



INFOTEC CENTRO DE INVESTIGACIÓN E
INNOVACIÓN EN TECNOLOGÍAS DE LA
INFORMACIÓN Y COMUNICACIÓN

DIRECCIÓN ADJUNTA DE INNOVACIÓN Y
CONOCIMIENTO
GERENCIA DE CAPITAL HUMANO
POSGRADOS

“UN ANÁLISIS HEURÍSTICO DE LA COMPLEJIDAD DEL SISTEMA GEOELECTORAL MEXICANO”

REPORTE ANALÍTICO DE EXPERIENCIA LABORAL
Que para obtener el grado de MAESTRO EN
CIENCIA DE DATOS E INFORMACIÓN

Presenta:

Miguel David Álvarez Hernández

Asesores:

Dr. Edgar González Fernández

Dr. Raúl Figueroa Romero

Ciudad de México, diciembre, 2020.

Autorización de impresión



AUTORIZACIÓN DE IMPRESIÓN Y NO ADEUDO EN BIBLIOTECA

MAESTRÍA EN CIENCIA DE DATOS E INFORMACIÓN

Ciudad de México, 21 de diciembre de 2020
INFOTEC-DAIC-GCH-SE-0652/2020.

La Gerencia de Capital Humano / Gerencia de Investigación hacen constar que el trabajo de titulación intitulado

UN ANÁLISIS HEURÍSTICO DE LA COMPLEJIDAD DEL SISTEMA GEOELECTORAL MEXICANO

Desarrollado por el alumno **Miguel David Álvarez Hernández** y bajo la asesoría del **Dr. Edgar González Fernández** y el **Dr. Raúl Figueroa Romero**; cumple con el formato de biblioteca. Por lo cual, se expide la presente autorización para impresión del proyecto terminal al que se ha hecho mención.

Asimismo se hace constar que no debe material de la biblioteca de INFOTEC.

Vo. Bo.

A handwritten signature in blue ink, appearing to read "Julieta Alcibar".

Mtra. Julieta Alcibar Hermosillo
Coordinadora de Biblioteca

Anexar a la presente autorización al Inicio de la versión impresa del trabajo referido que ampara la misma.

Agradecimientos

En palabras de mi padre: el agradecimiento es un valor tan soslayado que es necesario recuperar.

Agradezco, en forma general, a todas aquellas personas que, a través de sus enseñanzas o investigaciones, han contribuido positivamente en mi formación académica. Especialmente, agradezco las contribuciones de todos aquellos librepensadores y futuristas que, no conformes con su realidad, han mostrado a otros que el pensamiento crítico y científico es el mejor camino que puede seguir la humanidad. De igual forma, agradezco a todas las personas que han luchado por el acceso universal a la educación, un derecho sin el cual no hubiera sido posible realizar esta investigación.

En particular, agradezco el asesoramiento y el apoyo de los profesores Edgar González y Raúl Figueroa, ya que sus comentarios y consejos han sido decisivos para llevar a buen puerto la conclusión de mis estudios de maestría. Asimismo, agradezco a mis colegas del Instituto Nacional Electoral su apoyo y comprensión, ya que, sin su beneplácito, la tarea simultánea de realizar una maestría, y mantener un trabajo de tiempo completo, hubiera resultado ardua e inalcanzable.

Finalmente, y de manera especial, agradezco profundamente a mi familia por el apoyo que me han proporcionado, incondicionalmente, durante toda mi trayectoria de vida. A mi madre le doy mi más efusivo agradecimiento, ya que su inefable energía y afecto han hecho posible mi existencia. A mi hermana le otorgo mi más afectuoso agradecimiento, ya que su permanente compañía siempre ha proporcionado alegría a mi vida. Y a mi padre mi más eterno agradecimiento, ya que su sabiduría, e inquebrantable fuerza de voluntad, han sido siempre una fuente de inspiración, y por ello lo considero como el mejor maestro de vida que he tenido.

Tabla de contenido

Introducción.....	1
Capítulo 1. Consideraciones teóricas sobre la complejidad	9
1.1 Las metodologías modernas de análisis de sistemas complejos.....	13
1.2 El estudio de la complejidad en las ciencias sociales	15
Capítulo 2. Marco Metodológico	21
2.1 Definición del concepto de Complejidad Electoral	21
2.2 Los indicadores estadísticos para cuantificar la Complejidad Electoral	23
2.3 Transformación de datos y elaboración de los Índices de Complejidad Electoral	31
2.4 Agrupamiento de datos y elaboración de la Tipología de Complejidad Electoral	36
2.5 Identificación de casos atípicos de complejidad electoral	39
Capítulo 3. Resultados de los Índices de Complejidad Electoral a nivel municipal.....	43
3.1 Análisis exploratorio y construcción de los ICE municipales.....	44
3.2 Interpretación de los ICE municipales	51
3.3 <i>Ranking</i> municipal de complejidad electoral	55
Capítulo 4. Resultados de los Índices de Complejidad Electoral a nivel seccional	63
4.1 Análisis exploratorio y construcción de los ICE seccionales	64
4.2 Interpretación de los ICE seccionales	70
4.3 <i>Ranking</i> seccional de complejidad electoral	73
Capítulo 5. Tipología de la Complejidad Electoral: municipios	81
5.1 Análisis de silueta e implementación de K-medias: caso municipal.....	81
5.2 Estadísticas de los grupos de la TCE municipal.....	90
5.3 Identificación de casos atípicos a nivel municipal.....	94
Capítulo 6. Tipología de la Complejidad Electoral: secciones electorales ..	101
6.1 Análisis de silueta e implementación de K-medias: caso seccional.....	101
6.2 Estadísticas de los grupos de la TCE seccional	109
6.3 Identificación de casos atípicos a nivel seccional	113
Conclusiones.....	119

Bibliografía.....	126
ANEXO I: Pseudocódigo de Python para implementación del Análisis de Componentes Principales	133
ANEXO II: Pseudocódigo de Python para implementación de K-medias y Análisis de Coeficientes de Silueta	135

Índice de figuras

Figura 1. Ranking municipal de complejidad electoral.....	60
Figura 2. Ranking municipal de complejidad electoral en la zona metropolitana de la Ciudad de México.....	61
Figura 3. Ranking seccional de complejidad electoral.....	78
Figura 4. Ranking seccional de complejidad electoral en la zona metropolitana de la Ciudad de México.....	79
Figura 5. Mapa nacional de municipios con su respectiva clasificación en la TCE municipal	88
Figura 6. Mapa de los municipios de la zona metropolitana de la Ciudad de México con su respectiva clasificación en la TCE municipal.....	89
Figura 7. Mapa nacional de municipios con una complejidad electoral atípica	99
Figura 8. Mapa nacional de secciones electorales con su respectiva clasificación en la TCE seccional	107
Figura 9. Mapa de las secciones electorales de la zona metropolitana de la Ciudad de México con su respectiva clasificación en la TCE seccional.....	108
Figura 10. Mapa nacional de secciones electorales con una complejidad electoral atípica.....	117

Índice de gráficos

Gráfico 1. Panel de gráficas de correlación y distribución de los indicadores base que conforman los ICE municipales	45
Gráfico 2. Mapa de calor de la matriz de correlación de los indicadores municipales	46
Gráfico 3. ICE municipales: Proporción de varianza explicada por Componente Principal	48
Gráfico 4. Distribución de los municipios en el subespacio CP1-CP2.....	50
Gráfico 5. Distribución de los municipios en el subespacio CP1-CP2-CP3.....	51
Gráfico 6. ICE municipales: Ponderadores de los indicadores base en cada Componente Principal	55
Gráfico 7. Panel de gráficas de correlación y distribución de los indicadores que conforman los ICE seccionales	65
Gráfico 8. Mapa de calor de la matriz de correlación de los indicadores seccionales.....	66
Gráfico 9. ICE seccionales: Proporción de varianza explicada por Componente Principal	67
Gráfico 10. Distribución de las secciones electorales en el subespacio CP1-CP2	69
Gráfico 11. Proyección de las secciones electorales en el subespacio CP1-CP2-CP3	70
Gráfico 12. ICE seccionales: Ponderadores de los indicadores base en cada Componente Principal	73
Gráfico 13. TCE municipal: Diagramas de silueta para configuraciones de 2, 3, 4, 5 y 6 grupos.....	83
Gráfico 14. TCE municipal con K-medias. Subespacio CP1-CP2	84
Gráfico 15. TCE municipal con K-medias. Subespacio CP1-CP2-CP3.....	85
Gráfico 16. Diagrama de violín del ICE principal (con agrupación en función de la TCE municipal).....	91

Gráfico 17. Panel de diagramas de violín de los siete indicadores base (con agrupación en función de la TCE municipal)	93
Gráfico 18. TCE seccional: Diagramas de silueta para configuraciones de 2, 3, 4, 5 y 6 grupos.....	103
Gráfico 19. TCE seccional con K-medias. Subespacio CP1-CP2	104
Gráfico 20. TCE seccional con K-medias. Subespacio CP1-CP2-CP3.....	105
Gráfico 21. Diagrama de violín del ICE principal (con agrupación en función de la TCE seccional).....	110
Gráfico 22. Panel de diagramas de violín de los seis indicadores base (con agrupación en función de la TCE seccional)	111

Índice de cuadros

Cuadro 1. <i>Ejemplos de medidas de complejidad propuestas</i>	15
Cuadro 2. <i>Indicadores base para cuantificar la Complejidad Electoral.....</i>	30
Cuadro 3. <i>ICE municipales: Composición de las tres primeras Componentes Principales en función de los indicadores base.....</i>	52
Cuadro 4. <i>Ranking municipal de complejidad electoral: primeros 10 municipios por valoración en el ICE principal.....</i>	56
Cuadro 5. <i>Municipios de mayor complejidad electoral por entidad federativa.....</i>	57
Cuadro 6. <i>ICE seccionales: Composición de las tres primeras Componentes Principales en función de los indicadores base.....</i>	71
Cuadro 7. <i>Ranking seccional de complejidad electoral: primeras 10 secciones electorales por valoración en el ICE principal.....</i>	74
Cuadro 8. <i>Secciones electorales de mayor complejidad electoral por entidad federativa.....</i>	76

Siglas y abreviaturas

PCA	Análisis de Componentes Principales
CP	Componente Principal
CPV	Credencial para Votar
DERFE	Dirección Ejecutiva del Registro Federal de Electores
ICE	Índice de Complejidad Electoral
IIMA	Integración, Identificación, Mantenimiento y Actualización
INE	Instituto Nacional Electoral
INEGI	Instituto Nacional de Estadística y Geografía
LEGIPE	Ley General de Instituciones y Procedimientos Electorales
LNE	Lista Nominal de Electores
PE	Padrón Electoral
IQR	Rango Intercuartílico
RFE	Registro Federal de Electores
TCE	Tipología de Complejidad Electoral

Glosario

“A”

Análisis de Coeficientes de Silueta: Método matemático de interpretación y validación que utiliza coeficientes de silueta para evaluar la coherencia dentro de los cúmulos de datos, y también para cuantificar cuán cerca está cada punto de un cúmulo de los puntos de los cúmulos vecinos. Proporciona una forma de valorar visualmente parámetros como el número de cúmulos.

Análisis de Componentes Principales: Método matemático de transformación lineal que permite proyectar un conjunto mayor de observaciones a un nuevo sistema de coordenadas ortonormal, de tal forma que se obtiene un número menor o igual de variables no correlacionadas, conocidas como Componentes Principales. El método se utiliza ampliamente para destacar la variación y captar patrones fuertes en los conjuntos de datos, así como para reducir su dimensionalidad y facilitar su análisis visual.

Análisis de Correlación: Método estadístico utilizado para evaluar la fuerza de la relación entre dos variables cuantitativas. Una alta correlación significa que dos o más variables tienen una fuerte relación entre sí, mientras que una correlación débil significa que las variables apenas están relacionadas. En otras palabras, es el proceso de estudiar la fuerza de esa relación con los datos estadísticos disponibles.

Autocorrelación: Correlación de una serie de tiempo con una copia retardada de sí misma en función del retardo. Es decir, es la similitud entre las observaciones en función del retardo entre ellas.

“C”

Credencial para Votar: Identificación que contiene los datos de nombre, edad, sexo, domicilio y Clave Única de Registro de Población del ciudadano.

Coeficiente de Silueta: Medida de cuán similar es un objeto a su propio cúmulo (cohesión) en comparación con otros cúmulos (separación). Se utiliza frecuentemente en los algoritmos de agrupamiento para estudiar la distancia de separación entre los cúmulos de una tipología.

Coeficiente de Variación: Es una medida estadística estandarizada de la dispersión de una distribución de probabilidad o de frecuencia. Se suele expresar como un porcentaje, y se define como el cociente de la desviación estándar entre la media (“Coefficient of variation”, 2020).

Componente Principal: *Eigenvector* asociado a la matriz de correlación de un conjunto de observaciones. Proporciona las coordenadas de los datos sobre la base de un nuevo sistema de coordenadas ortonormal.

Complejidad Electoral: Conjunto de factores geoespaciales, demográficos y estructurales que impactan directamente en el desarrollo y ejecución de las tareas de Identificación, Integración, Mantenimiento y Actualización del Registro Federal de Electores.

“D”

Diagrama de violín: Método de trazado de datos numéricos que muestra la distribución de los datos cuantitativos a través de varios niveles de una (o más) variables categóricas, de manera que esas distribuciones puedan ser comparadas.

Distribución leptocúrtica: Distribución con una curtosis más grande que la de una distribución normal. Una distribución normal tiene una curtosis de tres. Por lo tanto, una distribución con curtosis mayor que tres es etiquetada como una distribución leptocúrtica.

“I”

Indicadores socioeconómicos: Conjunto de métricas, índices, *rankings* y clasificaciones elaborados con el fin de cuantificar el comportamiento global, o parcial, de un cierto sistema social y humano.

Índice de Complejidad Electoral: Número índice que valora el grado de complejidad electoral presente en una unidad territorial del sistema geoelectoral a partir de la ponderación de la información contenida en un grupo diverso de indicadores.

“K”

K-medias: Método de agrupamiento que tiene como objetivo la partición de un conjunto de n observaciones en k grupos, en el que cada observación pertenece al grupo cuyo valor medio es más cercano (“K-medias”, 2020).

“L”

Lista Nominal de Electores: Registro de todos aquellos ciudadanos mexicanos que cuentan con una Credencial para Votar vigente, y, por lo tanto, están en capacidad de sufragar.

“M”

Mapa de calor: Técnica de visualización de datos que muestra la magnitud de un fenómeno como color en dos dimensiones. La variación del color puede ser por tono o intensidad, dando pistas visuales obvias al lector sobre cómo el fenómeno se agrupa o varía respecto a categorías discretas.

Municipio: Unidad político-administrativa que sirve de base a la división territorial y organización política tripartita del Estado Mexicano (municipios, estados y federación) (Instituto de Investigaciones Jurídicas, 1984).

Módulo de Atención Ciudadana: Oficina del Instituto Nacional Electoral encargada de atender las solicitudes de credencialización de la ciudadanía.

“P”

Padrón Electoral: Registro de todos aquellos ciudadanos mexicanos que en algún momento tramitaron y recibieron una Credencial para Votar.

“R”

Rango intercuartílico: Medida de dispersión estadística igual a la diferencia entre los percentiles Q3 y Q1. En otras palabras, es una medida de la variabilidad basada en la división de un conjunto de datos en cuartiles. Los cuartiles dividen un conjunto de datos ordenados en cuatro partes iguales. Los valores que separan las partes se denominan primer, segundo y tercer cuartil, y se denotan por Q1, Q2 y Q3, respectivamente.

Ranking de complejidad electoral: Clasificación de complejidad electoral entre un conjunto de unidades territoriales, de tal manera que, para dos unidades cualesquiera (municipios o secciones), el primero está "clasificado más alto que", "clasificado más bajo que" o "clasificado igual a" el segundo.

Registro Federal de Electores: Base de datos conformada por el Padrón Electoral y la Lista Nominal de Electores.

Resecciónamiento: Proceso de modificación de los límites geográficos de una sección electoral.

“S”

Sección Electoral: Unidad político-electoral que sirve de base a la división territorial y organización electoral cuatripartita del Estado Mexicano (secciones, distritos, estados y circunscripciones).

Sistema complejo: Sistema compuesto por un vasto número de elementos constituyentes, el cual no posee un control o principio central rector, y que da muestras de respuesta al entorno; de comportamiento colectivo, de adaptación, de resiliencia y de evolución.

Sistema geoelectoral: Sistema compuesto por un vasto número de unidades político-administrativas-electorales (estados, distritos, municipios, y secciones electorales), que se caracteriza por la heterogeneidad de sus propiedades

sociodemográficas y geoespaciales, así como por el variado dinamismo de las interrelaciones entre las unidades del sistema.

“T”

Tipología de Complejidad Electoral: Sistema de agrupamiento de los municipios y secciones electorales en función de la similitud que muestran sus grados de complejidad electoral.

Introducción



Introducción

Motivación

Los listados de electores son una parte elemental y básica de cualquier sistema geoelectoral¹, ya que son instrumentos imprescindibles para proporcionar certidumbre y legitimidad a los procesos político-electorales que se desarrollan en un país. La tarea de construir estos instrumentos, aparentemente sencilla, adquiere diferentes grados de complejidad dependiendo de las características específicas de los sistemas geoelectorales, ya que entran en juego factores como la normatividad electoral, las condiciones sociopolíticas, y las características sociodemográficas y geográficas particulares de cada nación.

En términos generales, un listado de electores (o sistema registral) constituye una lista de ciudadanos que se integra a partir de una serie de condiciones que impone el marco normativo electoral que rige su conformación, particularmente en lo que respecta a (Welti Chanes, 2019):

- 1) Quiénes la integrarán y qué requisitos deben cumplir para formar parte de ella.
- 2) Cómo serán identificados sus integrantes.
- 3) Cuáles serán los medios físicos de resguardo en que se mantenga esta lista.
- 4) Cómo serán los mecanismos de actualización y consolidación de la información.

En el contexto específico del sistema geoelectoral mexicano, su listado de electores se encuentra representado en la figura del Registro Federal de Electores

¹ El concepto de sistema geoelectoral hace referencia a dos cuestiones primordiales: 1) las características político-electORALES y sociodemográficas propias de cada unidad territorial del país; y 2) las interrelaciones existentes entre dichas unidades territoriales (Álvarez Hernández y Álvarez Texocotilla, 2020). En particular, los municipios y las secciones electorales se pueden considerar como las unidades fundamentales del sistema geoelectoral mexicano, ya que los municipios representan la entidad base desde el punto de vista político/administrativo, mientras que las secciones electorales representan la entidad base desde el punto de vista electoral.

(RFE),² figura que es administrada exclusivamente por el Instituto Nacional Electoral (INE), ya que este organismo es la autoridad que elabora y supervisa la aplicación de los lineamientos normativos que rigen al listado de electores. En particular, el INE determina los procedimientos de empadronamiento (Integración), credencialización (Identificación), resguardo de las bases de datos (Mantenimiento) y consolidación de la información georreferenciada (Actualización).³

Es importante considerar que, en el caso particular de México, el RFE ha contribuido significativamente a otorgar mayor certeza a los procesos político-electorales que se desarrollan en el país, y también ha coadyuvado a otorgar una mayor legitimidad a los resultados de las elecciones, ya que uno de los propósitos principales del RFE es garantizar el libre acceso y ejercicio de los derechos políticos (Welti Chanes, 2019).⁴ Por lo tanto, las tareas IIMA son fundamentales para garantizar la calidad y consistencia del RFE.

Considerando este contexto, resulta primordial contar con herramientas analíticas que permitan planear, dirigir y evaluar los programas y acciones institucionales enfocados a las tareas IIMA del RFE, y que además tomen en consideración la diversidad geográfica del país, y la pluralidad de las características sociodemográficas de la población. Particularmente, resulta crucial contar con métricas e indicadores que sinteticen y cuantifiquen los factores que inciden en la complejidad electoral de los municipios y las secciones electorales de México,⁵ ya que dichas unidades territoriales son los bloques básicos que conforman al sistema geoelectoral mexicano.

² El RFE está conformado por dos instrumentos interrelacionados: el Padrón Electoral, y la Lista Nominal de Electores. El Padrón Electoral representa el registro de todos aquellos ciudadanos que en algún momento tramitaron y recibieron una Credencial para Votar (CPV). En contraste, la Lista Nominal representa el registro de todos aquellos ciudadanos que cuentan con una credencial para votar vigente y, por lo tanto, están en capacidad de votar. Ambos instrumentos son de carácter permanente (es decir, son registros que perduran independientemente de los calendarios electorales), pero están sujetos continuamente a procedimientos de depuración y actualización.

³ En lo subsiguiente, y para simplificar la escritura, a este grupo de cuatro tareas se le asociará la sigla IIMA.

⁴ Ver Ley General de Instituciones y Procedimientos Electorales [LEGIPE] (2020, artículo 54).

⁵ Como se explicará con mayor detalle en el Capítulo 1, la presente investigación define el concepto de complejidad electoral como el conjunto de factores geoespaciales, demográficos y estructurales que dificultan u obstaculizan las labores de integración, identificación, mantenimiento y actualización (tareas IIMA) del RFE.

Problema de investigación

A pesar de la necesidad institucional de conocer con precisión y actualidad la dinámica de los factores demográficos, geoespaciales y estructurales que determinan la complejidad electoral de los municipios y secciones electorales de México, no existe actualmente una metodología unificadora que permita su análisis y monitoreo. Asimismo, hay una falta acuciante de estudios que exploten en forma adecuada la información cartográfica, demográfica y electoral que genera continuamente el INE a nivel municipal y seccional, sobre todo considerando que dicha información se caracteriza por tener un mejor nivel de actualización respecto a otras fuentes de información oficial.⁶

Consecuentemente, la presente investigación busca solventar la actual falta de estudios en torno a la evaluación y monitoreo de las tareas IIMA mediante la propuesta de una metodología de análisis cuantitativo que permita analizar y evaluar los diversos factores que afectan el desarrollo de las tareas IIMA y, por ende, se pueda cuantificar la complejidad electoral específica de cada una de las unidades territoriales que conforman al sistema geoelectoral mexicano.

Resumiendo, el problema de investigación consiste en elaborar un conjunto de índices/métricas que caractericen y evalúen, de forma unificadora, los factores geoespaciales, demográficos y estructurales que inciden en el desarrollo de las tareas IIMA, lo cual permitirá cuantificar la complejidad electoral específica de cada municipio y sección electoral, así como elaborar una clasificación de las unidades territoriales del sistema geoelectoral en función de sus grados de complejidad electoral.

Objetivos

El objetivo principal del presente trabajo de investigación consiste en definir y cuantificar la complejidad del sistema geoelectoral, asociada a la realización de las

⁶ La actualización de la información cartográfica, sociodemográfica y electoral que genera el INE, respecto a otras instituciones como el INEGI, se debe primordialmente a la realización continua de los procesos de credencialización y verificación de los datos personales/georreferenciales de la ciudadanía mexicana. Dichos procesos implican la revisión y verificación periódica de las bases de datos, lo cual representa una ventaja con respecto a las fuentes de información que proceden de ejercicios censales o muestrales.

tareas IIMA del RFE, para posteriormente realizar una tipificación de los municipios y secciones electorales en función de la similitud observada en sus grados de complejidad electoral. Para alcanzar este objetivo se elaborarán dos conjuntos de Índices de Complejidad Electoral (ICE) a nivel municipal y seccional, los cuales utilizarán y se sustentarán en información procedente de diversos indicadores geoespaciales, demográficos y estructurales. Asimismo, los resultados estadísticos obtenidos a partir de los ICE se complementarán con la elaboración de una Tipología de Complejidad Electoral (TCE) a escala municipal y seccional, la cual permitirá tipificar a los municipios y secciones electorales en función de su complejidad electoral.

Respecto a los objetivos específicos de la investigación, estos se enumeran a continuación:

1) Elaborar un *ranking* de complejidad electoral, a nivel municipal y seccional, con la información proporcionada por los ICE. La intención de este *ranking* es proporcionar una clasificación simplificada de los municipios y secciones electorales.

2) Identificar los municipios y secciones electorales que presentan una complejidad electoral atípica de acuerdo con los resultados obtenidos en los ICE y la TCE. Este ejercicio pretende ubicar aquellas unidades territoriales del sistema geoelectoral que necesitan una mayor atención para cumplir adecuadamente con las tareas IIMA del RFE.

3) Contextualizar la metodología de cuantificación de la complejidad electoral frente a las metodologías modernas de estudio de los sistemas complejos.

Respecto a los alcances que tendrán los resultados de la investigación, se considera que los ICE y la TCE tendrán un efecto positivo y trascendente en las tareas fundamentales del INE, ya que estos instrumentos proporcionan una forma novedosa y efectiva para analizar y monitorear la evolución de los factores demográficos, geoespaciales y estructurales que afectan las tareas IIMA del RFE. De esta forma, el INE contará con mayores elementos técnicos que le permitirán mejorar y robustecer la eficacia de sus procesos de planeación presupuestal y organización electoral. Asimismo, también se considera que los instrumentos que

se presentan en este trabajo de investigación tendrán un efecto benéfico en la ciudadanía, ya que, al mejorar la eficiencia de los procesos administrativos del INE, la institución podrá ofrecer mejores servicios de atención a la población, sobre todo en el tema de la cobertura de los Módulos de Atención Ciudadana.⁷

Por otra parte, la utilización de los ICE y la TCE permitirá evaluar, a una escala precisa, la situación demográfica de México, ya que estos instrumentos posibilitan el análisis de la variabilidad del comportamiento poblacional de los municipios y las secciones electorales del país, lo cual a su vez facilitaría la creación de modelos de pronóstico demográfico locales mejor preparados para proporcionar proyecciones demográficas precisas. Consecuentemente, los ICE y la TCE ayudarán a solventar la actual falta de estudios que tengan como propósito monitorear la evolución del sistema geoelectoral mexicano a escala municipal y seccional.⁸

Estructura de la investigación

Para alcanzar los objetivos de la presente investigación, se inicia el Capítulo 1 con una breve revisión del debate contemporáneo sobre los sistemas complejos; se examinan algunas metodologías modernas de cuantificación de la complejidad y se discute el impacto de éstas en el campo de las ciencias sociales.

Posteriormente, en el Capítulo 2 se presenta el marco metodológico de la investigación. Específicamente, se propone una definición heurística de complejidad electoral enfocada a los objetivos de la investigación. Asimismo, se detalla el proceso de selección de los indicadores base que sirven para cuantificar la complejidad electoral, se expone la metodología de elaboración de los ICE –

⁷ Los Módulos de Atención Ciudadana son las oficinas del INE encargadas de atender las solicitudes de credencialización de la ciudadanía.

⁸ Actualmente el INE cuenta con algunas herramientas de análisis a nivel seccional, como es el caso de la tipología de secciones electorales (Instituto Federal Electoral [IFE], 1995) fundada en el número de localidades rurales (dicha tipología distingue tres tipos de categoría: secciones urbanas, secciones mixtas y secciones rurales). No obstante, esta tipología de secciones presenta las siguientes deficiencias: 1) es demasiado simple para capturar el concepto de complejidad electoral, ya que no toma en consideración otros indicadores sociodemográficos o geoespaciales; y 2) carece de una metodología que justifique la determinación de los grupos de la tipología.

fundamentada en el Análisis de Componentes Principales—, y se explica el algoritmo de agrupamiento –K-medias— utilizado en la construcción de la TCE.

En los Capítulos 3 y 4 se presentan los resultados estadísticos y gráficos del análisis exploratorio de datos, de la aplicación de la técnica de Análisis de Componentes Principales –detallando la capacidad explicativa de los ICE—, y de los patrones de agrupamiento discernidos en la proyección de los municipios y secciones electorales en el espacio de los ICE.

Luego, en los Capítulos 5 y 6 se exponen los resultados gráficos de la implementación del método de agrupamiento K-medias para la generación de la TCE a escala municipal y seccional. Se presentan los procedimientos de generación de diferentes escenarios de agrupamiento, la determinación de la agrupación óptima, los patrones de conglomeración de la TCE (con sus respectivas estadísticas a escala municipal y seccional), y la identificación de los casos atípicos de complejidad electoral.

Por último, se presentan las conclusiones y reflexiones que retoman, desde una perspectiva holística, los hallazgos obtenidos en cada una de las fases de la investigación. Asimismo, se plantean algunas líneas de investigación futuras, encaminadas a potenciar las capacidades de la metodología de los ICE y la TCE.

Recursos de la investigación

Para la construcción de los ICE y la TCE se utiliza primordialmente información procedente de las bases de datos del Padrón Electoral (PE) y la Lista Nominal de Electores (LNE); en particular, se usan los cortes de información del año 2019 para la extracción de los indicadores base. La información extraída del PE y la LNE se complementa con otros indicadores geoespaciales, obtenidos a partir de la base de datos cartográfica del INE.⁹

Respecto a las herramientas tecnológicas y computacionales que se implementan para el análisis de la información, la construcción de los ICE y la TCE;

⁹ Los datos utilizados en la investigación son de dominio público (tras previa solicitud formal a las oficinas de enlace público del INE). Los datos fuente de esta investigación se ponen a disposición del lector en los archivos anexos ubicados en el repositorio de GitHub, véase Álvarez Hernández (2020).

y la presentación de resultados, se hace un uso extensivo del lenguaje de programación *Python*, complementado con librerías especializadas en algoritmos de aprendizaje automático (*Scikit-learn*) y elaboración de gráficos estadísticos y animaciones (*Seaborn*, *Plotly*). Asimismo, también se utiliza el programa *QG/S* para la elaboración de mapas que muestran geográficamente los resultados derivados a partir de los ICE y la TCE.

Capítulo 1

Consideraciones teóricas sobre la complejidad



Capítulo 1. Consideraciones teóricas sobre la complejidad

No es errado considerar al mundo contemporáneo como un conjunto de sistemas altamente interconectados, cada uno de ellos caracterizado por la presencia de diversos procesos emergentes y fenómenos evolutivos de difícil predicción. Ejemplos de sistemas con estas características se pueden encontrar sobre todo en el ámbito de lo social, ya que la naturaleza impredecible del comportamiento individual y colectivo sienta las bases de la complejidad inherente que exhiben los sistemas sociales (Érdi, 2008).

En consecuencia, el tema de la complejidad ha tenido un impacto significativo en la forma como se concibe el conocimiento científico, lo cual ha generado algunas controversias, ya que en opinión de algunos expertos (por ejemplo, Waldrop, 1992; Wolfram, 2002), los estudios sobre la complejidad representan el posible inicio de una revolución científica; en contraste, la complejidad representa para otros un paradigma sin bases científicas sólidas que justifiquen sus indagaciones (por ejemplo, Horgan, 2015).

Esta diversidad de opiniones ha enriquecido el campo de estudio de los sistemas complejos, pero también ha ocasionado el surgimiento de decenas de definiciones, teorías y líneas de investigación sobre complejidad, sin que ninguna logre un consenso en la comunidad científica.

En este marco de falta de integración en el campo de la complejidad, se pueden distinguir dos corrientes principales de investigación (Phelan, 2001, p. 132; Becerra, 2020, p. 2): las autonombadas “ciencias de la complejidad”, caracterizadas por el enfoque algorítmico que siguen para analizar el comportamiento de sistemas adaptativos, dinámicos, no-equilibrados y no-lineales,¹⁰ y la tradición sistémico-cibernética enfocada al estudio de fenómenos de varios dominios desde una perspectiva de sistemas.¹¹ Ambas corrientes de investigación

¹⁰ Algunos de los principales exponentes de esta corriente son: Edward Lorenz, Benoit Mandelbrot e Ilya Prigogine.

¹¹ Algunos de los principales autores referentes de esta corriente de investigación son: Ludwig von Bertalanffy, Ross Ashby, Norbert Wiener, Gregory Bateson y Heinz von Foerster.

no comparten un sentido unívoco en torno al concepto de “complejidad”, ni incluso al interior de ellas –por ejemplo, existen más de 40 definiciones formuladas solamente en el ámbito de las ciencias de la complejidad— (Lloyd, 2001, p.7).

Desde la perspectiva de las ciencias de la complejidad se ha considerado que la forma adecuada de abordar la cuestión de la complejidad es encontrando una definición formal que explique las condiciones necesarias y suficientes que permitan inequívocamente distinguir los sistemas que son complejos de aquellos que no lo son. Alternativamente, la tradición sistémico-cibernetica señala que se puede definir el concepto de complejidad a través de una definición extensiva que tenga como base un enfoque más fenomenológico, el cual busque caracterizar al conjunto de sistemas complejos en términos de sus elementos constitutivos (Álvarez Hernández y Álvarez Texocotitla, 2020). No obstante, se ha encontrado que ambos enfoques presentan problemáticas desde el punto de vista metodológico (Flores Valdés y Martínez Mekler, 2011).

Para el caso de una definición formal, el problema principal radica en la obtención de una definición demasiado restrictiva o general. Por ejemplo, en las ciencias de la computación se define un concepto de complejidad computacional que sólo puede ser utilizado adecuadamente en el ámbito de su disciplina, ya que dicha definición considera únicamente aquellos sistemas que funcionan en términos de problemas computacionales (Flores Valdés y Martínez Mekler, 2011).¹²

Con respecto a establecer una definición extensiva, la cuestión principal reside en discriminar en forma no arbitraria aquellos sistemas que son complejos desde el punto de vista fenomenológico. Sin embargo, cada disciplina da una prioridad distinta a los fenómenos que consideran relevantes, lo cual añade un sesgo subjetivo a la definición del concepto de complejidad. Por ejemplo, para las ciencias sociales, ciertos fenómenos indican que un sistema es complejo, mientras que, desde la perspectiva de la matemática y de las ciencias naturales, otros fenómenos son más relevantes al momento de distinguir un sistema complejo de uno que no lo es (Álvarez Hernández y Álvarez Texocotitla, 2020).

¹² Para obtener un panorama más amplio y detallado de la Teoría de la Complejidad Computacional se recomienda revisar el texto de Goldreich (2008).

Tomando en consideración la falta de consenso sobre el concepto de complejidad, Mitchell (2009) ha propuesto una definición heurística de sistema complejo sustentada en un conjunto de propiedades que la mayoría de los investigadores ha considerado como inherentes y universales a los sistemas complejos (Ladyman, Lambert y Wiesner, 2013). Es importante puntualizar que estas propiedades universales de los sistemas complejos —que se explican a continuación— no aluden a un solo marco teórico o campo de estudio específico, ya que trascienden las barreras disciplinarias y establecen afinidades entre multitud de fenómenos y sistemas de diversa naturaleza (Mitchell, 2009):

La composición jerárquica a partir de elementos constitutivos simples.

Los sistemas complejos están frecuentemente compuestos por un vasto número de elementos individuales más simples que el sistema en su conjunto. Dichos elementos interactúan en formas complejas, dando pie a la formación de estructuras jerárquicas a diferentes escalas dentro del sistema.¹³ Por lo tanto, los sistemas complejos exhiben diferentes niveles de organización que tienen dinámicas individuales propias, las cuales en conjunto producen la dinámica global del sistema.

La capacidad de respuesta al entorno. Los sistemas complejos muestran capacidad de respuesta al entorno. Es decir, los mecanismos internos de estos sistemas son capaces de captar y procesar la información procedente del entorno, lo cual les permite adaptar y variar su comportamiento, e incluso su constitución, en función de los estímulos externos recibidos.

El comportamiento colectivo. Las interacciones que ocurren al interior de los sistemas complejos, así como la capacidad de captación y procesamiento de información externa, crean condiciones que facilitan el surgimiento de comportamientos colectivos entre los elementos constitutivos del sistema, los

¹³ El enfoque de la composición jerárquica ha sido ampliamente utilizado en el campo de la biología y la ecología, ya que en dichos campos es más evidente encontrar sistemas que presentan una jerarquización en las interrelaciones que ocurren entre los diferentes componentes de los sistemas. Un ejemplo claro es la dinámica molecular de las células (Oltvai y Barabási, 2002), ya que se conoce que las interacciones fisicoquímicas que ocurren entre los componentes básicos de las células originan comportamientos complejos, los cuales a su vez influyen en la dinámica de mayor escala de los sistemas biológicos.

cuales pueden cambiar totalmente el comportamiento global de éste; comportamiento que se distingue por tener, o no tener, un principio central rector.

La adaptación, resiliencia y evolución. Los sistemas complejos presentan dinámicas internas que, en conjunto, conservan el estado actual del sistema frente a cambios exógenos. Esta característica es una reminiscencia directa del comportamiento evolutivo que se observa en numerosos sistemas físicos, biológicos y sociales, ya que mediante un proceso de aprendizaje (consciente o inconsciente), estos sistemas se adaptan a las condiciones cambiantes de la realidad con el propósito de incrementar su perdurabilidad.

Por consiguiente, con base en las anteriores propiedades de los sistemas complejos, Mitchell (2009, p. 13) propone la siguiente definición heurística de un sistema complejo:

“Un sistema complejo es aquel sistema que esté compuesto por un vasto número de elementos constituyentes, que no tenga un control o principio central rector y que dé muestras de respuesta al entorno, de comportamiento colectivo, de adaptación, de resiliencia y de evolución”.

La anterior definición ha sido adaptada y modificada en numerosas ocasiones, ya que algunos estudios han relajado o enfatizado algunas de las propiedades enumeradas –principalmente en el ámbito de las propiedades relacionadas con el comportamiento colectivo—, esto con el propósito de delimitar aspectos particulares de los sistemas bajo estudio. Al respecto, es importante señalar que el estudio de los sistemas complejos es un campo interdisciplinario que actualmente se encuentra activo y en desarrollo y, por consiguiente, las propiedades y definiciones que se han propuesto hasta el momento deben considerarse temporales, ya que los avances y estudios que se realicen en el futuro mejorarán la comprensión sobre el comportamiento y las características de los sistemas complejos.¹⁴

¹⁴ Se recomienda al lector revisar los textos de Northrop (2011), Mitchell (2009), Érdi (2008), y Flores Valdés y Martínez Mekler (2011) para profundizar sobre el estado de arte de las ciencias de la complejidad.

1.1 Las metodologías modernas de análisis de sistemas complejos

A mediados del siglo XX, algunos científicos empezaron a señalar que el estudio de los fenómenos asociados a sistemas complejos no podía encasillarse en una sola disciplina, ya que, para poder modelar y comprender fehacientemente su funcionamiento, hacía falta construir un marco teórico amplio e interdisciplinario (Horgan, 2015). En este contexto, se gestaron diversas teorías, corrientes y campos científicos; por ejemplo, la cibernetica, la sinergética, la ciencia de sistemas, las ciencias cognitivas, la teoría de catástrofes y, más recientemente, las ciencias de la complejidad.

Como se comentó previamente, la comunidad científica reconoce la existencia de ciertas propiedades universales inherentes a los sistemas complejos; sin embargo, es frecuente observar que cada teoría o modelo sobre la complejidad propone alguna conceptualización específica y particular del concepto. Esta diversidad de conceptualizaciones se manifiesta sobre todo cuando se analizan las múltiples fórmulas y métodos que se han propuesto para cuantificar la complejidad. El problema de la cuantificación de la complejidad es primordial, ya que para entender cabalmente el funcionamiento de los sistemas complejos, es necesario generar cuantificadores que permitan caracterizar a los sistemas en función de su dinámica y composición, y que también permitan realizar comparaciones entre sistemas de diversa naturaleza.

Al respecto, es importante hacer notar que la problemática actual en torno al problema de la cuantificación de la complejidad también se ha suscitado en otros campos científicos. Por ejemplo, análogo al problema de medir la complejidad fue el problema de describir el electromagnetismo antes de las ecuaciones de Maxwell.¹⁵ En el caso del electromagnetismo, las cantidades tales como las fuerzas eléctricas y magnéticas —que se descubrieron en diferentes contextos experimentales— se consideraron en un inicio como fenómenos fundamentalmente diferentes. No obstante, con el tiempo los físicos dilucidaron que la electricidad y el magnetismo eran, de hecho, aspectos estrechamente relacionados del mismo

¹⁵ Las ecuaciones de Maxwell son un conjunto de cuatro ecuaciones diferenciales que forman la base teórica clásica para describir todos los fenómenos electromagnéticos.

fenómeno físico, el campo electromagnético (Sánchez Ron, 2007). De manera similar, los investigadores contemporáneos en biología, ciencias de la computación, física, genética, sistemas dinámicos, etc., han definido diferentes medidas de complejidad especializadas para cada campo y, por lo tanto, existe la impresión de que cada una de estas medidas o cuantificadores explican en realidad aspectos parciales de un fenómeno más amplio.

Por consiguiente, a medida que se han formulado diversas teorías y modelos sobre la complejidad, el número de fórmulas y métodos para cuantificarla ha crecido paralelamente. No obstante, es posible encontrar patrones en las muchas medidas de complejidad, ya que los investigadores tienden a realizar preguntas similares sobre la complejidad en sus diferentes métodos de investigación y, por lo tanto, sus medidas comparten factores en común.

En este contexto, Lloyd (2001) clasificó 40 medidas o cuantificadores de complejidad agrupados en 3 cuestiones principales: dificultad de descripción, dificultad de creación y grado de organización.

La primera cuestión señala que la complejidad de un sistema a menudo se asocia con el grado de dificultad que implica la descripción completa del sistema bajo estudio. La segunda cuestión está relacionada con el grado de dificultad involucrado en la construcción o duplicación de un sistema, el cual puede formar la base de una medida de complejidad. Y la tercera cuestión está relacionada con aspectos organizativos; por ejemplo, la cantidad de información compartida entre las partes de un sistema como resultado de su estructura organizativa, o la dificultad de describir las estructuras y jerarquías organizacionales (Álvarez Hernández y Álvarez Texocotitla, 2020).

En el Cuadro 1 se muestran algunas de las medidas de complejidad clasificadas por Lloyd (2001). Es importante puntualizar que estas medidas de complejidad abordan aspectos parciales o específicos de los sistemas complejos, por lo que ninguna es considerada actualmente como una medida universal de

complejidad.¹⁶ Por lo tanto, la aplicación de estas medidas está limitada por las características particulares del sistema estudiado.

Descripción	Creación	Organización
Dimensión fractal ¹⁷	Profundidad lógica ¹⁸	Complejidad jerárquica ¹⁹
Entropía ²⁰	Profundidad termodinámica ²¹	Complejidad estocástica ²²
Complejidad algorítmica ²³	Complejidad computacional ²⁴	Complejidad efectiva ²⁵
Descripción de longitud mínima ²⁶		Información mutua ²⁷

Cuadro 1. Ejemplos de medidas de complejidad propuestas

Fuente: Elaboración propia

1.2 El estudio de la complejidad en las ciencias sociales

Actualmente, las ciencias sociales distinguen dos enfoques de investigación fundamentales: el cuantitativo y el cualitativo (Haynes, 2018; Hernández-Sampieri y Mendoza Torres, 2018). Estos dos enfoques se caracterizan por tener diferentes fortalezas y debilidades en su aplicación y, por lo tanto, difieren en su capacidad explicativa y predictiva.

