

# La complejidad del sistema geoelectoral mexicano a nivel municipal

*The Complexity of the Mexican Geoelectoral System at the Municipal Level*

M. David Álvarez Hernández\*  
Shaní Eneida Álvarez Hernández\*\*  
Miguel Álvarez Texcotitla\*\*\*

## Resumen

Para determinar la complejidad del sistema geoelectoral mexicano, la presente investigación analiza a nivel municipal los factores que influyen en los procesos de integración, mantenimiento y actualización del Registro Federal de Electores (RFE). El análisis parte de una definición heurística del concepto de *complejidad electoral* para construir un grupo de índices de complejidad electoral que permitirán cuantificar la complejidad de los

**Artículo** recibido el 11 de abril de 2021 y aceptado para su publicación el 30 de noviembre de 2021. La **dictaminación** de este trabajo fue realizada por evaluadores externos al Instituto Electoral del Estado de México.

■ pp. 167-215



\* Maestro en Ciencia de Datos e Información por el INFOTEC-Centro de Investigación e Innovación en TIC. Jefe del Departamento de Soporte Estadístico del Instituto Nacional Electoral (INE). Sus líneas de investigación son ciencia de datos aplicada a fenómenos sociales, movimientos migratorios y cambio demográfico, y progreso tecnológico en los sistemas electorales. Correo electrónico: mdalvarezh@gmail.com

\*\* Maestra en Ciencias de la Computación por la University of York. Adscrita al Centro de Excelencia para Información Estadística de Gobierno, Seguridad Pública, Victimization y Justicia. Líneas de investigación: inteligencia artificial, geoestadística, modelación matemática. Correo electrónico: shanieneida@gmail.com

\*\*\* Doctor en Ciencias Sociales por la Universidad Autónoma Metropolitana (UAM)-Xochimilco. Profesor-investigador de tiempo completo del Departamento de Economía de la UAM-Iztapalapa. Sus líneas de investigación son deuda pública, crisis financiera, desarrollo económico, instituciones políticas y procesos electorales. Correo electrónico: atm@xanum.uam.mx

municipios a partir de información procedente de diversos indicadores estadísticos.

**Palabras clave:** sistemas complejos, geoestadística, indicadores socioeconómicos, análisis de componentes principales.

## Abstract

In order to determine the complexity of the Mexican geo-electoral system, this paper delves into the factors that influence the processes of integration, maintenance and updating of the Federal Register of Voters at a municipal level. Its analysis begins with a heuristic definition of the concept of electoral complexity, from which a group of electoral complexity indices are built, which will then allow the complexity of municipalities to be quantified based on information from various statistical indicators.

**Keywords:** complex systems, geostatistics, socioeconomic indicators, principal component analysis.

## Introducción

Los listados de electores son fundamentales para cualquier sistema geoelectoral,<sup>1</sup> ya que son instrumentos que proporcionan certidumbre y legitimidad a los procesos político-electorales que se desarrollan en un país. La tarea de construir estos instrumentos adquiere diferentes grados de complejidad dependiendo de las características específicas de los sistemas geoelectorales, ya que entran en juego la normatividad electoral, así como las condiciones sociopolíticas, sociodemográficas y geográficas particulares de cada nación.

---

<sup>1</sup> El concepto de *sistema geoelectoral* hace referencia a 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 Texcotitla, 2020). En particular, los municipios 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.

En general, un listado de electores 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: 1) requisitos que deben cumplir los ciudadanos para formar parte del listado, 2) medios de identificación de los integrantes del listado, 3) medios físicos de resguardo de la información del listado, y 4) mecanismos de actualización y consolidación de la información del listado (Welti Chanes, 2019).

En el contexto específico del sistema geoelectoral mexicano, su listado de electores se encuentra representado por la figura del RFE, que es administrada por el Instituto Nacional Electoral (INE), ya que este organismo es la autoridad que elabora y supervisa los procedimientos de empadronamiento (integración), credencialización (identificación), resguardo de las bases de datos (mantenimiento) y consolidación de la información (actualización).<sup>2</sup> Al respecto, es importante señalar que el RFE está conformado por dos instrumentos registrales distintos: el padrón electoral (PE) y la lista nominal de electores (LNE). El PE representa el registro de todos aquellos ciudadanos que en algún momento tramitaron y recibieron una credencial para votar (CPV). En contraste, la LNE representa el registro de todos aquellos ciudadanos que cuentan con una credencial para votar vigente y que por tanto están en capacidad de votar.

La calidad y eficacia de los instrumentos registrales es uno de los principales factores que pueden favorecer o inhibir la participación ciudadana en los ejercicios electorales y en la calidad de una democracia, ya que estos sistemas determinan quiénes tienen derecho a participar en los comicios, además que también determinan en buena parte los procedimientos de emisión y registro del voto. En consecuencia, un listado electoral consolidado favorecerá la participación electoral, ya que brindará confianza a la ciudadanía de que sus derechos político-electorales serán resguardados, además de que habrá certeza y seguridad en la emisión y conteo de los

---

<sup>2</sup> En lo subsiguiente, a este grupo de cuatro tareas se identificará como HMA.

sufragios. En el caso particular de México, el RFE ha contribuido significativamente a otorgar certeza a los procesos político-electorales y a dar legitimidad a los resultados de las elecciones, ya que uno de sus propósitos principales es garantizar el libre acceso y ejercicio de los derechos políticos (Welti Chanes, 2019).

Considerando este contexto, las tareas de IIMA son fundamentales para garantizar la calidad y consistencia del RFE, por consiguiente, resulta primordial contar con métricas que permitan planear, dirigir y evaluar los programas y acciones institucionales enfocados a las tareas de IIMA del RFE, y que además tomen en consideración la diversidad geográfica del país, así como la pluralidad de las características sociodemográficas de la población. Particularmente, resulta crucial contar con indicadores que sintetizen y cuantifiquen los factores que inciden en la complejidad electoral<sup>3</sup> de los municipios, esto con el fin de lograr una asignación diferenciada y eficiente de los recursos institucionales para la organización e implementación de los procesos electorales.

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 de México, no existe actualmente una metodología unificadora que permita su análisis y monitoreo. Asimismo, hay una falta acuciante de estudios que exploren en forma adecuada la información cartográfica, demográfica y electoral que genera continuamente el INE a nivel municipal, sobre todo considerando que dicha información se caracteriza por tener un mejor nivel de actualización respecto a otras fuentes de información oficial. Por consiguiente, el objetivo principal de la presente investigación consiste en elaborar, a partir de un grupo diverso de indicadores geoespaciales, demográficos y estructurales, un conjunto de índices de complejidad electoral (ICE) a escala municipal.

<sup>3</sup> Como se explicará con mayor detalle en la siguiente sección, se 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 de IIMA) del RFE.

Por otro lado, es importante destacar las contribuciones de este trabajo para valorar su relevancia. Una aportación capital es que utiliza una metodología unificadora que permite conocer con precisión y actualidad, la dinámica de los factores demográficos, geoespaciales y estructurales que determinan la complejidad electoral de los municipios. Otra contribución es instrumental, ya que los resultados provenientes de los ICE coadyuvan significativamente a robustecer la calidad del RFE, al proporcionar una forma novedosa y efectiva para analizar y monitorear la evolución de los factores demográficos, geoespaciales y estructurales que afectan las tareas de IIMA que realiza el INE. Asimismo, los ICE también suministran una base cuantitativa para realizar ejercicios de clasificación y estratificación de los municipios en función de sus características, lo cual permite evaluar, a una escala precisa, la situación demográfica de México, ya que los ejercicios de clasificación y estratificación posibilitan el análisis de la variabilidad del comportamiento poblacional de los municipios, facilitando la elaboración de modelos de pronóstico demográfico locales.<sup>4</sup>

Considerando que la metodología utilizada para la construcción de los ICE se fundamenta en las ciencias de la complejidad, la estrategia de investigación adoptada inicia con una breve revisión del estado del arte sobre los sistemas complejos. Posteriormente, se examinan algunas metodologías modernas de cuantificación de la complejidad y su impacto en las ciencias sociales.

En la segunda sección se presenta el marco metodológico de la presente investigación. Se propone una definición heurística de complejidad electoral, se detalla el proceso de selección de los indicadores base que sirven para cuantificar la complejidad electoral y se expone la metodología de elaboración de los ICE y su correspondiente estratificación.

---

<sup>4</sup> En una investigación previa, Álvarez Hernández y Álvarez Texocotitla (2020) presentaron un estudio de la complejidad electoral similar al desarrollado en el presente artículo; sin embargo, ese estudio se realizó con un mayor grado de desagregación geográfica, ya que la información de los factores demográficos, geoespaciales y estructurales se extrajo a nivel de sección electoral, utilizando una metodología distinta para la tipificación de la complejidad electoral.

En la tercera sección se presentan los resultados estadísticos y gráficos relacionados con la construcción de los ICE municipales. Se muestran los resultados del análisis exploratorio de correlación de los indicadores base que conforman los índices, así como los resultados derivados de la implementación del análisis de componentes principales (PCA por *Principal Component Analysis*), técnica que permite transformar y reducir la dimensionalidad del conjunto de indicadores base en un nuevo grupo de indicadores: los componentes principales (CP). Cada CP representará un ICE municipal, por lo que se analizan sus capacidades explicativas y sus composiciones con respecto a los indicadores base. Asimismo, se analizan los resultados de la estratificación municipal de complejidad electoral, la cual se genera a partir de la utilización del método de agrupamiento jerárquico de Ward en el subespacio de los primeros tres componentes principales.

Finalmente, el trabajo presenta las conclusiones y reflexiones que retoman, desde una perspectiva holística, los hallazgos obtenidos en cada una de las fases de esta investigación.

## Consideraciones teóricas sobre la complejidad

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 (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 (Horgan, 2015). Esta diversidad de opiniones ha enriquecido el campo de estudio de los sistemas complejos, pero también ha ocasionado el surgimiento de múltiples 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): la tradición sistémico-cibernética enfo-

cada al estudio de fenómenos de varios dominios desde una perspectiva de sistemas, y las autonombradas ciencias de la complejidad, caracterizadas por el enfoque algorítmico que siguen para analizar el comportamiento de sistemas adaptativos, dinámicos, no equilibrados y no lineales. Ambas corrientes de investigación no tienen un sentido unívoco en torno al concepto de *complejidad*, ni al interior de ellas. Por ejemplo, hay más de 40 definiciones propuestas solamente en el ámbito de las ciencias de la complejidad (Lloyd, 2001, p. 7).

