consideraciones respecto al comportamiento del parámetro λ :

- Si λ toma valores muy grandes, $\lambda \rightarrow \infty$, la curva ajustada \mathcal{X} se convierte en un polinomio de grado uno, teniendo que $PEN_2(x) = 0$.
- Por otro lado, cuando λ toma valores muy pequeños, $\lambda \rightarrow 0$, \mathcal{X} tiende a volverse cada vez más variable ya que la penalización tiende a ser nula, por lo que se aproxima a ser un interpolador de los datos.

2.1.2.5. Criterio de validación cruzada generalizada GCV

Para realizar el suavizamiento de los datos funcionales se debe seleccionar el número de bases K y el parámetro de penalización λ . No existe una regla general que indique como seleccionar estos parámetros de manera óptima, teniendo en cuenta que la selección de las bases depende de la naturaleza de los datos y el objetivo de estudio.

De esta manera, para realizar la selección de estos parámetros, se utiliza el criterio de validación cruzada generalizada (GCV por sus siglas en inglés), el cual está definido de la siguiente manera [18]:

$$GCV_i(K, \lambda) = \frac{1}{J} \sum_{j=1}^J (x_{ij} - \widehat{\mathcal{X}}_i(t_j))^2 \cdot (1 - \text{tr}(S_\lambda)/J)^2, \quad (2.12)$$

en donde $tr(S_\lambda)$ representa la traza de la matriz de suavizamiento de la expresión (2.11).

2.1.3. Profundida de datos

El término profundidad hace referencia al grado de centralidad de un punto respecto a un conjunto de datos, generando un orden natural del centro hacia fuera; es decir, los puntos que se encuentren más cerca del centro tendrán una mayor profundidad en comparación de aquellos que se encuentren más alejados. Por esta razón, si se considera una muestra de datos univariantes, el dato más profundo es alguna medida de tendencia central de la muestra, y mientras más se alejan los datos de este valor central, la profundidad de los datos empieza a disminuir. Para este caso, una alternativa es tomar la mediana como punto más profundo y empezar a determinar la profundidad a partir del cuantil que representa cada punto respecto de la distribución de probabilidad empírica [5].

Para el caso multivariante, es posible tomar como dato más profundo a una generalización multivariante de la mediana. Para este caso, la profundidad empezará a disminuir a lo largo de todos los rayos que surgen del punto más profundo. Siguiendo la lógica del caso univariante y dependiendo de la forma del conjunto de datos, la profundidad de los puntos disminuirá a diferente ritmo, según la dirección que se tome para alejarse del centro. A partir de esta noción de profundidad y de distribución de probabilidad, se puede definir el orden en el espacio euclídeo multidimensional que dependerá del grado de centralidad respecto a la distribución de los datos [30].

Tukey fue uno de los primeros en utilizar el término de profundidad. Propone una construcción que permite obtener un orden entre las observaciones de una muestra multivariante, a partir de este, se han propuesto varias alternativas para el cálculo de profundidades [55].

Dyckerhoff [14], describe formalmente que fijada una distribución de probabilidad F en \mathbb{R}^d , una función de profundidad es una función acotada y no negativa $D(\cdot; F) : \mathbb{R}^d \rightarrow \mathbb{R}$, que asigna a cada punto de \mathbb{R}^d un grado de centralidad respecto a F , la cual debe cumplir las siguientes propiedades:

- **D1 Invarianza afín:** La profundidad de un punto no tiene dependencia del sistema de coordenadas que se haya elegido. Es decir, para cualquier matriz no singular $A \in \mathbb{R}^{d \times d}$ y $b \in \mathbb{R}^d$, se tiene que $D(Ax + b; F_{AX+b}) = D(x; F_X)$,

donde P_X denota la probabilidad inducida por el vector aleatorio X .

- **D2 Se anula en el infinito:** Se tiene que $\lim_{\|x\| \rightarrow \infty} D(x; F) = 0$.
- **D3 Semicontinuidad superior:** El conjunto de puntos que tiene una profundidad de al menos α , $\{x \in \mathbb{R}^d : D(x; F) \geq \alpha\}$ es cerrado para cualquier α .
- **D4 Monotonía respecto del punto más profundo:** Si se tiene que θ es un punto de profundidad máxima, entonces $D(x; F) \leq D(\theta + \lambda(x - \theta); F)$ para cualquier $0 \leq \lambda \leq 1$.

Estas propiedades fueron investigadas e introducidas para el desarrollo de la función de profundidad simplicial descrita en [30]. Dichas propiedades se convirtieron, con muy pocos cambios, en las bases requeridas para cualquier función de profundidad, dando inicio a distintas investigaciones al rededor del tema. En [5, 58] se puede encontrar una revisión de las diferentes funciones de profundidad como la profundidad: semiespacial de Tukey [55], simplicial de Liu [30], Mahalanobis propuesta por Liu y Singh [32] que está basada en la profundidad de Mahalanobis [36], entre otras.

2.1.3.1. Funciones de profundidad funcionales

Para el caso de datos funcionales, la idea básica del cálculo de profundidad, es medir la centralidad de una curva \mathcal{X}_i respecto de las curvas muestrales $\{\mathcal{X}_1, \dots, \mathcal{X}_n\}$, las cuales son generadas a partir de un proceso estocástico S y se encuentran definidas sobre un intervalo $T = [a, b] \subset \mathbb{R}$ [20].

Existe diversos métodos que permiten calcular la profundidad de datos funcionales. Por ejemplo, la profundidad modal [9], en la cual la curva más profunda es aquella que se encuentra más densamente rodeada por las curvas muestrales. La profundidad de proyección aleatoria [10], proyecta los datos en una dirección aleatoria y la profundidad de las curvas muestrales está dada por la proyección univariada, que corresponde a su proyección unidimensional, entre otras.

Para el desarrollo de este trabajo se hará uso de dos profundidades: la profundidad por bandas, definida por López-Pintado y Romo [35]; y la profundidad definida por Fraiman y Muniz [21], la cual es una de las primeras medidas de profundidad introducidas para datos funcionales.

Definición 4. Sea $\{\mathcal{X}_1, \dots, \mathcal{X}_n\}$ ua sucesión de realizaciones sobre $L^2(T)$. La gráfica de una función \mathcal{X} es el subconjunto de \mathbb{R}^2 dado por $G(\mathcal{X}) = \{(t, \mathcal{X}(t)) : t \in T\}$.

La banda en \mathbb{R}^2 delimitada por las curvas $\mathcal{X}_{i_1} \dots \mathcal{X}_{i_n}$ es:

$$B(\mathcal{X}_{i_1} \dots \mathcal{X}_{i_n}) = \{(t, \mathcal{X}) : t \in T; \min_{r=1, \dots, k} \{\mathcal{X}_{ik}(t)\} \leq \mathcal{X} \leq \max_{i=1, \dots, r} \{\mathcal{X}_{ik}(t)\}\} \quad (2.13)$$

Sea J un parámetro exógeno tal que $2 \leq J \leq n$, que indica el máximo número de curvas usadas para construir una banda. La profundidad de banda para cualquier función \mathcal{X} en $\mathcal{X}_1 \dots \mathcal{X}_n$ se define como [34]:

$$BD_{n,J} = \sum_{j=2}^J BD_n^{(j)}(\mathcal{X}),$$

donde $BD_n^{(j)}$ está definida como :

$$BD_n^{(j)}(\mathcal{X}) = \binom{n}{j}^{-1} \sum_{1 < i_1 < \dots < i_j < n} I(G(\mathcal{X}) \subseteq B(\mathcal{X}_{i_1}, \dots, \mathcal{X}_{i_j}))$$

teniendo en cuenta que $I(\cdot)$ es la función indicatriz. $BD_n^{(j)}(\mathcal{X})$ expresa la proporción de bandas $B(\mathcal{X}_{i_1}, \dots, \mathcal{X}_{i_j})$ determinadas por j curvas diferentes $\mathcal{X}_{i_1} \dots \mathcal{X}_{i_j}$ contenidas en todo el gráfico de \mathcal{X} , por tanto la curva mas profunda será aquella \mathcal{X}_i donde $BD_{n,J}$ es máxima.

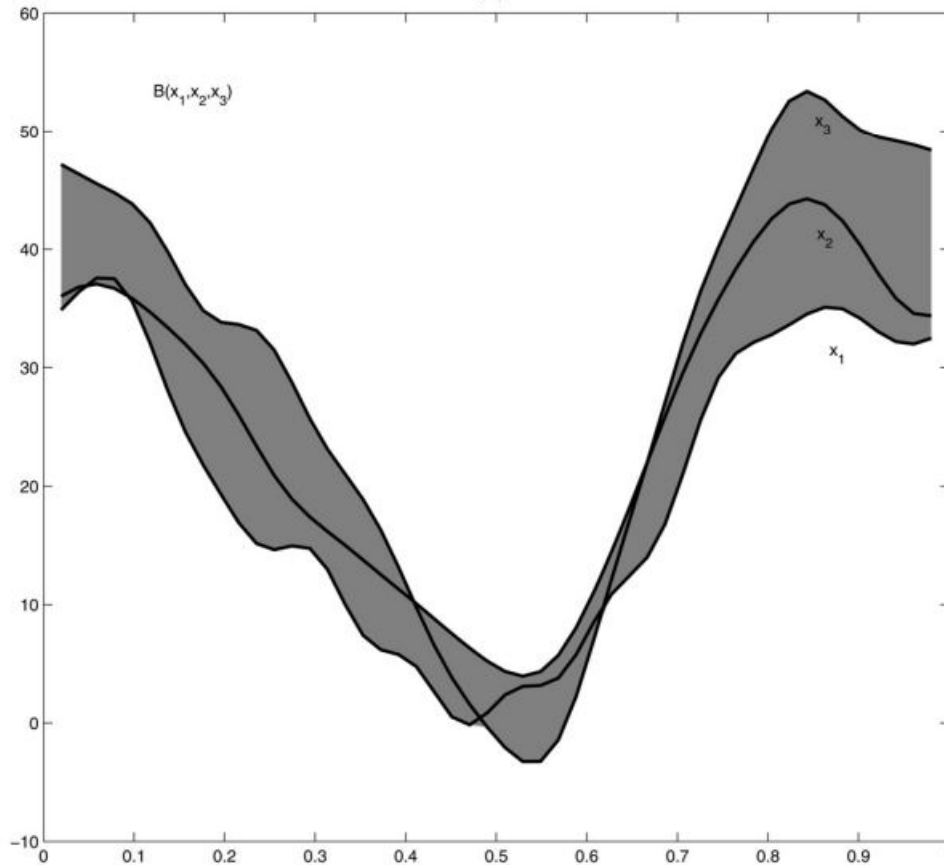


Figura 2.2: Banda definida por tres curvas, contenidas completamente en la banda.
Fuente: Lopéz-Pintado, S. y Romo, J. (2009) [34].

Definición 5. Sea $\{\mathcal{X}_1(t), \dots, \mathcal{X}_n(t)\}$ realizaciones independientes e idénticamente distribuidas de la variable \mathcal{X} definida sobre el intervalo $[a, b]$ y sea D_n una medida de profundidad definida en \mathbb{R} . Para cada $t \in [a, b]$, se considera $Z_i(t) = D_n(\mathcal{X}_i(t))$ como la profundidad univariada de $\mathcal{X}_i(t)$ en el tiempo t respecto de $\mathcal{X}_1(t) \dots \mathcal{X}_n(t)$.

Se define la profundidad de Fraiman y Muniz (FMD) [21] para el i -ésimo dato como:

$$FMD(\mathcal{X}_i) = \int_a^b Z_i(t) dt, \quad 1 \leq i \leq n$$

De esta manera, se obtienen estadísticas de orden para las curvas $\mathcal{X}_1(t) \dots \mathcal{X}_n(t)$ en función de los valores calculados y por tanto, la mediana funcional será aquella curva $\mathcal{X}_i(t)$ que tenga el valor $FMD(\mathcal{X}_i)$ máximo.

Se hace uso de la siguiente función de profundidad univariante:

$$Z_i(t) = D_n(\mathcal{X}_i(t)) = 1 - \left| \frac{1}{2} - F_{n,t}(\mathcal{X}_i(t)) \right|, \quad (2.14)$$

obteniendo así que:

$$FMD(\mathcal{X}_i) = 1 - \int_a^b \left| \frac{1}{2} - F_{n,t}(\mathcal{X}_i(t)) \right|, \quad (2.15)$$

donde $F_{n,t}$ es la distribución empírica definida como:

$$F_{n,t}(\mathcal{X}_i(t)) = \frac{1}{n} \sum_{k=1}^n I(\mathcal{X}_k(t) \leq \mathcal{X}_i(t))$$

donde $I(\cdot)$ representa la función indicatriz.

Adicionalmente, Fraiman y Muniz [21] realizan una extensión funcional para el cálculo de la media recortada usando la función (2.14), la cual es una medida de localización robusta.

Definición 6. Sea $\mathcal{X}_{(1)} \dots \mathcal{X}_{(n)}$, de tal manera que $\mathcal{X}_{(1)}$ representa la curva más profunda y $\mathcal{X}_{(n)}$ es la curva menos profunda. Se define la media recortada funcional como:

$$\hat{\mu}_{TM,\alpha} = \frac{1}{n - [\alpha n]} \sum_{i=1}^{n - [\alpha n]} \mathcal{X}_{(i)}, \quad (2.16)$$

donde $0 \leq \alpha \leq \frac{(n-1)}{n}$ y $[.]$ denota la parte entera.

Si $\alpha = 0$, entonces se tiene que $\hat{\mu}_{TM,\alpha}$ es la media funcional. Adicionalmente si solo una curva es la más profunda se tiene que $FMD(\mathcal{X}_1, \dots, \mathcal{X}_n) = \mathcal{X}_{(1)}$ es decir esta curva representaría la mediana funcional, de otra manera la mediana funcional $FMD(\mathcal{X}_i)$ esta definida como la media de las curvas que maximiza la profundidad [20].

2.1.3.2. P-regiones centrales funcionales

La región central se asocia a una noción de profundidad, en este trabajo para calcular dicha región, se utilizará las profundidades previamente revisadas.

Liu, Parelius y Singh [31], introducen varios conceptos sobre profundidad para el caso multivariante, entre ellos, los referentes a contornos y regiones centrales; mientras que, Zuo y Serfling [58] realizan un estudio sobre las propiedades de las mismas; por otro lado, López-Pintado y Romo [34] proponen una extensión de estos estudios para datos funcionales.