Por lo general, el enfoque cuantitativo es preferido cuando los problemas de investigación se encuentran bien definidos, o cuando hay una alta posibilidad de éxito de obtener resultados generales de los conceptos que se están problematizando a partir de métodos de modelado matemático/estadístico. En contraste, el enfoque cualitativo es más adecuado en un contexto donde las

¹⁶ Es importante mencionar que la diversidad de medidas de la complejidad ha sido considerada por algunos escépticos de las ciencias de la complejidad como un indicio de confusión y falta de madurez en el campo (Horgan, 2015).

¹⁷ Ver Mandelbrot (1967).

¹⁸ Ver Bennett (1988).

¹⁹ Ver Simon (1962).

²⁰ Ver Sethna (2011).

²¹ Ver Lloyd y Pagels (1988).

²² Ver Rissanen (1986).

²³ Ver Lambalgen (1989).

²⁴ Ver Goldreich (2008).

²⁵ Ver Gell-Mann y Lloyd (1996).

²⁶ Ver Rissanen (1986).

²⁷ Ver Northrop (2011).

preguntas primordiales de investigación tienen poca claridad, o la definición de las preguntas y los problemas de interés se cuestionan y por lo tanto requieren de más debate y reflexión (Hernández-Sampieri y Mendoza Torres, 2018).

Sin embargo, existen algunos enfoques alternativos de investigación que disputan el aparente dualismo entre los métodos cuantitativo y cualitativo. Estos enfoques alternativos, inspirados en ideas provenientes del campo de las ciencias de la complejidad, argumentan que es necesario crear una síntesis metodológica que retome las principales fortalezas de los enfoques cuantitativo y cualitativo (Haynes, 2018, pp. 5-33). El argumento principal que manejan estos enfoques alternativos es que la gran mayoría de los sistemas sociales y humanos muestran las principales propiedades que caracterizan a los sistemas complejos y, por ende, son sistemas que no pueden ser analizados fidedignamente sólo a través de teorías clásicas o modelos reduccionistas.

Por consiguiente, es necesario construir, en el ámbito de las ciencias sociales, técnicas y metodologías adaptativas que sean capaces de ajustarse y combinarse con el pragmatismo de las ciencias de la complejidad. Asimismo, dichas técnicas y metodologías deben mantener un balance entre lo cualitativo y lo cuantitativo, donde los juicios subjetivos y las observaciones sean tan importantes como las medidas cuantitativas y los modelos matemáticos/estadísticos (Álvarez Hernández y Álvarez Texocotitla, 2020).

Como se ha mencionado, los sistemas sociales son dinámicos, y la característica central de este dinamismo es la interacción entre personas y grupos colectivos de personas. Esta dinámica de interacciones no está en un estado perpetuo de aleatoriedad o desorden, ya que los sistemas sociales muestran patrones de jerarquización y orden, los cuales a menudo no tienen un origen o una explicación causal diáfana. Por ende, la tarea principal de la investigación social es hallar dichos patrones y entender los procesos que los originan, diseñando instrumentos que logren capturar y cuantificar los aspectos primordiales y la complejidad de los sistemas sociales (Byrne, 2001).

Por otra parte, es importante explicar cómo las metodologías e indicadores que tratan de analizar y cuantificar la complejidad de los sistemas físicos y

biológicos se han extendido al ámbito de los sistemas sociales. Al respecto, ha habido dos planteamientos generales de cómo trasladar y adaptar estas metodologías al ámbito específico de lo social.

El primer planteamiento, defendido principalmente por los científicos de las ciencias naturales, propone trasladar directamente las actuales metodologías de análisis y medición de la complejidad al ámbito de los sistemas sociales, sin recurrir prácticamente al uso de las metodologías y teorías tradicionales de las ciencias sociales. Es decir, desde esta perspectiva, la intención principal es encontrar una síntesis o teoría unificadora que explique, en un solo marco teórico, el funcionamiento de los sistemas físicos complejos y el funcionamiento de los sistemas sociales complejos.

Esta postura ha tenido poco impacto en las ciencias sociales, principalmente porque desconoce o descalifica gran parte del cuerpo teórico e ideológico acumulado en las disciplinas sociales. Asimismo, porque tradicionalmente no hay una comunicación circular entre los científicos naturales y sociales (ya que no comparten un mismo estilo o lenguaje de análisis), esto ha conllevado a que numerosos estudios de complejidad queden relegados únicamente a los círculos académicos especializados.

El segundo planteamiento, promovido por los investigadores adeptos a los métodos cuantitativos de las ciencias sociales, promulga la adopción de los conceptos provenientes de las ciencias de la complejidad, sin abandonar completamente las estructuras teóricas e ideológicas ya establecidas en las ciencias sociales.

La idea general de esta postura es crear nuevos métodos cuantitativos en las ciencias sociales, así como potenciar los actualmente existentes, a partir de una síntesis de ideas provenientes tanto de los cuerpos teóricos tradicionales de las disciplinas sociales, como de los cuerpos teóricos de las ciencias naturales. De esta forma, se podrá estudiar en forma no reduccionista, y en toda su extensión, la complejidad de los sistemas sociales.

Es importante mencionar que las ramas cuantitativas de las ciencias sociales ya han ideado numerosos métodos e indicadores que permiten la cuantificación y

caracterización de los fenómenos sociales, a tal punto que actualmente se puede reconocer que los sistemas de indicadores socioeconómicos son una base fundamental para cualquier descripción sólida de los hechos sociales y para la formulación de políticas normativas (Cooley y Snyder, 2015).²⁸

No obstante, la forma tradicional de construcción de estos indicadores ha estado fuertemente influenciada por el programa reduccionista y mecanicista, por lo que muchos métodos de cuantificación en las ciencias sociales tienden a generar medidas agregadas de los fenómenos, con el fin de simplificar el análisis de éstos (Álvarez Hernández y Álvarez Texocotitla, 2020).²⁹

Lo anterior tiene una implicación. Cualquier descripción "promedio" es intrínsecamente ineficaz para captar fehacientemente la dinámica de un sistema social complejo. Considerando que los indicadores socioeconómicos tradicionales tienden a presentar visiones agregadas de la realidad, éstos tienen un limitado grado de validez como instrumentos de estudio o de formulación y evaluación de políticas normativas (Maggino, 2017, pp. vii-viii). Es decir, la descripción y cuantificación de los sistemas sociales y humanos debe preservar de alguna manera la complejidad inherente de estos sistemas. La complejidad se puede simplificar y reducir, pero no se puede borrar sin privar a los productos estadísticos finales de su contenido de información y de su utilidad tanto para los científicos sociales como para los creadores de políticas (Álvarez Hernández y Álvarez Texocotitla, 2020).

Al respecto, es pertinente describir la complejidad existente al elaborar indicadores socioeconómicos adecuados. Dependiendo del objeto de estudio, la información estilizada puede requerir un indicador simple (índices, estadísticos, mapas u otros métodos de visualización y exploración de datos) y procesos estadísticos adecuados capaces de alimentarlos con información significativa.³⁰ En

²⁸ Por indicadores socioeconómicos se entiende aquellos indicadores (métricas, índices, *rankings*, clasificaciones) elaborados con el fin de cuantificar el comportamiento global o parcial de un cierto sistema social y humano.

²⁹ Hay que tener presente que proporcionar visiones sintéticas de los fenómenos sociales no implica necesariamente computar indicadores agregados.

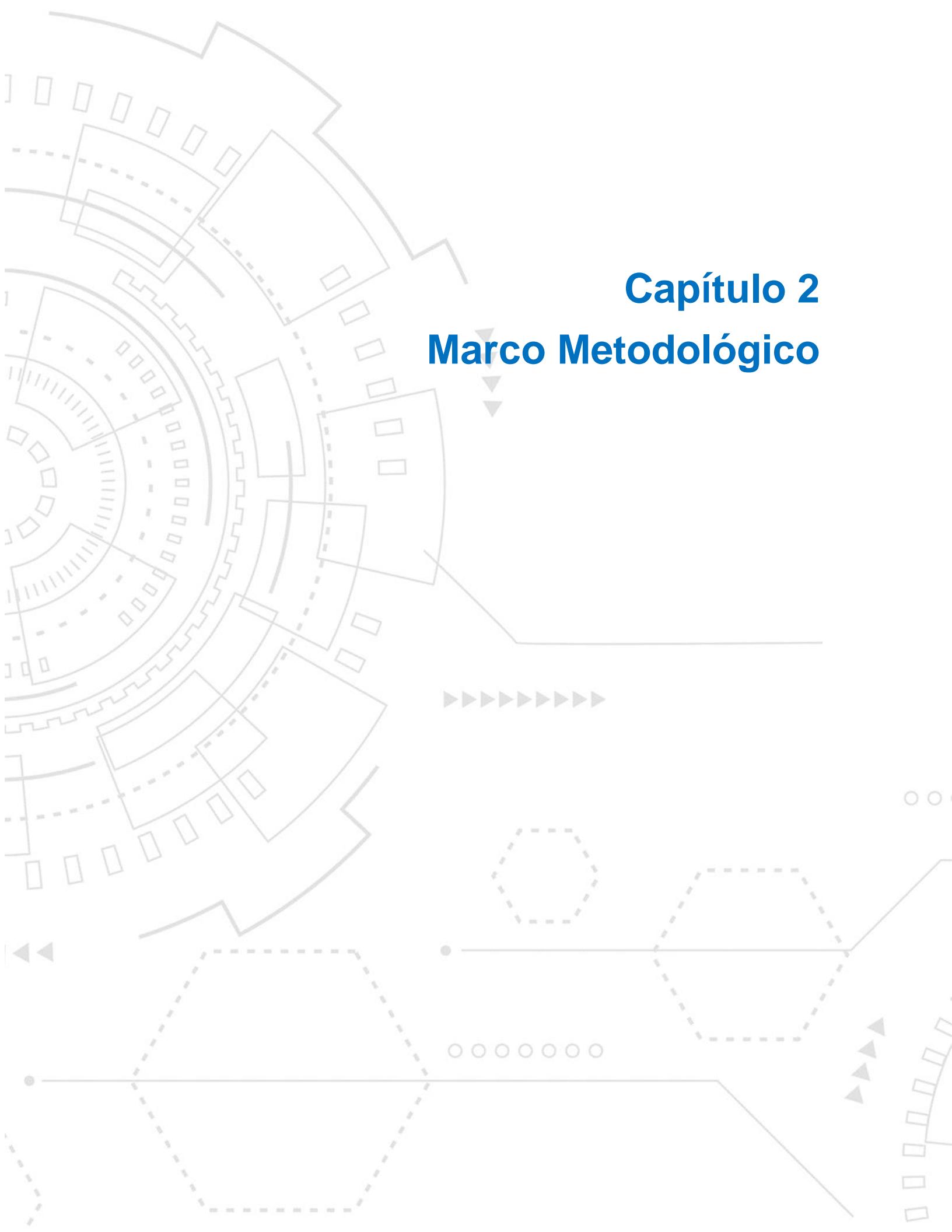
³⁰ La estilización se puede definir como la capacidad de un objeto matemático/estadístico para reproducir la estructura y las formas fundamentales de los fenómenos bajo investigación.

cualquier caso, el valor informativo de los procesos y resultados estadísticos depende de su capacidad para lograr tales estilizaciones estructurales. Asimismo, la arquitectura tradicional de los indicadores socioeconómicos debe ampliarse y complementarse con otros objetos matemáticos/estadísticos, a fin de lograr una imagen más integral y fidedigna de los sistemas sociales/humanos.

En términos generales, se pueden identificar al menos dos aspectos principales que contribuyen a la complejidad de elaborar indicadores socioeconómicos adecuados (Maggino, 2017):

El aspecto temporal. Una de las principales características de los sistemas sociales y humanos es su rápida, y a menudo, errática dinámica temporal. Como consecuencia, es de fundamental importancia interceptar rápidamente las señales de cambios estructurales, con el propósito de elaborar con precisión indicadores socioeconómicos fidedignos y actuales. Asimismo, dado que la formulación de políticas normativas responde constantemente a la información provista por los indicadores, es necesario que los indicadores construidos sean sumamente adaptables y reproducibles, esto con el fin de garantizar un flujo continuo de información y retroalimentación.

El aspecto geoespacial. Los sistemas sociales y humanos pueden mostrar una alta variabilidad en el aspecto geoespacial (por ejemplo, las condiciones de vida en diferentes áreas de un mismo país pueden presentar una alta variabilidad). Por lo tanto, no sólo los científicos sociales –y también los creadores de políticas— deben adaptar sus acciones y métodos de análisis a lo largo del tiempo, también deben ajustarlos de acuerdo con el contexto geoespacial que estén tratando. Es decir, los indicadores socioeconómicos deben proporcionar información desagregada, capaz de capturar y resolver efectivamente las diferencias cuantitativas relacionadas con el contexto geoespacial.



Capítulo 2

Marco Metodológico

Capítulo 2. Marco Metodológico

2.1 Definición del concepto de Complejidad Electoral

A pesar de que el concepto de complejidad se ha exportado y utilizado en diversos campos de las ciencias sociales, el uso de éste en los temas político-electorales ha sido escaso o con un enfoque marcadamente computacional.³¹ La presente investigación difiere de estos enfoques, ya que enfatiza un aspecto particular, pero de suma importancia, de los sistemas geoelectORALES: los listados de electores.³²

Como se mencionó previamente en la Introducción, en el sistema geoelectoral mexicano el listado de electores (RFE) está conformado por dos instrumentos: el Padrón Electoral (PE), y la Lista Nominal de Electores (LNE). Para Integrar, Identificar, Mantener y Actualizar estos instrumentos (tareas IIMA) es necesario que el INE efectúe periódicamente procesos de recolección de datos personales, depuración de bases de datos, implementación de estudios de campo para actualizar la cartografía electoral y de elaboración de análisis sobre la dinámica demográfica (Álvarez Hernández y Álvarez Texocotitla, 2020).

Por otra parte, para lograr una adecuada ejecución de las tareas IIMA del RFE, el INE debe considerar y ponderar –durante los procesos de planeación y presupuesto— todos aquellos factores que complican u obstruyen la realización de dichas tareas. Es decir, se debe analizar y cuantificar la complejidad que se encuentra inherente en el sistema geoelectoral mexicano, específicamente la complejidad que se encuentra asociada a las características geoespaciales, demográficas y estructurales que presentan cada una de las unidades territoriales base del sistema (Álvarez Hernández y Álvarez Texocotitla, 2020).

³¹ Por ejemplo, Faliszewski, Hemaspandra, Hemaspandra y Rothe (2006) proporcionan una exposición extensa de los estudios sobre complejidad en los sistemas electORALES, realizados desde una perspectiva computacional y teórica. Asimismo, en Alvarez Hernández y Alvarez Texocotitla (2020) se presenta un análisis del sistema geoelectoral mexicano desde una perspectiva de sistemas complejos.

³² En las conclusiones se abundará con mayor detalle en las conexiones que se pueden establecer entre el concepto de complejidad electoral, y otras definiciones de complejidad propuestas en diversos campos de estudio.

Consecuentemente, y considerando los antecedentes y la motivación del problema de investigación planteado en este trabajo, se puede definir el concepto de complejidad electoral como: **el conjunto de factores geoespaciales, demográficos y estructurales que impactan directamente en el desarrollo y ejecución de las tareas IIMA del RFE.**³³

Es decir, el concepto heurístico de complejidad electoral que se está acotando en esta investigación incorpora primordialmente los siguientes factores: las características geoespaciales, la dinámica demográfica y la actualidad de los instrumentos registrales.

Es importante mencionar que el énfasis en los factores geoespaciales, demográficos y estructurales es producto de las limitaciones de información (principalmente a escala seccional), ya que el objetivo primordial de esta investigación es proporcionar una definición concreta y cuantificable del concepto de complejidad electoral. Por lo tanto, en esta investigación se dejan de lado aspectos importantes del sistema geoelectoral mexicano (por ejemplo, la participación electoral de la ciudadanía, o el grado de competencia partidista) debido a la falta de indicadores estadísticos detallados que permitan incorporar estos aspectos en el concepto de complejidad electoral definido.³⁴

Considerando la definición de complejidad electoral propuesta, surge la problemática de cómo cuantificar el concepto para cada una de las componentes base del sistema geoelectoral mexicano. Para solucionar este problema se identifican, en primera instancia, los indicadores base que describen la dinámica de los factores que determinan la complejidad electoral, ya que estos serán los insumos de información principales para construir un nuevo conjunto de índices/métricas diseñado expresamente para cuantificar el grado de complejidad electoral

³³ En Álvarez Hernández y Álvarez Texocotitla (2020) se realiza un primer planteamiento de esta definición de complejidad electoral. Asimismo, en Instituto Nacional Electoral [INE] (2017) también se plantea una conceptualización sobre la complejidad electoral; sin embargo, dicha conceptualización tiene un enfoque puramente operativo para propósitos de clasificación de los distritos electorales.

³⁴ Asimismo, se dejan de lado factores coyunturales que impactan directamente en las tareas IIMA del RFE (por ejemplo, los eventos de la pandemia ocasionada por el SARS-COV-2), ya que incorporar dichos factores incrementaría la dificultad del análisis significativamente. No obstante, esto no implica que deba descartarse una futura investigación en esta dirección.

observado en los municipios y secciones electorales. Asimismo, estos índices/métricas servirán como base para la elaboración de una tipología de complejidad electoral a escala municipal y seccional.

2.2 Los indicadores estadísticos para cuantificar la Complejidad Electoral

La definición heurística de complejidad electoral que se propone en esta investigación permite clasificar los indicadores estadísticos que el INE monitorea constantemente para la planeación y realización de las tareas IIMA del RFE. En particular, se pueden distinguir las siguientes tres categorías de indicadores (Álvarez Hernández y Álvarez Texocotitla, 2020):

Indicadores geoespaciales. Estos indicadores describen primordialmente factores como la densidad y la dispersión geográfica de la población empadronada. Dichos factores deben ser continuamente considerados en las labores de integración y actualización del RFE, ya que proporcionan información crucial para la realización de tareas institucionales relacionadas con los procesos electorales (por ejemplo, los operativos de instalación de casillas, actualización de la cartografía electoral, levantamiento de encuestas y consultas, y repartimiento de los Módulos de Atención Ciudadana).³⁵

Indicadores demográficos. Estos indicadores se encuentran asociados fundamentalmente con factores de la dinámica demográfica de los municipios y las secciones electorales; concretamente, con los cambios demográficos advertidos en la LNE y el PE. El análisis de la dinámica demográfica es un elemento fundamental para determinar el grado de complejidad electoral de los municipios y secciones electorales, sobre todo considerando que los programas de identificación, actualización y mantenimiento del RFE (por ejemplo, las campañas de credencialización, la identificación de flujos migratorios anómalos y los programas

³⁵ Es importante señalar que a nivel municipal hay otros indicadores que podrían incluirse (por ejemplo, la cobertura de la red vial o estimaciones de tiempo de traslado entre localidades y la cabecera municipal). Sin embargo, a nivel de secciones electorales, no hay elaboración de indicadores geoespaciales similares a los encontrados a nivel municipal. Consecuentemente, la investigación se realizó sobre un conjunto de indicadores estadísticos estandarizado tanto a nivel municipal como a nivel seccional.

de resecciónamiento³⁶) deben acoplarse correctamente a las distintas dinámicas demográficas que se observan en el sistema geoelectoral.

Indicadores estructurales. Estos indicadores están relacionados fundamentalmente con las actividades de mantenimiento y actualización del RFE (por ejemplo, el nivel de correspondencia entre la LNE y el PE es un indicador que refleja el grado de actualización del RFE). Es importante advertir que las unidades territoriales que presentan un diagnóstico deficiente en estos indicadores reciben usualmente atención prioritaria y una mayor asignación de recursos institucionales, esto con el fin de solventar las deficiencias detectadas en los indicadores, y con ello garantizar la consistencia del RFE para la realización de los procesos electorales. En consecuencia, los indicadores estructurales son fundamentales para determinar la complejidad electoral en las escalas municipal y seccional.

A partir de estas tres categorías de indicadores se puede identificar un conjunto base de siete indicadores que aportan información crítica para la cuantificación de la complejidad electoral en las escalas municipal y seccional.³⁷ A continuación se presenta una breve descripción, así como las fórmulas, de estos siete indicadores base que forman la base de datos, a partir de la cual, se construye el conjunto de índices/métricas que cuantifican la complejidad electoral:³⁸

1) Magnitud del Electorado (LNE). Este indicador demográfico refleja el número de ciudadanos que residen en los municipios y secciones electorales, y que presentan además una CPV vigente (es decir, el número de ciudadanos que están en la posibilidad de ejercer su voto). Este indicador es constantemente monitoreado por el INE, ya que las actividades relacionadas con la integración y actualización del RFE dependen de la dinámica demográfica para la elaboración de los planes presupuestales y de organización electoral; y en particular, para la delimitación

³⁶ Resecciónamiento se refiere al proceso de modificar los límites geográficos de las secciones electorales, esto con el fin de cumplir la normatividad de tener un máximo de 3,000 ciudadanos residentes por sección (LEGIPE, 2020, artículo 147).

³⁷ Es importante puntualizar que los indicadores se estimaron en dos niveles de desagregación territorial (es decir, a nivel municipal y seccional), con excepción del indicador Grado de Urbanización Municipal, el cual sólo corresponde al análisis realizado a escala municipal.

³⁸ Véase Álvarez Hernández (2020) para consultar los archivos anexos del repositorio de GitHub, donde se explicitan los detalles técnicos del procesamiento de los indicadores base.

territorial de las secciones electorales, ya que sus límites geográficos se diseñan considerando un límite máximo de 3,000 ciudadanos por sección electoral.

En consecuencia, desde la perspectiva de las tareas IIMA, se considera que las unidades territoriales que poseen un electorado cuantioso presentan un mayor grado de complejidad electoral en comparación con las unidades territoriales que poseen un electorado de menor magnitud, ya que es necesario dedicar mayores recursos institucionales para manejar adecuadamente los altos volúmenes de trabajo asociados a la mayor cantidad de ciudadanos que deben ser atendidos.

Para la obtención de este indicador se utiliza el corte de información de la LNE a diciembre de 2019.³⁹

$$LNE^l \equiv L_{12,2019}^l. \quad (1)$$

2) Densidad del Electorado (Densidad_LNE). La densidad poblacional del electorado es un indicador geoespacial que informa sobre el número de ciudadanos registrados en la LNE por kilómetro cuadro de área urbanizada;⁴⁰ es decir, este indicador aporta información sobre el grado de dispersión geográfica del electorado.

Desde la perspectiva de las tareas IIMA se considera que aquellos municipios y secciones electorales que poseen una densidad notoriamente baja (es decir, una alta dispersión geográfica del electorado) presentan un alto grado de complejidad electoral, ya que su baja densidad permite deducir que las unidades territoriales son de naturaleza rural, lo cual implica potencialmente resolver diversas dificultades logísticas para ejecutar adecuadamente las tareas IIMA. En contraste, para el caso de aquellos municipios o secciones electorales que poseen una densidad notoriamente alta (es decir, una alta concentración geográfica del electorado), se puede deducir que las unidades territoriales son de naturaleza urbana, lo cual conlleva diversas ventajas logísticas para el desarrollo de las tareas IIMA.⁴¹

³⁹ $L_{12,2019}^l$ ≡ número de ciudadanos inscritos en la LNE del municipio o sección electoral l , al corte de diciembre de 2019.

⁴⁰ $A_{12,2019}^l$ ≡ área superficial (kilómetros cuadrados) del amranamiento urbano del municipio o sección electoral l , al corte de diciembre de 2019.

⁴¹ Por ejemplo, en las zonas con una alta dispersión geográfica del electorado es necesario habilitar Módulos de Atención Ciudadana itinerantes (es decir, que visiten periódicamente las localidades

Para obtener este indicador se utilizaron las estimaciones de la LNE, así como la información cartográfica de los municipios y secciones electorales catalogados en las bases de datos del INE (específicamente, se utiliza la estimación del área superficial cubierta por el amanazamiento para obtener el área urbanizada por municipio y por sección electoral):⁴²

$$Densidad_LNE^l \equiv \frac{L_{12,2019}^l}{A_{12,2019}^l}. \quad (2)$$

3) Coeficiente de Actualización del RFE (Razon_LNE_PE). Este indicador estructural refleja el nivel de actualización de la LNE y el PE, ya que indica la proporción de ciudadanos que mantienen un registro georreferenciado vigente para votar (es decir, es la razón entre el número de ciudadanos registrados en la LNE, y el número de ciudadanos registrados en el PE).⁴³ Consiguentemente, este indicador ayuda a explicar el nivel de complejidad electoral asociado al grado de actualización del RFE, ya que se considera que aquellos municipios y secciones electorales que presentan una baja valoración en este indicador poseen un mayor grado de complejidad electoral, pues es necesario implementar con mayor prontitud mecanismos institucionales que refuercen la realización de las tareas de mantenimiento y actualización del RFE.

Para la obtención de este indicador se utiliza el corte de información de la LNE y el PE a diciembre de 2019:

$$Razon_LNE_PE^l \equiv \frac{L_{12,2019}^l}{P_{12,2019}^l}. \quad (3)$$

4) Tasa de Crecimiento Anual de la LNE (TC_LNE_2019). Este indicador da cuenta del crecimiento (o decremento) demográfico que se observa en los municipios y secciones electorales. Se espera una mayor complejidad electoral en

rurales), esto con el fin de garantizar a todos los ciudadanos el acceso igualitario y universal a los trámites de credencialización y actualización de datos.

⁴² Al utilizar la información cartográfica del amanazamiento se obtiene una estimación más precisa de la superficie que se encuentra urbanizada en los municipios y secciones electorales.

⁴³ $P_{12,2019}^l \equiv$ número de ciudadanos registrados en el PE del municipio o sección electoral l , al corte de diciembre de 2019.

los municipios y secciones electorales de rápido crecimiento demográfico, ya que dicho fenómeno impacta en los procesos integración, identificación y actualización del RFE. Asimismo, es de particular importancia monitorear el crecimiento demográfico en las secciones electorales, ya que esto puede provocar potencialmente el rebase del límite demográfico establecido en la normatividad electoral. Por otra parte, la tasa de crecimiento de la LNE también puede reflejar el impacto de otros fenómenos sociales, como es el caso de la migración forzada por circunstancias económicas o de seguridad pública, fenómenos que también contribuyen a generar una mayor complejidad electoral.

El cálculo de este indicador se realizó en forma anualizada; es decir, se calculó la tasa de crecimiento demográfico observada entre enero y diciembre de 2019:⁴⁴

$$TC_LNE_{2019}^l \equiv \frac{L_{12,2019}^l - L_{1,2019}^l}{L_{1,2019}^l}. \quad (4)$$

5) Coeficiente de Variación de la LNE (Coef_Var_LNE). Este indicador demográfico, obtenido a partir del análisis de las series de tiempo de la LNE (a escala municipal y seccional),⁴⁵ proporciona información acerca de la variación demográfica observada en los municipios y secciones electorales.⁴⁶ En concreto, se considera que hay un mayor grado de complejidad electoral en los municipios y secciones electorales sujetos a un mayor nivel de variación poblacional, lo cual inherentemente afecta la planeación institucional referente a los procesos de identificación y actualización del RFE. Asimismo, las unidades territoriales que

⁴⁴ $L_{1,2019}^l$ ≡ número de ciudadanos inscritos en la LNE del municipio o sección electoral l , al corte de enero de 2019.

⁴⁵ $L_{i,2019}^l$ ≡ número de ciudadanos inscritos en la LNE del municipio o sección electoral l , al corte del mes i de 2019.

⁴⁶ Existen otros indicadores que también se utilizan para describir las variaciones temporales de la dinámica poblacional; en particular, una medida comúnmente usada en el campo de la ecología es el Coeficiente de Variación Proporcional. Para más detalles se recomienda ver Heath y Borowski (2013), y Fernández-Martínez et al. (2018). En esta investigación se analizó la pertinencia de incorporar el Coeficiente de Variación Proporcional; sin embargo, los resultados del análisis exploratorio de datos mostraron que había una alta correlación entre dicho coeficiente y el Coeficiente de Variación, por lo que se consideró innecesario incorporarlo al conjunto de indicadores base. Los detalles del análisis exploratorio de datos se pueden consultar en los archivos adjuntos del repositorio de GitHub de Álvarez Hernández (2020).

presentan una mayor variación poblacional son más complejas desde un punto de vista electoral, ya que es más difícil obtener estimaciones y pronósticos demográficos certeros para efectos de la planeación institucional.

$$Coef_Var_LNE^l \equiv \frac{\sigma_{L,2019}^l}{\bar{L}_{2019}^l} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (L_{i,2019}^l - \bar{L}_{2019}^l)^2}}{\sum_{i=1}^N \frac{L_{i,2019}^l}{N}}. \quad (5)$$

6) Autocorrelación Mensual de la LNE (Autocorr_LNE). La autocorrelación es una métrica que mide el grado de similitud entre una serie temporal y una versión pasada de sí misma. Es decir, la autocorrelación es similar a calcular la correlación entre dos series de tiempo diferentes, excepto que la autocorrelación usa la misma serie de tiempo dos veces, una en su forma original y la otra rezagada uno o más períodos de tiempo. Para los propósitos de cuantificar la complejidad electoral, es de particular interés calcular la autocorrelación de la LNE con un rezago mensual (es decir, 1 retraso, o el grado de similitud entre el comportamiento de un mes respecto al anterior), ya que la elaboración de los pronósticos de la LNE (para determinar, entre otras cosas, la demanda potencial de CPV) se realizan ordinariamente con una periodicidad mensual.⁴⁷

En resumen, se considera que los municipios y secciones electorales con un nivel de autocorrelación mensual bajo presentan un mayor nivel de complejidad electoral en comparación con los municipios y secciones electorales con un nivel de autocorrelación mensual alto, ya que el comportamiento temporal de su LNE es más irregular.

$$Autocorr_LNE^l \equiv r_k^l = \frac{\sigma_k^l}{\sigma_{L,2019}^l} = \frac{\frac{1}{N} \sum_{i=1}^{N-k} (L_{i,2019}^l - \bar{L}_{2019}^l)(L_{i+k,2019}^l - \bar{L}_{2019}^l)}{\frac{1}{N} \sum_{i=1}^N (L_{i,2019}^l - \bar{L}_{2019}^l)^2} \text{ } ^{48} \quad (6)$$

⁴⁷ Esto debido a la calendarización del sistema de información de los Módulos de Atención Ciudadana.

⁴⁸ $N \equiv 12$ meses.

7) Grado de Urbanización Municipal (NUM_SECC_RUR). Este indicador geoespacial da cuenta del grado de urbanización observado a escala municipal, ya que cuantifica el número secciones electorales de tipo rural que posee cada municipio.⁴⁹ En consecuencia, se considera que los municipios con un número elevado de secciones rurales presentan un mayor nivel de complejidad electoral, debido a que la baja magnitud del electorado, junto con la alta dispersión geográfica, trae consigo dificultades logísticas para la adecuada realización de las tareas IIMA del RFE.

$$NUM_SECC_RUR^l \equiv cardinalidad(secc_rur \in l) \quad (7)$$

En el Cuadro 2 se resume la información y características de los siete indicadores base descritos previamente.

⁴⁹ El INE define las secciones rurales como aquellas secciones electorales que están conformadas enteramente por localidades con una población menor a los 1,500 ciudadanos. Véase INE (2017).

Tipo	Indicador	Observaciones
Geoespacial	Grado de Urbanización Municipal (NUM_SECC_RUR)	Calculada con información cartográfica, corte diciembre de 2019
	Densidad del Electorado (Densidad_LNE)	Calculada con información cartográfica y de LNE, corte diciembre de 2019
Demográfica	Magnitud del Electorado (LNE)	Corte diciembre de 2019
	Tasa de Crecimiento Anual de la LNE (TC_LNE_2019)	Calculada con base en los cortes de la LNE de enero y diciembre de 2019
	Coeficiente de Variación de la LNE (Coef_Var_LNE)	Calculado con base en la serie de tiempo de la LNE, entre enero de 2019 y diciembre de 2019
Estructural	Coeficiente de Actualización del RFE (Razon_LNE_PE)	Calculado con información del PE y de la LNE, corte diciembre de 2019
	Autocorrelación Mensual de la LNE (Autocorr_LNE)	Calculado con base en la serie de tiempo de la LNE entre enero de 2019 y diciembre de 2019

Cuadro 2. *Indicadores base para cuantificar la Complejidad Electoral*⁵⁰

Fuente: Elaboración propia

Posterior a la selección de indicadores base, se realiza un análisis exploratorio de datos para estos indicadores, con el propósito de caracterizar sus comportamientos individuales y evaluar la pertinencia de su inclusión para la cuantificación de la complejidad electoral.

Específicamente, el análisis exploratorio de datos que se desarrolla en este trabajo se aboca primordialmente a la realización de análisis de correlación (por medio de la elaboración de correlogramas y mapas de calor),⁵¹ los cuales permiten

⁵⁰ Como se señaló anteriormente, seis de los siete indicadores base (Densidad_LNE, LNE, TC_LNE_2019, Coef_Var_LNE, Razon_LNE_PE y Autocorr_LNE) se estimaron tanto a escala municipal como a escala seccional. Únicamente el indicador NUM_SECC_RUR se calculó a escala municipal.

⁵¹ En el Capítulo 3 se muestran los resultados el análisis exploratorio a nivel municipal, y en el Capítulo 4 se muestran los resultados el análisis exploratorio a nivel seccional.

analizar las interrelaciones existentes entre los indicadores base.⁵² Asimismo, se realiza un análisis visual de las distribuciones de datos (por medio de la elaboración de histogramas y diagramas de violín) con el propósito de caracterizar el comportamiento individual de cada indicador, e identificar aquellos indicadores que poseen distribuciones leptocúrticas,⁵³ lo cual es un indicio de la probable existencia de casos atípicos de complejidad electoral.

2.3 Transformación de datos y elaboración de los Índices de Complejidad Electoral

Como se comentó con detalle en el Capítulo 1, el paradigma actual de la elaboración de indicadores socioeconómicos se caracteriza por incorporar aspectos de la perspectiva holística, la cual consiste en preservar la naturaleza multifactorial y multidimensional de los sistemas, con el propósito de definir indicadores globales de cuantificación que hagan un uso mínimo de métodos de agregación o ponderación subjetiva. Asimismo, estos indicadores deben ser capaces de soportar cambios estructurales o de escala en la información, a la vez que deben permitir un monitoreo continuo de la evolución temporal y geoespacial de los sistemas (Álvarez Hernández y Álvarez Texocotitla, 2020).

Por lo tanto, para obtener un conjunto de métricas que cuantifiquen la complejidad del sistema geoelectoral a escala municipal y seccional, y que además se apegue al paradigma actual de elaboración de indicadores socioeconómicos, se construye, a partir de la información contenida en los siete indicadores base, un conjunto de Índices de Complejidad Electoral (ICE) a nivel municipal y seccional.

Los ICE se pueden definir formalmente como: “**números índice que valoran el grado de complejidad electoral presente en una unidad territorial del sistema geoelectoral a partir de la síntesis (ponderación) de la información contenida en un grupo diverso de indicadores**” (Álvarez Hernández y Álvarez

⁵² En función de los resultados obtenidos en los ICE, se analizaron diferentes escenarios de indicadores. En particular, el grupo de siete indicadores seleccionados reportó los mejores resultados. Para más detalles se recomienda revisar los archivos anexos mostrados en Álvarez Hernández (2020).

⁵³ Una distribución más pronunciada alrededor de la media, y con colas más gruesas, que la distribución normal.

Texocotitla, 2020, p. 227). Es decir, los ICE cuantifican en un solo valor numérico el grado de complejidad electoral asociado al comportamiento (representado en los siete indicadores base) de los factores geoespaciales, demográficos y estructurales que determinan la complejidad electoral (Álvarez Hernández y Álvarez Texocotitla, 2020).

Considerando la anterior definición, se deriva que los ICE son indicadores que permiten establecer relaciones de orden y magnitud entre las diferentes unidades territoriales del sistema geoelectoral, ya que posibilitan la elaboración de una lista ordenada (*ranking*) de municipios y secciones electorales en función de su grado de complejidad electoral. Asimismo, los ICE también posibilitan el análisis temporal de la complejidad electoral (por ejemplo, si se recalcularán los ICE en un periodo de tiempo posterior), ya que permiten medir las variaciones de complejidad electoral ocasionadas por cambios en el comportamiento de los indicadores base.

Para la elaboración de los ICE es necesario utilizar una técnica de reducción de dimensionalidad, ya que a partir del conjunto de siete indicadores base se desea obtener un conjunto de índices que ponderen, en una forma no arbitraria, la influencia de cada indicador en la determinación de la complejidad electoral.

Consecuentemente, para construir los ICE a escala municipal y seccional, se utiliza la técnica de Análisis de Componentes Principales (PCA por *Principal Component Analysis*). Esta técnica de análisis multidimensional de datos es bastante estándar en los campos de la estadística multivariada y la minería de datos, y ha sido ampliamente utilizada en diversos contextos de estudio que implican la reducción o transformación de datos para la obtención de índices sintéticos.⁵⁴ Algunas de las características del PCA son las siguientes (Jackson, 1991; Abdi y Williams, 2010):

1) Reduce la dimensionalidad de grandes conjuntos multivariados de datos a través de un proceso de transformación lineal, el cual consiste en mapear el conjunto original de variables a un nuevo conjunto transformado de variables (los llamados Componentes Principales). Es decir, el PCA permite retener y reflejar el

⁵⁴ Véase Shin et al. (2020), o Consejo Nacional de Población [CONAPO] (2002), para ver una ejemplificación de la implementación del Análisis de Componentes Principales para un problema de construcción de indicadores sociales.

máximo de información en un subespacio dimensional conformado por un nuevo conjunto de variables, el cual sintetiza las características y patrones esenciales del conjunto original de datos (Álvarez Hernández y Álvarez Texocotitla, 2020).

2) Estima en una forma no arbitraria los coeficientes de ponderación de las variables originales, ya que las Componentes Principales representan una combinación lineal de las variables originales. Es decir, las Componentes Principales pueden interpretarse como una suma ponderada de todas las variables, cuyos coeficientes de ponderación se estiman a partir de la diagonalización de la matriz de correlación y la posterior obtención de los *eigenvalores* y *eigenvectores* correspondientes a la matriz.

3) Simplifica el análisis y la interpretación visual de los conjuntos de datos de alta dimensionalidad (tres o más dimensiones), ya que se puede utilizar un subconjunto de las Componentes Principales para proyectar el conjunto de datos original a un subespacio de variables transformado más reducido e interpretable (Álvarez Hernández y Álvarez Texocotitla, 2020).⁵⁵

4) Genera indicadores concisos que permiten tipificar los datos. Específicamente, se puede utilizar la primera Componente Principal como índice de ordenación cardinal, ya que esta nueva variable explica, comúnmente, la mayor parte de la varianza encontrada en los datos (Álvarez Hernández y Álvarez Texocotitla, 2020).

Resumiendo, el PCA es un método de transformación lineal que permite manejar la multidimensionalidad del fenómeno estudiado, a la vez que proporciona mediante la consideración de las Componentes Principales (CP), un conjunto de índices que sintetizan las características esenciales del conjunto original de datos, y que ponderan en una forma no subjetiva las contribuciones de cada variable original.

⁵⁵ El PCA, junto con el método *Singular Value Decomposition* (SVD), son las técnicas de proyección lineal más utilizadas para la generación de índices e indicadores. Por otra parte, existe otro conjunto de técnicas conocidas como métodos de proyección no-lineal; entre los que se encuentra el método de Principal Curves (este método se puede considerar como una generalización del PCA para hallar Componentes Principales no-lineales). Para una exposición más completa y rigurosa sobre los métodos de proyección lineales y no-lineales, se recomienda revisar Gniazdowski (2017), Jackson (1991), Abdi y Williams (2010), Hastie y Stuetzle (1989) y Jolliffe (2002).

Para el caso concreto de la problemática tratada en esta investigación – cuantificar la complejidad del sistema geoelectoral a escala municipal y seccional—, el PCA permite cuantificar la naturaleza multifactorial del concepto de complejidad electoral, ya que el cálculo de las Componentes Principales permite obtener una valoración que pondera el efecto de los factores geoespaciales, demográficos y estructurales sobre el grado de complejidad electoral observado en un municipio o sección electoral. En otras palabras, cada Componente Principal representa un ICE (es decir, una suma ponderada específica de los indicadores base), y la valoración que obtenga cada municipio o sección electoral en los ICE representa un grado de complejidad electoral exclusivo a dicho índice. En notación matemática, el grado de complejidad electoral se puede representar como:

$$ICE_i^l \equiv \sum_{j=1}^n L_i^j x_j^l.^{56} \quad (8)$$

Al respecto, es importante señalar que el método PCA puede proporcionar un número de Componentes Principales menor o igual al número de variables originales. Cada Componente Principal logra explicar un determinado porcentaje de la varianza encontrada en el conjunto de datos original. Para la elección de la cantidad de Componentes Principales extraídos a partir del PCA, existen diversos criterios. En particular, la presente investigación utiliza el criterio porcentual de la parte de la varianza explicada por Componente Principal, ya que dicho criterio permite evaluar la capacidad de síntesis de las componentes y, por ende, su respectiva importancia.⁵⁷

⁵⁶ ICE_i^l ≡ grado de complejidad electoral del municipio o sección electoral i , correspondiente al índice de complejidad electoral i ,

L_i^j ≡ ponderador correspondiente a la variable original j , del índice de complejidad electoral i ,

x_j^l ≡ valor del municipio o sección electoral i en la variable original j ,

n ≡ número de variables originales introducidas en el PCA.

⁵⁷ Dado que cada uno de los criterios puede sugerir diferentes números de componentes, la decisión final sobre sus números es tomada generalmente por el investigador, ya que este debe considerar el contexto y las características particulares de los datos analizados para determinar el subconjunto adecuado de Componentes Principales. Para ver los criterios existentes se recomienda revisar Gniazdowski (2017).

Como se mostrará con detalle en los Capítulos 3 y 4, los resultados del criterio de varianza explicada permiten justificar la elección de la primera Componente Principal como el principal ICE –tanto a escala municipal como a escala seccional—, y por lo tanto, se puede construir un *ranking* nacional de complejidad electoral a partir de dicho índice.⁵⁸ Las subsiguientes Componente Principales (es decir, los subsiguientes ICE) también se utilizan para obtener valoraciones alternativas de la complejidad electoral, ya que cada ICE aporta una diferente estructura de ponderación de los indicadores base.⁵⁹

A pesar de las ventajas que ofrece el PCA en el aspecto de la reducción de dimensionalidad –y en la proyección de los datos en un subespacio de indicadores más consistente— es importante examinar cuidadosamente los resultados obtenidos con dicho método, esto con el fin de valorar la capacidad explicativa de las Componentes Principales y su interpretación en función de las variables originales. Asimismo, se debe tomar en cuenta que los resultados inferenciales que se obtienen a partir del PCA corresponden, normalmente, a un modelo de efectos fijos; es decir, las observaciones se consideran la población de interés, y las conclusiones se limitan a estas observaciones específicas.⁶⁰

Como se mostrará en los siguientes capítulos, los resultados de la utilización del método PCA –tanto para el caso municipal, como para el caso seccional— indican que no se puede usar únicamente el ICE principal para generar un *ranking* de complejidad electoral consistente, ya que dicho índice no alcanza un nivel adecuado de síntesis de información para los indicadores base. Es decir, considerando el porcentaje de varianza explicada por la Primera Componente Principal, no está plenamente justificado utilizar exclusivamente el ICE principal

⁵⁸ Es muy frecuente observar que la primera Componente Principal sea la componente que muestre el mayor porcentaje de varianza explicada, ya que por construcción, el PCA ordena las Componentes Principales en función de la magnitud de sus *eigenvalores*. Es decir, la primera Componente Principal es el *eigenvector* que tiene asociado el mayor *eigenvalor*.