A consecuencia de la falta de consenso sobre el concepto de *complejidad*, Mitchell (2009, p. 13) ha propuesto una definición heurística de *sistema complejo*, sustentada en un conjunto de propiedades que la mayoría de los investigadores ha considerado como inherentes a los sistemas complejos (véanse Ladyman *et al.*, 2013; Mitchell, 2009; Byrne, 2001; y Simon, 1962, para una explicación más detallada):

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.

Por otra parte, la cuestión de cómo cuantificar la complejidad también es relevante, ya que para entender sus implicaciones y características es necesario responder las preguntas siguientes: ¿cómo medir la complejidad?, ¿cómo comparar la complejidad entre un sistema y otro?

Conforme los estudios sobre complejidad han ido avanzando, el número de formas de cuantificar la complejidad también ha crecido paralelamente. Lloyd (2001) clasificó 40 medidas o cuantificadores de complejidad, agrupados en tres cuestiones principales: 1) dificultad de descripción, 2) dificultad de creación y 3) grado de organización.

En la tabla 1 se muestran algunas de las medidas de complejidad clasificadas por Lloyd (2001) y Northrop (2011). Es importante puntualizar que estas medidas abordan aspectos parciales o específicos de los sistemas complejos, por lo que ninguna es considerada actualmente como una medida universal de complejidad.

**Tabla 1. Algunas medidas de complejidad propuestas**

Dificultad de descripción	Dificultad de creación	Grado de organización
Dimensión fractal	Profundidad lógica	Complejidad jerárquica
Entropía	Profundidad termodinámica	Complejidad estocástica
Complejidad algorítmica	Complejidad computacional	Complejidad efectiva
Descripción de longitud mínima		Información mutua

Fuente: Elaboración propia.

Respecto al papel de la complejidad en las ciencias 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 se puede reconocer que los sistemas de indicadores socioeconómicos son la base fundamental para cualquier descripción sólida y sintética de los hechos sociales que permita la formulación de políticas normativas (Cooley y Snyder, 2015). No obstante, hay que tener presente que el proporcionar visiones sintéticas de los fenómenos sociales no implica necesariamente computar indicadores agregados.

La forma tradicional de construcción de estos indicadores ha estado fuertemente influida por el paradigma 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 (Haynes, 2018, pp. 5-33). Lo anterior tiene una implicación. En un sistema social complejo y estructuralmente dinámico, cualquier descripción promedio de los fenómenos es intrínsecamente ineficaz para comprender y controlar las dinámicas de estos sistemas. Es decir, los indicadores socioeconómicos tradicionales tienden a presentar visiones agrega-



das de la realidad y, por lo tanto, son intrínsecamente imprecisos ya que no pueden captar fehacientemente la complejidad de los fenómenos sociales. Esto limita su grado de validez como instrumentos de estudio, o como instrumentos de formulación y evaluación de políticas normativas (Maggino, 2017, pp. vii-viii), ya que la complejidad no puede anularse sin privar a los productos estadísticos finales de su contenido de información y de su utilidad tanto para los científicos sociales como para los hacedores de políticas.

## Marco metodológico

### *Conceptualización de la complejidad electoral*

A pesar de que el concepto de *complejidad* se ha adaptado y esgrimido en diversos campos de las ciencias sociales, el uso de éste en los temas electorales ha sido escaso o con enfoques ceñidos exclusivamente al ámbito de los problemas computacionales que surgen en los procesos de votación (Faliszewski *et al.*, 2006), o la interdependencia entre las características de los sistemas geoelectorales y la complejidad en los sistemas de partidos (Klingemann, 2009). La presente investigación difiere de estos enfoques, ya que analiza un componente específico, aunque de suma importancia, de los sistemas geoelectorales: los registros de electores.

Como se puntualizó anteriormente, el registro de electores de México, el RFE, está conformado por dos listados: el PE y la LNE. Para integrar, identificar, mantener y actualizar estos instrumentos (tareas de IIMA) es necesario que el INE realice constantemente procesos de recolección de datos personales, depuración de bases de datos, implementación de estudios de campo para actualizar la cartografía electoral y elaboración de análisis sobre la dinámica demográfica (Álvarez Hernández y Álvarez Texocotitla, 2020).

Por otra parte, para lograr una adecuada ejecución de las tareas de 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 mexica-

no, 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).

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 la ejecución de las tareas de IIMA del RFE.<sup>5</sup>

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

Considerando la definición de *complejidad electoral* propuesta, surge la problemática de cómo cuantificar el concepto para cada uno de los componentes base del sistema geoelectoral mexicano, específicamente a nivel municipal. Para solucionar este problema se identifican, en primer lugar, los indicadores base que describen la dinámica de los factores determinantes de la complejidad electoral a nivel municipal, ya que éstos serán los insumos de información principales para construir un nuevo conjunto de índices, diseñado expresamente para cuantificar el grado de complejidad electoral

---

<sup>5</sup> En Álvarez Hernández y Álvarez Texocotitla (2020) se realiza un primer planteamiento de esta definición de *complejidad electoral*. Asimismo, en INE (2017) también se plantea una conceptualización sobre la complejidad electoral, sin embargo, tiene un enfoque puramente operativo para propósitos de clasificación de los distritos electorales. Además de los trabajos anteriormente mencionados, los autores de esta investigación no han identificado estudios previos que aborden la cuestión de la complejidad de los sistemas geoelectorales (en México o en otros países) desde una perspectiva sustentada en el análisis conjunto de indicadores geoespaciales, demográficos o estructurales.

observado en los municipios. Asimismo, estos índices servirán como base para la elaboración de una estratificación de complejidad electoral a escala municipal.

La definición heurística de *complejidad electoral* que se propone permite clasificar los indicadores estadísticos que el INE monitorea constantemente para la planeación y realización de las tareas de IIMA del RFE. En particular, se analizan tres grupos de indicadores: geoespaciales, demográficos y estructurales.

Los indicadores geoespaciales describen factores como la magnitud y dispersión geográfica de la población registrada en el RFE. Estos factores son 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, así como el repartimiento de los módulos de atención ciudadana<sup>6</sup>).

Los indicadores demográficos se encuentran asociados a factores de la dinámica demográfica de los municipios, 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, sobre todo considerando que los programas de identificación, actualización y mantenimiento del RFE (por ejemplo, las campañas de credencialización o la identificación de flujos migratorios anómalos) deben acoplarse correctamente a las distintas dinámicas demográficas que se observan en el sistema geoelectoral.

Los indicadores estructurales 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

---

<sup>6</sup> Los módulos de atención ciudadana son las oficinas de atención que el INE pone a disposición de la ciudadanía. En éstas se efectúan los trámites de credencialización, corrección de datos personales y levantamiento de estudios muestrales.

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.

### *Los indicadores estadísticos para cuantificar la complejidad electoral*

Para la conformación y construcción de los ICE, se consideró un conjunto de siete indicadores base que sintetizan las características geoespaciales, demográficas y estructurales esenciales que muestran cada uno de los municipios de México.

- 1) **Magnitud del electorado (LNE).** Este indicador demográfico refleja el número de ciudadanos que residen en los municipios y que están en la posibilidad de ejercer su voto. El INE monitorea constantemente este indicador, ya que se considera que éste incide en el grado de complejidad electoral que exhibe un municipio, debido a que las actividades relacionadas con la integración y la actualización del RFE dependen de la dinámica demográfica para la elaboración de los planes presupuestales y de organización electoral. Por lo tanto:

$$LNE' \equiv L'_{12,2019}$$

Donde  $L'_{12,2019}$  es el número de ciudadanos inscritos en la LNE del municipio  $l$ , al corte de diciembre de 2019.

- 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 área urbanizada, es decir, este indicador aporta información sobre el grado de dispersión geográfica del electorado. Desde la perspectiva de las tareas de IIMA se considera que aquellos municipios que poseen una densidad baja (es decir, una alta dispersión geográfica del electorado) presentan

un mayor grado de complejidad electoral, ya que su baja densidad permite deducir que los municipios son de naturaleza rural, lo cual implica potencialmente resolver diversas dificultades logísticas para ejecutar adecuadamente las tareas de IIMA. En contraste, para el caso de aquellos municipios que poseen una densidad alta (es decir, una alta concentración geográfica del electorado), se puede derivar que los municipios son de naturaleza urbana, lo cual conlleva diversas ventajas logísticas para el desarrollo de las tareas de IIMA.<sup>7</sup>

Para obtener este indicador se utilizaron las estimaciones de la LNE y la estimación del área superficial cubierta por el amanzamiento para obtener el área urbanizada por municipio:<sup>8</sup>

$$\text{Densidad\_LNE}^l \equiv \frac{L^l_{12,2019}}{A^l_{12,2019}}$$

Donde  $A^l_{12,2019}$  es el área superficial (kilómetros cuadrados) del amanzamiento urbano del municipio  $l$ , al corte de diciembre de 2019.

- 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 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). Este indicador ayuda a explicar el nivel de complejidad electoral asociado al grado de actualización del RFE, ya que se considera que aquellos municipios que presentan una baja valoración en este indicador poseen un mayor grado de complejidad electoral, pues es necesario implementar con mayor prontitud mecanismos

<sup>7</sup> 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 rurales), con el fin de garantizar a todos los ciudadanos el acceso a los trámites de credencialización.

<sup>8</sup> Al utilizar la información cartográfica del amanzamiento (calculado por el INE) se obtiene una estimación más precisa de la superficie que se encuentra urbanizada en los municipios.

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:

$$Razon\_LNE\_PE^l \equiv \frac{L^l_{12,2019}}{P^l_{12,2019}}$$

Con  $P^l_{12,2019}$  igual al número de ciudadanos registrados en el PE del municipio  $l$ , al corte de diciembre de 2019.