Inicialmente, es importante definir lo que es un contorno, para ello se toma lo descrito en [34]:

Definición 7. Sea X_1, \dots, X_n , procesos estocásticos independientes e idénticamente distribuidos con distribución F , tales que toman valores en el espacio L^2 definido por (2.1), sea una función de profundidad $D(\cdot, F) = D(\cdot)$ y su respectiva versión muestral definida por $D(\cdot, F_n) = D_n(\cdot)$ y $\alpha \geq 0$, el conjunto definido por:

$$\partial D^\alpha = \{\mathcal{X} \in L^2 : D(\mathcal{X}) = \alpha\},$$

se conoce como α -contorno o como contorno de profundidad α . Su versión muestral es:

$$\partial D_n^\alpha = \{\mathcal{X} \in L^2 : D_n(\mathcal{X}) = \alpha\},$$

es decir, el α -contorno, está compuesto por aquellas curvas que tienen una profundidad igual a α .

Definición 8. Sea X_1, \dots, X_n , procesos estocásticos independientes e idénticamente distribuidos con distribución F , tales que toman valores en el espacio L^2 definido por la expresión (2.1), sea una función de profundidad $D(\cdot, F) = D(\cdot)$ y su respectiva versión muestral definida por $D(\cdot, F_n) = D_n(\cdot)$ y $\alpha \geq 0$, el conjunto definido por:

$$D^\alpha = \{\mathcal{X} \in L^2 : D(\mathcal{X}) \geq \alpha\},$$

se denomina α -región recortada y es la región delimitada por el contorno de profundidad α , es decir, es el conjunto de aquellas funciones cuya profundidad es mayor o igual a α , y su respectiva versión muestral está dada por:

$$D_n^\alpha = \{\mathcal{X} \in L^2 : D_n(\mathcal{X}) \geq \alpha\}.$$

Las propiedades de la α -región recortada se enuncian en el siguiente resultado:

TEOREMA 1. Sea F una distribución de probabilidad en L^2 , entonces:

- $D^\alpha(F_{fX+g}) = f * D^\alpha(F_X) + g$, y $D_n^\alpha(F_{fX+g}) = f * D_n^\alpha(F_X) + g$ su versión muestral, donde f y g son funciones en L^2 y f es distinta de cero para todo $t \in T$.
- Sea $\alpha_1, \alpha_2 \geq 0$, para $\alpha_1 \geq \alpha_2$ se tiene que $D^{\alpha_1} \subset D^{\alpha_2}$ y $D_n^{\alpha_1} \subset D_n^{\alpha_2}$.

La primera propiedad está relacionada con la invarianza afín del caso multivariante, mientras que la segunda propiedad es directa a partir de la definición de región recortada.

Una vez se ha realizado una revisión sobre α -regiones recortadas, se puede definir lo que es una región central.

Definición 9. Sea F una distribución de probabilidad en L^2 , y P una medida de probabilidad, entonces el conjunto:

$$CR_p = \bigcap_{\alpha} \{D^\alpha : P(D^\alpha) \geq p\}, \quad 0 \leq p \leq 1,$$

se denomina p -región central. Este es el conjunto más pequeño determinado por regiones recortadas que contengan al menos probabilidad p .

En el caso de multivariante, la p -región central es un conjunto compacto y conectado, donde su límite es el p -contorno que se denota por ∂CR_p . Por otro lado, si se tiene que la función de densidad es distinta de cero en el espacio \mathbb{R}^d , entonces ∂CR_p es el contorno de $\{x \in \mathbb{R}^d : D(x) = \alpha_p\}$, en donde $P\{x \in \mathbb{R}^d : D(x) \geq \alpha_p\} = p$, por lo tanto, el p -contorno (∂CR_p) puede ser considerado como una superficie cuantil.

Considerando que las regiones recortadas muestrales D_n^α no son observables directamente y deben ser aproximadas, Liu, Parelius y Singh [31] proponen una estimación para el caso multivariante, tanto para las regiones recortadas D^α como la p -regiones centrales CR_p .

De esta manera, sea $x_{(1)}, \dots, x_{(r_\alpha)}$ en \mathbb{R}^d los datos muestrales, tales que se encuentran ordenados de tal manera que $x_{(1)}$ es el dato más profundo y $x_{(r_\alpha)}$ el menos profundo, además estos datos tienen una profundidad mayor o igual que α ($D(x_{(i)}) \geq \alpha$), es decir, se tiene que $D_n(x_{(1)}) \geq \dots \geq D_n(x_{(r_\alpha)})$, donde $D_n(x_{(r_\alpha)}) \geq \alpha$ y $D_n(x_{(r_\alpha+1)}) < \alpha$. D^α es estimada como la envolvente convexa de los puntos que tienen una profundidad mayor o igual a α :

$$\hat{D}_n^\alpha = \text{envolvente convexa}\{x_{(1)}, \dots, x_{(r_\alpha)}\},$$

en donde para esta estimación se cumple que, si $\alpha_1 \geq \alpha_2$, $\hat{D}_n^{\alpha_1} \subset \hat{D}_n^{\alpha_2}$.

Para el caso de la p -región central, se tiene la estimación mediante la envolvente convexa de la proporción p de datos más profundos de la muestra, por lo que se define como:

$$CR_{n,p} = \text{envolvente convexa}\{x_{(1)}, \dots, x_{(x_{|np|})}\},$$

donde

$$|np| = \begin{cases} np & \text{si } np \text{ es entero,} \\ 1 + [np] & \text{en otros casos,} \end{cases} \quad (2.17)$$

y $[.]$ representando la parte entera del número np . Así, $CR_{n,p}$ aproxima a la p -región central muestral y su límite $\partial CR_{n,p}$ es el p -contorno o la superficie cuantil p .

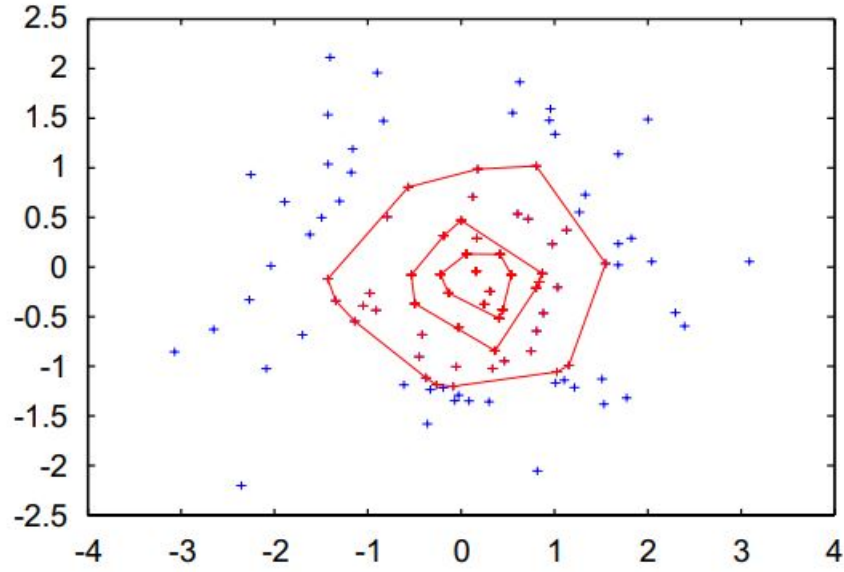


Figura 2.3: Región central estimada $C_{n,p}$. **Fuente:** López-Pintado, S. y Romo, J. (2009) [34].

En la figura 2.3, se muestra las estimaciones de las regiones centrales mediante $CR_{n,p}$ con una muestra de tamaño $n = 100$, que proviene de una distribución normal bivariada con media cero y matriz de covarianza $\Sigma = \text{diag}(2, 1)$, los valores para p que se consideran son 0,1, 0,2, 0,5; los puntos de color rojo representan el 50 % datos más profundos. Existen diversas investigaciones donde se proponen otras formas de estimar regiones recortadas y regiones centrales, estas pueden ser revisadas en [1].

Para el caso funcional, López-Pintado y Romo [34] definen la siguiente estimación para regiones recortadas y regiones centrales, que se realizan mediante el cálculo de bandas definida en la ecuación (2.13).

Definición 10. Sea $\{\mathcal{X}_1, \dots, \mathcal{X}_n\}$ una muestra de funciones en el espacio L^2 . Sea $\mathcal{X}_{(1)} \dots \mathcal{X}_{(n)}$ la muestra ordenada de manera decreciente de acuerdo a su profundidad, calculada mediante D_n una función de profundidad. Sean $\mathcal{X}_{(1)}, \dots, \mathcal{X}_{(r_\alpha)}$ las observaciones que tienen una profundidad mayor o igual a α . La estimación para la α -región recortada D^α está dada por:

$$\tilde{D}_n^\alpha = \{(t, y) \in I \times \mathbb{R} : \min_{i=1, \dots, r_\alpha} \{\mathcal{X}_{(i)}(t)\} \leq y \leq \max_{i=1, \dots, r_\alpha} \{\mathcal{X}_{(i)}(t)\}\}, \quad (2.18)$$

que es la banda delimitada por las curvas de la muestra con una profundidad mayor o igual a α .

Y, para la p -región central se tiene que la estimación está dada por:

$$B_{n,p} = B(\mathcal{X}_1, \dots, \mathcal{X}_{(\lfloor np \rfloor)}), \quad (2.19)$$

que en este caso, es la banda definida por la proporción p de curvas más profundas de la muestra.

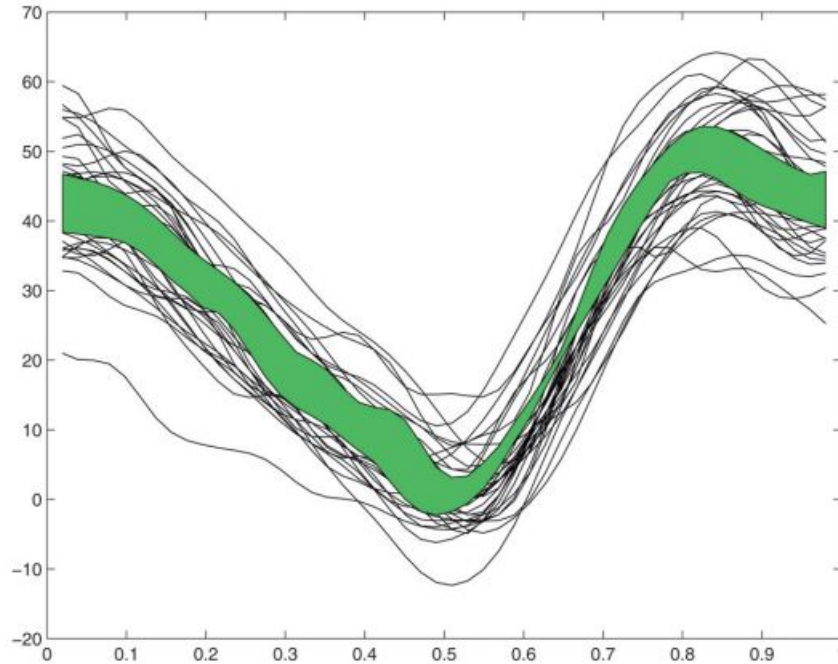


Figura 2.4: Región central con $p = 0.1$ **Fuente:** López-Pintado, S. y Romo, J. (2009) [34].

En la figura 2.4 se tiene un conjunto de funciones que representan el ángulo en el plano sagital formadas por la cadera de 39 niños durante el ciclo de caminado; la franja verde representa la región central de estas curvas con el 10% de curvas más profundas, es decir se tomó un valor de $p = 0,1$ [49].

2.2. Índices de capacidad de proceso

El control y la optimización continua de procesos, engloba todo tipo de acción que posibilita replantearlos, detectar actividades de mejora de las pautas de trabajo y crear ventajas productivas para la empresa, todo con el fin de generar valor agregado para sus clientes [39].

La capacidad del proceso puede ser evaluada por el cálculo de diversas rela-

ciones e índices de capacidad. Los índices de capacidad de proceso (PCI por sus siglas en inglés) son indicadores estadísticos muy eficaces, ampliamente aceptados y utilizados en la industria. Específicamente en el monitoreo estadístico del proceso para determinar si la capacidad del sistema y el rendimiento de los productos alcanzan los requisitos planteados [13]; es decir, un proceso se considerará capaz, si este puede producir elementos dentro de los límites de especificación que se han determinado previamente.

Un aspecto importante que se debe resaltar del análisis de capacidad es que existen PCI univariados y multivariados. Por una parte, los PCI univariados pueden utilizarse de manera independiente para obtener la capacidad relacionada a una sola característica, para un producto en particular; mientras que, los PCI multivariados describen la capacidad de un proceso, en el cual se consideran varias o todas las características de un producto de una manera global. De esta manera, los PCI son indicadores de la bondad del proceso relacionado con la posición y la variabilidad de las medidas dentro de los límites de especificación, que dependiendo de la característica que se esté analizando, pueden ser utilizados para especificaciones unilaterales o para especificaciones bilaterales.

2.2.1. Índices de capacidad de proceso univariados tradicionales

Los índices de capacidad de un proceso univariado son indicadores estadísticos, que son utilizados para cuantificar la bondad de un proceso relacionando la variabilidad de las medidas de una sola característica del producto con las medidas admisibles o planteadas. Se acepta que las medidas obtenidas, de las características del producto en las pruebas de calidad, siguen una distribución normal en la mayoría de los procesos de producción [39].

Para describir los índices de capacidad univariados necesarios para este trabajo, se utilizará la enciclopedia de índices de capacidad realizada por Pearn y Kots [45].

2.2.1.1. Índice de capacidad C_p

Aparentemente, el primer índice de capacidad de proceso que aparece en la literatura, es el índice C_p , definido por Kane [27]:

$$C_p = \frac{USL - LSL}{6\sigma}, \quad (2.20)$$

donde USL representa el límite de especificación superior, LSL el límite de especificación inferior y σ es la desviación estándar del proceso. El numerador de la expresión (2.20), se conoce como la región de especificación.