⁵⁹ Por ejemplo, considerando el caso municipal de los siete indicadores base, el método PCA puede proporcionar un máximo de siete ICE. Cada uno de estos índices proporcionará ponderadores específicos y únicos a los siete indicadores base.

⁶⁰ Es el caso de la presente investigación, ya que la población de interés corresponde al total de datos (en los siete indicadores base) de los municipios y secciones electorales existentes en el sistema geoelectoral mexicano.

como métrica de comparación de los grados de complejidad electoral que presenta cada municipio y sección electoral.

En este contexto, se puede solventar la falta de solidez estadística del ICE principal mediante el uso simultáneo de la información contenida en el conjunto total de índices de complejidad electoral (o en un subconjunto que logre una mayor explicación de la varianza observada en los indicadores base). Es decir, la proyección de los municipios y secciones electorales en el espacio de transformación de una o más Componentes Principales permite elaborar una tipología de municipios y secciones electorales, la cual ofrece un método de clasificación más robusto de la complejidad electoral, en comparación con un *ranking* de complejidad sustentado únicamente en el ICE principal.⁶¹

2.4 Agrupamiento de datos y elaboración de la Tipología de Complejidad Electoral

La agrupación de datos es el proceso de clasificar diferentes objetos en grupos, de tal manera que la similitud entre dos objetos sea máxima si pertenecen al mismo grupo, y sea mínima en caso contrario. La agrupación de datos no hace ninguna distinción entre variables dependientes e independientes, ya que examina todo el conjunto de relaciones interdependientes para establecer relaciones de similitud entre los datos. Los algoritmos de agrupación de datos se pueden agrupar en tres categorías principales (Dean, 2014):

Agrupamiento no supervisado (*unsupervised clustering*). El objetivo de la agrupación no supervisada es maximizar la similitud intra-conglomerado y minimizar la similitud inter-conglomerado dada una medida de similitud/disimilitud. Estos algoritmos utilizan funciones objetivo específicas (por ejemplo, una función que minimiza las distancias intra-clase para encontrar conglomerados concisos). Asimismo, utilizan un conjunto de datos que no poseen variables objetivo (es decir, son datos no etiquetados). Las técnicas de agrupamiento no supervisado más ampliamente utilizadas son K-medias y Agrupamiento Jerárquico.

⁶¹ En los Capítulos 3 y 4 se argumenta con mayor detalle la decisión de tomar el conjunto total de Componentes Principales para la elaboración de la tipología de complejidad electoral.

Agrupamiento semi-supervisado (*semisupervised clustering*). Además de la medida de similitud que también se usa en la agrupación no supervisada, la agrupación semi-supervisada utiliza información adicional de guía/ajuste para mejorar los resultados de la conglomeración. Esta información se utiliza para ajustar la medida de similitud o para modificar la búsqueda de conglomerados.

Agrupamiento supervisado (*supervised clustering*). El objetivo de la agrupación supervisada es identificar agrupaciones que tienen altas densidades de probabilidad con respecto a las clases individuales (conglomerados de clase uniforme). Se usa cuando hay una variable objetivo y un conjunto de entrenamiento que incluye las variables a conglomerar.

Para la cuestión específica de los resultados obtenidos en los ICE —tanto para el caso municipal, como para el caso seccional— se encuentra implícito un problema de agrupamiento no supervisado, ya que el conjunto de datos proveniente de los ICE no se encuentra etiquetado, es decir, carece de variables objetivo.⁶²

Por lo tanto, para la creación de la TCE (a escala municipal y seccional) se analizaron diferentes algoritmos de tipo no supervisado, ya que las características —y el volumen— de los datos implícitos en los ICE municipales y seccionales trae consigo problemas de implementación respecto a las capacidades de procesamiento computacional.⁶³ Favorablemente, la única familia de algoritmos de agrupamiento que no mostró restricción durante las pruebas experimentales fueron los algoritmos de tipo particional, específicamente el método K-medias. Consecuentemente, para la elaboración de las TCE, se utilizó un método de agrupamiento K-medias para clasificar a los municipios y secciones electorales en función de su respectiva valoración en los ICE.

⁶² Para una exposición más completa y rigurosa sobre los algoritmos de agrupamiento se recomienda revisar Dean (2014); Everitt, Landau, Leese y Stahl (2011). Asimismo, para una explicación y demostración amplia de la implementación de los algoritmos de aprendizaje no supervisado en *Python*, se recomienda revisar Pedregosa et al. (2011).

⁶³ Por ejemplo, en el caso de las secciones electorales se encuentra implícito el análisis de agrupamiento para una matriz de datos con 68,390 instancias (renglones) y 6 indicadores distintos (columnas). En el caso de los municipios, se encuentra implícita una matriz de datos con 2,458 instancias y 7 indicadores distintos.

K-medias es un algoritmo iterativo que permite dividir el conjunto de datos en un número predefinido K de subgrupos (cúmulos) distintos y no superpuestos, en los que cada punto de datos (es decir, cada municipio o sección electoral) pertenece a un solo subgrupo. La forma de operar de este algoritmo se fundamenta primordialmente en la asignación de puntos de datos a un cúmulo, de tal manera que la suma de la distancia cuadrada entre los puntos de datos y el centroide del cúmulo (media aritmética de todos los puntos de datos que pertenecen a ese cúmulo) sea mínima.⁶⁴

Para determinar el número óptimo de cúmulos es necesario realizar un análisis de *silhouette score* (también conocido como análisis de coeficientes de silueta), ya que éste permite analizar en forma gráfica (a través del llamado diagrama de silueta), la distancia de separación entre los grupos de una configuración específica,⁶⁵ la cual se especifica por medio de los coeficientes de silueta.⁶⁶ Por ejemplo, si un objeto (perteneciente a un grupo específico) presenta un coeficiente de silueta cercano a +1, esto indica que dicho objeto se encuentra muy alejado de otros objetos pertenecientes a grupos vecinos. En el caso de un coeficiente de silueta igual a 0, significa que el objeto se encuentra en el límite de decisión entre dos grupos vecinos. Finalmente, para el caso de un objeto con un coeficiente de silueta cercano a -1, implica que dicho objeto se encuentra más cercano a algún objeto perteneciente a otro grupo en comparación a los objetos que se encuentran dentro de su mismo grupo de pertenencia.⁶⁷

En resumen, la implementación conjunta del algoritmo K-medias con el análisis de coeficientes de silueta permite determinar el número y la configuración óptima de grupos para tipificar los municipios y secciones electorales en función de

⁶⁴ En los archivos anexos de Álvarez Hernández (2020) se explicitan los detalles técnicos de la inicialización de los parámetros del algoritmo K-medias en Python.

⁶⁵ Existen diferentes tipos de métricas de distancia, pero las más frecuentemente usadas en el algoritmo K-medias son las métricas Euclidiana y Manhattan. Los resultados del análisis exploratorio de los indicadores base no proveyó indicios que favorecieran el uso de alguna métrica específica. En consecuencia, durante el proceso de agrupamiento y análisis de coeficientes de silueta se decidió utilizar la métrica Euclidiana para establecer las distancias entre los puntos de datos, ya que su interpretación geométrica es más intuitiva respecto a otras métricas de distancia.

⁶⁶ Estos coeficientes tienen un rango entre [-1, +1].

⁶⁷ Para una explicación técnica y detallada del funcionamiento del análisis de coeficientes de silueta se recomienda ver Rousseeuw (1987).

las valoraciones que obtuvieron en los ICE. En otras palabras, el análisis de agrupamiento permite construir, en el espacio de proyección de los ICE, una tipología de complejidad electoral tanto a nivel municipal como a nivel seccional.

2.5 Identificación de casos atípicos de complejidad electoral

Desde la perspectiva institucional del INE es importante identificar aquellos municipios y secciones electorales que presentan niveles atípicos de complejidad electoral, ya que su atipicidad es un indicio de que dichas unidades territoriales presentan, por diversas razones, características anómalas respecto al resto de las unidades fundamentales del sistema geoelectoral. Consecuentemente, estos municipios y secciones electorales deben recibir una atención prioritaria, con el fin de optimizar el adecuado desarrollo de las tareas IIMA en un contexto geoespacial, demográfico y estructural anómalo.

Por otra parte, la identificación de casos atípicos de complejidad electoral también tiene una componente de interés para la investigación demográfica, ya que en los ICE y la TCE se encuentra codificado el comportamiento de la dinámica poblacional a nivel municipal y seccional. Es decir, la detección de los casos atípicos también permite identificar aquellas unidades territoriales que presentan un mayor dinamismo poblacional, el cual puede ser ocasionado por los movimientos migratorios internos y/o externos, o por choques exógenos en las tasas de natalidad y mortalidad.⁶⁸

El proceso de identificación de los casos atípicos que se desarrolla en esta investigación consiste en utilizar las agrupaciones generadas por la TCE –tanto a nivel municipal, como a nivel seccional— para posteriormente analizar las distribuciones individuales de cada ICE y determinar, mediante algún criterio estadístico, la condición de atipicidad en la distribución. Es decir, por cada grupo de la TCE se analiza la distribución de valores obtenida por los municipios y secciones

⁶⁸ Una posible línea de investigación puede aprovechar la información codificada en los ICE para estudiar el impacto de la pandemia de SARS-COV-2 en la dinámica poblacional observada a nivel municipal y seccional. Respecto a este tema, ya se han formulado diversas investigaciones, por ejemplo, se recomienda revisar el estudio de Marois, Muttarak y Scherbov (2020).

electorales en los ICE, y se identifican aquellos casos que muestran un grado de complejidad electoral atípico.⁶⁹

Este proceso de identificación de casos atípicos aprovecha el agrupamiento previamente generado por la tipología –el cual clasifica a los municipios y secciones en función de los grados de complejidad electoral obtenidos en cada uno de los ICE—, ya que el objetivo de este ejercicio consiste en distinguir específicamente aquellas unidades territoriales que muestran características notoriamente anómalas, incluso entre aquellos grupos de municipios y secciones que muestran, en promedio, características que propician un alto grado de complejidad electoral. Al proceder de esta forma, se logra incorporar el contexto geoespacial, demográfico y estructural particular de las unidades territoriales en el proceso de separación entre los casos promedio y los casos anómalos.

Por otra parte, existen diferentes criterios estadísticos para determinar la condición de atipicidad de las observaciones en una distribución (Barbato, Barini, Genta y Levi, 2011). En el caso particular de esta investigación se utiliza la regla del rango intercuartílico, ya que su implementación es bastante directa una vez generadas las distribuciones de los ICE por cada grupo de la TCE. La regla del Rango Intercuartílico (IQR) se fundamenta en el cálculo de los cuartiles de una distribución (específicamente los cuartiles Q1 y Q3) para establecer la condición de atipicidad, ya que la regla establece que las observaciones atípicas son aquellas que se encuentran fuera de los siguientes rangos inferior y superior:

$$\text{Rango inferior} \equiv Q1 - 1.5 * IQR = Q1 - 1.5(Q3 - Q1) \quad (9)$$

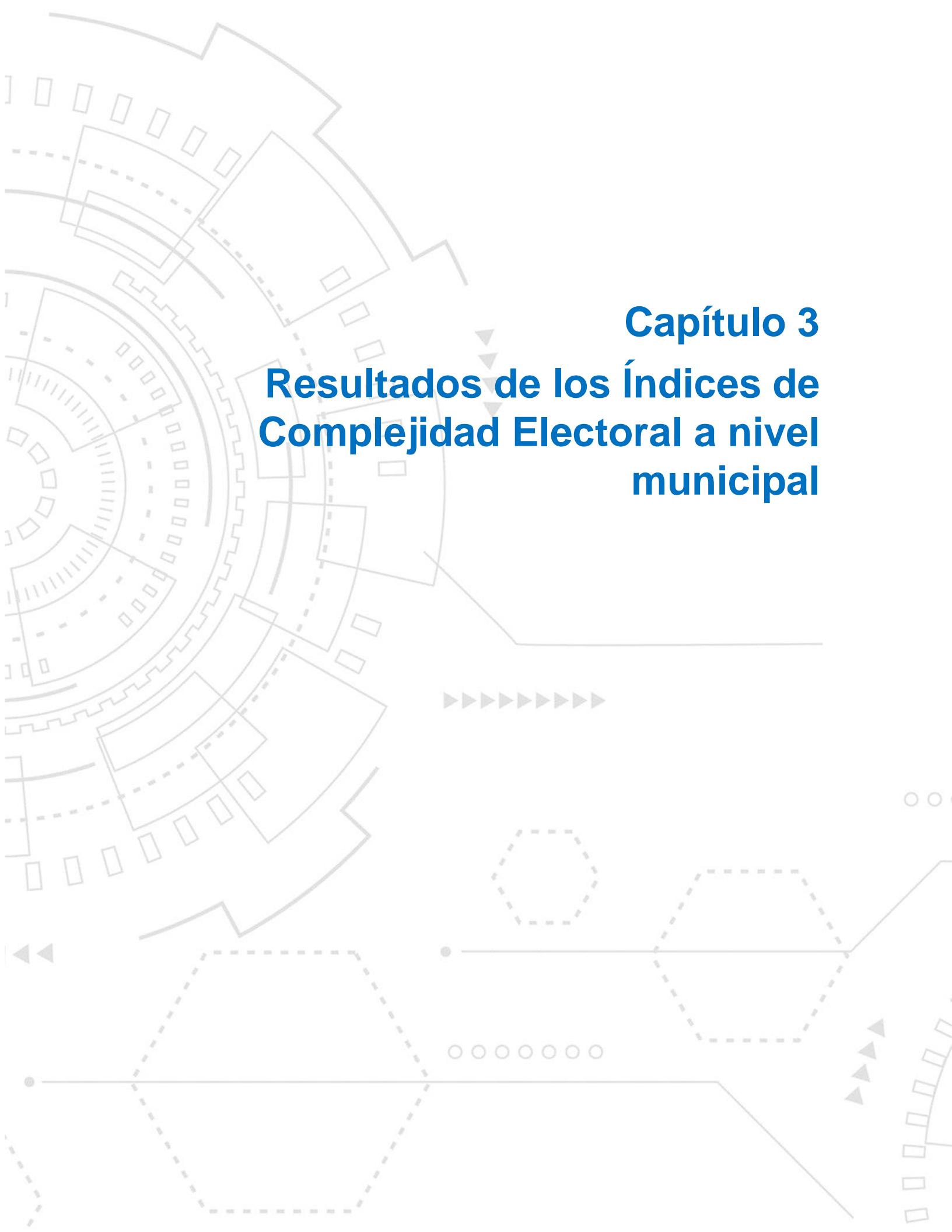
$$\text{Rango superior} \equiv Q3 + 1.5 * IQR = Q3 + 1.5(Q3 - Q1). \quad (10)$$

Es importante puntualizar que la justificación de la utilización de la regla del rango intercuartílico depende en gran medida de la condición de normalidad de la distribución analizada. Asimismo, es elemental considerar que la presencia misma

⁶⁹ Específicamente, se analizan sólo las distribuciones de los tres primeros ICE, ya que, como se explicará con detalle en los Capítulos 3 y 4, los tres primeros ICE concentran cerca del setenta por ciento de la varianza explicada por las Componentes Principales.

de observaciones atípicas puede alterar significativamente los resultados de la regla. En esta investigación no se utilizan métodos más sofisticados para la identificación de observaciones atípicas, ya que el análisis de los resultados obtenidos en la identificación de casos atípicos (ver Capítulos 5 y 6) resultó favorable considerando la información disponible en los indicadores base.⁷⁰

⁷⁰ Se recomienda revisar la exposición de Jackson y Chen (2004), y Sim, Gan y Chang (2005), para obtener un panorama general sobre los diversos métodos existentes para la identificación de observaciones atípicas.



Capítulo 3

Resultados de los Índices de Complejidad Electoral a nivel municipal

Capítulo 3. Resultados de los Índices de Complejidad Electoral a nivel municipal

La construcción de los ICE municipales inicia con un análisis exploratorio de los siete indicadores base.⁷¹ Posteriormente, se aplica la técnica PCA para transformar el conjunto de indicadores a un nuevo grupo de variables (las Componentes Principales). Una vez calculadas las Componentes Principales –y utilizando como métrica de evaluación la varianza explicada— se analizan sus capacidades para sintetizar la información de los indicadores base, y se asignan los ICE municipales a las Componentes Principales resultantes. De esta forma, se obtienen las proyecciones bidimensional y tridimensional de los municipios en el nuevo espacio de transformación conformado por las Componentes Principales (lo cual permite analizar la emergencia de patrones distinguibles de agrupamiento en las proyecciones), se determina la composición de los ICE municipales –es decir, la ponderación que recibe cada indicador base—, y se analizan los resultados geográficos del *ranking* municipal de complejidad electoral (obtenido a partir de la primera Componente Principal).

Es importante mencionar que durante el proceso de construcción de los ICE municipales se analizaron diferentes escenarios y transformaciones de datos con el grupo de indicadores base utilizado a nivel municipal. Específicamente, se analizaron los efectos de algunas transformaciones logarítmicas implementadas en los indicadores con distribuciones altamente asimétricas.⁷² En términos generales, la implementación de estas transformaciones elimina el alto sesgo presente en las distribuciones, pero dicha corrección no aporta alguna mejora significativa en la capacidad de síntesis de información de las Componentes Principales.⁷³ Por lo tanto, en este capítulo sólo se presentan los resultados correspondientes al escenario que mejor resultados proporcionó en el proceso de construcción de los ICE municipales;

⁷¹ Para el caso municipal, se pudo recuperar la información completa de los siete indicadores base para el universo total de 2,458 municipios existentes en la base de datos geoelectoral, de 2019, del INE.

⁷² Es el caso de los indicadores de Magnitud y Densidad del Electorado, caso que se analiza gráficamente a través de histogramas en la siguiente sección del capítulo.

⁷³ Por otra parte, como se explicará con mayor detalle en el Capítulo 5, la utilización de transformaciones logarítmicas para normalizar los indicadores base tiene como consecuencia indirecta, durante el proceso de construcción de la TCE municipal, la eliminación de los casos atípicos de complejidad electoral.

no obstante, en los archivos anexos del repositorio de GitHub se proporcionan al lector los resultados de los escenarios alternativos para su consulta.⁷⁴

3.1 Análisis exploratorio y construcción de los ICE municipales

En el Gráfico 1 se muestra un panel de gráficas (correlogramas) que permite realizar un análisis exploratorio gráfico del conjunto de indicadores base analizado a nivel municipal. En particular, el panel de correlogramas permite relacionar pares de indicadores para identificar la existencia, o ausencia, de correlaciones positivas o negativas. Asimismo, este panel de gráficas también muestra la distribución que presenta cada indicador base (histogramas ubicados sobre la diagonal).⁷⁵ Por ejemplo, el Gráfico 1 permite destacar el comportamiento altamente sesgado de algunos de los indicadores base, como es el caso de los indicadores Grado de Urbanización Municipal (NUM_SECC_RUR), Magnitud del Electorado (LNE) y Densidad del Electorado (Densidad_LNE), ya que estos tres indicadores base presentan una asimetría positiva en sus distribuciones, lo cual implica que en promedio, los municipios se caracterizan por tener un número bajo de secciones electorales rurales, o una baja magnitud y densidad en sus electorados (es decir, municipios de carácter rural); no obstante, hay un número considerable de municipios atípicos que no cumplen con este patrón (municipios de carácter urbano).

Asimismo, el análisis exploratorio gráfico mostrado en el Gráfico 1 también permite apreciar la presencia de correlaciones significativas entre los indicadores.⁷⁶

⁷⁴ Véase Álvarez Hernández (2020).

⁷⁵ A manera de ejemplo, considérese la gráfica del panel localizada en la primera fila y columna (histograma del indicador NUM_SECC_RUR). Dicho histograma muestra que la gran mayoría de municipios tiene un número reducido de secciones rurales (distribución sesgada hacia la izquierda).

⁷⁶ Un ejemplo es la gráfica ubicada en la segunda fila, primera columna (diagrama de dispersión de la LNE graficada contra NUM_SECC_RUR). Este diagrama muestra que los municipios con un número reducido de secciones rurales tienden a tener un mayor electorado, comparado con los municipios que tienen un mayor número de secciones rurales.

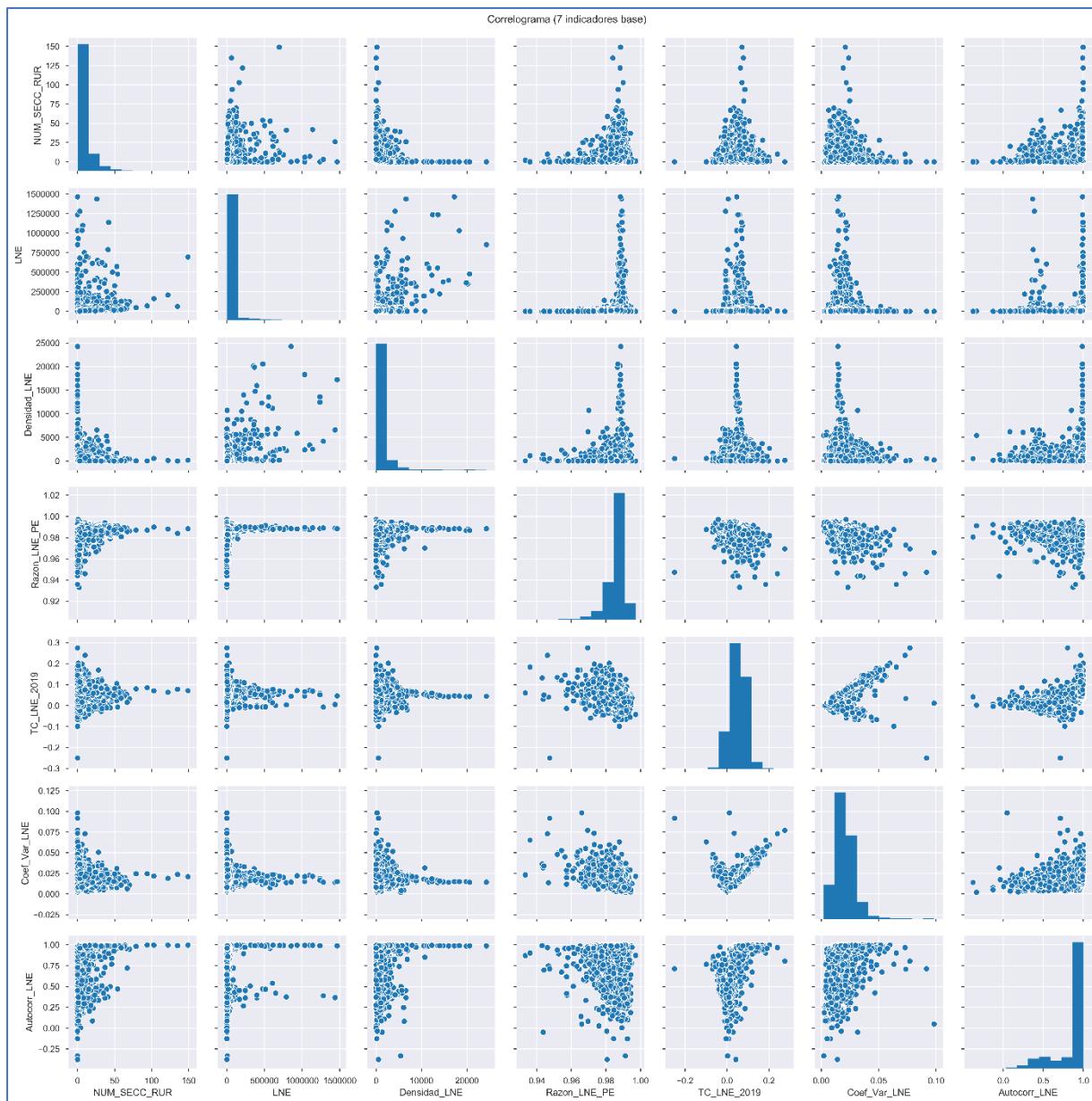


Gráfico 1. Panel de gráficas de correlación y distribución de los indicadores base que conforman los ICE municipales⁷⁷

Fuente: Elaboración propia con la librería Python-Pandas-Seaborn

Por ejemplo, se puede observar la presencia de correlaciones positivas entre la Tasa de Crecimiento Anual de la LNE (TC_LNE_2019) y el indicador de Coeficiente de Varianza (Coef_Var_LNE), y también entre los indicadores LNE y Densidad_LNE.

⁷⁷ Se recomienda revisar el Cuadro 2 para identificar adecuadamente la nomenclatura de los indicadores base.

Por otra parte, también se observa la presencia de correlaciones negativas entre el Coeficiente de Actualización del RFE (Razón_LNE_PE) con los indicadores de variabilidad demográfica (TC_LNE_2019 y Coef_Var_LNE).⁷⁸

La fuerza de las correlaciones se puede visualizar más claramente con un mapa de calor para cada par de indicadores (matriz de correlación), como el que se muestra en el Gráfico 2.

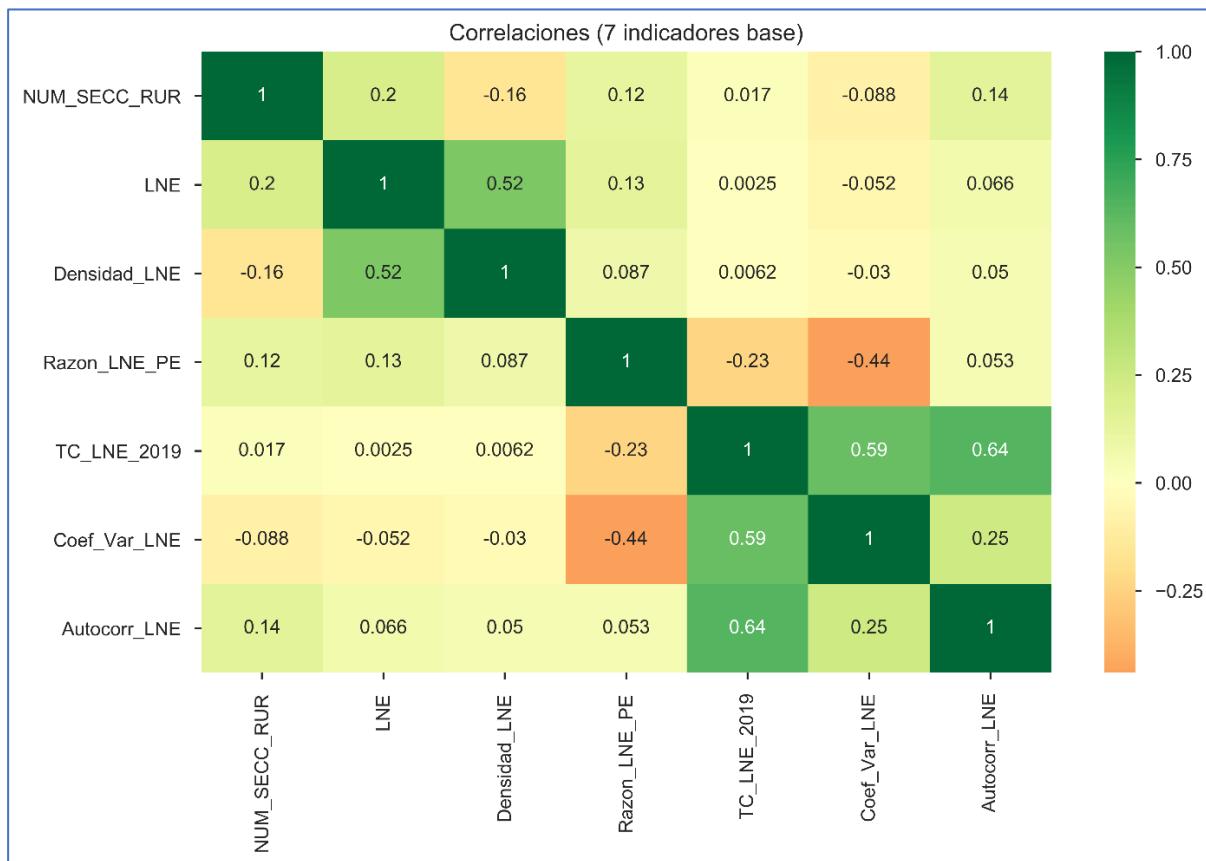


Gráfico 2. Mapa de calor de la matriz de correlación de los indicadores municipales⁷⁹

Fuente: Elaboración propia con la librería Python-Pandas-Seaborn

⁷⁸ Se recomienda ver el archivo anexo en Álvarez Hernández (2020) para visualizar con detalle los análisis comparativos y gráficos entre los indicadores base, y sus correspondientes estadísticas de tendencia central.

⁷⁹ Los números de cada recuadro indican el coeficiente de correlación entre el par de indicadores. Y la escala de colores resalta la fuerza de la correlación.

En particular, el Gráfico 2 permite destacar que, a nivel municipal, no hay indicios de que la magnitud y densidad del electorado esté relacionada positivamente —o negativamente— con la variabilidad demográfica, o con el nivel de actualización del RFE. Esto sugiere que los municipios altamente densificados o poblados (municipios ubicados en las zonas urbanas de México) no necesariamente presentarán una mayor complejidad electoral, al menos desde la perspectiva institucional relacionada con la realización de las tareas de identificación y actualización del RFE.

Recapitulando, los resultados del análisis exploratorio indican la presencia de correlaciones significativas entre los indicadores, lo cual da una justificación fundamental para la implementación del PCA, ya que dicha técnica es especialmente útil en reducir la dimensionalidad de conjuntos de indicadores que se encuentran altamente correlacionados.

La primera fase de la implementación del PCA es el preprocesamiento de datos (normalización y centralizado).⁸⁰ Posteriormente, se calcula la matriz de covarianza, con sus respectivos *eigenvalores* y *eigenvectores*, para generar las Componentes Principales. El resultado de la varianza explicada por cada Componente Principal se presenta en el Gráfico 3.

⁸⁰ El propósito principal de aplicar un preprocesamiento de datos es reducir la varianza observada en los indicadores base, ya que estos se encuentran en diferentes unidades de medida, lo cual induce una falsa variabilidad en los datos.

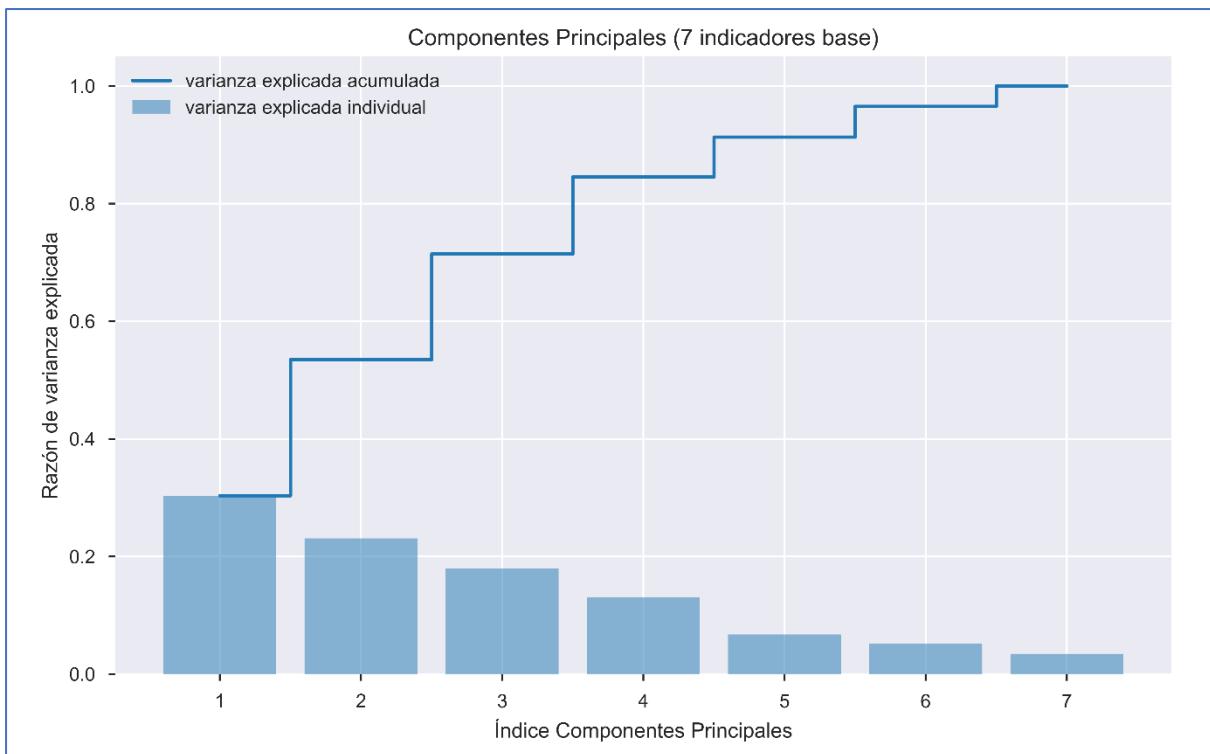


Gráfico 3. ICE municipales: Proporción de varianza explicada por Componente Principal

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn–Seaborn

El Gráfico 3 muestra que la primera Componente Principal (CP1) alcanza un porcentaje de varianza explicada aproximado del 30%; es decir, esta variable transformada presenta el mejor nivel de varianza explicada dentro del grupo de Componentes Principales y, en consecuencia, se toma como el principal ICE municipal. Sin embargo, la CP1 debe considerarse como un indicador límite, en el sentido de que no puede sintetizar completamente las características de los siete indicadores base analizados a nivel municipal, y, consecuentemente, debe usarse con cautela para la elaboración de un *ranking* que permita establecer comparaciones de complejidad electoral entre los municipios (de haber presentado la CP1 un mayor porcentaje de varianza explicada, se encontraría más justificado su uso como único ICE municipal).

Por otra parte, en el Gráfico 3 también se observa la contribución decreciente de las subsiguientes Componentes Principales en el porcentaje de varianza explicada. Por ejemplo, la segunda Componente Principal (CP2 o segundo ICE municipal)

contribuye con un 23% al porcentaje de varianza explicada, mientras que la tercera Componente Principal (CP3 o tercer ICE municipal) contribuye sólo con un 18%. Por lo tanto, se justifica la construcción de una tipología de municipios basada en el conjunto total de las Componentes Principales, ya que, en forma individual, cada ICE municipal no alcanza un nivel satisfactorio de síntesis de la información. Es decir, una TCE que se fundamente en los resultados de las siete Componentes Principales puede aportar más información para establecer comparaciones consistentes de complejidad electoral entre los municipios.⁸¹

En el Gráfico 4 se muestra la proyección de los municipios en el subespacio de CP1 y CP2. Esta gráfica permite apreciar algunos patrones de agrupamiento en la representación bidimensional, como es el caso de los municipios que obtienen una valoración de más de 4.0 en el ICE principal, o más de 2.5 en el segundo ICE (puntos en los extremos derecho y superior de la gráfica).⁸²

No obstante, la representación bidimensional del Gráfico 4 presenta algunas limitantes para mostrar claramente todos los patrones de agrupamiento existentes en el espacio de los ICE municipales, ya que la concentración de municipios alrededor del origen de la gráfica (valores cercanos a cero en el ICE principal y en el segundo ICE), impide distinguir patrones de agrupamiento adicionales. Para solventar estas limitantes, es necesario analizar la proyección tridimensional de los municipios en el espacio de los ICE municipales.

⁸¹ Hasta cierto punto es subjetiva la elección de cuántas Componentes Principales (ICE municipales) se deben elegir para proyectar los indicadores de los municipios en un espacio dimensional más reducido. Considerando los resultados obtenidos en el PCA a nivel municipal, y los objetivos de la investigación, se considera pertinente sustentar el *ranking* de complejidad electoral en el ICE principal, y utilizar el conjunto total de ICE municipales para elaborar posteriormente la TCE municipal.

⁸² La separación de estos municipios, respecto a la mayoría, es un indicio de que presentan características atípicas en alguno de los indicadores base. Para detectar en cuáles indicadores base se presentan estas características atípicas es necesario analizar la ponderación que recibió cada indicador en los ICE municipales (tema que se aborda en la siguiente sección de este capítulo).

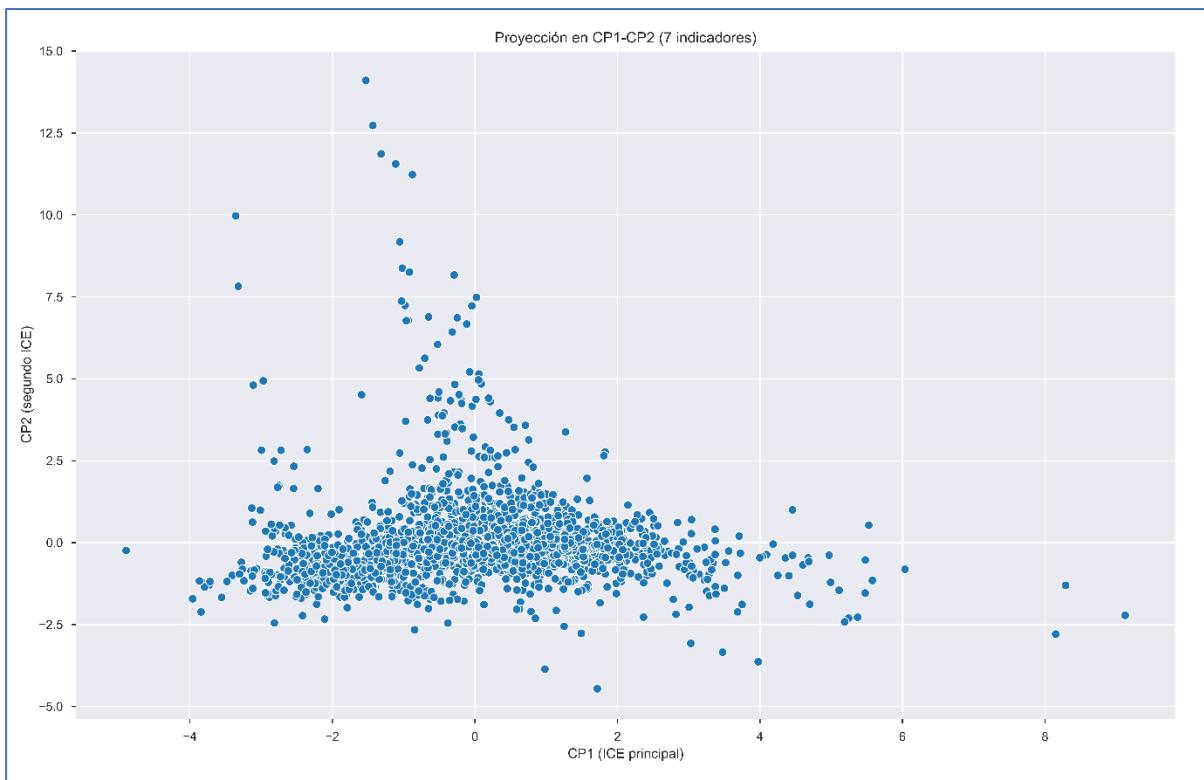


Gráfico 4. Distribución de los municipios en el subespacio CP1-CP2

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn–Seaborn

En el Gráfico 5 se presenta la proyección tridimensional de los municipios en el subespacio de las componentes CP1, CP2 y CP3. En esta gráfica se puede apreciar con mayor facilidad los patrones de agrupamiento en la nube de dispersión de los municipios (cúmulos en las partes externas de la nube de dispersión). Por ejemplo, se observa que la aparente acumulación de municipios alrededor del origen de la gráfica muestra en realidad un patrón de separación en la dirección del tercer ICE; es decir, hay municipios que presentan valoraciones similares en CP1 y CP2, pero se distinguen notoriamente en su valoración respecto a la CP3 (tercer ICE). Estas estructuras de agrupamiento se hacen más patentes cuando se rota la proyección tridimensional a lo largo de la dirección de alguna de las Componentes Principales.⁸³

⁸³ Se recomienda revisar la animación anexa en Álvarez Hernández (2020) para visualizar adecuadamente los patrones de agrupamiento de los municipios en el espacio de los ICE municipales.

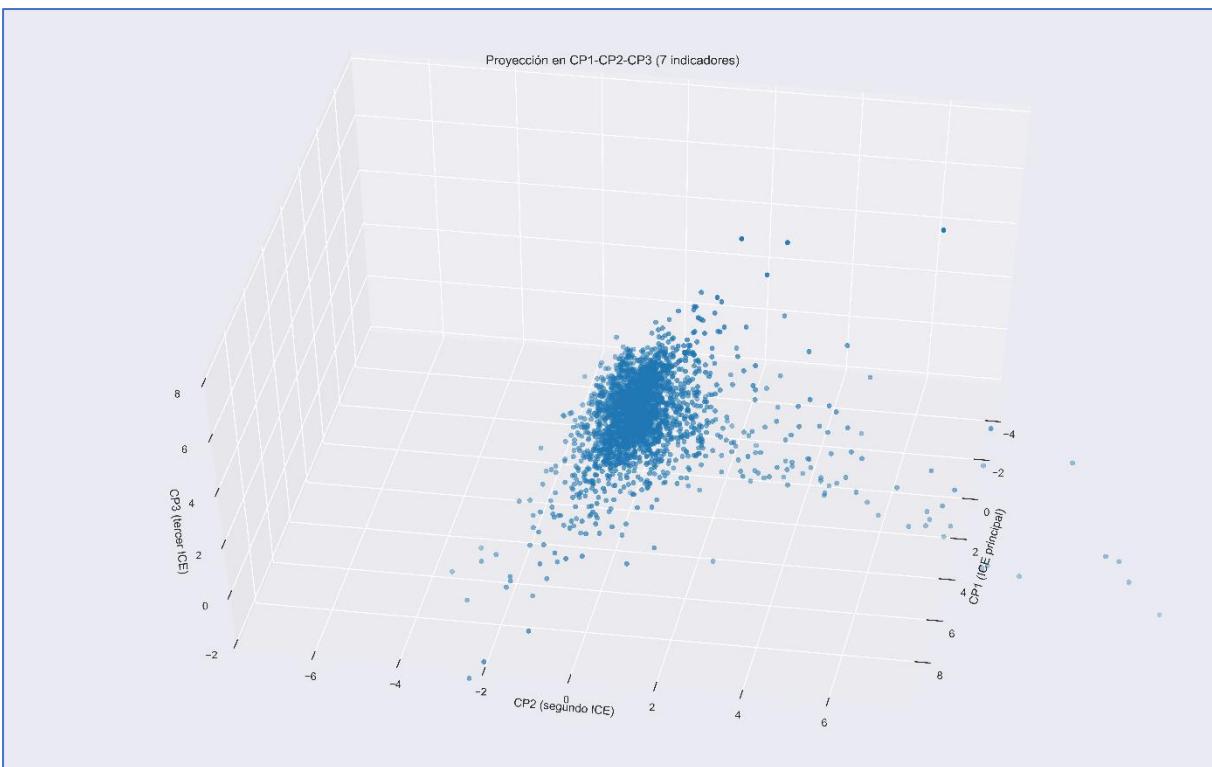


Gráfico 5. Distribución de los municipios en el subespacio CP1-CP2-CP3

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn–Seaborn

3.2 Interpretación de los ICE municipales

Para interpretar adecuadamente la información que aporta cada uno de los ICE municipales se analiza su composición respecto a los indicadores base; es decir, se analiza la ponderación que recibió cada indicador en la conformación de las Componentes Principales.⁸⁴

Para exemplificar, se muestra en el Cuadro 3 la composición de las primeras tres Componentes Principales. En dicho cuadro se puede observar que los indicadores que mayor peso tienen en la CP1 (ICE principal) son: el Coeficiente de Variación de la LNE, la Tasa de Crecimiento Anual de la LNE y el grado de Autocorrelación Mensual de la LNE. Respecto a la segunda Componente Principal, el mayor peso lo tienen los indicadores de Magnitud del Electorado, Densidad del Electorado, y el Coeficiente de Actualización del RFE. Finalmente, la tercera Componente Principal da mayor peso a

⁸⁴ Es importante recordar que los ICE municipales son una combinación lineal de los indicadores base (suma ponderada), donde los coeficientes de ponderación se encuentran determinados por el cálculo de los *eigenvectores* de la matriz de covarianza.

los indicadores: Grado de Urbanización Municipal, Densidad del Electorado, y Coeficiente de Actualización del RFE.