- 4) **Tasa de crecimiento anual de la LNE (TC\_LNE\_2019).** Este indicador demográfico da cuenta del crecimiento (o decremento) poblacional que se observa en los municipios. Se espera una mayor complejidad electoral en los municipios de rápido crecimiento demográfico, ya que dicho fenómeno impacta en los procesos de integración, identificación y actualización del RFE. 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:

$$TC\_LNE\_2019^l \equiv \frac{L^l_{12,2019} - L^l_{1,2019}}{L^l_{1,2019}}$$

Con  $L^l_{1,2019}$  igual al número de ciudadanos inscritos en la LNE del municipio  $l$ , al corte de enero de 2019.

- 5) **Coefficiente de variación de la LNE (Coef\_Var\_LNE).** Este indicador demográfico, obtenido a partir del análisis de las series de tiempo de la LNE,<sup>9</sup> proporciona información acerca de la variación demo-

<sup>9</sup> En este indicador, el análisis de las series de tiempo se realizó de julio de 2019 a diciembre de 2019, ya que en ese año hubo procesos electorales en las entidades de Aguascalientes, Baja California, Durango, Puebla, Quintana Roo y Tamaulipas, lo cual implicó que en dichos estados se detuvieran temporalmente las actividades de IIMA en los meses previos a la jornada electoral.

gráfica observada en los municipios.<sup>10</sup> En concreto, se considera que hay mayor grado de complejidad electoral en los municipios sujetos a mayor nivel de variación poblacional, lo cual inherentemente afecta la planeación institucional referente a los procesos de identificación y actualización del RFE. Asimismo, los municipios que presentan una mayor variación poblacional son más complejos desde un punto de vista electoral, ya que es más difícil obtener estimaciones y pronósticos demográficos certeros para efectos de la planeación institucional.

$$\text{Coef\_Var\_LNE}^l \equiv \frac{\sigma_{L,2019}^l}{\bar{L}_{L,2019}^l} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (L_{i,2019}^l - \bar{L}_{L,2019}^l)^2}}{\sum_{i=1}^N \frac{L_{i,2019}^l}{N}}$$

Con  $L_{i,2019}^l$  igual al número de ciudadanos inscritos en la LNE del municipio  $l$ , al corte del mes de 2019.

- 6) Autocorrelación mensual de la LNE (Autocorr\_LNE).** Esta métrica estructural mide el grado de similitud entre una serie temporal y una versión pasada de sí misma. Para los propósitos de cuantificar la complejidad electoral, es de particular interés calcular la autocorrelación de la LNE con un rezago mensual (es decir, 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.<sup>11</sup>

En resumen, se considera que los municipios con un nivel de autocorrelación mensual bajo presentan un mayor nivel de complejidad

<sup>10</sup> Existen otros indicadores que también se utilizan para describir las variaciones temporales de la dinámica poblacional, por ejemplo, el coeficiente de variación proporcional (Fernández-Martínez *et al.*, 2018). En esta investigación se analizó la pertinencia de incorporar el coeficiente de variación proporcional, sin embargo, el análisis exploratorio de datos mostró innecesario incorporarlo al conjunto de indicadores base. Los detalles del análisis se pueden consultar en Álvarez Hernández (2021).

<sup>11</sup> En este indicador, el análisis de las series de tiempo también se realizó de julio de 2019 a diciembre de 2019, a fin de no considerar los meses previos a la jornada electoral.

electoral en comparación con los municipios con un nivel de autocorrelación mensual alto, ya que el comportamiento temporal de su LNE es más irregular.

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

**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 de secciones electorales de tipo rural que posee cada municipio.<sup>12</sup> 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 de IIMA del RFE.

### *Transformación de datos y elaboración de los ICE*

El paradigma moderno de la construcción 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 (Álvarez Hernández y Álvarez Texocotitla, 2020).

Por lo tanto, siguiendo los preceptos del paradigma moderno de elaboración de indicadores socioeconómicos, se construye un conjunto de índices de complejidad electoral (ICE) a partir de la información contenida en los siete indicadores base. Estos índices permitirán cuantificar la complejidad del sistema geoelectoral a escala municipal.

<sup>12</sup> En INE (2017), se definen las secciones rurales como aquellas secciones electorales que están conformadas enteramente por localidades con una población menor a 1500 ciudadanos.



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 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 permiten establecer relaciones de orden y magnitud entre las diferentes unidades territoriales del sistema geoelectoral, ya que posibilitan la elaboración de estratificaciones y listas ordenadas (*ranking*) de municipios en función de su grado de complejidad electoral. Asimismo, los ICE también 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 subjetiva, la influencia de cada indicador en la determinación de la complejidad electoral.

Consecuentemente, para construir los ICE se utiliza la técnica PCA. 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 la transformación de datos para la obtención de índices sintéticos (véase Consejo Nacional de Población [Conapo], 2002, para ejemplificación de la implementación del PCA a un problema de construcción de indicadores socioeconómicos). 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 mapea

el conjunto original de variables a un nuevo conjunto transformado de variables (los llamados componentes principales).

- 2) Estima en una forma no arbitraria los coeficientes de ponderación de las variables originales, ya que los componentes principales representan una combinación lineal (suma ponderada) de las variables originales. Específicamente, se puede utilizar el primer componente principal como índice de ordenación cardinal, ya que esta nueva variable explica, comúnmente, la mayor parte de la varianza encontrada en los datos.

Para el caso concreto de esta investigación, el PCA 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. 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 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:

$$ICE_i^l \equiv \sum_{j=1}^n L_i^j x_j^l$$

Donde  $ICE_i^l$  es el grado de complejidad electoral del municipio  $l$ , correspondiente al índice de complejidad electoral  $i$ , es el ponderador correspondiente a la variable original  $j$  del índice de complejidad electoral  $i$ ,  $x_j^l$  es el valor del municipio  $l$  en la variable original  $j$ , mientras que  $n$  es el número de variables originales introducidas en el PCA.

Para la elección de la cantidad de componentes principales extraídos a partir del PCA, existen diversos criterios (Gniazdowski, 2017). En particular, la presente investigación utiliza el criterio porcentual de la parte de la varianza explicada por componente principal, ya que este criterio permite evaluar la capacidad de síntesis de los componentes.

Como se mostrará a detalle en la siguiente sección del artículo, los resultados del criterio de varianza explicada permiten justificar la elección del primer componente principal como el principal ICE y, por lo tanto, se puede construir un *ranking* nacional de complejidad electoral a partir de dicho índice. Los subsiguientes componentes principales (es decir, los subsiguientes ICE) también se utilizan para obtener valoraciones alternativas de la complejidad electoral, ya que cada ICE aporta una estructura de ponderación diferente a los indicadores base.

### *Ranking y estratificación de complejidad electoral*

A pesar de las ventajas que ofrece el PCA, es importante examinar cuidadosamente los resultados obtenidos con dicho método, esto con el fin de valorar la capacidad explicativa de los ICE y su interpretación en función de las variables originales. Más adelante se mostrará que, para nuestro caso de estudio, los resultados del método PCA 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 (reflejado en el porcentaje de varianza explicada por el primer componente principal), y, por lo tanto, no está plenamente justificado utilizar exclusivamente el ICE principal como métrica de comparación del grado de complejidad electoral que presentan los municipios.

En este contexto, se puede solventar la falta de solidez estadística del *ranking* de complejidad electoral (fundamentado sólo en el ICE principal) mediante una estratificación de complejidad electoral a nivel municipal, la cual hace uso simultáneo de la información contenida en un subconjunto seleccionado de ICE (específicamente los primeros tres ICE), a fin de lograr un mayor porcentaje de varianza explicada.

Para generar la estratificación se decide utilizar el método de Ward de agrupamiento jerárquico, el cual es utilizado frecuentemente en el campo de los algoritmos de aprendizaje no supervisado, ya que proporciona un enfoque sencillo para aproximar, de cualquier número de estratos, la parti-

ción que minimiza la varianza total dentro de los estratos.<sup>13</sup> Además de su simplicidad y el hecho de que se basa en un criterio de optimización intuitivo, el método de Ward suele acompañarse con una representación gráfica llamada dendrograma, la cual se utiliza como soporte para la selección de la configuración óptima de estratos y la interpretación de los resultados (Randriamihamison *et al.*, 2021; Everitt *et al.*, 2011).

Por otra parte, para determinar el número óptimo de estratos 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 estratos de una configuración específica. Por ejemplo, si un objeto (perteneciente a un estrato específico) presenta un coeficiente de silueta cercano a 1, dicho objeto se encuentra muy alejado de otros pertenecientes a estratos 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 estratos vecinos. Finalmente, para el caso de un objeto con un coeficiente de silueta cercano a -1, esto implica que dicho objeto se encuentra más cercano a otro estrato que al que le fue asignado.

En resumen, la implementación conjunta del método de Ward, la inspección visual del dendrograma y el análisis de los coeficientes de silueta permite determinar el número y la configuración óptima de grupos para estratificar los municipios en función de las valoraciones que se obtuvieron en un subconjunto de los ICE. En otras palabras, el análisis de agrupamiento permite construir una estratificación de complejidad electoral a nivel municipal en el espacio de proyección de los ICE.

---

<sup>13</sup> Es importante puntualizar que existen diferentes algoritmos y métodos de estratificación. Por ejemplo, una alternativa que es usada frecuentemente para abordar problemas de estratificación univariada es el método de Dalenius y Hodges. En este artículo no se implementa dicho método por dos razones: 1) los resultados del PCA indican que los ICE no alcanzan un nivel aceptable de síntesis de información, por lo que cualquier estratificación que se fundamente sólo en uno de estos índices fallará en la tarea de clasificar adecuadamente el grado de complejidad de los municipios, 2) la estratificación se realiza sobre un conjunto de variables, es decir, es necesario resolver un problema de estratificación multivariada sobre los ICE.