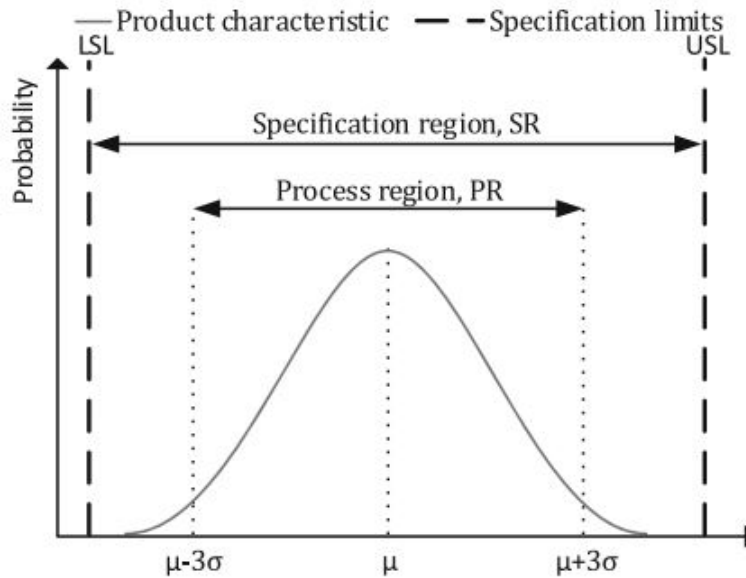


Figura 2.5: Región de especificación y región de proceso. **Fuente:** de Felipe, D. y Benedito, E. (2017) [13].

Como se observa en la figura 2.5, esta región se encuentra definida como el intervalo delimitado por los límites de especificación. Por otro lado, el denominador (6σ) se denomina como la región del proceso, se encuentra definida como el intervalo que incluye el 99,73% de los valores extraídos de una distribución normal y que se centra en el valor medio de la característica del producto medido. Así para el caso univariante, como se observa en la figura 2.5, el punto más bajo se encuentra en $\mu - 3\sigma$ y el punto más alto en $\mu + 3\sigma$. Generalmente, la varianza del proceso σ es desconocida, por lo que es posible estimarla mediante una muestra ($\hat{\sigma}$).

De acuerdo con [54], es posible reconocer entre varios niveles de capacidad cuando se realiza el uso del índice C_p :

Valores índice C_p	Descripción del Proceso
$C_p \geq 2,00$	Excelente
$1,67 \leq C_p < 2,00$	Satisfactorio
$1,33 \leq C_p < 1,67$	Bueno
$1,00 \leq C_p < 1,33$	Capaz
$0,67 \leq C_p < 1,00$	Inadecuado
$C_p < 0,67$	Pobre

Tabla 2.1: Niveles del proceso de acuerdo a los valores de C_p . **Fuente:** Tsai, C. y Chen, C. (2006) [54]. **Elaboración:** Autor

Como se observa en la tabla 2.1, es deseable tener valores grandes para el índice C_p ; ya que, el proceso se considera más capaz a medida que los valores crecen, lo cual, se logra cuando la variabilidad del proceso o región del proceso es pequeña en relación a la región de especificación, tal como se contempla en la expresión (2.20).

Uno de los problemas del índice de capacidad de proceso C_p , es que no tiene en cuenta donde se encuentra la media del proceso en relación con los límites de especificación; este índice mide la dispersión de las especificaciones en relación con la dispersión del proceso, es decir solo mide la capacidad potencial del proceso. Así, si la región del proceso no se encuentra centrada respecto a la región de especificación, es posible tener un porcentaje de productos con características fuera de los límites de especificación, aunque el índice C_p tenga un valor alto.

Para tener en consideración el centrado del proceso, los índices C_{pu} , C_{pl} , C_{pk} fueron introducidos y son descritos a continuación.

2.2.1.2. Índices de capacidad C_{pu} y C_{pl}

Los índices de capacidad C_{pu} y C_{pl} se lo conoce como índices de capacidad superior e inferior respectivamente, y son medidas unilaterales de la capacidad del proceso. Estos índices están formulados de la siguiente manera:

$$C_{pu} = \frac{USL - \mu}{3\sigma}, \quad (2.21)$$

$$C_{pl} = \frac{\mu - LSL}{3\sigma}. \quad (2.22)$$

El índice C_{pu} describe la relación que existe entre la mitad superior de la región de especificación ($USL - \mu$) y la mitad de la región del proceso (3σ). Análogamente, el

índice C_{pl} describe la relación entre la mitad inferior de la región de especificación ($\mu - LSL$) y la mitad de la región de proceso (3σ). Al igual que el índice C_p , μ y σ son estimadas ($\bar{x}, \hat{\sigma}$) mediante una muestra y son válidas para los índices descritos.

2.2.1.3. Índice de capacidad C_{pk}

Este índice toma en cuenta donde se encuentra la media del proceso, para lo cual, hace el uso de los índices definidos por las expresiones (2.21) y (2.22), y se define de la siguiente manera:

$$C_{pk} = \text{mín}\{C_{pu}, C_{pl}\}, \quad (2.23)$$

Este índice representa la relación de la distancia que existe entre el centro del proceso y el límite de especificación más cercano a la mitad de la variabilidad del proceso (3μ).

En [39] se explica que, si se cumple la igualdad $C_p = C_{pk}$, el proceso se encuentra centrado en el punto medio de los límites de especificación, mientras que si se tiene la desigualdad $C_{pk} < C_p$, el proceso se encuentra fuera del punto medio de los límites de especificación. Por otro lado, se explica que, la magnitud del índice C_{pk} en relación del índice C_p , es una medida directa de que tan lejos del centro está operando el proceso; siendo, un parámetro que mide la capacidad actual o real del proceso. Por lo general, si el valor del índice C_{pk} es mayor que 1, el proceso se considera como capaz, aunque Bothe [3] define que el valor más aceptado para considerar al proceso capaz debe ser de 1,33 o superior.

Todos los índices de capacidad descritos previamente están definidos sobre el supuesto de que la distribución del proceso es normal y tiene una tolerancia simétrica; es decir, que el valor objetivo coincida con el punto medio del intervalo de especificación. Sin embargo, la mayoría de procesos producen datos no normales y el uso de estos índices de capacidad pueden producir resultados engañosos, como se describe en [6, 23].

Una aproximación para solventar el problema de la no normalidad de los datos, es transformarlos mediante funciones, de tal manera que se conviertan en datos con una distribución normal y por tanto, usar los índices de capacidad descritos previamente. Siguiendo este enfoque, existen varios estudios como el de Johnson [26], que elaboró un sistema de distribuciones basado en el método de momentos, Somerville y Montgomery [52] utilizaron una transformación de raíz cuadrada o Box y Cox [4] que presentaron una familia de transformación de potencia.

Una segunda aproximación es el uso de percentiles no normales para modificar los índices tradicionales. Clements [8], propuso calcular los índices C_p y C_{pk} para distribuciones de cualquier forma, haciendo uso de la familia de curvas de Pearson. La ventaja principal al utilizar este método, es que no se requiere de un ajuste complicado de la distribución y es sencillo de utilizar. Estos índices se describen de la siguiente manera:

$$C_{p(q)} = \frac{USL - LSL}{x_{0,99865} - x_{0,00135}}, \quad (2.24)$$

y

$$C_{pk(q)} = \min\{C_{pu(q)}, C_{pl(q)}\}, \quad (2.25)$$

donde

$$C_{pu(q)} = \frac{USL - x_{0,5}}{x_{0,99865} - x_{0,5}}, \quad (2.26)$$

$$C_{pl(q)} = \frac{x_{0,5} - LSL}{x_{0,5} - x_{0,00135}}, \quad (2.27)$$

teniendo en cuenta que $x_p = p * 100$ representa el percentil p de los datos.

2.2.2. Índices de capacidad de proceso funcionales

Guevara y Vargas [22], proponen una extensión de los índices tradicionales C_p , C_{pu} , C_{pl} , C_{pk} descritos anteriormente hacia el campo funcional, utilizando la propuesta de Clements [8], que implica que los índices no dependan de la distribución que tengan los datos. Se consideran las funciones $\mathcal{X}_{0,5}$, $\mathcal{X}_{0,99865}$ y $\mathcal{X}_{0,00135}$ como las versiones poblacionales de la curva más profunda y de los percentiles 0,99865 y 0,00135 respectivamente.

De esta manera, se proponen dos métodos distintos para el cálculo de los índices de capacidad. El primer método o método A, se considera que el proceso es evaluado en J puntos (t_j con $j = 1, \dots, J$), así, se definen los índices de capacidad como el promedio de los índices de capacidad de proceso univariantes evaluados en cada uno de los puntos t_j , teniendo las siguientes expresiones para cada índice:

$$C_{p(prf)}^{(A)} = \frac{1}{J} \sum_{j=1}^J \frac{USL(t_j) - LSL(t_j)}{\mathcal{X}_{0,99865}(t_j) - \mathcal{X}_{0,00135}(t_j)} \quad (2.28)$$

$$C_{pu(prf)}^{(A)} = \frac{1}{J} \sum_{j=1}^J \frac{USL(t_j) - \mathcal{X}_{0,5}(t_j)}{\mathcal{X}_{0,99865}(t_j) - \mathcal{X}_{0,5}(t_j)} \quad (2.29)$$

$$C_{pl(prf)}^{(A)} = \frac{1}{J} \sum_{j=1}^J \frac{\mathcal{X}_{0,5}(t_j) - LSL(t_j)}{\mathcal{X}_{0,5}(t_j) - \mathcal{X}_{0,00135}(t_j)} \quad (2.30)$$

$$C_{pk(prf)}^{(A)} = \frac{1}{J} \sum_{j=1}^J \min \left[\frac{USL(t_j) - \mathcal{X}_{0,5}(t_j)}{\mathcal{X}_{0,99865}(t_j) - \mathcal{X}_{0,5}(t_j)}, \frac{\mathcal{X}_{0,5}(t_j) - LSL(t_j)}{\mathcal{X}_{0,5}(t_j) - \mathcal{X}_{0,00135}(t_j)} \right] \quad (2.31)$$

donde si $n = 1$ se tienen los índices de capacidad univariados dados por las ecuaciones (2.24), (2.26), (2.27), (2.25).

El segundo método o método B utiliza el hecho de que las funciones LSL, USL, $\mathcal{X}_{0,00135}$, $\mathcal{X}_{0,5}$, y $\mathcal{X}_{0,99865}$ son continuas, definiendo los índices de la siguiente manera:

$$C_{p(prf)}^{(B)} = \frac{\int_I (USL(t) - LSL(t)) dt}{\int_I (\mathcal{X}_{0,99865}(t) - \mathcal{X}_{0,00135}(t)) dt} \quad (2.32)$$

$$C_{pu(prf)}^{(B)} = \frac{\int_I (USL(t) - \mathcal{X}_{0,5}(t)) dt}{\int_I (\mathcal{X}_{0,99865}(t) - \mathcal{X}_{0,5}(t)) dt} \quad (2.33)$$

$$C_{pl(prf)}^{(B)} = \frac{\int_I (\mathcal{X}_{0,5}(t) - LSL(t)) dt}{\int_I (\mathcal{X}_{0,5}(t) - \mathcal{X}_{0,00135}(t)) dt} \quad (2.34)$$

$$C_{pk(prf)}^{(B)} = \min \left(C_{pu(prf)}^{(B)}, C_{pl(prf)}^{(B)} \right) \quad (2.35)$$

donde I es un intervalo compacto en \mathbb{R} .

Es importante tener en cuenta que los dos métodos (A y B) cuantifican la relación que existe entre el rendimiento real del proceso y los límites de especificación.

Capítulo 3

Aplicación de índices de capacidad de proceso funcionales y Resultados

En este capítulo se describirá la metodología para aplicar los índices de capacidad de proceso funcionales univariados propuestos en el capítulo 2, específicamente en la subsección 2.2.2. Se evaluará la capacidad climática para la producción de maíz amarillo duro, es decir, se quiere entender si existe la capacidad climática que permita llevar a cabo la producción del maíz y obtener rendimientos óptimos. Para esto se considera al clima como un proceso, el cual se asume se encuentra en control, mientras que los límites de especificación USL y LSL son definidos en función de los requerimientos climáticos para el desarrollo del maíz.

3.1. Datos y descripción de las variables climáticas

Para el desarrollo del trabajo se consideró como punto de estudio a La Concordia. La Concordia es un cantón de la provincia de Santo Domingo de los Tsáchilas, ubicada en la parte norte de la región litoral del Ecuador, a una altitud de 217 metros sobre el nivel del mar (msnm). En este cantón los principales cultivos son: palma africana, plátano, cacao, abacá, maíz duro seco, arroz, café y maracuyá. Este cantón aumentó su economía en los últimos años gracias a su actividad agrícola¹. Su clima, es lluvioso tropical con precipitaciones que oscilan entre 2000 y 3000 milímetros de agua (mm) al año. Su temperatura promedio varía entre 23°C y 25,5°C [37].

¹<https://www.elcomercio.com/actualidad/produccion-quevedo-santodomingo-concordia-agricultura/>

3.1.1. Datos

Los datos necesarios para este trabajo fueron tomados de la página NASA POWER. Este es un proyecto se inició para crear conjuntos de datos a partir de nuevos sistemas de satélites; el cual contiene datos sobre: energías renovables, sistemas sostenidos y agroclimatología. Utilizando este recurso, basta con ingresar los datos de latitud y longitud del punto de estudio, seleccionar las variables de interés y seleccionar el rango de fechas en el que se desean los datos para el estudio.

Según Segura y Andrade [50], la siembra del maíz debe ser realizada cuando se está iniciando la época lluviosa en la costa ecuatoriana, con preferencia entre el 15 de diciembre al 30 de enero. Para el desarrollo de este trabajo, se consideró como fecha de inicio de siembra el 15 de diciembre.

En [43, 50], se establece que el tiempo de siembra-cosecha, es de 120 días para los tipos de híbrido de maíz, que se suelen utilizar en la zona, como: el Brasilia 8501, INIAP H-551, INIAP H-552, Trueno NB 7443, Vencedor R-8330, entre otros; por lo que, se consideró 120 días para el período de siembra y cosecha.

De esta manera, los datos fueron tomados diariamente entre 1984 y 2020, empezando desde el 15 de diciembre y terminando con la cosecha en el mes de abril del siguiente año. Se consideraron observaciones semanales, es decir, cada punto de discretización para las funciones será la suma de 7 días. De esta manera, los datos para la aplicación se componen por 36 funciones o curvas, cada una con 17 puntos de discretización, esto para cada variable considerada para el estudio.