Indicador	CP1 (ICE principal)	CP2 (segundo ICE)	CP3 (tercer ICE)
Grado de Urbanización Municipal	-0.01939582	0.20642117	0.63877786
Magnitud del Electorado	-0.06200948	0.65606456	-0.15557432
Densidad del Electorado	-0.04362646	0.57276414	-0.4884409
Coeficiente de Actualización del RFE	-0.33486759	0.29680151	0.38620809
Tasa de Crecimiento Anual de la LNE (2019)	0.61225314	0.1327594	0.10449908
Coeficiente de Variación de la LNE	0.55376069	-0.07962829	-0.20403254
Autocorrelación Mensual de la LNE (1 retraso)	0.44747541	0.29470998	0.35704191

Cuadro 3. ICE municipales: Composición de las tres primeras Componentes Principales en función de los indicadores base

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn

Considerando que los indicadores que más pesan en el ICE principal están relacionados con el crecimiento y las variaciones poblacionales observadas en la LNE, dicho índice describe la complejidad electoral asociada a la variabilidad demográfica, y, por lo tanto, a los factores que inciden primordialmente en las tareas de identificación y actualización del RFE. Es decir, si un municipio presenta una alta valoración positiva en el ICE principal, éste se caracterizará por tener una mayor variabilidad y crecimiento demográfico en su LNE, junto con un alto nivel de autocorrelación mensual en el comportamiento temporal de su LNE.⁸⁵ Y en sentido contrario, si un municipio presenta una valoración altamente negativa en el ICE principal, éste se caracterizará por tener

⁸⁵ Recordando que, en el caso del Coeficiente de Variación de la LNE, la mayoría de los valores observados a nivel municipal se encuentran dentro del intervalo [0, 0.1]. En el caso de la Tasa de Crecimiento Anual de la LNE, los valores observados varían entre los intervalos [-0.2, 0.2]. Y en el caso del indicador Autocorrelación Mensual de la LNE, este sólo puede variar dentro del intervalo [0, 1].

menores niveles de variabilidad demográfica y de autocorrelación mensual, y también se caracterizará por presentar un decrecimiento demográfico en su LNE.⁸⁶

Por otra parte, el segundo ICE municipal está influenciado principalmente (y en forma positiva) por los indicadores de Magnitud y Densidad del Electorado. Esto implica que este índice describe fundamentalmente la complejidad electoral asociada a la magnitud y dispersión geográfica del electorado. Esto es, entre mayor sea la valoración de un municipio en el segundo ICE municipal, mayor será el número de ciudadanos inscritos en su LNE, y menor será la dispersión geográfica del electorado (densidad alta). Asimismo, el Coeficiente de Actualización del RFE también influye positivamente —pero en menor medida— en el segundo ICE municipal. En consecuencia, entre mayor sea la valoración de un municipio en el segundo ICE, mayor será el nivel de actualización entre la LNE y el PE de dicho municipio.⁸⁷ Respecto al comportamiento contrario, si un municipio presenta una valoración altamente negativa en el segundo ICE, éste se caracterizará por presentar un menor número de ciudadanos inscritos en la LNE, una mayor dispersión del electorado (densidad baja) y un menor nivel de actualización entre la LNE y el PE.

Finalmente, el tercer ICE municipal tiene tres influencias principales. En primer lugar (ponderación positiva), se encuentran los indicadores Grado de Urbanización Municipal y Coeficiente de Actualización del RFE. Y en segundo lugar (ponderación negativa), se encuentra el indicador Densidad del Electorado. Por consiguiente, este ICE municipal describe principalmente la complejidad electoral asociada a factores geográficos y estructurales. Es decir, los municipios que obtengan una alta valoración positiva en el tercer ICE se caracterizarán por poseer un elevado número de secciones rurales, una alta dispersión geográfica de su electorado (densidad baja), y un alto nivel de actualización entre la LNE y el PE.

Considerando las características de la composición del segundo ICE municipal, puede esperarse que, en general, los municipios de naturaleza mayoritariamente

⁸⁶ También obsérvese la ponderación negativa que posee el Coeficiente de Actualización del RFE en la composición del ICE principal. Esto implica que hay una correlación negativa entre el ICE principal y el Coeficiente de Actualización del RFE.

⁸⁷ Recordando que los indicadores NUM_SECC_RUR, LNE y Densidad_LNE sólo pueden presentar valores positivos (y enteros para el caso específico de los indicadores NUM_SECC_RUR y LNE).

urbana (distinguidos por una alta magnitud del electorado) presenten una valoración positiva en este índice. Algo similar ocurre respecto al tercer ICE municipal, ya que, considerando su composición, también se puede prever que los municipios de carácter mayoritariamente rural (distinguidos por una baja magnitud del electorado) presenten una valoración positiva en este índice. Respecto al ICE principal, no se puede predecir un patrón geográfico concluyente a partir de su composición, ya que, como se mostró en el análisis exploratorio del Gráfico 1, existen municipios urbanos y rurales con una alta variabilidad demográfica.

En el Gráfico 6 se representa, mediante un mapa de calor, las composiciones del conjunto total de Componentes Principales. En particular, entre mayor sea la intensidad de color verde, más fuerte es la ponderación positiva del indicador en el ICE municipal. Y contrariamente, entre mayor sea la intensidad de color rojo, más fuerte es la ponderación negativa del indicador en el ICE municipal. Consecuentemente, el Gráfico 6 permite interpretar los subsiguientes ICE municipales (CP4, CP5, CP6 y CP7) en función de la ponderación que reciben los indicadores base dentro de las Componentes Principales.⁸⁸

⁸⁸ En algunos casos (por ejemplo, la componente CP5) puede ser complicada la obtención de una interpretación clara respecto a qué factor de la complejidad electoral se está representando en la Componente Principal. Sin embargo, esto no afecta en general la información proporcionada por los primeros ICE municipales, sobre todo considerando que los primeros tres ICE municipales concentran aproximadamente el 72% del porcentaje de varianza explicada.

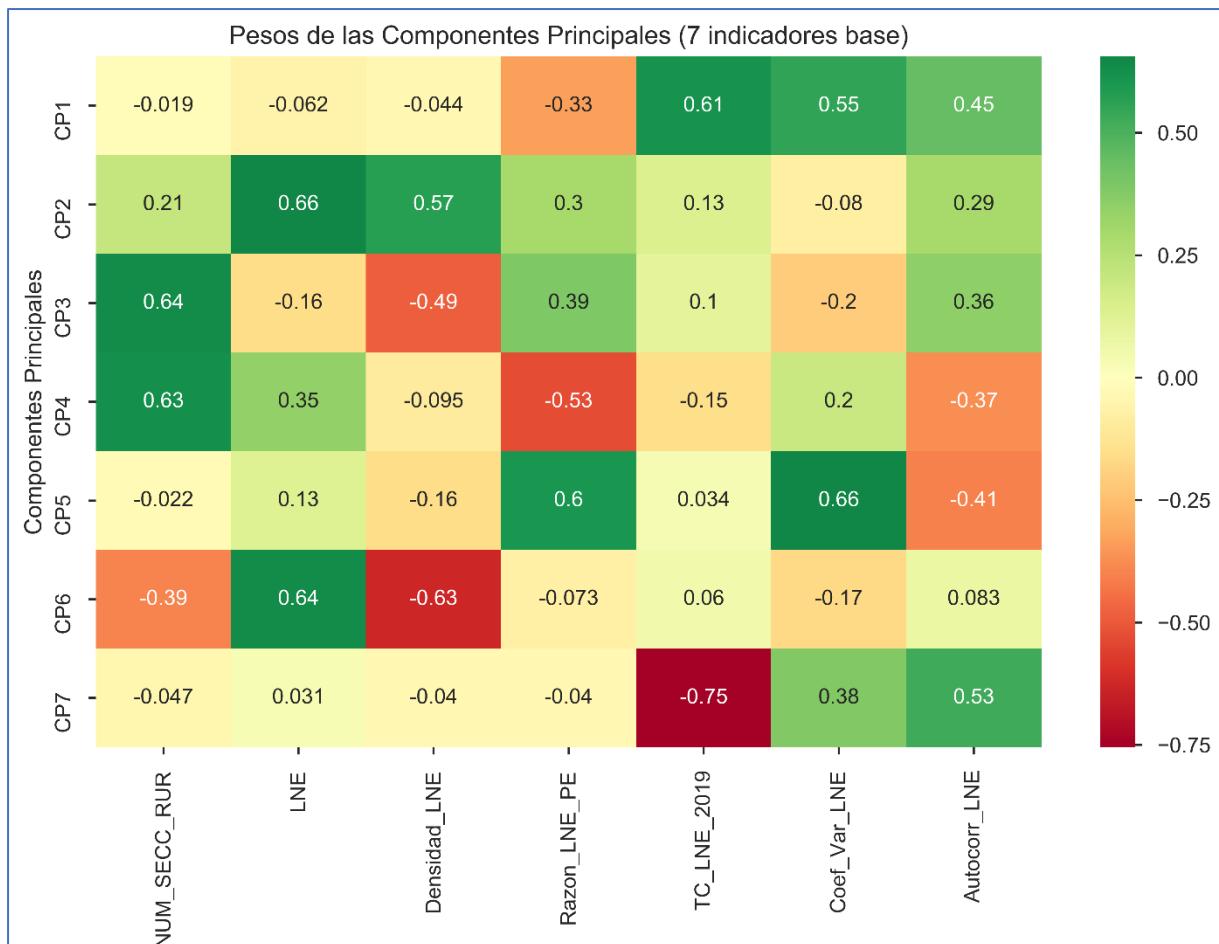


Gráfico 6. ICE municipales: Ponderadores de los indicadores base en cada Componente Principal⁸⁹

Fuente: Elaboración propia con la librería Python-Pandas-Seaborn

3.3 Ranking municipal de complejidad electoral

Es importante analizar la valoración que obtuvo cada municipio en los ICE municipales. Sin embargo, dada la cantidad de municipios examinados (2,458 municipios), no es posible ofrecer en este documento un desglose completo de los resultados obtenidos en el conjunto total de los ICE municipales.⁹⁰ Por lo tanto, en el Cuadro 4 se presenta una muestra del ranking municipal de complejidad electoral obtenido a partir de los resultados de la proyección de los municipios en el ICE principal.

⁸⁹ En este caso, los números que se encuentran en los recuadros indican los coeficientes de ponderación de cada indicador base en los ICE municipales.

⁹⁰ Se recomienda revisar el estadístico incorporado en los archivos de Álvarez Hernández (2020).

Ranking	Entidad	Municipio	CP1 (ICE principal)
1	Chihuahua	Uruachi	9.1234
2	Oaxaca	San Pedro Tidaa	8.2881
3	Oaxaca	San Juan Tabaa	8.1483
4	Jalisco	Mezquitic	6.0350
5	Chihuahua	Bocoyna	5.5773
6	Nuevo León	Carmen	5.5248
7	Yucatán	Kopoma	5.4768
8	Sonora	Nacori Chico	5.4745
9	Oaxaca	San Juan Petlapa	5.3696
10	Chihuahua	Maguarichi	5.2426

Cuadro 4. *Ranking municipal de complejidad electoral: primeros 10 municipios por valoración en el ICE principal*

Fuente: Elaboración propia

En el Cuadro 5 se presenta el *ranking* de complejidad electoral a nivel municipal por entidad federativa.

Entidad	Municipio	CP1 (ICE principal)
Aguascalientes	San Francisco de los Romo	-1.06189886
Baja California	Playas de Rosarito	-2.16744867
Baja California Sur	Mulege	1.17542173
Campeche	Escárcega	0.39042274
Coahuila	General Cepeda	3.55861987
Colima	Comala	1.31919868
Chiapas	Pantepec	3.25091229
Chihuahua	Uruachi	9.12348663

Ciudad de México	Milpa Alta	0.04783023
Durango	Canelas	-0.15619294
Guanajuato	Tarandacuaao	1.48154929
Guerrero	Tlalixtaquilla de Maldonado	1.21289799
Hidalgo	Atotonilco de Tula	1.57404423
Jalisco	Mezquitic	6.03503772
México	Villa del Carbón	2.4889961
Michoacán	Nahuatzen	2.50919005
Morelos	Huitzilac	1.65366449
Nayarit	Bahía de Banderas	1.68877378
Nuevo León	Carmen	5.5248378
Oaxaca	San Pedro Tidaa	8.28817981
Puebla	Chigmecatitlan	0.12606381
Querétaro	El Marques	2.24567824
Quintana Roo	Tulum	-0.124101
San Luis potosí	Villa de Arriaga	2.06128739
Sinaloa	Elota	2.31385085
Sonora	Nacori Chico	5.47451808
Tabasco	Teapa	0.55540671
Tamaulipas	Guerrero	-0.4800136
Tlaxcala	Ixtenco	1.14039603
Veracruz	Miahuatlan	3.24296649
Yucatán	Kopoma	5.47689392
Zacatecas	Trinidad García de la Cadena	3.05916005

Cuadro 5. Municipios de mayor complejidad electoral por entidad federativa

Fuente: Elaboración propia

La información mostrada en los Cuadros 4 y 5 permite señalar que los municipios de mayor complejidad electoral (respecto al ICE principal) se encuentran localizados principalmente en zonas geográficas alejadas notoriamente de los principales centros urbanos de México (por ejemplo, 7 de los primeros 10 municipios de mayor complejidad electoral se encuentran ubicados en las serranías del poniente y sur de México), o son municipios de las entidades federativas que tienen un carácter semi rural (como es el caso de la alcaldía de Milpa Alta en la Ciudad de México).

En la Figura 1 se muestra con mayor detalle el mapeo del *ranking* municipal de complejidad electoral en la totalidad del territorio nacional. La figura permite apreciar que hay amplias zonas del territorio nacional que presentan valoraciones negativas en el ICE principal, especialmente en las entidades de Baja California, Durango, Guerrero, Puebla, Quintana Roo y Tamaulipas. De acuerdo con estos resultados, un número significativo de los municipios ubicados en estas entidades federativas se caracterizan por tener un decrecimiento demográfico en su LNE, lo cual también podría explicar los bajos niveles de autocorrelación mensual de la LNE que se observan en estos municipios, y que coadyuvan a la valoración negativa obtenida en el ICE principal.

Por otra parte, hay otras zonas del territorio nacional (principalmente en los estados de Chihuahua, Coahuila, Jalisco y Sonora) que presentan una alta valoración positiva en el ICE principal. Esto implica que los municipios localizados en dichas zonas geográficas presentan un alto crecimiento demográfico en su electorado, junto con un alto nivel de variabilidad demográfica en la LNE, el cual impacta negativamente en los niveles de actualización de la LNE y el PE.

Otro patrón geográfico relevante del *ranking* se hace patente en las zonas urbanas. Por ejemplo, en la Figura 2 se muestra con detalle el mapeo del *ranking* municipal de complejidad electoral en la zona metropolitana de la Ciudad de México. En dicha figura se puede advertir que los municipios pertenecientes a la mancha urbana presentan en promedio una valoración negativa en el ICE principal. Es decir, estos municipios se caracterizan por presentar un alto nivel de actualización entre su LNE y el PE; y, en general, presentan tasas de crecimiento demográfico nulas o negativas (características concordantes con la naturaleza urbana de estos municipios).

En contraste, los municipios limítrofes a la mancha urbana presentan características opuestas –hecho reflejado en la valoración positiva que exhiben en el ICE principal—, ya que estos muestran un menor nivel de actualización en su LNE, y además presentan tasas de crecimiento demográfico más elevadas (características de los municipios que están en proceso de urbanización, debido a su cercanía a una urbe importante).

Para complementar los resultados y explicaciones presentados en este capítulo –principalmente los detalles técnicos de la obtención de las Componentes Principales— se proporciona, en el Anexo I, el pseudocódigo de Python utilizado para la implementación del PCA y la elaboración de las proyecciones de municipios en el espacio de los ICE municipales. Asimismo, en los archivos adjuntos del repositorio de GitHub se incluye una animación de la proyección de los municipios en el espacio de las Componentes Principales, así como un estadístico con el listado completo de los municipios de México con su respectiva valoración en los ICE municipales.⁹¹

⁹¹ Véanse archivos anexos de Álvarez Hernández (2020).

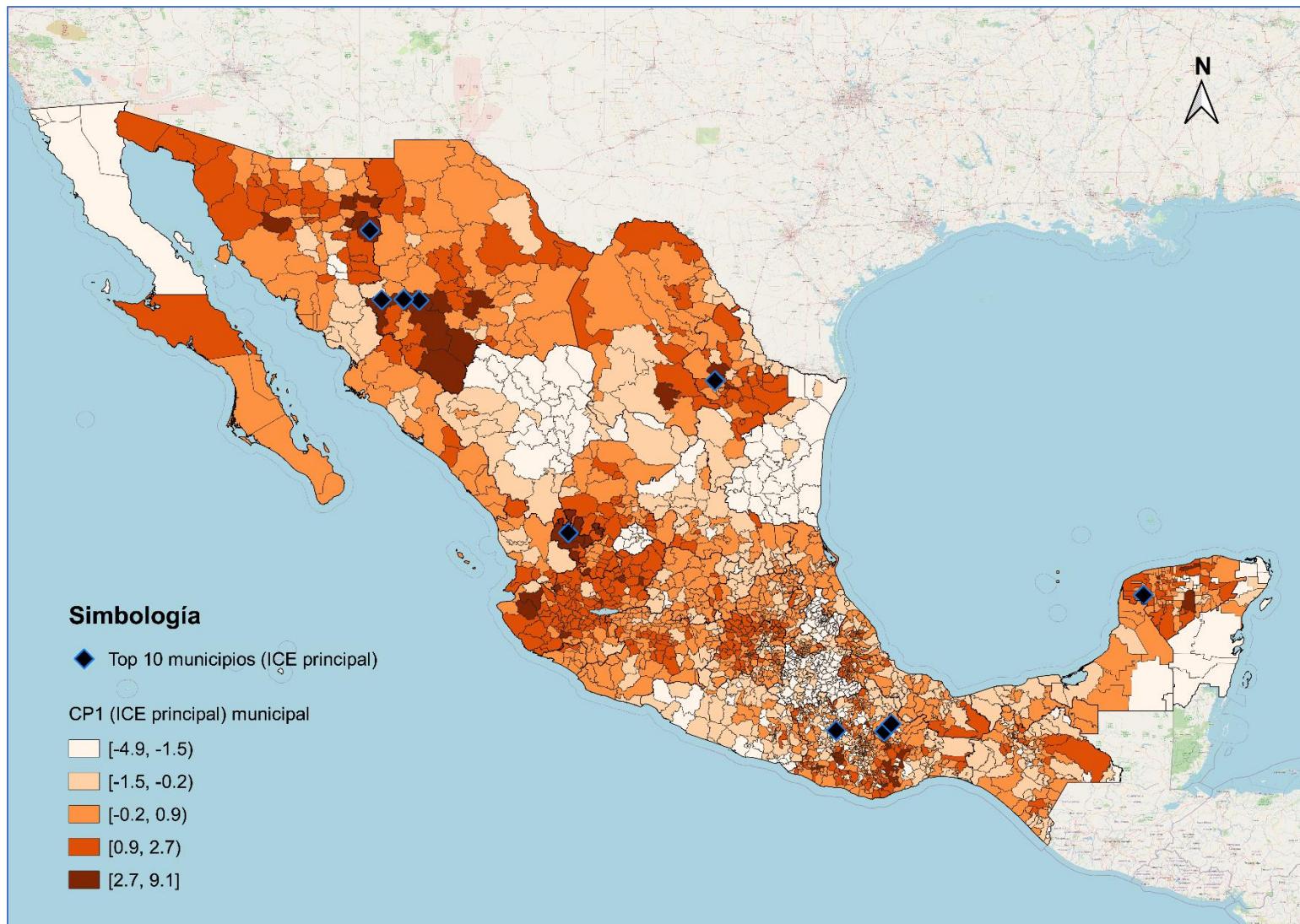


Figura 1. Ranking municipal de complejidad electoral

Fuente: Elaboración propia con QGIS

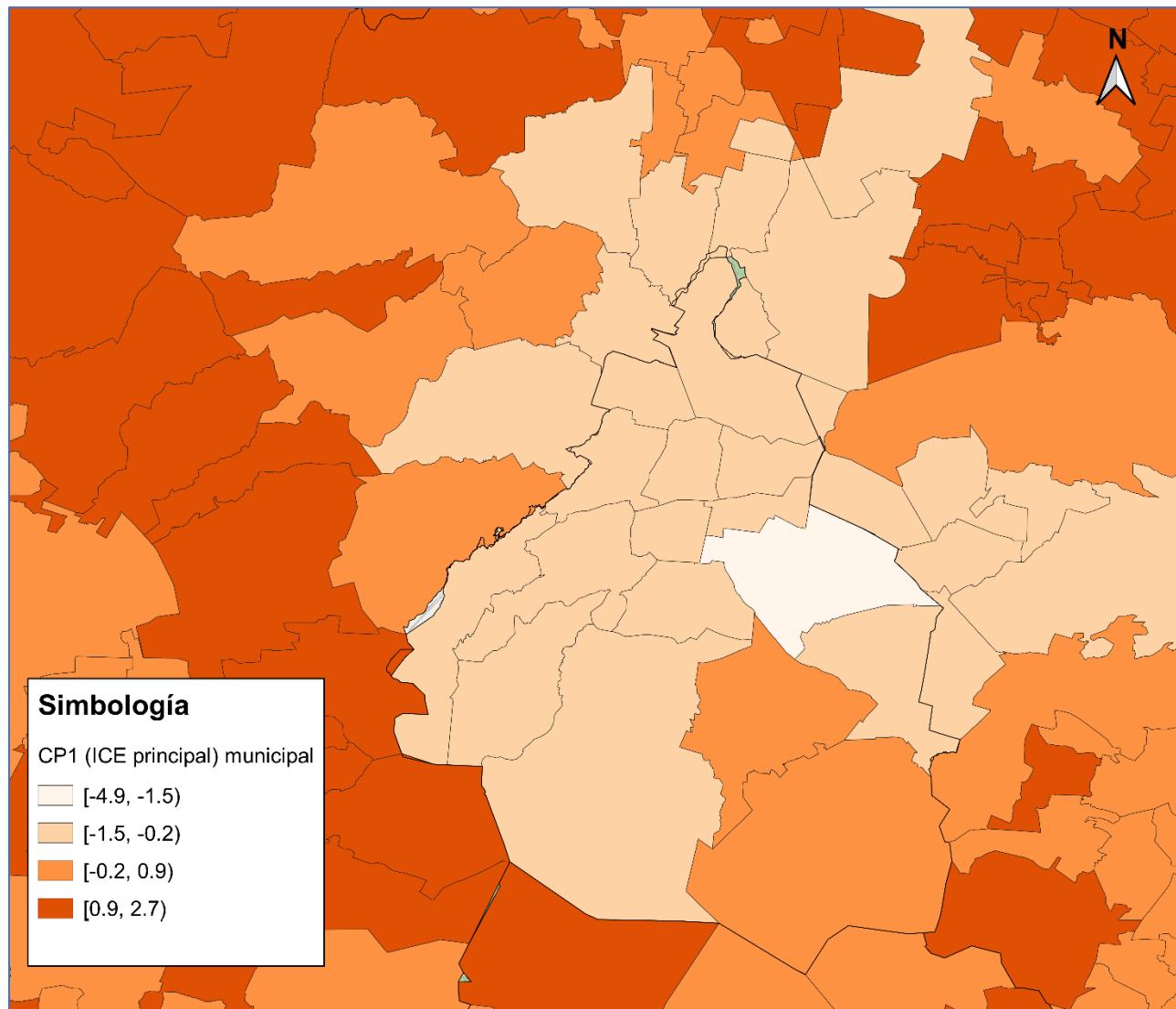
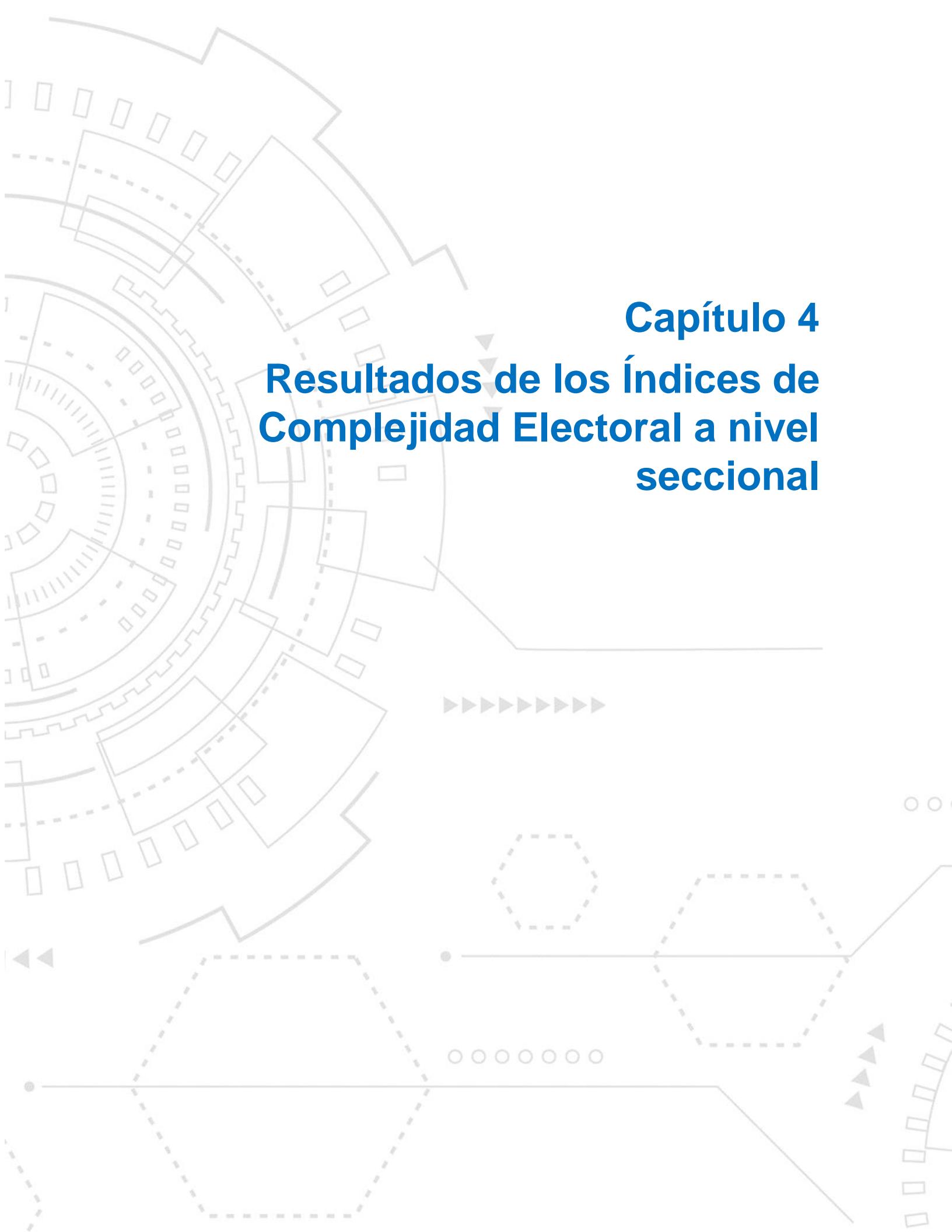


Figura 2. Ranking municipal de complejidad electoral en la zona metropolitana de la Ciudad de México

Fuente: Elaboración propia con QGIS



Capítulo 4

Resultados de los Índices de Complejidad Electoral a nivel seccional

Capítulo 4. Resultados de los Índices de Complejidad Electoral a nivel seccional

En forma similar a lo que se realizó en el Capítulo 3, se inicia la construcción de los ICE seccionales con un análisis exploratorio de los seis indicadores base analizados a nivel seccional.⁹² Luego, se implementa el método PCA para transformar el conjunto de indicadores a un nuevo grupo de variables (las Componentes Principales), y utilizando como métrica de evaluación la varianza explicada, se analizan sus capacidades de síntesis de información para establecer la correspondencia entre los ICE seccionales y las Componentes Principales. Finalmente, se obtienen las proyecciones bidimensional y tridimensional de las secciones en el nuevo espacio de transformación conformado por las Componentes Principales (lo cual permite analizar la emergencia de patrones de agrupamiento en las proyecciones), se determina la composición de los ICE seccionales –es decir, la ponderación que recibe cada indicador base en la conformación de las Componentes Principales— y se analizan los resultados geográficos del *ranking* seccional de complejidad electoral (obtenido a partir de la primera Componente Principal).

Al igual que en el proceso de construcción de los ICE municipales, la elaboración de los ICE seccionales también conllevó a un análisis de diferentes escenarios y transformaciones de datos en el grupo de indicadores base. Específicamente, se analizaron los efectos de algunas transformaciones logarítmicas implementadas en los indicadores con distribuciones altamente asimétricas.⁹³ La implementación de estas transformaciones elimina el alto sesgo presente en las distribuciones, pero dicha corrección no aporta alguna mejora significativa en la capacidad de síntesis de información de las Componentes Principales. Consecuentemente, en este capítulo sólo se presentan los resultados

⁹² De un universo total de 68,436 secciones electorales existentes en 2019, sólo se pudo recuperar la información completa de los seis indicadores base para un sub-universo de 68,390 secciones electorales. Las secciones electorales no incluidas presentaban campos vacíos, o datos inconsistentes, en alguno de los indicadores base.

⁹³ Mismo caso de los indicadores de Magnitud y Densidad del Electorado.

correspondientes al escenario que mejor resultados proporcionó en el proceso de construcción de los ICE seccionales.⁹⁴

4.1 Análisis exploratorio y construcción de los ICE seccionales

Como se explicó previamente en el Capítulo 3, los correlogramas permiten analizar gráficamente las relaciones entre pares de indicadores, con el propósito de identificar la existencia o ausencia de correlaciones positivas o negativas entre ellos. Asimismo, los histogramas permiten analizar gráficamente las distribuciones que presentan cada uno de los indicadores base.

En el Gráfico 7 se presentan los resultados del análisis exploratorio (panel de correlogramas e histogramas) realizado sobre el conjunto de seis indicadores considerados en el estudio a nivel seccional.⁹⁵ Los histogramas del Gráfico 7 (mostrados sobre la diagonal del panel) permiten constatar que los seis indicadores base —a excepción de los indicadores Coeficiente de Actualización del RFE (Razon_LNE_PE) y Coeficiente de Variación (Coef_Var_LNE)— presentan distribuciones asimétricas positivas, lo cual denota que hay un número significativo de secciones electorales que poseen características geoespaciales, demográficas y estructurales atípicas. En particular, destaca el comportamiento altamente sesgado de los indicadores Magnitud del Electorado (LNE) y Densidad del Electorado (Densidad_LNE), ya que estos dos indicadores base presentan una notoria asimetría positiva en sus distribuciones, lo cual implica que, en promedio, las secciones electorales se caracterizan por tener una baja magnitud y densidad en sus electorados (es decir, son secciones electorales que cumplen con la normativa constitucional de un máximo de 3,000 ciudadanos por sección electoral); sin embargo, hay un número considerable de secciones electorales atípicas que no cumplen con este patrón.⁹⁶

⁹⁴ Los resultados de los escenarios alternativos se pueden consultar en los archivos anexos de Álvarez Hernández (2020).

⁹⁵ Cada punto en los diagramas de dispersión representa alguna de las 68,390 secciones electorales estudiadas.

⁹⁶ A manera de ejemplo, considérese la gráfica del panel localizada en la primera fila y columna (histograma del indicador LNE). Este histograma muestra que la gran mayoría de secciones electorales tiene un número reducido de ciudadanos inscritos en la LNE, y esto se debe primordialmente a la normatividad constitucional de mantener un límite máximo de 3,000 ciudadanos por sección electoral. Sin embargo, el histograma también muestra que hay algunas secciones que rebasan significativamente el

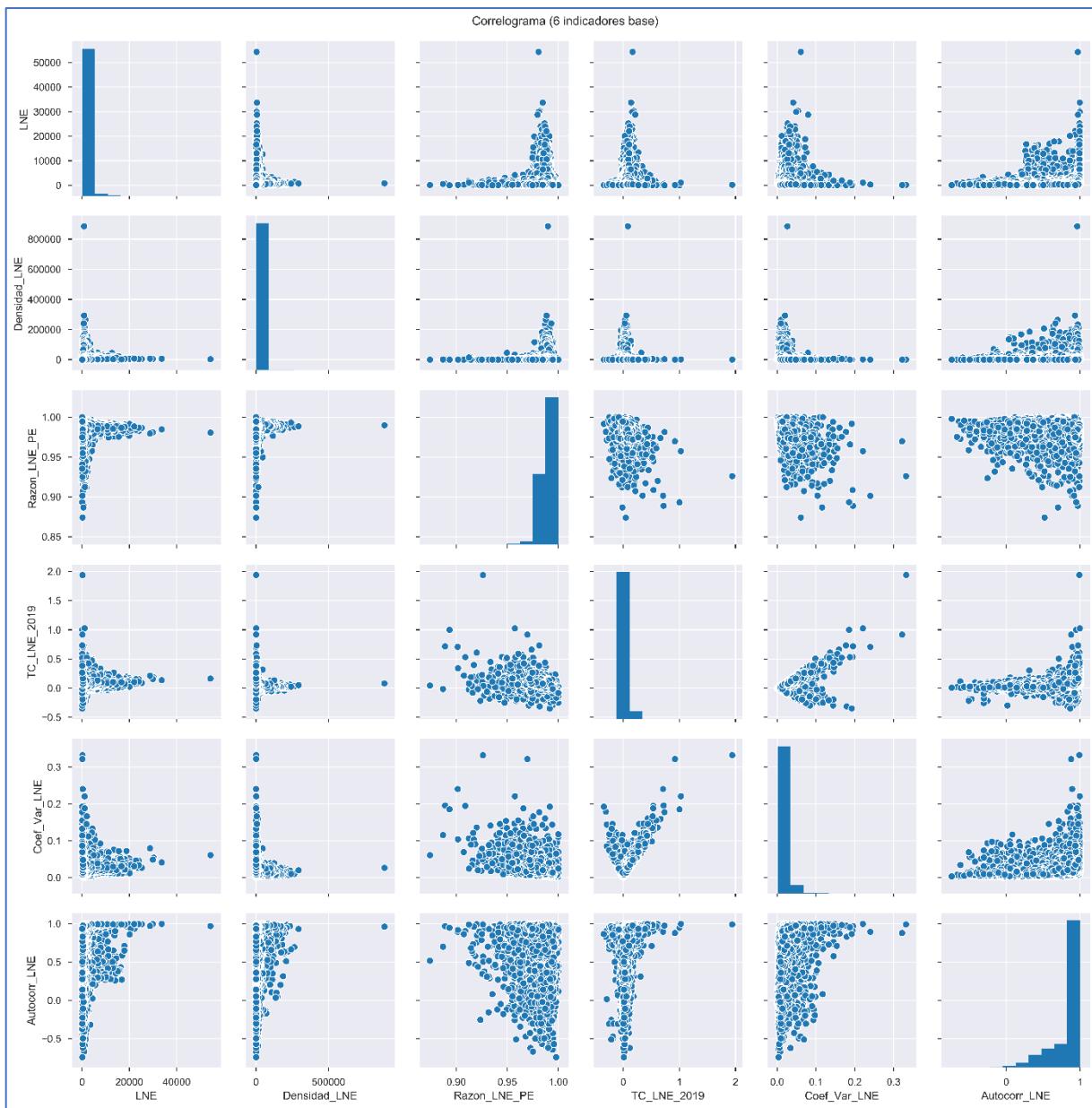


Gráfico 7. Panel de gráficas de correlación y distribución de los indicadores que conforman los ICE seccionales⁹⁷

Fuente: Elaboración propia con la librería Python-Pandas-Seaborn

Por otra parte, los correogramas del Gráfico 7 exhiben que los indicadores asociados a la variabilidad demográfica (TC_LNE_2019, Coef_Var_LNE) presentan

límite establecido (se identifican algunas secciones con más de 30,000 ciudadanos registrados en la LNE).

⁹⁷ Se recomienda revisar el Cuadro 2 para identificar adecuadamente la nomenclatura de los indicadores base.

correlaciones positivas entre sí. En contraste, el indicador asociado al nivel de actualización del RFE (Razon_LNE_PE) se encuentra correlacionado negativamente con los indicadores asociados a la variabilidad demográfica.⁹⁸ La fuerza de dichas correlaciones se visualiza más claramente en el mapa de calor del Gráfico 8.

El Gráfico 8 permite destacar que, a nivel seccional, no hay indicios de que la magnitud y densidad del electorado estén relacionadas positivamente —o negativamente— con la variabilidad demográfica o el nivel de actualización del RFE. Esto sugiere que las secciones electorales altamente densificadas y pobladas (secciones ubicadas en las zonas urbanas de México) no necesariamente son las secciones de mayor complejidad electoral, ya que existen otras secciones electorales que presentan un mayor dinamismo demográfico y, por ende, su grado de complejidad electoral también es mayor.

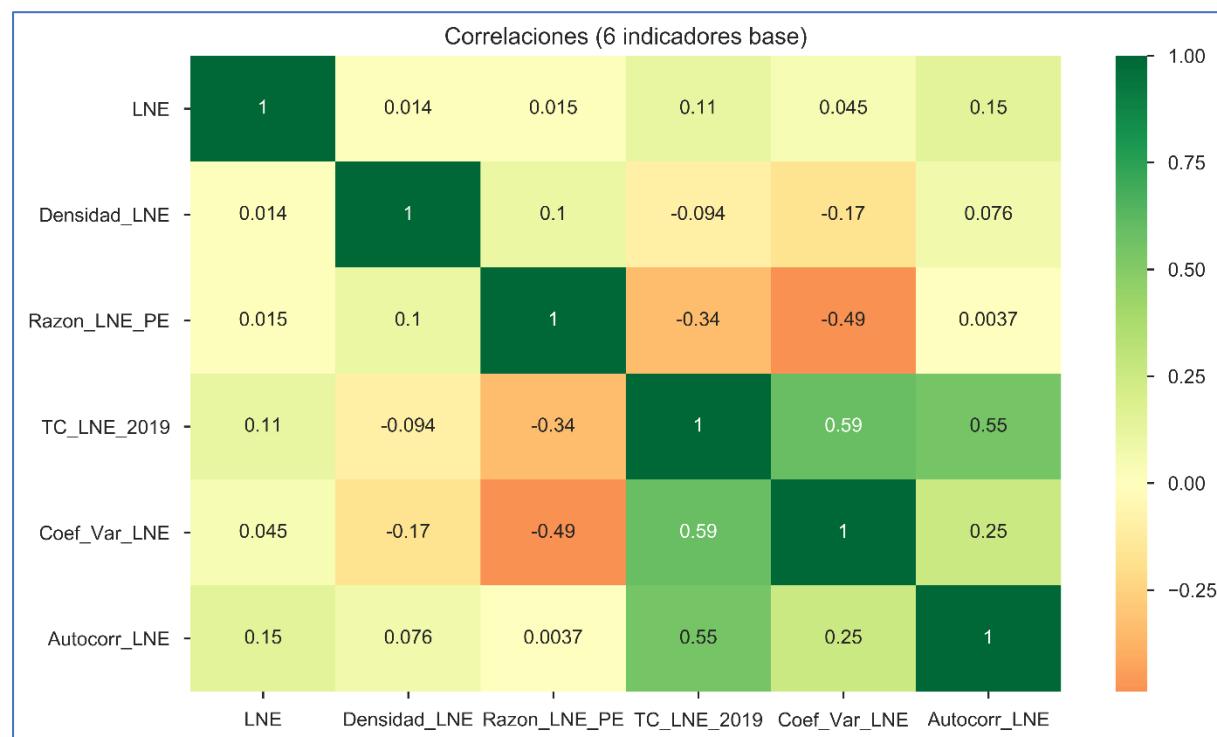


Gráfico 8. Mapa de calor de la matriz de correlación de los indicadores seccionales

Fuente: Elaboración propia con la librería Python-Pandas-Seaborn

⁹⁸ Comportamiento similar a lo que se observó previamente en el caso municipal.

Consecuentemente, los resultados del análisis exploratorio de datos justifican la pertinencia de aplicar el método PCA al conjunto de indicadores base seccional con el fin de obtener un nuevo y más sintético conjunto de índices. Como se explicó en el Capítulo 3, la primera fase de la implementación del PCA es el preprocesamiento del conjunto de indicadores base (normalizar y centralizar). Posteriormente, se calcula la matriz de covarianza con sus respectivos *eigenvalores* y *eigenvectores*, siendo estos últimos las Componentes Principales del caso seccional. El resultado del porcentaje de varianza explicada por cada Componente Principal se muestra en el Gráfico 9.