## Índices de complejidad electoral a nivel municipal

La construcción de los ICE municipales inicia con un análisis exploratorio de los siete indicadores base.<sup>14</sup> Posteriormente, se aplica la técnica PCA para transformar el conjunto de indicadores a un nuevo grupo de variables (los componentes principales). Una vez calculados los componentes principales, se analizan sus capacidades para sintetizar la información de los indicadores base y se asignan los ICE a los componentes principales resultantes. De esta forma, se obtienen las proyecciones de los municipios en el nuevo espacio de transformación conformado por los ICE (lo cual permite analizar la emergencia de patrones distinguibles de agrupamiento en las proyecciones), se determina la composición de los ICE, es decir, la ponderación que recibe cada indicador base, y se analizan los resultados del *ranking* municipal de complejidad electoral y de la estratificación municipal generada por el método de Ward.<sup>15</sup>

### *Análisis exploratorio y construcción de los ICE*

En la figura 1 se muestra un panel de gráficas (correlogramas) que permite realizar un análisis exploratorio gráfico del conjunto de indicadores base. 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).<sup>16</sup>

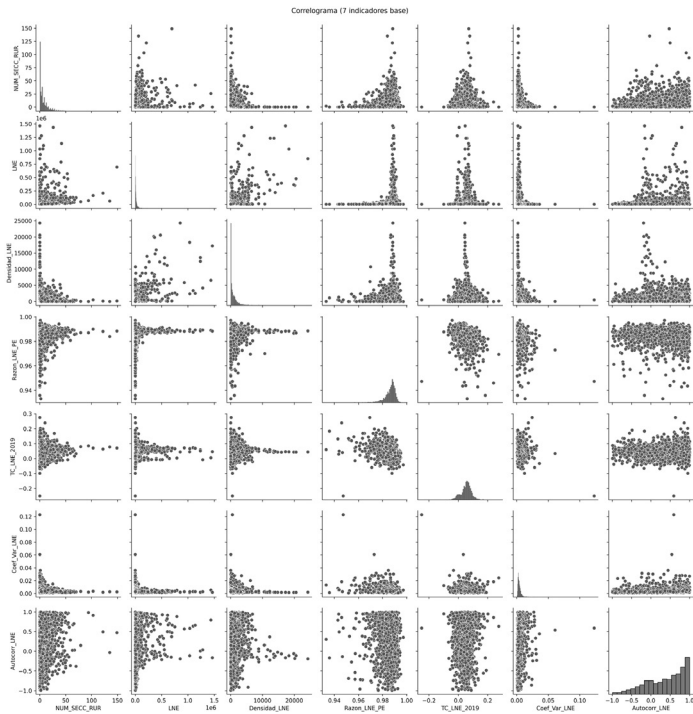
---

<sup>14</sup> Se utilizó la información completa de los siete indicadores base para el universo total de 2456 municipios existentes en la base de datos geoelectoral del INE (2019).

<sup>15</sup> Para una explicación y demostración amplia de la implementación de los métodos de análisis exploratorio y transformación de datos en Python, se recomienda revisar Pedregosa *et al.* (2011).

<sup>16</sup> 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).

**Figura 1. Panel de gráficas de correlación y distribución de los indicadores base que conforman los ICE**



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

Por ejemplo, la figura 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 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).

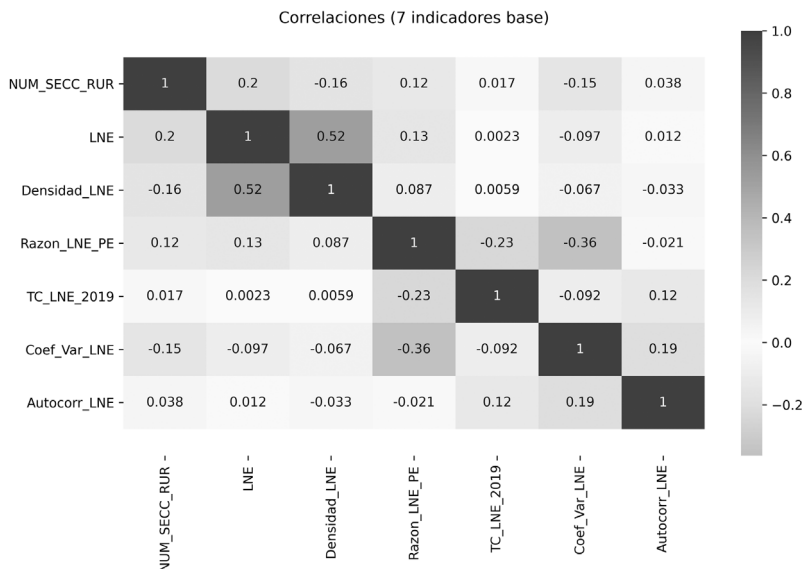
Asimismo, el análisis exploratorio mostrado en la figura 1 también permite apreciar la presencia de correlaciones significativas entre los indicadores.<sup>17</sup> Por ejemplo, se observa la presencia de correlaciones positivas entre los indicadores LNE y Densidad\_LNE, además de correlaciones negativas entre el Coeficiente de Actualización del RFE (Razón\_LNE\_PE) y los indicadores de variabilidad demográfica (TC\_LNE\_2019 y Coef\_Var\_LNE). 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 la figura 2.

En particular, la figura 2 permite destacar que a nivel municipal no hay indicios de que la magnitud y la densidad del electorado estén relacionadas 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.

---

<sup>17</sup> 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.

**Figura 2. Mapa de calor de la matriz de correlación de los indicadores municipales<sup>18</sup>**



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

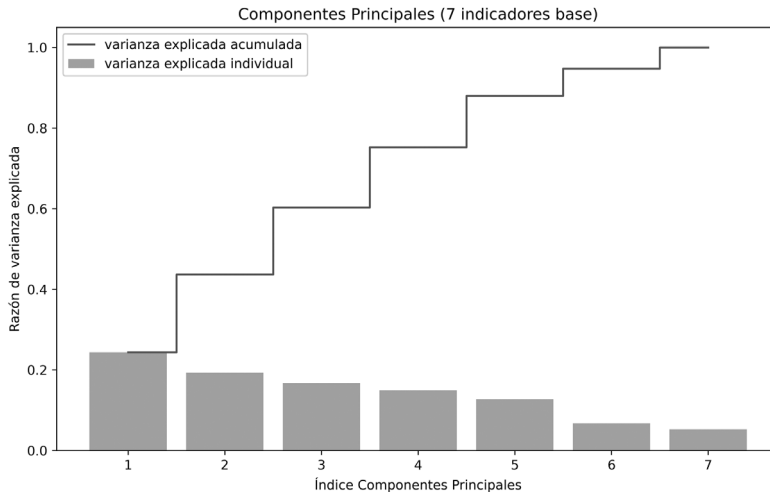
Recapitulando, los resultados del análisis exploratorio indican la presencia de correlaciones significativas entre los indicadores, lo cual da una justificación fundamental para la implementación del PCA. La primera fase de la implementación del PCA es el preprocesamiento de datos (normalización y centralizado).<sup>19</sup> Posteriormente, se calcula la matriz de covarianza, con sus respectivos *eigenvalores* y *eigenvectores*, para generar los componentes principales. El resultado de la varianza explicada por cada componente principal se presenta en la figura 3.

<sup>18</sup> Los números de cada recuadro indican el coeficiente de correlación entre el par de indicadores. Y la escala de color resalta la fuerza de la correlación.

<sup>19</sup> El propósito principal de aplicar un preprocesamiento de datos es reducir la varianza observada en los indicadores base, ya que éstos se encuentran en diferentes unidades de medida, lo cual induce una falsa variabilidad en los datos.



**Figura 3. ICE: proporción de varianza explicada por componente principal**



Fuente: Elaboración propia a partir de la librería Python-Scikit-learn-Seaborn.

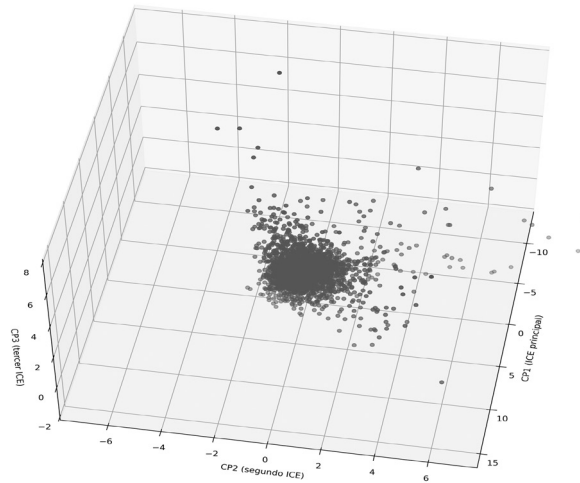
La figura 3 muestra que el primer componente principal (CP1) alcanza un porcentaje de varianza explicada aproximado de 24 %, es decir, esta variable transformada presenta el mejor nivel de varianza explicada dentro del grupo de componentes principales y, en consecuencia, se toma como el principal ICE municipal. Sin embargo, el 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 (no alcanza un mínimo de 50 % de la varianza explicada), por lo que debe usarse con cautela para la elaboración de un *ranking* que permita establecer comparaciones de complejidad electoral entre los municipios.

Por otra parte, en la figura 3 también se observa la contribución decreciente de los subsiguientes componentes principales en el porcentaje de varianza explicada. Por ejemplo, el segundo componente principal (CP2 o segundo ICE municipal) contribuye con 20 % al porcentaje de varianza explicada, mientras que el tercer componente principal (CP3 o tercer ICE municipal) contribuye sólo con 17 %.

Consecuentemente, los resultados obtenidos en el PCA indican que, en forma individual, cada ICE no alcanza un nivel satisfactorio de síntesis de la información para establecer comparaciones consistentes de complejidad electoral entre los municipios. Sin embargo, se puede observar que, si se consideran los primeros tres ICE, se alcanza un porcentaje de varianza explicada aproximado de 61 %. En consecuencia, se justifica la realización de un análisis de agrupamiento, mediante el método de Ward, para obtener una estratificación de complejidad electoral a partir de los resultados de los primeros tres ICE.

En la figura 4 se presenta la proyección tridimensional de los municipios en el subespacio de los componentes CP1, CP2 y CP3, en donde se pueden apreciar los patrones de agrupamiento en la nube de dispersión de los municipios (cúmulos en las partes externas de la nube de dispersión). En particular, en la figura 4 se observa que la aparente acumulación de municipios alrededor del origen de la figura 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 en su valoración respecto con el CP3 (tercer ICE). Estas estructuras de agrupamiento se hacen más patentes cuando se rota la proyección tridimensional a lo largo de la dirección de alguno de los componentes principales.