3.1.2. Variables climáticas

Las variables climáticas consideradas para evaluar la capacidad climática para la producción del maíz son:

- **Precipitación:** se considera como cualquier forma de hidrometeoro² que cae desde la atmósfera hasta llegar a la superficie terrestre, entre los cuales se encuentran fenómenos como: lluvia, llovizna, nieve, aguanieve, entre otros [Meteoblue].
- **Tiempos Térmicos:** Es la acumulación de temperatura que requiere la planta

²Un hidrometeoro es un fenómeno meteorológico que está compuesto por una agrupación de partículas sólidas, líquidas o acuosas que caen desde la atmósfera.

para su desarrollo. Para el cálculo de los tiempos térmicos se utiliza el método residual descrito en [51], y está dado por la siguiente ecuación:

$$TT = \left[\frac{T_{\text{máx}} + T_{\text{mín}}}{2} \right] - TB, \quad (3.1)$$

donde $T_{\text{máx}}$ es la temperatura máxima, $T_{\text{mín}}$ es la temperatura mínima y TB es la temperatura base. Para la temperatura base se considerará lo establecido en [56], que para el caso del maíz es de un valor igual a 10 °C.

- **Radiación solar fotosintéticamente activa:** también se conoce como radiación PAR por sus siglas en inglés, y es la que abarca las longitudes de onda del espectro visible. Esta radiación favorece a la planta en su proceso de fotosíntesis, además de ayudar en la producción de carbohidratos. Una buena interceptación de PAR por parte de las plantas logra una mayor obtención de materia seca (MS) [48].

3.2. Límites de especificación

Los límites de especificación son aquellos límites entre los que debe operar el producto o servicio. Para el desarrollo de este trabajo, como se mencionó anteriormente, estos límites son establecidos en función de las necesidades que tiene el maíz para obtener un óptimo desarrollo en el periodo de siembra-cosecha. Estos límites son fijados para cada una de las variables consideradas.

3.2.1. Límites de especificación para la precipitación

Norman, Pearson y Searle [42] manifiestan que el cultivo del maíz duro seco en tierras bajas necesita una precipitación entre 600 mm a 1300 mm distribuidas a lo largo del período de cultivo. Este requerimiento de precipitación resulta ser el mismo para cualquier híbrido de maíz en tierras bajas. Segura y Andrade [50] establecen los porcentajes de precipitación que se necesitan en cada fase del desarrollo y estos son:

1. **Fase de germinación:** se necesita un 5 % del requerimiento total y se encuentra comprendida entre 0 - 5 días luego de la siembra.
2. **Fase de desarrollo vegetativo:** se necesita el 23 % del total requerido y se encuentra comprendida 5 - 35 días luego de la siembra.

3. **Fase de prefloración:** se necesita el 23 % del total requerido y se encuentra comprendida entre 35 - 42 días luego de la siembra.
4. **Fase de floración:** se necesita el 14 % del requerimiento total y se encuentra comprendida entre 42 - 48 días luego de la siembra.
5. **Fase de llenado de grano:** se necesita el 34 % del total requerido y se encuentra comprendida entre 48 - 90 días luego de la siembra.
6. **Fase de llenado de grano:** para la fase final se necesita menos del 1 % del total requerido y empieza desde el día 90 hasta el día 120 luego de la siembra.

Sin embargo, para el desarrollo de este trabajo, se distribuye el requerimiento de precipitación equitativamente a lo largo del período siembra - cosecha, es decir, se establecieron los límites de especificación, distribuyendo esta precipitación a lo largo de las 17 semanas consideradas, obteniendo los siguientes límites de especificación:

Límite de especificación	Valor establecido
USL_p	76,5 mm.
LSL_p	35,5 mm.

Tabla 3.1: Límites de especificación para la precipitación. **Elaboración:** Autor.

Los valores de la tabla 3.1, representan los valores acumulados de una semana para los límites de especificación USL_p y LSL_p .

3.2.2. Límites de especificación para los tiempos térmicos

Los límites de especificación para la variable tiempo térmico, deben ser establecidos mediante el uso de la fórmula (3.1), es decir, se necesita información sobre la temperatura mínima y máxima requeridas para la producción del maíz duro seco. Leng y Aldrich [29] establecen que la temperatura mínima para el maíz duro seco debe encontrarse en un rango de 15°C a 20°C, la temperatura máxima debería variar entre 25°C a 30°C, y la temperatura base de 10°C. Es importante mencionar que las temperaturas mencionadas son las mismas para cualquier híbrido de maíz duro seco, por tanto, los tiempos térmicos resultan ser una variable general para cualquier híbrido de maíz al igual que la precipitación.

De esta manera, se tienen los siguientes valores:

Límite de especificación	Valor establecido
USL_{TT}	122,5 °C
LSL_{TT}	52,5 °C

Tabla 3.2: Límites de especificación para los tiempos térmicos. **Elaboración:** Autor.

Los valores establecidos en la tabla 3.2 para USL_{TT} y LSL_{TT} , representan los valores acumulados de una semana.

3.2.3. Límites de especificación para la radiación solar fotosintéticamente activa

Para definir los límites de especificación para la radiación fotosintéticamente activa, se estableció el nivel de producción del maíz, donde cada megajoul (MJ) interceptado ayuda a producir entre 2 y 4 gramos de maíz [2, 44]. Por esta razón, se consideran tres tipos de híbrido de maíz que tienen mayor demanda en la región norte del litoral, estableciendo así tres límites de especificación distintos.

El INIAP H-551 e INIAP H-552, son dos híbridos de maíz que se caracterizan por ser altamente adaptables en la región litoral, comparten ciertas características como el período de siembra y cosecha de 120 días; mientras que sus diferencias pueden ser revisadas en [15]. En esta investigación, se toma en cuenta que el potencial de rendimiento de producción, para los híbridos INIAP H-551 e INIAP H-552, es aproximadamente 160 y 134 quintales por hectárea, respectivamente. El tercer híbrido considerado es el maíz Vencedor R-8330, el cual cuenta con grandes cualidades agronómicas y tiene un rendimiento aproximado de 165 quintales por hectárea con un período de siembra-cosecha de 120 días [15, 50]. De este modo, se establecieron los límites de especificación para los diferentes tipos de maíz, obteniendo los siguientes valores:

Tipo de maíz	USL	LSL
INIAP H-552	39,2 MJ/m ²	19,6 MJ/m ²
INIAP H-551	43,8 MJ/m ²	21,9 MJ/m ²
Vencedor R-8330	48,2 MJ/m ²	24,1 MJ/m ²

Tabla 3.3: Límites de especificación para la PAR. **Elaboración:** Autor.

3.3. Curvas $\mathcal{X}_{0,99865}$, $\mathcal{X}_{0,00135}$, $\mathcal{X}_{0,5}$

Como se puede observar en la subsección 2.2.2, para la aplicación de los índices del método A y B; se requiere el cálculo de las curvas $\mathcal{X}_{0,99865}$, $\mathcal{X}_{0,00135}$ y $\mathcal{X}_{0,5}$. En primera instancia, es necesario obtener la representación funcional de los datos, para luego aplicar los métodos descritos como el cálculo de la curva más profunda, la región central y de esta manera obtener las curvas mencionadas.

3.3.1. Suavizamiento de datos

La reconstrucción de las curvas a partir de los datos discretizados de cada variable, se lo realizará mediante el uso de bases de Fourier, ya que la variable climática tienen un comportamiento periódico. Para esto se hace uso del software estadístico R³ [53] y del paquete `fda.usc` [17], el cual contiene una gran diversidad de funciones que permiten el tratamiento y análisis de datos funcionales.

Como se revisó en la sección 2.1.2.3, es necesario establecer el número óptimo K de bases a utilizar, esto se lo realiza mediante el criterio de validación cruzada generalizada (GCV) definido por la fórmula 2.12. El paquete posee la función `optim.basis`, la cual recibe como parámetros un vector de valores K de bases y un vector de valores de penalización λ , dando como resultado el número de bases K y de penalización λ óptimos que minimiza el GCV.

Se aplicaron diferentes criterios para el uso de estos índices, teniendo mejor resultado al aplicar un ajuste sobre todas las curvas y luego tomando aquellas curvas necesarias; los otros escenarios son descritos en la parte de anexos, al igual que los distintos valores para el GCV con sus respectivos pares (K, λ) .

3.3.1.1. Suavizamiento de la precipitación

Aplicando la función `optim.basis` a los datos de precipitación, se obtuvo que el GCV mínimo es 385,27, que se logra con un número de $K = 10$ bases y $\lambda = 7,2457$. En la figura 3.1, se muestra tanto los datos discretos como los datos funcionales de la variable precipitación.

³R es un entorno de software libre orientado en el análisis estadístico de datos

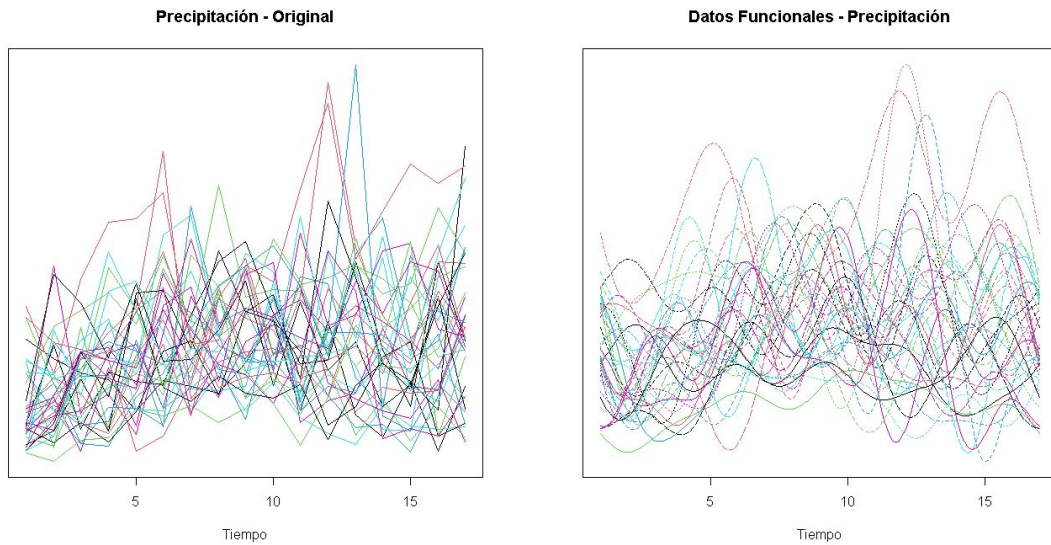


Figura 3.1: Datos discretos y Datos funcionales para la precipitación. **Elaboración:** Autor.

Las curvas funcionales presentadas a la derecha de la figura 3.1 son utilizadas para el cálculo de la región central de la precipitación.

3.3.1.2. Suavizamiento de los tiempos térmicos

Al igual que la precipitación, se hace uso de la función `optim.basis`, obteniendo para el caso de tiempos térmicos un valor GCV mínimo de 8,58, el cual se alcanza para un número de $K = 11$ bases y $\lambda = 2,43$, obteniendo los siguientes resultados:

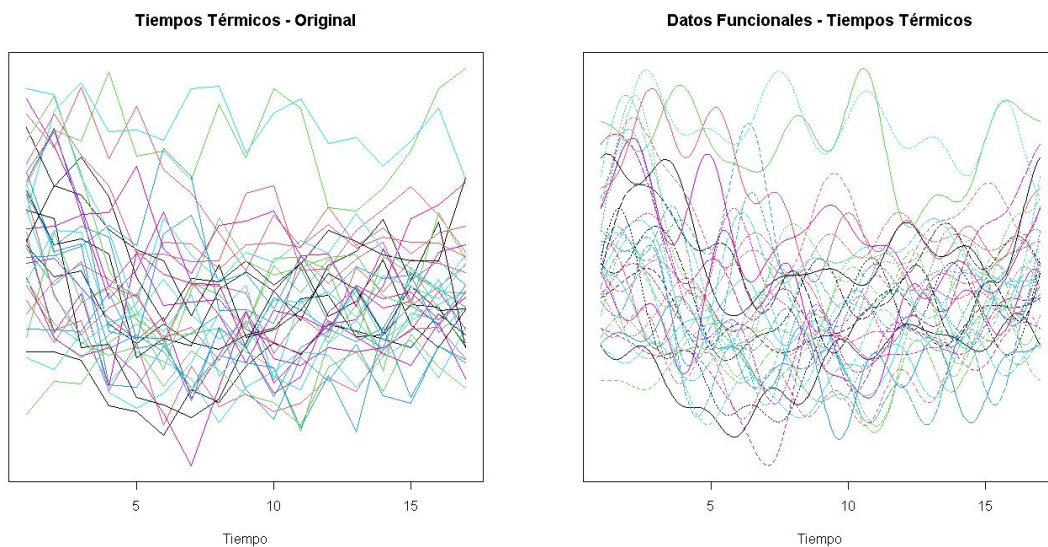


Figura 3.2: Curvas originales y ajustadas para los tiempos térmicos

Como se observa, a la derecha de la figura 3.2 se tienen los datos funcionales, los

cuales se usarán para obtener la región central de los tiempos térmicos.

3.3.1.3. Suavizamiento de la radiación solar fotosintéticamente activa PAR

Finalmente, se realizó el suavizamiento de la radiación solar fotosintéticamente activa y se obtuvo un GCV mínimo de 24,51, con un número de $K = 11$ bases y $\lambda = 122,68$ (Figura 3.3).

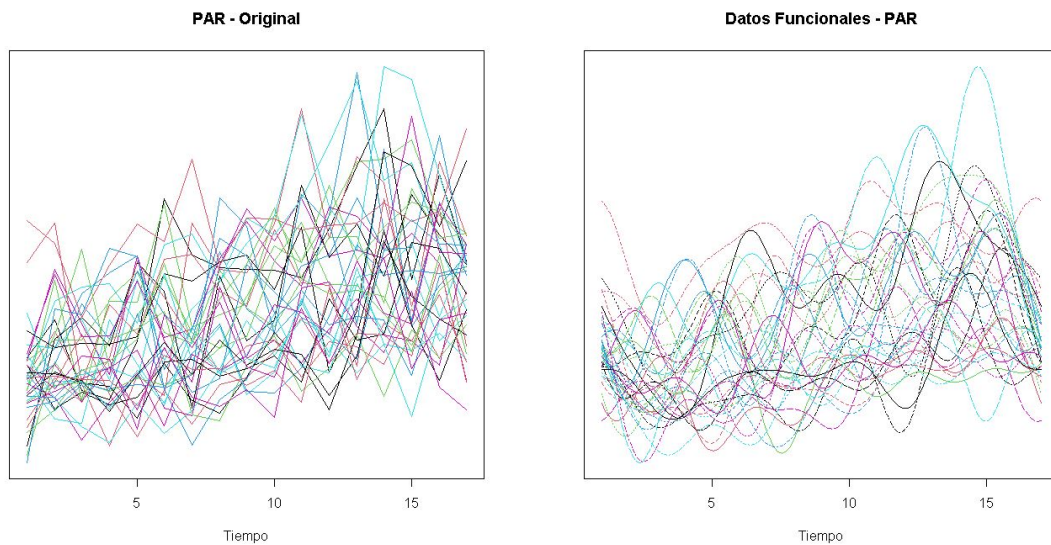


Figura 3.3: Curvas originales y ajustadas para la PAR

Por tanto, se obtienen los datos funcionales para la PAR a la derecha de la figura 3.3, que servirán para el cálculo de la región central de esta variable.