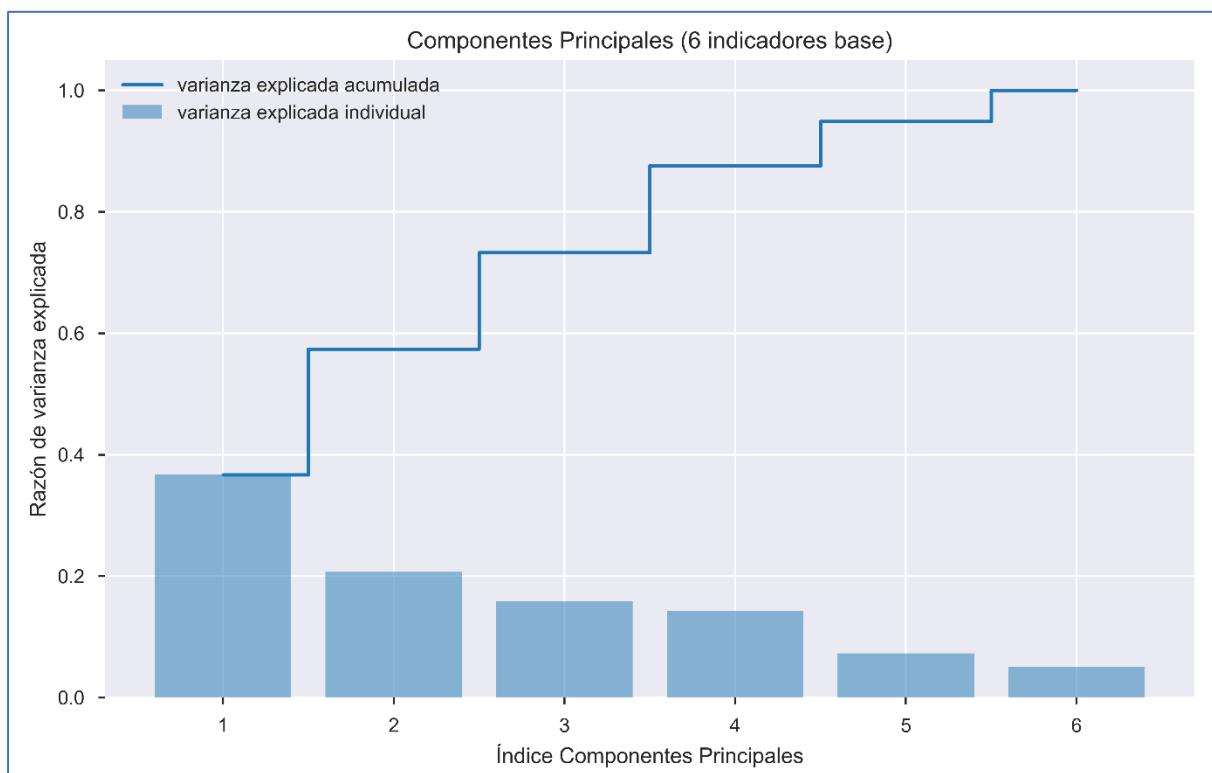


Gráfico 9. ICE seccionales: Proporción de varianza explicada por Componente Principal

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn–Seaborn

Se puede observar en el Gráfico 9 que la primera Componente Principal (CP1), alcanza un porcentaje de varianza explicada del 36% y consecuentemente, la CP1 se toma como el principal ICE seccional. Sin embargo, tal como ocurrió en el caso de los ICE municipales, la CP1 seccional no es un indicador que sintetice plenamente las características de los seis indicadores base y, por lo tanto, debe usarse con cautela

para la elaboración de un *ranking* que permita establecer comparaciones de complejidad electoral entre las secciones electorales.

Por otra parte, en el Gráfico 9 también se observa la contribución decreciente de las subsiguientes Componentes Principales en el porcentaje de varianza explicada. Por ejemplo, la segunda Componente Principal (CP2 o segundo ICE seccional) contribuye con un 20% al porcentaje de varianza explicada, mientras que la tercera Componente Principal (CP3 o tercer ICE seccional) contribuye sólo con un 15%. Por lo tanto, para realizar comparaciones consistentes de complejidad entre las secciones electorales, se justifica la construcción de una tipología de secciones electorales basada en el conjunto total de las Componentes Principales, considerando que, en forma individual, cada ICE seccional no consigue un nivel satisfactorio de síntesis de información.⁹⁹

En el Gráfico 10 se muestra la proyección de las secciones electorales en el subespacio de CP1 y CP2. Esta gráfica permite advertir algunos patrones de agrupamiento en la representación bidimensional, como es el caso de aquellas secciones electorales que obtienen una valoración de más de 10.0 en el ICE principal (puntos ubicados en el extremo derecho de la nube de dispersión). A partir de estos patrones de agrupamiento se puede inferir la existencia de grupos de secciones electorales con características similares (en lo que concierne a los factores geoespacial, demográfico y estructural), o con características anómalas respecto al comportamiento promedio observado en el conjunto total de secciones electorales.¹⁰⁰

⁹⁹ Como se mencionó previamente en el Capítulo 3, la elección de cuántas Componentes Principales se debe utilizar para proyectar las unidades territoriales en un espacio dimensional reducido es, hasta cierto punto, subjetiva. Por lo tanto, considerando los objetivos y las características específicas de los indicadores analizados en el caso seccional, se elige el conjunto total de Componentes Principales como los respectivos ICE seccionales, los cuales fundamentarán la construcción de la TCE seccional.

¹⁰⁰ En el Capítulo 6 se analiza con detalle el tema de la identificación de secciones electorales atípicas.

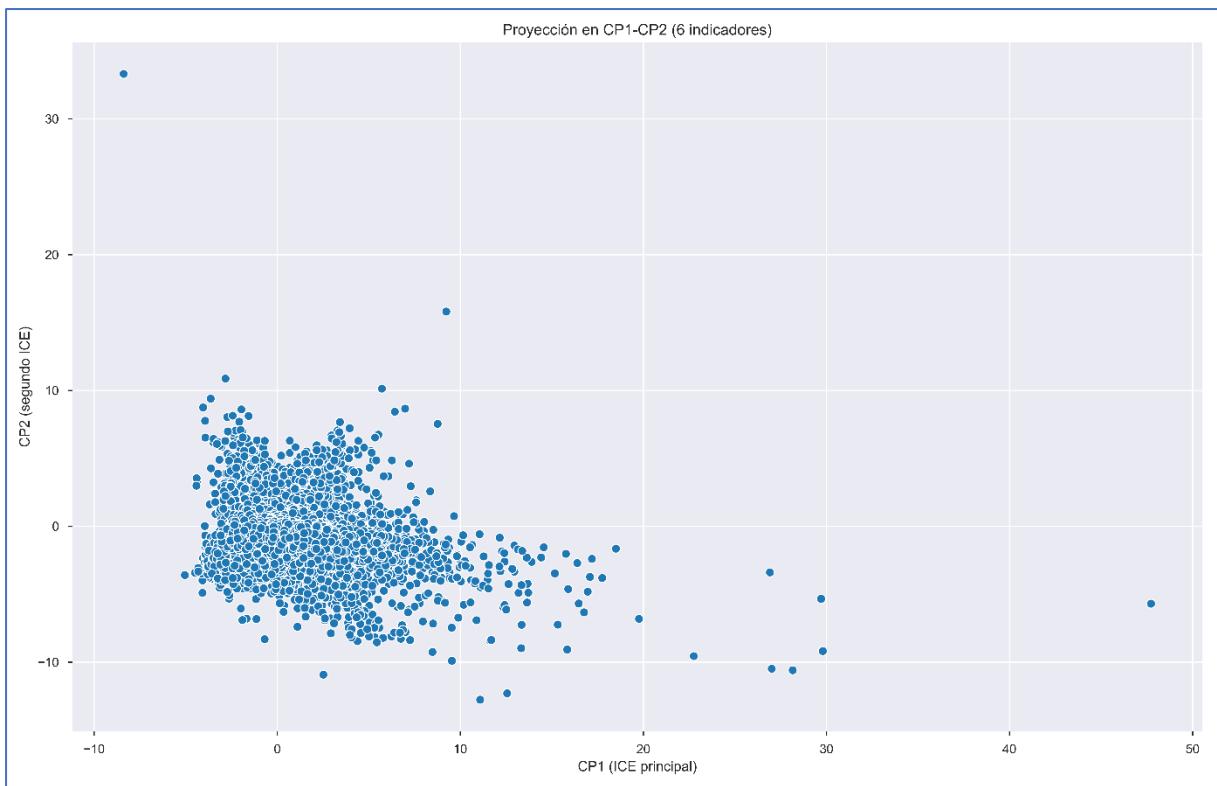


Gráfico 10. Distribución de las secciones electorales en el subespacio CP1-CP2

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn–Seaborn

Sin embargo, la representación bidimensional presenta algunas limitantes para mostrar claramente todos los patrones de agrupamiento existentes en el espacio de los ICE seccionales. Por ejemplo, la concentración de secciones electorales observada alrededor del origen del Gráfico 10 (valores cercanos a cero en el ICE principal y en el segundo ICE), implica la existencia de patrones de agrupamiento adicionales. En consecuencia, es necesario analizar la proyección tridimensional de las secciones electorales en el espacio completo de los ICE seccionales.

En el Gráfico 11 se proyecta el espacio de los componentes CP1, CP2 y CP3. En esta gráfica se pueden distinguir patrones de agrupamiento adicionales en la dispersión de las secciones electorales (véanse los picos de la nube de dispersión). Por ejemplo, se observa que la aparente aglomeración de secciones electorales cerca del origen de la gráfica posee en realidad un patrón de separación en la dirección del tercer ICE seccional. Las estructuras de aglomeración se hacen más patentes cuando

se rota la proyección tridimensional a lo largo de la dirección de alguna de las Componentes Principales.¹⁰¹

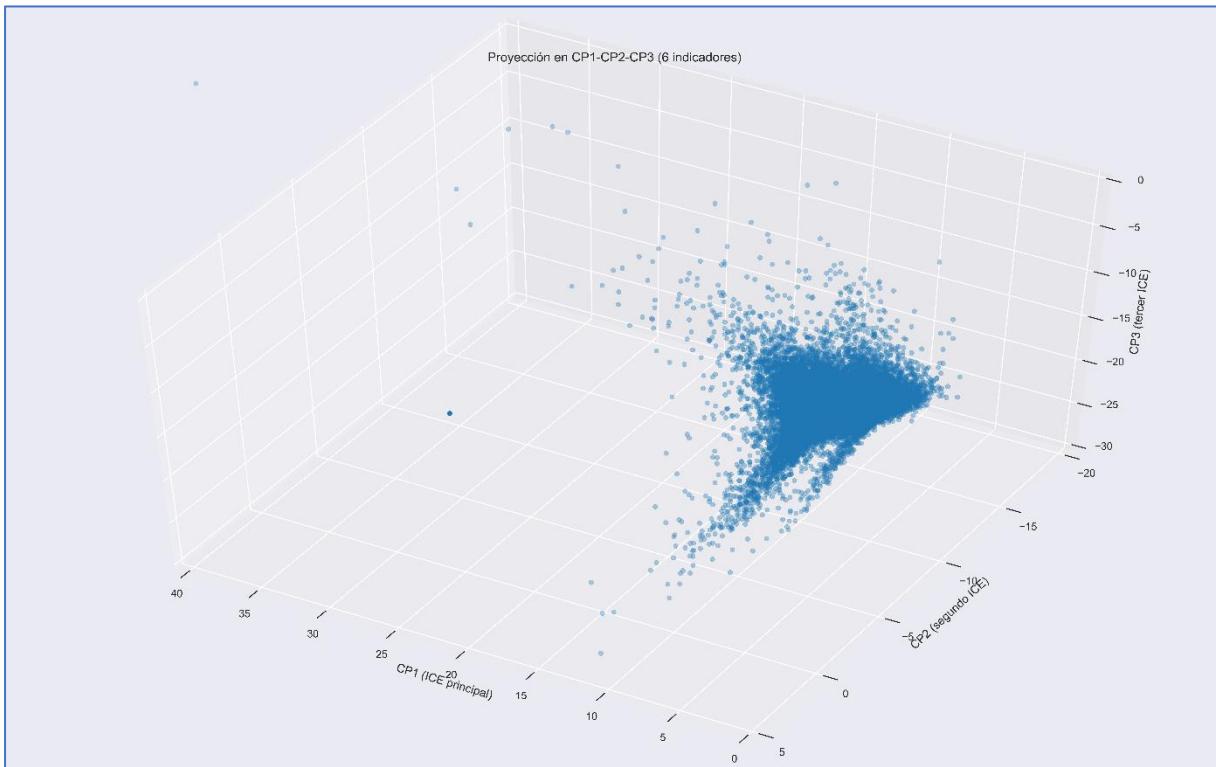


Gráfico 11. Proyección de las secciones electorales en el subespacio CP1-CP2-CP3

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn–Seaborn

4.2 Interpretación de los ICE seccionales

Con relación a la composición de los ICE seccionales (en función de los indicadores base),¹⁰² se muestra en el Cuadro 6 la composición de las primeras tres Componentes Principales para determinar su interpretación.

Indicador	CP1 (ICE principal)	CP2 (segundo ICE)	CP3 (tercer ICE)
Magnitud del Electorado	0.11835223	0.43815545	-0.72711206

¹⁰¹ Para visualizar adecuadamente los patrones de agrupamiento de las secciones electorales en el espacio de los ICE seccionales se recomienda revisar la animación anexa en el repositorio de GitHub de Álvarez Hernández (2020).

¹⁰² Es importante mencionar nuevamente que los ICE seccionales son una combinación lineal (suma ponderada) de los indicadores base, donde los coeficientes de ponderación se encuentran determinados por el cálculo de los *eigenvectores* de la matriz de covarianza.

Densidad del Electorado	-0.13463196	0.48142498	0.65542368
Coeficiente de Actualización del RFE	-0.40170491	0.44442913	-0.14182563
Tasa de Crecimiento Anual de la LNE (2019)	0.58735786	0.1514139	0.07721292
Coeficiente de Variación de la LNE	0.55730014	-0.20617036	0.04885994
Autocorrelación Mensual de la LNE (1 lag)	0.38849382	0.55973221	0.1151704

Cuadro 6. ICE seccionales: Composición de las tres primeras Componentes Principales en función de los indicadores base

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn

A partir del Cuadro 6 se puede constatar que los indicadores base que mayor peso (positivo) tienen en la CP1 (ICE principal) son: el Coeficiente de Variación de la LNE y la Tasa de Crecimiento Anual de la LNE. Asimismo, el Coeficiente de Actualización del RFE también presenta una ponderación (negativa) significativa dentro de la composición del ICE principal. Respecto a la segunda Componente Principal CP2 (segundo ICE), el mayor peso lo presentan los indicadores de Autocorrelación Mensual de la LNE, Densidad del Electorado, y el Coeficiente de Actualización del RFE. Por último, la tercera componente CP3 (tercer ICE) da mayor peso a los indicadores de Densidad del Electorado (ponderación positiva), Magnitud del Electorado, y el Coeficiente de Actualización del RFE (estos dos últimos indicadores presentan una ponderación negativa dentro del índice).

Considerando que los indicadores base que más pesan en la composición del ICE principal seccional están relacionados con el crecimiento demográfico y las variaciones temporales observadas en la LNE, éste índice describe la complejidad electoral asociada al factor demográfico. Es decir, entre mayor sea la valoración (positiva) de una sección electoral en el ICE principal, dicha sección electoral presentará una mayor variabilidad temporal en su LNE, un crecimiento demográfico positivo, y un menor nivel de actualización entre su LNE y el PE. Por otro lado, se tiene que entre menor sea la valoración (negativa) de una sección electoral en el ICE principal, menor será la variabilidad de su LNE, mayor será el nivel de actualización de su LNE respecto al PE, y presentará decrecimiento demográfico en su LNE.

Por otra parte, el segundo ICE seccional está influenciado primordialmente por el nivel de Autocorrelación de la LNE respecto al comportamiento mensual previo (autocorrelación a 1 retraso). Por lo tanto, este índice se puede interpretar como una métrica que mide la correlación entre el comportamiento reciente de la LNE con el comportamiento futuro de corto plazo. Asimismo, los indicadores Densidad del Electorado y Coeficiente de Actualización del RFE también influyen positivamente en el segundo ICE. En consecuencia, entre mayor sea la valoración (positiva) de una sección electoral en el segundo ICE, mayor será la regularidad del comportamiento de corto plazo de su LNE, menor la dispersión geográfica del electorado (densidad alta), y mayor el nivel de actualización de su LNE respecto al PE. Por el contrario, entre menor sea la valoración (negativa) de una sección electoral en el segundo ICE, más irregular será el comportamiento de corto plazo de su LNE, mayor será la dispersión geográfica del electorado (densidad baja), y el nivel de actualización de su LNE respecto al PE será menor.

Finalmente, el tercer ICE seccional presenta dos influencias principales. La primera influencia (ponderador positivo) proviene del indicador Densidad del Electorado, y la segunda influencia (ponderador negativo) proviene del indicador Magnitud del Electorado. Es importante puntualizar que hay una correlación directa entre ambos indicadores, ya que la densidad es igual al indicador de magnitud del electorado multiplicado por un valor escalar (superficie urbanizada de la sección electoral). Por consiguiente, las secciones electorales que obtengan una valoración alta (positiva) en el tercer ICE seccional se caracterizarán por presentar un bajo número de ciudadanos inscritos en la LNE, o por tener una reducida superficie urbanizada (y por ende, una alta densidad).

En el Gráfico 12 se representa, mediante un mapa de calor, la ponderación de cada indicador base en la composición de los ICE seccionales. Específicamente, en cada recuadro se indica el coeficiente de ponderación, y la tonalidad de colores señala la fuerza y polaridad de la ponderación. Por lo tanto, el Gráfico 12 permite interpretar

los subsiguientes ICE seccionales (CP4, CP5 y CP6) en función de la ponderación que reciben los indicadores base dentro de las Componentes Principales.¹⁰³

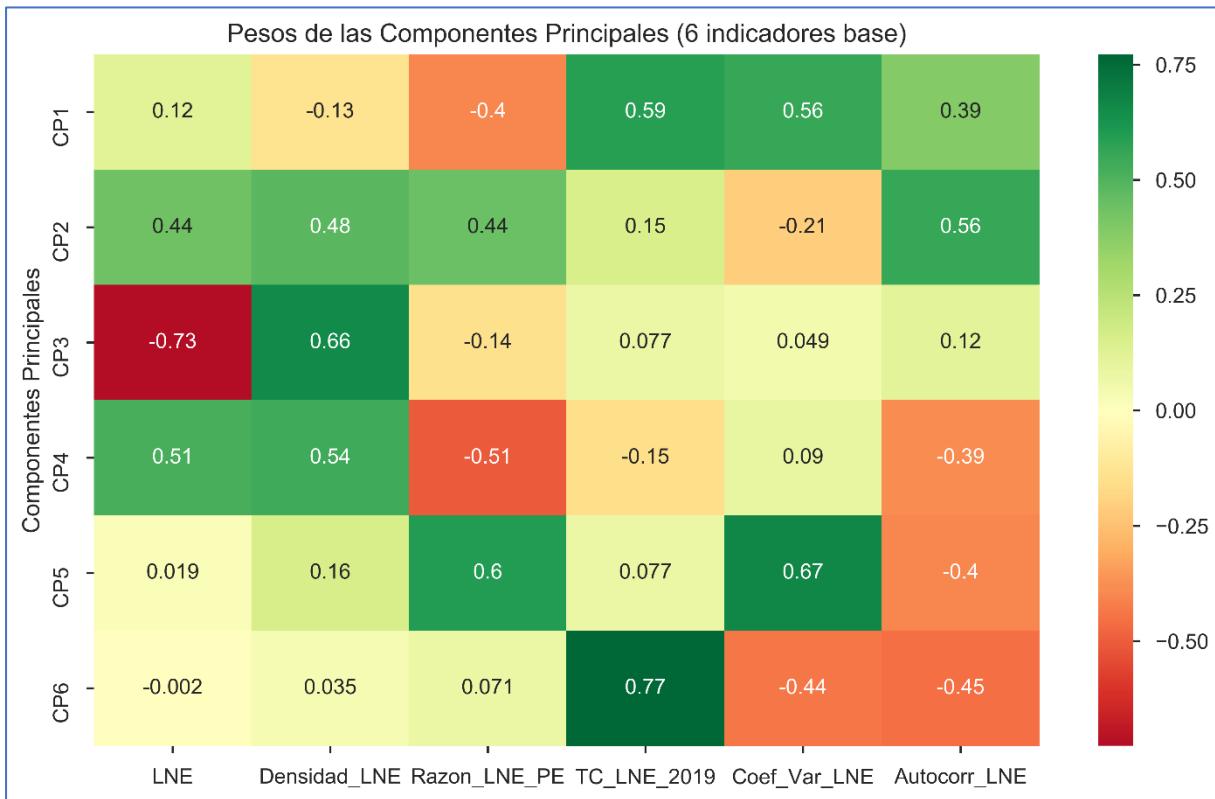


Gráfico 12. ICE seccionales: Ponderadores de los indicadores base en cada Componente Principal

Fuente: Elaboración propia con la librería Python-Pandas-Seaborn

4.3 Ranking seccional de complejidad electoral

Retomando la observación realizada en el Capítulo 3, es importante analizar la valoración que obtuvo cada sección electoral en los ICE seccionales. Sin embargo, la cantidad de secciones electorales analizadas (68,390 secciones) impide ofrecer en este documento un desglose completo de los resultados obtenidos con los seis ICE

¹⁰³ En algunos casos (por ejemplo, la componente CP6) puede ser complicada la obtención de una interpretación clara respecto a cuál factor de la complejidad electoral se está modelando en el ICE seccional. Sin embargo, esto no afecta en general la información proporcionada por los primeros ICE seccionales, sobre todo considerando que los primeros tres índices concentran aproximadamente el 71% del porcentaje de varianza explicada.

seccionales.¹⁰⁴ En su lugar, se presenta en el Cuadro 7 una muestra del *ranking* seccional de complejidad electoral, el cual se obtiene a partir del ICE principal.

Ranking	Entidad	Municipio	Sección	CP1 (ICE principal)
1	México	Otzolotepec	6517	47.7232
2	Chihuahua	Uruachi	2696	29.7973
3	Sinaloa	Sinaloa	3427	29.7040
4	Chihuahua	Uruachi	2687	28.1519
5	Chihuahua	Bocoyna	147	27.0129
6	Nuevo León	García	2547	26.9122
7	Chihuahua	Uruachi	2695	22.7558
8	Coahuila	San Juan de Sabinas	1711	19.7604
9	Nuevo León	Gral. Escobedo	2737	18.4925
10	Quintana Roo	Tulum	948	17.7509

Cuadro 7. Ranking seccional de complejidad electoral: primeras 10 secciones electorales por valoración en el ICE principal

Fuente: Elaboración propia

Asimismo, en el Cuadro 8 se presenta el *ranking* de complejidad electoral a nivel seccional por entidad federativa.

Entidad	Municipio	Sección	CP1 (ICE principal)
Aguascalientes	Aguascalientes	594	8.0005
Baja California	Tijuana	1869	8.5829
Baja California Sur	Mulege	105	8.4430
Campeche	Candelaria	295	7.6279

¹⁰⁴ Se recomienda revisar el estadístico incorporado en los archivos adjuntos de Álvarez Hernández (2020).

Coahuila	San Juan de Sabinas	1711	19.7604
Colima	Manzanillo	224	4.3147
Chiapas	Siltepec	1192	11.4813
Chihuahua	Uruachi	2696	29.7973
Ciudad de México	Iztapalapa	2509	4.7421
Durango	Otaez	956	8.5081
Guanajuato	León	1508	8.0231
Guerrero	Chilapa de Álvarez	1193	13.3199
Hidalgo	Tizayuca	1768	7.0814
Jalisco	San Martín de Bolaños	2114	12.9482
México	Otzolotepec	6517	47.7232
Michoacán	Marcos Castellanos	940	13.6410
Morelos	Atlatlahucan	20	3.5031
Nayarit	Huajicori	203	9.1026
Nuevo León	García	2547	26.9122
Oaxaca	Nejapa de Madero	467	16.4652
Puebla	Puebla	1492	5.6551
Querétaro	Querétaro	754	7.5828
Quintana Roo	Tulum	948	17.7509
San Luis Potosí	Santa María del Rio	1199	9.5429
Sinaloa	Sinaloa	3427	29.7040
Sonora	Bacanora	50	15.8922
Tabasco	Huimanguillo	715	5.9137
Tamaulipas	Reynosa	1772	15.7696
Tlaxcala	Santa Cruz Tlaxcala	609	3.74500

Veracruz	Llamatlan	1698	10.2852
Yucatán	Mérida	1110	12.3958
Zacatecas	Valparaíso	1542	12.1285

Cuadro 8. Secciones electorales de mayor complejidad electoral por entidad federativa

Fuente: Elaboración propia

La información mostrada en los Cuadros 7 y 8 permite señalar que las secciones electorales de mayor complejidad electoral (respecto al ICE principal) se encuentran localizadas principalmente en zonas geográficas alejadas de los principales centros urbanos de México (por ejemplo, 6 de las primeras 10 secciones electorales de mayor complejidad electoral se encuentran ubicadas en las zonas montañosas del occidente de México), o en los límites de zonas urbanas (como es el caso de las secciones 2547 y 2737 del estado de Nuevo León).

En la Figura 3 se muestra con detalle el mapeo del *ranking* seccional de complejidad electoral en la totalidad del territorio nacional. Se puede observar que hay amplias zonas del territorio nacional (especialmente en las entidades de Baja California, Durango, Puebla, Quintana Roo y Tamaulipas) que presentan valoraciones negativas en el ICE principal. Esto implica que en dichas zonas geográficas se encuentran secciones electorales que presentan poca variabilidad poblacional en su LNE, o que se caracterizan por tener tasas negativas de crecimiento demográfico.

En contraste, hay otras zonas del territorio nacional (principalmente en los estados de Chihuahua, Coahuila, Jalisco y Sonora) que concentran un número significativo de secciones electorales con una alta valoración positiva en el ICE principal. Esto significa que las secciones electorales ubicadas en estas zonas muestran, en general, tasas de crecimiento demográfico positivas, y además se caracterizan por presentar un bajo nivel de actualización en su LNE y PE.

Los resultados obtenidos a nivel seccional son concordantes con los resultados alcanzados con anterioridad en el análisis a nivel municipal, ya que se observa una coincidencia significativa entre las zonas geográficas de alta complejidad electoral, descritas por el *ranking* municipal, y las zonas geográficas de alta complejidad electoral descritas por el *ranking* seccional. La coincidencia de resultados se vuelve más patente

al analizar el mapeo del *ranking* seccional de complejidad electoral en la zona metropolitana de la Ciudad de México, el cual se muestra en la Figura 4.

A través de la Figura 4 se puede apreciar que las secciones electorales ubicadas dentro de la mancha urbana presentan en promedio una valoración negativa en el ICE principal. Es decir, estas secciones electorales se caracterizan por ostentar un alto nivel de actualización entre su LNE y el PE, consecuencia parcial de presentar también tasas de crecimiento demográfico nulas o negativas (características comunes en las secciones electorales que se encuentran completamente urbanizadas). En contraste, las secciones electorales limítrofes a la mancha urbana presentan características opuestas –hecho reflejado en la valoración positiva que exhiben en el ICE principal—, ya que éstas muestran un menor nivel de actualización en su LNE y, además, presentan tasas de crecimiento demográfico positivas (es decir, son secciones electorales que todavía cuentan con superficie disponible para urbanizar).

Para complementar los resultados y explicaciones presentados en este capítulo, se incluye en el Anexo I el pseudocódigo de Python utilizado para la implementación del PCA. Asimismo, en los archivos adjuntos se encuentra una animación de la proyección de las secciones en el espacio de las Componentes Principales, así como un estadístico con el listado completo de las secciones electorales con su respectiva valoración en los ICE seccionales.¹⁰⁵

¹⁰⁵ Ver archivos adjuntos del repositorio de GitHub de Álvarez Hernández (2020).

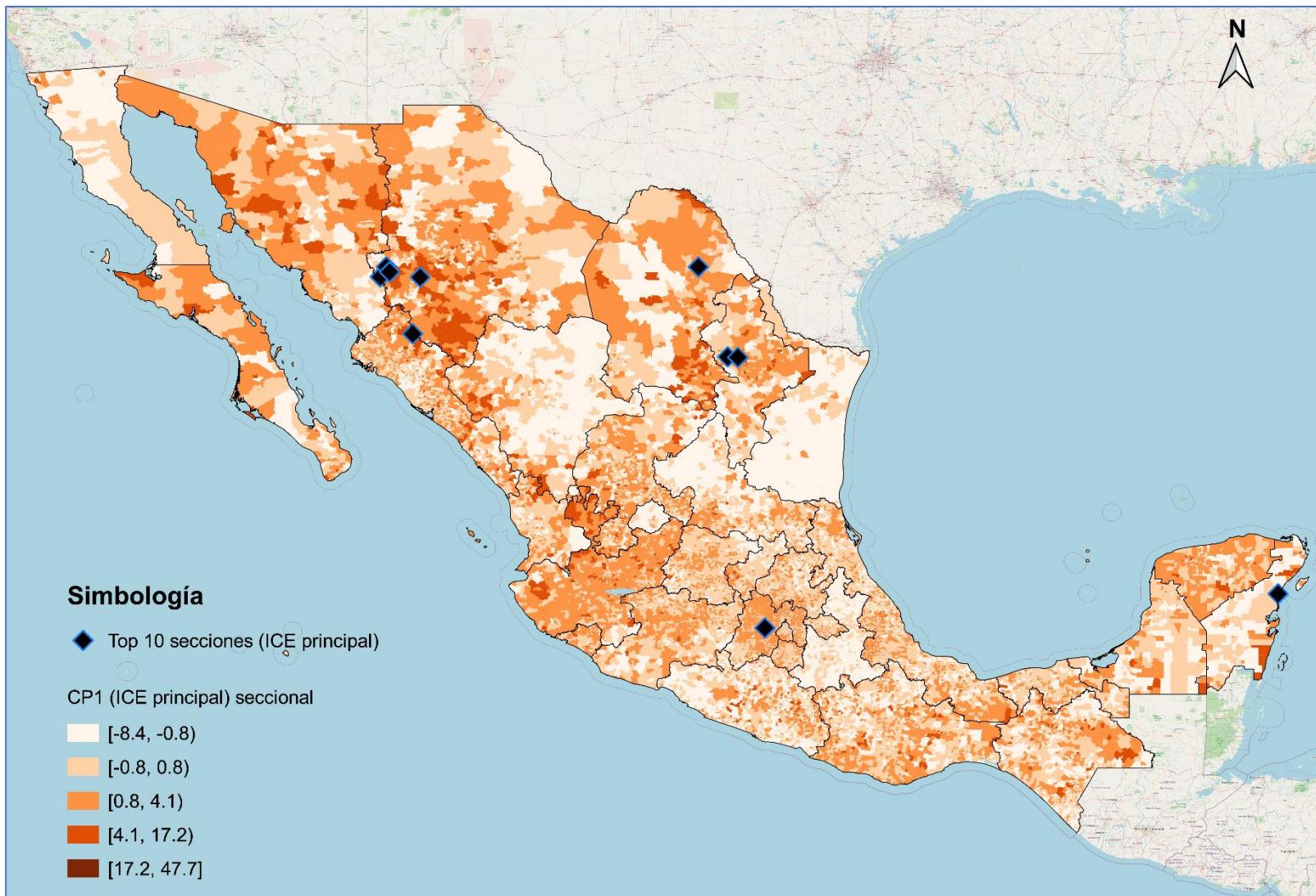


Figura 3. Ranking seccional de complejidad electoral

Fuente: Elaboración propia con QGIS

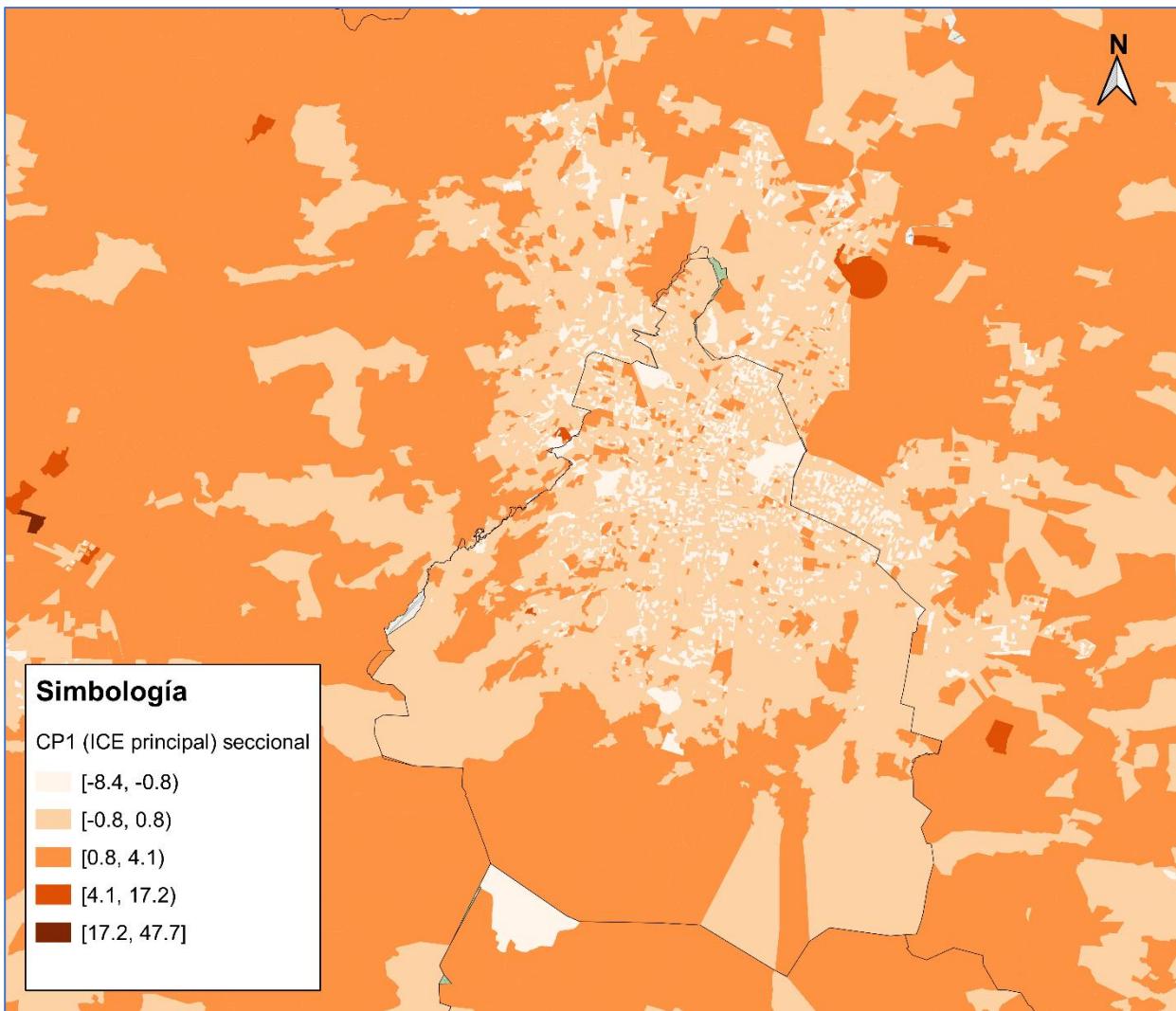
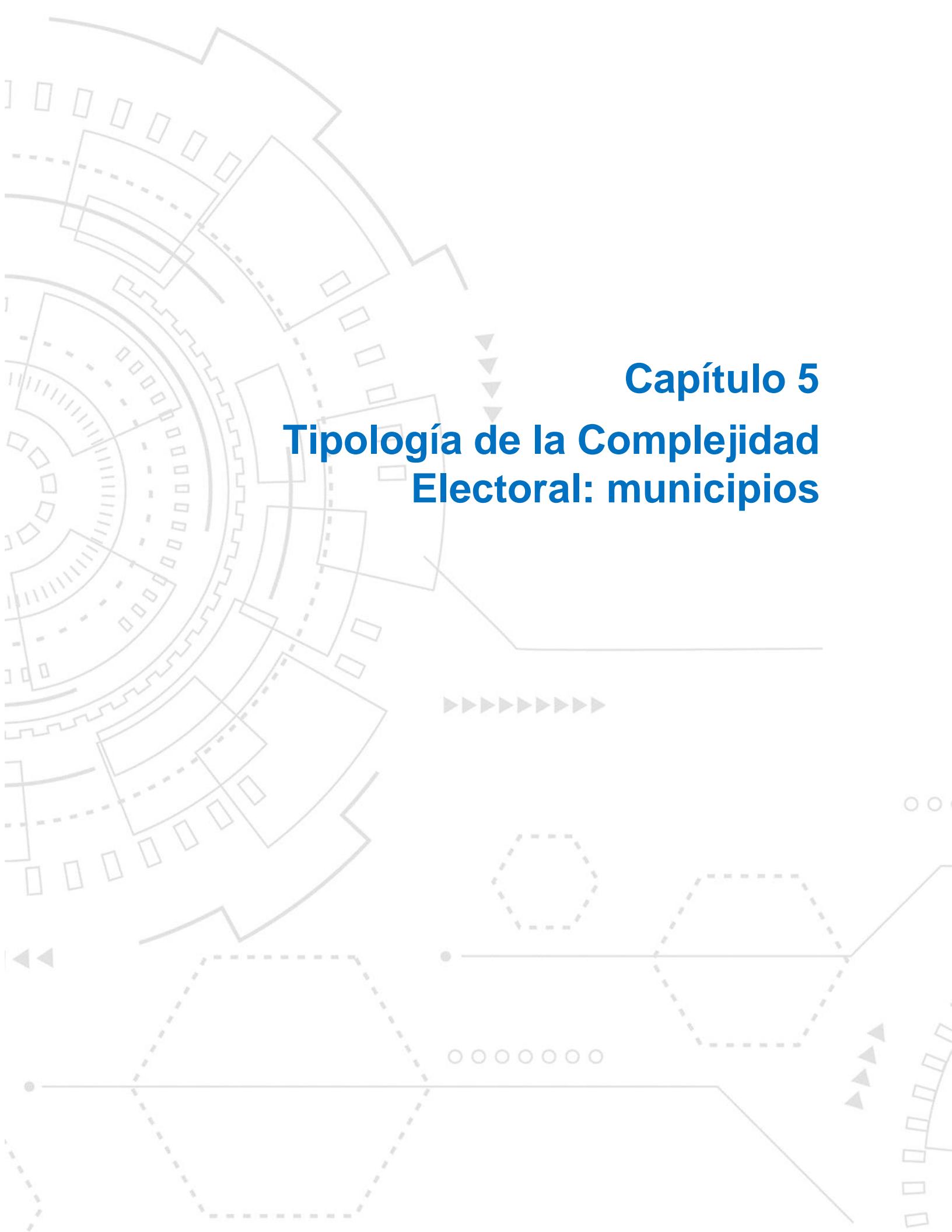


Figura 4. Ranking seccional de complejidad electoral en la zona metropolitana de la Ciudad de México

Fuente: Elaboración propia con QGIS.



Capítulo 5

Tipología de la Complejidad Electoral: municipios

Capítulo 5. Tipología de la Complejidad Electoral: municipios

En este capítulo se presentan los resultados de la implementación del algoritmo de partición K-medias. Dicho algoritmo utiliza los resultados de la proyección de los municipios en el espacio de los ICE municipales para realizar un análisis de agrupamiento, el cual permite identificar grupos de municipios similares en su grado de complejidad electoral. Es decir, los resultados del análisis de agrupamiento permiten constituir una TCE a escala municipal.

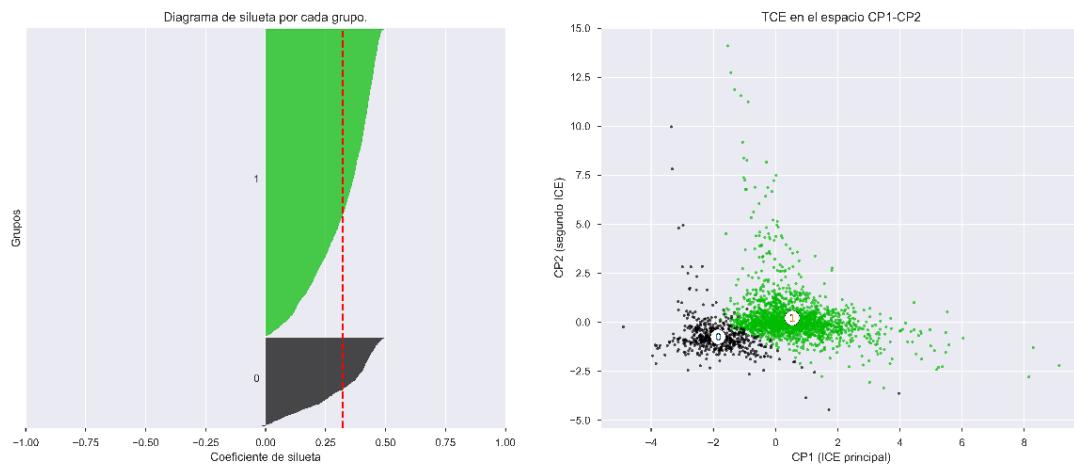
Para iniciar el método de K-medias se realiza un análisis de coeficientes de silueta para determinar el número óptimo de grupos en la TCE municipal. Una vez especificada la configuración de grupos de la TCE municipal, se clasifican los municipios en función de dicha tipología y se analizan las estadísticas de tendencia central de cada grupo de la TCE municipal para identificar sus rasgos característicos –en términos de los siete indicadores base—, e identificar los casos atípicos de complejidad electoral.

5.1 Análisis de silueta e implementación de K-medias: caso municipal

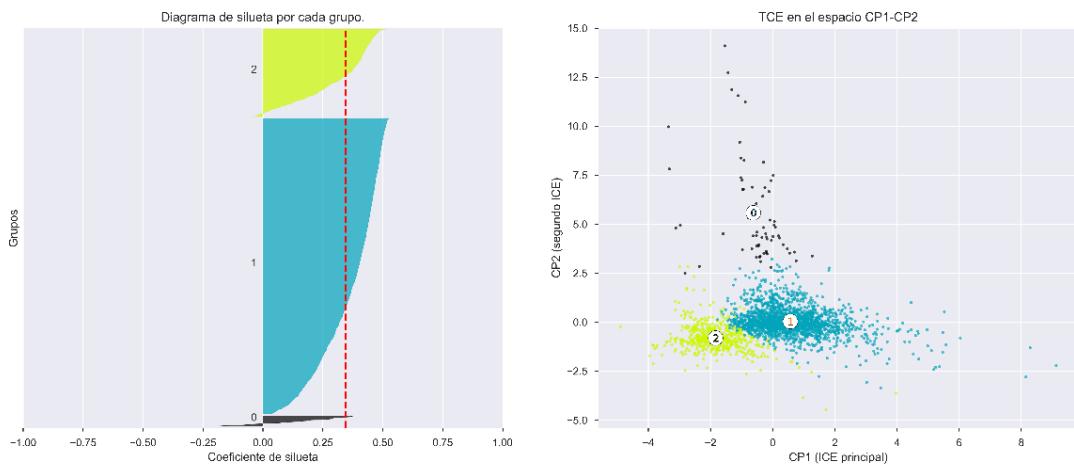
En el Gráfico 13 se muestran los diagramas de silueta correspondientes a cinco configuraciones de grupos analizadas con el algoritmo K-medias (configuraciones de 2, 3, 4, 5 y 6 grupos)¹⁰⁶. Los diagramas señalan que las configuraciones de 4 y 5 grupos son adecuadas para construir la TCE municipal, ya que presentan un número limitado de municipios agrupados incorrectamente, muestran el valor de silueta promedio más elevado y, además, las cardinalidades de sus grupos (es decir, el número de municipios contenidos en cada grupo) son relativamente equitativas.