**Figura 4. Distribución de los municipios en el subespacio CP1-CP2-CP3**

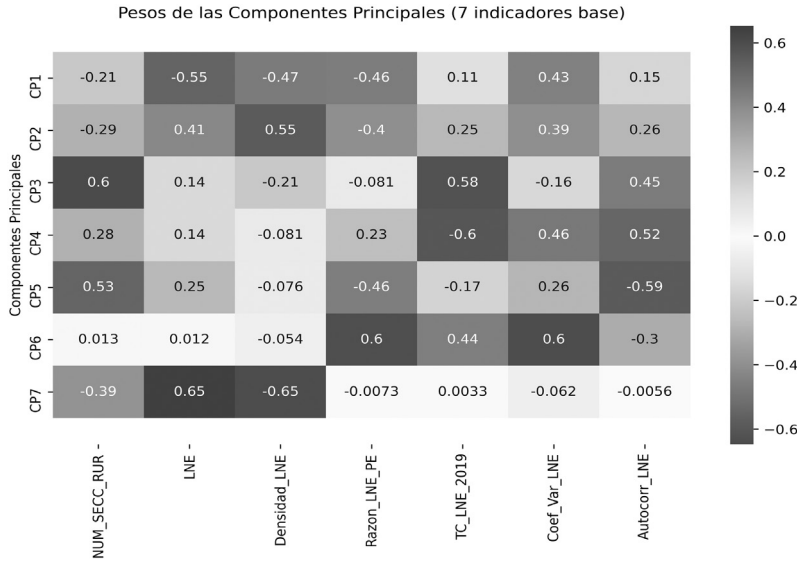


Fuente: Elaboración propia a partir de la librería Python-Scikit-learn-Seaborn.

### *Interpretación de los ICE*

Para interpretar adecuadamente la información que aporta cada uno de los ICE se analiza la ponderación que recibió cada indicador base en la conformación de los componentes principales mediante un mapa de calor (figura 5). En particular, entre mayor sea la intensidad de color, más fuerte es la ponderación positiva (o negativa) del indicador en el ICE municipal.

**Figura 5. ICE: Ponderadores de los indicadores base en cada componente principal<sup>20</sup>**



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

Para comprender adecuadamente la figura 5, se explicará con detalle la composición de los primeros tres componentes principales. La figura 5 muestra que los indicadores que mayor peso tienen en la CP1 (ICE principal) son: magnitud del electorado, densidad del electorado, coeficiente de actualización del RFE (con pesos negativos) y coeficiente de variación de la LNE (con peso positivo). Respecto al segundo componente principal, el mayor peso lo tienen los indicadores de magnitud del electorado, densidad del electorado, coeficiente de variación de la LNE (con pesos positivos), y el coeficiente de actualización del RFE (con peso negativo). Finalmente, el tercer componente principal da mayor peso a los indicadores: grado de urbanización

<sup>20</sup> En este caso, los números que se encuentran en los recuadros indican los coeficientes de ponderación de cada indicador base en los ICE.

municipal, tasa de crecimiento anual de la LNE y autocorrelación mensual de la LNE (con pesos positivos).

Considerando que los indicadores que más pesan en el ICE principal (en forma negativa) son los 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 ICE principal, menor será el número de ciudadanos inscritos en su LNE y mayor será la dispersión geográfica del electorado (densidad baja). Asimismo, el coeficiente de actualización del RFE también influye negativamente, mientras que el coeficiente de variación de la LNE presenta una ponderación positiva. En consecuencia, entre mayor sea la valoración de un municipio en el ICE principal, menor será el nivel de actualización entre la LNE y el PE de dicho municipio y será mayor la variabilidad demográfica en la LNE.

Por otra parte, el segundo ICE municipal está influido principalmente por los indicadores que describen las variaciones poblacionales observadas en la LNE 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 segundo ICE municipal, éste se caracterizará por tener una mayor variabilidad en su LNE, junto con un menor nivel de actualización en la LNE y el PE. Asimismo, en este índice los indicadores asociados a la magnitud y dispersión geográfica del electorado presentan una ponderación positiva significativa, lo cual implica que los municipios que presentan una alta valoración positiva en el segundo ICE municipal también se caracterizan por presentar un alto número de ciudadanos inscritos en la LNE, junto con una menor dispersión geográfica del electorado (densidad alta).

Finalmente, el tercer ICE municipal tiene cuatro influencias principales. En primer lugar (ponderación positiva), se encuentran los indicadores: grado de urbanización municipal, tasa de crecimiento de la LNE y autocorrelación mensual de la LNE. En segundo lugar (ponderación negativa), se encuentra el indicador densidad del electorado. Por consiguiente, este ICE 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 y tasas positivas de crecimiento poblacional.

### *Ranking y estratificación municipal de complejidad electoral*

Dada la cantidad de municipios examinados (2456 municipios), no es posible ofrecer en este documento un desglose completo de los resultados obtenidos en el conjunto total de los ICE. Por lo tanto, en la tabla 2 se presenta una muestra del *ranking* municipal de complejidad electoral obtenida a partir de los resultados de la valoración de los municipios en el ICE principal. En particular, se muestran los primeros 10 municipios con mayor nivel de complejidad electoral a nivel nacional, junto con los 10 municipios con menor grado de complejidad.

**Tabla 2. Muestra del *ranking* municipal de complejidad electoral: municipios con mayor y menor valoración en el ICE principal**

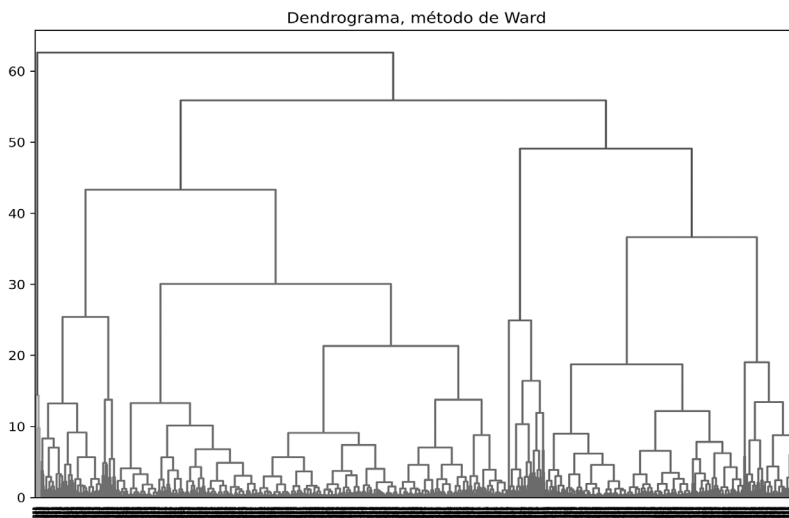
<i>Ranking</i>	Entidad	Municipio	CP1 (ICE principal)	<i>Ranking</i>	Entidad	Municipio	CP1 (ICE principal)
1	Oaxaca	Santiago Lalopa	9.1234	2447	Ciudad de México	Benito Juárez	-7.0663
2	Oaxaca	Santa Magdalena Jicotlan	8.2881	2448	Ciudad de México	Iztacalco	-7.2352
3	Chihuahua	Uruachi	8.1483	2449	Puebla	Puebla	-7.7739
4	Oaxaca	San Baltazar Yatzaqui el Bajo	6.0350	2450	Ciudad de México	Cuauhtémoc	-7.8410
5	Oaxaca	Abejones	5.5773	2451	Baja California	Tijuana	-9.0917
6	Oaxaca	San Pedro Tidaá	5.5248	2452	Jalisco	Guadalajara	-9.6251
7	Chiapas	Jitotol	5.4768	2453	México	Ecatepec de Morelos	-9.9246
8	Oaxaca	San Pedro Mártir Quiechapa	5.4745	2454	Ciudad de México	Gustavo A. Madero	-10.1779
9	Oaxaca	San Juan Juquila Vijianos	5.3696	2455	México	Nezahualcóyotl	-10.8453
10	Oaxaca	San Juan Tabaa	5.2426	2456	Ciudad de México	Iztapalapa	-12.0489

Fuente: Elaboración propia.

La información mostrada en la tabla 2 permite señalar que los municipios de mayor complejidad electoral (respecto al ICE principal) se encuentran localizados principalmente en zonas geográficas alejadas de los principales centros urbanos de México (por ejemplo, 9 de los primeros 10 municipios de mayor complejidad electoral se encuentran ubicados en los estados de Chiapas y Oaxaca). Esto contrasta con el caso de los municipios con menor grado de complejidad electoral (respecto al ICE principal), los cuales se ubican principalmente en el área metropolitana de la Ciudad de México.

Con el objetivo de analizar otros patrones que surgen a partir de los ICE, se presentan los resultados de la estratificación de complejidad electoral obtenida a partir del método de agrupamiento jerárquico de Ward. Como se comentó previamente, se utiliza la proyección de los municipios en el subespacio de los primeros tres ICE para realizar el análisis de agrupamiento, el cual permite establecer estratos de municipios con grados similares de complejidad electoral. Para determinar el número óptimo de estratos se realiza un análisis visual del dendrograma generado por el método de Ward, y se complementa con un análisis de coeficientes de silueta.



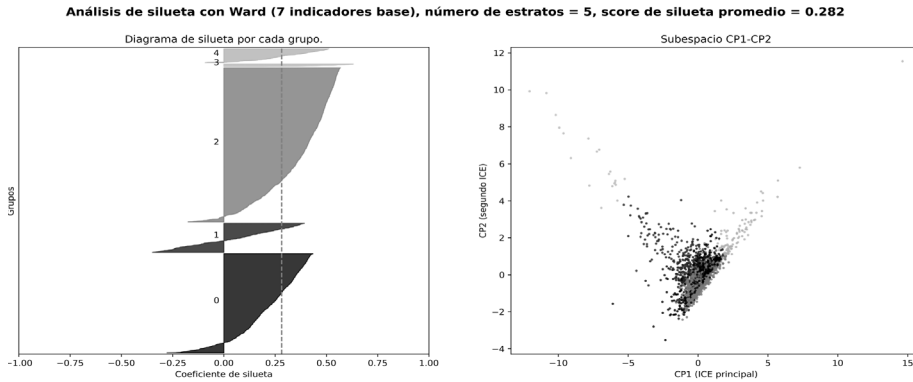
**Figura 6. Dendrograma generado por el método de Ward**

Fuente: Elaboración propia con la librería Python-Scikit-learn.