3.3.2. Cálculo de la región central

El cálculo de la región central permite obtener las curvas $\mathcal{X}_{0,99865}$ y $\mathcal{X}_{0,00135}$, que representan las curvas percentil 0,99865 y 0,00135 respectivamente, y son aquellas curvas en donde varía el proceso. Para esto, se hace uso del método descrito por la ecuación (2.19), utilizando las curvas muestrales de cada variable. En este caso, se considera el valor de $n = 36$ y $p = 0,9973$, ya que el 99,73% de las curvas se encuentran delimitadas por $\mathcal{X}_{0,99865}$ y $\mathcal{X}_{0,00135}$.

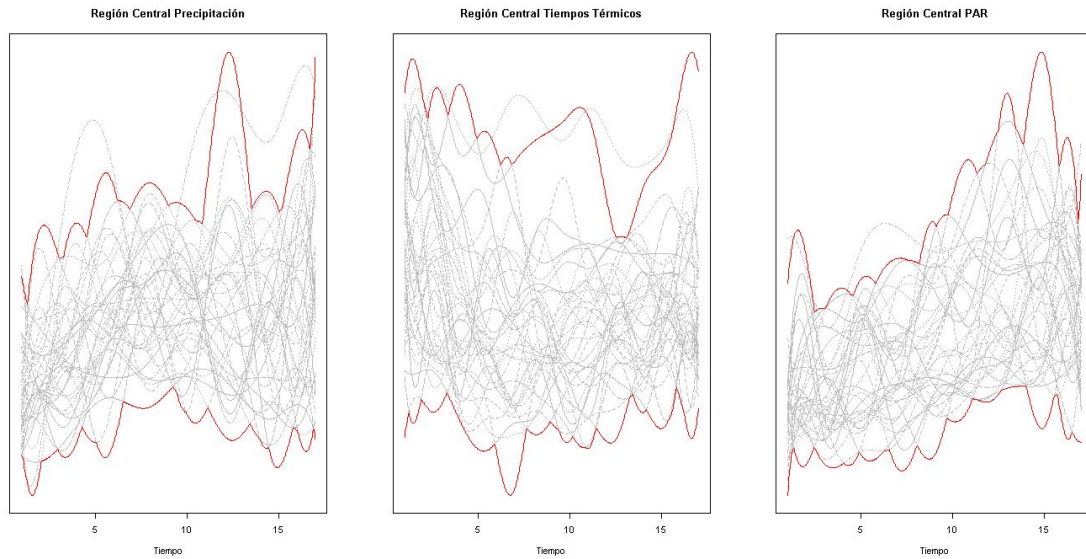


Figura 3.4: Región Central para Precipitación, Tiempos térmicos y PAR. **Elaboración:** Autor.

Una vez calculada la región central para cada variable, se extraen las curvas $\mathcal{X}_{0,99865}$ y $\mathcal{X}_{0,00135}$, que como se observa en la figura 3.4, son las curvas representadas por líneas de color rojo que serán utilizadas en el cálculo de los índices.

3.3.3. Curva más profunda

Para calcular la curva más profunda o curva $\mathcal{X}_{0,5}$, se utilizará la función `depth.FM` contenida en el paquete `fda.usc`. Esta medida de profundidad está descrita por la ecuación (2.16). Esta función recibe el conjunto de datos y permite ingresar el parámetro de recorte α . Se hará uso de esta medida de profundidad, ya que en el estudio realizado por Guevara y Vargas [22], se obtuvo mejores resultados. Al ser una medida robusta, evita problemas en el cálculo de los índices cuando la curva $\mathcal{X}_{0,5}$ coincide en algún punto con los límites de especificación.

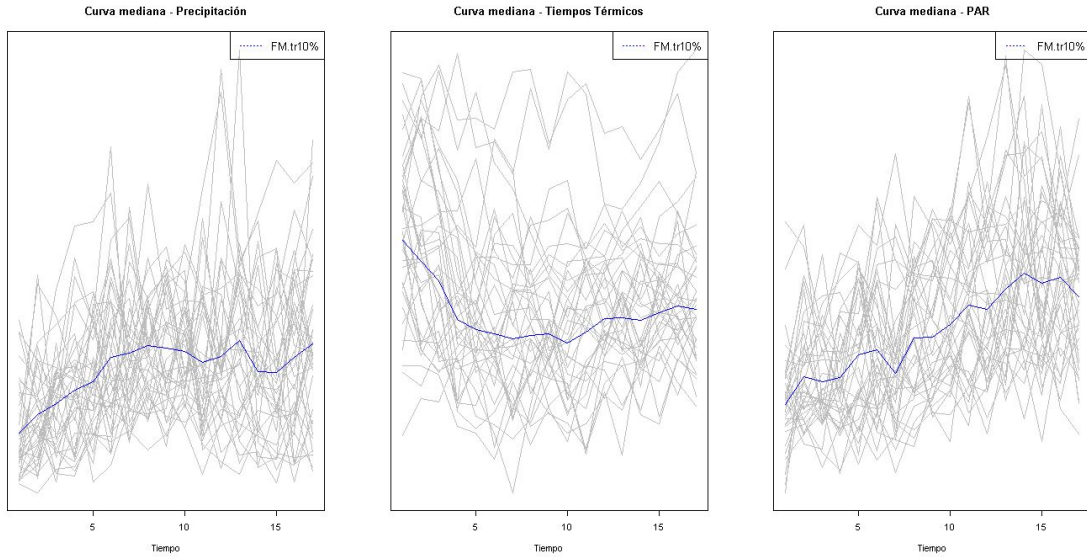


Figura 3.5: Curva media para Precipitación, Tiempos térmicos y PAR. **Elaboración:** Autor.

En la figura 3.5, la curva de color azul representa la media para el conjunto de curvas muestrales de cada variable, con un parámetro de recorte $\alpha = 0,1$. Como se menciona en [49], esta curva se obtiene sobre las curvas muestrales, sin representación en base, puesto que los resultados estarían condicionados por el error de la representación.

3.4. Aplicación y Resultados

Una vez se han fijado los límites de especificación y se ha realizado el cálculo de las curvas $\mathcal{X}_{0,99865}$, $\mathcal{X}_{0,00135}$, $\mathcal{X}_{0,5}$ para cada una de las variables de estudio, se procede a la aplicación de los índices definidos por las ecuaciones (2.28) y (2.35).

Adicionalmente, se incluyen los gráficos que permiten observar la relación entre la región del proceso, de especificación y la media recortada. Teniendo en cuenta que los límites de control $\mathcal{X}_{0,99865}$ y $\mathcal{X}_{0,00135}$ son las curvas de color rojo, los límites de especificación las curvas de color negro y la media de color azul.

3.4.1. Precipitación

Para la variable climática precipitación, se puede observar en la figura 3.6 la relación entre la región de especificación y la región del proceso. La región de proceso tiene una variabilidad más grande, se puede notar que se encuentra por fuera de la

región de especificación. Para entender si existe la capacidad de esta variable para la producción de maíz, se aplican los índices antes mencionados.

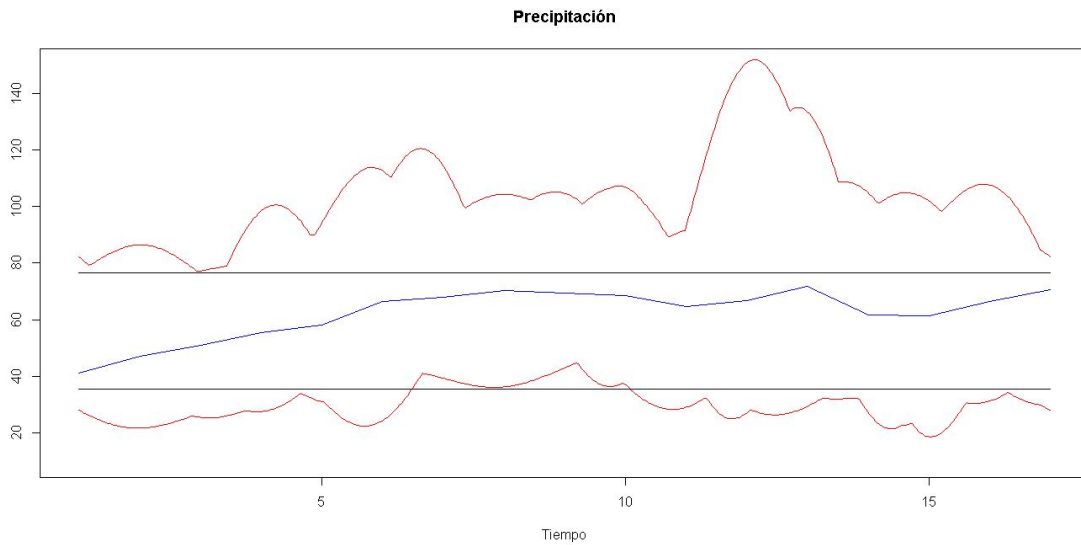


Figura 3.6: Región de especificación y región del proceso para la precipitación. **Elaboración:** Autor.

En la tabla 3.4, se presentan los valores obtenidos para los índices del método A y B, teniendo que el valor $C_{p(prf)}$ es menor a 0,67 para ambos métodos; sin embargo, se obtiene un mejor valor para este índice mediante el método A. De acuerdo con la tabla 2.1, el proceso se considera como pobre. Por otro lado, observando el valor del índice $C_{pk(prf)}$ se tiene que es menor a 1,33 para los dos métodos, con un mejor valor en el método A, evidenciando que no existe la capacidad de precipitación necesaria para la producción del maíz. Además, la media de la región del proceso no se encuentra centrada en los límites de especificación.

Método	$C_{p(prf)}$	$C_{pl(prf)}$	$C_{pu(prf)}$	$C_{pk(prf)}$
A	0,574	0,827	0,381	0,336
B	0,549	0,825	0,338	0,205

Tabla 3.4: Valores de los índices $C_{p(prf)}$ y $C_{pk(prf)}$ para los métodos A y B para la precipitación. **Elaboración:** Autor.

Analizando los límites $C_{pl(prf)}$ y $C_{pu(prf)}$, no se tiene la capacidad de precipitación, tanto mínima como máxima; es decir, a lo largo del ciclo, se tendría precipitación por debajo de los límites así como por encima de los mismos; sin embargo, se puede notar que el índice $C_{pu(prf)}$ es menor que el índice $C_{pl(prf)}$, teniendo una mayor posibilidad de precipitaciones por encima del nivel máximo requerido que

obtenerlas por debajo del mínimo requerido. Una posible solución cuando se obtiene niveles por debajo del requerido, es recurrir a riegos programados para cubrir la necesidad de agua; mientras que, cuando se tiene una precipitación mayor a la requerida, se procede al desecamiento del suelo [47].

3.4.2. Tiempos Térmicos

Para los tiempos térmicos se puede observar en la figura 3.5 que, la región del proceso es poco variable, puesto que las curvas $\mathcal{X}_{0,99865}$ y $\mathcal{X}_{0,00135}$ se encuentran relativamente cerca, mientras que la región de especificación es relativamente muy amplia. Además, se puede observar que la región del proceso está contenida completamente en la región de especificación.

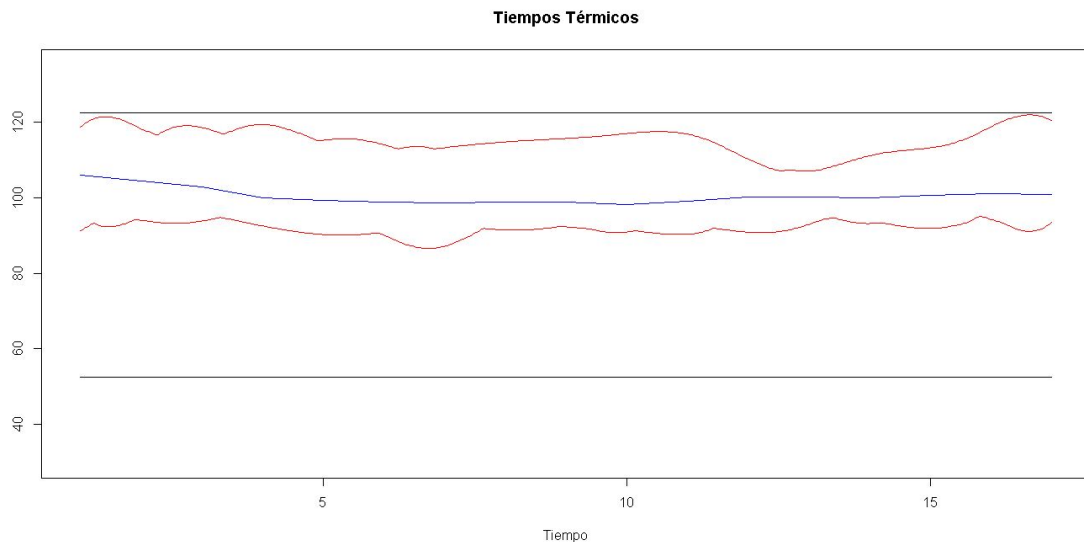


Figura 3.7: Región de especificación y región del proceso para los tiempos térmicos.
Elaboración: Autor.

Aplicando los índices de capacidad considerados, se puede observar en la tabla 3.5 que los valores del índice $C_{p(prf)}$ para los métodos A y B es mayor a 2, teniendo un valor más alto para el método A. Mediante lo descrito en la tabla 2.1, se tiene un proceso excelente, esto se debe a que la región de proceso es poco variable en comparación a la región de especificación.

Para el valor del índice $C_{pk(prf)}$, se tiene un valor mayor a 1,33 en ambos métodos, teniendo un mejor valor para el índice del método A, este valor indica que el proceso es capaz, es decir, existe la capacidad de los tiempos térmicos y no representan un problema para la producción del maíz, ya que se suplen las necesidades para el desarrollo óptimo del mismo.