¹⁰⁶ Para facilitar la visualización de los diagramas de silueta, y el correspondiente agrupamiento en el espacio de las Componentes Principales, sólo se grafica la representación bidimensional. Por otra parte, se recuerda que para la determinación de la distancia entre los centroides y los puntos de datos se utilizó la métrica Euclíadiana.

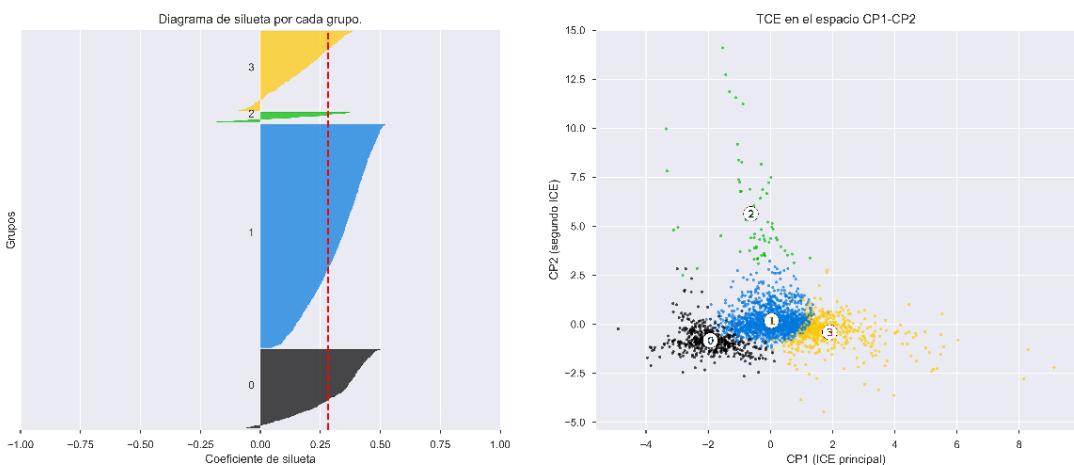
Análisis de silueta con K-medias (7 indicadores base). Número de grupos = 2



Análisis de silueta con K-medias (7 indicadores base). Número de grupos = 3



Análisis de silueta con K-medias (7 indicadores base). Número de grupos = 4



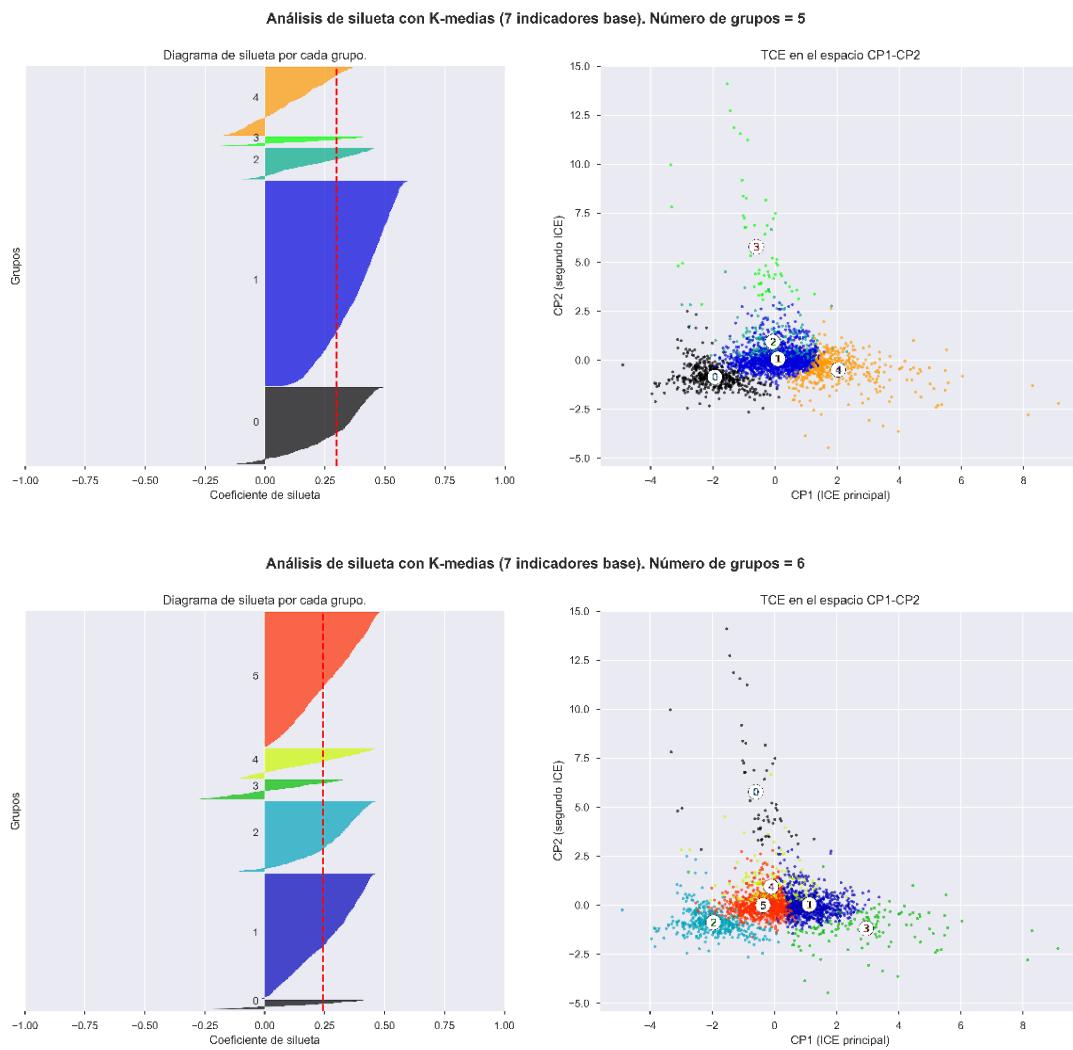


Gráfico 13. *TCE municipal: Diagramas de silueta para configuraciones de 2, 3, 4, 5 y 6 grupos*

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn

Un análisis más detallado de las configuraciones de 4 y 5 grupos muestra que la configuración de 5 grupos se puede considerar óptima, ya que ofrece una distribución más equitativa de los municipios en los grupos formados y separa en un grupo particular aquellos municipios que presentan una gran separación, a lo largo del eje CP2, respecto al conjunto mayoritario de municipios.¹⁰⁷ En consecuencia, los

¹⁰⁷ Se recomienda ver el archivo adjunto en Álvarez Hernández (2020) para observar con detalle los resultados del análisis de coeficientes de silueta.

resultados del análisis de coeficientes de silueta sugieren que una configuración de 5 grupos es la más adecuada para formar la TCE municipal.

En el Gráfico 14 se presenta el resultado de la TCE municipal en el subespacio CP1-CP2; específicamente, la configuración óptima de 5 grupos (etiquetas G0, G1, G2, G3, G4).¹⁰⁸ Se puede observar que la tipología distingue claramente, a lo largo de la dirección del eje CP1, a los municipios que tienen una valoración positiva en el ICE principal (municipios clasificados en los grupos G0, G1 y G4). Asimismo, a lo largo de la dirección del eje CP2, la tipología separa los municipios que tienen una valoración positiva en el segundo ICE (grupos G2, G3 y, en menor medida, el grupo G1) de los municipios que presentan una valoración negativa (en su mayoría, los municipios clasificados en los grupos G0 y G4).

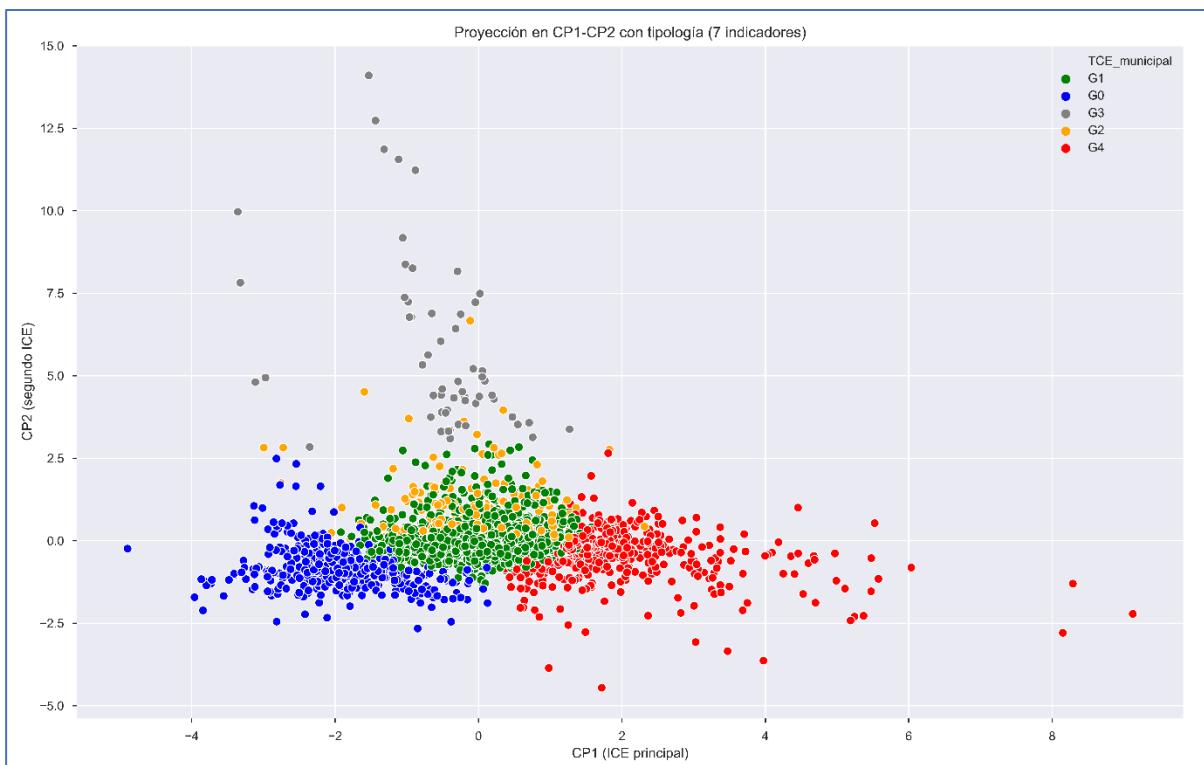


Gráfico 14. *TCE municipal con K-medias. Subespacio CP1-CP2*

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn-Seaborn

¹⁰⁸ El número de municipios por grupos de la TCE municipal es el siguiente: 1,288 municipios en el grupo G1, 485 municipios en el grupo G0, 428 municipios en el grupo G4, 199 municipios en el grupo G2, y 58 municipios en el grupo G3.

Por otra parte, en el Gráfico 15 se muestra la TCE municipal en el espacio tridimensional CP1-CP2-CP3. En este gráfico se observa más claramente la estructura dimensional que presenta la dispersión de los municipios en el espacio de los ICE municipales, así como el ajuste de la tipología a la estructura tridimensional de la proyección. Por ejemplo, se puede apreciar que los municipios clasificados en el grupo G2 (puntos en color amarillo) presentan valores superiores en el tercer ICE municipal, a diferencia de los municipios clasificados en el grupo G1 (puntos en color verde). Esta característica no se puede apreciar en la proyección CP1-CP2 (mostrada en el Gráfico 14), ya que en el espacio bidimensional se observa una aparente yuxtaposición entre los grupos G1 y G2, la cual sólo es un efecto que emerge de la reducción de dimensionalidad.¹⁰⁹

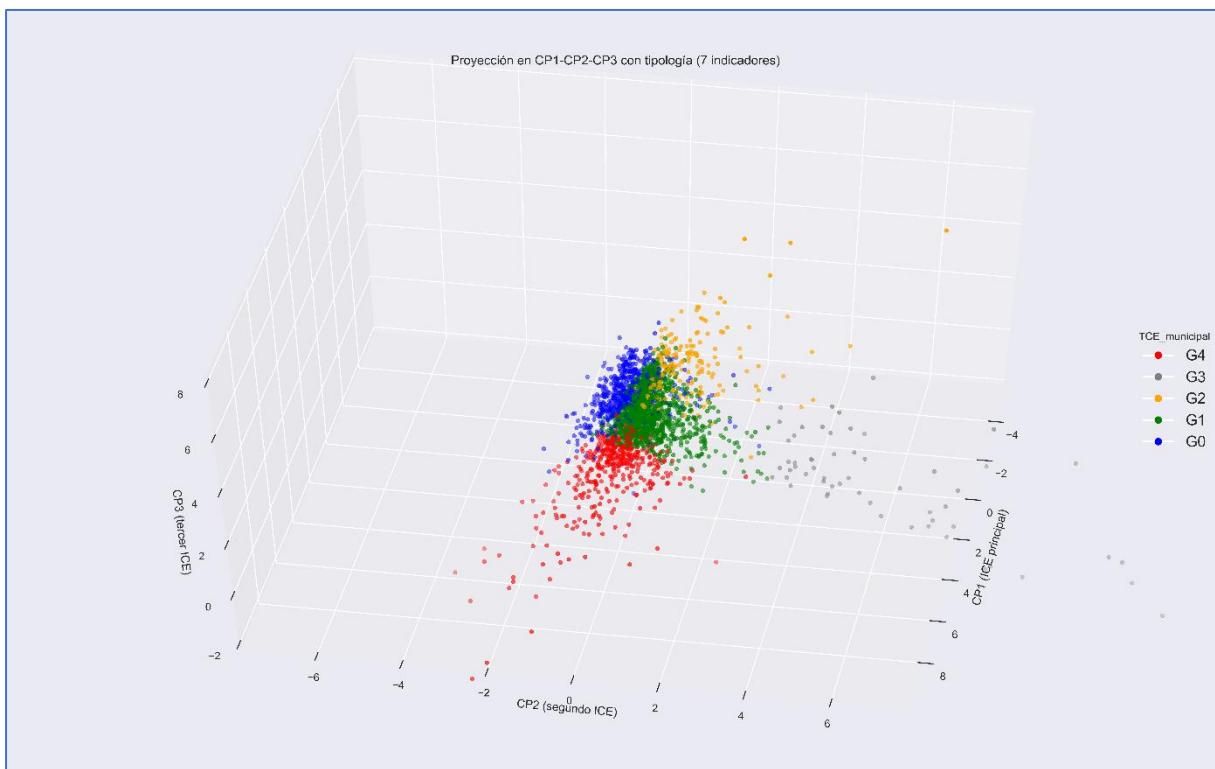


Gráfico 15. TCE municipal con K-medias. Subespacio CP1-CP2-CP3

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn–Seaborn

¹⁰⁹ Se recomienda revisar el archivo anexo en el repositorio de GitHub de Álvarez Hernández (2020) para visualizar adecuadamente las agrupaciones de la TCE municipal.

A partir de los resultados visuales mostrados en los Gráficos 14 y 15, y utilizando la interpretación de los ICE municipales (explicada con detalle en el Capítulo 4), se pueden deducir *a priori* algunas propiedades generales de los grupos de la TCE municipal.¹¹⁰ Por ejemplo, los municipios clasificados en el grupo G4 (y en menor medida, algunos municipios de los grupos G1 y G2) muestran características de alta complejidad electoral asociada al factor demográfico (alta variabilidad poblacional en la LNE, crecimiento demográfico positivo en la LNE y bajos niveles de actualización entre la LNE y el PE). En contraste, los municipios clasificados en los grupos G0 y G3 (junto con algunos municipios de los grupos G1 y G2) muestran una complejidad electoral inferior, ya que su valoración negativa en el ICE principal implica una baja variabilidad poblacional, un decrecimiento demográfico en la LNE y mayores niveles de actualización entre la LNE y el PE.

Asimismo, los municipios clasificados en los grupos G2 y G3 destacan por su valoración positiva en el segundo ICE municipal, lo cual implica que son municipios de carácter urbano (distinguidos por su electorado numeroso, poca dispersión geográfica del electorado, y altos niveles de actualización entre la LNE y el PE). Por el contrario, los municipios pertenecientes a los grupos restantes (G0, G1 y G4) presentan características propias de municipios de carácter rural o semiurbano (cercanos geográficamente a zonas metropolitanas).

Por último, destacan las características de los municipios clasificados en el grupo G1, ya que en general estos municipios presentan una valoración nula en los primeros tres ICE municipales (municipios conglomerados alrededor del centro de las proyecciones bidimensional y tridimensional).¹¹¹ Es decir, estos municipios no presentan tasas positivas o negativas de crecimiento demográfico en la LNE (son municipios estables desde el punto de vista demográfico), no tienen un electorado numeroso, y presentan una dispersión geográfica del electorado intermedia entre los municipios de carácter urbano y los municipios de carácter rural.

¹¹⁰ En la siguiente sección del capítulo se analizarán con detalle las estadísticas de los grupos de la TCE municipal.

¹¹¹ Cabe destacar que en este grupo de la TCE municipal se concentra aproximadamente el 52% (1,288) de los municipios clasificados.

En la Figura 5 se muestra el mapeo de la TCE municipal a nivel nacional, y, en la Figura 6, se muestra un acercamiento del mapeo de la TCE municipal en la zona metropolitana de la Ciudad de México.

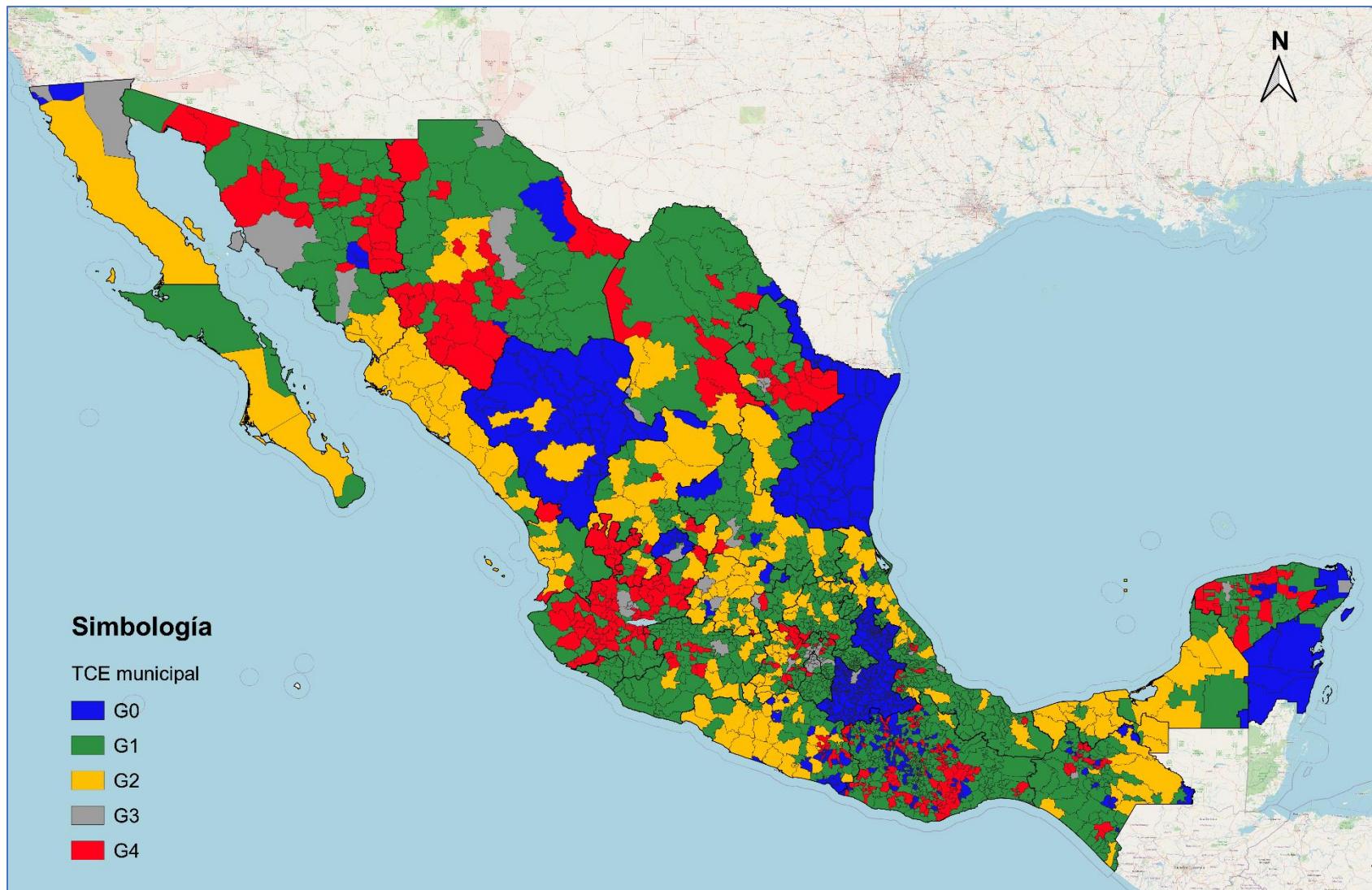


Figura 5. Mapa nacional de municipios con su respectiva clasificación en la TCE municipal

Fuente: Elaboración propia con QGIS

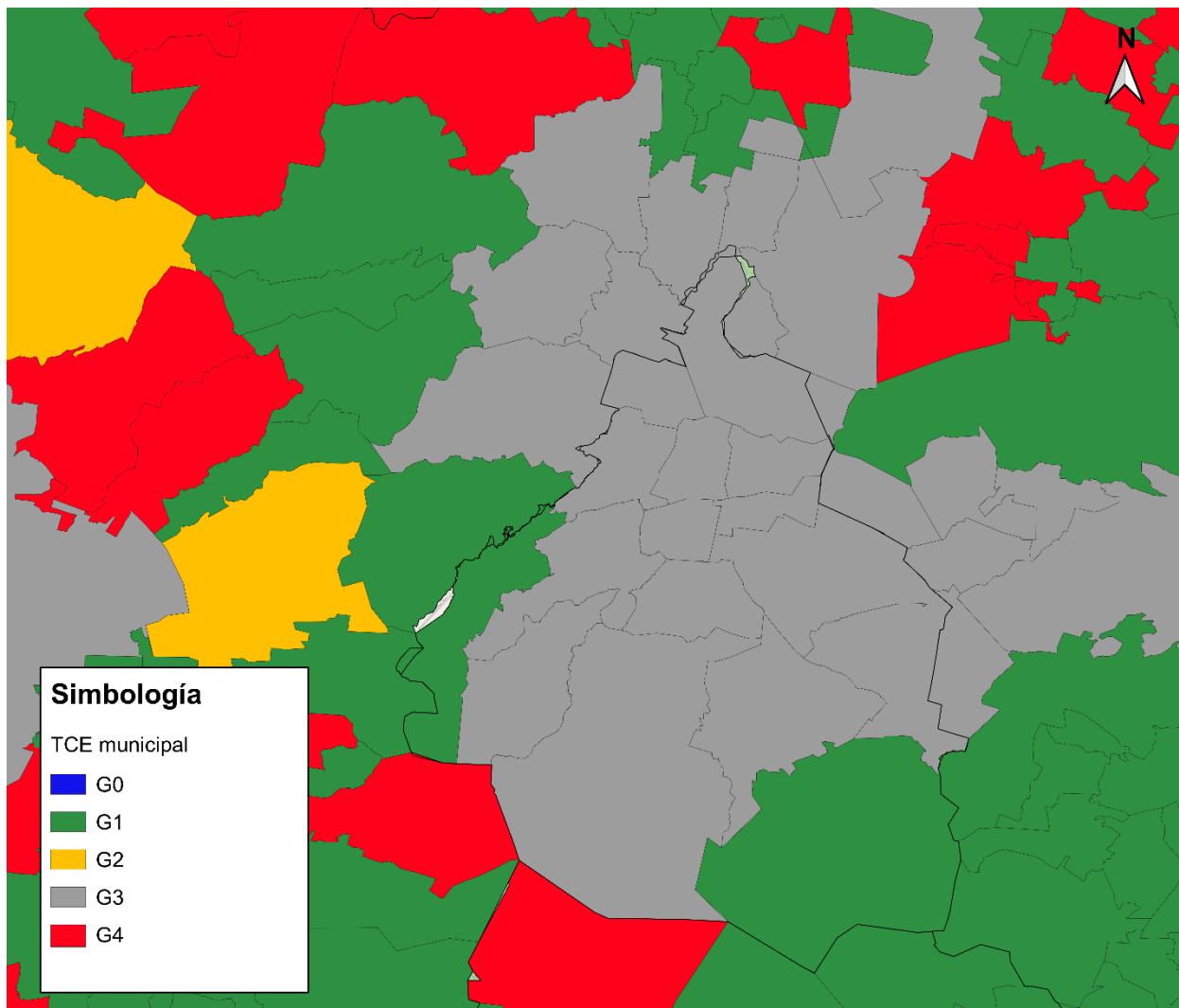


Figura 6. Mapa de los municipios de la zona metropolitana de la Ciudad de México con su respectiva clasificación en la TCE municipal

Fuente: Elaboración propia con QGIS

Las Figuras 5 y 6 permiten apreciar que las características deducidas *a priori* de la TCE municipal —auxiliados con la interpretación de los ICE municipales— se corresponden con las características geográficas y demográficas que se observan en el territorio mexicano. Por ejemplo, se observa que los municipios ubicados en las principales zonas urbanas de México se encuentran clasificados dentro del grupo G3 (municipios coloreados en color gris), lo cual concuerda con las características identificadas previamente. Asimismo, se observa que los municipios ubicados en zonas de poca densidad poblacional (por ejemplo, en las zonas montañosas del occidente, oriente y sur de México) son correctamente clasificados dentro de los grupos G0, G1 y G4; en particular, los municipios G4 muestran una coincidencia con las zonas sujetas a procesos de expulsión o atracción poblacional (como pudo ser el caso de los municipios G4 colindantes al área metropolitana de la Ciudad de México).

Para finalizar esta sección, es importante mencionar que, en los archivos adjuntos al presente trabajo se proporciona el listado completo de los municipios con su respectiva clasificación en la TCE municipal.¹¹² Asimismo, en el Anexo II se proporciona el pseudocódigo de Python implementado para generar el análisis de coeficientes de silueta y el algoritmo de K-medias.

5.2 Estadísticas de los grupos de la TCE municipal

En esta sección se analizan y comparan —mediante la elaboración de gráficas de violín—¹¹³ las distribuciones y estadísticas de tendencia central de los siete indicadores base, considerando la clasificación de grupos generada en la TCE municipal.

Con el propósito de ayudar a la comprensión del panel de diagramas de violín —presentado en el Gráfico 17—, se analiza en primera instancia el caso particular del diagrama de violín correspondiente al ICE principal (ver Gráfico 16).

En el Gráfico 16 se puede observar que los grupos G1, G2 y G3 de la TCE municipal presentan estadísticas de tendencia central similares, ya que exhiben un valor medio equiparable en el ICE principal (las medias se encuentran indicadas por

¹¹² Véase el estadístico adjunto en el repositorio de GitHub de Álvarez Hernández (2020).

¹¹³ Para más detalles sobre la interpretación de las gráficas de violín se recomienda revisar Hintze y Nelson (1998).

los puntos blancos dentro de las figuras de violín), y algo similar se observa respecto al tamaño de los rangos intercuartílicos (representados por los rectángulos negros dentro de las figuras de violín), y la longitud de los intervalos del 95% de confianza (descritos por las líneas verticales que salen de los rectángulos negros). En contraste, los grupos G0 y G4 de la TCE municipal muestran características significativamente diferentes en el ICE principal, ya que el grupo G4 exhibe una media más alta respecto a los demás grupos, mientras que, comparativamente, el grupo G0 presenta el menor valor promedio en el ICE principal. Es decir, el diagrama de violín indica que, comparativamente, los municipios clasificados en el grupo G4 presentan una complejidad electoral más alta en el ICE principal.

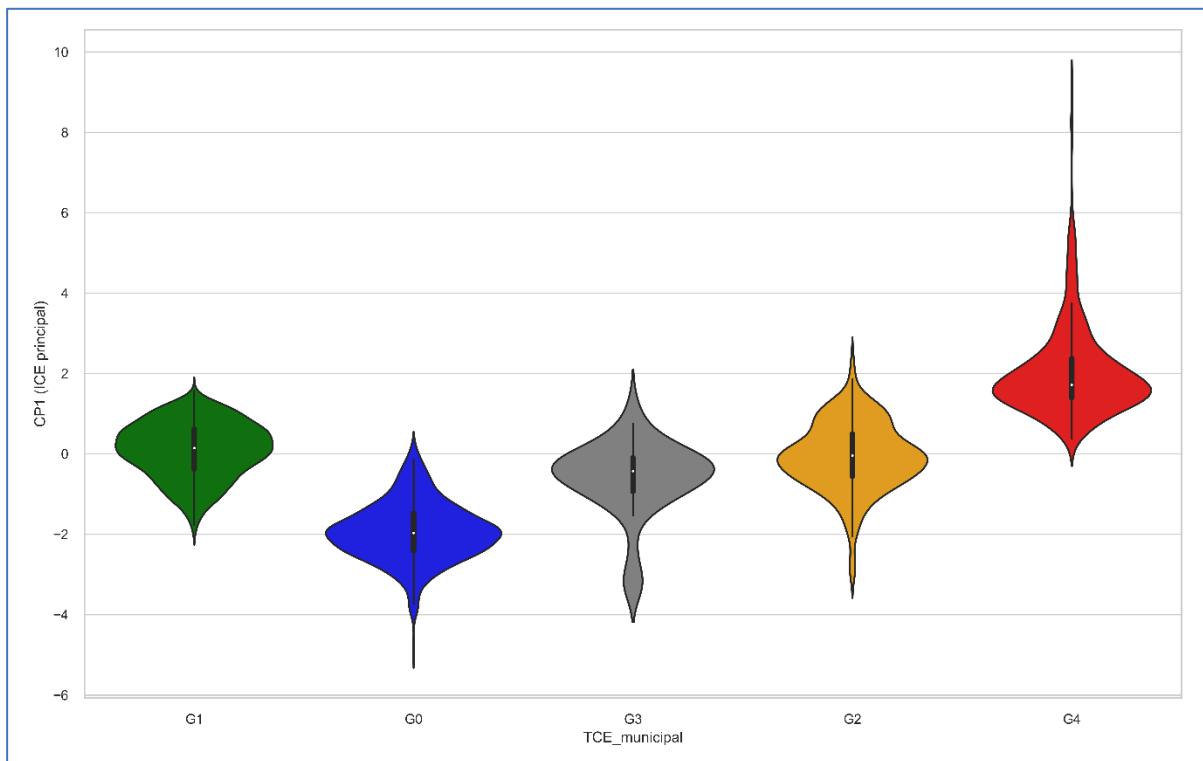


Gráfico 16. Diagrama de violín del ICE principal (con agrupación en función de la TCE municipal)

Fuente: Elaboración propia a partir de la librería Python–Seaborn

Por otra parte, el Gráfico 16 también muestra, a través de las figuras de violín, la forma de las distribuciones de valores correspondientes a cada grupo de la TCE municipal. En particular, se puede observar que los grupos G0, G1, G2 y G3 presentan

pocos valores atípicos (figuras de violín cortas), mientras que el grupo G4 exhibe un mayor número de valores atípicos (figura de violín alargada); es decir, la distribución del grupo G4 presenta un mayor comportamiento leptocúrtico respecto a los demás grupos de la TCE municipal.¹¹⁴

Teniendo presente la interpretación de la información contenida en el Gráfico 16, se procede al análisis del panel de diagramas de violín mostrado en el Gráfico 17.

Como se puede observar en el Gráfico 17, los cinco grupos de la TCE municipal muestran estadísticas similares (media, rango intercuartílico e intervalos de confianza) en los indicadores Nivel de Actualización del RFE (Razon_LNE_PE), Tasa de Crecimiento Anual de la LNE (TC_LNE_2019) y Coeficiente de Variación de la LNE (Coef_Var_LNE). Sin embargo, en los indicadores de Grado de Urbanización Municipal (NUM_SECC_RUR), Magnitud del Electorado (LNE), Densidad del Electorado (Densidad_LNE) y Autocorrelación Mensual de la LNE (Autocorr_LNE), se aprecian diferencias significativas entre las estadísticas de tendencia central de los grupos G0, G2 y G3, y las estadísticas de los grupos restantes.

Por ejemplo, se puede apreciar que los municipios clasificados en el grupo G3 poseen en promedio un electorado más numeroso, y presentan una menor dispersión geográfica en su población (densidad alta). Por otra parte, los municipios pertenecientes al grupo G0 se caracterizan por presentar un bajo nivel de actualización entre la LNE y el PE, aunque no presentan características asociadas a la variabilidad demográfica. Y finalmente, los municipios del grupo G2 muestran en promedio un mayor número de secciones rurales comparado con los municipios de otros grupos, lo cual sugiere que los municipios de clasificación G2 son en esencia rurales.

¹¹⁴ Un efecto indirecto y a veces indeseado de la utilización de transformaciones logarítmicas en los datos es la desaparición de los valores atípicos, ya que dichas transformaciones normalizan en forma natural las distribuciones sesgadas. Para el caso específico de los indicadores base analizados, se encontró que no es pertinente aplicar trasformaciones logarítmicas, ya que las distribuciones de los indicadores son por naturaleza no normales; es decir, existen municipios que presentan intrínsecamente características anómalas, las cuales no representan errores de información.

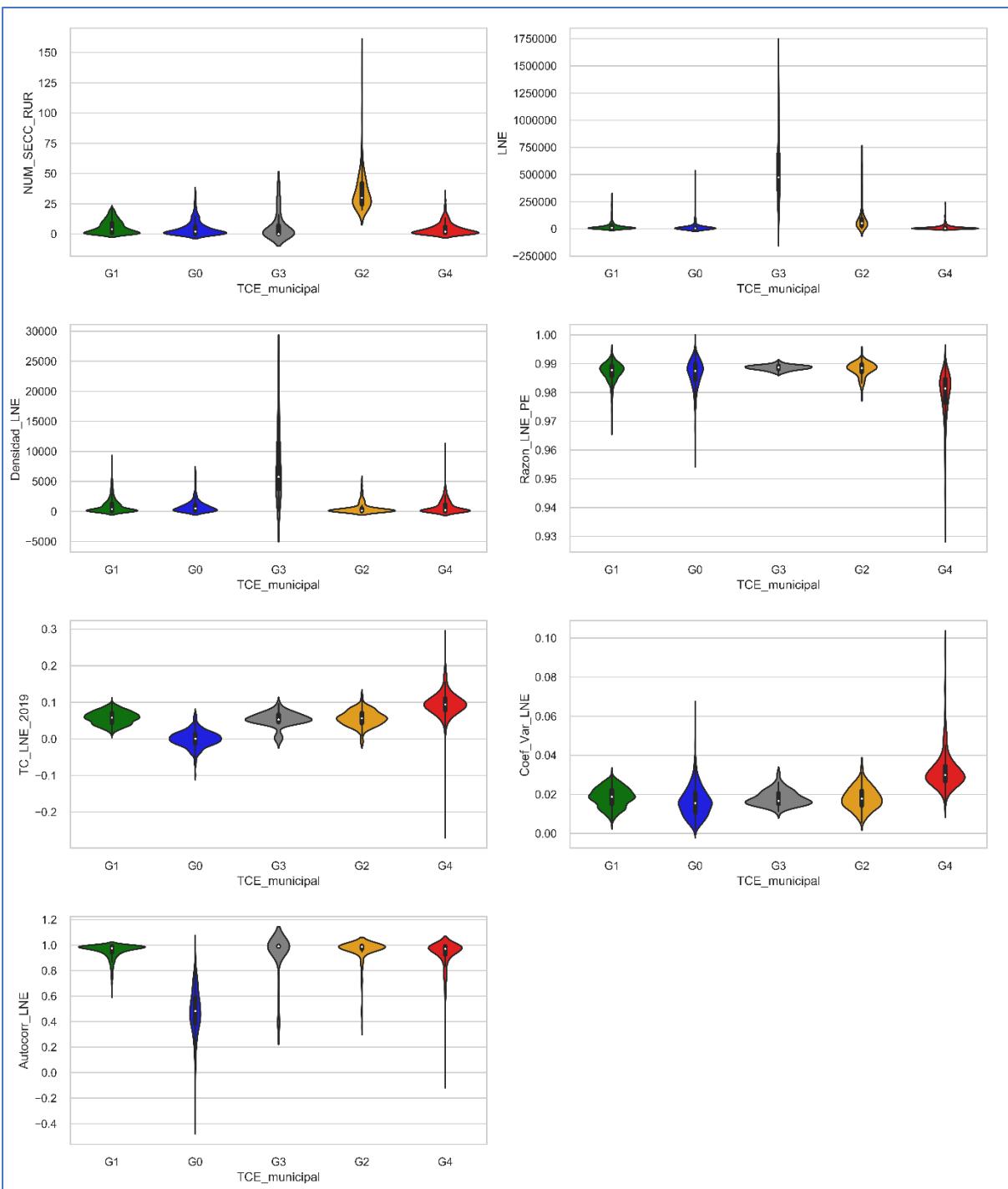


Gráfico 17. Panel de diagramas de violín de los siete indicadores base (con agrupación en función de la TCE municipal)

Fuente: Elaboración propia a partir de la librería Python–Seaborn

Respecto a las distribuciones de cada indicador base, el Gráfico 17 muestra la existencia de diferencias comparativas en los cinco grupos de la tipología municipal.

En particular, los diagramas de violín indican que todas las distribuciones de los indicadores son asimétricas, con diferentes grados y direcciones de sesgo. En relación con la presencia o ausencia de colas largas en las distribuciones de los indicadores (agrupados en función de la TCE municipal), se puede observar que en los cinco grupos de la tipología municipal hay un comportamiento leptocúrtico, ya que hay una cantidad significativa de municipios fuera de rango (valores atípicos).

Por otra parte, es importante analizar las estadísticas y las distribuciones observadas en los ICE municipales con el fin de identificar los municipios con niveles atípicos de complejidad electoral, ya que, desde una perspectiva institucional y electoral, estos municipios presentan características demográficas, geográficas y estructurales anómalas que pueden afectar potencialmente el desarrollo de las tareas IIMA. En la siguiente sección se abunda más sobre el proceso de identificación de los casos atípicos, así como las características de los municipios identificados en esta categoría.

5.3 Identificación de casos atípicos a nivel municipal

Como se explicó en el Capítulo 2, el proceso de identificación de municipios con características atípicas de complejidad electoral consiste en examinar, mediante la regla del rango intercuartílico y conforme a las agrupaciones generadas en la TCE municipal, las distribuciones de los ICE municipales.¹¹⁵ Los resultados del análisis de atipicidad a nivel municipal (ver mapa en la Figura 7) indican la existencia de tres grupos de municipios atípicos que, dadas sus características demográficas y geográficas, son de particular interés para la identificación de las zonas territoriales de México que presentan factores de riesgo para el desarrollo de las tareas IIMA. A continuación, se describen las características de los tres grupos identificados.¹¹⁶

El primer grupo está conformado por municipios que presentan una valoración atípica en el ICE principal y que además se encuentran clasificados en los grupos G0

¹¹⁵ En particular, sólo se realiza un análisis de atipicidad para los primeros tres ICE municipales, considerando que estos explican cerca del 72% de la varianza mostrada en los indicadores municipales base.

¹¹⁶ En el estadístico adjunto en el repositorio de GitHub de Álvarez Hernández (2020) se identifica, con la variable-columna Atipicidad, aquellos municipios que muestran una complejidad electoral anómala en los primeros tres ICE municipales.

y G4. Estos municipios (51 en total) se caracterizan por tener un electorado pequeño, una densidad de electores baja (comparado con municipios ubicados en zonas metropolitanas), una alta variabilidad demográfica (altas tasas de crecimiento/decrecimiento poblacional), y por concentrarse principalmente en las regiones rurales de las entidades de Chihuahua, Jalisco y Oaxaca. Por consiguiente, considerando las características propias de este grupo de municipios, se ofrecen las siguientes propuestas, encaminadas a evaluar y/o resolver las potenciales problemáticas asociadas a la condición de atipicidad del grupo:

1) Evaluar la calidad de los servicios que el INE presta a la ciudadanía residente en estos municipios atípicos, especialmente en lo que respecta a los procesos de empadronamiento y votación. La variabilidad demográfica de estos municipios atípicos tiene un efecto directo en los niveles de actualización del PE y la LNE, lo cual hace necesario otorgar a dichos municipios una mayor asignación de recursos financieros y humanos orientados a la realización de campañas de credencialización y/o actualización de datos personales.

2) Identificar los factores primordiales que provocan el dinamismo demográfico observado en estos municipios; por ejemplo, si los flujos migratorios (internos y/o externos) o la evolución natural de la población (envejecimiento) tienen el papel preponderante en la dinámica demográfica. Asimismo, es tema de análisis identificar en estos municipios atípicos el porcentaje de solicitudes de CPV realizados con actas de nacimiento extemporáneas,¹¹⁷ ya que esto permitiría identificar presuntas inscripciones irregulares al PE.

3) Analizar la cobertura geográfica de los Módulos de Atención Ciudadana ubicados en los municipios de este grupo, ya que se debe considerar que estos municipios se encuentran alejados considerablemente de los principales centros urbanos y presentan una alta dispersión del electorado, por lo que es necesario

¹¹⁷ Como su nombre lo indica, las actas de nacimiento extemporáneas son actas que se expediten tiempo después al momento del nacimiento, y usualmente, son tramitadas por el ciudadano en cuestión. Este documento es utilizado frecuentemente para mostrar, o tramitar, la ciudadanía mexicana y la residencia en una entidad federativa particular, requisito esencial para poder tramitar una CPV. Por estas razones, el INE monitorea el uso de las actas extemporáneas, ya que dicho documento se puede prestar para la realización de trámites irregulares de inscripción al RFE.

aumentar la cobertura del sistema de módulos mediante la implementación de estrategias de empadronamiento móvil.¹¹⁸

El segundo grupo está conformado por municipios que presentan una valoración atípica en el segundo ICE municipal y que además se encuentran clasificados en los grupos G2 y G3. Estos municipios (14 en total) se caracterizan por tener un electorado cuantioso, una alta densidad poblacional, y un crecimiento demográfico prominente comparado con el promedio observado a nivel nacional. Asimismo, se observa un patrón en la distribución geoespacial de los municipios de este grupo, ya que la mayoría se encuentra localizado en zonas urbanas correspondientes a ciudades de mediana magnitud poblacional (por ejemplo, Acapulco, Celaya, Culiacán y Saltillo), y sólo tres alcaldías/municipios (Iztapalapa, Nezahualcóyotl y García) se encuentran ubicados en las zonas metropolitanas de la Ciudad de México y Monterrey.

Por consiguiente, para este grupo de municipios atípicos se proponen las siguientes medidas:

1) Dar seguimiento periódico a las estadísticas de opinión sobre la calidad del servicio brindado en los Módulos de Atención Ciudadana, ya que la magnitud del electorado puede representar una mayor carga de trabajo para el personal de módulos, lo cual implicaría asignar una mayor cantidad de recursos financieros y humanos para mantener en un nivel óptimo la eficiencia de las tareas de credencialización y actualización.