En la figura 6 se muestra el dendrograma obtenido. En el eje vertical se encuentra representada la distancia o disimilitud entre estratos, mientras que en el eje horizontal se representan los municipios y los estratos. Cada unión (fusión) de dos estratos se representa en el dendrograma mediante la división de una línea vertical en dos líneas verticales. La posición vertical de la división, mostrada por las barras horizontales, da la distancia (disimilitud) entre los dos estratos. En particular, el dendrograma de la figura 6 sugiere que una configuración de cinco estratos puede considerarse adecuada para estratificar los municipios.

Posteriormente, en la figura 7 se muestran los resultados del análisis de coeficientes de silueta para la configuración de cinco estratos. En particular, la figura 7 señala que la configuración de cinco estratos puede considerarse óptima (comparado con escenarios de 3 a 7 estratos), ya que ésta reporta el mayor coeficiente de silueta promedio (que en este caso fue de 0.282). Los detalles de las pruebas con diferentes escenarios se pueden consultar en Álvarez Hernández (2021).

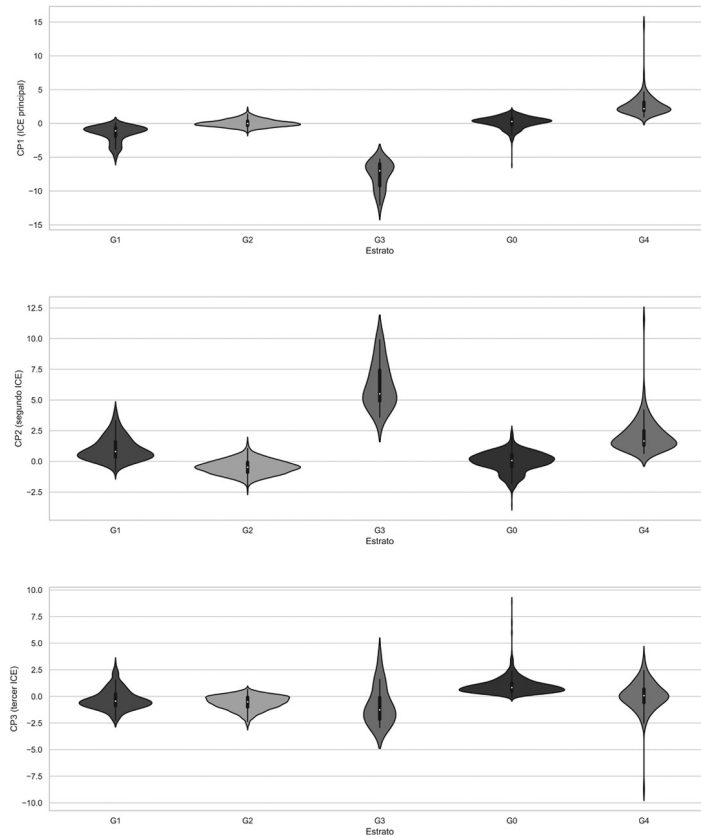
**Figura 7. Diagrama de silueta para la configuración de 5 estratos, y estratificación en el subespacio CP1-CP2**



Fuente: Elaboración propia con la librería Python-Scikit-learn.

Con la configuración óptima de estratos (que reciben las denominaciones G0, G1, G2, G3 y G4), se agrupan los municipios y se analizan las estadísticas de tendencia central de cada estrato (mediante los diagramas de violín mostrados en la figura 8) para identificar sus rasgos característicos e identificar algunos patrones geográficos (véase Hintze y Nelson, 1998, para más detalles en la interpretación de las gráficas de violín).

**Figura 8. Panel de diagramas de violín de los primeros tres ICE (con estratificación)**



Fuente: Elaboración propia con la librería Python-Scikit-learn.

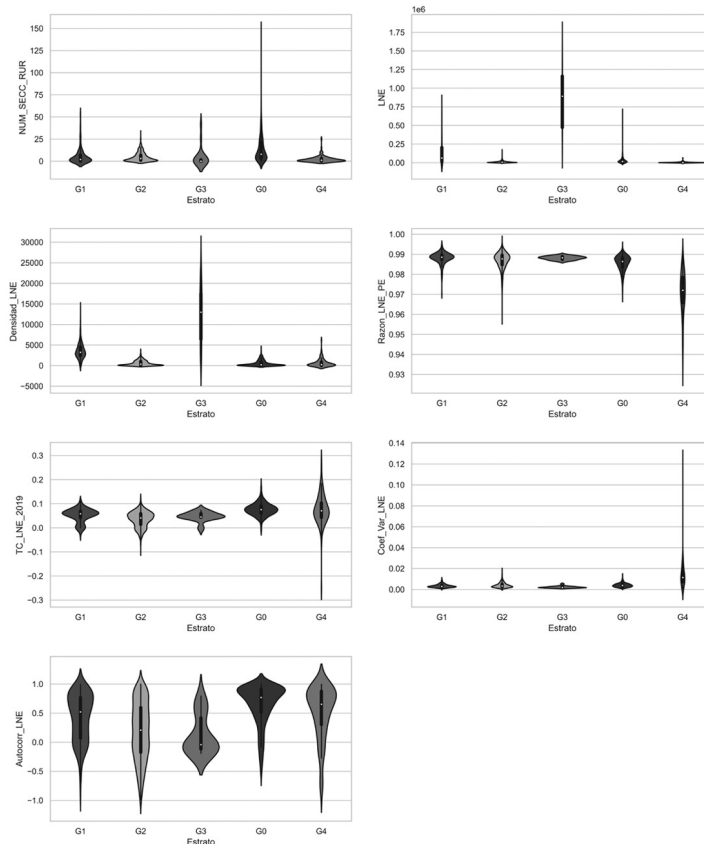
En la figura 8 se puede observar que los estratos G1, G2 y G0 presentan estadísticas de tendencia central similares en el ICE principal, ya que exhiben un valor medio equiparable (las medias se encuentran indicadas por los puntos blancos dentro de las figuras de violín), y algo similar se observa con respecto al tamaño de los rangos intercuartílicos (representados por los rectángulos negros dentro de las figuras de violín) y la longitud de los

intervalos de confianza de 95 % (descritos por las líneas verticales que salen de los rectángulos negros). En contraste, el estrato G4 exhibe una media más alta con respecto a los demás estratos, mientras que, comparativamente, el estrato G3 presenta el menor valor promedio en el ICE principal. Es decir, los municipios clasificados en el estrato G4 presentan una complejidad electoral más alta, mientras que los municipios pertenecientes a los estratos G0, G1 y G2 muestran un nivel de complejidad electoral similar, y los municipios clasificados en el estrato G3 presentan una complejidad electoral baja.

Por otra parte, la figura 9 también muestra que los municipios de los estratos G3 y G4 presentan características diferentes en el segundo ICE, ya que ambos estratos muestran una media más alta respecto a los demás (es decir, una mayor complejidad electoral). De igual forma, en el tercer ICE se observa un valor medio equiparable en todos los estratos, a excepción de los estratos G3 y G4, los cuales presentan un valor promedio negativo o cercano a cero.

Un elemento más que se puede observar en la figura 9, a través de las figuras de violín, es la forma de las distribuciones de valores de cada estrato. En particular, se puede observar que los estratos G1 y G2 presentan pocos municipios atípicos (figuras de violín cortas), mientras que los estratos G0 y G4 exhiben un mayor número de valores atípicos (figura de violín alargada); es decir, las distribuciones de estos estratos presentan un mayor comportamiento leptocúrtico con respecto a los demás grupos de la estratificación.

**Figura 9. Panel de diagramas de violín de los siete indicadores base (con estratificación)**



Fuente: Elaboración propia con la librería Python-Scikit-learn.

Teniendo presente la interpretación de la información contenida en la figura 8, se procede al análisis de la figura 9, la cual muestra un panel de diagramas de violín generado (en función de la estratificación) para cada indicador base. Se puede observar que los cinco estratos muestran estadísticas similares (media, rango intercuartílico e intervalos de confianza) en los indicadores TC\_LNE\_2019, NUM\_SECC\_RUR, Razon\_LNE\_PE y

Coef\_Var\_LNE. Sin embargo, hay algunas diferencias notables en las distribuciones (que ya se identificaban en la figura 8); por ejemplo, el estrato G0 posee municipios que muestran una cantidad atípica de secciones rurales y el estrato G4 tiene municipios con un elevado grado de variabilidad demográfica, lo cual puede relacionarse a un crecimiento (o decremento) poblacional pronunciado, el cual impacta negativamente en el nivel de actualización del RFE.

Con respecto a los indicadores de LNE, Densidad\_LNE y Autocorr\_LNE, se aprecian algunas diferencias entre las estadísticas de tendencia central de los estratos, especialmente en el G3, ya que los municipios de este estrato presentan en promedio un electorado más numeroso y una menor dispersión geográfica en su población.

### *Patrones geográficos de la estratificación*

A partir de la información mostrada en las figuras 8 y 9, se pueden deducir *a priori* algunas tendencias geográficas de la estratificación. Por ejemplo, la figura 8 muestra que los municipios clasificados en el estrato G3 manifiestan una baja complejidad electoral asociada a los factores demográfico y geoespacial, ya que éstos presentan una valoración negativa en el ICE principal (lo cual indica un nivel alto de actualización en el RFE, así como un electorado de gran magnitud con poca dispersión geográfica) junto con una valoración positiva en el segundo ICE, y negativa en el tercer ICE (lo cual implica una baja variabilidad y crecimiento demográfico en la LNE, así como un número mínimo de secciones rurales). En consecuencia, los municipios del estrato G3 deben corresponder a municipios plenamente urbanizados situados en las principales zonas metropolitanas de México.