Método	$C_{p(prf)}$	$C_{pl(prf)}$	$C_{pu(prf)}$	$C_{pk(prf)}$
A	3,094	5,721	1,604	1,604
B	3,004	5,591	1,505	1,505

Tabla 3.5: Valores de los índices $C_{p(prf)}$ y $C_{pk(prf)}$ para los métodos A y B para los tiempos térmicos. **Elaboración:** Autor.

Por otro lado, gráficamente se puede notar que el proceso no se encuentra centrado, lo cual se afirma observando que los índices $C_{p(prf)}$ y $C_{pk(prf)}$ tienen diferente valor. Revisando los valores de los índices $C_{pl(prf)}$ y $C_{pu(prf)}$, el valor del índice $C_{pl(prf)}$ es mayor al valor del índice $C_{pu(prf)}$, siendo más factible suplir la necesidad de tiempos térmicos mínimas a las máximas; aun así, al ser valores mayores a 1,33 se suplen ambas necesidades, como se indicó con los valores de los índices $C_{p(prf)}$ y $C_{pk(prf)}$. Cabe mencionar que al igual que los anteriores índices, el método A tiene mejores resultados que el método B.

3.4.3. Radiación solar fotosintéticamente activa - PAR

Para la radiación solar fotosintéticamente activa (PAR), se tienen distintos límites de especificación que se los define en función del nivel de producción del maíz. Como se observa en la figura 3.8, la región del proceso se encuentra por fuera de la región de especificación en los tres casos presentados, aunque se tienen relaciones distintas puesto que las regiones de especificación varían según el tipo de híbrido de maíz.

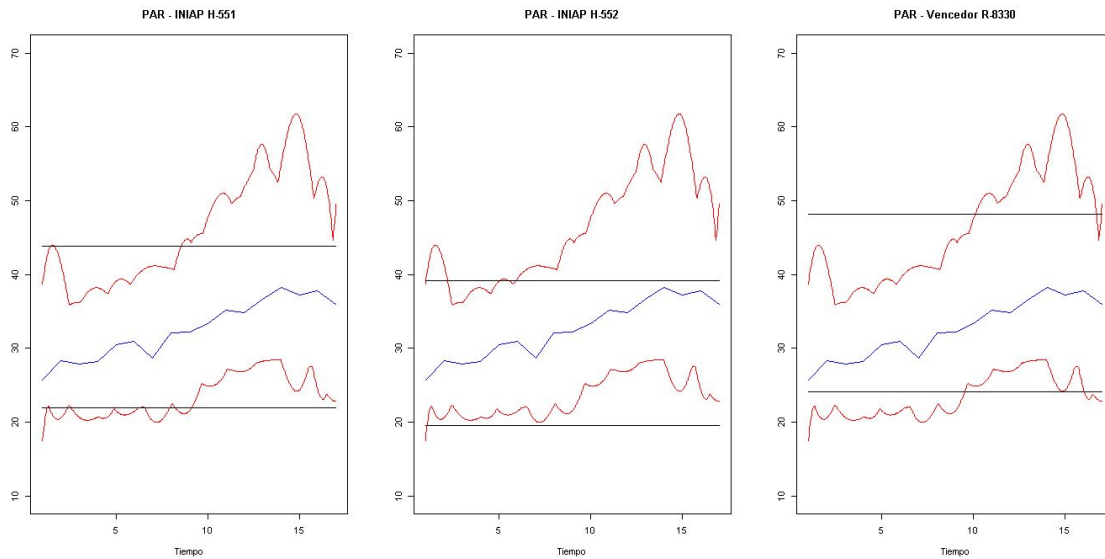


Figura 3.8: Región de especificación y región de proceso para la PAR y cada tipo de maíz **Elaboración:** Autor.

Mediante la aplicación de los métodos A y B, teniendo en cuenta los diferentes límites de especificación, se puede observar en la tabla 3.6 que el índice $C_{p(prf)}$ tiene mejores valores mediante el método A aunque la diferencia es baja respecto al método B; el valor de dicho índice para los híbridos INIAP H-551 y Vencedor R-8330 mayor a uno; mientras que, para el maíz INIAP H-552 es menor a uno. Utilizando la tabla 2.1 de valores para este índice, el proceso es capaz para los dos primeros híbridos mencionados e inadecuado para el tercero. Es decir, si tomamos en cuenta únicamente este índice, existe la capacidad de PAR para los híbridos INIAP H-551 y Vencedor R-8330.

Por otro lado, observando el valor del índice $C_{pk(prf)}$, para los tres híbridos de maíz es menor a 1,33 más aún menor a uno, teniendo en cuenta que el método B genera mejores resultados para los híbridos INIAP H-551 y Vencedor R-8330. Dado que los índices $C_{p(prf)}$ y $C_{pk(prf)}$ son distintos, se tiene que el proceso no se encuentra centrado, es decir, la radiación obtenida estaría por fuera de las especificaciones establecidas en ciertos períodos y por tanto no existe la capacidad para suplir las necesidades del maíz respecto a esta variable climática.

Maíz	Método	$C_{p(prf)}$	$C_{pl(prf)}$	$C_{pu(prf)}$	$C_{pk(prf)}$
INIAP H-551	A	1,006	1,113	0,988	0,732
	B	0,951	1,116	0,840	0,840
INIAP H-552	A	0,900	1,369	0,615	0,590
	B	0,851	1,364	0,505	0,505
Vencedor R-8330	A	1,107	0,868	1,344	0,693
	B	1,047	0,879	1,160	0,879

Tabla 3.6: Valores de los índices $C_{p(prf)}$ y $C_{pk(prf)}$ para los métodos A y B para la PAR. **Elaboración:** Autor.

Al igual que con las otras variables, revisando los valores de los índices $C_{pl(prf)}$ y $C_{pu(prf)}$, se tiene que para los híbridos INIAP H-551 e INIAP H-552 el valor $C_{pl(prf)}$ es mayor que el valor del índice $C_{pu(prf)}$; por lo que, para estos dos tipos de híbrido de maíz se tiene la capacidad de cumplir con el requerimiento mínimo de PAR; sin embargo, existe una mayor posibilidad que la radiación obtenida supere el límite de especificación superior, teniendo un exceso de radiación. Para el híbrido Vencedor R-8330 el escenario es el contrario, el valor del índice $C_{pl(prf)}$ es menor al valor del índice $C_{pu(prf)}$, es decir, existe una mayor posibilidad de que la radiación no alcance el requerimiento mínimo a que supere el requerimiento máximo.

De acuerdo a los resultados obtenidos en esta investigación, al considerar la variable clima como un proceso, permite tener una visión general de como varia respecto a las necesidades climatológicas óptimas para la producción de maíz. Como ya se mencionó anteriormente, cultivar bajo condiciones poco óptimas afecta al rendimiento final de producción, por lo que se debe tomar acciones para suplir las necesidades, o como una opción alterna considerar otros lugares donde las condiciones sean mejores.

Capítulo 4

Conclusiones y Recomendaciones

En el presente trabajo de investigación se aplicó una extensión de los índices de capacidad de proceso tradicionales C_p , C_{pu} , C_{pl} y C_{pk} hacía datos funcionales, en donde se consideró dos métodos para su cálculo; el método A, que se basa en un promedio de los índices tradicionales evaluados en cada punto de observación t_j ; mientras el método B, aprovecha la continuidad de las funciones. Esta aplicación se la hizo considerando las variables climáticas: tiempos térmicos, precipitación y radiación solar fotosintéticamente activa (PAR), con la finalidad de evaluar la capacidad climática para la producción de maíz duro seco, el cual, es el segundo cultivo transitorio con mayor producción en el Ecuador. Para este caso, se consideró la zona de La Concordia ubicada en la provincia de Santo Domingo de los Tsáchilas. Se tomaron datos climáticos de esta zona desde el año 1984 hasta el año 2020, tomando el tiempo entre la siembra y la cosecha de 120 días, los cuales, fueron distribuidos en 17 semanas iniciando el 15 de diciembre, obteniendo un total de 36 períodos.

El estudio de las variables orientado al análisis de datos funcionales (FDA) permite reducir la dimensionalidad de los datos recolectados, es así que, se construyó una forma funcional, de tal manera que cada período se representó como un solo dato. Esta representación, se la realizó mediante el suavizamiento de los datos discretos de la base original obtenida, para lo cual se hizo uso de bases de Fourier. Aunque no existe un método general que ayude a determinar el número de bases óptimo, se pudo concluir que el método de validación cruzada generalizada solventa este problema, y aumentando una penalización por rugosidad λ , se obtuvo un método más robusto. Para construir la mejor representación funcional de la variable precipitación se tomó $K = 10$ bases funcionales de Fourier y un valor de penalización $\lambda = 7,064$, donde se obtiene una estabilidad en el suavizamiento con un

valor para el método GCV de 385,269. Para la variable tiempos térmicos, se obtiene una estabilidad en el suavizamiento con un valor de GCV igual a 8,580, el cual se obtiene considerando $K = 11$ bases funcionales de Fourier y un valor de penalización $\lambda = 2,431$. Finalmente, para la variable radiación solar fotosintéticamente activa se consideró $K = 11$ bases funcionales de Fourier y un valor de penalización $\lambda = 121,354$, alcanzando una estabilidad en el suavizamiento con un valor GCV igual a 24,520.

Se consideró un proceso estable para cada variable, por esta razón se aplicaron directamente los índices de capacidad. Se realizó el cálculo de los límites de especificación. Para las variables precipitación y tiempos térmicos, el límite superior (USL) e inferior (LSL) son los mismos para cada híbrido de maíz considerado, ya que estos requerimientos son generales para el maíz duro seco en tierras bajas. Para el caso de la variable PAR, las necesidades varían para cada híbrido de maíz, por lo que se consideraron distintos límites de especificación. Definidos los límites de especificación, se pudo obtener la región de especificación, la cual es una parte importante para el cálculo de los índices.

Para obtener la región de proceso, se usó el método de Clements que se basa en percentiles, que resultó ser de gran utilidad en este trabajo, ya que este método no exige una distribución específica de los datos. Así, la región de proceso está limitada por las curvas percentil 0,99865 y 0,00135, las cuales se obtienen mediante el cálculo de la región central. Por otro lado, existen diversos métodos para el cálculo de la curva más profunda, en esta investigación se hizo uso de la media recortada, pues al ser una medida robusta ayuda a evitar problemas cuando la mediana se aproxima demasiado a los límites de especificación.

Mediante la aplicación de los índices de capacidad de proceso funcionales a los datos de precipitación, se concluye que el método A obtuvo mejores resultados, además, las necesidades hídricas de los tres híbridos de maíz no se satisfacen, es decir, no existe la capacidad de precipitación, ya que el índice $C_{p(prf)}$ tiene un valor 0,574, describiendo un proceso inadecuado de acuerdo a lo descrito en la tabla 2.1. Por otro lado, el valor para el índice $C_{pk(prf)}$ es 0,336, y por tanto, es menor a 1,33, afirmando que no se tiene la capacidad de precipitación. Así mismo, se concluye que es un proceso no centrado, ya que los índices $C_{p(prf)}$ y $C_{pk(prf)}$ son distintos, por lo que la precipitación obtenida estaría por fuera de los límites de especificación en ciertas semanas. Analizando los límites unilaterales, se pudo observar que hay una mayor posibilidad de que los valores de precipitación estén por encima de lo requerido, ya que se tiene la desigualdad $C_{pu(prf)} < C_{pl(prf)}$. Para solventar este pro-

blema, se requiere aplicar técnicas para realizar un desecamiento del suelo, aun así, la posibilidad de tener períodos con valores de precipitación por debajo del mínimo requerido es alta, por lo que también sería necesario recurrir a riegos para suplir las necesidades hídricas.

Para los tiempos térmicos, el método A entregó mejores resultados. Se pudo observar que las necesidades de esta variable se satisfacen para los tres híbridos de maíz, es decir, existe la capacidad de tiempos térmicos, ya que el índice $C_{p(prf)}$ tiene un valor de 3,094, que según la tabla 2.1, describe un proceso excelente. Por otra parte, el índice $C_{pk(prf)}$ tiene un valor de 1,604, lo cual afirma que el proceso es capaz. En este caso, el proceso no está centrado, ya que el índice $C_{p(prf)}$ no es igual al índice $C_{pk(prf)}$, sin embargo la región de proceso se encuentra completamente contenida en la región de especificación, es decir, que los datos de tiempos térmicos se encuentran dentro de los límites de especificación.

En el caso de la radiación solar fotosintéticamente activa, se obtuvo mejores resultados con el método A. Se pudo observar, en general, que no se satisfacen las necesidades de esta variable para los híbridos de maíz, es decir, no existe capacidad de radiación solar fotosintéticamente activa. Para los híbridos de maíz INIAP H-551 y Vencedor R-8830, el índice $C_{p(prf)}$ obtuvo los valores de 1,006 y 1,107 respectivamente, describiendo un proceso capaz según la tabla 2.1. Sin embargo, al complementar con el índice $C_{pk(prf)}$, el cual obtuvo los valores de 0,732 y 0,693 respectivamente, se observa que no existe capacidad, ya que son valores menores a 1,33. Además, el proceso no está centrado, puesto que los índices $C_{p(prf)}$ y $C_{pk(prf)}$ no coinciden. Por otro lado, para el híbrido de maíz INIAP H-552, el índice $C_{p(prf)}$ obtuvo un valor de 0,900, el cual describe un proceso inadecuado, es decir no existe capacidad, lo cual se complementa con el valor del índice $C_{pk(prf)}$, ya que tiene un valor de 0,693 y es menor a 1,33. Además, al igual que los híbridos de maíz INIAP H-551 y R-8830, el proceso no está centrado. Analizando los límites unilaterales, en el caso de los de los híbridos de maíz INIAP H-551 e INIAP H-552, hay más posibilidad que la radiación solar fotosintéticamente activa este por encima de lo requerido, mientras que para el híbrido R-8830, hay más posibilidad que este por debajo de lo requerido, ya que se tiene la desigualdad $C_{pu(prf)} > C_{pl(prf)}$. Un exceso de radiación puede causar daños en el desarrollo del maíz; una posible solución para evitar este problema es la colocación de invernaderos, aunque puede resultar muy costoso.

Con los resultados obtenidos, la capacidad climática en la zona La Concordia no resulta ser óptima para el cultivo de maíz; bajo estas condiciones, el rendimiento del maíz puede resultar muy por debajo del que se esperaría de acuerdo al potencial

de cada híbrido; sin embargo, se puede realizar acciones que permitan incrementar dicho rendimiento.