2) Evaluar las campañas de difusión, relacionadas con el trámite o la renovación de la Credencial para Votar, que se implementan en los municipios atípicos de este grupo, ya que éstos presentan un mayor nivel de desactualización en el PE y la LNE,

¹¹⁸ Actualmente, el sistema de Módulos de Atención Ciudadano está conformado por tres tipos de módulo: los módulos fijos (tienen una ubicación geográfica fija, y cuentan con un mayor número de equipos y personal de operación), los módulos semifijos (también tienen una ubicación geográfica fija, pero su operación está sujeta a la disponibilidad de recursos y personal de operación), y los módulos móviles (vagonetas que cuentan con un solo equipo de cómputo, y que están dedicadas a atender exclusivamente las localidades muy alejadas de los grandes centro de población).

por lo que es necesario incentivar a la ciudadanía a que mantenga actualizados sus datos de registro electoral.

3) Identificar los factores primordiales que provocan el dinamismo demográfico observado en estos municipios, y determinar el porcentaje de solicitudes de CPV realizados con actas de nacimiento extemporáneas, a fin de identificar presuntas inscripciones irregulares al PE.

Finalmente, el tercer grupo está conformado por municipios que presentan una valoración atípica en el tercer ICE municipal y que además se encuentran clasificados en los grupos G1 y G2. Estos municipios (10 en total) se caracterizan por tener un electorado pequeño, una densidad poblacional baja, y un elevado número de secciones electorales rurales. Respecto a la ubicación geoespacial de estos municipios, no se discierne un patrón geográfico claro, ya que estos se encuentran dispersados en diferentes zonas del territorio mexicano.

En consecuencia, y considerando las características de los municipios de este grupo, se propone:

1) Analizar la calidad de los servicios que el INE presta a la ciudadanía residente en estos municipios atípicos, especialmente en lo que concierne a los procesos de empadronamiento y votación, ya que la dispersión geográfica del electorado puede estar afectando directamente la realización de las campañas de credencialización y/o actualización de datos personales.

2) Evaluar la cobertura geográfica de los Módulos de Atención Ciudadana ubicados en los municipios de este grupo, ya que se debe considerar que estos municipios presentan un elevado número de secciones electorales rurales, por lo que puede ser necesario aumentar la cobertura del sistema de módulos mediante la implementación de estrategias de empadronamiento móvil.

Para finalizar esta sección, en la Figura 7 se presenta el mapeo geográfico de los municipios identificados como atípicos, y en los archivos adjuntos se hace un desglose completo de los municipios con su respectiva valoración en los ICE

municipales, su agrupación en la TCE municipal, y su tipo de atipicidad en los ICE municipales.¹¹⁹

¹¹⁹ Véanse los archivos [Resultados_Mun_Indic_PCA_Kmeans_Atip_ind7.csv](#), y [Pyplotly_Proyección3d_Mun_ind7.html](#) adjuntos en el repositorio de GitHub de Álvarez Hernández (2020).

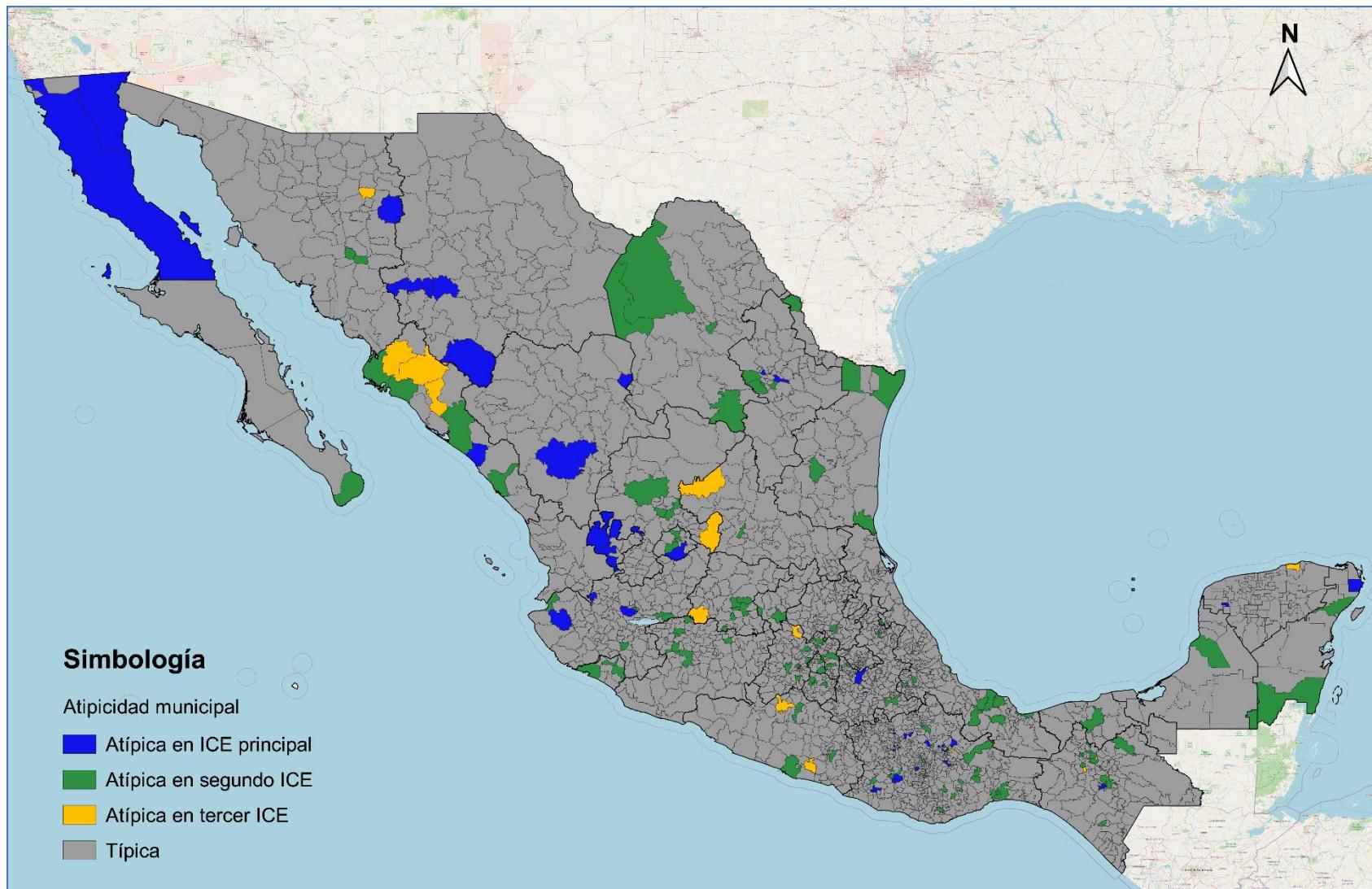


Figura 7. Mapa nacional de municipios con una complejidad electoral atípica

Fuente: Elaboración propia con QGIS

Capítulo 6

Tipología de la Complejidad Electoral: secciones electorales

Capítulo 6. Tipología de la Complejidad Electoral: secciones electorales

En este capítulo se presentan los resultados de la implementación del algoritmo de partición K-medias para construir la TCE a escala seccional. En forma similar a lo que se realizó en el Capítulo 5, el algoritmo K-medias utiliza los resultados de la proyección de las secciones electorales en el espacio de los ICE seccionales para realizar un análisis de agrupamiento, el cual permite formar grupos de secciones con grados similares de complejidad electoral.¹²⁰

Por consiguiente, se realiza un análisis de coeficientes de silueta para determinar la configuración óptima de grupos en la TCE seccional. Se clasifica el conjunto total de secciones electorales en función de la TCE seccional y se analizan las estadísticas de tendencia central de cada grupo de la tipología, esto con el fin de identificar los rasgos característicos de cada grupo –en términos de los seis indicadores base—, así como los casos atípicos de complejidad electoral.

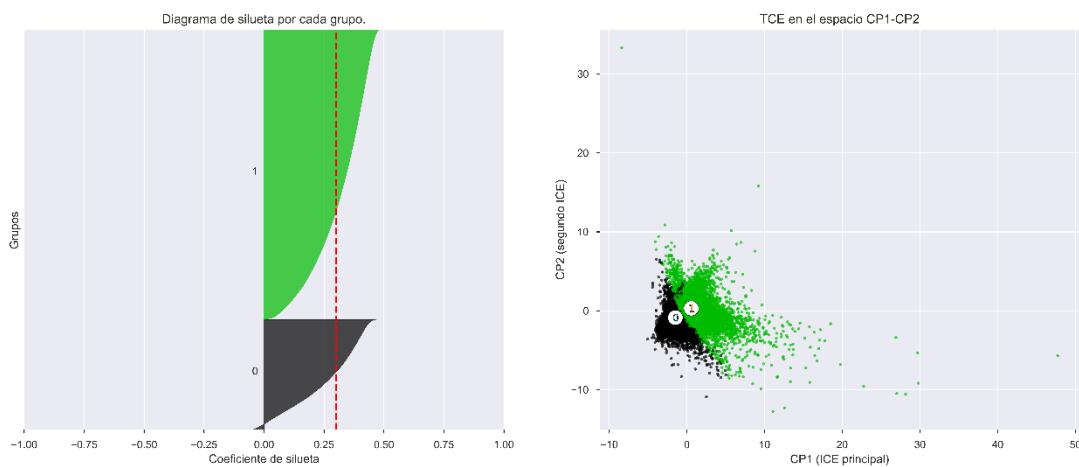
6.1 Análisis de silueta e implementación de K-medias: caso seccional

En el Gráfico 18 se muestran los diagramas de silueta –obtenidos con el algoritmo K-medias— de cinco posibles configuraciones de grupos (configuraciones de 2, 3, 4, 5 y 6 grupos). Los diagramas señalan que las configuraciones de 2, 5 y 6 grupos no son óptimas, ya que muestran un número significativo de secciones electorales agrupadas incorrectamente,¹²¹ o el tamaño de los conglomerados (número de secciones electorales en el grupo) es significativamente inequitativo. En contraste, las configuraciones de 3 y 4 grupos son superiores, ya que presentan un número bajo de secciones electorales agrupadas incorrectamente y el número de secciones electorales por conglomerado se encuentra mejor distribuido.

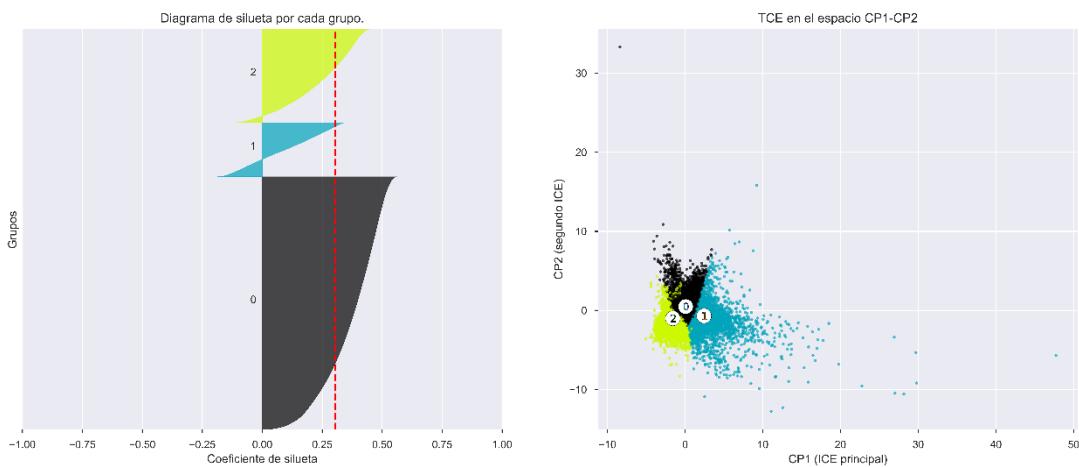
¹²⁰ Al igual que en el análisis de agrupamiento realizado en el caso municipal, se utiliza la métrica Euclíadiana como medida de distancia entre los centroides de los grupos y los puntos de datos. Asimismo, sólo se grafica la proyección bidimensional de los ICE seccionales para facilitar la visualización de los diagramas de silueta.

¹²¹ Secciones electorales con coeficientes de silueta negativos.

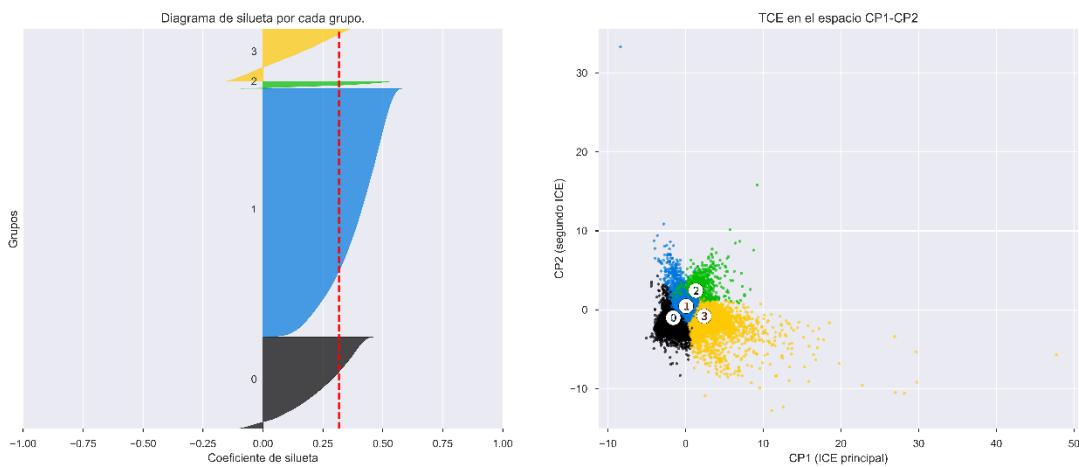
Análisis de silueta con K-medias (6 indicadores base). Número de grupos = 2



Análisis de silueta con K-medias (6 indicadores base). Número de grupos = 3



Análisis de silueta con K-medias (6 indicadores base). Número de grupos = 4



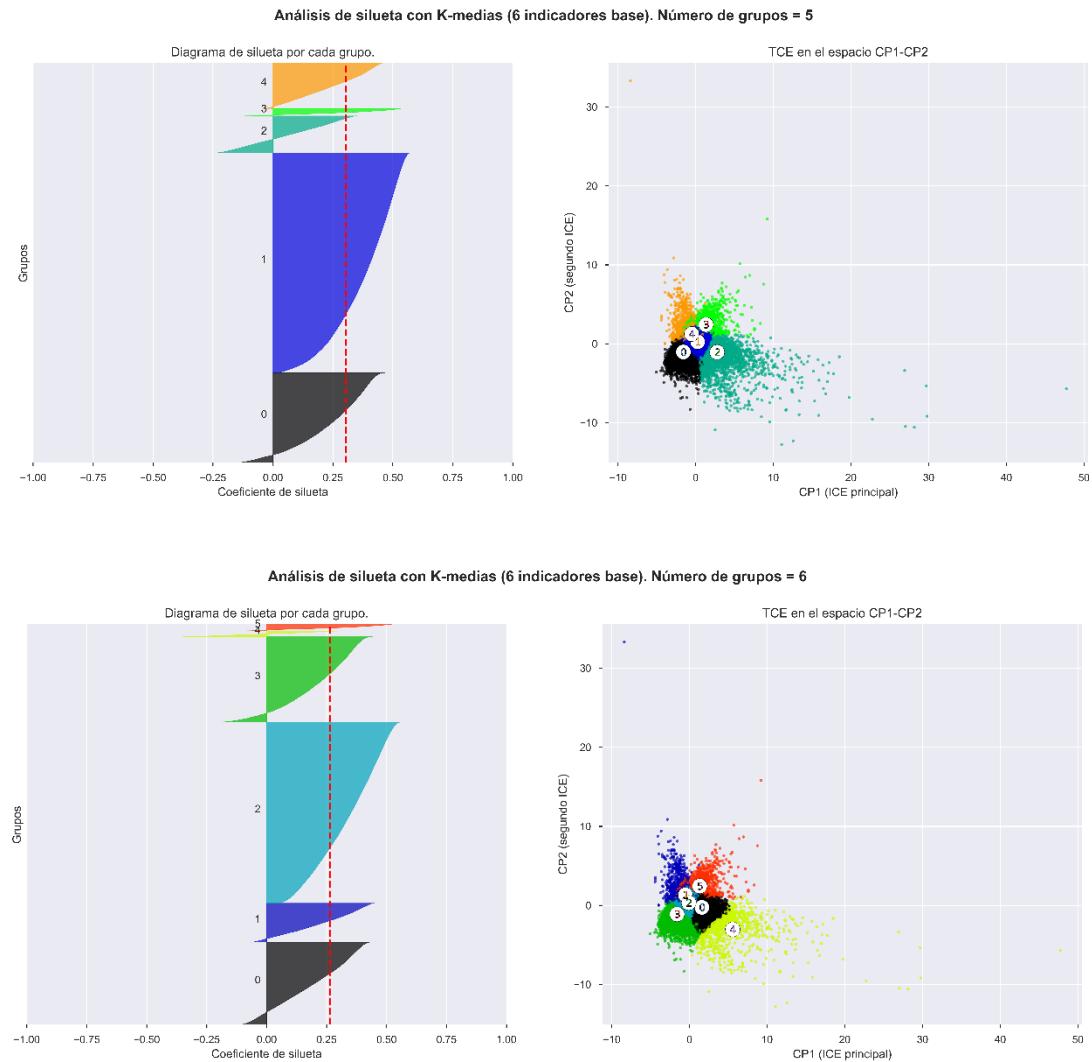


Gráfico 18. TCE seccional: Diagramas de silueta para configuraciones de 2, 3, 4, 5 y 6 grupos

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn

Un análisis más detallado de la configuración de 4 grupos muestra que en el caso seccional, esta configuración es óptima, ya que ofrece una distribución más equitativa de las secciones electorales en los grupos de la tipología. Es decir, los resultados obtenidos a partir del análisis de coeficientes de silueta indican que la configuración de 4 grupos es óptima para conformar la TCE seccional.

En el Gráfico 19 se presenta el resultado de la TCE seccional en el subespacio CP1-CP2, especificando la configuración óptima de 4 grupos (etiquetas G0, G1, G2 y

G3).¹²² En esta gráfica se puede observar que la TCE seccional distingue claramente, a lo largo de la dirección del eje CP1, las secciones electorales que tienen una valoración positiva en el ICE principal (secciones pertenecientes al grupo G2 y G3), y las secciones que obtienen una valoración negativa (secciones pertenecientes a los grupos G0 y G1). Asimismo, a lo largo de la dirección del eje CP2, la TCE seccional separa las secciones electorales que tienen una valoración positiva en el segundo ICE (grupos G1 y G2) de las secciones que presentan una valoración negativa (secciones pertenecientes al grupo G0 y G3).



Gráfico 19. TCE seccional con K-medias. Subespacio CP1-CP2

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn-Seaborn

Por otra parte, en el Gráfico 20 se muestra la tipología en el espacio de los primeros tres ICE seccionales (subespacio CP1-CP2-CP3). En esta gráfica se observa

¹²² El número de secciones electorales por grupo de la TCE seccional es el siguiente: 42,479 secciones en el grupo G1, 15,687 secciones en el grupo G0, 8,981 secciones en el grupo G3 y 1,243 secciones en el grupo G2.

más claramente la estructura dimensional que presenta la dispersión de las secciones electorales en el espacio de los ICE, y el ajuste de la TCE seccional a la estructura tridimensional de la proyección.¹²³

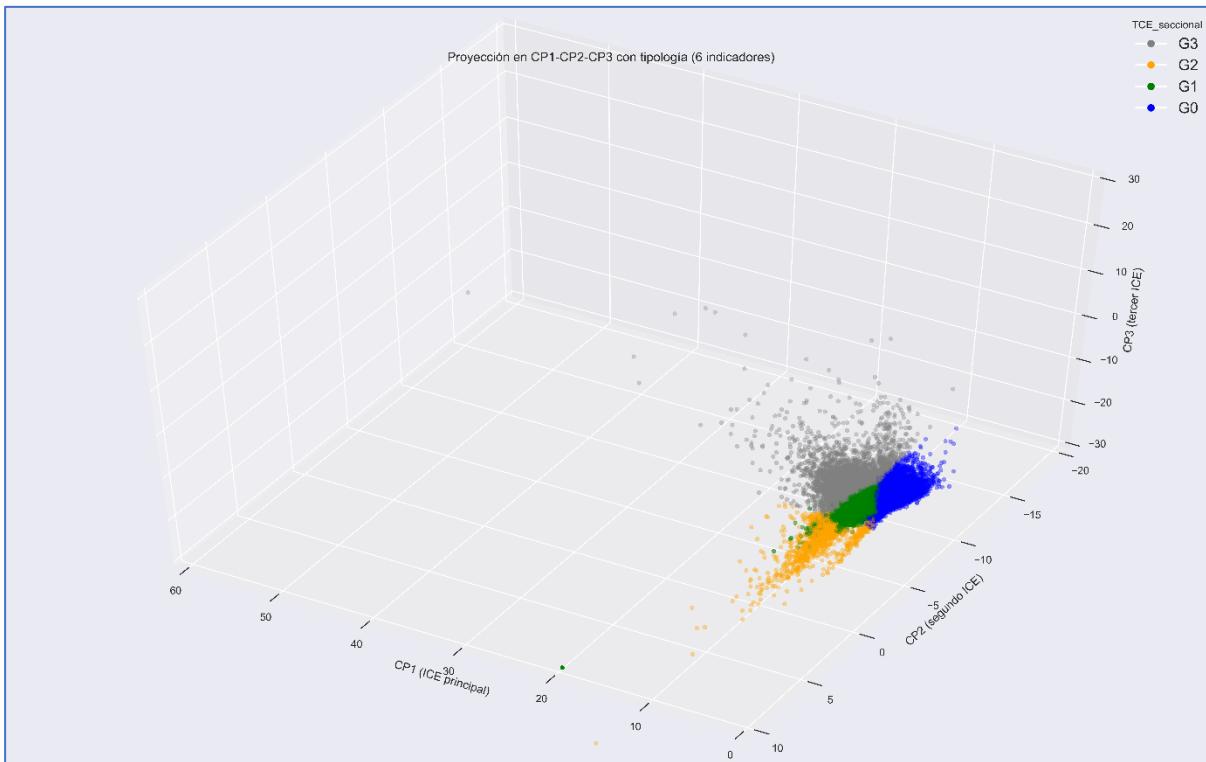


Gráfico 20. TCE seccional con *K*-medias. Subespacio CP1-CP2-CP3

Fuente: Elaboración propia a partir de la librería Python–Scikitlearn–Seaborn

A partir de los resultados visuales mostrados en los Gráficos 19 y 20, y recurriendo a la interpretación de los ICE seccionales (explicada con detalle en el Capítulo 5), se pueden derivar *a priori* algunas propiedades generales de los grupos de la TCE seccional.¹²⁴ Por ejemplo, las secciones electorales clasificadas en el grupo G3 (y en menor medida, algunas secciones del grupo G2) muestran características de alta complejidad electoral asociada al factor demográfico (alta variabilidad poblacional

¹²³ Se recomienda revisar la animación anexa en el repositorio de GitHub de Álvarez Hernández (2020), para ver en forma interactiva la proyección de las secciones electorales en el espacio de los ICE seccionales, junto con su correspondiente clasificación en la TCE seccional.

¹²⁴ En la siguiente sección del capítulo se analizarán con detalle las estadísticas de los grupos de la TCE seccional.

en la LNE, crecimiento demográfico positivo en la LNE y bajos niveles de actualización entre la LNE y el PE). En contraste, las secciones clasificadas en los grupos G0 y G1 muestran una complejidad electoral inferior, ya que su valoración negativa en el ICE principal implica una baja variabilidad poblacional, un decrecimiento demográfico en la LNE, y niveles más altos de actualización entre la LNE y el PE.

Por otra parte, las secciones clasificadas en los grupos G1 y G2 destacan por su valoración positiva en el segundo ICE seccional, lo cual implica que son secciones electorales de carácter urbano (distinguidas por tener un electorado numeroso, poca dispersión geográfica del electorado, y altos niveles de actualización entre la LNE y el PE). Por el contrario, las secciones electorales pertenecientes a los grupos restantes (G0 y G3) presentan características propias de municipios de carácter rural o semiurbano (un electorado pequeño y una mayor dispersión geográfica del electorado).

Por último, destacan las características de las secciones electorales clasificadas en el grupo G1, ya que, en general, estas secciones presentan una valoración nula en los primeros tres ICE seccionales (secciones conglomeradas alrededor del centro de las proyecciones bidimensional y tridimensional).¹²⁵ Es decir, estas secciones no presentan tasas positivas o negativas de crecimiento poblacional en la LNE (son secciones estables desde el punto de vista demográfico), no tienen un electorado numeroso (generalmente, respetan el límite máximo establecido de 3,000 electores por sección), y presentan una dispersión geográfica del electorado intermedia entre las secciones electorales de carácter urbano y las secciones electorales rurales.

En la Figura 8 se muestra el mapeo de la TCE seccional a nivel nacional y en la Figura 9 se muestra un acercamiento del mapeo de la TCE seccional en los alrededores de la zona metropolitana de la Ciudad de México.

¹²⁵ Cabe destacar que en este grupo de la TCE seccional se concentra aproximadamente el 62% (42,479) de las secciones electorales clasificadas.

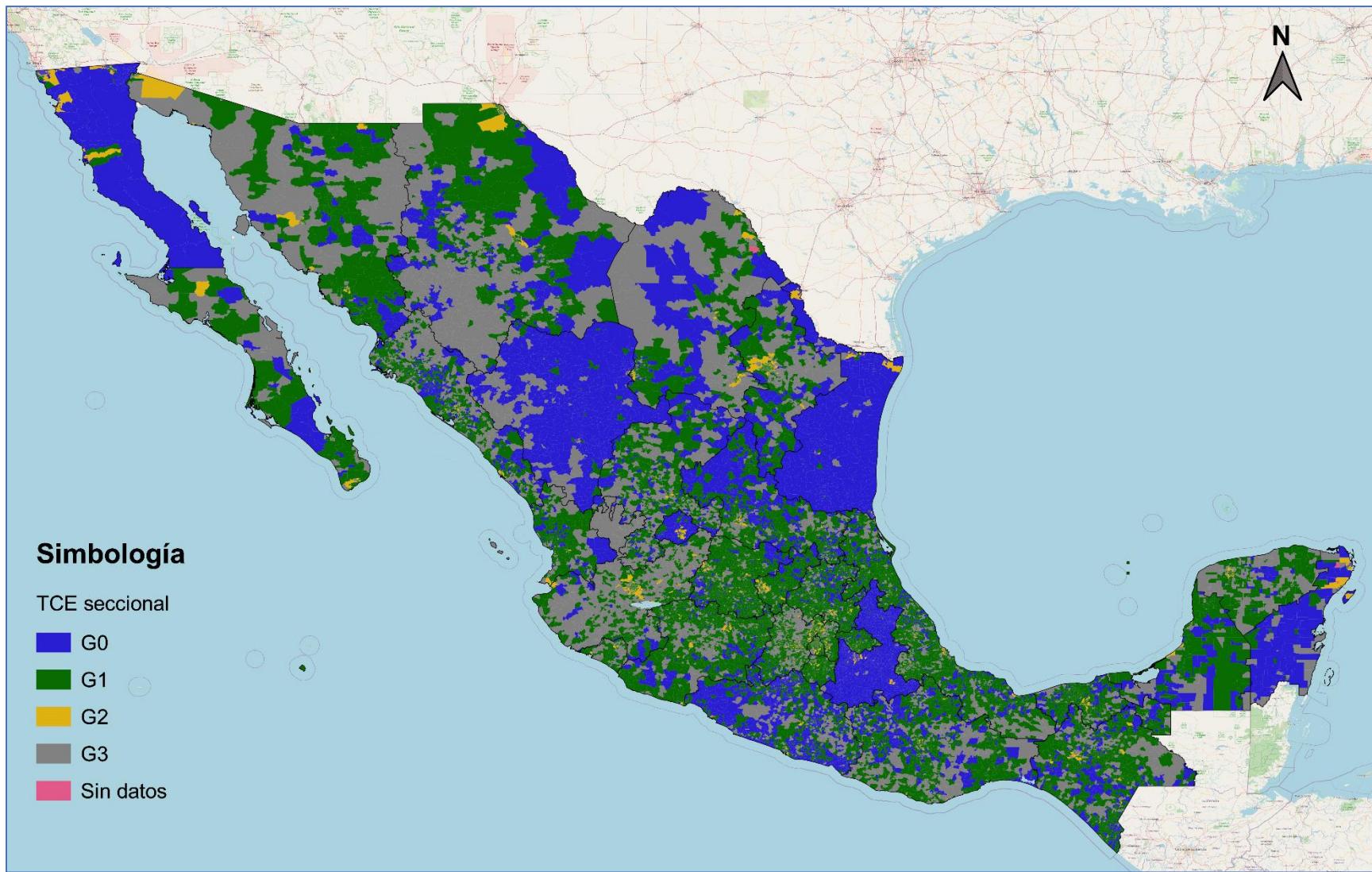


Figura 8. Mapa nacional de secciones electorales con su respectiva clasificación en la TCE seccional

Fuente: Elaboración propia con QGIS

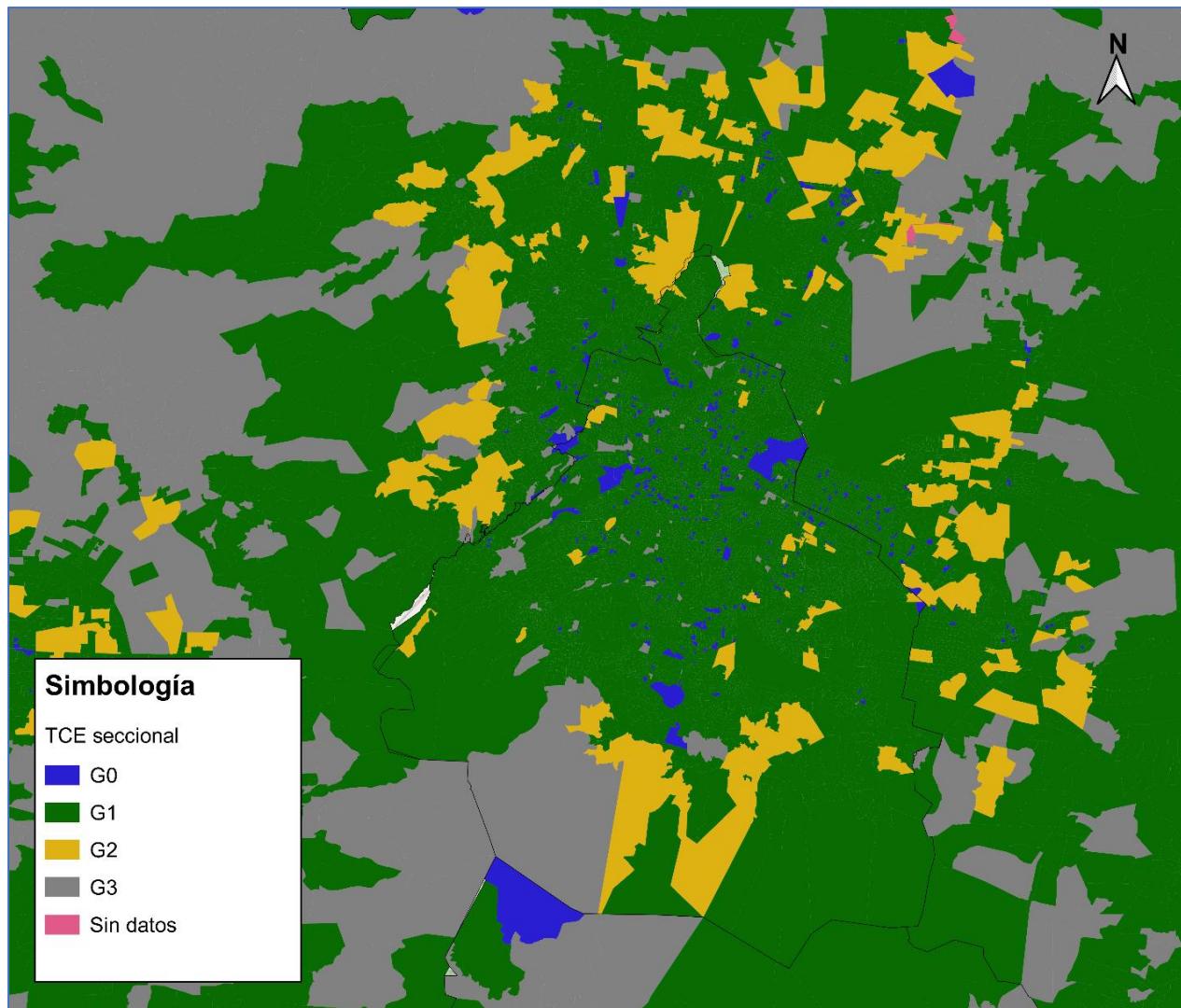


Figura 9. Mapa de las secciones electorales de la zona metropolitana de la Ciudad de México con su respectiva clasificación en la TCE seccional

Fuente: Elaboración propia con QGIS

Las Figuras 8 y 9 permiten observar que las características, deducidas *a priori*, de los grupos de la TCE seccional son concordantes con las características geográficas y demográficas que se entrevén en el territorio mexicano. Por ejemplo, se observa que las secciones electorales ubicadas en las principales zonas urbanas de México se encuentran clasificadas dentro de los grupos G1 y G2 (municipios coloreados en color verde y amarillo), lo cual concuerda con las características identificadas previamente; en particular, las secciones G2 muestran una coincidencia con las zonas sujetas a procesos de expulsión o atracción poblacional.¹²⁶ Asimismo, se observa que las secciones ubicadas en zonas de poca densidad poblacional (por ejemplo, en las zonas montañosas del occidente, norte y sur de México) son clasificadas dentro de los grupos G0 y G3.

Por último, se advierte que en los archivos adjuntos se proporciona el listado completo de las secciones electorales con su respectiva clasificación en la TCE seccional.¹²⁷ Asimismo, en el Anexo II se proporciona el pseudocódigo de Python con el desarrollo del algoritmo de K-medias y del análisis de coeficientes de silueta.

6.2 Estadísticas de los grupos de la TCE seccional

En esta sección se analizan y comparan —mediante gráficas de violín— las distribuciones y estadísticas de tendencia central de los seis indicadores base analizados a nivel seccional. Para la estimación de las estadísticas de tendencia central se agruparon las secciones electorales en función de su clasificación en la TCE seccional. Esta estrategia permite identificar los rasgos característicos de cada grupo de la tipología y a su vez permite identificar y clasificar aquellas secciones electorales con un grado de complejidad electoral atípico.

Con el propósito de ayudar a la comprensión del panel de diagramas de violín —presentado en el Gráfico 22—, se analiza en primera instancia el caso particular del diagrama de violín correspondiente al ICE principal (ver Gráfico 21).

¹²⁶ Esto también se observó previamente en la TCE municipal, ya que los municipios G4 (colindantes al área metropolitana de la Ciudad de México) también muestran un mayor dinamismo demográfico.

¹²⁷ Véase archivo [Resultados_Secc_Indic_PCA_Kmeans_Atip_ind6.csv](#) adjunto en el repositorio de GitHub de Álvarez Hernández (2020).

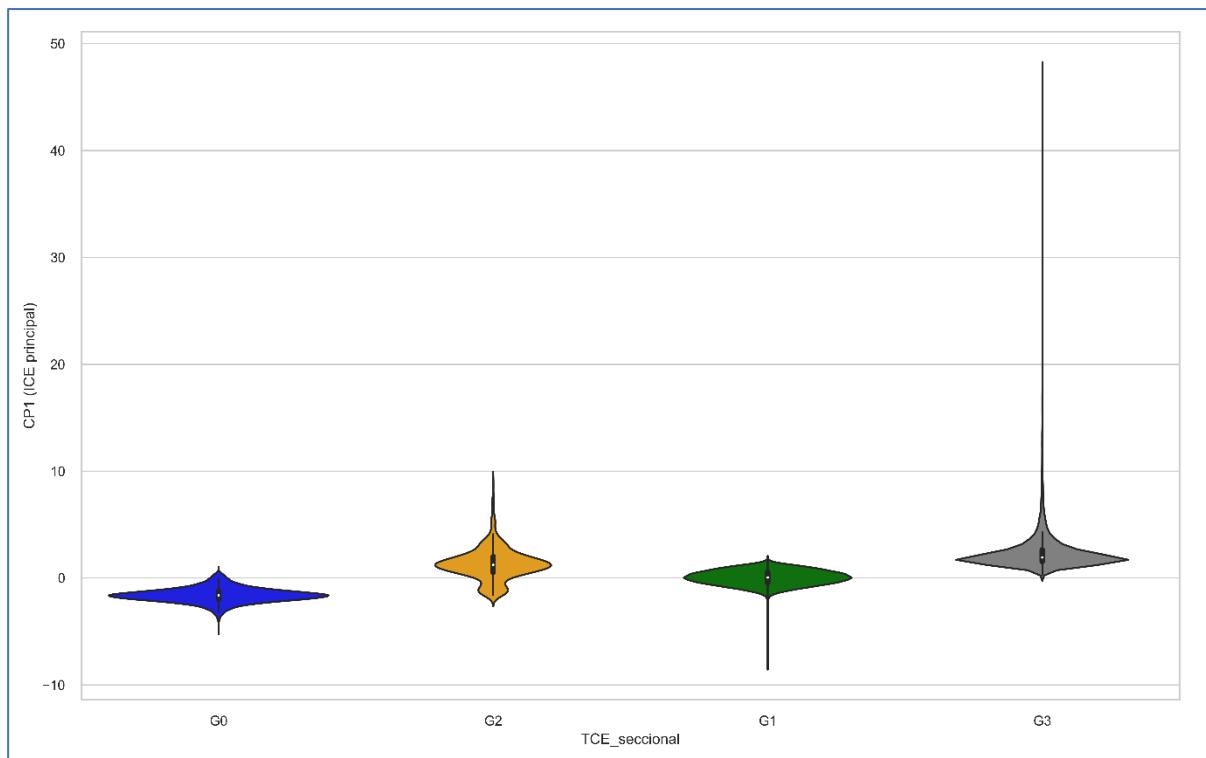


Gráfico 21. *Diagrama de violín del ICE principal (con agrupación en función de la TCE seccional)*

Fuente: Elaboración propia a partir de la librería Python–Seaborn

En el Gráfico 21 se puede observar que los grupos G0, G1, G2 y G3 exhiben estadísticas de tendencia central similares; por ejemplo, los cuatro grupos muestran medias equiparables (indicadas por los puntos blancos dentro de las figuras de violín), y también poseen rangos intercuartílicos (representados por los rectángulos negros) e intervalos de confianza (líneas verticales) similares.

Por otra parte, el Gráfico 21 también muestra, a través de las figuras de violín, las distribuciones de valores correspondientes a cada grupo de la TCE seccional. En particular, se puede observar que los cuatro grupos de la TCE seccional presentan valores atípicos (figuras de violín alargadas), especialmente el grupo G3, ya que este exhibe una distribución de marcado comportamiento leptocúrtico. El origen de estos valores atípicos se puede rastrear a la distribución que muestra el grupo G3 en los indicadores relacionados con la variabilidad demográfica (ver Gráfico 22).

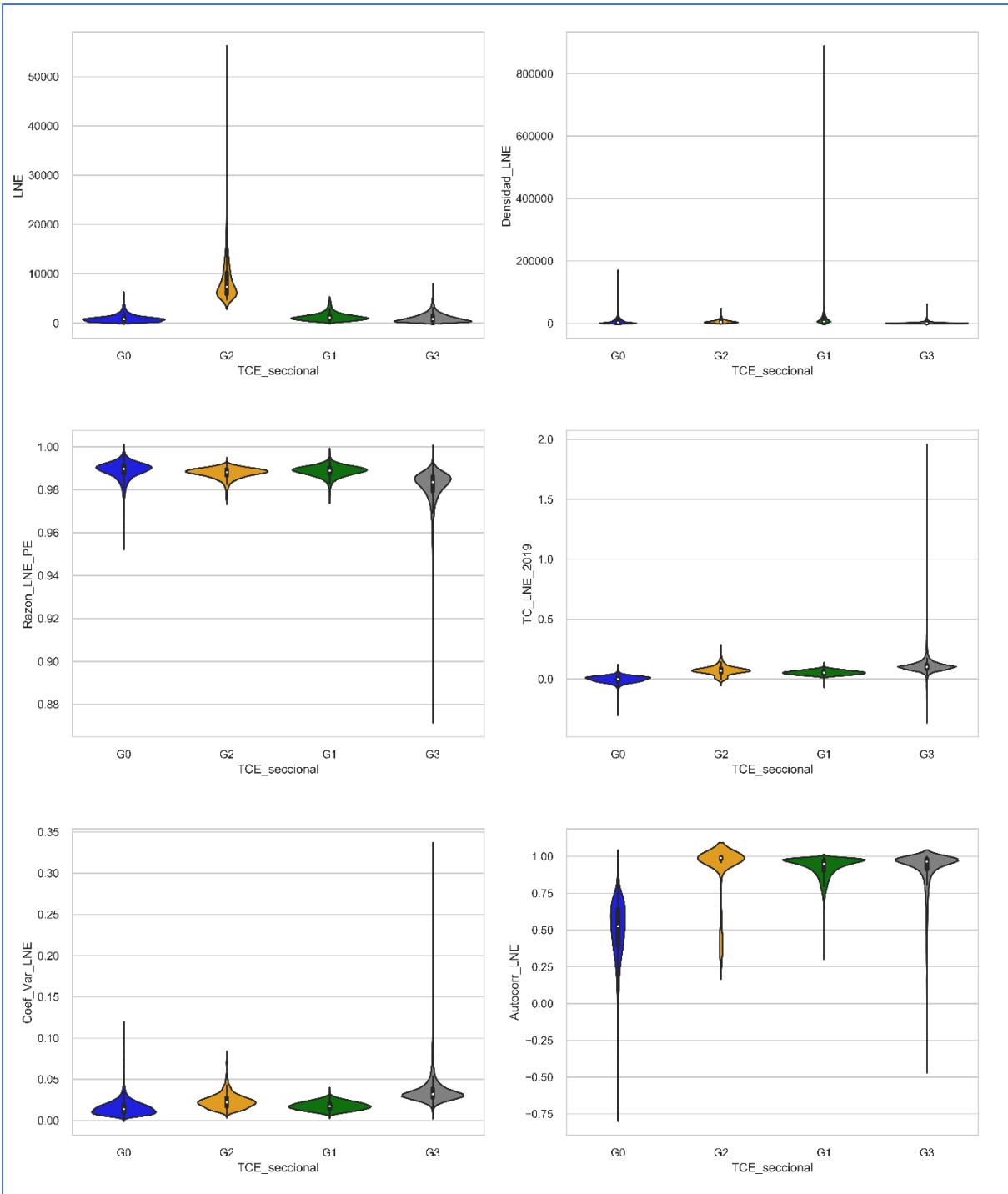


Gráfico 22. Panel de diagramas de violín de los seis indicadores base (con agrupación en función de la TCE seccional)

Fuente: Elaboración propia a partir de la librería Python–Seaborn

Teniendo presente la interpretación de la información contenida en el Gráfico 21, se procede al análisis del panel de diagramas de violín mostrado en el Gráfico 22. Los diagramas de violín muestran que los cuatro grupos de la TCE seccional se caracterizan por presentar estadísticas similares (media, rango intercuartílico e intervalo de confianza) en cada uno de los seis indicadores base, a excepción de los indicadores Magnitud de Electorado (LNE) y Autocorrelación Mensual de la LNE (Autocorr_LNE), ya que en estos indicadores se pueden apreciar diferencias significativas entre las estadísticas de tendencia central de los grupos G0 y G2, y las estadísticas de tendencia central de los grupos restantes.