En contraste, los municipios clasificados en el estrato G4 muestran una complejidad electoral alta asociada al factor demográfico y geoespacial, ya que su valoración negativa en el ICE principal implica un electorado de poca magnitud con dispersión geográfica significativa (no obstante, la densidad promedio del estrato es la segunda más baja, ya que el estrato G0 presenta la menor densidad promedio). Asimismo, los municipios del estrato

G4 también muestran una valoración positiva en el segundo ICE, por lo cual dichos municipios también se caracterizan por una alta variabilidad poblacional en la LNE (que parece no afectar el comportamiento mensual de la LNE, pero sí el grado de actualización del RFE), aunque esta variabilidad no es provocada por el crecimiento demográfico (como se puede observar en la figura 10). Por lo tanto, se infiere que estos municipios son de poca extensión territorial y están ubicados en regiones rurales que presentan una dinámica migratoria activa.

Con respecto a los municipios clasificados en los estratos G0, G1 y G2, se puede considerar que éstos presentan un nivel de complejidad electoral similar e intermedio comparado con los municipios del estrato G3 (complejidad baja) o del estrato G4 (complejidad alta). No obstante, en algunos indicadores base hay algunas disimilitudes importantes entre los estratos mencionados.

Por ejemplo, los municipios del estrato G0 se caracterizan en promedio por un alto número de secciones rurales y una densidad baja del electorado (a pesar de que la magnitud promedio de la LNE es la segunda más baja de los cinco estratos), lo cual sugiere que son municipios rurales de gran extensión territorial. Asimismo, este estrato presenta la segunda tasa promedio de crecimiento de la LNE más alta, pero al mismo tiempo presenta poca variabilidad en el comportamiento de la LNE, siendo esta última característica la disimilitud con respecto de sus pares del estrato G4. Por lo tanto, se puede considerar que el estrato G0 exhibe el segundo mayor grado de complejidad electoral en la estratificación.

En lo concerniente al estrato G2 (estrato que mayor cantidad de municipios concentra), se observan características muy similares a las observadas en el estrato G0, a excepción del grado de dispersión del electorado y la tasa de crecimiento de la LNE, ya que en el estrato G2, los municipios muestran una densidad del electorado más elevada, una menor cantidad de secciones rurales, así como poco o nulo crecimiento y variabilidad poblacional. Esto permite concluir que los municipios G2 son de carácter rural, tienen una extensión territorial contenida, exhiben estabilidad demográfica y presentan en promedio un menor grado de complejidad electoral comparado con los municipios de los estratos G4 y G0.

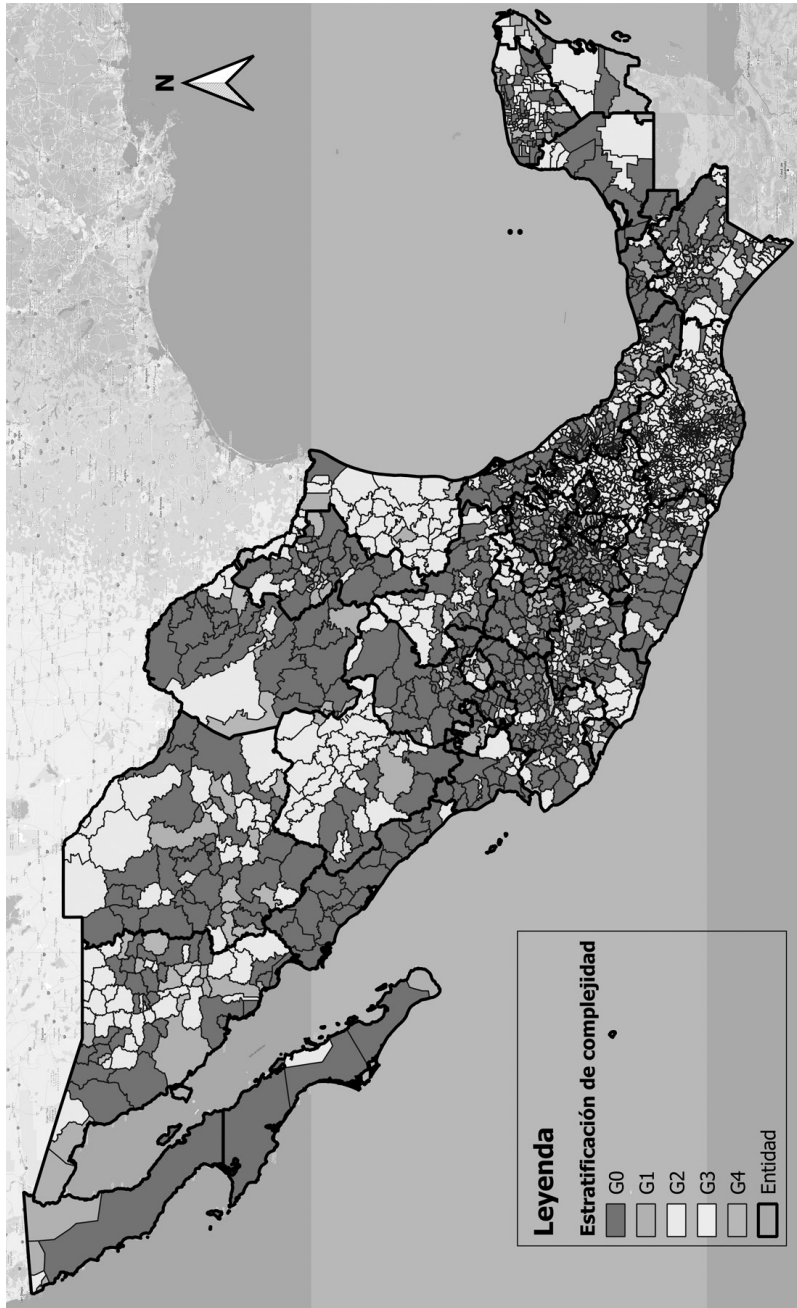
Finalmente, los municipios del estrato G1 también muestran un cierto grado de estabilidad demográfica (la tercera tasa de crecimiento de la LNE más alta), pero la densidad y magnitud promedio del electorado es más elevada en comparación con los estratos G0, G2 y G4 (es decir, una dispersión del electorado intermedia entre los municipios de carácter rural y los de carácter urbano). Consecuentemente, se infiere que en este estrato se clasifican los municipios que presentan una dinámica migratoria relativamente activa, la cual podría ser explicada por el proceso de conurbación debido a la cercanía de una zona metropolitana. Este dinamismo poblacional les proporciona a los municipios del estrato G1 un grado de complejidad electoral promedio más elevado con respecto a los municipios del estrato G3.

En la figura 10 se muestra el mapeo de la estratificación de complejidad electoral en la totalidad del territorio nacional y en la figura 11, un acercamiento de la estratificación en la zona metropolitana de la Ciudad de México. Ambas figuras permiten apreciar que las inferencias obtenidas a partir de las estadísticas de tendencia central son en efecto congruentes.

Por ejemplo, se observa que los municipios de gran extensión territorial ubicados en zonas rurales se clasifican en los estratos G0 (es el caso de algunos municipios de Baja California y Baja California Sur). Los municipios clasificados en el estrato G2 se distribuyen en la mayor parte del territorio nacional, predominando especialmente en los estados de Durango, Puebla, Oaxaca y Tamaulipas. Los municipios de mayor complejidad electoral, clasificados en el estrato G4, se concentran principalmente en las zonas rurales de los estados de Chiapas, Oaxaca y Sonora, mientras que los municipios de menor complejidad electoral, clasificados en los estratos G1 y G3, se concentran en los alrededores de las principales ciudades de México, como es el caso de la zona metropolitana de la Ciudad de México.

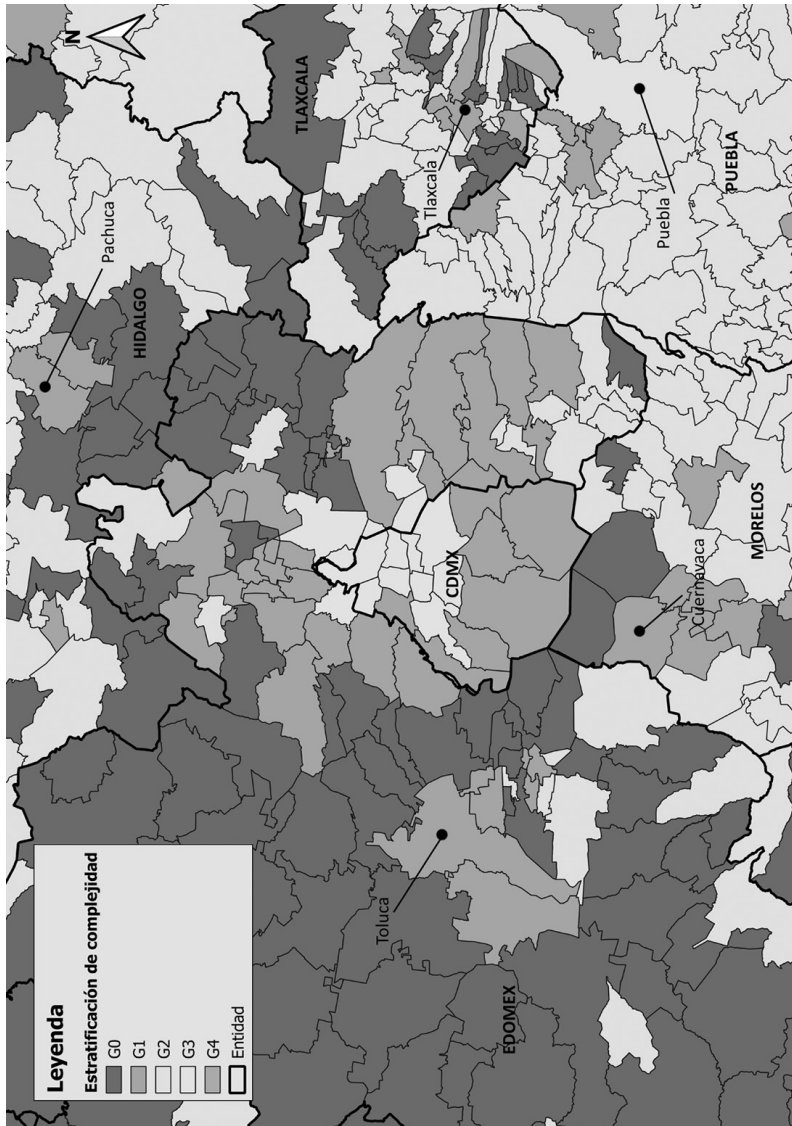


Figura 10. Estratificación de complejidad electoral



Fuente: Elaboración propia con QGIS.