Finalmente, con el propósito de extender este criterio de uso de índices de capacidad a la evaluación de capacidad climática, se presentan las siguientes recomendaciones:

- Como se mencionó anteriormente, las necesidades hídricas para el maíz varían dependiendo de la fase de desarrollo en la que se encuentra, por lo que se recomienda la construcción de los límites de especificación como una función escalonada que permita tener una evaluación más cercana a las necesidades.
- En los casos prácticos, el uso de índices de capacidad de proceso univariados otorga resultados limitados al evaluar cada característica por separado. El uso de índices de capacidad multivariados es una opción más robusta y por tanto más recomendable, pues considera todas las características a la vez, ayudando a entender de manera conjunta la capacidad climática para la producción de maíz.

Bibliografía

- [1] Akritas, M. y Politis, D. *Recent Advances and Trends in Nonparametric Statistics*. Elsevier, Ámsterdam, 1era edición, 2003.
- [2] Birch, C., Hammer, G., y Rickert, K. Dry matter accumulation and distribution in five cultivars of maize (*Zea mays*): relationships and procedures for use in crop modeling. *The Australian Journal of Agricultural and Resource Economics*, 50(5):13–527, 1999.
- [3] Bothe, D. Composite capability index for multiple product characteristics. *Quality Engineering*, 12(2):253–258, 1999.
- [4] Box, G. y Cox, D. An analysis of transformation. *Journal Royal Statistical Society B*, 26:211–243, 1964.
- [5] Cascos, I., López, A., y Romo, J. Data depth in multivariate statistics. *Boletín de Estadística e Investigación Operativa*, 27(3):151–174, 2011.
- [6] Chan, L., Xiong, Z., y Zhang, D. On the asymptotic distributions of some process capability indices. *Communications in Statistics: Theory and Methods*, 19(1):11–18, 1990.
- [7] Chen, K., Chen, H., y Chang, T. The construction and application of six sigma quality indices. *International Journal of Production Research*, 55:2365–2384, 2016.
- [8] Clements, J. Process capability calculations for non-normal distributions. *Quality Progress*, 22:95–97, 1989.
- [9] Cuevas, A., Febrero-Bande, M., y Fraiman, R. On the use of the bootstrap for estimating functions with functional data. *Computational Statistics & Data Analysis*, 51:1063 – 1074, 2006.

- [10] Cuevas, A., Febrero-Bande, M., y Fraiman, R. Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22:481 – 496, 2007.
- [11] de Boor, C. On calculating with b-splines. *Journal of approximation theory*, 6(1):50–62, 1972.
- [12] de Boor, C. *A Practical Guide to Splines*. Springer, Estados Unidos, 2001.
- [13] de Felipe, D. y Benedito, E. A review of univariate and multivariate process capability indices. *The International Journal of Advanced Manufacturing Technology*, 92:1687–1705, 2017.
- [14] Dyckerhoff, R. Data depths satisfying the projection property. *Allgemeines Statistisches Archiv*, 88(2):163–190, 2004.
- [15] Díaz, G., Sabando, F., Ávila, S., y Váscenez, G. Evaluación productiva y calidad del grano de cinco Híbridos de Maíz (*Zea mays* L.) en dos localidades de la provincia de Los Ríos. Tesis de grado, Universidad Técnica Estatal de Quevedo., Ecuador, 2009. Repositorio Institucional de la Universidad Técnica Estatal de Quevedo. http://www.uteq.edu.ec/revista_cyt/archivos/2009/v2_013/articulo_3.pdf.
- [16] Estrada, A. Nociones de profundidad en análisis de datos funcionales. Tesis de grado, Universidad de los Andes, Colombia, 2014. Repositorio Institucional-Universidad de los Andes.
- [17] Febrero-Bande, M. y Oviedo de la Fuente, M. fda.usc: Functional Data Analysis and Utilities for Statistical Computing. R package version 2.0.2, 2020. URL: <https://cran.r-project.org/web/packages/fda.usc/fda.usc.pdf>.
- [18] Febrero-Bande, M. y Oviedo de la Fuente, M. Statistical computing in functional data analysis: The r package fda.usc. *Journal of Statistical Software*, 51(4):1–28, 2012.
- [19] Ferraty, J. y Vieu, P. *Nonparametric Functional Data Analysis*. Springer, Estados Unidos, 2006.
- [20] Flores, M. *Nuevas Aportaciones del Análisis de Datos Funcionales en el Control Estadístico de Procesos*. Tesis de doctorado, Universidade da Coruña, España, 2018. Repositorio Institucional-Universidade da Coruña.

- [21] Fraiman, R. and Muniz, G. Trimmed means for functional data. *Test*, 10(2):419–440, 2001.
- [22] Guevara, R. y Vargas, J. Process capability analysis for nonlinear profiles using depth functions. *Quality and Reliability Engineering International.*, 31:465–487, 2015.
- [23] Gunter, B. The use and abuse of cpk. *Quality Progress*, 22(1):72–73, 1989.
- [24] Instituto Nacional de Estadísticas y Censos. Encuesta de superficie y producción agropecuaria continua. Reporte técnico, Quito, 2020.
- [25] Intergovernmental Panel on Climate Change (IPCC). Cambio climático. informe técnico V, Ginebra, 2014.
- [26] Johnson, N. System of frequency curves generated by methods of translation. *Biometrika*, 36:149–176, 1949.
- [27] Kane, V. Process capability indices. *Journal of Quality Technology*, 18:41–52, 1986.
- [28] Khadse, K. y Khadse, A. On properties of probability-based multivariate process capability indices. *Quality and Reliability Engineering International*, 36(5):1768–1785, 2020.
- [29] Leng, M. y Aldrich, S. Producción moderna del maíz. *Agencia para el desarrollo internacional (AID)*. El Batán, México.
- [30] Liu, R. On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414, 1990.
- [31] Liu, R., Parelius, J., y Singh, K. Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics*, 27(3):783–858, 1999.
- [32] Liu, R. y Singh, K. A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88:252–260, 1995.
- [33] López, A. y Hernández, D. Cambio climático y agricultura: una revisión de la literatura con énfasis en América Latina. *El Trimestre Económico*, 84(4):459–496, 2016.
- [34] López-Pintado, S. y Romo, J. Depth-based inference for functional data. *Computational Statistics & Data Analysis*, 51:4957–4968, 2007.

- [35] López-Pintado, S. y Romo, J. On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718 – 734, 2009.
- [36] Mahalanobis, P. On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2:49–55, 1936.
- [37] Ministerio de Agricultura (MAG). Levantamiento de cartografía temática escala 1:25.000, lote 1. Memoria técnica, Quito, 2015.
- [38] Ministerio del Ambiente (MAE). *Segunda Comunicación del Ecuador al UFGC*. Quito, 2010.
- [39] Montgomery, D. *Introduction to Statistical Quality Control*. John Wiley & Sons, New York, 3rd edición, 1996.
- [40] Müller, HG. *Functional data analysis*. In *International Encyclopedia of Statistical Science*. Miodrag, L., 554–555. Heidelberg, Ger.: Springer, 2011.
- [41] Nelson, G., Rosegrant, M., Koo, J., Robertson, R., Sulser, T., Zhu, T., Ringler, C., Msangi, S., Palazzo, A., Batka, M., Magalhae, M., Valmonte, R., Ewing, M., y y Lee, D. Cambio Climático: El impacto en la agricultura y los costos de adaptación., 2009.
- [42] Norman, J., Pearson, C., y Searle, P. *The ecology of tropical food crops*. Cambridge University Press, New York, USA, 1995.
- [43] Orozco, J. Evaluación bioagronómica de una variedad y cinco híbridos de maíz duro (*Zea mays* L.), en el sector La Colombina, cantón Alausí. Tesis de grado, Escuela Superior Politécnica de Chimborazo, Ecuador, 2010. Repositorio Institucional de la Escuela Superior Politécnica de Chimborazo. <http://dspace.esPOCH.edu.ec/bitstream/123456789/581/1/13T0665%20.pdf>.
- [44] Otegui, M., Andrade, F., y Suero, E. Growth, water use, and kernel abortion of maize subjected to drought at silking. *Field Crops Research*, 40(2):87–94, 1995.
- [45] Pearn, W. y Kotz, S. *Encyclopedia and Handbook of Process Capability Indices*, volume 12. World Scientific Publishing, USA, 2006.
- [46] Pérez, P. Análisis de datos funcionales: representación en bases y regresión funcional en scikit-fda. Tesis de grado, Universidad Autónoma de Madrid., España, 2019. Repositorio Institucional de la Universidad Autónoma de Madrid. <https://repositorio.uam.es/handle/10486/688860>.

- [47] Pérez, P., Rodríguez, J., Guevara, D., Mompie, E., y Sánchez, M. Métodos y modos de desecamiento de terrenos húmedos por excesivas precipitaciones y acumulación de agua. *Cultivos Tropicales*, 39(3):93–98, 2018.
- [48] Raffo, D. La radiación solar y las plantas: un delicado equilibrio. *Fruticultura & Diversificación*, 74:41–44, 2014.
- [49] Ramsay, J. y Silverman, B. *Functional Data Analysis*. Springer Verlag, New York, 2da edición, 2005.
- [50] Segura, M. y Andrade, L. Efecto de las condiciones agrometeorológicas sobre un cultivar criollo y dos híbridos de maíz en cuatro fechas de siembra. Tesis de grado, Escuela Politécnica del Ejército, Ecuador, 2011. Repositorio Institucional de la Escuela Politécnica del Ejército <http://repositorio.espe.edu.ec/xmlui/handle/21000/3827>.
- [51] Snyder, R. Hand calculating degree days. *Agricultural and Forest Meteorology*, 35(1):353–358, 1985.
- [52] Somerville, S. y Montgomery, D. Process capability indices and non-normal distributions. *Quality Engineering*, 19(2):305–316, 1996.
- [53] Team, R. C. R: A language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2019. URL: <https://www.R-project.org/>.
- [54] Tsai, C. y Chen, C. Making decision to evaluate process capability index cp with fuzzy numbers. *The International Journal of Advanced Manufacturing Technology*, 30:334–339, 2006.
- [55] Tukey, J. Mathematics and the picturing of data. *Proceedings of the international congress of mathematicians*, 2:523–531, 1975.
- [56] Villalpando, J. y Ruíz, J. *Observaciones agrometeorológicas y su uso en la agricultura*. Limusa, México, 1993.
- [57] Yepes, A. y Silveira, M. Respuestas de las plantas ante los factores ambientales del cambio climático global. *Colombia Forestal*, 14(2):213–232, 2011.
- [58] Zuo, Y. y Serfling, R. General notions of statistical depth function. *Annals of Statistics*, 28(2):461–482, 2000.

Anexos

Anexo A

Tablas de valores GCV para los pares (K, λ)

K/λ	0,0010	0,0014	0,002	0,003	0,004	0,006	0,008	0,012	0,017
5	441,669	441,657	441,640	441,615	441,580	441,530	441,459	441,357	441,213
6	488,982	488,931	488,858	488,754	488,606	488,395	488,097	487,674	487,077
7	512,402	512,243	512,016	511,694	511,238	510,593	509,686	508,416	506,653
8	585,345	584,884	584,230	583,306	582,006	580,188	577,666	574,208	569,539
9	653,195	652,029	650,386	648,085	644,885	640,483	634,513	626,575	616,293
10	702,946	700,318	696,658	691,611	684,746	675,578	663,637	648,579	630,348
11	827,631	821,638	813,447	802,437	787,973	769,530	746,897	720,392	690,973
12	893,712	881,254	864,787	843,632	817,439	786,471	751,801	715,244	678,899
13	1099,215	1069,583	1032,582	988,381	938,324	884,933	831,346	780,324	733,373
14	1397,595	1324,158	1241,052	1152,550	1064,034	980,349	904,302	836,260	775,007
15	1831,813	1649,909	1476,784	1321,078	1186,235	1070,878	971,064	882,839	803,827
K/λ	0,024	0,034	0,049	0,070	0,099	0,142	0,202	0,288	0,411
5	441,009	440,719	440,312	439,740	438,943	437,844	436,348	434,351	431,750
6	486,238	485,067	483,447	481,231	478,253	474,341	469,353	463,235	456,083
7	504,232	500,959	496,626	491,046	484,120	475,904	466,682	456,951	447,318
8	563,369	555,443	545,630	534,023	521,015	507,275	493,580	480,569	468,552
9	603,417	587,959	570,307	551,234	531,749	512,810	495,083	478,870	464,223
10	609,330	586,397	562,757	539,626	517,910	498,073	480,263	464,511	450,833
11	660,114	629,402	600,040	572,594	547,118	523,513	501,821	482,263	465,086
12	644,465	612,756	583,743	557,027	532,335	509,709	489,389	471,590	456,361
13	690,588	651,261	614,713	580,779	549,748	522,030	497,879	477,302	460,096
14	719,132	667,982	621,775	581,078	546,210	516,982	492,806	472,945	456,693
15	733,460	672,177	620,328	577,526	542,653	514,267	491,024	471,876	456,073
K/λ	0,587	0,837	1,195	1,704	2,432	3,470	4,951	7,064	10,079
5	428,472	424,514	419,974	415,074	410,125	405,457	401,342	397,944	395,304
6	448,185	439,998	432,053	424,800	418,479	413,082	408,426	404,288	400,526
7	438,313	430,229	423,086	416,734	411,009	405,840	401,251	397,302	394,022
8	457,505	447,235	437,589	428,562	420,285	412,931	406,621	401,378	397,133
9	451,116	439,522	429,406	420,700	413,298	407,081	401,934	397,737	394,364
10	439,202	429,470	421,367	414,569	408,798	403,870	399,683	385,269	393,286
11	450,392	438,062	427,808	419,280	412,161	406,221	401,300	397,273	394,025
12	443,560	432,896	424,008	416,544	410,227	404,872	400,369	396,635	393,591
13	445,913	434,311	424,814	416,974	410,438	404,964	400,400	396,638	393,582
14	443,426	432,582	423,656	416,222	409,960	404,664	400,213	396,522	393,511
15	443,052	432,332	423,471	416,075	409,842	404,571	400,142	396,468	393,471
K/λ	14,382	20,520	29,279	41,776	59,608	85,051	121,354	173,152	247,060
5	393,343	391,882	390,703	389,629	388,611	387,742	387,213	387,200	387,763
6	397,123	394,138	391,629	389,602	388,041	386,966	386,446	386,550	387,261
7	391,396	389,364	387,833	386,698	385,882	385,391	385,310	385,732	386,671
8	393,772	391,167	389,182	387,688	386,594	385,894	385,658	385,969	386,830
9	391,698	389,630	388,053	386,864	385,997	385,463	385,350	385,749	386,674
10	390,956	389,119	387,701	386,621	385,829	385,347	396,175	385,693	386,635
11	391,451	389,450	387,922	386,769	385,928	385,414	385,315	385,724	386,656
12	391,156	389,250	387,787	386,677	385,866	385,372	385,286	385,705	386,643
13	391,144	389,238	387,776	386,669	385,860	385,367	385,283	385,702	386,641
14	391,099	389,210	387,759	386,659	385,853	385,363	385,280	385,701	386,640
15	391,071	389,190	387,745	386,648	385,846	385,358	385,276	385,698	386,638
K/λ	352,514	502,981	717,671	1024					
5	388,814	390,156	391,577	392,910					
6	388,447	389,899	391,401	392,790					
7	388,020	389,590	391,176	392,628					
8	388,127	389,660	391,223	392,658					
9	388,015	389,581	391,167	392,619					
10	387,988	389,562	391,153	392,609					
11	388,002	389,571	391,160	392,614					
12	387,993	389,565	391,156	392,611					
13	387,992	389,564	391,155	392,610					
14	387,991	389,564	391,155	392,610					
15	387,990	389,563	391,154	392,610					

Tabla A.1: Valores GCV para la precipitación. **Elaboración:** Autor.