Por ejemplo, se puede apreciar que las secciones clasificadas en el grupo G2 poseen en promedio un electorado más numeroso, pero al mismo tiempo presentan una densidad poblacional similar a la observada en otros grupos de la tipología. Por otra parte, las secciones pertenecientes al grupo G0 se caracterizan por presentar un bajo nivel de autocorrelación mensual, aunque no presentan características asociadas a la variabilidad demográfica. Y finalmente, algunas de las secciones electorales clasificadas en el grupo G3 (ver la figura alargada de la distribución) muestran una notoria variabilidad demográfica, así como un bajo nivel de actualización entre el PE y la LNE.

Por otra parte, los diagramas de violín del Gráfico 22 muestran que todas las distribuciones de los indicadores base son asimétricas, con diferentes grados y direcciones de sesgo. En particular, las distribuciones observadas en los grupos G2 y G3 exhiben la presencia de colas largas, ya que hay una alta cantidad de secciones electorales fuera de rango (valores atípicos), especialmente en los indicadores relacionados con la magnitud del electorado y con la variabilidad demográfica (LNE, TC_LNE_2019, Coef_Var_LNE y Autocorr_LNE).

Es importante destacar la situación atípica de algunas secciones electorales, ocasionada por la valoración que obtuvieron en los ICE seccionales. Dichas secciones electorales son relevantes de monitorear, ya que presentan características demográficas y geoespaciales que aumentan significativamente su nivel de complejidad electoral. En la siguiente sección se abunda más sobre las características de estas secciones electorales atípicas, lo cual permite elaborar algunas

recomendaciones específicas de actividades institucionales que puede emprender el INE.¹²⁸

6.3 Identificación de casos atípicos a nivel seccional

En forma similar al proceso de identificación de casos atípicos realizado a nivel municipal (ver Capítulo 5), se utiliza la regla del rango intercuartílico, y las agrupaciones de la TCE seccional, para identificar aquellas secciones electorales que presentan una valoración atípica en los primeros tres ICE seccionales.¹²⁹ Los resultados obtenidos a partir del análisis de atipicidad señalan la existencia de tres conjuntos de interés. Las características de estos conjuntos, el mapeo de su ubicación geoespacial (ver Figura 10), y las propuestas derivadas de su análisis, se presentan a continuación.

El primer conjunto está conformado por secciones electorales (718 en total) que presentan una valoración atípica en el ICE principal, y que además se encuentran clasificadas dentro de los grupos G2 y G3. Estas secciones se caracterizan por tener un electorado considerable (algunas significativamente por encima del límite máximo de 3,000 ciudadanos por sección) y una alta variabilidad demográfica (altas tasas de crecimiento y variación de la LNE), factores que tienen un efecto directo en el nivel de actualización del PE y la LNE (valoración baja en el indicador Razón LNE_PE). Por otra parte, destaca que estas secciones electorales atípicas presentan en promedio una baja densificación del electorado en comparación con otras secciones electorales, lo cual permite inferir que estas secciones atípicas se concentran mayormente en la periferia de las zonas metropolitanas. En consecuencia, para este conjunto de secciones electorales atípicas se sugiere:

1) Monitorear constantemente la variabilidad demográfica de la LNE, ya que es probable que, en el corto plazo, las secciones de este conjunto rebasen (o ya lo hayan

¹²⁸ En los archivos anexos en el repositorio de GitHub de Álvarez Hernández (2020) se identifica con la variable/columna Atipicidad aquellas secciones electorales que muestran una complejidad electoral atípica.

¹²⁹ Se realiza únicamente el análisis de atipicidad en los primeros tres ICE seccionales, ya que estos índices logran explicar aproximadamente el 72% de la varianza mostrada en los indicadores base (ver Capítulo 4).

hecho) el límite demográfico de 3,000 ciudadanos por sección electoral. Es decir, se debe considerar las secciones electorales de este conjunto como candidatas potenciales de los futuros programas de resecciónamiento.

2) Estudiar el contexto geográfico y social de estas secciones electorales para determinar las causas de su rápido crecimiento demográfico. Específicamente, elaborar pronósticos con métodos y modelos idóneos para manejar una alta variabilidad en las series de tiempo de la LNE, con el fin de obtener estimaciones demográficas certeras que permitan establecer límites territoriales propicios para los programas de resecciónamiento.

3) Evaluar las campañas de difusión —relacionadas con el trámite o la renovación de la CPV— que se implementan en las secciones atípicas de este grupo, ya que éstas presentan un mayor nivel de desactualización en el PE y la LNE y, en consecuencia, se hace necesario reforzar las estrategias y campañas de actualización de datos.

El segundo conjunto de secciones electorales atípicas está compuesto por secciones que presentan una valoración atípica en el segundo ICE seccional y que se encuentran clasificadas en los grupos G1 y G2 (837 secciones). Estas secciones se caracterizan por tener un electorado numeroso y altamente densificado, pero presentan poca variabilidad demográfica (tasas nulas de crecimiento demográfico y un bajo coeficiente de variabilidad en la LNE). Esto permite inferir que estas secciones electorales se encuentran ubicadas en zonas completamente urbanizadas y presentan estabilidad demográfica, pero sus actuales límites geográficos son inadecuados, ya que rebasan el límite máximo de 3,000 ciudadanos. Consecuentemente, para este conjunto de secciones electorales atípicas se propone:

1) Considerar estas secciones para un programa de resecciónamiento, ya que se encuentran fuera del límite demográfico de 3,000 ciudadanos por sección electoral. Asimismo, tomar en consideración —durante la planeación de los programas de resecciónamiento— el bajo o nulo crecimiento demográfico de estas secciones para establecer límites territoriales permanentes de largo plazo.

Finalmente, el tercer conjunto de secciones electorales atípicas está conformado por secciones que presentan una valoración atípica en el tercer ICE seccional, y además se encuentran clasificadas en los grupos G0 y G1 (2,778 secciones). Las secciones electorales de este conjunto se caracterizan por tener: un electorado pequeño (la mayoría menor al límite de 3,000 ciudadanos), una alta densidad, altos niveles de actualización del PE y la LNE, y poca variabilidad demográfica.

Estos resultados permiten inferir que las secciones de este conjunto se concentran en el interior de las zonas metropolitanas por la alta densidad de su LNE, lo cual sugiere que la totalidad de su superficie se encuentra urbanizada. Asimismo, estas secciones son estables desde el punto de vista demográfico y presentan niveles de actualización adecuados en el PE y la LNE. Por lo tanto, para este grupo de secciones electorales no se recomienda la implementación de alguna acción institucional particular, ya que son secciones atípicas por su estabilidad en los factores geoespacial, demográfico y estructural.

En relación con las secciones electorales que no tienen una valoración en los ICE seccionales –y por ende, tampoco poseen una clasificación en la TCE seccional— se identifica un patrón en su distribución geoespacial, ya que la mayoría de estas secciones se encuentran localizadas en las zonas rurales de los estados de Chihuahua, Coahuila, Durango, Hidalgo, Michoacán, Nayarit y Sinaloa (ver Figura 10).

La razón principal de esta falta de información se encuentra en la metodología de cálculo de los indicadores base, específicamente, en la estimación del indicador Densidad del Electorado. Para estimar dicho indicador es necesario contar con los datos de amazanamiento para estimar el área urbanizada de la sección; sin embargo, existen secciones electorales que no cuentan con este dato, ya sea porque la base cartográfica del INE no está actualizada en su registro de amazanamiento, o porque

son zonas rurales que, efectivamente, no cuentan con un amanzanamiento discernible.¹³⁰

Se finaliza esta sección señalando que en los archivos adjuntos se hace un desglose completo de las secciones electorales con su respectiva valoración en los ICE seccionales, su agrupación en la TCE seccional y su tipo de atipicidad.¹³¹

¹³⁰ El INE podría implementar una metodología alternativa —que no dependa de la determinación de manzanas— para estimar el área urbanizada de las secciones electorales. Por ejemplo, podría explotarse la posibilidad de usar imágenes satelitales para estimar directamente el área superficial cubierta por los asentamientos humanos.

¹³¹ Véanse los archivos [Resultados_Secc_Indic_PCA_Kmeans_Atip_ind6.csv](#), y [Pyplotly_Proyección3d_Secc_ind6.html](#), adjuntos en el repositorio de GitHub de Álvarez Hernández (2020).

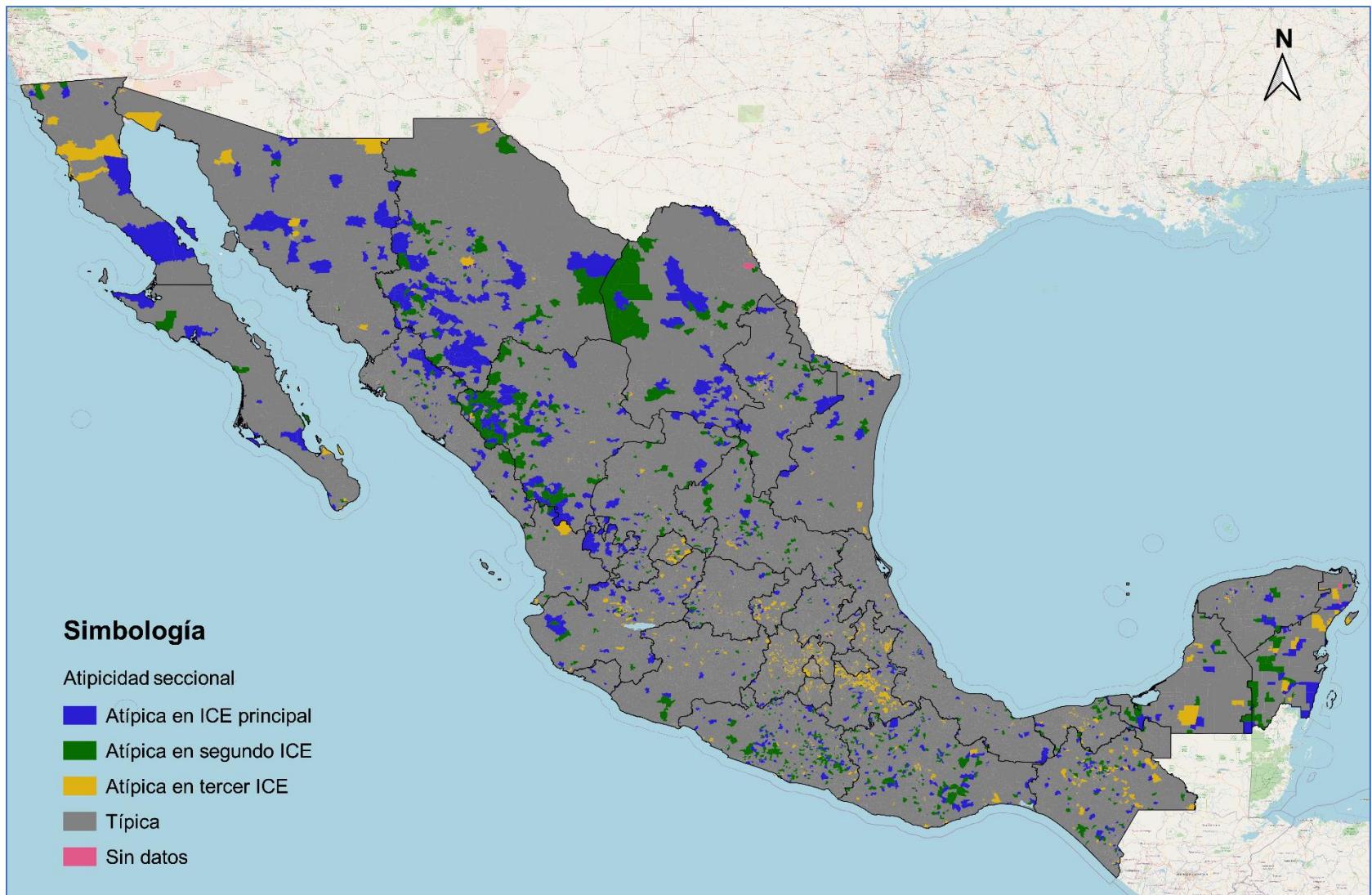


Figura 10. Mapa nacional de secciones electorales con una complejidad electoral atípica

Fuente: Elaboración propia con QGIS

Conclusiones



Conclusiones

En conformidad con el objetivo principal de la presente investigación, se han diseñado un conjunto de instrumentos analíticos que permiten cuantificar la complejidad del sistema geoelectoral mexicano en las escalas municipal y seccional. Con este conjunto de instrumentos se dota al INE de herramientas analíticas avanzadas que le permitirán evaluar y monitorear, de forma más eficaz, el estado del RFE, considerando el contexto geoespacial, demográfico y estructural de cada municipio y sección electoral de México.

Respecto al Capítulo 1, los temas analizados permiten resaltar dos puntos importantes, en los cuales la presente investigación realiza aportaciones. Primero, no obstante que el tema de la complejidad ha sido ampliamente difundido en los campos de la ciencias naturales y sociales, no existen estudios que aborden el tema de la complejidad electoral desde la perspectiva de la presente investigación. El estudio que aquí se presenta puede considerarse pionero en el tema.

Segundo, se puede argumentar que el paradigma actual para la elaboración de indicadores sociodemográficos debe ampliarse y complementarse con metodologías matemáticas y estadísticas alternativas para captar los aspectos geoespaciales, temporales y estructurales que determinan la complejidad inherente de los sistemas sociales. La presente investigación ha realizado esta tarea.

En el Capítulo 2 se construyó un marco metodológico fundamentado en tres elementos:

- 1) el concepto de complejidad electoral,
- 2) el conjunto de indicadores base seleccionado para caracterizar los factores determinantes de la complejidad electoral en las escalas municipal y seccional y
- 3) las técnicas de análisis multivariante y minería de datos, para la construcción de los ICE y la TCE.

El concepto de complejidad electoral se definió como el conjunto de factores geoespaciales, demográficos y estructurales que dificultan u obstaculizan las

labores de Integración, Identificación, Mantenimiento y Actualización del RFE (tareas IIMA). Es decir, la definición de complejidad electoral se focalizó en atender aquellos factores que tienen un impacto directo en las actividades institucionales del INE, dejando de lado otros factores –por ejemplo, la participación electoral o los mecanismos de voto— que pudieran también considerarse parte –directa o indirecta— del tema de la complejidad electoral.¹³²

El conjunto de indicadores base fue seleccionado utilizando la delimitación conceptual del problema de estudio –cuantificar la complejidad electoral—, y la información estadística disponible a escala municipal y seccional. Al respecto, es importante señalar que esta investigación aprovechó la información estadística que genera continuamente el INE, ya que varios de los indicadores base utilizados poseen un mejor nivel de actualidad y detalle que otras fuentes de información oficial.¹³³ Los resultados obtenidos al utilizar estos indicadores otorgaron una descripción más fidedigna y precisa de las actuales características sociodemográficas de los municipios y secciones electorales de México.

Con respecto a los métodos de análisis multivariante y minería de datos –utilizados en la elaboración de los ICE y la TCE— el PCA proporcionó un método matemático robusto para manejar la multidimensionalidad implícita en el conjunto de indicadores base tal que se modeló, con un menor número de variables, la información contenida en los indicadores base. Asimismo, el PCA también permitió obtener un conjunto de indicadores sintetizadores (los ICE) para ponderar –en una forma no subjetiva— el impacto de cada uno los indicadores base en el grado de complejidad electoral. De esta manera, cada uno de los ICE a escala municipal y seccional está fundamentado en las Componentes Principales resultantes del análisis efectuado en los conjuntos de indicadores base.

¹³² No obstante, a pesar de la acotación estricta que la investigación está haciendo del concepto de complejidad electoral, es posible considerar a futuro una descripción más amplia que incorpore otros factores (si la disponibilidad de información estadística sobre dichos factores lo permite).

¹³³ Por ejemplo, la información del Censo de Población 2010, o la Encuesta Intercensal 2015, tienen la limitante inherente de la antigüedad de sus datos. A escala seccional hay un problema aún mayor, ya que no hay dependencia gubernamental o privada que genere estadísticas sociodemográficas a dicha escala, a excepción de la información demográfica y cartográfica que recopila el INE.

En los Capítulos 3 y 4, a través de la presentación de los diagramas de varianza explicada, se mostraron las capacidades de síntesis de información de cada una de las Componentes Principales, tanto a escala municipal como a escala seccional. Los resultados de los análisis de varianza explicada permiten concluir que las Componentes Principales, consideradas individualmente, presentan algunas limitaciones para sintetizar adecuadamente la información contenida en los indicadores base; sin embargo, cuando se consideran en conjunto – específicamente, cuando se considera el conjunto total de Componentes Principales— se obtiene un nivel de síntesis de información superior al obtenido en forma individual por cada Componente Principal. En consecuencia, se justifica la selección del conjunto total de Componentes Principales como los correspondientes ICE.

Al respecto, es importante reiterar que los ICE son instrumentos analíticos que resuelven la problemática de cuantificar la complejidad del sistema geoelectoral mexicano, ya que éstos proporcionan una valoración numérica, y ponderada, de todos los factores que determinan la complejidad electoral presente en los municipios y secciones electorales de México. Además, la valoración que proporcionaron los ICE municipales y seccionales permitió elaborar comparaciones de magnitud y orden entre las diferentes unidades territoriales del sistema geoelectoral; es decir, se obtuvo un *ranking* de complejidad electoral a escala municipal y seccional (primordialmente, a partir del ICE principal). Estas aportaciones de la investigación contribuyen a solventar la actual falta de indicadores que presenten una imagen actual de la evolución demográfica y estructural de los municipios y secciones electorales, ya que los ICE proporcionan información no muestral, multifactorial y actualizada.

A través de gráficas y animaciones se analizó, en los Capítulos 3 y 4, la estructura bidimensional y tridimensional de las proyecciones de los municipios y secciones electorales en el espacio de transformación construido a partir de los ICE. Este análisis arrojó un resultado importante: se identificaron patrones de agrupamiento. Se comprobó la existencia de grupos particulares de municipios y secciones electorales con características geoespaciales, demográficas y

estructurales similares y, por ende, con similitudes en sus grados de complejidad electoral. Así, las estructuras observadas en las proyecciones justifican la elaboración de la TCE a escala municipal y seccional, ya que, para establecer mejores comparaciones de complejidad entre las diversas unidades territoriales del sistema geoelectoral, es necesario contar con una herramienta analítica que complemente el *ranking* de complejidad electoral obtenido con los ICE.

Con respecto al análisis e interpretación de la composición de los ICE municipales y seccionales, se identificó que los índices asignan diferentes ponderaciones a los indicadores base, y, por lo tanto, deben interpretarse separadamente conforme a sus composiciones específicas.¹³⁴

En los Capítulos 5 y 6 se diseñó la TCE con un análisis de coeficientes de silueta, el cual permitió determinar el número óptimo de grupos a considerar en las tipologías, y también permitió evaluar la pertinencia de usar el algoritmo de agrupamiento K-medias. El análisis de coeficientes de silueta nos permite concluir que la configuración óptima de la TCE municipal es de 5 grupos, mientras que en el caso de la TCE seccional, la configuración óptima es de 4 grupos. Es decir, todos los municipios son clasificados en alguno de los 5 grupos que conforman la TCE municipal, mientras que, en el caso de las secciones electorales, éstas son clasificadas en alguno de los 4 grupos de la TCE seccional.

Con base en esas configuraciones de grupos, se implementó el algoritmo K-medias en las proyecciones de los municipios y secciones electorales en el espacio de los ICE. El resultado final es la clasificación de todas las unidades territoriales básicas que conforman el sistema geoelectoral mexicano.

Es importante señalar que los resultados obtenidos con la TCE deben interpretarse como un ejercicio de clasificación, no como un ejercicio de comparaciones de magnitud u orden (esta tarea corresponde a los ICE municipales y seccionales). Es decir, la TCE no permite establecer comparaciones cardinales de

¹³⁴ Es conveniente recordar que los ICE son instrumentos multifactoriales, en el sentido que sintetizan la dinámica de un diverso grupo de factores e indicadores y, por lo tanto, para interpretar adecuadamente la valoración que asignan dichos índices a cada municipio y sección electoral, es necesario analizar cuidadosamente su composición respecto a los indicadores base originales.

la complejidad electoral, únicamente permite analizar el grado de similitud que presentan los municipios y las secciones electorales con relación a su nivel de complejidad electoral y, por consiguiente, la TCE es una herramienta que complementa los resultados del *ranking* de complejidad electoral formado a partir de los ICE.

En relación con la identificación de los casos atípicos de complejidad electoral, es importante mencionar que ésta es una aportación significativa de la presente investigación, ya que puede influir positivamente en el mejoramiento de las actividades institucionales del INE. El análisis de los casos atípicos (a escala municipal y seccional) permitió encontrar clústeres específicos de municipios y secciones electorales que presentan características anómalas respecto a la evolución de su LNE, o respecto a sus características geoespaciales.¹³⁵

Para finalizar, se ofrecen algunas recomendaciones y se señalan algunas limitaciones relacionadas con la operatividad y construcción de los ICE y la TCE; asimismo, se plantean algunas perspectivas sobre la expansión de dichos instrumentos para analizar otros factores y sistemas sociales complejos.

La metodología de cálculo de los ICE y la TCE se fundamentó primordialmente en el uso del PCA, y en la implementación del método de agrupamiento K-medias. Dicha elección de métodos de análisis estuvo supeditada en parte al volumen de datos examinado, y a las capacidades de cómputo disponibles, lo cual ocasionó que no se pudieran aprovechar otros métodos alternativos –y potencialmente más eficientes– de proyección y agrupamiento de datos. Por ejemplo, se experimentó infructuosamente con métodos de agrupamiento de tipo jerárquico, ya que la capacidad de cómputo disponible no fue suficiente para implementar el cálculo de la matriz de distancias. Asimismo, se experimentó sin éxito con diferentes algoritmos de agrupamiento basados en modelos (Gaussian Mixture, DBSCAN).

¹³⁵ En la presente investigación se analizaron las características de estos clústeres, y se elaboraron algunas recomendaciones sobre monitoreo y revisión de las acciones institucionales relacionadas con los servicios que el INE presta en dichas unidades territoriales.

Por consiguiente, es materia de investigación futura el utilizar otras tecnologías de análisis y procesamiento de datos que permitan remontar las limitantes que se encontraron en la presente investigación. Por ejemplo, considerando que se analiza un volumen significativo de municipios y secciones electorales, se podrían ampliar las capacidades de procesamiento de datos de los métodos de agrupamiento utilizando tecnologías de cómputo distribuido. Asimismo, también se podría mejorar la eficiencia de los ICE y la TCE al automatizar su elaboración y actualización, lo cual implicaría adoptar técnicas de *Big Data* para extraer, procesar y transformar la información procedente de los procesos de identificación y actualización del RFE.

Otra línea de investigación futura sería ampliar la utilización de los ICE y la TCE para incluir otro tipo de indicadores sociodemográficos, ya que es plausible ampliar el concepto de complejidad electoral para incorporar otros factores que potencialmente influyen en la complejidad del sistema geoelectoral. Por ejemplo, se podrían incluir indicadores relacionados con la estructura demográfica, los flujos de migración (interna o externa), la accesibilidad geográfica, el grado de adopción de las Tecnologías de la Información y la Comunicación, el nivel de participación electoral, o el comportamiento histórico de las preferencias de votación.

Sin embargo, la incorporación de factores e indicadores adicionales en el concepto de complejidad electoral se encontrará sujeta siempre a la disponibilidad de información, especialmente en un análisis del sistema geoelectoral a escala seccional. Por ende, la presente investigación hace un señalamiento en torno a la necesidad de recabar más información en las escalas básicas del sistema geoelectoral mexicano, tarea que el INE podría realizar eventualmente dado que posee una infraestructura de Módulos de Atención Ciudadana que le permite estar en contacto directo con la ciudadanía, y porque cuenta con procesos institucionales diseñados expresamente para la recopilación y resguardo de datos personales. Así, los instrumentos y resultados generados en esta investigación pueden potenciarse significativamente si se genera más información.

En este contexto, se sugiere realizar actualizaciones periódicas de los ICE y la TCE (por ejemplo, en la etapa inicial de los procesos electorales federales), para monitorear continuamente la evolución de los indicadores base, y los efectos de las estrategias institucionales encaminadas al mejoramiento de los servicios de empadronamiento, votación y actualización de los instrumentos registrales.

La presente investigación finaliza con una observación académica. El tema de la complejidad electoral está lejos de concluir y, por lo tanto, este análisis heurístico de la complejidad del sistema geoelectoral mexicano podría considerarse como un punto de partida para futuros estudios sobre complejidad en los sistemas sociales.

Bibliografía

Abdi, Hervé y Williams, Lynne J. (2010). Principal Component Analysis. *WIREs Comp Stat*, 2, 433-459. <https://doi.org/10.1002/wics.101>

Álvarez Hernández, M.D. (11 de octubre de 2020). *Proyecto-MCDI_Complejidad-sistema-geo-electoral*. GitHub. https://github.com/MDAlvarezH/Proyecto-MCDI_Complejidad-sistema-geo-electoral

Álvarez Hernández, M. David y Álvarez Texocotitla, Miguel. (2020). El sistema geoelectoral mexicano desde la perspectiva de los sistemas complejos. *Apuntes Electorales*, 19 (63), 213-258.

Barbato, G., Barini, E.M., Genta, G. y Levi, R. (2011). Features and performance of some outlier detection methods. *Journal of Applied Statistics*, 38 (10), 2133–2149. <https://doi.org/10.1080/02664763.2010.545119>

Becerra, Gastón. (2020). La Teoría de los Sistemas Complejos y la Teoría de los Sistemas Sociales en las controversias de la complejidad. *Convergencia*, 27, 1-23. <https://doi.org/10.29101/crcs.v27i83.12148>

Bennett, Charles H. (1988). Logical Depth and Physical Complexity: 227-257 en: *The Universal Turing Machine – a Half-Century Survey*, London: Oxford University Press.

Byrne, David. (2001). *Complexity theory and the social sciences: an introduction*. London: Routledge.

Coefficient of variation. (29 de noviembre de 2020). En Wikipedia. https://en.wikipedia.org/w/index.php?title=Coefficient_of_variation&oldid=991248694

Consejo Nacional de Población. (2002). *Índices de marginación a nivel localidad*. México: Autor.

Cooley, Alexander y Snyder, Jack. (2015). *Ranking the World*. New York: Cambridge University Press.

Dean, Jared (2014). *Big data, data mining, and machine learning: value creation for business leaders and practitioners*. New Jersey: Wiley.

Érdi, Peter (2008). *Complexity Explained*. Berlin: Springer.

Everitt, Brian S., Landau, Sabine, Leese, Morven y Stahl, Daniel. (2011). *Cluster Analysis*. London: Wiley.

Faliszewski, Piotr, Hemaspaandra, Edith, Hemaspaandra, Lane A. y Rothe, Joerg. (2006). A Richer Understanding of the Complexity of Election Systems. Recuperado de: <https://arxiv.org/abs/cs/0609112>

Fernández-Martínez, M., Vicca, Sara, Janssens, Ivan A., Carnicer, J., Martín-Vide, J. y Peñuelas, J. (2018). The consecutive disparity index, D: a measure of temporal variability in ecological studies. *Ecosphere* 9 (12), 1-13. <https://doi.org/10.1002/ecs2.2527>

Flores Valdés, Jorge y Martínez Mekler, Gustavo. (Comp.). (2011). *Encuentros con la complejidad*. México: XXI Siglo Veintiuno Editores.

Gell-Mann, Murray y Lloyd, Seth. (1996). Information Measures, Effective Complexity, and Total Information. *Complexity*, 2 (1), 44-52. [https://doi.org/10.1002/\(SICI\)1099-0526\(199609/10\)2:1<44::AID-CPLX10>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1099-0526(199609/10)2:1<44::AID-CPLX10>3.0.CO;2-X)

Gniazdowski, Zenon. (2017). New Interpretation of Principal Components Analysis. *Zeszyty Naukowe WWSI*, 16 (11), 43-65. <https://doi.org/10.26348/znwwsi.16.43>

Goldreich, Oded. (2008). *Computational Complexity: A Conceptual Perspective*, New York: Cambridge University Press.

Hastie, Trevor y Stuetzle, Werner. (1989). Principal Curves. *Journal of the American Statistical Association*, 84 (406), 502-516. <https://doi.org/10.1080/01621459.1989.10478797>

Haynes, Philip (2018). *Social Synthesis: Finding Dynamic Patterns in Complex Social Systems*. New York: Routledge.

Heath, Joel P. y Borowski, Peter. (2013). Quantifying Proportional Variability. *PLoS ONE*, 8 (12): e84074. <https://doi.org/10.1371/journal.pone.0084074>

Hernández-Sampieri, Roberto y Mendoza Torres, Christian Paulina. (2018). *Metodología de la investigación*. México: McGraw Hill Education.

Hintze, J. L. y Nelson, R.D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, 52(2), 181–184.

Horgan, John. (2015). *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age*. New York: Addison-Wesley.

Instituto de Investigaciones Jurídicas. (1984). *Diccionario Jurídico Mexicano: Tomo VI L-O*, (1er ed.). México: Universidad Nacional Autónoma de México.

Instituto Federal Electoral. (1995). *Tipología de secciones*, México: Autor.

Instituto Nacional Electoral. (2017). *Tipología de los distritos electorales para la aplicación de las actividades de la DERFE*, México: Autor.

Jackson, J.E. (1991). *A User's Guide to Principal Components*. New Jersey: Wiley.

Jackson, Donald A. y Chen, Yong (2004). Robust principal component analysis and outlier detection with ecological data. *Environmetrics*, 15 (2), 129-139.
<https://doi.org/10.1002/env.628>

Jolliffe, I.T. (2002). *Principal Component Analysis*. New York: Springer.

K-medias. (20 de noviembre de 2020). En Wikipedia.
<https://es.wikipedia.org/w/index.php?title=K-medias&oldid=131091731>.

Ladyman, James, Lambert, James y Wiesner, Karoline. (2013). What is a complex system? *Euro Jnl Phil Sci*, 3, 33–67. <https://doi.org/10.1007/s13194-012-0056-8>

Lambalgen, Michiel van. (1989). Algorithmic Information Theory, *Journal of Symbolic Logic*, 54 (4), 1389-1400. <https://doi.org/10.1017/S0022481200041153>

Ley General de Instituciones y Procedimientos Electorales. (2020). Recuperado de:
http://www.diputados.gob.mx/LeyesBiblio/pdf/LGIEPE_130420.pdf

Lloyd, Seth (2001). Measures of Complexity: A Nonexhaustive List, *IEEE Control Systems Magazine*, 21 (4), 7-8. <https://doi.org/10.1109/MCS.2001.939938>

Lloyd, Seth y Pagels, Heinz. (1988). Complexity as Thermodynamic Depth. *Annals of Physics*, 188 (1), 186-213. [https://doi.org/10.1016/0003-4916\(88\)90094-2](https://doi.org/10.1016/0003-4916(88)90094-2)

Maggino, Filomena (2017). *Complexity in Society: From Indicators Construction to their Synthesis*. Switzerland: Springer.

Mandelbrot, B. (1967). How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science*, 156 (3775), 636–638. <https://www.doi.org/10.1126/science.156.3775.636>

Marois, Guillaume, Muttarak, Raya y Scherbov, Sergei. (2020). Assessing the potential impact of COVID-19 on life expectancy. *PLoS ONE*, 15 (9), 1-12. <https://doi.org/10.1371/journal.pone.0238678>

Mitchell, Melanie. (2009). *Complexity: A Guided Tour*, London: Oxford University Press.

Northrop, Robert B. (2011). *Introduction to Complexity and Complex Systems*, New York: CRC Press.

Oltvai, Zoltán N. y Barabási, Albert-László. (2002). Life's Complexity Pyramid. *Science*, 298 (5594), 763-764. <https://www.doi.org/10.1126/science.1078563>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel V., Thirion, B., Grisel, O., ... Blondel, M. (2011). Scikit-learn: Machine Learning in Python. *JMLR*, 12, 2825-2830. Recuperado de: https://scikit-learn.org/stable/auto_examples/index.html

Phelan, Steven. (2001). What is complexity science, really? *Emergence*, 3 (1), 120-136. https://doi.org/10.1207/S15327000EM0301_08

Rissanen, Jorma. (1986). Stochastic Complexity and Modeling. *The Annals of Statistics*, 14 (3), 1080-1100. <https://www.doi.org/10.1214/aos/1176350051>.

Rousseeuw, Peter J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20 (1), 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)

Sánchez Ron, José Manuel. (2011). *El poder de la ciencia: Historia social, política y económica de la ciencia (siglos XIX y XX)*. Madrid: Crítica.

Sethna, James P. (2011). *Statistical Mechanics: Entropy, Order Parameters, and Complexity*. London: Oxford University Press.

Shin, J., Price, M.H., Wolpert, D.H., Shimao, H., Tracey, B. y Kohler, T.A. (2020). Scale and information-processing thresholds in Holocene social evolution. *Nature Communications*, 11 (2394), 1-8. <https://doi.org/10.1038/s41467-020-16035-9>

Simon, Herbert A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106 (6), 467-482.

Sim, C. H., Gan, F. y Chang, T.C. (2005) Outlier Labeling with Boxplot Procedures. *Journal of the American Statistical Association*, 100 (470), 642-652. <https://doi.org/10.1198/016214504000001466>

Waldrop, Mitchel. (1992). *Complexity. The Emerging Science at the Edge of Order and Chaos*. Estados Unidos: Simon & Schuster.

Welti Chanes, Carlos. (2019). Calidad del Padrón Electoral y elecciones confiables. *Papeles de Población*, 100, 13-52. <https://doi.org/10.22185/24487147.2019.100.12>

Wolfram, Stephen. (2002). *A New Kind of Science*. Estados Unidos: Wolfram Media.

ANEXOS



ANEXO I: Pseudocódigo de Python para implementación del Análisis de Componentes Principales

El siguiente pseudocódigo corresponde al análisis efectuado a escala municipal.

```
#Librerias
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from matplotlib import pyplot
from mpl_toolkits.mplot3d import Axes3D
from matplotlib.lines import Line2D

#Se seleccionan 7 indicadores
ind7 = ['NUM_SECC_RUR','LNE', 'Densidad_LNE', 'Razon_LNE_PE', 'TC_LNE_2019', 'Coef_Var_LNE', 'Autocorr_LNE']

#únicamente los valores numéricos
x = datos_mun_limpio_f.loc[:, ind7].values

#Se estandariza los valores de ind7 (normalización)
x = StandardScaler().fit_transform(x)
print(x.shape)

#cálculo de PCA (con las 7 componentes), y obtención de los valores CP para cada municipio
pca_7cp = PCA(n_components=7)
pca_ind7 = pca_7cp.fit_transform(x)

#razones de varianza explicada por cada CP
print('Razón de varianza explicada por cada CP (n_components=7): %s'
      % str(pca_7cp.explained_variance_ratio_))

#dataframe con los valores de los 7 CP por municipio
df_pca_ind7 = pd.DataFrame(data = pca_ind7
                            , columns = ['CP1', 'CP2', 'CP3','CP4', 'CP5', 'CP6','CP7'])

#obtención de los pesos por cada variable
pca_pesos_ind7 = pca_7cp.components_
print(pca_pesos_ind7)

#transformación a un dataframe:
df_pca_pesos_ind7 = pd.DataFrame(pca_pesos_ind7, columns=ind7)
principal_components_pesos = pd.DataFrame(['CP1', 'CP2', 'CP3','CP4', 'CP5', 'CP6','CP7'])
df_pca_pesos_ind7.insert(0, 'Componentes Principales', principal_components_pesos)

#se guarda el último dataframe en un csv
df_pca_pesos_ind7.to_csv(r'Resultados_Mun_PCA_pesos_ind7.csv', index = False)

#cálculo de la matriz de covarianza y sus correspondientes eigenvalores y eigenvectores
cov_mat = np.cov(x.T)
```

```

eigen_vals, eigen_vecs = np.linalg.eig(cov_mat)

#varianza explicada individual y cumulativa
tot = sum(eigen_vals)
var_exp = [(i / tot) for i in sorted(eigen_vals, reverse=True)]
cum_var_exp = np.cumsum(var_exp)

#gráfica de porcentajes de varianza explicada
plt.figure(figsize=(10,6), dpi= 200)
plt.bar(range(1,8), var_exp, alpha=0.5,
        align='center', label='varianza explicada individual')
plt.step(range(1,8), cum_var_exp, where='mid',
         label='varianza explicada acumulada')
plt.ylabel('Razón de varianza explicada')
plt.xlabel('Índice Componentes Principales')
plt.legend(loc='best')
plt.title('Componentes Principales (7 indicadores base)')
plt.savefig('Pyplot_PCA_variance_ind7.png',dpi=600)

#Proyección en espacio CP1-CP2
plt.figure(figsize=(16, 10), dpi= 200, facecolor='w', edgecolor='k')
sns.scatterplot(x="CP1", y="CP2",
                 data=df_pca_results_ind7)
# Set x-axis label
plt.xlabel('CP1 (ICE principal)')
# Set y-axis label
plt.ylabel('CP2 (segundo ICE)')
plt.title('Proyección en CP1-CP2 (7 indicadores)')

plt.savefig('Pyplot_PCA_projection_ind7.png',dpi=600)

```

ANEXO II: Pseudocódigo de Python para implementación de K-medias y Análisis de Coeficientes de Silueta

El siguiente pseudocódigo corresponde al análisis efectuado a escala municipal.

```
#Librerías
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_samples, silhouette_score
import matplotlib.cm as cm

#Se forma un np-array con las columnas_pca para la sección de métodos de agrupamiento
columnas_7pca = ['CP1', 'CP2', 'CP3', 'CP4', 'CP5', 'CP6', 'CP7']
x_pca = df_pca_results_ind7.loc[:, columnas_7pca].values

#Datos en el espacio de las componentes principales
X=x_pca

#número de clusters/grupos que serán analizados
range_n_clusters = [2, 3, 4, 5, 6, 7]

for n_clusters in range_n_clusters:
    # Create a subplot with 1 row and 2 columns
    fig, (ax1, ax2) = plt.subplots(1, 2)
    fig.set_size_inches(18, 7)

    # The 1st subplot is the silhouette plot
    # The silhouette coefficient can range from -1, 1
    ax1.set_xlim([-1, 1])
    # The (n_clusters+1)*10 is for inserting blank space between
    # Silhouette plots of individual clusters, to demarcate them clearly.
    ax1.set_ylim([0, len(X) + (n_clusters + 1) * 10])

    # Initialize the clusterer with n_clusters value and a random
    # generator
    # seed of 0 for reproducibility.
    clusterer = KMeans(n_clusters=n_clusters, random_state=0)
    cluster_labels = clusterer.fit_predict(X)

    # The silhouette_score gives the average value for all the samples.
    # This gives a perspective into the density and separation of the
    # formed clusters
    silhouette_avg = silhouette_score(X, cluster_labels)
    print("For n_clusters =", n_clusters,
          "The average silhouette_score is :", silhouette_avg)

    # Compute the silhouette scores for each sample
    sample_silhouette_values = silhouette_samples(X, cluster_labels)

    y_lower = 10

    for i in range(n_clusters):
        # Aggregate the silhouette scores for samples belonging to
        # cluster i, and sort them
```

```

    ith_cluster_silhouette_values = \
        sample_silhouette_values[cluster_labels == i]

    ith_cluster_silhouette_values.sort()

    size_cluster_i = ith_cluster_silhouette_values.shape[0]
    y_upper = y_lower + size_cluster_i

    color = cm.nipy_spectral(float(i) / n_clusters)
    ax1.fill_betweenx(np.arange(y_lower, y_upper),
                      0, ith_cluster_silhouette_values,
                      facecolor=color, edgecolor=color, alpha=0.7)

    # Label the silhouette plots with their cluster numbers at the
    # middle
    ax1.text(-0.05, y_lower + 0.5 * size_cluster_i, str(i))

    # Compute the new y_lower for next plot
    y_lower = y_upper + 10 # 10 for the 0 samples

    ax1.set_title("Diagrama de silueta por cada grupo.")
    ax1.set_xlabel("Coeficiente de silueta")
    ax1.set_ylabel("Grupos")

    # The vertical line for average silhouette score of all the values
    ax1.axvline(x=silhouette_avg, color="red", linestyle="--")

    ax1.set_yticks([]) # Clear the yaxis labels / ticks

    # 2nd Plot showing the actual clusters formed
    colors = cm.nipy_spectral(cluster_labels.astype(float) / n_clusters)
    ax2.scatter(X[:, 0], X[:, 1], marker='.', s=30, lw=0, alpha=0.7,
                c=colors, edgecolor='k')

    # Labeling the clusters
    centers = clusterer.cluster_centers_
    # Draw white circles at cluster centers
    ax2.scatter(centers[:, 0], centers[:, 1], marker='o',
                c="white", alpha=1, s=200, edgecolor='k')

    for i, c in enumerate(centers):
        ax2.scatter(c[0], c[1], marker='$%d$' % i, alpha=1,
                    s=50, edgecolor='k')

    ax2.set_title("TCE en el espacio CP1-CP2")
    ax2.set_xlabel("CP1 (ICE principal)")
    ax2.set_ylabel("CP2 (segundo ICE)")

    plt.suptitle(("Análisis de silueta con K-medias (7 indicadores base). "
                 "Número de grupos = %d" % n_clusters),
                 fontsize=14, fontweight='bold')
    plt.savefig('Pyplot_silhouette_score_ind7_Kmeans_'+str(n_clusters)+'.png', dpi=600)

#se obtienen las etiquetas predichas por K-medias para la configuración
con el mayor valor promedio de silueta(n=5)

```

```

num_clusters_ind7 = 5
clusterer_K = KMeans(n_clusters=num_clusters_ind7, random_state=0)
clusterer_K_labels = clusterer_K.fit_predict(X)

#se obtienen las etiquetas predichas por K-medias
y_pred_kmeans = clusterer_K_labels

#se guarda el np-array en un dataframe
y_pred_kmeans = pd.DataFrame(y_pred_kmeans, columns=['TCE_municipal'])

#se transforman las etiquetas numéricas a categorías
etiquetas_grupos = {0:'G0', 1:'G1', 2:'G2', 3:'G3', 4:'G4', 5:'G5'}
y_pred_kmeans['TCE_municipal']=y_pred_kmeans['TCE_municipal'].apply(lambda
a x: etiquetas_grupos[x])

#se concatenan los resultados de PCA con las etiquetas de K-medias
df_pca_results_ind7_kmeans = pd.concat([df_pca_results_ind7, y_pred_kmeans], axis = 1)

#se guardan los resultados en un csv
df_pca_results_ind7_kmeans.to_csv(r'Resultados_Mun_PCA_Kmeans_ind7.csv',
index = None)

```