Figura 11. Estratificación de complejidad electoral en la zona metropolitana de la Ciudad de México



Fuente: Elaboración propia con QGIS.

Para complementar los resultados y explicaciones presentados en esta sección, se recomienda revisar los archivos en Álvarez Hernández (2021), dado que en éstos se presenta un estadístico con el listado completo de municipios con su valoración en los ICE y respectiva estratificación. Asimismo, también se incluye el pseudocódigo de Python utilizado para la implementación del PCA y el método de Ward, además de los mapas generados con la estratificación de complejidad electoral.

## Conclusiones

La presente investigación ha permitido diseñar un conjunto de instrumentos analíticos que permiten cuantificar la complejidad del sistema geoelectoral de México a escala municipal, esto con el objetivo de monitorear y evaluar de forma más precisa la realización de las tareas de integración, identificación, mantenimiento y actualización del RFE. Asimismo, a partir del análisis holístico de los indicadores base, los resultados obtenidos otorgan una descripción más fidedigna y precisa de las actuales características geoespaciales, demográficas y estructurales presentes en el sistema geoelectoral mexicano.

Estos resultados se alcanzaron gracias a una metodología fundamentada en tres elementos: el concepto de *complejidad electoral*, el conjunto de indicadores base seleccionado para caracterizar los factores determinantes de la complejidad electoral a escala municipal y las técnicas de análisis multivariante para la construcción de los ICE y la estratificación de complejidad electoral.

El método PCA proporcionó un procedimiento matemático robusto para manejar la multidimensionalidad implícita en el conjunto de indicadores base, así como para modelar de forma holística la información contenida en los indicadores base. De esta manera, el PCA permitió obtener un conjunto de indicadores sintetizadores (ICE) para combinar y ponderar cada uno de los indicadores base. Sin embargo, los resultados obtenidos en esta investigación permiten concluir que los ICE, considerados indivi-

dualmente, presentan algunas limitaciones para sintetizar adecuadamente la información contenida en los indicadores base.

En consecuencia, para establecer mejores comparaciones de complejidad electoral entre las diversas unidades territoriales del sistema geoelectoral, es necesario contar con una estratificación que complemente los *ranking* de complejidad obtenidos a partir de los ICE.

El primer paso para obtener la estratificación de complejidad electoral fue seleccionar el método adecuado de estratificación, para posteriormente determinar el número óptimo de estratos que ésta debería de tener. En este contexto se utilizaron los análisis de coeficientes de silueta y dendrogramas, obteniendo resultados que permiten afirmar lo siguiente: el método de agrupamiento de Ward es adecuado para generar una estratificación de complejidad electoral consistente, siendo la configuración óptima la estratificación de cinco estratos. Es decir, todos los municipios son clasificados en alguno de los cinco estratos generados por el método de Ward, obteniendo como resultado final la estratificación del sistema geoelectoral mexicano en función de los grados de complejidad electoral que presentan los municipios.

A partir de los resultados generados por los ICE y la estratificación, se obtuvieron y analizaron algunos patrones geográficos relevantes que emergen a consecuencia de las distintas características demográficas y geospaciales que presentan los municipios. Por ejemplo, se determinaron algunos municipios de especial interés para el INE (caso de los estratos G0 y G4), ya que éstos presentan un grado de complejidad electoral notoriamente más elevado que el promedio de los municipios, y por ende es necesario monitorear y atender con más detalle estos municipios para garantizar la calidad y el funcionamiento de las tareas de IIMA implementadas en estos municipios. Asimismo, se mostró que, en promedio, los municipios localizados en zonas urbanas (estratos G1 y G3) exhiben características demográficas y geospaciales que facilitan el desarrollo de las tareas de IIMA y, por ende, presentan niveles de complejidad electoral bajos.

Para concluir, se señalan algunas recomendaciones y líneas de investigación futura. Ya que la metodología de cálculo de los ICE estuvo subordinada al volumen de datos examinado y a las capacidades de cómputo disponibles,

se sugiere la utilización de otros métodos alternativos de transformación de datos para avanzar en el análisis. Asimismo, una línea de investigación futura importante sería incluir en los ICE otro tipo de indicadores socioeconómicos y político-electorales, ya que es plausible ampliar el concepto de *complejidad electoral* para incorporar otros factores que potencialmente pudieran determinar la complejidad del sistema geoelectoral. Por ejemplo, se podrían incluir indicadores relacionados con la estructura demográfica, los flujos de migración (interna o externa), los resultados y preferencias de votación, el grado de desarrollo social, o la evolución epidemiológica de la pandemia de SARS-CoV-2. De esta forma, se podría hacer un análisis más detallado de las características sociales y políticas de las poblaciones que se agrupan en los diferentes estratos generados a partir de los ICE.

Igualmente, se sugieren actualizaciones periódicas de los ICE y de la estratificación (por ejemplo, en la etapa de inicio 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 y actualización del RFE.

Por último, es importante reiterar que el tema de la complejidad es vasto y difícil, ya que implica analizar los factores fundamentales que determinan el comportamiento de los sistemas, cuya dificultad se incrementa si se considera que aún falta determinar la efectividad de los enfoques de las ciencias de la complejidad frente a los enfoques tradicionales de análisis de las ciencias sociales. Los autores de la presente investigación aspiran que su trabajo sea un punto de partida para otros estudios sobre la complejidad en los sistemas sociales.





## Fuentes de consulta

- Abdi, Hervé y Williams, Lynne J. (2010). “Principal Component Analysis”. *WIREs Comp Stat*, 2, 433-459. doi: <https://doi.org/10.1002/wics.101>
- 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.
- Álvarez Hernández, M. D. (2021, 10 de agosto). *Artículo-Apuntes-Complejidad-sistema-geo-electoral-municipal*. GitHub. Recuperado el 10 de agosto de 2021, de <https://github.com/MDAlvarezH/Artículo-Apuntes-Complejidad-sistema-geo-electoral-municipal>
- 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. doi: <https://doi.org/10.29101/crcs.v27i83.12148>
- Byrne, David (2001). *Complexity Theory and the Social Sciences: An Introduction*. Londres: Routledge.
- Consejo Nacional de Población (2002). *Índices de marginación a nivel localidad*. Ciudad de México: Autor.
- Cooley, A. y Snyder, J. (2015). *Ranking the World*. Nueva York: Cambridge University Press.
- Everitt, Brian S.; Landau, Sabine; Leese, Morven; y Stahl, Daniel (2011). *Cluster Analysis*. Londres: Wiley.
- Faliszewski, Piotr; Hemaspaandra, Edith; Hemaspaandra, Lane A.; y Rothe, Joerg (2006). “A Richer Understanding of the Complexity of Election Systems”. Recuperado el 10 de febrero de 2019, 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. doi: <https://doi.org/10.1002/ecs2.2527>
- Gniazdowski, Zenon (2017). “New Interpretation of Principal Components Analysis”. *Zeszyty Naukowe WWSI*, 16 (11), 43-65. doi: <https://doi.org/10.26348/znwwsi.16.43>

- Haynes, Philip (2018). *Social Synthesis: Finding Dynamic Patterns in Complex Social Systems*. Nueva York: Routledge.
- 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*. Nueva York: Addison-Wesley.
- Instituto Nacional Electoral (2017). *Tipología de los distritos electorales para la aplicación de las actividades de la DERFE*. Ciudad de México: Autor.
- Instituto Nacional Electoral (2019). Cartografía Electoral. Recuperado el 10 de febrero de 2019 de <https://cartografia.ife.org.mx/>
- Jackson, J. E. (1991). *A User's Guide to Principal Components*. Nueva Jersey: Wiley.
- Klingemann, Hans-Dieter (2009). *The Comparative Study of Electoral Systems*. Nueva York: Oxford University Press.
- Ladyman, James; Lambert, James; y Wiesner, Karoline (2013). “What is a Complex System?”. *Euro Jnl Phil Sci*, 3, 33-67. doi: <https://doi.org/10.1007/s13194-012-0056-8>
- Lloyd, Seth (2001). “Measures of Complexity: A Nonexhaustive List”. *IEEE Control Systems Magazine*, 21 (4), 7-8. doi: <https://doi.org/10.1109/MCS.2001.939938>
- Maggino, Filomena (2017). *Complexity in Society: From Indicators Construction to their Synthesis*. Suiza: Springer.
- Mitchell, Melanie (2009). *Complexity: A Guided Tour*. Londres: Oxford University Press.
- Northrop, Robert B. (2011). *Introduction to Complexity and Complex Systems*. Nueva York: CRC Press.
- Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O; Blondel, M. *et al.* (2011). “Scikit-learn: Machine Learning in Python”. *JMLR*, 12, 2825-2830. Recuperado el 18 de junio de 2018, de [https://scikit-learn.org/stable/auto\\_examples/index.html](https://scikit-learn.org/stable/auto_examples/index.html)
- Phelan, Steven (2001). “What is Complexity Science, Really?”. *Emergence*, 3 (1), 120-136. doi: [https://doi.org/10.1207/S15327000EM0301\\_08](https://doi.org/10.1207/S15327000EM0301_08)
- Randriamihamison, N.; Vialaneix, N.; y Neuvial, P. (2021). “Applicability and Interpretability of Ward’s Hierarchical Agglomerative Clustering



- With or Without Contiguity Constraints”, *J Classif*, 38, 363-389. DOI: <https://doi.org/10.1007/s00357-020-09377-y>
- Simon, Herbert A. (1962). “The Architecture of Complexity”. *Proceedings of the American Philosophical Society*, 106 (6), 467-482.
- 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. DOI: <https://doi.org/10.22185/24487147.2019.100.12>
- Wolfram, Stephen (2002). *A New Kind of Science*. Estados Unidos: Wolfram Media.