K/λ	0,0010	0,0014	0,002	0,003	0,004	0,006	0,008	0,012	0,017
5	10,126	10,125	10,125	10,124	10,124	10,123	10,121	10,119	10,116
6	10,319	10,318	10,316	10,314	10,311	10,307	10,300	10,291	10,279
7	10,417	10,414	10,409	10,403	10,393	10,380	10,361	10,335	10,299
8	10,968	10,959	10,947	10,930	10,905	10,871	10,824	10,760	10,674
9	12,329	12,307	12,277	12,234	12,174	12,093	11,983	11,838	11,653
10	13,723	13,672	13,601	13,504	13,372	13,196	12,969	12,685	12,346
11	14,815	14,708	14,562	14,367	14,112	13,789	13,396	12,940	12,442
12	17,893	17,642	17,309	16,880	16,346	15,711	14,993	14,227	13,458
13	20,178	19,629	18,941	18,116	17,176	16,167	15,147	14,172	13,276
14	23,099	21,920	20,599	19,211	17,845	16,574	15,437	14,425	13,509
15	27,248	24,986	22,929	21,160	19,670	18,378	17,186	16,018	14,852
K/λ	0,024	0,034	0,049	0,070	0,099	0,142	0,202	0,288	0,411
5	10,111	10,105	10,096	10,084	10,066	10,042	10,010	9,967	9,911
6	10,261	10,237	10,203	10,158	10,097	10,018	9,920	9,803	9,670
7	10,249	10,183	10,096	9,985	9,850	9,696	9,530	9,366	9,218
8	10,562	10,421	10,249	10,053	9,843	9,636	9,446	9,285	9,155
9	11,425	11,158	10,863	10,556	10,256	9,980	9,732	9,513	9,318
10	11,961	11,549	11,135	10,738	10,373	10,043	9,745	9,480	9,249
11	11,930	11,433	10,971	10,552	10,175	9,836	9,536	9,277	9,064
12	12,721	12,040	11,423	10,867	10,373	9,942	9,579	9,283	9,052
13	12,470	11,752	11,114	10,555	10,079	9,686	9,372	9,126	8,940
14	12,662	11,879	11,169	10,553	10,041	9,634	9,319	9,080	8,902
15	13,713	12,650	11,712	10,928	10,301	9,815	9,447	9,171	8,966
K/λ	0,5867	0,8372	1,1945	1,7044	2,4318	3,4698	4,9509	7,0641	10,0794
5	9,841	9,757	9,664	9,569	9,482	9,414	9,378	9,383	9,430
6	9,530	9,394	9,272	9,171	9,094	9,041	9,009	9,001	9,022
7	9,098	9,008	8,945	8,899	8,785	8,843	8,835	8,851	8,898
8	9,049	8,962	8,888	8,828	8,866	8,762	8,763	8,791	8,853
9	9,145	8,998	8,879	8,792	8,736	8,711	8,716	8,751	8,819
10	9,055	8,902	8,788	8,713	8,671	8,659	8,675	8,719	8,795
11	8,896	8,772	8,687	8,635	8,580	8,616	8,644	8,697	8,778
12	8,879	8,754	8,671	8,623	8,604	8,609	8,639	8,693	8,776
13	8,802	8,703	8,638	8,601	8,589	8,600	8,632	8,689	8,773
14	8,772	8,680	8,621	8,589	8,613	8,593	8,628	8,686	8,771
15	8,817	8,712	8,643	8,604	8,591	8,601	8,633	8,690	8,774
K/λ	14,382	20,520	29,279	41,776	59,608	85,051	121,354	173,152	247,060
5	9,517	9,635	9,777	9,939	10,131	10,367	10,662	11,021	11,431
6	9,080	9,180	9,328	9,523	9,767	10,067	10,426	10,843	11,301
7	8,985	9,114	9,287	9,502	9,760	10,067	10,430	10,847	11,304
8	8,951	9,089	9,268	9,487	9,748	10,058	10,422	10,841	11,299
9	8,923	9,067	9,250	9,473	9,738	10,050	10,416	10,837	11,296
10	8,905	9,053	9,240	9,466	9,732	10,046	10,413	10,834	11,294
11	8,893	9,045	9,234	9,461	9,729	10,043	10,412	10,833	11,294
12	8,892	9,044	9,233	9,461	9,729	10,043	10,412	10,833	11,294
13	8,890	9,042	9,233	9,460	9,728	10,043	10,411	10,833	11,294
14	8,888	9,042	9,232	9,460	9,728	10,043	10,411	10,833	11,294
15	8,890	9,043	9,233	9,461	9,729	10,043	10,412	10,833	11,294
K/λ	352,514	502,981	717,671	1024					
5	11,8645	12,2870	12,670	12,996					
6	11,7712	12,2212	12,624	12,964					
7	11,7736	12,2225	12,624	12,964					
8	11,7695	12,2192	12,622	12,962					
9	11,7671	12,2174	12,620	12,961					
10	11,7659	12,2166	12,620	12,960					
11	11,7654	12,2162	12,620	12,960					
12	11,7654	12,2162	12,619	12,960					
13	11,7653	12,2162	12,619	12,960					
14	11,7654	12,2162	12,619	12,960					
15	11,7655	12,2162	12,620	12,960					

Tabla A.2: Valores GCV para los tiempos térmicos. Elaboración: Autor.

K/λ	0,0010	0,0014	0,002	0,003	0,004	0,006	0,008	0,012	0,017
5	29,395	29,394	29,393	29,391	29,388	29,385	29,380	29,373	29,363
6	31,708	31,705	31,700	31,694	31,684	31,671	31,652	31,625	31,587
7	35,533	35,522	35,507	35,484	35,453	35,408	35,345	35,256	35,133
8	40,122	40,090	40,046	39,982	39,893	39,767	39,593	39,354	39,029
9	46,446	46,363	46,246	46,082	45,855	45,541	45,115	44,548	43,812
10	48,824	48,641	48,388	48,038	47,564	46,931	46,110	45,078	43,834
11	54,091	53,701	53,170	52,459	51,529	50,350	48,917	47,259	45,448
12	59,982	59,152	58,059	56,660	54,939	52,922	50,687	48,364	46,091
13	69,878	68,035	65,753	63,059	60,057	56,921	53,850	51,002	48,433
14	64,181	61,248	58,102	54,986	52,148	49,741	47,758	46,054	44,438
15	88,712	82,016	76,104	71,181	67,131	63,611	60,243	56,770	53,133
K/λ	0,024	0,034	0,049	0,070	0,099	0,142	0,202	0,288	0,411
5	29,349	29,329	29,300	29,260	29,205	29,129	29,024	28,885	28,703
6	31,533	31,458	31,355	31,214	31,024	30,776	30,459	30,073	29,623
7	34,964	34,735	34,431	34,036	33,543	32,952	32,279	31,556	30,825
8	38,598	38,040	37,341	36,503	35,548	34,519	33,469	32,451	31,499
9	42,887	41,770	40,486	39,085	37,634	36,199	34,827	33,541	32,348
10	42,410	40,868	39,289	37,754	36,312	34,980	33,753	32,619	31,574
11	43,583	41,766	40,059	38,478	37,001	35,602	34,271	33,018	31,861
12	43,971	42,039	40,267	38,612	37,041	35,553	34,166	32,896	31,752
13	46,101	43,922	41,831	39,817	37,912	36,156	34,576	33,176	31,945
14	42,775	41,039	39,287	37,607	36,064	34,680	33,448	32,344	31,348
15	49,449	45,918	42,725	39,970	37,662	35,748	34,153	32,804	31,645
K/λ	0,5867	0,8372	1,1945	1,7044	2,4318	3,4698	4,9509	7,0641	10,0794
5	28,473	28,195	27,874	27,525	27,168	26,827	26,520	26,258	26,043
6	29,128	28,619	28,128	27,685	27,301	26,974	26,690	26,433	26,192
7	30,123	29,475	28,886	28,349	27,854	27,396	26,976	26,599	26,269
8	30,625	29,826	29,096	28,434	27,842	27,321	26,873	26,493	26,173
9	31,250	30,252	29,362	28,585	27,920	27,359	26,891	26,502	26,179
10	30,621	29,763	28,999	28,328	27,744	27,241	26,812	26,449	26,142
11	30,814	29,885	29,071	28,366	27,762	27,247	26,812	26,446	26,138
12	30,732	29,830	29,039	28,350	27,755	27,246	26,814	26,449	26,142
13	30,867	29,924	29,104	28,394	27,785	27,266	26,827	26,457	26,147
14	30,447	29,634	28,905	28,260	27,694	27,205	26,786	26,430	26,129
15	30,637	29,755	28,982	28,309	27,726	27,226	26,800	26,439	26,135
K/λ	14,382	20,520	29,279	41,776	59,608	85,051	121,354	173,152	247,060
5	25,869	25,722	25,587	25,446	25,294	25,133	24,5383	24,832	24,718
6	25,964	25,752	25,556	25,374	25,201	25,038	24,889	24,761	24,662
7	25,985	25,744	25,538	25,357	25,188	25,030	24,884	24,759	24,661
8	25,906	25,683	25,491	25,321	25,162	25,011	24,870	24,749	24,654
9	25,910	25,685	25,493	25,322	25,163	25,011	24,870	24,749	24,654
10	25,885	25,668	25,481	25,314	25,157	25,007	24,867	24,747	24,653
11	25,881	25,664	25,478	25,311	25,155	25,005	24,5202	24,746	24,652
12	25,884	25,666	25,480	25,313	25,156	25,006	24,867	24,746	24,652
13	25,887	25,669	25,481	25,314	25,157	25,007	24,5206	24,747	24,653
14	25,875	25,661	25,476	25,310	25,154	25,005	24,866	24,746	24,652
15	25,879	25,664	25,478	25,311	25,155	25,006	24,866	24,746	24,652
K/λ	352,514	502,981	717,671	1024					
5	24,6358	24,5832	24,5532	24,974					
6	24,5935	24,5521	24,5307	24,5222					
7	24,5934	24,5522	24,5309	24,5224					
8	24,5885	24,5488	24,5285	24,5207					
9	24,5884	24,5487	24,5284	24,867					
10	24,5875	24,5480	24,5280	24,5203					
11	24,5870	24,5478	24,5278	24,866					
12	24,5873	24,5479	24,5279	24,5202					
13	24,5874	24,5480	24,5279	24,5203					
14	24,5870	24,5477	24,5278	24,5202					
15	24,5872	24,5479	24,5278	24,5202					

Tabla A.3: Valores GCV para la PAR. **Elaboración:** Autor.

Anexo B

Valores de índices para escenarios considerados

B.1. Escenario 1: Aplicación sin suavizamiento de datos

Método	$C_{p(prf)}$	$C_{pl(prf)}$	$C_{pu(prf)}$	$C_{pk(prf)}$
A	0,524	0,797	0,345	0,195
B	0,484	0,817	0,268	0,268

Tabla B.1: Valores índices de capacidad proceso para métodos A y B - precipitación. **Elaboración:** Autor.

Método	$C_{p(prf)}$	$C_{pl(prf)}$	$C_{pu(prf)}$	$C_{pk(prf)}$
A	3,037	5,763	1,567	1,567
B	2,955	5,567	1,463	1,463

Tabla B.2: Valores índices de capacidad proceso para métodos A y B - tiempos térmicos. **Elaboración:** Autor.

Maíz	Método	$C_{p(prf)}$	$C_{pl(prf)}$	$C_{pu(prf)}$	$C_{pk(prf)}$
INIAP H-551	A	0,896	0,995	0,839	0,647
	B	0,859	1,024	0,742	0,742
INIAP H-552	A	0,802	1,224	0,511	0,493
	B	0,768	1,243	0,435	0,435
Vencedor R-8330	A	0,987	0,776	1,154	0,639
	B	0,945	0,815	1,036	0,815

Tabla B.3: Valores índices de capacidad proceso para métodos A y B - PAR. **Elaboración:** Autor.

B.2. Escenario 2: Suavizamiento de las curvas $\mathcal{X}_{0,99865}$, $\mathcal{X}_{0,00135}$

Método	$C_{p(prf)}$	$C_{pl(prf)}$	$C_{pu(prf)}$	$C_{pk(prf)}$
A	0,511	0,811	0,313	0,295
B	0,484	0,819	0,268	0,268

Tabla B.4: Valores índices de capacidad proceso para métodos A y B - precipitación. **Elaboración:** Autor.

Método	$C_{p(prf)}$	$C_{pl(prf)}$	$C_{pu(prf)}$	$C_{pk(prf)}$
A	3,031	5,637	1,590	1,590
B	2,959	5,540	1,477	1,477

Tabla B.5: Valores índices de capacidad proceso para métodos A y B - tiempos térmicos. **Elaboración:** Autor.

Maíz	Método	$C_{p(prf)}$	$C_{pl(prf)}$	$C_{pu(prf)}$	$C_{pk(prf)}$
INIAP H-551	A	0,891	0,998	0,826	0,676
	B	0,860	1,018	0,749	0,749
INIAP H-552	A	0,797	1,223	0,504	0,504
	B	0,769	1,236	0,441	0,441
Vencedor R-8330	A	0,980	0,783	1,134	0,657
	B	0,946	0,809	1,043	0,809

Tabla B.6: Valores índices de capacidad proceso para métodos A y B - PAR. **Elaboración:** Autor.